Inferring the connectivity of coupled oscillators from event timing analysis

Raul P. Aristides[a,c], Hilda A. Cerdeira[a,b], Cristina Masoller[c], Giulio Tirabassi[c]

[a] *Instituto de Física Teórica, Universidade Estadual Paulista, R. Dr. Bento Teobaldo Ferraz, 271 - Várzea da Barra Funda, 01140-070, São Paulo, Brazil*

[b] *Epistemic, Gomez & Gomez Ltda. ME, Rua Paulo Franco 520, Vila Leopoldina, 05305-031, São Paulo, Brazil*

[c] *Departament de Fisica, Universitat Politecnica de Catalunya,, Rambla St. Nebridi, 22, Terrassa, 08222, Barcelona, Spain*

Abstract

Understanding the coupling structure of interacting systems is an important open problem, and many methods have been proposed to reconstruct a network from observed data. Most require continuous observation of the nodes' dynamics; however, in many situations, we can only monitor the times when some events occur (e.g., in neural systems, spike times). Here, we propose a method for network reconstruction based on the analysis of event times at the network's nodes. First, from the event times, we generate phase time series. Then, we assimilate the phase time series to the Kuramoto model by using the unscented Kalman filter (UKF) that returns the inferred coupling coefficients. Finally, we use a clustering algorithm to discriminate the coupling coefficients into two groups that we associate with existing and non-existing links. We demonstrate the method with synthetic data from small networks of Izhikevich neurons, where we analyze the spike times, and with experimental data from a larger network of chaotic electronic circuits, where the events are voltage threshold-crossings. We also compare the UKF with the performance of the cross-correlation (CC), and the mutual information (MI). We show that, for neural network reconstruction, UKF often outperforms CC and MI, while for electronic network reconstruction, UKF shows similar performance to MI, and both methods outperform CC. Altogether, our results suggest that when event times are the only information available, the UKF can give a good reconstruction of small networks. However, as the network size increases, the method becomes computationally demanding.

1. Introduction

Inferring the structure of a network from observed data is an important open challenge in complexity science, with multiple practical applications [1–7]. When using time series data, the main problems come from the fact that usually only one or a few variables can be observed (often not simultaneously), the observed variables can be recorded during a limited time interval, with limited temporal resolution, and with often unavoidable observational noise. Many network inference (or network reconstruction) methods have been proposed [8–27], whose success depends not only on prior knowledge of the system, but also on the characteristics of the data.

The Kalman filter is a well-known data assimilation technique to infer the parameters of a model given uncertain observations [28, 29], and therefore, knowing the model that describes the network, the Kalman filter can be used to infer the coupling coefficients.

In [30] two of us used a nonlinear version of the Kalman filter, the unscented Kalman filter (UKF) [31], to reconstruct a network of mutually coupled Rössler-type chaotic oscillators by analyzing, in each oscillator, the dynamics of only one of the three variables that define the phase space of the individual oscillators. We assumed that we knew the model that described the oscillators and all the parameters except the coupling coefficients. We demonstrated the methodology using simulated data because UKF was unable to provide a good reconstruction from the analysis of the experimental data. This was interpreted as due to the fact that we had incomplete infor-

mation about the model describing the experimental oscillators and/or about the filtering performed by the data acquisition system, which was not taken into account in the UKF model. That is, to reconstruct the network from the analysis of a single variable, observed experimentally in each node, we not only need a good knowledge of the model that describes the dynamics of the nodes, but we must also take into account the bandwidth of the data acquisition system, which generally does not detect the instantaneous values of the observed variables.

In a follow up study [32], we used UKF to reconstruct small networks of Izhikevich neurons, also assuming that we knew the model and parameters of individual neurons, except the coupling coefficients. However, here we assumed that we had complete information of the activity of all neurons (i.e., for each neuron, we knew the time evolution of the two variables that define the phase space of the neuron). Using this information, we demonstrated that UKF can infer the structural connectivity – that is, the adjacency matrix – even if the network is directed and evolves over time. Unlike bivariate link inference methods such as cross-correlation, CC, or mutual information, MI, the UKF method is multivariate because it simultaneously analyzes the activity of N neurons and returns the inferred $N \times N$ coupling coefficient matrix.

However, a typical experimental limitation is that only neuronal spikes can be detected. Furthermore, the model that describes neuronal activity is unknown. Therefore, here we aim to determine whether neural connectivity can be inferred if we only know the spike times.

This problem was addressed by Cestnik and Rosenblum [23], who considered pulse-coupled oscillators whose phases are instantaneously reset by incoming pulses. Using an iterative procedure to recover, from the inter-spike intervals (ISIs), the phase response curves (PRCs), the authors were able to recover all the properties of all the nodes, including the strengths of all connections. This approach has the advantage that the data requirements are not demanding (sequences of 200 spikes allowed to reconstruct a network of 20 Morris-Lecar neurons), but it assumes that the coupling between neurons is mediated only by the pulses, and that it is sufficiently weak to justify the description of the neural activity in terms of PRCs. Another network inference method based on the analysis of spike times was proposed by Casadiego and coworkers [24]. By approximating the ISIs of each neuron i by an unknown, locally smooth function of K_i cross-spike intervals, the inference problem was mapped to event spaces yielding linear equations that allowed least squares solutions for the topology.

Machine learning (ML) network inference methods have also been proposed. Panaggio and coworkers [25] considered networks of phase oscillators and showed that ML can reconstruct the interaction network, given sufficient data on the transient evolution of each oscillator.

However, data limitations degrade the performance of ML algorithms. Banerjee and coworkers [26] considered the performance of reservoir computing, on synthetic data simulated with coupled Lorenz chaotic systems and on experimental data of *C. elegans* neural activity, with known structural network. The authors assigned scores that reflected the link probability between pairs of nodes and studied how various limitations of data acquisition, such as observational noise, affected the scores.

Complementing these efforts, here we analyze the situation in which the only information available for reconstructing the network are event times (e.g., spikes times). To use the UKF method, we need to assume a model and a straightforward choice is the Kuramoto model of coupled phase oscillators [33]. While we could use the coupling coefficients inferred by UKF to define scores that reflect the link probability (as in [26]), here we reconstruct the network by applying a clustering algorithm to the inferred coupling coefficients, to discriminate them into two groups that correspond to existing and non existing links (as in [32]). This allows us to quantify the reconstruction performance by calculating the F_1 score and the area under the receiver operating characteristic curve (ROC).

To demonstrate this methodology, we analyze neural activity simulated with the Izhikevich model [34], as well as experimental data recorded from coupled Rössler-like chaotic electronic circuits with known structural network [35]. To reconstruct the circuits' network, we assume that the only information we have are event times –specifically, the times when the circuits' voltages change signs, from negative to positive values.

2. Models, datasets and methods

2.1. Izhikevich model

To generate the spikes data, we simulate a neural network using the Izhikevich model, a neuron model capable

Table 1: Dimensionless parameters used to simulate the Izhikevich model.

a	b	c	d	I	K	σ_Z
0.2	2	−56	−16	−99	(0, 0.1)	0.02

of exhibiting many neuronal dynamics while being computationally efficient [34]. Letting $[x_1, y_1, \ldots, x_N, y_N]^T$ be the state vector describing the state of the neural network, the equations governing the system are

$$\begin{aligned} \dot{x}_i &= 0.04\, x_i^2 + 5x_i + 140 - y_i + I + E_i + \sigma_Z \xi_i^x, \\ \dot{y}_i &= a\,(b\, x_i - y_i) + \sigma_Z \xi_i^y, \end{aligned} \tag{1}$$

with the after-spike reset condition:

$$\text{if} \quad x_i > 30, \quad \text{then} \begin{cases} x_i \to c, \\ y_i \to y_i + d. \end{cases} \tag{2}$$

Here x_i represents the voltage and y_i the membrane recovery variable accounting for the activation of the ionic currents of the i-th neuron. For the constant parameters a, b, c, d, and I, specified in Table 1, we selected values such that the neurons display chaotic spiking dynamics [36]. E_i represents the electrical coupling between neurons and is given by

$$E_i = K \sum_{j=1}^{N} A_{ij}(x_j - x_i), \tag{3}$$

where K is the coupling conductance and A_{ij} are the coefficients of the adjacency matrix: $A_{ij} = 1$ whenever neuron i is connected to neuron j, otherwise $A_{ij} = 0$. Since electric coupling is symmetric, the adjacency matrix is symmetric, $A_{ij} = A_{ji}$. Finally, ξ_i^x and ξ_i^y represent Gaussian noise terms, and σ_Z is the square root of the noise variance.

We solved the model equations using a fourth-order Runge-Kutta method with an integration step of $dt = 0.01$ and initial conditions drawn from a normal distribution centered at $x = -56.25$ and $y = -112.5$, with standard deviation $\sigma = 3$. A transient of 80000 time steps was discarded, and the times of the spikes that occurred in the following 40000 time steps were saved for analysis. The spike times were defined as the times in which the variable x reaches the reset condition, Eq. (2), and were linearly interpolated such that $t_i : x(t_i) = 30$.

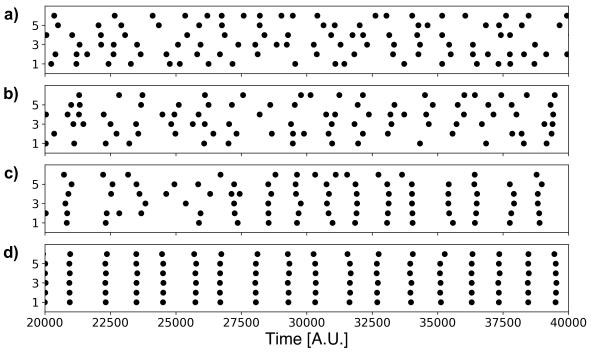

Figure 1: Spiking activity of a random network of 6 Izhikevich neurons and 8 links when the neurons are (a) uncoupled ($K = 0$), (b) weakly coupled ($K = 0.01$), (c), (d) strongly coupled (K =0.04 and 0.1 respectively).

Figure 1 displays an example of the spikes of a random network with 6 neurons and 8 links, for different coupling strengths. As can be observed, as the coupling increases, the spikes tend to become synchronized.

2.2. Chaotic electronic circuits

We also analyze a freely available experimental data set [35] that contains measurements from a network of $N = 28$ chaotic (Rössler-like) mutually coupled electronic circuits; the circuits are coupled with 42 links whose strength (the same for all the links) varies from 0 to 100. For each coupling strength, the voltage in each circuit was recorded three times. Each time, 30000 data points were recorded. We selected the data set recorded for the network $R7$, which is shown in Fig. 2. For our analysis, we used the following set of coupling strengths: {0, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51}; for stronger coupling the circuits become phase synchronized and their phase dynamics does not change significantly from $K = 51$.

We show in Fig. 3 the time evolution of the 28 oscillators for different coupling strengths. As for the Izhikevich neurons, we consider different levels of synchronization, up to a phase synchronized state.

We define events by considering the crossings of the $x = 0$ axis from $x < 0$. As in the neurons' case, the crossing times are linearly interpolated.

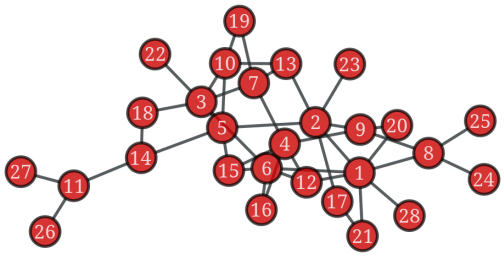

Figure 2: Network $R7$ of electronic circuits [35] that has $N = 28$ circuits (nodes) and 42 undirected links.

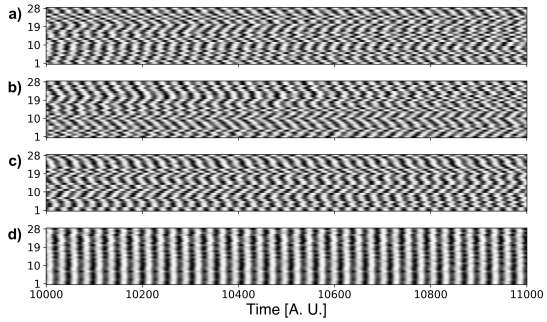

Figure 3: Oscillatory activity of 28 electronic circuits (in color code, experimentally recorded x variable) when the circuits are (a) uncoupled ($K = 0$), (b) weakly coupled ($K = 1$) and (c, d) strongly coupled ($K = 10$ and $K = 50$ respectively).

2.3. Phase description

From now on, we consider that the event times are the only data available to reconstruct the network, that is, for the neurons, we only know the spike times, and for the electronic circuits, we only know the times of $x = 0$ crossings. Interpreting each neuron/circuit as a phase oscillator, we associate a phase time series to each neuron/circuit by assuming that the phase increases linearly by 2π between consecutive spikes/crossings [37].

An example of the procedure is shown in Fig. 4. Regarding time resolution of the phases, we decided to use the model integration step for the neurons while, for the electronic circuits, we downsampled the original time-series to 1/5 of the measurement frequency, as we have seen that this has little impact on the accuracy but, at the same time, it speeds up the calculation significantly. Since the Kalman filter is designed to assimilate noisy observations, we added to the piece-wise linear phases a stochastic term as a Gaussian white noise with variance σ_ϕ. In the analysis hereby we fixed $\sigma_\phi = 0.12$. We tested different values for σ_ϕ and found that the UKF performance is not significantly impacted if the noise intensity is not too large. While further studies are needed to determine the role of this parameter, our preliminary analysis suggests that a large noise variance might decrease the UKF performance.

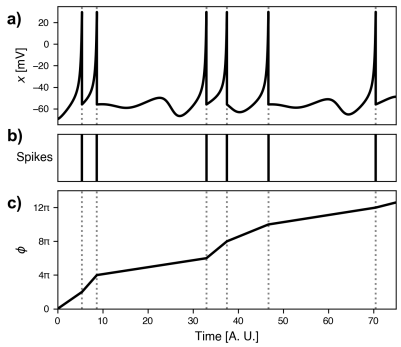

Figure 4: Example of a phase time series, $\phi(t)$, obtained from a voltage time series, $x(t)$. (a) $x(t)$. (b) Spikes occur when x reaches the reset condition ($x > 30$mV). (c) For each spike time, t_n, we define the phase as $\phi(t_n) = 2n\pi$ and interpolate linearly between consecutive spikes times.

2.4. UKF assimilation to the Kuramoto model

We assume that the phases' dynamics can be described, at least partially, in terms of a phase reduction equation of the form

$$\dot{\phi}_i = \bar{\omega} + \sum_{j=1}^{N} m_{ij} F(\phi_j, \phi_i) + \sigma_i \xi_i, \tag{4}$$

with F being a periodic function of ϕ_i and ϕ_j. A simple option for F is $F(\phi_j, \phi_i) = \sin(\phi_j - \phi_i)$, which gives the well-known Kuramoto model [33]:

$$\dot{\phi}_i = \bar{\omega} + \sum_{j=1}^{N} m_{ij} \sin(\phi_j - \phi_i) + \sigma_i \xi_i, \tag{5}$$

where $\bar{\omega}$ is the natural frequency of the oscillators, which is assumed to be the same for all oscillators, $m_{ij} = m_{ji}$ are symmetric coupling coefficients, ξ_i is a Gaussian noise

term, and σ_i is the noise standard deviation, representing, in the context of UKF, the uncertainty in the model used to fit the data.

The Kuramoto model successfully describes coupled phase oscillators; however, we remark that our approach is flexible because another, more advanced model could be used. For instance, in Eq.(5), non-instantaneous, lagged interactions could have been considered; however, this has the drawback that additional parameters need to be estimated.

UKF is able to estimate the state of N phase oscillators, denoted by $\bar{u} = [\phi_1; ...; \phi_N]^T$, by assimilating N phase time series (defined from the spikes or event times) into the Kuramoto model. Since our goal is to retrieve the oscillators' connectivity, we extend the state of the system to include the $N(N-1)/2$ coupling parameters, m_{ij}. Therefore, the state vector, $\bar{u} = [\phi_1; ...; \phi_N; m_{1,2}; ...; m_{N-1,N}]^T$, is a vector of dimension $M = N + N(N-1)/2$. Since the network structure remains constant in time, the equations governing the evolution of the coupling coefficients are $dm_{ij}/dt = 0$.

The natural frequency of the oscillators, $\bar{\omega}$, can be estimated as the average frequency of the oscillators when they are uncoupled. However, this information might not always be available; therefore, to demonstrate the general applicability of our methodology, we estimate $\bar{\omega}$ for every coupling strength, K, directly from the phases' time series as the average frequency of the oscillators. A comparison of different ways to estimate $\bar{\omega}$ is presented in *Appendix A*.

We implemented the UKF algorithm using the *Python* package *FilterPy*, with parameters given in Table 2. To initialize the algorithm, we have to choose initial conditions for the state vector $\bar{u}$. For the phases, we used the values at $t = 0$ while for the coupling coefficients, we set them all to an initial value K_0. The covariance matrices associated with the process and measurement processes, which are also evolved by the UKF algorithm, were initialized with the standard deviations σ_Z and σ_ν. The covariance estimate matrix was initialized as $\mathbb{I}\sigma_P$, where $\mathbb{I}$ is an $M \times M$ identity matrix.

The UKF provides an estimation for ϕ_i, σ_i, and m_{ij} at each time. An example of the evolution of the inferred coupling coefficients is presented in Fig. 5. Since the inference at the last available time integrates the information of all the other time steps, we consider the last estimation of the m_{ij} coefficients as the UKF estimation of the coefficients.

Table 2: Parameters used in the UKF algorithm [31, 38] for assimilating the phase time series of the neurons (of the chaotic circuits) to the Kuramoto model.

K_0	σ_Z	σ_ν	σ_P	α
0.1(0.05)	0.01(0.1)	0.05(0.15)	0.05(0.01)	0.0001

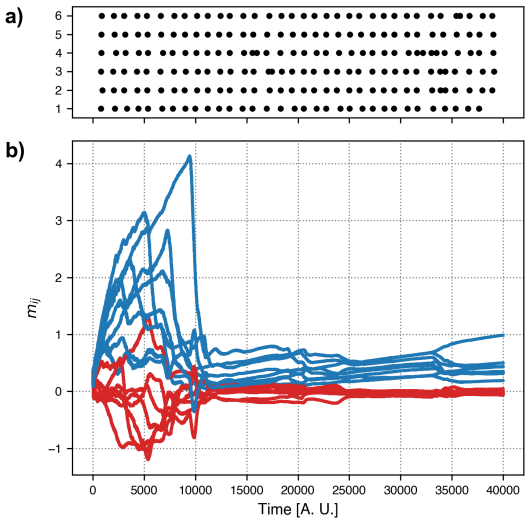

Figure 5: (a) Spike times of a random network of $N = 6$ Izhikevich neurons with 8 links and coupling $K = 0.04$. (b) Inferred coupling coefficients using the UKF. The blue lines represent the coefficients of existing links and the red lines, those of non-existing links.

We remark that the UKF approach is multivariate, in the sense that the algorithm uses N $\phi_i(t)$ time series to infer at once the m_{ij} coefficients.

2.5. Network inference

To transform the set of m_{ij} values returned by UKF into an inferred binary adjacency matrix, we used a k-means clustering algorithm [39, 40] that classifies the coefficients in three clusters. Other choices are possible, but we found that three clusters is the minimum number that gives good performance. The clustering algorithm puts low coupling values in the first cluster, high values in the second, and outliers in the third. The first cluster is interpreted as composed by non-existing links (zeros in the

adjacency matrix), while the other two clusters are interpreted as composed by existing links (ones in the adjacency matrix). Alternative approaches can be used to binarize the coefficients m_{ij}, which include, for instance, using a threshold defined in terms of the mean value and the standard deviation of the m_{ij} values.

2.6. Performance quantification

To evaluate the performance of the network reconstruction, for the neuron network we will use the F_1 score [41],

$$F_1 = \frac{2TP}{2TP + FN + FP}, \tag{6}$$

where TP is the number of true positives, FP the number of false positives, and FN the number of false negatives. The F_1 score represents the harmonic average between two other scores: *precision*, which is the fraction of predicted links that are real links, and *recall*, which is the fraction of real links recovered by the algorithm. Therefore, a perfect reconstruction will have $F_1 = 1$. Moreover, being a harmonic average, a low value of one score (precision or recall) will give a low F_1, regardless of the value of the other score.

In the case of the 28 electronic circuits, we will evaluate the performance in terms of the area under the receiver operating characteristic curve (AUC). We use the AUC instead of F_1 because the clustering algorithm does not provide a good differentiation of the coefficients into existing and non-existing links. An AUC= 1 means that there is a threshold that can perfectly classify the inferred coefficients into high coefficients, representing links, and low coefficients representing non-existing links. An AUC= 0.5, instead, represents the expected performance of a random classifier.

We compare the UKF performance with the performance of two bivariate link inference methods: the Pearson cross-correlation coefficient (CC) and the Mutual Information (MI) [42]. We calculate CC and MI from the phase times series, because we assume that we only know the event times, from which we derive phase time series. CC_{ij} is calculated as:

$$CC_{ij} = \frac{1}{T}\left|\sum_{t=0}^{T} a_i(t)a_j(t)\right| \tag{7}$$

where $a_i(t)$ and $a_j(t)$ are the detrended and normalized (to zero mean and unit variance) phase time series of oscillators i as j, and T is the length of the time series. MI_{ij} is calculated as

$$MI_{ij} = \iint p_{ij}(a_i, a_j) \log \frac{p_{ij}(a_i, a_j)}{p_i(a_i)p_j(a_j)}\, \mathrm{d}a_i \mathrm{d}a_j, \tag{8}$$

where the probability distributions p_i and p_j are derived from $a_i(t)$ and $a_j(t)$, and p_{ij} is the joint distribution.

For simplicity, in Eqs. (7) and (8) we neglect a possible lag between a_i and a_j; taking into account a lag may improve the reconstruction for weak coupling, while no significant effect is expected at strong coupling because the oscillators synchronize with zero lag.

The clustering procedure described in Sec. 2.5 for inferring the network from the m_{ij} coefficients returned by UKF, will also be applied to the $\{CC_{ij}\}$ and $\{MI_{ij}\}$ values.

3. Results

3.1. Izhikevich neurons

The results obtained by applying the network inference algorithms based on UKF, MI, and CC are displayed in Fig. 6 as a function of the coupling strength K. Here, for each K, we simulated 15 random networks with 6 neurons and 8 links.

An F_1 value was calculated for each reconstructed network, and we used the 15 F_1 values to compute the F_1 statistics.

We can observe that, for weak and intermediate coupling strengths, UKF tends to outperform the bivariate inference based on MI or CC, whereas the three methods perform similarly for strong coupling. An inspection of the Kuramoto order parameter, $R = \langle|\frac{1}{N}\sum_k e^{i\phi_k(t)}|\rangle_t$, displayed in Fig. 7 reveals that the performance of UKF improves with the coupling strength until the phases become partially synchronized at about $K \simeq 0.04$, and afterward the performance fluctuates around $F_1 \simeq 0.8$. The synchronization of the phases is never perfect ($R < 1$) and it is not stationary, as shown by the shaded area representing the interquartile range. This allows the algorithm to keep a good performance even at relatively high K values [43].

We also note a large dispersion of F_1 values for the three inference methods, regardless of the coupling strength. Perfect reconstruction of the adjacency matrix

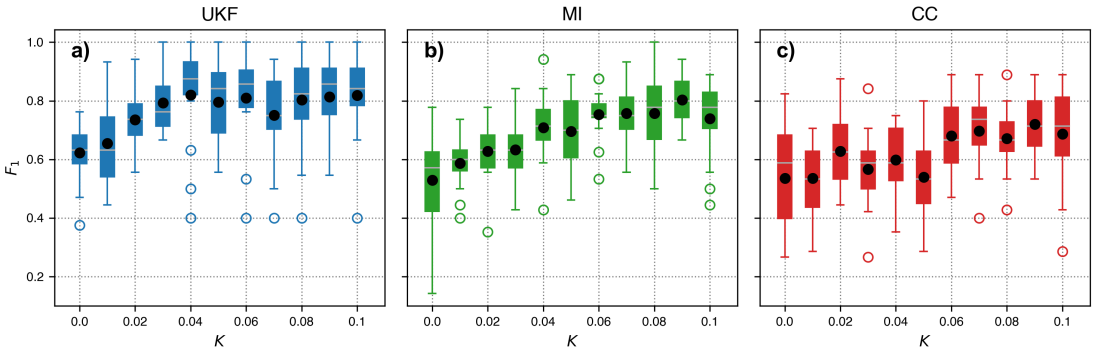

Figure 6: Performance of the inference methods based on UKF, *MI*, and *CC* as a function of the coupling strength, K, for a network of Izhikevich neurons. Box-plot statistics are obtained by inferring 15 different network topologies of 6 nodes and 8 links. Black circles represent the average score for each coupling coefficient.

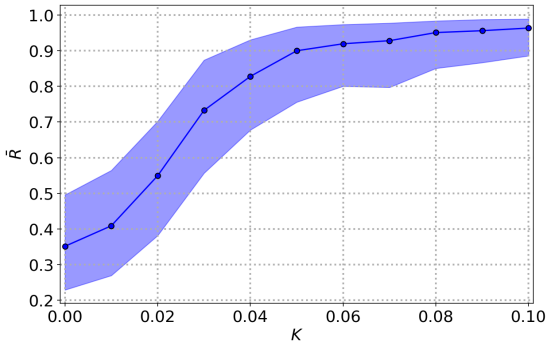

Figure 7: Kuramoto order parameter of the phases of the Izhikevich neurons vs. the coupling strength. Blue shading represents the interquartile range while the blue markers represent the median; the parameters are as in Fig. 6.

is, in general, not possible, as the F_1 is usually less than 1, except for a few perfect reconstructions obtained with the UKF algorithm.

To try to improve the performance of the inference methods, we simulated the model a number of times, s, with different initial conditions but the same adjacency matrix. Alternatively, we could simulate long enough time series and divide them into s non-overlapping segments.

From each set k ($k = 1 \dots s$) of spike sequences we derived the phase time series and applied to them the UKF algorithm, which returned the m_{ij}^k coefficients of the kth set, and we also calculated the corresponding CC_{ij}^k and MI_{ij}^k values. Then, the network was reconstructed by applying the same clustering procedure as before, but to $\langle m_{ij} \rangle_s$, $\langle CC_{ij} \rangle_s$ and $\langle MI_{ij} \rangle_s$, where $\langle \dots \rangle_s = (\sum_{k=1}^{s} \dots)/s$ represents the average over s simulations. Then, an F_1 value was calculated for each reconstructed network. Repeating the procedure with 15 different coupling topologies, we obtained 15 F_1 values that were used to compute the F_1 statistics.

The results are presented in Fig. 8, where the top (bottom) row presents the F_1 statistics for $s = 36$ ($s = 99$). We can see that averaging the inferred coefficients improves the UKF performance, whereas the performance of *MI* is only slightly improved, and the *CC* performance is left almost unaffected. However, despite the overall good performance, UKF presents some negative outliers, even after averaging $s = 99$ inferred coefficients, while *MI* and *CC* seem to be more robust. Examples of good and bad UKF reconstructions are presented in Fig. 9, where panel a) presents the real network, b) presents a good reconstruction ($F_1 = 0.933$), and c) presents a poor reconstruction ($F_1 = 0.5$).

We have analyzed the reasons for the negative UKF outliers and found that the m_{ij} coefficients are correctly

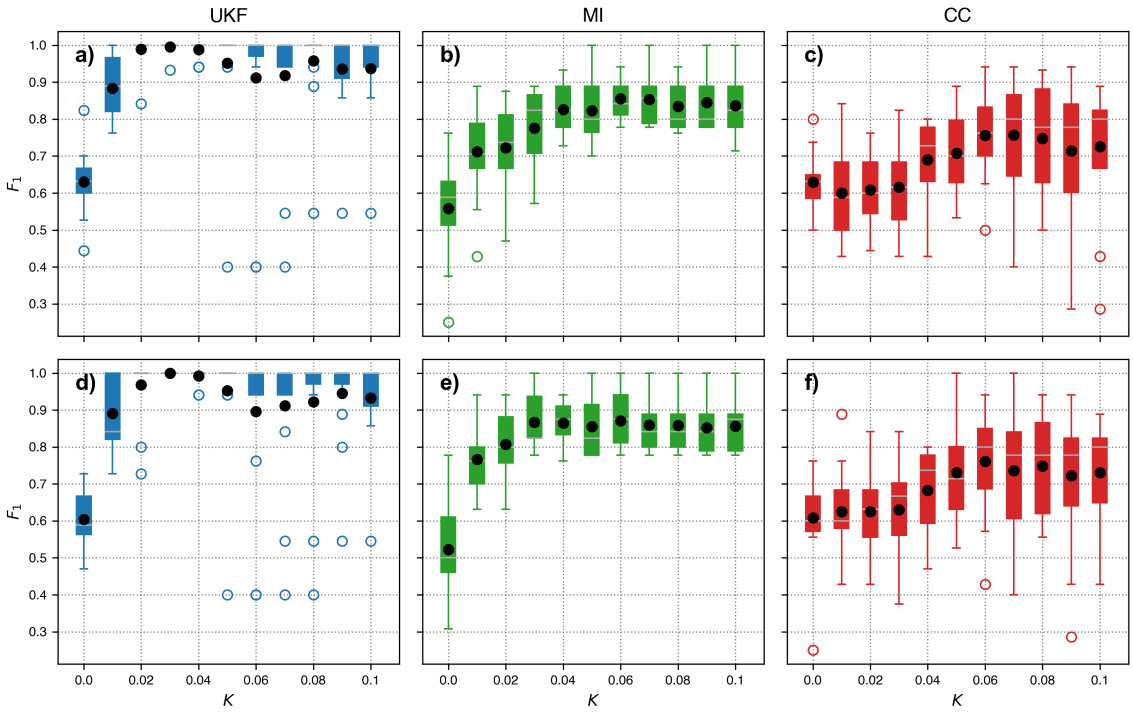

Figure 8: As Fig. 6 but the network is inferred from $\langle m_{ij} \rangle_s$, $\langle MI_{ij} \rangle_s$ and $\langle CC_{ij} \rangle_s$ averaged over $s = 36$ (a, b, c) or $s = 99$ (d, e, f) simulations.

ordered (the highest coefficients corresponding to existing links and the lowest coefficients, to missing links), but the clustering algorithm does not find the right threshold to separate them. As seen in Fig. 9(d), that displays the Kuramoto order parameter (red line) together with the degree of phase synchronization between three pairs of nodes (black lines show $R_{jk} = |\exp(i(\phi_j - \phi_k)|$) is above the average of the network. In this situation, the clustering algorithm puts these links in the "1" class and all the other links in the "0" class, which results in a low F_1 score. For the specific network topologies in which this occurs, UKF inference does not benefit from the averaging procedure. While more work is needed to identify better clustering strategies, we stress that UKF performance is low for high enough coupling; for weaker coupling the reconstruction can even be perfect, as seen in Fig. 8 (d), where for $K = 0.03$ the 15 networks are perfectly reconstructed (for $K = 0.03$ and $s = 99$, UKF $F_1 = 1$).

It is important to note that, even though a general performance improvement is expected from the averaging procedure, as using more simulations provides more data to infer the links, the efficiency with which the three methods leverage the new information is clearly different. We speculate that the statistical similarity of the neurons' time series across various time series (or segments) is high, leading to high correlations between the MI-based and CC-based inferences obtained from different simulations, leading to a small improvement or even no improvement of their performance. On the other hand, providing a model such as the Kuramoto model to the UKF allows us to extract more information from small datasets, e. g. discriminating small coupling from no coupling. These arguments are of course heuristics and more work is needed to understand the effect of the averaging procedure.

We compared the UKF performance when different ways of estimating the natural frequency of the oscilla-

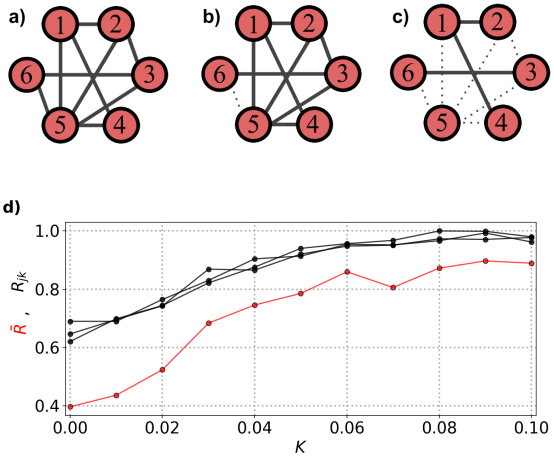

Figure 9: Examples of good (b) and poor (c) UKF reconstruction of the network shown in (a). In (b) and (c) the dashed lines indicate links that were not inferred (false negatives). (d) Kuramoto order parameter (red) vs. the coupling strength, and phase synchronization degree, $R_{jk} = |\exp(i(\phi_j - \phi_k)|$, for three pairs of nodes (1-2, 1-4, and 3-6) in black which exhibit a higher degree of phase synchronization.

tors, $\bar{\omega}$, were used (see *Appendix A*). We found that when the dynamics of the oscillators is highly regular (the spike/crossing times are almost periodic), setting $\bar{\omega} = 0$ gives the best performance. In addition, we analyzed the effect of the network size and link density, the results are presented in *Appendix B*. As could be expected, for all methods, increasing the network size decreases the performance. On the other hand, increasing the link density facilitates the inference. Overall, UKF is the most precise method (see Figs. A.14 and A.15 in *Appendix B*). The effect of the time series length is discussed in *Appendix C*, where we show that, as expected, the performance of the three methods decreases for shorter time series, but still UKF yields the best performance. Finally, the computational cost of the algorithms as a function of the network size is analyzed in *Appendix D*. Inevitably, the UKF has the largest computational costs, and it increases rapidly with the size of the network. This comes as no surprise since the UKF is a multivariate estimator, while *MI* and *CC* are bivariate measures.

Taken together, the results of the analysis of synthetic spike sequences generated with the Izhikevich model indicate that UKF is the optimal choice to reconstruct small networks of weakly coupled neurons.

3.2. Chaotic electronic circuits

In this section, we present the results obtained for the reconstruction of the network of 28 chaotic, Rössler-like electronic circuits. As the time series have 30000 data points (much longer than the length needed for UKF to produce a stable inference), we divided the time series into six non-overlapping segments so that each one contains approximately 40 $x = 0$ crossing events, as in the neurons' case. Since, for each coupling strength, three time series are available, we have 18 segments to analyze, and therefore, we have 18 AUC values to compute the score statistics. As explained in Sec. 2.6, we quantify performance in terms of the AUC rather than F_1 because the clustering algorithm does not provide good differentiation of the coefficients between existing and non-existing links and this occurs not only for the m_{ij} set, but also, for the MI_{ij} and CC_{ij} sets.

As in the neurons' case, we can also use the 18 segments to obtain 18 estimations of the values of m_{ij}, MI_{ij} and CC_{ij}, and we can average them to obtain the final estimations, i.e. to obtain the matrices $\langle m_{ij}\rangle_s$, $\langle MI_{ij}\rangle_s$ and $\langle CC_{ij}\rangle_s$ that are binarized to reconstruct the network.

Figure 10 displays the performance of UKF, *MI*, and *CC* in the reconstruction of the network. For UKF reconstruction, we have found that when the oscillators' dynamics is highly regular, i.e., when the timing of $x = 0$ crossings is nearly periodic, the best UKF performance is obtained by setting $\bar{\omega} = 0$ (the influence of the estimation of $\bar{\omega}$ is discussed in *Appendix A*). This occurs when $K \geq 13$ and, as it can be seen in Fig. 11 above this coupling strength there is a large increase in phase synchronization of the oscillators. Comparing Fig. 11 with Fig. 7, which depicts the synchronization parameter for the neurons, we notice that for the circuits the increase in synchronization is accompanied by a decrease of the temporal variance of the values of R, represented by the shaded area. This reveals not only an increased phase synchronization but also a small variance in time of the circuits' phase dynamics. In contrast, for the phases of the Izhikevich neurons, the increase of R with the coupling strength is more gradual, and even for high values of R, its variance remains high. An inspection of the time series of the neurons' phases reveals the existence of time intervals

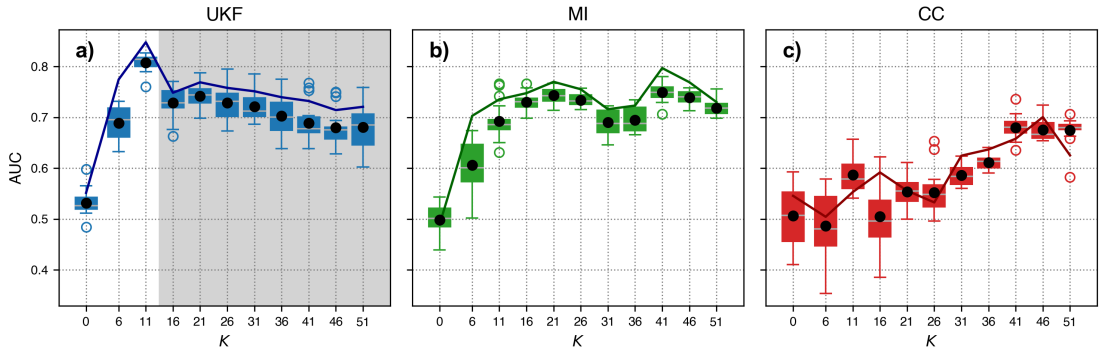

Figure 10: Performance of UKF, *MI*, and *CC* inference methods for the network of 28 chaotic electronic circuits. The horizontal axis displays the coupling strength in arbitrary units. Box-plot statistics are obtained from 18 data segments and the black circles indicate the average score. The solid lines represent the AUC when the network is inferred from $\langle m_{ij} \rangle_s$, $\langle MI_{ij} \rangle_s$ and $\langle CC_{ij} \rangle_s$, where the average is over $s = 18$ non-overlapping segments. The shaded area in panel a) represents the region where the oscillators have nearly identical frequencies and the best UKF performance is obtained by selecting $\bar{\omega} = 0$.

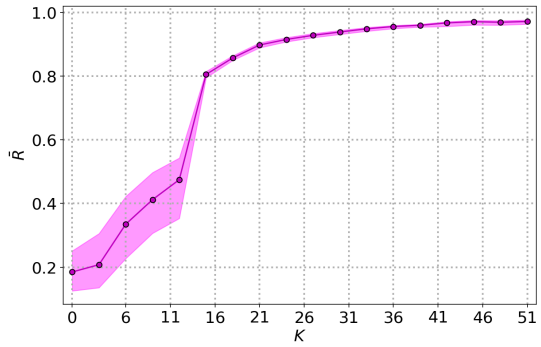

Figure 11: Kuramoto order parameter of the phases of the electronic circuits vs. the coupling strength in arbitrary units. Magenta shading represents the inter-quartile range while the magenta markers represent the median.

where the neurons lose synchronization. We believe that this difference may be, at least in part, responsible for the different performance of UKF in the two systems.

We can see in Fig. 10 that the averaging procedure improves the performance of both UKF and *MI*, while it does not affect the *CC* performance, which remains poor.

We also note that *MI*'s performance is comparable to UKF's, but UKF performs slightly better for low coupling. Specifically, UKF's best performance is at $K = 11$, and for this coupling strength, the best binarization threshold gives, for UKF, $F_1 = 0.416$, while for *MI*, $F_1 = 0.281$. On the other hand, *MI*'s best performance is at $K = 41$, and for this K, the best binarization threshold gives, for UKF, $F_1 = 0.344$, while for *MI*, $F_1 = 0.407$.

At high enough coupling the better performance of *MI* with respect to UKF may either be because the approximation of the function $F(\phi_j, \phi_i)$ in Eq. (4) by $F(\phi_j, \phi_i) = \sin(\phi_j - \phi_i)$ is not appropriated (as discussed in *Appendix A)* or because the phase description is not sufficiently precise and amplitude effects need to be taken into account. To check this point, we inspected the shape of the reconstructed attractor of an electronic circuit (using Takens embedding, as only one of the three variables is available), as a function of the coupling strength, and found that it continues to be phase coherent in the range of coupling strengths analyzed, and approaches the shape of a limit cycle for high couplings. Because we did not observe relevant variations in the amplitude of the oscillators, we believe that the phase description is still valid, but the simple approximation of the function F is not.

The good performance of the UKF in terms of AUC means that the inferred coefficients have the right relative magnitude: on average, those corresponding to existing links are larger than those corresponding to non-existing links. However, further studies are needed to determine

the optimal strategy to choose an appropriate threshold to classify the $\langle m_{ij} \rangle_s$ values in two distinct populations that correlate as much as possible with the adjacency matrix.

4. Conclusions

Understanding the coupling structure of interacting systems is an important open problem and many methods have been proposed to infer underlying interactions, from the analysis of observed data. However, most methods require model information or a good knowledge of the dynamics of the elements of the system. Here we have proposed a model-free inference method that assumes that the only information available are the times when events such as spikes or threshold-crossings occur. We have demonstrated the UKF-based method with simulated neuronal spikes and with experimental data recorded from chaotic electronic circuits, and compared its performance with the performance of the mutual information, and of the cross-correlation. Specifically, we have considered neural spike times simulated with the Izhikevich model, and experimental data recorded from chaotic electronic circuits, consisting of the times when the circuits' voltages change sign, from negative to positive values. In both cases, we have used the event times (spikes or threshold crossings) to obtain phase time series, and used UKF to assimilate the phases' time series to the Kuramoto model.

On synthetic data from small neural networks, we have found that averaging the UKF prediction over multiple realizations can allow obtaining a perfect reconstruction of the coupling topology. This averaging procedure also benefits the inference when using *MI* and *CC*, but their performance is in general lower than that of UKF.

In the experimental data, no method could fully recover the network. We have found a comparable performance of UKF and *MI*, with *CC*'s performance being significantly lower. Both, UKF and *MI* benefit from the averaging procedure before inferring the network, with UKF performing better at weak coupling and *MI* performing better at higher coupling.

The downside is that the improved performance of UFK comes with a high computational cost, which makes it unfeasible to use in the case of large networks.

We believe that the UKF method proposed here can be very useful for inferring the topology of small networks of coupled oscillators, when the oscillators can be approximately described as phase oscillators, because in this case UKF may provide better performance than *MI* or *CC*.

Future work will be devoted to searching for a theoretical foundation to better understand the UKF performance in the case of strong coupling, for which not only the phase but also the amplitude dynamics may have to be considered. In addition, we plan to test the UKF ability of infer directed pulsed interactions such as chemical couplings. It would also be interesting to analyze if a more detailed, tailored phase model –instead of the generic Kuramoto model– or if a more advanced phase-extraction procedure (for example, using the Hilbert phase) could improve the performance. It will also be interesting to consider whether other methods to analyze the set of inferred coupling coefficients (rather than k-means clustering) could lead to improved performance.

Acknowledgments

R.P.A. acknowledges financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior–Brasil (CAPES), Finance Code 001. H.A.C. thanks ICTP-SAIFR and FAPESP grants 2021/14335-0 and 2021/11754-2. G.T. and C.M. acknowledge the support of ICREA ACADEMIA, AGAUR (2021 SGR 00606) and Ministerio de Ciencia e Innovación, Spain (Project No. PID2021-123994NB-C21).

Appendix A. Estimation of the average frequency $\bar{\omega}$

As explained in the main text, Sec. 2.4, we aim to assimilate the phase time series to the Kuramoto model (Eq. (5)) and an important step is to select the natural frequency of the oscillators, $\bar{\omega}$. Here we discuss three methods to estimate $\bar{\omega}$ from the data and how they impact the final result of the network inference.

Method 1. For each coupling strength K_m, we estimate $\bar{\omega}$ as $\bar{\omega}_{K_m} = \bar{\phi}(t_f)/t_f$, where $\bar{\phi}(t)$ represents the average phase of the N oscillators at time t_f.

This method was used in Sec. 3.1 with the Izhikevich neurons data and partially with the experimental data in Sec. 3.2. As discussed in the main text, for the Rössler data, this approach only works for small coupling. We verified this in Fig. A.12(a), where we measure the reconstruction accuracy in terms of AUC. As we can see, the

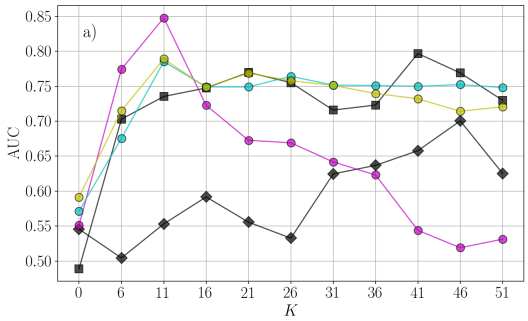

Figure A.12: Comparison of the UKF performance with three methods for estimating the average frequency $\bar{\omega}$, in colored lines. The grey lines display the *MI* and *CC* performances. (a) For the electronic circuits, AUC is used to quantify performance. (b) For the Izhikevich neurons, F_1 is used to quantify performance.

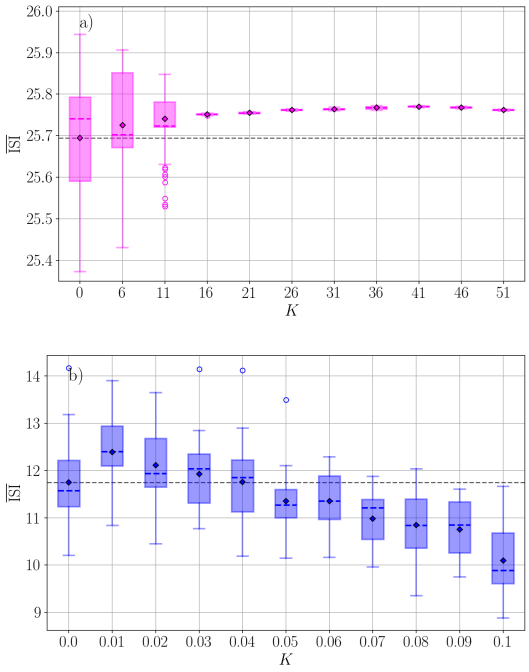

Figure A.13: Distribution of average inter-event-intervals of the electronic circuits (a) and average inter-spike-intervals of a network of $N = 6$ Izhikevich neurons with 8 links (b) as a function of the coupling strength K. The horizontal dashed line marks the average value at zero coupling.

algorithm performance decreases as the coupling strength increases beyond $K = 11$. On the contrary, in the case of the Izhikevich neurons (b), the method displays good performance.

An inspection of the distribution of the average inter-event intervals (spikes or crossings) allows us to better understand why the UKF method does not perform for high coupling in the electronic circuits. As we see in Fig. A.13(a), in the case of electronic circuits, for high coupling, the intervals become nearly identical. Thus, all circuits have approximately the same frequency, there is no need for the coupling term $\sum_{j=1}^{N} m_{ij} \sin(\phi_j - \phi_i)$. In other words, when the oscillators have very similar frequencies, the model $\dot{\phi} = \bar{\omega}$ is sufficient for the assimilation of the data. In contrast, for a network of $N = 6$ Izhikevich neurons with 8 links, Fig. A.13(b), the distribution of the average inter-spike intervals remains broad in the range of coupling values considered. Hence, the coupling term $\sum_{j=1}^{N} m_{ij} \sin(\phi_j - \phi_i)$ is needed to assimilate the phase time series to the Kuramoto model. The same behavior was observed in other neural networks.

Method 2. We select $\bar{\omega}$ as the average frequency computed from the dynamics of the uncoupled oscillators ($K = 0$). As we see in Fig. A.12, for low couplings the UKF performance is not as good as with Method 1, but it is stable around AUC = 0.75 for high coupling. The drawback of this method is that it requires the availability of data with no coupling. However, one can always estimate $\bar{\omega}$ from the lowest coupling available.

Method 3. We assume that all the oscillators have a natural frequency $\bar{\omega} = 0$. This means that Eq. (5) is reduced to the coupling term. We see in Fig. A.12 that this method performs similarly as Method 2, being slightly worse for

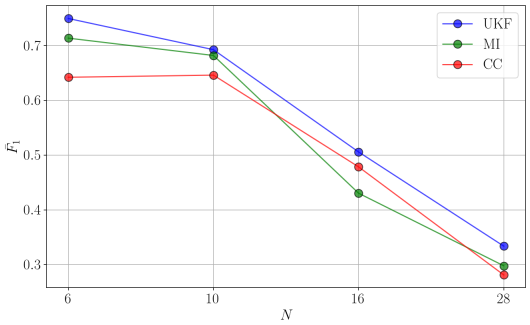

Figure A.14: Performance of UKF, *MI*, and *CC* for Izhikevich neurons, considering different network sizes. The values of p and K are such that, for each network size N, the average Kuramoto order parameter is $\bar{R} = 0.85$.

high coupling values.

Taken together, the results in Figs. A.12 and A.13 show that, if the nodes (the neurons or the chaotic circuits) have not identical frequencies, the UKF approach, estimating $\bar{\omega}$ either with Methods 1, 2, or 3, generally outperforms *CC* and *MI*.

As a final remark, $\bar{\omega}$ could have been an additional parameter to estimate using UKF, with the governing equation given by $d\bar{\omega}/dt = 0$. However, we found that this approach did not improve the performance.

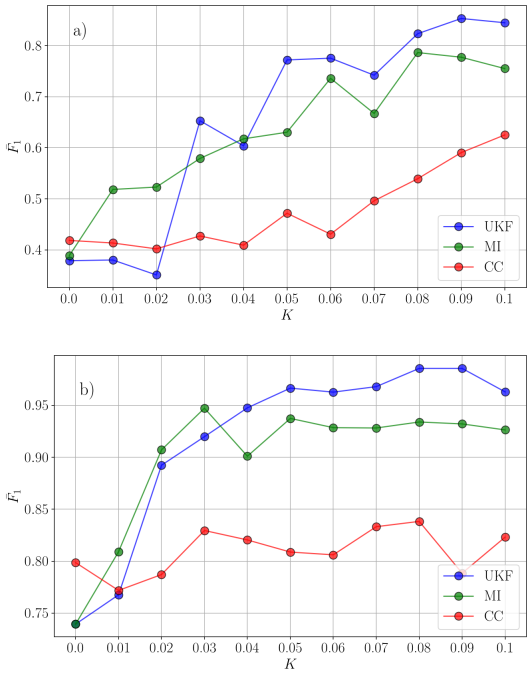

Figure A.15: Performance of UKF, *MI*, and *CC* algorithms for two link densities as a function of the coupling strength. Each curve represents the F_1 score averaged over the different networks and different trials. (a) Results for link density $p = 0.25$ and 7 networks, (b) for $p = 0.75$ and 9 networks.

Appendix B. Influence of the network size and link density

Here we compare the results of the network inference using UKF, *MI*, and *CC* for different network sizes $N = [6, 10, 16, 28]$ and link densities $p = [0.5, 0.5, 0.25, 0.15]$. For each set (N, p) we generated 50 random networks and simulated the Izhikevich model with a coupling strength K such that the networks have a similar level of synchronization, as measured by the average Kuramoto order parameter, $\bar{R} = 0.85$. For each network, we calculated the F_1 and averaged the result over the different networks. As the computational cost of the UKF algorithm grows rapidly (see Sec. Appendix D), we did not repeat the simulations to estimate the F_1 statistics.

The results are displayed in Fig. A.14. Although UKF yields better results for all network sizes considered, we see that the performance of all three methods decreases as N increases.

Regarding the role of the link density, p, we show in Fig. A.15 the average F_1 score for networks with low and higher p. For $p = 0.25$, for low coupling $K < 0.03$, the three methods have poor performance, with the *MI* yielding the best results. For higher coupling, both UKF and *MI* yield a generally good F_1 score, with UKF being the best method across the coupling range. For $p = 0.75$, given the higher number of links (11), to be detected even the *CC* yields a good result. It also comes as a result of the clustering algorithm, which sets the highest values as links. Again, here for weak coupling *MI* yields the best results, while for higher coupling, UKF yields the best performance.

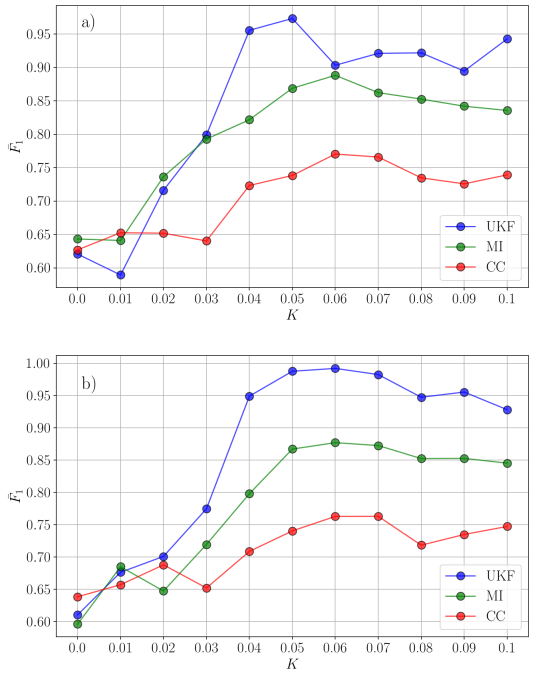

Figure C.16: Performance of UKF, *MI*, and *CC* as a function of the coupling strength K for networks of $N = 6$ Izhikevich neurons, when the time series length is (a) $L = 10000$, (b) $L = 20000$.

Appendix C. Influence of the length of the time series

In the main text, for Izhikevich neurons, we presented results of the analysis of time series of $L = 40000$ data points. Here, we report the results for shorter time series, and therefore, for a smaller number of spikes. To do so, we considered the same networks, coupling strengths, and transient as in the main text, the only difference being the length of the time series analyzed.

Figure C.16 display the F_1 score (averaged over 50 simulations) for $L = 20000$ (a) and for $L = 10000$ (b). In both cases, the UKF algorithm yields better results for $K \geq 0.03$.

Appendix D. Computational costs

Here we report a brief analysis of the computational cost (tested on Intel Xeon Processor Skylake - 2.2GHz)

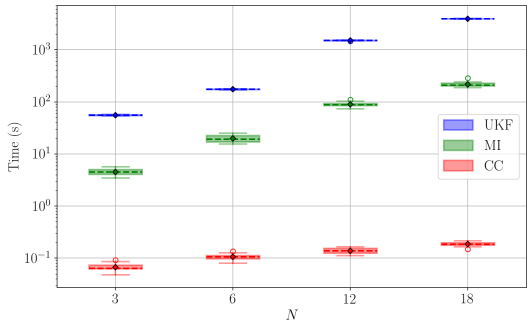

Figure D.17: Comparison of the computational time of UKF, *MI* and *CC* algorithms for three network sizes, N. For each N, 20 networks with link density $p = 0.5$ were used. Each boxplot represents the CPU time required to infer m_{ij}, MI_{ij}, and CC_{ij}.

of the network inference for different network sizes. The results are shown in Fig. D.17, where for each network size N, 20 networks with link density $p = 0.5$ were generated, and their dynamics were recorded in time series of length $L = 40000$ points. The time spent by each algorithm (UKF, MI, CC) to analyze the time series (without counting the binarization of the output) was measured using the *Python* library *time*. As expected, *CC* and *MI*, have a much smaller computational cost, since they only measure pairwise relations between the $N(N-1)/2$ pairs of time series. Meanwhile, UKF analyzes N time series simultaneously, assimilating them to a model in order to obtain the best fit for each model's parameters. For this reason, it requires a significantly longer time than the other two algorithms.

References

[1] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, A. V. Apkarian, Scale-free brain functional networks, Phys. Rev. Lett. 94 (2005) 018102.

[2] E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, Nat. Rev. Neurosci. 10 (2009) 186.

[3] D. Marbach, et al., Wisdom of crowds for robust gene network inference, Nature Methods 9 (2016) 796.

[4] L. Peel, T. P. Peixoto, M. D. Domenico, Statistical inference links data and theory in network science, Nat. Comm. 13 (2022) 6794.

[5] O. Fajardo-Fontiveros, R. Guimera, M. Sales-Pardo, Node metadata can produce predictability crossovers in network inference problems, Phys. Rev. X 12 (2022) 011010.

[6] J. F. Donges, Y. Zou, N. Marwan, J. Kurths, Complex networks in climate dynamics, Eur. Phys. J. ST 174 (2009) 157–179.

[7] H. A. Dijkstra, E. Hernandez-Garcia, C. Masoller, M. Barreiro, Networks in Climate, Cambridge University Press, 2019.

[8] M. Timme, Revealing network connectivity from response dynamics, Phys. Rev. Lett. 98 (2007) 224101.

[9] M. A. Kramer, U. T. Eden, S. S. Cash, E. D. Kolaczyk, Network inference with confidence from multivariate time series, Phys. Rev. E 79 (2009) 061916. doi:10.1103/PhysRevE.79.061916.
URL https://link.aps.org/doi/10.1103/PhysRevE.79.061916

[10] J. Runge, J. Heitzig, V. Petoukhov, J. Kurths, Escaping the curse of dimensionality in estimating multivariate transfer entropy, Phys. Rev. Lett. 108 (2012) 258701.

[11] E. S. C. Ching, P.-Y. Lai, C. Y. Leung, Reconstructing weighted networks from dynamics, Phys. Rev. E 91 (2015) 030801.

[12] M. Zanin, P. Sousa, D. Papo, R. Bajo, J. García-Prieto, F. del Pozo, E. Menasalvas, S. Boccaletti, Optimizing functional network representation of multivariate time series, Sci. Rep. 2 (2012) 630.

[13] N. Rubido, A. C. Marti, E. Bianco-Martinez, C. Grebogi, M. S. Baptista, C. Masoller, Exact detection of direct links in networks of coupled maps, New J. of Phys. 16 (2014) 093010.

[14] J. Sun, D. Taylor, E. M. Bollt, Causal network inference by optimal causation entropy, SIAM Journal on Applied Dynamical Systems 14 (1) (2015) 73–106.

[15] S. Mukherjee, T. P. Speed, Network inference using informative priors, PNAS 105 (2008) 14313–14318.

[16] T. P. Peixoto, Network reconstruction and community detection from dynamics, Phys. Rev. Lett. 123 (2019) 128301.

[17] F. Hassanibesheli, R. Donner, Network inference from the timing of events in coupled dynamical systems, Chaos 29 (2019) 083125.

[18] T. Rings, T. Bröhl, K. Lehnertz, Network structure from a characterization of interactions in complex systems, Sci. Rep. 12 (2022) 11742.

[19] F. Mori, H. Kori, Noninvasive inference methods for interaction and noise intensities of coupled oscillators using only spike time data, PNAS 19 (2022) e2113620119.

[20] S. Leng, Z. Xu, H. Ma, Reconstructing directional causal networks with random forest: Causality meeting machine learning, Chaos 29 (2019) 093130.

[21] A. Banerjee, J. Pathak, R. Roy, J. G. Restrepo, E. Ott, Using machine learning to assess short term causal dependence and infer network links, Chaos 29 (2019) 121104.

[22] A. Banerjee, J. D. Hart, R. Roy, E. Ott, Machine learning link inference of noisy delay-coupled networks with optoelectronic experimental tests, Phys. Rev. X 11 (2021) 031014.

[23] R. Cestnik, M. Rosenblum, Reconstructing networks of pulse-coupled oscillators from spike trains, Phys. Rev. E 96 (2017) 012209.

[24] J. Casadiego, D. Maoutsa, M. Timme, Inferring network connectivity from event timing patterns, Phys. Rev. Lett. 121 (2018) 054101.

[25] M. J. Panaggio, M. V. Ciocanel, L. Lazarus, C. M. Topaz, B. Xu, Model reconstruction from temporal data for coupled oscillator networks, Chaos 29 (2019) 103116.

[26] A. Banerjee, S. Chandra, E. Ott, Network inference from short, noisy, low time-resolution, partial measurements: Application to c. elegans neuronal calcium dynamics, PNAS 120 (2023) e2216030120.

[27] M. Rosenblum, A. Pikovsky, Inferring connectivity of an oscillatory network via the phase dynamics reconstruction, Front. Netw. Physiol. 3 (2023) 1298228.

[28] E. R. Kalman, A new approach to linear filtering and prediction problem, J. Basic Eng. 82 (1960) 35–45.

[29] C. M. Trudinger, M. R. Raupach, P. J. Rayner, I. G. Enting, Using the kalman filter for parameter estimation in biogeochemical models, Environmetrics 19 (2008) 849.

[30] E. Forero-Ortiz, G. Tirabassi, C. Masoller, A. J. Pons, Inferring the connectivity of coupled chaotic oscillators using kalman filtering, Sci. Rep. 11 (2021) 22376.

[31] S. J. Julier, J. K. Uhlmann, New extension of the Kalman filter to nonlinear systems, in: I. Kadar (Ed.), Signal Processing, Sensor Fusion, and Target Recognition VI, Vol. 3068, International Society for Optics and Photonics, SPIE, 1997, pp. 182 – 193.

[32] R. P. Aristides, A. Pons, H. Cerdeira, C. Masoller, G. Tirabassi, Parameter and coupling estimation in small groups of izhikevich's neurons, Chaos 33 (2023) 043123.

[33] Y. Kuramoto, Chemical oscillations, waves and turbulence, Springer, New York, 1984.

[34] E. M. Izhikevich, Simple model of spiking neurons, IEEE Trans. Neural Net. 14 (2003) 1569.

[35] V. P. Vera-Ávila, R. Sevilla-Escoboza, A. A. Lozano-Sánchez, R. R. Rivera-Durón, J. M. Buldú, Experimental datasets of net- works of nonlinear oscillators: Structure and dynamics during the path to synchronization, Data Brief (2020) 105012.

[36] S. Nobukawa, H. Nishimura, T. Yamanishi, J.-Q. Liu, Analysis of chaotic resonance in Izhikevich neuron model, PLOS ONE 10 (2015).

[37] A. Pikovsky, M. Rosenblum, J. Kurths, Synchronization: a universal concept in nonlinear science (2002).

[38] R. Labbe, Filterpy, https://filterpy.readthedocs.io/en (2015).

[39] S. Lloyd, Least squares quantization in pcm, IEEE Transactions on Information Theory 28 (2) (1982) 129–137. doi:10.1109/TIT.1982.1056489.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.

[41] C. J. V. Rijsbergen, Information Retrieval, Butterworths, London, 1979.

[42] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, Cambridge University Press, 1997.

[43] N. Rubido, C. Masoller, Impact of lag information on network inference, The European Physical Journal Special Topics 227 (2018) 1243–1250.