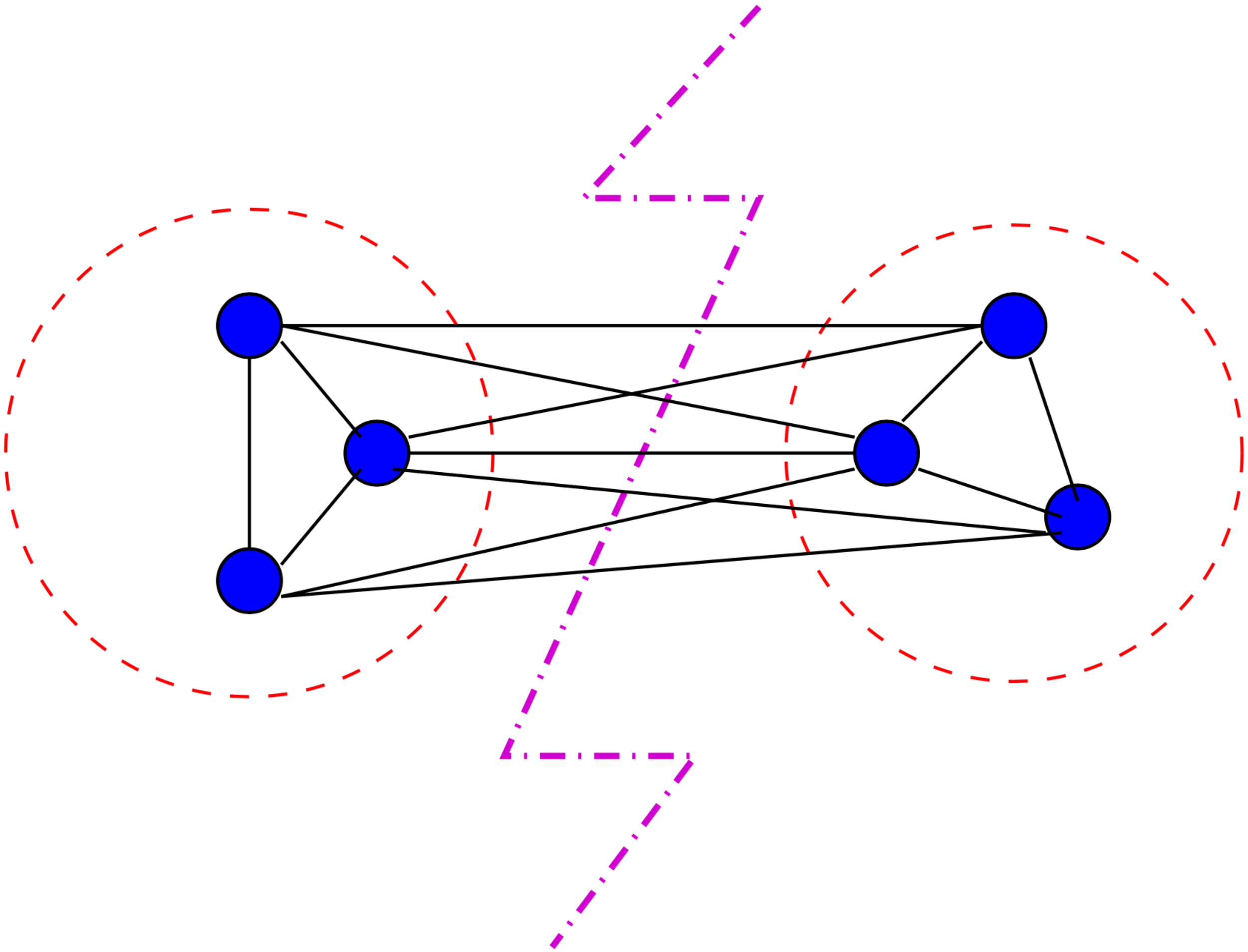
Spectral Clustering

What is Spectral Clustering?

- Easy to implement
- Reasonably fast especially for sparse data
- Spectral clustering treats the data clustering as a graph partitioning problem without making any assumption on the form of the data clusters using the spectrum of the Laplacian matrix

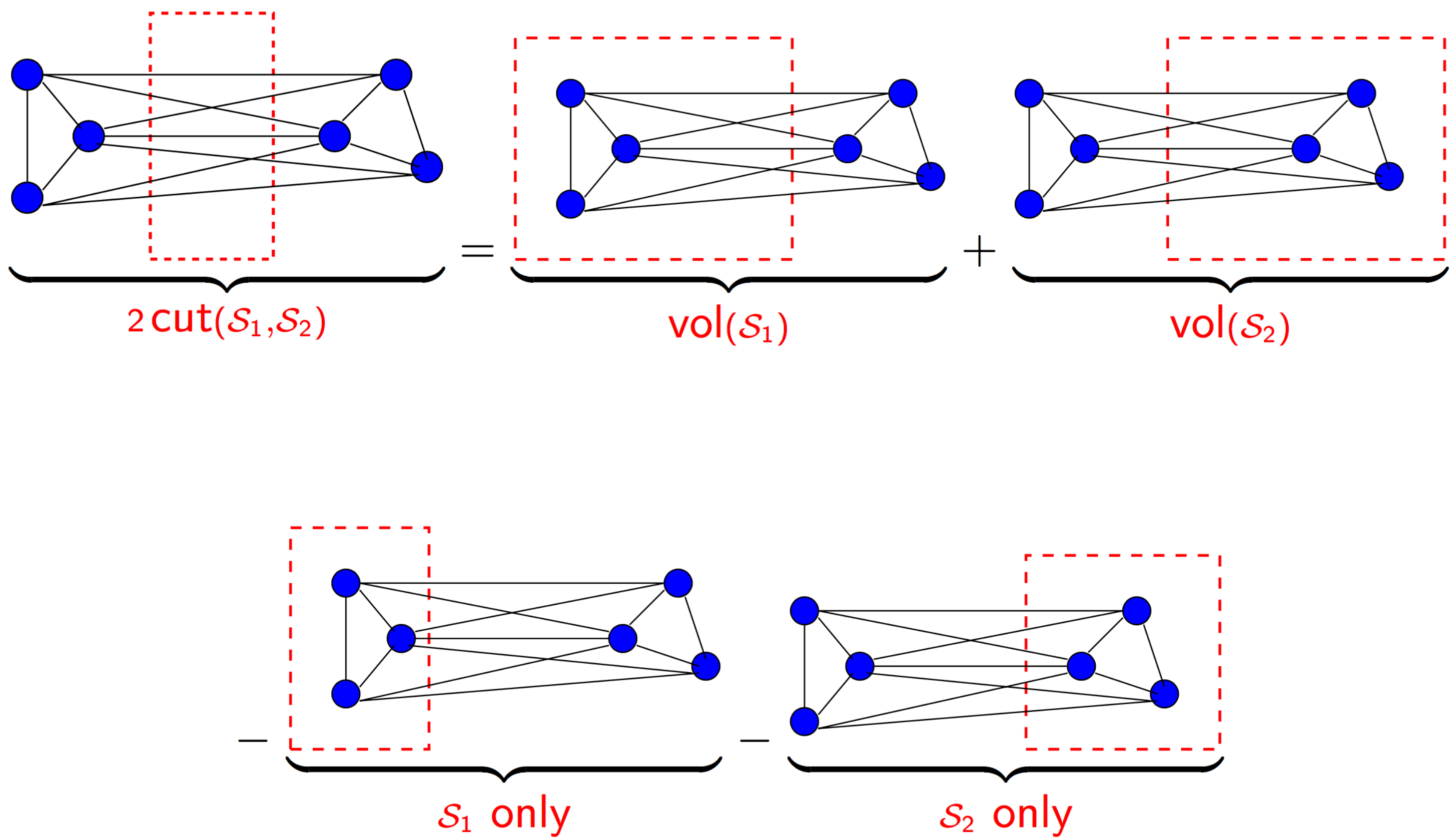
http://www.cmi.ac.in/~madhavan/courses/datamining12/reading/em-tutorial.pdf

Graph Bipartitioning



Graph Bipartitioning

- Affinity matrix $W = [W_{ij}] \in \mathbb{R}^{n\times n}$
- Degree of nodes $d_i = \sum_j W_{ij}$
- Volume $\mathrm{vol}(\mathcal{S}_1) = d_{\mathcal{S}_1} = \sum_{i\in\mathcal{S}_1} d_i$



Graph Bipartitioning

$$
\begin{aligned}
x^\top L x &= x^\top D x - x^\top W x \\
&= \sum_{i=1}^{n} d_i x_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} x_i x_j \\
&= \frac{1}{2} \left(\sum_i d_i x_i^2 - 2 \sum_i \sum_j W_{ij} x_i x_j + \sum_j d_j x_j^2 \right) \\
&= \frac{1}{2} \sum_i \sum_j W_{ij} (x_i - x_j)^2 .
\end{aligned}
$$

Algorithm

① Construct affinity matrix

$$W_{ij} = \begin{cases} \exp\{-\beta \|v_i - v_j\|^2\} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

② Calculate graph Laplacian $L = D - W$ where $D = \mathrm{diag}\{d_1, \ldots, d_n\}$ and $d_i = \sum_j W_{ij}$ is the degree of node v_i

③ Compute second smallest eigenvector of graph Laplacian (denoted by $u = [u_1 \ \cdots \ u_n]^\top$, Fiedler vector)

④ Partition u_i's by pre-specified threshold value and assign data point to cluster v_i