CS-621 Theory Gems	October 24, 2012

Lecture 11

Lecturer: Aleksander Mądry *Scribes: Carsten Moldenhauer and Robin Scheibler*

1 Introduction

In Lecture 10, we introduced a fundamental object of spectral graph theory: the graph Laplacian, and established some of its basic properties. We then focused on the task of estimating the value of eigenvalues of Laplacians. In particular, we proved the Courant-Fisher theorem that is instrumental in obtaining upper-bounding estimates on eigenvalues.

Today, we continue by showing a technique – so-called graph inequalities – that enables us to establish eigenvalue lower-bounding estimates. Then, we use these upper- and lower-bounding tools to obtain estimates on second-smallest eigenvalues of some important graphs.

Later, we discuss how the second-smallest eigenvalue of a (normalized) Laplacian of a graph relates to the mixing time of a random walk in that graph, as well as, its connectivity structure.

2 Brief Recap of Lecture 10

In Lecture 10, we defined the Laplacian matrix L of a weighted and undirected graph $G = (V, E, w)$ to be

$$L := D - A,$$

where A is the adjacency matrix of G given by

$$A_{u,v} = \begin{cases} 1 & \text{if } (u,v) \in E \\ 0 & \text{otherwise.} \end{cases}$$

and D is its weighted degree matrix defined as

$$D_{v,u} = \begin{cases} d_w(u) & \text{if } u = v \\ 0 & \text{otherwise,} \end{cases}$$

with $d_w(u) := \sum_{e=(u,v)\in E} w_e$.

Alternatively, one can view the Laplacian L of a graph G as a simple sum of the Laplacians corresponding to each of its edges. Formally,

$$L = \sum_{e \in E} L^e, \tag{1}$$

where L^e is a Laplacian of graph on vertex set V that has only one edge e and the weight of this edge is w_e.

We also introduced a normalized Laplacian matrix $\widehat{L}$ of G defined as

$$\widehat{L} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}. \tag{2}$$

Now, if $\lambda_1 \leq \ldots \leq \lambda_n$ are the eigenvalues of L (in non-decreasing order) and $\hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_n$ are the eigenvalue of $\widehat{L}$, we proved the following theorem.

Theorem 1 *We have* $0 = \lambda_1 \leq \ldots \leq \lambda_n$, $0 = \hat{\lambda}_1 \leq \ldots \leq \hat{\lambda}_n$, *and if the graph G is connected then* $\lambda_2 > 0$ *and* $\hat{\lambda}_2 > 0$.

Finally, we established the Courant-Fisher theorem.

Theorem 2 (Courant-Fisher) *Let $\mathcal{S}_k$ denote the set of k-dimensional subspaces of $\mathbb{R}^n$, then*

$$\lambda_k = \min_{S \in \mathcal{S}_k} \max_{y \in S} \frac{y^T L y}{y^T y}. \tag{3}$$

As well as, its corollary that is specialized for the case of estimating λ_2.

Corollary 3

$$\lambda_2 = \min_{x \in \mathbb{R}^n : x \perp \mathbf{1}} \frac{x^T L x}{x^T x}.$$

As we mentioned last time, the importance of this theorem (and of the corollary) is that it allows us to obtain upperbounds on the eigenvalues without the need of figuring out their exact values (which might be quite challenging). In particular, in case of λ_2, Corollary 3 tells us that exhibiting any vector x – sometimes called a *test vector* – that is orthogonal to all-ones vector, immediately shows that $\lambda_2 \leq \frac{x^T L x}{x^T x}$.

3 Graphic Inequalities

Last time, we introduced the notion of positive semi-definiteness. Recall that a symmetric matrix M is *positive semi-definite* – denoted $M \succeq 0$ – iff, for any vector x, $x^T M x \geq 0$.

Now, we would like to extend the positive semi-definite relation $\succeq$ to make it induce a partial ordering on symmetric matrices. Namely, for two symmetric matrices M and M', we say that $M \succeq M'$ iff $M - M' \succeq 0$, i.e., the matrix $M - M'$ is positive semi-definite.

We intend to use this ordering mainly in the context of graph Laplacians. Therefore we slightly abuse the notation and for two graph G and H, we write $G \succeq H$ iff $L^G \succeq L^H$, where L^G and L^H are Laplacians of the respective graphs. We will call such relation between two graphs *graphic inequality.*

Note that the fact that $\lambda_1 \geq 0$ (cf. Theorem 1) implies that $L^G \succeq 0$ for any graph G. This allows us to show a simple relation between a graph and any of its subgraphs.

Lemma 4 *For any graph $G = (V, E, w)$ and any of its subgraphs $G' = (V, E', w)$ with $E' \subseteq E$, we have*

$$G \succeq G'.$$

Proof Using (1), we have

$$L^G - L^{G'} = \sum_{e \in E \setminus E'} w_e L^e = L^{G \setminus G'},$$

where $G \setminus G'$ is the graph obtained from G by removing all edges of G'. Since every Laplacian is positive semi-definite, the lemma follows. ∎

Now, our main motivation for studying the graphic inequalities is captured by the following lemma.

Lemma 5 *Let G and H be two graphs such that $G \succeq c \cdot H$, for some $c > 0$. Then, for any k,*

$$\lambda_k(G) \geq c \cdot \lambda_k(H),$$

where $\lambda_k(G)$ and $\lambda_k(H)$ are k-th smallest eigenvalues of the Laplacians L^G and L^H of the respective graphs.

Proof Using the Courant-Fischer theorem (cf. Theorem 2) and the above definitions we have

$$\lambda_k(G) = \min_{S \in \mathcal{S}_k} \max_{y \in S} \frac{y^T L^G y}{y^T y} \geq \min_{S \in \mathcal{S}_k} \max_{y \in S} \frac{c \cdot y^T L^H y}{y^T y} = c \cdot \lambda_k(H).$$

∎

In the light of the above lemma, we can view graphic inequalities as a technique for proving eigenvalue lowerbounds.[1] Namely, if for a given graph G we are able to find another graph H such that $G \succeq c \cdot H$, for some $c > 0$, and we know accurate lowerbounds on some of the eigenvalues of H, then we immediately get lowerbounds on the corresponding eigenvalues of G.

Of course, the crux here is to be able to come up with appropriate lower-bounding graph H that should be – on one hand – sufficiently simple so we can understand its eigenvalues, but – on the other hand – flexible enough to allow us to relate it to a diverse set of graphs. Somewhat surprisingly, just taking H to be a complete graph and employing a simple graphic inequality that we introduce below, turns out to be a quite powerful tool for eigenvalue estimation.

Lemma 6 *Let P_n be a graph corresponding to a length-n path with u and v being its endpoints, then*

$$(n-1)P_n \succeq G^{(u,v)},$$

where $G^{(u,v)}$ is a graph on the same vertex set as P_n, but containing only edge (u, v) (that joins the two endpoints of P_n).

Proof Following the definition of graphic inequalities, we have to show that $(n-1)P_n - G^{(u,v)} \succeq 0$. To this end, let us fix an arbitrary $x \in \mathbb{R}^n$. We will show that $(n-1)x^T P_n x - x^T G^{(u,v)} x \geq 0$.

Let us use a naming convention in which the vertex set of both graphs is $\{1, \ldots, n\}$ and the endpoints u and v correspond to 1 and n respectively. In this case, we need to show that

$$(n-1)\sum_{i=1}^{n-1}\left[(x_{i+1}-x_i)^2\right] - (x_n - x_1)^2 \geq 0.$$

Substituting $\delta_i := x_{i+1} - x_i$ allows us to rewrite this inequality as

$$(n-1)\sum_{i=1}^{n-1}\delta_i^2 \geq \left(\sum_{i=1}^{n-1}\delta_i\right)^2.$$

But it is easy to see that by the quadratic-arithmetic-mean inequality (which is a special case of Cauchy-Schwarz inequality) we get

$$\sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n-1}\delta_i^2} \geq \frac{1}{n-1}\left|\sum_{i=1}^{n-1}\delta_i\right|,$$

which is equivalent to the condition we wanted to establish. ■

Side Remark: As pointed out in the class by Javad Ebrahimi, there is a simple argument that shows that the converse of Lemma 5 does not hold, i.e., that $\lambda_k(G) \geq c \cdot \lambda_k(H)$, for all k, does not necessarily imply that $G \succeq H$. (In other words, the relation $G \succeq H$ is stronger than just the corresponding inequalities for eigenvalues.)

To see this, note that if $e = (u', v')$ is one of the edges of the path graph P_n, then the graph G^e is a subgraph of P_n and thus, by Lemma 4, $P_n \succeq G^e$. This, by Lemma 5, implies that $\lambda_k(P_n) \geq \lambda_k(G^e)$, for all k.

Now, if we consider the graph $G^{(u,v)}$, for u and v being the endpoints of P_n, then we also have that $\lambda_k(P_n) \geq \lambda_k(G^{(u,v)})$, for all k, as the graphs $G^{(u,v)}$ and G^e are isomorphic and thus their eigenvalues (but not eigenvectors!) have to be the same.

However, looking at the proof of Lemma 6, one can see that the inequality from its statement is tight and thus, in particular, $P_n \not\succeq G^{(u,v)}$.

[1] Of course, in principle, one can also use these inequalities for proving eigenvalue upperbounds, but we will be interested only in the lower-bounding aspect.

4 Estimation of λ_2 of Certain Graphs

We will now show how one can use Courant-Fisher theorem (or, more precisely, Corollary 3) and graphic inequalities to obtain quite precise estimates on the second-smallest eigenvalues of some important graphs.

As we mentioned, when applying graphic inequalities, we will be comparing our graphs to the complete graph K_n on n vertices. One can easily check that

$$\lambda_1(K_n) = 0 \qquad \lambda_k(K_n) = n \quad \text{for } k \geq 2.$$

4.1 Path Graph

Let P_n be a graph describing a path $(1, 2, \ldots, n)$ on n vertices. We first provide an upperbound on $\lambda_2(P_n)$. To this end, let us consider a test vector x defined as

$$x_i = n + 1 - 2i,$$

for each $1 \leq i \leq n$. We can verify that

$$x^T \cdot \mathbf{1} = \sum_{i=1}^{n} n + 1 - 2i = n(n+1) - 2\sum_{i=1}^{n} i = 0,$$

hence $x \perp \mathbf{1}$. So, by Corollary 3, we can conclude that

$$\lambda_2(P_n) \leq \frac{x^T L^{P_n} x}{x^T x} = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{x^T x} = \frac{4(n-1)}{\sum_{i=1}^{n} (n+1-2i)^2} = O(\frac{1}{n^2}),$$

as $\sum_{i=1}^{n} (n+1-2i)^2 = \Theta(n^3)$.

To establish a lowerbound, we will show that $cP_n \succeq K_n$ for some (still to be determined) $c > 0$. Clearly, the complete graph is a union of all the possible edges, i.e.,

$$L^{K_n} = \sum_{i,j,i<j} L^{(i,j)}.$$

Looking at the proof of Lemma 6, one can easily see that, for any i and j, $i < j$, we have

$$(j - i) \cdot P_n \succeq G^{(i,j)}.$$

Summing over all pairs (i, j) we obtain

$$\sum_{i,j,i<j} (j - i) \cdot P_n \succeq \sum_{i,j,i<j} G_{(i,j)} = K_n.$$

This means that we can set $c = \sum_{i,j,i<j} (i - j) = \Theta(n^3)$ and obtain that $c \cdot P_n \succeq K_n$. So, using Lemma 5, we get that $c\lambda_2(P_n) \geq \lambda_2(K_n) = n$, which implies that $\lambda_2(P_n) = \Omega\left(\frac{1}{n^2}\right)$.

Putting the upperbound and lowerbound together, we obtain that $\lambda_2(P_n) = \Theta\left(\frac{1}{n^2}\right)$, which is an estimate that is tight up to a constant.

4.2 Binary Tree

We consider now a binary tree T_n as depicted on Figure 1. As before, we exhibit first a test vector x that will provide an upperbound on the value of $\lambda_2(T_n)$. Let x be defined as

$$x_i = \begin{cases} 0 & \text{if } i \text{ is the root node,} \\ 1 & \text{if } i \text{ is left of the root node,} \\ -1 & \text{if } i \text{ is right of the root node.} \end{cases}$$

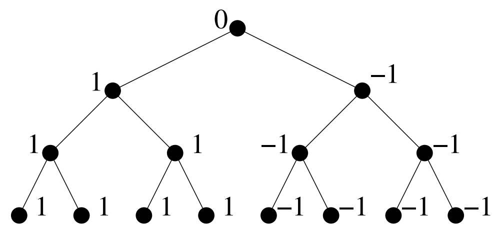

Figure 1: An example of a binary tree T_n with the choice of the value of x next to the nodes.

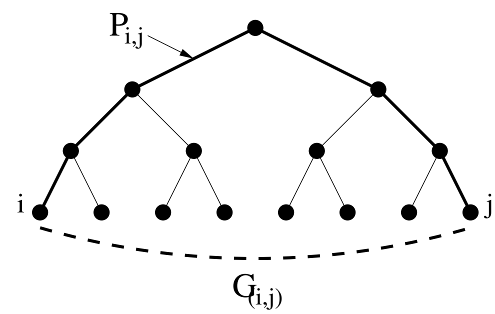

Figure 2: The path graph on the binary graph used to find the upper bound on $\lambda_2(T_n)$.

It is easy to see that $x \perp \mathbf{1}$. Therefore, Corollary 3 implies

$$\lambda_2(T_n) \leq \frac{x^T L^{T_n} x}{x^T x} = \frac{2}{n-1} = O\left(\frac{1}{n}\right).$$

Now, we proceed to establishing a lowerbound on $\lambda_2(T_n)$. Again, we use the complete graph K_n here. Note that $K_n \succeq T_n$ since T_n is a subgraph of K_n (cf. Lemma 4). But, similarly as before, we would like to find some $c > 0$ such that $c \cdot T_n \succeq K_n$.

To this end, we first note that for any i and j, $i < j$,

$$2 \log_2 n \; T_n \succeq 2 \log_2 n \; P_{i,j} \succeq G^{(i,j)},$$

where $P_{i,j}$ is the subgraph of T_n corresponding to the (unique) path in T_n that connects i and j. (Note that this path has length of at most $2 \log_2 n$ – cf. Figure 2.)

Summing over all pairs (i,j) we get

$$\sum_{i,j,i<j} 2 \log_2 n \; T_n \succeq \sum_{i,j,i<j} G^{(i,j)} = K_n.$$

Therefore, Lemma 5 and the fact that $\lambda_2(K_n) = n$ implies that

$$\lambda_2(T_n) = \Omega\left(\frac{1}{n \log_2 n}\right).$$

As we see, when we put together our lower- and upper-bound, there is a gap of $\Theta(\log n)$ between them. On the problem set, you will be asked to tighten this gap and show that $\lambda_2(T_n) = \Theta(\frac{1}{n})$.

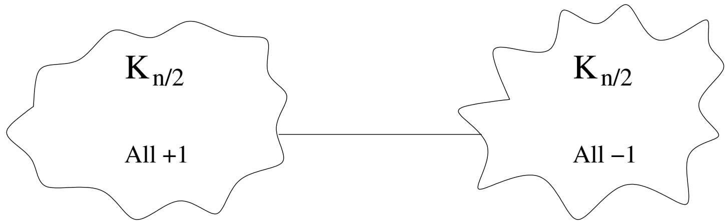

Figure 3: The dumbell graph D_n is composed of two complete graphs $K_{n/2}$ connected by a single bridge edge. We choose a test vector x such that all the nodes in the left (resp. right) copy of $K_{n/2}$ are assigned value of $+1$ (resp. -1).

4.3 Dumbell Graph

Our last example is the Dumbell graph D_n which is a graph composed of two complete graphs $K_{n/2}$ on $n/2$ vertices that are connected by a single bridge edge – cf. Figure 3.

We can easily get an upperbound on the value of $\lambda_2(D_n)$ by employing Corollary 3 with a test vector x such that all nodes in one copy of $K_{n/2}$ are assigned a value of $+1$, and all the nodes in the other copy are assigned a value of -1. Clearly, $x \perp \mathbf{1}$ and thus we have that

$$\lambda_2(D_n) \leq \frac{x^T L^{D_n} x}{x^T x} = \frac{4}{n}, \tag{4}$$

as the only contribution in the enumerator is coming from the bridge edge.

Now, to derive a lowerbound on $\lambda_2(D_n)$, we again compare our graph to the complete graph K_n. First, we observe that for any two given vertices i and j, there is a path of length at most 3 connecting them in D_n. Hence, following our usual approach, we can write for any i, j that

$$3\ D_n \succeq G_{(i,j)}.$$

Summing for all i, j, with $i > j$ we obtain

$$\frac{3}{2}n(n-1)\ D_n \succeq \sum_{i>j} G_{(i,j)} = K_n,$$

and thus by Lemma 5 we get that

$$\lambda_2(D_n) \geq \frac{2}{3(n-1)},$$

which – together with our upperbound (4) – shows that $\lambda_2(D_n) = \Theta\left(\frac{1}{n}\right)$.

5 Eigenvalues and Obstructions to Mixing of Random Walks

Once we developed tools for estimating the values of Laplacian eigenvalues and established good estimated for λ_2 for a variety of basic graphs, it is natural to ask why knowing λ_2 of a graph is even useful.

The reason for our interest in this value is that it is related to a variety of important graph properties such as the graph connectivity structure, its diameter, and behavior of random walks in it. We will be unveiling some of these connections in the coming lectures.

In particular, today we describe one of the most fundamental among them: the relationship between the Laplacian eigenvalues and mixing properties of random walks. As we will see, this relationship extends even further and, in particular, there is a link – captured by so-called Cheeger's Inequality – between the eigenvectors and the connectivity/cut structure of the graph.

5.1 L vs. $\widehat{L}$

Before we proceed further, there is a confession to be made. So far, in our treatment of Laplacians, we were mainly focused on the Laplacian matrix L of a graph and its eigenvalues $\lambda_1, \ldots, \lambda_n$, while treating the normalized Laplacian $\widehat{L}$ (cf. (2)) and its eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ as objects of only secondary importance.

However, the truth is that from the point of view of applications of spectral graph theory, it is the normalized Laplacian and its eigenvalues – especially, $\hat{\lambda}_2$ – that are the more natural objects that play the more prominent role.

The reason why we kept this misleading focus on the Laplacian L was twofold. Firstly, the matrix L and its eigenvalues are slightly easier to work with and thus are better when introducing all the main concepts. Secondly, and more importantly, actually all the machinery we developed for the Laplacian L transfers over – essentially unchanged – to the normalized Laplacian case. In particular, one can verify that

$$\frac{1}{d_{\min}} L \succeq \widehat{L} \succeq \frac{1}{d_{\max}} L,$$

where $d_{\min} := \min_{u \in V} d(u)$ and $d_{\max} := \max_{u \in V} d(u)$. So, by applying reasoning that is completely analogous to the one used in the proof of Courant-Fisher Theorem and Lemma 5, one can conclude that

$$\frac{1}{d_{\min}} \lambda_k \geq \hat{\lambda}_k \geq \frac{1}{d_{\max}} \lambda_k,$$

for any $1 \leq k \leq n$. That is, in case of a graph G being almost-regular – i.e., having *degree ratio* $\bar{d}(G) := \frac{d_{\max}}{d_{\min}}$ small, say $O(1)$ – the value of $\frac{\lambda_k}{d_{\max}}$ is a $O(1)$-approximation to the value of $\hat{\lambda}_k$. As for all the examples of graphs we considered above, $\bar{d}(G)$ was indeed $O(1)$, we can conclude that

$$\begin{aligned}
\hat{\lambda}_2(K_n) &= 1 \\
\hat{\lambda}_2(P_n) &= \Theta\left(\frac{1}{n^2}\right) \\
\hat{\lambda}_2(T_n) &= \Theta\left(\frac{1}{n}\right) \\
\hat{\lambda}_2(D_n) &= \Theta\left(\frac{1}{n^2}\right).
\end{aligned}$$

From now on, we will focus our attention on the normalized Laplacian $\widehat{L}$ and its eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$.

5.2 Mixing Time and $\hat{\lambda}_2$

We proceed to exploring the connections between the mixing properties of random walks in a graph and the second-smallest eigenvalue of its normalized Laplacian. To avoid the need for dealing with possibility of the graph being bipartite (or close to it), we will focus our treatment on the lazy variant of the random walks – see Lecture 9 for its definition.

To this end, recall from Lecture 9 that the vector $\hat{p}^t$ denotes the vertex probability distribution induced by t steps of lazy random walk that started from some starting distribution $\hat{p}^0$. That is, $\hat{p}^t_v$, for a given vertex v, is the probability that after taking t steps of lazy random walk, we end up at vertex v.

In Lecture 9, we showed that, as long as, the underlying graph is connected, the distribution $\hat{p}^t$ always converges to a stationary distribution π given by

$$\pi_u := \frac{d(u)}{\sum_v d(v)},$$

for each vertex u.

The rate of this convergence can be estimated using the following lemma (that we stated in Lecture 9 without proof)

Lemma 7 *For any graph $G = (V, E)$, any starting distribution $\hat{p}^0$, and any $t \geq 0$,*

$$\|\hat{p}^t - \pi\| \leq \sqrt{\bar{d}(G)}\hat{\omega}_2^t,$$

where $\bar{d}(G) := \frac{d_{\max}}{d_{\min}}$ is the degree ratio of G and $\hat{\omega}_2$ is the second-largest eigenvalue of lazy walk matrix $\widehat{W}$.

Note that the bound in the above lemma is mainly dependent on how much smaller than one is the value of $\hat{\omega}_2$. To make this precise, in Lecture 9 we defined the *spectral gap* $\hat{\lambda}(G)$ of a graph G to be

$$\hat{\lambda}(G) := 1 - \omega_2 = 2(1 - \hat{\omega}_2),$$

where ω_2 is the second-largest eigenvalue of the walk matrix W.

Now, if for given $\varepsilon > 0$, t_ε denotes the number of steps needed for the lazy random walk distribution $\hat{p}^t$ to be within ε of the stationary distribution π, then by Lemma 7 we have

$$t_\varepsilon = t_{1/2} \log_2 \frac{1}{\varepsilon} = O\left(\frac{1}{\hat{\lambda}(G)} \log \frac{\bar{d}(G)}{\varepsilon}\right), \tag{5}$$

where $t_{1/2}$ is called the *mixing time* of G and can be viewed as a parameter that describes the convergence rate of random walks in G.

To see why (5) holds, note that

$$\sqrt{\bar{d}(G)}\hat{\omega}_2^t = \sqrt{\bar{d}(G)}(1 - (1 - \hat{\omega}_2))^t \leq \sqrt{\bar{d}(G)}e^{(1-\hat{\omega}_2)t} = \sqrt{\bar{d}(G)}e^{\hat{\lambda}(G)t/2},$$

as $(1 - x)^{\frac{1}{x}} \leq e^{-1}$.

Now, to see how the mixing time of G is connected to the second-smallest eigenvalue of its normalized Laplacian, we note that in Lecture 10 (cf. Claim 4 there) we showed that

$$\hat{\lambda}_2 = 1 - \omega_2,$$

which implies that the spectral gap $\hat{\lambda}(G)$ of G is equal to $\hat{\lambda}_2$. (In fact, one usually defines it in this way.)

In the light of the above, we see that in a graph that has large spectral gap $\hat{\lambda}(G) = \hat{\lambda}_2$ (for example, the complete graph), we can expect the random walk to convergence fast – roughly, with mixing time being inverse-proportional to $\hat{\lambda}(G)$. Furthermore, in some sense, this connection is tight. Namely, from our proof of convergence in Lecture 9, one can see that if a graph has a small spectral gap and the starting distribution $\hat{p}^0$ of the lazy random walk has a large intersection with the eigenspace of $\widehat{W}$ that corresponds to $\hat{\omega}_2$, then the rate of convergence will necessarily be slow (and proportional to $\frac{1}{(1-\hat{\omega}_2)} = \frac{2}{\hat{\lambda}(G)}$).

5.3 Spectral Gap, Cuts, and Obstructions to Rapid Mixing

Although the connection between the spectral gap and mixing time we developed above is already pretty satisfying, we still would like to pursue it further. In particular, we would like to understand how graphs with large/small mixing time (and thus respectively small/large spectral gap) look like from the point of view of their connectivity/cut structure.

To make this precise, we need first to introduce some graph-theoretic definitions. To this end, for a given weighted and undirected graph $G = (V, E, w)$ and any non-empty and proper subset S of the vertex set V – we will call such S a *cut* of G – define

$$\partial S := \{(i, j) \in E \mid i \in S, j \notin S\},$$

i.e., ∂S is the set of all the edges of G that have exactly one endpoint in the cut S.

Next, let us define the *conductance* $\Phi(S)$ of the cut S as

$$\Phi(S) := \frac{w(\partial S) d(V)}{d(S) d(V \setminus S)},$$

where, for any $F \subseteq E$, $w(F) := \sum_{e \in F} w_e$, and, for any $U \subseteq V$, $d(U)$ is the *volume* of U equal to $\sum_{v \in U} d(v)$.

Intuitively, the conductance $\Phi(S)$ of the cut S measures how much connectivity there is between S and $V \setminus S$. That is, $\Phi(S)$ is smallest when there is only few (low-weight) edges between S and $V \setminus S$, and volumes of these two sets are roughly balanced (so they both are large).[2]

Finally, we define the *conductance* Φ_G of the graph G as

$$\Phi_G := \min_S \Phi(S),$$

i.e., Φ_G is equal to the conductance of the minimal-conductance cut of G. We can view Φ_G, as a measure of the connectivity of G, the smaller it is the larger connectivity bottleneck G has.

Now, we will be interested in tying the conductance of a graph to its mixing time (and thus its spectral gap). To get a hint of what this connection could be, note that the conductance $\Phi(S)$ of a cut S can be seen as a measure of how much probability mass can escape from S in one step of a (lazy) random walk. To make this more precise, consider a starting distribution $\hat{p}^0$ of the random walk to be

$$\hat{p}^0_u = \pi_S(u) = \begin{cases} \frac{d(u)}{d(S)} & \text{if } u \in S, \\ 0 & \text{o.w.} \end{cases},$$

i.e., $\hat{p}^0 = \pi_S$ is given by conditioning the stationary distribution π on being in the cut S. It is not hard to see that in such random walk the total amount of probability mass that can cross in one step from the cut S to $S \setminus V$ is at most $\Phi(S)$.

Given the above intuition, one would expect that if the conductance Φ_G of G is small then the mixing time will be large (and thus the spectral gap will be small) as – at least for some starting distribution – it will take the random walk a long time to move probability mass through the bottlenecking cuts.

As it turns out, this intuition can be confirmed – one can show that $\hat{\lambda}(G)$ is always upperbounded by Φ_G. What is, however, even more striking is that this relationship between $\hat{\lambda}(G)$ and Φ_G goes also the other way, as stated in the following theorem.

Theorem 8 (Cheeger's inequality) *For any graph G,*

$$\frac{\Phi_G^2}{4} \le \hat{\lambda}(G) \le \Phi_G.$$

We will discuss this very important theorem in more detail (as well as, prove one of these inequalities) in the next lecture, but for now we just want to note that both of these inequalities are tight (up to a constant). To see that the left inequality is tight, consider G to be a path graph P_n. As we already computed, $\hat{\lambda}(P_n) = \Theta(1/n^2)$, while one can easily see that $\Phi_{P_n} = \Theta(1/n)$. To see that the right inequality is also tight, consider G to be a binary tree graph T_n. We have $\hat{\lambda}(T_n) = \Theta(1/n)$ and $\Phi_{T_n} = \Theta(1/n)$.

[2]The conductance $\Phi(S)$ is sometimes also defined as $\frac{|\partial S|}{\min\{d(S), d(V \setminus S)\}}$. Note that both these definitions are equivalent up to a multiplicative factor of two.