School on Applications of Nonlinear Systems to Socio-Economic Complexity, Oct. 17 – Oct. 22 2022

Nonlinear time series analysis

Cristina Masoller

Departamento de Fisica
Universitat Politecnica de Catalunya

Class 1: From dynamical systems to complex systems
Class 2: Univariate time series analysis
Class 3: Univariate time series analysis
Class 4: Bivariate and Multivariate analysis

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

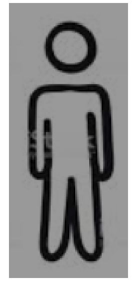
Campus d'Excel·lència Internacional

ICTP SAIFR International Centre for Theoretical Physics South American Institute for Fundamental Research

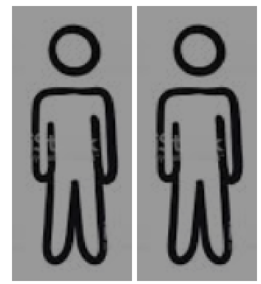
cristina.masoller@upc.edu @cristinamasoll1

Methods of time series analysis are classified as:

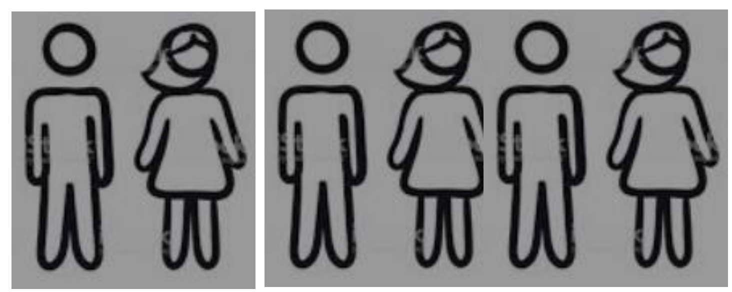
- Univariate analysis



- Bivariate analysis
  - Cross Correlation
  - Mutual Information
  - Event synchronization
  - Causality



- Multivariate analysis



cristina.masoller@upc.edu @cristinamasoll1

Cross-correlation

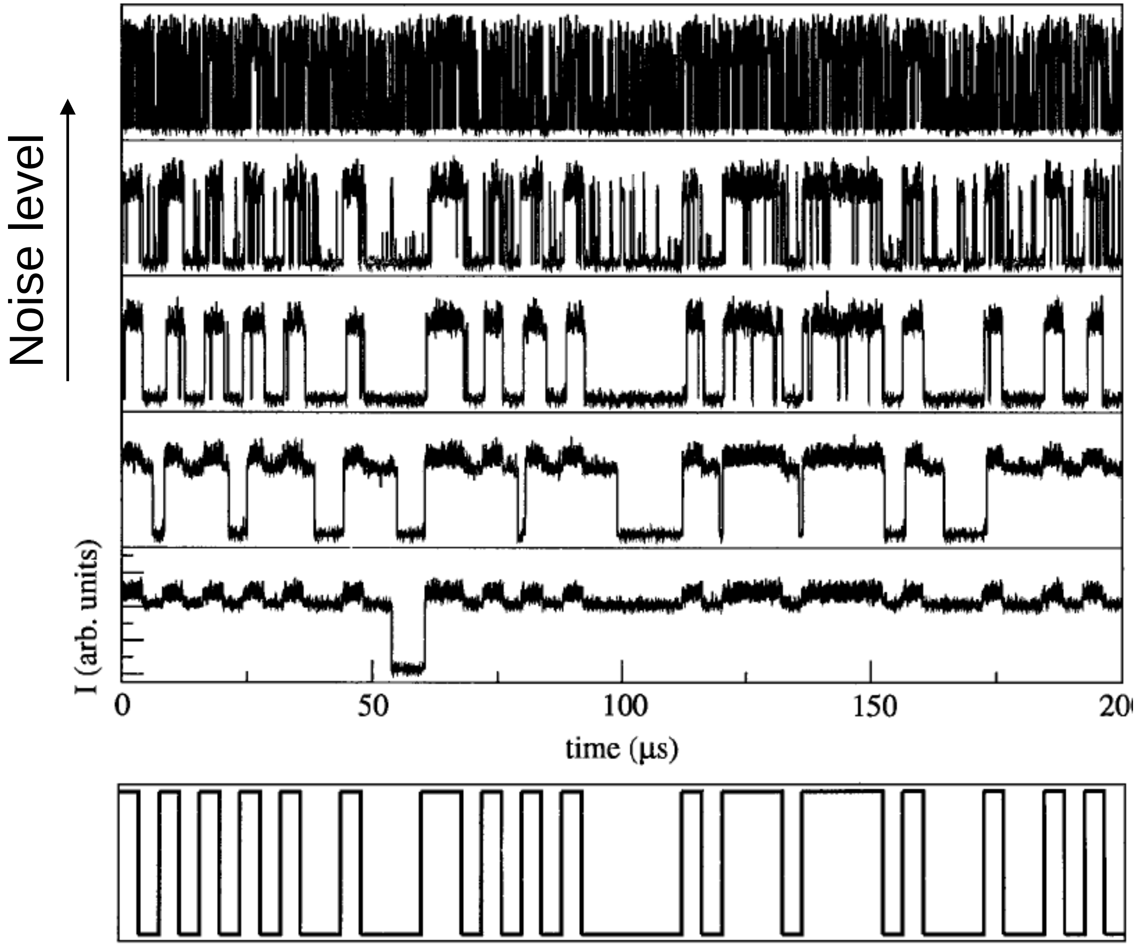
$$C_{xy}(\tau) = \frac{1}{N-\tau} \sum_{k=1}^{N-\tau} x(k+\tau) y(k)$$

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

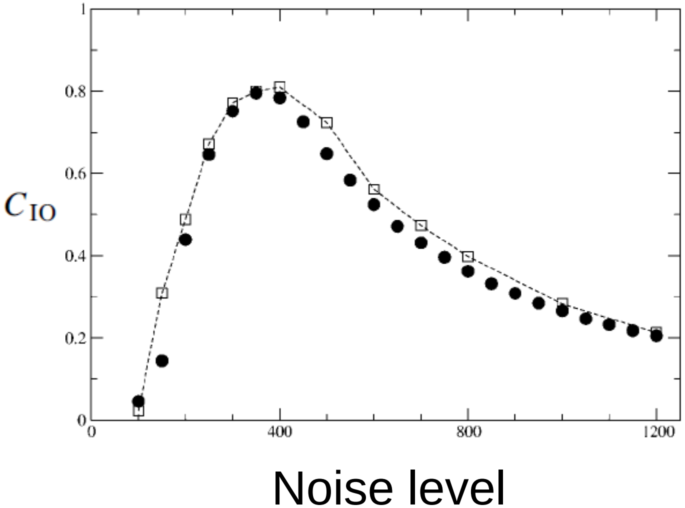
- $-1 \leq C_{X,Y} \leq 1$
- $C_{X,Y} = C_{Y,X}$
- The maximum of $C_{X,Y}(\tau)$ indicates the **lag** that renders the time series *X* and *Y* best aligned.
- Pearson coefficient: $\rho = |C_{X,Y}(0)|$

cristina.masoller@upc.edu @cristinamasoll1

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



Cross-correlation between input and output signal.



Barbay et al, PRL 85, 4652 (2000)

cristina.masoller@upc.edu @cristinamasoll1

Lags need to be taken into account when analyzing statistical similarities

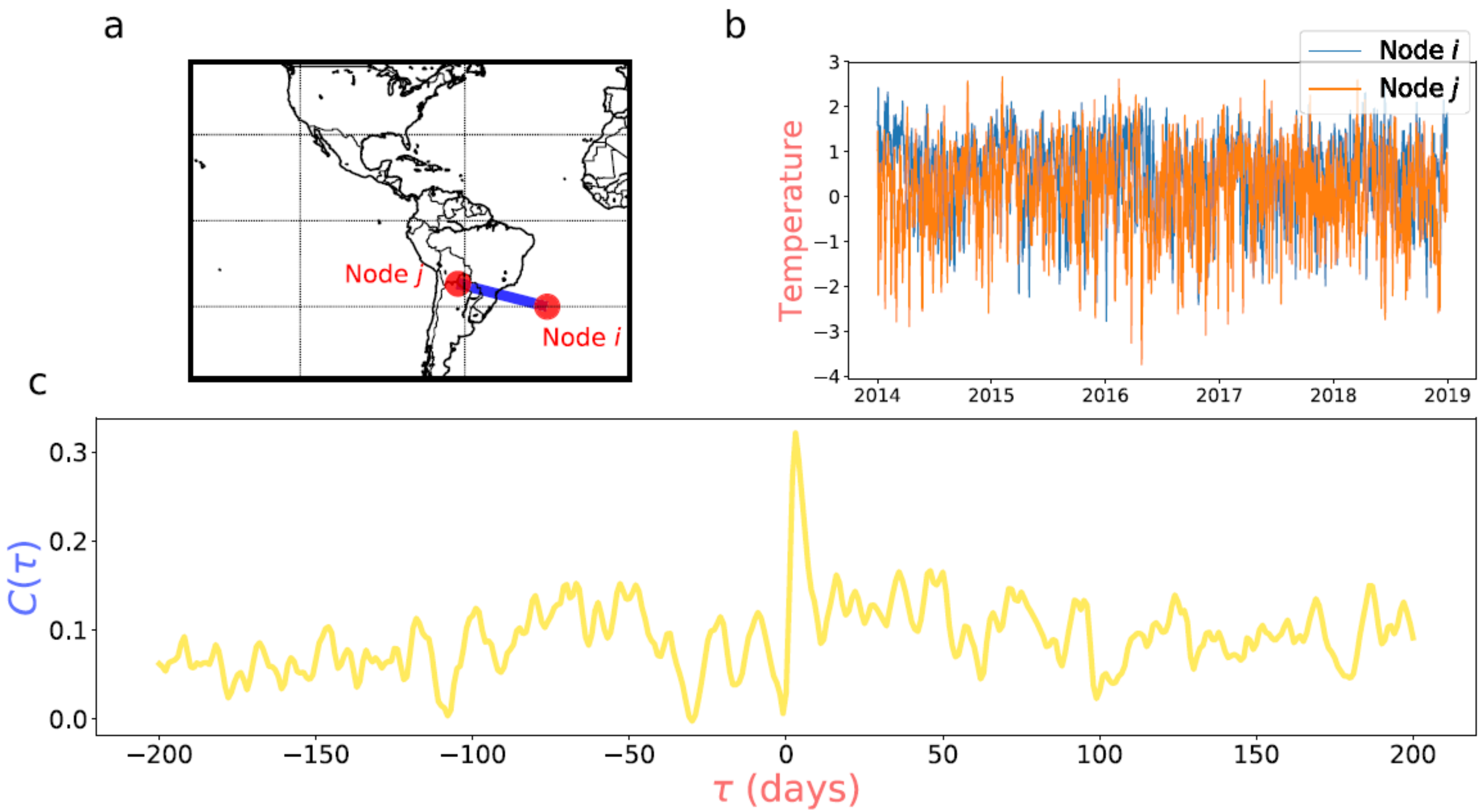
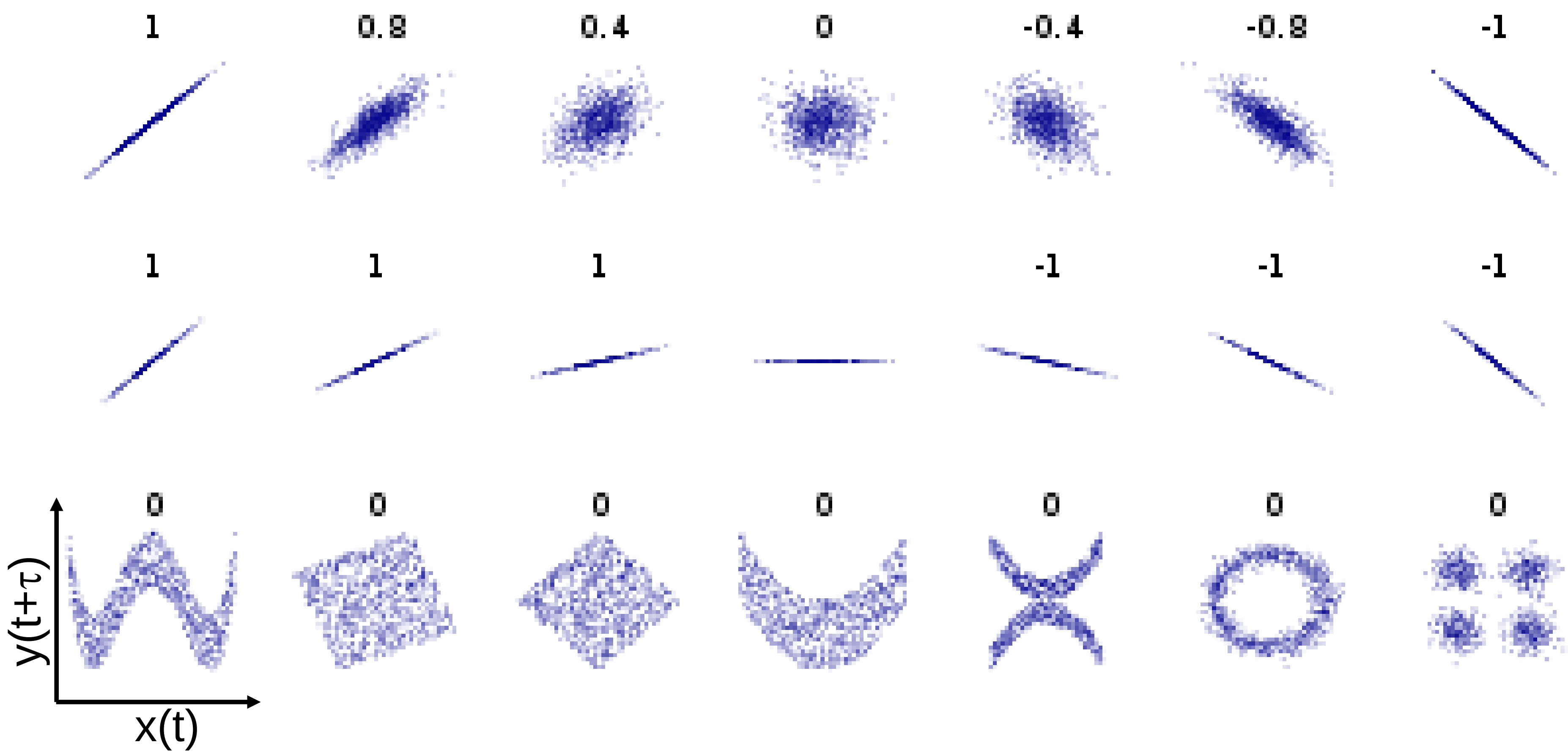


Fig. 6. Typical weighted and directed link in a Pearson Correlation Climate Network. (a) Node i is located on the Southwest Atlantic and node j is in the South American continent. (b) The near surface daily air temperature anomalies for the period [2014,2018]. (c) The cross-correlation function between the time series shown in (b). The direction of this link is from j to i with weight $W_{i,j}^{+} = 5.71$.

J. Fan et al. Phys. Rep. 896, 1 (2021).

cristina.masoller@upc.edu @cristinamasoll1

Cross-correlation detects linear relationships only

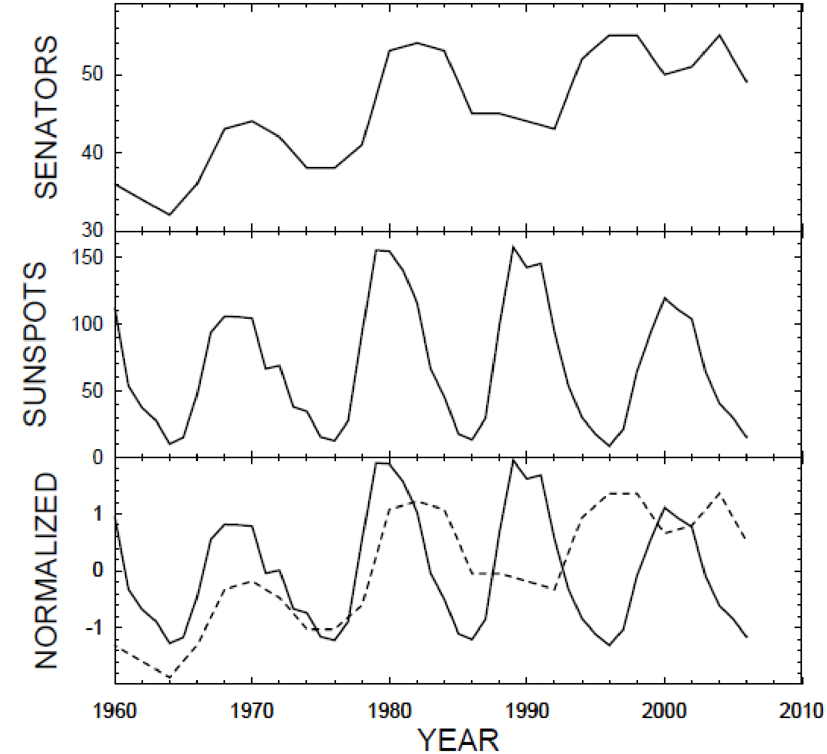


Source: wikipedia

cristina.masoller@upc.edu @cristinamasoll1

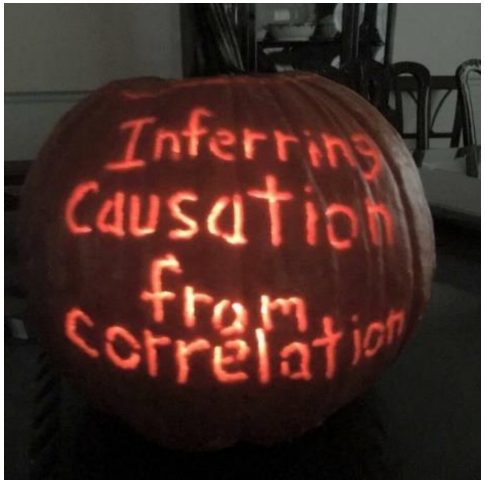
Correlation is NOT causality

An illustrative example: the number of sunspots and the number of the Republicans in the U.S. Senate in the years 1960 to 1986 (biannual sampling, 14 points).



C=0.52

Appropriate significance test needed!



M. Palus, Contemporary Physics 48, 307 (2007).

http://tylervigen.com/spurious-correlations

cristina.masoller@upc.edu @cristinamasoll1

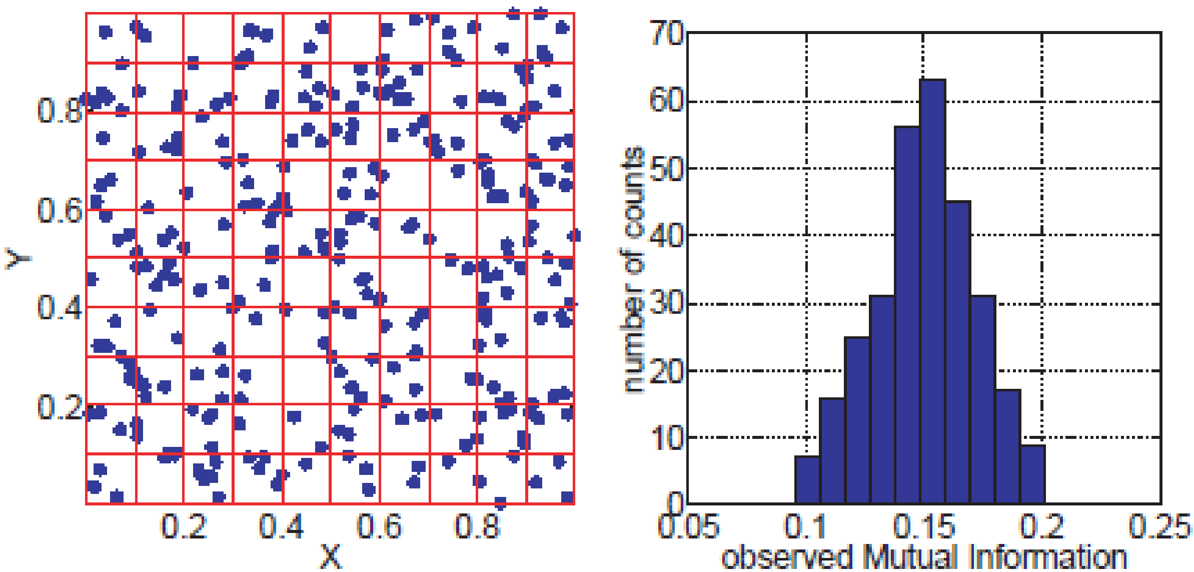
The Mutual Information: a nonlinear correlation measure

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

- *MI (x,y) = MI (y,x)*
- *p(x,y) = p(x) p(y)* $\Rightarrow$ *MI = 0*, else ***MI >0***
- *MI* can also be computed with a lag-time.
- *MI* can also be computed from symbolic probabilities (e.g., probabilities of ordinal patterns).

cristina.masoller@upc.edu @cristinamasoll1

MI values are systematically overestimated



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

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

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.

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

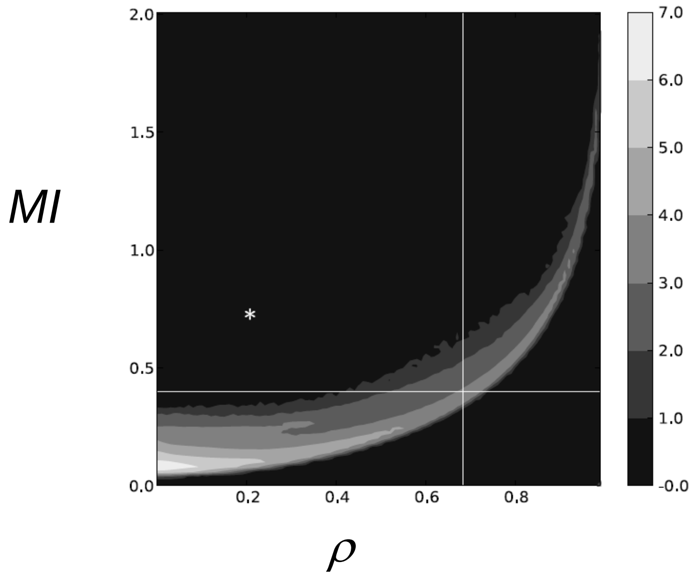
cristina.masoller@upc.edu @cristinamasoll1

Relation between cross-correlation and mutual information

- Depends on the data.
- If the two processes are Gaussian, the MI and the Pearson correlation coefficient are related as

$$MI = -1/2 \log(1-\rho^2).$$

- Example: values computed from 6816 x 6816 surface air temperature (SAT) anomaly series:

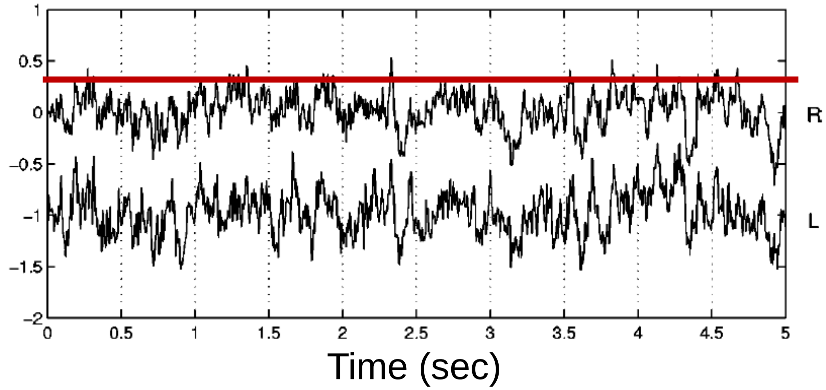


2D histogram; the color represents the number of elements in each bin in log scale.

Donges et al, Eur. Phys. J. Special Topics 174, 157 (2009).

cristina.masoller@upc.edu @cristinamasoll1

How to find “synchronized events” in two time series?



Rat EEG signals from right and left cortical intracranial electrodes. For a better visualization, the left signal is plotted with an offset.

- Define “events” in each time series.
- Count $c^{\tau}(x|y)$ = number of times an event appears in x shortly after an event appears in y. Analogous for $c^{\tau}(y|x)$.
- Calculate:

$$Q_{\tau} = \frac{c^{\tau}(y|x) + c^{\tau}(x|y)}{\sqrt{m_x m_y}} \qquad q_{\tau} = \frac{c^{\tau}(y|x) - c^{\tau}(x|y)}{\sqrt{m_x m_y}}$$

m_x, m_y are the number of events in each time series.

- $Q_{\tau} = 1$: the events of the signals are fully synchronized.
- $q_{\tau} = 1$: the events in x always precede those in y.

Quian Quiroga et al, PRE 66, 041904 (2002).

cristina.masoller@upc.edu @cristinamasoll1

Example

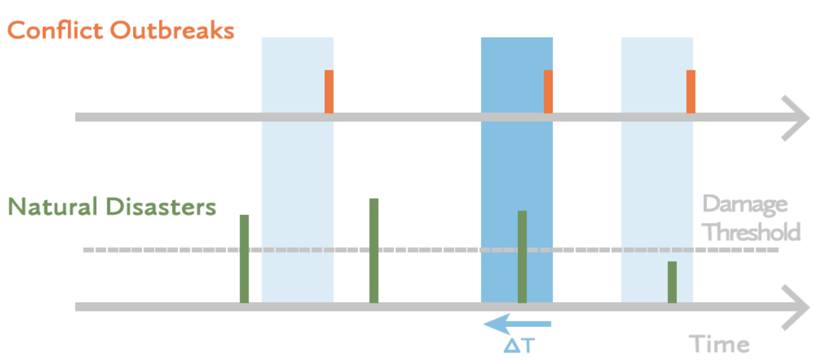
Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries

Carl-Friedrich Schleussner[a,b,c,1], Jonathan F. Donges[a,d], Reik V. Donner[a], and Hans Joachim Schellnhuber[a,e,1]

[a]Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany; [b]Climate Analytics, 10969 Berlin, Germany; [c]Integrative Research Institute on Transformations of Human–Environment Systems, Humboldt University, 10099 Berlin, Germany; [d]Stockholm Resilience Centre, Stockholm University, 114 19 Stockholm, Sweden; and [e]Santa Fe Institute, Santa Fe, NM 87501

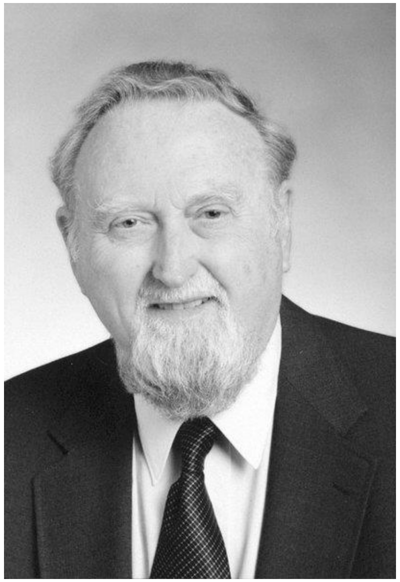
9216–9221 | PNAS | August 16, 2016 | vol. 113 | no. 33

Social and political tensions keep on fueling armed conflicts around the world. Although each conflict is the result of an individual context-specific mixture of interconnected factors, ethnicity appears to play a prominent and almost ubiquitous role in many of them. This overall state of affairs is likely to be exacerbated by anthropogenic climate change and in particular climate-related natural disasters. Ethnic divides might serve as predetermined conflict lines in case of rapidly emerging societal tensions arising from disruptive events like natural disasters. Here, we hypothesize that climate-related disaster occurrence enhances armed-conflict outbreak risk in ethnically fractionalized countries. Using event coincidence analysis, we test this hypothesis based on data on armed-conflict outbreaks and climate-related natural disasters for the period 1980–2010. Globally, we find a coincidence rate of 9% regarding armed-conflict outbreak and disaster occurrence such as heat waves or droughts. Our anal-



cristina.masoller@upc.edu @cristinamasoll1

Granger Causality



Hypothesis: X_1 and X_2 can be described by stationary autoregressive linear models.

past of X_1 — Residual error

$$X_1(t) = \sum_{j=1}^{p} A_{11,j}\, X_1(t-j) + E_1(t)$$

past of X_1 — past of X_2 — Residual error

$$X_1(t) = \sum_{j=1}^{p} A_{11,j}\, X_1(t-j) + \sum_{j=1}^{p} A_{12,j} X_2(t-j) + E'_1(t)$$

If $\langle E'_1(t)\rangle < \langle E_1(t)\rangle$ ➡ $X_2 \rightarrow X_1$

C. W. J. Granger *Investigating causal relations by econometric models and cross-spectral methods*. *Econometrica* 37, 424–438 (1969) **(> 10000 citations)**

cristina.masoller@upc.edu @cristinamasoll1

Transfer Entropy (TE) and Directionality Index (DI)

- TE: is the *Conditional* Mutual Information, given the "past" of one of the variables.

$$TE\,(x,y) = MI\,(x, y|x_\tau)$$

$$TE\,(y,x) = MI\,(y, x|y_\tau)$$

- MI (x,y) = MI (y,x) but TE (x,y) $\neq$ TE(y,x)

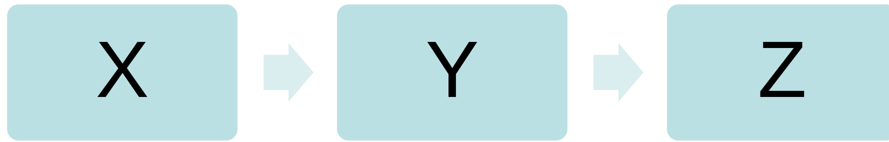
- Directionality Index: TE(x,y)-TE(y,x)
- TE and GC are equivalent for Gaussian processes.
- TE can be computed from the probabilities of symbols (symbolic TE).

T. Schreiber, Measuring information transfer, Phys. Rev. Lett. 85, 461 (2000).

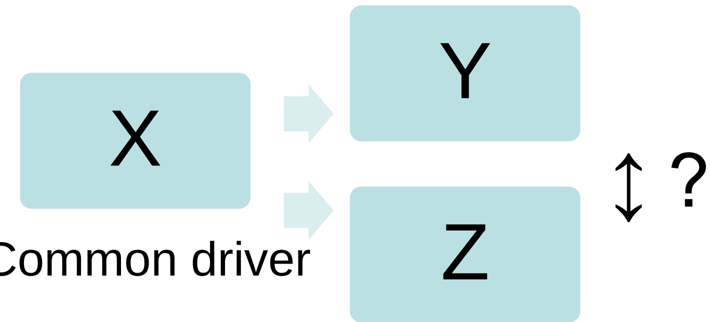
K. Hlaváčková-Schindler et al. / Physics Reports 441 (2007) 1–46

cristina.masoller@upc.edu @cristinamasoll1

Problems:



Indirect link X→Z?



Common driver

In addition: Transfer Entropy is computationally demanding.

A “simple” solution: to use the expression that is valid for Gaussian processes [$MI = -1/2 \log(1-\rho^2)$]

Does this work? Check it out:

R. Silini, C. Masoller “*Fast and effective pseudo transfer entropy for bivariate data-driven causal inference*”, Sci. Rep. 11, 8423 (2021).

https://doi.org/10.1038/s41598-021-87818-3

cristina.masoller@upc.edu @cristinamasoll1

Data Generating Processes and significance analysis

DGPs: We know whether X and Y are independent or not.

	Model
Y X	M0, M1, M2
Y → X	M3, M4, M5, M6, M7, M8, M9, M10, M11, M12
Y ⇆ X	M13, M14

Y X (M0–M2):

$$x_t = (0.01 + 0.5\, x_{t-1}^2)^{0.5} + E_{1t} \qquad y_t = 0.5\, y_{t-1} + E_{2t}$$

Y → X (M3–M12):

$$x_t = 0.6\, x_{t-1} + 0.5\, y_{t-1} + E_{1t} \qquad y_t = 0.5\, y_{t-1} + E_{2t}$$

Y ⇆ X (M13–M14):

$$x_t = 0.15\, x_{t-1} + 0.7\, y_{t-1} + E_{1t}$$

$$y_t = 0.1\, y_{t-1} + 0.8\, x_{t-1} + E_{2t}$$

Significance analysis: time-shifted surrogates (cheap for causality testing)

Quiroga et al., Phys. Rev. E 65, 041903 (2002).

cristina.masoller@upc.edu @cristinamasoll1

Results

Power: there is causality and we find causality (True Positives)
Size: there is no causality but we find causality (False Positives)

	Model	pTE	
		Y → X	X → Y
Y X	M0	3.8	3.9
	M1	2.3	2.6
	M2	4.2	4.7
Y → X	M3	100	4.5
	M4	80.7	3.8
	M5	100	2.2
	M6	100	1.8
	M7	100	2.8
	M8	100	4.5
	M9	100	0.1
	M10	62.6	3.1
	M11	46.1	43.1
	M12	99.9	1.0
Y ⇆ X	M13	100	100
	M14	100	100

cristina.masoller@upc.edu @cristinamasoll1

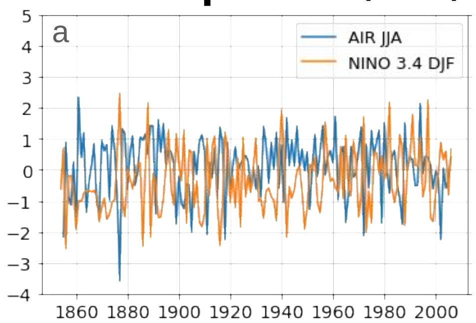
Comparison with Granger Causality and Transfer Entropy

	Model	pTE		GC		TE		DI		
		Y → X	X → Y	Y → X	X → Y	Y → X	X → Y	pTE	GC	TE
Y X	M0	3.8	3.9	5.1	5.0	4.4	4.4	−0.01	0.01	0.00
	M1	2.3	2.6	3.3	3.1	100	100	−0.06	0.03	0.00
	M2	4.2	4.7	5.5	5.9	4.7	4.9	−0.06	−0.04	−0.02
Y → X	M3	100	4.5	100	4.8	70.2	5.6	0.91	0.91	0.85
	M4	80.7	3.8	84.2	4.9	96.0	4.7	0.91	0.89	0.91
	M5	100	2.2	100	3.1	100	3.8	0.96	0.94	0.93
	M6	100	1.8	100	2.8	100	4.3	0.96	0.95	0.92
	M7	100	2.8	100	3.4	100	4.0	0.95	0.93	0.92
	M8	100	4.5	100	5.6	100	100	0.91	0.89	0.00
	M9	100	0.1	100	0.1	100	100	1.00	1.00	0.00
	M10	62.6	3.1	67.3	4.3	12.2	4.5	0.91	0.88	0.46
	M11	46.1	43.1	53.1	49.8	37.8	45.0	0.03	0.03	−0.09
	M12	99.9	1.0	100	0.9	100	0	1.0	1.0	1.0
Y ⇆ X	M13	100	100	100	100	100	100	0.00	0.00	0.00
	M14	100	100	100	100	100	100	0.00	0.00	0.00

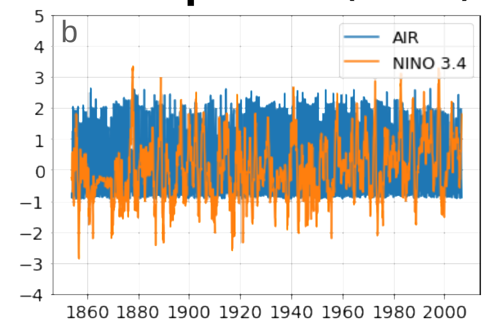
cristina.masoller@upc.edu @cristinamasoll1

Application to real data NINO3.4 ← → All India Rainfall

Yearly sampled (152)



Monthly sampled (1836)



pTE

NINO3.4 ⟶ AIR — 0.04 s

NINO3.4 ⟵ AIR — 0.5 s

GC

NINO3.4 ⟶ AIR — 0.4 s

NINO3.4 ⟵ AIR — 0.9 s

TE

NINO3.4 ⇆ AIR — 1 s

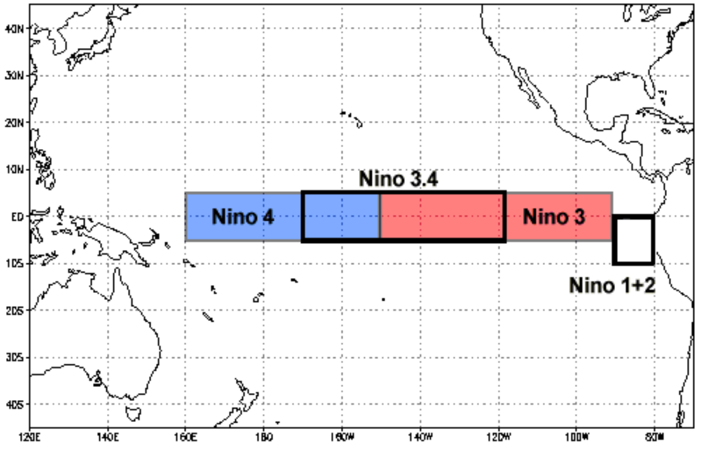
NINO3.4 ⇆ AIR — 68 s

cristina.masoller@upc.edu @cristinamasoll1

How much time we save by using the "pseudo Transfer Entropy"?

For two time-series of 500 data points (1 data point per month, 40 years):

TE:**112 ms** but pTE: **4 ms**



8000 grid points (high resolution)
$\Rightarrow$ 64 x 10^6 pairs

$\Rightarrow$ **829 days** (TE) vs. **29 days** (pTE).

(without "surrogate" analysis)

But, there is a price to pay, no "free lunch".

https://github.com/riccardosilini/pTE

cristina.masoller@upc.edu @cristinamasoll1

Besides Granger causality and transfer entropy, many methods have proposed (and extensively used in social, economic and financial data).

- Symbolic Transfer Entropy
- Partial Correlation
- Partial Directed Coherence
- Cross Mapping
- Partial Cross Mapping
- Etc.

Read more: A. Krakovska et al., *Comparison of six methods for the detection of causality in a bivariate time series*, Phys. Rev. E 97, 042207 (2018)

cristina.masoller@upc.edu @cristinamasoll1

Example

RESEARCH ARTICLE

INFORMATION SCIENCE

10.1126/sciadv.1501158

The dynamics of information-driven coordination phenomena: A transfer entropy analysis

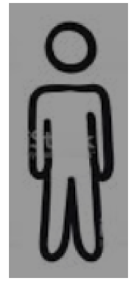
Javier Borge-Holthoefer,[1*†] Nicola Perra,[2*] Bruno Gonçalves,[3‡] Sandra González-Bailón,[4] Alex Arenas,[5*] Yamir Moreno,[6,7,8*] Alessandro Vespignani[2,8,9*]

Data from social media provide unprecedented opportunities to investigate the processes that govern the dynamics of collective social phenomena. We consider an information theoretical approach to define and measure the temporal and structural signatures typical of collective social events as they arise and gain prominence. We use the symbolic transfer entropy analysis of microblogging time series to extract directed networks of influence among geolocalized subunits in social systems. This methodology captures the emergence of system-level dynamics close to the onset of socially relevant collective phenomena. The framework is validated against a detailed empirical analysis of five case studies. In particular, we identify a change in the characteristic time scale of the information transfer that flags the onset of information-driven collective phenomena. Furthermore, our approach identifies an order-disorder transition in the directed network of in-

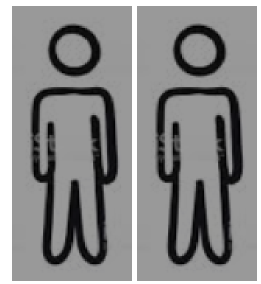
Borge-Holthoefer et al. Sci. Adv. 2, e1501158 (2016).

Methods of time series analysis are classified as:

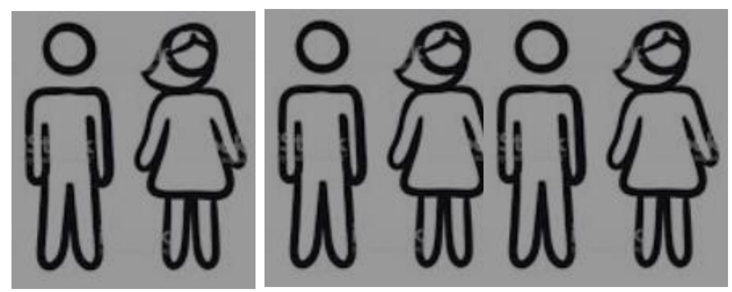
- Univariate analysis



- Bivariate analysis



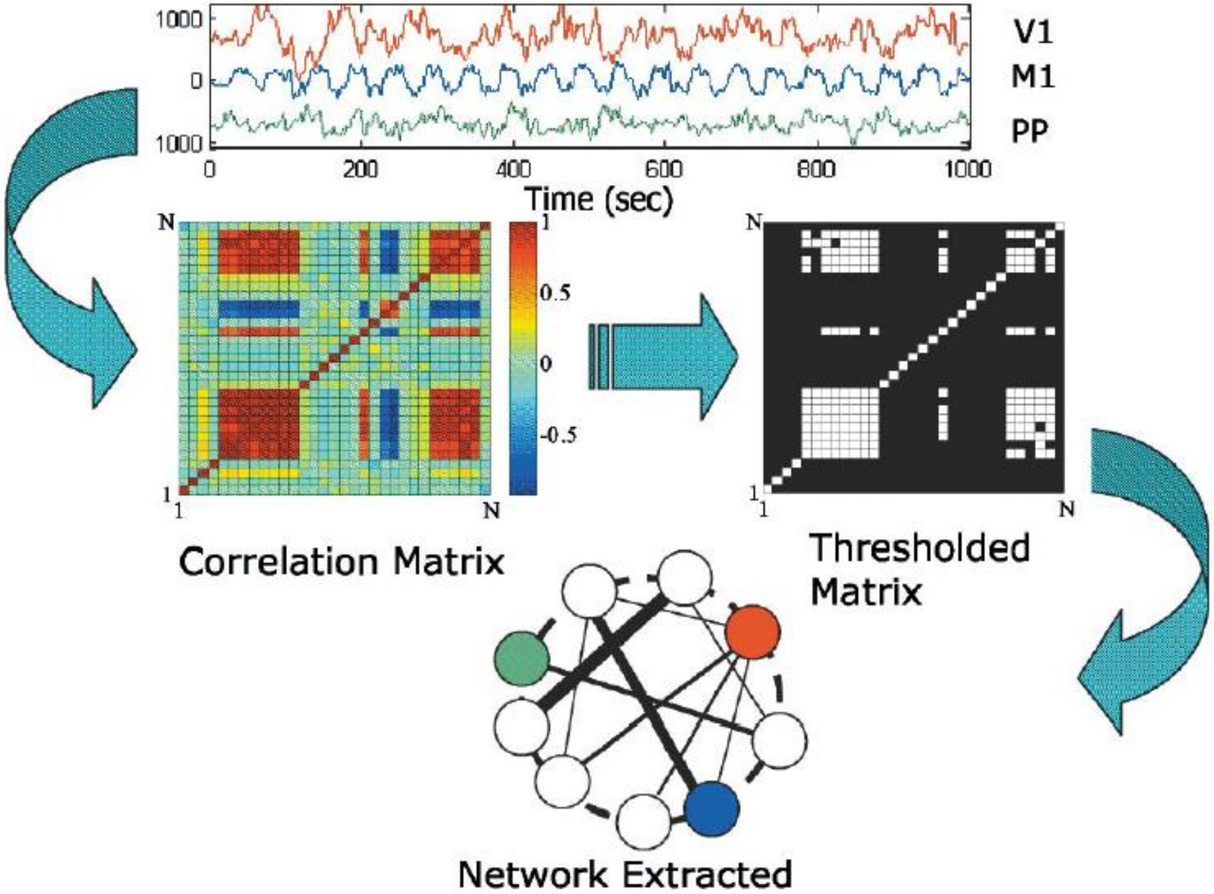
- Multivariate analysis
  - Functional networks
  - Network inference

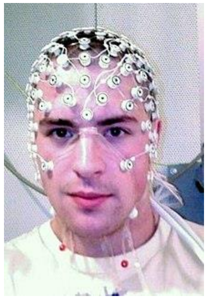


cristina.masoller@upc.edu @cristinamasoll1

“Functional networks” are obtained by using bivariate correlation or causality measures

Example: brain functional network





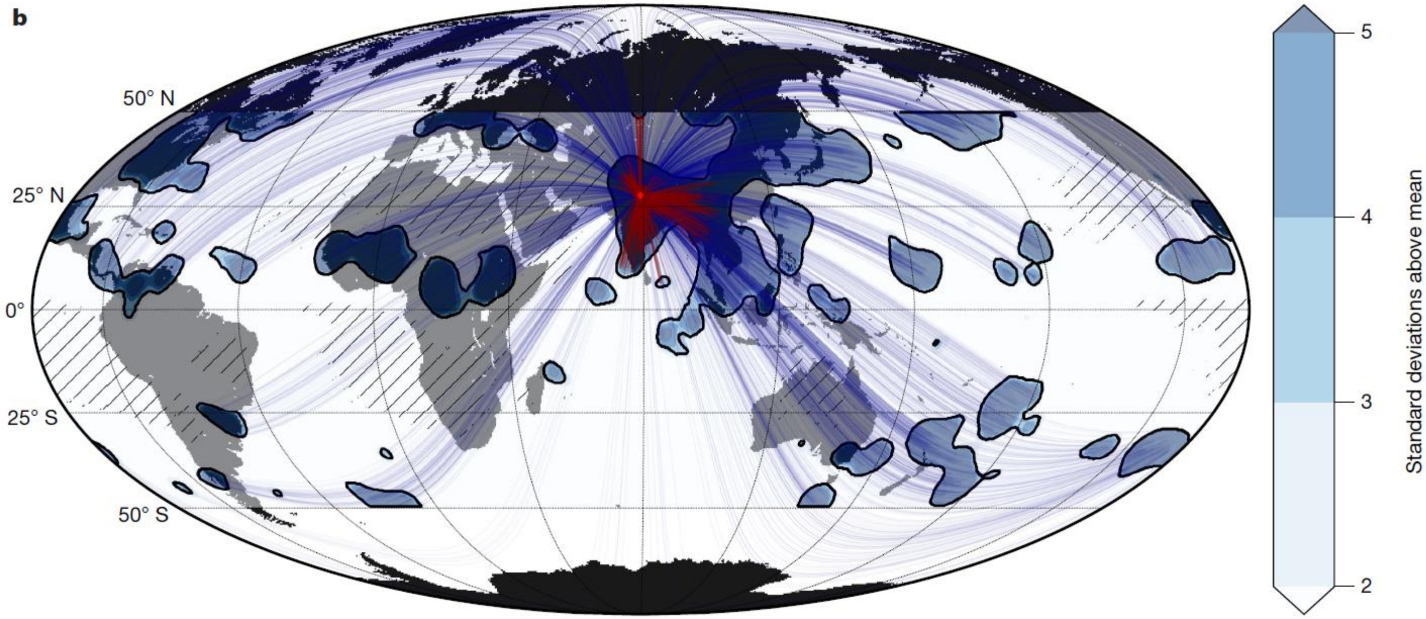
The adjacency matrix is obtained by “thresholding”

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

V. M. Eguiluz et al, Phys. Rev. Lett. 94, 018102 (2005).

cristina.masoller@upc.edu @cristinamasoll1

Example: functional network constructed from extreme rainfall events, using the "event synchronization" measure



Teleconnection pattern for south-central Asia for events above the 95th percentile (with a maximum delay of ten days). Links shorter (longer) than 2,500 km are shown in red (blue).

N. Boers et al, "Complex networks reveal global pattern of extreme-rainfall teleconnections", Nature 566, 373 (2019).

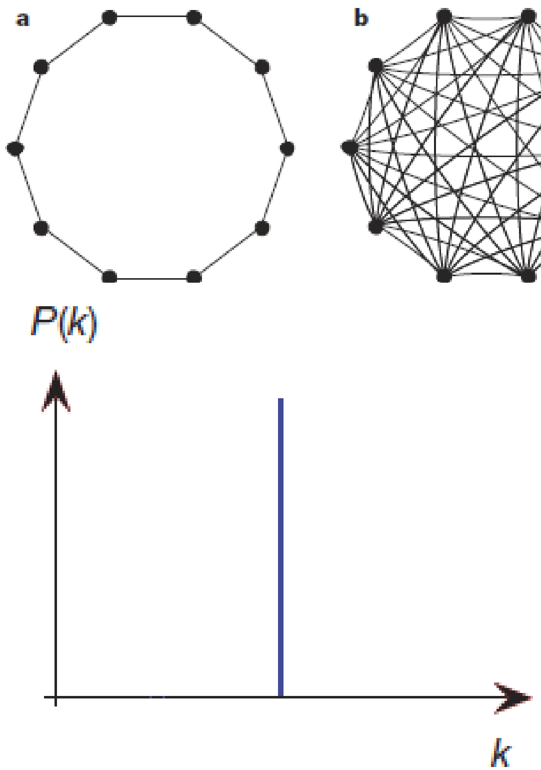
cristina.masoller@upc.edu @cristinamasoll1

How to characterize the graph?

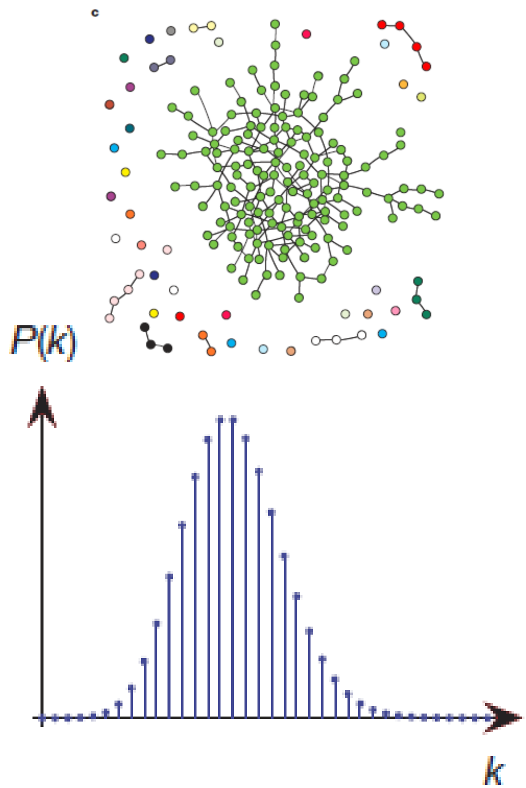
Begin with the degree distribution

Degree of node i: $k_i = \Sigma_j A_{ij}$

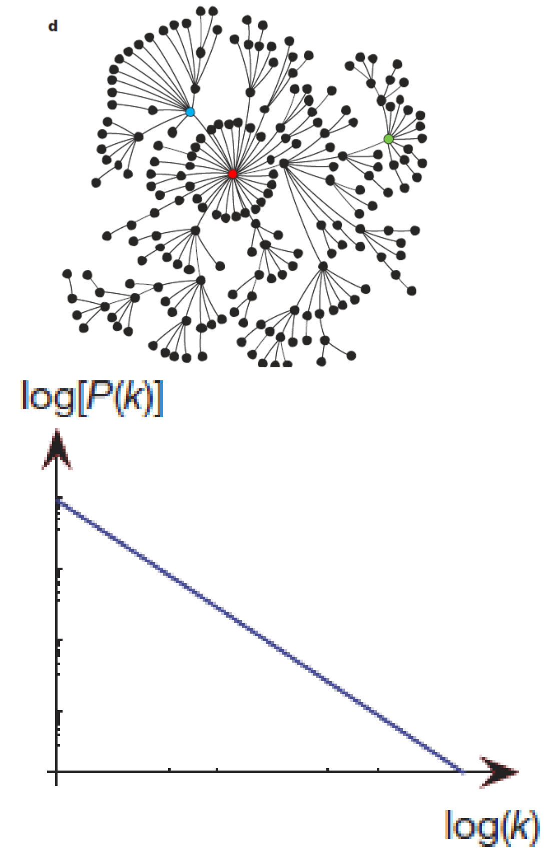
Regular



Random



Scale-free



S. H. Strogatz, Nature 410, 268 (2001).

cristina.masoller@upc.edu @cristinamasoll1

How to characterize the degree distribution?

- **Mean** degree: μ

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

- **Skewness**: $Z = \dfrac{X-\mu}{\sigma}$ $\qquad S = \mathrm{E}[Z^3]$

- **Kurtosis**: measures the "tailedness“ of the distribution. Are there “hubs”, i.e., nodes with lots of links?

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

cristina.masoller@upc.edu @cristinamasoll1

How to compare two distributions?

Distance between two distributions P and P_e

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

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

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

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

cristina.masoller@upc.edu @cristinamasoll1

Other distributions that characterize a network
(unweighted and undirected)

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

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

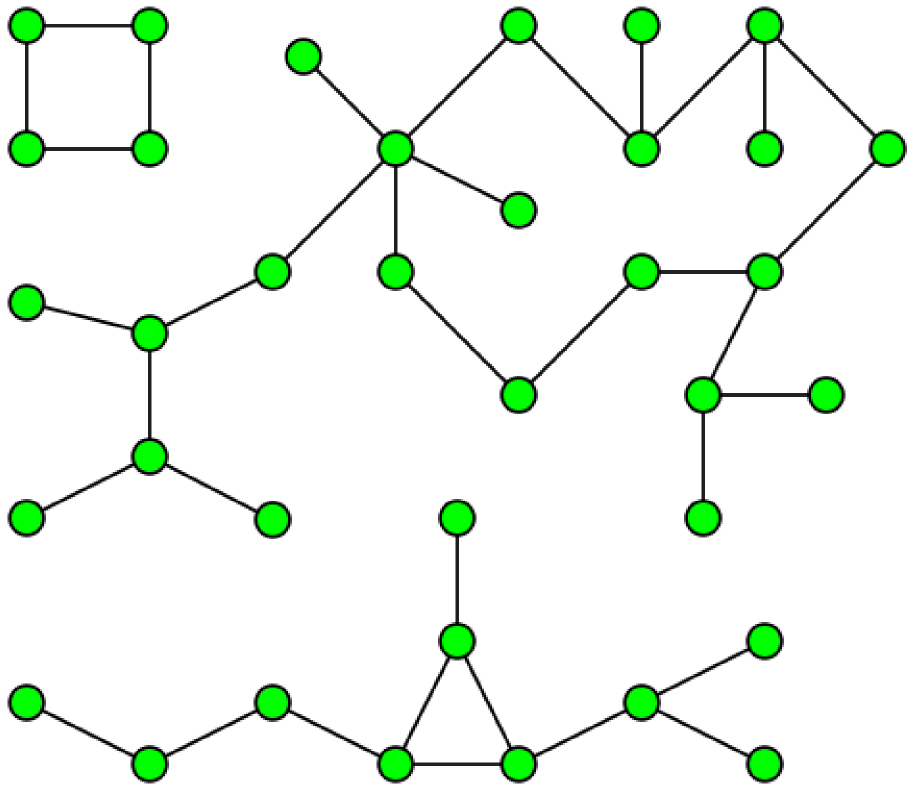
R_i is the number of connected pairs in the set of neighbors of node *i*

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

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

cristina.masoller@upc.edu @cristinamasoll1

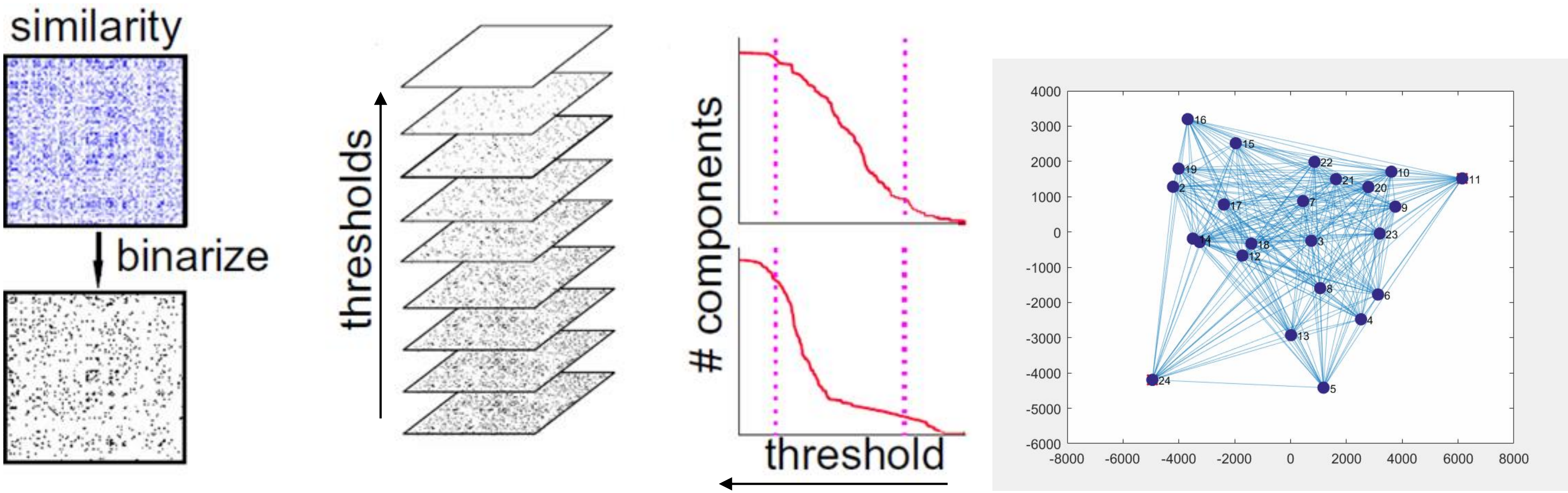
Number of connected components ("communities")



A graph with three connected components.
Source: Wikipedia

cristina.masoller@upc.edu @cristinamasoll1

Problems with networks obtained by thresholding



- The variation of the number of connected components with the threshold reveals different structures.
- Thresholding near the dotted lines would suggest inaccurately that these two networks have similar structures.
- Persistent “features” are "true" features.

Giusti et al., J. Comput. Neurosci. 4, 1 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Network inference: How to reconstruct the network from observations?

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

- How to select the "optimal" threshold?

- How to keep **weak-but-significant** links?

- A classification problem:
  - the interaction exists (is significant)
  - the interaction does not exists (or is not significant)

cristina.masoller@upc.edu @cristinamasoll1

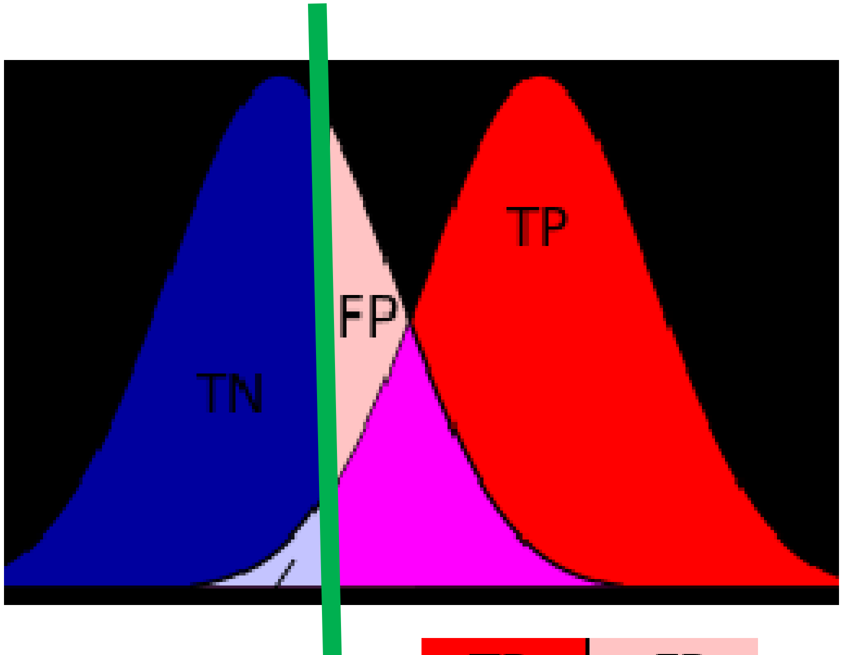
Confusion matrix

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

- ***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 or Recall): 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***: 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**

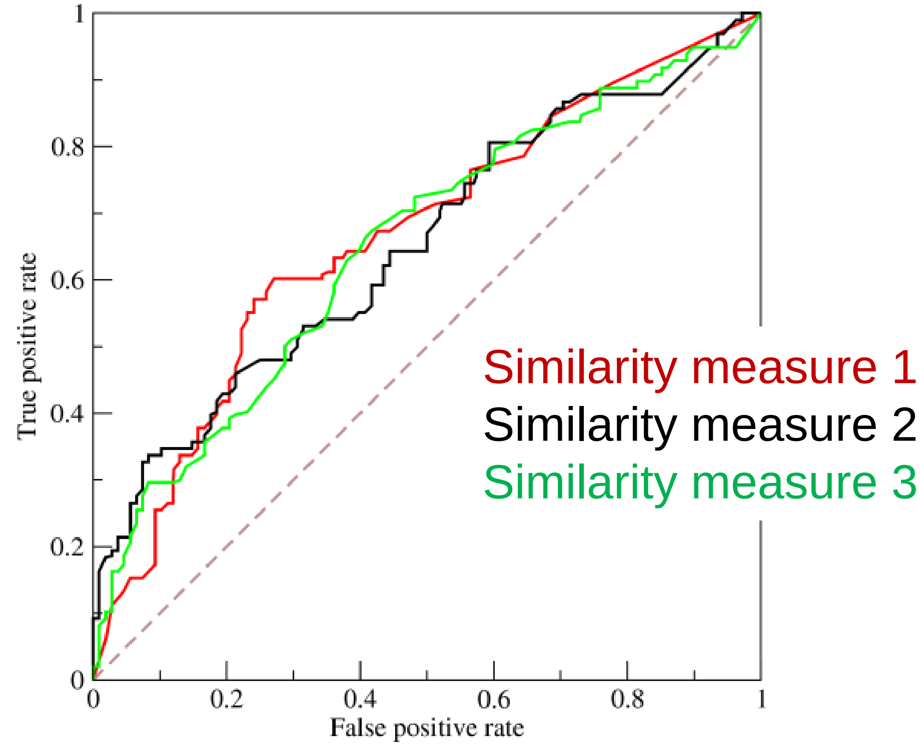
cristina.masoller@upc.edu @cristinamasoll1

Receiver operating characteristic (ROC curve) and Precision-Recall (PR curve)



Source: *Wikipedia*

TP	FP
FN	TN



For reconstructing sparse networks (with a small % of links) the "Precision-Recall" curve is more informative because it does not depend on the # of true negatives.

Precision =TP / predicted yes (TP+FP)
Recall = TP / actual yes (TP+FN)

cristina.masoller@upc.edu @cristinamasoll1

How to compare the performance of different statistical similarity measures for inferring interactions from data?

- Use a “toy model” where **we know** the “ground truth”, i.e., we know the underlying equations and interactions and so we can check the performance of the different measures in inferring the interactions.

- Problem: results will depend on the “toy model” used as the performance of the statistical similarity measure depends on the characteristics of the data.

cristina.masoller@upc.edu @cristinamasoll1

Kuramoto oscillators in a random network

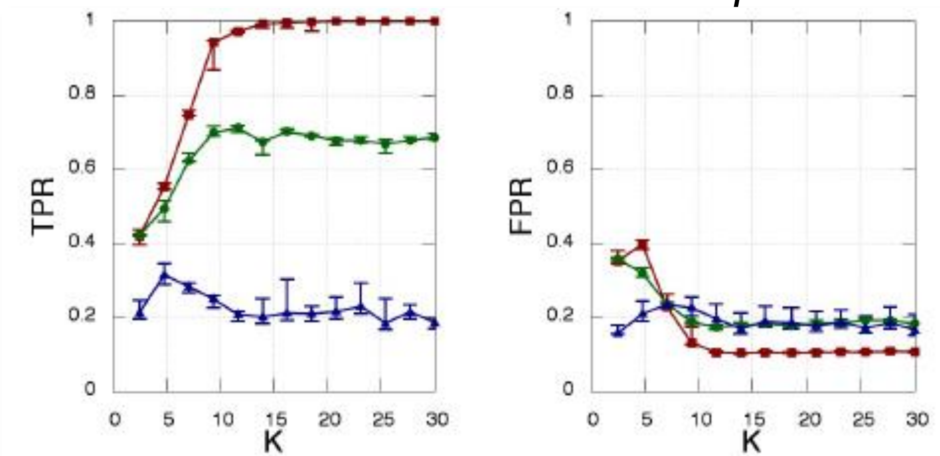
$$d\theta_i = \omega_i dt + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_j - \theta_i)dt + D\, dW_t^i$$

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

CC MI MIOP

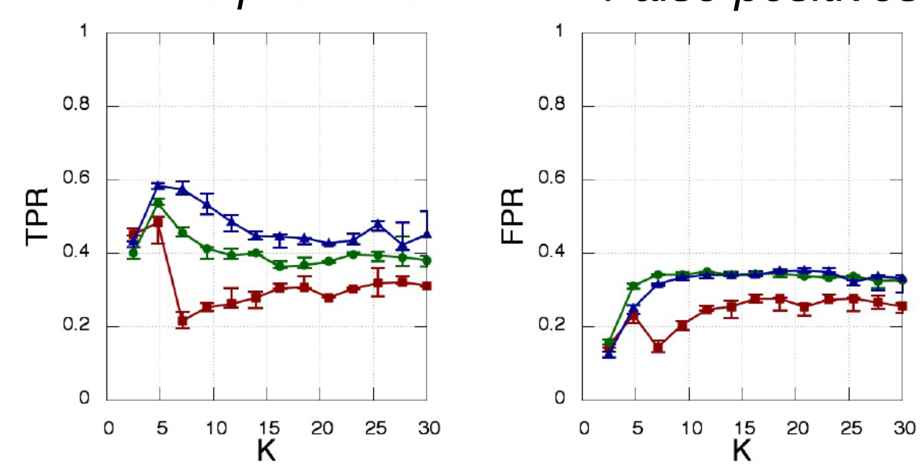
Phases (θ)

True positives *False positives*



"Observable" Y=sin(θ)

True positives *False positives*



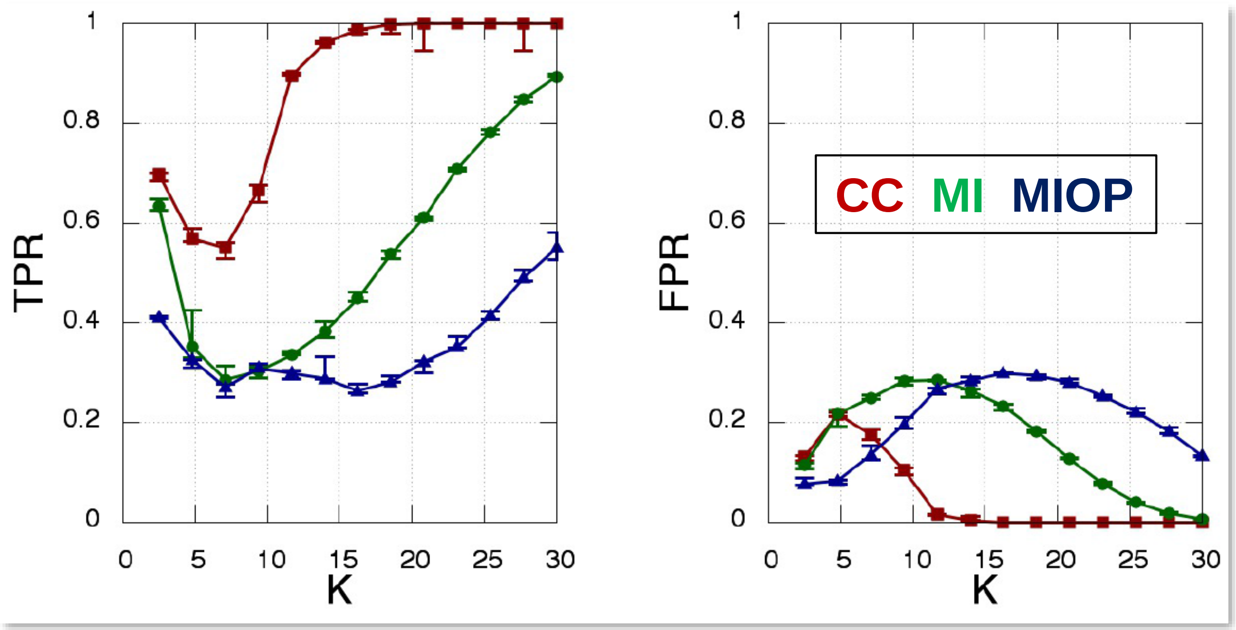
Results of a 100 simulations with different oscillators' frequencies, random matrices, noise realizations and initial conditions.

For each K, the threshold was varied to obtain optimal reconstruction.

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

cristina.masoller@upc.edu @cristinamasoll1

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).

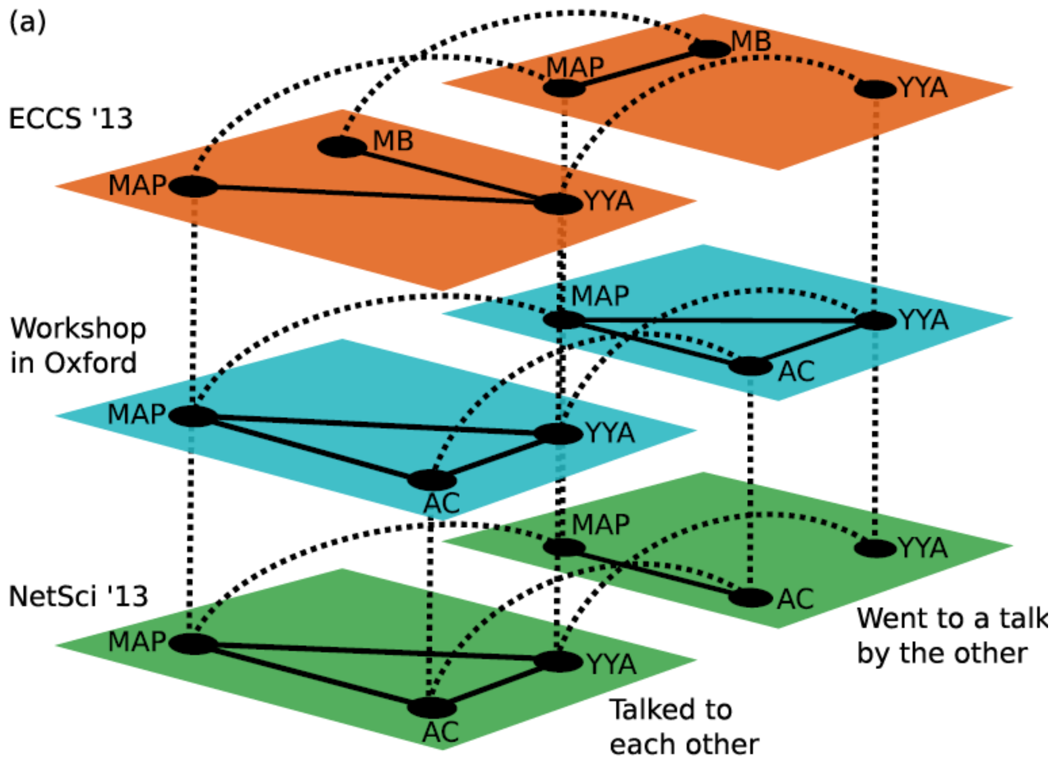
cristina.masoller@upc.edu @cristinamasoll1

Complicated structures (such as multilayer networks), require of advanced methods for characterization / reconstruction.

Facebook

Twitter

Linkedin



Kivela et al, J. Complex Netw. 2, 203 (2014).

cristina.masoller@upc.edu @cristinamasoll1

Multilayer networks in social systems and economy

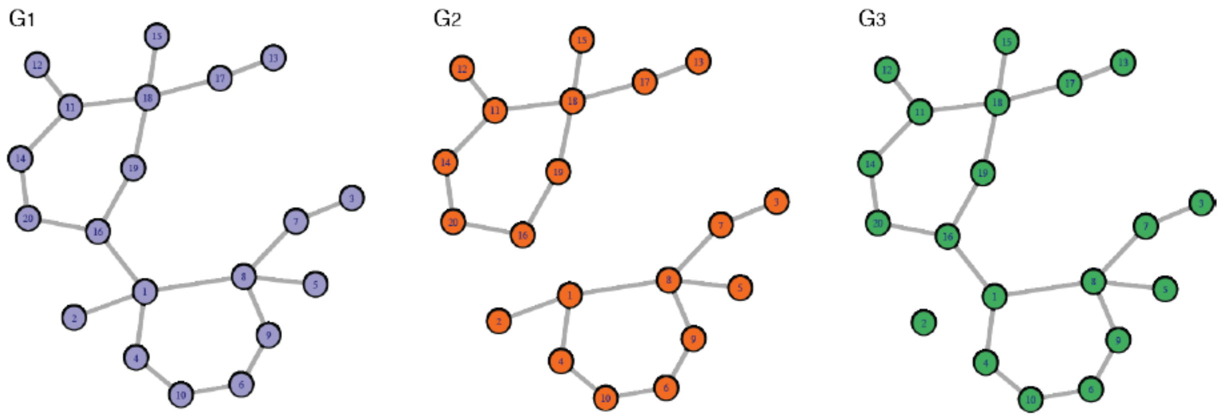
Resume of topics and references		
Field	Topic	References
Social	Online communities	Pardus: [63, 421–424]
		Netflix: [425, 426]
		Flickr: [66, 88, 427]
		Facebook: [68, 428–430]
		Youtube: [431]
		Other online communities: [54, 89, 432]
		Merging multiple communities: [122, 123, 433, 434]
	Internet	[109, 110, 435]
	Citation networks	DBLP: [31, 33, 436–441]
	Other social networks	Scottish Community Alliance: [442]
		Politics: [68, 443]
		Terrorism: [23]
		Bible: [444]
		Mobile communication: [445]
Economy	Trade networks	International Trade Network: [70, 71, 450]
		Maritime flows: [451]
	Interbank market	[452]
	Organizational networks	[453–455]

Boccaletti et al., Phys. Reps. 544, 1 (2014).

cristina.masoller@upc.edu @cristinamasoll1

How similar are two graphs?

- Hamming distance $$d_{\text{Hamming}}(y_1, y_2) = \sum_{i \neq j}^{N} \left[A_{ij}^{(1)} \neq A_{ij}^{(2)} \right]$$
- Can be used to compare two graphs of the same size.
- Main problem: not all the links have the same importance.

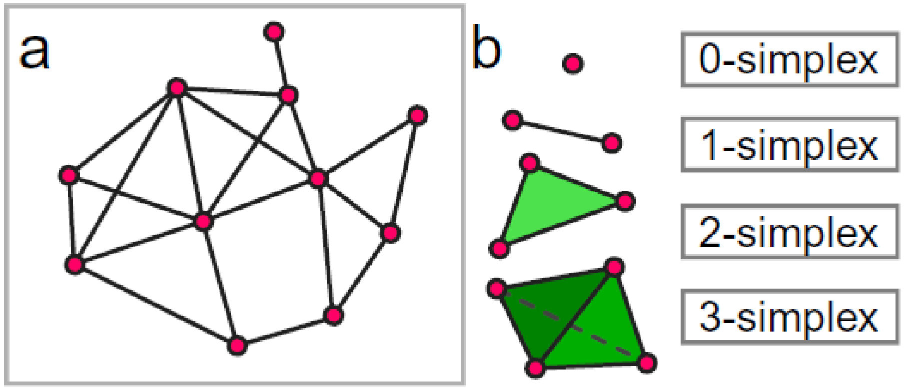


- A “dissimilarity” measure that can be used to compare graphs with different sizes, based in the comparison of distributions extracted from the graphs: Schieber et al, Nat. Comm. 8, 13928 (2017).

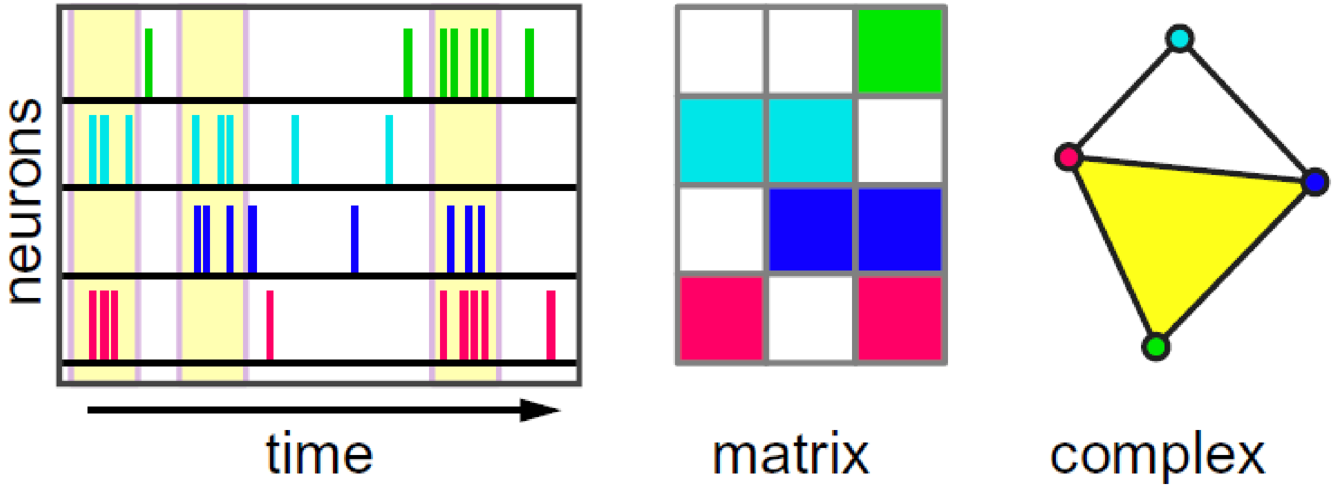
cristina.masoller@upc.edu @cristinamasoll1

A basic limitation of network analysis

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



Example:



Giusti et al., J Comput Neurosci 41, 1 (2016).
Battiston et al., Phys. Rep. 874, 1–92 (2020).

cristina.masoller@upc.edu @cristinamasoll1

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, unobserved variables, hidden “nodes” can difficult or make impossible the inference of the network structure.
- Network science: fast evolving field, with many applications and challenges!

cristina.masoller@upc.edu @cristinamasoll1

Collaborators, references and announcement

Andrés Aragoneses

Carlos Quintero

Riccardo Silini

Giulio Tirabassi

G. Tirabassi et al, Sci. Rep. 5 10829 (2015).
A. Aragoneses et al., Phys. Rev. Lett. 116, 033902 (2016).
M. Panozzo, C. Quintero-Quiroz et al, Chaos 27, 114315 (2017).
C. Quintero-Quiroz et al., Chaos 28, 106307 (2018).
R. Silini and C. Masoller, Sci. Rep. 11, 8423 (2021).

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

ICREA

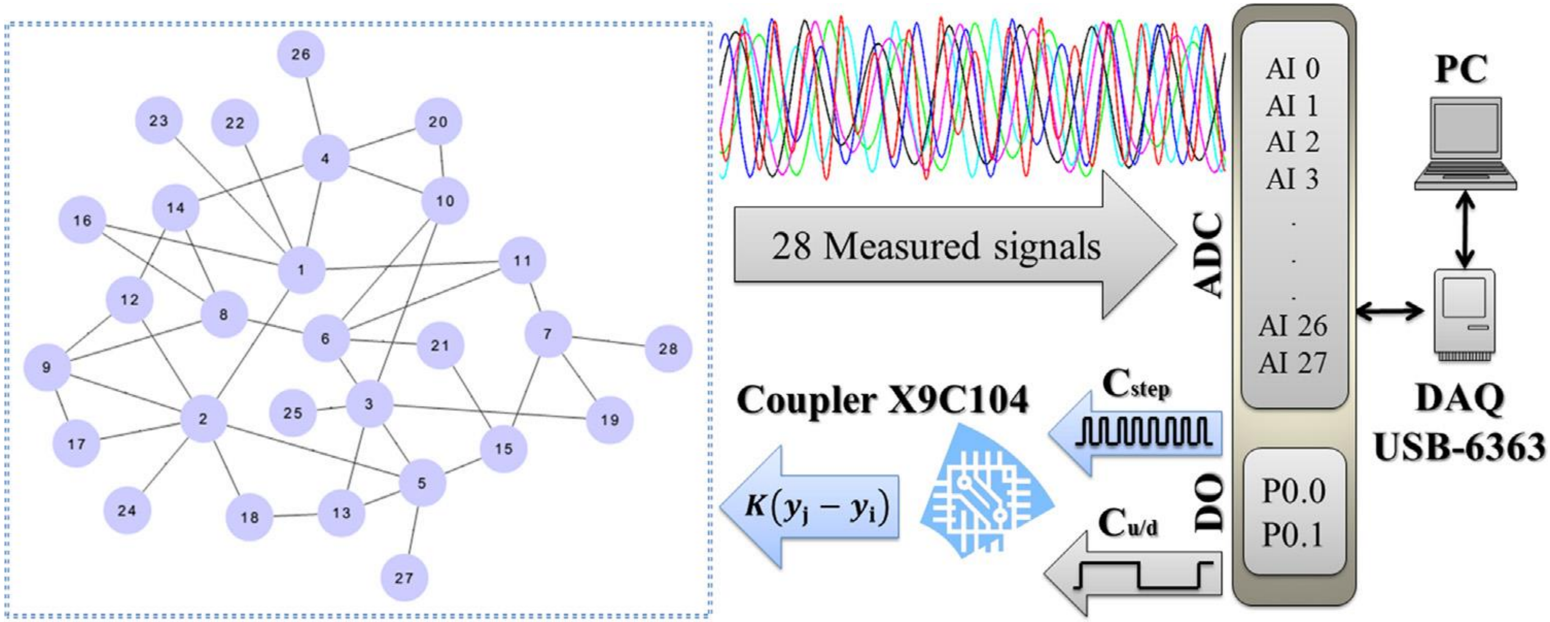
GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

MEDYFINOL Conference, Colonia del Sacramento, Uruguay, Dec. 8-10, 2022.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
UPC

cristina.masoller@upc.edu **@cristinamasoll1**

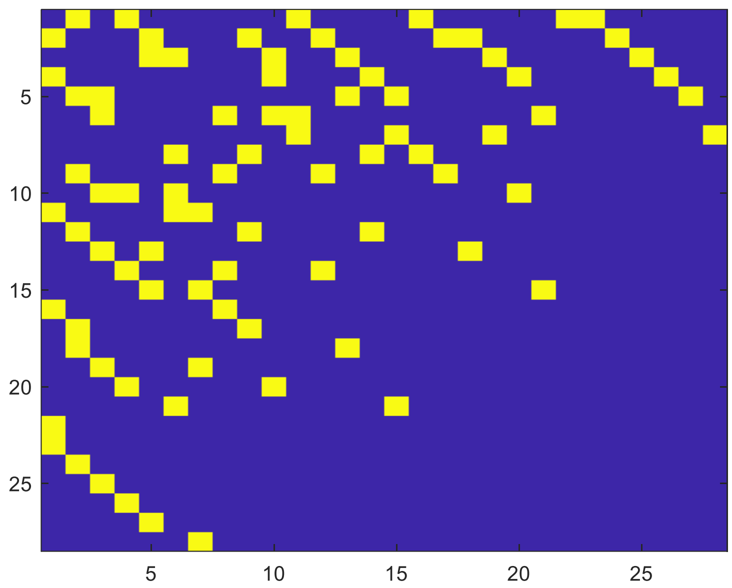
Hands-on exercise 4: analyze a random network of chaotic Rossler-like electronic oscillators

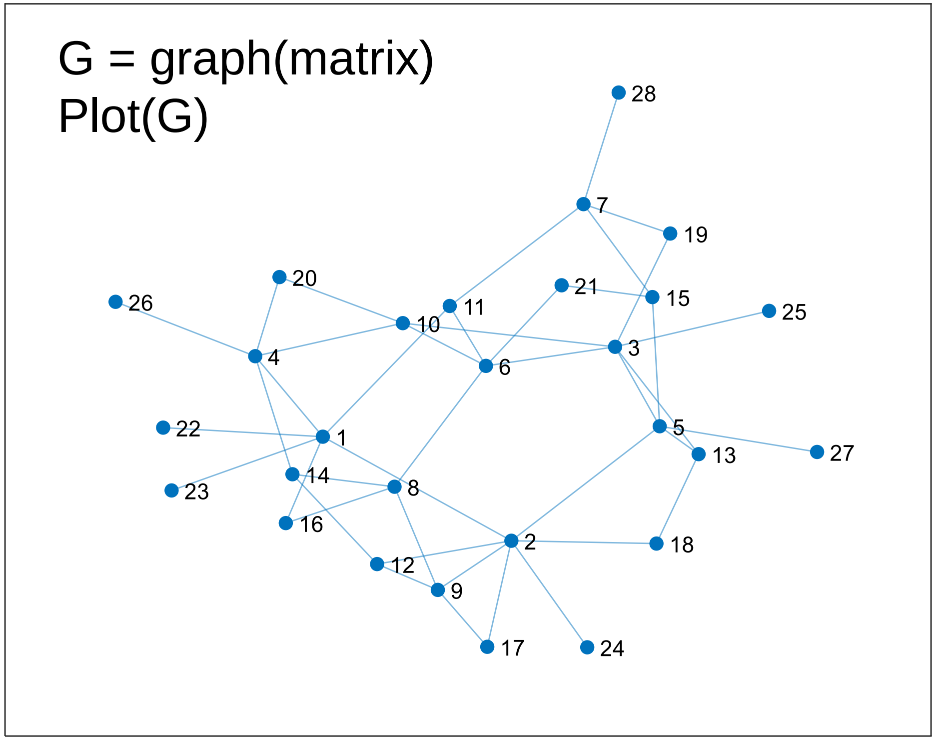


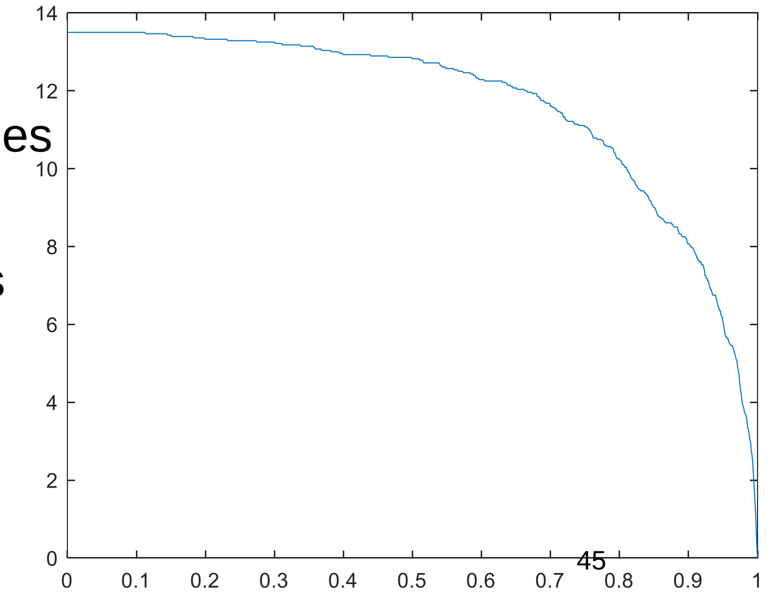
Reference: Data in Brief 7 (2016)1185–1189

1. Represent graphically the structural network.
2. Calculate the degree distribution of the structural network.
3. Reconstruct the network using the Pearson coefficient (ρ=|CC(0)|) as statistical similarity measure, keeping only the 10% strongest links.
4. Reconstruct the network using the threshold condition ρ>0.5.
5. Calculate the mean degree as a function of the threshold ρ_{th} = 0…1.

cristina.masoller@upc.edu @cristinamasoll1







Mean degree = number of links / number of nodes

Ts_25, only first 5000 datapoints