Graphs for data science and ML

Machine Learning for graphs and with graphs

P. Borgnat, CNRS, LP ENSL (3)
Acknowledgements: some slides taken from P. Vandergheynst (EPFL)

Exploit the properties of the matrices of graphs

Second: find clusters, cut the graph

Cuts, clustering and communities

The good, the bad and the ugly

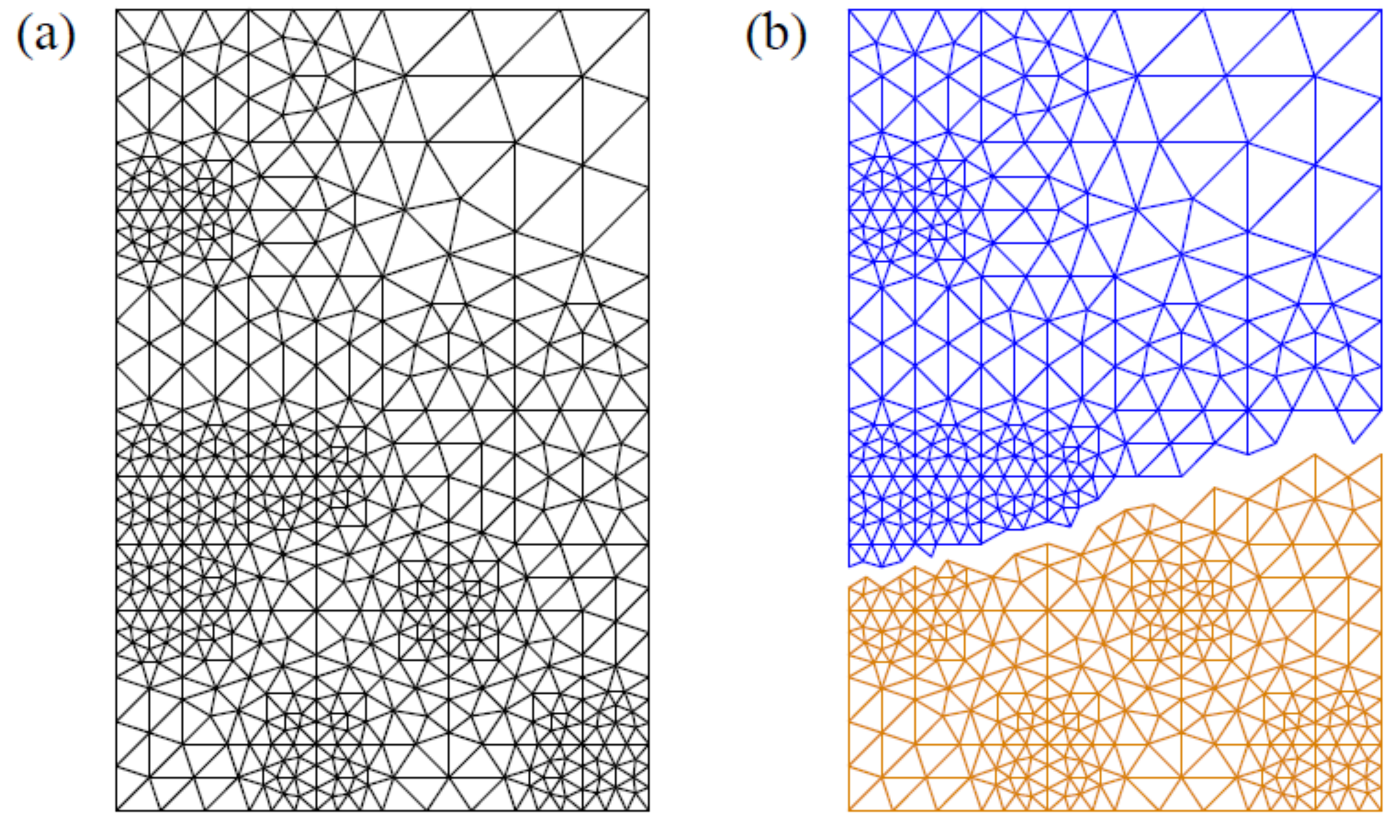
- Networks are often inhomogeneous, with important links, hubs, clusters, or communities (modules)
- These are observed in various types of data on networks: social, technological, biological,...
- Importance of cuts: the min-cut max-flow theorem.
  These are two primal-dual linear programs.
  The max value of a flow = the min capacity over all cuts.

- For clusters and communities, see the extensive surveys:

  [S. Fortunato, *Physic Reports*, 2010]

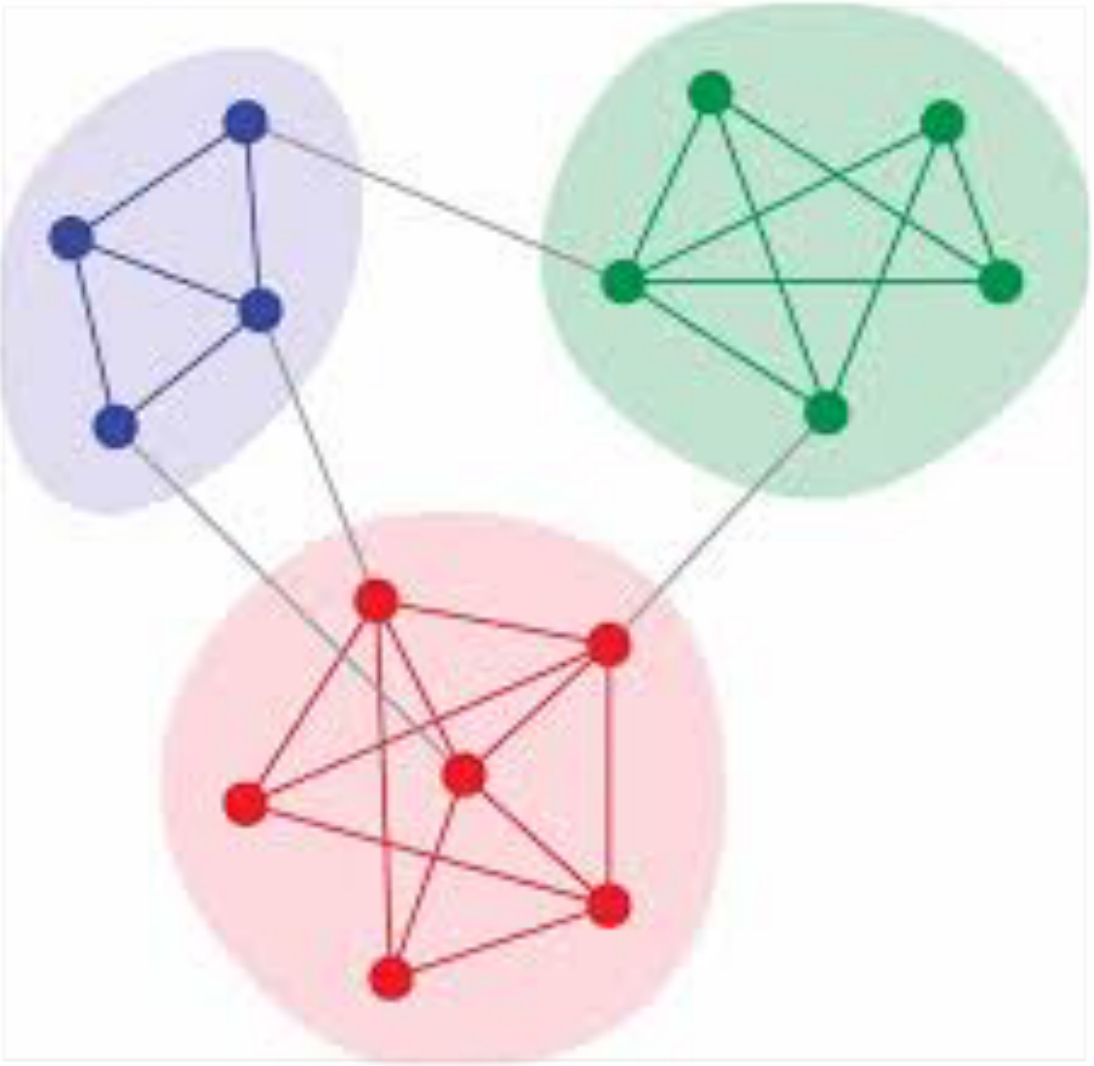
  [von Luxburg, *Statistics and Computating*, 2007]

- Example of (spectral) bisection on an irregular mesh



Exploit the properties of the matrices of graphs

Second: find clusters, cut the graph



Cuts, clustering and communities

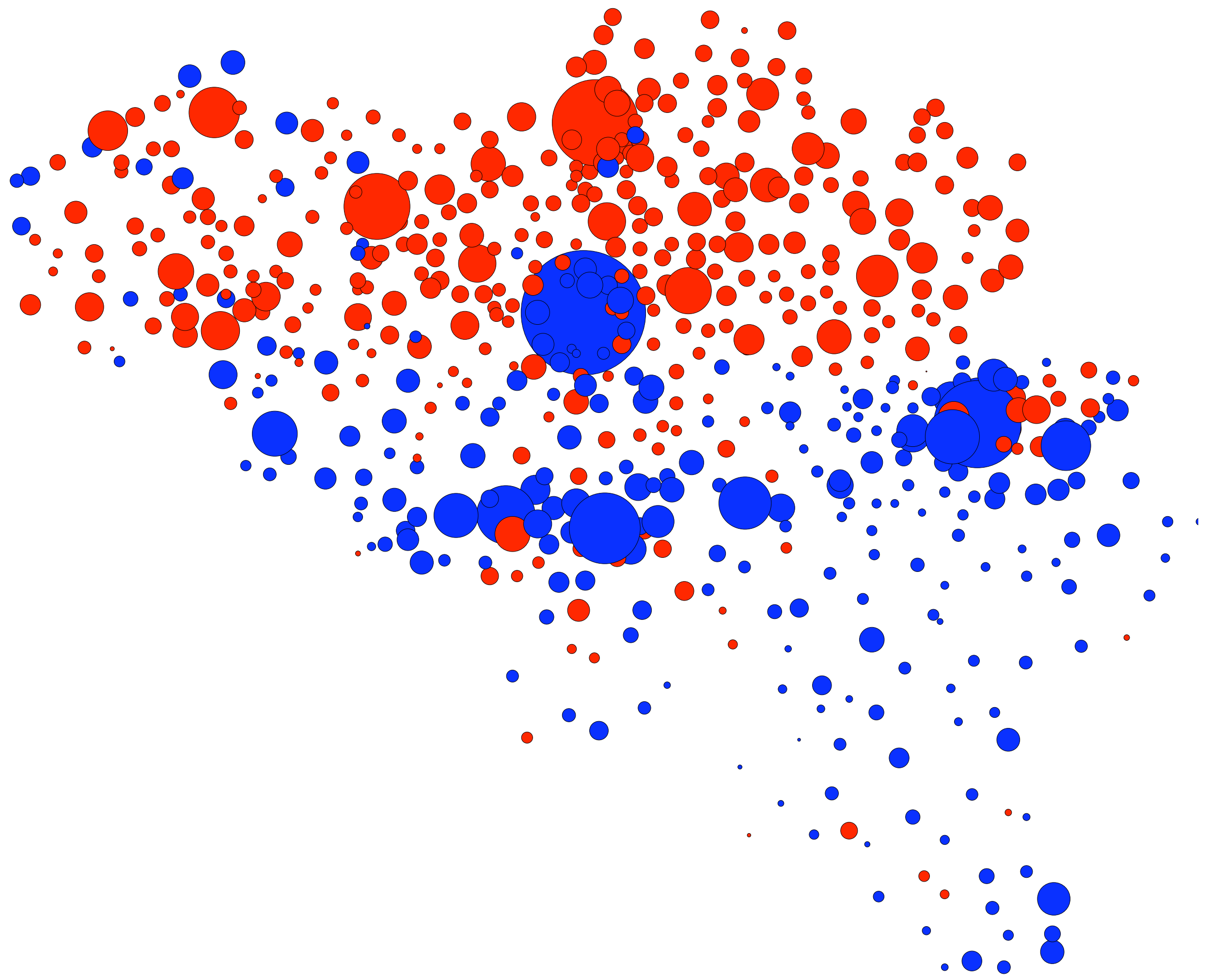
The good, the bad and the ugly

- Networks are often inhomogeneous, with important links, hubs, clusters, or communities (modules)
- These are observed in various types of data on networks: social, technological, biological,...
- Importance of cuts: the min-cut max-flow theorem.
  These are two primal-dual linear programs.
  The max value of a flow = the min capacity over all cuts.

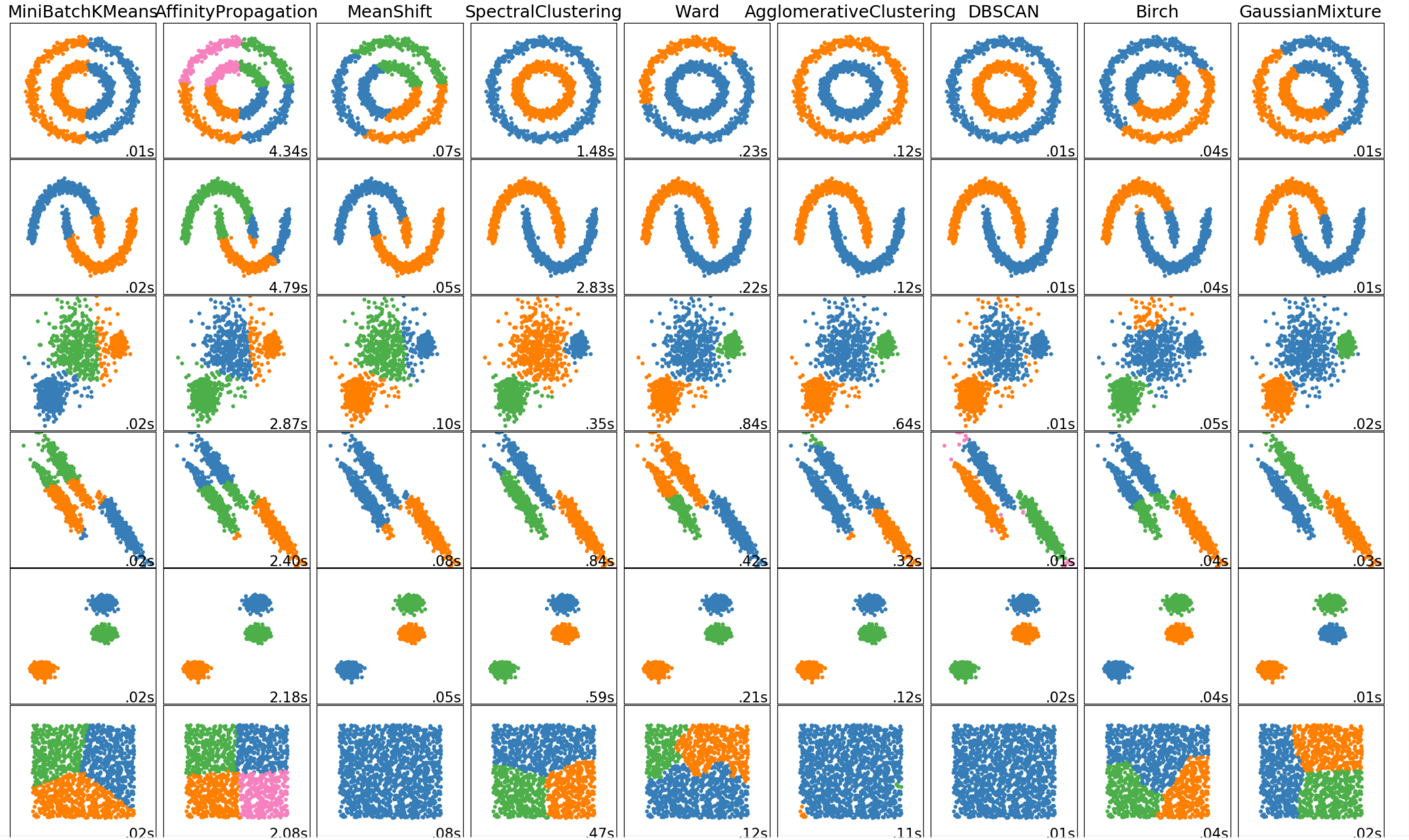
- For clusters and communities, see the extensive surveys:

  [S. Fortunato, *Physic Reports*, 2010]

  [von Luxburg, *Statistics and Computing*, 2007]

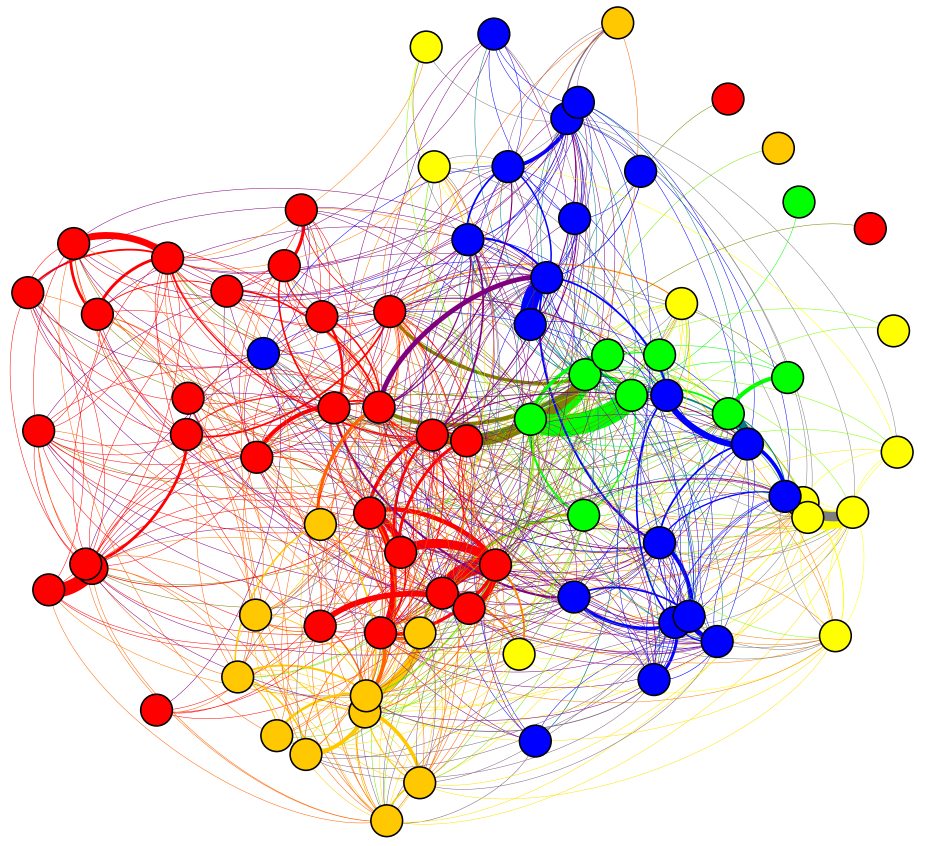


Clustering: a well known topic in data analysis

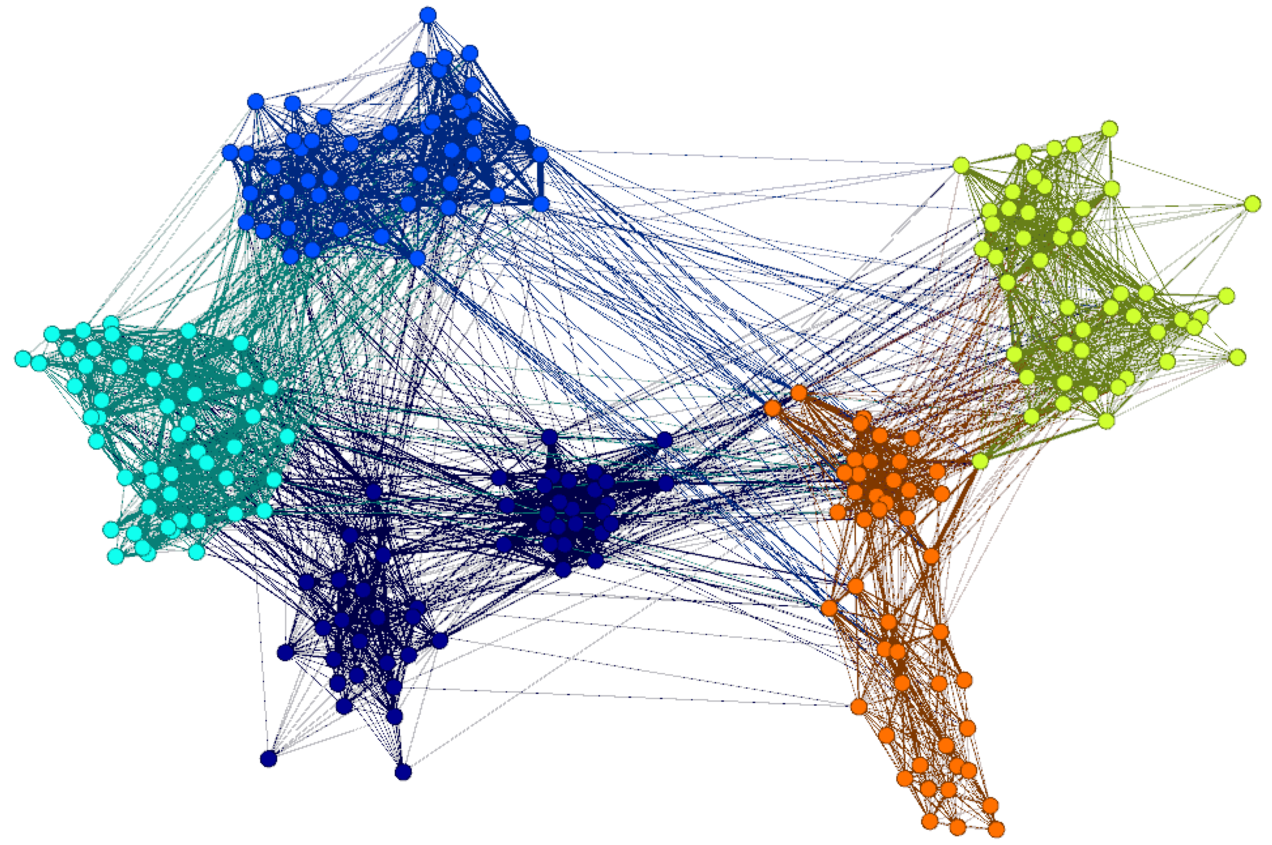


Clustering of graphs : a well known topic in DA

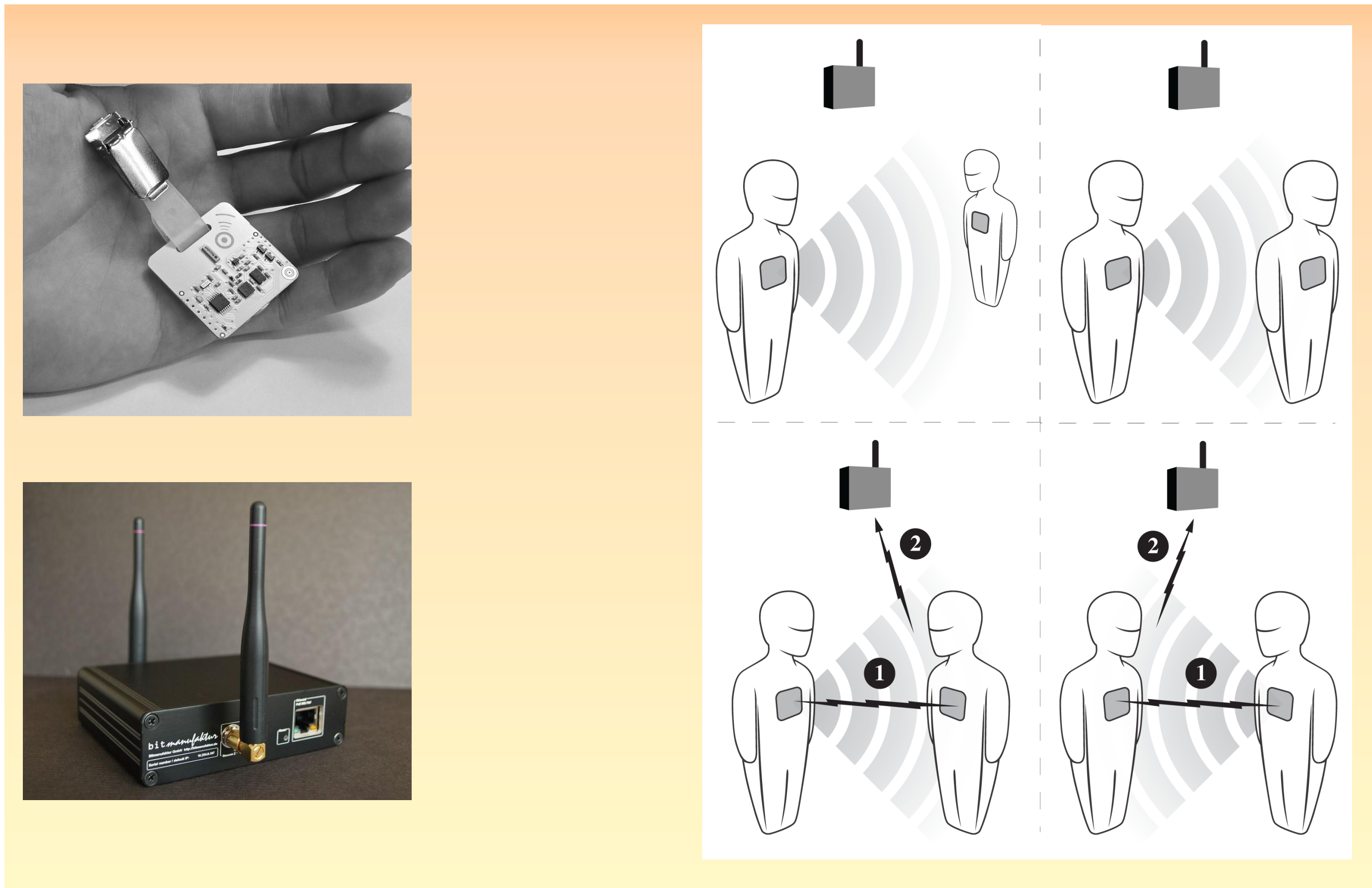
- Social face-to-face interaction networks



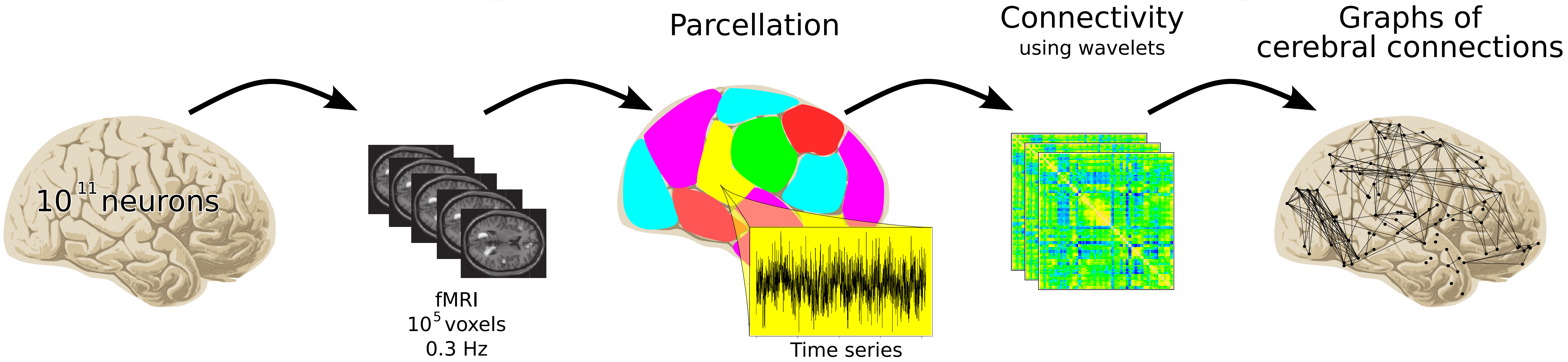
(Lab. physique, ENSL, 2013)



(école primaire, Sociopatterns)



- Brain networks [Bullmore, Achard, 2006]



GRAPHSIP project challenges

Challenge 1: Robustness and hierarchical analysis of brain connectivity

Challenge 2: Brain networks clustering

Challenge 3: Longitudinal study of brain networks

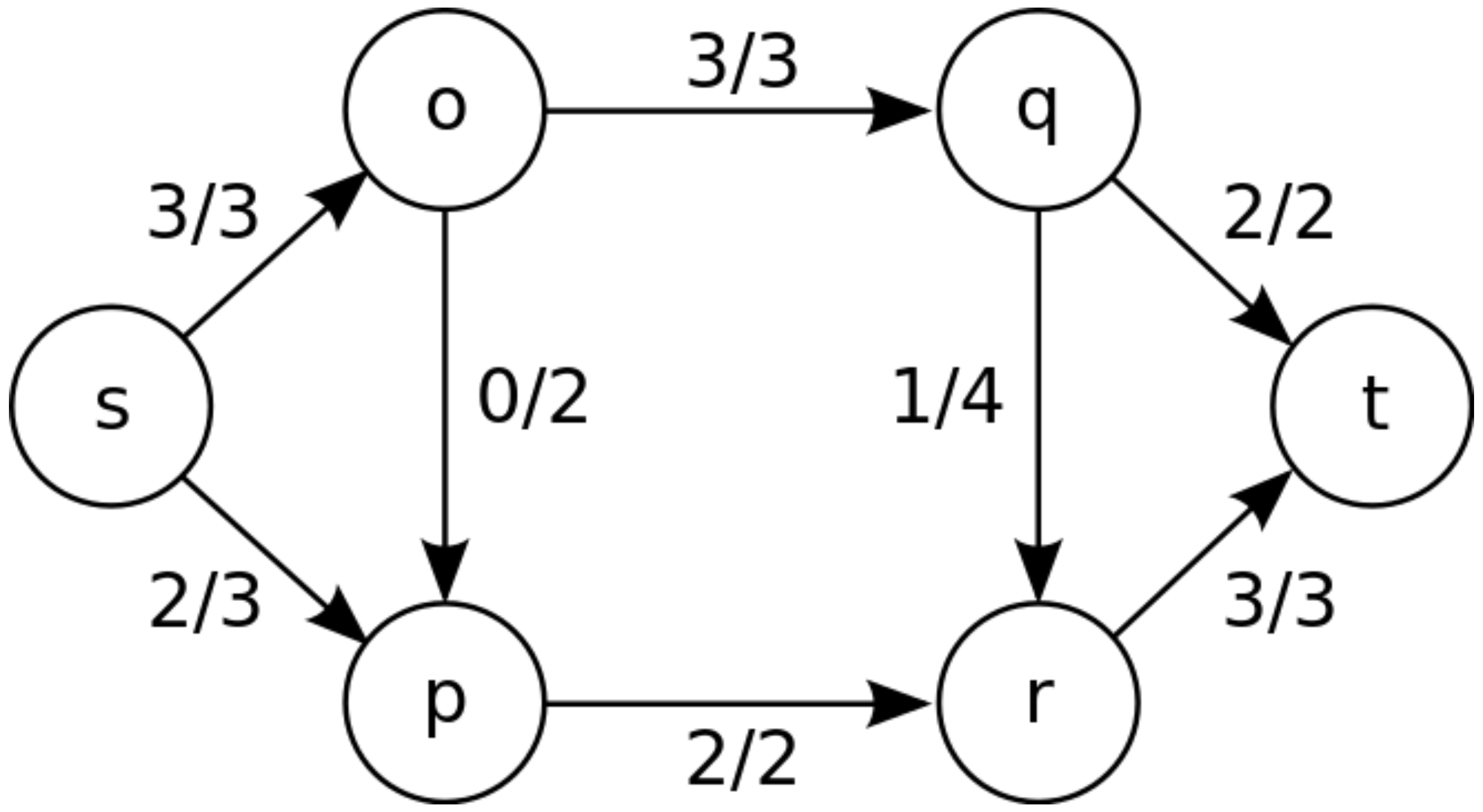
Clustering of graphs : also well known in Graph Theory

Min-Cut and Max-Flow

Graph cuts

- Graph cuts in 2 (or bisection): find the partition in two groups of nodes that minimize the cut size (i.e., the number of links cut)
- Exhaustive search can be computationally challenging
- About the problem of cuts:
  An important result is the min-cut max-flow theorem.

  Min-cut pb and Max-flow pb are two primal-dual problems
  The max value of a flow = the min capacity over all cuts
  One possible solution from linear program



Clustering of graphs : a well known topic in DA

Some theoretical properties: Algebraic Connectivity

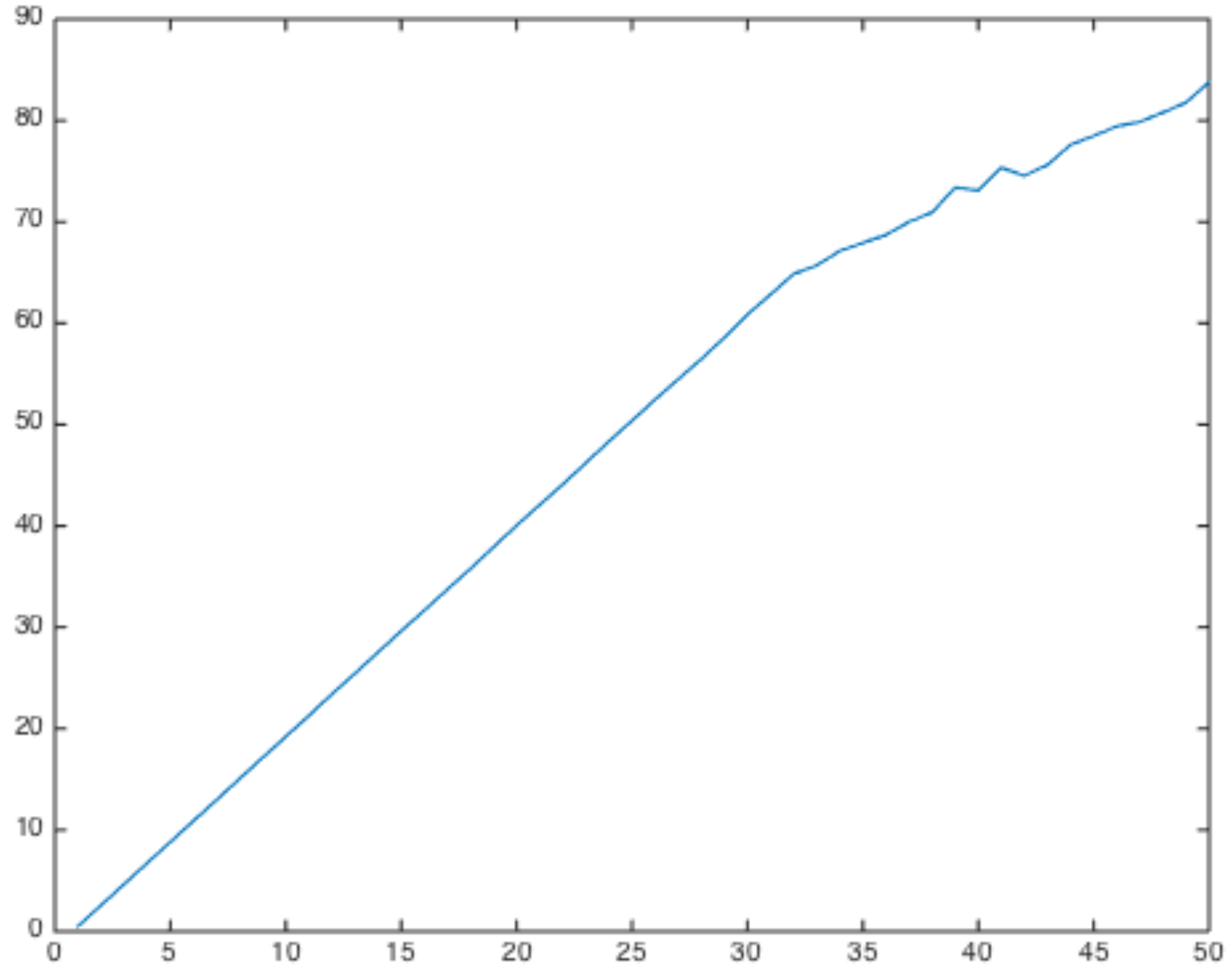
Multiplicity of eigenvalue 0 gives connectedness of graph

What if $\lambda_2 > 0$?

Experiment:

Gradually increase connections

between two Erdos-Renyi subgraphs



$$\lambda_2 \geqslant \frac{1}{\mathrm{vol}(G)d(G)}$$ where $d(G)$ is the diameter of the graph

Clustering of graphs : a well known topic in DA

Some theoretical properties: The Cheeger Constant

Cheeger constant measures presence of a "bottleneck"

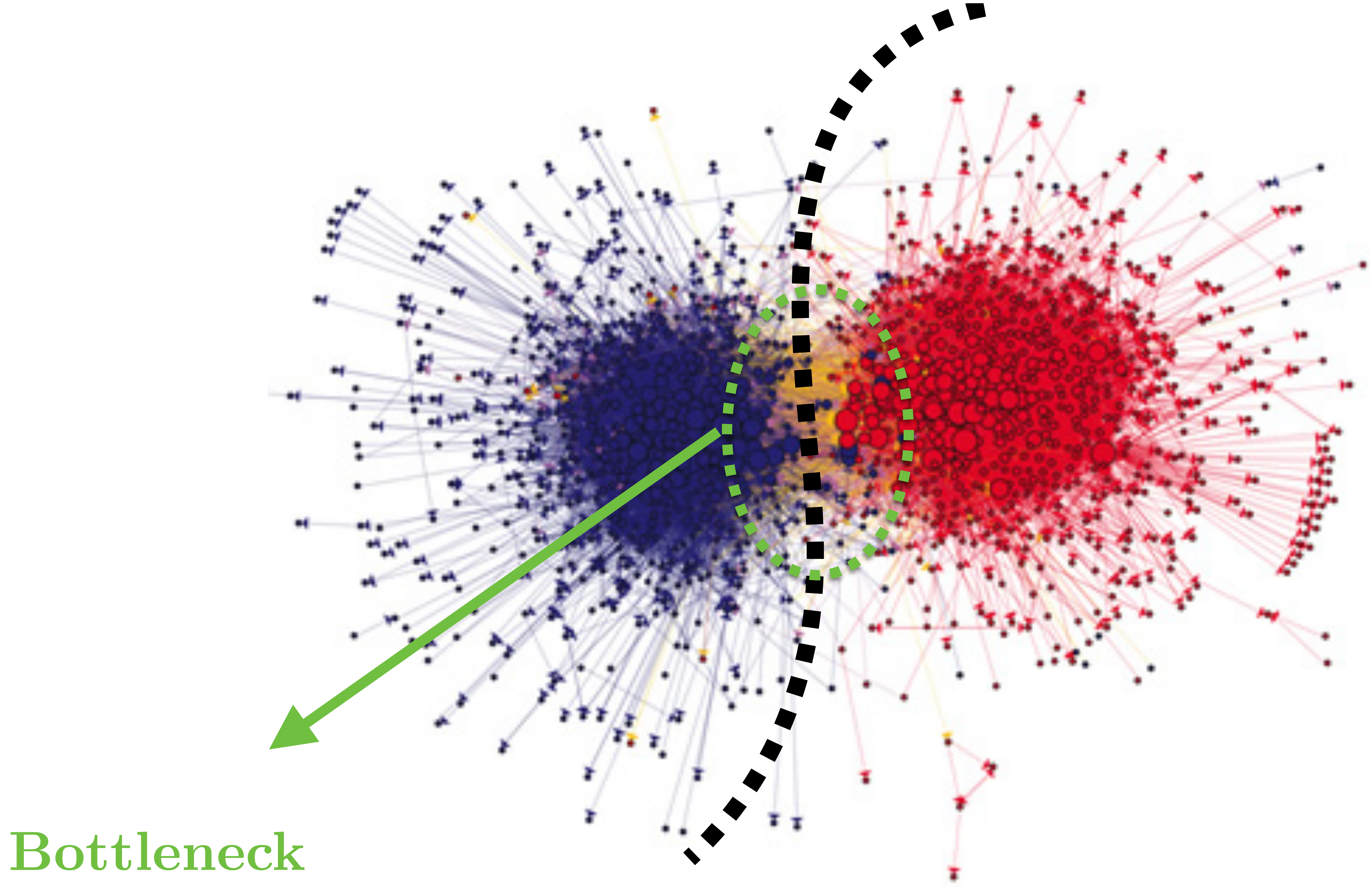
$$A \subset V \qquad \partial A = \Big\{(u,v) \in E \text{ s.t. } u \in A,\, v \in \overline{A}\Big\}$$

$$\mathrm{vol}(A) = \sum_{u \in A} d(u)$$

$$h(G) = \min_{A \subset V} \left\{ \frac{|\partial A|}{\min\Big(\mathrm{vol}(A), \mathrm{vol}(\overline{A})\Big)} \text{ s.t. } 0 < |A| < \frac{1}{2}|V| \right\}$$

Clustering of graphs : a well known topic in DA

Some theoretical properties: The Cheeger Constant



Clustering of graphs : a well known topic in DA

Some theoretical properties: The Cheeger Constant

The Cheeger constant and algebraic connectivity are related by Cheeger inequalities. A simple example:

Theorem: Cheeger Inequality[Polya, Szego]

For a general graph G,

$$2h(G) \geqslant \lambda_2 \geqslant \frac{h^2(G)}{2}$$

<u>Remark:</u> the eigenvector associated to the algebraic connectivity is called the Fiedler vector

Clustering of graphs : a well known topic in DA

Some theoretical properties: The Fiedler Vector

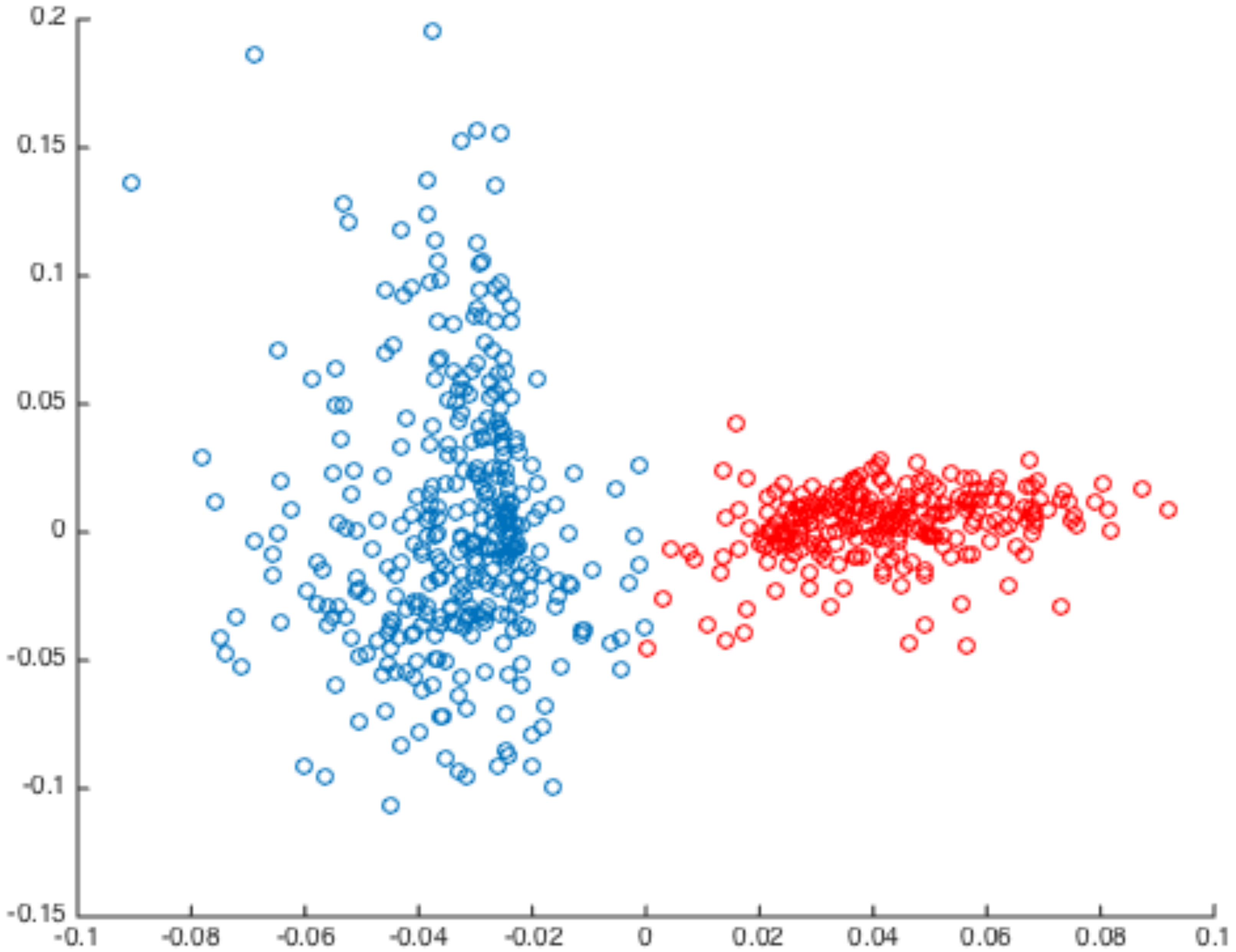
Set of 1490 US political blogs, labelled “Dem” or “Rep”

Hyperlinks among blogs

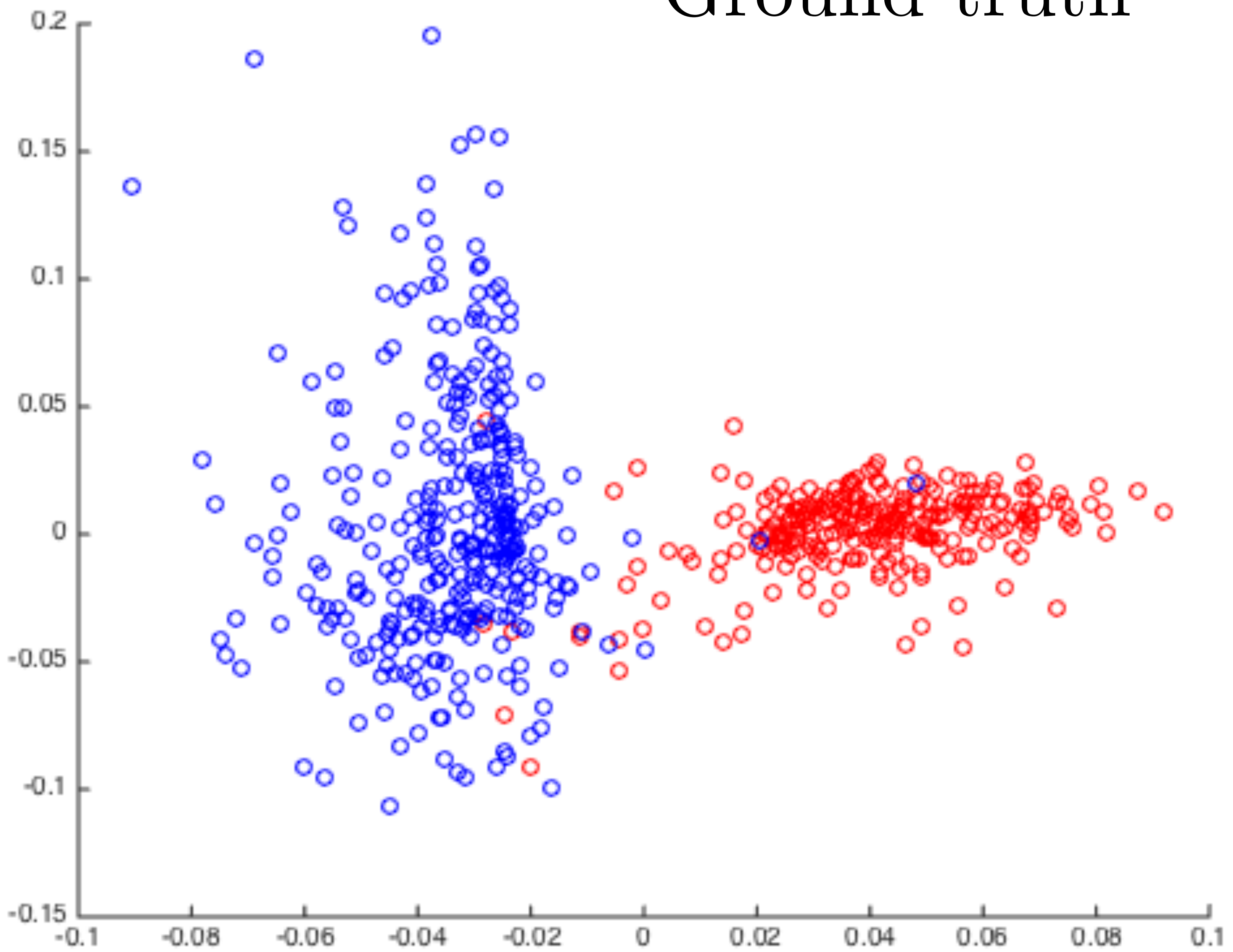
Removed small degrees (<12), keep N = 622 vertices

Compute normalised Laplacian, Fiedler vector

Assign attributes +1, -1 by sign of Fiedler vector



Ground truth

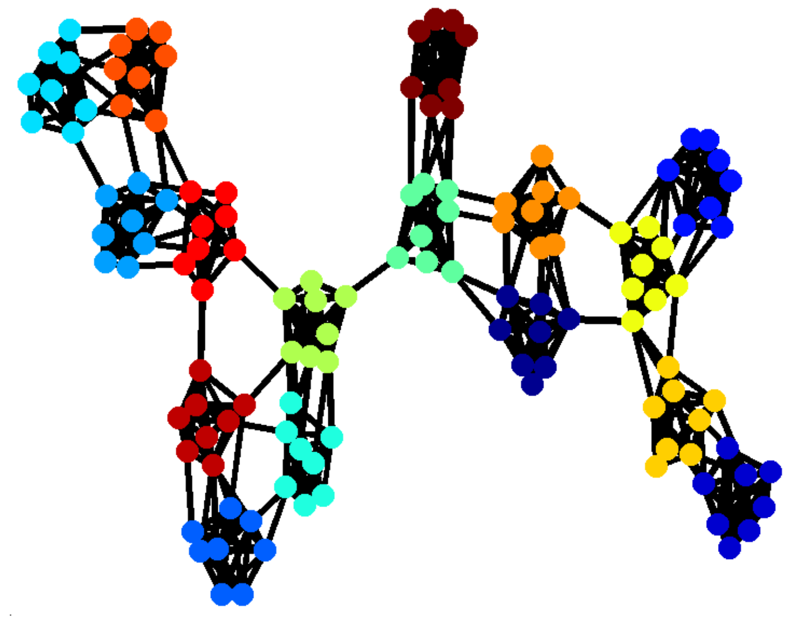


Clustering of graphs or Communities

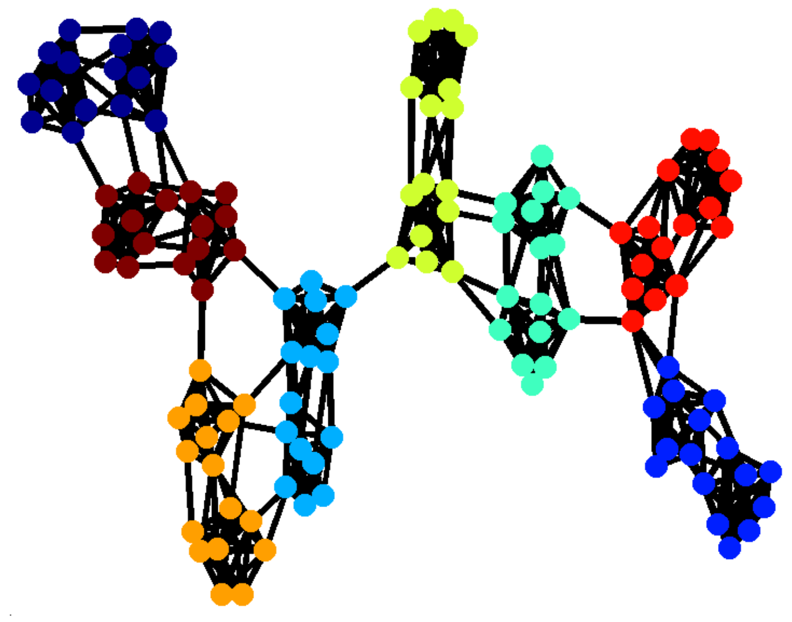
and their relation to the Laplacian eigenvectors

Example: graph with multiscale communities

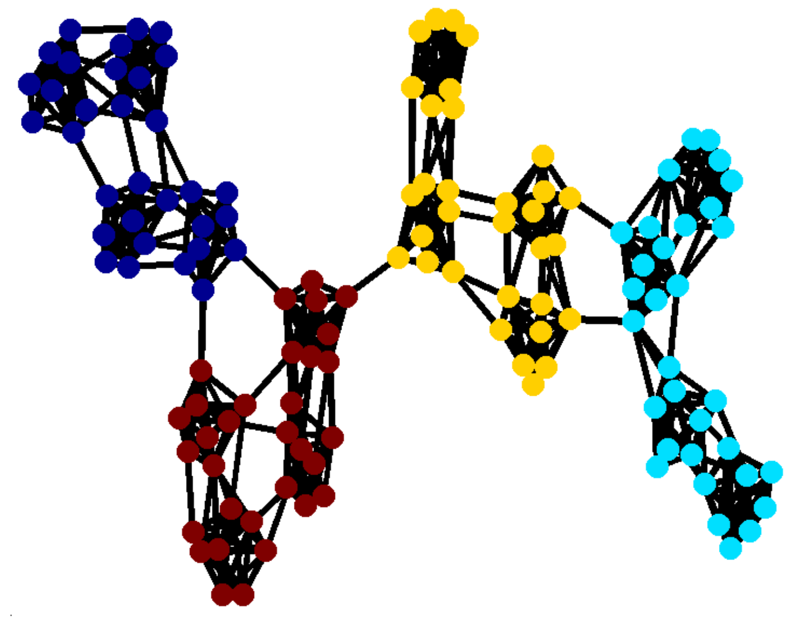
finest scale (16 com.):



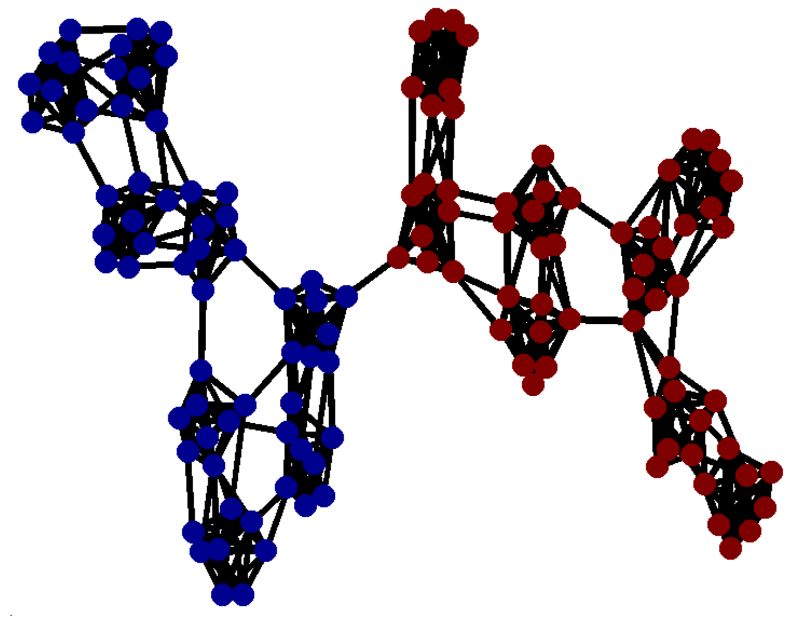
fine scale (8 com.):



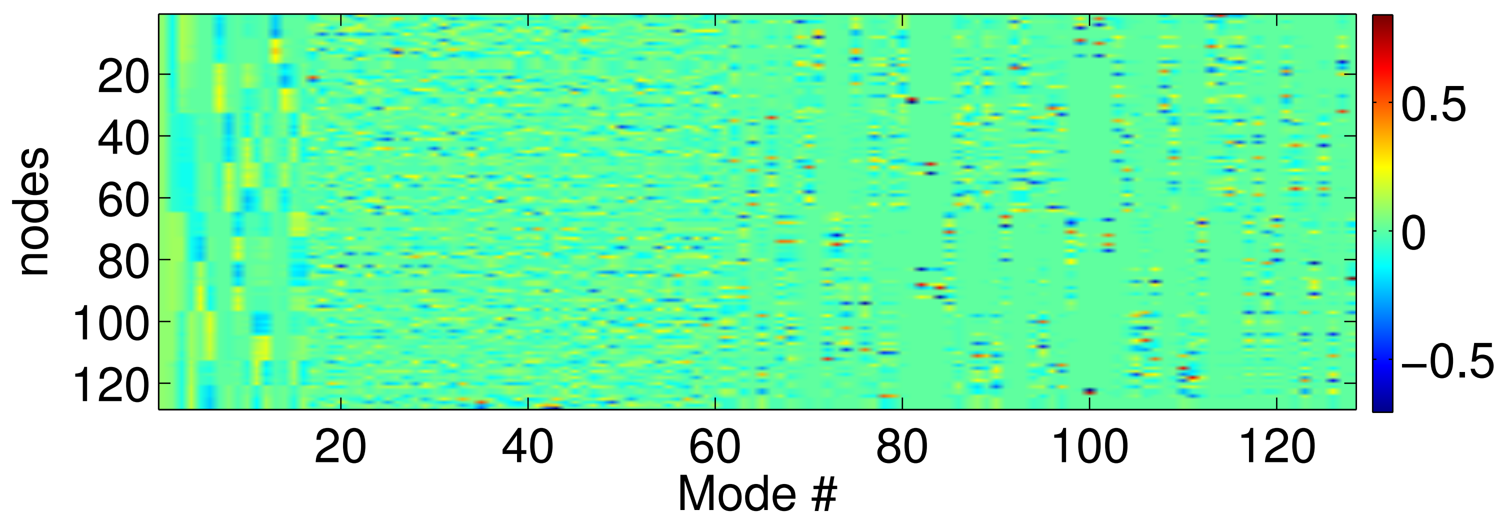
coarser scale (4 com.):

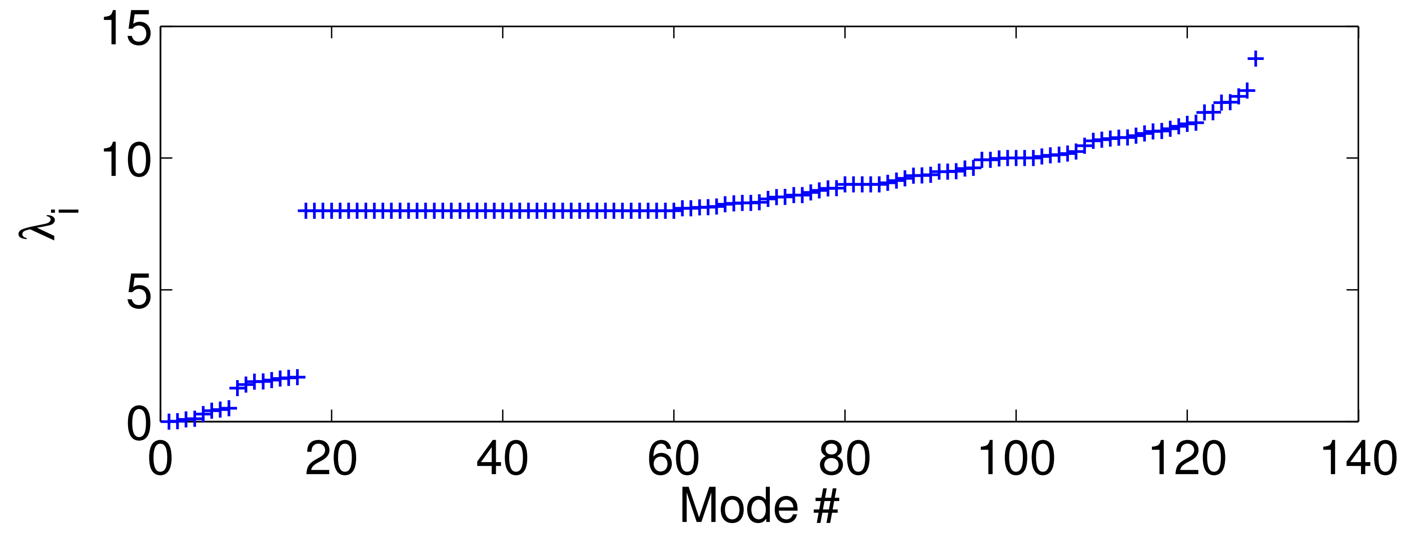


coarsest scale (2 com.):



Spectral analysis: the χ_i and λ_i of this multiscale graph

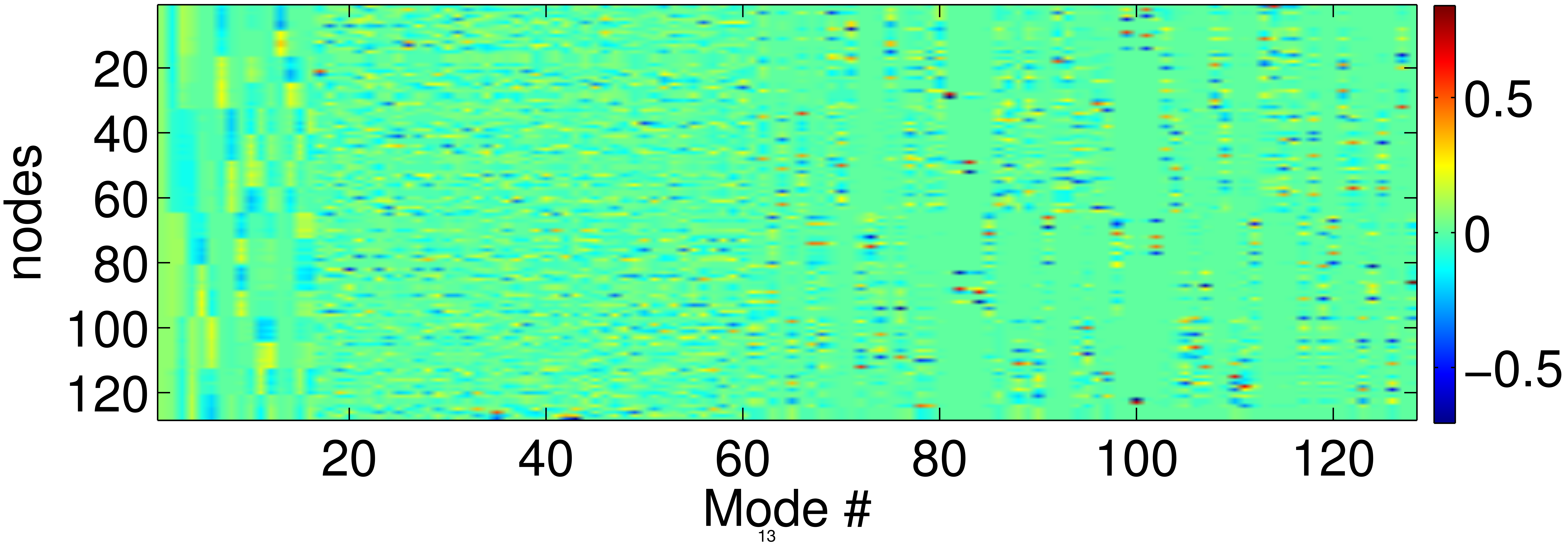




Clustering of graphs or Communities

and their relation to the Laplacian eigenvectors

Spectral analysis: the χ_i and λ_i of this multiscale graph



Clustering of graphs: the spectral approach

U. von Luxburg, “A Tutorial on spectral clustering”, Stat. Comput., 2007

- To cut a graph, one has to measure the size of the cut = # of cut edges
- then associate a cost function inspired by the Cheeger constant

$$C(A,B) := \sum_{i\in A, j\in B} \mathbf{W}[i,j]$$

$$\mathrm{RatioCut}(A,\overline{A}) := \frac{1}{2}\frac{C(A,\overline{A})}{|A|} + \frac{1}{2}\frac{C(A,\overline{A})}{|\overline{A}|}$$

$$\mathrm{NormalizedCut}(A,\overline{A}) = \frac{1}{2}\frac{C(A,\overline{A})}{\mathrm{vol}(A)} + \frac{1}{2}\frac{C(A,\overline{A})}{\mathrm{vol}(\overline{A})}$$

Clustering of graphs: the spectral approach

How to minimise RatioCut ? (a combinatorial problem)

- Re-write the problem with features indicative of each set :
  - $f(i) = +\sqrt{\frac{|\bar{A}|}{|A|}}$ if $i \in A$ and $f(i) = -\sqrt{\frac{|\bar{A}|}{|A|}}$ if $i \in \bar{A}$
  - then $||f||_2 = \sqrt{|V|}$ and $f^\top \mathbf{1} = 0$
- One can compute the RatioCut : $f^\top \mathbf{L} f = |V| \cdot \text{RatioCut}(A, \bar{A})$.
- The problem can be written as:

$$\min_{\mathbf{f}} \ \mathbf{f}^\top L\, \mathbf{f}$$
$$\text{such that } \mathbf{f}^\top \mathbf{1} = 0, \ ||\mathbf{f}||_2 = \sqrt{|V|}$$

  - With $\text{sign}(f)$ an indicator function! <= still combinatorics

Clustering of graphs: the spectral approach

How to minimise RatioCut ? (a relaxed, spectral, problem)

- The exact problem with f indicator function of A is (NP-)hard
- The same problem with any f is a relaxed version: looking for a smooth partition function

$$\arg\min_f f^T \mathbf{L} f \text{ subject to } \|f\| = \sqrt{N}, \quad \langle f, 1 \rangle = 0$$

Solution (G connected): eigenvector of λ_2

Warning: recover partition after thresholding $f = \mathrm{sign}(u_2)$

So we are back to the Fiedler vector !!!

Clustering of graphs: the spectral approach

RatioCut : generalization to k >2

For more than two components, we look for a set of partition functions

$$F \in \mathbb{R}^{N\times k} \qquad F[i,j] = f_j[i] = \begin{cases} 1/\sqrt{|A_j|} & \text{if } v_i \in A_j \\ 0 & \text{otherwise} \end{cases}$$

Observe: $$f_j^T \mathbf{L} f_j = \frac{\mathrm{Cut}(A_j, \overline{A_j})}{|A_j|} \qquad F^T F = \mathbb{I}$$

$$\mathrm{RatioCut}(A_1, \dots, A_k) = \mathrm{Tr}(F^T \mathbf{L} F)$$

Suggests the relaxed problem:

$$\arg\min_{F\in\mathbb{R}^{N\times k}} \mathrm{Tr}(F^T \mathbf{L} F) \text{ such that } F^T F = \mathbb{I}$$

Unnormalized Spectral Clustering

This form of relaxed RatioCut = Unnormalized Spectral Clustering

$$\arg\min_{F \in \mathbb{R}^{N \times k}} \mathrm{Tr}(F^T \mathbf{L} F) \text{ such that } F^T F = \mathbb{I}$$

Algorithm: Unnormalized Spectral Clustering

Compute the matrix F of first k eigenvectors of L

Apply k-means to rows of F to obtain cluster assignments

Normalized Spectral Clustering

consider Normalized Cut, k=2

$$\text{NormalizedCut}(A, \overline{A}) = \frac{1}{2}\frac{C(A, \overline{A})}{\text{vol}(A)} + \frac{1}{2}\frac{C(A, \overline{A})}{\text{vol}(\overline{A})}$$

Then :

$$f[i] = \begin{cases} \sqrt{\text{vol}(\overline{A})/\text{vol}(A)} & \text{if } v_i \in A \\ -\sqrt{\text{vol}(A)/\text{vol}(\overline{A})} & \text{otherwise} \end{cases}$$

$$\text{Check that: } \langle \mathbf{D}f, 1\rangle = 0 \qquad f^T\mathbf{D}f = \text{vol}(G)$$

$$f^T\mathbf{L}f = \text{vol}(V)\text{NormalizedCut}(A, \overline{A})$$

$$\arg\min_f f^T\mathbf{L}f \text{ subject to } f^T\mathbf{D}f = \text{vol}(G), \quad \langle \mathbf{D}f, 1\rangle = 0$$

$$\Updownarrow \quad g = \mathbf{D}^{1/2}f$$

$$\arg\min_g g^T\mathbf{L}_{\text{norm}}g \text{ subject to } \|g\|^2 = \text{vol}(G), \quad \langle g, \mathbf{D}^{1/2}1\rangle = 0$$

Normalized Spectral Clustering

consider Normalized Cut, k>2

Then

$$F[i,j] = f_j[i] = \begin{cases} 1/\sqrt{\mathrm{vol}(A_j)} & \text{if } v_i \in A_j \\ 0 & \text{otherwise} \end{cases}$$

$$f_j^T \mathbf{L} f_j = \frac{\mathrm{Cut}(A_j, \overline{A_j})}{\mathrm{vol}(A_j)} \qquad F^T F = \mathbb{I} \qquad f_j^T \mathbf{D} f_j = 1$$

$$\arg \min_{H \in \mathbb{R}^{N \times k}} \mathrm{Tr}(H^T \mathbf{L}_{\mathrm{norm}} H) \text{ such that } H^T H = \mathbb{I} \qquad H = \mathbf{D}^{1/2} F$$

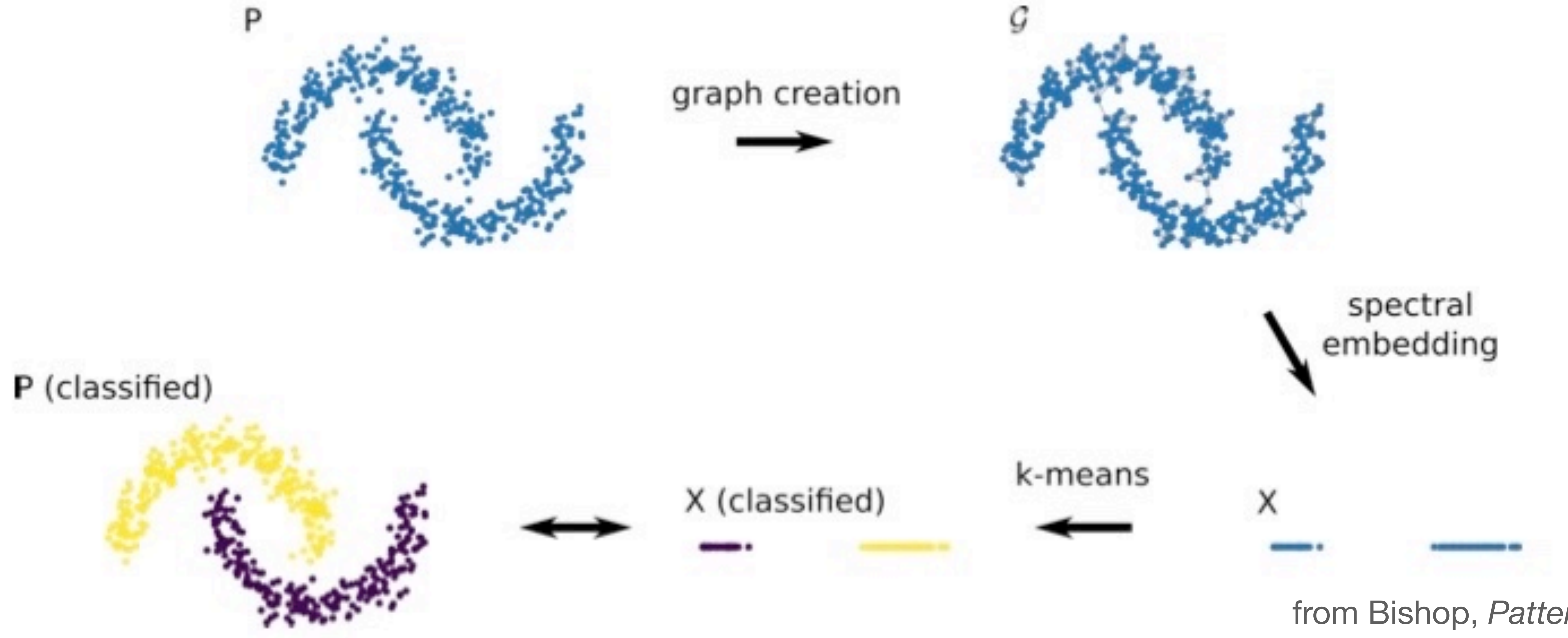
Algorithm: Normalized Spectral Clustering

Compute the matrix H of first k eigenvectors of $\mathbf{L}_{\mathrm{norm}}$

Apply k-means to rows of H to obtain cluster assignments

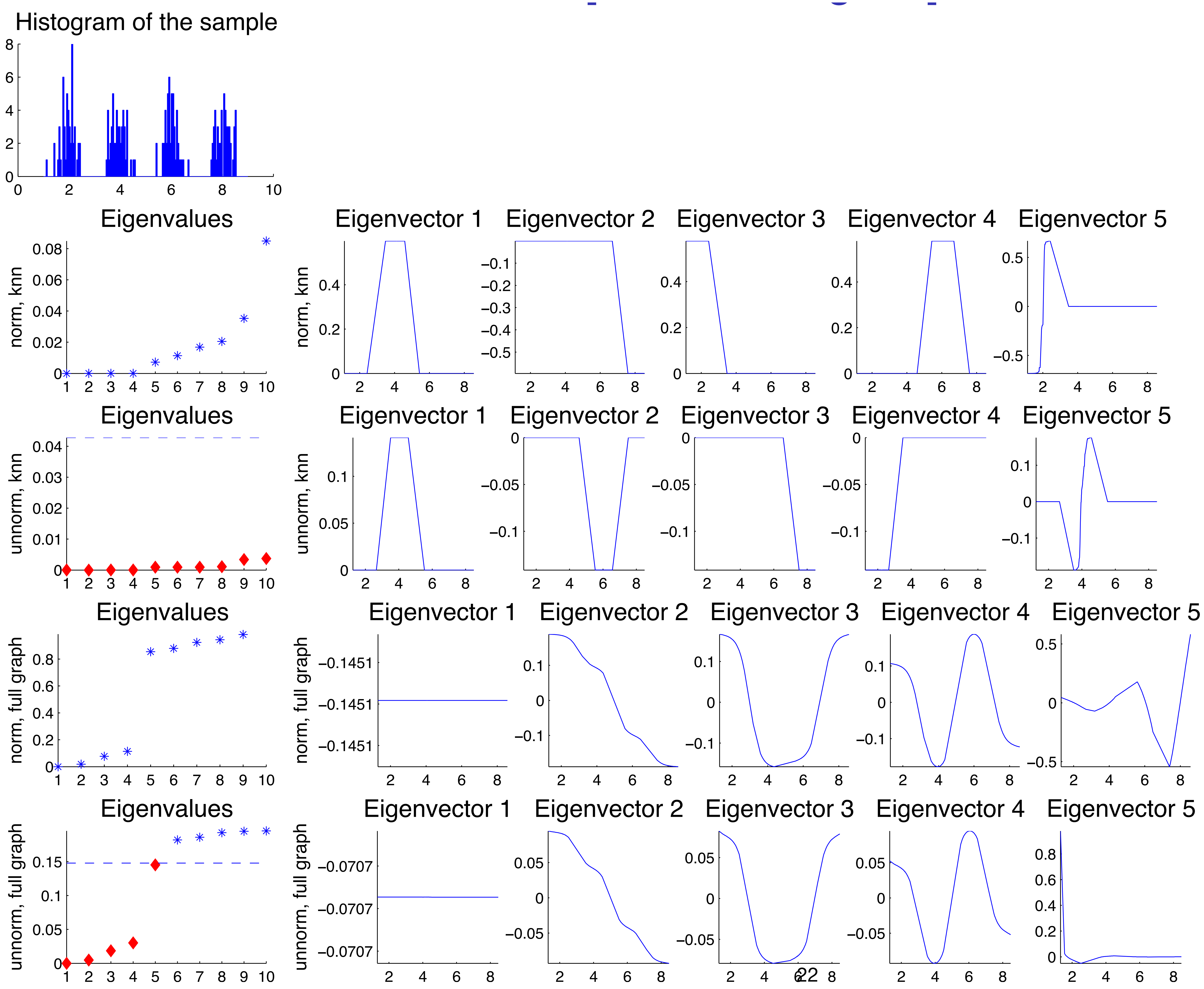
Spectral Clustering in a nutshell

1. **Graph construction.** A sparse similarity graph is built between the n points.
2. **Spectral embedding.** The first k eigenvectors of a graph representative matrix (such as the Laplacian) are computed.
3. **Clustering.** k-means is performed on these spectral features, to obtain k clusters.



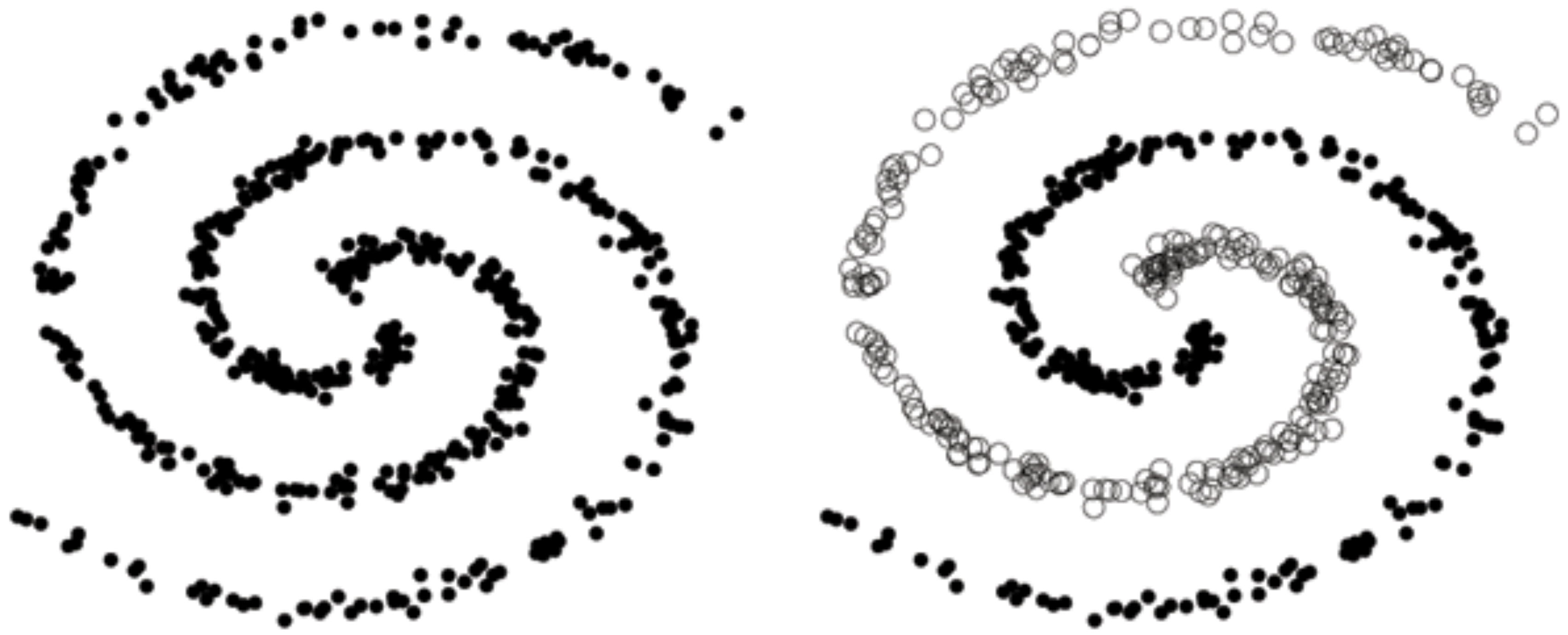
from Bishop, *Pattern Recognition and Machine Learning*, 2006 (chapt. 6)

Spectral Clustering in action



von Luxburg, 2007

Spectral Clustering in action



In practice normalised spectral clustering is often preferred

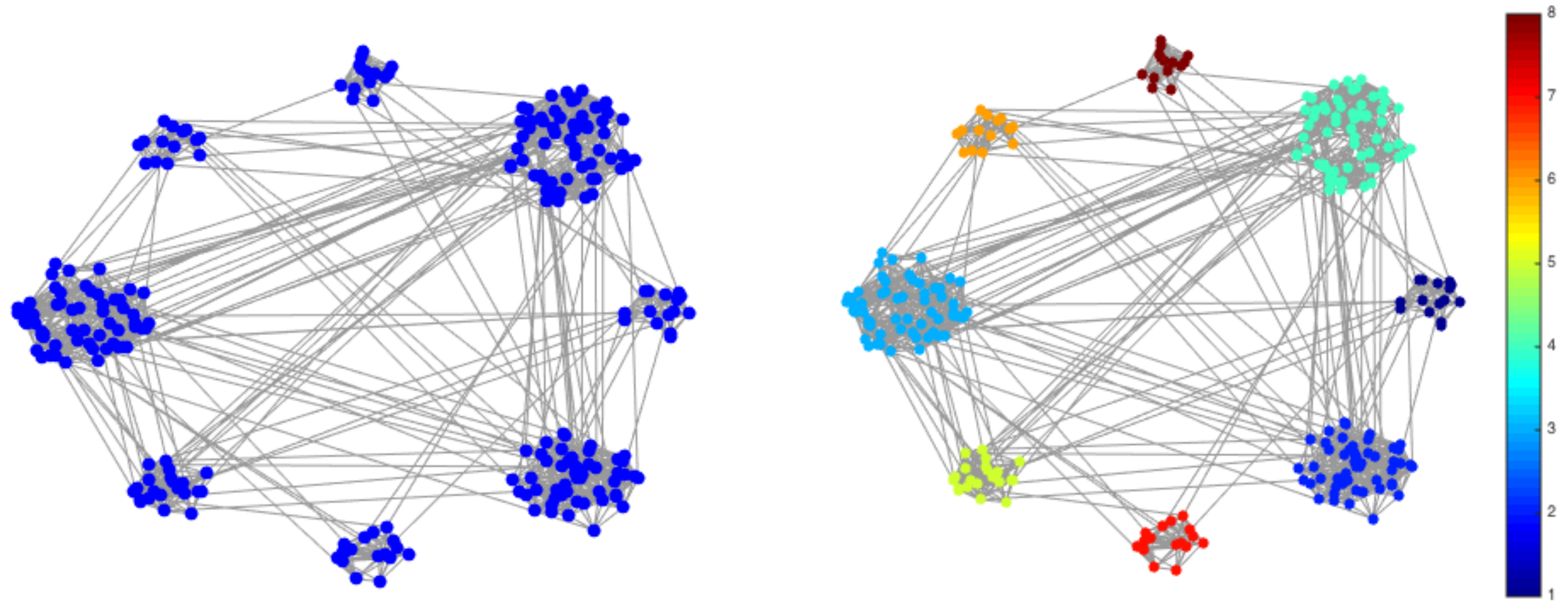
In practice the eigenvectors are "re-normalized" by the degrees $F = \mathbf{D}^{-1/2} H$

before k-means, because these are real cluster assignments

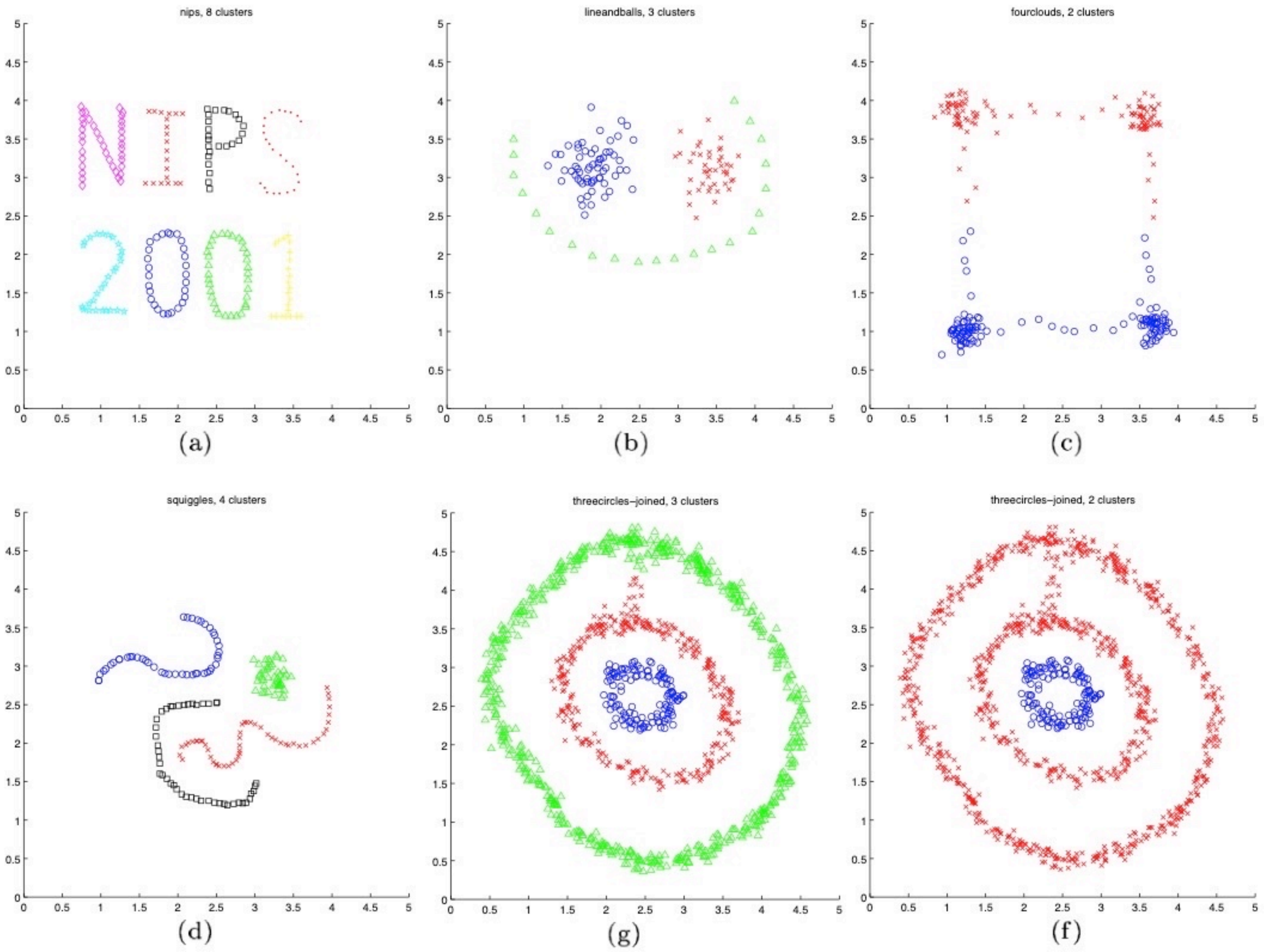
Rem: this is equivalent to using the "random walk Laplacian"

$$\mathbf{L}_{\text{rw}} = \mathbf{D}^{-1}\mathbf{L}$$

If data has k clear clusters, there will be a gap in the Laplacian spectrum after the k-th eigenvalue. Use to choose k.



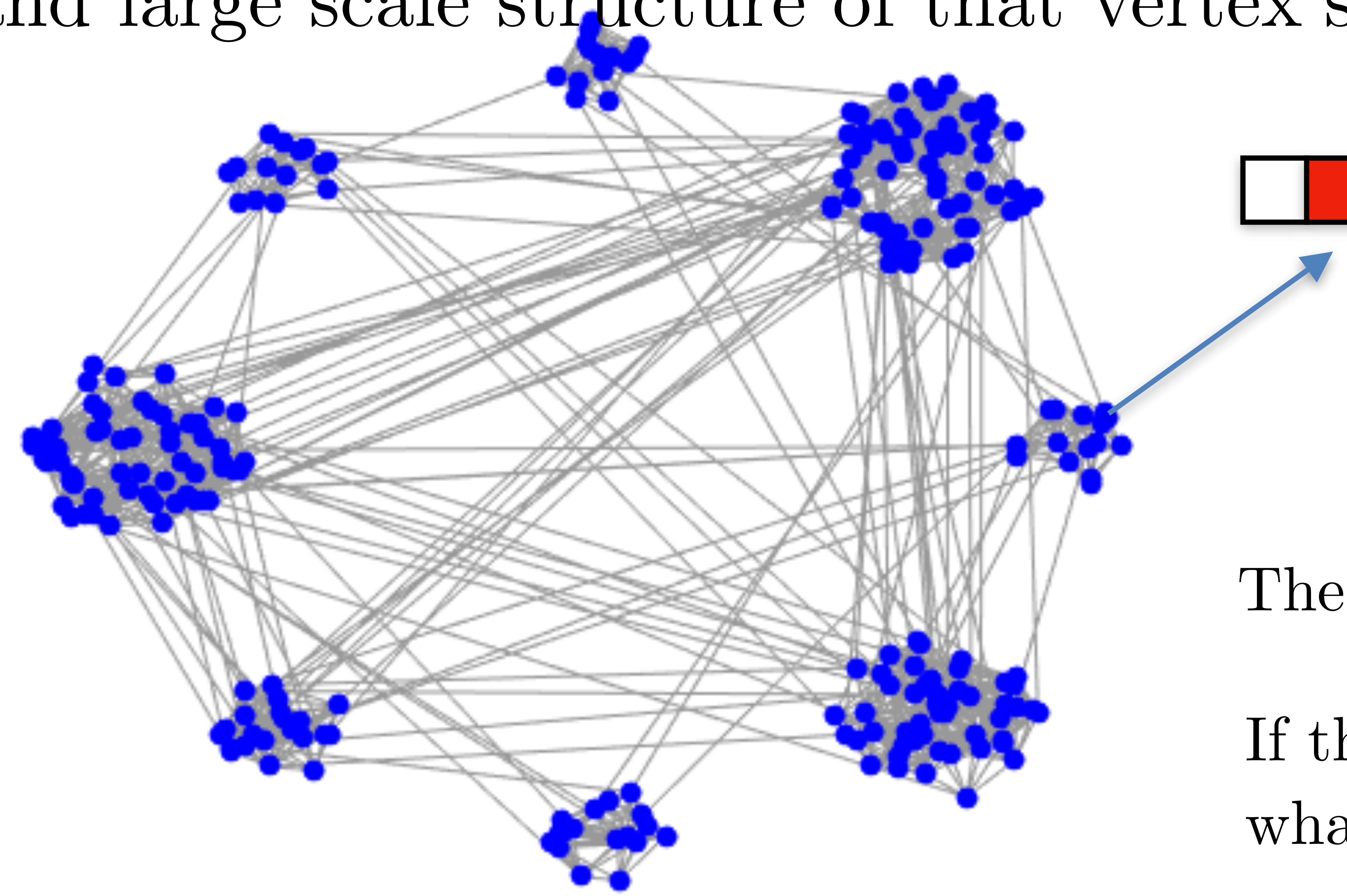
Spectral Clustering in action

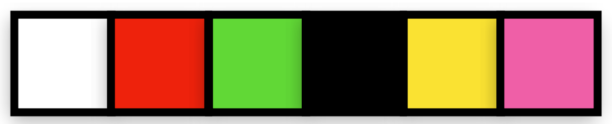


Ng, Jordan, Weiss, NIPS'01

Spectral Clustering interpretation 1

At each vertex the algorithm associates a feature vector that represents the fine and large scale structure of that vertex's neighbourhood in the graph





vertex $[i]$, feature $f_i \in \mathbb{R}^k$

$$f_i[\ell] = u_\ell[i], \quad \ell = 1, \ldots, k$$

These are the lines of the matrix F

If the graph has k disconnected components, what does F look like ?

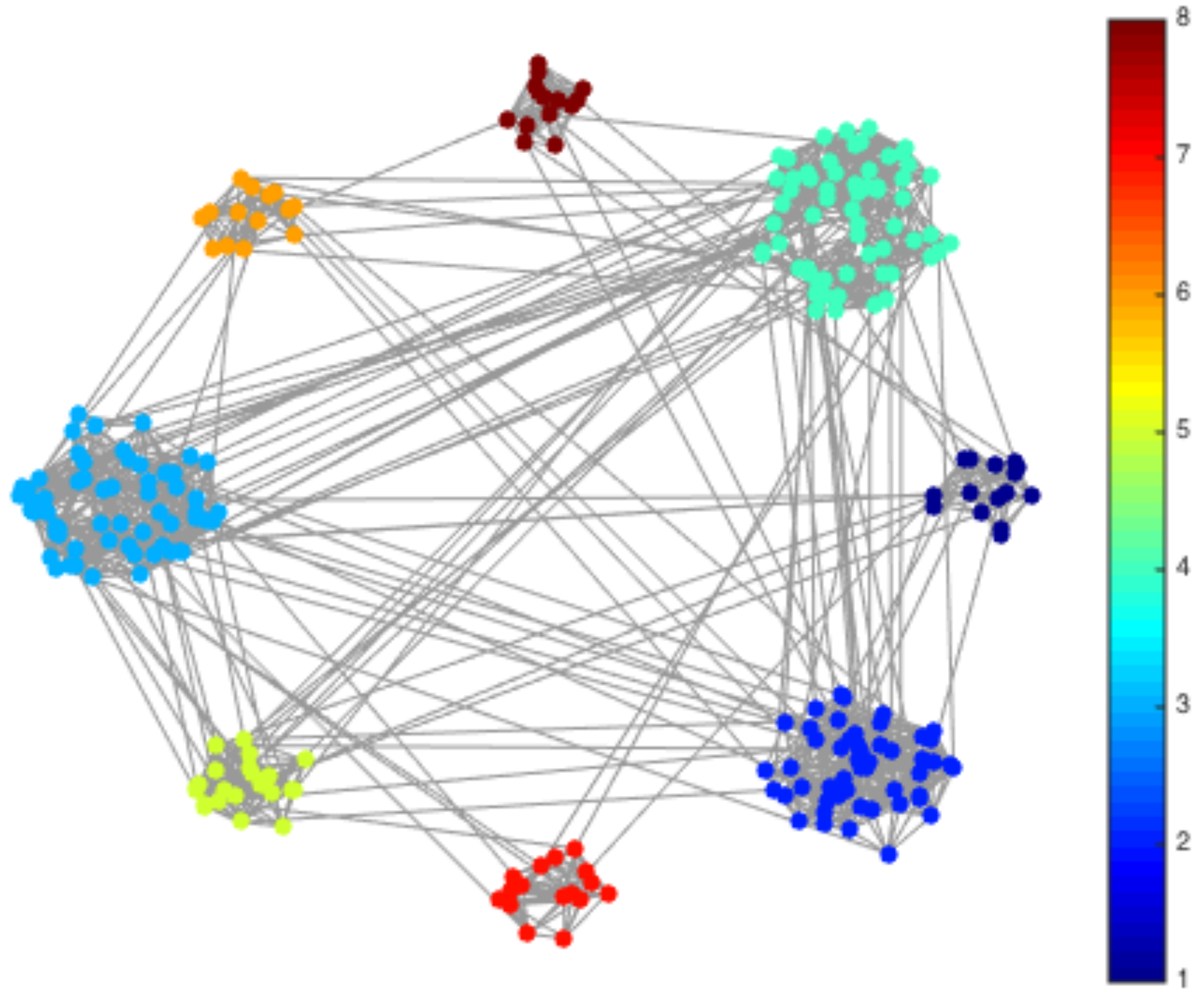
k-means is then applied to these vectors to cluster into k clusters

In short, we transform the graph into a feature matrix and partition it.

Spectral Clustering interpretation 2

We are looking for k "partition signals (functions)"

$$f_\ell : V \mapsto \mathbb{R}$$



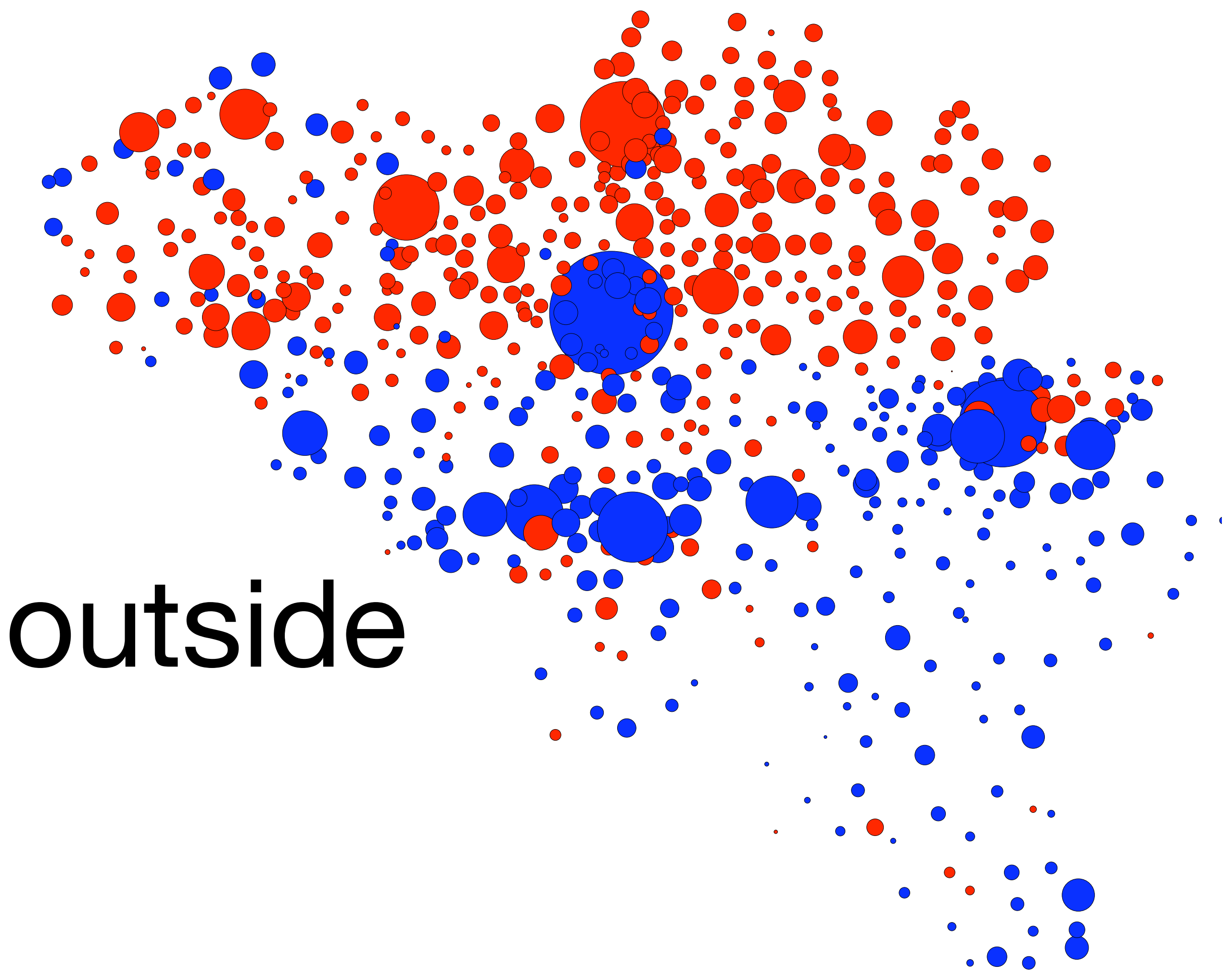
In the ideal case (k disconnected components)

$$f_\ell[i] = \begin{cases} c_i & \text{if } i \in \text{cluster } \ell \\ 0 & \text{otherwise} \end{cases}$$

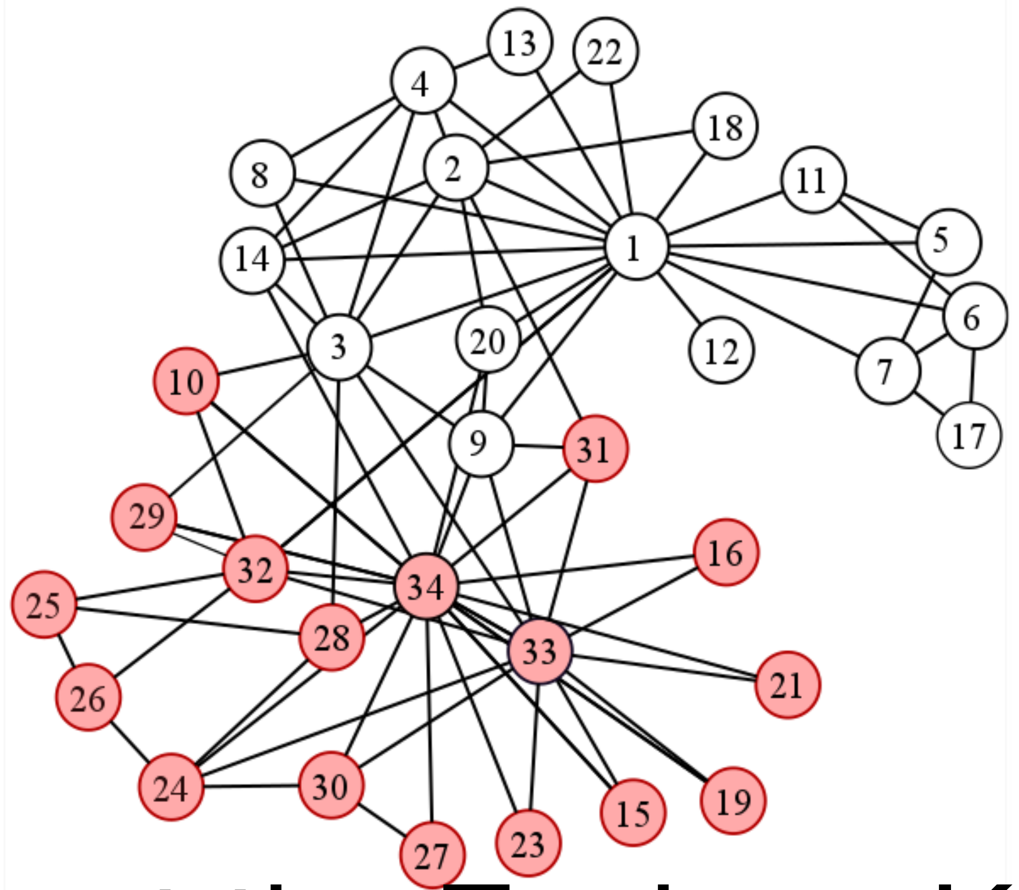
These are maximally smooth graph signals: $f_\ell^T \mathbf{L} f_\ell = 0$

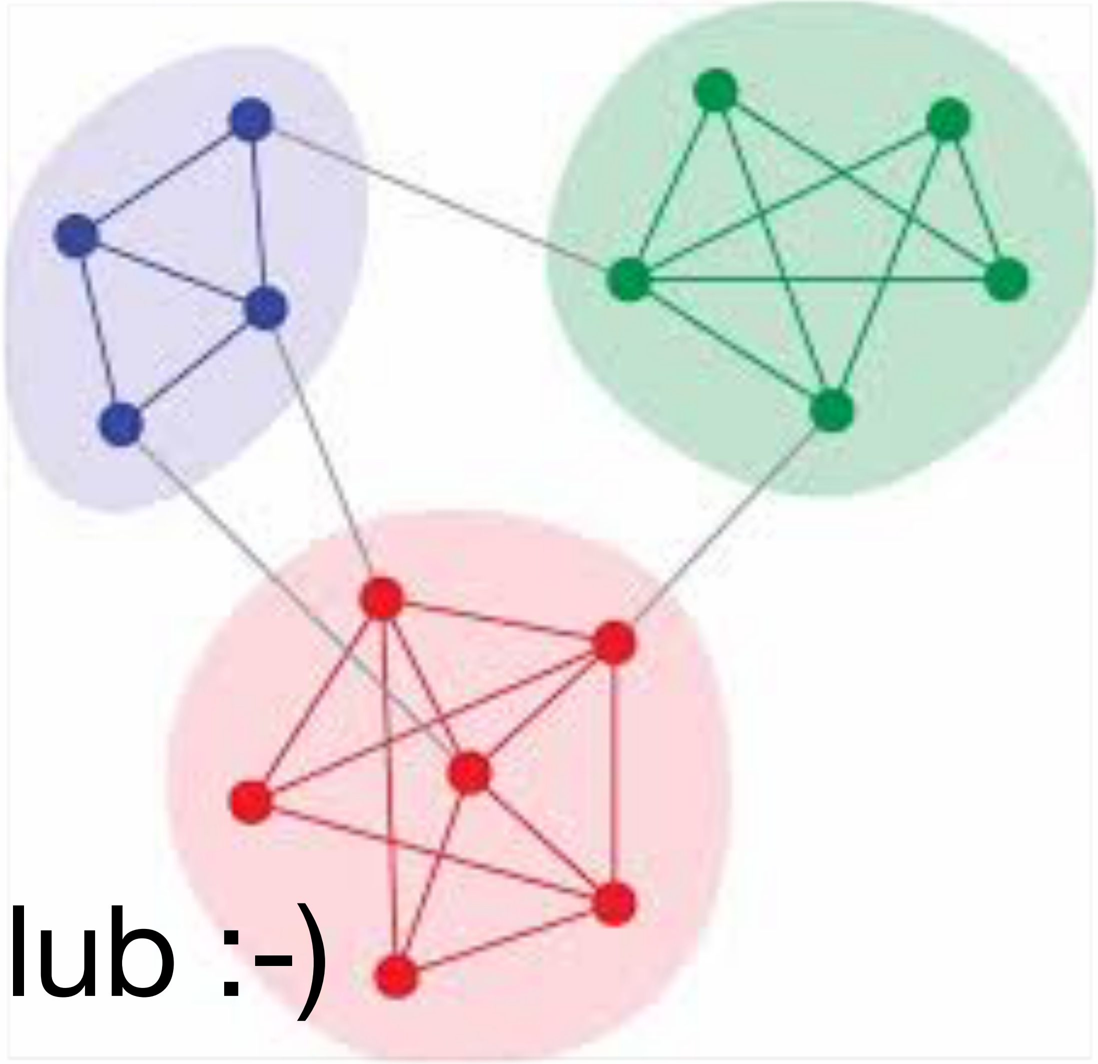
Exploit the properties of the matrices of graphs

Third: find communities in complex networks



- Communities are more loosely defined:
  - usually = nodes more connected together than with the outside
- Many methods (thousands of papers):
  - Modularity [Girvan, Newman, 2004]
  - Infomap [Rosvall, Bergstrom, 2009]
  - Stochastic Block Models





- Win a prize at NetSci: be 1st to talk about the Zachary Karate Club :-)

Summary up to now:

Exploit the spectral analysis of graphs

- Graph spectral analysis: eigenvectors and eigenvalues of matrices A, L, L_{rw},…
- They have interesting properties, especially for L: new basis of representation, analog for Fourier oscillations, study of smoothness on graphs,…
- Centralities: can be studied with these matrices
- Clusters: can be found with them as well (in a relaxed way)

- More processing methods to come!