Inferring the connectivity of a complex system from data

Cristina Masoller
Universitat Politecnica de Catalunya, Terrassa, Barcelona, Spain

Cristina.masoller@upc.edu
www.fisica.edu.uy/~cris

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

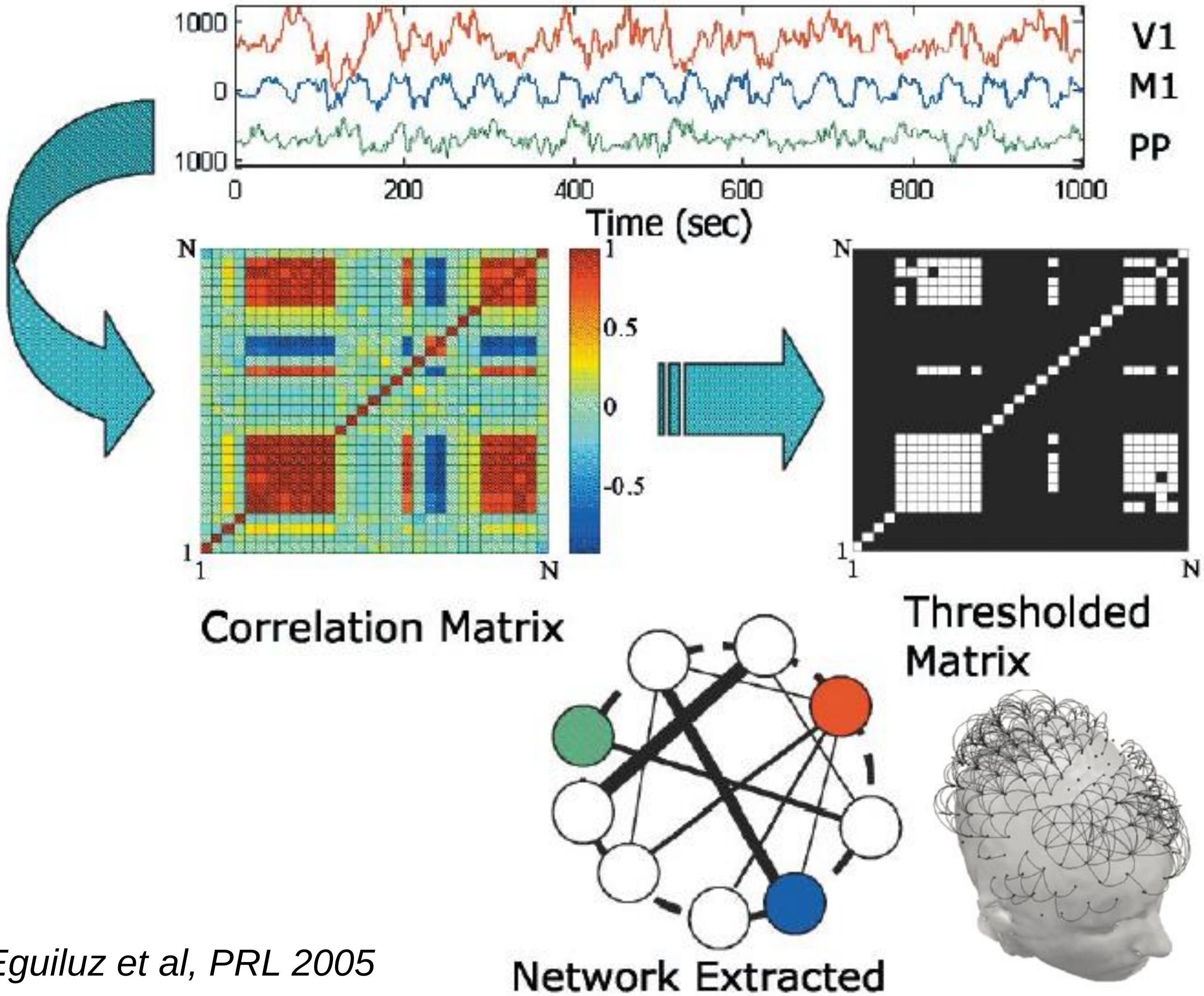
Campus d'Excel·lència Internacional

Workshop on Control of self-organizing nonlinear systems, Wittenberg 22/8/19

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

ICREA

Inferring the connectivity of a complex system from data. First example: brain functional network

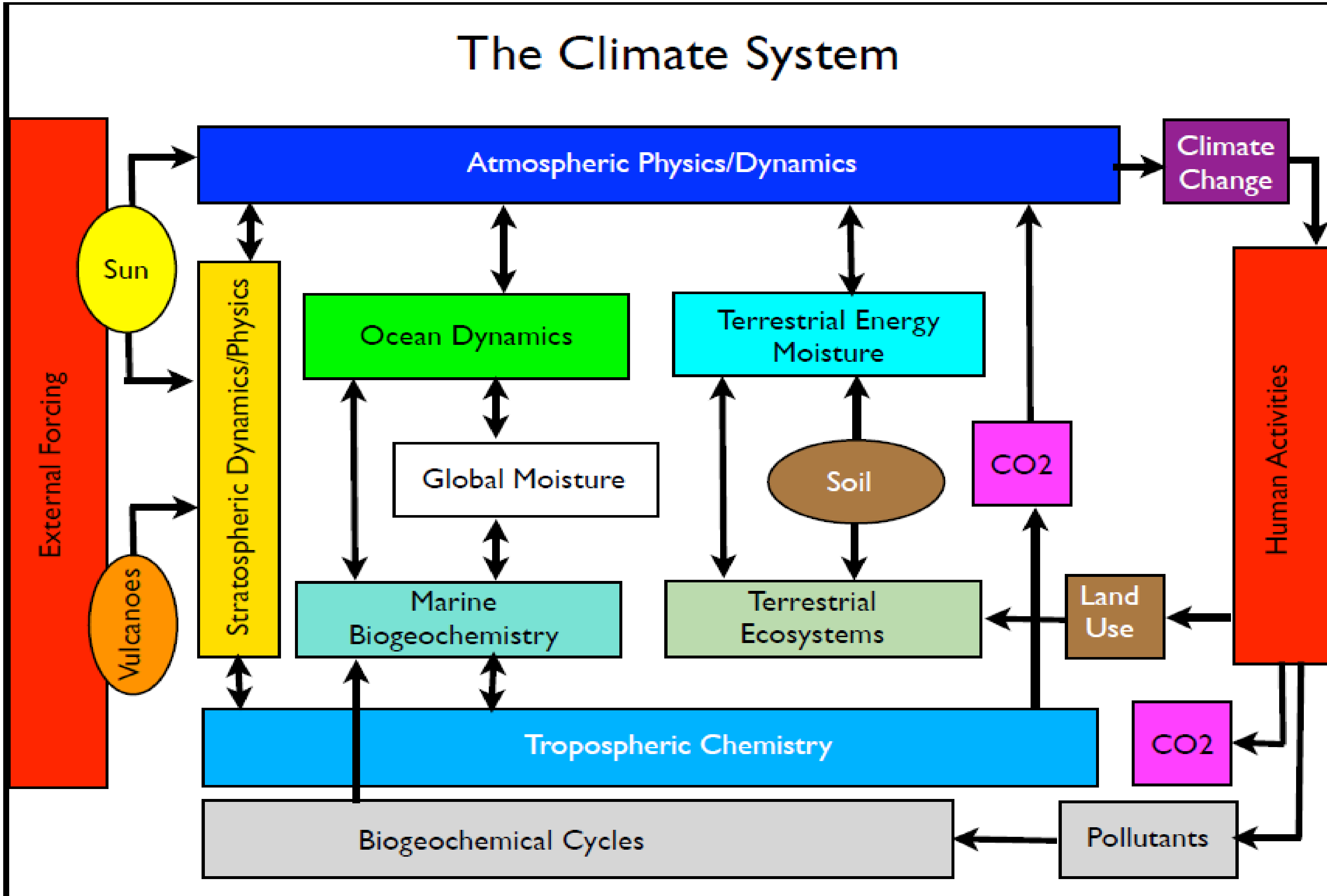


Adjacency matrix

$S_{ij} > Th$
$\Rightarrow A_{ij} = 1,$
else A_{ij}=0

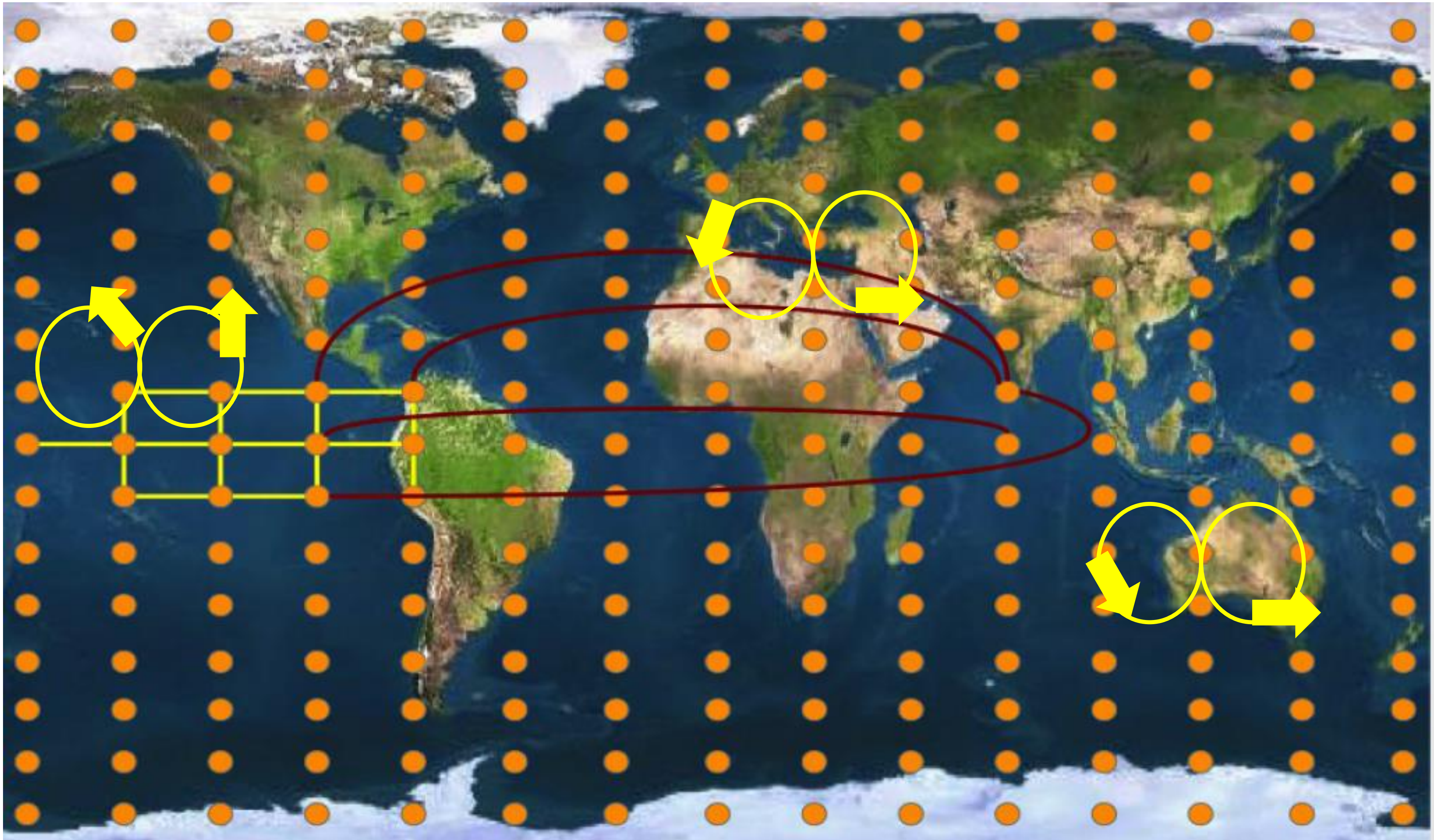
Eguiluz et al, PRL 2005

Second example: climate networks

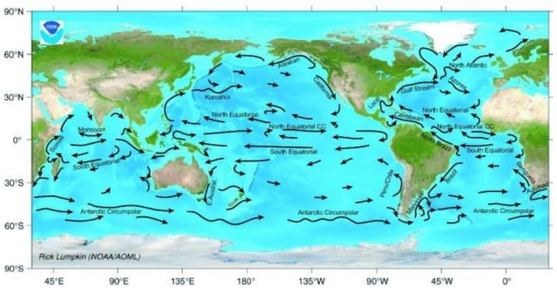


Henk Dijkstra (Universidad de Ultrech)

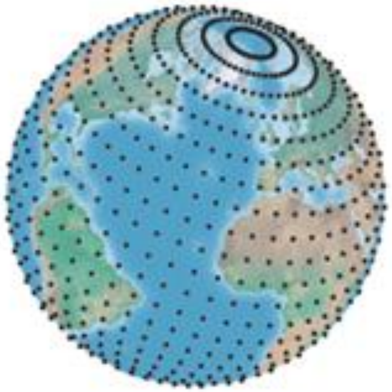
The climate system as a complex network



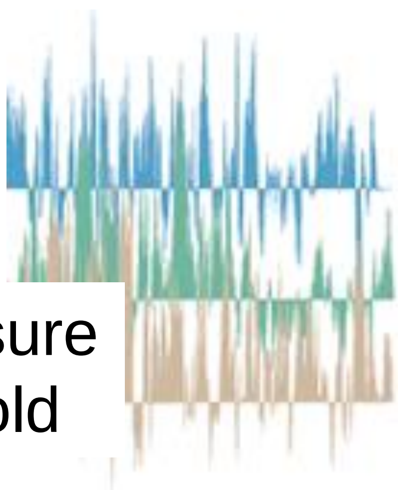
Complex network representation of the climate system



Back to the climate system: interpretation (currents, winds, etc.)



Nodes

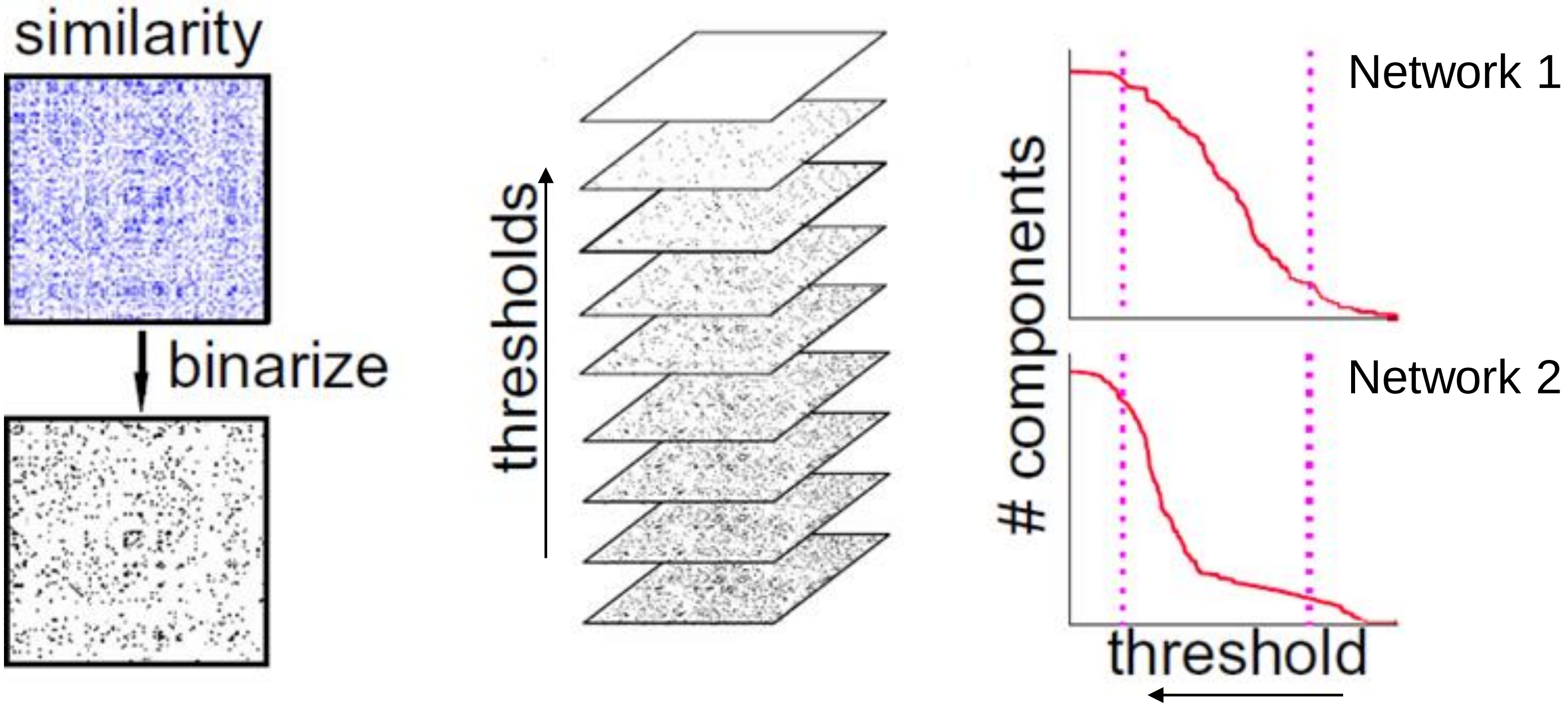


Time series in each node (e.g. air temperature)

Sim. measure + threshold

Donges et al, Chaos 2015

Problems with thresholding



- The number of connected components as a function of threshold reveals different structures.
- But thresholding near the dotted lines indicates (inaccurately) that networks 1 and 2 have similar structures.

Giusti et al., J Comput Neurosci (2016) 41:1–14

How to “infer” significant interactions from observed data?

How to “reconstruct” the network?

A classification problem

$$\boxed{S_{ij} > \mathrm{Th} \Rightarrow A_{ij} = 1, \text{ else } A_{ij}=0}$$

- How to select the threshold?
- In “spatially embedded networks”, nearby nodes have the strongest links.
- How to keep **weak-but-significant** links?

- There are many **statistical similarity measures** to infer interactions from observations, i.e., to classify:
  - the interaction exists (is significant)
  - the interaction does not exists (or is not significant)

Goal: use a system with known connectivity to test the performance of statistical similarity measures

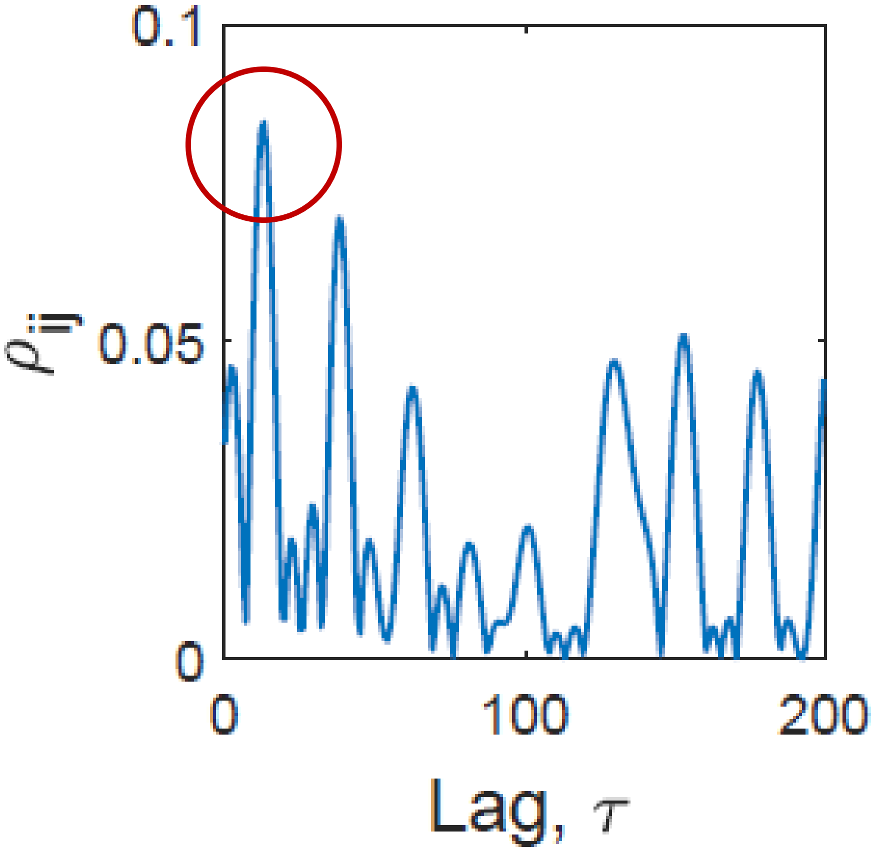
Observed time series in nodes i and j: $a_i(t)$, $a_j(t)$, $t=1, \ldots, T$ *(normalized* $\mu=0$, $\sigma=1$*)*

Lagged |cross correlation|: $$\rho_{ij}(\tau) = \left| \left\langle a_i(t) a_j(t-\tau) \right\rangle_t \right|$$

$$\rho_{ij}(\tau) = \frac{1}{T - \tau_{\max}} \left| \sum_{t=\tau_{\max}}^{T} \left\langle a_i(t) a_j(t-\tau) \right\rangle \right|$$

Statistical Similarity Measure:

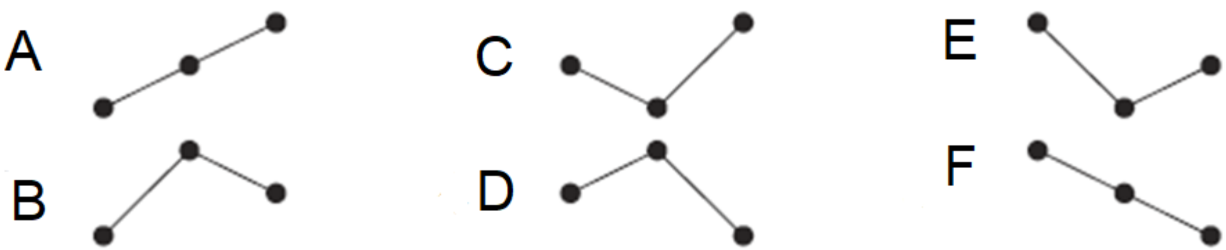
$$S_{ij} = \max |\rho_{ij}(\tau)|$$
$$= |\rho_{ij}(\tau_{ij})| \qquad \tau_{ij} \text{ in } [0, \tau_{\max}]$$

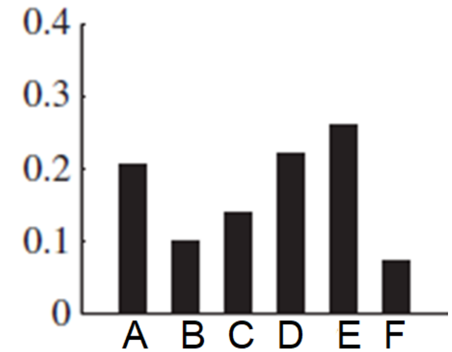


We compare with the Mutual Information, computed from probabilities of “raw” values and from ordinal probabilities

Ordinal analysis: a method to find patterns in data

- Consider a time series $x(t)=\{\ldots x_i,\ x_{i+1},\ x_{i+2},\ \ldots\}$
- Which are the possible order relations among three data points?





- Calculate ordinal probabilities by counting how many times each “ordinal pattern” appears.
- Advantages: allows to identify temporal structures & is robust to noise.
- Drawback: information about actual data values is lost.

Bandt and Pompe PRL 88, 174102 (2002)

The Mutual Information: a nonlinear correlation measure

$$MI = \sum_{i \in x} \sum_{j \in y} p(x,y) \, \log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

- *MI (x,y) = MI (y,x)*

- *p(x,y) = p(x) p(y)* $\Rightarrow$ *MI = 0*, else ***MI >0***

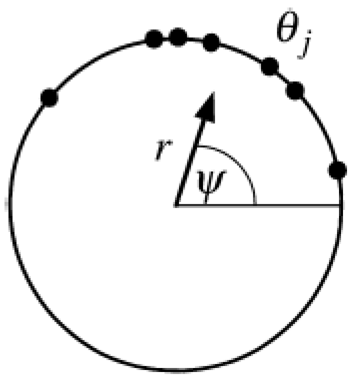
- *MI value significant?* $\Rightarrow$ *Analysis of surrogate data*

Dynamical system:
Kuramoto phase oscillators

Kuramoto model

(Japanese physicist, 1975)

Model of **all-to-all** coupled **phase oscillators**.



$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N}\sum_{j=1}^{N}\sin(\theta_j - \theta_i) + \xi_i, \qquad i = 1 \ldots N$$

K = coupling strength, ξ_i = stochastic term (noise)

Describes the emergence of collective behavior

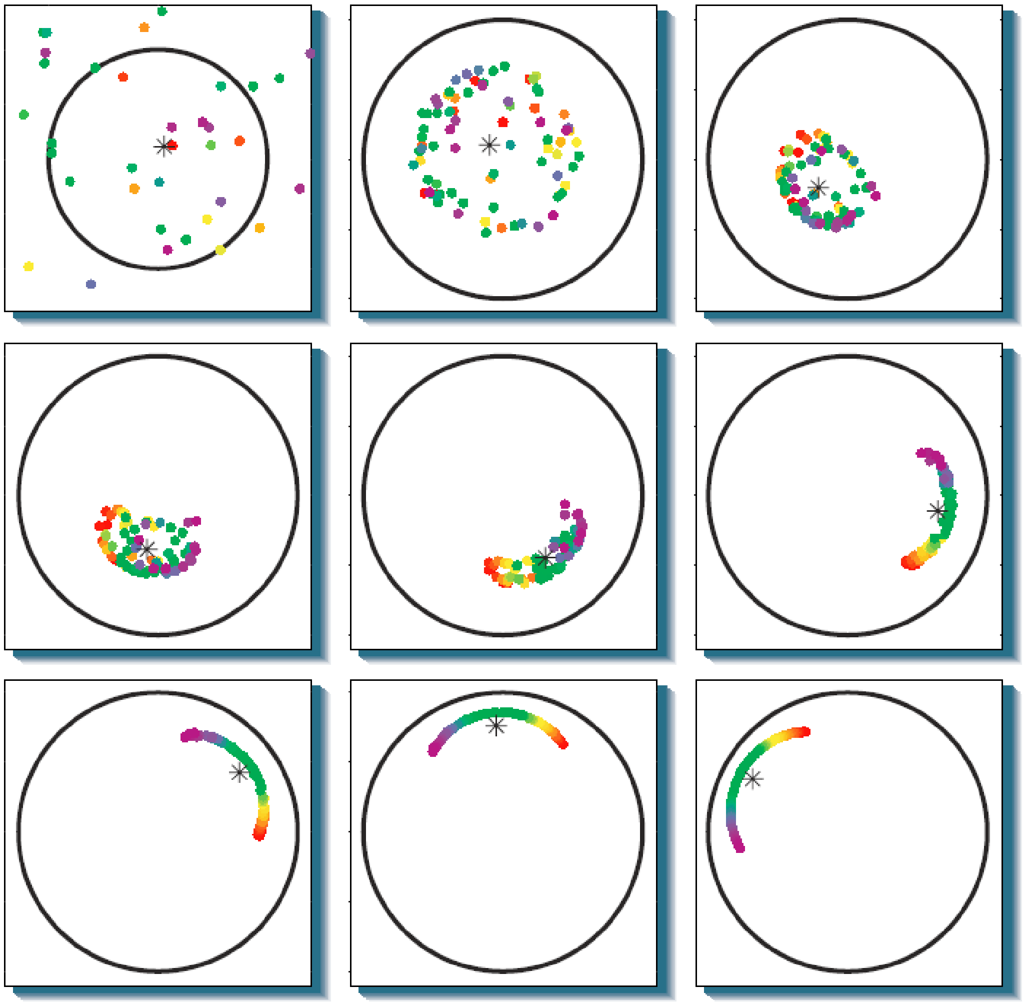
How to quantify?

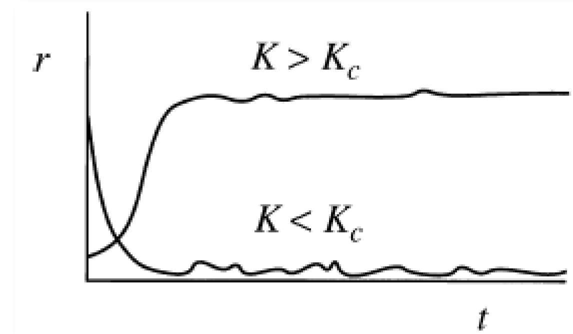
With the **order parameter**: $$re^{i\psi} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j}$$

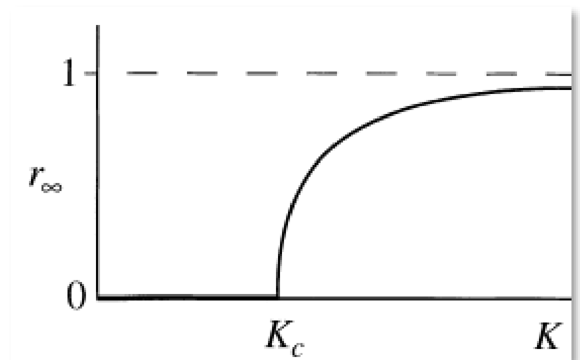
r =0 incoherent state (oscillators scattered in the unit circle)

r =1 all oscillators are in phase ($\theta_i=\theta_j \ \forall$ i,j)

Synchronization transition as the coupling strength increases







Inferring the links is not possible if the oscillators are synchronized

Strogatz, Nature 2001

Kuramoto oscillators in a random network

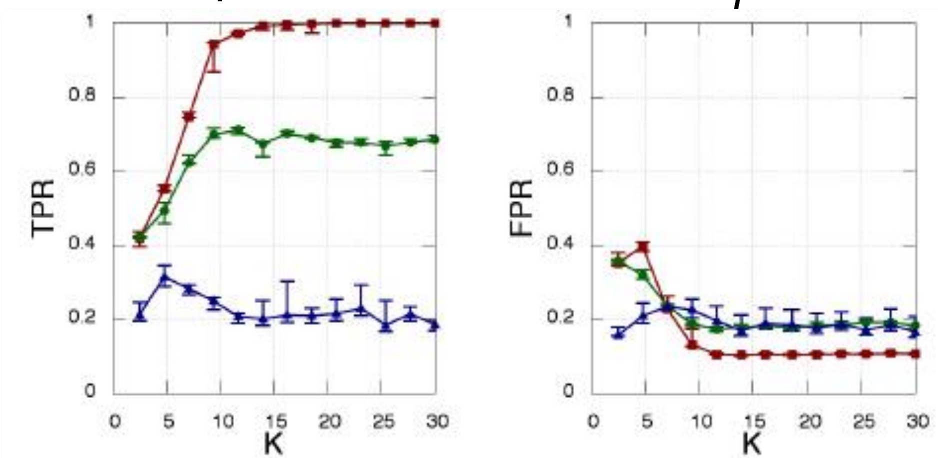
$$d\theta_i = \omega_i dt + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_j - \theta_i)dt + D\, dW_t^i$$

A_{ij} is a symmetric random matrix; N=12 time-series, each with 10^4 data points.

CC MI MIOP

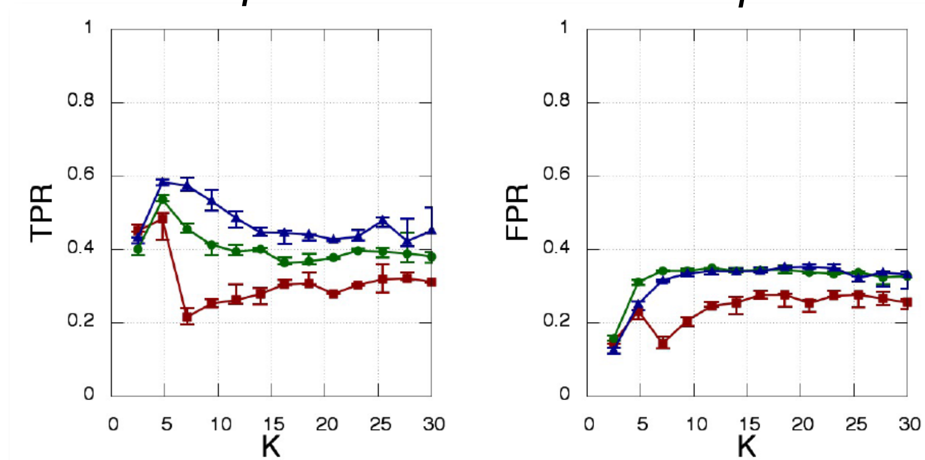
Phases (θ)

True positives *False positives*



"Observable" Y=sin(θ)

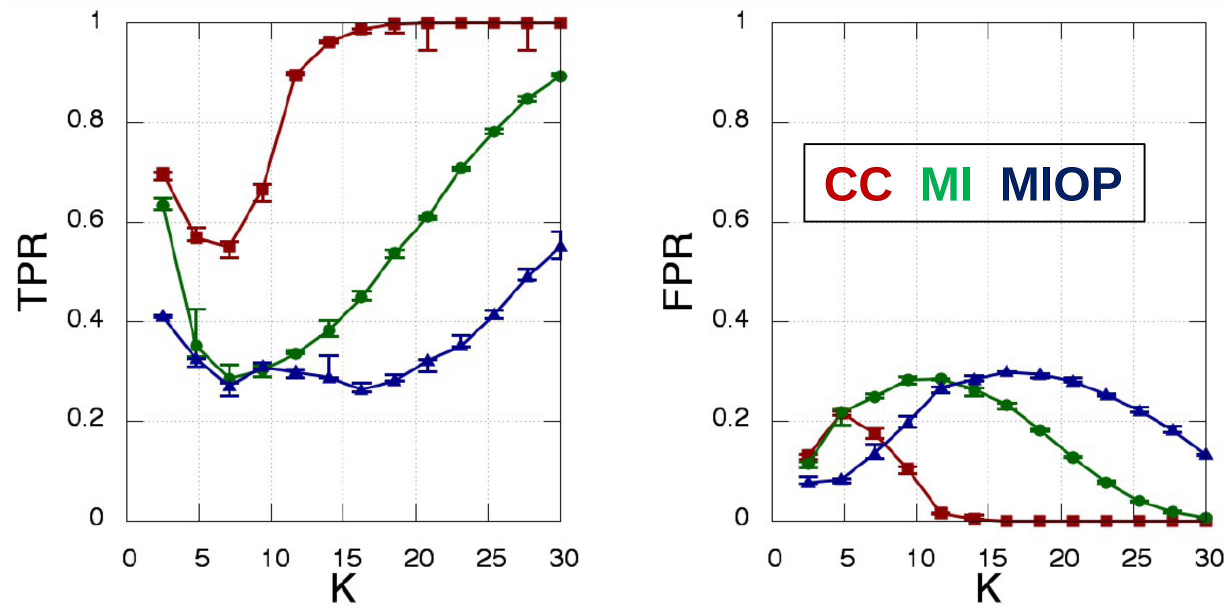
True positives *False positives*



Results of a 100 simulations with different oscillators' frequencies, random matrices, noise realizations and initial conditions.
For each K, the threshold was varied to obtain optimal reconstruction.

G. Tirabassi et al., "*Inferring the connectivity of coupled oscillators from time-series statistical similarity analysis*", Sci. Rep. **5** 10829 (2015).

Instantaneous frequencies (dθ/dt)



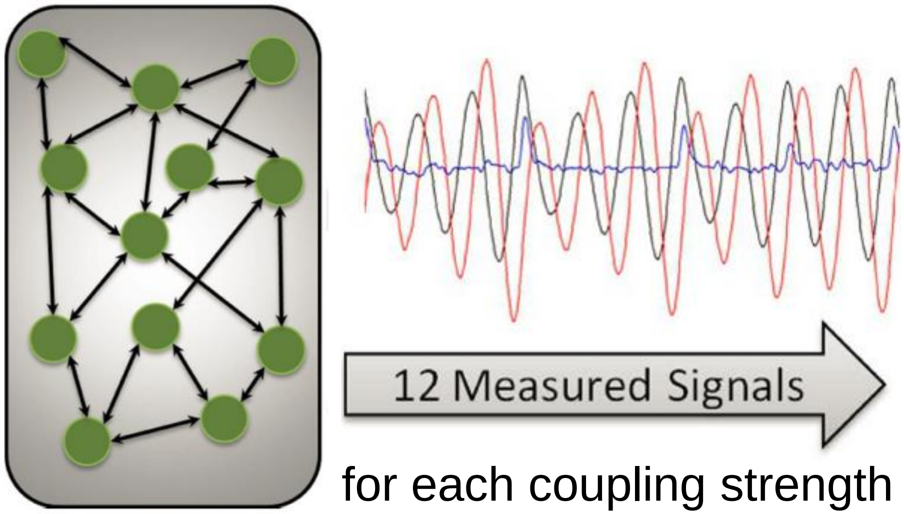
Perfect network inference is possible!

BUT

- the number of oscillators is small (12),
- the coupling is symmetric ($\Rightarrow$ only 66 possible links) and
- the data sets are long (10^4 points)

*G. Tirabassi et al, Sci. Rep. **5** 10829 (2015)*

We also analyzed experimental data recorded from 12 chaotic Rössler electronic oscillators (symmetric and random coupling)



for each coupling strength

The Hilbert Transform was used to obtain phases from experimental data

- Kuramoto Oscillators' Network

$$\theta_i$$

$$f_i = \dot{\theta}_i$$

$$Y_i = \sin(\theta_i)$$

- Rössler Oscillators' Network

$$\varphi_i = HT(x_i)$$

$$f_i = \dot{\varphi}_i$$

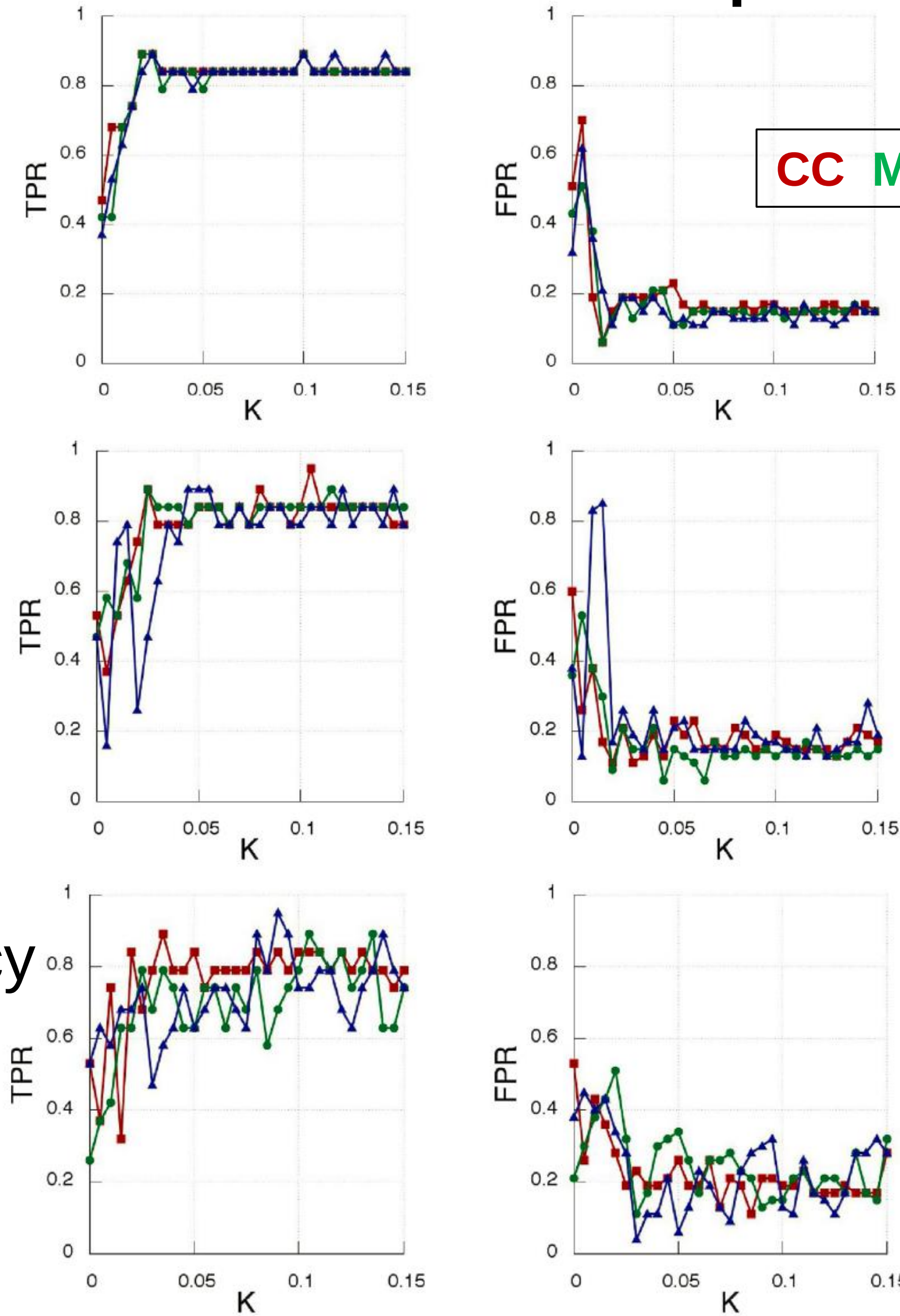
$$x_i$$

Results obtained with experimental data

Observed variable (x)

Hilbert phase

Hilbert frequency

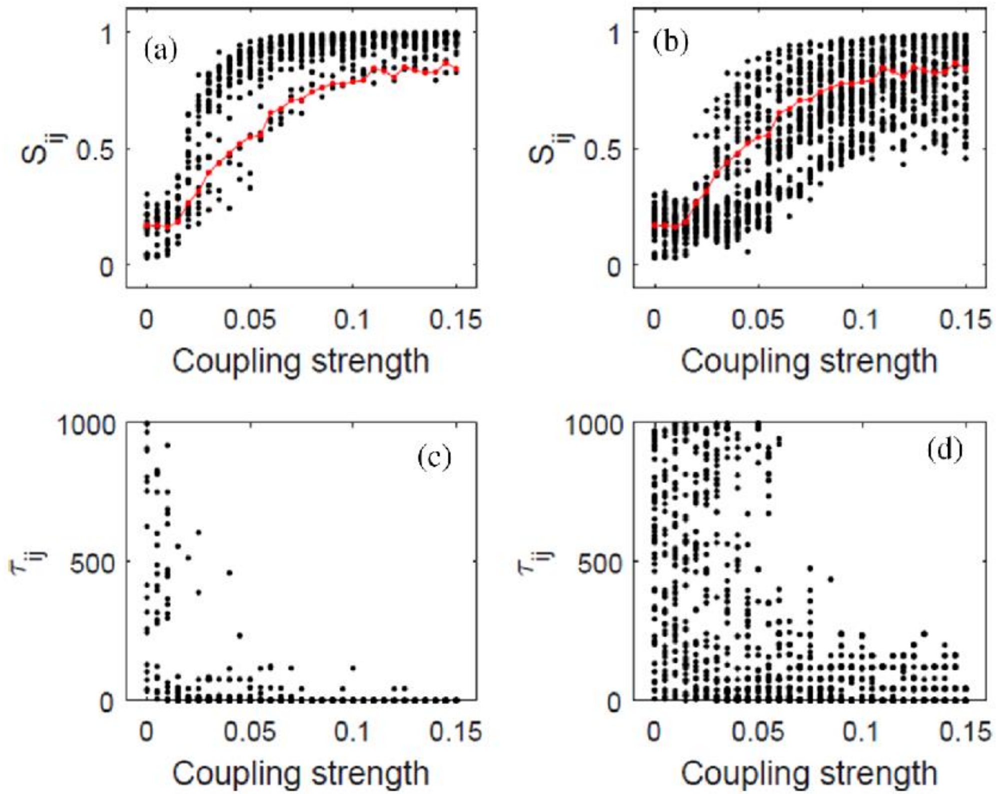


– No perfect reconstruction

– No important difference among the 3 methods & 3 variables

How the similarity values and lag times depend on the coupling strength?

12 electronic chaotic circuits



N. Rubido and C. Masoller, "*Impact of lag information on network inference*", Eur. Phys. J. Special Topics **227**, 1243-1250 (2018).

Can we use lag-time information to infer the links?

If $\mathbf{S_{ij} > TH}$ the link i $\longleftrightarrow$ j exists, otherwise, it does not exist

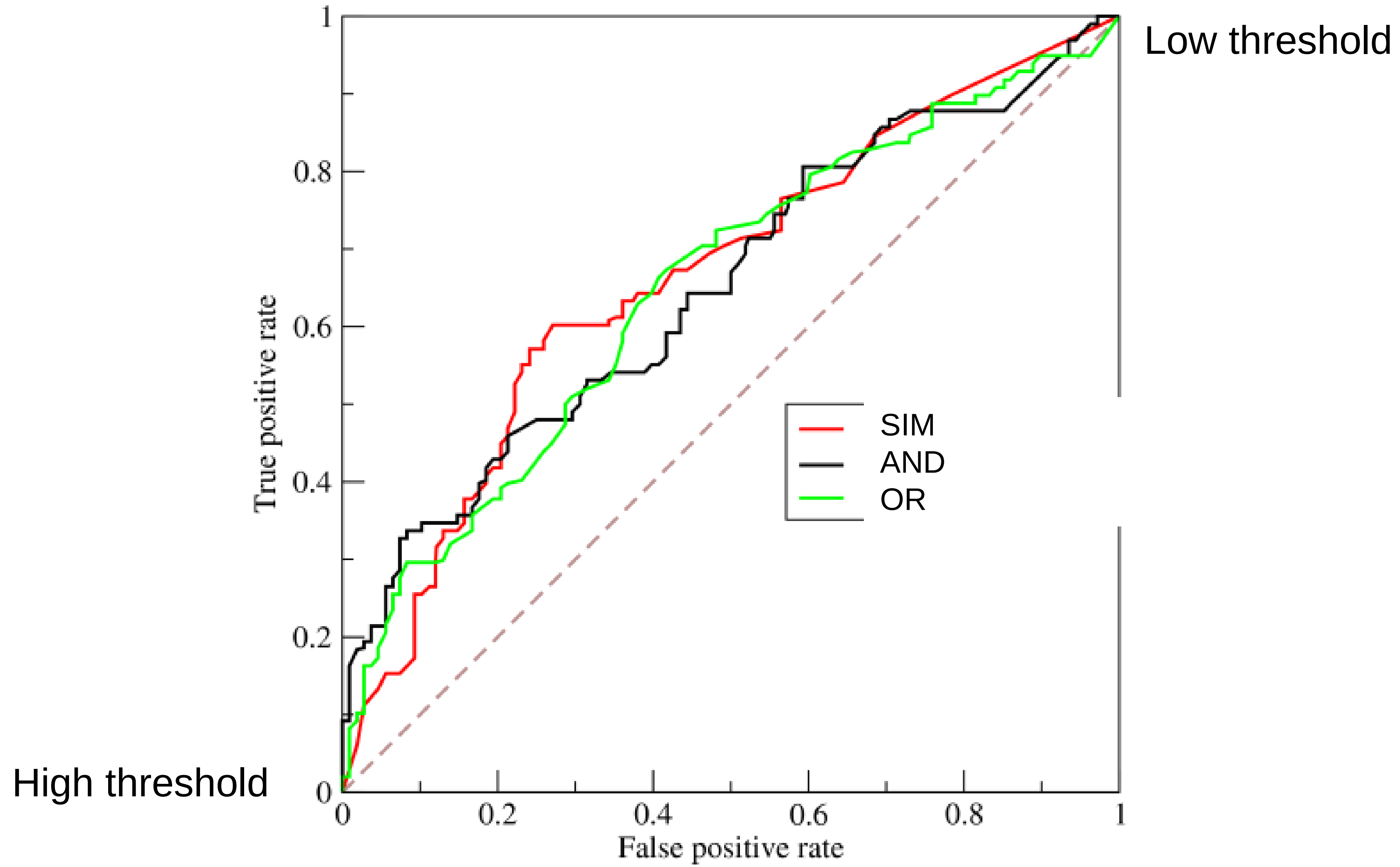
If $\tau_{ij} < \tau_{TH}$ the link i $\longleftrightarrow$ j exists, otherwise, it does not exist

Three possible rules:

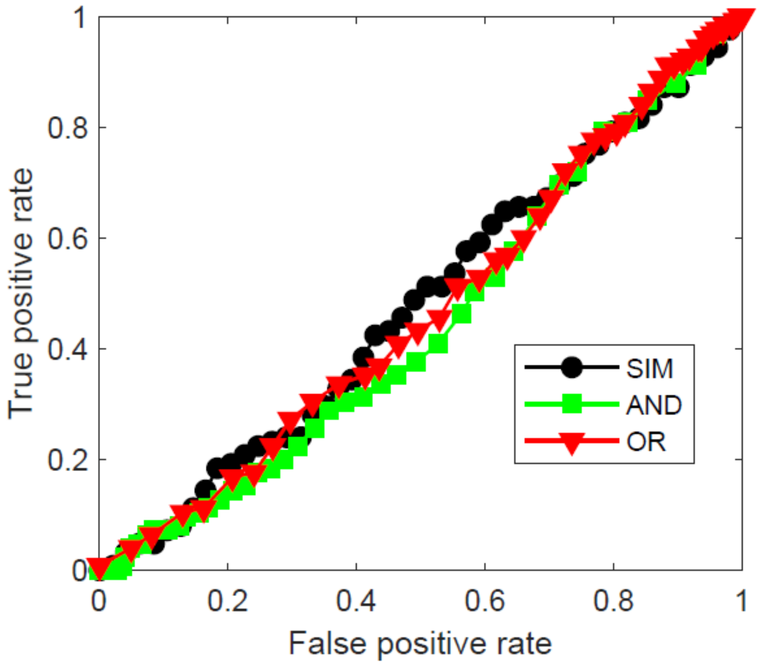
The link i $\longleftrightarrow$ j exists if

- SIM : only the first criterion holds ($\mathbf{S_{ij} > TH}$)
- AND: both criteria hold ($\mathbf{S_{ij} > TH}$ and $\tau_{ij} < \tau_{TH}$)
- OR: at least one criteria holds ($\mathbf{S_{ij} > TH}$ or $\tau_{ij} < \tau_{TH}$)

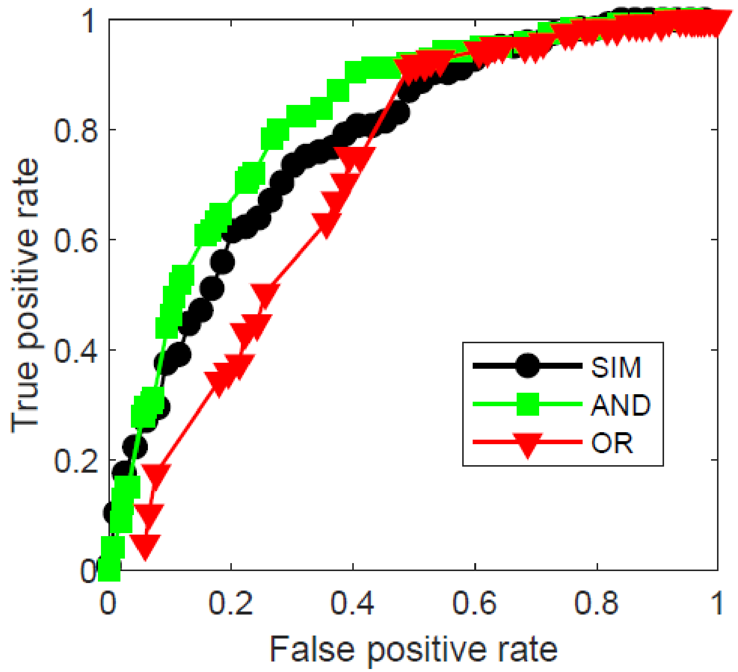
To quantify how good these rules are we use the area under the receiver operating characteristic (ROC) curve



Uncoupled oscillators



Coupled oscillators



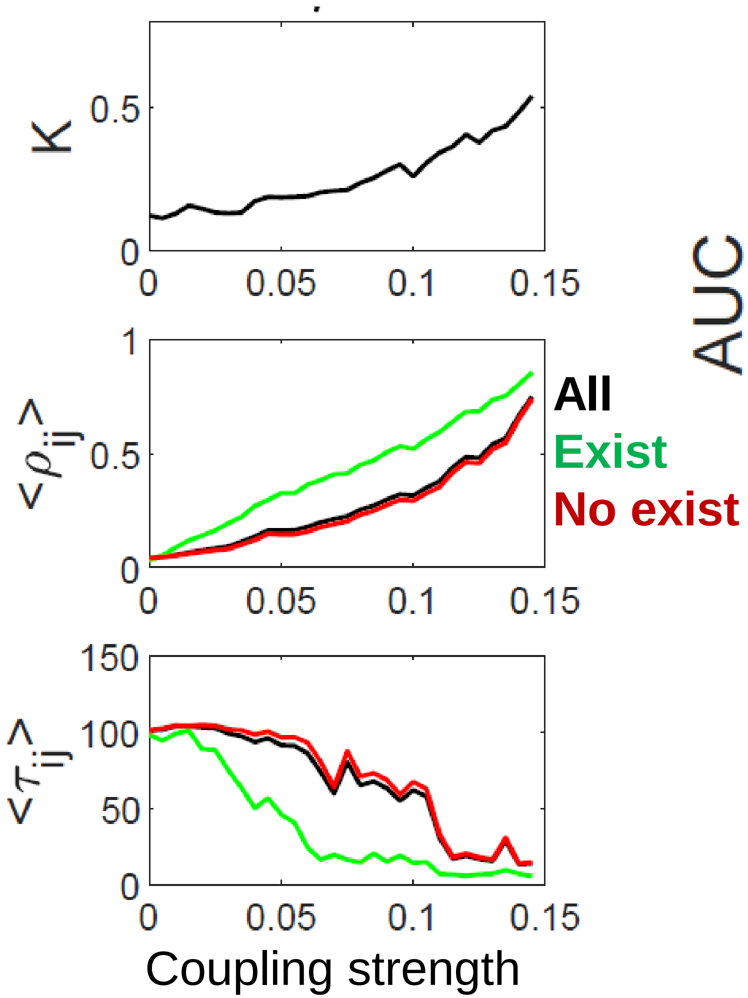
Results

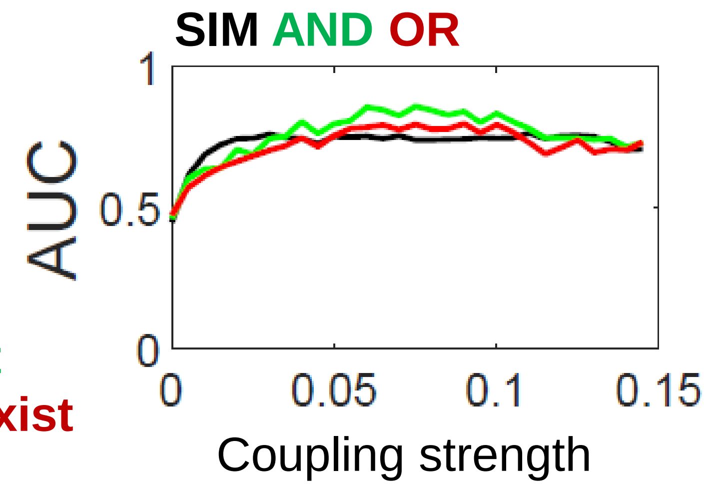
50 Kuramoto phase oscillators, 10% existing links,
Similarity $\rho_{ij} = \max_{\tau}$ cross-correlation of $\cos(\phi_i)$, $\cos(\phi_j)$

Order parameter

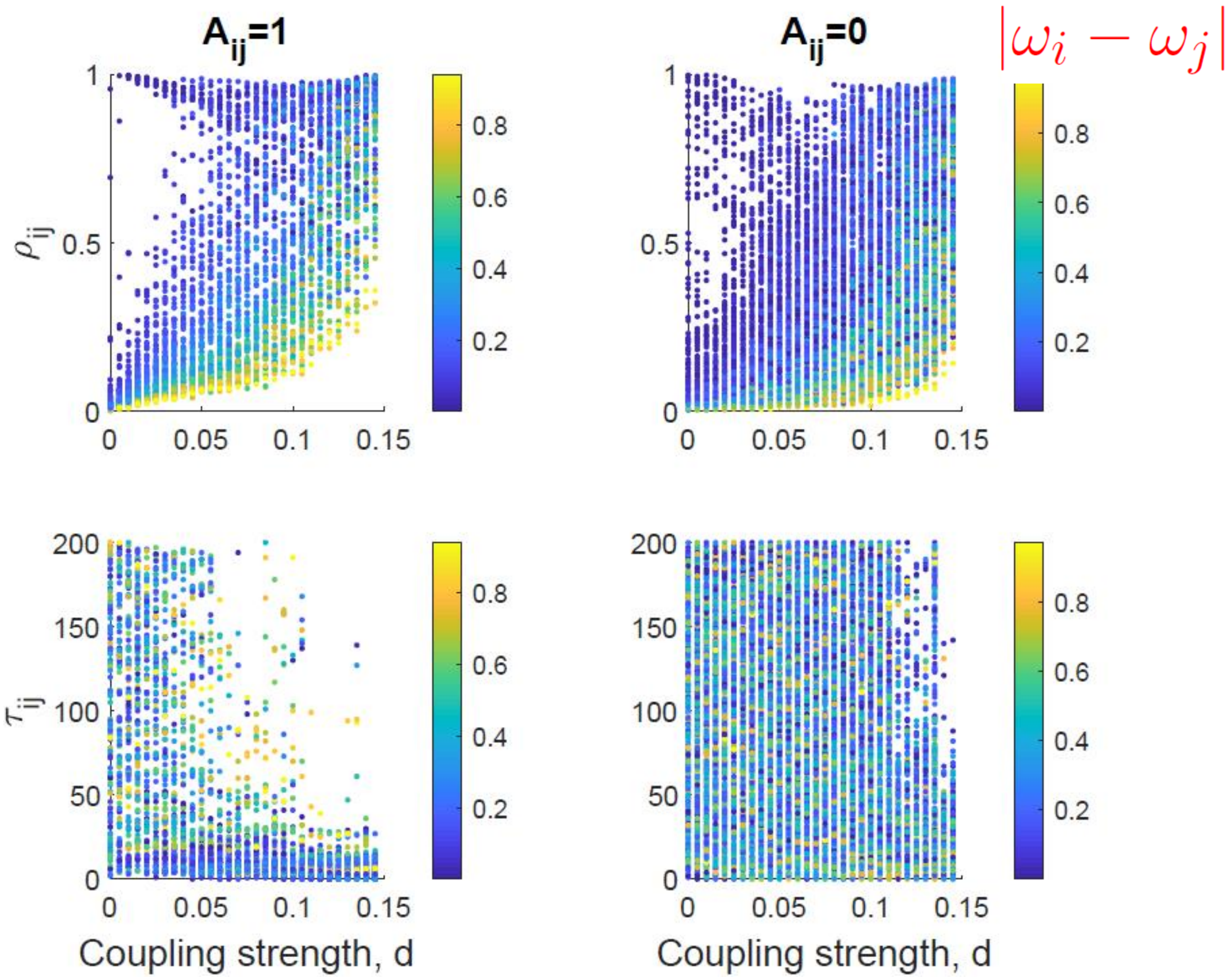
$$K = \left\langle \frac{1}{N} \left| \sum_{i=1}^{N} e^{i\phi_i(t)} \right| \right\rangle_T$$

$$\left\langle \cos[\phi_i(t)] \cos[\phi_j(t-\tau)] \right\rangle$$



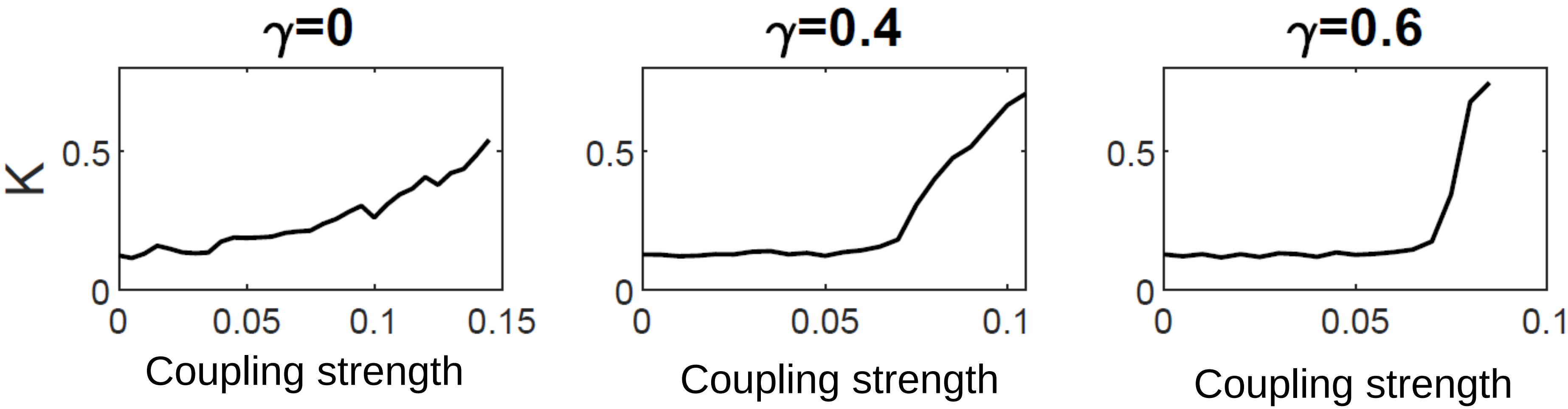


Variation of similarity and τ_{ij} values with the coupling



Explosive transition to synchrony

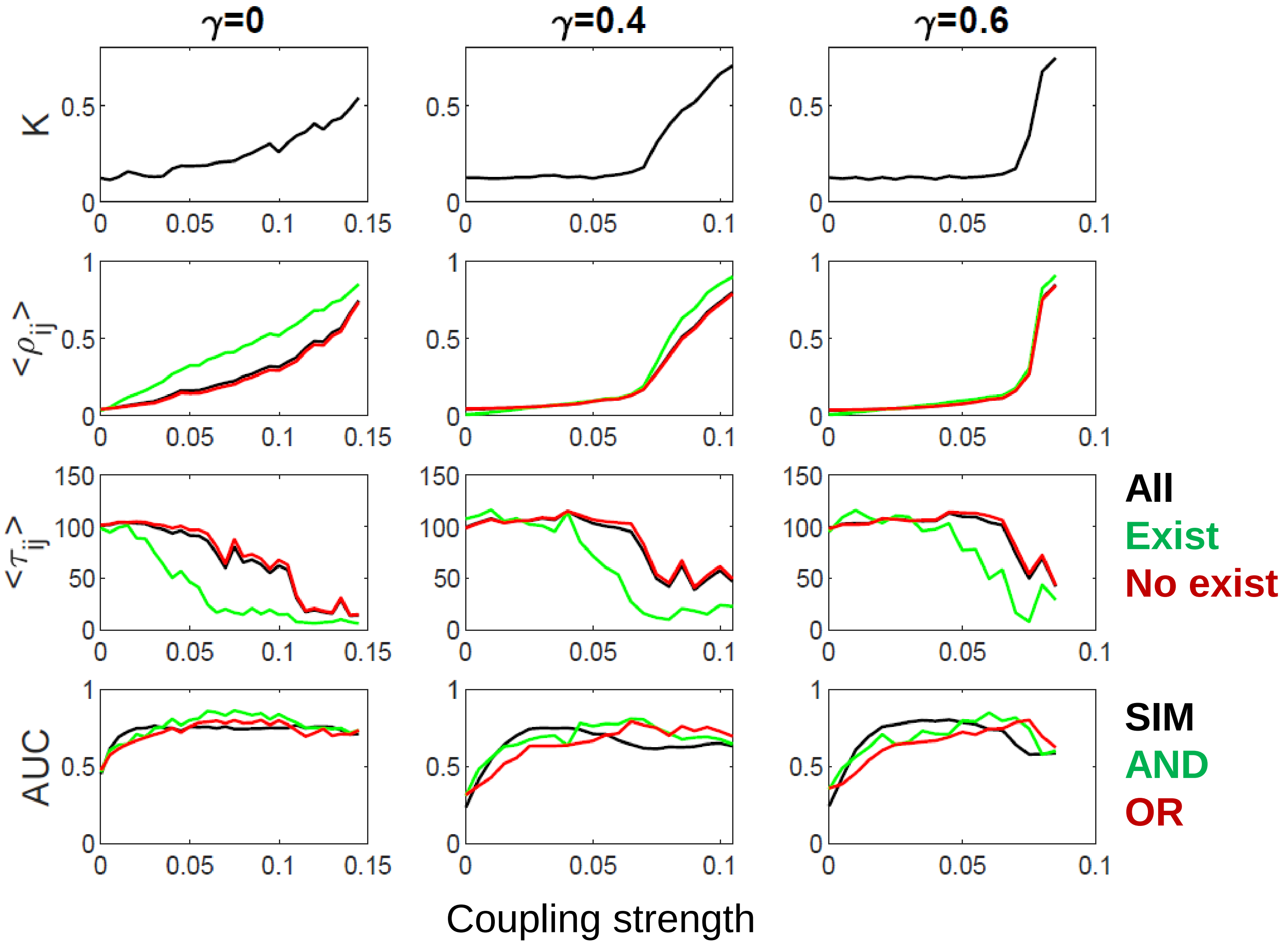
Oscillators can be linked only if they have different frequencies: $|\omega_i - \omega_j| > \gamma$



Order parameter $K = \left\langle \frac{1}{N} \left| \sum_{i=1}^{N} e^{i\phi_i(t)} \right| \right\rangle_T$

I. Leyva et al. Explosive transitions to synchronization in networked phase oscillators. Scientific Reports 3 (2013) 1281

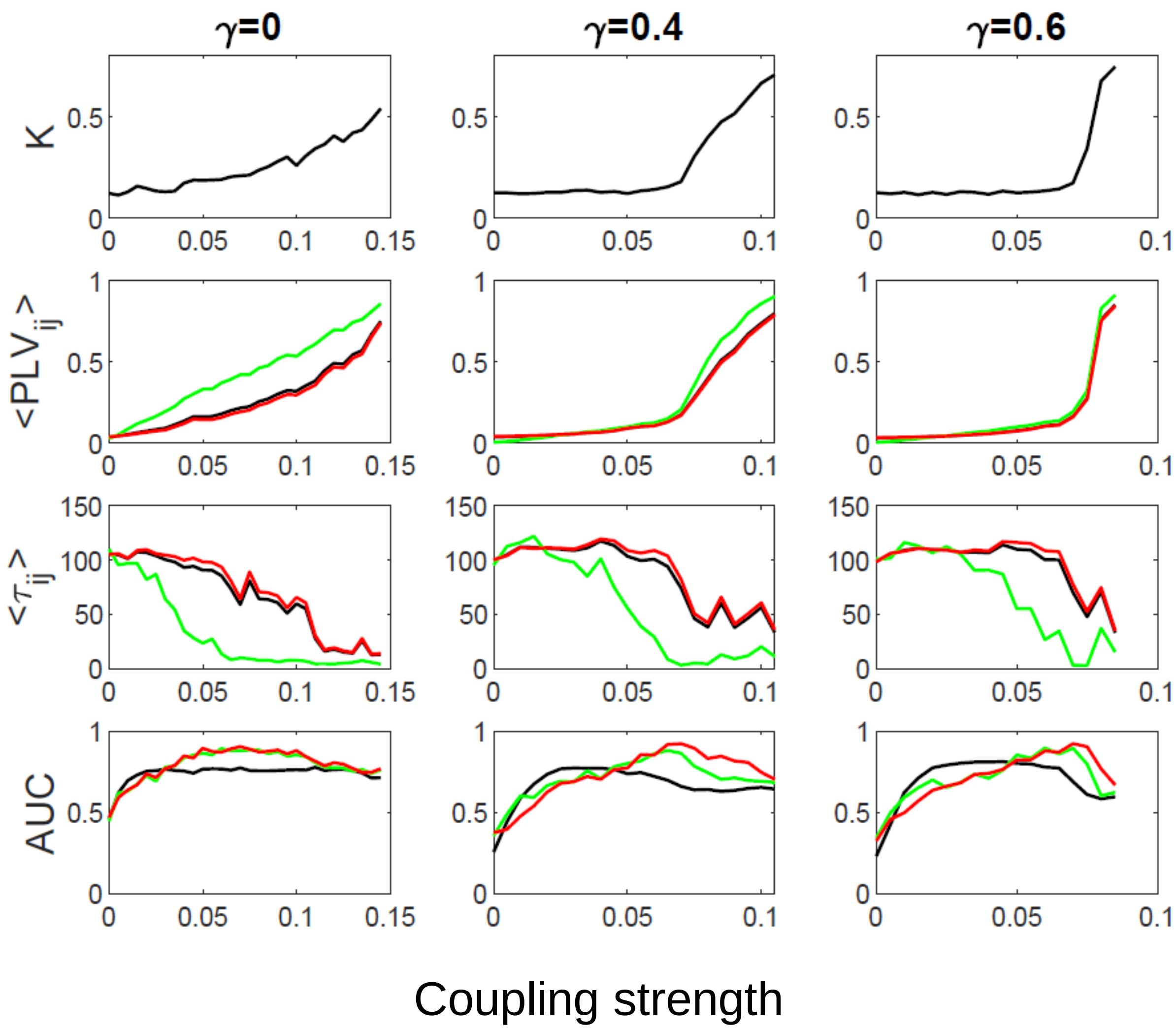
Results



γ=0
γ=0.4
γ=0.6
K
<ρij>
<τij>
AUC
All
Exist
No exist
SIM
AND
OR
Coupling strength

Results obtained using the *Phase Locking Value* as a measure of the similarity of two oscillators

$$\left\langle e^{i[\phi_i(t)-\phi_j(t-\tau)]} \right\rangle_t$$

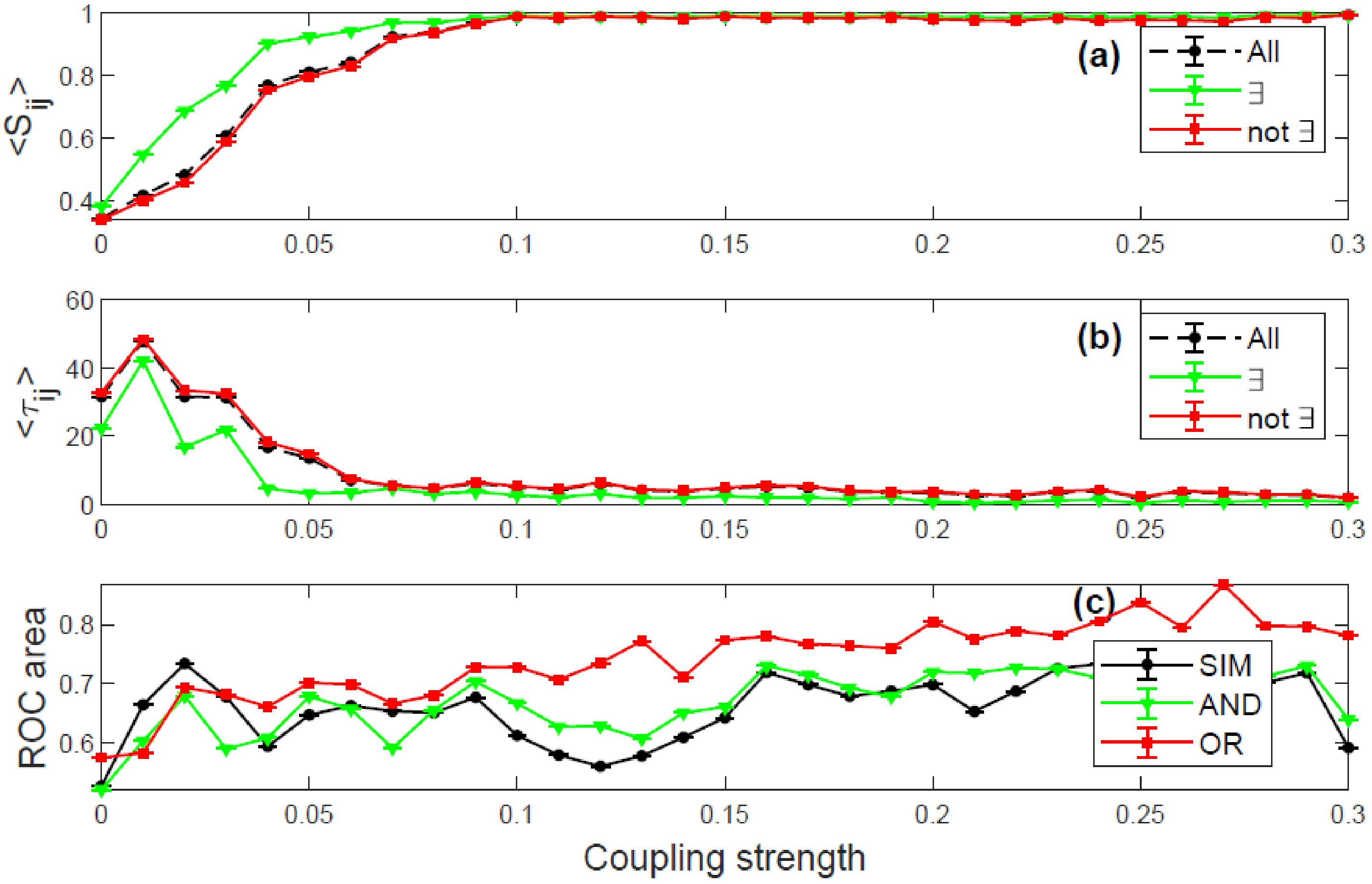


All
Exist
No exist

SIM AND OR

Results obtained from experimental data

28 electronic chaotic circuits, randomly connected



Data from: R. Sevilla-Escoboza & J. M. Buldu, Synchronization of networks of chaotic oscillators: Structural and dynamical data sets. Data in Brief 7 (2016) 1185–1189

Summary

- If we know the system's connectivity, lag information seems to be useful for anticipating the transition to synchronization (explosive or not).

- If we don't know the system's connectivity, lag information is not useful for inferring the links (but it could be useful to reduce certain types of mistakes – the false positives or the false negatives).

THANK YOU FOR YOUR ATTENTION !

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

ICREA