Network-based data analysis tools for identifying and characterizing regime transitions in complex systems

Cristina Masoller
Terrassa, Barcelona, Spain

Cristina.masoller@upc.edu

www.fisica.edu.uy/~cris

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

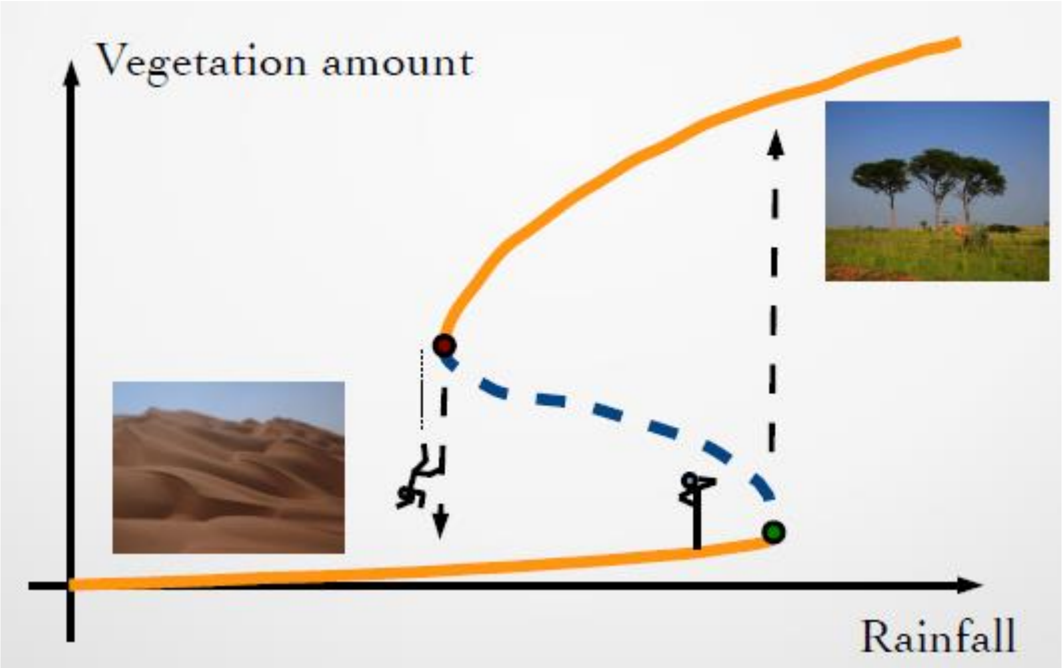
Campus d'Excel·lència Internacional

NDES 2017
Zernez, June 2017

ICREA

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

Tipping points in ecosystems



Bangladesh,
Nature 2014

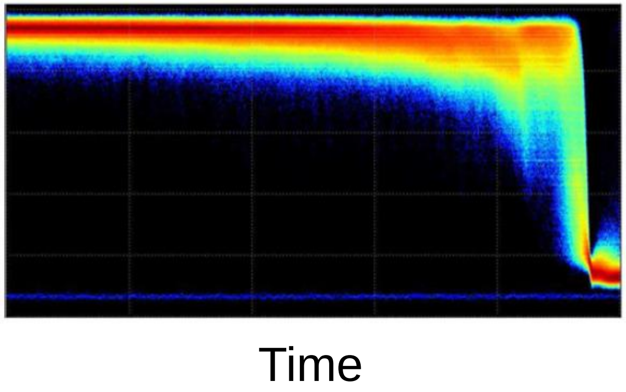
Is there a way to quantify how close we are to the transition point?

Goal: to develop reliable early warning indicators

Dynamical transitions in optical systems

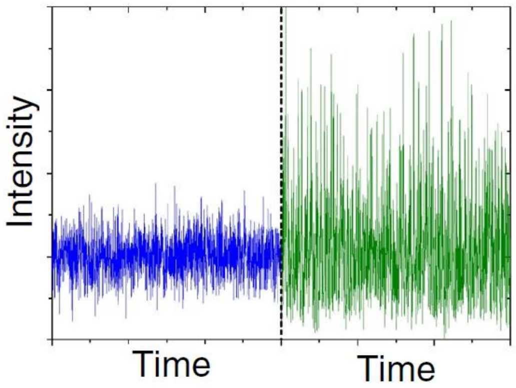
- **Polarization switching**

  Semiconductor laser output as the pump current increases



- **Transition to turbulence**

  Fiber laser output as the pump power increases

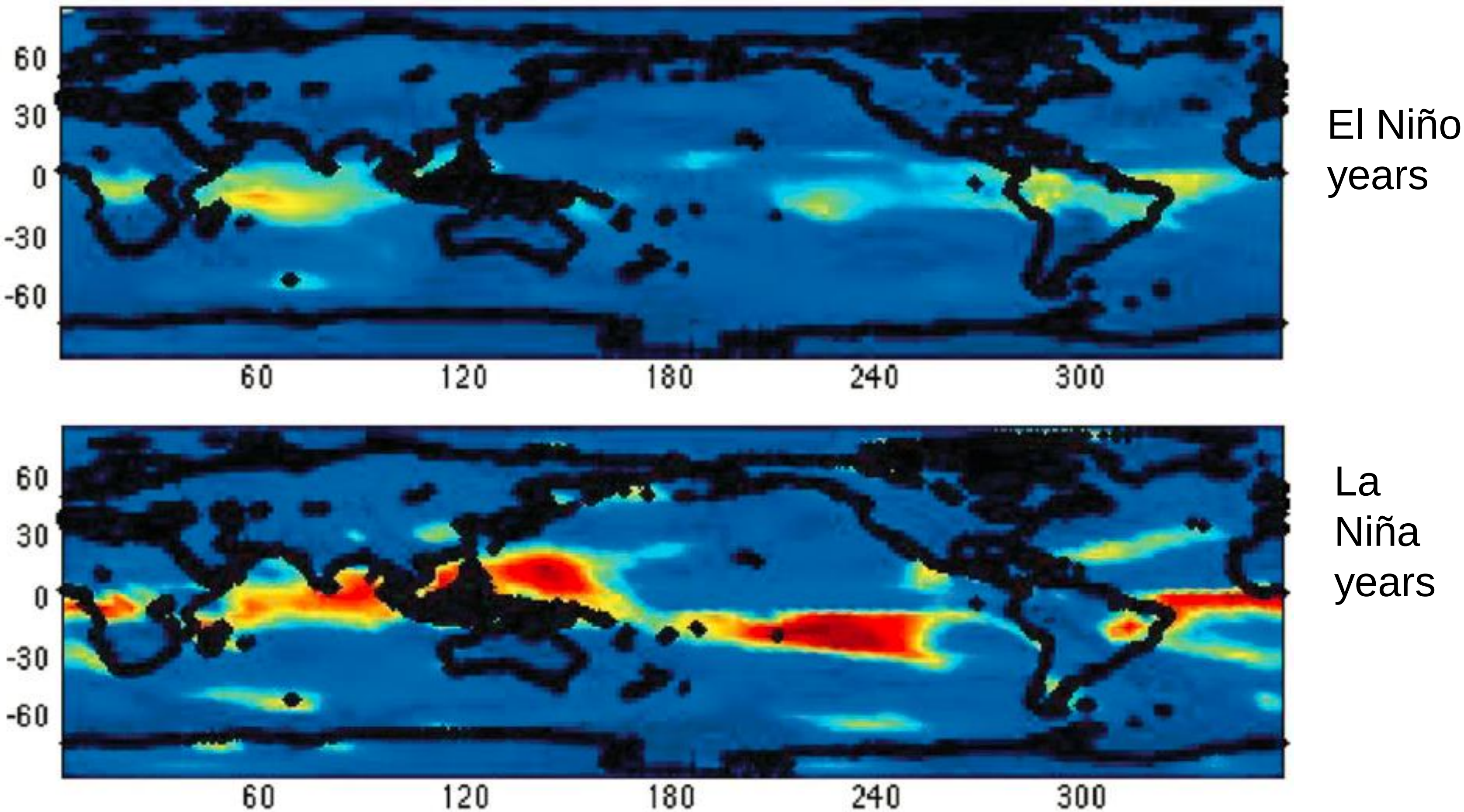


Goal: convince you that

- Network-based data analysis tools provide new insights into these phenomena.
- Optical data can be useful for testing novel analysis tools.

How to detect transitions in complex networks?

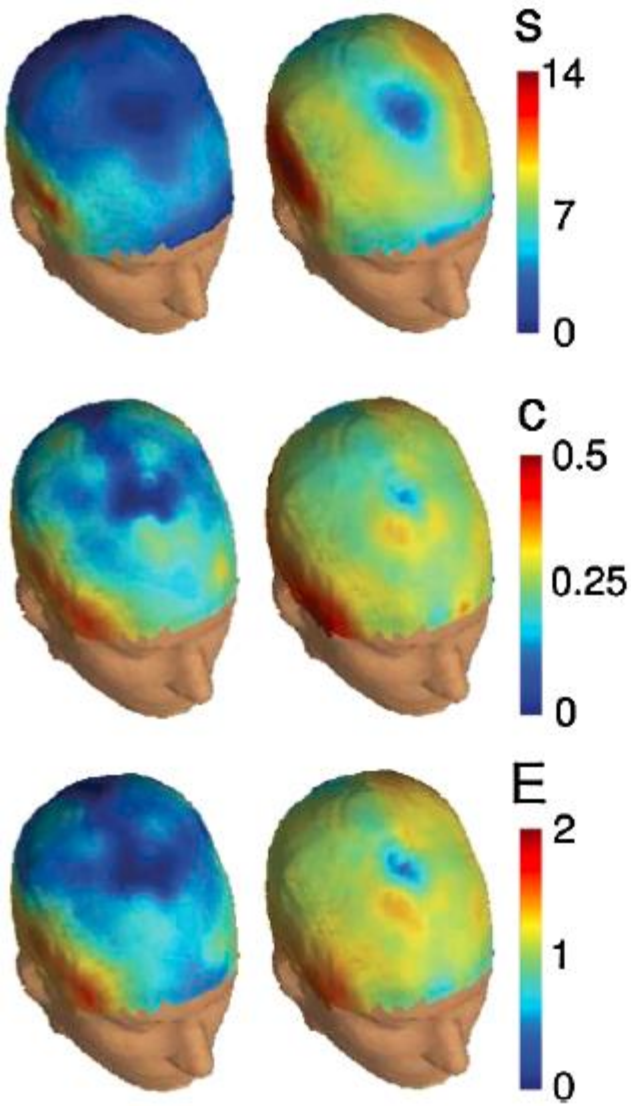
- How to compare time-evolving climate networks?



Tsonis and Swanson, PRL 100, 228502 (2008)

Functional brain networks

Healthy subjects Epileptic patients



averaged node Strength

averaged weighted Clustering coefficient

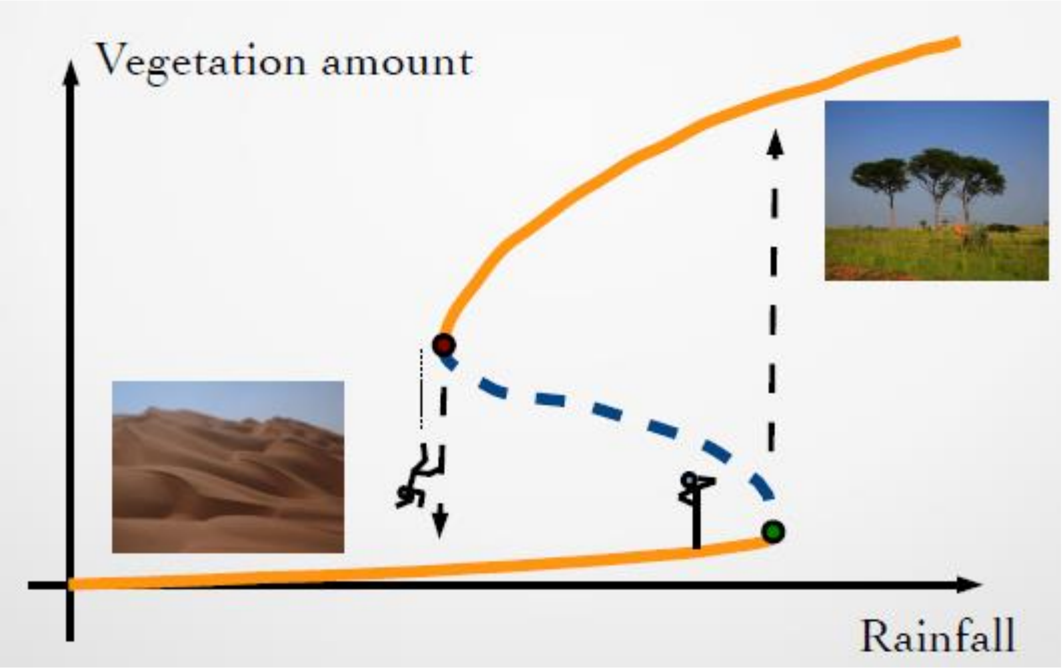
global Efficiency

Goal: a tool to quantify network dissimilarities.

M. Chavez et al., PRL 104, 118701 (2010)

Outline

- Early-warning indicators of desertification transition
- Quantifying sudden changes using symbolic networks
- Emergence of temporal correlations in the optical laminar-turbulence transition
- Quantifying network dissimilarities



Early-warning indicators of desertification transition

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Campus d'Excel·lència Internacional

Early warning indicators

- Bifurcation → eigenvalue 'a' with 0 real part
- → long recovery time of perturbations
- **Critical Slowing Down** (CSD)
- CSD → High autocorrelation, variance, spatial correlation, etc.

- Can we use network tools to detect tipping points?
- Yes!

Interaction network based early warning indicators for the Atlantic MOC collapse

Mirjam van der Mheen,[1] Henk A. Dijkstra,[1] Avi Gozolchiani,[2] Matthijs den Toom,[1] Qingyi Feng,[1] Jürgen Kurths,[3] and Emilio Hernandez-Garcia[4]

Received 14 March 2013; revised 26 April 2013; accepted 26 April 2013; published 4 June 2013.

GEOPHYSICAL RESEARCH LETTERS, VOL. 40, 2714–2719, doi:10.1002/grl.50515, 2013

Desertification transition: A simple model

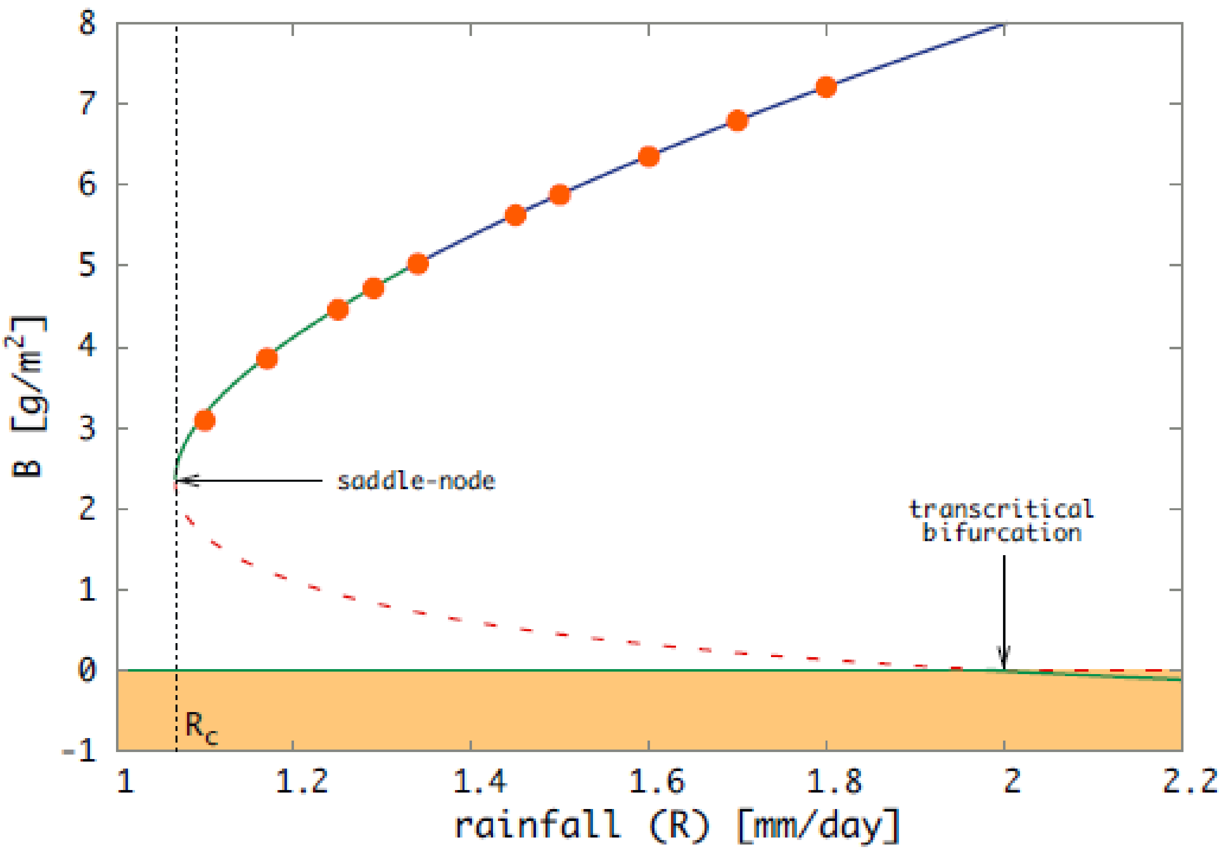
$$dw_t = \left(R - \frac{w}{\tau_w} - \Lambda w B + D \Delta w \right) dt + \sigma_w dW_t$$

$$dB_t = \left(\rho B \left(\frac{w}{w_0} - \frac{B}{B_0} \right) - \mu \frac{B}{B + B_o} + D \Delta B \right) dt + \sigma_B dW_t$$

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

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

Saddle-node bifurcation



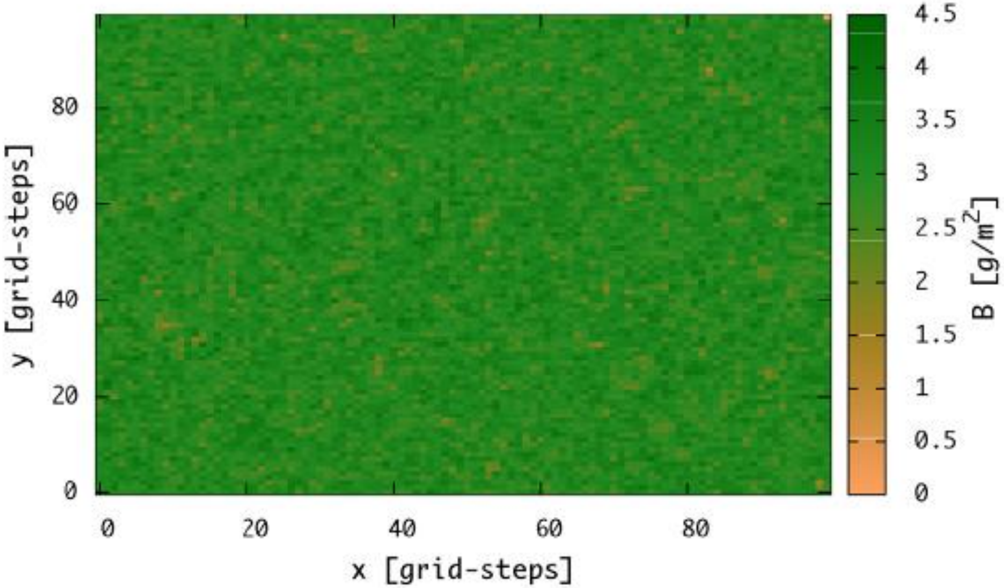
R<R_c: only desert-like solution (B=0)

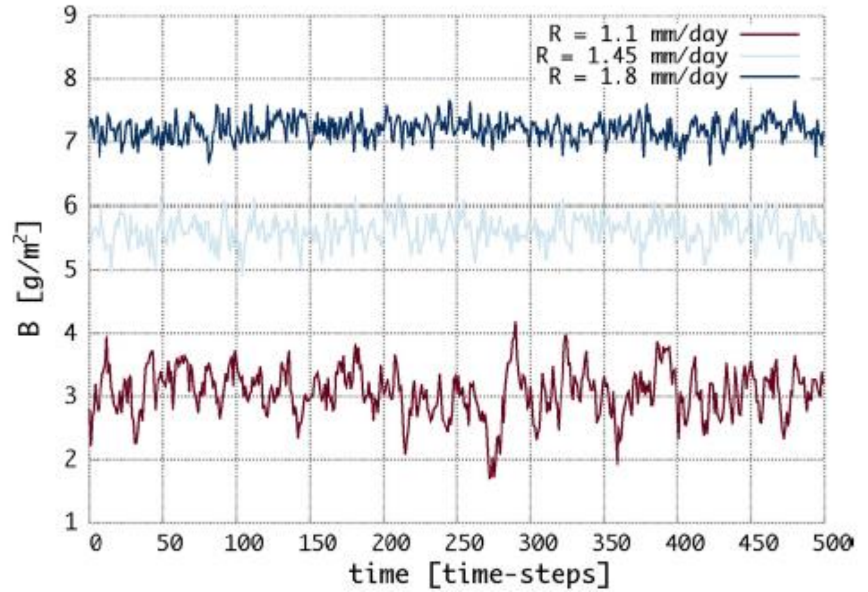
R_c = 1.067 mm/day

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

Biomass B at T=5 days when R=1.1 mm/day





100 m x 100 m → 100x100 = 10^4 grid cells
Simulation time 5 days in 500 time steps
Periodic boundary conditions

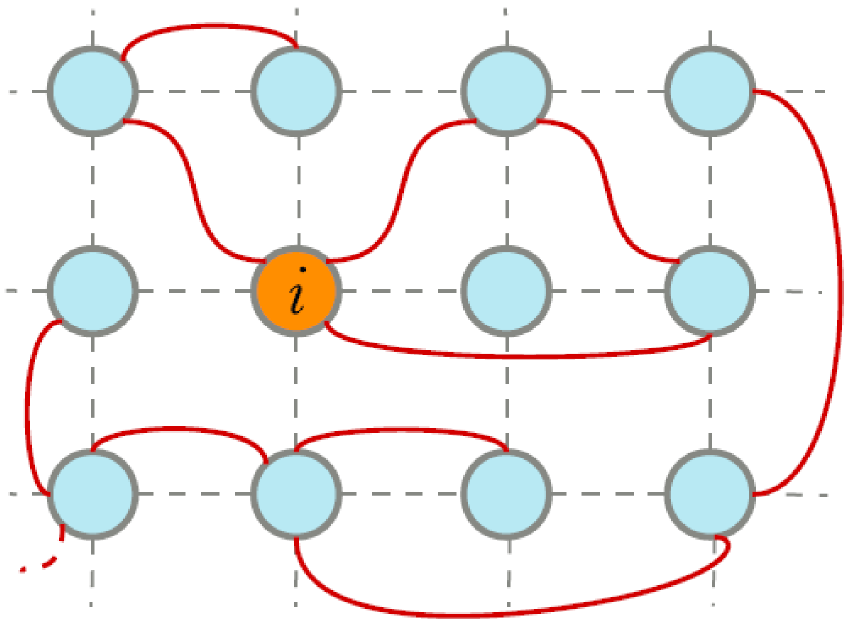
Correlation Network

$$A_{ij} = \mathrm{H}(|\mathcal{C}(B_i, B_j)| - \theta)$$

Adjacency matrix

Zero-lagged cross-correlation

Threshold $\theta=0.2$ gives $p<0.05$



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

Network analysis

- **Degree** (number of links of a node)

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

- **Assortativity** (average degree of the neighbors of a node)

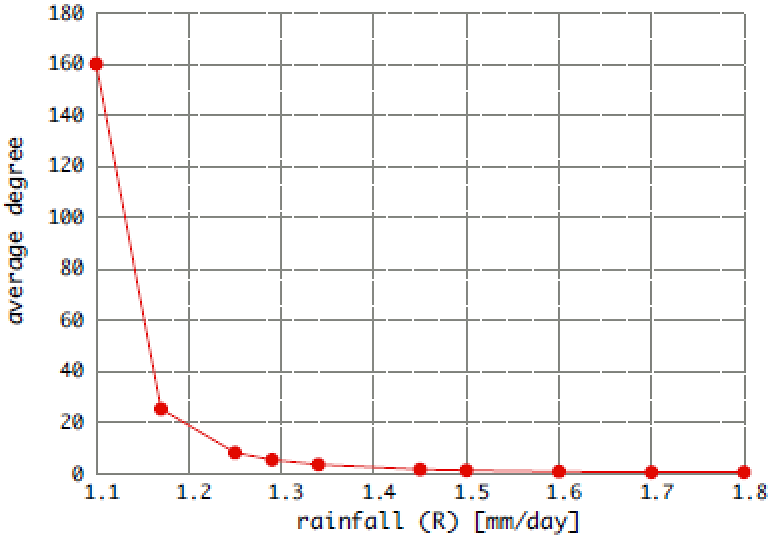
$$a_i \equiv \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j$$

- **Clustering** (fraction of neighbors of a node that are also neighbors among them)

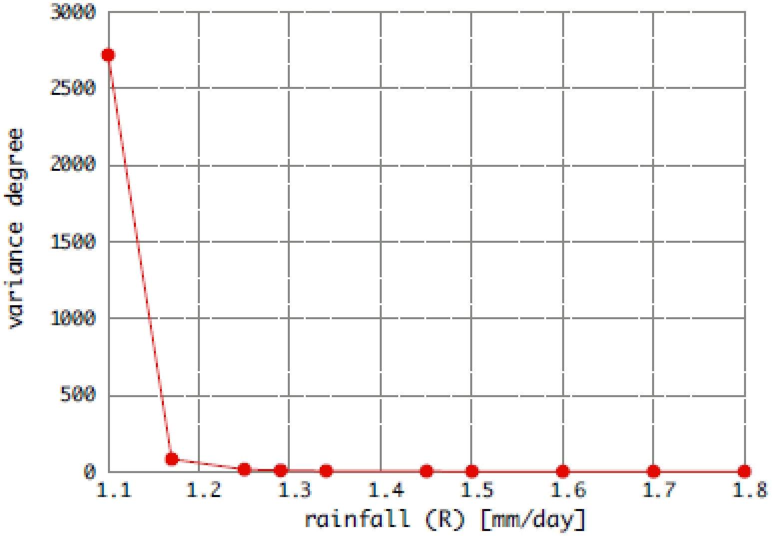
$$c_i \equiv \frac{1}{k_i(k_i - 1)} \sum_{j=1}^{N} \sum_{l=1}^{N} A_{ij} A_{jl} A_{li}$$

Results

Mean degree



Standard deviation of the degree distribution

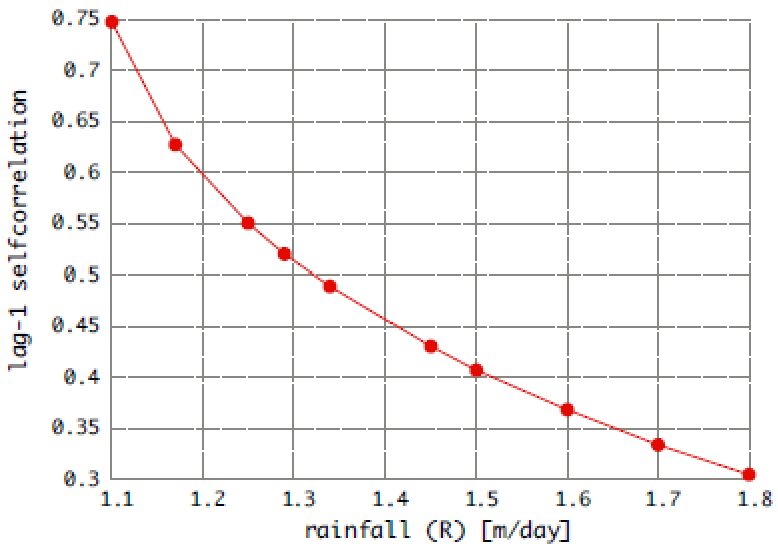


Sharp increase close to the transition captures the emergence of spatial correlations

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

Comparison with classical indicators

Spatially averaged lag-1 autocorrelation, calculated at the final simulation time

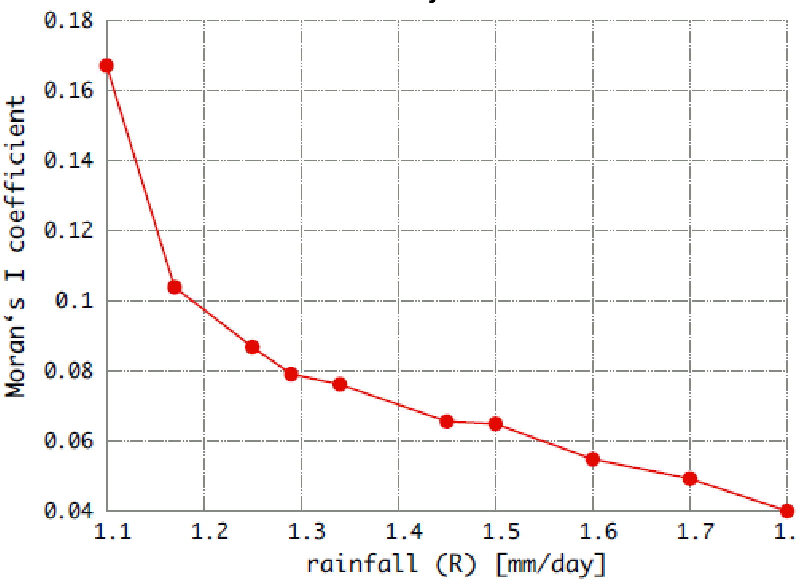


Smooth increase does not capture the vicinity to the transition point

Moran's coefficient

$$I \equiv \frac{N}{\sum_{ij} g_{ij}} \frac{\sum_{ij} g_{ij}(B_i - \bar{B})(B_j - \bar{B})}{\sum_i (B_i - \bar{B})^2}$$

$g_{ij} = 1$ if i and j are adjacent grid cells, else $g_{ij} = 0$

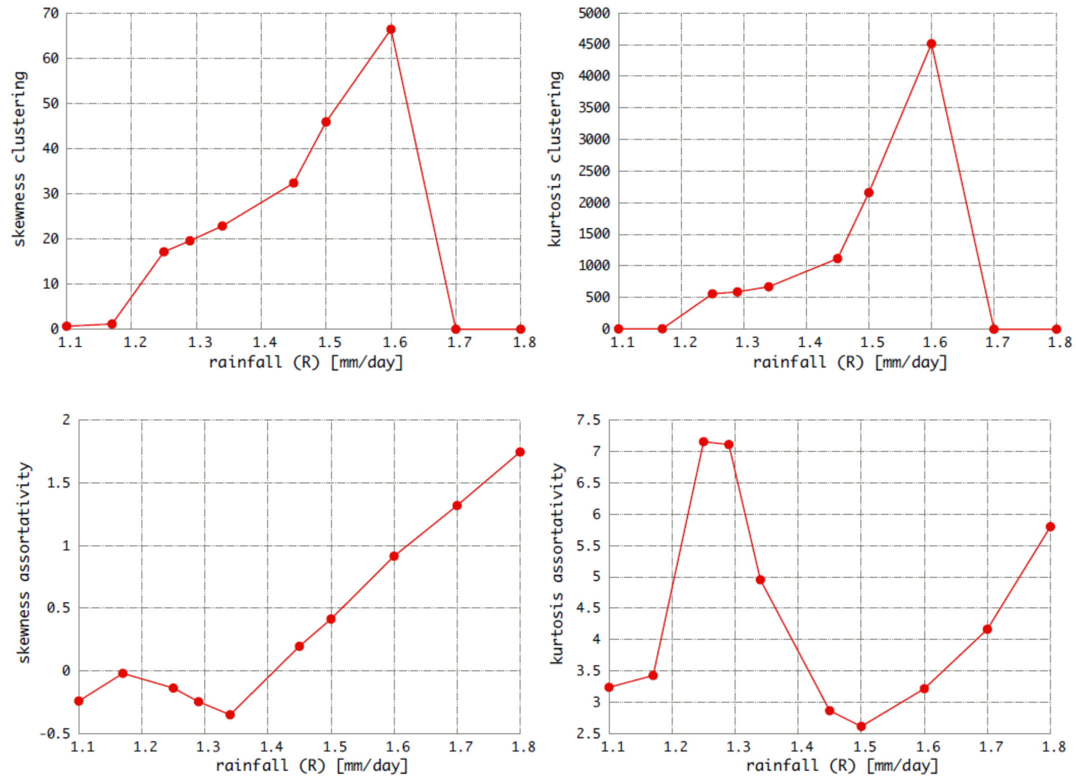


Network-based indicators

skewness | kurtosis

clustering

assortativity

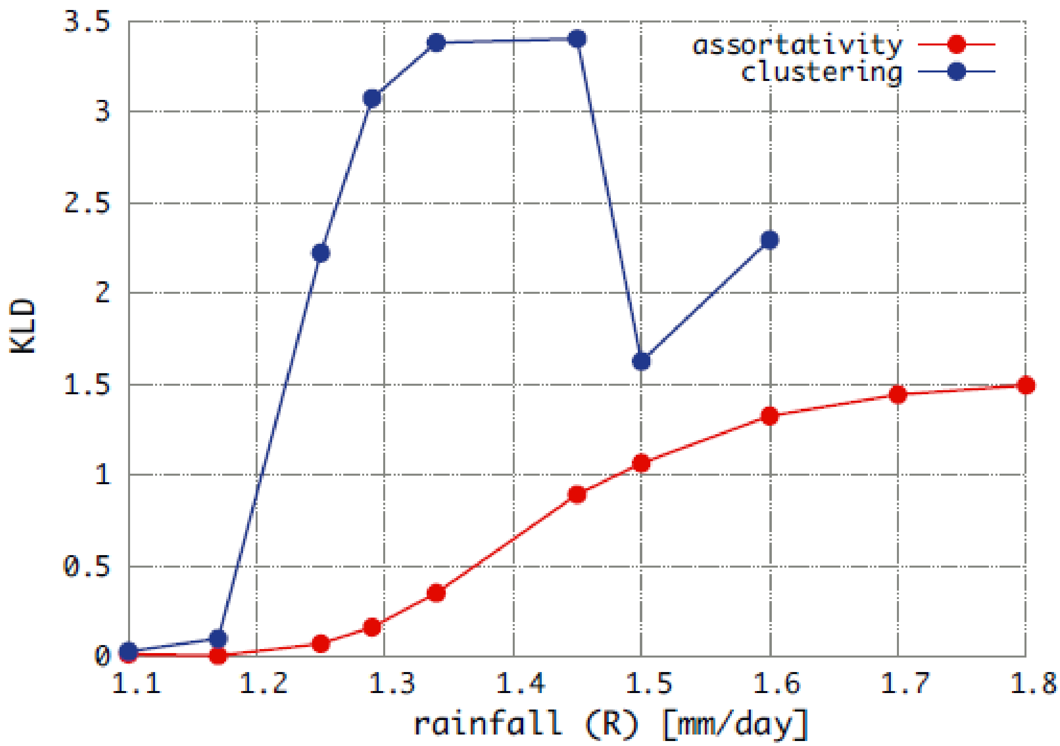


"Gaussianisation" of the clustering and of the assortativity distributions when approaching the tipping point

How to quantify "Gaussianisation"?

Kullback–Leibler Distance (KLD) between 2 PDFs

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

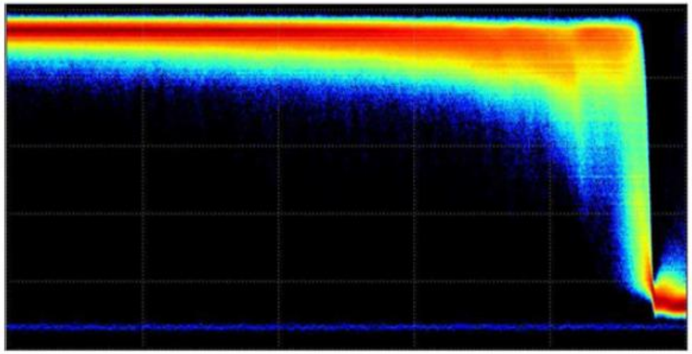


G. Tirabassi et al., Ecological Complexity 19, 148 (2014)

Summary Part 1

- Network-based indicators can identify desertification transition in advance.
- Perform better than the classical ones.
- Open issue: the Gaussianisation might a model-specific feature.

G. Tirabassi et al., *Interaction network based early-warning indicators of vegetation transitions*, Ecological Complexity 19, 148 (2014)



Quantifying sudden changes using symbolic networks

- “optical big data”: provides new insight & is useful for testing novel diagnostic tools

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d’Excel·lència Internacional

Experimental data from INLN & Bangor University
(S. Barland & Y. Hong)

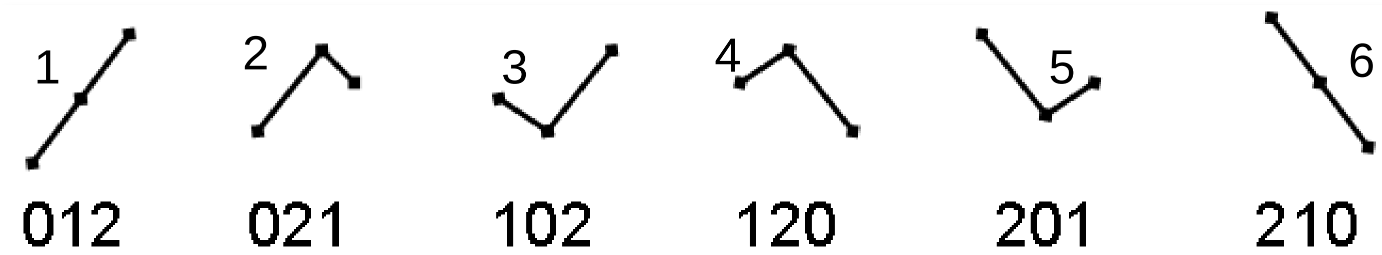
UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

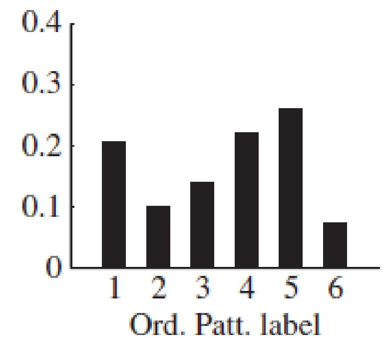
Campus d'Excel·lència Internacional

Method of nonlinear **symbolic** time-series analysis: ordinal patterns

Brandt & Pompe, PRL 88, 174102 (2002)

- $X = \{\ldots x_i, x_{i+1}, x_{i+2}, \ldots\}$





The OP probabilities allow identifying <u>more expressed and/or infrequent patterns</u> in the <u>*order*</u> of the sequence of data values.

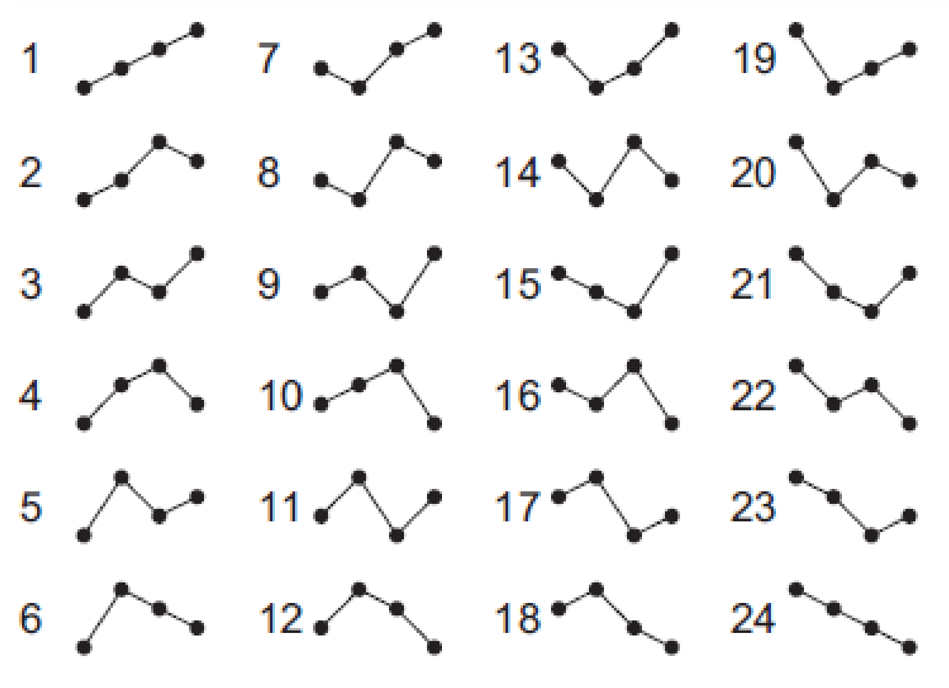
Random data? (OPs equally probable)

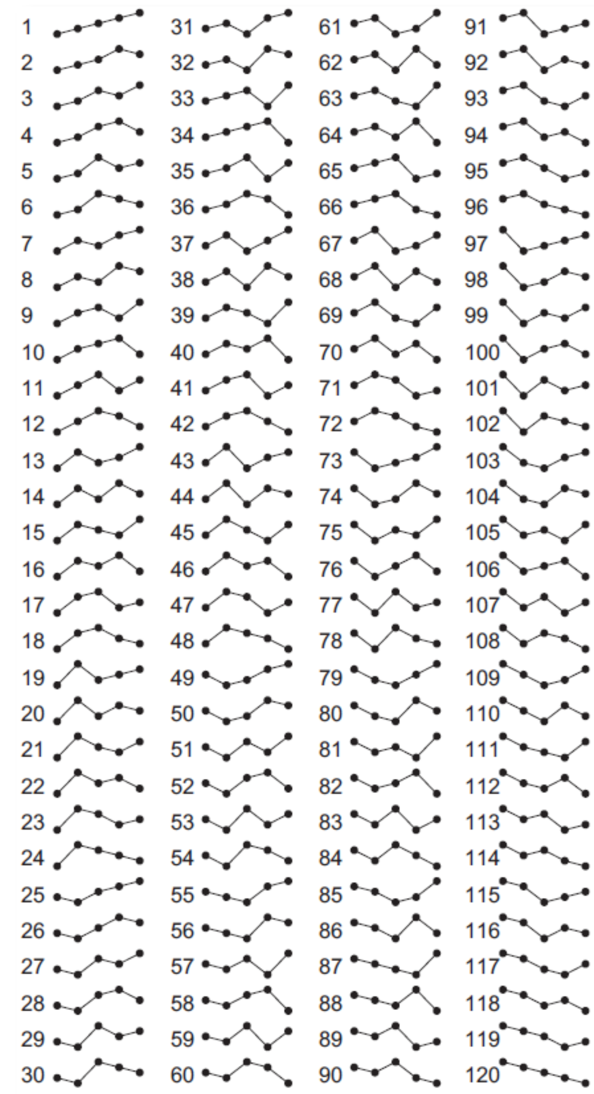
– Advantage: the probabilities uncover temporal correlations.

– Drawback: we lose information about the actual values.

$\Rightarrow$ Ordinal analysis gives **complementary information** to that gained with other analysis tools.

The number of patterns increases as D!



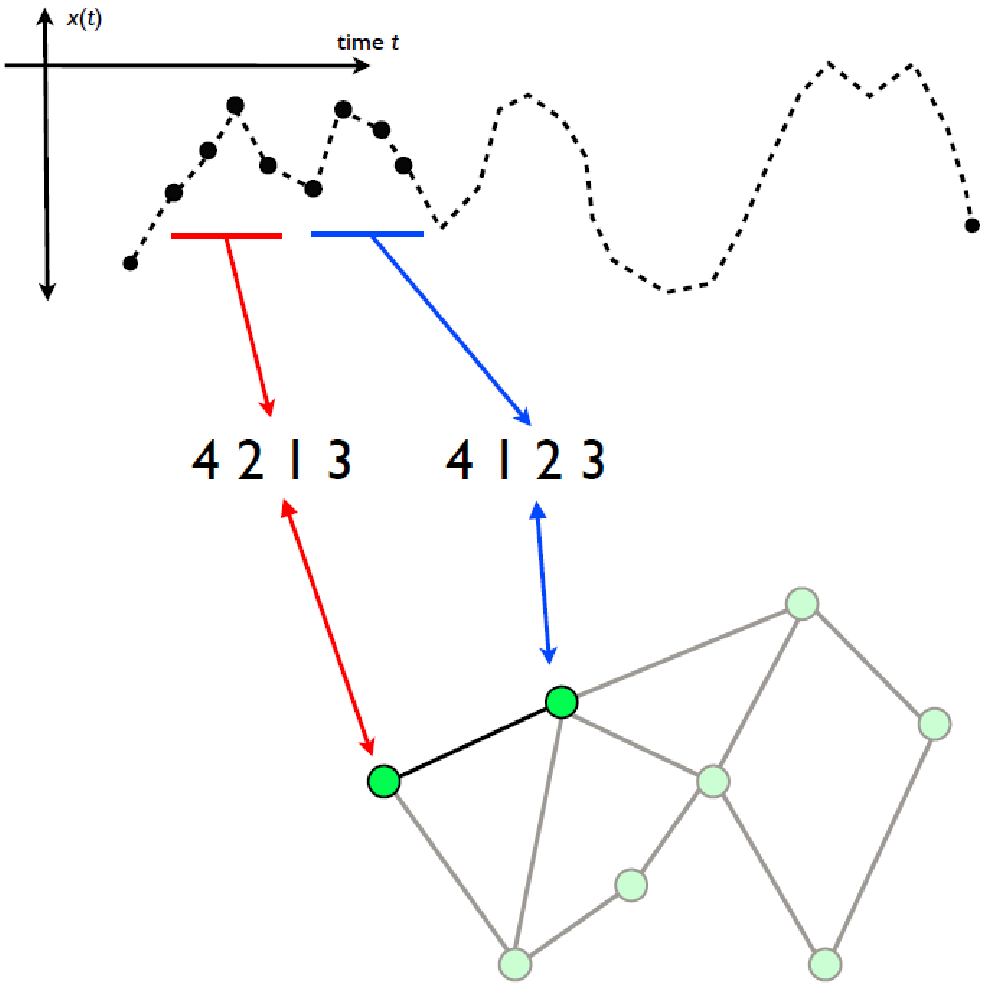


Opportunity: turn a time-series into a network by using the patterns as the "nodes" of the network.

U. Parlitz et al. / Computers in Biology and Medicine 42 (2012) 319–327

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Campus d'Excel·lència Internacional

The network nodes are the "ordinal patterns", and the links?



- The links are defined in terms of the probability of pattern "β" occurring after pattern "α".
- <u>Weighs of nodes</u>: the probabilities of the patterns ($\sum_i p_i=1$).
- <u>Weights of links</u>: the probabilities of the transitions ($\sum_j w_{ij}=1 \ \forall i$).

$\Rightarrow$ ***Weighted and directed network***

Adapted from M. Small (The University of Western Australia)

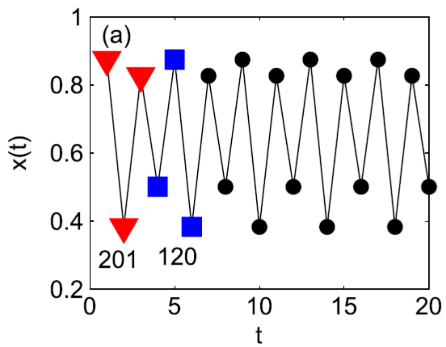
Three network-based diagnostic tools

- Entropy computed from the weights of the nodes (**permutation entropy**)

$$s_p = -\sum p_i \log p_i$$

- Entropy computed from weights of the links (**transition probabilities**, '01'→ '01', '01'→ '10', etc.)

$$w_{ij} = \frac{\sum_{t=1}^{L-1} n\left[s(t) = i,\, s(t+1) = j\right]}{\sum_{t=1}^{L-1} n\left[s(t) = i\right]}$$



- Asymmetry coefficient: normalized difference of transition probabilities, P('01'→ '10') - P('10'→ '01'), etc.

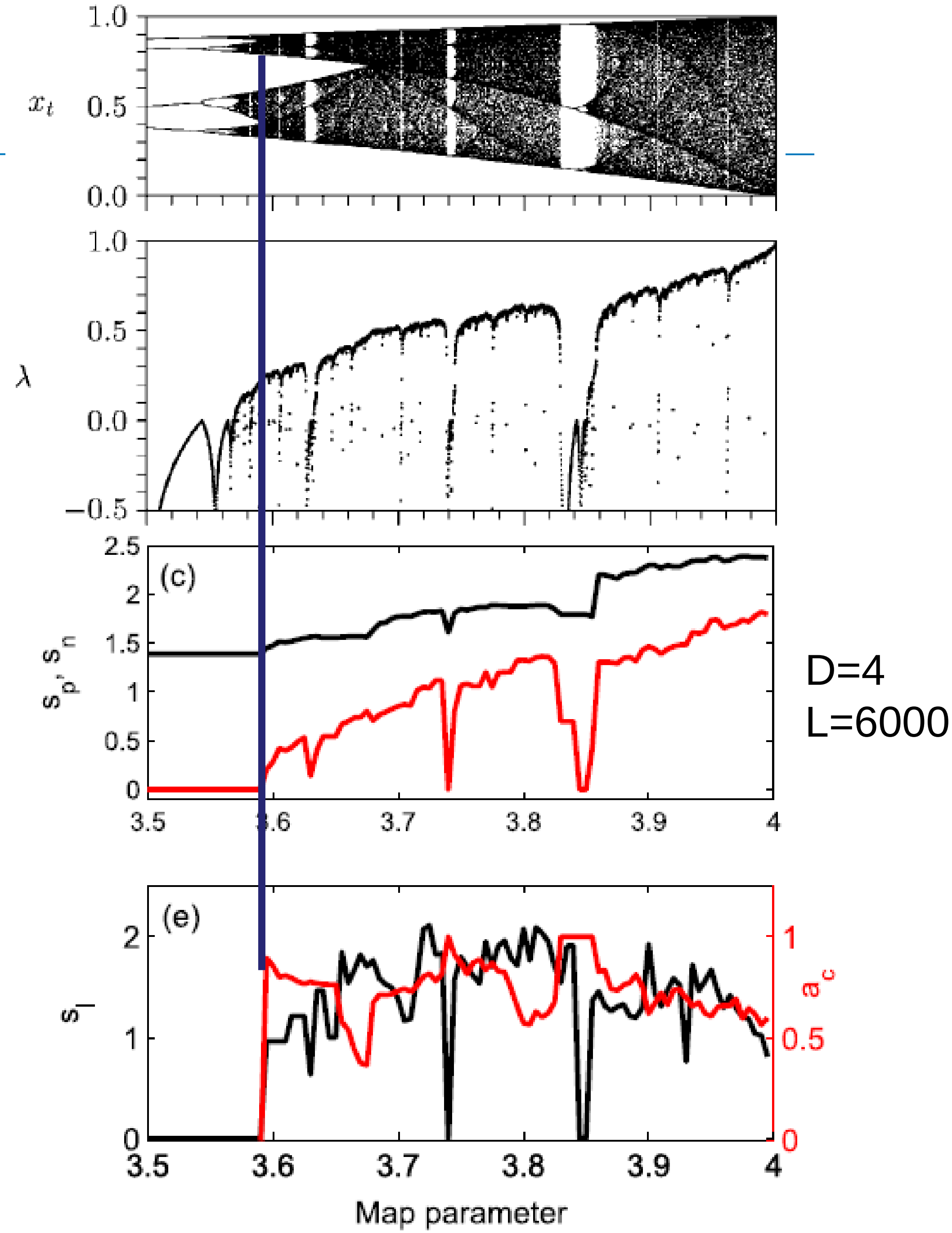
$$a_c = \frac{\sum_i \sum_{j \neq i} \left| w_{ij} - w_{ji} \right|}{\sum_i \sum_{j \neq i} \left(w_{ij} + w_{ji} \right)}$$

(0 in a fully symmetric network; 1 in a fully directed network)

First test the method with synthetic data: the logistic map

- **x(i+1)=r x(i)[1-x(i)]**

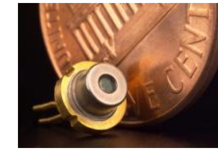
⇒ Detects transitions (not bifurcations) that are not detected by Lyapunov exponent.



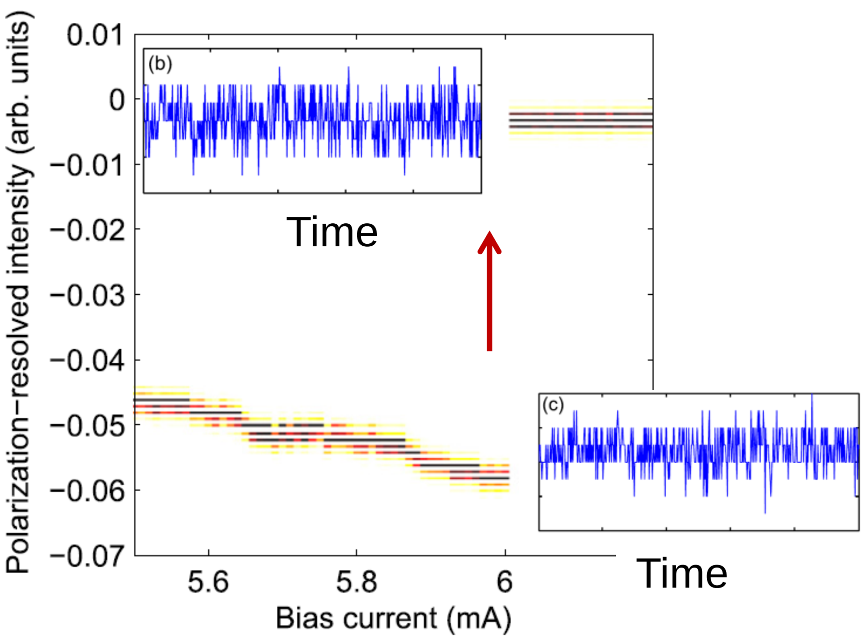
C. Masoller et al, New J. Phys. 17, 023068 (2015)

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Campus d'Excel·lència Internacional

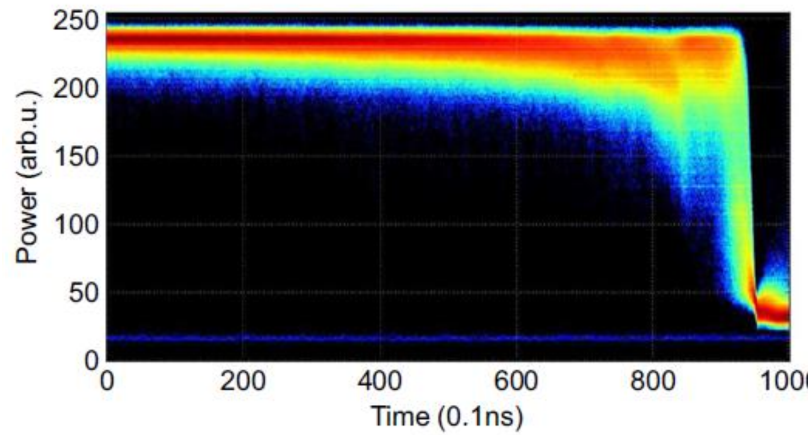
A laser polarization-resolved intensity: two sets of experiments



- Time series recorded with laser current constant in time.
- Record the <u>turn-on</u> of the orthogonal mode.



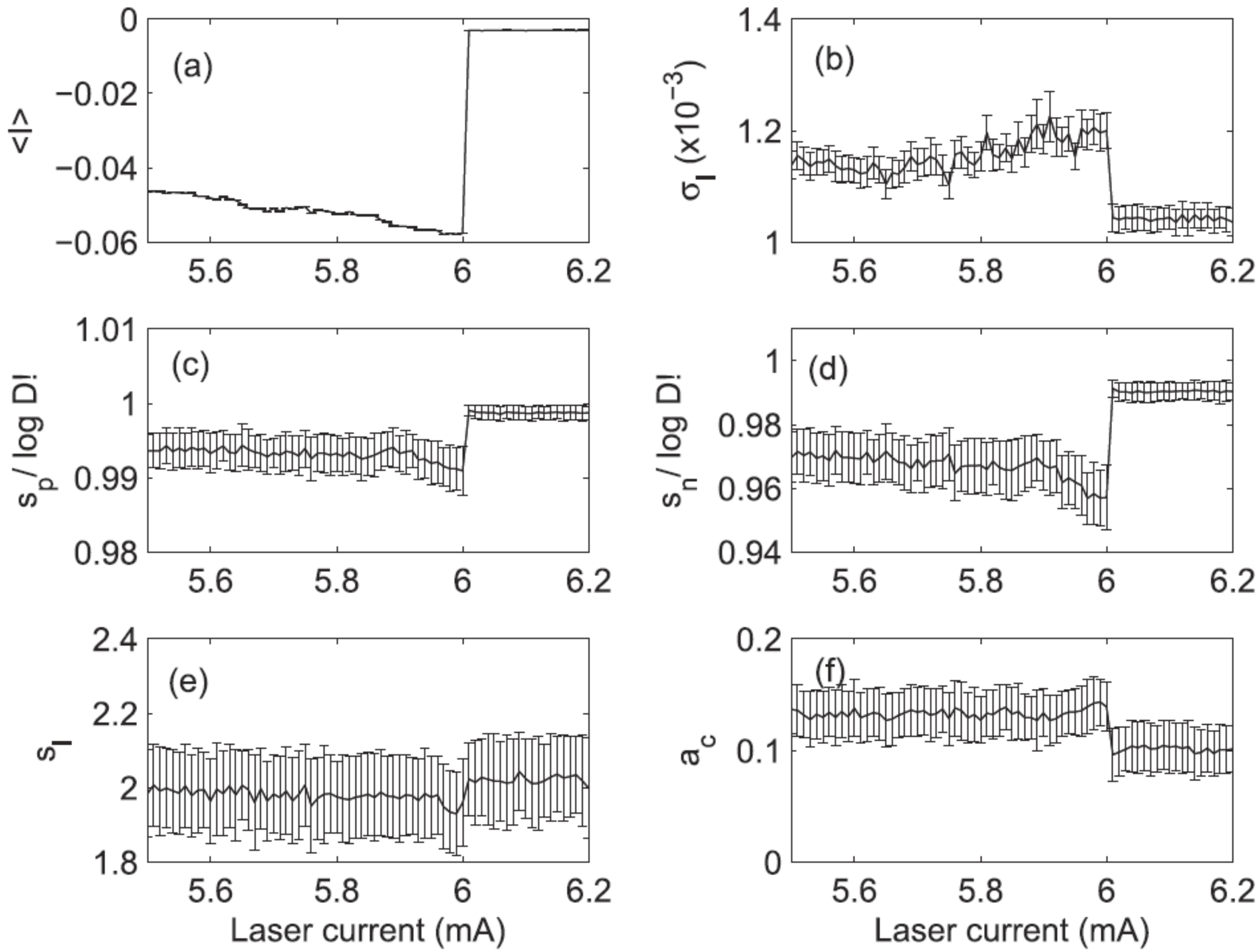
- Time series recorded with laser current varying in time.
- Record the <u>turn-off</u> of the fundamental mode.



Is it possible to anticipate the PS?

No if the mechanisms that trigger the PS are fully stochastic.

Results for constant pump current & turn-on of the orthogonal mode



- ⇒ Despite of the stochasticity of the time-series, the measures "anticipate" the PS.
- ⇒ Deterministic mechanisms involved.

Error bars computed from 100 non-overlapping windows with L=1000 data points each. Length of the pattern D=3.

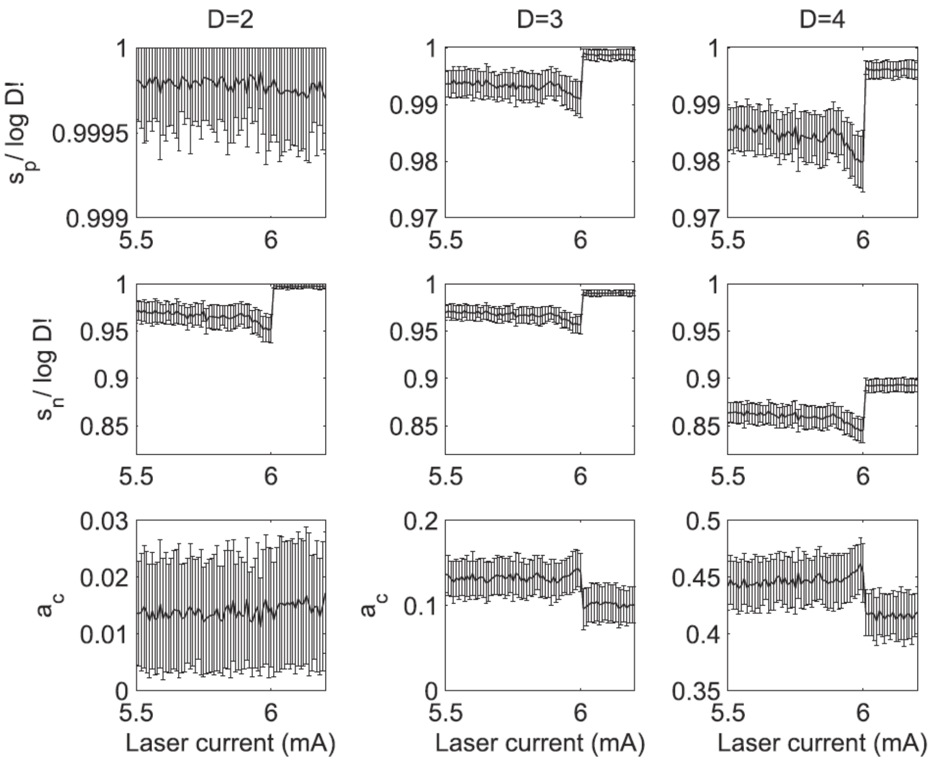
C. Masoller et al, New J. Phys. 17 (2015) 023068

Influence of the length of the pattern (D) and of length of the time-series (L)

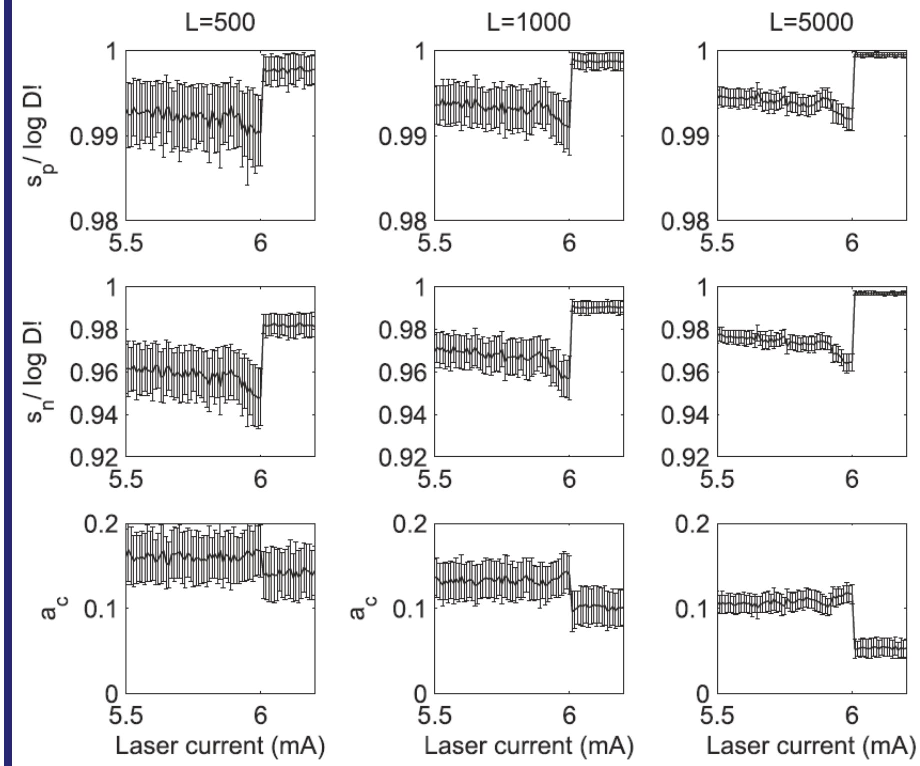
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

L=1000



D=3



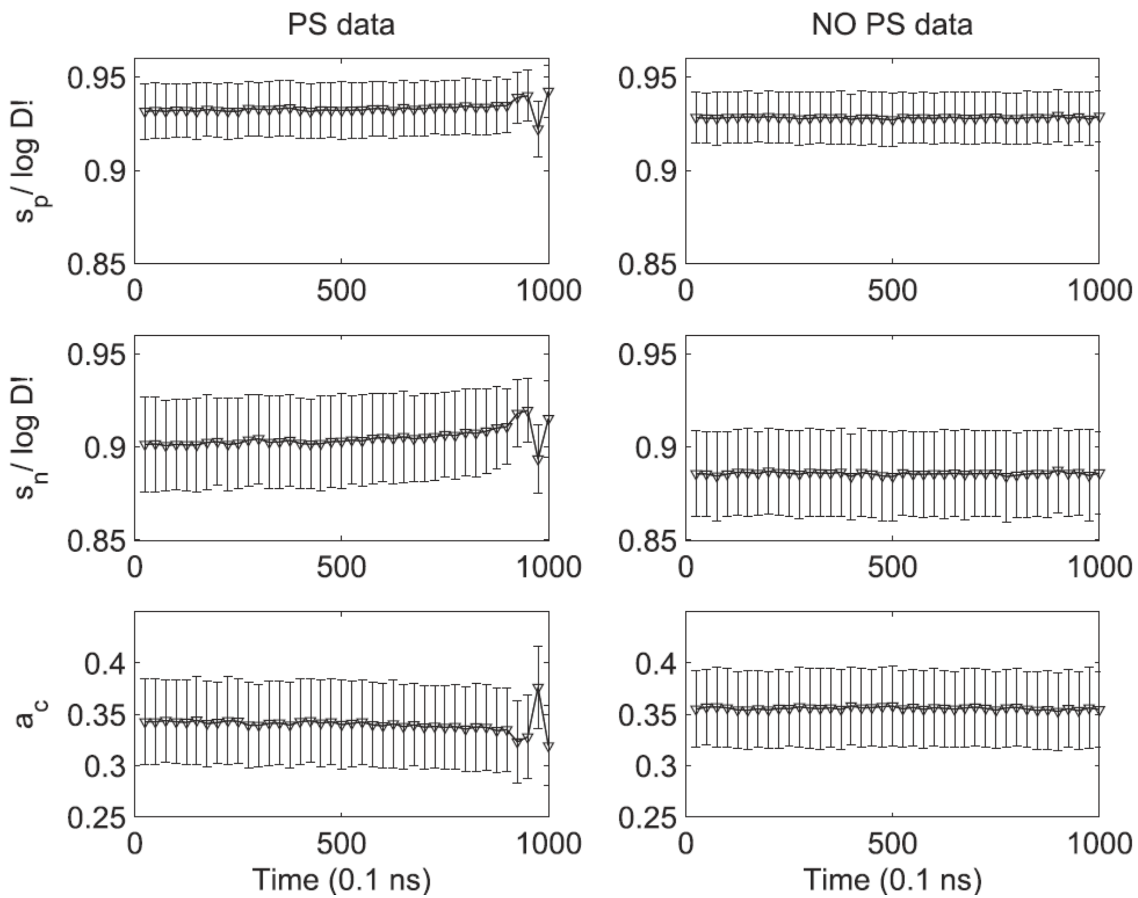
$\Rightarrow$ Transition detected even for short dataset (L=500 with D=3).

C. Masoller et al, New J. Phys. 17 (2015) 023068

Results for time-varying pump current & turn-off of the fundamental mode

Slightly different optical feedback conditions result in PS or no PS.

Analysis done with D=3, error bars computed with 1000 time series L=500.



C. Masoller et al, New J. Phys. 17 (2015) 023068

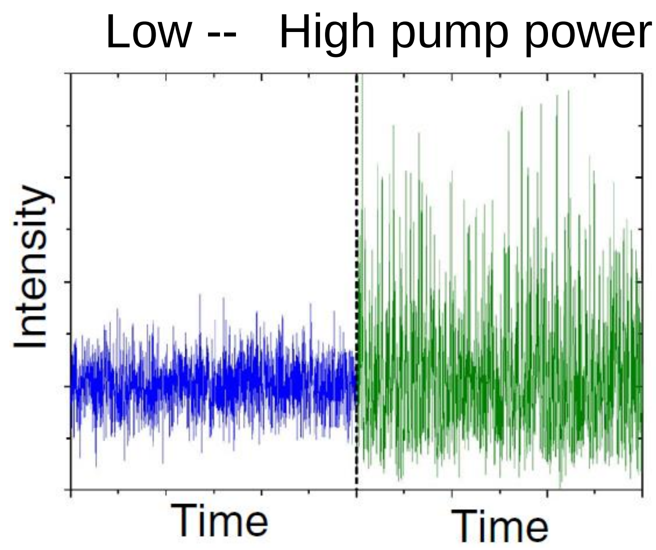
Summary Part 2

- In synthetic data: indicators based in symbolic networks detect transitions which are not captured by Lyapunov analysis.

- In empirical data: they provide early warning indicators of polarization-switching.

- Open issue: comparison with other diagnostic tools.

C. Masoller et al, "Quantifying sudden changes in dynamical systems using symbolic networks", New J. Phys. 17, 023068 (2015).



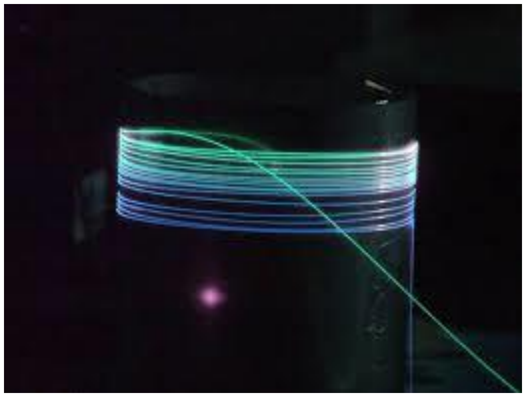
Characterizing the laminar-turbulence transition in a fiber laser

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

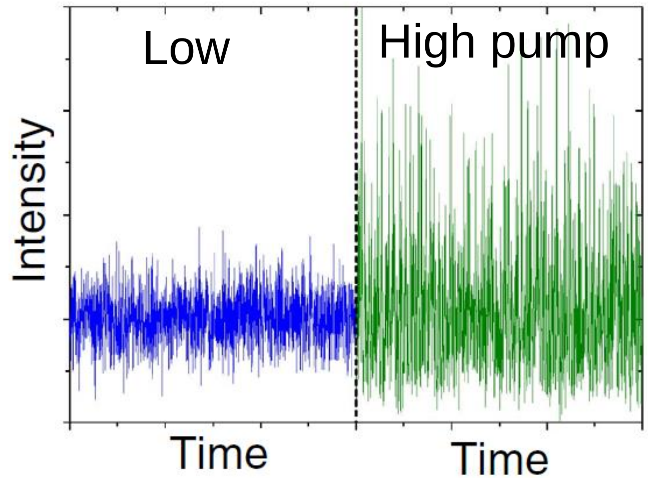
Campus d'Excel·lència Internacional

Experimental data from Aston University, UK (Prof. Turitsyn' group)

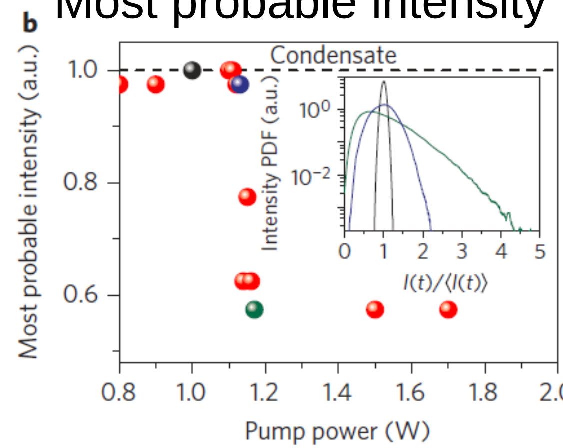
Fiber laser



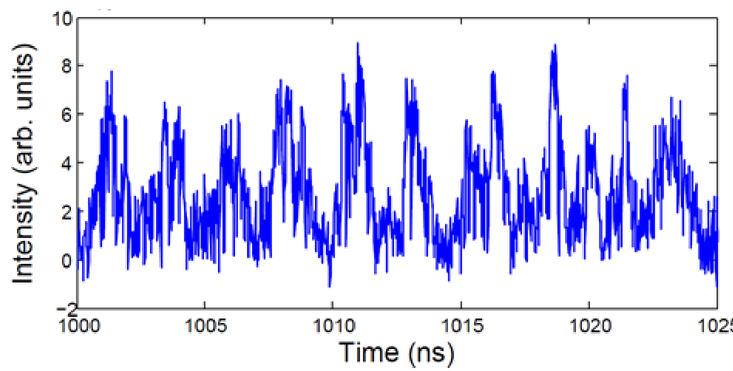
L=1 km,
millions of modes



Most probable intensity



At the transition:

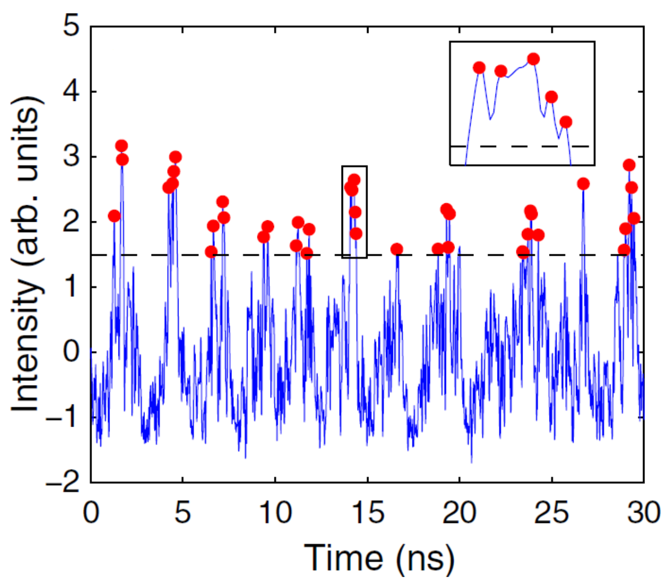


Experimental data from Prof. Turitsyn' group (Aston University, UK)

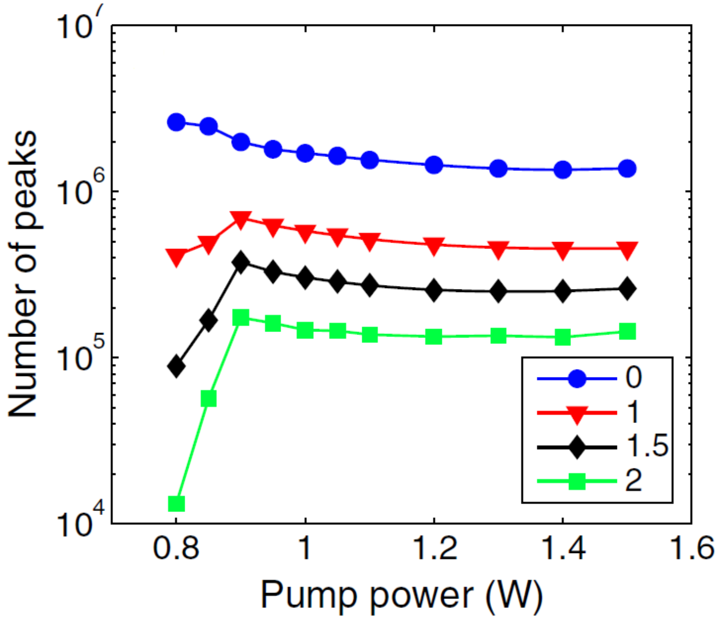
E. G. Turitsyna et al. Nat. Phot. 7, 783 (2013)

Analysis of the intensity peaks higher than a threshold

Each time series is first normalized to ⟨I⟩=0 and σ=1



$$\{I_{\max,i}\}$$



L= $5\text{x}10^7$ data points.

Sampling time dt = 12.5 ps

Th = 2: number of peaks $>10^4$ for all values of the pump power

Diagnostic tool: horizontal visibility graph (HVG)

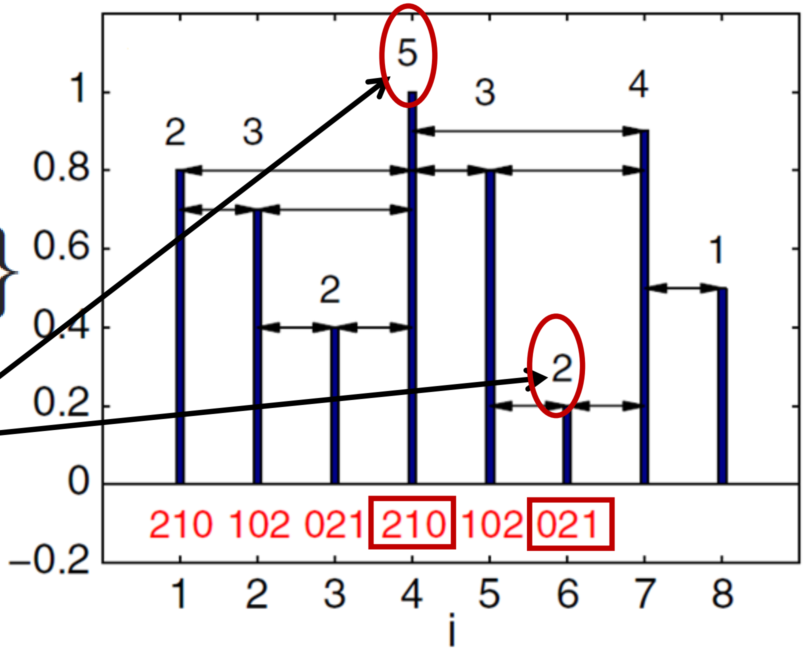
A time-series is represented as a graph, where each data point is a node

intensity peaks above Th = 2

$$\{I_{\max,i}\}$$

Number of links

Ordinal pattern D=3

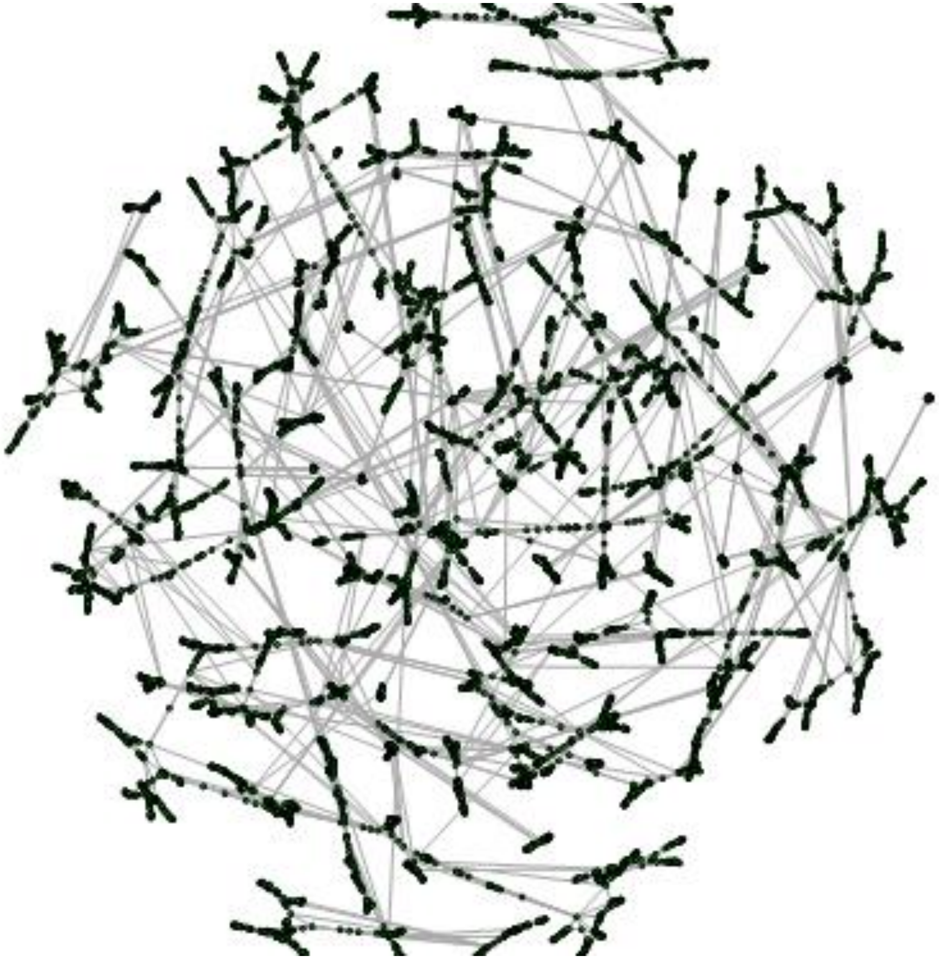


- Rule: data points *i* and *j* are connected if there is "visibility" between them: $I_{\max,i}$ and $I_{\max,j} > I_{max,n}$ for all n, $i<n<j$

$\Rightarrow$ ***Unweighted and undirected graph***

HVG method: B. Luque et al, PRE 80, 046103 (2009)

The resulting network

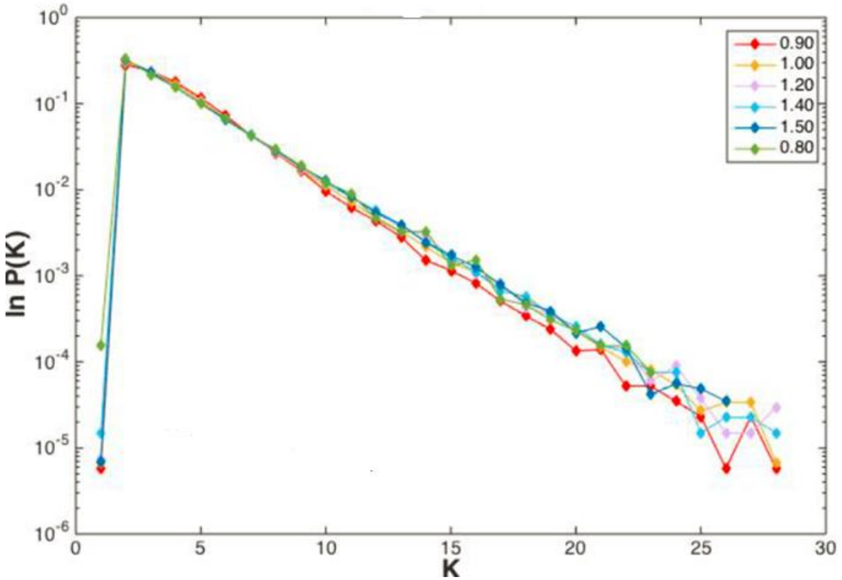


How to characterize this network?

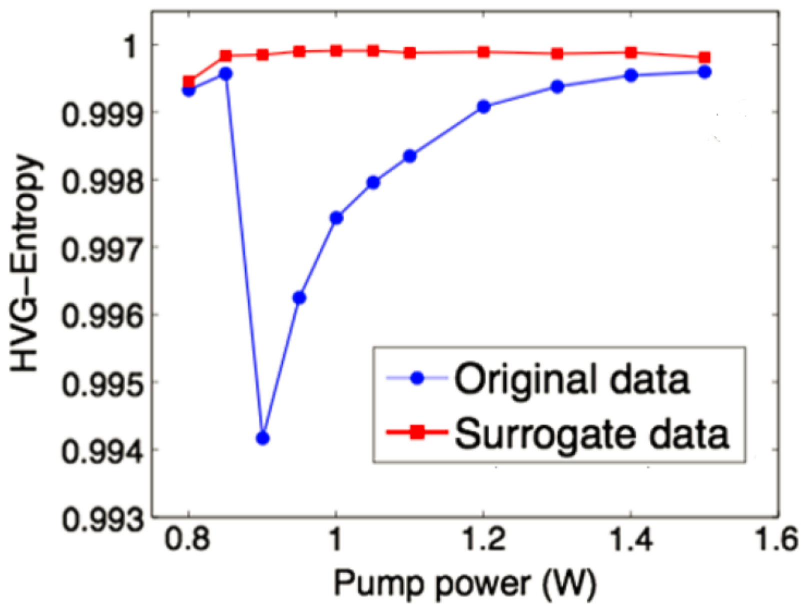
HVG analysis

⇒ Degree Distribution (distribution of the number of links)

- Degree distribution for various pump powers using Th=2.



- Entropy of the degree distribution (normalized to the entropy of Gaussian white noise)

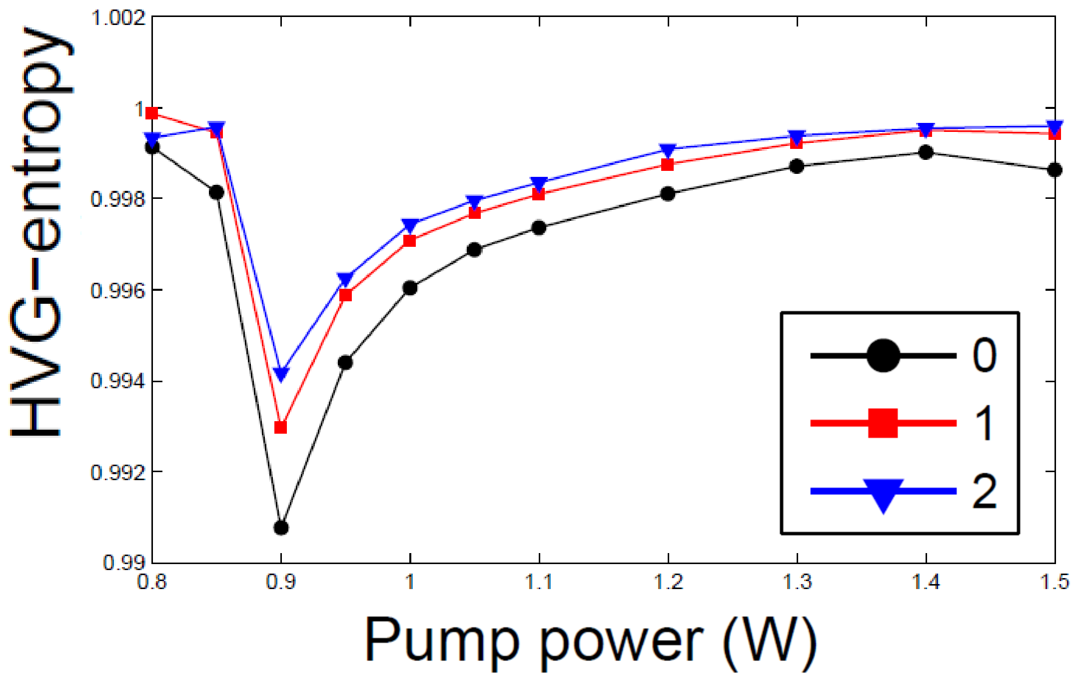


⇒ sharp transition detected.

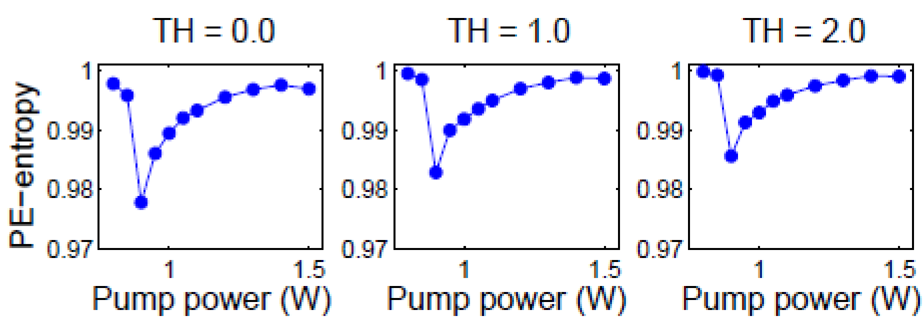
Aragoneses et al, PRL 116, 033902 (2016)

Influence of the threshold

Raw data $\{\ldots I_i \ldots\} \Rightarrow \text{Th} \Rightarrow \{\ldots I_{max,i} \ldots\}$



$\Rightarrow$ sharp transition detected with different thresholds.



Aragoneses et al, PRL 116, 033902 (2016)

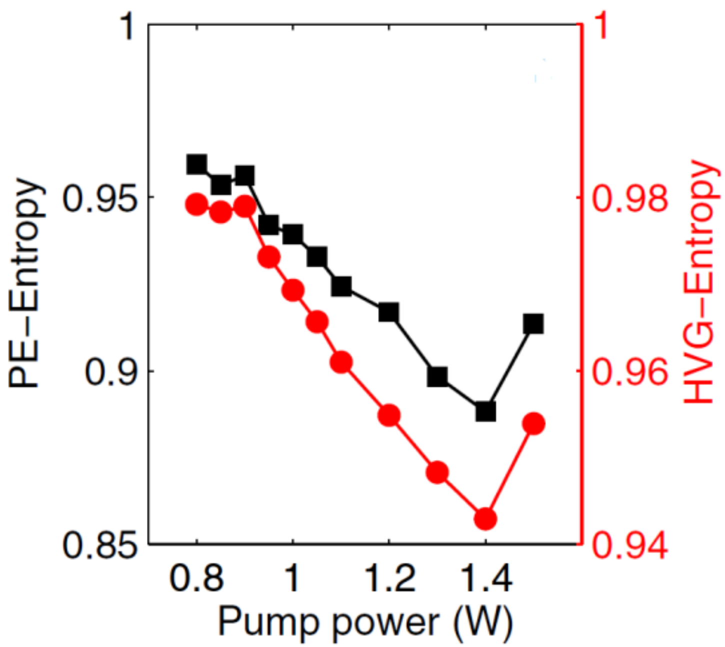
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

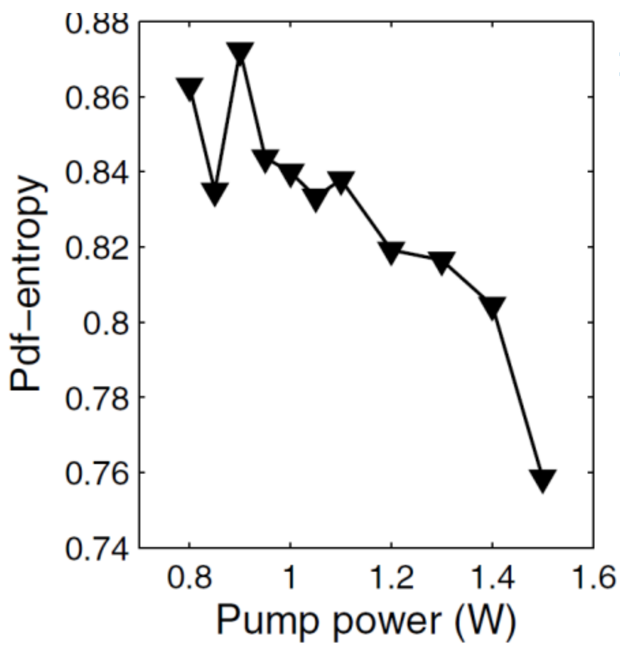
Campus d'Excel·lència Internacional

Influence of the threshold

Raw data $\{\ldots I_i \ldots\} \Rightarrow$ Th $\Rightarrow \{\ldots I_{max,i} \ldots\}$

With the raw data





$\Rightarrow$ sharp transition not detected.

Can we obtain more info. from the raw data?

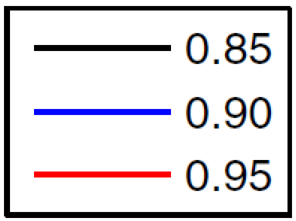
Aragoneses et al, PRL 116, 033902 (2016)

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

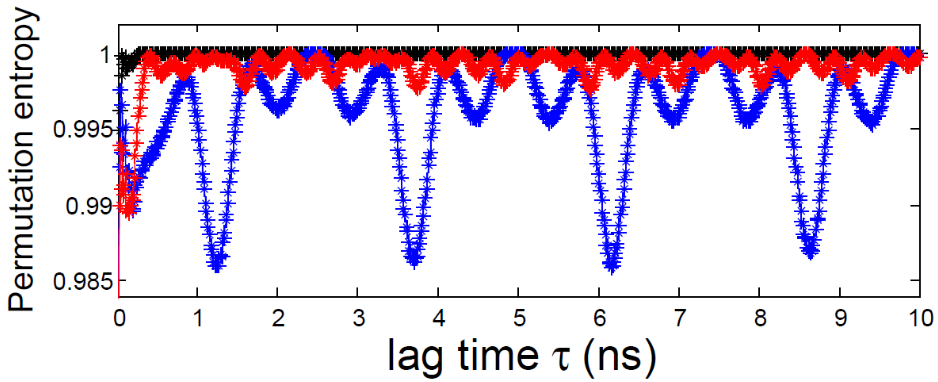
Ordinal analysis of lagged intensity data

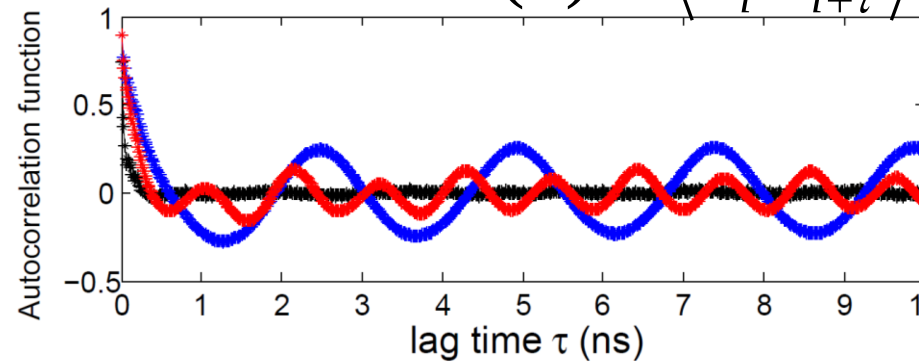
$$\{I_i, I_{i+\tau}, I_{i+2\tau}, \dots\};$$



Pump power below, at, and above the transition.

$$ACF(\tau) = \langle I_i \, I_{i+\tau} \rangle$$





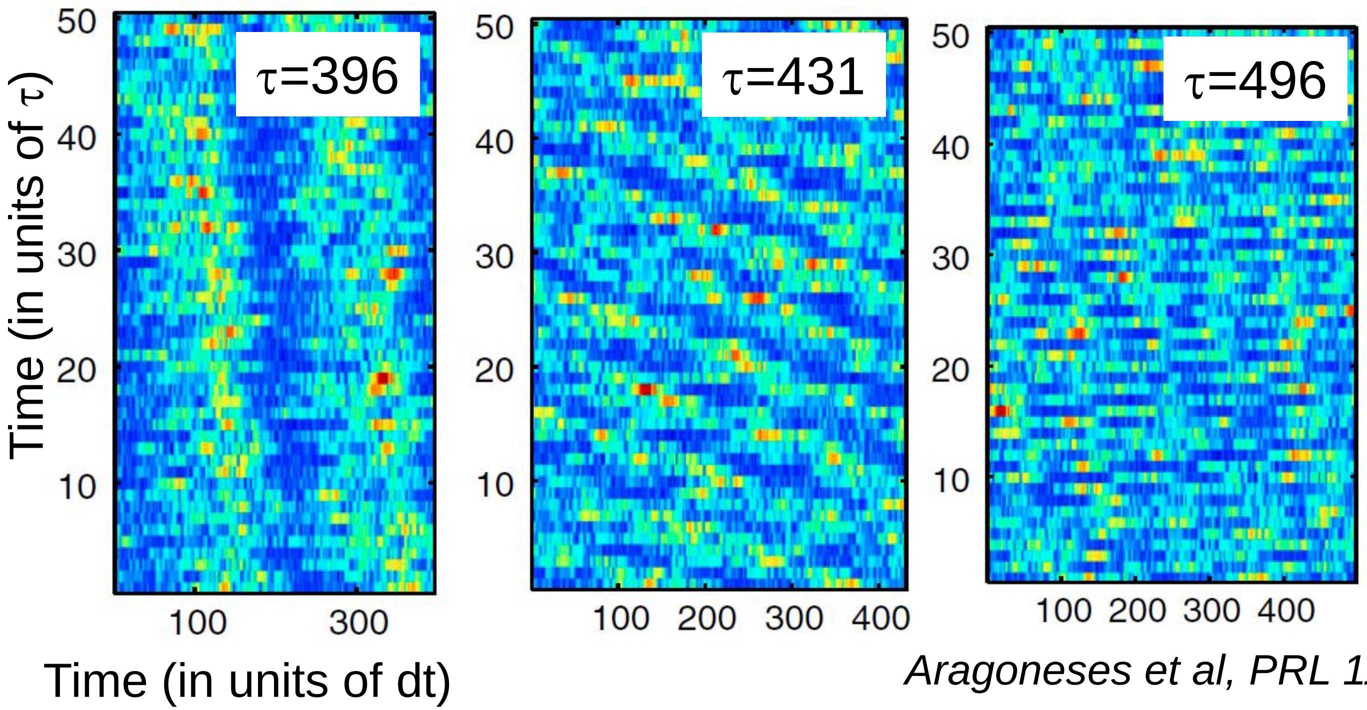
$\Rightarrow$ Sharp variations not captured by linear correlation analysis.

Aragoneses et al, PRL 116, 033902 (2016)

Space time representation

$$\{I_1, I_2, I_3 \ldots,\ I_\tau, I_{\tau+1}, \ldots,\ I_{2\tau}, I_{2\tau+1}, \ldots\} \Rightarrow \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ I_{2\tau+1} & I_{2\tau+2} & \cdots & I_{3\tau} \\ I_{\tau+1} & I_{\tau+2} & \cdots & I_{2\tau} \\ I_1 & I_2 & \cdots & I_\tau \end{bmatrix}$$

Color: I_i



$\Rightarrow$ Different coherent structures uncovered with different lags (sampling times).

Aragoneses et al, PRL 116, 033902 (2016)

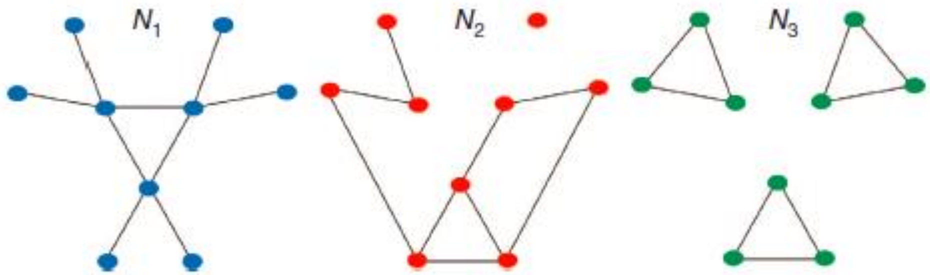
Summary Part 3

- The laser intensity dynamics was mapped to a complex network.

- Sharp transition seen in thresholded data but not in raw data.

- Specific time-scales detected at the transition, not captured by linear correlation analysis.

A. Aragoneses et al, "Unveiling temporal correlations characteristic of a phase transition in the output intensity of a fiber laser"
PRL 116, 033902 (2016).



Quantifying network dissimilarities

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

Complex network measures

- Degree distribution, closeness centrality, betweenness centrality, average path length, etc.
- Provide *partial* information.
- How to define a measure that contains detailed information about the *global topology* of a network, in a *compact* way?

$\Rightarrow$ Node Distance Distributions (NDDs)

- $p_i(j)$ is the fraction of nodes that are connected to node i at distance j
- NDDs = vector of N pdfs $\{p_1, p_2, \ldots, p_N\}$
- If two networks have the same set of distance distributions $\Rightarrow$ 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 $\{p_1, p_2, \ldots, p_N\}$
- Quantifies the heterogeneity of connectivity distances.

$$\mathrm{NND}(G) = \frac{\mathcal{J}(\mathbf{P}_1, \ldots, \mathbf{P}_N)}{\log(d+1)}$$

d = diameter

$$\mathcal{J}(\mathbf{P}_1, \ldots, \mathbf{P}_N) = \frac{1}{N} \sum_{\mathrm{i,j}} p_{\mathrm{i}}(j) \log\left(\frac{p_{\mathrm{i}}(j)}{\mu_{\mathrm{j}}}\right)$$

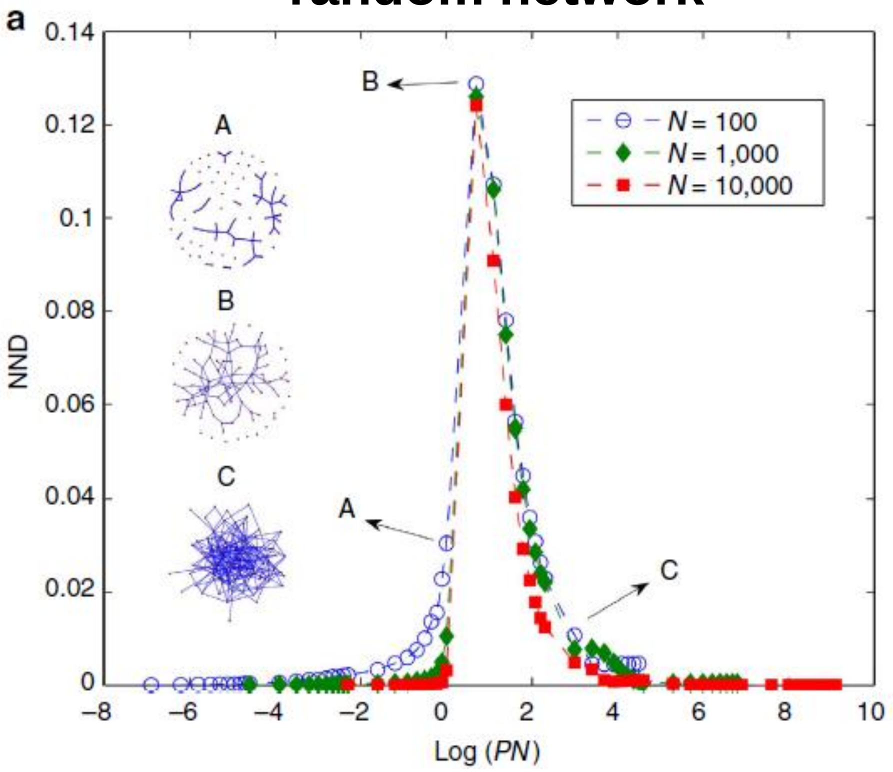
$$\mu_{\mathrm{j}} = \left(\sum_{\mathrm{i}=1}^{N} p_{\mathrm{i}}(j)\right) / N$$

Reminder: distance between P and Z

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

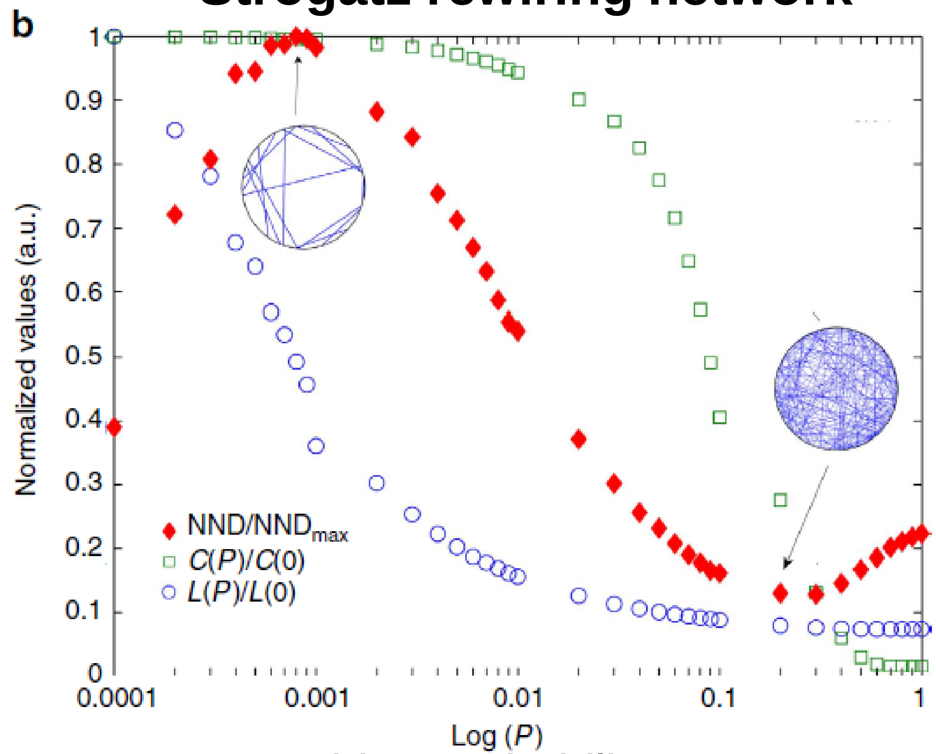
Application.
The Network Node Dispersion detects:

Percolation transition in a random network



P=connection probability

Small-world region in the Watts-Strogatz rewiring network



P=rewiring probability
N=1000, average degree = 10
L=path length, C=clustering coef.

T. A. Schieber et al, Nat. Comm. 8:13928 (2017)

Dissimilarity between two networks

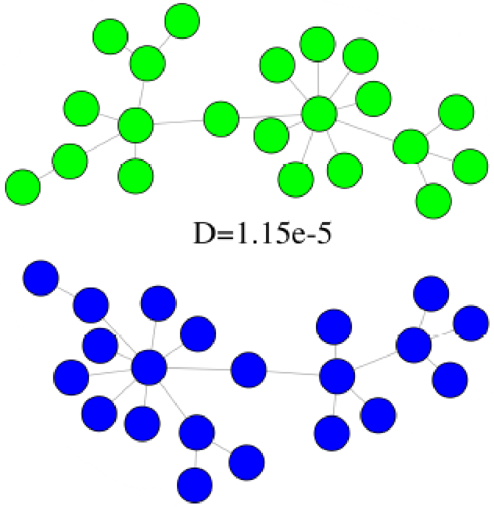
$$D(G,G') = w_1 \sqrt{\frac{\mathcal{J}(\mu_G, \mu_{G'})}{\log 2}} + w_2 \left| \sqrt{\mathrm{NND}(G)} - \sqrt{\mathrm{NND}(G')} \right|$$

$w_1 = w_2 = 0.5$

compares the averaged connectivity

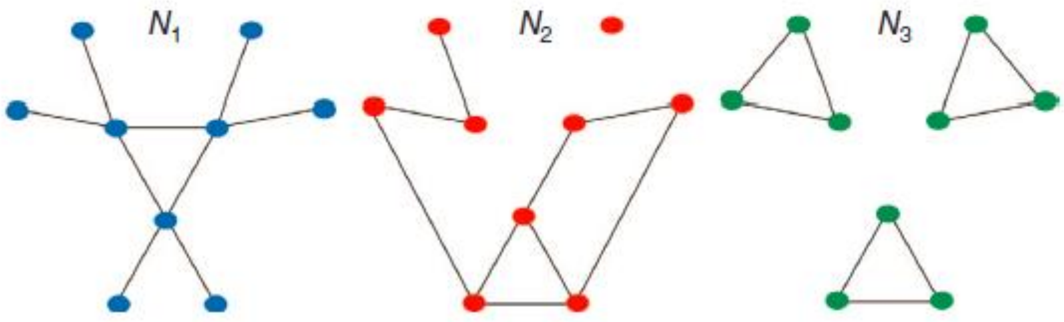
compares the heterogeneity of the connectivity distances

- Extensive numerical experiments demonstrate that Isomorphic graphs return ***D=0***



T. A. Schieber et al, Nat. Comm. 8:13928 (2017)

Comparing three networks with the same number of nodes and links



	D	Hamming	Graph Edit Distance
N_1,N_2	0.25	12	6
N_1,N_3	0.56	12	6
N_2,N_3	0.47	12	6

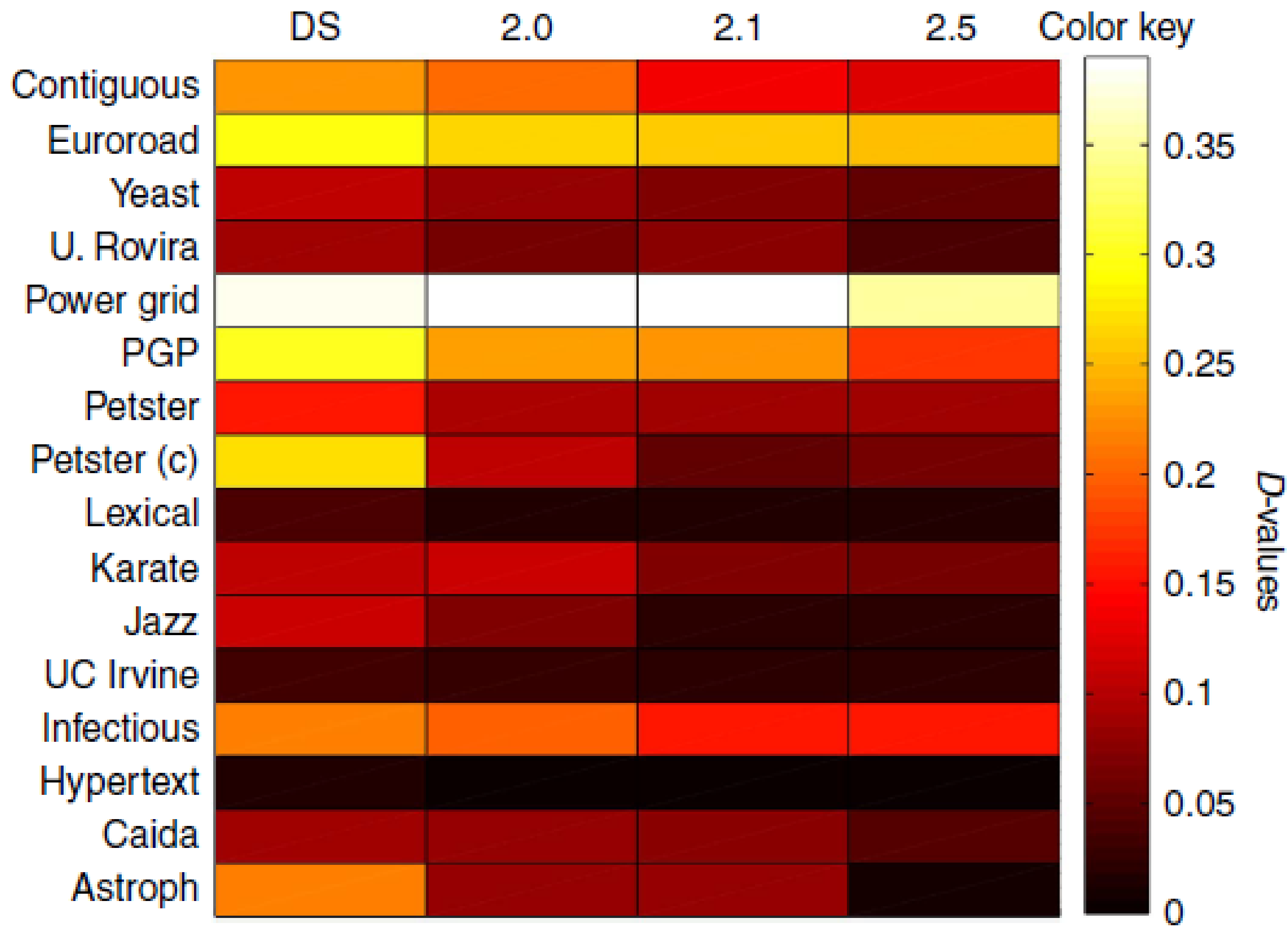
T. A. Schieber et al, Nat. Comm. 8:13928 (2017)

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

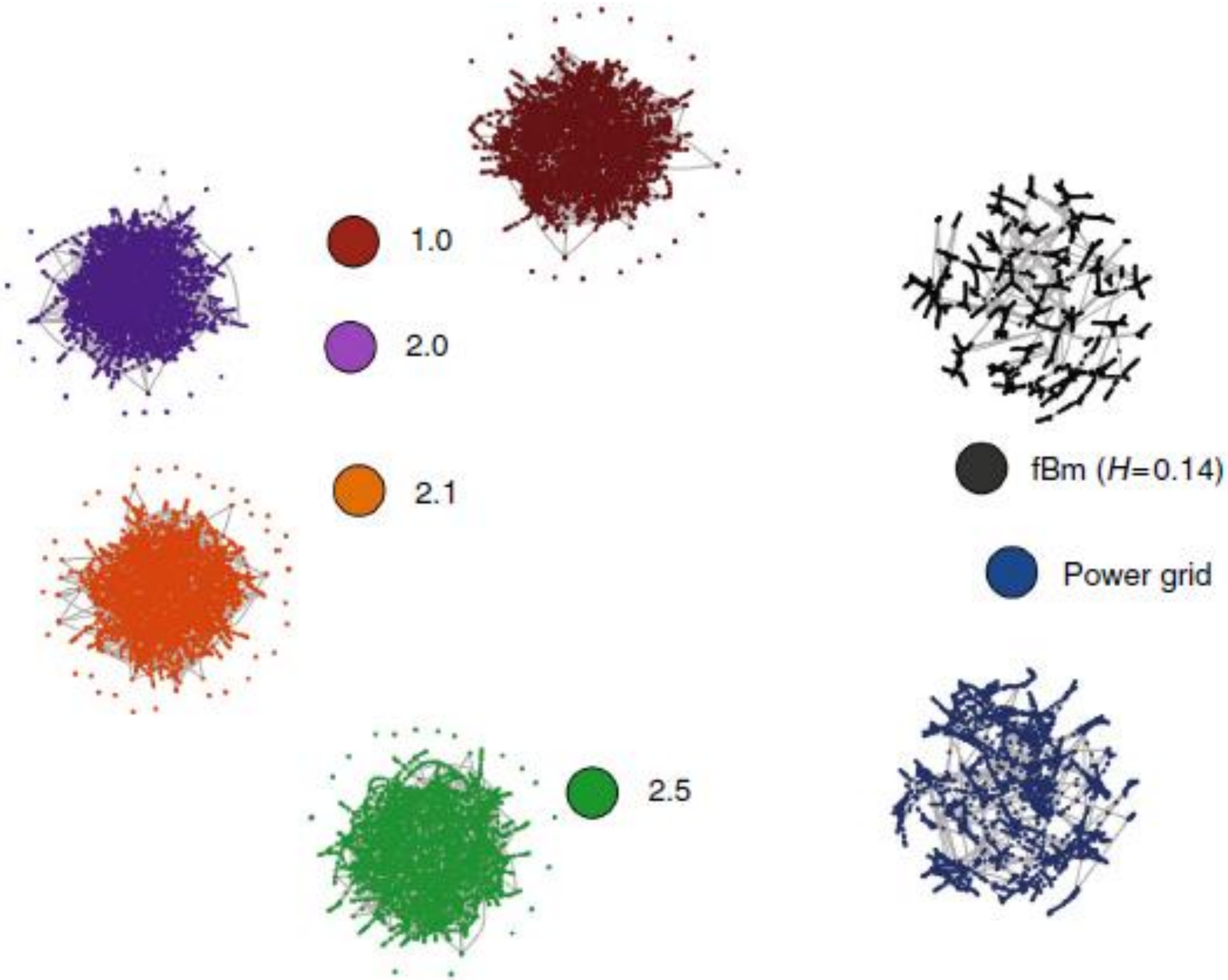
Comparing real networks to null models

DS preserves the degree sequence; 2.0 also preserves the degree correlation; 2.1 also preserves the clustering coefficient; 2.5 includes the clustering spectrum



T. A. Schieber et al, Nat. Comm. 8:13928 (2017)
Details in the supplementary information

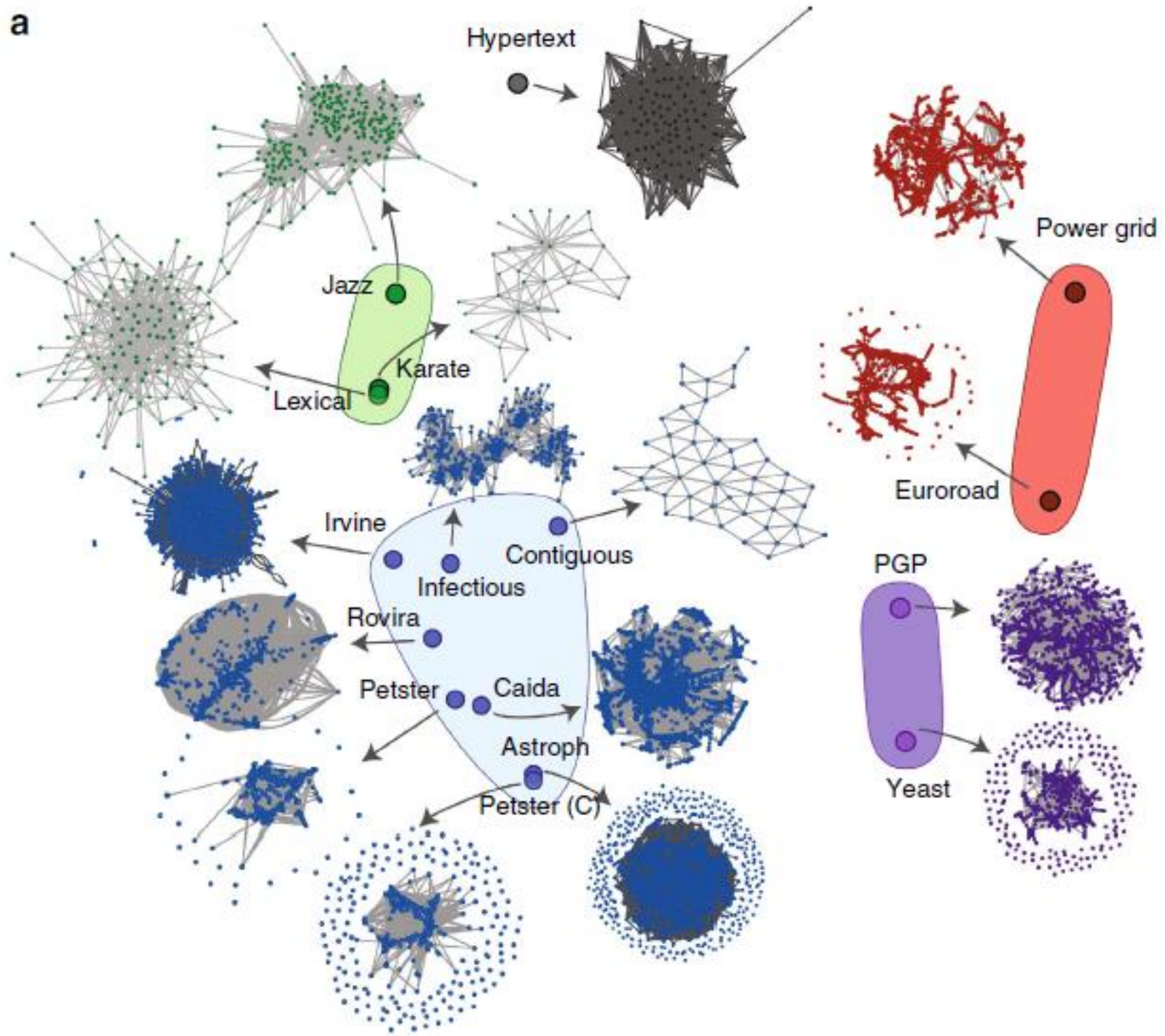
Best model of Power Grid Network?



HVG

T. A. Schieber et al, Nat. Comm. 8:13928 (2017)
Details in the supplementary information

Comparing real networks among them



T. A. Schieber et al, Nat. Comm. 8:13928 (2017)
Details in the supplementary information

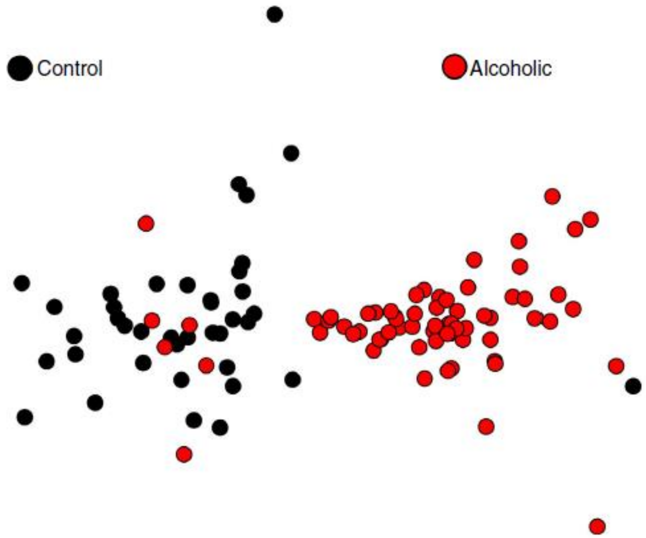
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Campus d'Excel·lència Internacional

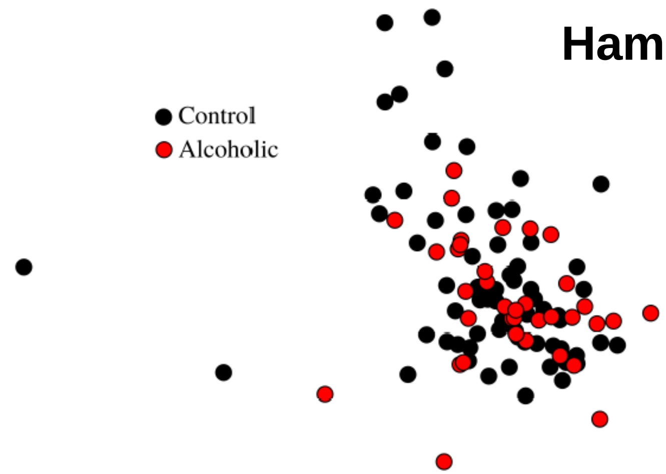
Comparing brain networks

- Use HVG to transform EEG time-series into networks.
- Weight between two brain regions given by 1-*D*(G,G')
- Identify two brain regions (called 'nd' and 'y'), where the weight of the connections between these regions is higher in control than in alcoholic networks

D



Hamming



T. A. Schieber et al, Nat. Comm. 8:13928 (2017)
Details in the supplementary information

Summary Part 4

- New measure to quantify the heterogeneity of the connectivity of a network.
  - Uses a set of probability distributions that contain full information of the connectivity distances.
  - Detects percolation transition in random networks and the limits of the small-world region in rewiring networks.

- New measure to compare two networks
  - Allows to compare directed/undirected networks, same or different sizes.
  - Returns D=0 only if the two networks are isomorphic.
  - Wide applications to real-world networks.

T. A. Schieber et al, "Quantification of network structural dissimilarities", Nat. Comm. 8:13928 (2017)

Coauthors

At UPC:

- **Giulio Tirabassi**
- **Andres Aragoneses**
- **Laura Carpi**
- **Antonio Pons**
- **Carme Torrent**

Experimental data:

- Polarization swithching data from **S. Barland** (Nice, France) and **Y. Hong** (Bangor University, UK)
- Fiber laser data from **S.K. Turitsyn, N. Tarasov & D.V. Churkin** (Aston University, UK)

Elsewhere:

- **J. Viebahn, V. Dakos , H.A. Dijkstra, M. Rietkerk & S.C. Dekker** (Utrecht University)
- **Sergio Gomez & Alex Arenas** (Universidad Rovira Virgil, Tarragona)
- **Albert Diaz-Guilera** (Universidad de Barcelona)
- **T. A. Schieber & M. G. Ravetti** (Universidade Federal de Minas Gerais, Brazil)
- **Panos M. Pardalos** (University of Florida)

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Campus d'Excel·lència Internacional

THANK YOU FOR YOUR ATTENTION !

<cristina.masoller@upc.edu>

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

- G. Tirabassi et al, "*Interaction network based early-warning indicators of vegetation transitions*", Ecological Complexity 19, 148 (2014).
- C. Masoller et al, "*Quantifying sudden changes in dynamical systems using symbolic networks*", New J. Phys. 17, 023068 (2015).
- A. Aragoneses et al, "*Unveiling temporal correlations characteristic of a phase transition in the output intensity of a fiber lase*r",

  PRL 116, 033902 (2016).
- T. A. Schieber et al, "*Quantification of network structural dissimilarities*", Nat. Comm. 8:13928 (2017).

ICREA

GOBIERNO DE ESPAÑA

MINISTERIO DE CIENCIA E INNOVACIÓN