Lecture 15 - Spectral Clustering

Renan Monteiro Barbosa

2025

Spectral Clustering

Let G be a graph, $G(V,E)$, where $V = \{1,2,3,\cdots,n\}$

The Laplacian matrix:

$$L_G = \underbrace{D_G}_{\substack{\text{Diagonal matrix}\\ \text{with degrees of}\\ \text{each vertex}}} - \underbrace{A_G}_{\text{adjancency matrix}}$$

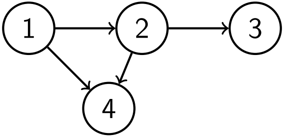
L_G is a PSD Matrix whose eigenvalues are

$$0 = \lambda_1 \leq \underbrace{\lambda_2}_{\text{Fiedler value}} \leq \lambda_3 \leq \cdots \leq \lambda_n$$

An eigenvector with eigenvalue λ_2 is called the Fiedler vector of $\vec{W}$.

Spectral Clustering

We will use $\vec{W}$, L_G to understand clusters in a graph.



$A = \{4, 5, 6, 7, 8\}, |A| = 5$
Notation
If X, Y are sets $X \ Y = \{x \in X | x \notin Y\}$
$|x| =$ size of x

Cut

Definition

A cut is G is a partition of V into two sets, A and V A, when AV. (The cut induced by A)

Notation

If X, YV, let E(X,Y) denote all edges with one certex in X and one vertex in Y.

$E(A, V \ A) = \{\{1, 6\}\}$

Definition

The density of the cut induced by A is:

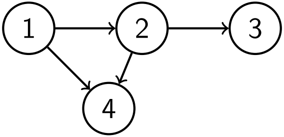
Sparsest cut

Definition:

Let denote the smallest possible desnity of ca cut in G. Any cut with density is called a sparesest cut in G.

Define BG

Define the matrix B_G, the "directed" node-edge incidence matrix of G.



pick a direction for each edge
rows indexed by nodes
columns indexed by edges of G

$$B_G = \begin{array}{c c} & \begin{array}{cccc} 12 & 14 & 23 & 24 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix} \end{array}$$

In the column indexed by the directed edge ij, we put 1 in row i and -1 in row j.

$L_G = B^T B$

Define BG

$L_G = B_G \cdot B_G^T$

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

If you get i^{th} row and i^{th} column, 1^2 or $(-1)^2$ for every edge incident to i. So the sum d_i.
i^{th} row and j^{th} column (if $i \neq j$) if ij is on edge, you will get $1(-1) = -1$
Therefore, L_G is PSD.

Lemma

Lemma:

If G is a graph, the Laplacian L_G is PSD and its eigenvalues are:

$$0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$$

and $\vec{1}$ is an eigenvector of L_G with eigenvalue 0.
Since L_G is PSD, the $\vec{x}^T L_G \vec{x} \geq 0$, $\forall \vec{x} \in \mathbb{R}^n$

Lemma

PwP:

Let $G = ([n], E)$ be the graph with vertices [n] and edge set E. If $\vec{x} = (x_1, \cdots, x_n)^T$.

$$\vec{x}^T L_G \vec{x} = \sum_{\{i,j\} \in E} (x_i - x_j)^2$$

Proof

We know that $L_G = B_G B_G^T$. So $\vec{x}^T L_G \vec{x} = \vec{x}^T B_G B_G^T \vec{x}$.

Lets understand $\vec{x}^T B_G$.

So the entries of $\vec{x}^T B_G$ are all $x_i - xj$ when ij is an edge of G.

Lemma

PwP:

So $\vec{x}^T L_G \vec{x} = \left(\vec{x}^T B_G\right) \cdot \left(B_G^T \vec{x}\right) = \sum_{\{i,j\}\in E} (x_i - x_j)^2$.

Do this for the example:

Connectivity

Definition:

A Graph is conencted if there is a way to walk from any vertex to any other vertex along the edges of G.
Theorem:
G is connected if and only if $\lambda_2 > 0$.
Proof:
If G is not connected the Graph Laplacian will look like this:

Connectivity

We sort of proved if G is not connected, then $\lambda_2 = 0$. (if G has k components, $\lambda_1 = \lambda_2 = \cdots \lambda_4 = 0$)

We still need t show that if G is connected, $\lambda_2 > 0$.
For a symmetric matrix, $AM(\lambda) = GM(\lambda)$ for an eigenvalues.
So we need to show $GM(0) = 1$. Let $\vec{u}$ be an eigenvector with eigenvalue of 0. Then $L_G \vec{u} = \vec{0}$.

$$0 = \vec{u}^T L_G \vec{u} = \sum_{\{i,j\} \in E} (u_i - u_j)^2$$

If $u_1 = C$ and 1,k is an edge, the $u_k = c$. for all edge $ij \in E$.
Similarly if kj is an edge $u_j = c$. Since G is connected $\vec{u} = c\vec{1}$.
AM, algebraic multiplicity.

GM, geometric multiplicity.

Conclusion

Slight Generalization:

The number of times 0 is on eigenvalue of L_G counts the number of conencted components of G.
Fact:

The second eigenvalue λ_2 is called the spectra gap or the Fiedler value of G.

Testing Latex stuff

$$\alpha = \begin{array}{c} \\ k_1 \\ k_2 \\ k_3 \\ k_4 \end{array} \begin{array}{c} \begin{array}{ccc} \bar{f}_1 & \bar{f}_2 & \bar{f}_3 \end{array} \\ \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

$$\begin{bmatrix} 2 & -1 & 0 & -1 \\ -1 & 3 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

$$\beta = \begin{array}{c c} & \begin{array}{ccc} \bar{f}_1 & \bar{f}_2 & \bar{f}_3 \end{array} \\ \begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} & \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & 2 \\ 5 & 3 & 1 \end{pmatrix} \end{array}$$

$$\begin{array}{c c} & \begin{array}{cccc} C_1 & C_2 & \dots & C_n \end{array} \\ \begin{array}{c} N_1 \\ N_2 \\ \\ N_n \end{array} & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \end{array}$$

$$\gamma = \begin{array}{c c c} & \begin{array}{cccc} \bar{f}_1 & \bar{f}_2 & \dots & \bar{f}_n \end{array} & \\ \begin{array}{c} k_1 \\ k_2 \\ \\ k_n \end{array} & \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} & \begin{array}{c} = 1 \\ = 1 \\ \\ = 1 \end{array} \end{array}$$