Graph cuts

- $G$ graph with (weighted) adjacency matrix $W = (w_{ij})$.
- We define:
  $$W(A,B) := \sum_{i \in A, j \in B} w_{ij}.$$  
- $|A| := \text{number of vertices in } A$.
- $\text{vol}(A) := \sum_{i \in A} d_i$.

Given a partition $A_1, \ldots, A_k$ of the vertices of $G$, we let
  $$\text{cut}(A_1, \ldots, A_k) := \frac{1}{2} \sum_{i=1}^k W(A_i, \overline{A_i}).$$

The min-cut problem consists of solving:
  $$\min_{V = A_1 \cup \cdots \cup A_k} \text{cut}(A_1, \ldots, A_k).$$

Balanced cuts

The two most common objective functions that are used as a replacement to the min-cut objective are:

- RatioCut (Hagen and Kahng, 1992):
  $$\text{RatioCut}(A_1, \ldots, A_k) := \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \overline{A_i})}{|A_i|} = \sum_{i=1}^k \frac{\text{cut}(A_i, \overline{A_i})}{|A_i|}.$$  

- Normalized cut (Shi and Malik, 2000):
  $$\text{Ncut}(A_1, \ldots, A_k) := \frac{1}{2} \sum_{i=1}^k \frac{W(A_i, \overline{A_i})}{\text{vol}(A_i)} = \sum_{i=1}^k \frac{\text{cut}(A_i, \overline{A_i})}{\text{vol}(A_i)}.$$  

- Note: both objective functions take larger values when the clusters $A_i$ are “small”.
- Resulting clusters are more “balanced”.
- However, the resulting problems are NP hard - see Wagner and Wagner (1993).
Spectral clustering provides a way to relax the RatioCut and the Normalized cut problems.

Strategy:
1. Express the original problem as a linear algebra problem involving discrete/combinatorial constraints.
2. Relax/remove the constraints.

RatioCut with $k = 2$: solve

$$\min_{A \subset V} \text{RatioCut}(A, \overline{A}).$$

Given $A \subset V$, let $f \in \mathbb{R}^n$ be given by

$$f_i := \begin{cases} \sqrt{|A|/|\overline{A}|} & \text{if } v_i \in A \\ -\sqrt{|A|/|\overline{A}|} & \text{if } v_i \not\in A. \end{cases}$$

Relaxing RatioCut

Let $L = D - W$ be the (unnormalized) Laplacian of $G$. Then

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2$$

$$= \frac{1}{2} \sum_{i \in A, j \in \overline{A}} w_{ij} \left( \sqrt{|A|/|\overline{A}|} + \sqrt{|\overline{A}|/|A|} \right)^2 + \frac{1}{2} \sum_{i \in A, j \in A} w_{ij} \left( -\sqrt{|A|/|\overline{A}|} - \sqrt{|\overline{A}|/|A|} \right)^2$$

$$= W(A, \overline{A}) \left( 2 + \frac{|A|}{|\overline{A}|} + \frac{|\overline{A}|}{|A|} \right)$$

$$= W(A, \overline{A}) \left( \frac{|A| + |\overline{A}|}{|A|} + \frac{|A| + |\overline{A}|}{|\overline{A}|} \right)$$

$$= |V| \cdot \frac{1}{2} \left( \frac{W(A, \overline{A})}{|A|} + \frac{W(\overline{A}, A)}{|\overline{A}|} \right)$$

$$= |V| \cdot \text{RatioCut}(A, \overline{A}).$$

since $|A| + |\overline{A}| = |V|$, and $W(A, \overline{A}) = W(\overline{A}, A)$.

Relaxing RatioCut (cont.)

We have:

$$\min_{A \subset V} \text{RatioCut}(A, \overline{A}).$$

subject to $f \perp \mathbb{1}$, $\|f\| = \sqrt{n}$, $f_i$ defined as above.

- This is a discrete optimization problem since the entries of $f$ can only take two values: $\sqrt{|\overline{A}|/|A|}$ and $-\sqrt{|A|/|\overline{A}|}$.

- The problem is NP hard.

The natural relaxation of the problem is to remove the discreteness condition on $f$ and solve

$$\min_{f \in \mathbb{R}^n} f^T L f$$

subject to $f \perp \mathbb{1}$, $\|f\| = \sqrt{n}$. 

Thus, we have showed that the Ratio-Cut problem is equivalent to

$$\min_{A \subset V} \text{RatioCut}(A, \overline{A}).$$

subject to $f \perp \mathbb{1}$, $\|f\| = \sqrt{n}$, $f_i$ defined as above.
Relaxing RatioCut

Using properties of the Rayleigh quotient, it is not hard to show that the solution of

\[
\min_{f \in \mathbb{R}^n} f^T Lf
\]

subject to \( f \perp 1, \|f\| = \sqrt{n} \)
is an eigenvector of \( L \) corresponding to the second eigenvalue of \( L \).

Clearly, if \( \tilde{f} \) is the solution of the problem, then \( \tilde{f}^T L \tilde{f} \leq \min_{A \subseteq V} \text{RatioCut}(A, A) \).

How do we get the clusters from \( \tilde{f} \)?

We could set

\[
\left\{ \begin{array}{ll}
v_i \in A & \text{if } f_i \geq 0 \\
v_i \in \overline{A} & \text{if } f_i < 0.
\end{array} \right.
\]

More generally, we cluster the coordinates of \( f \) using \( K \)-means.

This is the unnormalized spectral clustering algorithm for \( k = 2 \).

Relaxing RatioCut : \( k > 2 \)

Now,

\[
\tilde{h}_i^T L \tilde{h}_i = (H^T LH)_{ii}.
\]

Thus,

\[
\text{RatioCut}(A_1, \ldots, A_k) = \sum_{i=1}^{k} \frac{\text{cut}(A_i, \overline{A_i})}{|A_i|} = \sum_{i=1}^{k} \tilde{h}_i^T L \tilde{h}_i = \text{Tr}(H^T LH).
\]

So the problem

\[
\min_{V = A_1 \cup \cdots \cup A_k} \text{RatioCut}(A_1, \ldots, A_k)
\]
is equivalent to

\[
\min_{H \in \mathbb{R}^{n \times k}} \text{Tr}(H^T LH)
\]

subject to \( H^T H = I_{k \times k} \), \( H \) defined as above.

As before, we consider a natural relaxation of the problem:

\[
\min_{H \in \mathbb{R}^{n \times k}} \text{Tr}(H^T LH)
\]

subject to \( H^T H = I_{k \times k} \).
The unnormalized spectral clustering algorithm:

Unnormalized spectral clustering

Input: Similarity matrix \( S \in \mathbb{R}^{n \times n} \), number \( k \) of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let \( W \) be its weighted adjacency matrix.
- Compute the unnormalized Laplacian \( L \).
- Compute the first \( k \) eigenvectors \( u_1, \ldots, u_k \) of \( L \).
- Let \( U \in \mathbb{R}^{n \times k} \) be the matrix containing the vectors \( u_1, \ldots, u_k \) as columns.
- For \( i = 1, \ldots, n \), let \( y_i \in \mathbb{R}^k \) be the vector corresponding to the \( i \)-th row of \( U \).
- Cluster the points \( y_1, \ldots, y_n \) in \( \mathbb{R}^k \) with the \( k \)-means algorithm into clusters \( C_1, \ldots, C_k \).
- Output: Clusters \( A_1, \ldots, A_k \) with \( A_i = \{ j \mid y_j \in C_i \} \).


Normalized spectral clustering according to Shi and Malik (2000)

Input: Similarity matrix \( S \in \mathbb{R}^{n \times n} \), number \( k \) of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let \( W \) be its weighted adjacency matrix.
- Compute the unnormalized Laplacian \( L \).
- Compute the first \( k \) eigenvectors \( u_1, \ldots, u_k \) of \( L \).
- Let \( U \in \mathbb{R}^{n \times k} \) be the matrix containing the vectors \( u_1, \ldots, u_k \) as columns.
- For \( i = 1, \ldots, n \), let \( y_i \in \mathbb{R}^k \) be the vector corresponding to the \( i \)-th row of \( U \).
- Cluster the points \( y_1, \ldots, y_n \) in \( \mathbb{R}^k \) with the \( k \)-means algorithm into clusters \( C_1, \ldots, C_k \).
- Output: Clusters \( A_1, \ldots, A_k \) with \( A_i = \{ j \mid y_j \in C_i \} \).


- Relaxing the RatioCut leads to unnormalized spectral clustering.
- By relaxing the Ncut problem, we obtain the normalized spectral clustering algorithm of Shi and Malik (2000).

Note: The solutions of \( Lu = \lambda Du \) are the eigenvectors of \( L_{rw} \).

The normalized clustering algorithm of Ng et al.

Another popular variant of the spectral clustering algorithm was provided by Ng, Jordan, and Weiss (2002).

The algorithm uses \( L_{sym} \) instead of \( L \) (unnormalized clustering) or \( L_{rw} \) (Shi and Malik’s normalized clustering).

Normalized spectral clustering according to Ng, Jordan, and Weiss (2002)

Input: Similarity matrix \( S \in \mathbb{R}^{n \times n} \), number \( k \) of clusters to construct.

- Construct a similarity graph by one of the ways described in Section 2. Let \( W \) be its weighted adjacency matrix.
- Compute the normalized Laplacian \( L_{sym} \).
- Compute the first \( k \) eigenvectors \( u_1, \ldots, u_k \) of \( L_{sym} \).
- Let \( U \in \mathbb{R}^{n \times k} \) be the matrix containing the vectors \( u_1, \ldots, u_k \) as columns.
- Form the matrix \( T \in \mathbb{R}^{k \times n} \) from \( U \) by normalizing the rows to norm 1, that is in set \( t_{ij} = u_j / (\sum_i u_{ij})^{1/2} \).
- For \( i = 1, \ldots, n \), let \( y_i \in \mathbb{R}^k \) be the vector corresponding to the \( i \)-th row of \( T \).
- Cluster the points \( y_1, \ldots, y_n \) with the \( k \)-means algorithm into clusters \( C_1, \ldots, C_k \).
- Output: Clusters \( A_1, \ldots, A_k \) with \( A_i = \{ j \mid y_j \in C_i \} \).
