CS 270 Combinatorial Algorithms and Data Structures, Spring 2015

Lecture 21: Spectral graph theory

We'll do an introduction to basic spectral graph theory today.

Real symmetric matrices

Our starting point would be the following spectral theorem for real symmetric matrices.

Theorem Let A be a real symmetric $n \times n$ matrix. Then there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A, and the corresponding eigenvalues are real numbers.

We will not prove this theorem; see e.g. the book "algebraic graph theory" by Godsil and Royle.

The above theorem applies to adjacency matrices of undirected graphs, but not for directed graphs, and this is the reason that the spectral theory is much more developed in undirected graphs.

Power of matrices

Having an orthonormal set of eigenvectors is very good for computation.

For example, let V be the matrix whose columns form an orthonormal basis of eigenvectors of A.

Then $AV = VD$ where D is the diagonal matrix of eigenvalues, and hence $A = VDV^{-1} = VDV^T$.

(since $V^TV = I$ and so $V^{-1} = V^T$)

To compute A^k, we observe that it is just $\overbrace{(VDV^T)(VDV^T)\ldots(VDV^T)}^{k \text{ times}} = VD^kV^T$, and D^k can be computed readily since D is a diagonal matrix, i.e. if $D = \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix}$, $D^k = \begin{bmatrix} d_1^k & & \\ & \ddots & \\ & & d_n^k \end{bmatrix}$

This is very useful, for example, in analyzing random walk of a graph, which we'll see later.

We have also seen that it is useful when we study matrix exponential last time.

Eigen-decomposition

Also, let $\{v_1, v_2, \ldots, v_n\}$ be a basis of orthonormal eigenvectors.

Then any vector x can be written as $c_1v_1 + c_2v_2 + \ldots + c_nv_n$.

By orthonormality, $\langle x, v_i \rangle = \langle c_1v_1 + \cdots + c_nv_n, v_i \rangle = c_1\langle v_1, v_i\rangle + \ldots + c_i\langle v_i, v_i\rangle + \ldots + c_n\langle v_n, v_i\rangle = c_i$.

Therefore, $x = \langle x, v_1\rangle v_1 + \langle x, v_2\rangle v_2 + \ldots + \langle x, v_n\rangle v_n$

$$= v_1v_1^Tx + v_2v_2^Tx + \ldots + v_nv_n^Tx$$

$$= (v_1 v_1^T + v_2 v_2^T + \dots + v_n v_n^T)\, x \,.$$

This is true for all x, and hence $v_1 v_1^T + v_2 v_2^T + \dots + v_n v_n^T = I$.

Multiplying both sides by A, we get

$$Ax = A(v_1 v_1^T + v_2 v_2^T + \dots + v_n v_n^T)\, x$$

$$= (\lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T + \dots + \lambda_n v_n v_n^T)\, x \,.$$

Thus, $A = \lambda_1 v_1 v_1^T + \dots + \lambda_n v_n v_n^T$.

Finally, we claim that $A^{-1} = \frac{1}{\lambda_1} v_1 v_1^T + \frac{1}{\lambda_2} v_2 v_2^T + \dots + \frac{1}{\lambda_n} v_n v_n^T$ if $\lambda_i \neq 0$ for all i.

because $(\lambda_1 v_1 v_1^T + \dots + \lambda_n v_n v_n^T)(\frac{1}{\lambda_1} v_1 v_1^T + \dots + \frac{1}{\lambda_n} v_n v_n^T) = v_1 v_1^T + \dots + v_n v_n^T = I$.

Later on, we will use this idea to define the "pseudo-inverse" of a matrix A, when A is not of full rank.

Trace = sum of eigenvalues

Fact: Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of A. Then $\sum_{i=1}^n \lambda_i = \text{trace}(A)$ where trace of A is defined as the sum of diagonal entries of A.

Proof: Consider $\det(\lambda I - A)$. Its roots are the eigenvalues of A, and so

$\det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2) \dots (\lambda - \lambda_n)$.

Note that the coefficients of $\lambda^{n-1} = -\sum_{i=1}^n \lambda_i$.

On the other hand, $\det(\lambda I - A) = \det \begin{pmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{pmatrix}$.

By the (expansion) definition of the determinant, the coefficient of λ^{n-1} only appears in the term $(\lambda - a_{11})(\lambda - a_{22}) \dots (\lambda - a_{nn})$, which is $-\sum_{i=1}^n a_{ii}$.

Therefore, $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii} = \text{trace}(A)$. $\square$

Graph spectrum

Spectral graph theory relates the combinatorial properties of a graph to the eigenvalues and eigenvectors of its associated matrix (e.g. adjacency matrix, Laplacian matrix).

Let's look at some examples and compute their spectrums.

Complete graph What is the spectrum of a complete graph K_n?

If G is a complete graph, then $A(G) = J - I$, where J denotes the all-one matrix.

Any vector is an eigenvector of I with eigenvalue 1.

Hence the eigenvalues of A are one less than that of J.

Since J is of rank 1, there are $n-1$ eigenvalues of 0.

The all-one vector is an eigenvector of J with eigenvalue n.

So, A has one eigenvalue of $n-1$, and $n-1$ eigenvalues of -1.

<u>**Complete bipartite graph**</u> What is the spectrum of a complete bipartite graph $K_{m,n}$?

The adjacency matrix of $K_{m,n}$ looks like this: $\begin{array}{c} m\{ \\ n\{ \end{array}\underbrace{\begin{bmatrix} 0 & \boxed{1} \\ \boxed{1} & 0 \end{bmatrix}}_{m\quad n}$.

It is of rank 2, so there are $n+m-2$ eigenvalues of 0, and two non-zero eigenvalues λ_1, λ_2.

As $\sum_{i=1}^{n+m} \lambda_i = \text{trace}(A) = 0$, we have $\lambda_1 = -\lambda_2$. Let this value be k.

Thus $\det(\lambda I - A) = \lambda^{n+m} - k^2 \lambda^{n+m-2}$

To determine k, we use the (expansion) definition of determinant of $\begin{pmatrix} \lambda & & \boxed{-1} \\ & \lambda & \\ \boxed{-1} & & \ddots \\ & & & \lambda \end{pmatrix}$.

Any term that contributes to λ^{n+m-2} must have $n+m-2$ diagonal entries,

and the remaining two entries must be $-a_{ij}$ and $-a_{ji}$ for some i,j.

There are totally mn such terms, where the sign of each term is -1.

So, $k^2 = mn$, and thus $k = \sqrt{mn}$.

To conclude, there are $n+m-2$ eigenvalues of 0, and one eigenvalue of $\sqrt{mn}$, and one of $-\sqrt{mn}$.

<u>**Bipartite graphs**</u> We can characterize bipartite graphs by the spectrums.

<u>Claim</u> If G is a bipartite graph and λ is an eigenvalue of $A(G)$ with multiplicity k,

then $-\lambda$ is an eigenvalue of $A(G)$ with multiplicity k.

<u>Proof</u> If G is a bipartite graph, then we can permute the rows and columns of $A(G)$ to obtain the form $A(G) = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$.

Suppose $u = \begin{pmatrix} x \\ y \end{pmatrix}$ is an eigenvector of $A(G)$ with eigenvalue λ.

Then $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$ which implies $B^T x = \lambda y$ and $By = \lambda x$.

This implies that $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}\begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} -By \\ B^T x \end{pmatrix} = \begin{pmatrix} -\lambda x \\ \lambda y \end{pmatrix} = -\lambda \begin{pmatrix} x \\ -y \end{pmatrix}$, and thus $\begin{pmatrix} \vdots \\ -y \end{pmatrix}$ is an eigenvector of $A(G)$ with eigenvalue $-\lambda$.

k linearly independent eigenvectors with eigenvalue λ would give k linearly independent with eigenvalue $-\lambda$, hence the claim. $\square$

The above result shows that the spectrum of a bipartite graph is symmetric around the origin. We now prove that the converse is also true.

<u>Claim</u> If the nonzero eigenvalues occur in pairs λ_i, λ_j with $\lambda_i = -\lambda_j$, then G is bipartite.

<u>Proof</u> Let k be an odd number.

Then $\sum_{i=1}^{n} \lambda_i^k = 0$.

Note that $\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k$ are the eigenvalues of A^k, because if $Au = \lambda u$ then $A^k u = \lambda^k u$.

So, we have $\mathrm{trace}(A^k) = \sum_{i=1}^{n} \lambda_i^k = 0$.

Observe that $A^k_{i,j}$ is the number of length k walks from i to j in G. (by induction)

If G has an odd cycle of length k, then $A^k_{ii} > 0$ for some i and $\mathrm{trace}(A^k) > 0$.

So, since $\mathrm{trace}(A^k) = 0$, G must have no odd cycles and thus bipartite. $\square$

<u>Laplacian Matrices</u>

Given an undirected graph G, the <u>Laplacian matrix</u> $L(G)$ is defined as $D(G) - A(G)$, where $D(G) = \begin{pmatrix} d_1 & & 0 \\ & d_2 & \\ 0 & & \ddots \\ & & & d_n \end{pmatrix}$ is a diagonal matrix with d_i = degree of vertex i in G.

When G is a regular graph, then $D = \begin{pmatrix} d & & 0 \\ & d & \\ 0 & & \ddots \\ & & & d \end{pmatrix}$ and $L = D - A$. Any eigenvector of A with eigenvalue λ is an eigenvector of L with eigenvalue with eigenvalue $d - \lambda$, and vice versa.

So in this case the spectrums of the adjacency matrix and the Laplacian matrix are basically equivalent, but when G is non-regular it may not be easy to relate their eigenvalues.

Let's try to understand more about the spectrum of the Laplacian matrices.

Let $\vec{1}$ be the all-one vector. Then it can be easily checked that $L\vec{1} = 0$.

So L has 0 as an eigenvalue.

Can L have a smaller eigenvalue?

Let $e = ij$ be an edge in G.

call this L_e

Let $e = ij$ be an edge in G.

Then it can be verified that $L(G) = L(G-e) + \begin{array}{c} \\ i \\ j \end{array}\begin{array}{c} \begin{array}{cc} i & j \end{array} \\ \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{array} = L(G-e) + L_e = L(G-e) + \begin{array}{c} i \\ j \end{array}\begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{array}{c} \begin{array}{cc} i & j \end{array} \\ [\,1 \;\; -1\,] \end{array}$

call this L_e

Let $e = ij$. Let b_e be the column vector with the i-th position $= +1$ and the j-th position $= -1$, and 0 o.w.

By induction, we can write $L(G) = \sum_{e \in G} L_e = \sum_{e=ij \in G} b_e b_e^T$.

Let $B = \begin{pmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_m \\ | & | & & | \end{pmatrix}$ be the matrix whose columns are $\{ b_e \mid e \in G \}$. Then $L = BB^T$.

This shows that L is positive semidefinite, and thus 0 is the smallest eigenvalue.

Connectedness

Claim. A graph is connected if and only if 0 is an eigenvalue of $L(G)$ with multiplicity 1.

Proof. If G is disconnected, then the vertex set can be partitioned onto two sets S_1 and S_2 such that there are no edges between them. Then $L(G) = \begin{pmatrix} L(G_1) & 0 \\ 0 & L(G_2) \end{pmatrix}$ and so $\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ are both eigenvectors of $L(G)$ with eigenvalue 0, hence multiplicity ≥ 2.

If G is connected, consider $x^T L x = x^T \left(\sum_{e \in G} L_e \right) x = x^T \left(\sum_{e \in G} b_e b_e^T \right) x = \sum_{e=ij \in G} (x_i - x_j)^2 \geq 0$

If x is an eigenvector with eigenvalue 0, then $Lx = 0$ and thus $x^T L x = 0$.

For $x^T L x = \sum_{e=ij \in G} (x_i - x_j)^2 \geq 0$, we must have $x_i = x_j$ for every edge ij.

Since G is connected, it implies that $x = c \cdot \vec{1}$ for some c, i.e. a multiple of $\vec{1}$.

Hence the eigenvalue 0 has multiplicity one, as required $\square$

Actually, the proof can be used to prove the following (exercise).

Claim. The Laplacian matrix $L(G)$ has 0 as its eigenvalue with multiplicity k if and only if the graph G has k connected components.

Generalizations

So far we have just used the graph spectrum to deduce some simple properties of the graph, like bipartiteness or connectedness, which are easy to deduce by other methods (e.g. BFS).

But the nice thing about these spectral characterizations is that they can be generalized nontrivially:

- λ_2 is "small" iff the graph is "close" to disconnected (i.e. existence of a "sparse" cut).

- λ_k is "small" iff the graph is "close" to having k connected components (i.e. k disjoint "sparse" cuts).
- α_n is "close" to $-\alpha_1$ (adjacency matrix) iff the graph has a component "close" to bipartite.

We will prove the first item and mention the next two items next time.

Rayleigh Quotient

The main tool in relating eigenvalues and eigenvectors to optimization problem is the Rayleigh quotient,

which is defined as $\frac{x^TAx}{x^Tx} = \frac{\sum_{i,j} a_{ij}x_ix_j}{\sum_i x_i^2}$

Let A be a real symmetric matrix with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, and orthonormal eigenvectors $v_1, v_2, \ldots, v_n$.

Claim $\quad \lambda_1 = \max_x \frac{x^TAx}{x^Tx}$

proof $\quad$ Let $x = c_1v_1 + c_2v_2 + \ldots + c_nv_n$, as $v_1, \ldots, v_n$ form a basis

Then $x^TAx = (c_1v_1 + \ldots + c_nv_n)^T A (c_1v_1 + \ldots + c_nv_n)$

$= (c_1v_1 + \ldots + c_nv_n)^T (c_1\lambda_1v_1 + c_2\lambda_2v_2 + \ldots + c_n\lambda_nv_n)$

$= \sum_{i=1}^n c_i^2\lambda_i \quad$ (because $v_1, \ldots, v_n$ are orthonormal)

Similarly, $x^Tx = (c_1v_1 + \ldots + c_nv_n)^T(c_1v_1 + \ldots + c_nv_n) = \sum_{i=1}^n c_i^2$.

So, $\frac{x^TAx}{x^Tx} = \frac{\sum_{i=1}^n c_i^2\lambda_i}{\sum_{i=1}^n c_i^2} \leq \frac{\lambda_1 \sum_{i=1}^n c_i^2}{\sum_{i=1}^n c_i^2} = \lambda_1$.

Since v_1 attains the maximum, the claim follows. $\square$

This can be extended to characterize other eigenvalues.

Let T_k be the set of vectors that are orthogonal to $v_1, v_2, \ldots, v_{k-1}$.

Claim $\quad \lambda_k = \max_{x \in T_k} \frac{x^TAx}{x^Tx}$

proof $\quad$ Let $x \in T_k$. Write $x = c_1v_1 + \ldots + c_nv_n$.

Recall that $c_i = \langle x, v_i \rangle$. Since $x \in T_k$, we have $c_1 = c_2 = \ldots = c_{k-1} = 0$.

Then, $\frac{x^TAx}{x^Tx} = \frac{\sum_{i=k}^n c_i^2\lambda_i}{\sum_{i=k}^n c_i^2} \leq \frac{\lambda_k \sum_{i=k}^n c_i^2}{\sum_{i=k}^n c_i^2} = \lambda_k$.

Since $v_k \in T_k$ and $\frac{v_k^TAv_k}{v_k^Tv_k} = \lambda_k$, the claim follows. $\square$

The above result gives a characterization of λ_k, but it requires the knowledge of the previous eigenvectors.

The result below gives a characterization without knowing the eigenvectors, and is more useful in giving bounds on eigenvalues.

Courant-Fischer Theorem $\lambda_k = \max_{\substack{S \subseteq R^n \\ \dim(S)=k}} \min_{x \in S} \frac{x^T A x}{x^T x} = \min_{\substack{S \subseteq R^n \\ \dim(S)=n-k+1}} \max_{x \in S} \frac{x^T A x}{x^T x}$

Proof We first consider the max-min term.

Let S_k be the k-dimensional subspace spanned by $v_1, \ldots, v_k$, i.e. $\{x \mid x = c_1 v_1 + \ldots + c_k v_k \text{ for some } c_1, \ldots, c_k\}$.

For any $x \in S_k$, $\frac{x^T A x}{x^T x} = \frac{(c_1 v_1 + \ldots + c_k v_k)^T A (c_1 v_1 + \ldots + c_k v_k)}{(c_1 v_1 + \ldots + c_k v_k)^T (c_1 v_1 + \ldots + c_k v_k)} = \frac{\sum_{i=1}^k c_i^2 \lambda_i}{\sum_{i=1}^k c_i^2} \geq \frac{\lambda_k \sum_{i=1}^k c_i^2}{\sum_{i=1}^k c_i^2} = \lambda_k$.

So, $\max_{\substack{S \subseteq R \\ \dim(S)=k}} \min_{x \in S} \frac{x^T A x}{x^T x} \geq \min_{x \in S_k} \frac{x^T A x}{x^T x} \geq \lambda_k$.

To prove that the maximum cannot be greater that λ_k, observe that any k-dimensional subspace must intersect the n-k+1 dimensional subspace T_k spanned by $\{v_k, v_{k+1}, \ldots, v_n\}$.

For any $x \in T_k$, $\frac{x^T A x}{x^T x} = \frac{\sum_{i=k}^n c_i^2 \lambda_i}{\sum_{i=k}^n c_i^2} \leq \lambda_k$.

So, $\max_{\substack{S \subseteq R \\ \dim(S)=k}} \min_{x \in S} \frac{x^T A x}{x^T x} \leq \max_{\substack{S \subseteq R \\ \dim(S)=k}} \min_{x \in S \cap T_k} \frac{x^T A x}{x^T x} \leq \lambda_k$. $\square$

One consequence of the Courant-Fischer theorem is the eigenvalue interlacing theorem.

Eigenvalue Interlacing Theorem Let A be an $n \times n$ symmetric matrix and let B be a principle submatrix of dimension $n-1$ (i.e. B is obtained from A by deleting the same row and column from A). Then

$$\alpha_1 \geq \beta_1 \geq \alpha_2 \geq \beta_2 \geq \ldots \geq \alpha_{n-1} \geq \beta_{n-1} \geq \alpha_n,$$

when $\alpha_1 \geq \alpha_2 \geq \ldots \geq \alpha_n$ are the eigenvalues of A and $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_{n-1}$ are the eigenvalues of B.

Proof It should be clear that $\alpha_i \geq \beta_i$, because $\alpha_i = \max_{\substack{S \subseteq R^n \\ \dim(S)=i}} \min_{x \in S} \frac{x^T A x}{x^T x} \geq \max_{\substack{S \subseteq R^{n-1} \\ \dim(S)=i}} \min_{x \in S} \frac{x^T A x}{x^T x}$

$$= \max_{\substack{S \subseteq R^{n-1} \\ \dim(S)=i}} \min_{x \in S} \frac{x^T B x}{x^T x} = \beta_i.$$

Next we prove $\beta_i \geq \alpha_{i+1}$. For any $S \subseteq R^n$ with $\dim(S) = i+1$, its restriction to the first n-1 coordinates (i.e. $S \cap R^{n-1}$) is of dimension at least i. Let S^* be the set that attains maximum for α_{i+1}. So, $\alpha_{i+1} = \min_{x \in S^*} \frac{x^T A x}{x^T x} \leq \min_{x \in S^* \cap R^{n-1}} \frac{x^T A x}{x^T x} \leq \max_{\substack{S \subseteq R^{n-1} \\ \dim(S) \geq i}} \min_{x \in S} \frac{x^T A x}{x^T x} = \max_{\substack{S \subseteq R^{n-1} \\ \dim(S)=i}} \min_{x \in S} \frac{x^T A x}{x^T x}$

$$= \max_{\substack{S \subseteq R^{n-1} \\ \dim(S)=i}} \min_{x \in S} \frac{x^T B x}{x^T x} = \beta_i. \quad \square$$

First Eigenvalue

First Eigenvalue

Let A be the adjacency matrix of an undirected graph. Let λ_1 be its largest eigenvalue.

Claim $\lambda_1 \le d_{max}$, where d_{max} denotes the maximum degree in G.

Proof Let v_1 be an eigenvector with eigenvalue λ_1.

Let j be the vertex with $v_1(j) \ge v_1(i)$ for all i.

$$\lambda_1 v_1(j) = (Av_1)(j) = \sum_{i:ij \in E(G)} v_1(i) \le \sum_{i:ij\in E(G)} v_1(j) = \deg(j)\, v_1(j) \le d_{max}\, v_1(j).$$

Therefore, $\lambda_1 \le d_{max}$. $\square$

In fact, if $\lambda_1 = d_{max}$, then the above inequalities must hold as equalities, i.e. $v_1(i) = v_1(j)$ for every neighbor i of j and $\deg(j) = d_{max}$. It implies that when G is connected and $\lambda_1 = d_{max}$, then G must be d_{max}-regular and the eigenvalue λ_1 is of multiplicity 1, since the eigenvectors for λ_1 must be of the form $c \cdot \vec{1}$ for some constant c.

Claim $\lambda_1 \ge d_{avg}$, where d_{avg} denotes the average degree of G.

Proof $$\lambda_1 = \max_x \frac{x^T A x}{x^T x} \ge \frac{\vec{1}^T A \vec{1}}{\vec{1}^T \vec{1}} = \frac{\sum_{i,j} a_{ij}}{n} = \frac{2m}{n} = d_{avg}. \quad \square$$

More generally, λ_1 is at least the average degree of the densest induced subgraph.

The Perron-Frobenius theorem for non-negative matrices tell us more about the first eigenvalue.

Theorem Let G be a connected undirected graph. Then

1. the first eigenvalue is of multiplicity one
2. $|\lambda_i| \le \lambda_1$ for all i.
3. all entries of the first eigenvector are non-zero and have the same sign.

We will not prove it in class See Chapter 8 of Godsil-Royle for proofs.

References

You are referred to the course notes on "spectral graph theory" by Dan Spielman for more background.

The book "Algebraic graph theory" by Godsil and Royle is also a good reference.

The course notes "graph partitioning and expanders" by Luca Trevisan is closely related and well written.