Estimating topological entropy using ordinal partition networks

Konstantinos Sakellariou [1,2,*] Thomas Stemler [1,†] and Michael Small [1,3,‡]

[1]*Complex Systems Group, Department of Mathematics & Statistics, The University of Western Australia, Crawley WA 6009, Australia*
[2]*Nodes & Links Ltd, Leof. Athalassas 176, Strovolos, Nicosia 2025, Cyprus*
[3]*Mineral Resources, CSIRO, Kensington WA 6151, Australia*

We propose a computationally simple and efficient network-based method for approximating topological entropy of low-dimensional chaotic systems. This approach relies on the notion of an ordinal partition. The proposed methodology is compared to the three existing techniques based on counting ordinal patterns—all of which derive from collecting statistics about the symbolic itinerary—namely (i) the gradient of the logarithm of the *number of observed patterns* as a function of the pattern length, (ii) direct application of the definition of *topological permutation entropy*, and (iii) the *outgrowth ratio* of patterns of increasing length. In contrast to these alternatives, our method involves the construction of a sequence of complex networks that constitute stochastic approximations of the underlying dynamics on an increasingly finer partition. An *ordinal partition network* can be computed using any scalar observable generated by multidimensional ergodic systems, provided the measurement function comprises a monotonic transformation if nonlinear. Numerical experiments on an ensemble of systems demonstrate that the logarithm of the spectral radius of the connectivity matrix produces significantly more accurate approximations than existing alternatives—despite practical constraints dictating the selection of low finite values for the pattern length.

I. INTRODUCTION

Topological entropy is a dynamical invariant that constitutes one of many tools aimed at measuring the complexity of a specified dynamical regime. Intuitively it quantifies the exponential rate at which the number of *distinguishable* orbits grows with partition refinement, i.e., with respect to increasing sampling resolution. This notion was originally developed by Adler *et al.* [1] via assignment of non-negative numbers to open covers of compact topological spaces in order to measure their size, inspired by ideas of Kolmogorov and Tihomirov [2] and the definition of Kolmogorov-Sinai (KS) entropy [3–5]. The variational principle relates topological entropy to KS entropy [6–9]. A different definition that applies to metric spaces was introduced independently by Dinaburg [8] and Bowen [10]. Equivalence between the two definitions was established shortly after [11].

Estimating topological entropy is a notoriously difficult pursuit, significantly more so in comparison to other dynamical invariants such as the maximal Lyapunov exponent or fractal dimensions. A number of studies have investigated techniques to tackle this estimation problem from a range of viewpoints; kneading theory [12,13], periodic and cyclic properties [14,15], Markov models [16,17], sofic shifts [18], and length growth algorithms [19]—all of which involve symbolic dynamics. Additionally, many of these methods essentially reduce to constructing a matrix and examining its spectrum. Originally, Chomsky and Miller [20] used this idea to calculate the number of grammatical strings of any given length within the context of finite-state languages, inspired by the notion of channel capacity by Shannon [21].

In this study our focus is on the approximation of the topological entropy by means of *ordinal patterns*. This term refers to a computationally fast approach that became popular among practitioners over the last decade. In particular with *permutation entropy*, the metric and topological version of which were originally introduced by Bandt *et al.* [22]. The type of dynamical regimes that we are concerned with are deterministic, low-dimensional—i.e., low number of active degrees of freedom even if the dimension of the ambient space is high—*measure-preserving* systems that display chaotic behavior (sensitivity to initial conditions, topological mixing, existence of dense periodic orbits). The main requirement underpinning the Markovian formalism employed here, established in [23], is for the dynamics to fulfill the ergodicity property, which guarantees a form of "typicality." The collection of iteration maps that we study are all regular maps, with evolution rules mostly given by polynomial forms. Several publications on real-world applications followed the analysis performed on an audio recording in the seminal study [24], including EEG [25–29] and ECG interbeat interval [30–32] recordings to name a few.

Symbolic dynamics has traditionally proved itself a useful tool in the study of dynamical systems. It involves discretization of the underlying state space into a finite number of regions, each of which is uniquely labeled by means of a finite *alphabet* of symbols. A symbol sequence is generated

*kostas@nodeslinks.com
†thomas.stemler@uwa.edu.au
‡michael.small@uwa.edu.au

by recording each region visited by the specific orbit the system adheres to given a specified initial condition. Thus, every dynamical orbit may be associated with an *itinerary* (of infinite length), a sequence of labels that represent the order in which successive coarse-grained states of the system manifest as it evolves in time. The resulting symbolic orbit shares properties with the original dynamics. Moreover, in the case of generating covers, the association between orbits and itineraries is injective.

The Bandt-Keller-Pompe paradigm [22,24,25,33,34] involves counting the frequency of occurrence of types of subsequences in the symbolic itinerary. The alphabet of choice comprises the set of permutations of the natural numbers $\{1, 2, 3, \ldots, m - 1, m\}$, which are assigned to equal-length subsequences of successive time series points. Symbolization is performed on the basis of *ordinal relations between points within each segment* and is, therefore, representative of the visual pattern formed by the respective segment. This paradigm rapidly gained popularity in the nonlinear time series analysis community due to its simplicity and ease of computation.

A slightly different approach (in the order of an iterated limit between pattern length and length of the orbit) by Amigó and colleagues was developed later. Amigó *et al.* [35] generalized the equality between (i) metric entropy and (ii) metric permutation entropy for piecewise monotone maps on one-dimensional intervals, established in [24], to higher dimensional *ergodic* systems. Amigó and Kennel [36] established a parallel result for the topological permutation entropy, generalizing the other main finding of [24], by requiring that the deterministic map is *expansive*.

There are three existing ordinal methods for estimating topological entropy. The first involves the exponential growth rate of the number of patterns observed in the symbolic itinerary as their length increases. The second uses the definition of topological permutation entropy [22,36]. The third revolves around the concept of the outgrowth ratio of increasingly longer patterns. We propose a more accurate estimator by exploiting the use of complex networks as representations of the coarse-grained dynamics. We use the ordinal partition network formalism introduced by the authors in [23].

A commonplace challenge in time series analysis is that statistical estimates computed from a single long deterministic trajectory can sometimes produce misleading results. A robust workaround transpires from perturbing a dynamical system by a small amount of noise, which has been recognized as an alternative to direct analysis for some time [37]. This is achieved by means of a symbolic approach. The key factor to a successful probabilistic description of deterministic dynamics hinges on finding an appropriate finite *covering* of state space. The germane notion is that of a *Markov partition* [38]. Typically a regular-grid partitioning of state space is used in practice, an approach which derives from Ulam's method [39]. The Markov model approximates the evolution of densities in state space by the action of the so-called *Perron-Frobenius* operator [40] and is easily calculated numerically only for systems of low dimension [18].

Instead of employing regular-grid or more complicated partitions that necessitate engagement with more sophisticated techniques or require partial knowledge about the dynamics, we use a standardized partitioning method [33],

namely an ordinal partition computed according to the Bandt-Keller-Pompe paradigm. In [23] it was demonstrated that counting patterns is equivalent to defining an ordinal partition that becomes finer as the observed pattern size is augmented. A Markovian framework is, thus, formulated to capture topological and transitivity properties. The adjacency matrix of the unweighted directed Markov graph, i.e., the binary *connectivity* matrix, encapsulates all possible transitions between state space regions that the symbolic trajectory can exhibit. Adapting the method established by Chomsky and Miller [20] to estimate lengths of words composed of binary symbols, we approximate the topological entropy of the underlying system by computing the logarithm of its spectral radius for increasingly finer ordinal partitions. Numerical results indicate that network representations, which incorporate information on transitions between symbols (i.e., the space of ordered symbol pairs)—as opposed to simply recording statistics on the symbol space itself—demonstrate merits and increased reliability in topological entropy estimates over traditional ordinal analysis approaches.

II. METHODOLOGY

Let $\{\hat{x}_n\}_{n=1}^N$ denote an arbitrary scalar time series of N temporally ordered measurements of the state of a multidimensional discrete dynamical system. The system is defined by the pair $(\mathcal{M}, \phi)$, where the single-valued C^r function $\phi : \mathcal{M} \to \mathcal{M}$ represents the action of a rule for temporal evolution within the system's state space, the compact manifold $\mathcal{M}$. Dynamics in d-dimensional space are described in terms of forward time iterates of the governing equation from a specified initial state via functional composition. In general, this equation is given by $\boldsymbol{x}_{n+1} = \boldsymbol{f}(\boldsymbol{x}_n; \boldsymbol{\theta})$, where $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}^d$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ is an ordered d-tuple, $\boldsymbol{\theta}$ corresponds to a set of parameters, and $n \in \mathbb{Z}$. The discrete-time *flow* which specifies the solution to this equation is defined by

$$\phi(\boldsymbol{x}_0, n; \boldsymbol{\theta}) = \phi_n(\boldsymbol{x}_0; \boldsymbol{\theta}) = \boldsymbol{f}^n(\boldsymbol{x}_0; \boldsymbol{\theta}) \qquad (1)$$

for some initial state $\boldsymbol{x}_0$. We assume that $\hat{x}_n$ comprises either a state variable or a nonlinear monotonic transformation thereof, i.e.,

$$\hat{x}_n = h \circ f_i^n(\boldsymbol{x}_0; \boldsymbol{\theta}), \qquad (2)$$

where the symbol $\circ$ denotes functional composition, f_i represents the ith component of $\boldsymbol{f}$ and the measurement function $h : \mathbb{R} \to \mathbb{R}$ fulfils the condition $x_1 < x_2 \Rightarrow h(x_1) < h(x_2)$.

Ordinal patterns can be computed by defining an ordinal partition [24,35] and recording the evolution of the associated symbolic trajectory. Statistics of this record are then used to formulate transition probabilities for a Markovian model acting as a perturbation to the original dynamics. The ordinal network encodes the rule under which all allowed Markovian orbits are generated. We briefly outline the details of the method developed by the authors in [23] and, following this, elaborate on how to estimate topological entropy using ordinal information.

A. Ordinal patterns and symbolic itinerary

Ordinal symbolization [24] consists of a coarse-grained segmentation of observable $\hat{x}_n$ into elements of equal length m. Within each such element—often referred to as a temporal *window* since it encapsulates a finite history of the reference trajectory at discrete moments in time—consecutive points may be separated by τ recordings in the time series. This procedure is equivalent to embedding the time series into m-dimensional space using a *time delay* τ. A third parameter, the *slide lag*, is introduced to modulate the level of correlation between windows—in a similar manner to how τ modulates correlation among points within each window. Successive windows are *not* overlapping at $w = w(m) \in \{1, m\}$ points, the two extremes of the spectrum being $w_{\max} = 1$ (almost full overlap) and $w_{\max} = m$ (zero overlap).

The sequence of windows obtained via this process can be formally represented by an indexed sequence of m-tuples of the form

$$\hat{\boldsymbol{x}}_i^{(m)} = (\hat{x}_i, \hat{x}_{i+\tau}, \dots, \hat{x}_{i+(m-1)\tau}) \in \mathbb{R}^m, \tag{3}$$

where the index takes the values $i = 1, w+1, 2w+1, \dots, \lfloor \frac{N-(m-1)\tau}{w} \rfloor + 1$ to preserve the original temporal order. The total number involves the floor function due to the constraint $\max i \leqslant \frac{N-(m-1)\tau}{w}$ imposed by the time series length. For instance, to demonstrate the procedure on a concrete example, consider a 12-point long observable, $\hat{x}_{n=1}^{12}$, segmented into windows of length $m = 3$ with, say, delay $\tau = 2$. Window intercorrelation is minimal in the *nonoverlapping* scheme, whereby $w = w_{\max} = 3$ is maximal, and leads to the ordered window sequence

$$\hat{\boldsymbol{x}}_1^{(3)} = (\hat{x}_1, \hat{x}_3, \hat{x}_5),$$
$$\hat{\boldsymbol{x}}_2^{(3)} = (\hat{x}_4, \hat{x}_6, \hat{x}_8),$$
$$\hat{\boldsymbol{x}}_3^{(3)} = (\hat{x}_7, \hat{x}_9, \hat{x}_{11}).$$

We remark that the slide lag parameter can have its definition extended to $w = w(m, \tau) \in \{1, \tau, \dots, m\tau\}$, so that $w_{\max} = m\tau$ corresponds to zero overlap in a *temporal* sense as well. We refer to $w_{\max} = m$ as the *nonoverlapping scheme* in reference to a "sliding window" procedure over segments—if $\tau = 1$, this scheme is nonoverlapping in a temporal sense too. However, should one opt for a higher delay, there is a distinction to be made between *window* and *temporal* overlapping and the slide lag should be defined accordingly. This would, for instance, lead to the predecessor-successor pairs of the form

$$(\hat{x}_1, \hat{x}_3, \hat{x}_5) \rightarrow (\hat{x}_7, \hat{x}_9, \hat{x}_{11}),$$
$$(\hat{x}_2, \hat{x}_4, \hat{x}_6) \rightarrow (\hat{x}_8, \hat{x}_10, \hat{x}_{12}),$$

in the above example.

Performing the same operation using the *maximally overlapping* scheme, whereby $w = w_{\min} = 1$ is minimal, leads to the richer collection

$$\hat{\boldsymbol{x}}_1^{(3)} = (\hat{x}_1, \hat{x}_3, \hat{x}_5), \quad \hat{\boldsymbol{x}}_2^{(3)} = (\hat{x}_2, \hat{x}_4, \hat{x}_6), \quad \hat{\boldsymbol{x}}_3^{(3)} = (\hat{x}_3, \hat{x}_5, \hat{x}_7),$$
$$\hat{\boldsymbol{x}}_4^{(3)} = (\hat{x}_4, \hat{x}_6, \hat{x}_8), \quad \hat{\boldsymbol{x}}_5^{(3)} = (\hat{x}_5, \hat{x}_7, \hat{x}_9), \quad \hat{\boldsymbol{x}}_6^{(3)} = (\hat{x}_6, \hat{x}_8, \hat{x}_{10}),$$
$$\hat{\boldsymbol{x}}_7^{(3)} = (\hat{x}_7, \hat{x}_9, \hat{x}_{11}), \quad \hat{\boldsymbol{x}}_8^{(3)} = (\hat{x}_8, \hat{x}_{10}, \hat{x}_{12}),$$

TABLE I. Number of windows obtained for arbitrary window length m and time series length N for four distinct segmentation schemes.

Scheme	Zero overlap $(w = m)$	Maximal overlap $(w = 1)$
Minimal delay ($\tau = 1$)	$\lfloor \frac{N-m}{w} \rfloor + 1$	$N - m + 1$
Larger delay ($\tau \geqslant 1$)	$\lfloor \frac{N-(m-1)\tau}{w} \rfloor + 1$	$N - (m-1)\tau$

where correlation between windows is at its highest. Table I displays the total number of windows obtained for the two extreme cases of *interwindow* correlation (A) $w = w_{\min}$ and (B) $w = w_{\max}$, using a (I) single-point delay $\tau = 1$ or (II) a larger value $\tau \geqslant 1$ to distinguish between maximal and lower *intrawindow* correlation, for arbitrary window length m.

In [23] the authors prescribed a formulation based on the premise that that maximal intra- and interwindow correlation is preferable. Parameter selection herein is made according to this standard, we therefore choose a delay $\tau = 1$ and slide lag $w = 1$. This particular slide lag selection was motivated by a range of numerical results of [23]. First of all, periodicity of a periodic input trajectory can only be retrieved, i.e., via a random walk, from the network topology *exactly* (not at a lower multiple of the period) in the maximally overlapping scheme. Second, networks produced even from simple (periodic) dynamics are less sensitive to changes in the main parameter, the pattern length m, when $w = 1$. Third, the set of admissible patterns that are *eventually* observed is the same irrespective of w, but reaching the true number of patterns *for a fixed pattern length m* occurs for a much smaller sample size N in the case of the maximal overlap. Fourth, Markovianity of the underlying deterministic dynamics is not being fully preserved in the projection to the symbol space described below. The conditioning on predecessor-successor pairs leads to a reduction of undesired effects generated as a consequence, as was shown via numerical experiments on an ensemble of discrete maps and continuous-time flows [23]. In addition, the nonoverlapping variant $w = 1$ (with $\tau = 1$) produces poor estimates of topological entropy (method IV introduced in Sec. II C) being sensitive to the pattern length m and exhibiting no clear scaling (results omitted).

Delay selection is motivated by the fact that the value of unity constitutes a natural choice for discrete-time flows (since sample points are generated at a precision close to the maximum numerical limit) and coincides with the global minimum of the (self-)mutual information in chaotic maps. Additionally, this choice enables the prevention of the presence of false admissible patterns (due to undesired *aliasing* effects [41]). We propose using a Poincaré surface of section in the case of continuous-time flows (see [23]), otherwise the election of $\tau = 1$ can lead to false forbidden patterns. Both choices are further motivated by the richer collection of windows produced by the segmentation, as Table I portrays, which leads to a larger sample of symbols to draw statistics from.

Each window is mapped to a symbol drawn from a finite alphabet, in particular the set of permutations of the natural numbers $\{1, 2, \dots, m\}$. The *symbolic ordering* assigned to
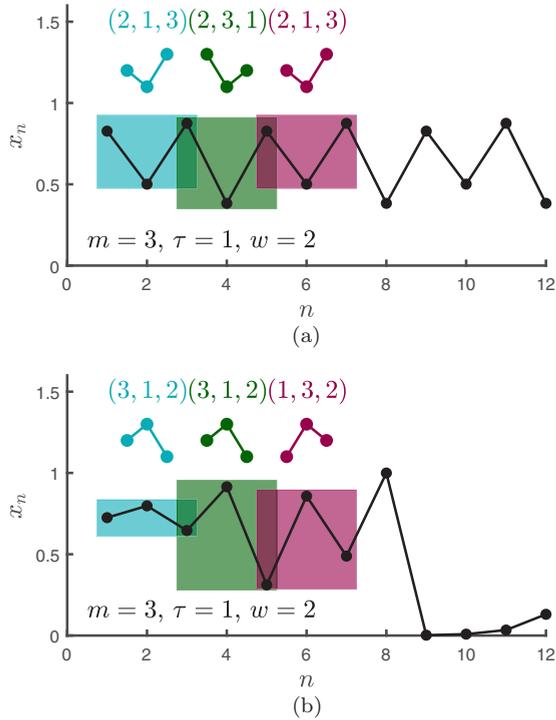
FIG. 1. Counting ordinal patterns with $m = 3$, $\tau = 1$, and $w = 2$ from a periodic (top $r = 3.5$) and a chaotic (bottom $r = 4$) trajectory generated by the logistic equation $x_{n+1} = r x_n (1 - x_n)$. (a) 4-periodic regime and (b) chaotic regime.

each window reflects the ordinal relations between points within. The mapping is defined by a chronological index ranking, according to which shifted time indices of points within a window are stored in a vector indexed by the corresponding amplitude rank in ascending order. For example, to identify the symbols assigned to windows $\hat{\boldsymbol{x}}_1^{(3)}$ and $\hat{\boldsymbol{x}}_2^{(3)}$ from the overlapping segmentation above, and assuming that the relative magnitude relations are given by, e.g.,

$$\hat{x}_3 < \hat{x}_{10} < \hat{x}_5 < \hat{x}_8 < \hat{x}_{11} < \hat{x}_9 < \hat{x}_7$$
$$< \hat{x}_1 < \hat{x}_2 < \hat{x}_6 < \hat{x}_{12} < \hat{x}_4,$$

we shift index sets $\{1, 3, 5\}$ and $\{2, 4, 6\}$ to the set $\{1, 2, 3\}$. Given that $\hat{x}_3 < \hat{x}_1 < \hat{x}_2$ and $\hat{x}_2 < \hat{x}_6 < \hat{x}_4$, the associated permutations are, respectively, (3,1,2) and (1,3,2). Figure 1 depicts the counting procedure on periodic and chaotic series of the logistic map, using $m = 3$, $\tau = 1$, and $w = 2$ to illuminate the role of each methodological parameter in the symbol sequence construction.

Formally, an *ordinal pattern* of length m for the window labeled by $\hat{\boldsymbol{x}}_i^{(m)}$

$$\boldsymbol{\pi}\left(\hat{\boldsymbol{x}}_i^{(m)}\right) = (\pi_1, \pi_2, \ldots, \pi_j, \ldots, \pi_m), \quad \pi_j \in \{1, \ldots, m\}$$

may be defined as the permutation $\boldsymbol{\pi} \in \mathcal{S}_m$ which arranges the points within according to their order, i.e., $\hat{x}_{\pi_1} < \hat{x}_{\pi_2} < \cdots < \hat{x}_{\pi_m}$. In the event that two elements of $\hat{\boldsymbol{x}}_i^{(m)}$ are equal, we arbitrarily pick the one that occurs first chronologically as the smallest [24]. This occurrence is quite rare in continuous-time dynamics, but may, in principle, be present in discrete dynamics depending on the system. The number of "ties"

recorded in our test systems seems to be negligible, but we suggest caution and some testing when faced with a new time series. If ties are frequent, this could impact the estimated empirical distribution of patterns.

A symbol sequence that corresponds to a given time series is generated via the mapping of windows to ordinal patterns. Note that in the case of the partition being generating, the association between orbits, of which a time series comprises a finite subset, and itineraries is injective. The ordinal symbolic *itinerary* of time series $\{\hat{x}_n\}_{n=1}^N$ is denoted by

$$\{s_n\}_{n=1}^{N-m+1}, \quad s_n \in \mathcal{S}_m, \quad \forall n \in \mathbb{N}. \tag{4}$$

The empirical distribution of ordinal patterns is defined via their relative frequency of occurrence within s_n,

$$\mathbb{P}[\boldsymbol{\pi}] = \frac{|\{n \text{ such that } s_n = \boldsymbol{\pi}\}|}{N - m + 1}, \quad \boldsymbol{\pi} \in \mathcal{S}_m, \tag{5}$$

where $|\cdot|$ denotes the cardinality of a set.

The set of all permutations of order m, $\mathcal{S}_m$, can be subdivided into two groups. The first contains all the *distinct* occurring patterns observed in itinerary s_n, called *admissible*. The second comprises the complementary set of all patterns that *cannot* be realized in a time series due to deterministic constraints, referred to as *forbidden* [42]. The number of admissible patterns is

$$\mathcal{N}(s_n, m) = |\{\boldsymbol{\pi} \in \mathcal{S}_m \text{ such that } \mathbb{P}[\boldsymbol{\pi}] \neq 0\}|. \tag{6}$$

Type I errors (false positives, i.e., false admissible patterns) can occur due to aliasing when $\tau > 1$. Type II errors (false negatives, i.e., false forbidden) can occur due to undersampling, e.g., if m is chosen too large for specified N—akin to overembedding.

B. Ordinal partition network

To construct an *ordinal partition network* (OPN), bijective mappings $\kappa : \mathcal{S}_m \to \mathcal{V}$ and $\psi : \mathcal{S}_m \times \mathcal{S}_m \to \mathcal{E}$ are applied to the (i) set of admissible symbols and (ii) product set of *ordered pairs* of symbols. The image of κ is the set of nodes (or vertices) $\mathcal{V} = \{1, 2, 3, \ldots, V\}$ of a graph, say $G(\mathcal{V}, \mathcal{E})$. The set of directed links (or edges) $\mathcal{E}$ is created via assignment of connections from source to target patterns occurring in temporal succession, i.e., nodes mapped to consecutive symbols in itinerary s_n are linked.

The set of links corresponds one-to-one to admissible *forward-time transitions* between regions of the ordinal partition (for details see [23]) as observed in the given time series. Note that the parameter choice $w = 1$ implies that the stochastic model represented by the graph is a *finite-memory* Markov process of order $m - 1$. False admissible transitions (type I error in the product space of symbol pairs) can occur if successive patterns are completely uncorrelated when selecting $w \geq m$. False forbidden transitions (type II) occur due to undersampling, as with the space of symbols, but effects are more pronounced in the distribution of symbol pairs for given partition refinement (dictated by m) and the same (finite) input trajectory.

(a)

(b)
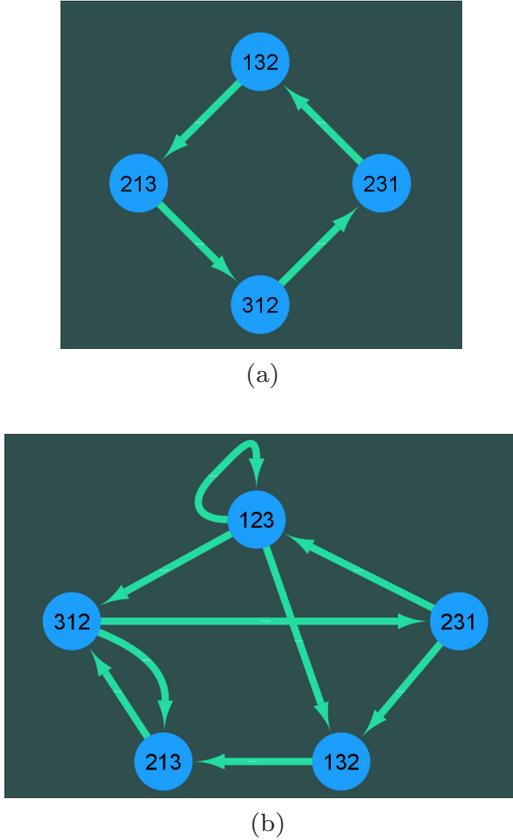
FIG. 2. Ordinal networks that represent coarse-grained dynamics on the $m = 3$ partition for the periodic and chaotic time series of Fig. 1. Parameter selection according to the recommendations in [23], i.e., $\tau = 1$ and $w = 1$ (instead of $w = 2$ as in the schematic of Fig. 1). (a) cycle graph C_4 and (b) complex cyclic network.

C. Topological entropy estimators

By construction the network consists of a single giant component without any disconnected islands or isolated nodes (Fig. 2 displays networks computed from the logistic trajectories of Fig. 1). The network size equals $|\mathcal{V}| = V(m) = \mathcal{N}(s_n, m)$. In practice, this is an estimator of the true number for the case $N < \infty$ since both experiments and numerical simulations can only record a trajectory but not the full orbit. The first method for approximating topological entropy tested in this study is the simplest, namely estimate the exponential growth rate of $V(m)$ for increasing m [42]. This estimator quantifies the rate of growth of the set of admissible patterns observed in the itinerary, or equivalently the network size, by increasingly refining the resolution of the ordinal partition. For piecewise monotone one-dimensional maps it has been proved [22] that $\mathcal{N}(s_n, m) \propto e^{mh_{\mathrm{TOP}}(f)}$. For finite m and N where the infinite limit cannot be computed, the estimator reduces to computing the slope of the semi-logarithmic $V - m$ graph

$$\gamma = \frac{\partial}{\partial m}(\ln \mathcal{N}(s_n, m)) \quad [\text{method I}] \quad (7)$$

while discarding outlier data points for sufficiently large values of m where $\mathcal{N}(s_n, m) \sim O(N)$.

Application of the direct definition of the *topological permutation entropy* (TPE) [22,36]

$$h_0^{\mathrm{PE}}(m) = \frac{1}{m-1} \log_2 \mathcal{N}(s_n, m) \quad [\text{method II}] \quad (8)$$

comprises the second method that we examine. In contrast to method I, this estimator generates different approximations for each value of m and is expected to be more sensitive to under-sampling when the partition is too fine (large m).

To introduce the third method, note that the absence of an m pattern, if truly forbidden, pervades all longer patterns for larger m in the form of outgrowth forbidden patterns whose growth is superexponential in contrast to admissible m patterns. The *outgrowth ratio* function of patterns of order m [36] is defined as the number of admissible patterns of order $m + 1$ that contain the given m-pattern as a "prefix." Formally, for $\boldsymbol{\pi} \in \mathcal{S}_m$

$$\mathfrak{g}(\boldsymbol{\pi}) = |\boldsymbol{\xi} \in \mathcal{S}_{m+1} \quad \text{such that} \quad \boldsymbol{\xi} = (\boldsymbol{\pi}, \pi_{m+1}),$$
$$\mathbb{P}[\boldsymbol{\xi}] \neq 0, \quad \pi_{m+1} \in \{1, 2, \ldots, m+1\}|. \quad (9)$$

Since the additional element π_{m+1} of $\boldsymbol{\xi}$, as compared to its prefix $\boldsymbol{\pi}$, is necessarily part of the alphabet $\{1, 2, \ldots, m+1\}$, at most $m + 1$ patterns of order $m + 1$ can be conditioned on the prefix pattern of order m—irrespective of the underlying dynamics. The third estimator is the natural logarithm of $\mathcal{G}(m)$, the outgrowth ratio averaged uniformly over all extant prefix patterns, namely

$$\mathcal{G}(m) = \frac{1}{\mathcal{N}(s_n, m)} \sum_{\boldsymbol{\pi} \in \mathcal{S}_m} \mathfrak{g}(\boldsymbol{\pi}) \quad [\text{method III}], \quad (10)$$

where $\boldsymbol{\pi}$ is required to be admissible. Note that if pattern $\boldsymbol{\pi}$ is forbidden, i.e., $\mathbb{P}[\boldsymbol{\pi}] = 0$, then necessarily $\mathbb{P}[\boldsymbol{\xi}] = 0$ for all $\boldsymbol{\xi} = (\boldsymbol{\pi}, \cdot)$. Similarly for all $\boldsymbol{\psi} = (\boldsymbol{\xi}, \cdot) = (\boldsymbol{\pi}, \cdot, \cdot)$ and, by induction, all patterns of the form $(\boldsymbol{\pi}, \cdot, \ldots)$ must be forbidden. These are precisely the outgrowth forbidden patterns (with respect to $\boldsymbol{\pi}$) mentioned above.

The method proposed herein is based on the $\mathcal{N} \times \mathcal{N}$ connectivity matrix $\boldsymbol{A}(m)$ of the ordinal partition network [23]. This is essentially the adjacency matrix of the network with all nonzero entries mapped to unity (to remove weights representing the frequency of occurrence of each transition in s_n). The fourth estimator is defined as the natural logarithm of the spectral radius

$$\rho[\boldsymbol{A}(m)] = \max(\lambda_1, \lambda_2, \ldots, \lambda_{\mathcal{N}}) \quad [\text{method IV}], \quad (11)$$

where $\lambda_1, \ldots, \lambda_{\mathcal{N}}$ are the eigenvalues of $\boldsymbol{A}(m)$.

TABLE II. The four estimators of topological entropy examined within our numerical experiments. Note that TPE is defined in terms of logarithm base 2; for a fair comparison a change of basis is applied.

Estimator	Notation
Growth rate of admissible patterns	γ
Topological permutation entropy	$h_0^{\mathrm{PE}}(m) \times \ln 2$
Outgrowth ratio logarithm	$\ln \mathcal{G}(m)$
Spectral radius logarithm	$\ln \rho[\boldsymbol{A}(m)]$

TABLE III. Numerical experiments were conducted on a range of synthetic time series generated by these systems.

System	$f(\boldsymbol{x})$	Parameters	Regime	DIM	Volume	h_{TOP}
Logistic	$rx(1-x)$	$r = 4$	chaos	1	expanding	0.693
Cubic	$(a-1)x - ax^3$	$a = 4$	chaos	1	expanding	1.089
Bent baker	$\frac{4}{3}\sqrt{6}x^3 - 2\sqrt{6}x^2 + (2 + \frac{2}{3}\sqrt{6})x \mod 1$	–	chaos	1	expanding	0.693
Gauss	$e^{-\alpha x^2} + \beta$	$(\alpha, \beta) = (6.2, -0.486)$	chaos	1		
Dyadic (or Bit-shift)	$\mu x \mod 1$	$\mu = 2$	chaos	1	expanding	0.693
Tent	$\begin{cases} \mu x, & \text{if } x < 1/2 \\ \mu(1-x), & \text{if } x \geqslant 1/2 \end{cases}$	$\mu = 2$	chaos	1	expanding	0.693
Lozi	$(x_2, 1 + bx_1 - a\lvert x_2\rvert)$	$(a, b) = (6/5, -2/15)$	chaos	2	dissipative	0.300
Henon	$(1 - ax_1^2 + x_2, bx_1)$	$(a, b) = (1.4, 0.3)$	chaos	2	dissipative	0.465
Folded towel	$\big(ax_1(1-x_1) - 0.05(x_2+0.35)(1-2x_3),$ $0.1[(x_2+0.35)(1+2x_3) - 1](1-1.9x_1),$ $3.78x_3(1-x_3) + bx_2\big)$	$(a, b) = (3.8, 0.2)$	hyperchaos	3	dissipative	–

In summary, the four methods that we test as estimators of the topological entropy of discrete maps are displayed in Table II.

III. NUMERICAL EXPERIMENTS

We conduct a range of numerical experiments on an ensemble of well-studied discrete deterministic systems within parameter regimes known to exhibit chaotic behavior (see Table III for details). All input time series were simulated for twice the number of iterations signified by their stated length to remove transients. The accuracy and computational efficiency of the four estimators described in Sec. II C is evaluated on the basis of five separate tests presented below.

A. Test I: Logistic map

For a first comparison between methods I–IV, we consider a chaotic trajectory generated via iteration of the logistic map with $r = 4$, where topological entropy is known to be exactly equal to $\ln 2 \simeq 0.6931$. Figure 3 displays the estimates produced by each technique by sequential application of ordinal symbolization for different pattern lengths, in the range $2 \leqslant m \leqslant 20$.

Evidently the network-based estimator (method IV; circle markers) outperforms the three pattern-based methods.

It constitutes the only technique that produces an accurate estimate. Method III (square markers), which is based on the outgrowth ratio, is superior to directly applying the TPE definition (method II; diamonds), but only marginally better than the approximation produced by the growth rate of admissible symbols (method I; red dashed line). In fact, γ is close to the average over the estimates of $\ln \mathcal{G}(m)$ within the region $7 \leqslant m \leqslant 15$.

In addition to accuracy, method IV exhibits a clear scaling region in parameter space, for values where the pattern length (or, alternatively, embedding dimension) is within $4 \leqslant m \leqslant 17$. If $m < 4$, the refinement of the partition is not sufficiently detailed to allow the Markov approximation to unveil the true complexity of the dynamics. Recall that for $m = 3$ the partition consists of $3! = 6$ elements, only 5 of which are visited in this case (for elaborate explanations see [23], and specifically Figs. 5 and 19(b) therein). Typically if the partition is too coarse and the underlying dynamics is volume expanding, the spectral radius of $\boldsymbol{A}(m)$ severely underestimates topological entropy. This also occurs if m is too large due to undersampling. The length of the logistic trajectory used to compute the symbolic itinerary for Fig. 3 is equal to 10 million points.

FIG. 3. Methods I–IV applied to a time series of length $N = 10^7$ points generated by the logistic map $x_{n+1} = rx_n(1-x_n)$ with $r = 4$.
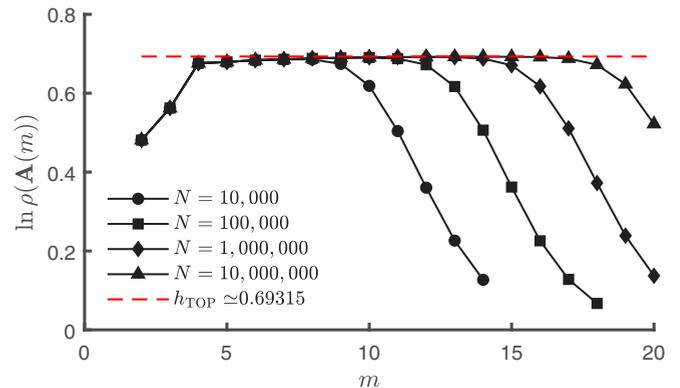
FIG. 4. Juxtaposition of method IV estimates applied to the logistic trajectory of Fig. 3 and shorter subseries thereof (of length $N = 10^4$, 10^5, 10^6). The scaling region of this estimator elongates as the sample size increases, improving the approximation's accuracy for a larger range of pattern lengths.
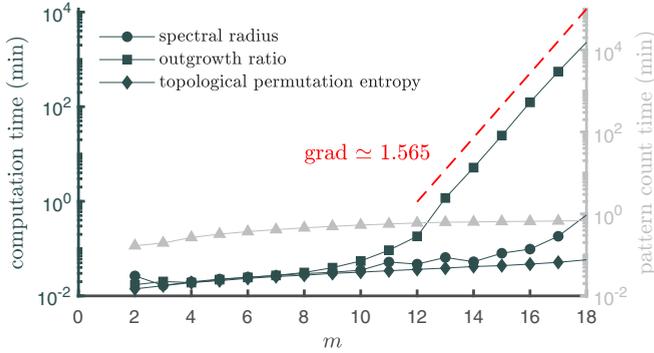
FIG. 5. Left axis: Required computation time (in minutes) for methods II–IV to produce the results displayed in Fig. 3 on an Intel Core i7-8700 CPU @ 3.20 GHz processor with 16 GB of available RAM. Right axis: Time required to compute the symbolic itinerary from the input series of length $N = 10^7$ for patterns of length up to $m = 18$.

For the complexity of the $r = 4$ chaotic regime, partitions with $m \geqslant 18$ are finer than constraints (dictated by this sample size) will allow for sufficient sampling of all their elements. Regions of the ordinal partition may not be visited by the trajectory simply due to not allowing adequate time.

To test further the dependence of method IV on the sample size and evaluate how its accuracy is affected, we computed networks from increasingly longer segments that comprise subsequences of the same trajectory as in Fig. 3. In particular, we used shortened input series of length $N = 10^4$, 10^5, and 10^6. The results, shown in Fig. 4, are indicative of the length and robustness of the scaling region in parameter space.

As expected, the longer the trajectory, the smaller the effects of undersampling. The scaling region elongates and an accurate estimate for a wider spectrum of values of m is obtained. In fact, even for $m = 17$ where the ordinal partition is very refined (17! elements, $\mathcal{N}(\{s_n\}_{n=1}^{N=10^7}, 17) = 486\,689$ visited regions), the estimate is very close to the ones computed with lower values; this is more important if higher-dimensional dynamics are under examination because higher refinement is required for the Markov model to capture the full complexity of the underlying dynamics. We remark that given a time series of specified length, there exists a threshold value beyond which m is too large. In this case, for $m \geqslant 19$ several regions of the partition are left unpopulated due to insufficient data, hence the Markovian approximation lacks faithfulness, the estimator performs poorly and the topological entropy estimates gradually decay to zero as m increases.

Overall, the results of Fig. 4 are very encouraging as far as method IV is concerned. The ordinal partition, even if not generating for finite values of m, can produce rather faithful stochastic descriptions of one-dimensional chaotic dynamics. This is true even for shorter time series, e.g., of $N = 10\,000$ points, as long as the selected value of m lies within the scaling region ($4 \leqslant m \leqslant 9$ here). Given that the theoretical results in [22–35] involve infinite limits, it is unequivocally surprising that small finite values such as $m = 4$ can lead to ordinal networks whose connectivity structure encapsulates the complexity of the $r = 4$ logistic dynamics to such an extent as to enable accurate topological entropy estimates—even

for shorter trajectories whose length is well within the usual measurements procured by real-world experimental setups.

Another facet of comparison between the methods is the total computational time required by each. Figure 5 illustrates the difference (in minutes) between methods II, III, and IV (diamond, square, and circle markers, respectively) on a semi-logarithmic axis for the parameter range $2 \leqslant m \leqslant 18$ (method I only requires a linear curve fit on the m-$\ln \mathcal{N}(s_n, m)$ graph, hence results are omitted due to the negligible time requirement). The time necessary for computing the symbol sequence from the input time series is common to all methods and is, hence, not included in the displayed results. For completeness we show it on the right axis of Fig. 5 (triangle markers). The short times ($\leqslant 1$ min), even for values on the right end of the spectrum, demonstrate the computational efficiency of ordinal symbolization (recall that the trajectory consists of $N = 10^7$ points).

When $m \leqslant 14$, all three estimators (II–IV) require less than 10 min, which is quite fast given the length of the input series. Beyond this value, the outgrowth ratio (method III) is by far the most computationally costly; clearly the required time grows exponentially when $m \geqslant 12$ at an approximate rate of 1.56, i.e., almost double the rate of growth of observed patterns as m increases ($\gamma \simeq 0.79$). This is presumably due to the fact that $\mathcal{G}(m)$ requires a search for *subsets of length m* within the set of admissible patterns of order $m + 1$—whose cardinality $\mathcal{N}(s_n, m + 1)$ is about twice as large (more precisely, around $e^{0.79} \simeq 2.203$) as the set of admissible patterns of order m, $\mathcal{N}(s_n, m)$. In contrast, the computation of all the other estimators is fast for all values of m.

We note that methods I and II are rather light timewise since they constitute very simple computations on statistics of the itinerary s_n that has already been computed. Method IV is also very efficient since less than a minute is required to arrive at an estimate even for $m = 18$ for this $N = 10\,000\,000$-long observable—where $\text{DIM}[A(m)] \sim O(10^6)$. The short time is owing to the sparsity of the connectivity matrix and the fact that modern tools for calculating the largest eigenvalue of an irreducible matrix (see Sec. IV in [23] for details of the ordinal Markov process, in particular Secs. IV B and IV C on ergodicity and irreducibility) are very efficient.

B. Test II: One-dimensional expansive maps

We extend our analysis to a collection of one-dimensional expansive dynamics. In particular we examine an additional five well-known maps. For the dyadic (or doubling and bit-shift) map and the tent map, we consider dynamical regimes that are topologically semiconjugate and conjugate, respectively, to the logistic $r = 4$ chaos. Topological entropy is an invariant for topological conjugacy [10]. Consequently, their topological entropy is known to be equal to $\ln 2$. This result has been proved to be true in greater generality for the case of shift maps, such as the dyadic, of the form $f(x) = \mu x \mod 1$ for $\mu \in \mathbb{N}$, where $h_{\text{TOP}}(f) = \ln \mu$ [43]. In the case of the family of tent maps parametrized by μ, $h_{\text{TOP}}(f) = \ln \mu$ provided $1 < \mu \leqslant 2$ and identically zero for $0 \leqslant \mu \leqslant 1$ [13]. The remaining three examples that we present comprise chaotic regimes drawn from the cubic, bent baker, and Gauss maps (see Table III), whose entropy is at present not found within the literature.
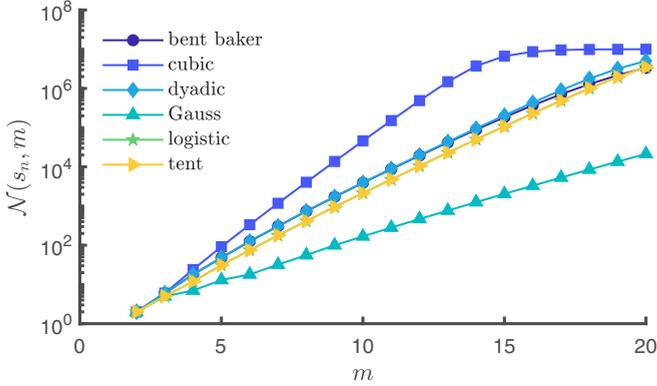
FIG. 6. Growth of the number of admissible patterns (i.e., network size) as patterns become longer for six one-dimensional expansive chaotic maps.

Results on all five maps are computed from samples equal in size to our primary example series, the logistic trajectory, and juxtaposed for comparison. Figure 6 shows the growth of $\mathcal{N}(s_n, m)$ (and hence the network size) with m on a semi-logarithmic scale. Intuitively, one expects that the more complex the dynamical regime, the larger the growth rate of admissible patterns. If the dynamics in a pair of regimes is topologically conjugate, then the curves should coincide, provided that sampling of the orbits is sufficient.

Evidently the logistic (green star markers), dyadic (light blue diamonds), and tent (yellow right arrows) regimes admit ordinal networks whose size is almost identical for all m. This is a first indication of their complexity being "similar," at least from a macroscopic point of view. The bent baker regime (dark blue circles) exhibits a very similar growth. Clearly the cubic regime (blue squares; whose state space $[-1, 1]$ is double the size of the logistic state space in the Lebesgue measure) produces networks of larger size as compared to all other dynamics for the same pattern length. Moreover, the cubic map's network size exhibits a faster saturation to its upper bound, i.e., the total number of time series points, whereby the network has collapsed to a long serial connectivity chain—a trivial structure that does not encode any useful information about the underlying dynamics. In this situation, the Markov model is no longer useful since the partition is too refined and undersampling has reached a maximum. The Gauss regime, on the other hand, seems to be less complex and sufficiently well sampled for any $2 \leqslant m \leqslant 20$.

Figure 7 depicts the topological entropy estimates using methods II (dark blue diamond markers) and IV (light blue circles) for all six one-dimensional dynamics. Method III requires significant computation time and has not been computed, except for two examples (1D logistic, 2D Lozi). The ordinal spectral radius not only leads to more accurate estimates than directly computing topological permutation entropy, but also generates long scaling regions, which verifies the results observed on the logistic map. Furthermore, we can confirm that the estimate for the logistic, dyadic, and tent regimes is approximately equal—and the same holds true for the bent baker trajectory.

While their ordinal structure and the associated networks for these four regimes are similar, they are not identical even

for the same pattern length, since the trajectories are finite. In theory, one would expect identical networks in the double limit $m \to \infty$ and $N \to \infty$ for topologically equivalent dynamics. It is, therefore, remarkable that method IV performs so well for such small values of m. Additionally, comparing these four regimes, we notice that the scaling region is not equally long for all, which implies that undersampling may appear for lower m in a subset of them. This effect is more pronounced in the bent baker regime. The underlying dynamics could be also characterized by a topological entropy equal to $\ln 2$, as with the other three regimes, as results indicate. However, it is unknown if it is indeed topologically conjugate to the other dynamics. Dynamical complexity manifests in several properties, such as sensitivity to initial conditions or fractal dimensionality (active degrees of freedom), hence the discrepancy in undersampling for a fixed value of m in the bent baker map could be the result of a difference in nontopological (e.g., metric) properties.

Comparing the three equivalent regimes, the dyadic seems to incur undersampling slightly earlier, e.g., the estimate reads 0.4 for $m = 20$, whereas the logistic and tent estimates are higher. In theory larger m, i.e., a finer partition, is necessary to obtain the true topological entropy not only to sufficient accuracy but also to arbitrary *precision*. Hence, the minor discrepancies between the three regimes could be attributed to this limitation of lower m values and would disappear as m grows, provided a sufficiently long sample for fixed m.

Another important aspect worth noting is that while the ordinal partition can be a *generating partition* in the $m \to \infty$ limit for these 1D maps [22], it is not necessarily so in the case of finite m. This holds true only in the case of the dyadic map, even for $m = 2$, since the diagonal line segment $\{(x, y) \in \mathbb{R}^2$ such that $x = y\}$ constitutes a generating partition by definition. It also appears to be true for the bent baker map; in fact the $m = 2$ estimates are equal to $\ln 2$ to numerical precision for these two maps, and arbitrarily close (>12 decimal digits) for all $m \leqslant 8$ in the bent baker case and $m \leqslant 13$ in the dyadic case. Consequently, discrepancies in estimates among topologically conjugate dynamics are possible for finite m depending on how close to generating the ordinal partition is.

Finally, as expected, the estimate for the cubic map is much higher ($\simeq 1.095$) than all other dynamics, whereas much lower for the Gauss regime ($\simeq 0.407$), confirming the suspicion of, respectively, higher and lower complexity than the four maps examined above. Notice the clear, albeit shorter, scaling region generated by the computed networks in the cubic case. The usual culprit, undersampling, shortens the scaling with m because longer trajectories are required in order to populate all the visited elements of the partition. Many more elements are occupied in the cubic regime due to both higher complexity as well as a larger support of the natural invariant measure (in the Lebesgue sense).

C. Test III: Bifurcations

To compare method IV to previous work on topological entropy *not* based on ordinal patterns (e.g., see [13,15,18,44]), we applied it with $2 \leqslant m \leqslant 16$ to the $r \in [3.5, 4]$ family of logistic mappings $x_{n+1} = f_r(x_n) = r x_n (1 - x_n)$. Selected results for $m = 12$ are displayed on the right axis of Fig. 8), superimposed on the orbit diagram.
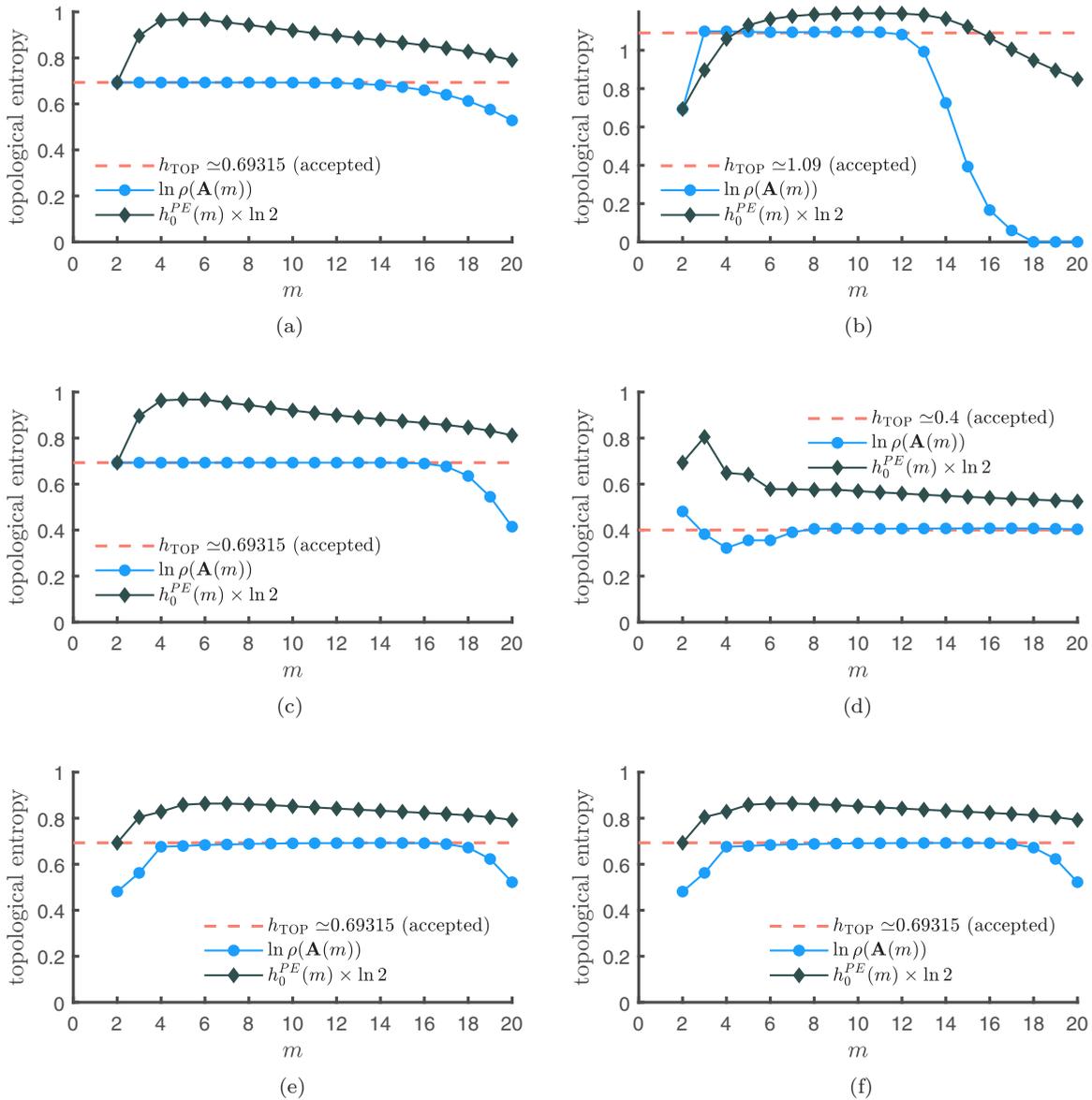
FIG. 7. Methods II and IV applied to time series obtained from six one-dimensional maps in a chaotic regime. The estimated exponential growth rate of admissible patterns for each map is $\gamma = 0.8378$ (bent baker), $\gamma = 1.2147$ (cubic), $\gamma = 0.8419$ (dyadic), $\gamma = 0.5438$ (Gauss), $\gamma = 0.8232$ (logistic), and $\gamma = 0.8232$ (tent). (a) bent baker, (b) cubic, (c) dyadic, (d) Gauss, (e) logistic, and (f) tent.

The discretization of this parameter range into 1000 equispaced points provides a high refinement (spatial step $\Delta r = 0.0005$) for investigating the efficacy of the method, especially around bifurcation points where Pommeau-Manneville intermittent regimes can occur. The curve in Fig. 8 is consistent with previous results (see Fig. 7 in [13], Fig. 1 in [15], Fig. 1 in [18], Fig. 2 in [44]) except at periodic windows.

It is known that the entropy should be zero for any $r < 3.5$ [13], however Block *et al.* [13] state that "on those intervals of μ [i.e., r here] where there are attracting periodic points for the map f_μ, the topological entropy is constant. This is known for theoretical reasons to be true." Clearly this is verified by all aforementioned techniques, whereas the ordinal spectral radius indicates that the entropy is zero for every

periodic window within $r \in [3.5, 4]$. An explanation for this behavior is given in Sec. IV. We remark that it is unclear why topological entropy should be constant rather than exactly zero within periodic regimes and no details or citations are given in [13] (moreover, none of the other references even discuss this issue). According to the definition based on ϵ-separated points [9,10], if distance between neighboring points is preserved under sequential iteration of the map, the number of distinguishable orbits remains constant and hence the topological entropy ought to be zero. Finally, we note that for this archetypal one-dimensional system, method IV proved a great predictor of regime switching between periodic and chaotic dynamics—whether due to an artifact of the technique or because entropy is indeed zero for periodic dynamics.
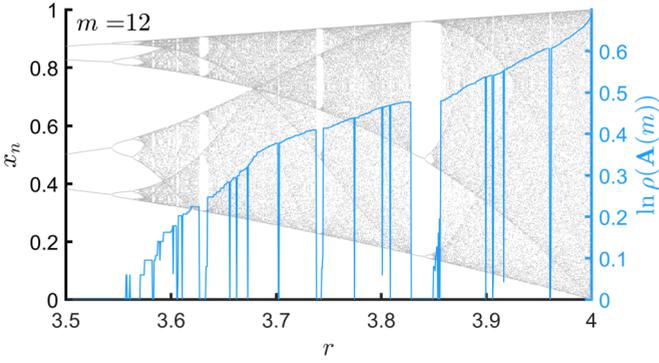
FIG. 8. Logarithm of the ordinal spectral radius applied to the logistic family of maps $f_r(x)$ for $r \in [3.5, 4]$. The results displayed correspond to the $m = 12$ ordinal partition.

D. Test IV: Two-dimensional dissipative maps

Next we consider two famous examples of dissipative dynamics in two dimensions, the Lozi and the Henon map. The topological entropy of both is known to be equal to $h_{\mathrm{TOP}}(f_{\mathrm{LOZI}}) = 0.3$ and approximately equal to $h_{\mathrm{TOP}}(f_{\mathrm{HENON}}) \simeq 0.465$, respectively.

Figure 9 shows that methods I, III, and IV (diamonds, squares, and circles, respectively) can all produce relatively accurate estimates for some values of m in the chaotic Lozi regime. For the latter method, the quality of the partition is poorer for odd values of m, especially when the partition is not very refined. For $m = 23$, eigenvalues do not even converge. The ordinal spectral radius exhibits oscillatory behavior rather than the type of smooth scaling region observed for the one-dimensional maps, which is not something that we have observed with any other test system. In fact, both method III and method IV display oscillations in this case (see, for example, Fig. 1 in Amigó and Kennel [36] as it shows this effect more clearly than Fig. 9 here due to the scale), however their amplitude is much more pronounced for the latter. Therefore, it is clear that for lower m values, odd m produce poor partitioning. The situation slightly improves as m increases.

We conjecture that false ordinal transitions appear due to the poor partitioning which may be due to the peculiarities
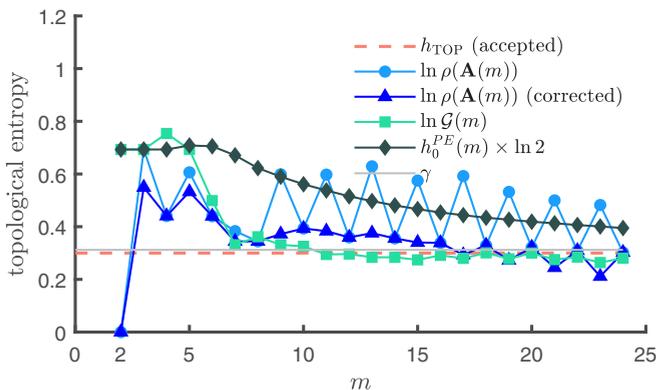
FIG. 9. Methods I–IV applied to a time series of length $N = 10^7$ points generated by the chaotic Lozi map regime $(a, b) = (6/5, -2/15)$.
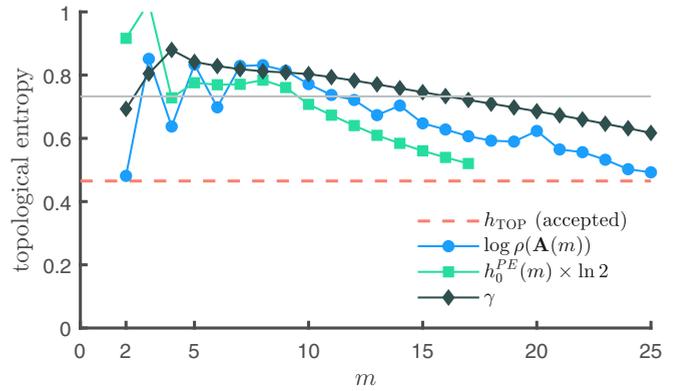
FIG. 10. Methods I, II, and IV applied to a time series of length $N = 10^7$ points generated by the chaotic Henon map regime $(a, b) = (1.4, 0.3)$.

of this specific attractor, particularly the spatial symmetries that arise in state space. In order to investigate this further, we recorded all occurring transitions and observed that certain transitions appear much more infrequently in comparison to others. By removing the particularly infrequent ones—using a threshold of relative frequency larger than 10^{-4} as a proof of concept—we recomputed the connectivity matrix. This is simply a practical heuristic; the threshold should ideally be dependent on m rather than constant. Essentially, the threshold can be set by detecting the largest gap in the relative frequency histogram of transitions. Even with our coarse heuristic, the results (dark blue triangles on Fig. 9) display significant improvement for lower values of m. Further improvement for higher m necessitates that the threshold is *not* independent of m.

Note that longer patterns (i.e., finer partitions) are required by all methods to obtain accurate estimates than for the 1D maps. The outgrowth ratio provides a smoother curve with reduced oscillation amplitude. TPE converges much more slowly, as is the case with all studied systems. Method I performs surprisingly well, probably due to the extremely thin support of the invariant measure of the Lozi attractor. Only a very small number of patterns are admissible even when m is large, therefore no undersampling occurred for the parameter range that we investigate in the Lozi regime.

The Henon trajectory on which computations are performed to generate the estimates of Fig. 10 is drawn from the original parameter regime in [45]. Method III is very expensive computationally—required time is in the order of hours and days even for $m \leqslant 20$—hence we omit estimates for $m > 12$ and do not present results for lower values. None of the methods lead to an accurate estimate in this case. Method IV seems to be the most reliable, producing values closer to the accepted one for all m (except $m = 3$). Additionally, for $m = 24, 25$ some scaling begins to appear, but more pronounced undersampling for any $m \geqslant 26$ prevents us from establishing a clear deduction. A longer time series needs to be examined to test whether the partition refinement $m > 23$ is such that the Henon dynamics can be fully encapsulated by the ordinal network.
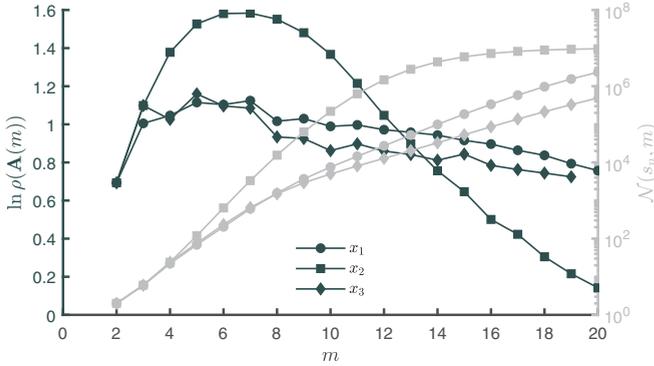
FIG. 11. Left axis: Method IV applied to three $N = 10^7$-long time series, each corresponding to a component (x_1, x_2, x_3) of the trajectory generated by the hyperchaotic regime $(a, b) = (3.8, 0.2)$ in the folded-towel map. Right axis: Network size produced for each component by the ordinal mapping.

E. Test V: Hyperchaos in a three-dimensional dissipative map

The final test to which the network-based method IV is subjected is a hyperchaotic regime, in the sense of two positive Lyapunov exponents. We use the the three-dimensional folded-towel map, originally introduced in [46] as a simpler prototypic map of a Poincaré' cross section through the hyperchaotic four-dimensional continuous-time Rössler flow. The dynamics may be understood via consideration of a rectangular cube that will evolve into a "folded-towel" shape after compression, stretching and folding over itself in two directions, then placed back onto itself during each iteration. Pairs of adjacent points diverge exponentially in both the x and the z direction, a phenomenon termed lateral instability in two directions by Rössler. Its topological entropy is unknown, but presumably it is higher than in all prior examples.

Figure 11 shows the entropy estimates obtained for each state variable (x_1, x_2, x_3) by method IV with $2 \leqslant m \leqslant 20$, depicted by circle, square, and diamond markers, respectively. This is the first example where different components of the map do not yield identical estimates; the main reason being undersampling since longer trajectories than $N = 10^7$ are required to sample the given orbit sufficiently. In fact, by checking the critical sample size beyond which the network size, i.e., $\mathcal{N}(s_n, m)$, saturates as N increases (results omitted), we can confirm that at least some undersampling effects are present for any ordinal partition with $m \geqslant 4$. This is evident by observing the right axis of Fig. 11 that displays the evolution of the network size as the pattern length increases. Networks generated by the three components diverge in size if $m > 5$. The higher complexity of this map in comparison to all one- and two-dimensional examples presented herein implies that finer partitions are necessary to capture the true dynamics. However, due to the trade-off between refinement and sample size, $N = 10^7$ is not adequate for most values in the interval $m \in [2, 20]$.

Estimates produced by observables x_1 and x_3 do not reveal any faithful scaling with m, except for when $3 \leqslant m \leqslant 7$. The average values within this range are, respectively, $\ln \rho_{\mathrm{AVG}}[A(m; x_1)] \simeq 1.079$ and $\ln \rho_{\mathrm{AVG}}[A(m; x_3)] \simeq 1.093$. Undersampling is less substantial for these two ob-

servables than for x_2, with the network size coinciding for all $m \leqslant 8$.

It is particularly interesting to examine the x_2 component as there is no exponential divergence between neighboring points in this direction (contrary to the focal point if looking for Lyapunov exponents). In addition, the nature of this attractor dictates that transitions between the two "towel branches" correlate with the x_2-component series. The growth of admissible patterns displays a clear exponential growth in the interval $4 \leqslant m \leqslant 11$ for this component, at a significantly higher rate in comparison to x_1, x_3. Undersampling sets in more rapidly in this case—in fact for any $m > 11$ the network is close to trivial—hence we deem only partitions with $m \leqslant 11$ to be reliable. With $5 \leqslant m \leqslant 8$, x_2 estimates manifest some scaling $(\ln \rho_{\mathrm{AVG}}[A(m; x_2)] \simeq 1.561)$, however it the scaling region is rather short and it is hard to reach conclusions.

Ultimately, if undersampling is the pertinent factor, estimates obtained from x_1 and x_3 are more reliable. Furthermore, the entropy estimates decay more slowly as m increases than for x_3, another indication of increased robustness of results. For example, computing the average estimate in the parameter region $3 \leqslant m \leqslant 11$ obtained using x_1 yields $\ln \rho_{\mathrm{AVG}}[A(m; x_2)] \simeq 1.047$, a value quite close to the one calculated for $3 \leqslant m \leqslant 7$ whereby the scaling is very clear. If, on the other hand, the complexity factor is more crucial, for small pattern lengths (i.e., coarse partitions), x_2 may be a preferable observable as its symbolic dynamics may enable superior estimates due to its revealing a "wilder" higher-mixing evolution closer to the true underlying dynamics.

Whether the x_1, x_3 series yield more reliable estimates or the x_2 series is preferable, all estimates produce higher values than in the case of all lower-dimensional (and lower-complexity) maps, which is in agreement with our expectations. Additionally, all estimates are higher than the approximate metric entropy of the map equal to 0.807 (positive Lyapunov exponents $\lambda_1 \simeq 0.430$, $\lambda_2 \simeq 0.377$). This is consistent with the theoretical result that topological entropy is the supremum of metric entropy attained when a partition is generating.

Finally, we remark that another source of inaccuracy may be due to the map not necessarily being ergodic. Judging by the gamut of systems examined for the purposes of this study, we note that apart from high complexity or attractors occupying a large volume—both of which lead to more stringent requirements of partition refinement and sample size—lack of ergodicity also contributes to the unfaithfulness of the ordinal representation. A suitable example that can convey this point is the Ikeda map under the most studied chaotic regime ($\mu = 0.9$). It is known to contain homoclinic orbits [47] that prevent the system from being fully mixing. The ordinal statistics collected from a time series (of length $N \sim 10^6$) generated by this dynamical regime led to less faithful representations. This is evidenced by the lack of irreducibility of ordinal networks computed from the Ikeda trajectory with $m > 12$ (results omitted). In Sakellariou et al. [23] the authors prescribed a Markovian framework for ordinal networks under the assumption of ergodicity in the underlying dynamics. None of the ordinal methods that we examined is expected to be applicable to such systems since several building arguments no longer hold> true.

IV. SPECTRAL RADIUS OF CONNECTIVITY MATRIX

Why is the network-based method IV more effective in estimating topological entropy as compared to the other ordinal analysis techniques? The main difference lies on the Markovian description of the dynamics. It encapsulates information about the transitions between patterns within the symbolic itinerary.

Markov processes date back to the first half of the 20th century. Since the early days accompanying methods that utilize the transition graph as a natural representation were developed. In the context of finite-state languages, Chomsky and Miller [20] established a method for estimating lengths of words composed of binary symbols. Since then, the logarithm of the largest eigenvalue of the connectivity graph associated with a specified Markov model has been used to estimate the topological entropy of dynamics on cellular automata [48] and chaotic unimodal self-interval maps [49]. Block *et al.* [13], Góra and Boyarsky [15], and Balmforth *et al.* [16] developed improved algorithms based on it for more general one-dimensional settings (e.g., nonunimodal or discontinuous maps), focusing on convergence results and computational efficiency. Froyland *et al.* [18] proposed an analytically rigorous and sophisticated approach based on right-resolving representations and sofic shifts.

We have essentially adapted the idea to the setting of an ordinal partition. This allowed us to (a) place ordinal representations under a more stringent quantitative test of faithfulness to the underlying dynamics and (b) investigate whether building a complex network produces superior approximations than simply counting patterns. As demonstrated, our findings indicate significantly improved performance.

To provide some intuition with respect to the effectiveness of this approach, consider the space of infinite sequences of natural numbers up to $\mathcal{J} \in \mathbb{N}$, namely $\{1, 2, \ldots, \mathcal{J}\}^{\mathbb{N}}$. Denote the set

$$s = \{(s_0 s_1 s_2 s_3 \cdots) \mid s_j \in \mathbb{N}, \ 1 \leqslant s_j \leqslant \mathcal{J}\}, \quad (12)$$

by $\Sigma_{\mathcal{J}}$. In the case of ordinal symbolization, the permutation indices can be mapped bijectively to the natural numbers up to $\mathcal{J} = m!$ for pattern length $m \in \mathbb{N}$, $m \geqslant 2$. For instance, when $m = 2$, Σ_2 is the binary sequence space examined by [20].

A natural metric on $\Sigma_{\mathcal{J}}$ is defined by

$$d_{\mathcal{J}}(s, t) = \sum_{i=0}^{\infty} \frac{\delta(s_i, t_i)}{\mathcal{J}^i},$$

$$\delta(s_i, t_i) = \begin{cases} 0 & \text{if } s_i = t_i, \\ 1 & \text{if } s_i \neq t_i \end{cases} \quad (13)$$

since $d_{\mathcal{J}} \geqslant 0$, $d_{\mathcal{J}}(s, t) = d_{\mathcal{J}}(t, s)$ (symmetry) and $d_{\mathcal{J}}(s, t) \leqslant d_{\mathcal{J}}(s, r) + d_{\mathcal{J}}(r, t)$ (triangle inequality). Furthermore, $d_{\mathcal{J}}(s, t) = 0 \Leftrightarrow s = t$. Note that

$$d_{\mathcal{J}}(s, t) < \frac{1}{\mathcal{J}^k} \Rightarrow s_i = t_i \quad \text{for } 0 \leqslant i \leqslant k \quad (14)$$

and

$$s_i = t_i \text{ for } 0 \leqslant i \leqslant k \Rightarrow d_{\mathcal{J}}(s, t) \leqslant \frac{1}{\mathcal{J}^k}. \quad (15)$$

Additionally, $d_{\mathcal{J}}(s, t)$ is bounded by the geometric sequence

$$d_{\mathcal{J}}(s, t) \leqslant \sum_{i=0}^{\infty} \frac{1}{\mathcal{J}^i} = \frac{1}{1 - \frac{1}{\mathcal{J}}} = \frac{\mathcal{J}}{\mathcal{J} - 1}, \quad (16)$$

with the upper bound attained when $s_i \neq t_i \ \forall i \in \mathbb{N}$. The (one-sided) 1-shift map $\sigma : \Sigma_{\mathcal{J}} \to \Sigma_{\mathcal{J}}$ given by

$$\sigma(s) = \sigma(s_0 s_1 s_2 s_3 \cdots) = s_1 s_2 s_3 s_4 \cdots \quad (17)$$

is continuous in the metric $d_{\mathcal{J}}$.

A *connectivity* matrix of size $\mathcal{N} \times \mathcal{N}$, say $A = (a_{ij})$ where $\mathcal{N} \leqslant \mathcal{J}$ and $a_{ij} \in \{0, 1\}$, describes a subset of $\Sigma_{\mathcal{J}}$, denoted by Σ_A, whereby member sequences obey the following rule:

$$\Sigma_A = \{s = (s_0 s_1 s_2 s_3 \cdots) \mid a_{s_i s_{i+1}} = 1, \ \forall i \in \mathbb{N}\}. \quad (18)$$

As an example consider the simple connection matrix of a 3-node network

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}. \quad (19)$$

Since it contains no nonzero diagonal elements, there are no links connecting a node to itself. Therefore, sequences of the constant form $(1111 \ldots)$, $(2222 \ldots)$, or $(3333 \ldots)$ are forbidden. The only possible links between successive elements of sequences in Σ_A are $1 \to 2$, $2 \to 1$, $2 \to 3$, and $3 \to 1$. Hence, member sequences include the periodic itineraries $(121212 \ldots)$, $(123123 \ldots)$, $(1212312123 \ldots)$, etc., as well as aperiodic trajectories, e.g., $(12312121233121121212123 \ldots)$ which in this case is composed of several periodic 2 and 3 blocks.

The *restriction* of σ to the set Σ_A is denoted by σ_A. This makes sense as Σ_A is a closed subset of $\Sigma_{\mathcal{J}}$. Furthermore, Σ_A is invariant under the action of σ_A since the σ image of any infinite sequence $s \in \Sigma_A$ necessarily also comprises an element of Σ_A due to the constraints imposed by Eq. (18). In traditional terminology, σ_A is a *subshift of finite type*. Therefore, ordinal networks constitute a subshift of finite type on the space of infinite sequences *composed of indices associated with each element of the ordinal partition.*

For instance, consider the restriction of the shift map σ in the case of the chaotic ($r = 4$) logistic trajectory and a 4-periodic trajectory ($r = 3.5$). We construct networks with a pattern length equal to $m = 3$ whereby the ordinal partition consists of $3! = 6$ elements. Figure 2 illustrates the topology of these networks; node labels refer to the specific pattern or permutation.

The symbolic itinerary in the 4-periodic case only visits four out of the six partition elements since patterns (1,2,3) and (3,2,1) are forbidden. The Markovian connectivity matrix for this regime is given by

$$A_{r=3.5} = \begin{matrix} & \begin{matrix} (1,3,2) & (2,1,3) & (2,3,1) & (3,1,2) \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} & \begin{matrix} (1,3,2) \\ (2,1,3) \\ (2,3,1) \\ (3,1,2) \end{matrix} \end{matrix}. \quad (20)$$

Only four transitions are allowed, namely $1 \to 2$, $2 \to 4$, $3 \to 1$, and $4 \to 3$ (using the index notation instead of the

permutation or pattern), and no element is connected to itself. Consequently, Σ_A constitutes of periodic itineraries of the specific form $(12431243\ldots)$ only; no other infinite sequence is possible, be it constant, e.g., $(1111\ldots)$, periodic of a different form, e.g., $(12341234\ldots)$, or aperiodic. The eigenvalues of $\boldsymbol{A}_{r=3.5}$ are $1, -1, i, -i$, hence

$$\ln\rho(\boldsymbol{A}_{r=3.5}(m)) = \ln 1 = 0. \tag{21}$$

Refining the partition by counting increasingly longer m patterns yields the exact same connectivity matrix for this regime. In fact, this invariance holds true for any 2^k-periodic regime, provided that $m \geqslant 2^{k-1}$ (and also that $\tau = 1, w = 1$; see Sec. V A in [23] for details). The number of distinguishable orbits prescribed by $\boldsymbol{A}_{r=3.5}$ remains constant as m increases.

The chaotic itinerary, on the other hand, visits five regions [only pattern $(3,2,1)$ is forbidden]. In addition, nine transitions are possible, specifically $1 \to 1, 1 \to 2, 1 \to 5, 2 \to 3, 3 \to 5, 4 \to 1, 4 \to 2, 5 \to 3$, and $5 \to 4$. The connectivity matrix is

$$\boldsymbol{A}_{r=4} = \begin{array}{c} \\ \end{array}\begin{matrix}(1,2,3)\ (1,3,2)\ (2,1,3)\ (2,3,1)\ (3,1,2)\\ \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}\begin{matrix}(1,2,3)\\(1,3,2)\\(2,1,3)\\(2,3,1)\\(3,1,2)\end{matrix}\end{matrix}. \tag{22}$$

As a result of the larger number of allowed transitions prescribed by $\boldsymbol{A}_{r=4}$, Σ_A comprises a much richer set of member sequences. Examples include the constant sequence of the form $(1111\ldots)$, 2-periodic itineraries, e.g., $(3535\ldots)$, 3-periodic itineraries, e.g., $(154154\ldots)$, 4-periodic itineraries, e.g., $(23542354\ldots)$, 5-periodic itineraries, e.g., $(1235412354\ldots)$, etc. as well as aperiodic trajectories, e.g., $(12353542351115411111123\ldots)$. Three eigenvalues of $\boldsymbol{A}_{r=4}$ are real, namely $1.7549, -1, 0$, while the complex conjugate pair $0.1226 \pm 0.7449i$ has a nonzero imaginary part and a modulus approximately equal to 0.7549. Hence we have

$$\ln\rho[\boldsymbol{A}_{r=4}(3)] = \ln 1.7549 = 0.5624 > 0. \tag{23}$$

The approximation is poor in this case; it generates an underestimate of the true entropy. The pattern length $m = 3$ does not induce a sufficiently fine partition to fully capture the complexity of this expansive regime. However, partition refinement by merely increasing the pattern length to one additional point ($m = 4$), leads to a significantly more accurate value ($\ln\rho[\boldsymbol{A}_{r=4}(4)] \simeq 0.676$) as Fig. 3 shows.

In general, by the Perron-Frobenius theorem for nonnegative matrices, there exists a non-negative eigenvalue $\lambda_{\max}$ such that no other eigenvalue of A has absolute value greater than it. Corresponding to $\lambda_{\max}$ there exist unique non-negative left and right eigenvectors. Moreover, if A is irreducible and aperiodic, $\lambda_{\max}$ is a simple eigenvalue and the eigenvectors are strictly positive. Note that the set Σ_A consists of symbolic itineraries which are the manifestation of all possible random walks from any starting node on A. Consequently, a leading eigenvector corresponds intuitively to a set of regions in state space which remain invariant (up to scaling) under a single update of the evolution rule and, furthermore, endure the

maximum possible scaling (in absolute value), i.e., regions where mass tends to concentrate the most after an update of the dynamics. Topological entropy is then related to the growth rate of the number of possible paths from all starting nodes, which in the infinite limit tends to grow as the natural logarithm of $\lambda_{\max}$, i.e., the maximum scaling factor.

V. DISCUSSION

Amigó and Kennel [36] argue that attempting to estimate topological entropy from a single long trajectory is very difficult. Metric entropy is also not captured easily due to the requirement for very long input time series that enables faithful representations by observing longer patterns, as reflected in undersampling heuristics such as $N \gg m! + m - 1$ proposed in the literature [42]. Quoting [36], "... but topological entropy is worse yet, because it weights each pattern equally. This means that patterns which are exceptionally infrequent on the natural measure of the attractor can still have a significant influence on the result. Attempting to estimate the same quantities using empirical occurrences of order patterns is even more difficult, requiring more data than would a good, low-alphabet generating partition for ordinary symbolic dynamics." These invariant estimation challenges associated with all other symbol-only methods seem to be somewhat addressed by employing the network and matrix characterization. The reason is that the largest eigenvalue of the ordinal Markov discrete operator—that acts as an approximation to the evolution operator of the underlying system—measures the accumulated effect of matrix multiplication. Vectors on eigendirections—or equivalently "mass" distributions in the symbol space—are preserved up to a scaling factor by the action of the ordinal Markov operator. This scaling factor lowers the effect of infrequent admissible patterns on ordinal estimates mentioned in [36].

The five tests that we conducted on the ensemble of systems and dynamical regimes examined herein suggest that ordinal partition networks can prove of great practical use. The ability to approximate accurately an invariant such as topological entropy by merely observing rather short patterns (low finite m, far from the thermodynamic limit $m \to \infty$) on scalar observables suggests that the proposed method is effective. Furthermore, these results shed light on the faithfulness of ordinal finite-memory Markov processes as a coarse approximation to chaotic dynamics.

Our study will hopefully spark some interest for further theoretical and numerical investigations of the properties of an ordinal partition in chaotic maps. In particular, determining the maximum finite value of m for which faithful computations are feasible, examining how the proposed method performs if the input time series is generated by a map of infinite entropy, establishing convergence rates for m in relation to dynamical invariants or known features of a specified system, e.g., dimensionality, volume preservation, ergodicity, etc. This shall determine more clearly the type of data for which this methodology can be applicable in practice. McCullough *et al.* [41] showed that heart variability state distinction as well as characterisation of complexity of cardiac dynamics may be pursuits amenable to ordinal analysis. Moreover, Downarowicz [50] developed a rich theory on the concept of *entropy*

structure, a kind of master invariant for the entropy theory of topological dynamical systems, in order to shed light on the complex interplay between entropy and scale. It would be interesting to determine whether ordinal partitioning along with successive refinement can potentially extract information about the entropy structure.

[1] R. L. Adler, A. G. Konheim, and M. H. McAndrew, Trans. Amer. Math. Soc. **114**, 309 (1965).

[2] A. N. Kolmogorov and V. M. Tihomirov, Am. Math. Soc. Transl. Ser. 2 **17**, 277 (1961) [Uspekhi Mat. Nauk **14**, 3 (1959)].

[3] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **119**, 861 (1958).

[4] A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **124**, 754 (1959).

[5] Y. G. Sinai, Dokl. Akad. Nauk SSSR **124**, 768 (1959).

[6] L. W. Goodwyn, Proc. Am. Math. Soc. **23**, 679 (1969).

[7] T. N. T. Goodman, Bull. London Math. Soc. **3**, 176 (1971).

[8] E. I. Dinaburg, Dokl. Acad. Nauk **190**, 19 (1970).

[9] E. I. Dinaburg, Izv. Akad. Nauk SSSR **5**, 337 (1971).

[10] R. Bowen, Trans. Am. Math. Soc. **153**, 401 (1971).

[11] R. Bowen, Trans. Am. Math. Soc. **154**, 377 (1971).

[12] P. Collet, J. P. Crutchfield, and J. P. Eckmann, Commun. Math. Phys. **88**, 257 (1983).

[13] L. Block, J. Keesling, S. Li, and K. Peterson, J. Stat. Phys. **55**, 929 (1989).

[14] Y. Takahashi, Sci. Pap. Coll. Gen. Educ. **30**, 11 (1980).

[15] P. Góra and A. Boyarsky, Trans. Am. Math. Soc. **323**, 39 (1991).

[16] N. J. Balmforth, E. A. Spiegel, and C. Tresser, Phys. Rev. Lett. **72**, 80 (1994).

[17] G. Froyland, R. Murray, and D. Terhesiu, Phys. Rev. E **76**, 036702 (2007).

[18] G. Froyland, O. Junge, and G. Ochs, Physica D **154**, 68 (2001).

[19] S. Newhouse and T. Pignataro, J. Stat. Phys. **72**, 1331 (1993).

[20] N. Chomsky and G. A. Miller, Inform. Control **1**, 91 (1958).

[21] C. E. Shannon, Bell Syst. Tech. J. **27**, 623 (1948).

[22] C. Bandt, G. Keller, and B. Pompe, Nonlinearity **15**, 1595 (2002).

[23] K. Sakellariou, T. Stemler, and M. Small, Phys. Rev. E **100**, 062307 (2019).

[24] C. Bandt and B. Pompe, Phys. Rev. Lett. **88**, 174102 (2002).

[25] K. Keller and H. Lauffer, Int. J. Bifurcation Chaos **13**, 2657 (2003).

[26] Y. Cao, W. W. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, Phys. Rev. E **70**, 046217 (2004).

[27] G. Ouyang, C. Dang, D. A. Richards, and X. Li, Clin. Neurophysiol. **121**, 694 (2010).

[28] K. Schindler, H. Gast, L. Stieglitz, A. Stibal, M. Hauf, R. Wiest, L. Mariani, and C. Rummel, Epilepsia **52**, 1771 (2011).

[29] C. Rummel, E. Abela, M. Hauf, R. Wiest, and K. Schindler, Eur. Phys. J. Special Topics **222**, 569 (2013).

[30] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, Comput. Biol. Med. **42**, 319 (2012).

[31] D. Cysarz, A. Porta, N. Montano, P. V. Leeuwen, J. Kurths, and N. Wessel, Eur. Phys. J. Special Topics **222**, 487 (2013).

[32] G. Graff, B. Graff, A. Kaczkowska, D. Makowiec, J. M. Amigó, J. Piskorski, K. Narkiewicz, and P. Guzik, Eur. Phys. J. Special Topics **222**, 525 (2013).

[33] K. Keller and M. Sinn, Nonlinearity **22**, 2417 (2009).

[34] K. Keller, T. Mangold, I. Stolz, and J. Werner, Entropy **19**, 134 (2017).

[35] J. M. Amigó, M. B. Kennel, and L. Kocarev, Physica D **210**, 77 (2005).

[36] J. M. Amigó and M. B. Kennel, Physica D **231**, 137 (2007).

[37] G. Nicolis and C. Nicolis, Phys. Rev. A **38**, 427 (1988).

[38] R. Bowen, Am. J. Math. **92**, 725 (1970).

[39] S. Ulam, Fund. Math. **16**, 140 (1930).

[40] G. Nicolis and P. Gaspard, Chaos Soliton. Fract. **4**, 41 (1994).

[41] M. McCullough, M. Small, H. H. C. Iu, and T. Stemler, Philos. Trans. R. Soc. London Sect. A **375**, 20160292 (2017).

[42] J. M. Amigó, S. Zambrano, and M. A. F. Sanjuán, Europhys. Lett. **79**, 50001 (2007).

[43] K. Butt, Ph.D. thesis, University of Chicago, 2014.

[44] S. J. Canovas, Entropy **15**, 3100 (2013).

[45] M. Hénon, Commun. Math. Phys. **50**, 69 (1976).

[46] O. Rössler, Phys. Lett. A **71**, 155 (1979).

[47] K. Ikeda, Opt. Commun. **30**, 257 (1979).

[48] S. Wolfram, Commun. Math. Phys. **96**, 15 (1984).

[49] S. Isola and A. Politi, J. Stat. Phys. **61**, 263 (1990).

[50] T. Downarowicz, J. D'Analyse Math. **96**, 57 (2005).