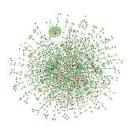


Applications of Complex Networks

Javier M. Buldú

http://www.complex.etsit.urjc.es/jmbuldu
Complex Systems Group
Universidad Rey Juan Carlos (Madrid, Spain)







OUTLINE OF THE COURSE

- 0.- Bibliography
- 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 of Complex Networks
 - 2.1.- Social Networks
 - 2.2.- Technological Networks
 - 2.3.- Biological Networks
- 3.- Future trends (and paranoias)





2.- Applications of Complex Networks

2.4.- Biological Networks

- 2.4.1.- Introduction
- 2.4.2.- Metabolic, protein and genetic networks
- 2.4.3.- RNA neutral networks
- 2.4.4.- Brain functional networks





■ 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	2115	2240	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	2359	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).





■ Biological Networks:

General properties (if they exist!):

- ☐ 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.





□ 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}	$C_{ m actual}$	$C_{ m random}$
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
<i>C. elegans</i>	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:	
Biological networks are in fact a wide area of st	tudy
☐ Metabolic, protein and genetic networks	
☐ Neuron networks	
☐ Mutation networks of virus	
☐ Functional brain networks	
☐ Food webs in ecosystems	
☐ Animal grouping and swarm movement	
□ and many others	



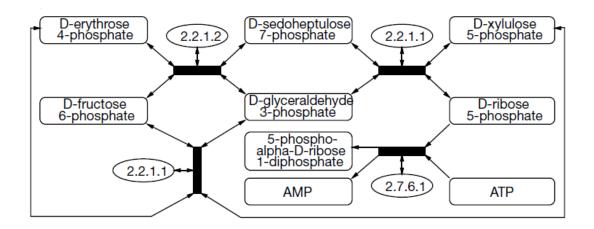






Metabolic networks:

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.



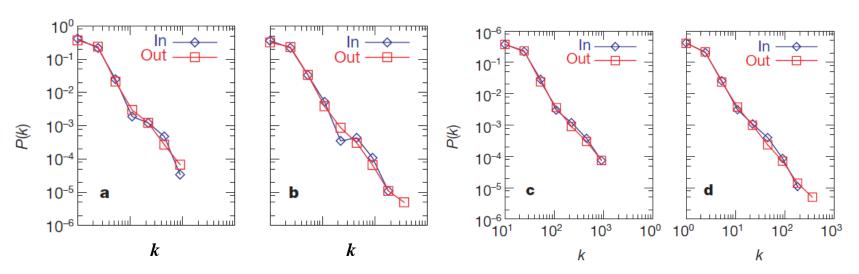
A portion of the WIT database for *E. coli*. Each substrate can be represented as a node of the graph, linked through temporary educt-educt complexes (black boxes) from which the products emerge as new nodes (substrates). The enzymes, which provide the catalytic scaffolds for the reactions, are shown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





Metabolic networks:

Metabolic networks have scale-free degree distribution.



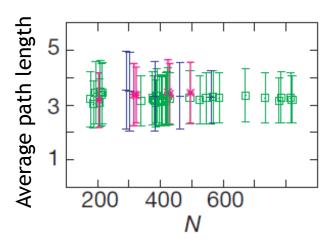
Connectivity distributions P(k) for: (a) Archaeoglobus fulgidus (archae); (b) E. coli (bacterium); (c) Caenorhabditis elegans (eukaryote), counting separately the incoming (In) and outgoing links (Out) for each substrate. k_{in} (k_{out}) corresponds to the number of reactions in which a substrate participates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



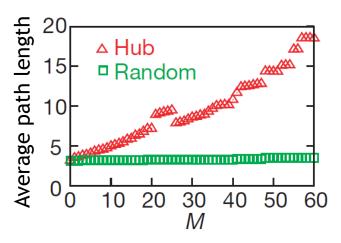


■ Metabolic networks:

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 \sim 8% of the network metabolites. From Jeong et al., Nature, 407, 651 (2000).





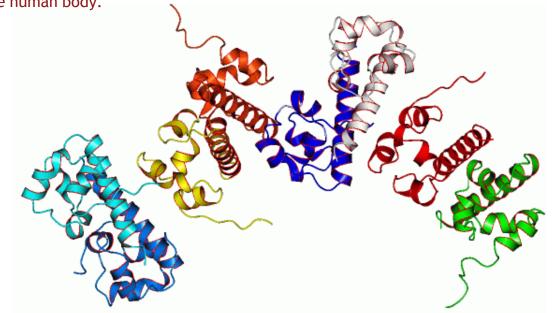


□ Protein networks:

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)



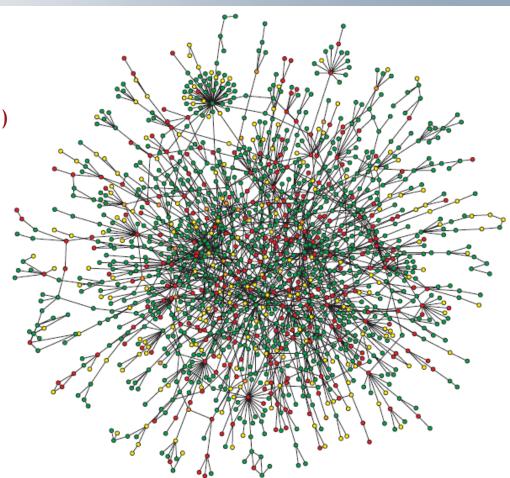




☐ Protein networks:

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).



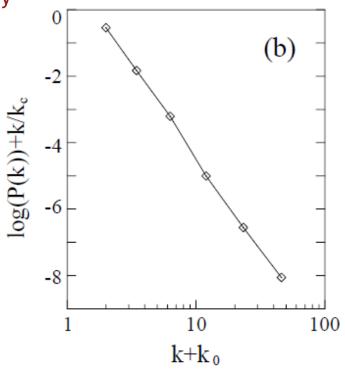




☐ Protein networks:

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 distribution is scale-free with an exponential cut-off (around $k_c \sim 20$). From Jeong et al., Nature, 411, 41 (2001).









☐ 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.

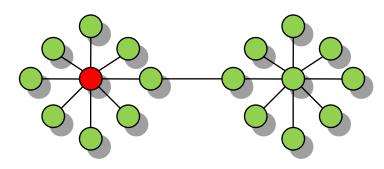
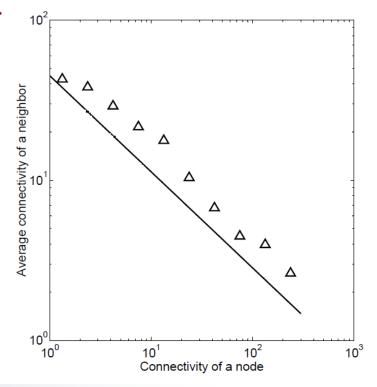


Figure: Distribution of the average neighbor connectivity for the yeast protein-protein interaction network. Here, N=3278 and M= 4549. From Maslov et al., Science., 296, 910 (2002).



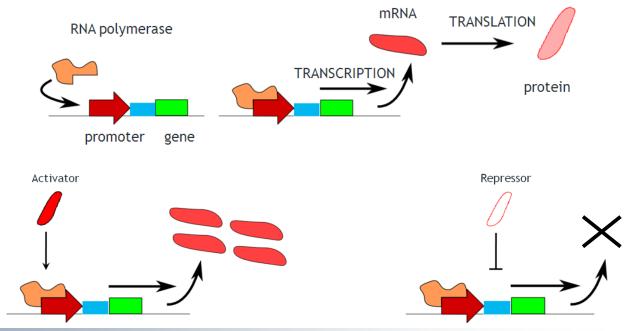






☐ Genetic networks:

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

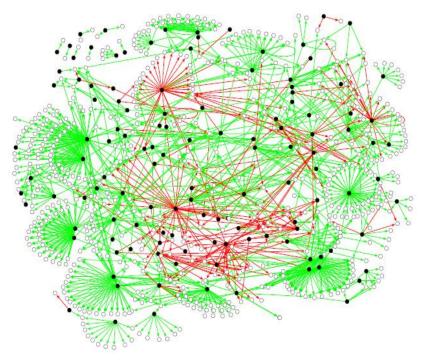






☐ Genetic networks:

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



- transcription factor
- o 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)



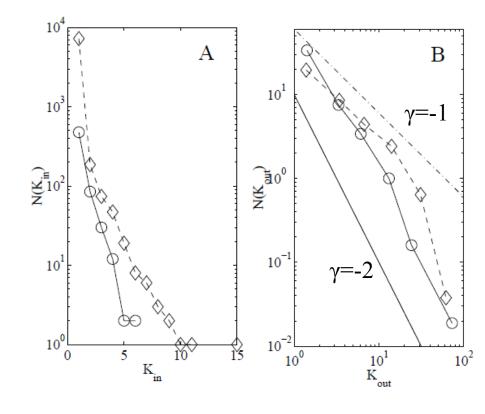




☐ Genetic networks:

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' indegrees 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).





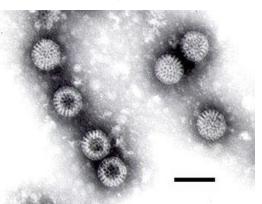






RNA neutral networks:

An RNA virus is a virus that has RNA (ribonucleic acid) as its genetic material



A: adenina

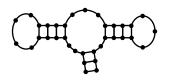
C: citosina

G: guanina

U: uracilo (en lugar de tinina)

The map between sequence and structure is degenerated: the same structure can be attained with a very large number of different chains

```
(((..((((...))))...((((...)))).))
(((..((((...))))...((((...)))).))
(((..((((...))))...((((...)))).))
(((..((((...))))...((((...)))).))
```







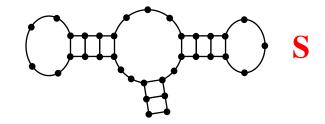


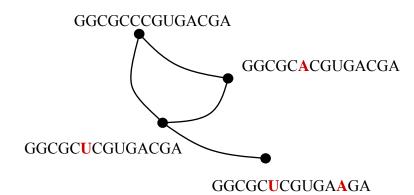
☐ RNA neutral networks:

CONSTRUCTION OF THE NETWORK

- Fix a secondary structure \$
- A *node* corresponds to a sequence that has **S** as a secondary structure
- A *link* is drawn between two nodes if they are at a Hamming distance of one

(A sequence of length **n** is linked to at most 3**n** other nodes. The maximum size of such network is **n**⁴.)



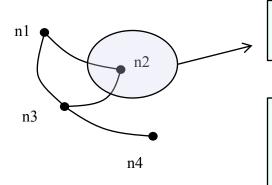






■ How does the virus population evolve in the network?

RNA virus have high mutation rates µ



At the duplication process, each sequence mutates (at one base) with a probability μ

If the mutation EXISTS (it should be a neighbor), it increases the population of the neighbor sequence

If the mutation DOES NOT EXIST, it dies.

The virus population is constant





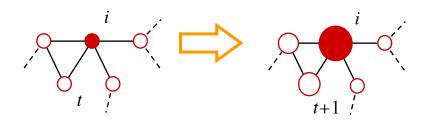
■ How does the virus population evolve in the network?

$$n_i(t+1) = (2-\mu)n_i(t) + \frac{\mu}{3l} \sum_{\{nn\}_i} n_j(t)$$
.

n_i: population at node i μ: mutation rate



The topology is contained in the adjacency matrix **C**



Knowledge of C permits to calculate the final state (population in each node i) and the time required to attain equilibrium:

The final state only depends on C

Time to equilibrium depends on C and on the mutation rate





□ How does the virus population evolve in the network?

$$n_i(t+1) = (2-\mu)n_i(t) + \frac{\mu}{3l} \sum_{\{nn\}_i} n_j(t).$$

$$\vec{n}(t+1) = M\vec{n}(t)$$

$$\mathbf{M} = (2 - \mu)\mathbf{I} + \frac{\mu}{3l}\mathbf{C}.$$

M=Transition matrix
C=Adjacency matrix (topology)

$$\lambda_i = (2 - \mu) + \frac{\mu}{3l} \gamma_i$$

 λ_i =eigenvalues of M γ_i =eigenvalues of C

$$\vec{u}_i = \vec{w}_i$$

 w_i =eigenvalues of M u_i =eigenvalues of C





■ How does the virus population evolve in the network?

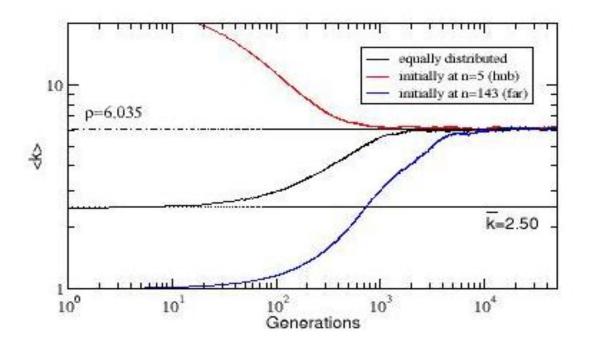
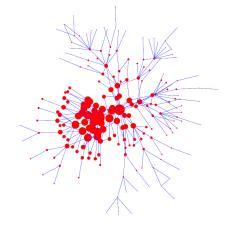


Figure: Average degree of the population as a function of time for a scale-free network. The final value ρ corresponds to the spectral radius of the adjacency matrix. Here μ =0.1 (N=200).







■ How does the virus population evolve in the network?

- \square The final distribution of the population (t $\rightarrow \infty$) is given by \vec{u}_1 (the eigenvector of the largest eigenvalue). The topology tells us the final state!
- \Box The mean degree of the population is $\langle k \rangle = \gamma_1$

$$k_{min} < \langle k \rangle < \gamma_1 < k_{max}$$

indicating that the population evolves to the more connected regions of the network.

☐ The time to reach the equilibrium is approximated by

$$t_{\epsilon}^{1} \simeq \frac{\ln |\alpha_{2}/\alpha_{1}| - \ln \epsilon}{\ln |\lambda_{1}/\lambda_{2}|}$$

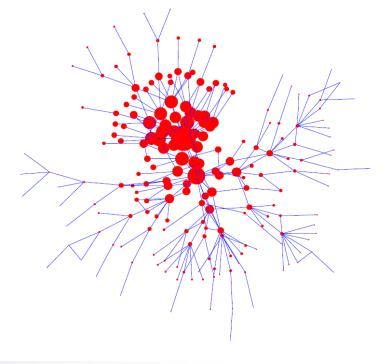
 \Box For a fixed topology and initial conditions, the time to reach the equilibrium goes with μ^{-1}



■ How does the virus population evolve in the network?

No matter where the initial sequence is (in the network), if the virus has enough time, it will evolve to the same final distribution

The population evolves to the more connected areas, in this way is more robust to mutation. This property is known as **neutrality**.







Energy versus topology

$$n_i(t+1) = \left[(2-\mu)n_i(t) + \frac{\mu}{3l} \sum_{\{nn\}_i} n_j(t) \right] e^{-\beta(E_i - E_{min})}$$

The probability to stay at node *i* is higher the lower its energy

$$p_i = \exp\{-\beta(E_i - E_{min})\}.$$

$$\mathbf{M}' = \mathbf{E}\left[(2-\mu)\mathbf{I} + \frac{\mu}{3l}\mathbf{C}\right] = \mathbf{E}\mathbf{M}$$

$$E_i = -(3N_{
m GC} + 2N_{
m AU} + N_{
m GU}) \,,$$
 A=U - 2 Kcal/mol

G≡C - 3 Kcal/mol G-U - 1 Kcal/mol

The parameter β quantifies the relative importance of high connectivity versus low energy

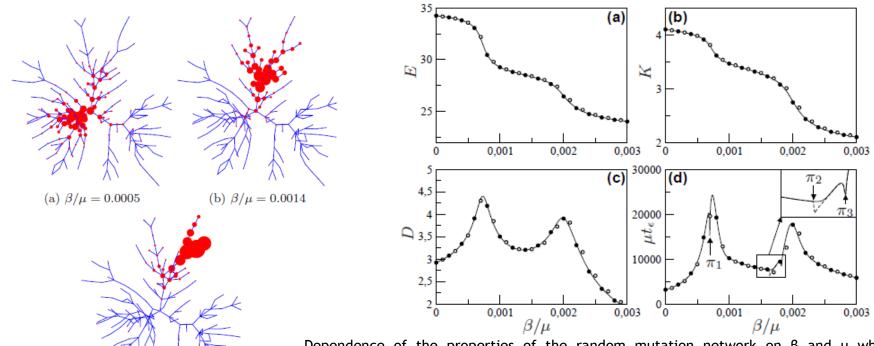
 $\beta \rightarrow 0$ the population evolves to the most connected nodes (connectivity).

 $\beta \rightarrow \infty$ the population evolves to nodes with lower energy (stability).





☐ Energy versus topology in random networks



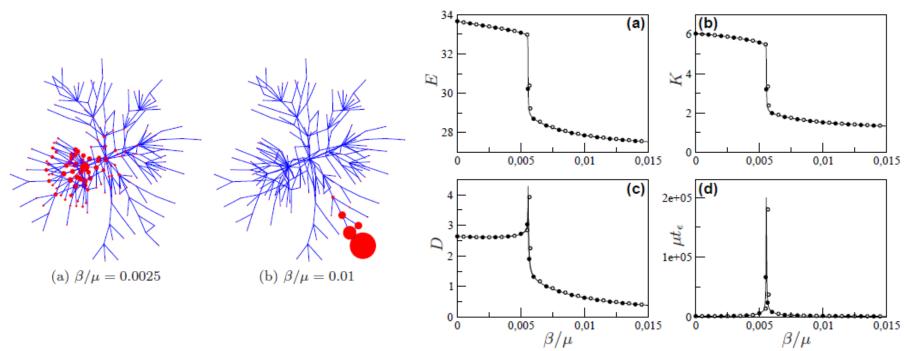
Dependence of the properties of the random mutation network on β and μ when neutrality and energetic stability are negatively correlated (NS-). Each curve is plotted for μ = 0.001 (•), 0.01 (solid line), and 0.05 (). (a) Average energy E, (b) Average degree K, (c) Average dispersion D, (d) Dependence of the rescaled time to equilibrium





(c) $\beta/\mu = 0.0027$

☐ Energy versus topology in scale-free networks



Dependence of the properties of the preferential mutation network on B and μ when neutrality and energetic stability are negatively correlated (NS-). Each curve is plotted for μ = 0.001 (•), 0.01 (solid line), and 0.05 (). (a) Average energy E, (b) Average degree K, (c) Average dispersion D, (d) dependence of the rescaled time to equilibrium







■ Evolution of RNA virus

	Evolut	ion	ary dynami	cs on neutr	al networks	leads	populations	to h	ighly conne	ected area	s in
the	space	of	genomes:	neutrality	(connectivity	y) is	optimized,	thus	increasing	robustness	to
mut	ations										

- ☐ When the **energy** of the folded state is taken into account, the population concentrates around sequences of **minimal energy**, thus increasing robustness to perturbations
- □ Robustness arises as a compromise between minimizing the effect of mutations and maximizing structural stability
- ☐ The **time** required to reach the asymptotic state has to be shorter than the time between changes in the environment; diversity (related to adaptability) depends on the relation between energy and connectivity of sequences





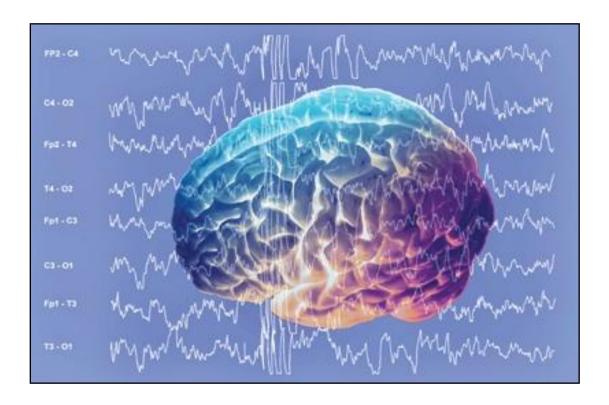
2.4.4.- Brain functional networks







■ What does the brain when it is working?

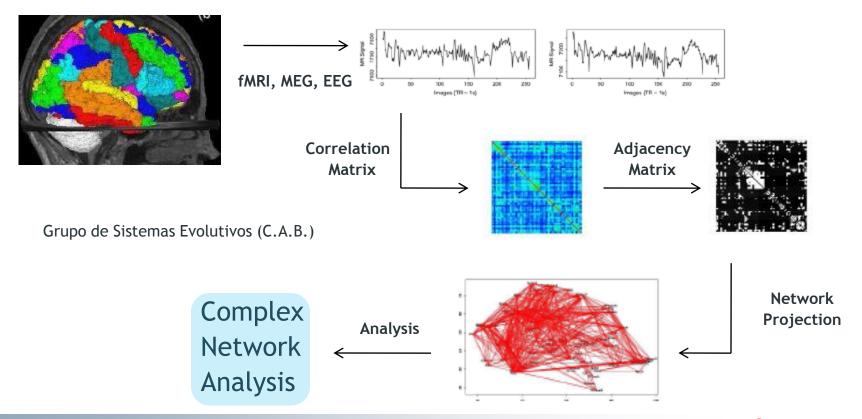






2.4.4.- Brain functional networks

☐ How to obtain a network by measuring the brain activity







2.4.4.- Brain functional networks



- ☐ How to measure the brain activity
 - Functional MRI (fMRI). The detection of changes in regional brain activity through their effects on blood flow and blood oxygenation (which, in turn, affect magnetic susceptibility and tissue contrast in magnetic resonance images). High spatial resolution (~mm³) but low temporal resolution (~ seconds).
 - □ Electroencephalography (EEG). A technique used to measure neural activity by monitoring electrical signals from the brain, usually through scalp electrodes. EEG has good temporal resolution but relatively poor spatial resolution.
 - ☐ Magnetoencephalography (MEG). A method of measuring brain activity by detecting perturbations in the extracranial magnetic field that are generated by the electrical activity of neuronal populations.





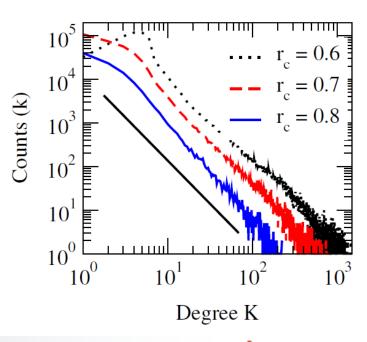


☐ Scale-free brain functional networks

- ☐ Two activities: finger tapping and listening to music
- □ ~ 400 events every 2.5 seconds
- ☐ 36 x 64 x 64 brain sites (147456 voxels)
- ☐ The linear cross-correlation is measured

$$r(x_1, x_2) = \frac{\langle V(x_1, t)V(x_2, t)\rangle - \langle V(x_1, t)\rangle\langle V(x_2, t)\rangle}{\sigma(V(x_1))\sigma(V(x_2))}$$

☐ Several thresholds are considered in order to obtain the adjacency matrix.





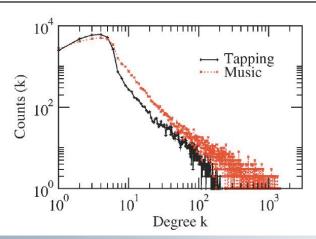


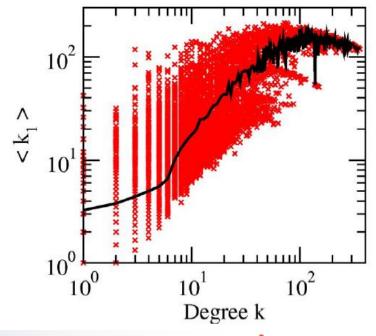


■ Scale-free brain functional networks

fMRI functional networks are small-world, scale-free and assortative

r_c	N	С	L	$\langle k \rangle$	γ	C_{rand}	$L_{ m rand}$
0.6	31 503	0.14	11.4	13.41	2.0	4.3×10^{-4}	3.9
0.7	17 174	0.13	12.9	6.29	2.1	3.7×10^{-4}	5.3
0.8	4891	0.15	6.0	4.12	2.2	8.9×10^{-4}	6.0









☐ How to measure the brain activity

MEG (and EEG) allows the band decomposition of the signal into frequency bands

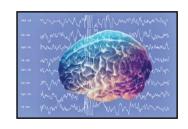
Wavelet decomposition level	Frequency range, Hz	Corr	τ	k	L	С	σ	ζ	S (×10 ⁻³)
Resting									
1	37.5-75	0.18 ± 0.02	0.50 ± 0.05	16.3 ± 5.1	4.5 ± 0.5	0.23 ± 0.02	1.9 ± 0.2	61 ± 14	9.7 ± 1.9
2	18.8-37.5	0.26 ± 0.02	0.74 ± 0.04	12.6 ± 3.1	5.2 ± 0.5	0.21 ± 0.02	1.9 ± 0.1	70 ± 41	8.2 ± 2.3
3	9.4-18.8	0.30 ± 0.03	0.81 ± 0.03	12.4 ± 1.8	5.4 ± 0.4	0.20 ± 0.01	1.9 ± 0.2	100 ± 72	6.3 ± 2.7
4	4.7-9.4	0.30 ± 0.03	0.82 ± 0.03	12.3 ± 2.0	5.4 ± 0.4	0.21 ± 0.01	1.9 ± 0.2	106 ± 75	6.4 ± 3.3
5	2.3-4.7	0.30 ± 0.02	0.81 ± 0.02	12.5 ± 2.0	5.2 ± 0.4	0.21 ± 0.01	2.0 ± 0.1	118 ± 71	7.6 ± 2.9
6	1.1-2.3	0.33 ± 0.05	0.83 ± 0.02	13.7 ± 3.3	5.1 ± 0.4	0.23 ± 0.02	1.9 ± 0.1	137 ± 62	6.0 ± 2.4
Tapping									
1	37.5-75	0.18 ± 0.03	0.49 ± 0.09	16.9 ± 5.1	4.4 ± 0.6	0.23 ± 0.02	1.8 ± 0.2	132 ± 21	10.2 ± 3.6 /
2	18.8-37.5	0.23 ± 0.02	0.69 ± 0.04	13.0 ± 2.6	5.0 ± 0.4	0.21 ± 0.01	2.0 ± 0.1	105 ± 9	9.8 ± 2.7
3	9.4-18.8	0.27 ± 0.02	0.77 ± 0.03	12.2 ± 1.7	5.2 ± 0.4	0.21 ± 0.01	2.0 ± 0.1	118 ± 27	8.4 ± 2.8
4	4.7-9.4	0.28 ± 0.03	0.79 ± 0.02	12.7 ± 2.1	5.2 ± 0.5	0.21 ± 0.01	1.9 ± 0.2	116 ± 35	8.2 ± 2.9
5	2.3-4.7	0.30 ± 0.05	0.81 ± 0.01	13.8 ± 4.9	5.1 ± 0.5	0.21 ± 0.01	1.9 ± 0.2	137 ± 47	7.2 ± 2.6
6	1.1-2.3	0.34 ± 0.06	0.82 ± 0.01	16.7 ± 8.3	4.9 ± 0.8	0.22 ± 0.02	1.7±0.2	144 ± 55	5.2 ± 1.6

Corr: average correlation of the whole brain network before thresholding; τ : threshold applied to wavelet correlation matrices; k: average degree of the network; L: average path length; C: average clustering; σ : smallworld scalar value; ζ , characteristic length scale in millimeters; S, synchronizability. (N=275)





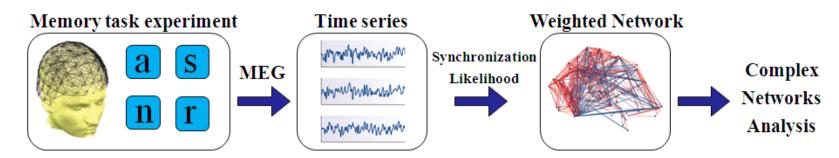
Brain functional networks



- □ Comprehension of the structural properties of the networks. It is interesting to unveil the functional structure of the brain and understand the implications on the cognitive processes.
- □ Comprehension of brain diseases. The analysis of the functional networks of patients with a certain pathology can help us to understand an localize malfunctions in the brain activity.
- Detection and treatment of brain disease. The final application of the analysis of brain functional networks is the early detection of brain diseases (which will help in their treatment).



- ☐ A practical application: Mild Cognitive Impairment (MCI)
 - Mild Cognitive Impairment (MCI). A brain disorder in which thinking abilities are mildly impaired. Individuals with mild cognitive impairment are able to function in everyday activities but have difficulty with memory, trouble remembering the names of people they met recently, remembering the flow of a conversation, and a tendency to misplace things.
 - Experiment: We performed magnetoencephalograms (MEG) to a group of 11 MCI's patients and 9 control subjects during a memory task. By means of the synchronization likelihood (SL) we quantify the interaction between the 148 channels of the MEG system and we obtain a weighted connectivity matrix between cortical areas.







(weighted) Network parameters:

- \square Matrix normalization (*P*): We map the weights of the synchronization matrix ωij into a continuous bijective map M:R \rightarrow [0,1]. The obtained matrix P {pij} can be interpreted as a matrix of probabilities that tell us how probable the existence of a link between node i and j is.
- $p_{ij} = \frac{\omega_{ij} \min(\omega_{ij})}{\max(\omega_{ii}) \min(\omega_{ii})}$
- \square Mean Shortest path (L): it measures the shortest topological (not Euclidean) distance lij between any pair of nodes in the network.
- \square Mean Clustering (C): it measures the probability that two neighbors of a certain node are also connected with each other.
- $c_i = \frac{\sum_{j,k} p_{ij} p_{jk} p_{ik}}{\sum_{j,k} p_{ij} p_{ik}}$
- □ Network outreach (O): The outreach O_i balances the distance of the connections of a node i with their probability.

$$O_i = \sum_{j \in V(i)} p_{ij} d_{ij}$$

 \square Network modularity (Q): Takes into account the number of links between the nodes of the same community and measures the statistical deviation from a random assignment of nodes between communities



□ MCI Network parameters:

There exist significant differences between the networks of both groups (control and MCI)

CONTROL	$k(\hat{k})$	$L(\hat{L})$	$C(\hat{C})$	$O(\hat{O})$	Q
α_1	11.4463 (1)	$10.1574 \ (2.12557)$	$0.149065 \ (1.91366)$	$0.0522277 \ (0.572173)$	0.282328
α_2	12.2643 (1)	$10.0241 \ (2.13698)$	$0.146967\ (1.76083)$	$0.0585542 \ (0.598713)$	0.260331
eta_1	10.1545 (1)	$11.1015 \ (2.13231)$	$0.139872\ (2.02415)$	0.0453014 (0.559416)	0.29265
eta_2	9.03521 (1)	$12.7187 \ (2.07625)$	$0.121496 \ (1.97666)$	0.0404702 (0.561634)	0.300408
M.C.I.	$k(\hat{k})$	$L(\hat{L})$	$C(\hat{C})$	$O\left(\hat{O}\right)$	Q
α_1	15.9807 (1)	$9.08057 \ (2.07814)$	$0.15655 \ (1.43985)$	$0.0850967 \ (0.667935)$	0.217019
α_2	14.179 (1)	$9.32649\ (2.07442)$	$0.156551\ (1.62274)$	$0.0710385 \ (0.628381)$	0.240754
eta_1	11.7071 (1)	$10.4313 \ (2.11228)$	$0.14822 \ (1.86043)$	$0.0548219 \ (0.587238)$	0.271003
eta_2	11.5259 (1)	$12.0867 \ (2.06726)$	$0.119362 \ (1.52228)$	$0.0600801 \ (0.653776)$	0.238308
M.C.I. vs. CONTROL	$k(\hat{k})$	$L(\hat{L})$	$C(\hat{C})$	O (Ô)	Q
α_1	+39.61 %	-10.60 %	+5.02~%	+62.93 %	-23.13 %
α_2	+15.61 %	-6.96 $\%$	+6.52~%	+21.32~%	-7.52 %
β_1	+15.29 %	-6.04 $\%$	+5.97~%	+21.02~%	-7.40 %
eta_2	+27.57~%	-4.97 $\%$	-1.76 %	+48.46~%	-20.67 %

Average network parameters for the Control and MCI group and the percentage of variation. The probability matrix has been obtained for four frequency bands. Mean degree kp, average shortest path Lp, average clustering coefficient Cp, average outreach Op and average modularity Qp. The symbol ^ indicates normalization over 100 realizations of a randomized version of the networks.

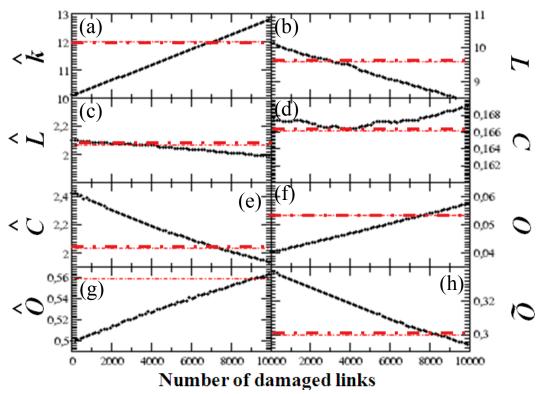




■ Modelling MCI Damage:

Instead of random failures, we simulate random increases of the synchronization between brain regions

Figure: Evolution of the network parameters predicted by the model. Red dashed lines indicate the value of the parameter of the MCI average network. (a) average degree of the network, (b) average shortest path, (c), normalized average shortest path, (d) clustering, (e) normalized clustering, (f) outreach, (g) normalized outreach, and (h) modularity.

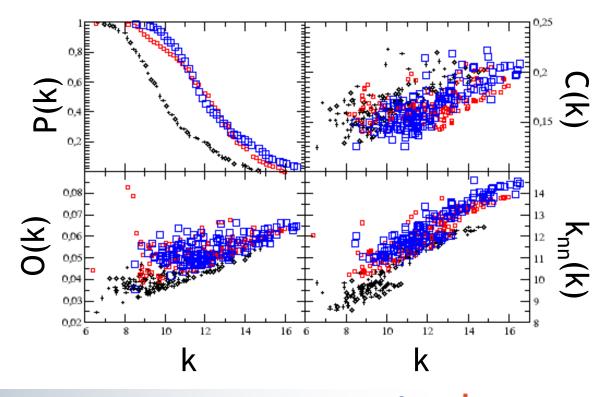






■ Modelling MCI Damage:

Figure: Several parameter distributions for the control group (black circles), MCI group (red squares) and random model evolution (blue squares): (a) probability distribution of finding a node with a degree higher than k, (b) clustering coefficient C(k), (c) outreach O(k) and (d) average neighbor degree knn(k).







Complex Network analyses of brain diseases
□ Alzeheimer.
☐ The overall synchronization of the network is decreased.
$\ \square$ The average path length increases (probably as a consequence of the reduction of the synchronization).
$\hfill\Box$ The clustering coefficient is significantly reduced (the network evolves to random topologies).
☐ Mild Cognitive Impairment.
☐ The average synchronization increases.
$\hfill \square$ Network outreach increases as a consequence of an unbalanced increase of the synchronization in the long-range connections.
☐ The network becomes more random.





- Complex Network analyses of brain diseases
 - ☐ Schizophrenia.
 - ☐ The small-world properties of the network are impaired (specially at low-frequency bands).
 - ☐ Clustering and average path length are shifted to random configurations.
 - ☐ The hierarchical configuration of the network is also affected.
 - ☐ Epilepsia.
 - ☐ Synchronization increases during the epileptic episodes.
 - ☐ As a consequence, clustering coefficient increases and average path length decreases.
 - ☐ Changes are more significant at delta, theta and alpha bands.





2.1.- SOCIAL NETWORKS



Structure matters, but we are far from understanding its relation with function





