CME 305: Discrete Mathematics and Algorithms

Instructor: Professor Aaron Sidford (sidford@stanford.edu)

February 8, 2018

Lecture 11 - Spectral Graph Theory[1]

The central goals of Todays lecture is to prove Cheeger's Inequality, one of the most central and important results in spectral graph theory.

Theorem 1 (Cheeger's Inequality)**.** *For simple, undirected, positive weighted, connected G with maximum degree $d_{\max}$ it is the case that*

$$\frac{\sigma(G)^2}{2 \cdot d_{\max}} \leq \lambda_2(\mathcal{L}(G)) \leq 2 \cdot \sigma(G)$$

where $d_{\max}$ is the largest degree in the graph.

In the last class we proved the upper bound on $\lambda_2(\mathcal{L}(G))$. This known as the "easy direction" of Cheeger's inequality as it essentially follows from viewing λ_2 as a relaxation of the sparsest cut problem. This class we prove the "difficult direction" of Cheeger's inequality, the lower bound on λ_2. Importantly, we provide a constructive proof showing how to use vector with small Rayleigh quotient to find a sparse cut. In the next lecture we'll discuss computational aspects of this inequality as well as extensions.

1 Recap

In the remainder of this lecture our focus will be on a simple, undirected, positively weighted, connected graph $G = (V, E, w)$ with $w \in \mathbb{R}^E_{>0}$. As we noted, spectral graph theory for broader classes of matrices is possible, but this will provide a good taste of the field of spectral graph theory. Recall the following fundamental matrices associated with G, the (weighted) adjacency matrix $\mathbf{A}(G) \in \mathbb{R}^{V \times V}$, the (weighted) degree matrix $\mathbf{D}(G) \in \mathbb{R}^{V \times V}$, and the Laplacian matrix $\mathcal{L}(G) \in \mathbb{R}^{V \times V}$ each defined for all $i, j \in V$ by

$$\mathbf{A}(G)_{ij} \stackrel{\text{def}}{=} \begin{cases} w_{\{i,j\}} & \{i,j\} \in E \\ 0 & \text{otherwise} \end{cases} , \quad \mathbf{D}(G)_{ij} \stackrel{\text{def}}{=} \begin{cases} \deg(i) & i = j \\ 0 & \text{otherwise} \end{cases} , \quad \mathcal{L}(G)_{ij} = \stackrel{\text{def}}{=} \begin{cases} \deg(i) & i = j \\ -w_{\{i,j\}} & \{i,j\} \in E \\ 0 & \text{otherwise} \end{cases} .$$

Also recall that we saw last class that

$$\lambda_2(\mathcal{L}) = \min_{x \perp \vec{1}} R_{\mathcal{L}}(x) = \frac{x^\top \mathcal{L} x}{x^\top x} \text{ and } \sigma(G) = \min_{S \notin \{\emptyset, V\}} \phi(S) = \frac{w(\partial(S))}{\min\{|S|, |V \setminus S|\}} .$$

Finally, recall that we saw that $\sigma(S) \leq d_{\max}$ for all S and and $\sigma(G) \leq d_{\min}$ though there exists graphs with $\sigma(G) = \Omega(d_{\max})$. Consequently if we want to normalize Cheeger's inequality, we should divide through by $d_{\max}$ in which case we get

$$\frac{1}{2} \cdot \left(\frac{\sigma(G)}{d_{\max}}\right)^2 \leq \frac{\lambda_2(\mathcal{L}(G))}{d_{\max}} \leq 2 \cdot \frac{\sigma(G)}{d_{\max}}$$

and thus see that Cheeger's inequality let's us approximate $\sigma(G)/d_{\max}$ up to a multiplicative factor which gets worse as $\sigma(G)/d_{\max}$ decreases.

[1] These lecture notes are a work in progress and there may be typos, awkward language, omitted proof details, etc. Moreover, content from lectures might be missing. These notes are intended to converge to a polished superset of the material covered in class so if you would like anything clarified, please do not hesitate to post to Piazza and ask.

2 Another Perspective

Before we get to proving Cheeger's inequality, we provide a little more notation and machinery that lets us obtain another perspective on it. We begin with the following definition.

Definition 2 (Loewner order)**.** For symmetric matrices $\mathbf{A}, \mathbf{B}$ we say that $\mathbf{B}$ is larger than $\mathbf{A}$ *in the Loewner order*, denoted $\mathbf{A} \preceq \mathbf{B}$, if and only if $x^\top \mathbf{A} x \leq x^\top \mathbf{B} x$ for all x. We define $\succeq, \succ, \prec$ analogously.

There are many equivalent phrases we will use for this. For example, we'll say that $\mathbf{B}$ spectrally dominates $\mathbf{A}$, or $\mathbf{A}$ is spectrally less than or equal to $\mathbf{B}$, etc. Note that with this notation we have that $\mathbf{A}$ is PSD if and only if $\mathbf{A} \succeq \mathbf{0}$.

In the case of PSD matrices the Loewner order has a nice geometric notation. For PSD matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ let us define $E_\mathbf{A} \subseteq \mathbb{R}^n$ as follows

$$E_\mathbf{A} \stackrel{\text{def}}{=} \left\{x \in \mathbb{R}^n \,:\, x = z \cdot R_\mathbf{A}(z) \,,\, \|z\|_2 \leq 1\right\} .$$

It is not to hard to see that $E_\mathbf{A}$ is the interior of an ellipse where the axis are the eigenvectors and the radius of each axis is the eigenvalue. To see this, simply imagine performing the change of basis to make $\mathbf{A}$ diagonal and then consider the ellipse.

Now, as we saw last lecture $\mathcal{L}(K_n) = n\mathbf{I} - \vec{1}\vec{1}^\top$ where K_n is the complete graph on n vertices. Consequently,

$$\lambda_2(\mathcal{L}(G)) = \min_{x \perp \vec{1}} \frac{x^\top \mathcal{L} x}{x^\top x} = n \cdot \min_{x \perp \vec{1}} \frac{x^\top \mathcal{L}(G) x}{x^\top (\mathcal{L}(K_n)) x} .$$

This actually implies that $\lambda_2(\mathcal{L}(G))$ is the largest value of α such that $n\mathcal{L}(G) \succeq \alpha \mathcal{L}(K_n)$. In other words, λ_2 is a measure how exactly how much larger is $\mathcal{L}(G)$ then the complete graph spectrally.

So the Loewner ordering on Laplacians is one naturally way we could choose to compare graphs. Another natural way to compare would be in terms of cuts. Suppose we define $\preceq_{cut}$ by $\mathbf{A} \preceq_{cut} \mathbf{B}$ if and only if $x^\top \mathbf{A} x \leq x^\top \mathbf{B} x$ for all $x \in \{0,1\}^n$. Then we have that for two graph G and H on the same vertices it is the case that $\mathcal{L}(G) \preceq_{cut} \mathcal{L}(H)$ if and only if $w_G(\partial_G(S)) \leq w_H(\partial_H(S))$ for every set S, i.e. all cuts in H are larger in size then the corresponding cut in G. Now as we saw last class for some $c_G \in [\frac{1}{2}, 1]$

$$\sigma(G) = \min_{S \notin \{\emptyset, V\}} \frac{w(\partial(S))}{\min\{|S|, |V \setminus S|\}} = c_G n \cdot \min_{S \notin \{\emptyset, V\}} \frac{w(\partial(S))}{w(\partial(K_n))}$$

due to the fact that $\min\{|S|, |V \setminus S|\} \cdot \max\{|S|, |V \setminus S|\} = |S| \cdot |V \setminus S| = w(\partial(K_n))$ and $c_G = \max\{|S|, |V \setminus S|\}/n \in [\frac{1}{2}, 1]$. From this perspective we see that up to a mulitplicative factor between $\frac{1}{2}$ and 1 we have that $\sigma(G)$ is the largest value of α such that $n\mathcal{L}(G) \succeq_{cut} \alpha \mathcal{L}(K_n)$. In other words, $\sigma(G)$ is a measure how exactly how much larger is $\mathcal{L}(G)$ then the complete graph in terms of cuts.

Thinking about these bounds shows that Cheeger's inequality is a question about the relationship between spectral approximation of graph and the cut approximation of graphs. It is about connecting how much larger a graph is to the complete graph spectrally (i.e. λ_2) to how much larger a graph is than the complete graph in terms of cuts (i.e. $\sigma(G)$).

3 Proving Cheeger's Inequality

In the remainder of these notes our primary goal is to prove Cheeger's inequality. We assume that G is a fixed simple, undirected, positively weighted, connected graph $G = (V, E, w)$ with $w \in \mathbb{R}^E_{>0}$ and write

$\mathcal{L} = \mathcal{L}(G)$. For simplicity of notation we also assume that $V = [n]$. As we have discussed we have already proven that

$$\lambda_2(\mathcal{L}) \leq 2 \cdot \sigma(G)$$

which is known as the "easy direction." What remains is to prove the difficult direction

$$\lambda_2(\mathcal{L}) \geq \frac{1}{2} \cdot \frac{\sigma(G)^2}{d_{\max}}$$

where $d_{\max}$ is the largest weighted degree in the graph. Since

$$\lambda_2(\mathcal{L}(G)) = \min_{x \perp \vec{1}} R_{\mathcal{L}}(x)$$

it suffices to show that given $x \perp \vec{1}$ with $R_{\mathcal{L}}(x)$ small, we can use x to find a sparse cut. In fact, we will show something stronger, that given any $x \perp \vec{1}$ we can find a cut whose sparsity depends only on $R_{\mathcal{L}}(x)$ and $d_{\max}$.

In particular, we will show the following.

Theorem 3. *If $x \perp \vec{1}$ then there exists a non-trivial set $S \subseteq V$ with $\sigma(S) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$.*

The proof of this theorem is the main result of this section and we will prove it in steps. However, first we note that proving this suffices to prove Cheegers inequality.

Proof of Theorem 1 using Theorem 3. We showed that $\lambda_2(\mathcal{L}) \leq 2 \cdot \sigma(G)$ last lecture. Now, let v be such that $v \perp \vec{1}$ and $R_{\mathcal{L}}(v) = \lambda_2$. Such a v must exist by the variational characterization of eigenvalues and the fact that $\vec{1} \in \ker(\mathcal{L})$. Now, by Thoerem 3 we have that there exists a non-trivial set $S \subseteq V$ with $\sigma(S) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(v)} = \sqrt{2 \cdot d_{\max} \cdot \lambda_2}$. Since clearly $\sigma(G) \leq \sigma(S)$ we have $\lambda_2 \geq \frac{1}{2}\sigma(G)^2/d_{\max}$. □

Consequently, all that remains is to prove Theorem 3.

3.1 Sweep Cuts

So how should we prove Theorem 3? Note that this theorem says that given any vector $x \perp \vec{1}$ we can find a set S whose sparsity $\sigma(S)$ is upper bounded by $R_{\mathcal{L}}(x)$. Thus, to prove this theorem a natural strategy would be to use x to produce the cut and then from this procedure the sparsity of the cut would be tied to the Rayleigh quotient of x. So how should we use x to find the cut?

To think about this, let us consider the extreme case where if $x \in \{0, 1\}$ with more 0's then $1's$.[2] In this case, x is the indicator vector, $x = \vec{1}_S$, for some set S, with $|S| \leq V/2$. Consequently, we have that $R_{\mathcal{L}}(x) = w(\partial(S))/|S| = \sigma(S)$. Furthermore, if we simply sort the coordinates, into 0's and 1's and output the set consisting of all the 1's then we get a cut with sparsity exactly σ.

This suggests a more general strategy. We could sort the coordinates so that they are monotonically increasing, i.e. $x_1 \leq x_2 \leq ... \leq x_n$ and then consider the sets, $[1] = \{1\}$, $[2] = \{1, 2\}$, $[3] = \{1, 2, 3\}$, etc. These are known as the *sweep sets* and the cuts they induce are known as the sweep cuts.

On natural idea to prove Theorem 3 would then simply be to show the sweep cut with smallest sparsity is sufficiently sparse. In fact we will show this works and prove the following.

Theorem 4. *For $x \in \mathbb{R}^n$ with $x_1 \leq x_2 \leq ... \leq x_n$ with $x \perp \vec{1}$ it is the case that $\min_{i \in [n-1]} \sigma([i]) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$.*

[2] Note such an x is never perpendicular to $\vec{1}$ but this case will still be informative.

Note if we can prove Theorem 4 this immediately will prove Theorem 3.

Proof of Theorem 3 using Theorem 4. Given a vector $x \perp \vec{1}$ we can assume without loss of generality that $x_1 \leq x_2 \leq ... \leq x_n$ by simply re-naming the coordinates. Consequently, by Theorem 4 there exists $i \in [n-1]$ with $\sigma([i]) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$ as desired. Note that, $[i]$ for $i \in n-1$ is a non-trivial set, i.e. not $\emptyset$ or $V = [n]$. □

Consequently, it simply remains to prove Theorem 4.

As a quick aside, note that we can easily compute $\min_{i \in [n-1]} \sigma([i])$ in $O(|E|)$ time. As we considering the sweep cuts $\{1\}$, $\{1,2\}$, $\{1,2,3\}$, etc, each edge will start contributing to $w(\partial([i]))$ for some i and then stop at some j, thus if we simply keep track of the change in $w(\partial([i]))$ we can compute this quantity for all i in $O(|E|)$ time and therefore find the minimum $\sigma([i])$ in the same amount of time.

3.2 Proof Strategy

Note to prove Theorem 4 it suffices to show that there exists some $i \in [n-1]$ with $\sigma([i]) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$. Since the theorem only needs us to bound the minimum this is sufficient.

To do this we consider a slightly more restrictive class of the sweet sets. For all $c \in \mathbb{R}$ let

$$S_c(x) \stackrel{\text{def}}{=} \{i \in [n] \,:\, x_i \leq c\}$$

in other words, let $S_c(x)$ denote the coordinates of x which are at most some c.Note that when $x_1 \leq x_2 \leq ... \leq x_n$ each $S_c(x) = [i_c]$ for some $i_c \in [n]$ which depends on c and that $[i] = S_{c_i}(x)$ for some c_i depending on i whenever there isn't an $j > i$ with $x_j = x_i$. In other words, all the $S_c(x)$ are sweep sets and each sweep set is a $S_c(x)$ so long as it doesn't separate coordinates of the same value. Now, we will prove something even more restrictive then the Theorem 4 we will show that there exists a $S_c(x)$ with sufficient sparsity, i.e. we will prove the following.

Theorem 5. *For $x \in \mathbb{R}^n$ with $x_1 \leq x_2 \leq ... \leq x_n$ with $x \perp \vec{1}$ there exists $c \in (x_1, x_n)$ such that $\sigma(S_c(x)) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$.*

Note if we can prove Theorem 5 this immediately will prove Theorem 4.

Proof of Theorem 4 using Theorem 5. By Theorem 5 we have that there exists $c \in (x_1, x_n)$ such that $\sigma(S_c(x)) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$. Note that since $x \perp \vec{1}$ clearly $x_1 < x_n$ and since $c \in (x_1, x_n)$ we have $1 \in S_c(x)$ and $n \notin S_c(x)$ and therefore $S_c(x) = [i]$ for $i \in [n-1]$ giving the result. □

Consequently, it simply remains to prove Theorem 5. The way we will do this is through a version of the *probabilistic method* we will provide a distribution $\mathcal{D}$ over $c \in (x_1, x_n)$ so that

$$\Pr_{c \sim \mathcal{D}}\left[\frac{w(\partial(S_c(x)))}{\min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}\right] > 0\,.$$

If this happens, then we know there $\exists c \sim \mathcal{D}$ with $\sigma(S_c(x)) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$ as desired.

The way we will try to prove this is by showing the following.

$$\frac{\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(x)))}{\mathbb{E}_{c\sim\mathcal{D}} \min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}\,.$$

However, it is not immediately clear that this suffices. Why does bounding the ratio of expectations of two random variables allow us to the value of the ratio of the random variables. In particular, when the random variables could be correlated, this seems tricky to reason about. Fortunately, the following very simple lemma shows that reasoning about the ratio of expectations is sufficient.

Lemma 6. *Let X and Y be random real valued random variables where $Y > 0$ with probability 1 with then* $\Pr[\frac{X}{Y} \leq \frac{\mathbb{E}X}{\mathbb{E}Y}] > 0$.

Proof. Let $r = \frac{\mathbb{E}X}{\mathbb{E}Y}$. We have $\mathbb{E}[Yr - X] \geq 0$ and thus $\Pr[Yr - X \geq 0] > 0$ which yields the result as $Y > 0$. □

From this lemma we see that if we can bound the ratio of the expectations of the random variables then there is a realization of the random variables where the ratio is comparably bounded. Putting this all together, we can further reduce our problem to showing the following theorem.

Theorem 7. *For $x \in \mathbb{R}^n$ with $x_1 \leq x_2 \leq ... \leq x_n$ with $x \perp \vec{1}$ there exists a distribution $\mathcal{D}$ over $c \in (x_1, x_n)$ such that*

$$\frac{\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(x)))}{\mathbb{E}_{c\sim\mathcal{D}} \min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}\,.$$

Note if we can prove Theorem 7 this immediately will prove Theorem 5.

Proof of Theorem 5 using Theorem 7. By Theorem 7 we have that there exists a distribution $\mathcal{D}$ over $c \in (x_1, x_n)$ such that

$$\frac{\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(x)))}{\mathbb{E}_{c\sim\mathcal{D}} \min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}\,.$$

By Lemma 6 we have

$$\Pr_{c\sim\mathcal{D}}\left[\frac{w(\partial(S_c(x)))}{\min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}\right] > 0\,.$$

and therefore there is some $c \in (x_1, x_n)$ such that $\sigma(S_c(x)) \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(x)}$. □

Consequently, it simply remains to prove Theorem 7.

3.3 Some Preprocessing

We are nearly ready the prove Cheeger's inequality. All that we need to do is produce our distribution $\mathcal{D}$ over $[x_1, x_n]$ and bound how well it does. However, to simplify our analysis we perform some simple modifications to x that we can do without loss of generality that make our analysis easier. Instead of working with $x \perp \vec{1}$ as typically stated we will scale and shift x, i.e. work with $y = \alpha(x - \beta\vec{1})$ chosen so that $y_{\lceil \frac{n}{2} \rceil} = 0$ and $y_1^2 + y_n^2 = 1$. The reason for these conditions will be made clear in the next section. Here we show formally that it suffices to work with such a y vector. In particular, we show that if suffices to prove the following Thoerem.

Theorem 8. *For $y \in \mathbb{R}^n$ with $y_1 \leq y_2 \leq ... \leq y_n$ with $y_{\lceil \frac{n}{2} \rceil} = 0$ and $y_1^2 + y_n^2 = 1$ there exists a distribution $\mathcal{D}$ over $c \in (y_1, y_n)$ such that*

$$\frac{\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(y)))}{\mathbb{E}_{c\sim\mathcal{D}} \min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(y)}\,.$$

In the remainder of this subsection we show that if we can prove Theorem 8 this immediately will prove Theorem 7.

Proof of Theorem 7 using Theorem 8. Given $x \perp \vec{1}$ let $y = \alpha(x - \beta\vec{1})$ for $\alpha > 0$ and β chosen so that $y_{\lceil \frac{n}{2} \rceil} = 0$ and $y_1^2 + y_n^2 = 1$. Note that β can be chosen always to make $y_{\lceil \frac{n}{2} \rceil} = 0$ hold. Furthermore, since for $x \perp \vec{1}$ we have $x - \beta\vec{1} \neq 0$ for all β it is the case that $\alpha > 0$ can be then chosen to make $y_1^2 + y_n^2 = 1$. Now, note that for this vector y it is still the case that $y_1 \leq y_2 \leq ... \leq y_n$ since we have only shifted coordinates and then scaled them by a positive amount. Furthermore we see that

$$R_{\mathcal{L}}(y) = \frac{y^\top \mathcal{L} y}{y^\top y} = \frac{\alpha \cdot x^\top \mathcal{L} x}{\alpha \cdot \left[x^\top x - 2\beta x^\top \vec{1} + \beta^2 \vec{1}^\top \vec{1}\right]} = \frac{x^\top \mathcal{L} x}{x^\top x + \beta^2 n} \leq R_{\mathcal{L}}(x)$$

where we used that $x \perp \vec{1}$. Now, by Theorem 8 there is a distribution $\mathcal{D}$ over $c \in (y_1, y_n)$ such that

$$\frac{\mathbb{E}_{c \sim \mathcal{D}} w(\partial(S_c(y)))}{\mathbb{E}_{c \sim \mathcal{D}} \min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(y)}$$

but since y was obtained just by scaling and shifting x there is a corresponding distribution D' on $c \in (x_1, x_n)$ with

$$\frac{\mathbb{E}_{c \sim \mathcal{D}'} w(\partial(S_c(x)))}{\mathbb{E}_{c \sim \mathcal{D}'} \min\{|S_c|, |V \setminus S_c|\}} \leq \sqrt{2 \cdot d_{\max} \cdot R_{\mathcal{L}}(y)} \leq \sqrt{2 \cdot d_{\max} \cdot R_L(x)}$$

□

Consequently, it simply remains to prove Theorem 8.

3.4 The Distribution

All the remains is to produce $\mathcal{D}$ and prove Theorem 8. We will pick $\mathcal{D}$ so that

$$\mathbb{E}_{c \sim \mathcal{D}} \min\{|S_c|, |V \setminus S_c|\} = y^\top y = \sum_{i \in [n]} y_i^2 \,.$$

However, note that

$$\mathbb{E}_{c \sim \mathcal{D}} \min\{|S_c|, |V \setminus S_c|\} = \sum_{i \in [n]} \Pr\left[i \text{ is in the smaller of } S_c \text{ and } V \setminus S_c\right] .$$

However, , the reason we performed the preprocessing to make $y_{\lceil \frac{n}{2} \rceil} = 0$ is so we can easily reason about whether or not i is the smaller of S_c of $V \setminus S_c$. If $c \leq \lceil \frac{n}{2} \rceil$ then $|S_c| \leq |V \setminus S_c|$ and if $c \geq \lceil \frac{n}{2} \rceil$ then $|S_c| \geq |V \setminus S_c|$ and consequently we have that i is in the smaller of S_c or $V \setminus S_c$ if and only if c and y_i have the same sign with $|y_i| \geq |c|$. Therefore if we let $p : (x_1, x_n) \to [0, 1]$ denote the probability density of $\mathcal{D}$ and suppose that $p(x) = p(-x)$ for all x defined we have that

$$\Pr\left[i \text{ is in the smaller of } S_c \text{ and } V \setminus S_c\right] = \int_0^{|y_i|} p(c) dc$$

and therefore if we wan the right-hand side to be y_i^2 we should pick $p(c) = 2|c|$. However, this would not necessarily imply the p is a valid density function since it might not be normalized. However since $y_1 \leq 0 \leq y_n$ we have

$$\int_{y_1}^{y_n} 2|c| dc = \int_0^{y_1} 2|c| dc + \int_0^{y_n} 2|c| dc = y_1^2 + y_n^2 = 1$$

and therefore $p(c) = 2|c|$ is a probability distribution precisely by our normalization that $y_1^2 + y_n^2 = 1$ (which is why we chose this normalization).

We now have everything we need to prove Theorem 8 and thereby prove Cheeger's inequality is well. We let p be the pdf of $\mathcal{D}$ and prove this works below.

Proof of Theorem 8. Let $p(c) = 2|c|$ be the pdf of the distribution $\mathcal{D}$ as we discussed in the previous section we have that

$$\mathbb{E}_{c\sim\mathcal{D}} \min\{|S_c|, |V \setminus S_c|\} = \sum_{i\in[n]} \Pr\left[i \text{ is in the smaller of } S_c \text{ and } V \setminus S_c\right]$$
$$= \sum_{i\in[n]} \int_0^{|y_i|} p(c)dc = \sum_{i\in[n]} \int_0^{|y_i|} 2|c|dc = \sum_{i\in[n]} y_i^2 = y^\top y\,.$$

Next, we just need to bound $\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(y)))$. Clearly,

$$\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(y))) = \sum_{\{i,j\}\in E} w_{ij} \Pr\left[\{i,j\} \in \partial(S_c)\right]$$

and therefore we just need to upper bound the probability and edge $\{i,j\}$ is in the cut S_c. We analyze this in two cases. First suppose that y_i and y_j have the same sign and $|y_i| \leq |y_j|$. In this case we have that $\{i,j\}$ is in the cut S_c when c has the same sign as y_i and y_j and has $|y_i| < c < |y_j|$. Therefore,

$$\Pr\left[\{i,j\} \in \partial(S_c)\right] = \int_{|y_i|}^{|y_j|} 2|c|dc = 2\left[y_j^2 - y_i^2\right] = |y_i - y_j| \cdot (|y_i| + |y_j|)\ .$$

In the second case suppose that y_i and y_j have the different signs and $y_i \leq y_j$. In this case we have $\{i,j\}$ is in the cut S_c when $c \in (y_i, y_j)$ and therefore

$$\Pr\left[\{i,j\} \in \partial(S_c)\right] = \int_{|y_i|}^{|y_j|} 2|c|dc = \int_0^{|y_i|} 2|c|dc + \int_0^{|y_j|} 2|c|dc = y_i^2 + y_j^2 \leq |y_i - y_j| \cdot (|y_i| + |y_j|)\ .$$

Consequently, we have that by Cauchy Schwarz

$$\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(y))) = \sum_{\{i,j\}\in E} w_{ij} \Pr\left[\{i,j\} \in \partial(S_c)\right] \leq \sum_{\{i,j\}\in E} w_{ij} \left(|y_i - y_j| \cdot (|y_i| + |y_j|)\right)$$
$$\leq \sqrt{\sum_{\{i,j\}\in E} w_{ij}(y_i - y_j)^2 \cdot \sum_{\{i,j\}\in E} w_{ij} \left(|y_i| + |y_j|\right)^2}\ .$$

Now, since $(x+y)^2 \leq 2x^2 + 2y^2$ we have that

$$\sum_{\{i,j\}\in E} w_{ij} \left(|y_i| + |y_j|\right)^2 \leq 2 \sum_{\{i,j\}\in E} w_{ij} \left(y_i^2 + y_j^2\right) = 2 \sum_{i\in[n]} \deg(i) y_i^2 \leq 2 \cdot d_{\max} \cdot y^\top y\,.$$

Combining our two bounds we have that

$$\frac{\mathbb{E}_{c\sim\mathcal{D}} w(\partial(S_c(y)))}{\mathbb{E}_{c\sim\mathcal{D}} \min\{|S_c|, |V \setminus S_c|\}} \leq \frac{\sqrt{y^\top \mathcal{L} y \cdot 2 \cdot d_{\max} \cdot y^\top y}}{y^\top y} = \sqrt{2 d_{\max} R_{\mathcal{L}}(y)}\,.$$

□

Now that we have proven Theorem 8, this implies Theorem 7 is true, which in turn implies Theorem 5, which in turn implies Theorem 4, which in turn implies Theorem 3, which in turn proves Cheeger's inequality, Theorem 1.

In the next class we will discuss a bit more computational aspects of Cheeger's inequality and extensions.