L1~~0~~3: Spectral Clustering

Jeff M. Phillips

February 24, 2025

Clustering

Input: Data set $X=\{x_1, x_2, \dots$

~~Distance~~ $D: X \times X \to \mathbb{R}$

Graph / similarity

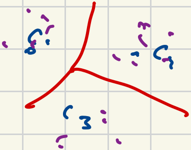
1. Hierarchical Agglomerative

   Density-based

   bottom up : find two closest sets merge.

2. Assignment-based

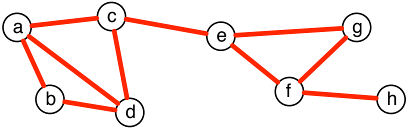
   Find Centers $C=\{c_1 \dots c_k\}$



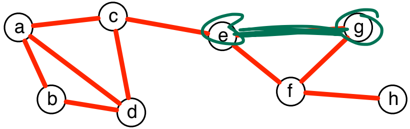
3. Spectral Clustering

   - top down
   - input Graph (similarity)
   - Dimensionality Reduction.

Graphs



Graphs



Vertex/Edge Set-style: $G = (V, E)$ where

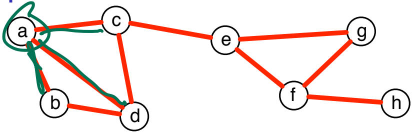
$V = \{a, b, c, d, e, f, g\}$ and

Data points

No order

$$E = \Big\{\{a, b\}, \{a, c\}, \{a, d\}, \{b, d\}, \{c, d\}, \{c, e\}, \{e, f\}, \boxed{\{e, g\}}, \{f, g\}, \{f, h\}\Big\}.$$

Graphs



Vertex/Edge Set-style: $G = (V, E)$ where

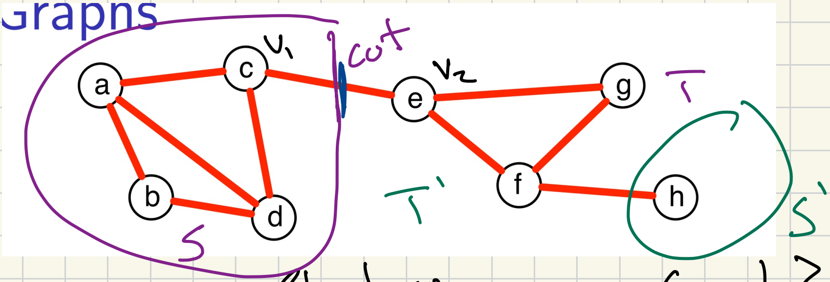
$V = \{a, b, c, d, e, f, g\}$ and

$$E = \Big\{\{a,b\},\{a,c\},\{a,d\},\{b,d\},\{c,d\},\{c,e\},\{e,f\},\{e,g\},\{f,g\},\{f,h\}\Big\}.$$

Matrix-Style: As a matrix with 1 if there is an edge, and 0 otherwise. (For a directed graph, it may not be symmetric).

$$G = \begin{array}{c|cccccccc} & a & b & c & d & e & f & g & h \\ \hline a & 0 & \boxed{1} & \boxed{1} & \boxed{1} & 0 & 0 & 0 & 0 \\ b & \boxed{1} & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ c & \boxed{1} & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ d & \boxed{1} & 1 & 1 & 0 & 0 & \boxed{0} & 0 & 0 \\ e & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ f & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ g & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ h & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{array} = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

sparse



Clustering on Graph?

Cluster vertex set V.

Split $V \to S, T$ $\quad T = V / S$

use edges to measure split.

$Cut(S,T) = \#$ edges in E s.t. $e \in \{v_1, v_2\}$, $v_1 \in S$, $v_2 \in T$

$Cut(S',T') = Cut(S,T) = 1$

Normalized Cut

$Vol(S)$ = # edges with at least 1 endpoint in S.

$Vol(S) = 6$ $Vol(T) = 5$

$Vol(S') = 1$ $Vol(T') = 10$

$$NCut(S,T) = \frac{Cut(S,T)}{Vol(S)} + \frac{Cut(S,T)}{Vol(T)}$$

$$NCut(S,T) = \frac{1}{6} + \frac{1}{5} \approx 0.367$$

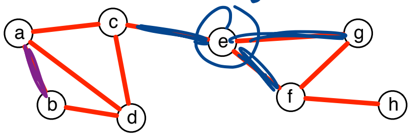
$$NCut(S',T') = \frac{1}{1} + \frac{1}{10} = 1.1$$

Goal: Find $S \subset V$, minimize $NCut(S, V/S)$

Recurse!

Laplacian Matrix

degree(e) = 3



adjacency

$$A = \begin{array}{cc} & a \\ b & \end{array}\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

degree

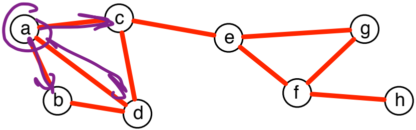
$$D = \begin{array}{c|cccccccc} & & & & & e & & & \\ \hline & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ e & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}.$$

Unnormalized Laplacian Matrix

$$L_0 = U \Lambda U^T$$

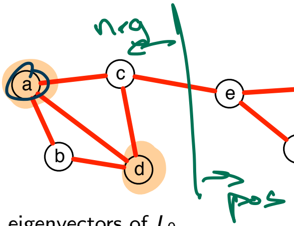
eigenvectors

eigenvalues

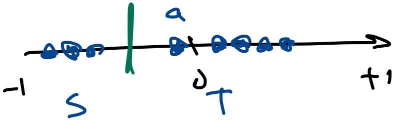


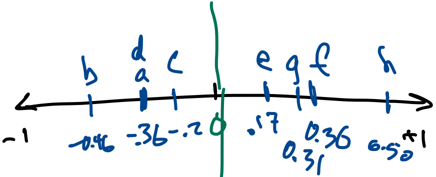
$$L_0 = D - A = \begin{pmatrix} 3 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 3 & -1 & -1 & 0 & 0 & 0 \\ -1 & -1 & -1 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 3 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}.$$

Unnormalized Laplacian Matrix



eigenvectors of L_0





λ_2

λ	0	**0.278**	1.11	2.31	3.46	4	4.82
V	$1/\sqrt{8}$	−.36	0.08	0.10	0.28	0.25	$1/\sqrt{2}$
	$1/\sqrt{8}$	−.42	0.18	0.64	−.38	0.25	0
	$1/\sqrt{8}$	−.20	−.11	0.61	0.03	−.25	0
	$1/\sqrt{8}$	−.36	0.08	0.10	0.28	0.25	$-1/\sqrt{2}$
	$1/\sqrt{8}$	0.17	−.37	0.21	−.54	−.25	0
	$1/\sqrt{8}$	0.36	−.08	−.10	−.28	0.75	0
	$1/\sqrt{8}$	0.31	−.51	−.36	−.56	0.56	0
	$1/\sqrt{8}$	0.50	0.73	0.08	0.11	0.11	0

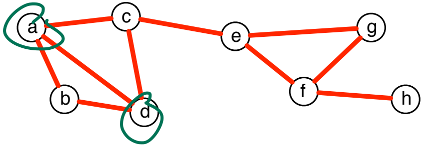
3 eigenvalues

eigen vectors

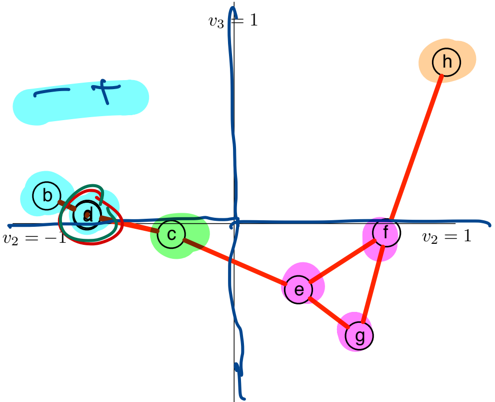
V_2

Fiedler vector | How best to map to 1 dim

Unnormalized Laplacian Matrix



λ	**0.278**	1.11	
V	−.36	0.08	*a*
	−.42	0.18	*b*
	−.20	−.11	*c*
	−.36	0.08	*d*
	0.17	−.37	*e*
	0.36	−.08	*f*
	0.31	−.51	*g*
	0.50	0.73	*h*
	v_2	v_3	



Affinity Matrix

A = Adjacency Matrix $\qquad A_{ij} = 1$ if $e_{ij} \in E$

0 o.w.

$e_{ij} \in E$ if i "follows" j

or j "follows" i

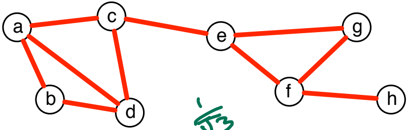
Affinity Matrix $\qquad A_{ij} \in [0, 1]$

not connected $\qquad$ Everything connected

$$A_{ij} = \mathrm{sim}(x_i, x_j)$$

$$L_0 = D - A \qquad D_{ii} = \sum_{j=1}^{n} A_{ij}$$

Laplacian Matrix

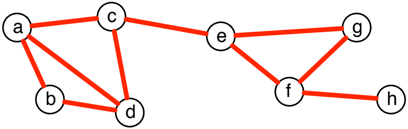


$= \frac{1}{\sqrt{3}}$ $= \frac{1}{\sqrt{2}}$

$$D^{-1/2} = \begin{pmatrix} 0.577 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.707 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.577 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.577 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.577 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.577 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.707 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

$= \frac{1}{\sqrt{1}}$

Laplacian Matrix

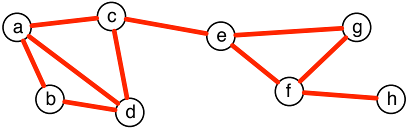


normalized Laplacian

$L = I - D^{-1/2}AD^{-1/2} =$ $D^{-1/2} L_0 D^{-1/2}$

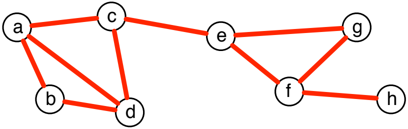
$$\begin{pmatrix} 1 & -0.408 & -0.333 & -0.333 & 0 & 0 & 0 & 0 \\ -0.408 & 1 & 0 & -0.408 & 0 & 0 & 0 & 0 \\ -0.333 & 0 & 1 & -0.333 & -0.333 & 0 & 0 & 0 \\ -0.333 & -0.408 & -0.333 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.333 & 0 & 1 & -0.333 & -0.408 & 0 \\ 0 & 0 & 0 & 0 & -0.333 & 1 & -0.408 & -0.577 \\ 0 & 0 & 0 & 0 & -0.408 & -0.408 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.577 & 0 & 1 \end{pmatrix}.$$

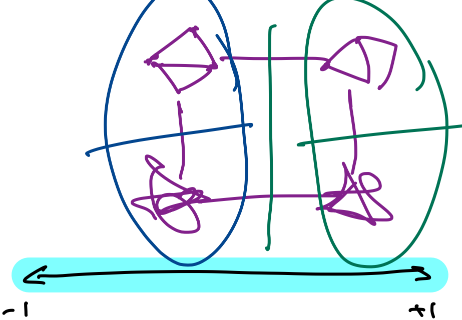
Laplacian Matrix



eigenvectors of L

Laplacian Matrix

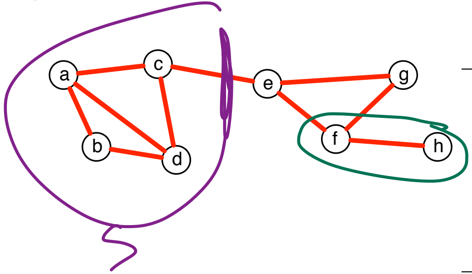




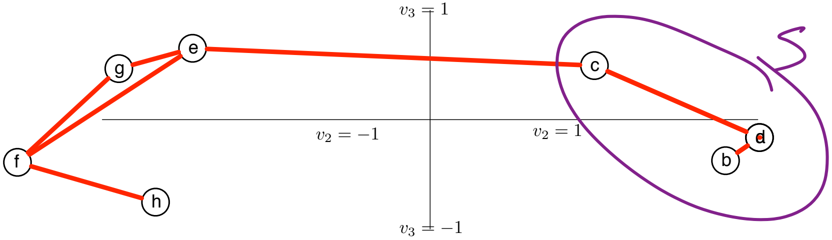
eigenvectors of L

λ	0	**0.125**	0.724	1.00	1.33	1.42	1.66	1.73
V	−.39	0.38	−.09	0.00	0.71	0.26	−.32	0.16
	−.32	0.36	−.27	0.50	0.00	−.51	0.38	−.18
	−.39	0.18	0.36	−.61	0.00	0.03	0.47	−.29
	−.39	0.38	−.09	0.00	−.71	0.26	−.32	0.16
	−.39	−.28	0.48	0.00	0.00	−.57	0.31	0.33
	−.39	−.48	−.29	0.00	0.00	0.05	−.31	−.65
	−.31	−.36	0.27	0.50	0.00	0.51	0.38	−.18
	−.22	−.32	−.61	−.35	0.00	−.07	0.27	0.51

Laplacian Matrix



λ	**0.125**	0.724	
V	0.38	$-.09$	*a*
	0.36	$-.27$	*b*
	0.18	0.36	*c*
	0.38	$-.09$	*d*
	$-.28$	0.48	*e*
	$-.48$	$-.29$	*f*
	$-.36$	0.27	*g*
	$-.32$	$-.61$	*h*
	v_2	v_3	



$\longleftrightarrow$
$v_2 \cdot \frac{1}{\lambda_2}$