Randomised Algorithms

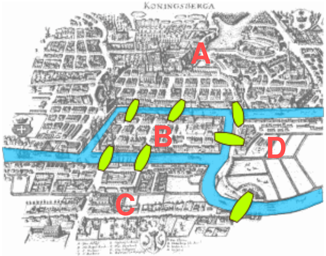
Lecture 11-12: Spectral Graph Theory and Clustering

Thomas Sauerwald (tms41@cam.ac.uk)

Lent 2022

Outline

Origin of Graph Theory



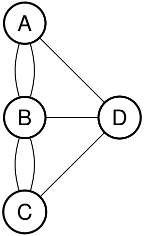
Source: Wikipedia

Seven Bridges at Königsberg 1737

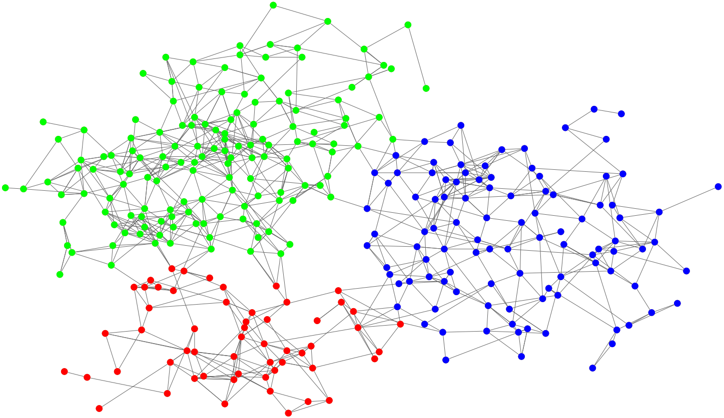
Source: Wikipedia

Leonhard Euler (1707-1783)

Is there a tour which crosses each bridge **exactly once**?



Graphs Nowadays: Clustering



Goal: Use spectrum of graphs (unstructured data) to extract clustering (communitites) or other structural information.

Graph Clustering (applications)

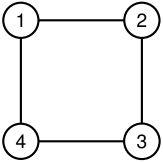
- Applications of Graph Clustering
  - Community detection
  - Group webpages according to their topics
  - Find proteins performing the same function within a cell
  - Image segmentation
  - Identify bottlenecks in a network
  - ...
- Unsupervised learning method
  (there is no ground truth (usually), and we cannot learn from mistakes!)

- Different formalisations for different applications
  - Geometric Clustering: partition points in a Euclidean space
    - k-means, k-medians, k-centres, etc.

  - Graph Clustering: partition vertices in a graph
    - modularity, conductance, min-cut, etc.

Graphs and Matrices

Graphs



- Connectivity
- Bipartiteness
- Number of triangles
- Graph Clustering
- Graph isomorphism
- Maximum Flow
- Shortest Paths
- ...

Matrices

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

- Eigenvalues
- Eigenvectors
- Inverse
- Determinant
- Matrix-powers
- ...

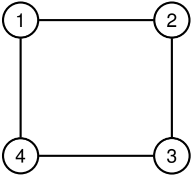
Outline

Adjacency Matrix

Adjacency matrix

Let $G = (V, E)$ be an undirected graph. The adjacency matrix of G is the n by n matrix $\mathbf{A}$ defined as

$$\mathbf{A}_{u,v} = \begin{cases} 1 & \text{if } \{u, v\} \in E \\ 0 & \text{otherwise.} \end{cases}$$



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Properties of $\mathbf{A}$:

- The sum of elements in each row/column i equals the degree of the corresponding vertex i, $\deg(i)$
- Since G is undirected, $\mathbf{A}$ is symmetric

Eigenvalues and Graph Spectrum of A

Eigenvalues and Eigenvectors

Let $\mathbf{M} \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{M}$ if and only if there exists $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

$$\mathbf{M}x = \lambda x.$$

We call x an eigenvector of $\mathbf{M}$ corresponding to the eigenvalue λ.

An undirected graph G is d-regular if every degree is d, i.e., every vertex has exactly d connections.

Graph Spectrum

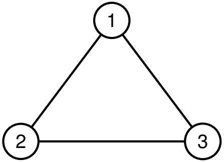
Let $\mathbf{A}$ be the adjacency matrix of a d-regular graph G with n vertices. Then, $\mathbf{A}$ has n real eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and n corresponding orthonormal eigenvectors $f_1, \ldots, f_n$. These eigenvalues associated with their multiplicities constitute the spectrum of G.

For symmetric matrices: algebraic multiplicity = geometric multiplicity

Exercise 1

Bonus: Can you find a short-cut to $\det(\mathbf{A} - \lambda \cdot \mathbf{I})$?

Exercise: What are the Eigenvalues and Eigenvectors?



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Solution:

- The three eigenvalues are $\lambda_1 = \lambda_2 = -1, \lambda_3 = 2$.
- The three eigenvectors are (for example):

$$f_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad f_2 = \begin{pmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix}, \quad f_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

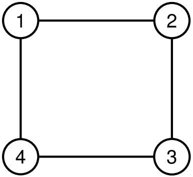
Laplacian Matrix

Laplacian Matrix

Let $G = (V, E)$ be a d-regular undirected graph. The (normalised) Laplacian matrix of G is the n by n matrix $\mathbf{L}$ defined as

$$\mathbf{L} = \mathbf{I} - \frac{1}{d}\mathbf{A},$$

where $\mathbf{I}$ is the $n \times n$ identity matrix.



$$\mathbf{L} = \begin{pmatrix} 1 & -1/2 & 0 & -1/2 \\ -1/2 & 1 & -1/2 & 0 \\ 0 & -1/2 & 1 & -1/2 \\ -1/2 & 0 & -1/2 & 1 \end{pmatrix}$$

Properties of $\mathbf{L}$:

- The sum of elements in each row/column equals zero
- $\mathbf{L}$ is symmetric

Relating Spectrum of Adjacency Matrix and Laplacian Matrix

Correspondence between Adjacency and Laplacian Matrix

$\mathbf{A}$ and $\mathbf{L}$ have the same eigenvectors.

Proof:

- Let λ and f be an eigenvalue and eigenvector of $\mathbf{A}$, i.e., $\mathbf{A} \cdot f = \lambda \cdot f$.
- Then:

$$\begin{aligned} \mathbf{L} \cdot f &= \left(\mathbf{I} - \frac{1}{d} \mathbf{A} \right) \cdot f \\ &= \mathbf{I} \cdot f - \frac{1}{d} \mathbf{A} \cdot f \\ &= f - \frac{1}{d} \lambda \cdot f \\ &= \left(1 - \frac{\lambda}{d} \right) \cdot f. \end{aligned}$$

- Hence $(1 - \frac{\lambda}{d}, f)$ is an eigenvalue and eigenvector pair of $\mathbf{L}$. □

Eigenvalues and Graph Spectrum of L

Eigenvalues and eigenvectors

Let $\mathbf{M} \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{M}$ if and only if there exists $x \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ such that

$$\mathbf{M}x = \lambda x.$$

We call x an eigenvector of $\mathbf{M}$ corresponding to the eigenvalue λ.

Graph Spectrum

Let $\mathbf{L}$ be the Laplacian matrix of a d-regular graph G with n vertices. Then, $\mathbf{L}$ has n real eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ and n corresponding orthonormal eigenvectors $f_1, \ldots, f_n$.

Useful Facts of Graph Spectrum

Lemma

Let $\mathbf{L}$ be the Laplacian matrix of an undirected, regular graph $G = (V, E)$ with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$.

1. $\lambda_1 = 0$ with eigenvector $\mathbf{1}$
2. the multiplicity of the eigenvalue 0 is equal to the number of connected components in G
3. $\lambda_n \leq 2$
4. $\lambda_n = 2$ iff there exists a bipartite connected component.

The proof of these properties is based on a powerful characterisation of eigenvalues/vectors!

A Min-Max Characterisation of Eigenvalues and Eigenvectors

Courant-Fischer Min-Max Formula

Let $\mathbf{M}$ be an n by n symmetric matrix with eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. Then,

$$\lambda_k = \min_{\substack{x^{(1)},\ldots,x^{(k)} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, \\ x^{(i)} \perp x^{(j)}}} \max_{i \in \{1,\ldots,k\}} \frac{{x^{(i)}}^T \mathbf{M} x^{(i)}}{{x^{(i)}}^T x^{(i)}}.$$

The eigenvectors corresponding to $\lambda_1, \ldots, \lambda_k$ minimise such expression.

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

minimised by an eigenvector f_1 for λ_1

$$\lambda_2 = \min_{\substack{x \in \mathbb{R}^n \setminus \{\mathbf{0}\} \\ x \perp f_1}} \frac{x^T \mathbf{M} x}{x^T x}$$

minimised by f_2

Quadratic Forms of the Laplacian

Lemma

Let $\mathbf{L}$ be the Laplacian matrix of a d-regular graph $G = (V, E)$ with n vertices. For any $x \in \mathbb{R}^n$,

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

Proof:

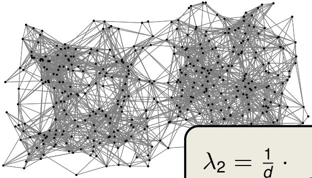
$$\begin{aligned} x^T \mathbf{L} x = x^T \left(\mathbf{I} - \frac{1}{d} \mathbf{A} \right) x &= x^T x - \frac{1}{d} x^T \mathbf{A} x \\ &= \sum_{u \in V} x_u^2 - \frac{2}{d} \sum_{\{u,v\} \in E} x_u x_v \\ &= \frac{1}{d} \sum_{\{u,v\} \in E} (x_u^2 + x_v^2 - 2 x_u x_v) \\ &= \sum_{\{u,v\} \in E} \frac{(x_u - x_v)^2}{d}. \end{aligned}$$

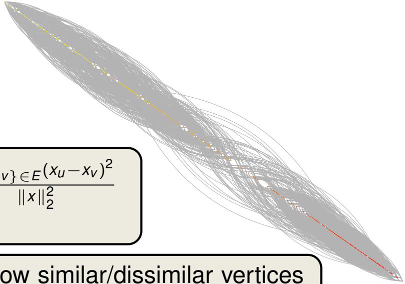
Visualising a Graph

Question: How can we visualize a complicated object like an unknown graph with many vertices in low-dimensional space?

Embedding onto Line

Coordinates given by x





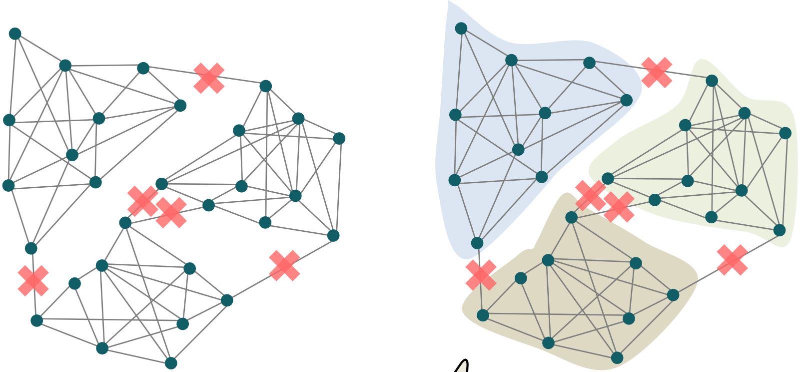
$$\lambda_2 = \frac{1}{d} \cdot \min_{\substack{x \in \mathbb{R}^n \setminus \{\mathbf{0}\} \\ x \perp f_1}} \frac{\sum_{\{u,v\} \in E} (x_u - x_v)^2}{\|x\|_2^2}$$

The coordinates in the vector **x** indicate how similar/dissimilar vertices are. Edges between dissimilar vertices are penalised quadratically.

Outline

A Simplified Clustering Problem

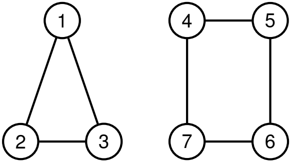
*Partition the graph into **connected components** so that any pair of vertices in the same component is connected, but vertices in different components are not.*



We could obviously solve this easily using DFS/BFS, but let's see how we can tackle this using the spectrum of **L**!

Exercise 2

Exercise: What are the Eigenvectors with Eigenvalue 0 of **L**?



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & 0 & 0 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \\ 0 & 0 & 0 & -\frac{1}{2} & 0 & -\frac{1}{2} & 1 \end{pmatrix}$$

Solution:

- The two smallest eigenvalues are $\lambda_1 = \lambda_2 = 0$.
- The corresponding two eigenvectors are:

$$f_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \left(\text{or} \quad f_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} - \\ -1/3 \\ \vdots \\ \vdots \\ \vdots \\ 1/4 \\ 1/4 \end{pmatrix} \right.$$

Thus we can easily solve the simplified clustering problem by computing the eigenvectors with eigenvalue 0

Next section: A fine-grained approach works even if the clusters are **sparsely** connected!

Useful Facts of Graph Spectrum (Proof of 2)

Let us generalise and formalise the example before!

Proof of 2 (multiplicity of 0 equals the no. of connected components):

1. ("$\Longrightarrow$" $cc(G) \leq \mathrm{mult}(0)$). We will show:
   G has exactly k connected comp. $C_1, \ldots, C_k \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_k = 0$
   - Take $\chi_{C_i} \in \{0,1\}^n$ such that $\chi_{C_i}(u) = \mathbf{1}_{u \in C_i}$ for all $u \in V$
   - Clearly, the χ_{C_i}'s are orthogonal
   - $\chi_{C_i}^T \mathbf{L} \chi_{C_i} = \frac{1}{d} \cdot \sum_{\{u,v\} \in E} (\chi_{C_i}(u) - \chi_{C_i}(v))^2 = 0 \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_k = 0$

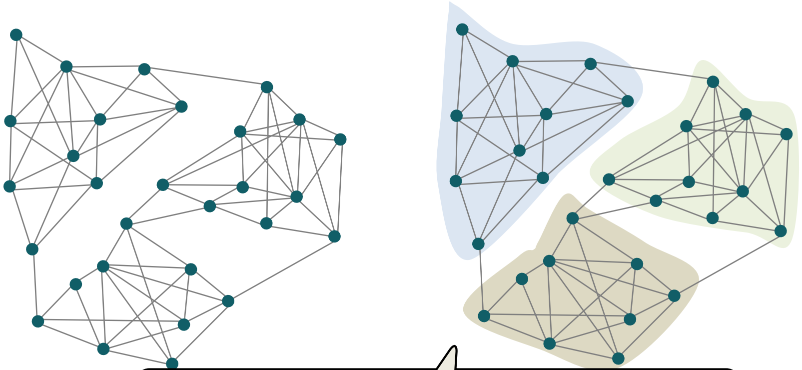
2. ("$\Longleftarrow$" $cc(G) \geq \mathrm{mult}(0)$). We will show:
   $\lambda_1 = \cdots = \lambda_k = 0 \quad \Rightarrow \quad G$ has at least k connected comp. $C_1, \ldots, C_k$
   - there exist $f_1, \ldots, f_k$ orthonormal such that $\sum_{\{u,v\} \in E} (f_i(u) - f_i(v))^2 = 0$
   - $\Rightarrow f_1, \ldots, f_k$ constant on connected components
   - as $f_1, \ldots, f_k$ are pairwise orthogonal, G must have k different connected components.

□

Outline

Graph Clustering

*Partition the graph into **pieces (clusters)** so that vertices in the same piece have, on average, more connections among each other than with vertices in other clusters*



Let us for simplicity focus on the case of **two clusters**!

Conductance

Conductance

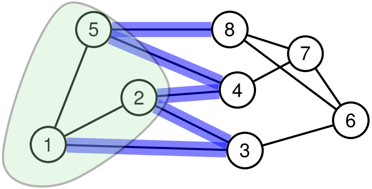
Let $G = (V, E)$ be a d-regular and undirected graph and $\emptyset \neq S \subsetneq V$. The conductance (edge expansion) of S is

$$\phi(S) := \frac{e(S, S^c)}{d \cdot |S|}$$

Moreover, the conductance (edge expansion) of the graph G is

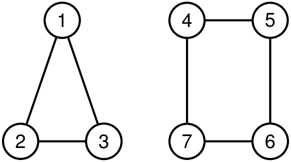
$$\phi(G) := \min_{S \subseteq V:\ 1 \leq |S| \leq n/2} \phi(S)$$

NP-hard to compute!



- $\phi(S) = \frac{5}{9}$
- $\phi(G) \in [0, 1]$ and $\phi(G) = 0$ iff G is disconnected
- If G is a complete graph, then $e(S, V \setminus S) = |S| \cdot (n - |S|)$ and $\phi(G) \approx 1/2$.

λ_2 versus Conductance (1/2)



$$\phi(G) = 0 \quad \Leftrightarrow \quad G \text{ is disconnected} \quad \Leftrightarrow \quad \lambda_2(G) = 0$$

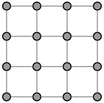
What is the relationship between $\phi(G)$ and $\lambda_2(G)$ for **connected** graphs?

λ_2 versus Conductance (2/2)

1D Grid

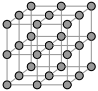
$$\lambda_2 \sim n^{-2}$$
$$\phi \sim n^{-1}$$

2D Grid



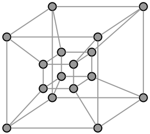
$$\lambda_2 \sim n^{-1}$$
$$\phi \sim n^{-1/2}$$

3D Grid



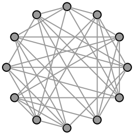
$$\lambda_2 \sim n^{-2/3}$$
$$\phi \sim n^{-1/3}$$

Hypercube



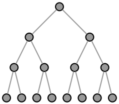
$$\lambda_2 \sim (\log n)^{-1}$$
$$\phi \sim (\log n)^{-1}$$

Random Graph (Expanders)



$$\lambda_2 = \Theta(1)$$
$$\phi = \Theta(1)$$

Binary Tree



$$\lambda_2 \sim n^{-1}$$
$$\phi \sim n^{-1}$$

Relating λ_2 and Conductance

Cheeger's inequality

Let G be a d-regular undirected graph and $\lambda_1 \leq \cdots \leq \lambda_n$ be the eigenvalues of its Laplacian matrix. Then,

$$\frac{\lambda_2}{2} \leq \phi(G) \leq \sqrt{2\lambda_2}.$$

Spectral Clustering:

1. Compute the eigenvector x corresponding to λ_2
2. Order the vertices so that $x_1 \leq x_2 \leq \cdots \leq x_n$ (embed V on $\mathbb{R}$)
3. Try all $n-1$ sweep cuts of the form $(\{1, 2, \ldots, k\}, \{k+1, \ldots, n\})$ and return the one with smallest conductance

- It returns cluster $S \subseteq V$ such that $\phi(S) \leq \sqrt{2\lambda_2} \leq 2\sqrt{\phi(G)}$
- no constant factor worst-case guarantee, but usually works well in practice (see examples later!)
- very fast: can be implemented in $O(|E| \log |E|)$ time

Proof of Cheeger's Inequality (non-examinable)

Proof (of the easy direction):

Optimisation Problem: Embed vertices on a line such that sum of squared distances is minimised

- By the Courant-Fischer Formula,

$$\lambda_2 = \min_{\substack{x \in \mathbb{R}^n \\ x \neq 0, x \perp 1}} \frac{x^T \mathbf{L} x}{x^T x} = \frac{1}{d} \cdot \min_{\substack{x \in \mathbb{R}^n \\ x \neq 0, x \perp 1}} \frac{\sum_{u \sim v} (x_u - x_v)^2}{\sum_u x_u^2}.$$

- Let $S \subseteq V$ be the subset for which $\phi(G)$ is minimised. Define $y \in \mathbb{R}^n$ by:

$$y_u = \begin{cases} \frac{1}{|S|} & \text{if } u \in S, \\ -\frac{1}{|V \setminus S|} & \text{if } u \in V \setminus S. \end{cases}$$

- Since $y \perp 1$, it follows that

$$\lambda_2 \leq \frac{1}{d} \cdot \frac{\sum_{u \sim v} (y_u - y_v)^2}{\sum_u y_u^2} = \frac{1}{d} \cdot \frac{|E(S, V \setminus S)| \cdot (\frac{1}{|S|} + \frac{1}{|V \setminus S|})^2}{\frac{1}{|S|} + \frac{1}{|V \setminus S|}}$$

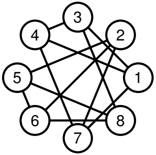
$$= \frac{1}{d} \cdot |E(S, V \setminus S)| \cdot \left(\frac{1}{|S|} + \frac{1}{|V \setminus S|} \right)$$

$$\leq \frac{1}{d} \cdot \frac{2 \cdot |E(S, V \setminus S)|}{|S|} = 2 \cdot \phi(G). \qquad \square$$

Outline

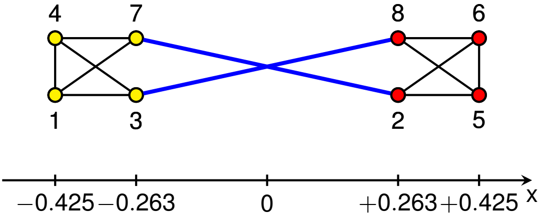
Illustration on a small Example

$$\mathbf{A} = \begin{pmatrix} 0&0&1&1&0&0&1&0\\ 0&0&0&0&1&1&1&0\\ 1&0&0&1&0&0&0&1\\ 1&0&1&0&0&0&1&0\\ 0&1&0&0&0&1&0&1\\ 0&1&0&0&1&0&0&1\\ 1&1&0&1&0&0&0&0\\ 0&0&1&0&1&1&0&0 \end{pmatrix} \quad \mathbf{L} = \begin{pmatrix} 1&0&-\frac{1}{3}&-\frac{1}{3}&0&0&-\frac{1}{3}&0\\ 0&1&0&0&-\frac{1}{3}&-\frac{1}{3}&-\frac{1}{3}&0\\ -\frac{1}{3}&0&1&-\frac{1}{3}&0&0&0&-\frac{1}{3}\\ -\frac{1}{3}&0&-\frac{1}{3}&1&0&0&-\frac{1}{3}&0\\ 0&-\frac{1}{3}&0&0&1&-\frac{1}{3}&0&-\frac{1}{3}\\ 0&-\frac{1}{3}&0&0&-\frac{1}{3}&1&0&-\frac{1}{3}\\ -\frac{1}{3}&-\frac{1}{3}&0&-\frac{1}{3}&0&0&1&0\\ 0&0&-\frac{1}{3}&0&-\frac{1}{3}&-\frac{1}{3}&0&1 \end{pmatrix}$$



$$\lambda_2 = 1 - \sqrt{5}/3 \approx 0.25$$

$$v = (-0.425, +0.263, -0.263, -0.425, +0.425, +0.425, -0.263, +0.263)^T$$



Sweep: 4

Conductance: 0.166

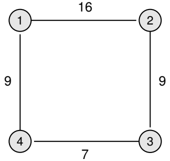
Let us now look at an example of a non-regular graph!

The Laplacian Matrix (General Version)

The (normalised) Laplacian matrix of $G = (V, E, w)$ is the n by n matrix

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$$

where $\mathbf{D}$ is a diagonal $n \times n$ matrix s.t. $\mathbf{D}_{uu} = deg(u) = \sum_{\{u,v\}\in E} w(u,v)$, and $\mathbf{A}$ is the weighted adjacency matrix of G.



$$\mathbf{L} = \begin{pmatrix} 1 & -16/25 & 0 & -9/20 \\ -16/25 & 1 & -9/20 & 0 \\ 0 & -9/20 & 1 & -7/16 \\ -9/20 & 0 & -7/16 & 1 \end{pmatrix}$$

- $\mathbf{L}_{uv} = \frac{w(u,v)}{\sqrt{d_u d_v}}$ for $u \neq v$
- $\mathbf{L}$ is symmetric
- If G is d-regular, $\mathbf{L} = \mathbf{I} - \frac{1}{d} \cdot \mathbf{A}$.

Conductance and Spectral Clustering (General Version)

Conductance (General Version)

Let $G = (V, E, w)$ and $\emptyset \subsetneq S \subsetneq V$. The conductance (edge expansion) of S is

$$\phi(S) := \frac{w(S, S^c)}{\min\{\mathrm{vol}(S), \mathrm{vol}(S^c)\}},$$

where $w(S, S^c) := \sum_{u \in S, v \in S^c} w(u, v)$ and $\mathrm{vol}(S) := \sum_{u \in S} d(u)$. Moreover, the conductance (edge expansion) of G is

$$\phi(G) := \min_{\emptyset \neq S \subsetneq V} \phi(S).$$

Spectral Clustering (General Version):

1. Compute the eigenvector x corresponding to λ_2 **and** $y = \mathbf{D}^{-1/2} x$.
2. Order the vertices so that $y_1 \leq y_2 \leq \cdots \leq y_n$ (embed V on $\mathbb{R}$)
3. Try all $n - 1$ sweep cuts of the form $(\{1, 2, \ldots, k\}, \{k + 1, \ldots, n\})$ and return the one with smallest conductance

Stochastic Block Model and 1D-Embedding

Stochastic Block Model

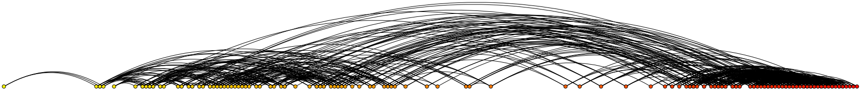
$G = (V, E)$ with clusters $S_1, S_2 \subseteq V$, $0 \leq q < p \leq 1$

$$\mathbf{P}[\{u, v\} \in E] = \begin{cases} p & \text{if } u, v \in S_i, \\ q & \text{if } u \in S_i, v \in S_j, i \neq j. \end{cases}$$

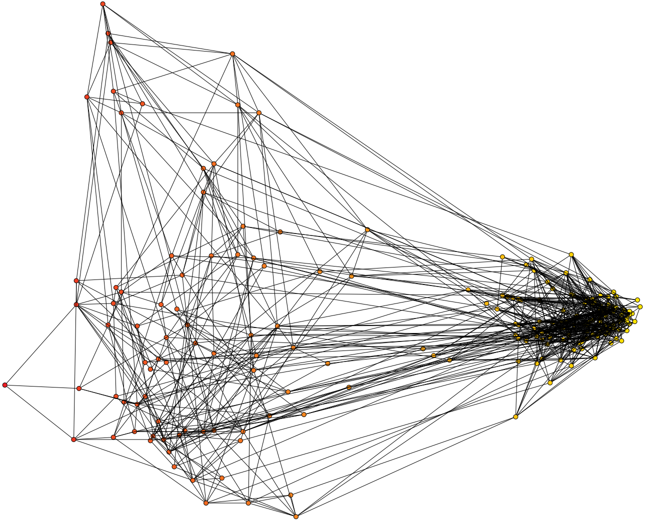
Here:

- $|S_1| = 80$, $|S_2| = 120$
- $p = 0.08$
- $q = 0.01$

```
Number of Vertices: 200
Number of Edges:  919
Eigenvalue  1  :  -1.1968431479565368e-16
Eigenvalue  2  :  0.1543784937248489
Eigenvalue  3  :  0.37049909753568877
Eigenvalue  4  :  0.39770640242147404
Eigenvalue  5  :  0.4316114413430584
Eigenvalue  6  :  0.44379221120189777
Eigenvalue  7  :  0.4564011652684181
Eigenvalue  8  :  0.4632911204500282
Eigenvalue  9  :  0.474638606357877
Eigenvalue  10  :  0.4814019607292904
```



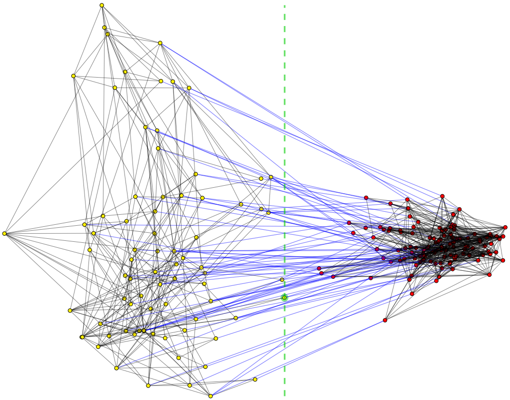
Drawing the 2D-Embedding



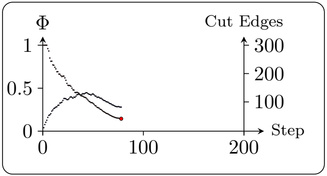
Spectral Clustering

For the animation, see the full slides.

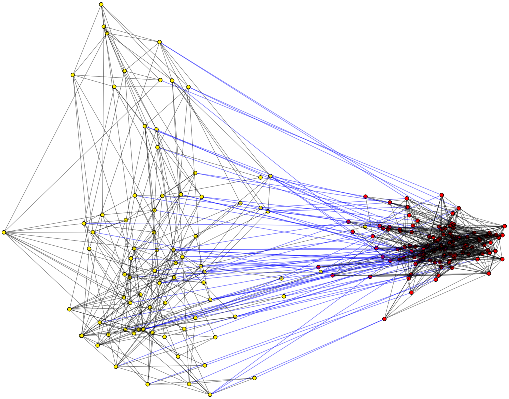
Best Solution found by Spectral Clustering



- Step: 78
- Threshold: −0.0268
- Partition Sizes: 78/122
- Cut Edges: 84
- Conductance: 0.1448



Clustering induced by Blocks



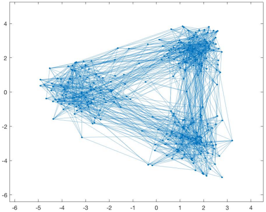
- Step: 1
- Threshold: 0
- Partition Sizes: 80/120
- Cut Edges: 88
- Conductance: 0.1486

Additional Example: Stochastic Block Models with 3 Clusters

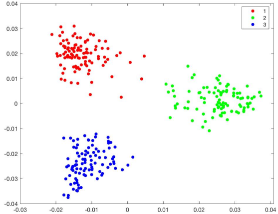
Graph $G = (V, E)$ with clusters
$S_1, S_2, S_3 \subseteq V; \quad 0 \leq q < p \leq 1$

$$\mathbf{P}\left[\{u, v\} \in E\right] = \begin{cases} p & u, v \in S_i \\ q & u \in S_i, v \in S_j, i \neq j \end{cases}$$

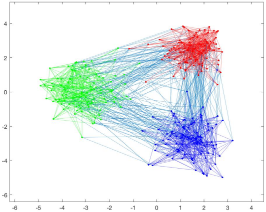
$|V| = 300, |S_i| = 100$
$p = 0.08, q = 0.01.$



Spectral embedding



Output of Spectral Clustering



Choosing the Cluster Number k

- If k is unknown:
  - small λ_k means there exist k sparsely connected subsets in the graph (recall: $\lambda_1 = \ldots = \lambda_k = 0$ means there are k connected components)
  - large λ_{k+1} means all these k subsets have "good" inner-connectivity properties

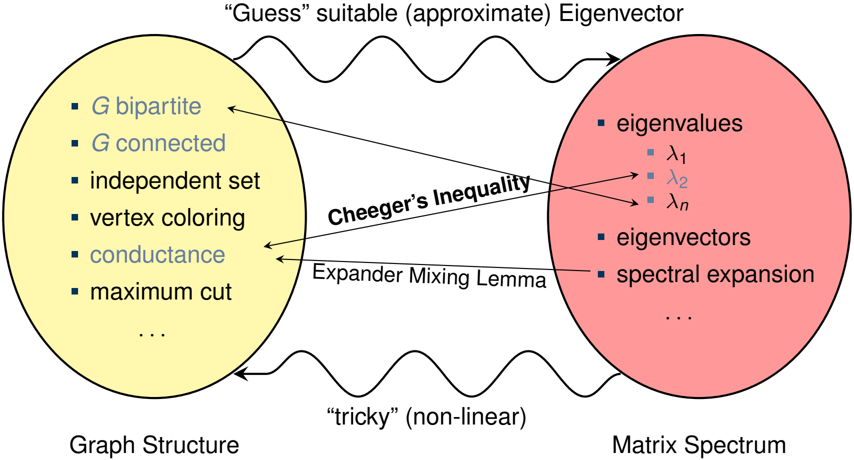
  $\Rightarrow$ choose smallest $k \geq 2$ so that the spectral gap $\lambda_{k+1} - \lambda_k$ is "large"

- In the latter example $\lambda = \{0, 0.20, 0.22, 0.43, 0.45, \ldots\} \implies k = 3.$

- In the former example $\lambda = \{0, 0.15, 0.37, 0.40, 0.43, \ldots\} \implies k = 2.$

- For $k = 2$ use sweep-cut extract clusters. For $k \geq 3$ use embedding in k-dimensional space and apply k-means (geometric clustering)

Summary (1/2): Graph Structure vs. Matrix Spectrum

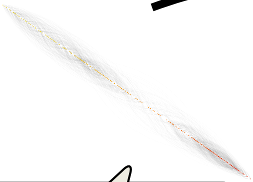


Summary (2/2): Spectral Clustering

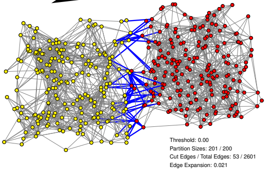
Spectral Embedding onto Line

Compute Sweep Cuts

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$



$$\min_{\substack{x \in \mathbb{R}^n \setminus \{\mathbf{0}\} \\ x \perp 1}} \frac{\sum_{u \sim v} (x_u - x_v)^2}{\sum_u x_u^2}$$

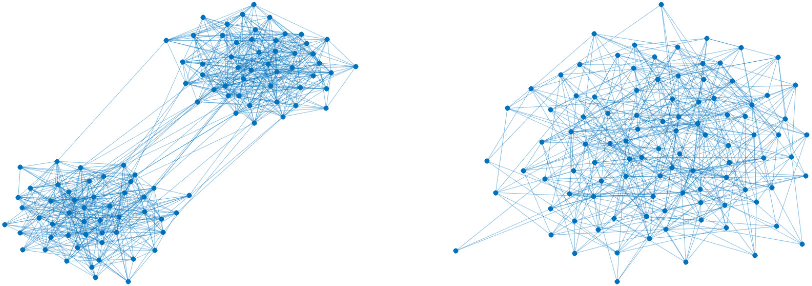


- Given any graph (adjacency matrix)
- Graph Spectrum (computable in poly-time)
  - λ_2 (relates to connectivity)
  - λ_n (relates to bipartiteness)
  - ...
- Cheeger's Inequality
  - relates λ_2 to conductance
  - unbounded approximation ratio
  - effective in practice

Outline

Relation between Clustering and Mixing

- Which graph has a “cluster-structure”?
- Which graph mixes faster?



Convergence of Random Walk

Recall: If the underlying graph G is connected, undirected and d-regular, then the random walk converges towards the stationary distribution $\pi = (1/n, \ldots, 1/n)$, which satisfies $\pi \mathbf{P} = \pi$.

Here all vector multiplications (including eigenvectors) will always be from the left!

Lemma

Consider a lazy random walk on a connected, undirected and d-regular graph. Then for any initial distribution x,

$$\left\| x\mathbf{P}^t - \pi \right\|_2 \leq \lambda^t,$$

with $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ as eigenvalues and $\lambda := \max\{|\lambda_2|, |\lambda_n|\}$.
$\Rightarrow$ This implies for $t = \mathcal{O}\left(\frac{\log n}{\log(1/\lambda)}\right) = \mathcal{O}\left(\frac{\log n}{1-\lambda}\right)$,

due to laziness, $\lambda_n \geq 0$

$$\left\| x\mathbf{P}^t - \pi \right\|_{tv} \leq \frac{1}{4}.$$

Proof of Lemma

- Express x in terms of the orthonormal basis of $\mathbf{P}$, $v_1 = \pi, v_2, \ldots, v_n$:

$$x = \sum_{i=1}^{n} \alpha_i v_i.$$

- Since x is a probability vector and all $v_i \geq 2$ are orthogonal to π, $\alpha_1 = 1$.

$\Rightarrow$

$$\begin{aligned}
\|x\mathbf{P} - \pi\|_2^2 &= \left\| \left(\sum_{i=1}^{n} \alpha_i v_i \right) \mathbf{P} - \pi \right\|_2^2 \\
&= \left\| \pi + \sum_{i=2}^{n} \alpha_i \lambda_i v_i - \pi \right\|_2^2 \\
&= \left\| \sum_{i=2}^{n} \alpha_i \lambda_i v_i \right\|_2^2 \\
&= \sum_{i=2}^{n} \|\alpha_i \lambda_i v_i\|_2^2 \\
&\leq \lambda^2 \sum_{i=2}^{n} \|\alpha_i v_i\|_2^2 = \lambda^2 \left\| \sum_{i=2}^{n} \alpha_i v_i \right\|_2^2 = \lambda^2 \|x - \pi\|_2^2
\end{aligned}$$

since the v_i's are orthogonal

since the v_i's are orthogonal

- Hence $\|x\mathbf{P}^t - \pi\|_2^2 \leq \lambda^{2t} \cdot \|x - \pi\|_2^2 \leq \lambda^{2t} \cdot 1.$

$\|x - \pi\|_2^2 + \|\pi\|_2^2 = \|x\|_2^2 \leq 1$

Outline

Similarity graph

Given $X = \{x_1, \ldots, x_n\} \in \mathbb{R}^d$, construct $G = (V, E, w)$:

- $x_i \in X \mapsto v_i \in V$
- $E = \binom{V}{2}$
- $w(v_i, v_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$ (Gaussian similarity function)

Remarks:

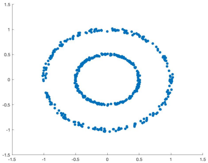
- $w(v_i, v_j)$ is large if x_i is close to x_j
- value of $\sigma \geq 0$ depends on the application (choose it by trial and error, usually $\sigma \in (0.05, 10)$)
- large σ if, on average, pairwise nearest neighbours are far apart

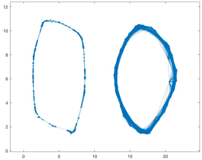
Problem: Since G is complete, from $\Theta(dn)$ to $\Theta(n^2)$ space.

Possible solution: r-nearest neighbour graph ($v_i \sim v_j$ iff x_j is one of the r-nearest neighbours of x_i or vice versa)

From geometric to graph clustering!

Example

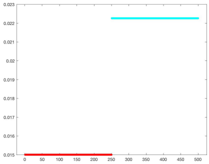




Similarity graph: Gaussian with $\sigma = 0.1$. Only edges with weight ≥ 0.01 shown.

Spectral Clustering (variant for non-regular graphs)

1. Compute the eigenvector x corresponding to λ_2 and $y = \mathbf{D}^{-1/2}x$.
2. Order the vertices so that $y_1 \leq y_2 \leq \cdots \leq y_n$
3. Choose "sweep" cut $(\{1, 2, \ldots, i\}, \{i+1, \ldots, n\})$ with smallest conductance



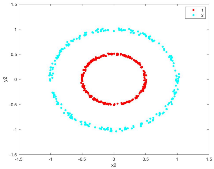


Image segmentation

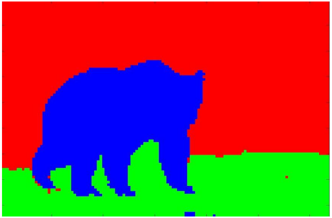
Goal: identify different objects in an image

Construct similarity graph as follows:

- A pixel p is characterised by its position in the image and by its RGB value
- map pixel p in position (x, y) to a vector $v_p = (x, y, r, g, b)$
- construct similarity graph as explained earlier

Original image

Output SC (Gaussian, $\sigma = 10$)



References

Fan R.K. Chung.
Graph Theory in the Information Age.
Notices of the AMS, vol. 57, no. 6, pages 726–732, 2010.

Fan R.K. Chung.
Spectral Graph Theory.
Volume 92 of CBMS Regional Conference Series in Mathematics, 1997.

S. Hoory, N. Linial and A. Widgerson.
Expander Graphs and their Applications.
Bulletin of the AMS, vol. 43, no. 4, pages 439–561, 2006.

Daniel Spielman
Chapter 16, Spectral Graph Theory
Combinatorial Scientific Computing

Luca Trevisan.
Lectures Notes on Expansion, Sparsest Cut, and Spectral Graph Theory, 2016.
https://lucatrevisan.github.io/books/expanders-2016.pdf