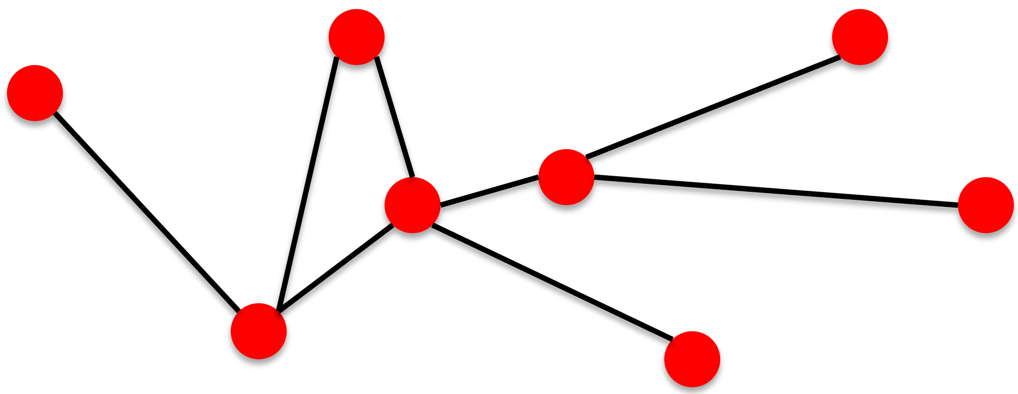
Ciencia de Redes (Humanas y Sociales) #2

Carlos Sarraute

Instituto de Cálculo, Abril-Junio 2019

Conceptos fundamentales de Teoría de Grafos

COMPONENTES DE UN SISTEMA COMPLEJO



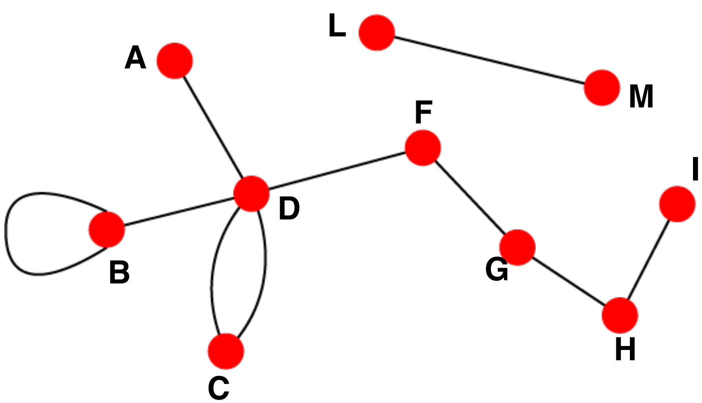
- **componentes**: nodos, vertices — N
- **interacciones**: vínculos, enlaces, aristas — L
- **sistema**: red, grafo — (N,L)

REDES DIRIGIDAS VS. NO DIRIGIDAS

No dirigido

Enlaces: no dirigidos (simétricos)

Grafo:

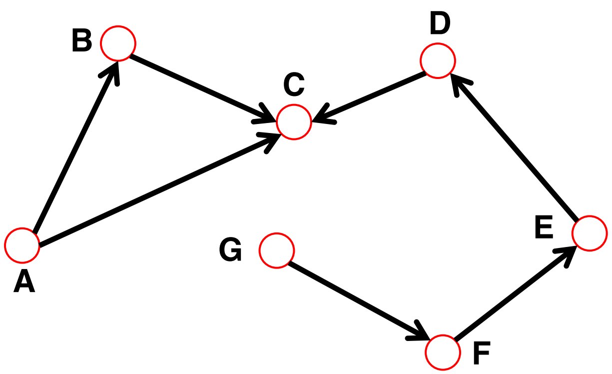


Enlaces no dirigidos:
Vínculo de coautor
Red de actors
Interacciones entre proteinas

Dirigido

Links: directed (*arcs*).

Digrafo = directed graph:

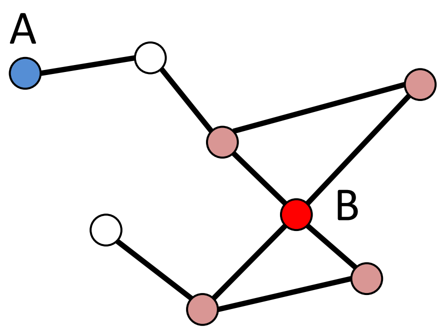


Enlaces dirigidos :
URLs en la web
Llamados telefónicos
Reacciones metabolicas

Distribución de grados

GRADO DE UN NODO

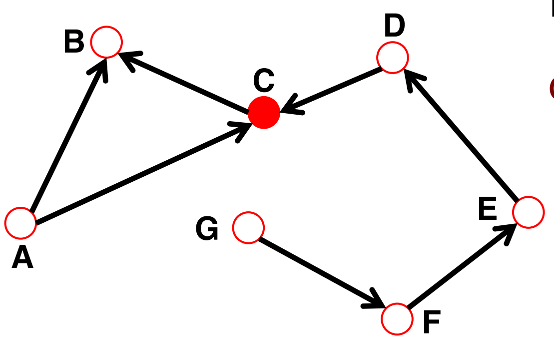
No dirigido



Grado del nodo: cantidad de enlaces que conectan con el nodo

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

Dirigido



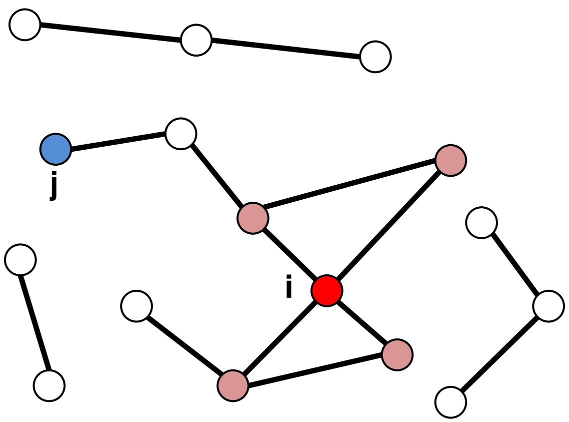
En los *grafos dirigidos* se puede definir un in-degree y out-degree. El grado (total) es la suma de in- y out-degree.

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

Source: nodo con $k^{in} = 0$; Sink: nodo con $k^{out} = 0$.

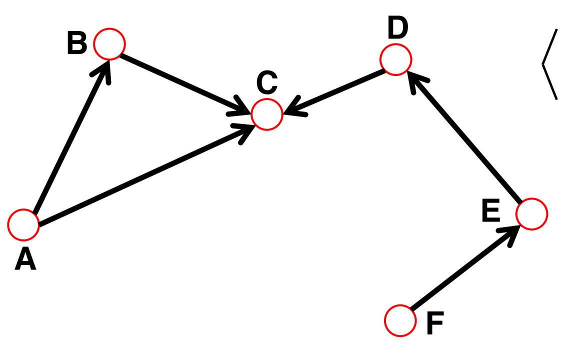
GRADO PROMEDIO

No dirigido



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

Dirigido



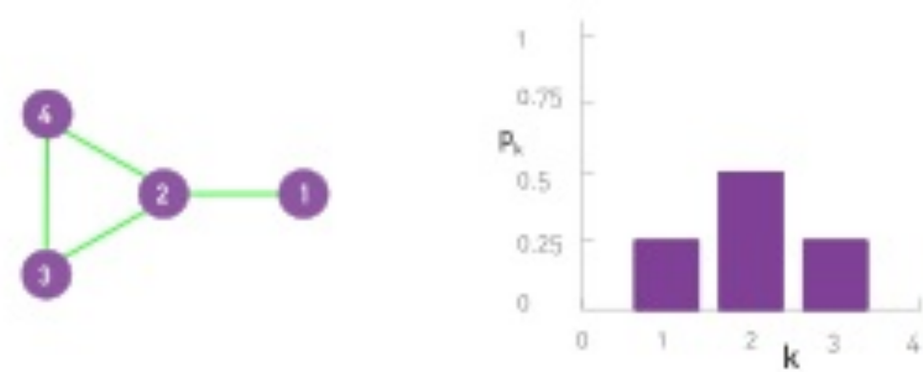
$$\langle k^{in} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

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

DISTRIBUCIÓN DE GRADOS

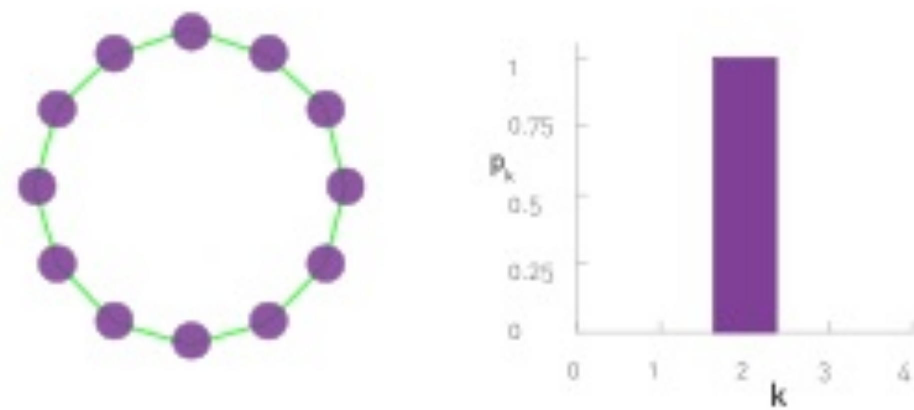
Distribución de grados

P(k): probabilidad de que un nodo al azar tenga grado *k*



N_k = # nodos con grado k

$P(k) = N_k / N$ ➔ plot



DISTRIBUCIÓN DE GRADOS

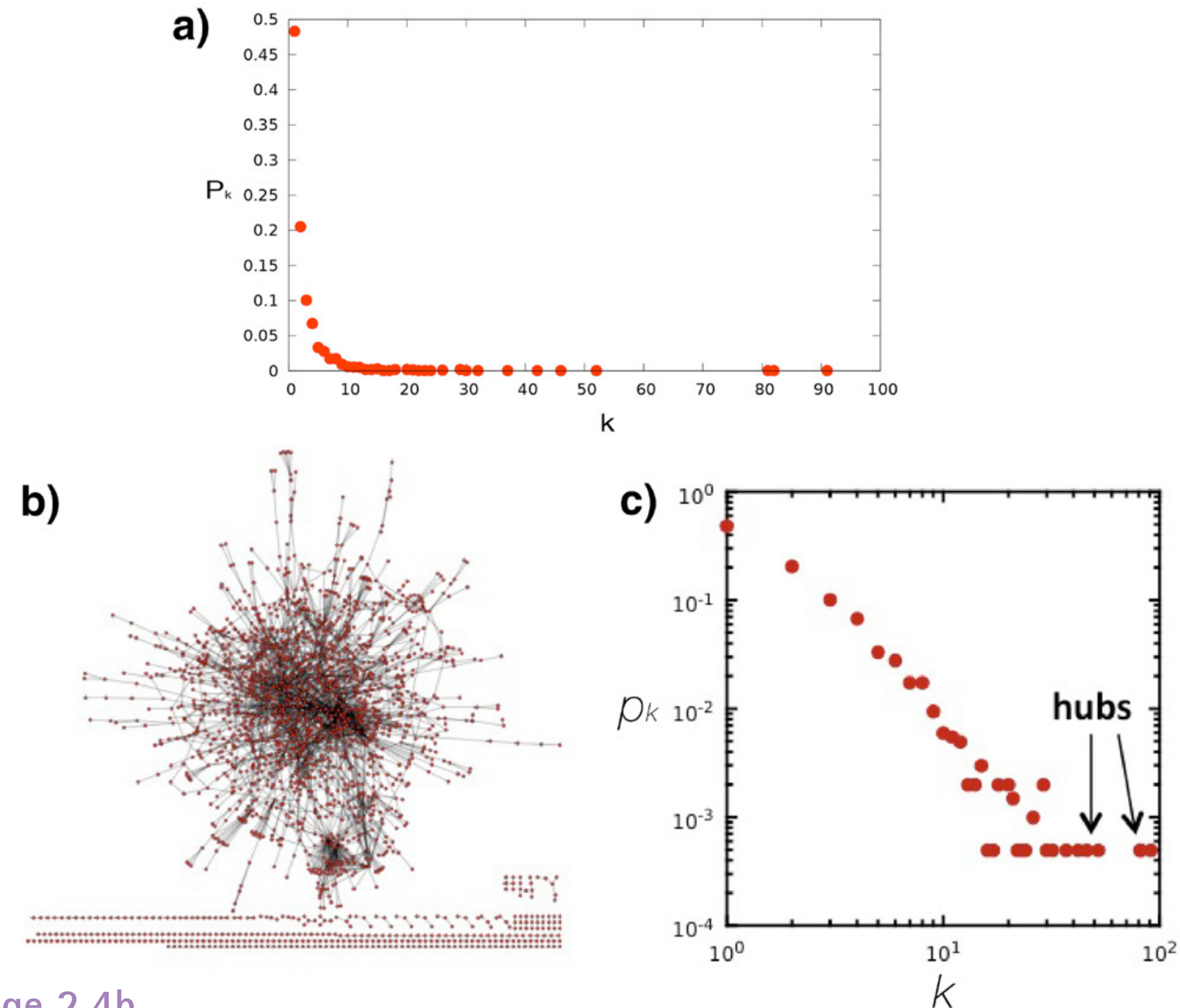
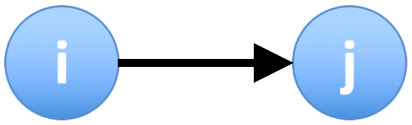


Image 2.4b

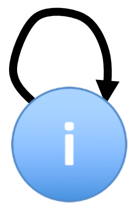
Matriz de adyacencia

Matriz de adyacencia

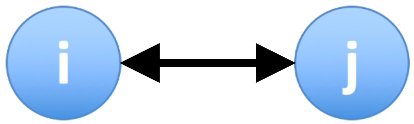
- Representa enlaces como matriz
  - $A_{ij} = 1$ si nodo i tiene enlace hacia nodo j
    $= 0$ sino



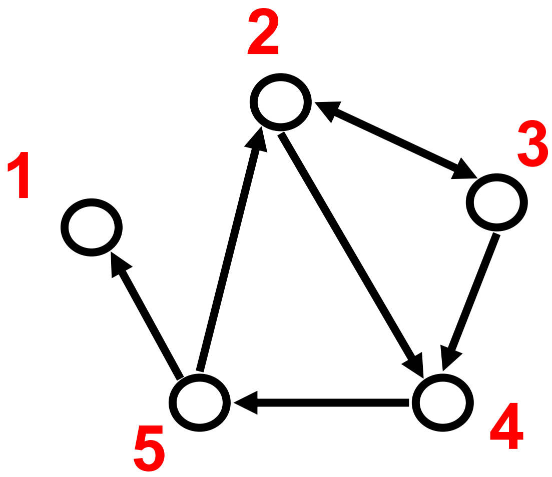
  - $A_{ii} = 0$ salvo que el grafo tenga "self-loops"



  - $A_{ij} = A_{ji}$ si el grafo es no dirigido,
    o si i y j tiene un enlace recíproco

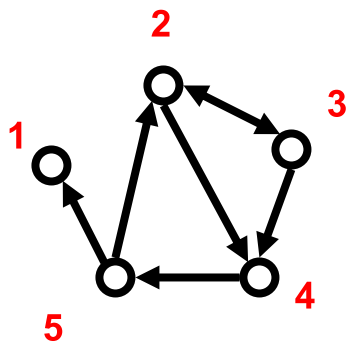


Ejemplo de matriz de adyacencia



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

Grados de nodos usando matriz



Outdegree = $\sum_{j=1}^{n} A_{ij}$

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ \boxed{0 \quad 1 \quad 0 \quad 1 \quad 0} \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

ejemplo: outdegree para nodo 3
sumamos la 3er *fila*

$$\sum_{j=1}^{n} A_{3j}$$

Indegree = $\sum_{i=1}^{n} A_{ij}$

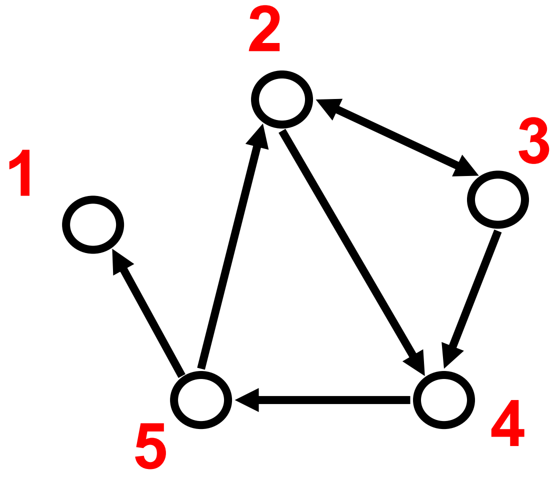
$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

ejemplo: indegree para nodo 3
Sumamos la 3er *columna*

$$\sum_{i=1}^{n} A_{i3}$$

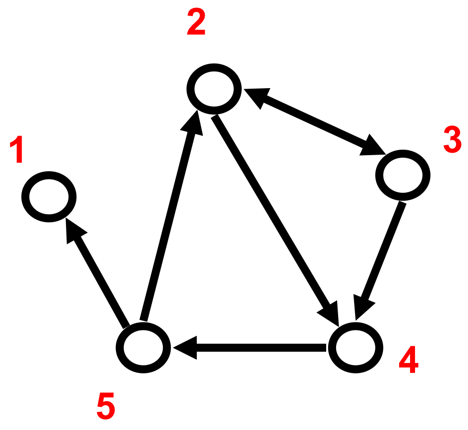
Lista de aristas

- Lista de aristas
  - 2, 3
  - 2, 4
  - 3, 2
  - 3, 4
  - 4, 5
  - 5, 2
  - 5, 1

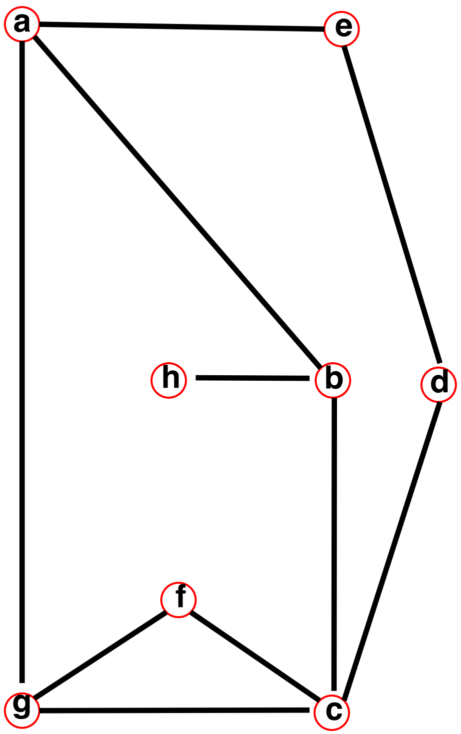


Lista de adyacencia

- Lista de adyacencia
  - Mas facil de usar para redes
    - grandes
    - ralas (sparse)
  - Recuperar facilmente los vecinos de un nodo
    - 1:
    - 2: 3 4
    - 3: 2 4
    - 4: 5
    - 5: 1 2



Ejemplo de matriz de adyacencia



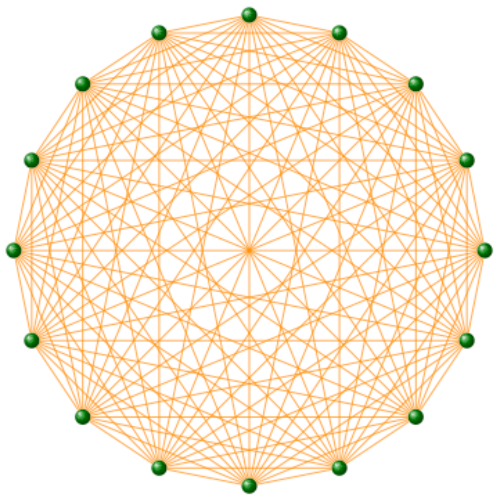
	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	1	0	0
h	0	1	0	0	0	0	0	0

Las redes reales son ralas (sparse)

Grafo completo

La cantidad máxima de vínculos en una red con N nodos:

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



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

LAS REDES REALES SON RALAS

La mayoría de las redes observadas en sistemas reales son ralas (sparse):

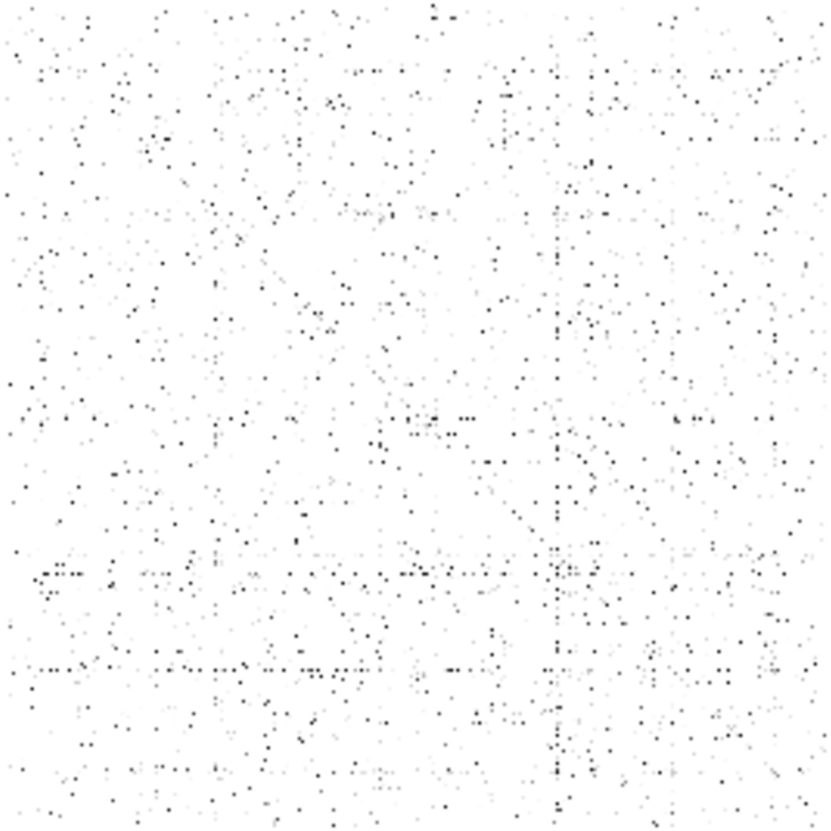
$$L \ll L_{max}$$

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

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)

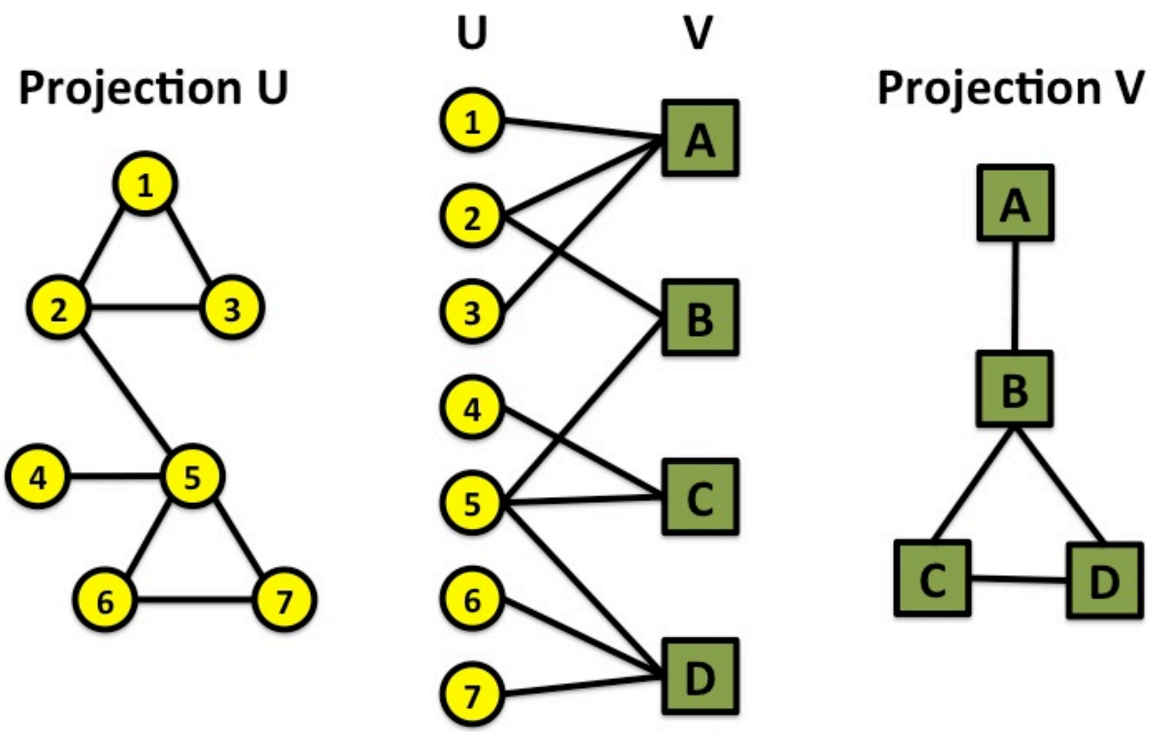
MATRICES DE ADYACENCIA SON RALAS



REDES BIPARTITAS

GRAFO BIPARTITO

grafo bipartito es un grafo 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**.

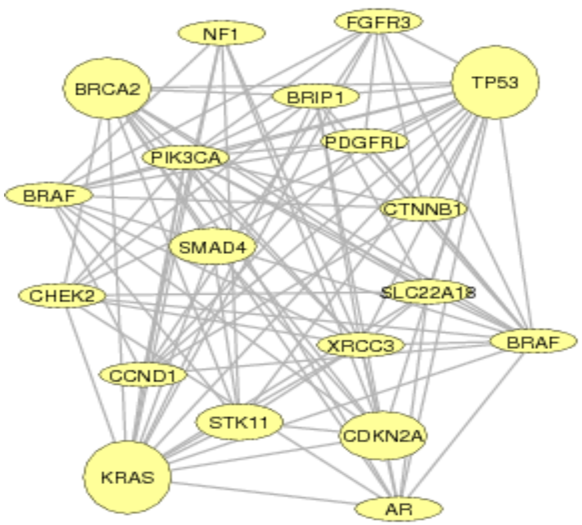


Ejemplos:

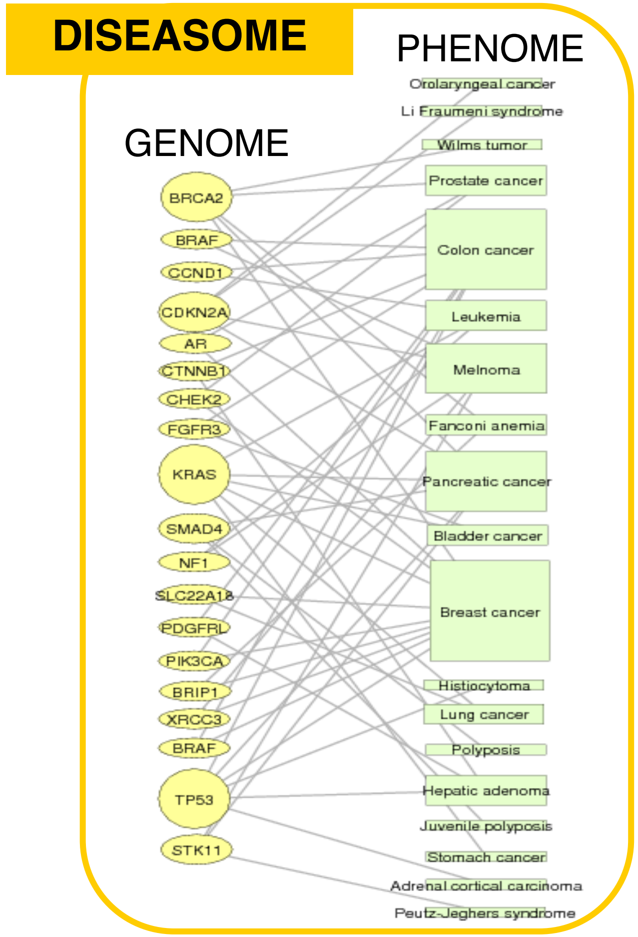
Red de actores del cine argentino

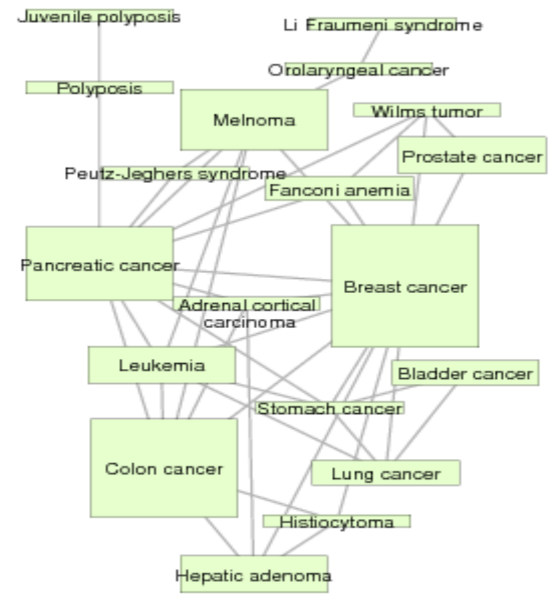
Red de enfermedades

RED DE GENES Y RED DE ENFERMEDADES



Gene network

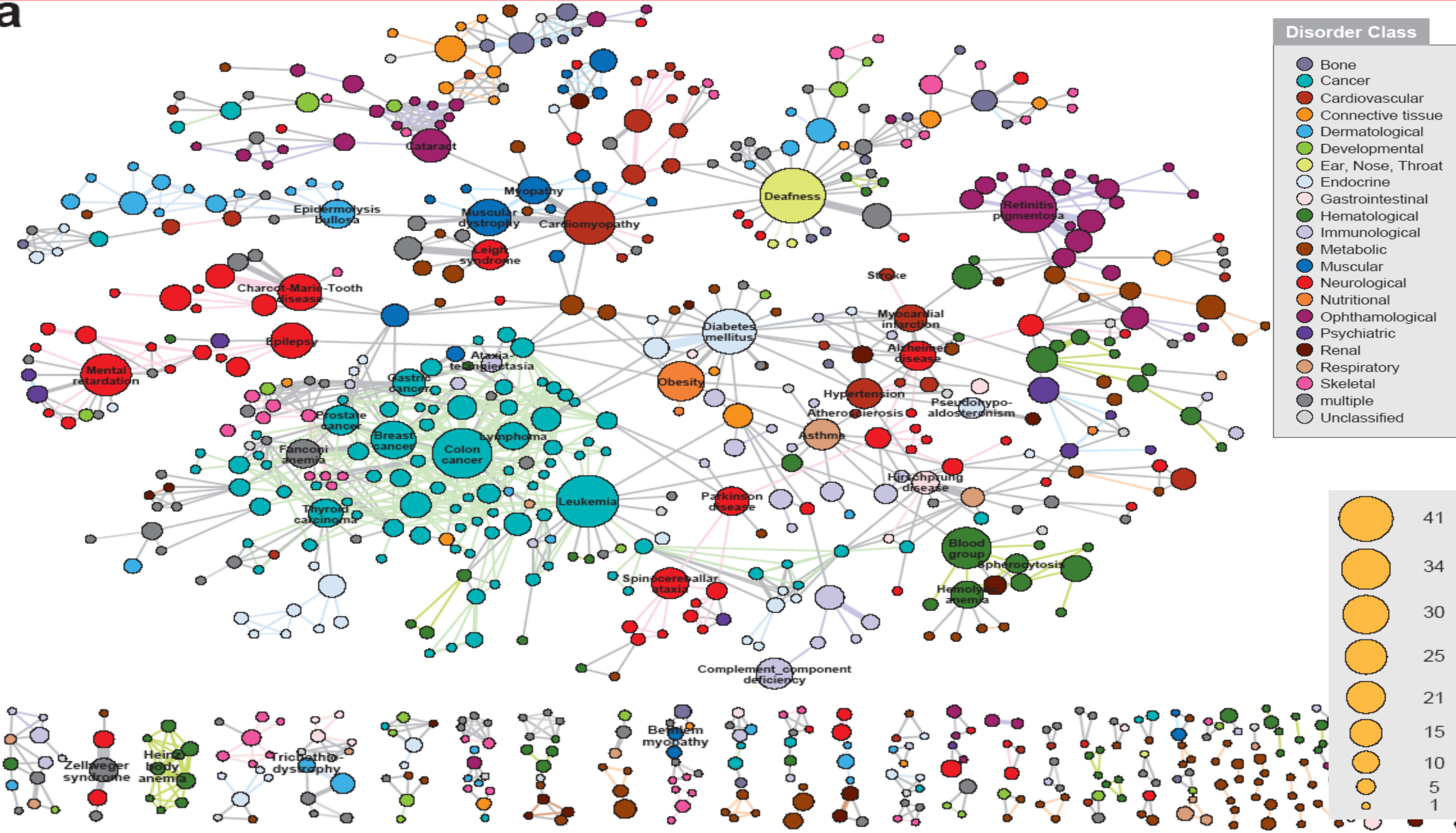




Disease network

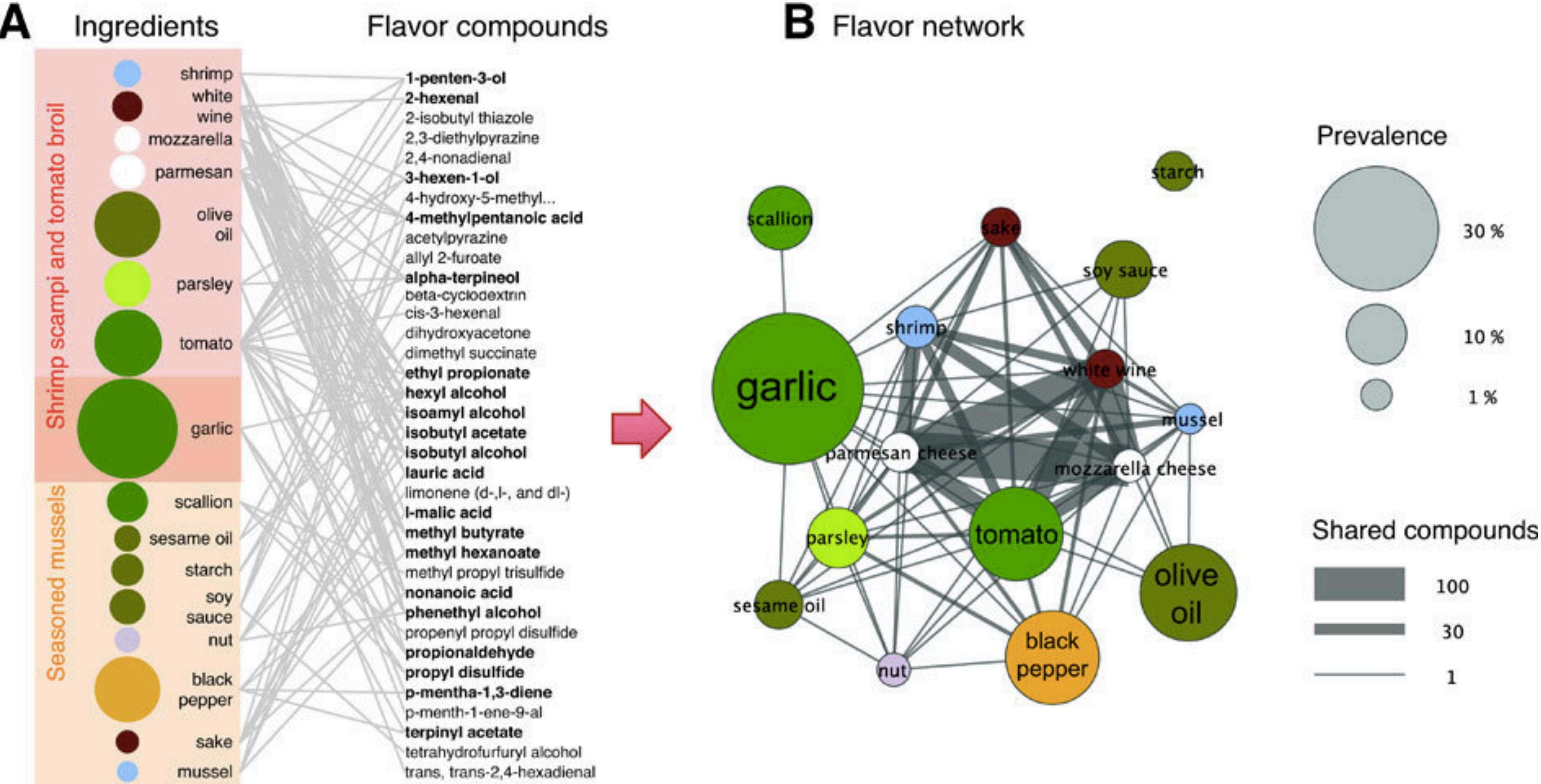
Goh, Cusick, Valle, Childs, Vidal & Barabási, PNAS (2007)

RED DE ENFERMEDADES HUMANAS

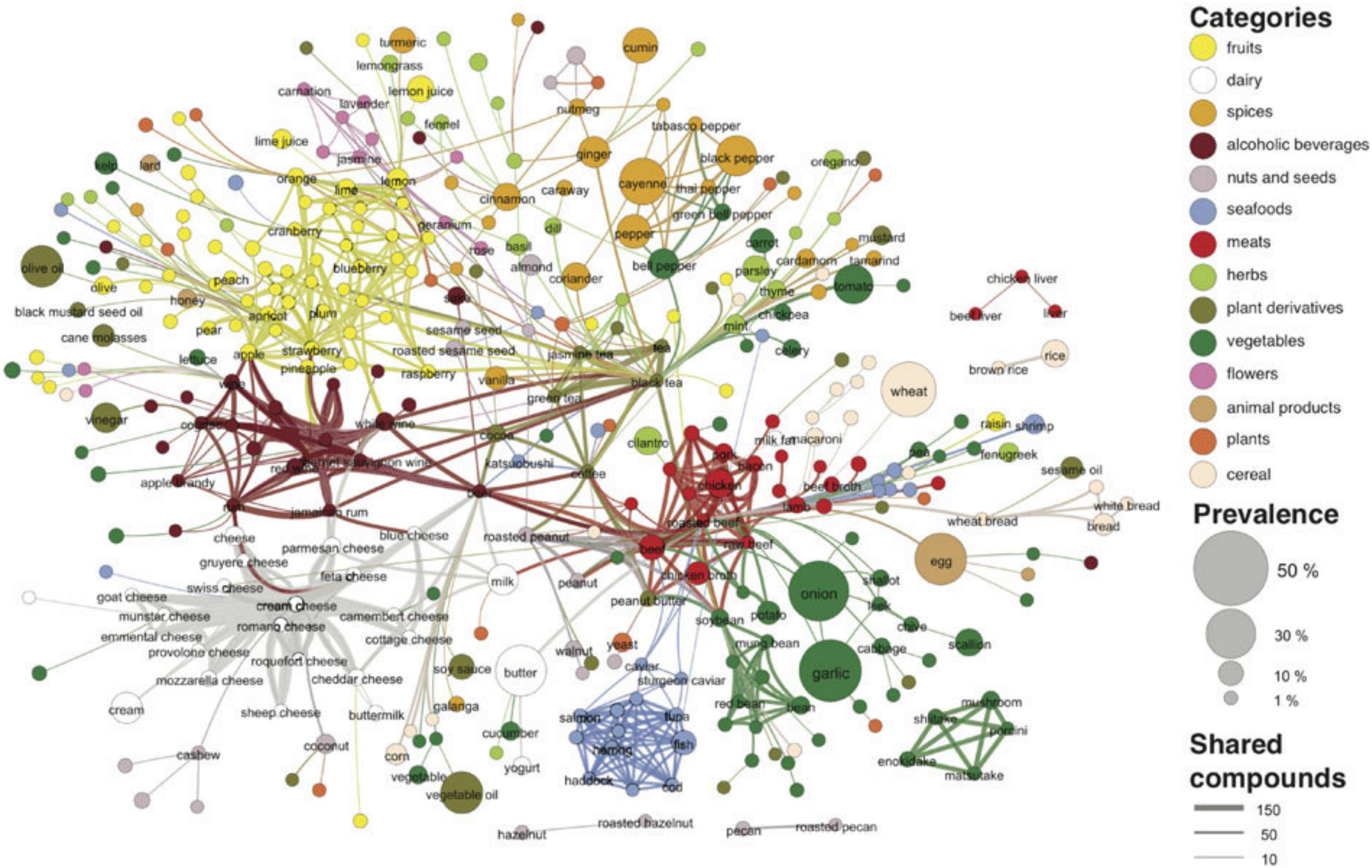


https://archive.nytimes.com/www.nytimes.com/interactive/2008/05/05/science/20080506_DISEASE.html?ref=health

RED BIPARTITA DE INGREDIENTES Y SABORES

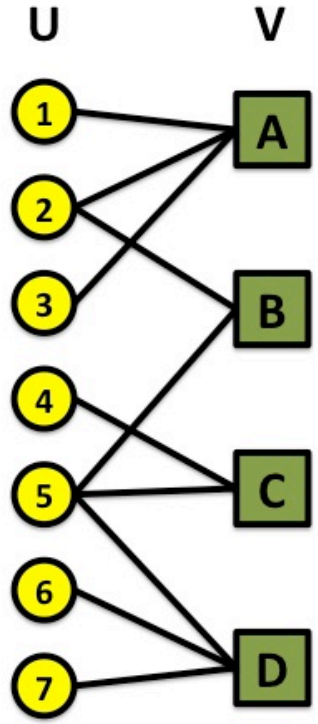


Y.-Y. Ahn, S. E. Ahnert, J. P. Bagrow, A.-L. Barabási Flavor network and the principles of food pairing , Scientific Reports 196, (2011).



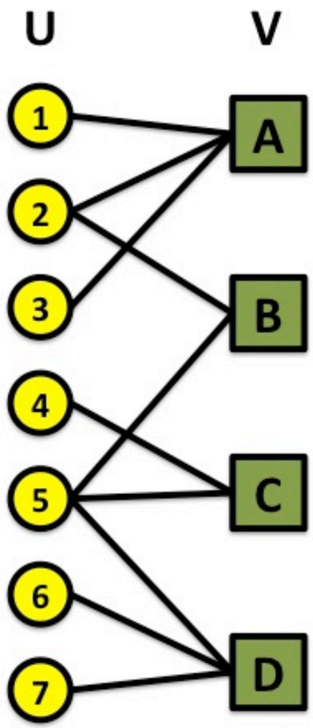
Categories
fruits
dairy
spices
alcoholic beverages
nuts and seeds
seafoods
meats
herbs
plant derivatives
vegetables
flowers
animal products
plants
cereal
Prevalence
50 %
30 %
10 %
1 %
Shared compounds
150
50
10
turmeric
lemongrass
cumin
carnation
lavender
lemon juice
nutmeg
fennel
tabasco pepper
lime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

Ejemplos de grafos bipartitos



Ejemplos de grafos bipartitos

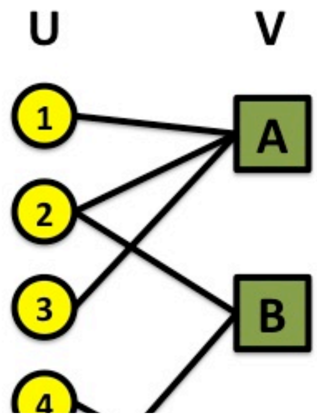
- Científicos
- Actores
- Músicos



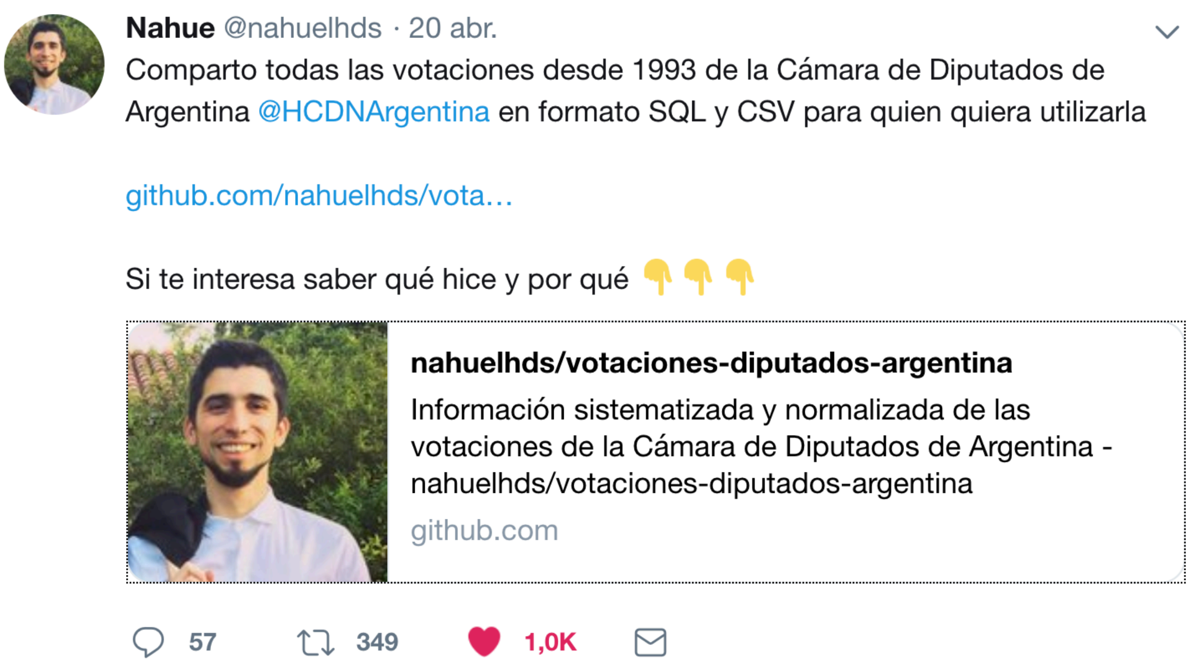
- Papers
- Películas
- Bandas, conciertos

Ejemplos de grafos bipartitos

Legisladores

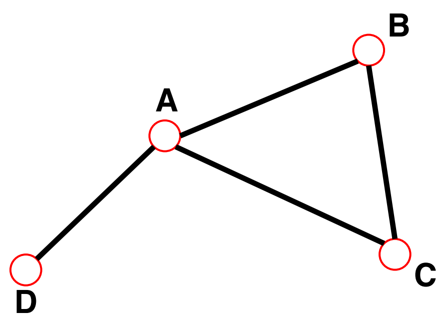


Leyes



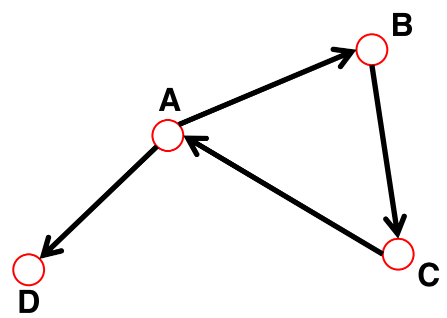
Sección 5

CAMINOS



The *distance (shortest path, geodesic path)* between two nodes is defined as the number of edges along the shortest path connecting them.

*If the two nodes are disconnected, the distance is infinity.



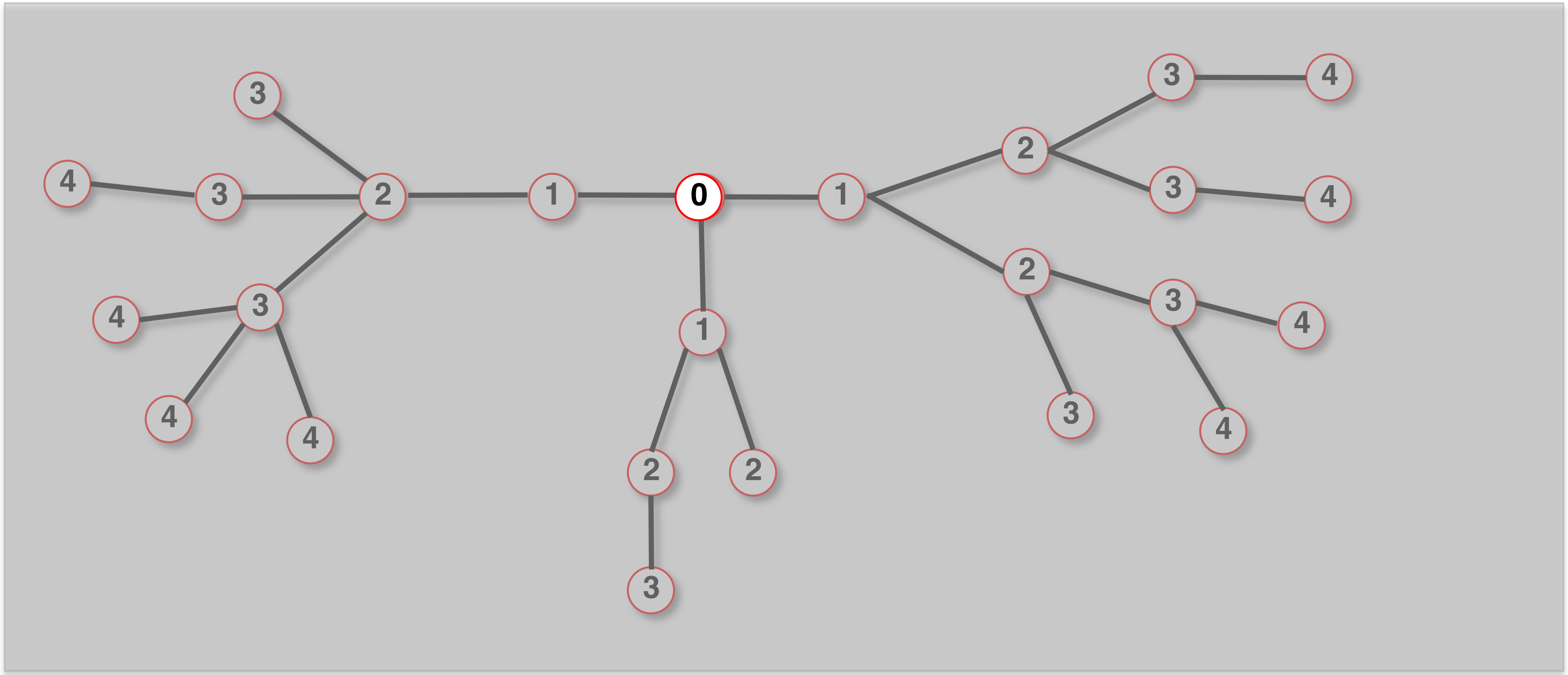
In directed graphs each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

CALCULANDO DISTANCIAS: BREADTH FIRST SEARCH

Distance between node 0 and node 4:

1.Start at 0.

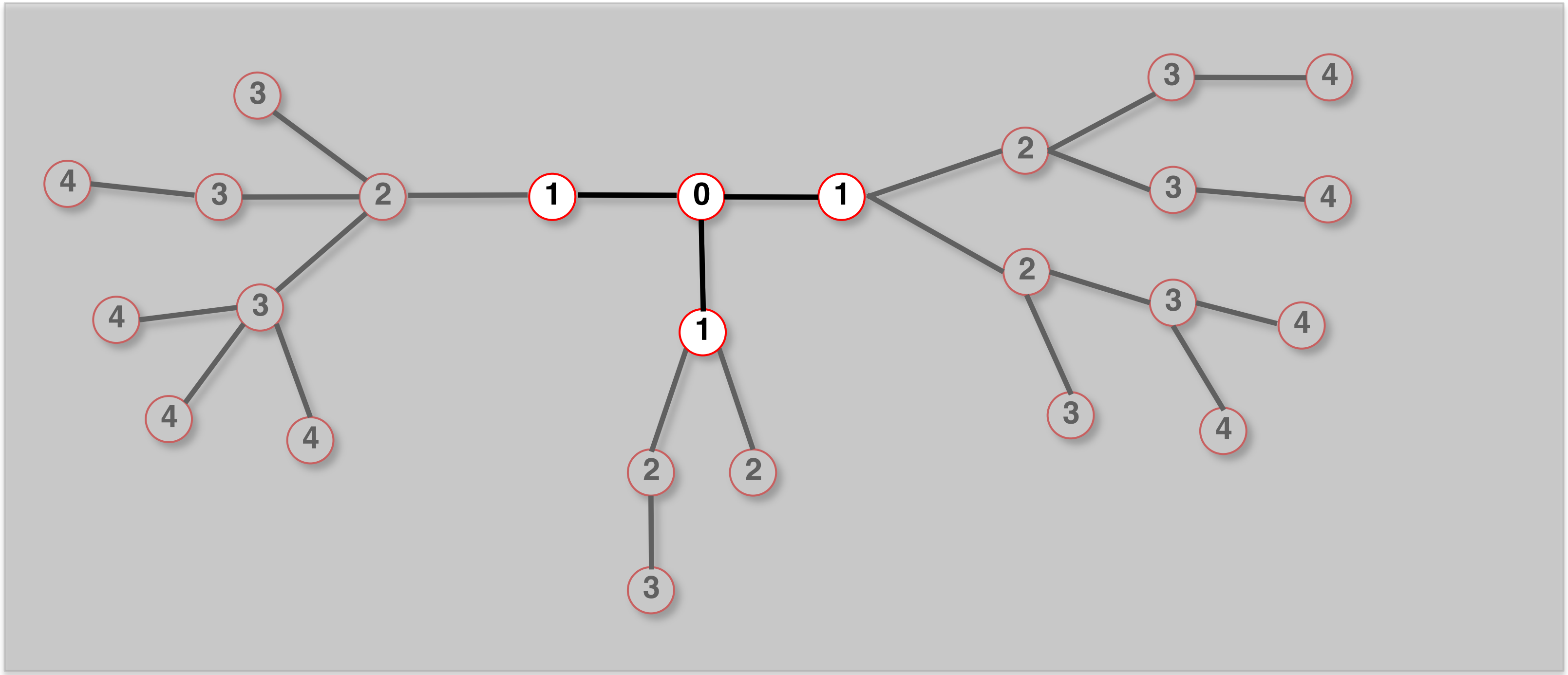


CALCULANDO DISTANCIAS: BREADTH FIRST SEARCH

Distance between node 0 and node 4:

1.Start at 0.

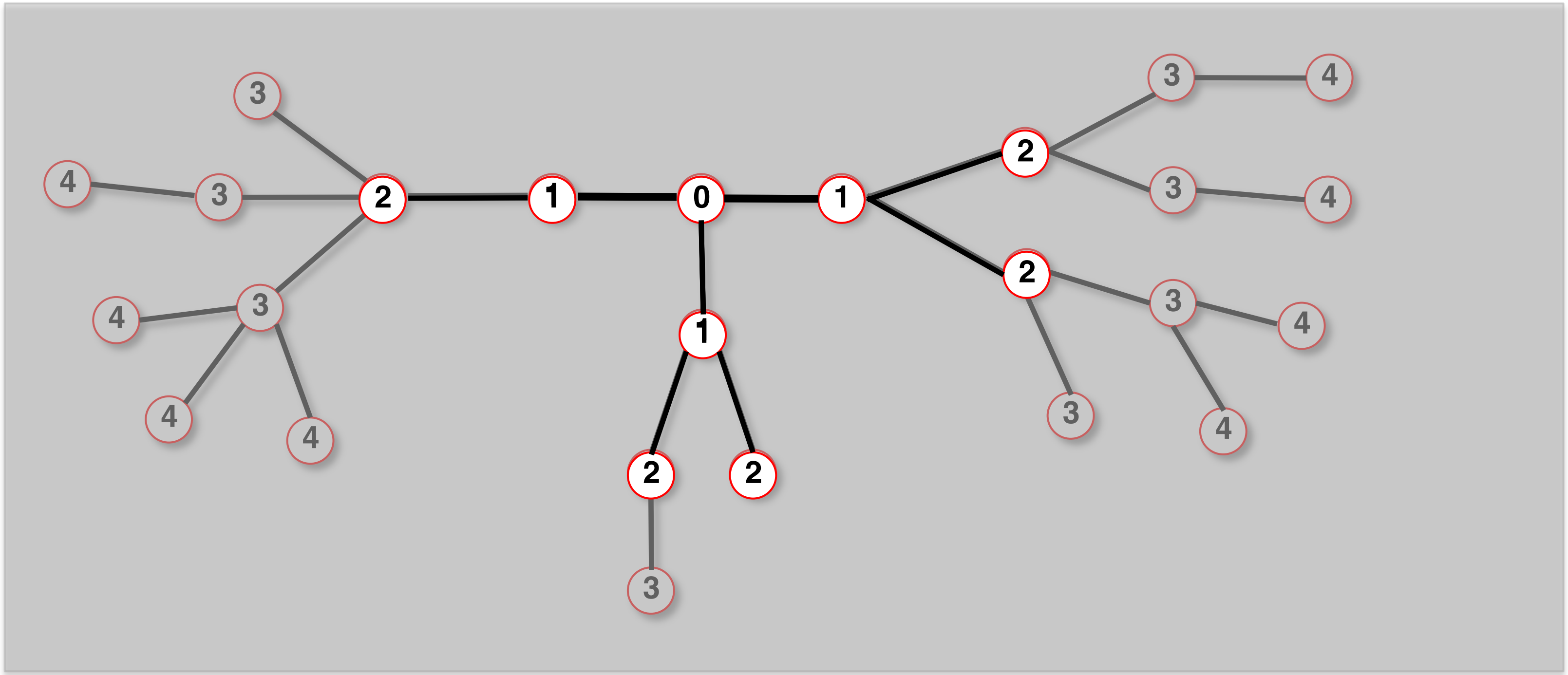
2.Find the nodes adjacent to 1. Mark them as at distance 1. Put them in a queue.



CALCULANDO DISTANCIAS: BREADTH FIRST SEARCH

Distance between node 0 and node 4:

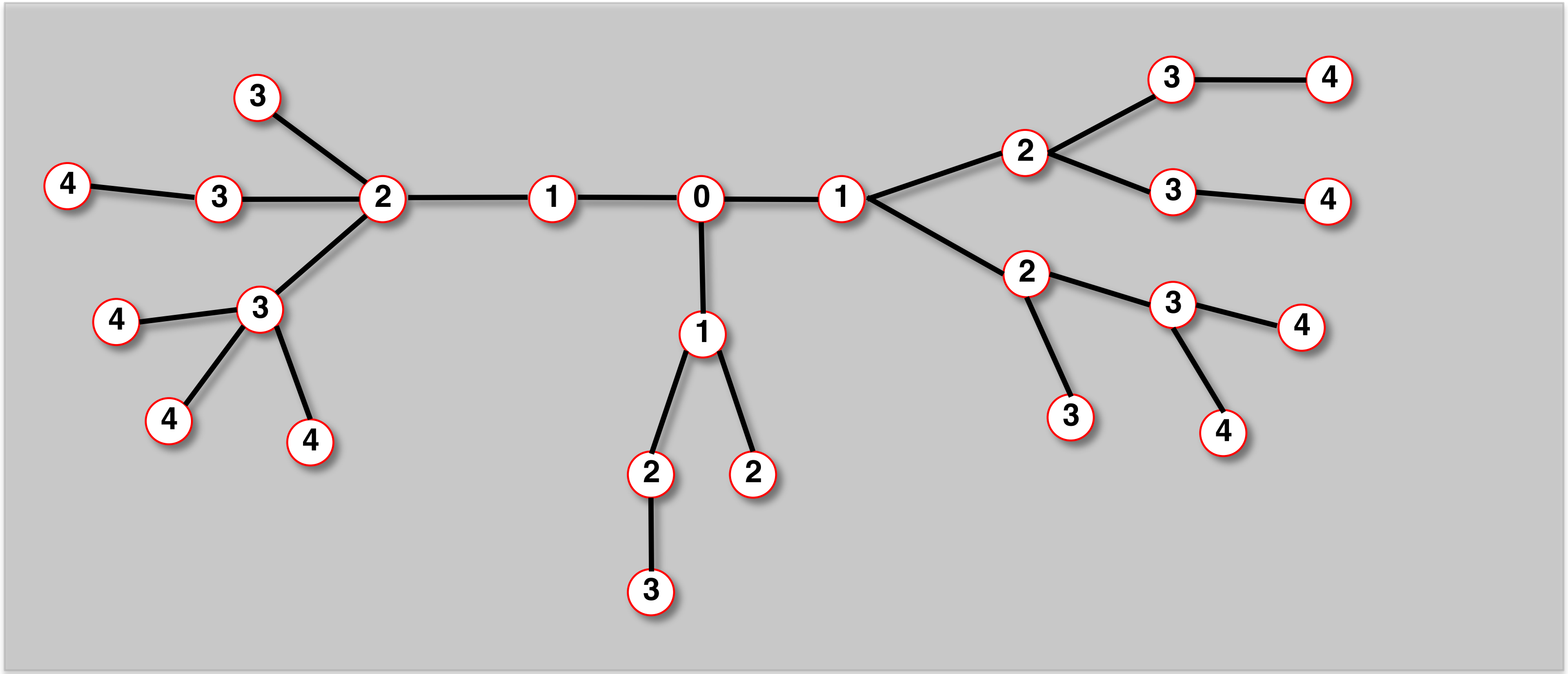
1. Start at 0.
2. Find the nodes adjacent to 0. Mark them as at distance 1. Put them in a queue.
3. Take the first node out of the queue. Find the unmarked nodes adjacent to it in the graph. Mark them with the label of 2. Put them in the queue.



CALCULANDO DISTANCIAS: BREADTH FIRST SEARCH

Distance between node 0 and node 4:

1.Repeat until you find node 4 or there are no more nodes in the queue.
2.The distance between 0 and 4 is the label of 4 or, if 4 does not have a label, infinity.



DIAMETRO DE LA RED Y DISTANCIA PROMEDIO

Diameter: $\boldsymbol{d_{max}}$ the maximum distance between any pair of nodes in the graph.

Average path length/distance, <d>, for a connected graph:

where d_{ij} is the distance from node *i* to node *j*

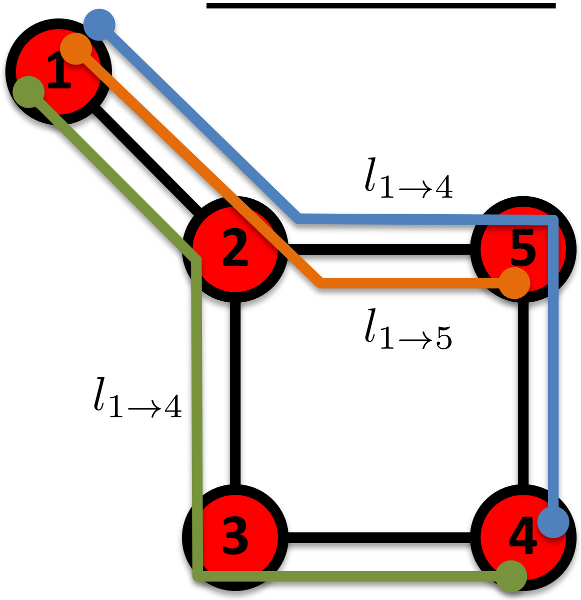
$$\langle d \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} d_{ij}$$

In an *undirected graph* $d_{ij} = d_{ji}$, *so* we only need to count them once:

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

CAMINOS: RESUMEN

Shortest Path



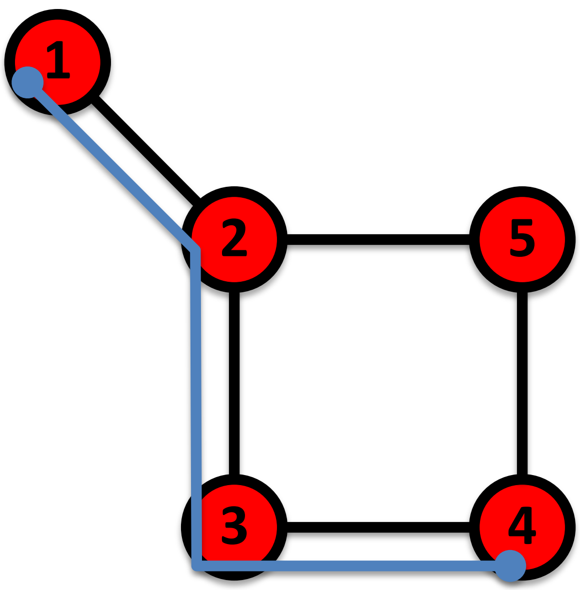
$$l_{1\to4} = 3$$

$$l_{1\to5} = 2$$

The path with the shortest length between two nodes (distance).

CAMINOS: RESUMEN

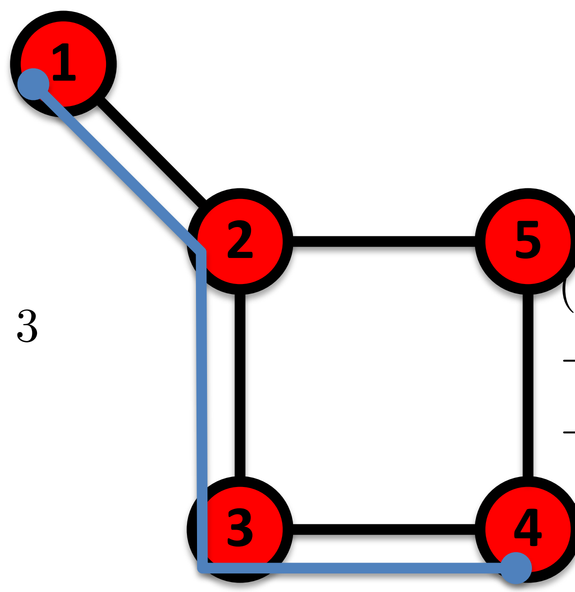
Diameter



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

The longest shortest path in a graph

Average Path Length

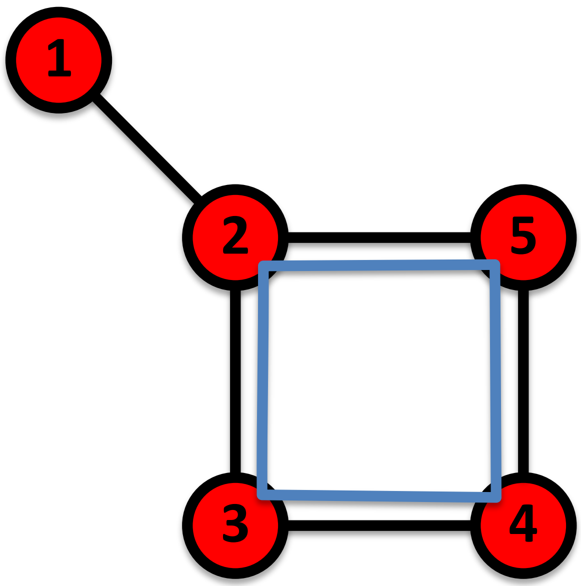


$$(l_{1\to2} + l_{1\to3} + l_{1\to4} + l_{1\to5} + l_{2\to3} + l_{2\to4} + l_{2\to5} + l_{3\to4} + l_{3\to5} + l_{4\to5})/10 = 1.6$$

The average of the shortest paths for all pairs of nodes.

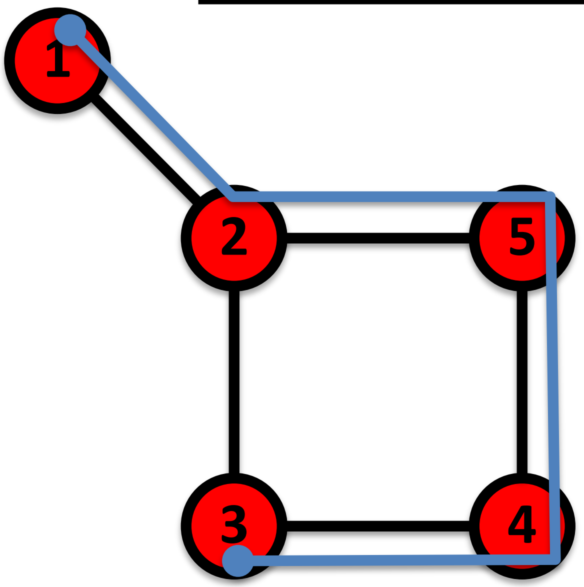
CAMINOS: RESUMEN

<u>Cycle</u>



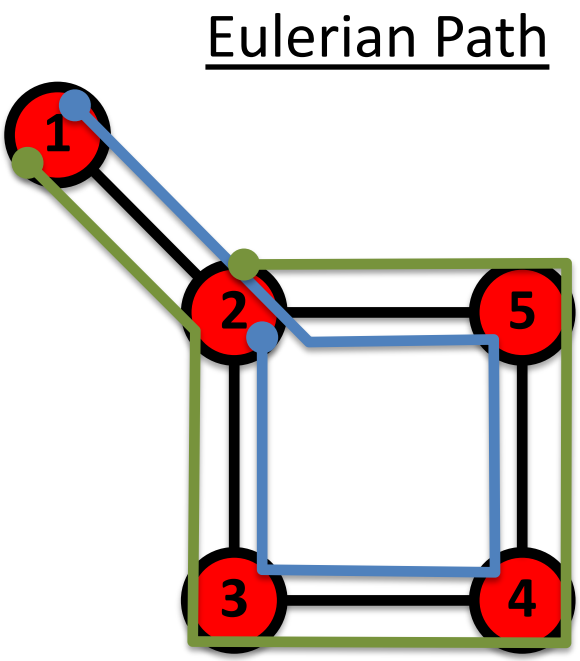
A path with the same start and end node.

<u>Self-avoiding Path</u>

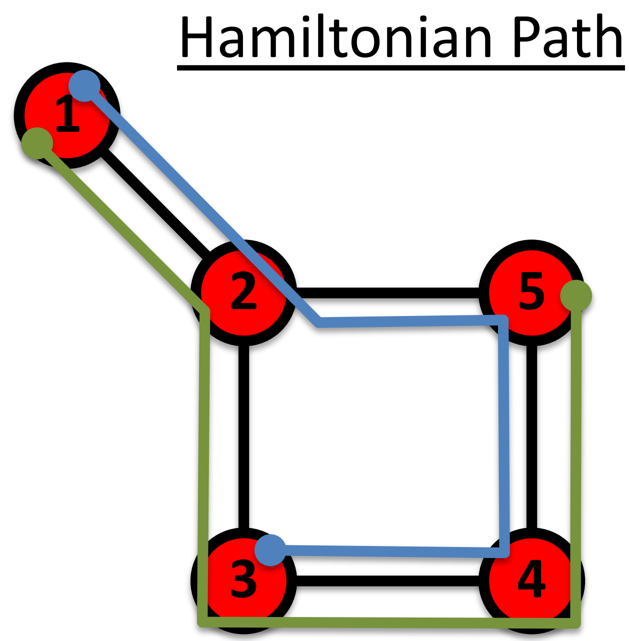


A path that does not intersect itself.

CAMINOS: RESUMEN



A path that traverses each link exactly once.

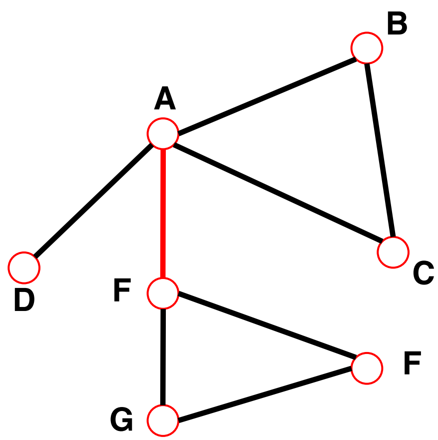


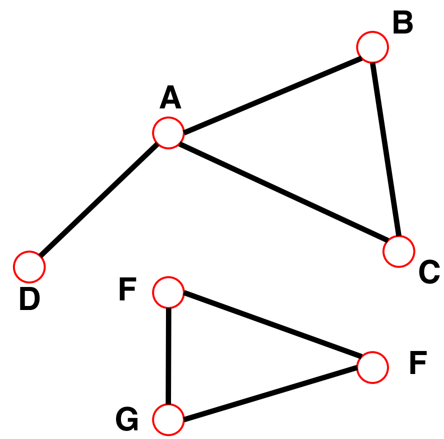
A path that visits each node exactly once.

CONECTIVIDAD

CONECTIVIDAD EN GRAFOS NO DIRIGIDOS

Connected (undirected) graph: any two vertices can be joined by a path.
A disconnected graph is made up by two or more connected components.





Largest Component:
Giant Connected Component

The rest: **Isolates**

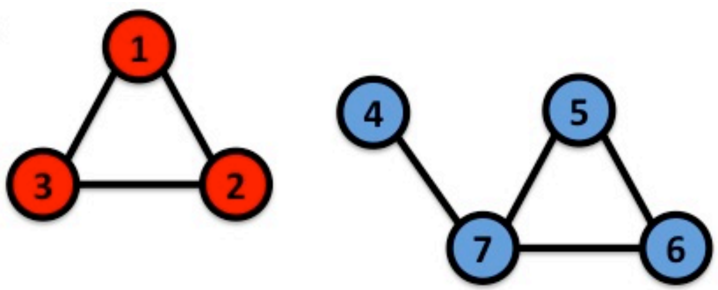
Bridge: if we erase it, the graph becomes disconnected.

CONECTIVIDAD EN GRAFOS NO DIRIGIDOS

Matriz de Adyacencia

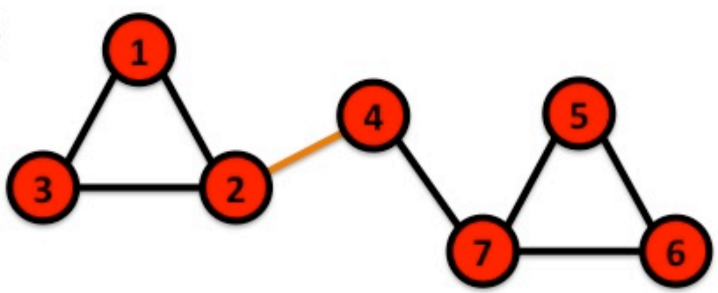
The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero:

(a)



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

(b)



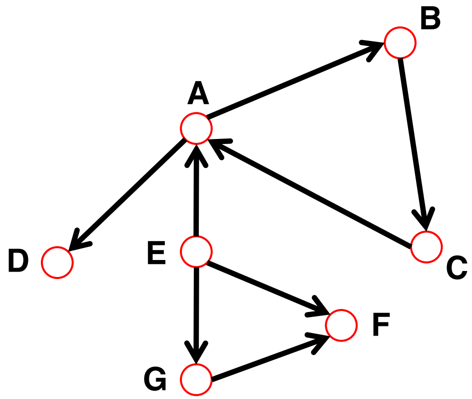
$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

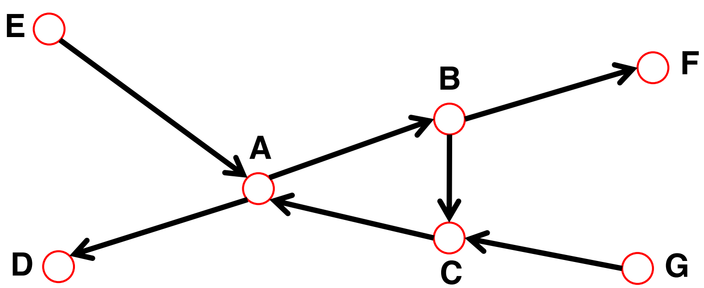
CONECTIVIDAD EN GRAFOS DIRIGIDOS

Strongly connected directed graph: has a path from each node to every other node and vice versa (e.g. AB path and BA path).

Weakly connected directed graph: it is connected if we disregard the edge directions.

Strongly connected components can be identified, but not every node is part of a nontrivial strongly connected component.





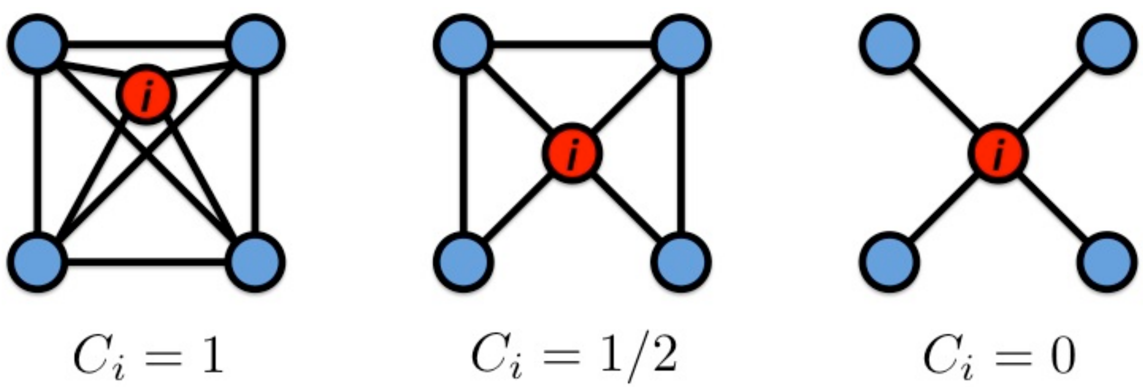
Coeficiente de Clustering

Coeficiente de Clustering:

qué fracción de tus vecinos están conectados?

- Nodo i con grado k_i
- C_i en [0,1]

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



Watts & Strogatz, Nature 1998.

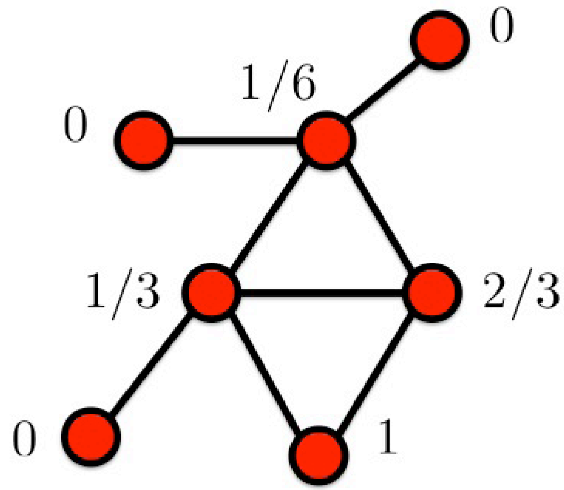
COEFICIENTE DE CLUSTERING

✱ Coeficiente de Clustering:

qué fracción de tus vecinos están conectados?

- ✱ Nodo i con grado k_i
- ✱ C_i en [0,1]

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



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

$$C = \frac{3}{8} = 0.375$$

Watts & Strogatz, Nature 1998.

RESUMEN

TRES MÉTRICAS CENTRALES EN CIENCIA DE REDES

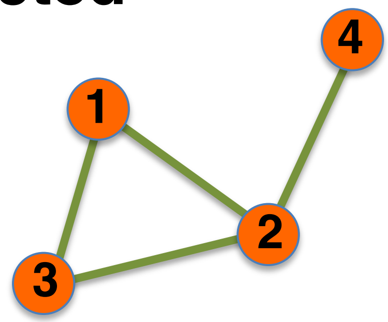
Distribución de grados: **P(k)**

Longitud de caminos: ***<d>***

Coeficiente de Clustering:

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

Undirected



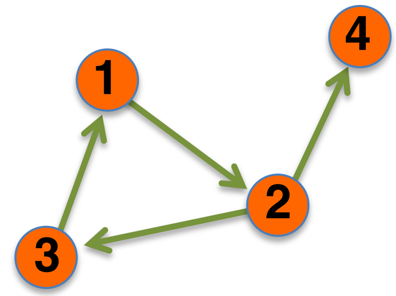
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

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

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

Actor network, protein-protein interactions

Directed



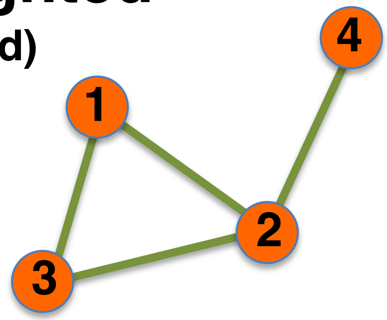
$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

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

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

WWW, citation networks

Unweighted
(undirected)



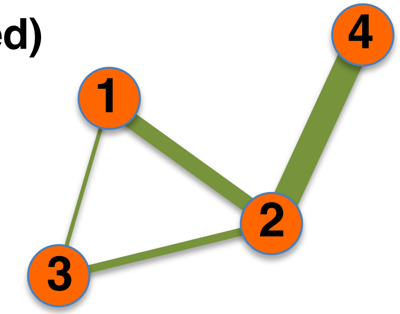
$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

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

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

protein-protein interactions, www

Weighted
(undirected)



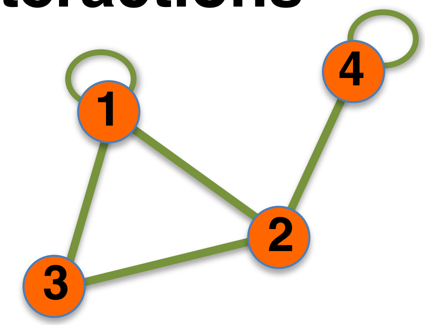
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

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

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

Call Graph, metabolic networks

Self-interactions



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

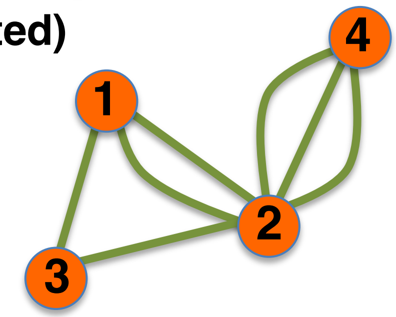
$$A_{ii} \neq 0 \qquad A_{ij} = A_{ji}$$

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

Protein interaction network, www

Multigraph

(undirected)



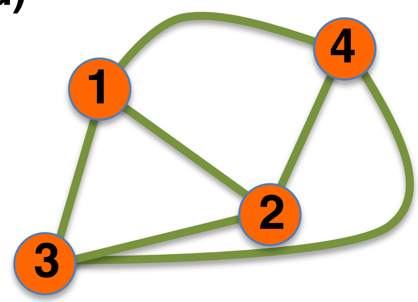
$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

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

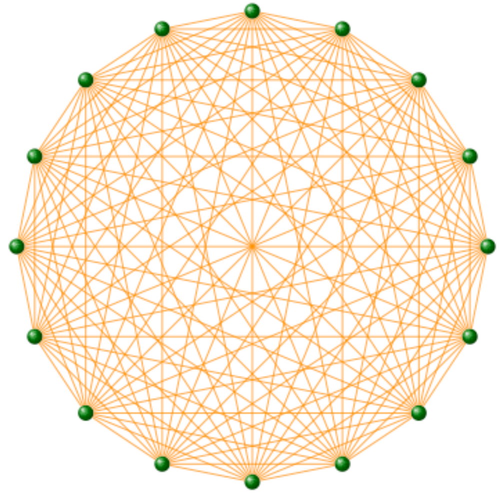
$$L = \frac{1}{2}\sum_{i,j=1}^{N} nonzero(A_{ij}) \qquad < k >= \frac{2L}{N}$$

Social networks, collaboration networks

Complete Graph
(undirected)



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



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

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

Actor network, protein-protein interactions