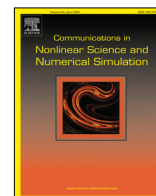




Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns

Research paper

Network inference combining mutual information rate and statistical tests

Chris G. Antonopoulos

Department of Mathematical Sciences, University of Essex, Wivenhoe Park, UK



ARTICLE INFO

Article history:

Received 3 January 2022

Received in revised form 8 September 2022

Accepted 15 September 2022

Available online 21 September 2022

Keywords:

Network inference

Mutual information rate

Mutual information

Statistical tests

False discovery rate

Shannon entropy

Complex systems

Complex networks

Time-series data

ABSTRACT

In this paper, we present a method that combines information-theoretical and statistical approaches to infer connectivity in complex networks using time-series data. The method is based on estimations of the Mutual Information Rate for pairs of time-series and on statistical significance tests for connectivity acceptance using the false discovery rate method for multiple hypothesis testing. We provide the mathematical background on Mutual Information Rate, discuss the statistical significance tests and the false discovery rate. Further on, we present results for correlated normal-variables data, coupled circle and coupled logistic maps, coupled Lorenz systems and coupled stochastic Kuramoto phase oscillators. Following up, we study the effect of noise on the presented methodology in networks of coupled stochastic Kuramoto phase oscillators and of coupling heterogeneity degree on networks of coupled circle maps. We show that the method can infer the correct number and pairs of connected nodes, by means of receiver operating characteristic curves. In the more realistic case of stochastic data, we demonstrate its ability to infer the structure of the initial connectivity matrices. The method is also shown to recover the initial connectivity matrices for dynamics on the nodes of Erdős-Rényi and small-world networks with varying coupling heterogeneity in their connections. The highlight of the proposed methodology is its ability to infer the underlying network connectivity based solely on the recorded datasets.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Complex systems are ubiquitous in nature and are the subject of intense research over the last decades [1,2], which has led to the emergence of the new paradigm of complexity or “complex systems” science [2]. A complex system consists of many units which may interact in non-trivial ways, and whose aggregate behaviour is undetermined from the behaviour of the individual units [2,3]. Prominent examples are organisms, the human brain, power grids, transportation and communication systems, ecological systems, etc., [2]. What is common in these systems is that their behaviour is intrinsically difficult to model due to dependencies, competitions, relationships, or other types of interactions among their units or between the system and its environment. What is fascinating is that they exhibit distinct properties that arise from these relationships, such as nonlinearity, emergence, self-organisation, spontaneous order, adaptation, feedback loops, etc. Due to their appearance in a variety of fields and disciplines, common approaches to study them have become prominent and topics of their own importance [2]. Interestingly, one can represent complex systems as a mathematical model of a graph that consists of nodes that represent the units in the system and edges that represent interactions among the nodes, a structure that is commonly referred to as a network.

E-mail address: canton@essex.ac.uk.

Considering these units as nodes in the network, and their underlying physical interactions as links, a step forward in understanding complex systems is to infer the topological structure of their network representation, i.e., their connectivity. This is driven by the fact that in most cases, the connectivity is not known, and that one might only have access to data recorded on the nodes of the network, for example in experiments. Usually, the connectivity of the units is not known or is difficult to detect by physical methods due to large system-sizes. Consequently, it becomes clear that if one wants to study complex systems when only recorded data are available, one first has to infer the underlying structure, namely to establish whether there are undirected links among the pairs of nodes in the network.

These hurdles highlight the importance of the development of mathematical approaches to infer network connectivity using recorded data, i.e., methodologies that can reveal interconnectivity among nodes in a network. Depending on the research field, some of these approaches might be more applicable than others. For example, in [4], the authors developed HERMES, a Matlab software package with a collection of methods to infer functional and effective brain connectivity from neurophysiological data such as from multivariate electroencephalography (EEG) and magnetoencephalography (MEG) recordings. In [5], a review of methods and assessment of tools and techniques is discussed for biological network inference and in [6], the author provides a review of advanced techniques for the estimation of brain connectivity measured with EEG and MEG. The authors in [7] compared connectivity analyses for resting state EEG data pointing to the advantages of nonlinear methods, and indicating a relationship between the flow of information and the level of synchronisation in the brain. In [8], it is discussed how to improve accuracy of complex network modelling using maximum likelihood estimation and expectation-maximisation.

Network inference in nonlinear systems, physics, engineering, neuroscience, biology, etc. has been extensively studied in recent years using, for example, cross-correlation (CC) or mutual information (MI) [9–11], recurrences [12–14], functional, response and complex dynamics [15–18], Granger and multivariate Granger causality [19–22], rank-based connectivity measures [23], rank statistics [24], mutual prediction [25], phase transfer entropy [26], etc. In [27], the authors present a method that can reconstruct directed and weighted topologies of networks of heterogeneous nonlinear oscillators. In [28], a novel approach for the recovery of the directional connectivity of a small oscillator network by means of the phase dynamics reconstruction from multivariate time-series data is presented, where the authors use a triplet analysis instead of the traditional pairwise. In [29], the authors present a general framework for dynamic complex networks applicable for characterising dynamic complex networks. They discuss that whether a dynamic complex network evolves toward synchrony or deterioration and collapse can be determined by tracking ensemble entropy in time. They found that, in addition to misrepresenting the true network dynamics, static network structures do not differentiate in resolving network properties such as average path-lengths and degree distributions. Even though the progress made, network inference still presents open challenges. For one, any similarity measure usually results in a non-zero value [13,30] as, in finite datasets, persistent trends or deterministic recurrent oscillations result in spurious correlations [31,32]. Also, nonlinearities and asymmetries in the structure or noise level of the dynamics can affect these methods [33–36].

In this paper, we are using an information-theoretical methodology and statistical significance tests complemented by the false discovery rate (FDR) method for multiple hypothesis testing to infer connectivity in complex networks using only time-series data [37,38]. The data are recorded on the nodes of the network. The equations of motions and initial connectivity matrices are only used to compute and record the solutions, and to compare the inferred with the initial connectivity matrices by means of receiver operating characteristic (ROC) curves for different parameter values. This allows to quantify how successful the network inference is. Our methodology is based on: (a) estimations of the Mutual Information Rate (MIR) for pairs of time-series (that represent pairs of nodes in the network) [37,38] and (b) on statistical significance tests to accept or reject connectivity using the false discovery rate method for multiple hypothesis testing.

MIR is the rate of information exchanged per unit of time between pairs of time-series [37–39] and is an appropriate measure to quantify the exchange of information in systems with correlation [33,39,40]. A normalised version of MIR has been shown to perform well in inferring the structure of networks in different cases of dynamics [37] and, in financial and stock markets data [38]. The authors in [37] discuss how to calculate MIR for Markov partitions. Even though it is a very interesting approach, these partitions are in general difficult or impossible to find. The methodology presented here is applicable in the case where only recorded data are available. In [9], two inference methods based on CC and MI are compared and it was shown that, for the class of studied systems, when an abrupt change in the ordered CC or MI values exists, it is possible to infer, without errors, the underlying network topology from the time-series data, even in the presence of observational noise, non-identical units, and coupling heterogeneity. However, this approach might not be able to infer the initial topology in situations where the gaps in the ordered CC or MI values are absent or when there are more than one gaps as it happens for the financial and stock market data in [38]. In these cases, a (unique) threshold cannot be chosen to infer the network structure [9]. The approach presented here does not make use of thresholds: an inferred network will be computed based on the recorded data using pairwise computed MIR values and surrogate data to test the null hypotheses for connectivity using the false discovery rate method for multiple hypothesis testing.

In [37], the authors used the bin method and computed the probabilities for the estimation of MIR in partitions of e.g., equally sized cells on the probability space generated by a pair of variables or units X and Y . Here, we use the same approach. This method leads to the overestimation of MIR for random systems or non-Markovian partitions [30,41]. This is also the case here and is due to the finite resolution of non-Markovian partitions and to the finite length of the recorded time-series. According to [30,41], these errors are systematic and always present in the computation of MIR for arbitrary

non-Markovian partitions. In [37], these systematic errors were mitigated by double-normalising MIR, first over the MIR values of the pairs of variables X and Y for a particular grid size, and then, over the number of grid sizes on the probability space. Here, we show that this double-normalisation is not necessary and that even though, the resulting MIR values are overestimated, they are enough to allow for network inference when combined with the use of surrogate data to test the null hypotheses for connectivity and the false discovery rate method for multiple hypothesis testing.

It is worth it to note that in this work, we are interested in identifying undirected connections among nodes as MIR is a symmetric quantity with respect to X and Y , i.e., $\text{MIR}_{XY} = \text{MIR}_{YX}$. Thus, we assume throughout the paper only undirected connections, and that such connections are due to linear or nonlinear, dynamical interactions. This is tested by using surrogate data to test the null hypotheses for connectivity and the false discovery rate method for multiple hypothesis testing. Our results show that this combination of methods can be effective in inferring network connectivity.

The paper is organised as follows: In Section 2, we provide the mathematical background on information and network connectivity, on MI and MIR, on the estimation of the correlation-decay time and on network inference based on MIR and statistical significance tests using the false discovery rate for multiple hypothesis testing. In Section 3, we report on the results for different types of processes and, for discrete and continuous dynamical systems, such as correlated normal-variates, coupled circle maps, coupled logistic maps, coupled Lorenz systems and coupled stochastic Kuramoto phase oscillators. We also study the effect of noise on the proposed methodology in networks of coupled stochastic Kuramoto phase oscillators and of the effect of coupling heterogeneity degree on the method in networks of coupled circle maps. In Section 4, we discuss our work in light of its advantages and disadvantages with respect to other approaches in the literature. Finally, in Section 5, we conclude our work and discuss possible extensions that would be worth studying further.

2. Theoretical background

2.1. Information and network connectivity

Complex systems are networks made of a number of units that interact with each other, typically in nonlinear ways [2], without excluding the possibility of highly ordered processes that follow simple rules.[42]. These systems may arise and evolve through self-organisation, such that they are neither completely regular nor completely random, permitting the development of emergent behaviour at macroscopic scales [2]. In this context, a system, composed of units, can produce information that can be transferred among them through its connectivity [33,39,43–47], involving at least two interacting units. In general, these interactions may be direct or indirect and linear or nonlinear [2,42].

Particularly, the units of the system are coupled and are represented as nodes in a network. Each of these units, labelled by $i = 1, \dots, M$, can “observe” the behaviour of some neighbours $j = 1, \dots, J$ by a measurement function $h(X_j)$, where X_j is the state of unit j . For example, each unit i could be a Rössler oscillator, and the measurement function could be $h(X_j) = y_j$, meaning that the Rössler oscillators are coupled by their y_j units. Here, we assume that the units are represented by recorded time-series, i.e., by the evolution of y_j in time.

In what follows, we use MIR to estimate the amount of information exchanged per unit of time between pairs of units in a system, namely to determine whether the units are connected bidirectionally. In this context, bidirectional connectivity means an undirected connection among the units attributed to their interactions, either linear or nonlinear. Particularly, MIR measures the amount of information exchanged per unit of time between two units. Its application to time-series data is of primordial importance and will be used to determine if an undirected connection between pairs of units exists, and thus pairs of nodes in the network. In this framework, the strengths of the connections, given by their MIR values, will be estimated and compared with the MIR values of all other pairs of units in the network by means of statistical significance tests and the false discovery rate method for multiple hypothesis testing. The idea is to find out whether the connections among units, and thus among nodes exist due to their dynamics.

2.2. Mutual information

MI and MIR were originally introduced by Shannon in his seminal paper in 1948 [48]. In particular, the MI of two random variables, X and Y , is a measure of their mutual dependence and quantifies the “amount of information” obtained about X after observing Y or vice versa. Formally, it is defined by [48,49]

$$I_{XY}(N) = H_X + H_Y - H_{XY}, \quad (1)$$

where N is the number of random events in X and Y . H_X and H_Y are the marginal entropies of X and Y , respectively, i.e., their Shannon entropies, with H_X given by

$$H_X = - \sum_{i=1}^N P_X(i) \log P_X(i),$$

where $P_X(i)$ is the probability of a random event i to occur in X (similarly for H_Y). The joint entropy, H_{XY} , in Eq. (1) measures the amount of uncertainty in X and Y when taken together and is defined by

$$H_{XY} = - \sum_{i=1}^N \sum_{j=1}^N P_{XY}(i, j) \log P_{XY}(i, j), \quad (2)$$

where $P_{XY}(i, j)$ is the joint probability of events i and j to occur simultaneously in X and Y , respectively.

Equivalently, one can define MI as

$$I_{XY}(N) = \sum_{i=1}^N \sum_{j=1}^N P_{XY}(i, j) \log \left(\frac{P_{XY}(i, j)}{P_X(i)P_Y(i)} \right). \quad (3)$$

This equation provides a measure of the strength of the dependence between X and Y , i.e., the amount of information X contains about Y and vice versa [49]. When MI is zero, i.e., $I_{XY} = 0$, the strength of the dependence is null, and thus X and Y are independent variables. This means knowing X does not provide any information about Y and vice versa. MI is a symmetric quantity as $I_{XY} = I_{YX}$ and thus, not suited to study causal effects between X and Y .

The computation of $I_{XY}(N)$ requires the calculation of probabilities on a suitably defined 2D probability space generated by X and Y on which a partition based on the N^2 events can be defined (see for example Eqs. (2) and (3)). The probabilities are defined in terms of the frequency of occurrence of events over all events on the 2D probability space and thus, what will be considered as an event is crucial for the definition of the probability space and its partitioning. I_{XY} can be computed for any pair of nodes, X and Y , in the same network and can be compared with I_{XY} of other pairs in the same network. However, I_{XY} is not suitable to compare data from different systems as they may exhibit different correlation-decay times and time scales [50–52].

There are many methods to compute MI, and all depend on the approach to calculate the probabilities in Eq. (3). These are the bin method [53], the density-kernel method [54] and the method of the estimation of probabilities from distances between closest neighbours [55]. Here, we use the bin method, and particularly, grids of N^2 equally sized cells [37]. This method tends to overestimate MI because of the finite length of recorded time-series data, and the finite resolution of non-Markovian partitions [30,56]. Even though these errors are systematic and always present for any given non-Markovian partition, we will show in Section 3 that when it comes to network inference based on MIR, they still allow for correct network inference when combined with statistical significance tests and the false discovery rate method for multiple hypothesis testing.

2.3. Mutual information rate

MIR provides a method that bypasses problems associated with the resolution of non-Markovian partitions. In [37], the authors have shown how to calculate MIR for two, finite-length time-series, independently of the partitions on the probability space. For infinitely long time-series, MIR is theoretically defined as the MI exchanged per unit of time between the random variables X and Y [48,57,58]

$$\begin{aligned} \text{MIR}_{XY} &= \lim_{N \rightarrow \infty} \lim_{L \rightarrow \infty} \sum_{i=1}^{L-1} \frac{I_{XY}(i+1, N) - I_{XY}(i, N)}{L} \\ &= \lim_{N \rightarrow \infty} \lim_{L \rightarrow \infty} \frac{I_{XY}(L, N) - I_{XY}(1, N)}{L} \\ &= \lim_{N \rightarrow \infty} \lim_{L \rightarrow \infty} \frac{I_{XY}(L, N)}{L}, \end{aligned} \quad (4)$$

where $I_{XY}(L, N)$ is the MI between X and Y in Eq. (1). In this framework, we consider trajectories of length L that follow an itinerary over cells on a grid of infinitely many cells N . Clearly, $I_{XY}(1, N)/L$ tends to zero in the limit of infinitely long trajectories, i.e., when $L \rightarrow \infty$.

For finite-length time-series X and Y , the definition in Eq. (4) can be further simplified, as demonstrated in [37], to

$$\text{MIR}_{XY}(N) = \frac{I_{XY}(N)}{T(N)}, \quad (5)$$

where $I_{XY}(N)$ is defined as in Eqs. (1) and (3), i.e., it is the MI between X and Y , and N is the number of cells in a Markov partition of order T . In particular, the probabilities to calculate I_{XY} are calculated in a Markov partition of order T where $T(N)$ is the shortest time for the correlation between X and Y to be lost in the particular Markov partition [37]. In the next section, we discuss $T(N)$ in more detail and elaborate on its estimation, showing how it can be approximated for finite-size times-series.

2.4. Estimation of the correlation-decay time

To calculate MI and MIR for a pair of variables X and Y , we define a 2D probability space Ω generated by X and Y . The space Ω is partitioned into a grid of $N \times N$ equally sized cells following the bin method [37,53]. The probability of an event i in X is then given

$$P_X(i) = \frac{\text{number of data points in column } i}{\text{total number of data points in } \Omega},$$

and of an event j in Y

$$P_Y(j) = \frac{\text{number of data points in row } j}{\text{total number of data points in } \Omega}.$$

Similarly, the joint probability can be defined by the ratio of points in cell (i, j) of the same partition in Ω and is expressed by

$$P_{XY}(i, j) = \frac{\text{number of data points in cell } (i, j)}{\text{total number of data points in } \Omega}.$$

Therefore, MI can be calculated from Eq. (3) for a grid size N , and is thus partition-dependent as it results in different values for different N .

To ensure there is always a sufficiently large number of data points in the cells of the partition of Ω , we require the average number of points in all occupied cells to be sufficiently larger than the number of occupied cells,

$$\langle N_0(N) \rangle \geq N_{oc}.$$

Here, N_{oc} is the number of occupied cells and $\langle N_0(N) \rangle$ the average number of points in all occupied cells in Ω .

To compute MIR using Eq. (5), we assume that an underlying Markov partition exists and estimate the correlation-decay time $T(N)$, which is the time when X and Y lose memory from their initial states. In systems with sensitive dependence on initial conditions, such as in chaotic systems, the system becomes unpredictable after $T(N)$ time. The correlation-decay time $T(N)$ can be calculated in different ways, e.g., by using the Lyapunov exponents of the dynamics, the largest expansion rates [39] or the diameter of an associated itinerary graph [37]. These ways exploit the fact that the dynamics in X and Y is chaotic and thus, points expand to the whole extend of Ω after about $T(N)$ time. In the following, we will focus on the itinerary graph method [37], and will discuss the details of its calculation and its relation to $T(N)$. The method is computationally fast and allows for successful network inference, as we will show in Section 3. These characteristics makes it a reasonable choice.

2.5. The itinerary graph method for the estimation of the correlation-decay time

The correlation-decay time $T(N)$ depends on quantities such as Lyapunov exponents and expansion rates [37], which are computationally expensive [39]. Here, we estimate it by using the diameter of an associated itinerary graph G [37], i.e., by computing the number of iterations it takes points in cells in Ω to expand and cover it completely. This is a necessary condition to determine the shortest time for the correlation to decay to zero. Following [37], we calculate $T(N)$ from the diameter of the network G , based on the dynamics of points mapped from one cell in Ω to another, namely, on a network where its connectivity is given by the transition of points from cell to cell in Ω , hence the term itinerary graph.

In particular, G can be constructed as follows: We assume that each equally sized cell in Ω , occupied by at least one point, represents a node in G . Following the transition of points from one cell to another (itinerary), we can form the connections between nodes in G . Specifically, a link between i and j nodes in G exists if points in Ω travel from cell i to cell j , where $i, j = 1, \dots, N^2$, i.e., they are the labels of the cells that partition Ω . If the link exists, the entry in G is equal to 1, otherwise it is set equal to 0. Therefore, G is defined as the corresponding binary, $N^2 \times N^2$, matrix of zeros and ones. Consequently, we define $T(N)$ as the diameter of G , i.e., as the longest of all shortest paths in G . This amounts to the shortest distance between the two most distant nodes in the network. In the context of the expansion of points in Ω , this is the minimum number of iterations it takes for points in cells in Ω to spread and cover it. Hence, the method transforms the calculation of the correlation-decay time $T(N)$ to the calculation of the diameter of G . In our work, we use Johnson's algorithm [59] to estimate the diameter of G and, thus to estimate $T(N)$.

Since we are using fixed-size, non-Markovian cells to compute $T(N)$ and MIR, errors will occur, causing a systematic bias toward larger MIR values. This means that MIR is partition-dependent and that it is overestimated. To account for these errors, and to render MIR partition-independent, the authors in [37] came up with two normalisation procedures, one with respect to the grid-sizes N and another one with respect to all pairs X and Y . This is because bigger N (smaller-size cells) are less likely to produce significantly different partitions to Markovian as opposed to smaller N (bigger-size cells). Moreover, using the fact that $\text{MIR}_{XY} = \text{MIR}_{YX}$ and $\text{MIR}_{XX} = 0$, one can narrow down the number of X, Y pairs of nodes from M^2 to $M(M - 1)/2$, which reduces considerably the computational time by almost a half.

2.6. Network inference based on MIR and statistical significance tests

Here, we adopt a simpler approach and do not normalise MIR as in [37,38]. Even though the errors due to the use of non-Markovian partitions will be present, they still allow for successful network inference when combined with statistical significance tests and the false discovery rate method for multiple hypothesis testing. This involves the use of the itinerary graph method in Section 2.5 to estimate the correlation-decay time $T(N)$ and Eq. (5) to compute MIR combined with the false discover rate method.

To infer the structure of a network from data, the authors in [37,38] used a threshold $\tau \in [0, 1]$ and considered that the pair X, Y is connected if the normalised MIR value is bigger or equal than the threshold. If so, the corresponding entry in the inferred adjacency matrix A^i is 1, otherwise it is 0. In that sense, the choice of an appropriate τ becomes crucial in recovering successfully the structure of the initial connectivity matrix. If τ is set too high, actual connections might be missed, and if set too low, spurious connections might appear in the inferred network. Usually, there is no a priori knowledge how to best set τ , which is addressed here by using a statistical approach which does not make use of thresholds.

In [9], the authors came up with a way to overcome this problem by determining τ , first by sorting all normalised MIR values in ascending order, and then by identifying the first X, Y pair for which the normalised MIR increases more than 0.1. This is an increase which accounts for an abrupt change in the normalised MIR values. Then, they set the threshold τ as the middle value of the normalised MIR for the identified pair and that for the immediately previous pair of nodes. This is based on the observation that there are two main groups of normalised MIR values, i.e., one for the connected and one for the unconnected groups of nodes in the network. However, this classification might not always be possible as no such gaps always appear or more than one gaps may appear (see for example panels (c) and (d) in Fig. 6 in [38] for financial and stock market data).

Here, assuming only access to recorded data and to overcome such difficulties, we introduce a statistical approach that does not require a threshold: For each pair of nodes X, Y , we compute MIR_{XY} using Eq. (5) and calculate a symmetric matrix of MIR values complemented by the use of the false discovery rate method for multiple hypothesis testing to accept or reject connections.

2.6.1. Calculation of surrogate data

The statistical approach followed here amounts to creating a number of surrogate data N^{SD} for each pair X, Y and to computing their MIR values using Eq. (5). Since we want to infer the network structure in datasets, we make use of random or twin surrogate data [60], depending on the properties of the original dataset, which is discussed next. In both cases, the surrogate data have the same length and dimensionality with the recorded time-series.

In particular, when producing random surrogate datasets, their entries are uniformly distributed random numbers in $(0, 1)$ and when producing twin surrogate datasets following [60], the method calculates blocks of surrogate data with the same second order properties as the original dataset by transforming the original data into the frequency domain, randomising simultaneously the phases across the time-series and converting the data back into the time domain. The first method destroys any possible relationship among the original data and the second destroys only the phase relationship in the original data, if there is any, preserving at the same time their correlation.

To decide whether to use random or twin surrogate data, we estimate: (i) the number of statistically significant correlated X, Y pairs by computing the correlation coefficients of the original dataset and their p -values for testing the hypothesis there is no relationship between X and Y (null hypothesis). If the p -values are smaller than 0.05 and the absolute value of the correlation coefficient is bigger than 0.5, then the correlation coefficient is considered statistically significant (correlation metric). (ii) the amount of additive noise in the original data assuming they are composed of smoothly varying functions added to Gaussian noise with mean zero and variance σ^2 . At any point, X can be represented by a low order polynomial, i.e., a truncated local Taylor series approximation. Applying finite differences, they kill off low-order polynomial terms, while producing a correlated series of points, where its variance is unchanged at σ^2 . Thus, the method is estimating the variance of additive noise in the original data (noise metric). (iii) the global phase-synchronisation in the original dataset by computing the order parameter $\rho \in [0, 1]$ following [61]. Here, $\rho = 0$ means complete desynchronisation and $\rho = 1$, complete synchronisation (global phase-synchronisation metric). If the mean of the noise metric over the nodes X is bigger than 10^{-5} or the number of statistically significant correlated X, Y pairs over all pairs is bigger than the estimated global phase-synchronisation metric, random surrogate datasets will be used, otherwise twin surrogate datasets will be used. This means if the original data are noisy or linearly correlated, random surrogate data will be used whereas if the original data are highly phase-synchronised, twin surrogate data will be used. The idea is that in any case, we use surrogate data with similar properties with the original data, removing the cause that leads to potential connectivity among the nodes, testing whether connectivity comes from the dynamics or due to chance.

The number of surrogate data, N^{SD} , is determined by setting a desired significance level p^{sl} , such that

$$N^{SD} = \frac{1}{p^{sl}}.$$

For example, if we are interested in a 95% confidence interval, then $p^{sl} = 0.05$ and we will use $N^{SD} = 1/0.05 = 20$ surrogate datasets. We quantify how successful the network inference is by means of ROC curves, using the p^{sl} values

10^{-1} , 10^{-2} and 10^{-3} that correspond to $N^{SD} = 10$, 100 and 1000 surrogate datasets, respectively. These curves quantify the true positive rate (TPR) and false positive rate (FPR) as a function of the number of time samples. TPR is the ratio between the number of links that are correctly inferred and the number of links that actually exist in the network. On the other hand, FPR is the ratio between the number of links that are incorrectly inferred and the actual number of non-existing links in the network. The ROC curves provide further information about the type of errors made in the inference method. In particular, TPR and FPR are defined by

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{FN}},$$

where TP are the true positive, FN the false negative and FP the false positive connections in the inferred connectivity matrix when compared with the initial connectivity matrix. Both TPR and FPR take values in $[0, 1]$. When plotting FPR on the horizontal axis and TPR on the vertical axis, perfect inference corresponds to the point with $\text{TPR} = 1$ and $\text{FPR} = 0$, i.e., to the point at the upper left corner on the ROC plot with coordinates $(0, 1)$. Positive FPR means inferred connections that are not present in the initial connectivity matrix (false positives). Thus, for successful inference, one would typically require TPR to be as close as possible to 1 and FPR as close as possible to 0, or ideally TPR to be equal to 1 and FPR equal to 0, which corresponds to perfect inference.

2.6.2. Statistical approach to test for connectivity and the inferred adjacency matrix

To test whether nodes X , Y are connected due to their dynamics, we start by defining the same significance level p^{sl} for all pairs of nodes and the null hypothesis that they are unconnected. Then, we count the number of pairs of surrogate data their MIR value (obtained test statistic) is bigger or equal than the MIR value of X and Y (observed test statistic) and denote it by c_{XY} . We use this to compute the probability $p_{XY} = c_{XY}/N^{SD}$, where N^{SD} is the number of surrogate data. p_{XY} is the probability of obtaining a test statistic that is as or more extreme than the observed one, assuming the null hypothesis is true. For example, if $p_{XY} = 0.03$, it would mean that if the null hypothesis is true, there would be a 3% chance of obtaining the observed test statistic or a more extreme one. Since this is a small probability, we would reject the null hypothesis and would say that nodes X and Y are connected.

Since we are conducting multiple comparisons for the $M(M-1)/2$ pairs of nodes, there is an increased probability of accepting false positives or type I errors, i.e., accepting the rejection of true null hypotheses. Equivalently, this would mean that even though X and Y would be unconnected, we would accept them as connected. To account for such errors, we perform a false discovery rate multiple hypothesis testing [62] based on the probabilities p_{XY} and a desired false discovery rate of $p^{sl}/100$ to guarantee that the probability of having one or more false positives is p^{sl} or less. The FDR procedure described in [62] is accurate for any test dependency structure (e.g., for Gaussian variables with any covariance matrix). FDR controls the false discovery rate of a family of hypothesis tests and is the expected proportion of rejected hypotheses that are mistakenly rejected (i.e., the null hypothesis is actually true for those tests). FDR is generally a somewhat less conservative and more powerful method for correcting for multiple comparisons than the Bonferroni correction that provides strong control of the family wise error rate, i.e., the probability that one or more null hypotheses are mistakenly rejected.

If the output of the application of the FDR method for a pair X , Y is 1, then the test statistic that produced the p_{XY} value is statistically significant and the null hypothesis of the test for the pair X , Y must be rejected. This means if the null hypothesis is true (i.e., nodes X and Y are unconnected), there would be a p_{XY} (small) chance of obtaining the observed test statistic or a more extreme. In this case, since the null hypothesis must be rejected, nodes X and Y are connected and the corresponding entry in the inferred adjacency matrix A^i is set equal to 1. If the null hypothesis cannot be rejected, then nodes X and Y remain unconnected and the corresponding entry in the inferred adjacency matrix A^i is set equal to 0. Following this approach for all $M(M-1)/2$ pairs of nodes, we compute the inferred, symmetric, binary adjacency matrix A^i , where its entries A_{kl}^i are given by

$$A_{kl}^i = \begin{cases} 0, & \text{when no connection between nodes } k \text{ and } l \text{ has been inferred from the statistical approach} \\ 1, & \text{when a connection between nodes } k \text{ and } l \text{ has been inferred from the statistical approach} \end{cases}.$$

Consequently, if a connection between nodes k and l in A^i has been inferred, then a connection between nodes l and k is implied (undirected connections). Finally, A^i is compared with the initial connectivity matrix A^0 to compute the ROC curves as a function of the number of time samples in the datasets.

3. Results

We start by applying the proposed methodology to correlated normal-variates data from stochastic processes in the absence of underlying equations of motion, and then, to data from deterministic, chaotic systems of coupled maps and coupled ordinary differential equations (ODEs), which we call chaotic dynamical units. In all cases, we aim to recover the initial connectivity matrices that were used to couple the processes or chaotic dynamical units.

The coupling is given by undirected networks in the form of binary, symmetric, adjacency matrices (initial connectivity matrices) that are compared with those that result from the application of the method proposed in Section 2 (inferred matrices). By binary matrices, we mean that any two nodes in the network can be either bidirectionally connected, which corresponds to an entry equal to 1 or unconnected, which corresponds to an entry equal to 0 in the matrix. Since the connections are undirected, the adjacency matrices (initial connectivity and inferred ones) are symmetric.

We couple the deterministic, chaotic dynamical units according to the initial connectivity matrices, record the evolution of their dynamics and produce time-series data. We use the recorded time-series in a pairwise fashion and the method proposed in Section 2 to infer the initial connectivity matrices and quantify how successful the inference is as a function of the number of time-samples. We compute ROC curves, where the horizontal axis is the rate of false positive connections (FPR) and the vertical the rate of true positive connections (TPR). In that sense, perfect inference corresponds to $\text{TPR} = 1$ and $\text{FPR} = 0$, i.e., to the point $(0, 1)$ on the ROC plane. When $\text{TPR} = 1$, it means the proposed methodology was able to infer correctly all connections in the initial connectivity matrix, and when $\text{FPR} = 0$, it means no incorrectly inferred connections were found when compared to the initial connectivity matrix. Finally, we plot TPR and FPR as a function of time-sample sizes to show convergence to the point with coordinates $(0, 1)$ on the ROC plane, which corresponds to perfect network inference.

3.1. Correlated normal-variates data

We start by applying the methodology in Section 2 to correlated normal-variates data motivated by situations where data can only be recorded, as the underlying dynamics and equations of motion are not known. For example, data from global financial markets of different countries can be recorded over time and are found to be correlated [38], however the underlying equations that govern their evolution are not known [63].

To demonstrate the effectiveness of the method, we used the data of three groups of correlated normal-variates in [38]. Each group $i = 1, 2, 3$ consists of three correlated normal-variates specified by a covariance matrix Σ_i , where the three groups are uncorrelated with each other. Fig. 1(A) shows the scatter matrix of the three groups. The first group (i.e., $i = 1$) consists of variables x_1, x_2, x_3 , the second (i.e., $i = 2$) of variables x_4, x_5, x_6 and the third (i.e., $i = 3$) of variables x_7, x_8, x_9 . The covariance matrices Σ_i of the three groups are given by

$$\Sigma_1 = \begin{bmatrix} 3.40 & -2.75 & -2.00 \\ -2.75 & 5.50 & 1.50 \\ -2.00 & 1.50 & 1.25 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 0.5 & 0.3 \\ 0.3 & 0.3 & 0.3 \end{bmatrix}, \quad \Sigma_3 = \begin{bmatrix} 1.40 & -2.75 & -2.00 \\ -2.75 & 5.50 & -1.00 \\ -2.00 & -1.00 & 3.25 \end{bmatrix},$$

respectively, where the subscripts denote the groups.

An ellipsoidal or circular pattern in Fig. 1(A) indicates the degree of independence of the datasets, i.e., if they are weakly correlated or uncorrelated. A cigar-shaped cloud of points shows stronger correlation, which can be either positive (correlation) or negative (anti-correlation). This is in agreement with the signs of the entries in the covariance matrices. The plots in the diagonal of Fig. 1(A) show the histograms of the datasets with themselves, from which it results they are normally distributed. For example, one can observe that datasets x_8 and x_9 are weakly anti-correlated and would not expect to see a connection in the inferred network. In contrast, datasets x_1 and x_3 are strongly anti-correlated and one would expect to see a connection in the inferred network. Fig. 1(B) shows the initial connectivity network, which is the same with the inferred network that resulted from the application of the proposed methodology. The three groups are apparent in the network in the form of the three disconnected subnetworks.

This is a case where there is a clear distinction between correlated and uncorrelated pairs of nodes as they were constructed per se, what leads to the three disconnected subnetworks in Fig. 1(B). Here, we used random surrogate data for the computations of the ROC curves in Fig. 1 for $p^{sl} = 10^{-1}, 10^{-2}$ and 10^{-3} as the computation of the metrics showed the original data are noisy and correlated, consistent with what is expected from correlated normal-variates data. Fig. 1(C) shows the ROC curves (left column of plots), FPR as a function of time-sample sizes (middle column) and TPR as a function of time-sample sizes (right column) for $p^{sl} = 10^{-1}, 10^{-2}$ and 10^{-3} , from top to bottom. One can see that for $p^{sl} = 10^{-1}$ (top row of plots in panel (C)), FPR fluctuates between 10^{-1} and 10^{-2} for about 20,000 time samples and then drops to 0 for bigger time-sample sizes. TPR on the other hand remains fixed at 1 for about 20,000 time samples showing that all actual connections were inferred correctly. The computations stopped after about 20,000 time samples as FPR became zero (i.e., no false-positive connections were inferred, as shown in Fig. 1(B)) and TPR became 1 (i.e., all true positive connections were inferred correctly, as shown in Fig. 1(B)) for four consecutive time-sample sizes. These results demonstrate that about 20,000 time samples are enough to successfully infer the network in Fig. 1(B) for a significance level $p^{sl} = 10^{-1}$. A similar conclusion can be drawn for the smaller significance levels $p^{sl} = 10^{-2}$ and 10^{-3} by looking at the results in the second and third rows of panel (C) (from top to bottom), respectively. In that case, a much smaller size (of about 4000) time samples is enough to guarantee perfect network inference, i.e., $\text{TPR} = 1$ and $\text{FPR} = 0$. The absence of points in the FPR vs. time samples plots (middle column) in panel (C) means that $\text{FPR} = 0$ as the scale is logarithmic, i.e., no false positive connections were inferred. The results in Fig. 1(C) show that the proposed methodology inferred successfully the network from correlated normal-variates data. One can see that the data were successfully classified into the three disconnected groups. For example, the connection between variables x_8 and x_9 is missing as they are weakly anti-correlated and the connection between x_1 and x_3 is present as they are strongly anti-correlated. This is also corroborated by the results in

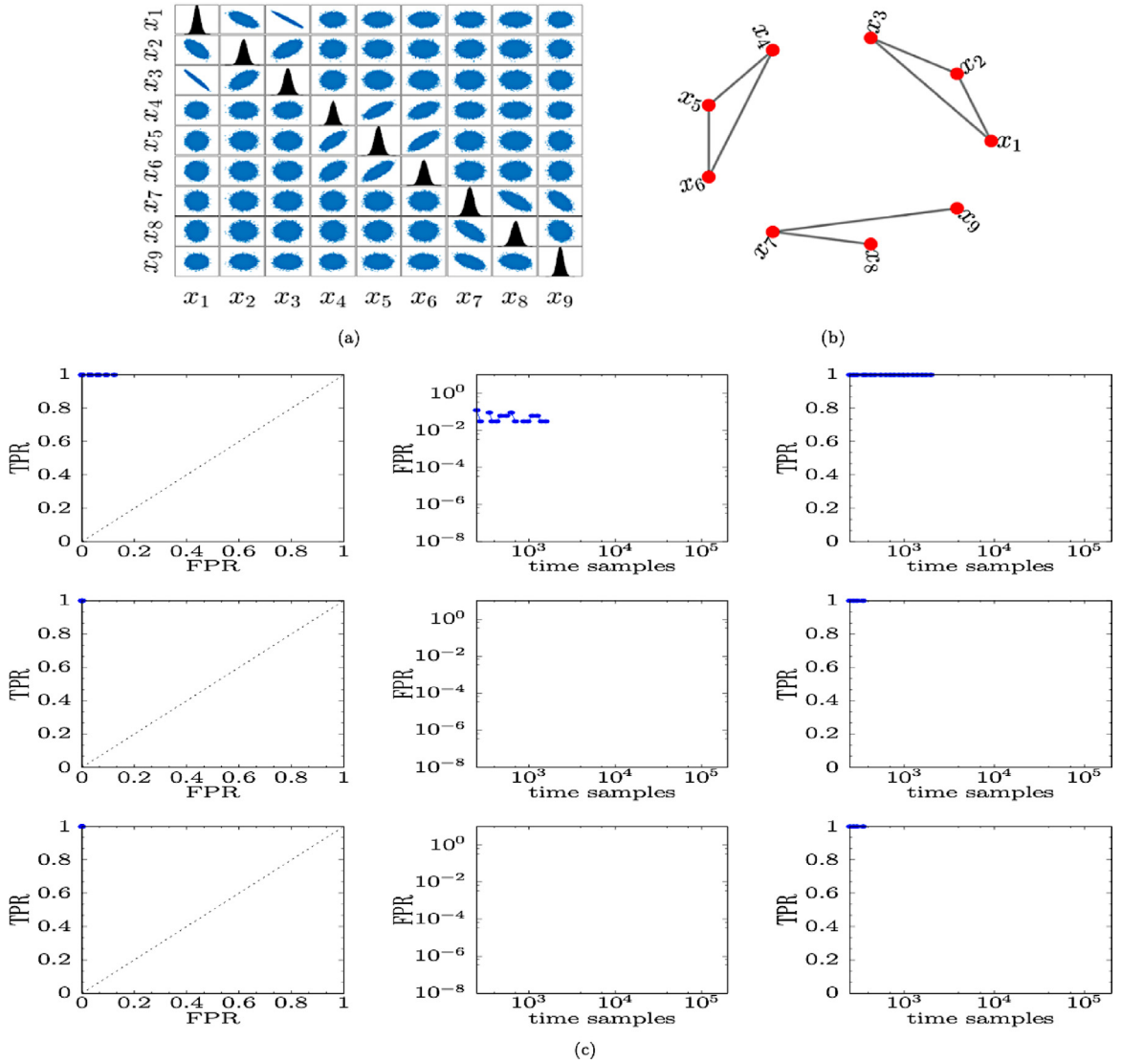


Fig. 1. Application of the proposed methodology to correlated normal-variables data. Panel (A) shows the scatter matrix of the nine datasets that split into the groups x_1, x_2, x_3 (first), x_4, x_5, x_6 (second) and x_7, x_8, x_9 (third). Each group consists of three correlated normal-variables, and the groups are uncorrelated with each other. An ellipsoidal or circular pattern in panel (A) indicates the degree of independence of the datasets, whereas a cigar-shaped cloud of points shows stronger correlation, which can be either positive (correlation) or negative (anti-correlation). The plots in the diagonal of panel (A) show the histograms of the datasets themselves, revealing they are normally distributed. Panel (B) shows the network configuration that results from panel (A), which is the same with the inferred network that resulted from the application of the proposed methodology. Panel (C) shows the ROC curves (left column of plots), FPR vs. time samples (middle column) and TPR vs. time samples (right column) for $p^{sl} = 10^{-1}, 10^{-2}$ and 10^{-3} , from top to bottom. Note that the vertical axes in the middle column and, horizontal axes in the middle and right columns of plots in panel (C) are in logarithmic scale and that 0 FPR values in the plots in the middle column of panel (C) are not plotted as the scales in the vertical axes are logarithmic.

panel (B) which shows the initial connectivity network with all nodes and connections depicted, being the same with the inferred network that resulted from the application of the proposed methodology. Concluding, we have shown that the proposed methodology can be used to recover the correct number and pairs of correlated normal-variables data and thus, successfully infer the underlying network structure.

3.2. Network of coupled circle maps

Next, we applied the method to data obtained from a system of deterministic, chaotic, coupled nonlinear maps and in particular, from coupled circle maps. The network that was used to couple them is shown in Fig. 2(A) and is composed

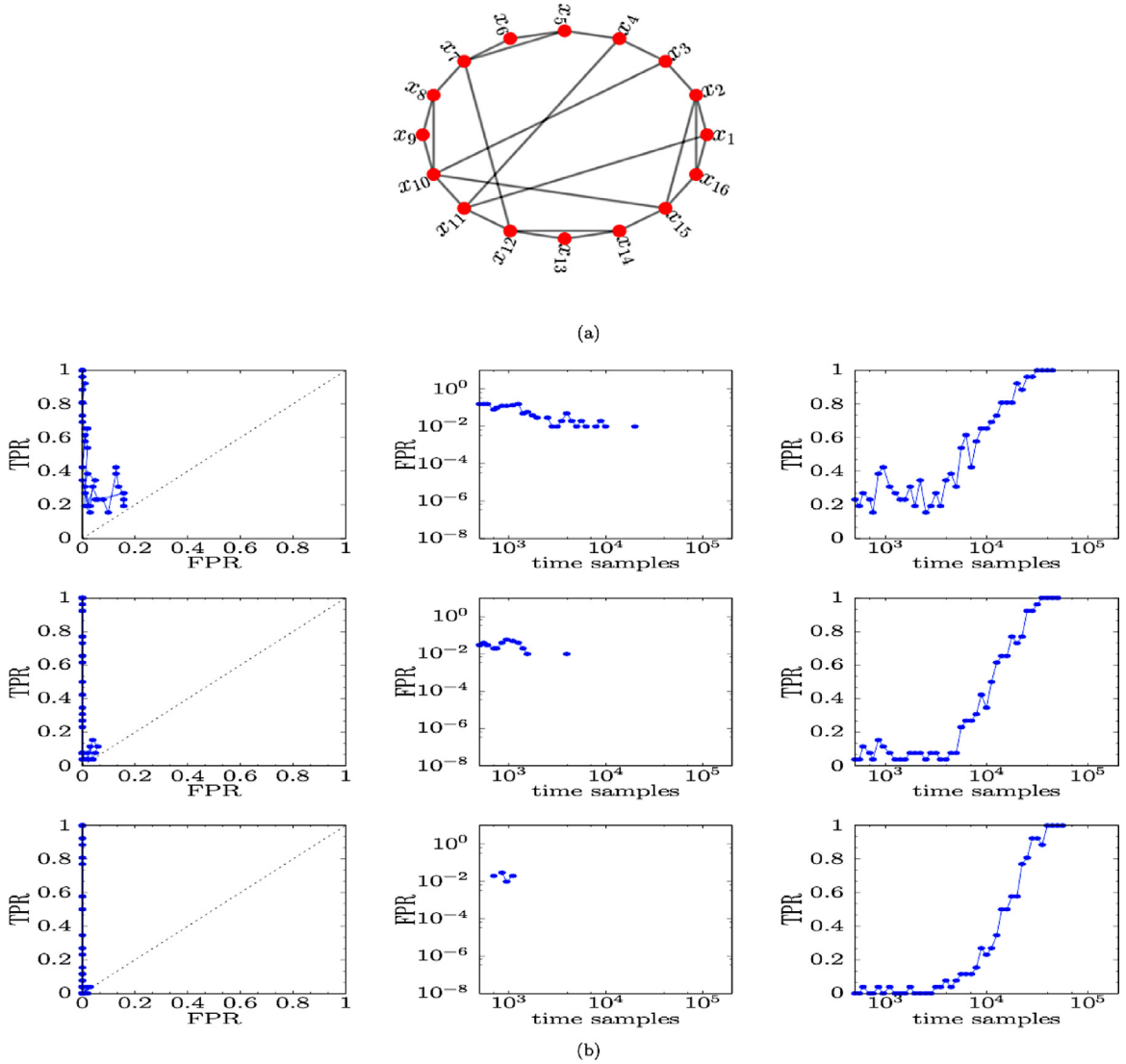


Fig. 2. Application of the proposed methodology to coupled circle maps data for $\alpha = 0.03$ and $K = 6.9115$. Panel (A) shows the network of the 16 nodes. Panel (B) shows the ROC curves (left column of plots), FPR vs. time samples (middle column) and TPR vs. time samples (right column) for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. Note that the vertical axes in the middle column and, horizontal axes in the middle and right columns of plots in panel (B) are in logarithmic scale and that 0 FPR values in the plots in the middle column of panel (B) are not plotted as the scales in the vertical axes are logarithmic.

of $M = 16$ nodes (circle maps) with dynamics given by [64]

$$x_{n+1}^i = (1 - \alpha)f(x_n^i, r) + \frac{\alpha}{k_i} \sum_{j=1}^M A_{ij}f(x_n^j, r), \quad (6)$$

where x_n^i is the n th iterate of map $i = 1, \dots, M$ and $\alpha \in [0, 1]$ is the coupling strength. (A_{ij}) is the binary, symmetric, adjacency matrix of the network shown in Fig. 2(A). The degree of node i is $k_i = \sum_{j=1}^M A_{ij}$ and r is the parameter of each circle map, which assumes the same value for all M maps. In particular, in Eq. (6), $f(x_n, r)$ is the circle map, defined by

$$f(x_n, r) = x_n + r - \frac{K}{2\pi} \sin(2\pi x_n) \mod 1. \quad (7)$$

Following [37], we used $\alpha = 0.03$ to build a weakly interacting system of coupled circle maps with $r = 0.35$ and $K = 6.9115$. These values correspond to fully developed chaos for each circle map (see Eq. (7)). The initial conditions

x_0^i were initialised randomly, drawn from the uniform distribution, in $[0, 1]$ and the system was run for a total of 10^6 iterations. We discarded the first 8×10^5 iterations and kept the subsequent 2×10^5 iterations to use as the dataset to infer the initial connectivity matrix. Here, we used random surrogate data for the computations of the ROC curves in Fig. 2(B) for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} as the computation of the metrics showed the original data are mainly noisy, what is expected from chaotic datasets.

Fig. 2(A) shows the actual network used to generate the data for the coupled circle maps (7) in Eq. (6). Fig. 2(B) shows the ROC curves (left column of plots), FPR as a function of time-sample sizes (middle column) and TPR as a function of time-sample sizes (right column) for the significance levels $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. One can see that for $p^{sl} = 10^{-1}$ (top row of plots in panel (B)), FPR fluctuates between 10^{-1} and 10^{-2} for about 20,000 time samples after which it drops to 0 for bigger time-sample sizes. TPR on the other hand fluctuates between 0.2 and 0.5 until 50,000 time samples, after which it increases and reaches 1 at about 30,000 time samples, showing that all actual connections in the initial connectivity matrix were inferred correctly. The computations stopped after about 50,000 time samples as FPR became zero (i.e., no false-positive connections were inferred as shown in panel (B)) and TPR became 1 (i.e., all true positive connections were inferred as shown in panel (B)) for four consecutive time-sample sizes. These results demonstrate that about 50,000 time samples are enough to successfully infer the network shown in panel (B) for $p^{sl} = 10^{-1}$. As expected, a similar conclusion can be drawn for the smaller significance levels $p^{sl} = 10^{-2}$ and 10^{-3} by looking at the results in the second and third rows of panel (B) (from top to bottom), respectively. Again, about 50,000 time samples are enough for perfect network inference, i.e., TPR = 1 and FPR = 0. The absence of points in the FPR vs. time samples plots (middle column) in panel (B) for smaller time-sample sizes means that FPR becomes 0 faster than it takes TPR to become 1.

The results in Fig. 2(B) suggest that the proposed methodology inferred successfully the network of deterministic, chaotic, dynamics.

3.3. Network of coupled logistic maps

Here, we apply our methodology to data from coupled logistic maps [37]. Again, they come from a system of deterministic, coupled nonlinear equations that give rise to chaotic dynamics. The network used comprises $M = 30$ nodes as shown in Fig. 3(A), and we show that the method works for larger-size networks. The dynamics is given by Eq. (6), whereby each logistic map f_i is given by

$$f_i(x_n^i, r) = rx_n^i(1 - x_n^i), \quad (8)$$

where x_n^i is the n th iterate of map $i = 1, \dots, M$ and r the parameter of each logistic map, which is the same for all M maps. In this study, we have used $r = 4$ which corresponds to fully developed chaos for each individual logistic map.

Following [37], we used $\alpha = 0.06$ to build a weakly interacting system of coupled, chaotic, logistic maps. The initial conditions x_0^i were initialised randomly in the interval $[10^{-10}, 0.3]$ and the system was run for 10^6 iterations. We discarded the first 8×10^5 iterations and kept the next 2×10^5 to use as the dataset to infer the network structure, shown in Fig. 3(A). We used random surrogate data for the computations of the ROC curves in Fig. 3(B), for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} as the computation of the metrics showed the original data are mainly noisy, what is expected from chaotic datasets.

Fig. 3(A) shows the network of $M = 30$ nodes used to produce the data for the coupled logistic maps (8) in Eq. (6). Fig. 3(B) shows the ROC curves (left column of plots), FPR (middle column) and TPR (right column) as a function of time-sample sizes for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. One can see that for $p^{sl} = 10^{-1}$ (top row of plots in panel (B)), FPR fluctuates between 10^{-1} and 10^{-3} for the first 1200 time samples after which it plummets to 0 as bigger time-sample sizes are considered, with a single exception at 8000 time samples where it attains a value between 10^{-2} and 10^{-3} . TPR on the other hand fluctuates between 0 and 0.2 for the first 2000 time samples, after which it increases quickly with a linear trend, reaching TPR = 1 at about 40,000 time samples, implying that all actual connections in the initial connectivity matrix were inferred correctly. The computations stopped after about 60,000 time samples as FPR becomes zero (i.e., no false-positive connections were inferred as shown in panel (B)) and TPR becomes 1 (i.e., all true positive connections were inferred as shown in the same panel) for four consecutive time-sample sizes. These results demonstrate that about 40,000 time samples are enough to successfully infer the network shown in panel (A) for the significance level $p^{sl} = 10^{-1}$. As expected, similar conclusions can be drawn for the smaller significance levels $p^{sl} = 10^{-2}$ and 10^{-3} by looking at the results in the second and third rows of panel (B) (from top to bottom), respectively. Again, about 40,000 time samples are enough to guarantee perfect network inference of TPR = 1 and FPR = 0. The absence of points in the FPR vs. time-samples plots (middle column) in panel (B) for smaller time-sample sizes means that FPR = 0, i.e., no false positive connections were found.

The results in Fig. 3(B) suggest that the proposed methodology can be used to successfully infer the network of deterministic dynamics and chaotic datasets, and thus, successfully infer the underlying network structure for larger-size networks.

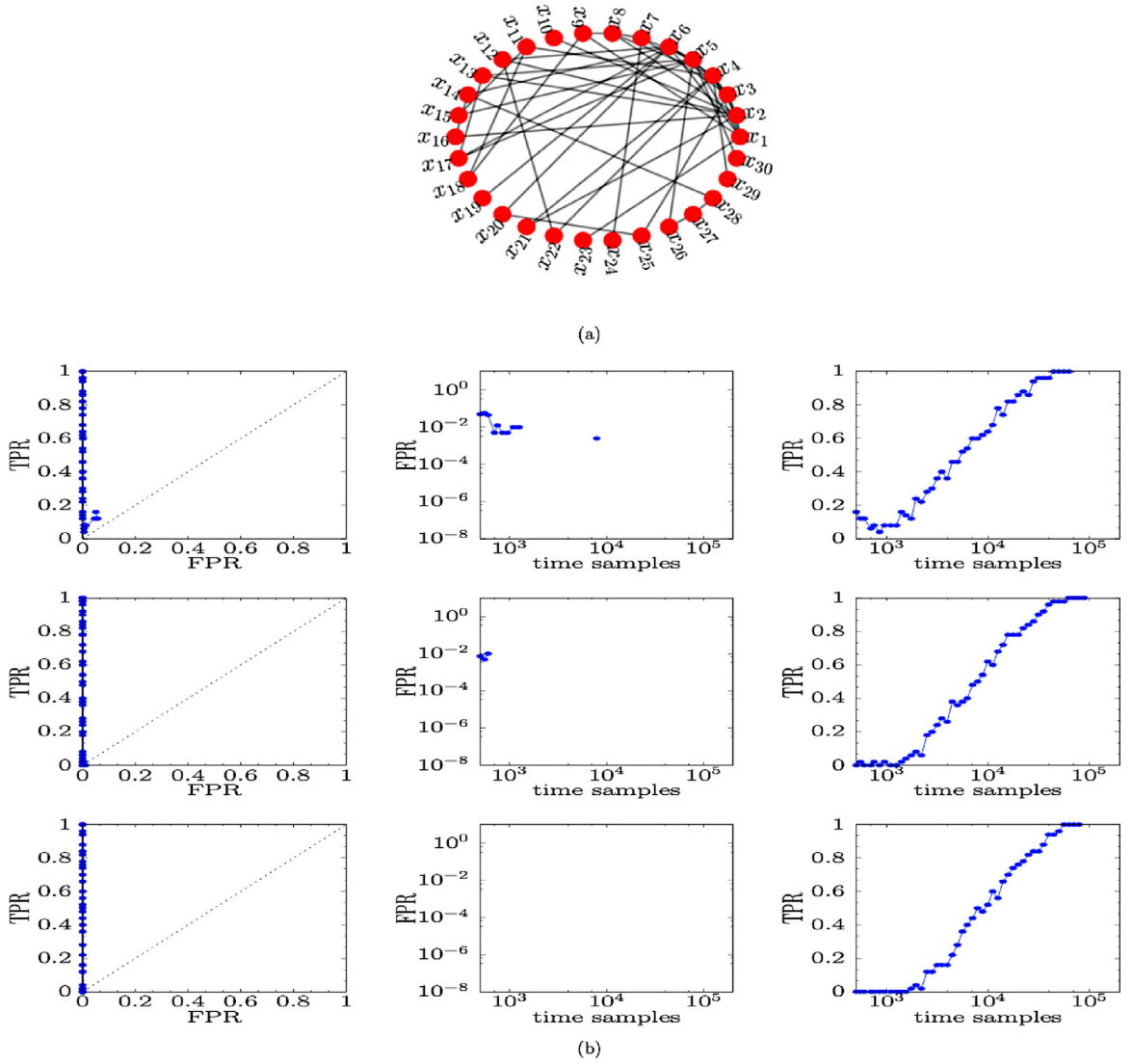


Fig. 3. Application of the proposed methodology to coupled logistic maps for $\alpha = 0.06$. Panel (A) shows the network configuration for the system of 30 coupled logistic maps. Panel (B) shows the ROC curves (left column of plots), FPR vs. time samples (middle column) and TPR vs. time samples (right column) for $p^sl = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. Note that the vertical axes in the middle column and, horizontal axes in the middle and right columns of plots in panel (B) are in logarithmic scale and that 0 FPR values in the plots in the middle column of panel (C) are not plotted as the scales in the vertical axes are logarithmic.

3.4. Network of coupled Lorenz systems

Next, we apply the proposed methodology to data obtained from $M = 16$ coupled Lorenz systems (see Eqs. (9)), similar to the system studied in [23]. The authors therein, studied the system in the presence of independent Gaussian noise with zero mean whereas here, we study the same system in the absence of noise as a similar analysis will be presented in Section 3.5. The network is given in Fig. 2(A) and the adjacency matrix (A_{ij}) in Eq. (9) is the adjacency matrix of that network.

In particular, we considered an undirected network with Lorenz dynamics at the nodes, which are connected via a diffusive coupling with strength K through the x units

$$\begin{aligned}\dot{x}_i &= \sigma(y_i - x_i) + K \sum_{j=1}^M A_{ij}(x_j - x_i), \\ \dot{y}_i &= x_i(\rho - z_i) - y_i,\end{aligned}\tag{9}$$

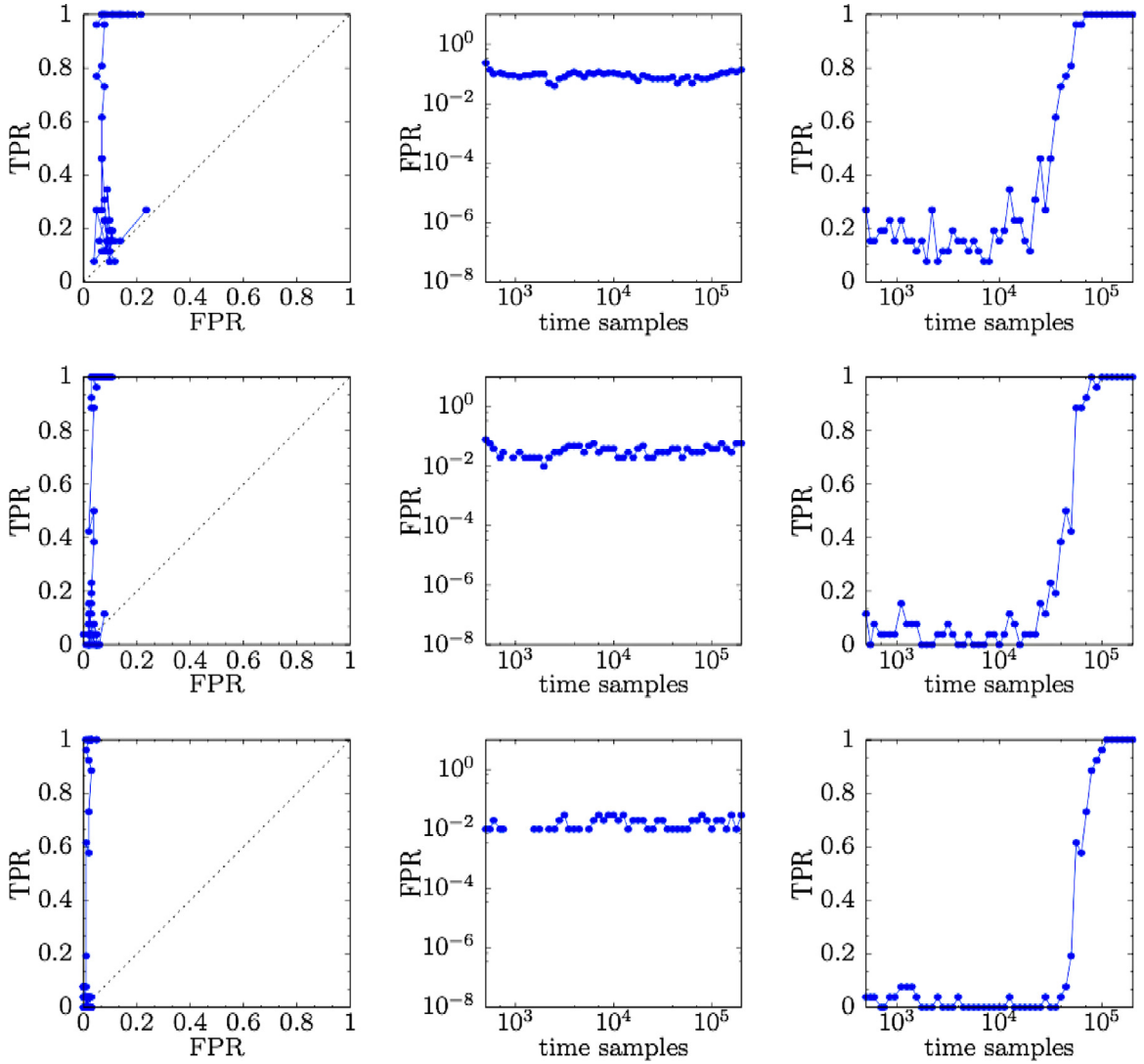


Fig. 4. Application of the proposed methodology to data from $M = 16$ coupled Lorenz systems for $K = 1.2$. The ROC curves (left column of plots), FPR vs. time samples (middle column) and TPR vs. time samples (right column) for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. Note that the vertical axes in the middle column and, horizontal axes in the middle and right columns of plots are in logarithmic scale and that 0 FPR values in the plots in the middle column are not plotted because of the logarithmic scale used in the vertical axes of FPR.

$$\dot{z}_i = x_i y_i - \beta z_i,$$

where $i = 1, \dots, M$, $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. We have chosen $K = 1.2$ as the fixed coupling term for all Lorenz systems as it leads to chaotic dynamics. The initial conditions for $i = 1, \dots, M$ are given by

$$\begin{aligned} x_i &= 0.1 + 0.5\xi_i^x, \\ y_i &= -0.19 + 0.5\xi_i^y, \\ z_i &= -0.27 + 0.5\xi_i^z, \end{aligned}$$

where ξ_i^x , ξ_i^y and ξ_i^z are uniformly random numbers in $[0, 1]$. The system was solved using the Euler integration (first-order) method with a time step $h = 0.01$ and was run up to final integration time $t_f = 1,004,000$. Similar results were obtained using the classic Runge–Kutta fourth order integration method. We considered the last 4×10^5 data points to use as the dataset to infer the network structure (A_{ij}). Here, we used twin surrogate data [60] for the computations of the ROC curves in Fig. 4, for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} (from top to bottom) as the computation of the metrics showed the original data were more phase correlated than noisy. This method [60] calculates blocks of surrogate data with the same second order properties as the original time-series dataset by transforming the original data into the frequency domain,

randomising the phases simultaneously across the time-series and converting the data back into the time domain. The main effect of the method is that it destroys the phase relationship in the original data (if there is any) preserving at the same time the correlation among the datasets.

Fig. 4 shows the ROC curves (left column of plots), FPR as a function of time-sample sizes (middle column) and TPR as a function of time-sample sizes (right column) for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. One can see that for $p^{sl} = 10^{-1}$ (top row of plots), FPR fluctuates around 10^{-1} , which shows the method identified connections as true ones even though they are not present in the initial connectivity matrix, i.e., false positive connections. Also, there is an apparent trend for FPR to increase slightly after about 130,000 time-samples. TPR on the other hand fluctuates between 0 and 0.4 for the first 30,000 time samples, after which it increases with a linear trend to TPR = 1 at about 70,000 time samples, implying that all actual connections in the initial connectivity matrix were inferred correctly. These results are better appreciated in the TPR vs. FPR plot in the first row, where it is clear that the ROC curve tends to stay close to the point (0, 1) of perfect inference, even though there are still connections that are falsely identified. The results for $p^{sl} = 10^{-1}$ in the first row demonstrate that about 70,000 time samples are enough to infer all connections in the original network shown in Fig. 2(A), even though at the same time a significant amount of false positive connections is identified. Similar conclusions can be drawn for the smaller significance levels $p^{sl} = 10^{-2}$ and 10^{-3} by looking at the results in the second and third rows (from top to bottom), respectively. Evidently, about 70,000 to 100,000 time samples are enough to infer the original network structure with FPR values decreasing for decreasing significance levels, implying smaller identification errors of false positive connections. The absence of points in the FPR vs. time-samples plots (middle column) means that FPR = 0, i.e., no false positive connections were inferred. The ROC curves approach the vertical axis at FPR = 0 decreasing by one order of magnitude as the significance level decreases by the same order of magnitude, with TPR approaching 1. Concluding, the quality of the network inference based on MIR improves by one order of magnitude when the significance level decreases by one order of magnitude.

3.5. Network of coupled stochastic Kuramoto phase oscillators

In this section, we applied the proposed methodology to data obtained from $M = 16$ coupled stochastic Kuramoto phase oscillators (see Eqs. (10)) [10]. The network is shown in Fig. 2(A) and the adjacency matrix (A_{ij}) in Eqs. (10) is the adjacency matrix of that network. We consider an undirected network of stochastic Kuramoto phase-oscillators dynamics at the nodes, which are connected through the system of equations

$$d\theta_i = \omega_i dt + \frac{K}{M} \sum_{j=1}^M A_{ij} \sin(\theta_j - \theta_i) dt + D dW_t^i, \quad (10)$$

where θ_i and ω_i are respectively, the phase and natural frequency of oscillator $i = 1, \dots, M$, and K the fixed coupling constant, which is assumed the same for all oscillators. The term dW_t^i is a Weiner process with 0 mean and variance tuned by the parameter D . The distribution of the natural frequencies ω_i is considered to be centred around 0, since the equations are invariant under the phase translation $\theta_i \rightarrow \theta_j - \langle \omega_i \rangle t$, where $\langle \omega_i \rangle$ is the average natural frequency over all oscillators in the network. Also, it is taken symmetric around 0 [65] and we assume that the frequencies of the oscillators are uniformly, randomly distributed in $[-2\pi, 2\pi]$.

The initial conditions used for $i = 1, \dots, M$ are given by

$$\theta_i = 2\pi\xi_i,$$

where ξ_i is a uniformly random number in $[0, 1]$. The system was solved by a weak Runge–Kutta method of order 2 with time step $h = 0.05$ and was run up to final integration time $t_f = 25,000$. We considered the second half 250,000 data points to use as the dataset to infer the network structure (A_{ij}). We used random surrogate data for the computations of the ROC curves in Fig. 5, for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} (from top to bottom) as the computation of the metrics showed the original data are mainly noisy, what is expected from stochastic dynamics. Following [10], for noisy data, one can derive two quantities (or probes) from the phases θ_i that can be used to infer the structure of the network, namely the instantaneous frequencies $\dot{\theta}_i(t)$ and Y -projections $\sin(\theta_i(t))$, i.e., the vertical projections of the unitary vectors of the phases $\theta_i(t)$. Our analysis showed that the quality of inference when using the solution to Eqs. (10) itself or the Y -projections is inferior to that obtained from the instantaneous frequencies. Thus, in Fig. 5, we present a representative example of network inference for the case $K = 4$ and $D = 0.05$ using the instantaneous frequencies as a probe.

Fig. 5 shows the ROC curves (left column of plots), FPR as a function of time-sample sizes (middle column) and TPR as a function of time-sample sizes (right column) for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. One can see that for $p^{sl} = 10^{-1}$ (top row of plots), FPR fluctuates between 10^{-1} and 10^{-2} up to 8000 time samples, after which it becomes 0. In other words, no more any false positive connections were inferred as the number of time samples increased. TPR on the other hand fluctuates between 0.55 and 0.9 for the first 2500 time samples, after which it reaches up to TPR = 1 at about 3000 time samples, implying that all actual connections in the initial connectivity matrix were inferred correctly.

These results are better appreciated in the TPR vs. FPR plot in the first row, where it is evident that the points on the ROC curve converge to the point (0, 1) of perfect inference. The results for $p^{sl} = 10^{-1}$ in the first row demonstrate that about 8000 time samples are enough to infer correctly all connections in the original network shown in Fig. 2(A). Similar

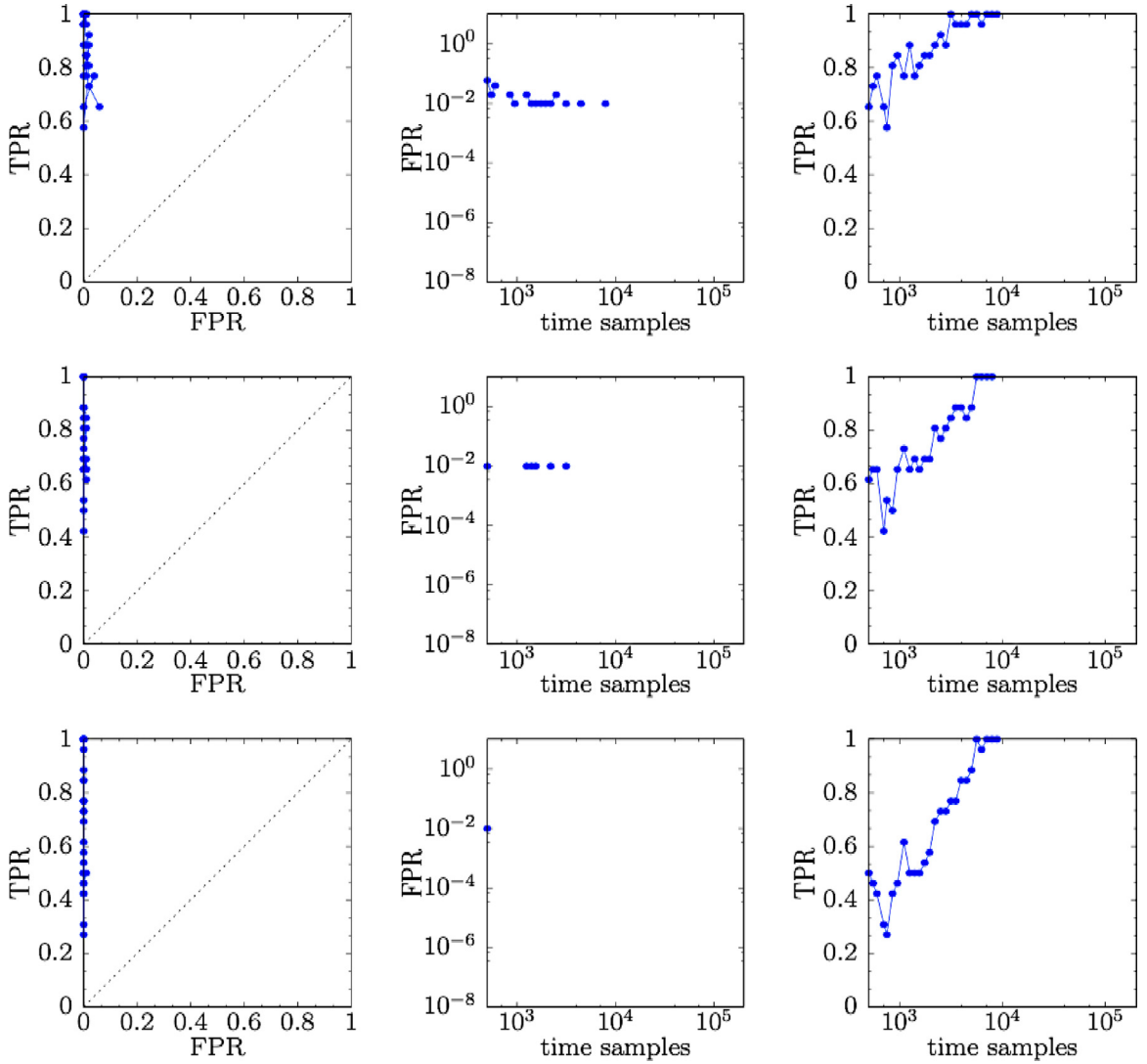


Fig. 5. Application of the proposed methodology to coupled stochastic Kuramoto phase-oscillators data for $K = 4$ and $D = 0.05$. The ROC curves (left column of plots), FPR vs. time samples (middle column) and TPR vs. time samples (right column) for $p^{sl} = 10^{-1}$, 10^{-2} and 10^{-3} , from top to bottom. Note that the vertical axes in the middle column and, horizontal axes in the middle and right columns of plots are in logarithmic scale and that 0 FPR values in the plots in the middle column are not plotted because of the logarithmic scale used in the vertical axes of FPR.

conclusions can be drawn for the smaller significance levels $p^{sl} = 10^{-2}$ and 10^{-3} when looking at the results in the second and third rows (from top to bottom), respectively. These findings show that when considering smaller significance levels, the quality of inference improves. The absence of points in the FPR vs. time-samples plots (middle column) means that $FPR = 0$. The ROC curves approach the vertical axis at $FPR = 0$ decreasing by one order of magnitude as the significance level decreases by one order of magnitude, with TPR approaching 1. Concluding, the quality of the network inference improves by one order of magnitude as the significance level decreases by one order of magnitude.

3.6. Effect of noise on the proposed inference method in networks of coupled stochastic Kuramoto phase oscillators

We now move to the study of the effect of additive noise as a function of time-sample sizes on the ability of the proposed methodology to successfully infer the structure of four types of networks. We consider the network of 16 nodes in Fig. 6(A), which is the same with the network in Fig. 2(A), the scale-free network in Fig. 6(B), the Erdős-Rényi network with rewiring probability 0.1 in Fig. 6(C) and the small-world network in Fig. 6(D). In all cases, the time-sample sizes range between 250 and 100,000 and the strength of the additive noise D is in $[0, 0.2]$. Also, in all cases, the dynamics is given by Eq. (10) of coupled stochastic Kuramoto phase oscillators with all parameter values and initial conditions as in Section 3.5. We used 10 surrogate datasets to demonstrate the ability of the proposed methodology to successfully infer the structure

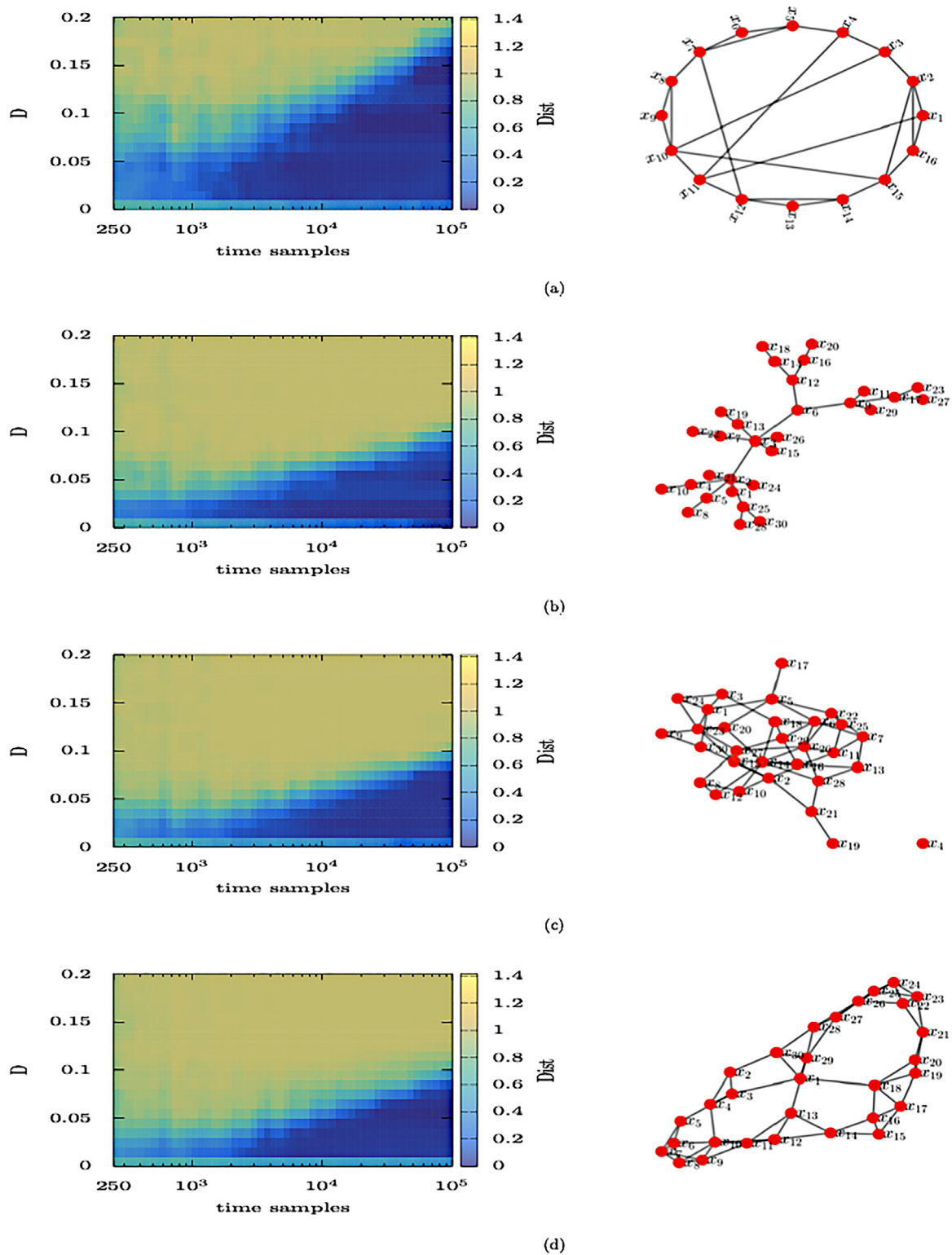


Fig. 6. Effect of noise strength D on the performance of the proposed methodology in: (A) the network of 16 nodes in Fig. 2(A), (B) a scale-free network of 30 nodes, (C) an Erdős-Rényi network of 30 nodes with rewiring probability 0.1, and (D) a small-world network of 30 nodes. Note that in all plots, the horizontal axes are in logarithmic scale and the vertical in linear. The colour bars encode the values of the Euclidean distances between the points on the ROC curves and $(0, 1)$ and take values in $[0, \sqrt{2}]$. A zero-distance denotes perfect inference, i.e., the point $(0, 1)$ on the ROC plane (not shown).

of the 4 networks for a small such number, which corresponds to $p^{sl} = 10^{-1}$. In this study, random surrogate datasets have been used as the metrics showed the original data are mainly noisy, what is expected from stochastic dynamics.

As discussed in Section 3.5, an appropriate probe for coupled stochastic Kuramoto phase oscillators is the instantaneous frequencies, which is used in this analysis. This is also backed by the fact that in experiments or systems in which only a projection of a multidimensional dynamics is available for monitoring, as in brain and climate networks, such projections seem to be a good choice [10].

In Fig. 6, for each pair of time-sample sizes (horizontal axes) and D values (vertical axes), the proposed methodology for network inference produces a point in the ROC plane by comparing the inferred network with the initial, shown in the left-hand-side plots. We compute the Euclidean distance, denoted by Dist in the figure, between that point and point $(0, 1)$ of perfect network inference. These distances are then encoded in the colour bars and lie in $[0, \sqrt{2}]$. The maximum distance $\sqrt{2}$ on the ROC plane corresponds to the lengths of the two diagonals. The smaller the distance, the better the resulting network inference, with perfect inference corresponding to a distance equal to 0, i.e., the adjacency matrices of the inferred and original connectivity matrices are equal.

We plot the Euclidean distance, Dist , for the four networks in Fig. 6. Panel (A) shows the results of the analysis for the network of 16 nodes. For small variance of the Wiener process, D , as the time-sample size increases, the method infers the structure of the network correctly (blue corresponds to Dist equal to 0). Expectedly, as D increases, larger and larger time-sample sizes are required for successful network inference (as seen in the right-hand-side) of the panel. For D values larger than about 0.2, even larger time-sample sizes would be required for the method to successfully infer the original network, a behaviour reminiscent of the results in [37].

Apparently, one can arrive at qualitatively similar conclusions when looking at the results in panels (B), (C) and (D) for the scale-free, Erdős-Rényi and small-world networks, respectively. What is different though, is the fact that for D larger than about 0.1, even larger time-sample sizes would be required for the method to successfully infer the initial connectivity matrix, again reminiscent of the results in [37].

It is worth it to mention here that scale-free networks, such as the one in panel (B), have heterogeneous topologies in their node-degree distributions, unlike the heterogeneity used in Section 3.7 where the connectivities are given by random weights to existing links [9]. In panel (B), for $D = 0$, the dynamics is deterministic, and the network is the Barabási scale-free network of 30 nodes in the right-hand-side in the same panel. Hence the 30 Kuramoto models are coupled through this network. As the number of time samples increases, the Euclidean distance, Dist , approaches 0 of successful inference, whereby all connections in the network have been inferred and no incorrectly inferred ones have been found. Particularly, for 10^5 time samples, the Euclidean distance, Dist , is found to be about 0.077. Hence one would expect the proposed method to work for scale-free networks and, deterministic, chaotic dynamics similarly to the results in Fig. 7 for random and small-world networks.

Concluding, here we demonstrated the ability of the proposed methodology to infer the structure of the original networks with stochastic dynamics as long as large enough time-series are available for large noise strength D , even if only 10 surrogate datasets are used. Based on the results in the previous subsections, had a smaller significance level been used, smaller datasets would be required for perfect network inference, or the performance of the method would be better for the same pairs of noise strengths and time-sample sizes.

3.7. Effect of coupling heterogeneity degree on the proposed inference method in networks of coupled circle maps

Finally, we study the performance of the proposed methodology in networks with varying coupling heterogeneity in their connections which is expected to affect the dynamics of the systems and recorded data [9]. We want to understand how successful the method is in inferring the initial connectivity matrix for Erdős-Rényi and small-world networks, where their connectivities are given by setting random weights to existing links.

Following [9], after fixing the initial connectivity matrix by producing a symmetric, binary adjacency matrix (A_{ij}) of $M = 16$ nodes with rewiring probability 0.3 for both networks, random weights are associated to each connection by considering a new symmetric adjacency matrix with entries

$$W_{ij} = A_{ij}(1 + g\xi_{ij}), \quad (11)$$

for $j > i$, where $0 \leq g < 1$ is the coupling heterogeneity degree parameter and ξ_{ij} an uncorrelated, zero-mean, uniformly distributed random number in $[-1, 1]$. Symmetry in A and W is imposed by considering $A_{ij} = A_{ji}$ and $W_{ij} = W_{ji}$, resulting in undirected connections.

The equation of motion of the i th unit is given by [66]

$$\dot{x}_{n+1}^i = (1 - \alpha)f(r_i, x_n^i) + \alpha \sum_{j=1}^M \frac{W_{ij}}{k_i} f(r_j, x_n^j), \quad (12)$$

where, the coupling strength $\alpha = 0.1$, $M = 16$, $f(r, x) = x + r - 1.1 \sin(2\pi x) \bmod 1$ (i.e., circle maps), r_i is the parameter of the i th circle map, W_{ij} is given by Eq. (11) and $k_i = \sum_{j=1}^M W_{ij}$ is the weighted degree of node i . The parameters r_i , α and the underlying topology (W_{ij}) determine the dynamics of the circle maps. Here, we set $r_i = r^* - \xi_i \delta r > 0$, where $r^* = 0.35$ corresponds to the chaotic regime of the circle maps, ξ_i is a random number uniformly distributed in $[0, 1]$

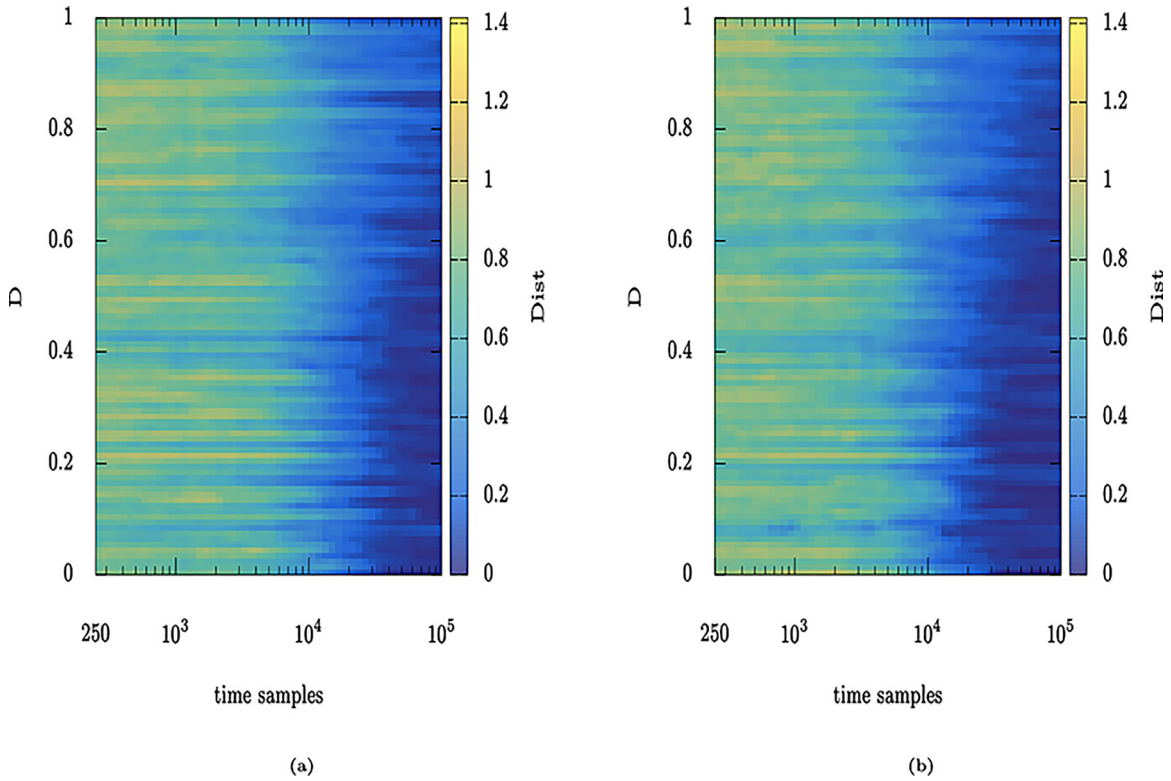


Fig. 7. Effect of coupling heterogeneity degree g on the performance of the proposed methodology in: (A) Erdős-Rényi networks and (B) small-world networks. In both panels, the networks consist of 16 nodes and the rewiring probability is 0.3. Note that all horizontal axes are in logarithmic scale and all vertical in linear. For each g , a different Erdős-Rényi or small-world network is used. The colour bars encode the values of the Euclidean distances, Dist , between the points on the ROC curves and $(0, 1)$, and take values in $[0, \sqrt{2}]$. A zero-distance denotes perfect inference, i.e., the point $(0, 1)$ on the ROC plane (not shown).

and $0 \leq \delta r < r^*$ the distribution width that controls the degree heterogeneity in the dynamics. In the following, we considered $\delta r = 0.34$ that corresponds to random values r_i in $[0.01, 0.35]$.

Next, we obtained the $M = 16$ time-series from Eq. (12) for 100, equally spaced values of g in $[0, 1]$. For each value, a different network of $M = 16$ nodes was used to couple the circle maps, in both cases of network types. Consequently, for each case, 100 different networks of 16 nodes were used, all with rewiring probability 0.3, associated to the 100 equally spaced g values in $[0, 1]$. We used 10 surrogate datasets to demonstrate the ability of the proposed methodology to successfully infer the structure of the 2 types of networks for a small such number, which corresponds to $p^{\text{sl}} = 10^{-1}$. Also, the metrics showed the original data are mainly noisy, what is expected from chaotic dynamics, thus random surrogate datasets were used.

The results of this analysis are shown in Fig. 7, where we report the effect of coupling heterogeneity degree g (see vertical axes in linear scale) as a function of the length of the time-series (see horizontal axes in logarithmic scale), on the performance of the proposed methodology for Erdős-Rényi networks in panel (A) and small-world networks in panel (B). For each pair of time-sample sizes (horizontal axes) and g values (vertical axes), the proposed methodology produced a point on the ROC plane by comparing the inferred network with the initial connectivity matrix (A_{ij}) . The Euclidean distance, Dist , was computed between that point and point $(0, 1)$. The distances are then encoded in the colour bars and lie in $[0, \sqrt{2}]$.

Our findings show that for both types of networks, the proposed methodology inferred successfully the initial connectivity matrix for data produced by dynamics with coupling heterogeneity as a function of time-sample sizes (see, the blue regions on the right hand-side of the plots which correspond to a distance very close to 0). Concluding, the performance of the proposed methodology is robust for coupling heterogeneity degree g as big as 1, for both, Erdős-Rényi and small-world networks.

4. Discussion

Complex systems are the subject of intense research in the last few decades, which has led to the emergence of complexity science. As they consist of many units which may interact in non-trivial ways, and whose aggregate behaviour

may be undetermined from the behaviour of the individual units, one can realise their structure by representing them as graphs made up of nodes and edges, i.e., as networks. Thus, recording the evolution of the dynamics on their nodes (i.e., as time-series) is key to inferring their connectivity, in the sense of inferring relations among the time-series, which amounts to finding whether links between pairs of nodes in the networks exist.

This highlights the importance of developing mathematical approaches that can infer network connectivity based on recorded data. Even though, network inference in nonlinear systems, physics, neuroscience, biology, economy, etc., has been studied extensively in recent years using correlation coefficients, cross-correlation, mutual information, Granger causality, rank-based connectivity measures, rank statistics, mutual prediction, phase transfer entropy, etc., it still presents open challenges! Possible reasons might be non-zero values resulting from similarity measures and, nonlinearities and asymmetries in the structure or noise level of the recorded dynamics.

Here, following recent advances in [37,38], we presented an information-theoretical methodology to infer connectivity in complex systems using time-series data in a pairwise fashion and statistical approaches. The methodology does not depend on the equations of motion, on the processes that produced the data and on the initial connectivity matrix. The only input required is the recorded data. As such, the equations of motions and initial connectivity matrices are only used to record the data and compare the inferred with the initial connectivity matrices by computing receiver operating characteristic curves. The presented method is based on: (a) estimations of the Mutual Information Rate for pairs of time-series and (b) on statistical significance tests for the acceptance of connectivity using the false discovery rate method for multiple hypothesis testing.

The Mutual Information Rate is the rate of information exchanged per unit of time between pairs of time-series and turns out to be an appropriate measure to quantify the exchange of information in systems with correlation. For example, a normalised version of the Mutual Information Rate has been shown to successfully infer the structure of networks in many different cases of toy-dynamical models [37] and, has also been used to infer financial and stock markets data [38]. In [37], it is shown how to calculate the Mutual Information Rate for Markov partitions, a challenging task as these partitions are difficult or impossible to find. In [9], two inference methods were compared based on cross-correlation and mutual information. It was shown that when an abrupt change in the ordered values of these measures exists, it is possible to infer correctly the underlying network topology from time-series data. These approaches can be used when well-defined gaps are present. However, they might not be able to infer the correct topology when gaps in the ordered values of the measures are absent or when more than one gaps appear. The presented methodology can be used when only recorded data are available and does not make use of thresholds: an inferred network will be computed based on the recorded data using pairwise computed Mutual Information Rate values and surrogate data to test the null hypotheses for connectivity using the false discovery rate method for multiple hypothesis testing.

The proposed methodology uses the bin or histogram method to compute probabilities for the estimation of the Mutual Information Rate. This is known to lead to overestimation of the Mutual Information Rate for random systems or non-Markovian partitions [30,41], as is the case in this work. The reason is the finite resolution of non-Markovian partitions and the finite length of recorded time-series. These errors are systematic and always present in the computation of the Mutual Information Rate for arbitrary non-Markovian partitions [30,41], e.g., when using the bin method. In [37], these errors were mitigated by double-normalising the computed Mutual Information Rate values. Here, we showed that this is actually not necessary. Even though, the non-normalised Mutual Information Rate values are overestimated, this is still enough to allow for network inference when combined with the use of significance tests and the false discovery rate method for the acceptance or rejection of connectivity. Comparisons between the MIR values of the dynamics of pairs of nodes can still be made. This is another reason that makes the proposed methodology a reasonable approach to use in network inference from artificial and experimental data.

5. Conclusions

In this paper, we presented a methodology that combines information-theoretical and statistical, significance tests using the false discovery rate method for multiple hypothesis testing to infer the structure of networks. We provided the mathematical background on information and network connectivity, and on Mutual Information and Mutual Information Rate. Then, we showed how to estimate the correlation-decay time used in the computations of the Mutual Information Rate and how to combine it with statistical significance tests and the false discovery method to accept those connections that result from the dynamics of the nodes. Following on, we have shown results of network inference for correlated normal-variates data and for coupled circle maps, coupled logistic maps, coupled Lorenz systems and coupled stochastic Kuramoto phase oscillators. Lastly, we have shown the effect of noise on the presented methodology in networks of coupled stochastic Kuramoto phase oscillators and of coupling heterogeneity degree on the presented methodology in networks of coupled circle maps.

We showed that the proposed methodology can infer the correct number and pairs of connected nodes, for correlated normal-variates data and, deterministic and stochastic dynamics. Thus, it can successfully infer the initial connectivity matrix. We also showed that the level of successful network inference improves by an order of magnitude as the significance level decreases by an order of magnitude. This demonstrates the potential of the method in studying scenarios and experimental cases. In the more realistic case of stochastic data, we demonstrated its ability to infer the structure of the initial connectivity matrices as long as large enough time-series are available. Interestingly, had a smaller significance

level been used, smaller datasets would be required for perfect network inference, or the performance of the method would be better for the same pairs of noise strengths and time-sample sizes. Moreover, we have shown that the method is able to recover the initial connectivity matrices for Erdős-Rényi and small-world networks with varying coupling heterogeneity in their connections, which is affecting the dynamics of the systems and thus, the datasets.

Finally, having shown the robust performance of the proposed methodology for different types of computer-generated datasets, dynamics and networks, it would be interesting to study in the future its performance for data coming from experiments or measurements, for example for electroencephalography and magnetoencephalography recordings, financial and stock market data, etc., where the equations of motion are unknown.

CRediT authorship contribution statement

Chris G. Antonopoulos: Conceptualisation, Investigation, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank Goh Yong Kheng (UTAR) for fruitful discussions and for providing the correlated normal-variables data used in the analysis. The author acknowledges the use of the High Performance Computing Facility (Ceres) and its associated support services at the University of Essex in the completion of this work.

References

- [1] Thurner S, Hanel R, Klimek P. Introduction to the theory of complex systems. Oxford University Press; 2018.
- [2] Sayama H. Introduction to the modeling and analysis of complex systems. SUNY Binghamton; 2015.
- [3] Barrat A, Barthelemy M, Vespignani A. Dynamical processes on complex networks. Cambridge University Press; 2008.
- [4] Niso G, Bruna R, Pereda E, Gutiérrez R, Bajo R, Maestú F, del Pozo F. HERMES: Towards an integrated toolbox to characterize functional and effective brain connectivity. *Rev Modern Phys* 2013;11:405–34.
- [5] Omony J. Biological network inference: A review of methods and assessment of tools and techniques. *Annu Res Rev Biol* 2014;4(4):577–601.
- [6] Sakkalis V. Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Comput Biol Med* 2011;41(12):1110–7.
- [7] Olejarczyk E, Marzetti L, Pizzella V, Zappasodi F. Comparison of connectivity analyses for resting state EEG data. *J Neural Eng* 2017;14(3):036017.
- [8] Jahanpour E, Chen X. Improving accuracy of complex network modeling using maximum likelihood estimation and expectation-maximization. *Interdiscip J Discontin, Nonlinearity, Complex* 2014;3:169–221.
- [9] Rubido N, Martí AC, Bianco-Martínez E, Grebogi C, Baptista MS, Masoller C. Exact detection of direct links in networks of interacting dynamical units. *New J Phys* 2014;16:093010.
- [10] Tirabassi G, Sevilla-Escoboza R, Buldú JM, Masoller C. Inferring the connectivity of coupled oscillators from time-series statistical similarity analysis. *Sci Rep* 2015;5:10829.
- [11] Butte A, Kohane IS. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Pac Symp Biocomput* 2000;5:415–26.
- [12] Zou Y, Romano MC, Thiel M, Marwan N, Kurths J. Inferring indirect coupling by means of recurrences. *Chaos* 2011;21(4):1099–111.
- [13] Nawrath J, Romano MC, Thiel M, Kiss IZ, Wickramasinghe M, Timmer J, Kurths J, Schelter B. Distinguishing direct from indirect interactions in oscillatory networks with multiple time scales. *Phys Rev Lett* 2010;104:038701.
- [14] Hempel S, Koseska A, Kurths J, Nikoloski Z. Inner composition alignment for inferring directed networks from short time series. *Phys Rev Lett* 2011;107:054101.
- [15] Ta XH, Yoon CN, Holm L, Han SK. Inferring the physical connectivity of complex networks from their functional dynamics. *BMC Sys Biol* 2010;4(70):1–12.
- [16] Timme M. Revealing network connectivity from response dynamics. *Phys Rev Lett* 2007;98:224101.
- [17] Shandilya SG, Timme M. Inferring network topology from complex dynamics. *New J Phys* 2011;13:013004.
- [18] Timme M, Casadiego J. Revealing networks from dynamics: An introduction. *J Phys A* 2014;47:343001.
- [19] Granger CWJ. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 1969;37:424–38.
- [20] Guo S, Ladrone C, Feng J. Granger causality: theory and applications. In: *Front. comput. sys. biol.*, vol. 15. 2010.
- [21] Bressler SL, Seth AK. Wiener-Granger causality: A well established methodology. *NeuroImage* 2010;58(2):323–9.
- [22] Barnett L, Seth AK. The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J Neurosci Methods* 2014;223:50–68.
- [23] Leguia MG, Martínez CGB, Malvestio I, Campo AT, Rocamora R, Levnajić Z, Andrzejak RG. Inferring directed networks using a rank-based connectivity measure. *Phys Rev E* 2019;99(012319):012319.
- [24] Chicharro D, Andrzejak RG. Reliable detection of directional couplings using rank statistics. *Phys Rev E* 2009;80:026217.
- [25] Schiff SJ, So P, Chang T, Burke RE, Sauer T. Detecting dynamical interdependence and generalized synchrony through mutual prediction in a neural ensemble. *Phys Rev E* 1996;54:6708.

- [26] Lobier M, Siebenhühner F, Palma S, Palva JM. Phase transfer entropy: A novel phase-based measure for directed connectivity in networks coupled by oscillatory interactions. *NeuroImage* 2014;85:853–72.
- [27] Alderisio F, Fiore G, di Bernardo M. Reconstructing the structure of directed and weighted networks of nonlinear oscillators. *Phys Rev E* 2017;95:042302. <http://dx.doi.org/10.1103/PhysRevE.95.042302>, URL <https://link.aps.org/doi/10.1103/PhysRevE.95.042302>.
- [28] Kralemann B, Pikovsky A, Rosenblum M. Reconstructing effective phase connectivity of oscillator networks from observations. *New J Phys* 2014;16(8):085013. <http://dx.doi.org/10.1088/1367-2630/16/8/085013>.
- [29] Yang C-L, Suh CS. A general framework for dynamic complex networks. *J Vib Test Syst Dyn* 2021;5(1):87.
- [30] Steuer R, Kurths J, Daub CO, Weise J, Selbig J. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics* 2002;18:S231–40.
- [31] Paluš M. From nonlinearity to causality: Statistical testing and inference of physical mechanisms underlying complex dynamics. *Contemp Phys* 2007;48(6):307–48.
- [32] Paluš M, Hartman D, Hlinka J, Vejmelka M. Discerning connectivity from dynamics in climate networks. *Nonlinear Proc Geophys* 2011;18:751–63.
- [33] Schreiber T. Measuring information transfer. *Phys Rev Lett* 2000;85(2):461–4.
- [34] Quiroga RQ, Arnhold J, Grassberger P. Learning driver-response relationships from synchronization patterns. *Phys Rev E* 2000;61:5142–8.
- [35] Smirnov DA, Bezruchko BP. Estimation of interaction strength and direction from short and noisy time series. *Phys Rev E* 2003;68:046209.
- [36] Nalatore H, Ding M, Rangarajan G. Mitigating the effects of measurement noise on Granger causality. *Phys Rev E* 2007;75:031123.
- [37] Bianco-Martínez E, Rubido N, Antonopoulos CG, Baptista MS. Successful network inference from time-series data using Mutual Information Rate. *Chaos* 2016;26(4):043102.
- [38] Goh YK, Hasim HM, Antonopoulos CG. Inference of financial networks using the normalised mutual information rate. *PLoS One* 2018;13(2):1–21.
- [39] Baptista MS, Rubinger RM, Viana ER, Sartorelli JC, Parlitz U, Grebogi C. Mutual information rate and bounds for it. *PLoS One* 2012;7(10):e46745.
- [40] Palus M. Coarse-grained entropy rates for characterization of complex time series. *Physica D* 1996;93:64–77.
- [41] Herzog H, Schmitt AO, Ebeling W. Finite sample effects in sequence analysis. *Chaos Solitons Fractals* 1994;4(1):97–113.
- [42] J. C. Erratum: Between order and chaos. *Nat Phys* 2013;9:382.
- [43] Paluš M. Coarse-grained entropy rates for characterization of complex time series. *Physica D* 1996;93(1–3):64–77.
- [44] Marchiori MA, Fariello R, de Aguiar MAM. Energy transfer dynamics and thermalization of two oscillators interacting via chaos. *Phys Rev E* 2012;85(4):041119.
- [45] Mandal D, Quan HT, Jarzynski C. Maxwell's refrigerator: An exactly solvable model. *Phys Rev Lett* 2013;111(3):030602.
- [46] Antonopoulos CG, Bianco-Martínez E, Baptista MS. Production and transfer of energy and information in Hamiltonian systems. *PLoS One* 2014;9(2):e89585.
- [47] Antonopoulos CG, Srivastava S, Pinto SES, Baptista MS. Do brain networks evolve by maximizing their information flow capacity? *PLOS Comput Biol* 2015;11(8):e1004372.
- [48] Shannon CE. A mathematical theory of communication. *Bell Labs Tech J* 1948;27(3):379–423.
- [49] Kullback S. *Information theory and statistics*. Dover Publications; 1997.
- [50] Eckmann JP, Ruelle D. Ergodic theory of chaos and strange attractors. *Rev Modern Phys* 1985;57(3):617.
- [51] Baptista MS, Carvalho JXD, Hussein MS. Finding quasi-optimal network topologies for information transmission in active networks. *PLoS One* 2008;3(10):e3479.
- [52] Pinto PRF, Baptista MS, Labouriau IS. Density of first , Poincaré returns, periodic orbits, and, Kolmogorov-Sinai entropy. *Commun Nonlinear Sci Numer Simul* 2011;16(2):863–75.
- [53] Moddemeijer R. On estimation of entropy and mutual information of continuous distributions. *Signal Proc* 1989;16(3):233–48.
- [54] Moon YI, Rajagopalan B, Lall U. Estimation of mutual information using kernel density estimators. *Phys Rev E* 1995;52(3):2318.
- [55] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E* 2004;69(6):066138.
- [56] Butte AJ, Kohane IS. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. In: *Pacific symposium on biocomputing*, vol. 5. World Scientific; 2000, p. 418–29.
- [57] Dobrushin RL. General formulation of Shannon's main theorem in information theory. *Am Math Soc Transl* 1963;33(2):323–438.
- [58] Gray RM, Kieffer JC. Asymptotically mean stationary measures. *Ann Probab* 1980;8(5):962–73.
- [59] Johnson DB. Efficient algorithms for shortest paths in sparse networks. *J ACM* 1997;24(1–3):1–13.
- [60] Prichard D, Theiler J. Generating surrogate data for time series with several simultaneously measured variables. *Phys Rev Lett* 1994;73(7):951–4.
- [61] Gómez-Gardeñes J, Zamora-López G, Moreno Y, Arenas A. From modular to centralized organization of synchronization in functional areas of the cat cerebral cortex. *PLoS One* 2010;5:e12313.
- [62] Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Statist* 2001;29:1165–88.
- [63] Junior LS, Mullokandov A, Kenett DY. Dependency relations among international stock market indices. *J Risk Financ Manag* 2015;8(2):227–65.
- [64] Kaneko K. Clustering, coding, switching, hierarchical ordering, and control in a network of chaotic elements. *Physica D* 1990;41(2):137–72.
- [65] Acebrón JA, Bonilla LL, Vicente CJP, Ritort F, Spigler R. The Kuramoto model: A simple paradigm for synchronization phenomena. *Rev Modern Phys* 2005;77:137.
- [66] Kaneko K. Overview of coupled map lattices. *Chaos* 1992;2(3):279–82.