FACULTAD DE INGENIERÍA

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Teoría Básica de Grafos

Paola Bermolen

paola@fing.edu.uy

Definiciones y conceptos básicos

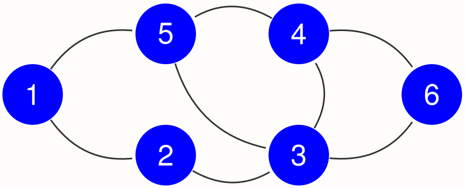
De redes a grafos

- Redes son sistemas complejos de elementos interconectados.
- Grafos son una representación matemática de tales sistemas.

  $\Rightarrow$ brinda un lenguaje formal para hablar de redes

- Componentes:
  nodos, vértices V
- Interconexiones:
  aristas E
- Sistema: redes,
  grafos $G(V, E)$

Grafos



Definición (Grafo)

Llamamos grafo $G(V, E)$ a un conjunto V de vértices o nodos conectados por un conjunto E de aristas. Los elementos de $E \subset V \times V$ son pares no ordenados de vértices distintos (u, v), $u, v \in V$

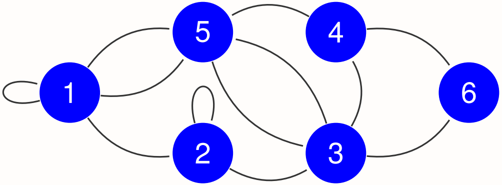
- En la figura $\Rightarrow$ Vértices $V = \{1, 2, 3, 4, 5, 6\}$
  $\Rightarrow$ Aristas $E = \{(1, 2), (1, 5), (2, 3), (3, 4), (3, 5), (3, 6), (4, 5), (4, 6)\}$
- Usualmene se dice que el grafo G tiene orden $N_v := |V|$, y tamaño $N_e := |E|$

Vértices y aristas en redes reales

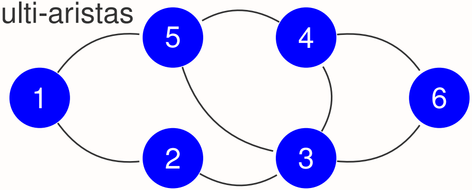
red	vértice	arista
Internet	Computer/router	Cable or wireless link
Metabolic web	Metabolite	Metabolic reaction
WWW	Web page	Hyperlink
Food web	Species	Predation
Gene-regulatory web	Gene	Regulation of expression
Friendship web	Person	Friendship or acquaintance
Power grid	Substation	Transmission line
Affiliation web	Person and club	Membership
Protein interaction	Protein	Physical interaction
Citation web	Article/patent	Citation
Neural web	Neuron	Synapse
⋮	⋮	⋮

Grafos simples y multi-grafos

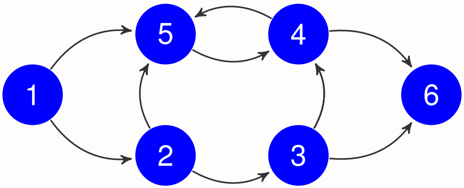
- En general, los grafos pueden tener self-loops y multi-aristas
  - $\Rightarrow$ Un grafo con esas características se llama multi-grafo



- ¿Cómo cambia la definición de grafo?
- La mayoría de los trabajos y lo que vamos a hacer se basa en grafos simples, sin self-loops ni multi-aristas



Grafos dirigidos



- En grafos dirigidos, los elementos de E son pares ordenados (u, v), $u, v \in V$
  - $\Rightarrow$ implica que la arista (u, v) es distinta de la arista (v, u)
  - $\Rightarrow$ las aristas dirigidas se llaman semiaristas o arcos
- Y a los grafos dirigidos, para abreviar, les decimos digrafos
  - $\Rightarrow$ Por convención, el arco (u, v) sale de u y apunta a v (los arcos se representan con flechas)
- Ej: relaciones no simétricas como seguidores en Twitter o citas en artículos

Grafos con pesos

- Muchas veces los vértices o las aristas tienen asociados valores numéricos

  $\Rightarrow$ tales grafos se llaman grafos con pesos

- Los valores pueden corresponder a medidas de un proceso definido en la red
- Ej: tráfico que cursa un enlace, cantidad de autos en una cierta ruta, nodo/persona infectado o no por covid, etc.
- Las multi-aristas en general se codifican mediante pesos

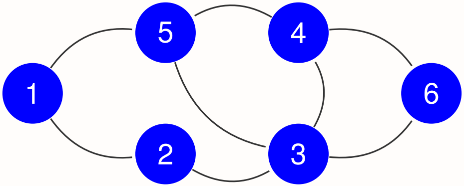
Representaciones de algunas redes típicas

Red	Grafo asociado
WWW	Multi-grafo (con loops), dirigido, sin pesos
Amistades en Facebook	No dirigido, sin pesos
Red de Citas	Dirigido, sin pesos, *acíclico*
Red de Colaboración	No dirigido, sin pesos
Llamadas telefónicas	Dirigido, con pesos
Interacción de proteínas	Multi-grafo (con loops), dirigido, sin pesos
⋮	⋮

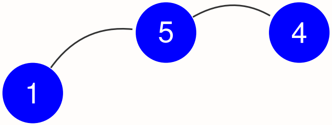
Subgrafos

Definición (Subgrafo)

Dado un grafo $G = (V, E)$ y un subconjunto de vértices $V' \subseteq V$, el subgrafo inducido por V' es el grafo formado por los vértices V' y todas las aristas de G que unan vértices en V'.



- Ejemplo: Grafo inducido por $V' = \{1, 4, 5\}$

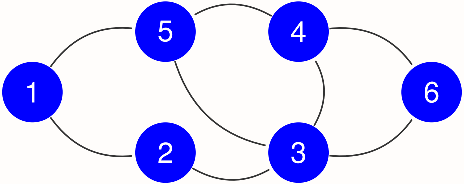


Relación de adyacencia

- Concepto necesario para estudiar la conectividad de un grafo
- La adyacencia es un concepto local muy simple

  $\Rightarrow$ dos vértices $u, v \in V$ son adyacentes si están conectados por una arista, es decir, si $(u, v) \in E$

  $\Rightarrow$ dos aristas $e_1, e_2 \in E$ son adyacentes si comparten algún vértice $v \in V$



- Ejemplo $\Rightarrow$ los vértices 1 y 5 son adyacentes; 2 y 4 no lo son

  $\Rightarrow$ la arista $(1, 2)$ es adyacente a $(1, 5)$, pero no a $(4, 6)$

Grado

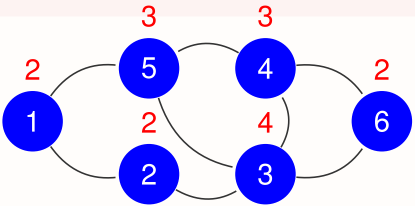
- Una arista (u, v) es incidente a los vértices u y v

Definición (Grado)

El grado d_v del vértice v es el número de su aristas incidentes

Definición (Secuencia de grados)

La secuencia de grados ordena los grados de todos los vértices en orden no-decreciente



- Ejemplo $\Rightarrow$ grados de los vértices en rojo, e.g., $d_1 = 2$ y $d_5 = 3$
  $\Rightarrow$ secuencia de grados: 2,2,2,3,3,4

Propiedades del grado

- Los valores de los grados van de 0 a $N_v - 1$
- La suma de la secuencia de grados es el doble del tamaño del grafo

$$\sum_{v=1}^{N_v} d_v = 2|E| = 2N_e$$

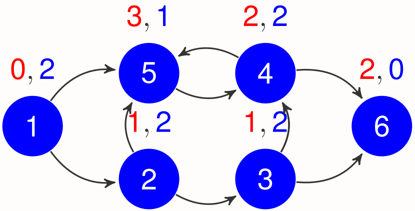
$\Rightarrow$ El número de vértices con grado impar es par (¡Euler!)

Definición (Grado en digrafos)

Para grafos dirigidos se definen los grados de entrada d_v^{in} y grados de salida d_v^{out} como el número de aristas entrantes y salientes a v, respectivamente

- ¿Se cumple la propiedad anterior?

Grados en digrafos



- Ejemplo $\Rightarrow$ grado entrante en rojo y grado saliente en azul
  $\Rightarrow$ Ejemplo, $d_1^{in} = 0, d_1^{out} = 2$ y $d_5^{in} = 3, d_5^{out} = 1$
- Suma de los grados:

$$\sum_{v=1}^{N_v} d_v^{in} = \sum_{v=1}^{N_v} d_v^{out} = |E| = N_e$$

Densidad

- Grado medio del grafo:

$$c = \frac{1}{N_v}\sum_{v=1}^{N_v} d_v = \frac{2N_e}{N_v}$$

- La máxima cantidad de aristas (grafo simple) es $\frac{N_v(N_v-1)}{2}$
- Se define la densidad ρ del grafo como

$$\rho = \frac{N_e}{\frac{N_v(N_v-1)}{2}} = \frac{2N_e}{N_v(N_v-1)} = \frac{c}{N_v-1} \approx \frac{c}{N_v}$$

- La densidad ($0 \leq \rho \leq 1$) se puede pensar como la probabilidad de que dos vértices elegidos al azar sean adyacentes

Sparsity

- Se dice que un grafo es sparse si cuando $n \to \infty$, se cumple que $\rho \to 0$
- No hay definición no asintótica: informalmente se usa ρ chico:
  - hay pocas aristas presentes comparadas con el máximo posible
- Muy importante para los aspectos computacionales
- Las redes reales en general son “sparse”
- Si al agregar nodos a la red el grado medio permanece constante entonces se cumple que $\rho \to 0$ y la red es sparse
- Lo mismo sucede si al agregar nodos el grado medio crece a menor velocidad

Movimientos en un grafo y conectividad

Movimientos en un grafo

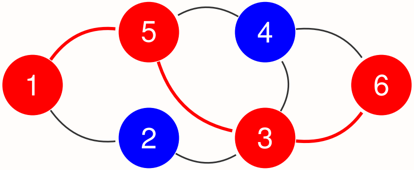
Definición (Paseo)

Un paseo de largo l entre v_0 y v_l es una secuencia alternada de vértices y aristas adyacentes entre sí

$$\{v_0, e_1, v_1, \ldots, v_{l-1}, e_l, v_l\}, \text{ donde } e_i \text{ es incidente a } v_{i-1}, v_i$$

Definición (Camino)

Un camino es un paseo sin nodos ni aristas repetidas



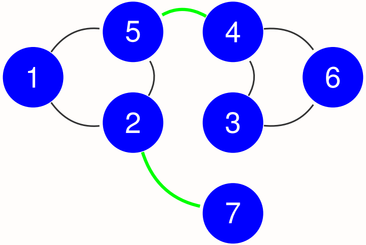
- Un paseo es cerrado si $v_0 = v_l$.
- Un ciclo es un paseo cerrado sin nodos repetidos excepto $v_0 = v_l$

Conectividad

- Un vértice v es alcanzable desde u si existe un camino de u a v

Definición (Grafo conexo o conectado)

Un grafo es G es conexo (o conectado) si cada vértice es alcanzable desde cualquier otro vértice



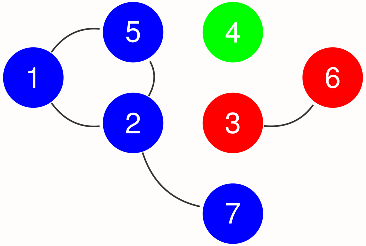
- Si las aristas puente son removidas, el grafo se desconecta

Componentes conexas

Definición (Componente)

Una componente es un subgrafo conexo maximal

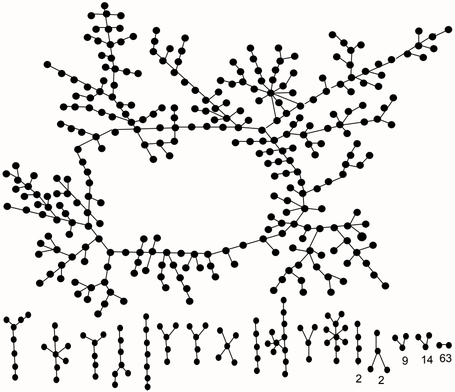
- Maximal significa que no se puede agregar más nodos sin que deje de ser conexo



- Ejemplo $\Rightarrow$ Las componentes son $\{1, 2, 5, 7\}$, $\{3, 6\}$ y $\{4\}$
  $\Rightarrow$ El subgrafo $\{3, 4, 6\}$ no es conexo, y el $\{1, 2, 5\}$ no es maximal
- Los grafos desconectados tienen 2 o más componentes
  $\Rightarrow$ La componente más grande se denomina componente gigante

Componentes gigantes

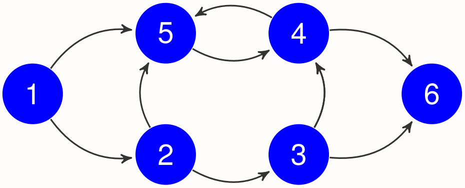
- Las redes reales de gran tamaño típicamente tienen una componente gigante
- Ejemplo: relaciones románticas en un liceo de US [Bearman et al'04]



- P: ¿Porque es razonable que haya una sola componente gigante?
- R: Basta una sola arista para unir dos componentes gigantes

Conectividad para digrafos

- Conectividad es un poco más sútil para digrafos. Dos nociones:
- **Def:** Un digrafo es fuertemente conectado si para cada par $u, v \in V$, u es alcanzable desde v (vía un camino dirigido) y vice versa
- **Def:** Un digrafo es débilmente conectado si es conectado al ignorar las direcciones de las aristas

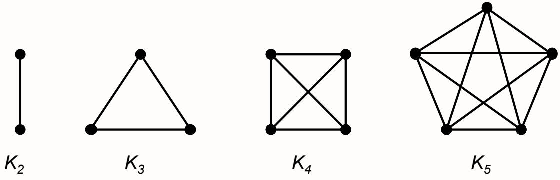


- El grafo el ejemplo débilmente conectado pero no fuertemente conectado

Familias de grafos

Grafos completos y cliques

- Un grafo completo K_n de orden n tiene todas las posibles aristas

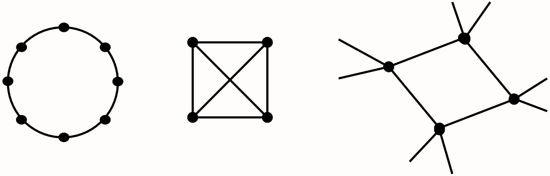


- P: ¿Cuál es el tamaño de K_n? ¿Grado medio? ¿Densidad?
- R: Número de aristas en $K_n = \frac{n(n-1)}{2}$, $c = n - 1$ y $\rho = 1$
- En análisis de redes aparecen como objeto de interés los cliques, i.e., subgrafos completos

  $\Rightarrow$ Noción extrema de comunidad o grupo cohesivo

Grafos regulares

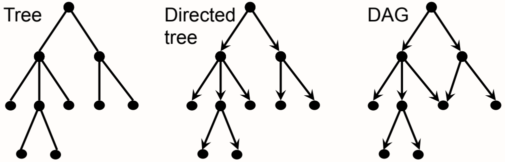
- Un grafo regular de orden d tiene todos sus vértices de grado d



- El grafo completo K_n es $(n-1)$-regular
  - Los ciclos son (sub) grafos 2-regulares
- Los grafos regulares aparecen con frecuencia en:
  - física y química, en el estudio de estructuras de cristales
  - referencias geo-espaciales y pixel-models en procesamiento de imágenes

Árboles y grafos dirigidos acíclicos

- Un grafo acíclico es un grafo sin ciclos (camino cerrado con las aristas todas dirigidas en el mismo sentido)
- Un árbol es un grafo no dirigido conexo acíclico (no hay loops).
- Ejemplos: ríos, evolución, red de citas de papers

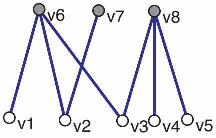


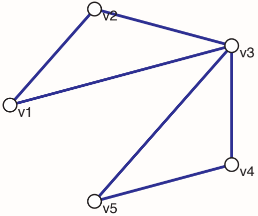
- Un árbol dirigido es un grafo dirigido cuyo grafo no dirigido subyacente es un árbol

  $\Rightarrow$ La raíz es el único vértice con caminos a todos los otros vértices
- Pensar en un grafo dirigido acíclico que no sea un árbol

Grafos bipartitos

- Un grafo $G(V, E)$ se dice bipartito si
  - $\Rightarrow$ V se puede dividir en dos conjuntos disjuntos V_1 y V_2 tales que
  - $\Rightarrow$ cada arista en E tiene un extremos en V_1, y el otro en V_2

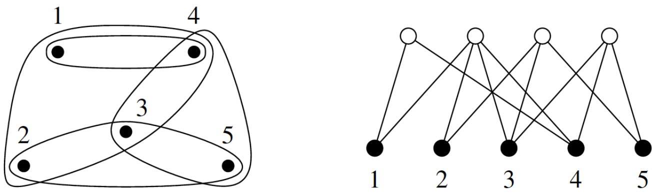




- Útil para representar por ejemplo, membresías o redes de afiliación
  - $\Rightarrow$ Vértices en V_1 pueden ser personas, y vértices en V_2 clubs
  - $\Rightarrow$ El grafo inducido $G(V_1, E_1)$ reúne miembros del mismo club

Hipergrafos y grafos bipartitos

- ➤ En algunos casos las aristas unen más de dos vértices (*affiliation networks*)
  - $\Rightarrow$ Un grafo con esas características se llama hipergrafo

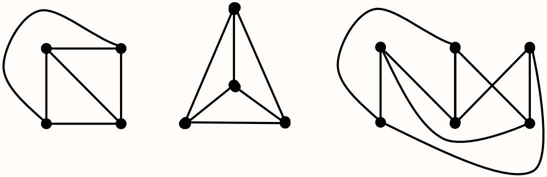


- ➤ Una manera más sencilla de representarlos es con grafos bipartitos
  - $\Rightarrow$ Se crean nuevos vértices (tantos como grupos) y se conectan los vértices a los grupos a los que pertenecen.

Grafos planares

- Un grafo $G(V, E)$ se dice planar si puede ser dibujado en el plano de modo tal que no haya dos aristas que se crucen entre ellas

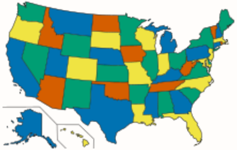
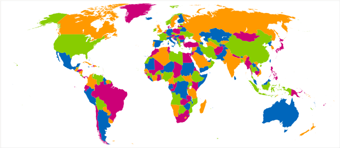
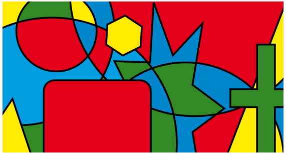
  $\Rightarrow$ Los árboles son grafos planares r



- Los grafos planares se pueden dibujar en el plano usando solo líneas rectas
- Útil para representar o mapear redes con componentes espaciales: ríos, carreteras (sacando puentes)
- Los árboles son planares

Grafos planares

- Difícil determinar si un grafo es planar pero hay algunas condiciones necesarias que relacionan, vértices, aristas y caras (ej: $v - e + f = 2$, $c < 6$)
- Teorema de los cuatro colores:

Teoría algebraica de grafos

Matriz de adyacencia

- Teoría algebraica de grafos refiere a la representación matricial de grafos
- ¿Cómo representamos la conectividad de un grafo $G(V, E)$ en una matriz?

Definición (Matriz de adyacencia)

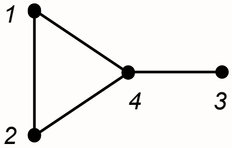
Dado un grafo simple $G(V, E)$, se define la matriz de adyacencia a una matriz binaria y simétrica $\mathbf{A} \in \mathcal{M}_{N_v \times N_v}$ tal que:

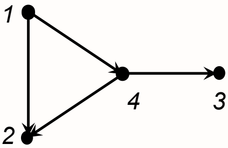
$$A_{ij} = \begin{cases} 1, & \text{si } (i,j) \in E \\ 0, & \text{en otro caso} \end{cases} .$$

- Los vértices los numeramos con valores $1, \ldots, N_v$
- Para grafos dirigidos, la matriz de adyacencia puede no ser simétrica
- Para grafos con pesos la matriz no es binaria, se define $A_{ij} = w_{ij}$

Ejemplos matriz de adyacencia

- Ejemplos para grafos simples y digrafos





$$\mathbf{A}_u = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad \mathbf{A}_d = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

- Ejercicio: ¿Cómo queda la matriz de adyacencia asociada a un grafo bipartito? ¿y a un grafo dirigido acíclico?

Propiedades de la matriz de adyacencia

Dada una matriz de adyacencia $\mathbf{A} \in \mathcal{M}_{n\times n}$ ¿qué puedo decir del G asociado? Supongamos binaria y simétrica (G simple)

- Cantidad de vértices $N_v = n$
- Cantidad de aristas $\sum_{i,j} A_{ij} = 2N_e$
- La suma por filas (o columnas) nos da los grados de cada vértice $\sum_{j=1}^{N_v} A_{ij} = d_i$
- Para digrafos $\mathbf{A}$ no es simétrica y las sumas por filas o columnas difieren:

$$\sum_{j=1}^{N_v} A_{ij} = d_i^{out}, \qquad \sum_{i=1}^{N_v} A_{ij} = d_j^{in}$$

$\Rightarrow$ se obtienen los grados entrantes y salientes

Matriz de adyacencia y conectividad

- Caminos: ¿podemos deducir si existe un camino entre dos vértices a partir de la matriz asociada? ¿y si existe un camino de largo r?

Teorema (Caminos de largo r)

Sea $\mathbf{A}^r$ la potencia r-ésima de $\mathbf{A}$, con entradas $A_{ij}^{(r)}$, entonces $A_{ij}^{(r)}$ da el número de caminos de i a j de largo r in G

- Caminos de largo 2 entre el mismo vértice:

$$\mathrm{tr}(\mathbf{A}^2)/2 = N_e$$

- Caminos de largo 3 que empiezan y terminan en el mismo vértice (triángulos):

$$\mathrm{tr}(\mathbf{A}^3)/6 = \#\triangle$$

Espectro de la matriz de adyacencia

- Si A es simétrica, entonces tiene n valores propios reales y se cumple que

$$A = PDP^t \quad \text{con } D = \mathrm{diag}(\lambda_1, \cdots, \lambda_n) \text{ y } P^t = P^{-1}$$

- Relación con la cantidad de aristas y triángulos:

$$N_e = \mathrm{tr}(\mathbf{A}^2)/2 = (\lambda_1^2 + \cdots + \lambda_n^2) \quad \text{y} \quad \#\triangle = \mathrm{tr}(\mathbf{A}^3)/6 = (\lambda_1^3 + \cdots + \lambda_n^3)$$

- G es d-regular si y sólo si $\mathbf{1}$ es un vector propio de $\mathbf{A}$ asociado al valor propio d i.e.,

$$\mathbf{A1} = d\mathbf{1}$$

- ¿Espectros iguales $\Leftrightarrow$ grafos iguales?

  $\Rightarrow$ ¡Lamentablemente NO! Ver ejemplo

Espectro de matrices

Definición (Valores y vectores propios)

Dada una matriz $A \in \mathcal{M}_{n\times n}(\mathbb{R}^n)$ se dice que $\lambda \in \mathbb{R}$ es un valor propio de A si existe un vector $v \in \mathbb{R}^n$ no nulo tal que $Av = \lambda v$. El vector v se llama vector propio de A asociado al valor propio λ

- Los vap's son raíces del polinomio característico $\chi(\lambda) = \det(A - \lambda I) = 0$
- Los veps's son un subespacio vectorial $S_\lambda = \mathrm{N}(A - \lambda I)$.

  $\Rightarrow$ se cumple que $1 \leq \mathrm{mg}(\lambda) \leq \mathrm{mg}(\lambda) \leq n$ donde $\mathrm{mg}(\lambda)$ es la dimensión de S_λ y $\mathrm{ma}(\lambda)$ es la multiplicidad como raíz de $\chi(\lambda)$

- A es diagonalizable $\Leftrightarrow A = PDP^{-1}$ con D diagonal (vaps en la diagonal) $\Leftrightarrow$ existe una base de $\mathbb{R}^n$ formada por vep's de $A \Leftrightarrow \mathrm{ma}(\lambda) = \mathrm{mg}(\lambda)$ para todo vap λ de A

Espectro de matrices

- Si $A \in \mathcal{M}_{n\times n}(\mathbb{R}^n)$ es simétrica, es diagonalizable y se cumple $A = PDP^t$ con D diagonal (vaps en la diagonal) y P ortogonal
  - $\Rightarrow$ existe una base ortonormal de $\mathbb{R}^n$ formada por vep's de A
- Las matrices de adyacencia de grafos no dirigidos son simétricas
- Las matrices de adyacencia de grafos dirigidos no son necesariamente simétricas
- Si $A \in \mathcal{M}_{m\times n}(\mathbb{R}^n)$ (no cuadrada) entonces se utiliza la descomposición en valores singulares: $A = USV^t$ con S diagonal, U y V ortogonales
  - $\Rightarrow$ D está formada por los valores singulares $\sigma_i = \sqrt{\lambda_i}$ con λ_i vap de A^tA
  - $\Rightarrow$ el número de valores singulares no nulos es el rango de A

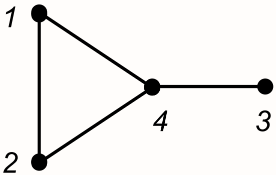
Laplaciano del grafo

Definición (Laplaciano del grafo)

Dado un grafo simple $G = (V, E)$, se define el laplaciano del grafo a una matriz simétrica $\mathbf{L} \in \mathcal{M}_{N_v \times N_n}$ tal que

$$L_{ij} = \begin{cases} d_i, & \text{si } i = j \\ -1, & \text{si } (i,j) \in E \\ 0, & \text{en otro caso} \end{cases},$$

- Definiendo $\mathbf{D}$ matriz diagonal con $D_{ii} = d_i$ de grados, resulta que $\mathbf{L} = \mathbf{D} - \mathbf{A}$



$$\mathbf{D} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \qquad \mathbf{L} = \begin{pmatrix} 2 & -1 & 0 & -1 \\ -1 & 2 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ -1 & -1 & 0 & 3 \end{pmatrix}$$

Propiedades del Laplaciano

- Se puede extender a grafos no dirigidos con pesos, usando la matriz de pesos y el grado la suma de los pesos de las aristas adyacentes.
- No hay una forma natural de extender la definición a digrafos
- Dada una matriz $\mathbf{L} \in \mathcal{M}_{n \times n}$ laplaciano del grafo G, ¿qué puedo decir de G?
- Cantidad de vértices: $n = N_v$
- Cantidad de aristas: $\mathrm{tr}(\mathbf{L}) = \sum_{v \in V} d_v = 2|N_e|$
- Suma por filas

$$\sum_{j=1}^{n} \mathbf{L}_{ij} = d_i - \sum_{j=1}^{n} \mathbf{A}_{ij} = d_i - d_i = 0$$

$\Rightarrow$ $\mathbf{1}$ es un vector propio de $\mathbf{L}$ asociado al valor propio 0

$\Rightarrow$ $\mathbf{L}$ no es invertible y $\mathrm{rango}(\mathbf{L}) < n$

Espectro del Laplaciano

- **Regularidad - smoothness:** Dado un vector $\mathbf{x} \in \mathbb{R}^{N_v}$ de "valores en los vértices" de un grafo G, se cumple que

$$\mathbf{x}^\top \mathbf{L}\mathbf{x} = \sum_{(i,j)\in E} (x_i - x_j)^2$$

$$\mathbf{x}^\top \mathbf{L}\mathbf{x} = ||D\mathbf{x}||^2 = (D\mathbf{x})^t(D\mathbf{x}) = \mathbf{x}^\top D^t D\mathbf{x}$$

con D el operador diferenciación. Es una medida de regularidad de funciones definidas en G.

- **Semidefinida positiva:** se deduce de lo anterior ya que $\mathbf{x}^\top \mathbf{L}\mathbf{x} \geq 0$ para todo $\mathbf{x} \in \mathbb{R}^{N_v}$

  $\Rightarrow$ Todos los valores propios de **L** son no-negativos

  $\Rightarrow$ Ya vimos que 0 es siempre valor propio y es entonces el más chico

Teoría espectral de grafos

- Espectro y conectividad: El segundo valor propio más chico λ_2 es distinto de 0 ($\lambda_2 \neq 0$), si y solo si G es conexo
  - $\Rightarrow$ si 0 tiene multiplicidad algebraica 1, entonces G es conexo.
  - $\Rightarrow$ λ_2 se denomina “spectral gap”
- **L** tiene K valores propios nulos (multiplicad algebraica K) si y solo si G tiene K componentes conectadas
  - $\Rightarrow$ forma de bloques de **L**
- El vector propio asociado a λ_2 se llama vector de Fiedler y se usa como un primer método de clusterización
- Ejericio: grafo de Barbell de tamaño $n = 3$ (dos grafos completos K_n unidos por una arista). Hallar valores y vectores propios del Lapaciano. ¿Qué observan?