School on Nonlinear Dynamics, Complex Networks, Information Theory and Machine Learning in Neuroscience, 22-26 May 2023

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: Bivariate and multivariate analysis

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

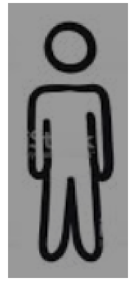
Campus d'Excel·lència Internacional

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

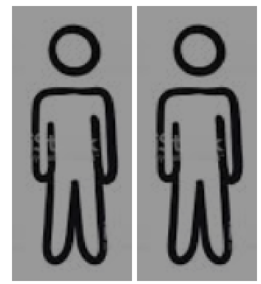
cristina.masoller@upc.edu @cristinamasoll1

Outline

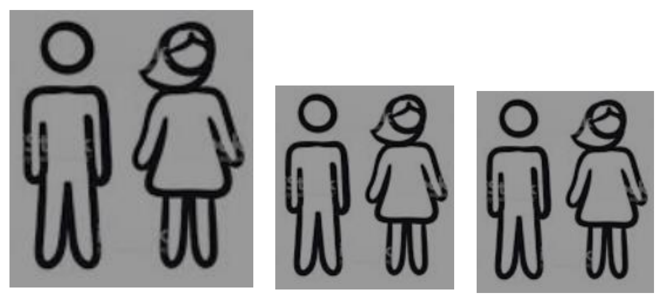
- Univariate analysis



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

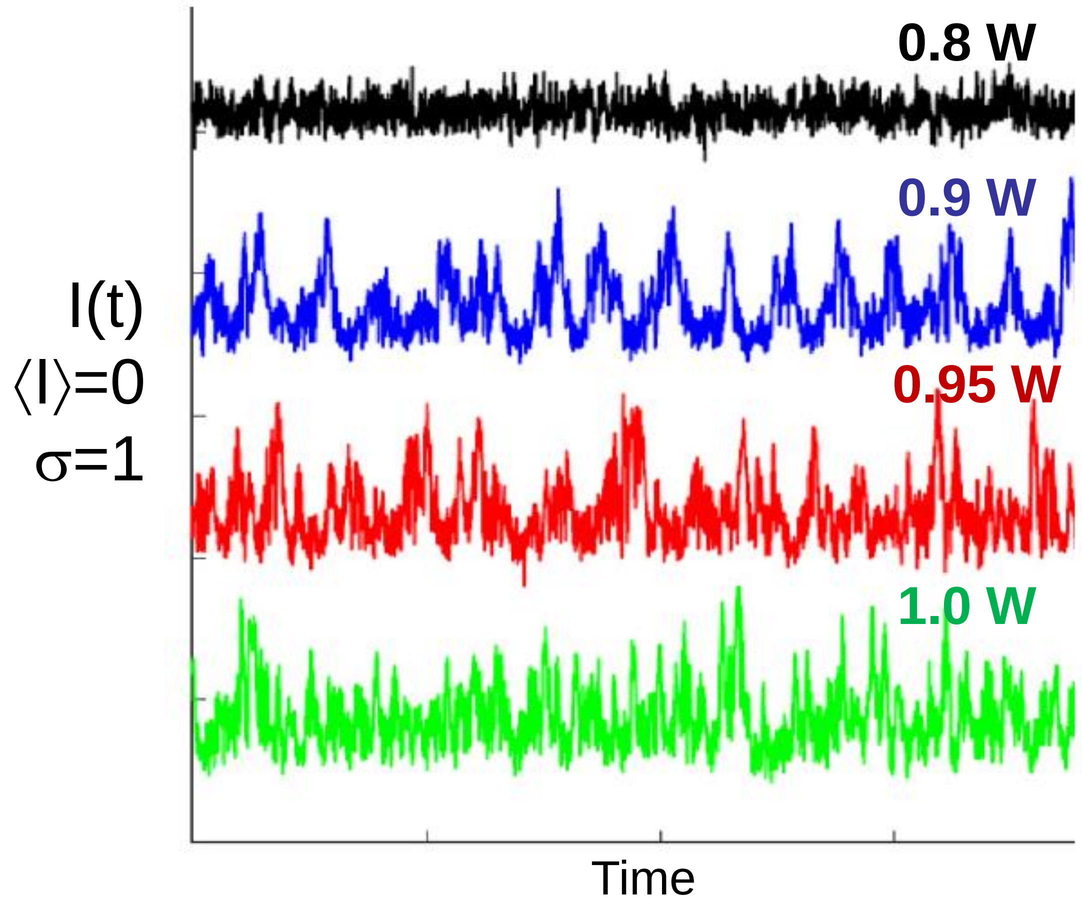


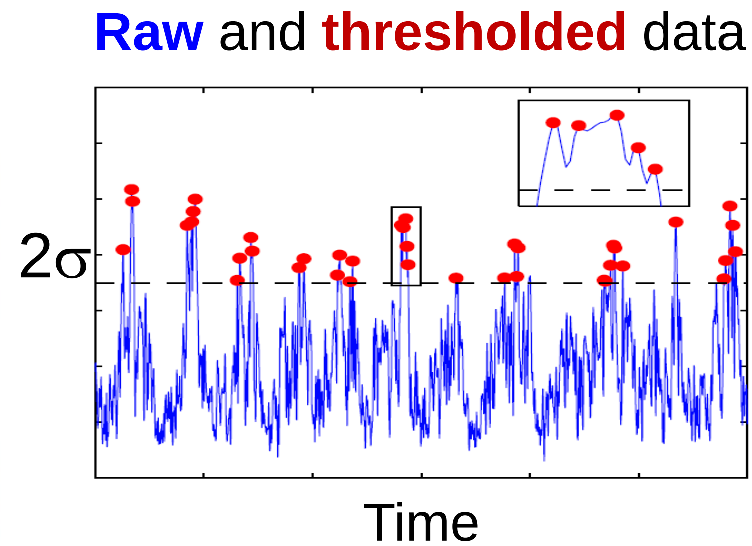
- Multivariate analysis



cristina.masoller@upc.edu @cristinamasoll1

Univariate analysis of “noisy” time series



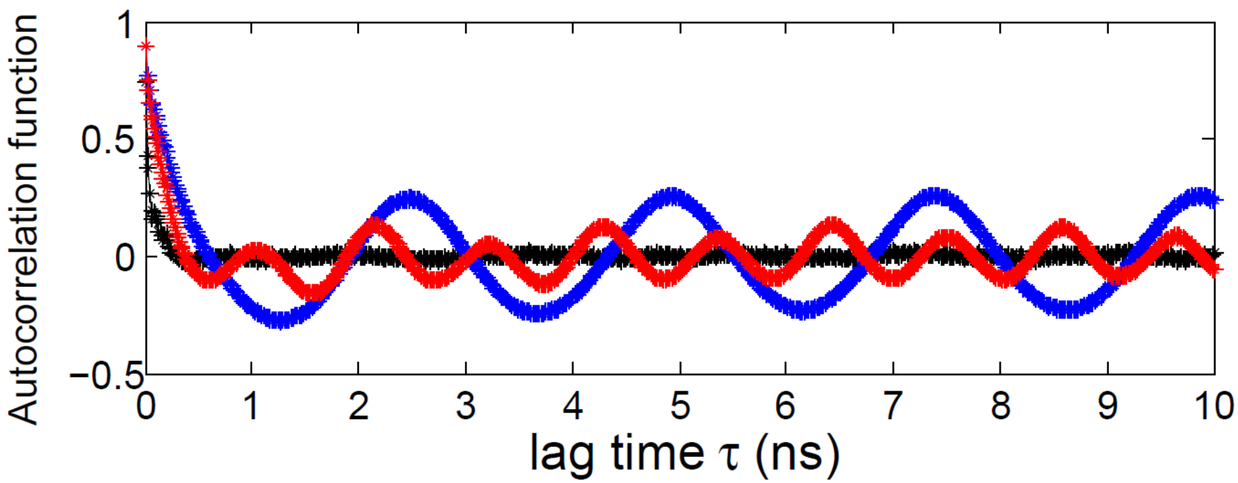


Can give very different results!

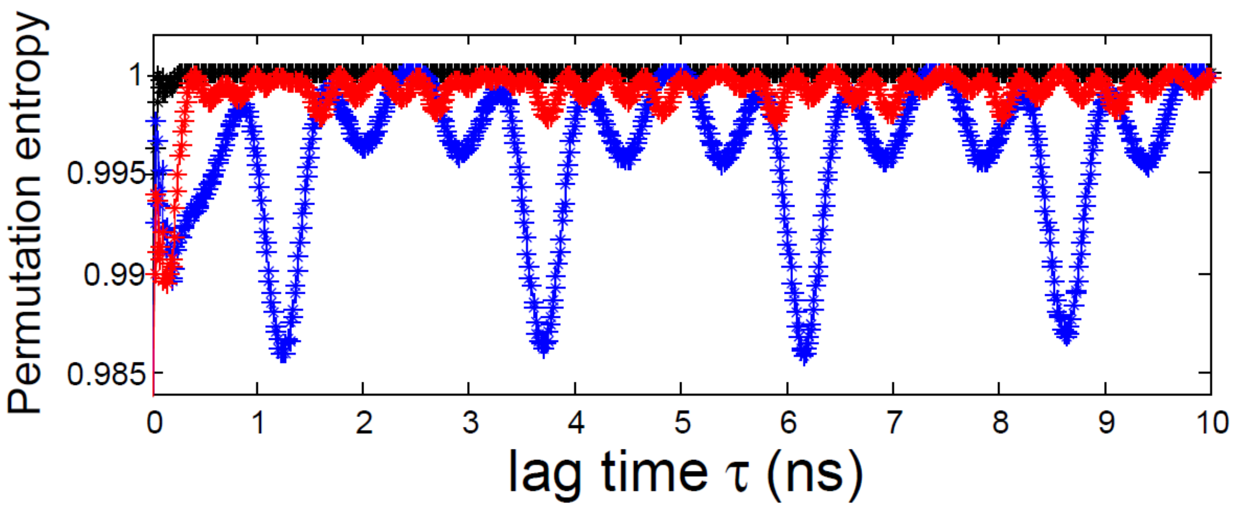
A. Aragoneses et al., “Unveiling temporal correlations characteristic to phase transition in the intensity of fibre laser radiation”, Phys. Rev. Lett. 116, 033902 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Entropy vs. lag finds hidden periodicities in data.



$$C(\tau) = \frac{\langle [I_i - \langle I \rangle][I_{i-\tau} - \langle I \rangle] \rangle}{\sigma^2}$$



$$\{I_i, I_{i+\tau}, I_{i+2\tau}, \ldots\}$$

$$H = -\sum_i p_i \log_2 p_i$$

A. Aragoneses et al., Phys. Rev. Lett. 116, 033902 (2016).

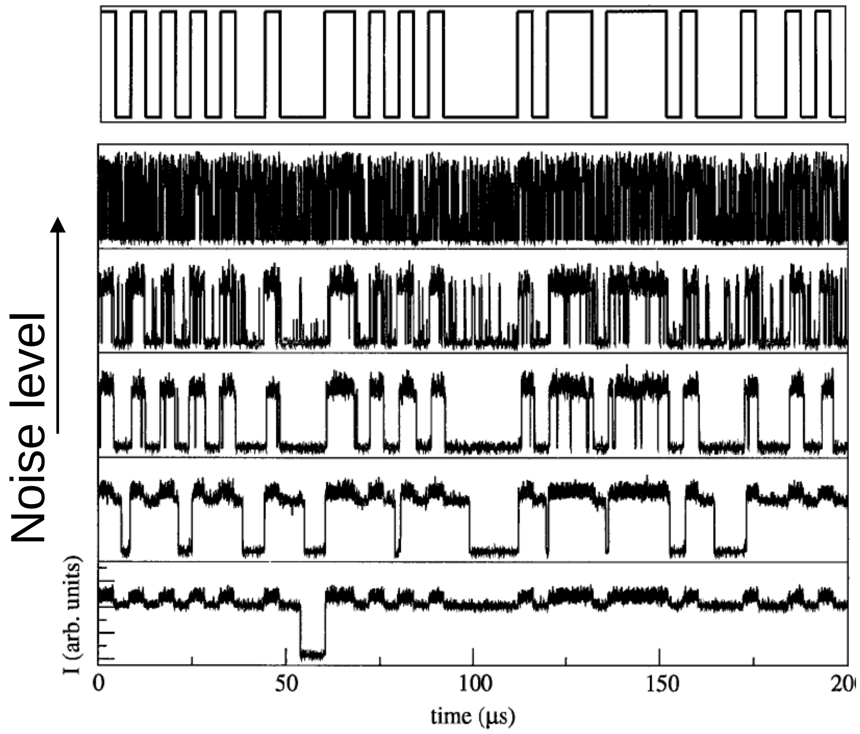
cristina.masoller@upc.edu @cristinamasoll1

Bivariate time series analysis: response of a bistable system to an aperiodic signal

Quantifying stochastic resonance using cross-correlation

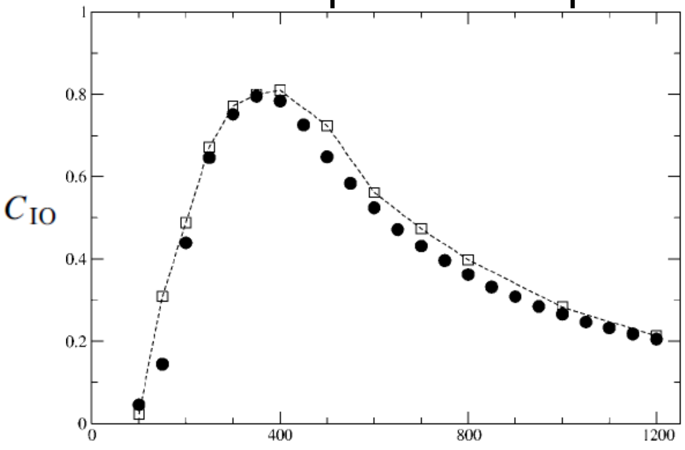
$$C_{xy}(\tau) = \frac{1}{N-\tau}\sum_{k=1}^{N-\tau} x(k+\tau)y(k)$$

x(t) and y(t) are normalized to **μ=0** and **σ=1**



Barbay et al, PRL 85, 4652 (2000)

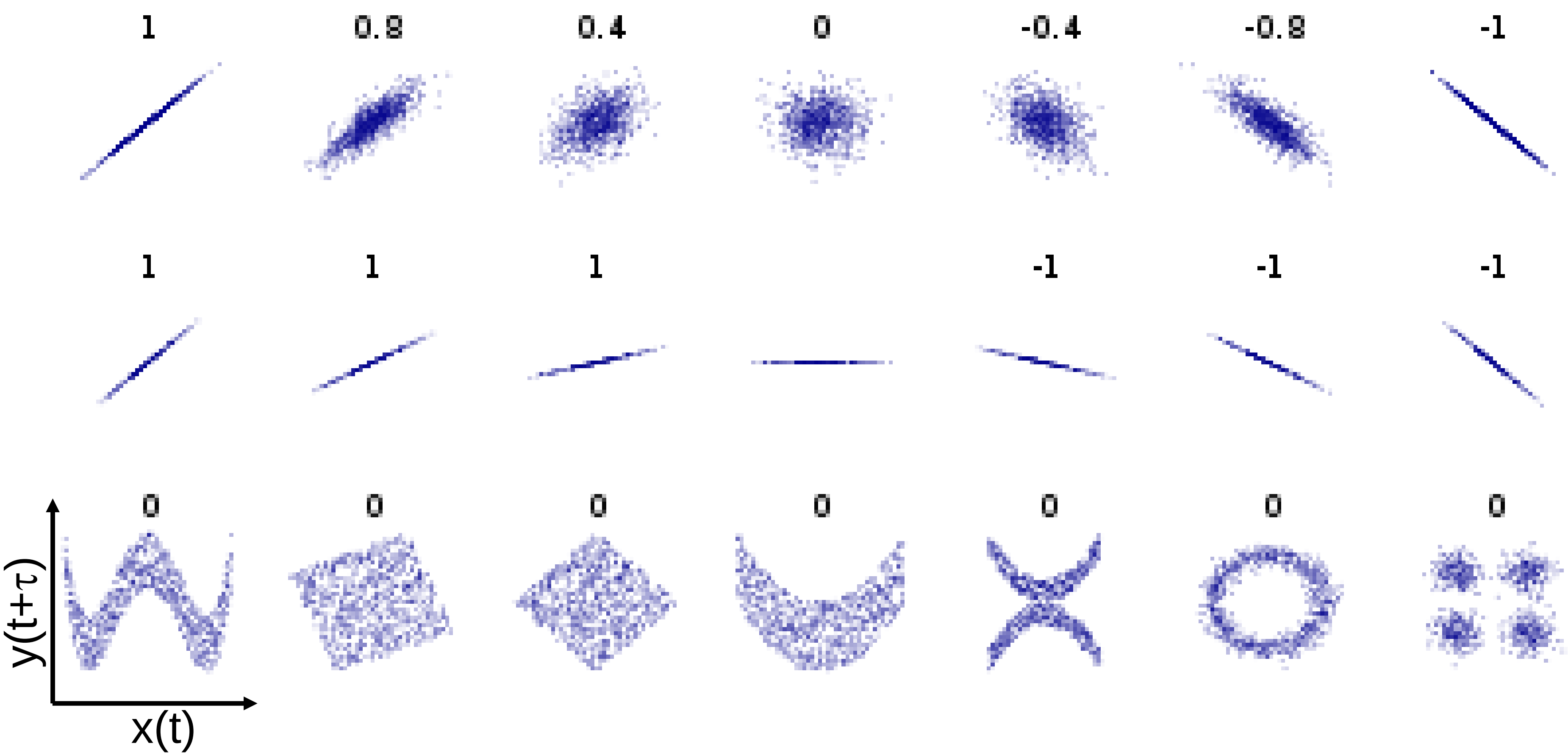
Zero-lag cross-correlation between input and output.



Noise level

cristina.masoller@upc.edu @cristinamasoll1

Cross-correlation detects linear relationships only



Source: wikipedia

cristina.masoller@upc.edu @cristinamasoll1

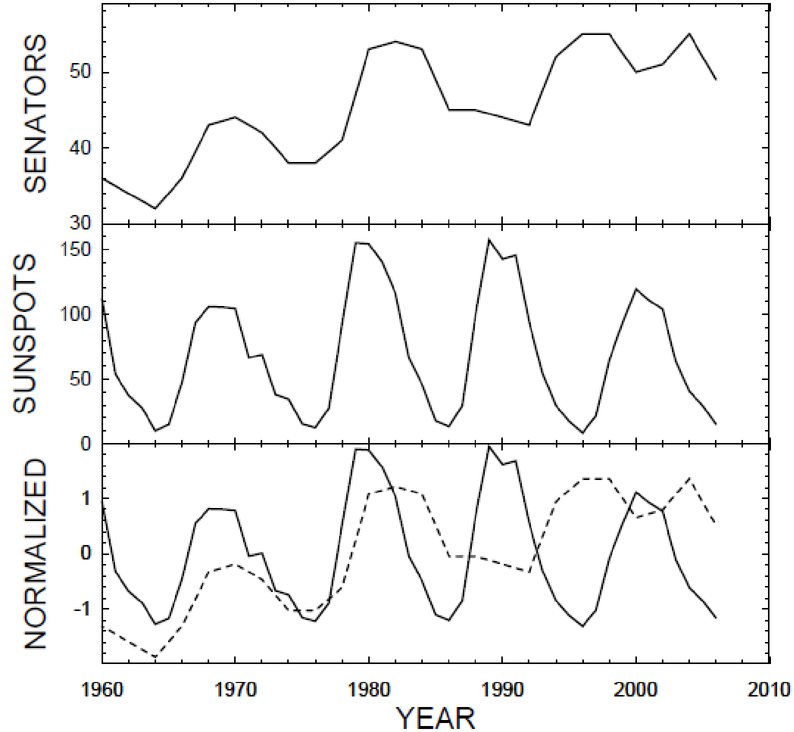
Correlation is NOT causality

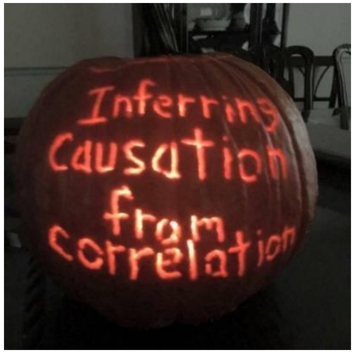
Example: the number of sunspots and the number of the Republicans in the U.S. Senate in the years 1960 to 1986 (biannual sampling, 14 points): **C=0.52**

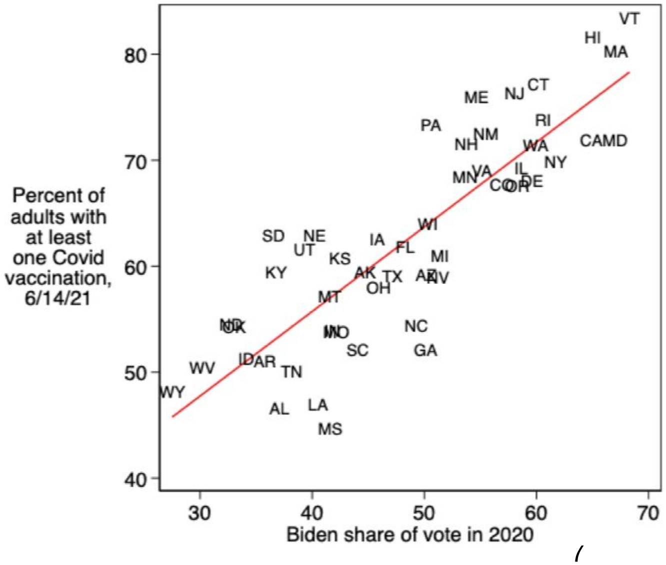
Appropriate significance test needed!

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*** ***!!! (significance test needed)***

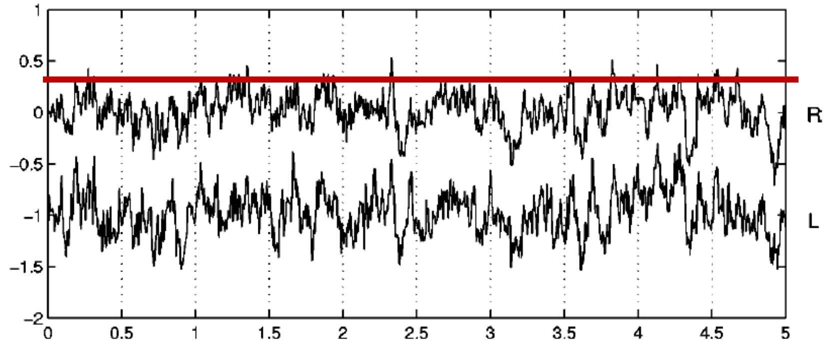
- *MI* can be computed with a lag-time.

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

- If x and y are Gaussian processes *MI* = -1/2 log(1-ρ^2) (ρ=cross-correlation).

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 <u>shortly after</u> (within interval τ) an event appears in y. Idem 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 occur before those in y.
- $q_\tau = -1$: the events in x always occur after those in y.

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

cristina.masoller@upc.edu @cristinamasoll1

Many other measures are available to quantify synchronization of two time series

PHYSICAL REVIEW E, VOLUME 65, 041903

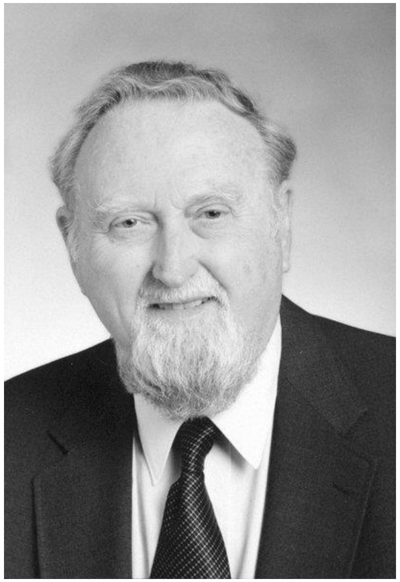
Performance of different synchronization measures in real data: A case study on electroencephalographic signals

R. Quian Quiroga,[1,*] A. Kraskov,[1] T. Kreuz,[1,2] and P. Grassberger[1]
[1]*John von Neumann Institute for Computing, Forschungszentrum Jülich GmbH, D-52425 Jülich, Germany*
[2]*Department of Epileptology, University of Bonn, Sigmund-Freud Strasse 25, D-53105 Bonn, Germany*
(Received 18 September 2001; published 15 March 2002)

V. CONCLUSIONS

We applied several linear and nonlinear measures of synchronization to three typical EEG signals. Besides mutual information, which was not robust due to the low number of data points, all these measures gave a similar tendency in the synchronization levels. A similar analysis would have been

Granger Causality



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

$$X_1(t) = \underbrace{\sum_{j=1}^{p} A_{11,j}\, X_1(t-j)}_{\text{past of } X_1} + \underbrace{E_1(t)}_{\text{Residual error}}$$

$$X_1(t) = \underbrace{\sum_{j=1}^{p} A_{11,j}\, X_1(t-j)}_{\text{past of } X_1} + \underbrace{\sum_{j=1}^{p} A_{12,j} X_2(t-j)}_{\text{past of } X_2} + \underbrace{E'_1(t)}_{\text{Residual error}}$$

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

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)

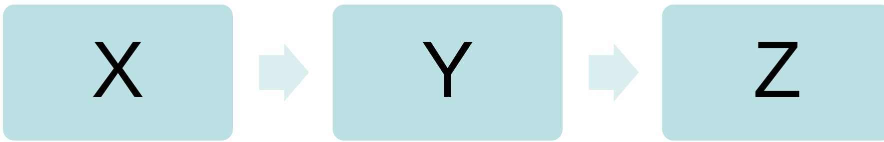
- Directionality Index: TE(x,y)-TE(y,x)
- TE and GC are equivalent for Gaussian processes.
- TE can be computed from the probabilities of symbols (symbolic TE).

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

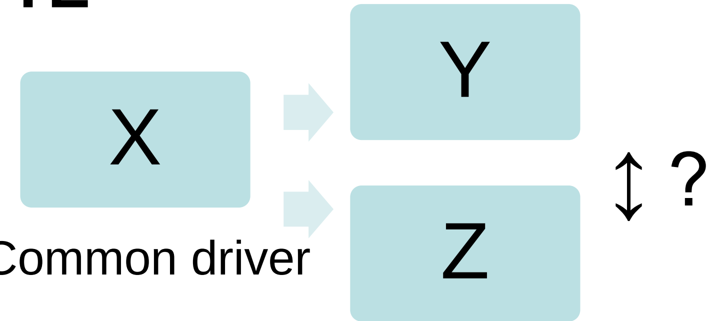
K. Hlaváčková-Schindler et al. / Physics Reports 441 (2007) 1–46

cristina.masoller@upc.edu @cristinamasoll1

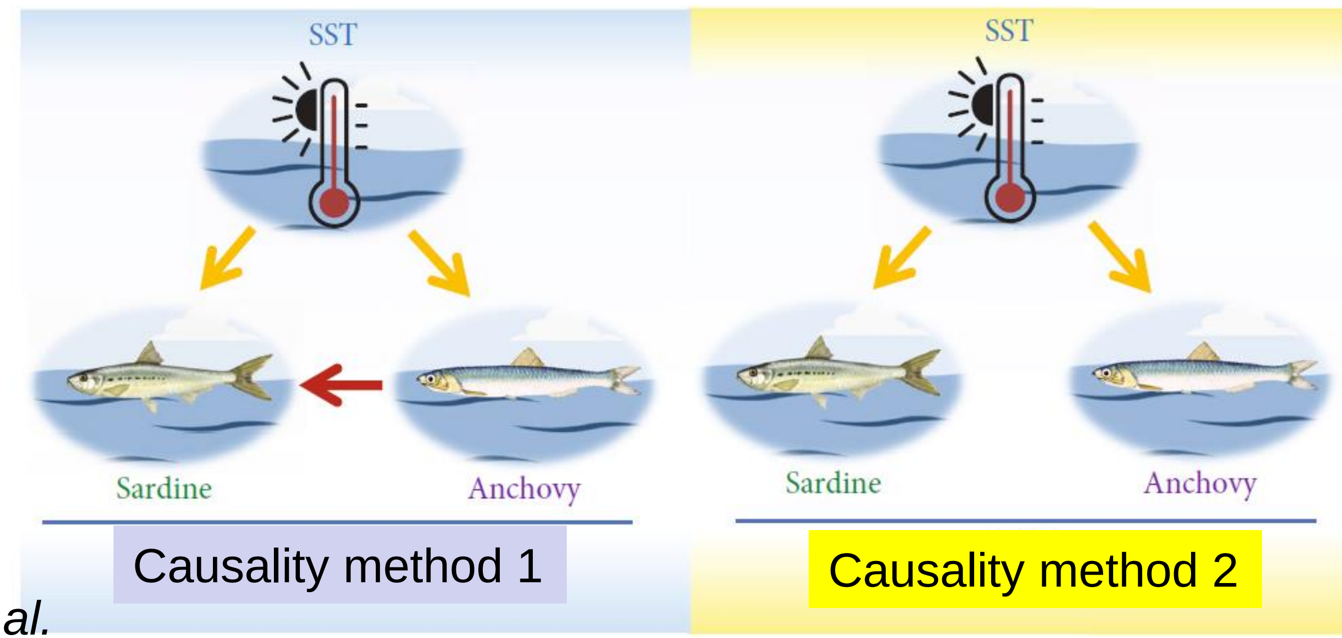
Problems of Grange Causality and TE



Indirect link X→Z?



Common driver



X. Ying et al.
AAAS Research 2022

In addition: Transfer Entropy is computationally demanding.

cristina.masoller@upc.edu @cristinamasoll1

A “simple” solution

Use the expression of Transfer Entropy that is valid for Gaussian processes [MI = $-1/2 \log(1-\rho^2)$]

Does this work?

Sometimes

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

Besides Granger Causality and Transfer Entropy, many methods have proposed

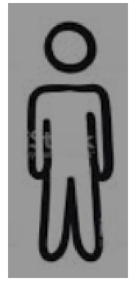
- Symbolic Transfer Entropy
- Partial Correlation
- Partial Directed Coherence
- Cross Mapping
- Partial Cross Mapping
- Etc.

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)

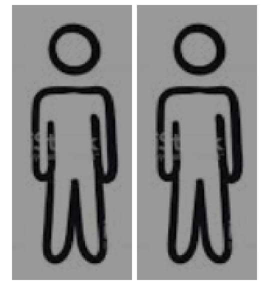
cristina.masoller@upc.edu @cristinamasoll1

Outline

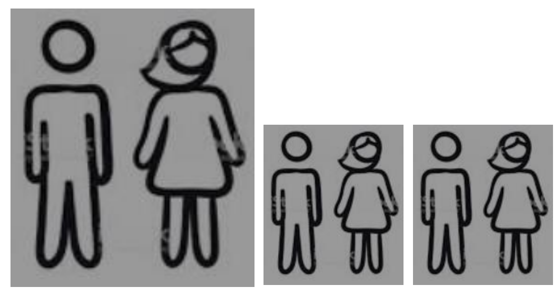
- Univariate analysis



- Bivariate analysis



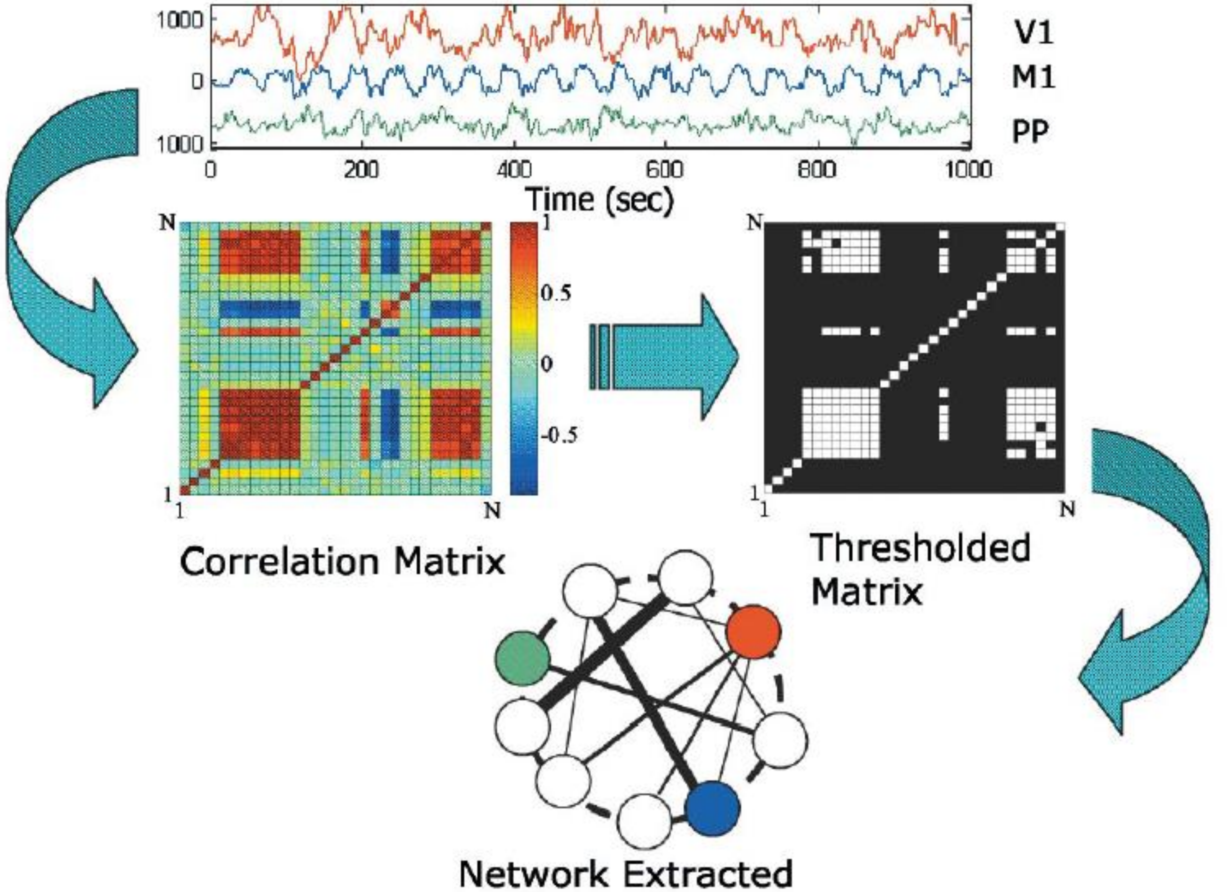
- **Multivariate analysis**
  - Functional networks
  - Network inference

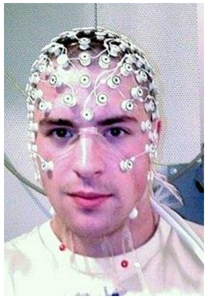


cristina.masoller@upc.edu @cristinamasoll1

"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).

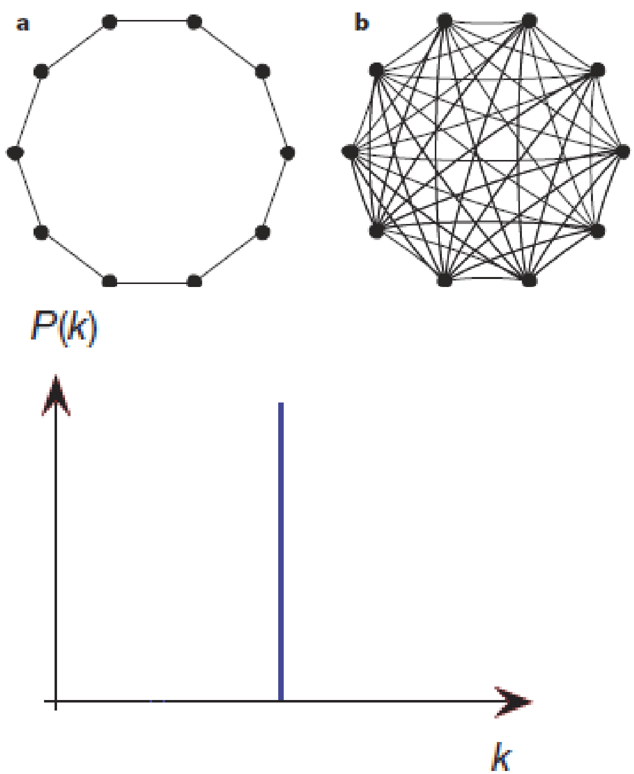
cristina.masoller@upc.edu @cristinamasoll1

How to characterize the graph?

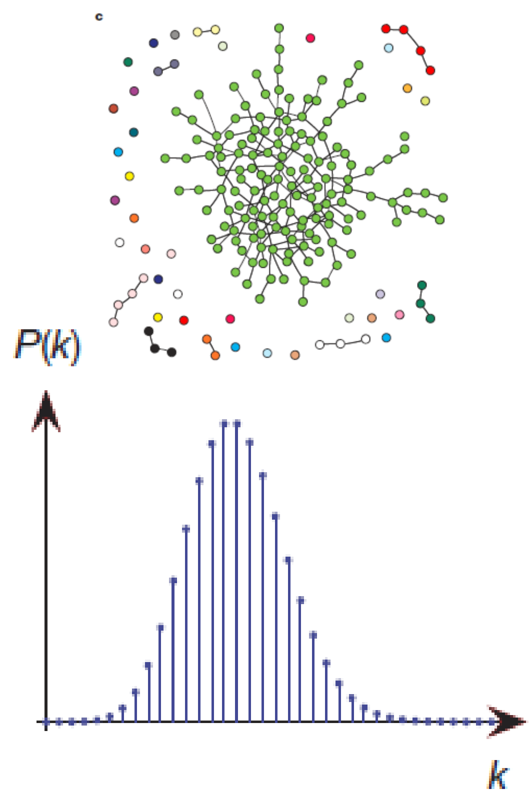
Begin with the degree distribution

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

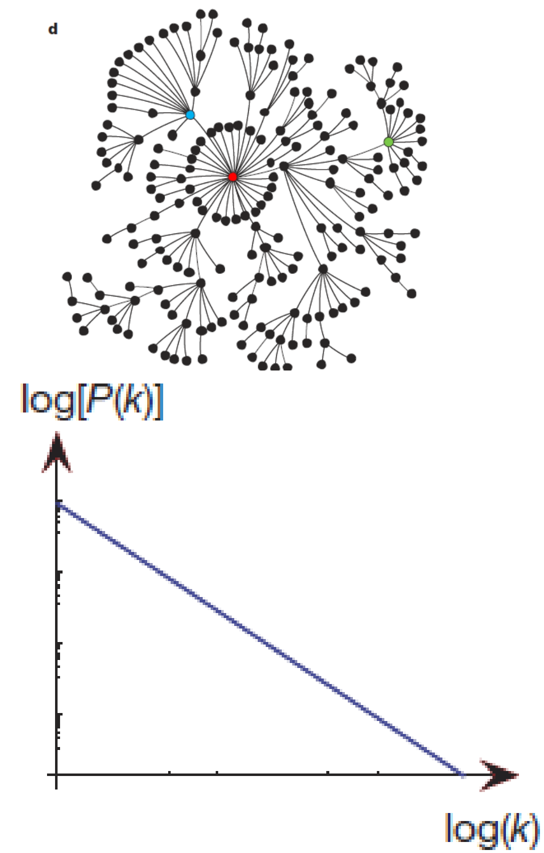
Regular



Random



Scale-free



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

cristina.masoller@upc.edu @cristinamasoll1

How to compare two distributions? (Prof. Rosso's talk)

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

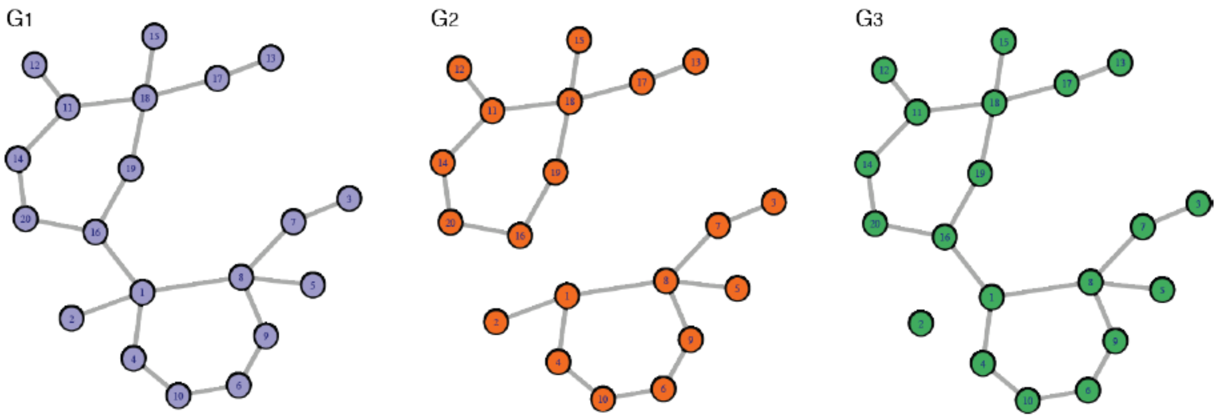
Jensen divergence $$D_J[P, P_e] = \frac{K\left[P|P_e\right] + K\left[P_e|P\right]}{2}$$

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

How different 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 of different sizes, based in distances between distributions extracted from the graphs:

T. A. Schieber, L. Carpi, A. Diaz-Guilera, P. M. Pardalos, C. Masoller, M. G. Ravetti, "*Quantification of network structural dissimilarities*", Nature Communications **8**, 13928 (2017).

cristina.masoller@upc.edu @cristinamasoll1

Network inference: How to reconstruct the network from observations?

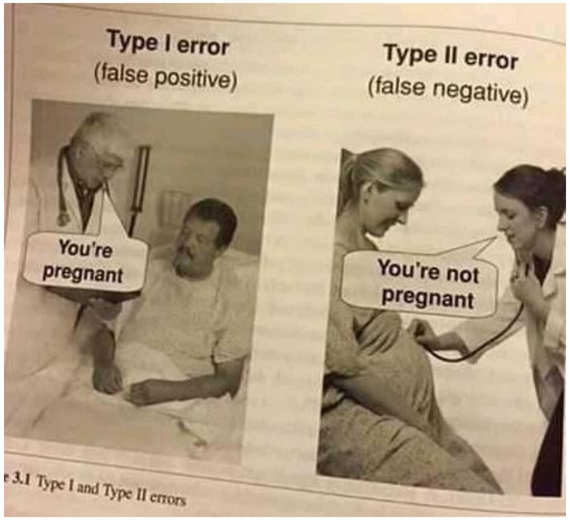
$$S_{ij} > \mathrm{Th} \Rightarrow A_{ij} = 1 \text{ else } A_{ij}=0$$

- 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

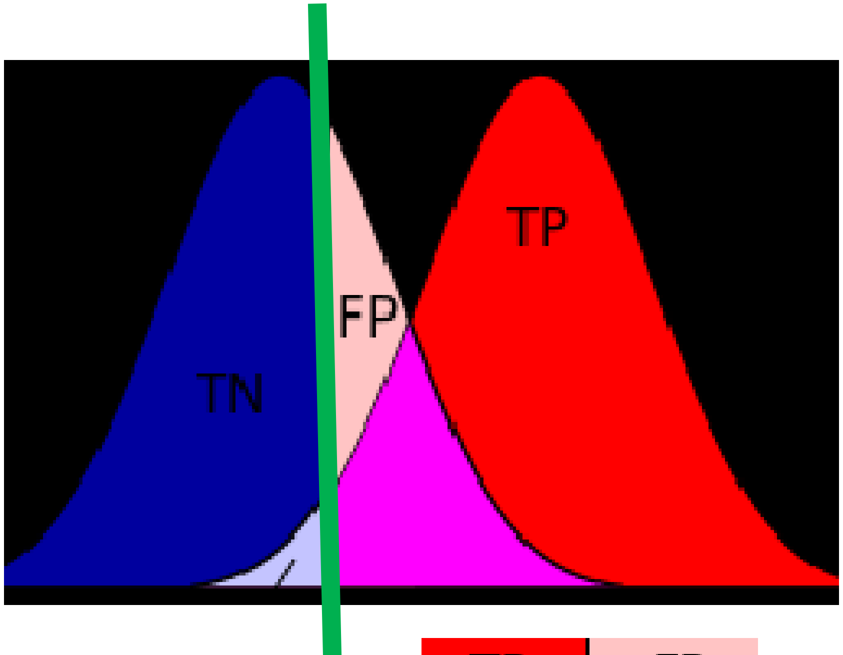
	Predicted: NO	Predicted: YES
Actual: NO	TN	FP
Actual: YES	FN	TP



- ***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**

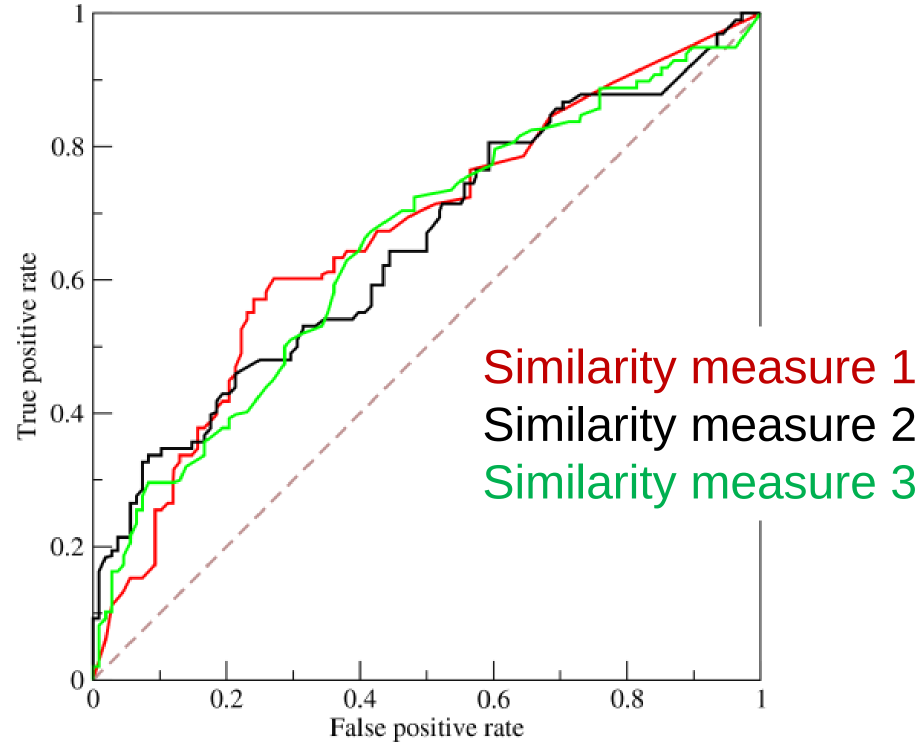
cristina.masoller@upc.edu @cristinamasoll1

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



Source: Wikipedia

TP	FP
FN	TN



For <u>unbalanced sets</u> 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

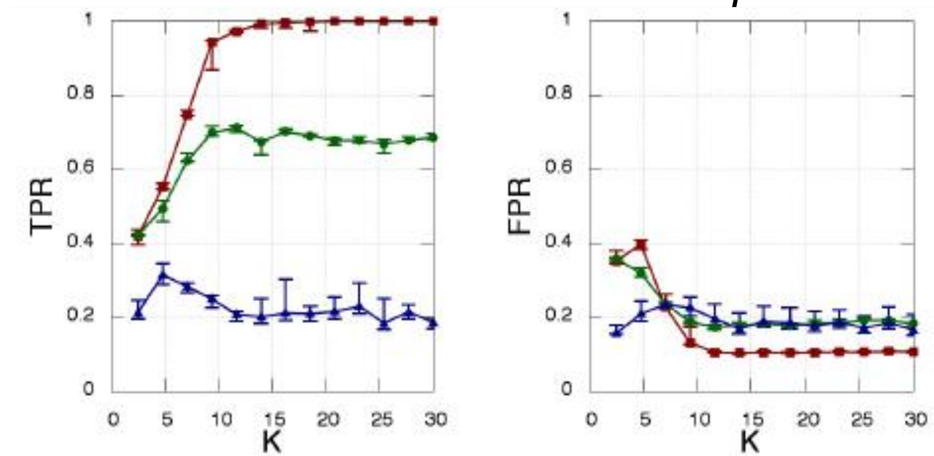
$$d\theta_i = \omega_i dt + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_j - \theta_i)dt + D\, dW_t^i$$

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

CC MI MIOP

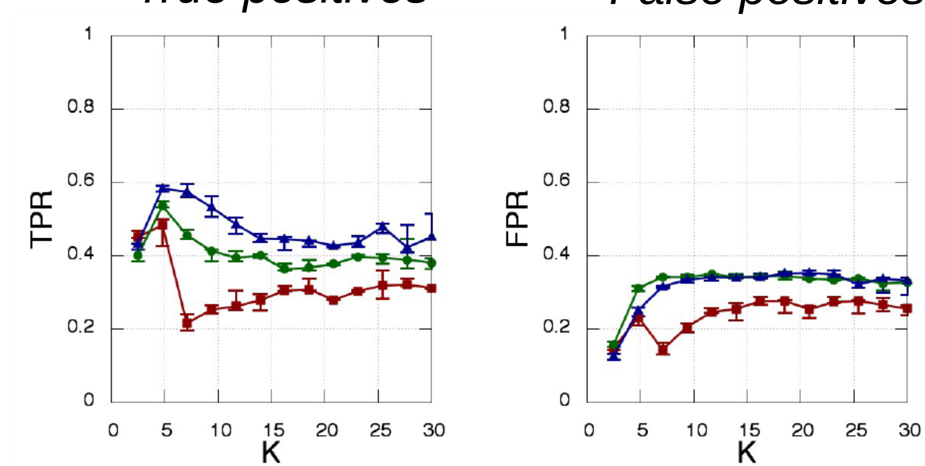
Phases (θ)

True positives *False positives*



"Observable" Y=sin(θ)

True positives *False positives*

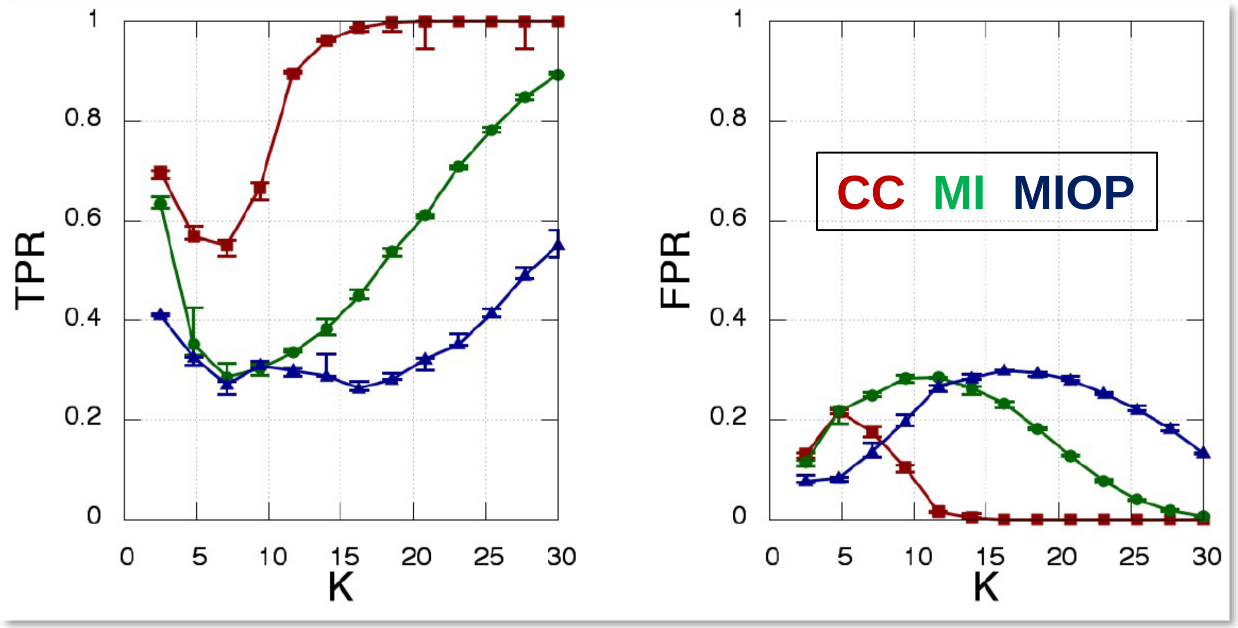


For each K, the threshold was varied to obtain optimal reconstruction.

G. Tirabassi, R. Sevilla-Escoboza, J. M. Buldú, C. Masoller, "Inferring the connectivity of coupled oscillators from time-series statistical similarity analysis", Sci. Rep. 5 10829 (2015).

cristina.masoller@upc.edu @cristinamasoll1

Instantaneous frequencies (dθ/dt)



Perfect network inference is possible!

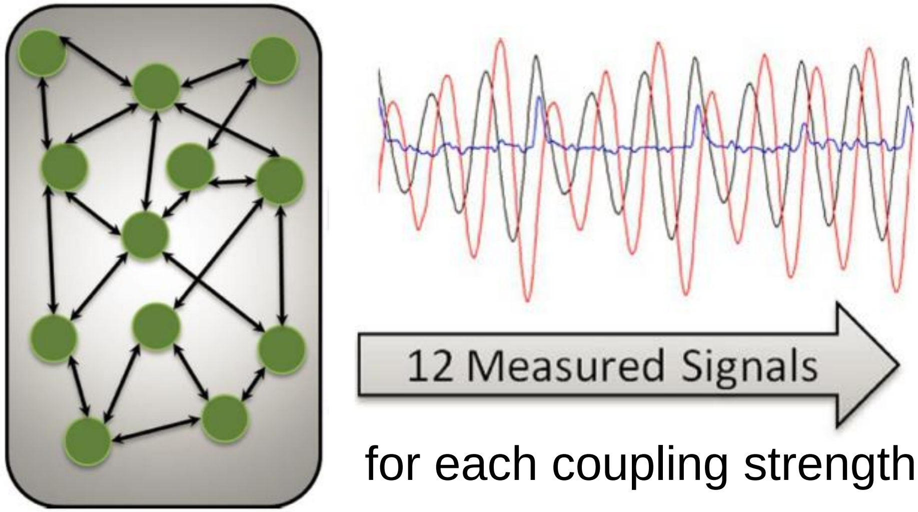
BUT

- the number of oscillators is small (12),
- the coupling is symmetric (⇒ 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

Test with experimental data recorded from 12 chaotic electronic oscillators (symmetric and random coupling)



for each coupling strength

The Hilbert Transform was used to obtain phases from experimental data

- Kuramoto Oscillators' Network

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

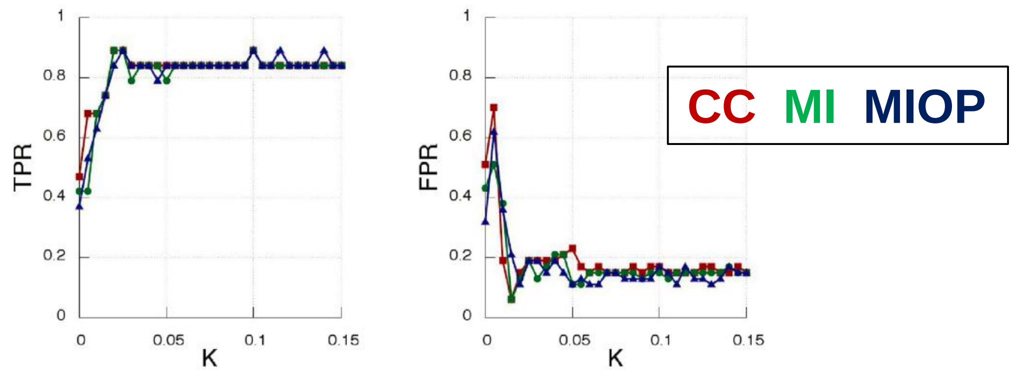
- Rössler Oscillators' Network

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

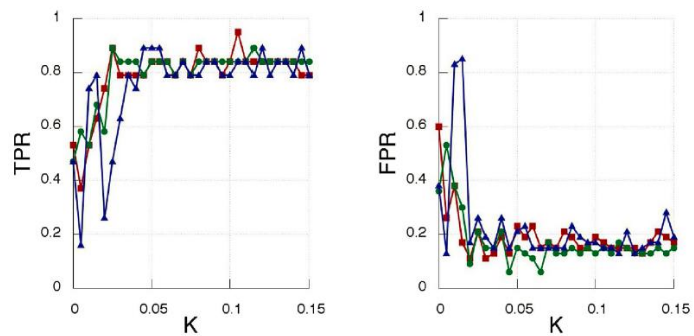
cristina.masoller@upc.edu @cristinamasoll1

Results obtained with experimental data

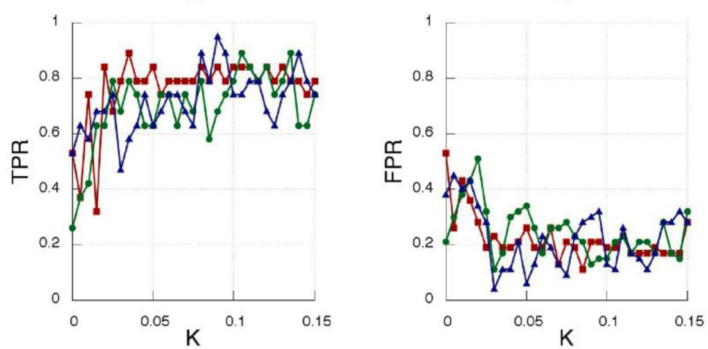
Observed variable (x)



Hilbert phase



Hilbert frequency



– No perfect reconstruction

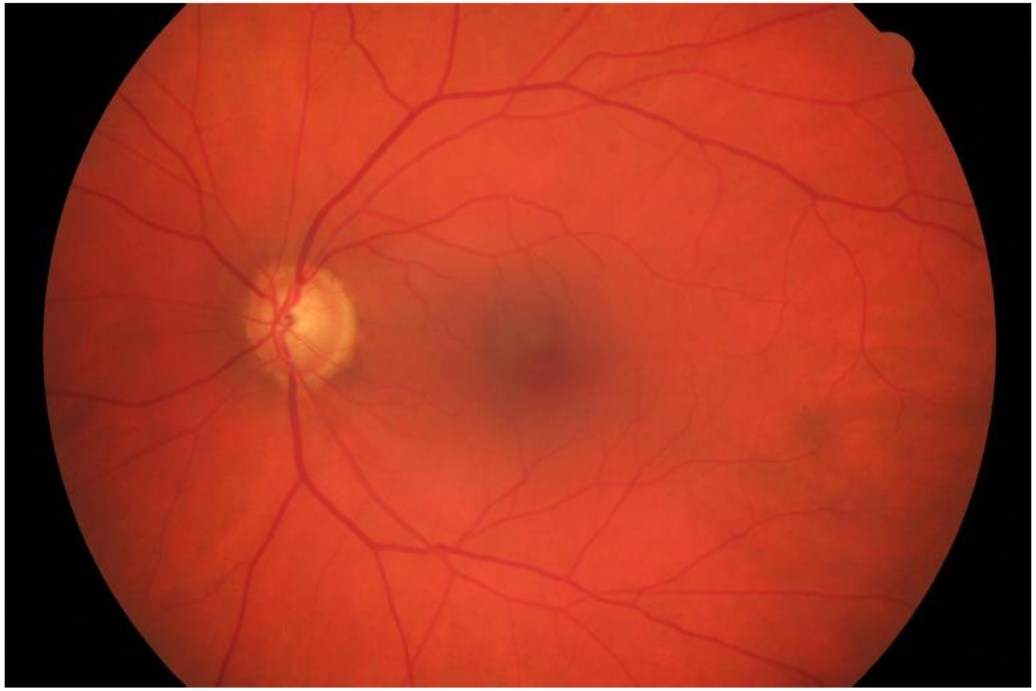
– No important difference among the 3 methods & 3 variables

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

cristina.masoller@upc.edu @cristinamasoll1

Machine learning, network-based analysis of retina fundus images

- For the diagnosis of eye diseases & follow up of treatments.
- Biometric identity identification.
- The retina is a window to the brain.
- Opportunity to detect other diseases: alterations in retina network may reflect alterations in other arterial systems.



BE-OPTICAL

Advanced Biomedical Optical Imaging and Data Analysis

H2020-675512

cristina.masoller@upc.edu @cristinamasoll1

Brief introduction to machine learning algorithms

What is a ML algorithm?

A computer program that learns from experience E with respect to some task T and performance measure P: its performance at task T, as measured by P, improves with experience E.

(T. Mitchell, Machine Learning, 1997)

Example

- Suppose your email program observes which emails you mark as spam and which you do not, and based on that information learns how to better filter spam.
- Task T: Classifying emails as spam or not spam
- Experience E: Observing you label emails as spam or not spam
- Performance P: The number (or fraction) of emails correctly classified as spam/not spam

(taken from T. Eliassi-Rad)

cristina.masoller@upc.edu @cristinamasoll1

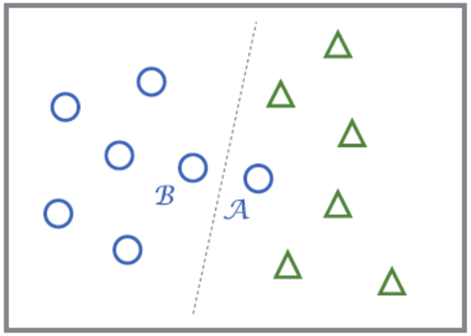
Main types of ML algorithms

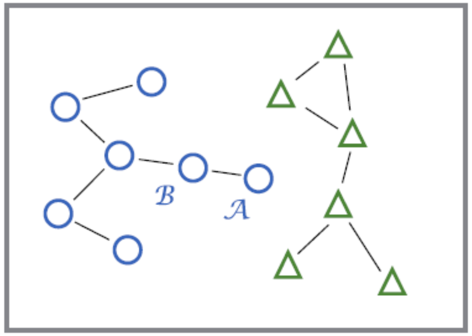
- **Supervised learning**: Given a labeled training set, can we accurately predict/classify new data points?
  - Classification
  - Regression

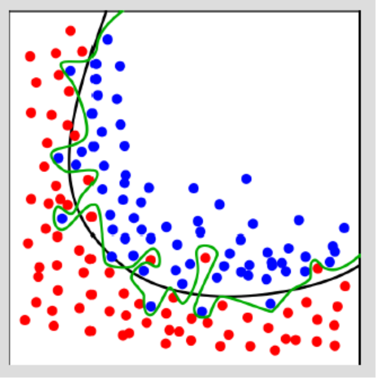
- **Unsupervised learning**: Can we discover structure in unlabeled data?
  - Clustering

Output of community detection algorithm

Problem: overfitting



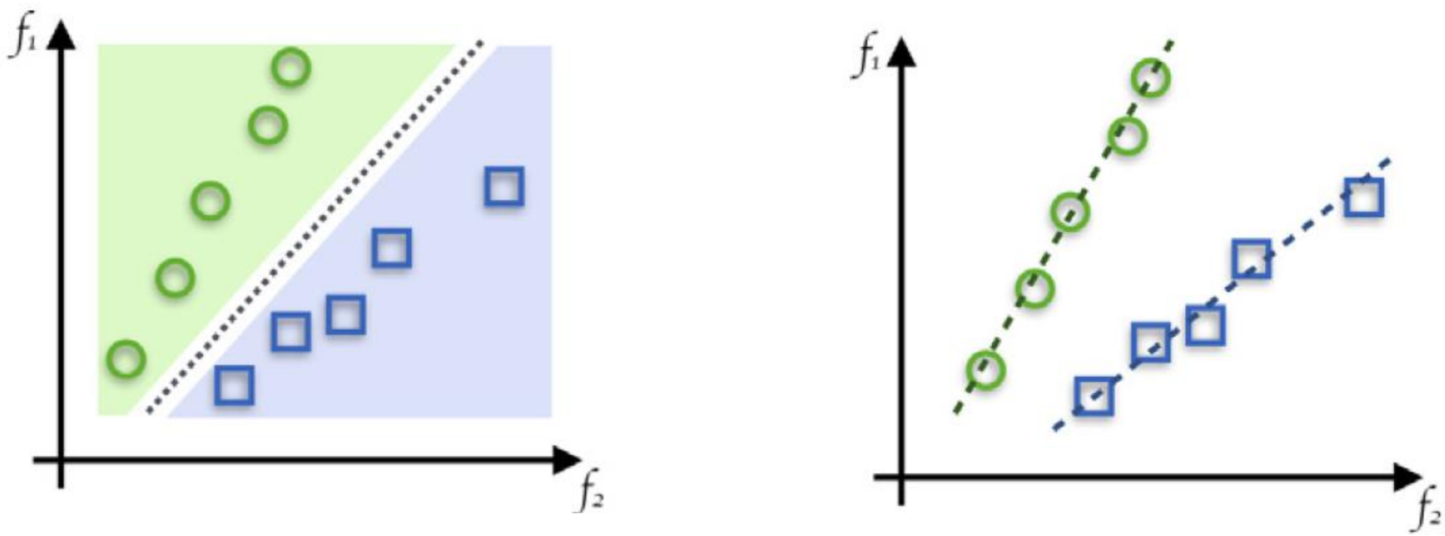




M. Zanin et al., Physics Reports 635, 1 (2016).

cristina.masoller@upc.edu @cristinamasoll1

Classification and regression



More in general:

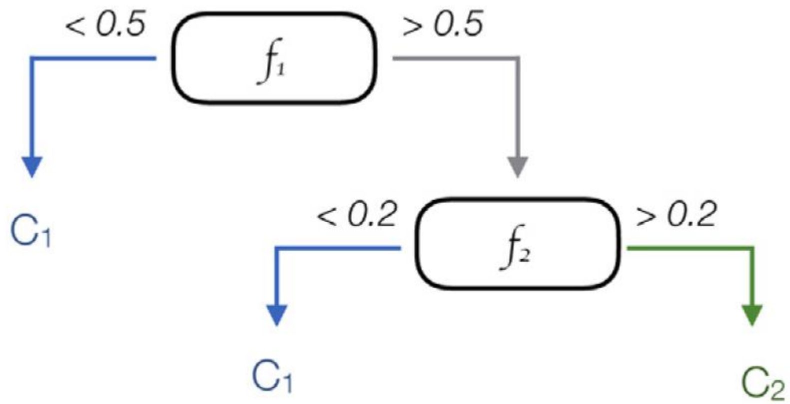
Input → System → Output

M. Zanin et al., Physics Reports 635, 1 (2016).

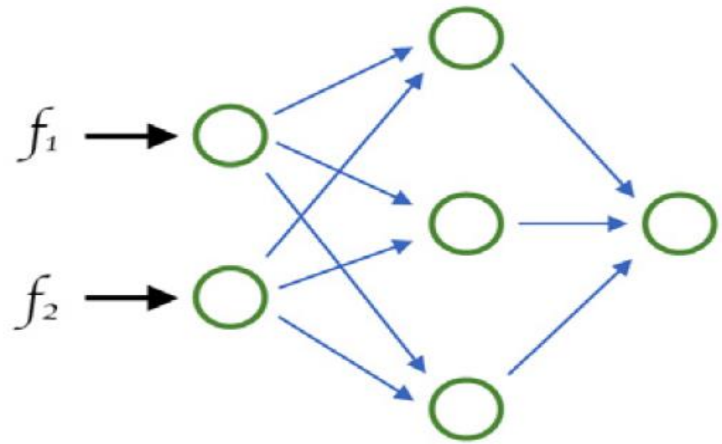
cristina.masoller@upc.edu @cristinamasoll1

Supervised learning algorithms

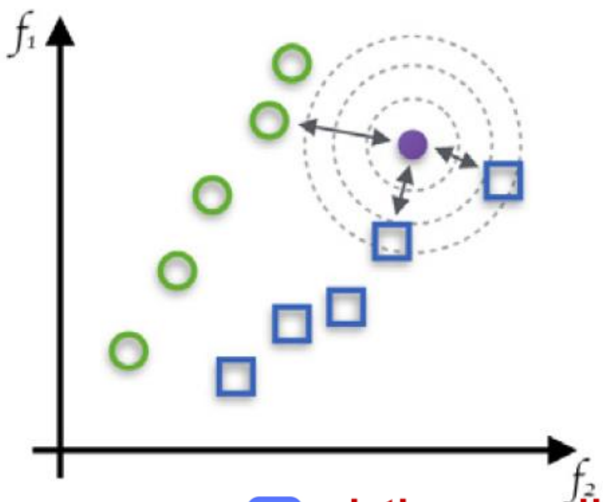
Decision Tree



Artificial Neural Network

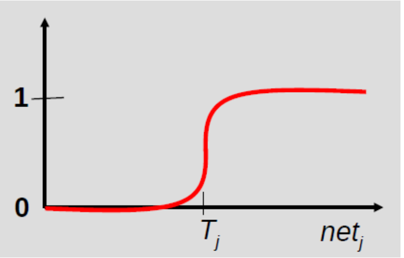


k nearest neighbors



Sigmoidal activation function

$$o_j = \frac{1}{1+e^{-(net_j - T_j)}}$$



cristina.masoller@upc.edu @cristinamasoll1

Forecasting methods

$$\mathbf{x}_i = \left(y_{i-n}, \ldots, y_{i-1}\right)$$

- Learn function $\tilde{f}\left(\mathbf{x}_i\right) = \tilde{y}_i$

- Evaluate performance $\mathrm{MARE} = \frac{1}{N}\sum_{i=1}^{i=N}\frac{\left|\tilde{y}_i - y_i\right|}{y_i}$

  mean absolute relative error, cross correlation, etc.

- kNN: $\tilde{y} = \frac{1}{k}\sum_{j\in\mathcal{N}} y_j$

- Support Vector Machine (SVM): inner product of points in the set is used to approximate the response function

$$\tilde{f}\left(\mathbf{x}_i\right) = \sum_j \beta_j \left\langle \mathbf{x}_j, \mathbf{x}_i \right\rangle + b$$

Parameters are obtained with optimization techniques

Linear $\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle = \mathbf{x}_i^t \mathbf{x}_j$

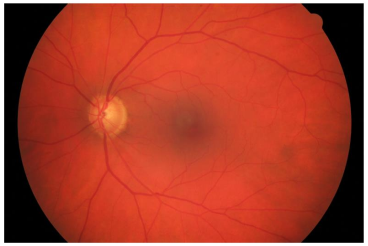
Nonlinear $\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|}{2\sigma^2}\right)$

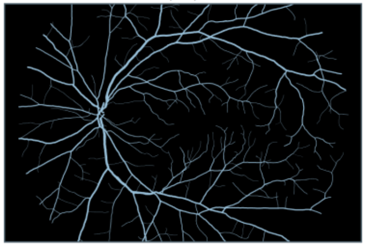
P. Amil, M. C. Soriano, and C. Masoller, "*Machine learning algorithms for predicting the amplitude of chaotic laser pulses*", Chaos 29, 113111 (2019).

cristina.masoller@upc.edu @cristinamasoll1

Data and image analysis steps

- 45 high resolution images (3504 × 2336 pixels)
  15 healthy subjects
  15 glaucoma
  15 diabetic retinopathy
- For every subject we had:
  - fundus photography
  - <u>manual</u> segmentation done by an expert ophthalmologist.





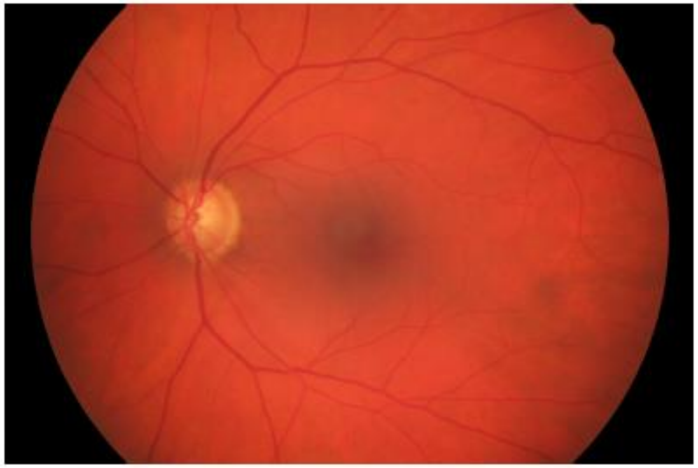
Steps:

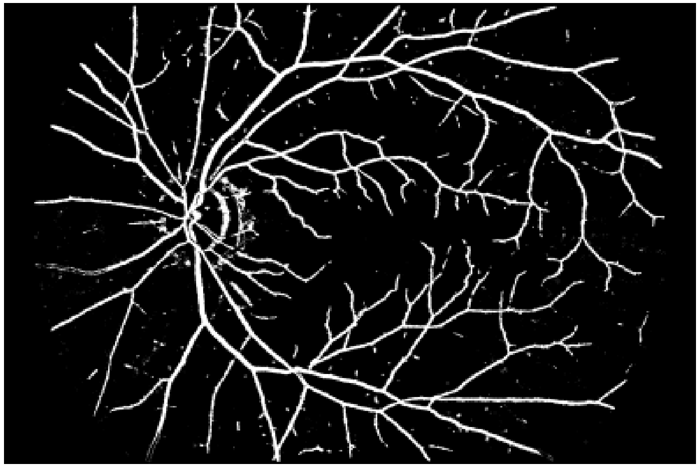
1. Pre-process and unsupervisely, segment the images.
2. Extract network.
3. Extract features by comparing networks obtained from different images.
4. Classify the images.

https://www5.cs.fau.de/research/data/fundus-images/

cristina.masoller@upc.edu @cristinamasoll1

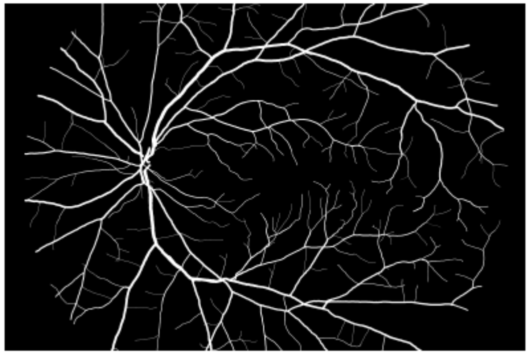
Step 1: Pre-process and segmentation





We adapted an *unsupervised* algorithm, originally developed for segmenting images of **cultured neural networks**.

Manual segmentation

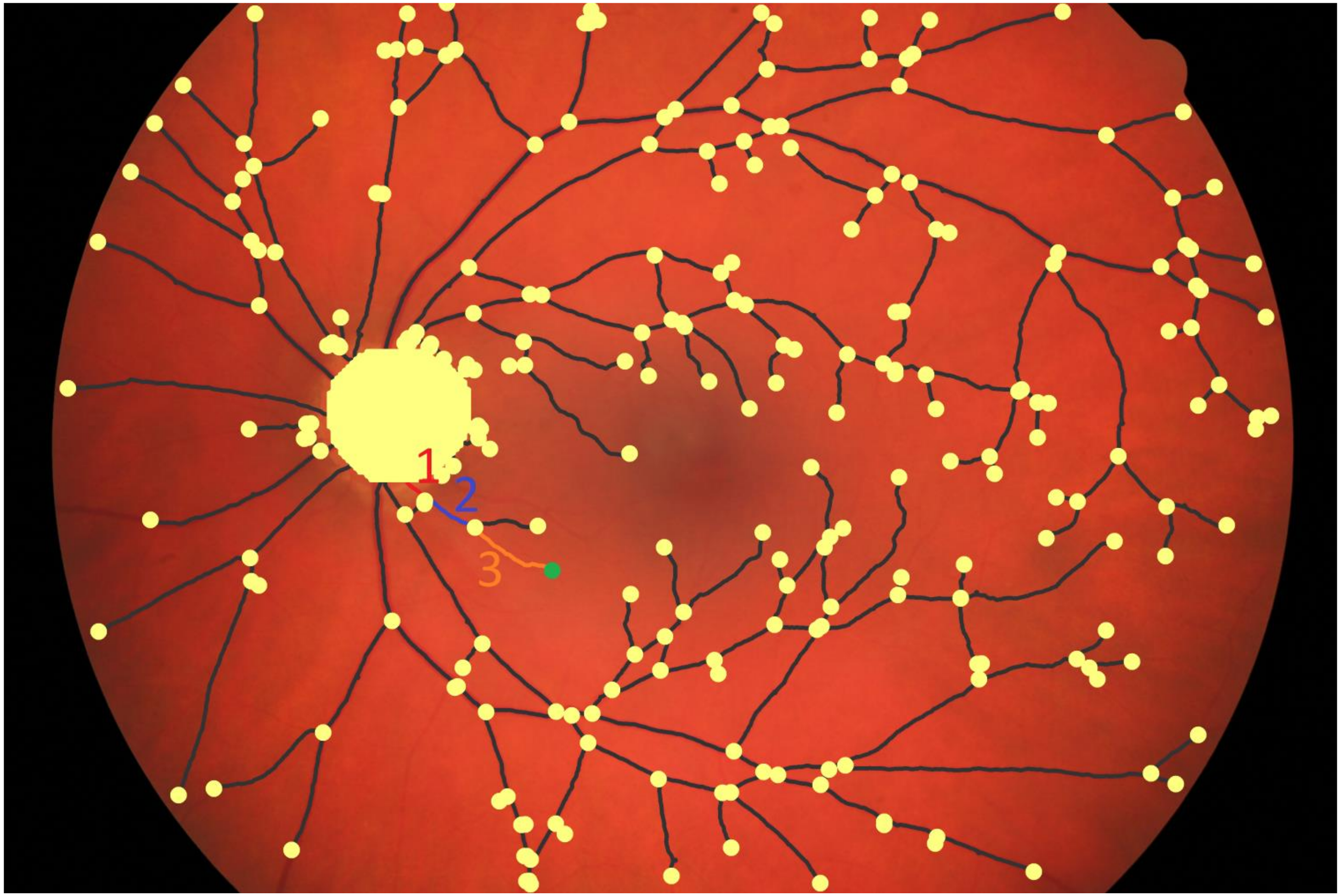


D. Santos-Sierra, I. Sendiña-Nadal, I. Leyva et al. Cytometry Part A. 87, 513 (2015).

P. Amil, F. Reyes-Manzano, L. Guzmán-Vargas, I. Sendiña-Nadal, C. Masoller, "*Network-based features for retinal fundus vessel structure analysis*", PLoS ONE 14, e0220132 (2019).

cristina.masoller@upc.edu @cristinamasoll1

Step 2: extract the network (identification of the optical nerve, nodes and links and assign weights to the links).

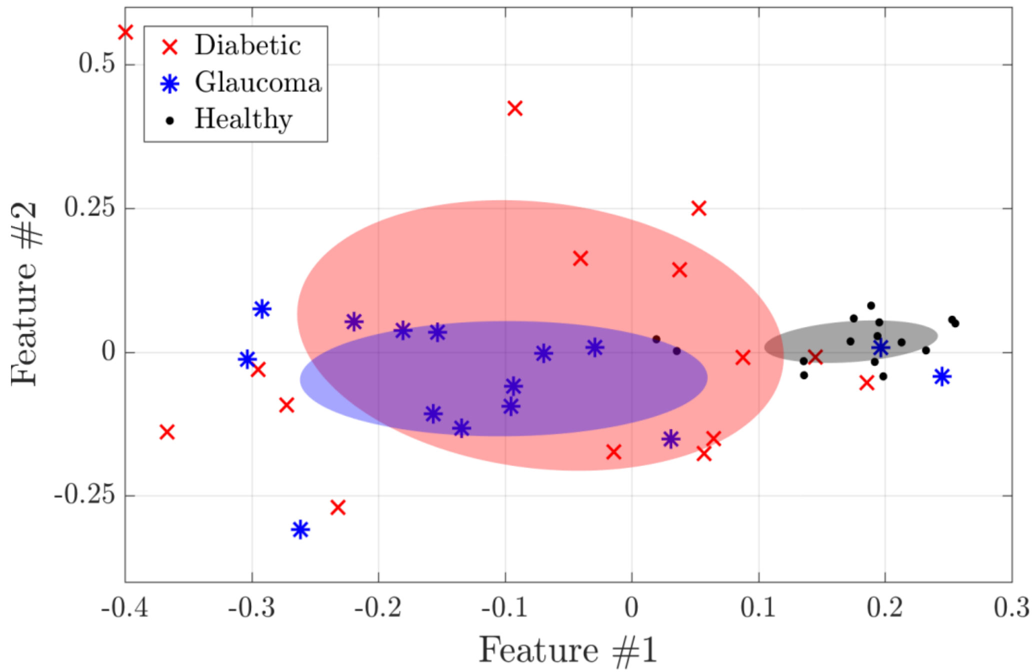


cristina.masoller@upc.edu @cristinamasoll1

Steps 3 and 4: Compare the networks extracted from different images and classify the images.

- $\{p_{i,j}\}$: distances between probability distributions that characterize the networks obtained from images i and j.
- We used nonlinear dimensionality reduction (*Isomap*) to reduce the set of 45x45 $\{p_{i,j}\}$ values to only two features.

Distance distribution to the central node in the *manual* segmentation

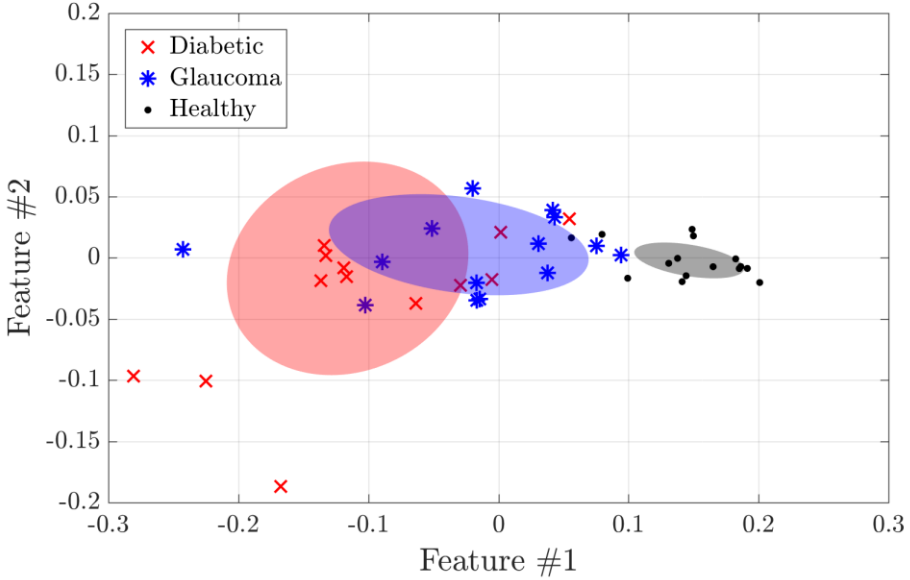


P. Amil et al, Network-based features for retinal fundus vessel structure analysis, PLoS ONE 14 e0220132 (2019).

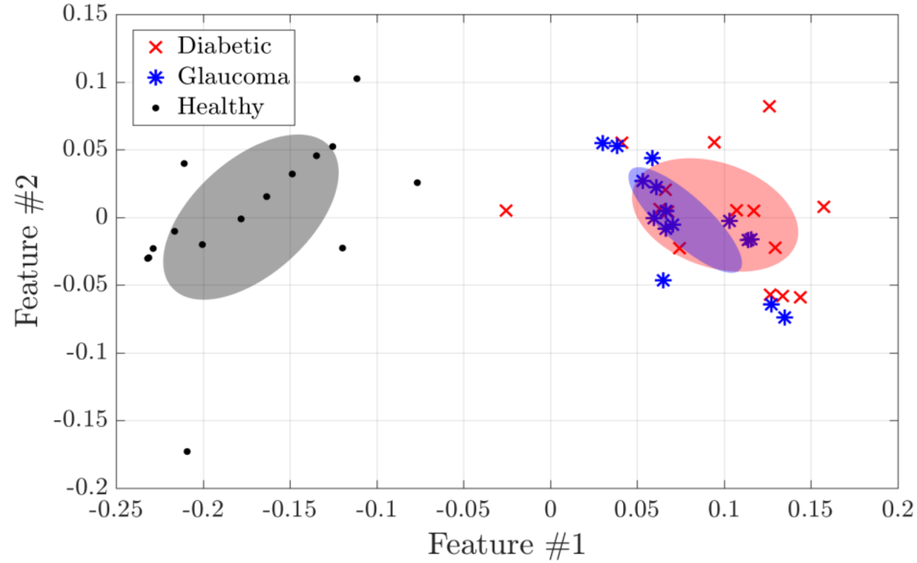
cristina.masoller@upc.edu @cristinamasoll1

Performance of network features in the *manual* segmentation

Distribution of weights along the shortest path to central node



Distribution of weighted degrees

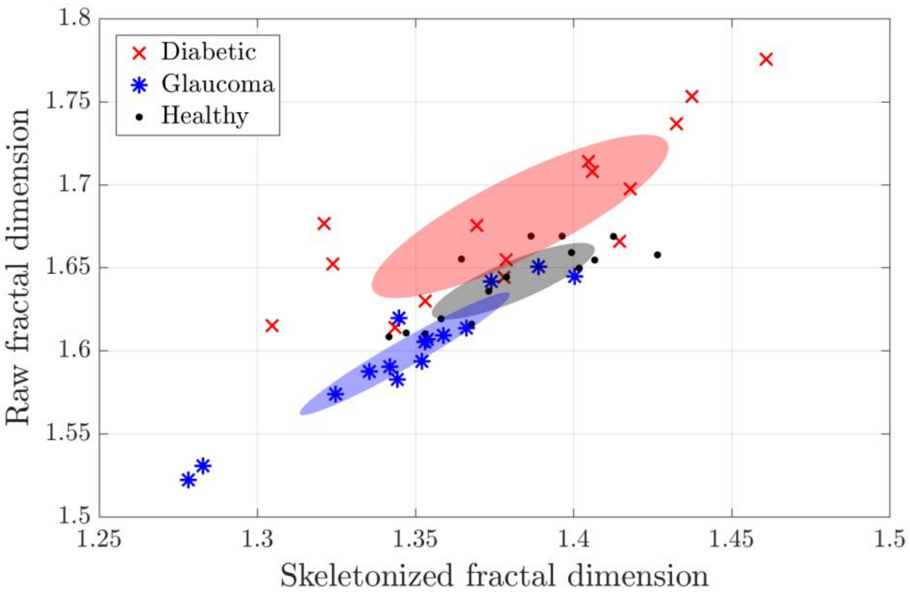


P. Amil et al, Network-based features for retinal fundus vessel structure analysis, PLoS ONE 14 e0220132 (2019).

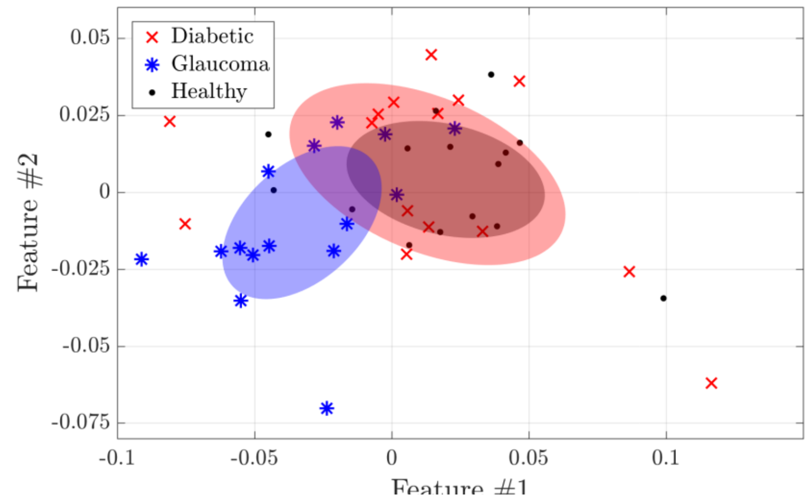
cristina.masoller@upc.edu @cristinamasoll1

In the automated segmentation

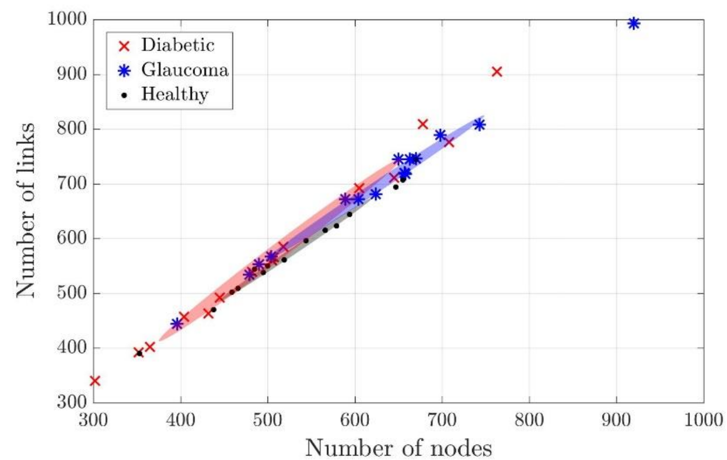
Fractal dimension $D = \lim_{\varepsilon \to 0} \frac{\log(N(\varepsilon))}{\log(1/\varepsilon)}$



Mean weight distribution along the shortest path to central node



Simple network features do not differentiate



P. Amil et al, Network-based features for retinal fundus vessel structure analysis, PLoS ONE 14 e0220132 (2019).

cristina.masoller@upc.edu @cristinamasoll1

Summary

- Bivariate and multivariate analyses uncover inter-relationships in datasets
- Different similarity measures are available for inferring the connectivity of a complex system from observations.
- Different methods can uncover different properties.
- Methods can be adapted to analyze different types of data.

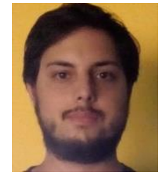
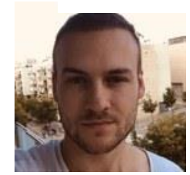
cristina.masoller@upc.edu @cristinamasoll1

Thanks!







A. Aragoneses et al., *“Unveiling temporal correlations characteristic to phase transition in the intensity of fibre laser radiation*”, Phys. Rev. Lett. 116, 033902 (2016).

J. A. Reinoso et. al, “*Emergence of spike correlations in periodically forced excitable systems*”, Phys. Rev. E. 94, 032218 (2016).

J. Tiana-Alsina et. al, “*Comparing the dynamics of periodically forced lasers and neurons*”, New J. of Phys. 21, 103039 (2019).

M. Masoliver and C. Masoller, “*Neuronal coupling benefits the encoding of weak periodic signals in symbolic spike patterns*”, Commun. Nonlinear Sci. Numer. Simulat. 88, 105023 (2020).

C. Quintero-Quiroz et al., “*Differentiating resting brain states using ordinal symbolic analysis*”, Chaos 28, 106307 (2018).

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

G. Tirabassi and C. Masoller, “*Entropy-based early detection of critical transitions in spatial vegetation fields*”, PNAS 120, e2215667120 (2022).

P. Amil et al., “*Network-based features for retinal fundus vessel structure analysis*”, PLoS ONE 14, e0220132 (2019).

P. Amil et al., “*Machine learning algorithms for predicting the amplitude of chaotic laser pulses*”, Chaos 29, 113111 (2019).

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