Graph Theory

Spectral clustering

Instructor: *Nicolò Cesa-Bianchi* version of April 13, 2022

The material in this handout is mostly taken from: Luca Trevisan, Lecture Notes on Graph Partitioning, Expanders and Spectral Methods, 2016.

Intuitively, a graph is clusterable if its vertices can be partitioned (in a non-trivial way) so that the number of edges across the elements of the partition is small. A key notion is therefore that of cut between disjoint subsets of vertices. We study the clusterability of a graph through the algebraic properties of its adiacency matrix.

Given two disjoints subsets S, T of vertices of a graph $G = (V, E)$, let $E(S, T)$ be the set of edges having one endpoint in S and one endpoint in T. Also, let $\neg S = V \setminus S$.

A cut is any partition $(S, \neg S)$ such that $S \not\equiv V$ and $S \not\equiv \emptyset$. The sparsity of a cut $(S, \neg S)$ is

$$\sigma(S) = \frac{\bigl|E(S, \neg S)\bigr|}{|S|\,|\neg S|}$$

This the fraction of edges in the cut among all potential edges between the two subset of vertices. Note that sparsity is small when there are few edges in the cut and the cut is balanced (that is S and $\neg S$ have about the same size). The sparsity of a graph is the sparsity of the sparsest cut,

$$\sigma(G) = \min_{S\,:\,(S,\neg S)\text{ is a cut}} \sigma(S)$$

In what follows, we focus on d-regular graphs for simplicity. In a d-regular graph, the **expansion** of a set $S \subseteq V$ is defined by

$$\mathrm{xpn}(S) = \frac{\bigl|E(S, \neg S)\bigr|}{d|S|}$$

This is the fraction of edges in the cut among all all potential edges leaving the set S. Note that, unlike sparsity, expansion is not symmetric. The **conductance** of a cut $(S, \neg S)$ is the symmetrized version of expansion and is defined by

$$\phi(S) = \max\bigl\{\mathrm{xpn}(S), \mathrm{xpn}(\neg S)\bigr\} = \frac{\bigl|E(S, \neg S)\bigr|}{d\,\min\bigl\{|S|, |\neg S|\bigr\}}$$

Finally, the conductance of a graph is given by

$$\phi(G) = \min_{S \subseteq V} \phi(S) \tag{1}$$

As $\alpha(1-\alpha) \le \min\{\alpha, 1-\alpha\} \le 2\alpha(1-\alpha)$ for all $\alpha \in [0,1]$, we have

$$\frac{1}{n}|S|\,|\neg S| \le \min\bigl\{|S|\,|\neg S|\bigr\} \le \frac{2}{n}|S|\,|\neg S|\ .$$

Therefore, $\phi(S) \le (n/d)\sigma(S) \le 2\phi(S)$ for all cuts $(S, \neg S)$, and we have

$$\phi(G) \le \frac{n}{d}\sigma(G) \le 2\phi(G) \ .$$

Hence, in d-regular graphs, finding the sparsest cut is equivalent—up to a factor of two—to finding a cut minimizing conductance. We now study the relationships between conductance and algebraic properties of the adjacency matrix.

Laplacian matrix. The Laplacian matrix of a d-regular graph $G = (V, E)$ is the symmetric matrix $L = I - \frac{1}{d}A$, where A is the adiacency matrix with entries $A_{i,j} = \mathbb{I}\{(i,j) \in E\}$. For any $\boldsymbol{x} \in \mathbb{R}^n$, we have that

$$\begin{aligned}
\boldsymbol{x}^\top L\boldsymbol{x} &= \sum_{i\in V} x_i^2 - \frac{1}{d}\sum_{i\in V}\sum_{j\in V} A_{i,j}x_i x_j \\
&= \frac{1}{d}\sum_{i\in V}\sum_{j\,:\,(i,j)\in E} x_i^2 - \frac{1}{d}\sum_{i\in V}\sum_{j\,:\,(i,j)\in E} x_i x_j \\
&= \frac{1}{d}\sum_{i\in V}\sum_{j\,:\,(i,j)\in E} \left(x_i^2 - x_i x_j\right) \\
&= \frac{1}{d}\sum_{(i,j)\in E} \left(x_i^2 + x_j^2 - 2x_i x_j\right) \\
&= \frac{1}{d}\sum_{(i,j)\in E} (x_i - x_j)^2
\end{aligned}$$

Therefore, the Laplacian matrix is positive semidefinite.

Since the rows and columns of L sum to zero (verify that),

$$\lambda_1 = \min_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{u}^\top L\boldsymbol{u}}{\boldsymbol{u}^\top\boldsymbol{u}} = 0$$

where the minimum is attained by $\boldsymbol{u} = \boldsymbol{1}$, where we write $\boldsymbol{1} = (1, \dots, 1)$. Hence, $\boldsymbol{u}_1 = \frac{1}{\sqrt{n}}\boldsymbol{1}$ is the eigenvector of λ_1, while the remaining eigenvalues of L are all nonnegative because L is positive semidefinite. Note also that any other eigenvector $\boldsymbol{u}_i$ of L with $i > 1$ is such that $\boldsymbol{u}_i^\top\boldsymbol{u}_1 = 0$. Therefore, $\boldsymbol{u}_i^\top\boldsymbol{1} = 0$ for all $i = 2, \dots, n$.

This helps us characterize λ_2,

$$\lambda_2 = \min_{\substack{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\} \\ \boldsymbol{u}^\top\boldsymbol{1}=0}} \frac{\boldsymbol{u}^\top L\boldsymbol{u}}{\boldsymbol{u}^\top\boldsymbol{u}} = \min_{\substack{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\} \\ \boldsymbol{u}^\top\boldsymbol{1}=0}} \frac{\sum_{(i,j)\in E}(u_i - u_j)^2}{d\sum_{i\in V} u_i^2} \ .$$

If $G = (V, E)$ has two connected components $X, Y \subset V$, then we can choose $\boldsymbol{u} \in \mathbb{R}^n$ such that $u_i = 1/|X|$ for all $i \in X$ and $u_j = -1/|Y|$ for all $j \in Y$. This ensures that $\boldsymbol{u}^\top\boldsymbol{1} = 0$. Moreover, $(i,j) \in E$ if and only if $(u_i - u_j)^2 = 0$. So $\boldsymbol{u}^\top L\boldsymbol{u} = 0$ and therefore $\boldsymbol{u}/\sqrt{2}$ is an eigenvector with eigenvalue $\lambda_2 = 0$. More generally, it can be proven that $\lambda_k = 0$ if and only if G has k connected components.

We now look at the largest eigenvalue,

$$\begin{aligned}
\lambda_n &= \max_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{u}^\top L\boldsymbol{u}}{\boldsymbol{u}^\top\boldsymbol{u}} \\
&= \max_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{\sum_{(i,j)\in E}(u_i-u_j)^2}{d\sum_{i\in V} u_i^2} \\
&= \max_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{d\sum_{i\in V} u_i^2 - \sum_{(i,j)\in E} 2u_iu_j}{d\sum_{i\in V} u_i^2} \\
&= \max_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{2d\sum_{i\in V} u_i^2 - d\sum_{i\in V} u_i^2 - \sum_{(i,j)\in E} 2u_iu_j}{d\sum_{i\in V} u_i^2} \\
&= \max_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{2d\sum_{i\in V} u_i^2 - \sum_{(i,j)\in E}(u_i+u_j)^2}{d\sum_{i\in V} u_i^2} \\
&= 2 - \min_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}} \frac{\sum_{(i,j)\in E}(u_i+u_j)^2}{d\sum_{i\in V} u_i^2} \ .
\end{aligned}$$

So $\lambda_n \le 2$ and $\lambda_n = 2$ if G has at least a bipartite component (X,Y). Indeed, in this case we can pick $\boldsymbol{u}\in\mathbb{R}^n$ such that $u_i = 1$ for all $i\in X$, $u_j = -1$ for all $j\in Y$, and $u_k = 0$ for all remaining k. Then $(u_i+u_j)^2 = 0$ for all $(i,j)\in E$, and so $\boldsymbol{u}/\left\|\boldsymbol{u}\right\|$ is an eigenvector of G with eigenvalue 2.

Cheeger's inequalities. This important inequality connects the second eigenvalue with the conductance, revealing the importance of λ_2 in clustering,

$$\frac{\lambda_2}{2} \le \phi(G) \le \sqrt{2\lambda_2} \ .$$

We begin by proving the first inequality. From now on we write $\sum_{i=1}^n$ instead of $\sum_{i\in V}$.

Lemma 1 *For any connected and d-regular graph G, $\lambda_2 \le 2\phi(G)$.*

PROOF. We start noticing that, for any $\boldsymbol{u}\in\mathbb{R}^d$ such that $\boldsymbol{u}^\top\boldsymbol{1} = 0$,

$$\sum_{i=1}^n\sum_{j=1}^n (u_i-u_j)^2 = 2n\sum_{i=1}^n u_i^2 - 2\sum_{i=1}^n\sum_{j=1}^n u_iu_j = 2n\sum_{i=1}^n u_i^2 - 2\left(\sum_{i=1}^n u_i\right)^2 = 2n\sum_{i=1}^n u_i^2 \tag{2}$$

Therefore, we have that

$$\begin{aligned}
\lambda_2 &= \min_{\substack{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}\\ \boldsymbol{u}^\top\boldsymbol{1}=0}} \frac{\sum_{(i,j)\in E}(u_i-u_j)^2}{d\sum_{i=1}^n u_i^2} \\
&= \min_{\substack{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\}\\ \boldsymbol{u}^\top\boldsymbol{1}=0}} \frac{\sum_{(i,j)\in E}(u_i-u_j)^2}{\frac{d}{2n}\sum_{i=1}^n\sum_{j=1}^n (u_i-u_j)^2} && (3)\\
&= \min_{\boldsymbol{u}\in\mathbb{R}^n\setminus\{\boldsymbol{0},\boldsymbol{1}\}} \frac{\sum_{(i,j)\in E}(u_i-u_j)^2}{\frac{d}{2n}\sum_{i=1}^n\sum_{j=1}^n (u_i-u_j)^2} && (4)
\end{aligned}$$

To understand the last equality: if $\boldsymbol{u}$ satisfies $\boldsymbol{u} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $\boldsymbol{u}^\top \boldsymbol{1} = 0$, then it also satisfies $\boldsymbol{u} \notin \{\boldsymbol{0}, \boldsymbol{1}\}$, and so (4) is not larger than (3). Vice versa, if $\boldsymbol{u} \notin \{\boldsymbol{0}, \boldsymbol{1}\}$, then $\boldsymbol{u}'$ defined by $u'_i = u_i - \frac{1}{n}\sum_j u_j$ satisfies $\boldsymbol{u} \neq \boldsymbol{0}$ and $\boldsymbol{u}^\top \boldsymbol{1} = 0$. Hence, the value of (3) is not larger than (4) because the shift by $\frac{1}{n}\sum_j u_j$ cancels out in the numerator and the denominator of the obective function.

For any $S \subseteq V$, let $\boldsymbol{u} \in \{0,1\}^n$ be the incidence vector of the set S, that is $u_i = \mathbb{I}\{i \in S\}$ for $i = 1, \ldots, n$. Then $\big|E(S, \neg S)\big| = \sum_{(i,j)\in E}(u_i - u_j)^2$. Also, using $u_i = u_i^2$ for all i and an argument similar to the one used in (2),

$$|S|\,|\neg S| = \left(\sum_{i=1}^n u_i^2\right)\left(n - \sum_{j=1}^n u_j^2\right) = n\sum_{i=1}^n u_i^2 - \sum_{i=1}^n\sum_{j=1}^n u_i u_j = \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n (u_i - u_j)^2$$

Therefore,

$$\sigma(G) = \min_{S \subsetneq V \,:\, S \not\equiv \emptyset} \frac{\big|E(S,\neg S)\big|}{|S|\,|\neg S|} = \min_{\boldsymbol{u}\in\{0,1\}^n\setminus\{\boldsymbol{0},\boldsymbol{1}\}} \frac{\sum_{(i,j)\in E}(u_i - u_j)^2}{\frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n (u_i - u_j)^2}$$

which implies $\frac{d}{n}\lambda_2 \le \sigma(G)$. Since $\sigma(G) \le \frac{2d}{n}\phi(G)$, the proof is concluded. $\square$

The proof of the second inequality of Cheeger is based on the analysis of Fiedler's algorithm, the simplest algorithm for spectral clustering. The algorithm finds a cut of small conductance by looking at the $n-1$ cuts induced by the ranked components of the input vector $\boldsymbol{x}$. As we see in the analysis, the algorithm works well when $\boldsymbol{x}$ is the eigenvector of λ_2.

Algorithm 1 (Fiedler)

Input: Graph $G = (V, E)$, vector $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$.

1: Sort V according to the values x_v and let $v_1, \ldots, v_n$ be the vertices of V in the sorted order

2: Find $k \in \{1, \ldots, n-1\}$ minimizing the conductance $\phi\big(\{v_1, \ldots, v_k\}\big)$

Output: k

Note that Fiedler's algorithm can be implemented in time $\mathcal{O}(|E| + |V|\ln|V|)$, because it takes time $\mathcal{O}(|V|\ln|V|)$ to sort the vertices, and the cut of minimal expansion that respects the sorted order can be found in time $\mathcal{O}(|E|)$. On the other hand, minimizing conductance over all exponentially many cuts $S \subset V$ is NP-hard. Cheeger's inequalities imply that the conductance of the cut S_F found by Fiedler algorithm satisfies

$$\phi(S_F) \le \sqrt{\frac{8}{\lambda_2}}\phi(G)$$

We move on to the analysis of the algorithm, which gives us the second inequality of Cheeger as an immediate consequence. Let

$$R_L(\boldsymbol{x}) = \frac{\sum_{(i,j)\in E}(x_i - x_j)^2}{d\sum_{i=1}^n x_i^2}$$

be the Rayleigh quotient for L evaluated at $\boldsymbol{x} \in \mathbb{R}^n$, and recall that

$$\lambda_2 = \min_{\substack{\boldsymbol{x}\in\mathbb{R}^n\setminus\{\boldsymbol{0}\} \\ \boldsymbol{x}^\top\boldsymbol{1}=0}} R_L(\boldsymbol{x})$$

We now prove the following result, which implies $\phi(G) \leq \sqrt{2\lambda_2}$.

Theorem 2 *Let $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ be such that $\boldsymbol{x}^\top \mathbf{1} = 0$, and let $S_F \subset V$ be the cut found by Fiedler's algorithm with input $\boldsymbol{x}$. Then $\phi(S_F) \leq \sqrt{2R_L(\boldsymbol{x})}$.*

Indeed, when the input $\boldsymbol{x}$ is the eigenvector of λ_2, using (1) we get that

$$\phi(G) = \min_{S \subseteq V} \phi(S) \leq \phi(S_F) \leq \sqrt{2\lambda_2}$$

In order to prove Theorem 2, we need to prove two auxiliary lemmas first.

Lemma 3 *Let $\boldsymbol{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ be such that $\boldsymbol{x}^\top \mathbf{1} = 0$. Then there exists a nonnegative vector $\boldsymbol{y}$ such that $R_L(\boldsymbol{y}) \leq R_L(\boldsymbol{x})$. Furthermore, for every $0 < t \leq \max_{v \in V} y_v$, the cut*

$$\big(\{v \in V \,:\, y_v \geq t\}\,,\, \{v \in V \,:\, y_v < t\} \big)$$

is one of the cuts considered in line 2 of Fiedler's algorithm on input $\boldsymbol{x}$.

Proof. Let m be the median value of the entries of $\boldsymbol{x}$. Let $\boldsymbol{x}^+, \boldsymbol{x}^-$ have components $x_v^+ = \big[x_v - m\big]_+$ and $x_v^- = \big[m - x_v\big]_+$, where $[z]_+ = z\,\mathbb{I}\{z > 0\}$. Note that $\boldsymbol{x}^+, \boldsymbol{x}^-$ are both nonnegative. Now, for every $t > 0$,

$$\big\{v \in V \,:\, x_v^+ \geq t\big\} = \Big\{v \in V \,:\, \big[x_v - m\big]_+ \geq t\Big\} = \{v \in V \,:\, x_v \geq m + t\}$$

is one of the cuts considered by Fiedler's algorithm on input $\boldsymbol{x}$. Similarly, for every $t > 0$,

$$\big\{v \in V \,:\, x_v^- \geq t\big\} = \Big\{v \in V \,:\, \big[m - x_v\big]_+ \geq t\Big\} = \{v \in V \,:\, x_v \leq m - t\}$$

is also one of the cuts considered by Fiedler's algorithm on input $\boldsymbol{x}$. It remains to show that $R_L(\boldsymbol{y}) \leq R_L(\boldsymbol{x})$ for some nonnegative $\boldsymbol{y} \in \mathbb{R}^n$. We set

$$\boldsymbol{y} = \operatorname*{argmin}_{\boldsymbol{z} \in \{\boldsymbol{x}^+, \boldsymbol{x}^-\}} R_L(\boldsymbol{z})$$

Let $\boldsymbol{x}' = \boldsymbol{x} - m\,\mathbf{1} = \boldsymbol{x}^+ - \boldsymbol{x}^-$ and observe that, for every constant c, $R_L(\boldsymbol{x} + c\,\mathbf{1}) \leq R_L(\boldsymbol{x})$. Indeed, the numerator of $R_L(\boldsymbol{x}+c\,\mathbf{1})$ and the numerator of $R_L(\boldsymbol{x})$ are the same. Moreover, the denominator of $R_L(\boldsymbol{x} + c\,\mathbf{1})$ is $\|\boldsymbol{x} + c\,\mathbf{1}\|^2 = \|\boldsymbol{x}\|^2 + \|c\,\mathbf{1}\|^2 \geq \|\boldsymbol{x}\|^2$. Therefore $R_L(\boldsymbol{x}') \leq R_L(\boldsymbol{x})$ and we are left to show that $R_L(\boldsymbol{y}) \leq R_L(\boldsymbol{x}')$. To this end we write

$$\begin{aligned}
R_L(\boldsymbol{y}) &= \min\big\{R_L(\boldsymbol{x}^+), R_L(\boldsymbol{x}^-)\big\} \\
&\leq \frac{\|\boldsymbol{x}^+\|^2 R_L(\boldsymbol{x}^+) + \|\boldsymbol{x}^-\|^2 R_L(\boldsymbol{x}^-)}{\|\boldsymbol{x}^+\|^2 + \|\boldsymbol{x}^-\|^2} && \text{(using } \min\{a,b\} \leq \alpha a + (1-\alpha)b\text{)} \\
&= \frac{\sum_{(i,j)\in E}(x_i^+ - x_j^+)^2 + \sum_{(i,j)\in E}(x_i^- - x_j^-)^2}{\|\boldsymbol{x}^+\|^2 + \|\boldsymbol{x}^-\|^2} \\
&\leq \frac{\sum_{(i,j)\in E}\Big((x_i^+ - x_j^+) - (x_i^- - x_j^-)\Big)^2}{\|\boldsymbol{x}^+\|^2 + \|\boldsymbol{x}^-\|^2} && \text{(this is shown below)} \\
&= \frac{\sum_{(i,j)\in E}(x_i' - x_j')^2}{\|\boldsymbol{x}'\|^2} && \text{(using } \boldsymbol{x}' = \boldsymbol{x}^+ + \boldsymbol{x}^- \text{ and } (\boldsymbol{x}^+)^\top \boldsymbol{x}^- = 0\text{)} \\
&= R_L(\boldsymbol{x}')
\end{aligned}$$

To finish the proof, we need to verify that for each $(i,j) \in E$,

$$(x_i^+ - x_j^+)^2 + (x_i^- - x_j^-)^2 \leq \Big((x_i^+ - x_j^+) - (x_i^- - x_j^-)\Big)^2 \tag{5}$$

If $(x_i^+ - x_j^+)(x_i^- - x_j^-) = 0$, then (5) holds with equality. This happens when $x_i^+ = x_j^+ = 0$ or $x_i^- = x_j^- = 0$. Equivalently, when $x'_i, x'_j \geq 0$ or $x'_i, x'_j \leq 0$.

Now assume $(x_i^+ - x_j^+)(x_i^- - x_j^-) \neq 0$ and, without loss of generality, assume that $x'_i < 0 < x'_j$. Then $x_i^+ = 0$ and $x_j^- = 0$, so (5) becomes $(x_j^+)^2 + (x_i^-)^2 \leq \big(x_j^+ + x_i^-\big)^2$, which is clearly true. This concludes the proof. □

The following observation is used in the proof of the next lemma.

Fact 4 *For all random variables X, Y such that $Y > 0$ and $\mathbb{E}[X], \mathbb{E}[Y] < \infty$,*

$$\mathbb{P}\left(\frac{X}{Y} \leq \frac{\mathbb{E}[X]}{\mathbb{E}[Y]}\right) > 0$$

Proof. Let $r = \mathbb{E}[X]/\mathbb{E}[Y]$. Because of linearity of expectation, $\mathbb{E}\big[X - rY\big] = 0$. Since the expected value is zero, the random variable $X - rY$ must be nonpositive with probability bigger than zero, $\mathbb{P}\big(X - rY \leq 0\big) > 0$. Dividing both sides of $X - rY \leq 0$ by $Y > 0$, we get the desired result. □

We are now ready to prove the second auxiliary lemma.

Lemma 5 *For all nonnegative vectors $\boldsymbol{y} \in \mathbb{R}^n$ there exists $0 < t \leq \max_v y_v$ such that*

$$\operatorname{xpn}\big(\,\{v \in V \,:\, y_v \geq t\}\,\big) \leq \sqrt{2R_L(\boldsymbol{y})}$$

Proof. Since rescaling does not affect the Rayleigh quotient, we may assume $\max_v y_v = 1$. The proof uses the probabilistic method. Let t be a random variable such that $\mathbb{P}\big(t \leq \sqrt{a}\big) = a$, which means that t^2 is uniformly distributed in $[0,1]$, and define $S_t = \{v \in V \,:\, y_v \geq t\}$. Note that $|S_t|$ is positive for all $t \in [0,1]$. Note also that $\mathbb{P}\big(t \leq 0\big) = 0$, so that $t > 0$. Therefore, we can write

$$\begin{aligned}
\operatorname{xpn}\big(\,\{v \in V \,:\, y_v \geq t\}\,\big) &= \frac{\big|E(S_t, \neg S_t)\big|}{d|S_t|} && \text{(by definition of xpn and } S_t\text{)} \\
&\leq \frac{\mathbb{E}\Big[\big|E(S_t, \neg S_t)\big|\Big]}{d\,\mathbb{E}\big[|S_t|\big]} && \text{(with probability} > 0\text{, by Fact 4)}
\end{aligned}$$

This implies that there exists some $t \in (0,1]$ such that the above holds. To conclude the proof, we show that

$$\frac{\mathbb{E}\Big[\big|E(S_t, \neg S_t)\big|\Big]}{d\,\mathbb{E}\big[|S_t|\big]} \leq \sqrt{2R_L(\boldsymbol{y})}$$

We start to bound the denominator. Using that t is uniformly distributed in $[0,1]$,

$$E\big[|S_t|\big] = \sum_{i=1}^{n} \mathbb{P}(i \in S_t) = \sum_{i=1}^{n} \mathbb{P}(t \leq y_i) = \sum_{i=1}^{n} y_i^2 \tag{6}$$

Now pick any $(i,j) \in E$ and assume $y_j \le y_i$. Then

$$\begin{aligned}
\mathbb{P}\big(i \in S_t,\, j \in \neg S_t\big) &= \mathbb{P}\big(y_j < t \le y_i\big) \\
&= \Big(\mathbb{P}(t \le y_i) - \mathbb{P}(t \le y_j)\Big) \\
&= y_i^2 - y_j^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\Big[\big|E(S_t, \neg S_t)\big|\Big] &= \sum_{(i,j)\in E} \Big(\big(y_i^2 - y_j^2\big)\mathbb{I}\{y_j \le y_i\} + \big(y_j^2 - y_i^2\big)\mathbb{I}\{y_i \le y_j\}\Big) \\
&= \sum_{(i,j)\in E} \big|y_i^2 - y_j^2\big| \\
&= \sum_{(i,j)\in E} \big|y_i - y_j\big|(y_i + y_j) \\
&\le \sqrt{\sum_{(i,j)\in E} (y_i - y_j)^2} \sqrt{\sum_{(i,j)\in E} (y_i + y_j)^2}
\end{aligned}$$

where we applied the Cauchy-Schwartz inequality $\boldsymbol{u}^\top \boldsymbol{v} \le \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$ in the last step. Using now the elementary inequality $(a+b)^2 \le 2(a^2+b^2)$ we may write

$$\sum_{(i,j)\in E} (y_i + y_j)^2 \le 2 \sum_{(i,j)\in E} \big(y_i^2 + y_j^2\big) = 2d \sum_{i=1}^n y_i^2$$

Combining the above with (6) we obtain

$$\frac{\mathbb{E}\Big[\big|E(S_t, \neg S_t)\big|\Big]}{d\,\mathbb{E}\big[|S_t|\big]} \le \frac{\sqrt{\Big(\sum_{(i,j)\in E}(y_i - y_j)^2\Big)\big(2d\sum_{i=1}^n y_i^2\big)}}{d\sum_{i=1}^n y_i^2} = \sqrt{\frac{2\sum_{(i,j)\in E}(y_i - y_j)^2}{d\sum_{i=1}^n y_i^2}}$$

concluding the proof. □

We can now prove Theorem 2.

Proof of Theorem 2. Let $\boldsymbol{x} \in \mathbb{R}^n$ be such that $\boldsymbol{x}^\top \mathbf{1} = 0$ and let $\big(S_F, \neg S_F\big)$ be the cut found by Fiedler's algorithm on input $\boldsymbol{x}$. Lemma 3 states that:

1. there exists a nonnegative vector $\boldsymbol{y}$ such that $R_L(\boldsymbol{y}) \le R_L(\boldsymbol{x})$;
2. for this $\boldsymbol{y}$ and for any $0 < t \le \max_{v\in V} y_v$, the set $S_t = \{v \in V : y_v \ge t\}$ has at most $\frac{n}{2}$ vertices (because $\boldsymbol{y}$ has at most $\frac{n}{2}$ nonzero components, as we defined it using the median) implying $\phi(S_t) = \mathrm{xpn}(S_t)$;
3. the cut $(S_t, \neg S_t)$ is one of the cuts considered by Fiedler's algorithm on input $\boldsymbol{x}$, which implies $\phi(S_F) \le \phi(S_t)$ for all t.

Then, Lemma 5 ensures there exists a threshold $0 < t \le \max_{v \in V} y_v$ such that $\mathrm{xpn}(S_t) \le \sqrt{2R_L(\boldsymbol{y})}$. We can thus write

$$\phi(S_F) \le \phi(S_t) = \mathrm{xpn}(S_t) \le \sqrt{2R_L(\boldsymbol{y})} \le \sqrt{2R_L(\boldsymbol{x})}$$

concluding the proof. $\square$

Nonregular graphs. What is the correct generalization of the Laplacian matrix $I - \frac{1}{d}A$ when G is not d-regular? As we want to preserve the spectral properties, we look at the Rayleigh quotient for the d-regular case:

$$\frac{\sum_{(i,j)\in E}(x_i - x_j)^2}{d\sum_{i=1}^n x_i^2}$$

The natural generalization to nonregular graphs is then

$$\frac{\sum_{(i,j)\in E}(x_i - x_j)^2}{\sum_{i=1}^n d(i)x_i^2} = \frac{\sum_{(i,j)\in E}(x_i - x_j)^2}{\sum_{i=1}^n \left(\sqrt{d(i)}x_i\right)\left(\sqrt{d(i)}x_i\right)} = \frac{\boldsymbol{x}^\top (D-A)\boldsymbol{x}}{\left(D^{1/2}\boldsymbol{x}\right)^\top \left(D^{1/2}\boldsymbol{x}\right)}$$

where where $D^{1/2} = \mathrm{diag}\left(\sqrt{d(1)}, \ldots, \sqrt{d(n)}\right)$ and $d(i)$ is the degree of i. If we now set $\boldsymbol{u} = D^{1/2}\boldsymbol{x}$, the above becomes

$$\frac{\left(D^{-1/2}\boldsymbol{u}\right)^\top (D-A)\left(D^{-1/2}\boldsymbol{u}\right)}{\boldsymbol{u}^\top \boldsymbol{u}} = \frac{\boldsymbol{u}^\top D^{-1/2}(D-A)D^{-1/2}\boldsymbol{u}}{\boldsymbol{u}^\top \boldsymbol{u}} = \frac{\boldsymbol{u}^\top \left(I - D^{-1/2}AD^{-1/2}\right)\boldsymbol{u}}{\boldsymbol{u}^\top \boldsymbol{u}}$$

The matrix $L_{\mathrm{norm}} = I - D^{-1/2}AD^{-1/2}$ is known as the normalized Laplacian. Since the mapping $\boldsymbol{x} \to D^{1/2}\boldsymbol{x}$ is bijective (because $D^{1/2}$ is full rank), for any sublinear space $S \subseteq \mathbb{R}^d$,

$$\min_{\boldsymbol{x}\in S\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{x}^\top (D-A)\boldsymbol{x}}{\left(D^{1/2}\boldsymbol{x}\right)^\top \left(D^{1/2}\boldsymbol{x}\right)} = \min_{\boldsymbol{u}\in S\setminus\{\boldsymbol{0}\}} \frac{\boldsymbol{u}^\top L_{\mathrm{norm}}\boldsymbol{u}}{\boldsymbol{u}^\top \boldsymbol{u}}$$

which implies that all the spectral properties which we proved for d-regular graphs, including Cheeger's inequalities, continue to hold for the normalized Laplacian of arbitrary graphs.

Acknowledgements. Thanks to Alberto Boggio and Alessia Galdeman for flagging mistakes and typos in earlier versions of this handout.