ECCS Warm-up

Alex Arenas
Universitat Rovira i Virgili (Tarragona, Spain)

Mesoscopic description of complex networks: theory and applications

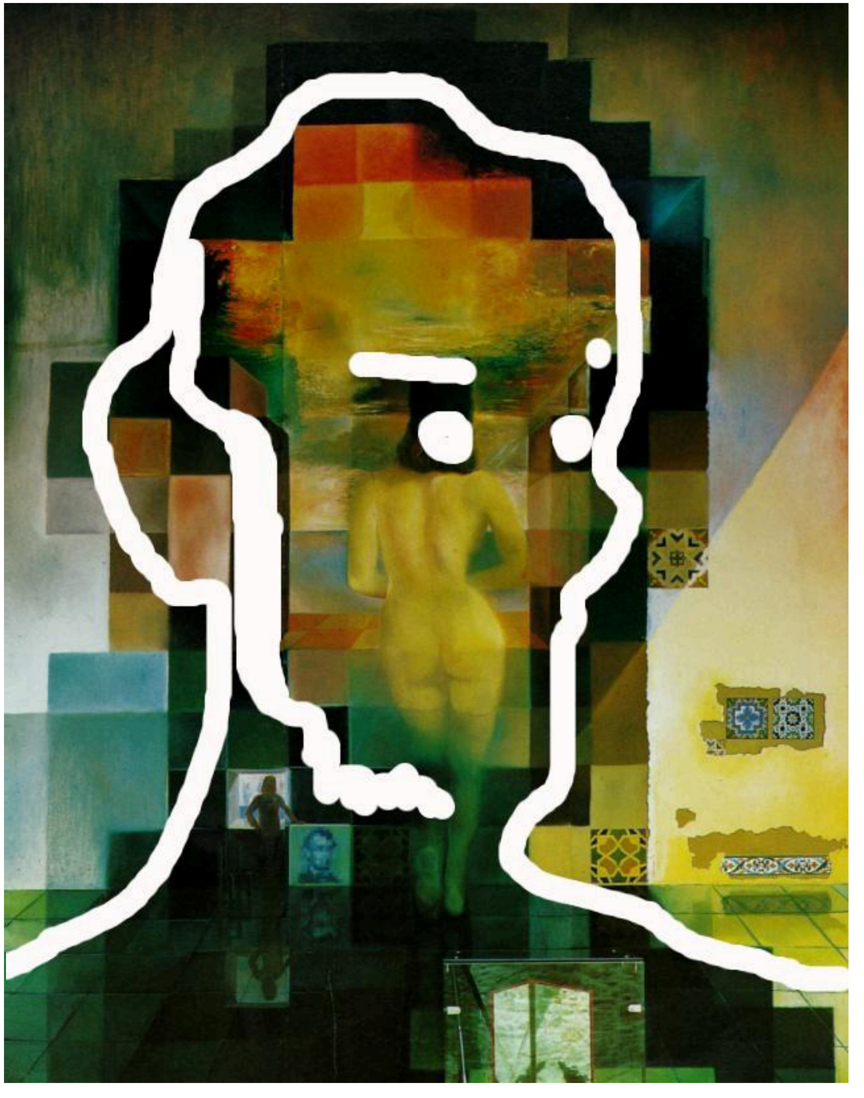
Summary

- Introduction
- Community detection problem
- Dynamic mesoscales
- Static mesoscales
- Applications

What it means “mesoscopic”?

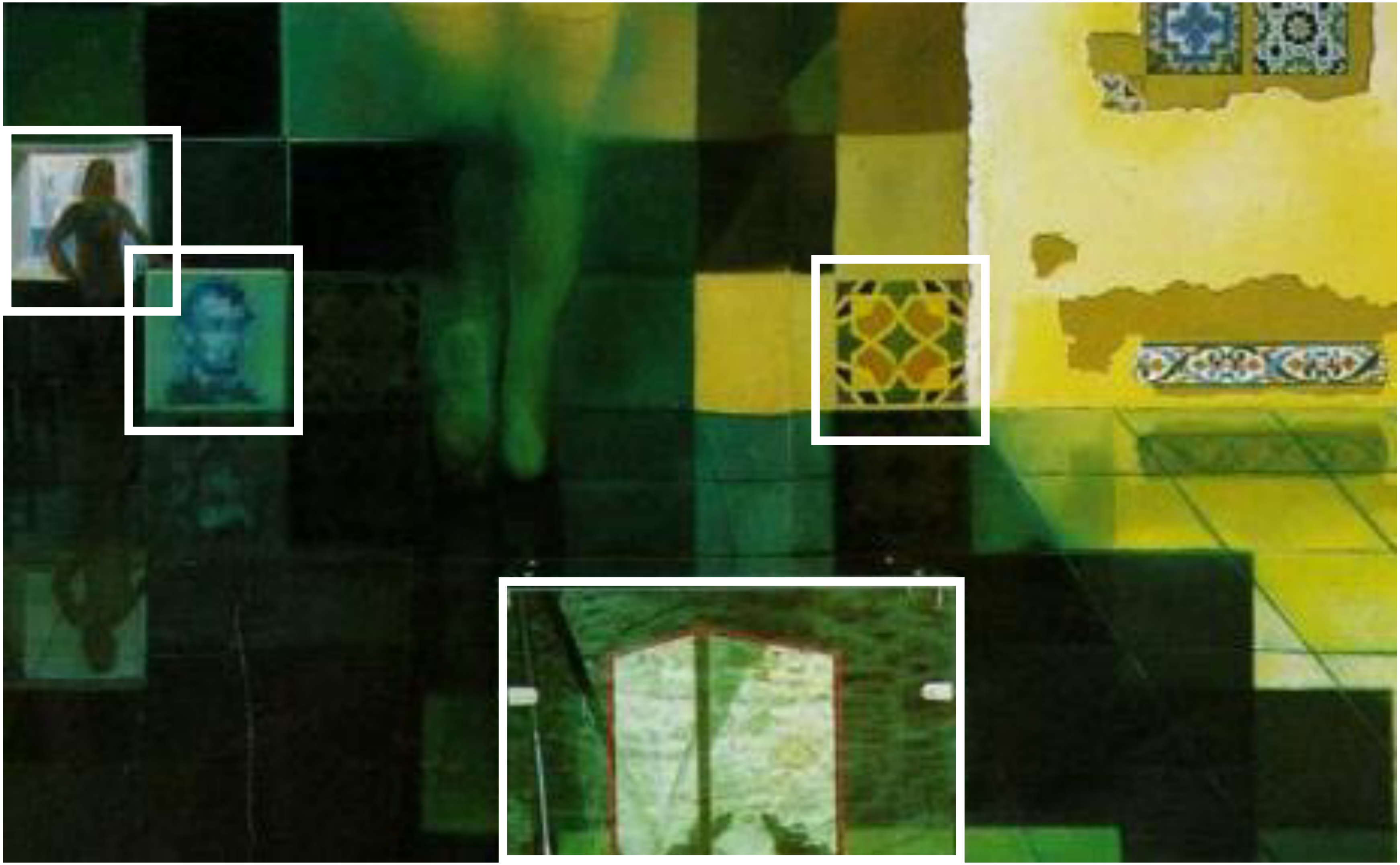
- Difficult to assess an accurate definition:
  - Statics’ perspective: consideration of length scales ranging between max and min levels of experimental resolution
  - Dynamics’ perspective: consideration of dynamic phenomena that appear at time or length scales that range between the max and min levels of description for the evolution

Static example: Dali -Lincoln portrait-



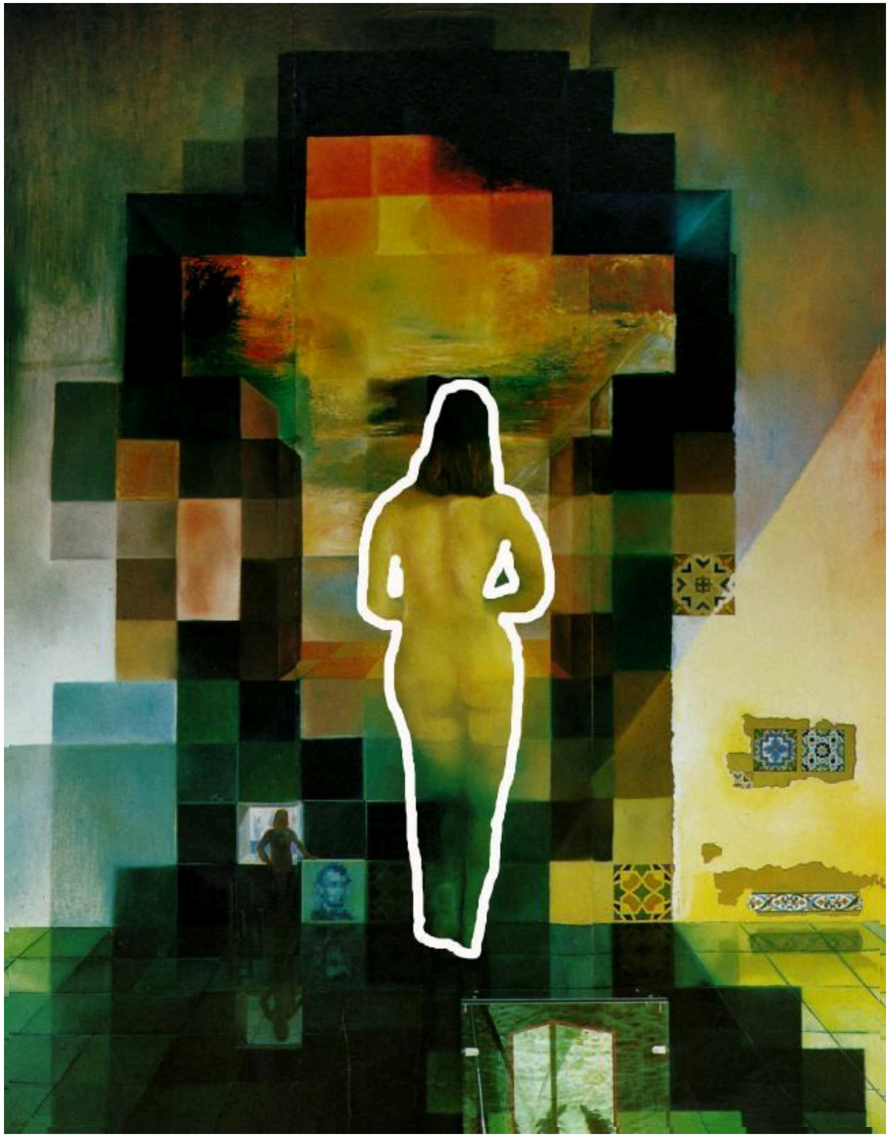
Macroscale

Static example: Dali -Lincoln portrait-



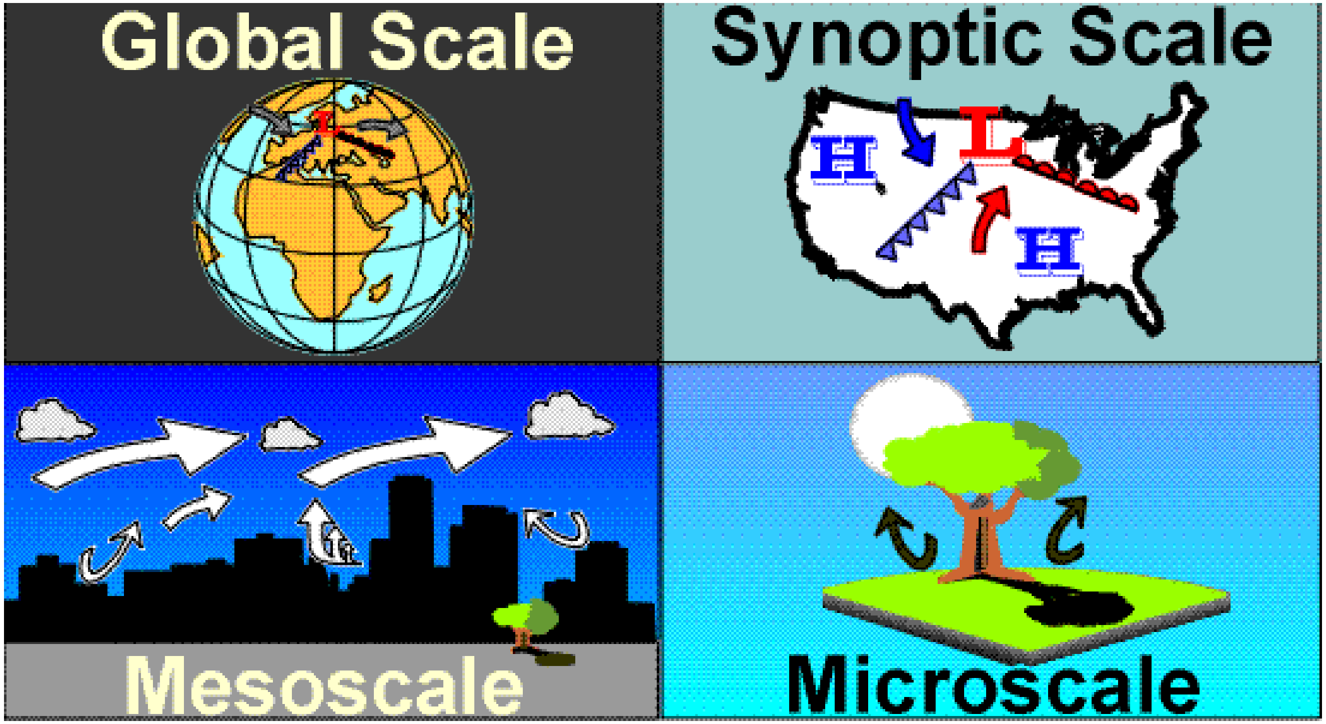
Microscale

Static example: Dali -Lincoln portrait-

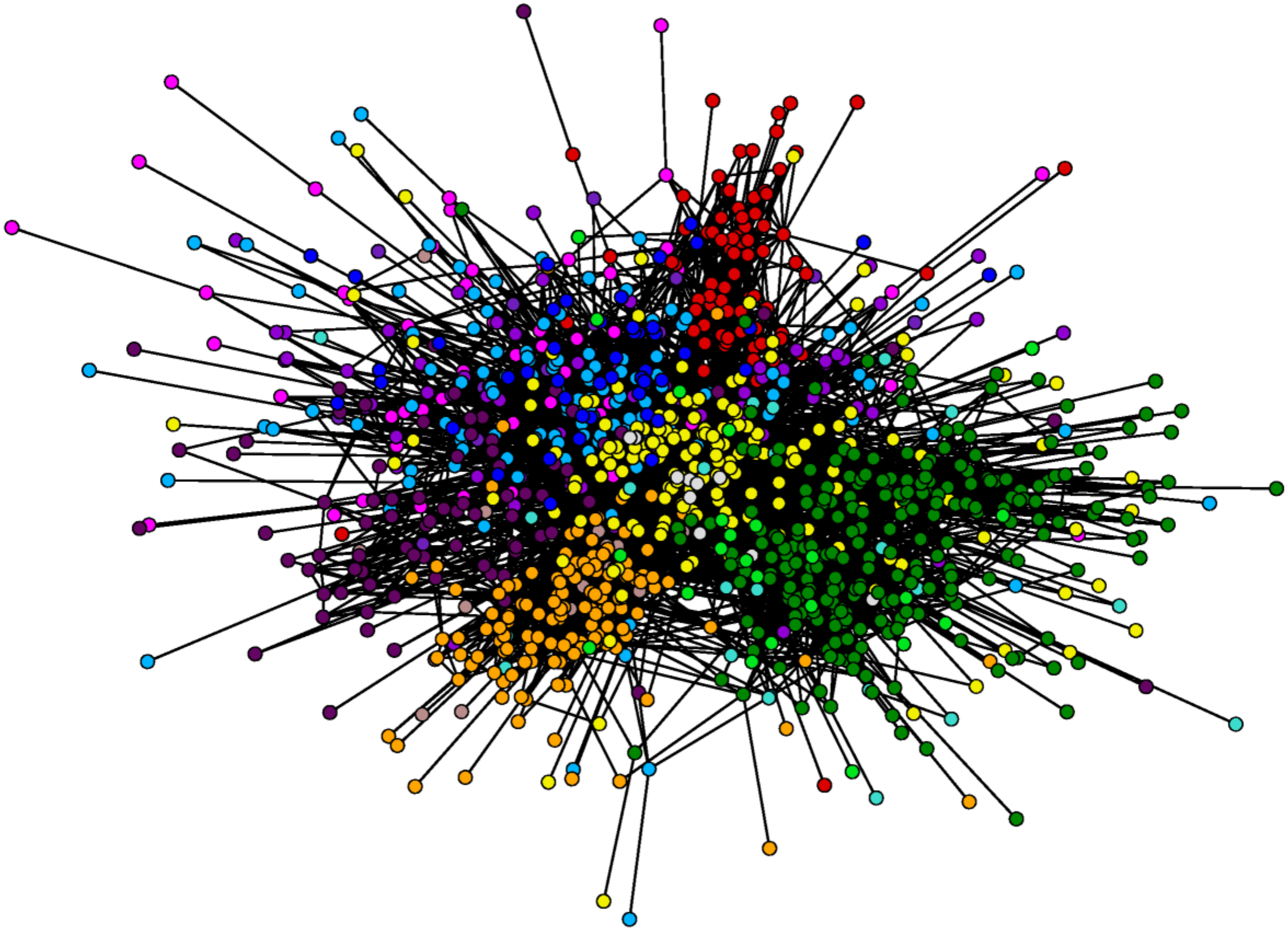


Mesoscale

Dynamic example: Weather prediction models

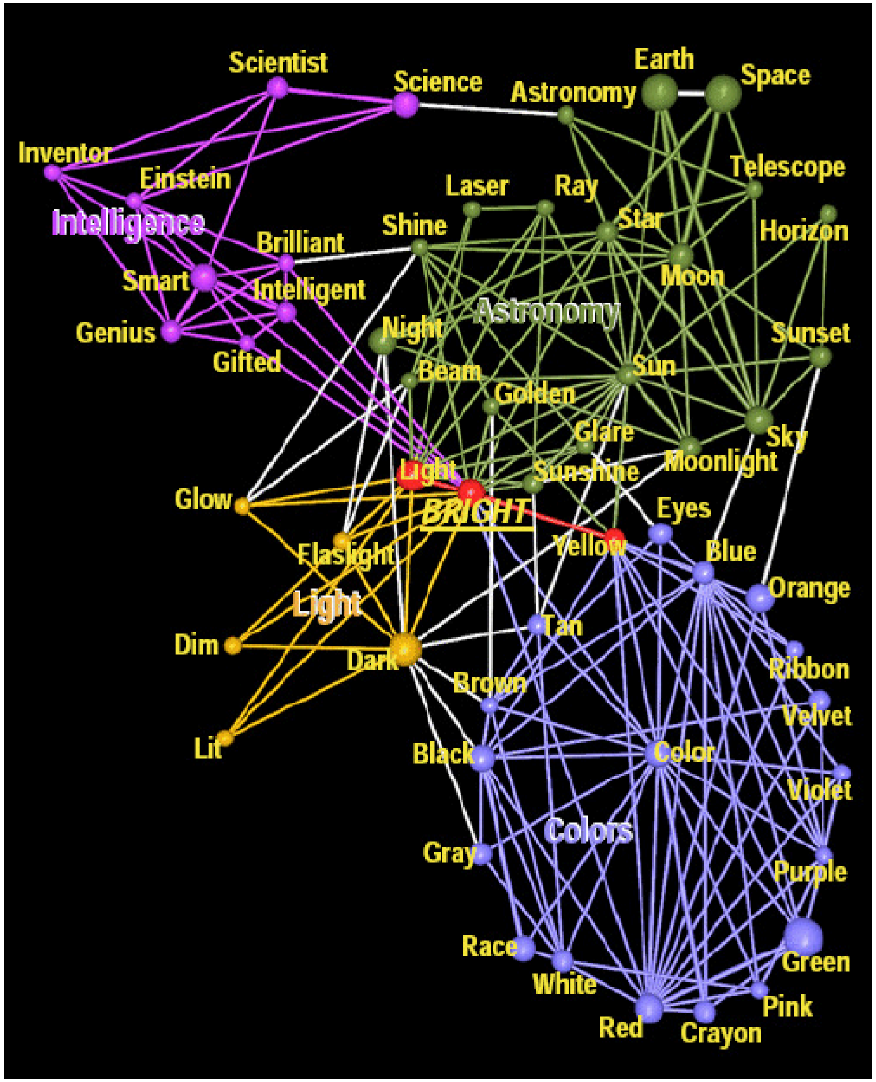


Complex networks of interactions

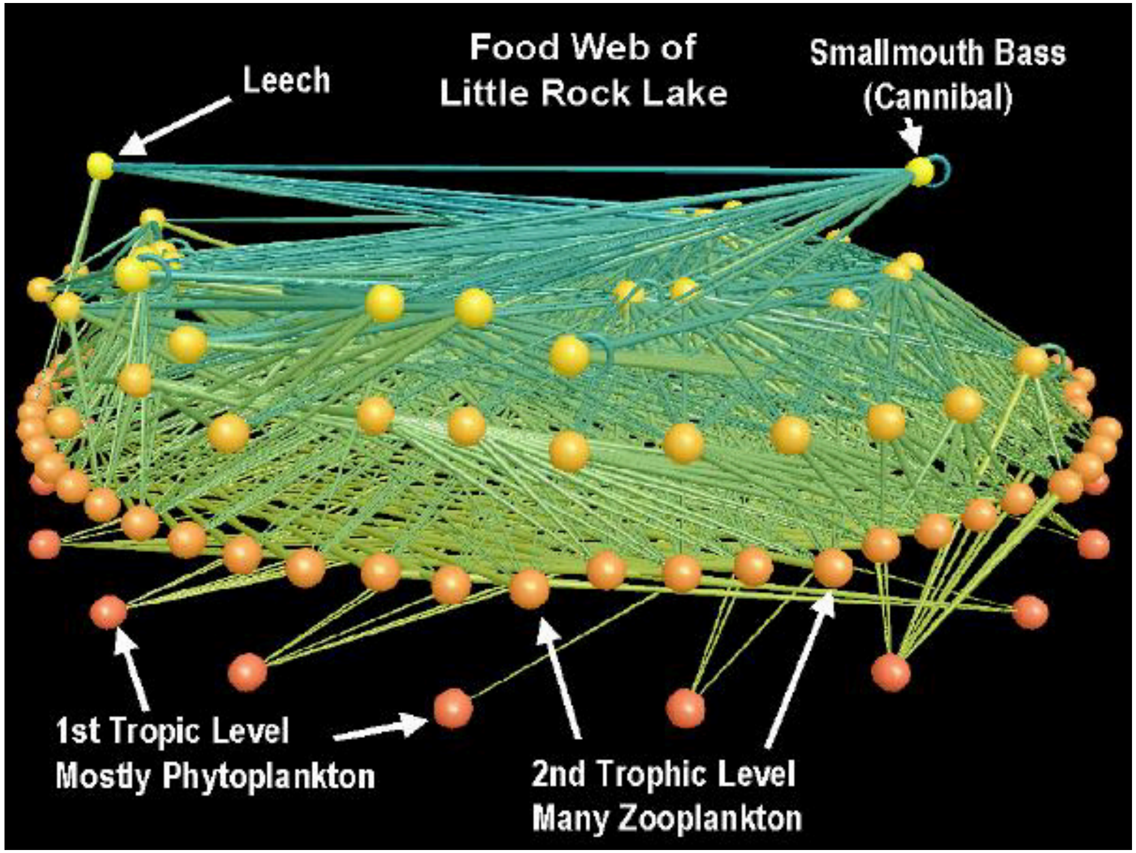


URV (Tarragona, Spain) e-mails inter-exchange network

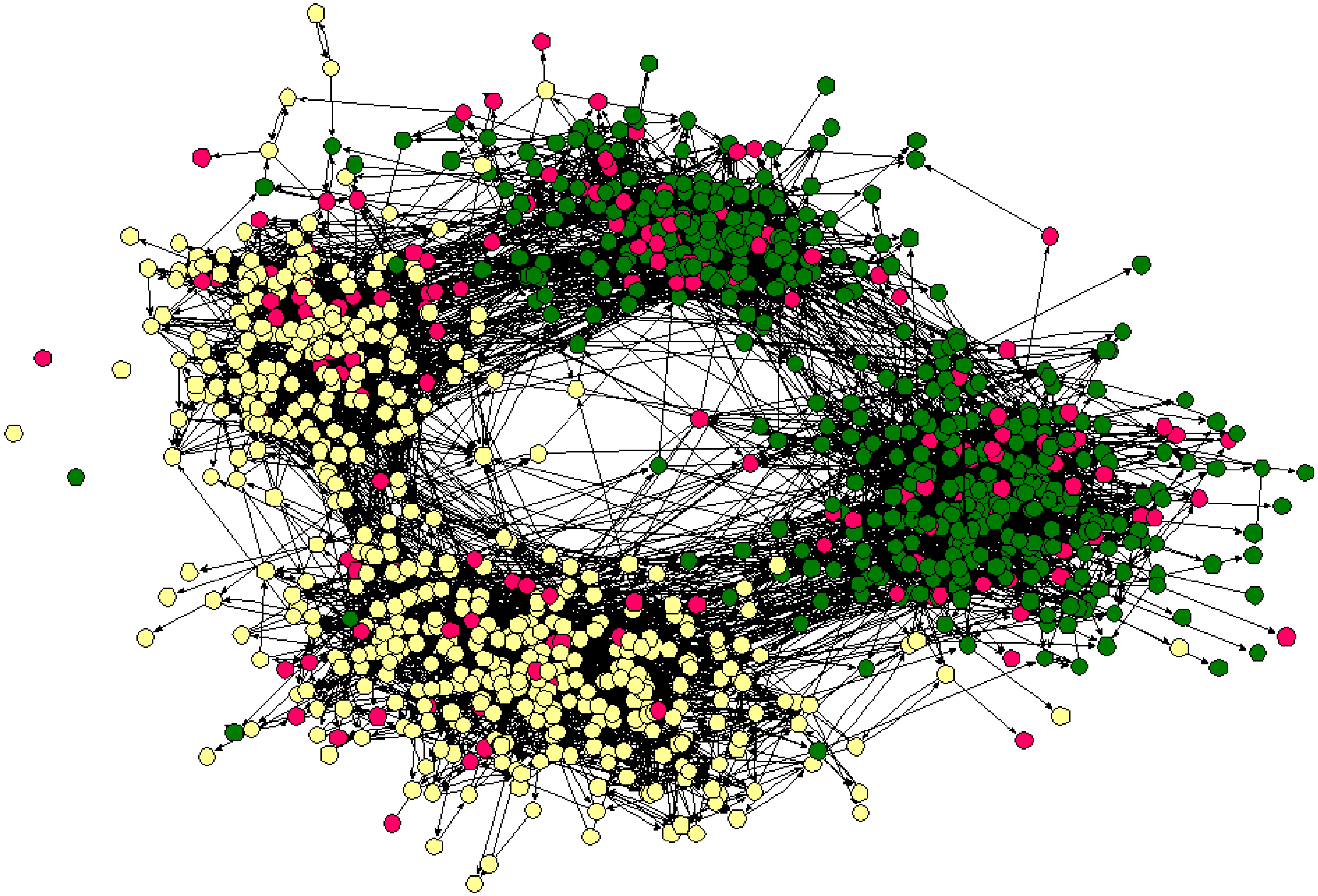
Word-association network



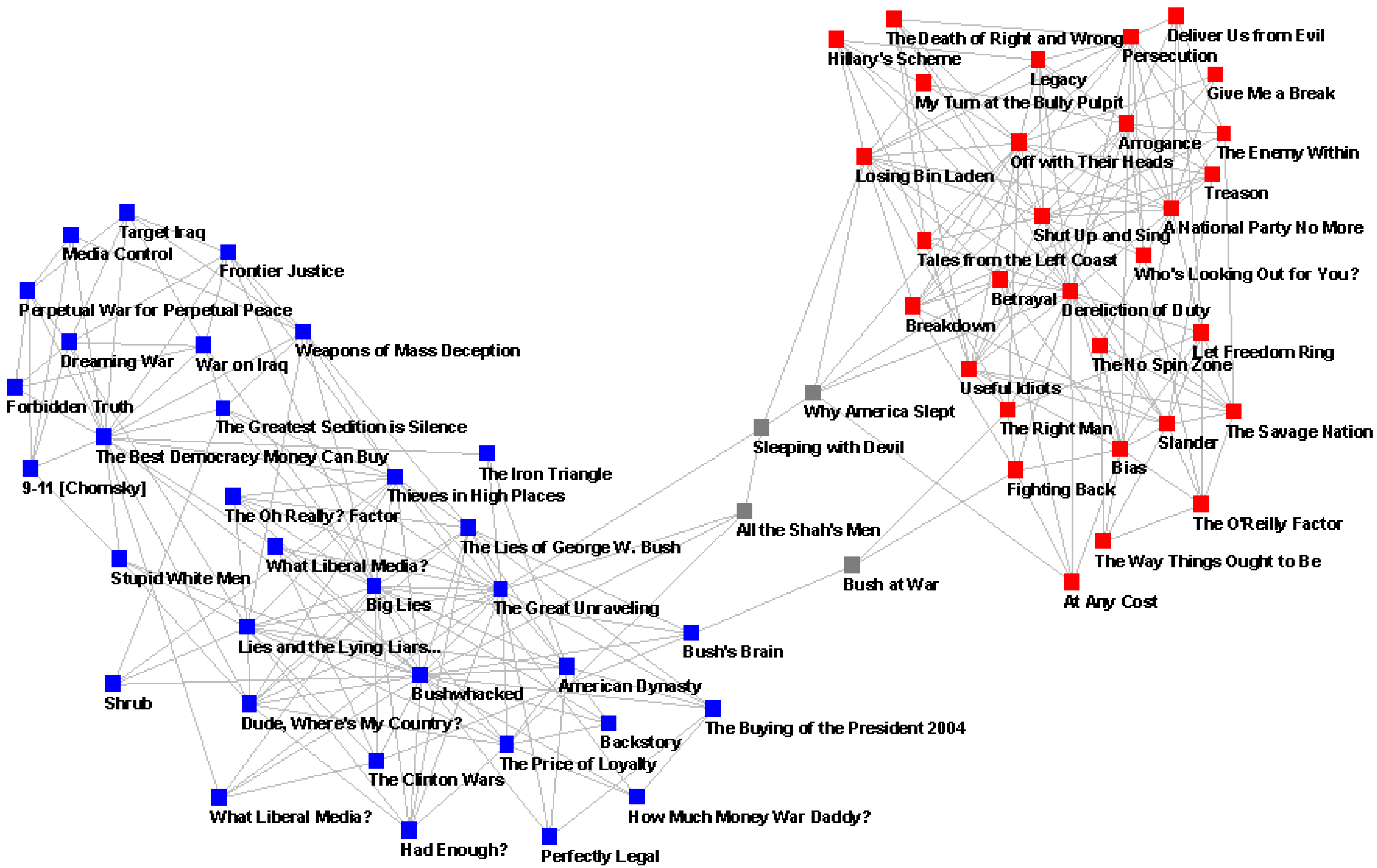
Food-web network



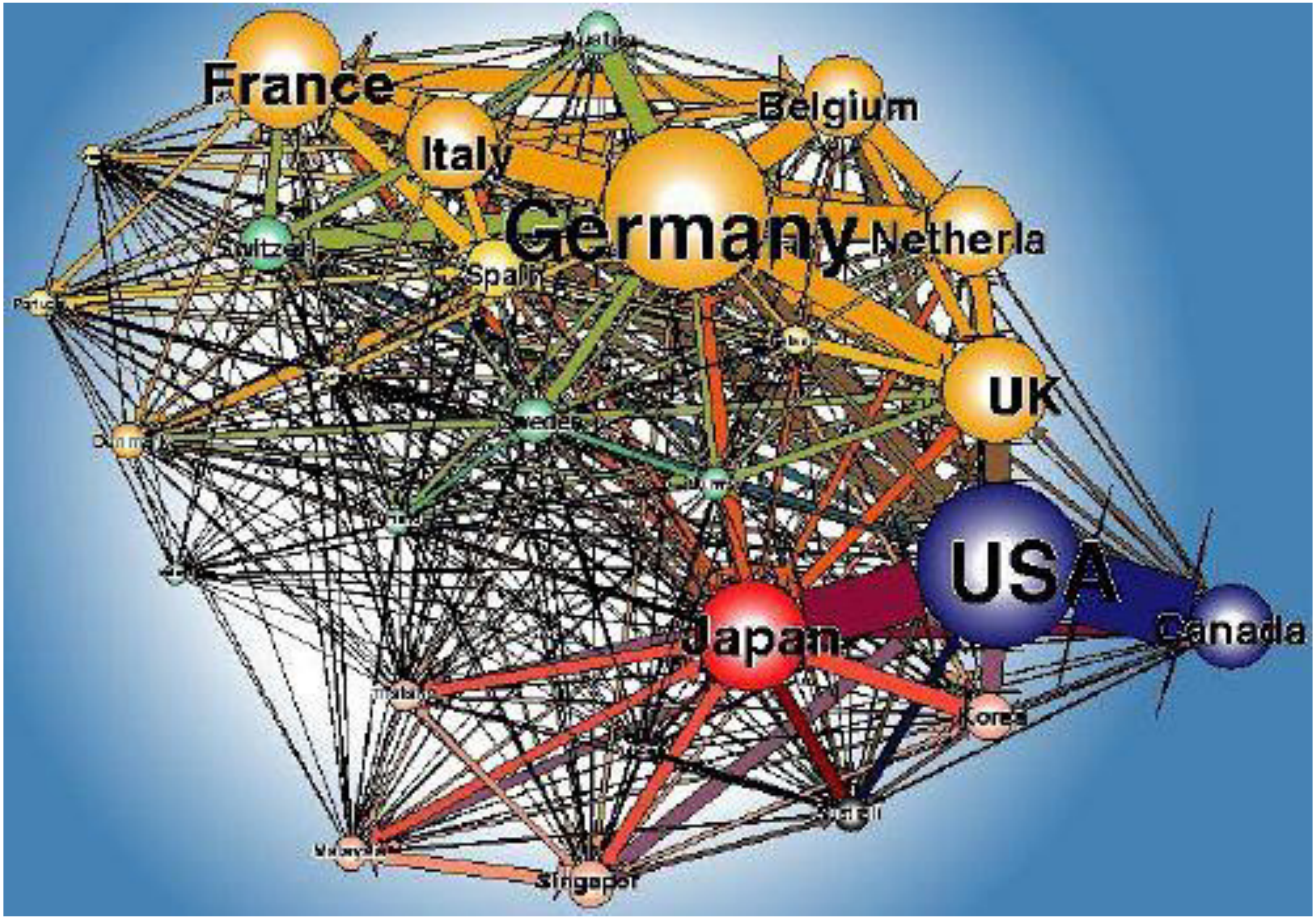
High-school friendship network



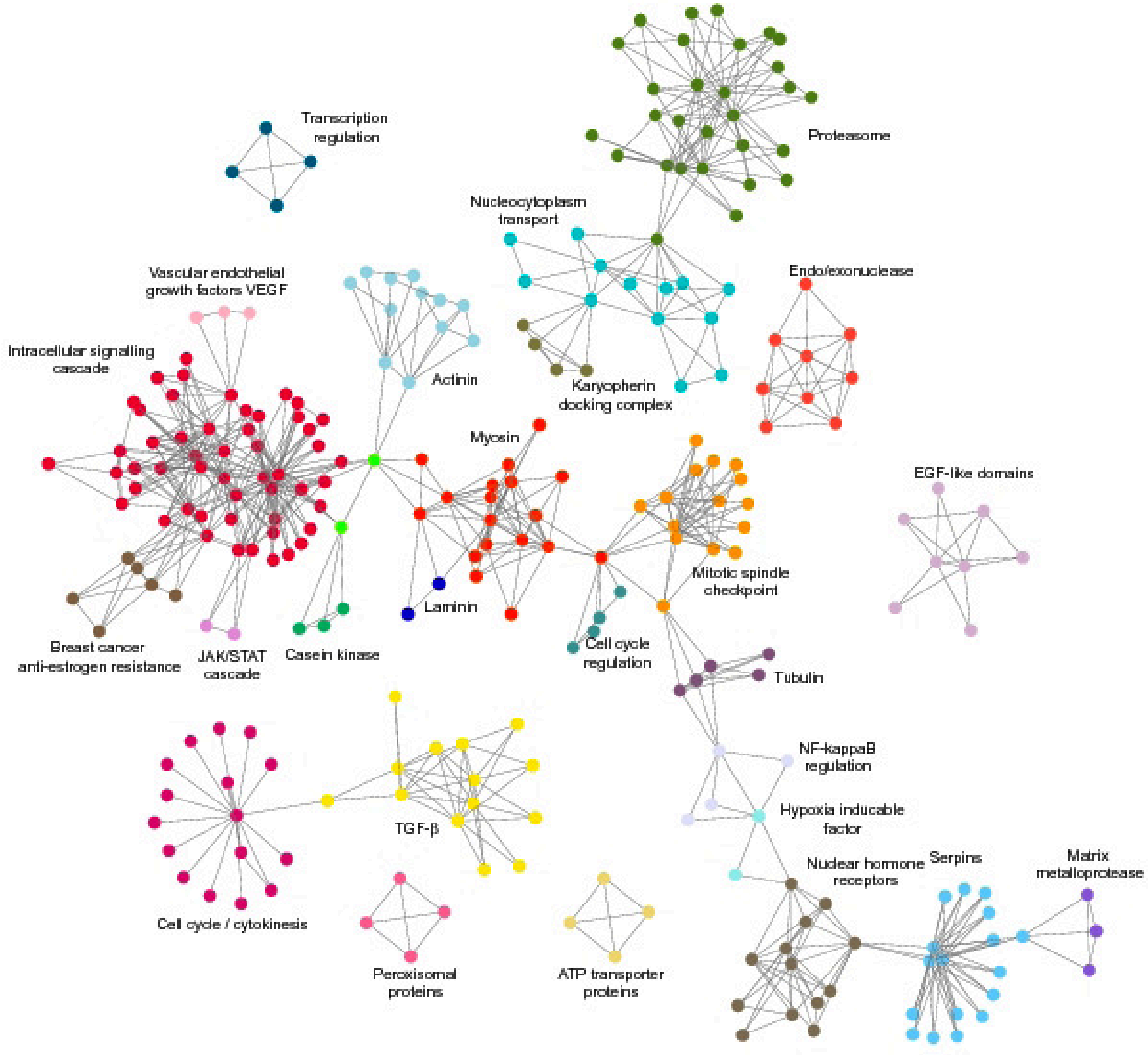
Books on politics network



World Trade Web network



Protein-protein interaction network



Different analysis at different scales

- Microscale: properties of individual nodes or motifs
- Macroscale: statistical properties of the whole network
- Mesoscale: sub-structure of networks "Communities" (also k-cores, n-cliques and n-loops)

Why it is interesting the mesoscale?

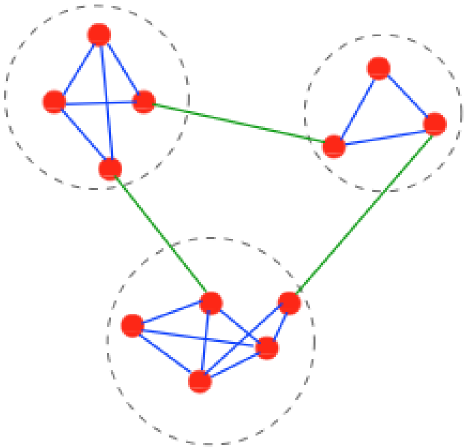
- Similitude with usual coarse-graining in physics
- New statistical properties at this level of description
- Promising way to unravel the interplay between topology and dynamics

Community Detection

- Real complex networks present internal structure (clusters, modules, groups, communities) that can be identified as denser subgraphs within the whole network
- Behind these sub-structures there is information about the functionality of the system that we would like to discover
- The automatic detection of communities is a difficult problem

The problem with the definition of module

- The first problem we face, to determine the sub-structure of networks, is its very definition
- What is a module or community?



Different definitions of modules

- Strong: Every node within a module satisfy the relation k_in>k_out
- Weak: For all nodes in a module the \sum k_in > \sum k_out
- Modules are cliques
- Modules are nodes reachable at distance X
- Etc....

Implicit definition of communities

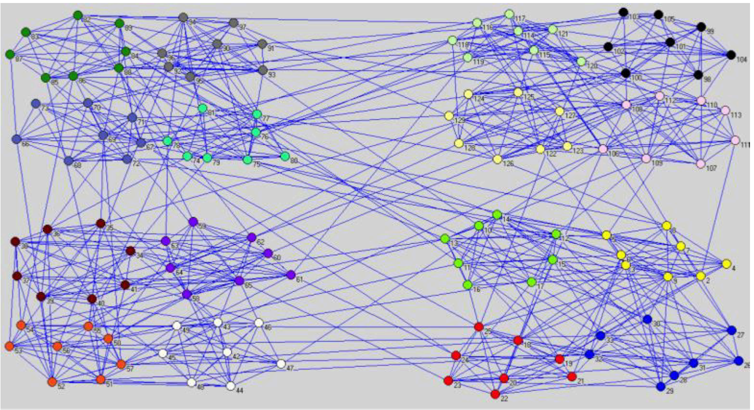
- Modules are the result of maximizing the modularity function (Newman and Girvan)
  - Given a certain disjoint partition of the network, compute the probability of having links between the nodes within a group, minus the same probability if nodes were connected at random but preserving its degree

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(C_i, C_j).$$

What is the role of topological structure in the dynamics?

- Syncronization in complex networks
- Our approach:
  - Topology: mesoscale description, community structure
  - Dynamics: transient towards synchronization

Transient towards synchronization

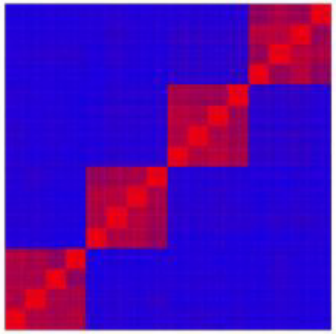


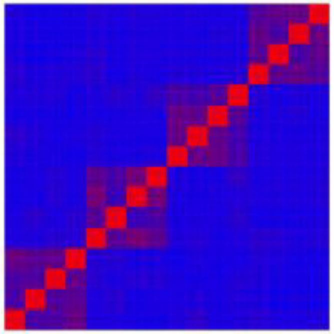
Kuramoto's model

$$\frac{d\ {}_i}{dt} = K \quad {}_{j \quad (i)} \sin(\ {}_j \qquad {}_i)$$

Correlation matrix

$$\rho_{ij}(t) = < cos(\theta_i(t) - \theta_j(t)) >$$





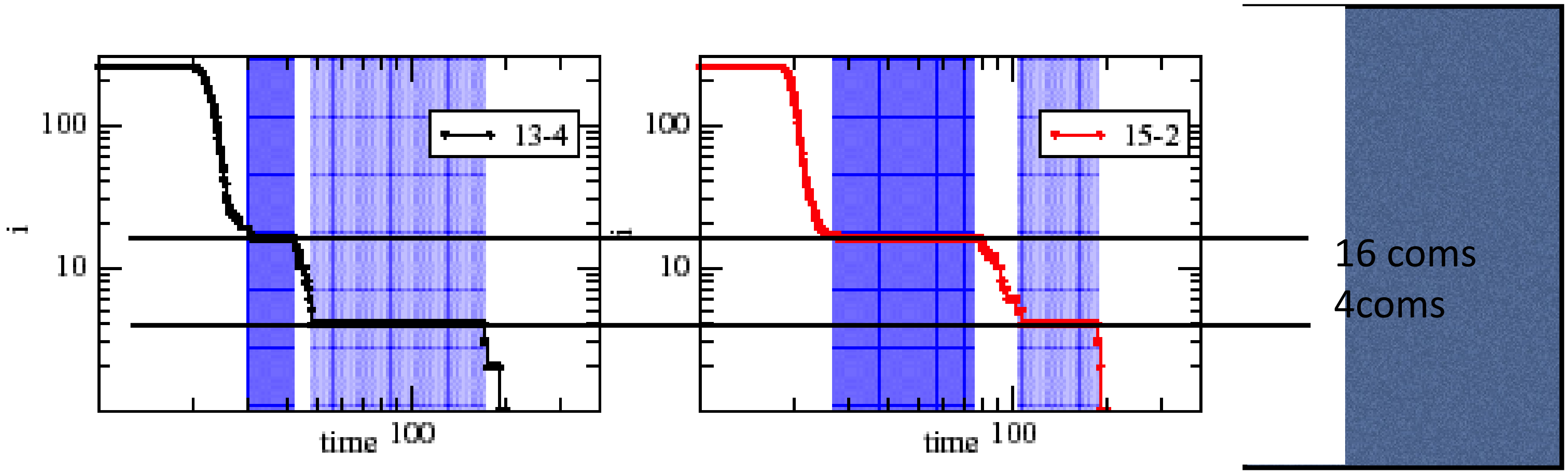
All hierarchical partitions are revealed through time !

13-4-1 15-2-1

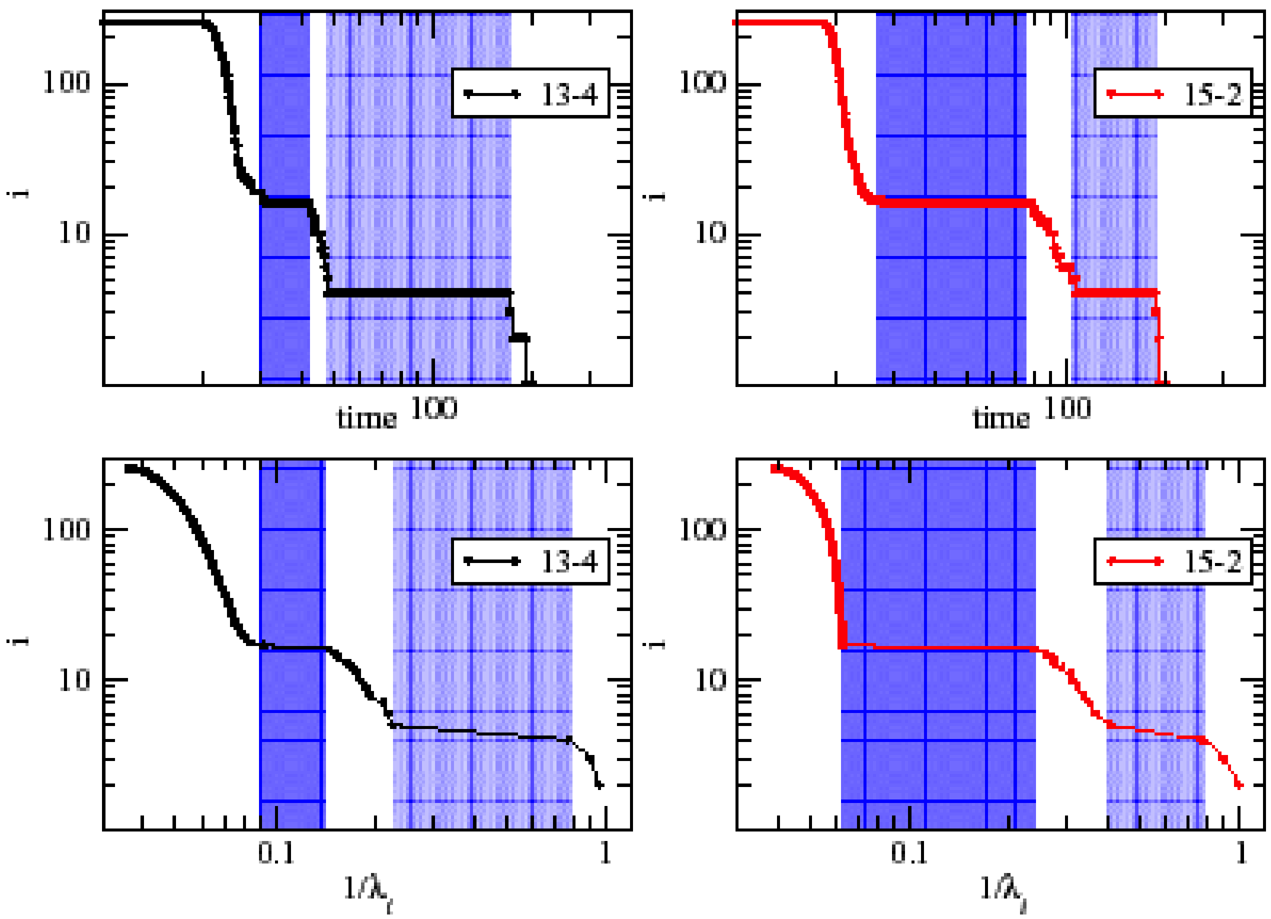
Analysis of the correlation matrix

Dynamic connectivity matrix

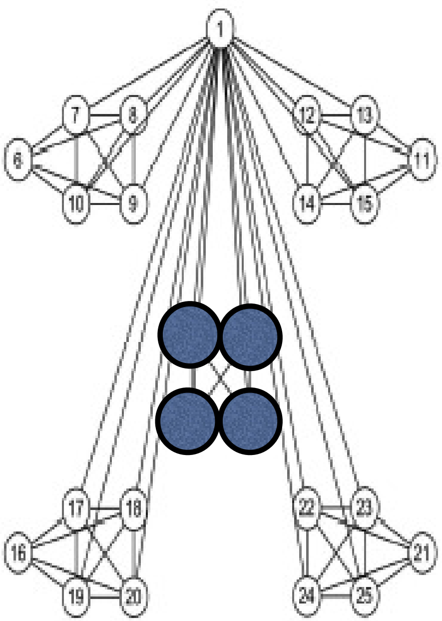
$$\mathcal{D}_T(t)_{ij} = \begin{cases} 1 & \text{if } \rho_{ij}(t) > T \\ 0 & \text{if } \rho_{ij}(t) < T \end{cases}$$

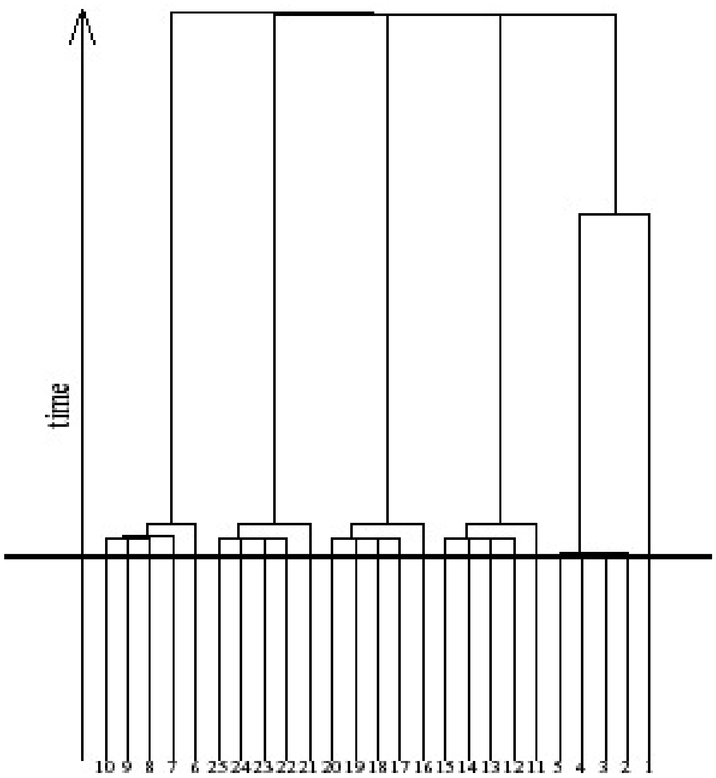


The connection between topology and dynamics

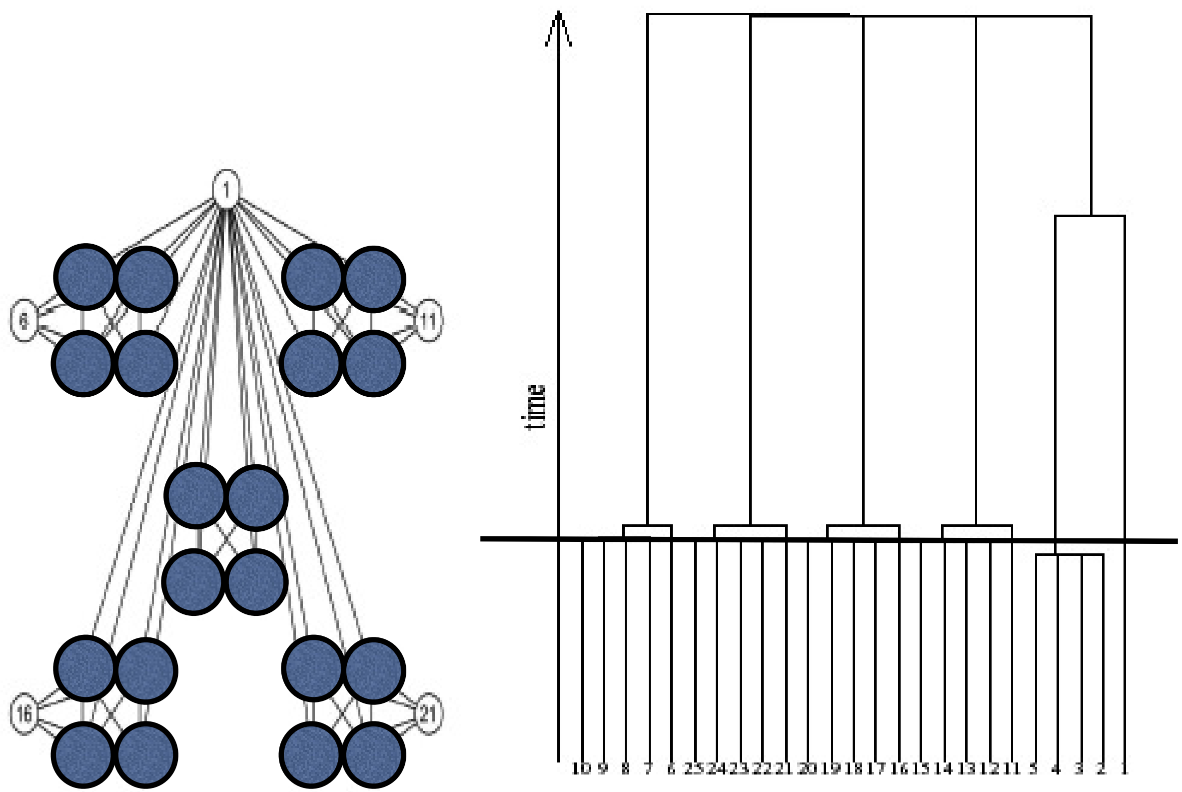


The connection between topology and dynamics

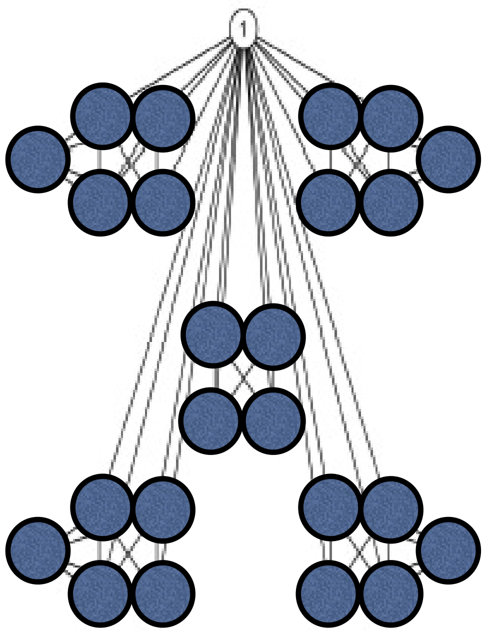


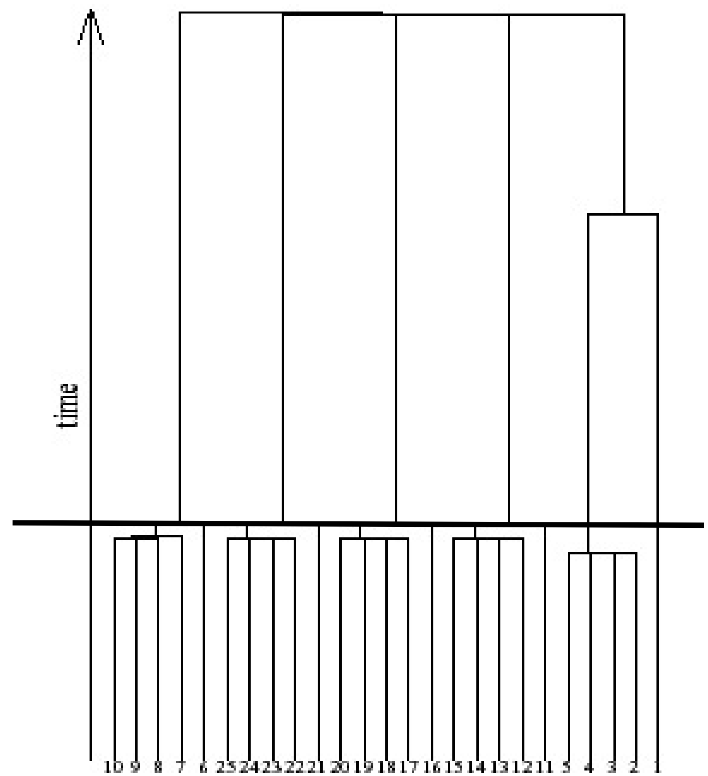


The connection between topology and dynamics

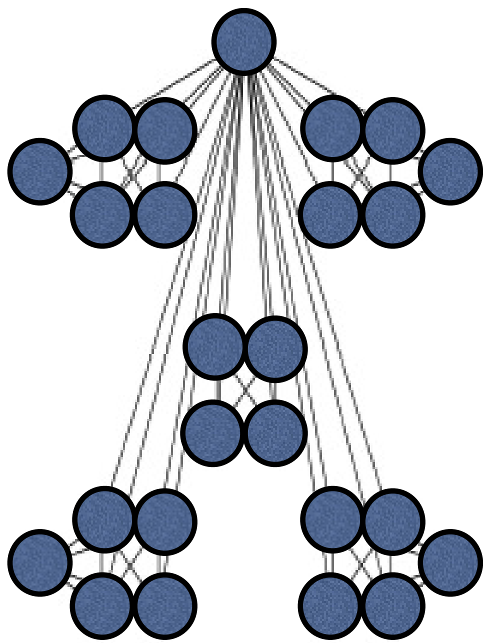


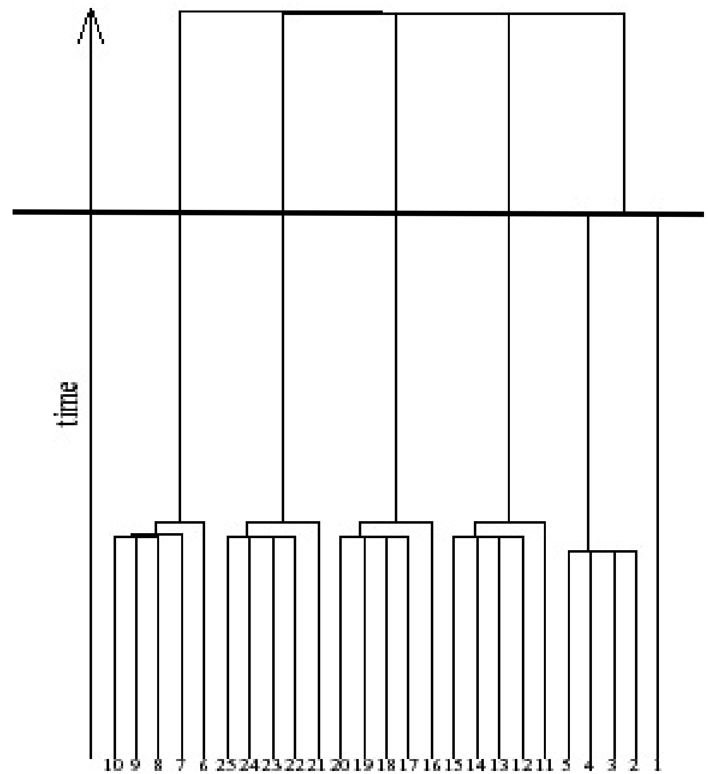
The connection between topology and dynamics





The connection between topology and dynamics





Understanding the problem

Linearized version of the problem

$$\frac{d\theta_i}{dt} = -k \sum_j L_{ij}\theta_j$$

Solution in terms of normal modes

$$\varphi_i(t) = \sum_j B_{ij}\theta_j = \varphi_i(0)e^{-\lambda_i t}$$

Understanding the problem

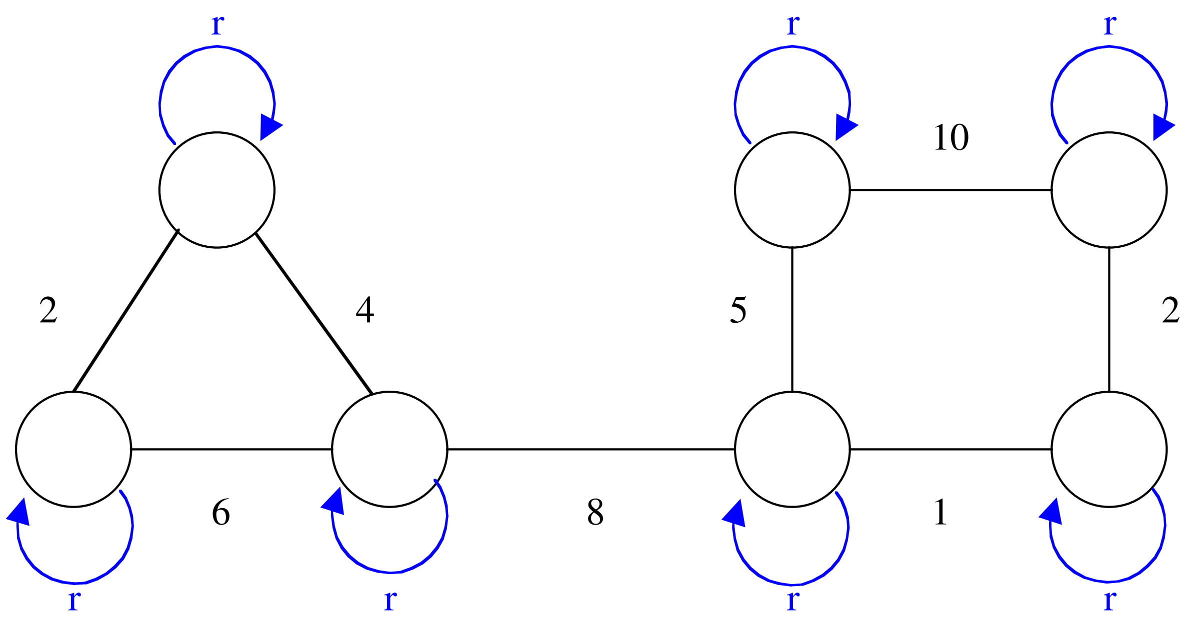
- If we rank the equations from fastest to lower modes we observe a cascade of synchronization restrictions
- The largests eigenvalues involve “topological average” equations
- Similar eigenvalues involve closed equations that synchronize groups of oscillators at the same time scale

How to unravel the mesoscale from the topology?

- Idea:
- To provide each node with a global magnitude that will tune its strength at different intermediate topological scales without affecting its structural topology, i.e. preserving P(k), vertex-vertex correlations and spectrum
- Find community structure at each resolution level. We overcome the resolution limit first found by Fortunato & Barthelemy

How to unravel the mesoscale from the topology?

We introduce a resistance "r" of nodes to become part of a group



Some notation

Weighted matrix

$$w'_{ij} = \begin{cases} w_{ij} & \text{if} \quad i \neq j\,, \\ r & \text{if} \quad i = j\,. \end{cases}$$

Node strength

$$w'_i = \sum_j w'_{ij} = w_i + r$$

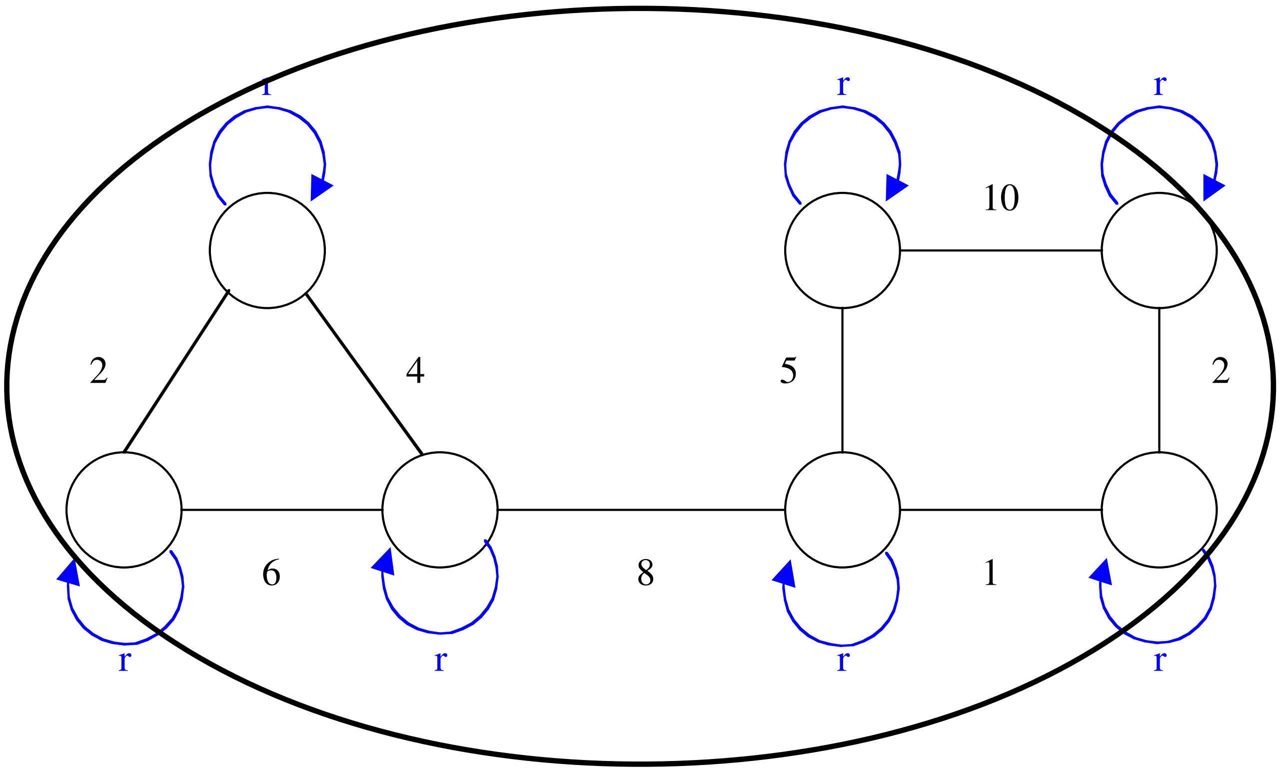
Total strength

$$2w' = \sum_i w'_i = 2w + Nr$$

$$Q(r) \;\; = \;\; \frac{1}{2w'} \sum_i \sum_j \left(w'_{ij} - \frac{w'_i w'_j}{2w'} \right) \delta(C_i, C_j)$$

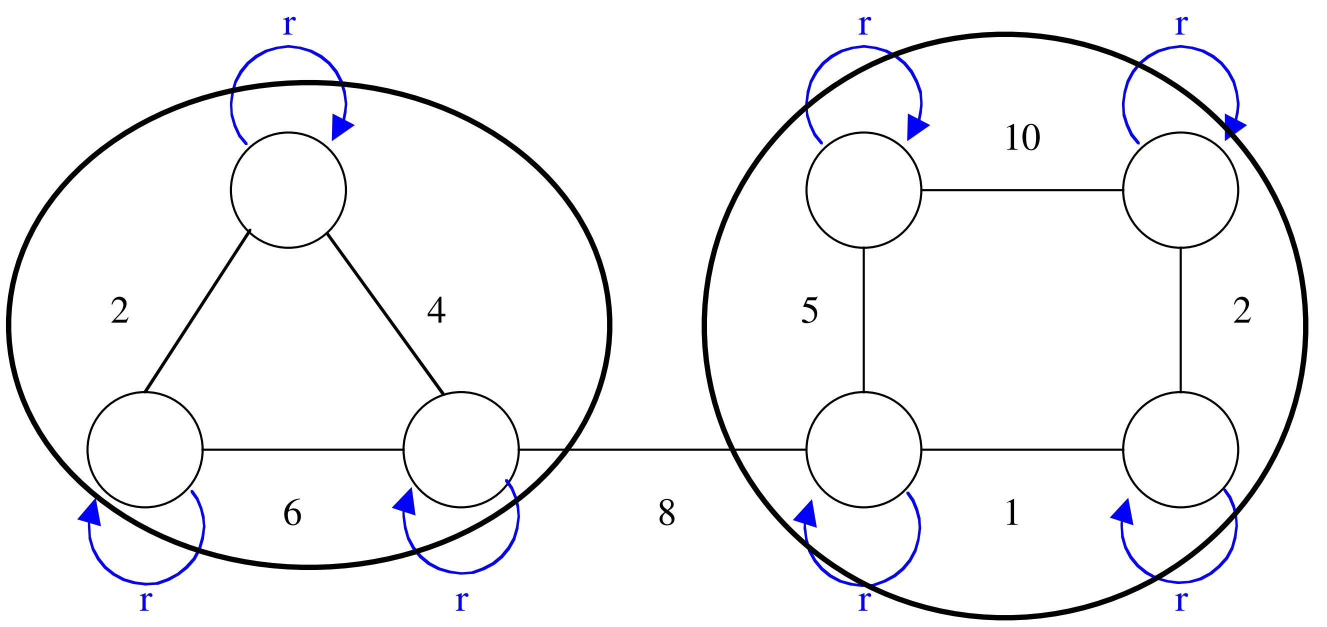
Modularity, original formulation

Community structure at different values of r



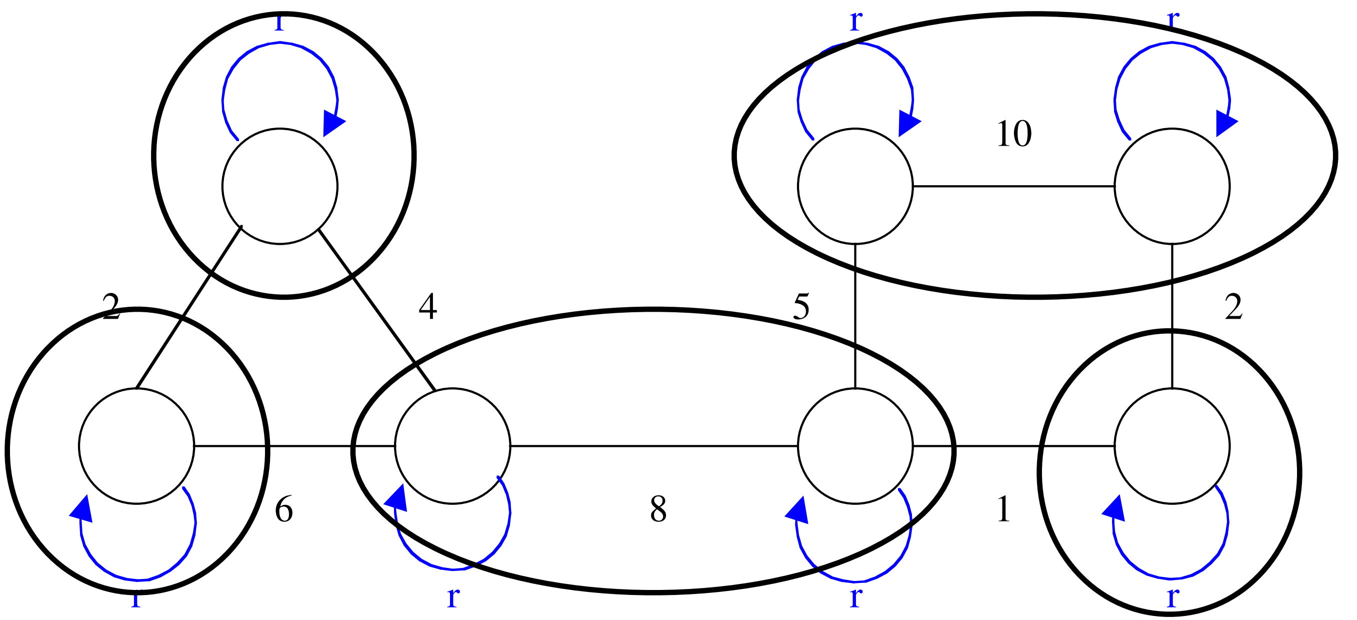
r_min

Community structure at different values of r



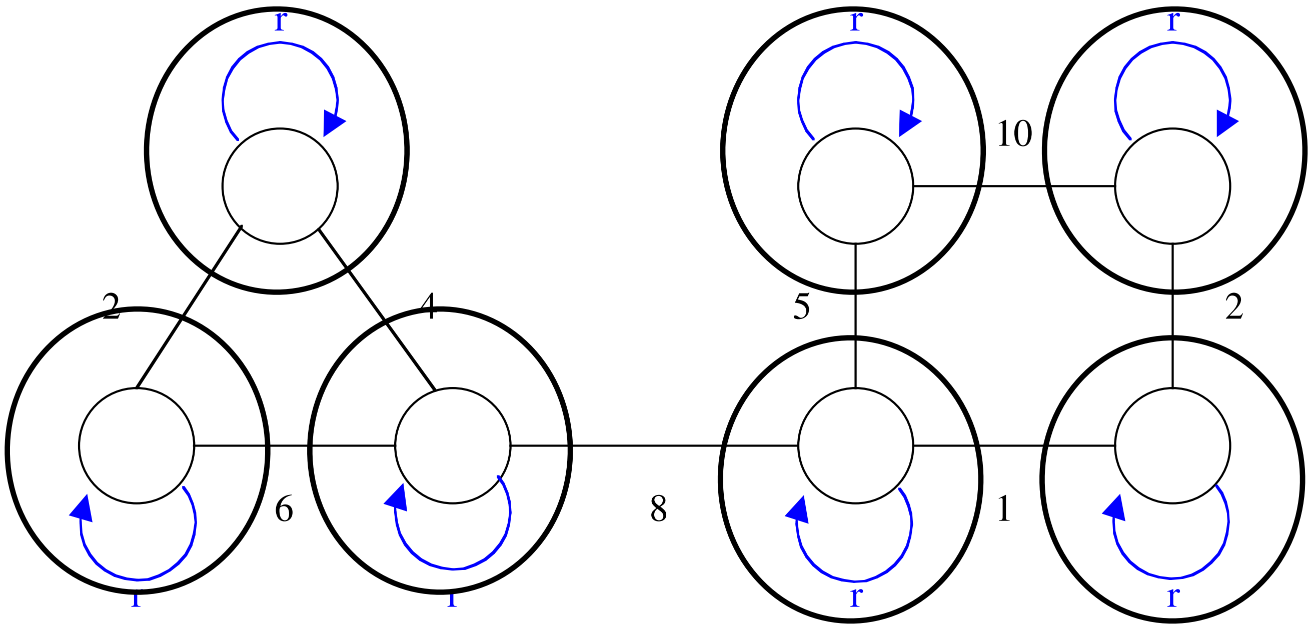
r1 intermediate

Community structure at different values of r



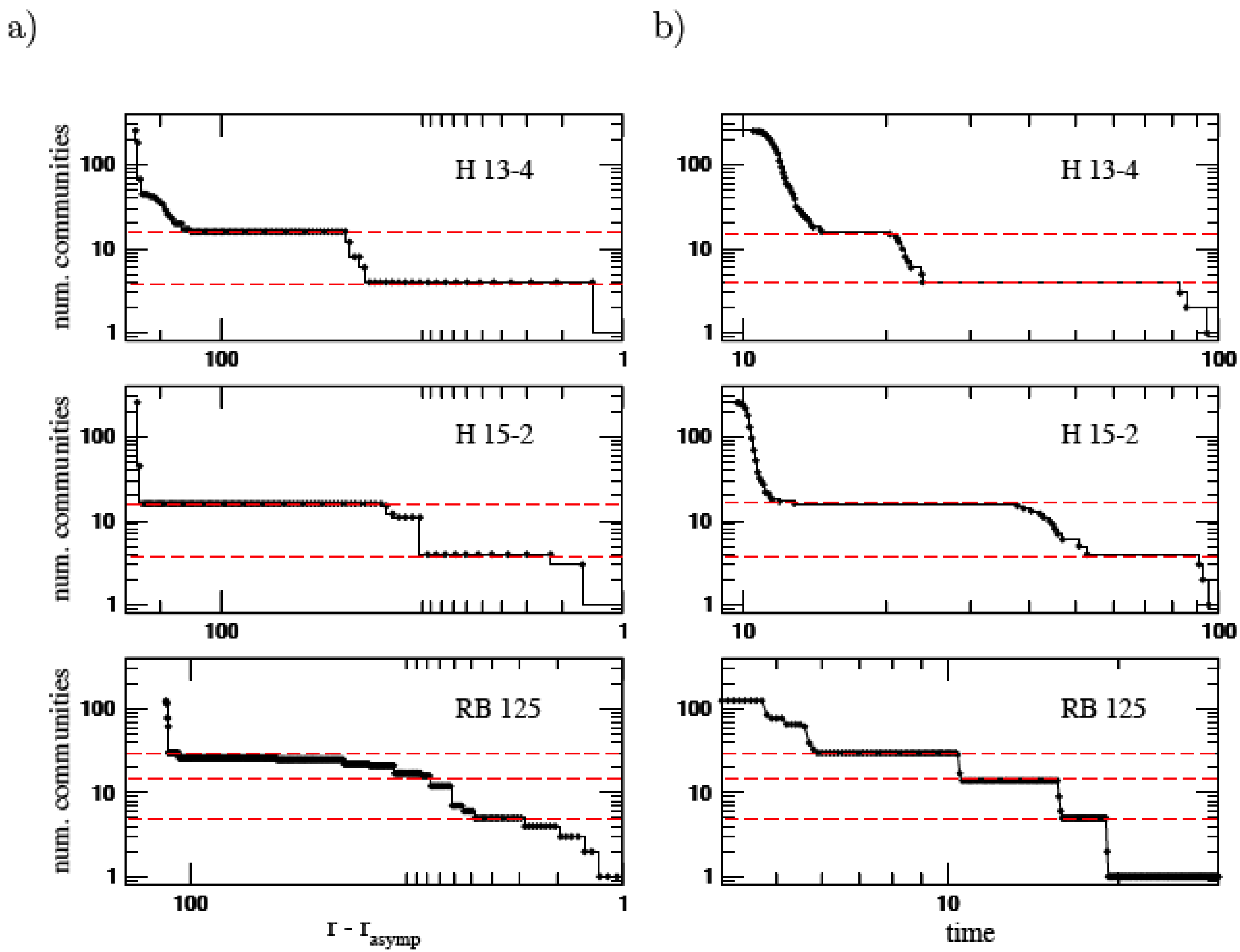
r2 intermediate

Community structure at different values of r

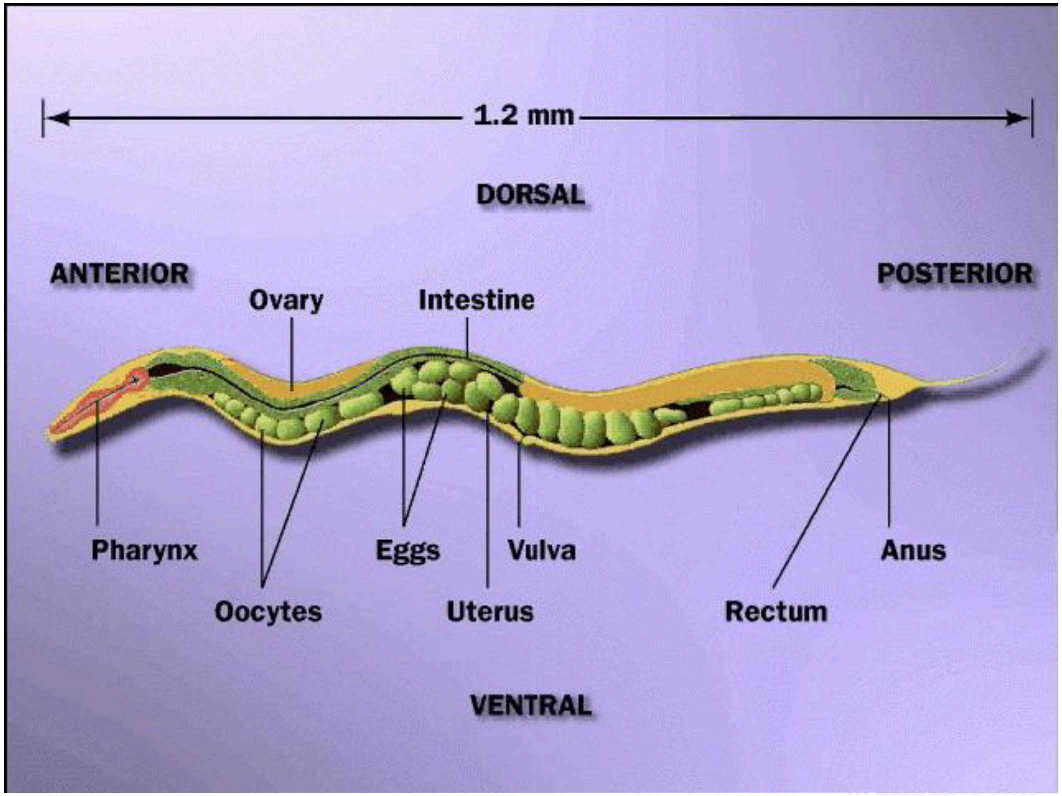


r_max

Comparing the dynamic and static mesoscale



Applications: mesoscale of C.elegans neural network

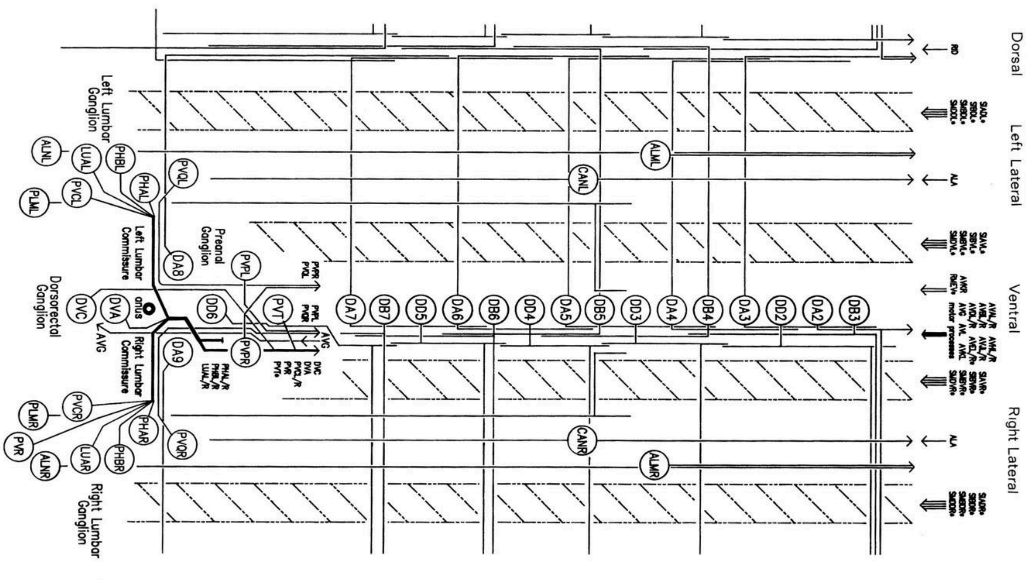


Why to study C. elegans neural network?

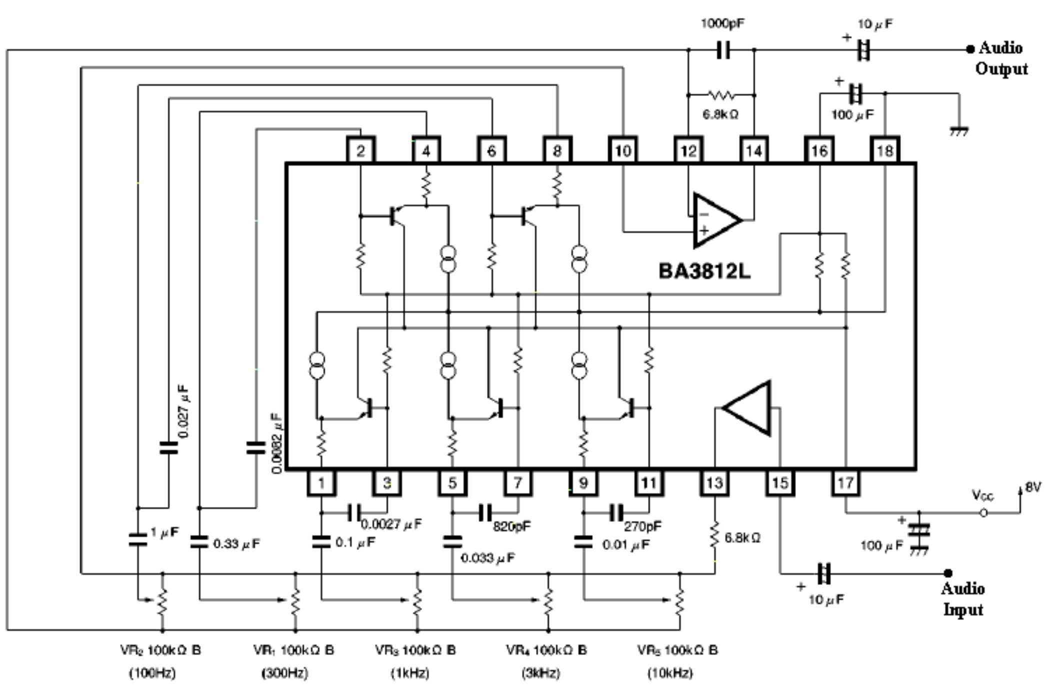
- C. elegans is the most studied living organism in history, the detail of knowledge is exhaustive
- Neural system not plastic, genetically designed
- Still missing the structure-functionality relationship

Kind of reverse engineering

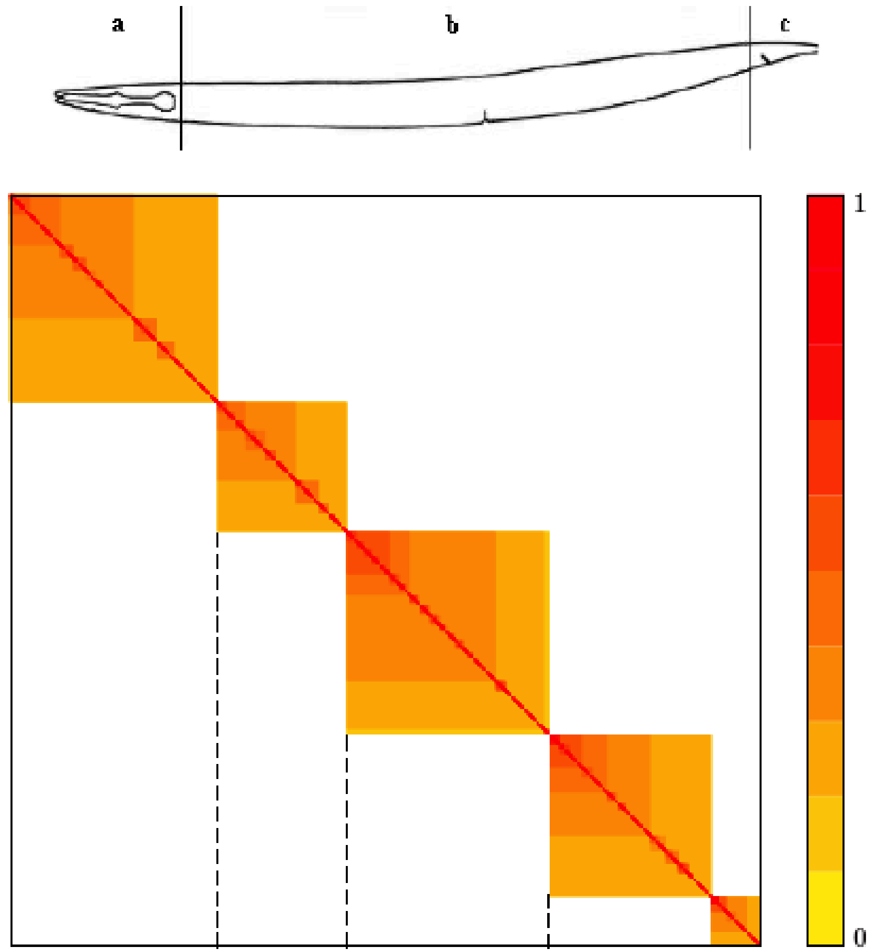
C.elegans neural system



Chip electronic circuit

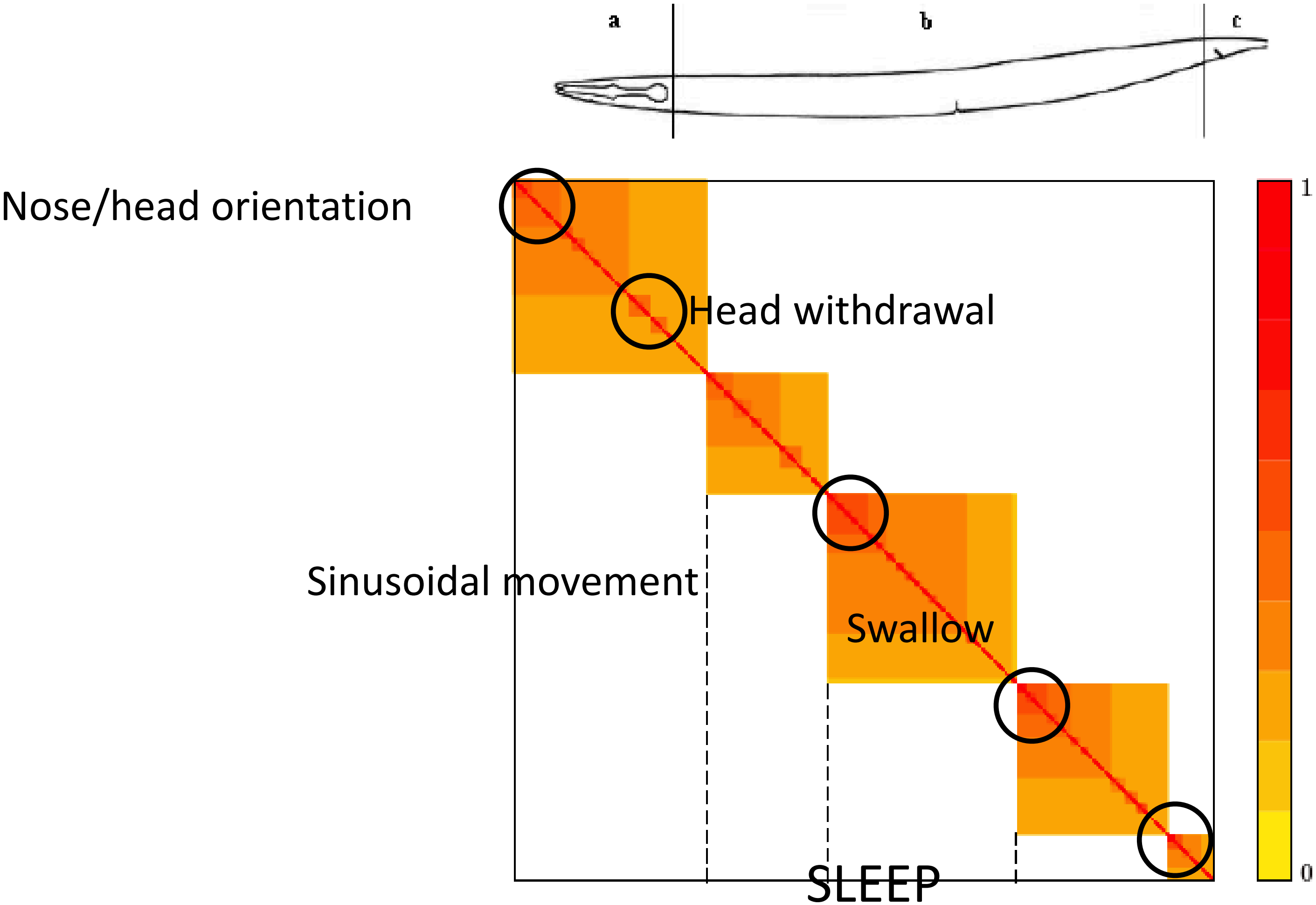


Mesoscale: location correlations



a	96.4 %	92.0 %	15.2 %	40.6 %	36.8 %
b	1.2 %	0.0 %	82.3 %	32.8 %	26.4 %
c	2.4 %	8.0 %	2.5 %	26.6 %	36.8 %

Mesoscale: functional correlations



a	96.4 %	92.0 %	15.2 %	40.6 %	36.8 %
b	1.2 %	0.0 %	82.3 %	32.8 %	26.4 %
c	2.4 %	8.0 %	2.5 %	26.6 %	36.8 %

Applications: Data Clustering

- Clustering is the unsupervised classification of patterns (observations, data items, or feature vectors) into groups (clusters).

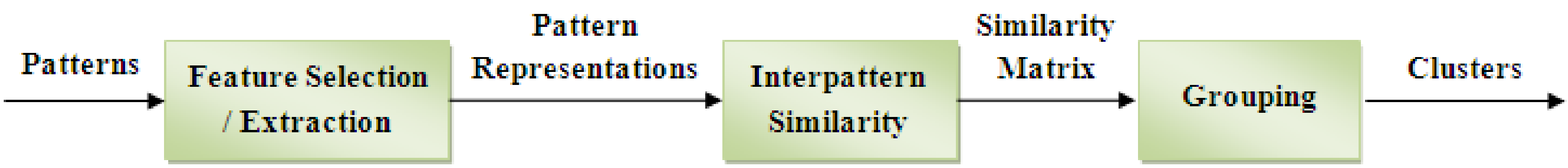
Some applications:

Marketing: finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records.

WWW: document classification; clustering weblog data to discover groups of similar access patterns.

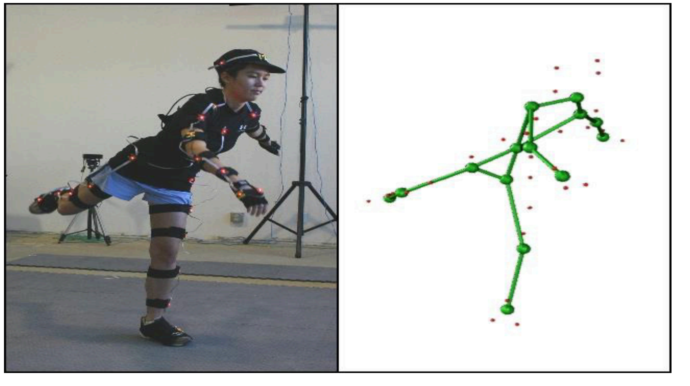
Biology: classification of plants and animals given their features.

Applications: Data Clustering



Feature selection / extraction

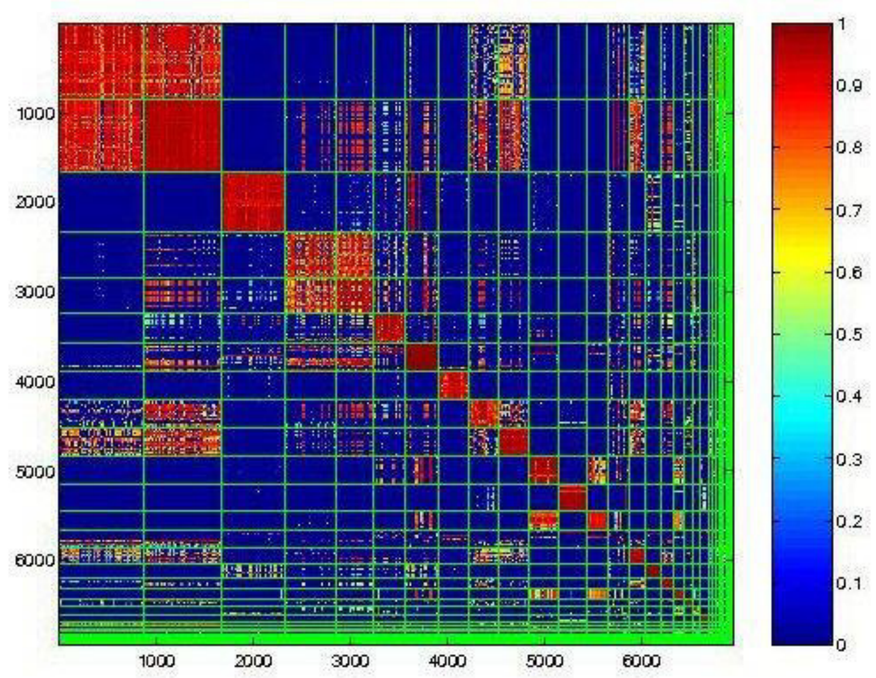
- Stands for the process of selecting the most representative variables for the classification process.
- Many popular search approaches use heuristics which iteratively evaluates a candidate subset of features, then modifies the subset and evaluates if the new subset is an improvement over the old using a scoring metrics.



Interpattern Similarity

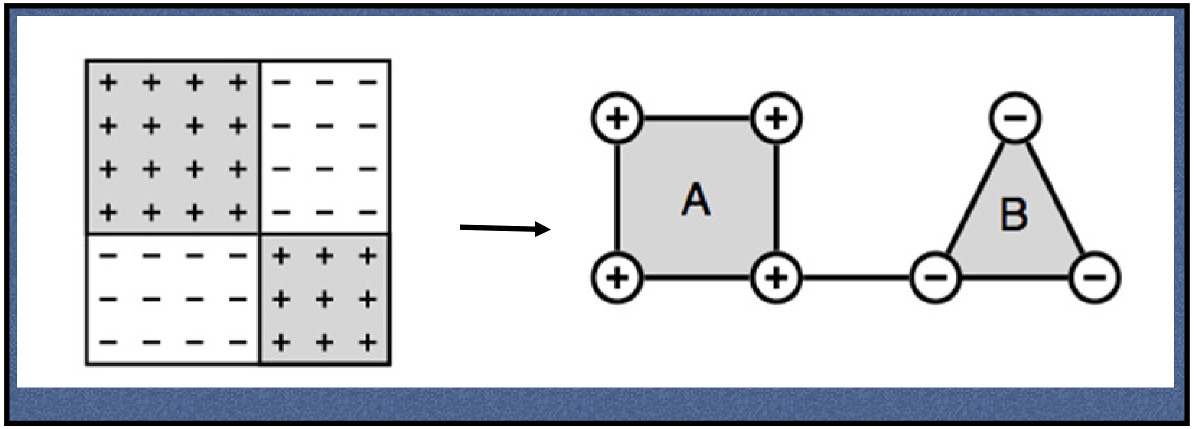
- Similarity Matrix: A matrix of scores expressing the similarity between two data points in terms of distances (e.g Minkowski distance).

$$d_p(x_i, x_j) = \left(\sum_{k=1}^{d} \|x_{i,k} - x_{j,k}\|^p \right)^{\frac{1}{p}}$$



Modularity based community analysis

- A representation of a matrix in terms of networks is likely to be analyzed using community detection algorithm.
- Using a reformulation of modularity for signed links



IRIS flower data set
Introduced by Sir R. A. Fisher (1936)

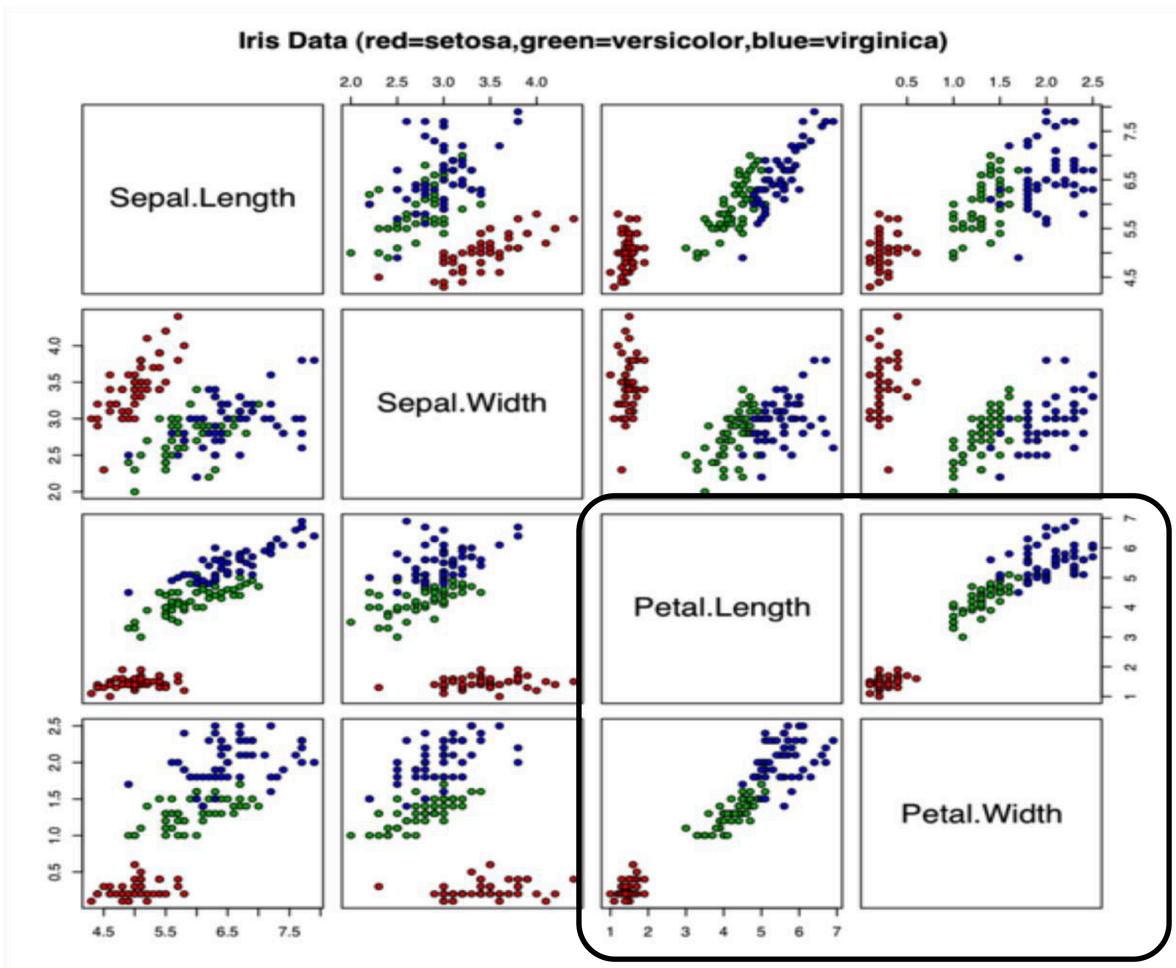
Setosa

Versicolor

Virginica

- Problem: to automatically classify the three types, using a database of:
  - 50 samples of each one, characterized by 4 dimensional vectors, corresponding to length and width of sepal and petal

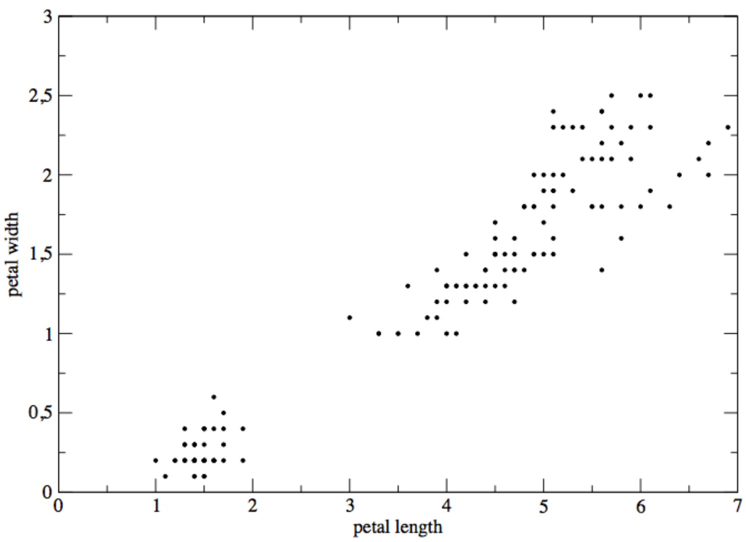
Variable selection

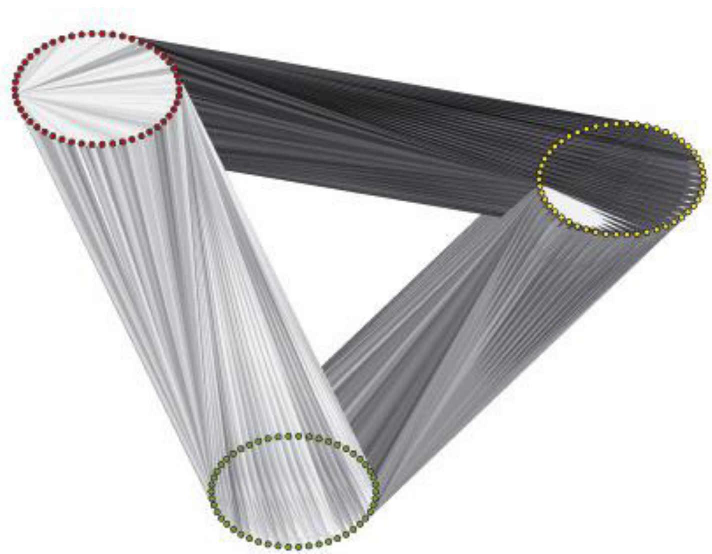


Similarity matrix

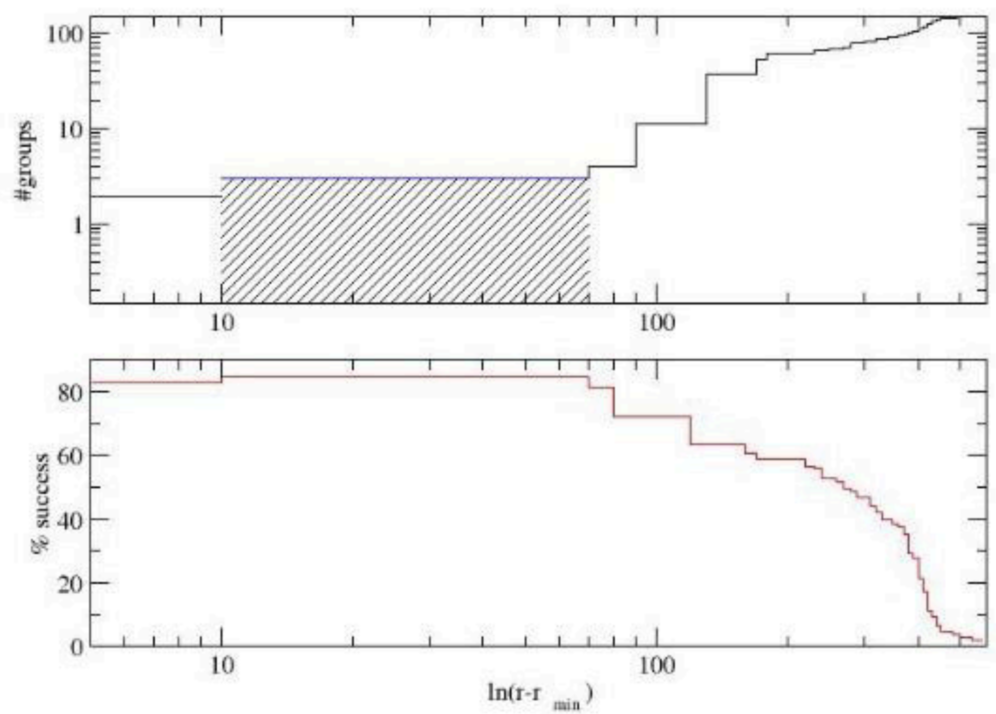
We propose to compute the similarity between data using distances in PCA

$$s_{ij} = d_{ij} - \langle d \rangle$$





Classification results using multi-resolution of modularity



Conclusions

- We have proposed a method to compute the topological mesoscale in complex networks
- The method provide with substructures compatible with those meta-stable patterns observed in the synchronization dynamics
- The substructures obtained in the C.elegans case study allow for a tentative classification of structure-function relations
- The use of community analysis in data clustering is a promising tool.

http://deim.urv.cat/~aarenas

References

Mesoscopic analysis of networks: applications to exploratory analysis and data clustering C. Granell, S. Gómez and A. Arenas, Chaos: An Interdisciplinary Journal of Nonlinear Science, 21, 016102 (2011)

Optimal map of the modular structure of complex networks A. Arenas, J. Borge-Holthoefer, S. Gomez and G. Zamora-Lopez, New Journal of Physics, 12, 053009 (2010)

Categorizing words through semantic memory navigation J. Borge-Holthoefer and A. Arenas, European Physical Journal B, 74(2), 265 (2010)

Analysis of community structure in networks of correlated data S. Gómez, P. Jensen and A. Arenas, Physical Review E 80, 016114 (2009)

Analysis of the structure of complex networks at different resolution levels A. Arenas, A. Fernández and S. Gómez, New Journal of Physics 10, 053039 (2008)

Diffusion processes on multiplex networks

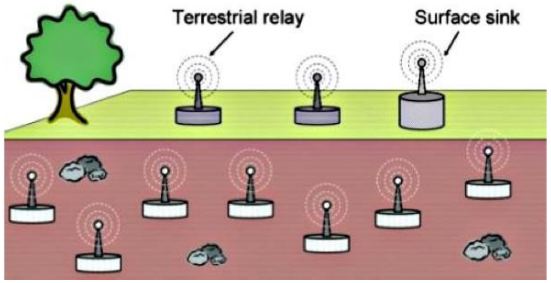
Alex Arenas
Universitat Rovira i Virgili (Tarragona, Spain)

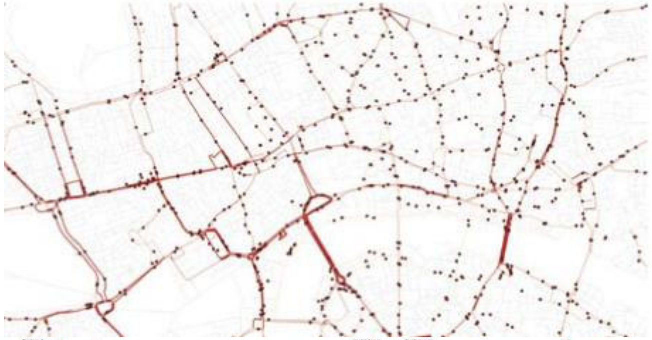
Alephsys

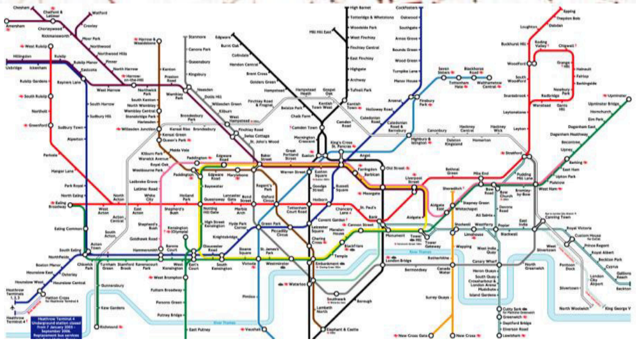
Multiplex networks (EU projects: PLEXMATH, LASAGNE and MULTIPLEX)

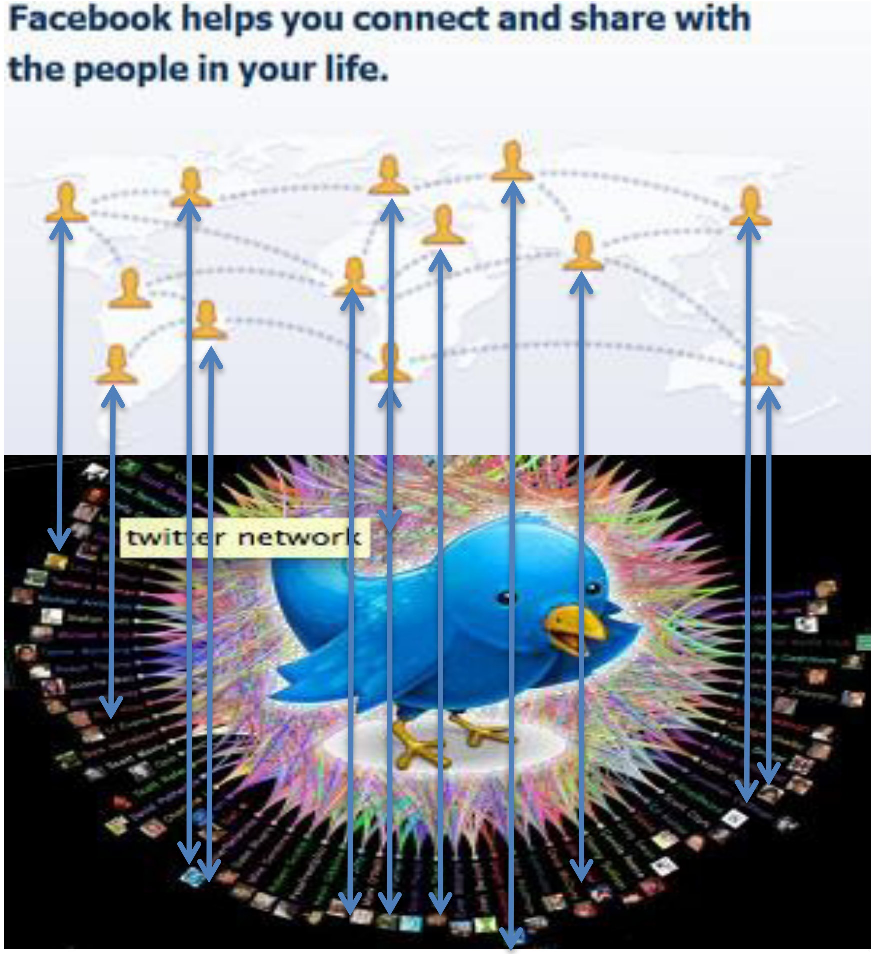
- Definitions
  - *Simple network* (or *graph*)
  - *Multigraph*: multiple links between nodes, equivalent to a weighted network with integer weights
  - *Hypergraph*: links may connect any number of nodes
  - *Simple network with colored edges*: links of different classes
  - *Multiplex*: network in layers, and with connections between layers; the interconnections between layers are only between a node and its counterpart in the other layer (i.e. the same node)

Instances of multiplex

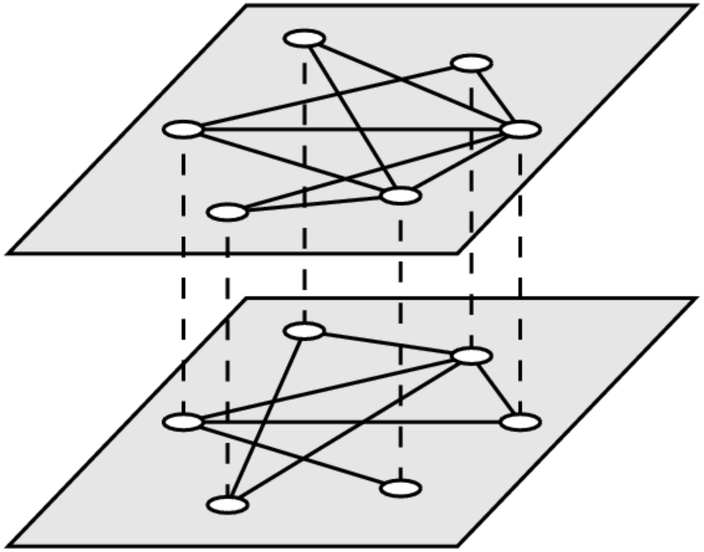




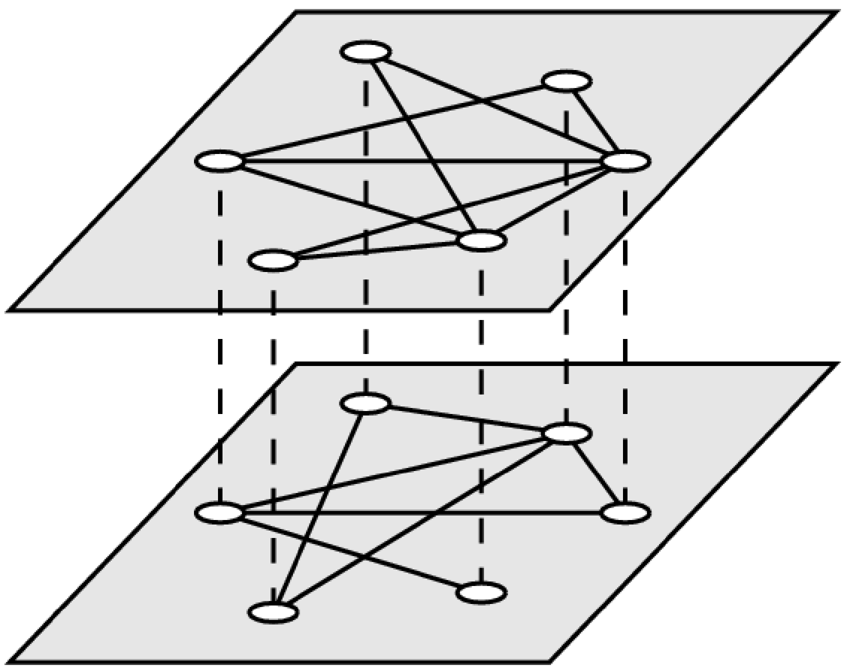




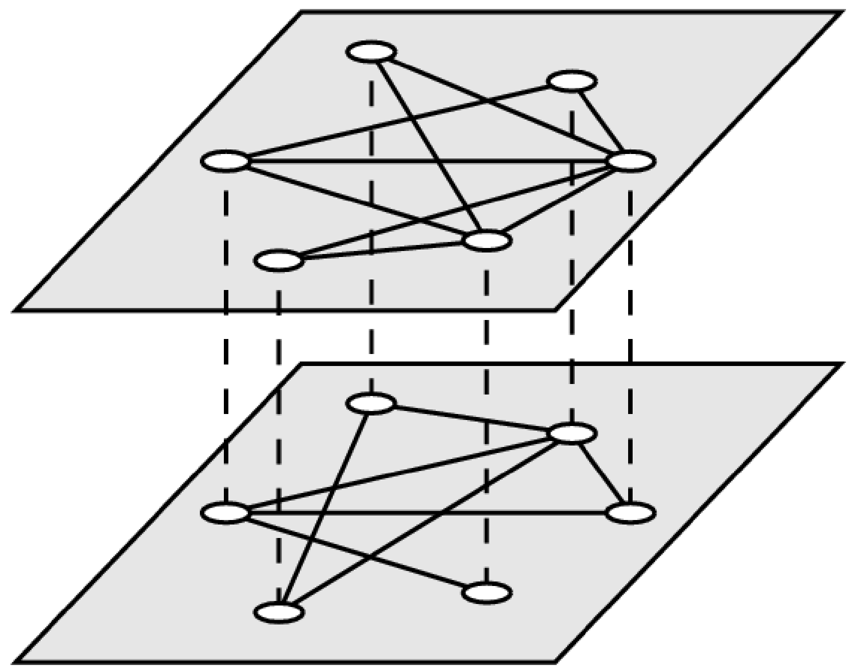
- Interest of multiplex networks:
  - Represent a “next step” in the study of the complexity of networked systems
  - Non trivial correlations emerge between different layers
  - The preceding studies on percolation in interdependent networks show that coupling networks can have unexpected emergent behaviors
  - Understood dynamical processes can show new characteristics because of the multiplex structure
  - The theory of multiplex must cover the current theory developed for “simplex” networks



- Dynamic processes on multiplex networks
  - Physics: Diffusion processes
  - Social sciences: Prisoner dilemma
  - Biology: Boolean dynamics



- **Diffusion in a multiplex network**
  - Objective
    - Which is the diffusion time scale in the multiplex?
      - Diffusion on multiplex networks, S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C.J. Perez-Vicente, Y. Moreno and A. Arenas, Physical Review Letters, 110, 028701 (2013)
  - Hypothesis
    - Only two layers
    - Connected weighted networks (W_1 and W_2) at each layer
    - Diffusion rates: D_1, D_2, D_x



- Mathematical description
  - Connectivity of the multiplex

$$\mathcal{W} = \left(\begin{array}{c|c} D_1 W_1 & 0 \\ \hline 0 & D_2 W_2 \end{array}\right) + D_x \left(\begin{array}{c|c} 0 & I \\ \hline I & 0 \end{array}\right)$$

  - Laplacians at each layer

$$L_K = S_K - W_K \qquad (S_K)_{ii} = s_i^K = \sum_j w_{ij}^K$$

  - Supra-Laplacian of the multiplex

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{D} = \left(\begin{array}{c|c} D_1 L_1 & 0 \\ \hline 0 & D_2 L_2 \end{array}\right) + D_x \left(\begin{array}{c|c} I & -I \\ \hline -I & I \end{array}\right)$$

  - Diffusion equation

$$\dot{\mathbf{x}} = -\mathcal{L}\mathbf{x}$$

  - Simplification (without loss of generalization)

$$D_1 = D_2 = 1$$

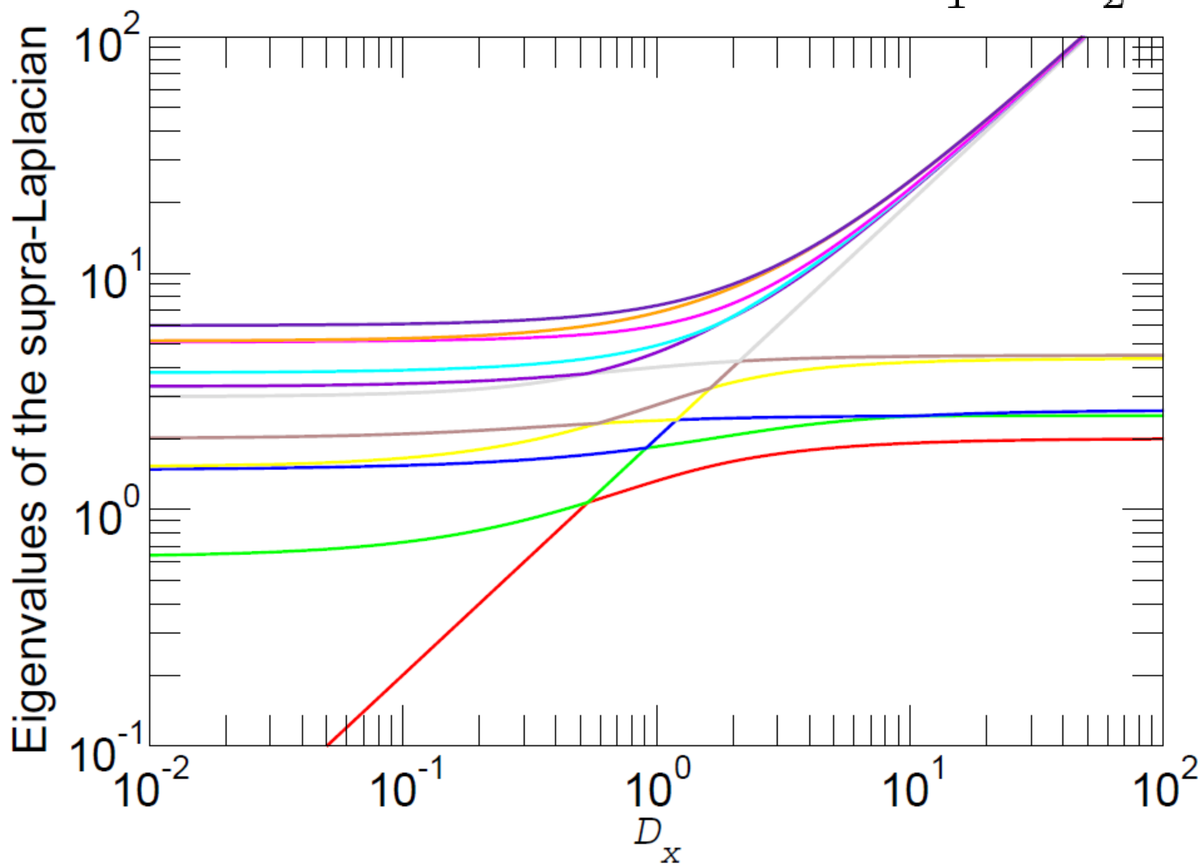
- Eigenvalues

$$0 = \lambda_1 < \lambda_2 \le \lambda_3 \le \ldots \le \lambda_{2n} \qquad D_x \neq 0$$

- Diffusion time scale

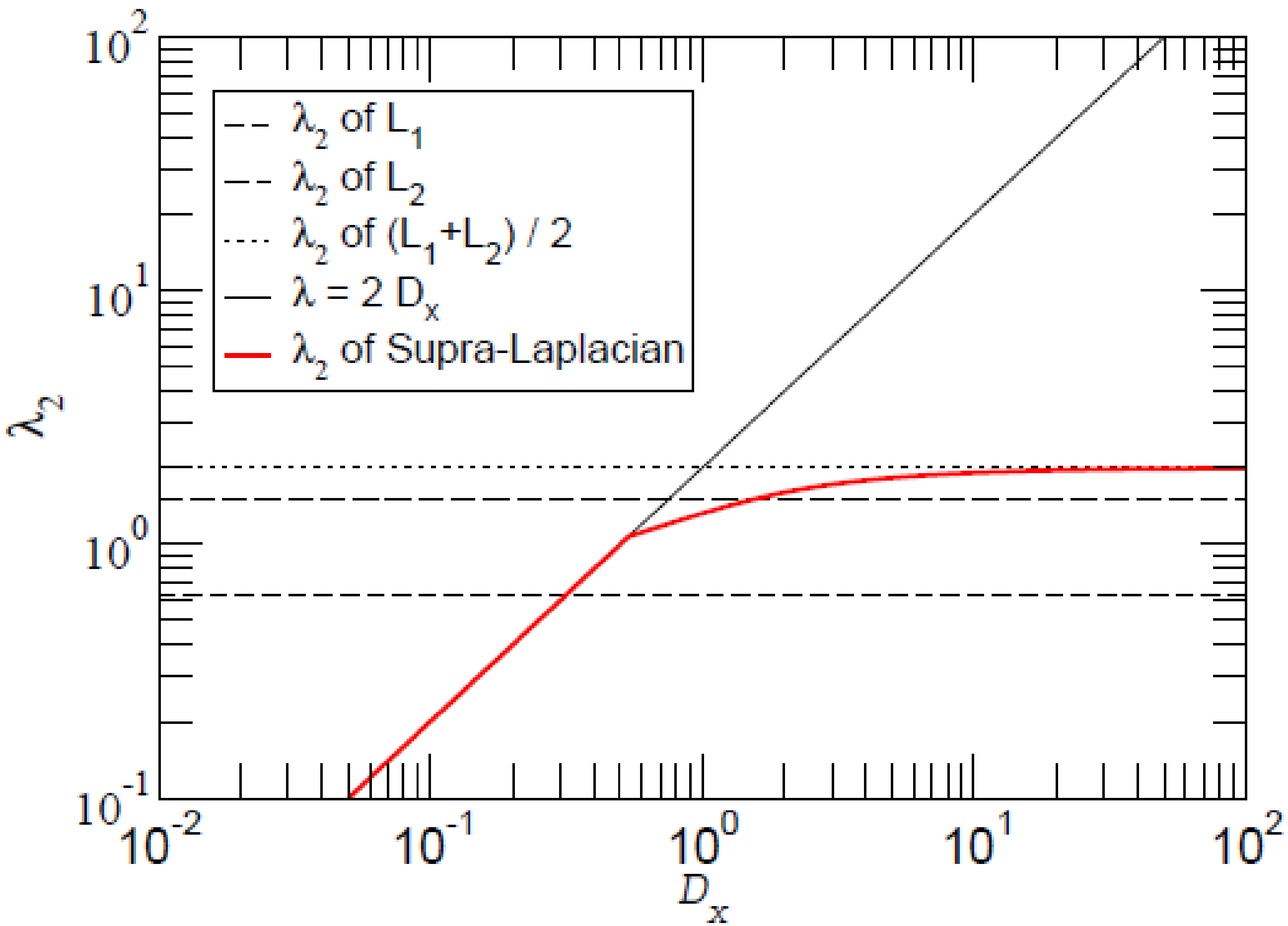
$$\tau = 1/\lambda_2$$

$$0 = \lambda_1^1 < \lambda_2^1 \le \ldots \lambda_n^1$$
$$0 = \lambda_1^2 < \lambda_2^2 \le \ldots \lambda_n^2$$



- Eigenvalues of supra-Laplacian for $D_x << 1$

$$L_1 \rightarrow \lambda_1^1 = 0 \rightarrow (1\cdots1|0\cdots0) \qquad (1\cdots1|1\cdots1) \rightarrow \lambda_1 = 0$$
$$L_2 \rightarrow \lambda_1^2 = 0 \rightarrow (0\cdots0|1\cdots1) \Rightarrow (1\cdots1|-1\cdots-1) \rightarrow \lambda = 2D_x$$



- Eigenvalues of supra-Laplacian for $D_x >> 1$

$$D_x = 1/\epsilon$$

$$\mathcal{L} = D_x \left[\left(\begin{array}{c|c} I & -I \\ \hline -I & I \end{array} \right) + \epsilon \left(\begin{array}{c|c} L_1 & 0 \\ \hline 0 & L_2 \end{array} \right) \right] = D_x \tilde{\mathcal{L}}$$

$$\left(\begin{array}{c|c} I & -I \\ \hline -I & I \end{array} \right) \Rightarrow \begin{array}{l} (\mathbf{u}|\mathbf{u}) \quad \rightarrow \tilde{\lambda} = 0 \\ (\mathbf{u}|-\mathbf{u}) \rightarrow \tilde{\lambda} = 2 \end{array}$$

At first order $O(\epsilon)$

$$\mathcal{L} \rightarrow \mathbf{v} = \left(\frac{\mathbf{u} + \epsilon \mathbf{u}'}{\mathbf{u} - \epsilon \mathbf{u}'} \right) \rightarrow \frac{\lambda_s}{2}$$

$$(L_1 + L_2)\mathbf{u} = \lambda_s \mathbf{u}$$

$$\mathbf{u}' = \frac{1}{4}(L_2 - L_1)\mathbf{u}$$

- Results
  - For $D_x << 1$

$$\tau = \frac{1}{2D_x}$$

  - For $D_x >> 1$

$$\tau \sim \frac{2}{\lambda_s}$$

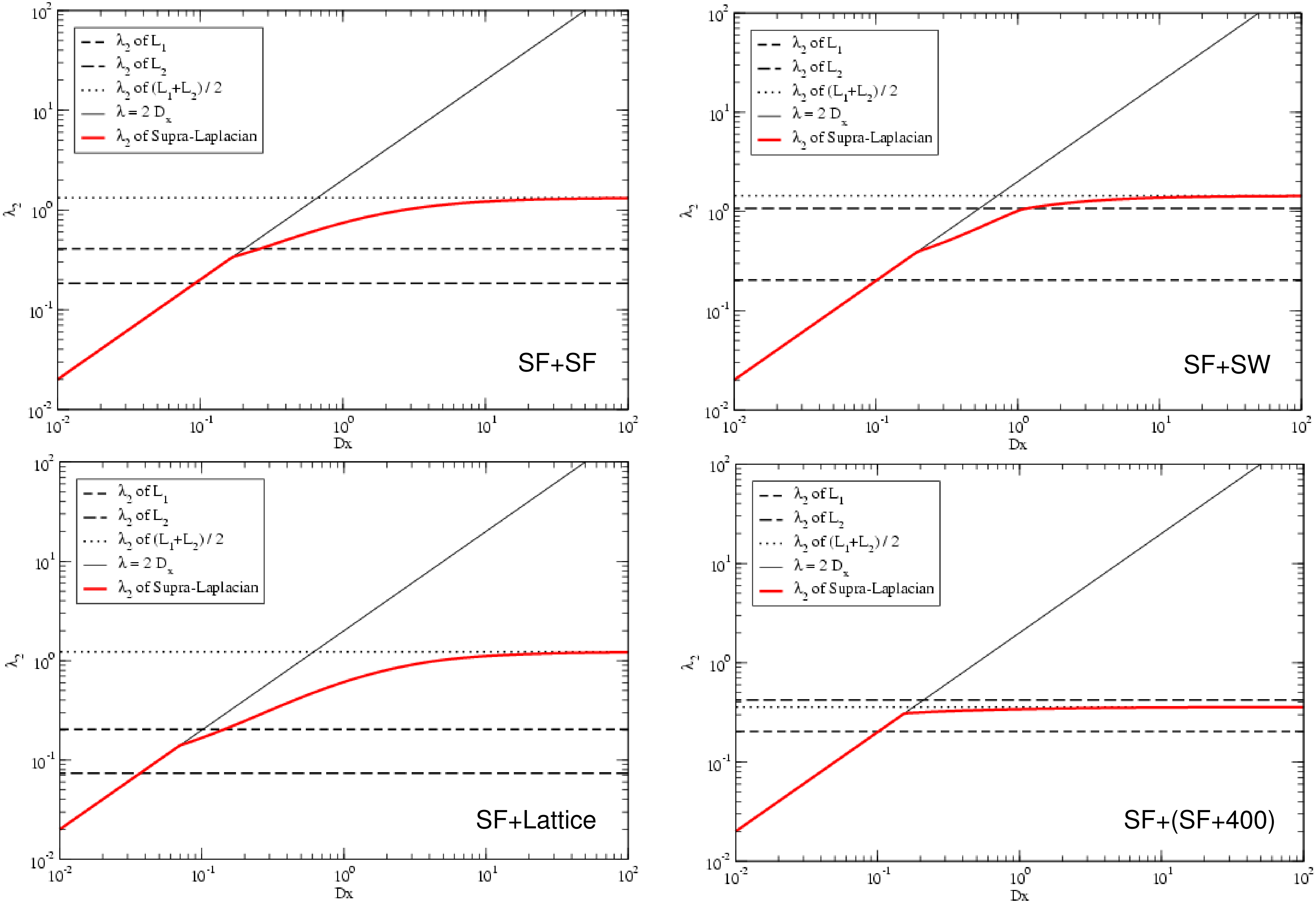
Multiplex faster than the slowest layer

$$\frac{\lambda_s}{2} \geq \frac{\lambda_2^1 + \lambda_2^2}{2} \geq \min\left(\lambda_2^1, \lambda_2^2\right)$$

Superdiffusion if multiplex faster than fastest layer

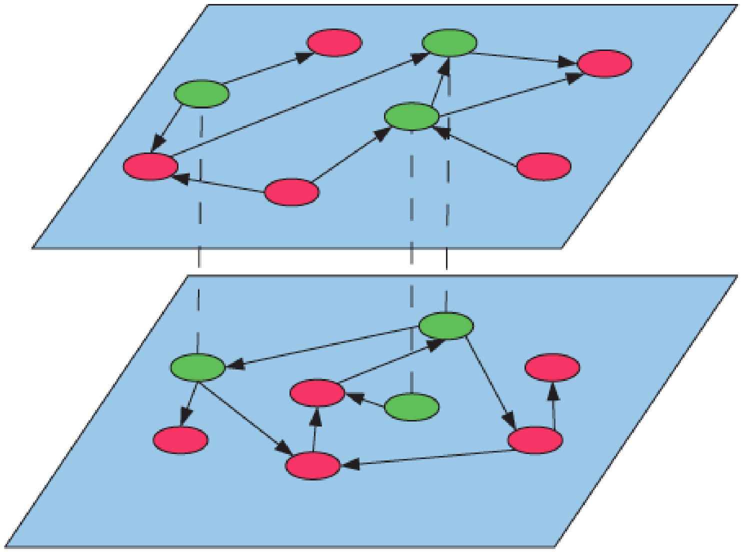
$$\frac{\lambda_s}{2} \leq \frac{2N}{2N-1} \min_i \left(\frac{s_i^1 + s_i^2}{2} + D_x \right)$$

Multiplex networks



- **Social dilemma on multiplex networks**
  - Objective
    - Understand prisoner’s dilemma on multiplex?
      - Evolution of Cooperation in Multiplex Networks, J. Gómez-Gardeñes, I. Reinares, A. Arenas and L. M. Floria, Scientific Reports, 10.1038 srep00620 (2012)
  - Hypothesis
    - Each node plays a PD

	C	D
C	1, 1	0, b
D	b, 0	0, 0

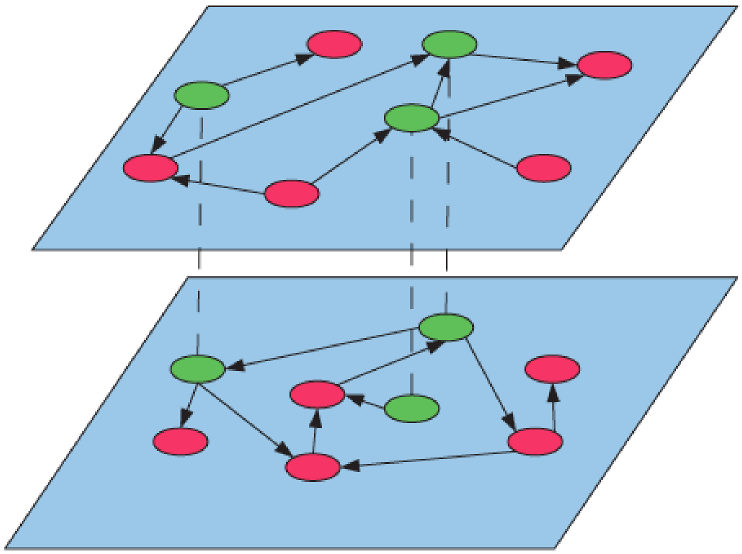


- Payoffs are accumulated through layers, update type Replicator-rule

- How it works:
- Each of the players, say *i*, chooses a layer, say *l*, at random among the *m* possible networks and a neighbor *j* (also randomly) among its k^l_i acquaintances. Then it compares their total payoffs, $P_i(t)$ and $P_j(t)$, obtained in the last round of the game.
- If $P_i(t) > P_j(t)$ nothing happens and *i* will use the same strategy within the network layer *l* in the next round of the PD game, $s_i(t + 1) = s_i(t)$.
- However, when $P_i(t) < P_j(t)$ agent *i* will take the strategy of *j* at layer *l*

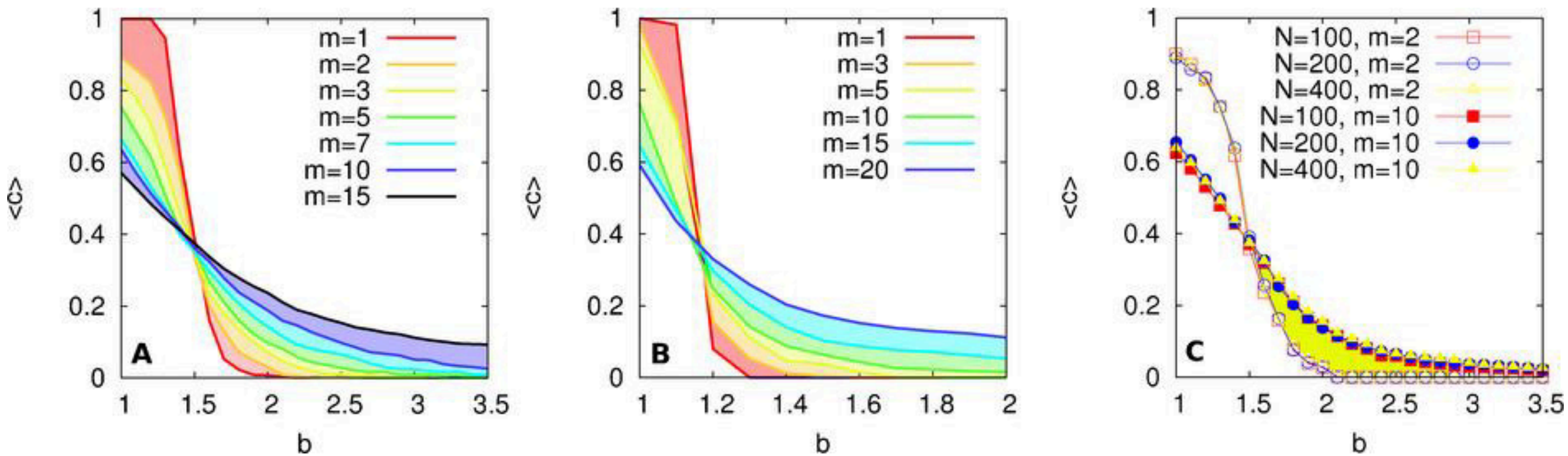
with a probability proportional to their payoff difference:

$$\Pi^l_{i \to j} = \frac{P_j(t) - P_i(t)}{b\, max(\sum_l k^l_i, \sum_l k^l_j)}$$



- Results:

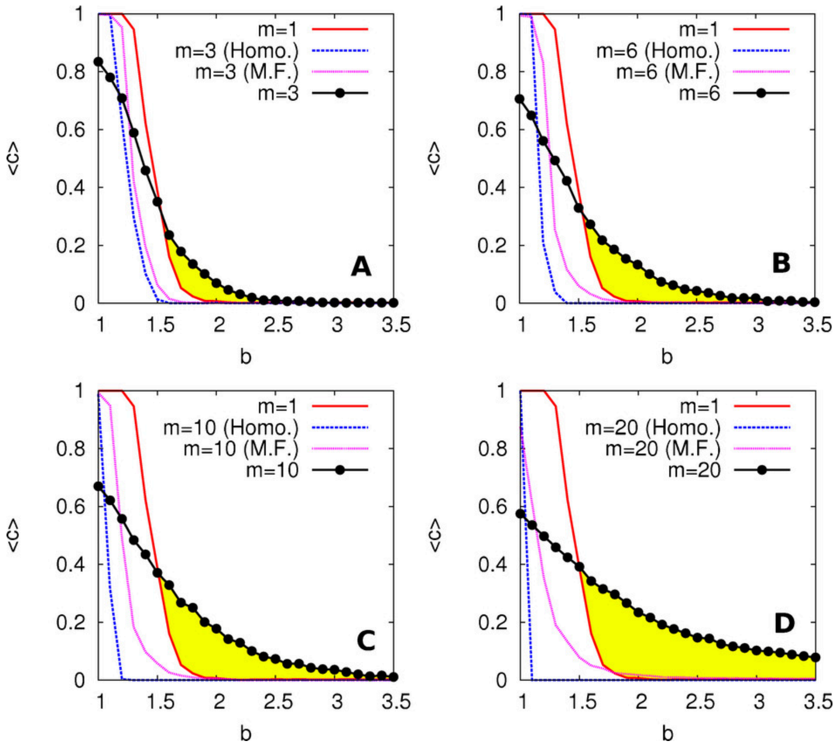
Average level of cooperation <*c*> as a function of the temptation to defect *b* for several multiplex with different number of layers *m*.



- In panel **A** the network layers are ER graphs with <*k*> = 3 (sparse graphs) while in panel **B** we have <*k*> = 20. In both cases *N* = 250 nodes. As can be observed, the resilience of cooperation increases remarkably as the number of layers *m* grows. Finally, panel **C** shows the curves <*c*>(*b*) for ER graphs with *k* = 3 (as in panel **A**) for *m* = 2 and *m* = 10 and different network sizes *N* = 100, 200 and 400.

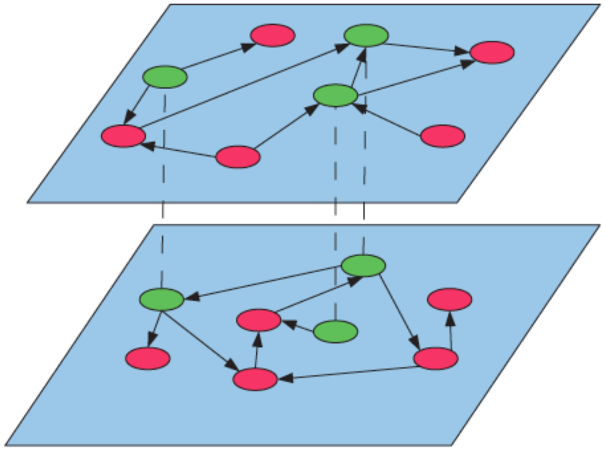
- Results

Average degree of cooperation *c* as a function of *b* (solid line with filled circles) for *m* = 3 (**A**), 6 (**B**), 10 (**C**) and 20 (**D**).



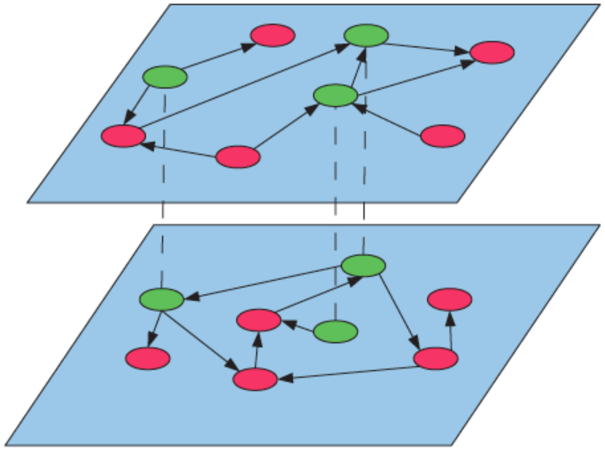
In each panel we show the case of a simplex (*m* = 1) network, the evolution <*c*>(*b*) for homogeneous strategists' populations corresponding to each value of *m* (Homo.) and the curve <*c*>(*b*) corresponding to the mean-field assumption for the coupling between layers (M.F.). Each multiplex network has *N* = 250 nodes while the interdependent layers are ER graphs with <*k*> = 3.

- **Stability of boolean multiplex networks**
  - Objective
    - Understand the interplay between structure and boolean dynamics?
      - Stability of Boolean multilevel networks, E. Cozzo, A. Arenas and Y. Moreno, PRE 86, 036115 (2012)
    - In cellular biochemical networks, many different signaling channels do work in parallel i.e., the same gene or biochemical specie can be involved in a regulatory interaction, in a metabolic reaction or in another signaling pathway.
  - Hypothesis
    - Boolean nodes (on-off)
    - Canalizing rules boolean functions
    - Observation of one layer with hidden connectivity of the degree in other layers



- Stability of boolean multiplex networks
  - Objective
    - Understand the interplay between structure and boolean dynamics?
      - Stability of Boolean multilevel networks, E. Cozzo, A. Arenas and Y. Moreno, PRE 86, 036115 (2012)
    - In cellular biochemical networks, many different signaling channels do work in parallel i.e., the same gene or biochemical specie can be involved in a regulatory interaction, in a metabolic reaction or in another signaling pathway.
  - Hypothesis
    - Boolean nodes (on-off), not all nodes in all layers, Ñ(t) number of different nodes

$$\tilde{\mathbf{x}}(t) = (\tilde{x}_1(t), ..., \tilde{x}_{\tilde{N}(t)})$$

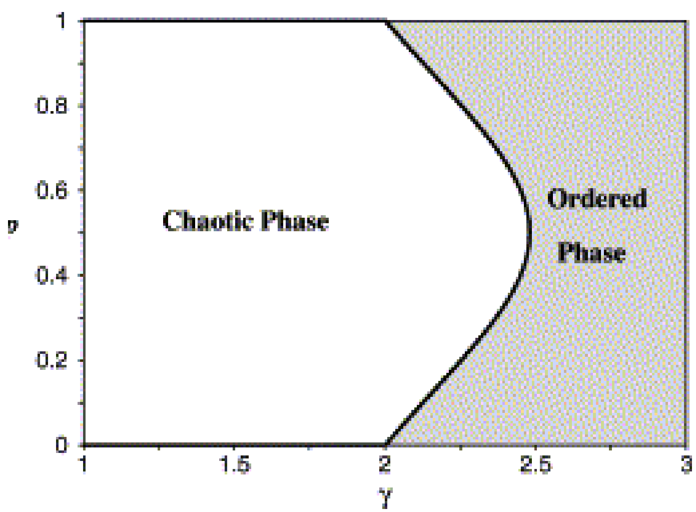


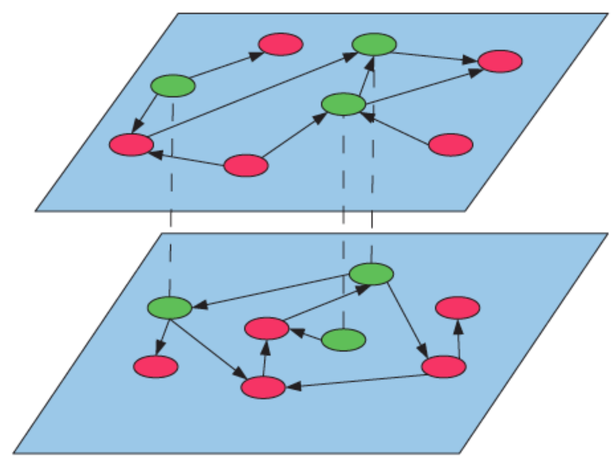
- Hypothesis
  - Update functions

$$\tilde{x}_i(t) = \tilde{f}_i(\tilde{x}_{j\in\Gamma^{\text{in}}_{\alpha}(i)}(t-1))$$

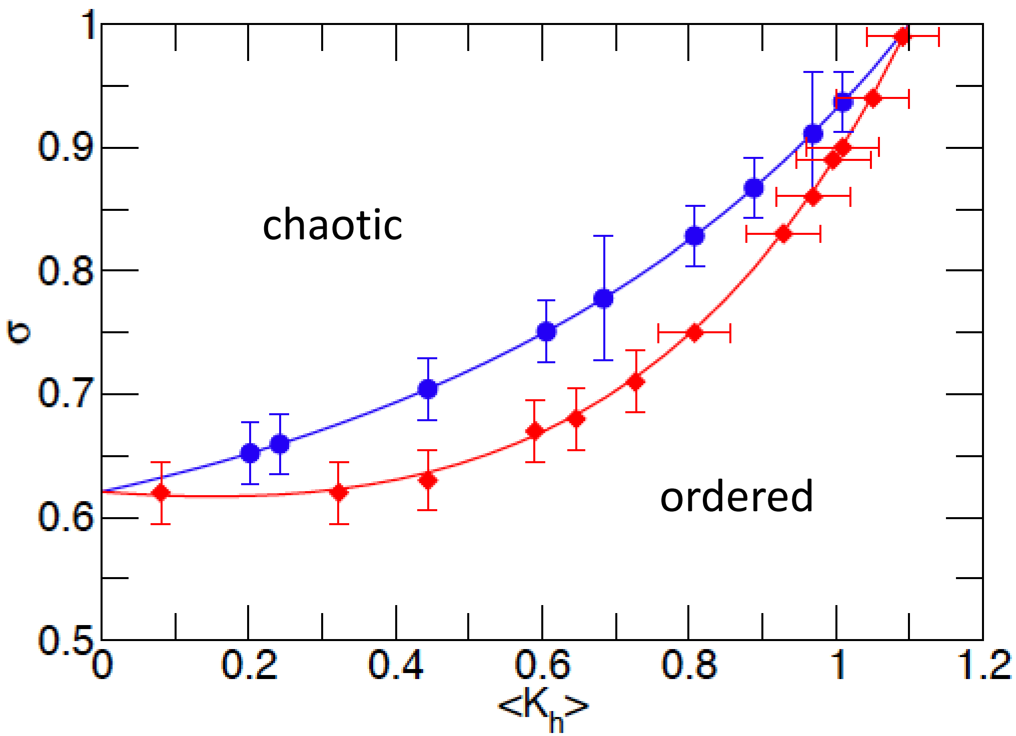
What is the effect of the multiplex connectivity?

Typical phase diagram of boolean networks





- Stabilizing effect of having several layers



Blue (mutiplex), red (simplex.)
Remarkably, the results show that a single ingredient − the multilevel nature of the system could explain why there are biologically stable systems that are however theoretically expected to operate in the unstable regime.

Summary:

- *Diffusion processes* on multiplex networks can show a non-trivial super-diffusion effect

- *The prisoner's dilemma* on top of multiplex networks coupled by the cumulative payoff prove to be more robust to cooperation

- *Boolean networks* in multiplex structures are able to stabilize the whole system while separate layers will show unstable regimes

- Papers and software available at
  - http://deim.urv.cat/~aarenas
  - http://deim.urv.cat/~alephsys