Chaos
An Interdisciplinary Journal of Nonlinear Science

RESEARCH ARTICLE | JULY 30 2024

Stepwise reconstruction of higher-order networks from dynamics

Yingbang Zang ; Ziye Fan ; Zixi Wang ; Yi Zheng ; Li Ding ; Xiaoqun Wu

Chaos 34, 073156 (2024)
https://doi.org/10.1063/5.0210741

Stepwise reconstruction of higher-order networks from dynamics

Cite as: Chaos 34, 073156 (2024); doi: 10.1063/5.0210741
Submitted: 27 March 2024 · Accepted: 2 July 2024 ·
Published Online: 30 July 2024

Yingbang Zang,[1] Ziye Fan,[1] Zixi Wang,[2] Yi Zheng,[3,a)] Li Ding,[4] and Xiaoqun Wu[3,a)]

AFFILIATIONS

[1]School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
[2]School of Journalism and Communication, Wuhan University, Wuhan 430072, China
[3]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[4]School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China

[a)]**Authors to whom correspondence should be addressed:** xymather@whu.edu.cn and xqwu@whu.edu.cn

ABSTRACT

Higher-order networks present great promise in network modeling, analysis, and control. However, reconstructing higher-order interactions remains an open problem. A significant challenge is the exponential growth in the number of potential interactions that need to be modeled as the maximum possible node number in an interaction increases, making the reconstruction exceedingly difficult. For higher-order networks, where higher-order interactions exhibit properties of lower-order dependency and weaker or fewer higher-order connections, we develop a reconstruction scheme integrating a stepwise strategy and an optimization technique to infer higher-order networks from time series. This approach significantly reduces the potential search space for higher-order interactions. Simulation experiments on a wide range of networks and dynamical systems demonstrate the effectiveness and robustness of our method.

Previous research on network reconstruction has mainly focused on pairwise interactions, thereby constraining the scope of their applications. Recently, higher-order interactions have attracted widespread attention among researchers, and concurrently, the reconstruction of higher-order structures has emerged as a pressing scientific imperative. A natural challenge in higher-order network reconstruction is the curse of dimensionality, which refers to the phenomenon where an immense number of potential interactions need to be modeled due to the incorporation of higher-order interactions, making it extremely difficult to determine which interactions truly exist. To date, only a few methods have been developed for reconstructing higher-order networks, and even fewer methods can address the dimensionality curse in higher-order reconstruction. Focusing on networks characterized by lower-order dependency and weaker or fewer higher-order connections, we develop a data-driven stepwise reconstruction scheme to reconstruct higher-order interactions. This approach provides a possible perspective for tackling the dimensionality curse in higher-order structure reconstruction. Simulation experiments have validated our method's effectiveness, even when these two properties are not fully met.

I. INTRODUCTION

Complex networks are widely present in various fields, such as World Wide Web,[1] power grids,[2] and transportation systems.[3] Networks with different topological structures exhibit varied evolutionary characteristics,[4,5] propagation patterns,[6] and levels of information security.[7] Moreover, the study of brain networks reveals a similar complexity, where the topological organization of brain regions profoundly influences cognitive functions and neurological disorders.[8] Therefore, knowing the topological structure is the prerequisite for understanding and explaining the networks' evolutionary mechanisms and functional behaviors. However, in practice, acquiring network structures is challenging, specifically in complex networks with a large number of nodes, connections, and diverse interaction characteristics. Therefore, reconstructing complex networks' topology holds significant theoretical and practical significance in network science. Focusing on networks with pairwise interactions over the past two decades, researchers have developed diverse methods for reconstructing network structures, including adaptive synchronization approaches,[9–13] Granger causality-based techniques,[14] statistical inference,[15,16] compressed sensing-based strategies,[17–22]

deep learning-based approaches,[23] and other optimization-based methods.[24–27]

Beyond pairwise interaction, researchers have recently realized the importance of higher-order structures for network representation and interaction.[28,29] Based on higher-order structures, many new insights have been developed in various fields, such as ecosystem,[30] social contagion,[31] and game theory.[32] Simultaneously, the introduction of higher-order structures brings significant challenges to network reconstruction. Specifically, to reconstruct higher-order interactions, more potential connections will inevitably be modeled[33] and the number will grow exponentially, making it more challenging to determine which interactions exist. This paper refers to this phenomenon as *the dimensionality curse in higher-order structure reconstruction*. The previous methods for inferring networks with only pairwise interactions are not directly applicable to higher-order networks. So far, there have been some approaches for reconstructing higher-order interactions, including Bayesian or probability-based inference methods,[34–38] optimization-based techniques,[39,40] and adaptive synchronization-based strategies.[41] However, few techniques have been developed to deal with the higher-order structure reconstruction's dimensionality curse.

In many cases, higher-order structures display a certain dependence on lower-order structures (referred to as the "*lower-order dependency*" property), and their overall impact on network evolution is relatively minor compared to lower-order structures, especially in scenarios with fewer or less intense higher-order interactions (referred to as the "*weaker or fewer higher-order connections*" property). For example, in a friendship network, the formation of friendships among three individuals often relies on the existence of friendships between pairs, indicating a certain dependence of higher-order interactions on lower-order connections (lower-order dependency). Compared to friendships between two individuals, most people tend to have fewer friendships among three individuals (fewer), or, despite possibly having more three-individual friendships, the time and energy investment is relatively less intense (weaker). To reconstruct higher-order networks with "lower-order dependency" and "weaker or fewer higher-order connections" properties, we propose a data-driven reconstruction scheme combining stepwise strategy and optimization techniques, which can tackle the dimensionality curse in higher-order structure reconstruction. It is worth noting that while our method is primarily designed for networks with these two properties, even for more general networks that may not fully adhere to these characteristics, our approach still demonstrates significant superiority over baseline methods in various simulations.

The paper is organized as follows. Section II introduces the proposed method. Section III provides some simulations to verify the validity of the proposed method. Finally, Sec. IV presents the conclusion and discussion.

II. METHOD

A. Network model and problem definition

For a higher-order complex network consisting of n nodes and with a maximum interaction order of K, the dynamics of the ith node can be written as

$$\dot{\boldsymbol{x}}_i = \boldsymbol{f}_i(\boldsymbol{x}_i) + \sum_{k=2}^{K} \sigma_k \sum_{j_2,\ldots,j_k=1}^{n} a^{(k)}_{ij_2,\ldots,j_k} \boldsymbol{g}^{(k)}_i(\boldsymbol{x}_i, \boldsymbol{x}_{j_2}, \ldots, \boldsymbol{x}_{j_k}), \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the state of the ith node with self-dynamics function $\boldsymbol{f}_i : \mathbb{R}^d \to \mathbb{R}^d$. $\sigma_k \in \mathbb{R}^+$ and $\boldsymbol{g}^{(k)}_i : \mathbb{R}^{kd} \to \mathbb{R}^d$ $(k = 2, \ldots, K)$ are the coupling strength and pattern of k-body interaction, respectively. The topology of the higher-order complex network is represented by adjacency tensors $\boldsymbol{A}^{(2)} = \left(a^{(2)}_{ij}\right) \in \mathbb{R}^{n\times n}, \ldots, \boldsymbol{A}^{(K)} = \left(a^{(K)}_{ij_2,\ldots,j_K}\right) \in \mathbb{R}^{\overbrace{n\times\cdots\times n}^{K}}$. For instance, $a^{(k)}_{ij_2,\ldots,j_k} \neq 0$ if there exists interaction $\boldsymbol{g}^{(k)}_i\left(\boldsymbol{x}_i, \boldsymbol{x}_{j_2}, \ldots, \boldsymbol{x}_{j_k}\right)$, but 0 otherwise. We assume that the network has no self-loops to prevent misidentifying connections involving self-loops as higher-order interactions,[42] so $a^{(k)}_{ij_2,\ldots,j_k} = 0$ if the cardinality of the set $\{i, j_2, \ldots, j_k\}$ is not equal to k. We further assume that the node self-dynamics function $\boldsymbol{f}_i$ and interaction forms $\boldsymbol{g}^{(k)}_i$ $(k = 2, \ldots, K)$ are known for the ith node $(i = 1, \ldots, n)$.

For simplicity, we primarily focus on unweighted and undirected networks with the maximum interaction order of $K = 3$. For the sake of illustration, we adopt node dimension d of 1. The situation where $d > 1$ can be similarly addressed. Then, Eq. (1) can be rewritten as

$$\dot{x}_i = f_i(x_i) + \sigma_2 \sum_{j=1}^{n} a^{(2)}_{ij} g^{(2)}_i\left(x_i, x_j\right) + \sigma_3 \sum_{j=1}^{n}\sum_{k=1}^{n} a^{(3)}_{ijk} g^{(3)}_i\left(x_i, x_j, x_k\right).$$

Define $\boldsymbol{x}(t) = [x_1(t), \ldots, x_n(t)]^\top \in \mathbb{R}^n$. Our goal is to reconstruct the topologies at different interaction orders (i.e., $\boldsymbol{A}^{(2)}$ and $\boldsymbol{A}^{(3)}$) based on the node states' time-series data at m equally spaced time points (the data are denoted as $\boldsymbol{X} = [\boldsymbol{x}(t_1), \ldots, \boldsymbol{x}(t_m)]^\top \in \mathbb{R}^{m\times n}$). Let $\tilde{y}_i = \dot{x}_i - f_i(x_i) \in \mathbb{R}$, then

$$\tilde{y}_i = \sigma_2 \sum_{j=1}^{n} a^{(2)}_{ij} g^{(2)}_i(x_i, x_j) + \sigma_3 \sum_{j=1}^{n}\sum_{k=1}^{n} a^{(3)}_{ijk} g^{(3)}_i(x_i, x_j, x_k), \tag{2}$$

where $\dot{x}_i$ can be numerically derived from time-series data of the ith node through the five-point approximation,[43] $f_i(x_i)$, $g^{(2)}_i(x_i, x_j)$, and $g^{(3)}_i(x_i, x_j, x_k)$ can be obtained by substituting data into the known self-dynamics f_i and interaction forms $g^{(2)}_i$ and $g^{(3)}_i$. Considering that many interaction dynamics often exhibit some form of symmetry related to external nodes, such as Kuramoto oscillators (Appendix A) and Hindmarsh–Rose neurons (Appendix B), we assume that $g^{(3)}_i$ satisfies the condition $g^{(3)}_i(x_i, x_j, x_k) \equiv g^{(3)}_i(x_i, x_k, x_j)$ for any x_k, x_j, termed as symmetric dynamics condition. Denote $g^{(k)}_i(x_i, x_{j_2}, \ldots, x_{j_k})$ as $g^{(k)}_{ij_2,\ldots,j_k}$.

Denote

$$\bar{\boldsymbol{\mathcal{G}}}^{(2)}_i = \left[g^{(2)}_{i1}, \ldots, g^{(2)}_{i,i-1}, g^{(2)}_{i,i+1}, \ldots, g^{(2)}_{in}\right]^\top \in \mathbb{R}^{n-1},$$

$$\bar{\boldsymbol{\mathcal{G}}}^{(3)}_i = \left[g^{(3)}_{i12}, \ldots, g^{(3)}_{i1n}, g^{(3)}_{i23}, \ldots, g^{(3)}_{i2n}, \ldots, g^{(3)}_{i,n-1,n}\right]^\top \in \mathbb{R}^{(n-1)(n-2)/2},$$

$$\bar{\boldsymbol{\mathcal{G}}}_i = \left[\left(\bar{\boldsymbol{\mathcal{G}}}^{(2)}_i\right)^\top, \left(\bar{\boldsymbol{\mathcal{G}}}^{(3)}_i\right)^\top\right]^\top \in \mathbb{R}^{n-1+(n-1)(n-2)/2},$$

$$\bar{A}_i = \left[\sigma_2\left(a_{i1}^{(2)},\ldots,a_{i,i-1}^{(2)},a_{i,i+1}^{(2)},\ldots,a_{in}^{(2)}\right),2\sigma_3\left(a_{i12}^{(3)},\ldots,a_{i1n}^{(3)},a_{i23}^{(3)},\right.\right.$$
$$\left.\left.\ldots,a_{i2n}^{(3)},\ldots,a_{i,n-1,n}^{(3)}\right)\right]^\top \in \mathbb{R}^{n-1+(n-1)(n-2)/2}.$$

B. Direct approach and its challenge

To reconstruct the network, a natural idea is to employ optimization techniques, transforming the network reconstruction problem into solving a system of equations, as outlined below. By writing Eq. (2) in a compact form, we obtain

$$\tilde{y}_i = \bar{\mathcal{G}}_i^\top \bar{A}_i. \tag{3}$$

Further, by substituting all time-series data $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ into Eq. (3), we have

$$\boldsymbol{y}_i = \bar{\boldsymbol{G}}_i \bar{\boldsymbol{A}}_i, \tag{4}$$

where $\boldsymbol{y}_i = [\tilde{y}_i(t_1),\ldots,\tilde{y}_i(t_m)]^\top \in \mathbb{R}^m$, $\bar{\boldsymbol{G}}_i = [\bar{\mathcal{G}}_i(t_1),\ldots,\bar{\mathcal{G}}_i(t_m)]^\top \in \mathbb{R}^{m\times(n-1+(n-1)(n-2)/2)}$ can be directly obtained, and $\bar{\boldsymbol{A}}_i \in \mathbb{R}^{n-1+(n-1)(n-2)/2}$, representing the interaction of the ith node with others, is the parameter to be determined. In this way, the network reconstruction problem is transformed into solving $\bar{\boldsymbol{A}}_i$ for each node individually. Based on the relationship between the number of rows and columns of $\bar{\boldsymbol{G}}_i$, this problem can be addressed using some classical methods. Recently, Federico Malizia *et al.* have pioneered this direct strategy in reconstructing higher-order networks[39] and proposed methods to solve Eq. (4) within the context of higher-order networks, including adding additional constraints and extending the signal lasso method to handle higher-order interactions. However, in this direct approach, the dimension of potential interactions $\bar{\mathcal{G}}_i$ (i.e., the number of columns of $\bar{\boldsymbol{G}}_i$) significantly increases with the number of nodes and the maximum interaction order, posing a great challenge to the reconstruction. It should be noted that, although an example in the paper by Federico Malizia *et al.* proposed a system of equations solving method that can reduce the data requirements for reconstruction to some extent, this method does not attempt to reduce the number of potential interactions that need to be modeled. Moreover, to achieve satisfactory reconstruction results, the required data volume for their solving method still remains approximately proportional to the number of modeled interactions, thus fundamentally failing to surmount the curse of dimensionality. To address the dimensionality curse in higher-order structure reconstruction, we will propose a stepwise reconstruction strategy for higher-order networks with specific properties, which can effectively reduce the number of potential interactions requiring modeling, thereby fundamentally alleviating the curse of dimensionality in reconstructing the higher-order structures.

C. Stepwise reconstruction scheme

For higher-order networks, where higher-order interactions exhibit "lower-order dependency" and "weaker or fewer higher-order connections" properties, we propose a method to tackle the dimensionality curse in higher-order structure reconstruction. To be specific, we develop a reconstruction scheme combining stepwise strategy and optimization techniques to infer higher-order networks from time-series data, as described below.

First, we temporarily treat the network evolution to be solely driven by pairwise interactions, leading to the following approximation:

$$\boldsymbol{y}_i \approx \boldsymbol{G}_i^{(2)} \hat{\boldsymbol{A}}_i^{(2)}, \tag{5}$$

where $\boldsymbol{G}_i^{(2)} = \left[\mathcal{G}_i^{(2)}(t_1),\ldots,\mathcal{G}_i^{(2)}(t_m)\right]^\top \in \mathbb{R}^{m\times(n-1)}$, $\mathcal{G}_i^{(2)} = \left[g_{i1}^{(2)},\ldots,g_{i,i-1}^{(2)},g_{i,i+1}^{(2)},\ldots,g_{in}^{(2)}\right]^\top \in \mathbb{R}^{n-1}$ is the set representing the potential pairwise interactions of the ith node, and $\hat{\boldsymbol{A}}_i^{(2)} = \left[\sigma_2\left(\hat{a}_{i1}^{(2)},\ldots,\hat{a}_{i,i-1}^{(2)},\hat{a}_{i,i+1}^{(2)},\ldots,\hat{a}_{in}^{(2)}\right)\right]^\top \in \mathbb{R}^{n-1}$ represents approximate two-body interaction relationship of the ith node. Given the inevitable influence of higher-order structures on reconstruction, the identified coefficients $\hat{\boldsymbol{A}}_i^{(2)}$ inevitably contain some errors, the magnitude of which depends on the extent to which higher-order structures influence network dynamics. Due to the "weaker or fewer higher-order connections" properties of higher-order structures, the impact of higher-order interactions on network dynamics is relatively small, thus, the identification accuracy of lower-order structures is relatively high. By setting a threshold or using the prior degrees of nodes, we determine which edges are more likely to exist. Perform this step independently for each node, then the approximate two-body interaction matrix $\hat{\boldsymbol{A}}^{(2)}$ can be obtained.

Based on the "lower-order dependency" properties of higher-order interactions, we can contract the initially high-dimensional three-body interaction space $\bar{\mathcal{G}}_i^{(3)} \in \mathbb{R}^{O(n^2/2)}$ into a smaller space $\mathcal{G}_i^{(3)} \in \mathbb{R}^{O(s)}$, where $s \le n^2/2$. When all the two-body interactions exist, $s = n^2/2$, which implies that the stepwise strategy degenerates into the direct strategy. For notably sparse two-body structures, $s \ll n^2/2$. Various methods exist for screening potential higher-order interactions to be modeled, using already identified approximate lower-order structures. For simplicity, a typical strategy is adopted, where higher-order interactions are incorporated only when all corresponding lower-order interactions already exist. Therefore, in the network reconstruction process described here, the prerequisite for modeling a three-body interaction is the existence of all three pairs of corresponding two-body interactions in the identified two-body interaction matrix. A specific example of $\mathcal{G}_i^{(3)}$ is illustrated in Figs. 1(d)–1(e), where $g_{123}^{(3)}$ is included in $\mathcal{G}_1^{(3)}$ only if $g_{12}^{(2)}$, $g_{13}^{(2)}$, and $g_{23}^{(2)}$ are all considered to exist. It should be noted that the form of s may exhibit slight variations depending on different higher-order structure screening rules employed. However, generally speaking, s is mainly influenced by the network's sparsity. The sparser the network, the smaller s becomes, resulting in a reduced number of interactions that need to be modeled. Since real-world networks are generally sparse, the number of interactions modeled by the stepwise strategy will be significantly fewer than those modeled by the direct strategy. Detailed analysis regarding the number of potential interactions that need to be modeled can be found in Appendix C.

Then, substitute time series data into screened potential pairwise and triplet interactions, we get

$$\boldsymbol{y}_i \approx \boldsymbol{G}_i \boldsymbol{A}_i, \tag{6}$$

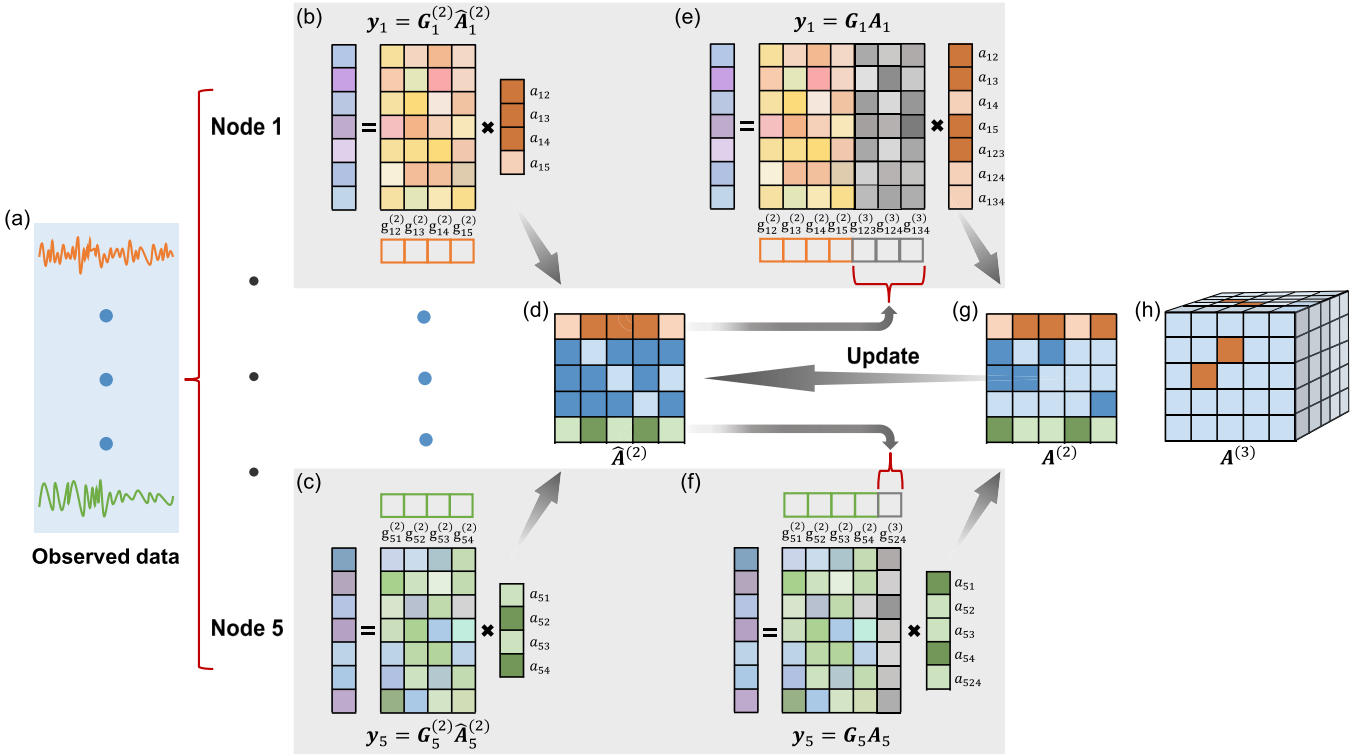

FIG. 1. The schematic diagram of stepwise reconstruction of higher-order networks from dynamics. The complex network consists of n nodes, with a maximum interaction order of 3, where n is specified as 5 for illustrative purposes. (a) The observed time-series data. (b) and (c) For each node, construct potential pairwise interactions, i.e., the interactions with all other nodes, as depicted by the orange boxes for node 1, $\mathcal{G}_1^{(2)} = [g_{12}^{(2)}, g_{13}^{(2)}, g_{14}^{(2)}, g_{15}^{(2)}]^\top$ and green boxes for node 5, $\mathcal{G}_5^{(2)} = [g_{51}^{(2)}, g_{52}^{(2)}, g_{53}^{(2)}, g_{54}^{(2)}]^\top$. Substitute time series data into potential pairwise interactions, establish, and solve equations to obtain the approximate two-body interaction relationships of each node, then (d) the approximate two-body interaction matrix $\hat{\boldsymbol{A}}^{(2)}$ can be derived. (e) and (f) Based on the preliminary two-body interaction matrix, for each node, potential three-body interactions are screened out, as depicted by the gray boxes $\left(\text{for node 1, } \mathcal{G}_1^{(3)} = [g_{123}^{(3)}, g_{124}^{(3)}, g_{134}^{(3)}]^\top; \text{ for node 5 , } \mathcal{G}_5^{(3)} = g_{524}^{(3)}\right)$. Substitute time series data into potential pairwise and triplet interactions, establish, and solve equations to obtain the approximate two- and three- body interaction relationships of each node $\Big($for node 1, $\boldsymbol{A}_1^{(2)} = [a_{12}, a_{13}, a_{14}, a_{15}]^\top, \boldsymbol{A}_1^{(3)} = [a_{123}, a_{124}, a_{134}]^\top$ in this iteration round, $\boldsymbol{A}_1 = \left[\left(\boldsymbol{A}_1^{(2)}\right)^\top, \left(\boldsymbol{A}_1^{(3)}\right)^\top\right]^\top$; for node 5, $\boldsymbol{A}_5^{(2)} = [a_{51}, a_{52}, a_{53}, a_{54}]^\top, \boldsymbol{A}_5^{(3)} = a_{524}$ in this iteration round, $\boldsymbol{A}_5 = \left[\left(\boldsymbol{A}_5^{(2)}\right)^\top, \left(\boldsymbol{A}_5^{(3)}\right)^\top\right]^\top\Big)$, then the approximate (g) two-body interaction matrix $\boldsymbol{A}^{(2)}$ and (h) three-body interaction matrix $\boldsymbol{A}^{(3)}$ can be derived. Update the initial two-body interaction matrix (d) with the newly acquired one (g). Iterate through the steps, from (d) to (h) and from (g) to (d), until the termination condition is satisfied.

where $\boldsymbol{G}_i = [\mathcal{G}_i(t_1), \ldots, \mathcal{G}_i(t_m)]^\top \in \mathbb{R}^{m\times(n-1+O(s))}$, $\mathcal{G}_i = \left[\left(\mathcal{G}_i^{(2)}\right)^\top, \left(\mathcal{G}_i^{(3)}\right)^\top\right]^\top \in \mathbb{R}^{n-1+O(s)}$ is the set representing the potential two- and three-body interactions of the ith node, $\boldsymbol{A}_i = \left[\left(\boldsymbol{A}_i^{(2)}\right)^\top, \left(\boldsymbol{A}_i^{(3)}\right)^\top\right]^\top$, $\boldsymbol{A}_i^{(2)} = \left[\sigma_2\left(a_{i,1}^{(2)}, \ldots, a_{i,i-1}^{(2)}, a_{i,i+1}^{(2)}, \ldots, a_{in}^{(2)}\right)\right]^\top \in \mathbb{R}^{n-1}$ represents two-body interactions of the ith node, and $\boldsymbol{A}_i^{(3)} \in \mathbb{R}^{O(s)}$ represents the three-body interactions, the examples of which are given in Fig. 1. Then, the approximate two- and three-body interaction matrices $\boldsymbol{A}^{(2)}$ and $\boldsymbol{A}^{(3)}$ can be derived. Update $\hat{\boldsymbol{A}}^{(2)}$ with $\boldsymbol{A}^{(2)}$, generate potential three-body interactions, solve to obtain $\boldsymbol{A}^{(2)}$ and $\boldsymbol{A}^{(3)}$, and iterate these steps until the termination condition is met, e.g., reaching the maximum number of iterations or achieving convergence of the reconstructed result. The pseudocode and schematic diagram of our stepwise reconstruction scheme are provided in Algorithm 1 and Fig. 1, respectively.

Additionally, the reason for using an iterative loop is as follows: Since the network exhibits the property of weaker or fewer higher-order connections, according to the proposed algorithm, the influence of higher-order structures on dynamic evolution is relatively small. However, considering the inevitable influence of higher-order structures, the initially obtained approximate two-body interaction matrix $\hat{\boldsymbol{A}}^{(2)}$ [Fig. 1(d)] is relatively accurate but not necessarily completely accurate. By incorporating the screened potential higher-order interactions and solving the equations [Figs. 1(e) and 1(f)],

ALGORITHM 1. Stepwise reconstruction scheme.

Input: Time-series data $\boldsymbol{X}$, self-dynamics f_i $(i = 1, \ldots, n)$, and interaction forms $g_i^{(2)}, g_i^{(3)}$ $(i = 1, \ldots, n)$;
Output: Adjacency tensors $\boldsymbol{A}^{(2)}$ and $\boldsymbol{A}^{(3)}$;
for $i = 1, \ldots, n$ **do**
  Compute $\boldsymbol{y}_i$;
  Construct the potential two-body interactions $\mathcal{G}_i^{(2)}$;
  Substitute $\boldsymbol{X}$ into $\mathcal{G}_i^{(2)}$ to obtain $\boldsymbol{G}_i^{(2)}$;
  Solve $\hat{\boldsymbol{A}}_i^{(2)}$ in Eq. (5);
end
while *the termination condition is not satisfied* **do**
  for $i = 1, \ldots, n$ **do**
    Construct $\mathcal{G}_i^{(3)}$ based on $\hat{\boldsymbol{A}}_i^{(2)}$;
    Update $\mathcal{G}_i = [(\mathcal{G}_i^{(2)})^\top, (\mathcal{G}_i^{(3)})^\top]^\top$;
    Substitute $\boldsymbol{X}$ into $\mathcal{G}_i$ to obtain $\boldsymbol{G}_i$;
    Solve $\boldsymbol{A}_i = [(\boldsymbol{A}_i^{(2)})^\top, (\boldsymbol{A}_i^{(3)})^\top]^\top$ in Eq. (6);
  end
  Update $\hat{\boldsymbol{A}}^{(2)}$ with $\boldsymbol{A}^{(2)}$;
end

the identified two-body interactions are theoretically more accurate than when only considering two-body interactions, as the potential interaction space includes some three-body interactions (although it may not include all real three-body interactions). Thus, updating $\hat{\boldsymbol{A}}^{(2)}$ with the more accurate two-body interaction matrix $\boldsymbol{A}^{(2)}$ [Fig. 1(g)], we can then construct potential three-body interactions again. This time, the screened potential three-body interactions will theoretically be more accurate, and the identified results will be more precise [Figs. 1(e) and 1(f)]. Ideally, the identification performance will gradually improve by iterating several times until all the interactions are successfully reconstructed or the termination condition is satisfied.

In general, compared to the direct strategy, which requires modeling all possible interactions, the stepwise strategy can effectively reduce the number of parameters required to be modeled due to the screening effect of the lower-order dependency property on potential interactions. Consequently, the time and space complexity of our proposed reconstruction strategy is comparatively lower. For a detailed analysis and comparison of the algorithm complexity between the direct and stepwise strategies, please refer to Appendix C. Additionally, since the network reconstruction problem is transformed into an equation-solving problem, solving the equations can become exceedingly challenging when numerous interactions need to be modeled. Therefore, besides reducing computational resources, our method also enables the resolution of problems that were previously intractable due to the extensive number of interactions that required modeling.

III. NUMERICAL SIMULATIONS

In this section, we verify the performance of our stepwise method through various simulation experiments and provide further applications. In the subsequent simulations, our stepwise reconstruction approach assumes prior knowledge of node degrees, incorporating higher-order interactions only when all corresponding lower-order interactions exist. For example, $g_{123}^{(3)}$ is included in $\mathcal{G}_1^{(3)}$ only if $g_{12}^{(2)}, g_{13}^{(2)}$, and $g_{23}^{(2)}$ are all considered present. Additionally, considering the nature of network reconstruction using a stepwise strategy, the identified approximate lower-order interaction matrix may exhibit non-symmetry. For example, node 1 might be connected to node 2, yet node 2 may not reciprocate this. In the context of undirected networks, before screening higher-order interactions that need to be modeled, we will symmetrize the approximate lower-order interaction matrix, potentially allowing us to include both existent and non-existent interactions within the search space as much as possible, thereby improving the accuracy and robustness of the reconstruction. The termination condition for the stepwise strategy, as outlined in Algorithm 1, is set to reach the maximum number of iterations set at 30. The reconstruction performance is assessed using the F1 score,[44] calculated as $\mathrm{F1} = 2\mathrm{PR}/(\mathrm{P} + \mathrm{R})$, where $\mathrm{P} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP})$ and $\mathrm{R} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$. Here, TP, FP, and FN represent the counts of true positives, false positives, and false negatives, respectively. The F1 score ranges from 0 to 1, with higher values signifying superior accuracy. An F1 score of 1 indicates a perfect reconstruction of the original network.

A. Our method's performance

1. Our stepwise method vs direct approach

To test the performance of our method, we will compare our approach (Stepwise) with the direct method (Direct) in two typical scenarios: c1 and c2. Case 1 (c1) refers to networks with properties of "lower-order dependency" and "weaker or fewer higher-order connections," while case 2 (c2) denotes networks deviating significantly from these properties. For detailed simulation parameters, please refer to the caption of Fig. 2. It is worth noting that the focus of the comparison lies in contrasting the direct and stepwise strategies, rather than the equation-solving methods (such as the equation-solving method for higher-order networks proposed by Federico Malizia *et al.*). In fact, our stepwise strategy is a plug-and-play module independent of the specific equation-solving method used. From this perspective, even the equation-solving method proposed by Federico Malizia *et al.* could be used within our framework. However, considering that their method is primarily designed for underdetermined systems—where the amount of collected data is less than the number of potential interactions—for a fair and strategy-focused comparison (detailed reasons can be found in Appendix E), unless otherwise specified, we will focus on Eq. (4) (Direct) and Eqs. (5)–(6) (Stepwise) being overdetermined, meaning that the amount of collected data exceeds the number of potential interactions. Accordingly, the sequential thresholded least-squares algorithm,[45] a classic method for overdetermined systems of equations, is employed as the equation-solving method for both the direct and stepwise strategies.

The simulation results are depicted in Fig. 2. It can be observed that the performance of the direct method is similar in both case 1

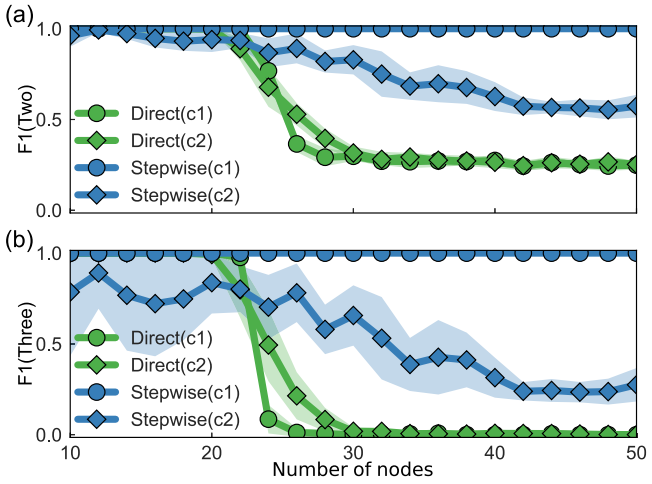

FIG. 2. Comparison of reconstruction performance between direct and stepwise approaches. Data are from random networks of Kuramoto oscillators (Appendix A) with a maximum interaction order of 3. The two-body structure is ER with an edge probability of 0.2. Denote ρ_3 ($\rho_3 \in [0, 1]$) as the probability that three-body interactions are independent of two-body interactions: If $\rho_3 = 0$, the existence of three pairwise interactions is a necessary condition for the potential presence of corresponding three-body interaction. If $\rho_3 = 0.4$, then 40% of three-body interactions are completely independent of corresponding lower-order interactions. The default setting for the two-body coupling strength is $\sigma_2 = 1$. In case 1, the number of three-body edges E_3 is set to 0.2 times the number of two-body edges E_2 (fewer higher-order connections), $\sigma_3 = 0.2$ (weaker higher-order connections), and $\rho_3 = 0$ (lower-order dependency). In case 2, set $E_3 = E_2, \sigma_3 = 1$, and $\rho_3 = 0.2$, which deviate significantly from "lower-order dependency" and "weaker or fewer higher-order connections" properties. The plotted points are the average of 10 independent experiments, where initial state values and network topology are randomly generated. The lines represent the means, and shading indicates standard deviation across ensembles of experiments. Dynamic data are collected from 20 runs with different initial values in each experiment. The duration of each run is 1 unit, and the sampling interval is 0.01. With a fixed amount of collected dynamic data, the reconstruction performance for (a) two-body (F1 Two) and (b) three-body (F1 Three) interactions gradually decreases as the network size increases.

and case 2. With a given amount of observation data, as the network size increases, the reconstruction performance of the direct method is excellent at smaller scales with 10–20 nodes, but sharply declines at 20 nodes, then reaches a very poor level at 30 nodes. This suggests that the direct method is unable to overcome the dimensionality curse in higher-order structure reconstruction. The network in case 1 is perfectly reconstructed by our stepwise method, demonstrating our method's effectiveness. In case 2, due to significant deviations from the two properties, the performance of our approach gradually declines as the network size increases. However, overall, our method still demonstrates acceptable performance compared to the direct method, especially when the network size is larger (25–50 nodes), where our method exhibits better reconstruction performance compared to the direct method. In addition, in the stepwise method, the F1 score of three-body interactions is generally lower than that of two-body interactions because our method depends on accurately identifying lower-order interactions to select potential higher-order ones. Thus, reconstructing higher-order structures becomes more challenging if lower-order reconstruction performance is not satisfactory. Overall, our method excels when the network fully satisfies the two properties, far surpassing the direct method. In cases where these two properties are severely violated, the direct method performs better for the case with fewer nodes, but ours remains viable; as the number of nodes increases, our method outperforms the direct method.

2. Further tests

We further analyze the influence of two properties ("lower-order dependency" and "weaker or fewer higher-order connections") on stepwise reconstruction performance. The parameter associated with the lower-order dependency property is ρ_3, and the parameters related to weaker and fewer higher-order connections are σ_3 and E_3/E_2, respectively [Fig. 3(a)]. In Figs. 3(b)–3(d), we, respectively, vary one parameter among ρ_3, σ_3, and E_3/E_2, while randomly generating the others within the lower and upper bounds corresponding to the parameters in cases 1 and 2 for generalizability. The experimental results present that as ρ_3 gradually increases, reconstruction performance deteriorates rapidly, while for σ_3 and E_3/E_2, their increase within a certain range maintains acceptable performance. This indicates that whether the

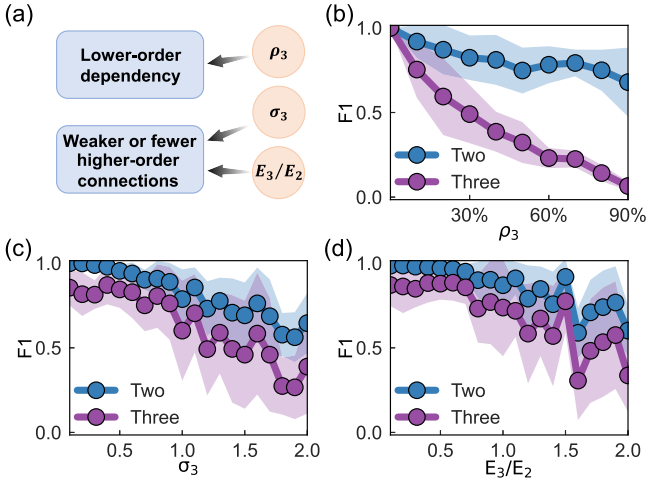

FIG. 3. The influence of the two properties ("lower-order dependency" and "weaker or fewer higher-order connections") on stepwise reconstruction performance. Data are from random networks of Kuramoto oscillators with a maximum interaction order of 3 and a fixed network size of 50 nodes. (a) A higher value of $\rho_3 \in [0, 1]$ indicates less satisfaction of the lower-order dependency property. The weaker higher-order connections property, expressed by σ_3/σ_2, with σ_2 defaulted to 1, can be characterized simply as σ_3. A higher value of σ_3 indicates less satisfaction of this property. The property of fewer higher-order connections can be described by E_3/E_2. Once the two-body edge probability is determined, E_2 remains relatively constant around a certain value, which can be approximated as a constant. Therefore, the variation of E_3/E_2 primarily depends on E_3. A higher value of E_3/E_2 indicates less satisfaction of this property. In (b)–(d), all other conditions remain the same as in Fig. 2 unless otherwise specified. (b) $\sigma_3 \sim U(0.2, 1), E_3/E_2 \sim U(0.2, 1)$. (c) $\rho_3 \sim U(0, 0.2), E_3/E_2 \sim U(0.2, 1)$. (d) $\rho_3 \sim U(0, 0.2), \sigma_3 \sim U(0.2, 1)$.

"lower-order dependency" property is satisfied has a significant impact on the performance of our method, while the presence of "weaker or fewer higher-order connections" can be relaxed to some extent.

B. More applications

Finally, we apply our method in more scenarios, including scenarios with observation noise, situations with higher node dimensions, larger-scale networks, and higher-order networks ($K = 4$). The values of parameters ρ_3, σ_3, and E_3/E_2 are randomly generated to generalize the testing scenarios, with specific values in the caption of Fig. 4. The experimental results are shown in Fig. 4. The experimental results show that our method can still effectively reconstruct networks even under 40 dB observation noise [Fig. 4(a)]. Our method is suitable for networks composed of systems with multidimensional state variables, and simulations on networks of Hindmarsh–Rose neurons illustrate its effectiveness [Fig. 4(b)]. Experiments indicate that our method can fully reconstruct networks even when the network size reaches 100 [Fig. 4(c)]. Moreover, our method can be applied to even larger-scale networks by increasing the amount of data. For higher-order networks, taking the network of Kuramoto oscillators with maximum order of four (Appendix A) as an example, our method still exhibits good reconstruction performance for interactions of all orders [Fig. 4(d)].

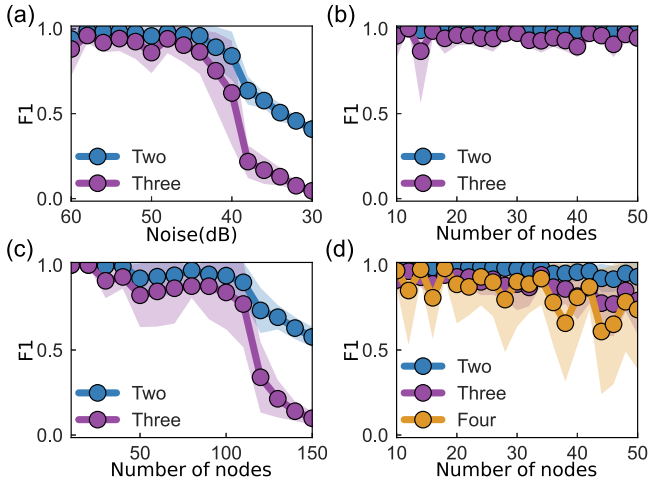

FIG. 4. More application of stepwise reconstruction scheme. Unless otherwise stated, all network and parameter settings remain consistent with those in Fig. 2. For parameters related to the two properties, $\sigma_3 \sim U(0.2, 1)$, $E_3/E_2 \sim U(0.2, 1)$, and $\rho_3 \sim U(0, 0.1)$ are set for (a) noisy scenarios with a fixed network size of 50, (b) networks of Hindmarsh–Rose neurons demonstrated in Appendix B with multidimensional state variables, and (c) larger-scale networks. In scenario (d) with a maximum interaction order of 4, fewer four-body interactions are generated within the original two-body structure type (ER networks), regardless of the edge probability. Therefore, BA networks with an average node degree of 8 are employed for the two-body interaction structure in scenario (d). Moreover, σ_3 and $\sigma_4 \sim U(0.2, 0.3)$, $E_3/E_2 = 2$ (considering that a too-small value would result in few four-body interactions), $E_4/E_2 \sim U(0.2, 1)$, and ρ_3 and $\rho_4 \sim U(0, 0.1)$, where ρ_4 represents the probability of four-body interactions being independent of three-body interactions.

IV. CONCLUSIONS

The curse of dimensionality in higher-order structure reconstruction poses a persistent challenge in network science. To reconstruct higher-order interactions, the number of interactions that need to be modeled exponentially increases, placing a significant burden on computational and data resources, thereby making it more challenging to determine which interactions truly exist. In this paper, we have developed a data-driven, stepwise reconstruction framework for higher-order networks with "lower-order dependency" and "weaker or fewer higher-order connections" properties. This approach offers a promising solution to address the dimensionality curse in higher-order structure reconstruction. Compared to the direct strategy, which models all possible interactions, our stepwise strategy effectively reduces the number of interactions that need to be modeled, resulting in lower time and space complexity correspondingly. Simulation experiments on a wide range of networks and dynamical systems have demonstrated the effectiveness and robustness of our method. It is worth noting that, although our method is primarily designed for networks with these two properties, simulation experiments indicate that even for more general networks that may not fully adhere to these characteristics, our approach still presents satisfactory performance.

For simplicity, the networks used in this paper for simulation are unweighted and undirected. Similarly, our stepwise strategy is expected to be directly applicable to weighted and directed networks by making corresponding modifications to the proposed scheme. It is important to note that in weighted networks with a significant variance in weights among interactions, smaller weights may be considered reconstruction errors stemming from temporarily ignoring higher-order interactions. Regarding the "weaker higher-order connections" property, it is essential to mention that it is difficult to reconstruct network interactions when the coupling strength between nodes is exceedingly weak.[27] Consequently, if the higher-order interaction strength is very weak, the reconstruction of higher-order structures indeed becomes very challenging. However, this does not render this property meaningless or suggest a contradiction within our stepwise scheme. Please refer to Appendix F for detailed reasons.

Additionally, our method could be optimized in the following five aspects in the future: (1) This paper assumes that the self-dynamics and interaction forms are known. In cases where the self-dynamics and interaction forms are unknown, the ARNI technique[26] may be combined with our stepwise approach to achieve model-free reconstruction. (2) Simulations assume that the degrees of nodes are known, and threshold methods[46] are possible improvements to remove the degree prior. For scale-free networks, some techniques may be used to distinguish between hubs and low-degree nodes.[27] (3) Our method heavily relies on the "lower-order dependency" property. A possible approach to relax the requirement on this property is to incorporate Bayesian or related sampling techniques to expand the range of potential interactions. (4) For the "weaker or fewer higher-order connections" property, although our approach performs acceptably even with a considerable influence of

higher-order interactions, future research aims further to improve performance under even more significant influence from higher-order interactions. A possible solution is to expand the number of potential interactions for each node gradually and even include seemingly unlikely interactions in the potential interactions. This approach aims to incorporate lower-order connections that the influence of higher-order interactions may overshadow. Ensuring algorithm stability and convergence will be a significant challenge in this process. (5) In our stepwise reconstruction method simulations, Algorithm 1 is set to terminate after 30 iterations. Ideally, the reconstruction performance would improve with constant iterations. However, empirical results demonstrate that an increase in iterations does not invariably result in better outcomes. If the initial approximate two-body interactions are inaccurate, the reconstruction quality might remain poor regardless of the iteration count. Additionally, if the optimal reconstruction is achieved after a certain number of iterations, additional iterations will not yield any further benefits. Furthermore, the identification performance may experience sudden changes or oscillations with iterations. Thus, designing an adaptive termination criterion is a future research direction. One possible approach is to use the known interactions of certain nodes as a supervisory signal and determine the terminating point based on the identification performance of these known structures.

ACKNOWLEDGMENTS

This work is supported by the Major Research Plan of the National Natural Science Foundation of China (No. 92267101), the Fundamental Research Funds for Major Program (JD) of Hubei Province (No. 2023BAA005), and the Startup grant from Shenzhen University.

NOMENCLATURE

In this paper, vectors and matrices are bolded. $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote the n- and $m \times n$-dimensional Euclidean space, respectively, and these notations can be extended and applied to higher-dimensional spaces.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yingbang Zang: Conceptualization (lead); Software (lead); Writing – original draft (lead); Writing – review & editing (lead). **Ziye Fan:** Writing – review & editing (lead). **Zixi Wang:** Writing – review & editing (equal). **Yi Zheng:** Conceptualization (equal); Writing – review & editing (equal). **Li Ding:** Writing – review & editing (supporting). **Xiaoqun Wu:** Conceptualization (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the first author upon reasonable request.

APPENDIX A: NETWORKS OF KURAMOTO OSCILLATORS

If two- and three-body interactions are considered, then the dynamics of each oscillator x_i is given by[47]

$$\dot{x}_i = \omega_i + \sigma_2 \sum_{j=1}^{n} a_{ij}^{(2)} \sin(x_j - x_i) + \sigma_3 \sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk}^{(3)} \sin(x_j + x_k - 2x_i),$$

where ω_i is the intrinsic frequency (self-dynamics) of the ith oscillator. If four-body interactions are also considered, then

$$\begin{aligned}\dot{x}_i = \omega_i &+ \sigma_2 \sum_{j_2=1}^{n} a_{ij_2}^{(2)} \sin(x_{j_2} - x_i)\\ &+ \sigma_3 \sum_{j_2=1}^{n}\sum_{j_3=1}^{n} a_{ij_2j_3}^{(3)} \sin(x_{j_2} + x_{j_3} - 2x_i)\\ &+ \sigma_4 \sum_{j_2=1}^{n}\sum_{j_3=1}^{n}\sum_{j_4=1}^{n} a_{ij_2j_3j_4}^{(4)} \sin(x_{j_2} + x_{j_3} + x_{j_4} - 3x_i).\end{aligned}$$

APPENDIX B: NETWORKS OF HINDMARSH–ROSE NEURONS

Let $\boldsymbol{x}_i = [x_i, y_i, z_i]^\top$. Consider Hindmarsh–Rose neurons with two-body interaction $\boldsymbol{g}^{(2)}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left[\tanh\left(\frac{x_j - x_i}{0.5}\right), 0, 0\right]^\top$ and three-body interaction $\boldsymbol{g}^{(3)}(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k) = \left[\tanh\left(\frac{x_j + x_k - 2x_i}{0.5}\right), 0, 0\right]^\top$, the dynamics of the ith node $\boldsymbol{x}_i$ is given by[48]

$$\begin{aligned}\dot{x}_i &= y_i + 3x_i^2 - x_i^3 - z_i + I + \sigma_2 \sum_{j=1}^{n} a_{ij}^{(2)} \tanh\left(\frac{x_j - x_i}{0.5}\right)\\ &+ \sigma_3 \sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk}^{(3)} \tanh\left(\frac{x_j + x_k - 2x_i}{0.5}\right),\\ \dot{y}_i &= 1 - 5x_i^2 - y_i,\\ \dot{z}_i &= -rz_i + rs(x_i + 1.6).\end{aligned}$$

In simulation, we set $I = 10$, $r = 3$, and $s = 2$.

APPENDIX C: ALGORITHM COMPLEXITY ANALYSIS

This section will analyze and compare the algorithm complexity of the direct and stepwise strategies. Since this paper uses undirected networks to present our method and does not provide strategies for filtering higher-order interactions in directed networks, the following analysis will focus on undirected networks. Additionally, while some steps in our stepwise strategy involve aggregating information from all nodes, the computational resources required are negligible. Moreover, all other steps of both the stepwise and direct strategies can be executed in parallel for each node. Therefore, the following complexity analysis will be based on a single node (taking node i as an example). Unless otherwise specified, the total complexity mentioned hereafter refers to that for node i.

In the main text, both the theoretical and simulated aspects of the direct and stepwise strategies are based on symmetric dynamics, i.e., $g_i^{(k)}(x_i, x_{j_2'}, \ldots, x_{j_k'}) \equiv g_i^{(k)}(x_i, x_{j_2''}, \ldots, x_{j_k''})$ for any $x_{j_2}, \ldots, x_{j_k}$, where $j_2', \ldots, j_k'$ and $j_2'', \ldots, j_k''$ are numerical reorderings of $j_2, \ldots, j_k$ (here, $i, j_2, \ldots, j_k$ are not equal to each other). If the symmetric dynamics condition is not met, the direct and stepwise methods can still be employed by adding the corresponding potential higher-order interactions that need to be modeled. This adjustment does not fundamentally affect the analysis and comparison of the two strategies. To illustrate this point, we will also provide an analytical perspective on symmetric and non-symmetric dynamics in the algorithm complexity analysis of direct and stepwise strategies. Moreover, the algorithm complexities can be considered approximately proportional to the number of modeled interactions for both the direct and stepwise methods. Thus, in the following, we will first analyze the number of potential interactions that need to be modeled for the direct and stepwise strategies from both symmetric and non-symmetric dynamics perspectives and then evaluate the time and space complexities.

Let K denote the highest possible interaction order. Since not all orders may be modeled, let $\mathcal{K}$ represent the highest order analyzed, and we have $3 \le \mathcal{K} \le K \le n$. To simplify notation for subsequent analysis, denote $\mathcal{K}' = \mathcal{K} - 1$.

1. Analysis of the direct strategy

Under non-symmetric dynamics condition: Considering that the real-world maximum interaction order K is typically smaller relative to number of nodes n, for the $(k+1)$-body interaction, the number of potential interactions to be modeled is $\prod_{\tilde{k}=1}^{k}(n-\tilde{k}) \approx n^k$. Thus, the total number

$$S_{direct}^{non} = \sum_{k=1}^{\mathcal{K}'} n^k,$$

with the maximum analyzed order $\mathcal{K}$. Accordingly, we have

$$n^{\mathcal{K}'} \le \sum_{k=1}^{\mathcal{K}'} n^k \le 2n^{\mathcal{K}'}.$$

Under symmetric dynamics condition: The analysis of symmetric dynamics scenarios is similar to that of non-symmetric scenarios. The difference lies in that under symmetric conditions, certain interactions have equivalent impacts on node i, and in undirected networks, the presence or absence of these interactions among corresponding nodes is uniform—either all are present or all are absent. Therefore, these interactions can be modeled collectively as a single interaction. For example, as $g_i^{(k)}(x_i, x_{j_2'}, \ldots, x_{j_{k+1}'}) \equiv g_i^{(k)}(x_i, x_{j_2''}, \ldots, x_{j_{k+1}''})$ for any $x_{j_2}, \ldots, x_{j_{k+1}}$, we can, therefore, model the interaction $g_i^{(k)}(x_i, x_{j_2'}, \ldots, x_{j_{k+1}'})$ alone to represent the potential $k!$ interactions. Therefore, for the $(k+1)$-body interaction, the number of potential interactions to be modeled can be approximated by $\frac{n^k}{k!}$. Then, the total number

$$S_{direct}^{sym} = \sum_{k=1}^{\mathcal{K}'} \frac{n^k}{k!},$$

and we have

$$\left(\frac{n}{K}\right)^{\mathcal{K}'} \le \frac{n^{\mathcal{K}'}}{\mathcal{K}'^{\mathcal{K}'}} \le \frac{n^{\mathcal{K}'}}{\mathcal{K}'!} \le \sum_{k=1}^{\mathcal{K}'} \frac{n^k}{k!} \le \mathcal{K}' \frac{n^{\mathcal{K}'}}{\mathcal{K}'!} \le n^{\mathcal{K}'}.$$

Based on the analysis above, we can conclude that for non-symmetric dynamics scenarios, if n is fixed and $\mathcal{K}$ increases, the number of interactions that need to be modeled grows exponentially; if $\mathcal{K}$ is fixed and n increases, the number increases polynomially, with larger values of $\mathcal{K}$ resulting in faster growth. For symmetric dynamics scenarios, the situation is analogous, but the lower bound is related to n/K instead of n. Given that K is generally much smaller than n, this difference has a relatively small impact on the scale of interactions to be modeled. In summary, for the direct strategy, regardless of whether the interaction dynamics are symmetric or not, the lower bounds on the number of interactions that need to be modeled are all of the form $\left(\frac{1}{\zeta} \cdot n\right)^{\mathcal{K}-1}$, where ζ is 1 for non-symmetric cases or K for symmetric ones. As the number of nodes n and the highest order analyzed $\mathcal{K}$ increase, the numbers will rapidly increase.

2. Analysis of the stepwise strategy

For the stepwise strategy, the higher-order interactions that need to be modeled depend on the identified lower-order interactions. Denote $E_{k,i}$ $(k = 2, \ldots, \mathcal{K})$ the number of k-body interactions identified of node i, and define $\theta_{k\to k+1}(E_{k,i})$ as the function that determines the number of $(k+1)$-body interactions to be modeled based on identified k-body interactions. Then, the total number

$$S_{stepwise} = n + \sum_{k=2}^{\mathcal{K}'} \theta_{k\to k+1}(E_{k,i})$$

with the maximum analyzed order $\mathcal{K}$. Depending on prior knowledge or specific strategies for screening potential higher-order interactions, $\theta_{k\to k+1}(E_{k,i})$ can take different forms. Next, we will provide several typical examples of $\theta_{k\to k+1}(E_{k,i})$ for both symmetric and non-symmetric dynamics scenarios. Overall, the number of higher-order interactions that need to be modeled in symmetric dynamics scenarios will be translated into a problem of selecting nodes/lower-order edges. Once the approximate k-body interactions are identified, calculating the number of $(k+1)$-body interactions to be modeled primarily requires selecting k additional nodes and some k-body edges, apart from the ith node itself. The different ways of selecting these nodes and edges correspond to different rules for screening higher-order interactions. The analysis in non-symmetric dynamics scenarios requires only slight modifications based on the symmetric scenario. Therefore, our focus is to analyze the upper bound of $\theta_{k\to k+1}(E_{k,i})$, beginning with $\theta_{k\to k+1}^{sym}$ for symmetric dynamics scenarios, and then $\theta_{k\to k+1}^{non}$ for non-symmetric dynamics scenarios.

Under symmetric dynamics condition: First, we analyze a typical strategy for screening potential higher-order interactions

from lower-order ones, where higher-order interactions are modeled only when all corresponding lower-order interactions exist. To illustrate this more clearly, we provide a specific example. For the ith node (here, $i, j_1, \ldots, j_6$ are distinct from each other), suppose $a^{(3)}_{ij_1j_2} \neq 0, a^{(3)}_{ij_1j_3} \neq 0, a^{(3)}_{ij_2j_3} \neq 0$. When screening potential four-body interactions based on the identified three-body interactions ($k = 3$), we follow these three steps:

- Select 3 three-body interactions connected to the ith node (e.g., $a_{ij_1j_2}, a_{ij_1j_3}, a_{ij_2j_3}$) from the approximate three-body interaction matrix.
- Analyze whether these three edges could potentially form a four-body interaction.
- Calculate the probability of the existence of the fourth three-body interaction $a_{j_1j_2j_3}$.

Specifically, first, the number of combinations to select three types of three-body connections from the identified three-body interaction types of the ith node is given by the combination $C^3_{E_{3,i}/(3-1)!}$, where $E_{3,i}/(3-1)!$ represents the number of available types of three-body connections. Since interactions of the same connection type in $E_{3,i}$ are counted multiple times, the number of available types is approximated by $E_{3,i}$ divided by $(3-1)!$. For example, $a_{ij_1j_2}$ and $a_{ij_2j_1}$ belong to the same type but are counted twice; thus, the actual number of selectable edge types is $E_{3,i}/(3-1)!$. For simplicity, we assume $E_{k,i}/(k-1)!$ is an integer, which essentially does not affect the subsequent analyses. Then, despite having selected three types of three-body connections associated with the ith node, it is possible that these three edges cannot form a four-body connection (e.g., $a_{ij_1j_2}, a_{ij_3j_4}, a_{ij_5j_6}$). $\mu_3 \in [0,1]$ ($\mu_k \in [0,1]$) is defined to characterize the probability of not contradicting the formation of four-body ($(k+1)$-body) interactions. Finally, we analyze the probability of the existence of the fourth three-body edge. It should be noted that although we have known whether the fourth edge exists based on the approximate three-body interaction matrix, for formal analysis, we adopt the mean-field approach, where the probability of the fourth edge's existence is approximated by the probability of each interaction type's existence, represented as $\frac{E_3/3!}{n^3/3!} = \frac{E_3}{n^3}$ (here, E_3 represents the number of identified three-body interactions in the entire network).

Combining the above three steps, we have

$$\theta^{sym}_{3\to4}(E_{3,i}) = \mu_3 \cdot C^3_{E_{3,i}/(3-1)!} \cdot \frac{E_3/3!}{n^3/3!}.$$

Accordingly, for $\theta^{sym}_{k\to k+1}$, we have $\theta^{sym}_{k\to k+1}(E_{k,i}) = \mu_k \cdot C^k_{E_{k,i}/(k-1)!} \cdot \frac{E_k/k!}{n^k/k!}$. If $k \le E_{k,i}/(k-1)!$, from Theorem 1 in Appendix D, we have $C^k_{E_{k,i}/(k-1)!} < \left(\frac{(E_{k,i}/(k-1)!)\cdot e}{k}\right)^k = \left(\frac{e}{k} \cdot \frac{E_{k,i}}{(k-1)!}\right)^k$, where e is a mathematical constant approximately equal to 2.718 28. Otherwise, if $k > E_{k,i}/(k-1)!$, then $\theta^{sym}_{k\to k+1} = 0$. Furthermore, we have

$$\theta^{sym}_{k\to k+1}(E_{k,i}) = \mu_k \cdot C^k_{E_{k,i}/(k-1)!} \cdot \frac{E_k/k!}{n^k/k!} < \left(\frac{e}{k} \cdot \frac{E_{k,i}}{(k-1)!}\right)^k \cdot \frac{E_k}{n^k}. \tag{C1}$$

The number of $(k+1)$-body interactions modeled by the direct method is $\frac{n^k}{k!}$, which increases rapidly with the number of nodes and the interaction order. Compared to the direct method, the number of interactions that need to be modeled here depends more on the number of k-body edge types of the ith node $\left(\text{i.e., } \frac{E_{k,i}}{(k-1)!}\right)$ and the sparsity of k-body edges in the entire network $\left(\text{i.e., } \frac{E_k}{n^k}\right)$. As real-world networks are generally sparse, our stepwise strategy can significantly reduce the number of $(k+1)$-body interactions that need to be modeled.

Equation (C1) is derived under the rule that higher-order interactions are modeled only when all corresponding lower-order interactions exist. Next, we can analyze other situations similarly based on the "three-step" analysis and incorporating the perspective of determining nodes. The determining nodes perspective refers to the process , where when screening for potential $(k+1)$-body interactions from k-body interactions, the primary focus should be identifying which nodes might be included in the potential $(k+1)$-body interactions. Specifically, since the ith node itself is already determined, the primary task is to identify the remaining k nodes. Denote $\tilde{k}$ ($\tilde{k} = 0, \ldots, k$) as the number of sampled k-body interactions from the identified k-body interactions when screening potential $(k+1)$-body interactions. If $\tilde{k}$ (here, $\tilde{k} = 2, \ldots, k$) k-body interactions are required to depend on the identified ones when screening for $(k+1)$-body interactions, and the rest can be chosen arbitrarily, then all $k+1$ nodes have already been determined. For example, if only two types of k-body edges are determined, each type involving $k-1$ nodes other than the ith node, and considering that at least one node differs between the two types of k-body edges, then all necessary nodes are already determined. If the determined number of nodes (including ith node) is greater than $k+1$, then this edge is contradictory and thus will not be created. As before, $\mu_k \in [0,1]$ is used to characterize the probability of not contradicting the formation of $(k+1)$-body interactions. Then, we have

$$\theta^{sym}_{k\to k+1}(E_{k,i}) = \mu_k \cdot C^{\tilde{k}}_{E_{k,i}/(k-1)!} < \left(\frac{e}{\tilde{k}} \cdot \frac{E_{k,i}}{(k-1)!}\right)^{\tilde{k}}, \quad \tilde{k} = 2, \ldots, k,$$

where only the connectivity information of the ith node itself matters, with no need to consider the sparsity of the entire network. If $\tilde{k} = 1$, then we have only determined $k-1$ nodes other than the ith node, so one more node still needs to be determined. The number of selectable nodes (precisely $n-k$) is approximately n, thus,

$$\theta^{sym}_{k\to k+1}(E_{k,i}) = \mu_k \cdot C^1_{E_{k,i}/(k-1)!} \cdot n < \frac{E_{k,i}}{(k-1)!} \cdot n,$$

where the number of modeled $(k+1)$-body interactions depends not only on the connectivity information of the ith node itself but also linearly on the total number of nodes in the network. If $\tilde{k} = 0$, then the stepwise strategy degenerates to the direct strategy, and the number of $(k+1)$-body interactions that need to be modeled will become

$$\theta^{sym}_{k\to k+1}(E_{k,i}) = \frac{n^k}{k!}.$$

Under non-symmetric dynamics condition: The analysis of $\theta^{sym}_{k\to k+1}$ has been given in the context of symmetric dynamics. Next, we analyze the number of $(k+1)$-body interactions that need to be modeled, $\theta^{non}_{k\to k+1}$, for non-symmetric dynamics scenarios. In fact, for

symmetric dynamics scenarios, $k!$ $(k+1)$-body interactions belonging to the same type are modeled as one edge, so we only need to analyze the number of types of $(k+1)$-body interactions that need to be modeled. However, in non-symmetric dynamics, interactions of the same type cannot be modeled as one edge. Apart from this, the other steps are the same as in the symmetric scenarios. Therefore, the number of $(k+1)$-body interactions that need to be modeled in non-symmetric dynamics scenarios only needs to be multiplied by $k!$ based on the results of the analysis of symmetric dynamics scenarios, and we have

$$\theta^{non}_{k\to k+1} = k! \cdot \theta^{sym}_{k\to k+1}.$$

So far, we have analyzed and compared the number of higher-order interactions that need to be modeled for both the direct and stepwise strategies. Based on the above analysis, we can conclude that for the stepwise strategy, whether with symmetric or non-symmetric dynamics, as the dependence of higher-order interactions on lower-order interactions is reduced, more higher-order interactions will need to be modeled. However, since the number of higher-order interactions that need to be modeled primarily depends on the sparsity of the network, and real networks are generally sparse, the numbers of $(k+1)$-body interactions $\theta^{sym}_{k\to k+1}$ and $\theta^{non}_{k\to k+1}$ that need to be modeled are significantly smaller compared to $\frac{n^k}{k!}$ and n^k, respectively. Accordingly, it follows that $S_{stepwise}$ is smaller than S_{direct}.

Next, we will analyze the computational complexity, including time and space complexity. For simplicity, we approximate the time complexity of solving the system of equations by the number of variables in the equations (i.e., the number of potential interactions being modeled). However, the time complexity of solving a system of equations is generally not linearly related to the number of variables (it could also be quadratic or cubic), which can lead to the calculated ratio of time complexities for the direct and stepwise strategies being smaller than the actual ratio. Nevertheless, as we will see next, even with this approximation, our stepwise strategy still demonstrates a lower time complexity. Then, the time complexity of the direct strategy can be directly approximated as the total number of interactions that need to be modeled, S_{direct}, and for that of the stepwise strategy, we have

$$\begin{aligned} &n + \mathcal{L}(n + \theta_{2\to 3}) + \cdots + \mathcal{L}(n + \theta_{2\to 3} + \cdots + \theta_{\mathcal{K}'\to\mathcal{K}}) \\ &\quad = \left[(\mathcal{K}'-1)\mathcal{L}+1\right] n + (\mathcal{K}'-1)\mathcal{L}\theta_{2\to 3} + \cdots + \mathcal{L}\theta_{\mathcal{K}'\to\mathcal{K}} \\ &\quad < \mathcal{K}\mathcal{L}\cdot S_{stepwise}, \end{aligned}$$

where $\mathcal{L}$ represents the maximum number of iterations, and $\mathcal{K}$ represents the highest order analyzed. As $\mathcal{K}\mathcal{L}$ is a constant, the time complexity of the stepwise strategy primarily hinges on $S_{stepwise}$. Therefore, consistent with the previous conclusion, when the network is sparse, the stepwise strategy exhibits significantly lower time complexity than the direct strategy. Regarding space complexity, for simplicity, the space requirements of both the direct and stepwise strategies are approximated by the number of interactions that need to be modeled. In sparse networks, the stepwise strategy also demonstrates significantly better space complexity than the direct strategy.

APPENDIX D: PROOF OF AN INEQUALITY FROM APPENDIX C

First, we state and prove Lemma 1. Then, in Theorem 1, we present the required inequality and its proof.

Lemma 1: *If k is a bounded positive integer, then $\frac{1}{k!} < \left(\frac{e}{k}\right)^k$, where e is a mathematical constant approximately equal to 2.718 28.*

Proof. Let $\{x_i, i = 1, 2, \ldots\}$ be the sequence with $x_i = \left(\frac{i}{e}\right)^i$, and we have

$$\frac{x_{i+1}}{x_i} = \frac{(i+1)^{i+1}}{i^i e} = \frac{(1+\frac{1}{i})^i}{e}\cdot(i+1). \tag{D1}$$

Let $\{y_i, i = 1, 2, \ldots\}$ be the sequence with $y_i = (1+\frac{1}{i})^i$. We can get $y_i < y_{i+1}$ and $\lim_{i\to+\infty} y_i = e$. It follows that

$$y_k < e \tag{D2}$$

since k is bounded. Combining Eqs. (D1) and (D2), we get $\frac{x_{i+1}}{x_i} < i+1$. Moreover, as $x_1 = \frac{1}{e} < 1$, we get $x_k = x_1\frac{x_2}{x_1}\cdots\frac{x_k}{x_{k-1}} < k!$, i.e., $\left(\frac{k}{e}\right)^k < k!$. Accordingly, we have $\frac{1}{k!} < \left(\frac{e}{k}\right)^k$. □

Theorem 1: *If k and $\mathcal{N}$ are positive integers with $k \le \mathcal{N}$, then*

$$\mathrm{C}^k_{\mathcal{N}} < \left(\frac{\mathcal{N}e}{k}\right)^k,$$

where $\mathrm{C}^k_{\mathcal{N}}$ denotes the combination and e is a mathematical constant approximately equal to 2.718 28.

Proof. By the definition of combinations, we have $\mathrm{C}^k_{\mathcal{N}} = \frac{\mathcal{N}!}{k!(\mathcal{N}-k)!} = \left(\frac{1}{k!}\right)\cdot\left(\mathcal{N}(\mathcal{N}-1)\cdots(\mathcal{N}-k+1)\right)$. From Lemma 1, we get $\frac{1}{k!} < \left(\frac{e}{k}\right)^k$. Moreover, since $\mathcal{N}(\mathcal{N}-1)\cdots(\mathcal{N}-k+1) \le \mathcal{N}^k$, it follows that $\mathrm{C}^k_{\mathcal{N}} < \left(\frac{e}{k}\right)^k\cdot\mathcal{N}^k = \left(\frac{\mathcal{N}e}{k}\right)^k$. □

APPENDIX E: WHY CONSIDER OVERDETERMINED SCENARIOS IN SIMULATIONS

Only overdetermined scenarios are considered for simulation primarily because of the potential variability in over-determination and under-determination when using the stepwise strategy. For instance, it might be overdetermined at the first step but become underdetermined after adding potential three-body interactions in the next step. Moreover, variations in the number of potential interactions per round during iterative processes can also lead to changes between over-determination and under-determination. While we could potentially attempt to automatically switch solving methods and configure adaptive hyperparameters based on the relative amount of data to potential interactions, the effectiveness of direct and stepwise strategies may vary due to the method used for solving and hyperparameter selection, which complicates the comparison of direct and stepwise strategies.

Additionally, a natural question arises: Can we consider only underdetermined scenarios? In fact, this is not a preferable option. On the one hand, the same amount of data may be underdetermined for the direct strategy, but it is likely overdetermined for the stepwise strategy. On the other hand, even if less data is used to ensure that both the direct and stepwise strategies are underdetermined, the stepwise method is likely slightly underdetermined, while the

direct method is highly underdetermined. Since solving underdetermined systems implicitly imposes certain requirements on the minimum amount of data, the data available for the direct strategy may consistently fall far below the minimum threshold required. In the comparison process, both of these situations disadvantage the direct strategy.

Based on the reasons stated above, all simulations are conducted in the overdetermined scenario.

APPENDIX F: NOTES ON WEAKER HIGHER-ORDER CONNECTIONS PROPERTY

It should be noted that it is difficult to reconstruct network interactions when the coupling strength between nodes is extremely weak.[27] Accordingly, if the higher-order interaction strength is very weak, the reconstruction of higher-order structures indeed becomes very challenging. Then, does this imply that the property of "weaker higher-order connections" is meaningless? In other words, our stepwise strategy requires weaker or fewer higher-order interactions to become applicable, but weaker higher-order interactions will lead to identification failure. Is this contradictory?

In fact, the "weaker higher-order connections" property is meaningful and not contradictory. The reason is as follows: First, we acknowledge that extremely weak higher-order interactions are difficult to be identified successfully. However, the aim of having weaker or fewer higher-order connections is to ensure that the overall impact of higher-order interactions (in terms of both the number of higher-order edges and the strength of higher-order interactions) is relatively smaller compared to lower-order interactions, which will facilitate more accurate initial estimation of the lower-order structure by temporarily disregarding the higher-order effects. This smaller impact of higher-order interactions can be satisfied by either weaker or fewer higher-order connections, so even if the connections are not that weak, having fewer higher-order connections would suffice. In other words, from a theoretical perspective, extremely weak higher-order connections are not strictly necessary. So, the property of weaker higher-order connections is meaningful, primarily reflecting an ideal relative trend where the strength of higher-order interactions is weaker than that of lower-order interactions. Higher-order interactions' strength does not need to be extremely weak and can even be relatively strong. Therefore, having not-so-weak higher-order interactions is theoretically feasible and can still be successfully identified, so this is not contradictory.

Moreover, as shown in Fig. 3(c), simulations also demonstrate that when the higher-order strength is one-tenth of the lower-order strength, even with a significant amount of higher-order interactions ($E_3/E_2 \sim U(0.2, 1)$), the identification performance remains highly accurate. In other words, in practice, even when the higher-order interactions are not particularly weak and their quantity is not as small as analyzed, our reconstruction can still maintain satisfactory performance. Nonetheless, as long as the higher-order interaction strength is above the identifiable threshold, networks with weaker higher-order interactions are easier to reconstruct.

REFERENCES

[1] R. Albert, H. Jeong, and A.-L. Barabási, "Diameter of the world-wide web," Nature **401**, 130–131 (1999).
[2] M. Rohden, A. Sorge, M. Timme, and D. Witthaut, "Self-organized synchronization in decentralized power grids," Phys. Rev. Lett. **109**, 064101 (2012).
[3] A. Kirkley, H. Barbosa, M. Barthelemy, and G. Ghoshal, "From the betweenness centrality in street networks to structural invariants in random planar graphs," Nat. Commun. **9**, 2501 (2018).
[4] Y. Luan, X. Wu, and B. Liu, "Maximizing synchronizability of networks with community structure based on node similarity," Chaos **32**, 083106 (2022).
[5] Z. Fan, X. Wu, B. Mao, and J. Lü, "Output discernibility of topological variations in linear dynamical networks," IEEE Trans. Autom. Control (published online, 2024).
[6] X. Wu, X. Wu, C.-Y. Wang, B. Mao, J.-A. Lu, J. Lü, Y.-C. Zhang, and L. Lü, "Synchronization in multiplex networks," Phys. Rep. **1060**, 1–54 (2024).
[7] S. Yu, M. Zhao, C. Fu, J. Zheng, H. Huang, X. Shu, Q. Xuan, and G. Chen, "Target defense against link-prediction-based attacks via evolutionary perturbations," IEEE Trans. Knowl. Data Eng. **33**, 754–767 (2021).
[8] M. Pope, C. Seguin, T. F. Varley, J. Faskowitz, and O. Sporns, "Co-evolving dynamics and topology in a coupled oscillator model of resting brain function," NeuroImage **277**, 120266 (2023).
[9] H. Liu, J.-A. Lu, J. Lü, and D. J. Hill, "Structure identification of uncertain general complex dynamical networks with time delay," Automatica **45**, 1799–1807 (2009).
[10] S. Zhang, X. Wu, J.-A. Lu, H. Feng, and J. Lü, "Recovering structures of complex dynamical networks based on generalized outer synchronization," IEEE Trans. Circuits Syst. I: Regul. Pap. **61**, 3216–3224 (2014).
[11] S. Zhu, J. Zhou, G. Chen, and J.-A. Lu, "A new method for topology identification of complex dynamical networks," IEEE Trans. Cybern. **51**, 2224–2231 (2021).
[12] Y. Zheng, X. Wu, G. He, and W. Wang, "Topology identification of fractional-order complex dynamical networks based on auxiliary-system approach," Chaos **31**, 043125 (2021).
[13] Y. Zheng, X. Wu, Z. Fan, and W. Wang, "Identifying topology and system parameters of fractional-order complex dynamical networks," Appl. Math. Comput. **414**, 126666 (2022).
[14] X. Wu, W. Wang, and W. X. Zheng, "Inferring topologies of complex networks with hidden variables," Phys. Rev. E **86**, 046106 (2012).
[15] C. Ma, H.-S. Chen, X. Li, Y.-C. Lai, and H.-F. Zhang, "Data based reconstruction of duplex networks," SIAM J. Appl. Dyn. Syst. **19**, 124–150 (2020).
[16] L. Peel, T. P. Peixoto, and M. De Domenico, "Statistical inference links data and theory in network science," Nat. Commun. **13**, 6794 (2022).
[17] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, "Predicting catastrophes in nonlinear dynamical systems by compressive sensing," Phys. Rev. Lett. **106**, 154101 (2011).
[18] X. Han, Z. Shen, W.-X. Wang, and Z. Di, "Robust reconstruction of complex networks from sparse data," Phys. Rev. Lett. **114**, 028701 (2015).
[19] G. Mei, X. Wu, Y. Wang, M. Hu, J.-A. Lu, and G. Chen, "Compressive-sensing-based structure identification for multilayer networks," IEEE Trans. Cybern. **48**, 754–764 (2018).
[20] G. Li, N. Li, S. Liu, and X. Wu, "Compressive sensing-based topology identification of multilayer networks," Chaos **29**, 053117 (2019).
[21] X. Wang, J. Lü, and X. Wu, "Recovering network structures with time-varying nodal parameters," IEEE Trans. Syst. Man Cybern.: Syst. **50**, 2588–2598 (2020).
[22] X. Wu, Z. Fan, J. He, W. Wang, and J. Lü, "Reconstruction and layer division of unknown multilayer networks," IEEE Trans. Syst. Man Cybern.: Syst. **53**, 7794–7804 (2023).
[23] Y. Zhang, Y. Guo, Z. Zhang, M. Chen, S. Wang, and J. Zhang, "Universal framework for reconstructing complex networks and node dynamics from discrete or continuous dynamics data," Phys. Rev. E **106**, 034315 (2022).
[24] M. Schmidt and H. Lipson, "Distilling free-form natural laws from experimental data," Science **324**, 81–85 (2009).
[25] M. Nitzan, J. Casadiego, and M. Timme, "Revealing physical interaction networks from statistics of collective dynamics," Sci. Adv. **3**, e1600396 (2017).
[26] J. Casadiego, M. Nitzan, S. Hallerberg, and M. Timme, "Model-free inference of direct network interactions from nonlinear collective dynamics," Nat. Commun. **8**, 2192 (2017).
[27] I. Topal and D. Eroglu, "Reconstructing network dynamics of coupled discrete chaotic units from data," Phys. Rev. Lett. **130**, 117401 (2023).

[28]L. Torres, A. S. Blevins, D. Bassett, and T. Eliassi-Rad, "The why, how, and when of representations for complex systems," SIAM Rev. **63**, 435–485 (2021).
[29]Q. Xuan, J. Wang, M. Zhao, J. Yuan, C. Fu, Z. Ruan, and G. Chen, "Subgraph networks with application to structural feature space expansion," IEEE Trans. Knowl. Data Eng. **33**, 2776–2789 (2021).
[30]J. Grilli, G. Barabás, M. J. Michalska-Smith, and S. Allesina, "Higher-order interactions stabilize dynamics in competitive network models," Nature **548**, 210–213 (2017).
[31]I. Iacopini, G. Petri, A. Barrat, and V. Latora, "Simplicial models of social contagion," Nat. Commun. **10**, 2485 (2019).
[32]J. Li, X. Wu, S. Zhong, R. Fan, J. Hu, and J. Lü, "A novel game investment model on uniform hypergraphs," IEEE Trans. Network Sci. Eng. **10**, 3480–3490 (2023).
[33]H. Schaeffer, G. Tran, and R. Ward, "Extracting sparse high-dimensional dynamics from limited data," SIAM J. Appl. Math. **78**, 3279–3295 (2018).
[34]J.-G. Young, G. Petri, and T. P. Peixoto, "Hypergraph reconstruction from network data," Commun. Phys. **4**, 135 (2021).
[35]H. Wang, C. Ma, H.-S. Chen, Y.-C. Lai, and H.-F. Zhang, "Full reconstruction of simplicial complexes from binary contagion and ising data," Nat. Commun. **13**, 3043 (2022).
[36]M. Contisciani, F. Battiston, and C. De Bacco, "Inference of hyperedges and overlapping communities in hypergraphs," Nat. Commun. **13**, 7229 (2022).
[37]S. Lizotte, J.-G. Young, and A. Allard, "Hypergraph reconstruction from uncertain pairwise observations," Sci. Rep. **13**, 21364 (2023).
[38]A. Kirkley, "Inference of dynamic hypergraph representations in temporal interaction data," Phys. Rev. E **109**(5), 054306 (2024).
[39]F. Malizia, A. Corso, L. V. Gambuzza, G. Russo, V. Latora, and M. Frasca, "Reconstructing higher-order interactions in coupled dynamical systems," Nat. Commun. **15**(1), 5184 (2024).
[40]R. Delabays, G. D. Pasquale, F. Dörfler, and Y. Zhang, "Hypergraph reconstruction from dynamics," arXiv:2402.00078 (2024).
[41]Z. Fan and X. Wu, "Identifying partial topology of simplicial complexes," Chaos **32**, 113128 (2022).
[42]L. Neuhäuser, A. Mellor, and R. Lambiotte, "Multibody interactions and nonlinear consensus dynamics on networked systems," Phys. Rev. E **101**, 032310 (2020).
[43]T.-T. Gao and G. Yan, "Autonomous inference of complex network dynamics from incomplete and noisy data," Nat. Comput. Sci. **2**, 160–168 (2022).
[44]D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation," J. Mach. Learn. Technol. **2**, 37–63 (2011).
[45]S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl. Acad. Sci. U.S.A. **113**, 3932–3937 (2016).
[46]B.-B. Xiang, C. Ma, H.-S. Chen, and H.-F. Zhang, "Reconstructing signed networks via ising dynamics," Chaos **28**, 123117 (2018).
[47]X. Wang, C. Xu, and Z. Zheng, "Phase transition and scaling in Kuramoto model with high-order coupling," Nonlinear Dyn. **103**, 2721–2732 (2021).
[48]L. V. Gambuzza, F. Di Patti, L. Gallo, S. Lepri, M. Romance, R. Criado, M. Frasca, V. Latora, and S. Boccaletti, "Stability of synchronization in simplicial complexes," Nat. Commun. **12**, 1255 (2021).