AGRUPAMIENTO ESPECTRAL

3 pasos básicos en agrupamiento espectral:

1. Construir un grafo de semejanza
   - → Varios criterios para construir el grafo { → Umbral, → KNN }
   - → Construir la matriz de adyacencias

2. Proyectar los datos a un espacio de menor dimensión
   - → Utilizando eigenvectores de la matriz Laplaciana { diferentes matrices Laplacianas

3. Agrupar los datos
   - ↳ utilizaremos K-means sobre el espacio transformado

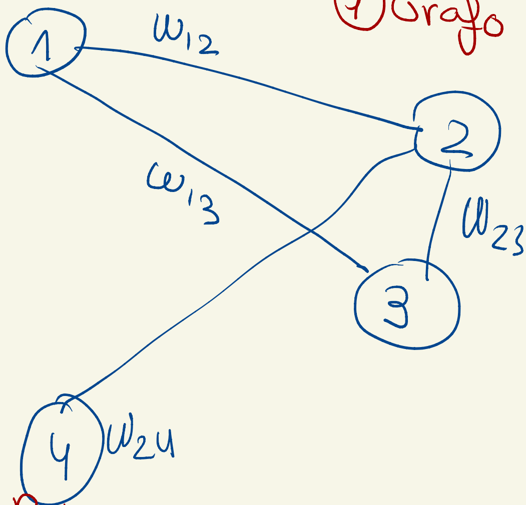
Como hacemos el proceso

la matriz de adyacencia es una matriz simétrica de $n \times n$

2 Matriz de adyacencia

	1	2	3	4
1	0	w_{21}	w_{31}	0
2	w_{12}	0	w_{23}	w_{24}
3	w_{13}	w_{23}	0	0
4	0	w_{24}	0	0

1 Grafo



3 Construimos la laplaciana

Una vez que obtenemos A, sacamos

$$d_i = \sum_{\forall (i,j) \in E} w_{ij}$$

obtenemos matriz laplaciana

$$L = D - A$$

donde D es la matriz de grados

$$d_1 = w_{12} + w_{13}$$

$$d_2 = w_{12} + w_{23} + w_{24}$$

$$d_3 = w_{13} + w_{23}$$

$$d_4 = w_{24}$$

$$L_{ij} = \begin{cases} d_i & \text{si } i=j \\ -w_{ij} & \text{si } (i,j) \text{ es un arista} \\ 0 & \text{si no hay arista} \end{cases}$$

|| la matriz de grados es diagonal

$L =$

	1	2	3	4
1	d_1	$-w_{12}$	$-w_{13}$	0
2	$-w_{12}$	d_2	$-w_{23}$	w_{24}
3	$-w_{13}$	$-w_{23}$	d_3	0
4	0	$-w_{24}$	0	d_4

Propiedades de la matriz Laplaciana

- $\mathcal{L}(G)$ es simétrica $(\mathcal{L}(G) = \mathcal{L}^{T}(G))$
- Los eigenvalores de L son reales no negativos
- $\bar{x} = [1, \dots 1]$ es un eigenvector de $\mathcal{L}(G)$

  La suma de los renglones es 0

  por lo que si hago $\boxed{\mathcal{L} \cdot \bar{x} = 0}$ es la suma

$\lambda = 0$ es un autovalor

$\lambda = 0$ es un autovalor

la multiplicidad del 0 como eigenvalor es igual al número de componentes conexas de la red

$\det(A - \lambda I) = 0$

El segundo autovalor más pequeño es la conectividad del grafo (no nulo)

Ese autovalor es mayor que 0 solo si es un grafo conexo

4 Eigenvalor y Eigenvector del Grafo

Para una matriz A, λ es un eigenvalor de A si para algún vector v, $\boxed{Av = \lambda v}$ diferente de 0 *
y es el eigenvector de A correspondiente a λ

Para un grafo G con n nodos, su adyacencia tiene n eigenvalores $\{ \mu_1, \mu_2 \ldots \mu_n \}$ donde

$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n$

y n correspondientes eigenvectores (vectores propios)

$\{ x_1, x_2, x_n \}$

$$\boxed{\mu_1 \geq \mu_2 \geq \ldots \geq \mu_n}$$

* En los ejemplos sencillos, λ se puede aproximar mucho a 0.

El grafo Laplaciano de G, L_G, tb tiene eigenvalores $\{\lambda_1, \lambda_2 \dots \lambda_n\}$, donde

$$0 \le \lambda_1 \le \lambda_2 \le \dots \lambda_n$$

y eigenvectores

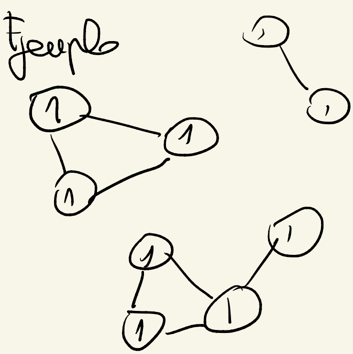
$$\{v_1, v_2, \dots, v_n\}$$

Los eigenvalores revelan propiedades globales de un graf que no son obvias en el mismo

$$\boxed{0 = \lambda_1 \le \lambda_2 \le \dots \le \lambda_n}$$

Si el eigenvalor de L_G es 0 con k diferentes vectores, entonces, G tiene k componentes conectados

Ejemplo



L_G has $0 = \lambda_1 = \lambda_2 = \lambda_3$ and $\lambda_4 > 0$

Si el grafo está conectado y $\lambda_2 > 0$ λ_2 es la conectividad algebraica

3 ¿Como se dividen los datos?

Utilizamos varios eigenvectores

x_1, x_2, x_3 asociados a los 3 eigenvalores más pequeños * en los ejemplos fáciles serán igual a 0

a cada vertice de la red (punto)

$i \in V(G) = (a_1, a_2, a_3)$

a_1: i-esima coordenada de x_1

a_2: i-ésima coordenada de x_2

a_3: i-ésima coordenada de x_3

i: recordamos que nuestra matriz de adyacencia era de $n \times n$ por tanto cada eigenvector es de $1 \times n$

K-means en el nuevo espacio

Por tanto, ahora, cada punto de mi espacio viene definido por 3 coordenadas, hemos disminuido la dimensión, trasladandolo a otro espacio que es fácil de diferenciar.