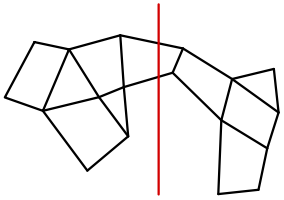
Lecture 35: Finding Clusters in Graphs



2 clusters

Problem: Find x y

to minimize

$$\sum \|a_i - x\|^2 + \sum \|b_i - y\|^2$$

① a's $\cup$ b's = all nodes

② a's $\cap$ b's = empty set

Given a's $= a_1, \cdots, a_k$

what is best x?

x = centroid of a's $= \dfrac{\text{sum of } a\text{'s}}{\text{number of } a\text{'s}}$

① K-means (here 2-means)

① Given a's b's find centroids

② Given x, y form best clusters

Each node goes with the closer of x, y.

② "Spectral clustering"

Start with graph Laplacian matrix

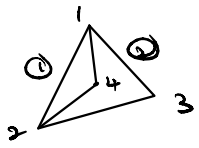
$$L = A^T A = D - B$$

A: Incidence matrix

L: positive semidefinite

B: adjacency matrix

D: degree matrix (diagonal)



Incidence matrix

$$\begin{matrix}1\\2\\3\\4\\5\end{matrix}\begin{bmatrix}-1 & 1 & 0 & 0\\ -1 & 0 & 1 & 0\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot\end{bmatrix} \qquad A$$

$$D = \begin{bmatrix}3 & & & \\ & 3 & & \\ & & 2 & \\ & & & 2\end{bmatrix} \qquad B = \begin{bmatrix}0 & 1 & 1 & 1\\ 1 & 0 & 1 & 1\\ 1 & 1 & 0 & 0\\ 1 & 1 & 0 & 0\end{bmatrix}$$

$$L = D - B = \begin{bmatrix}3 & -1 & -1 & -1\\ -1 & 3 & -1 & -1\\ -1 & -1 & 2 & 0\\ -1 & -1 & 0 & 2\end{bmatrix}$$

$$Lx = \begin{bmatrix}0\\0\\0\\0\end{bmatrix} \text{ for } x = \begin{bmatrix}c\\c\\c\\c\end{bmatrix}$$

$\dim N(L) = 1$

L has $\lambda_1 = 0$

$$x_1 = \begin{bmatrix}c\\c\\c\\c\end{bmatrix}$$

FIEDLER

x_2 = eigenvector for λ_{min} of L

positive and negative components of its eigenvector indicate the 2 clusters of nodes.