Lecture 1

In which we describe what this course is about and we review some linear algebra.[1]

1 Overview

This class is about the following topics:

1. *Approximation algorithms for graph partitioning problems.* We will study approximation algorithms for the *sparsest cut* problem, in which one wants to find a cut (a partition into two sets) of the vertex set of a given graph so that a minimal number of edges cross the cut compared to the number of pairs of vertices that are disconnected by the removal of such edges.

   This problem is related to estimating the edge expansion of a graph and to find *balanced separators*, that is, ways to disconnect a constant fraction of the pairs of vertices in a graph after removing a minimal number of edges.

   Finding balanced separators and sparse cuts arises in *clustering* problems, in which the presence of an edge denotes a relation of similarity, and one wants to partition vertices into few clusters so that, for the most part, vertices in the same cluster are similar and vertices in different clusters are not. For example, sparse cut approximation algorithms are used for *image segmentation*, by reducing the image segmentation problem to a graph clustering problem in which the vertices are the pixels of the image and the (weights of the) edges represent similarities between nearby pixels.

   Balanced separators are also useful in the design of divide-and-conquer algorithms for graph problems, in which one finds a small set of edges that disconnects the graph, recursively solves the problem on the connected components,

[1] 

and then patches the partial solutions and the edges of the cut, via either exact methods (usually dynamic programming) or approximate heuristic. The sparsity of the cut determines the running time of the exact algorithms and the quality of approximation of the heuristic ones.

We will study three approximation algorithms:

(a) The *Spectral Partitioning Algorithm*, based on linear algebra;

(b) The *Leighton-Rao Algorithm*, based on a linear programming relaxation;

(c) The *Arora-Rao-Vazirani (ARV) Algorithm*, based on a semidefinite programming relaxation.

The three approaches are related, because the continuous optimization problem that underlies the Spectral Partitioning algorithm is a relaxation of the ARV semidefinite programming relaxation, and so is the Leighton-Rao relaxation. Rounding the Leighton-Rao and the ARV relaxations raise interesting problems in metric geometry, some of which are still open.

2. *Explicit Constructions of Bounded-Degree Expanders.* Expander graphs are graphs with very strong connectivity and "pseudorandomness" properties. Constructions of constant-degree expanders are useful in a variety of applications, from the design of data structures, to the derandomization of algorithms, from efficient cryptographic constructions to being building blocks of more complex quasirandom objects.

   There are two families of approaches to the explicit (efficient) construction of bounded-degree expanders. One is via algebraic constructions, typically ones in which the expander is constructed as a Cayley graph of a finite group. Usually these constructions are easy to describe but rather difficult to analyze. The study of such expanders, and of the related group properties, has become a very active research program, involving mostly ergodic theorists and number theorists. There are also combinatorial constructions, which are somewhat more complicated to describe but considerably simpler to analyze.

3. *Bounding the Mixing Time of Random Walks and Approximate Counting and Sampling.* If one takes a random walk in a regular graph that is connected and not bipartite, then, regardless of the starting vertex, the distribution of the t-th step of the walk is close to the uniform distribution over the vertices, provided that t is large enough. It is always sufficient for t to be quadratic in the number of vertices; in some graphs, however, the distribution is near-uniform even when t is just poly-logarithmic. Among other applications, the study of the "mixing time" (the time that it takes to reach the uniform distribution) of random walks has applications to analyzing the convergence time of certain randomized algorithms.

The design of approximation algorithms for *combinatorial counting* problems, in which one wants to count the number of solutions to a given NP-type problem, can be reduced to the design of *approximately uniform sampling* in which one wants to approximately sample from the set of such solutions. For example, the task of approximately counting the number of perfect matchings can be reduced to the task of sampling almost uniformly from the set of perfect matchings of a given graph. One can design approximate sampling algorithms by starting from an arbitrary solution and then making a series of random local changes. The behavior of the algorithm then corresponds to performing a random walk in the graph that has a vertex for every possible solution and an edge for each local change that the algorithm can choose to make. Although the graph can have an exponential number of vertices in the size of the problem that we want to solve, it is possible for the approximate sampling algorithm to run in polynomial time, provided that a random walk in the graph converges to uniform in time poly-logarithmic in its size.

The study of the mixing time of random walks in graphs is thus a main analysis tool to bound the running time of approximate sampling algorithms (and, via reductions, of approximate counting algorithms).

These three research programs are pursued by largely disjoint communities, but share the same mathematical core.

One direction of *Cheeger's inequality*, for example, which is a basic result in algebraic graph theory, is useful in the construction of expanders because it establishes that good edge-expansion (the property that one is usually looking for, but that is coNP-complete, and thus rather hard to certify) is implied by good spectral expansion (a property that is usually easier to establish and that is in P and thus has short certificates); the other direction of Cheeger's inequality, that good edge expansion implies good spectral expansion, is often used in the study of random walks, because spectral expansion is the property that controls the mixing time of random walks, and in some cases it is easier to prove that a graph has good spectral expansion by proving that it has good edge expansion. Both directions are equivalent to the statement that the nearly-linear-time *spectral partitioning algorithm* achieves a non-trivial approximation for the sparsest cut problem.

In this course we will study these three research programs back-to-back, emphasizing the connections, and providing, when necessary, a "dictionary" to translate the ways the same mathematical facts are thought about in the three communities.

2 Expander Graphs and Sparse Cuts

Before giving the definition of expander graph, it is helpful to consider examples of graphs that are not expanders, in order to gain intuition about the type of "bad examples" that the definition is designed to avoid.

Suppose that a communication network is shaped as a path, with the vertices representing the communicating devices and the edges representing the available links. The clearly undesirable feature of such a configuration is that the failure of a single edge can cause the network to be disconnected, and, in particular, the failure of the middle edge will disconnect half of the vertices from the other half.

This is a situation that can occur in reality. Most of Italian highway traffic is along the highway that connect Milan to Naples via Bologna, Florence and Rome. The section between Bologna and Florence goes through relatively high mountain passes, and snow and ice can cause road closures. When this happens, it is almost impossible to drive between Northern and Southern Italy. Closer to California, I was once driving from Banff, a mountain resort town in Alberta which hosts a mathematical institute, back to the US. Suddenly, traffic on Canada's highway 1 came to a stop. People from the other cars, after a while, got out of their cars and started hanging out and chatting on the side of the road. We asked if there was any other way to go in case whatever accident was ahead of us would cause a long road closure. They said no, this is the only highway here. Thankfully we started moving again in half an hour or so.

Now, consider a two-dimensional $\sqrt{n} \times \sqrt{n}$ grid. The removal of an edge cannot disconnect the graph, and the removal of a constant number of edges can only disconnected a constant number of vertices from the rest of the graph, but it is possible to remove just $\sqrt{n}$ edges, a $1/O(\sqrt{n})$ fraction of the total, and have half of the vertices be disconnected from the other half.

A k-dimensional hypercube with $n = 2^k$ is considerably better connected than a grid, although it is still possible to remove a vanishingly small fraction of edges (the edges of a "dimension cut," which are a $1/k = 1/\log_2 n$ fraction of the total number of edges) and disconnect half of the vertices from the other half.

Clearly, the most reliable network layout is the clique; in a clique, if an adversary wants to disconnect a p fraction of vertices from the rest of the graph, he has to remove at least a $p \cdot (1-p)$ fraction of edges from the graph.

This property of the clique will be our "gold standard" for reliability. The expansion and the sparsest cut parameters of a graph measure how worse a graph is compared with a clique from this point.

Definition 1 (Sparsest Cut) *Let $G = (V, E)$ be a graph and let $(S, V - S)$ be a partition of the vertices (a cut). Then the (normalized) sparsity of the cut is*

$$\sigma(S) := \frac{E(S, V-S)}{|E|} \cdot \left(\frac{|S| \cdot |V-S|}{|V|^2/2} \right)^{-1}$$

where $E(S, V-S)$ is the number of edges in E that have one endpoint in S and one endpoint in $V-S$.

The sparsest cut problem is, given a graph, to find the set of minimal sparsity. The sparsity of a graph $G = (V, E)$ is

$$\sigma(G) := \min_{S \subseteq V : S \neq \emptyset, S \neq V} \sigma(S)$$

That is, we are looking at the ratio between the fraction of edges that need to be removed in order to disconnect S from $V - S$ and the fraction of pairs of vertices that would be so disconnected.

It is more common to define the sparsity as

$$\frac{E(S, V-S)}{|S| \cdot |V-S|}$$

without the normalizing factor $(V^2/2|E|)$; the normalized definition used above yields simpler formulas in some of the applications that we will discuss later.

Note that if G is a d-regular graph, then

$$\sigma(S) := \frac{E(S, V-S)}{\frac{d}{|V|} \cdot |S| \cdot |V-S|}$$

In a d-regular graph, the *edge expansion* of a set of vertices $S \subseteq V$ is the related quantity

$$\phi(S) := \frac{E(S, V-S)}{d \cdot |S|}$$

in which we look at the ratio between the number of edges between S and $V - S$ and the obvious upper bound given by the total number of edges incident S.

The edge expansion $\phi(G)$ of a graph is

$$\phi(G) := \min_{S : |S| \leq \frac{|V|}{2}} \phi(S)$$

the minimum of $\phi(S)$ over all partitions $(S, V - S)$, where $|S| \leq |V - S|$.

(It is common to define the edge expansion without the normalizing factor of d in the denominator.)

We note that for every regular graph G we have that, for every set S such that $|S| \leq |V|/2$,

$$\frac{1}{2}\sigma(S) \leq \phi(S) \leq \sigma(S)$$

and we have $\sigma(S) = \sigma(V - S)$, hence

$$\frac{1}{2}\sigma(G) \leq \phi(G) \leq \cdot\sigma(G)$$

A family of constant degree expanders is a family of (multi-)graphs $\{G_n\}_{n \geq d}$ such that each graph G_n is a d-regular graph with n vertices and such that there is an absolute constant $\phi > 0$ such that $\phi(G_n) \geq \phi$ for every n.

Constant-degree graphs of constant expansion are sparse graphs with exceptionally good connectivity properties. For example, we have the following observation.

Lemma 2 *Let $G = (V, E)$ be a regular graph of expansion ϕ. Then, after an $\epsilon < \phi$ fraction of the edges are adversarially removed, the graph has a connected component that spans at least $1 - \epsilon/2\phi$ fraction of the vertices.*

PROOF: Let d be the degree of G, and let $E' \subseteq E$ be an arbitrary subset of $\leq \epsilon|E| = \epsilon \cdot d \cdot |V|/2$ edges. Let $C_1, \ldots, C_m$ be the connected components of the graph $(V, E - E')$, ordered so that $|C_1| \geq |C_2| \geq \cdots \geq |C_m|$. We want to prove that $|C_1| \geq |V| \cdot (1 - 2\epsilon/\phi)$. We have

$$|E'| \geq \frac{1}{2}\sum_{i \neq j} E(C_i, C_j) = \frac{1}{2}\sum_i E(C_i, V - C_i)$$

If $|C_1| \leq |V|/2$, then we have

$$|E'| \geq \frac{1}{2}\sum_i d \cdot \phi \cdot |C_i| = \frac{1}{2} \cdot d \cdot \phi \cdot |V|$$

but this is impossible if $\phi > \epsilon$.

If $|C_1| \geq |V|/2$, then define $S := C_2 \cup \cdots \cup C_m$. We have

$$|E'| \geq E(C_1, S) \geq d \cdot \phi \cdot |S|$$

which implies that $|S| \leq \frac{\epsilon}{2\phi} \cdot |V|$ and so $C_1 \geq \left(1 - \frac{\epsilon}{2\phi}\right) \cdot |V|$. $\square$

In words, this means that, in a d-regular expander, the removal of k edges can cause at most $O(k/d)$ vertices to be disconnected from the remaining "giant component." Clearly, it is always possible to disconnect k/d vertices after removing k edges, so the reliability of an expander is essentially best possible.

3 Eigenvalues and Eigenvectors

Spectral graph theory studies how the eigenvalues of the adjacency matrix of a graph, which are purely algebraic quantities, relate to combinatorial properties of the graph.

We begin with a brief review of linear algebra.

If $x = a + ib$ is a complex number, then we let $\bar{x} = a - ib$ denote its *conjugate*. If $M \in \mathbb{C}^{m \times n}$ is a matrix, then M^* denotes the conjugate transpose of M, that is, $(M^*)_{i,j} = \overline{M_{j,i}}$. If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ are two vectors, then their inner product is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^* \mathbf{y} = \sum_i \overline{x_i} \cdot y_i \tag{1}$$

Notice that, by definition, we have $\langle \mathbf{x}, \mathbf{y} \rangle = (\langle \mathbf{x}, \mathbf{y} \rangle)^*$ and $\langle \mathbf{x}, \mathbf{x} \rangle = ||\mathbf{x}||^2$.

If $M \in \mathbb{C}^{n \times n}$ is a square matrix, $\lambda \in \mathbb{C}$ is a scalar, $\mathbf{x} \in \mathbb{C}^n - \{\mathbf{0}\}$ is a non-zero vector and we have

$$M\mathbf{x} = \lambda \mathbf{x} \tag{2}$$

then we say that λ is an *eigenvalue* of M and that $\mathbf{x}$ is *eigenvector* of M corresponding to the eigenvalue λ.

When (2) is satisfied, then we equivalently have

$$(M - \lambda I) \cdot \mathbf{x} = \mathbf{0}$$

for a non-zero vector $\mathbf{x}$, which is equivalent to

$$\det(M - \lambda I) = 0 \tag{3}$$

For a fixed matrix M, the function $\lambda \to \det(M - \lambda I)$ is a univariate polynomial of degree n in λ and so, over the complex numbers, the equation (3) has exactly n solutions, counting multiplicities.

If $G = (V, E)$ is a graph, then we will be interested in the adjacency matrix A of G, that is the matrix such that $A_{ij} = 1$ if $(i, j) \in E$ and $A_{ij} = 0$ otherwise. If G is a

multigraph or a weighted graph, then A_{ij} is equal to the number of edges between (i, j), or the weight of the edge (i, j), respectively.

The adjacency matrix of an undirected graph is symmetric, and this implies that its eigenvalues are all real.

Definition 3 *A matrix $M \in \mathbb{C}^{n \times n}$ is Hermitian if $M = M^*$.*

Note that a real symmetric matrix is always Hermitian.

Lemma 4 *If M is Hermitian, then all the eigenvalues of M are real.*

PROOF: Let M be an Hermitian matrix and let λ be a scalar and $\mathbf{x}$ be a non-zero vector such that $M\mathbf{x} = \lambda\mathbf{x}$. We will show that $\lambda = \lambda^*$, which implies that λ is a real number.

We first see that

$$\begin{aligned} & \langle M\mathbf{x}, \mathbf{x} \rangle \\ = & \sum_i \sum_j M^*_{ij} x^*_j x_i \\ = & \sum_i \sum_j M_{ji} x_i x^*_j \\ = & \langle \mathbf{x}, M\mathbf{x} \rangle \end{aligned}$$

where we use the fact that M is Hermitian. Then we note that

$$\langle M\mathbf{x}, \mathbf{x} \rangle = \langle \lambda\mathbf{x}, \mathbf{x} \rangle = \lambda^* ||x||^2$$

and

$$\langle \mathbf{x}, M\mathbf{x} \rangle = \langle \mathbf{x}, \lambda\mathbf{x} \rangle = \lambda ||x||^2$$

so that $\lambda = \lambda^*$. $\square$

We also note the following useful fact.

Fact 5 *If M is an Hermitian matrix, and $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors of different eigenvalues, then $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.*

PROOF: Let $\mathbf{x}$ be an eigenvector λ and $\mathbf{y}$ be an eigenvector of λ', then, from the fact that M is Hermitian, we get

$$\langle M\mathbf{x}, \mathbf{y}\rangle = (M\mathbf{x})^*\mathbf{y} = \mathbf{x}^*M^*\mathbf{y} = \mathbf{x}^*M\mathbf{y} = \langle \mathbf{x}, M\mathbf{y}\rangle$$

but

$$\langle M\mathbf{x}, \mathbf{y}\rangle = \lambda \cdot \langle \mathbf{x}, \mathbf{y}\rangle$$

and

$$\langle \mathbf{x}, M\mathbf{y}\rangle = \lambda' \cdot \langle \mathbf{x}, \mathbf{y}\rangle$$

so that

$$(\lambda - \lambda') \cdot \langle \mathbf{x}, \mathbf{y}\rangle = 0$$

which implies that $\langle \mathbf{x}, \mathbf{y}\rangle = 0$, that is, that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. □

We will be interested in relating combinatorial properties of a graph G, such as connectivity and bipartiteness, with the values of the eigenvalues of the adjacency matrix of G.

A step in this direction is to see the problem of computing the eigenvalues of a real symmetric matrix as the solution to an *optimization* problem.

Theorem 6 (Variational Characterization of Eigenvalues) *Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and $\lambda_1 \leq \ldots \leq \lambda_n$ be its real eigenvalues, counted with multiplicities and sorted in nondecreasing order. Let $\mathbf{x}_1, \cdots, \mathbf{x}_k$, $k < n$, be orthonormal vectors such that $M\mathbf{x}_i = \lambda_i\mathbf{x}_i$ for $i = 1, \ldots, k$. Then*

$$\lambda_{k+1} = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}:\mathbf{x}\perp\mathbf{x}_1,\ldots,\mathbf{x}\perp\mathbf{x}_k} \frac{\mathbf{x}^TM\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

and any minimizer is an eigenvector of λ_{k+1}.

In particular, Theorem 6 implies that

$$\lambda_1 = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}} \frac{\mathbf{x}^TM\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

and, if we call $\mathbf{x}_1$ a minimizer of the above expression, then

$$\lambda_2 = \min_{\mathbf{x}\in\mathbb{R}^n - \{\mathbf{0}\}:\mathbf{x}\perp\mathbf{x}_1} \frac{\mathbf{x}^TM\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

and a minimizer $\mathbf{x}_2$ of the above expression is an eigenvector of $\mathbf{x}_1$, and so on.

In order to prove Theorem 6, we first prove the following result.

Lemma 7 *Let Let $M \in \mathbb{R}^{n\times n}$ be a real symmetric matrix, and let $\mathbf{x}_1, \ldots, \mathbf{x}_k$, $k < n$ be orthogonal eigenvectors of M. Then there is an eigenvector $\mathbf{x}_{k+1}$ of M that is orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$.*

PROOF: Let V be the $(n-k)$-dimensional subspace of $\mathbb{R}^n$ that contains all the vectors orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$. We claim that for every vector $\mathbf{x} \in V$ we also have $M\mathbf{x} \in V$. Indeed, for every i, the inner product of $M\mathbf{x}$ and $\mathbf{x}_i$ is

$$\langle \mathbf{x}_i, M\mathbf{x}\rangle = \mathbf{x}_i^T M\mathbf{x} = (M^T\mathbf{x}_i)^T\mathbf{x} = (M\mathbf{x}_i)^T\mathbf{x} = \lambda_i \mathbf{x}_i^T\mathbf{x} = \lambda \cdot \langle \mathbf{x}_1, \mathbf{x}\rangle = 0$$

Let $B \in \mathbb{R}^{n\times(n-k)}$ be the matrix that computes a bijective map from $\mathbb{R}^{n-k}$ to V. (If $\mathbf{b}_1, \ldots, \mathbf{b}_{n-k}$ are an orthonormal for basis for V, then B is just the matrix whose columns are the vectors $\mathbf{b}_i$.) Let also $B' \in \mathbb{R}^{(n-1)\times n}$ be the matrix such that, for every $\mathbf{y} \in V$, $B'\mathbf{y}$ is the $(n-k)$-dimensional vector such that $BB'\mathbf{y} = \mathbf{y}$. (Let $B' = B^T$ where B is as described above.) Let λ be a real eigenvalue of the real symmetric matrix

$$M' := B'MB \in \mathbb{R}^{(n-k)\times(n-k)}$$

and $\mathbf{y}$ be a real eigenvector of M'.

Then we have the equation

$$B'MB\mathbf{x} = \lambda\mathbf{y}$$

and so

$$BB'MB\mathbf{y} = \lambda B\mathbf{y}$$

Since $B\mathbf{y}$ is orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$, it follows that $MB\mathbf{y}$ is also orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$, and so

$$BB'MB\mathbf{y} = MB\mathbf{y}\ ,$$

which means that we have

$$MB\mathbf{y} = \lambda B\mathbf{y}$$

and, defining $\mathbf{x}_{k+1} := B\mathbf{y}$, we have

$$M\mathbf{x}_{k+1} = \lambda B\mathbf{x}_{k+1}$$

□

We note that Lemma 7 has the following important consequence.

Corollary 8 (Spectral Theorem) *Let $M \in \mathbb{R}^{n\times n}$ be an real symmetric matrix, and $\lambda_1, \ldots, \lambda_n$ be its real eigenvalues, with multiplicities; then there are orthonormal vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^n$ such that $\mathbf{x}_i$ is an eigenvector of λ_i.*

We are now ready to prove Theorem 6.

PROOF:[Of Theorem 6] By repeated applications of Lemma 7, we find $n-k$ orthogonal eigenvectors which are also orthogonal to $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The eigenvalues of this system of n orthogonal eigenvectors must include all the eigenvalues of M, because if there was any other eigenvalue, its eigenvector would be orthogonal to our n vectors, which is impossible. Let us call the additional $n-k$ vectors $\mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$, where $\mathbf{x}_i$ is an eigenvector of λ_i. Now consider the minimization problem

$$\min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\} : \mathbf{x} \perp \mathbf{x}_1, \ldots, \mathbf{x} \perp \mathbf{x}_k} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

The solution $\mathbf{x} := \mathbf{x}_{k+1}$ is feasible, and it has cost λ_{k+1}, so the minimum is at most λ_{k+1}.

Consider now an arbitrary feasible solution $\mathbf{x}$. We can write

$$\mathbf{x} = a_{k+1}\mathbf{x}_{k+1} + \cdots + a_n \mathbf{x}_n$$

and we see that the cost of such a solution is

$$\frac{\sum_{i=k+1}^n \lambda_i a_i^2}{\sum_{i=k+1}^n a_i^2} \geq \lambda_{k+1} \cdot \frac{\sum_{i=k+1}^n a_i^2}{\sum_{i=k+1}^n a_i^2}$$

and so the minimum is also at least λ_{k+1}. Notice also that if $\mathbf{x}$ is a minimizer, that is, if the cost of $\mathbf{x}$ is λ_{k+1}, then we must $a_i = 0$ for every i such that $\lambda_i > \lambda_{k+1}$, which means that $\mathbf{x}$ is a linear combination of eigenvectors of λ_{k+1}, and so it is itself an eigenvector of λ_{k+1}. □

Sometimes it will be helpful to use the following variant of the variational characterization of eigenvalues.

Corollary 9 *Let $M \in \mathbb{R}^{n \times n}$ be a real symmetric matrix, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ its eigenvalues, counted with multiplicities and sorted in nondecreasing order. Then*

$$\lambda_k = \min_{V \ k-\text{dimensional subspace of } \mathbb{R}^n} \max_{\mathbf{x} \in V - \{\mathbf{0}\}} \frac{\mathbf{x}^T M \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

4 The Basics of Spectral Graph Theory

From the discussion so far, we have that if A is the adjacency matrix of an undirected graph then it has n real eigenvalues, counting multiplicities of the number of solutions to $\det(A - \lambda I) = 0$.

If G is a d-regular graph, then instead of working with the adjacency matrix of G it is somewhat more convenient to work with the *normalized Laplacian matrix* of G, which is defined as $L := I - \frac{1}{d} \cdot A$.

In the rest of this section we shall prove the following relations between the eigenvalues of L and certain purely combinatorial properties of G.

Theorem 10 *Let G be a d-regular undirected graph, and $L = I - \frac{1}{d} \cdot A$ be its normalized Laplacian matrix. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ be the real eigenvalues of L with multiplicities. Then*

1. *$\lambda_1 = 0$ and $\lambda_n \leq 2$.*
2. *$\lambda_k = 0$ if and only if G has at least k connected components.*
3. *$\lambda_n = 2$ if and only if at least one of the connected components of G is bipartite.*

Note that the first two properties imply that the multiplicity of 0 as an eigenvalue is precisely the number of connected components of G.

4.1 Proof of Theorem 10

We will make repeated use of the following identity, whose proof is immediate: if L is the normalized Laplacian matrix of a d-regular graph G, and $\mathbf{x}$ is any vector, then

$$\mathbf{x}^T L \mathbf{x} = \frac{1}{d} \cdot \sum_{\{u,v\} \in E} (x_u - x_v)^2 \tag{4}$$

and so

$$\lambda_1 = \min_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}:} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \geq 0$$

If we take $\mathbf{1} = (1, \ldots, 1)$ to be the all-one vector, we see that $\mathbf{1}^T L \mathbf{1} = 0$, and so 0 is the smallest eigenvalue of L, with $\mathbf{1}$ being one of the vectors in the eigenspace of 1.

We also have the following formula for λ_k:

$$\lambda_k = \min_{S \text{ } k-\text{dimensional subspace of } \mathbb{R}^n} \max_{\mathbf{x} \in S - \{\mathbf{0}\}} \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{d \sum_v x_v^2}$$

So, if $\lambda_k = 0$, there must exist a k-dimensional space S such that for every $\mathbf{x} \in S$ we have

$$\sum_{\{u,v\} \in E} (x_u - x_v)^2 = 0 \ ,$$

but this means that, for every $\mathbf{x}$, and for every edge $(u, v) \in E$ of positive weight, we have $x_u = x_v$, and so $x_u = x_v$ for every u, v which are in the same connected component. This means that each $\mathbf{x} \in V$ must be constant within each connected component of G, and so the dimension of V can be at most the number of connected components of G, meaning that G has at least k connected components.

Conversely, if G has at least k connected components, we can let S be the space of vectors that are constant within each component, and S is a space of dimension at least k such that for every element $\mathbf{x}$ of S we have

$$\sum_{\{u,v\} \in E} (x_u - x_v)^2 = 0$$

meaning that S is a witness of the fact that $\lambda_k = 0$.

Finally, to study λ_n, we first note that we have the formula

$$\lambda_n = \max_{\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}} \frac{\mathbf{x}^T L \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$$

which we can prove by using the variational characterization of the eigenvalues of $-L$ and noting that $-\lambda_n$ is the smallest eigenvalue of $-L$.

We also observe that for every vector $\mathbf{x} \in \mathbb{R}^n$ we have

$$2 - \mathbf{x}^T L \mathbf{x} = \frac{1}{d} \sum_{\{u,v\} \in E} (x_u + x_v)^2$$

and so

$$\lambda_n \leq 2$$

and if $\lambda_n = 2$ then there must be a non-zero vector $\mathbf{x}$ such that

$$\sum_{\{u,v\} \in E} (x_u + x_v)^2 = 0$$

which means that $x_u = -x_v$ for every edge $(u, v) \in E$.

Let v be a vertex such that $x_v = a \neq 0$, and define the sets $A := \{v : x_v = a\}$, $B := \{j : x_v = -a\}$ and $R = \{v : x_v \neq \pm a\}$. The set $A \cup B$ is disconnected from the rest of the graph, because otherwise an edge with an endpoint in $A \cup B$ and an endpoint in R would give a positive contribution to $\sum_{u,v} A_{u,v}(x_u + x_v)^2$; furthermore, every edge incident on a vertex on A must have the other endpoint in B, and vice versa. Thus, $A \cup B$ is a connected component, or a collection of connected components, of G which is bipartite, with the bipartition A, B.