CSE 575: Statistical Machine Learning (Spring 2021)
Instructor: Nupur Thakur

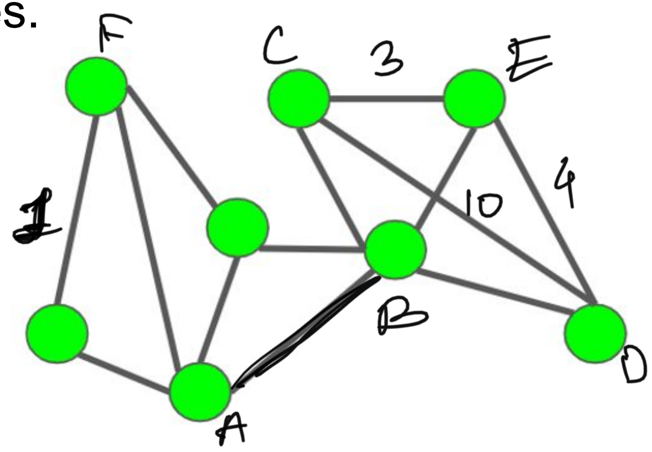
Spectral Clustering

Arizona State University

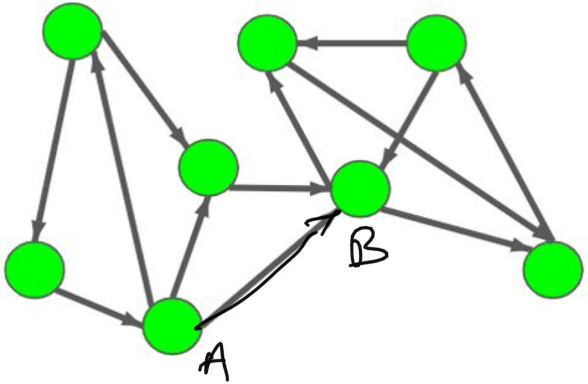
Table of contents

What is a graph?

- A graph G = (V, E) is defined by V, a set of N vertices, and E, a set of edges.



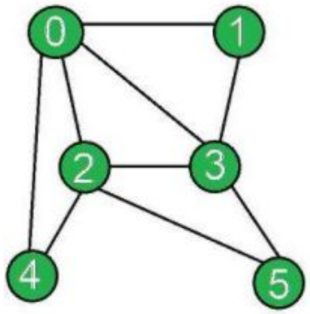
Undirected graph



Directed graph

Graph Representation - Adjacency Matrix W

- For an undirected graph - N vertices



NXN

$$\begin{array}{cc} & N \\ N & \begin{bmatrix} \quad & \quad & \quad \\ & & \\ & & \end{bmatrix} \end{array}$$

Adjacency matrix — 6X6

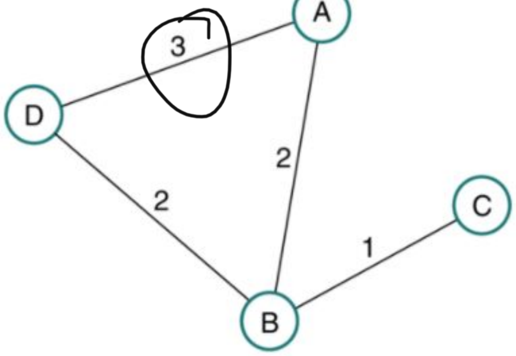
$$\begin{array}{cc} & \begin{matrix}0 & 1 & 2 & 3 & 4 & 5\end{matrix} \\ \begin{matrix}0\\1\\2\\3\\4\\5\end{matrix} & \begin{bmatrix}0 & 1 & 1 & 1 & 1 & 0 \\ & 0 & & & & \\ & & 0 & & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0\end{bmatrix} \end{array}$$

Graph Representation - Weighted Adjacency Matrix

NXN matrix

4X4 matrix

$$\begin{array}{c c} & \begin{array}{cccc} A & B & C & D \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \end{array} & \begin{bmatrix} 0 & 2 & 0 & 3 \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} \end{array}$$



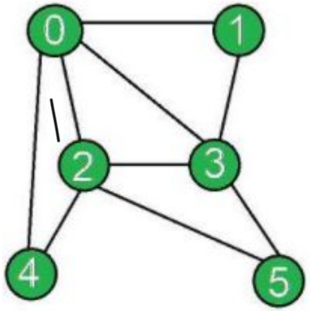
Graph Representation - Degree Matrix D

- Degree of a node is the number of edges incident to the node.

- NXN matrix for N nodes.
- Diagonal matrix

$$\begin{array}{c|c} & \begin{matrix} 0 & 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 4 & 0 & 0 & 0 & & \\ 0 & 2 & & & & \\ & & 4 & & & \\ 0 & & & 4 & & \\ 0 & & & & 2 & \\ 0 & & & & & 2 \end{bmatrix} \end{array}$$

— Undirected graph.



Graph Representation - Graph Laplacian L

Degree matrix

- **L = D - W** — Unnormalized

Adjacency matrix

- Properties of **L** -
  - **L** is symmetric and positive semi-definite
  - L has n non-negative, real-valued eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$
  - The smallest eigenvalue is 0, the corresponding eigenvector is the 1-vector (all elements being 1)

Refer for all more details on the properties: https://arxiv.org/pdf/0711.0189.pdf and https://sites.cs.ucsb.edu/~veronika/SpectralClustering.pdf

Different ways to build similarity graphs

- ε-neighborhood graph $dist(A, B) < \varepsilon \Rightarrow$ Edge betn A & B
  - unweighted graph.
- k-nearest neighbor graph   –

  mutual k-nearest neighbor - property is satisfied both ways
- fully connected graph

Clustering – Increase inter-cluster distance
Decrease intra-cluster distance.

Clustering as a Graph Partition

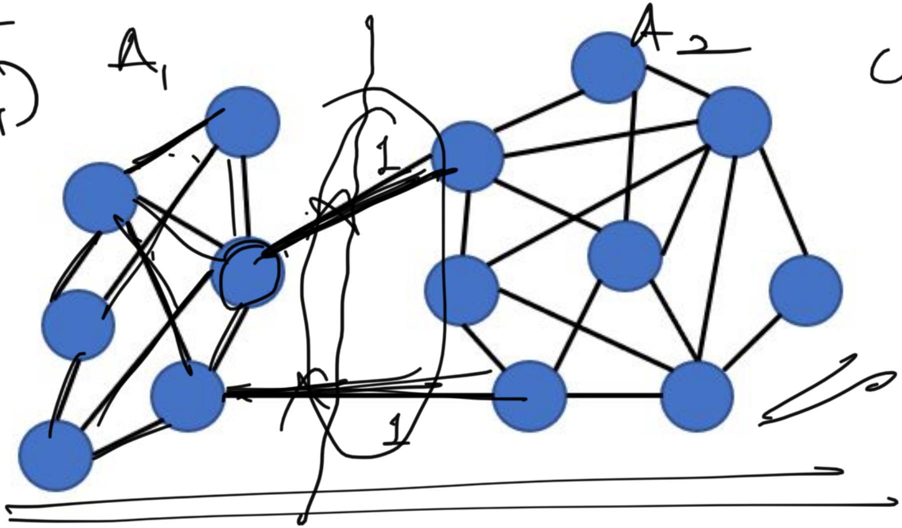
- Find a partition of a graph such that the edges between different groups have a very low weight while the edges within a group have high weight.

$w(A_1, A_1)$

$|A_1| = 6$

$|A_2| = 8$

$vol(A_1) =$



cluster = 2

MinCut Problem

- Given a similarity graph with adjacency matrix W, the simplest and most direct way to construct a partition of the graph is to solve the mincut problem.

i-th V

complement of A_i

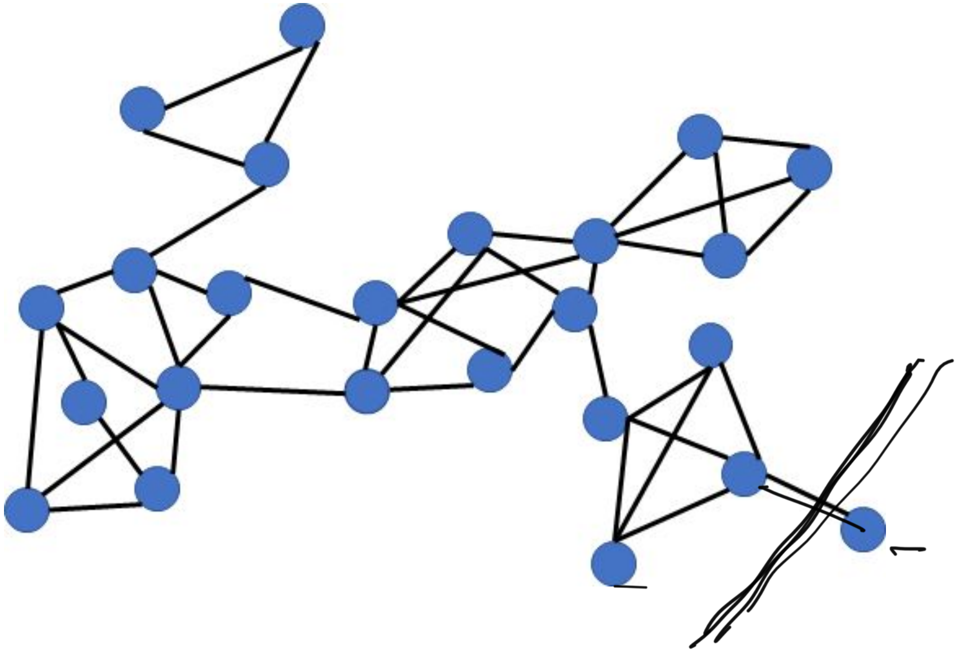
$$\text{minimize} \left\{ \mathrm{cut}(A_1, \ldots, A_k) := \frac{1}{2} \sum_{i=1}^{k} W(A_i, \overline{A}_i) \right\}$$

$k=2$

$$\mathrm{cut}(A_1, A_2) = \frac{1}{2} \sum_{i=1}^{2} W(A_{1,2}, \overline{A}_1)$$

MinCut Problem - Major Drawback

- It does not balance the size of the partitions



$|A_i| \Rightarrow$ No. of vertices/nodes in a cluster A_i.

Other types of cuts

- To overcome the drawback of MinCut, one obvious solution is to specify that the partitions are reasonably large.

$$\mathrm{RatioCut}(A_1, \ldots, A_k) := \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \overline{A}_i)}{|A_i|} = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A}_i)}{|A_i|}$$

Normalized cut. $\leftarrow$

$$\mathrm{Ncut}(A_1, \ldots, A_k) := \frac{1}{2}\sum_{i=1}^{k} \frac{W(A_i, \overline{A}_i)}{\mathrm{vol}(A_i)} = \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A}_i)}{\mathrm{vol}(A_i)}.$$

$$\mathrm{MinMaxCut}(A_1, \ldots, A_k) := \sum_{i=1}^{k} \frac{\mathrm{cut}(A_i, \overline{A}_i)}{W(A_i, A_i)}.$$

Questions?