Spectral clustering: an overview

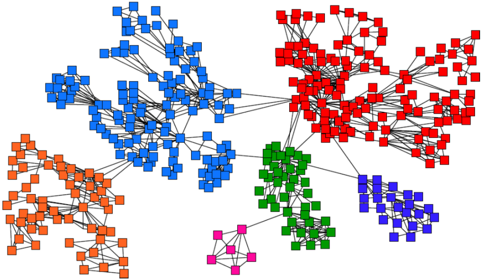
Maxim Panov

Institute for Information Transmission Problems

2015

Community detection

- **Graph**:
  - Nodes v_j
  - Edge weights $w_{ij} > 0$.
- **Problem**: Want to partition graph such that edges between groups have low weights.

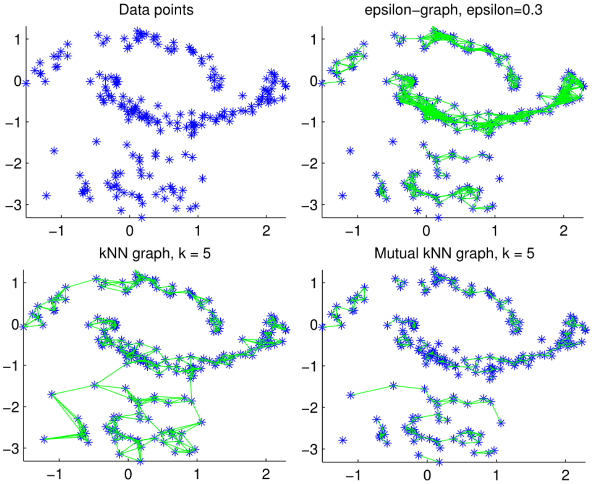


Similarity graphs

Types of graphs:

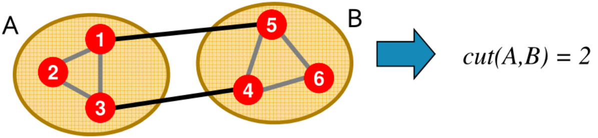
- ε-**neighborhood**:
  - Only include edges with distances $< \varepsilon$;
  - Treat as unweighted: $w_{ij} = Const$.
- **k-NN**:
  - Connect v_i and v_j if v_j is a k-NN of v_i.
  - Weighted by similarity $w_{ij} = s_{ij}$.
  - Directed or undirected.
- **Mutual k-NN**:
  - Same as k-NN, but only include mutual k-NN.

Similarity graphs



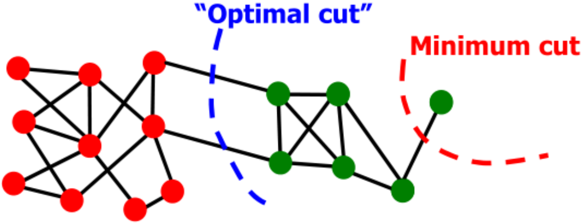
Graph cuts

- **Problem:** Partition graph such that edges between groups have low weights
- **Define:** $W(A, B) = \sum_{i \in A, j \in B} w_{ij}$.
- **MinCut problem:** $Cut(A_1, \ldots, A_k) = \sum_{i=1}^{k} W(A_i, \bar{A}_i)$.
- **Choose:** $A_1, \ldots, A_k = \arg\min_{A_1, \ldots, A_k} Cut(A_1, \ldots, A_k)$.



MinCut

Problem: MinCut favors isolated clusters



Solution:

- Ratio cuts (RatioCut)
- Normalized cuts (Ncut)
- Lead to "balanced" clusters

Graph terminology

Two measures of size of a subset:

- Cardinality:

$$|A| = \# \text{ of vertices in } A.$$

- Volume:

$$vol(A) = \sum_{i \in A} \sum_{j=1}^{N} w_{ij}.$$

Cuts Accounting for Size

- Ratio cuts (RatioCut)
  - $k = 2$: $RatioCut(A, \bar{A}) = Cut(A, \bar{A})\left(\frac{1}{|A|} + \frac{1}{|\bar{A}|}\right)$.
  - General k: $RatioCut(A_1, \ldots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{Cut(A_i, \bar{A}_i)}{|A_i|}$.
- Normalized cuts (Ncut)
  - $k = 2$: $NCut(A, \bar{A}) = Cut(A, \bar{A})\left(\frac{1}{Vol(A)} + \frac{1}{Vol(\bar{A})}\right)$.
  - General k: $NCut(A_1, \ldots, A_k) = \frac{1}{2}\sum_{i=1}^{k} \frac{Cut(A_i, \bar{A}_i)}{Vol(A_i)}$
- Problem is NP-hard!
- We need to look at relaxation.

Graph Laplacian

Definition: $L = D - W$.

Facts:

- Symmetric, positive semi-definite
- Eigenvalues:

$$0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N.$$

- λ_1 corresponds to eigenvector $\mathbf{u} = (1, \ldots, 1)^T$.
- Invariance to self-edges:

$$L_{ii} = d_i - w_{ii}, L_{ij} = -w_{ij}.$$

- Norm in L space:

$$\forall \mathbf{f} \in \mathbb{R}^N : \mathbf{f}^T L \mathbf{f} = \frac{1}{2} \sum_{i,j=1}^{N} w_{ij}(f_i - f_j)^2.$$

Relationship to Identifying Connected Components

Theorem

The multiplicity k of eigenvalue 0 of L is equal to the number of connected components.

Spectral clustering

Three basic stages:

1. Pre-processing
   - Construct a matrix representation of the graph.
2. Decomposition
   - Compute eigenvalues and eigenvectors of the matrix.
   - Map each point to a lower-dimensional representation based on one or more eigenvectors.
3. Grouping
   - Assign points to two or more clusters, based on the new representation.
     - Naive: thresholding (works for $k = 2$).
     - K-means in projected space (works for any $k \geq 2$).

Graph Laplacians and Ratio cuts

Ratio cuts for $k = 2$:

- Define cluster indicator variables:

$$f_i = \begin{cases} \sqrt{|\bar{A}|/|A|}, & v_i \in A, \\ -\sqrt{|A|/|\bar{A}|}, & v_i \notin A, \end{cases} \tag{1}$$

- Properties:

$$\sum_{i=1}^{N} f_i = |A|\sqrt{|\bar{A}|/|A|} - |\bar{A}|\sqrt{|A|/|\bar{A}|} = 0,$$

$$\|\mathbf{f}_A\|_2^2 = N.$$

- RatioCut

$$RatioCut(A, \bar{A}) = \frac{\mathbf{f}_A^T L \mathbf{f}_A}{|V|}.$$

Relaxation

- Reformulating RatioCut problem

$$\min_{A \subset V} \mathbf{f}_A^T L \mathbf{f}_A \text{ s.t. } \mathbf{f}_A \text{ is def. by Eq. (1)}, \ \mathbf{f}_A \perp \mathbf{1}, \ \|\mathbf{f}_A\| = \sqrt{N}.$$

- Still NP-hard!
- **Relaxation:**

$$\min_{\mathbf{f} \in \mathbb{R}^N} \mathbf{f}^T L \mathbf{f} \text{ s.t. } \mathbf{f} \perp \mathbf{1}, \ \|\mathbf{f}\| = \sqrt{N}.$$

- **Solution:** given by the vector $\mathbf{f}$ which is the eigenvector corresponding to the second smallest eigenvalue of L:

$$\lambda_2 = \min_{\mathbf{f} \in \mathbb{R}^N} \frac{\mathbf{f}^T L \mathbf{f}}{\mathbf{f}^T \mathbf{f}}.$$

Ratio Cuts for General k

- Define cluster indicator variables:

$$F_{ij} = \begin{cases} 1/\sqrt{|A_j|}, & v_i \in A_j, \\ 0, & v_i \notin A_j. \end{cases}$$

- RatioCut: define $F_A = \left(F_{ij}, i \in \overline{1, N}, j \in \overline{1, k}\right) \in \mathbb{R}^{N \times k}$; $F_A^T F_A = \mathrm{I}$:

$$RatioCut(A_1, \ldots, A_k) = \sum_{j=1}^{k} \mathbf{f}_{A_j}^T L \mathbf{f}_{A_j} = Tr\left(F_A^T L F_A\right).$$

- Reformulating RatioCut problem

$$\min_{A_1, \ldots, A_k} = Tr\left(F_A^T L F_A\right), \text{s.t. } F_A \text{ is defined above and } F_A^T F_A = \mathrm{I}.$$

- Relaxation: $\min_{F \in \mathbb{R}^{N \times k}} = Tr\left(F^T L F\right), \text{s.t.} F^T F = \mathrm{I}.$

Graph Laplacians and Norm cuts

Ratio cuts for $k = 2$:

- Define cluster indicator variables:

$$f_i = \begin{cases} \sqrt{vol(\bar{A})/vol(A)}, & v_i \in A, \\ -\sqrt{vol(A)/vol(\bar{A})}, & v_i \notin A, \end{cases} \tag{2}$$

- Properties:

$$(D\mathbf{f}_A) \perp \mathbf{1}, \quad \mathbf{f}_A^T D\mathbf{f}_A = vol(V).$$

- NCut

$$NCut(A, \bar{A}) = \frac{\mathbf{f}_A^T L\mathbf{f}_A}{vol(V)}.$$

Relaxation

- Reformulating NCut problem

$$\min_{A \subset V} \mathbf{f}_A^T L \mathbf{f}_A \text{ s.t. } \mathbf{f}_A \text{ is def. by Eq. (2)}, \; D\mathbf{f}_A \perp \mathbf{1}, \; \mathbf{f}_A^T D \mathbf{f}_A = vol(V).$$

- Still NP-hard!
- **Relaxation:**

$$\min_{\mathbf{f} \in \mathbb{R}^N} \mathbf{f}^T L \mathbf{f} \text{ s.t. } D\mathbf{f} \perp \mathbf{1}, \; \mathbf{f}^T D \mathbf{f} = vol(V).$$

or equivalently for $\mathbf{g} = D^{1/2}\mathbf{f}$

$$\min_{\mathbf{g} \in \mathbb{R}^N} \mathbf{g}^T D^{-1/2} L D^{-1/2} \mathbf{g} \text{ s.t. } \mathbf{g} \perp D^{1/2}\mathbf{1}, \; \|\mathbf{g}\|^2 = vol(V).$$

- **Solution:** given by the vector $\mathbf{f}$ which is the eigenvector corresponding to the second smallest eigenvalue of $L_{sym} = D^{-1/2} L D^{-1/2}$.

Norm Cuts for General k

- Define cluster indicator variables:

$$F_{ij} = \begin{cases} 1/\sqrt{vol(A_j)}, & v_i \in A_j, \\ 0, & v_i \notin A_j. \end{cases}$$

- Reformulating NCut problem

$$\min_{A_1,\dots,A_k} = Tr\left(F_A^T L F_A\right), \text{s.t. } F_A \text{ is defined above and } F_A^T D F_A = \mathrm{I}.$$

- Relaxation:

$$\min_{H \in \mathbb{R}^{N \times k}} = Tr\left(H^T D^{-1/2} L D^{-1/2} H\right), \text{s.t.} H^T H = \mathrm{I}.$$

Spectral clustering

Which graph Laplacian to use?

- If degrees in graph vary significantly, then Laplacians are quite different.
- In general, L_{rw} behaves the best.
- Volume gives better measure of within-cluster similarity than cardinality.
- Normalized cuts has consistency results, Ratio cuts does not.