CME 305: Discrete Mathematics and Algorithms
Instructor: Professor Aaron Sidford (sidford@stanford.edu)
February 8, 2018

Lecture 10 - Spectral Graph Theory[1]

In the last lecture we began our study of spectral graph theory, the study of connections between combinatorial properties of graphs and linear algebraic properties of the fundamental matrices associated with them. In this lecture we continue this study making connections between eigenvalues of and eigenvectors of the Laplacian matrix associated with a simple, undirected, positively weighted graph and combinatorial properties of these graphs.

1 Recap

In the remainder of this lecture our focus will be on a simple, undirected, positively weighted 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. Also note that while we might define things with G explicitly in the definition, e.g. $\mathcal{L}(G)$ or $\deg_G$ we may drop the G when it is clear from context.

Now, last class we introduced several fundamental matrices associated with a graph, 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 as we pointed out a matrix $\mathbf{M}$ is an adjacency matrix of some G if and only if $\mathbf{A}_{ji} = \mathbf{A}_{ij} \geq 0$ for all $i, j \in V$ and $\mathbf{A}_{ii} = 0$ for all $i \in V$ and a matrix a Laplacian matrix of some G if and only if $\mathcal{L}_{ij} = \mathcal{L}_{ji} \leq 0$ and $\mathcal{L}\vec{1} = \vec{0}$. Furthermore, each gives a bijection from these matrices to simple, undirected, weighted graphs with positive edge weights.

2 Laplacian Quadratic Form

Now why do we focus on Laplacian matrices? As we will show the quadratic form of the Laplacian has a number of nice properties.

2.1 Laplacians are PSD

One nice properties of Laplacians are that they are positive semidefinite (PSD).

Definition 1 (Positive Semidefinite (PSD))**.** A symmetric matrix $\mathbf{M} = \mathbf{M}^\top$ is *positive semidefinite (PSD)* if and only if for all x we have $x^\top \mathbf{M} x \geq 0$.

We refer to $x^\top \mathbf{M} x$ as the quadratic form of $\mathbf{M}$ and this definition says a symmetric matrix is PSD if and only if the quadratic form is non-negative. As we will see this is equivalent to saying that all eigenvalues of $\mathbf{M}$ are non-negative.

[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.

Here we will prove that $\mathcal{L}(G)$ is PSD, proving several interesting properties of Laplacians and their quadratic form along the way.

We begin, by defining the sum of two graphs as follows:

Definition 2 (Graph Sums)**.** For simple, undirected, weighted graphs $G_1 = (V, E_1, w_1)$ and $G_2 = (V, E_2, w_2)$ we define the graph sum $G_1 + G_2 = (V, E_+, w_+)$ by $E_+ \stackrel{\text{def}}{=} E_1 \cup E_2$ and for $\{i, j\} \in E_+$

$$w_+(i,j) = \begin{cases} w_1(i,j) + w_2(i,j) & \{i,j\} \in E_1 \cap E_2 \\ w_1(i,j) & \{i,j\} \in E_1 \setminus E_2 \\ w_2(i,j) & \{i,j\} \in E_2 \setminus E_1 \\ 0 & \text{otherwise} \end{cases}.$$

In other words, the sum of two graphs with the same vertices we are defining as the graph on the same vertices with the union of their edges and the weight of the edges defined as the sum of the weights in each graph (where if an edge does not exist in a graph we are summing it with weight 0).

One nice property of this definition is that it directly corresponds to summing graph Laplacians. In other words, it can easily be checked that

$$\mathcal{L}(G_1 + G_2) = \mathcal{L}(G_1) + \mathcal{L}(G_2)\,.$$

Consequently, if for $e = \{i, j\} \in E$ we define $\mathcal{L}(e)$ to be the Laplacian on the graph with vertices V and 1 edge of weight 1 between i and j then we see that

$$\mathcal{L}(G) = \sum_{e \in E} w_e \mathcal{L}(e)$$

and therefore

$$x^\top \mathcal{L} x = \sum_{e \in E} w_e \cdot x^\top \mathcal{L}(e) x\,.$$

Thus, to reason about the quadratic form of $\mathcal{L}$ it suffices to reason about the quadratic form of a single edge.

Now $\mathcal{L}(e)$ for $e = \{i, j\}$ simply has a 1 in coordinates ii and jj and a -1 in ij and ji. All other coordinates are 0. Consequently, if we let $\vec{1}_i$ be the vector the is 0 at all coordinates except for i which is a 1, i.e.

$$\vec{1}_i(j) = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}$$

then we see that

$$\mathcal{L}(e) = \left(\vec{1}_i - \vec{1}_j\right)(\vec{1}_i - \vec{1}_j)^\top = \vec{\delta}_{ij}\vec{\delta}_{ij}^\top$$

where we let $\vec{\delta}_{ij} \stackrel{\text{def}}{=} \vec{1}_i - \vec{1}_j$ for shorthand.Consequently, we have that

$$x^\top \mathcal{L}(e) x = (x_i - x_j)^2 \geq 0$$

and

$$x^\top \mathcal{L}(G) x = \sum_{\{i,j\} \in E} w_{\{i,j\}} (x_i - x_j)^2 \geq 0$$

and therefore $\mathcal{L}(G)$ is PSD as claimed.

2.2 Laplacian Quadratic Form

Note that the quadratic form the Laplacian has a nice interpretation. Imagine each edge $\{i,j\}$ is a spring with stiffness $w_{\{i,j\}}$ and that the vertices designate which springs are linked together. Now given a vector $x \in \mathbb{R}^V$we could imagine putting one edge of spring $\{i,j\}$ at position x_i on a line and the other end at position x_j for each edge. If we do this, then $x^\top \mathcal{L}(G)x = \sum_{\{i,j\}\in E} w_{ij}(x_i - x_j)^2$ would be the total potential energy of the springs. In other words, $x^\top \mathcal{L}(g)x$, corresponds to the energy to pin each connection point $i \in V$ at point x_i on the line, i.e. the enrgy to layout the graph along a line.

There are several nice implications of this analysis. First, suppose that for some $S \subseteq V$ we have $x = \vec{1}_S$, i.e. the indicator vector for set S, where for all $i \in V$ we have

$$\vec{1}_S(i) = \begin{cases} 1 & i \in S \\ 0 & i \notin S \end{cases}.$$

What is $x^\top \mathcal{L}x$ in this case? We have

$$x^\top \mathcal{L}x = \vec{1}_S^\top \mathcal{L}\vec{1}_S = \sum_{\{i,j\}\in E} w_{\{i,j\}}(x_i - x_j)^2 = \sum_{\{i,j\}\in\partial(S)} w_{ij} = w(\partial(S))$$

in other words, it is the total weight of the edges cut by S. Consequently, the quadratic form of $\mathcal{L}$ specialized to boolean vectors, i.e. $x \in \{0,1\}^V$, gives the size of all cuts in the graph.

2.3 Incidence Matrix Decomposition

Another nice implication of our analysis in deriving the quadratic form of a Laplacian is that it gives us another way of decomposing Laplacian matrices. Let $e_1, ..., e_m$ denote the edges the graph and suppose we pick an arbitrary orientation of the edges so that $e_i = (u_i, v_i)$ for all $i \in [m]$. Now, we define the *incidence matrix* of G to be $\mathbf{B}(G) \in \mathbb{R}^{E\times V}$ where

$$\mathbf{B}(G) = \begin{pmatrix} \vec{\delta}_{e_1}^\top \\ \vec{\delta}_{e_2}^\top \\ \vdots \\ \vec{\delta}_{e_m}^\top \end{pmatrix} = \begin{pmatrix} (\vec{1}_{u_1} - \vec{1}_{v_1})^\top \\ (\vec{1}_{u_2} - \vec{1}_{v_2})^\top \\ \vdots \\ (\vec{1}_{u_m} - \vec{1}_{v_m})^\top \end{pmatrix}$$

and we define the *weight matrix* of G to be $\mathbf{W}(G) \in \mathbb{R}^{E\times E}$, the diagonal matrix associated with w, i.e.

$$\mathbf{W}(G)_{e_1e_2} = \begin{cases} w_e & e = e_1 = e_2 \\ 0 & \text{otherwise} \end{cases}.$$

Now, with these definitions established we have

$$\mathbf{B}^\top\mathbf{W}\mathbf{B} = \sum_{e\in E} w_e \vec{\delta}_e \vec{\delta}_e^\top = \sum_{e\in E} w_e \mathcal{L}(e) = \mathcal{L}(G).$$

Note that if we define $\mathbf{W}^{1/2}$ to be $\mathbf{W}$ where we take the square root of each diagonal entry then we have

$$\mathcal{L} = \left(\mathbf{W}^{1/2}\mathbf{B}\right)^\top (\mathbf{W}^{1/2}\mathbf{B})$$

and therefore clearly $\mathcal{L}$ is PSD as $x^\top\mathcal{L}x = \|\mathbf{W}^{1/2}\mathbf{B}x\|_2^2$. Consequently, we can prove that $\mathcal{L}$ is PSD by efficiently providing an explicit factorization of $\mathcal{L} = \mathbf{M}^\top\mathbf{M}$ for some $\mathbf{M}$. It can be shown that all PSD matrices have such a factorization, however that for Laplacians it can be compute so easily is a particularly nice property.

3 Linear Algebra Review

Now that we have a basic understanding of the Laplacian matrix and its quadratic form. Our next step is to take a closer look at the eigenvectors and eigenvalues of this matrix. Before we do this, here we review some linear algebra theory regarding eigenvalues and eigenvectors of symmetric matrices.

First, we recall the definition of an eigenvector and an eigenvalue.

Definition 3 (Eigenvalues and Eigenvectors)**.** A vector v is an *eigenvector* of matrix $\mathbf{M}$ with *eigenvalue* λ if and only if $\mathbf{M}v = \lambda v$.

Next, we recall some basic properties of eigenvectors of symmetric matrices

Lemma 4. *Let $\mathbf{M} \in \mathbb{R}^{n\times n}$ be a symmetric matrix and let $v_1, v_2 \in \mathbb{R}^n$ be eigenvectors of $\mathbf{M}$ with eigenvalues λ_1 and λ_2 respectively. If $\lambda_1 \neq \lambda_2$ then $v_1 \perp v_2$ and if $\lambda_1 = \lambda_2$ then for all $\alpha, \beta \in \mathbb{R}$ we have that $\alpha v_1 + \beta v_2$ is an eigenvector of $\mathbf{A}$ of eigenvalue λ_1.*

Proof. Note that $v_1^\top \mathbf{M} v_2 = \lambda_2 v_1^\top v_2$ by assumption. Furthermore, since $\mathbf{M}$ is symmetric we have $v_1^\top \mathbf{M} = \lambda_1 v_1^\top$ and thus $v_1^\top \mathbf{M} v_2 = \lambda_1 v_1^\top v_2$. Consequently

$$0 = v_1^\top \mathbf{M} v_2 - v_1^\top \mathbf{M} v_2 = (\lambda_1 - \lambda_2) \cdot v_1^\top v_2$$

and if $\lambda_1 \neq \lambda_2$ then $v_1^\top v_2 = 0$ yielding the first fact. The second follows as by linearity $\mathbf{A}(\alpha v_1 + \beta v_2) = \alpha\lambda_1 v_1 + \beta\lambda_2 v_2$. □

From this lemma it is clear that for a symmetric matrix $\mathbf{M}$ the set of eigenvectors of a particular eigenvalue form a linear subspace, known as an *eigenspace*, and each eigenspace is orthogonal from other eigenspaces. Consequently, if the eigenspaces spanned $\mathbb{R}^n$ then we could just think of the effect of $\mathbf{M}$ on each eigenspace separately. However, one well known fact about symmetric linear operators is that this always the case and moreoever, the eigenvalues are real.

Theorem 5 (Spectral Theorem for Real Symmetric Matrices)**.** *If $\mathbf{M} \in \mathbb{R}^{n\times n}$ is symmetric then there is an orthonormal basis of eigenvectors, $v_1(\mathbf{M}), ..., v_n(\mathbf{M})$, associated with real eigenvalues $\lambda_1(\mathbf{M}) \leq \lambda_2(\mathbf{M}) \leq ... \leq \lambda_n(\mathbf{M}) \in \mathbb{R}$, i.e for all $i, j \in [n]$*

$$v_i^\top v_j = \begin{cases} 1 & i = j \\ 0 & \textit{otherwise} \end{cases} \quad \textit{and } \mathbf{M}v_i = \lambda_i \,.$$

Proof. Note that v is an eigenvalue of $\mathbf{M}$ of eigenvalue λ if and only if $\mathbf{M}v = \lambda v$ or equivalently, $(\mathbf{M} - \lambda\mathbf{I})v = 0$. Consequently, λ is an eigenvalue of $\mathbf{M}$ if and only if $\det(\mathbf{M} - \lambda\mathbf{I}) = 0$. Now, $\det(\mathbf{M} - \lambda\mathbf{I})$ is a polynomial in λ with real coefficients. Consequently it has some complex root $r \in \mathbb{C}$, i.e. $\det(\mathbf{M} - r\mathbf{I}) = 0$, and for some $v \in \mathbb{C}^n$ with $v \neq 0$ we have $\mathbf{M}v = r \cdot v$. However, as $\mathbf{M}\overline{v} = \overline{\mathbf{M}v} = \overline{r} \cdot \overline{v}$ and $\mathbf{M}$ is symmetric we have that $v^\top \mathbf{M} \overline{v} = r v^\top \overline{v} = \overline{r} \cdot v^\top \overline{v}$. Since $v \neq 0$ we have $v^\top \overline{v} \neq 0$ and thus $r = \overline{r}$, i.e. r is real, $r \in \mathbb{R}$. Consequently, $\mathbf{M}$ has a real eigenvalue r and associated real eigenvector v. Since for all $x \perp v$ by the symmetry of $\mathbf{A}$ we have that $v^\top \mathbf{A} x = r \cdot v^\top x = 0$ we see that $\mathbf{M}$ is an invariant linear operator on the subspace orthogonal to v. Consequently, we can repeat the argument on this space yielding the result by induction. □

This means that given any symmetric matrix $\mathbf{M}$ we can efficiently *diagonalize* it, i.e. write $\mathbf{A} = \mathbf{PDP}^{-1}$ for invertible $\mathbf{P}$ and diagonal $\mathbf{D}$. In other words, $\mathbf{M}$, is diagonal under some change of bases and therefore many questions about $\mathbf{M}$ to questions about diagonal matrices.

Theorem 6 (Diagonalization of Symmetric Matrices)**.** *For symmetric* $\mathbf{M} \in \mathbb{R}^{n \times n}$ *let* $\mathbf{V}(\mathbf{M}) \in \mathbb{R}^{n \times n}$ *be the matrix where column* i *is* $v_i(\mathbf{M})$ *and let* $\mathbf{\Lambda}(\mathbf{M})$ *be the diagonal matrix where* $[\mathbf{\Lambda}(\mathbf{M})]_{ii} = \lambda_i(\mathbf{M})$*. Then,*

$$\mathbf{I} = \mathbf{V}^\top \mathbf{V} = \mathbf{V}\mathbf{V}^\top = \sum_{i \in [n]} v_i v_i^\top \text{ and } \mathbf{V}^{-1} = \mathbf{V}^\top$$

and

$$\mathbf{M} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top = \sum_{i \in [n]} \lambda_i \cdot v_i v_i^\top .$$

Proof. Since, by Theorem 5 we have that $v_i^\top v_i = 1$ and $v_i^\top v_j = 0$ for $i \neq j$ clearly $\mathbf{I} = \mathbf{V}^\top \mathbf{V}$. Consequently $\mathbf{V}^{-1} = \mathbf{V}^\top$ and $\mathbf{I} = \mathbf{V}\mathbf{V}^\top$. To obtain the second fact, note that $\mathbf{M} v_i = \lambda_i v_i$ and therefore

$$\mathbf{M} = \mathbf{M}\mathbf{I} = \mathbf{M} \sum_{i \in [n]} v_i v_i^\top = \sum_{i \in [n]} \lambda_i v_i v_i^\top = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top .$$

□

4 Smallest Eigenvaleus of Laplacian

With the spectral theorem for real symmetric matrices established, our next question is what do the eigenvalues of a Laplacian look like? As we have eluded to before, from the fact that $\mathcal{L}$ is PSD, we can show that its eigenvalues are all non-negative.

Lemma 7. *A real symmetric matrix* $\mathbf{M}$ *is PSD if and only if* $\lambda_1(\mathbf{M}) \geq 0$*.*

Proof. If $\mathbf{M}$ is PSD then for $v_1 = v_1(\mathbf{M})$ and $\lambda_1 = \lambda_1(\mathbf{M})$ we have

$$0 \leq v_1^\top \mathbf{M} v_1 = \lambda_1 \cdot v_1^\top v_1 = \lambda .$$

On the other hand if $\lambda_1(\mathbf{M}) \geq 0$ then for all x we have

$$x^\top \mathbf{M} x = \sum_{i \in [n]} \lambda_i(\mathbf{M}) \cdot \left[v_i(\mathbf{M})^\top x\right]^2 \geq \lambda_1(\mathbf{M}) \sum_{i \in [n]} \left[v_i(\mathbf{M})^\top x\right]^2 \geq 0 .$$

□

With this in mind, what is $\lambda_1(\mathcal{L})$? Well, we have already shown that $\mathcal{L}\vec{1} = \vec{0}$ and consequently $\mathcal{L}$ has a 0 eigenvalue. Furthermore, since all eigenvalues of $\mathcal{L}$ are non-negative this implies that $\lambda_1(\mathcal{L}) = 0$.

Now what is $\lambda_2(\mathcal{L})$? Can it be the case that $\lambda_2(\mathcal{L}) = 0$? Or more broadly, what are the 0 eigenvalues of $\mathcal{L}$, or what is $\ker(\mathcal{L})$?

A common starting point in spectral graph theory to address such questions is to look at the quadratic form of $\mathcal{L}$. Suppose $x \in \ker(\mathcal{L})$ with $x \neq 0$, i.e. $\mathcal{L}x = \vec{0}$. In this case

$$0 = x^\top \mathcal{L} x = \sum_{\{i,j\} \in E} w_{ij} (x_i - x_j)^2 .$$

Now, since $x \neq 0$ there must be some $i \in V$ with $x_i \neq 0$. Given this, what are the values of x_j for $j \in N(i)$. Note that if $x_j \neq x_i$ then since $\{i,j\} \in E$, $w_{ij} > 0$, and $(x_i - x_j)^2 > 0$ we have that $x^\top \mathcal{L} x > 0$. Consequently, $x_j = x_i$ for all $j \in N(i)$. Repeating this argument we see that $x_j = x_i$ for all j that are neighbors of neighbors of i. As we have seen, repeating this argument ultimately implies that $x_j = x_i$ for all j that are in the same connected component as i. Moreover, it is not to hard to see that this property suffices for a vector to be in the kernel of $\mathcal{L}$.

Lemma 8. *$x \in \ker(\mathcal{L}(G))$ if and only if $x_i = x_j$ for all i and j in the same connected component of G.*

Proof. If $x \in \ker(\mathcal{L}(G))$ then as we have seen $x_i = x_j$ for all i that are neighbors of neighbors of neighbors etc. of j, i.e. when i and j are in the same connected component. On the other hand if $x_i = x_j$ for all i and j in the same connected component of G then since $\{i,j\} \in E$ implies that i and j are in the same connected component we have that $(x_i - x_j)^2 = 0$ for such $\{i,j\} \in E$ and therefore $\sum_{\{i,j\}\in E} w_{ij}(x_i - x_j)^2 = x^\top \mathcal{L} x = 0$. □

Thinking about this just a little bit more, we see that the connected components of G essentially characterize the kernel of $\mathcal{L}$.

Lemma 9. *Let $S_1, ..., S_k$ be a partition of V into its connected components. Then $\{\vec{1}_{s_i}, ..., \vec{1}_{s_k}\}$ is a basis for $\ker(\mathcal{L})$ and therefore $\dim(\ker(\mathcal{L}))$ is the number of connected components in G.*

Proof. As we have seen each $\vec{1}_{s_i} \in \ker(\mathcal{L})$. Furthermore, the $\vec{1}_{s_i}$ are clearly orthogonal. Finally, for $x \in \ker(\mathcal{L})$ since x has the same value on all vertices in each connected component, if we let α_i be the value of x_i on some $v \in S_i$ we have that $x = \sum_{i\in[k]} \alpha_i \vec{1}_{S_i}$ and therefore $x \in \mathrm{span}(\vec{1}_{S_1}, ..., \vec{1}_{S_k})$. □

Consequently, we have that $\lambda_2(\mathcal{L}(G)) \neq 0$ if and only if G is connected.

5 λ_2 and Graph Connectivity

In the last section we saw that $\lambda_2(\mathcal{L}(G))$ is non-zero if and only if G is connected. Consequently, if G is connected and λ_2 is "large" in some sense, we might hope to show that G is "well connected" in some sense. Alternatively, when λ_2 is small we might hope to find a cut in the graph that "proves" that G is not well connected. So what notion of connectivity or what kind of cuts should we hope to find?

5.1 Sparsest Cut

To think about this, let's think a little bit about our intuitive notations of graph clustering and connectedness. Intuitively, we might think that well connected graphs have no trivial subdivision, everything is connected to everything so there is no great way to cut it up. Therefore, when a graph is not well connected, we could look for a good non-trivial cut to decompose the graph into natural small pieces.

One type of cut we have seen to compute before is the global minimum cut in a graph. However, if we have a graph that consists of two complete graphs on n vertices and an isolated vertex and then we add a small number of edges between the complete graphs but only one edge to the isolated vertex, the global minimum cut would just remove the edge to the isolated vertex. However, intuitively we would think that a more natural cut would be to remove just the edges connecting the two complete graphs.

The problem with minimum cut in the above example is that we did not try to keep the size of our cuts balanced, i.e. having about the same number of vertices. Consequently, we could instead look for an approximate bisection of the graph, i.e. we could look for the cut that removes the least edges where the number of the vertices in the cut is between $|V|/2$ and $|V|/c$ for some constant c. However, if our graph consists of the complete graph on n vertices and the complete graph on $\sqrt{n}$vertices with these two graphs connected by a single edge, then any approximate bisection will cut a large chunk of the complete graph. This is problematic, as just removing the single edge seems more desirable.

This thought experiments motivates the following natural definition of a good clustering cut, known as the *sparsest cut*. Intuitively we could ask for a cut that minimizes the ratio of the weight of the edges cut, to the size of the smaller side of the cut. This notion of cut sparsity and sparsest cut we define formally below.

Definition 10 (Cut Sparsity)**.** For $S \subseteq V$ with $S \notin \{\emptyset, V\}$ we define the *cut sparsity* of S by

$$\sigma(S) = \frac{w(\partial(S))}{\min\{|S|, |V \setminus S|\}}$$

and we define the value of the *sparsest cut* in G by

$$\sigma(G) = \min_{S \notin \{\emptyset, V\}} \sigma(S) \,.$$

Another way we could have reasoned about well-connected graphs would be by comparing our graph to an idealized graph, i.e. a graph that is as well connected as possible. A natural candidate for this would be the *complete graph on n-vertices*, denoted K_n. Thus we could have asked what do the size of the cuts in G look like compared to the size of cuts in K_n? Now recall that the complete graph has an edge between every pair of vertices in the graph with weight one. Consequently, for $S \subseteq V$ we have

$$w_{K_n}(\partial_{K_n}(S)) = \sum_{i \in S} \sum_{j \notin S} 1 = |S| \cdot |V \setminus S| \,.$$

Now, note that trivially $|S| \cdot |V \setminus S| = \min\{|S|, |V \setminus S|\} \cdot \max\{|S|, |V \setminus S\}$. Furthermore, $\max\{|S|, |V \setminus S\} \in [\frac{n}{2}, n]$. Therefore, up to a factor of 2 we see that the value of the sparsest cut in G is simply $\frac{1}{n}$ times how much larger all cuts are in G then in K_n. Consequently, sparsest cut is approximately capturing the ratio of cut sizes in G to idealized cut sizes, as the would be in the complete graph.

Consequently, computing the sparsest cut in a graph, or certifying that the value is large and therefore that a graph seems well-connected seems useful. However, unfortunately there is no known polynomial time algorithm for computing $\sigma(G)$. Moreoever, as we will see later in the class there is a line of work providing evidence that computing $\sigma(G)$ exactly in the worst case is incredibly computationally expensive. That said, even approximating $\sigma(G)$ could provide a useful characterization of the connectivity of a graph or ways of clustering the graph.

Can we possibly use λ_2 and v_2 for this purpose? We will show that this is possible. This result is a classic result in the area of spectral graph theory known as Cheeger's inequality. We will prove this formally next lecture. In the remainder of this lecture we lay the groundwork for this result and formally state it.

5.2 Variational Characterization of Eigenvalues

One immediate roadblock towards connecting λ_2 and v_2 towards the connectivity of a graph or problems like sparsest cut is that so far we have only characterized λ_2 as the solution to a particular linear system $\mathcal{L}v_2 = \lambda_2 v_2$ for some value of λ_2. However, the sparsest cut problem is a natural optimization problem over the space of cuts. Thus, as a first step towards connecting the spectrum of a Laplacian to optimization problems (and therefore combinatorial properties of graphs more broadly) we would like to view eigenvalues and eigenvectors as the solution to optimization problems. Fortunately, there is such a characterization known as the variational characterization of eigenvalues.

The idea behind this characterization is to look at the ratio of the quadratic form of a real symmetric matrix to the quadratic form of the identity matrix (i.e. the vectors ℓ_2 norm). This ratio is known as the *Rayleigh quotient*

Definition 11 (Rayleigh quotient)**.** For a symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$ the Rayleigh quotient is defined as

$$R_{\mathbf{A}}(x) \stackrel{\text{def}}{=} \frac{x^\top \mathbf{M} x}{x^\top x}$$

By the spectral theorem we know that

$$R_{\mathbf{M}}(x) = \frac{x^\top \mathbf{M} x}{x^\top \mathbf{I} x} = \frac{x^\top \left[\sum_{i\in[n]} \lambda_i(\mathbf{M}) \cdot v_i(\mathbf{M}) v_i(\mathbf{M})^\top\right] x}{x^\top \left[\sum_{i\in[n]} v_i(\mathbf{M}) \cdot v_i(\mathbf{M})^\top\right] x} = \frac{\sum_{i\in[n]} \lambda_i(\mathbf{M}) \cdot (v_i(\mathbf{M})^\top x)^2}{\sum_{i\in[n]} (v_i(\mathbf{M})^\top x)^2} .$$

This, shows that the Rayleigh quotient is essentially the ratio of the the contribution of x to each eigendirection as scaled by λ_i as compared to it being scaled by 1.

Now, clearly, $R_{\mathbf{A}}(v_1(\mathbf{A})) = \lambda_1(\mathbf{A})$ and thus $\min_{x\neq 0} R_{\mathbf{A}}(x) \leq \lambda_1(\mathbf{A})$. However, we can in fact show that this inequality is tight by the following simple lemma.

Lemma 12 (Dan's Favorite Inequality[2])**.** *For any $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n_{>0}$ we have that*

$$\min_{i\in[n]} \frac{a_i}{b_i} \leq \frac{\sum_{i\in[n]} a_i}{\sum_{i\in[n]} b_i} \leq \max_{i\in[n]} \frac{a_i}{b_i}$$

Proof. Clearly

$$\min_{i\in[n]} \frac{a_i}{b_i} \cdot \sum_{i\in[n]} b_i \leq \sum_{i\in[n]} b_i \frac{a_i}{b_i} = \sum_{i\in[n]} a_i = \sum_{i\in[n]} b_i \frac{a_i}{b_i} \leq \max_{i\in[n]} \frac{a_i}{b_i} \cdot \sum_{i\in[n]} b_i .$$

□

From this lemma we can immediately characterize the range of the Rayleigh quotient for non-zero input.

Lemma 13. *For symmetric $\mathbf{M} \in \mathbb{R}^{n\times n}$ we have that the image of $R_{\mathbf{M}}(x)$ for non-zero $x \in \mathbb{R}^n$ is the interval $[\lambda_1(\mathbf{M}), \lambda_n(\mathbf{M})]$.*

Proof. By Lemma 12 and the fact that

$$R_{\mathbf{M}}(x) = \frac{\sum_{i\in[n]} \lambda_i(\mathbf{M}) \cdot (v_i(\mathbf{M})^\top x)^2}{\sum_{i\in[n]} (v_i(\mathbf{M})^\top x)^2}$$

immediately we have that the range of the Rayleigh quotient for non-zero input lie with the eigenvalues, i.e. $R_{\mathbf{M}}(x) \in [\lambda_1(\mathbf{M}), \lambda_n(\mathbf{M})]$ for all $x \neq 0$. Furthermore, since for $x = \alpha \cdot v_1(\mathbf{M}) + (1-\alpha) \cdot v_n(\mathbf{M})$ for $\alpha \in [0,1]$ we have

$$R_{\mathbf{A}}(x) = \frac{\alpha^2 \cdot \lambda_1 + (1-\alpha)^2 \cdot \lambda_n}{\alpha^2 + (1-\alpha)^2}$$

and consequently $R_{\mathbf{A}}(x)$ can take on any value in the interval $[\lambda_1(\mathbf{M}), \lambda_n(\mathbf{A})]$. □

From this reasoning we also see that maximizing and minimizing the Rayleigh quotient corresponds to computing the smallest and largest eigenvalues of a matrix respectively. With a little more thought we can also derive similar optimization problems for computing intermediate eigenvalues this gives the variational characterization of eigenvalues we were looking for.

[2]This name is chosen as the $n = 2$ case (which immediately implies this case) is called "Dan's Favorite Inequality" in lectures notes by Daniel Spielman. I must confess, I do not know if this is actually Dan's favorite inequality.

Lemma 14 (Variational Characterization of Eigenvalues). *For symmetric* $\mathbf{M} \in \mathbb{R}^{n \times n}$ *we have that*

$$\lambda_1(\mathbf{M}) = \min_{x \neq 0} \frac{x^\top \mathbf{M} x}{x^\top x} \quad \text{and } \lambda_n(\mathbf{M}) = \max_{x \neq 0} \frac{x^\top \mathbf{M} x}{x^\top x}$$

and for $i \in [n]$ *with* $i \neq 1$ *and* $i \neq j$

$$\lambda_i(\mathbf{M}) = \min_{\substack{x \neq 0 \\ x \perp v_1(\mathbf{M}), ..., \lambda_{i-1}(\mathbf{M})}} \frac{x^\top \mathbf{M} x}{x^\top x} = \max_{\substack{x \neq 0 \\ x \perp \lambda_{i+1}(\mathbf{A}), ..., \lambda_n(\mathbf{A})}} \frac{x^\top \mathbf{M} x}{x^\top x}.$$

Furthermore, for all $i \in [n]$ *we have*

$$\lambda_i = \min_{i \text{ dimensional supace } S \subseteq \mathbb{R}^n} \left[\max_{x \in S : x \neq 0} \frac{x^\top \mathbf{M} x}{x^\top x} \right] = \max_{n+1-i \text{ dimensional supace } S \subseteq \mathbb{R}^n} \left[\min_{x \in S : x \neq 0} \frac{x^\top \mathbf{M} x}{x^\top x} \right]$$

Furthermore, the associated eigenvectors v_i *are maximizing and minimizing arguments for* x *in these programs.*

This gives us several ways to look at computing the eigenvalues as minimization or maximization problems.

5.3 λ_2 as Sparsest Cut Relaxation

With the variational characterization of λ_2 in hand we can view view λ_2 as a relaxation of the problem of computing the value of the sparsest cut, i.e. a similar optimization problem over a larger domain. To see this, suppose that our graph is connected. In that case $v_1 = \vec{1}$ and we have

$$\lambda_2(\mathcal{L}) = \min_{x \perp \vec{1}} \frac{x^\top \mathcal{L} x}{x^\top x}$$

Now we want to relate this to the complete graph as discussed. However, the Laplacian of the complete graph has a particularly nice form. Note that

$$\mathcal{L}(K_n)_{ij} = \begin{cases} -1 & i \neq j \\ n-1 & \text{otherwise} \end{cases}$$

and therefore

$$\mathcal{L}(K_n) = n \cdot \mathbf{I} - \vec{1}\vec{1}^\top$$

in other words, the Laplacian of the complete graph is simply a rank one update by the outter product of the all ones vector from the identity matrix. This is interesting as we know that a Laplacian has to have a kernel and this shows that the complete graph is as close to the identity as you could be in some sense, while still having the all ones vector in the Kernel. Our hope is that we can characterize graphs with large λ_2 similarly.

Note this also implies that if $x \perp \vec{1}$ then $x^\top \mathcal{L}(K_n) x = n \cdot x^\top x$ and therefore

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

where in the last step we used that adding a multiple of the all ones vector to x does not change $x^\top \mathcal{L} x$ or $x^\top \mathcal{L}(K_n) x$. Now, if we want to relate this to $\sigma(G)$ we can simply restrict the set of vectors we are optimizing over. Note that the set $x \in \{0,1\}^n$ with $x \notin \{\vec{0}, \vec{1}\}$ is a subset of $x \notin \text{span}(\vec{1})$. Consequently, we have

$$\lambda_2(\mathcal{L}) \leq n \cdot \min_{x \in \{0,1\}^n : x \notin \text{span}(\vec{1})} \frac{x^\top \mathcal{L} x}{x^\top \mathcal{L}(K_n) x} = n \cdot \min_{S \subseteq V : S \notin \{\emptyset, V\}} \frac{w(\partial(S))}{w_{K_n}(\partial_{K_n}(S))}$$

However, as we have discussed, we know

$$w_{K_n}(\partial_{K_n}(S)) = |S| \cdot |V \setminus S| \geq \frac{n}{2} \cdot \min\{|V|, |V \setminus S|\}$$

and therefore

$$\lambda_2(\mathcal{L}) \leq 2 \cdot \min_{S \subseteq V : S \notin \{\emptyset, V\}} \frac{w(\partial(S))}{\min\{|S|, |V \setminus S|\}} = 2 \cdot \sigma(G)\,.$$

Thus we have shown that up to a factor of 2 it is the case that $\lambda_2(\mathcal{L})$ is less than the sparsity of the graph. However, it can be shown that λ_2 is not too much less than $\sigma(G)$, this is known as Cheeger's inequality and we will prove variants of it next class.

Theorem 15 (Cheeger's Inequality)**.** *For simple, undirected, positive weighted, connected G with maximum degree $d_{\max}$ it is the case that*

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

Interestingly, our proof of this will be constructive. We will show how to provide a cut of sparsity $\frac{\sigma(G)^2}{2 \cdot d_{\max}}$ given v_2 and therefore, a multiplicative approximation the sparsest cut of a graph can be computed in polynomial time. In the remainder of this note we provide some bounds on these quantities so that we can better think about the quality of this approximation.

5.4 Eigenvalue and Sparsity Bounds

To understand the quality of Cheeger's inequality, let's first look at the range of values $\sigma(S)$ can take. Let $d_{\max} = \max_{i \in V} \deg(i)$ be the largest degree of any vertex and let $d_{\min} = \min_{i \in V} \deg(i)$ be the smallest degree of any vertex. Now, note that

$$\sigma(S) = \frac{w(\partial_G(S))}{\min\{|S|, |V \setminus S|\}} \leq \frac{\min\{|S|, |V \setminus S|\} \cdot d_{\max}}{\min\{|S|, |V \setminus S|\}} \leq d_{\max}$$

as in the very worst all the edges leaving each vertex in $|S|$ or $|V \setminus S|$ cross the induced cut. Furthermore, clearly for some $\{i\}$ it is the case that $\sigma(\{i\}) = d_{\max}$. Lastly, in the complete graph we have

$$\sigma(K_n) = \min_{S \subseteq V \,:\, S \notin \{\emptyset, V\}} \sigma(S) = \frac{|S| \cdot |V \setminus S|}{\min\{|S|, |V \setminus S|\}} = \min_{S \subseteq V \,:\, S \notin \{\emptyset, V\}} \cdot \max\{|S|, |V \setminus S\} = \lceil \frac{n}{2} \rceil\,.$$

as $n-1$ is the degree of each vertex in K_n we have that it can actually be the case that $\sigma(G) = \Omega(d_{\max})$. Consequently, if we want to normalize $\sigma(G)$ to make it at most 1 we could divide by $d_{\max}$. Then Cheeger's inequality yields

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

Consequently, we are saying that given $\lambda_2(G)$ we can approximate $\sigma(G)/d_{\max}$ up to a constant factor times $(\sigma(G)/d_{\max})$. However,

$$\sigma(G) = \min_{S \subseteq V \,:\, S \notin \{\emptyset, V\}} \sigma(S) \leq \min_{i \in V} \sigma(i) = \min_{i \in V} \deg(\{i\}) = d_{\min}\,.$$

Thus, when $d_{\min}/d_{\max}$ is a constant, Cheeger's inequality gives a constant approximation when $\sigma(G)$ is nearly as large as it could possibly be. However, the approximation quality decays as $\sigma(G)$ decreases and when $d_{\min}/d_{\max}$ is small there is no hope of a constant approximation. As we will see next lecture, there is a related notion of connectivity that we can also use to reason about the quality of well-connected graphs with non-uniform degrees without such a loss.

Lastly, for completeness, let's take a look at how large λ_n can be. We have

$$\lambda_n(\mathcal{L}(G)) = \max_{x \neq 0} \frac{x^\top \mathcal{L} x}{x^\top x} = \max_{x \neq 0} \frac{\sum_{\{i,j\} \in E} w_{ij}(x_i - x_j)^2}{\sum_{i \in V} x_i^2} \leq \max_{x \neq 0} \frac{\sum_{\{i,j\} \in E} w_{ij}(2x_i^2 + 2x_j^2)}{\sum_{i \in V} x_i^2}$$
$$= \max_{x \neq 0} \frac{2 \cdot \sum_{i \in V} \deg(i) \cdot x_i^2}{\sum_{i \in V} x_i^2} = 2 \cdot d_{\max} \,.$$

where we used that $(x_i - x_j)^2 \leq 2x_i^2 + 2x_j^2$ as $2x_i^2 + 2x_j^2 - (x_i - x_j)^2 = (x_i + x_j)^2 \geq 0$. Furthermore, clearly $\lambda_n(\mathcal{L}(G)) \geq d_{\max}$ by looking at the Rayleigh quotient of indicator vectors, as $R_{\mathcal{L}}(\vec{1}_i) = \deg(\{i\})$. This should start to give you a sense of how to manipulate the quadratic form of Laplacian.

We conclude by noting that it can be shown that $\lambda_n(\mathcal{L}(G)) = d_{\max}$ when all these inequalities are tight, which happens precisely when G contains a bipartite connected component where all vertices in that component have degree $d_{\max}$ and there is even a Cheeger-like inequality relating λ_n in this case to the largest cut in a graph. It is an active area of research to further extend these sorts of relations and we will discuss all of this more next class.