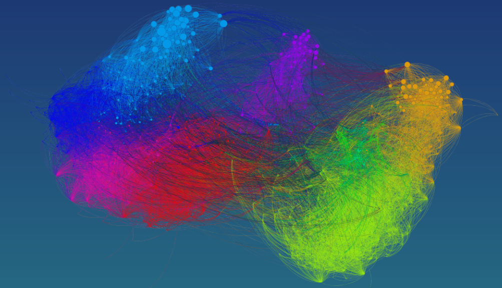
Graphs in Machine Learning

Michal Valko

Inria Lille - Nord Europe, France

TA: Pierre Perrault

Partially based on material by: Ulrike von Luxburg, Gary Miller, Mikhail Belkin



Previous Lecture

- similarity graphs
  - different types
  - construction
  - sources of graphs
  - practical considerations
- spectral graph theory
- Laplacians and their properties
  - symmetric and asymmetric normalization
- random walks
- recommendation on a bipartite graph
- resistive networks
  - recommendation score as a resistance?
  - Laplacian and resistive networks
  - resistance distance and random walks

This Lecture

- geometry of the data and the connectivity
- spectral clustering
- manifold learning with Laplacians eigenmaps
- Gaussian random fields and harmonic solution
- graph-based semi-supervised learning and manifold regularization
- transductive learning
- inductive and transductive semi-supervised learning

Next Class: Lab Session

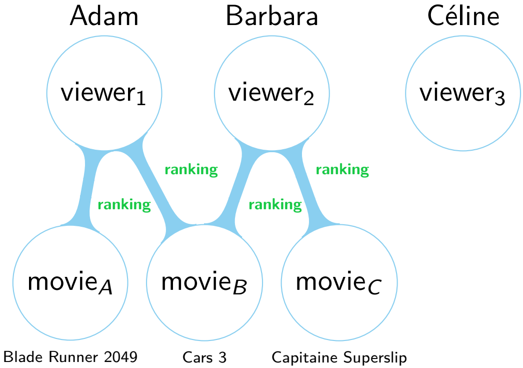
- 23. 10. 2017 by Pierre Perrault
- cca. 10h00-10h30 optional help with setup, 10h30-12h30: TD
- Salle Condorcet
- Download the image and set it up BEFORE the class
- Matlab/Octave
- Short written report (graded)
- All homeworks together account for 40% of the final grade
- Content
  - Graph Construction
  - Test sensitivity to parameters: σ, k, ε
  - Spectral Clustering
  - Spectral Clustering vs. k-means
  - Image Segmentation

PhD student position on the topic of *sequential learning* (mathematical statistics and machine learning) at Uni Magdeburg. PhD candidate will focus on developing sequential learning algorithms with mathematical guarantees for learning on given non-stationary processes that are relevant in the context of recommendation systems, and on implementation of the algorithms that will be developed. S/He will also work on the eye tracker based application of the project. A degree in machine learning or in mathematics with an interest in theoretical computer science will be preferred. Uni Magdeburg (90 minutes away from Berlin, Germany, by public transports) and universities in nearby Berlin offer a highly motivating and rich research environment.

The PhD candidates will be advised by Dr. Alexandra Carpentier (contact https://www.uni-potsdam.de/fileadmin01/projects/sfb1294/Ausschreibung_SFB1294_A03_2ndround-4-2.pdf for more info).

Use of Laplacians: Movie recommendation

How to do movie recommendation on a bipartite graph?



Question: *Do we recommend Capitaine Superslip to Adam?*

Let's compute some $\mathrm{score}(v, m)$!

Use of Laplacians: Movie recommendation

How to compute the $\mathrm{score}(v, m)$? Using some **graph distance**!

Idea_1: maximally weighted path

$$\mathrm{score}(v, m) = \max_{vPm} \mathrm{weight}(P) = \max_{vPm} \sum_{e \in P} \mathrm{ranking}(e)$$

Idea_2: change the path weight

$$\mathrm{score}_2(v, m) = \max_{vPm} \mathrm{weight}_2(P) = \max_{vPm} \min_{e \in P} \mathrm{ranking}(e)$$

Idea_3: consider everything

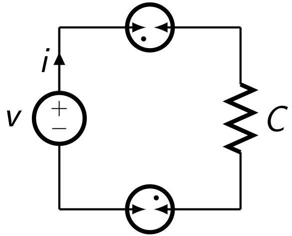
$$\mathrm{score}_3(v, m) = \text{max flow from } m \text{ to } v$$

Laplacians and Resistive Networks

How to compute the $\text{score}(v, m)$?

Idea$_4$: view edges as conductors

$\text{score}_4(v, m)$ = effective resistance between m and v



$C \equiv$ conductance

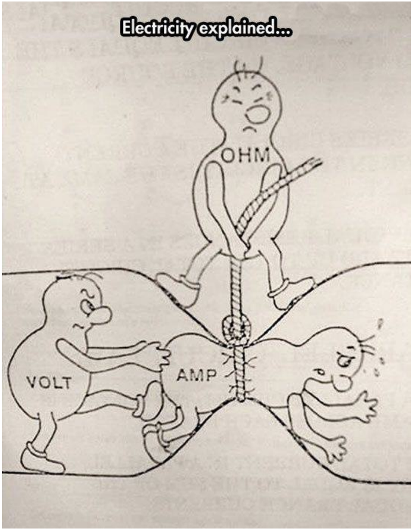
$R \equiv$ resistance

$i \equiv$ current

$V \equiv$ voltage

$$C = \frac{1}{R} \qquad i = CV = \frac{V}{R}$$

Resistive Networks: Some high-school physics



Resistive Networks

resistors in series

$$R = R_1 + \cdots + R_n \qquad C = \frac{1}{\frac{1}{C_1} + \cdots + \frac{1}{C_N}} \qquad i = \frac{V}{R}$$

conductors in parallel

$$C = C_1 + \cdots + C_N \qquad i = VC$$

Effective Resistance on a graph

Take two nodes: $a \neq b$. Let V_{ab} be the voltage between them and i_{ab} the current between them. Define $R_{ab} = \frac{V_{ab}}{i_{ab}}$ and $C_{ab} = \frac{1}{R_{ab}}$.

We treat the entire graph as a resistor!

Resistive Networks: Optional Homework (ungraded)

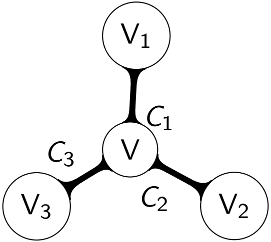
Show that R_{ab} is a metric space.

1. $R_{ab} \geq 0$
2. $R_{ab} = 0$ iff $a = b$
3. $R_{ab} = R_{ba}$
4. $R_{ac} \leq R_{ab} + R_{bc}$

The effective resistance is a distance!

How to compute effective resistance?

Kirchhoff's Law $\equiv$ flow in $=$ flow out



$V = \frac{C_1}{C}V_1 + \frac{C_2}{C}V_2 + \frac{C_3}{C}V_3$ (convex combination)

residual current $= CV - C_1V_1 - C_2V_2 - C_3V_3$

Kirchhoff says: This is zero! **There is no residual current!**

Resistors: Where is the link with the Laplacian?

General case of the previous! $d_i = \sum_j c_{ij}$ = sum of conductances

$$\mathbf{L}_{ij} = \begin{cases} d_i & \text{if } i = j, \\ -c_{ij} & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

$\mathbf{v}$ = **voltage setting** of the nodes on graph.

$(\mathbf{Lv})_i$ = residual current at $\mathbf{v}_i$ — as we derived

Use: setting voltages and getting the current

Inverting $\equiv$ injecting current and getting the voltages

The net injected has to be zero $\equiv$ Kirchhoff's Law.

Resistors and the Laplacian: Finding R_{ab}

Let's calculate R_{1N} to get the **movie recommendation score**!

$$\mathbf{L}\begin{pmatrix} 0 \\ v_2 \\ \vdots \\ v_{n-1} \\ 1 \end{pmatrix} = \begin{pmatrix} i \\ 0 \\ \vdots \\ 0 \\ -i \end{pmatrix}$$

$$i = \frac{V}{R} \qquad V = 1 \qquad R = \frac{1}{i}$$

Return $R_{1N} = \dfrac{1}{i}$

Doyle and Snell: Random Walks and Electric Networks

https://math.dartmouth.edu/~doyle/docs/walks/walks.pdf

Resistors and the Laplacian: Finding R_{1N}

$\mathbf{Lv} = (i, 0, \ldots, -i)^\mathsf{T} \equiv$ **boundary valued problem**

For R_{1N}

V_1 and V_N are the **boundary**

$(v_1, v_2, \ldots, v_N)$ is **harmonic:**

$V_i \in$ **interior** (not boundary)

V_i is a **convex combination of its neighbors**

Resistors and the Laplacian: Finding R_{1n}

From the properties of electric networks (cf. Doyle and Snell) we inherit the useful properties of the Laplacians!

Example: Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions (later in the course)

Maximum Principle

If $\mathbf{f} = \mathbf{v}$ is harmonic then min and max are on the boundary.

Uniqueness Principle

If $\mathbf{f}$ and $\mathbf{g}$ are harmonic with the same boundary then $\mathbf{f} = \mathbf{g}$

Resistors and the Laplacian: Finding R_{1N}

Alternative method to calculate R_{1N}:

$$\mathbf{L}\mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ -1 \end{pmatrix} \stackrel{\text{def}}{=} \mathbf{i}_{\text{ext}} \quad \text{Return} \quad R_{1N} = v_1 - v_N$$

Why?

Question: Does $\mathbf{v}$ exist? $\mathbf{L}$ does not have an inverse :(.
Not unique: $\mathbf{1}$ in the nullspace of $\mathbf{L}$: $\mathbf{L}(\mathbf{v} + c\mathbf{1}) = \mathbf{L}\mathbf{v} + c\mathbf{L}\mathbf{1} = \mathbf{L}\mathbf{v}$
Moore-Penrose pseudo-inverse solves LS
Solution: Instead of $\mathbf{v} = \mathbf{L}^{-1}\mathbf{i}_{\text{ext}}$ we take $\mathbf{v} = \mathbf{L}^{+}\mathbf{i}_{\text{ext}}$
We get: $R_{1N} = v_1 - v_N = \mathbf{i}_{\text{ext}}^{\mathsf{T}}\mathbf{v} = \mathbf{i}_{\text{ext}}^{\mathsf{T}}\mathbf{L}^{+}\mathbf{i}_{\text{ext}}$.
Notice: We can reuse $\mathbf{L}^{+}$ to get resistances for any pair of nodes!

What? A pseudo-inverse?

Eigendecomposition of the Laplacian:

$$\mathbf{L} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{\mathsf{T}} = \sum_{i=1}^{N} \lambda_i \mathbf{q}_i \mathbf{q}_i^{\mathsf{T}} = \sum_{i=2}^{N} \lambda_i \mathbf{q}_i \mathbf{q}_i^{\mathsf{T}}$$

Pseudo-inverse of the Laplacian:

$$\mathbf{L}^{+} = \mathbf{Q}\boldsymbol{\Lambda}^{+}\mathbf{Q}^{\mathsf{T}} = \sum_{i=2}^{N} \frac{1}{\lambda_i} \mathbf{q}_i \mathbf{q}_i^{\mathsf{T}}$$

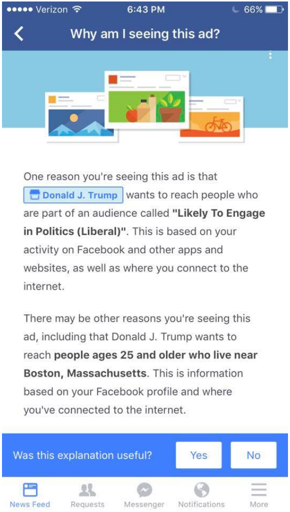
Moore-Penrose pseudo-inverse solves a least squares problem:

$$\mathbf{v} = \arg\min_{\mathbf{x}} \|\mathbf{L}\mathbf{x} - \mathbf{i}_{\text{ext}}\|_2 = \mathbf{L}^{+}\mathbf{i}_{\text{ext}}$$

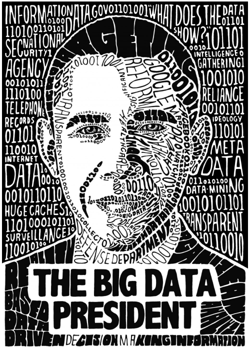
How to rule the world?

Let's make France great again!

How to rule the world?



How to rule the world: “AI” is here

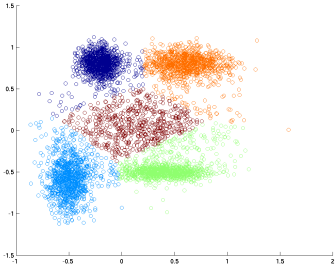


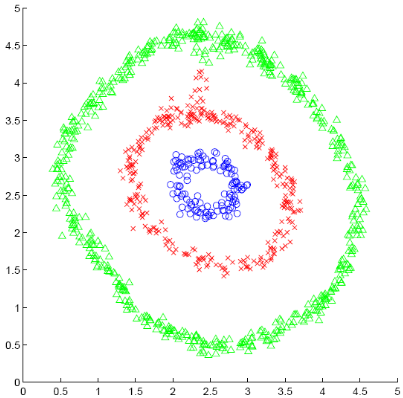
https://www.washingtonpost.com/opinions/obama-the-big-data-president/2013/06/14/1d71fe2e-d391-11e2-b05f-3ea3f0e7bb5a_story.html

https://www.technologyreview.com/s/509026/how-obamas-team-used-big-data-to-rally-voters/

Talk of Rayid Ghaniy: https://www.youtube.com/watch?v=gDM1GuszM_U

Application of Graphs for ML: Clustering

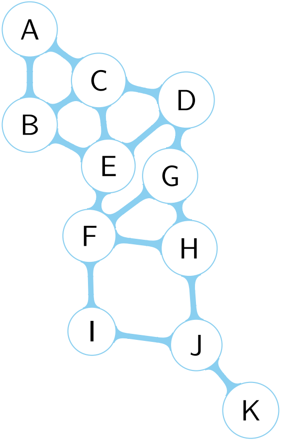




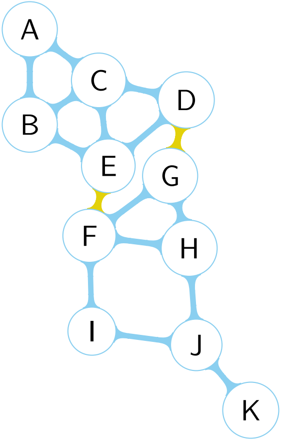
Application: Clustering - Recap

- What do we know about the **clustering** in general?
  - ill defined problem (different tasks $\rightarrow$ different paradigms)
  - "I know it when I see it"
  - inconsistent (wrt. Kleinberg's axioms)
  - number of clusters k need often be known
  - difficult to evaluate
- What do we know about k-**means**?
  - "hard" version of EM clustering
  - sensitive to initialization
  - optimizes for **compactness**
  - yet: algorithm-to-go

Spectral Clustering: Cuts on graphs

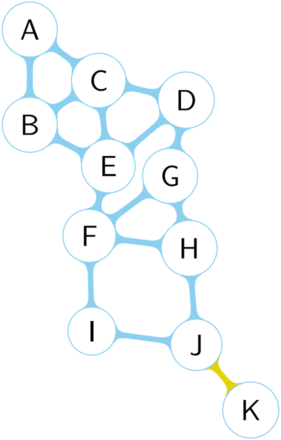


Spectral Clustering: Cuts on graphs



Defining the cut objective we get the clustering!

Spectral Clustering: Cuts on graphs



MinCut: $\text{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij}$

Are we done?

Can be solved efficiently, but maybe not what we want . . . .

Spectral Clustering: Balanced Cuts

Let's balance the cuts!

MinCut

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij}$$

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Normalized Cut

$$\mathrm{NCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)} \right)$$

Spectral Clustering: Balanced Cuts

$$\mathrm{RatioCut}(A, B) = \mathrm{cut}(A, B)\left(\frac{1}{|A|} + \frac{1}{|B|}\right)$$

$$\mathrm{NCut}(A, B) = \mathrm{cut}(A, B)\left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)}\right)$$

Can we compute this? RatioCut and NCut are NP hard :(

Approximate!

Spectral Clustering: Relaxing Balanced Cuts

Relaxation for (simple) balanced cuts for 2 sets

$$\min_{A,B} \mathrm{cut}(A, B) \quad \text{s.t.} \quad |A| = |B|$$

Graph function $\mathbf{f}$ for cluster membership: $f_i = \begin{cases} 1 & \text{if } V_i \in A, \\ -1 & \text{if } V_i \in B. \end{cases}$

What it is the cut value with this definition?

$$\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4} \sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

What is the relationship with the **smoothness** of a graph function?

Spectral Clustering: Relaxing Balanced Cuts

$$\mathrm{cut}(A,B) = \sum_{i \in A, j \in B} w_{i,j} = \frac{1}{4}\sum_{i,j} w_{i,j}(f_i - f_j)^2 = \frac{1}{2}\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f}$$

$$|A| = |B| \implies \sum_i f_i = 0 \implies \mathbf{f} \perp \mathbf{1}_N$$

$$\|\mathbf{f}\| = \sqrt{N}$$

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i = \pm 1, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Still NP hard :($\rightarrow$ Relax even further!

$$\cancel{f_i = \pm 1} \quad \rightarrow \quad f_i \in \mathbb{R}$$

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Rayleigh-Ritz theorem

If $\lambda_1 \leq \cdots \leq \lambda_N$ are the eigenvectors of real symmetric $\mathbf{L}$ then

$$\lambda_1 = \min_{\mathbf{x} \neq 0} \frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} = \min_{\mathbf{x}^{\mathsf{T}}\mathbf{x}=1} \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$$

$$\lambda_N = \max_{\mathbf{x} \neq 0} \frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} = \max_{\mathbf{x}^{\mathsf{T}}\mathbf{x}=1} \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$$

$\frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} \equiv$ **Rayleigh quotient**

How can we use it?

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Generalized Rayleigh-Ritz theorem (Courant-Fischer-Weyl)

If $\lambda_1 \leq \cdots \leq \lambda_N$ are the eigenvectors of real symmetric $\mathbf{L}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_N$ the corresponding orthogonal eigenvalues, then for $k = 1 : N - 1$

$$\lambda_{k+1} = \min_{\mathbf{x} \neq 0, \mathbf{x} \perp \mathbf{v}_1, \ldots \mathbf{v}_k} \frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} = \min_{\mathbf{x}^{\mathsf{T}}\mathbf{x} = 1, \mathbf{x} \perp \mathbf{v}_1, \ldots \mathbf{v}_k} \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$$

$$\lambda_{N-k} = \max_{\mathbf{x} \neq 0, \mathbf{x} \perp \mathbf{v}_n, \ldots \mathbf{v}_{N-k+1}} \frac{\mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}}{\mathbf{x}^{\mathsf{T}}\mathbf{x}} = \max_{\mathbf{x}^{\mathsf{T}}\mathbf{x} = 1, \mathbf{x} \perp \mathbf{v}_N, \ldots \mathbf{v}_{N-k+1}} \mathbf{x}^{\mathsf{T}}\mathbf{L}\mathbf{x}$$

Rayleigh-Ritz theorem: Quick and dirty proof

When we reach the extreme points?

$$\frac{\partial}{\partial \mathbf{x}}\left(\frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}}\right) = \frac{\partial}{\partial \mathbf{x}}\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) = 0 \iff f'(\mathbf{x})g(\mathbf{x}) = f(\mathbf{x})g'(\mathbf{x})$$

By matrix calculus (or just calculus):

$$\frac{\partial \mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{L}\mathbf{x} \quad \text{and} \quad \frac{\partial \mathbf{x}^\mathsf{T}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$$

When $f'(\mathbf{x})g(\mathbf{x}) = f(\mathbf{x})g'(\mathbf{x})$?

$$\mathbf{L}\mathbf{x}\,(\mathbf{x}^\mathsf{T}\mathbf{x}) = (\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x})\,\mathbf{x} \iff \mathbf{L}\mathbf{x} = \frac{\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{x}}{\mathbf{x}^\mathsf{T}\mathbf{x}}\mathbf{x} \iff \mathbf{L}\mathbf{x} = \lambda\mathbf{x}$$

Conclusion: Extremes are the eigenvectors with their eigenvalues

Spectral Clustering: Relaxing Balanced Cuts

objective function of spectral clustering

$$\min_{\mathbf{f}} \mathbf{f}^{\mathsf{T}}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Solution: **second eigenvector** How do we get the clustering?

The solution may not be integral. What to do?

$$\text{cluster}_i = \begin{cases} 1 & \text{if } f_i \geq 0, \\ -1 & \text{if } f_i < 0. \end{cases}$$

Works but this heuristics is often too simple. In practice, cluster $\mathbf{f}$ using k-means to get $\{C_i\}_i$ and assign:

$$\text{cluster}_i = \begin{cases} 1 & \text{if } i \in C_1, \\ -1 & \text{if } i \in C_{-1}. \end{cases}$$

Spectral Clustering: Approximating RatioCut

Wait, but we did not care about approximating mincut!

RatioCut

$$\mathrm{RatioCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{|A|} + \frac{1}{|B|} \right)$$

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} = \frac{1}{2} \sum_{i,j} w_{i,j}(f_i - f_j)^2 = (|A| + |B|)\mathrm{RatioCut}(A, B)$$

Spectral Clustering: Approximating RatioCut

Define graph function $\mathbf{f}$ for cluster membership of RatioCut:

$$f_i = \begin{cases} \sqrt{\frac{|B|}{|A|}} & \text{if } V_i \in A, \\ -\sqrt{\frac{|A|}{|B|}} & \text{if } V_i \in B. \end{cases}$$

$$\sum_i f_i = 0$$

$$\sum_i f_i^2 = N$$

objective function of spectral clustering (same - it's magic!)

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{f} \perp \mathbf{1}_N, \quad \|\mathbf{f}\| = \sqrt{N}$$

Spectral Clustering: Approximating NCut

Normalized Cut

$$\mathrm{NCut}(A, B) = \sum_{i \in A, j \in B} w_{ij} \left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)} \right)$$

Define graph function **f** for cluster membership of NCut:

$$f_i = \begin{cases} \sqrt{\frac{\mathrm{vol}(A)}{\mathrm{vol}(B)}} & \text{if } V_i \in A, \\ -\sqrt{\frac{\mathrm{vol}(B)}{\mathrm{vol}(A)}} & \text{if } V_i \in B. \end{cases}$$

$$(\mathbf{Df})^\mathsf{T} \mathbf{1}_n = 0 \qquad \mathbf{f}^\mathsf{T} \mathbf{Df} = \mathrm{vol}(\mathcal{V}) \qquad \mathbf{f}^\mathsf{T} \mathbf{Lf} = \mathrm{vol}(\mathcal{V}) \mathrm{NCut}(A, B)$$

objective function of spectral clustering (NCut)

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T} \mathbf{Lf} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{Df} \perp \mathbf{1}_N, \quad \mathbf{f}^\mathsf{T} \mathbf{Df} = \mathrm{vol}(\mathcal{V})$$

Spectral Clustering: Approximating NCut

objective function of spectral clustering (NCut)

$$\min_{\mathbf{f}} \mathbf{f}^\mathsf{T}\mathbf{L}\mathbf{f} \quad \text{s.t.} \quad f_i \in \mathbb{R}, \quad \mathbf{D}\mathbf{f} \perp \mathbf{1}_N, \quad \mathbf{f}^\mathsf{T}\mathbf{D}\mathbf{f} = \mathrm{vol}(\mathcal{V})$$

Can we apply Rayleigh-Ritz now? Define $\mathbf{w} = \mathbf{D}^{1/2}\mathbf{f}$

objective function of spectral clustering (NCut)

$$\min_{\mathbf{w}} \mathbf{w}^\mathsf{T}\mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}\mathbf{w} \quad \text{s.t.} \quad w_i \in \mathbb{R}, \mathbf{w} \perp \mathbf{D}^{1/2}\mathbf{1}_N, \|\mathbf{w}\|^2 = \mathrm{vol}(\mathcal{V})$$

objective function of spectral clustering (NCut)

$$\min_{\mathbf{w}} \mathbf{w}^\mathsf{T}\mathbf{L}_{\mathrm{sym}}\mathbf{w} \quad \text{s.t.} \quad w_i \in \mathbb{R}, \quad \mathbf{w} \perp \mathbf{v}_{1,\mathbf{L}_{\mathrm{sym}}}, \quad \|\mathbf{w}\|^2 = \mathrm{vol}(\mathcal{V})$$

Spectral Clustering: Approximating NCut

objective function of spectral clustering (NCut)

$$\min_{\mathbf{w}} \mathbf{w}^\mathsf{T} \mathbf{L}_{\mathrm{sym}} \mathbf{w} \quad \text{s.t.} \quad w_i \in \mathbb{R}, \quad \mathbf{w} \perp \mathbf{v}_{1,\mathbf{L}_{\mathrm{sym}}}, \quad \|\mathbf{w}\| = \mathrm{vol}(\mathcal{V})$$

Solution by Rayleigh-Ritz? $\mathbf{w} = \mathbf{v}_{2,\mathbf{L}_{\mathrm{sym}}}$ $\mathbf{f} = \mathbf{D}^{-1/2}\mathbf{w}$

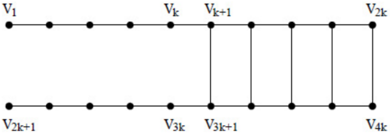
$\mathbf{f}$ is a the second eigenvector of $\mathbf{L}_{\mathrm{rw}}$!

tl;dr: Get the second eigenvector of $\mathbf{L}/\mathbf{L}_{\mathrm{rw}}$ for RatioCut/NCut.

Spectral Clustering: Approximation

These are all approximations. How bad can they be?

Example: cockroach graphs

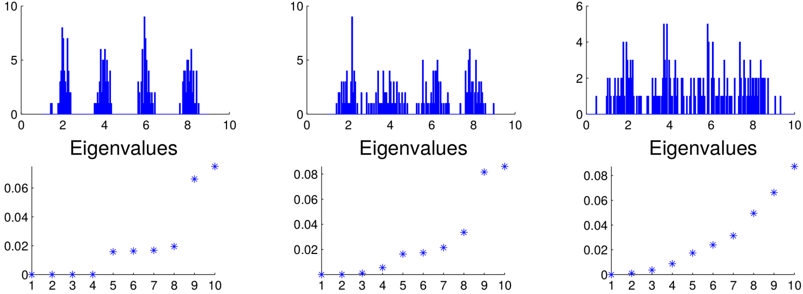


No efficient approximation exist. Other relaxations possible.

`https://www.cs.cmu.edu/~glmiller/Publications/Papers/GuMi95.pdf`

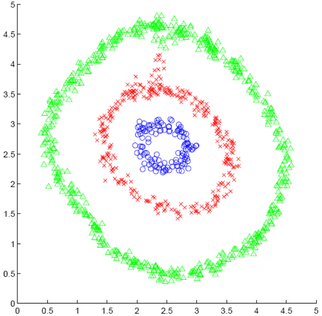
Spectral Clustering: 1D Example

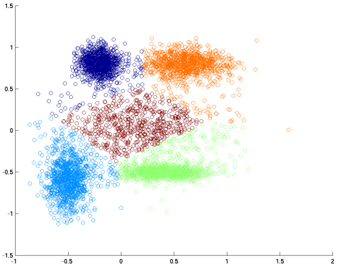
Elbow rule/EigenGap heuristic for number of clusters



Spectral Clustering: Understanding

Compactness vs. **Connectivity**

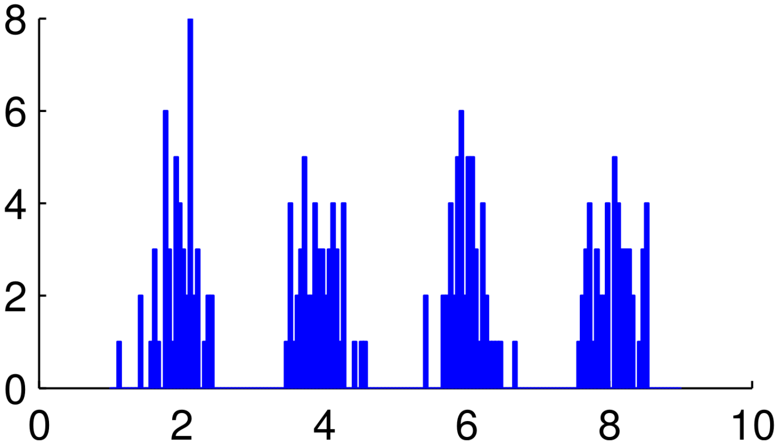




For which kind of data we can use one vs. the other?

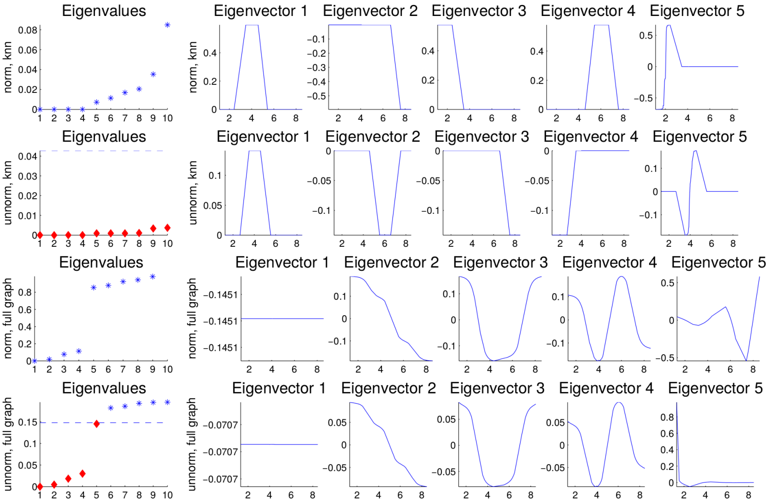
Any disadvantages of spectral clustering?

Spectral Clustering: 1D Example - Histogram



http://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/Luxburg07_tutorial.pdf

Spectral Clustering: 1D Example - Eigenvectors



Spectral Clustering: Bibliography

- M. Meila et al. "A random walks view of spectral segmentation". In: *International Conference on Artificial Intelligence and Statistics* (2001)
- $\mathbf{L}_{\mathrm{sym}}$ Andrew Y Ng, Michael I Jordan, and Yair Weiss. "On spectral clustering: Analysis and an algorithm". In: *Neural Information Processing Systems*. 2001
- $\mathbf{L}_{\mathrm{rm}}$ J Shi and J Malik. "Normalized Cuts and Image Segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (2000), pp. 888–905
- Things can go wrong with the relaxation: Daniel A. Spielman and Shang H. Teng. "Spectral partitioning works: Planar graphs and finite element meshes". In: *Linear Algebra and Its Applications* 421 (2007), pp. 284–305

Michal Valko
michal.valko@inria.fr
ENS Paris-Saclay, MVA 2017/2018
SequeL team, Inria Lille — Nord Europe
https://team.inria.fr/sequel/