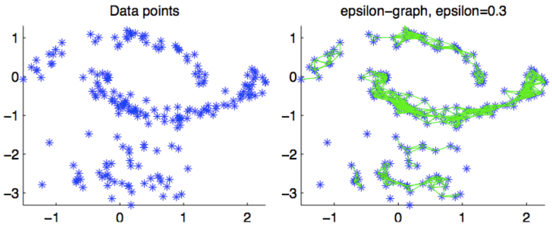
Freie Universität Berlin



A Tutorial on Spectral Clustering
Ulrike von Luxburg

Eric Kadikowski
Freie Universität Berlin

Numerik IVc, 1/30/13

Clustering

- set of n data points $x_1, \ldots, x_n$
- with similarites $s_{i,j} \geq 0$ between all pairs of data points $x_i, \ldots, x_j$

$\Updownarrow$ represent in a similarity graph $G = (V, E)$

- each vertex v_i represents a data point x_i
- two vertices are connected, if $s_{i,j}$ is larger than a certain threshold, and the edge ist weighted by $s_{i,j}$

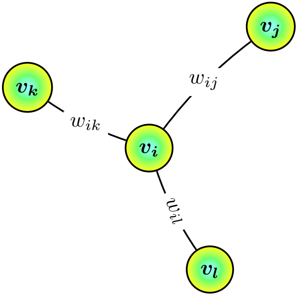
Graph Notation

The
weighted *adjacency matrix*
is the symmetric matrix

$$W := (w_{ij})_{i,j=1,\dots,n} \geq 0.$$

The *degree* of a
vertex $v_i \in V$ is definded as

$$d_i := \sum_{j=1}^{n} w_{ij} \text{ and}$$



$$d_i = w_{ij} + w_{ik} + w_{il}$$

the *degree matrix* as the diagonal matrix with $d_1 \dots d_n$ on the diagonal

$$D := \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{pmatrix}.$$

graph Laplacian $L := D - W$

Proposition (Properties of L)

1. *For every* $f \in \mathbb{R}^n$ *we have*

$$f^t L f = \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} \left(f_i - f_j\right)^2 .$$

2. *L is symmetric and positiv semi-definite.*
3. *The smallest eigenvalue of L is 0, the corresponding eigenvector is the constant vector $\mathbb{1}$.*
4. *L has n non-negative, real-valued eigenvalues* $0 = \lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$.

Proposition (Number of connected components and the spectrum of L)

Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of L equals the number of connected components $A_1, \ldots, A_k$ in the graph. The eigenspace of 0 is spanned by the indicator vectors $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k}$ of those components.

Proposition (Number of connected components and the spectrum of L)

Let G be an undirected graph with non-negative weights. Then the multiplicity k of the eigenvalue 0 of L equals the number of connected components $A_1, \ldots, A_k$ in the graph. The eigenspace of 0 is spanned by the indicator vectors $\mathbb{1}_{A_1}, \ldots, \mathbb{1}_{A_k}$ of those components.

Theorem (Davis-Kahan theorem from matrix perturbation theory)

In a "nearly ideal case" where we still have distinct clusters, but the between-cluster similarity is not exactly 0, we consider L to be a perturbed version of the ideal case. As the eigenvectors in the ideal case are piecewise constant on the connected components, the same will approximately be true in the perturbed case.

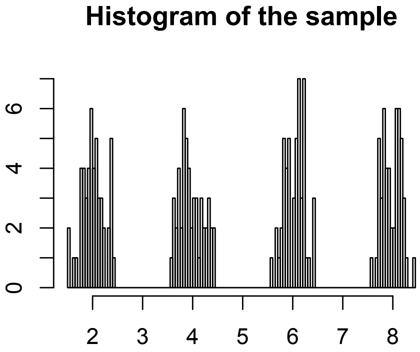
Spectral Clustering Algorithm

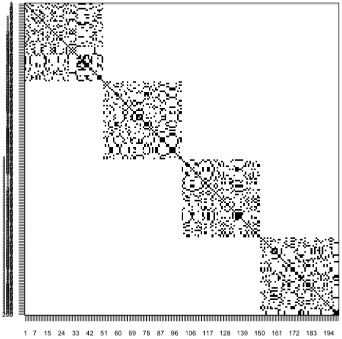
Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.

- Construct a similarity graph. Let W be its weighted adjacency matrix and D the degree matrix.
- Compute the Laplacian $L = D - W$.
- Compute the first k eigenvectors $u_1, \dots, u_k$ of L.
- Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \dots, u_k$ as columns.
- For $i = 1, \dots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.
- Cluster the points $(y_i)_{i=1,\dots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1, \dots, C_k$.

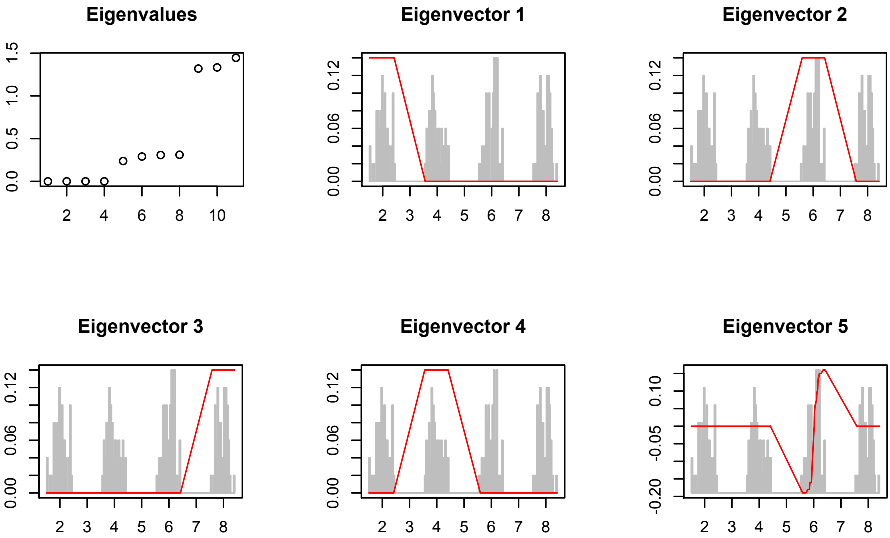
Output: Clusters $A_1, \dots, A_k$ with $A_i = \{j \mid y_j \in C_i\}$.

Toy Example


Histogram of the sample
0
2
4
6
2
3
4
5
6
7
8


1 7 15 24 33 42 51 60 69 78 87 96 106 117 128 139 150 161 172 183 194

Toy Example



Eigenvalues
0.0
0.5
1.0
1.5
2
4
6
8
10
Eigenvector 1
0.00
0.06
0.12
2
3
4
5
6
7
8
Eigenvector 2
0.00
0.06
0.12
2
3
4
5
6
7
8
Eigenvector 3
0.00
0.06
0.12
2
3
4
5
6
7
8
Eigenvector 4
0.00
0.06
0.12
2
3
4
5
6
7
8
Eigenvector 5
-0.20
-0.05
0.10
2
3
4
5
6
7
8

Spectral Clustering: Pros

spectral clustering

- does not make strong assumptions on the form of the clusters
  $\Rightarrow$ can solve very general problems like intertwined spirals
- can be implemented efficiently even for large data sets (as the adjacency matrix is sparse)
- no issues of getting stuck in local minima or restarting the algorithm for several times with different initializations

Spectral Clustering: Cons

but

- choosing a good similarity graph is not trivial
- spectral clustering can be quite unstable under different choices of the parameters for the similarity graph

⇒ Spectral clustering cannot serve as a “black box algorithm” which automatically detects the correct clusters in any given data set. But it can be considered as a powerful tool which can produce good results if applied with care.

End

Thanks!