

Hypergraph reconstruction from dynamics

Robin Delabays,^{1,*} Giulia De Pasquale,^{2,†} Florian Dörfler,^{2,‡} and Yuanzhao Zhang^{3,§}

¹*School of Engineering, University of Applied Sciences of Western Switzerland HES-SO, Sion, Switzerland*

²*Department of Information Technology and Electrical Engineering, ETH Zürich, Zürich, Switzerland*

³*Santa Fe Institute, Santa Fe, NM, USA*

A plethora of methods have been developed in the past two decades to infer the underlying network structure of an interconnected system from its collective dynamics. However, methods capable of inferring nonpairwise interactions are only starting to appear. Here, we develop an inference algorithm based on sparse identification of nonlinear dynamics (SINDy) to reconstruct hypergraphs and simplicial complexes from time-series data. Our model-free method does not require information about node dynamics or coupling functions, making it applicable to complex systems that do not have reliable mathematical descriptions. We first benchmark the new method on synthetic data generated from Kuramoto and Lorenz dynamics. We then use it to infer the effective connectivity among seven brain regions from resting-state EEG data, which reveals significant contributions from non-pairwise interactions in shaping the macroscopic brain dynamics.

INTRODUCTION

Hypergraphs and simplicial complexes have emerged as versatile and powerful tools for modeling and analyzing complex systems, offering flexible and expressive representations of higher-order relationships and dependencies that traditional graphs cannot capture [1–4]. These higher-order structures are important for a wide range of processes, from brain dynamics [5, 6] to communications in social systems [7, 8] and competitions in ecological systems [9].

The analysis of networked dynamical systems often comes as a two-sided coin. On the one hand, it is important to predict the possible dynamical behaviors based on model equations. For example, given the underlying network structure, what can we say about the dynamics? Many recent studies have taken this approach and investigated how higher-order interactions can influence collective dynamics [10–16], such as diffusion [17], consensus [18, 19], contagion [20, 21], synchronization [22–26], and controllability [27]. On the other hand, the inverse problem is equally important. Namely, is it possible to recover the network structure from dynamics? This network inference problem is important in many fields, including neuroscience [28–31] (e.g., infer effective connectivity between brain regions from fMRI data) and epidemiology [32, 33] (e.g., reconstruct social interactions from COVID infection data). For example, in the context of neuroscience, hypergraph inference techniques can help identify altered or dysfunctional network patterns, aiding in the understanding of the underlying mechanisms of these disorders and potentially suggesting novel therapeutic interventions [34].

When considering only pairwise interactions, network inference is a relatively mature field [35–44] and has been

approached from numerous perspectives utilizing diverse tools such as sparse regression [45], Bayesian inference [46], reservoir computing [47], causation entropy [48], transfer entropy [49], and Granger causality [50]. In comparison, the study of hypergraph inference is still in its infancy [4]. Luckily, progress is happening rapidly. Some recently proposed methods include probabilistic inference based on prior network data [51–53] or binary contagion data [54], and optimization-based methods that apply to time-series data [55]. These methods, however, are limited by their strong assumptions of the generative process and/or reliance on knowledge of the model dynamics. When the underlying model is unknown, the literature on hypergraph inference is rather scarce. To the best of our knowledge, the only model-free method in this space is the *Algorithm for Revealing Network Interactions (ARNI)* [41]. Although the original version of ARNI mostly focused on pairwise interactions, it can be adapted to infer higher-order interactions.

Here, we propose the *Taylor-based Hypergraph Inference using SINDy (THIS)* algorithm, which reconstructs hypergraphs and simplicial complexes from time-series data. Our method is system-agnostic and noninvasive. In particular, it does not require knowledge of the node dynamics or coupling functions and works with generic basis functions (i.e., no need to curate different nonlinear feature libraries for each application). Neither does it require perturbing the system in precise ways (e.g., via control input injection). Importantly, the method takes advantage of the intrinsic sparsity of most real-world hypergraphs through the use of the SINDy algorithm [56]. We apply THIS to both synthetic data generated from canonical dynamical systems and resting-state EEG data from 109 human subjects. With the synthetic data, we show that THIS can be extremely data efficient and compares favorably against ARNI in accuracy. From the EEG data, we find that higher-order interactions play a key role in shaping the macroscopic brain dynamics (despite the fact that most physical connections between brain regions are pairwise [30]).

* robin.delabays@hevs.ch

† degiulia@ethz.ch

‡ dorfler@ethz.ch

§ yzhang@santafe.edu

RESULTS

When is hypergraph reconstruction possible?

We consider n nonlinear systems coupled through a hypergraph, as described by the following equations:

$$\begin{aligned} \dot{x}_i = F_i(\mathbf{x}) = & f_i(x_i) + \sum_{j=1}^n a_{ij}^{(2)} h^{(2)}(x_i, x_j) \\ & + \sum_{j,k=1}^n a_{ijk}^{(3)} h^{(3)}(x_i, x_j, x_k) + \dots, \quad i = 1, \dots, n, \end{aligned} \quad (1)$$

where f_i describes the intrinsic dynamics of node i . The adjacency tensor $\mathbf{A}^{(p)} = \{a_{ij\ell\dots\ell}^{(p)}\}$ determines which nodes are coupled through the p -th order interaction function $h^{(p)}$.

Before developing an inference method, it is helpful to characterize when is inference theoretically possible. For this purpose, we assume that it is possible to observe all nodes in the system for many different initial conditions at high resolution without the interference of noise. In this perfect and abundant data limit, the question becomes: Can we determine the hypergraph structure uniquely from the vector field? When will there be ambiguity? Below, for simplicity and without loss of generality, we consider systems with interactions up to the third order (i.e., $\mathbf{A}^{(p)} = \mathbf{0}$ for $p > 3$).

Neuhäuser et al. [18] recently pointed out that, for linear consensus dynamics, pairwise and higher-order interactions are theoretically indistinguishable from each other. Here, we generalize this observation to any dynamical systems on hypergraphs. If the triadic interaction function $h^{(3)}(x_i, x_j, x_k)$ can be written as a linear combination of pairwise interaction functions $h^{(3)}(x_i, x_j, x_k) = h'(x_i, x_j) + h''(x_i, x_k) + h'''(x_j, x_k)$, then there is no formal difference between a 2-simplex and a closed triangle of links. In other words, for any hypergraph with linearly decomposable interaction functions, we can always find a corresponding network that generates the same dynamics. As a consequence, hypergraph reconstruction would be impossible due to such ambiguity. More generally, if we only have data from a localized region of the state space (e.g., by observing the system respond to small noise around a fixed point), inside which the dynamics can be effectively linearized, then it is not possible to reconstruct the hypergraph from dynamics even when the interactions are nonlinear.

On the other hand, if a higher-order interaction cannot be decomposed into a linear combination of pairwise interactions, then it will produce a vector field that is different from any vector field generated with only pairwise interactions. This difference can then, in principle, be used to infer the existence of higher-order interactions. Next, we propose a simple strategy to extract this information, which can then be used to reconstruct the hypergraph.

Taylor-based Hypergraph Inference using SINDy (THIS)

In Eq. (1), one can write the Taylor expansion of the dynamics F_i around a chosen point \mathbf{x}_0 , leading to

$$\begin{aligned} \dot{x}_i \approx & F_i(\mathbf{x}_0) + \sum_j \partial_j F_i(\mathbf{x}_0) \Delta x_j \\ & + \frac{1}{2!} \sum_{j,k} \partial_{j,k} F_i(\mathbf{x}_0) \Delta x_j \Delta x_k + \dots \end{aligned} \quad (2)$$

with $\Delta x_i = x_i - x_{0,i}$, $\forall i \in \{1, \dots, n\}$. One realizes that, for $i \neq j \neq k$, if the coefficient $\partial_{j,k} F_i(\mathbf{x}_0)$ is nonzero, then there is necessarily a triadic interaction involving nodes i , j , and k (more precisely, nodes j and k would influence node i jointly through a directed triadic interaction). One can see this by noticing that $\partial_{j,k} h^{(2)}(x_i, x_j) = 0$ for $i \neq k$, whereas $\partial_{j,k} h^{(3)}(x_i, x_j, x_k)$ is generally nonzero. Our approach utilizes this observation and aims to infer the nonzero coefficients $\partial_{j,k} F_i$ from time series.

This Taylor-based approach is intrinsically system-agnostic and does not rely on knowing the node dynamics f_i or coupling functions $h^{(p)}$. Moreover, the Taylor expansion can be computed around any point \mathbf{x}_0 where the vector field is differentiable, rendering the approach flexible in terms of where data are collected. In rare circumstances (usually with zero probability), a Taylor coefficient could vanish at \mathbf{x}_0 even when the corresponding interaction exists. We can easily circumvent this problem by choosing a different base point \mathbf{x}_0 .

To recover the coefficients of the Taylor expansion based on the time series $\mathbf{x}(t)$, we use SINDy [56] with a library of monomials up to a chosen degree (see Methods section for details). The purpose of SINDy is to find a parsimonious linear combination of the library functions that best explains the time-series data $\mathbf{X} = \{\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_K)\}$. Each nonzero coefficient obtained by SINDy selects a monomial from the library, which can in turn be used to infer the corresponding hyperedge. Specifically, we consider a hyperedge to exist if the coefficient of the corresponding monomial is above a prespecified threshold ϵ . The threshold ϵ has overlapping functionalities with the sparsity parameter in SINDy. We introduce ϵ here because it facilitates the computation of the Receiver Operating Characteristic (ROC) curves, which measure the quality of the inference. Note that since we don't know the form of the coupling functions, there is no point in inferring the weights of the hyperedges, so we focus on the binary inference problem. Namely, is there a hyperedge or not? We illustrate our inference procedure in Fig. 1.

Our main rationale for using SINDy is that many real-world hypergraphs are intrinsically sparse. As each nonzero coefficient corresponds to a hyperedge, an algorithm promoting sparsity is desirable in most cases. Of course, dense hypergraphs do exist. However, in such cases, the actual hypergraph structure has only marginal importance, as most phenomena on dense hypergraphs

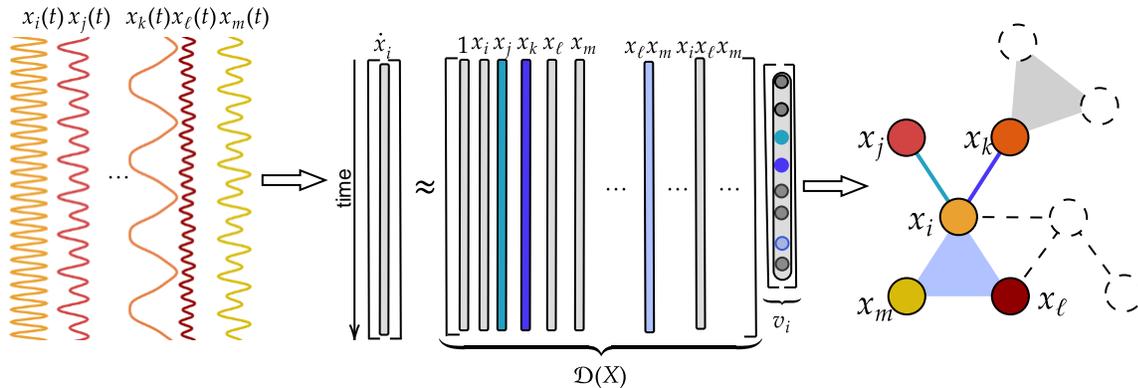


FIG. 1. **Illustration of Taylor-based Hypergraph Inference using SINDy (THIS).** Here, we focus on inferring the couplings received by node i . Inference for any other node can be done independently using the same procedure. For each inference task, the input is the time series measured for all nodes (left), and the output is the inferred connections pointing towards node i (right). The key step of the inference is solving the matrix equation $\dot{\mathbf{x}}_i = \mathcal{D}(\mathbf{X})\mathbf{v}_i$ (middle). On the left-hand side, $\dot{\mathbf{x}}_i = [\dot{x}_i(t_1), \dot{x}_i(t_2), \dots, \dot{x}_i(t_K)]^\top$ is a column vector consisting of the derivative of x_i at different time points. On the right-hand side, $\mathcal{D}(\mathbf{X})$ is the data matrix obtained by applying nonlinear features (i.e., the monomials) to the time series, and \mathbf{v}_i is a sparse vector that approximately solves the matrix equation. The nonzero elements of \mathbf{v}_i are used to infer the existence of (hyper)edges. In this example, there is a triadic coupling from nodes ℓ and m to node i because the coefficient for $x_\ell x_m$ is nonzero.

are well captured by mean-field approximations. On top of that, there are three additional advantages of using SINDy. First, we can naturally control the trade-off between sensitivity and specificity of the reconstruction by tuning the sparsity parameter in SINDy. Second, inference can be done independently for each node, rendering our method easily parallelizable, which is important when dealing with high-dimensional time series from large systems. Finally, the SINDy-based approach also allows us to use mature, off-the-shelf implementations of the algorithm that are user-friendly and highly optimized [57]. For these reasons, we focus on the SINDy implementation of our approach in this paper. However, it is possible to combine our Taylor-based approach with other data-driven methods, such as EDMD [58].

We remark that THIS has a few similarities with ARNI [41]. Indeed, the basic idea of both approaches is to represent the unknown coupling functions as linear combinations of a set of nonlinear basis functions. Namely, monomials for THIS and a user-specified function library for ARNI. However, the ways in which THIS and ARNI identify the coefficients of the decomposition (i.e., the optimization strategies) are drastically different. For ARNI, the coefficients are determined by consecutively projecting the time series onto the subspaces spanned by the basis functions. THIS, on the other hand, performs inference through a sparsity-promoting optimization procedure. Next, we compare the performance of the two algorithms under identical conditions using synthetic data.

Benchmark on synthetic data

Kuramoto oscillators with nonpairwise coupling. We

first test ARNI and THIS using a generalization of the Kuramoto model:

$$\begin{aligned} \dot{\theta}_i = & \omega_i + \sum_{j=1}^n a_{ij}^{(2)} \sin(\theta_j - \theta_i) \\ & + \sum_{j,k=1}^n a_{ijk}^{(3)} \sin(\theta_j + \theta_k - 2\theta_i), \quad i = 1, \dots, n, \end{aligned} \quad (3)$$

where $\theta_i \in S^1$ represents the phase of oscillator i and ω_i is its natural frequency. We use the seven-node hypergraph shown in Fig. 2 for this test (see Supplementary Fig. 5 for results on random graphs). Samples are taken randomly and uniformly inside a hypercube of side length 0.1 centered at the origin (one can use hypercubes of side length up to 1.0 without affecting the results, see Supplementary Fig. 6). For each data point, we compute the derivatives directly using Eq. (3).

The output of both inference methods (THIS and ARNI with power-series basis) is a weighted adjacency tensor $\mathbf{A}^{(p)}$ for each interaction order p . One then needs to choose a threshold ϵ such that a hyperedge is considered present if and only if the corresponding coefficient in $\mathbf{A}^{(p)}$ is larger than ϵ . For a given ϵ , the true positive rate (TPR) [resp. false positive rate (FPR)] is the percentage of existing [resp. nonexistent] hyperedges that were inferred. The ROC curves plot TPR against FPR for all possible threshold values.

In Fig. 2, we show the ROC curves for THIS and ARNI applied to samples of varying sizes. Among the different ROC curves, we vary the sample size from 10 to 150 (curves with brighter colors use more data points). Along each ROC curve, we gradually decrease the inference threshold ϵ from ∞ to 0. Thus, each curve starts at the lower left corner (0, 0), where no (hyper)edge is inferred,

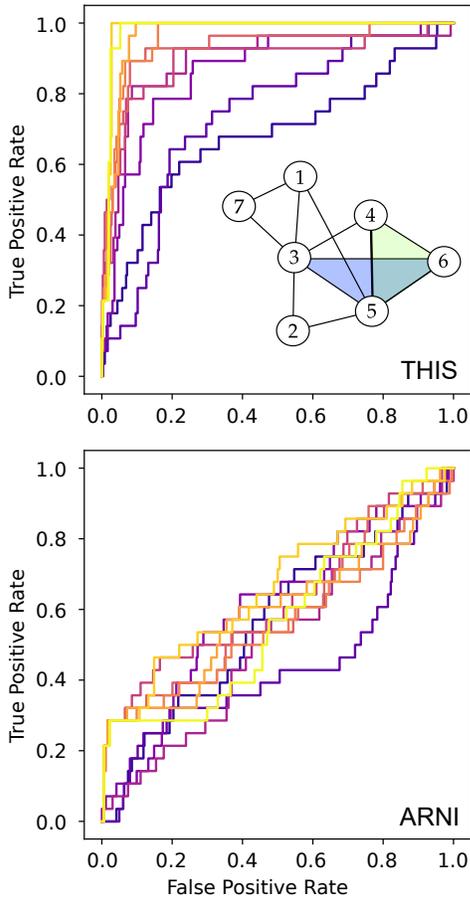


FIG. 2. **THIS and ARNI applied to Kuramoto dynamics.** We infer the hypergraph shown in the top panel. The ROC curves measure the inference quality considering all orders of interactions (both pairwise and triadic). Each curve corresponds to a different sample size used in inference, ranging from 10 data points (dark purple) to 150 data points (bright yellow). The quicker the ROC curve rises to the upper left corner, the better the inference.

and ends at the upper right corner (1,1), where all possible (hyper)edges are inferred. Since we want high TPR and low FPR, an ideal ROC curve should rise steeply from the lower left corner to the upper left corner. A diagonal curve is indicative of performance comparable to random guesses. We see that even with just 10 data points, THIS achieves good accuracy in reconstructing the hypergraph. In contrast, ARNI barely outperforms random guesses even with 150 data points.

Lorenz oscillators with nonpairwise coupling. Next, we apply THIS to Lorenz oscillators coupled through pairwise and triadic interactions:

$$\begin{aligned} \dot{x}_i &= \sigma(y_i - x_i) + \sum_{j=1}^n a_{ij}^{(2)}(x_j - x_i) + \sum_{j,k=1}^n a_{ijk}^{(3)}(x_j x_k^2 - x_i^3), \\ \dot{y}_i &= x_i(\rho - z_i) - y_i, \\ \dot{z}_i &= x_i y_i - \beta z_i, \end{aligned}$$

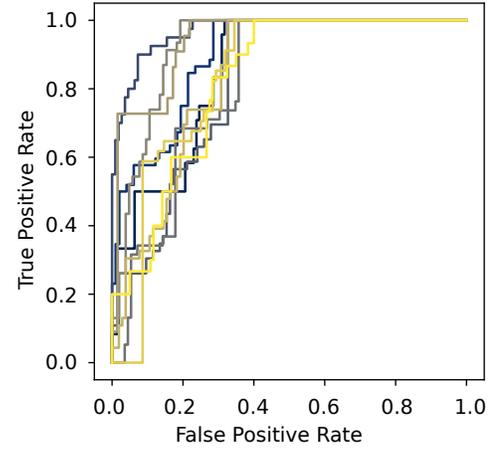


FIG. 3. **THIS applied to Lorenz oscillators on random hypergraphs.** Each ROC curve corresponds to the inference of a different five-node random hypergraph. To test the robustness of THIS, we made the hypergraphs non-sparse ($p = 0.5$ for both pairwise and triadic connections) and estimated derivatives from numerical differentiation. Since each Lorenz oscillator has three degrees of freedom, we are essentially performing inference on a 15-dimensional system.

where we set $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. Here, each node has three degrees of freedom. We perform inference on the whole system as if it were a $3n$ -node system. Then, we aggregate the cross-node interactions: For each pair of nodes, say i and j , we use the largest inferred coefficient between any degree of freedom of node i and any degree of freedom of node j as $a_{ij}^{(2)}$. We perform the same procedure to determine $a_{ijk}^{(3)}$.

In Fig. 3, we show that our method performs well even for chaotic dynamics and nodes with more than one dimension. Each ROC curve corresponds to a different five-node hypergraph generated using XGI's `random_hypergraph` function [60], with a probability of 0.5 for both 2-edges and 3-edges. For each hypergraph, we use 10 independent time series for the inference, each consisting of 150 time steps with a step size $\delta_t = 0.01$, and we approximate the derivatives through finite differences. The time series are all initialized within a hypercube of side length 0.7, centered at a random point chosen uniformly from $[-1, 1]^{3n}$. For most hypergraphs, we can reach over 80% TPR with less than 20% FTR. The performance can be further improved if we use more data points, optimize hyperparameters such as the sampling box size, or calculate the derivatives directly from the underlying Lorenz equations.

We note that for THIS, the existence of higher-order interactions could mask the existence of lower-order interactions. This stems from an unavoidable drawback of relying on the Taylor expansion: Higher-order interactions also contribute to lower-order Taylor coefficients. Indeed, the existence of interactions of order p implies that the corresponding coefficients of order lower than p can be nonzero as well. In other words, when a p -edge e

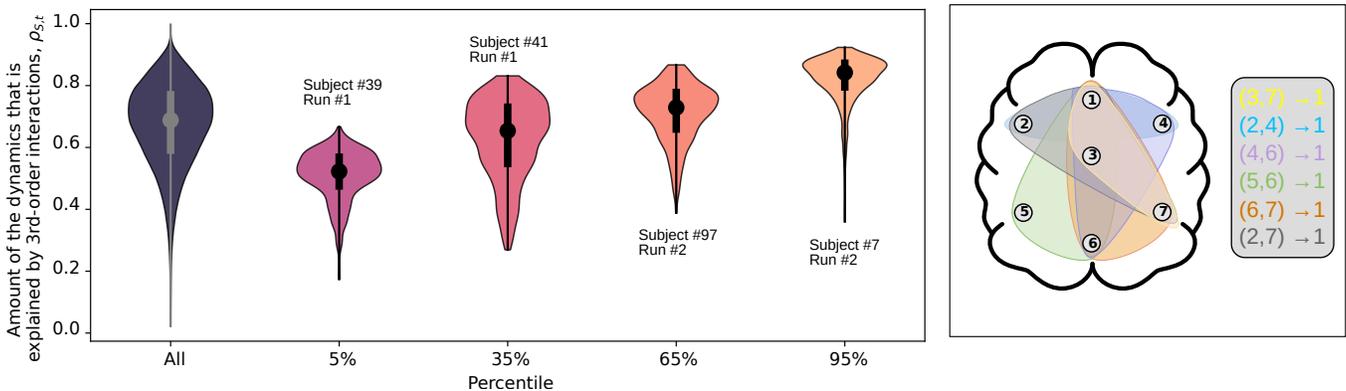


FIG. 4. **Higher-order interactions play a significant role in shaping macroscopic brain dynamics.** Left panel: The percentage of dynamics (i.e., derivatives) contributed by the inferred triadic couplings, $\rho_{s,t}$. See Eq. (4) for the mathematical definition of $\rho_{s,t}$. The left-most violin shows the distribution of $\rho_{s,t}$ aggregating all 218 time series s over all time t . The five other violins show the distribution of $\rho_{s,t}$ for one time series each. The time series displayed are the ones whose median ratio is at the 5, 35, 65, and 95 percentiles, respectively. Right panel: Illustration of the seven brain areas and the six most frequently inferred 3-edges (they all appear in about 73% of the 218 inferred hypergraphs). Interestingly, the top six hyperedges are all pointed towards area 1, which roughly corresponds to the prefrontal cortex (Supplementary Fig. 7). This makes sense because the prefrontal cortex is highly interconnected with the rest of the brain, known to be involved in a wide range of higher-order cognitive functions, and considered one of the key information processing hubs in the brain [59].

is present in the hypergraph, THIS will likely infer some lower-order edges contained in e even if they did not exist, potentially leading to false positives. Despite this limitation, THIS performed well in our benchmark with random hypergraphs (Fig. 3). We will discuss practical strategies to mitigate this issue in the Discussion section.

How important are higher-order interactions in shaping macroscopic brain dynamics?

An important special case of the hypergraph reconstruction problem is the following: Given time-series data, can we tell whether networks are adequate to capture the observed dynamics, or are higher-order interactions truly needed? Answering this question has important applications in neuroscience: Although different brain regions are mostly connected through (pairwise) anatomical axons and nerve fibers, the release of neurotransmitters can potentially induce higher-order interactions among neuronal populations. So how important are non-pairwise interactions in shaping the macroscopic brain dynamics?

Below, we apply THIS to reconstruct the effective connectivity between seven brain regions using resting-state EEG data from 109 human subjects. For each subject, two independent recordings were collected, giving a total of 218 time series (see Methods section for details of the analysis protocol). In order to limit the computational burden, we restrict our focus to second and third-order interactions. Interestingly, we find that non-pairwise interactions account for more than 60% of the EEG dynamics, and this is robust across brain regions and individual subjects.

To measure the amount of dynamics that is explained by triadic interactions for time series s , $s \in \{1, \dots, 218\}$, at each time t , we compute the relative contribution of the triadic coupling terms to the derivative. Namely, if $\hat{a}_{s,ij}^{(2)}$ and $\hat{a}_{s,ijk}^{(3)}$ are the inferred adjacency tensors for time series s , we compute, for each time step t , the third-order contribution ratio

$$\rho_{s,t} = \frac{\sum_{i,j,k;i \neq j \neq k} |\hat{a}_{s,ijk}^{(3)} x_j x_k|}{\sum_{i,j;i \neq j} |\hat{a}_{s,ij}^{(2)} x_j| + \sum_{i,j,k;i \neq j \neq k} |\hat{a}_{s,ijk}^{(3)} x_j x_k|}. \quad (4)$$

We then sort the time series according to their median ratio (calculated from $\rho_{s,t}$ using all time steps). In Fig. 4, we show the distribution of $\rho_{s,t}$. The left-most violin plot shows the distribution of $\rho_{s,t}$ aggregated over all time series and all time steps, whereas the other violin plots show the distribution for individual time series from different percentiles (ranked by their median ratios). Our inference results are robust and consistent across different subjects: For the majority of the time series, between 60% and 70% of the dynamics are attributed to triadic interactions. The right panel of Fig. 4 shows the six most prominent hyperedges inferred by THIS. Interestingly, all of them involve the prefrontal cortex, which matches the expectation that the prefrontal cortex acts as a major information processing center in the brain [59].

Our finding raises the question about the neurological basis of higher-order interactions in the brain. Some potential mechanisms for creating nonpairwise interactions among neuronal populations include heterosynaptic plasticity, ephaptic transmission, extracellular fields, and metabolic regulation [61, 62]. It is worth emphasizing that the goal of THIS is not inferring functional connectivity or structural connectivity. Instead, we are

inferring effective connectivity, in the same spirit of the dynamic causal modeling paradigm [28]. Even from purely pairwise structural connectivity, nonpairwise effective connectivity can naturally emerge (e.g., through coarse-graining).

We note that currently there is no consensus on the relative importance of pairwise versus nonpairwise interactions in shaping the macroscopic brain dynamics [63–70]. Due to the intrinsic complexities of the brain, different methods applied to different neurophysiological datasets can easily lead to seemingly contradictory results. For example, a recent work [67] supports the hypothesis that linear auto-regressive models provide the best fit for macroscopic brain dynamics in resting-state conditions both in terms of one-step predictive power and computational complexity. This hypothesis was validated on both iEEG and fMRI data. Ref. [70], on the other hand, supports the existence of higher-order interactions based on partial entropy decomposition to resting-state fMRI data from human brains. Their study also suggests higher-order interactions encode significant bio-makers that distinguish between healthy and pathological states associated with anesthesia or brain injuries, and they mirror changes associated with aging. Further studies are needed to clarify the role of higher-order interactions in the brain. We hope our inference method can provide a new tool and a novel perspective on this important open question in neuroscience.

DISCUSSION

The hypergraph reconstruction problem considered here is a special (but very interesting and important) case of data-driven equation discovery. Instead of inferring the full equations, we only aim to determine which variables are linked (i.e., inferring the hyperedges). Because we only need to know the causal relationship among the variables, our algorithm is much less sensitive to the choice of the nonlinear feature library (when applying SINDy to infer the full underlying equations, it is crucial to have the correct terms in the library). This allows us to work with interconnected dynamical systems in a model-free fashion that is currently out of reach for standard equation discovery methods.

The main strength of THIS is that it can perform robust hypergraph inference with minimal prior knowledge of the underlying system—we only assume that it can be modeled as a set of coupled differential equations. Despite the system-agnostic nature of the method, we also can incorporate additional information about the system as it becomes available. For example, if we know that the underlying hypergraph is a simplicial complex, then we can utilize the downward inclusion condition to automatically infer all the lower-order interactions, circumventing one of the limitations of THIS. The same is true for hypergraphs for which no lower-order interactions overlap with higher-order interactions (i.e., anti-simplicial com-

plexes and, to first approximation, uncorrelated sparse hypergraphs). Also, if one knows the coupling functions, then they can be passed onto the SINDy algorithm to improve inference.

The price to pay for THIS’s generality and flexibility is that the sampling area for data needs to be well-balanced. Namely, the area should be large enough to adequately explore the system’s dynamics (i.e., go beyond the linearizable region) but also not too large in order to preserve the accuracy of the Taylor approximation. As long as this balance is achieved, THIS is fully flexible in terms of where data came from in the state space. In our experiments on Kuramoto and Lorenz oscillators, we found that this balance can be achieved for a wide range of sampling box sizes (Supplementary Fig. 6). A practical strategy for picking the right box size is the following: Instead of guessing one using intuition, we can test the algorithm over a wide range of box sizes and find the plateau for which the predictions stay stable. Moreover, we often have partial information about the connectivity in many applications. For instance, in the context of predicting missing links [71], the majority of the (hyper)edges are known. We can utilize this information and tune the box size to maximize the match between the inference and the ground truth for the known (hyper)edges.

There are several promising directions to further improve our method in the future. First, the size of the monomial library increases rapidly with the dimension of the system. Currently, THIS can handle systems with up to about 20 nodes. Future studies have the opportunity to refine the monomial library (e.g., through the library bagging technique [72]) and improve the computational complexity of the algorithm. Second, we can make THIS more robust against noise by taking advantage of the recent advances in SINDy algorithms, most notably the Ensemble-SINDy approach [72]. Third, in many applications, one often has to work with partial measurements. In those cases, it would be important to combine THIS with data assimilation techniques (e.g., ensemble Kalman filter [73]) to tackle missing variables. We hope this work will further stimulate the development of hypergraph inference methods so that we can leverage the torrents of time-series data gathered in diverse fields (neuroscience, genetics, sociology, etc.) to uncover previously hidden (higher-order) interactions in complex systems.

METHODS

Implementation of SINDy

In THIS, SINDy is implemented with a library of N monomials:

$$\mathcal{D} = \{1, x_1, x_2, \dots, x_n, x_1^2, x_1x_2, x_1x_3, \dots, x_n^2, x_1^3, x_1^2x_2, \dots, x_1x_2x_3, x_1x_2x_4, \dots\}.$$

The purpose of SINDy is to find a parsimonious model that best explains the time-series data \mathbf{X} =

$\{\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_K)\}$. The basic idea is to solve the matrix equations

$$\dot{\mathbf{x}}_i = \mathcal{D}(\mathbf{X})\mathbf{v}_i, \quad i = 1, \dots, n \quad (5)$$

with sparse, N -dimensional column vectors \mathbf{v}_i . The n matrix equations (one for each node) are independent of each other and can be solved in parallel. Here, $\dot{\mathbf{x}}_i = [\dot{x}_i(t_1), \dot{x}_i(t_2), \dots, \dot{x}_i(t_K)]^\top$ is a K -dimensional column vector consisting of the derivative of x_i at different time points t_k (either given as part of the data or inferred from \mathbf{x}_i). Each column of the $K \times N$ matrix $\mathcal{D}(\mathbf{X})$ corresponds to a monomial of the variables, whereas the rows encode different time points. In solving the matrix equation, we want to balance the goodness-of-fit (i.e., minimization of the mismatch between $\dot{\mathbf{x}}_i$ and $\mathcal{D}(\mathbf{X})\mathbf{v}_i$) with the sparsity of \mathbf{v}_i , which is achieved by performing sequential thresholded least square regression on Eq. (5). After solving the optimization problem, each nonzero element in \mathbf{v}_i is linked to a monomial from the library, which can in turn be used to infer the existence of hyperedges pointing towards node i .

Analysis of the EEG time series

We apply THIS to the resting-state time series from the dataset reported in Refs. [74–76], consisting of 109 subjects, each with two independent recordings. Each recording lasts one minute, with a sampling rate of 160 points per second. For each inference, we start with a 64-dimensional signal obtained from 64 sensors distributed across the scalp (Supplementary Fig. 7). For the sake of tractability and noise reduction, we divided the sensors into seven groups according to their proximity and took the average within each group, reducing the 64-dimensional signal to a 7-dimensional signal (Supplementary Fig. 7). Each dimension captures the macroscopic dynamics of one of the seven brain regions. To facilitate approximating the time derivatives through finite differ-

ences (i.e., $\dot{x}_i(t_k) = (x_i(t_{k+1}) - x_i(t_k))/\Delta t$), we apply a low-pass filter to the 7-dimensional signal, keeping only part of the signal whose frequency is below ≈ 5 [Hz]. Additionally, we normalize the time series so their standard deviation is 1. This last step is important because having variances much larger or smaller than 1 induces large discrepancies in the order of magnitude for the monomials of different degrees, rendering SINDy’s thresholding ill-conditioned. Finally, as THIS is valid in a restricted domain of the state space, we apply it only to the 1000 data points that are closest to the median. With the above protocol, we were able to achieve an average relative fitting error below 50%. We also tested applying our algorithm only to the 34 time series with a relative fitting error below 20%, and the results remain qualitatively unchanged.

ACKNOWLEDGMENTS

We thank Federico Battiston, Maxime Lucas, Mason Porter, Andrew Stier, Marc Timme, and David Wolpert for insightful discussions. We are grateful to Benedetta Franceschiello and Giovanni Petri for their guidance in analyzing and interpreting the EEG data. R.D. was supported by the Swiss National Science Foundation under grants P400P2_194359 and 200021_215336. Y.Z. acknowledges support from the Omidyar Fellowship. F.D. was supported by the National Centre of Competence in Research “Dependable, ubiquitous automation”: <https://nccr-automation.ch/>.

DATA AVAILABILITY

The EEG data were retrieved from the PhysioNet database [74–76].

CODE AVAILABILITY

All codes are available online from the repository <https://doi.org/10.5281/zenodo.10530471>.

-
- [1] Lambiotte, R., Rosvall, M. & Scholtes, I. From networks to optimal higher-order models of complex systems. *Nature Physics* **15**, 313–320 (2019).
- [2] Battiston, F. *et al.* Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports* **874**, 1–92 (2020).
- [3] Torres, L., Blevins, A. S., Bassett, D. & Eliassi-Rad, T. The Why, How, and When of Representations for Complex Systems. *SIAM Review* **63**, 435–485 (2021).
- [4] Battiston, F. *et al.* The physics of higher-order interactions in complex systems. *Nature Physics* **17**, 1093–1098 (2021).
- [5] Petri, G. *et al.* Homological scaffolds of brain functional networks. *Journal of the Royal Society Interface* **11**, 1–10 (2014).
- [6] Giusti, C., Christ, R. & Bassett, D. S. Two’s company, three (or more) is a simplex. *Journal of Computer Neuroscience* **41**, 1–14 (2016).
- [7] Benson, A. R., Abebe, R., Schaub, M. T., Jadbabaie, A. & Kleinberg, J. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences* **115**, E11221–E11230 (2018).
- [8] Cencetti, G., Battiston, F., Lepri, B. & Karsai, M. Temporal properties of higher-order interactions in social networks. *Scientific Reports* **11**, 7028 (2021).
- [9] Grilli, J., Barabás, G., Michalska-Smith, M. J. & Allesina, S. Higher-order interactions stabilize dynamics in competitive network models. *Nature* **548**, 210–213

- (2017).
- [10] Bick, C., Ashwin, P. & Rodrigues, A. Chaos in generically coupled phase oscillator networks with nonpairwise interactions. *Chaos* **26**, 094814 (2016).
 - [11] Skardal, P. S. & Arenas, A. Higher order interactions in complex networks of phase oscillators promote abrupt synchronization switching. *Communications Physics* **3**, 218 (2020).
 - [12] Arnaudon, A., Peach, R. L., Petri, G. & Expert, P. Connecting hodge and sakaguchi-kuramoto through a mathematical framework for coupled oscillators on simplicial complexes. *Communications Physics* **5**, 211 (2022).
 - [13] Zhang, Y., Lucas, M. & Battiston, F. Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. *Nature Communications* **14**, 1605 (2023).
 - [14] Sun, H., Radicchi, F., Kurths, J. & Bianconi, G. The dynamic nature of percolation on networks with triadic interactions. *Nature Communications* **14**, 1308 (2023).
 - [15] Ferraz de Arruda, G., Petri, G., Rodriguez, P. M. & Moreno, Y. Multistability, intermittency, and hybrid transitions in social contagion models on hypergraphs. *Nature Communications* **14**, 1375 (2023).
 - [16] Zhang, Y., Skardal, P. S., Battiston, F., Petri, G. & Lucas, M. Deeper but smaller: Higher-order interactions increase linear stability but shrink basins. *arXiv:2309.16581* (2023).
 - [17] Carletti, T., Battiston, F., Cencetti, G. & Fanelli, D. Random walks on hypergraphs. *Physical Review E* **101**, 022308 (2020).
 - [18] Neuhäuser, L., Mellor, A. & Lambiotte, R. Multibody interactions and nonlinear consensus dynamics on networked systems. *Physical Review E* **101**, 032310 (2020).
 - [19] DeVille, L. Consensus on simplicial complexes: Results on stability and synchronization. *Chaos* **31**, 023137 (2021).
 - [20] Iacopini, I., Petri, G., Barrat, A. & Latora, V. Simplicial models of social contagion. *Nature Communications* **10**, 2485 (2019).
 - [21] Ferraz de Arruda, G., Tizzani, M. & Moreno, Y. Phase transitions and stability of dynamical processes on hypergraphs. *Communications Physics* **4**, 1–9 (2021).
 - [22] Millán, A. P., Torres, J. J. & Bianconi, G. Explosive Higher-Order Kuramoto Dynamics on Simplicial Complexes. *Physics Review Letters* **124**, 218301 (2020).
 - [23] Lucas, M., Cencetti, G. & Battiston, F. Multiorder Laplacian for synchronization in higher-order networks. *Physics Review Research* **2**, 033410 (2020).
 - [24] Zhang, Y., Latora, V. & Motter, A. E. Unified treatment of synchronization patterns in generalized networks with higher-order, multilayer, and temporal interactions. *Communications Physics* **4**, 195 (2021).
 - [25] Salova, A. & D'Souza, R. M. Cluster synchronization on hypergraphs. *arXiv:2101.05464* (2021).
 - [26] Gambuzza, L. V. *et al.* Stability of synchronization in simplicial complexes. *Nature Communications* **12**, 1255 (2021).
 - [27] Chen, C., Suran, A., Bloch, A. M. & Rajapakse, I. Controllability of hypergraphs. *IEEE Transaction Network Science and Engineering* **8**, 1646–1657 (2021).
 - [28] Friston, K. J., Harrison, L. & Penny, W. Dynamic causal modelling. *Neuroimage* **19**, 1273–1302 (2003).
 - [29] Hagmann, P. *et al.* Mapping the structural core of human cerebral cortex. *PLoS Biology* **6**, e159 (2008).
 - [30] Bullmore, E. & Sporns, O. The economy of brain network organization. *Nature Reviews Neuroscience* **9**, 336–349 (2009).
 - [31] Faskowitz, J., Betzel, R. F. & Sporns, O. Edges in brain networks: Contributions to models of structure and function. *Network Neuroscience* **6**, 1–28 (2022).
 - [32] Bradshaw, W. J., Alley, E. C., Huggins, J. H., Lloyd, A. L. & Esvelt, K. M. Bidirectional contact tracing could dramatically improve COVID-19 control. *Nature Communications* **12**, 232 (2021).
 - [33] Kojaku, S., Hébert-Dufresne, L., Mones, E., Lehmann, S. & Ahn, Y.-Y. The effectiveness of backward contact tracing in networks. *Nature Physics* **17**, 652–658 (2021).
 - [34] Zu, C. *et al.* Identifying disease-related subnetwork connectome biomarkers by sparse hypergraph learning. *Brain Imaging and Behaviour* **13**, 879–892 (2018).
 - [35] Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
 - [36] Yu, D., Righero, M. & Kocarev, L. Estimating topology of networks. *Physics Review Letters* **97**, 1–4 (2006).
 - [37] Timme, M. Revealing network connectivity from response dynamics. *Physics Review Letters* **98**, 224101 (2007).
 - [38] Lü, L. & Zhou, T. Link prediction in complex networks: A survey. *Physica A* **390**, 1150–1170 (2011).
 - [39] Kralemann, B., Pikovsky, A. & Rosenblum, M. Reconstructing phase dynamics of oscillator networks. *Chaos* **21**, 025104 (2011).
 - [40] Han, X., Shen, Z., Wang, W.-X. & Di, Z. Robust reconstruction of complex networks from sparse data. *Physics Review Letters* **114**, 028701 (2015).
 - [41] Casadiego, J., Nitzan, M., Hallerberg, S. & Timme, M. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nature Communications* **8**, 2192 (2017).
 - [42] Newman, M. E. Network structure from rich but noisy data. *Nature Physics* **14**, 542–545 (2018).
 - [43] Eroglu, D., Tanzi, M., van Strien, S. & Pereira, T. Revealing dynamics, communities, and criticality from data. *Physics Review X* **10**, 021047 (2020).
 - [44] Tyloo, M., Delabays, R. & Jacquod, P. Reconstructing network structures from partial measurements. *Chaos* **31**, 103117 (2021).
 - [45] Mangan, N. M., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* **2**, 52–63 (2016).
 - [46] Peixoto, T. P. Network reconstruction and community detection from dynamics. *Physics Review Letters* **123**, 128301 (2019).
 - [47] Banerjee, A., Hart, J. D., Roy, R. & Ott, E. Machine learning link inference of noisy delay-coupled networks with optoelectronic experimental tests. *Physics Review X* **11**, 031014 (2021).
 - [48] Sun, J., Taylor, D. & Bollt, E. M. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems* **14**, 73–106 (2015).
 - [49] Novelli, L., Wollstadt, P., Mediano, P., Wibral, M. & Lizier, J. T. Large-scale directed network inference with multivariate transfer entropy and hierarchical statistical testing. *Network Neuroscience* **3**, 827–847 (2019).

- [50] Deshpande, A., Chu, L.-F., Stewart, R. & Gitter, A. Network inference with granger causality ensembles on single-cell transcriptomics. *Cell Reports* **38**, 110333 (2022).
- [51] Young, J.-G., Petri, G. & Peixoto, T. P. Hypergraph reconstruction from network data. *Communications Physics* **4**, 135 (2021).
- [52] Contisciani, M., Battiston, F. & De Bacco, C. Inference of hyperedges and overlapping communities in hypergraphs. *Nature Communications* **13**, 7229 (2022).
- [53] Wang, Y. & Kleinberg, J. Supervised hypergraph reconstruction. *arXiv preprint arXiv:2211.13343* (2022).
- [54] Wang, H., Ma, C., Chen, H.-S., Lai, Y.-C. & Zhang, H.-F. Full reconstruction of simplicial complexes from binary contagion and ising data. *Nature Communications* **13**, 3043 (2022).
- [55] Malizia, F. *et al.* Reconstructing higher-order interactions in coupled dynamical systems. *arXiv:2305.06609* (2023).
- [56] Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **113**, 3932–3937 (2016).
- [57] Kaptanoglu, A. A. *et al.* PySINDy: A comprehensive python package for robust sparse system identification. *arXiv:2111.08481* (2021).
- [58] M. O. Williams, C. W. R., I. G. Kevrekidis. A data-driven approximation of the koopman operator: Extending dynamic mode decomposition. *Journal of Nonlinear Science* **25**, 1307–1346 (2015).
- [59] Fuster, J. *The prefrontal cortex* (Academic press, 2015).
- [60] Landry, N. W. *et al.* XGI: A Python package for higher-order interaction networks. *Journal of Open Source Software* **8**, 5162 (2023).
- [61] Dudek, F. E., Yasumura, T. & Rash, J. E. ‘Non-synaptic’ mechanisms in seizures and epileptogenesis. *Cell Biology International* **22**, 793–805 (1998).
- [62] Ikeda, K., Kataoka, M. & Tanaka, N. K. Nonsynaptic transmission mediates light context-dependent odor responses in drosophila melanogaster. *Journal of Neuroscience* **42**, 8621–8628 (2022).
- [63] Tang, A. *et al.* A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *Journal of Neuroscience* **28**, 505–518 (2008).
- [64] Watanabe, T. *et al.* A pairwise maximum entropy model accurately describes resting-state human brain networks. *Nature Communications* **4**, 1370 (2013).
- [65] van der Plas, T. L. *et al.* Neural assemblies uncovered by generative modeling explain whole-brain activity statistics and reflect structural connectivity. *eLife* **12**, e83139 (2023).
- [66] Rosch, R. E., Burrows, D. R., Lynn, C. W. & Ashourvan, A. Spontaneous brain activity emerges from pairwise interactions in the larval zebrafish brain. *arXiv:2309.05939* (2023).
- [67] Nozari, E. *et al.* Macroscopic resting-state brain dynamics are best described by linear models. *Nature Biomedical Engineering* (2023).
- [68] Tanaka, T. & Aoyagi, T. Multistable attractors in a network of phase oscillators with three-body interactions. *Physics Review Letters* **106**, 224101 (2011).
- [69] Santoro, A., Battiston, F., Petri, G. & Amico, E. Higher-order organization of multivariate time series. *Nature Physics* **19**, 221–229 (2023).
- [70] Varley, T. F., Pope, M., Grazia, M., Joshua & Sporns, O. Partial entropy decomposition reveals higher-order information structures in human brain activity. *Proceedings of the National Academy of Sciences* **120**, e2300888120 (2023).
- [71] Clauset, A., Moore, C. & Newman, M. E. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).
- [72] Fasel, U., Kutz, J. N., Brunton, B. W. & Brunton, S. L. Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proceedings Royal Society A* **478**, 20210904 (2022).
- [73] Evensen, G. *et al.* *Data assimilation: the ensemble Kalman filter*, vol. 2 (Springer, 2009).
- [74] Physionet data base. URL <https://physionet.org/content/eegmmidb/1.0.0/>.
- [75] Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N. & Wolpaw, J. R. BCI2000: A general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering* **51**, 1034–1043 (2004).
- [76] Goldberger, A. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]* **101**, e215–e220 (2000).

AUTHOR CONTRIBUTIONS

R.D., G.D.P., and Y.Z. designed the research, developed the methodology, and derived the theoretical results. All authors analyzed the numerical results and wrote the manuscript. R.D. wrote the code and ran the simulations.

COMPETING INTERESTS

The authors declare no competing interests.

Supplementary Information

Hypergraph reconstruction from dynamics

Robin Delabays, Giulia De Pasquale, Florian Dörfler, and Yuanzhao Zhang

I. NETWORK INFERENCE, ARNI VS. THIS

Here, we compare ARNI [41] and THIS using synthetic data generated by Kuramoto oscillators on random graphs. The ROC curves are shown in Fig. 5 for THIS (top left) and ARNI with five different function bases. The different curves are colored according to the length of the time series used (the darker the shorter). We see that with suitable basis functions, ARNI can perform as well as THIS, but a non-ideal choice of basis functions significantly reduces ARNI's performance.

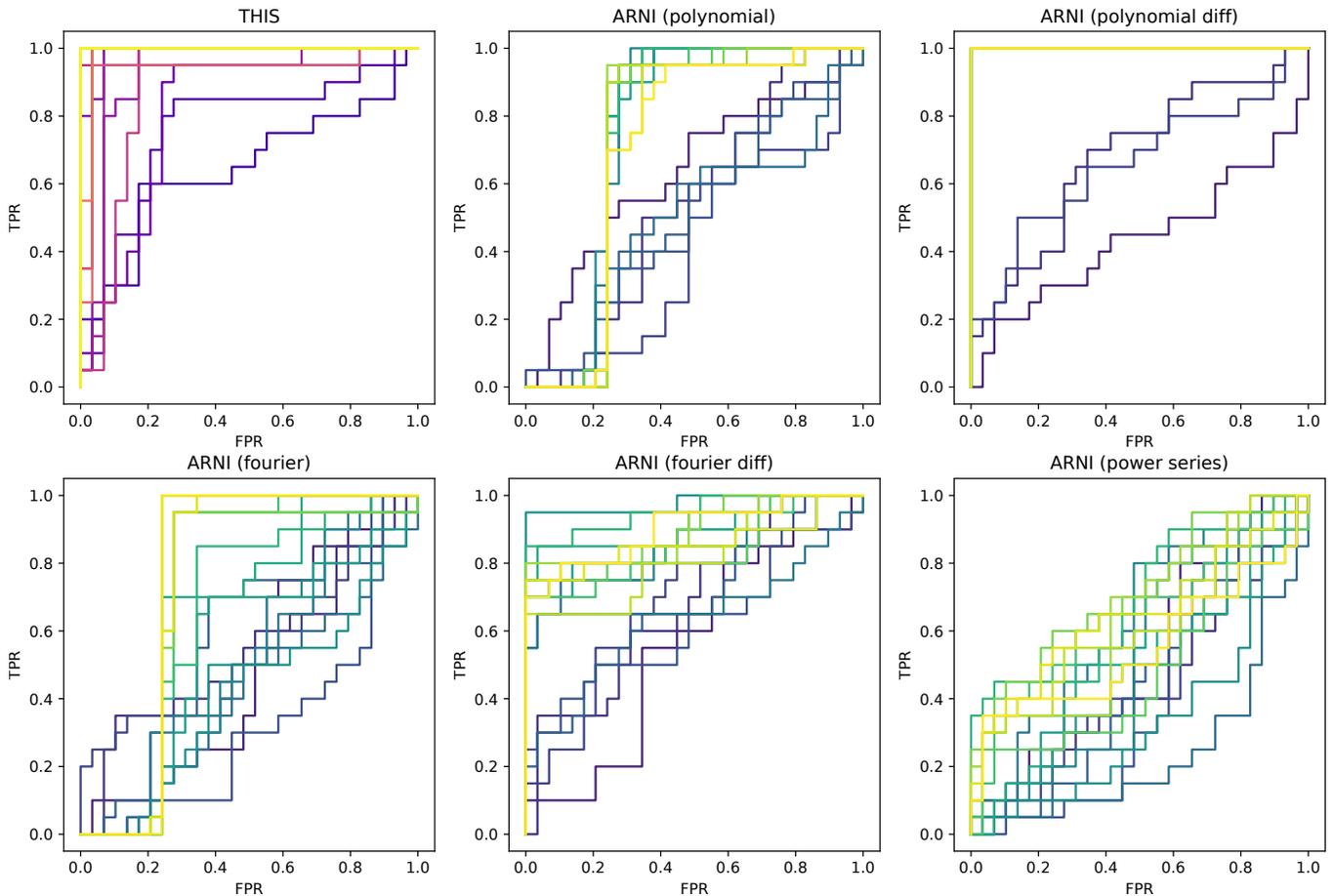


FIG. 5. **Comparison between ARNI and THIS on the task of inferring pairwise interactions.** Each panel shows the ROC curves for the inference of the adjacency matrix, based on a time series generated by seven Kuramoto oscillators coupled through an Erdős-Rényi random graph ($p = 0.5$). The same time series are used in each panel. Each curve corresponds to a different length of the time series (the darker the shorter), ranging from 8 to 80 data points. Each point is sampled uniformly from a box of size 1.5 (radians) centered at the origin and the exact time derivative is computed from the ODE at each time point. The top left panel is our method, the others are ARNI with the five main function bases proposed in Ref. [41]. Namely, polynomial basis, polynomial basis of differences, Fourier basis, Fourier basis of differences, and power series basis (see [41] for details). Our method achieves similar performance as the best version of ARNI, but much better than ARNI with a non-ideal basis (there is no way to know which basis is better a priori when using ARNI). Moreover, while we focused on inferring pairwise interactions here when using ARNI, we made the task harder for THIS by allowing it to infer third-order interactions as well.

II. TUNING THE BOX SIZE

As mentioned in the main text, THIS requires the time series to be sampled in a domain large enough to sufficiently explore the dynamics, but not too large so as to preserve the accuracy of the Taylor approximation. In Fig. 6, we show the AUC of the inference as a function of the size of the sampling domain. We observe that there is a wide range of box sizes for which THIS performs well. For the sake of comparison, we tested ARNI with the same data and found that the box size does not influence its performance.

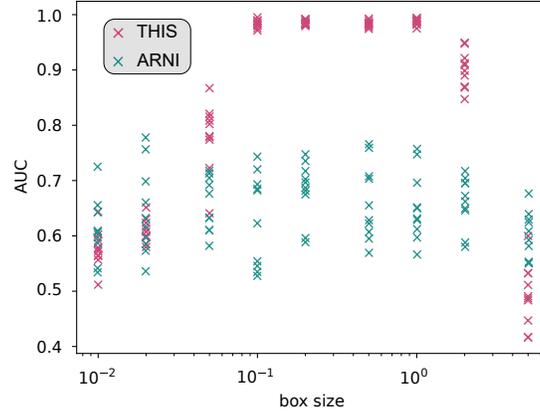


FIG. 6. **Quality of the inference with respect to the size of the sampling area.** The AUC is the area under the ROC curve for the inference (1 means perfect inference while 0.5 means performance comparable to random guesses). For each box size, we generate ten hypergraphs with Kuramoto-like dynamics. For each test, the inference is performed with 100 initial conditions and perfect measurement of the time derivatives. For THIS (pink), we see a trade-off between a box that is too small to accurately represent the dynamics and a box that is too large to admit a faithful Taylor approximation. For ARNI (green), the box size does not influence its performance.

III. SENSOR PLACEMENT AND SIGNAL AGGREGATION

Fig. 7 shows the location of 64 sensors on the scalp, and their aggregation into seven groups, inside which the time series were averaged before the inference.

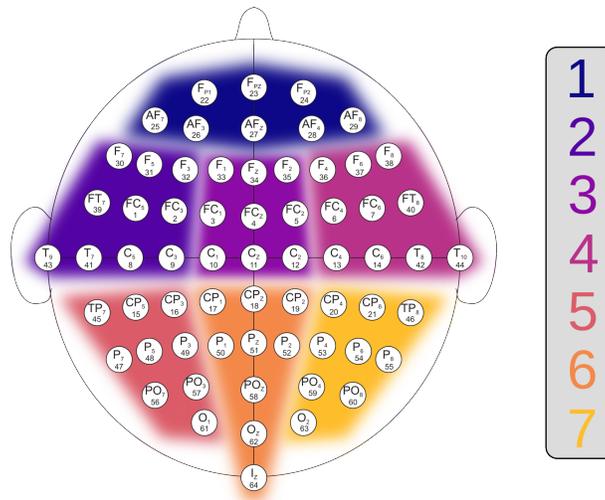


FIG. 7. **Location of the 64 sensors for the EEG measurements.** The colors indicate the seven aggregation areas used in our analysis.