Quantifying sudden changes in dynamical systems using symbolic networks

This content has been downloaded from IOPscience. Please scroll down to see the full text.
2015 New J. Phys. 17 023068
(http://iopscience.iop.org/1367-2630/17/2/023068)

View the table of contents for this issue, or go to the journal homepage for more

Download details:

IP Address: 193.147.222.244
This content was downloaded on 09/10/2015 at 14:48

Please note that terms and conditions apply.
Quantifying sudden changes in dynamical systems using symbolic networks

Cristina Masoller¹, Yanhua Hong², Sarah Ayad³, Francois Gustave³, Stephane Barland³, Antonio J Pons¹, Sergio Gómez⁴ and Alex Arenas⁴

¹ Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, 08222 Terrassa, Spain
² School of Electronic Engineering, Bangor University, Bangor, Gwynedd LL57 1UT, Wales, UK
³ Université de Nice Sophia Antipolis, Institut Non-Linéaire de Nice, UMR 6618, 06560 Valbonne, France
⁴ Departament d’Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain

E-mail: cristina.masoller@upc.edu, y.hong@bangor.ac.uk, stephane.barland@inln.cnrs.fr, a.pons@upc.edu, sergio.gomez@urv.cat and alexandre.arenas@urv.cat

Keywords: complex networks, time series analysis, nonlinear dynamical systems

Abstract
We characterize the evolution of a dynamical system by combining two well-known complex systems’ tools, namely, symbolic ordinal analysis and networks. From the ordinal representation of a time series we construct a network in which every node weight represents the probability of an ordinal pattern (OP) to appear in the symbolic sequence and each edge’s weight represents the probability of transitions between two consecutive OPs. Several network-based diagnostics are then proposed to characterize the dynamics of different systems: logistic, tent, and circle maps. We show that these diagnostics are able to capture changes produced in the dynamics as a control parameter is varied. We also apply our new measures to empirical data from semiconductor lasers and show that they are able to anticipate the polarization switchings, thus providing early warning signals of abrupt transitions.

1. Introduction
Symbolic time series analysis is a powerful technique able to extract hidden features such as the presence of frequent recurrent patterns [1–5], or the existence of missing/forbidden patterns [6–8] from stochastic, high-dimensional signals. Symbolic analysis has also proven to be useful for classifying different types of signals [9–12] and, in bivariate analysis, for inferring the direction of information flow [13, 14]. The symbolic approach involves the transformation of a time series, x(t), into a sequence of symbols, s(t), by using an appropriated codification rule. A significant advantage of symbolization is its reduced sensitivity to observational noise. Complexity measures have been proposed to characterize the resulting symbolic sequence, a very popular one being the permutation entropy (PE) [15–21], computed with an ordinal symbolization rule by comparing neighboring values in the time series.

Complex networks have also been successfully employed for time series analysis [22–27]. Correlation graphs [28, 29], recurrence graphs [30–32], and visibility graphs [33–37] have been shown to provide relevant information, for example, of early warning indicators of qualitative changes and abrupt transitions.

Here we characterize the evolution of a dynamical system by combining symbolic ordinal analysis and network representation in a hybrid methodology. Specifically, a time series x(t) (raw data) is transformed in a sequence s(t) of symbols (ordinal patterns, OPs) that is used to construct a directed and weighted graph where the different symbols that appear in s(t) constitute the network nodes. Each of these nodes has an associated probability of occurrence. The transitions between consecutive symbols constitute the network links: any pair of nodes, i and j, are connected by a directed link if the sequence of symbols (i, j) occurs in the symbolic sequence, and the weight of the directed link is the probability of j occurring after i (referred to as the transition probability i → j, TP).
Our approach for defining the links as the TP between symbols was originally proposed by Nicolis et al [22] and the use of ordinal analysis was proposed by Sun et al [27]; however, in [27] the weights of the links were not defined in terms of TPs (as in [22]), but rather, they were defined as the number of times the pair of consecutive symbols \((i, j)\) occurred in the time series. A major advantage of the use of TPs is that they are normalized in each node, and thus, allow for computing an entropy for each node, in the following referred to as node entropy. The node entropies can then be used to identify changes in the network, which reflect changes in the time series when the control parameter is varied.

Exploiting this network representation we propose several novel measures to characterize a time series:

- the nodes’ entropy, which is the mean entropy of the distribution of the weights of the outgoing links of a node (in other words, the node entropy averaged over all the nodes of the network);
- the links’ entropy, which is the entropy of the distribution of links’ weights; and
- the asymmetry coefficient, which is computed from the difference of the weights of the links \(j \rightarrow i\) and \(i \rightarrow j\), averaged over all the links of the network.

Note that the nodes’ and the links’ entropies are computed according to the classical Shannon entropy definition.

By analyzing numerical data generated from the logistic map, the tent map, the circle map, and experimental data, recorded from the intensity of semiconductor lasers under different operation conditions, we show that the entropies defined from the symbolic network vary over a wider range of values and outperform the PE in providing early warning signals of sudden changes in the time series.

This paper is organized as follows. In sections 2 and 3 the method of network construction and the network-based measures proposed for characterizing the time series are described. In section 4 results are presented and, finally, we provide the discussion and some conclusions in section 5.

2. Network construction

We transform a time series \(x(t)\) into a sequence of symbols \(s(t)\) by using the OP representation with symbols of length \(D\). In this case, symbols are defined by considering groups of \(D\) consecutive values in the time series [15]. There exist \(D!\) different OPs of length \(D\). For example, for \(D = 2\) there are two OPs: \(x(t) < x(t + 1)\) gives symbol ‘01’ and \(x(t) > x(t + 1)\) gives symbol ‘10’; for \(D = 3\) there are six possible OPs: \(x(t) < x(t + 1) < x(t + 2)\) gives ‘012’, \(x(t + 2) < x(t + 1) < x(t)\) gives ‘210’, etc.

In general, the ordinal representation of \(x(t)\) gives a sequence \(s(t)\) of OPs with \(M\) different symbols because depending on the system’s dynamics not all possible symbols will be present in the symbolic sequence, \(s(t)\), because either the dynamics do not allow for some of the (the so-called ‘forbidden patterns’), or the time series \(x(t)\) is of finite length and simply does not appear (‘missing patterns’) [6–8]. Thus, the number of nodes in the network corresponds only to those symbols appearing in the symbolic sequence and is \(M \leq D!\).

The weight of a node \(i\) is the relative number of times the symbol \(i\) occurs in the sequence \(s(t)\):

\[
p_i = \frac{1}{L} \sum_{t=1}^{L} n[s(t) = i],
\]

where \(n\) is a count of the number of occurrences and \(L\) is the length of the sequence \(s(t)\). The weights are normalized, \(\sum_{i=1}^{M} p_i = 1\), and the PE \(s_p\) is the entropy of the distribution of node weights [15],

\[
s_p = -\sum_{i=1}^{M} p_i \log p_i.
\]

We note that \(0 \leq s_p \leq \log M\), with \(s_p = 0\) if \(p_i^* = 1\) for \(i = i^*\) and \(p_i = 0 \forall i \neq i^*\) (only one symbol, \(i^*\), appears in the symbolic sequence and thus the network has only one node); \(s_p = \log M\) if \(p_i = 1/M \forall i\) (all the symbols appear in the symbolic sequence with the same probability and thus all the nodes have the same weight).

The weight of a link \(i \rightarrow j\), \(w_{ij}\), is the relative number of times, in the sequence \(s(t)\), the symbol \(i\) is followed by the symbol \(j\):

\[
w_{ij} = \frac{\sum_{t=1}^{L-1} n[s(t) = i, s(t + 1) = j]}{\sum_{t=1}^{L-1} n[s(t) = i]}.
\]
With this definition the link weights are normalized in each node, i.e., \( \sum_{j=1}^{M} w_{ij} = 1, \forall i \). We note that this definition allows for the presence of self-loops.

It is illustrative to describe the type of networks that are generated by simple dynamics. A periodic time series will give a regular symbolic sequence, whose periodicity will depend on the length of the OP, \( D \). The resulting symbolic network will then depend on \( D \). To focus the ideas, let’s consider the time series shown in figure 1(a), with period 4, generated from the logistic map. With \( D = 3 \) we obtain the following sequence of symbols

\[
201, 021, 102, 120, 201, 021, 102, \ldots \tag{4}
\]

Because the TPs are computed from OPs that are formed by non-superposed values (i.e., for \( D = 3 \), \{\( x(t) \), \( x(t+1) \), \( x(t+2) \)\} define one pattern and \{\( x(t+3) \), \( x(t+4) \), \( x(t+5) \)\} define the next one), the network obtained has four nodes connected as follows:

\[
201 \rightarrow 120, 021 \rightarrow 201, 102 \rightarrow 021, 120 \rightarrow 102. \tag{5}
\]

If we use \( D = 4 \) OPs, then the sequence of symbols is

\[
3021, 0213, 2130, 1302, 3021, 0213, 2130, \ldots \tag{6}
\]

and the resulting network has four nodes with self-loops,

\[
3021 \rightarrow 3021, 0213 \rightarrow 0213, 2130 \rightarrow 2130, 1302 \rightarrow 1302. \tag{7}
\]

It is important to note that, in the ordinal encoding scheme, depending on the value of \( D \), an irregular time series can have an associated regular symbolic sequence, and thus, a regular network. To fix the ideas, let’s consider the time series shown in figure 1(b), also generated from the logistic map. With values of \( D \) equal to 2, 3, or 4, this time series gives the same symbolic sequence as the period-4 time series shown in panel (a). This feature is a consequence of the ordinal codification rule. Other codification choices can result in two different symbolic sequences.

On the other hand, if the time series is fully random, the symbolic sequence will also be fully random, and then the network will have a regular, all-to-all topology, regardless of the value of \( D \) (assuming the time series is sufficiently long).

3. Network-based diagnostics for time series analysis

The symbolic network allows defining several novel measures:

1) Since \( \sum_{j=1}^{M} w_{ij} = 1, \forall i \), we can compute an entropy at each node,

\[
s_i = - \sum_{j=1}^{M} w_{ij} \log w_{ij}, \tag{8}
\]

and then characterize the network heterogeneity in terms of the distribution of the \( s_i \) values.

We note that, on one hand, \( s_i = 0 \) if and only if node \( i \) has only one outgoing link, \( j^* \); in this case, \( w_{ij^*} = 1 \) and \( w_{ij} = 0, \forall j \neq j^* \), and the symbol \( j^* \) occurring after symbol \( i \) is always the same. In this case, in the symbolic sequence, the symbol that follows \( i \) is fully predictable. On the other hand, if all the outgoing links of node \( i \) have the same weight, then \( w_{ij} = 1/k_i \) and \( s_i = \log k_i \) (here \( k_i \) is the out-degree of node \( i \), \( k_i = \sum_{j=1}^{M} A_{ij} \), where \( A_{ij} \) are
the components of the adjacency matrix, \( A_{ij} = 1 \) if \( w_{ij} \neq 0 \), otherwise \( A_{ij} = 0 \). The maximum \( s_i \) value is \( \log M \) and corresponds to a node that is connected to all the other nodes—including itself—with links that have uniform weights. In this case, in the symbolic sequence any symbol can follow \( i \) with the same probability.

We propose the use of the first moment of the distribution of \( s_i \) values, the network-averaged node entropy,

\[
s_n = \frac{1}{M} \sum_{i=1}^{M} s_i,
\]

as a novel measure for the analysis of a time series. We note that \( 0 \leq s_n \leq \log M \). If \( s_n = 0 \), then \( s_i = 0 \) \( \forall i \), and all the nodes have only one outgoing link; therefore, the symbol \( j \) that occurs after symbol \( i \) is fully predictable, \( \forall i \). On the contrary, the largest \( s_n \) value, \( \log M \), occurs when \( s_i = \log M \) \( \forall i \). In this case the symbolic sequence is fully random and after any symbol, \( i \), any symbol in the sequence can follow with the same probability \( 1/M \).

We note that the range of variation of the node entropies, \( s_j \), and of their average, \( s_n \), is the same as that of the PE, \( s_p \).

2) From the probability distribution function (PDF) of the weights of all the links, \( p(w) \), we define the link’s entropy as:

\[
s_l = -\int_0^1 p(w) \log p(w) \, dw.
\]

We have estimated \( p(w) \) from an histogram of the weights of the existing links \( (w_{ij} \neq 0) \). Large values of \( s_j \) indicate a heterogeneous distribution of link weights, while small values of \( s_j \) indicate a narrow distribution of link weights. The delta distribution, \( w_{ij} = w, \forall i, j \) that gives \( s_j = 0 \) includes two limits, one in which the symbolic sequence is fully predictable (each node has only one outgoing link: \( \forall i, w_{ii}^* = 1 \) and \( w_{ij} = 0 \) if \( j \neq j_i^* \)) and the other in which the symbolic sequence is fully unpredictable (the nodes are all-to-all connected with uniform weights: \( w_{ij} = 1/M, \forall i, j \)). Thus, \( s_j \) quantifies the complexity of the symbolic sequence, as \( s_j = 0 \) corresponds to both, perfectly regular and fully random sequences.

3) To quantify the asymmetry in the direction of the network links, we introduce the asymmetry coefficient, which is defined as:

\[
a_c = \frac{\sum_{j \neq i} \sum_{j \neq i} |w_{ij} - w_{ji}|}{\sum_{j \neq i} (w_{ij} + w_{ji})},
\]

\( a_c = 0 \) in a fully symmetric network \( (w_{ij} = w_{ji}, \forall i, j) \) and \( a_c = 1 \) in a fully directed network (either \( w_{ij} = 0 \) or \( w_{ji} = 0, \forall i, j \)). This asymmetry coefficient is closely related to the concept of reciprocity in graphs [38].

We note that to compute these measures, only the nodes and the links that are present in the network are taken into account. While absent nodes and links, corresponding to missing or forbidden symbols and transitions, could have been taken into account with zero weights, we preferred to use this more generic approach because, depending on the symbolic transformation used, the complete set of symbols can be unknown or even infinite.

### 4. Results

#### 4.1. Analysis of synthetic data

We start by presenting the results of the analysis of simulated time series for the logistic map, the tent map and the stochastic circle map. The equations and parameters are

- **logistic map**: \( x_{i+1} = rx_i(1 - x_i) \);
- **tent map**: \( x_{i+1} = rx_i \) if \( x_i < 0.5, x_{i+1} = r (1 - x_i) \) if \( x_i \geq 0.5 \);
- **stochastic circle map**: \( \phi_{i+1} = \phi_i + \rho + k (\sin 2\pi \phi_i)/(2\pi) + \beta \xi \). Here \( \xi \) is a Gaussian white noise with unit standard deviation, \( k \) is the map parameter, \( \beta \) is the noise strength and \( \rho = -0.23 \) is kept fixed. We analyze a time series of \( x_i = \phi_i - \phi_{i-1} \).

The analysis is performed with a time series of length \( L = 6000 \) and patterns of length \( D = 4 \). The results are robust as long as \( L \) is much larger than the number of possible links. The value of \( D \) is chosen because it allows constructing symbolic networks with 24 possible nodes and 576 possible links; in contrast, \( D = 5 \) gives networks with 120 possible nodes and 14400 possible links, and therefore, computing the links’ weights with robust statistics requires long time series. We discuss below the influence of \( D \) and \( L \), when we analyze empirical data.
Regarding the logistic and tent maps, they have similar bifurcation sequences (except for the periodic windows in the logistic map) and thus, assuming similar bifurcation sequences gives rise to similar dynamical behaviors, we expect to obtain similar symbolic networks, which will depend on the control parameter. Indeed, this is observed in figure 2 that displays, for the logistic map (left column) and for the tent map (right column), the bifurcation diagrams (top row), the PE, $sp$, and the average node entropy, $sn$, (middle row) and the links' entropy, $sl$, and the asymmetry coefficient, $ac$ (bottom row). For both maps $sp$ and $sn$ increase in a similar way with the map parameter (except in periodic windows). One can note that $sn$ varies over a wider range of values in comparison to $sp$. We note that this difference in the actual range of variation of $sn$ and $sp$ can not simply be normalized over because, as discussed in section 3, both $sn$ and $sp$ vary within the same range of values. Also, $sn$ displays more abrupt variations, which can be understood in the following terms: as the map parameter increases, the dynamics become increasingly chaotic and new OPs appear in the symbolic sequence, which results in new nodes and links in the symbolic network; however, the frequency of occurrence of the new OPs is initially small, and their appearance does not produce abrupt variations in the PE. But the new nodes will not necessarily have small entropies; therefore, they can induce abrupt variations in the average node entropy.

Both $sn$ and $sp$ vary with the map parameter in a qualitatively similar way as the Lyapunov exponent (LE). For the logistic map it has previously been shown that for chaotic parameters and low $D$ values, $sp$ behaves as the LE [15]. In addition, for piecewise monotone interval maps, in the limit of $D \to \infty$, $sp$ tends to the topological entropy [16]. In the visibility graph, a direct relationship between the network entropy and the LE can be established by using the Pesin identity [36]. Therefore, these results suggest that the network entropy can indeed be expected to capture nontrivial properties of the dynamics; however, determining a direct correspondence between different network properties and different kinds of dynamics is not straightforward and is left for future work.

Figure 3 displays the results of the analysis of simulations of the stochastic circle map, when varying the parameter that represents the forcing amplitude (left column) or the noise strength (right column). Varying $k$ allows us to study the transition to locking. For very small values of $k$ the stochastic term dominates and the network contains all possible nodes and links, thus, $sn$ and $sp$ are maximum ($\approx \log 4! = 3.18$). As $k$ increases the dynamics become more deterministic, the number of nodes and links gradually decrease (not shown), and $sp$ and $sn$ also decrease. At $k \approx 1.5$ the transition to locking occurs (see the bifurcation diagram in figure 3(a)) and the presence of noise results in a symbolic sequence that is again fully random, thus, $sp$ and $sn$ again reach the maximum value. If $k$ is further increased, again an increase in the ‘order’ occurs and $sp$ and $sn$ again decrease. The

**Figure 2.** Analysis of the logistic map (left column) and tent map (right column). (a), (b) PDF of the variable $x$ (in color scale) versus the map parameter. (c), (d) The average node entropy, $sn$ (red line), and the PE, $sp$ (black line). Note that the range of variation of $sn$ is more than twice that of $sp$. (e), (f) The links' entropy, $sl$ (black line), and the asymmetry coefficient, $ac$ (red line). The analysis was performed with $D = 4$ and $L = 6000$. 

right column in figure 3 reveals that $s_n$ and $s_p$ capture the monotonic increase of ‘disorder’ in the symbolic sequence as the noise strength, $\beta$, increases.

Other network measures are consistent with these observations. As seen in figures 2(e) and (f), for the logistic and for the tent maps, when the dynamics are weakly chaotic ($r \approx 3.6$ for the logistic map; $r \approx 1.2$ for the tent map), the symbolic network has a low link entropy and is highly asymmetric (the asymmetry coefficient is close to one); as the parameter increases the network grows and the link’s entropy increases, accompanied by large variations of the asymmetry coefficient; for large parameter values (strong chaos) both the asymmetry coefficient and the links’ entropy decrease, suggesting that the network becomes more homogeneous. These results are consistent with a previous analysis of the complexity of the logistic map, where the peak values of the various complexity measures analyzed occurred at intermediate values of the map parameter $[39]$. For the stochastic circle map, $s_l$ and $s_c$ [figures 3(e) and (f)] reveal that the network is homogeneous (large link entropy accompanied by small asymmetry coefficient) for $k \approx 0$, at the locking transition, and for strong noise.

4.2. Analysis of empirical data
Let us next present the results of the analysis of two empirical data sets. The first data set was recorded from the output of a semiconductor laser at various values of the laser bias current ranging from 5.5 to 6.2 mA. The type of laser is a vertical-cavity surface-emitting laser (VCSEL) that can emit two orthogonal polarizations and, when varying the bias current, there is an abrupt polarization switch (PS) at about 6 mA: the dominant polarization turns off and the orthogonal one turns on, as shown in figure 4(a). The data consist of 71 time series of the intensity of one polarization mode recorded at fixed values of the bias current, below and above the polarization-switching point. Each time series contains more than $10^6$ data points. In order to identify clear trends, each time series was divided in several sections, each containing $L$ data points ($L$ varying in the range $10^4$–$10^5$ as discussed below), and the network indicators were computed by averaging the values in each section.

Since the empirical data is very noisy, see figures 4(b) and (c), all possible pairs of symbols $(i, j)$ occur in the symbolic sequence and the network is a regular, all-to-all graph. Nevertheless, we will show that the network measures adequately capture dynamical changes in the time series.

Figure 5 presents a comparison of the the network-based measures, with the mean value, $\langle I \rangle$, and the standard deviation, $\sigma$, of the intensity fluctuations: $\langle I \rangle$ and $\sigma$ vary linearly, and thus, they don’t provide an early warning for the sudden PS. It can also be seen that both $s_p$ and $s_n$ change the slope of the initial linear trend before the PS, but the $s_n$ trend decreases faster, providing a better indicator of the PS.
While the evidence of early warning in terms of a single network measure is not strong, taken together these results provide clear early warnings of abrupt switching. This suggests that the switching involves deterministic light–matter interactions that, despite of the highly stochastic character of the signals analyzed, the network measures are able to capture. In VCSELs the two linear polarizations are strongly anti-correlated and the

Figure 4. (a) PDF of the intensity of the linear polarization mode that turns on versus the laser bias current. Intensity time-traces just before (b) and after (c) the switch.

Figure 5. Early indicators of abrupt polarization switching (PS). The data analyzed is as in figure 4. (a) The mean value and (b) the standard deviation of the intensity fluctuations; (c) the PE, (d) the average node entropy, (e) the links entropy and (f) the asymmetry coefficient. $s_p$ and $s_n$ are normalized to the maximum value, $\log D$. The analysis was performed with $D = 3$. The error bars indicate the standard deviation of the distribution of values computed over 100 sections with $L = 1000$ data points each (see text for details). It can be seen that $s_p$, $s_n$, and $a_c$ are good indicators of the abrupt switching, as they change the linear slope before the PS; in contrast $\langle I \rangle$, $\sigma_I$, and $s_l$ are not good indicators as they either vary linearly or fluctuate with constant amplitude. $s_p$ varies nonlinearly before the PS but in a smaller range of values as compared with $s_n$.

While the evidence of early warning in terms of a single network measure is not strong, taken together these results provide clear early warnings of abrupt switching. This suggests that the switching involves deterministic light–matter interactions that, despite of the highly stochastic character of the signals analyzed, the network measures are able to capture. In VCSELs the two linear polarizations are strongly anti-correlated and the
polarization switching can be in part due to thermal effects (Joule heating) through a shift of the gain maximum [40]. Also, small cavity anisotropies can play a role, as the VCSEL circular transverse geometry provides no polarization selection mechanism. In addition, material birefringence, saturable dispersion, and a microscopic spin-flip relaxation mechanism can explain the PS [41]. Because various sources of noise are ubiquitous in laser systems (spontaneous emission noise, thermal and electrical noise), it is not possible to investigate experimentally the relative impact of stochastic and deterministic mechanisms that trigger polarization switchings. Various models suggest that several laser parameters can modify the linear stability of the two polarization modes, while small variations (deterministic or stochastic) are responsible for triggering the switch.

Figures 6 and 7 analyze the influence of the length $D$ of the OP. The data analyzed is as in figure 4. The left column presents results for $D=2$, the central column, for $D=3$ and the right column, for $D=4$. The top row displays the PE, the central row displays the average node entropy, and the bottom row displays the asymmetry coefficient. For easy comparison $s_p$ and $s_n$ are normalized to the maximum value, $\log D$. The error bars indicate the standard deviation of the distribution of values computed over 100 sections with $L=1000$ data points each (see text for details). The behavior of the network measures is robust for $D=3$ and $D=4$ and even for $D=2$, $s_n$ anticipates the polarization switching.

Figure 6. Influence of the length $D$ of the OP. The data analyzed is as in figure 4. The left column presents results for $D=2$, the central column, for $D=3$ and the right column, for $D=4$. The top row displays the PE, the central row displays the average node entropy, and the bottom row displays the asymmetry coefficient. For easy comparison $s_p$ and $s_n$ are normalized to the maximum value, $\log D$. The error bars indicate the standard deviation of the distribution of values computed over 100 sections with $L=1000$ data points each (see text for details). The behavior of the network measures is robust for $D=3$ and $D=4$ and even for $D=2$, $s_n$ anticipates the polarization switching.

To confirm these observations we now consider a second empirical data set, in which the intensity of one polarization mode was recorded while the control parameter, the laser pump current, was linearly scanned across the switching point. The data was recorded from the output of a different VCSEL, which was subject to polarized optical feedback. In this VCSEL, the polarization switching (shown in figure 8) occurred only if the feedback was strong enough; without feedback no polarization switching was observed in this device. This setup has therefore the advantage that, by finely tuning the feedback strength, we could record two sets of time series: one set in which a PS occurs, and, by decreasing slightly the feedback strength, another set in which no PS occurs.

We recorded 1000 independent realizations (each time series has 20000 data points), and in each time series the control parameter, the laser current, varied from a value below the PS to a value above the PS (see figure 8). For comparison, a second data set (also of 1000 time series with 20000 data points) was recorded with slightly weaker feedback, such that no switchings were observed.
We divided each time series in non-overlapping sections and computed the network measures in each section. As we are interested in anticipating the switching in real time, we want to consider sections as short as possible, but on the other hand, they can not be too short as we need to compute the weights of the links with good statistics. We found that using sections of \( L = 500 \) data points each, and OPs of \( D = 3 \), provided a good compromise, because we have 6 nodes, 36 links and their probabilities can be computed, in each section, with sufficient statistics.

The results are presented in figure 9, which displays the distribution of \( s_p, s_n \) and \( a_c \) values computed over the 1000 independent realizations: the left column corresponds to the data set in which PS occurs, while in the right column, no PS occurs. We observe that \( s_p \) varies over a wider range of values with respect to \( s_p \), and, when approaching the switching, both \( s_n \) and \( s_p \) increase reaching similar values. The asymmetry coefficient decreases

---

**Figure 7.** As in figure 6 but now we analyze the effect of the length of the time series: in the left column, \( L = 500 \); in the central column, \( L = 1000 \) and in the right column, \( L = 5000 \). In each panel \( D = 3 \). We note that, as the length of the data set increases, the size of the error bars diminishes.

**Figure 8.** Second empirical data set recorded from a semiconductor laser with orthogonal optical feedback: PDF of the intensity (in color code) of the linear polarization mode that turns off versus time. The PDF is computed by averaging over 1000 time series.
only slightly when approaching the PS and thus is not a good indicator of the switching ahead (this is probably because of the short section length, see the influence of $L$ in the analysis of the first empirical data set, figure 7).

5. Conclusions and discussion

We have used a hybrid methodology to characterize the evolution of a dynamical system that combines two mathematically grounded tools for time series analysis: symbolic ordinal analysis and network representation. Specifically, time series were transformed in sequences of OPs that were used to construct directed and weighted graphs. Exploiting this graph representation we propose several novel diagnostics to characterize time series: the averaged node entropy, the entropy of the distribution of links’ weights, and the asymmetry coefficient. The analysis of numerical data (generated from the logistic, tent, and circle maps) and experimental data (recorded from the output of semiconductor lasers under different external perturbations) demonstrated that these network-based measures are suitable for characterizing time series, adequately capturing the effects of parameter changes that result in subtle or sudden transitions.

The links entropy is an interesting quantity for measuring the complexity of the symbolic sequence associated to a time series, because it is equal to 0 when the weights of the links are delta-distributed, which occurs when the symbolic sequence is either perfectly regular or fully stochastic. In agreement with a previous study that found maximum complexity of the logistic map at intermediate values of the map parameter $r$ [39], here we found that the links’ entropy is maximum at intermediate values of $r$.

The analysis of the empirical data confirmed that network-based measures provide early warning signals of abrupt transitions, despite the highly stochastic character of the signals analyzed. These network-based networks can thus complement other indicators, such as the standard deviation or the PE. The methodology proposed

![Figure 9. Comparison between 1000 independent realizations in which polarization switchings occur (left column) and 1000 independent realizations in which no switching occurs (right column, see text for details). Top row: the PE; middle row: the average node entropy; bottom row: the asymmetry coefficient. The analysis is performed with $D = 3$, and each time series is divided in 40 non-overlapping sections of $L = 500$ data points each. $s_p$ and $s_n$ are normalized to the maximum value, $\log D!$. While the dispersion of values is quite large (due to the short length of each section), in the left column there is a clear trend in the distributions of $s_p$ and $s_n$ values when approaching the switching point. In comparison the variation of $a_c$ is small and thus, $a_c$ does not constitute a good indicator of the PS. We observe that $s_n$ varies over a wider range of values with respect to $s_p$ and, when approaching the switching, both $s_p$ and $s_n$ increase reaching similar values.](image)
here can be a valuable tool for the analysis of a wide range of systems where critical transitions might occur, including population extinctions, desertification, wetland degradation, and epileptic seizures, to name a few.

Our approach could allow identifying changes in the symbolic dynamics of these systems, for example, the appearance of new symbols or the appearance of new transitions, which result in variations of the network measures. Bifurcations from a chaotic attractor to another chaotic attractor could be identified if the two attractors have associated symbolic dynamics that have different statistical distribution of symbols and TPs. On the contrary, if a bifurcation does not result in changes in the symbolic dynamics of the system, the network measures proposed here won’t be able to capture the bifurcation (see, e.g., the two time series displayed in figure 1, which, with $D$ being 2, 3, or 4, OPs are encoded in the same symbolic sequence).

Another important drawback is that both, periodic and irregular time series could eventually be mapped to the same class of regular networks. Note that this methodology is not intended to classify between different time series, but to detect changes in the time series as a function of the tuning parameter. Thus, this methodology should be only used to complement other conventional tools of time series analysis if the goal is a classification. In spite of these drawbacks, our approach provides additional predictive power for time series analysis, as identifying the nodes that have one (or a reduced number) outgoing link will allow identifying ‘predictable symbols’ in the symbolic sequence, those for which we know which is the next most probable symbol in the symbolic sequence. If ordinal analysis is used to construct the symbolic network, then the information gained about the next most probable symbol will not allow inferring the variations of values in the time series, but it will allow inferring the shape of the future oscillation in the time series. In this sense, the symbolic network–based analysis proposed here will provide complementary information to that gained by using other well-established tools for time series analysis.

Acknowledgments

This work has been supported by the Spanish MINECO (FIS2012-37655-C02-01 to CM, FIS2012-38266 to AA and SG), the Generalitat de Catalunya (ICREA Academia to CM and AA), EOARD (Grant FA9550·14–1·0359) to CM, the European Commission FET-Proactive Project MULTIPLEX (Grant 317532) to AA and SG, and the James S McDonnell Foundation to AA.

References

[10] Li X L, Cui S Y and Voss L J 2008 Using permutation entropy to measure the electroencephalographic effects of sevoflurane Anesthesiology 109 448–56
[27] Sun X, Small M, Zhao Y and Xue X 2014 Characterizing system dynamics with a weighted and directed network constructed from time series data Chaos 24 024402
[34] Lacasa L and Toral R 2010 Description of stochastic and chaotic series using visibility graphs Phys. Rev. E 82 036120
[38] Garlaschelli D and Loffredo M I 2004 Patterns of link reciprocity in directed networks Phys. Rev. Lett. 93 268701
[40] Panajotov K and Prati F 2013 VCSELs (Springer Series in Optical Sciences) ed R Michalzik pp 181–231