Spectral Clustering: A Graph Cut Formulation

Ira A. Fulton Schools of Engineering
Arizona State University

Objective

Objective

Define the graph partition formulation

Objective

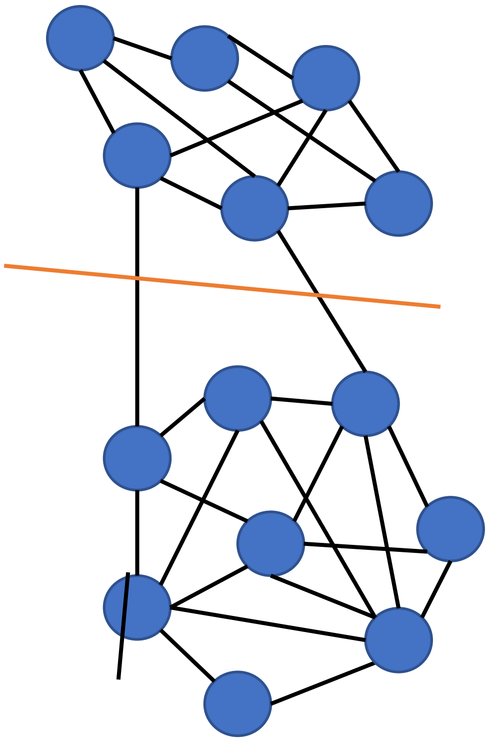
Learn how to solve simple graph partition

Clustering as Graph Partition/Cut

- Find a partition of a graph such that the edges between different groups have a very low weight while the edges within a group have high weight.

- E.g., minimum cut

CutSize = 2



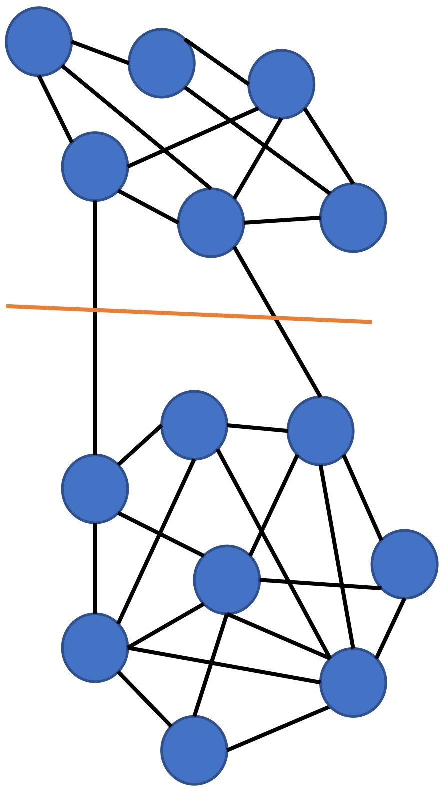
- More general, consider weighted edges.

2-way Spectral Graph Partitioning

- Weighted adjacency matrix **W**
  - $w_{i,j}$: the weight between two vertices *i* and *j*

- (Cluster) Membership vector **q**

$$q_i = \begin{cases} 1 & i \in Cluster\ A \\ -1 & i \in Cluster\ B \end{cases}$$



$$\mathbf{q} = \underset{\mathbf{q} \in [-1,\ 1]^n}{\operatorname{argmin}}\ CutSize\,,$$

$$CutSize = \mathrm{J} = \frac{1}{4} \sum_{i,j} (q_i - q_j)^2 w_{i,j}$$

Solving the Optimization Problem

$$\mathbf{q} = \underset{\mathbf{q}\in[-1,\ \ 1]^n}{\operatorname{argmin}} \frac{1}{4}\sum_{i,j}(q_i - q_j)^2 w_{i,j}\ ,$$

- Directly solving the above problem requires combinatorial search → exponential complexity
- How to reduce the computational complexity?

Relaxation Approach

- Key difficulty: q_i has to be either -1,1.
  - Relax q_i to be any real number.
  - Impose Constraint: $\sum_{i=1}^{n} q_i^2 = n$

$$J = \frac{1}{4}\sum_{i,j}(q_i - q_j)^2 w_{i,j} = \frac{1}{4}\sum_{i,j}(q_i^2 - 2q_i q_j + q_j^2) w_{i,j}$$

$$= \frac{1}{4}\sum_i 2q_i^2\left(\sum_j w_{i,j}\right) - \frac{1}{4}\sum_{i,j} 2q_i q_j w_{i,j}$$

$$= \frac{1}{2}\sum_i q_i^2 d_i - \frac{1}{2}\sum_{i,j} q_i (d_i \delta_{i,j} - w_{i,j}) q_j$$

where $d_i = \sum_j w_{i,j}$ and $D \equiv [d_i \delta_{i,j}]$

$$\Rightarrow \; J = \frac{1}{2}\mathbf{q}^T(\mathbf{D} - \mathbf{W})\mathbf{q}$$

Relaxation Approach (cont'd)

- The final problem formulation:

$$\mathbf{q} = \underset{\mathbf{q}}{\operatorname{argmin}}\ J = \underset{\mathbf{q}}{\operatorname{argmin}}\, \mathbf{q}^T(\mathbf{D} - \mathbf{W})\mathbf{q} \ ,$$

subject to $\sum_{i=1}^{n} q_i^2 = n$

- Solution: the second minimum eigenvector for **D-W**

$$(\mathbf{D} - \mathbf{W})\mathbf{q} = \lambda_2\ \mathbf{q}$$

Graph Laplacian

- $\mathbf{L} = \mathbf{D} - \mathbf{W}$
- $\mathbf{L}$ is semi-positive definitive matrix.
  - For any $\mathbf{x}$, we have $\mathbf{x}^T\mathbf{L}\mathbf{x} \geq 0$. (Why?)
- Minimum eigenvalue $\lambda_1 = 0$ (what is the eigenvector?)

$$0 = \lambda_1 < \lambda_2 < \lambda_3 \ldots < \lambda_k$$

- The eigenvector that corresponds to the second minimum eigenvalue λ_2 gives the best bipartite graph partition.

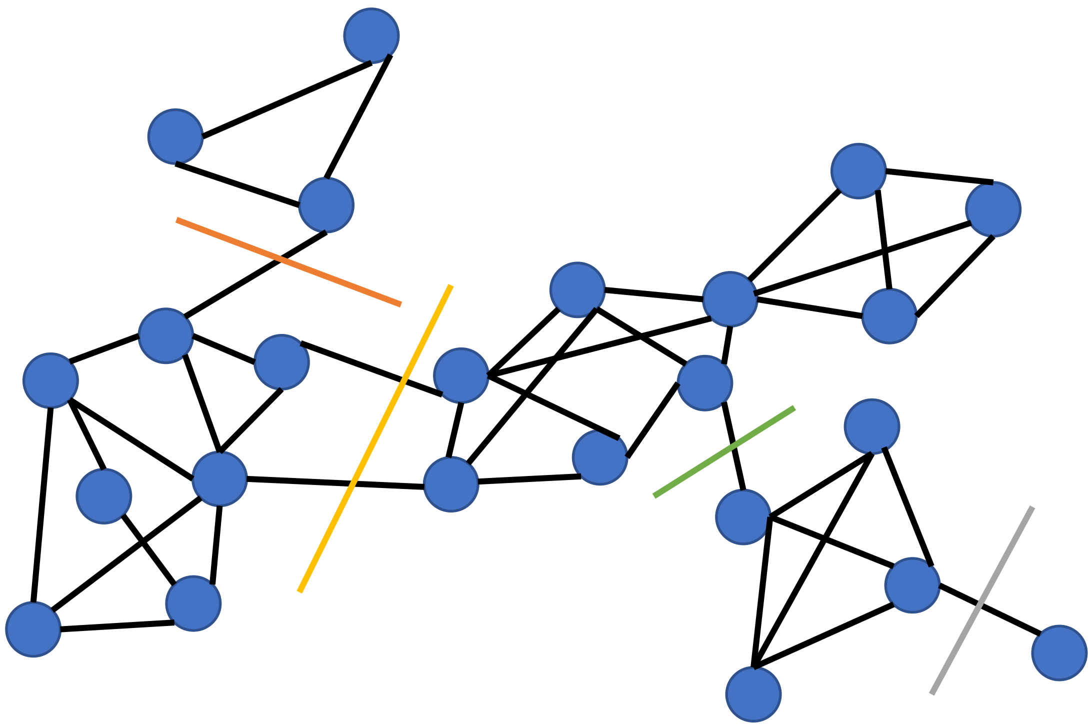
Recovering the Partitions

- Due to the relaxation, **q** can be any number (not just -1 and 1)
- How to construct the partition based on the eigenvector?
- A simple strategy :

$$A = \{i | q_i < 0\}, \qquad B = \{i | q_i \geq 0\}$$

One Obvious Drawback

- Minimum cut does not balance the size of bipartite graphs.



- How should we consider other factors like the sizes of the partitions?