

## RESEARCH ARTICLE

# Inferring Topology of Networks With Hidden Dynamic Variables

RAOUL SCHMIDT<sup>1,2</sup>, HAUKE HAEHNE<sup>3</sup>, LAURA HILLMANN<sup>4</sup>, JOSE CASADIEGO<sup>1,2</sup>, DIRK WITTHAUT<sup>5,6</sup>, BENJAMIN SCHÄFER<sup>1,2,7,8</sup>, AND MARC TIMME<sup>1,2,9</sup>

<sup>1</sup>Chair for Network Dynamics, Center for Advancing Electronics Dresden (cfaed), Technische Universität of Dresden, 01062 Dresden, Germany

<sup>2</sup>Institute for Theoretical Physics, Technische Universität of Dresden, 01062 Dresden, Germany

<sup>3</sup>Institute of Physics and ForWind, Carl von Ossietzky University Oldenburg, 26129 Oldenburg, Germany

<sup>4</sup>GFZ Potsdam German Research Centre for Geosciences, 14473 Potsdam, Germany

<sup>5</sup>Forschungszentrum Jülich, Institute for Energy and Climate Research—Systems Analysis and Technology Evaluation (IEK-STE), 52428 Jülich, Germany

<sup>6</sup>Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany

<sup>7</sup>School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, U.K.

<sup>8</sup>Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, 76344 Eggenstein-Leopoldshafen, Germany

<sup>9</sup>Lakeside Laboratories, 9020 Klagenfurt, Austria

Corresponding author: Marc Timme (marc.timme@tu-dresden.de)

This work was supported in part by the Research Cluster Center for Advancing Electronics Dresden (cfaed), Technische Universität (TU) Dresden; in part by the German Federal Ministry for Research and Education through Bundesministerium für Bildung und Forschung (BMBF, German Federal Ministry for Research and Education) under Grant 03SF0472 and Grant 03EK3055; in part by the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie under Grant 840825; in part by the Helmholtz Association under Grant VH-NG-1727; and in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2068—390729961—Cluster of Excellence Physics of Life at TU Dresden.

**ABSTRACT** Inferring the network topology from the dynamics of interacting units constitutes a topical challenge that drives research on its theory and applications across physics, mathematics, biology, and engineering. Most current inference methods rely on time series data recorded from all dynamical variables in the system. In applications, often only some of these time series are accessible, while other units or variables of all units are hidden, i.e. inaccessible or unobserved. For instance, in AC power grids, frequency measurements often are easily available whereas determining the phase relations among the oscillatory units requires much more effort. Here, we propose a network inference method that allows to reconstruct the full network topology even if all units exhibit hidden variables. We illustrate the approach in terms of a basic AC power grid model with two variables per node, the local phase angle and the local instantaneous frequency. Based solely on frequency measurements, we infer the underlying network topology as well as the relative phases that are inaccessible to measurement. The presented method may be enhanced to include systems with more complex coupling functions and additional parameters such as losses in power grid models. These results may thus contribute towards developing and applying novel network inference approaches in engineering, biology and beyond.

**INDEX TERMS** Complex networks, inverse problems, nonlinear dynamics, network inference, power grids.

## I. INTRODUCTION

Many complex real world systems, ranging from technical supply networks to biological processes such as gene regulation, consist of a multitude of interacting units and thus represent network dynamical systems. To understand, predict and control the full network dynamics typically requires the

The associate editor coordinating the review of this manuscript and approving it for publication was Ludovico Minati.

knowledge of the topology defining how the units interact. However, the topology underlying the current system dynamics might be unavailable, e.g. because it is difficult or expensive to observe. In addition, several dynamical variables may not be easily accessible to measurement and thus hidden.

The inference of network topology from time series recordings or other network features [1]–[3] and inverse problems in general [4] are active fields of interdisciplinary research on complex systems. Inference methods are applied

for example to obtain gene and protein interaction networks [5], observe epidemic spreading on networks [6] or determine the topology of coupled oscillator networks [7]–[9]. Depending on the field of interest, the available data and the (mathematical) understanding of the underlying system, different inference methods prove viable. Inference methods range from Bayesian networks and information theoretical methods based on mutual information or maximal entropy [10], [11] over compressed sensing [12], [13] and deep learning approaches [14] to network inference from response dynamics [7], see [2], [3] for comprehensive reviews. Such methods are useful to gain a better understanding of the structure and function of the network at hand, whether it is natural or engineered. The procedure for many inference approaches is very similar: Observing a system out of equilibrium reveals the interplay of its different units and may enable us to infer the underlying topology. Mathematically, time series of the system are combined with prior knowledge, e.g. of the intrinsic dynamics at each node, to estimate the interaction matrix via pseudo-inverses [2].

Recent advances in network inference have focused for example on revealing the physical interactions, in contrast to pure correlations or adopted model free inference approaches without any knowledge of the intrinsic dynamics, see, e.g., [7], [15]–[20]. Recent work also indicates how to infer the number of nodes in the network from observing only a few accessible nodes [21]. However, most theoretical approaches still require the full knowledge of the time series from all variables of each given unit. Uncovering the interacting topology of networks from time series in the presence of hidden dynamical variables has been successful under certain conditions by heuristic means, see for instance [18]. Aside from applications of inference methods on biological networks, such as gene-regulation [22] or neuronal networks [23], or communication networks [24], [25], there are also several methods for parameter estimation in power systems, see [13] for an overview.

As paradigmatic example systems, we consider models of electrical power grids [26] as these constitute particularly interesting applications for inference methods, for instance because they may get damaged or experience line failures. Inference techniques then serve to detect these changes in topology quickly and help localizing the fault [27], [28]. Furthermore, the power grid topology is changing rapidly, both in industrialized as well as developing countries, due to the introduction of distributed generators, charging infrastructure for electric vehicles and other factors. Combined with the fact that different parts of one grid are often operated and administered by different grid operators, the precise topology is not always fully known, specifically also of lower-voltage distribution grids [29]. These lower voltage levels become increasingly relevant for grid operators and power dispatch strategies because the number of generators connected to the low-voltage grid is growing, e.g. by the increasing installation of solar panels [30]. The additional feed-in of power on lower voltage levels thereby supports power flow from lower levels

to higher ones, in addition to the traditional, reverse direction. Hence, in particular for power grids, it is essential to have access to their full topology on all scales to ensure stable planning and operation [31].

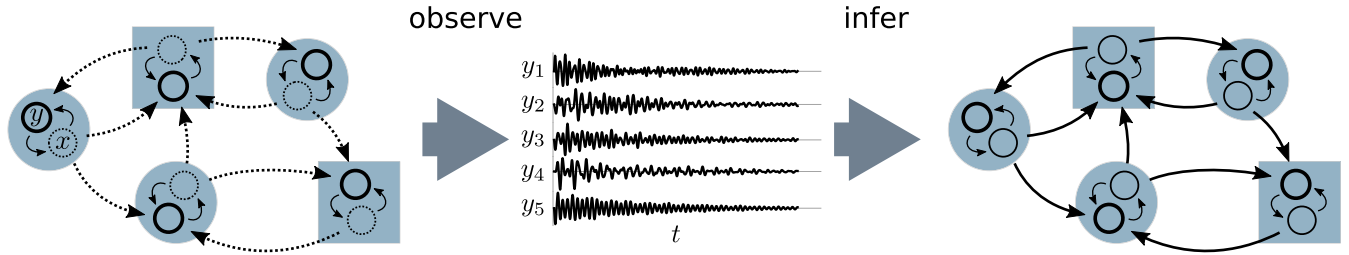
Applying inference on electrical power grids poses a special challenge. We typically have access to the voltage amplitude and frequency at each unit, since measurements are easy and cheap [32]. In contrast, determining the voltage angle (the oscillatory phase) via a PMU (phasor measurement unit) between different units [33] in a network is more expensive. Therefore, in many power grid applications, time series of voltage angles are not available. To put it in the general perspective of network dynamical systems, the time series of some variables of each unit are hidden. Simultaneously, knowledge about the voltage angle is valuable since the power flow along a transmission line is a direct function of the difference of voltage angles (phase differences), not the frequency [34].

Here we propose a method for revealing the interaction topology of networks with hidden dynamical variables as well as time series of the hidden variables themselves (relative to a base line). For instance, for an AC power grid model, the method only requires grid frequency measurements, not phase (voltage angle) or phase difference measurements, at all  $N$  nodes. We provide estimators for the network's underlying physical connectivity (which node is connected to which others) as well as power flows. We find that for a moderate number of measurements we obtain high prediction accuracy of both the network topology and the precise state of the network in terms of flows. Overall, our method is immediately applicable to power grid simulations and should also be useful for real-world power grid recordings if they follow a given model approximately. Furthermore, the same procedure may be extended to other applications in which certain variables are not directly accessible to measurements.

This article is structured as follows. We first introduce a class of model systems with two variables per unit, as well as a specific model approximately characterizing aspects of power grid dynamics. Further, we present the theoretical foundation to infer the network topology from time series available for one of the two variables (frequency-only measurements for the power grid model). We evaluate network reconstruction on differently sized network topologies and close with a discussion.

## II. THEORY

We start by explaining the theoretical concept and resulting procedures underlying the inference of network topology and the network state for systems where not all of the dynamical variables are recorded. We base our approach on the *direct method*, which has been introduced in [15] and is further discussed in [2], for inference tasks where the time series of all variables of all units are available. The *direct method* allows to infer a network's topology by separating the known and unknown terms of the dynamical system, evaluating the possibly nonlinear unit dynamics and coupling functions at



**FIGURE 1. Network inference with hidden variables.** Each node  $i$  in the network consists of two interacting dynamic variables ( $x_i$  and  $y_i$ ), one of which is hidden ( $x_i$ ). Nodal dynamics may differ between nodes, as long as they satisfy equations (1) and (2). In addition, the interactions between nodes are hidden as well. By recording the transient relaxation of the observable variable of each node after a small perturbation, we infer the topology as well as the differences between each pair of hidden variables, both in their initial conditions and in the fixed point.

a number of time points, and reformulating the problem as an optimization task of systems of linear equations. This linear system is overdetermined by recording state variables (or components of them) at sufficiently many time points at each unit and is then solved for the coupling matrix as a least squares problem (by calculating the Moore-Penrose pseudo-inverse).

Fig. 1 illustrates the concept. For a network dynamical system with each node  $i$  consisting of two interacting dynamical variables  $x_i$  and  $y_i$ , hidden topology and hidden time series  $x_i(t)$  for all  $i \in \{1 \dots N\}$ , the proposed method aims to infer both network topology and the time series of the hidden dynamic variables based on the observed time series of  $y_i$  and known dynamics, as detailed below.

**A. CLASS OF MODEL SYSTEMS**

Throughout this work, we focus on a class of simple network dynamical systems

$$\dot{x}_i = f_i(\mathbf{y}, t) \tag{1}$$

$$\dot{y}_i = g_i(\mathbf{y}, t) + h_i(\mathbf{x}) \tag{2}$$

with two variables ( $x_i$  and  $y_i$ ) per unit  $i \in \{1, \dots, N\}$ . The vectors  $\mathbf{x}(t) = (x_1(t), \dots, x_N(t))^T$  and  $\mathbf{y}(t) = (y_1(t), \dots, y_N(t))^T$  together make up the state of the entire systems at time  $t$ . The (possibly nonlinear) functions  $f_i$  and  $g_i$  describe the nodal dynamics, whereas the  $h_i$  mediate the coupling between units and the two variables  $x_i$  and  $y_i$  at node  $i$ . Here, we take the  $f_i$  and  $g_i$  to depend on  $\mathbf{y}$  and not on  $\mathbf{x}$ , such that inferring the topology becomes possible without access to  $\mathbf{x}$ -measurements. We remark that the third class of functions, the  $h_i$ , still may depend on  $\mathbf{x}$ . We expand the function

$$h_i(\mathbf{x}) = \sum_k \sum_{j=1}^N (A_k)_{ij} \sin(k(x_j - x_i)) + \sum_{k'} \sum_{j=1}^N (B_{k'})_{ij} \cos(k'(x_j - x_i)) \tag{3}$$

in terms of Fourier series with constant matrices  $A_k, B_{k'} \in \mathbb{R}^{N \times N}$  of coupling strengths in Fourier modes  $k, k'$ . In the examples illustrated below, we restrict ourselves to the

coupling function

$$h_i(\mathbf{x}) = \sum_{j=1}^N A_{ij} \sin(x_j - x_i) \tag{4}$$

for all  $i \in \{1, \dots, N\}$ , i.e., consider only the contribution of one Fourier term,  $A_1 = A$ , all other  $A_k = 0$  and all  $B_{k'} = 0$ . The extension to general  $h$  is conceptually identical but much more cumbersome. As for the *direct method* mentioned above, we take all functions  $g_i(\mathbf{y}, t)$  and  $f_i(\mathbf{y}, t)$  as well as the time series of  $y_i(t)$  to be known for all  $i \in \{1, \dots, N\}$ .

**B. INFERENCE OF INTERACTION TOPOLOGY**

The task is to find for which pairs  $(i, j)$ , the coupling matrix is non-zero, i.e.  $A_{ij} \neq 0$ , and to infer state differences  $x_j - x_i$  for the non-observed part of the state vector  $\mathbf{x}$ . If we were to apply the original direct method in a network with  $N$  units with two variables each, both the time series of  $\mathbf{x}$  and  $\mathbf{y}$  had to be recorded at each unit and at  $M > 2N$  measurement points  $t_m, n \in \{1, \dots, M\}$ . In contrast, the method we introduce in the following is applicable without knowledge of the time series of the  $x_i(t)$ . Given we have no access to  $\mathbf{x}(t)$ , we aim at eliminating the  $\mathbf{x}(t)$  from the dynamical system. Substituting the  $x_i(t)$  by integrating the first differential equation (1) that defines the dynamics yields

$$x_i(t) = \int_{t_0}^t f_i(\mathbf{y}(t'), t') dt' + x_i(t_0) \tag{5}$$

where we now introduce the  $N$  initial condition components  $x_i^0 = x_i(t_0) \in \mathbb{R}$ , which are unknown real parameters. Substituting these integral expressions into equations (2) (for our choice (4) of the  $h_i$ ) gives

$$\dot{y}_i = g_i(\mathbf{y}, t) + \sum_{j=1}^N A_{ij} \sin\left(\int_{t_0}^t \Delta f_{ji}(\mathbf{y}, t') dt' + \Delta x_{ji}^0\right), \tag{6}$$

with  $\Delta f_{ji}(\mathbf{y}, t) := f_j(\mathbf{y}, t) - f_i(\mathbf{y}, t)$  and  $\Delta x_{ji}^0 := x_j^0 - x_i^0$ .

Separating known and unknown variables results in a linear system of equations, which then is overdetermined by recording sufficiently many data of the partial state vectors  $\mathbf{y}(t)$ . Exploiting  $\sin(\alpha + \beta) = \sin \alpha \cos \beta + \cos \alpha \sin \beta$ ,

the system's coupling terms become

$$\sin(x_j - x_i) = \sin\left(\int_{t_0}^t \Delta f_{ji}(\mathbf{y}, t') dt'\right) \cos(\Delta x_{ji}^0) + \cos\left(\int_{t_0}^t \Delta f_{ji}(\mathbf{y}, t') dt'\right) \sin(\Delta x_{ji}^0). \quad (7)$$

We approximate both  $\dot{y}_i(t)$  and  $\int_{t_0}^t \Delta f_{ji}(\mathbf{y}, t') dt'$  by numerically differentiating and integrating the recorded time series of  $y_i(t)$ , respectively, compare [2]. Such approximation is faithful if the temporal sampling resolution of the measured trajectory  $\mathbf{y}(t)$  allows for robust numerical integration and differentiation.

For each unit  $i \in \{1, \dots, N\}$  and each time instant  $t_m$ ,  $m \in \{1, \dots, M\}$ , we define

$$\ell^i(t_m) := \dot{y}_i(t_m) - g_i(\mathbf{y}(t_m), t) \in \mathbb{R} \quad (8)$$

and  $\mathbf{r}^i(t_m) \in \mathbb{R}^{2N \times 1}$  via its components

$$\mathbf{r}_{j'}^i(t_m) = \begin{cases} \sin\left(\int_{t_0}^{t_m} \Delta f_{j'i}(\mathbf{y}, t') dt'\right), & j' \in \mathbb{N}_N \\ \cos\left(\int_{t_0}^{t_m} \Delta f_{(j'-N)i}(\mathbf{y}, t') dt'\right), & j' \in \mathbb{N}_{2N} \setminus \mathbb{N}_N \end{cases} \quad (9)$$

where  $\mathbb{N}_n = \{1, \dots, n\}$  for  $n \in \mathbb{N}$ . Moreover, we abbreviate

$$\mathbf{L}^i = \left(\ell^i(t_1), \dots, \ell^i(t_M)\right) \in \mathbb{R}^{1 \times M} \quad (10)$$

and

$$\mathbf{R}^i = \left(\mathbf{r}^i(t_1), \dots, \mathbf{r}^i(t_M)\right) \in \mathbb{R}^{2N \times M}. \quad (11)$$

The unknown, temporally constant terms that result from a combination of the unknown initial condition components  $x_i^0$  and the missing knowledge of the  $A_{ij}$  (coupling strength from unit  $j$  to unit  $i$ ) to be inferred are substituted by  $\mathbf{C}^i \in \mathbb{R}^{1 \times 2N}$  with

$$\mathbf{C}_{j'}^i = \begin{cases} A_{ij'} \cos(\Delta x_{ji}^0), & j' \in \mathbb{N}_N \\ A_{i(j'-N)} \sin(\Delta x_{(j'-N)i}^0), & j' \in \mathbb{N}_{2N} \setminus \mathbb{N}_N. \end{cases} \quad (12)$$

We hence rewrite both left-hand side and right-hand side of the dynamical system defined by (1) and (2), sampled at times  $t_m$  as a system of linear equations,

$$\mathbf{L}^i = \mathbf{C}^i \mathbf{R}^i. \quad (13)$$

For each unit  $i \in \{1, \dots, N\}$ , the quantities  $\mathbf{L}^i$  and  $\mathbf{R}^i$  are known, whereas  $\mathbf{C}^i$  is unknown and searched for. The system (13) is overdetermined if the number  $M$  of time points at which state data of  $\mathbf{y}$  is recorded exceeds  $2N$ , twice the number of units.

Due to inexact measurement data and numerical inaccuracies, we are searching for a robust solution of the overdetermined system of equations (13). Hence, we minimize the  $L_2$  norm  $\|\mathbf{L}^i - \mathbf{C}^i \mathbf{R}^i\|_2$  and obtain the solution  $\widehat{\mathbf{C}}^i$  by multiplying (13) with the Moore-Penrose-Pseudoinverse  $\mathbf{R}^{i+}$  [2],

$$\widehat{\mathbf{C}}^i = \mathbf{L}^i \mathbf{R}^{i+}. \quad (14)$$

By definition (12) the matrix  $\widehat{\mathbf{C}}^i$  already hints which of the  $A_{ij}$  might be zero, indicating which of the potential interactions

from a unit  $j$  to another  $i$  are not present. Beyond this qualitative binary information, we may quantitatively estimate the elements of the coupling matrix via

$$\begin{aligned} |A_{ij}| &= \sqrt{\left(A_{ij} \cos(\Delta x_{ji}^0)\right)^2 + \left(A_{ij} \sin(\Delta x_{ji}^0)\right)^2} \\ &= \sqrt{\left(\widehat{\mathbf{C}}_j^i\right)^2 + \left(\widehat{\mathbf{C}}_{j+N}^i\right)^2} \end{aligned} \quad (15)$$

for  $j \in \{1, \dots, N\}$ .

### C. FEATURES OF HIDDEN STATE VARIABLES

Given the dynamics, the values of the hidden variables  $\mathbf{x}$  cannot be obtained entirely. However, we can determine their differences for all times  $t$ , i.e., find estimates for all  $x_i$  up to a uniform additive shift for all units  $i$ . We identify the initial difference between  $x_j$  and  $x_i$  using Euler's formula and obtain their difference

$$\Delta x_{ji}^0 + 2\pi m = -\iota \ln \left[ \frac{1}{A_{ij}} \left( \widehat{\mathbf{C}}_j^i + \iota \widehat{\mathbf{C}}_{j+N}^i \right) \right], \quad m \in \mathbb{Z} \quad (16)$$

where  $\iota$  is the imaginary unit. The complex valued logarithm specifies the phase except for an additive term of an integer multiple of  $2\pi$ . Here, we take the terms  $m = 0$ , and restrict the phase differences to the interval  $[-\pi, \pi)$ . We calculate the phase differences at any arbitrary time  $t \in [t_0, t_\infty]$  via

$$\Delta x_{ji}(t) = \int_{t_0}^t [f_j(\mathbf{y}, t') - f_i(\mathbf{y}, t')] dt' + \Delta x_{ji}^0 \quad (17)$$

where  $t_0$  is the time at start of observation and  $t_\infty$  chosen such that  $t_\infty - t_0$  is much larger than the relevant relaxation time scales in the system.

Finally, combining (5) for two units ( $i$  and  $j$ ) yields the fixed point differences,

$$\Delta x_{ji}^* = \int_{t_0}^{t_\infty} [f_j(\mathbf{y}(t'), t') - f_i(\mathbf{y}(t'), t')] dt' + \Delta x_{ji}^0. \quad (18)$$

From the phase differences  $\Delta x_{ji}(t)$  we thus obtain all the phases  $x_i(t)$  up to one overall (and arbitrary) phase shift.

### D. UNDIRECTED NETWORKS

For undirected networks, i.e. systems with symmetric coupling matrices  $A_{ij} = A_{ji}$ , we modify the equations to assure symmetry in the coupling matrix and antisymmetry in the coupling functions by requiring that

$$\mathbf{C}_j^i = \mathbf{C}_i^j \quad \text{and} \quad \mathbf{C}_{j+N}^i = -\mathbf{C}_{i+N}^j. \quad (19)$$

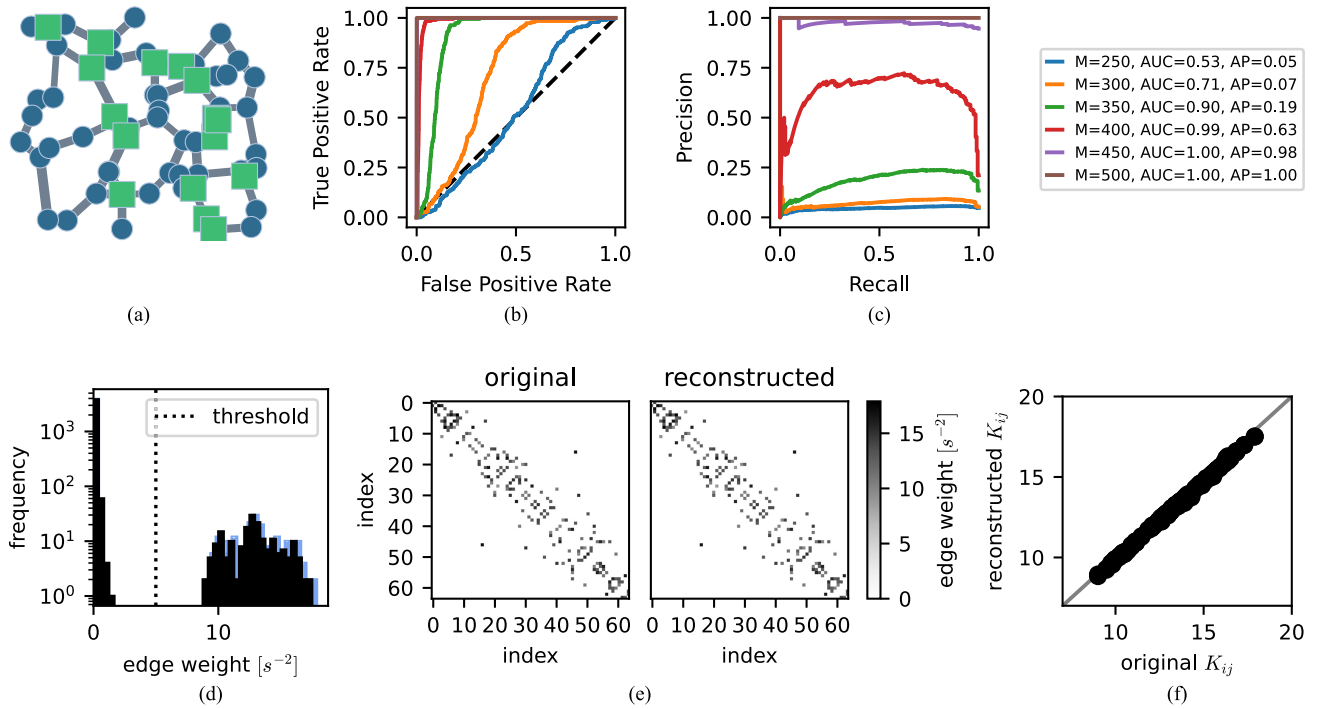
Taking

$$|A_{ij}| = |A_{ji}| = \frac{1}{2} \sqrt{(\widehat{\mathbf{C}}_j^i + \widehat{\mathbf{C}}_i^j)^2 + (\widehat{\mathbf{C}}_{j+N}^i - \widehat{\mathbf{C}}_{i+N}^j)^2}, \quad (20)$$

symmetrizes the reconstructed matrix, where the quantities on the right hand side are computed via (12).

For the differences in initial conditions we obtain

$$\Delta x_{ji}^0 = -\iota \ln \left[ \frac{1}{2A_{ij}} \left( \widehat{\mathbf{C}}_j^i + \widehat{\mathbf{C}}_i^j + \iota (\widehat{\mathbf{C}}_{j+N}^i - \widehat{\mathbf{C}}_{i+N}^j) \right) \right]. \quad (21)$$



**FIGURE 2. Reconstruction of an example model network.** (a) Exemplary synthetic power grid with  $N = 64$  units, generated using the algorithm published by Schultz et al. [35]. Green squares represent generators, and blue disks represent consumers. (b) Receiver operating characteristic (ROC) curves for the depicted network for various sample lengths. The AUC score is defined as the area below the ROC curve. (c) Precision-recall (PR) curves for various sample lengths. AP is defined as the area below the PR curve. Sample length, AUC score and AP for each curve are listed in the upper right box. (d) Histogram of the frequency of reconstructed edge weights (black). For comparison, the real edge distribution is also depicted in blue. To filter out false positives, we apply Otsu's threshold selection method [36]. Every edge with a weight below the calculated threshold, depicted as the dotted vertical line, is discarded. (e) Visualization of the original and reconstructed coupling matrix. (f) Plotting the reconstructed against the original edge weights shows good agreement, although the inferred edge weights are systematically slightly too small.

### III. RESULTS

#### A. APPLICATION ON A POWER GRID MODEL

We illustrate and discuss the proposed reconstruction method by applying it to a simple autonomous dynamical network model for power grid dynamics, namely the coupled second order swing equations. As opposed to static load-flow models, the coupled swing equations considered as an example here model power grid dynamics that include electromechanical dynamics [37]. Here, the voltage angle  $x_i(t) = \theta_i(t)$  and the angular frequency deviation from a reference frequency (e.g., 50 Hz)  $y_i(t) = \omega_i(t)$  are the components of the state variables,  $f_i(\omega, t) = \omega_i$  and  $g_i(\omega, t) = P_i - \gamma_i \omega_i$  both do not explicitly depend on time and  $h_i(\theta) = \sum_{j=1}^N K_{ij} \sin(\theta_j - \theta_i)$  is a coupling function exhibiting one Fourier mode. We thereby consider a lossless transmission grid for clarity of presentation. Including losses would introduce a cosine Fourier term in the coupling function. As discussed in section II A, including additional terms in the coupling function would make the network inference problem technically more cumbersome without contributing any qualitative novelty. The resulting dynamical system is given by [34], [38], [39]

$$\dot{\theta}_i = \omega_i, \tag{22}$$

$$I_i \dot{\omega}_i = P_i - \gamma_i \omega_i + \sum_{j=1}^N K_{ij} \sin(\theta_j - \theta_i), \tag{23}$$

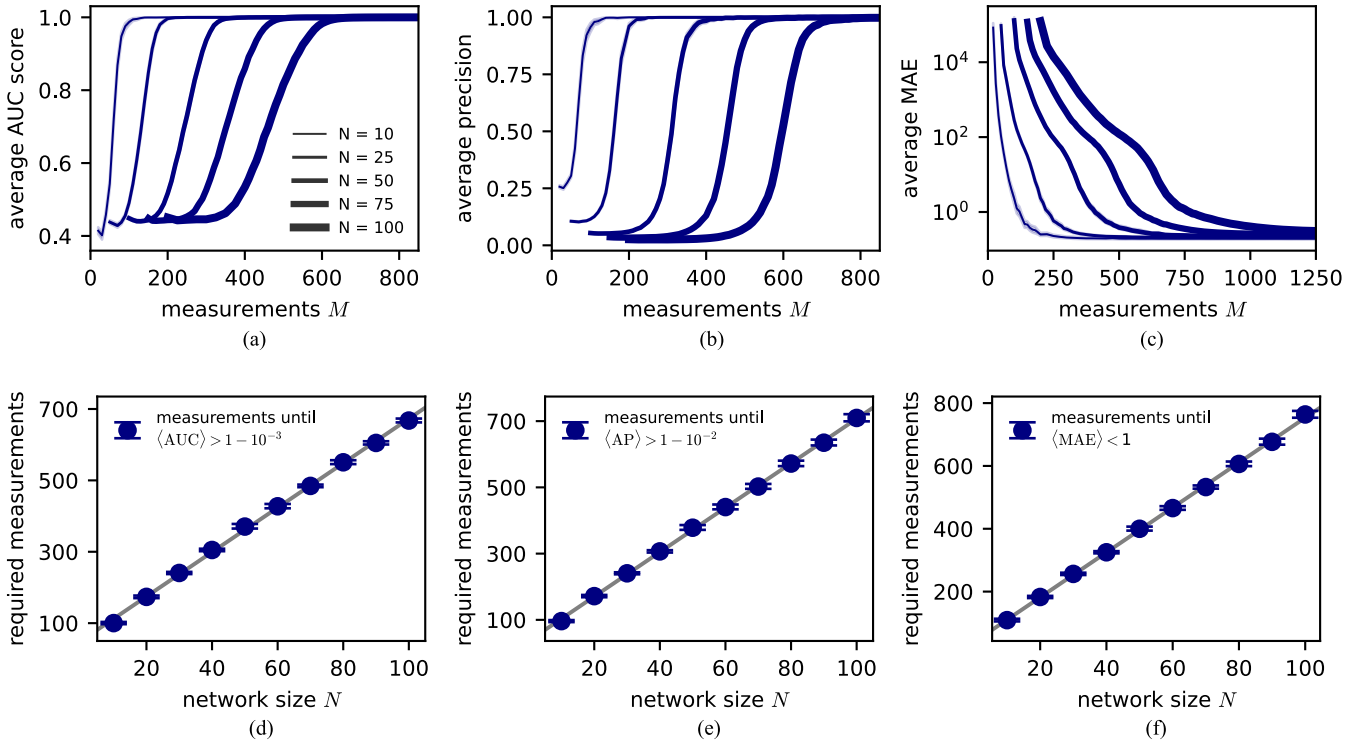
where  $I_i$  denotes the moment of inertia,  $P_i$  reflects the power consumed ( $P_i < 0$ ) or provided ( $P_i > 0$ ) at unit  $i$  and  $K_{ij}$  is the matrix of effective coupling strengths among units of the power grid.

The term  $-\gamma_i \omega_i$  in (23) includes both damping and primary (proportional) control. In this manuscript, we assume homogeneous damping,  $\gamma_i \equiv \gamma$  for all  $i \in \{1, \dots, N\}$ , and do not include secondary (integrative) control, again because this would make the presentation more cumbersome without adding novelty. Drawing moments of inertia  $I_i$  from a normal distribution with mean 1 and standard deviation 0.1 (almost certainly yielding  $I_i > 0$  for reasonable  $N$ ) creates a family of sample networks.

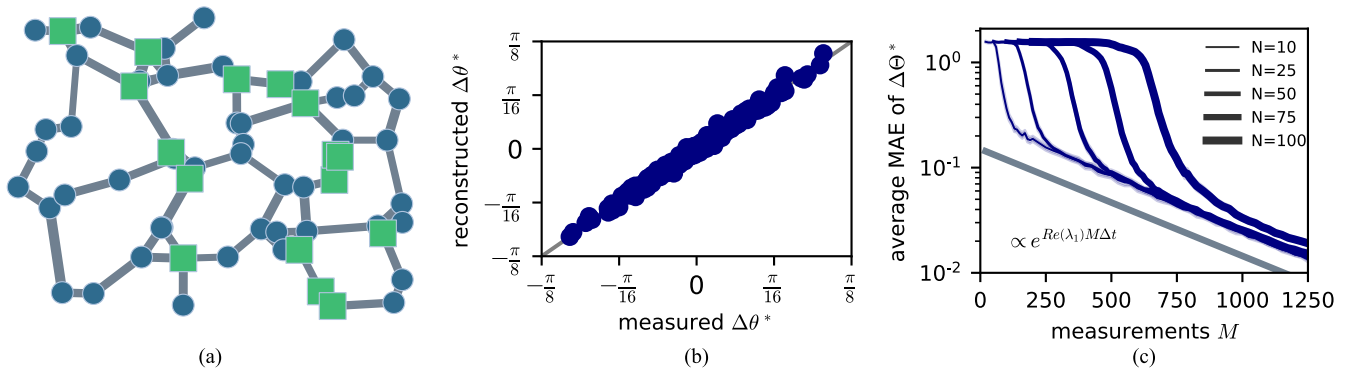
Dividing (23) by  $I_i > 0$  gives the functional form demanded in (2). Note that we thus infer the scaled matrix elements  $K'_{ij} = K_{ij}/I_i$ , and hence have to correct for the normalization afterwards. According to the conditions under which we aim to illustrate inference, in this system, we take  $\gamma$ ,  $P_i$  and  $I_i$  to be known for each unit  $i \in \{1, \dots, N\}$ . Measuring the transient frequencies at each node  $i$  as the system relaxes after a small perturbation yields the frequency time series  $\omega_i(t)$ . We take all the  $\theta_i(t)$  time series to be unknown.

Following the steps described in Section II, for each measured time instant  $t_m$ ,  $m \in \{1, \dots, M\}$ , we define

$$\ell^i(t_m) = \dot{\omega}(t_m) - \frac{1}{I_i} (P_i - \gamma \omega_i(t_m)), \tag{24}$$



**FIGURE 3. Scaling of topology reconstruction.** For various network sizes, measured by their number of units  $N$ , 100 random networks have been created using the model by Schultz *et al.* [35]. We apply the presented inference method as a function of the measurements  $M$ . (a) and (b) display how the AUC score and AP reach their ideal value of 1 with increasing number of measurements, respectively. (c) shows how the mean absolute error (MAE) decreases with increasing number of measurements. In (a) to (c) lines represent the mean value and the shaded areas emphasize the 95% confidence interval of the distribution. (d) to (f) report the scaling to reach a suitable AUC score, AP and MAE as a function of the network size, respectively. Error bars again indicate the 95% confidence interval, obtained via bootstrapping.



**FIGURE 4. Recovery of the hidden variable.** (a) For one random network drawn as above via the model by Schultz *et al.* (same power grid realization as in Fig. 2), we reconstructed the phase differences  $\Delta\theta_{ij}^*$  according to (29) from the recorded frequency time series, yielding reasonably agreement for all edges  $(i, j)$ , see (b). The scaling of the MAE and its 95% confidence interval with the number of measurements of an ensemble of 100 networks for each network size  $N$  is illustrated in (c).

$$r_{j'}^i(t_m) = \begin{cases} \sin \Delta\theta_{j'i}(t_m), & j' \in \mathbb{N}_N \\ \cos \Delta\theta_{(j'-N)i}(t_m), & j' \in \mathbb{N}_{2N} \setminus \mathbb{N}_N, \end{cases} \quad (25)$$

$$C_{j'}^i = \begin{cases} K'_{ij'} \cos \Delta\theta_{j'i}^0, & j' \in \mathbb{N}_N \\ K'_{i(j'-N)} \sin \Delta\theta_{(j'-N)i}^0, & j' \in \mathbb{N}_{2N} \setminus \mathbb{N}_N. \end{cases} \quad (26)$$

yields the elements of the connectivity matrix,

$$|K_{ij}| = I_i |K'_{ij}| = I_i \sqrt{(\widehat{C}_j^i)^2 + (\widehat{C}_{j+N}^i)^2}, \quad (27)$$

and the initial phase difference at the instant  $t = t_0$  of the perturbation,

$$\Delta\theta_{ji}^0 = -t \ln \left( \frac{I_i}{K_{ij}} \left( \widehat{C}_j^i + t \widehat{C}_{j+N}^i \right) \right), \quad (28)$$

where  $-\Delta\theta_{ij}^0 = \Delta\theta_{ji}^0$ . Next, we infer the phase differences in the fixed point via

$$\Delta\theta_{ji}^* = \int_{t_0}^{t_\infty} [\omega_j(t') - \omega_i(t')] dt' + \Delta\theta_{ji}^0. \quad (29)$$

As follows from linear response theory, the amplitude of the perturbation decays in first order approximation proportional to  $\exp(\text{Re}[\lambda_1]t)$ , where  $\lambda_1$  denotes the eigenvalue of the system's Jacobian with largest non-zero real part. In the simulations, we employ this feature and take  $t_\infty$  as the time, after which the perturbation amplitude has decayed to 1% of its initial amplitude.

### B. TECHNICAL DETAILS

To test our approach on a large number of networks with an adjustable number of nodes, which still display core characteristics of real power grid topology, we utilize the power grid topology generation algorithm introduced by Schultz et al. [35]. The unit locations are drawn randomly from the uniform distribution on  $[0, 1) \times [0, 1)$  and the distance between them is defined via the  $L_2$  norm. The edge weights  $K_{ij}$  are drawn from a normal distribution with mean 13 and standard deviation 2.0. The damping coefficient is chosen as  $\gamma = 0.1$ . In each network with  $N$  nodes,  $\lfloor \frac{N}{4} \rfloor$  units act as power generators, whereas the remaining  $N^C = N - \lfloor \frac{N}{4} \rfloor$  units act as consumers with  $P^C = -1$ . Since the system operates at a stable fixed point, the power of the generators satisfies  $\sum_{i=0}^N P_i = 0$ , hence  $P^G = -N^C P^C / (N - N^C)$ .

We consider the transient dynamics of the system after a small perturbation. In real systems, perturbations occur even without experimental interference; here we externally perturb the system randomly. Specifically, we perturb each unit with a phase shift drawn from a normal distribution with mean zero and standard deviation  $\pi/8$  and simulate the nodal frequencies  $\omega_i(t)$  at each unit  $i \in \{1, \dots, N\}$  as the system relaxes to its stable fixed point by integrating the dynamical system (22), (23). We assume that the frequencies are sampled in time steps of 0.05. We use these frequency time series to reconstruct both the network topology and phase differences following the procedure described above.

An exemplary reconstruction of a modeled power grid based on a frequency time series with 1500 data points is illustrated in Fig. 2.

### C. QUANTIFICATION AND ERROR MEASURES

Understanding and quantifying systematic errors occurring in the proposed reconstruction scheme, requires a systematic analysis of the initially obtained raw results. How many data points have to be measured to obtain a reliable reconstruction? How does this requirement scale with the number of nodes in the network? Answering these questions requires different methods and quantifiers, as detailed below.

The area under the receiver operating characteristic curves (AUC score) and the average precision (AP) quantify the overall reconstructability of the network given a recorded frequency time series, whereas the mean absolute error (MAE)

indicates the precision of the inferred topology and edge weights. The receiver operating characteristic (ROC) curve follows from comparing the reconstructed (RM) and original (OM) adjacency matrix. Edges present both in RM and OM count as true positives. If an edge is in RM but not in OM, it counts as a false positive. The ROC curve results from calculating the true positive rates and false positive rates of RM with respect to OM. Each unique value in RM successively is taken as a threshold, and all edges with weights below this threshold value are set to zero, values above are set to 1. Comparison with the original adjacency matrix yields the true and false positive rates. The ROC curve then follows from plotting the true positive rate against the false positive rate. The area below the ROC curve is denoted the AUC score and serves as a good indicator of the ability to reconstruct the network topology. An AUC score of 1 is equivalent to perfect reconstructability, with the true positive rate equaling 1 and the false positive rate equals zero. An AUC score of 0.5 indicates, that a detected edge may be true positive or false positive with equal probability [40], equivalent to randomly guessing the outcome. Similarly, the area under the precision-recall (PR) curve, denoted AUC-PR or average precision (AP), follows from plotting precision, defined as the ratio of true positives over the sum of true positives and false positives, as a function of recall, defined as the ratio of true positives over the sum of true positives and false negatives, for varying thresholds. In contrast to the area under the receiver operating characteristic curves, AP is less sensitive to unbalanced data, where positives occur much less frequently than negatives [41]. Fig. 2 (b) and (c) display exemplary ROC and PR curves of the displayed network at various sample lengths  $M$ , respectively. For the subsequent images in Fig. 2,  $M = 1500$  measurements have been used, leading to an AUC score and an AP of 1.

Still, calculating AP and the AUC score requires the knowledge of the original coupling matrix, which we take to be unknown. To identify correct edges and discard numerical artifacts, a different approach is needed. Plotting the distribution of the raw reconstructed edge weights (black) and the distribution of the real edge weights (blue), see Fig. 2 (d), reveals incorrectly detected edges with edge weights close to zero. Applying Otsu's method [36] yields a threshold to classify edges and artifacts. All edges with weights below this threshold are assumed to be false positives and hence discarded.

Whereas AP and the AUC score indicates the reconstructability and Otsu's method serves to distinguish between correct and incorrect edges, the mean absolute error (MAE) quantifies the accuracy of the inferred edge weights. The MAE between two matrices  $A, B \in \mathbb{R}^{N \times N}$  is defined by

$$\text{MAE}(A, B) = \frac{1}{N'} \sum_{i=0}^N \sum_{j=0}^N |A_{ij} - B_{ij}|. \quad (30)$$

To take into account network sparsity, we do not set the normalization constant  $N'$  to  $N^2$  as often is the default but

choose  $N'$  as the number of non-zero matrix elements in the original matrix. Removing the numeric artifacts (discarding edges below a threshold, see Fig. 2 (d)), before calculating the MAE neglects their contribution to the error. Adjusting  $N'$  accordingly prevents scaling of the MAE with the matrix dimensions.

#### D. NETWORK INFERENCE

Fig. 2 illustrates step by step the numerical results obtained by using the introduced method to infer a medium sized synthetic power grid based on its frequency time series induced by a perturbation. When recording a sufficient amount of data (sample length  $M = 1500$  in this example), the reconstructed matrix coincides very well with the original matrix, see Fig. 2 (e). More quantitatively, plotting the reconstructed against the original edge weights (Fig. 2 (f)) reveals a good agreement between the original and reconstructed coupling matrix. Although the method tends to slightly underestimate the edge weights with increasing sample length, most values lie very close to the diagonal that indicates perfect inference. For a systematic analysis of the reconstruction quality and the algorithm's scaling with the number of units in the network and the measurement count, we applied the method to an ensemble of 100 networks for each network size and evaluated the reconstructed matrix at different measurement counts. Fig. 3 (a) to (c) show the AUC score, AP and the average MAE as a function of the number of measurements for different network sizes. With the given parameter settings, both AUC and AP approach 1 for all tested networks. The average MAE also appears to converge to a fixed value, although larger networks require significantly more measurements before reaching this limit. Before slowing down at this tail, the required measurements for reaching a fixed AUC score, AP and MAE increase linearly with the network size, as Fig. 3 (d) to (f) illustrate. The cutoff MAE value of 1 corresponds to an average deviation between predicted and original edge weights of approximately 7.7% with the given simulation parameters. Sufficiently long measurements yield a relative edge deviation of approximately 2% with the chosen parameters. Rarely, we detected numerical outliers in the MAE due to a single very large inferred matrix element. Data of the corresponding network have been omitted in Fig. 3.

The maximal reconstructable network size based on data of a single perturbation is limited, since the amplitude of the initial perturbation decays exponentially with the largest real part of the linearized system's eigenvalues not equal to zero, which is dependent on the system's parameters. The reconstructability of a dynamical system's coupling matrix hence improves with larger initial perturbations and lower damping.

Comparing the introduced method with the direct method to infer network topology from time series measurements [15] reveals that our proposed method requires more samples to achieve a comparable inference quality in terms of F1-Score (the harmonic mean of the precision and recall value) and MAE, see Fig. 5. However, the direct method is incapable

of inferring network topology from incomplete information such as time series missing for any variables. We therefore used both the frequency and phase angle time series for inference and hence approximately twice the information available to our proposed method. With sufficient data available, both methods achieve a similarly high F1-Score and MAE.

#### E. PHASE RECOVERY

As discussed in Section II, the hidden phase differences  $\Delta\theta_{ij}$  along all lines in the network may be recovered from measurements of the local angular velocities  $\omega_i$ , see Fig. 4.

Given sufficiently many data points, the reconstructed phase differences in the fixed point  $\Delta\theta_{ij}^*$  lie close to the diagonal when plotted against the measured phase differences, i.e. the measured and reconstructed  $\Delta\theta_{ij}^*$  match well, see Fig. 4 (b) for the inferred phase difference from the network topology shown in Fig. 4 (a) (same power grid realization as in Fig. 2).

The scaling of the mean MAE for ensembles of 100 networks each for different measurement counts and network sizes is illustrated in Fig. 4 (c). For all networks, the MAE starts at a constant value, since phase angles differences are restricted to the interval  $[0, \pi)$ , leading to an average initial MAE of  $\pi/2$ . After the number of measurements  $M$  increases, the MAE decreases very fast and converges to an exponential decay afterwards. The initially large MAE is dominated by the reconstruction error of  $\Delta\theta_{ij}^0$ . As this error decreases, the accuracy still is restricted by the exponential damping of the dynamical system. Only as the system's state has reached its stable fixed point again, the inference error is minimal.

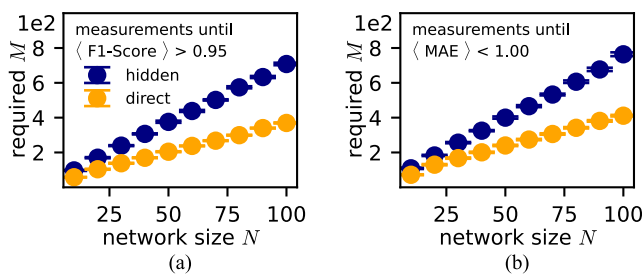
The amplitudes of the perturbed frequencies decay approximately exponentially and so does the error of the reconstructed phases. Adding the reconstructed initial phase difference to the integral over the frequency difference yields the phase differences in the fixed point, see (29). Note that we also could recover the fixed point phases by using the reconstructed edges to calculate the roots of the dynamical system given by (22) and (23). However, by using the introduced method we additionally obtain the unknown initial perturbation  $\Delta\theta_{ij}^0$  and also compute phase difference trajectories  $\Delta\theta_{ij}(t)$  for all edges, purely based on frequency measurements.

#### IV. DISCUSSION AND OUTLOOK

We have presented an inference method to obtain the network connectivity matrix without monitoring all nodal quantities, i.e., inference with hidden (not directly monitored) variables. Using frequency recordings of a simulated synthetic power grid, both the coupling matrix and the entire phase difference time series may be recovered. Thereby, we have demonstrated by an example model application that the theoretically derived method of reconstruction from partial information works well in a simulated dynamical system. The presented inference method may prove to be a powerful tool for sys-



tems with established dynamical models of appropriate functional forms (2). The inference quality is high already for a moderate number of measurements. Keeping the inference quality constant, the number of required measurements scales approximately linearly with the network size. Since we are not aware of other methods that infer topology from partial measurements, we compare the proposed method with the well-established *direct method* [15], which requires access to all time series (no hidden variables). As illustrated in Fig. 5, compared to the direct method, the method proposed here requires a larger number of measurements from the observable units to achieve similarly large F1-Score (the harmonic mean of the precision and the recall value) and MAE, respectively. However, since the direct method uses effectively twice as much information (hidden variable and observable variable), the total information required in both approaches is approximately equal.



**FIGURE 5. Comparison with direct method** Comparing the required measurements of our proposed method (blue) and the *direct method* (orange) [15] to achieve a given F1-Score (a) and a given MAE (b) reveals that the direct method (assuming full knowledge of all time series) infers more accurately at the same number of measurements. However, both methods achieve similarly good scores eventually.

Our advancement is twofold: First, we have extended general network inference methods by providing an approach that does not require all nodal quantities to be monitored directly. This further moves the boundary of what quantities and information of a dynamical system can be recovered. We add to model-free inference [19] and hidden node detection [21] a method of inference when facing incomplete variable observation. This could be useful to many other applications where monitoring all quantities is either impossible or very costly, as in many biological systems, such as metabolic or neuronal systems [17]–[19], [42], [43].

Second, we have applied network inference methods to basic dynamical systems models of power grids with its dynamics given by the second order model (coupled swing equations). We thereby provide first steps towards a tool that may be useful for engineers inferring the network topology of real-world systems. This tool could for example be applied in distribution grids, e.g., to locate a faulty line or determine the status of switches if documentation is unavailable.

The proposed method requires knowledge about the full model of the underlying dynamical system, making it applicable to systems whose dynamics can at least be approximately modeled by the required functional forms. One major open

question is how to infer topology from time series sampled from a subset of variables in the absence of any nonlinear dynamical systems' model. Furthermore, although there is no hard limit, in practice both finite memory and the exponential decay of the perturbation amplitude limit the maximal reconstructable network size (and thus number of variables). As discussed in section II, the dimensionality of the linear system which has to be optimized scales linearly with the number of measurements. Simultaneously, the perturbation amplitudes decay exponentially with the largest non-zero eigenvalue real part. Above some network size (or number of variables), the information contained in a single perturbation hence may not suffice to reconstruct the topology accurately. Assessing the theoretical approach with experimentally obtained data also remains a task for future research.

There remain many open questions. From the application point of view, it would be interesting to set up a laboratory experiment or see this inference methodology applied to a real distribution power grid. The method itself should be advanced to cover a broader spectrum of local dynamics  $f$  and  $g$ , and by overcoming the limit of effectively reconstructable network sizes, e.g., by using measurements of multiple perturbations. Finally, a long-term project could combine model-free approaches [19], the inference of hidden nodes [21] and our approach of determining inaccessible variables into one coherent and powerful inference framework. However, the introduced method may already contribute to the analysis of coupled dynamical systems whose topology previously could not be retrieved due to the inaccessibility of some dynamical variables.

#### AVAILABILITY OF CODE

Python code to infer the network topology and differences in the hidden variable and a short discussion of numerical properties and challenges is made available at <https://github.com/Network-Dynamics/network-inference-with-hidden-variables>.

#### ACKNOWLEDGMENT

(Raoul Schmidt, Hauke Haehne, Benjamin Schäfer, and Marc Timme contributed equally to this work.)

#### REFERENCES

- [1] R. De Smet and K. Marchal, "Advantages and limitations of current network inference methods," *Nature Rev. Microbiol.*, vol. 8, no. 10, pp. 717–719, 2010.
- [2] M. Timme and J. Casadiego, "Revealing networks from dynamics: An introduction," *J. Phys. A, Math. Theor.*, vol. 47, no. 34, 2014, Art. no. 343001.
- [3] W.-X. Wang, Y.-C. Lai, and C. Grebogi, "Data based identification and prediction of nonlinear and complex dynamical systems," *Phys. Rep.*, vol. 644, pp. 1–76, Jul. 2016.
- [4] R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems*. Amsterdam, The Netherlands: Elsevier, 2018.
- [5] P. D'haeseleer, S. Liang, and R. Somogyi, "Genetic network inference: From co-expression clustering to reverse engineering," *Bioinformatics*, vol. 16, no. 8, pp. 707–726, 2000.
- [6] T. P. Peixoto, "Network reconstruction and community detection from dynamics," *Phys. Rev. Lett.*, vol. 123, no. 12, Sep. 2019, Art. no. 128301.

- [7] M. Timme, "Revealing network connectivity from response dynamics," *Phys. Rev. Lett.*, vol. 98, no. 22, May 2007, Art. no. 224101.
- [8] Z. Levnajić and A. Pikovsky, "Network reconstruction from random phase resetting," *Phys. Rev. Lett.*, vol. 107, no. 3, Jul. 2011, Art. no. 034101.
- [9] M. Tyloo, R. Delabays, and P. Jacquod, "Reconstructing network structures from partial measurements," 2020, *arXiv:2007.16136*.
- [10] R. Aghdam, M. Ganjali, X. Zhang, and C. Eslahchi, "CN: A consensus algorithm for inferring gene regulatory networks using the SORDER algorithm and conditional mutual information test," *Mol. BioSyst.*, vol. 11, no. 3, pp. 942–949, 2015.
- [11] S. H. Mahmoodi, R. Aghdam, and C. Eslahchi, "An order independent algorithm for inferring gene regulatory network using quantile value for conditional independence tests," *Sci. Rep.*, vol. 11, no. 1, p. 7605, Dec. 2021.
- [12] W.-X. Wang, Y.-C. Lai, C. Grebogi, and J. Ye, "Network reconstruction based on evolutionary-game data via compressive sensing," *Phys. Rev. X*, vol. 1, no. 2, Oct./Dec. 2011, Art. no. 021021.
- [13] F. Basiri, J. Casadiego, M. Timme, and D. Witthaut, "Inferring power-grid topology in the face of uncertainties," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 98, no. 1, Jul. 2018, Art. no. 012305.
- [14] Z. Zhang, Y. Zhao, J. Liu, S. Wang, R. Tao, R. Xin, and J. Zhang, "A general deep learning framework for network reconstruction and dynamics learning," *Appl. Netw. Sci.*, vol. 4, no. 1, p. 110, Dec. 2019.
- [15] S. G. Shandilya and M. Timme, "Inferring network topology from complex dynamics," *New J. Phys.*, vol. 13, no. 1, Jan. 2011, Art. no. 013004.
- [16] E. S. C. Ching, P.-Y. Lai, and C. Y. Leung, "Reconstructing weighted networks from dynamics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 91, no. 3, Mar. 2015, Art. no. 030801.
- [17] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 15, pp. 3932–3937, Apr. 2016.
- [18] M. Nitzan, J. Casadiego, and M. Timme, "Revealing physical interaction networks from statistics of collective dynamics," *Sci. Adv.*, vol. 3, no. 2, Feb. 2017, Art. no. e1600396.
- [19] J. Casadiego, M. Nitzan, S. Hallerberg, and M. Timme, "Model-free inference of direct network interactions from nonlinear collective dynamics," *Nature Commun.*, vol. 8, no. 1, p. 2192, Dec. 2017.
- [20] I. T. Tokuda, Z. Levnajić, and K. Ishimura, "A practical method for estimating coupling functions in complex dynamical systems," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 377, no. 2160, Dec. 2019, Art. no. 20190015.
- [21] H. Haehne, J. Casadiego, J. Peinke, and M. Timme, "Detecting hidden units and network size from perceptible dynamics," *Phys. Rev. Lett.*, vol. 122, no. 15, Apr. 2019, Art. no. 158301.
- [22] D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, M. Kellis, J. J. Collins, and G. Stolovitzky, "Wisdom of crowds for robust gene network inference," *Nature Methods*, vol. 9, no. 8, pp. 796–804, 2012.
- [23] H. Lütcke, F. Gerhard, F. Zenke, W. Gerstner, and F. Helmchen, "Inference of neuronal network spike dynamics and topology from calcium imaging data," *Frontiers Neural Circuits*, vol. 7, p. 201, Dec. 2013.
- [24] X. Jin, W.-P.-K. Yiu, S.-H.-G. Chan, and Y. Wang, "Network topology inference based on end-to-end measurements," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 12, pp. 2182–2195, Dec. 2006.
- [25] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 3, pp. 467–483, Sep. 2017.
- [26] D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme, "Collective nonlinear dynamics and self-organization in decentralized power grids," *Rev. Mod. Phys.*, vol. 94, no. 1, Feb. 2022, Art. no. 015005.
- [27] A. N. Milioudis, G. T. Andreou, and D. P. Labridis, "Enhanced protection scheme for smart grids using power line communications techniques—Part II: Location of high impedance fault position," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1631–1640, Dec. 2012.
- [28] C. Hens, U. Harush, S. Haber, R. Cohen, and B. Barzel, "Spatiotemporal signal propagation in complex networks," *Nature Phys.*, vol. 15, no. 4, pp. 403–412, Jan. 2019.
- [29] L. Lampe and M. O. Ahmed, "Power grid topology inference using power line communications," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2013, pp. 336–341.
- [30] W. A. Omran, M. Kazerani, and M. M. A. Salama, "Investigation of methods for reduction of power fluctuations generated from large grid-connected photovoltaic systems," *IEEE Trans. Energy Convers.*, vol. 26, no. 1, pp. 318–327, Mar. 2011.
- [31] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Mag.*, vol. 7, no. 2, pp. 52–62, Mar./Apr. 2009.
- [32] M. Pospíšilík, P. Vařacha, and M. Adáček, "Low-cost rectifier for measuring of AC voltage or current frequency compensation proposal," *Int. J. Circuits, Syst. Signal Process.*, vol. 6, no. 3, pp. 394–401, 2012.
- [33] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 20–27, Jun. 2010.
- [34] J. Machowski, Z. Lubosny, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control*. Hoboken, NJ, USA: Wiley, 2020.
- [35] P. Schultz, J. Heitzig, and J. Kurths, "A random growth model for power grids and other spatially embedded infrastructure networks," *Eur. Phys. J. Special Topics*, vol. 223, no. 12, pp. 2593–2610, Oct. 2014.
- [36] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.
- [37] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power System Dynamics: Stability and Control*, 2nd ed. Chichester, U.K.: Wiley, 2008.
- [38] G. Filatrella, A. H. Nielsen, and N. F. Pedersen, "Analysis of a power grid using a Kuramoto-like model," *Eur. Phys. J. B*, vol. 61, no. 4, pp. 485–491, Feb. 2008.
- [39] M. Rohden, A. Sorge, M. Timme, and D. Witthaut, "Self-organized synchronization in decentralized power grids," *Phys. Rev. Lett.*, vol. 109, no. 6, Aug. 2012, Art. no. 064101.
- [40] C. D. Brown and H. T. Davis, "Receiver operating characteristics curves and related decision measures: A tutorial," *Chemometric Intell. Lab. Syst.*, vol. 80, no. 1, pp. 24–38, Jan. 2006.
- [41] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, "The area under the precision-recall curve as a performance metric for rare binary events," *Methods Ecol. Evol.*, vol. 10, no. 4, pp. 565–577, Apr. 2019.
- [42] M. Quach, N. Brunel, and F. d'Alché-Buc, "Estimating parameters and hidden variables in non-linear state-space models based on odes for biological networks inference," *Bioinformatics*, vol. 23, no. 23, pp. 3209–3216, 2007.
- [43] M. Helmstaedter, K. L. Briggman, and W. Denk, "High-accuracy neurite reconstruction for high-throughput neuroanatomy," *Nature Neurosci.*, vol. 14, no. 8, pp. 1081–1088, 2011.



**RAOUL SCHMIDT** received the B.Sc. and M.Sc. degrees in physics from Technical University Dresden, where he is currently pursuing the Ph.D. degree with the Center for Advancing Electronics Dresden. His research interests include network reconstruction and the analysis of nonlinear dynamical systems.



**HAUKE HAEHNE** received the B.Sc. and M.Sc. degrees in physics from the University of Göttingen, Germany, and the Ph.D. degree in physics from the University of Oldenburg, Germany, in 2019. His research interests include inverse problems in network dynamical systems and fluctuations in power grids.



**LAURA HILLMANN** received the degree in physics from the University of Göttingen. During her master's degree, she focused on network dynamics and wrote her thesis about network reconstruction of power grids. Currently, she is working as an Engineer at GEOFON, a seismological network operated by the German Research Center for Geosciences, Potsdam.



**JOSE CASADIEGO** received the degree in physics and majored in nonlinear dynamics and complex systems from the University of Carabobo, Venezuela, in 2009, and the Ph.D. degree in physics from the University of Goettingen, in 2016. In 2010, he moved to Goettingen, Germany, to pursue his Ph.D. degree. From 2016 to 2019, he worked as a Postdoctoral Researcher at the Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany, the Center of Advancing Electronics Dresden, and Technical University Dresden. Since 2019, he has been transitioned to the world of strategic consulting, taking a leading role in the areas of predictive modeling, optimization, machine learning, and statistics with the Advanced Analytics Division, Kearney. Among his duties, he is in-charge of developing novel mathematical approaches to solve industry problems and help shaping the strategy of advanced analytics across Europe.



**DIRK WITTHAUT** received the Diploma degree in physics and the Ph.D. degree from the Technical University of Kaiserslautern, Kaiserslautern, Germany, in 2004 and 2007, respectively. He has worked as a Postdoctoral Researcher with the Niels Bohr Institute, Copenhagen, Denmark, and the Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany. He has been a Guest Lecturer with the Kigali Institute for Science and Technology, Rwanda. Since 2014, he has been leading a research group at Forschungszentrum Jülich, Germany. He is currently a Professor with the University of Cologne. His research interests include data and network science and applications to energy systems.



**BENJAMIN SCHÄFER** received the Diploma degree in physics from the University of Magdeburg, Germany, in 2013, and the Ph.D. degree in physics from the University of Göttingen, in 2017. Pursuing his Ph.D. degree in Göttingen, Germany, London, U.K., and Tokyo, Japan. He has worked as a Postdoctoral Researcher at the Max Planck Institute for Dynamics and Self-Organization, Göttingen, Technical University Dresden, Germany, and as a Marie Skłodowska-Curie Research Fellow at the Queen Mary University of London. In 2021, he worked as an Associate Professor in physics at the Norwegian University of Life Sciences. Since 2022, he has been leading the Young Investigator Group on Data-Driven Analysis of Complex Systems for a Sustainable Future at KIT, Germany.



**MARC TIMME** received the degree in physics and mathematics in Würzburg, Stony Brook, USA, and Göttingen. After working as a Postdoctoral Researcher at the Max Planck Institute for Flow Research and as a Research Scholar at Cornell University, Ithaca, NY, USA, he was selected by the Max Planck Society to the Head of Broadly Cross-Disciplinary Max Planck Research Group on Network Dynamics, Max Planck Institute for Dynamics and Self-Organization. He held a Visiting Professorship at TU Darmstadt and was a Visiting Faculty at ETH Zürich. He is currently a Strategic Professor and the Head of the Chair for Network Dynamics at the Center for Advancing Electronics Dresden (cfaed), the Institute for Theoretical Physics, and the Cluster of Excellence Physics of Life, TU Dresden. He is also the Co-Chair of the Division of Socio-Economic Physics of the German Physical Society (DPG). Since 2018, he has been an Honorary Member of the Lakeside Laboratories, Klagenfurt. His research integrates first principles theory with data-driven modeling to establish generic fundamental insights that drive applications of complex dynamical systems, including bio-inspired information processing, energy systems, collective mobility and transport, and systemic sustainability.

...