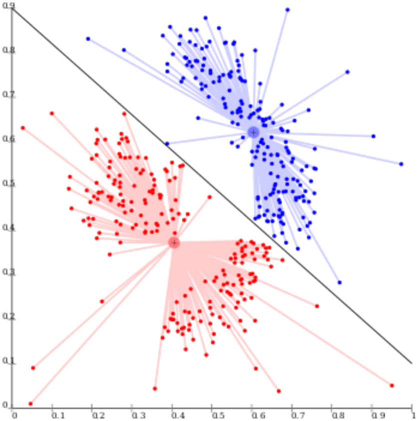
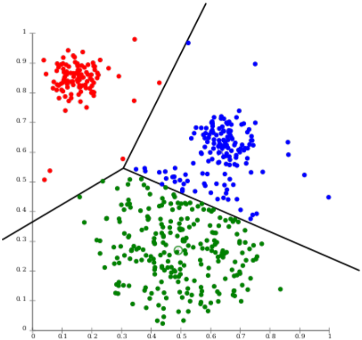
What are...clustering methods?

Or: Partitioning made easy

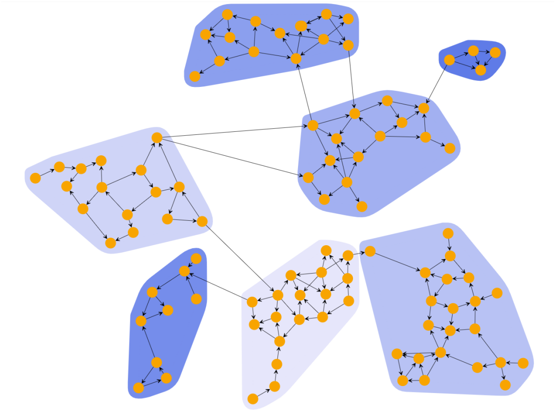
Analyzing data





- Given a large data set, one often wants to cluster it
- This means one wants to minimize the distance from (to be found) k centers
- There are good algorithms to do this in $\mathbb{R}^m$ (not optimal, but good)

Clustering in graphs



- If the data is given as a graph, one can compute eigenvectors
- Let $u_1, \ldots, u_m$ be the eigenvectors for the smallest Laplace eigenvalues
- Assign vertex x to $(u_i(x))_i$ and run an $(\mathbb{R}^m, k)$ cluster algorithm

Back to cutting

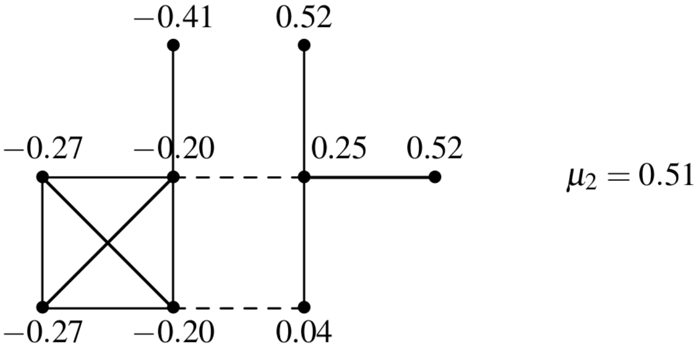


Fig. 3.2 Graph with 2nd Laplace eigenvector

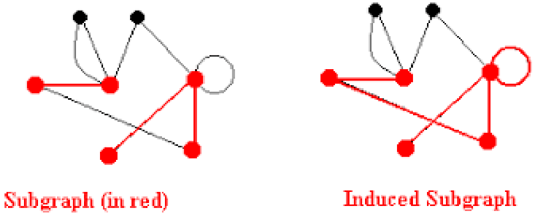
- Special case The cheapest cut of the graph into two relatively large pieces
- Use $x \mapsto u(x)$ for the eigenvector for the second smallest Laplace eigenvalue
- Use clustering for $(\mathbb{R} = \mathbb{R}^1, k = 2)$

For completeness: A formal statement

For $G = (V, E)$ suppose $W \subset V$ is such that the induced subgraph on $V \setminus W$ is disconnected, then:

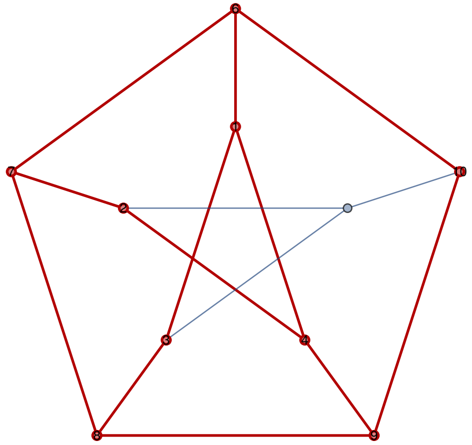
$$|W| \geq \mu_{-2}$$

- μ_{-2} = second smallest Laplace eigenvalue also called algebraic connectivity
- Induced subgraph = take all edges that make sense, *e.g.*:



- For the clustering algorithm I can only offer "it works well" – so take the above as a justification why one could expect it to work well

Algebraic connectivity



- The Petersen graph has $\mu_{-2} = 2$
- Removing an arbitrary vertex will not disconnect the graph
- The lower bound 2 is however not optimal (but we should not expect that anyway)

Thank you for your attention!

I hope that was of some help.