

TEMA 2: Redes y grafos, definiciones y conceptos

Lourdes Araujo y Juan Martinez-Romo
Dpto. Lenguajes y Sistemas Informáticos
UNED

Redes y grafos

- Una **red** es una colección de objetos en la que algunos pares de objetos están conectados por un enlace.
- Un **grafo** es una representación matemática de una red.

$$\text{RED} = \text{GRAFO} + \text{DATOS}$$

- En general utilizaremos los términos indistintamente.

Redes y grafos: Definición

Una **red** $N = (V, L, P, W)$ está formada por

- Un **grafo** $G = (V, L)$, siendo V con conjunto de nodos o vértices de tamaño $|V|$ y L un conjunto de enlaces de tamaño $|L|$.
- P son funciones o propiedades de los vértices.
 $P: V \rightarrow A.$
- W son funciones o pesos de los enlaces.
 $W: L \rightarrow A.$

Elección de la representación

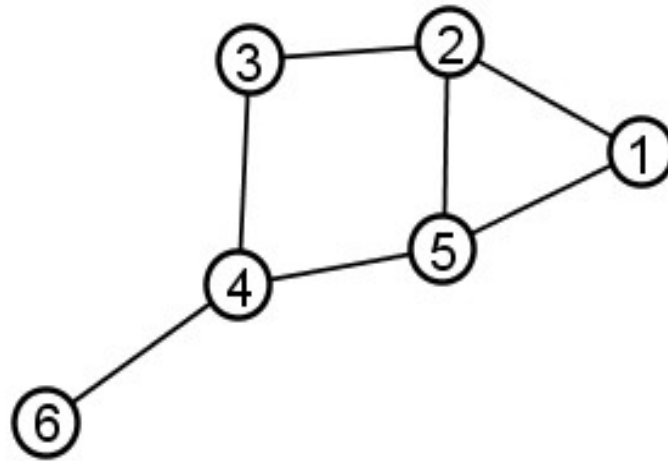
- Elegir la **representación apropiada** es determinante para la utilidad de las redes.
- A veces existe una única representación. Otras debemos elegir entre **distintas alternativas**.
- La forma en la que se asignan los enlaces determina el problema que queremos analizar.

Elección de la representación: Ejemplos

- Si conectamos personas que han hablado por teléfono construimos una **red de contactos**.
- Si conectamos artículos científicos que se citan construimos una **red de citas**.
- Si conectamos documentos con un alto número de palabras en común construimos una **red temática**.

Grafos no dirigidos

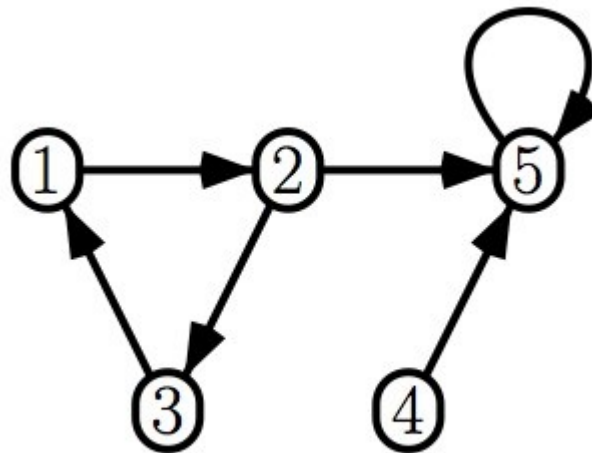
- Los grafos son **no dirigidos** cuando los enlaces son simétricos



- **Ejemplos:** amigos en Facebook, relaciones familiares.

Grafos dirigidos

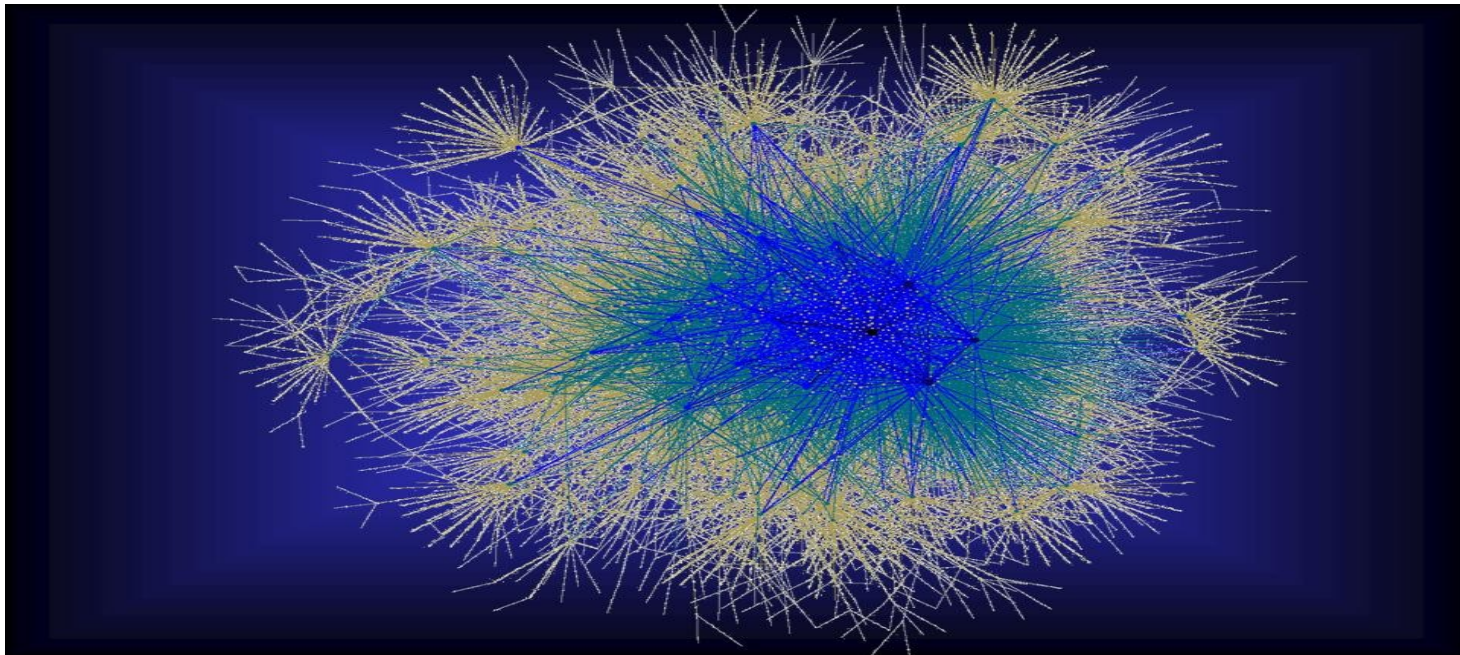
- Los grafos son **dirigidos** cuando los enlaces son flechas que apuntan a un nodo



- **Ejemplos:** llamadas de teléfono, seguidores en Twitter, la web

La web

- La web es un grafo dirigido
 - Los nodos son las páginas web
 - Los enlaces son los hiperenlaces de las páginas

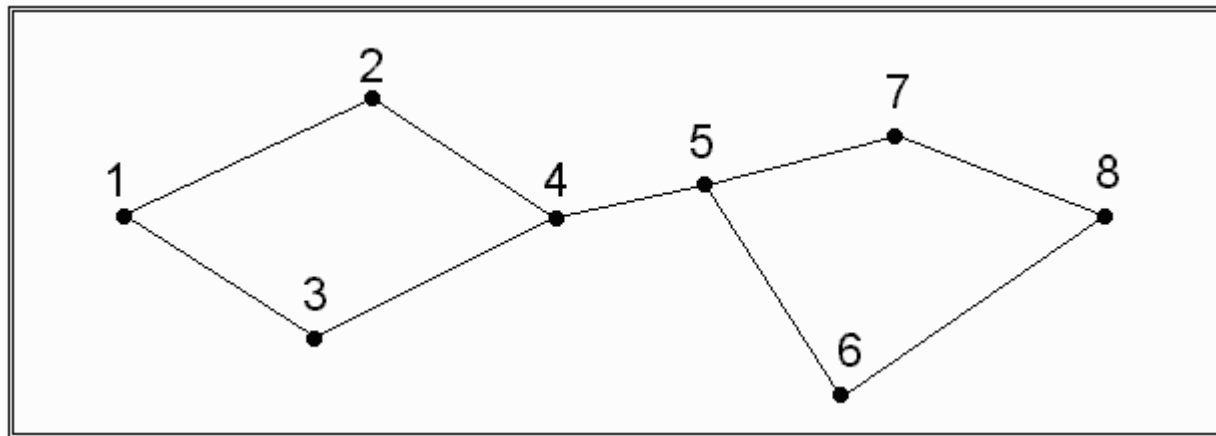


Propiedades de los nodos

- **Grado** (degree) de un nodo $\deg(v)$ es el número de enlaces con un extremo en v .
- **Grado de entrada** (indegree) de un nodo v , $\text{indeg}(v)$, es el número de enlaces que terminan en v .
- **Grado de salida** (outdegree) de un nodo v , $\text{outdeg}(v)$, es el número de enlaces que parten de v .

Grafo conexo

- Cualquier par de nodos está unido por un camino

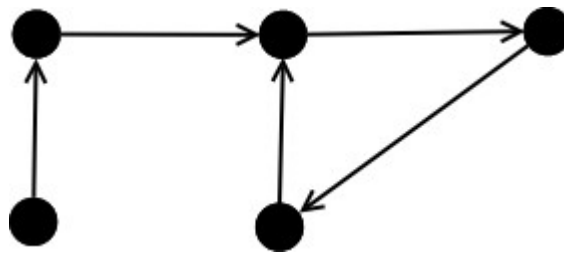


Conectividad de los grafos

- **Enlace puente**: si se borra el grafo pasa a ser desconexo (enlace 4-5 en el ejemplo)
- **Punto de articulación**: nodo que si se borra (junto con las aristas que inciden en él) el grafo pasa a ser desconexo (nodos 4 y 5 en el ejemplo)

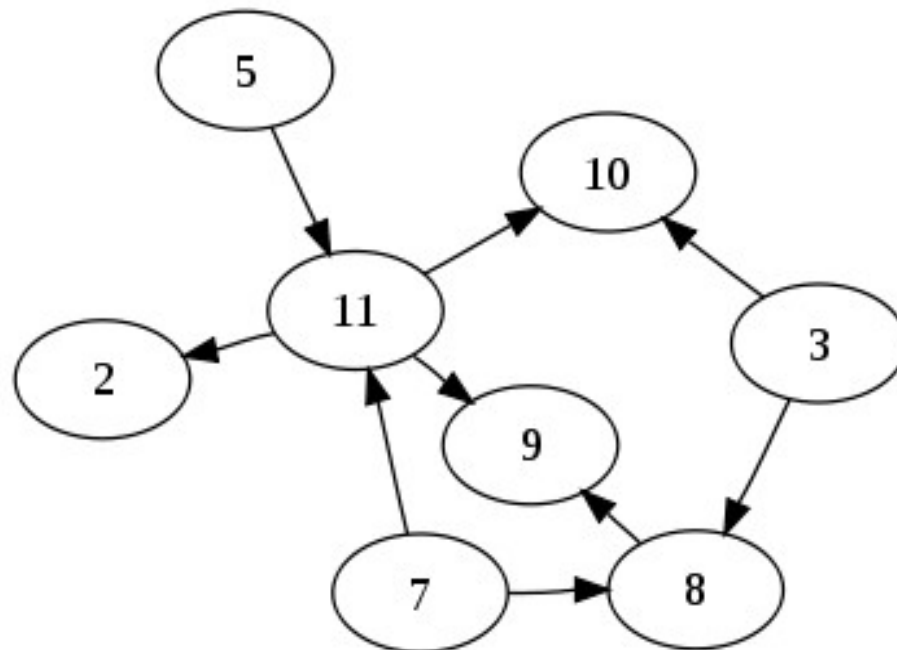
Fuerza de la conexión

- **Grafos dirigidos fuertemente conexos:** Existe un camino de cada nodo a cada otro nodo y viceversa (strongly connected directed graph)
- **Grafos dirigidos debilmente conexos:** es conexo si no tenemos en cuenta la dirección de los enlaces

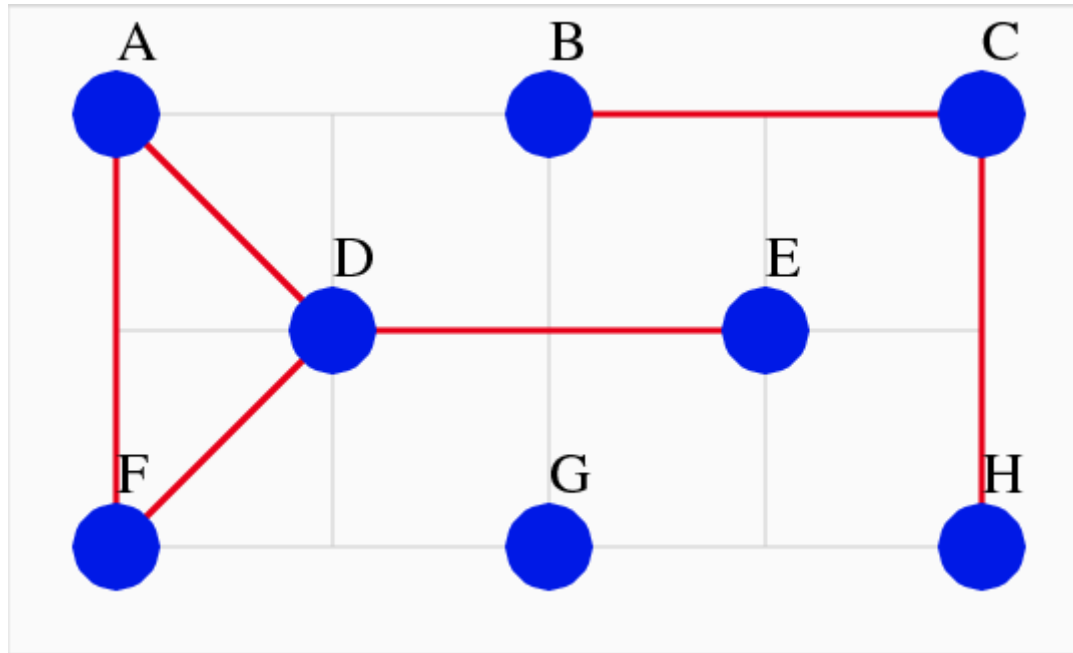


Grafos dirigidos acíclicos

- **Grafo dirigido acíclico**: no tiene ciclos (Directed acyclic Graph, DAG)



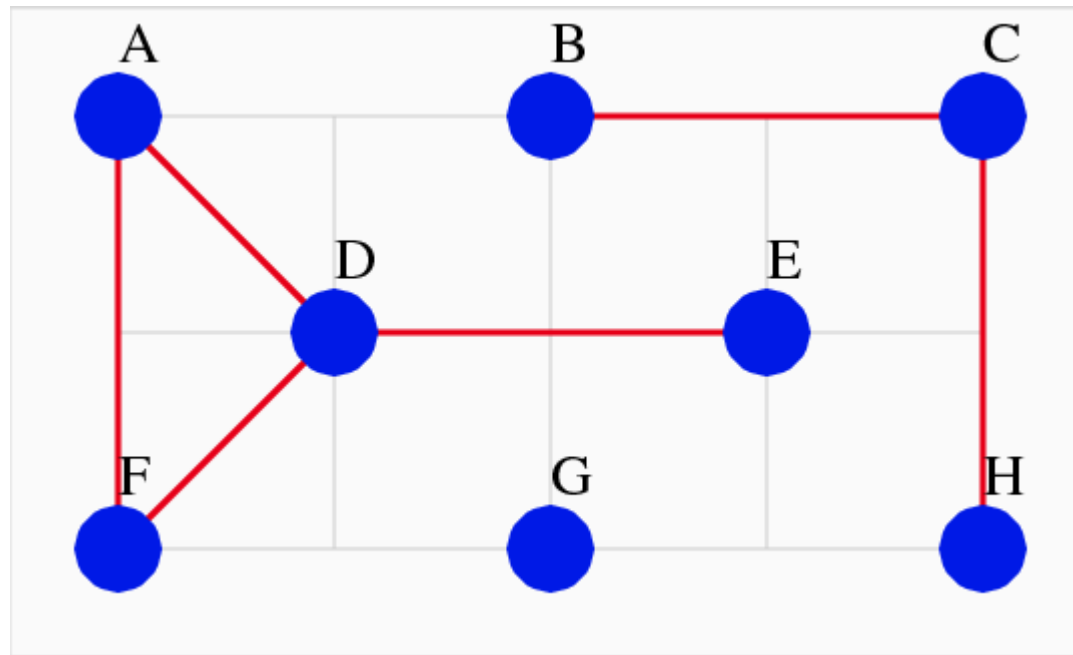
Componentes conexas



- Un grafo desconexo está compuesto por dos o más **componentes conexas**

Componente gigante

- La mayor componente se denomina **componente gigante**: A-F-D-E

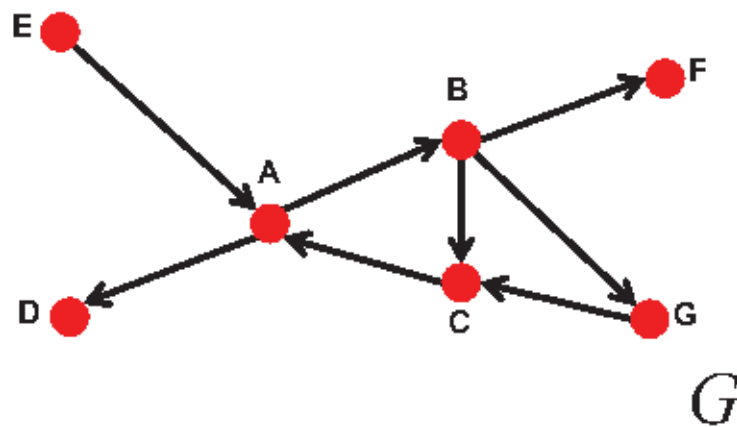


Algunos resultados(1)

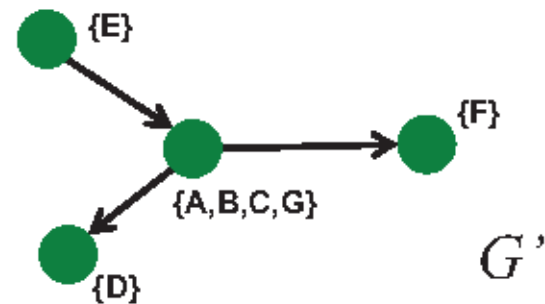
- Todo grafo dirigido es un DAG (Directed acyclic) en su SCC (Strong. Connected Comp.)
 - SCCs particionan los nodos de G : cada nodo está exactamente en un SCC
 - Si construimos un grafo G' cuyos nodos son SCCs y con un enlace entre ellos si lo hay entre los correspondientes SCCs en G , entonces G' es un DAG.

Algunos resultados (2)

- Ejemplo



- (1) Strongly connected components of graph G : $\{A, B, C, G\}$, $\{D\}$, $\{E\}$, $\{F\}$
- (2) G' is a DAG:



Algunos resultados (3)

- Los SCCs particionan los nodos de G : cada nodo de G pertenece exactamente a un SCC.

La web como un grafo dirigido

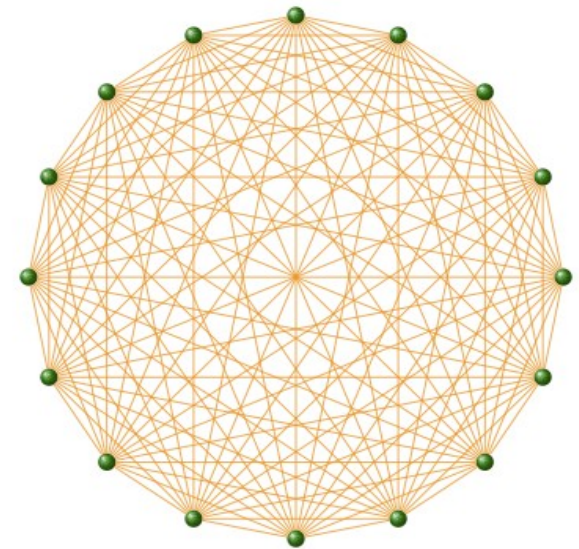
- 203 millones de páginas y 1.5 billones de enlaces [Broder et al. 2000]
- Mayor SCC: 28% de los nodos (56 millones)
- Alrededor de 44 millones de nodos con enlaces dirigidos a ese SCC
- Alrededor de 44 millones de nodos conectados a ese SCC por enlaces salientes.

Grafo completo

- El número máximo de enlaces en un grafo no dirigido de N nodos es

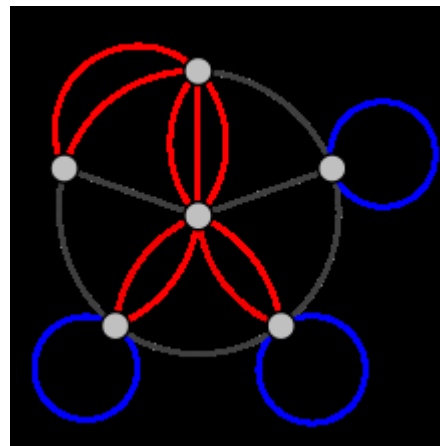
$$E_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$$

- Grafo completo:** grafo con un número de enlaces $E = E_{max}$
- Su grado medio es $N - 1$.



Otros tipos de grafos (1)

- **Multigrafo**: puede tener varios enlaces entre los mismos nodos
- **Grafos con autoenlaces**: enlaces que salen y llegan al mismo nodo

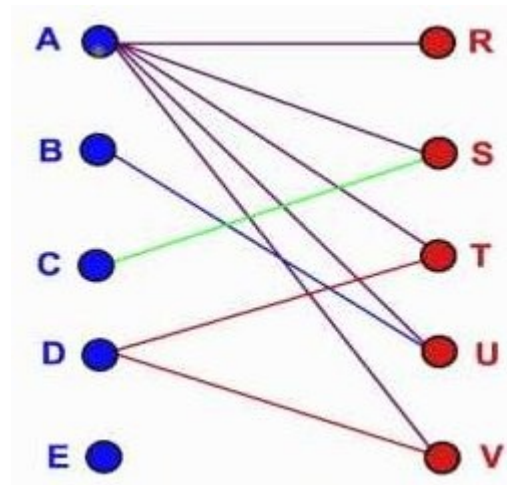


Representación de algunas redes

- WWW → multigrafo dirigido con autociclos
- Amistades de Facebook → no dirigido y sin pesos
- Redes de citas → sin pesos, dirigido, acíclico
- Llamadas a moviles → multigrafo dirigido

Otros tipos de grafos (2)

- **Grafo bipartito**: sus nodos pueden dividirse en dos conjuntos disjuntos U y V tales que cada enlace conecta a un nodo de U con un nodo de V (U y V son conjuntos independientes).



Otros tipos de grafos (3)

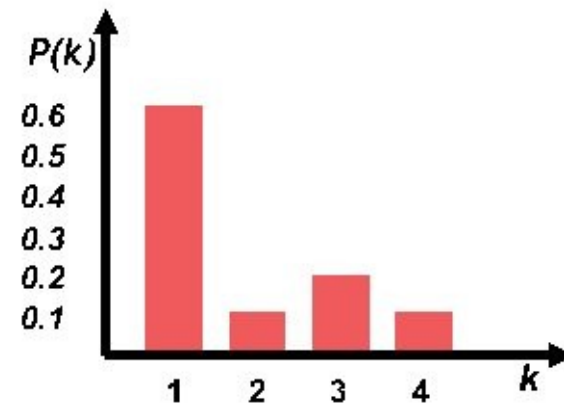
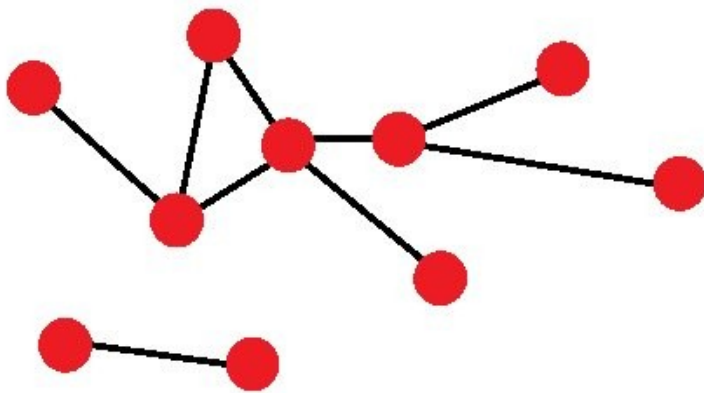
- Ejemplos de grafos bipartitos:
 - Autores-a-artículos (los escriben)
 - Actores-a-Peliculas (las protagonizan)
 - Sistemas de recomendación:
 - Usuarios-a-Peliculas (las puntúan), libros, etc.

Propiedades de las redes

- ¿como caracterizamos una determinada red?
 - Distribución del grado
 - Longitud de los caminos
 - Coeficiente de clustering

Distribución del grado $P(k)$

- **Distribución del grado $P(k)$** : Probabilidad de que un nodo tomado al azar tenga grado k
- $N_k = \#$ nodos de grado k
- Histograma normalizado $P(k) = N_k / N$

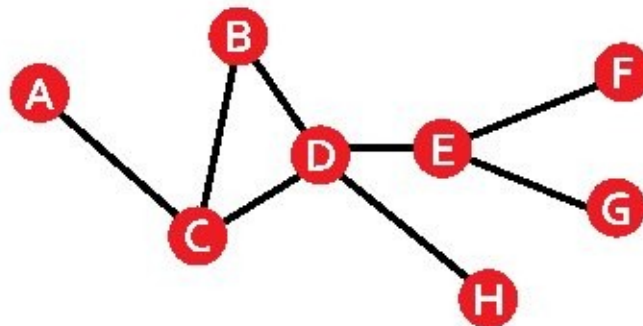


Caminos (walks)

- **Camino** (walk): secuencia de nodos de G en la que cada nodo está enlazado con el siguiente.
- Los caminos pueden intersecarse con ellos mismos y pasar por el mismo nodo varias veces.

Caminos (2)

- Ejemplo: ACBDCDEG



- En los grafos dirigidos un camino sólo puede seguir la dirección de las flechas.

Número de caminos entre dos nodos

- Número de caminos entre los nodos u y v : A_{uv}
- **de long 1**: si hay un enlace entre u y v , $A_{uv}=1$, sino $A_{uv} = 0$
- de long 2,
- ...
- La forma de calcularlos depende de la representación (Tema 3).

Distancia entre nodos h

- **Distancia (camino más corto o geodésica)** entre un par de nodos: número de enlaces en el camino más corto que conecta los nodos
- Si no hay camino se considera infinita
- En los grafos dirigidos la distancia no es simétrica: los caminos siguen la dirección de las flechas

Diámetro de una red

- **Diámetro**: distancia máxima (camino más corto) entre cualquier par de nodos en el grafo
- **Longitud media del camino** en un grafo (o componente) conexo o en un grafo (o componente) dirigido fuertemente conexo:

$$\bar{h} = \frac{1}{2 E_{max}} \sum_{i, j \neq i} h_{ij}$$

h_{ij} es la distancia del nodo i al j

- Se suele calcular sobre los nodos conectados, ignorando los caminos infinitos.

Coeficiente de clustering C

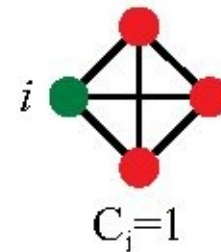
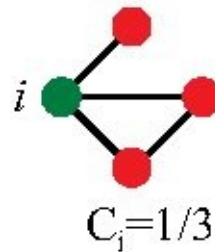
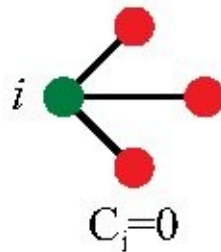
- Indica la proporción de vecinos del nodo i (con grado k_i) que están conectados.
- $C_i \in [0, 1]$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

- e_i es el número de enlaces entre los vecinos de i .

Coeficiente de clustering (2)

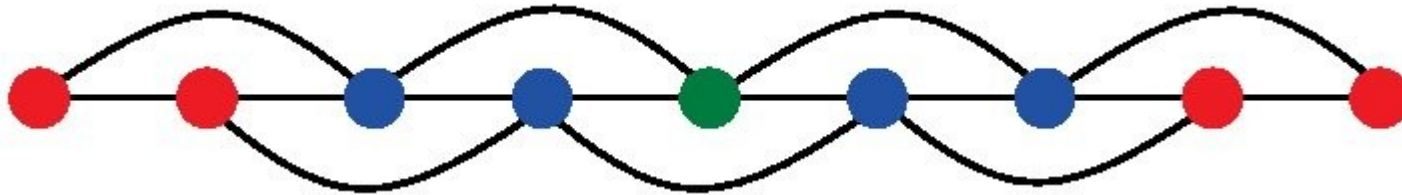
- Ejemplo:



- Coeficiente medio de clustering C :

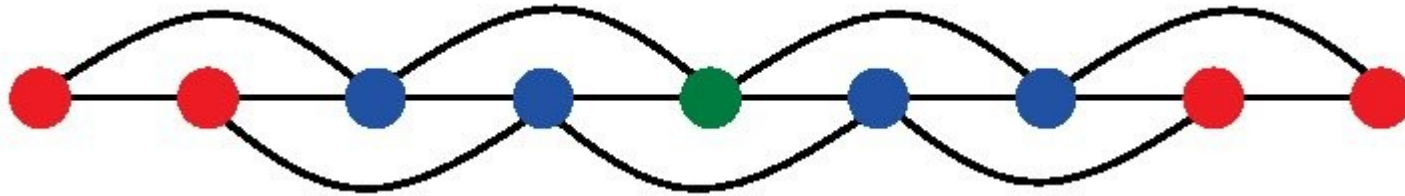
$$C = \frac{1}{N} \sum_i^N C_i$$

Ejemplo: reticulo 1 dimensión (1)



- Nos interesan cantidades cuando $N \rightarrow \infty$
- Usamos notación O: $f(x) = O(g(x))$ cuando $x \rightarrow \infty$ si $f(x) < g(x) \cdot c$ para todo $x > x_0$ y alguna constante c .
- $k_i = 4$ para todos los nodos (ignorando los extremos)
- $C = 1/2$

Ejemplo: reticulo 1 dimensión (2)

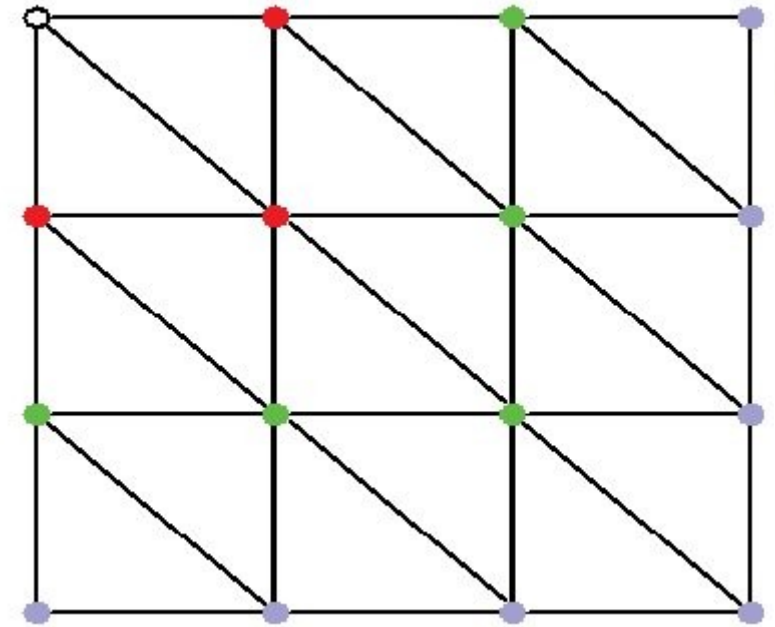


- $P(k) = \delta(k-4)$ ($\delta(x)=1$ si $x=0$, $\delta(x)=0$ sino, δ de Kronecker)
- Diámetro:
$$h_{max} = \left\lceil \frac{N-1}{2} \right\rceil = O(N)$$
- Long. media del camino más corto $\bar{h} = O(N)$

Ejemplo: retículo 2 dimensiones

- $K = 6$ para cada nodo interno
- $P(k) = \delta(k-6)$
- $C = 6/15$ para los nodos internos
- Diámetro

$$h_{max} = O(\sqrt{N})$$



Algunos modelos importantes de redes

- **Aleatorio**: generación aleatoria de enlaces entre nodos.
- **Regular**: todos los nodos tienen el mismo grado.
- **Mundo pequeño**: camino medio entre vértices pequeño y coeficiente de clustering alto.
- **Scale-free network**: la distribución de grados sigue una ley de potencias $P(k) = k^{-c}$.

Modelo de grafo aleatorio (Erdős-Renyi)

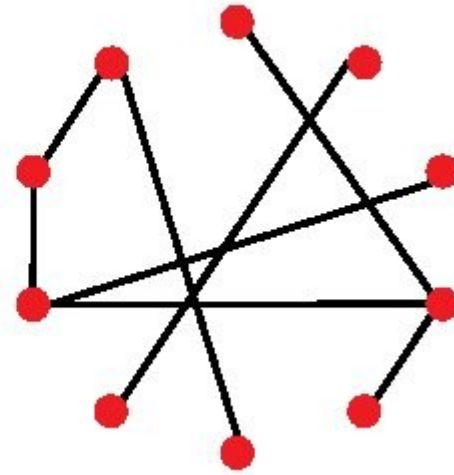
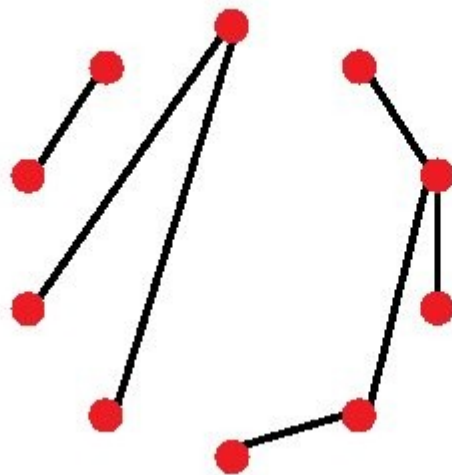
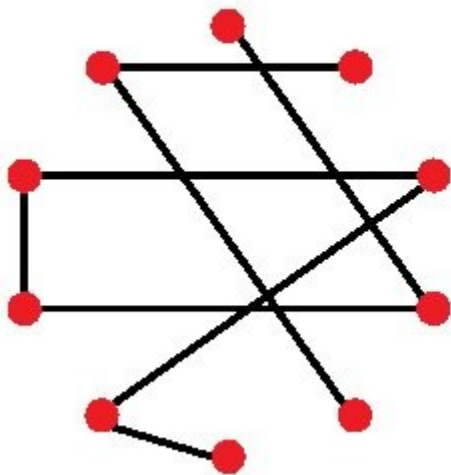
- **Grafos aleatorios**: se construyen generando enlaces aleatoriamente entre un conjunto de N vértices.
- Si los enlaces se generan con una determinada probabilidad p se generaría un tipo particular de grafo aleatorio conocido como **modelo de Erdős-Renyi**.

Modelo de Erdős-Renyi (1)

- Dos variantes:
 - $G_{n,p}$: grafo no dirigido con n nodos y cada enlace (u,v) aparece independiente e idénticamente distribuido con probabilidad p
 - $G_{n,m}$: grafo no dirigido con n nodos y m enlaces tomados al azar uniformemente.

Modelo de Erdős-Renyi (2)

- ¿qué tipo de redes producen estos modelos?
- N y p no determinan el grafo de forma unívoca



$n = 10$
 $p = 1/6$

Propiedades de G_{np} (1)

- La distribución de grados de G_{np} es binomial
- $P(k)$ denota la fracción de nodos con grado k :

$$p(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

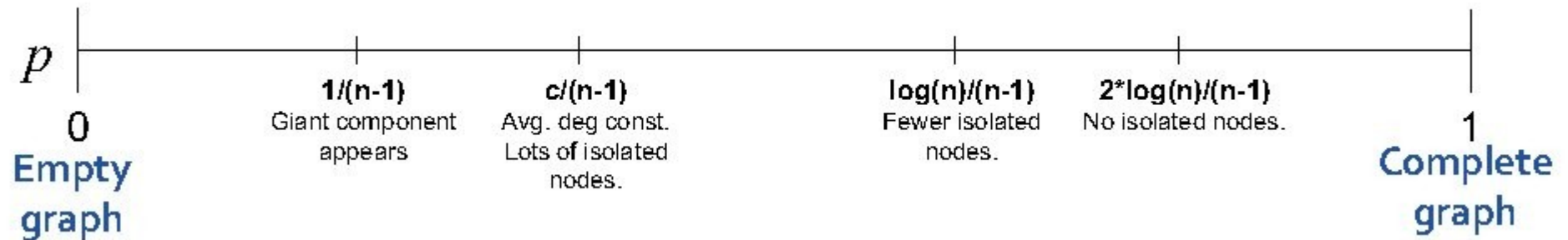
- p : prob de que un nodo tenga k enlaces
- Primer factor: formas de seleccionar k nodos de entre $n - 1$ (con los que se puede enlazar un nodo).
- Ultimo factor: prob de que le falten $n - 1 - k$ enlaces

Propiedades de G_{np} (2)

- Longitud del camino $O(\log n)$
- Coeficiente de clustering $C = p$

Evolución de la estructura de G_{np}

- Evolución de la estructura de G_{np} con p :



- $c = p(n-1)$

Correspondencia de Gnp con redes reales

- ¿Son las redes reales grafos aleatorios? **NO!**
 - Tienen un componente gigante y la longitud media de los caminos es similar **pero...**
 - El coeficiente de clustering y la distribución de grados son diferentes!
 - El coeficiente de clustering de las redes aleatorias es muy bajo al no tener estructura local.
- **Entonces ¿porqué estudiar redes aleatorias?**

Utilidad de las redes aleatorias

- Son un modelo de referencia para el resto de las clases.
- Nos ayudan a calcular muchas cantidades que podemos comparar con datos reales.
- Nos ayudan a estudiar hasta qué punto una propiedad es el resultado de un proceso aleatorio.


Modelo del mundo pequeño

- ¿Hay redes con caminos cortos y un coeficiente de clustering alto?
- Experimento de Milgram ([Milgram '67])
 - ¿cual es al camino más corto entre cualquier par de personas en una red social?

Experimento de Milgram (1)

- Se tomaron 300 personas de Omaha, Nebraska y Wichita, Kansas.
- Se les pidió que hicieran llegar una carta a un corredor de bolsa de Boston pasandola entre amigos.
- ¿cuantos pasos se necesitaron en media?

Experimento de Milgram (2)

- Se completaron 64 entregas
- Se necesitó una media de 6,2 pasos 
¡ 6 grados de separación!
- Algunas críticas al experimento
 - El número de muestras es pequeño
 - La gente podía tener información adicional ...
- Pero se ha confirmado en otros experimentos

Estudio de Columbia (2003)

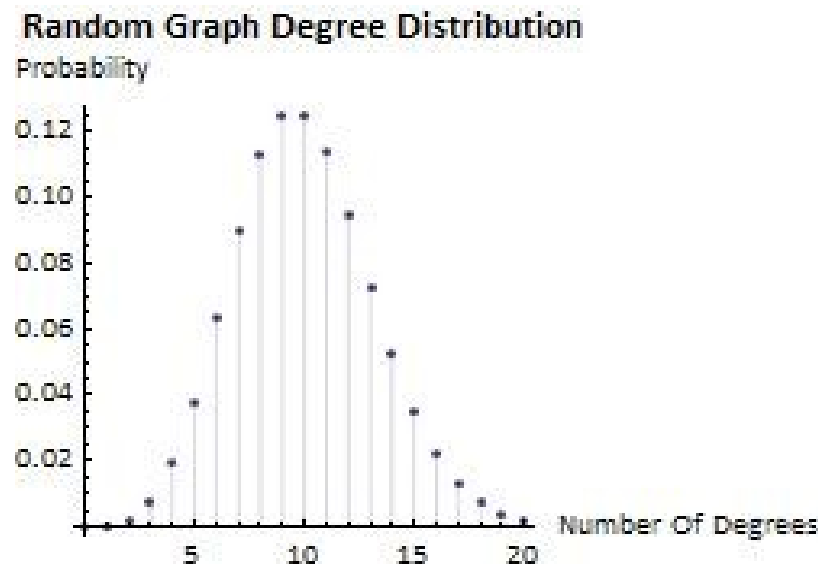
- Correo electrónico
- 18 destinos de distinto tipo
- 24000 puntos de partida
- 65% de abandono por paso
- Se completaron 384 entregas
- Longitud media de la cadena: 4.01

No tan inesperado!

- Si cada persona está conectada a otras 100:
 - 1 paso: conecta con 100 personas
 - 2 pasos: con 10000
 - 3 pasos: 1000000
 - 4 pasos: 100M
 - 5 pasos: 10 billones

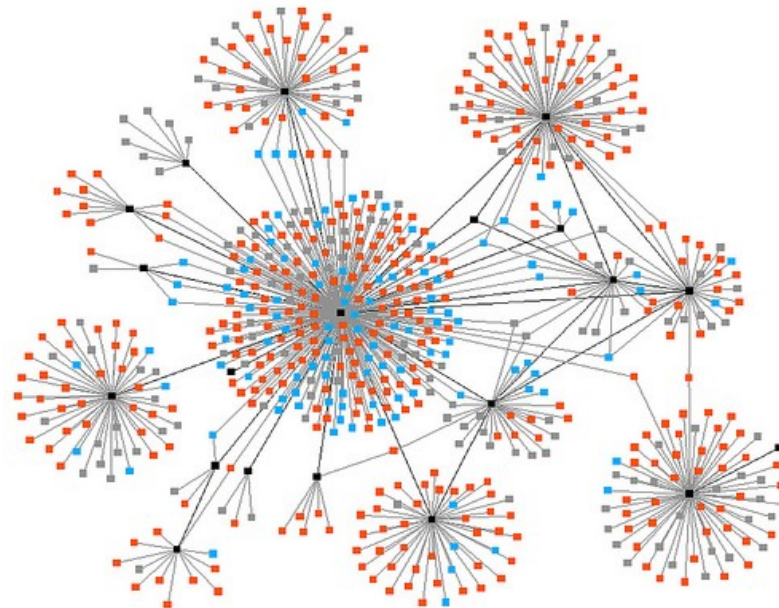
Resumen propiedades:

- **Redes Aleatorias:**
 - Distribución de grados binomial, Poisson
 - Distancia media entre nodos pequeña
 - Coeficiente de clustering muy pequeño (para redes grandes se anula)



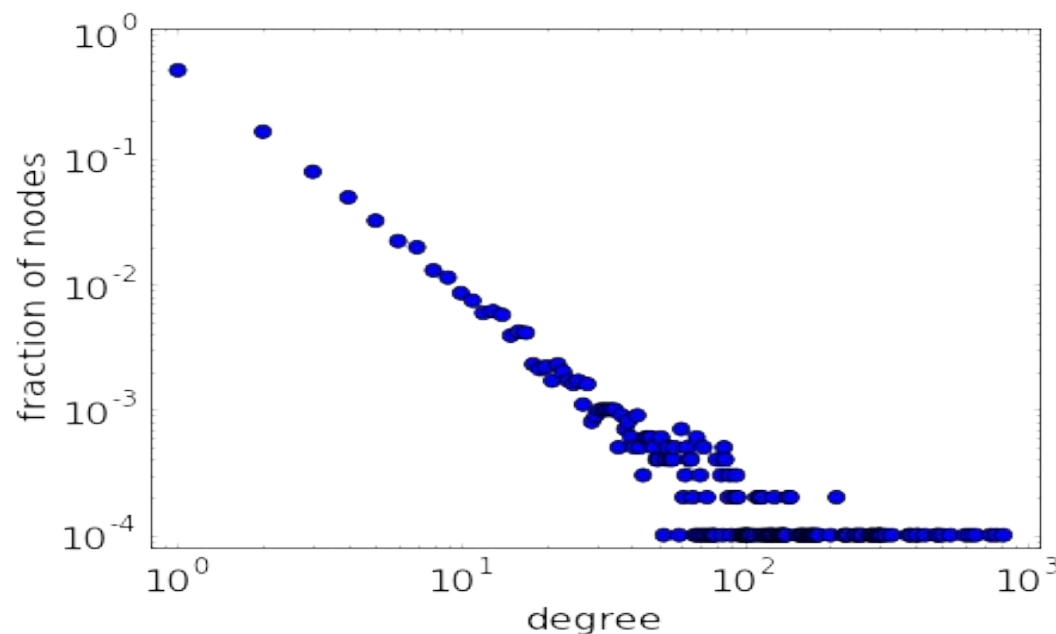
Resumen propiedades:

- **Mundos pequeños:**
 - Diámetro de la red pequeño
 - Distancia media entre nodos pequeña ($\approx \ln(N)$)
 - Coeficiente de clustering \gg Coef. en redes aleatorias



Resumen propiedades:

- Redes libres de escala (scale-free):
 - Distribución de grados sigue ley de potencias
 - Se ve más claramente en un representación en escala log-log



Resumen propiedades:

- ¿Puede una red libre de escala ser un mundo pequeño?
 - Si, depende del exponente asociado a la red

Referencias

- Networks, Crowds, and Markets:
Reasoning About a Highly Connected World
David Easley and Jon Kleinberg
<http://www.cs.cornell.edu/home/kleinber/networks-book/>
Capítulo 2, 13, 20
- Milgram, Stanley. (1963). "Behavioral Study of Obedience". Journal of Abnormal and Social Psychology 67, 371-378.
http://www.radford.edu/~jaspelme/_private/gradsoc_articles/obedience/Migram_Obedience.pdf

Subgrafos y particiones

- Un **subgrafo** $G'=(V',L')$ de un grafo $G=(V,L)$ con $V' \subseteq V$ y $L' \subseteq L$ con V' conteniendo todos los nodos de los extremos de L' .
- Un subgrafo es un **subgrafo de recubrimiento** (spanning) si y solo si $V' = V$.