



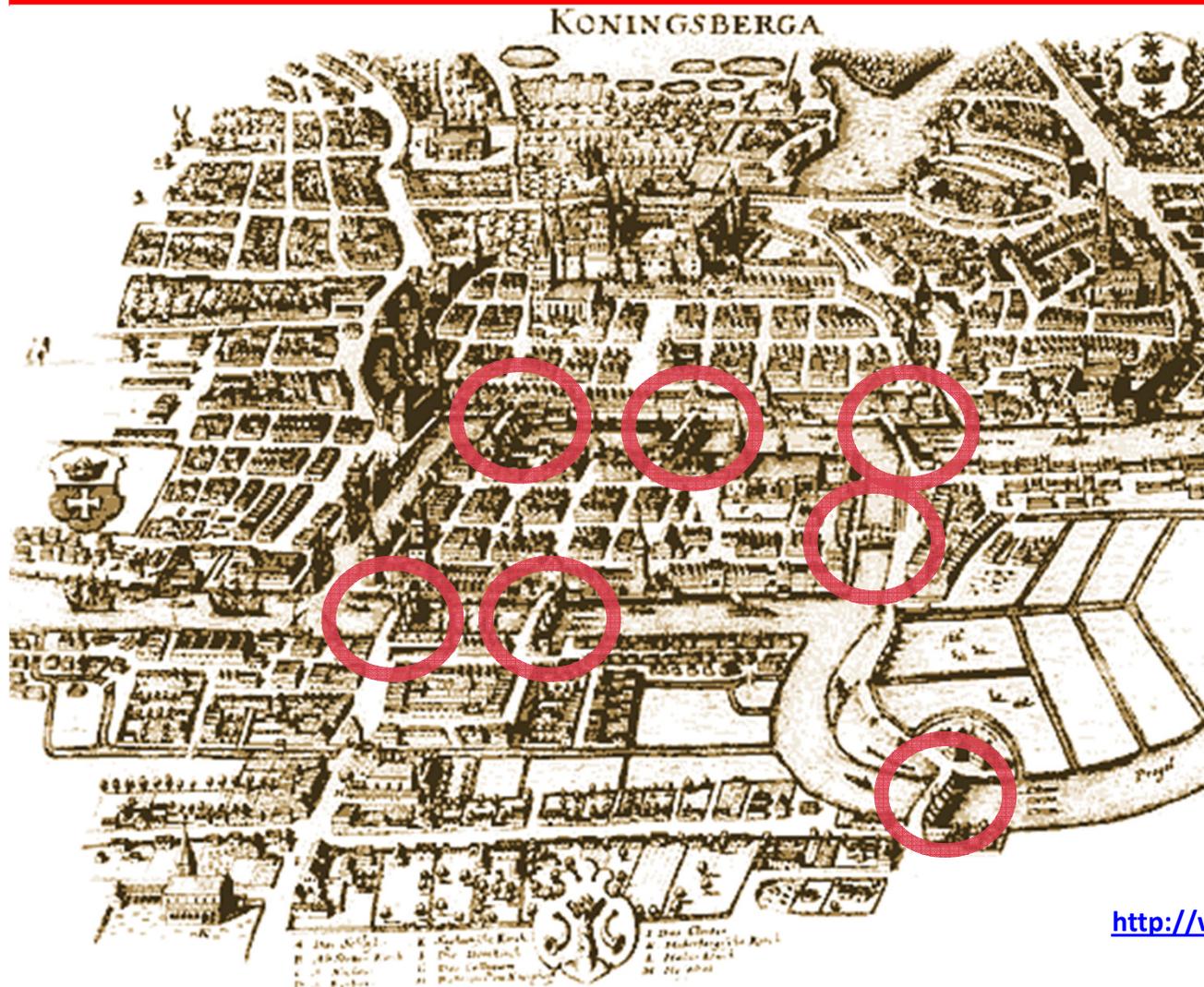
Redes y Sistemas Complejos
Cuarto Curso del Grado en Ingeniería Informática

Tema 2: Aspectos Básicos y Propiedades Estructurales de las Redes

Oscar Cordon García

Dpto. Ciencias de la Computación e Inteligencia Artificial
ocordon@decsai.ugr.es

LOS PUENTES DE KONIGSBERG (1)



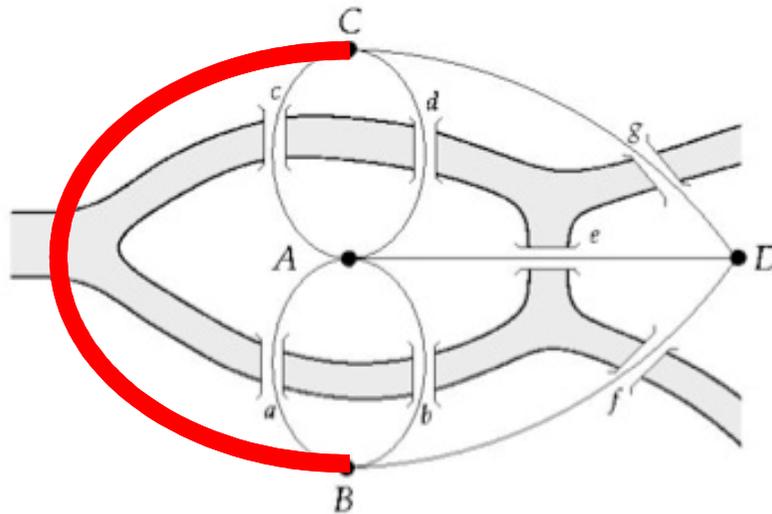
¿Es posible recorrer los siete puentes sin pasar dos veces por el mismo?



Kaliningrado, en la actualidad

<http://www.numericana.com/answer/graphs.htm>

LOS PUENTES DE KONIGSBERG (2)



¿Es posible recorrer los siete puentes sin pasar dos veces por el mismo?

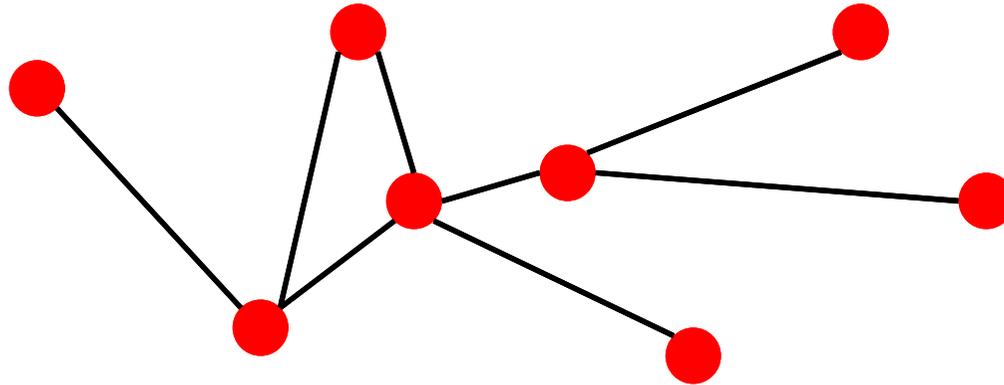
1735: Teorema de Leonhard Euler:

- (a) Si el grafo tiene más de dos nodos de grado impar, no existe dicho camino
- (b) Si el grafo es conexo y no tiene más de dos nodos de grado impar, al menos hay un camino

Nacimiento de la Teoría de Grafos: Primer problema resuelto usando un grafo. Prueba de que hay problemas que son más sencillos de resolver si se representan en forma de grafo

Teoría de Redes: La existencia del camino es una propiedad intrínseca del grafo. Las redes tienen ciertas propiedades ocultas en su estructura que limitan o mejoran su comportamiento

COMPONENTES DE UN SISTEMA COMPLEJO



- **componentes:** nodos, vértices N
- **interacciones:** enlaces, arcos L
- **sistema:** red, grafo (N,L)

NOMENCLATURA: ¿REDES O GRAFOS?

red se suele referir a sistemas reales

- www: red de páginas conectadas por URLs,
- red social: grupo de individuos conectados por una relación familiar, de amistad o profesional,
- red metabólica: conjunto de las reacciones químicas que tienen lugar en una célula

Nomenclatura: (Red, nodo, enlace)

grafo: representación matemática de una red

- grafo de la web,
- grafo social (término de Facebook)

Nomenclatura: (Grafo, vértice, arco)

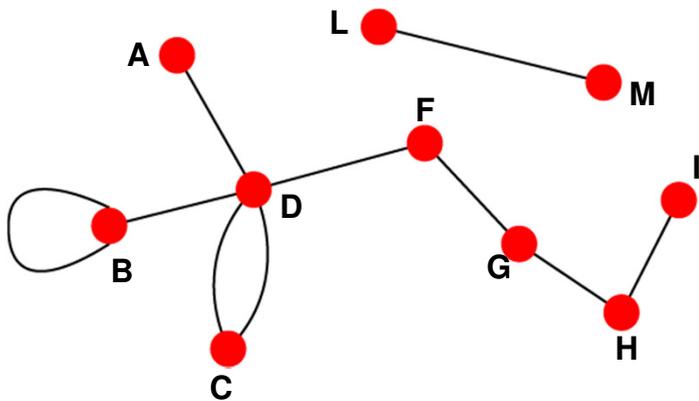
Haremos esta distinción cuando sea necesario pero en la mayoría de los casos usaremos los dos términos indistintamente

REDES DIRIGIDAS Y NO DIRIGIDAS

No dirigidas

Enlaces: no dirigidos (*simétricos*)

Grafo:



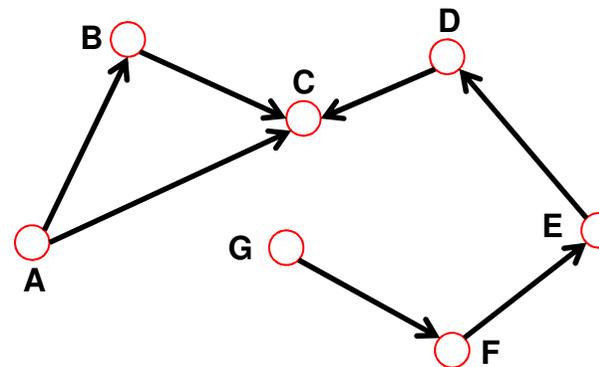
Enlaces no dirigidos :

- enlaces de coautoría
- red de actores de Hollywood
- interacciones entre proteínas
- relaciones de amistad en Facebook

Dirigidas

Enlaces: dirigidos

Digrafo = grafo dirigido:

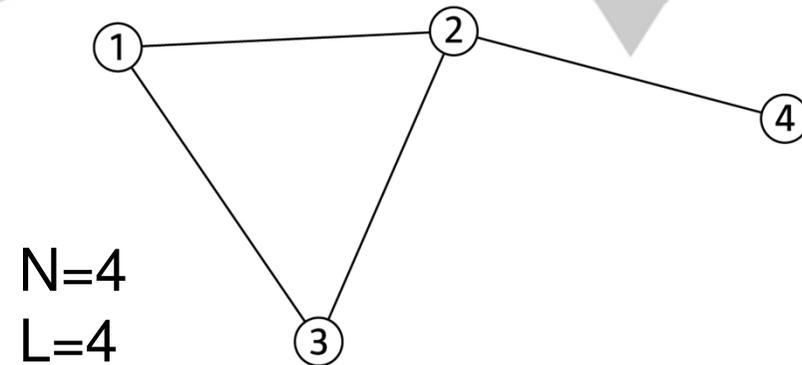
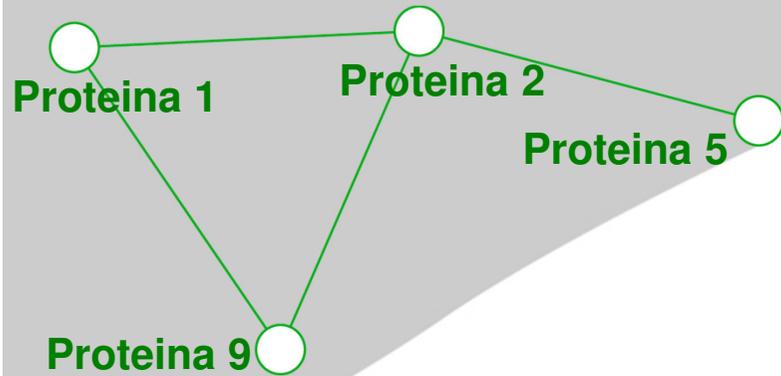
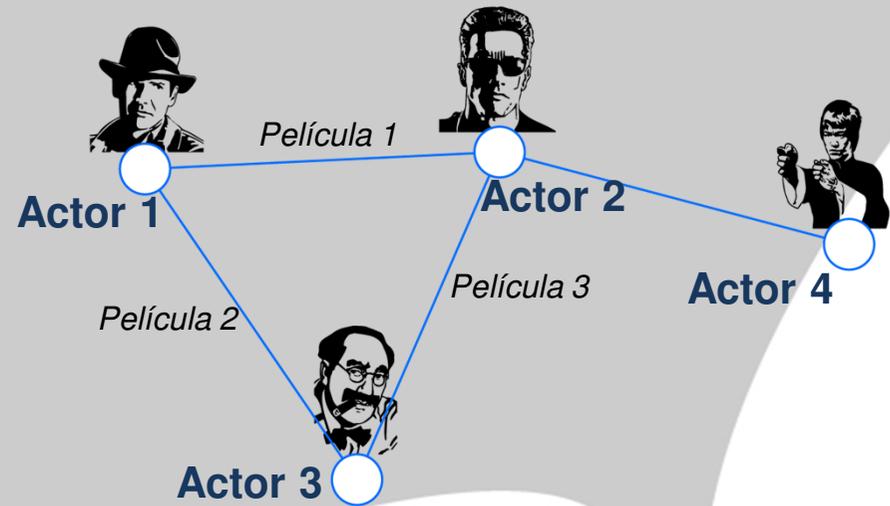
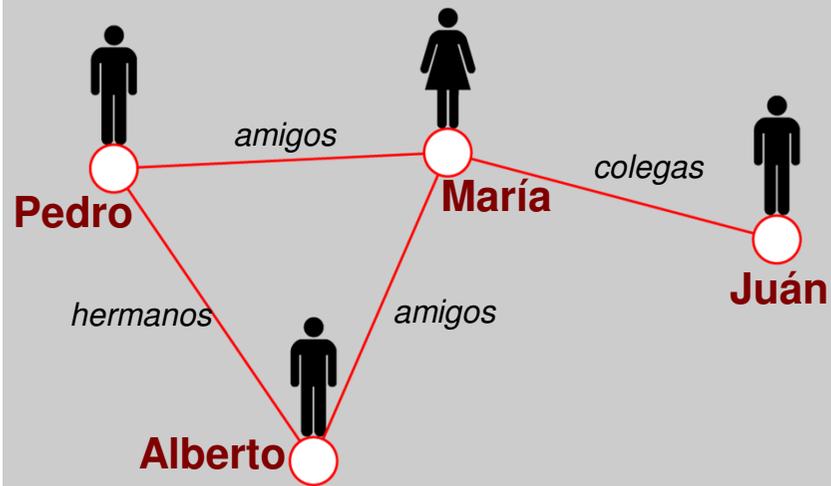


Un enlace no dirigido es la unión de dos enlaces dirigidos opuestos

Enlaces dirigidos :

- URLs en la WWW
- llamadas de teléfono
- reacciones metabólicas
- relaciones de seguimiento en Twitter

LAS REDES OFRECEN UN LENGUAJE COMÚN PARA ESTUDIAR SISTEMAS



LA IMPORTANCIA DE ELEGIR UNA REPRESENTACIÓN ADECUADA (1)

La elección de la representación adecuada para la red determina nuestra habilidad para usar la Teoría de Redes con éxito

En algunos casos, existe una representación única, sin ambigüedad

En otros, la representación no es única, ni mucho menos

Por ejemplo, la forma de asignar los enlaces entre un grupo de individuos determinará la naturaleza de la cuestión a estudiar

LA IMPORTANCIA DE ELEGIR UNA REPRESENTACIÓN ADECUADA (2)

They Rule

Josh On (2004)
<http://www.theyrule.net>

**si conectamos
individuos que trabajen
con otros, exploraremos
redes profesionales**



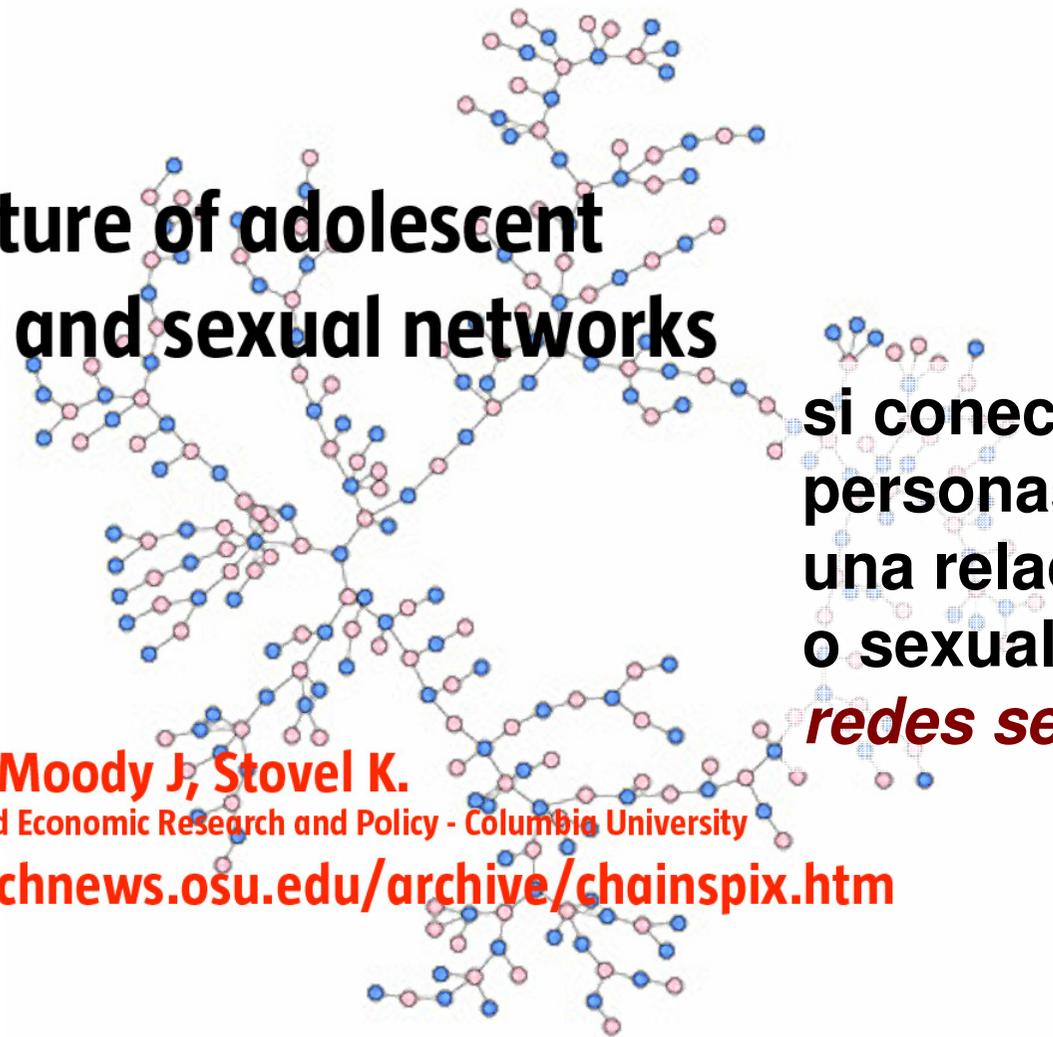
The structure of adolescent romantic and sexual networks

si conectamos
personas que tengan
una relación romántica
o sexual, exploraremos
redes sexuales

Bearman PS, Moody J, Stovel K.

Institute for Social and Economic Research and Policy - Columbia University

<http://researchnews.osu.edu/archive/chainspix.htm>



LA IMPORTANCIA DE ELEGIR UNA REPRESENTACIÓN ADECUADA (4)

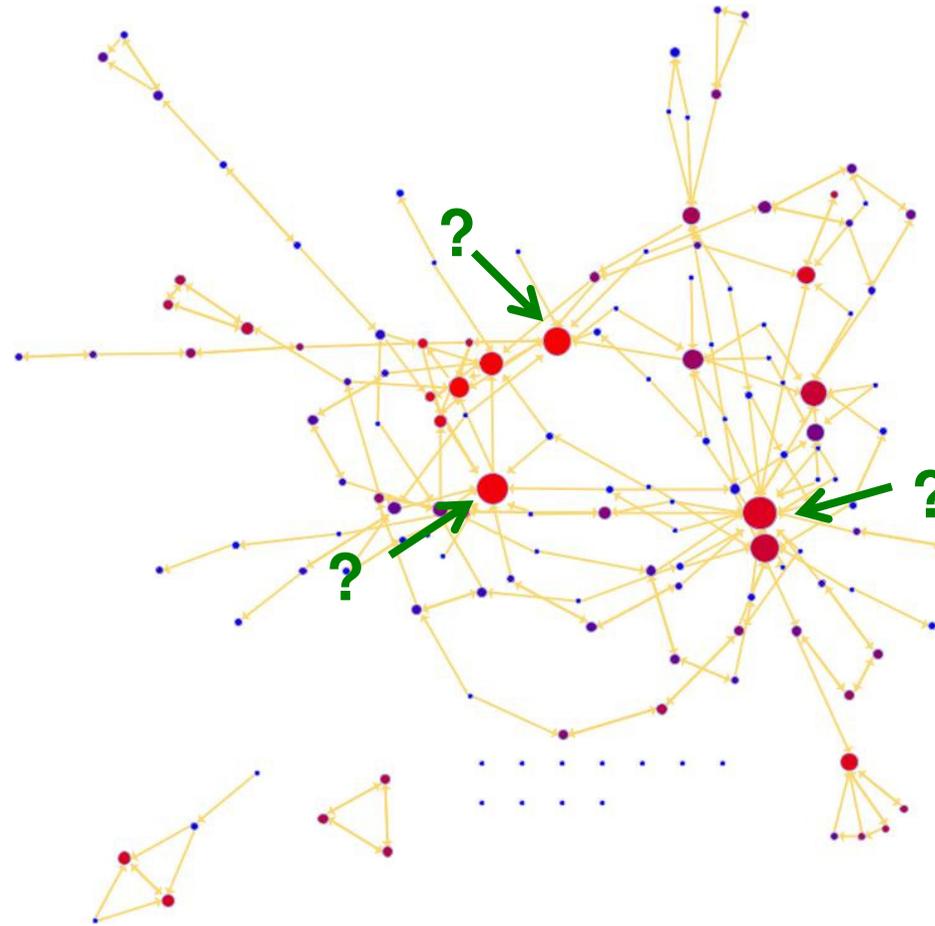
Si conectáramos individuos en función de su nombre propio (*todos los Pedros conectados entre sí*), ¿qué querríamos explorar? No tendría mucho sentido...

Aún así, **sería una red**

MEDIDAS EN REDES: EL GRADO

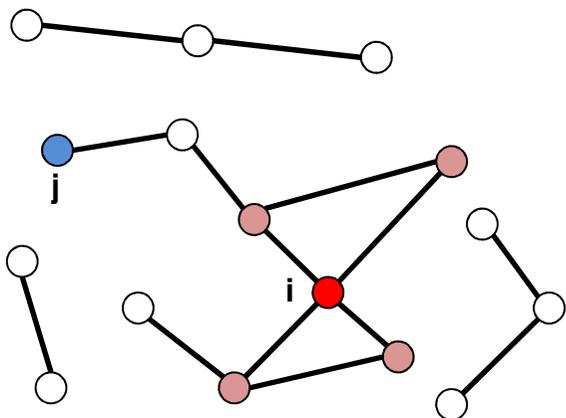
GRADO DE UN NODO (2): ¿PARA QUÉ SIRVE?

Medida de la importancia de un nodo individual en la red: ¿qué nodo tiene más enlaces?



GRADO MEDIO: MEDIDA GLOBAL

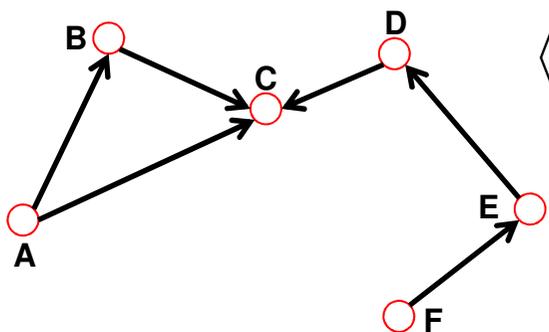
No dirigida



$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i \quad \langle k \rangle \equiv \frac{2L}{N}$$

N – número de nodos de la red
L – número de enlaces de la red

Dirigida



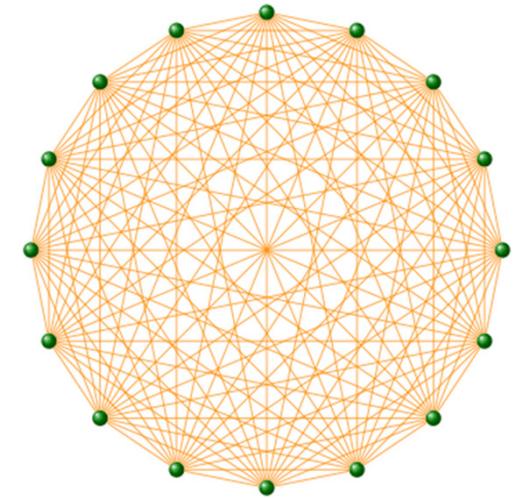
$$\langle k^{in} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

$$\langle k \rangle \equiv \frac{L}{N}$$

GRAFO COMPLETO

El número máximo de enlaces que puede tener una red de N nodos es:

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



Un grafo con $L=L_{\max}$ se denomina **grafo completo** y su grado promedio es **$\langle k \rangle = N-1$**

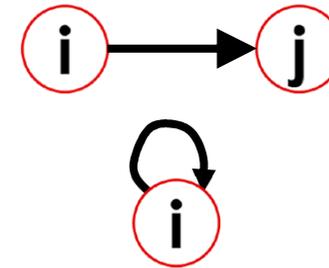
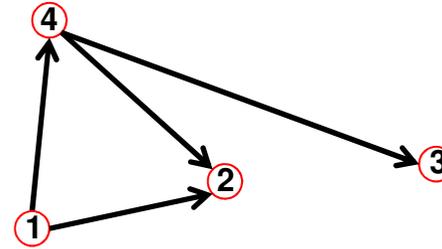
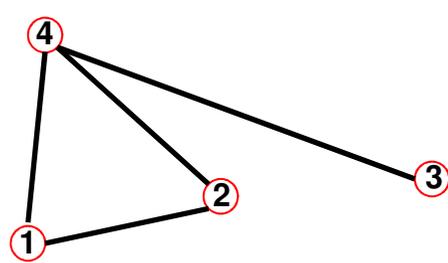
LAS REDES REALES SON DISPERSAS

La mayoría de las redes observadas en sistemas reales son dispersas:

$$L \ll L_{\max} \quad 0 < \langle k \rangle \ll N-1$$

NETWORK NAME	NODES	LINKS	DIRECTED/ UNDIRECTED	N	L	$\langle K \rangle$
Internet	routers	Internet Connections	Undirected	192,244	609,066	2.67
WWW	webpages	links	Directed	325,729	1,497,134	4.60
Power Grid	power plants, transformers	cables	Undirected	4,941	6,594	2.67
Mobile-Phone Calls	subscribers	calls	Directed	36,595	91,826	2.51
Email	email addresses	emails	Directed	57,194	103,731	1.81
Science Collaboration	scientists	co-authorships	Undirected	23,133	186,936	16.16
Actor Network	actors	co-acting	Undirected	212,250	3,054,278	28.78
Citation Network	papers	citations	Directed	449,673	4,707,958	10.47
E. coli Metabolism	metabolites	chemical reactions	Directed	1,039	5,802	5.84
Yeast Protein Interactions	proteins	binding interactions	Undirected	2,018	2,930	2.90

REPRESENTACIONES DE UNA RED: MATRIZ DE ADYACENCIA



$A_{ij}=1$ si existe un enlace entre los nodos i y j

$A_{ij}=0$ si los nodos i y j no están conectados

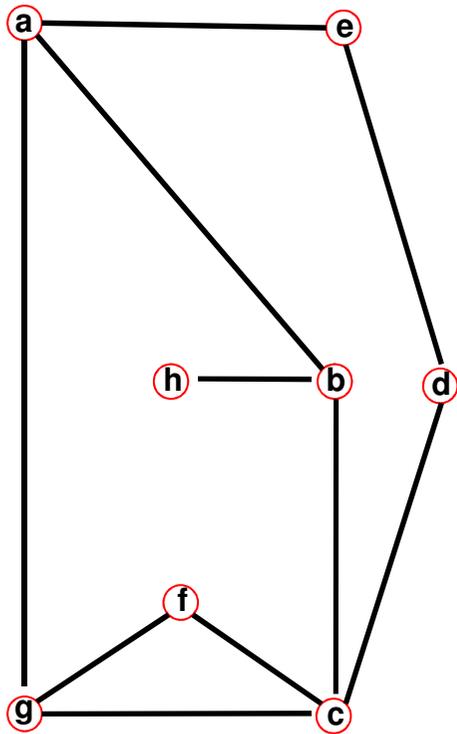
$A_{ii}=0$, a menos que la red tenga **auto-enlaces**

$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

En redes dirigidas (derecha) la matriz no es simétrica

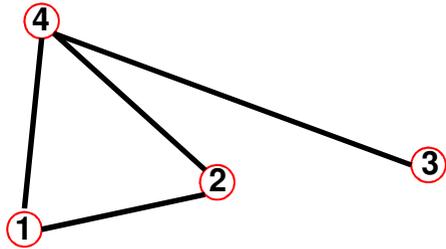
EJEMPLO DE MATRIZ DE ADYACENCIA



	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	0	0	0
h	0	1	0	0	0	0	0	0

MATRIZ DE ADYACENCIA Y CÁLCULO DEL GRADO DE LOS NODOS

No dirigida



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} = A_{ji}$$

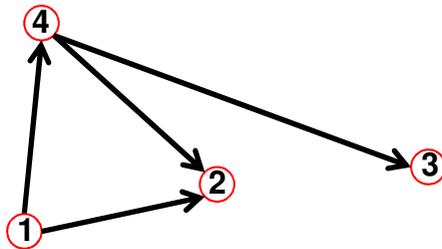
$$A_{ii} = 0$$

$$k_i = \sum_{j=1}^N A_{ij}$$

$$k_j = \sum_{i=1}^N A_{ij}$$

$$L = \frac{1}{2} \sum_{i=1}^N k_i = \frac{1}{2} \sum_{ij} A_{ij}$$

Dirigida



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ij} \neq A_{ji}$$

$$A_{ii} = 0$$

$$k_i^{in} = \sum_{j=1}^N A_{ji}$$

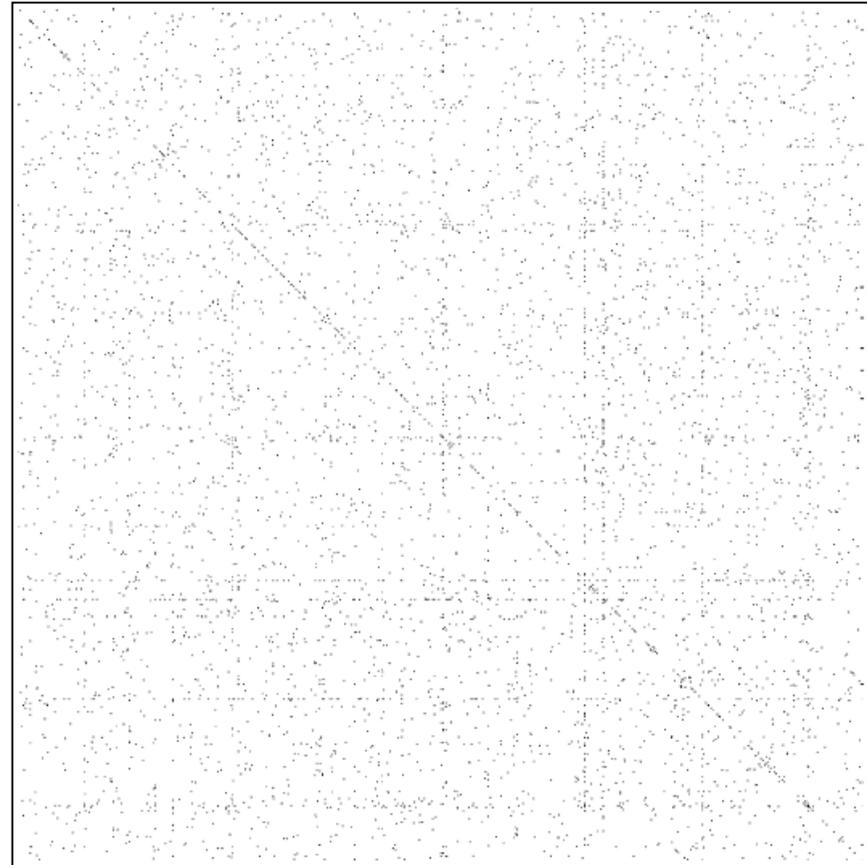
$$k_j^{out} = \sum_{i=1}^N A_{ij}$$

$$L = \sum_{i=1}^N k_i^{in} = \sum_{j=1}^N k_j^{out} = \sum_{i,j} A_{ij}$$

PROBLEMA DE LA REPRESENTACIONES POR MATRIZ DE ADYACENCIA

El problema principal de esta representación es que las redes reales suelen ser muy dispersas

De este modo, el uso de una matriz de adyacencia gasta mucha memoria de forma innecesaria

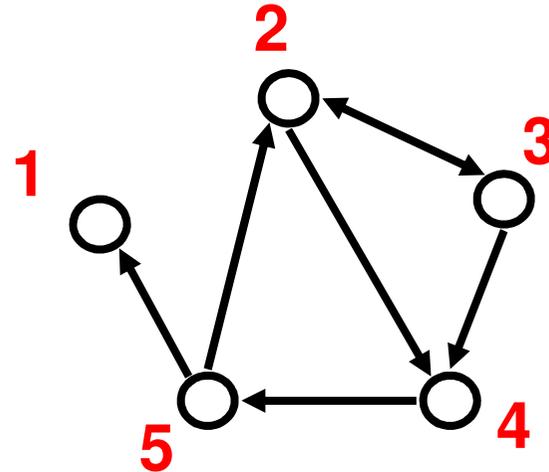


Matriz de adyacencia de la red de interacciones entre proteínas (N=2018)

REPRESENTACIONES DE UNA RED: LISTA DE ENLACES

Lista de enlaces (sólo se almacenan los elementos para los que $A_{ij} \neq 0$):

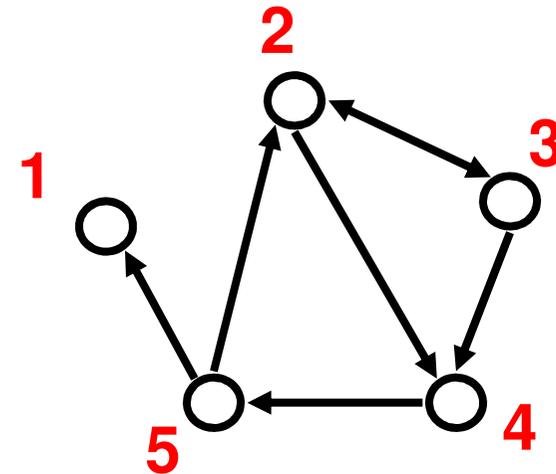
- 2, 3
- 2, 4
- 3, 2
- 3, 4
- 4, 5
- 5, 2
- 5, 1



REPRESENTACIONES DE UNA RED: LISTA DE ADYACENCIAS

Lista de adyacencias:

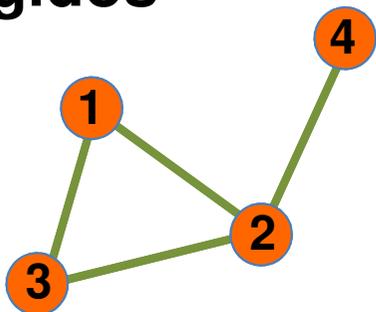
- 1:
- 2: 3 4
- 3: 2 4
- 4: 5
- 5: 1 2



Tiene la ventaja adicional de que se pueden recuperar los vecinos de cada nodo de una forma muy rápida

GRAFOS Y REDES

No dirigidos



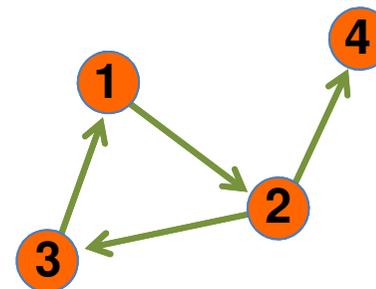
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

*Red de actores de Hollywood,
Redes de interacciones entre proteínas*

Dirigidos



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

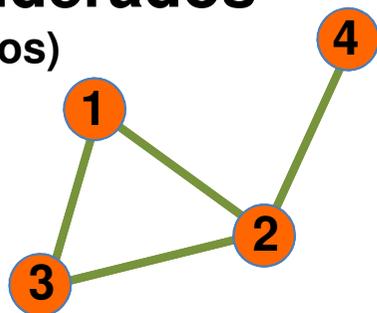
$$A_{ii} = 0 \quad A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{L}{N}$$

*WWW
Redes de citas en artículos científicos*

No ponderados

(no dirigidos)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

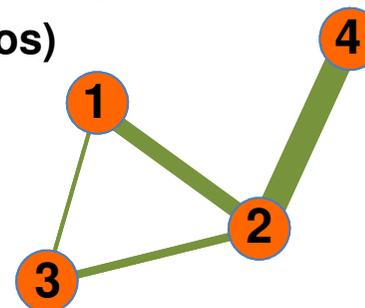
$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij} \quad \langle k \rangle = \frac{2L}{N}$$

Redes de interacciones entre proteínas
WWW

Ponderados

(no dirigidos)



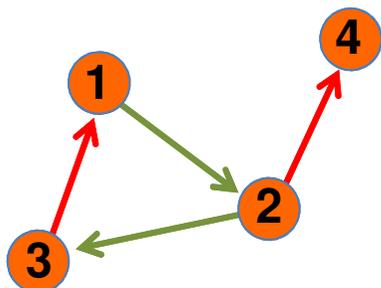
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij}) \quad \langle k \rangle = \frac{2L}{N}$$

Redes de llamadas
Redes metabólicas

Ponderados con signo (dirigidos)

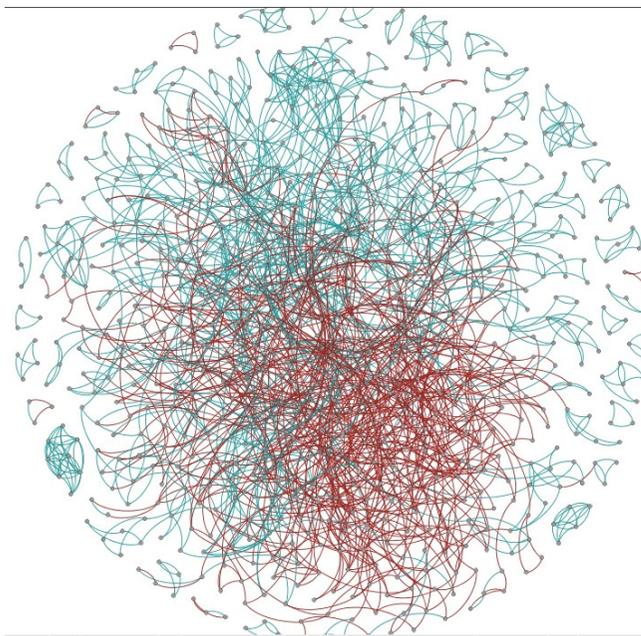


$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0 \quad A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^N |A_{ij}| \quad \langle k \rangle = \frac{L}{N}$$

Redes de confianza



Muestreo de las opiniones positivas y negativas de la red Epinions

<http://blogs.cornell.edu/info2040/2011/09/28/epinions-com-and-the-web-of-trust/>

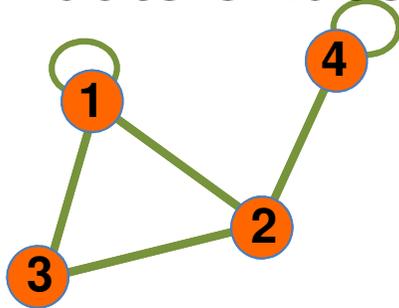
<http://snap.stanford.edu/data/soc-sign-epinions.html>

Opinión de la confianza de una persona sobre otra en el portal *Epinions.com*

RETO: Determinar cómo se propagan los sentimientos negativos en la red. ¿Es mi amigo el enemigo de mi enemigo?

Grafo con auto-enlaces

(no dirigido)



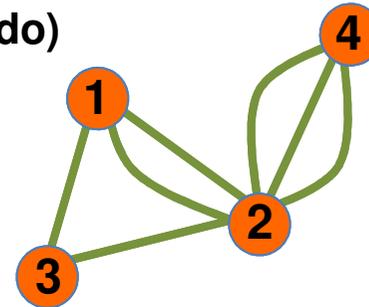
$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$L = \frac{1}{2} \sum_{i,j=1, i \neq j}^N A_{ij} + \sum_{i=1}^N A_{ii} \quad A_{ij} \neq 0 \quad A_{ij} = A_{ji} \quad ?$$

Redes de interacciones entre proteínas
WWW

Multigrafo

(no dirigido)



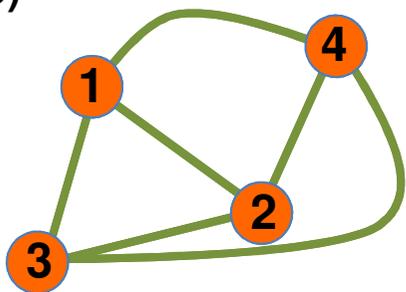
$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N \text{nonzero}(A_{ij}) \quad A_{ij} = 0 \quad A_{ij} = A_{ji} \quad \langle k \rangle = \frac{2L}{N}$$

Redes sociales
Redes de colaboración

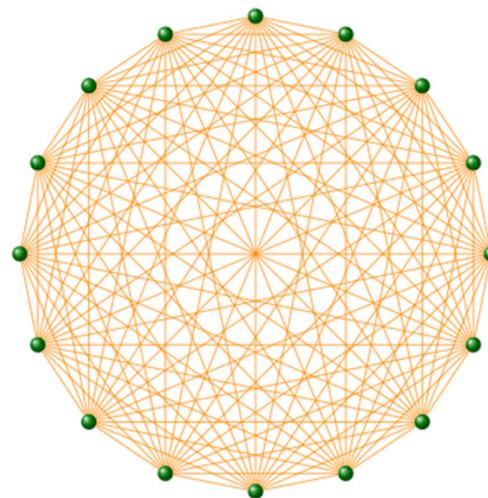
Grafo completo

(no dirigido)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$L = L_{\max} = \frac{N(N-1)}{2} \quad A_{ii} = 0 \quad A_{i \neq j} = 1 \quad \langle k \rangle = N-1$$



WWW > multigrafo dirigido con auto-enlaces

Interacciones entre proteínas > no dirigido, no ponderado, con auto-enlaces

Colaboraciones > multigrafo no dirigido o ponderado

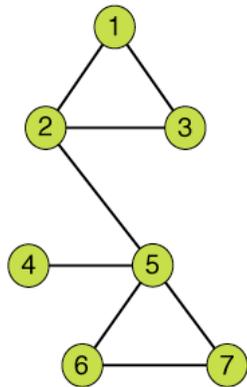
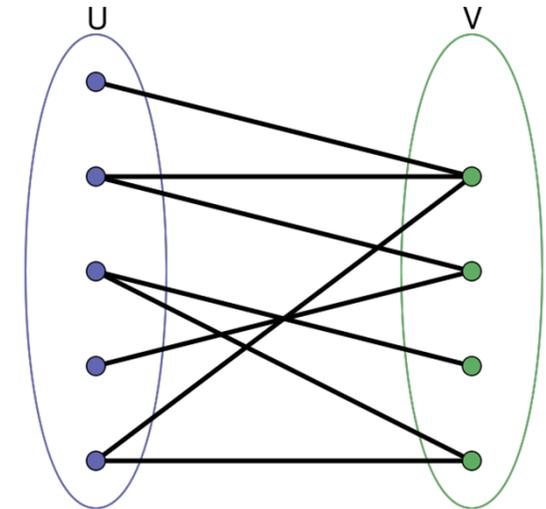
Llamadas a teléfonos móviles > dirigido, ponderado.

Relaciones de amistad en Facebook > no dirigido, no ponderado

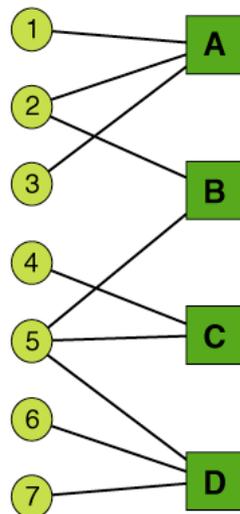
GRAFOS BIPARTITOS

Un **grafo bipartito** (o **bigrafo**) es un grafo cuyos nodos pueden dividirse en dos conjuntos disjuntos U y V de tal modo que cada enlace conecta un nodo de U con otro de V ; es decir, U y V son conjuntos independientes

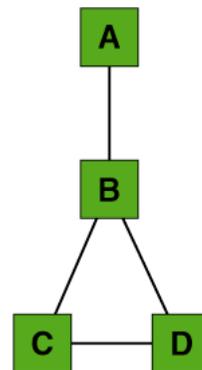
Se pueden generar **dos proyecciones** de cada red bipartita. La primera conecta los nodos de U que están conectados al mismo nodo de V . La segunda hace lo mismo con los nodos de V



proyección U



red bipartita



proyección V

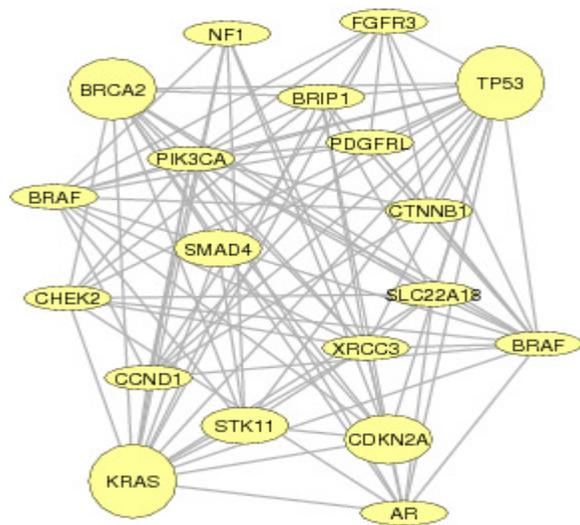
Ejemplos:

Red de actores de Hollywood

Redes de colaboración

Red de enfermedades (*diseasome*)

RED DE GENES – RED DE ENFERMEDADES HUMANAS



**Red de genes
(DGN)**

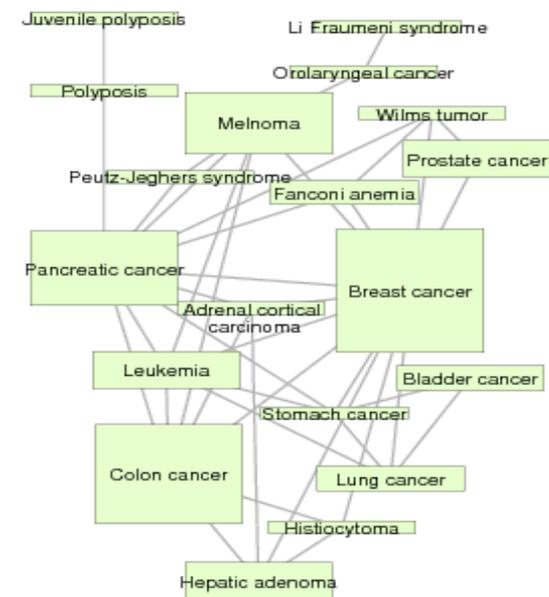
DISEASOME

GENOME

- BRCA2
- BRAF
- CCND1
- CDKN2A
- AR
- CTNNB1
- CHEK2
- FGFR3
- KRAS
- SMAD4
- NF1
- SLC22A18
- PDGFRL
- PIK3CA
- BRIP1
- XRCC3
- BRAF
- TP53
- STK11

PHENOME

- Orolaryngeal cancer
- Li Fraumeni syndrome
- Wilms tumor
- Prostate cancer
- Colon cancer
- Leukemia
- Melanoma
- Fanconi anemia
- Pancreatic cancer
- Bladder cancer
- Breast cancer
- Histiocytoma
- Lung cancer
- Polyposis
- Hepatic adenoma
- Juvenile polyposis
- Stomach cancer
- Adrenal cortical carcinoma
- Peutz-Jeghers syndrome

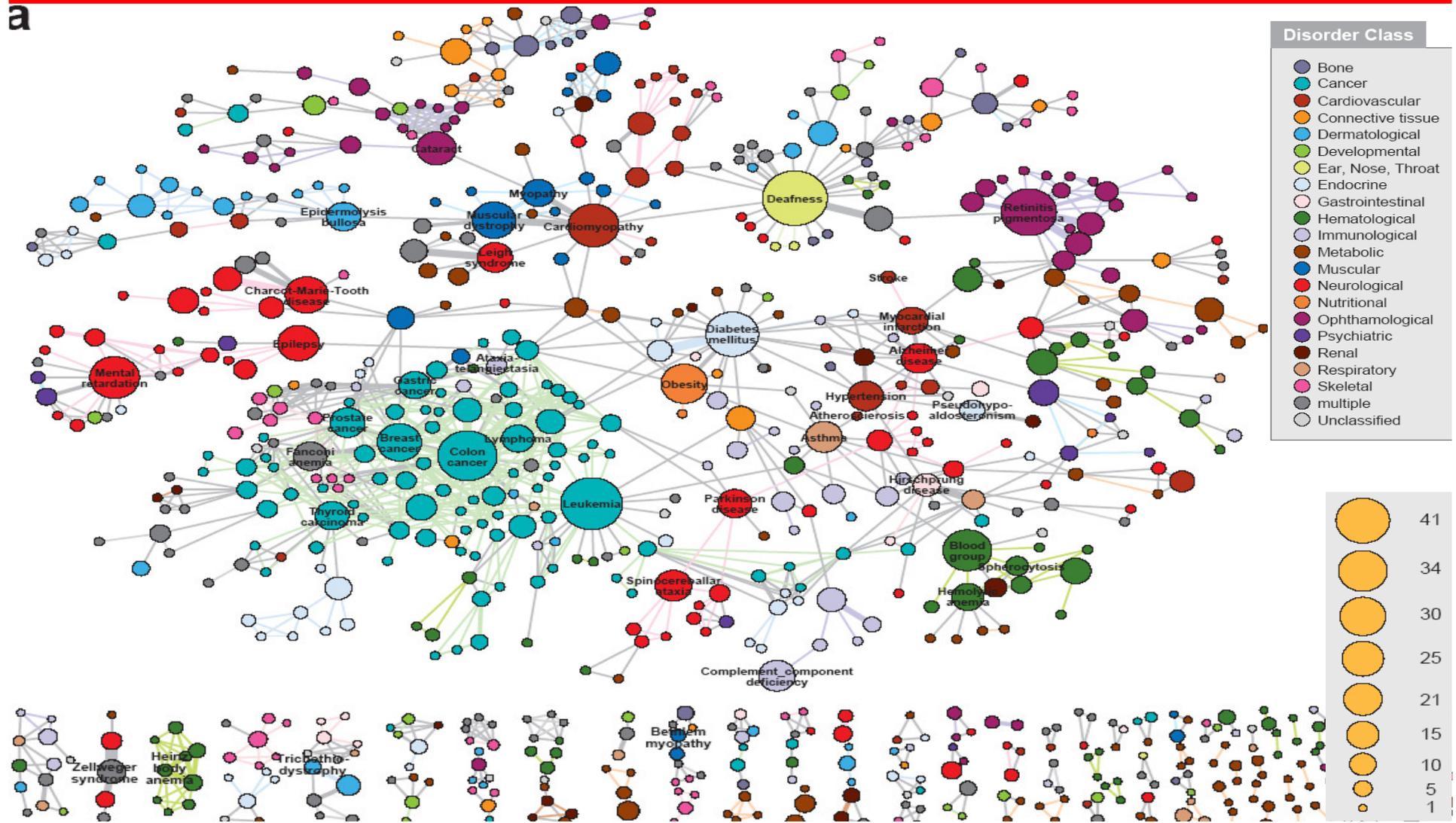


**Red de
enfermedades
(HDN)**

Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

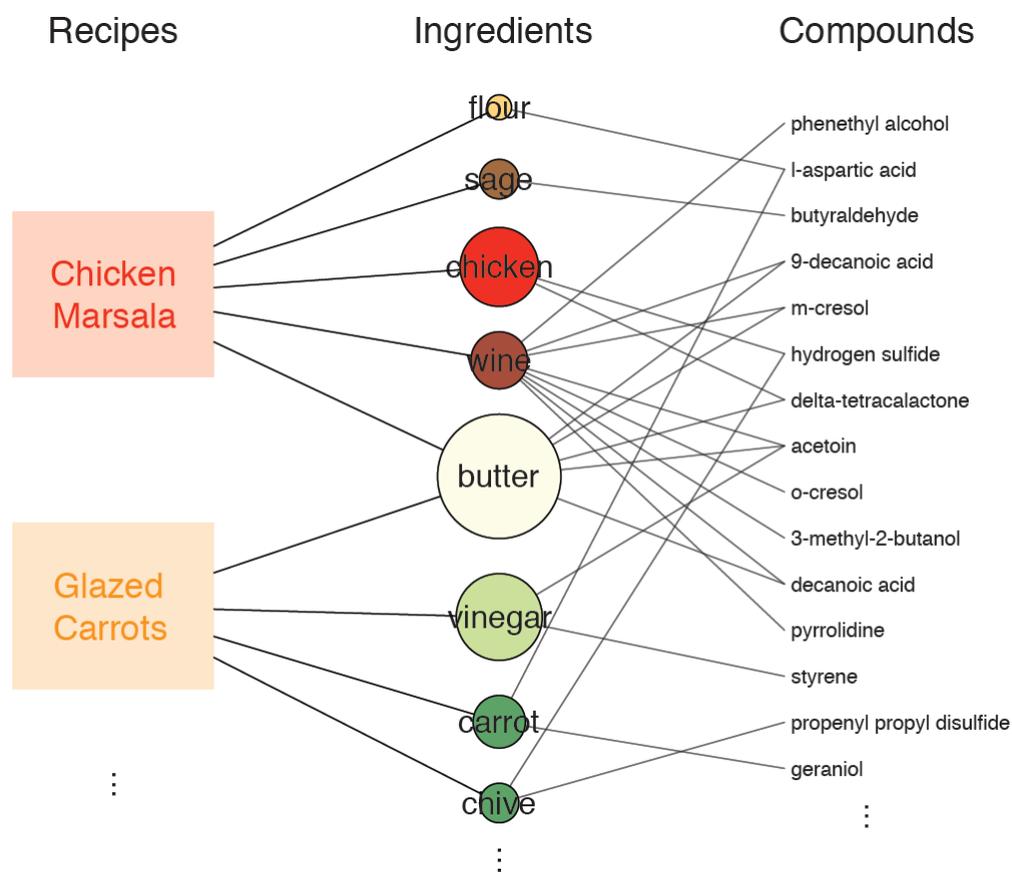
RED DE ENFERMEDADES HUMANAS (HDN)

a

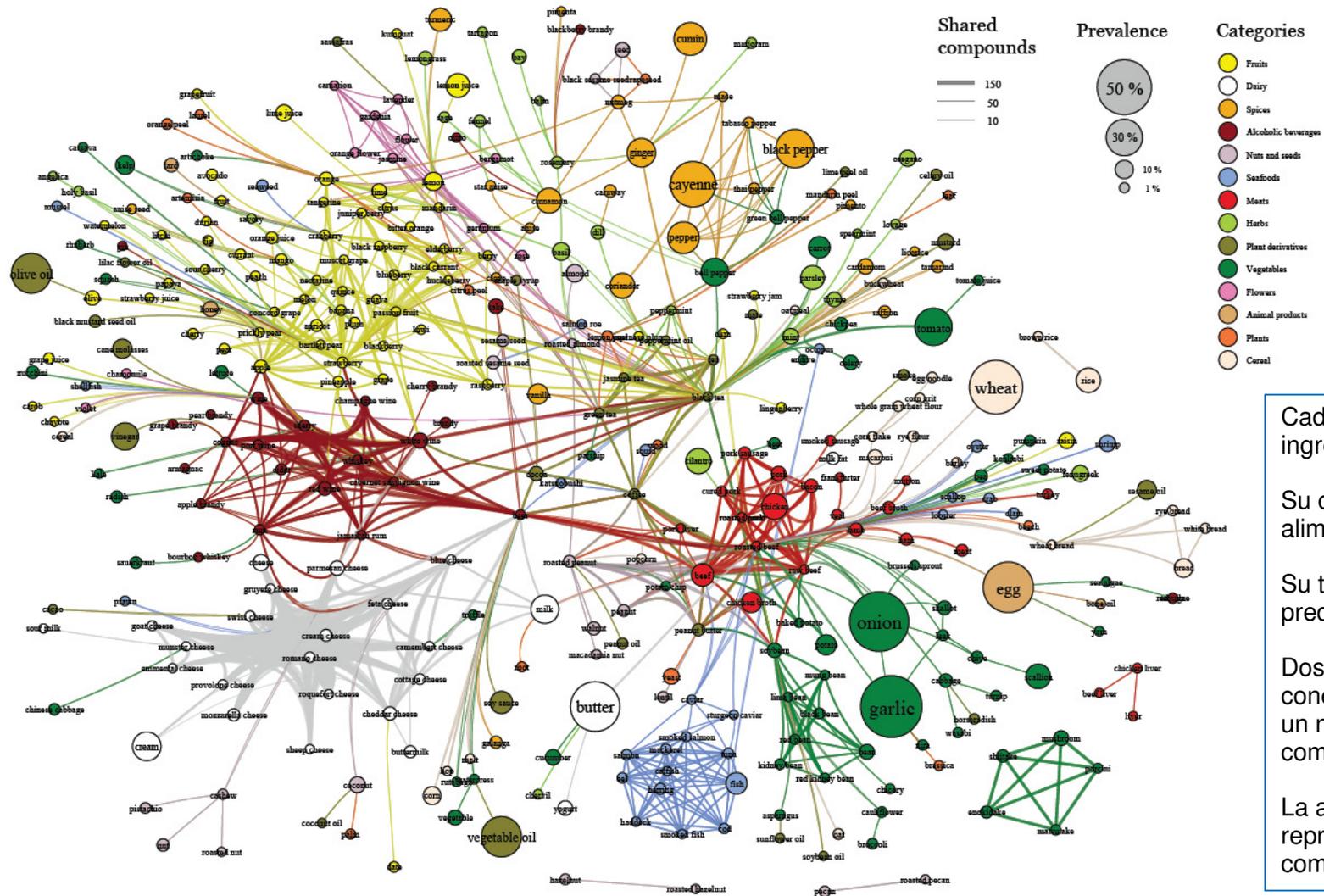


REDES MULTIPARTITAS

También existen las **redes multipartitas**, por ejemplo la red tripartita de recetas-ingredientes-compuestos



RED DE INGREDIENTES (Proyección de la red de recetas-ingredientes-compuestos)



Cada nodo representa un ingrediente

Su color indica el tipo de alimento

Su tamaño indica su predominio en las recetas

Dos ingredientes están conectados si comparten un número significativo de compuestos de sabor

La anchura del enlace representa el número de compuestos compartidos

CAMINOS EN GRAFOS

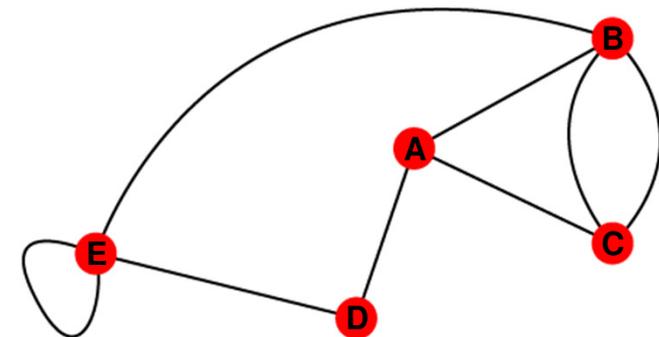
CAMINOS

Un *camino* es una secuencia de nodos en el que cada nodo es adyacente al siguiente

P_{i_0, i_n} denota un camino entre los nodos i_0 e i_n . Es una colección ordenada de $n+1$ nodos y n enlaces. Su longitud es n , el número de enlaces que lo componen

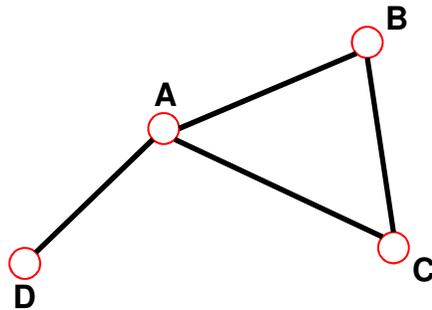
$$P_n = \{i_0, i_1, i_2, \dots, i_n\} \quad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_{n-1}, i_n)\}$$

- Un camino puede tener intersecciones consigo mismo y pasar a través del mismo enlace continuamente. Si un enlace se cruza varias veces, se cuenta en todas ellas
- Un camino válido en el grafo de la derecha podría ser: **ABCBCADEEBA**
- En una red dirigida, el camino tiene que seguir la dirección de los enlaces dirigidos que lo componen



DISTANCIA EN GRAFOS

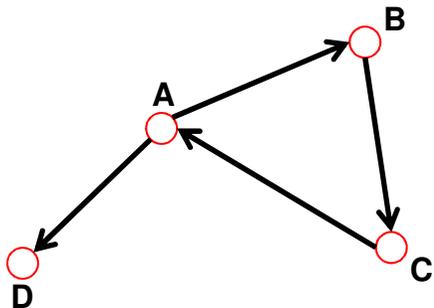
Camino Mínimo



La *distancia (camino mínimo, camino geodésico)* entre dos nodos se define como el número de ejes que contiene el camino más corto que los conecta

* Si no hay ningún camino que conecte los dos nodos, la distancia es infinito.

En **grafos dirigidos** cada camino necesita seguir la dirección concreta de los enlaces dirigidos



Así, en un digrafo, la distancia de un nodo A a otro B suele ser distinta a la distancia de B a A (en el ejemplo, $d_{AB}=1$ y $d_{BA}=2$)

De hecho, uno de los caminos podría existir y el otro no

El número de caminos mínimos entre cualesquiera dos nodos i y j , N_{ij} , se puede calcular del siguiente modo:

$d_{ij}=1$: Si existe un enlace directo entre i y j , entonces $A_{ij}=1$ (en caso contrario, $A_{ij}=0$)

$d_{ij}=2$: Si existe un camino de longitud 2 entre i y j , entonces $A_{ik}A_{kj}=1$ (en caso contrario, $A_{ik}A_{kj}=0$)

El número de caminos de longitud 2 entre i y j es:
$$N_{ij}^{(2)} = \sum_{k=1}^N A_{ik}A_{kj} = [A^2]_{ij}$$

$d_{ij}=d$: En general, si existe un camino de longitud d entre i y j , entonces $A_{ik}\dots A_{lj}=1$ (en caso contrario, $A_{ik}\dots A_{lj}=0$)

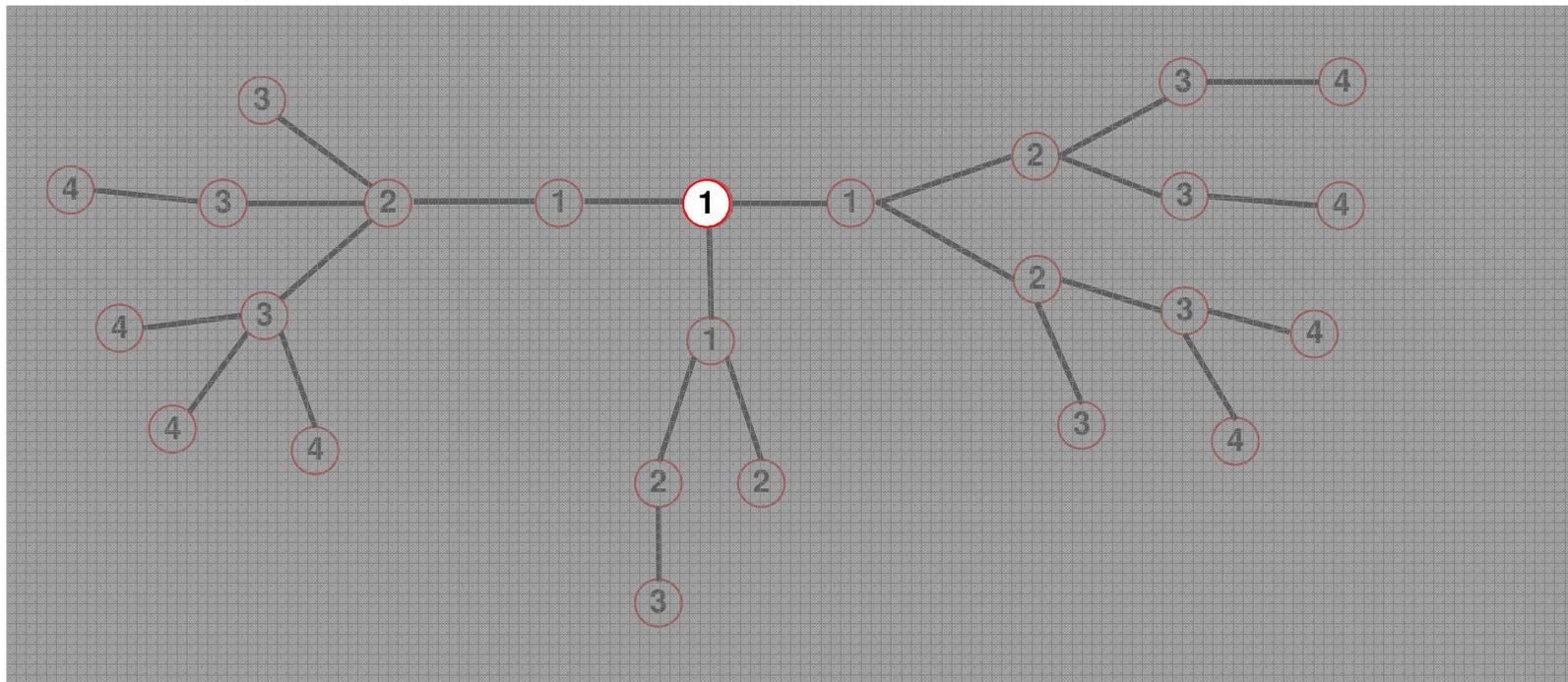
El número de caminos de longitud d entre i y j es:
$$N_{ij}^{(d)} = [A^d]_{ij}$$

- Esta ecuación es válida para redes dirigidas y no dirigidas
- El algoritmo se basa en Programación Dinámica. El algoritmo **Pathfinder** de poda de redes sociales usa un enfoque similar
- De este modo, es **muy costoso para redes grandes**, en las que es más eficiente usar una **búsqueda primero en anchura**

CÁLCULO DE LA DISTANCIA ENTRE DOS NODOS: BÚSQUEDA PRIMERO EN ANCHURA (1)

Distancia entre el nodo 1 y el nodo 4:

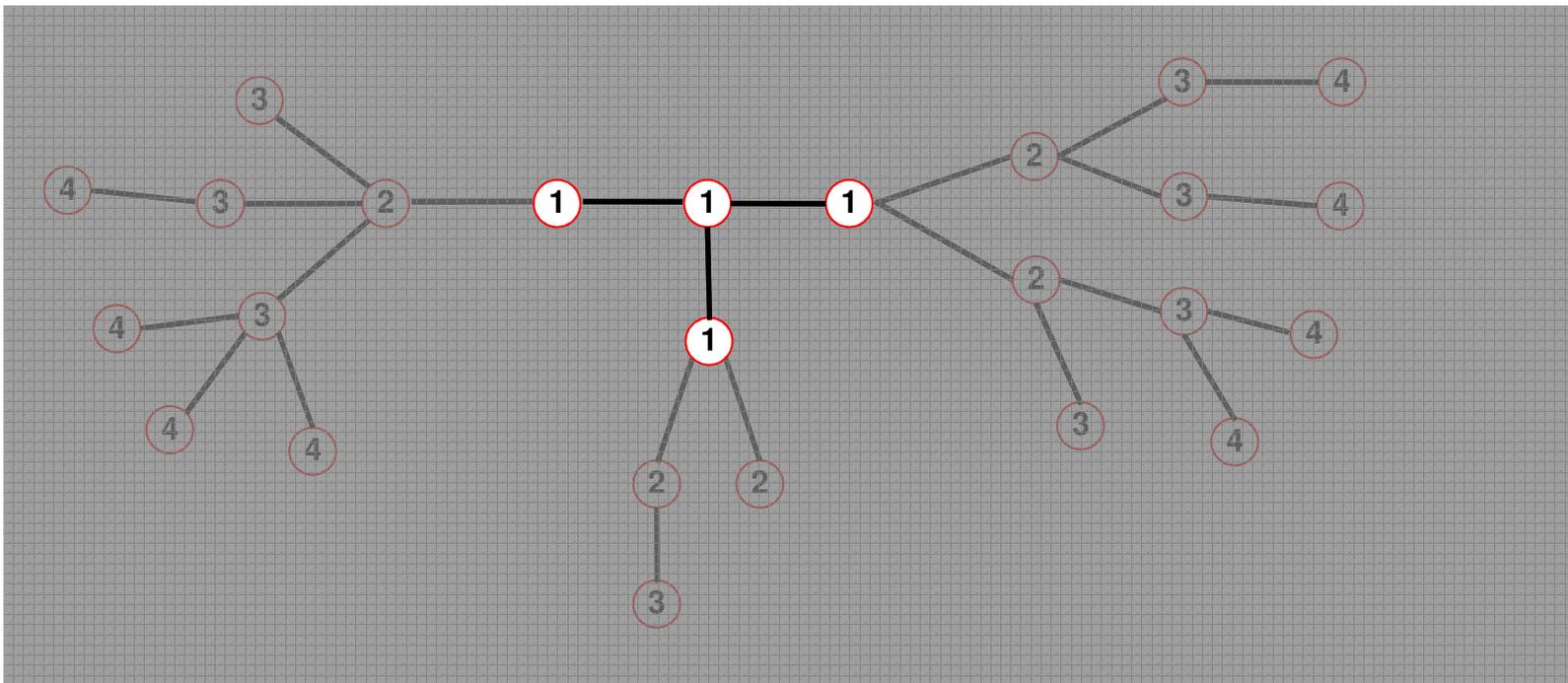
1. Se comienza en el nodo 1



CÁLCULO DE LA DISTANCIA ENTRE DOS NODOS: BÚSQUEDA PRIMERO EN ANCHURA (2)

Distancia entre el nodo 1 y el nodo 4:

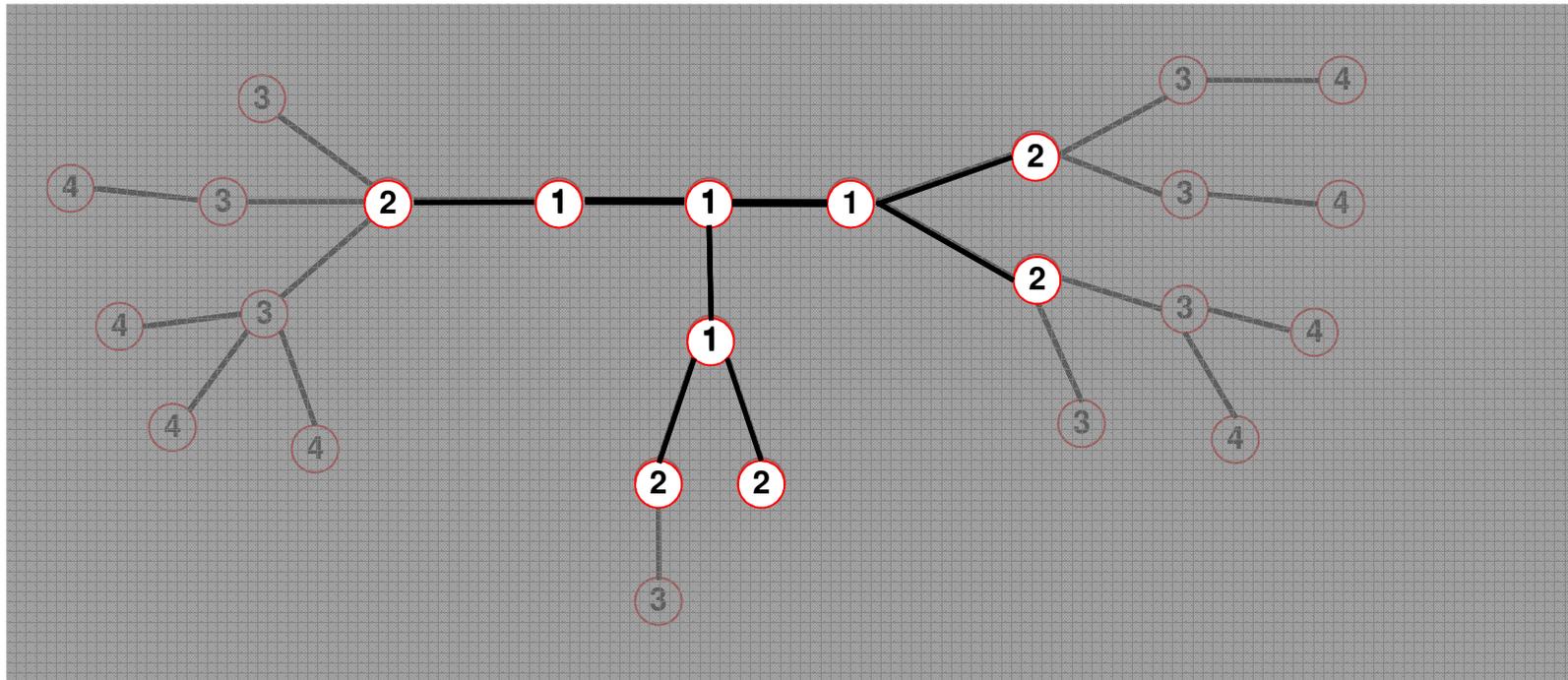
1. Se comienza en el nodo 1
2. Se etiquetan los nodos adyacentes a 1 con distancia 1. Se insertan en una cola



CÁLCULO DE LA DISTANCIA ENTRE DOS NODOS: BÚSQUEDA PRIMERO EN ANCHURA (3)

Distancia entre el nodo 1 y el nodo 4:

1. Se comienza en el nodo 1
2. Se marcan los nodos adyacentes a 1 con distancia 1. Se insertan en una cola
3. Se saca el primer nodo i de la cola. Se etiquetan los nodos adyacentes a i no etiquetados previamente con distancia 2. Se insertan en la cola

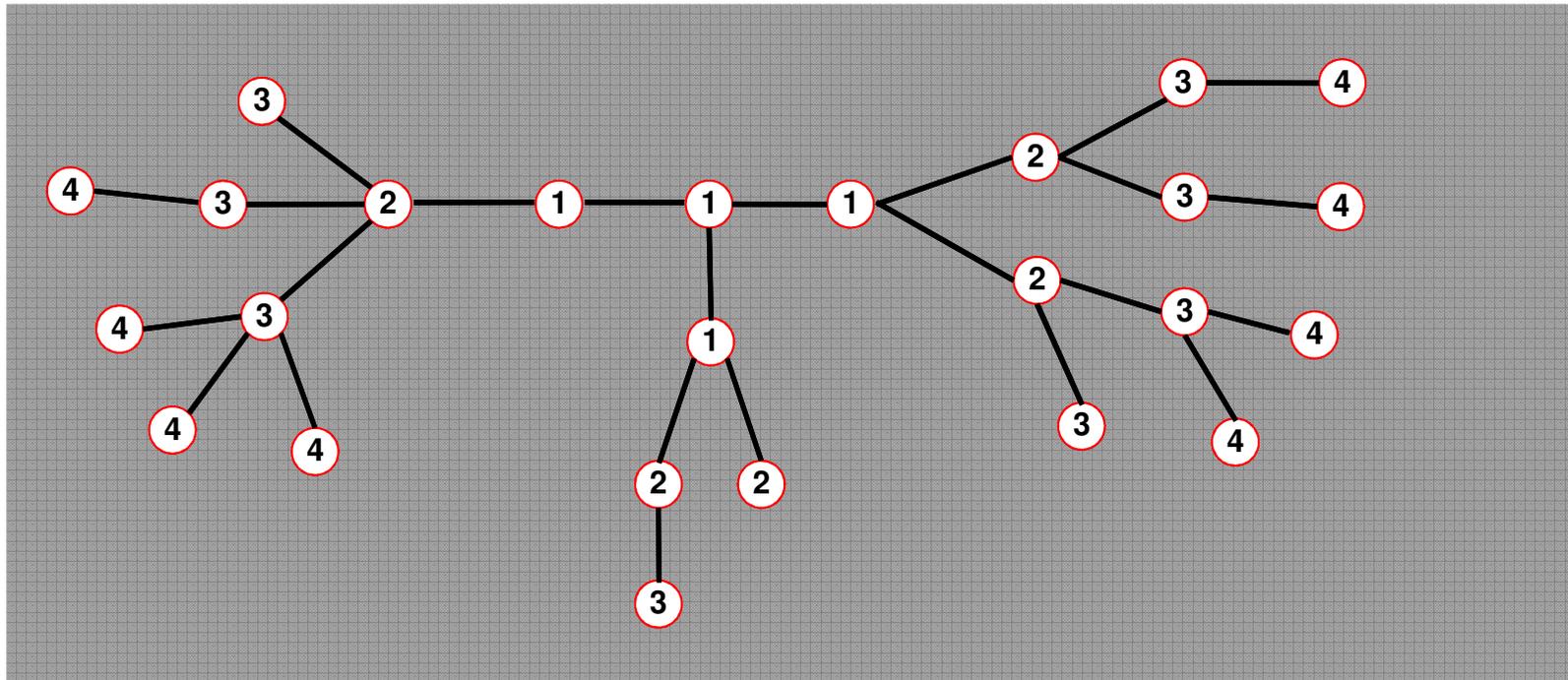


CÁLCULO DE LA DISTANCIA ENTRE DOS NODOS: BÚSQUEA PRIMERO EN ANCHURA (4)

Distancia entre el nodo 1 y el nodo 4:

4. Se itera hasta encontrar el nodo 4 o hasta que no queden más nodos en la cola

La distancia entre 1 y 4 corresponde a la etiqueta del nodo 4, si existe, o a infinito, en caso contrario



DIÁMETRO DE LA RED Y DISTANCIA MEDIA

Diámetro (d_{max}): longitud del camino mínimo más largo de la red (distancia entre los dos nodos más lejanos)

En redes grandes, se puede determinar con el algoritmo de búsqueda primero en anchura

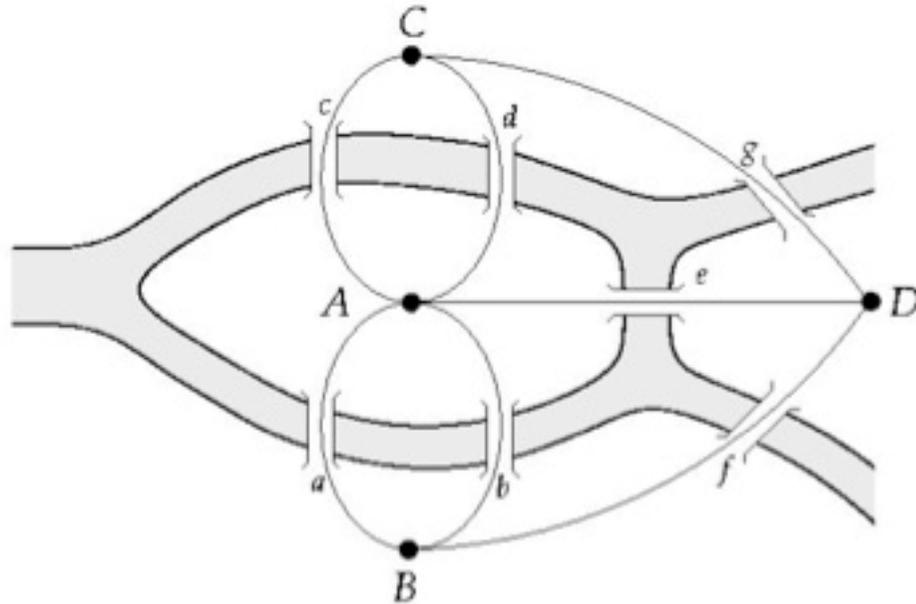
Distancia media/longitud media de los caminos ($\langle d \rangle$) para un grafo dirigido:

$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i, j \neq i} d_{ij} \quad \text{donde } d_{ij} \text{ es la distancia entre los nodos } i \text{ y } j$$

En un grafo no dirigido $d_{ij}=d_{ji}$. De este modo, sólo es necesario contar la longitud de los caminos una vez:

$$\langle d \rangle \equiv \frac{1}{L_{\max}} \sum_{i, j > i} d_{ij}$$

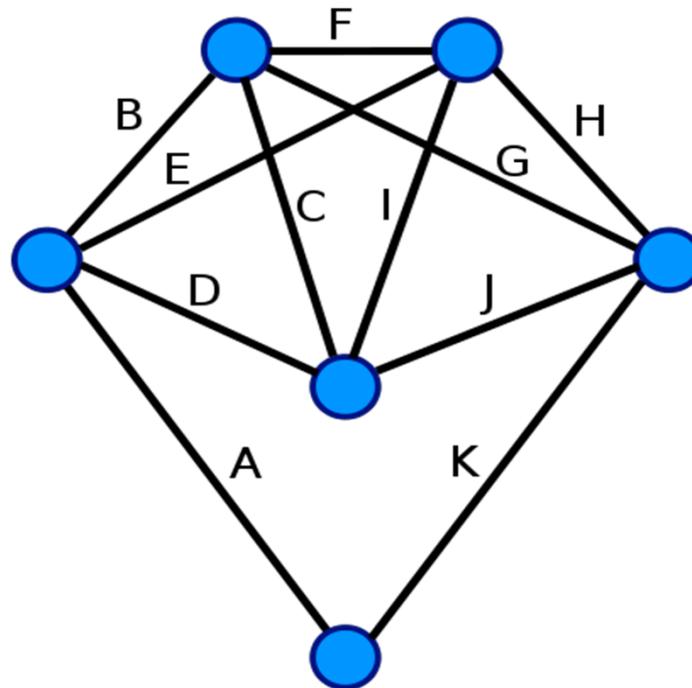
LOS PUENTES DE KONIGSBERG



¿Es posible recorrer los siete puentes sin pasar dos veces por el mismo?

Camino (o circuito) Euleriano : camino que retorna al punto inicial después de recorrer **cada enlace** del grafo una sola vez

GRAFO EULERIANO: aquel que tiene un camino Euleriano

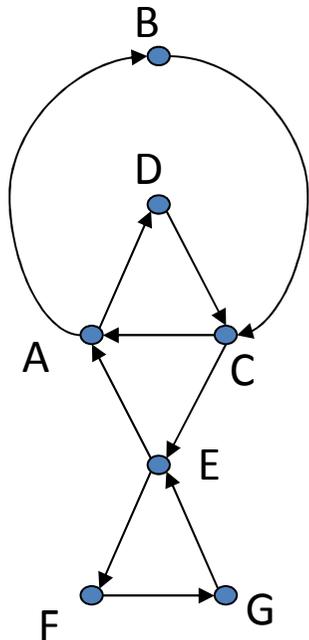


Puesto que todos los vértices de este grafo tienen un grado par, es un grafo Euleriano

Si se siguen los enlaces en orden alfabético obtenemos un circuito Euleriano

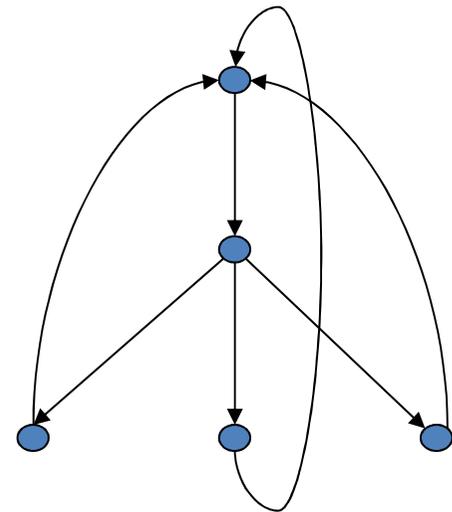
http://en.wikipedia.org/wiki/Euler_circuit

CIRCUITOS EULERIANOS EN GRAFOS DIRIGIDOS



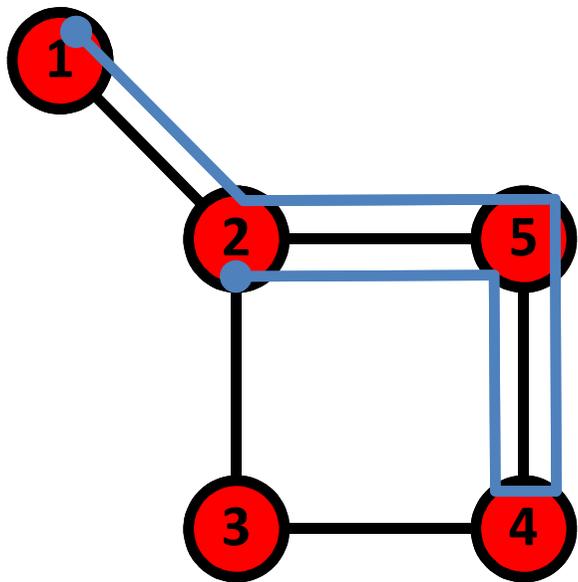
Si un digrafo es fuertemente conexo y el grado de entrada (in-degree) de cada nodo es igual al de salida (out-degree), existe un circuito Euleriano

En caso contrario, no existe circuito Euleriano
En un circuito, hay que entrar en cada nodo tantas veces como se sale de él



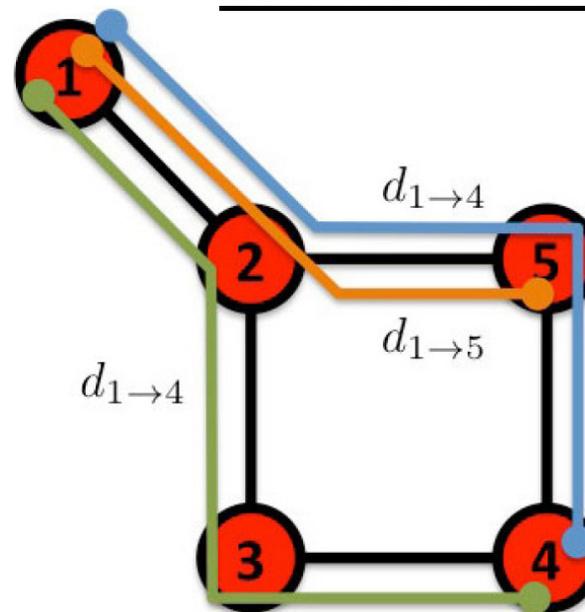
RESUMEN: CAMINOS EN GRAFOS (1)

Camino



Secuencia de nodos de forma que cada nodo está conectado al siguiente mediante un enlace

Camino mínimo



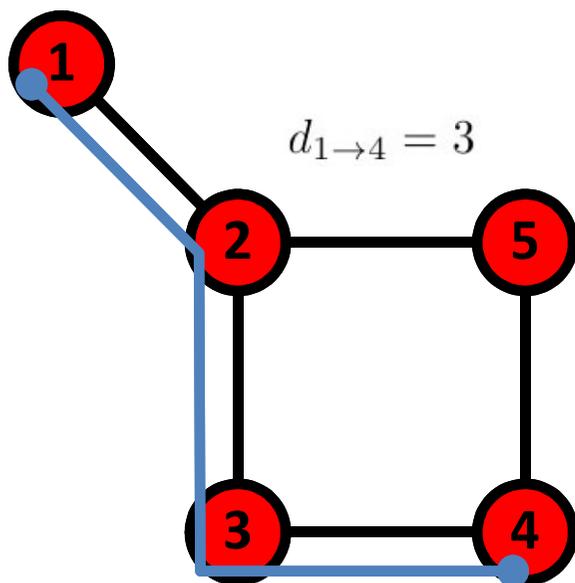
Camino más corto entre dos nodos (distancia)

$$d_{1 \rightarrow 4} = 3$$

$$d_{1 \rightarrow 5} = 2$$

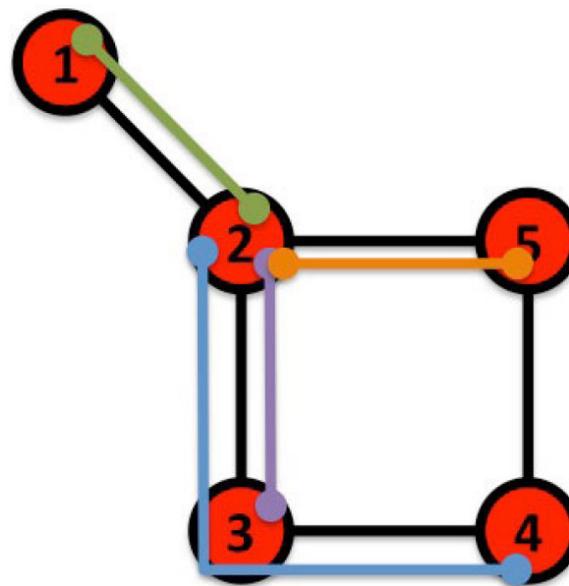
RESUMEN: CAMINOS EN GRAFOS (2)

Diámetro



La longitud del camino
mínimo más largo del grafo

Distancia media

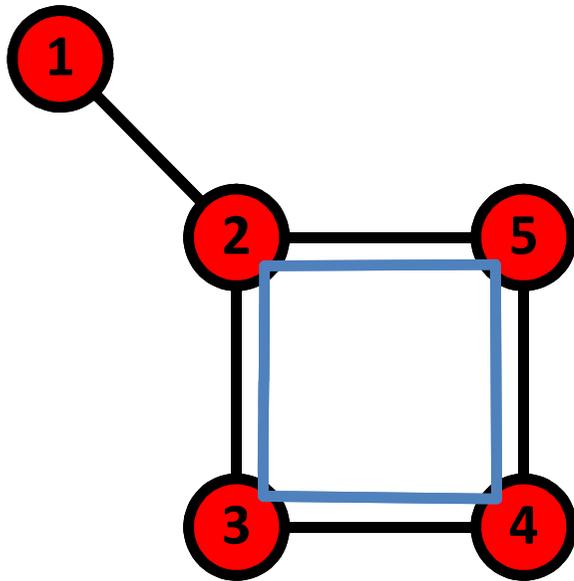


$$\begin{aligned} & (d_{1 \rightarrow 2} + d_{1 \rightarrow 3} + d_{1 \rightarrow 4} + \\ & + d_{1 \rightarrow 5} + d_{2 \rightarrow 3} + d_{2 \rightarrow 4} + \\ & + d_{2 \rightarrow 5} + d_{3 \rightarrow 4} + d_{3 \rightarrow 5} + \\ & + d_{4 \rightarrow 5}) / 10 = 1.6 \end{aligned}$$

Media de la longitud de los caminos
mínimos entre todos los pares de nodos

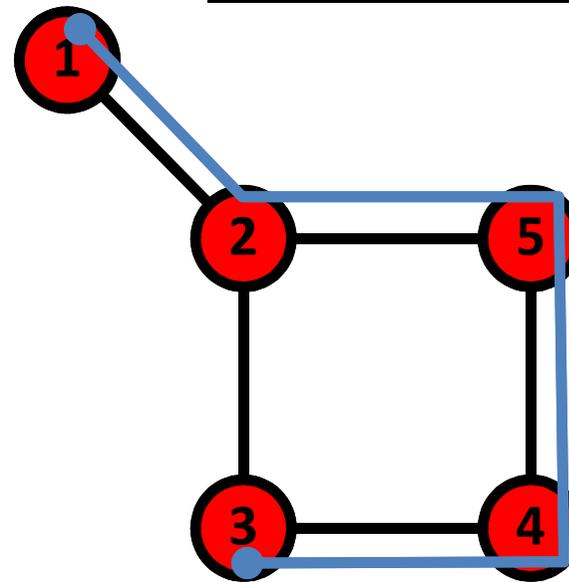
RESUMEN: CAMINOS EN GRAFOS (3)

Ciclo



Camino con el mismo nodo inicial y final

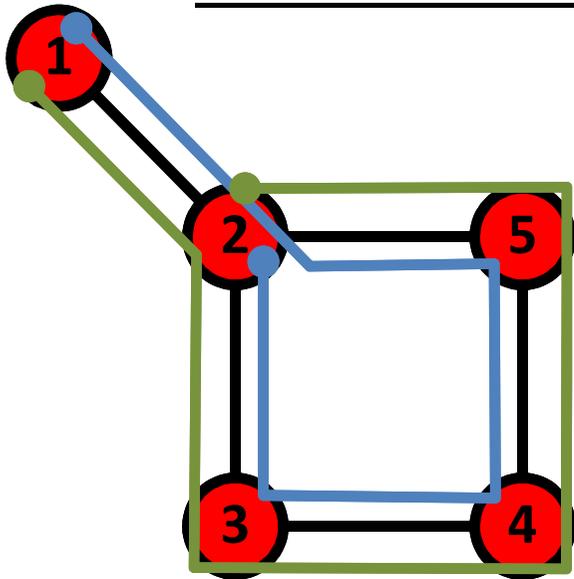
Camino auto-evitado



Camino que no interseca consigo mismo

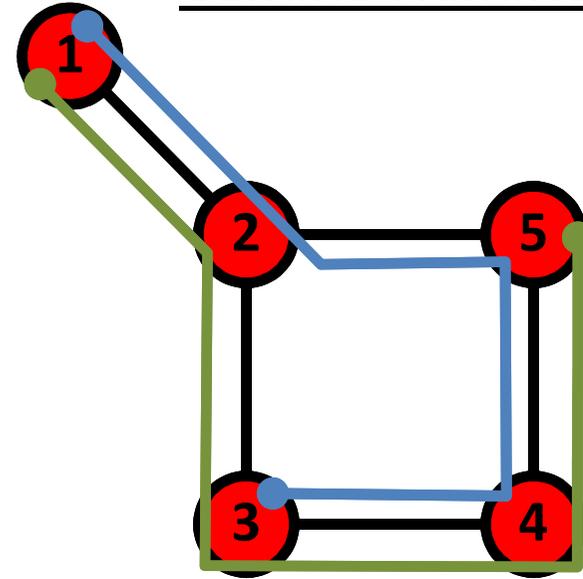
RESUMEN: CAMINOS EN GRAFOS (4)

Camino Euleriano



Camino que cruza cada enlace exactamente una vez

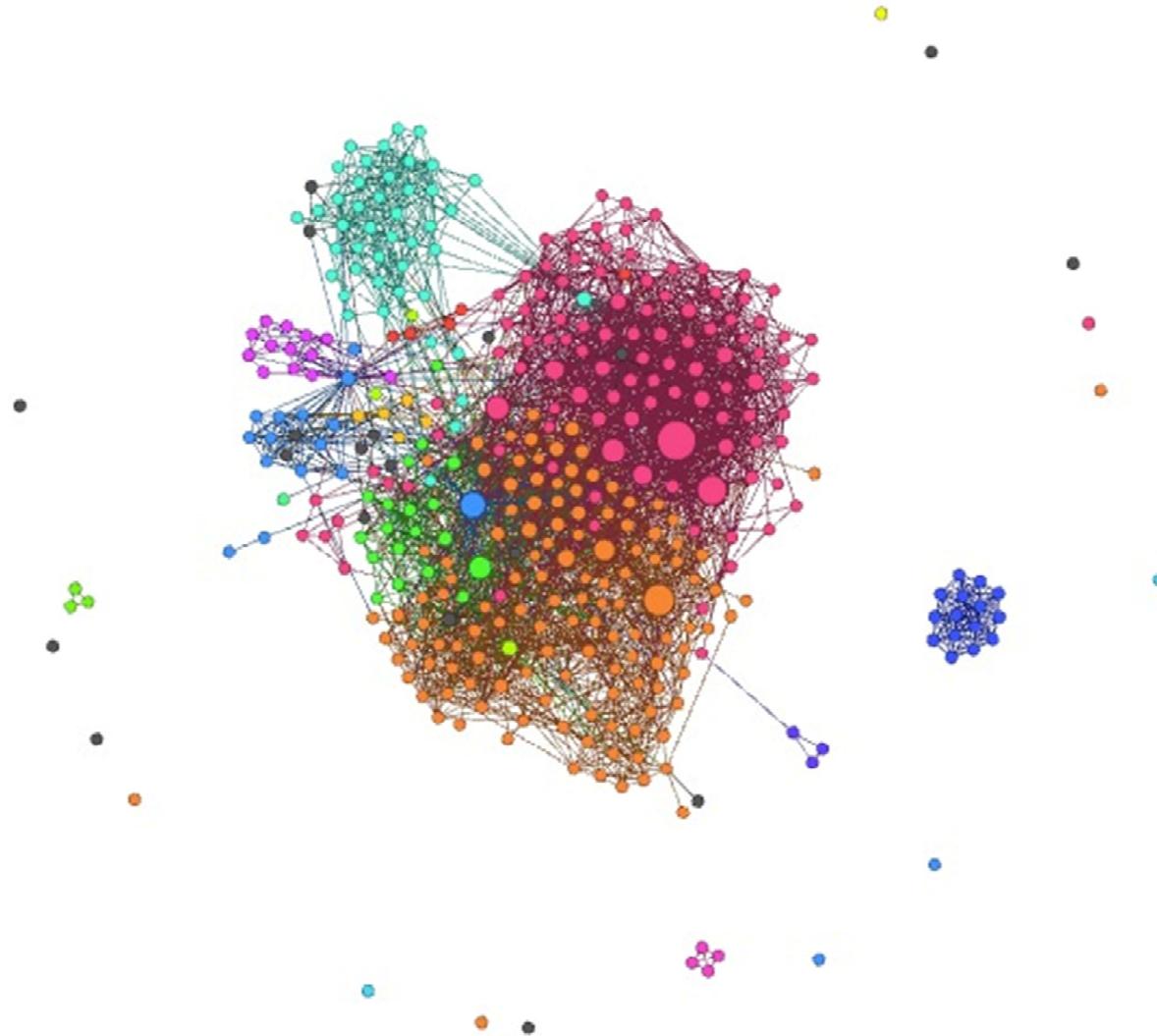
Camino Hamiltoniano



Camino que visita cada nodo exactamente una vez

CONECTIVIDAD Y COMPONENTES

¿ESTÁN CONECTADOS TODOS LOS NODOS EN UNA RED?

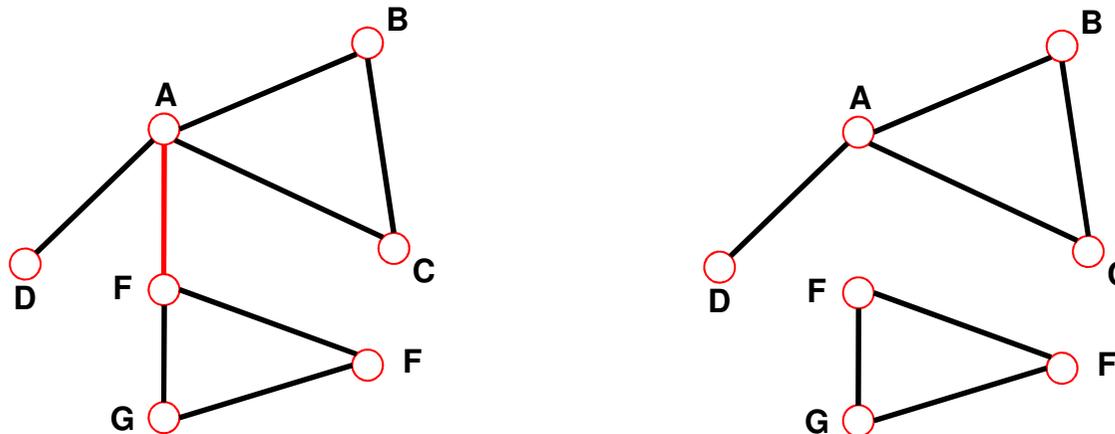


CONECTIVIDAD DE GRAFOS NO DIRIGIDOS (1)

La utilidad de muchas redes reales es conectar a sus componentes entre sí (p.ej., red telefónica). La formulación teórica de la conectividad en grafos se basa en los caminos. Dos nodos están conectados si existe al menos un camino entre ellos en el grafo

Grafo conexo (no dirigido): todo par de vértices está unido por al menos un camino

Componente conexa: subgrafo conexo del grafo principal al que no se pueden añadir más nodos. Un grafo no conexo está formado por dos o más componentes conexas



Puente: cualquier enlace que, en caso de ser borrado, hace que el grafo sea no conexo

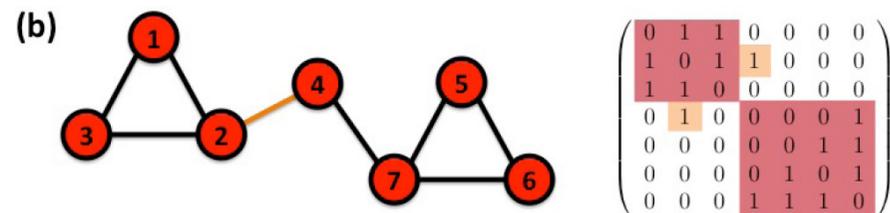
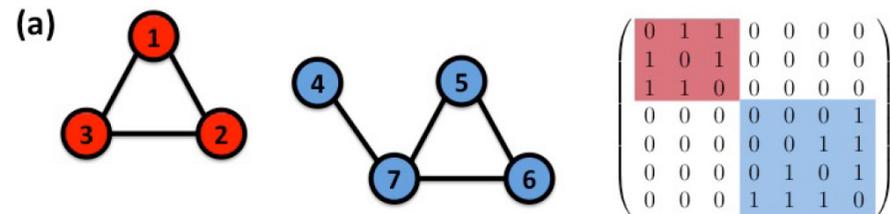
CONECTIVIDAD DE GRAFOS NO DIRIGIDOS (2)

Matriz de Adyacencia

En redes no dirigidas, las componentes conexas se pueden obtener a partir de la matriz de adyacencia

Si la red no es conexa, su matriz puede escribirse en forma **diagonal por bloques**, es decir, ordenarse de tal forma que los elementos no nulos se agrupan en bloques cuadrados en la diagonal y el resto valen cero. Cada bloque es una componente:

$$A = \begin{pmatrix} \text{[Red Square]} & 0 & \dots \\ 0 & \text{[Red Square]} & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

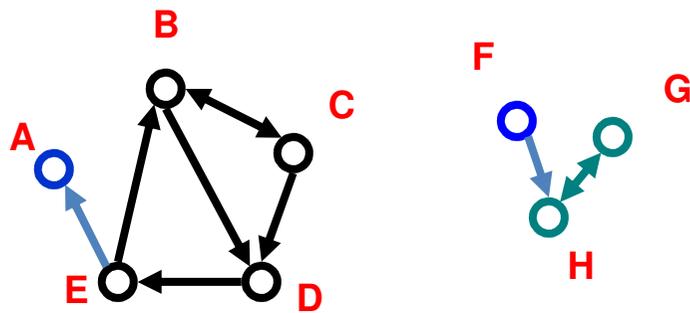


Existen métodos de algebra lineal para saber si una matriz es diagonalizable por bloques pero son costosos. En la práctica, **las componentes conexas de redes grandes se determinan con ejecuciones sucesivas del algoritmo de búsqueda primero en anchura**

CONECTIVIDAD DE GRAFOS DIRIGIDOS

Grafo (dirigido) fuertemente conexo: para cada par de nodos i y j , existe un camino de i a j y **viceversa** (otro de j a i), es decir, todo nodo del grafo es accesible desde cualquier otro siguiendo los enlaces dirigidos

Grafo (dirigido) débilmente conexo: todo nodo del grafo es accesible desde cualquier otro si no se tiene en cuenta la dirección de los enlaces



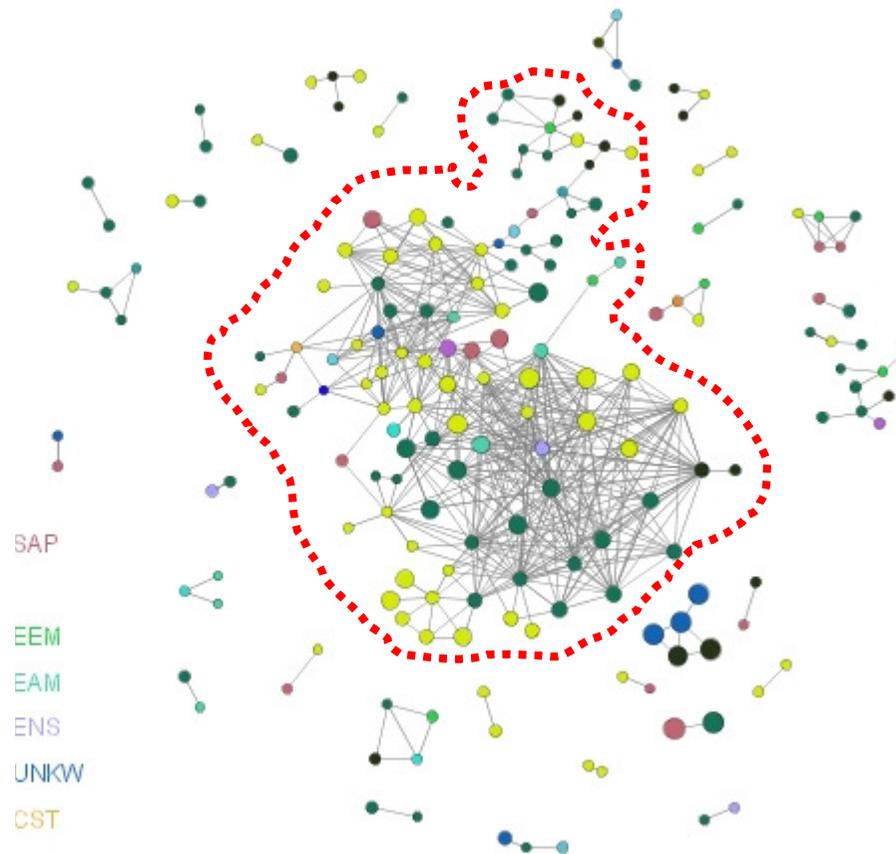
- Componentes fuertemente conexas (CFCs):
 - B C D E
 - A
 - G H
 - F
- Componentes débilmente conexas (CDCs):
 - A B C D E
 - G H F

Componente de entrada: nodos que pueden alcanzar la CFC

Componente de salida: nodos que pueden ser alcanzados desde la CFC

COMPONENTE GIGANTE

Si la mayor componente conexa de la red incluye una fracción significativa de la misma, se denomina **componente gigante**. El resto de componentes se denominan **aisladas**



MEDIDAS PRINCIPALES DE LAS REDES

MEDIDAS EN TEORÍA DE REDES

Las medidas permiten caracterizar una red para su análisis

Ya hemos estudiado algunas medidas de este tipo:

- **Locales:** grado de los nodos k_j
- **Globales:**
 - número de nodos N
 - número de enlaces L y número máximo de enlaces L_{max} . Densidad = L/L_{max}
 - grado medio $\langle k \rangle$
 - diámetro d_{max}
 - distancia media $\langle d \rangle$

LAS TRES MEDIDAS PRINCIPALES EN TEORÍA DE REDES

Distribución (de probabilidad)
de grados

$$p_k$$

Distancia media

$$\langle d \rangle$$

Coeficiente de Clustering

$$C_i$$

REPASO DE ESTADÍSTICA

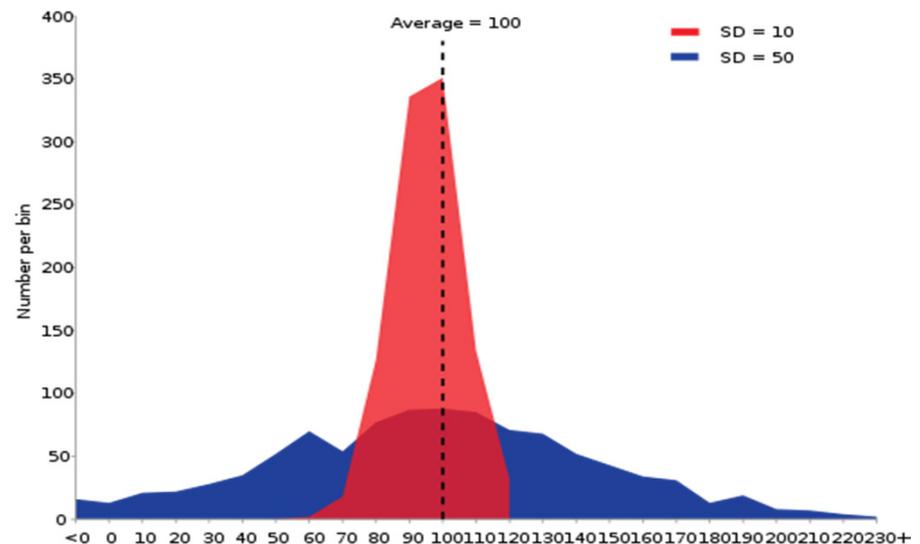
Dada una muestra de valores x_1, \dots, x_N , se puede definir la **Distribución** de x : probabilidad de que un valor aleatorio tome valor x

$$P(x) = (\text{número de valores de la muestra} = x) / N$$

(casos favorables/casos posibles)

$$\sum_i P(x_i) = 1 \quad \text{¡siempre!}$$

Histogramas >>>



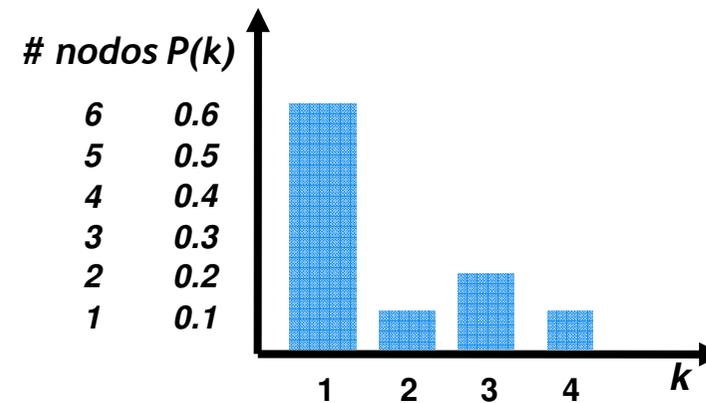
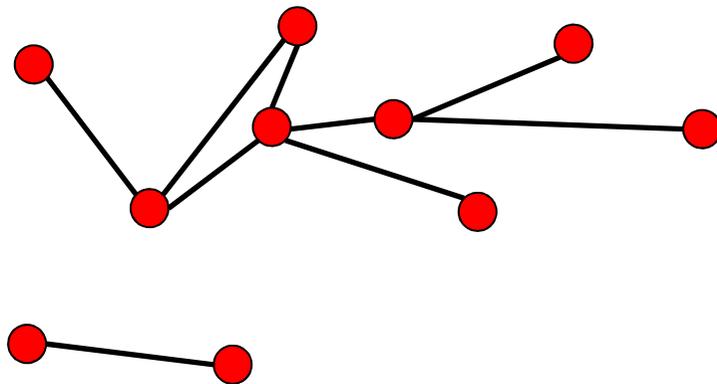
DISTRIBUCIÓN DE GRADOS (1)

Distribución de grados $P(k)$: probabilidad de que un nodo escogido aleatoriamente tenga grado k

Se construye a partir del histograma de grados de los nodos de la red:

$N_k = \# \text{ nodos con grado } k$

$P(k) = N_k / N$



DISTRIBUCIÓN DE GRADOS (2)

Representación discreta: p_k es la probabilidad de que un nodo tenga grado k

Descripción continua: $p(k)$ es la función de densidad de probabilidad de los grados, donde

$$\int_{k_1}^{k_2} p(k) dk$$

Representa la probabilidad de que el grado de un nodo esté entre k_1 y k_2

Condición de Normalidad:

$$\sum_0^{\infty} p_k = 1 \qquad \int_{K_{\min}}^{\infty} p(k) dk = 1$$

donde K_{\min} es el grado mínimo de los nodos de la red

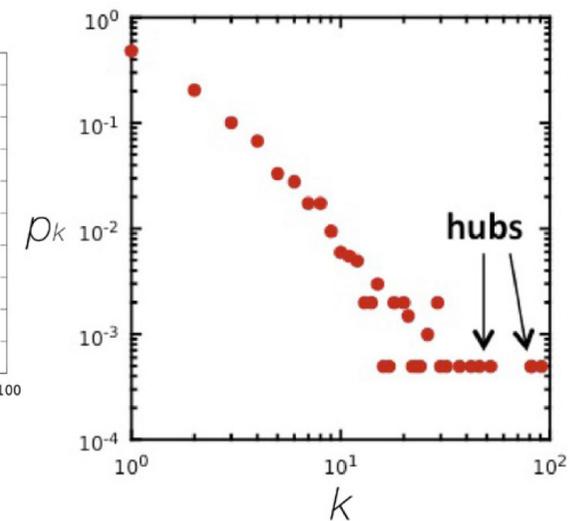
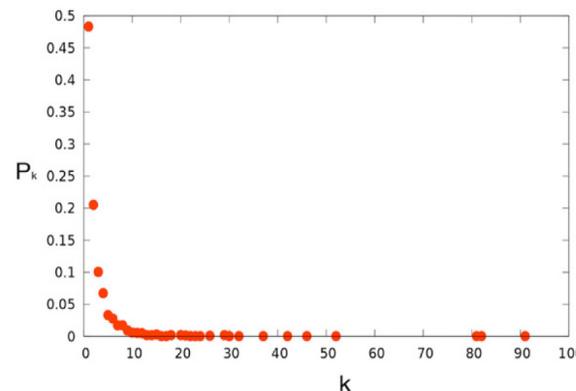
DISTRIBUCIÓN DE GRADOS (3)

La distribución de grados juega un papel central en Teoría de Redes. Permite caracterizar distintos modelos de redes, especialmente las **redes libres de escala** (*scale-free*), y determina muchos fenómenos de la red como su robustez y la difusión de virus

Además, el cálculo de muchas propiedades de la red requiere conocer p_k . Por ejemplo, el grado medio de una red se puede calcular como:

$$\langle k \rangle \equiv \sum_{k=0}^{\infty} k \cdot p_k$$

La distribución del grado suele representarse en forma de **gráficos log-log** usando ejes logarítmicos (ej. red de interacción de proteínas)



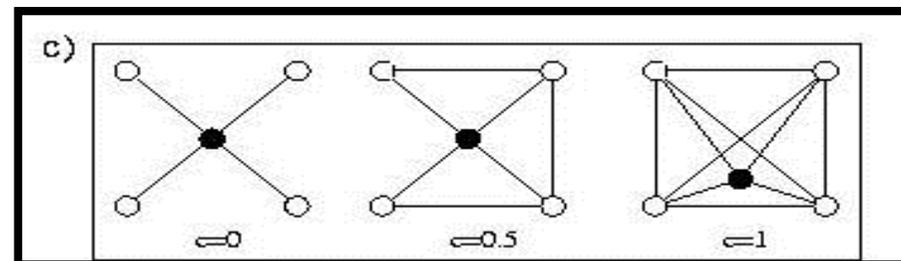
COEFICIENTE DE CLUSTERING (1)

* **Coeficiente de clustering:** Mide la **densidad local de la red**

¿qué proporción de los vecinos de cada nodo están conectados?

- * Nodo i con grado k_i
- * L_i = número de enlaces entre los vecinos del nodo i
- * $C_i \in [0,1]$. $C_i = 0$ indica que ninguno de los vecinos de i están conectados entre sí, $C_i = 1$ indica que todos están conectados

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

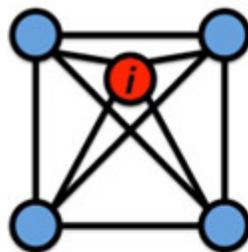


COEFICIENTE DE CLUSTERING (2)

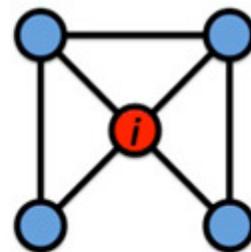
La **densidad global de la red** se mide con el **coeficiente de clustering medio**:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

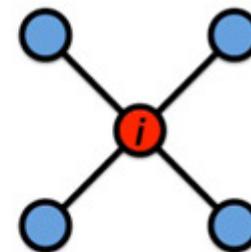
Interpretación probabilística:
 $\langle C \rangle$ mide la probabilidad de que dos vecinos de un nodo de la red escogido aleatoriamente estén conectados entre sí



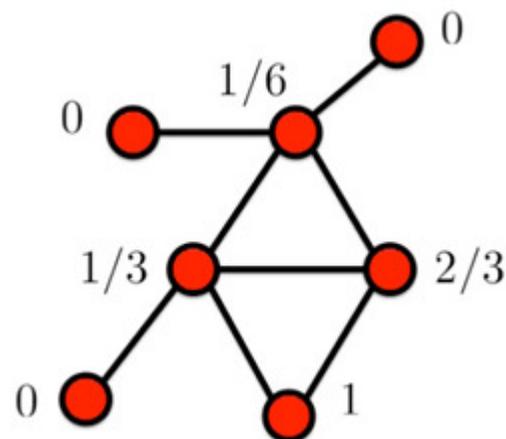
$$C_i = 1$$



$$C_i = 1/2$$

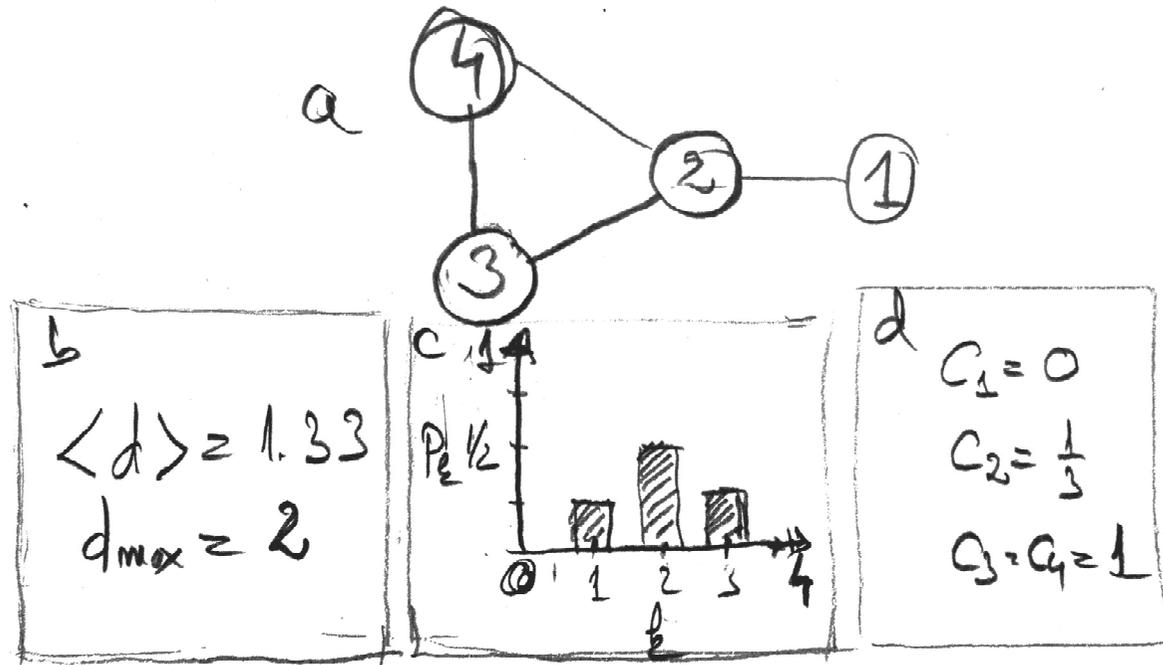


$$C_i = 0$$



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

PRIMEROS PASOS EN ANÁLISIS DE REDES (1)



A. Distribución de grados:

P_k

B. Distancia media:

$\langle d \rangle$

C. Coeficiente de clustering:

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

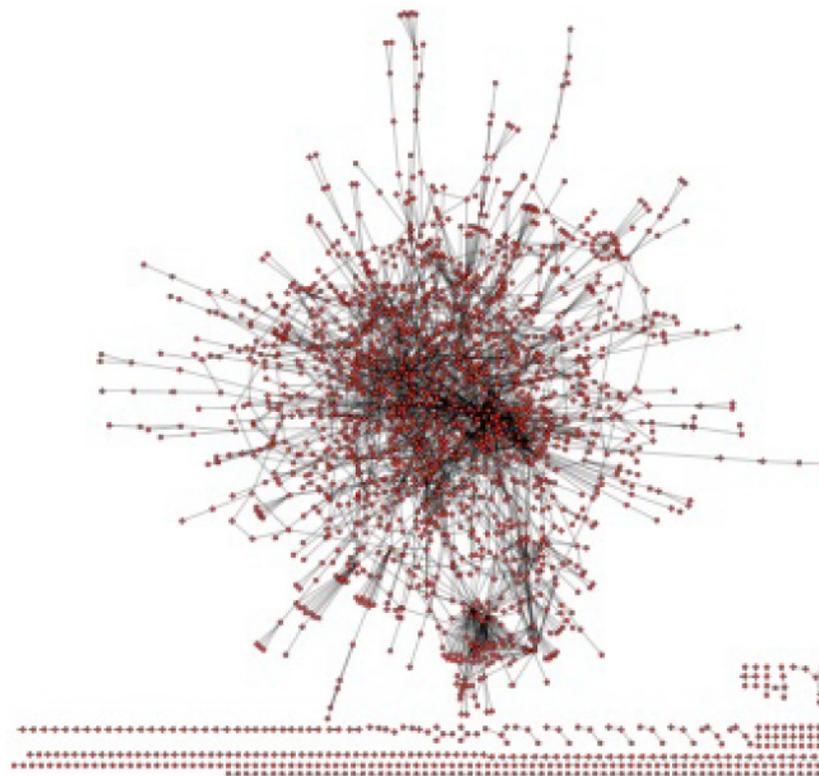
PRIMEROS PASOS EN ANÁLISIS DE REDES (2)

El objetivo principal de la Teoría de Redes es el análisis de redes reales

Estas redes suelen ser de gran tamaño (desde cientos a millones de nodos), lo que complica su análisis. Para llevarlo a cabo, se usan las **medidas** estudiadas

Ejemplo: Red de interacción entre proteínas (PPI) de la levadura. Los nodos son proteínas y están conectados si hay evidencia empírica de que interactúan

¡Demasiado compleja para analizarla visualmente o con la matriz de adyacencia!



PRIMEROS PASOS EN ANÁLISIS DE REDES (3)

Valores de las medidas: $N=2018$ nodos, $L=2930$ enlaces, $Densidad=0.00144$

Grado medio: $\langle k \rangle \equiv \frac{2L}{N} = 2.90 \rightarrow$

una proteína típica interactúa con otras 2-3

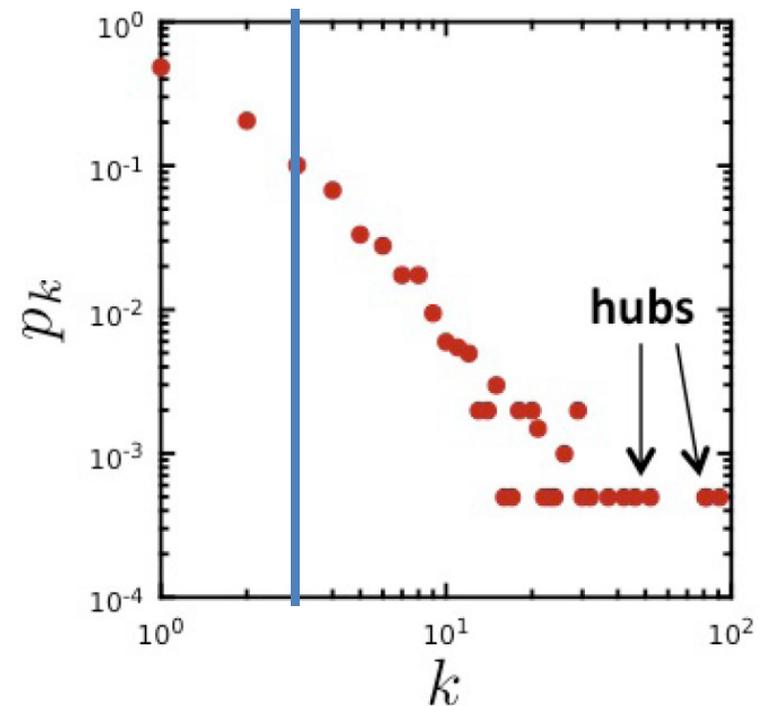
Además, la distribución de grados indica que no hay un nivel de interacción alto. Un 69% de los nodos tienen grado menor que 3 ($k_i < \langle k \rangle$)



Existen unos pocos nodos fuertemente conectados (hubs), el mayor con grado 91

Consecuencia de la **propiedad libre de escala (scale-free)**, muy común en redes reales (que estudiaremos en temas posteriores)

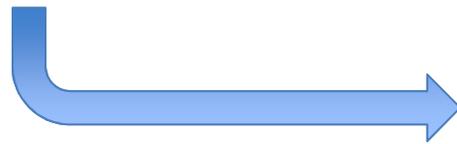
Distribución de grados



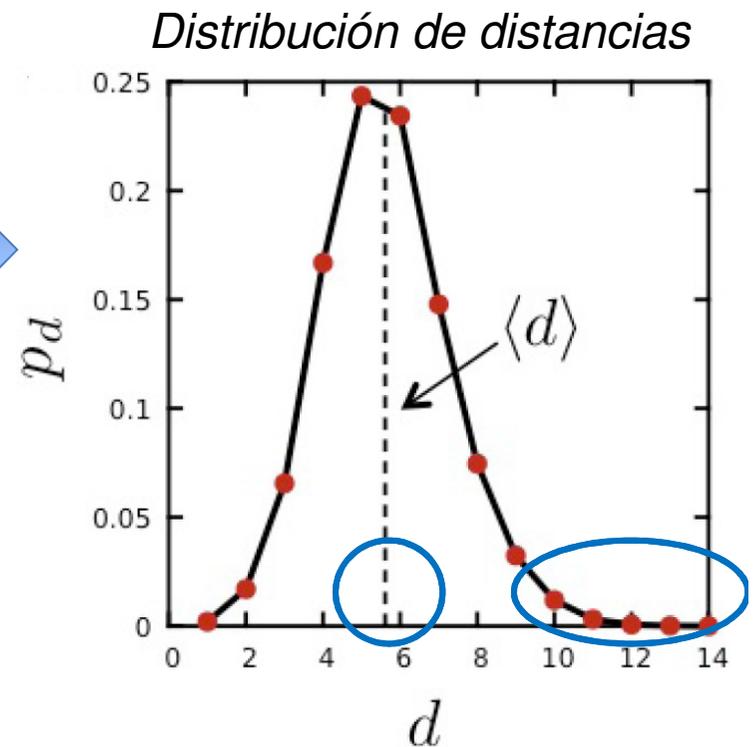
PRIMEROS PASOS EN ANÁLISIS DE REDES (4)

Diámetro: $d_{max}=14$. Viendo la red, pensaríamos que hay variaciones grandes en las distancias entre nodos pero la distribución tiene una media ($\langle d \rangle = 5.61$) y una desviación típica ($\sigma_d = 1.64$) bajas (muchos nodos con distancia cercana a la media)

p_d cae fuertemente para las distancias grandes, indicando que no existen distancias grandes



Consecuencia de la **propiedad de mundos pequeños (*small-world*)**, también común en redes reales



PRIMEROS PASOS EN ANÁLISIS DE REDES (5)

Conectividad: la red presenta 185 componentes conexas, muchas de ellas formados por una sola proteína aislada que no interactúa con el resto

Vemos una componente gigante que agrupa el 81% de los nodos: 1647 de los 2018. Hay también algunas componentes aisladas muy pequeñas. Esta **fragmentación** también es común en redes reales

Coefficiente de clustering medio: $\langle C \rangle = 0.12$. Aunque no lo parezca, es bastante **alto**, indicando un grado significativo de clustering local

El coeficiente de clustering es mucho mayor en los nodos poco conectados que en los hubs \rightarrow los nodos de grado bajo se sitúan en vecindarios localmente densos y viceversa

Consecuencia de la **jerarquía de redes**

Distribución de coef. clust.

