School on Biological Complex Networks: From the Cell to the Brain and beyond

Introduction to Biological Complex Networks

Javier M. Buldú

http://www.complexity.es/jmbuldu

Laboratory of Biological Networks
Center for Biomedical Technology (Madrid, Spain)

Complex Systems Group
Universidad Rey Juan Carlos (Madrid, Spain)

OUTLINE

1.- Introduction to Complex networks
      1.1.- What is a (complex) network?
      1.2.- Types of networks
      1.3.- Basic concepts about networks
      1.4.- Brief historical background
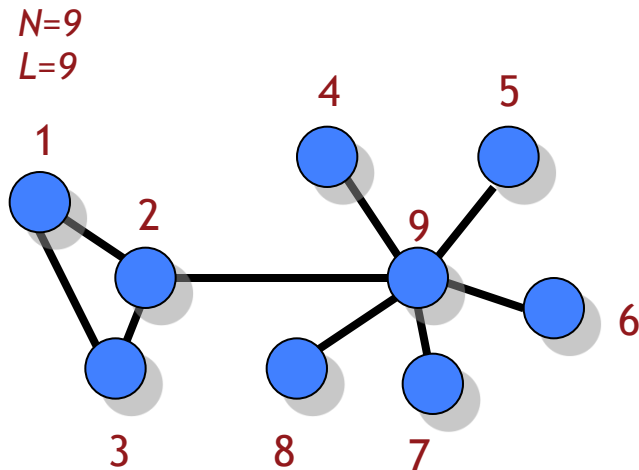
2.- Applications to Biological Networks
      2.1.- Genetic, protein and metabolic networks
      2.2.- Neuron and brain networks
      2.3.- Networks in ecology
      2.4.- Disease spreading

CTB centro de tecnologia biomédica   CSG   Universidad Rey Juan Carlos

1.- INTRODUCTION TO COMPLEX NETWORKS

1.1.- What is a (complex) network?

❑ **A Network is a set of elements with connections between them**
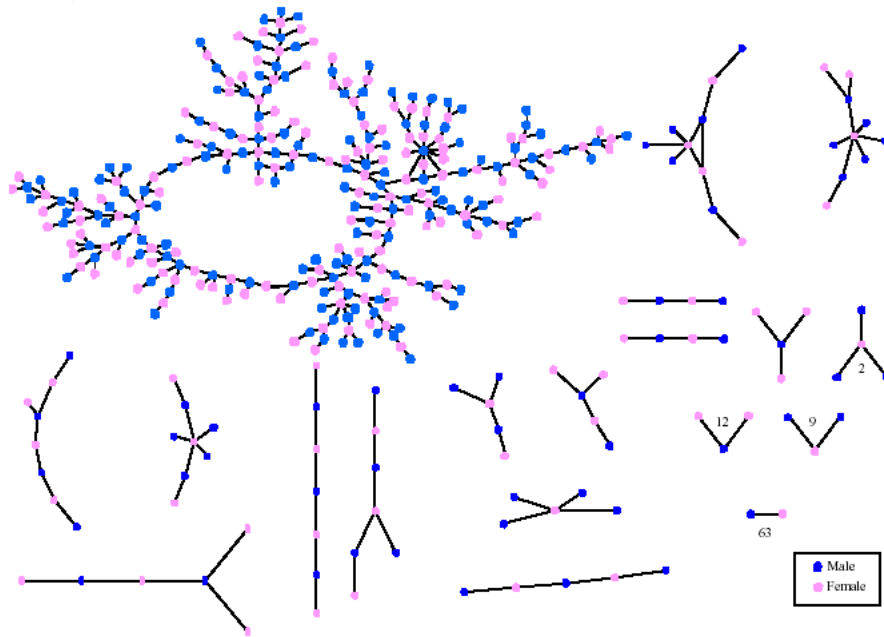
N=9
L=9

A network (graph) *G=(N,L)* consists of a set of $N=\{n_1, n_2, \dots , n_N\}$ nodes and a set of $L=\{l_1, l_2, \dots , l_M\}$ links.

A graph is the mathematical abstraction of a network. Despite it is not rigorous, we will use both terms, graph and network, as synonyms.

From this viewpoint, each element is represented by a site (physics), node (computer science), actor (sociology) or vertex (graph theory) and the interaction between two elements corresponds to a bond (physics), link (computer science), tie (sociology) or edge (graph theory).

1.1.- WHAT IS A (COMPLEX) NETWORK?

❑ **Nodes and links may arise from completely different contexts:**

Schematic representation of a network of hosts and routers.

Madrid Power Grid. From http://www.ree.es

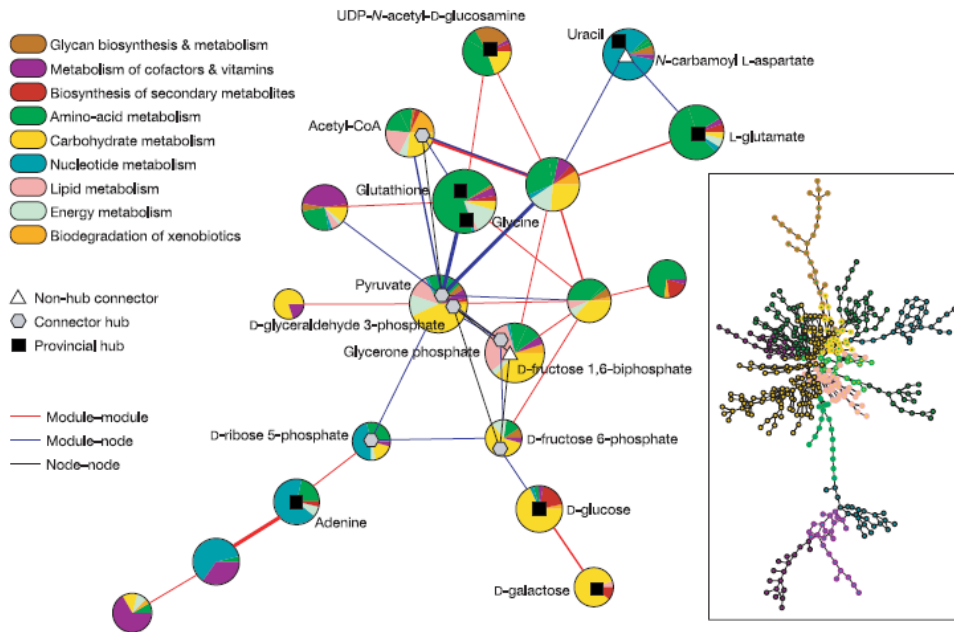# 1.1.- WHAT IS A (COMPLEX) NETWORK?

Structure of romantic and sexual contact at Jefferson High School
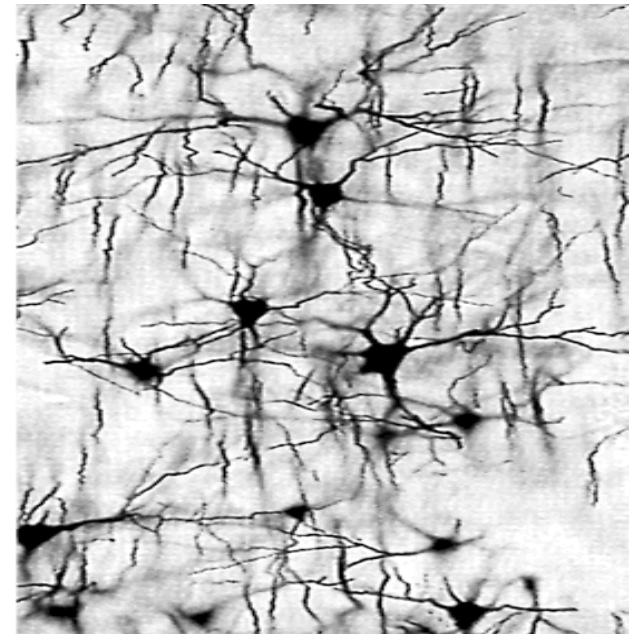From P.S. Bearman et al., AJS, 110, 44 (2004)
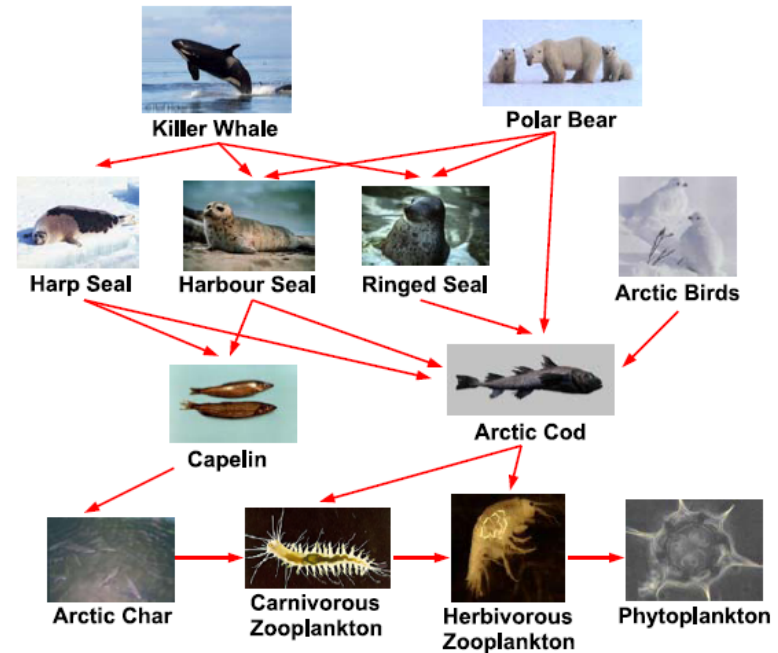
Simplified representation of the Arctic food web
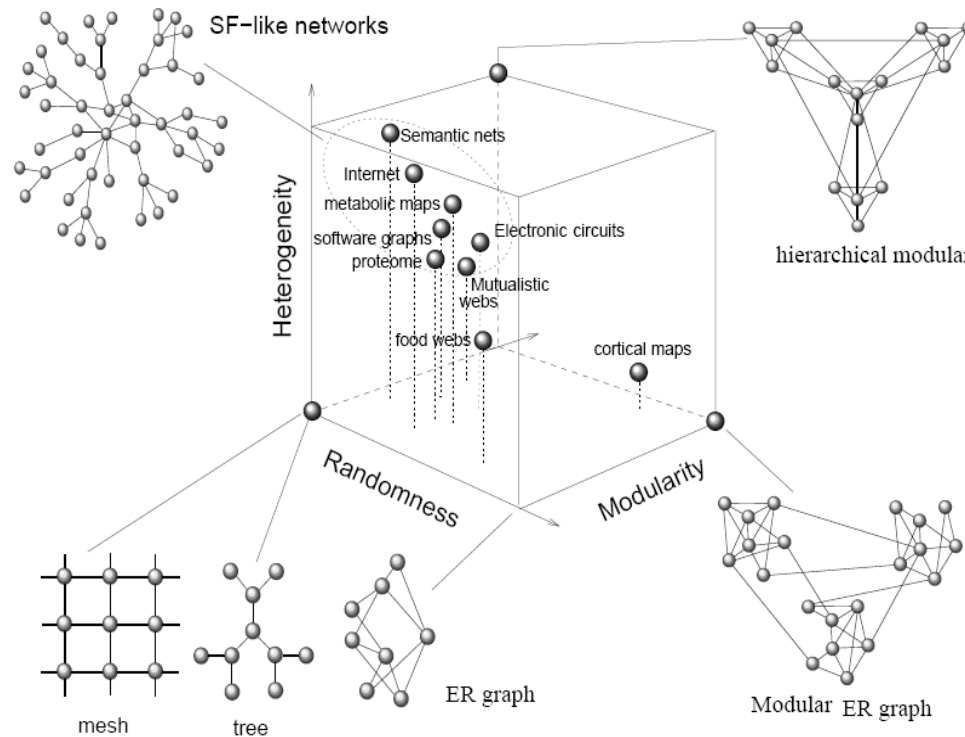
1.1.- WHAT IS A (COMPLEX) NETWORK?

Metabolic network of the *E. Coli*.
From Guimerà et al., Nature, 433, 895, 2005

Network of neurons

❑ A **Complex Network** is a network with non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random.

From: R.V. Solé and S. Valverde,
Lecture Notes in Physics, **650**, 189, 2004

1.2.- Types of networks

❑ **There exist different classifications of networks:**

    ❑   According to the direction of the links: **directed or undirected.**

    ❑   According to the kind of interaction: **weighted or unweighted**.

    ❑   According to the differences between nodes: **bipartite** or not.

    ❑   According to the evolution of their topology: **static or evolving.**

    ❑   According to the dynamics of the nodes: **with/without dynamics.**
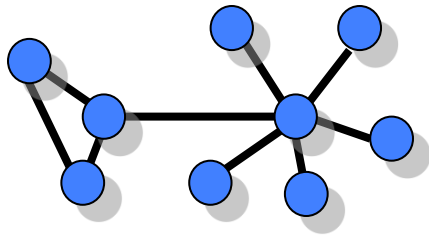
    ❑   …

1.2.- TYPES OF NETWORKS

more at: **M.E.J. Newman,** *Networks: An Introduction*
S. Boccaletti et al., Phys. Rep., 424, 175 (2006)
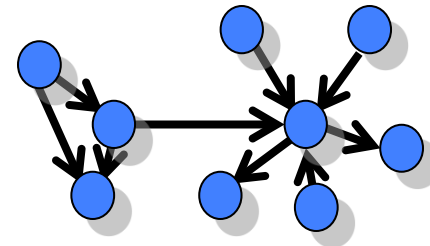
❑ **Directed and undirected networks:**

The relationship between nodes may be symmetric (undirected networks) or asymmetric (directed networks).

Undirected network   **Directed network (digraph)**

Examples: router network, power grids. collaboration networks, etc...

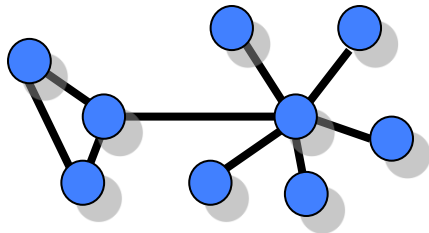**Examples: WWW, food webs, e-mail/telephone networks, etc...**

The direction of the links is crucial in dynamical processes ocurring in the network, such as information spreading, synchronization or network robustness.

1.2.- TYPES OF NETWORKS
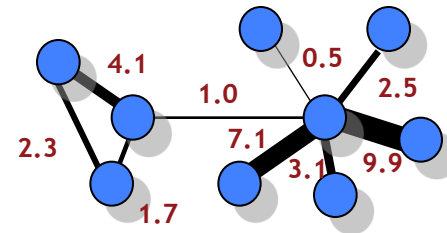
❑ **Weighted and unweighted networks:**

The capacity or intensity of the relationship between nodes may be heterogeneous (weighted networks).

Unweighted network

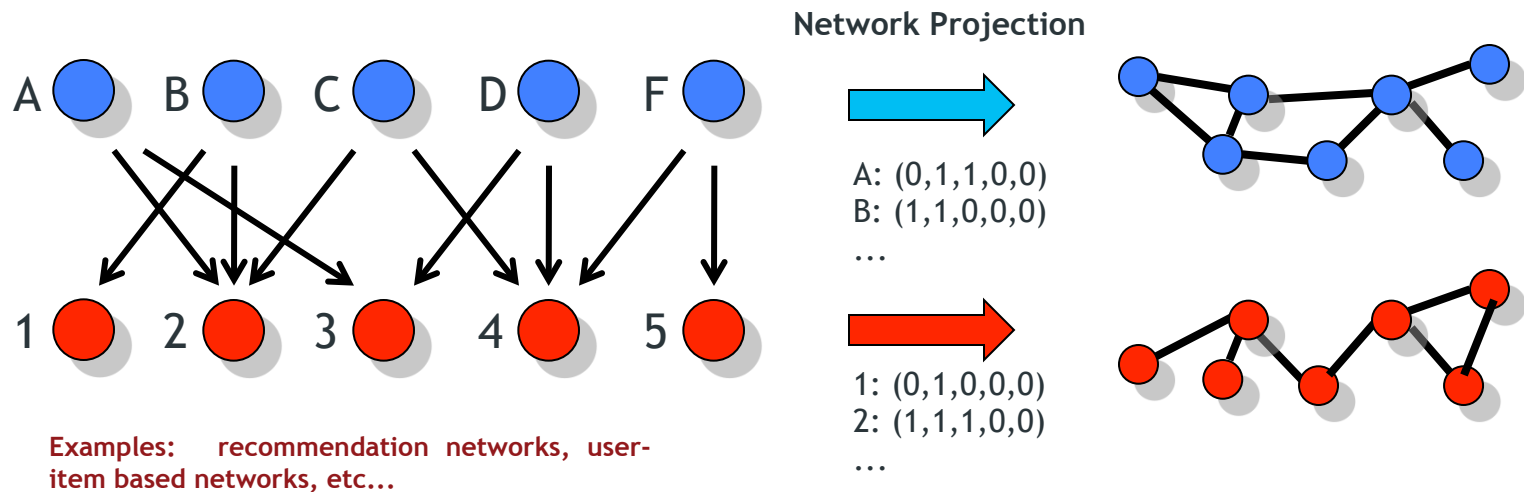Examples: citation network, Internet, etc...

Weighted network

Examples: e-mail/telephone networks, food webs, power grid, colaboration networks, etc...

Again, the weight of the links is crucial in dynamical processes ocurring in the network, such as information spreading, synchronization or network robustness.

1.2.- TYPES OF NETWORKS

❑ **Bipartite networks:**

Networks with two (or more) kind of nodes and links joining ONLY nodes of unlike type.

Network Projection

A: (0,1,1,0,0)
B: (1,1,0,0,0)
…

1: (0,1,0,0,0)
2: (1,1,1,0,0)
…

Examples: recommendation networks, user-item based networks, etc...

Despite being bipartite, it is possible to project the network.

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

1.2.- TYPES OF NETWORKS

☐ Static or evolving networks:

Networks do not appear suddenly. We have to know if the network that we are studying is static (its structure is stationary) or if it is still evolving
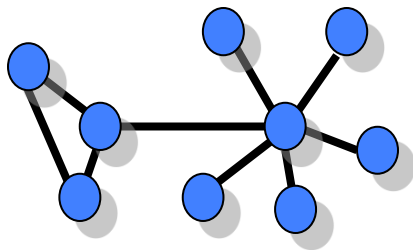

t=0

t=1

t=2

t=3
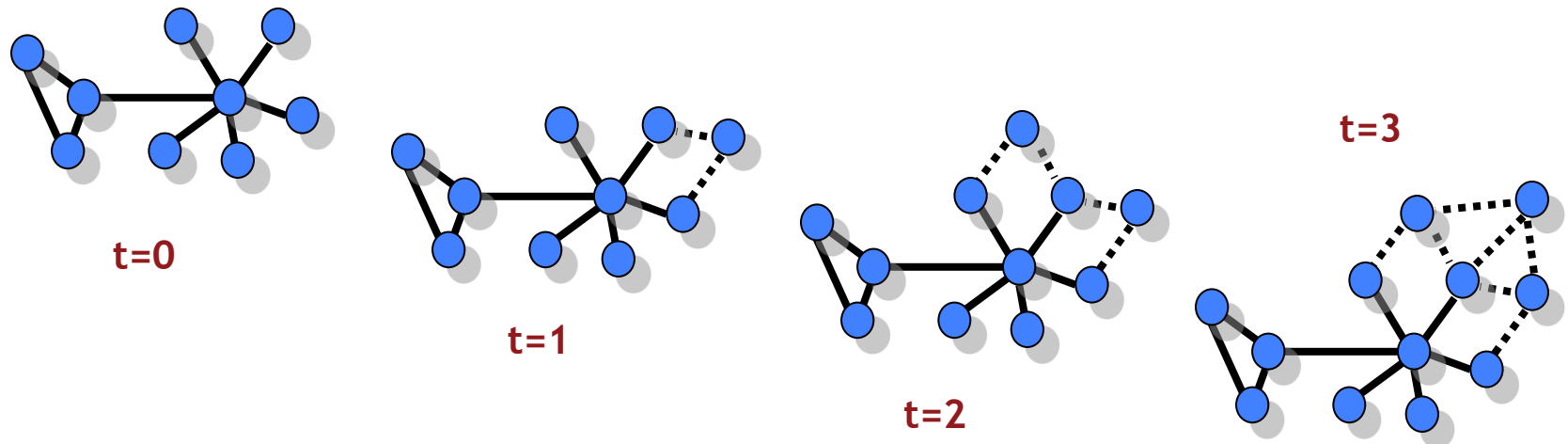
Two fundamental questions are addressed when working with evolving networks: what are the rules governing the evolution? What consequences have the rules on the final topology?
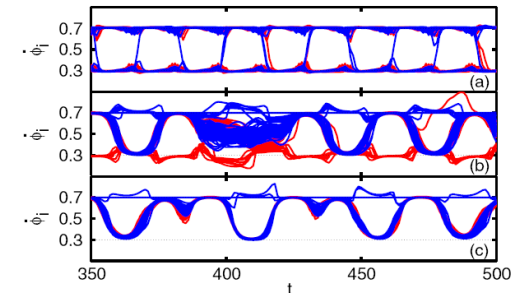
❑ **Networks of dynamical systems:**

Nodes are dynamical systems whose dynamics is influenced through the matrix of connections.

Nodes are (coupled) dynamical systems (periodic oscilators, excitable systems, chaotic oscilators, bistable systems, ...)

$$\dot{\phi}_i = \begin{cases} \omega_i + \frac{d}{(k_i+k_{p_i})} \sum_{j=1}^{N} a_{ij} \sin(\phi_j - \phi_i) \\ + \frac{d_p k_{p_i}}{(k_i+k_{p_i})} \sin(\phi_{p_i} - \phi_i), \end{cases}$$

In this case, we have to study the influence of the topology in the dynamical processes occurring in the network (synchronization, stochastic processes, etc..) ...   ... and vice-versa!

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

1.2.- TYPES OF NETWORKS

❑   Despite the different types of networks, which in turn are obtained from completely different interacting systems (people, neurons, proteins, routers,...) we will see that they share some universal properties
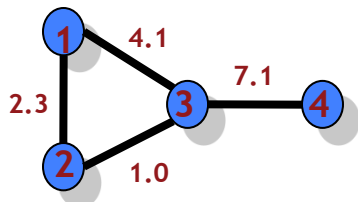
Is it a social network?
A technological network?
A biological network?
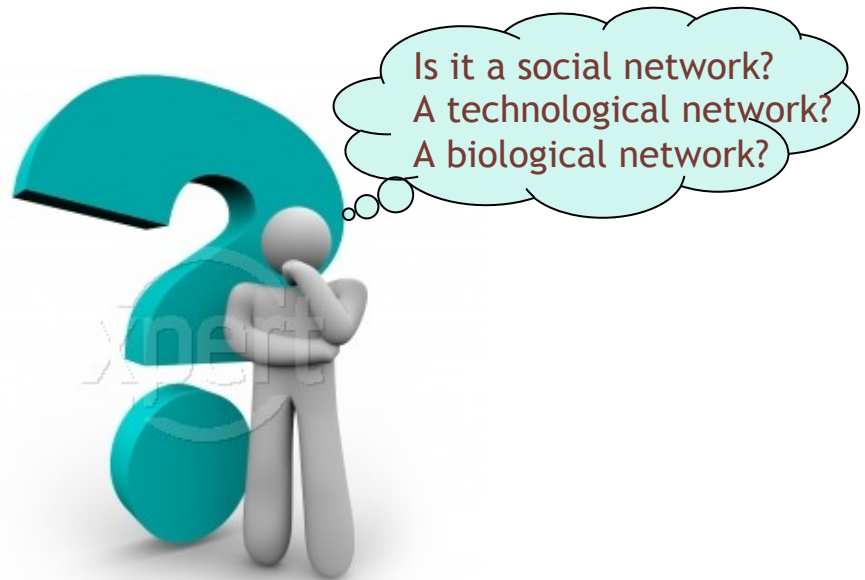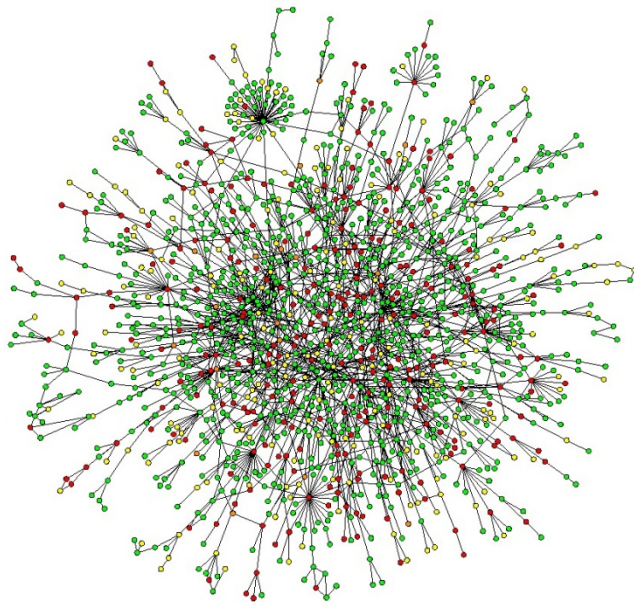
1.3.- Basic concepts about networks

❑ Adjacency, Weights and Laplacian Matrix:

All the former networks can be described using a matricial formalism.
Given a set of *N* nodes with *M* conections between them:

Weights Matrix (*W*):

Entries of the matrix are the weights w_{ij} (i,j=1, ..., N) of the connections

$$\begin{pmatrix} 0.0 & 2.3 & 4.1 & 0.0 \\ 2.3 & 0.0 & 1.0 & 0.0 \\ 4.1 & 1.0 & 0.0 & 7.1 \\ 0.0 & 0.0 & 7.1 & 0.0 \end{pmatrix}$$

Adjacency Matrix (*A*):

$a_{ij}=1$ if there exists a link between i and j, and $a_{ij}=0$ otherwise

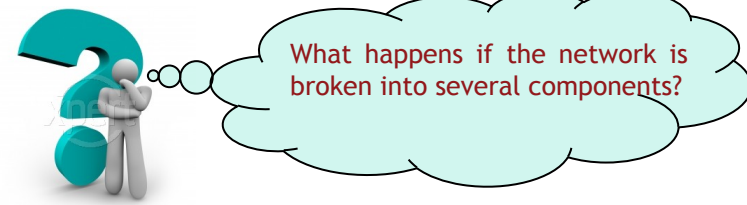$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Laplacian Matrix (*L*):

The Laplacian matrix is defined as *L=K-A*, where *K* is a diagonal matrix of elements $k_{ii}=\sum a_{ij}$. Thus, it has a zero-row sum.

$$\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 3 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

Matrices will be symmetric if networks are undirected.

What happens if the network is broken into several components?

❑ Shortest path, average path length and diameter:

Shortest path (d_{ij}):

The shortest path d_{ij} between nodes i and j corresponds to the minimal distance (or weight) between all paths that connect i and j

Average path length (l):

The average path length l is the average shortest path between all nodes in the network:

$$\ell = \langle d_{ij} \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

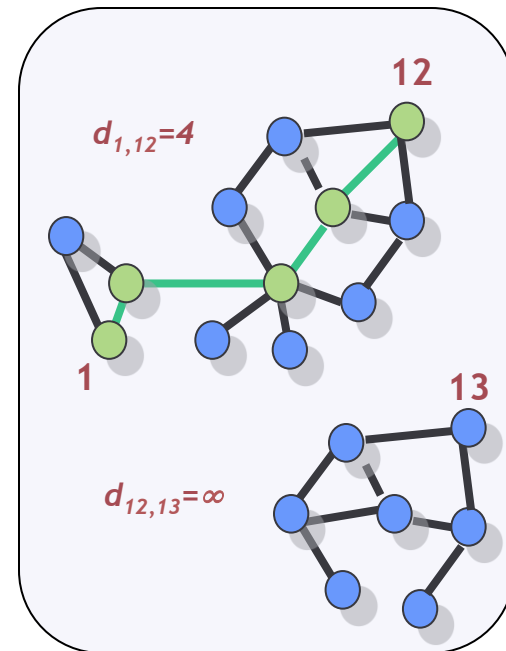when the network is not connected it is usefull to define the "harmonic mean"

$$\ell = \frac{1}{\langle d_{ij}^{-1} \rangle} = \left(\frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \right)^{-1}$$
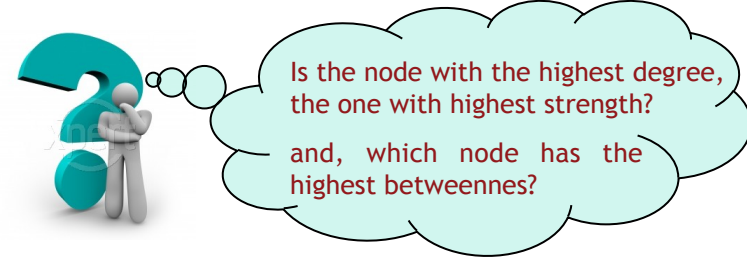
Diameter(D):

The maximum between all shortest paths $D=max(d_{ij})$

Component:

The set of nodes reachable from a given node.

$d_{1,12}=4$

12

1

13

$d_{12,13}=\infty$

Is the node with the highest degree, the one with highest strength?

and, which node has the highest betweennes?

□ **Degree, strength and betweenness:**

Degree (k_i):

The degree k_i of a node i is the number of connections of the node

Strength (s_i):

The strength s_i of a node i is the sum of the weigths of the connections to that node $s_i = \sum w_{ij}$

Betweenness (b_i):

The betweeennes of a node i (or a link) accounts for the number of shortest paths passing through that node (or link).

$$b_i = \sum_{j,k \in \mathcal{N}, j \neq k} \frac{n_{jk}(i)}{n_{jk}}$$

where n_{jk} is the number of shortest paths connecting j and k, and $n_{jk}(i)$ are those shortest paths between j and k that pass through i.
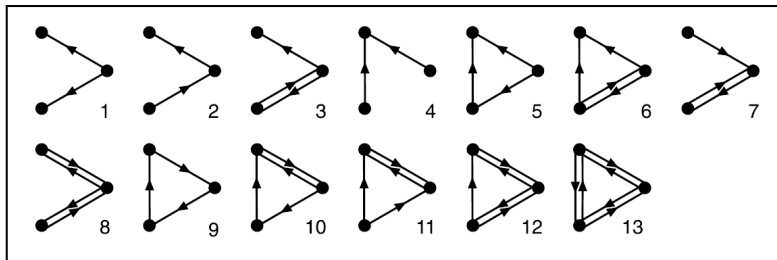
CTB centro de tecnologia biomédica

CSG
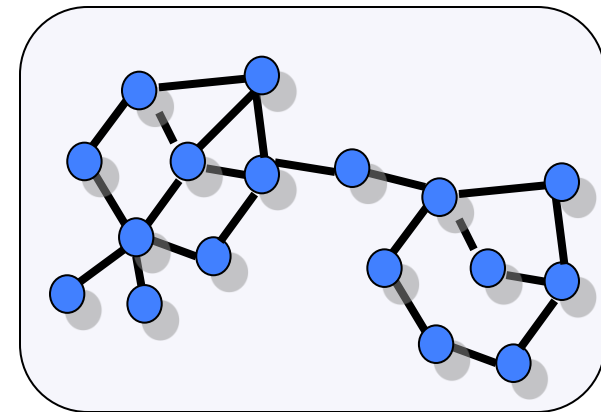
Universidad Rey Juan Carlos

❑ Network Motifs:

Network motifs are patterns (sub-graphs) that occur within a network much more often than expected at random.

Example: all 13 types of three-node connected subgraphs:

Each network motif can carry out specific information-processing functions

Figures from: Milo et al., Science, 298, 824 (2002)

Network	Nodes	Edges	N_{real}	$N_{rand} \pm$ SD	Z score
Gene regulation (transcription)				X ↓ Y ↓ Z	Feed-forward loop
E. coli	424	519	40	7 ± 3	10
*S. cerevisiae**	685	1,052	70	11 ± 4	14
Neurons				X ↓ Y ↓ Z	Feed-forward loop
C. elegans†	252	509	125	90 ± 10	3.7
Electronic circuits (digital fractional multipliers)				X ↗ ↘ Y ← Z	Three-node feedback loop
s208	122	189	10	1 ± 1	9
s420	252	399	20	1 ± 1	18
s838‡	512	819	40	1 ± 1	38
World Wide Web				X ↕ Y ↕ Z	Feedback with two mutual dyads
nd.edu§	325,729	1.46e6	1.1e5	$2e3 \pm 1e2$	800

Are my friends, friends of my friends?

❑ **Clustering coefficient:**

The clustering coefficient *C* accounts for the number of triangles in the network. Specifically, C_i is the ratio between the number of links *E* connecting the nearest neighbors of *i* and the total number of possible links between these neighbors.
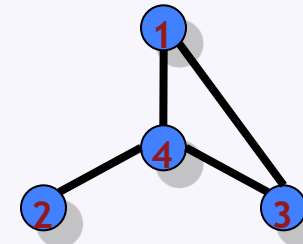
$$C_i = \frac{2E}{k_i(k_i - 1)}$$

The clustering coefficient of the network *C* is the average of C_i over all nodes.
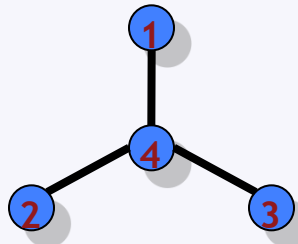
$C_{1,2,3,4} = \{0,0,0,0\}$
$C=0$
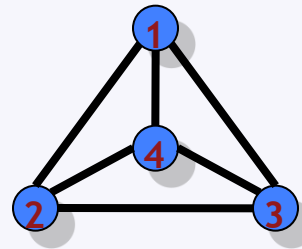
$C_{1,2,3,4} = \{1,1,1,1\}$
$C=1$

$C_{1,2,3,4} = \{1,0,1,1/3\}$
$C=7/12$

❏ Local and Global Efficiency:

The efficiency overcomes the divergence of the shortest paths if the graph is disconnected

Global Efficency (*E*):

The global efficiency is the harmonic mean of the geodesic paths between all nodes of the network:

$$E = \frac{1}{N(N-1)} \sum_{i,j \in \mathcal{N}, i \neq j} \frac{1}{d_{ij}}$$

Local Efficency (*E*$_i$):

The local efficiency E_i of a node i, measures the inverse of the shortest path length between the subset G_i of neighbors of the node i, when i is not present.

$$E_{\mathrm{loc}} = \frac{1}{N} \sum_{i \in \mathcal{N}} E(G_i)$$

The local efficiency is related, somehow, with the clustering cofficient

CTB centro de tecnologia biomédica

CSG
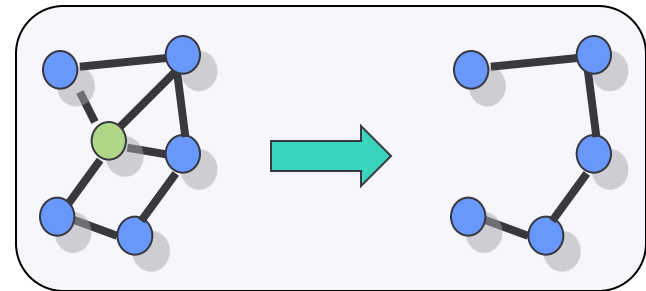
Universidad Rey Juan Carlos

❑ Graph Spectrum:

The spectrum of a graph is the set of eigenvalues of its adjacency (or Laplacian) matrix *A*. A graph $G_{N,M}$, has *N* eigenvalues $\mu_i=(\mu_1,\ \mu_2,...,\ \mu_N)$ and *N* associated eigenvectors $v_i=(v_1,v_2,...,\ v_N)$.

The eigenvalues and associated eigenvectors of a graph are intimately related to important topological features such as the diameter, the number of cycles, information transmission and the connectivity properties of the graph.
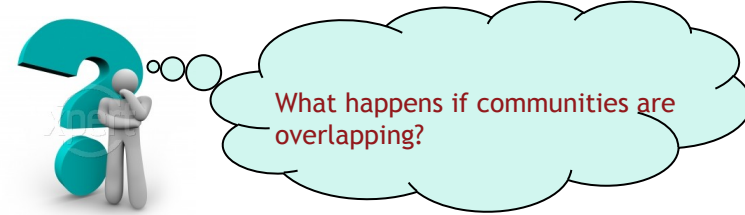
Spectral density:

$$\rho(\mu) = \frac{1}{N} \sum_{i=1}^{N} \delta(\mu - \mu_i)$$
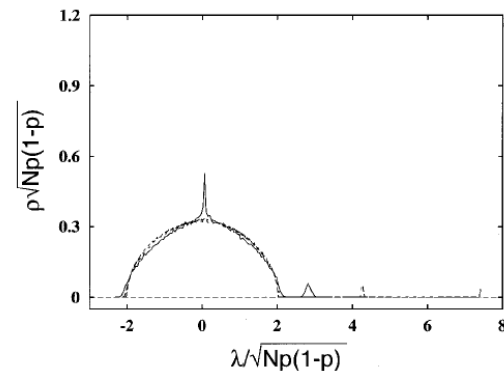
Rescaled spectral density of three random graphs having p=0.05 and size N=100, N=300, and N=1000. The isolated peak corresponds to the principal eigenvalue.

What happens if communities are overlapping?

❑ **Community Structure (I):**

Given a graph $G_{N,M}$, a community is a subgraph $G'_{N',M'}$ whose nodes are thightly connected (or at least, more connected than in a random equivalent network).
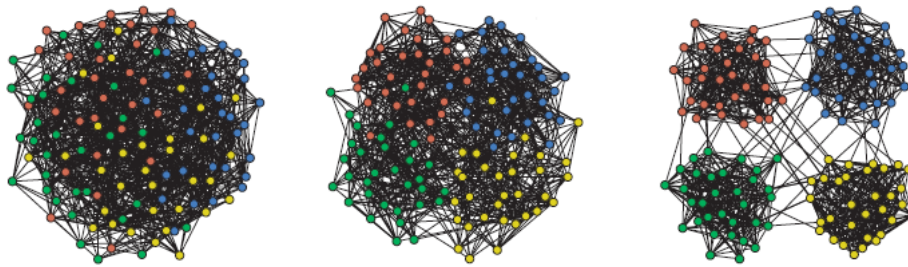

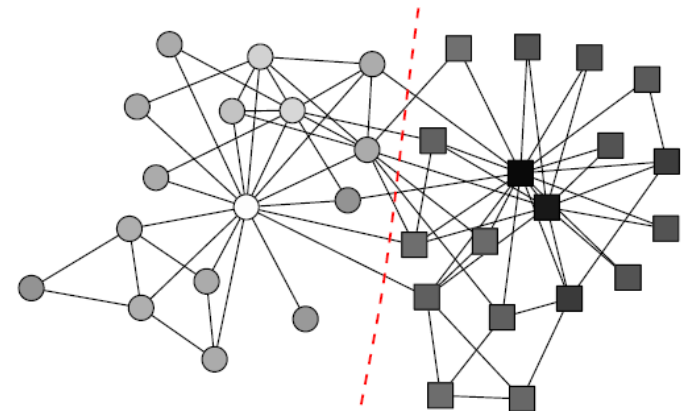Figure from: Guimerà et al., Nature, 433, 895(2005)

Zachary Karate Club

Figure from: M. E. J. Newman, Proc. Natl. Acad. Sci. USA 103, 8577 (200

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

❑ Community Structure (II):

Several algorithms have been proposed in order to split a sparse network into communities:
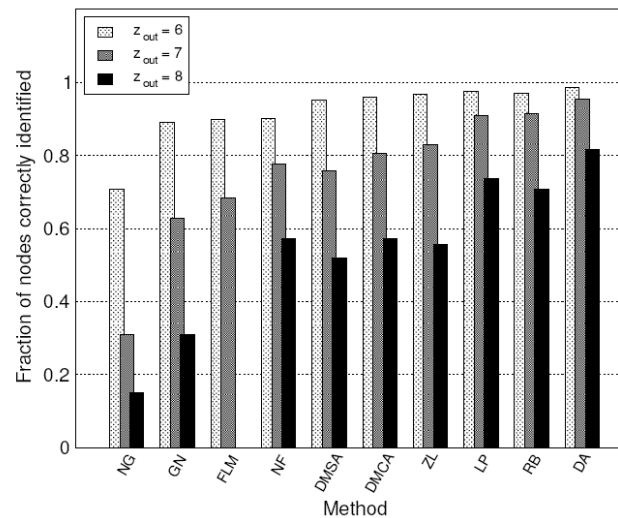

Figure from: L. Danon et al., World Scientific, 93-113 (2007)

Modularity M is and objective measure in order to evaluate community division:
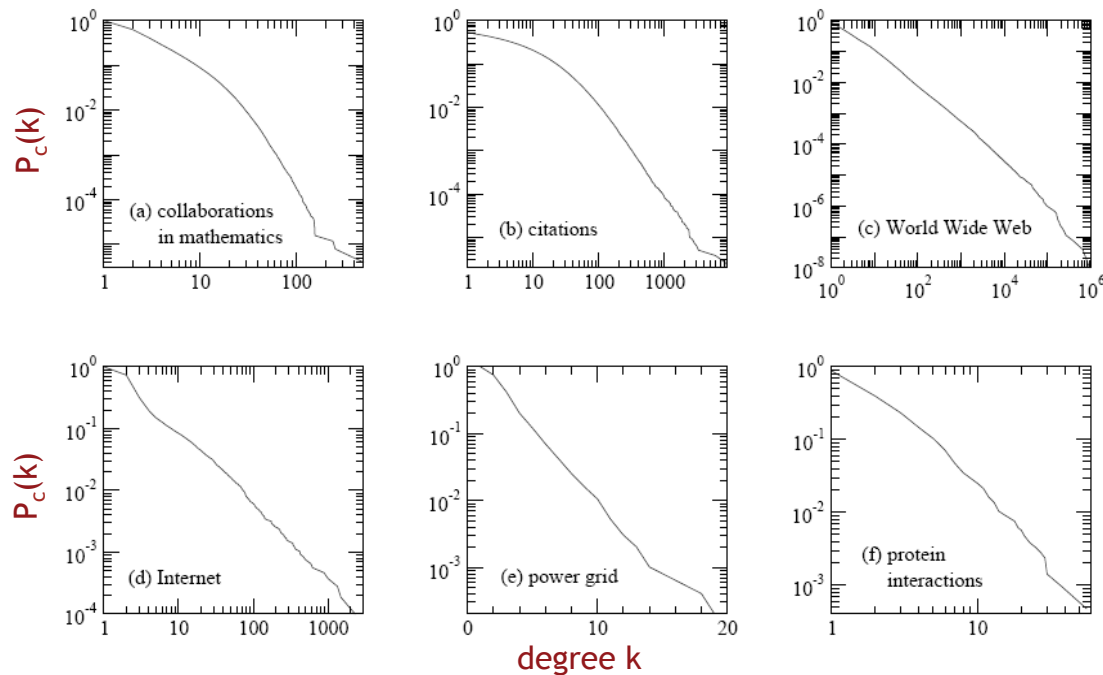
$$M \equiv \sum_{s=1}^{N_M} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right]$$

where N_M is the number of modules, L is the number of links in the network, l_s is the number of links between nodes in module s, and d_s is the sum of the degrees of the nodes in module s.

❑ Degree Distribution (I):

The [cumulative] degree distribution [$P_c(k)$] $p(k)$ accounts for the fraction of nodes in the network with a degree [higher than] equal to k.
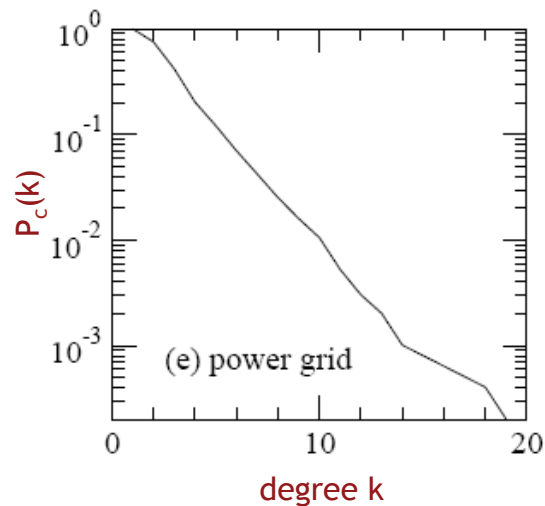
❑ Degree Distribution (II):
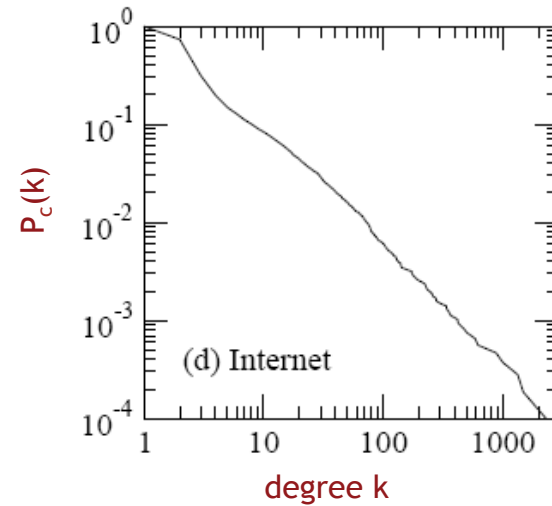
Two types of degree distribution appear more frequently in real networks :

Exponential decay: $P_c(k) \sim e^{-\alpha k}$

Power-law decay: $P_c(k) \sim k^{-\gamma}$

degree k

Typical in random networks

degree k

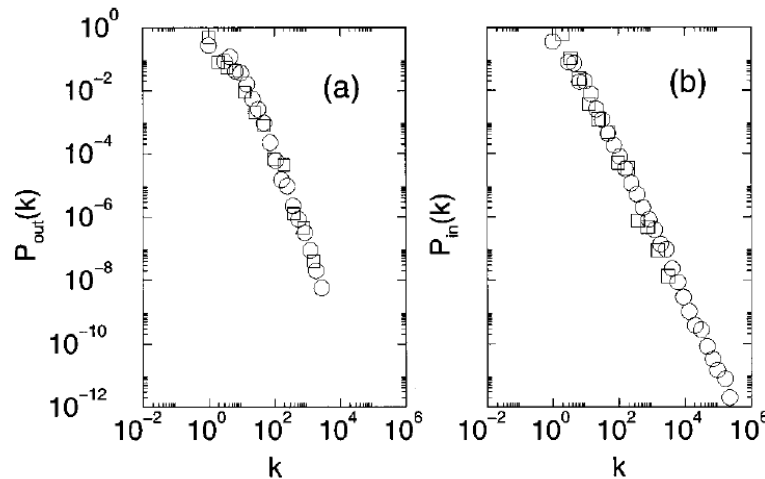Networks with power-law decay are called scale-free networks.

What is the relation between degree and strength?

❑ **Degree Distribution (III):**

Other related distributions are:

In/out degree distributions
(directed networks)

Strength distribution
(weighted networks)

In/out degree distributions of WWW (from two different samples: 325.729 and 200.000.000 nodes). From R. Albert et al., Rev. Mod. Phys. 74, 47 (2002).
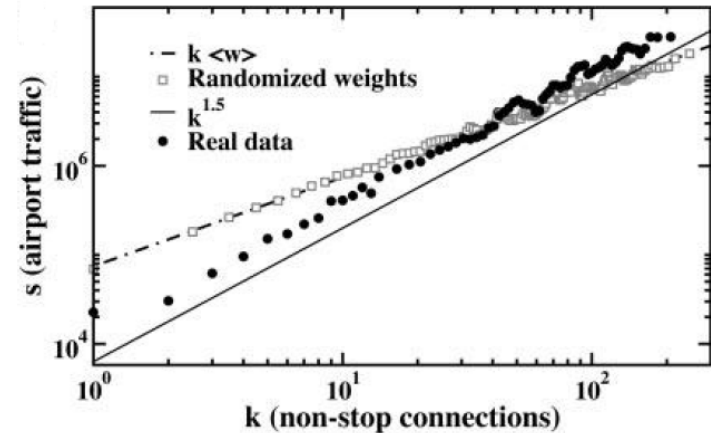
Strength distribution of the International Air Transportation Network (www.iata.org). From A. Barrat et al., PNAS, 101, 3747 (2004).

❑ **Clustering Distribution C(k):**

The clustering distribution has been related with the modularity and hierarchy of the network:

Figure: Clustering distribution in three organisms: Aquidex aeolicus (archaea) (C), Escherichia coli (bacterium) (D), and Saccharomyces cerevisiae (eukaryote) (E). (F) The C(k) curves averaged over all 43 organisms is shown, and the inset displays all 43 species together. Lines correspond to $C(k) \sim k^{-1}$, and diamonds represent the C(k) value expected for an equivalent scale-free network, indicating the absence of scaling
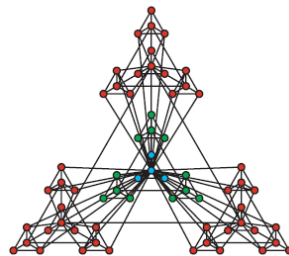
From E. Ravasz et al., Science, 297, 1551 (2002).
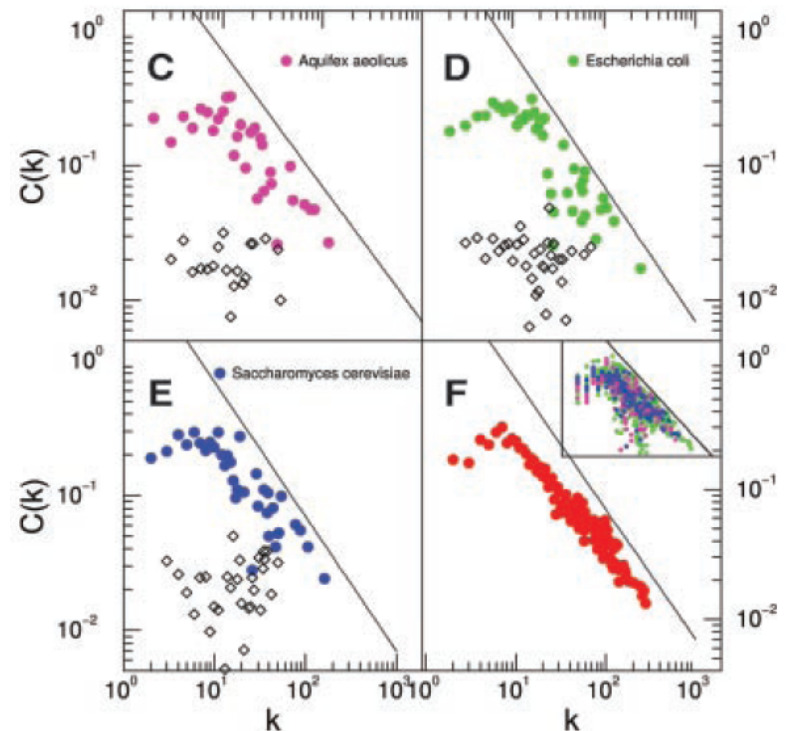
❑ **Nearest neighbor degree $k_{nn}(k)$ and assortativity**

The $k_{nn}(k)$ distribution measures the degree of the nearest neighbors. It is an indicator of the **assortativity** of the network.
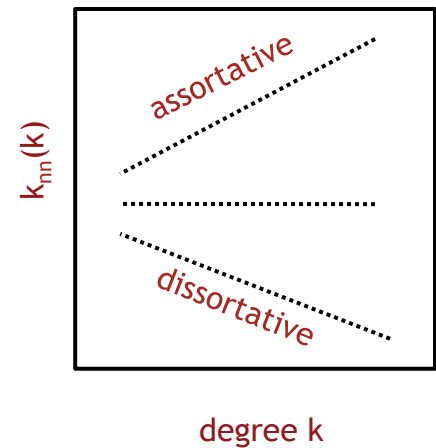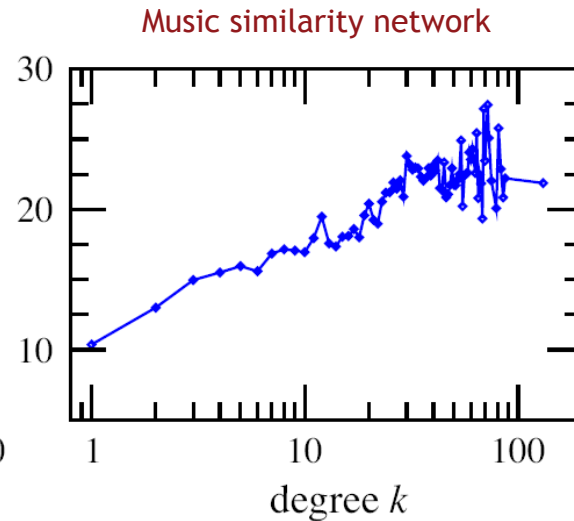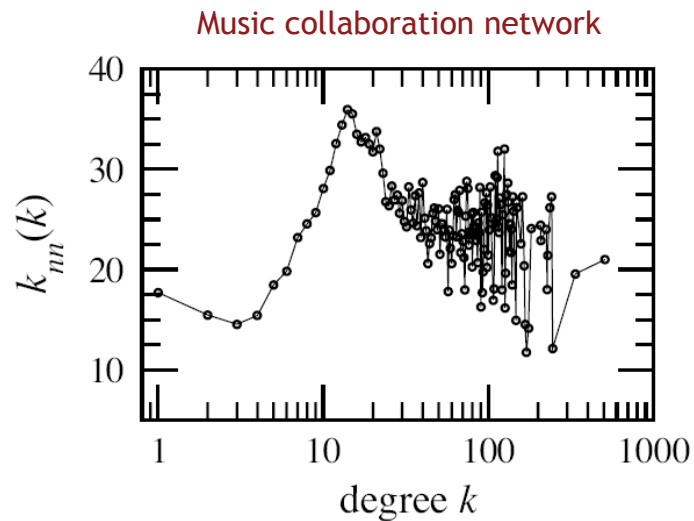
Colaboration and similarity network obtained from a music database (AllMusic Guide).
From J. Park et al., IJBC, 17, 2281 (2007).

1.4.- Brief historical background

1.4.- BRIEF HISTORICAL BRACKGROUND

❏ **Leonard Euler (Basel 1707 – St. Petersburg 1783)**

Some revealing data about Leo:

❏ Euler worked in almost all areas of mathematics: geometry, calculus, trigonometry, algebra, and number theory, as well as continuum physics, lunar theory and other areas of physics.

❏ Large number of topics of physics and mathematics are named in his honour (e.g., Eulers's function, Euler's Equation or Euler's formula).

❏ All his work is collected in *Opera Omnia*, which consists of 886 books.

❏ With one eye from 1738 and completely blind from 1766!

❏ And the most atonishing data: all of that with **13 children!**
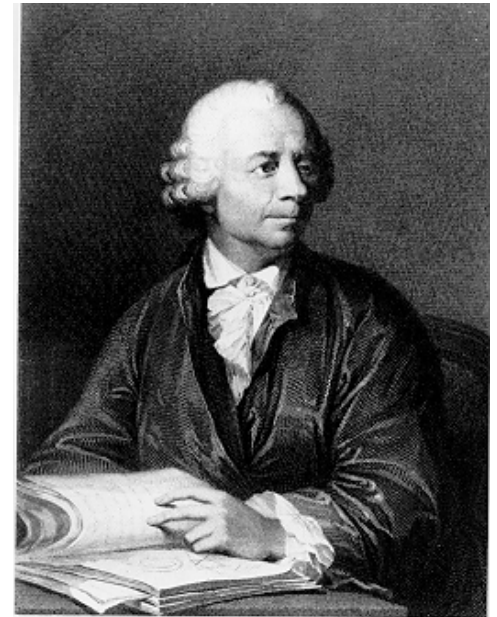
❑ Euler, the father of graph theory:

The seven bridges of Konïgsberg and the origin of graph theory:
Is it possible to cross the seven bridges only once?
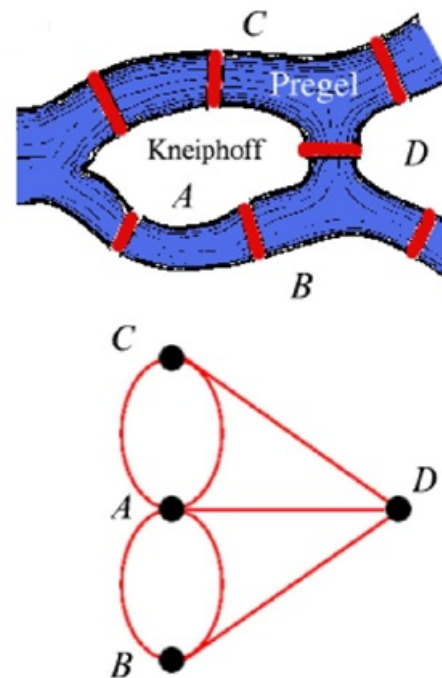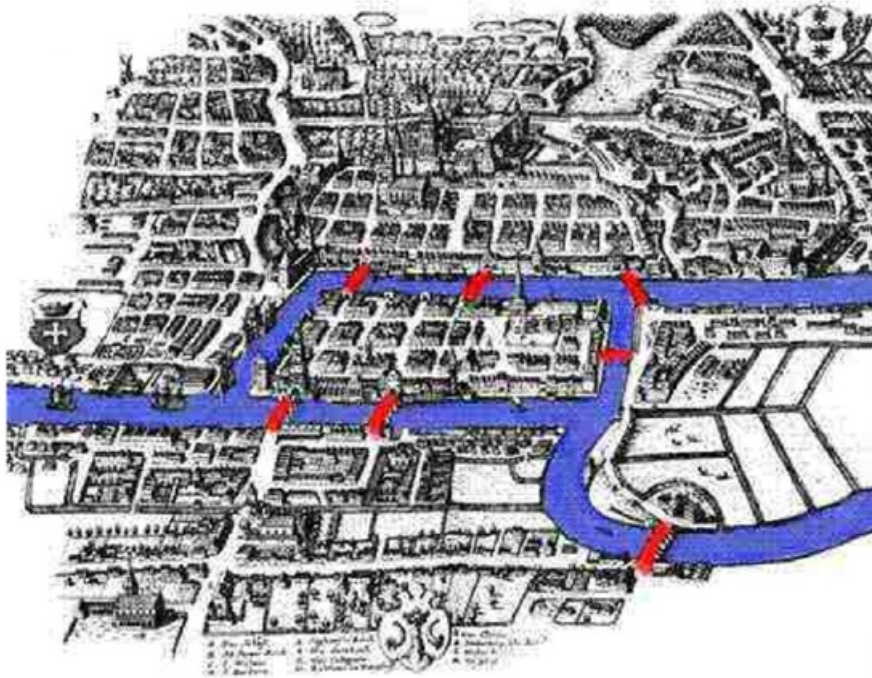
Euler's Solution:

N_0 = Number of nodes with odd degree

1.- If $N_0 > 2$, no solution.

2.- If $N_0 = 2$, only one solution starting from one of the odd nodes.

1.- If $N_0 < 2$, there are solutions starting from any node.

1.4.- BRIEF HISTORICAL BRACKGROUND
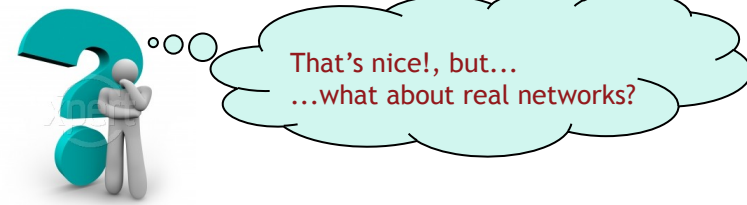
❑ Regular Graphs

❑ After the death of Euler, graph theory received many contributions from mathematicians such as Hamilton, Kirchhoff or Cayley.

❑ The core of graph theory focused on the study of regular graphs:

Regular graph: a graph where all nodes have the same degree.

Lattice: a regular network where all nodes are coupled to their nearest neighbors.

N = number of nodes
K = degree
C = clustering coefficient
d = dimension of the lattice
l = average path length

$$C = \frac{3(K - 2d)}{4(K - d)}$$

(if $K < 2N/3$)

$$\ell \sim \sqrt[d]{\frac{N}{K}}$$
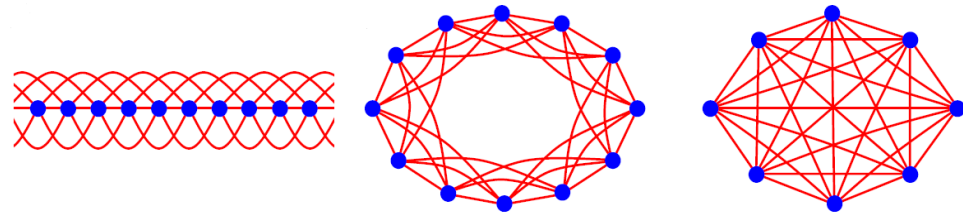
❑ **Paul Erdös (Budapest 1913 – Warsaw 1996)**

Some revealing data about Paul:

❑ Seminal contributions in combinatorics, graph theory, number theory, classical analysis, approximation theory, set theory, and probability theory.

❑ Paul wrote 1475 papers and collaborated with 511 scientists.

❑ Excentric person, he had an special vocabulary (children="epsilons", women="bosses", U.S.="samland", etc...)

❑ Paul offered small prizes for solutions to unresolved problems (from 25$ to some thousands), and there are still open problems!

❑ "You don't have to believe in God, but you should believe in The Book." (he recognized that he took amphetamines)
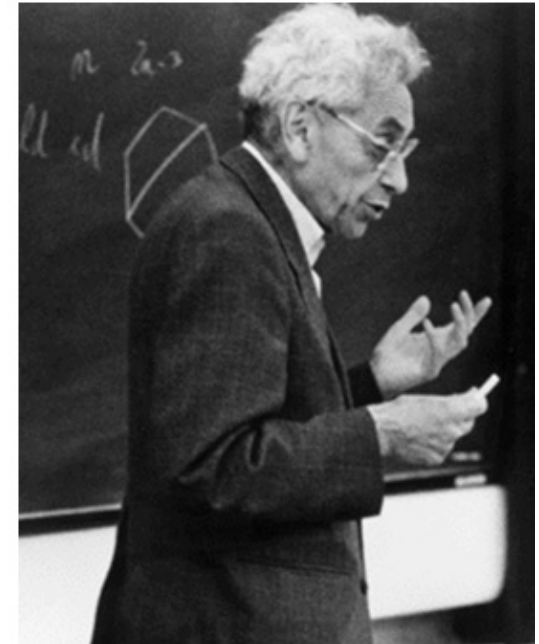
❑ Paul Erdös and Alfred Rényi

They worked on the analysis of social networks by finding analogies with the so-called *random graphs*, in which the existence of a link between a pair of nodes has a probability *p*.

p
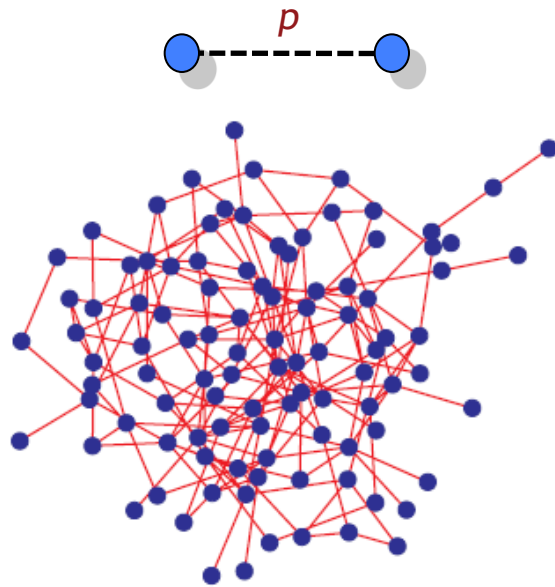
N = number of nodes
$<k>$ = mean degree
$<L>$ = number of random connections
p = probability of connection between two nodes

Mean degree of the network → $<k> = p(N-1) \cong pN$

Number of random connections → $<L> = ½ pN(N-1) \cong ½ <k>N$

1.4.- BRIEF HISTORICAL BRACKGROUND

❑ Emergence of a giant component

When propability p crosses a critical value p_c, there emerge a giant component that contains and extensive fraction of the nodes in the network

p increases ⟶
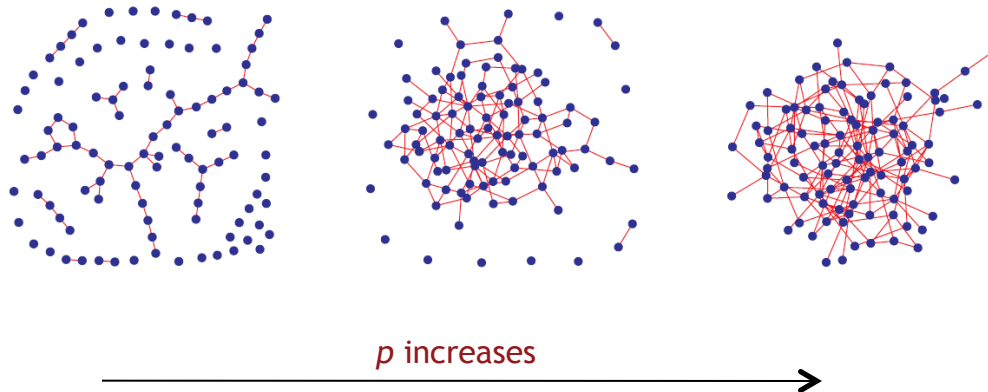
Critical probability (N→ ∞):

$$p_c \sim \frac{\ln N}{N}$$

Critical mean degree:

$$\langle k \rangle_c \sim \ln N$$

Clustering coefficient:

$$\overline{C} = p \cong \frac{\langle k \rangle}{N} \ll 1$$

N=1000
<k>=2
C ~ 0.002

Average shortest path:

$$\ell \sim \frac{\ln N}{\ln \langle k \rangle}$$

N=1000000
<k>=5
l ~ 8.6

centro de
tecnología
biomédica

CSG

Universidad
Rey Juan Carlos

1.4.- BRIEF HISTORICAL BRACKGROUND

☐ **Stanley Milgram (New York 1933 – New York 1984)**

Stanley Milgram was an American social psychologist most notable for his controversial studies on the obedience to authority.

Some Stanley's famous experiments:

☐ The Milgram experiment 18

☐ The lost-letter experiment

☐ The small-world experiment

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

1.4.- BRIEF HISTORICAL BRACKGROUND

❑ The small-world experiment

A group of people from Omaha (Nebraska) and Wichita (Kansas) was asked to send a letter to an unknown person in Boston (Massachussetts).

Basic Rule of the experiment:

❑ People should forward the letter to a person that they consider closer to the target person

Results of one experiment (in fact, there where several!):

❑ 232 out of 296 letters never reached the target

❑ 64 letters reached the target (with paths from 2 to 10)

❑ The average path length was …. 5.2 (steps)

CTB centro de tecnología biomédica   CSG   Universidad Rey Juan Carlos
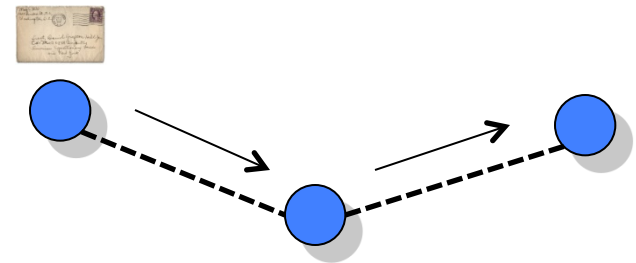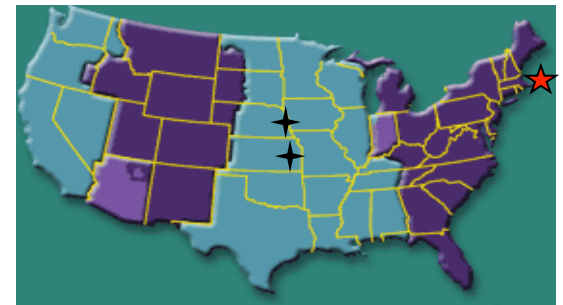
❑ **It's a small world!**

This is a big world

This is a small world

or in other words:

$$d_{ij} << N$$
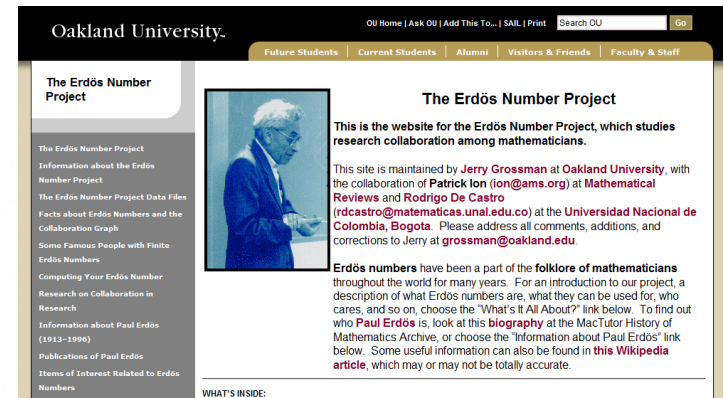
1.4.- BRIEF HISTORICAL BRACKGROUND

❑ **Let's go back to Erdös:**

You can measure the distance with Paul Erdös.
(http://www.oakland.edu/enp/)

❑ Mean Erdös number: ~5

❑ Largest Erdös number: ~13

Max von Laue	1914	4
Albert Einstein	1921	2
Niels Bohr	1922	5
Louis de Broglie	1929	5
Werner Heisenberg	1932	4
Paul A. Dirac	1933	4
Erwin Schrödinger	1933	8
Enrico Fermi	1938	3
Ernest O. Lawrence	1939	6
Otto Stern	1943	3
Isidor I. Rabi	1944	4
Wolfgang Pauli	1945	3
Frits Zernike	1953	6
Max Born	1954	3
Willis E. Lamb	1955	3

Are real networks a mixture between lattices and random networks?
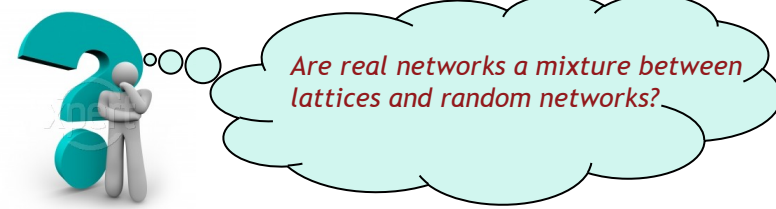
❑ **It's a small world everywhere!**

The small-world property has been reported in a large number of real networks of different origin.

NETWORK	SIZE	$\langle k \rangle$	ℓ	ℓ_{rnd}	C	C_{rnd}
1. Movie actors	225 226	61.0	3.65	2.99	0.79	0.00027
2. Power grid	4 941	2.67	18.7	12.4	0.08	0.00054
3. WWW site level (undir.)	153 127	35.2	3.10	3.35	0.11	0.00023
4. Words (co–ocurrence)	460 902	70.1	2.67	3.03	0.44	0.00015
5. LANL co–authorship	52 909	9.70	5.90	4.79	0.43	0.00018
6. MEDLINE co–authorship	1 520 251	18.1	4.60	4.91	0.07	0.00001
7. Math. co–authorship	70 975	3.90	9.50	8.21	0.59	0.00005

Average path length and clustering coefficient of some real networks. We compare the values in the real network with those of equivalent random networks

The average path length is similar in random networks (where $l \sim \ln N$) but the clustering coefficient is some orders of magnitude higher (and closer to the clustering coefficient of a lattice!).

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

❑ The Watts-Strogatz model (I)

Watts and Strogatz (PRL 1998) proposed a network model that conciliated the high clustering and short average path length of real networks

Starting from a regular ring, a certain (random) rewiring is introduced with a probability p

p increases
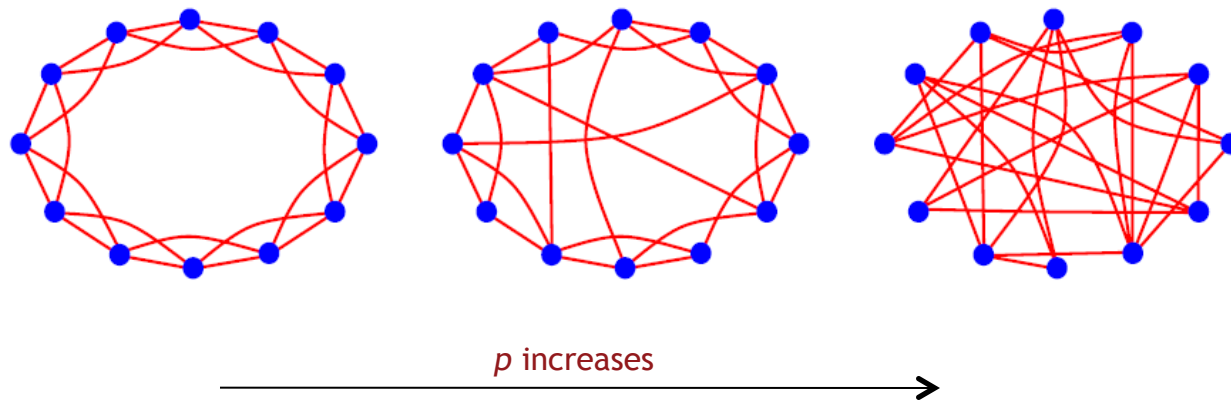
❑ The Watts-Strogatz model (II)

Small-world networks are characterized by a low average shortest path and high clustering
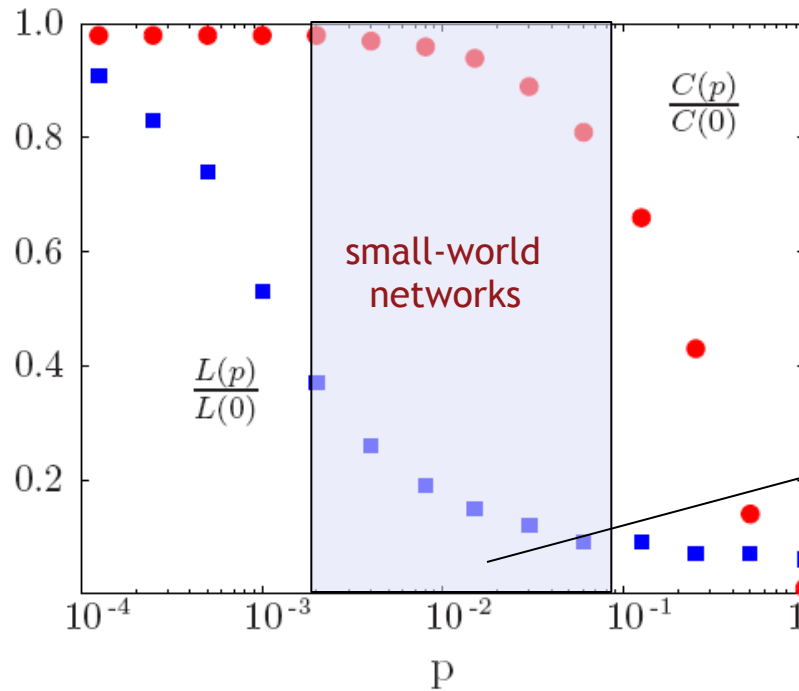
A low number of "shortcuts" reduces the distance between nodes without modifying the local properties

❑ The Watts-Strogatz model (II)

The larger the network, the higher probability to be small-wolrd.
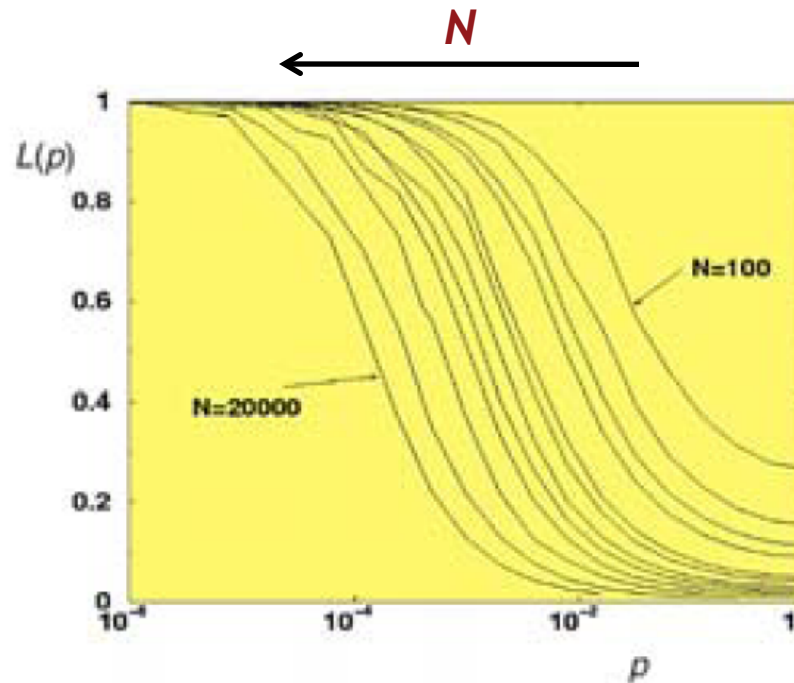
The rewiring of the links in order to enter the small world-region goes with:

$$p \sim 1/N$$

❏ Scale-free networks (I)

Unfortunatelly (or luckily!) many real networks are not exponential. On the contrary, they have a power-law decay (i.e., $P(k) \sim k^{-\gamma}$).
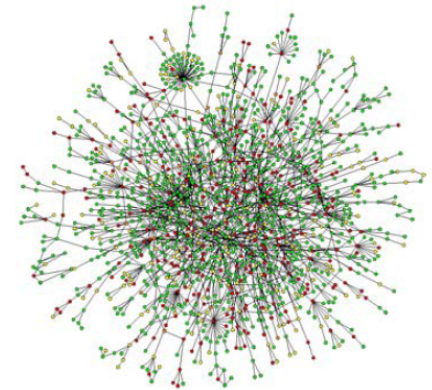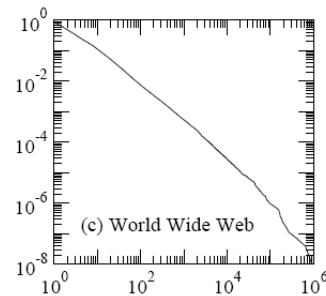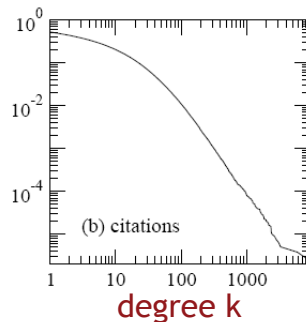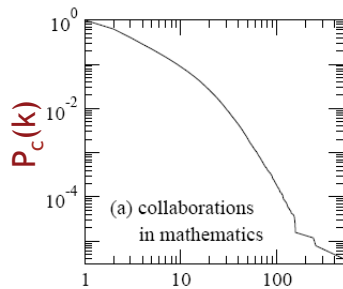
❏ Scale-free networks have power law decays $P(k) \sim k^{-\gamma}$

❏ Power laws are relatively slow decreasing functions (the probability of having highly connected nodes is much higher than in exponential networks).

❏ A power-law distribution has no peak at its average value (no characteristic scale).

❑ Scale-free networks (II)

NETWORK	SIZE	γ_{in}/γ_{out}
1. Movie actors [57]	212 250	2.3
2. WWW [59]	$2 \cdot 10^8$	2.7/2.1
3. Internet, router [60]	260 000	—/1.94
4. Words (co–ocurrence) [13]	460 902	2.7
5. Neuro. co–authorship [61]	209 293	2.1
6. SPIRES co–authorship [48]	56 627	1.2
7. E-mail messages [62]	59 912	1.5/2.0
8. Metabollic network [63]	778	2.2

Real networks with scale-free structure. From Almendral, PhD. Thesis

Interestingly, the exponent of the power laws range from 1.2 to 3, with the majority between 2 and 3.

centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

❑ The Barabási-Albert model (I)

They introduce a model in order to explain the origin of the power-law distributions of real networks. A network is constructed from scratch following two fundamental rules:

❑ *Growth. From an initial number of nodes N_0, new nodes are attached to the existing ones at discrete time steps. Thus, the number of nodes increases with time $N(t)= N_0+t$ and also the number of links $L(t)= mt$ (being m the number of links of each new node)*

❑ *Preferential attachment. The nodes to which the new node is attached are chosen following a preference function:*

$$p_i = \frac{k_i}{\sum_{j=1}^{N(t)} k_j}$$

CTB centro de tecnología biomédica   CSG   Universidad Rey Juan Carlos

❑ **The Barabási-Albert model (II)**

The BA model shows a power law decay independent of the number of links or the system size (with an exponent $\gamma=3$)

HUBS!

Evolution of the degree of two nodes
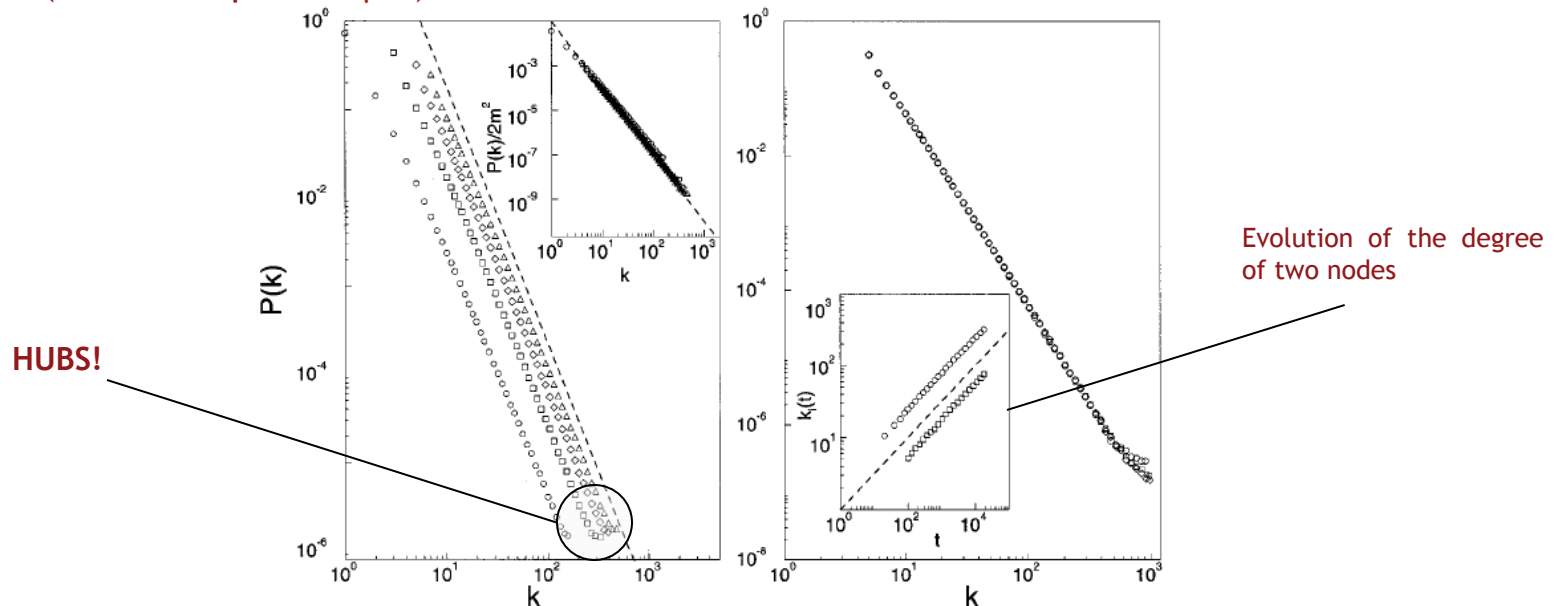
(Left) Degree distribution of the B-A model, with $N=m_0+t=300000$ and $m_0=1,3,5,7$. The dashed lines correspond to $P(k)=k^{-2.9}$. (Right) $P(k)$ for $m_0=5$ and different systems size: $m=100000$, 150000 and 200000. From R. Albert et al., Rev. Mod. Phys. 74, 47(2002).

Are networks obtained with the BA model small world networks?

☐ The Barabási-Albert model (III)

As in random networks, the clustering coefficient obtained with the BA model is low

Clustering coefficient C of the network as a function of the system size N. From R. Albert et al., Rev. Mod. Phys. 74, 47(2002)

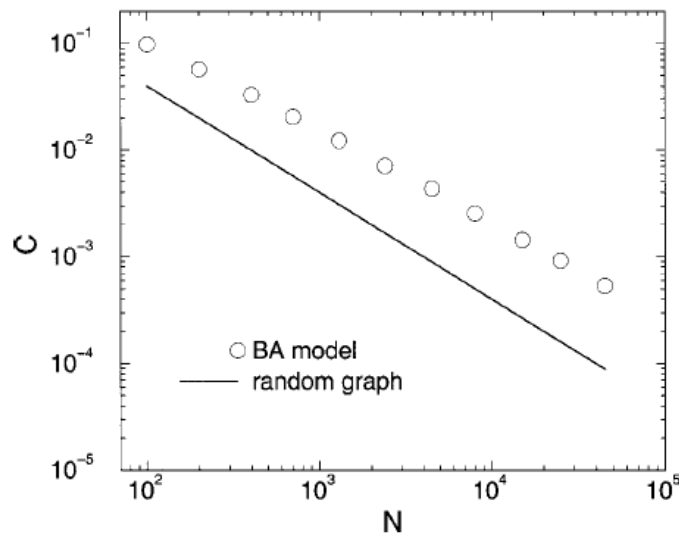|	network	type	n	m	C	ℓ
social	film actors	undirected	449 913	25 516 482	0.78	3.48
	company directors	undirected	7 673	55 392	0.88	4.60
	math coauthorship	undirected	253 339	496 489	0.34	7.57
	physics coauthorship	undirected	52 909	245 300	0.56	6.19
	biology coauthorship	undirected	1 520 251	11 803 064	0.60	4.92
	telephone call graph	undirected	47 000 000	80 000 000		
	email messages	directed	59 912	86 300	0.16	4.95
	email address books	directed	16 881	57 029	0.13	5.22
	student relationships	undirected	573	477	0.001	16.01
	sexual contacts	undirected	2 810			
technological	Internet	undirected	10 697	31 992	0.39	3.31
	power grid	undirected	4 941	6 594	0.080	18.99
	train routes	undirected	587	19 603	0.69	2.16
	software packages	directed	1 439	1 723	0.082	2.42
	software classes	directed	1 377	2 213	0.012	1.51
	electronic circuits	undirected	24 097	53 248	0.030	11.05
	peer-to-peer network	undirected	880	1 296	0.011	4.28
biological	metabolic network	undirected	765	3 686	0.67	2.56
	protein interactions	undirected	2 115	2 240	0.071	6.80
	marine food web	directed	135	598	0.23	2.05
	freshwater food web	directed	92	997	0.087	1.90

Clustering coefficient C and average path length of some real networks. From Newman, SIAM Rev, 45, 167 (2003)

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

❑ **The Barabási-Albert model (IV)**

Attractiveness, aging, capacity, … can modify the scale free behaviour of the BA model.

The Dorogovtsev-Mendes -Samukhin model

$$\prod_{j \to i} = \frac{k_i + k_0}{\sum_l (k_l + k_0)}$$

k_0= initial attractiveness
($-m < k_0 < \infty$)
m= number of new links

$$\gamma = 3 + k_0/m$$

($2 < \gamma < \infty$)

Dorogovtsev et al.,
PRL 85 4633 (2000)

The Kaprivsky et al. model

$$\prod_{j \to i} = \frac{k_i^\alpha}{\sum_l k_l^\alpha}$$

$\alpha < 1$: streched exponential decay
$\alpha > 1$: a single node dominates

Krapivsky et al.,
PRL, 4629 85 (2000)
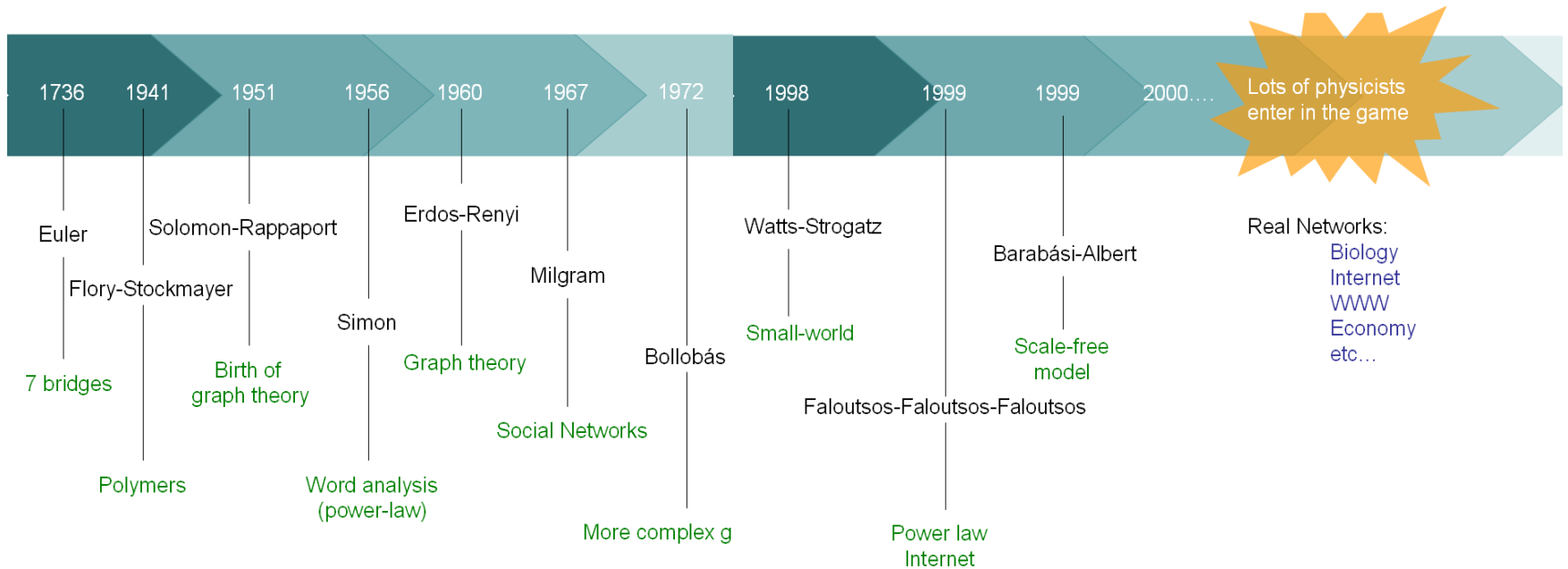
The Dorogovtsev-Mendes model

Probability of linking depends on $\tau^{-\alpha}$
(being τ the age of the node)
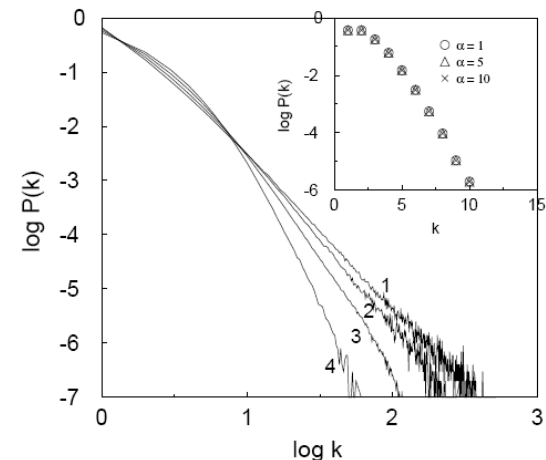
Probability distribution for serveral aging exponents:
1) 0.2, 2) 0.25, 3) 0.5 and 4) 0.75. α>1 exponential decay.
From PRE62, 1842 (2000)

❑ **Complex Networks time line:**

| 1736 | 1941 | 1951 | 1956 | 1960 | 1967 | 1972 | | 1998 | 1999 | 1999 | 2000…. | Lots of physicists enter in the game |

Euler

Flory-Stockmayer

Solomon-Rappaport

Simon

Erdos-Renyi

Milgram

Bollobás

Watts-Strogatz

Barabási-Albert

Real Networks:
Biology
Internet
WWW
Economy
etc…

7 bridges

Birth of graph theory

Graph theory

Small-world

Scale-free model

Polymers

Word analysis (power-law)

Social Networks

Faloutsos-Faloutsos-Faloutsos

More complex g

Power law Internet

(from J.F.F. Mendes presentation)

2.- Applications To Biological Networks

2.1.- Genetic, protein and metabolic networks
2.2.- Neuron and brain networks
2.3.- Networks in ecology
2.4.- Disease spreading

❑ **Biological Networks:**

	network	type	n	m	z	ℓ	α	C	r
biological	metabolic network	undirected	765	3 686	9.64	2.56	2.2	0.67	-0.240
	protein interactions	undirected	2 115	2 240	2.12	6.80	2.4	0.071	-0.156
	marine food web	directed	135	598	4.43	2.05	$-$	0.23	-0.263
	freshwater food web	directed	92	997	10.84	1.90	$-$	0.087	-0.326
	neural network	directed	307	2 359	7.68	3.97	$-$	0.28	-0.226

Network parameters of several biological networks: n, number of nodes; m, number of links; z, mean degree,; l average shortest path; α, power-law exponent; C, clustering coefficient, and r, assortativity. From Newman, SIAM, 45, 167 (2003).

❑ **How are Biological Networks?:**

  ❑ Biological networks are small-world.

  ❑ It is common to observe dissortative mixing (i.e., most connected nodes are not preferentially connected with each other).

  ❑ They are (typically) organized in sub-modules and, as a consequence, they have high modularity and community structures.

Nevertheless, each network deserves its own interpretation

2.0 Introduction to Biological Networks

❑ **Complex networks in biology:**

One of the first contributions of the Complex Network Theory to biological systems is the seminal paper of Watts and Strogatz

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

The small-world of *C. Elegans* neural network , with an edge joining two neurons if they are connected by either a synapse or a gap junction (n= 282, <k>= 14.). Table from Watts et al., 393, 440 (1998)

❑ **Biological networks comprise a wide set of different networks:**

    ❑ **Metabolic, protein and genetic networks**

    ❑ **Networks of neurons**

    ❑ **Functional and anatomical brain networks**

    ❑ **Food webs in ecosystems**

    ❑ **Animal grouping and swarm movement**

    ❑ **and many others …**

2.1.- Genetic, protein and metabolic networks

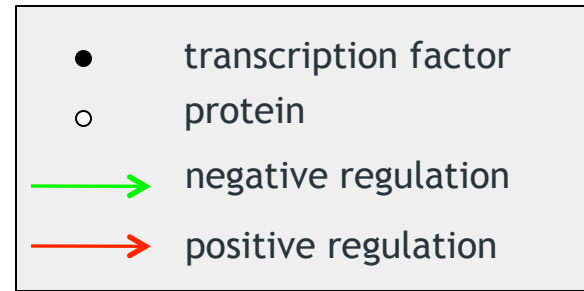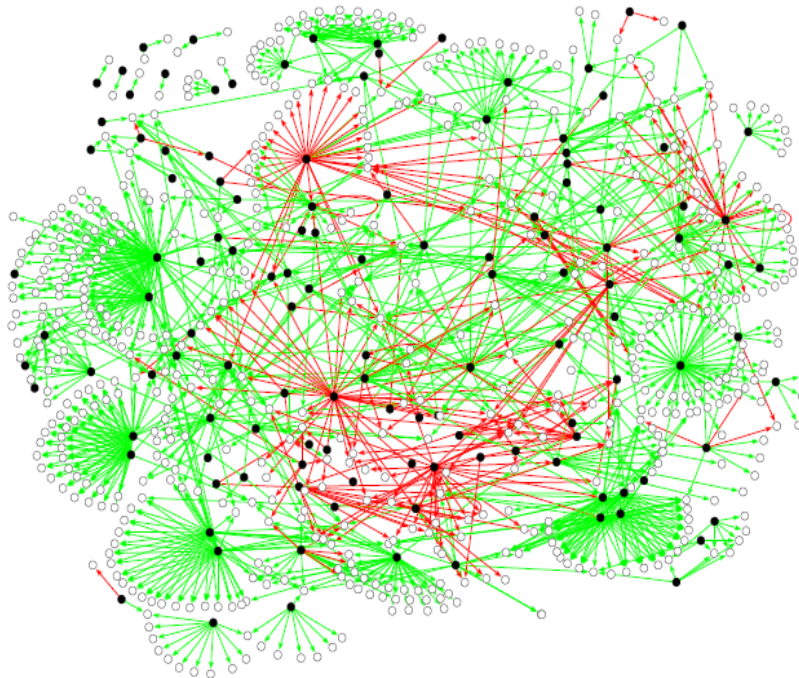# 2.1.- Genetic, protein and metabolic networks

❑ Genetic networks:

Interaction between genes (through transcription factors) lead to a nework of promotor/repressor interactions
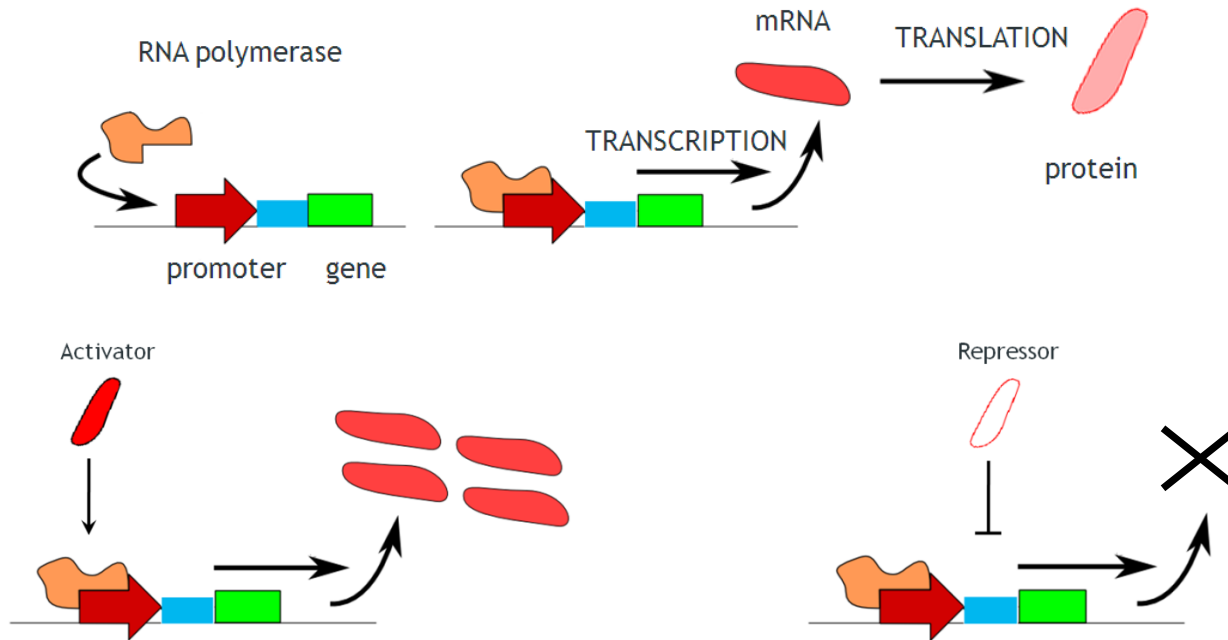
❑ Genetic networks:

Again, genetic transcription networks are directed (digraphs) with positive/negative regulations:

Legend:
- ● transcription factor
- ○ protein
- → negative regulation
- → positive regulation

Yeast (S. Cerevisiae) network of transcriptional regulation (N=682 proteins and M=1289 interactions). From Maslov et al., *Large-Scale Topological Properties of Molecular Networks* (Springer 2003)
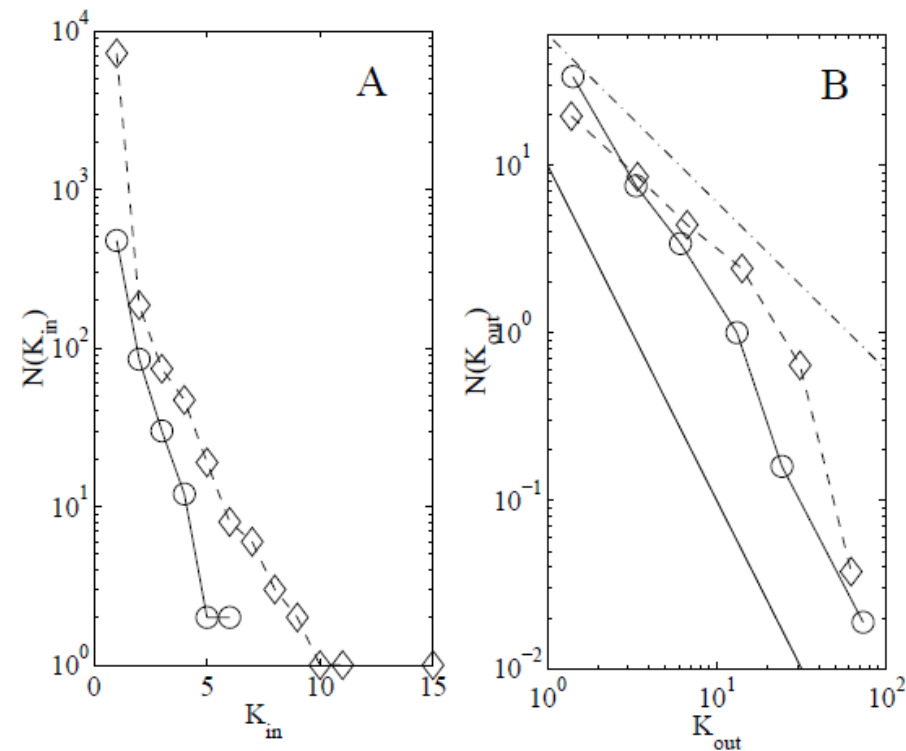
2.1.- Genetic, protein and metabolic networks

❑ **Genetic networks:**

The $P_{in}(k)$ distribution is limited by the system (due to the finite space of the promoter). $P_{out}(k)$ is not limited and, as a consequence, has a heavy tail.

Figure: (a) The histogram $N(K_{in})$ of nodes' in-degrees K_{in} in transcription regulatory networks of yeast (diamonds, dashed line), and E. coli (circles, solid line). (b) the same as (a) but considering the $N(K_{out})$. From Maslov et al., (2003).

2.1.- Genetic, protein and metabolic networks

❏ **Metabolic networks:**

Metabolic networks are obtained from the biochemical reactions involving the transformation of energy and matter in the cell. The participating substrates are called metabolites and are catalyzed and regulated by enzymes.
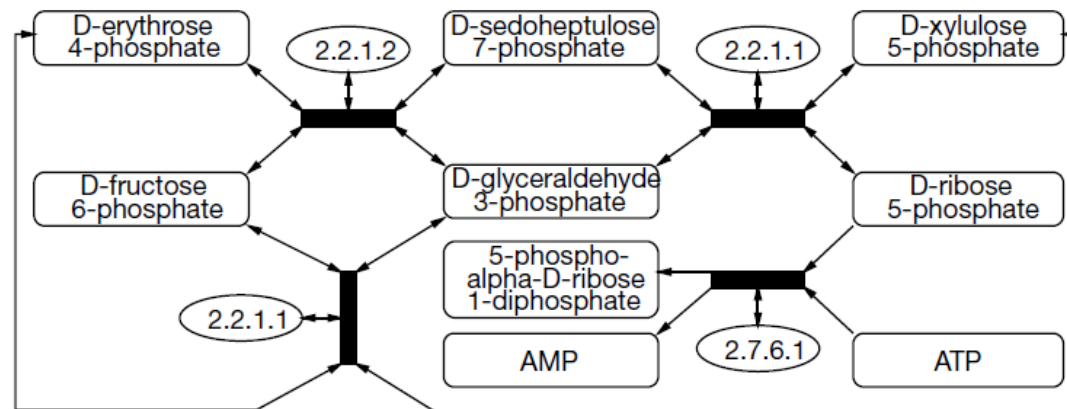

A portion of the WIT database for E. coli. Each substrate can be represented as a node of the graph, linked through temporary educt-educt complexes (black boxes) from which the products emerge as new nodes (substrates). The enzymes, which provide the catalytic scaffolds for the reactions, are shown by their EC numbers. From Jeong et al., Nature, 407.651 (2000).

❑ **Metabolic networks:**

Metabolic networks have scale-free degree distribution.
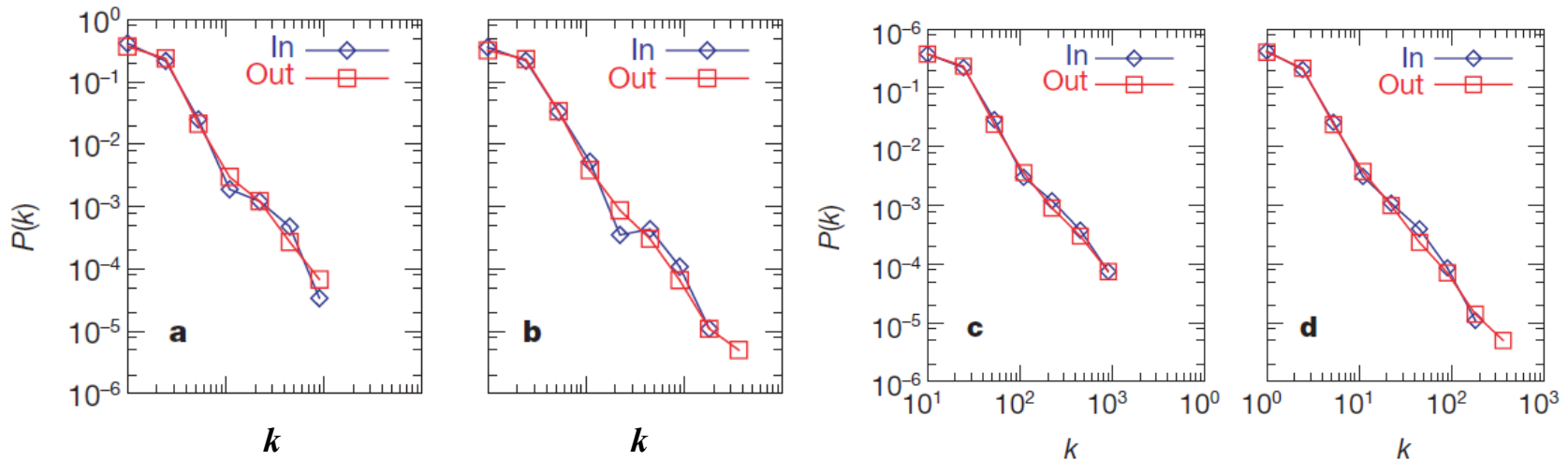
Connectivity distributions P(k) for: (a) Archaeoglobus fulgidus (archae); (b) E. coli (bacterium); (c) Caenorhabditis elegans (eukaryote), counting separately the incoming (In) and outgoing links (Out) for each substrate. k_{in} (k_{out}) corresponds to the number of reactions in which a substrate participates as a product (educt). (d) The connectivity distribution averaged over all 43 organisms. From Jeong et al., Nature, 407.651 (2000).
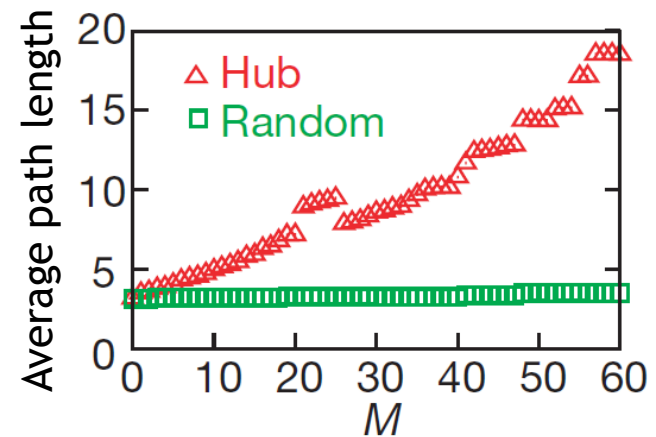
☐ **Metabolic networks:**

They also have the small-world property and the resilience to failures of scale-free networks:
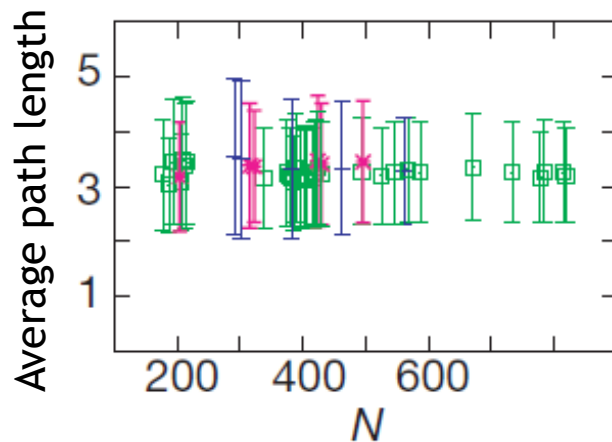
Average path length of the metabolic network of 43 organisms. From Jeong et al., Nature, 407, 651 (2000).

The effect of substrate removal on the metabolic network of *E. coli*. $M=60$ corresponds to the ~8% of the network metabolites. From Jeong et al., Nature, 407, 651 (2000).

❑ **Protein networks:**

They reflect physical or chemical interactions between proteins. It is estimated that even simple single-celled organisms such as yeast have their roughly 6000 proteins interacting by at least 3 interactions per protein, i.e. a total of 20,000 interactions or more. By extrapolation, there may be on the order of ~100,000 interactions in the human body.
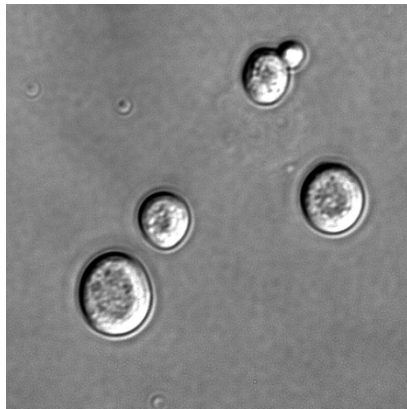

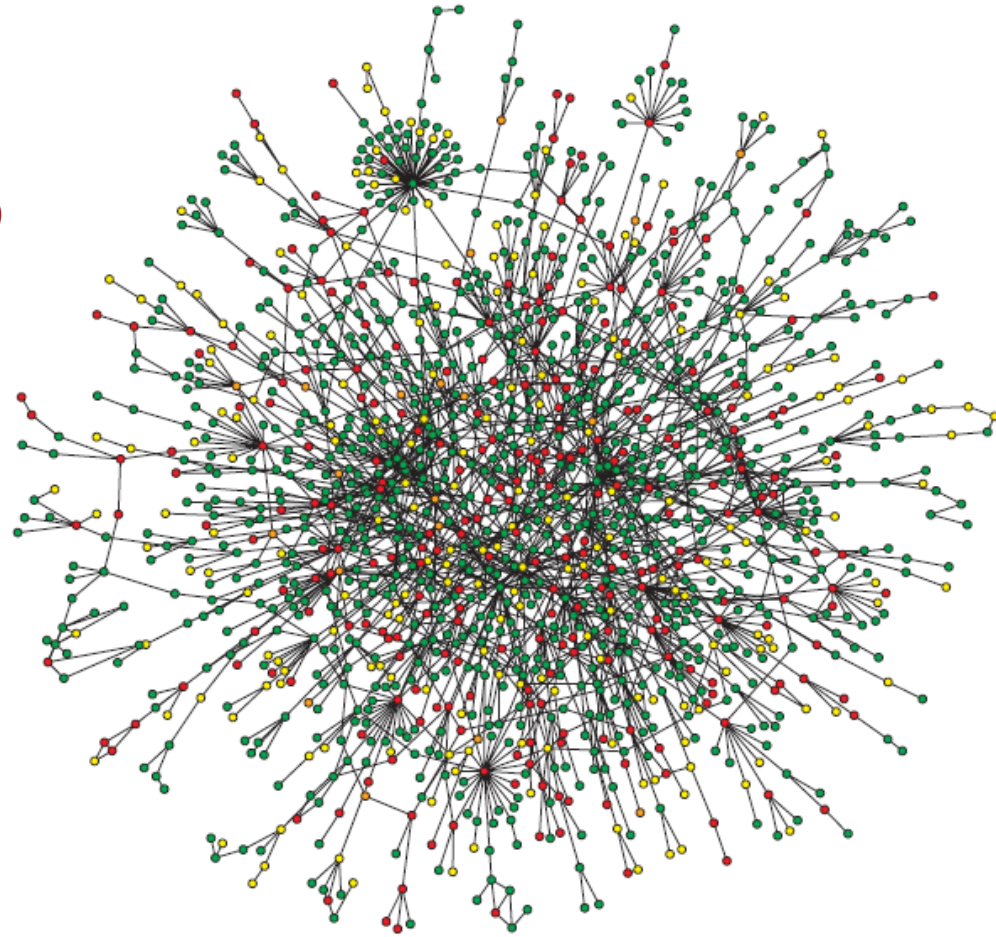Figure from Thanos, et al., Science, 283, 833 (1999)

2.1.- Genetic, protein and metabolic networks

❏ **Protein networks:**

The protein-protein (bidirectional) interactions are an example of a complex network.

Protein-protein interaction in the yeast *S. cerevisiae*, (N=1870 and M=2240). From Jeong et al., Nature, 411, 41 (2001).
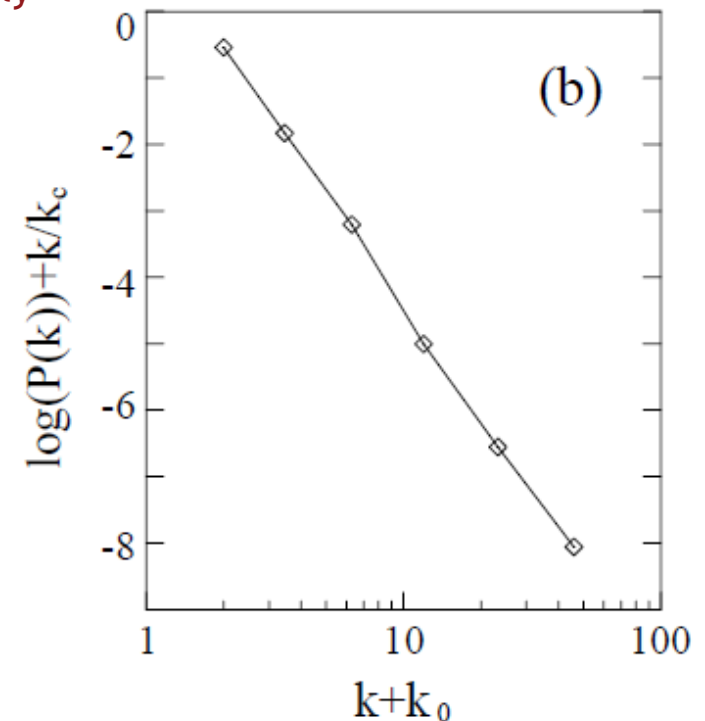
❑ Protein networks:

Protein-protein interaction networks are typically scale-free with an exponential cut-off:

Figure: Probability distribution of the protein-protein interaction in the yeast *S. cerevisiae,* (*N=1870* and *M=2240*). The distribtuion is scale-free with an exponential cut-off (around $k_c \sim 20$). From Jeong et al., Nature, 411, 41 (2001).

Read more at:
Maslov et al.,
Science., 296, 910 (2002).

2.1.- Genetic, protein and metabolic networks

❑ Protein networks:

Dissortative structure has been also reported.

Interestingly, dissortative structures are robust against failures of the hubs due to the reduced propagation to the neighbors.
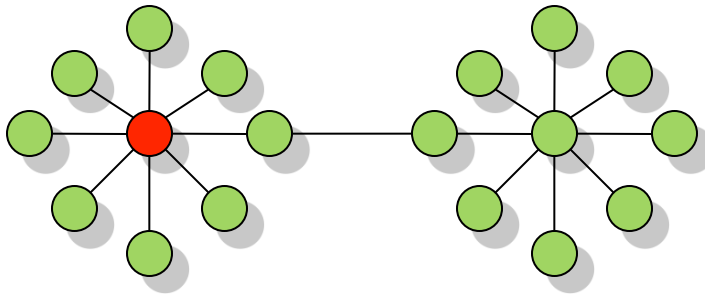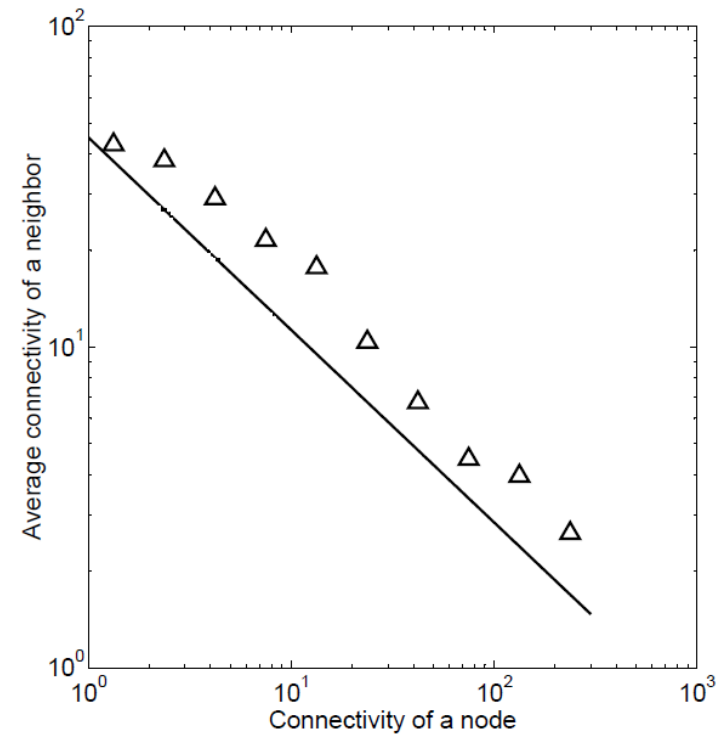

Figure: Distribution of the average neighbor connectivity for the yeast protein-protein interaction network. Here, N=3278 and M= 4549. From Maslov et al., Science., 296, 910 (2002).

2.2.- Neuron and brain networks

❑ **Networks of neurons:**

C. Elegans: It is the only living system that has been fully mapped. It has 302 neurons and average degree <k>≈29.

It has low shortest path and high clustering: it is a small-world network.

Existence of network motifs.

The tail of the distribution of degrees p(k) is power-law.

Gap juntions connections and chemical synapses of C. Elegans neurons. From Varshney, PLoS Comp. Biol, 7, 1001066 (2011)
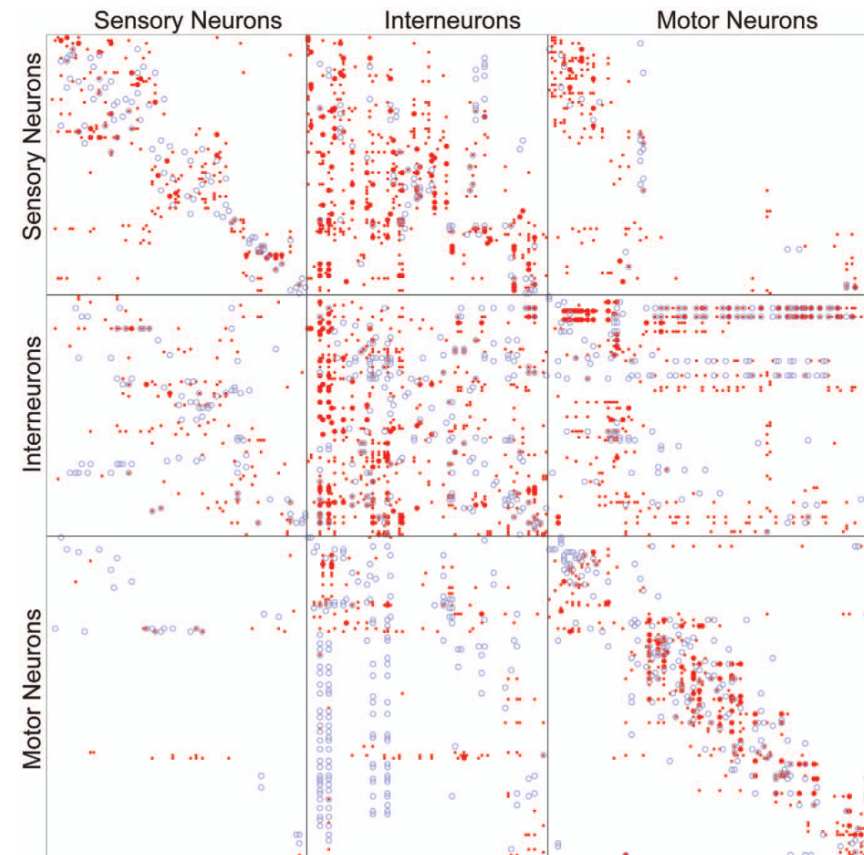
2.2.- Neuron and brain networks

❑ Are anatomical networks efficient in transmitting information?

Nonoptimal Component Placement, but Short Processing Paths, due to Long-Distance Projections in Neural Systems

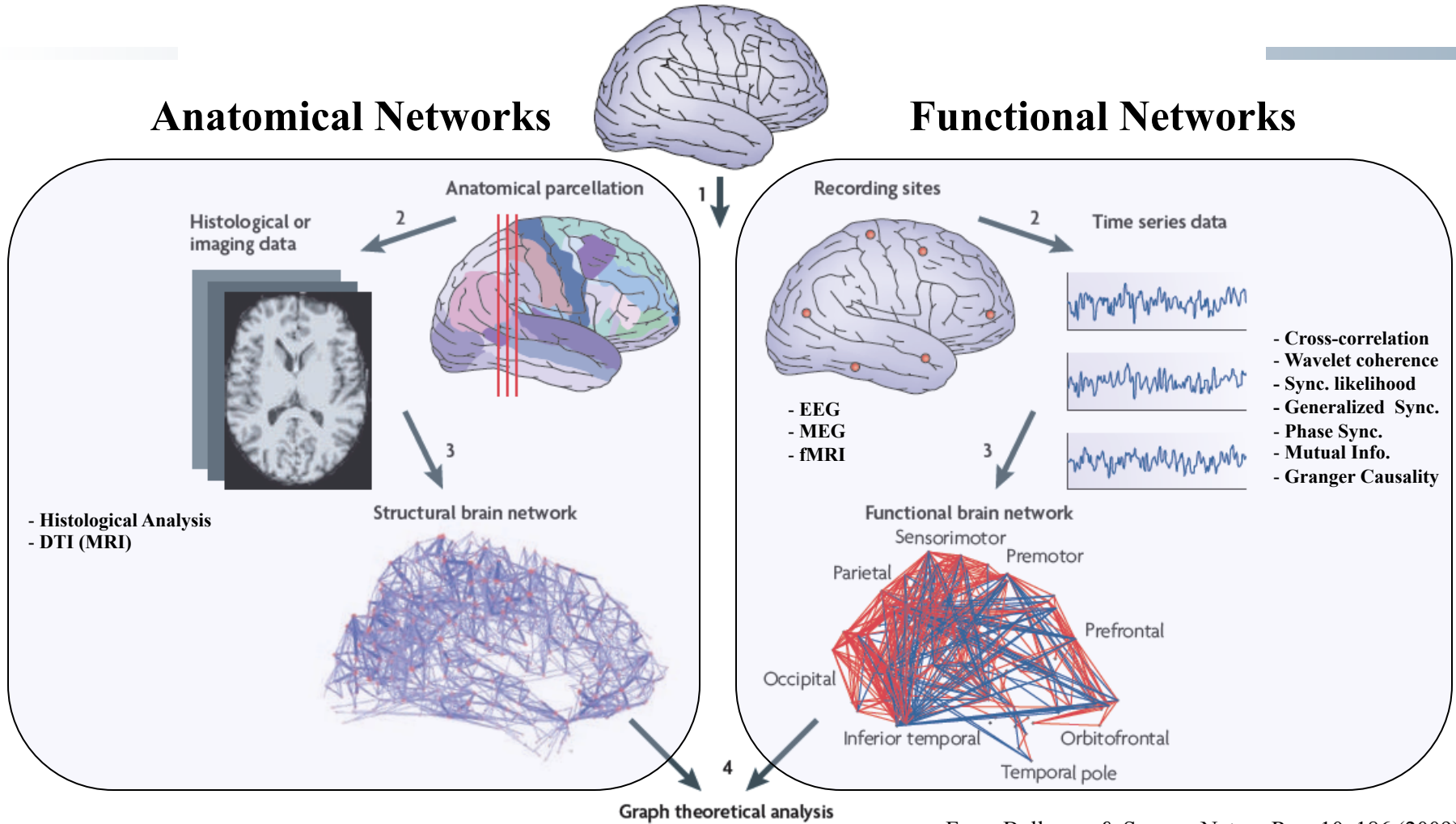Marcus Kaiser[1,2,3*], Claus C. Hilgetag[3,4]

1 School of Computing Science, University of Newcastle, Newcastle upon Tyne, United Kingdom, 2 Institute of Neuroscience, University of Newcastle, Newcastle upon Tyne, United Kingdom, 3 International University Bremen, School of Engineering and Science, Bremen, Germany, 4 Boston University, Sargent College, Department of Health Sciences, Boston, Massachusetts, United States of America

It has been suggested that neural systems across several scales of organization show optimal component placement, in which any spatial rearrangement of the components would lead to an increase of total wiring. Using extensive connectivity datasets for diverse neural networks combined with spatial coordinates for network nodes, we applied an optimization algorithm to the network layouts, in order to search for wire-saving component rearrangements. We found that optimized component rearrangements could substantially reduce total wiring length in all tested neural networks. Specifically, total wiring among 95 primate (Macaque) cortical areas could be decreased by 32%, and wiring of neuronal networks in the nematode Caenorhabditis elegans could be reduced by 48% on the global level, and by 49% for neurons within frontal ganglia. Wiring length reductions were possible due to the existence of long-distance projections in neural networks. We explored the role of these projections by comparing the original networks with minimally rewired networks of the same size, which possessed only the shortest possible connections. In the minimally rewired networks, the number of processing steps along the shortest paths between components was significantly increased compared to the original networks. Additional benchmark comparisons also indicated that neural networks are more similar to network layouts that minimize the length of processing paths, rather than wiring length. These findings suggest that neural systems are not exclusively optimized for minimal global wiring, but for a variety of factors including the minimization of processing steps.

2.2.- Neuron and brain networks

❑ Are anatomical networks efficient in transmitting information?

Macaque cortex:

real   optimized

C. Elegans:

real

optimized

2.2.- Neuron and brain networks

Anatomical Networks

Functional Networks

Anatomical parcellation

Histological or imaging data

2

3

- Histological Analysis
- DTI (MRI)

Structural brain network

Recording sites

2

Time series data

- Cross-correlation
- Wavelet coherence
- Sync. likelihood
- Generalized Sync.
- Phase Sync.
- Mutual Info.
- Granger Causality

- EEG
- MEG
- fMRI

3

Functional brain network

Sensorimotor
Premotor
Parietal
Prefrontal
Occipital
Inferior temporal
Orbitofrontal
Temporal pole

4

Graph theoretical analysis

From Bullmore & Sporns, Nature Rev. 10, 186 (2009)

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

2.2.- Neuron and brain networks

❑ **Anatomical networks in animals :**

❑ **Macaque cortex:**

N=71 Brain Areas and L=746
Small-world
No power-law

❑ **Cat cortex:**

N= 52 Brain Areas and L=820
Small-world
No power-law

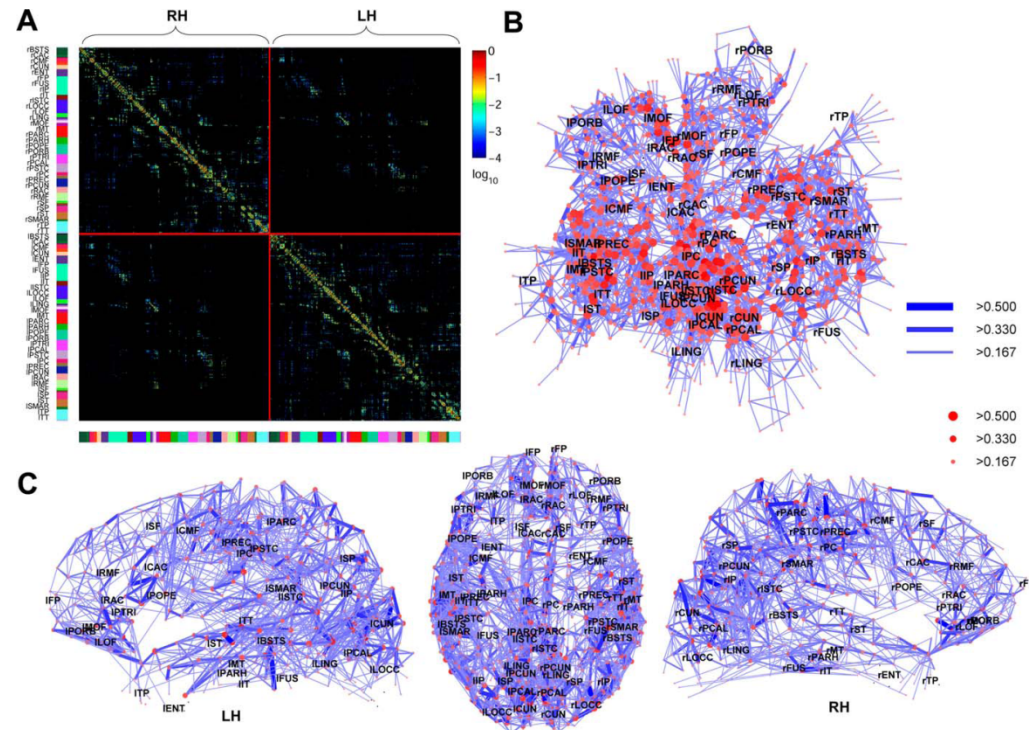# 2.2.- Neuron and brain networks

❑ **The anatomical structure of Human Brain:**

❑ **Exponential** (not scale-free) degree distribution (note that there are 66 subregions and 998 ROIs).

❑ **Small-world** attributes.

❑ **Multiple modules** interlinked by hub regions.

❑ **Positive assortativity**.

Hagmann et al. (2008) *PLoS Biol. 6, e159*
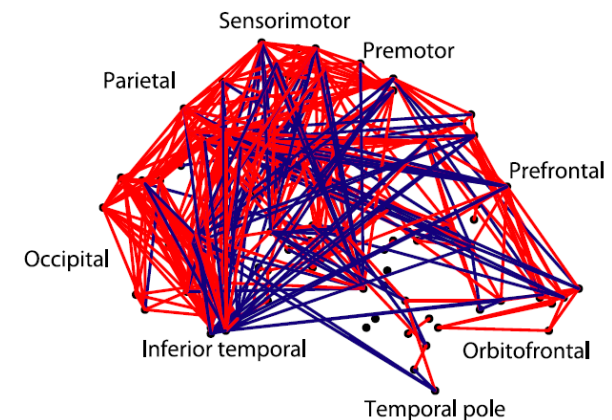
❑ How to obtain functional brain networks:

| STEP 1 | STEP 2 | STEP 3 |
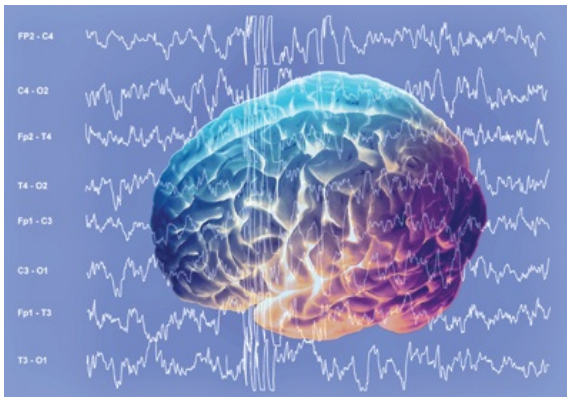
**Measurement of brain activity**   **Time series analysis**   **Network Analysis**
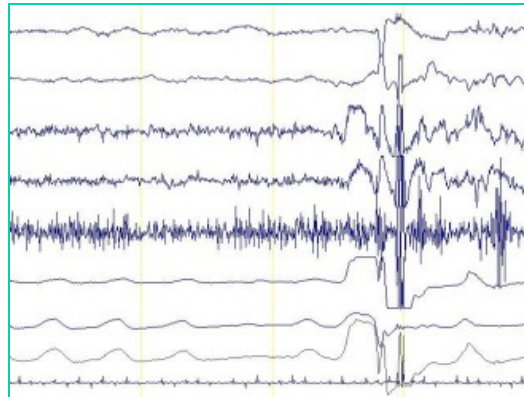
2.2.- Neuron and brain networks

❑ Complex networks methods give useful information at 3 different levels:

 ✓ **Characterize the topology of brain functional networks and its influence in the processes occurring in them:**
   ❑ Small-world topology -> High efficiency in information transmission.
   ❑ High clustering -> Good local resilience.
   ❑ Modularity -> Segregation & integration of information.

 ➢ **Identify differences between healthy brains and those with a certain pathology:**
   ❑ Quantify evolution towards random topologies.
   ❑ Evaluate the loss of modularity in the networks.
   ❑ Quantify the increase of energy expenses.

 ➢ **Develop models in order to explain the changes found in impaired functional networks:**
   ❑ Identify what are the rules that determine the network distortion.

CTB centro de tecnología biomédica   CSG   Universidad Rey Juan Carlos

❑ **Food Webs = Trophic interactions**
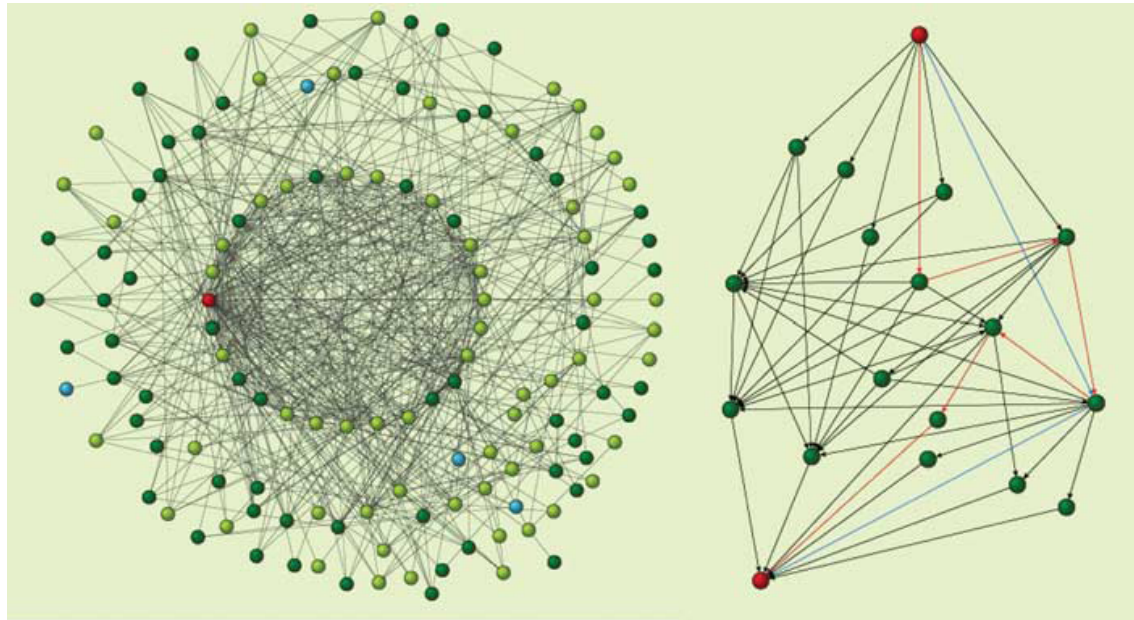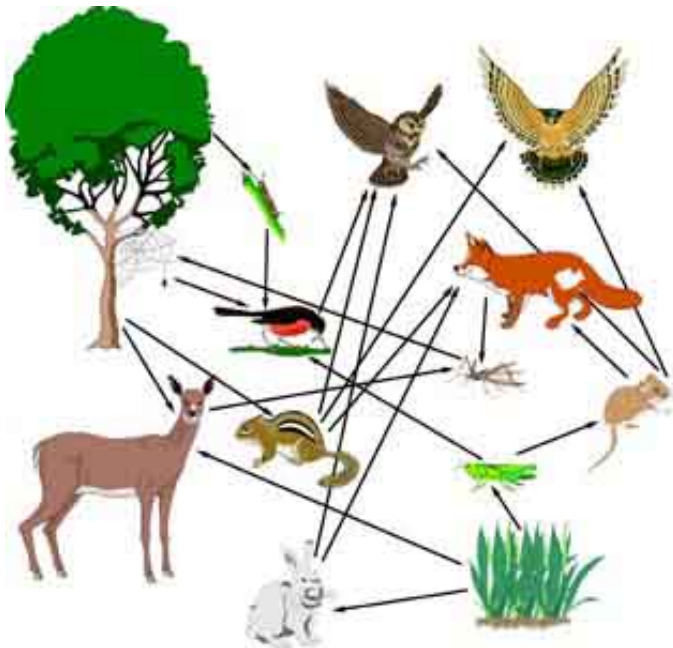
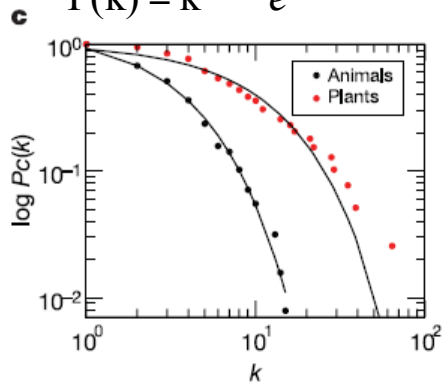Montoya, J, S L Pimm, R V Sole (2006) Ecological Networks and their fragility. *Nature*, 442

Montoya, J, S L Pimm, R V Sole *Nature*, 442 (2006)

An exponential

$P(k) = e^{-k/3.998}$
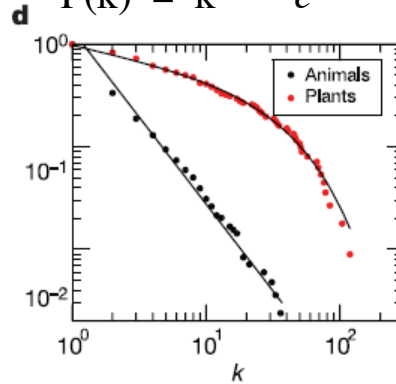
Pl truncated power law

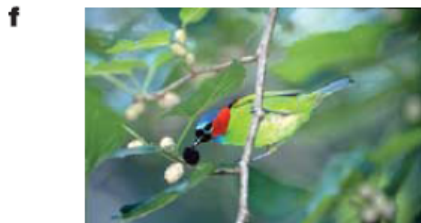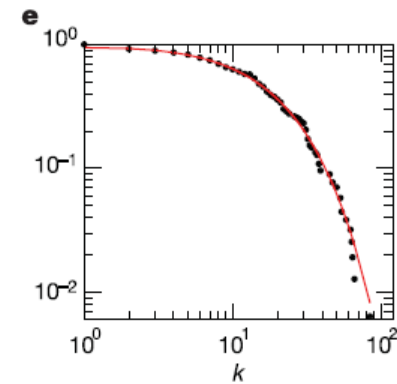$P(k) = k^{-0.013} e^{-k/11.22}$

A power law

$P(k) = k^{-1.512}$

Pl truncated power law

$P(k) = k^{-0.2822} e^{-k/42.55}$

exponential

$P(k) = e^{-k/8.861}$

Frugivore-plant

Pollinator-plant

El Verde rainforest

2.3.- Networks in Ecology

Dolphin network of Doubtful Sound ($\Omega=0.7$)

☐ **How did real networks play their cards?**

$C_A = 0.95$

50%

$C_B = 0.05$

☐ We define a competition parameter that indicates which network benefited from the structure of connections in real cases:
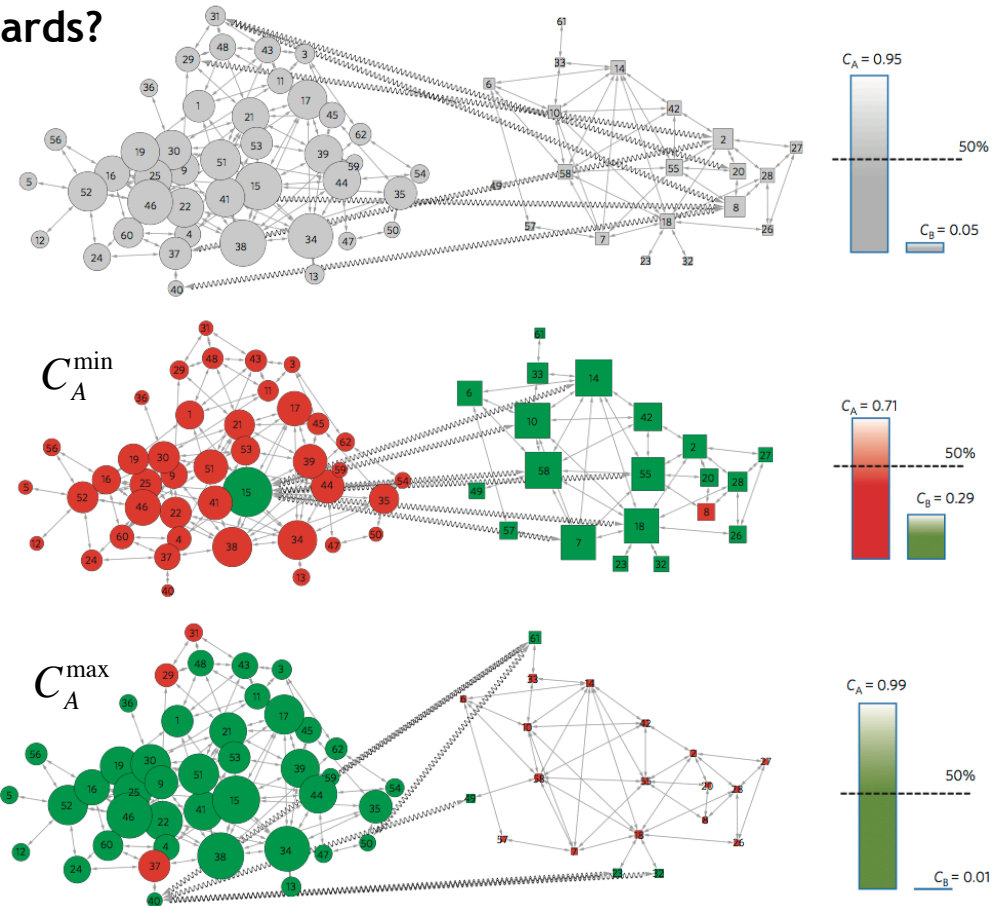
$$\Omega = \frac{2\left(C_A - C_A^{\min}\right)}{C_A^{\max} - C_A^{\min}} - 1$$

$C_A^{\min}$

$C_A = 0.71$

50%

$C_B = 0.29$

$\Omega = 1$ Strong network best case

$\Omega = -1$ Weak network best case

$\Omega \approx 0$ Trade-off solution

$C_A^{\max}$

$C_A = 0.99$

50%

$C_B = 0.01$
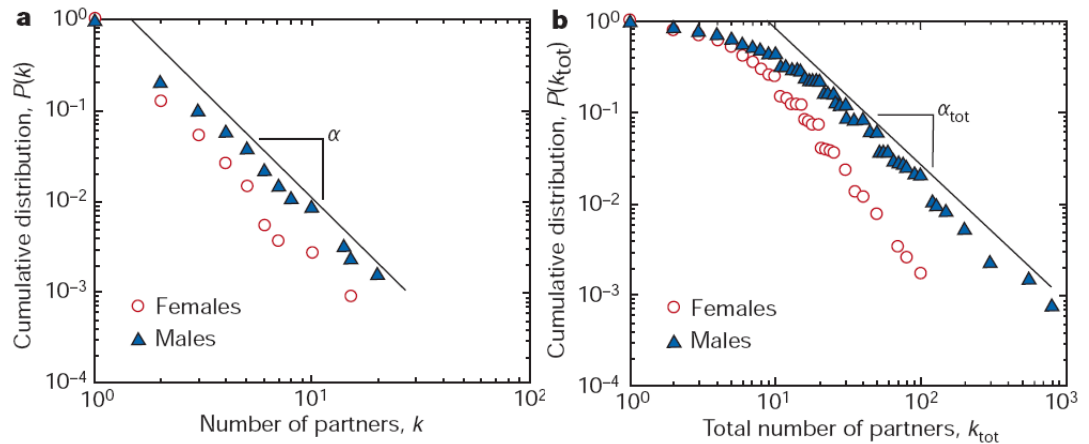
❑ Networks of social contacts are not homogeneous
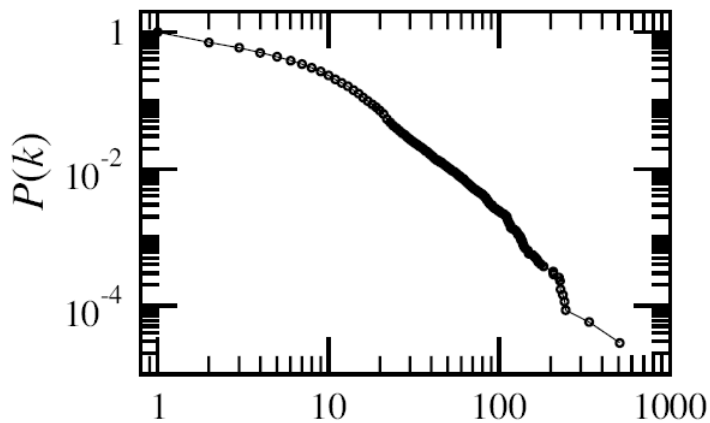
Collaboration network between musical artists. From Park, IJBC, 17, 2281 (2007).

Network of sexual contacts. (a) Partners during the last year and (b) number of partners during the entire lifetime. Sample: 4781 Swedes. From Liljeros, Nature, 411, 907 (2001)

Wilt Chamberlain (NBA basketball player) wrote that he had had sex with approximately 20,000 women. (that means having sex with 1.2 women a day, every day since he was fifteen years old.) (no comments)

Read more at:
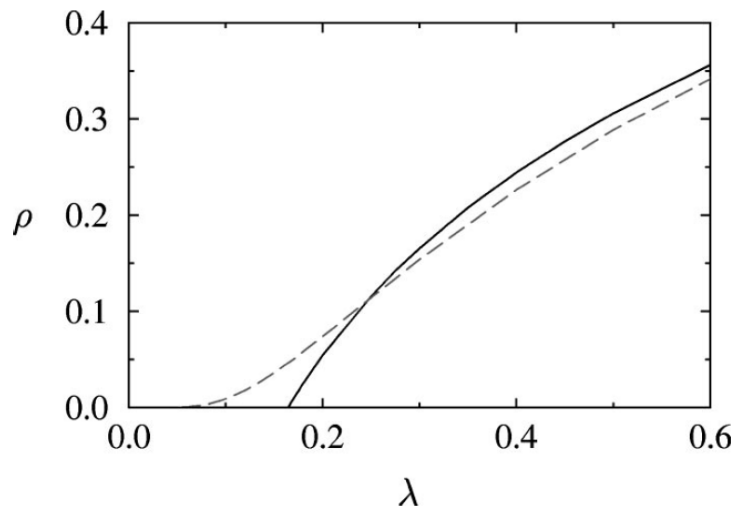*Pastor-Satorras et al.,
PRE, 65, 36104 (2002)*

2.4.- Disease spreading

❑ How can I fight the disease in scale-free networks?

Targeted immunization: We select the most connected nodes as the candidates for immunization.

ρ_g / ρ_0 = Reduced prevalence
g= ratio of immune individuals

Density of infected nodes in a random network (solid) and BA network (dashed). From Pastor-Satorras, PRE 63, 066117 (2001).

$g_c \simeq 0.16$

- Uniform Immunization
- Targeted Immunization
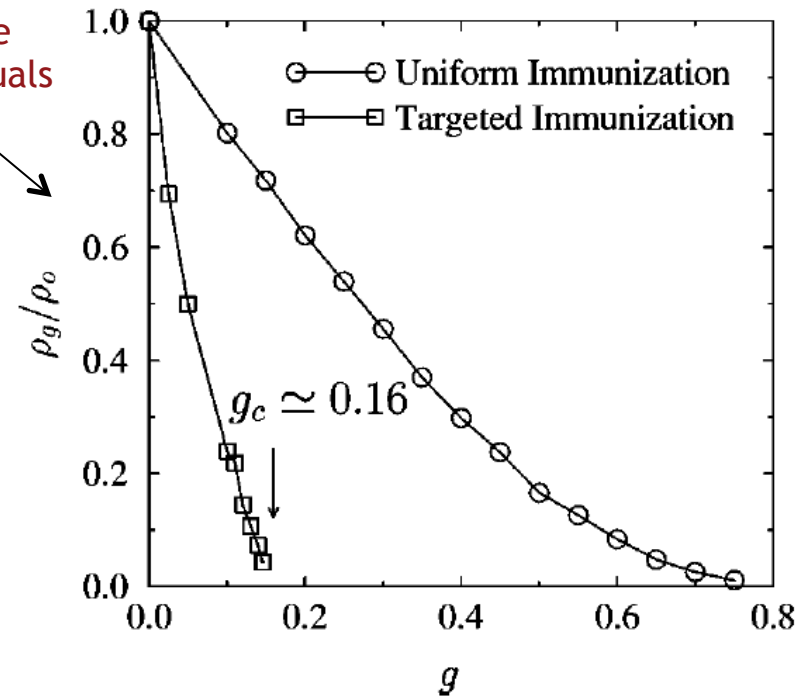
Summary and conclusions

1) *The structure of connections between biological units (genes, proteins, neurons, individuals) plays a fundamental role in any biological process.*

2) *It seems I am going to enjoy the School!*

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos

Thanks for your attention

Come to visit us in Madrid (great city, great people)

Contact: javier.buldu@urjc.es

CTB centro de tecnología biomédica

CSG

Universidad Rey Juan Carlos