Randomised Algorithms

Lecture 12: Spectral Graph Clustering

Thomas Sauerwald (tms41@cam.ac.uk)

Lent 2025

Outline

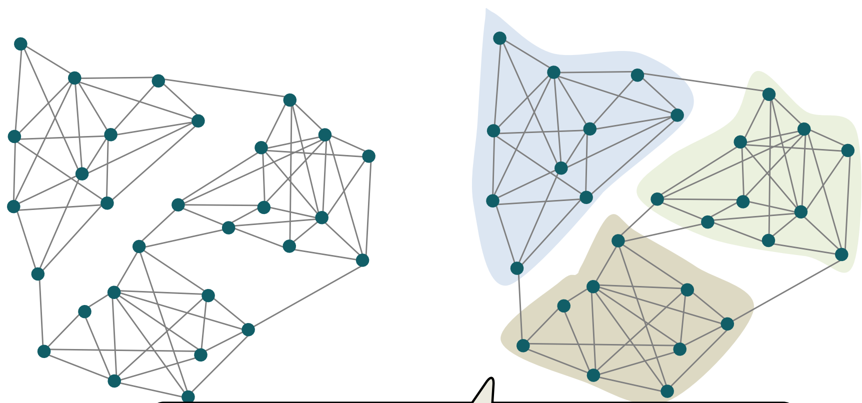
Conductance, Cheeger's Inequality and Spectral Clustering

Illustrations of Spectral Clustering and Extension to Non-Regular Graphs

Appendix: Relating Spectrum to Mixing Times (non-examinable)

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

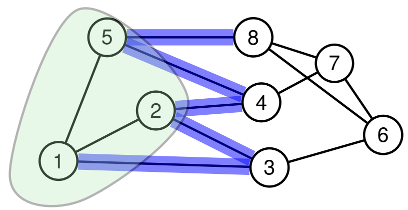
Let $G = (V, E)$ be a d-regular and undirected graph and $\emptyset \neq S \subsetneq V$. The conductance (edge expansion) of S is

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

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

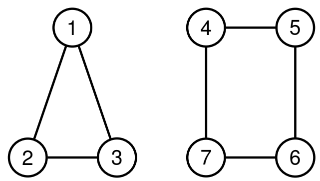
$$\phi(G) := \min_{S \subseteq V\colon 1 \leq |S| \leq n/2} \phi(S)$$

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?

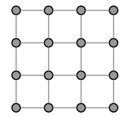
λ_2 versus Conductance (2/2)

1D Grid (Path)

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

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

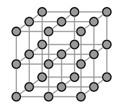
2D Grid



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

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

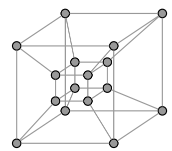
3D Grid



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

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

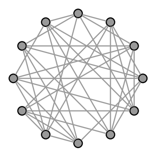
Hypercube



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

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

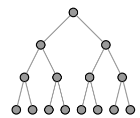
Random Graph (Expanders)



$\lambda_2 = \Theta(1)$

$\phi = \Theta(1)$

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

$$\begin{aligned} \lambda_2 \leq \frac{1}{d} \cdot \frac{\sum_{u \sim v} (y_u - y_v)^2}{\sum_u y_u^2} &= \frac{1}{d} \cdot \frac{|E(S, V \setminus S)| \cdot \left(\frac{1}{|S|} + \frac{1}{|V \setminus S|}\right)^2}{\frac{1}{|S|} + \frac{1}{|V \setminus S|}} \\ &= \frac{1}{d} \cdot |E(S, V \setminus S)| \cdot \left(\frac{1}{|S|} + \frac{1}{|V \setminus S|}\right) \\ &\leq \frac{1}{d} \cdot \frac{2 \cdot |E(S, V \setminus S)|}{|S|} = 2 \cdot \phi(G). \qquad \square \end{aligned}$$

Outline

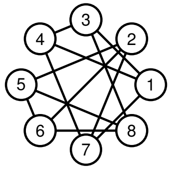
Conductance, Cheeger's Inequality and Spectral Clustering

Illustrations of Spectral Clustering and Extension to Non-Regular Graphs

Appendix: Relating Spectrum to Mixing Times (non-examinable)

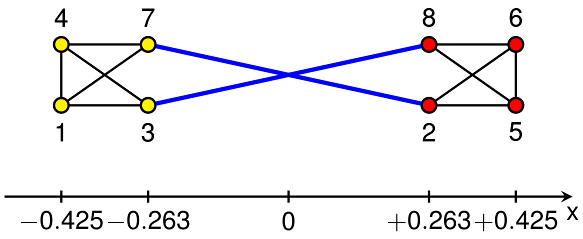
Illustration on a small Example

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



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

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



Sweep: 4

Conductance: 0.166

Physical Interpretation of the Minimisation Problem

- For each edge $\{u, v\} \in E(G)$, add spring between pins at x_u and x_v
- The potential energy at each spring is $(x_u - x_v)^2$
- Courant-Fisher characterisation:

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

- In our example, we found out that $\lambda_2 \approx 0.25$
- The eigenvector x on the last slide is normalised (i.e., $\|x\|_2^2 = 1$). Hence,

$$\lambda_2 = \frac{1}{3} \cdot \left((x_1 - x_3)^2 + (x_1 - x_4)^2 + (x_1 - x_7)^2 + \cdots + (x_6 - x_8)^2 \right) \approx 0.25$$

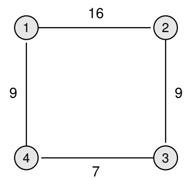
Let us now look at an example of a non-regular graph!

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 such that $\mathbf{D}_{uu} = \deg(u) = \sum_{v:\,\{u,v\}\in E} w(u,v)$, 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

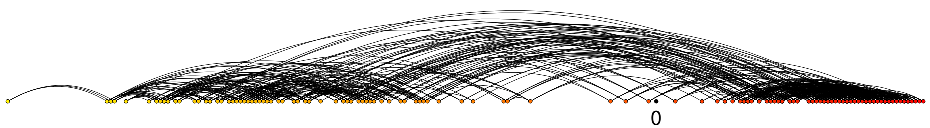
$G = (V, E)$ with clusters $S_1, S_2 \subseteq V$, $0 \leq q < p \leq 1$

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

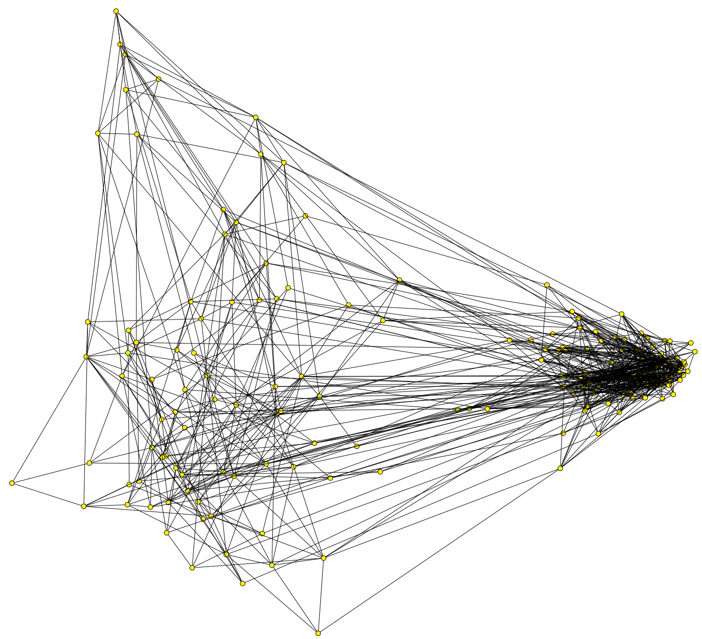
Here:

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

```
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
```

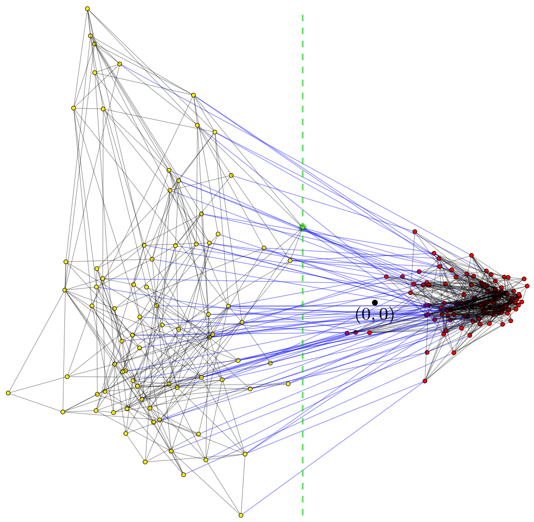


Drawing the 2D-Embedding

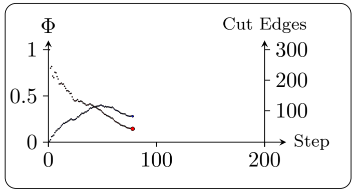


Best Solution found by Spectral Clustering

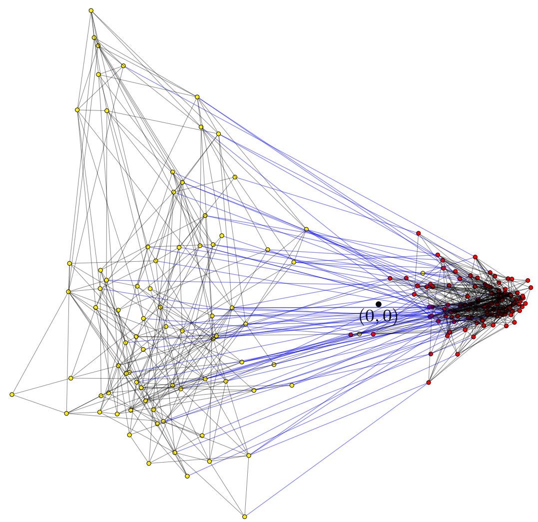
For the complete animation, see the full slides.



- Step: 78
- Threshold: −0.0336
- Partition Sizes: 78/122
- Cut Edges: 84
- Conductance: 0.1448



Clustering induced by Blocks



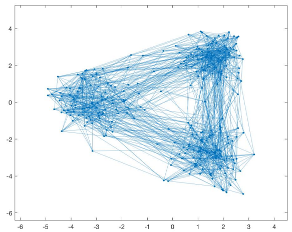
- Step: –
- Threshold: –
- Partition Sizes: 80/120
- Cut Edges: 88
- Conductance: 0.1486

Additional Example: Stochastic Block Models with 3 Clusters

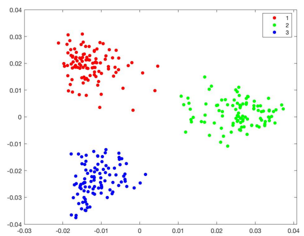
Graph $G = (V, E)$ with clusters
$S_1, S_2, S_3 \subseteq V; \quad 0 \leq q < p \leq 1$

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

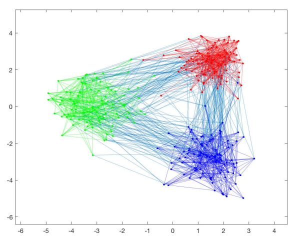
$|V| = 300, |S_i| = 100$
$p = 0.08, q = 0.01.$



Spectral embedding



Output of Spectral Clustering

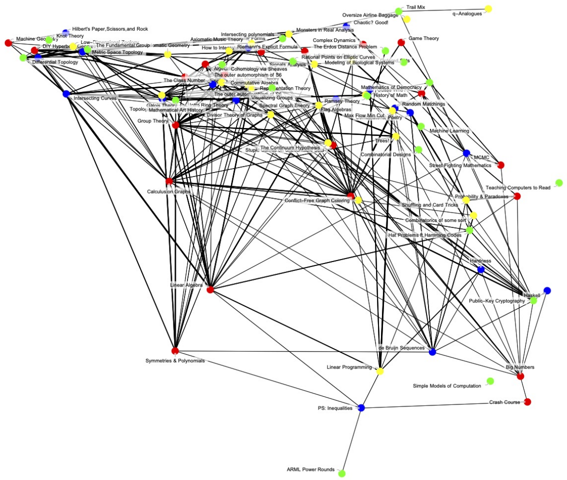


How to Choose the Cluster Number k

- 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 (cannot be divided further)

  ⇒ 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)

Another Example



(many thanks to Kalina Jasinska)

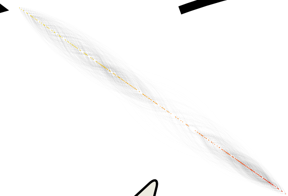
- nodes represent math topics taught within 4 weeks of a Mathcamp
- node colours represent to the week in which they thought
- teachers were asked to assign weights in 0 – 10 indicating how closely related two classes are

Summary: Spectral Clustering

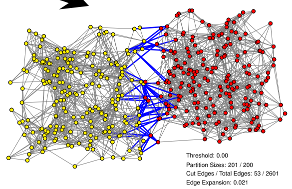
Spectral Embedding onto Line

Compute Sweep Cuts

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



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



- 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

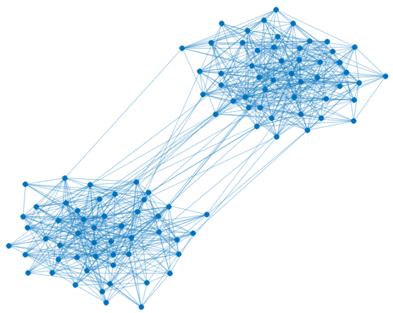
Conductance, Cheeger's Inequality and Spectral Clustering

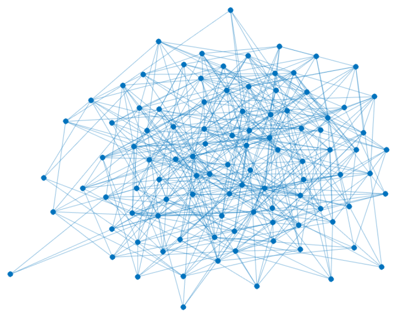
Illustrations of Spectral Clustering and Extension to Non-Regular Graphs

Appendix: Relating Spectrum to Mixing Times (non-examinable)

Relation between Clustering and Mixing (non-examinable)

- Which graph has a "cluster-structure"?
- Which graph mixes faster?





Convergence of Random Walk (non-examinable)

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

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

for lazy random walks, $\lambda_n \geq 0$

Proof of Lemma (non-examinable)

- 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, $\alpha_1 = 1$.

$\Rightarrow$

$$\|x\mathbf{P} - \pi\|_2^2 = \left\| \left(\sum_{i=1}^n \alpha_i v_i \right) \mathbf{P} - \pi \right\|_2^2$$

$$= \left\| \pi + \sum_{i=2}^n \alpha_i \lambda_i v_i - \pi \right\|_2^2$$

$$= \left\| \sum_{i=2}^n \alpha_i \lambda_i v_i \right\|_2^2$$

$$= \sum_{i=2}^n \|\alpha_i \lambda_i v_i\|_2^2$$

$$\leq \lambda^2 \sum_{i=2}^n \|\alpha_i v_i\|_2^2 = \lambda^2 \left\| \sum_{i=2}^n \alpha_i v_i \right\|_2^2 = \lambda^2 \, \|x - \pi\|_2^2$$

since the v_i's are orthogonal

since the v_i's are orthogonal

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

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

Some References on Spectral Graph Theory and Clustering

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, 2010.

Luca Trevisan.
Lectures Notes on Graph Partitioning, Expanders and Spectral Methods, 2017.
https://lucatrevisan.github.io/books/expanders-2016.pdf

The End...

Thank you and Best Wishes for the Exam!

I'm very interested to hear your feedback about the slides and the course more generally. You can use the student feedback form or send me an email during or after the course (tms41@cam.ac.uk).