Data-Efficient Inference of Nonlinear Oscillator Networks

Bharat Singhal, Minh Vu, Shen Zeng, and Jr-Shin Li *

* *Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA (e-mail: jsli@wustl.edu)*

Abstract: Decoding the connectivity structure of a network of nonlinear oscillators from measurement data is a difficult yet essential task for understanding and controlling network functionality. Several data-driven network inference algorithms have been presented, but the commonly considered premise of ample measurement data is often difficult to satisfy in practice. In this paper, we propose a data-efficient network inference technique by combining correlation statistics with the model-fitting procedure. The proposed approach can identify the network structure reliably in the case of limited measurement data. We compare the proposed method with existing techniques on a network of Stuart-Landau oscillators, oscillators describing circadian gene expression, and noisy experimental data obtained from Rössler Electronic Oscillator network.

Keywords: Network Inference, Data-driven Modeling, Nonlinear Oscillators, Time-series Analysis

1. INTRODUCTION

The majority of our physiological activities are determined not by a single component, but by the interaction of multiple elements. For example, large networks of co-regulated genes determine the cellular phenotype (Margolin et al., 2006), or interactions between cellular oscillators in the suprachiasmatic nucleus (SCN) generate circadian rhythms (Abel et al., 2016). Typically, such groups of interacting elements are modeled using a network, where elements form the vertices and their interactions are represented by edges. In practice, the time series data from each node might be accessible, and a common objective is to infer the network connectivity structure. Inferring the connections between the nodes is critical for understanding how the functionality of the whole system depends on its connectivity structure and thus allows us to effectively control the network function, which has implications across multiple disciplines (Singhal et al., 2023; Vu et al., 2023).

Owing to such practical significance, various techniques for decoding the network connectivity structures using measured data have been proposed. Information-theoretic measures such as correlation, dynamic time wrapping distance, and mutual information, have been used to determine the connectivity structure (Care et al., 2019; Philips et al., 2022; Margolin et al., 2006). From a system-theoretic viewpoint, rigorous conditions on exact identification have also been provided for networks of linear systems by assuming a fixed network topology such as tree topology (Materassi and Innocenti, 2010). For inference of complex large-scale networks, one common approach is to determine the connectivity structure by fitting the measured data to a preset dynamic model of the network (Casadiego et al., 2017; Wang et al., 2018; Panaggio et al., 2019). These approaches are referred to as parametric methods and offer an interpretable quantification of connectivity via the strength of the connection between the nodes.

Despite these advancements, two practical limitations are often overlooked: (i) As the network size grows, so does the amount of data required to fit the network model, which may be unattainable in certain applications. For example, the high dimensionality of gene transcripts often exceeds the number of available samples (Bühlmann et al., 2014), and long-term recordings are difficult to obtain in biological systems due to various physiological constraints (Abel et al., 2016) (ii) Commonly, only a subset of the nodes in a network can be measured (Hirata and Aihara, 2010), necessitating some adjustments to the current methodologies in order to properly recover the connectivity structure of the observed part by capturing the influence of the unobserved nodes on the observed network. Motivated by these practical considerations, we first propose a data-efficient inference technique that can infer the network connectivity more accurately than the conventional parametric approaches, given limited measured data. More specifically, we present a two-stage network inference method in which we use, in the first step, the sparsity assumption to eliminate low-probability connections in the network and thus reduce the effective network size, and in the second step, we fit the measured data to the reduced network model. Then, we modify the proposed technique to incorporate the effects of unobserved nodes when reconstructing the connectivity structure of a partially observable network. We compare the proposed technique with the standard parametric methods on a network of Stuart-Landau oscillators, and oscillators describing circadian gene expression. We also test the proposed technique on experimental data obtained from Rössler Electronic Oscillator network.

⋆ This work was supported in part by the National Science Foundation under the awards ECCS-1810202 and CMMI-1933976, and by the Air Force Office of Scientific Research under the award FA9550-21-1-0335.

2. THEORY AND ALGORITHM

2.1 Problem Formulation

Consider a network of N limit-cycle oscillators in which the dynamics of each oscillator i, $i = 1, \ldots, N$, consists of its own dynamics and the effect due to coupling with other oscillators, i.e.,

$$\dot{X}_i(t) = f_i(X_i(t)) + \sum_{\substack{j=1 \\ j\neq i}}^{N} K_{ij}(X_i(t), X_j(t)), \tag{1}$$

where $X_i(t) \in \mathbb{R}^n$ $(n \geq 2)$ is the state of the oscillator i at the time t; f_i and $K_{ij}, i, j = 1, \ldots, N$, denote the self dynamics of oscillator i and the coupling function between the oscillators i and j, respectively (K_{ij} can be different from K_{ji}). If there is no coupling between i and j, then K_{ij} and K_{ji} are trivially zero. Given that time series data from the measurement of the oscillators is available, the objective is to recover the (nonzero) coupling functions that describe the connectivity structure of the network.

2.2 2-Step network inference algorithm

One common practice to decode the network topology is to leverage the phase-model description of the oscillator network. For a network with sufficiently weak coupling inputs, the evolution of each oscillator can effectively be captured by a nonlinear 1-dimensional phase model (Pietras and Daffertshofer, 2019), i.e.,

$$\dot{\varphi}_i(t) = \omega_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} \alpha_{ij}(\varphi_j(t) - \varphi_i(t)), \tag{2}$$

where $\varphi_i, \omega_i \in \mathbb{R}$ are the phase and natural frequency of the oscillator i, and α_{ij} is a 2π-periodic coupling function between the oscillators i and j. For a smooth coupling function $\alpha_{ij}(\Delta\varphi_{ij})$, $\Delta\varphi_{ij} = \varphi_j - \varphi_i$, the function is often represented by a truncated Fourier series

$$\alpha_{ij}(\Delta\varphi_{ij}) \approx a_{ij}^{(0)} + \sum_{m=1}^{r} a_{ij}^{(m)} \cos(m\Delta\varphi_{ij}) + b_{ij}^{(m)} \sin(m\Delta\varphi_{ij}), \tag{3}$$

where $\{a_{ij}^{(m)}\}$ and $\{b_{ij}^{(m)}\}$ are the unknown coefficients, and r is the number of Fourier terms that can be chosen sufficiently large to accurately approximate the coupling function. From (2) and (3), the phase dynamics of the oscillator i can be written as

$$\dot{\varphi}_i(t) = \bar{\omega}_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} \sum_{m=1}^{r} a_{ij}^{(m)} \cos(m\Delta\varphi_{ij}) + b_{ij}^{(m)} \sin(m\Delta\varphi_{ij}), \tag{4}$$

where $\bar{\omega}_i := \omega_i + \sum_{j=1, j\neq i}^{N} a_{ij}^{(0)}$. This Fourier representation of coupling functions transforms the nonlinear dynamics into a linear in parameters (LIP) form. Given time series data of each oscillator (e.g., measured voltages) in the network, the oscillator's phase, $\varphi_i(t)$, can be estimated using standard phase-estimation algorithms such as peak-finding or wavelet transform (Mitrou et al., 2017). The derivative of the oscillator phase, $\dot{\varphi}_i$, can also be evaluated, for example, by using forward difference approximation. Now, with the estimated values, (4) becomes a linear equation with unknown coefficients. The coefficients $\bar{\omega}_i, a_{ij}^{(m)}$, and $b_{ij}^{(m)}$ can then be obtained by solving the least-square problem $Y_i = A_i Z_i + \epsilon_i$ for each node i in the network, where $Y_i \in \mathbb{R}^M$ is a vector of the estimated $\dot{\varphi}_i$ at each time step, with M being the number of steps, $A_i \in \mathbb{R}^{M\times 2(N-1)r+1}$ is a matrix formed by the Fourier basis functions evaluated at each sampled point, and $Z_i \in \mathbb{R}^{2(N-1)r+1}$ is a vector containing the unknown coefficients $\bar{\omega}_i, a_{ij}^{(m)}$, and $b_{ij}^{(m)}$. The first element of Z_i denotes $\bar{\omega}_i$, and the remaining $2(N-1)r$ elements are $a_{ij}^{(m)}$ and $b_{ij}^{(m)}$, respectively. The unmodeled dynamics and measurement noise are represented by ϵ_i. Once the coefficients are computed, we can readily obtain the coupling functions, α_{ij}, $i, j = 1, \ldots, N$, describing the network connectivity structure.

In one form or another, the above described method serves as the foundation for many parametric network inference approaches. The predominant distinction between different techniques arises from the types of models considered (in our approach, for example, we utilize the concept of phase models). Regardless of the considered model, as the network size N increases, so does the number of unknowns in the considered model, necessitating a large amount of data to obtain a reliable inference of the model parameters. In the case of limited data, one needs more effective and data-efficient approaches. To this end, we propose a 2−step method to infer the connectivity structure of the network that enhances the parametric inference by incorporating correlation statistics. More specifically, for each node i, we first obtain a set of nodes, $N_c^{(i)}$, that have a high probability of forming connections (with node i) by discarding the nodes with a low probability. We then solve the above described least-square problem by only considering the connections to set $N_c^{(i)}$. By doing so, we effectively reduce the number of unknown parameters for node i and improve the accuracy of the inferred coefficients when only limited measurement data is available.

Step-1: We estimate the correlation between the time series of node i and node j, $j = 1, \ldots, N$, $j \neq i$. Let $C^{(i)}$ be the resulting correlation vector, i.e., $C^{(i)} = \big[c(x_i, x_1), \ldots, c(x_i, x_N)\big]'$, where $c(x_i, x_j)$ is the Pearson correlation coefficient between x_i and x_j, the respective time series. Note that if two nodes i and j are not connected, then their respective time series data will be uncorrelated, resulting in a smaller $c(x_i, x_j)$. In light of this, we define $N_c^{(i)}$ as the set of all the nodes that have highly correlated time series data with node i, i.e., $N_c^{(i)} = \{j \ : |c(x_i, x_j)| \geq k, j \neq i\}$, where $k > 0$ is a user-defined threshold. If multiple time series recordings of the network are available, the set $N_c^{(i)}$ can be defined as $N_c^{(i)} = \{j \ : \frac{1}{L}\sum_{l=1}^{L} |c(x_{i,l}, x_{j,l})| \geq k, j \neq i\}$, where L is the number of recordings, and $x_{i,l}$ denotes the time series of node i in the recording l.

Using only correlation statistics, one might not be able to distinguish between direct and indirect interactions. For example, if there is a direct connection from, say, node i to node j, and node j and k in the network, then the time series data of node i and k will also be (indirectly) correlated. As a result, $N_c^{(i)}$ will contain the

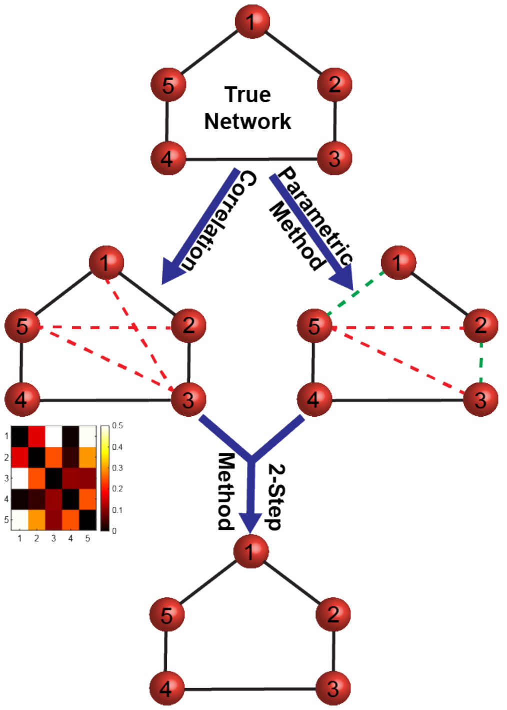

Fig. 1. An illustration of the proposed 2-step network inference method. (a) True network of 5 Stuart-Landau oscillators. (b) Inferred connectivity by estimating the correlation between the recorded time series. The heat-map displays the estimated correlation values. (c) Inferred connectivity via parametric methods, where insufficiency of data leads to missed links (dotted green) and falsely recovered links (dotted red). (d) The proposed method combines correlation statistics with parametric inference to reduce the number of estimated parameters and result in an accurate estimation of network connectivity.

nodes that are both directly and indirectly connected to node i, as illustrated in Figure 1. For this reason, the following additional analysis is considered to differentiate between directly and indirectly connected nodes.

Step-2: After identifying the set of potentially connected nodes, both directly and indirectly, to node i, we simplify the phase dynamics of the oscillator i as

$$\dot{\varphi}_i(t) = \bar{\omega}_i + \sum_{j \in N_c^{(i)}} \sum_{m=1}^{r} a_{ij}^{(m)} \cos(m\Delta\varphi_{ij}) + b_{ij}^{(m)} \sin(m\Delta\varphi_{ij}). \quad (5)$$

Given the estimated phases, (5) can be formulated as a least-square problem of $Y_i = A_i Z_i + \epsilon_i$, as previously described. It is important to note that due to step-1, the unknown coefficient vector here has a smaller dimension, i.e., $Z_i \in \mathbb{R}^{2\mathbf{card}(N_c^{(i)})r+1}$, where $\mathbf{card}(N_c^{(i)}) < N-1$ is the cardinality of set $N_c^{(i)}$. We then impose an l_1-regularization on the least-square problem to promote sparsity in the solution, i.e.,

$$\hat{Z}_i = \arg\min_{Z_i} \|Y_i - A_i Z_i\|_2^2 + \lambda_i \|Z_i\|_1, \quad (6)$$

where the regularization parameter λ_i can be estimated using cross validation. Note that a group penalty, such as group lasso (Yuan and Lin, 2006), can be employed in lieu of a direct l_1 penalty. However, the main objective of this study is not to identify the optimal penalty for network inference, but rather to combine non-parametric network approaches with parametric techniques. Using the estimated coefficient vector $\hat{Z}_i$, we compute the estimated strength of the connection between oscillator i and other oscillators j, i.e.,

$$\hat{\beta}_{ij} = \begin{cases} \sqrt{\sum_{m=1}^{r} (\hat{a}_{ij}^{m})^2 + (\hat{b}_{ij}^{m})^2} & \text{if } j \in N_c^{(i)} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

which will be used, along with some user-defined cutoff threshold, to determine the network's connectivity structure.

2.3 Performance Evaluation Metric

To evaluate the performance of various network inference techniques with our proposed method, we use the area under the receiver operating characteristic curve (ROC curve) which plots the true positive rate (TPR) against the false positive rate (FPR) for various classification thresholds; that is, various cut-off values used to convert the estimated coupling matrix $[\hat{\beta}_{ij}]$ into binary outputs, i.e. the presence or absence of a link.With this metric, a greater area under the ROC curve (AUC score) indicates that the inference method is better at recognizing true and false connections. TPR measures the proportion of correctly identified connections to the number of total connections in the true network, i.e.,

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

where TP (True Positive) is the number of correctly identified links and FN (False Negative) is the number of links in the true network that were missed by the inference algorithm; whereas FPR is defined as:

$$FPR = \frac{FP}{FP + TN}$$

where FP (False Positive) is the number of falsely detected connections and TN (True Negative) is the number of correctly identified negative links. For more details, readers are referred to (Hanley and McNeil, 1982).

3. NUMERICAL SIMULATIONS

In this section, we illustrate the effectiveness of the proposed network inference approach and compare its performance to that of other network inference techniques.

3.1 Stuart-Landau (SL) Oscillator Network

We consider a network of 100 SL oscillators, in which each oscillator is described by

$$\dot{x}_i(t) = ax_i - \omega y_i - (x_i^2 + y_i^2)x_i + \alpha_{ij} \sum_{j=1}^{100} (x_j - x_i),$$

$$\dot{y}_i(t) = \omega x_i + ay_i - (x_i^2 + y_i^2)y_i + \alpha_{ij} \sum_{j=1}^{100} (y_j - y_i),$$

where $x_i(t)$ and $y_i(t)$ are the states of oscillator i, $i = 1, \ldots, 100$, and α_{ij} denotes the coupling strength if there exists a connection between node i and j, otherwise $\alpha_{ij} = 0$. The coefficients a and ω are the amplitude and frequency of the oscillation, respectively. For this oscillator (Nakao, 2016), the phase of the oscillation at time t can be calculated by $\varphi_i(t) = \arctan\left(\frac{y_i(t)}{x_i(t)}\right)$.

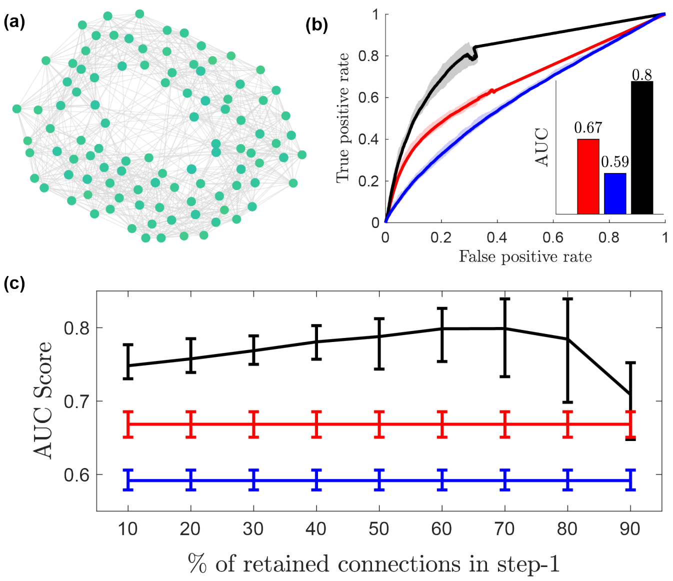

Fig. 2. Connectivity inference of SL oscillator networks. (a) Representation of a 100-node network. (b) Mean ROC curve and AUC scores (inset) employing the proposed approach (black), sparse regression (red), and ICON (blue); variation in the ROC curve, over 10 different networks, is shown using the shaded region. (c) Mean AUC scores (where the bars denote the minimum and maximum values) when the threshold in step-1 is varied to retain different percentages of highly correlated nodes for each node i.

We evaluate and compare the proposed method with the standard sparse-regression technique and the ICON method (Wang et al., 2018), where functional approximations along with truncated SVD were utilized to determine the network topology. Inference of network topology could become infeasible when the network is synchronized (Shandilya and Timme, 2011) since the observed phase differences are constant and no longer provide additional information. To avoid network synchronization, we generate data using 5 different initial conditions, and for each time, we record the data for 10 cycles with a sampling rate of 150 points per cycle.

Figure 2 compares the network inference accuracy of the three techniques when only a small data set is available, i.e., the measured data is not sufficient to recover the accurate network topology with an AUC score of 1. For this limited data set, we select the cutoff threshold for step-1, such that 70% of the most correlated nodes are retained for the inference in step-2. The proposed method achieves a mean AUC score of 0.8, an improvement of 76%, compared to the other best-performing method, which is based on enhancements above the baseline of 0.5 for random guesses. We also evaluate the effect of varying the threshold in step-1 on the inference accuracy of the proposed technique for different network structures. The threshold for step-1 is selected such that the x%, $x = 10, \dots, 90$, of the most highly correlated nodes, for each node, are retained. As illustrated in Figure 2(c), the proposed method outperforms ICON and sparse regression algorithms for all threshold levels.

3.2 Circadian Gene Expression Network

We consider a network of 100 oscillators that has been used to describe circadian oscillations (Leloup et al., 1999). A 3-dimensional model describing the dynamics of each oscillator is given by

$$\dot{M}_i(t) = v_{s_i} \frac{K_1^n}{K_1^n + P_{n_i}^n} - v_{m_i} \frac{M_i}{M_i + K_m}$$

$$\dot{P}_{c_i}(t) = k_s M_i - v_d \frac{P_{c_i}}{k_d + P_{c_i}} - k_1 P_{c_i} + k_2 P_{n_i}$$

$$\dot{P}_{n_i}(t) = k_1 P_{c_i} - k_2 P_{n_i},$$

where P_{c_i}, P_{n_i}, and M_i are the states of the oscillator i, $\{n, K_1, K_m, v_d, k_s, k_d, k_1, k_2\} = \{4, 1, 0.5, 1, 0.41, 0.13, 0.41, 0.5\}$ are the constant parameters, and v_{m_i} is uniformly distributed in $[0.3655, 0.3745]$. The connections of the oscillator i with other oscillators are represented by v_{s_i}, where $v_{s_i} = 0.83 + \sum_{j=1}^{100} \alpha_{ij}(M_j - M_i)$ and α_{ij} are the strengths of the connections.

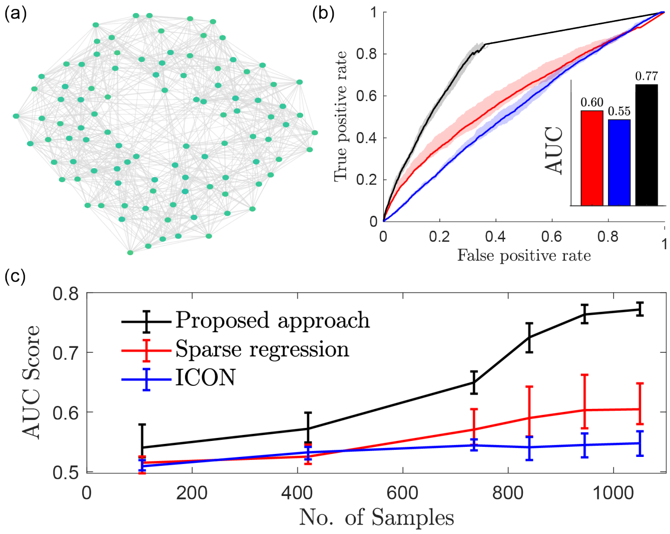

Fig. 3. Reconstruction of circadian oscillator networks. (a) A network of 100 circadian oscillators. (b) The Mean ROC curve with the variation, across 10 different networks, described by the shaded region and the corresponding AUC scores (inset) employing the proposed approach (black), ICON (blue), and sparse regression (red). (c) Quality of reconstruction as a function of the number of samples, with the bars denoting the (min-max) range of AUC scores.

We simulate the circadian 100-oscillator network for 3 different initial conditions, each for 10 cycles, to generate the measurement data (sampling rate: 48 points/cycle). The phase of the recorded signal M_{c_i}, $i = 1, \dots, 100$, is extracted by the wavelet transform using the open-source Python-based Biological Oscillations Analysis Toolkit (Schmal et al., 2022). The proposed approach, ICON, and sparse regression are used to reconstruct the network topology from the estimated phase data. This process is then repeated 10 times for different network structures. The resulting mean AUC scores along with the mean ROC curves are displayed in Figure 3(b). In our method, the threshold in step-1 is chosen to retain 50% of the most correlated nodes, for each node. By focusing on inferring the connection of the highly-correlated nodes, our results demonstrate a 170% improvement over the other best-performing method (sparse regression); the mean AUC is improved from 0.6 to 0.77 (i.e., from the enhancement of 0.1 to 0.27 above the baseline 0.5). We also evaluate the reconstruction accuracy across the 10 networks when different amounts of data are considered. Figure 3(c) illustrates that the proposed method achieves a more accurate inference than the other methods for all sample sizes.

3.3 Experimental Data-set for a Network of Rössler Electronic Oscillators

In experimental settings, measurement noise is an uncontrolled element that affects the recorded time series. To evaluate the robustness of the proposed approach against measurement noise, we consider the experimental time series recordings obtained from a network of 28 Rössler electronic oscillators. The publicly accessible data set comprises time series recordings for 20 distinct network topologies, and for each network configuration, time series recordings were obtained for varying coupling strengths between oscillators (Vera-Ávila et al., 2020).

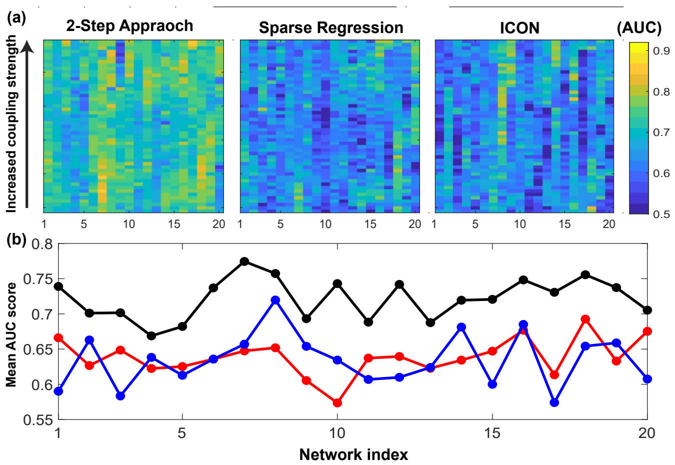

Fig. 4. Network reconstruction using noisy experimental data-sets. (a) Heat maps displaying estimated AUC values for 20 different network topologies and varying coupling strengths for each network topology. (b) The mean AUC scores across coupling strengths for each network topology, in the cases of the proposed method (black), ICON (blue), and sparse regression (red).

For all 20 network topologies and associated coupling strengths, we evaluate and compare the proposed technique with Sparse Regression and ICON. The results are summarized in Figure 4. The proposed method outperforms sparse regression and ICON across all network topologies and coupling strengths, as illustrated in Figure 4(a). The mean AUC scores across different coupling strengths for each network topology are plotted in Figure 4(b), where we observe a minimum improvement of 21% over the best-performing methods between sparse regression and ICON; the maximum improvement is 230%. The threshold in step-1 is chosen such that 50% of the most correlated nodes are retained, for each node.

4. PARTIALLY OBSERVED NETWORKS

In many applications, as described in Section 1, it may be unfeasible to record the trajectories of all the nodes due to sensing restrictions. When only a subset of nodes can be observed, the inference algorithm may perceive the effect from unobserved nodes as additional connections between the observable nodes, thus promoting the recovery of false interactions. For this reason, it is essential to account for the influence of the unobserved nodes when determining the connectivity structure of a partially observed network.

In an N-node network, we define N^o as the set of measurable nodes with $N^o < N$. Then, the coupling to node i will have two components: one due to the measurable nodes $\sum_{j\in N^o} \alpha_{ij}(\Delta\varphi_{ij})$ and one from the unmeasurable nodes $\sum_{j\in\{1,\ldots,N\}\setminus N^o} \alpha_{ij}(\Delta\varphi_{ij})$. In this scenario, the approach of leveraging phase measurements to write the dynamics of node i as a linear equation of unknown parameters will fall short due to the lack of measurement φ_j when $j \in \{1,\ldots,N\} \setminus N^o$. We handle this issue by viewing the coupling functions to node i from the unobservable nodes as an unknown input $u_i(t)$ and express the phase dynamics of node i in the form of

$$\dot{\varphi}_i(t) = \omega_i + \sum_{j\in N^o} \alpha_{ij}(\varphi_j(t) - \varphi_i(t)) + u_i(t). \tag{8}$$

Because each coupling function is a periodic function, the unknown input $u_i(t) = \sum_{j\in\{1,\ldots,N\}\setminus N^o} \alpha_{ij}(\Delta\varphi_{ij}(t))$ will be of periodic nature and can thus be captured by Fourier series, i.e.,

$$u_i(t) = c_{i0} + \sum_{m=1}^{r_i} c_{im}\cos(\frac{2\pi m}{T}t) + d_{im}\sin(\frac{2\pi m}{T}t), \tag{9}$$

where T is the length of the recorded time series. This simple modification allows us to estimate the effects of the unobserved nodes and thus account for their influences when determining network structures.

Now, by following the ideas introduced in Section 2, we estimate the set of highly correlated nodes to each node i, $N_c^{o(i)} \subset N^o$, and then leverage the equations (8), (9) and (5) to rewrite the node i dynamics as

$$Y_i = [A_i \;\; A_{i,u}] \begin{bmatrix} Z_i \\ Z_{i,u} \end{bmatrix} + \epsilon_i, \tag{10}$$

where the matrix $A_{i,u}$ contains the basis functions corresponding to the input u_i, and the vectors Z_i and $Z_{i,u}$ are the unknown coefficients which indicate connections to highly correlated nodes in $N_c^{o(i)}$ and the effects of the unknown input $u_i(t)$, respectively. We apply l_1-regularization to solve the above least-square problem and obtain the connections among the observed nodes.

We illustrate the idea through a numerical simulation on a network of 50 Kuramoto oscillators where the phase of oscillator i is described by $\dot{\varphi}_i(t) = \omega_i + \sum_{j=1}^{50} \alpha_{ij}\sin(\varphi_j - \varphi_i)$. The natural frequencies of the oscillators are uniformly distributed in $[3, 5]$, and the coupling strengths α_{ij} are

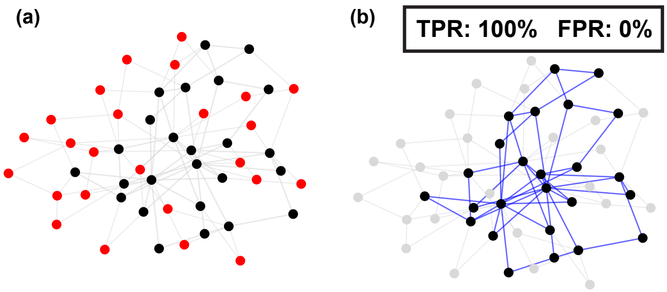

Fig. 5. Recovering the network connectivity when only a subset of the nodes can be measured. (a) The underlying network where only the nodes shown in black (red) can be (cannot be) measured. (b) The recovered network between the measured nodes.

normally distributed with a mean and variance of 0.04 and 0.004. For this 50-node network, we only record the phase trajectories of 25 randomly selected oscillators and apply the proposed framework to recover the connectivity structure among the recorded oscillators. The underlying network and the recovered network are shown in Figure 5.

5. CONCLUSION

In this paper, we introduced a data-efficient network reconstruction technique that combines correlation statistics with parametric inference and a simple modification of the technique to effectively incorporate the effects of unobserved nodes in the network. The combined development offers an effective computational framework for inference of oscillator networks, which has been demonstrated through a wide range of examples. Networks composed of limit-cycle oscillators are used as the mathematical setting for our investigation; however, the presented ideas are also applicable to networks of general nonlinear systems.

REFERENCES

Abel, J.H., Meeker, K., Granados-Fuentes, D., John, P.C.S., Wang, T.J., Bales, B.B., Doyle, F.J., Herzog, E.D., and Petzold, L.R. (2016). Functional network inference of the suprachiasmatic nucleus. *Proceedings of the National Academy of Sciences*, 113(16), 4512–4517.

Bühlmann, P., Kalisch, M., and Meier, L. (2014). High-dimensional statistics with a view toward applications in biology. *Annual Review of Statistics and Its Application*, 1(1), 255–278.

Care, M.A., Westhead, D.R., and Tooze, R.M. (2019). Parsimonious gene correlation network analysis (pgcna): a tool to define modular gene co-expression for refined molecular stratification in cancer. *NPJ systems biology and applications*, 5(1), 1–17.

Casadiego, J., Nitzan, M., Hallerberg, S., and Timme, M. (2017). Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature communications*, 8(1), 1–10.

Hanley, J.A. and McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143(1), 29–36.

Hirata, Y. and Aihara, K. (2010). Identifying hidden common causes from bivariate time series: A method using recurrence plots. *Physical Review E*, 81(1), 016203.

Leloup, J.C., Gonze, D., and Goldbeter, A. (1999). Limit cycle models for circadian rhythms based on transcriptional regulation in drosophila and neurospora. *Journal of biological rhythms*, 14(6), 433–448.

Margolin, A.A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006). Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, 1–15. BioMed Central.

Materassi, D. and Innocenti, G. (2010). Topological identification in networks of dynamical systems. *IEEE Transactions on Automatic Control*, 55(8), 1860–1871.

Mitrou, N., Laurin, A., Dick, T., and Inskip, J. (2017). A peak detection method for identifying phase in physiological signals. *Biomedical Signal Processing and Control*, 31, 452–462.

Nakao, H. (2016). Phase reduction approach to synchronisation of nonlinear oscillators. *Contemporary Physics*, 57(2), 188–214.

Panaggio, M.J., Ciocanel, M.V., Lazarus, L., Topaz, C.M., and Xu, B. (2019). Model reconstruction from temporal data for coupled oscillator networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(10), 103116.

Philips, R.T., Torrisi, S.J., Gorka, A.X., Grillon, C., and Ernst, M. (2022). Dynamic time warping identifies functionally distinct fmri resting state cortical networks specific to vta and snc: A proof of concept. *Cerebral Cortex*, 32(6), 1142–1151.

Pietras, B. and Daffertshofer, A. (2019). Network dynamics of coupled oscillators and phase reduction techniques. *Physics Reports*, 819, 1–105.

Schmal, C., Mönke, G., and Granada, A.E. (2022). Analysis of complex circadian time series data using wavelets. In *Circadian Regulation*, 35–54. Springer.

Shandilya, S.G. and Timme, M. (2011). Inferring network topology from complex dynamics. *New Journal of Physics*, 13(1), 013004.

Singhal, B., Vu, M., Zeng, S., and Li, J.S. (2023). An iterative approach to optimal control design for oscillator networks. In *2023 American Control Conference (ACC)*, (to appear). IEEE.

Vera-Ávila, V., Sevilla-Escoboza, R., Lozano-Sánchez, A., Rivera-Durón, R., and Buldú, J.M. (2020). Experimental datasets of networks of nonlinear oscillators: Structure and dynamics during the path to synchronization. *Data in brief*, 28, 105012.

Vu, M., Singhal, B., Zeng, S., and Li, J.S. (2023). Data-driven control of neuronal networks with population-level measurement. doi:10.21203/rs.3.rs-2600572/v1.

Wang, S., Herzog, E.D., Kiss, I.Z., Schwartz, W.J., Bloch, G., Sebek, M., Granados-Fuentes, D., Wang, L., and Li, J.S. (2018). Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks. *Proceedings of the National Academy of Sciences*, 115(37), 9300–9305.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.