Introduction to Computational Social Science

Redes Complejas: Teoría de Grafos

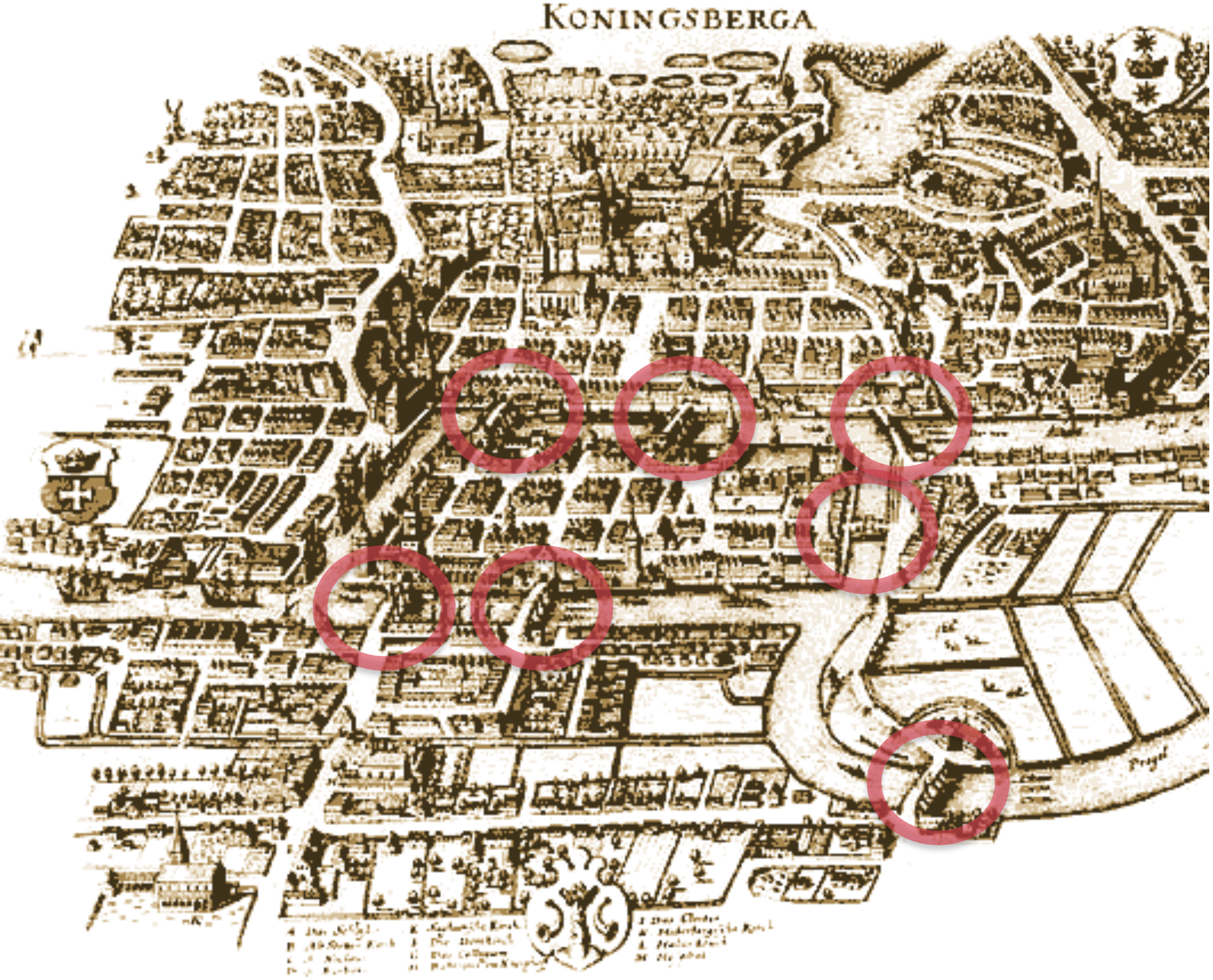
Cristian Candia-Castro Vallejos, Ph.D.

Assistant Professor, Facultad de Ingeniería, UDD
Director of the Data Science Master's Degree, UDD
External Faculty Northwestern Institute on Complex Systems, Northwestern University, USA
Founder at CINS-Data

Ayudantes: Melanie Oyarzún, Ph.D.(c) y Victor Landaeta, Ph.D.

Los puentes de Konigsberg

LOS PUENTES DE KONIGSBERG



¿Se puede cruzar los siete puentes y nunca cruzar el mismo puente dos veces?

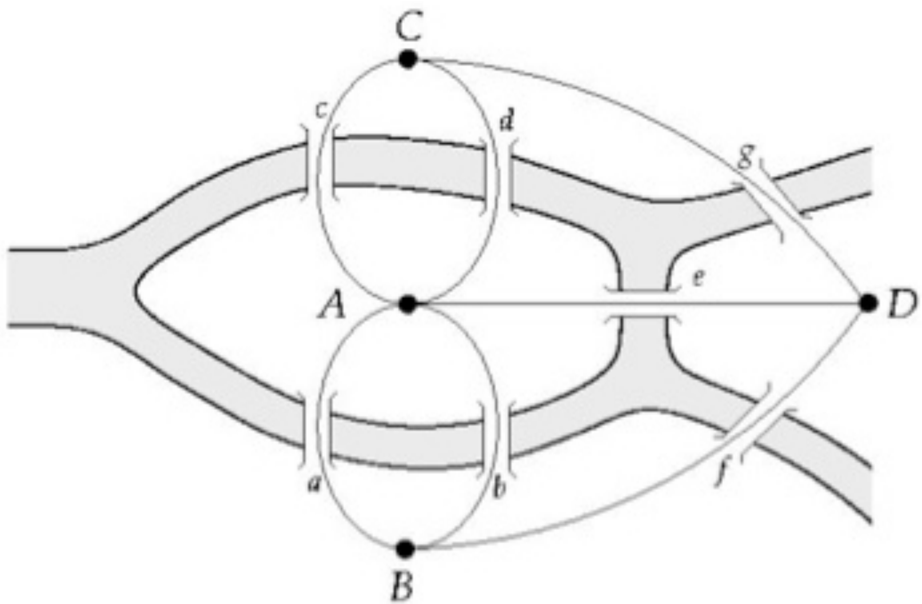
1736

LOS PUENTES DE KONIGSBERG

¿Se puede cruzar los siete puentes y nunca cruzar el mismo puente dos veces?

1736

LOS PUENTES DE KONIGSBERG



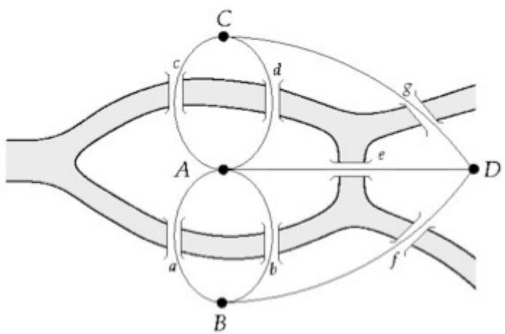
¿Se puede cruzar los siete puentes y nunca cruzar el mismo puente dos veces?

1735: **Teorema de Euler:**

(a) Si un grafo tiene **más de dos nodos de grado impar**, no hay ruta.

(b) **Si un grafo está conectado** y **no tiene nodos de grados impares**, tiene al menos una ruta.

LOS PUENTES DE KONIGSBERG



¿Se puede cruzar los siete puentes y nunca cruzar el mismo puente dos veces?

Hoy recordamos la prueba de Euler porque fue la primera vez que alguien resolvió un problema matemático convirtiéndolo en un grafo. En retrospectiva, la prueba tiene dos mensajes importantes:

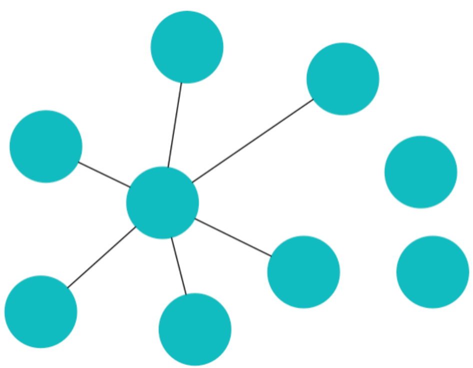
1. **Algunos problemas se vuelven más simples y tratables si se representan como un grafo.**

2. **La existencia del camino no depende de nuestro ingenio para encontrarlo.** Más bien, es una propiedad del grafo. De hecho, dado el diseño de los puentes Konigsberg, no importa lo inteligentes que seamos, nunca encontraremos el camino deseado.

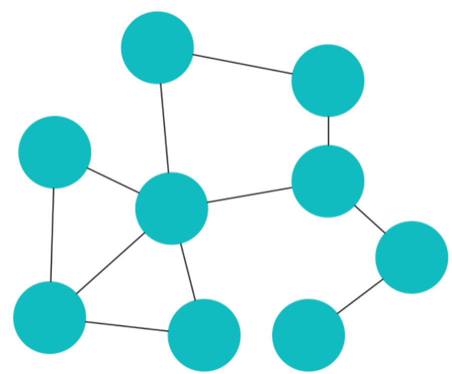
¿En cuál de estas redes es posible idear un paseo que atraviese cada enlace una vez y solo una vez?

Es decir, ¿en qué red es posible realizar un Camino Euleriano?

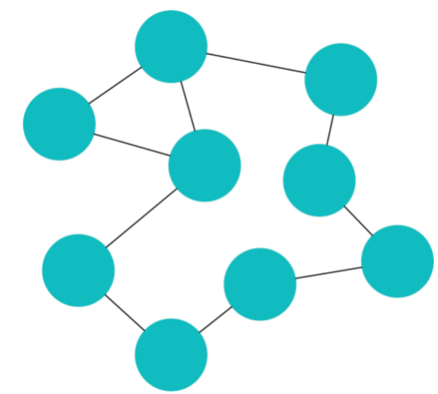
A



B

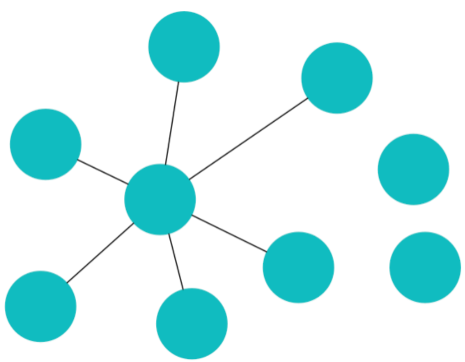


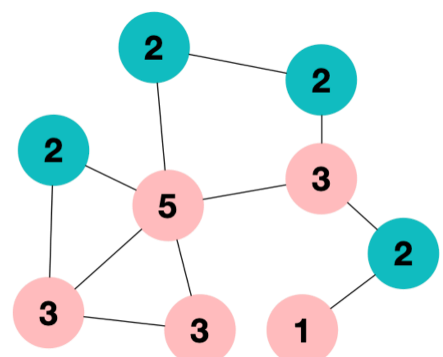
C

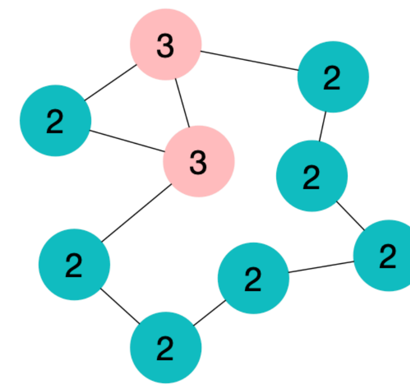


1. ¿Está conectado?

2. Cuente el número de nodos de grados impares, ¿es 0 o 2?



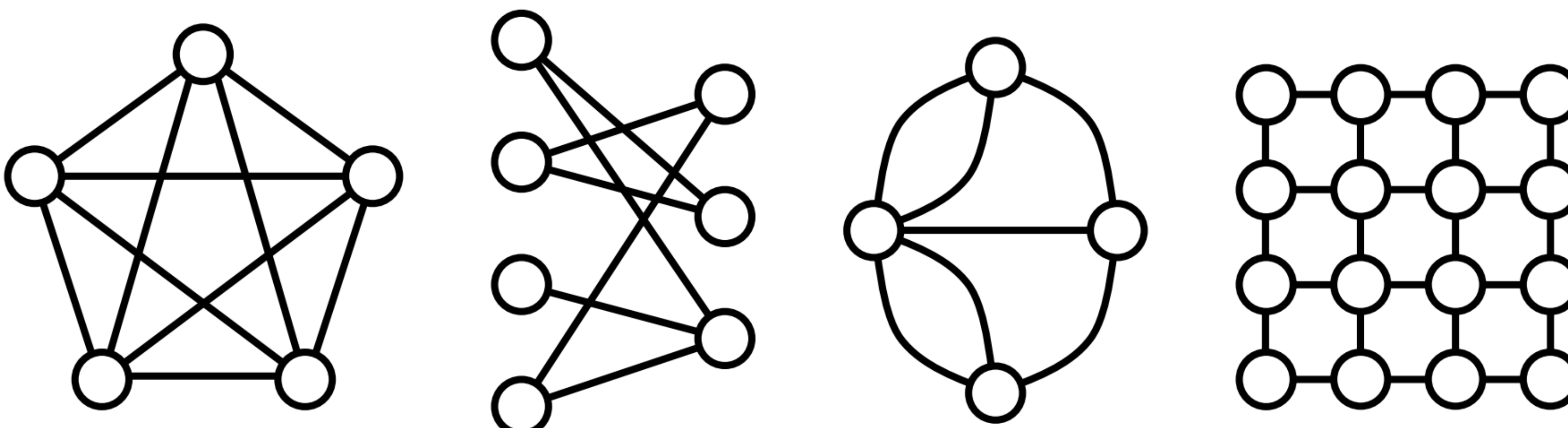




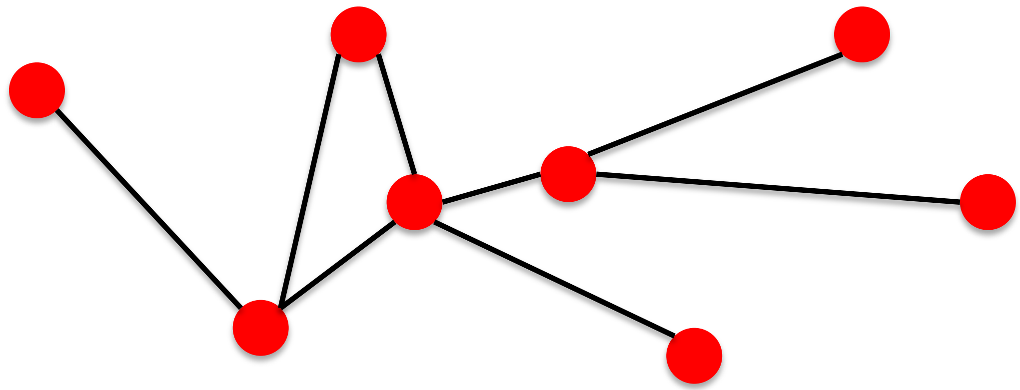
Redes y grafos

¿Qué es un grafo?

Estructura matemática que consta de "nodos" (o vértices) y "bordes" (o enlaces) que conectan a los nodos.



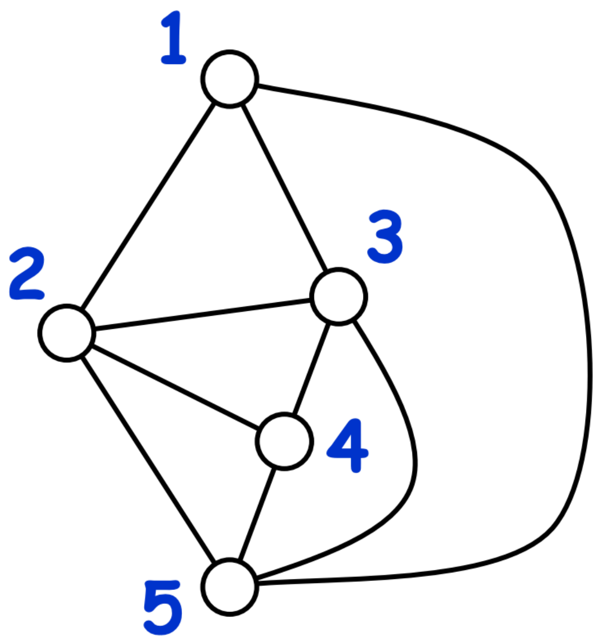
COMPONENTES DE UN SISTEMA COMPLEJO



- **componenetes**: nodos, vertices   N
- **interacciones**: enlaces, bordes   L
- **sistema**: red, grafo   (N,L)

Ejemplo

- G(V, E): graph (network)
  V: vertices (nodes), E: edges (links)



Nodos

1, 2, 3, 4, 5

Enlaces

1<->2, 1<->3, 1<->5,
2<->3, 2<->4, 2<->5,
3<->4, 3<->5, 4<->5

Los nodos pueden tener estados; los enlaces pueden tener direcciones y pesos

REDES O GRAFOS?

red a menudo se refiere a sistemas reales

•www,
•social network (red social)
•metabolic network. (red metabólica)

Lenguaje: (Red (Network), nodo (node), enlace (link))

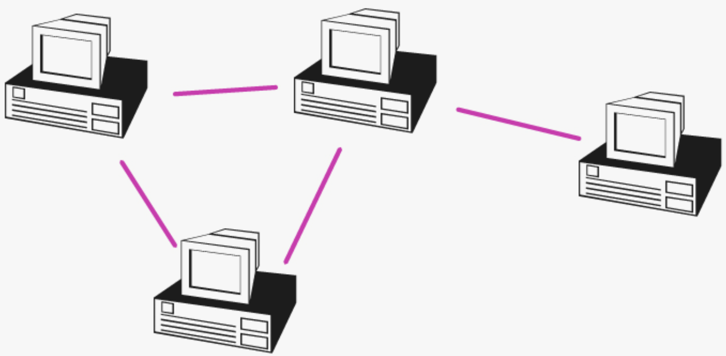
grafo: representación matemática de una red

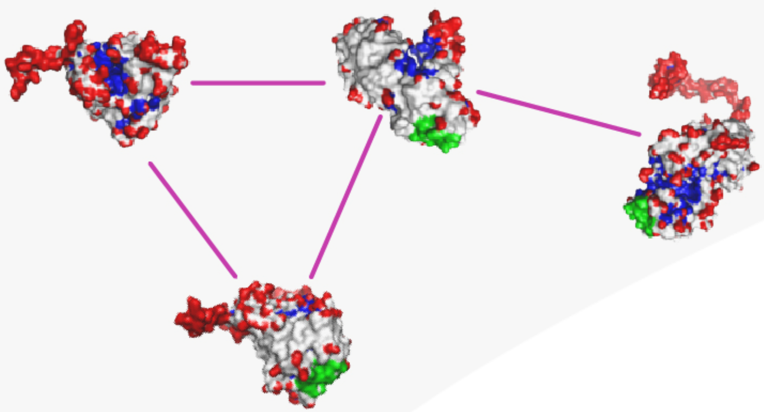
•web graph,
•social graph (a Facebook term)

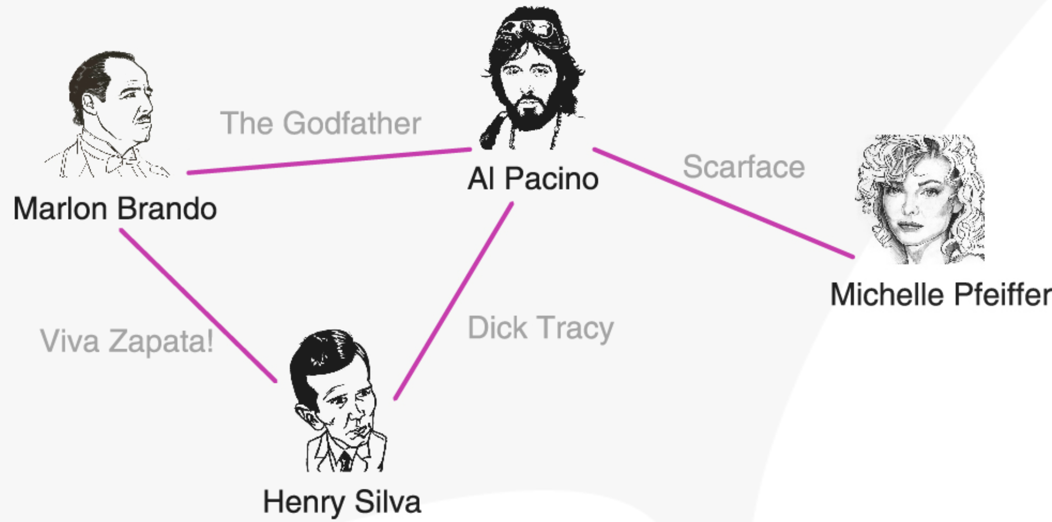
Lenguaje: (Grafo (graph), vertice (vertex), enlace (edge))

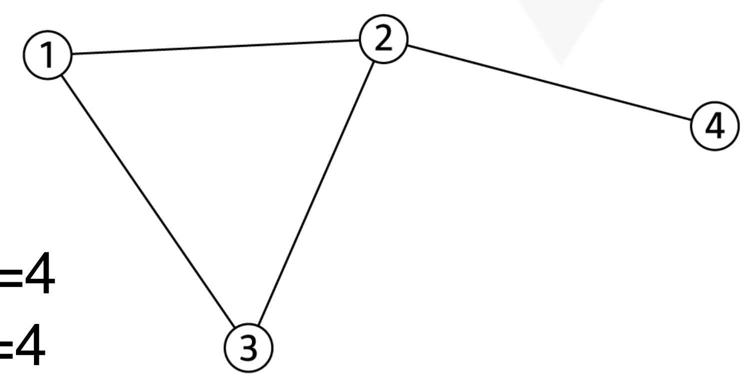
Trataremos de hacer esta distinción siempre que sea apropiado, pero en la mayoría de los casos usaremos los dos términos indistintamente.

UN LENGUAJE COMÚN









N=4

L=4

CHOOSING A PROPER REPRESENTATION

La elección de la representación de red adecuada determina nuestra capacidad para utilizar la teoría de redes con éxito.

En algunos casos hay una representación única e inequívoca. En otros casos, la representación no es de ninguna manera única.

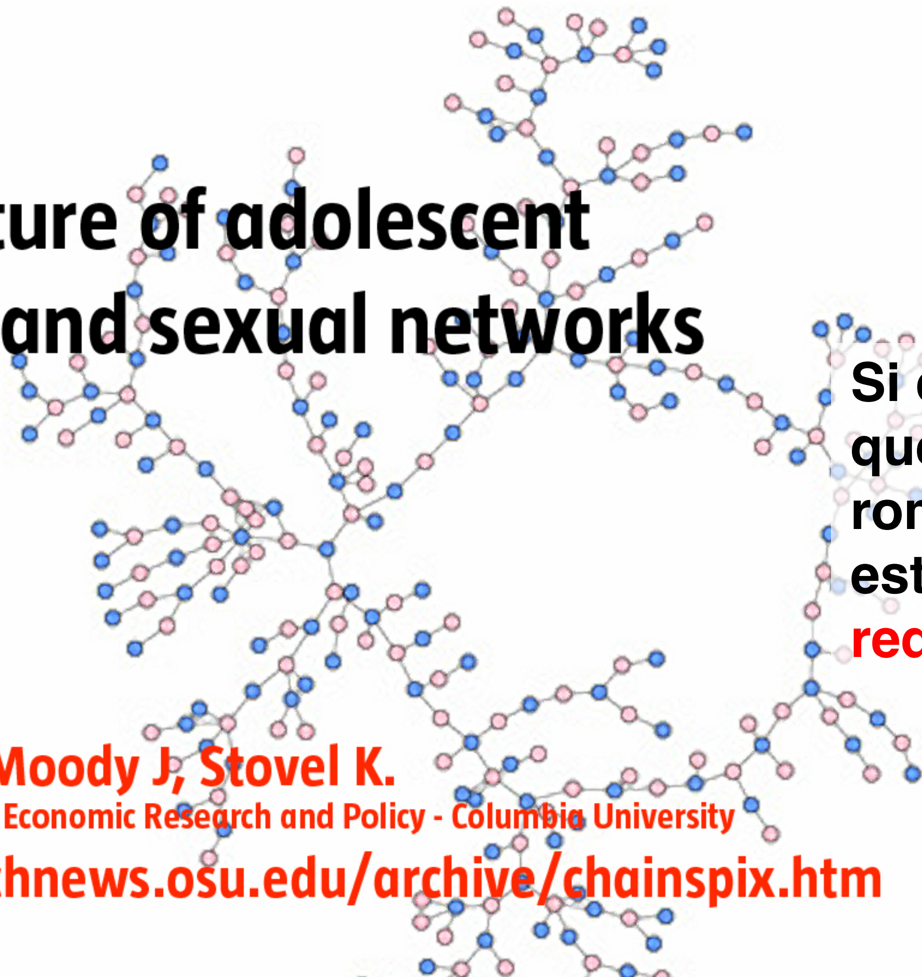
Por ejemplo, la forma en que asignamos los vínculos entre un grupo de individuos determinará la naturaleza de la pregunta que podemos estudiar.

ELIGIENDO UNA REPRESENTACIÓN ADECUADA

They Rule

Si conectas personas que trabajan entre sí, explorarás La red profesional.

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



The structure of adolescent romantic and sexual networks

Si conectas a aquellos que tienen una relación romántica y sexual, estarás explorando las **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

ELIGIENDO UNA REPRESENTACIÓN ADECUADA

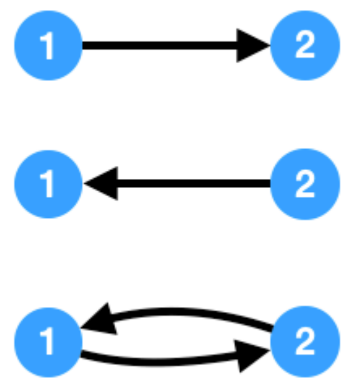
Si conecta a personas en función de su primer nombre (todos los Juanes se conectan entre sí), ¿qué explorarás?

Sin embargo, es una red.

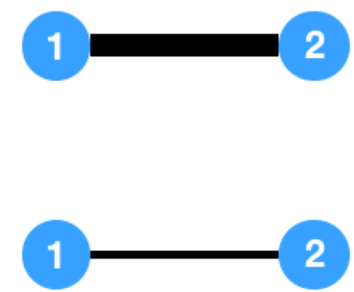
Redes dirigidas vs no dirigidas

Enlaces/Bordes

No-dirigida

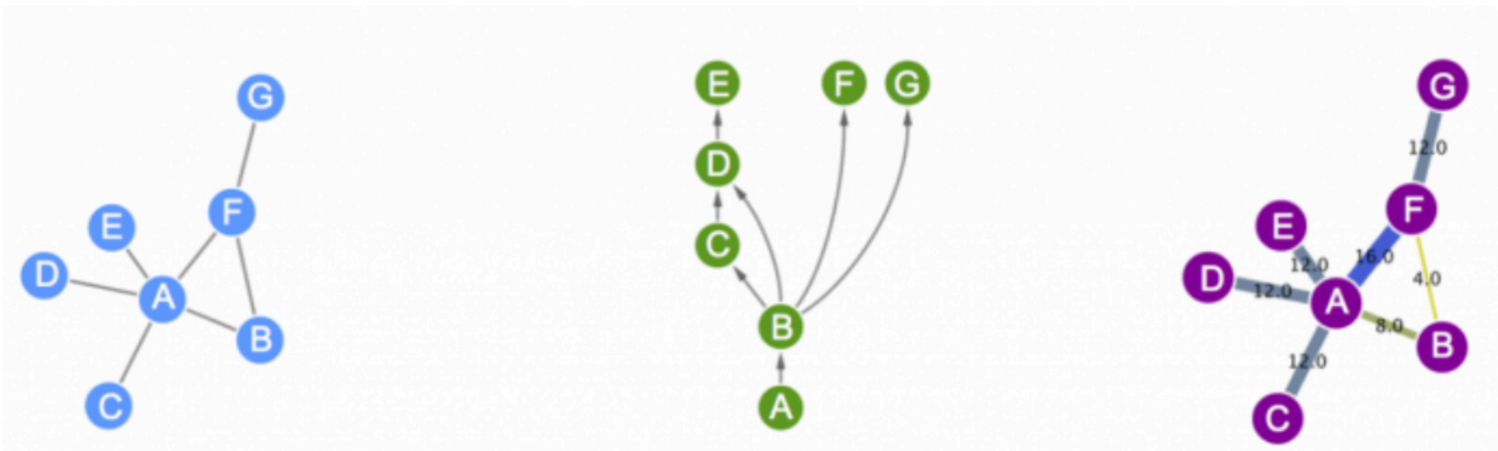


Dirigida



Con pesos

Tipos de Redes



Friendships, Influence

Parenthood, Dependences

Similarity, Financial Ties

No-dirigida

Dirigida

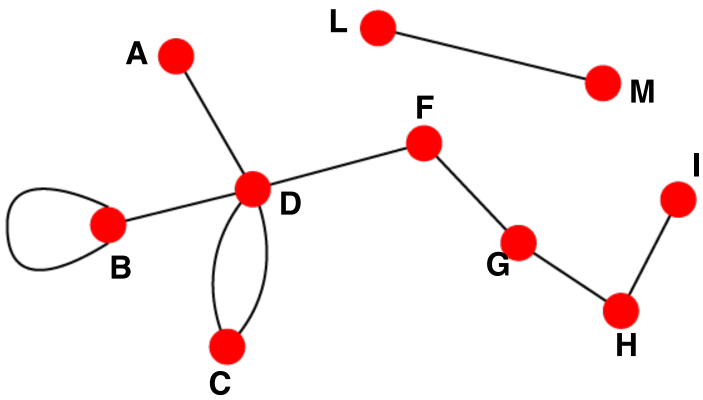
Con pesos

REDES DIRIGIDAS VS REDES NO-DIRIGIDAS

No-dirigidas

Links: no-dirigidos (*simétricos*)

Grafo:

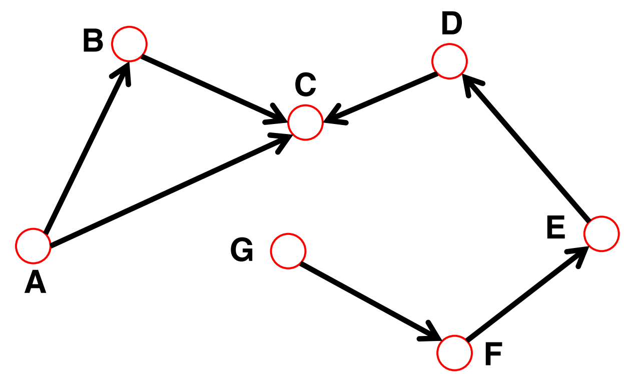


Links no-dirigidos:
Links de co-autores
Red de actores
Interacción de proteínas

Dirigidas

Links: dirigidos (*arcos*).

Digrafo = grafo dirigido:



Un enlace no dirigido es la superposición de dos enlaces dirigidos opuestos.

Links dirigidos:
URLs in el WWW
Llamdas de teléfono
Reacciones metabólicas

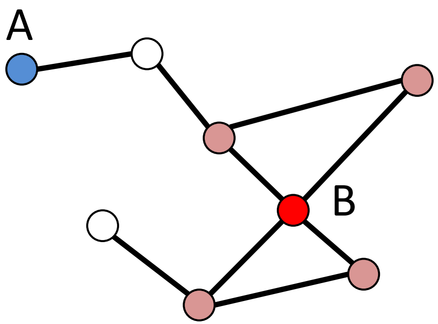
Reference Networks

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L
Internet	Routers	Internet connections	Undirected	192,244	609,066
WWW	Webpages	Links	Directed	325,729	1,497,134
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826
Email	Email addresses	Emails	Directed	57,194	103,731
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908
Citation Network	Paper	Citations	Directed	449,673	4,689,479
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930

Grado, Grado promedio y Distribución de grado

GRADOS DE NODOS

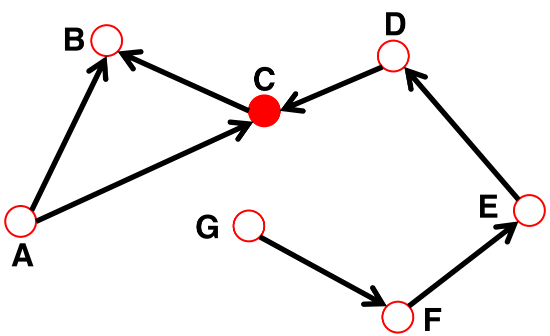
No-dirigido



Grado del nodo: el número de links conectados al nodo.

$$k_A = 1 \qquad k_B = 4$$

Dirigido



En ***redes dirigidas*** Podemos definer un grado de entrada (in-degree) y un grado de salida (out-degree). El grado total es la suma de ambos.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Fuente: un nodo con $k^{in} = 0$; Sumidero: un nodo con $k^{out} = 0$.

UN POCO DE ESTADISTICAS

Breve revisión estadística

Four key quantities characterize a sample of N values $x_1, \ldots, x_N$:

Promedio:

$$\langle x \rangle = \frac{x_1 + x_2 + \ldots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

El n-esimo momento:

$$\langle x^n \rangle = \frac{x_1^n + x_2^n + \ldots + x_N^n}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i^n$$

Desviación estándar

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_i - \langle x \rangle\right)^2}$$

Distribución de *x*

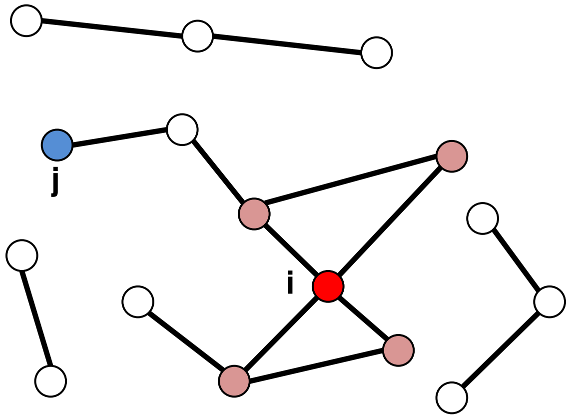
$$p_x = \frac{1}{N}\sum_i \delta_{x,x_i}$$

Donde *x* sigue:

$$\sum_i p_x = 1 \left(\int p_x \, dx = 1 \right)$$

GRADO PROMEDIO

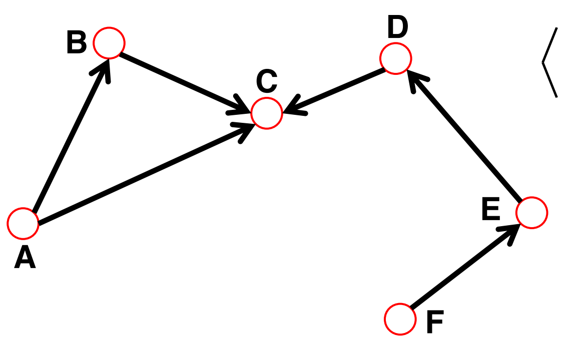
No-dirigido



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

N – el número de nodos en el grafo

Dirigido



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

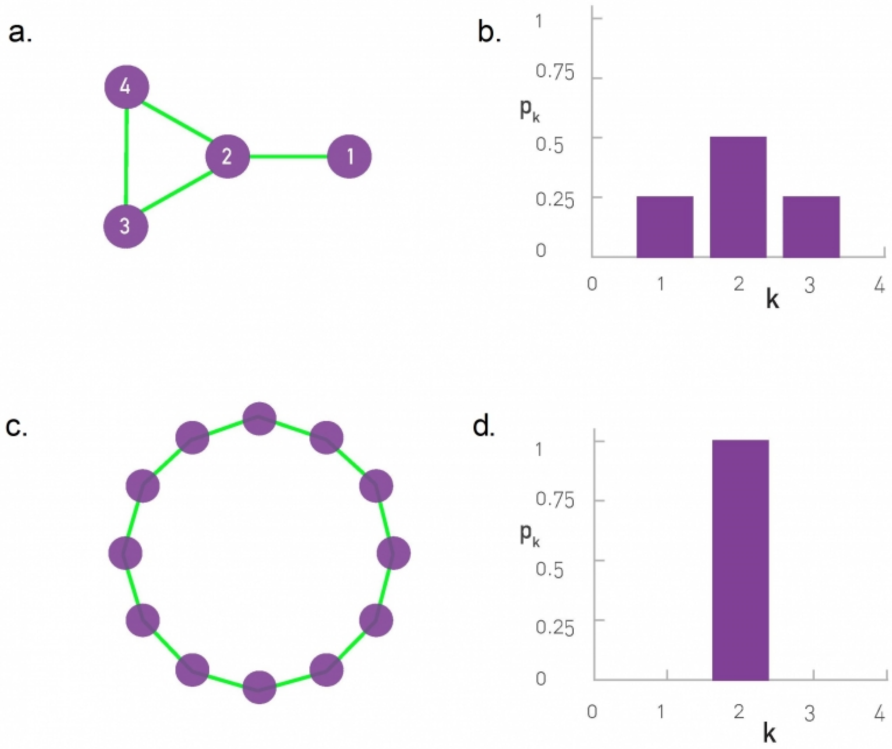
$$\langle k \rangle \equiv \frac{L}{N}$$

GRADO PROMEDIO

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	$\langle k \rangle$
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.33
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-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

Distribución de grado

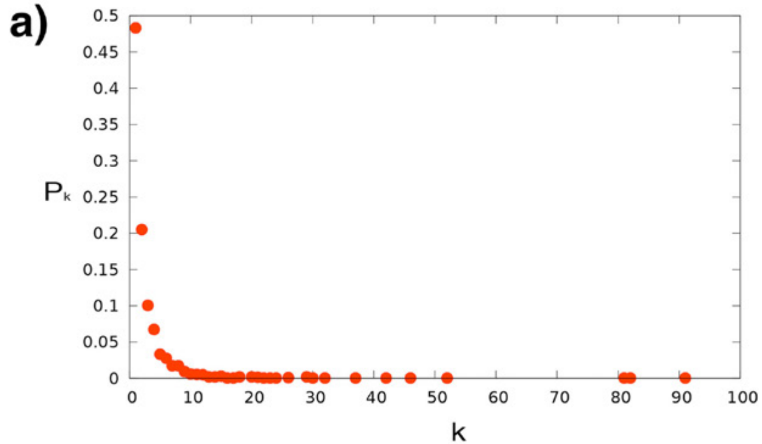
P(k): probabilidad de que un nodo elegído aleatoriamente tenga grado *k*



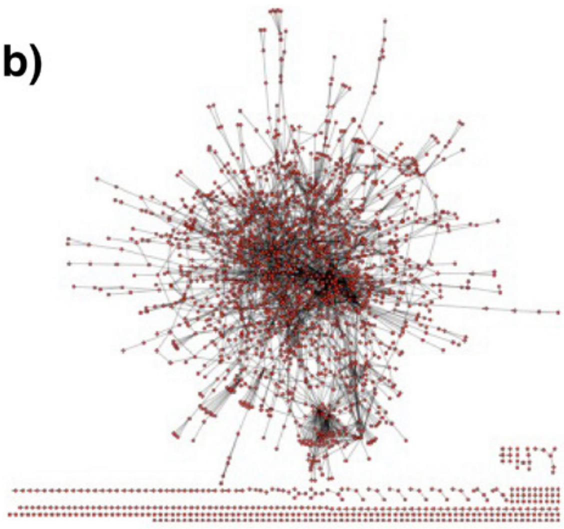
N_k = # nodos con grado k

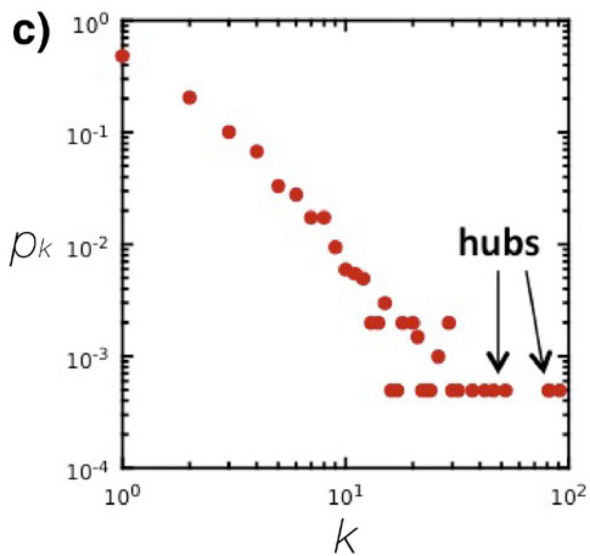
$P(k) = N_k / N$ ➔ plot

DISTRIBUCION DE GRADO



En muchas redes reales, el grado de nodo puede variar considerablemente.
Por ejemplo, como indica la distribución de grados (a), los grados de las proteínas en la red de interacción de proteínas que se muestran en (b) varían entre k = 0 (nodos aislados) y k = 92, que es el grado del nodo más grande, llamado un centro. También hay grandes diferencias en el número de nodos con diferentes grados: como muestra (a), casi la mitad de los nodos tienen grado uno (es decir, p1 = 0,48), mientras que solo hay una copia del nodo más grande, por lo tanto, p_92 = 1 / N = 0.0005. (c) La distribución de grados a menudo se muestra en el llamado gráfico log-log, en el que trazamos log p_k en función de log k, o, como hicimos en (c), usamos ejes logarítmicos.





ige 2.4b

DISTRIBUCION DE GRADO

Representación discreta: $\mathbf{p_k}$ Es la probabilidad de que un nodo tenga grado **k**.

Descripción continua: **p(k)** es la pdf de los grados, donde

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

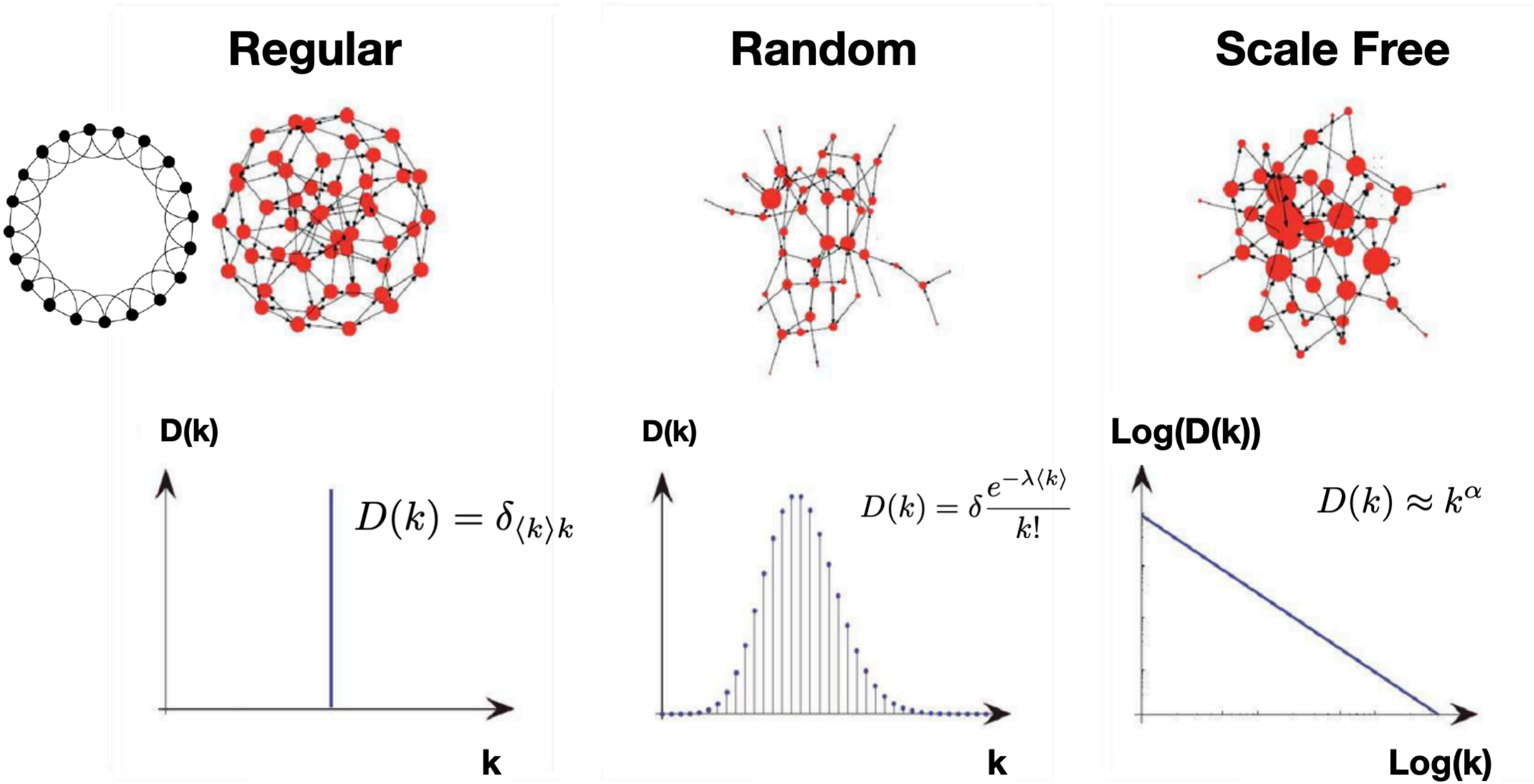
representa la probabilidad de que el grado de un nodo esté entre $\mathbf{k_1}$ y $\mathbf{k_2}$.

Condición de normalización:

$$\sum_{0}^{\infty} p_k = 1 \qquad\qquad \int_{K_{\min}}^{\infty} p(k)dk = 1$$

donde K_{min} es el mínimo grado de la red.

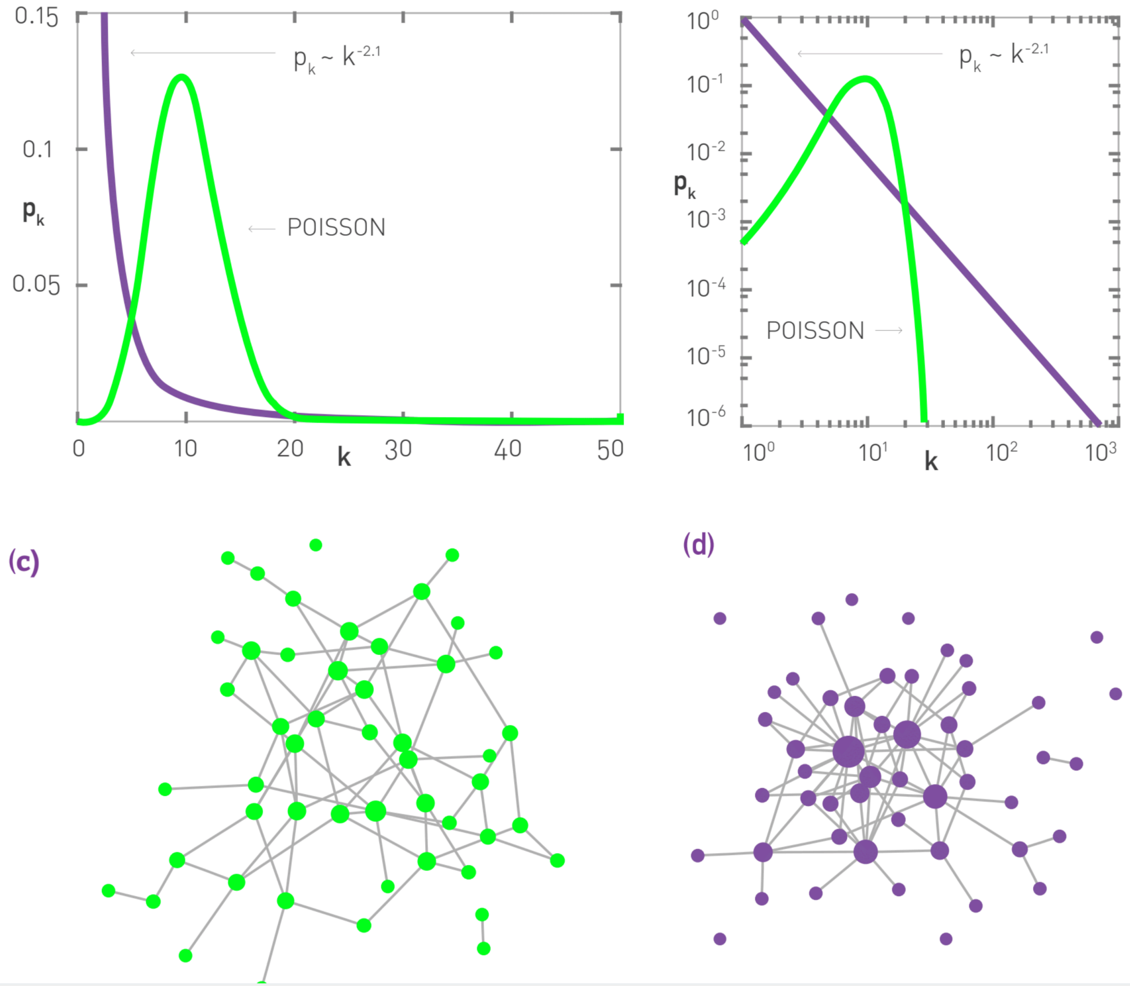
DISTRIBUCION DE GRADO



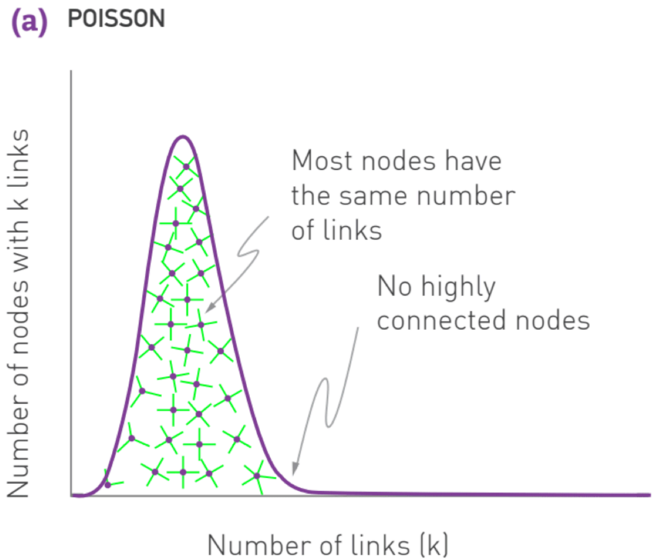
Piensen en los histogramas de fercuencias de grado

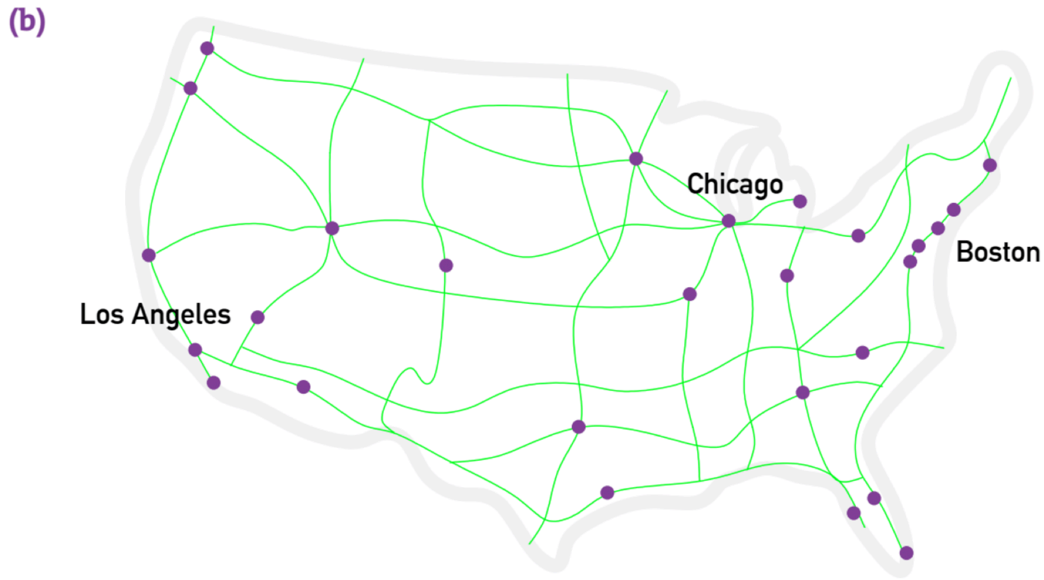
¿Cómo se comparan los promedios con los valores máximos?

Diferencia entre red aletorie y libre de escala

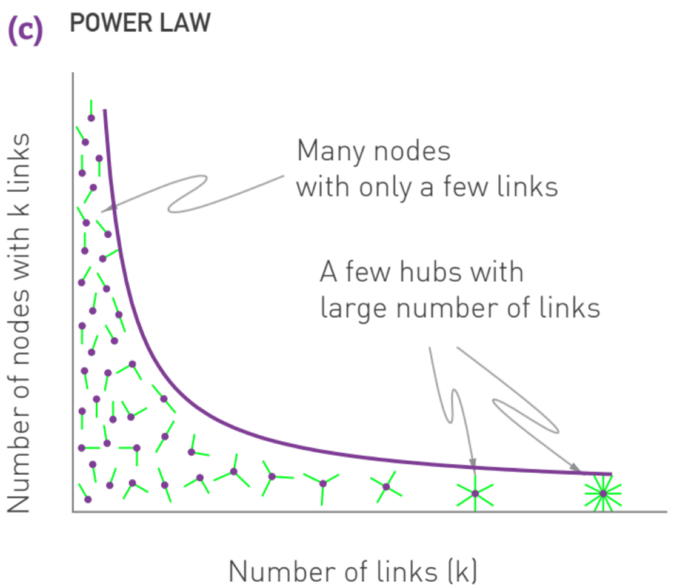


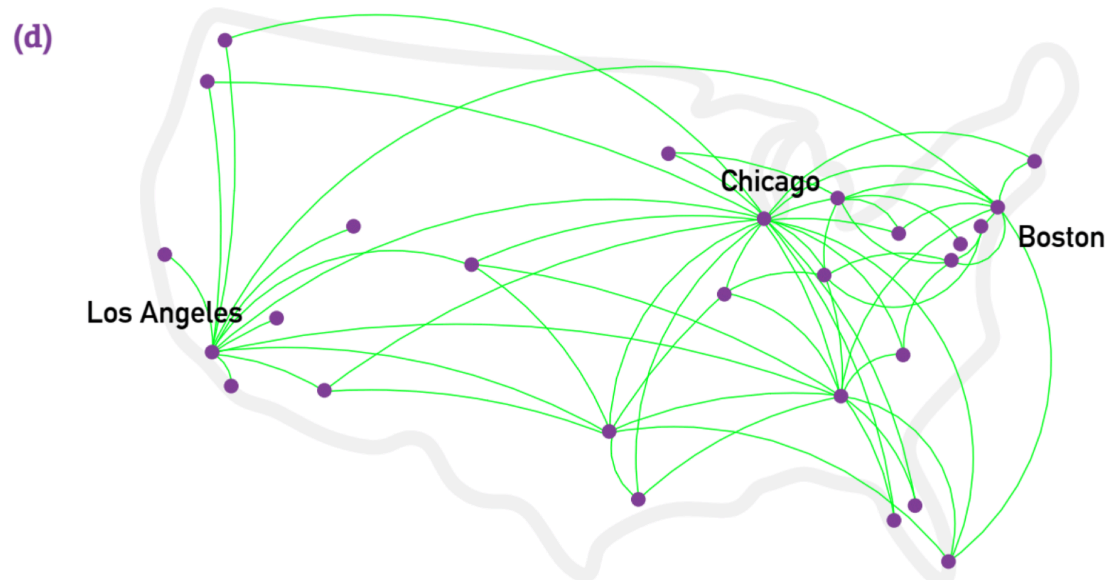
Diferencia entre red aletorie y libre de escala





Diferencia entre red aletorie y libre de escala

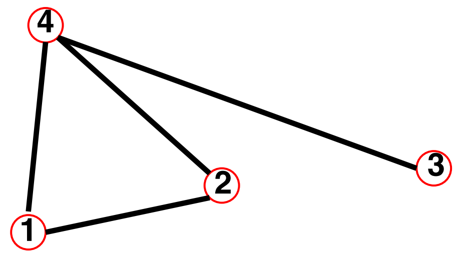


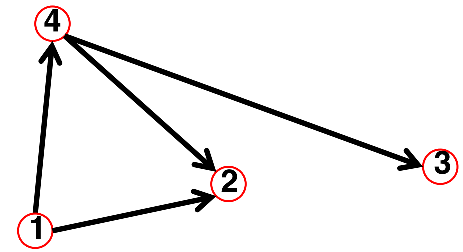


Estas estructuras tienen un impacto significativo en cómo se propaga la información en los sistemas físicos, biológicos, sociales, etc.

Matriz de adyacencia

MATRIZ DE ADYACENCIA





$\mathbf{A_{ij}=1}$ si hay un link entre el nodo *i* y *j*

$\mathbf{A_{ij}=0}$ si los nodos *i* y *j* no estan conectados entre si

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

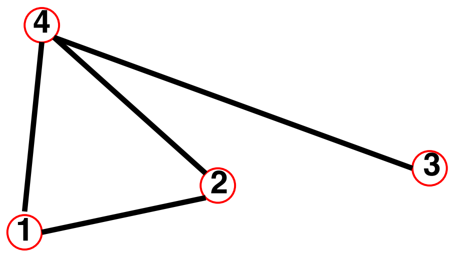
Tenga en cuenta que para un grafo dirigido (derecha) la matriz no es simétrica.

$A_{ij} = 1$ Si hay un link apuntando desde el nodo *j* al *i*

$A_{ij} = 0$ Si hay un link apuntando desde el nodo *i* al *j*

MATRIZ DE ADYACENCIA Y EL GRADO LOS NODOS

No-dirigido



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

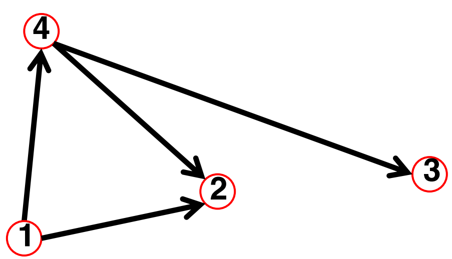
$$A_{ij} = A_{ji}$$
$$A_{ii} = 0$$

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

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

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

Dirigido



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

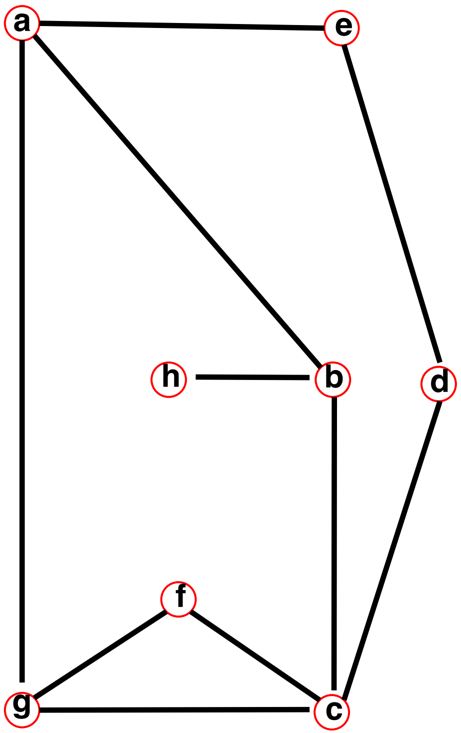
$$A_{ij} \neq A_{ji}$$
$$A_{ii} = 0$$

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

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

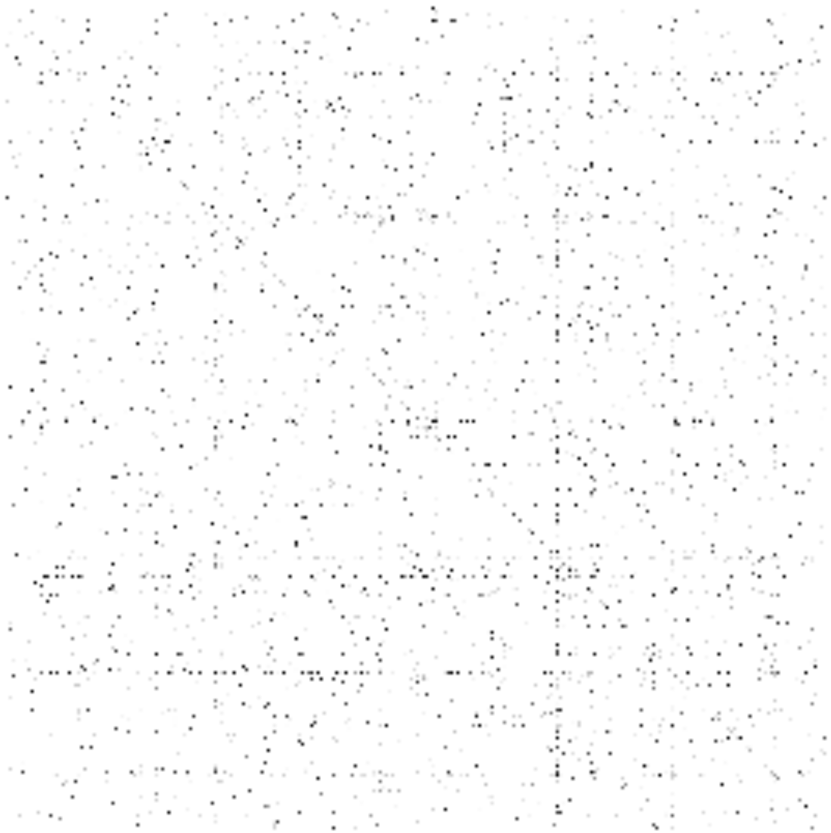
$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{j=1}^{N} k_j^{out} = \sum_{i,j}^{N} A_{ij}$$

MATRIZ DE ADYACENCIA

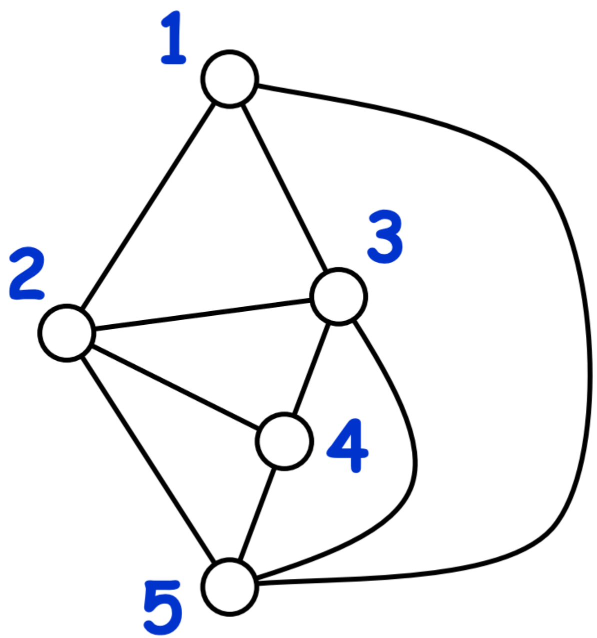


	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	0	0	0
h	0	1	0	0	0	0	0	0

LAS MARICES DE ADYACENCIA SON ¨SPARSE¨



Ejercicio



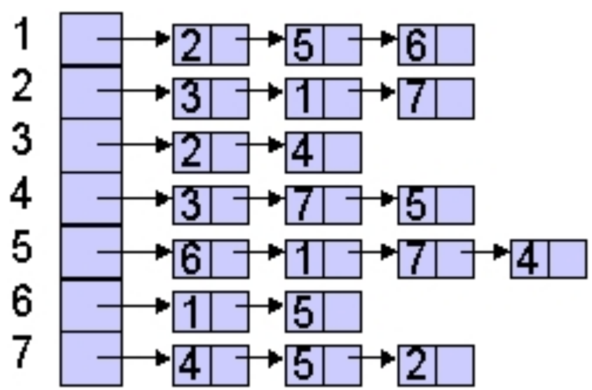
Representa el grafo como una matriz de adyacencia

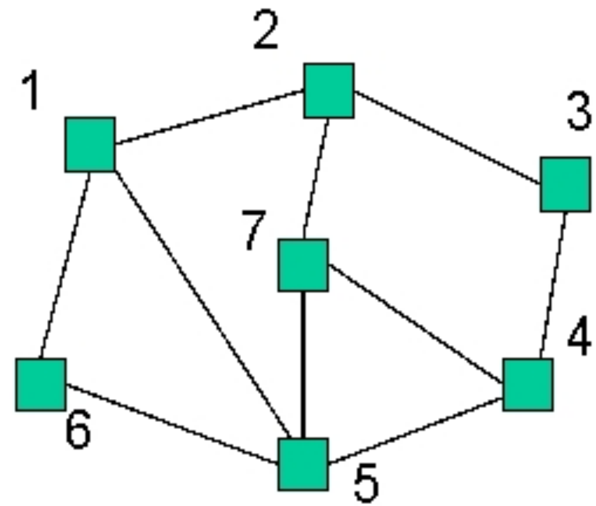
Lista de enlaces y lista de adyacencia

- List of edges

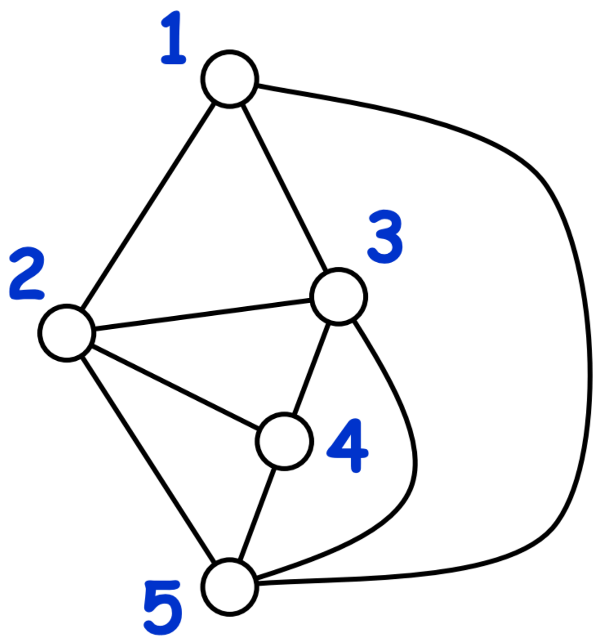
1	5	1	2	2	3	5	7	5	5
2	1	6	7	3	4	6	4	7	4

- Adjacency lists





Ejercicio



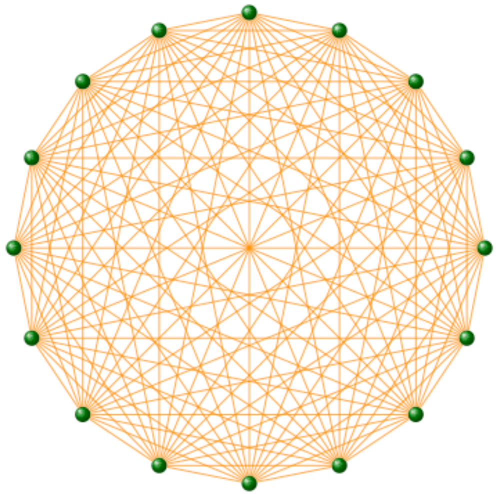
Representa el grafo como una lista de adyacencia
Representa el grafo como una lista de enlaces

Redes reales son poco densas (sparse)

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 grado L=L_{max} es llamado **grafo completo**, y su grado promedio es **<k>=N-1**

$$C_n^p = \binom{n}{p} = \frac{n!}{p!(n-p)!}\,.$$

REDES REALES SON "SPARSE"

La mayoría de las redes observadas en sistemas reales son poco densas:

$$L \ll L_{max}$$

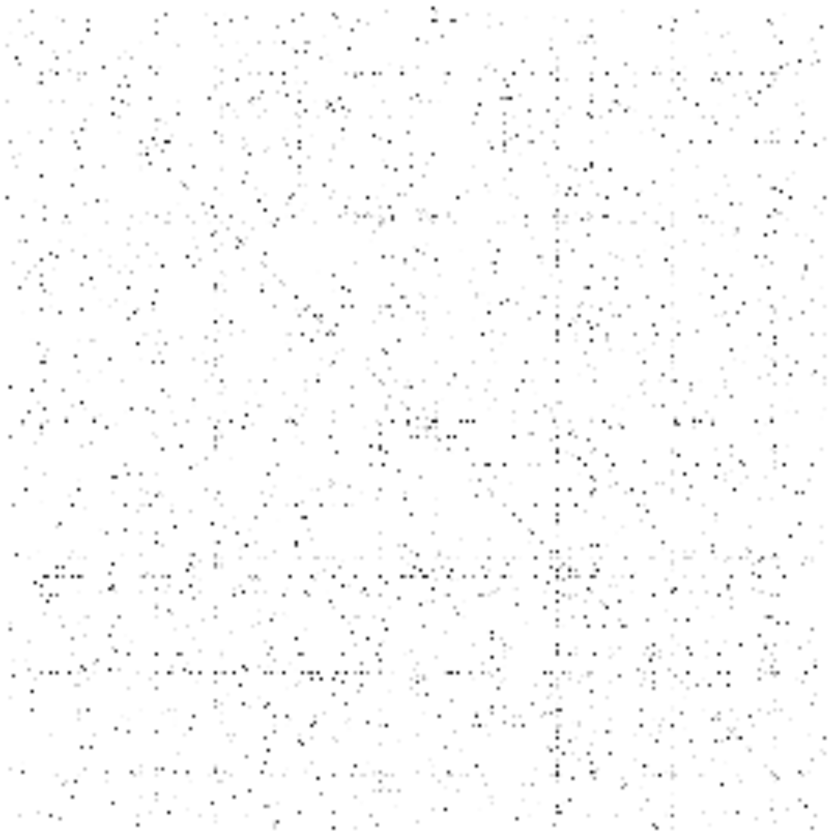
o

$$\langle k \rangle \ll N-1.$$

WWW (ND Sample):	N=325,729;	L=1.4 10^6	$L_{max}=10^{12}$	$\langle k \rangle=4.51$
Protein (*S. Cerevisiae*):	N= 1,870;	L=4,470	$L_{max}=10^7$	$\langle k \rangle=2.39$
Coauthorship (Math):	N= 70,975;	L=2 10^5	$L_{max}=3\ 10^{10}$	$\langle k \rangle=3.9$
Movie Actors:	N=212,250;	L=6 10^6	$L_{max}=1.8\ 10^{13}$	$\langle k \rangle=28.78$

(Source: Albert, Barabasi, RMP2002)

LAS MARICES DE ADYACENCIA SON "SPARSE"

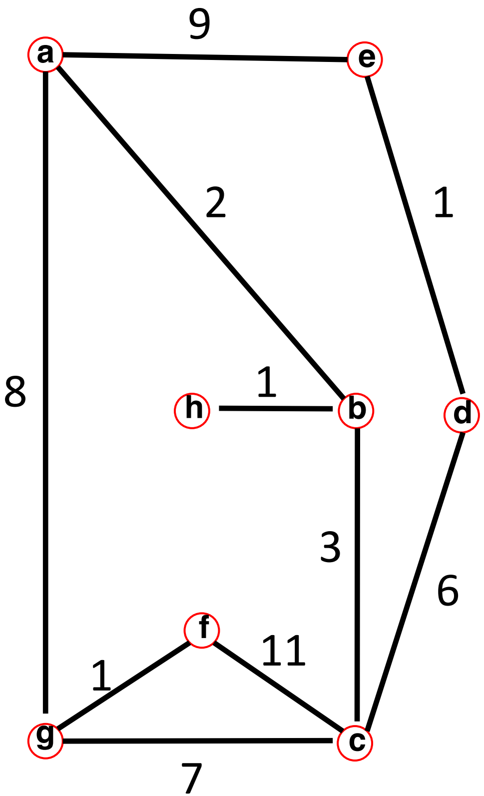


REDES CON Y SIN PESOS

REDES CON Y SIN PESOS

$$A_{ij} = w_{ij}$$

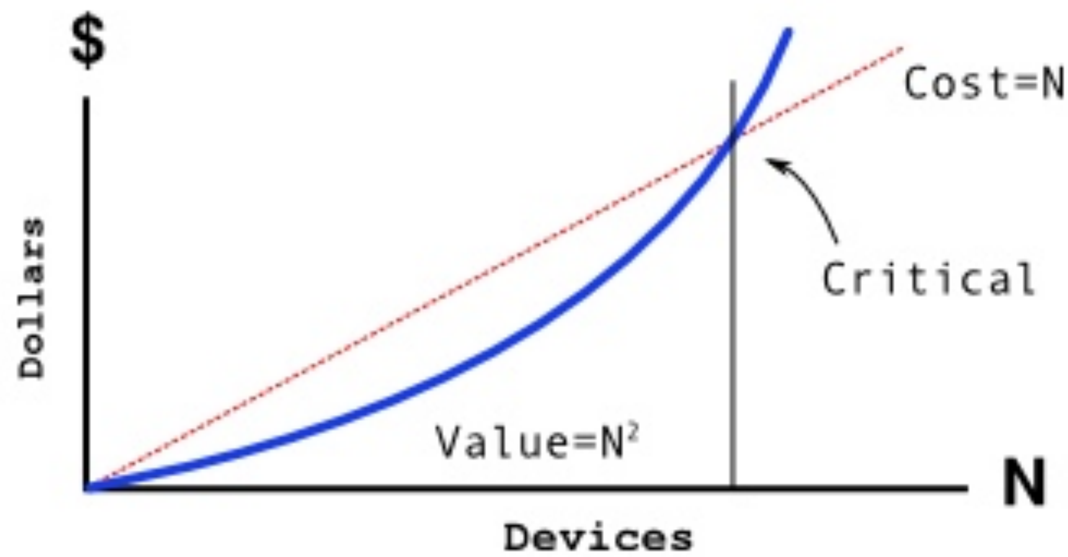
MATRIZ DE ADYACENCIA CON PESOS



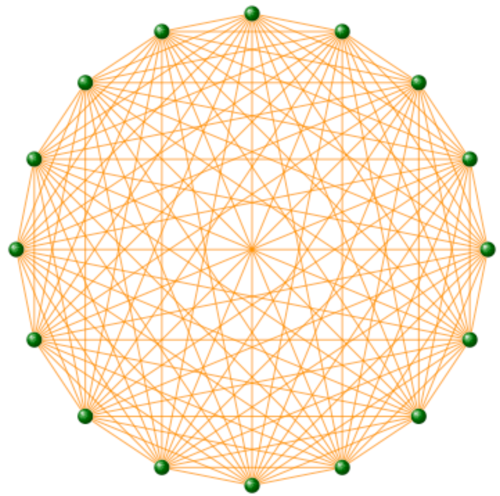
$$A_{ij} = w_{ij}$$

	a	b	c	d	e	f	g	h
a	0	2	0	0	9	0	8	0
b	2	0	3	0	0	0	0	1
c	0	3	0	6	0	11	7	0
d	0	0	6	0	1	0	0	0
e	9	0	0	1	0	0	0	0
f	0	0	11	0	0	0	1	0
g	8	0	7	0	0	1	0	0
h	0	1	0	0	0	0	0	0

METCALFE'S LAW



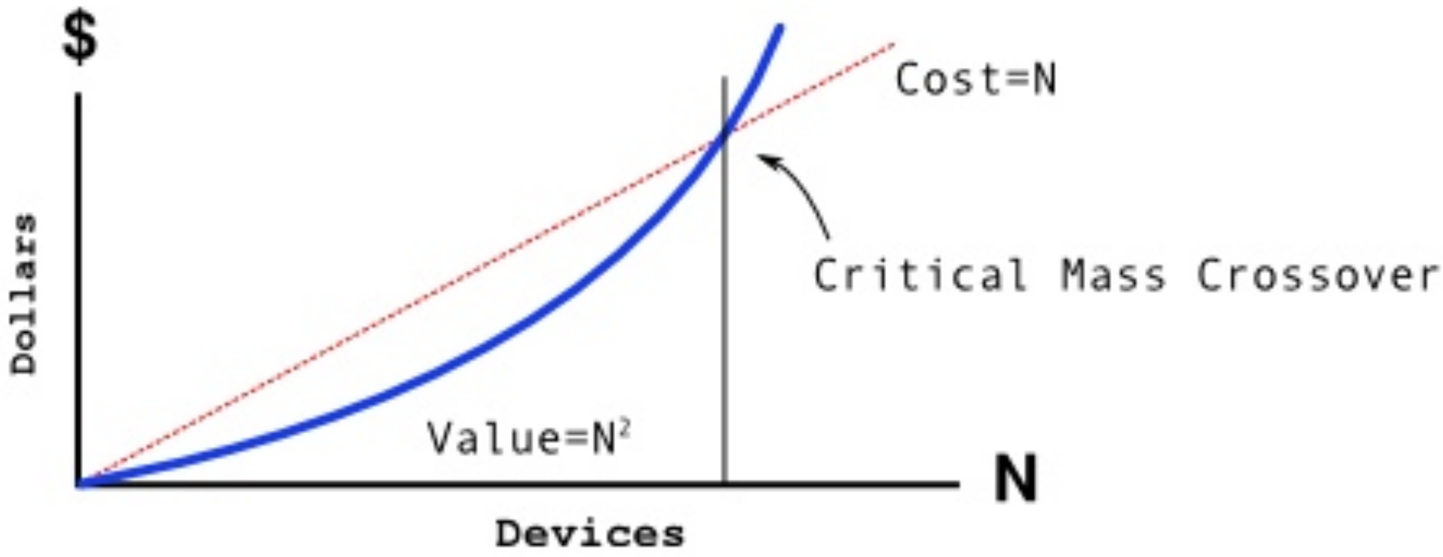
"mientras más personas usan una red, más valiosa se vuelve"



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

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

METCALFE'S LAW



Hay dos problemas fundamentales con la ley de Metcalfe:

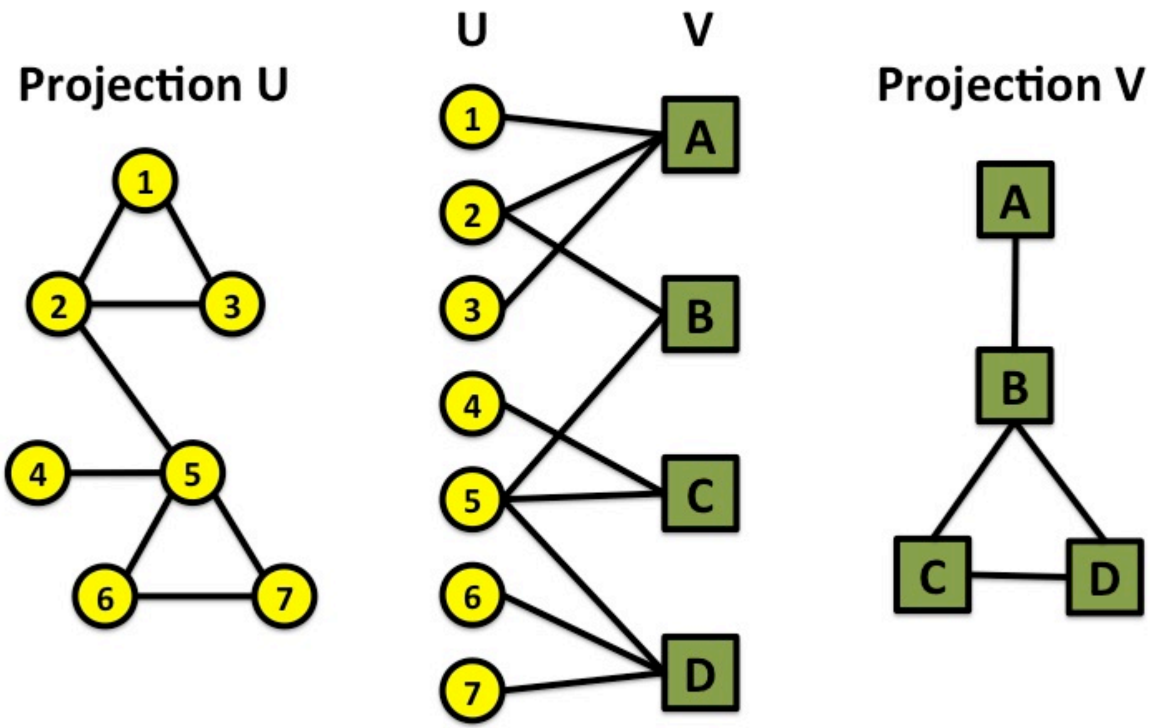
Si bien todos los enlaces son posibles, en redes reales no todos los enlaces están presentes. De hecho, la mayoría de las redes reales son dispersas, lo que significa que solo una pequeña fracción de los enlaces están presentes. Si asignamos un valor a cada enlace, entonces el valor total de la red crecerá más lentamente que N ^ 2, como veremos en los próximos capítulos.

No todos los enlaces son de igual valor (pesos). Algunos enlaces se usan mucho mientras que la mayoría de los enlaces son 'débiles', es decir, rara vez se utilizan.

REDES BIPARTITAS

GRAFOS BIPARTITOS

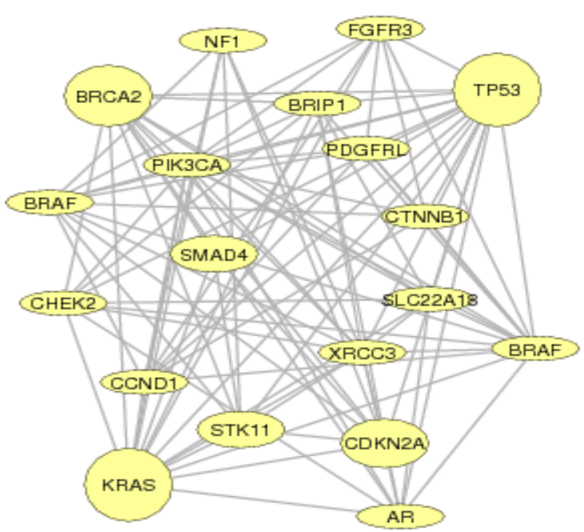
Un **grafo bipartito** (o bigrafo) es un gráfo cuyos nodos se pueden dividir en dos **conjuntos separados** U y V, de manera que cada enlace conecta un nodo en U con uno en V; es decir, U y V son conjuntos **independientes**.



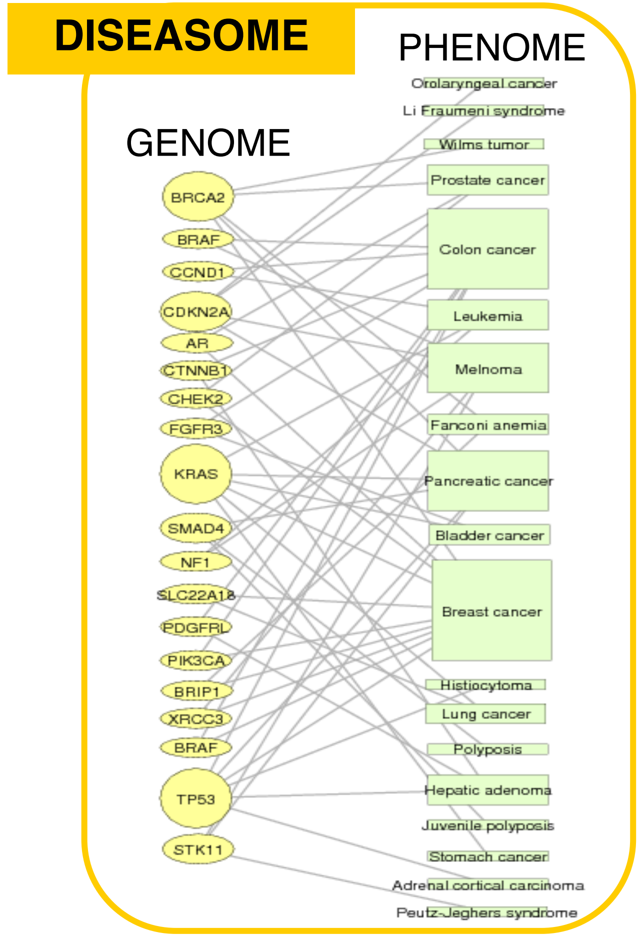
Ejemplos:

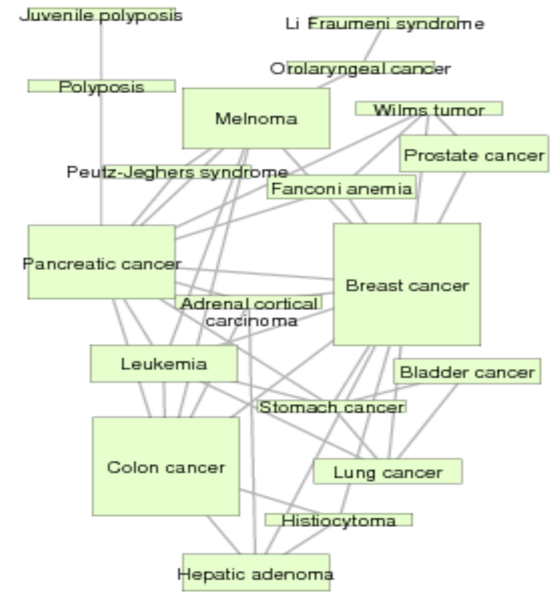
Red de actores de Hollywood
Redes de colaboración
Red de enfermedades (diseasome)

Red de genes– Red de enfermedades



Gene network



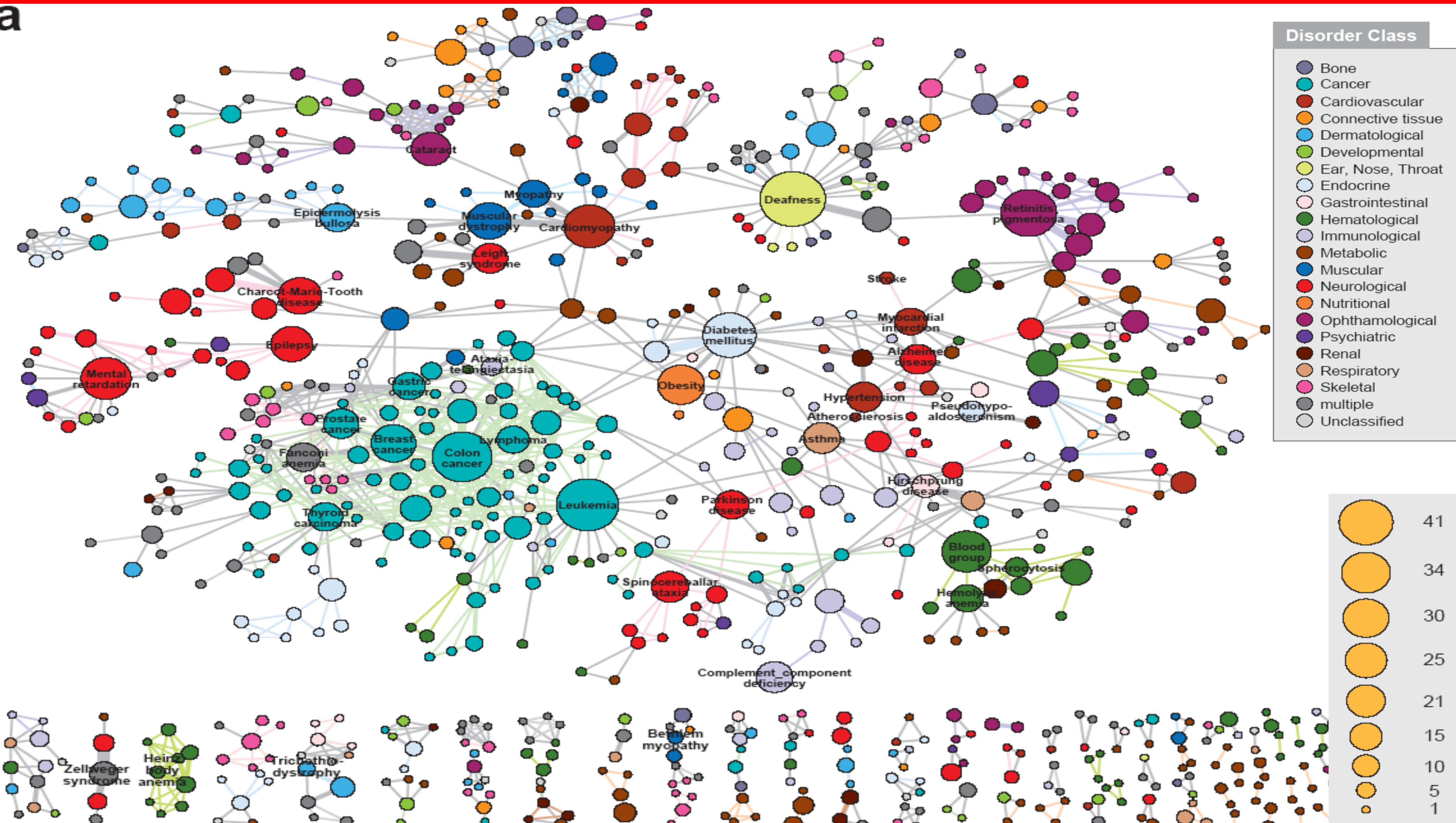


Disease network

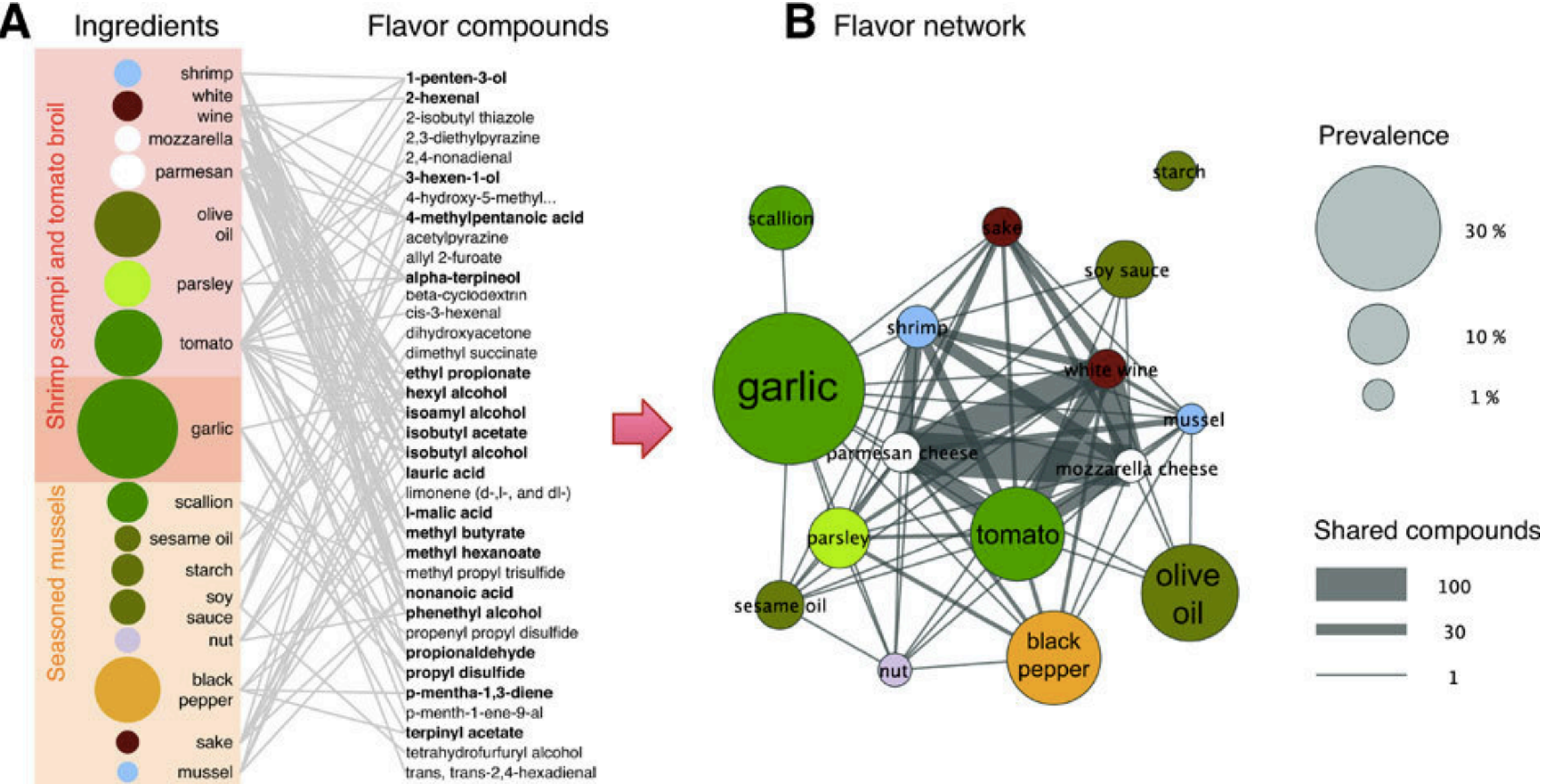
Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

RED DE ENFERMEDADES HUMANAS

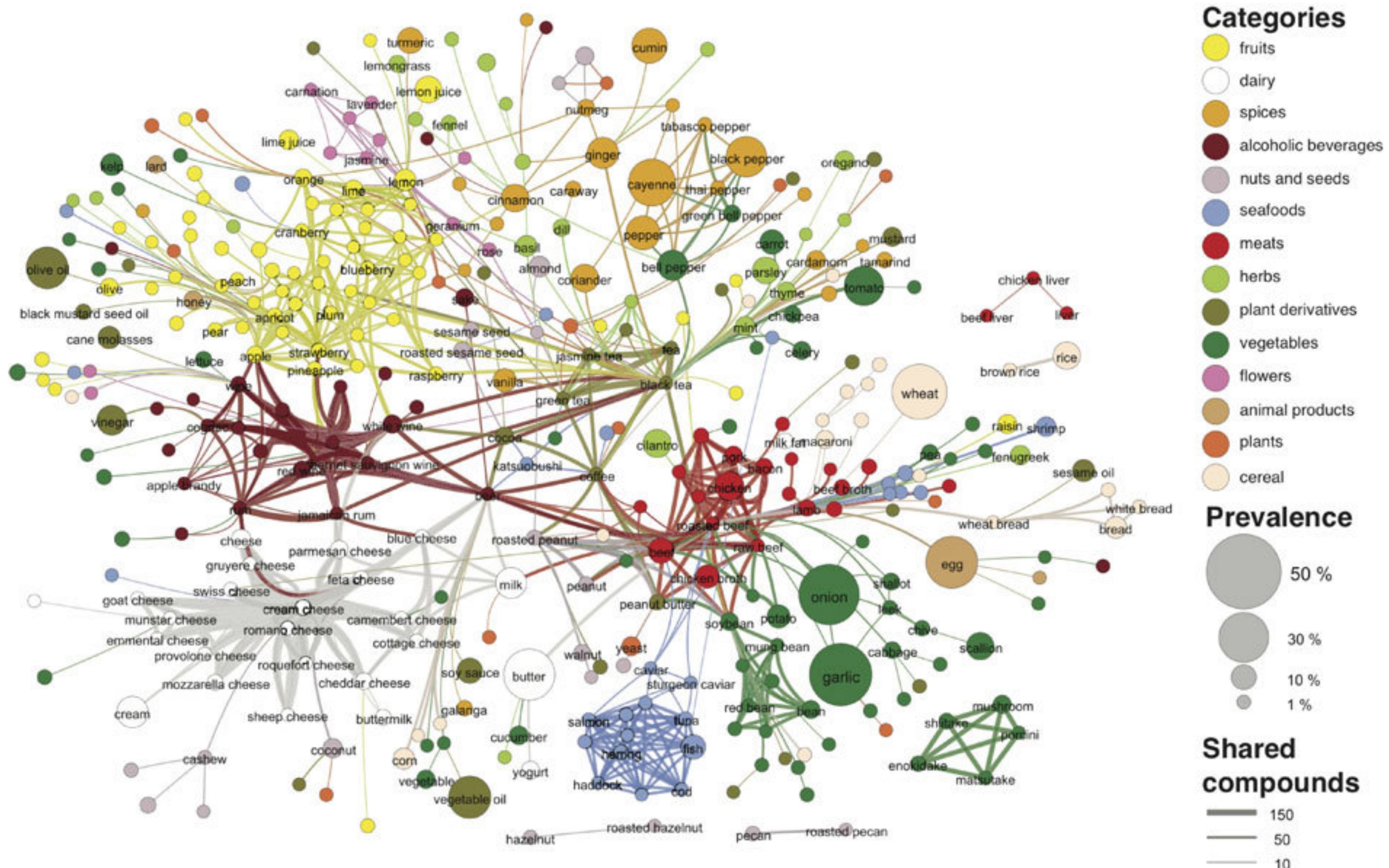
https://www.nytimes.com/2008/05/06/health/research/06dise.html



Red bipartita ingredient-sabor

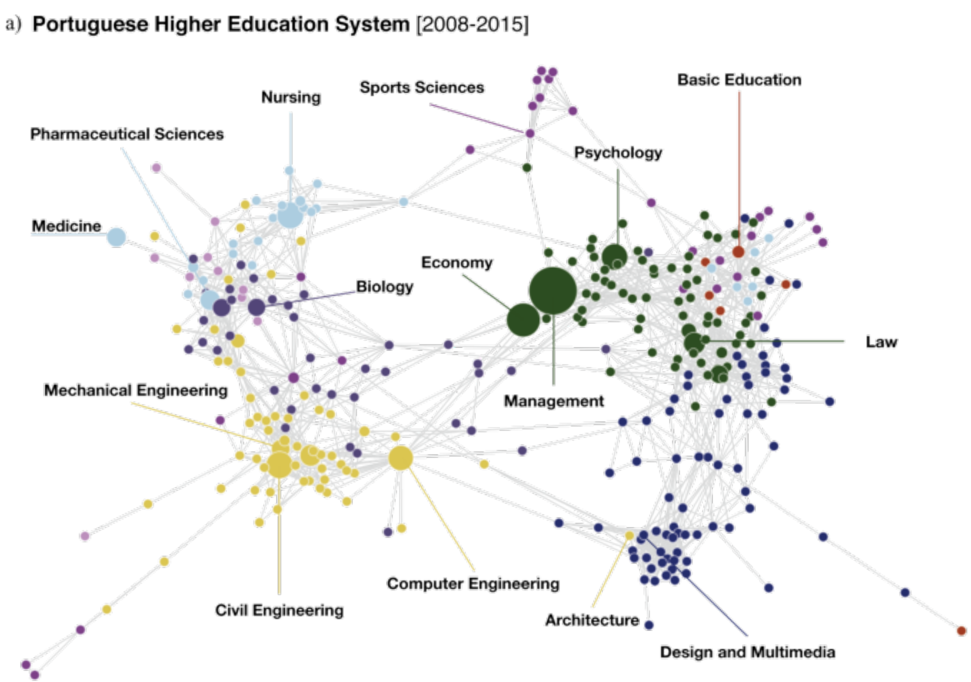


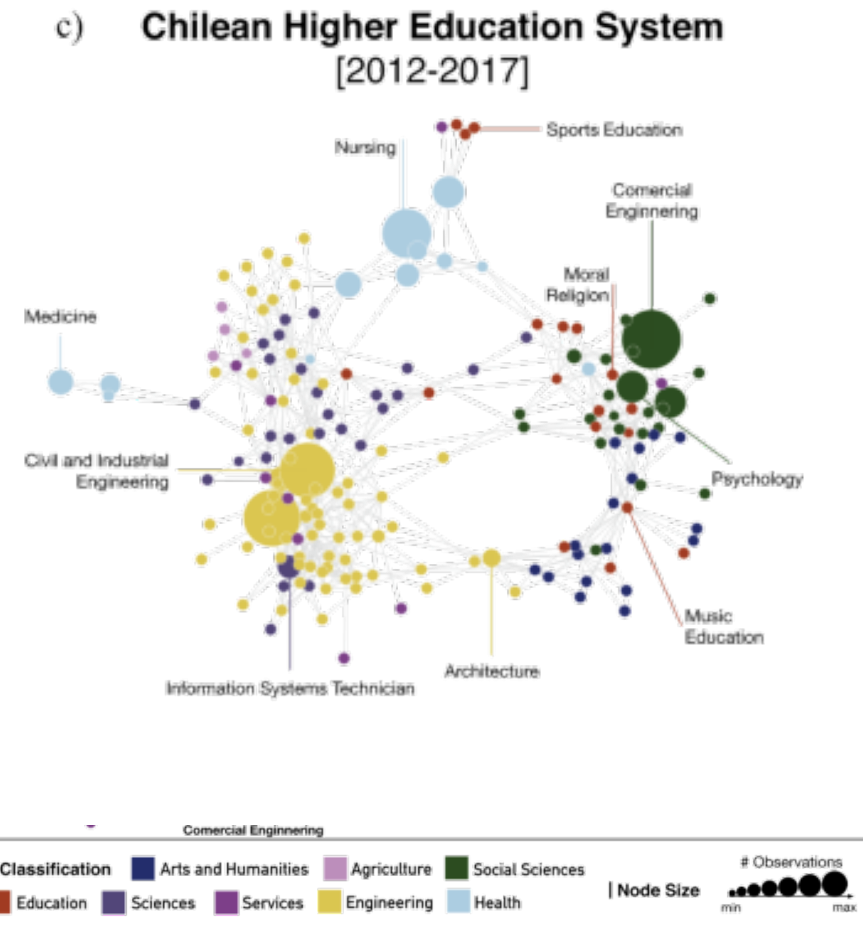
Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási Flavor network and the principles of food pairing , Scientific Reports 196, (2011).



Categories
fruits
dairy
spices
alcoholic beverages
nuts and seeds
seafoods
meats
herbs
plant derivatives
vegetables
flowers
animal products
plants
cereal
Prevalence
50 %
30 %
10 %
1 %
Shared compounds
150
50
10
turmeric
cumin
lemongrass
carnation
lavender
lemon juice
nutmeg
fennel
tabasco pepper
lime juice
jasmine
ginger
black pepper
kelp
lard
orange
lime
lemon
cayenne
thai pepper
oregano
cinnamon
caraway
green bell pepper
cranberry
geranium
pepper
dill
mustard
basil
rose
carrot
almond
bell pepper
cardamom
tamarind
olive oil
blueberry
coriander
parsley
peach
tomato
olive
thyme
chicken liver
honey
sake
black mustard seed oil
apricot
plum
chickpea
beef liver
liver
pear
sesame seed
mint
cane molasses
apple
strawberry
roasted sesame seed
tea
celery
lettuce
jasmine tea
rice
pineapple
raspberry
brown rice
wine
vanilla
black tea
wheat
green tea
vinegar
raisin
shrimp
cognac
white wine
cocoa
cilantro
milk fat
macaroni
pea
fenugreek
cabernet sauvignon wine
katsuobushi
pork
red wine
coffee
bacon
sesame oil
apple brandy
chicken
beef broth
beer
lamb
white bread
rum
jamaican rum
roasted beef
wheat bread
bread
blue cheese
roasted peanut
cheese
raw beef
parmesan cheese
beef
egg
gruyere cheese
feta cheese
chicken broth
milk
shallot
swiss cheese
peanut
goat cheese
onion
cream cheese
peanut butter
leek
munster cheese
camembert cheese
potato
soybean
romano cheese
emmental cheese
chive
cottage cheese
mung bean
scallion
yeast
walnut
provolone cheese
cabbage
roquefort cheese
caviar
soy sauce
garlic
butter
mozzarella cheese
sturgeon caviar
cheddar cheese
mushroom
cream
red bean
sheep cheese
tuna
galanga
bean
salmon
buttermilk
shiitake
cucumber
porcini
fish
coconut
herring
cashew
corn
enokidake
yogurt
vegetable
matsutake
haddock
cod
vegetable oil
roasted hazelnut
roasted pecan
pecan
hazelnut

Red bipartita carreras-postulantes

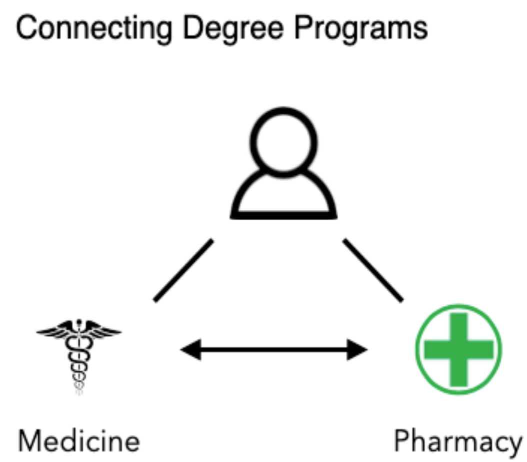




Candia, Encarnaçao, Pinheiro. EPJ Data Science 2019*

Red bipartita carreras-postulantes

The Higher Education Space



Lista de Preferencias

1.	Medicina
2.	Odontología
3.	Tecnología Médica
4.	Odontología
5.	Cs. Físicas y Astronómicas

Consideramos todas las preferencias de cada postulantes. Las preferencias repetidas indican una postulación a dos instituciones de educación superior distintas.

Pares de Carreras

Medicina	⟷	Odontología	Odontología	⟷	Odontología
Medicina	⟷	Tecnología Médica	Odontología	⟷	Cs. Físicas y Astronómicas
Medicina	⟷	Odonotología	Tecnología Médica	⟷	Odontología
Medicina	⟷	Cs. Físicas y Astronómicas	Tecnología Médica	⟷	Cs. Físicas y Astronómicas
Odontología	⟷	Tecnología Médica	Odontología	⟷	Cs. Físicas y Astronómicas

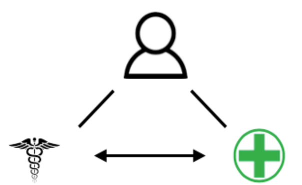
Creamos todos los pares de carreras posibles. Luego, descartamos los que contienen la misma carrera en ambos extremos (color gris).

Candia, Encarnaçao, Pinheiro. EPJ Data Science 2019*

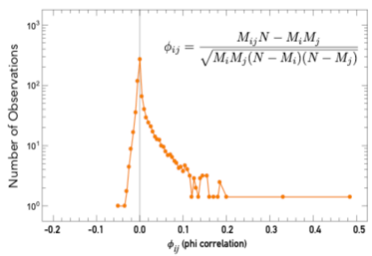
Red bipartita carreras-postulantes

The Higher Education Space

Connecting Degree Programs

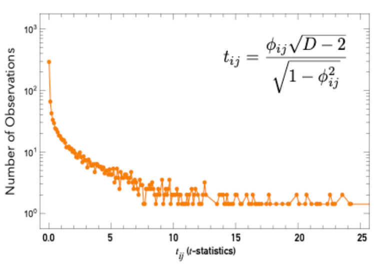


Finding Correlations Between Degrees

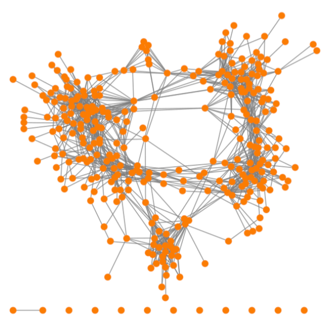


Discard all negative correlations

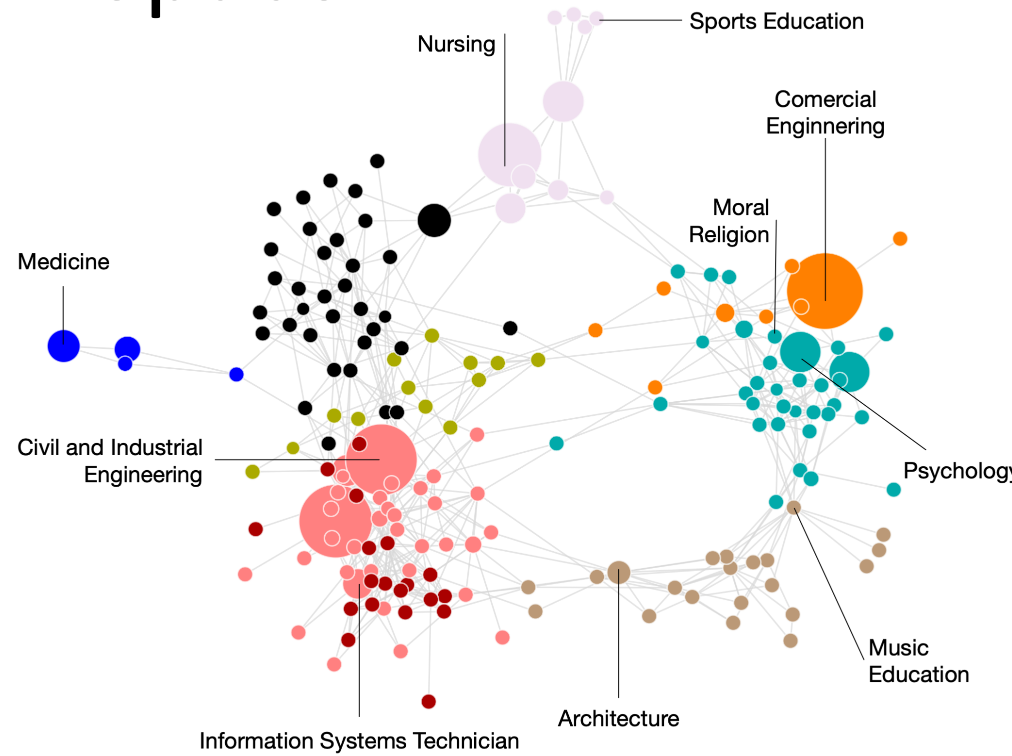
Finding Statistical Significance of Correlations



Discard all non-significant links



Discard all loose Nodes



Candia, Encarnaçao, Pinheiro. EPJ Data Science 2019*

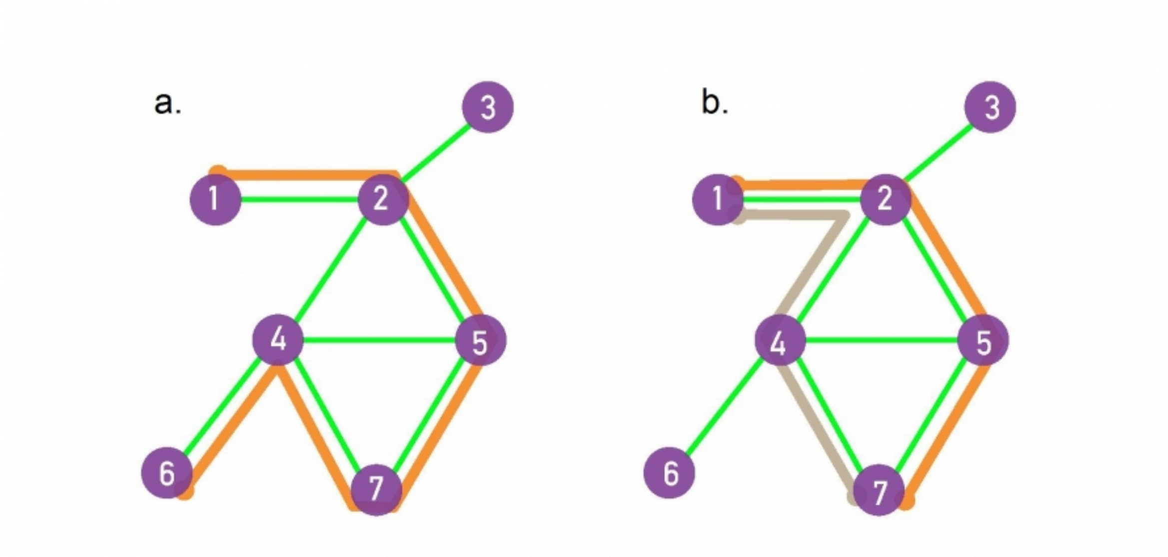
“CAMINOLOGIA”
(PATHOLOGY)

CAMINOS

Un *camino (path) es una secuencia de nodos en los que cada nodo es adyacente al siguiente*

$P_{i0,in}$ de longitud *n* entre los nodos i_0 y i_n es una colección ordenada de *n+1* nodos y *n* links

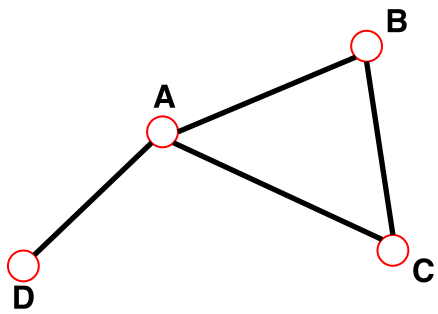
$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$



- En una red dirigida, la ruta solo puede seguir la dirección de una flecha.

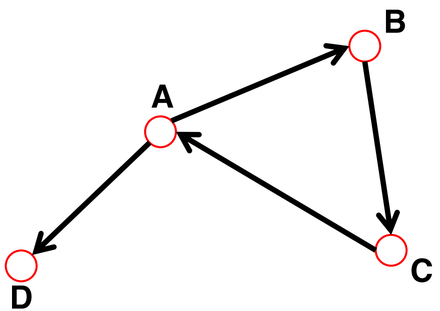
DISTANCIA EN UN GRAFO

Shortest Path, Geodesic Path



La distancia (ruta más corta, ruta geodésica) entre dos nodos se define como el número de enlaces a lo largo de la ruta más corta que los conecta.

* Si los dos nodos están desconectados, la distancia es infinita.



En los grafos dirigidos, cada ruta debe seguir la dirección de las flechas.

Así, en un grafo, la distancia desde el nodo A hasta B (en una ruta AB) es generalmente diferente de la distancia desde el nodo B hasta A (en una ruta BCA).

N_{ij}, númenro de caminos entre dos nodos cualqueira *i* y *j*:

Longitud n=1: Si existe un link entre *i* y *j*, entonces A_{ij}=1 y A_{ij}=0 en otro caso.

Longitud n=2: Si existe un camino de longitude dos entre *i* y *j*, entonces $A_{ik}A_{kj}$=1, y $A_{ik}A_{kj}$=0 en otro caso.
El número de caminos de longitud 2 es:

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

Longitud n: en general, si existe un camino de longitud *n* entre *i* y *j*, entonces $A_{ik}\ldots A_{lj}$=1 y $A_{ik}\ldots A_{lj}$=0 en otro caso.
El número de caminos de longitude *n* entre *i* y *j* es*

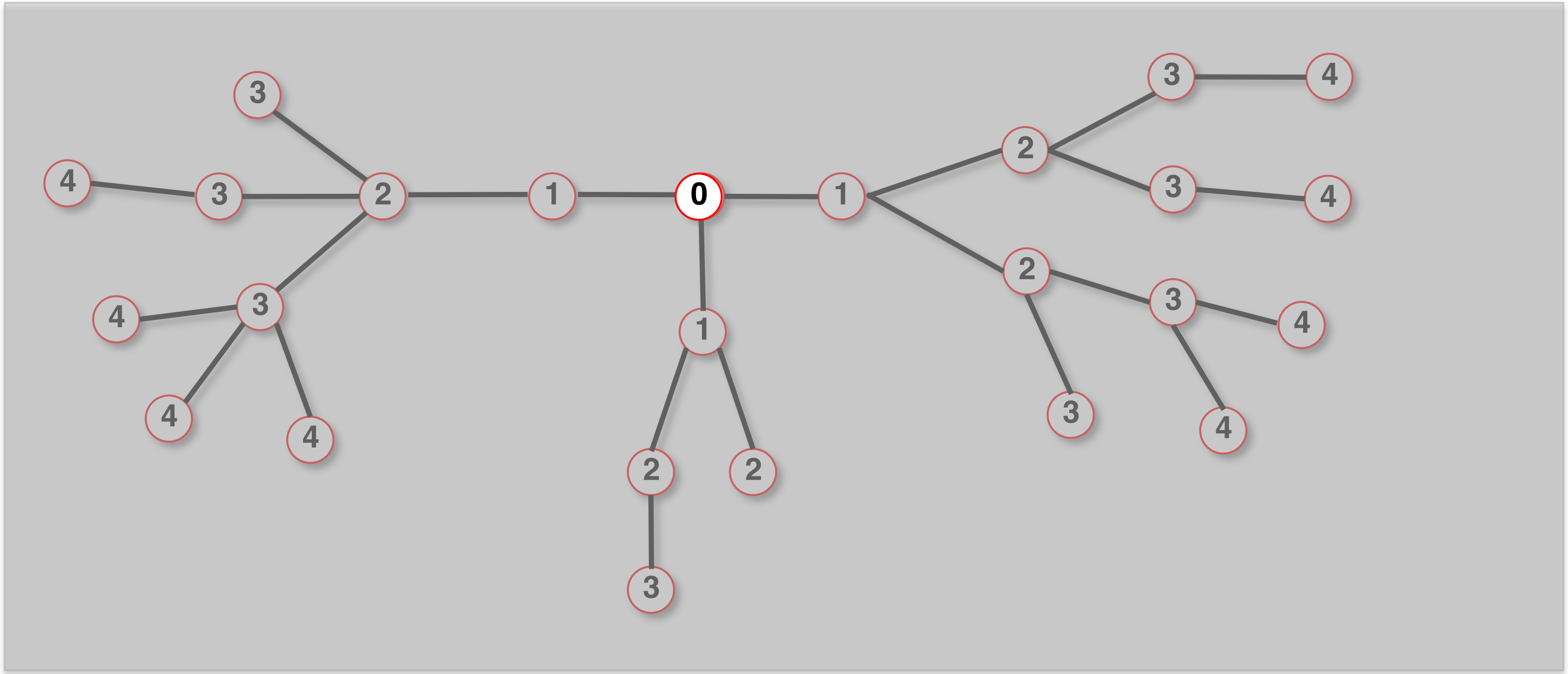
$$N_{ij}^{(n)} = [A^n]_{ij}$$

**Se aplica tanto a redes dirigidas como no dirigidas.*

ECONTRANDO DISTANCIAS: "BREADTH FIRST SEARCH"

La distancia entre el nodo 0 y nodo 4:

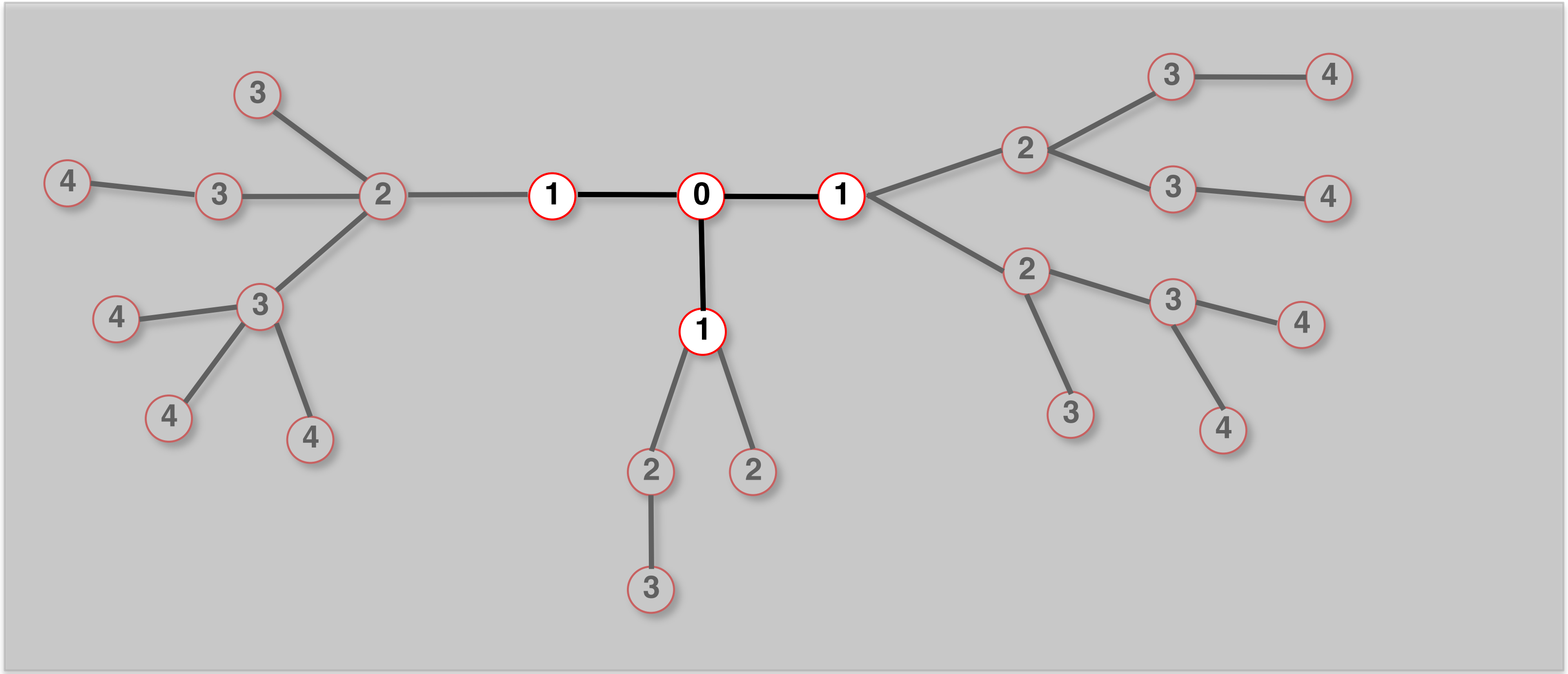
1. Comienza en 0.



ECONTRANDO DISTANCIAS: “BREADTH FIRST SEARCH”

La distancia entre el nodo 0 y nodo 4:

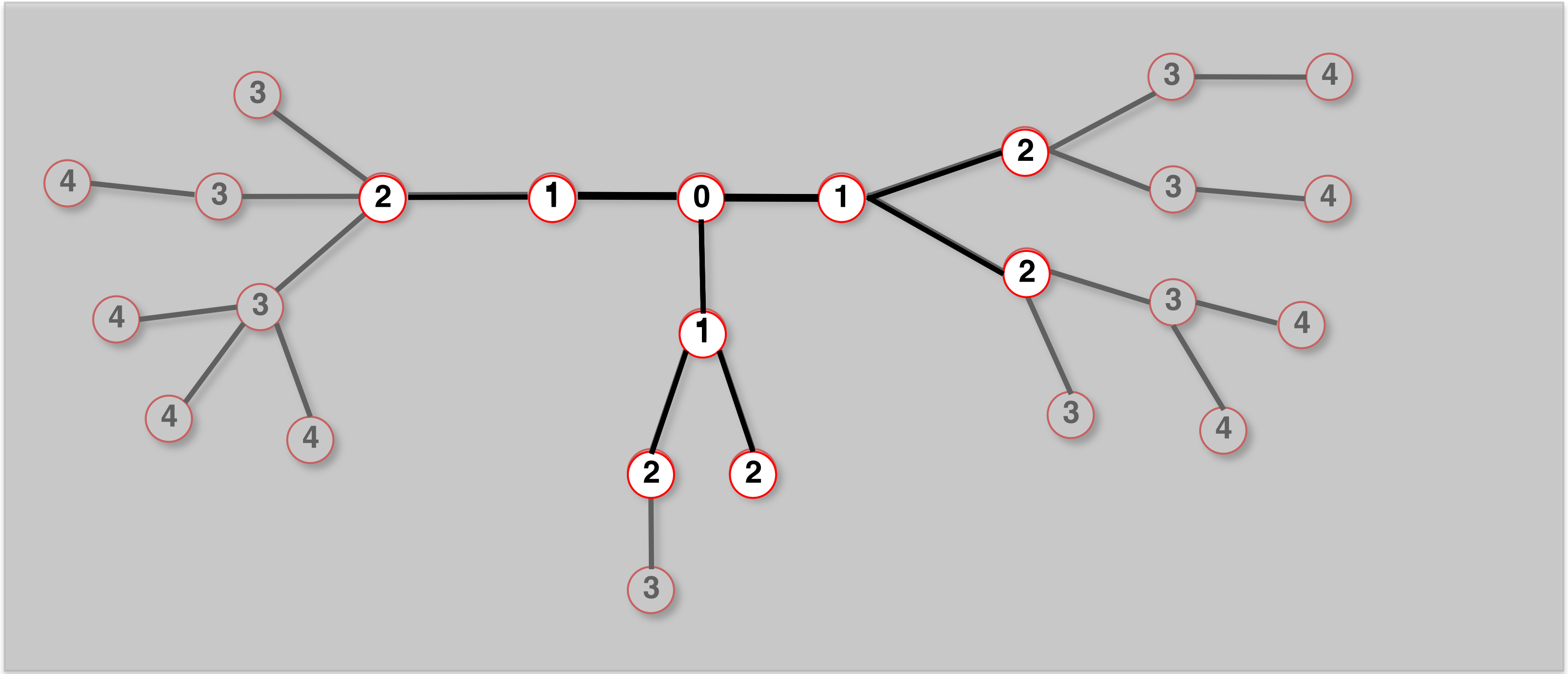
1.Comienza en 0.
2.Encuentra los nodos adyacentes a 0. Márcalos con la etiqueta 1. Ponlos en una fila.



ECONTRANDO DISTANCIAS: “BREADTH FIRST SEARCH”

La distancia entre el nodo 0 y nodo 4:

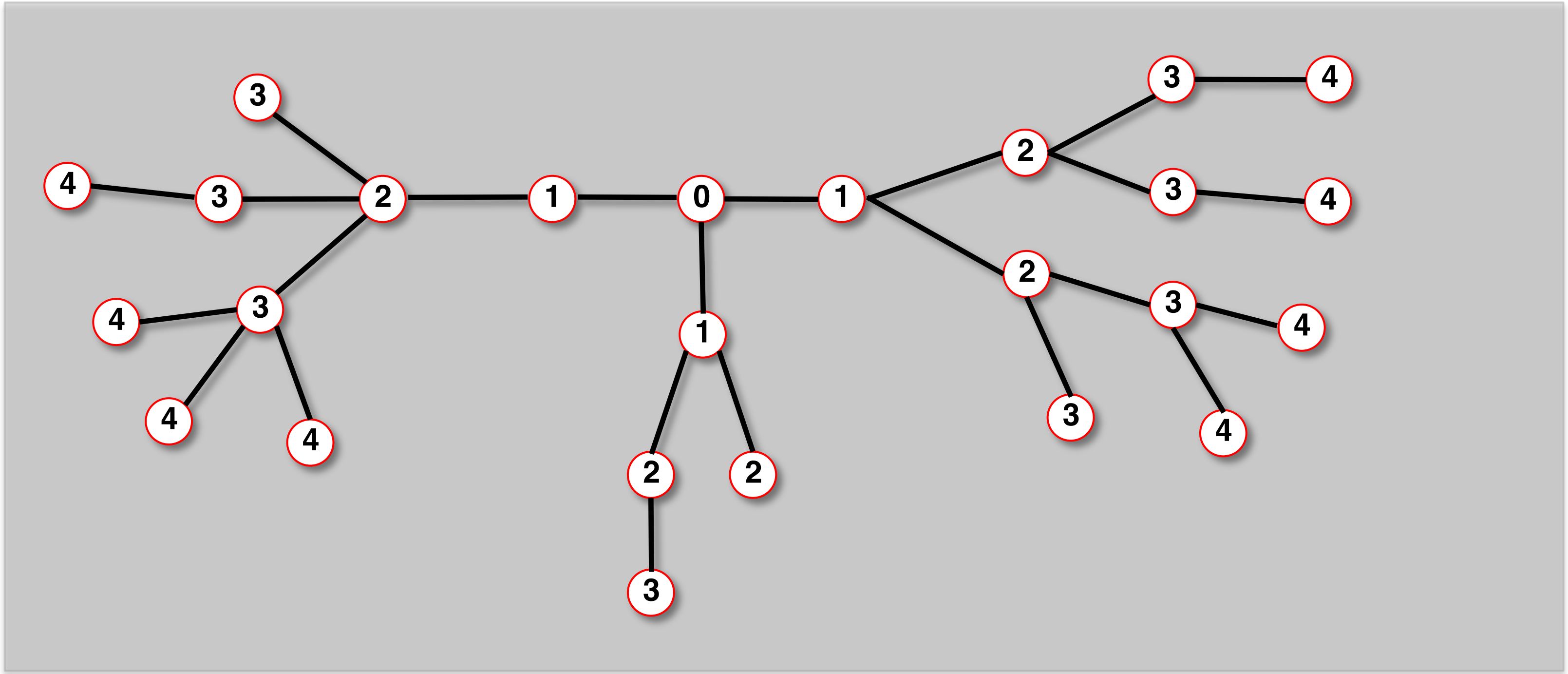
1.Comienza en 0.
2.Encuentra los nodos adyacentes a 1. Márcalos con la etiqueta 1. Pon los en una fila.
3.Toma el primer nodo de la fila. Encuentra los nodos no-marcados adyacentes en el grafo. Márcalos con la etiqueta 2. Ponlos en la fila.



ECONTRANDO DISTANCIAS: “BREADTH FIRST SEARCH”

La distancia entre el nodo 0 y nodo 4:

1. Repite hasta que encuentres el nodo 4 o no hayan mas nodos en la fila.
2. La distancia entre 0 y 4 es la etiqueta 4 o, si 4 no tiene una etiqueta, infinito.



DIÁMETRO DE RED Y DISTANCIA MEDIA

Diámetro: $\boldsymbol{d_{max}}$ La distancia máxima entre cualquier par de nodos en el grafo.

Longitud / distancia media del camino, <d>, para un ***grafo conectado****:*

dónde d_{ij} es la distancia desde el nodo *i* al nodo j

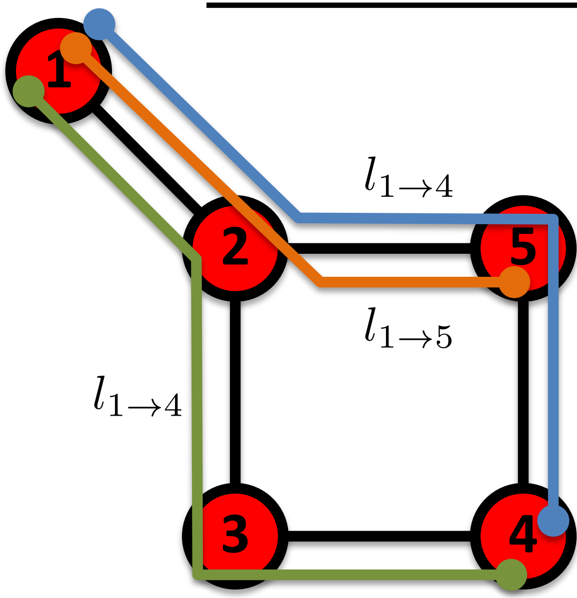
$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} d_{ij}$$

En un grafo no-dirigido $d_{ij} = d_{ji}$*, solo necesitamos contarlos una sola vez:*

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

CAMINOLOGIA: resumen

Camino más corto:



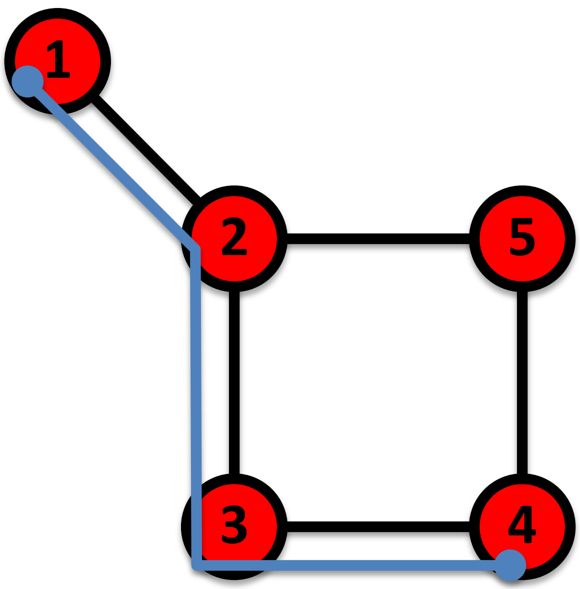
$$l_{1\to 4} = 3$$

$$l_{1\to 5} = 2$$

El camino con la longitud más corta entre dos nodos (distancia).

CAMINOLOGIA: resumen

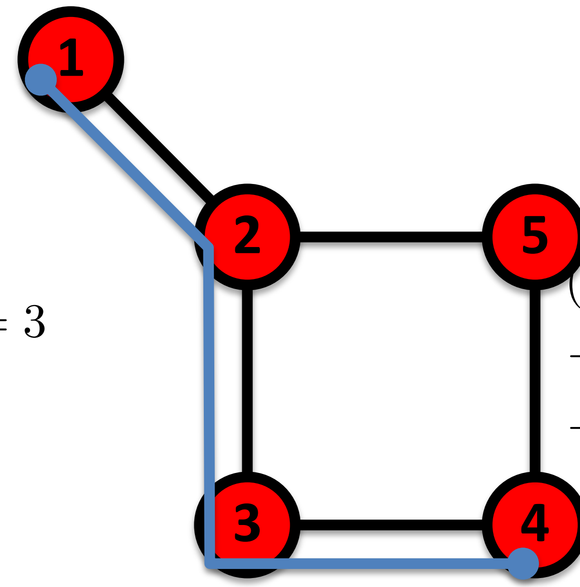
Diámetro



$$l_{1\to4} = 3$$

La distancia más larga del grafo

Longitud de camino promedio

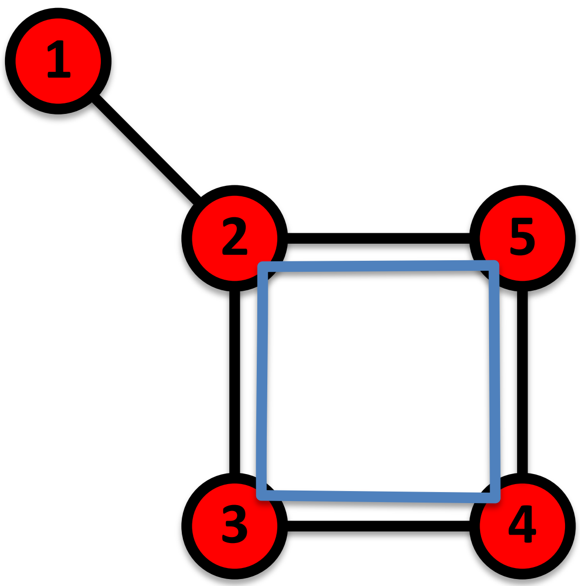


$$\begin{aligned}&(l_{1\to2} + l_{1\to3} + l_{1\to4} +\\ &+ l_{1\to5} + l_{2\to3} + l_{2\to4} +\\ &+ l_{2\to5} + l_{3\to4} + l_{3\to5} +\\ &+ l_{4\to5})/10 = 1.6\end{aligned}$$

El promedio de los caminos más cortos para todos los pares de nodos

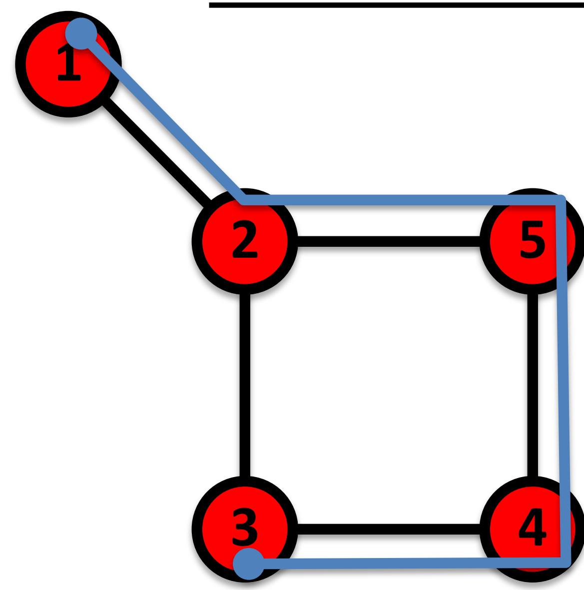
CAMINOLOGIA: resumen

Cíclo



Una ruta con el mismo nodo de inicio y final.

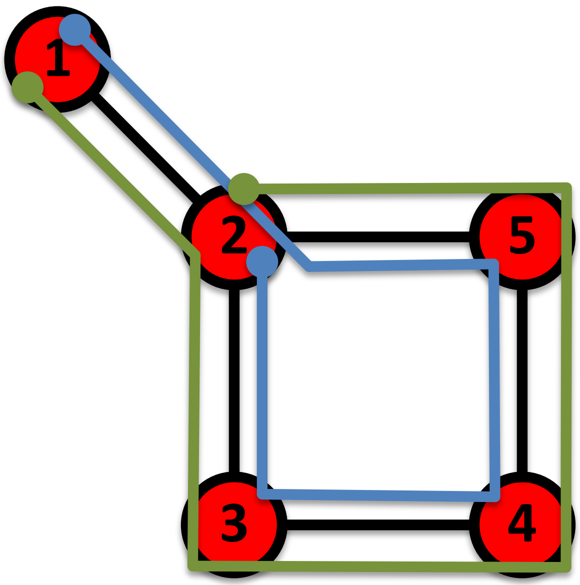
Camino de auto-evasión



Un camino que no se intersecta.

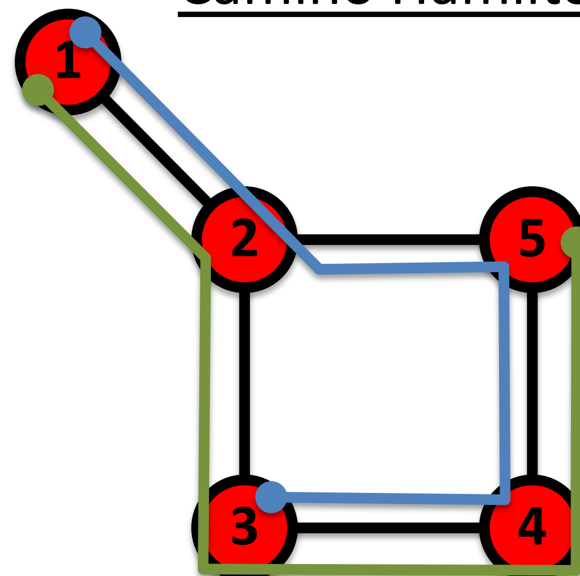
CAMINOLOGIA: resumen

Camino Euleriano



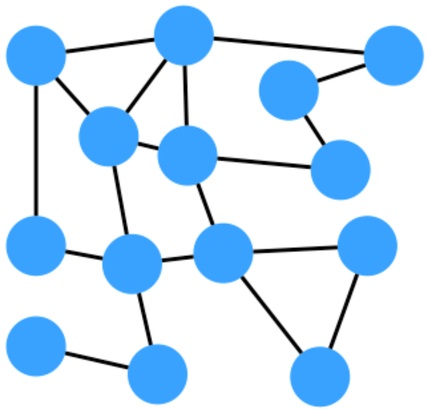
Un camino que atraviesa cada enlace exactamente una vez.

Camino Hamiltoniano



Una ruta que visita cada nodo exactamente una vez.

CAMINOLOGIA: resumen



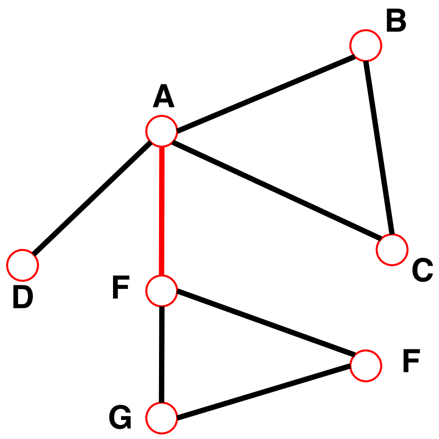
[**Importante**] La red proporciona una métrica natural para el sistema. Es decir, nos permite cuantificar distancias entre elementos.

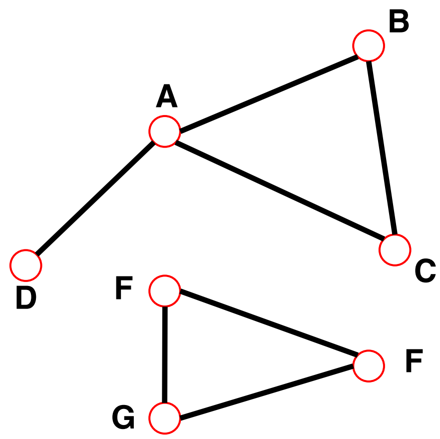
Esto es particularmente importante ya que trataremos con sistemas en los que las distancias / similitudes de cálculo no son naturales y deben realizarse a través de un proxy (la red).

CONECTIVIDAD

CONECTIVIDAD DE GRAFOS NO-DIRIGIDOS

Gráfico conectado (no dirigido): cualquiera de los dos vértices se puede unir por una ruta.
Un gráfico desconectado está compuesto por dos o más componentes conectados.





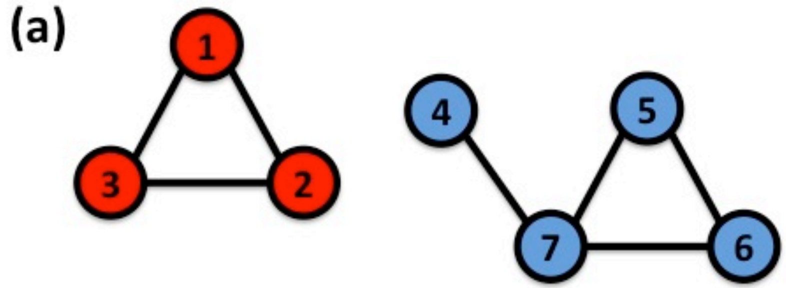
Componente más grande:
Componente gigante

El resto: **Aíslados**

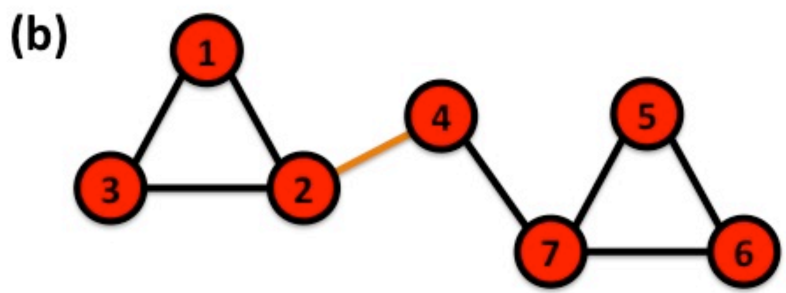
Puente: si lo borramos, la gráfica se desconecta.

CONECTIVIDAD DE GRAFOS NO-DIRIGIDOS. Matriz de adyacencia

La matriz de adyacencia de una red con varios componentes se puede escribir en forma de diagonal de bloques, de modo que los elementos distintos de cero se limiten a los cuadrados, y todos los demás elementos sean cero:



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

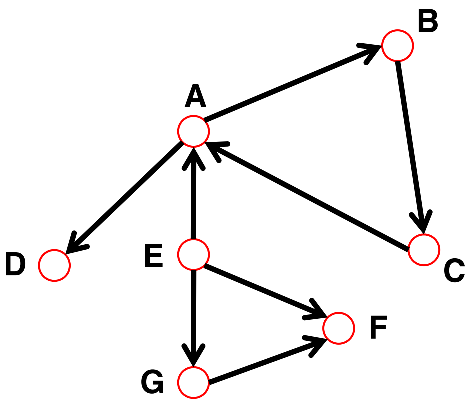


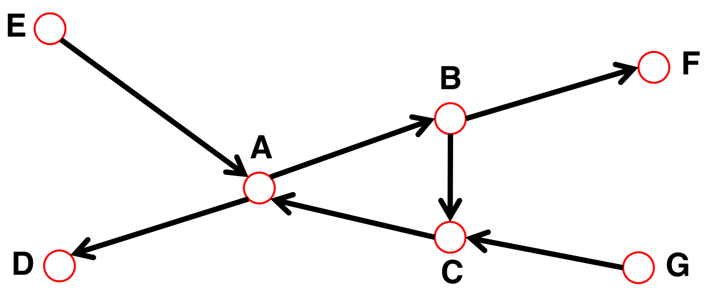
$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

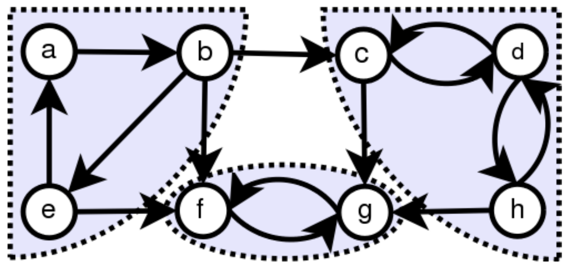
CONECTIVIDAD DE GRAFOS DIRIGIDOS

Grafo dirigido fuertemente conectado: tiene una ruta desde cada nodo a todos los demás nodos y viceversa (por ejemplo, ruta AB y ruta BA).

Grafo dirigido débilmente conectada: está conectada si ignoramos el direcciones de borde.







In-component: nodos que pueden alcanzar el scc (strongly connected component),
Out-component: Nodos a los que se puede acceder desde el scc.

Coeficiente de Clustering

Coeficiente de clustering:

¿Qué fracción de tus vecinos están conectados?

Nodo i con grado k_i

C_i entre [0,1]

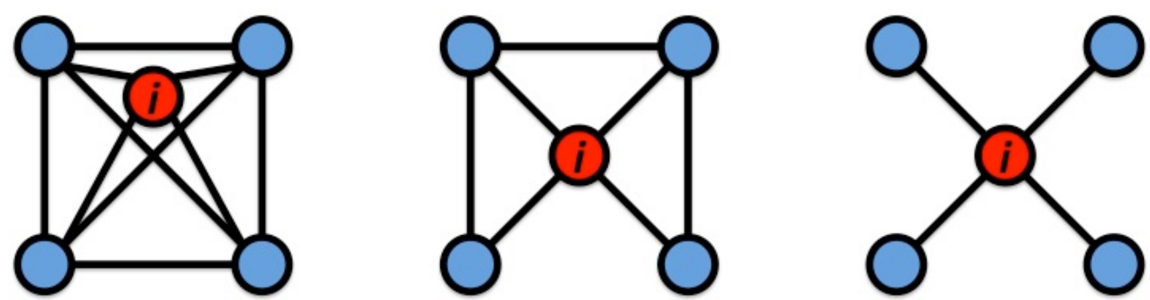
e_i (links entre los vecinos del nodo i)

Dirigido

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E.$$

No-Dirigido

$$C_i = \frac{2|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E.$$



$C_i = 1$ $\quad$ $C_i = 1/2$ $\quad$ $C_i = 0$

2*6/(4*3)=1 $\quad$ 2*3/(4*3)=1/2

Watts & Strogatz, Nature 1998.

Coeficiente de clustering:

¿Qué fracción de tus vecinos están conectados?

Nodo i con grado k_i

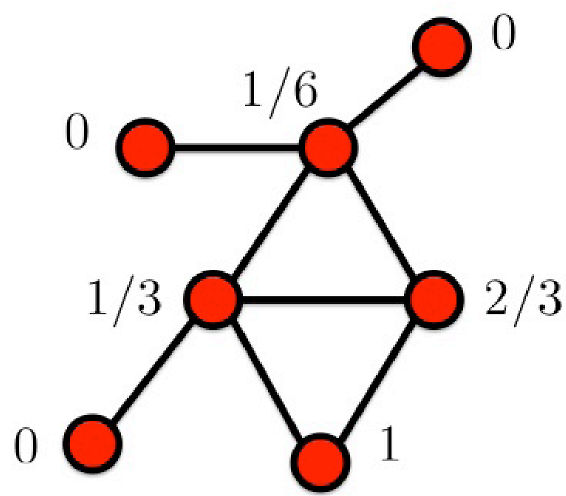
C_i entre [0,1]

e_i (links entre los vecinos del nodo i)

$$C_i = \frac{1}{k_i(k_i-1)} \sum_{j,k} A_{ij} A_{jk} A_{ki}$$

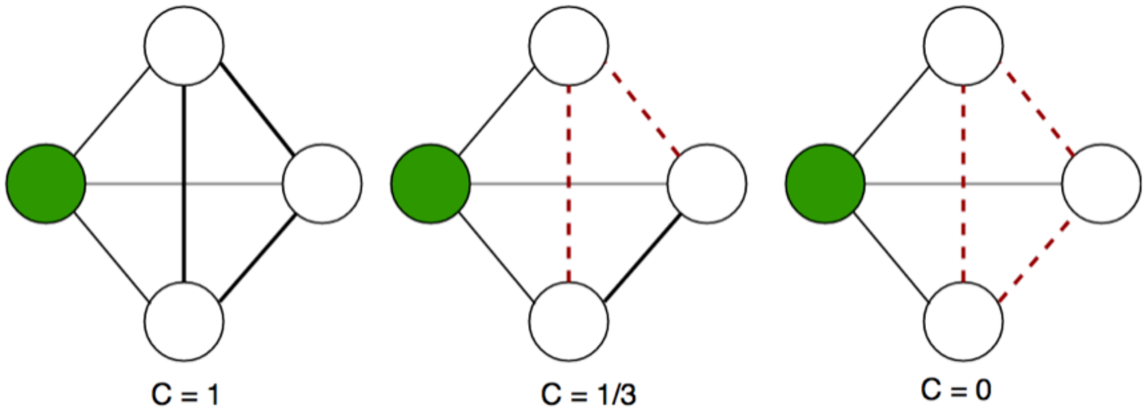
$$k_i = \sum_j A_{ij}$$

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i .$$

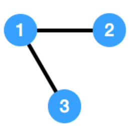


$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

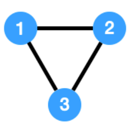
COEFICIENTE DE CLUSTERING



$$C = \frac{\text{number of closed triplets}}{\text{number of all triplets (open and closed)}}.$$



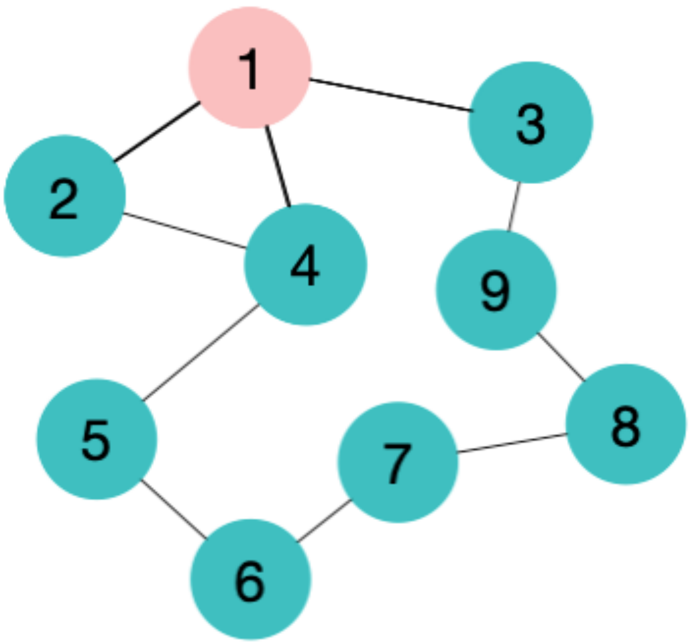
Open Triangle
(triplet)



Closed Triangle
(closed triplet)

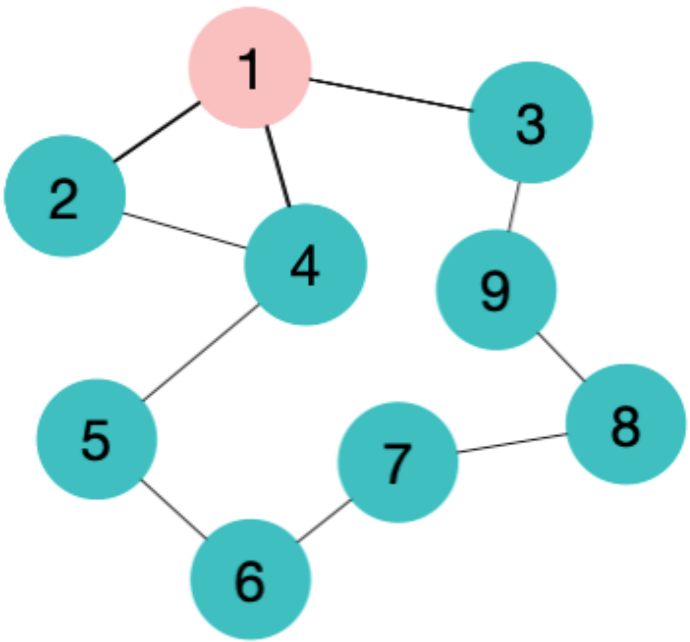
Medición de la densidad de los enlaces
Enlaces Fuertes entre nodos.

COEFICIENTE DE CLUSTERING

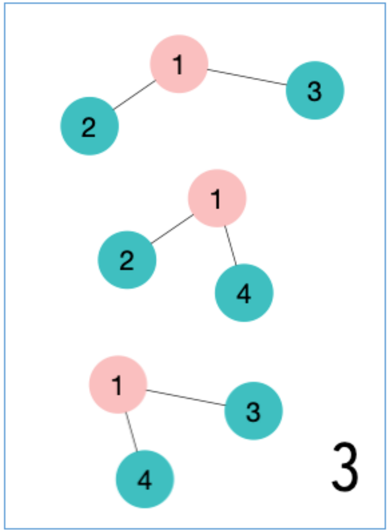


¿Cuál es el CC del nodo 1?

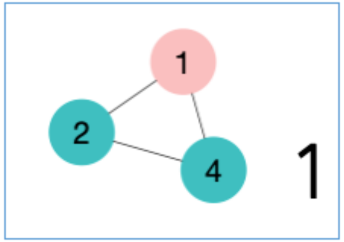
COEFICIENTE DE CLUSTERING



Tripletes



Triángulos Cerrados



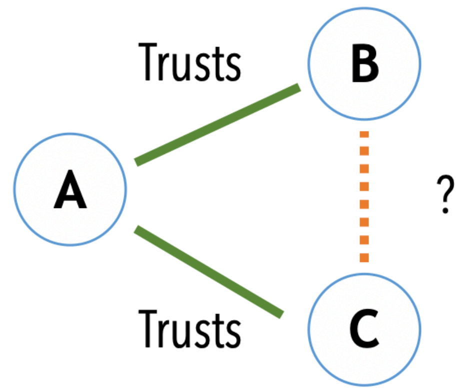
Coeficiente de Clustering:

1/3=0.33

Strength of Weak Ties

Mark Granovetter

Triadic closure in Social Relationships

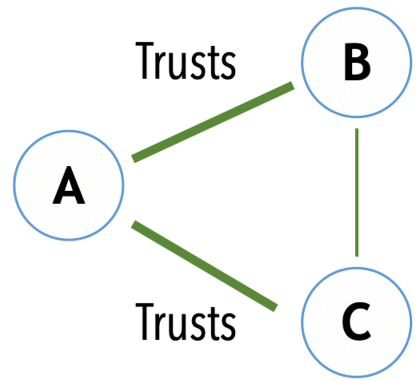


¿Podemos predecir vínculos entre individuos?
La idea de cierre triádico se vincula con la idea de Coeficiente de Clúster.

Strength of Weak Ties

Mark Granovetter

Cierre Triádico en las Relaciones Sociales



B y C deberían tener cierta confianza latente sobre la base de la relación con A

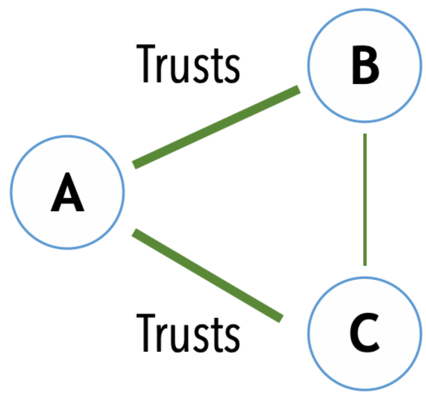
La primera premisa de la teoría es que cuanto más fuerte es el vínculo entre dos personas, más probable es que sus mundos sociales se superpongan.

Las causas subyacentes de la formación de vínculos tienen este tipo de transitividad incorporada.

Strength of Weak Ties

Mark Granovetter

Cierre Triádico en las Relaciones Sociales

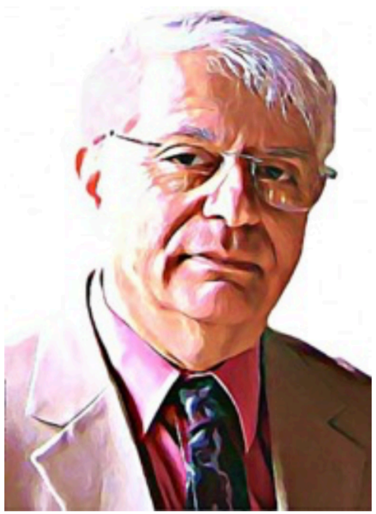


B y C deberían tener cierta confianza latente sobre la base de la relación con A

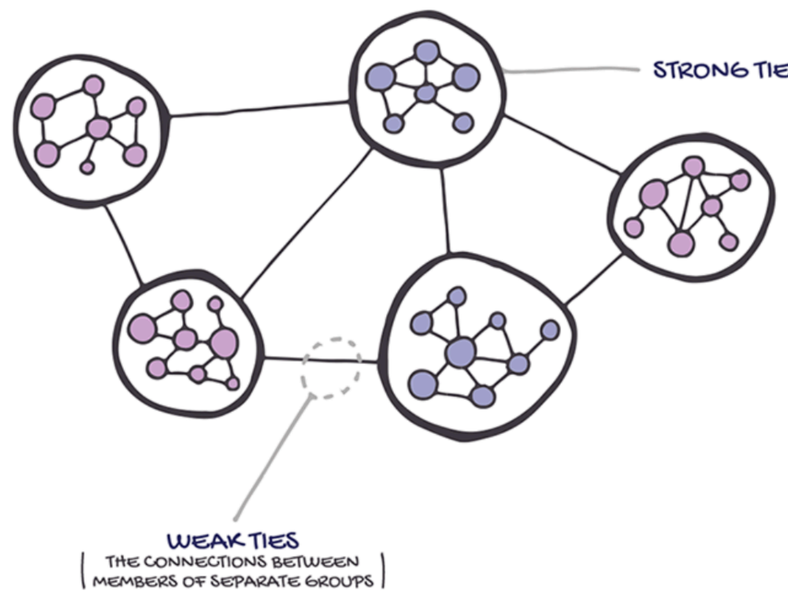
La segunda premisa de SWT es que los lazos de unión son una fuente potencial de ideas novedosas.

A través de un lazo puente, una persona puede escuchar cosas que aún no están circulando entre sus amigos cercanos.

Strength of Weak Ties



Mark Granovetter



Los lazos débiles son los responsables de la difusión de las innovaciones.

Mark Granovetter

The Strength of Weak Ties[1]

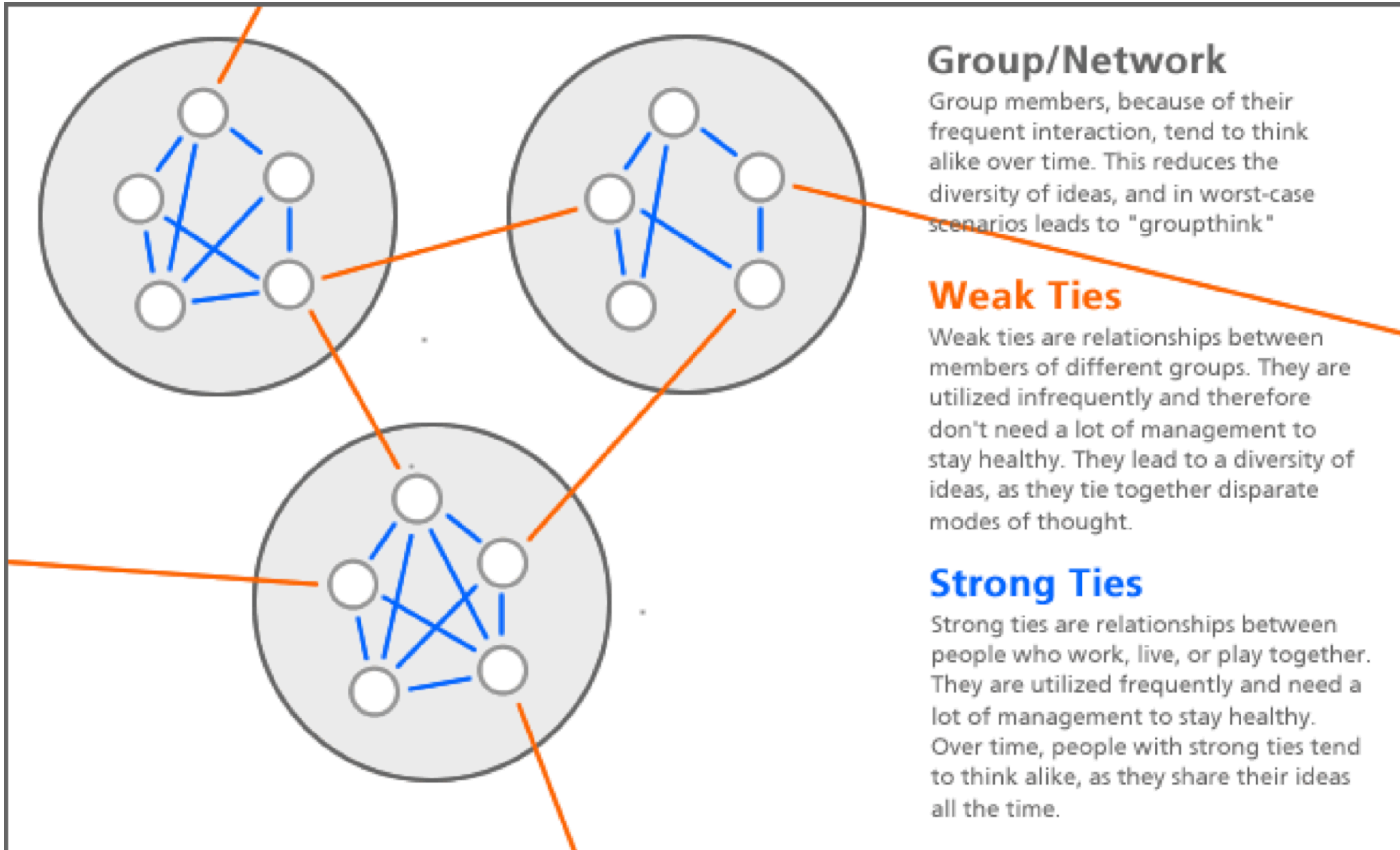
Mark S. Granovetter
Johns Hopkins University

Analysis of social networks is suggested as a tool for linking micro and macro levels of sociological theory. The procedure is illustrated by elaboration of the macro implications of one aspect of small-scale interaction: the strength of dyadic ties. It is argued that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. The impact of this principle on diffusion of influence and information, mobility opportunity, and community organization is explored. Stress is laid on the cohesive power of weak ties. Most network models deal, implicitly, with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups.

Uno de los artículos más citados en Sociología.

Los conceptos expuestos en su obra seminal son fundamentales…

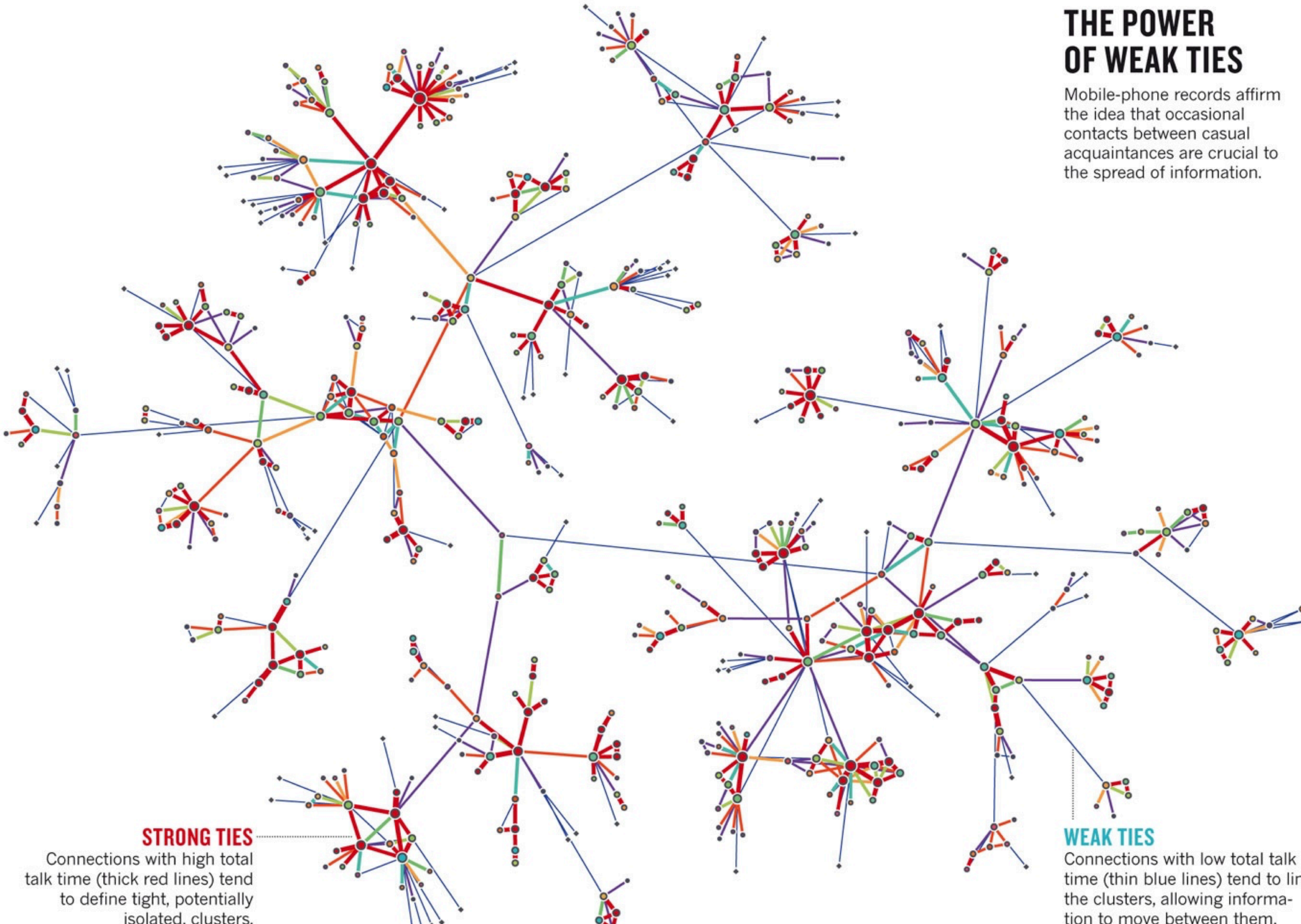
Altamente influyente en marketing, sistemas de información y política, ayudando a los investigadores a ofrecer productos que pueden llegar a audiencias poco probables.



http://bokardo.com/archives/weak-ties-and-diversity-in-social-networks/

THE POWER OF WEAK TIES

Mobile-phone records affirm the idea that occasional contacts between casual acquaintances are crucial to the spread of information.



REPORT | SOCIAL NETWORKS

A causal test of the strength of weak ties

KARTHIK RAJKUMAR, GUILLAUME SAINT-JACQUES, IAVOR BOJINOV, ERIK BRYNJOLFSSON, AND SINAN ARAL Authors Info & Affiliations

SCIENCE • 15 Sep 2022 • Vol 377, Issue 6612 • pp. 1304-1310 • DOI: 10.1126/science.abl4476

RELATED PERSPECTIVE

Weak ties, failed tries, and success

BY DASHUN WANG, BRIAN UZZI

The influence of weak associations

The strength of weak ties is an influential social-scientific theory that stresses the importance of weak associations (e.g., acquaintance versus close friendship) in influencing the transmission of information through social networks. However, causal tests of this paradoxical theory have proved difficult. Rajkumar *et al*. address the question using multiple large-scale, randomized experiments conducted on LinkedIn's "People You May Know" algorithm, which recommends connections to users (see the Perspective by Wang and Uzzi). The experiments showed that weak ties increase job transmissions, but only to a point, after which there are diminishing marginal returns to tie weakness. The authors show that the weakest ties had the greatest impact on job mobility, whereas the strongest ties had the least. Together, these results help to resolve the apparent "paradox of weak ties" and provide evidence of the strength of weak ties theory. —AMS

Resumen y Conceptos Importantes

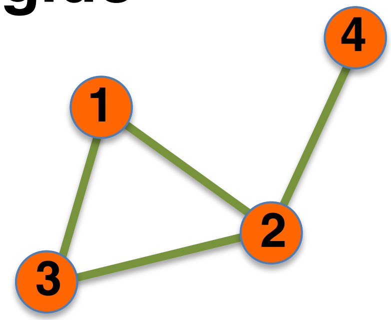
TRES PRINCIPALES CANTIDADES EN NETWORK SCIENCE

Distribución de grado: $P(k)$

Longitud de camino: $\langle d \rangle$

Coeficiente de clustering: $C_i = \dfrac{2e_i}{k_i(k_i - 1)}$

No-dirigido



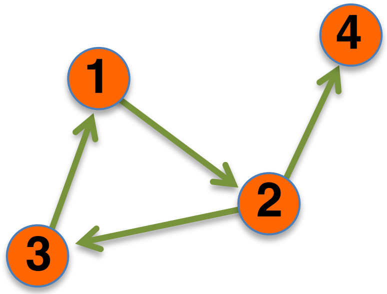
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

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

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad < k >= \frac{2L}{N}$$

Red de actores, interacciones proteina-proteina

Dirigido



In=columnas
Out=filas

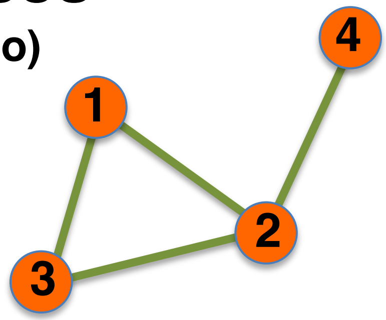
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

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

$$L = \sum_{i,j=1}^{N} A_{ij} \qquad < k >= \frac{L}{N}$$

WWW, red de citaciones

Sin pesos
(no-dirigido)



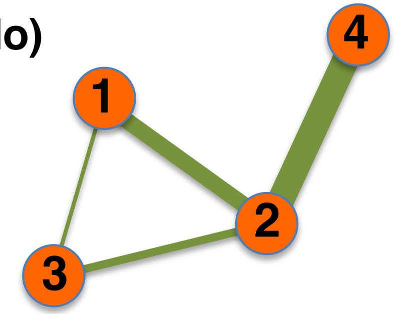
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

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

$$L = \frac{1}{2}\sum_{i,j=1}^{N} A_{ij} \qquad < k >= \frac{2L}{N}$$

Interacciones proteina-proteina, www

Con pesos
(no-dirigido)



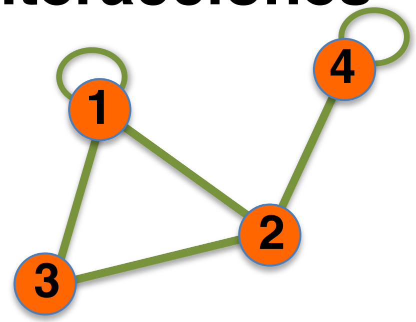
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

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

$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad < k >= \frac{2L}{N}$$

Grafo de llamadas, red metabólica

Auto-interacciones



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

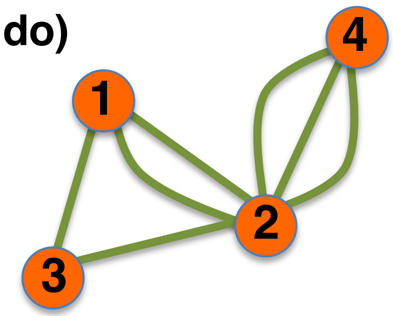
$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

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

Red de interacción de proteinas, www

Multigrafo

(no-dirigido)



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

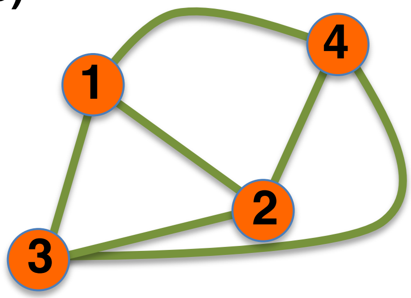
$$A_{ii} = 0 \qquad A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad <k> = \frac{2L}{N}$$

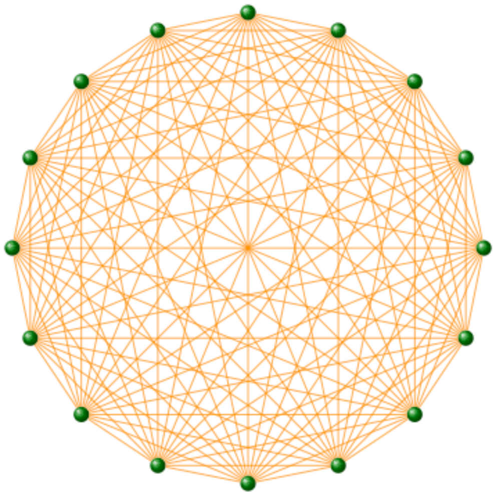
Redes sociales, Red de colaboraciones

Grafo completo

(no-dirigido)



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



$$A_{ii} = 0 \qquad A_{i \neq j} = 1$$

$$L = L_{\max} = \frac{N(N-1)}{2} \qquad < k >= N - 1$$

Red de actores, Interacciones proteina-proteina

GRAPHOLOGY: Las redes reales pueden tener multiples características

WWW > Multigrafo dirigido con auto-interacciones

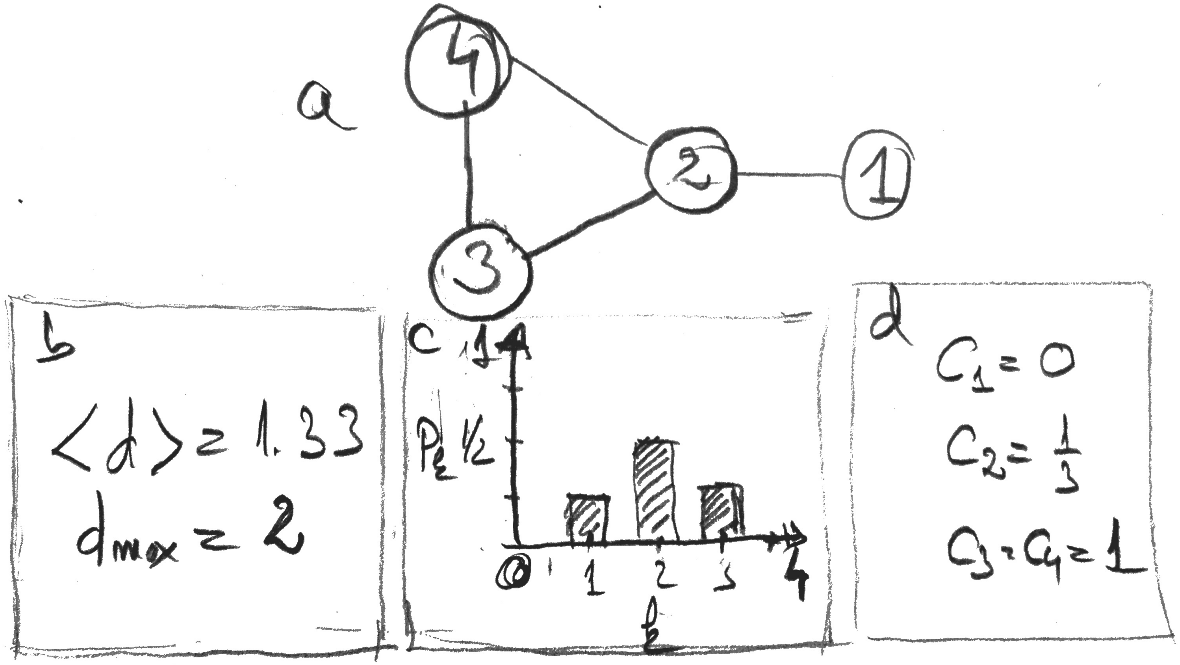
Protein Interactions > no-dirigido, sin pesos, con auto-interacciones

Collaboration network > no-dirigido, multigrafo or con pesos

Mobile phone calls > dirigido, con pesos.

Facebook Friendship links > no-dirigido, sin pesos.

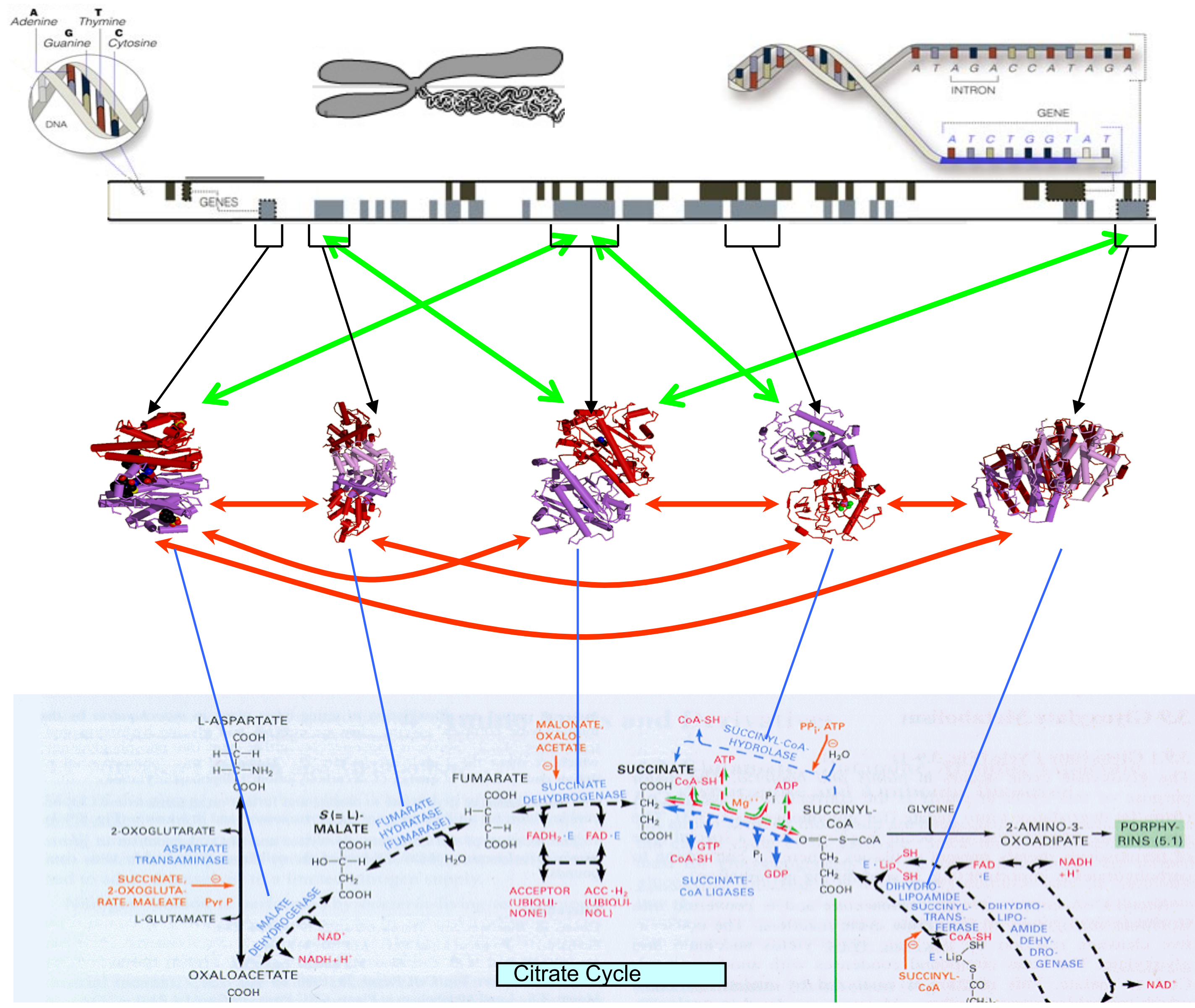
TRES CANTIDADES CENTRALES EN NETWORK SCIENCE



A. Distribución de grado: p_k

B. Longitud de camino: *<d>*

C. Coeficiente de clustering: $C_i = \frac{2e_i}{k_i(k_i - 1)}$



GENOME

Interacciones proteina-genes

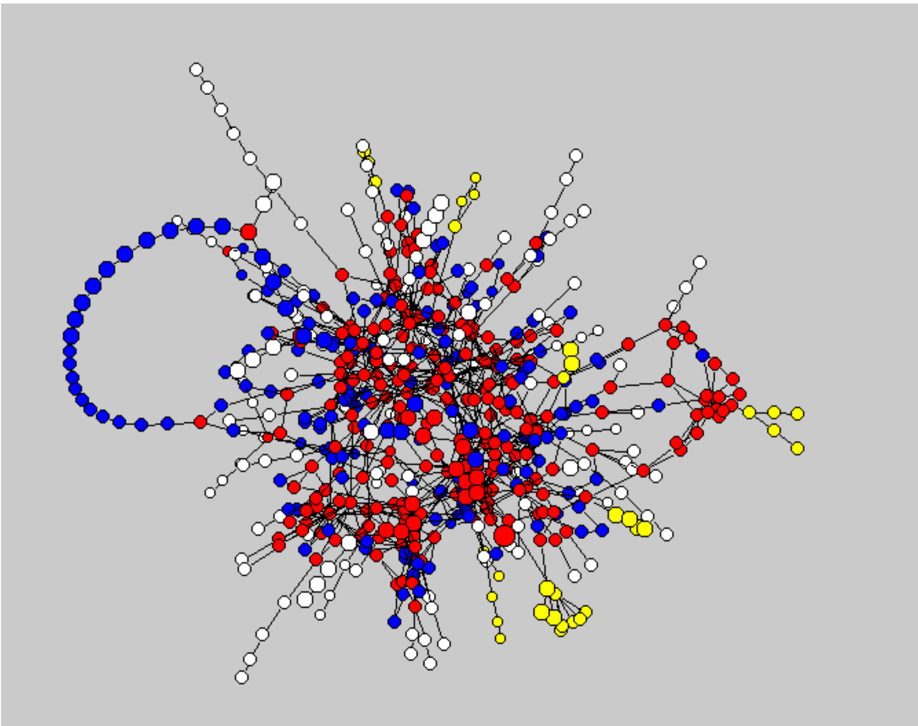
PROTEOME

Interacciones proteina-proteina

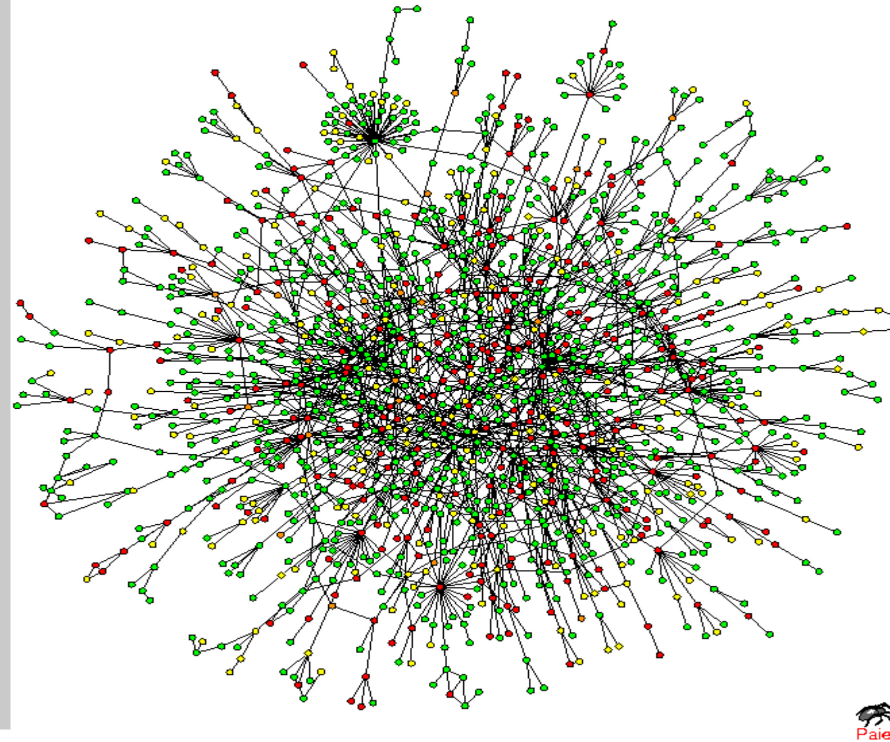
METABOLISM

Reacciones bio-químicas

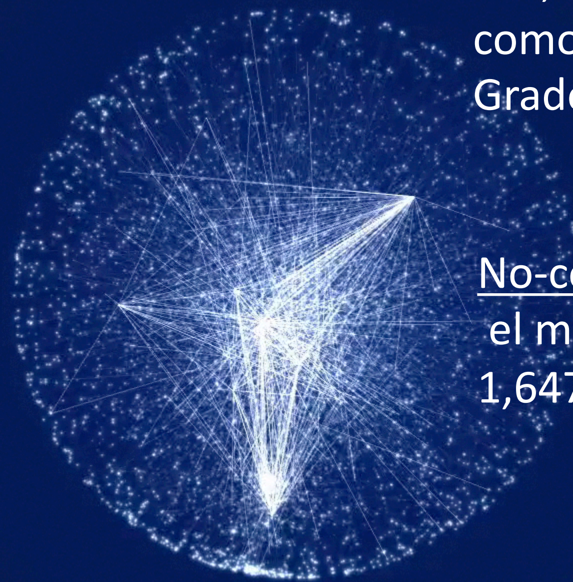
Red metabólica



Interacciones entre Proteinas



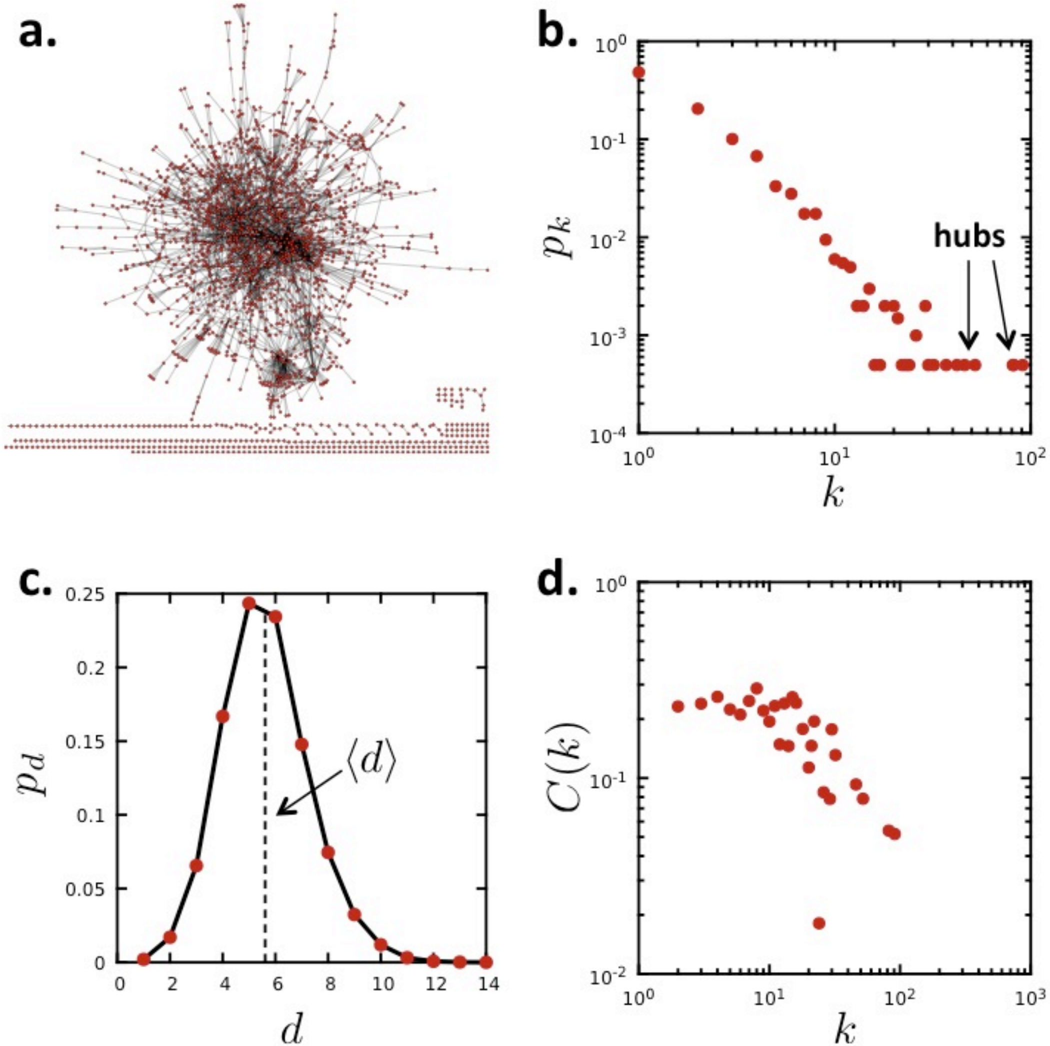
A CASE STUDY: PROTEIN-PROTEIN INTERACTION NETWORK



Red no-dirigida
N=2,018 proteinas como nodos
L=2,930 interacciones vinculantes como links.
Grado promedio <k>=2.90.

No-conectado: 185 componentes
el más grande (componente gigante)
1,647/2018 nodos

UN CASO DE ESTUDIO: INTERACCION PROTEINA-PROTEINA



<u>Red no-dirigida</u>

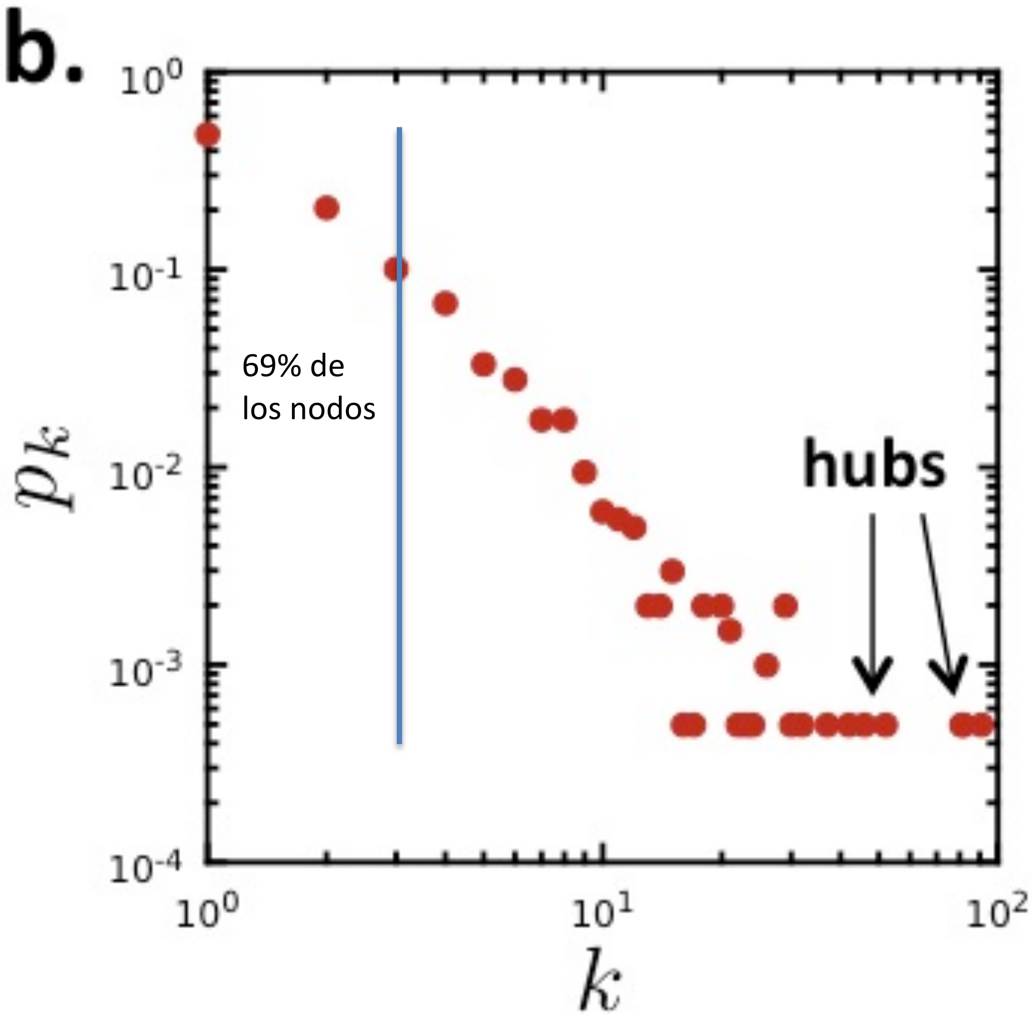
N=2,018 proteínas como nodos

L=2,930 interacciones vinculantes como links.

Grado promedio <k>=2.90.

<u>No conectado: </u>185 componentes las más grande (componente gigante) 1,647 nodos

UN CASO DE ESTUDIO: INTERACCION PROTEINA-PROTEINA



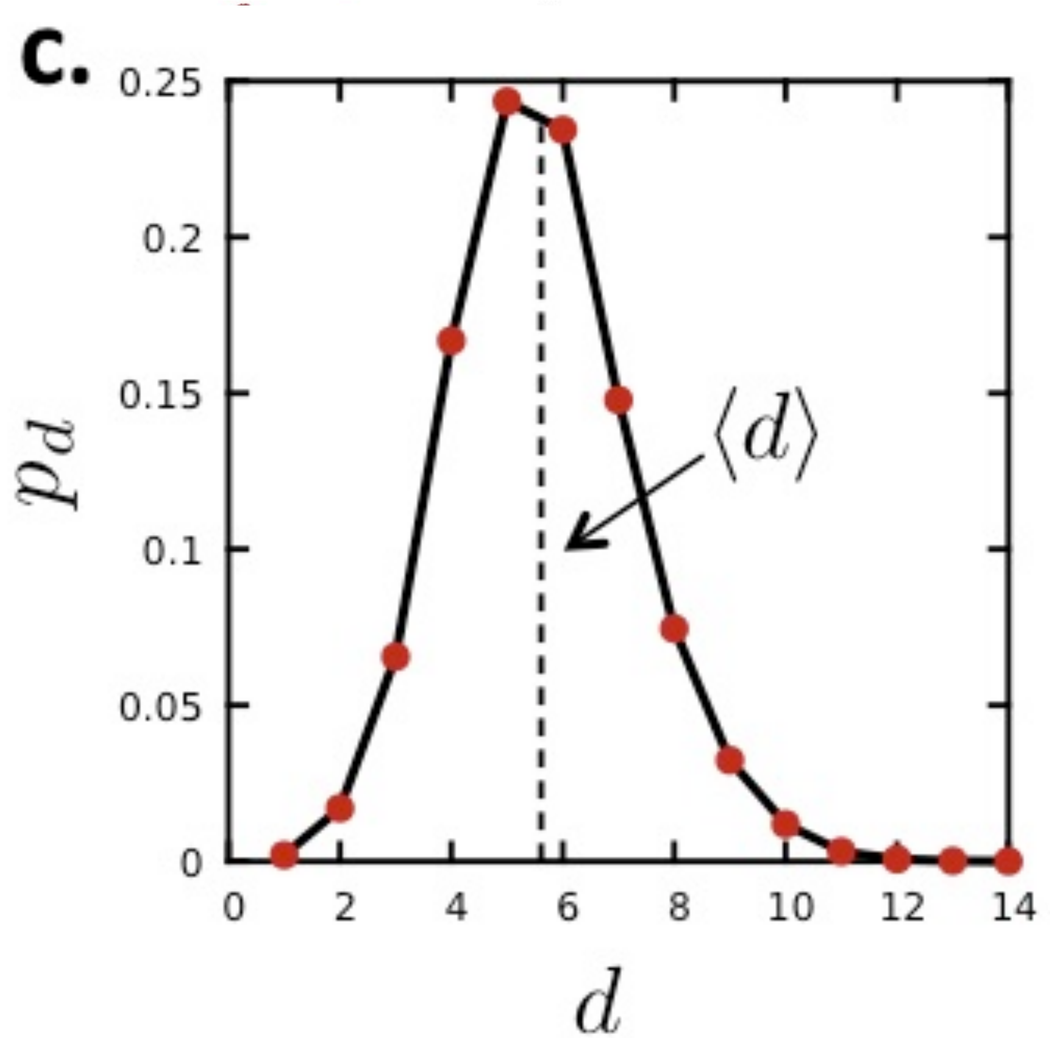
p_k es la probabilidad de que un nodo tenga grado **k**.

N_k = # nodos con grado k

$p_k = N_k / N$

Propiedad libre de escala (scale free)

UN CASO DE ESTUDIO: INTERACCION PROTEINA-PROTEINA

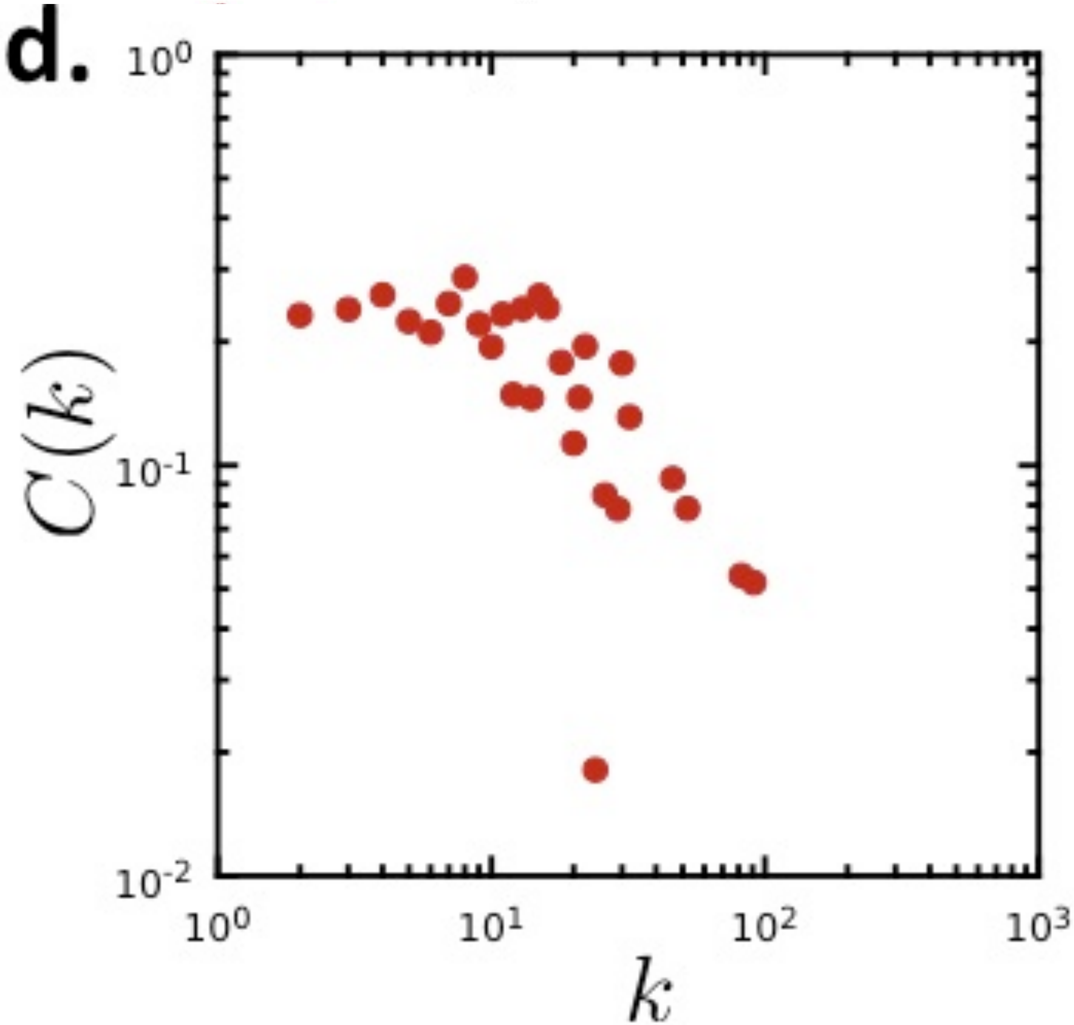


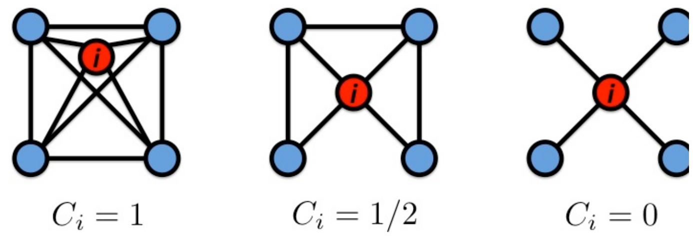
d_{max}=14

<d>=5.61

Propiedad de mundo pequeño

UN CASO DE ESTUDIO: INTERACCION PROTEINA-PROTEINA





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

<C>=0.12

Propiedad de jerarquía