Graph

A graph can be defined as an ordered pair

$$G = (V, E)$$

V is a set of vertices (nodes, points)

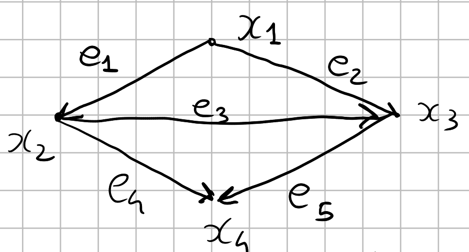
$E = \{(x,y) : x, y \in V \text{ and } x \neq y\}$ a set of edges which are unordered pair of vertices

G is an unordered graph

If we consider the edges ordered (x,y) $\quad x \to y$

G is a directed graph

The incidence matrix of an ordered graph with n nodes and m edges : $A \in \mathbb{R}^{m \times n}$



$$A = \begin{array}{c} \\ \\ \\ \\ \\ \\ \end{array}\begin{array}{c} \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \end{array} \begin{array}{c} \\ e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{array} \qquad B^{(o)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

L Laplacian matrix of ordered/unordered graph

$$L = A^T A = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{bmatrix} = D - B^{(u)}$$

D : diagonal degree matrix

degree : number of edges connecting each node

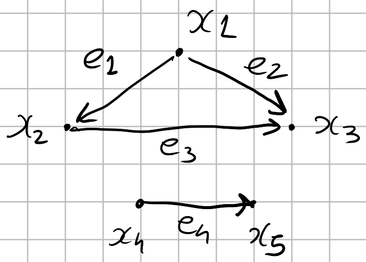
$B^{(u)}$: connectivity matrix

unordred graph, no self loops

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

Connectivity matrix for the unordered graph

$\mathbb{1} \in \mathbb{R}^n \qquad \mathbb{1} = (1, 1, 1, \dots, 1)^T \qquad \mathbb{1} \in \ker(A) \qquad A\mathbb{1} = 0$



$$A = \begin{array}{c} \\ \\ \\ \\ \end{array}\begin{array}{c c} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} & \\ \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} & \begin{array}{c} e_1 \\ e_2 \\ e_3 \\ e_4 \end{array} \end{array}$$

$$\ker(A) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$$

The vectors generating the $\ker(A)$ are also in the $\ker(L)$

$A \in \mathbb{R}^{m \times n}$ $\quad L = A^T A \quad$ $A^T \in \mathbb{R}^{n \times m}$

$x_0 \in \mathbb{R}^n \quad x_0 \in \ker(A) \quad A x_0 = 0_{\mathbb{R}^m} \in \mathbb{R}^m \quad L \in \mathbb{R}^{n \times n}$

$$L x_0 = A^T A x_0 = A^T 0_{\mathbb{R}^m} = 0_{\mathbb{R}^n}$$

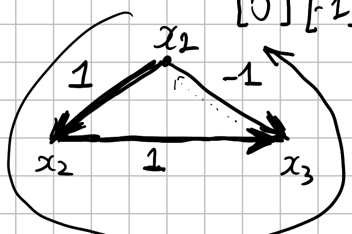
$A \in \mathbb{R}^{5 \times 4}$ (the first A introduced)

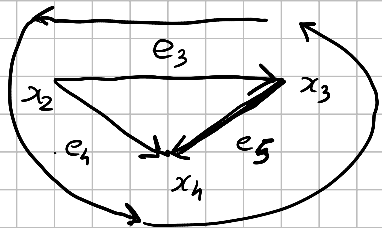
$\dim(\ker(A)) = 1$ $\qquad \dim(\text{Im}(A)) = 4 - 1 = 3 = \text{rank}(A)$

$\text{rank}(A^T) = 3$

$\dim(\ker(A^T)) = 5 - \text{rank}(A^T) = 5 - 3 = 2$

$$\ker(A^T) = \text{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ -1 \end{bmatrix} \right\} \qquad A^T \in \mathbb{R}^{4 \times 3} = \mathbb{R}^{n \times m}$$

$\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ 

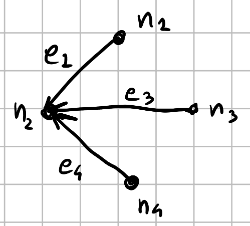
$\begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$ 

A^T is the Kirchoff matrix representing the currents balance at the nodes

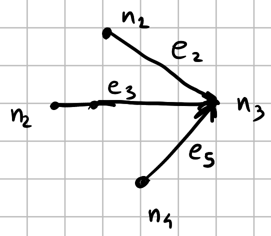
The $\ker(A^T)$ are the current flows self balanced

The number of independent rows of A^T is the number of trees covering the graph

$$\begin{bmatrix} 1 \\ 0 \\ -1 \\ -1 \\ 0 \end{bmatrix}$$



$$\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$$



Euler equation

Number of nodes − Number of edges + Number of loops = 1

$$4 - 5 + 2 = 1$$

L Laplacian matrix

1) L symmetric positive semidefinite

$$L = A^T A \qquad [\text{symmetric}]$$

$\mathbb{1} \in \mathbb{R}^n$ is eigenvector corresponding to $\lambda = 0$ [semi-definite]

2) The multiplicity of the eigenvalue $\lambda = 0$ is the number of connected components of the graph

---

Similarity graph

Given a set of points $x_1, \dots, x_n$ and a similarity notion $S_{ij} > 0$ between all pairs of data points x_i and x_j.

The intuitive idea of clustering is to divide the data points into several groups such that the points in the same group are similar.

Weighted graph

w_{ij} weight for the edge connecting x_i and x_j

$w_{ij} = 0$ $\qquad$ x_i is not connected to x_j

degree of a vertex $x_i \in V$

$$d_i = \sum_{j=1}^{n} w_{ij} \qquad\qquad D = \operatorname{diag}(d)$$

Let $A \subset V$ $\qquad$ A is a subset of nodes in V

$$\bar{A} = V \setminus A \qquad \text{complement of } A$$

$f = \mathbb{1}_A$ vector of 1 for the components f_i such that $x_i \in A$ and $f_i = 0$ if not.

$\mathbb{1}_A$ is the indicator vector of A

$|A|$ number of vertices in A

$Vol(A) = \sum_{i \in I_A} d_i$ $\qquad I_A = \{ i \in [1, \dots, n] : x_i \in A \}$

A possible similarity function can be

$$S_{ij} = S(x_i, x_j) = \exp\left(- \frac{\| x_i - x_j \|^2}{2\sigma^2} \right)$$

This similarity function is defining a fully connected graph

k-nearst neighbour graph

In the <u>connectivity matrix</u> we keep only the k largest wieghts in each row

symmetric connectivity → or $\quad S_{ij} = \max\{S_{ij}, S_{ji}\}$

symmetric connectivity → and $\quad S_{ij} = \min\{S_{ij}, S_{ji}\}$

$L = D - W$ $\qquad$ symmetric positive semi-definite matrix

Properties

1) For any vector $f \in \mathbb{R}^n$ we have

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

2) L is symmetric and positive semi-definite

3) The smallest eigenvalue of L is $\lambda = 0$, the corresponding eigenvector is the constant vector $\mathbb{1}_V$

Proof

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

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

$$= \frac{1}{2}\left(\sum_{i=1}^{n} d_{ii} f_i^2 - 2\sum_{i,j=1}^{n} w_{ij} f_i f_j + \sum_{j=1}^{n} d_{jj} f_j^2 \right) \qquad d_{ii} = \sum_{j=1}^{n} w_{ij}$$

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

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

$$= \frac{1}{2} \sum_{i,j=1}^{n} w_{ij} (f_i - f_j)^2 \geq 0$$

2) D, W symmetric $\qquad L: \quad f^T L f \geq 0$

3) $f = \mathbb{1}_V$ eigenvector of L corresponding to $\lambda = 0$

The indicator vectors of the connected components are eigenvectors of $\lambda = 0$

$f^T L f = 0 \qquad f$ is the eigenvector of $\lambda = 0$

$$R_\lambda(f) = \frac{f^T L f}{f^T f} \qquad \nabla R_\lambda(f) = \frac{2}{f^T f}\left[L f - R_\lambda(f) f \right]$$

$\nabla R_\lambda(f) = 0 \qquad L f = R_\lambda(f)\, f = \lambda_f\, f$

$$\frac{\partial f^T L f}{\partial f_i} = e_i^T L f + f^T L e_i = 2 e_i^T L f$$

$\nabla (f^T L f) = 2 L f \qquad L f = 0 \qquad L f = 0\, f$

$\uparrow$

$f^T L f$ is a C^∞ function of f

f is a minimum point if $\nabla (f^T L f) = 0 \qquad L f = 0$

Spectra Clustering

Similarity function $\Rightarrow W$ symmetric connectivity matrix of the graph

$W \Rightarrow D$

$$L = D - W$$

Compute the M smallest eigenvalues / eigenvectors

$u_m \quad m = 1, \dots, M$

Let $U \in \mathbb{R}^{n \times M}$ the matrix with columns $u_m \quad m = 1, \dots, M$

For $i = 1, \dots, n$ let $y_i = U(i, :) \qquad y_i \in \mathbb{R}^M$

Cluster the points y_i $i = 1, \dots, n$ with the k-means algorithm in k clusters

$$C_1, \dots, C_k$$

C_j is a set of points in $\mathbb{R}^M$

A_j of points in $\mathbb{R}^n$ corresponding to the clusters C_j

A_j $j = 1, \dots, k$ is one of my clusters defined by the spectral clustering

---

$$L \in \mathbb{R}^{n \times n}$$

$$L_M = U^T L U \in \mathbb{R}^{M \times M}$$