Spectral Clustering

References:

U. Von Luxburg, “A tutorial on spectral clustering,” *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

Spectral Clustering

- The study of Laplacian eigenvalues revealed the structure of graphs, in particular the existence of a partition.
- Eigenvectors reveal how to select partitions
- Can we make these insights more explicit and formulate a spectral theory of clustering ?

Reference: U. Von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

Back to the Start: Cut and Cluster

When cutting through edges, we can associate cost functions inspired by the Cheeger constant:

$$C(A,B) := \sum_{i\in A, j\in B} \mathbf{W}[i,j]$$

$$\mathrm{RatioCut}(A,\overline{A}) := \frac{1}{2}\frac{C(A,\overline{A})}{|A|} + \frac{1}{2}\frac{C(A,\overline{A})}{|\overline{A}|}$$

$$\mathrm{NormalizedCut}(A,\overline{A}) = \frac{1}{2}\frac{C(A,\overline{A})}{\mathrm{vol}(A)} + \frac{1}{2}\frac{C(A,\overline{A})}{\mathrm{vol}(\overline{A})}$$

Normalization seeks to impose **balanced** clusters

Exposing RatioCut

Let's try to solve: $$\min_{A \subset V} \mathrm{RatioCut}(A, \overline{A})$$

Observations:

$$f[i] = \begin{cases} \sqrt{|\overline{A}|/|A|} & \text{if } i \in A \\ -\sqrt{|A|/|\overline{A}|} & \text{if } i \in \overline{A} \end{cases}$$

f is the indicator of the partition

$\Leftrightarrow$ $$f^T \mathbf{L} f = |V| \mathrm{RatioCut}(A, \overline{A})$$

$$\|f\| = \sqrt{|V|} \text{ and } \langle f, 1 \rangle = 0$$

Exposing RatioCut

The following problem is equivalent to Ratiocut:

$$\arg\min_{\boxed{A \subset V}} f^T \mathbf{L} f \text{ subject to } \|f\| = \sqrt{N}, \quad \langle f, 1\rangle = 0 \text{ and } \boxed{f \text{ indicator of } A}$$

NP-hard

$$\arg\min_{f} f^T \mathbf{L} f \text{ subject to } \|f\| = \sqrt{N}, \quad \langle f, 1\rangle = 0$$

Relaxed problem: Looking for a **smooth** partition function!

Exposing RatioCut

$$\arg\min_f f^T \mathbf{L} f \text{ subject to } \|f\| = \sqrt{N}, \quad \langle f, 1 \rangle = 0$$

Solution (G connected): eigenvector of λ_2

Warning: recover partition after thresholding $f = \mathrm{sign}(u_2)$

So we are back to the Fiedler vector !!!

RatioCut: Generalizing to $k > 2$

For more than two components, we look for a set of partition functions

$$F \in \mathbb{R}^{N\times k} \qquad F[i,j] = f_j[i] = \begin{cases} 1/\sqrt{|A_j|} & \text{if } v_i \in A_j \\ 0 & \text{otherwise} \end{cases}$$

Observe: $$f_j^T \mathbf{L} f_j = \frac{\mathrm{Cut}(A_j, \overline{A_j})}{|A_j|} \qquad F^T F = \mathbb{I}$$

$$\mathrm{RatioCut}(A_1, \ldots, A_k) = \mathrm{Tr}(F^T \mathbf{L} F)$$

Suggests the relaxed problem:

$$\arg \min_{F \in \mathbb{R}^{N\times k}} \mathrm{Tr}(F^T \mathbf{L} F) \text{ such that } F^T F = \mathbb{I}$$

Unnormalized Spectral Clustering

This form of relaxed RatioCut = **Unnormalized Spectral Clustering**

$$\arg\min_{F \in \mathbb{R}^{N \times k}} \mathrm{Tr}(F^T \mathbf{L} F) \text{ such that } F^T F = \mathbb{I}$$

Algorithm: Unnormalized Spectral Clustering

Compute the matrix F of first k eigenvectors of $\mathbf{L}$

Apply k-means to rows of F to obtain cluster assignments

Normalized Cut, k=2

$$\text{NormalizedCut}(A,\overline{A}) = \frac{1}{2}\frac{C(A,\overline{A})}{\text{vol}(A)} + \frac{1}{2}\frac{C(A,\overline{A})}{\text{vol}(\overline{A})}$$

$$f[i] = \begin{cases} \sqrt{\text{vol}(\overline{A})/\text{vol}(A)} & \text{if } v_i \in A \\ -\sqrt{\text{vol}(A)/\text{vol}(\overline{A})} & \text{otherwise} \end{cases}$$

Check that: $\langle \mathbf{D}f, 1\rangle = 0 \qquad f^T\mathbf{D}f = \text{vol}(G)$

$$f^T\mathbf{L}f = \text{vol}(V)\text{NormalizedCut}(A,\overline{A})$$

$$\arg\min_f f^T\mathbf{L}f \text{ subject to } f^T\mathbf{D}f = \text{vol}(G), \quad \langle \mathbf{D}f, 1\rangle = 0$$

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

$$\arg\min_g g^T\mathbf{L}_{\text{norm}}g \text{ subject to } \|g\|^2 = \text{vol}(G), \quad \langle g, \mathbf{D}^{1/2}1\rangle = 0$$

Normalized Cut, k>2

$$F[i,j] = f_j[i] = \begin{cases} 1/\sqrt{\mathrm{vol}(A_j)} & \text{if } v_i \in A_j \\ 0 & \text{otherwise} \end{cases}$$

$$f_j^T \mathbf{L} f_j = \frac{\mathrm{Cut}(A_j, \overline{A_j})}{\mathrm{vol}(A_j)} \qquad F^T F = \mathbb{I} \qquad f_j^T \mathbf{D} f_j = 1$$

$$\arg \min_{H \in \mathbb{R}^{N \times k}} \mathrm{Tr}(H^T \mathbf{L}_{\mathrm{norm}} H) \text{ such that } H^T H = \mathbb{I} \qquad H = \mathbf{D}^{1/2} F$$

Algorithm: Normalized Spectral Clustering

Compute the matrix H of first k eigenvectors of $\mathbf{L}_{\mathrm{norm}}$

Apply k-means to rows of H to obtain cluster assignments

Applications

In practice normalised spectral clustering is often preferred

In practice the eigenvectors are "re-normalized" by the degrees

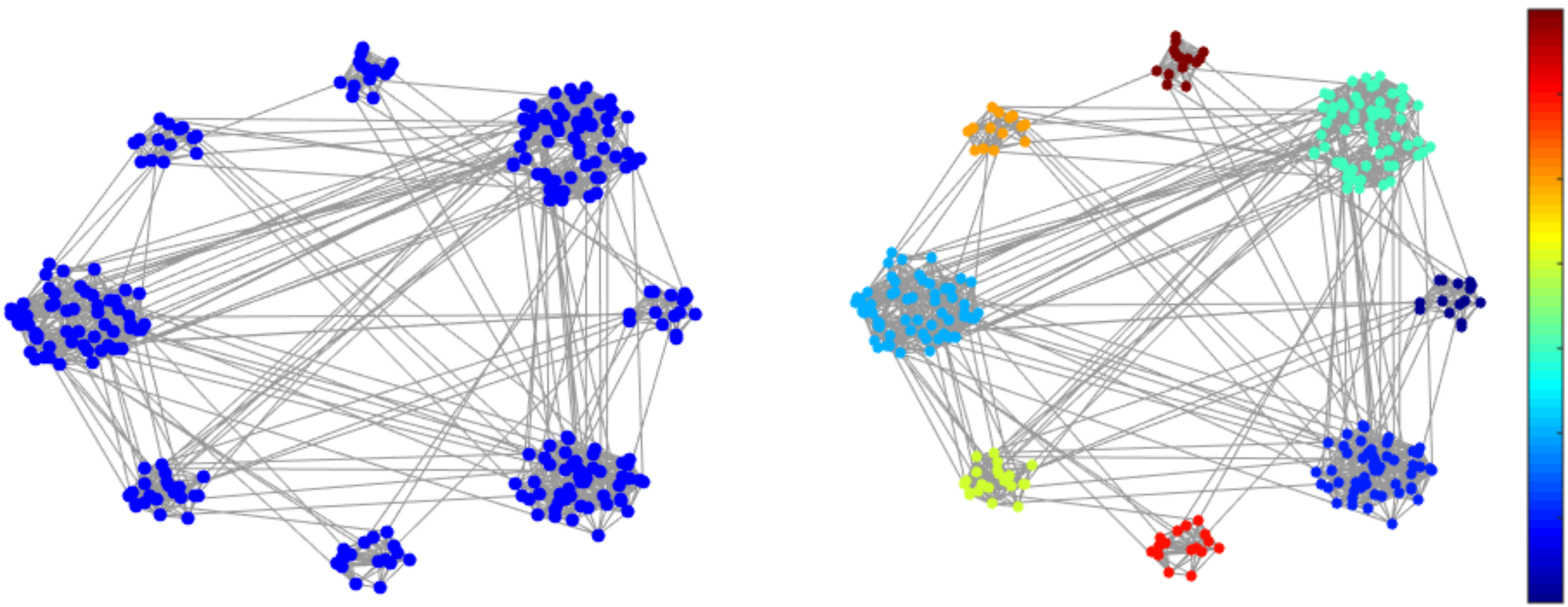
$$F = \mathbf{D}^{-1/2} H$$

before k-means, because these are real cluster assignments

Rem: this is equivalent to using the "random walk Laplacian"

$$\mathbf{L}_{\mathrm{rw}} = \mathbf{D}^{-1}\mathbf{L}$$

If data has k **clear** clusters, there will be a gap in the Laplacian spectrum after the k-th eigenvalue. Use to choose k.


8
7
6
5
4
3
2
1

Example