Learning the Causal Structure of Networked Dynamical Systems under Latent Nodes and Structured Noise

Augusto Santos[1*], **Diogo Rente**[2], **Rui Seabra**[2], **José M. F. Moura**[2]

[1] Instituto de Telecomunicações-IT, Lisbon, Portugal
[2] Department of Electrical and Computer Engineering at Carnegie Mellon University, Pittsburgh, PA, USA
augusto.santos@lx.it.pt, drobertd@andrew.cmu.edu, rvilarpo@andrew.cmu.edu, moura@andrew.cmu.edu.

Abstract

This paper considers learning the hidden causal network of a linear networked dynamical system (NDS) from the time series data at some of its nodes – partial observability. The dynamics of the NDS are driven by *colored* noise that generates spurious associations across pairs of nodes, rendering the problem much harder. To address the challenge of noise correlation and partial observability, we assign to each pair of nodes a feature vector computed from the time series data of observed nodes. The feature embedding is engineered to yield structural consistency: There exists an affine hyperplane that consistently partitions the set of features, separating the feature vectors corresponding to connected pairs of nodes from those corresponding to disconnected pairs. The causal inference problem is thus addressed via clustering the designed features. We demonstrate with simple baseline supervised methods the competitive performance of the proposed causal inference mechanism under broad connectivity regimes and noise correlation levels, including a real world network. Further, we devise novel technical guarantees of structural consistency for linear NDS under the considered regime.

Introduction

Complex Networked Dynamical Systems (NDS) are causally structured: the state of each unit or node evolves over time due to peer-to-peer interactions or local information exchange (Barrat, Barthélemy, and Vespignani 2008). Examples include: brain activity (Huang and Ding 2016; Liégeois et al. 2020); Gene Regulatory Networks (Aalto et al. 2020; Casadiego et al. 2017); or pandemics (Ganesh, Massoulié, and Towsley 2005; Wan et al. 2014). Understanding the underlying connectivity pattern in these applications is fundamental to forecast the long term behavior of the NDS as, for example, predicting the onset of criticality (Eroglu et al. 2020) or neurological crises (Douw et al. 2010); ascertaining whether a strain of virus persists or dies out in the long run (Lahmanovich and James 1976); or for structure-informed downstream tasks, e.g., the design of mitigation measures (Ren et al. 2019) in a pandemics.

In all these applications, usually, only the time series data at some of the nodes are readily available while the underlying causal structure linking these nodes lurks underneath.

The main goal of causal inference or structure identification is to map out the underlying causal architecture of a complex system, i.e., to enable consistent inference of the latent network structure from the observed data.

For the most part, causal inference is addressed in two phases: i) *Estimation Phase*— a scalar value is assigned to each pair of nodes as a proxy to the pair's coupling strength or affinity (e.g., mutual information); ii) *Classification Phase*— a thresholding or hypothesis testing is applied to the estimated couplings to draw the network, consistently.

Reference (Machado et al. 2023) departed from this approach and addressed the framework of partial observability (with *diagonal* or uncorrelated noise), under a novel feature based paradigm. Instead of a scalar-based estimation for the pairs' affinity, a feature vector is assigned to each pair of nodes, and structure is recovered by leveraging certain identifiability properties of the engineered features in a higher-dimensional space (instead of thresholding). Desirable characteristics of this feature based approach include: i) *Separability* —there exists a manifold that separates the features of connected pairs from those of disconnected pairs; ii) *Stability*—this separation manifold should not be *too sensitive* to differences in the regimes of connectivity, observability and noise correlation of the underlying complex system; and iii) *Locality*—the feature of each pair can be computed from the time series of each pair (or neighboring pairs thereof). *Separability* is a necessary property to consistently cluster the features, i.e., for structure inference; *stability* renders supervised ML tools (e.g., SVM) amenable to generalization; and *locality* of the features is crucial for large-scale systems where only a subset of nodes can be feasibly observable.

However, in most applications, the underlying excitation noise exhibits nontrivial correlation structure. Depending on the regime of observation or noise structure, information about the main object of inference may be fundamentally lost (Barbier et al. 2023). The level of noise plays a major role in the feasibility of statistical inference tasks in that the problem can be *impossible*, *hard* (possible, but no known polynomial-time algorithm), or *easy*, depending on the noise level (Barbier et al. 2019). It is thus, critical to understand the impact of the noise structure on inference problems.

In this work, we address causal inference under the delicate regime of partial observability and structured noise. For this structure identification problem, and defying common

wisdom in statistical physics, we show that structural information is preserved in the partially observed time series irrespective of the level of noise correlations, as long as they are sufficiently homogeneous across pairs (Theorem 2). Under greater correlation heterogeneity, we show that exogenous interventions help recover feasibility.

Our contribution is twofold: i) we establish a novel identifiability condition for linear networked dynamical systems (Theorem 2); ii) we propose novel features to devise a Machine Learning based causal inference mechanism exhibiting competitive performance under the referred regime. Point i) provides, in particular, novel guarantees for linear networked dynamical systems, where the information of the underlying network structure is contained in the observed time series data (i.e., it is not fundamentally lost).

Related Work

There is no universal method for causal inference, i.e., there does not exist an algorithm that is consistent and optimal across a broad range of attributes, including: 1) dynamical laws; 2) structural patterns – sparse, dense networks; 3) regimes of observability (full or partial); and 4) statistical properties of the input or perturbation noise. For example, if the observed data follow a Gaussian multivariate distribution, the *Precision* matrix – or regularized versions thereof, e.g., Graphical Lasso – is a structurally consistent estimator (Loh and Wainwright 2013); if the samples follow a ferromagnetic Ising model distribution, then the *correlation* matrix is a possible consistent estimator over sparse networks (Montanari and Pereira 2009); Chow-Liu's mutual information based seminal algorithm (Chow and Liu 1968) is structurally consistent for a great family of probability distributions generating the i.i.d. samples, but it is specialized to trees; Granger (Geiger et al. 2015; Matta, Santos, and Sayed 2020; Santos, Matta, and Sayed 2020; Chen, Wang, and Shen 2022) is consistent for linear dynamical systems (even under partial observability), but it exhibits compromised performance under colored noise and partial observability; conditional independence tests and corresponding measures seldom fit densely connected networked systems.

For a recent comprehensive discussion on related causal inference works, please refer to (Machado et al. 2023) or (Matta, Santos, and Sayed 2020). Most works focus on sparse networks – by methods primarily promoting model-parsimony – and/or *unstructured* excitation noise, i.e., independence of the exogenous noise across distinct nodes in the system. While this renders the inference task technically tractable, for the most part, it is an overall unrealistic assumption. For example, this noise independence across nodes is a fundamental assumption in Structural Causal Models (SCM) (Pearl 2009; Peters, Janzing, and Schlkopf 2017) and underlies the structural consistency of conditional independence tests based algorithms (Spirtes 2001; Spirtes, Glymour, and Scheines 2000; Anandkumar et al. 2013; Anandkumar and Valluvan 2013) as well as other ML-based approaches (Melnychuk, Frauen, and Feuerriegel 2022; Cao et al. 2023; Runge et al. 2019; Cheng et al. 2023; Marcinkevičs and Vogt 2021). Several works explore linear networked dynamical systems (Materassi and Salapaka
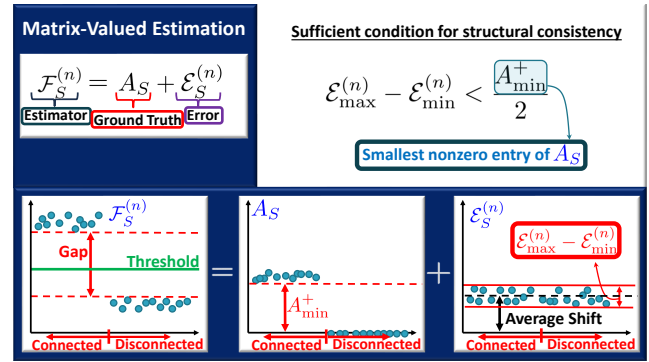

Figure 1: Sufficient condition for structural consistency. The plots represent the off-diagonal entries of the corresponding matrices: the estimator, the ground truth and the error.

2012, 2015; Pereira, Ibrahimi, and Montanari 2010; Santos, Matta, and Sayed 2020; Matta, Santos, and Sayed 2022) tailored to sparse networks and uncorrelated noise. The recent reference (Shaikh Veedu and Salapaka 2023) considers linear dynamic influence models (LDIMs), defined in the frequency domain, with correlated noise and develops a Wiener based approach specialized to sparse networks – leveraging on a sparse + low rank method. The reference makes important limiting assumptions on the noise correlation: i) it results from affine interactions; and, ii) the excitation noise is assumed independent across latent nodes.

In this work, we address the causal inference problem for linear networked dynamical systems (NDS) when the noise across nodes is correlated and under the partial observability regime (only a subset of nodes is observed). We provide novel technical guarantees of structural consistency or identifiability and show that the proposed approach exhibits significant superior performance than benchmark alternatives across distinct regimes of network connectivity, observability, and noise correlation. We assume no knowledge of the underlying noise covariance matrix or the network of interactions, and we impose no assumptions on the particular nature of the noise correlations.

Problem Formulation

The time series data streams from the linear Networked Dynamical System (NDS) given by

$$\mathbf{y}(n+1) = A\mathbf{y}(n) + \mathbf{x}(n+1), \tag{1}$$

where $\mathbf{y}(n) = [y_1(n) \ y_2(n) \ \dots \ y_N(n)]^\top \in \mathbb{R}^N$ represents the state-vector of the N-dimensional NDS collecting the state $y_i(n)$ of each node $i = 1, 2, \dots, N$ at time n; $\mathbf{x}(n)$ is the zero mean input noise at time n with finite first and second order moments, i.e., it admits a well-defined covariance matrix $\Sigma_x \in \mathbb{S}_+^N$, and $(\mathbf{x}(n))_{n \in \mathbb{N}}$ is independent across time n; $A \in \mathbb{R}_+^N$ is the non-negative symmetric interaction matrix describing the graph linking the nodes as its support. We assume that the spectral radius of A is smaller than 1, $\rho(A) < 1$, i.e., the dynamical system is assumed stable.

Namely, we assume that A is stochastic up to a multiplicative constant, i.e., $A = \rho \overline{A}$ where $\overline{A}$ is stochastic and $\rho < 1$.

The discrete-time linear NDS (1) emerges naturally from time discretization of continuous-time linear stochastic differential equations as, e.g., in (Pereira, Ibrahimi, and Montanari 2010; Aalto et al. 2020). These linear stochastic networked dynamical systems may also describe nonlinear networked dynamical systems operating close to an equilibrium state with *small* enough noise level (Ching and Tam 2017; Ren et al. 2010), in the spirit of the Hartman-Grobman linearization Theorem (Barreira and Valls 2012). Linearity can also be recovered via appropriate Koopman-based embedding in higher dimensional spaces (Mauroy and Goncalves 2016; Lusch, Kutz, and Brunton 2018; Hossain, Adesunkanmi, and Kumar 2023), change of variables, or statistical transformations like Box-Cox (Passemiers, Moreau, and Raimondi 2022). These render structural inference methods developed for the linear NDS (1) consistent or useful over a broad class of nonlinear NDS.

The goal of graph recovery under *full* observability is to infer consistently the support of the interaction matrix A from the time series data $\{\mathbf{y}(n)\}_{n=1}^{T}$. Under partial observability (Machado et al. 2023; Santos, Matta, and Sayed 2020; Mei and Moura 2018; Matta, Santos, and Sayed 2020; Geiger et al. 2015; Matta, Santos, and Sayed 2022), it is to recover the support of the submatrix A_S that links a set $S := \{m_1, m_2, \ldots, m_{|S|}\}$ of observable nodes from their time series $\{[\mathbf{y}(n)]_S\}_{n=1}^{T}$, where $[\mathbf{y}(n)]_S \triangleq \left[\mathbf{y}_{m_1}(n)\ \mathbf{y}_{m_2}(n) \cdots \mathbf{y}_{m_{|S|}}(n)\right]^{\top} \in \mathbb{R}^{|S|}$ is the $|S|$-dimensional vector collecting the states of the observed nodes at each time n, $|S| < N$.

Define the kth lag empirical covariance matrix $\widehat{R}_k(n) \triangleq \frac{1}{n} \sum_{\ell=0}^{n-1} \mathbf{y}(\ell + k)\mathbf{y}(\ell)^{\top}$. Similarly to as done in (Machado et al. 2023), we refer to a *matrix-valued estimator* as any map whose input is given by the (observed) time series data and the output is given by a matrix, namely,

$$
F_S^{(n)} : \quad
\begin{array}{ccc}
\mathbb{R}^{|S| \times n} & \longrightarrow & \mathbb{R}^{|S| \times |S|} \\
\{[\mathbf{y}(\ell)]_S\}_{\ell=0}^{n-1} & \longmapsto & \mathcal{F}_S^{(n)}
\end{array} , \quad (2)
$$

for any given $n \in \mathbb{N}$. The ij^{th} entry of the matrix $\mathcal{F}_S^{(n)}$ reflects the strength of the edge from j to i from n time series samples. That is, $\mathcal{F}_S^{(n)}$ is a matrix summarizing the estimated affinities between the pairs of nodes in the system. The estimator $F_S^{(n)}$ is *structurally consistent* whenever the estimated strength $\mathcal{F}_{ij}^{(n)}$ of any connected pair (i, j) lies above the estimated strength of any disconnected pair. In this case, the underlying network structure is recovered via proper thresholding. This is illustrated in Fig. 1 and it is formalized with the following definition presented in (Machado et al. 2023).

Definition 1 (Structural Consistency). *A matrix-valued estimator $F^{(n)}$ is structurally consistent with high probability, whenever there exists a threshold τ so that, $\mathbb{P}\left(\mathcal{F}_{ij}^{(n)} > \tau\right) \overset{n \to \infty}{\longrightarrow} 1 \iff j \to i$, i.e., j links to i if and only if the ij^{th} entry of the estimator matrix $\mathcal{F}^{(n)}$ lies above the threshold τ, given a large enough number of samples n.*

That is, up to a proper threshold $\tau \in \mathbb{R}$, the output matrix $\mathcal{F}^{(n)}$ reflects the underlying structure of the graph in that $[\text{Supp}(A_S)]_{ij} = \mathbf{1}_{\left\{\mathcal{F}_{ij}^{(n)} > \tau\right\}}$, for all pairs $i \neq j$ w.h.p.

We adopt the following three assumptions throughout.

Assumption 1 (NDS Stability). *The interaction matrix A is assumed symmetric, nonnegative and stable, i.e., $\rho(A) < 1$, where $\rho(A)$ is the spectral radius of A.*

Assumption 2 (Nodewise Homogeneity). *We assume that $\sigma^2 = \mathbb{E}\left[\mathbf{x}_i^2\right]$ for all i.*

Remark that $\sigma^2 \geq \mathbb{E}\left[\mathbf{x}_i \mathbf{x}_j\right] \forall i, j$, necessarily. Indeed,

$$
0 \leq E[(\mathbf{x}_i - \mathbf{x}_k)^2] = E[\mathbf{x}_i^2] + E[\mathbf{x}_k^2] - 2E[\mathbf{x}_i \mathbf{x}_k] \\
= 2\sigma^2 - 2E[\mathbf{x}_i \mathbf{x}_k]. \quad (3)
$$

Assumption 3 (Pairwise Distinguishability). *We assume that $\sigma^2 > \mathbb{E}\left[\mathbf{x}_i \mathbf{x}_j\right] = [\Sigma_x]_{ij}$ for all $i \neq j$. That is, the off-diagonal entries of the noise covariance matrix are strictly smaller than the diagonal.*

Assumption 1 makes the NDS stable and it is thus natural. Assumption 2 implies Assumption 3 whenever there are no pair of nodes with the same noise source, i.e., $\mathbb{P}\left(\mathbf{x}_i \neq \mathbf{x}_j\right) > 0$, for $i \neq j$. Indeed, from (3), $\sigma^2 = E\left[\mathbf{x}_i \mathbf{x}_k\right]$ if and only if $\mathbb{E}\left[(\mathbf{x}_i - \mathbf{x}_k)^2\right] = 0$ which implies $\mathbf{x}_i \overset{\text{a.s.}}{=} \mathbf{x}_k$. Therefore, Assumption 3 is mild under Assumption 2.

In view of Assumptions 2 and 3, we shall refer to $\sigma_{\text{gap}}^2 \overset{\Delta}{=} \sigma^2 - \max_{i \neq j} \mathbb{E}\left[\mathbf{x}_i \mathbf{x}_j\right] > 0$ as the gap between the diagonal and the off-diagonal entries of the noise covariance Σ_x.

Structural Consistency under Noise Structure and Partial Observability

We write a matrix-valued estimator, see equation (2), as $\mathcal{F}_S^{(n)} = A_S + \mathcal{E}_S^{(n)}$ where, under partial observability, A_S is the underlying ground truth interaction matrix, $\mathcal{E}_S^{(n)}$ is the (matrix) error term, and S is the set of observable nodes. As we now show, for structural consistency, we do not need the error term to be zero (or to converge to zero with n). Define $\text{Osc}\left(\mathcal{E}_S^{(n)}\right) \overset{\Delta}{=} \mathcal{E}_{\max}^{(n)} - \mathcal{E}_{\min}^{(n)}$ and let $A_{\min}^{+}$ be the smallest nonzero entry of A_S, where $\mathcal{E}_{\max}^{(n)}$ and $\mathcal{E}_{\min}^{(n)}$ are the maximum and minimum off-diagonal entries of the error matrix $\mathcal{E}_S^{(n)}$. Then, if, with high probability (w.h.p.),

$$
\text{Osc}\left(\mathcal{E}_S^{(n)}\right) \leq \frac{A_{\min}^{+}}{2}, \quad (4)
$$

it is easily seen that $F_S^{(n)}$ is structurally consistent. In fact, condition (4) precludes *inversion* and consequent loss of structural information in Definition 1 in the following sense: If ij is a connected pair and $k\ell$ is a disconnected pair, then constraint (4) implies that $\mathcal{F}_{ij}^{(n)} > \mathcal{F}_{k\ell}^{(n)}$ necessarily. This means that, under (4), structure can be recovered via properly thresholding the off-diagonal entries of the matrix $\mathcal{F}_S^{(n)}$, otherwise, structural information is lost. Further, when the

error matrix is *flat*, i.e., $\mathsf{Off}\left(\mathcal{E}_S^{(n)}\right) \propto \mathsf{Off}\left(\mathbf{11}^\top\right)$, all the entries of $F^{(n)}$ are the entries of A_S shifted by the same quantity. Therefore, again, structural information is preserved. Fig. 1 offers an illustration of this discussion.

Example: Granger estimator. Reference (Matta, Santos, and Sayed 2020) studied the error $\mathcal{E}_S$ for the Granger estimator $[R_1]_S ([R_0]_S)^{-1}$, under partial observability, where R_1 and R_0 are the limit one-lag and correlation matrices, respectively. Under full observability, Granger recovers the underlying interaction matrix A w.h.p. since $R_1 (R_0)^{-1} = A$. In other words, the limiting error matrix $\mathcal{E}$ associated with Granger under full observability is zero (trivially flat). However, under partial observability, the limiting error matrix associated with Granger is given by (Matta, Santos, and Sayed 2020) $\mathcal{E}_S = A_{SS'} \left([R_0^{-1}]_{S'}\right)^{-1} [A^2]_{S'S}$, where S' is the set of latent nodes (complement of S). For diagonal and correlated noise structures, this leads to

$$
\begin{aligned}
\mathcal{E}_S^{(\text{diagonal})} &= A_{SS'} \left(I_{S'} - [A^2]_{S'}\right)^{-1} [A^2]_{S'S} \\
\mathcal{E}_S^{(\text{colored})} &= A_{SS'} \left([(\sum_{i=0}^\infty A^i \Sigma_x A^i)^{-1}]_{S'}\right)^{-1} [A^2]_{S'S}
\end{aligned}
\tag{5}
$$

The error $\mathcal{E}_S^{(\text{diagonal})}$ has been thoroughly studied in (Matta, Santos, and Sayed 2020) that shows structural consistency of the Granger estimator under partial observability and diagonal noise. However, even for Granger, there is no known analysis of $\mathcal{E}_S^{(\text{colored})}$. In fact, under colored noise and partial observability, Granger's performance degrades. □

In summary, the behavior of the error matrix $\mathcal{E}_S^{(n)}$ determines whether structural information is preserved in the observed time series or fundamentally lost: an estimator $F_S^{(n)}$ is structurally consistent whenever the error matrix is flat enough w.h.p. However, given a particular estimator, this error term responds differently to distinct regimes of i) connectivity of the underlying networked dynamical system, ii) observability, or iii) noise structure. This discussion shows the importance of two delicate steps in structure estimation: i) the characterization of the error term; and ii) the analysis of the *flatness* of the error. Indeed, what matters is the *variability* and not the (average) drift of its entries. It also alludes to the difficulty in handling the noise structure.

This paper addresses these questions. The next section characterizes the error term for the estimator $\widehat{R}_1(n) - \widehat{R}_3(n)$ as a function of the covariance matrix Σ_x. The following section establishes a sufficient condition on Σ_x to guarantee that the error is flat enough, and so guaranteeing w.h.p. structural consistency of $\widehat{R}_1(n) - \widehat{R}_3(n)$. This provides a sufficient condition that fundamentally asserts when structural information is preserved in the observed time series data in the regime of partial observability and colored noise.

Structural Consistency: Error Characterization

The estimator $\widehat{R}_1(n) - \widehat{R}_3(n)$ was first proposed and studied recently in (Chen, Wang, and Shen 2022) under diagonal noise. In contrast, we characterize the limiting error term of

this estimator under very broad conditions of full and partial observability and correlated noise, see Theorem 1 below. This will be used in the next section to develop novel conditions of structural consistency under partial observability and colored noise for the networked dynamical system (1). These are conditions on the covariance matrix Σ_x that if verified guarantee that structural information is preserved and can be recovered from the observed time series.

Before proceeding to the main result, observe that, given a covariance matrix $\Sigma_x \in \mathbb{S}_+^N$ satisfying Assumptions 2 and 3, we can uniquely represent it as

$$
\Sigma_x := \sigma_{\text{gap}}^2 I + \beta \mathbf{11}^\top + \overline{\Sigma}
\tag{6}
$$

where we have defined $\sigma_{\text{gap}}^2 \triangleq \sigma^2 - \max_{i \neq j} \mathbb{E}[\mathbf{x}_i \mathbf{x}_j] > 0$ and $\beta \triangleq \frac{1}{N(N-1)} \mathbf{1}^\top \mathsf{Off}(\Sigma_x)$, with $\mathsf{Off}(\Sigma_x)$ representing the vector collecting the off-diagonal entries of the matrix Σ_x. Thus, β stands for the average of the off-diagonals of Σ_x.

The representation in equation (6) decomposes the covariance into: i) a diagonal matrix $\sigma_{\text{gap}}^2 I$; ii) the average *offset* matrix $\beta \mathbf{11}^\top$; and iii) the matrix $\overline{\Sigma}$ containing the variability of the off-diagonal entries of Σ_x. This will be useful.

Theorem 1 (Error Characterization). *Under Assumptions* 1, *2, and* 3, *for the NDS* (1), *we have*

$$
\frac{1}{\sigma_{\text{gap}}^2} \left(\widehat{R}_1(n) - \widehat{R}_3(n)\right) \xrightarrow{n \to \infty} A + \mathcal{E},
\tag{7}
$$

with

$$
\mathcal{E} \triangleq \frac{1}{\sigma_{\text{gap}}^2} \left[\beta \rho \mathbf{11}^\top + (I - A^2)\left(\sum_{i=0}^\infty A^{i+1} \overline{\Sigma} A^i\right)\right],
\tag{8}
$$

where the convergence holds in probability. Accordingly, under partial observability, we have

$$
\frac{1}{\sigma_{\text{gap}}^2} \left(\left[\widehat{R}_1(n)\right]_S - \left[\widehat{R}_3(n)\right]_S\right) \xrightarrow{n \to \infty} A_S + \mathcal{E}_S.
\tag{9}
$$

Proof. Equation (7) is entrywise, so (9) follows directly from (7). We prove (7). Recall that the limit covariance matrices are given in terms of the interaction matrix A as $R_0 \triangleq \lim_{n \to \infty} R_0(n) = \sum_{i=0}^\infty A^i \Sigma_x A^i$ (solution to the Lyapunov equation) and $R_k = A^k R_0$ for all $k \in \mathbb{N}$. These expressions for R_0 and R_k, with representation (6), lead to

$$
\begin{cases}
R_1 = \sigma_{\text{gap}}^2 \sum_{i=0}^\infty A^{2i+1} + \beta \sum_{i=0}^\infty A^{i+1} \mathbf{11}^\top A^i \\
\qquad\qquad + \sum_{i=0}^\infty A^{i+1} \overline{\Sigma} A^i \\
R_3 = \sigma_{\text{gap}}^2 \sum_{i=0}^\infty A^{2i+3} + \beta \sum_{i=0}^\infty A^{i+3} \mathbf{11}^\top A^i \\
\qquad\qquad + \sum_{i=0}^\infty A^{i+3} \overline{\Sigma} A^i,
\end{cases}
\tag{10}
$$

Now, $\sigma_{\text{gap}}^2 \sum_{i=0}^\infty A^{2i+1} - \sigma_{\text{gap}}^2 \sum_{i=0}^\infty A^{2i+3} = \sigma_{\text{gap}}^2 A$ and

$$
\beta \left(\sum_{i=0}^\infty A^{i+1} \mathbf{11}^\top A^i - \sum_{i=0}^\infty A^{i+3} \mathbf{11}^\top A^i\right)
$$

$$
= \beta (I - A^2) \left(\sum_{i=0}^\infty A^{i+1} \mathbf{11}^\top A^i\right)
$$

$$
= \beta (I - A^2) A \mathbf{11}^\top (\sum_{i=0}^\infty \rho^{2i}) = \frac{\beta \rho (1 - \rho^2)}{(1 - \rho^2)} \mathbf{11}^\top = \beta \rho \mathbf{11}^\top.
$$

Finally, the last terms in R_1 and R_3 lead to the second term in (8). This shows the limits (7) (for full observability) and (9) (for partial observability).

Define $R_k = \lim_{n\to\infty} R_k(n) = \lim_{n\to\infty} \mathbb{E}\left[\mathbf{y}(n+k)\mathbf{y}(n)^\top\right]$. By modifying the proof for diagonal noise $\Sigma_x = \sigma^2 I$ in (Chen, Wang, and Shen 2022), we can show that, for general (spatially) correlated noise Σ_x, but independent across time, the empirical covariances $\widehat{R}_1(n)$ or $\widehat{R}_3(n)$ converge in probability to the exact limit covariances R_1 and R_3. $\qquad\square$

Theorem 1 characterizes the error term in terms of the covariance matrix Σ_x. As a Corollary to Theorem 1, if Σ_x is *flat*, i.e., $\Sigma_x = \sigma_{\mathsf{gap}}^2 I + \beta \mathbf{1}\mathbf{1}^\top$, then the error term associated with the estimator $\widehat{R}_1(n) - \widehat{R}_3(n)$ is flat as well, implying that the estimator is structurally consistent w.h.p.

Theorem 1 further offers an intuitive interpretation for the impact of the noise structure Σ_x on the causal inference problem. In particular, the hyperparameter β controls the drift or shift that the underlying ground truth matrix A_S suffers under the estimation $\widehat{R}_1(n) - \widehat{R}_3(n)$. On the other hand, the *oscillation* of the off diagonal entries given in $\overline{\Sigma}$ implies a perturbation on the entries of A_S, possibly compromising structural consistency. Fig. 2 offers an illustration.

Features separability. Lemma 1 in (Machado et al. 2023) establishes that structural consistency of $\widehat{R}_1(n) - \widehat{R}_3(n)$ implies that the set of features $\{\mathcal{F}_{ij}\}_{i\neq j}$

$$\mathcal{F}_{ij}(n) = \left([R_D(n)]_{ij}, [R_{D+1}(n)]_{ij}, \dots, [R_M(n)]_{ij}\right), \tag{11}$$

with $D \leq 1$ and $M \geq 3$ are affinely separable, i.e., there exists a particular hyperplane (up to a shift) that consistently partitions this set. The previous discussion translates into the feature domain (in view of Lemma 1 in (Machado et al. 2023)): β controls the *drift* of the features separating hyperplane (compromising stability), while the *variation* $\overline{\Sigma}$ of the off-diagonal entries of Σ_x induces a spread of the features.

Identifiability under Noise Structure

The next theorem provides a sufficient condition on the noise covariance Σ_x and the interaction matrix A that lead to the linear separability and stability of the (centered) features.

Before proceeding to our main theorem, define $\mathsf{Osc}(\mathbf{z}) = \max_i z_i - \min_i z_i$ as the difference between the maximum and minimum entries of the vector $\mathbf{z}$ – we shall loosely refer to it as *oscillation*. We state some properties of the map $\mathsf{Osc} : \mathbb{R}^N \longrightarrow \mathbb{R}_+$ that will be useful to establish the result.

Property 1. [Contraction] Given a symmetric stochastic matrix $\overline{A}$, we have $\mathsf{Osc}\left(\overline{A}v\right) \leq \mathsf{Osc}(v)$, as each entry of $\overline{v} \triangleq \overline{A}v$ lies in the convex hull (Hiriart-Urruty and Lemaréchal 2001) of the set $\{v_1, v_2, \dots, v_N\}$ of the entries of $v \in \mathbb{R}^N$ and, in particular, $\overline{v}_i \in [v_{\min}, v_{\max}]$ for all i.

Property 2. [Scalar linearity] Observe that $\mathsf{Osc}(\alpha v) = |\alpha|\,\mathsf{Osc}(v)$ for any $v \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$. Indeed, if $\alpha > 0$, then we have $\alpha v_{\max} > \alpha v_i$ for any $i = 1, 2, \dots, N$.

Property 3. [Subadditivity] The map $\mathsf{Osc}(\cdot)$ obeys $\mathsf{Osc}(B + C) \leq \mathsf{Osc}(B) + \mathsf{Osc}(C)$, for any $B, C \in \mathbb{R}^N$.
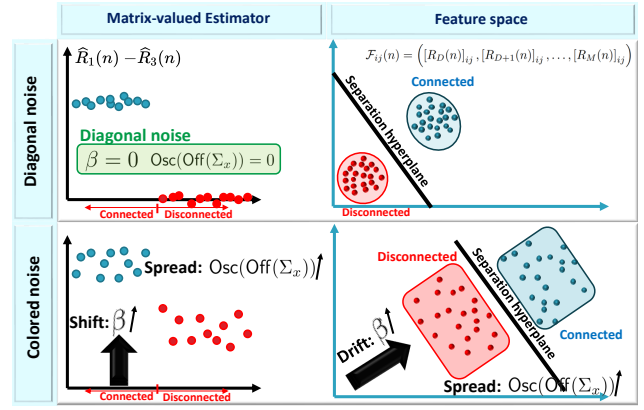

Figure 2: Impact of the correlation structure of the noise on the features. The overarching idea (Theorem 1) is that the average β across the off-diagonals of the covariance Σ_x yields a drift of the features, while the *oscillation* across the off-diagonals of Σ_x leads to a greater *spread* of the features.

Property 4. [Submultiplicativity] The map $\mathsf{Osc}(\cdot)$ obeys: If $\mathsf{Osc}(Bv) \leq k_b\mathsf{Osc}(v)$ and $\mathsf{Osc}(Cv) \leq k_c\mathsf{Osc}(v)$ for all $v \in \mathbb{R}^N$ then, $\mathsf{Osc}(CBv) \leq k_c\mathsf{Osc}(Bv) \leq k_bk_c\mathsf{Osc}(v)$ with $k_b, k_c > 0$.

Let $\mathcal{C}(\mathcal{K}) \triangleq \frac{1}{|\mathcal{K}|}\sum_{k\in\mathcal{K}} k$ be the centroid of the set $\mathcal{K}$.

Theorem 2 (Linear Separability & Stability). *Let $A = \rho\overline{A}$, where $\overline{A}$ is a stochastic matrix with $0 < \rho < 1$. Under Assumptions 2 and 3, if*

$$\frac{\mathsf{Osc}\left(\mathsf{Off}(\Sigma_x)\right)}{\sigma_{\mathsf{gap}}^2} \leq \frac{A_{\min}^+(1 - \rho^2)}{2\rho(\rho^2 + 1)}, \tag{12}$$

where $\sigma_{\mathsf{gap}}^2 \triangleq \sigma^2 - \max_{i\neq j}\mathbb{E}[\mathbf{x}_i\mathbf{x}_j] > 0$ and $A_{\min}^+$ is the smallest nonzero entry of the interaction matrix A, then the centered features $\{\mathcal{F}_{ij}(n)\}_{i\neq j} - \mathcal{C}\left(\{\mathcal{F}_{ij}(n)\}_{i\neq j}\right)$ are linearly separable and stable.

Proof. The properties of the map $\mathsf{Osc}(\cdot)$ yield

$$
\begin{aligned}
\mathsf{Osc}\left((I - A^2)V\right) &\leq \mathsf{Osc}(V) + \mathsf{Osc}(A^2 V) \\
&\leq \mathsf{Osc}(V) + \rho^2\mathsf{Osc}\left(\overline{A}^2 V\right), \\
&\leq (1 + \rho^2)\,\mathsf{Osc}(V)
\end{aligned} \tag{13}
$$

where the first inequality holds in view of Properties 2 and 3; the second inequality follows from Property 3; and the last inequality stems from Properties 1 and 4. Therefore, we have

$$
\begin{aligned}
\mathsf{Osc}\left((I - A^2)AV\right) &= \rho\,\mathsf{Osc}\left((I - A^2)\overline{A}V\right) \\
&\leq \rho(1 + \rho^2)\,\mathsf{Osc}(V).
\end{aligned} \tag{14}
$$

Now, observe that $\sum_{i=0}^\infty A^i\overline{\Sigma}A^i$ is bounded above and below as follows

$$
\frac{\overline{\Sigma}_{\min}}{1 - \rho^2}\mathbf{1}\mathbf{1}^\top = \overline{\Sigma}_{\min}\sum_{i=0}^\infty A^i\mathbf{1}\mathbf{1}^\top A^i \leq \sum_{i=0}^\infty A^i\overline{\Sigma}A^i \leq
$$

$$
\overline{\Sigma}_{\max}\sum_{i=0}^\infty A^i\mathbf{1}\mathbf{1}^\top A^i = \frac{\overline{\Sigma}_{\max}}{1 - \rho^2}\mathbf{1}\mathbf{1}^\top,
$$

and thus, $\mathsf{Osc}\left(\sum_{i=0}^{\infty} A^i \overline{\Sigma} A^i\right) \leq \frac{(\overline{\Sigma}_{\max} - \overline{\Sigma}_{\min})}{1-\rho^2}$. Therefore,

$$
\begin{aligned}
\mathsf{Osc}\left(\mathcal{E}\right) &= \mathsf{Osc}\left(\left(I - A^2\right) A \sum_{i=0}^{\infty} A^i \overline{\Sigma} A^i\right) \\
&\leq \frac{\rho(1+\rho^2)}{1-\rho^2}\left(\overline{\Sigma}_{\max} - \overline{\Sigma}_{\min}\right) \\
&= \frac{\rho(1+\rho^2)}{1-\rho^2}\mathsf{Osc}\left(\mathsf{Off}(\overline{\Sigma}_x)\right).
\end{aligned}
\tag{15}
$$

In view, of inequality (4) and Theorem 1, we have structural consistency (or features separability) whenever $\mathsf{Osc}\left(\mathcal{E}\right) \leq \frac{\sigma_{\text{gap}}^2 A_{\min}^+}{2}$, which holds whenever $\frac{\rho(1+\rho^2)}{1-\rho^2}\mathsf{Osc}\left(\mathsf{Off}(\overline{\Sigma}_x)\right) \leq \frac{\sigma_{\text{gap}}^2 A_{\min}^+}{2}$, in view of inequality (15). $\qquad\square$

Remark 1 (Exogenous interventions). *Under exogenous interventions to the NDS, i.e.,* $\mathbf{y}(n+1) = A\mathbf{y}(n) + \mathbf{x}(n+1) + \xi(n+1)$, *where* $\xi(n) \sim \mathcal{N}(0, \sigma_\xi^2)$ *is i.i.d., and independent of the process* $(\mathbf{x}(n))$, *equation* (12) *becomes* $\frac{\mathsf{Osc}(\mathsf{Off}(\Sigma_x))}{\sigma_{\text{gap}}^2 + \sigma_\xi^2} \leq \frac{A_{\min}(1-\rho^2)}{2\rho(\rho^2+1)}$, *implying that regardless of the covariance matrix of the input process* $(\mathbf{x}(n))$, *the features are linearly separable w.h.p. as soon as the level (or variance)* σ_ξ^2 *of the external intervention* $(\xi(n))$ *is high enough.*

We recall that due to the sensitivity of the separating hyperplane on the color of the noise, the method deployed in (Machado et al. 2023) degrades as the average of the absolute values of the off-diagonal entries of Σ_x increases (captured by β). To mitigate the shift of the separating hyperplane and allow the deployment of supervised methods, we renormalize the features applying *standard scaler*. Further, we observed empirically that incorporating novel matrix-valued estimators into the vector of features $\mathcal{F}_{ij}(n)$ of each pair ij, greatly helps to mitigate both the drift and the *oscillation* in the features. This is introduced next.

Novel Features

Define the set of features $\mathcal{T}^{(n)} = \left\{\mathcal{T}_{ij}^{(n)}\right\}_{ij}$

$$
\mathcal{T}_{ij}^{(n)} = \left(\left[\left(\left[\widehat{R}_0(n)\right]_S\right)^{-1}\right]_{ij}, \ldots, \left[\left(\left[\widehat{R}_M(n)\right]_S\right)^{-1}\right]_{ij}\right),
$$

built from the inverse of the observed part of the lag moments. We have empirically observed that a certain normalized version of these features exhibits nontrivial separability and stability properties across a broad range of connectivity regimes and noise correlation levels β. Namely, coupled with the features $\mathcal{F}^{(n)}$, see equation (11), they aid in mitigating the *oscillation* and drift caused by the noise structure. More concretely, define $\mathcal{K}_{ij}^{(n)} = \left(\mathcal{F}_{ij}^{(n)}, \mathcal{T}_{ij}^{(n)}\right)$ for each ij. Then, given Lemma 2 in (Machado et al. 2023), the identifiability gap separating the features of connected pairs from features of disconnected pairs (refer to its formal definition in (Machado et al. 2023)) is larger for the features $\mathcal{K}^{(n)}$ than for either the features $\mathcal{T}^{(n)}$ or $\mathcal{F}^{(n)}$ w.h.p. under the sufficient condition in Theorem 2. This tends to make the features $\mathcal{K}^{(n)}$ more amenable to generalization when using supervised methods to cluster them.

In summary, the proposed features lead to the appropriate identifiability properties if either the variability of the off-diagonal entries of Σ_x is not too large (Theorem 2) or the exogenous intervention is large enough (Remark 1).

Methodology

We adopt the following training set

$$
\mathsf{Tr}^{(n)} \triangleq \left\{\left(\overline{\mathcal{K}}_{ij}^{(n)}, \mathbf{1}_{\{A_{ij} \neq 0\}}\right)\right\}_{i \neq j},
\tag{16}
$$

where we define the *normalized* features $\overline{\mathcal{K}}_{ij}^{(n)} := \left[\mathsf{SC}\left(\mathcal{K}^{(n)}\right)\right]_{ij}$, with SC denoting the *standard-scaler*, i.e., the set $\mathcal{K}^{(n)} = \left\{\mathcal{K}_{ij}^{(n)}\right\}_{i \neq j}$ is centered and the entries of each feature $\mathcal{K}_{ij}^{(n)}$ are normalized to meet unit variance as described in (Pedregosa et al. 2011). The standard-scaler normalization helps to mitigate the drift of the features caused by the average noise cross-correlation β. That is, for training, we provide a normalized set of features $\overline{\mathcal{K}}^{(n)}$, with $\overline{\mathcal{K}}_{ij}^{(n)}$ being the feature of the pair ij used as input to a Feed Forward Neural Network (FFNN) whose training output should be the ground truth connectivity of the pair ij, i.e., $\mathbf{1}_{\{A_{ij} \neq 0\}}$.

We generate the time series data $\{\mathbf{y}(i)\}_{i=0}^n$ as done in (Machado et al. 2023), i.e., we first build a graph G on N nodes and then assign weights to the graph G to obtain an $N \times N$ interaction matrix A that is nonnegative and stable, i.e., $\rho(A) < 1$. Namely, we create G as a realization of an Erdős–Rényi random graph – without self-loops, i.e., $G_{ii} = 0$ for all i – with N nodes and probability of edge or arrow placing p. The value of p controls the sparsity of the underlying network. Then, the interaction matrix A is characterized by: $A_{ij} = \alpha \frac{G_{ij}}{d_{\max}(G)}$, for $i \neq j$; and $A_{ii} = \rho - \sum_{k \neq i} A_{ik}$, for all i where $d_{\max}(G)$ is the maximum *inwards*-degree of the underlying graph G and α, ρ are stability hyperparameters with $0 < \alpha \leq \rho < 1$. This weight assignment is often referred to as *Laplacian rule* (Chung 1994). This generative process renders the NDS (1) stable and with a support graph of interactions given by G – the main object of inference in our framework. To generate a directed graph G, we considered the realization of the Erdős–Rényi version for directed graphs also known as Binomial random graph model.

We train FFNNs over one realization of a random graph model with $p = 0.5$ and $N = 50$ and apply this trained FFNN to networked systems with distinct connectivities obtained with different p. Further, for training, we generate datasets with distinct average noise cross-correlation levels $\beta = 0, 5, 10, \ldots, 50$. Remarkably, our method generalizes to distinct connectivity regimes and different noise correlation levels as seen in the next Section.

Simulation Results

Fig. 3 displays the performance of distinct structure estimation methods. The performance is captured by the *accuracy* of the graph structure estimation as a function of the number of samples of the time series, known as empirical sample complexity. The accuracy is determined by the ratio of the
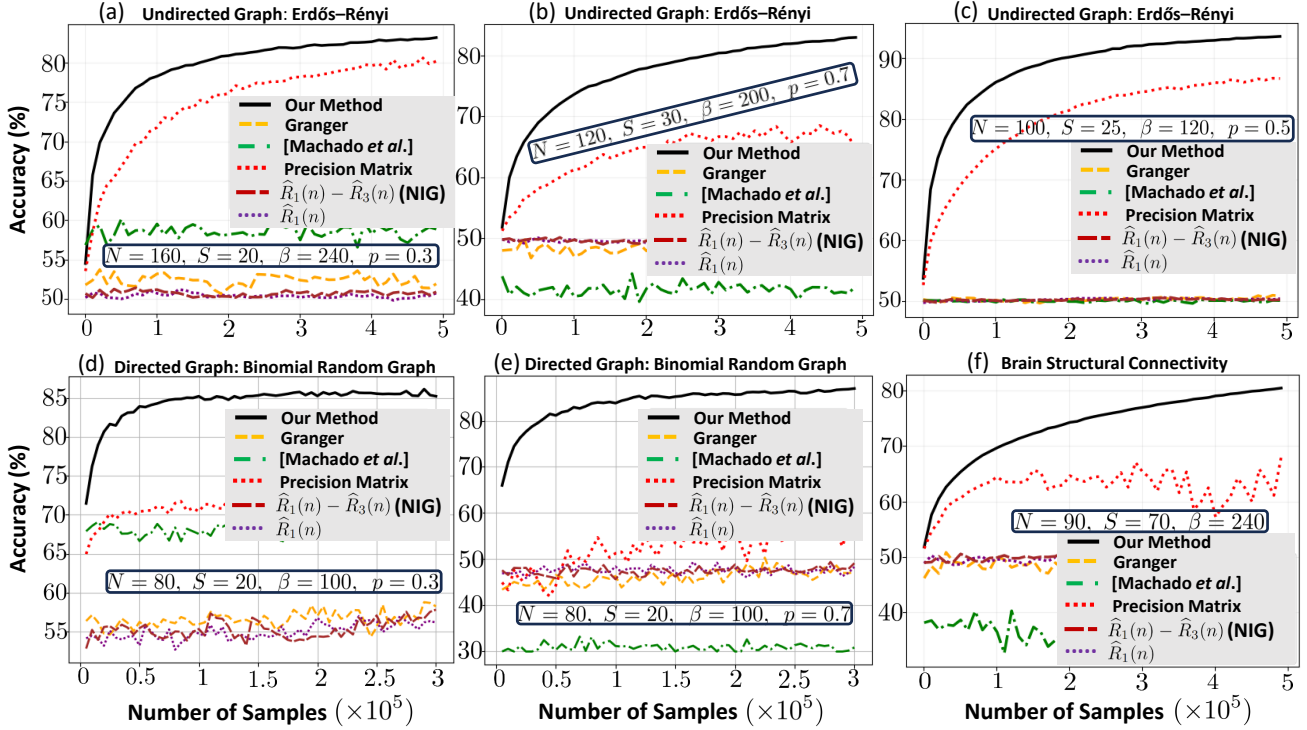
Figure 3: Plots (a)-(f) contrast the accuracy of distinct methods as a function of the number of time series samples: i) (a)-(c) undirected networks (graph learning); (d)-(e) directed networks (causal inference); and (f) refers to a directed Brain connectome matrix. The hyperparameters N, S, β and p displayed framed in the plots entail the underlying regime.

number of *directed* pairs correctly classified by each method over the total number of directed pairs comprising the graph.

We have chosen to compare our causal inference method with: i) the recently proposed feature based approach (Machado et al. 2023); ii) the Granger estimator (Geiger et al. 2015; Santos, Matta, and Sayed 2020); iii) the one-lag $\widehat{R}_1(n)$ estimator (Matta, Santos, and Sayed 2022); iv) the NIG $\widehat{R}_1(n) - \widehat{R}_3(n)$ estimator (Chen, Wang, and Shen 2022); v) the Precision matrix or Graphical Lasso for sparse regimes (interchangeably) (Friedman, Hastie, and Tibshirani 2007). While other graph learning or causal inference methods are optimal for specific regimes of observability or connectivity (mostly for *very* sparse networks), the methods $ii) - v)$ we choose as benchmarks are popular and can be applied to partial observability (with technical guarantees, e.g., (Mei and Moura 2018; Matta, Santos, and Sayed 2020)) and across various connectivity graphs (dense or sparse networks), thus, ideal to be contrasted with our method across distinct regimes of observability and connectivity. As in (Chen, Wang, and Shen 2022), we deploy Gaussian mixture on the estimators $ii) - v)$ to recover the underlying graph from the estimated affinity matrix.

Fig. 3 demonstrates the overall performance superiority of our method in the colored regimes considered. Fig. 3 (f) refers to a human brain *connectome*, obtained via tractography (Škoch et al. 2022). In this latter case, we generate the time series $\{\mathbf{y}(n)\}_{n=1}^{M}$ for this real world network to in-

fer its connectivity. The regimes (a)-(c) refer to undirected graphs obtained as realizations of an Erdős–Rényi random graph model, and (d),(e) refer to directed graphs, with the networks drawn from a Binomial random graph model.

The hyperparameters N, S, β and p represent the number of nodes in the NDS, number of observed nodes, average cross-correlation of the noise (average of the off-diagonal entries of Σ_x), and the probability of edge or arrow placing in the corresponding Erdős–Rényi or Binomial random graph model (where applicable). Additional experiments for undirected graphs are in (Santos et al. 2024).

Concluding Remarks

We studied the problem of network identification from the times series data streaming from a linear networked dynamical system (NDS). We considered the challenging framework of partial observability under colored noise. We first devised a novel identifiability condition for linear NDS under this regime. Then, departing from the scalar based methods that are dominant in causal inference, we considered a feature based paradigm: to each pair of nodes, we assign a feature vector. We studied the impact of the noise structure on the identifiability properties of the features of (Machado et al. 2023) to propose novel features. We trained FFNNs with our features to cluster them and recover the causal network structure. The numerical experiments across distinct regimes of connectivity, observability and noise correlation demonstrate the competitive performance of our approach.

Acknowledgments

References

Aalto, A.; Viitasaari, L.; Ilmonen, P.; Mombaerts, L.; and Goncalves, J. 2020. Gene regulatory network inference from sparsely sampled noisy data. *Nature Communications*, 11(3493).

Anandkumar, A.; Hsu, D.; Javanmard, A.; and Kakade, S. 2013. Learning Linear Bayesian Networks with Latent Variables. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, 249–257. Atlanta, Georgia, USA: PMLR.

Anandkumar, A.; and Valluvan, R. 2013. Learning loopy graphical models with latent variables: Efficient methods and guarantees. *Ann. Statist.*, 41(2): 401–435.

Barbier, J.; Camilli, F.; Mondelli, M.; and Sáenz, M. 2023. Fundamental limits in structured principal component analysis and how to reach them. *Proceedings of the National Academy of Sciences of the United States of America*, 120.

Barbier, J.; Krzakala, F.; Macris, N.; Miolane, L.; and Zdeborova, L. 2019. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116: 201802705.

Barrat, A.; Barthélemy, M.; and Vespignani, A. 2008. *Dynamical Processes on Complex Networks*. Cambridge University Press.

Barreira, L.; and Valls, C. 2012. *Ordinary Differential Equations: Qualitative Theory*. Graduate studies in mathematics. American Mathematical Society. ISBN 9780821887493.

Cao, D.; Enouen, J.; Wang, Y.; Song, X.; Meng, C.; Niu, H.; and Liu, Y. 2023. Estimating Treatment Effects from Irregular Time Series Observations with Hidden Confounders. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press. ISBN 978-1-57735-880-0.

Casadiego, J.; Nitzan, M.; Hallerberg, S.; and Timme, M. 2017. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature Communications*, 8(2192).

Chen, Y.; Wang, Z.; and Shen, X. 2022. An Unbiased Symmetric Matrix Estimator for Topology Inference under Partial Observability. *IEEE Signal Processing Letters*, 29(02): 1257–1261.

Cheng, Y.; Yang, R.; Xiao, T.; Li, Z.; Suo, J.; He, K.; and Dai, Q. 2023. CUTS: Neural Causal Discovery from Irregular Time-Series Data. In *The Eleventh International Conference on Learning Representations (ICLR)*.

Ching, E. S. C.; and Tam, H. C. 2017. Reconstructing links in directed networks from noisy dynamics. *Phys. Rev. E*, 95: 010301.

Chow, C.; and Liu, C. 1968. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3): 462–467.

Chung, F. 1994. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series. Conference Board of the Mathematical Sciences. ISBN 9780821889367.

Douw, L.; De Groot, M.; Dellen, E.; Heimans, J.; Ronner, H.; Stam, C.; and Reijneveld, J. 2010. 'Functional Connectivity' is a Sensitive Predictor of Epilepsy Diagnosis after the First Seizure. *PloS one*, 5.

Eroglu, D.; Tanzi, M.; van Strien, S.; and Pereira, T. 2020. Revealing Dynamics, Communities, and Criticality from Data. *Phys. Rev. X*, 10: 021047.

Friedman, J.; Hastie, T.; and Tibshirani, R. 2007. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3): 432–441.

Ganesh, A.; Massoulié, L.; and Towsley, D. 2005. The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, 1455–1466.

Geiger, P.; Zhang, K.; Schölkopf, B.; Gong, M.; and Janzing, D. 2015. Causal Inference by Identification of Vector Autoregressive Processes with Hidden Components. In *Proc. International Conference on Machine Learning*, volume 37, 1917–1925.

Hiriart-Urruty, J.-B.; and Lemaréchal, C. 2001. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer-Verlag Berlin Heidelberg. ISBN 978-3-540-42205-1.

Hossain, R. R.; Adesunkanmi, R.; and Kumar, R. 2023. Data-Driven Linear Koopman Embedding for Networked Systems: Model-Predictive Grid Control. *IEEE Systems Journal*, 17(3): 4809–4820.

Huang, H.; and Ding, M. 2016. Linking Functional Connectivity and Structural Connectivity Quantitatively: A Comparison of Methods. *Brain Connectivity*, 6(2): 99–108.

Lahmanovich, A.; and James, A. Y. 1976. A Deterministic Model for Gonorrhea in a Nonhomogeneous Population. *Mathematical Biosciences*, 28: 221–236.

Liégeois, R.; Santos, A.; Matta, V.; Van De Ville, D.; and Sayed, A. H. 2020. Revisiting correlation-based functional connectivity and its relationship with structural connectivity. *Network Neuroscience*, 4(4): 1235–1251.

Loh, P.-L.; and Wainwright, M. J. 2013. Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *The Annals of Statistics*, 41(6): 3022 – 3049.

Lusch, B.; Kutz, J.; and Brunton, S. 2018. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9.

Machado, S.; Sridhar, A.; Gil, P.; Henriques, J.; Moura, J. M. F.; and Santos, A. 2023. Recovering the Graph Underlying Networked Dynamical Systems under Partial-observability: a Deep Learning Approach. In *Proceedings of the 37th AAAI Coference on Artificial Intelligence*. AAAI.

Marcinkevičs, R.; and Vogt, J. E. 2021. Interpretable Models for Granger Causality Using Self-explaining Neural Networks. In *International Conference on Learning Representations*.

Materassi, D.; and Salapaka, M. V. 2012. Network reconstruction of dynamical polytrees with unobserved nodes. In *Proc. IEEE Conference on Decision and Control (CDC)*, 4629–4634. Maui, Hawaii.

Materassi, D.; and Salapaka, M. V. 2015. Identification of network components in presence of unobserved nodes. In *Proc. IEEE Conference on Decision and Control (CDC)*, 1563–1568. Osaka, Japan.

Matta, V.; Santos, A.; and Sayed, A. H. 2020. Graph Learning under Partial Observability. *Proceedings of the IEEE*, 108: 2049 – 2066.

Matta, V.; Santos, A.; and Sayed, A. H. 2022. Graph Learning over Partially Observed Diffusion Networks: Role of Degree Concentration. *IEEE Open Journal of Signal Processing*, 335–371.

Mauroy, A.; and Goncalves, J. 2016. Linear identification of nonlinear systems: A lifting technique based on the Koopman operator. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 6500–6505. Las Vegas, USA.

Mei, J.; and Moura, J. M. F. 2018. SILVar: Single Index Latent Variable Models. *IEEE Transactions on Signal Processing*, 66(11): 2790–2803.

Melnychuk, V.; Frauen, D.; and Feuerriegel, S. 2022. Causal Transformer for Estimating Counterfactual Outcomes. In *ICML*, 15293–15329.

Montanari, A.; and Pereira, J. 2009. Which graphical models are difficult to learn? In *Advances in Neural Information Processing Systems*, volume 22. Vancouver, Canada.

Passemiers, A.; Moreau, Y.; and Raimondi, D. 2022. Fast and accurate inference of gene regulatory networks through robust precision matrix estimation. *Bioinformatics*, 38(10): 2802–2809.

Pearl, J. 2009. *Causality*. Cambridge University Press, 2 edition.

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

Pereira, J.; Ibrahimi, M.; and Montanari, A. 2010. Learning Networks of Stochastic Differential Equations. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc.

Peters, J.; Janzing, D.; and Schlkopf, B. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press. ISBN 0262037319.

Ren, J.; Wang, W.-X.; Li, B.; and Lai, Y.-C. 2010. Noise Bridges Dynamical Correlation and Topology in Coupled Oscillator Networks. *Phys. Rev. Lett.*, 104: 058701.

Ren, X.-L.; Gleinig, N.; Helbing, D.; and Antulov-Fantulin, N. 2019. Generalized network dismantling. *Proceedings of the National Academy of Sciences*, 116(14): 6554–6559.

Runge, J.; Bathiany, S.; Bollt, E. M.; Camps-Valls, G.; Coumou, D.; Deyle, E. R.; Glymour, C.; Kretschmer, M.; Mahecha, M. D.; Muñoz-Marí, J.; van Nes, E. H.; Peters, J.; Quax, R.; Reichstein, M.; Scheffer, M.; Scholkopf, B.; Spirtes, P.; Sugihara, G.; Sun, J.; Zhang, K.; and Zscheischler, J. 2019. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10.

Santos, A.; Matta, V.; and Sayed, A. H. 2020. Local Tomography of Large Networks under the Low-Observability Regime. *IEEE Transactions on Information Theory*, 66: 587 – 613.

Santos, A.; Rente, D.; Seabra, R.; and Moura, J. M. F. 2024. Inferring the Graph of Networked Dynamical Systems under Partial Observability and Spatially Colored Noise. In *The IEEE International Conference on Acoustics, Speech and Signal Processing, 2024 (Accepted)*.

Shaikh Veedu, M.; and Salapaka, M. V. 2023. Topology identification under spatially correlated noise. *Automatica*, 156: 111182.

Spirtes, P. 2001. An Anytime Algorithm for Causal Inference. In Richardson, T. S.; and Jaakkola, T. S., eds., *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics*, volume R3 of *Proceedings of Machine Learning Research*, 278–285. PMLR. Reissued by PMLR on 31 March 2021.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT press, 2nd edition.

Wan, X.; Liu, J.; Cheung, K.-W.; and Tong, T. 2014. Inferring Epidemic Network Topology from Surveillance Data. *PloS one*, 9: e100661.

Škoch, A.; Rehák Bučková, B.; Mareš, J.; Tintera, J.; Sanda, P.; Jajcay, L.; Horacek, J.; Spaniel, F.; and Hlinka, J. 2022. Human brain structural connectivity matrices–ready for modelling. *Scientific Data*, 9: 486.