Lecture 9

1 Sparse cuts and Cheeger's inequality

Consider the problem of partitioning a given graph $G(V, E)$ into two or more 'large' pieces by deleting a 'few' edges. If edges in the graph represent some notion of similarity or closeness between the vertices, then such a decomposition yields a good clustering of the vertices in the graph. The quality of such a decomposition can be quantified in terms of the edge expansion which is defined next.

For simplicity we assume that G is a d-regular graph, let $(S, V - S)$ be a partition of the vertex set V. The edge expansion of the partition $(S, V - S)$ is,

$$h(S) = \frac{|E(S, V - S)|}{d \min(|S|, |V - S|)} \tag{1}$$

The edge expansion of the graph G is the minimum expansion over all partitions of V,

$$h(G) = \min_{S \subseteq V} h(S) \tag{2}$$

Another way to describe the edge expansion is the following: We wish to break off a large part of the graph G by cutting only a small number of edges. Given that we get to keep the smaller of the two pieces, the edge expansion is the ratio of the number of edges cut to the number of edges that we get to keep. (including the cut edges).

The conductance $\phi(S)$ of a partition $(S, V - S)$ is a measure closely related to the edge expansion $h(S)$. The conductance of G is the minimum conductance over all partitions of V,

$$\phi(G) = \min_{S \subseteq V} \frac{n|E(S, V - S)|}{d|S||V - S|} \tag{3}$$

The conductance of a partition $(S, V - S)$ is the edge expansion $h(S)$ multiplied by $n/|\overline{S}|$ where $|\overline{S}| \geq n/2$ is the size of the larger piece.

The conductance approximates the edge expansion within a factor of 2 for all partitions, so reasoning about the two measures is equivalent.

$$h(S) \leq \phi(S) \leq 2h(S) \tag{4}$$

The edge expansion of a graph is closely related to the second largest eigenvalue of its adjacency matrix. The connection was first discovered by Cheeger in differential geometry, relating the isoperimetric properties of Riemannian manifolds to the second eigenvalue of Laplacian operators defined on them. The results were introduced to computer science by Alon, and have found several applications since then including work by Jerrum and Sinclair on bounding the mixing time of Markov chains and by Malik and Shi on image segmentation.

1.1 The spectrum of a graph

The adjacency matrix of G is the $n \times n$ matrix with entry $A_{ij} = 1$ if (i, j) is an edge and 0 otherwise. The adjacency matrix is normalized to $M = A/d$ such that all the rows of M sum to 1. The matrix M is real and symmetric so it has an orthonormal set of eigenvectors v_i with corresponding eigenvalues λ_i for $i \in [n]$ by the spectral theorem. Wlog we can assume that the eigenvectors are real i.e. $v_i \in \mathbb{R}^n$.

Claim 1
If v, v' are eigenvectors of a real symmetric matrix M with distinct eigenvalues λ, λ' then $v^t v' = 0$.

Proof: The proof follows by evaluating the expression $v^t M v'$ in two different ways using the fact that $M^T = M$,

$$\begin{aligned} v^t M v' &= \lambda' v^t v' \\ v^t M v' &= M^T v^t v' = \lambda v^t v' \end{aligned}$$

As λ, λ' are distinct, the vectors v, v' must be orthogonal. □

The claim says that there is a basis of eigenvectors v_i in which the action of M is to shrink or expand the basis vectors. The matrix M is diagonal in the basis of eigenvectors,

$$M = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{5}$$

The basis of eigenvectors is uniquely determined if all the eigenvalues are distinct. If an eigenvalue λ_i has multiplicity d the space of vectors with eigenvalue λ_i has dimension d. A basis of eigenvectors can be defined by choosing any orthonormal basis for each of the eigenspaces.

1.2 Action of the adjacency matrix

It is useful to visualize the action of M as follows: A vector $v \in \mathbb{R}^n$ can be though of as assigning a weight v_i to vertex i in G. The action of M maps v to Mv, which replaces v_i by $\frac{1}{d} \sum_{j \sim i} v_j$. Therefore M acts by replacing the weight of each vertex by the average weight of its neighbors.

We have introduced three different concepts so far: sparse cuts in G, eigenvectors and eigenvalues of the normalized adjacency matrix M and the action of the normalized adjacency matrix. We will relate the three views, drawing on the different perspectives to prove results.

The vector v_1 with all entries equal to $\frac{1}{\sqrt{n}}$ is an eigenvector for a d regular graph with eigenvalue 1. The second largest eigenvalue λ_2 is strictly less than 1 if G is connected, Claim 2

$\lambda_1 = 1$, $v_1 = \frac{1}{\sqrt{n}}\vec{1}$ *and* $|\lambda_k| \leq 1$ *for a* d *regular graph* G*. If the graph* G *is connected,* $\lambda_2 < 1$.

Proof: The vector $v_1 = \frac{1}{\sqrt{n}}\vec{1}$ has unit length and $Mv = v$ by the averaging interpretation of the the action of M, showing that $\lambda_1 = 1$.

The coordinates of the k-th eigenvector v_k scale by a factor of λ_k under averaging over neighbors,

$$(\lambda_k v_k)_i = (Mv_k)_i = \frac{1}{d}\sum_{j\sim i} v_{kj} \tag{6}$$

Let i be a coordinate of v_k having the maximum absolute value, by the triangle inequality we have,

$$|\lambda_k||v_{ki}| \leq \frac{1}{d}\sum_{j\sim i} |v_{kj}| \leq |v_{ki}| \tag{7}$$

It follows that $|\lambda_k| \leq 1$ for all $k \in [n]$. Suppose v is an eigenvector with eigenvalue 1 for a connected graph G. Equality holds in (7) for v, so $v_j = v_i$ for all vertices j adjacent to i. All vertices can be reached by paths starting at i as G is a connected graph so $v_k = v_i$ for all k. It follows that a connected graph has a unique eigenvector with eigenvalue 1, hence $\lambda_2 < 1$.

□

The above claim was proved using the averaging interpretation of the action of M. The diagonal representation of M in the spectral basis (5) yields another characterization of the eigenvalues,

Claim 3

$$\lambda_1 = \sup_{x\in\mathbb{R}^n} \frac{x^T M x}{x^t x}$$

Proof: The matrix M is diagonal in the spectral basis (5) so $x^T Mx$ can be evaluated easily if x is represented in the spectral basis,

$$x^t Mx = \sum_i \lambda_i x_i^2 \leq \lambda_1 \sum_i x_i^2 = \lambda_1 x^t x \tag{8}$$

The inequality is tight for $x_1 = e_1$ i.e. when x is the first eigenvector. □

The quantity $\frac{x^T Mx}{x^T x}$ is called the Rayleigh quotient in the literature. An argument similar to (8) for vectors x such that $x_1 = 0$ in the spectral basis provides a characterization of the second eigenvalue,

$$\forall x \text{ s.t. } x_1 = 0, \qquad x^t Mx = \sum_{i>1} \lambda_i x_i^2 \leq \lambda_2 \sum_i x_i^2 = \lambda_2 x^t x \tag{9}$$

The inequality is tight for $x = e_2$ i.e. when x is the second eigenvector.

The first eigenvector for a d regular graph is $\frac{1}{\sqrt{n}}\vec{1}$ so $x_1 = 0$ in the spectral basis is equivalent to saying that $x \perp \vec{1}$ in the standard basis. Substituting in (9) we have the Rayleigh quotient characterization of the second eigenvalue,

$$\lambda_2 = \sup_{x \perp \vec{1}} \frac{x^T M x}{x^t x} \tag{10}$$

More generally, all the eigenvalues have a similar characterization given by the Courant Fischer theorem (the proof relies on a dimension argument in linear algebra, exercise)

$$\lambda_k = \max_{dim(S)=k} \min_{x \in S} \frac{x^T M x}{x^t x} \tag{11}$$

1.3 Cheeger's inequality

The first eigenvector of a d regular graph is $\frac{1}{\sqrt{n}}\vec{1}$ and does not reveal information about the graph structure. The spectral gap $\mu = \lambda_1 - \lambda_2$ is the difference between the first two eigenvalues. The spectral gap reveals information about the connectivity, for example $\mu = 0$ if and only if G has more than one connected component. Cheeger's inequality provides the connection between the spectral gap and edge expansion,

$$\frac{\mu}{2} = \frac{1-\lambda_2}{2} \leq h(G) \leq \sqrt{2(1-\lambda_2)} = \sqrt{2\mu} \tag{12}$$

We illustrate Cheeger's inequality with the examples of the d dimensional hypercube and the cycle, showing that both sides of the inequality are tight. The inequality will be proved over the next few lectures.

1.4 The hypercube

The d dimensional hypercube has $V = \{0,1\}^d$ with $(x, y) \in E$ if x and y represented as binary strings differ in exactly one bit. The number of vertices is $n = 2^d$, each vertex has degree d so the number of edges is $d2^{d-1}$. A way to picture the hypercube is the following: The d dimensional hypercube is built by taking two copies of a $d-1$ dimensional hypercube and connecting the corresponding vertices.

Edge expansion: The i-th coordinate cut in the hypercube is defined as $S_i := \{x \in \{0,1\}^d \mid x_i = 0\}$. The coordinate cuts achieve the minimum value for the edge expansion. This will follow from Cheeger's inequality once we obtain the spectrum of the hypercube,

$$h(G) = \frac{|E(S_i, \overline{S_i})|}{d|S_i|} = \frac{2^{d-1}}{d2^{d-1}} = \frac{1}{d} = \frac{1}{\log n} \tag{13}$$

Vertex expansion: The ball cut $S := \{x \in \{0,1\}^d \mid \sum x_i \leq d/2\}$ consists of strings that have at most $d/2$ ones. The cut S achieves the smallest possible vertex expansion: more generally an isoperimetric theorem for the hypercube says that a cut having $1 + \binom{n}{1} + \cdots + \binom{n}{k}$ vertices must have at least $\binom{n}{k}$ vertices at its boundary.

If the vertices of the hypercube are permuted randomly it turns out that is is difficult to reconstruct the coordinate cuts, however it is easy to recover good approximations to the ball cuts. What is the edge expansion achieved by the ball cuts?

A vertex at the boundary of S has exactly $d/2$ neighbors in S, so the edge expansion is given by,

$$h(S) = \frac{E(S, \overline{S})}{d|S_i|} = \frac{d|S|}{d2^d} \approx \frac{1}{\sqrt{d}} = \frac{1}{\sqrt{\log n}} \tag{14}$$

The size of S is equal to $\binom{d}{d/2} \approx \frac{2^d}{\sqrt{d}}$ where the approximation $\log d! \approx d \log d - d$ is used. [The approximation is known as Stirling's formula in the literature, exercise if you have not seen this before].

1.4.1 The spectrum of the hypercube:

Finding the spectrum: Recall that there are $\binom{d}{1}$ coordinate cuts on the hypercube given by $S_i = \{x \in \{0,1\}^d \mid x_i = 0\}$. Consider the characteristic vectors v_i of the coordinate cuts where $v_i(x) = 1$ if $x \in S_i$ and $v_i(x) = -1$ if $x \in \overline{S_i}$. Each vertex in S_i has $d-1$ neighbors in S_i and one neighbor in $\overline{S_i}$, the averaging interpretation of the action of the adjacency matrix shows that $Mv_i = (1 - 2/d)v_i$. The coordinate cuts v_i are eigenvectors of the hypercube with eigenvalue $1 - 2/d$.

Consider the $\binom{d}{2}$ cuts on the hypercube given by $S_{ij} = \{x \in \{0,1\}^d \mid (x_i \oplus x_j) = 1\}$. These cuts are obtained by considering the 4 hypercubes of dimension $d-2$ contained in the d dimensional hypercube. Each vertex in S_{ij} has $d-2$ neighbors in S_{ij} and two neighbors in $\overline{S_{ij}}$, the averaging interpretation of the action of the adjacency matrix shows that $Mv_{ij} = (1 - 4/d)v_{ij}$, where v_{ij} is the characteristic vector of S_{ij}. The vectors v_{ij} are mutually orthogonal, showing that there are $\binom{d}{2}$ eigenvectors for the hypercube with eigenvalue $1 - 4/d$.

Similarly by looking at the dimension $d-k$ hypercubes inside the d dimensional hypercube we find $\binom{d}{k}$ eigenvectors with eigenvalue $1 - 2k/d$. The sum of the binomial coefficients $\sum \binom{d}{i} = 2^n$, so we have found the complete spectrum of the hypercube.

A histogram of the spectral profile of the hypercube looks like a plot of the binomial distribution scaled to lie in $[-1, 1]$. A cleaner way to obtain the spectrum of the hypercube is to observe that the hypercube is a Cayley graph for the group $\mathbb{Z}_2^d$, the spectra of Cayley graphs can be determined easily as the eigenvectors are the Fourier basis vectors (exercise for theorists).

The second eigenspace: The second eigenspace of the hypercube has dimension d and is spanned by the coordinate cuts. If the hypercube is randomly permuted so that the vertex labels are lost, an eigenvalue finding program will output some linear combination of the coordinate cuts as the second eigenvector. The ball cut in (14) is a linear combination of the coordinate cuts with coefficients $1/2$. A random linear combination of the coordinate cuts has expansion similar to the ball cuts.

Tightness of Cheeger's inequality: The hypercube is an example for which the left side of Cheeger's inequality is tight. We showed that $\frac{1-\lambda_2}{2} = \frac{1}{d}$ and equation (13) shows that the edge expansion of the coordinate cuts is $h(S_i) = \frac{1}{d}$. Equality holds in the left side of Cheeger's inequality for the hypercube. It follows that the coordinate cuts are the sparsest cuts for the hypercube.

1.5 The cycle

The n cycle has vertex set $[n]$ and edges $(i, i+1) \mod n$ for $i \in [n]$. The edge expansion of the cycle $h(C) = 2/n$ and the sparsest cut is the partition of the cycle into two equal halves. We will show that $\lambda_2 \geq 1 - O(\frac{1}{n^2})$ for the cycle, implying that the right side of Cheeger's inequality $h(C) \leq \sqrt{2(1-\lambda_2)}$ is tight for the cycle.

Proof of the tightness of Cheeger: A quick way to show that the second eigenvector for the cycle is large is to find a vector $x \perp 1$ having a high Rayleigh quotient. A vector with high Rayleigh quotient will be approximately invariant under averaging over neighbors. Consider the following vector,

$$x_i = \begin{cases} i - n/4 & \text{if } i \leq n/2 \\ 3n/4 - i & \text{if } i > n/2 \end{cases}$$

The vector x is chosen to be orthogonal to $\vec{1}$ and the coordinates of x are approximately invariant under averaging over neighbors,

$$(Mx)_i = \begin{cases} -n/4 + 1/2 & \text{if } i = 1, n \\ n/4 - 1 & \text{if } i = n/2 \\ x_i & \text{otherwise} \end{cases}$$

Using $x^t x = \sum_{i \in [n/4]} 16i^2 = O(n^3)$ the value of the Rayleigh quotient can be computed as follows,

$$\frac{x^t M x}{x^t x} = \frac{x^t x - O(n)}{x^t x} = 1 - O\left(\frac{1}{n^2}\right)$$

The second eigenvalue λ_2 must be greater than the Rayleigh quotient of x, showing that the right side of Cheeger's inequality is tight for the cycle.

Alternative proof by computing the spectrum: We will write down the eigenvalues and eigenvectors for the n cycle explicitly. A cleaner way to obtain the spectrum of the n cycle is to observe that it is a Cayley graph for $\mathbb{Z}_n$ and the eigenvectors are Fourier basis vectors.

The eigenvalues of the n cycle are $\cos(2\pi k/n)$ for $0 \leq k \leq n-1$. The vector v with coordinates $v_i = \cos(2\pi ki/n)$ is an eigenvector of the cycle with eigenvalue $\cos(2\pi k/n)$. This follows from the trigonometric identity $\cos(A+B) = \cos A \cos B - \sin A \sin B$,

$$\cos\left(\frac{2\pi k(i+1)}{n}\right) + \cos\left(\frac{2\pi k(i-1)}{n}\right) = 2\cos\left(\frac{2\pi k}{n}\right)\cos\left(\frac{2\pi ki}{n}\right)$$

The vector w with coordinates $w_i = \sin(2\pi ki/n)$ is also an eigenvector of the cycle with eigenvalue $\cos(2\pi k/n)$, this can be seen using the identity $\sin(A+B) = \sin A \cos B + \cos A \sin B$.

The eigenspaces of the cycle corresponding to non ± 1 eigenvalues are two dimensional, this also follows from the fact that the eigenvectors of the cycle are the Fourier basis vectors. More generally graphs with symmetries have degenerate eigenspaces, the standard embedding of the cycle has a reflection symmetry about the x axis.

Tightness of Cheeger's inequality: As $n \to \infty$, the spectral gap for the cycle tends to $1 - 2\cos(2\pi/n) = O(1/n^2)$ using the Taylor series expansion $\cos(\delta) = 1 - \frac{\delta^2}{2} + o(\delta^2)$ for $\delta \to 0$.