

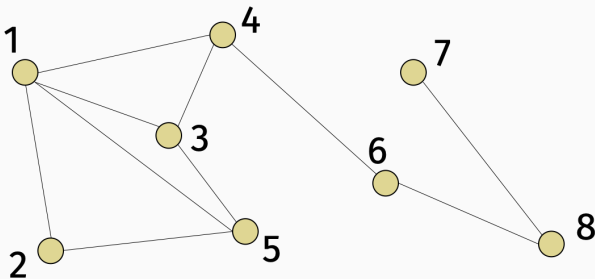
CS-GY 6763: Lecture 12

Spectral clustering, spectral graph theory.

NYU Tandon School of Engineering, Prof. Rajesh Jayaram

SPECTRAL GRAPH THEORY

Main idea: Understand graph data by constructing natural matrix representations, and studying that matrix's spectrum (eigenvalues/eigenvectors).

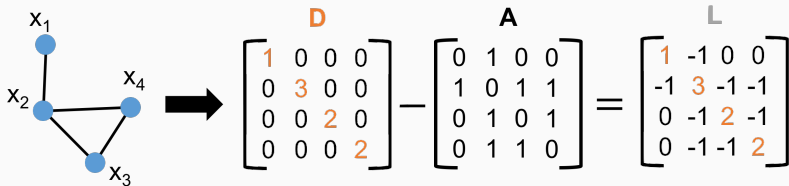


For now assume $G = (V, E)$ is an undirected, unweighted graph with n nodes.

MATRIX REPRESENTATIONS OF GRAPHS

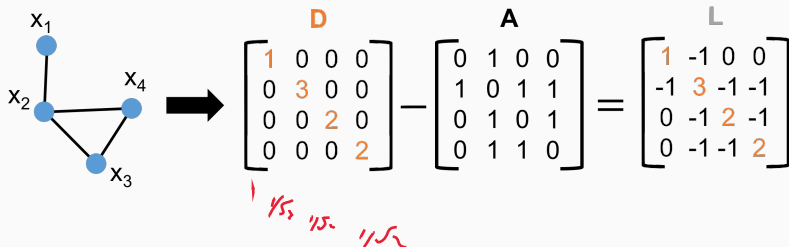
Two most common representations: $n \times n$ adjacency matrix $\mathbf{A}$ and graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is the diagonal degree matrix $\mathbf{D}_{i,i} = \deg(v_i)$, and

$$\mathbf{A}_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in E \\ 0 & \text{otherwise} \end{cases}$$



MATRIX REPRESENTATIONS OF GRAPHS

Two most common representations: $n \times n$ adjacency matrix $\mathbf{A}$ and graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is the diagonal degree matrix.



Also common to look at normalized versions of both of these:

$$\bar{\mathbf{A}} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2} \text{ and } \bar{\mathbf{L}} = \mathbf{I} - \mathbf{D}^{-1/2} \mathbf{A} \mathbf{D}^{-1/2}.$$

Lemma

For $k \geq 1$, the matrix $\mathbf{A}^k$ has the property that $\mathbf{A}_{i,j}^k$ is the number of walks of length k between vertices i and j (recall, a walk is a path where vertices and edges can be repeated).

SPECTRAL GRAPH THEORY TIDBITS

Lemma

For $k \geq 1$, the matrix $\mathbf{A}^k$ has the property that $\mathbf{A}_{i,j}^k$ is the number of walks of length k between vertices i and j (recall, a walk is a path where vertices and edges can be repeated).

Proof: Base case, $\mathbf{A}_{i,j}^2 = \sum_k \mathbf{A}_{i,k} \mathbf{A}_{k,j}$ = number of vertices k which are connected to both v_i and v_j .

In general $\mathbf{A}^k = \mathbf{A}^{k-1} \mathbf{A} = \sum_k \mathbf{A}_{i,k}^{k-1} \mathbf{A}_{k,j}$ = number of walks ending at a vertex k that is connected to v_j .



SPECTRAL GRAPH THEORY TIDBITS

$$A^3: \lambda_1^3 + \lambda_2^3 + \dots + \lambda_n^3 =$$

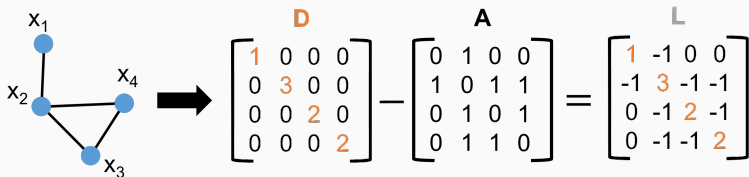
- If $\mathbf{L}$ have k eigenvalues equal to 0, then G has k connected components.
- Sum of cubes of $\mathbf{A}$'s eigenvalues is equal to number of triangles in the graph times 6.
- Sum of eigenvalues to the power q is proportional to the number of q cycles.



Do you see why the last two facts follow from the earlier Lemma?

$$n^w \approx n^{2,3,\dots}$$

THE LAPLACIAN VIEW



$L = B^T B$ where B is the signed “edge-vertex incidence” matrix.

$B =$



$B \in \mathbb{R}^{m \times n}$

$E = \{(i_1, j_1), (i_2, j_2), \dots, (i_m, j_m)\}$

Hand-drawn diagram showing the structure of the incidence matrix B for a specific edge (i_1, j_1) . The matrix is represented as a row vector with a 1 at column i_1 and a -1 at column j_1 , and zeros elsewhere.

$$[0 \dots 1 \dots 0 \dots -1 \dots 0 \dots]$$

$i_1 \quad j_1$

$(n \times m) \times (m \times n)$
 $n \times n$

THE LAPLACIAN VIEW

$$\mathcal{L} = \mathbf{B}^T \mathbf{B} = \mathbf{b}_1 \mathbf{b}_1^T + \mathbf{b}_2 \mathbf{b}_2^T + \dots + \mathbf{b}_m \mathbf{b}_m^T,$$

where $\mathbf{b}_i$ is the i^{th} row of $\mathbf{B}$ (each row corresponds to a single edge).

$$\begin{array}{c} \begin{array}{|c|} \hline 1 \\ \hline -1 \\ \hline \end{array} \begin{array}{|cc|} \hline 1 & -1 \\ \hline \end{array} \\ \mathbf{b}_1 \qquad \mathbf{b}_1^T \end{array} = \begin{array}{|cc|} \hline 1 & -1 \\ -1 & 1 \\ \hline \end{array}$$
$$\begin{array}{c} \begin{array}{|c|} \hline 1 \\ \hline -1 \\ \hline \end{array} \begin{array}{|cc|} \hline 1 & -1 \\ \hline \end{array} \\ \mathbf{b}_2 \qquad \mathbf{b}_2^T \end{array} = \begin{array}{|cc|} \hline 1 & -1 \\ -1 & 1 \\ \hline \end{array}$$

THE LAPLACIAN VIEW

Conclusions from $\mathbf{L} = \mathbf{B}^T \mathbf{B}$

$$(\mathbf{V} \underline{\Sigma} \mathbf{V}^T) (\mathbf{V} \underline{\Sigma} \mathbf{V}^T)$$

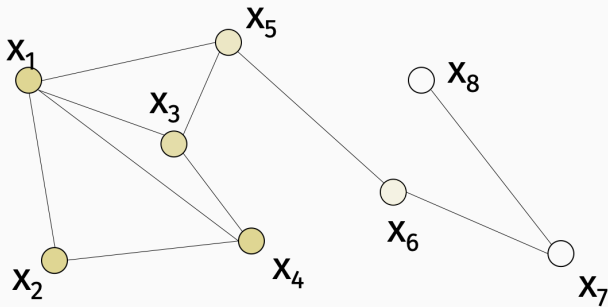
- $\mathbf{L}$ is positive semidefinite: $\mathbf{x}^T \mathbf{L} \mathbf{x} \geq 0$ for all $\mathbf{x}$.
- $\mathbf{L} = \mathbf{V} \underline{\Sigma}^2 \mathbf{V}^T$ where $\mathbf{U} \underline{\Sigma} \mathbf{V}^T$ is $\mathbf{B}$'s SVD. Columns of $\mathbf{V}$ are eigenvectors of $\mathbf{L}$.
- For any vector $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T \mathbf{L} \mathbf{x} = \sum_{(i,j) \in E} (\mathbf{x}(i) - \mathbf{x}(j))^2.$$

$$\mathbf{L} = \sum_i \mathbf{b}_i^T \mathbf{b}_i$$
$$\mathbf{x}^T \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \mathbf{x}$$

THE LAPLACIAN VIEW

$\mathbf{x}^T L \mathbf{x} = \sum_{(i,j) \in E} (\mathbf{x}(i) - \mathbf{x}(j))^2$. So $\mathbf{x}^T L \mathbf{x}$ is small if $\mathbf{x}$ is a “smooth” function with respect to the graph.



Eigenvectors of the Laplacian with small eigenvalues correspond to smooth functions over the graph.

SMALLEST LAPLACIAN EIGENVECTOR

Courant–Fischer min-max principle

Let $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ be the eigenvectors of $\mathbf{L}$.

$$\mathbf{v}_n = \arg \min_{\|\mathbf{v}\|=1} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

$$\mathbf{v}_{n-1} = \arg \min_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_n} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

$$\mathbf{v}_{n-2} = \arg \min_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_n, \mathbf{v}_{n-1}} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

$$\vdots$$

$$\mathbf{v}_1 = \arg \min_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_n, \dots, \mathbf{v}_2} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

LARGEST LAPLACIAN EIGENVECTOR

Courant–Fischer min-max principle

Let $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ be the eigenvectors of $\mathbf{L}$.

$$\mathbf{v}_1 = \arg \max_{\|\mathbf{v}\|=1} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

$$\mathbf{v}_3 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \mathbf{v}_2} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

$\vdots$

$$\mathbf{v}_n = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1, \dots, \mathbf{v}_{n-1}} \mathbf{v}^T \mathbf{L} \mathbf{v}$$

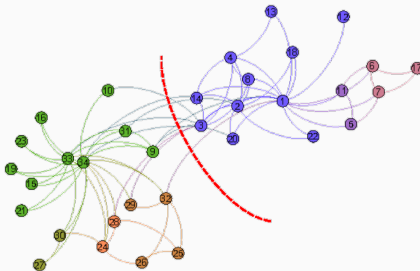
EXAMPLE APPLICATION OF SPECTRAL GRAPH THEORY

- Study graph partitioning problem important in 1) understanding social networks 2) nonlinear clustering in unsupervised machine learning (spectral clustering).
- See how this problem can be solved approximately using Laplacian eigenvectors.
- Give a full analysis of the method for a common random graph model.
- Use two tools: matrix concentration and eigenvector perturbation bounds.

BALANCED CUT

Common goal: Given a graph $G = (V, E)$, partition nodes along a cut that:

- Has few crossing edges: $|\{(u, v) \in E : u \in S, v \in T\}|$ is small.
- Separates large partitions: $|S|, |T|$ are not too small.



(a) Zachary Karate Club Graph

Important in understanding community structure in social networks.

SOCIAL NETWORKS IN THE 1970S

Wayne W. Zachary (1977). An Information Flow Model for Conflict and Fission in Small Groups.

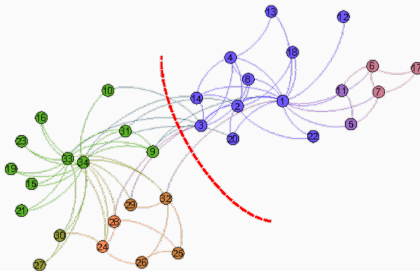
“The network captures 34 members of a karate club, documenting links between pairs of members who interacted outside the club. During the study a conflict arose between the administrator “John A” and instructor “Mr. Hi” (pseudonyms), which led to the split of the club into two. Half of the members formed a new club around Mr. Hi; members from the other part found a new instructor or gave up karate. Based on collected data Zachary correctly assigned all but one member of the club to the groups they actually joined after the split.” – Wikipedia

Beautiful paper – definitely worth checking out!

BALANCED CUT

Common goal: Given a graph $G = (V, E)$, partition nodes along a cut that:

- Has few crossing edges: $|\{(u, v) \in E : u \in S, v \in T\}|$ is small.
- Separates large partitions: $|S|, |T|$ are not too small.

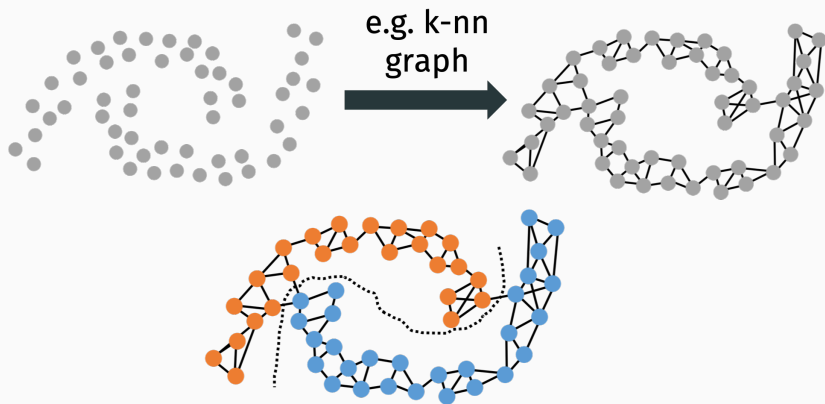


(a) Zachary Karate Club Graph

Important in understanding community structure in social networks.

SPECTRAL CLUSTERING

Idea: Construct synthetic graph for data that is hard to cluster.



Spectral Clustering, Laplacian Eigenmaps, Locally linear embedding, Isomap, etc.

SPECTRAL GRAPH THEORY

There are many ways to formalize Zachary's problem:

β -Balanced Cut:

$$\min_S \text{cut}(S, \underbrace{V \setminus S}_T) \text{ such that } \min(|S|, \underbrace{|V \setminus S|}_T) \geq \beta \text{ for } \beta \leq .5$$

Sparsest Cut:

$$\min_S \frac{\text{cut}(S, V \setminus S)}{\min(|S|, |V \setminus S|)}$$

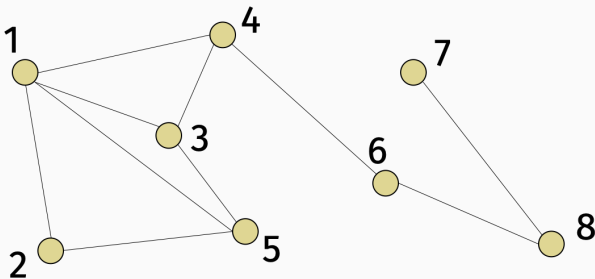
Most formalizations lead to computationally hard problems. Lots of interest in designing polynomial time approximation algorithms, but tend to be slow. In practice, much simpler methods based on the graph spectrum are used.

THE LAPLACIAN VIEW

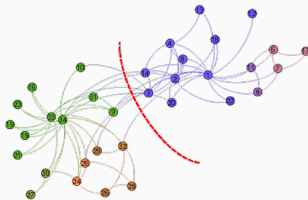
Another conclusion from $L = B^T B$:

For a cut indicator vector $\mathbf{c} \in \{-1, 1\}^n$ with $\mathbf{c}(i) = -1$ for $i \in S$ and $\mathbf{c}(i) = 1$ for $i \in T = V \setminus S$:

$$\mathbf{c}^T L \mathbf{c} = \sum_{(i,j) \in E} (\mathbf{c}(i) - \mathbf{c}(j))^2 = 4 \cdot \text{cut}(S, T). \quad (1)$$



THE LAPLACIAN VIEW



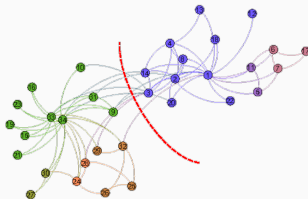
(a) Zachary Karate Club Graph

For a cut indicator vector $\mathbf{c} \in \{-1, 1\}^n$ with $\mathbf{c}(i) = -1$ for $i \in S$ and $\mathbf{c}(i) = 1$ for $i \in T$:

- $\mathbf{c}^T L \mathbf{c} = 4 \cdot \text{cut}(S, T)$.
- $\mathbf{c}^T \mathbf{1} = |T| - |S|$.

Want to minimize both $\mathbf{c}^T L \mathbf{c}$ (cut size) and $\mathbf{c}^T \mathbf{1}$ (imbalance).

THE LAPLACIAN VIEW



(a) Zachary Karate Club Graph

Equivalent formulation if we divide everything by $\sqrt{n}$ so that $\mathbf{c}$ has norm 1. Then $\mathbf{c} \in \{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$ and:

- $\mathbf{c}^T L \mathbf{c} = \frac{4}{n} \cdot \text{cut}(S, T).$
- $\mathbf{c}^T \mathbf{1} = \frac{1}{\sqrt{n}}(|T| - |S|).$

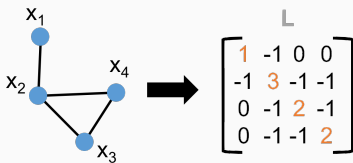
Want to minimize both $\mathbf{c}^T L \mathbf{c}$ (cut size) and $\mathbf{c}^T \mathbf{1}$ (imbalance).

SMALLEST LAPLACIAN EIGENVECTOR

The smallest eigenvector/singular vector $\mathbf{v}_n$ satisfies:

$$\mathbf{v}_n = \frac{1}{\sqrt{n}} \cdot \mathbf{1} = \arg \min_{\mathbf{v} \in \mathbb{R}^n \text{ with } \|\mathbf{v}\|=1} \mathbf{v}^T L \mathbf{v}$$

with $\mathbf{v}_n^T L \mathbf{v}_n = 0$.



SECOND SMALLEST LAPLACIAN EIGENVECTOR

By Courant-Fischer, $\mathbf{v}_{n-1}$ is given by:

$$\mathbf{v}_{n-1} = \arg \min_{\|\mathbf{v}\|=1, \mathbf{v}_n^T \mathbf{v}=0} \mathbf{v}^T L \mathbf{v}$$

If $\mathbf{v}_{n-1}$ were binary $\{-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}\}^n$ it would have:

- $\mathbf{v}_{n-1}^T L \mathbf{v}_{n-1} = \frac{1}{n} \text{cut}(S, T)$ as small as possible **given that**
 $\mathbf{v}_{n-1}^T \mathbf{1} = |T| - |S| = 0$.
- $\mathbf{v}_{n-1}$ would indicate the smallest perfectly balanced cut.

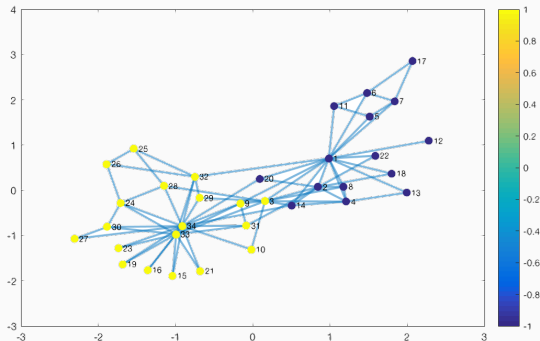
$\mathbf{v}_{n-1} \in \mathbb{R}^n$ is not generally binary, but a natural approach is to
'round' the vector to obtain a cut.

CUTTING WITH THE SECOND LAPLACIAN EIGENVECTOR

Find a good partition of the graph by computing

$$\mathbf{v}_{n-1} = \arg \min_{\mathbf{v} \in \mathbb{R}^n \text{ with } \|\mathbf{v}\|=1, \mathbf{v}^T \mathbf{1}=0} \mathbf{v}^T L \mathbf{v}$$

Set S to be all nodes with $\mathbf{v}_{n-1}(i) < 0$, and T to be all with $\mathbf{v}_{n-1}(i) \geq 0$.

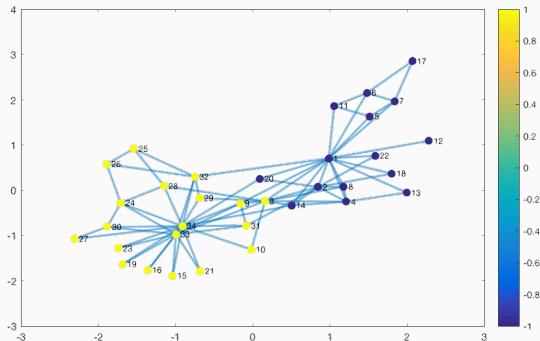


CUTTING WITH THE SECOND LAPLACIAN EIGENVECTOR

Find a good partition of the graph by computing

$$\mathbf{v}_{n-1} = \arg \min_{\mathbf{v} \in \mathbb{R}^n \text{ with } \|\mathbf{v}\|=1, \mathbf{v}^T \mathbf{1}=0} \mathbf{v}^T L \mathbf{v}$$

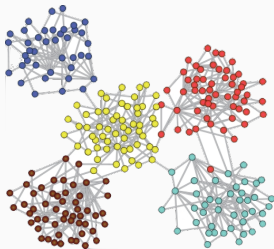
Set S to be all nodes with $\mathbf{v}_{n-1}(i) < 0$, and T to be all with $\mathbf{v}_{n-1}(i) \geq 0$.



SPECTRAL PARTITIONING IN PRACTICE

The Shi-Malik normalized cuts algorithm is one of the most commonly used variants of this approach, using the normalized Laplacian $\bar{\mathbf{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$.

Important consideration: What to do when we want to split the graph into more than two parts?

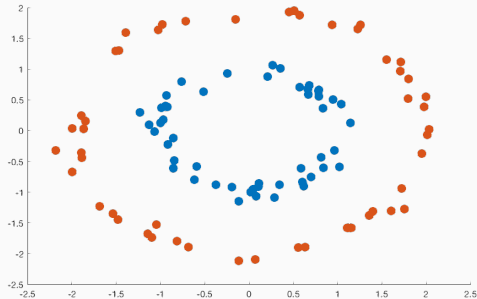


Spectral Clustering:

- Compute smallest ℓ eigenvectors $\mathbf{v}_{n-1}, \dots, \mathbf{v}_{n-\ell}$ of $\mathbf{L}$.
- Represent each node by its corresponding row in $\mathbf{V} \in \mathbb{R}^{n \times \ell}$ whose columns are $\mathbf{v}_{n-1}, \dots, \mathbf{v}_{n-\ell}$.
- Row $\mathbf{V}_{i,*}$ is the *embedding* of node i via ℓ -smallest eigenvalues.
- Cluster these rows using k -means clustering (or really any clustering method).
- Often we choose $\ell = k$, but not necessarily.

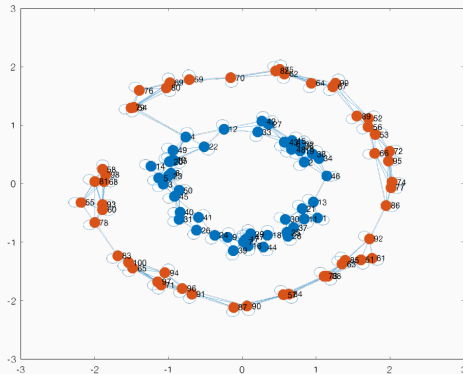
LAPLACIAN EMBEDDING

Original Data: (not linearly separable)



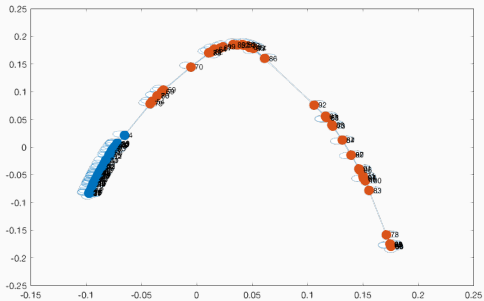
LAPLACIAN EMBEDDING

k -Nearest Neighbors Graph:



LAPLACIAN EMBEDDING

Embedding with eigenvectors $\mathbf{v}_{n-1}, \mathbf{v}_{n-2}$: (linearly separable)

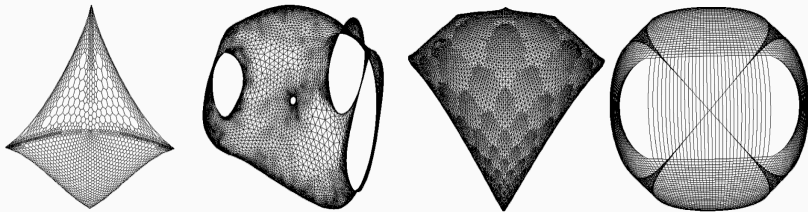


WHY DOES THIS WORK?

Intuitively, since $\mathbf{v} \in \mathbf{v}_1, \dots, \mathbf{v}_k$ are smooth over the graph,

$$\sum_{i,j \in E} (\mathbf{v}[i] - \mathbf{v}[j])^2$$

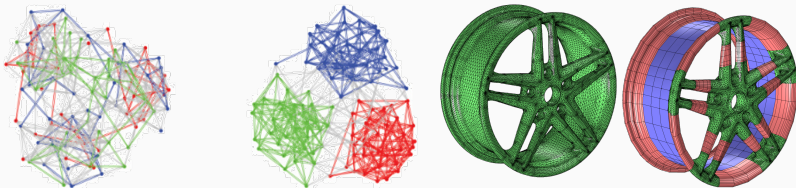
is small for each coordinate. I.e. this embedding explicitly encourages nodes connected by an edge to be placed in nearby locations in the embedding.



Also useful e.g., in graph drawing.

TONS OF OTHER APPLICATIONS!

Fast balanced partitioning algorithms are also use in distributing data in graph databases, for partitioning finite element meshes in scientific computing (e.g., that arise when solving differential equations), and more.



Lots of good software packages (e.g. METIS).

GENERATIVE MODELS

So far: Showed that spectral clustering partitions a graph along a small cut between large pieces.

- No formal guarantee on the ‘quality’ of the partitioning.
- Would be difficult to analyze for general input graphs.

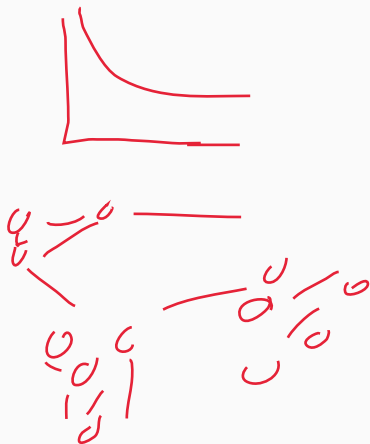
Common approach: Design a natural **generative model** that produces random but realistic inputs and analyze how the algorithm performs on inputs drawn from this model.

- Very common in algorithm design and analysis. Great way to start approaching a problem.
- This is also the whole idea behind Bayesian Machine Learning (can be used to justify ℓ_2 linear regression, k -means clustering, PCA, etc.)

STOCHASTIC BLOCK MODEL

Ideas for a generative model for **social network graphs** that would allow us to understand partitioning?

$$\begin{aligned} & \text{Erdős - Renyi} \\ & G(n, p) \\ & p \sim [0, 1] \\ & \mathbb{E}[|E|] = n^2 \cdot p \\ & p \geq \frac{\log n}{n} \end{aligned}$$



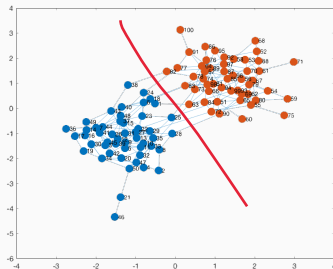
STOCHASTIC BLOCK MODEL

Stochastic Block Model (Planted Partition Model):

Let $G_n(p, q)$ be a distribution over graphs on n nodes, split equally into two groups B and C , each with $n/2$ nodes.

- Any two nodes in the **same group** are connected with probability p (including self-loops).
- Any two nodes in **different groups** are connected with prob.

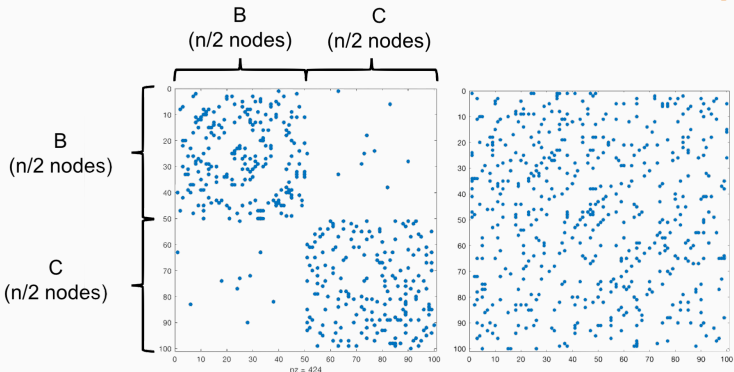
$$q < p.$$



LINEAR ALGEBRAIC VIEW

Let G be a stochastic block model graph drawn from $G_n(p, q)$.

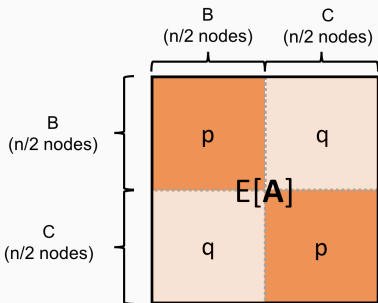
- Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be the adjacency matrix of G . What is $\mathbb{E}[\mathbf{A}]$?



Note that we are arbitrarily ordering the nodes in $\mathbf{A}$ by group. In reality $\mathbf{A}$ would look “scrambled” as on the right.

EXPECTED ADJACENCY SPECTRUM

Letting G be a stochastic block model graph drawn from $G_n(p, q)$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be its adjacency matrix. $(\mathbb{E}[\mathbf{A}])_{i,j} = p$ for i, j in same group, $(\mathbb{E}[\mathbf{A}])_{i,j} = q$ otherwise.



What are the eigenvectors and eigenvalues of $\mathbb{E}[\mathbf{A}]$?

EXPECTED LAPLACIAN

What is the expected Laplacian $G_n(p, q)$?

$$-A_{\mathbb{E}} = \begin{bmatrix} -p & -q \\ -q & -p \end{bmatrix}$$

$$D = \begin{bmatrix} p & & \\ & q & \\ & & p \\ & & & q \end{bmatrix}$$

$$d = \left(\frac{n}{2} - 1\right)p + \frac{n}{2}q$$

$$A \cdot \mathbf{1} = d$$

A and **L** have the same eigenvectors and eigenvalue are equal up to a shift.

EXPECTED ADJACENCY SPECTRUM

Letting G be a stochastic block model graph drawn from $G_n(p, q)$ and $\mathbf{A} \in \mathbb{R}^{n \times n}$ be its adjacency matrix, what are the eigenvectors and eigenvalues of $\mathbb{E}[\mathbf{A}]$?

$$\mathbb{E}[\mathbf{A}] \cdot \mathbf{1} = \left(\left(\frac{n}{2} - 1 \right) p + \frac{n}{2} q \right) \mathbf{1}$$

$$\mathbb{E}[\mathbf{A}] \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} = \left(\left(\frac{n}{2} - 1 \right) p - \frac{n}{2} q \right) \cdot \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix}$$

EXPECTED ADJACENCY SPECTRUM

Diagram illustrating the Expected Adjacency Spectrum. The matrix $E[A]$ is a 2×2 block matrix with blocks p and q , representing communities B and C (each with $n/2$ nodes). The matrix is equal to the product $V \Lambda V^T$.

The matrix V is a 2×8 matrix with columns representing eigenvectors:

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}$$

The matrix Λ is a diagonal matrix with eigenvalues:

$$\Lambda = \begin{bmatrix} \frac{p+q}{2} & & & & & & & \\ & \frac{p-q}{2} & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \\ & & & & & & & \end{bmatrix}$$

The matrix V^T is the transpose of V :

$$V^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}$$

- $\mathbf{v}_1 \sim \mathbf{1}$ with eigenvalue $\lambda_1 = \frac{(p+q)n}{2}$.
- $\mathbf{v}_2 \sim \chi_{B,C}$ with eigenvalue $\lambda_2 = \frac{(p-q)n}{2}$.
- $\chi_{B,C}(i) = 1$ if $i \in B$ and $\chi_{B,C}(i) = -1$ for $i \in C$.

If we compute $\mathbf{v}_2$ then we recover the communities B and C !

EXPECTED LAPLACIAN SPECTRUM

Upshot: The second smallest eigenvector of $\mathbb{E}[\mathbf{L}]$, equivalently the second largest of $\mathbb{E}[\mathbf{A}]$, is $\chi_{B,C}$ – the indicator vector for the cut between the communities.

- If the random graph G (equivilantly $\mathbf{A}$ and $\mathbf{L}$) were exactly equal to its expectation, partitioning using this eigenvector would exactly recover communities B and C .

How do we show that a matrix (e.g., $\mathbf{A}$) is close to its expectation?

Matrix concentration inequalities.

- Analogous to scalar concentration inequalities like Markovs, Chebyshevs, Bernsteins.

MATRIX CONCENTRATION

Matrix Concentration Inequality: If $p \geq O\left(\frac{\log^4 n}{n}\right)$, then with high probability

$$\|\mathbf{A} - \mathbb{E}[\mathbf{A}]\|_2 \leq O(\sqrt{pn}).$$

where $\|\cdot\|_2$ is the matrix **spectral** norm (operator norm).

For $\mathbf{X} \in \mathbb{R}^{n \times d}$, $\|\mathbf{X}\|_2 = \max_{\mathbf{z} \in \mathbb{R}^d: \|\mathbf{z}\|_2=1} \|\mathbf{X}\mathbf{z}\|_2$.

For the stochastic block model application, we want to show that the second eigenvectors of $\mathbf{A}$ and $\mathbb{E}[\mathbf{A}]$ are close. How does this relate to their difference in spectral norm?

Davis-Kahan Eigenvector Perturbation Theorem: Suppose $\mathbf{A}, \bar{\mathbf{A}} \in \mathbb{R}^{d \times d}$ are symmetric with $\|\mathbf{A} - \bar{\mathbf{A}}\|_2 \leq \epsilon$ and eigenvectors $v_1, v_2, \dots, v_d$ and $\bar{v}_1, \bar{v}_2, \dots, \bar{v}_d$. Letting $\theta(v_i, \bar{v}_i)$ denote the angle between v_i and $\bar{v}_i$, for all i :

$$\sin[\theta(v_i, \bar{v}_i)] \leq \frac{\epsilon}{\min_{j \neq i} |\lambda_i - \lambda_j|}$$

where $\lambda_1, \dots, \lambda_d$ are the eigenvalues of $\bar{\mathbf{A}}$.

The error gets larger if there are eigenvalues with similar magnitudes.

EIGENVECTOR PERTURBATION

$$\begin{array}{c} \mathbf{A} \\ \boxed{\begin{array}{cc} 1+\varepsilon & 0 \\ 0 & 1 \end{array}} \end{array} - \begin{array}{c} \bar{\mathbf{A}} \\ \boxed{\begin{array}{cc} 1 & 0 \\ 0 & 1+\varepsilon \end{array}} \end{array} = \begin{array}{c} \mathbf{A}-\bar{\mathbf{A}} \\ \boxed{\begin{array}{cc} \varepsilon & 0 \\ 0 & \varepsilon \end{array}} \end{array}$$

APPLICATION TO STOCHASTIC BLOCK MODEL

Claim 1 (Matrix Concentration): For $p \geq O\left(\frac{\log^4 n}{n}\right)$,

$$\|\mathbf{A} - \mathbb{E}[\mathbf{A}]\|_2 \leq O(\sqrt{pn}).$$

Claim 2 (Davis-Kahan): For $p \geq O\left(\frac{\log^4 n}{n}\right)$,

$$\sin \theta(v_2, \bar{v}_2) \leq \frac{O(\sqrt{pn})}{\min_{j \neq i} |\lambda_i - \lambda_j|} \leq \frac{O(\sqrt{pn})}{(p-q)n/2} = O\left(\frac{\sqrt{p}}{(p-q)\sqrt{n}}\right)$$

Recall: $\mathbb{E}[\mathbf{A}]$, has eigenvalues $\lambda_1 = \frac{(p+q)n}{2}$, $\lambda_2 = \frac{(p-q)n}{2}$, $\lambda_i = 0$ for $i \geq 3$.

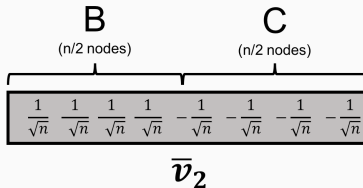
$$\min_{j \neq i} |\lambda_i - \lambda_j| = \min\left(qn, \frac{(p-q)n}{2}\right).$$

Assume $\left|\frac{(p-q)n}{2} - 0\right|$ will be the minimum of the two gaps. I.e. smaller than $\left|\frac{(p+q)n}{2} - \frac{(p-q)n}{2}\right| = qn$.

APPLICATION TO STOCHASTIC BLOCK MODEL

So Far: $\sin \theta(v_2, \bar{v}_2) \leq O\left(\frac{\sqrt{p}}{(p-q)\sqrt{n}}\right)$. What does this give us?

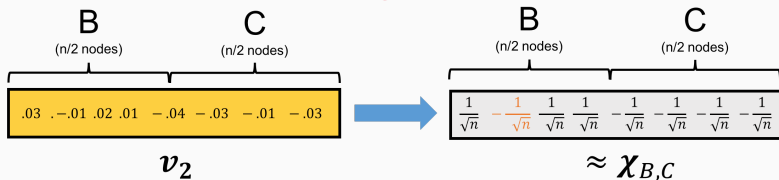
- Can show that this implies $\|v_2 - \bar{v}_2\|_2^2 \leq O\left(\frac{p}{(p-q)^2 n}\right)$ (exercise).
- $\bar{v}_2$ is $\frac{1}{\sqrt{n}}\chi_{B,C}$: the community indicator vector.



- Every i where $v_2(i)$, $\bar{v}_2(i)$ differ in sign contributes $\geq \frac{1}{n}$ to $\|v_2 - \bar{v}_2\|_2^2$.
- So they differ in sign in at most $O\left(\frac{p}{(p-q)^2}\right)$ positions.

APPLICATION TO STOCHASTIC BLOCK MODEL

Upshot: If G is a stochastic block model graph with adjacency matrix $\mathbf{A}$, if we compute its second large eigenvector v_2 and assign nodes to communities according to the sign pattern of this vector, we will correctly assign all but $O\left(\frac{p}{(p-q)^2}\right)$ nodes.



- Why does the error increase as q gets close to p ?
- Even when $p - q = O(1/\sqrt{n})$, assign all but an $O(n)$ fraction of nodes correctly. E.g., assign 99% of nodes correctly.