

Nonlinear time series analysis

Multivariate analysis

Cristina Masoller

Universitat Politècnica de Catalunya, Terrassa, Barcelona, Spain

Cristina.masoller@upc.edu

www.fisica.edu.uy/~cris



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

■ Introduction

- Historical developments: from dynamical systems to complex systems

■ Univariate analysis

- Methods to extract information from a time series.
- Applications to climate data.

■ Bivariate analysis

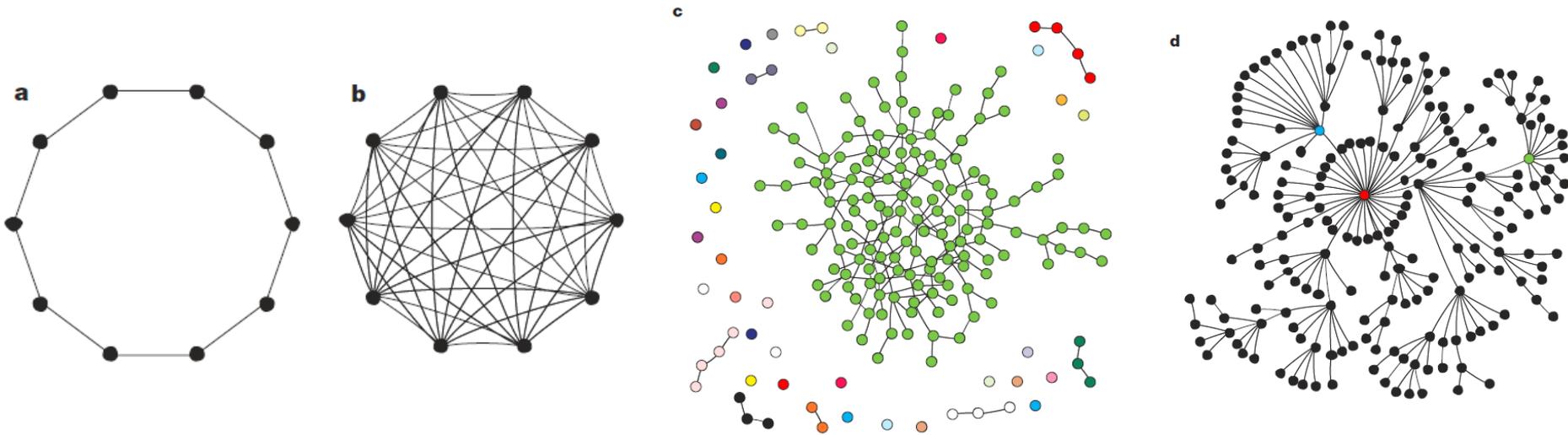
- Extracting information from two time series.
- Correlation, directionality and causality.
- Applications to climate data.

■ Multivariate analysis

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

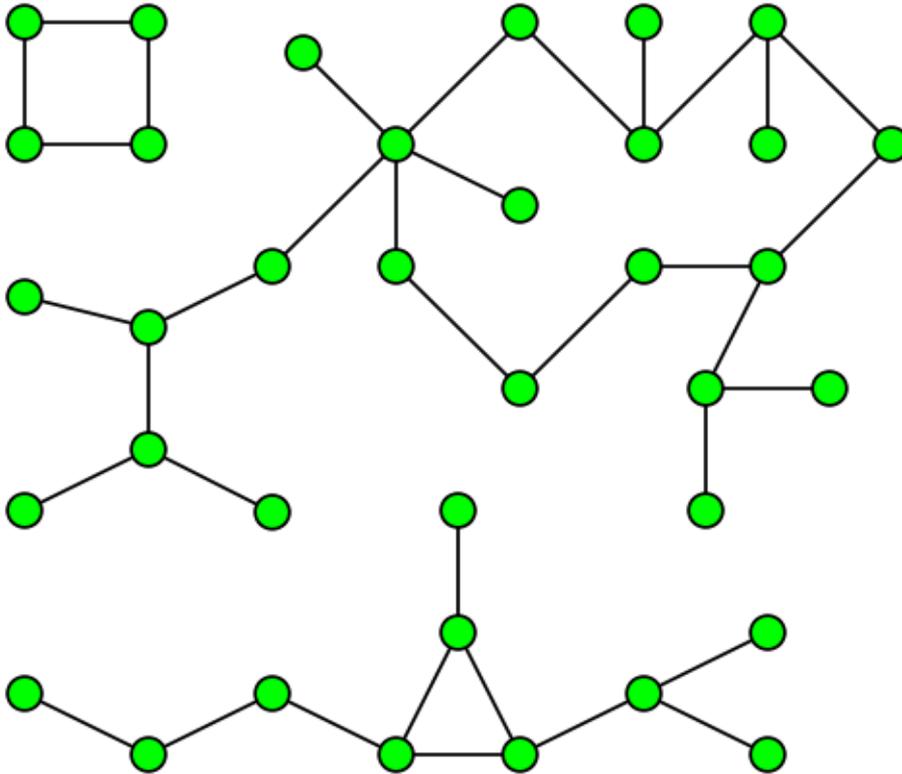
Networks or graphs

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



Source: Strogatz
Nature 2001

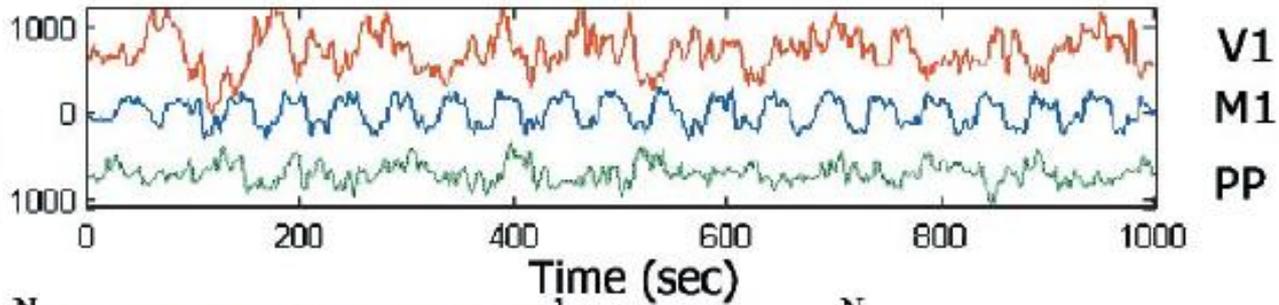
Connected components ("communities")



A graph with three connected components.
Source: Wikipedia

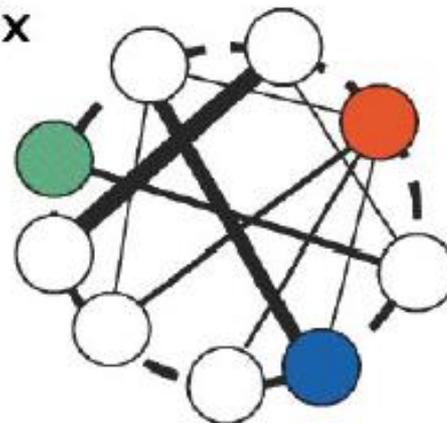
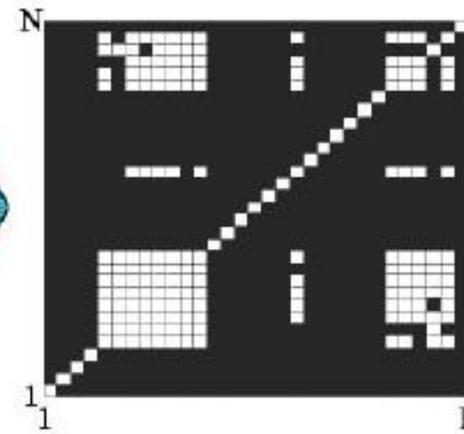
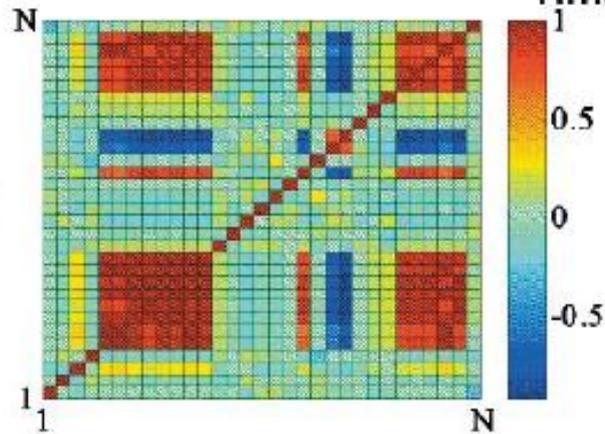
**Using statistical similarity
measures to infer interactions
from data: “functional networks”**

Brain functional network



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

Adjacency
matrix



Eguiluz et al, PRL 2005
Chavez et al, PRE 2008

Graphical representation

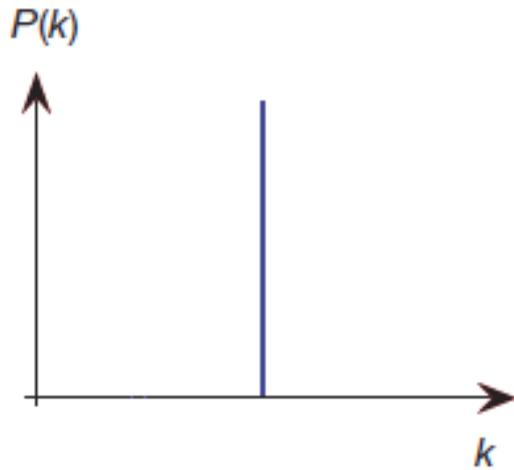
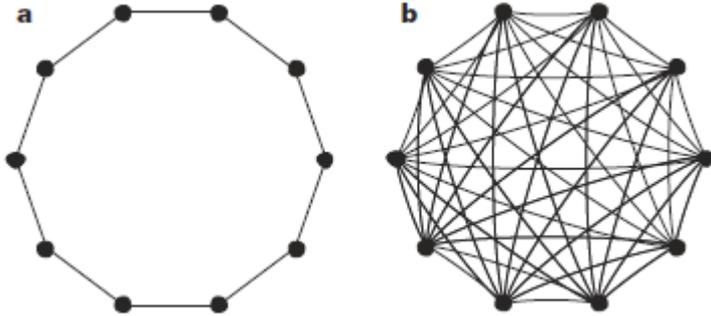


Thresholded matrix = inferred (“functional”) adjacency matrix

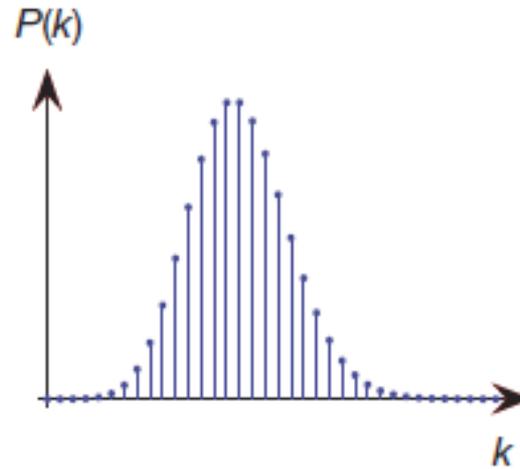
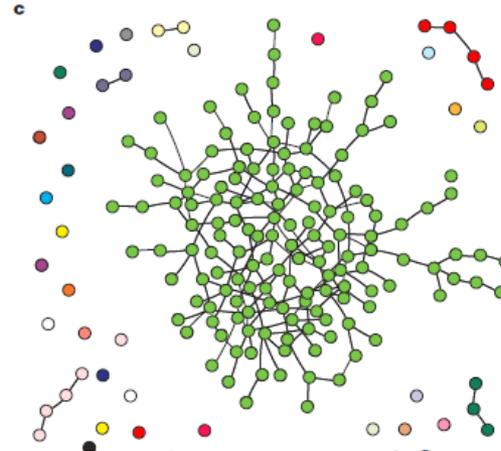
Degree of a node: number of links, $k_i = \sum_j A_{ij}$

The degree distribution: usual way to characterize a graph

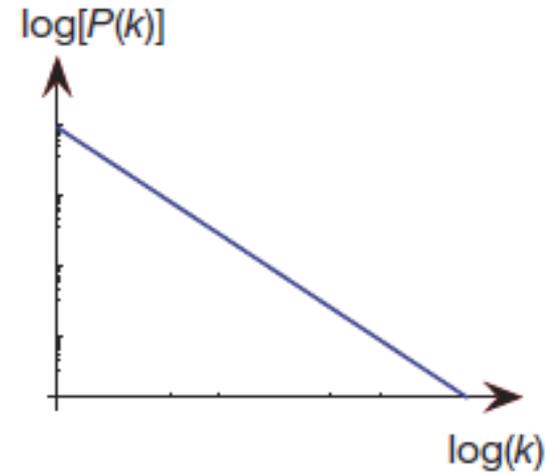
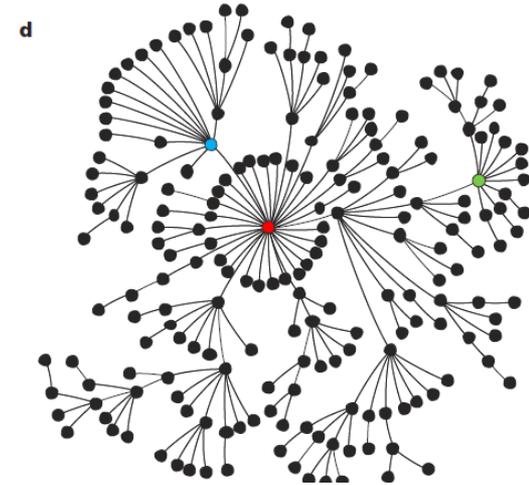
Regular



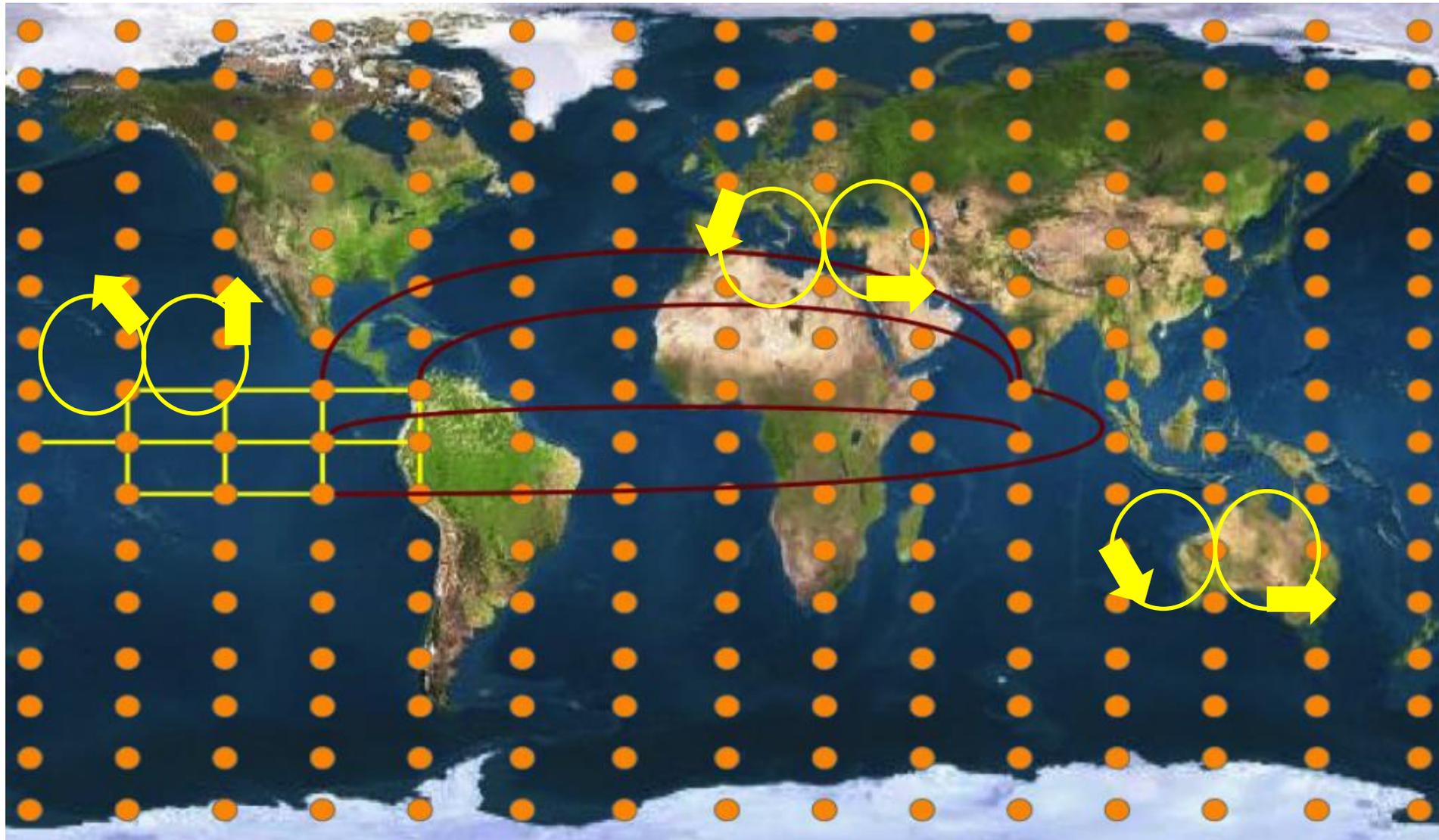
Random



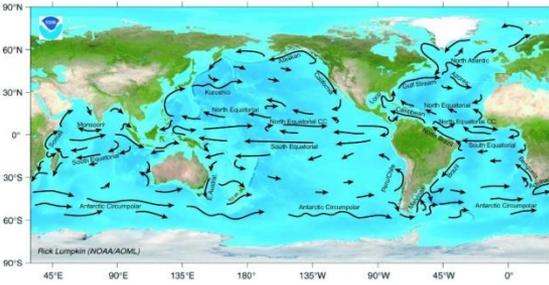
Scale-free



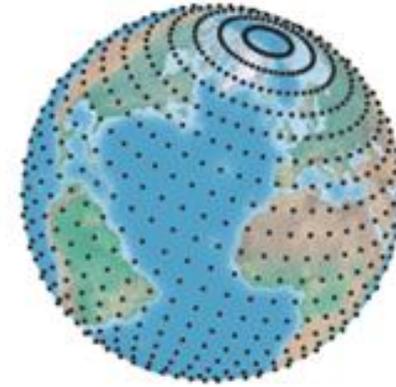
The climate system as a set of “interacting oscillators”



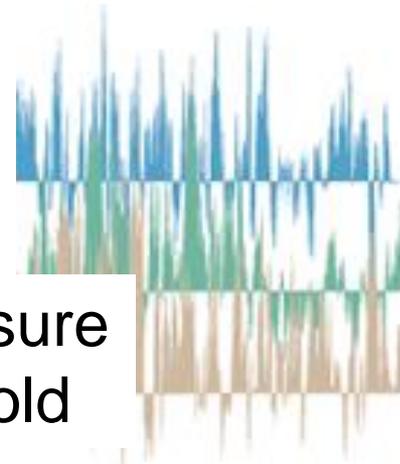
Complex network representation of the climate system



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



More than 10000 nodes (with different sizes).



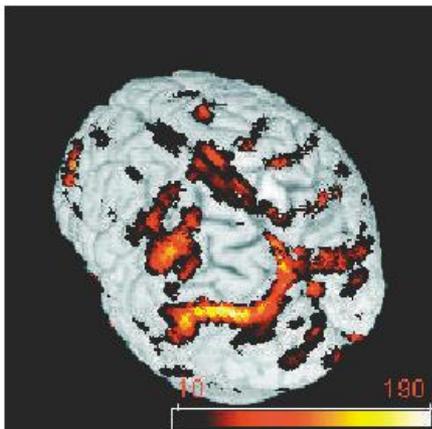
Daily resolution: more than 13000 data points in each TS



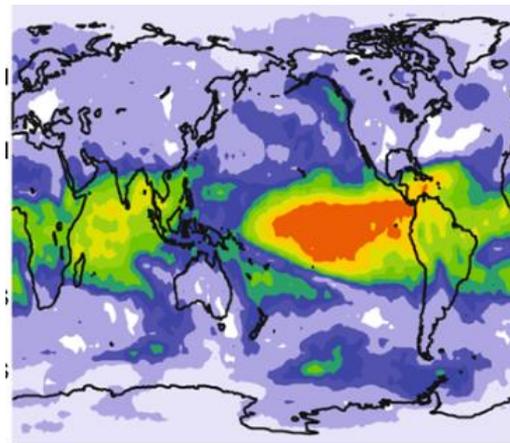
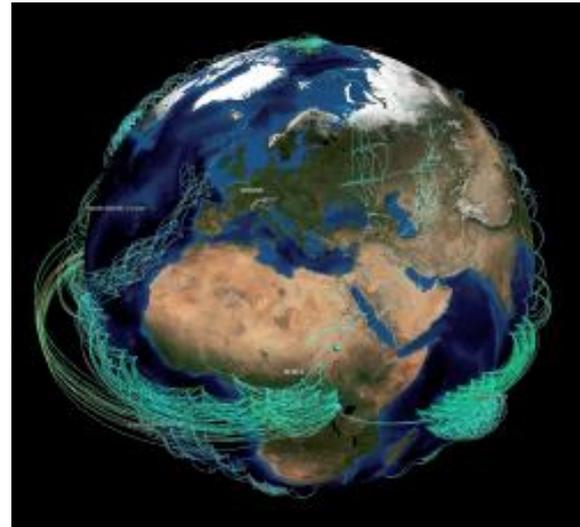
Sim. measure + threshold

Surface Air Temperature Anomalies (solar cycle removed)

Brain network



Climate network



Area weighted connectivity (AWC): is the weighted degree (nodes are areas with different sizes)

$$AWC_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

Thresholding

- Is the statistical similarity measure (S_{ij}) between the time series in two nodes (i,j) is “significant”, the nodes are linked, otherwise, they are not.

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

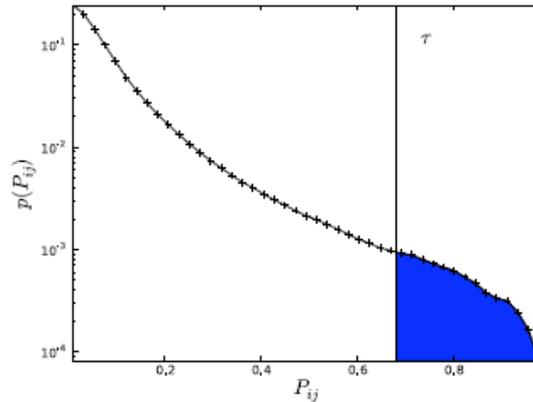
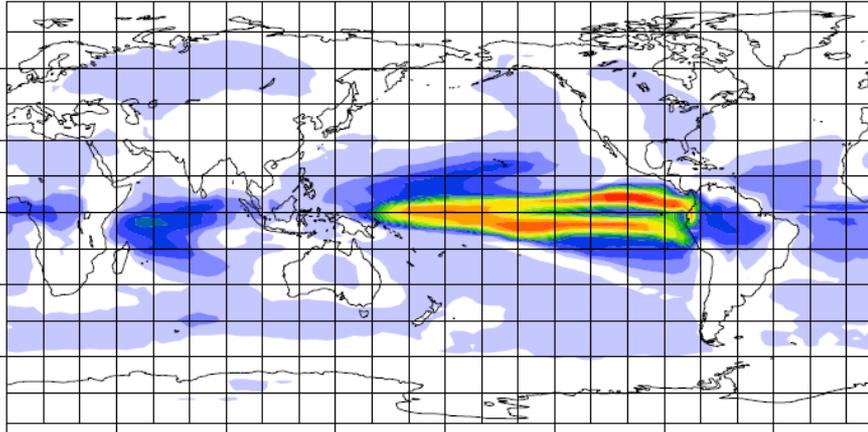
How to select the threshold?

Three criteria are typically used:

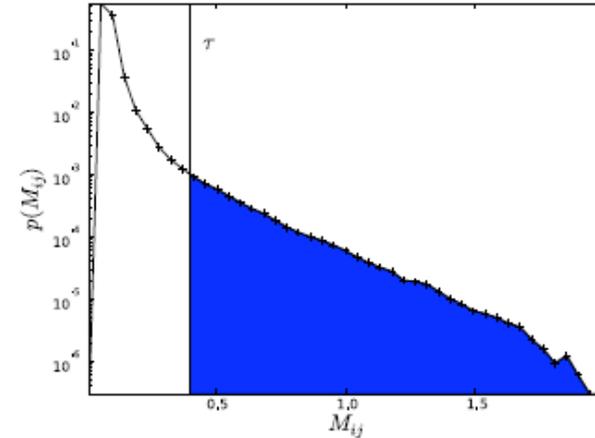
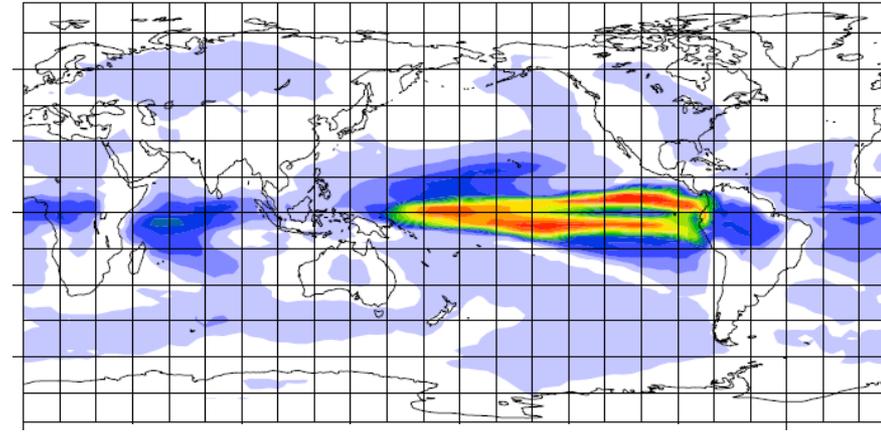
- A significance level is used (typically 5%) in order to omit connectivity values that can be expected by chance;
- We select an arbitrary value as threshold, such that it gives a certain pre-fixed number of links (or link density);
- We define the threshold as large as possible while guaranteeing that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

Comparison of area weighted connectivity with $|CC|$ and MI

AWC computed with $|$ cross-correlation $|$



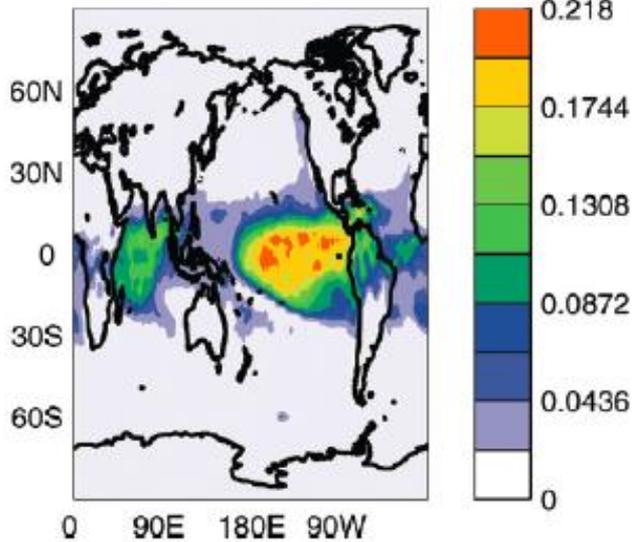
AWC computed with mutual information



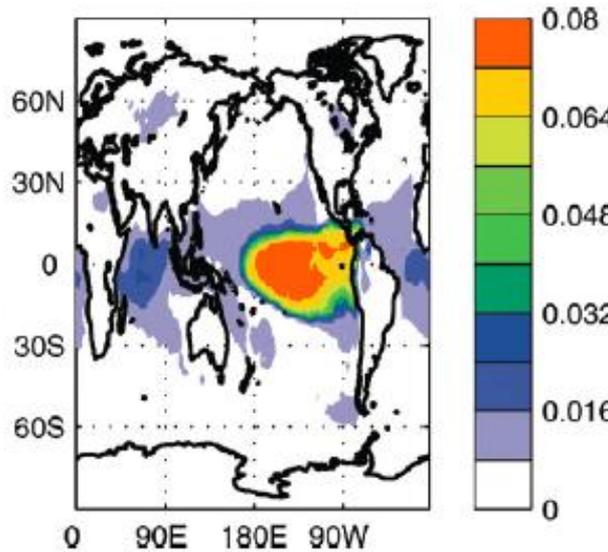
The threshold was selected to give a network with the same link density (0.005)

Influence of the threshold

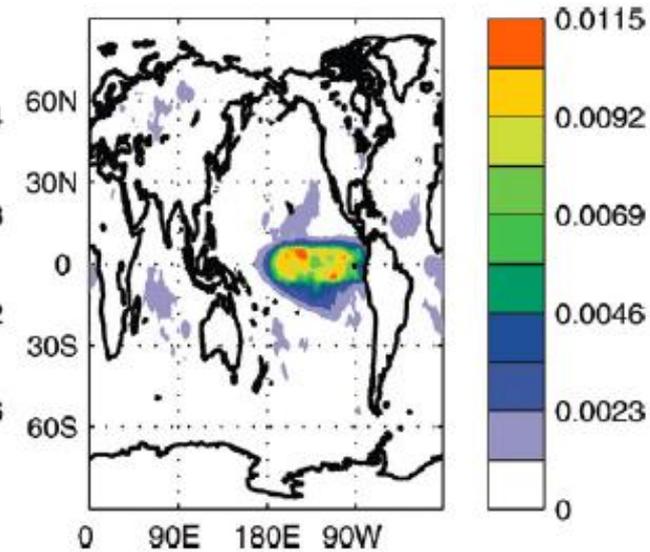
$\rho=0.027$



$\rho=0.01$

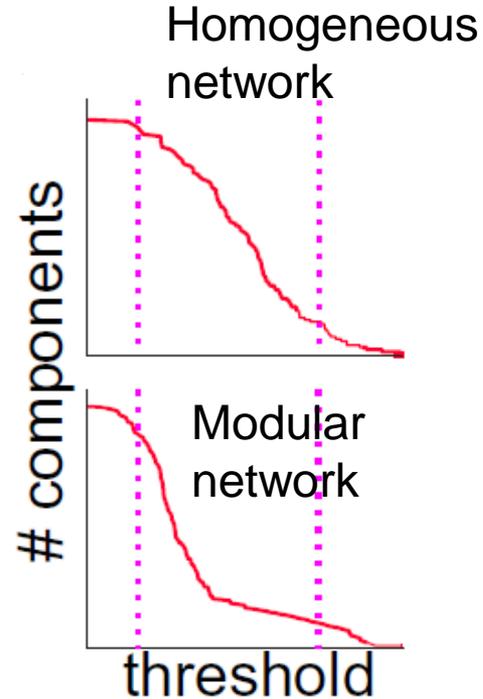
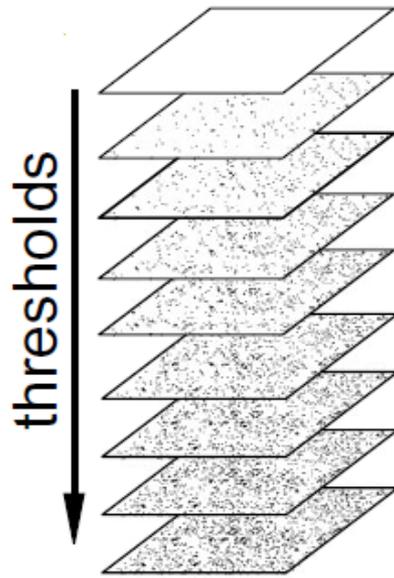
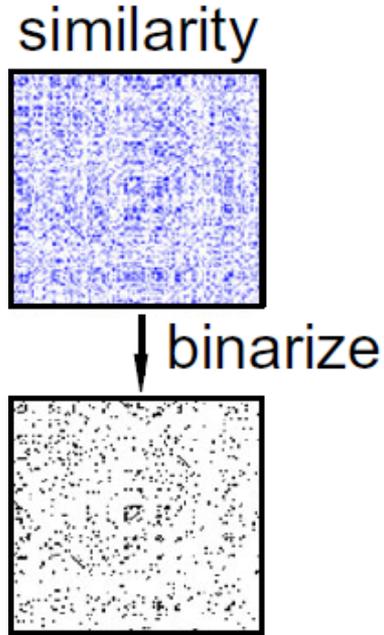


$\rho=0.001$



[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 would suggest inaccurately that these two networks have similar structures.
- “Features” that persist for a wide range of thresholds are “true” features.

Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package

Jonathan F. Donges^{*}, Jobst Heitzig, Boyan Beronov, Marc Wiedermann, Jakob Runge, Qing Yi Feng, Liubov Tupikina, Veronika Stolbova, Reik V. Donner, Norbert Marwan, Henk A. Dijkstra, and Jürgen Kurths

Citation: *Chaos* **25**, 113101 (2015); doi: 10.1063/1.4934554

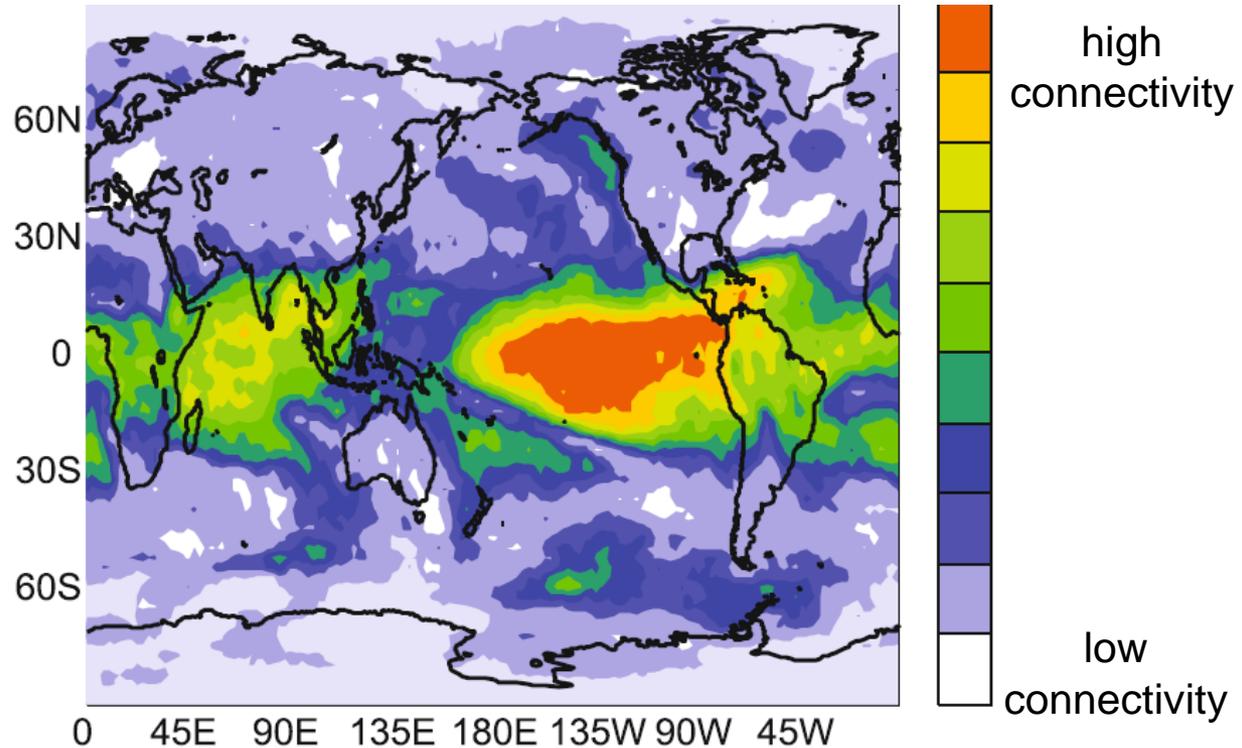
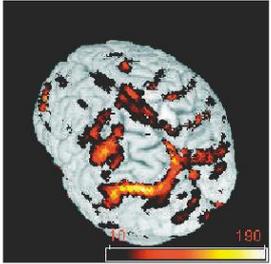
View online: <http://dx.doi.org/10.1063/1.4934554>

pyunicorn is available at <https://github.com/pik-copan/>

Climate network with mutual information computed with probabilities of ordinal patterns

$$AWC_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}$$

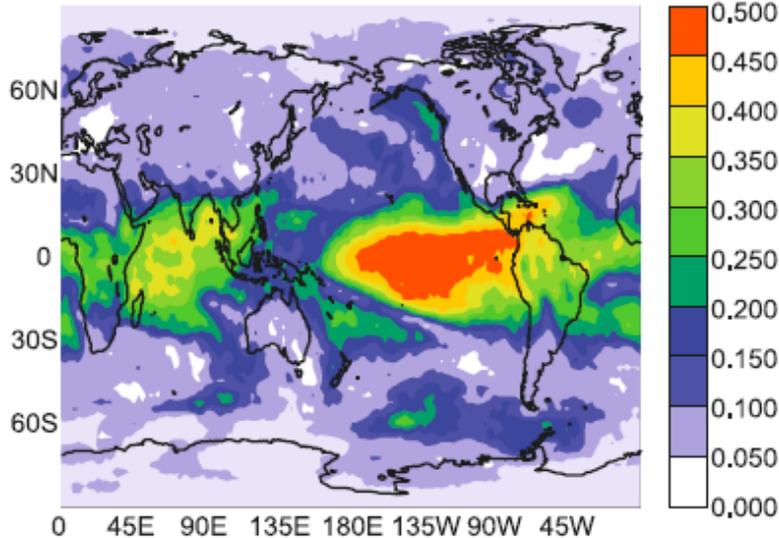
inter-annual time-scale (3 consecutive years). The color-code indicates the Area Weighted Connectivity (weighted degree)



Comparison: ordinal probabilities vs. histogram of data values

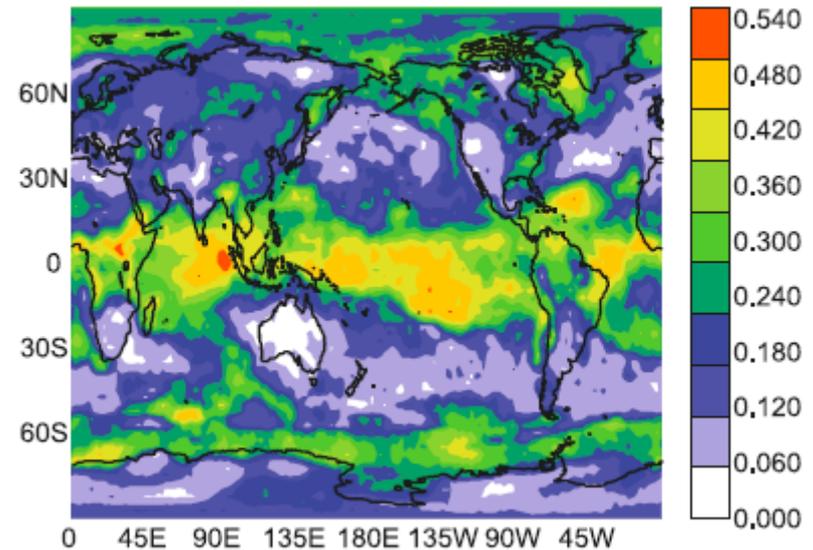
$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}$$

Network when the probabilities are computed with ordinal analysis

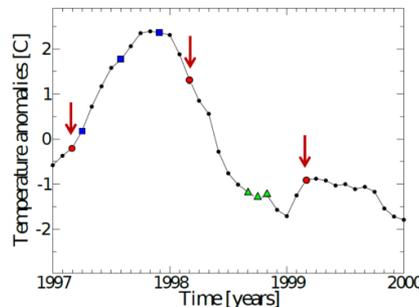


Color code indicates the area-weighted connectivity

Network when the probabilities are computed with histogram of values

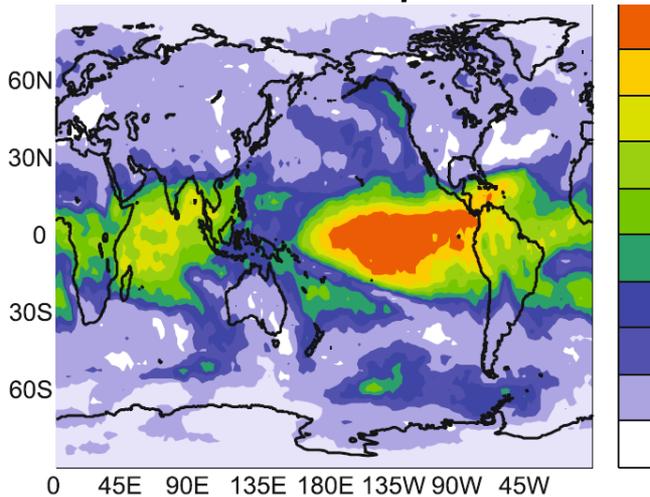


inter-annual time scale

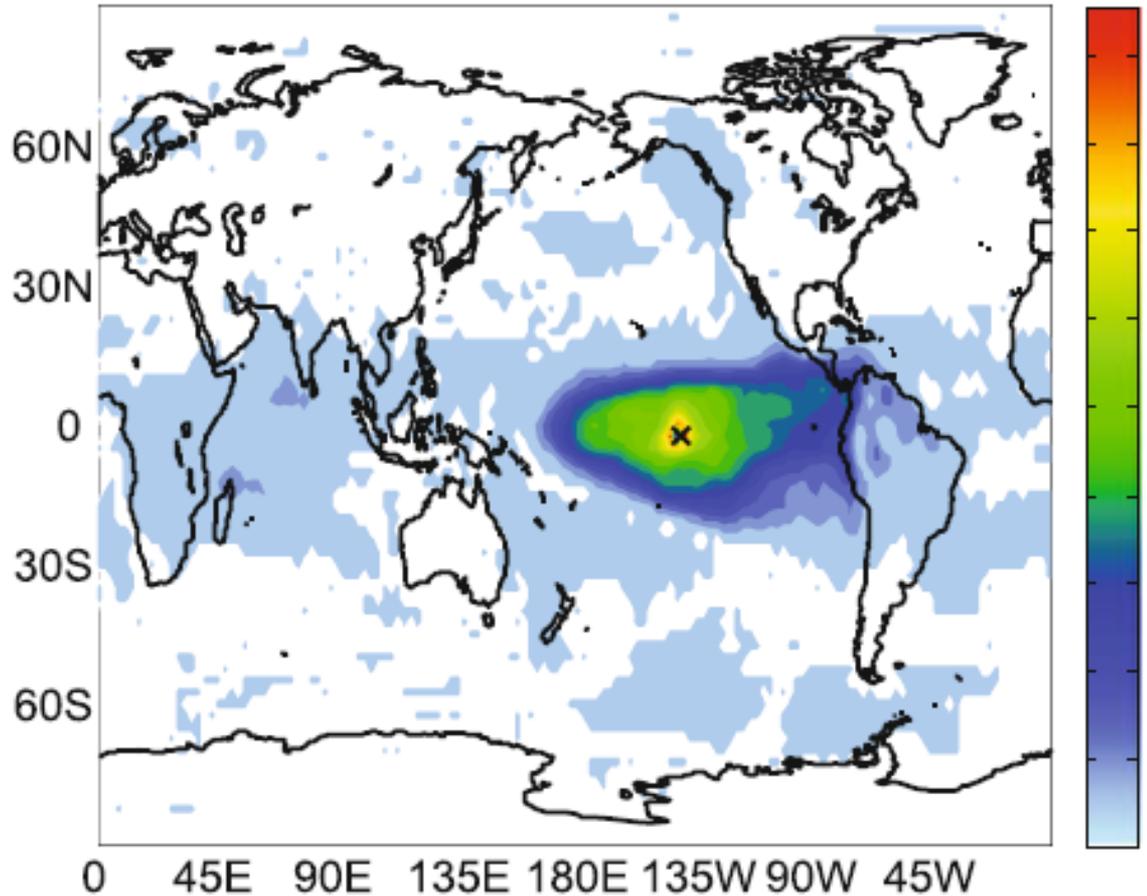


Who is connected to who?

AWC map



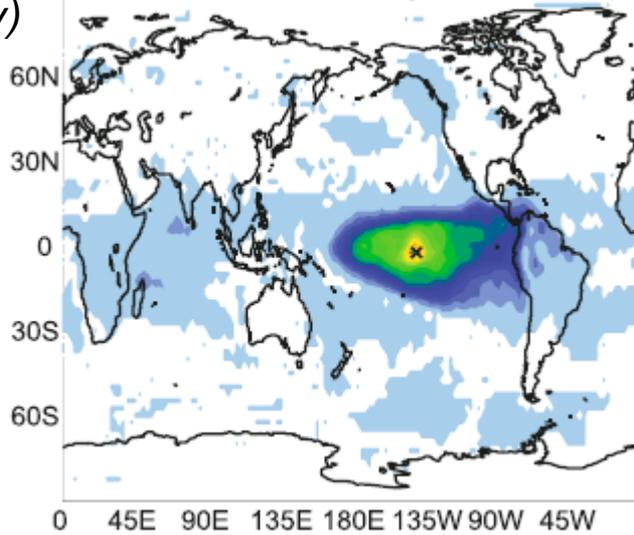
color-code indicates the MI values (only significant values)



Influence of the threshold

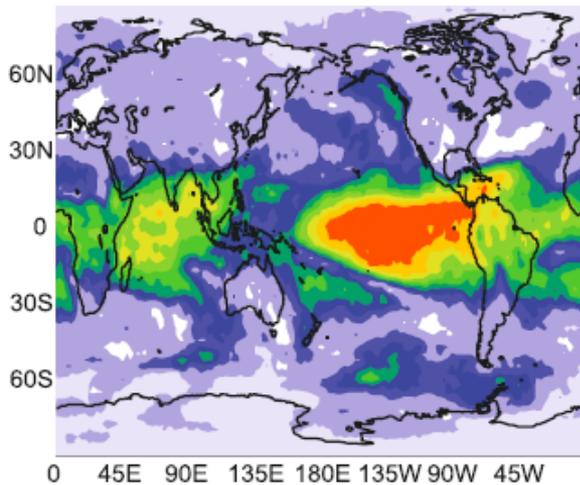
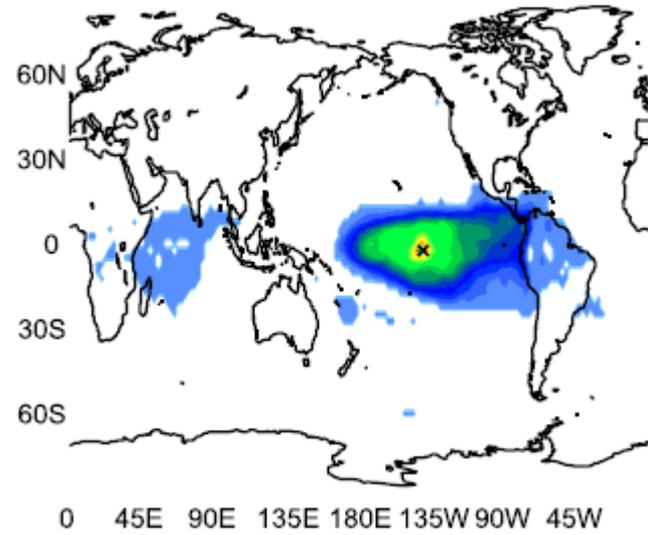
All significant links

(11% link density)

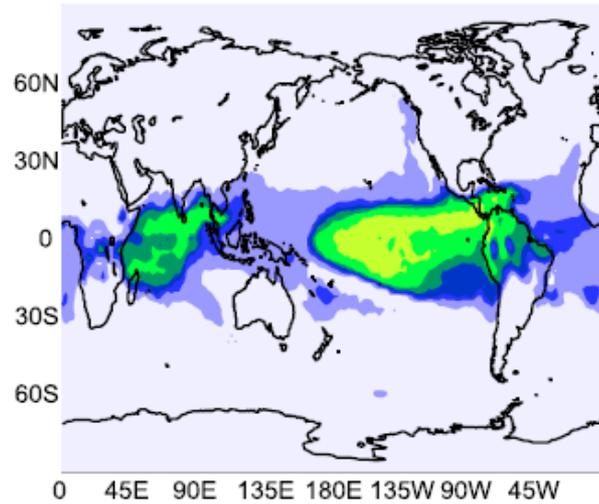


Color code:
MI

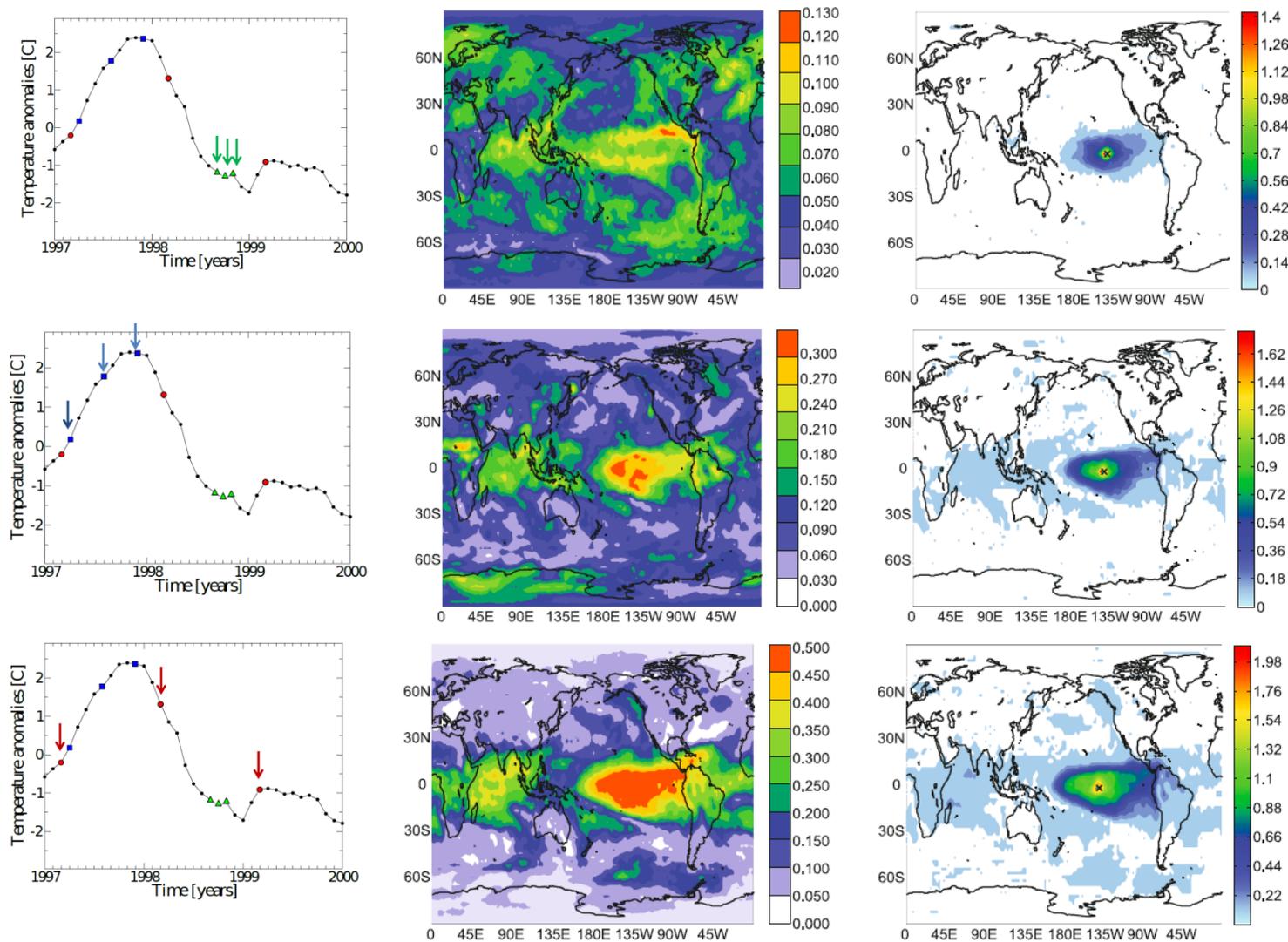
Higher threshold (3% link density)



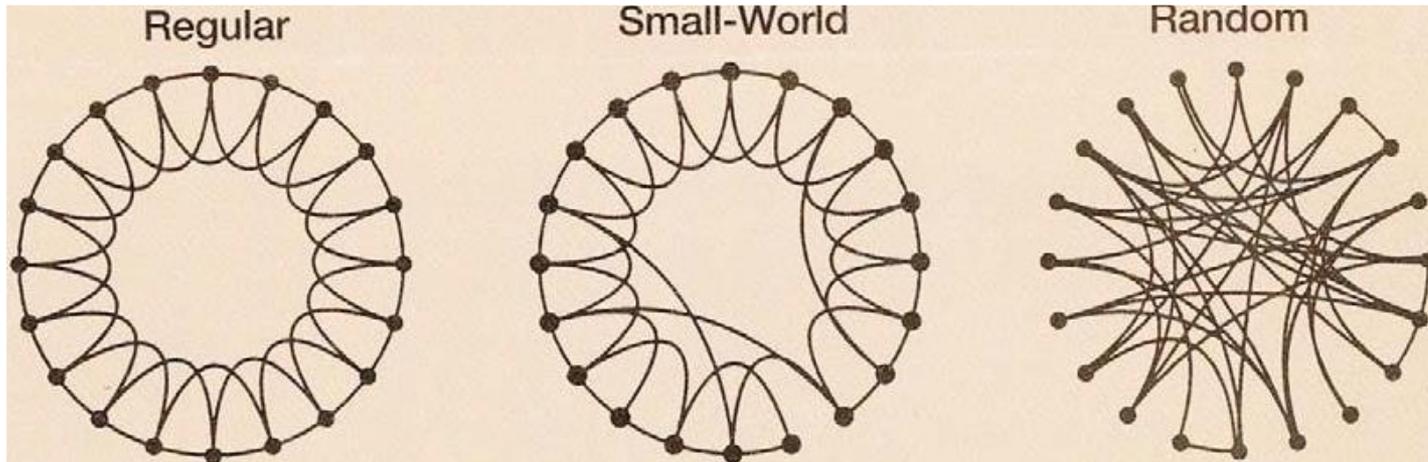
Color code:
AWC
[Video](#)



Influence of the time-scale of the pattern



Longer time-scale \Rightarrow increased connectivity



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 $k_i = \sum_j A_{ij}$

- **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): $\mu = E[X] = \sum_{i=1}^k x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_k p_k$

- **Variance**: $\sigma^2 = \text{Var}(X) = E[(X - \mu)^2]$

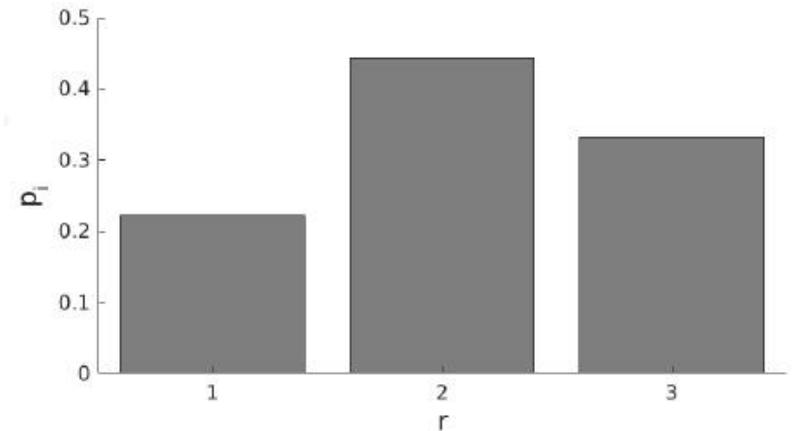
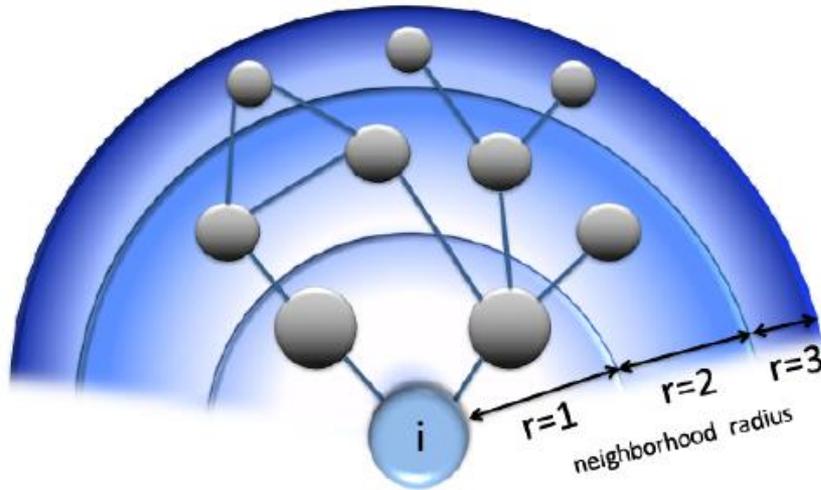
- **Skewness**: “measures” the asymmetry of the distribution

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

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

$$K = E[Z^4]$$

Diameter: longest shortest path



Node Distance Distribution (NDD) of node i : fraction of nodes that are connected to node i at distance j .

How to compare two distributions (degree, NDD, etc.)?

Distance between two distributions P and P_e

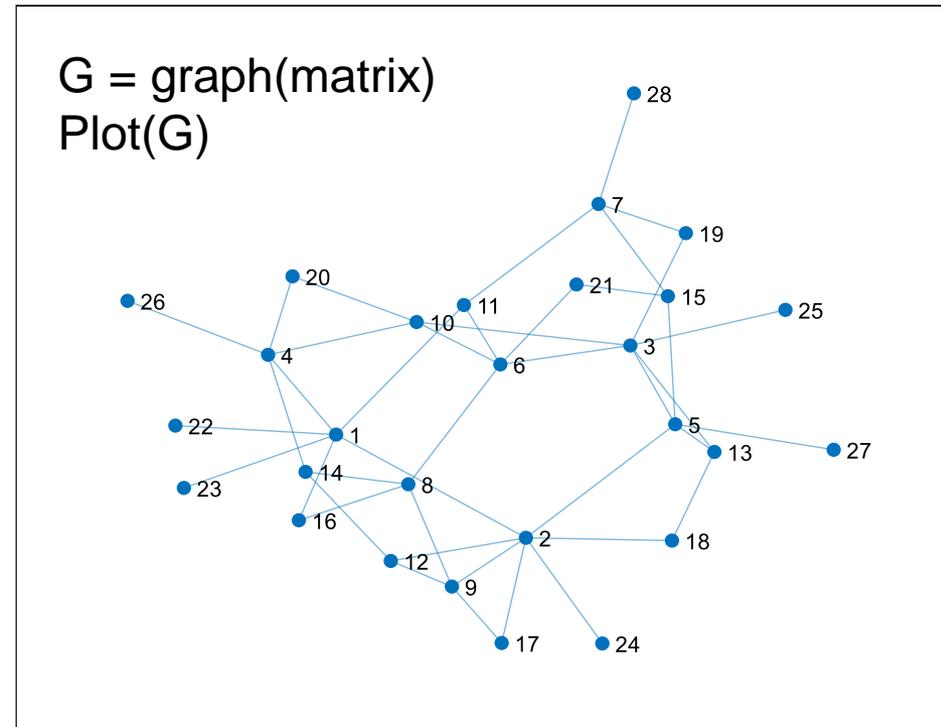
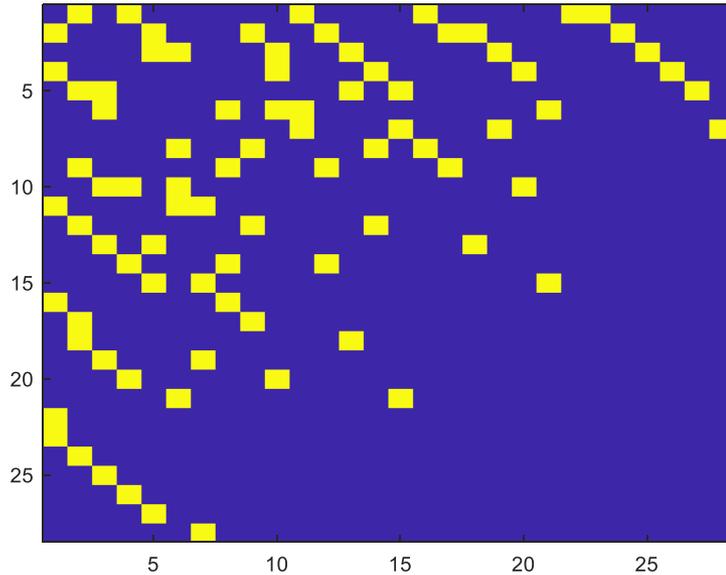
Euclidean $D_E[P, P_e] = \|P - P_e\|_E = \sum_i (p_i - p_{i,e})^2$

Kullback $D_K[P, P_e] = K[P|P_e] = I[P_e] - I[P]$

Jensen divergence $D_J[P, P_e] = \frac{K[P|P_e] + K[P_e|P]}{2}$

Read more: S-H Cha: *Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions*, Int. J of. Math. Models and Meth. 1, 300 (2007)

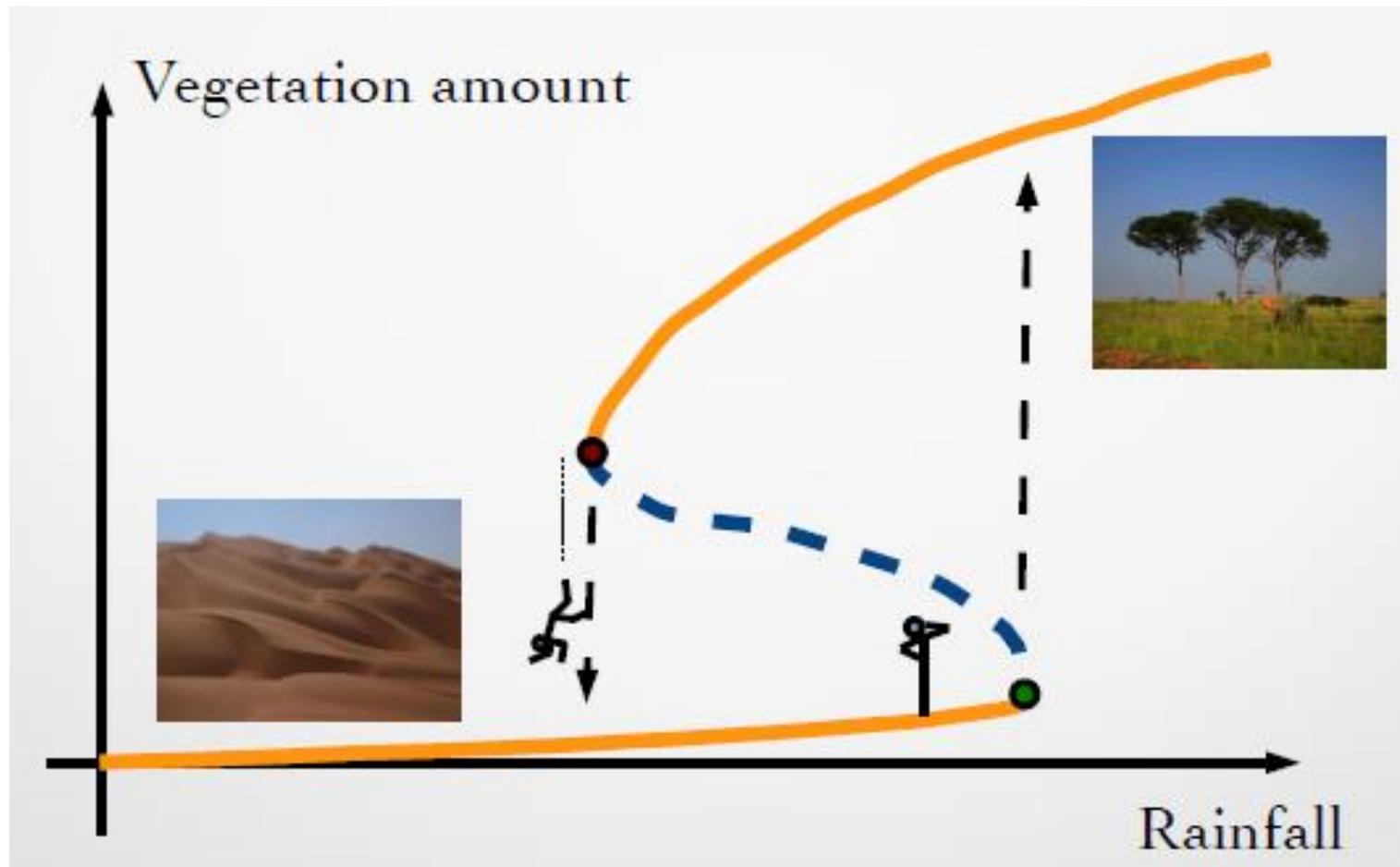
Exercise 1: graphical representation of the “structural” network of the electronic circuits and calculate the degree distribution



Exercise 2: analyze the functional network of the 28 electronic circuits

- Construct the network using the Pearson coefficient as statistical similarity measure.
- Calculate the mean degree for different values of the coupling for fixed TH.
- Calculate the mean degree for different values of the threshold for fixed coupling.

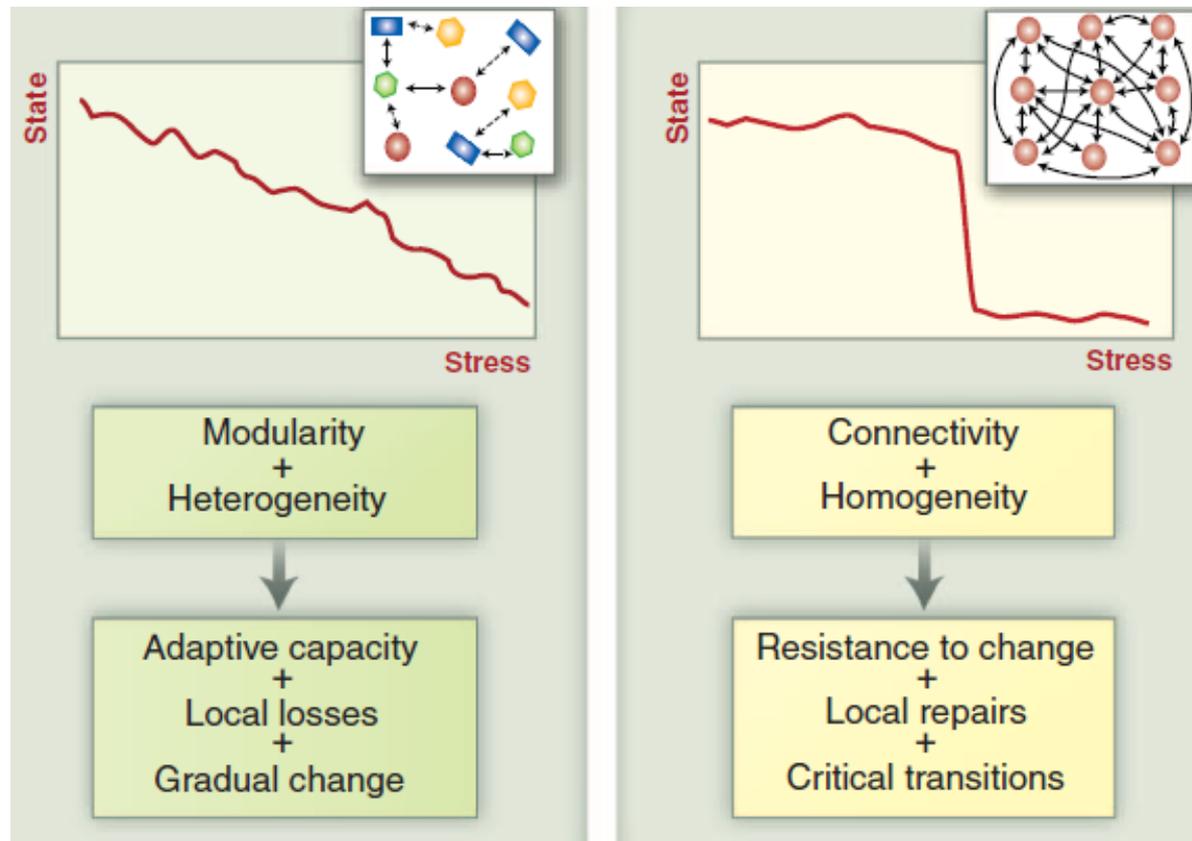
**Example:
desertification transition under
the lens of network analysis**



Our goal: to develop reliable early-warning indicators

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

Role of the network structure



Networks in which the components are heterogeneous and where incomplete connectivity causes modularity tend to gradually adjust to change.

In highly connected networks, local losses tend to be “repaired” by subsidiary inputs from linked units until at a critical stress level the system collapses.

Desertification transition: model

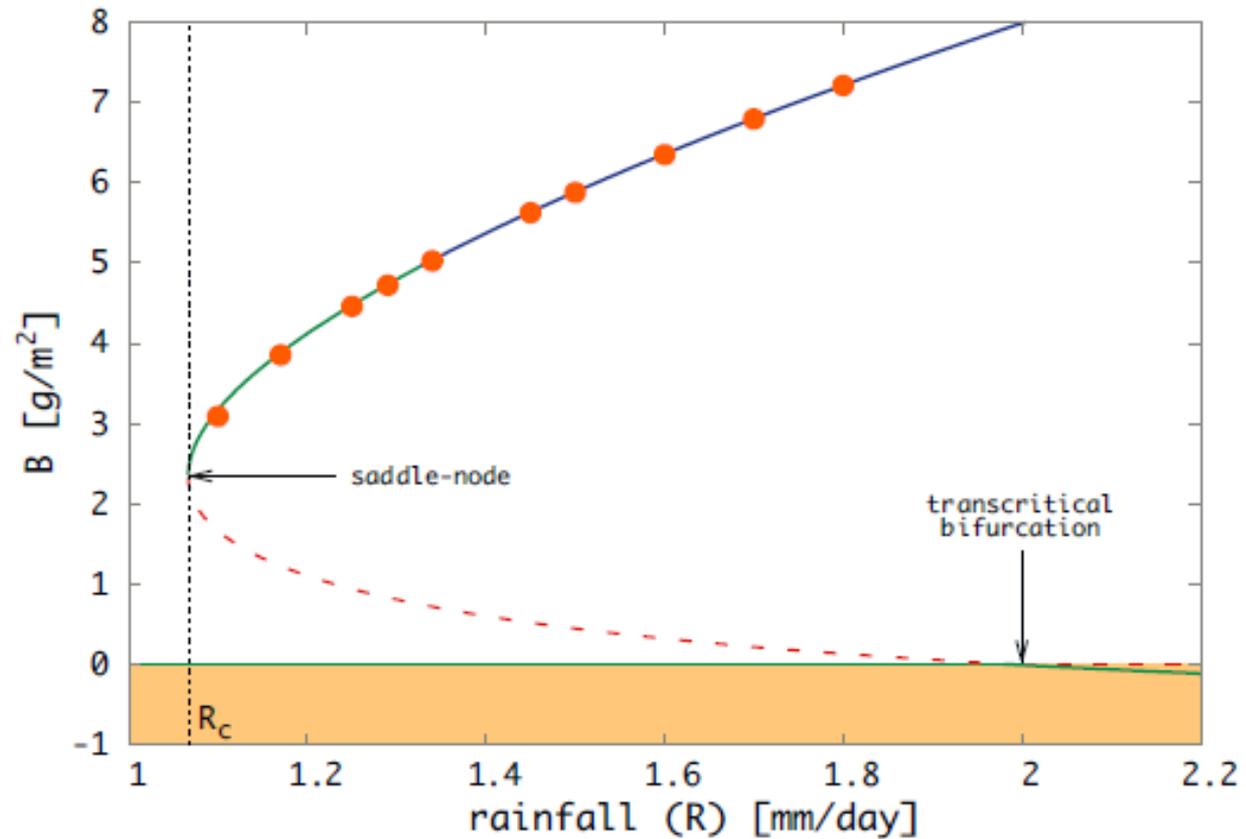
$$\frac{\partial w}{\partial t} = R - \frac{w}{\tau_w} - \Lambda w B + D \nabla^2 w + \sigma_w w_0 \xi^w(t),$$

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

- w (in mm) is the soil water amount
- B (in g/m²) 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

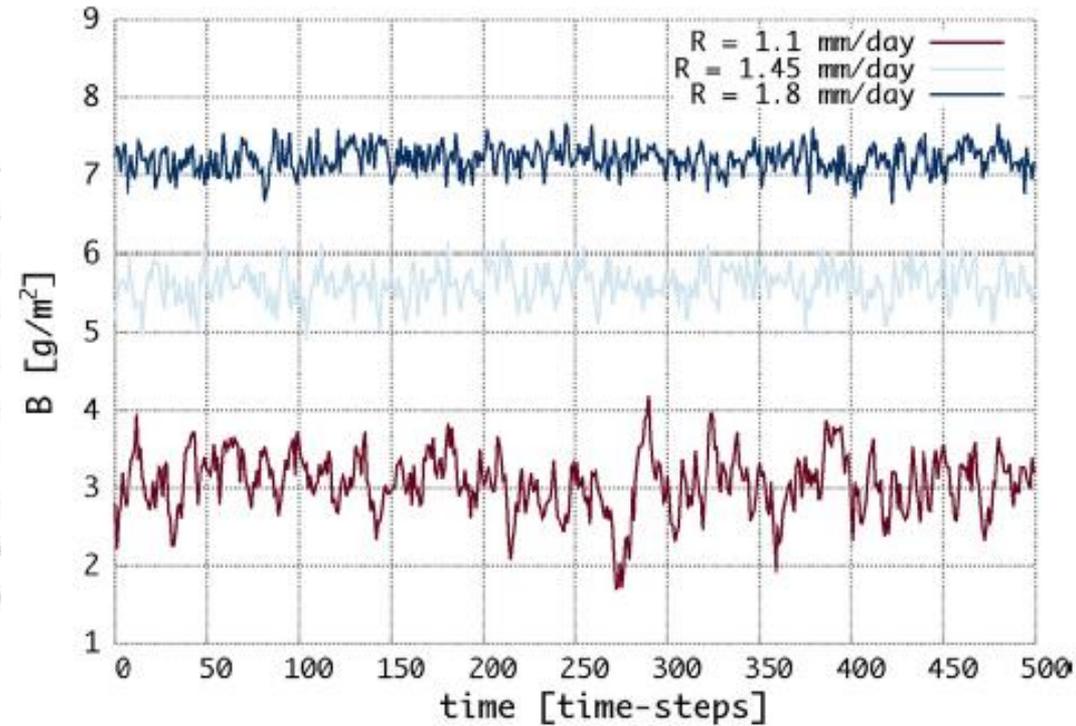
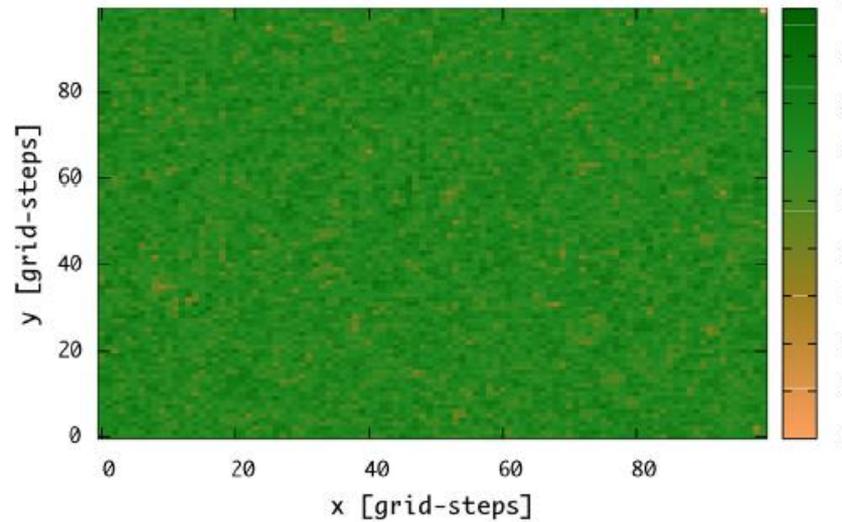


$R < R_c$: only desert-like solution ($B=0$)

$R_c = 1.067$ 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

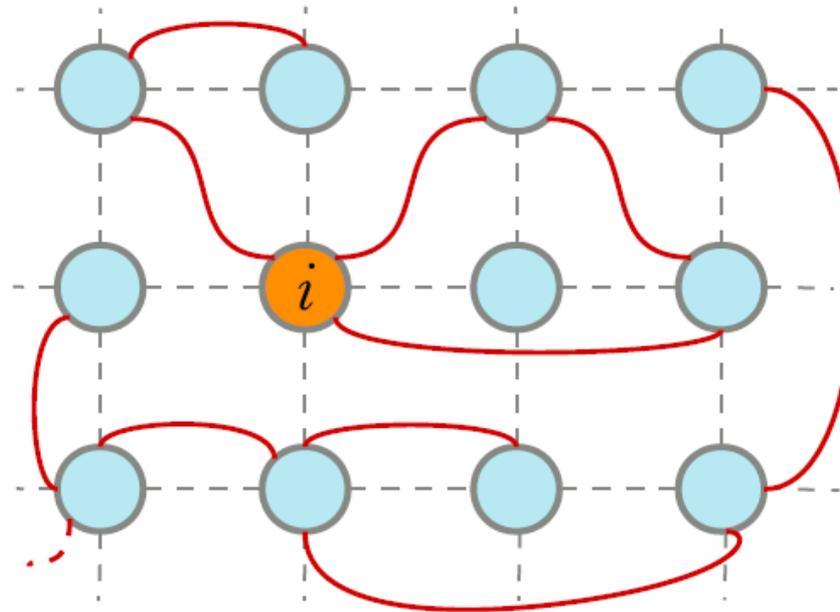
Correlation Network

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

A_{ij} is the adjacency matrix

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

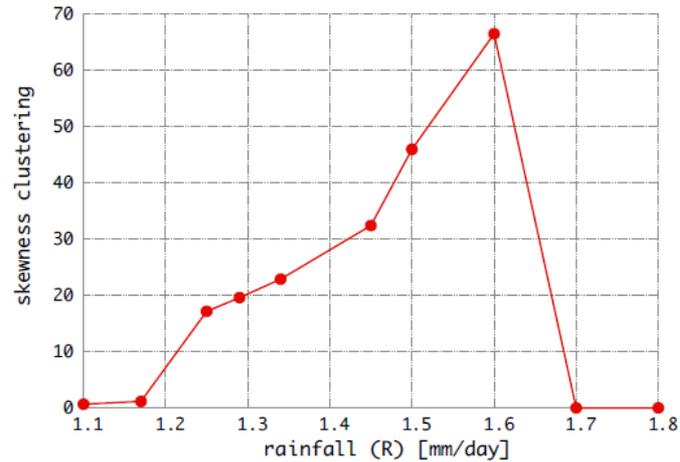
Threshold: $Th=0.2$ keeps only significant correlations ($p<0.05$)



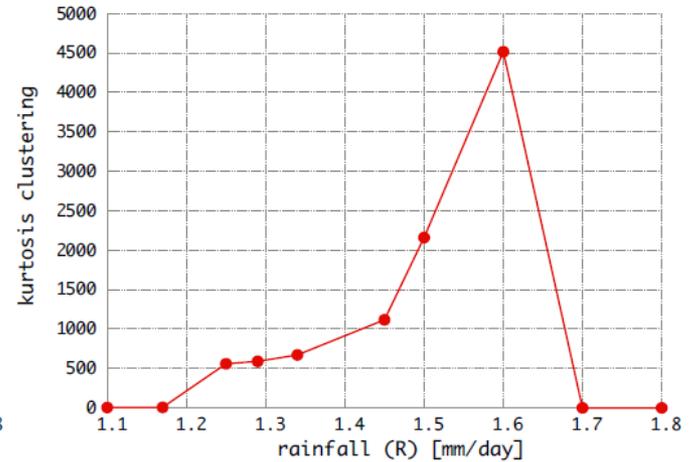
“Randomization” of the correlation network as the tipping point is approached

clustering

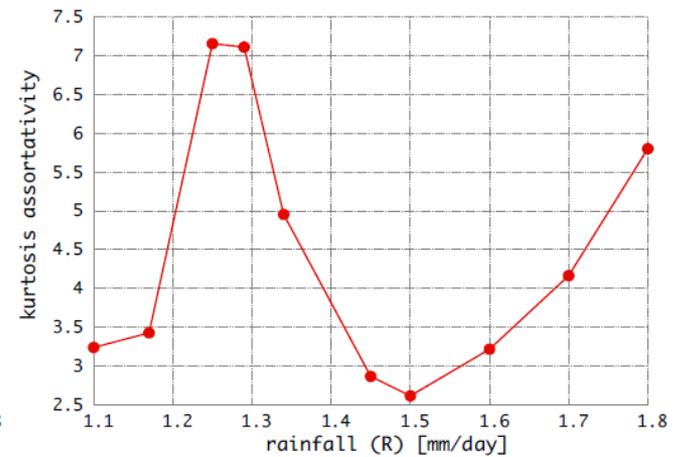
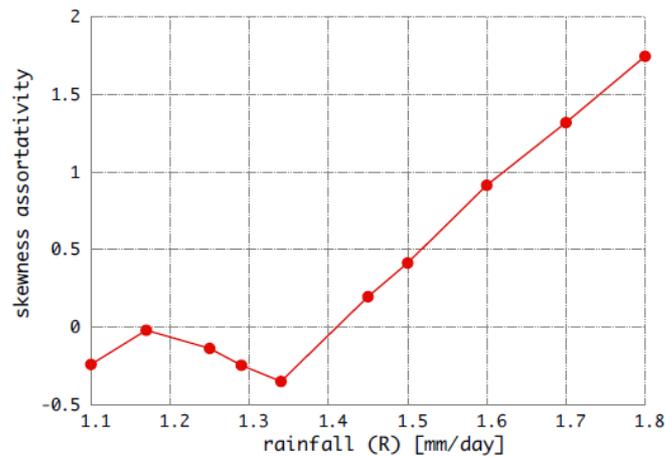
skewness



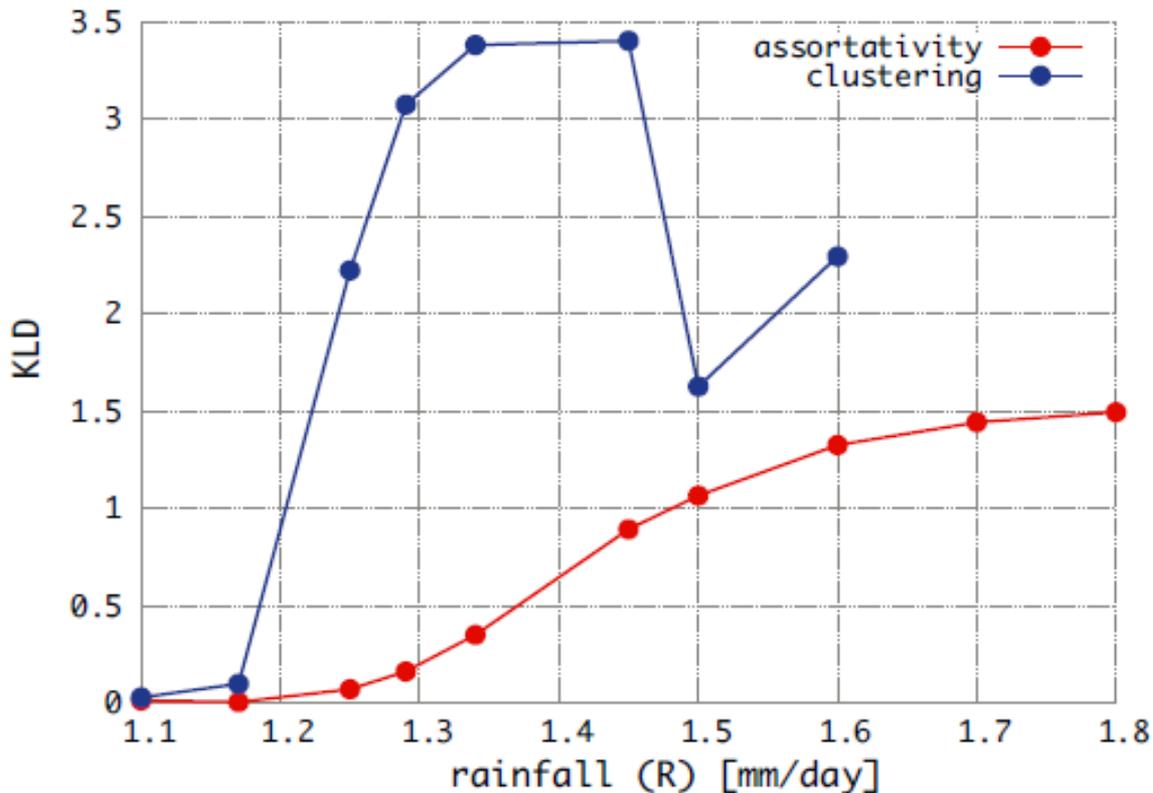
kurtosis



assortativity



The “Gaussianisation” of the distributions of a_i & c_i values is quantified by the Kullback–Leibler Distance



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

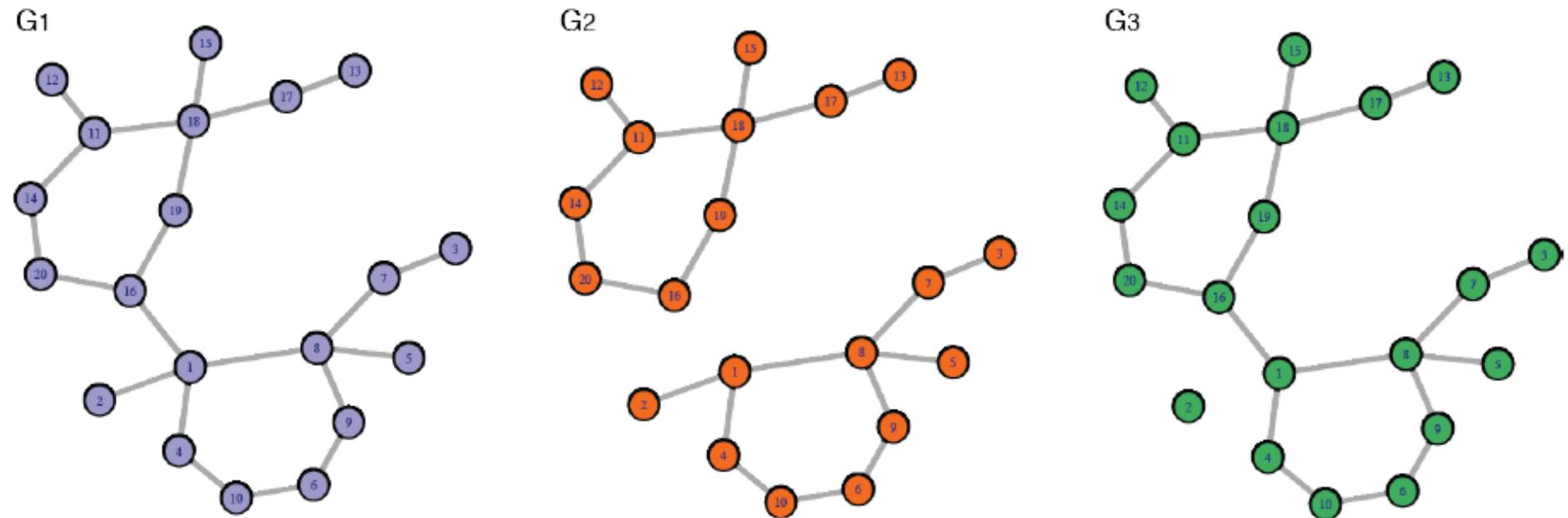
- Open issue: the “Gaussianisation” might be a model-specific feature.
- How to quantify the changes of the network?
- We need a distance to compare graphs.

**How to compare different
networks?**

Labelled networks with the same size

- Hamming distance $d_{\text{Hamming}}(\mathbf{y}_1, \mathbf{y}_2) = \sum_{i \neq j}^N [A_{ij}^{(1)} \neq A_{ij}^{(2)}]$

- Main problem: not all the links have the same importance.



In order to detect structural differences we need a precise measure to compare networks

- Degree, centrality, assortativity distributions etc. provide *partial* information.
- How to define a measure that contains detailed information about the *global topology* of a network, in a *compact way*?

⇒ Node Distance Distributions (NDDs)

- $p_i(j)$ of node “i” is the fraction of nodes that are connected to node i at distance j
- If a network has N nodes:

NDDs = vector of N pdfs $\{p_1, p_2, \dots, p_N\}$

- If two networks have the same set of NDDs ⇒ they have the same diameter, average path length, etc.

How to condense the information contained in the node distance distributions?

- The *Network Node Dispersion (NND)* measures the heterogeneity of the N pdfs $\{p_1, p_2, \dots, p_N\}$
- Quantifies the heterogeneity of connectivity distances.

$$\text{NND}(G) = \frac{\mathcal{J}(\mathbf{P}_1, \dots, \mathbf{P}_N)}{\log(d + 1)} \quad d = \text{diameter}$$

$$\mathcal{J}(\mathbf{P}_1, \dots, \mathbf{P}_N) = \frac{1}{N} \sum_{i,j} p_i(j) \log\left(\frac{p_i(j)}{\mu_j}\right)$$

$$\mu_j = \left(\sum_{i=1}^N p_i(j)\right) / N$$

Dissimilarity between two networks

$$D(G, G') = w_1 \sqrt{\frac{\mathcal{J}(\mu_G, \mu_{G'})}{\log 2}} + w_2 \left| \sqrt{\text{NND}(G)} - \sqrt{\text{NND}(G')} \right| \quad w_1=w_2=0.5$$

compares the
averaged
connectivity

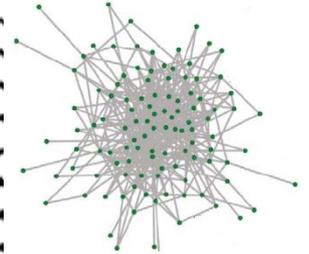
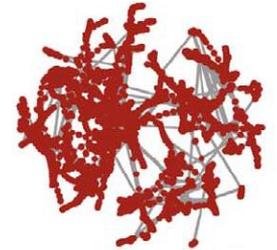
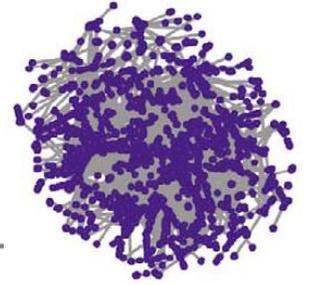
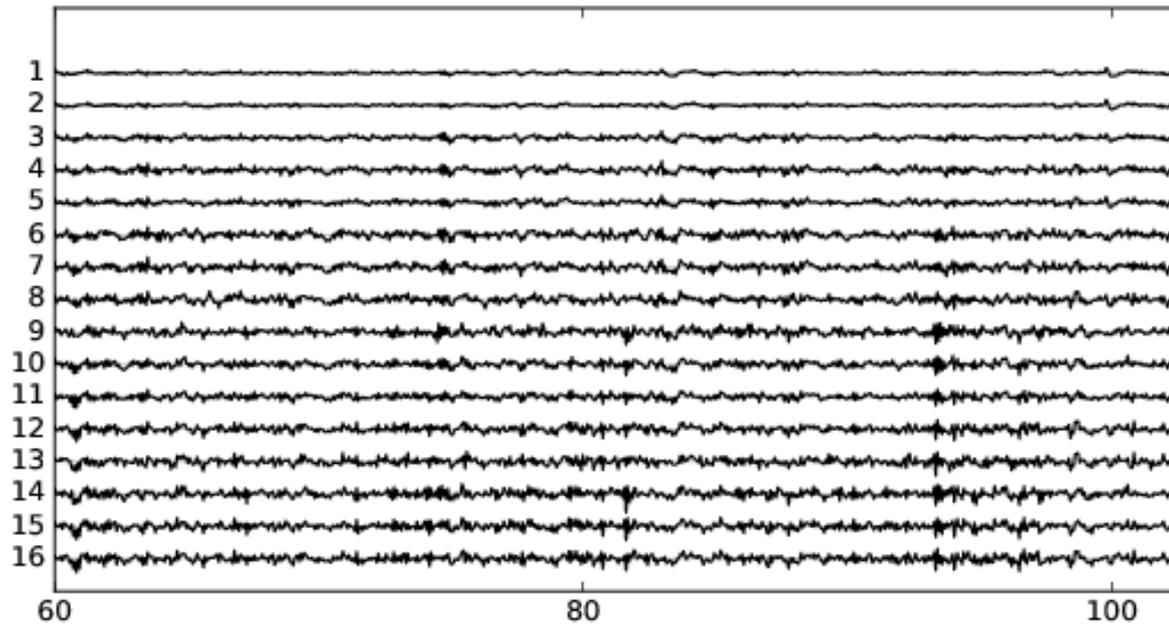
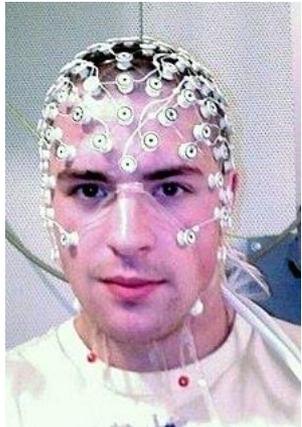
compares the
heterogeneity of the
connectivity distances

- Extensive numerical experiments demonstrate that isomorphic graphs return **$D=0$** .
- Computationally efficient.

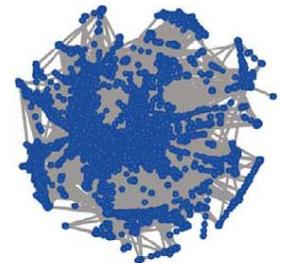
Application: comparing brain networks

- EEG data
 - <https://archive.ics.uci.edu/ml/datasets/eeg+database>
 - 64 electrodes placed on the subject's scalp sampled at 256 Hz during 1s
 - 107 subjects: 39 control and 68 alcoholic
- Use HVG to transform each EEG TS into a network G .
- Weight between two brain regions: $1-D(G,G')$
- The resulting network represents the weighted similarity between the brain regions of an individual.
 - ⇒ We can compare the different individuals.

For each subject, the time series recorded at each electrode is transformed into a graph



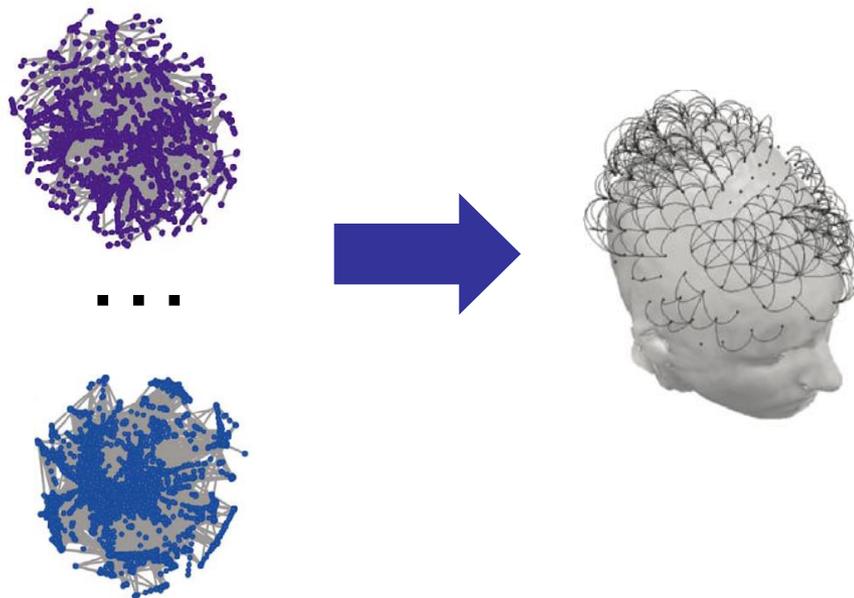
■ ■ ■



Dataset has 64 channels \Rightarrow 64 networks

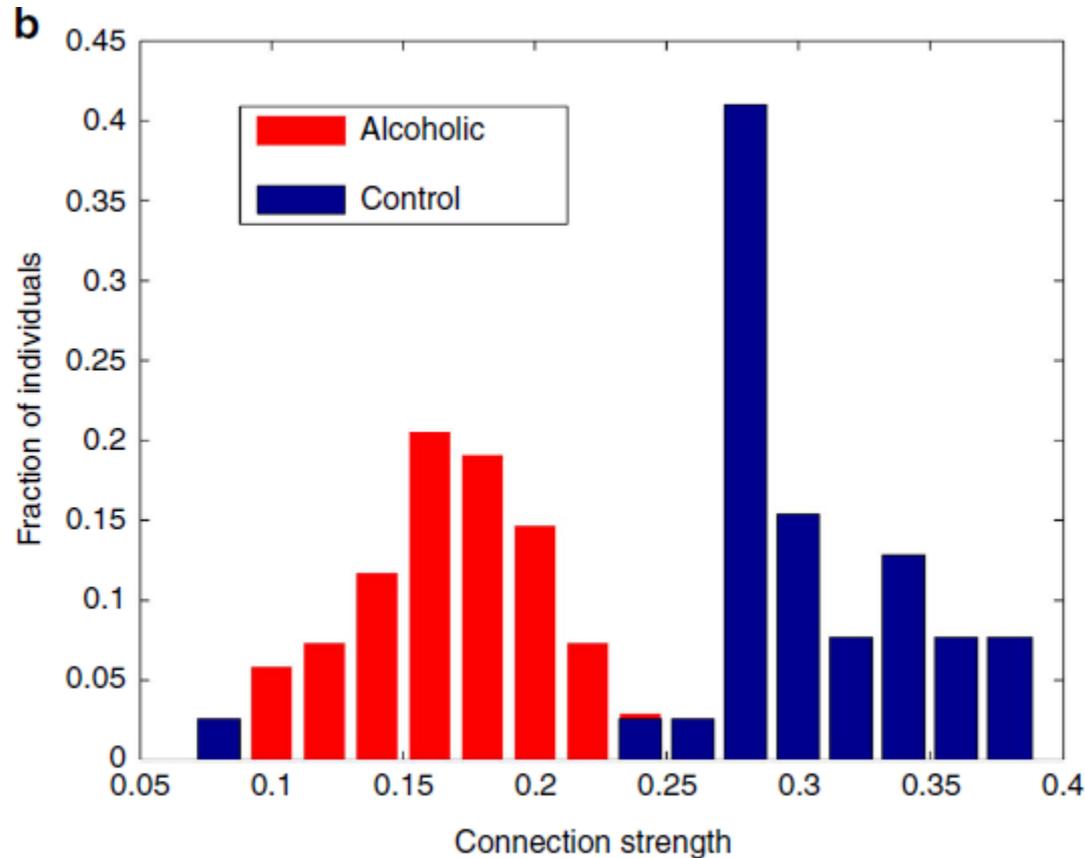
The brain network of each subject

- The weight of the link between two graphs (G, G') representing two brain regions is defined as: $1-D(G, G')$



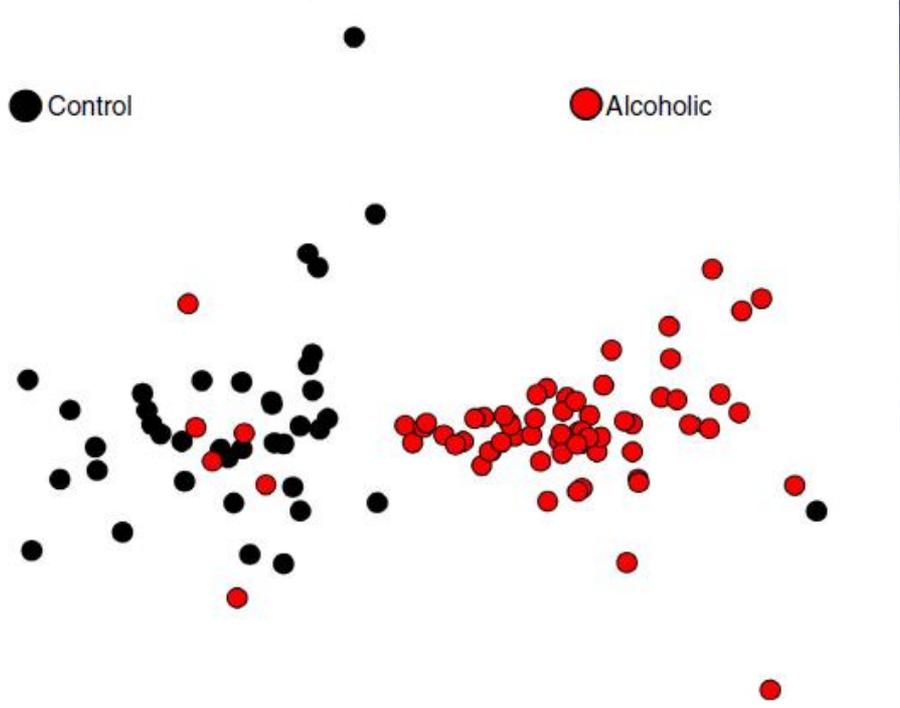
- The resulting network (with 64 nodes=electrodes, all-to-all coupled with weighted links) represents the similarity between the EEG signals in different brain regions of one subject.
⇒ We can then compare different subjects.

We identify two brain regions (called 'nd' and 'y'), where the connection strength between these regions is higher in control than in alcoholic subjects.

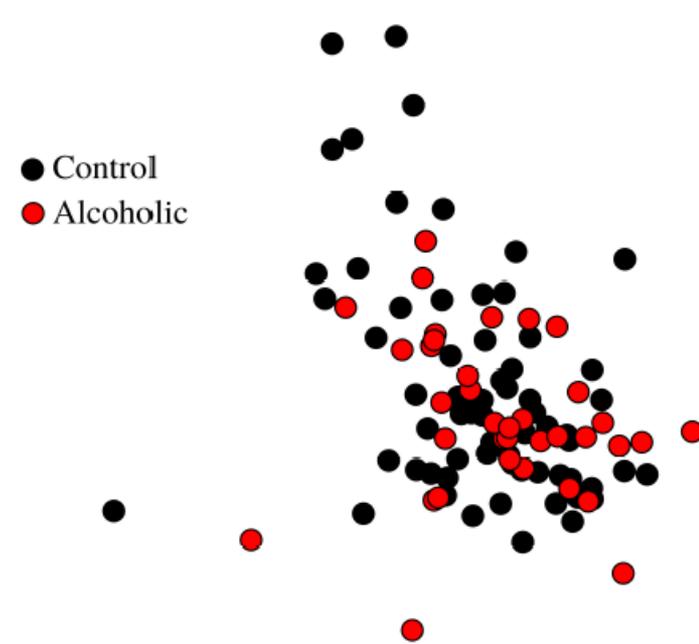


Using the Hamming distance we can not distinguish.

Dissimilarity measure



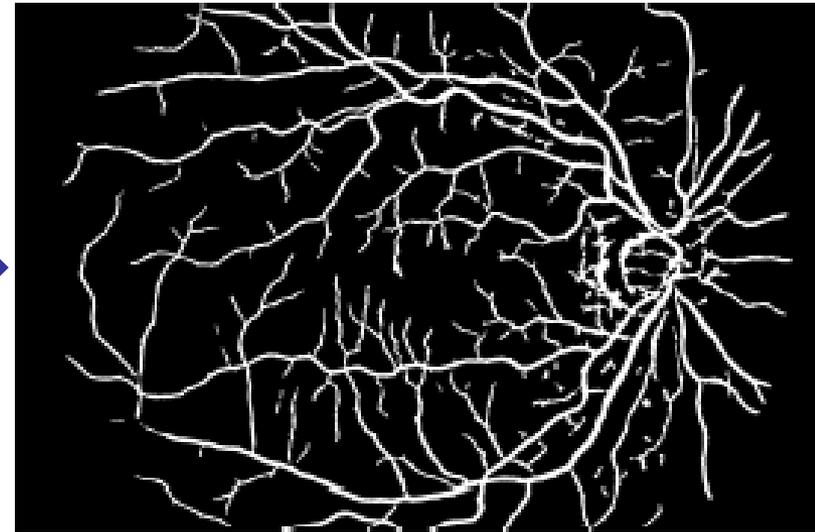
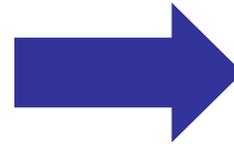
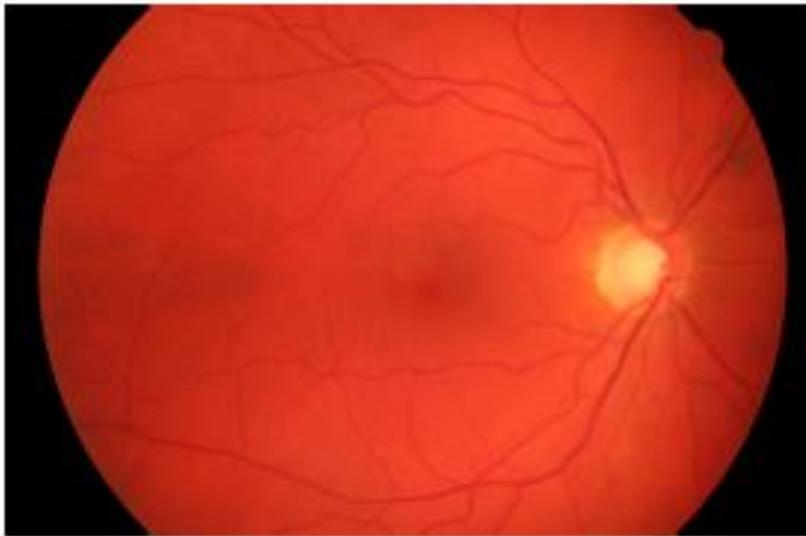
Hamming distance



[T. A. Schieber et al, Nat. Comm. 8, 13928 \(2017\)](#)

Retina image classification

Pablo Amil (UPC), **Fabian Reyes** (Mexico) & **Irene Sendiña** (Madrid)



ITN

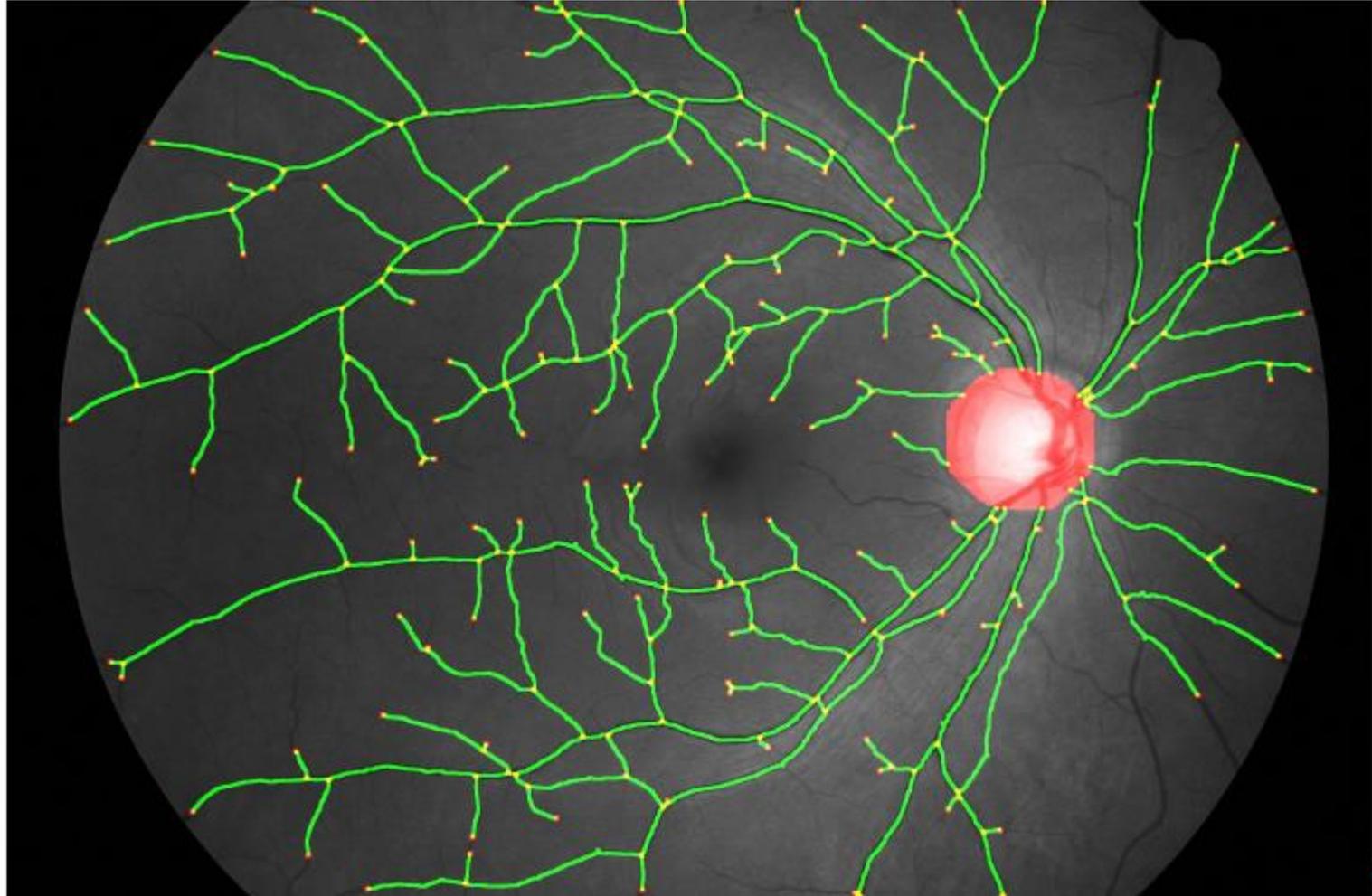


BE-OPTICAL

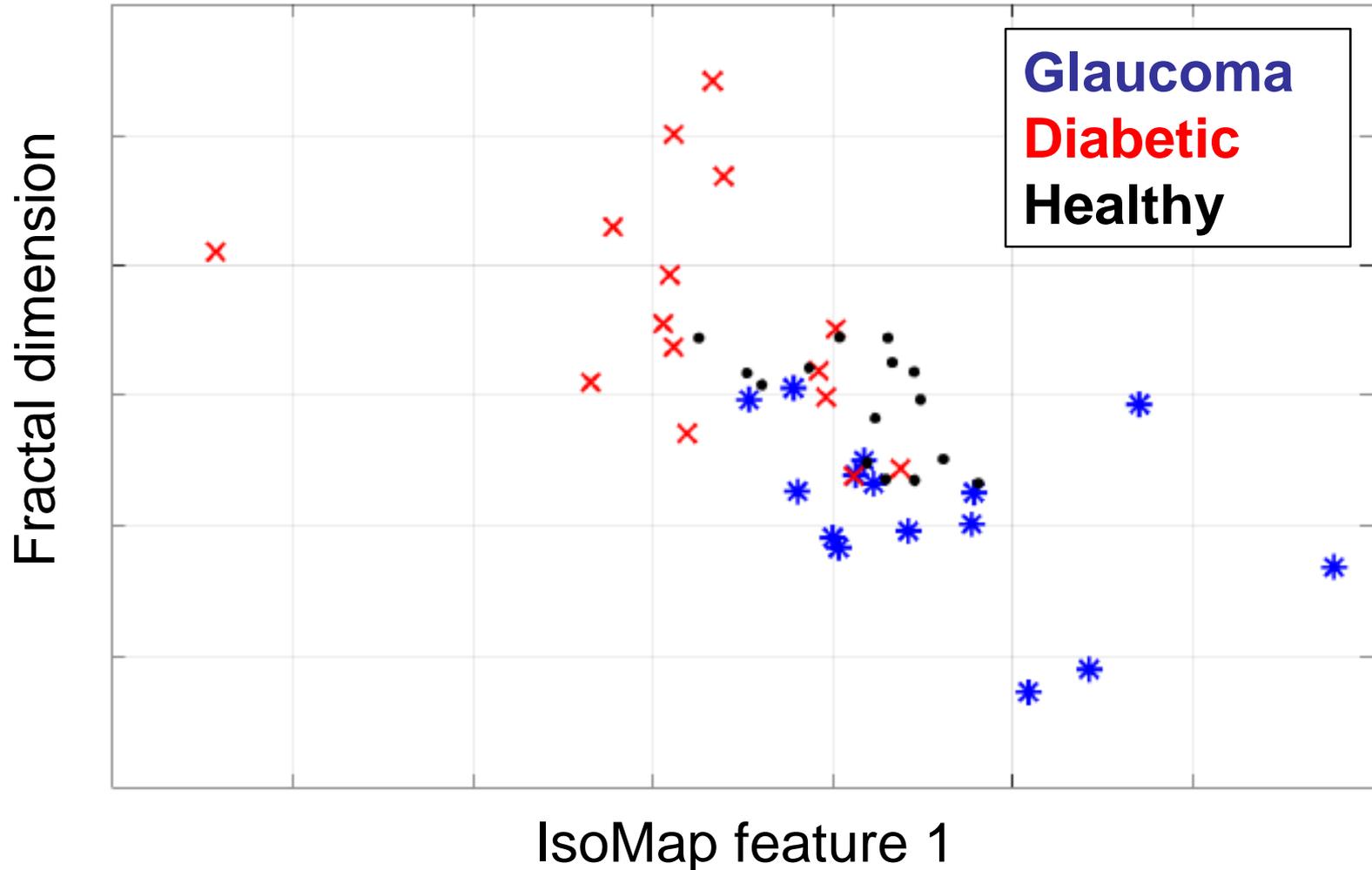
*Advanced Biomedical Optical
Imaging and Data Analysis*

Beoptical.eu

Network identification and analysis of the connectivity paths of the central node



Analysis of the node distance distribution (NDD) of the central node + fractal dimension analysis: promising classification tool



Exercise 3: for the electronic circuits calculate the distance between the functional and the structural networks

Compare the degree distributions of the functional networks constructed in Ex. 2 with the “ground truth”: the degree distribution of the structural network.

**Network inference:
how to infer the underlying
interactions from observed data?
a classification problem**

Main problem:

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

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

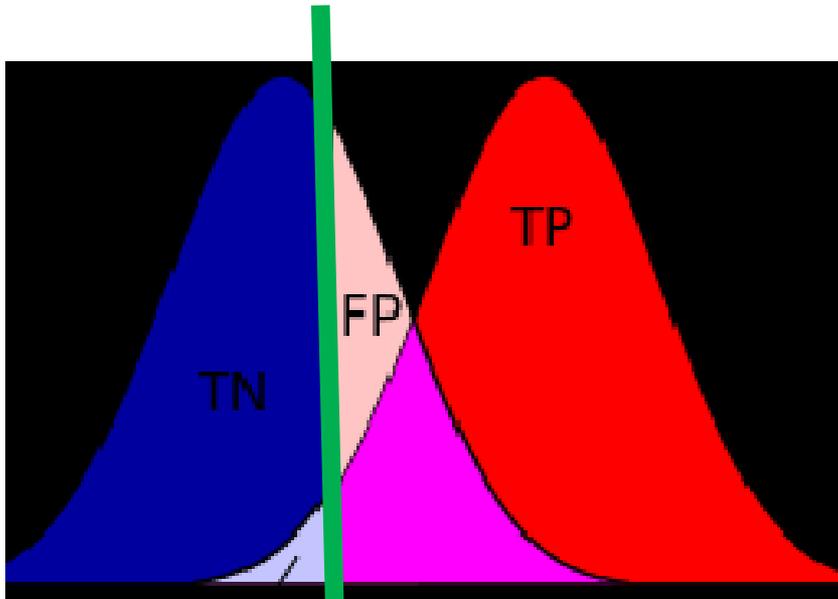
- There are many statistical similarity measures to infer bi-variate mutual interactions from observations, i.e., to classify:
 - the interaction exists (is significant)
 - the interaction does not exist (or is not significant)

	Predicted: NO	Predicted: YES
Actual: NO	TN	FP
Actual: YES	FN	TP

Confusion matrix

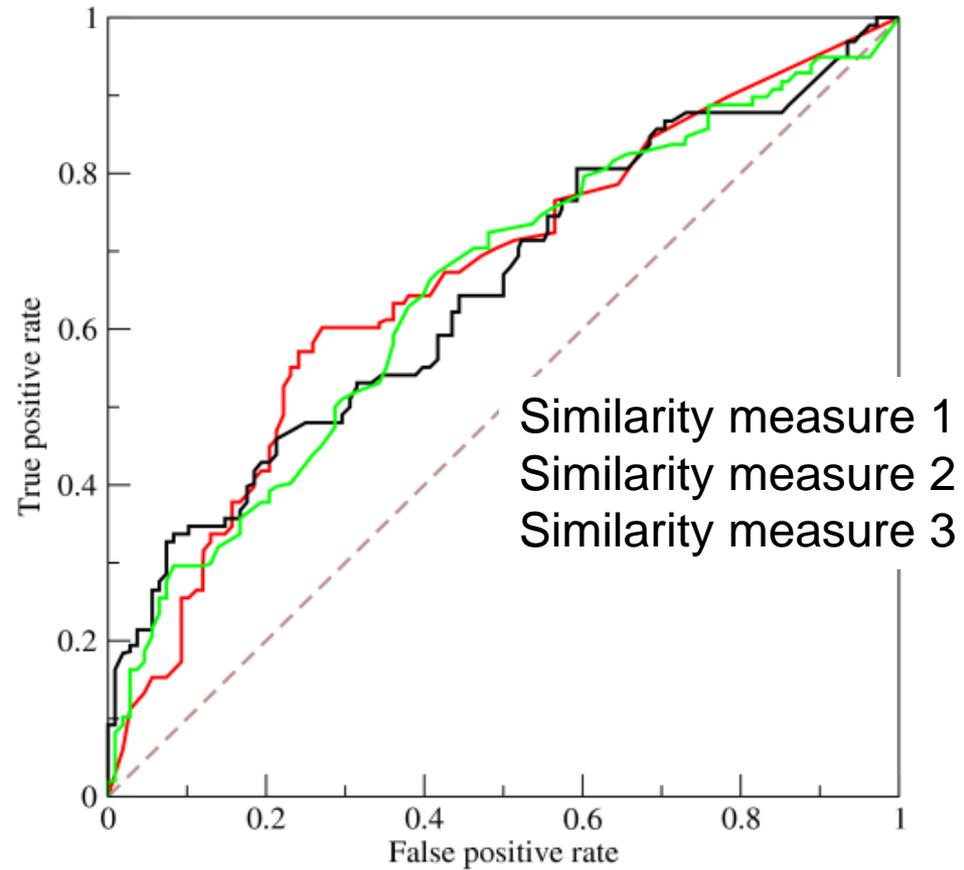
- **Accuracy**: How often is the classifier correct? **$(TP+TN)/total$**
- **Misclassification** (Error Rate): How often is it wrong? **$(FP+FN)/total$**
- **True Positive Rate** (TPR, Sensitivity): When it's yes, how often does it predict yes? **$TP/actual\ yes$**
- **False Positive Rate** (FPR) : When it's no, how often does it predict yes? **$FP/actual\ no$**
- **Specificity** ($1 - FPR$) : When it's no, how often it predicts no? **$TN/actual\ no$**
- **Precision** (Positive Predictive Value): When it predicts yes, how often is it correct? **$TP/predicted\ yes$**
- **Negative Predictive Value**: When it predicts no, how often is it correct? **$TN/predicted\ no$**
- **Prevalence**: How often does the yes condition actually occur in the sample? **$actual\ yes/total$**

Receiver operating characteristic (ROC curve)



TP	FP
FN	TN

Source: wikipedia



Our goal

- To compare the performance of different statistical similarity measures for inferring interactions from observations.
- Using a “toy model” where **we know** the underlying equations and interactions and so we can check the performance of the different measures in inferring the interactions.

Kuramoto oscillators in a random network

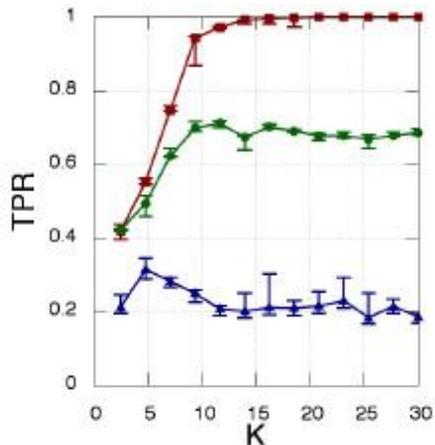
$$d\theta_i = \omega_i dt + \frac{K}{N} \sum_{j=1}^N A_{ij} \sin(\theta_j - \theta_i) dt + D dW_t^i$$

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

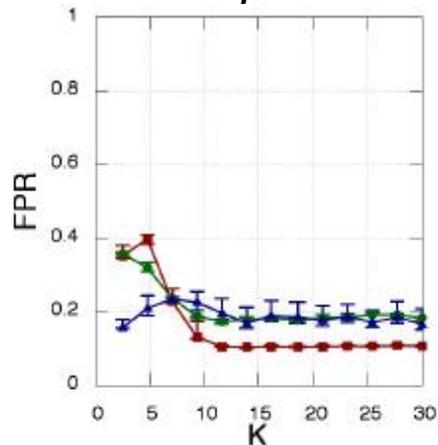
Phases (θ)

CC MI MIOP

True positives

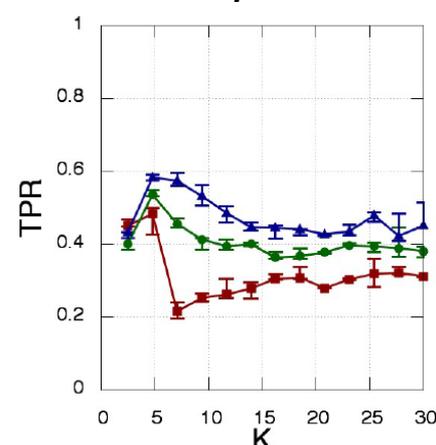


False positives

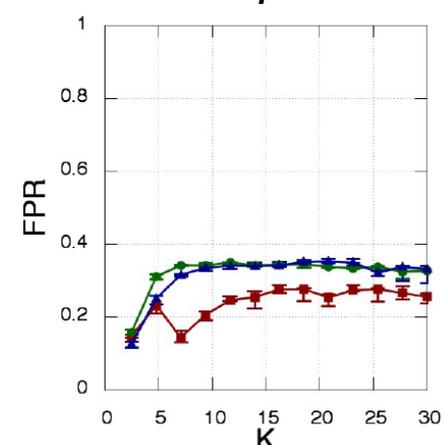


“Observable” $Y=\sin(\theta)$

True positives



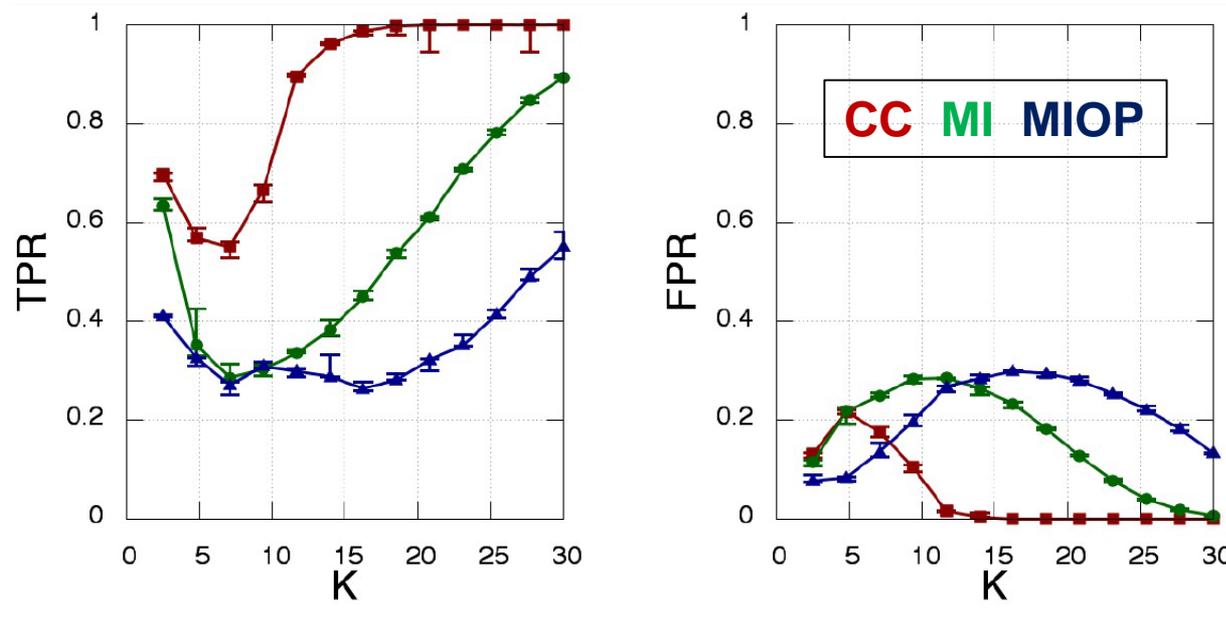
False positives



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 ($d\theta/dt$)



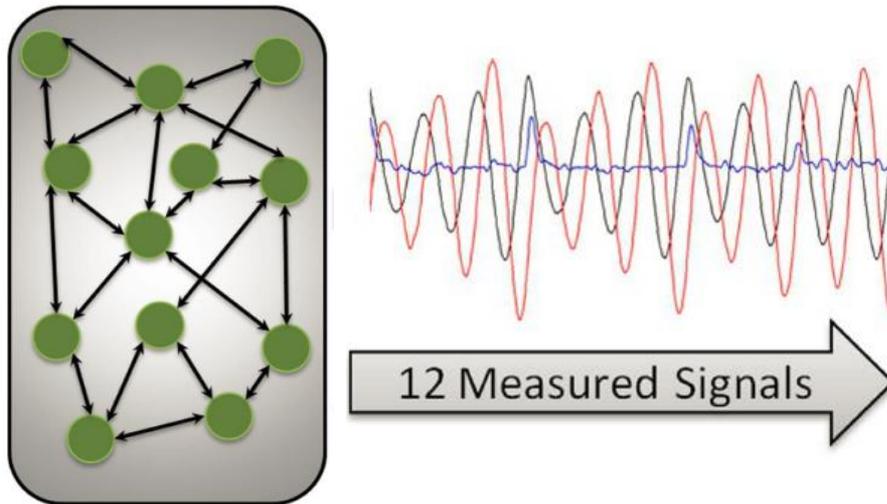
Perfect network inference is possible!

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)



The Hilbert Transform was used to obtain phases from experimental data

[G. Tirabassi et al, Sci. Rep. 5 10829 \(2015\)](#)

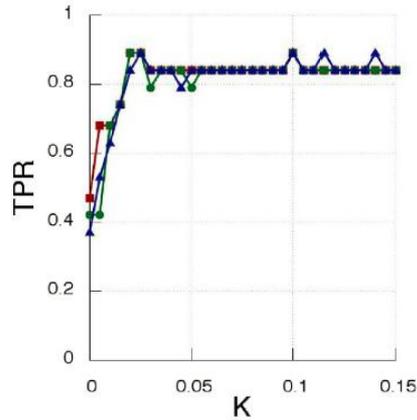
- Kuramoto Oscillators' Network

θ_i
 $f_i = \dot{\theta}_i$
 $Y_i = \sin(\theta_i)$

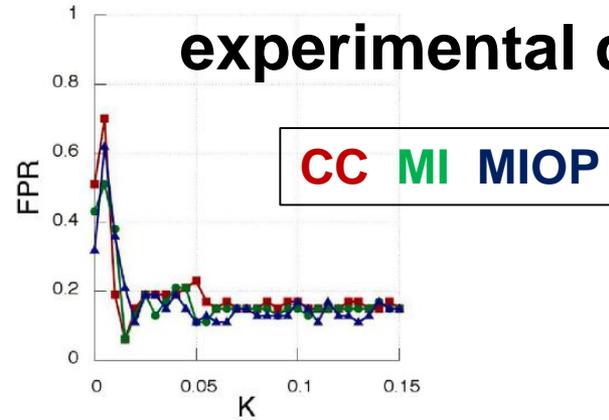
- Rössler Oscillators' Network

$\varphi_i = HT(x_i)$
 $f_i = \dot{\varphi}_i$
 x_i

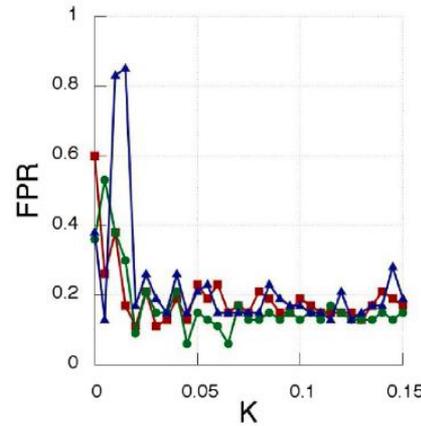
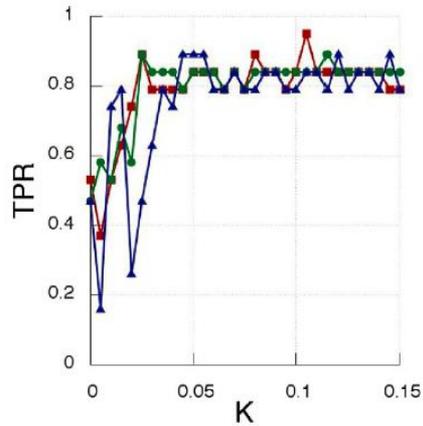
Observed variable (x)



Results obtained with experimental data



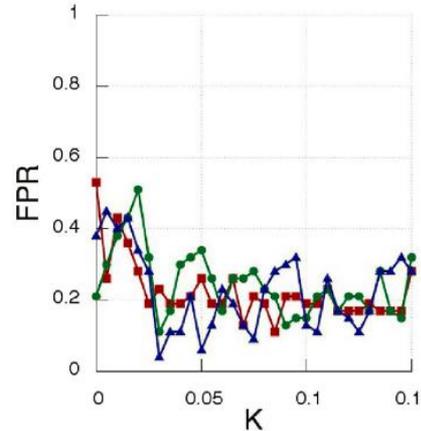
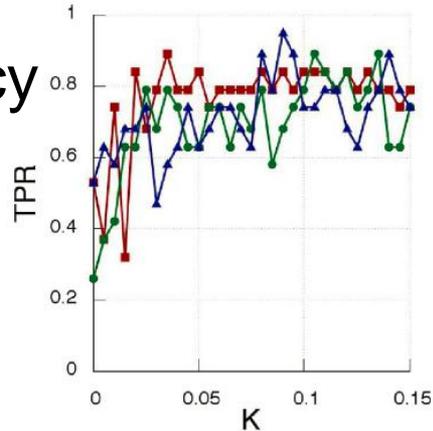
Hilbert phase



– No perfect reconstruction

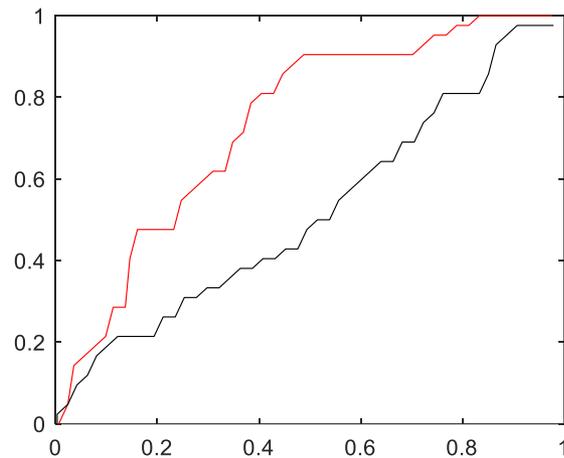
– No important difference among the 3 methods & 3 variables

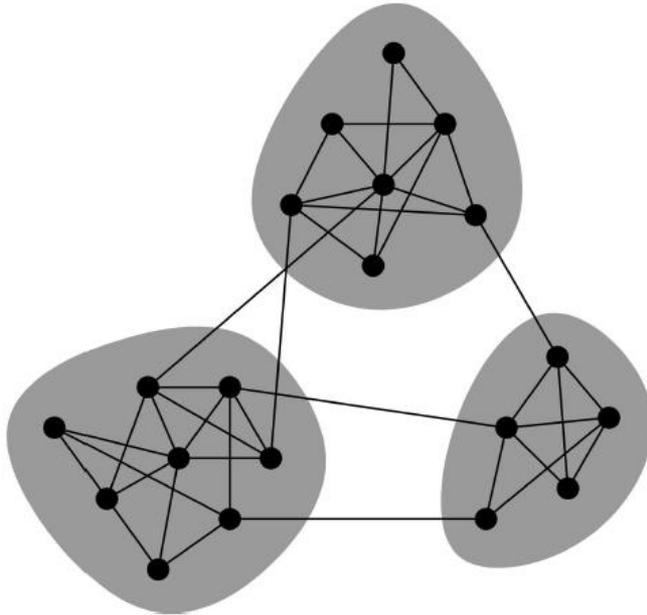
Hilbert frequency



Exercise 4: for the 28 electronic circuits, can we infer the structural network?

Calculate the ROC curve for different coupling strengths.



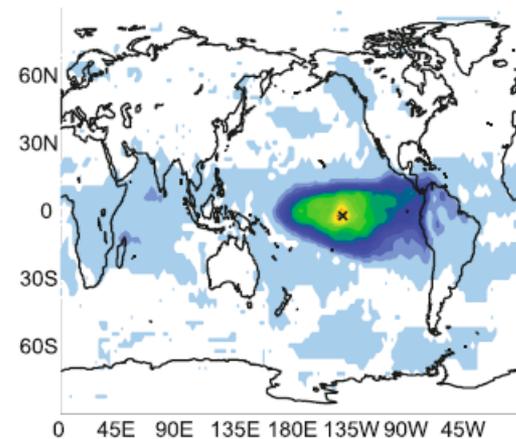
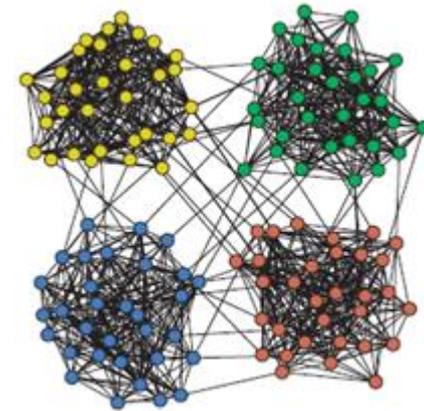


Community detection

Climate “communities”

How to identify regions with similar climate?

- Goal: to construct a network in which regions with similar climate (e.g., continental) are in the same “community”.
- Problem: not possible with the “usual” correlation-based method to construct the network because NH and SH are only indirectly connected.



Network construction based on similar symbolic dynamics

- Step 1: transform SAT anomalies in each node in a sequence of symbols (we use ordinal patterns)

$$s_i = \{012, 102, 210, 012, \dots\} \quad s_j = \{201, 210, 210, 012, \dots\}$$

- Step 2: in each node compute the transition probabilities

$$TP_{\alpha\beta}^i = \#(\alpha \rightarrow \beta) / N$$

- Step 3: define the weights

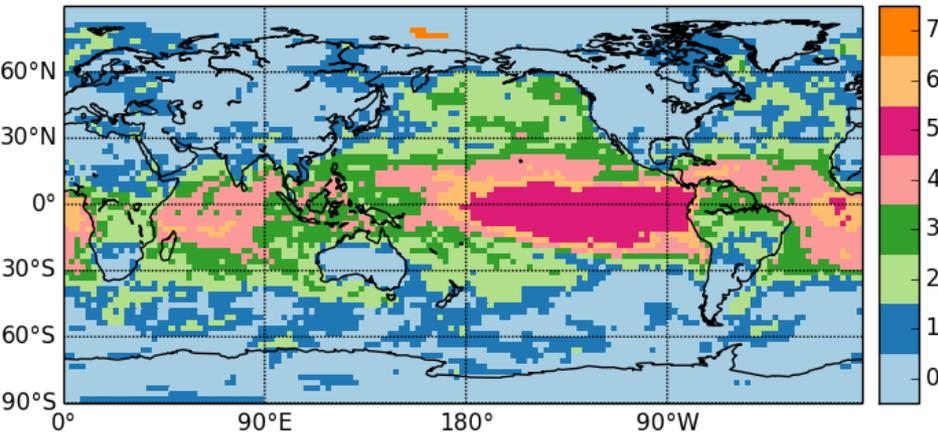
$$w_{ij} = \frac{1}{\sum_{\alpha\beta} (TP_{\alpha\beta}^i - TP_{\alpha\beta}^j)^2}$$

High weight
if similar
symbolic
“language”

- Step 4: threshold w_{ij} to obtain the adjacency matrix.
- Step 5: run a *community detection algorithm (Infomap)*.

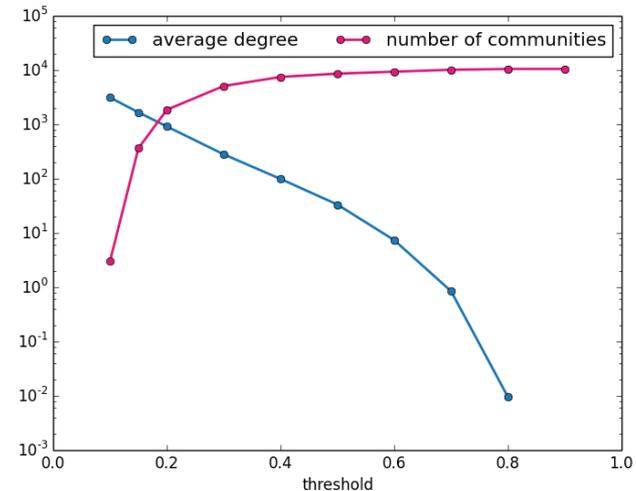
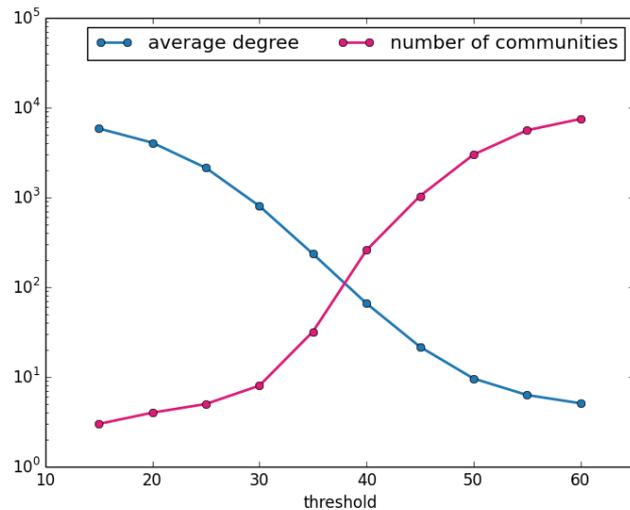
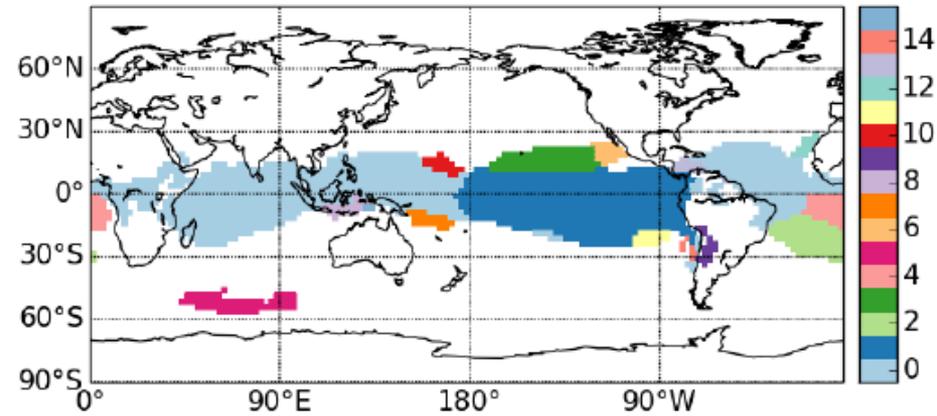
Results

TP Network



CC Network

(only the largest 16)



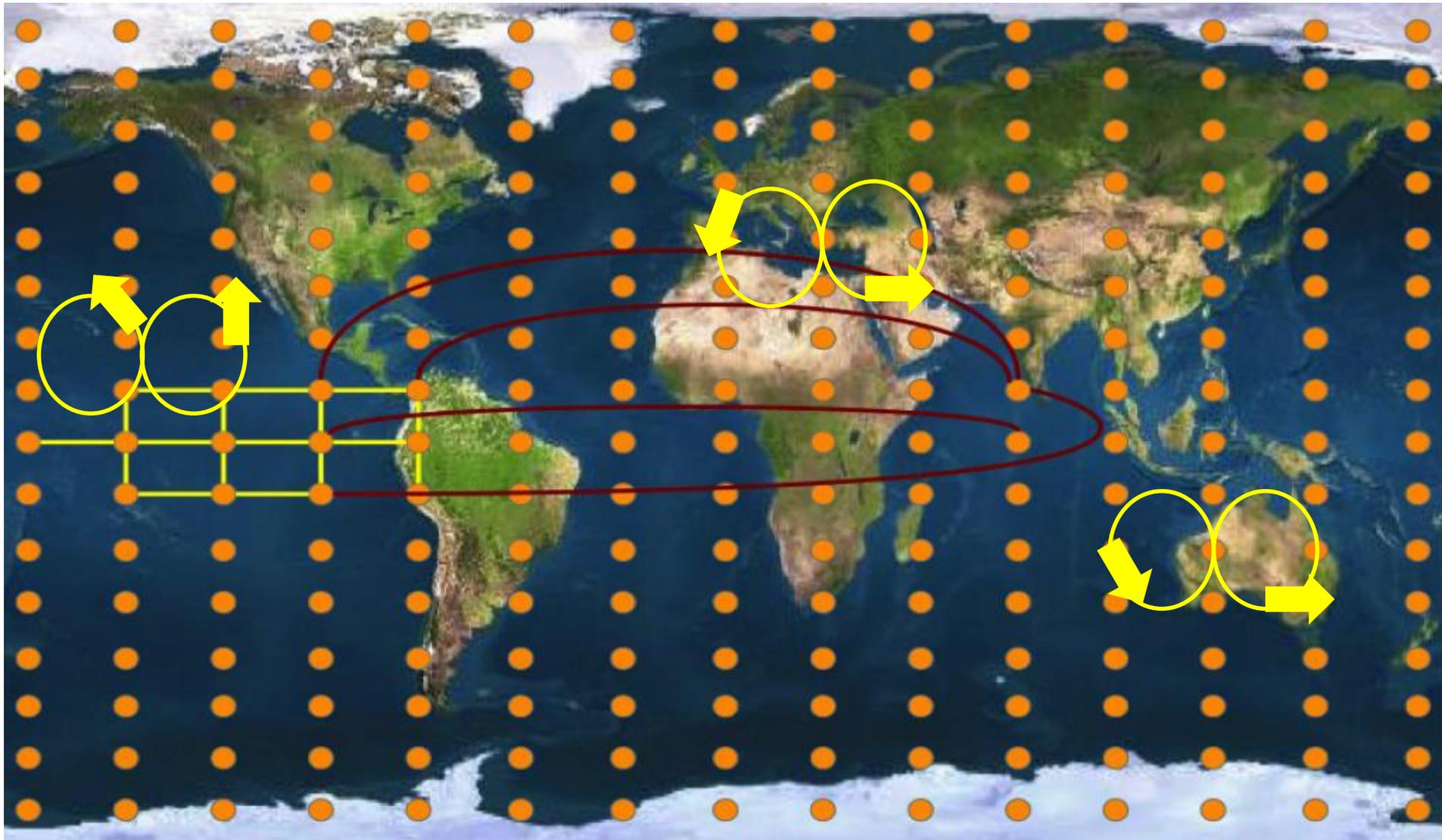
G. Tirabassi and C. Masoller, "Unravelling the community structure of the climate system by using lags and symbolic time-series analysis", [Sci. Rep. 6, 29804 \(2016\)](https://doi.org/10.1038/s41598-016-02980-4).

Community detection algorithms

- *Infomap* (<http://www.mapequation.org/code.html>) and many others.
- *Infomap* clusters tightly interconnected nodes into modules and detects nested modules.
- Many other algorithms have been proposed.
- Further reading: S. Fortunato, “*Community detection in graphs*”, Phys. Rep. 486, 75 (2010).

**How to detect phase
synchronization in climate data?**

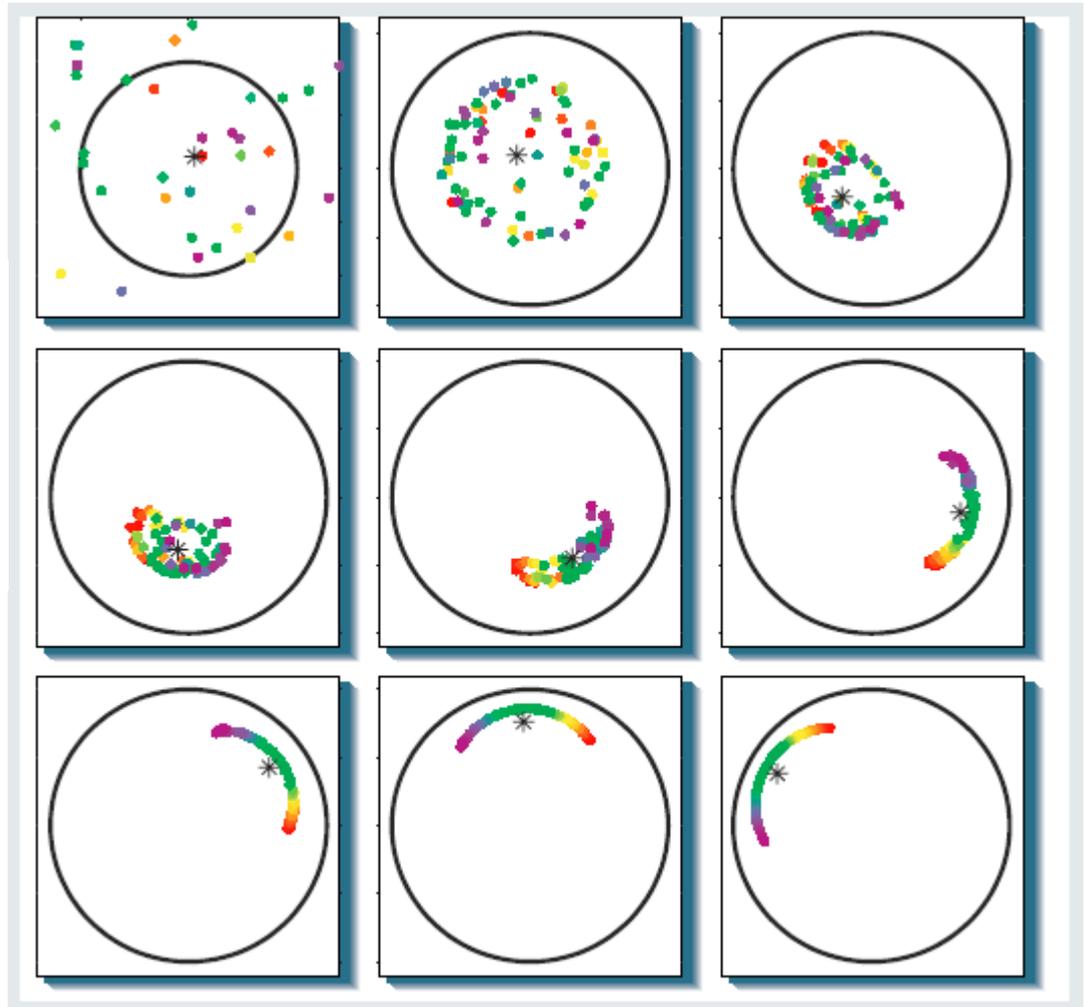
Network of individual oscillators



Quantifying phase synchronization

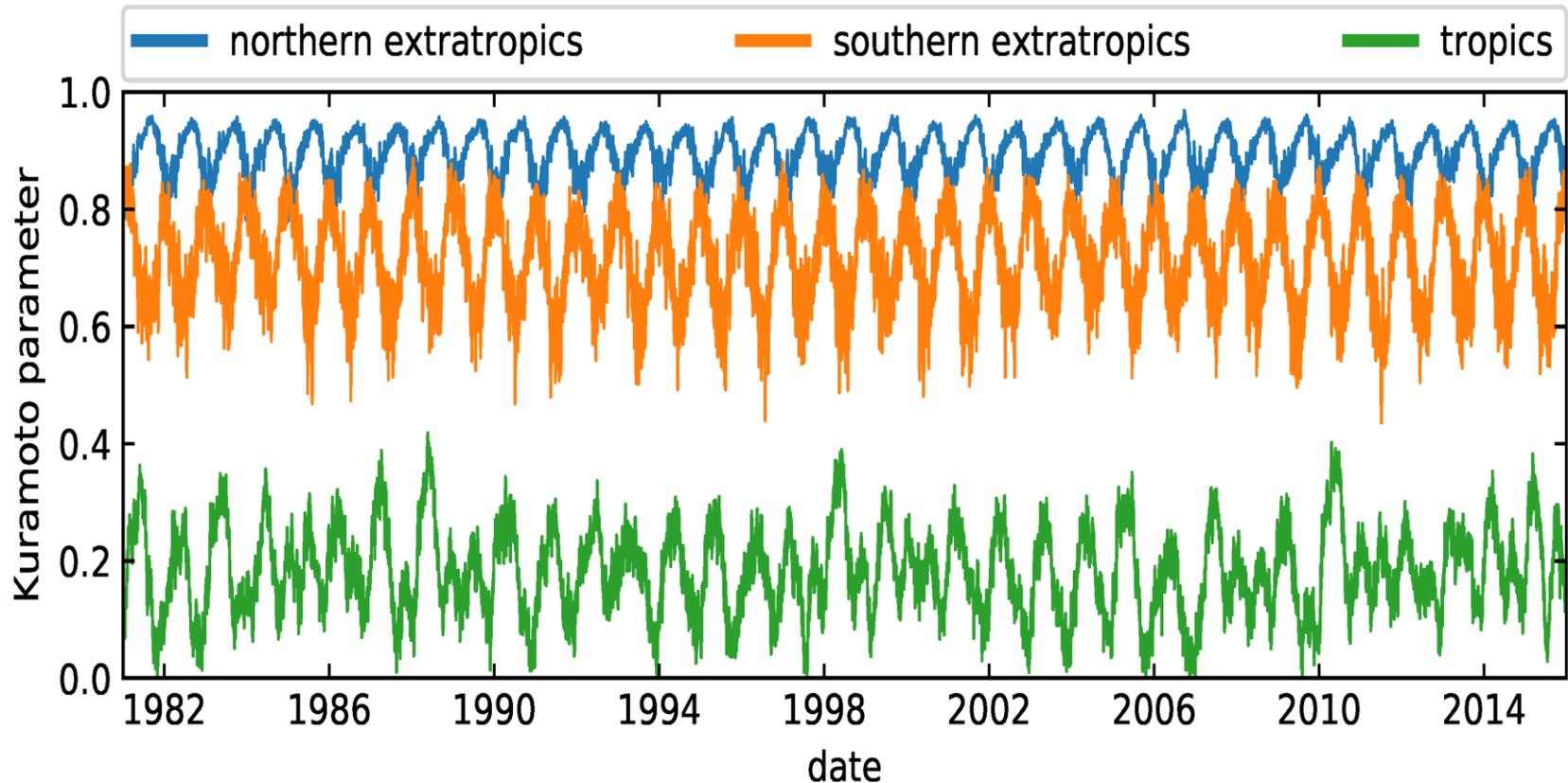
- Kuramoto order parameter

$$r(t) = \left| \frac{1}{N} \sum_{j=1}^N e^{i\theta_j(t)} \right|$$



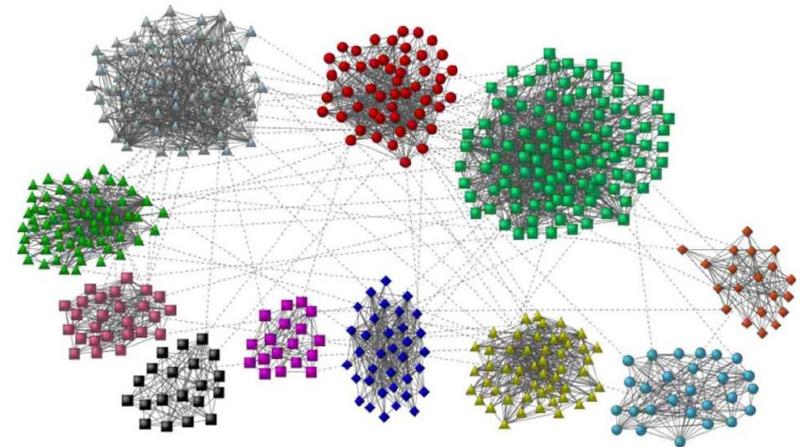
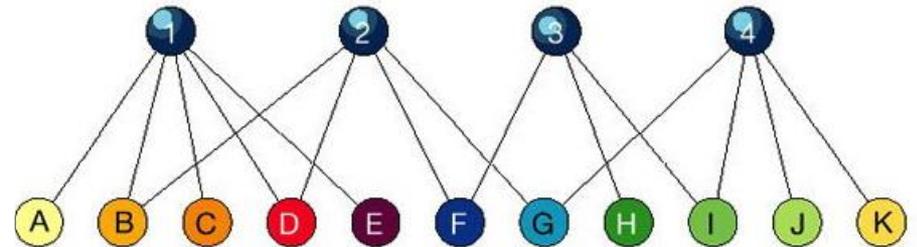
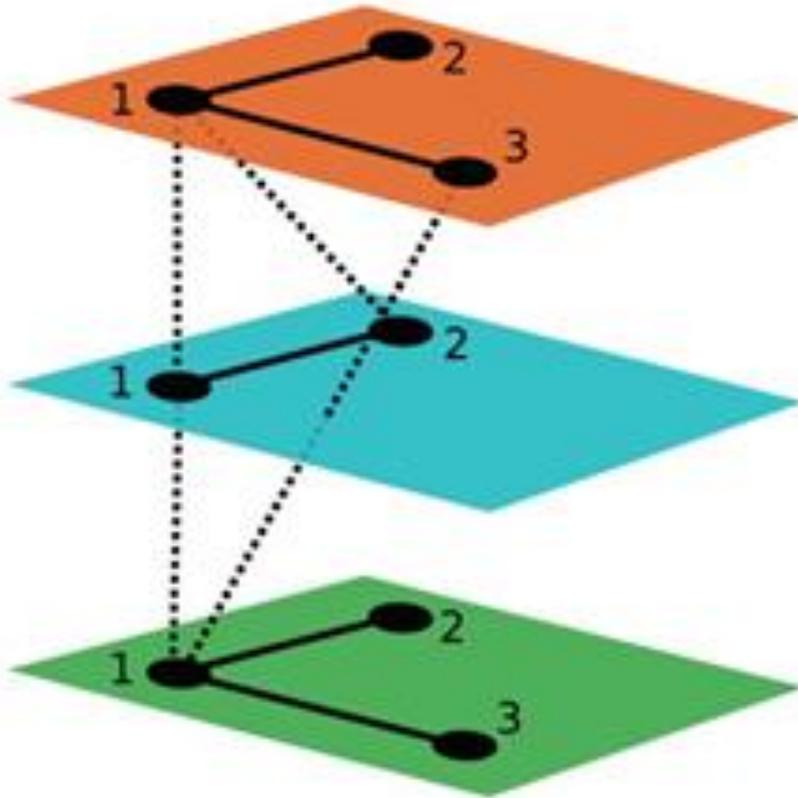
Hilbert transform \Rightarrow phase time series \Rightarrow area-weighted Kuramoto order parameter

$$r(t) = \left| \frac{\sum_{j \in S} w_j e^{i\psi_j(t)}}{\sum_{j \in S} w_j} \right|$$

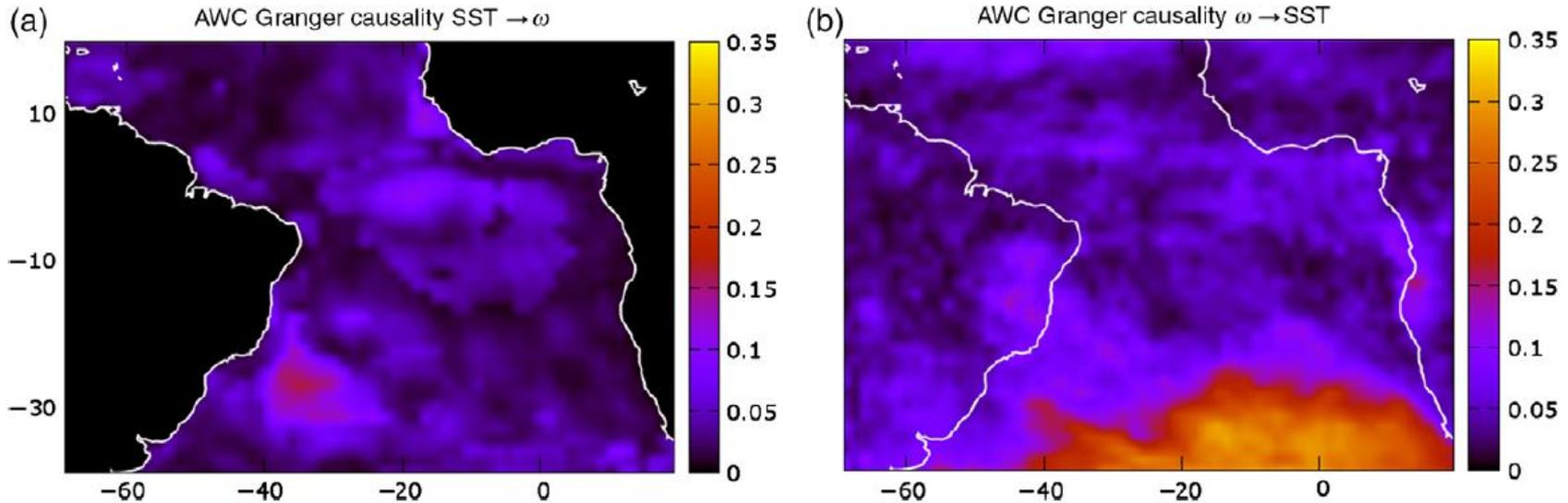


Generalizations of complex network analysis

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



Example of a bilayer climate network representing ocean-precipitation interactions



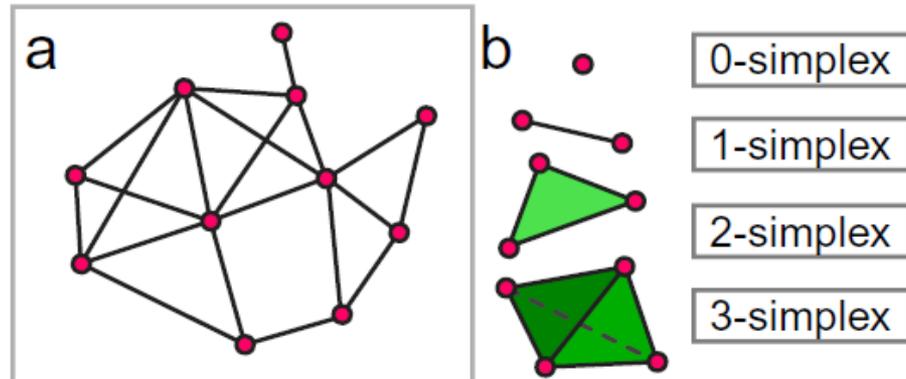
Color code shows the area-weighted connectivity (weighted degree) of a bilayer network where the links are defined using Granger causality (only GCE values at 99% confidence level have been considered).

SST = Surface sea temperature

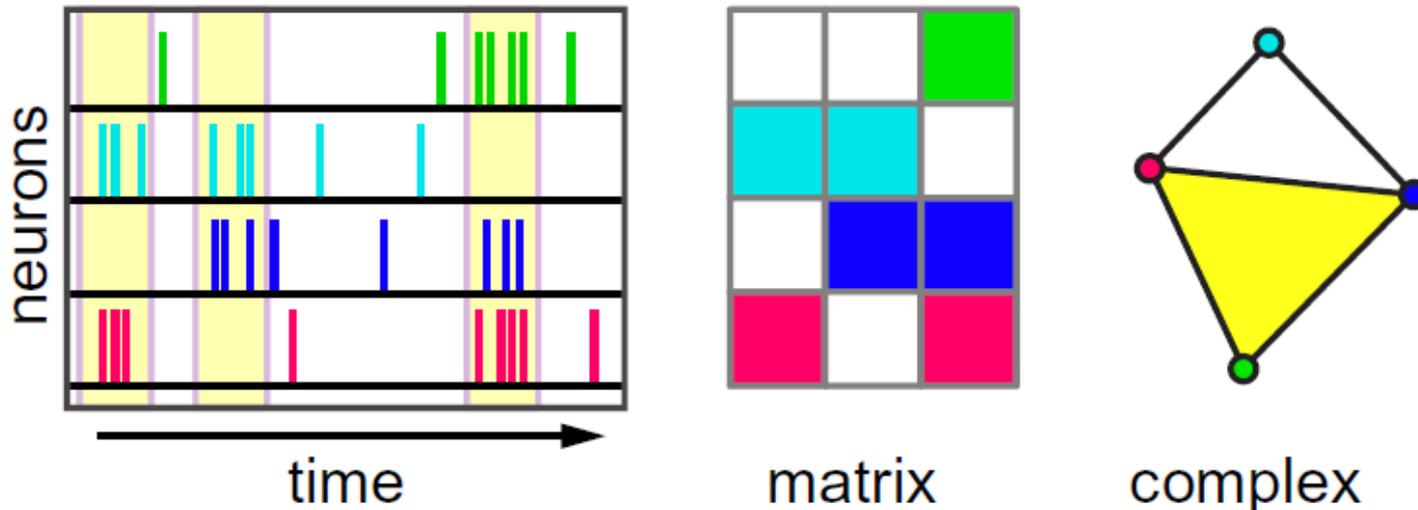
ω = vertical wind velocity at 500 hPa (precipitation proxy)

A basic limitation of network analysis

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



Example



Concluding

Take home messages

- There are many methods for inferring the underlying connectivity of a complex system from the observed output signals.
- Different methods infer different networks.
- Comparing (quantifying differences) between networks is challenging.
- Different sets of “communities” (clusters) can be uncovered depending on the property that is analyzed.

References

- [M. Barreiro, et. al, Chaos 21, 013101 \(2011\)](#)
- [Deza, Barreiro and Masoller, Eur. Phys. J. ST 222, 511 \(2013\)](#)
- [Tirabassi and Masoller, EPL 102, 59003 \(2013\)](#)
- [G. Tirabassi et al., Ecological Complexity 19, 148 \(2014\)](#)
- [Tirabassi et al, Sci. Rep. 5 10829 \(2015\)](#)
- [G. Tirabassi and C. Masoller, Sci. Rep. 6:29804 \(2016\)](#)
- [T. A. Schieber et al, Nat. Comm. 8, 13928 \(2017\)](#)

<crisrina.masoller@upc.edu>

<http://www.fisica.edu.uy/~cris/>