

Spectral Clustering

● ● ● | Families of Clustering Algorithms

- Partition-based methods
 - e.g., K-means
- Hierarchical clustering
 - e.g., hierarchical agglomerative clustering
- Probabilistic model-based clustering
 - e.g., mixture models, Gaussian Mixture Models
 - expectation maximization
- Spectral Clustering

✓

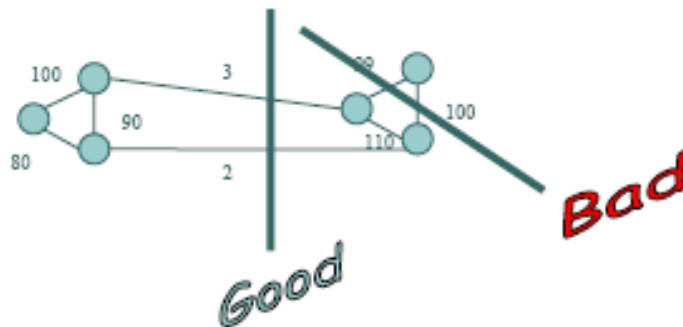
✓

... Apr 23

NOW!

● ● ● | Graph-Theoretic Clustering

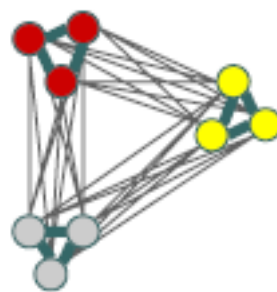
- Weighted graph
- Edge weights correspond to similarity
- Cut edges in the graph to form a good set of connected components—ideally the within component edges in the graph have large weights and the across component edges have small weights



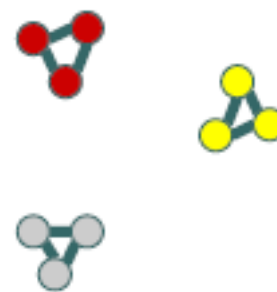
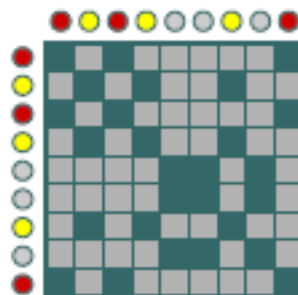
● ● ● | Spectral Clustering Overview



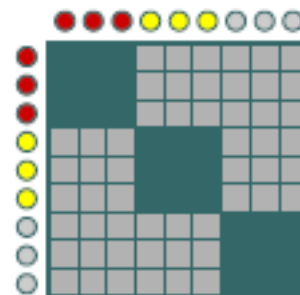
Data



Similarities



Block-Detection



● ● ● | aside: Linear Algebra Refresher

- **Eigen decomposition/matrix diagonalization** of a square $k \times k$ matrix A into eigenvalues and eigenvectors. Suppose A has nondegenerate eigenvalues $\lambda_1 \lambda_2 \dots \lambda_k$ and corresponding linearly independent eigenvectors $X_1 X_2 \dots X_k$. Let $D = \text{diag}(\lambda_1 \lambda_2 \dots \lambda_k)$ and $P = [X_1 X_2 \dots X_k]$ then
$$A = P D P^{-1}$$

- **Singular Value Decomposition (SVD)**: decomposition of $m \times n$ matrix A
$$A = U D V^T$$
where U and V are orthogonal $m \times m$ and $n \times n$ matrices and D is a diagonal matrix whose diagonal entries σ_i are called the singular values.

U =unitary matrix

D =singular values, non-negatives on the diagonal

V =orthonormal basis vector directions

● ● ● | Spectral Algorithm

- Algorithm:

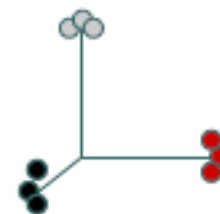
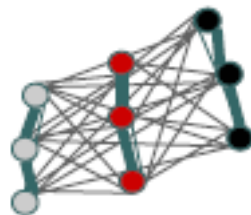
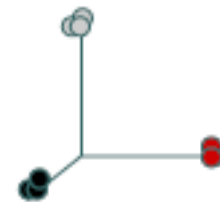
Find the top k right singular vectors v_1, \dots, v_k
Let C be the matrix whose j th column is given by $A v_j$
Place row i in cluster j if C_{ij} is largest entry in the i th row of C

- Interpretation:

Suppose the rows of A are points in a high-dimensional space. Then the subspace defined by the top k right singular vectors of A is the rank- k subspace that best approximates A . The spectral algorithm projects all the points onto this subspace. Each singular vector then defines a cluster; to obtain a clustering we map each project point to the cluster of the singular vector that is closest to it in angle.

● ● ● | The Spectral Advantage

- The key advantage of spectral clustering is the spectral space representation:



● ● ● | Spectral Clustering Example

Data



3 Sets of
News-
group
Postings

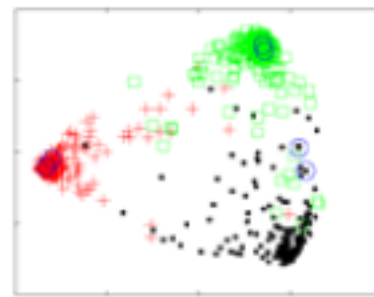
Similarities



Spectral Space



Cluster



Three main steps in Spectral clustering algorithms

1. Data preprocessing: normalization of the similarity matrix S
2. Spectral mapping
3. Clustering

Some Spectral clustering algorithms

Recursive spectral – split data into 2 partitions based on a single eigenvector, then recursively generate more partitions

Multiway spectral – use info from multiple eigenvectors to partition the data

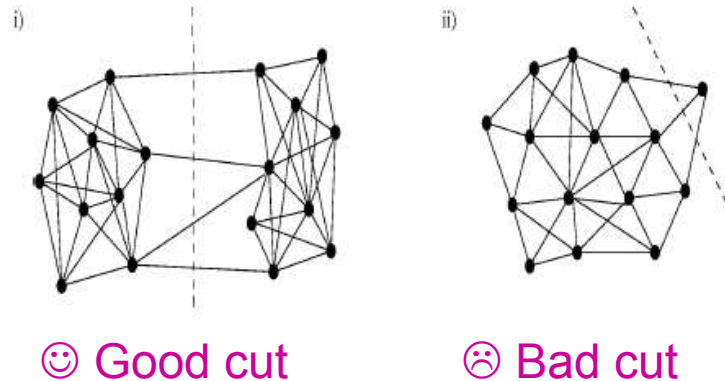
The Shi and Malik algorithm (SM)

1. Given an image or image sequence, set up a weighted graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ and set the weight on the edge connecting two nodes to be a measure of the similarity between the two nodes.
2. Solve $(\mathbf{D} - \mathbf{W})\mathbf{x} = \lambda\mathbf{D}\mathbf{x}$ for eigenvectors with the smallest eigenvalues.
3. Use the eigenvector with the second smallest eigenvalue to bipartition the graph.
4. Decide if the current partition should be subdivided and recursively repartition the segmented parts if necessary.

The Kannan, Vempala and Vetta Algorithm (KVV)

Similar to SM ... except it uses a measure for the goodness of a cut.

Optimal cuts are found with reference to a *conductance* measure ϕ for the cut.



Sum of sim scores
For all points in S

$$\phi(S) = \frac{\sum_{i \in S, j \notin S} a_{ij}}{\min(a(S), a(\bar{S}))}$$

Closest points
in S and not S

Meila-Shi algorithm

1. Compute $P = D^{-1}S$
2. Compute v^1, \dots, v^K the eigenvectors of P corresponding to the K largest eigenvalues. Form the matrix V whose columns are v^1, \dots, v^K .
3. Cluster the rows of V as points in a K -dimensional space.

Meila's paper 'A Comparison of Spectral Clustering Algorithms' shows some comparative results between various spectral clustering algorithms and delves into more theoretical details behind the algorithms.
(see course website for link to this paper)

Ng, Jordan and Weiss algorithm

S=similarity matrix. SVD \rightarrow S=UDV'

1. Set the diagonal elements S_{ii} to 0.
2. Compute the matrix

$$L = D^{-\frac{1}{2}} S D^{-\frac{1}{2}} \quad (2)$$

3. Let $1 = \mu_1 \geq \mu_2 \geq \dots \geq \mu_K$ be the K largest eigenvalues of L and u^1, u^2, \dots, u^K the corresponding eigenvectors² All eigenvectors are normalized to have unit length. Form the matrix $U = [u^1 u^2 \dots u^K]$ by stacking the eigenvectors in columns.
4. Form the matrix Y from U by renormalizing each of U 's rows to have unit length (i.e $Y_{ij} = U_{ij} / \sqrt{\sum_j U_{ij}^2}$).
5. K-Means-Orthogonal Treating each row of Y as a point in K dimensions, cluster them by the K-means algorithm to obtain the final clustering. The K-means algorithm is initialized by the Orthogonal-Initialization method described in [8].

Application of this algorithm to clustering genes to be presented by Nikita