UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

**Nonlinear time series analysis**
Master degree in Industrial Engineering
Master degree in Aeronautic Engineering
Course 2020-2021

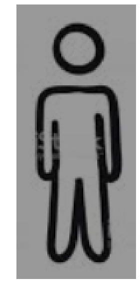
# Bivariate and multivariate analysis

Prof. Cristina Masoller
Universitat Politècnica de Catalunya
cristina.masoller@upc.edu
www.fisica.edu.uy/~cris

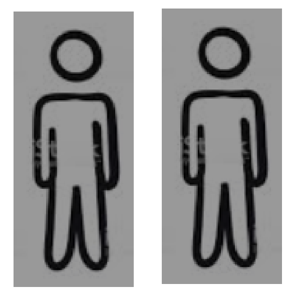
# Outline

- **Introduction**
  - Historical development: from dynamical systems to complex systems

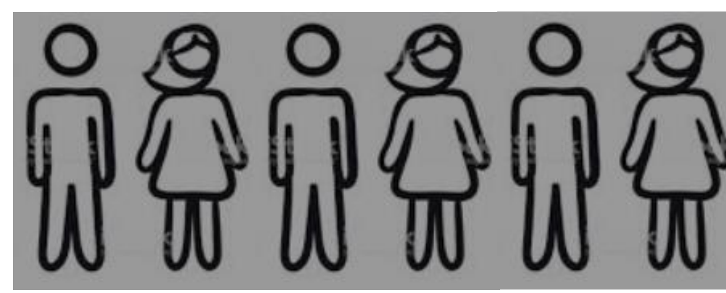
- **Univariate analysis**
  - Methods to extract information from a time series.
  - Applications.



- **Bivariate analysis**
  - Correlation, directionality and causality.
  - Applications.



- **Multivariate analysis**
  - Many time series: complex networks.
  - Network characterization and analysis.



# Cross-correlation of two time series *X* and *Y* of length *N*

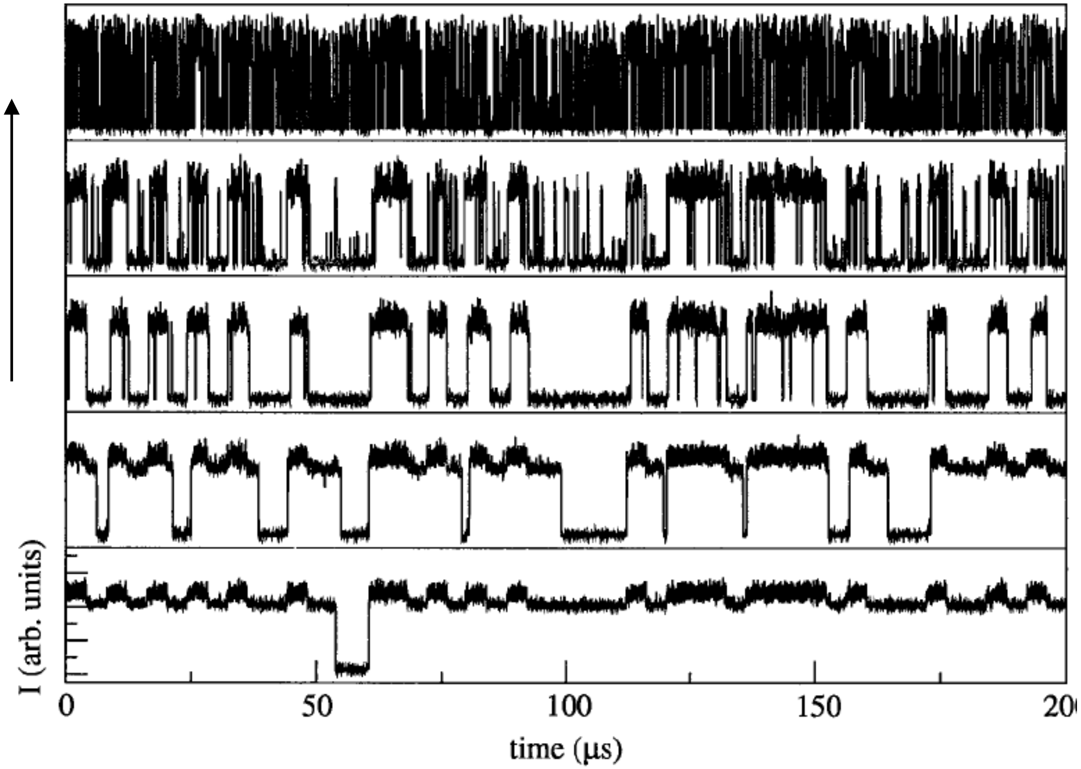
$$C_{xy}(\tau) = \frac{1}{N-\tau}\sum_{k=1}^{N-\tau} x(k+\tau)y(k)$$

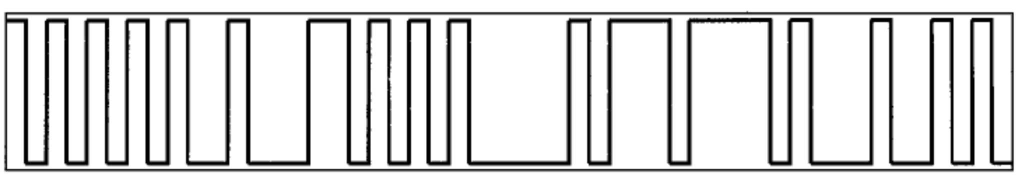
the two time series are normalized to zero-mean **$\mu$=0** and unit variance, **$\sigma$=1**

- $-1 \leq C_{X,Y} \leq 1$
- $C_{X,Y} = C_{Y,X}$
- The maximum of $C_{X,Y}(\tau)$ indicates the **lag** that renders the time series *X* and *Y* best aligned.
- Pearson coefficient: $\rho = C_{X,Y}(0)$

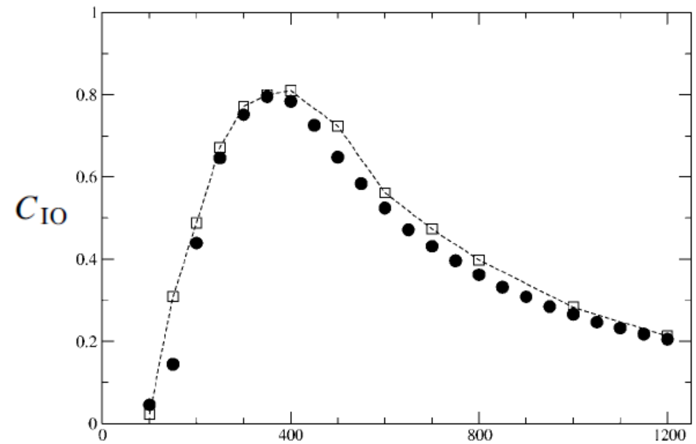
# Example: response of a bistable system to an aperiodic signal (stochastic resonance)

Noise level



Cross-correlation between input and output signal.

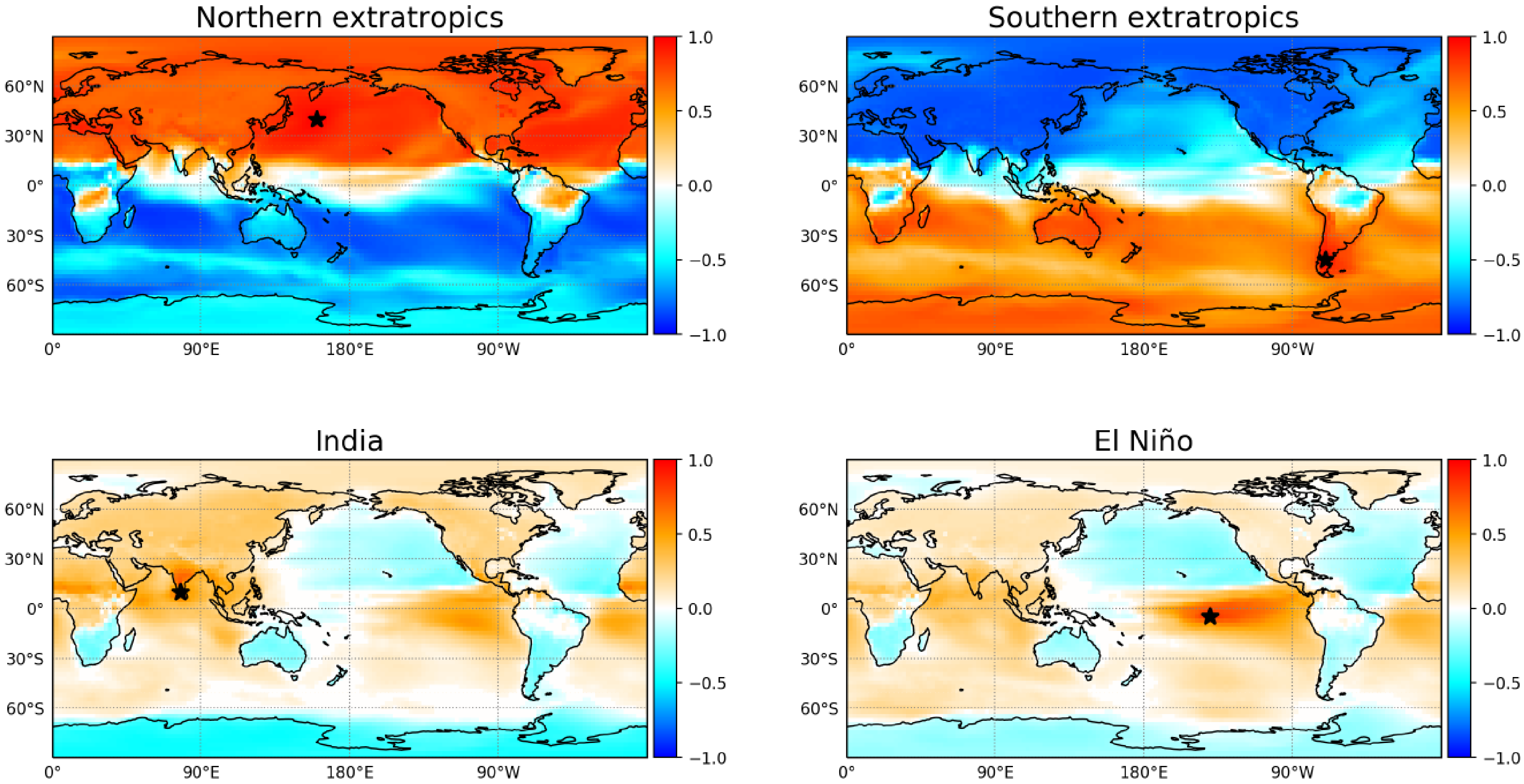



Noise level

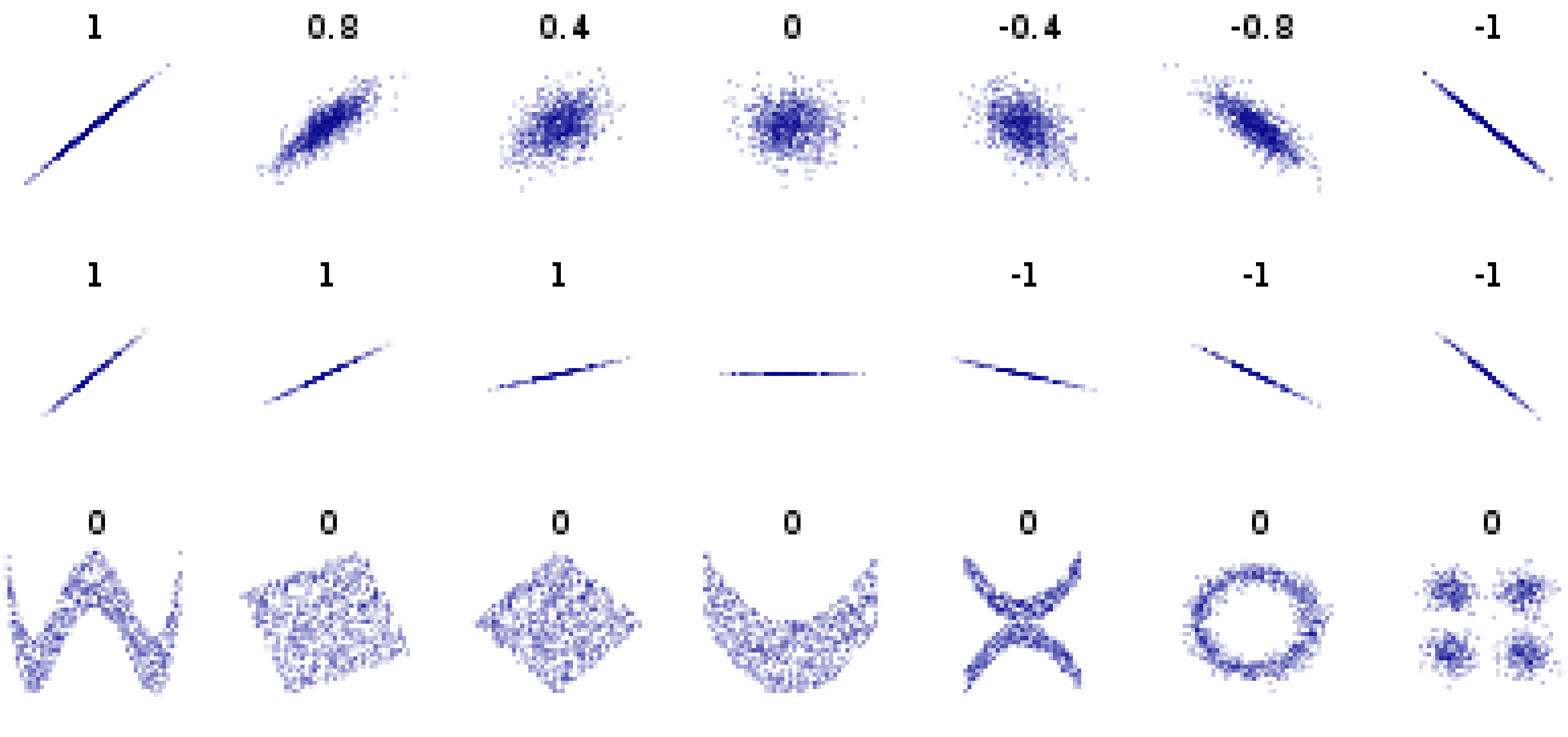


Barbay et al, PRL 85, 4652 (2000)

# Example: cross-correlation of cosine of Hilbert phase of SAT at a reference point (*), and all other regions



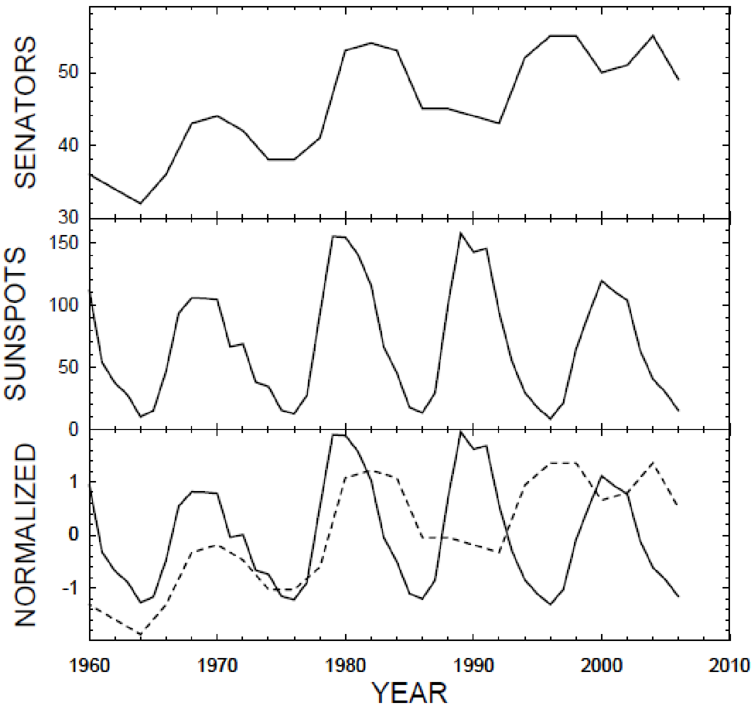
# Cross-correlation analysis detects linear relationships only




*Source: wikipedia*

## Correlation is NOT causality

An illustrative example: the number of sunspots and the number of the Republicans in the U.S. Senate in the years 1960-2006.



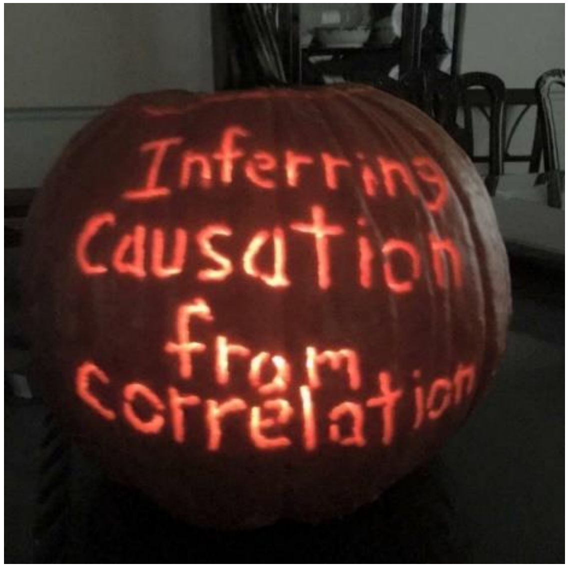

Interval 1960 to 1986 (biannual sampling, 14 points):

**C=0.52** **Is this significant?**

# Surrogate test

The significance of a correlation value is usually checked by calculating the cross-correlation from an ensemble of signals (**surrogates**) with the same **autocorrelation** than the original time series but completely independent from each other.

http://tylervigen.com/spurious-correlations




G. Lancaster et al, "*Surrogate data for hypothesis testing of physical systems*", Physics Reports 748 (2018) 1–60.

## Nonlinear correlation measure based on information theory: the mutual Information

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

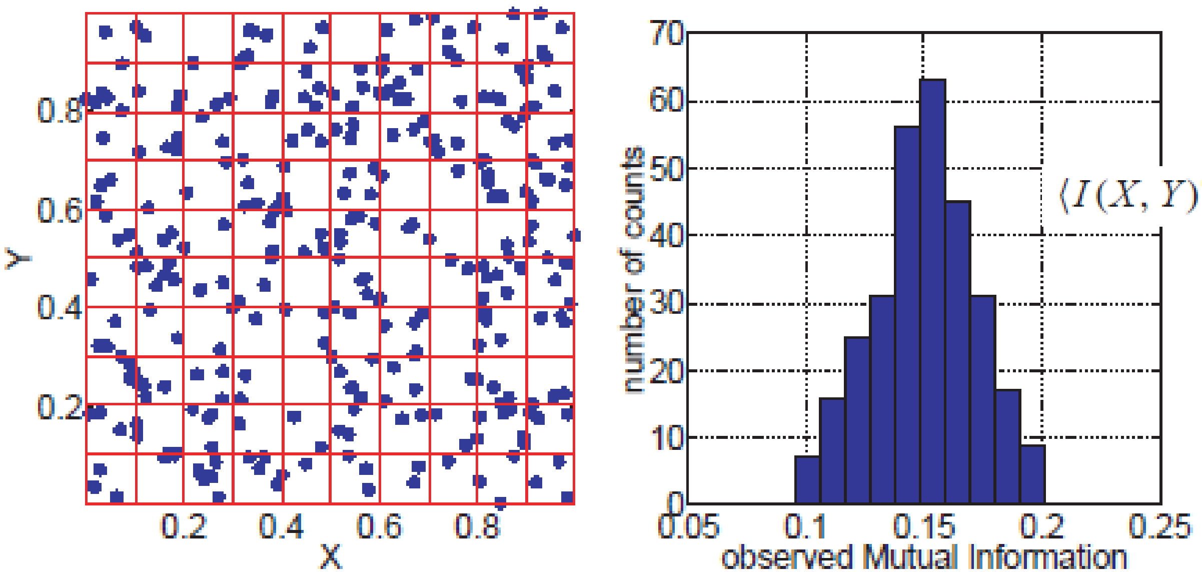
- *MI (x,y) = MI (y,x)*

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

- *MI* can also be computed with a lag-time.

- *MI* can also be computed from symbolic probabilities (e.g., probabilities of ordinal patterns).

# MI values are systematically overestimated




$\langle I(X, Y)^{estimated} \rangle \approx 0.15 \pm 0.02$

**Fig. 1.** Naive estimation of the mutual information for finite data. Left: The dataset consists of $N = 300$ artificially generated independent and equidistributed random numbers. The probabilities are estimated using a histogram which divides each axis into $M_x = M_y = 10$ bins. Right: The histogram of the estimated mutual information $I(X, Y)$ obtained from 300 independent realizations.

Main problem: a reliable estimation of MI requires a large amount of data

R. Steuer et al, Bioinformatics 18, suppl 2, S231 (2002).

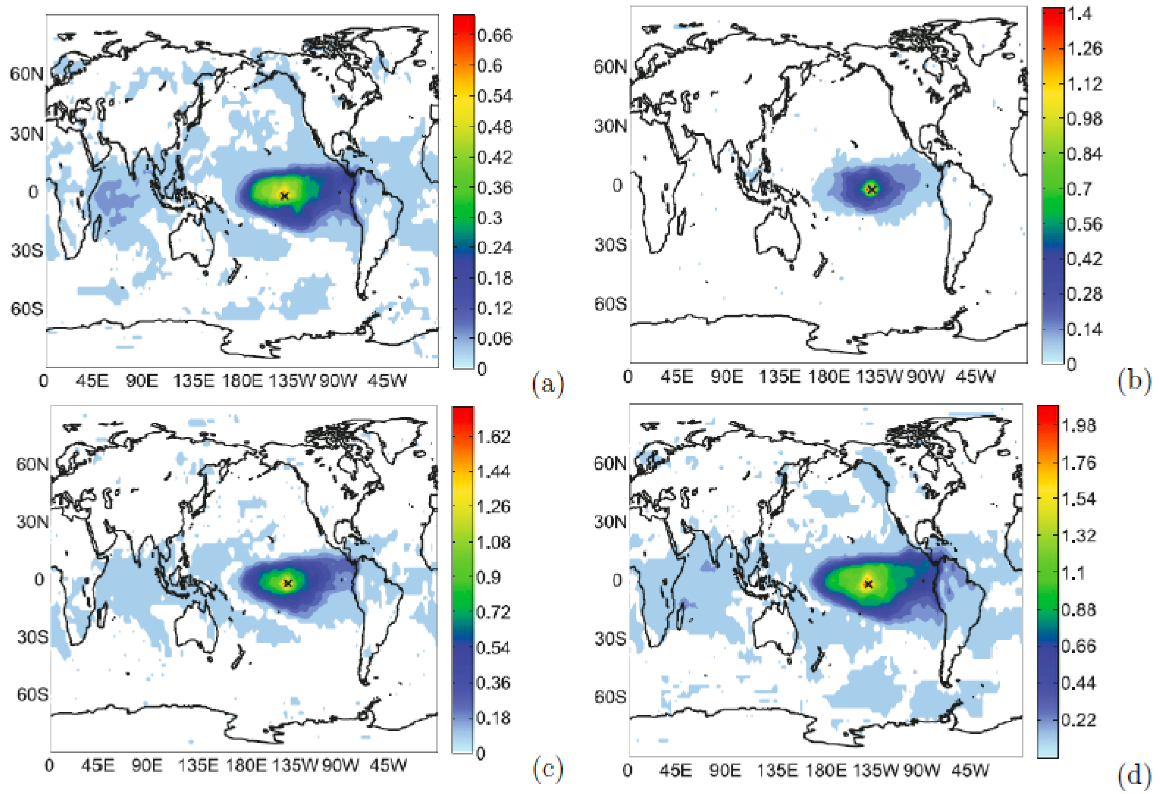
# Example: MI maps computed from SAT anomalies at a reference point located in El Niño, and all the other regions

Histograms

3 months ordinal patterns

Inter-annual ordinal patterns

3 years ordinal patterns




Ordinal analysis separates the times-scales of the interactions

Deza, Barreiro and Masoller, Eur. Phys. J. ST 222, 511 (2013)

# Direction of interaction?

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

*Campus d'Excel·lència Internacional*

# Conditional mutual information (CMI) and transfer entropy (TE)

- CMI measures the amount of information shared between two time series *i(t)* and *j(t)*, given the effect of a third time series, *k(t)*, over *j(t)*.

$$M_I(i;j|k) = \sum_{m,n,l} p_{ijk}(m,n,l) \log \frac{p_k(l) p_{ijk}(m,n,l)}{p_{ik}(m,l) p_{jk}(n,l)}$$

- Transfer entropy = CMI with the third time series, *k(t)*, replaced by the *past* of *i(t)* or *j(t)*.

$$\mathrm{TE}_{ij}(\tau) \equiv M_I(i;j|i_\tau) \qquad \mathrm{TE}_{ji}(\tau) \equiv M_I(j;i|j_\tau)$$

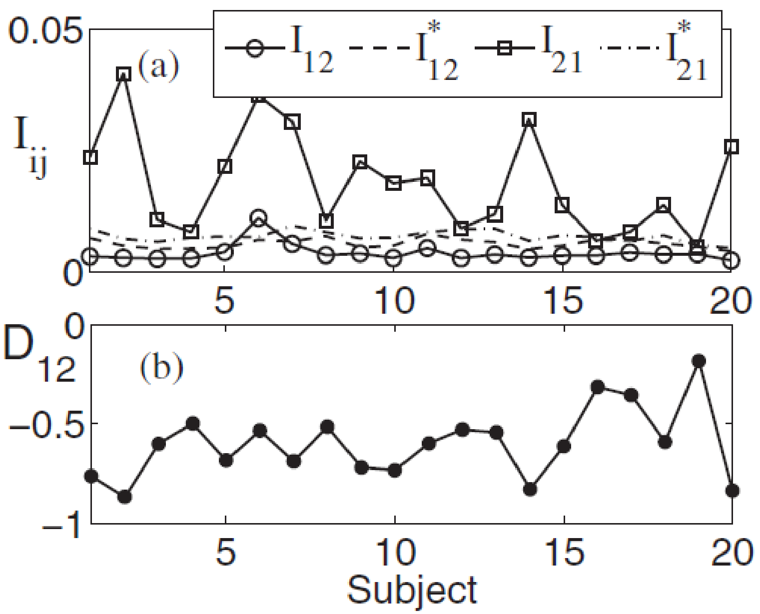
# Directionality index

- τ: *time-scale* of information transfer
- *DI*: *net direction* of information transfer
- $DI_{ij} > 0 \rightarrow i$ drives $j$.

$$\mathrm{DI}_{ij}(\tau) = \frac{\mathrm{TE}_{ij}(\tau) - \mathrm{TE}_{ji}(\tau)}{\mathrm{TE}_{ij}(\tau) + \mathrm{TE}_{ji}(\tau)}$$

Problem: $x \rightarrow i$, $x \rightarrow j$ $\quad i \leftrightarrow j$ ??

Application to cardiorespiratory data measured from 20 healthy subjects:
(a) TEs (dashed lines: surrogate data)
(b) $D_{12}$ (1 = heart; 2 = respiration).




$D_{12} < 0 \rightarrow$ respiration drives cardiac activity.

TEs were computed from ordinal probabilities and averaged over a short range of lags to decrease fluctuations.

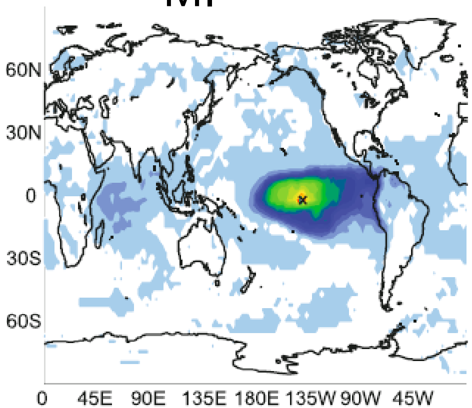
A. Bahraminasab et al., PRL 100, 084101 (2008)

# Application to climate data

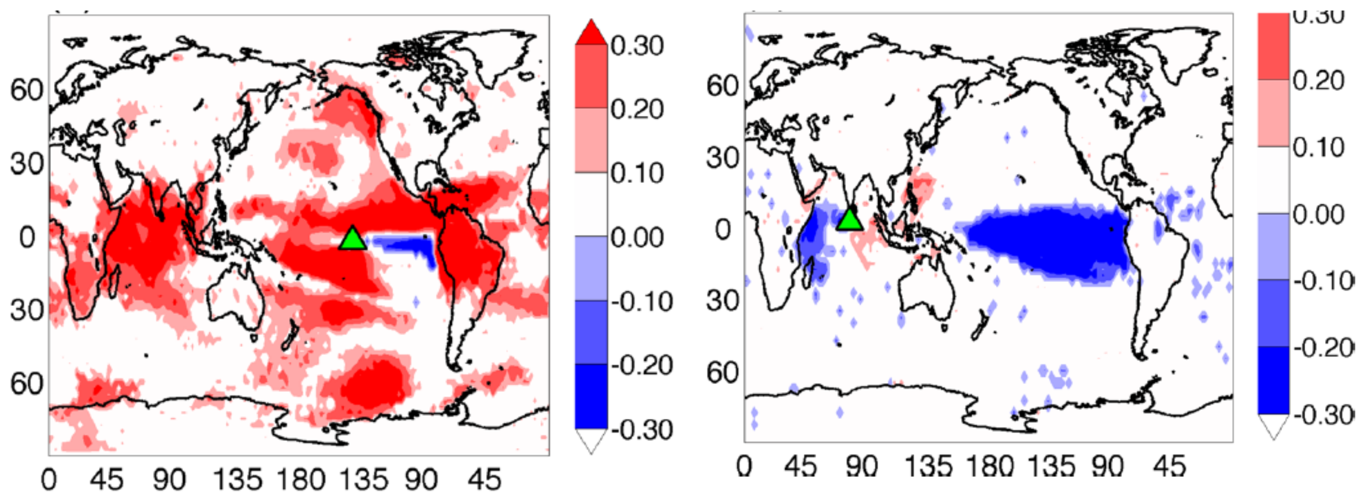
DI computed from daily SAT anomalies, PDFs estimated from histograms of values.

MI and DI are _both significant_ (>3$\sigma$, surrogates), $\tau$=30 days.

MI




DI




[J. I. Deza, M. Barreiro, and C. Masoller, "*Assessing the direction of climate interactions by means of complex networks and information theoretic tools*", Chaos 25, 033105 (2015).]

# Causality?

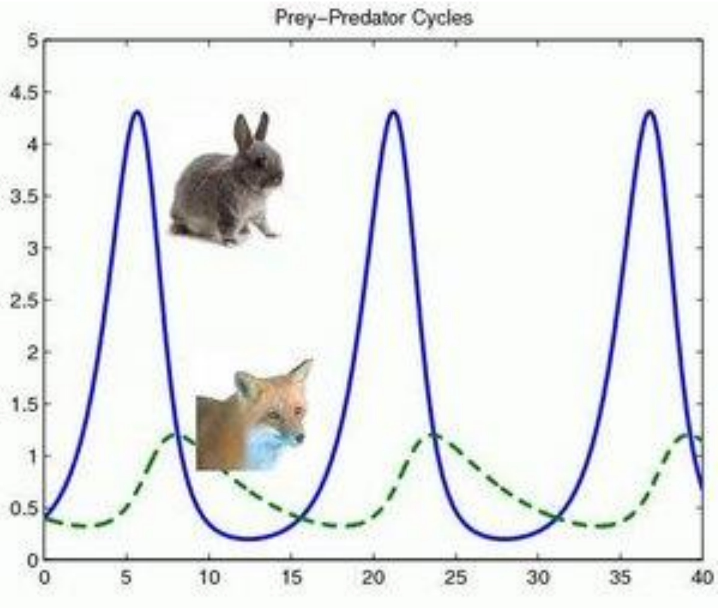
UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

# Main idea

- A time series *X* is Granger causal to a time series *Y* ($X \rightarrow Y$) if the information given by *X* allows for a more precise prediction of *Y*.

- Example: in the predator - prey system, information about variations in the predator population can reveal properties of the prey population.



# Granger causality: how to detect X→Y

- Model *Y* as a processes with memory forced by *X* with residual noise ε

$$Y_t = \sum_{k=1}^{D} a_k Y_{t-k} + \sum_{k=1}^{D} b_k X_{t-k} + \varepsilon_t$$

- Test the hypothesis *b* ≠ 0 against the null hypothesis *b*=0:
  - Fit vectors *a* and *b* with a linear regression and compute the variance of the residual: $\sigma^2_{\mathrm{coupled}}$
  - Repeat with *b*=0 and compute: $\sigma^2_{\mathrm{uncoupled}}$
  - Then calculate the Granger Causality Estimator

# Granger Causality Estimator

$$GCE = \frac{\sigma^2_{\text{uncoupled}} - \sigma^2_{\text{coupled}}}{\sigma^2_{\text{uncoupled}}}$$

- If *GCE*>0 the information given by *X* allowed for a more precise prediction of *Y*.

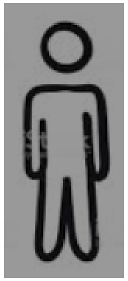
- Problems:
  - how to select the dimension *d*?
  - how to test the statistical significance of the GCE value?

# Summary

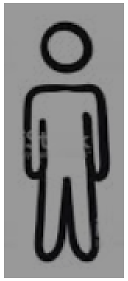
- Cross-correlation: detects linear interdependencies.
- Mutual information: detects nonlinear interdependencies.
- The MI computed from the probabilities of ordinal patterns allows to select the time-scale of the analysis.
- The directionality index detects the net direction of the information flow.
- Granger causality can "disentangle" mutual interactions.

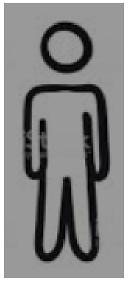
# Outline

- **Introduction**
  - Historical development: from dynamical systems to complex systems
- **Univariate analysis**
  - Methods to extract information from a time series.
  - Applications.

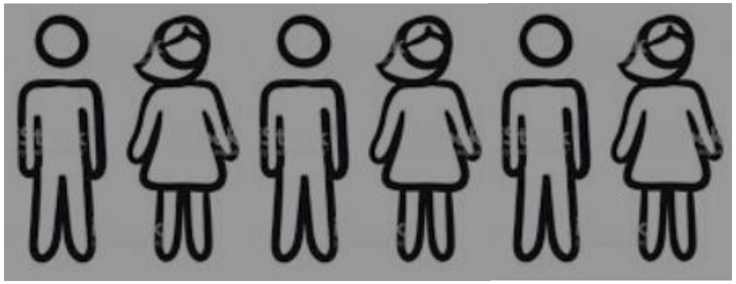


- **Bivariate analysis**
  - Correlation, directionality and causality.
  - Applications.



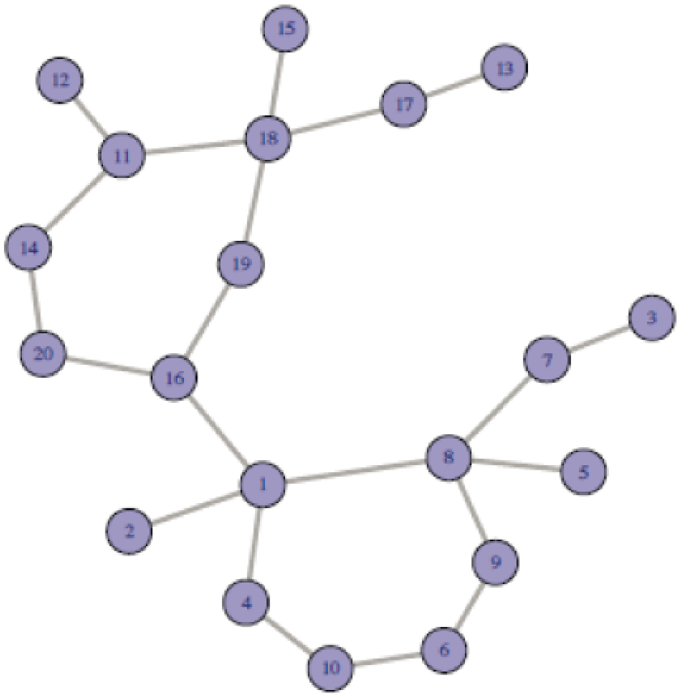


- **Multivariate analysis**
  - Many time series: complex networks.
  - Network characterization and analysis.



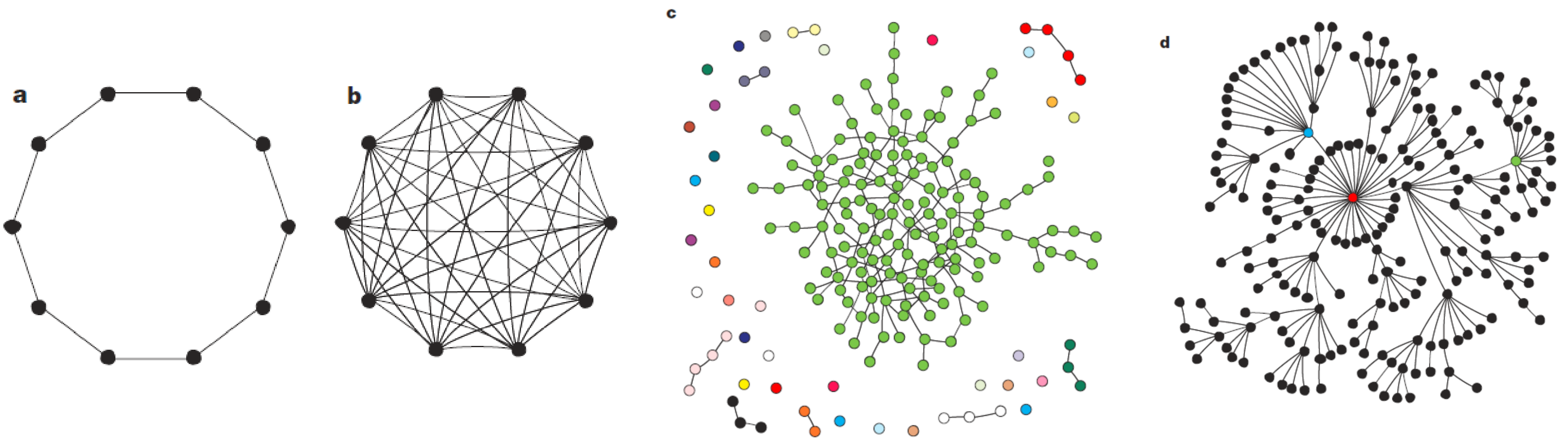
## What is a network?

- A graph: a set of "nodes" connected by a set of "links".
- Nodes and links can be weighted or unweighted.
- Links can be directed or undirected.



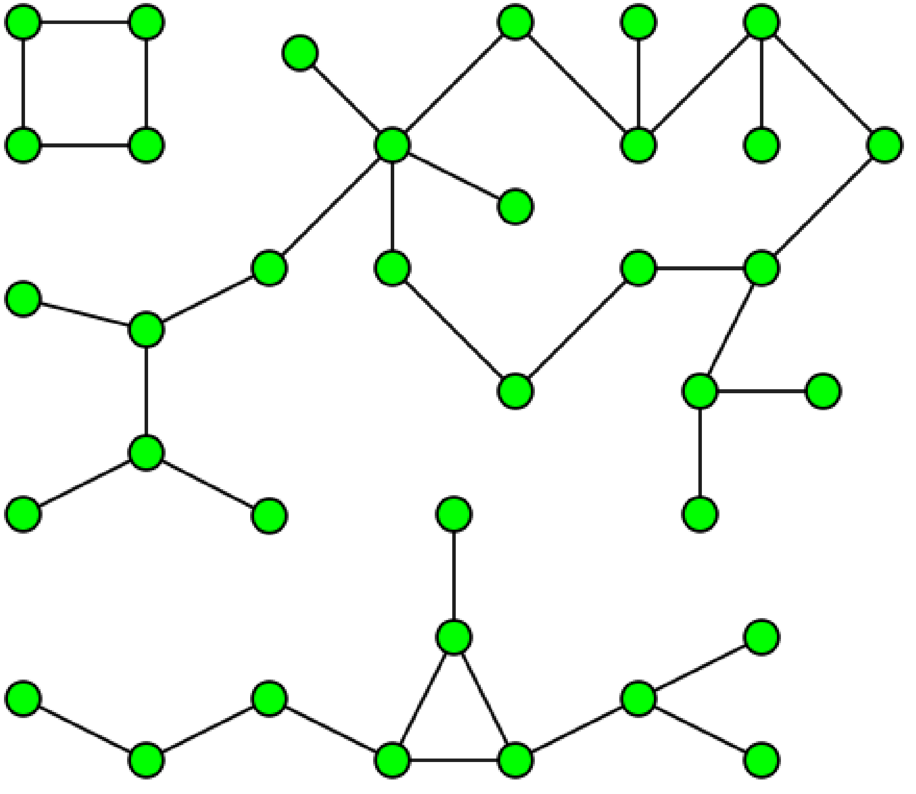
## Networks or graphs

The challenge in the context of time series analysis: to infer the underlying network structure from observed signals.




Source: Strogatz
Nature 2001

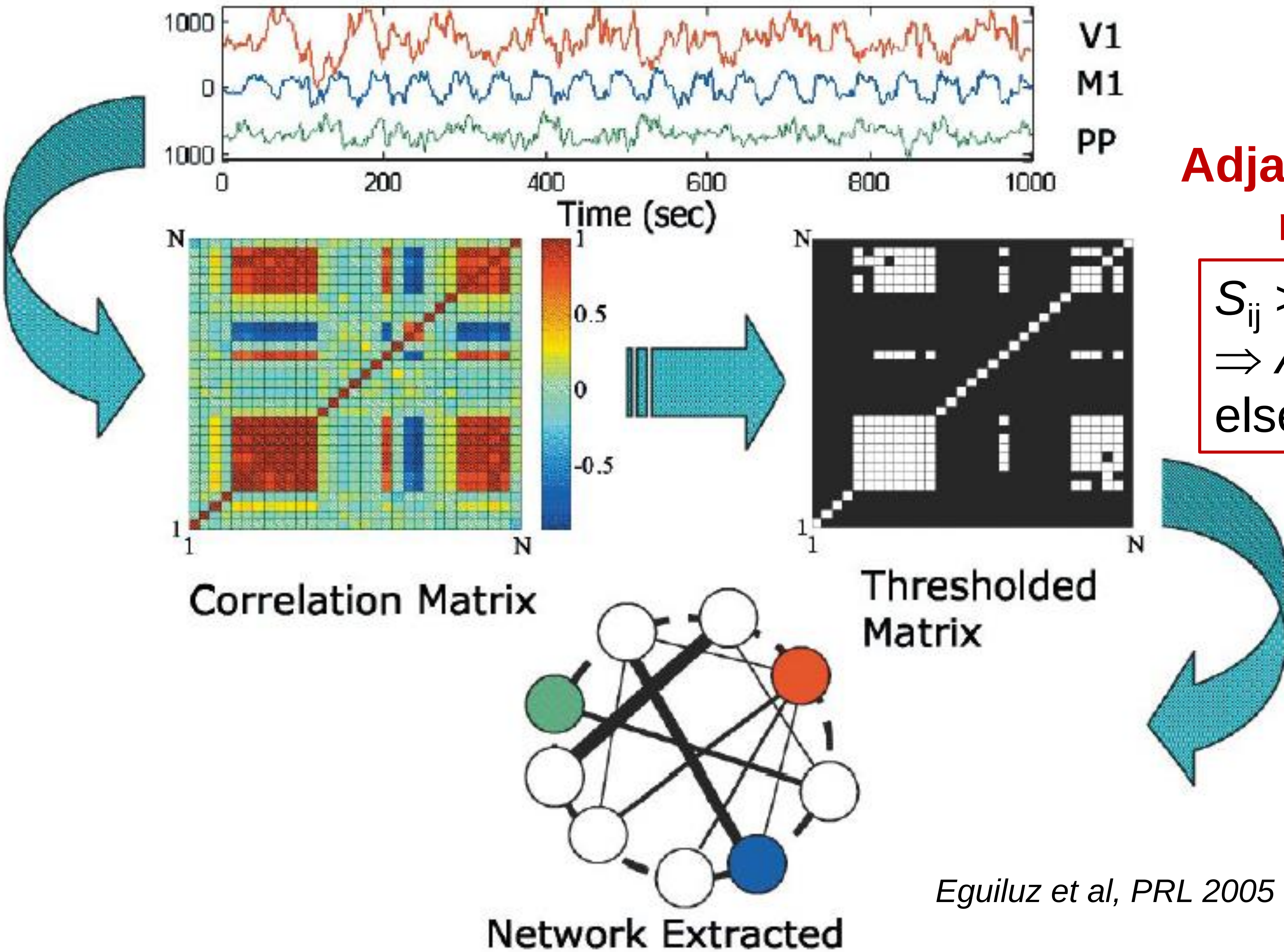
# Connected components (“communities”)



A graph with three connected components.
Source: Wikipedia

# Using bivariate measures to infer interactions from data: "functional networks"

# Brain functional network



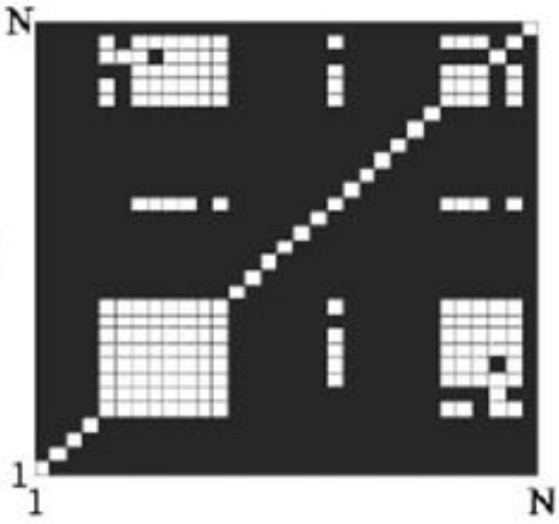

**Adjacency matrix**

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

*Eguiluz et al, PRL 2005*

# Graphical representation

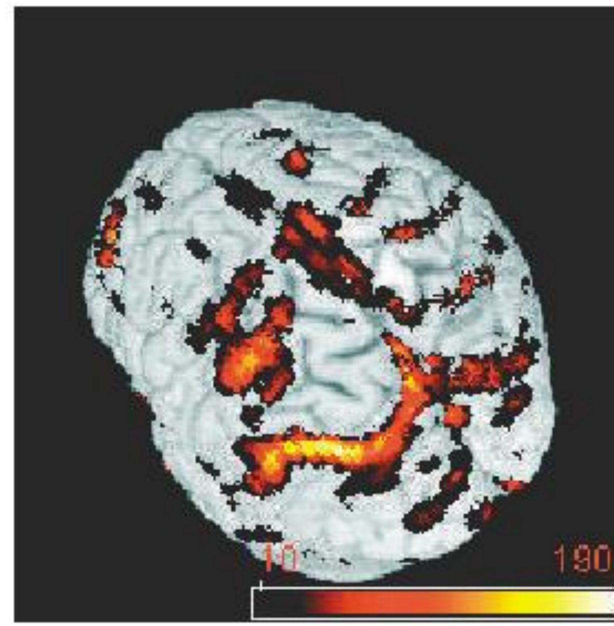
Adjacency matrix




Thresholded matrix = inferred (“functional”) network

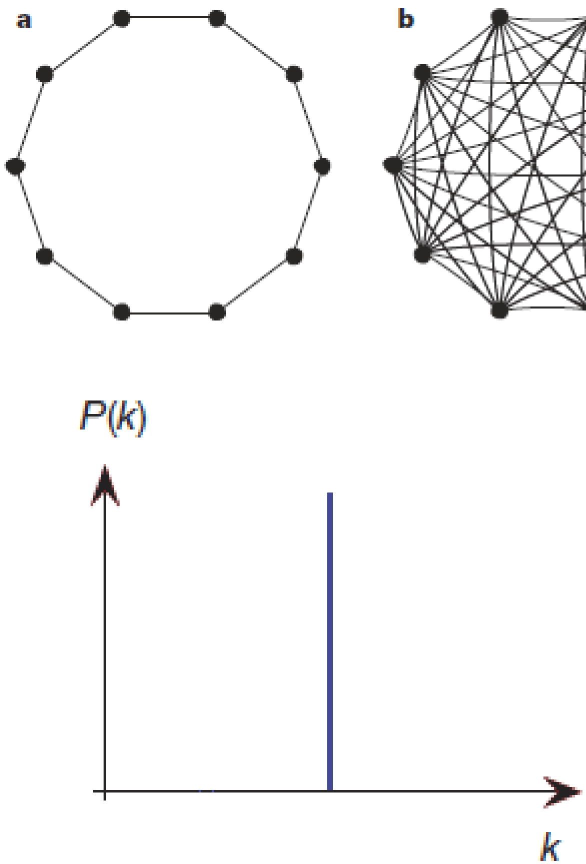
**Degree of a node**: number of links

$$k_i = \Sigma_j A_{ij}$$

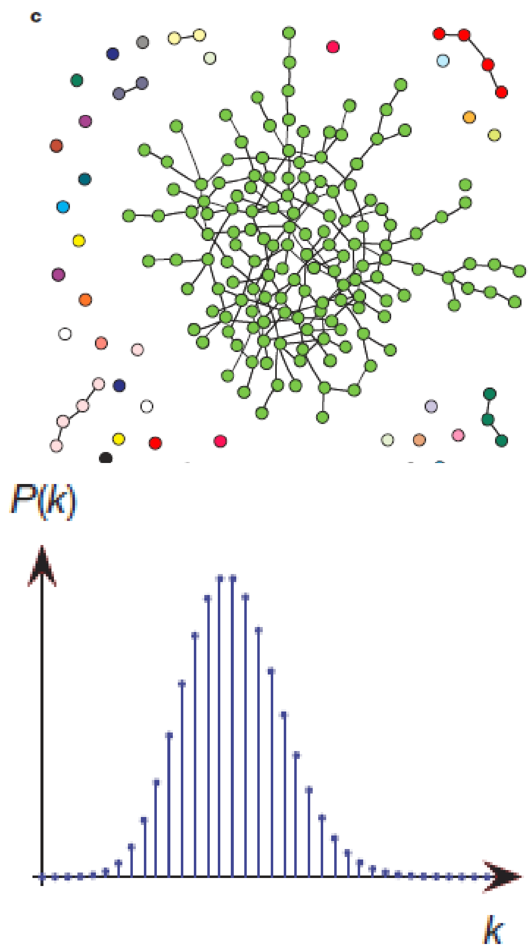


# The degree distribution: usual way to characterize a graph

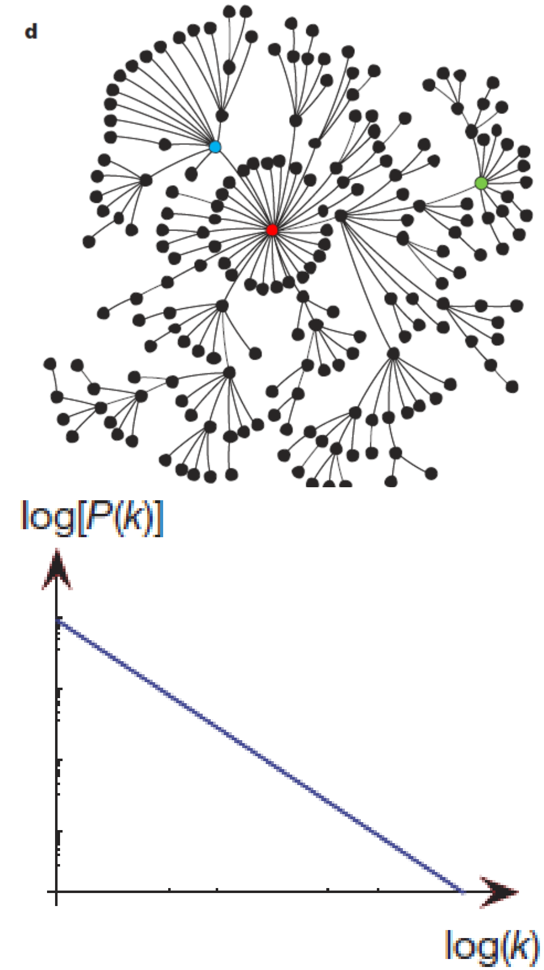
Regular




Random

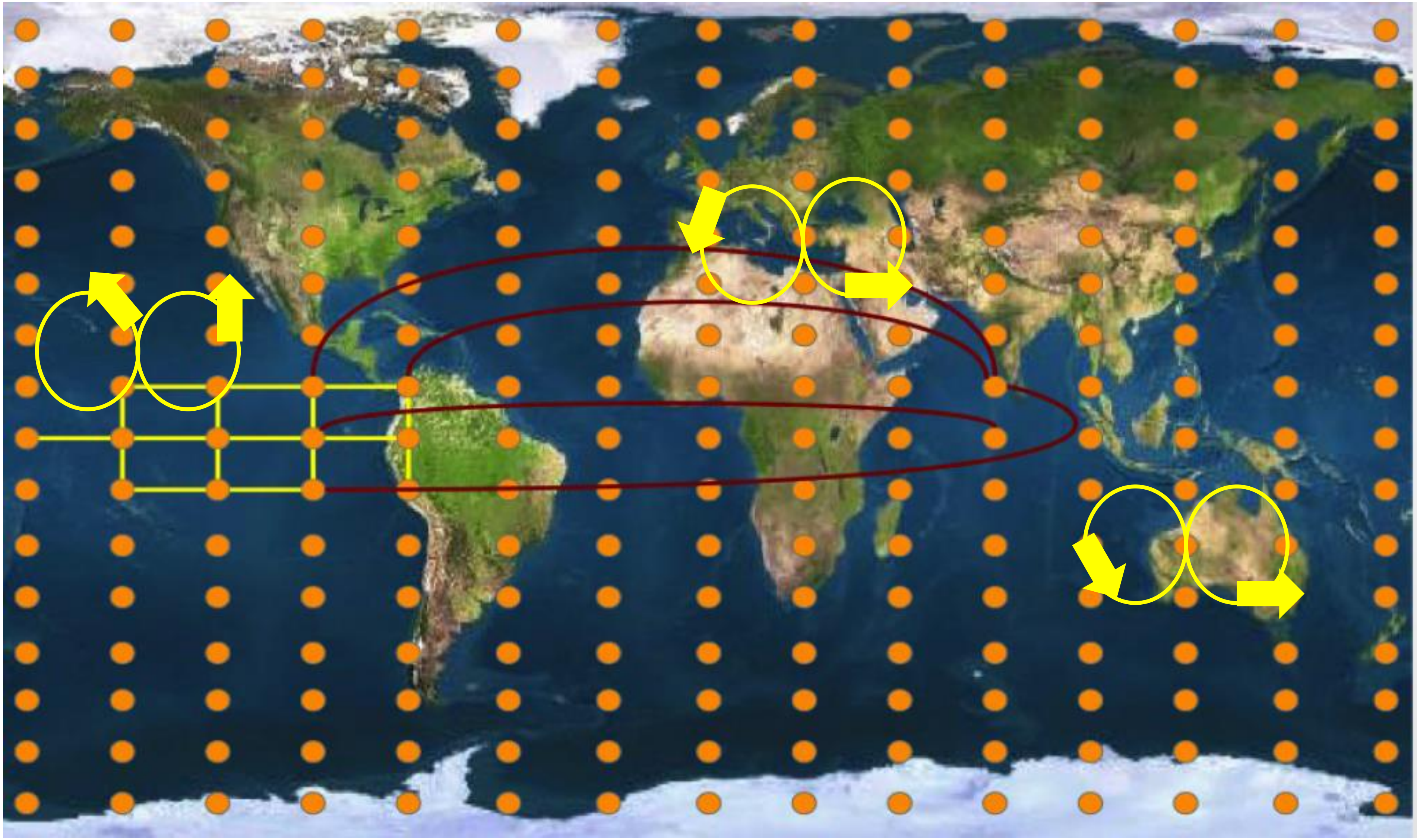



Scale-free

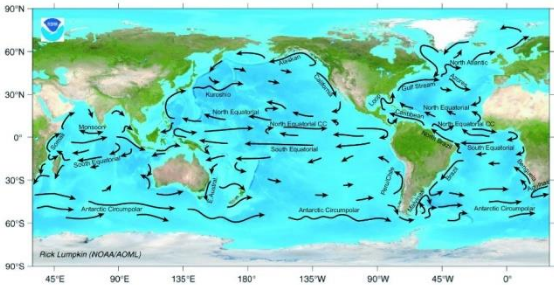



*Strogatz, Nature 2001*

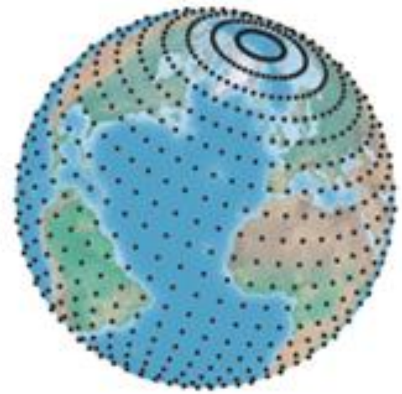
# The climate system as a set of “interacting oscillators”



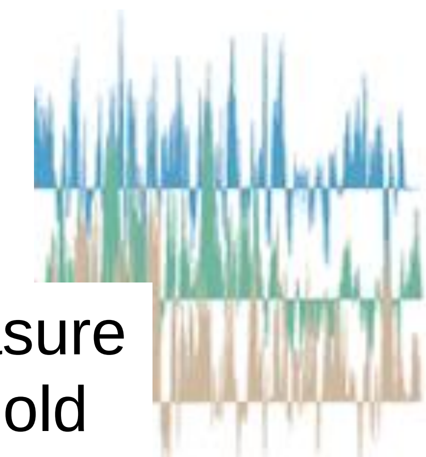
# Complex network representation of the climate system



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



More than 10000 nodes (with different sizes).



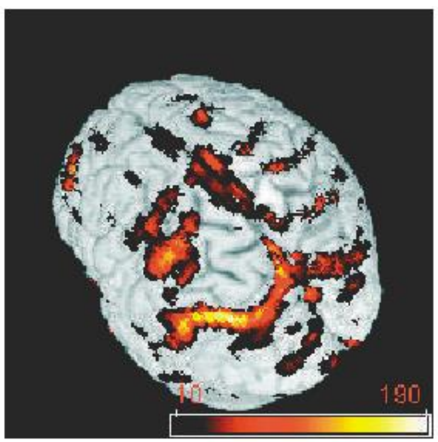
Daily resolution: more than 13000 data points in each TS

Sim. measure + threshold

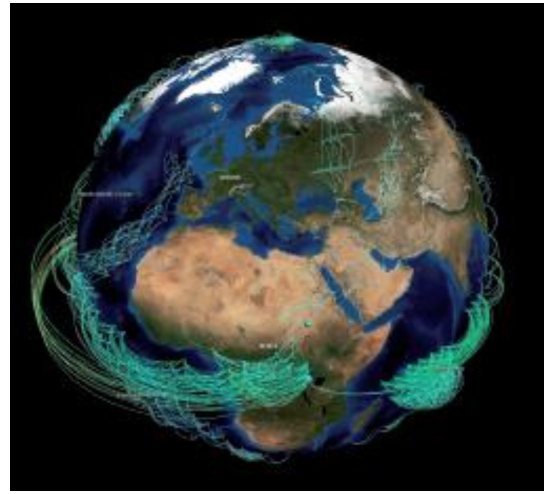
Surface Air Temperature <u>Anomalies</u> (solar cycle removed)

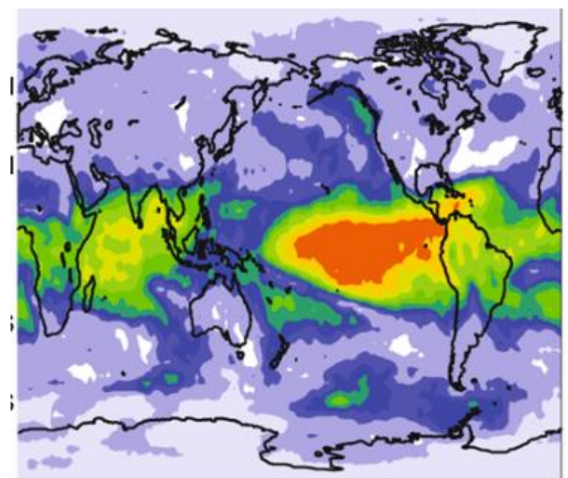
*Donges et al, Chaos 2015*

**Brain network**




**Climate network**





Area weighted connectivity (AWC):

weighted degree (nodes represent areas with different sizes)

$$\mathrm{AWC}_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

## How to select the threshold ?

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

Three criteria are typically used:

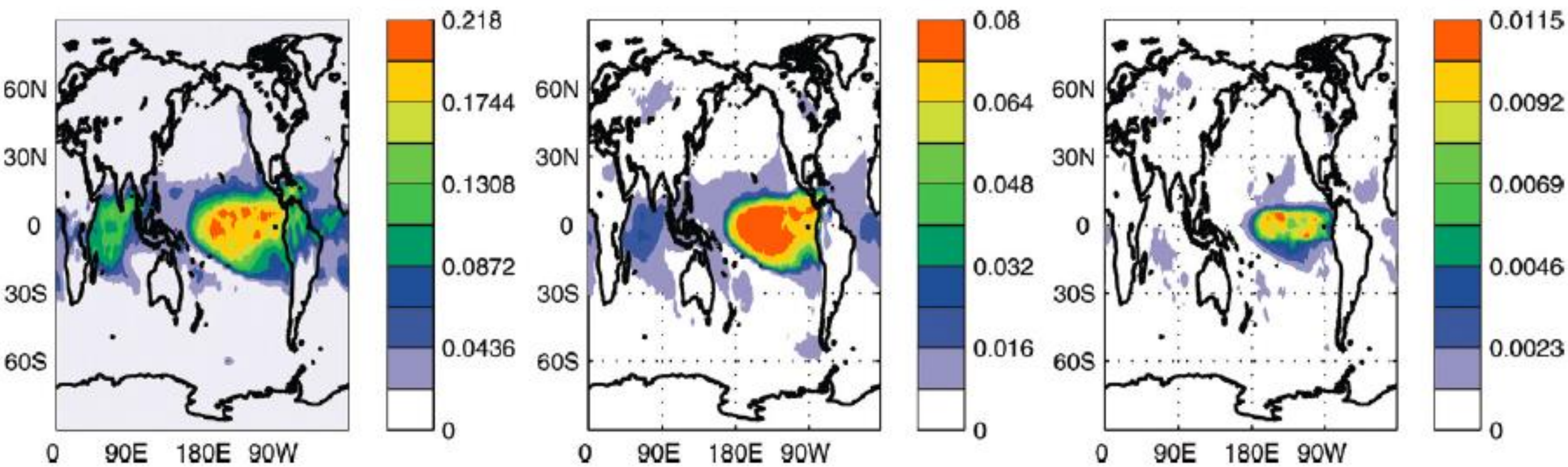
- A significance level is used (typically 5%) in order to omit connectivity values that can be expected by chance;
- We select an arbitrary value as threshold, such that it gives a certain pre-fixed number of links (or link density);
- We define the threshold as large as possible while guaranteeing that all nodes are connected (or a so-called “giant component” exists).

C. M. van Wijk et al., “*Comparing Brain Networks of Different Size and Connectivity Density Using Graph Theory*”, PLoS ONE 5, e13701 (2010)

**How to select the threshold ?**

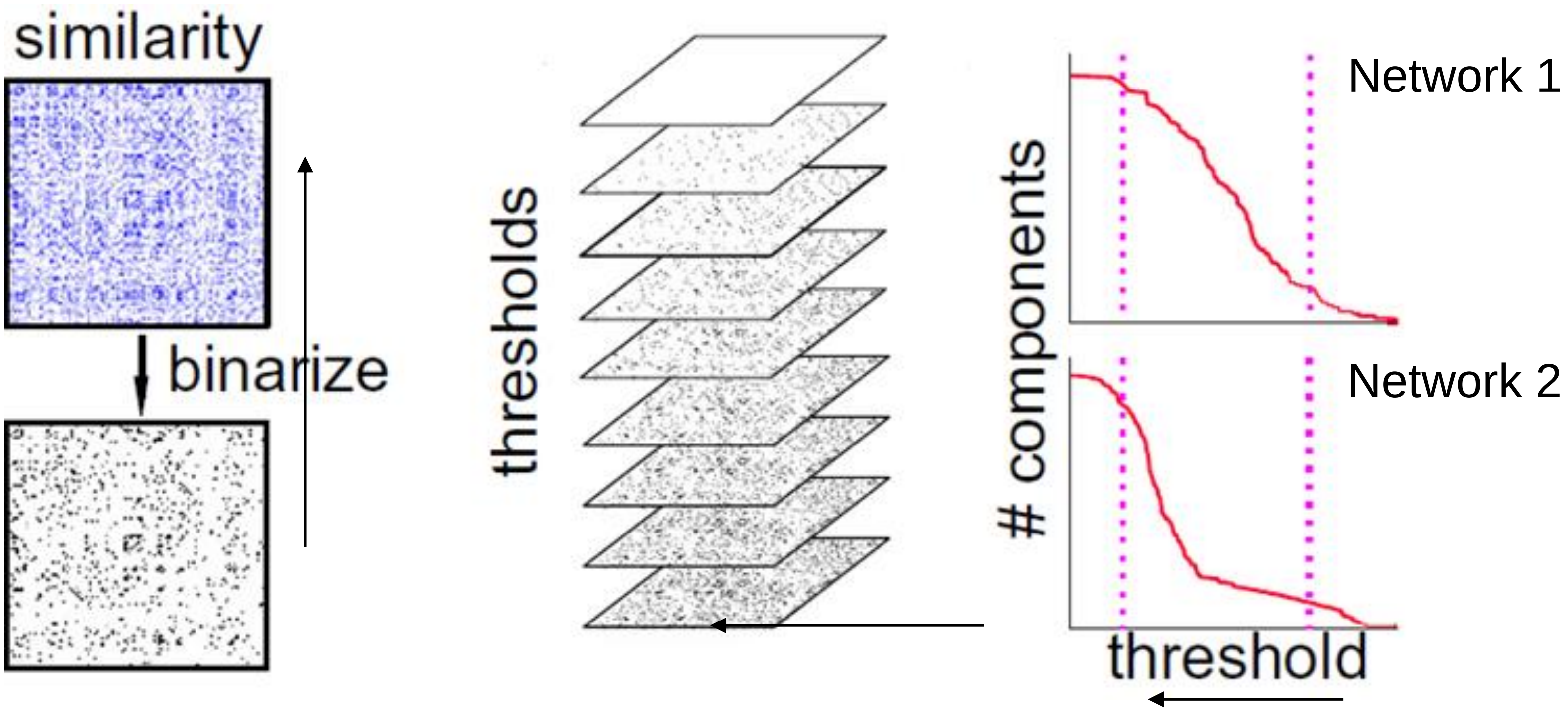
If $S_{ij}$ > Th $\Rightarrow$ $A_{ij}$ = 1,
else $A_{ij}$=0

Th $\Rightarrow$




M. Barreiro, et. al, Chaos 21, 013101 (2011)

# 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*

# Network characterization

# Definitions (for unweighted and undirected graphs)

- ***Adjacency matrix***: $A_{ij}$ = 1 if $i$ and $j$ are connected, else $A_{ij}$ = 0.

- ***Degree*** of a node $k_i = \Sigma_j A_{ij}$

- ***Clustering coefficient***: measures the fraction of a node's neighbors that are neighbors also among themselves

$$C_i = \frac{2R_i}{k_i(k_i-1)} = \frac{1}{k_i(k_i-1)} \sum_{j=1}^{N} \sum_{l=1}^{N} A_{ij} A_{jl} A_{li}$$

$R_i$ is the number of connected pairs in the set of neighbors of node $i$

- ***Assortativity***: measures the tendency of a node with high/low degree to be connected to other nodes with high/low degree

$$a_i \equiv \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j$$

# How to characterize the degree distribution?

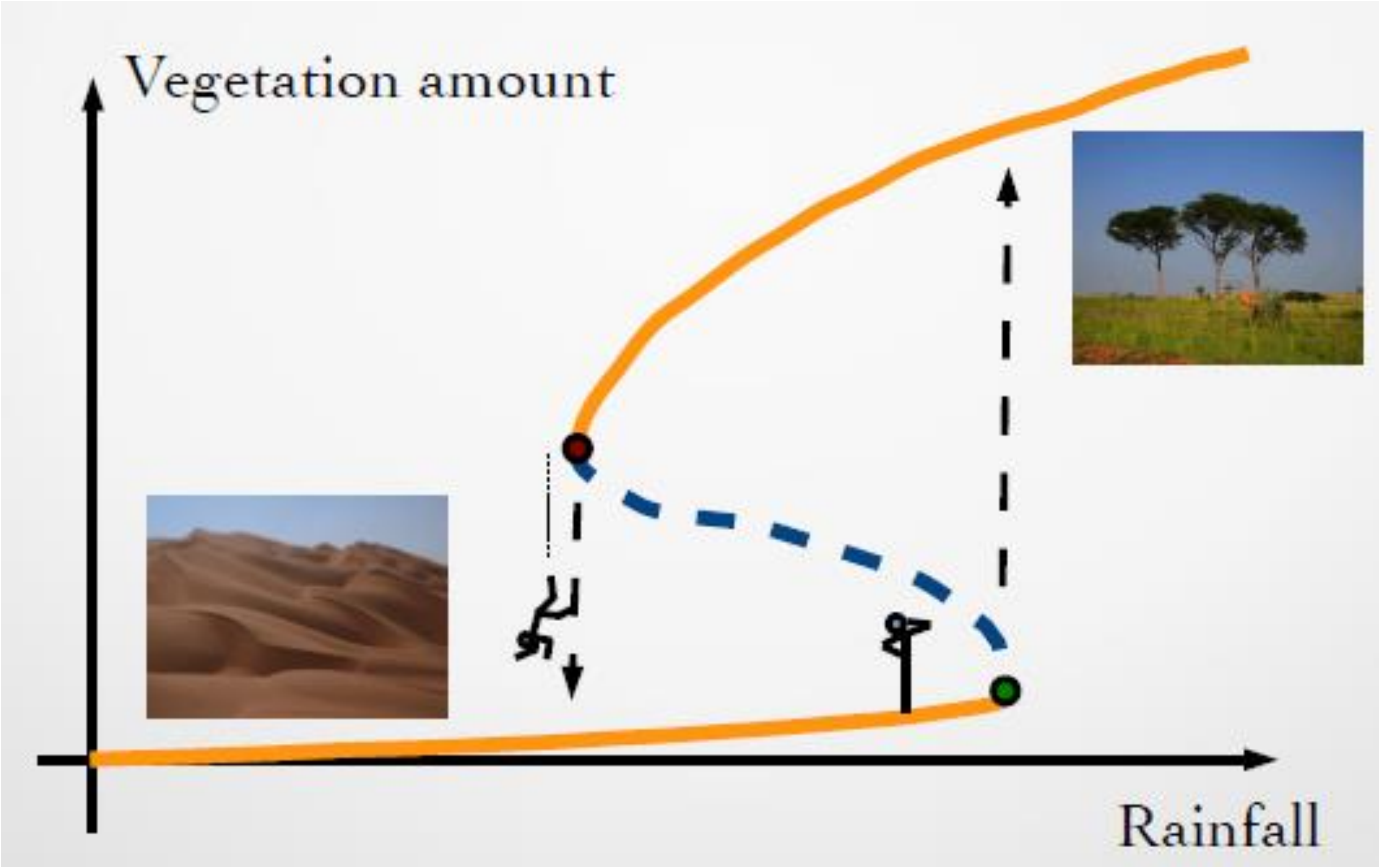
- **Mean** (expected value of $X$): $\mu = \mathrm{E}[X] = \sum_{i=1}^{k} x_i\, p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k$
- **Variance**: $\sigma^2 = Var\,(X) = \mathrm{E}[(X-\mu)^2]$
- **Skewness**: “measures” the asymmetry of the distribution

$$Z = \frac{X - \mu}{\sigma} \qquad S = \mathrm{E}[Z^3]$$

- **Kurtosis**: measures the "tailedness“ of the distribution. For a normal distribution $K$=3.

$$K = \mathrm{E}[Z^4]$$

# Example of application: desertification transition




Our goal: to develop reliable early-warning indicators

Can we use "correlation networks" to detect the approach to a tipping point?

# Model

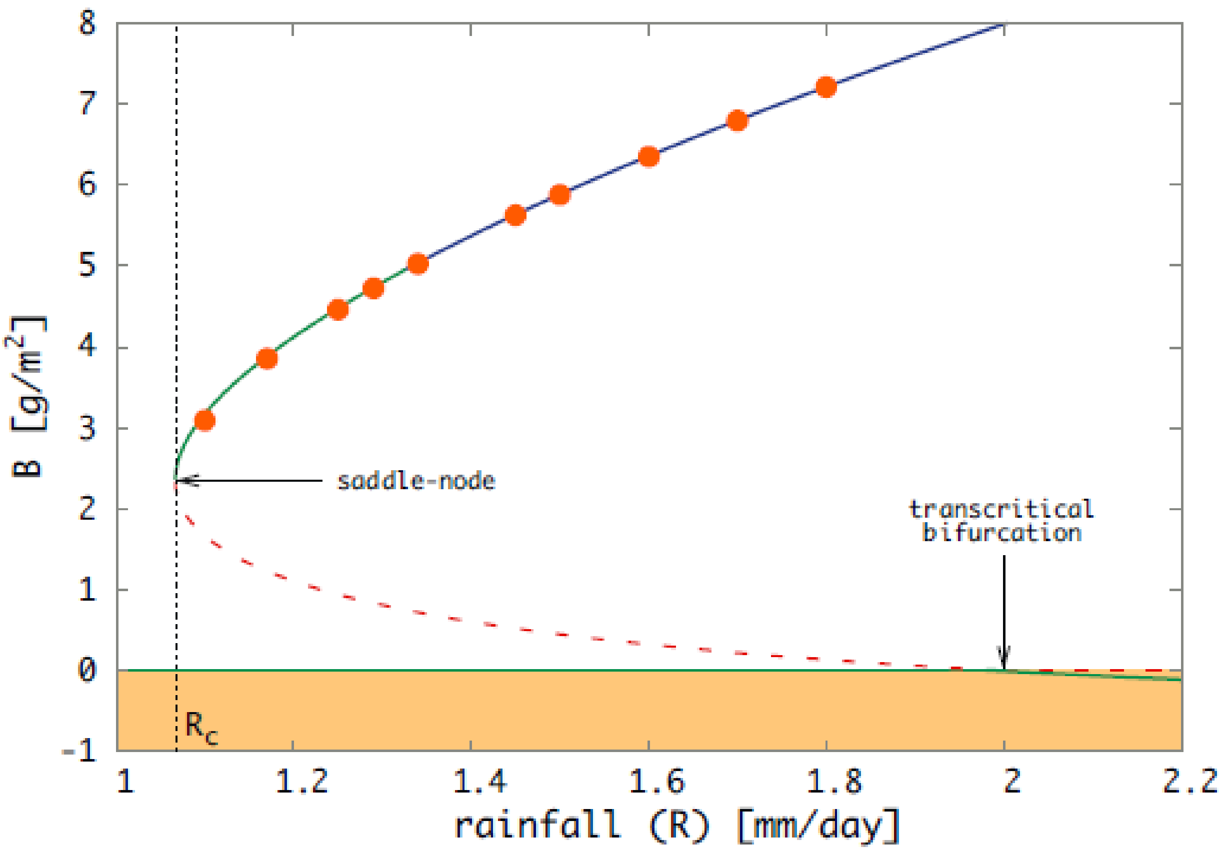
$$\frac{\partial w}{\partial t} = R - \frac{w}{\tau_w} - \Lambda wB + D\nabla^2 w + \sigma_w w_0 \xi^w(t),$$

$$\frac{\partial B}{\partial t} = \rho B\left(\frac{w}{w_0} - \frac{B}{B_c}\right) - \mu \frac{B}{B + B_0} + D\nabla^2 B + \sigma_B B_0 \xi^B(t)$$

- *w* (in mm) is the soil water amount
- *B* (in g/m$^2$) is the vegetation biomass
- Uncorrelated Gaussian white noise
- *R* (rainfall) is the bifurcation parameter

*Shnerb et al. (2003), Guttal & Jayaprakash (2007), Dakos et al. (2011)*

# Saddle-node bifurcation

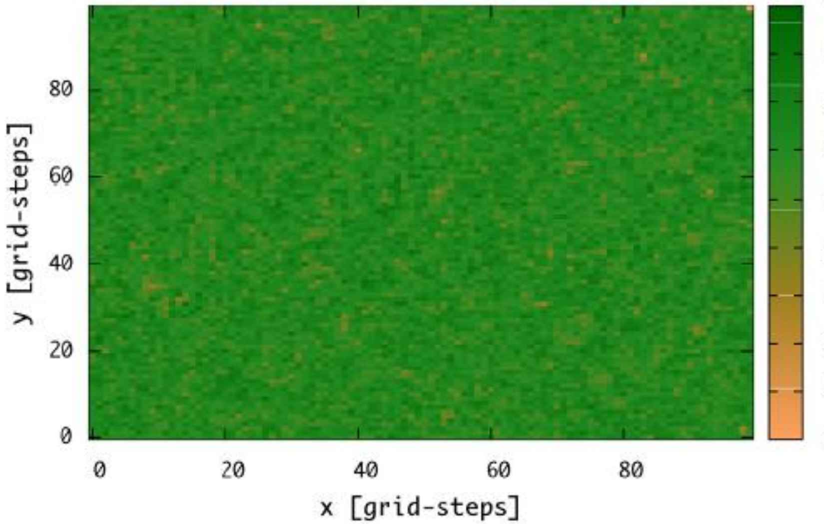


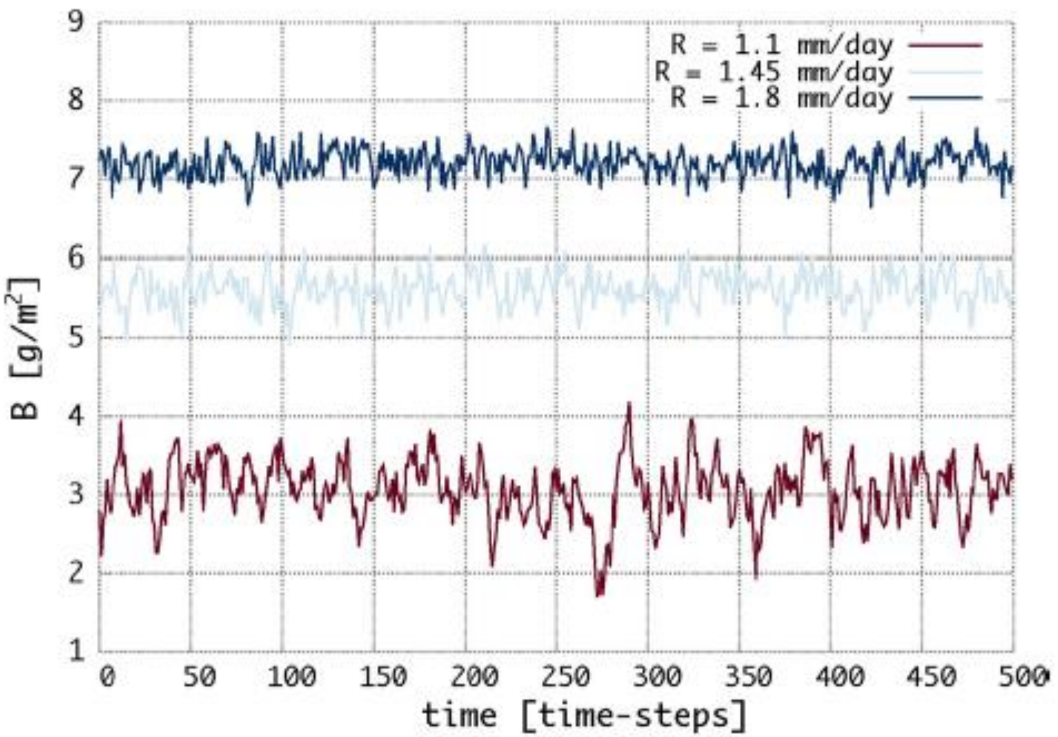

$R<R_c$: only desert-like solution (B=0)

**$R_c$ = 1.067 mm/day**

# Biomass time series

Biomass *B* when *R*=1.1 mm/day



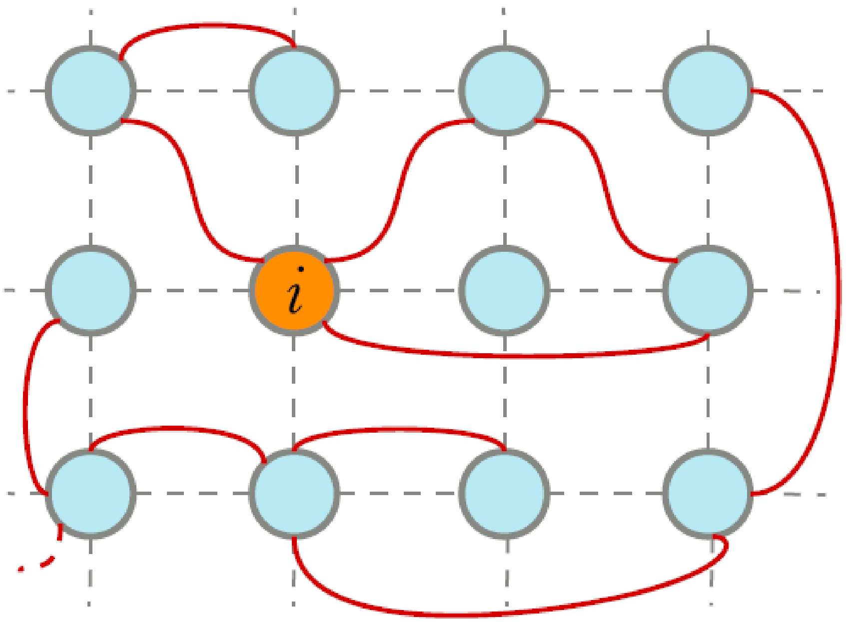

100 m x 100 m = $10^4$ grid cells
Simulation time 5 days in 500 time steps
Periodic boundary conditions

# Correlation Network

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

Statistical similarity measure: Pearson coef.= |zero-lag cross-correlation|

Threshold: Th=0.2 keeps only significant correlations ($p$<0.05)



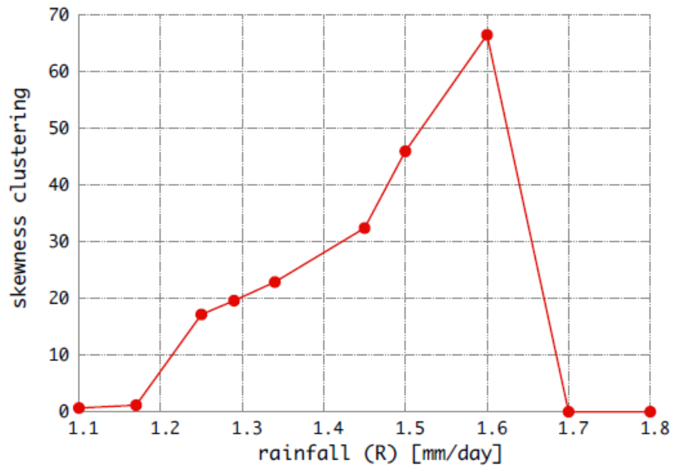

G. Tirabassi et al., Ecological Complexity (2014)

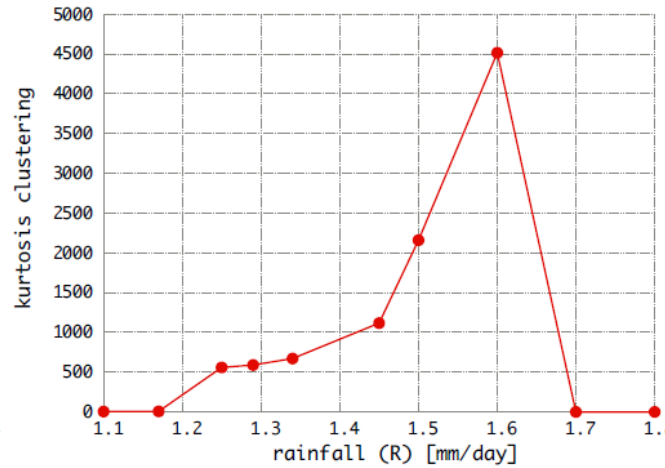
# “Gaussianization” of the distributions of $a_i$ & $c_i$ as the tipping point is approached

skewness

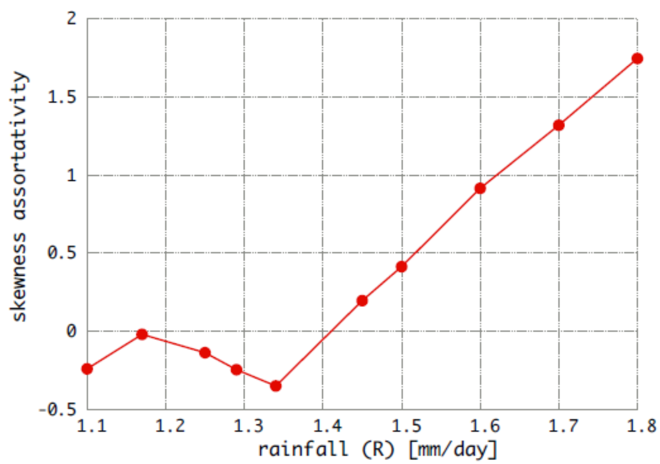
kurtosis

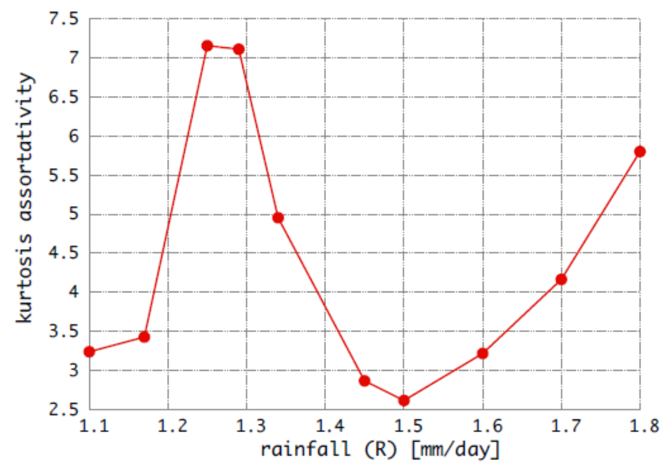
clustering



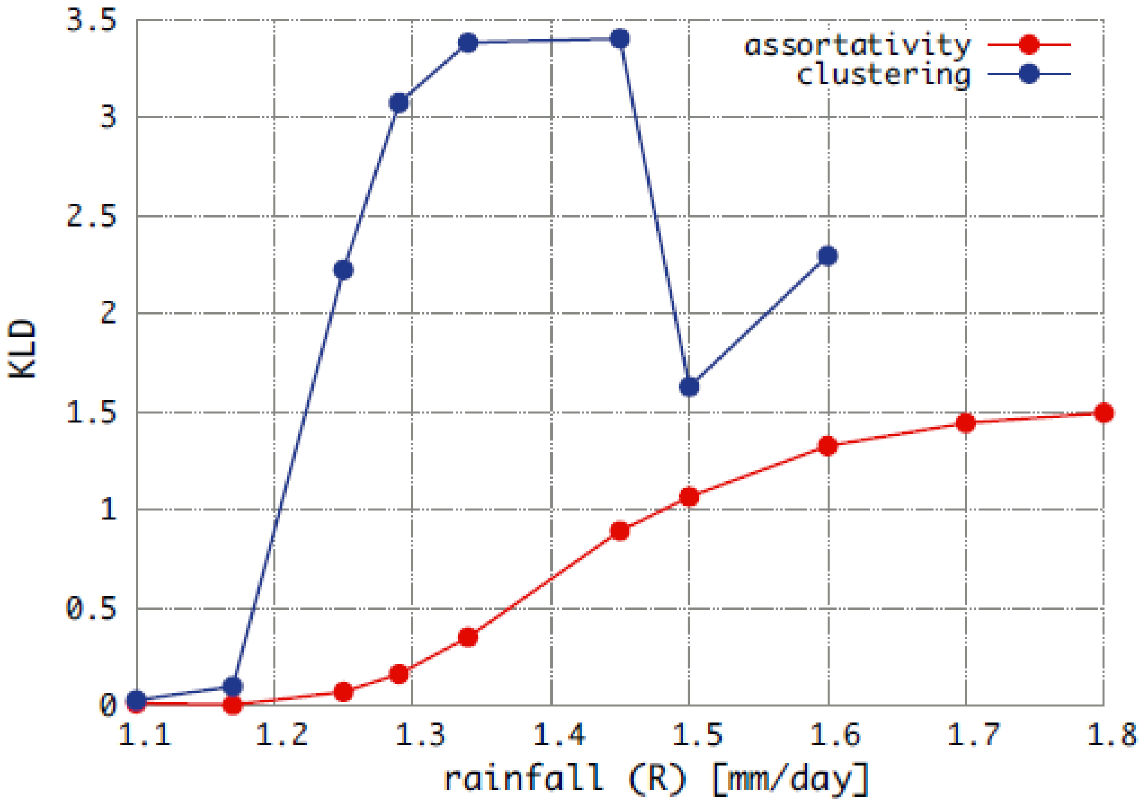




assortativity



**The "Gaussianisation" is quantified by the Kullback distance to a Gaussian (Z) distribution**




$$\mathrm{KLD} \equiv \int_{-\infty}^{\infty} \ln\left(\frac{P(x)}{Z(x)}\right) P(x)\,\mathrm{d}x.$$

- Open issue: the "Gaussianisation" might be a model-specific feature.
- How to precisely quantify changes of the network?
- We need a distance to compare graphs.

G. Tirabassi et al., Ecological Complexity 19, 148 (2014)

# How to “infer” interactions from observed data?

# A classification problem

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

- 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|: $$CC_{ij}(\tau) = \frac{1}{T-\tau_{max}} \left| \sum_{t=0}^{T-\tau_{max}} a_i(t) a_j(t+\tau) \right|$$

Statistical Similarity Measure:

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

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

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

# Kuramoto oscillators in a random network

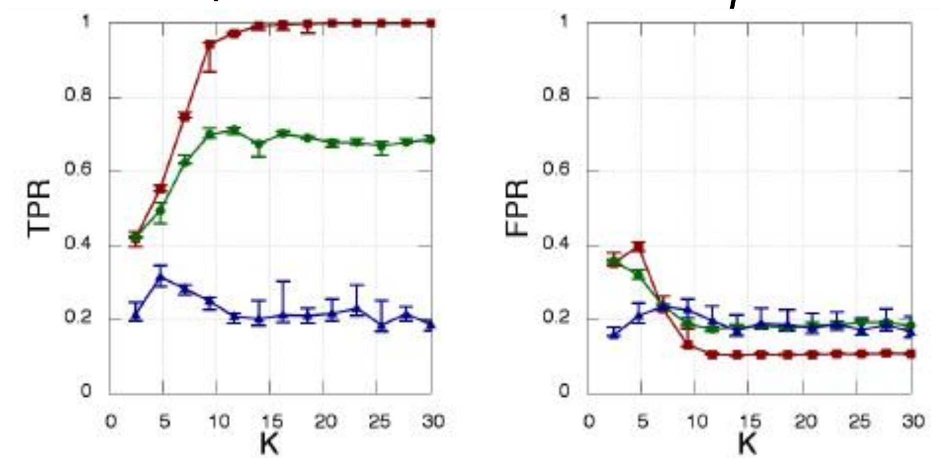
$$d\theta_i = \omega_i \mathrm{d}t + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_j - \theta_i)\mathrm{d}t + D\ \mathrm{d}W_t^i$$

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

**CC MI MIOP**

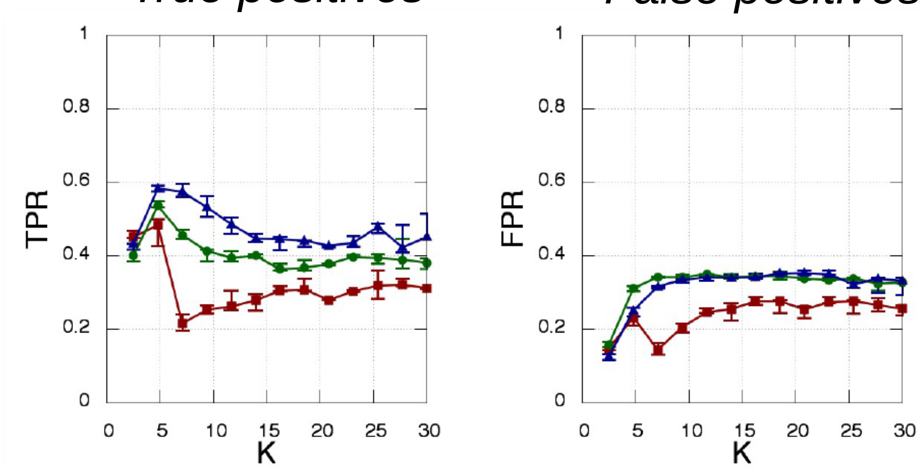
## Phases (θ)

*True positives* *False positives*



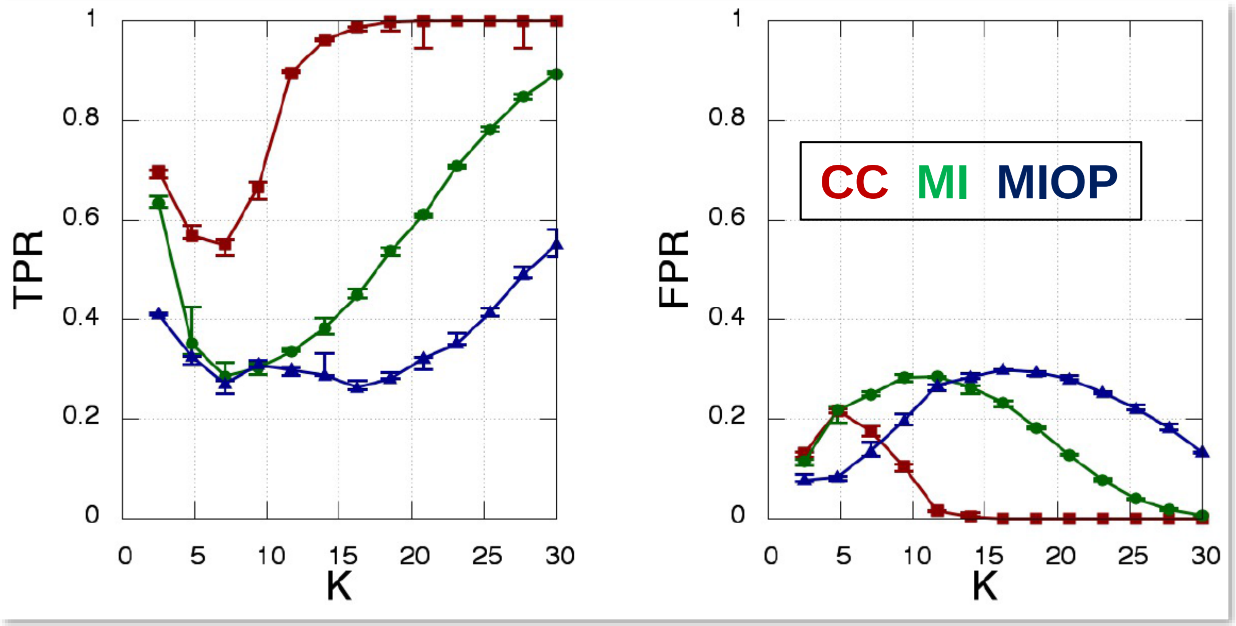

## "Observable" Y=sin(θ)

*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.

# Instantaneous frequencies (dθ/dt)



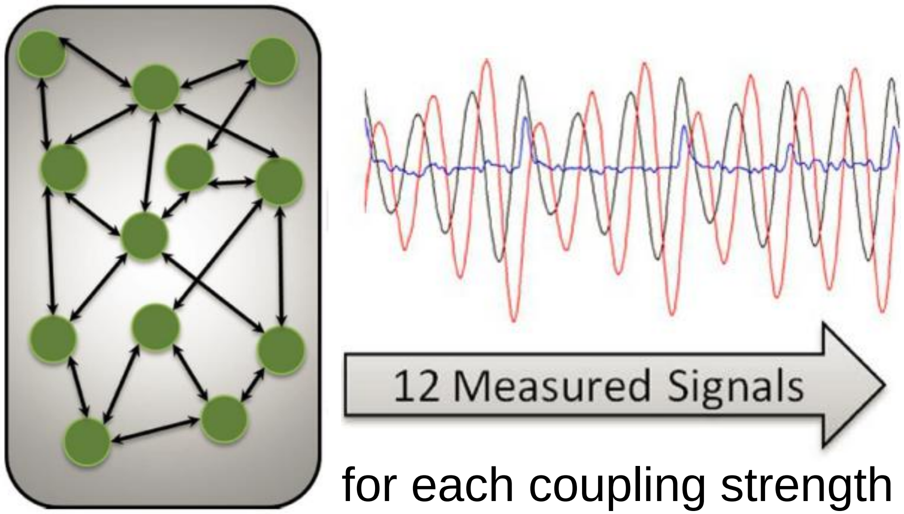

Perfect network inference is possible!

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

$$\varphi_i = HT(x_i)$$

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

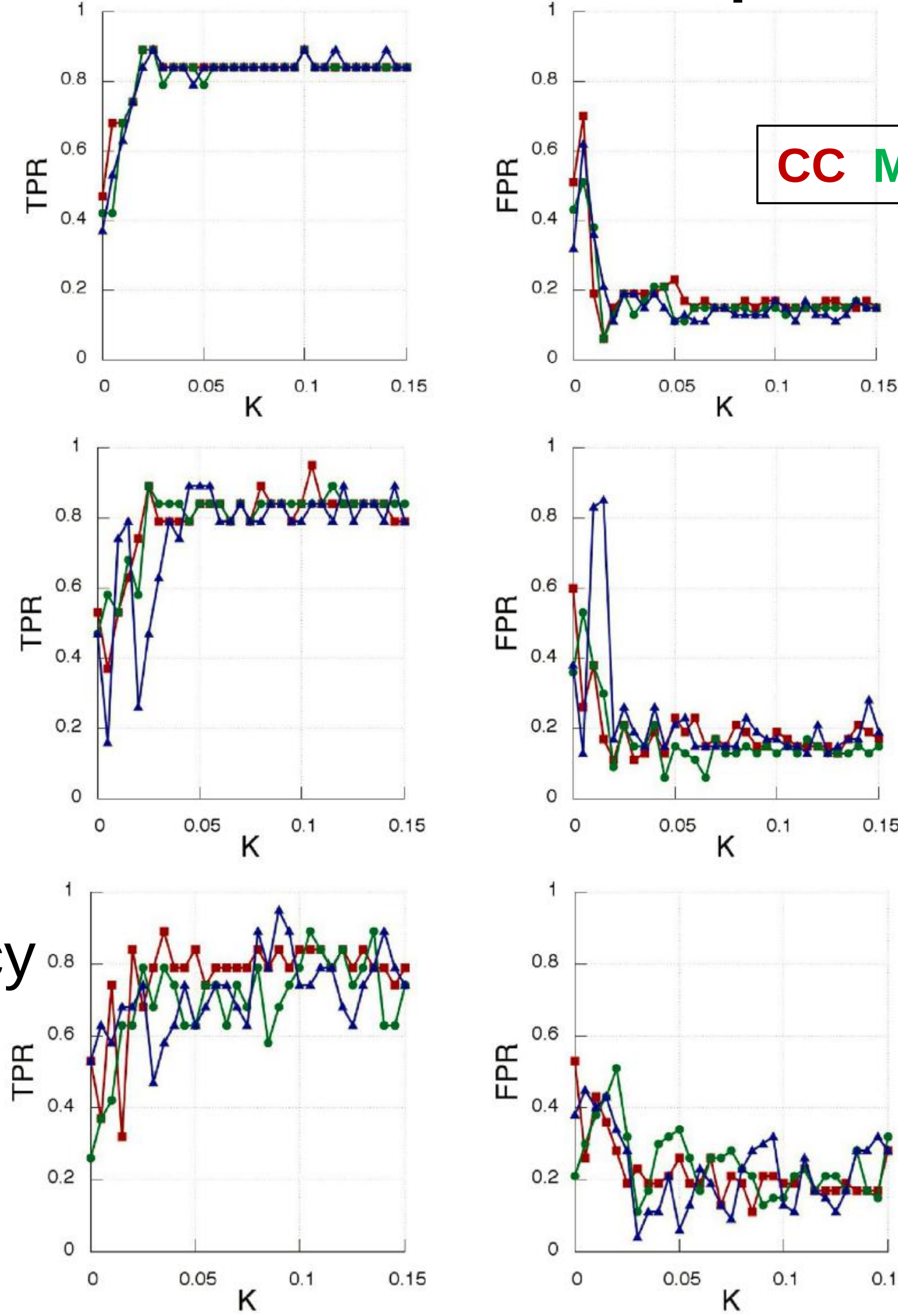
$$x_i$$

# Results obtained with experimental data

Observed variable (x)

Hilbert phase

Hilbert frequency



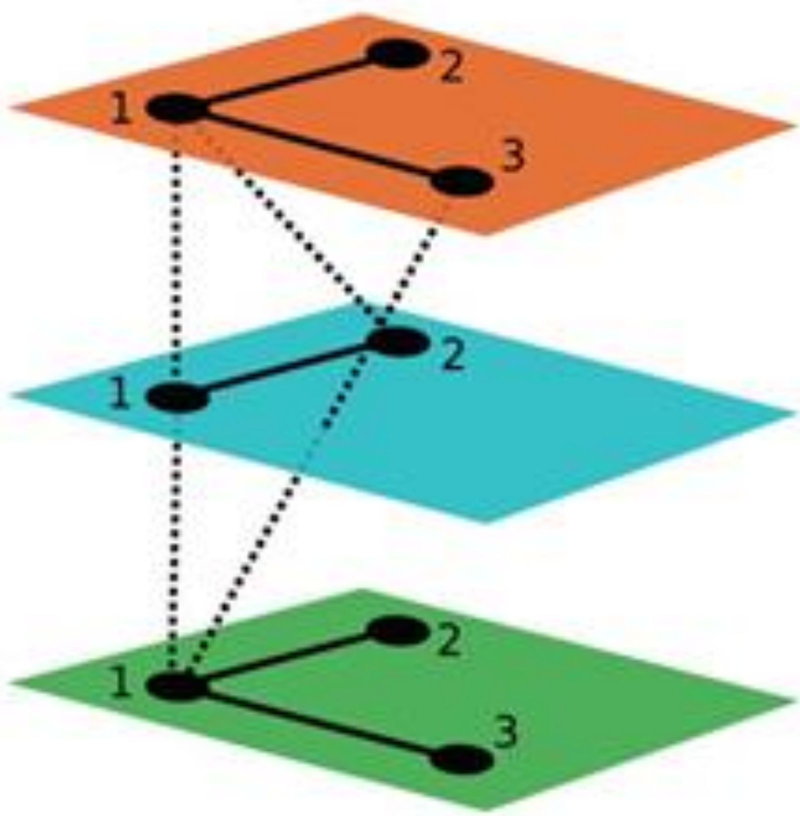

CC MI MIOP

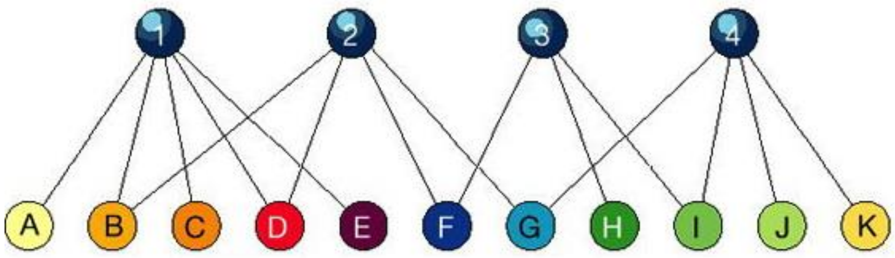
– No perfect reconstruction

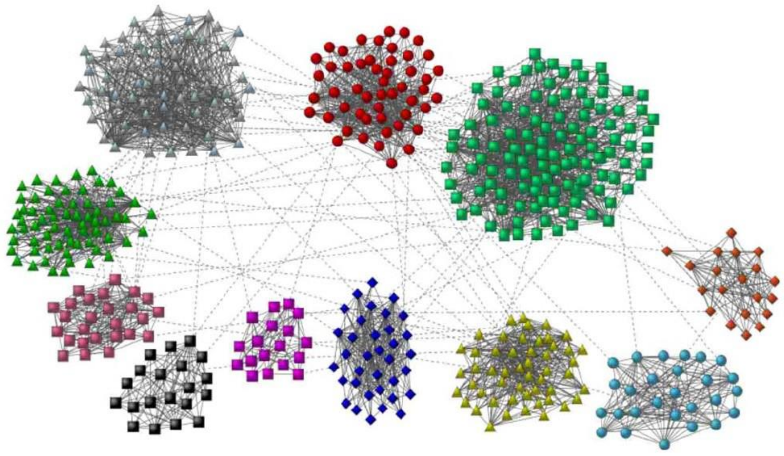
– No important difference among the 3 methods & 3 variables

# Generalizations of complex network analysis

# Network structures: Multilayer, multiplex, bipartite, networks of networks and many others



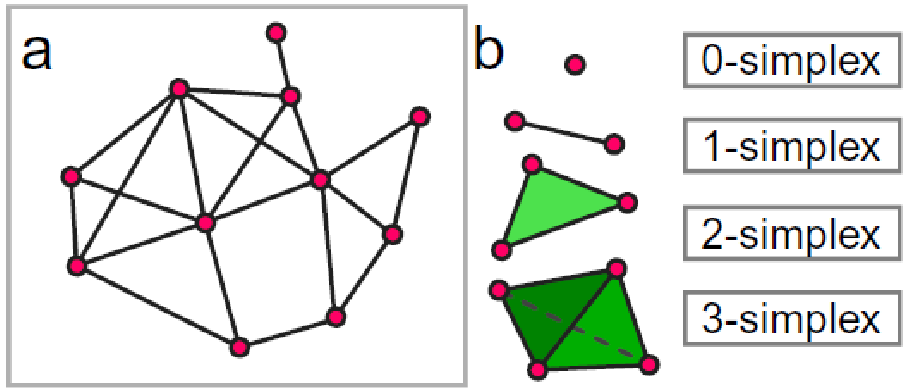



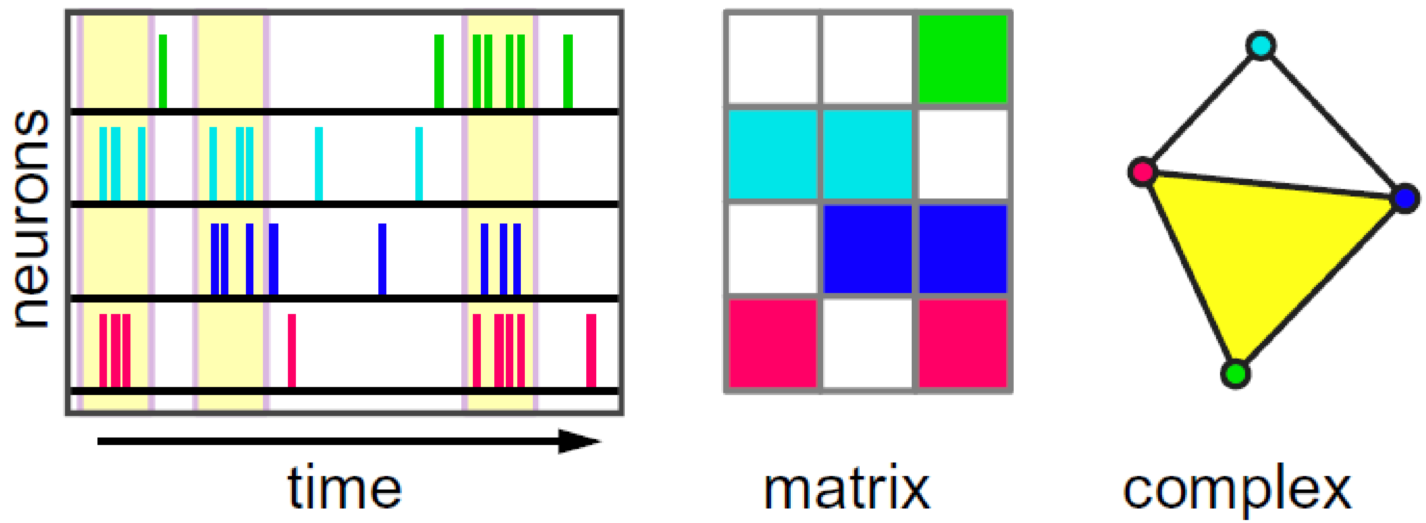
# Limitations of complex network analysis

# Interactions are not limited to pairs of elements

- Links represent interactions between pairs of nodes.
- ***Simplicial complexes*** represent interactions among several nodes.




Example






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

## Summary

- Multivariate analysis uncovers inter-relationships in datasets
- Different similarity measures are available for inferring the connectivity of a complex system from observations.
- Different measures can uncover different properties.
- Thresholding, hidden variables, hidden “nodes” can difficult or make impossible the inference of the network structure.
- Different sets of “communities” can be uncovered depending on the property that is analyzed.
- Network science: many applications and challenges!

<cristina.masoller@upc.edu>

Universitat Politecnica de Catalunya

**http://www.fisica.edu.uy/~cris/**

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH