

PAPER • OPEN ACCESS

Bayesian hypergraph inference from scarce and noisy dynamical observations

To cite this article: Katerina Tang *et al* 2026 *J. Phys. Complex.* **7** 045009

View the [article online](#) for updates and enhancements.

You may also like

- [Message-passing on hypergraphs: detectability, phase transitions and higher-order information](#)
Nicolò Ruggeri, Alessandro Lonardi and Caterina De Bacco
- [Data-driven inference of complex system dynamics: A mini-review](#)
Ting-Ting Gao and Gang Yan
- [Physics-inspired machine learning detects 'unknown unknowns' in networks: discovering network boundaries from observable dynamics](#)
Moshir Harsh, Leonhard Götz Vulpius and Peter Sollich



PAPER

OPEN ACCESS

RECEIVED
10 June 2026

REVISED
24 August 2026

ACCEPTED FOR PUBLICATION
29 September 2026

PUBLISHED
20 October 2026

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Bayesian hypergraph inference from scarce and noisy dynamical observations

Katerina Tang^{1,*} , Vivek Srikrishnan² and Jackson Kulik³

¹ Center for Applied Mathematics, Cornell University, Ithaca, NY 14853, United States of America

² Biological and Environmental Engineering, Cornell University, Ithaca, NY 14853, United States of America

³ Mechanical and Aerospace Engineering, Utah State University, Logan, UT 84322, United States of America

* Author to whom any correspondence should be addressed.

E-mail: kbt28@cornell.edu

Keywords: higher-order networks, hypergraph inference, network reconstruction, Bayesian inference, sparse regression, uncertainty quantification

Supplementary material for this article is available [online](#)

Abstract

Inferring higher-order interaction structure from observations of dynamics is a central challenge in complex systems, particularly when data are scarce, noisy, or concentrated in lower-dimensional regions of state space. We develop Bayes-THIS, a Bayesian extension of Taylor-based Hypergraph Inference using SINDy (THIS) that replaces threshold-based sparse regression with sparse Bayesian regression using automatic relevance determination. This yields adaptive, term-wise shrinkage and a posterior over inferred Taylor expansion coefficients that encode the presence or absence of interactions. Across numerical experiments, Bayes-THIS improves inference in low-data, high-noise, and ill-conditioned regimes. The Bayesian formulation also enables two capabilities absent from the original THIS framework: a posterior predictive diagnostic for assessing whether a dataset contains higher-order signal distinguishable from noise, and a posterior interval rule for pruning weakly supported hyperedges. Finally, we examine a separate, structural limitation of Taylor-based hypergraph reconstruction: higher-order interactions can produce spurious lower-order signatures that are indistinguishable from genuine lower-order edges. We find that the severity of this non-identifiability depends on cross-order support, showing that structural properties of the target hypergraph can be as important as data quality in determining what can be inferred from dynamics within this framework.

1. Introduction

The observed behaviors of many complex systems across the natural and social sciences arise from direct interactions of their component parts. Examples include gene regulatory mechanisms [20], large-scale brain activity [7], and the spread of epidemics [21, 25]. However, in many of these systems, the network of interactions is not easily measured or observed and must instead be inferred from often noisy and incomplete data [41]. Network inference, or network reconstruction, is the task of recovering the hidden interaction structure of a complex system from data.

To date, a broad array of approaches has been developed for the inference of pairwise networks. Sparse regression [32] and compressed sensing [17] methods exploit the structural sparsity of many real-world networks. Bayesian frameworks [25, 29, 36] provide probabilistic reconstructions with uncertainty quantification. Other methods recover functional connectivity by detecting statistical dependencies as quantified by, e.g. correlations [23], mutual information [44], and Granger causality [24, 46]. These methods differ in their data requirements but all share the simplifying assumption that interactions are exclusively dyadic.

However, a growing body of literature suggests that pairwise models are insufficient to capture the structure and dynamics of many real-world systems [2, 3]. In biological, physical, and social systems, a

node may be influenced by two or more other nodes in a nonlinear fashion; such higher-order interactions cannot be decomposed into a linear combination of pairwise couplings and are better represented by hypergraphs or simplicial complexes [3, 6]. Specifically, higher-order structure has been documented in social networks [4, 19, 35], ecological systems [1, 16], and the brain [37, 48]. Furthermore, recent studies have demonstrated that higher-order interactions can profoundly alter collective dynamics such as contagion [19], diffusion [9], and synchronization [38, 39], motivating efforts to model complex systems with hypergraphs and to develop methods for inferring their structure from data [2].

Existing approaches to hypergraph reconstruction fall into two broad categories. Probabilistic inference techniques attempt to infer hyperedges from, e.g. observations of pairwise interactions [26, 47] or contagion dynamics [45], typically under strong assumptions about the generative process. Optimization- and dictionary-learning methods [10, 30], in contrast, aim to recover higher-order interactions directly from time series; these methods typically require some prior knowledge or intuition about network dynamics. More recently, Delabays *et al* [11] proposed Taylor-based Hypergraph Inference using SINDy (THIS), which leverages the sparse identification of nonlinear dynamics [8] (SINDy) framework to reconstruct hypergraphs from multivariate time-series data without requiring prior knowledge of the underlying node dynamics or tailoring feature libraries to specific applications. In this context, SINDy approximates each node's dynamics by a truncated Taylor expansion about a reference point—expressed as a sparse linear combination of monomials in the state variables—so that identifying the active terms provides evidence for pairwise and higher-order interactions.

However, making THIS practical for limited, noisy data requires reliable estimation and selection of Taylor coefficients. In the original formulation, this depends on two thresholding steps: sequential thresholded least-squares (STLS) for sparse coefficient estimation, followed by magnitude-based thresholding to select which inferred hyperedges are retained. The corresponding thresholds play overlapping roles, and their appropriate values depend on the scale of the data, the noise level, and correlations among monomial library terms. As a result, small perturbations in data-limited or ill-conditioned regimes can produce large variability in the inferred coefficients and, consequently, in the reconstructed interaction structure [12, 14, 18]. Motivated by prior work showing that sparse Bayesian regression can improve SINDy's robustness under noise [14, 18, 34, 49], we replace STLS with sparse Bayesian regression using automatic relevance determination (ARD) [42] and develop Bayes-THIS, a Bayesian extension of Taylor-based hypergraph inference (section 2).

This reformulation retains the flexibility of THIS while replacing global hard thresholding with evidence-based, term-wise adaptive shrinkage that explicitly accounts for residual variance. This leads to consistent improvements in challenging regimes with limited data, high measurement noise, and ill-conditioned design matrices (section 3). The resulting posterior over Taylor coefficients also enables uncertainty-aware hyperedge selection: we use posterior intervals to identify supported interactions, substantially reducing sensitivity to post hoc magnitude thresholds (section 4). Beyond coefficient estimation and pruning, we use posterior predictive checks to address a complementary question not answered in the original THIS framework: whether a given dataset contains sufficient higher-order signal distinguishable from measurement noise to reliably support inference beyond a pairwise model (section 5). Finally, we characterize a fundamental limitation of the Taylor-based framework that is independent of the inference method: when higher-order interactions concentrate on node pairs lacking lower-order connections, spurious lower-order interactions are inferred that are structurally indistinguishable from genuine ones (section 6). This suggests that the overlap across interaction orders in a target system is practically relevant prior information for any practitioner considering Taylor-based hypergraph inference.

2. Taylor-based hypergraph inference

We consider a system of n nodes whose dynamics are coupled through a hypergraph. The dynamics of the i th node are given by

$$\dot{x}_i = F_i(\mathbf{x}) = f_i(x_i) + \sum_{j=1}^n a_{ij}^{(2)} g^{(2)}(x_i, x_j) + \sum_{j,k=1}^n a_{ijk}^{(3)} g^{(3)}(x_i, x_j, x_k) + \cdots, \quad (1)$$

where function f_i describes the intrinsic dynamics of node i . The p th order adjacency tensor $A^{(p)} = \{a_{i_1, \dots, i_p}^{(p)}\}$ encodes the topology of p -body interactions, which act via coupling functions $g^{(p)}$. For simplicity, we consider systems with up to third-order interactions, i.e. $A^{(p)} = \mathbf{0}$ for all $p > 3$.

One can write the Taylor expansion of the dynamics F_i in equation (1) about an arbitrary point $\mathbf{x}^*$:

$$\dot{x}_i = F_i(\mathbf{x}^*) + \sum_{j=1}^n \partial_j F_i(\mathbf{x}^*) \Delta x_j + \frac{1}{2} \sum_{j,k=1}^n \partial_{j,k} F_i(\mathbf{x}^*) \Delta x_j \Delta x_k + \dots, \quad (2)$$

where ∂_j denotes the partial derivative with respect to x_j and $\Delta x_j = x_j - x_j^*$ for $j = 1, \dots, n$. Delabays *et al* [11] note that for i, j , and k distinct, if $\partial_{j,k} F_i(\mathbf{x}^*)$ is nonzero, then—assuming there exist only up to third-order interactions in the underlying hypergraph—there is necessarily a triadic interaction between nodes i, j , and k . Thus, they propose reconstructing higher-order interactions by inferring the nonzero coefficients $\partial_{j,k} F_i(\mathbf{x}^*)$ from time-series data $\{\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_N)\}$. The time derivatives $\{\dot{x}_i(t_1), \dot{x}_i(t_2), \dots, \dot{x}_i(t_N)\}$ for each node i may be directly measured as part of the dataset or approximated numerically, e.g. via finite differencing.

2.1. Inference using SINDy

To recover the nonzero Taylor series coefficients, Delabays *et al* [11] employ the popular SINDy framework [8], arguing that many real-world hypergraphs of interest are intrinsically sparse. THIS first defines a library of candidate functions, $\mathcal{L} = \{\psi_1, \dots, \psi_M\}$, here consisting of monomials up to a chosen degree. Given multivariate time series data

$$\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_n] = \begin{bmatrix} x_1(t_1) & x_2(t_1) & \dots & x_n(t_1) \\ x_1(t_2) & x_2(t_2) & \dots & x_n(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_N) & x_2(t_N) & \dots & x_n(t_N) \end{bmatrix}$$

expressed in terms of deviations from the chosen base point $\mathbf{x}^*$ of the Taylor expansion, these library functions are evaluated at each observation to form the design matrix $\Theta \in \mathbb{R}^{N \times M}$ with entries $\Theta_{ij} = \psi_j(\mathbf{x}(t_i))$ so that, for example,

$$\Theta = [\mathbf{1} \ \mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_1^2 \ \mathbf{x}_1 \mathbf{x}_2 \ \mathbf{x}_2^2 \ \dots] \in \mathbb{R}^{N \times M}, \quad (3)$$

where products and powers of the columns of $\mathbf{X}$ are understood elementwise. Assuming that the Taylor expansion of F_i involves only a small number of candidate monomials, the vector ξ_i satisfying

$$\dot{\mathbf{x}}_i = \Theta \xi_i \quad (4)$$

will be sparse. Each nonzero coefficient in ξ_i corresponds to a monomial in library $\mathcal{L}$, which in turn implies a candidate hyperedge pointing towards node i .

SINDy traditionally recovers ξ_i through sparse regression methods such as LASSO [40] or STLS [8]. These techniques typically involve hyperparameters controlling how aggressively sparsity is promoted, and selecting appropriate values for these hyperparameters is nontrivial. A common heuristic is to sweep over the sparsity parameter and identify a Pareto front balancing model complexity and accuracy [8], but this approach can be ambiguous when the Pareto front does not have a sharp elbow. More principled statistical criteria for model selection, such as the Akaike or Bayesian information criterion, perform reliably only when sufficiently abundant, high-quality data are available [31].

In general, scarce and noisy data present a significant challenge for SINDy [12, 33], limiting its effectiveness when applied to real-world datasets. In particular, excessive noise in the data undermines the ability of STLS to reliably select the most appropriate model, in part because STLS does not account for uncertainty in the data—even though this uncertainty is critical to consider when deciding which terms to retain [14].

Because THIS is built on the SINDy framework using STLS, it inherits both of these challenges: the difficulty of tuning sparsity-promoting hyperparameters and a lack of robustness to noisy or limited data. Prior work [13, 14, 34, 49] has demonstrated that adopting a Bayesian perspective—specifically, integrating sparse Bayesian regression with SINDy—can substantially improve robustness by treating model selection and uncertainty quantification within a unified probabilistic framework. This motivates a Bayesian reformulation of equation (4).

2.2. Inference using sparse Bayesian regression with ARD

In practice, two sources of error arise in the reconstruction of coefficient vector ξ_i . First, the Taylor expansion of F_i is necessarily truncated to a finite number of terms, introducing an approximation error that depends on the degree of truncation and the region over which the data are sampled. Unlike measurement noise, this truncation error is a deterministic function of the data; it is therefore not addressed by the Bayesian framework developed here and not well modeled by the Gaussian noise assumption we discuss below. In practice, truncation error is controlled by the choice of sampling region; data sampled sufficiently close to the expansion point will keep higher-order terms small. Note that sampling too close to the expansion point introduces a separate difficulty: the triadic signal becomes indistinguishable from noise. Navigating this tradeoff is the subject of section 5.

Second, measurement error can enter the derivative data either directly or through numerical differentiation of noisy state observations; this is the dominant source of error when the sampling region is appropriately chosen. If time derivatives are directly observed, we assume additive, zero-mean Gaussian measurement noise with variance σ^2 , independent and identically distributed (i.i.d.) across nodes and observations. If instead time derivatives are approximated via finite differencing from trajectory data with the same noise structure—that is, i.i.d. Gaussian noise—then the variance of the resulting derivative noise scales as $1/(\Delta t)^2$, where Δt is the timestep in regularly sampled time-series data, and short-range temporal correlations are introduced across time points. In both cases, treating the derivative noise as i.i.d. Gaussian noise provides an analytically tractable approximation that is exact in the former setting and reasonable in the latter.

Accordingly, we augment the deterministic model in equation (4) with additive Gaussian noise:

$$\dot{\mathbf{x}}_i = \Theta \xi_i + \varepsilon_i, \quad (5)$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2 \mathbf{I}_N)$ is an N -vector of i.i.d. Gaussian noise with variance σ_i^2 , which may be specified beforehand—if known—but is more commonly estimated from the data. For simplicity, we will focus on one node and drop the subscript i in the following discussion.

This modeling of the measurement noise implies a multivariate Gaussian likelihood:

$$p(\dot{\mathbf{x}} | \xi, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{\|\dot{\mathbf{x}} - \Theta \xi\|^2}{2\sigma^2} \right\}.$$

The key to sparse Bayesian regression with ARD is then the definition of a sparsity-promoting, hyperparameterized prior over coefficient vector ξ of the form

$$p(\xi | \alpha) = \prod_{m=1}^M \left[(2\pi)^{-1/2} \alpha_m^{1/2} \exp \left\{ -\frac{\alpha_m \xi_m^2}{2} \right\} \right], \quad (6)$$

where each hyperparameter α_m controls the strength (i.e. precision) of the prior over its associated coefficient ξ_m [42, 43].

Rather than performing full Bayesian inference over these hyperparameters, we follow Tipping and Faul [43] and estimate the precision parameters by a type-II maximum likelihood procedure [5]. For fixed σ^2 , this involves (local) maximization with respect to α of the log-evidence [42, 43]:

$$\begin{aligned} \ln p(\dot{\mathbf{x}} | \alpha, \sigma^2) &= \ln \int p(\dot{\mathbf{x}} | \xi, \sigma^2) p(\xi | \alpha) d\xi, \\ &= -\frac{1}{2} [N \ln 2\pi + \ln |\mathbf{C}| + \dot{\mathbf{x}}^\top \mathbf{C}^{-1} \dot{\mathbf{x}}], \end{aligned} \quad (7)$$

where

$$\mathbf{C} = \sigma^2 \mathbf{I}_N + \Theta \mathbf{A}^{-1} \Theta^\top, \quad \mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_M).$$

Tipping and Faul [43] decompose $\mathbf{C}$ to isolate the contribution of each α_m (and the inclusion of the corresponding library function ψ_m in the model) to the log-evidence in equation (7). Their iterative algorithm can therefore ‘add’ or ‘delete’ individual terms from the model in a principled way—i.e. by comparing models by their evidence. Deleting the m th library function corresponds to taking $\alpha_m \rightarrow \infty$, in which case the prior $p(\xi_m | \alpha_m)$ is effectively a ‘spike’ at zero, while terms that are included in the model will have finite α_m , effectively a ‘slab’ prior. ARD thus approximates a spike-and-slab prior—typically considered the gold-standard sparsity-inducing prior—without introducing explicit discrete inclusion variables [14]. This avoids direct sampling over model structures (as in UQ-SINDy [18]),

which becomes prohibitively expensive for monomial libraries whose dimension grows combinatorially with n and the maximum interaction order. The ARD framework instead retains analytical tractability by exploiting the conjugacy property of Gaussian distributions [42, 43].

This evidence-maximization procedure admits a hierarchical interpretation by assigning independent Gamma hyperpriors to the coefficient precisions:

$$p(\alpha) \propto \prod_{m=1}^M \alpha_m^{a-1} e^{-b\alpha_m}, \quad (8)$$

where a and b are the shape and rate parameters, respectively [42]. For $a, b > 0$, integrating the Gaussian prior of equation (6) against equation (8) yields a Student- t marginal prior on each coefficient ξ_m , whose concentration near zero and heavy tails favor sparse coefficient vectors while allowing well-supported coefficients to remain large [42]. The Tipping–Faul procedure described above corresponds to the formal limit $a = b = 0$ under which $p(\alpha_m) \propto \alpha_m^{-1}$ is improper but flat in $\log \alpha_m$ so that maximizing the evidence over α is exactly maximum *a posteriori* (MAP) estimation of $\log \alpha$. Moreover, this limit is also what makes the spike-and-slab mechanism possible: for any $b > 0$, the Gamma hyperprior contributes a penalty to the hyperparameter log-posterior that diverges as $\alpha_m \rightarrow \infty$, foreclosing exact deletion. We emphasize, however, that the hierarchy is interpretive rather than operational in the procedure used here: α is replaced by a point estimate $\hat{\alpha}$ rather than integrated over.

If σ^2 is not known, the Tipping–Faul procedure also re-estimates the value of $\beta \equiv \sigma^{-2}$ between iterations to maximize the log-evidence [43]. Conditional on the type-II maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$, the coefficient posterior $p(\xi | \mathbf{x}, \hat{\alpha}, \hat{\beta})$ is multivariate Gaussian with mean

$$\mu = \hat{\beta} \Sigma \Theta^\top \mathbf{x} \quad (9)$$

and covariance

$$\Sigma = \left(\hat{\mathbf{A}} + \hat{\beta} \Theta^\top \Theta \right)^{-1}, \quad \hat{\mathbf{A}} = \text{diag}(\hat{\alpha}_1, \dots, \hat{\alpha}_M). \quad (10)$$

Coefficients associated with deleted library functions are assigned a point mass at zero [42].

We refer to the resulting Bayesian extension of Taylor-based hypergraph inference as Bayes-THIS. In practice, Bayes-THIS proceeds analogously to THIS: we construct the design matrix Θ and, for each node i , infer the coefficient vector ξ_i independently under the assumption that noise is independent across nodes. Unlike STLS's global hard thresholding, however, model selection is performed by maximizing the log-evidence with respect to hyperparameters α_i and β_i . We use the posterior mean μ_i (equivalently, the MAP estimate) as the point estimate for ξ_i while retaining posterior covariance Σ_i for uncertainty quantification. Bayes-THIS is therefore an empirical-Bayes procedure: the Gaussian coefficient posterior is conditional on the type-II maximum likelihood estimates $\hat{\alpha}_i$ and $\hat{\beta}_i$. It quantifies coefficient uncertainty conditional on these fitted quantities but does not propagate uncertainty in the hyperparameters or in the active-set selection itself. See the Supplementary Material (SM) for more details.

3. Robustness and data efficiency

We compare the performance of THIS and Bayes-THIS using a generalization of the Kuramoto model with both pairwise and triadic interactions:

$$\dot{\theta}_i = \omega + \sum_{j=1}^n a_{ij}^{(2)} [\sin(\theta_j - \theta_i + \varphi_2) - \sin \varphi_2] + \sum_{j,k=1}^n a_{ijk}^{(3)} [\sin(\theta_j + \theta_k - 2\theta_i + \varphi_3) - \sin \varphi_3], \quad (11)$$

where $\theta_i \in S^1$ represents the phase of oscillator i and ω its natural frequency. After transforming to a rotating reference frame with frequency ω , the system admits as an equilibrium any constant phase vector; without loss of generality, we consider motion in the vicinity of the origin. The phase lags φ_2 and φ_3 act as frustration parameters, shifting the effective coupling. For the connected, undirected hypergraphs considered here, the constant-phase manifold is locally asymptotically stable when $\cos \varphi_2, \cos \varphi_3 > 0$, so trajectories initialized near the origin remain close, making it a suitable base point for Taylor expansion. In our inference experiments, we set $\varphi_2 = \varphi_3 = \pi/4$, which guarantees stability and prevents parity-induced cancellations in the Taylor expansion at the constant-phase equilibrium.

In the unfrustrated case, the sine coupling is odd about the equilibrium, so genuine triadic interactions have vanishing second-order Taylor coefficients and would instead need to be detected through higher-order derivatives.

Both THIS and Bayes-THIS infer a collection of coefficient vectors $\Xi = [\xi_1 \dots \xi_n] \in \mathbb{R}^{M \times n}$ corresponding to monomial terms in the Taylor expansion of the dynamics. To assess inference quality, we follow Delabays *et al* [11] and introduce a threshold τ on coefficient magnitude: a hyperedge is retained if and only if the magnitude of its corresponding monomial's coefficient exceeds τ . Section 4 examines the practical limitations of magnitude-based thresholding and introduces a more principled alternative based on the conditional coefficient posterior. As τ decreases from ∞ to 0, the inferred structure interpolates between the empty hypergraph and the complete set of candidate hyperedges, and plotting the true positive rate (TPR) against the false positive rate across this range yields the receiver operating characteristic curve. We summarize overall inference quality—across pairwise and triadic interactions—by the area under this curve (AUROC), where 1 indicates perfect discrimination between true and nonexistent hyperedges and 0.5 indicates performance equivalent to random guessing.

For triadic interactions, we supplement AUROC with the area under the precision-recall curve (AUPRC), which plots precision—the fraction of inferred triadic interactions that are genuine—against recall (equivalently, TPR). This metric emphasizes performance at low false-positive rates and is thus more informative than AUROC in sparse settings, i.e. as the number of candidate interactions grows combinatorially with n and the inference problem is highly class-imbalanced. In these robustness benchmarks, we do not report AUPRC for pairwise interactions because, as noted by Delabays *et al* triadic interactions generally contribute to lower-order Taylor coefficients: a genuine triadic interaction between nodes i, j , and k will typically produce nonzero coefficients on the monomials x_j and x_k in the Taylor expansion for $\dot{x}_i$ even when no pairwise interactions exist. This is a structural false positive—a direct consequence of the Taylor expansion rather than an artifact of finite data or model misspecification—that we discuss more in section 6.

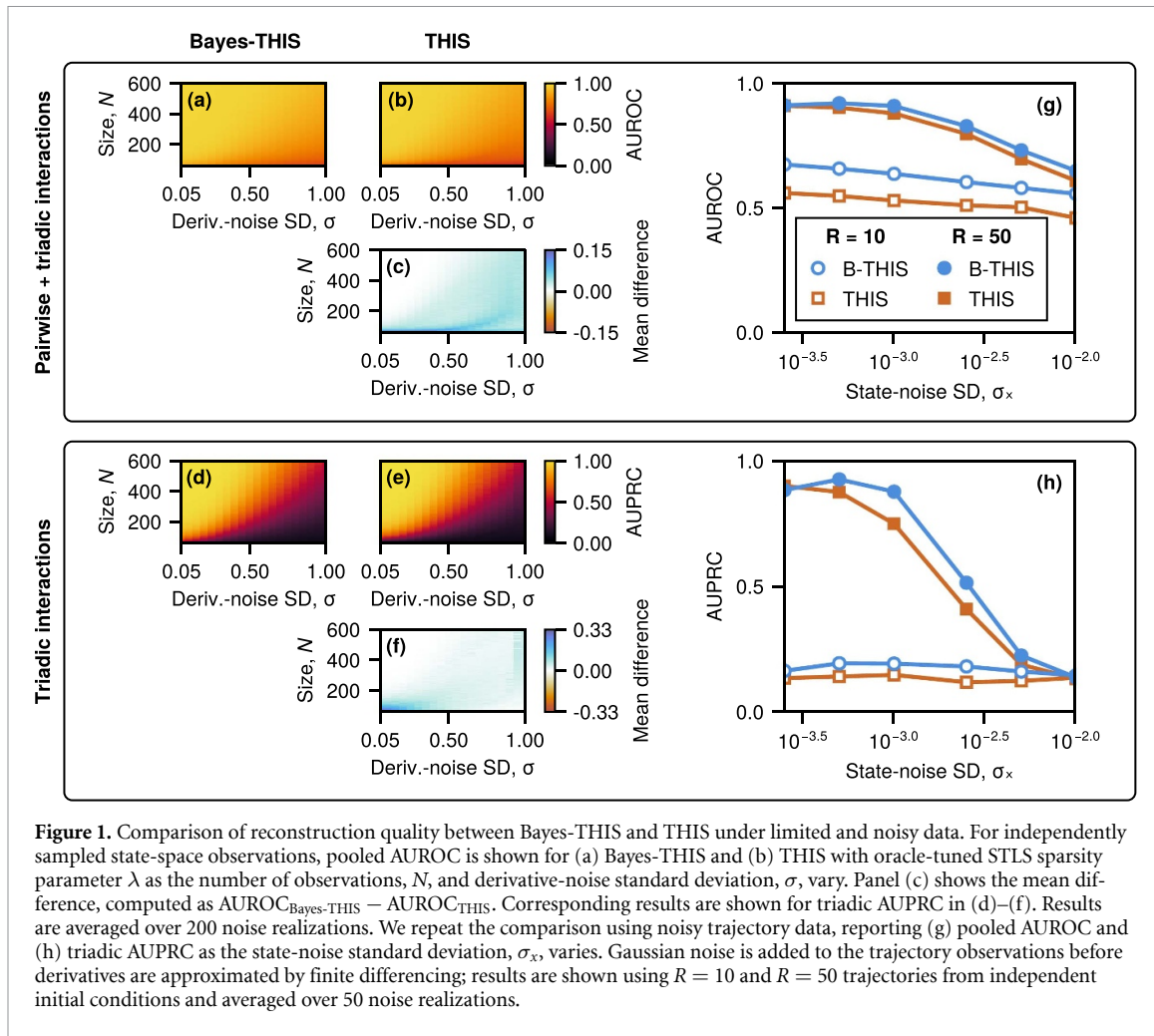
3.1. Data availability and measurement noise

We first compare the performance of THIS and Bayes-THIS as a function of data availability and measurement noise (figure 1). We fix a ten-node ($n = 10$) Erdős–Rényi (ER) hypergraph [6] with pairwise and triadic edge probabilities $\rho_2 = 0.35$ and $\rho_3 = 0.06$, respectively, yielding average pairwise degree $\langle k^{(2)} \rangle = 3.15$ and average triadic degree $\langle k^{(3)} \rangle = 2.16$. For each experiment, we draw N samples independently and uniformly at random from a ten-dimensional hypercube of side length 1.0 centered at the origin to form datasets $\mathbf{X}$ and then compute derivatives $\dot{\mathbf{X}}$ directly from equation (11). We corrupt $\dot{\mathbf{X}}$ with additive, zero-mean Gaussian noise with standard deviation σ , independent across nodes and observations—the same noise model as in section 2.2.

Figure 1 compares Bayes-THIS with oracle-tuned THIS as N and σ vary. For each (σ, N) setting, we select STLS sparsity parameter λ to maximize the mean pooled AUROC across realizations; the top block of figure 1 reports this average pooled AUROC for both methods. While the two methods perform similarly when data are abundant and noise is low, Bayes-THIS consistently outperforms THIS when data are limited and measurement noise increases (figure 1(c)).

This advantage persists when we focus specifically on triadic reconstruction: for small N and small to moderate σ , Bayes-THIS consistently achieves higher AUPRC (figure 1(f)). STLS is poorly suited to this sparse regime where most candidate triadic interactions are absent: with limited data, coefficient estimates are noisy, and a fixed threshold λ cannot simultaneously suppress spurious terms and retain genuine ones across all library monomials. Term-wise shrinkage, in contrast, concentrates regularization where the data provide little support, reducing false positives without sacrificing recall. At high σ , however, the triadic signal is too weak for reliable recovery by either method, and the advantage of Bayes-THIS diminishes.

The preceding experiment isolates coefficient estimation under a noise model that closely matches the assumptions described in section 2.2. In practice, however, inference is often performed with noisy trajectory data so that measurement error affects both the design matrix Θ and derivatives estimated by numerical differentiation. Figures 1(g) and (h) therefore show results from a benchmark using noisy trajectory data. We generate 50 trajectories by integrating equation (11) over $t \in [0, 0.6]$ with sampling interval $\Delta t = 0.01$ on a fixed, eight-node ER hypergraph with pairwise and triadic edge probabilities $\rho_2 = 0.45$ and $\rho_3 = 0.10$, respectively. Independent Gaussian noise with standard deviation σ_x is added to the state observations before derivatives are approximated by finite differencing. We perform inference using a subset of $R = 10$ or all $R = 50$ trajectories under different noise realizations to examine the effect of data availability. As above, results for THIS are shown for oracle-tuned λ .



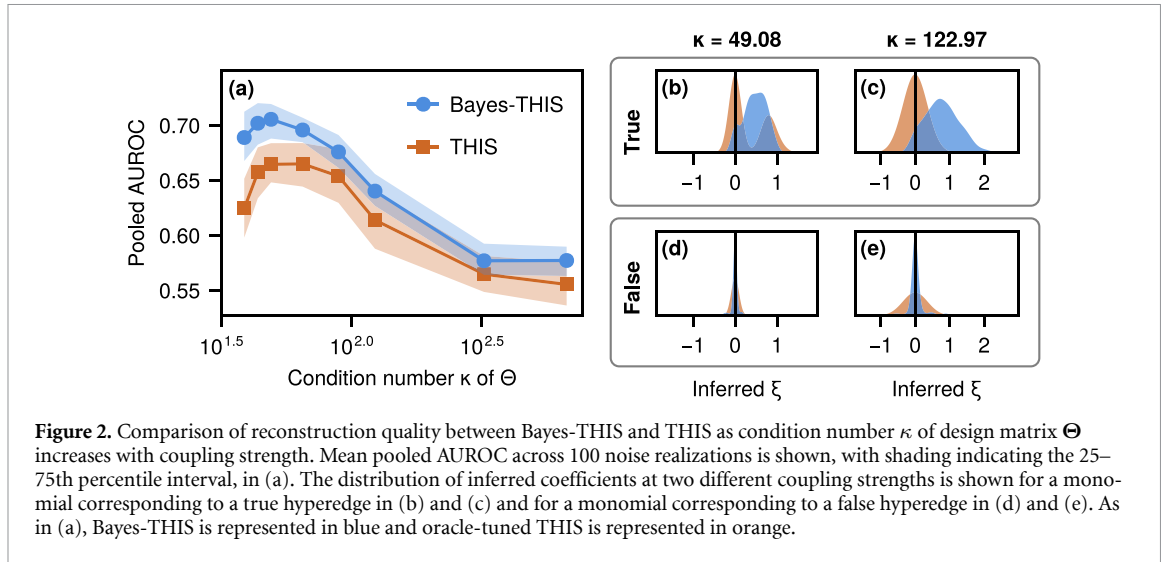
The trajectory benchmark is more challenging for both methods: reconstruction quality decreases significantly as state noise σ_x increases, with triadic inference particularly sensitive (figure 1(h)). Even so, Bayes-THIS matches or outperforms THIS over the range considered with the largest gains again occurring when fewer trajectories are available and at intermediate levels of noise. These results show that the advantages of sparse Bayesian regression are not confined to situations closely matching the assumptions described in section 2.2 but persist when inference is performed from noisy trajectories with numerically estimated derivatives.

Additional experiments in the SM show that the advantages of Bayes-THIS persist as system size increases and that its performance remains comparable to THIS across a broad range of interaction densities. Bayes-THIS is substantially more expensive per fit, though the SM comparison times THIS at a single oracle-tuned λ rather than the sweep over candidate values that practical tuning requires.

3.2. Ill-conditioning from restricted state-space sampling

As the coupling strengths in equation (11) increase, trajectories starting at the same initial conditions converge more rapidly toward the synchronized constant-phase manifold. Consequently, observations concentrate in an increasingly narrow, nearly one-dimensional region of state space so that monomials in the library—when evaluated on the data—are increasingly correlated. The resulting near-linear dependence among columns in the design matrix Θ makes the regression problem ill-conditioned. Thus, we next compare the performance of THIS and Bayes-THIS as this ill-conditioning becomes more severe.

We use a fixed ER hypergraph with $n = 10$, $\rho_2 = 0.35$, and $\rho_3 = 0.06$ but vary the pairwise and triadic coupling coefficients in equation (11) by a common multiplicative factor while keeping the initial conditions, observation window $t \in [0, 0.7]$, and timestep $\Delta t = 0.1$ fixed. Stronger coupling drives faster convergence toward synchronization. Derivatives along the sampled trajectories are computed analytically, and we add Gaussian noise with standard deviation $\sigma = 0.4$, which represents a growing noise-to-signal ratio as coupling strength increases and derivative magnitudes shrink.



Across increasing coupling strengths—and increasingly ill-conditioned design matrices—Bayes-THIS consistently outperforms oracle-tuned THIS, where λ is chosen independently for each coupling strength to maximize mean pooled AUROC across noise realizations (figure 2(a)). Ill-conditioning amplifies the sensitivity of coefficient estimates to noise in $\tilde{\mathbf{X}}$, and STLS compounds this by applying a hard, global threshold at each iteration. In contrast, sparse Bayesian regression with ARD regularizes coefficients in a term-wise, data-adaptive manner and explicitly models residual variance, generally reducing the variance of the inferred coefficients (figures 2(b)–(e)).

In particular, noise-induced fluctuations in the STLS coefficient estimates can drive monomial coefficients corresponding to true hyperedges below threshold λ (figure 2(b)) and those corresponding to nonexistent interactions above the threshold (figure 2(e)). In contrast, the distribution of coefficients inferred by sparse Bayesian regression remains peaked away from zero for true interactions (figures 2(b) and (c)) and at zero for false interactions (figures 2(d) and (e)). Even so, Bayes-THIS is not immune to the effects of more severe ill-conditioning: as coupling strength increases and columns of Θ become more correlated, the inferred coefficient distributions broaden (figures 2(c) and (e)). Overall, however, the stability of coefficient estimates from sparse Bayesian regression with ARD translates directly into improved reconstruction quality.

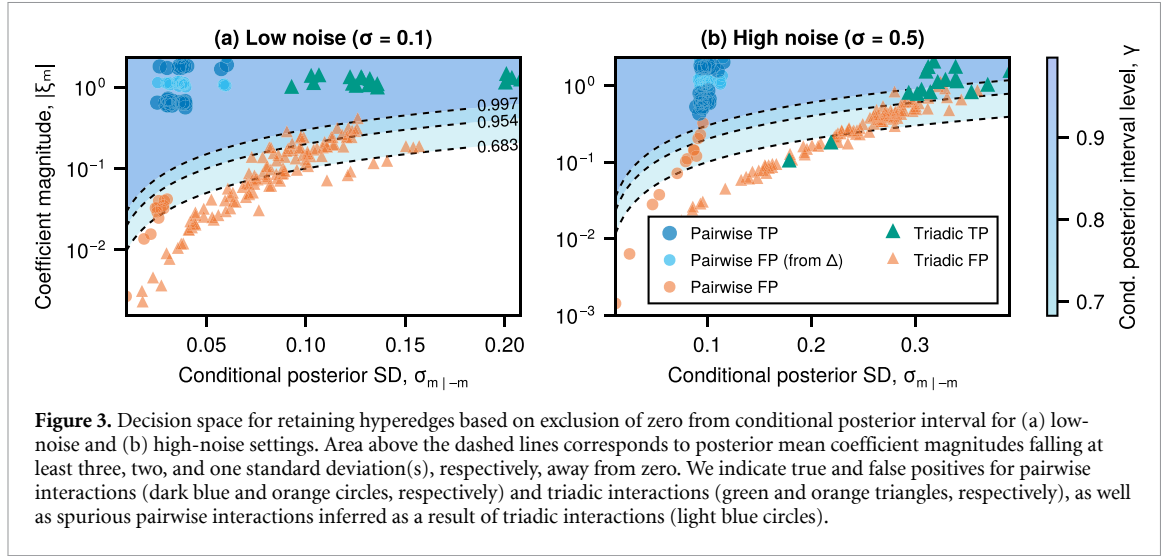
Taken together with the experiments in section 3.1, we see that Bayes-THIS offers meaningful improvements over THIS precisely in the regimes most relevant for empirical reconstruction: limited observations, high measurement noise, and ill-conditioned design matrices induced by trajectories that explore only a restricted region of state space. In each case, the improvement arises from a single mechanism—sparse Bayesian regression models residual variance and applies adaptive, term-wise shrinkage rather than a fixed global threshold—that yields more robust inference without manual tuning.

4. Uncertainty-aware pruning of weakly supported hyperedges

In addition to the STLS sparsity parameter λ , Delabays *et al* [11] introduce a thresholding parameter τ as a post-processing step in THIS in order to discard hyperedges whose inferred monomial coefficients have small magnitude. This introduces a related but new hyperparameter to tune: as with λ , no single threshold reliably separates true and spurious interactions across regimes because coefficient magnitudes depend on the scale of the data, the level of measurement noise, and correlations among library terms.

The Bayesian formulation of Bayes-THIS provides a natural alternative. Rather than filtering coefficients based solely on magnitude, we use uncertainty in the fitted coefficient posterior to assess whether each inferred coefficient is sufficiently separated from zero to retain its corresponding hyperedge. This yields an uncertainty-aware pruning rule whose effective threshold adapts to the inferred noise level and the scale of the corresponding design-matrix column.

For each node, after estimating hyperparameters $\hat{\alpha}$ (and possibly $\hat{\beta}$) and obtaining the Gaussian posterior $p(\xi | \tilde{\mathbf{x}}, \hat{\alpha}, \hat{\beta})$, we assess the relevance of the m th coefficient ξ_m using its conditional posterior distribution with the remaining coefficients fixed at their posterior means. We use the conditional rather than marginal posterior variance because it better isolates the uncertainty relevant to deciding whether



an individual coefficient ξ_m is supported within the fitted model: holding the remaining fitted coefficients fixed removes uncertainty arising from their joint variation with ξ_m .

Writing the posterior covariance—restricted to the ARD-selected active set, that is, those coefficients with α_m finite—in block form

$$\Sigma = \begin{bmatrix} \Sigma_{m,m} & \Sigma_{m,-m} \\ \Sigma_{-m,m} & \Sigma_{-m,-m} \end{bmatrix},$$

where $-m$ denotes all indices but the m th, the conditional distribution

$$p(\xi_m | \xi_{-m} = \mu_{-m}, \hat{\mathbf{x}}, \hat{\alpha}, \hat{\beta})$$

is Gaussian with mean μ_m and variance

$$\sigma_{m|-m}^2 = \Sigma_{m,m} - \Sigma_{m,-m} \Sigma_{-m,-m}^{-1} \Sigma_{-m,m} = \frac{1}{\hat{\alpha}_m + \hat{\beta} \|\Theta_{:,m}\|_2^2}. \quad (12)$$

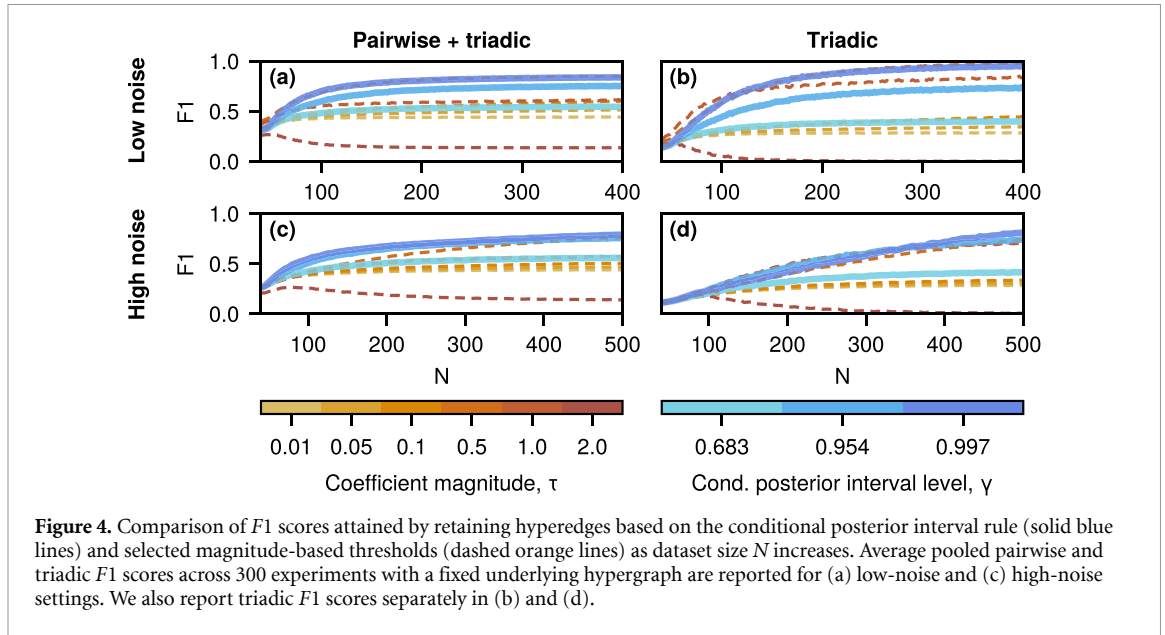
We retain the m th hyperedge if zero lies outside the equal-tailed γ -interval

$$\mu_m \pm z_{(1+\gamma)/2} \sigma_{m|-m},$$

where $z_{(1+\gamma)/2}$ denotes the $(1+\gamma)/2$ quantile of the standard normal distribution. This converts hyperedge selection from a magnitude heuristic into an uncertainty-aware decision rule for whether a coefficient is distinguishable from zero. Note that these intervals are conditional empirical-Bayes summaries because $\hat{\alpha}$, $\hat{\beta}$ and the ARD-selected active set are treated as fixed, as are the remaining coefficients when evaluating ξ_m . They therefore do not represent uncertainty propagated over hyperparameters or model structure, and the nominal interval level γ need not correspond to empirical coverage under repeated sampling. In the SM, we examine this calibration and compare conditional and marginal posterior intervals directly as pruning criteria; we find that conditional posterior intervals yield a more favorable precision-recall tradeoff, particularly in scarce or noisy regimes.

To illustrate how this pruning rule operates in practice, figure 3 shows the decision boundaries for a ten-node hypergraph with the same settings used to generate figure 1, here inferring coefficients from one dataset $\mathbf{X}$ of $N = 300$ points and with analytically-computed derivatives $\hat{\mathbf{X}}$ corrupted by Gaussian noise with standard deviation $\sigma = 0.1$ or $\sigma = 0.5$. Setting $\gamma = 0.683$, 0.954 , and 0.997 corresponds to retaining hyperedges whose corresponding monomial coefficients have posterior means at least one, two, and three conditional posterior standard deviations from zero, respectively. Thus, coefficients with greater posterior uncertainty must have larger magnitude to survive pruning [see equation (12)]. Note that the pairwise false positives that remain are almost entirely attributable to triadic interactions: as discussed in section 3, higher-order interactions generically contribute to lower-order Taylor coefficients, producing spurious pairwise terms that are indistinguishable from genuine pairwise interactions within the Taylor-based framework (see also section 6).

This uncertainty-adaptive threshold also makes the pruning rule less sensitive to the particular nominal interval level γ than a fixed coefficient-magnitude cutoff. Figure 4 tests this by comparing



magnitude-based thresholding and conditional posterior interval pruning in terms of $F1$ score—the harmonic mean of precision and recall—as a function of N at the same fixed noise levels ($\sigma = 0.1$ and $\sigma = 0.5$). The hypergraph settings and data generating process are the same as in figure 3. Posterior interval filtering (solid blue curves) consistently outperforms magnitude thresholding (dashed orange curves) across both noise levels except for some carefully tuned values of τ . Importantly, the posterior interval curves are more tightly clustered across values of γ , indicating that performance is more robust to the specific nominal interval level chosen. Magnitude thresholding, by contrast, is highly sensitive to the choice of τ : choosing τ too conservatively or too aggressively leads to poor performance even when data are abundant.

Together, these results suggest that conditional posterior interval pruning offers a principled and practically robust alternative to magnitude-based thresholding: with γ chosen sufficiently large—i.e. corresponding to two or three posterior standard deviations—it consistently achieves a high $F1$ score across noise levels and sample sizes and requires no manual tuning of a sensitive threshold parameter. In principle, an overly aggressive γ would suppress genuine interactions whose coefficients are small relative to posterior uncertainty—analogueous to choosing τ too aggressively in magnitude thresholding—but the range of well-performing γ values is fairly wide in these experiments, as the posterior uncertainty naturally adapts to the noise level.

5. Assessing data quality for higher-order inference

Finally, our Bayesian framework affords a principled approach to assessing data quality *before* inference. As noted by Delabays *et al* [11], Taylor-based hypergraph inference requires the sampling region to be appropriately scaled relative to the noise level: if the data are sampled too close to the Taylor expansion point, any triadic signal will be indistinguishable from noise. If sampled too far away, the Taylor approximation breaks down. In either case, proceeding with inference risks drawing conclusions from data that cannot reliably support them. We can use posterior predictive checks (PPCs) [15] to detect the former failure mode—that is, to assess whether the observed data contain sufficient triadic signal to warrant inference beyond a pairwise model.

The key idea is to ask whether the residuals from a pairwise model contain more structure in triadic directions than would be expected under the assumed measurement noise model. If the data are sampled from too small a region relative to the noise level, the pairwise null model will adequately explain the data and the residuals will be indistinguishable from noise, even when triadic interactions are present. Formally, suppose we partition the design matrix Θ as

$$\Theta = [\Theta_1 \mid \Theta_2],$$

where Θ_1 and Θ_2 contain the columns compatible with a pairwise model and genuine triadic directions, respectively. Then equation (5) becomes, dropping subscript i ,

$$\dot{\mathbf{x}} = \Theta_1 \xi_1 + \Theta_2 \xi_2 + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N).$$

Here we assume that $\beta^{-1} = \sigma^2$ is known—a reasonable assumption in many experimental settings where the measurement noise level can be estimated independently of the system dynamics. We fit a pairwise model using sparse Bayesian regression, obtaining conditional posterior $p(\xi_1 | \dot{\mathbf{x}}, \hat{\alpha}, \beta)$ with mean μ_1 and covariance Σ_1 .

For each posterior draw $\xi_1^{(s)} \sim p(\xi_1 | \dot{\mathbf{x}}, \hat{\alpha}, \beta)$, $s = 1, \dots, S$, we compute the observed residual:

$$\hat{\varepsilon}^{(s)} = \dot{\mathbf{x}} - \Theta_1 \xi_1^{(s)}. \quad (13)$$

To isolate triadic structure that is genuinely distinguishable from pairwise directions, let $\mathbf{M}_1$ denote the orthogonal projector onto the complement of the column space of Θ_1 and define

$$\tilde{\Theta}_2 = \mathbf{M}_1 \Theta_2.$$

We then let $\mathbf{P}_2$ denote the orthogonal projector onto the column space of $\tilde{\Theta}_2$. Our discrepancy measure is then

$$T^{(s)} = \hat{\varepsilon}^{(s)\top} \mathbf{P}_2 \hat{\varepsilon}^{(s)},$$

which measures the projected residual sum of squares in triadic directions orthogonal to the pairwise subspace—that is, the portion of the residual lying in triadic directions that cannot be represented by the pairwise model.

For posterior draw s , a replicate dataset is generated under the pairwise model as

$$\dot{\mathbf{x}}_{\text{rep}}^{(s)} = \Theta_1 \xi_1^{(s)} + \varepsilon_{\text{rep}}^{(s)}, \quad \varepsilon_{\text{rep}}^{(s)} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N),$$

so that the corresponding residual $\dot{\mathbf{x}}_{\text{rep}}^{(s)} - \Theta_1 \xi_1^{(s)}$ is exactly $\varepsilon_{\text{rep}}^{(s)}$. The replicated discrepancy measure is then

$$T_{\text{rep}}^{(s)} = \varepsilon_{\text{rep}}^{(s)\top} \mathbf{P}_2 \varepsilon_{\text{rep}}^{(s)}.$$

The posterior predictive p -value [15] is estimated as

$$p = \frac{1}{S} \sum_{s=1}^S \mathbf{1}[T_{\text{rep}}^{(s)} \geq T^{(s)}].$$

Note that because the same orthogonal projector $\mathbf{P}_2$ is applied to both the observed and replicated residuals, strong correlations among the orthogonalized triadic columns leave the posterior predictive comparison valid; near-collinearity instead reduces the number of distinguishable triadic directions and therefore the power of the check.

A large posterior predictive p -value indicates that observed residuals are consistent with the replicated residuals under the pairwise null model, suggesting that the data do not reliably support triadic inference. A p -value near 0, by contrast, suggests that the observed residuals are systematically more structured in the triadic directions than expected under noise alone, providing evidence that the data are informative about triadic interactions. Note that for joint inference across all nodes $i = 1, \dots, n$, we can sum the discrepancy measures for each i before computing the p -value because noise is assumed independent across nodes; this aggregates evidence across nodes and assesses whether the dataset as a whole supports inference beyond a pairwise model.

To demonstrate the value of such a heuristic, we study triadic reconstruction quality as a function of sampling region size and noise. Specifically, we fix the underlying ten-node hypergraph with $\rho_2 = 0.35$ and $\rho_3 = 0.06$ and sample $N = 300$ data points independently and uniformly at random from a ten-dimensional hypercube of side length 1.0 centered at the origin. By scaling the data down uniformly, we effectively shrink the sampling box size; derivatives $\dot{\mathbf{X}}$ are then computed analytically and corrupted with additive Gaussian noise with standard deviation $\sigma = 0.05$ (low-noise setting) or $\sigma = 0.2$ (high-noise setting). Figures 5(a) and (b) plot triadic $F1$ score—after conditional posterior interval filtering as described in section 4—as a function of sampling box size with each point representing a different noise realization and colored by posterior predictive p -value.

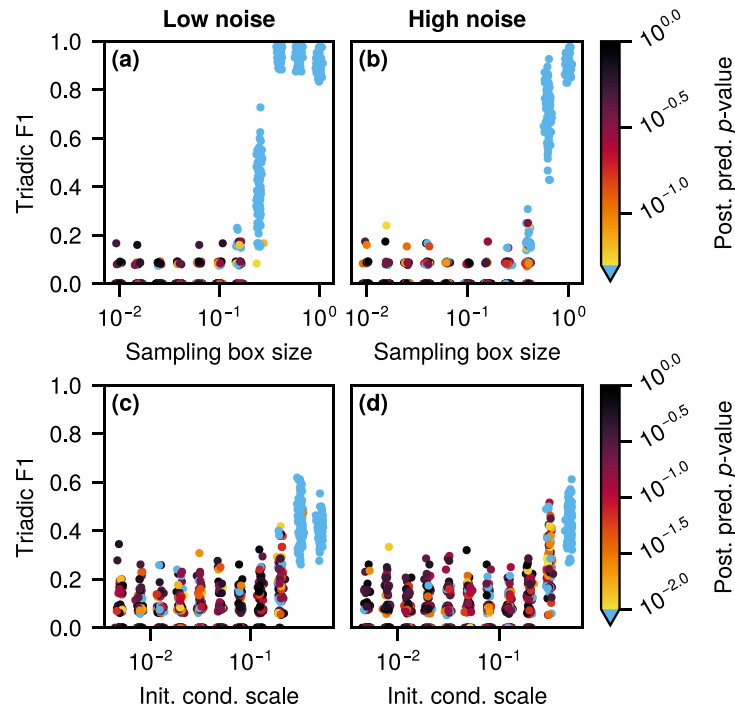


Figure 5. Posterior predictive p -value and triadic inference quality as sampling region grows for fixed underlying hypergraphs. We report triadic $F1$ score across different noise realizations after filtering with a 99.7% conditional posterior interval for low-noise and high-noise settings (left and right columns, respectively). In (a) and (b), derivatives are computed analytically at data points sampled from a hypercube centered at the origin and then corrupted with i.i.d. zero-mean Gaussian noise. In (c) and (d), derivatives are approximated via finite differencing from noisy trajectory data sampled from 30 initial conditions.

We see that with small sampling boxes, the data are sampled too close to the expansion point for triadic structure to be distinguishable from noise; reconstruction quality is poor, and p -values are generally large, correctly indicating that the data do not support inference beyond a pairwise model. As the box size increases, reconstruction quality improves and p -values decrease, as expected. This transition occurs at larger sampling box sizes under high noise, reflecting the need to sample from a wider region of the state space to overcome the increased difficulty of distinguishing triadic signal from noise. Note that while a small p -value provides evidence that the data contain detectable structure beyond the pairwise model, it is not sufficient to guarantee accurate reconstruction. The PPC should therefore be interpreted as a filter that identifies when it is unsafe to proceed with triadic inference rather than as a guarantee of high-quality triadic inference.

To assess whether the PPC remains informative when derivatives are estimated from data rather than computed analytically—the more realistic setting in practice—we repeat the experiment using noisy trajectory data with derivatives approximated via finite differencing (figures 5(c) and (d)). We observe trajectories for an eight-node hypergraph with pairwise edge probability $\rho_2 = 0.45$ and triadic edge probability $\rho_3 = 0.1$ from 30 different initial conditions. To mimic varying sampling box sizes, trajectories are initialized along the same directions in state space, but initial conditions are scaled down. We add i.i.d. zero-mean Gaussian noise with standard deviation $\sigma_x = 5 \times 10^{-4}$ (low-noise setting) or $\sigma_x = 10^{-3}$ (high-noise setting) to the trajectory observations before finite differencing. Thus, measurement noise affects both the design matrix and the estimated derivatives. Since finite differencing is a linear operation, the noise in estimated derivatives has covariance $\sigma_x^2 \mathbf{L}_{\partial t} \mathbf{L}_{\partial t}^T$, where $\mathbf{L}_{\partial t}$ is the matrix representing the finite differencing operation. We approximate this covariance matrix by its diagonal, treating the noise as independent across time points. The results show the same qualitative trend as before: small p -values concentrate at larger initial condition scales where triadic reconstruction quality is higher, and large p -values correctly flag settings where inference is unreliable (figures 5(c) and (d)). The overall reduction in reconstruction quality reflects the challenges of inference using trajectory data near synchronization—since less of the state space is explored—and derivative estimates from noisy data.

It is worth noting that this accounts for only one sampling failure mode—sampling too close—and that the other failure mode—sampling too far, when truncation error in the Taylor expansion grows and overwhelms triadic signal—cannot be detected. However, these results establish the posterior predictive p -value as a practical, preliminary diagnostic for assessing whether a given dataset can reliably support

hypergraph inference beyond a pairwise model. A large p -value reliably identifies settings where the sampling region is too small relative to the noise level for triadic structure to be detectable. A small p -value, on the other hand, indicates detectable residual structure beyond the pairwise model but does not guarantee accurate reconstruction, which ultimately depends on data availability and noise level. Importantly, the diagnostic remains informative when derivatives are estimated from trajectory data, suggesting it is applicable in realistic experimental settings where analytic derivatives are unavailable.

To further assess how coefficient uncertainty and measurement noise affect an empirical reconstruction, we revisit in the SM the resting-state EEG data analyzed by Delabays *et al* [11]. Bayes-THIS recovers a substantial higher-order contribution that is robust to conditional posterior interval pruning, while an adaptation of the projection-based diagnostic developed here quantifies the measurement-noise amplitude required to account for higher-order dynamics not present in a pairwise model.

6. Hypergraph structure & inferability

In sections 3 through 5, we examined how Bayes-THIS improves upon THIS in challenging data regimes—first offering a heuristic for identifying when data quality is insufficient for higher-order inference and then improving robustness in noisy, data-limited settings—and enables a principled, uncertainty-aware approach to pruning weakly supported hyperedges. The proposed inference workflow is outlined in full in the appendix. In this section, we consider a structural property of hypergraphs—namely, the extent to which higher-order interactions are supported by lower-order edges—that can make the inference problem itself harder, independent of data quality or inference method. Specifically, we show that when many of the node pairs contained in triadic interactions lack corresponding pairwise edges, spurious pairwise edges are structurally non-identifiable from genuine ones within the Taylor-based framework. This non-identifiability is not specific to THIS or Bayes-THIS but applies to any method operating within this framework.

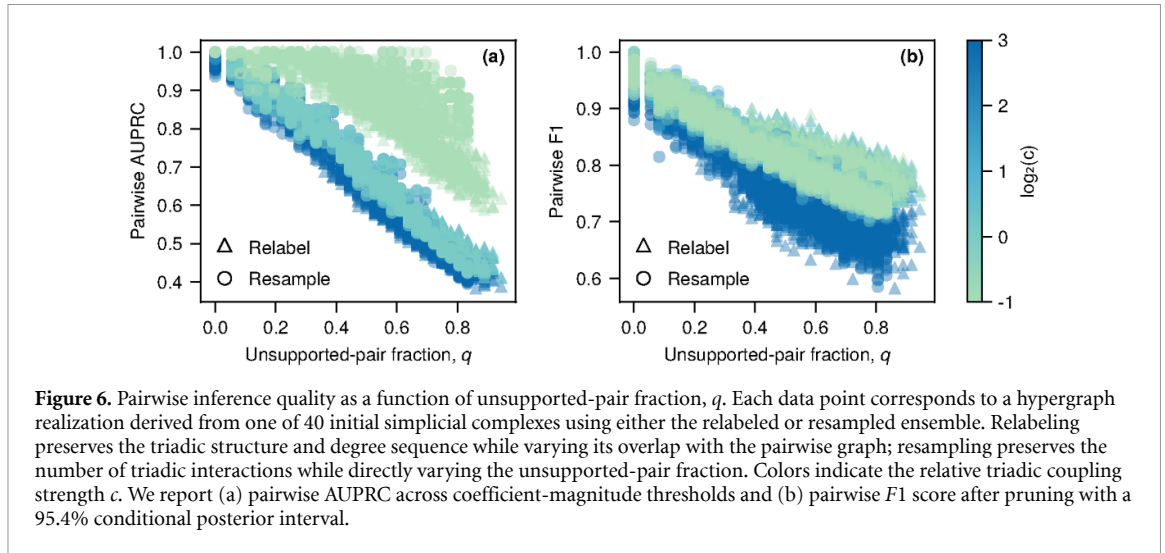
The source of this non-identifiability is the downward contribution of higher-order interactions to lower-order Taylor coefficients. As discussed in section 3 and noted by Delabays *et al* [11], triadic interactions generally contribute to coefficients in equation (2) corresponding to pairwise edges. The inferential consequences of this downward contribution depend on the nestedness of the underlying hypergraph. We quantify this here using the unsupported-pair fraction,

$$q \equiv \frac{1}{3|E^{(3)}|} \sum_{t \in E^{(3)}} \sum_{\substack{p \subset t \\ |p|=2}} \mathbf{1} [p \notin E^{(2)}],$$

where $E^{(2)}$ and $E^{(3)}$ are the existing pairwise and triadic edge sets, respectively. That is, q is the fraction of pair-triad incidences for which the pair is absent from the pairwise interaction set $E^{(2)}$. Note $0 \leq q \leq 1$ with $q = 0$ for a simplicial complex.

When this fraction is small—that is, when triadic interactions are concentrated on nodes that already share pairwise connections, as in a simplicial complex—the spurious pairwise contributions introduced by triadic interactions overlap with genuine pairwise signal. When we view hypergraph inference as a binary classification problem, these contributions therefore do not generate pairwise false positives. When the unsupported-pair fraction is large—when triadic interactions are concentrated among node pairs without pairwise connections—the downward contribution of triadic interactions to pairwise coefficients leads to systematic inference of nonexistent pairwise edges that are indistinguishable from true pairwise signal (see figure 3, for example). This effect is exacerbated when triadic interactions are strong relative to pairwise interactions. As shown in the SM, this effect also extends to higher orders: fourth-order interactions contribute to coefficients associated with triadic interactions and reduce triadic precision when the induced terms fall on absent triadic interactions.

To explore this phenomenon, we generate 40 random ER graphs with $n = 15$ nodes and pairwise edge probability $\rho_2 = 0.35$, then select 12 existing triangles in each graph to form an initial simplicial complex. We vary the alignment between pairwise and triadic structure using two complementary ensembles. In the first—which we will refer to as the ‘relabelled’ ensemble—we randomly shuffle which nodes carry the triadic interactions while leaving the pairwise structure unchanged. This preserves the triadic structure, including its degree sequence, but changes how it overlaps with the pairwise interactions. In the second, we instead resample the triadic interactions to target a prescribed fraction of unsupported pairs, keeping the pairwise graph and total number of triads fixed but allowing the triadic



degree sequence to vary; we refer to this ensemble as the ‘resampled’ ensemble. Comparing the two constructions allows us to test whether pairwise inference quality is governed by cross-order support or by structural features specific to either ensemble.

We repeat this procedure across a range of triadic coupling strengths c , keeping the pairwise coupling strength fixed. Following Lucas *et al* [27], we first normalize the pairwise and triadic coupling functions by the average pairwise and triadic degree, respectively, ensuring that c consistently controls the relative strength of pairwise and triadic coupling across hypergraph realizations. That is, equation (11) becomes

$$\dot{\theta}_i = \omega + \sum_{j=1}^n \frac{a_{ij}^{(2)}}{\langle k^{(2)} \rangle} [\sin(\theta_j - \theta_i + \varphi_2) - \sin \varphi_2] + \sum_{j,k=1}^n \frac{ca_{ijk}^{(3)}}{2\langle k^{(3)} \rangle} [\sin(\theta_j + \theta_k - 2\theta_i + \varphi_3) - \sin \varphi_3], \quad (14)$$

where the extra factor of 2 in the denominator of the triadic coupling coefficients is to avoid double counting the same simplex and where c varies from 0.5 to 8 in our experiment. We use a fixed dataset $\mathbf{X}$ throughout with $N = 600$ points sampled independently and uniformly at random from a hypercube with side length 1.0 centered at the origin, and we use equation (14) to compute derivatives analytically for each hypergraph realization, corrupting our measurements with additive Gaussian noise with standard deviation $\sigma = 0.01$. The size of the dataset and the level of noise should not be challenging for Bayes-THIS; we want to probe the effect of hypergraph structure here.

Figure 6 shows that, as expected, increasing the unsupported-pair fraction systematically degrades pairwise inference quality. Importantly, the relabeled and resampled ensembles follow the same overall relationship despite preserving different properties of the triadic structure. That this is a fundamental rather than incidental limitation is established by examining pairwise AUPRC—which measures the intrinsic separability of true and spurious pairwise coefficients across all possible magnitude thresholds τ —as a function of both unsupported-pair fraction and triadic coupling strength c (figure 6(a)). As c increases, pairwise AUPRC deteriorates more dramatically with increasing unsupported-pair fraction because spurious pairwise coefficients induced by triadic interactions grow in magnitude with c , increasing their overlap with coefficients associated with genuine pairwise edges. This reduces the global separability of true and false pairwise edges across all thresholds τ .

Notably, the conditional posterior interval pruning rule of section 4 is more robust to the triadic coupling strength c than magnitude-based thresholding (figure 6(b)). As c increases, the spurious pairwise contributions from triadic interactions grow in magnitude; however, so does the fitted residual variance $\hat{\beta}^{-1}$, widening the conditional posterior intervals and leaving the binary retention decision approximately stable across multiple magnitudes of c . The pairwise $F1$ score after posterior interval filtering therefore degrades more gradually with increasing unsupported-pair fraction. This provides further practical support for uncertainty-aware pruning. However, this robustness does not resolve the underlying non-identifiability: as the unsupported-pair fraction increases, false pairwise edges become genuinely indistinguishable from true ones in the posterior, and no coefficient-wise pruning rule can distinguish them.

This structural non-identifiability resonates with recent work on functional reducibility of higher-order networks [28]. Lucas *et al* propose an information-theoretic framework for determining whether

a hypergraph's higher-order structure can be discarded without sacrificing functional information, as encoded by diffusion dynamics. A hypergraph is reducible—well-described by a lower-order network—when higher-order edges are structurally and functionally redundant with lower-order ones, introducing no genuinely new spectral content in the multiorder Laplacian [27]. The structurally challenging regime we identify for Taylor-based inference—strong triadic interactions concentrated on node pairs without corresponding pairwise edges—is qualitatively similar to the irreducible regime: just as non-nested higher-order interactions introduce genuinely new spectral content that cannot be collapsed to lower orders without information loss, they introduce genuinely new first-order terms in the Taylor expansion that are indistinguishable from pairwise signal on absent node pairs. Both phenomena ultimately reflect a shared question of identifiability—namely, which structural features of a hypergraph can be reliably recovered from dynamical observations of a given type—though the precise relationship between functional reducibility and inferability remains an open question we flag for future work.

Practically, our findings imply a tradeoff that is intrinsic to the Taylor-based hypergraph inference framework. The flexibility that makes THIS and Bayes-THIS attractive—requiring no prior knowledge of the specific dynamical form of the coupling functions—comes at a cost: practitioners must have some knowledge of *structural* form instead. Specifically, whether higher-order interactions are approximately nested within lower-order structure (small unsupported-pair fraction) or are largely unsupported (large unsupported-pair fraction) is practically relevant prior information because it determines to what extent inferred pairwise—and more generally, lower-order interactions—should be trusted. That is, what the framework saves in assumptions about dynamics, it may require back in assumptions about structure.

7. Discussion

In this work, we have seen how adopting a Bayesian perspective on Taylor-based hypergraph inference improves robustness and data efficiency, enables uncertainty-aware pruning of weakly supported hyperedges, and allows—on a limited basis—assessment of data quality via PPCs. These three methodological contributions compose naturally into a coherent workflow: assess data quality first with the PPC, run Bayes-THIS, and then prune hyperedges whose conditional posterior intervals include zero. We summarize this workflow with reference to the relevant equations in the appendix. Importantly, Bayes-THIS requires very little manual tuning of hyperparameters—i.e. no sparsity parameter λ or coefficient-magnitude threshold τ —and can be parallelized across nodes like THIS.

There are limitations of this work that are worth noting. First, sparse Bayesian regression with ARD assumes residual noise is independent across observations, an approximation that is violated when derivatives are estimated from noisy trajectories. Nevertheless, the noisy-trajectory experiments in section 3 suggest that the relative advantage of Bayes-THIS over THIS is fairly robust to this mismatch. A more principled treatment could account for the derivative-noise covariance directly: if this covariance can be estimated independently and reliably, premultiplying equation (5) by a whitening transformation [22] can remove the induced residual correlations while retaining the existing Bayes-THIS inference machinery. More generally, extending Bayes-THIS to better account for the structure of derivative noise is a natural direction for future work. Fung *et al* [14] recently proposed an adaptation of sparse Bayesian regression with ARD that explicitly accounts for how measurement noise propagates through finite-difference estimates of time derivatives, and integrating this approach into Bayes-THIS would be a natural step toward reliable application to real trajectory data. More general applications may also require confronting non-equilibrium dynamics for which trajectories do not remain localized around any natural Taylor-expansion point. The PPC provides one preliminary tool for assessing data quality, but designing good experiments for hypergraph inference from dynamics remains an open problem.

Perhaps more importantly, as discussed in section 6, the contribution of higher-order interactions to lower-order monomial coefficients is an inherent structural feature of the Taylor-based framework, not an artifact of the regression method. Bayes-THIS does not resolve the ambiguity introduced by higher-order interactions with limited support from lower-order structure. Additional structural assumptions could be incorporated through a prior that treats coefficients across interaction orders jointly, potentially regularizing this ambiguity when such prior information is justified.

A recurring theme across the later sections of this work is that the question of *whether* higher-order structures can be reliably inferred from dynamics is as consequential as the question of *how* to infer them. The practical utility of any inference method depends on understanding and being able to recognize the conditions under which it can succeed—conditions that are shaped not only by data quality and quantity but also by structural and dynamical properties of the true hypergraph itself. By simultaneously advancing the inference methodology and exploring its fundamental limits, we hope this work contributes to placing hypergraph reconstruction on firmer empirical and theoretical footing.

Acknowledgments

Katerina Tang would like to acknowledge helpful conversations with Anastasia Bizyaeva and Daniel Kaiser and to thank Robin Delabays for generously sharing the code associated with their original work [11] on THIS. This research was conducted with support from the Cornell University Center for Advanced Computing.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://doi.org/10.5281/zenodo.20040617> [50].

Supplementary material available at <https://doi.org/10.1088/2632-072X/aeae02/data1>.

Author contributions

Katerina Tang  [0000-0002-9177-0830](https://orcid.org/0000-0002-9177-0830)

Conceptualization (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Visualization (lead), Writing – original draft (lead)

Vivek Srikrishnan  [0000-0003-0049-3805](https://orcid.org/0000-0003-0049-3805)

Conceptualization (supporting), Resources (lead), Supervision (equal), Writing – review & editing (equal)

Jackson Kulik  [0009-0007-0740-0436](https://orcid.org/0009-0007-0740-0436)

Conceptualization (supporting), Methodology (supporting), Supervision (equal), Writing – review & editing (equal)

Appendix. Algorithm for Bayes-THIS

Algorithm 1 summarizes the full Bayes-THIS workflow—with posterior predictive check and conditional posterior interval pruning—for a single node i . Since inference is independent across nodes, the algorithm can be parallelized in a straightforward manner.

Algorithm 1. Bayes-THIS with posterior predictive check and posterior interval pruning.

Input: state data $\mathbf{X}$ centered about expansion point, time derivatives $\dot{\mathbf{x}}_i$, noise precision $\beta = \sigma^{-2}$ (for PPC step), PPC significance threshold p^* , number of posterior draws S , nominal interval level γ

Output: sparse inferred coefficient vector ξ_i

1 Form design matrix Θ from equation (3)

// PPC described in section 5

2 Partition $\Theta = [\Theta_1 \mid \Theta_2]$ into pairwise and triadic monomials, respectively

3 Fit pairwise-only model via sparse Bayesian regression with ARD (with $\beta = \sigma^{-2}$ fixed) on $(\Theta_1, \dot{\mathbf{x}}_i)$ to get posterior mean μ_1 (equation (9)) and covariance Σ_1 (equation (10))

4 Compute triadic projection matrix $\mathbf{P}_2$

5 **foreach** $s = 1, \dots, S$ **do**

6 Draw $\xi_1^{(s)} \sim \mathcal{N}(\mu_1, \Sigma_1)$

7 Compute observed residual $\hat{\epsilon}^{(s)}$ from equation (13) and discrepancy:

$$T^{(s)} = \hat{\epsilon}^{(s)\top} \mathbf{P}_2 \hat{\epsilon}^{(s)}$$

8 Draw replicate residual $\epsilon_{\text{rep}}^{(s)} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_N)$ and compute discrepancy:

$$T_{\text{rep}}^{(s)} = \epsilon_{\text{rep}}^{(s)\top} \mathbf{P}_2 \epsilon_{\text{rep}}^{(s)}$$

9 **end**

10 Estimate posterior predictive p -value:

$$p = \frac{1}{S} \sum_{s=1}^S \mathbf{1}[T_{\text{rep}}^{(s)} \geq T^{(s)}]$$

11 **if** $p \geq p^*$ **then**

12 | warn that data do not reliably support inference beyond pairwise model

13 **end**

// full higher-order model inference

14 Fit full model via sparse Bayesian regression with ARD on $(\Theta, \dot{\mathbf{x}}_i)$ to get posterior mean μ (equation (9)) and covariance Σ (equation (10))

// conditional posterior interval pruning as described in section 4

15 **foreach** m in ARD-selected active set **do**

16 Compute conditional posterior variance $\sigma_{m|-m}^2$ from equation (12)

17 Apply posterior interval pruning:

$$\xi_{i,m} = \begin{cases} 0 & |\mu_{i,m}| \leq z_{(1+\gamma)/2} \sigma_{i,m|-m} \\ \mu_{i,m} & \text{else} \end{cases}$$

18 **end**

References

- [1] Arya S, George A B and O'Dwyer J P 2023 Sparsity of higher-order landscape interactions enables learning and prediction for microbiomes *Proc. Natl Acad. Sci.* **120** e2307313120
- [2] Battiston F et al 2021 The physics of higher-order interactions in complex systems *Nat. Phys.* **17** 1093–8
- [3] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G and Petri G 2020 Networks beyond pairwise interactions: structure and dynamics *Phys. Rep.* **874** 1–92
- [4] Benson A R, Abebe R, Schaub M T, Jadbabaie A and Kleinberg J 2018 Simplicial closure and higher-order link prediction *Proc. Natl Acad. Sci.* **115** E11221–30
- [5] Berger J O 1985 Statistical Decision Theory and Bayesian Analysis *Springer Series in Statistics* (Springer)
- [6] Bick C, Gross E, Harrington H A and Schaub M T 2023 What are higher-order networks? *SIAM Rev.* **65** 686–731
- [7] Breakspear M 2017 Dynamic models of large-scale brain activity *Nat. Neurosci.* **20** 340–52
- [8] Brunton S L, Proctor J L and Kutz J N 2016 Discovering governing equations from data by sparse identification of nonlinear dynamical systems *Proc. Natl Acad. Sci.* **113** 3932–7
- [9] Carletti T, Battiston F, Cencetti G and Fanelli D 2020 Random walks on hypergraphs *Phys. Rev. E* **101** 022308
- [10] Casadiego J, Nitzan M, Hallerberg S and Timme M 2017 Model-free inference of direct network interactions from nonlinear collective dynamics *Nat. Commun.* **8** 2192
- [11] Delabays R, De Pasquale G, Dörfler F and Zhang Y 2025 Hypergraph reconstruction from dynamics *Nat. Commun.* **16** 2691
- [12] Fasel U, Kutz J N, Brunton B W and Brunton S L 2022 Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control *Proc. R. Soc. A* **478** 20210904
- [13] Fuentes R, Nayek R, Gardner P, Dervilis N, Rogers T, Worden K and Cross E J 2021 Equation discovery for nonlinear dynamical systems: a Bayesian viewpoint *Mech. Syst. Signal Process.* **154** 107528
- [14] Fung L, Fasel U and Juniper M 2025 Rapid Bayesian identification of sparse nonlinear dynamics from scarce and noisy data *Proc. R. Soc. A* **481** 20240200

- [15] Gelman A, Meng X-L and Stern H 1996 Posterior predictive assessment of model fitness via realized discrepancies *Stat. Sin.* **6** 733–60
- [16] Grilli J, Barabás G, Michalska-Smith M J and Allesina S 2017 Higher-order interactions stabilize dynamics in competitive network models *Nature* **548** 210–3
- [17] Han X, Shen Z, Wang W-X and Di Z 2015 Robust reconstruction of complex networks from sparse data *Phys. Rev. Lett.* **114** 028701
- [18] Hirsh S M, Barajas-Solano D A and Kutz J N 2022 Sparsifying priors for Bayesian uncertainty quantification in model discovery *R. Soc. Open Sci.* **9** 211823
- [19] Iacopini I, Petri G, Barrat A and Latora V 2019 Simplicial models of social contagion *Nat. Commun.* **10** 2485
- [20] Karlebach G and Shamir R 2008 Modelling and analysis of gene regulatory networks *Nat. Rev. Mol. Cell Biol.* **9** 770–80
- [21] Keeling M J and Rohani P 2002 Estimating spatial coupling in epidemiological systems: a mechanistic approach *Ecol. Lett.* **5** 20–29
- [22] Kessy A, Lewin A and Strimmer K 2018 Optimal whitening and decorrelation *Am. Stat.* **72** 309–14
- [23] Kramer M A, Eden U T, Cash S S and Kolaczyk E D 2009 Network inference with confidence from multivariate time series *Phys. Rev. E* **79** 061916
- [24] Ladrone C, Guo S, Kendrick K and Feng J 2009 Beyond element-wise interactions: identifying complex interactions in biological processes *PLoS One* **4** e6899
- [25] Landry N W, Thompson W, Hébert-Dufresne L and Young J-G 2024 Reconstructing networks from simple and complex contagions *Phys. Rev. E* **110** L042301
- [26] Lizotte S, Young J-G and Allard A 2023 Hypergraph reconstruction from uncertain pairwise observations *Sci. Rep.* **13** 21364
- [27] Lucas M, Cencetti G and Battiston F 2020 Multiorder Laplacian for synchronization in higher-order networks *Phys. Rev. Res.* **2** 033410
- [28] Lucas M, Gallo L, Ghavasieh A, Battiston F and De Domenico M 2026 Reducibility of higher-order networks from dynamics *Nat. Commun.* **17** 1551
- [29] Ma C, Chen H-S, Lai Y-C and Zhang H-F 2018 Statistical inference approach to structural reconstruction of complex networks from binary time series *Phys. Rev. E* **97** 022301
- [30] Malizia F, Corso A, Gambuzza L V, Russo G, Latora V and Frasca M 2024 Reconstructing higher-order interactions in coupled dynamical systems *Nat. Commun.* **15** 5184
- [31] Mangan N M, Kutz J N, Brunton S L and Proctor J L 2017 Model selection for dynamical systems via sparse regression and information criteria *Proc. R. Soc. A* **473** 20170009
- [32] Mangan N M, Brunton S L, Proctor J L and Kutz J N 2017 Inferring biological networks by sparse identification of nonlinear dynamics *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2** 52–63
- [33] Messenger D A and Bortz D M 2021 Weak SINDy: Galerkin-based data-driven model selection *Multiscale Model. Simul.* **19** 1474–97
- [34] Niven R K, Cordier L, Mohammad-Djafari A, Abel M and Quade M 2024 Dynamical system identification, model selection and model uncertainty quantification by Bayesian inference *Chaos* **34** 083140
- [35] Patania A, Petri G and Vaccarino F 2017 The shape of collaborations *EPJ Data Sci.* **6** 18
- [36] Peixoto T P 2019 Network reconstruction and community detection from dynamics *Phys. Rev. Lett.* **123** 128301
- [37] Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer P J and Vaccarino F 2014 Homological scaffolds of brain functional networks *J. R. Soc. Interface* **11** 20140873
- [38] Skardal P S and Arenas A 2019 Abrupt desynchronization and extensive multistability in globally coupled oscillator simplexes *Phys. Rev. Lett.* **122** 248301
- [39] Skardal P S and Arenas A 2020 Higher order interactions in complex networks of phase oscillators promote abrupt synchronization switching *Commun. Phys.* **3** 218
- [40] Tibshirani R 1996 Regression shrinkage and selection via the lasso *J. R. Stat. Soc. B* **58** 267–88
- [41] Timme M and Casadiego J 2014 Revealing networks from dynamics: an introduction *J. Phys. A: Math. Theor.* **47** 343001
- [42] Tipping M E 2001 Sparse Bayesian learning and the relevance vector machine *J. Mach. Learn. Res.* **1** 211–44
- [43] Tipping M E and Faul A C 2003 Fast marginal likelihood maximisation for sparse Bayesian models *Proc. 9th Int. Workshop on Artificial Intelligence and Statistics* pp 276–83
- [44] Tirabassi G, Sevilla-Escoboza R, Buldú J M and Masoller C 2015 Inferring the connectivity of coupled oscillators from time-series statistical similarity analysis *Sci. Rep.* **5** 10829
- [45] Wang H, Ma C, Chen H-S, Lai Y-C and Zhang H-F 2022 Full reconstruction of simplicial complexes from binary contagion and Ising data *Nat. Commun.* **13** 3043
- [46] Wu X, Wang W and Zheng W X 2012 Inferring topologies of complex networks with hidden variables *Phys. Rev. E* **86** 046106
- [47] Young J-G, Petri G and Peixoto T P 2021 Hypergraph reconstruction from network data *Commun. Phys.* **4** 135
- [48] Yu S, Yang H, Nakahara H, Santos G S, Nikolić D and Plenz D 2011 Higher-order interactions characterized in cortical activity *J. Neurosci.* **31** 17514–26
- [49] Zhang S and Lin G 2018 Robust data-driven discovery of governing physical laws with error bars *Proc. R. Soc. A* **474** 20180305
- [50] Katerina T 2026 “kbtang28/bayesian-hypergraph-inference: v2.0” *Zenodo* (<https://doi.org/10.5281/zenodo.20040617>)