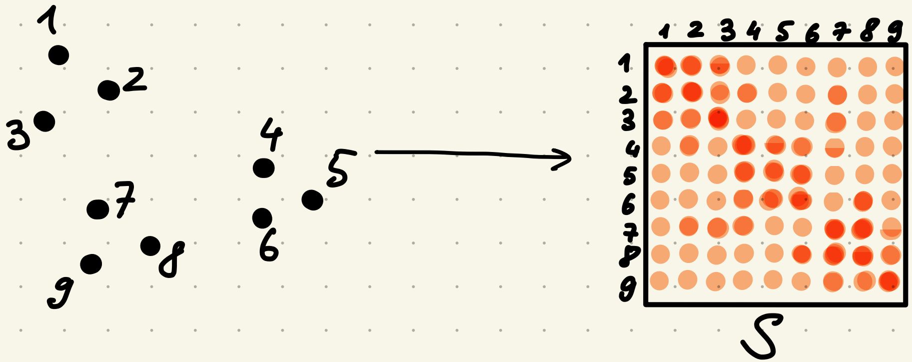
Spectral Clustering

Start with pairwise similarities between observations: $S \in \mathbb{R}^{n \times n}$ where S_{ij} represents similarity between x_i and x_j.

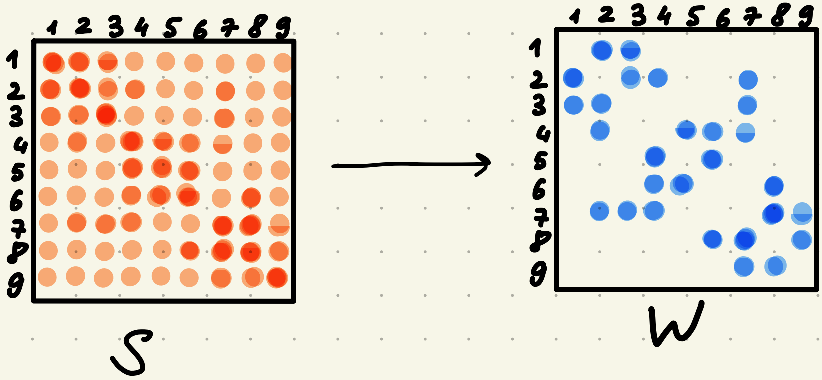
The elements in S are $S_{ij} = S_{ji} \geq 0$.



<u>Examples</u> : • gaussian kernel $S_{ij} = e^{-\gamma \|x_i - x_j\|^2}$

• general kernels $S_{ij} = K(x_i, x_j)$

Use S to derive ==weighted adjacency== matrix W with $W_{ii} = 0$ and $W_{ij} \geq 0$.



<u>Examples</u>: • Fully connected $W_{ij} = S_{ij}$

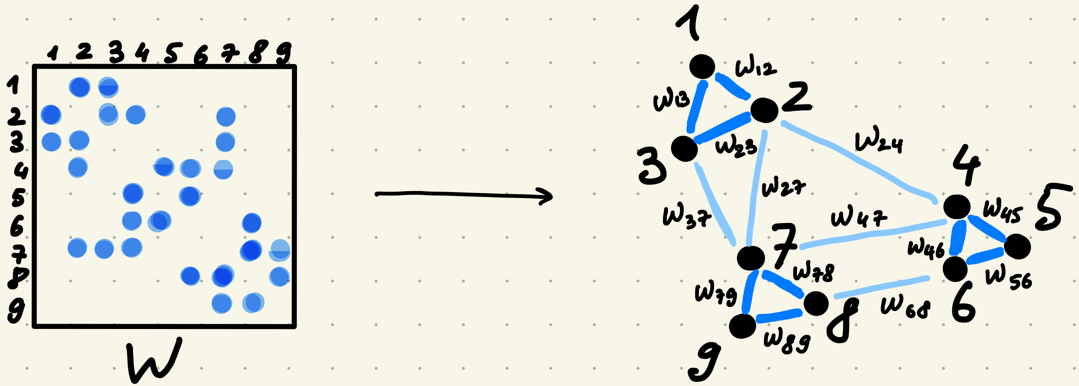
- Soft threshold $W_{ij} = (S_{ij} - \lambda)_+$
- The ε-neighborhood $W_{ij} = I(S_{ij} > \varepsilon)$
- Mutual K-nearest neighbor graph

$W_{ij} = S_{ij}$ if i is among K the most similar data point to j and vice versa.

View data as weighted graph

$$\mathcal{G} = (\underset{\text{vertices}}{V}, \underset{\text{edges}}{E}, \underset{\text{weights}}{W})$$

- vertices are data points.
- an edge connects i and j if $w_{ij} > 0$.



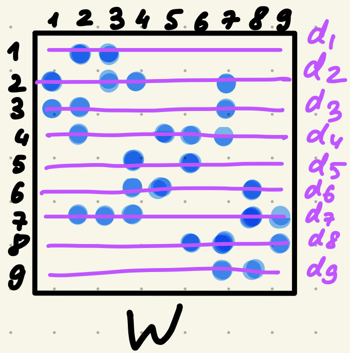
<u>Goal</u>: partition the graph so that edges within clusters have large weights and edges between clusters have small weights.

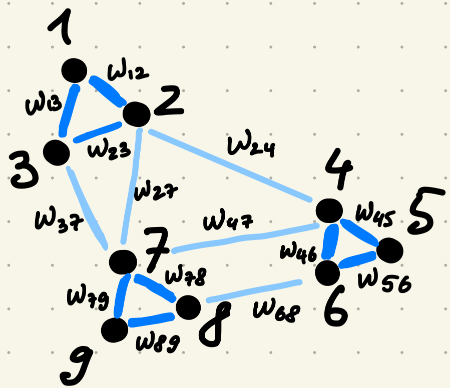
Graph Laplacian

Denote by $d_i = \sum_{j=1}^{n} w_{ij}$ the **degree** of vertex i and by $D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}$ the degree matrix.

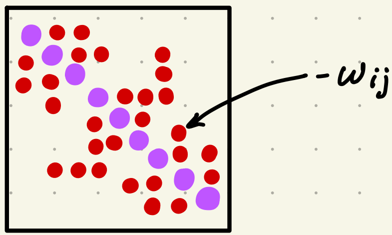
Graph Laplacian is $L = D - W$.

Example





$$\begin{cases} d_1 = w_{12} + w_{13} \\ d_2 = w_{12} + w_{23} + w_{24} + w_{27} \\ \quad \cdots \\ d_9 = w_{79} + w_{89} \end{cases}$$

$L =$ 

Properties of Laplacian

- Let $f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \in \mathbb{R}^n$ then $f^T L f = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (f_i - f_j)^2$

$$f^T L f = f^T (D - W) f = \sum_{i=1}^{n} d_i f_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} f_i f_j =$$

$$= \frac{1}{2} \left(\sum_{i=1}^{n} d_i f_i^2 - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} f_i f_j + \sum_{j=1}^{n} d_j f_j^2 \right) =$$

$$= \frac{1}{2} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} f_i^2 - 2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} f_i f_j + \sum_{j=1}^{n} \sum_{i=1}^{n} \overset{w_{ij}}{\overset{\shortparallel}{w_{ji}}} f_j^2 \right) =$$

$$= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (f_i^2 - 2 f_i f_j + f_j^2) = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (f_i - f_j)^2$$

- L is PSD.

- The smallest eigenvalue is 0 with e. vector $\mathbb{1}_n$

$$L \cdot 1_n = (D - W) \cdot 1_n = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} - \begin{pmatrix} \sum_{j=1}^{n} w_{1j} \\ \cdots \\ \sum_{j=1}^{n} w_{nj} \end{pmatrix} = 0$$

Spectral Clustering Algorithm

<u>Input</u> Similarity matrix $S \in \mathbb{R}^{n \times n}$,
number of clusters K.

<u>Step 1</u> Compute weighted adjacency matrix W

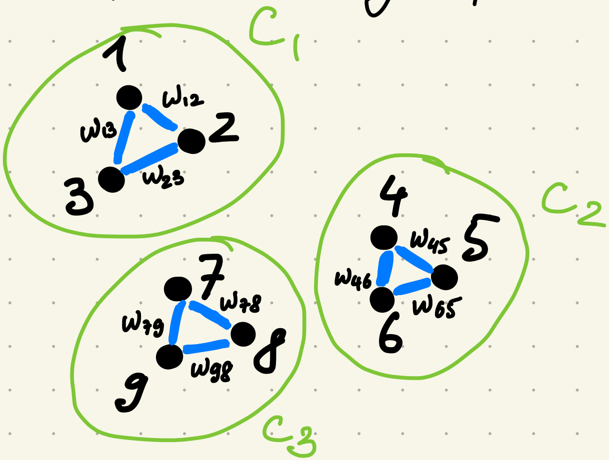
<u>Step 2</u> Compute Laplacian $L = D - W$

<u>Step 3</u> Pick K smallest eigenvectors of L.
$u_1 \dots u_K$ and construct $U = (u_1 \dots u_K)$

<u>Step 4</u> Cluster <u>rows</u> of U (using, for example, K-means)

① If graph has K connected components

Denote the components by $C_1 .. C_K$ and $n_k = |C_k|$.



- L has block-diagonal structure.

(after a proper permutation of columns & rows)

$$W = \begin{pmatrix} \boxed{W_1} & & \\ & \ddots & \\ & & \boxed{W_K} \end{pmatrix} \quad \text{then} \quad d_i = \sum_{j \in C_i} w_{ij} \quad \text{and}$$

$$L = \begin{pmatrix} \boxed{L_1} & & \\ & \ddots & \\ & & \boxed{L_K} \end{pmatrix} \quad \text{where } L_i \text{ is Laplacian for } C_i.$$

- Multiplicity of the zero eigenvalue is K.

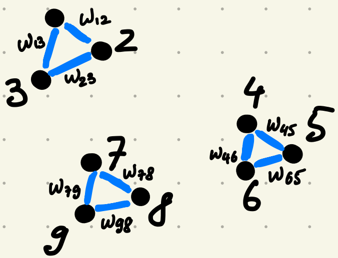
Since $L_k \mathbb{1}_{n_k} = 0$, then $L = \begin{pmatrix} \boxed{L_1} & & \\ & \boxed{L_k} & \\ & & \boxed{L_n} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ \mathbb{1}_{n_k} \\ \vdots \\ 0 \end{pmatrix} = 0$

Thus $\begin{pmatrix} \mathbb{1}_{n_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ \mathbb{1}_{n_2} \\ \vdots \\ 0 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ \vdots \\ \mathbb{1}_{n_k} \end{pmatrix}$ are all e.vectors corresponding to zero.

- Eigenvector f corresponding to 0 is such that $f_i = f_j$ if i, j are in the same cluster.

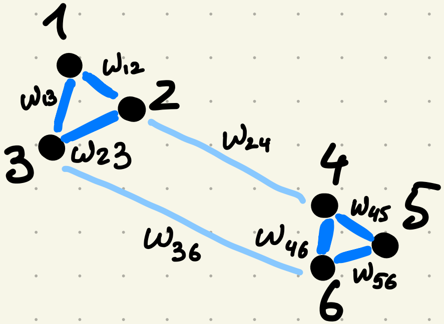
$$f = \beta_1 \cdot \begin{pmatrix} \mathbb{1}_{n_1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \beta_2 \begin{pmatrix} 0 \\ \mathbb{1}_{n_2} \\ \vdots \\ 0 \end{pmatrix} + \ldots + \beta_k \begin{pmatrix} 0 \\ \vdots \\ \mathbb{1}_{n_k} \end{pmatrix} = \begin{pmatrix} \boxed{\beta_1} \\ \vdots \\ \boxed{\beta_k} \end{pmatrix} \begin{matrix} \}n_1 \\ \\ \}n_k \end{matrix}$$

Also $f^T L f = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (f_i - f_j)^2 = 0$ if $f_i = f_j$ for $w_{ij} > 0$



$$\longrightarrow L = \begin{pmatrix} \boxed{L_1} & & \\ & \boxed{L_2} & \\ & & \boxed{L_3} \end{pmatrix} \longrightarrow U = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{aligned} C_1 &= \{1, 2, 3\} \\ C_2 &= \{4, 5, 6\} \\ C_3 &= \{7, 8, 9\} \end{aligned}$$

② If graph is connected loosely



- L is not block-diagonal (but almost)

$$L = \begin{pmatrix} \boxed{L_1} & * \quad * \\ * \quad * & \boxed{L_2} \end{pmatrix}$$

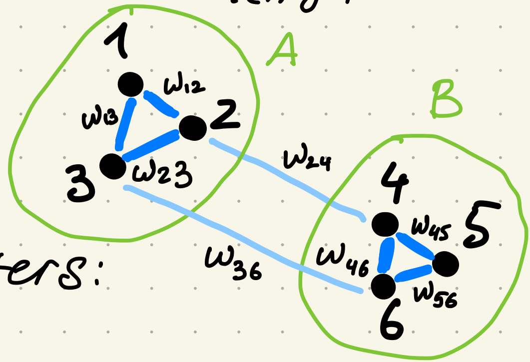
- The smallest eigenvector is $\mathbb{1}$, but the other eigenvectors can provide more insights.

graph - cut point of view (assume $K=2$)

Given cluster A define

n_A = number of vertices in $A = |A|$

$\mathrm{vol}(A) = \sum_{i \in A} d_i$ = all weight attached $A = \sum_{i\in A} \sum_{j=1}^{n} w_{ij}$



Example $A = \{1,2,3\}$ $n_A = 3$

$\mathrm{vol}(A) = w_{13} + w_{12} + w_{23} + w_{24}$

Define graph cut for A, B clusters:

$$\mathrm{cut}(A,B) = \sum_{i\in A} \sum_{j \in B} w_{ij}$$

ratio: $\quad r\mathrm{cut}(A,B) = \mathrm{cut}(A,B)\left(\frac{1}{|A|} + \frac{1}{|B|}\right)$

normalized: $\quad n\mathrm{cut}(A,B) = \mathrm{cut}(A,B)\left(\frac{1}{\mathrm{vol}(A)} + \frac{1}{\mathrm{vol}(B)}\right)$

Example: $\mathrm{cut}(A,B) = w_{24} + w_{36}$

Consider $f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$ with $f_i = \begin{cases} \sqrt{\frac{n_B}{n_A}} & \text{if } i \in A \\ -\sqrt{\frac{n_A}{n_B}} & \text{if } i \in B \end{cases}$

Denote this set of f vectors by $\mathcal{F}(A,B)$.

Then minimizing $rcut(A,B)$ is equivalent to

$$\underset{A,B}{\text{minimize}}\ f^T L f \quad \text{s.t.} \quad f \in \mathcal{F}(A,B) \text{ and } \quad (*)$$
$$f^T \mathbb{1} = 0, \quad \|f\|^2 = n$$

- $f^T L f = n \cdot rcut(A,B)$

$$f^T L f = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (f_i - f_j)^2 w_{ij} =$$

$$\frac{1}{2}\left[\sum_{i \in A} \sum_{j \in A} (f_i - f_j)^2 w_{ij} + \sum_{i \in A} \sum_{j \in B} \ldots + \sum_{i \in B} \sum_{j \in A} \ldots + \sum_{i \in B} \sum_{j \in B} \ldots\right] =$$

$$0 + \frac{1}{2} \sum_{i \in A} \sum_{j \in B} w_{ij} \left(\sqrt{\frac{n_B}{n_A}} + \sqrt{\frac{n_A}{n_B}}\right)^2 + \frac{1}{2} \sum_{i \in B} \sum_{j \in A} w_{ij} \left(\sqrt{\frac{n_A}{n_B}} + \sqrt{\frac{n_B}{n_A}}\right)^2 + 0 =$$

$$\sum_{i \in A} \sum_{j \in B} w_{ij} \frac{(n_B + n_A)^2}{n_A n_B} = cut(A,B) \cdot \frac{n^2}{n_A \cdot n_B} = n\, cut(A,B) \left(\frac{1}{n_A} + \frac{1}{n_B}\right)$$

- $f^T 1 = 0$

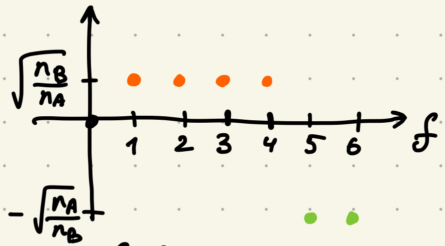
$$f^T 1 = \sum_{i \in A} \sqrt{\frac{n_B}{n_A}} + \sum_{i \in B} \left(-\sqrt{\frac{n_A}{n_B}}\right) = n_A \sqrt{\frac{n_B}{n_A}} - n_B \sqrt{\frac{n_A}{n_B}} = \sqrt{n_A n_B} - \sqrt{n_B n_A} = 0$$

- $\|f\|^2 = n$

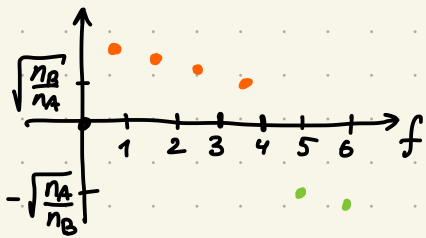
$$\|f\|^2 = \sum_{i=1}^{n} f_i^2 = \sum_{i \in A} \frac{n_B}{n_A} + \sum_{i \in B} \frac{n_A}{n_B} = n_B + n_A = n$$

Problem (*) is NP-hard. Relaxation:

$$\underset{f \in \mathbb{R}^n}{\text{minimize}}\ f^T L f \quad \text{s.t.} \quad f^T 1 = 0,\ \|f\|^2 = n \quad (**)$$



R Cut



Relaxation

- The new problem will search for the second smallest eigenvector of L.

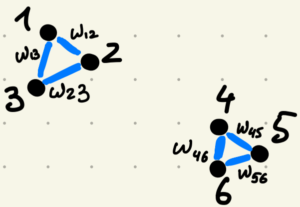
$\cancel{f^T 1 = 0}$, then f is the eigenvector corresponding to the smallest eigenvalue of L

$f^T 1 = 0$, then f is eigenvector corresponding to the second smallest eigenvalue of L.

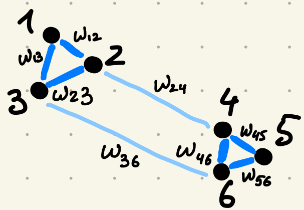
- The second eigenvector can be used to cluster points into A and B, for example,

$$i \in A \text{ if } f_i \geq 0 \qquad\qquad i \in B \text{ if } f_i < 0$$

- Alternatively, one can cluster f_i in two clusters. This is exactly what Spectral clustering would do.



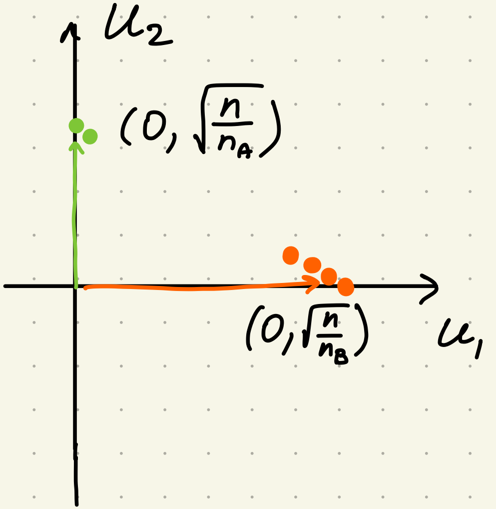
$$L = \begin{pmatrix} L_1 & \\ & L_2 \end{pmatrix}$$



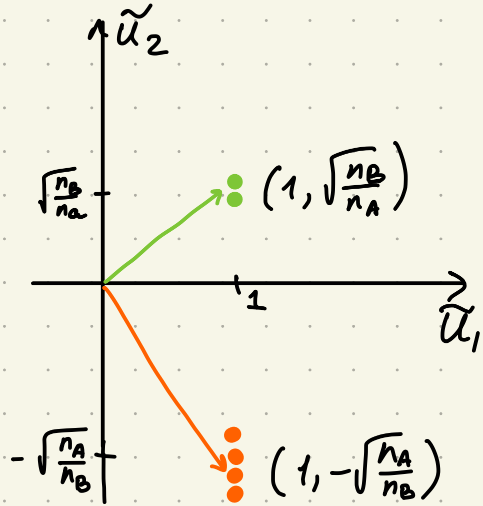
$$\tilde{L} = \begin{pmatrix} L_1 & * \\ * & L_2 \end{pmatrix}$$

rcut Solution

$$u = \sqrt{n} \cdot \begin{pmatrix} 0 & \frac{1}{\sqrt{n_A}} \\ \frac{1}{\sqrt{n_B}} & 0 \end{pmatrix} \quad \xleftrightarrow[\text{same subspace (Davis-Kahan)}]{\text{span the}} \quad \tilde{u} \approx \begin{pmatrix} 1 & \sqrt{\frac{n_B}{n_A}} \\ 1 & -\sqrt{\frac{n_A}{n_B}} \end{pmatrix}$$



rotate



Comments:

- For $k>2$ the relaxation of rcut could also be built.

(See Tutorial on Spectral clustering by Ulrike von Luxburg)

- Popular modification of Spectral Clustering

Normalized Laplacian $L' = D^{-1}L = I - D^{-1}W$

Second eigenvector corresponds to the ncut(A,B).