# Aprendizaje No Supervisado

## Maestría en Ciencia de Datos

Lucas Fernández Piana

Primavera 2022

Universidad de San Andrés

# CLUSTERS ESPECTRALES

# GRAFOS

# INTRO

Dado un conjunto de datos $\{x_1, \ldots, x_n\}$ y una similaridad $(s_{ij})_{i,j}$ el objetivo intuitivo de agrupar la información en clusters es que los puntos en el mismo grupo sean similares entre sí y puntos en diferentes grupos sean menos similares.

# INTRO

Dado un conjunto de datos $\{x_1, \ldots, x_n\}$ y una similaridad $(s_{ij})_{i,j}$ el objetivo intuitivo de agrupar la información en clusters es que los puntos en el mismo grupo sean similares entre sí y puntos en diferentes grupos sean menos similares.

Si la información con la que contamos viene dada solamente por la similaridad, una forma de representar los datos es un grafo de similaridades $G = (V, E)$ donde

- Cada vértice en el grafo representa un dato.
- Dos vértices están conectados por una arista si la similaridad es mayor que cierto umbral y el peso que tiene la unión esta dado por la similaridad.

# INTRO

Desde este nuevo punto de vista, el problema de encontrar clusters se puede reformular: encontrar una partición del grafo tal que los vértices de distintos grupo estén conectados por aristas débiles y los vértices dentro del mismo grupo estén conectados por aristas fuertes.

# FORMALIDADES

Un grafo no dirigido es un par $G = (V, E)$ donde

- $V$ es un conjunto finito, vértices.
- $E$ es un conjunto de pares no ordenados de vértices, aristas.

# FORMALIDADES

Un grafo no dirigido es un par $G = (V, E)$ donde

- $V$ es un conjunto finito, vértices.
- $E$ es un conjunto de pares no ordenados de vértices, aristas.

**Ejemplo:** $G = (V, E)$

- $V = \{1, 2, 3, 4, 5, 6, 7\}$.
- $E = \{(1, 2), (2, 4), (2, 7), (3, 5), (3, 6), (4, 7), (5, 6), (6, 7)\}$.

Sea $G = (V, E)$ un grafo no dirigido con vertices $V = \{v_1, ..., v_n\}$ diremos que es un grafo pesado si las aristas tienen asociado un peso $w_{ij} \geq 0$.

Sea $G = (V, E)$ un grafo no dirigido con vertices $V = \{v_1, ..., v_n\}$ diremos que es un grafo pesado si las aristas tienen asociado un peso $w_{ij} \geq 0$.

Los pesos podemos escribirlos en una matriz que se llama matriz de adyacencia pesada $W = (w_{ij})_{ij}$ $1 \leq i, j \leq n$.

Si $w_{ij} = 0$ entonces los vértices $v_i$ y $v_j$ no están conectados. Observar que la matriz $W$ es simétrica $w_{ij} = w_{ji}$.

# FORMALIDADES

El grado de un vértice $v_i$ se define como $d_i = \sum_{j=1}^{n} w_{ij}$.

Llamamos matriz de grados a la matriz $D = diag(d_1, \ldots, d_n)$.

# FORMALIDADES

El grado de un vértice $v_i$ se define como $d_i = \sum_{j=1}^{n} w_{ij}$.

Llamamos matriz de grados a la matriz $D = diag(d_1, \ldots, d_n)$.

Dado un subconjunto de vértices $A \subset V$ tenemos dos médidas posibles para medir el “tamaño”de $A$

- $|A| =$ número de vértices en A.
- $Vol(A) = \sum_{i \in A} d_i$.

# FORMALIDADES

Decimos que un subconjunto $A \subset V$ del grado es conexo si cualquier para de vértices pueden unirse mediante un camino. Es decir, si $a, b \ \in A$, existen $v_1 = a, v_2, \ldots, v_m = b$ vértices en $A$ tal que $w_{ii+1} > 0$.

Decimos que un subconjunto $A \subset V$ del grado es conexo si cualquier para de vértices pueden unirse mediante un camino. Es decir, si $a, b \ \in A$, existen $v_1 = a, v_2, \ldots, v_m = b$ vértices en $A$ tal que $w_{ii+1} > 0$.

Un subconjunto $A \subset V$ decimos que es una componente conexa si no hay conexiones entre los vértices de $A$ y los vértices de $V - A = \{v \in V : \ v \notin A\} = \overline{A}$.

# FORMALIDADES

Decimos que un subconjunto $A \subset V$ del grado es conexo si cualquier para de vértices pueden unirse mediante un camino. Es decir, si $a, b \;\in A$, existen $v_1 = a, v_2, \ldots, v_m = b$ vértices en $A$ tal que $w_{ii+1} > 0$.

Un subconjunto $A \subset V$ decimos que es una componente conexa si no hay conexiones entre los vértices de $A$ y los vértices de $V - A = \{v \in V : \; v \notin A\} = \overline{A}$.

Observar que las componentes conexas de un grafo forman un partición del mismo.

Existen varios tipos de construcciones en la literatura para transformar un conjunto de datos $\{x_1, \ldots, x_n\}$ con similaridad $s_{ij}$ o distancias $d_{ij}$ en un grafo.

# GRAFOS DE SIMILARIDAD

Existen varios tipos de construcciones en la literatura para transformar un conjunto de datos $\{x_1, \ldots, x_n\}$ con similaridad $s_{ij}$ o distancias $d_{ij}$ en un grafo.

El objetivo es modelar la relación local entre los datos que son vecinos utilizando las similaridades para construir la matriz de pesos.

Veremos a continuación …

# $\epsilon$-VECINDAD

También conocido en inglés como $\epsilon-$ neighborhood graph, para armar el grafo conectar los vértices cuyas distancias sean menores que el umbral $\epsilon$.

Como se supone que las distancias entre los puntos conectados están más o menos en la misma escala, agregar un peso a las aristas no incorporaría más información.

Normalmente, se concidera a este grado sin pesos.

# $k$-VECINOS MÁS CERCANOS

También conocido en inglés como $k-$ nearest neighbor graph. La idea es que el vértice $v_i$ se conecta con $v_j$ si $v_j$ está dentro de los $k-$ vecinos más cercanos.

Esto puede llevar a un grafo dirigido, dado que la relación de unirse los vecinos más cercanos es asimétrica. Hay un par de soluciones:

- Ignorar la dirección de las aristas.
- Sólo conectar a $v_i$ con $v_j$ si ambos están dentro de sus $k-$ vecinos más cercanos.

Finalmente, una vez terminada la conexión agregar los pesos basados en las similaridades entre los puntos unidos.

# COMPLETAMENTE CONECTADO

In english, for the friends, fully connected graph. Simplemente conectamos todos los puntos que tengan similaridad positiva.

Los pesos directamente los tomamos como $w_{ij} = s_{ij}$.

Esta construcción sólo es útil cuando la similaridad realmente representa la relación local que tiene los datos.

Un ejemplo muy utilizado es la similaridad Gausiana:

$$s(x_i, x_j) = exp\{-\frac{||x_i - x_j||^2}{2\sigma^2}\}.$$

# GRAFOS LAPLACIANOS

La matrix no normalizada de un grafo Laplaciano se define como $L = D - W$.

# GRAFOS LAPLACIANOS

La matrix no normalizada de un grafo Laplaciano se define como $L = D - W$.

**Lema**

La matriz $L$ satisface las siguientes propiedades:

1. para todo $u \in \mathbb{R}^n$,

$$\cancel{uLu' = \frac{1}{2}\sum_{i=1}^{n} w_{ij}(u_i - u_j)^2.}$$

2. $L$ es simétrica y semidefinida positiva.
3. El autovalor más chico de $L$ es 0 con autovector $(1, \ldots, 1)$.
4. $L$ tiene autovalores reales y no negativos.

**Proposición**

Sea $G$ un grafo indirecto con pesos no negativos. Entonces, la multiplicidad del autovalor 0 de $L$ es igual a la cantidad de componentes conexas del grafo. Más aún, el autoespacio correspondiente al autovalor 0 tiene como generadores a los vectores indicadores de cada componente conexa.

**Proposición**

Sea $G$ un grafo indirecto con pesos no negativos. Entonces, la multiplicidad del autovalor 0 de $L$ es igual a la cantidad de componentes conexas del grafo. Más aún, el autoespacio correspondiente al autovalor 0 tiene como generadores a los vectores indicadores de cada componente conexa.

**Observación:** una vez elegido el grado esta proposición nos dice que el problema de encontrar clusters se transforma en encontrar autovectores de 0.

## Algoritmo Espectral:

Dado un grafo con $n$ vértices y matriz de pesos $(w_{ij})_{ij}$ y $K$ el número de clusters deseados:

1. Construya la matriz Laplaciana $L = D - W$.
2. Encuentre los $K$ autovectores corresponientes a los autovalores más chicos.
3. Sea $U$ la matriz de tamaño $n \times K$ de autovectores.
4. Para $i = 1, \ldots, n$ sea $y_i$ el correspondiente a la i-ésima fila de $U$.
5. Clusterizar los vectores $y_i$ $i = 1, \ldots, n$ con kmeans en $K$ clusters $C_1, \ldots, C_K$.
6. Obtenemos las componentes conexas $A_i = \{v_j : \ y_j \in C_i\}$.

Hay dos matrices que se las llaman matrices normalizadas de un grafo Laplaciano en la literatura.

$$L_{sym} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2}.$$
$$L_{rw} = D^{-1} L = I - D^{-1} W.$$

Ambas matrices tiene un relación muy estrecha y cumplen propiedades similares a la $L$ original.

# Grafo Cortado

# INTRO

Recordemos que la idea en que nos basamos es separar datos en distintos grupos de acuerdo a sus similaridades.

# INTRO

Recordemos que la idea en que nos basamos es separar datos en distintos grupos de acuerdo a sus similaridades.

Para datos dados en forma de un grafo de similaridades queremos encontrar un partición del grafo tal que las aristas entre los distintos grupos tengan peso muy chico y las aristas dentro de un mismo grupo tenga peso elevado.

Recordemos que la idea en que nos basamos es separar datos en distintos grupos de acuerdo a sus similaridades.

Para datos dados en forma de un grafo de similaridades queremos encontrar un partición del grafo tal que las aristas entre los distintos grupos tengan peso muy chico y las aristas dentro de un mismo grupo tenga peso elevado.

Sea $G = (E, V)$ un grafo definimos para dos subconjuntos $A, B \subset V$,

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

Dado un grafo de similaridades con matriz de pesos $W$, nuestra intuición se traduce a lo que se llama el problema de mínimo corte.

Dado el número de grupos, $K$, encontrar la partición de $V$ $A_1, \ldots, A_K$ que minimiza:

$$cut(A_1, \ldots, A_K) = \frac{1}{2} \sum_{i=1}^{K} W(A_i, \overline{A}_i).$$

# ALTERNATIVAS

Ponderar por la cantidad de vértices en cada conjunto de la partición:

$$Radiocut(A_1, \ldots, A_K) = \frac{1}{2} \sum_{i=1}^{K} \frac{cut(A_i, \overline{A}_i)}{|A_i|}.$$

Ponderar por el volumen de cada conjunto de la partición:

$$Ncut(A_1, \ldots, A_K) = \frac{1}{2} \sum_{i=1}^{K} \frac{cut(A_i, \overline{A}_i)}{vol(A_i)}.$$

Supongamos que $K = 2$, nuestro objectivo es resolver el problema de optimización

$$\min_{A \subset V} Radiocut(A, \overline{A}).$$

Sea $u = (u_1, \ldots, u_n)$ el vector en $\mathbb{R}^n$ definido por

$$u_i = \begin{cases} \sqrt{\frac{|\overline{A}|}{|A|}} \text{ si } v_i \in A \\ -\sqrt{\frac{|A|}{|\overline{A}|}} \text{ si } v_i \notin A \end{cases} \tag{1}$$

Ahora la función objetivo *Radiocut* se puede reescribir utilizando la matriz no normalizada del grafo Laplaciano:

$$uLu' = \frac{1}{2}\sum_{i=1}^{n} w_{ij}(u_i - u_j)^2 = \cdots = |V| Radiocut(A, \overline{A}). \qquad (2)$$

Además,

$$\sum_{i=1}^{n} u_i = \sum_{i \in A} \sqrt{\frac{|\overline{A}|}{|A|}} - \sum_{i \in \overline{A}} \sqrt{\frac{|A|}{|\overline{A}|}} = |A|\sqrt{\frac{|\overline{A}|}{|A|}} - |\overline{A}|\sqrt{\frac{|A|}{|\overline{A}|}} = 0. \qquad (3)$$

Haciendo una cuenta análoga también tenemos que

$$||u||^2 = \sum_{i=1}^{n} u_i^2 = \cdots = |\overline{A}| + |A| = n. \tag{4}$$

Juntando lo que hicimos en (2), (3) y (4); el problema de minimización se convierte en:

$$\min_{A \subset V} uLu' \text{ sujeto a } u \perp (1, \ldots, 1) \text{ y } ||u|| = \sqrt{n}.$$

# COFFEE BREAK!