

Article

Robust Model-Free Identification of the Causal Networks Underlying Complex Nonlinear Systems

Guanxue Yang ^{1,*} , Shimin Lei ¹ and Guanxiao Yang ²¹ School of Electrical and Information Engineering, Jiangsu University, Zhenjiang 212013, China; 2222207067@stmail.ujs.edu.cn² College of Automation, Jiangsu University of Science and Technology, Zhenjiang 212100, China; dxxy@just.edu.cn

* Correspondence: gxyang@ujs.edu.cn

Abstract: Inferring causal networks from noisy observations is of vital importance in various fields. Due to the complexity of system modeling, the way in which universal and feasible inference algorithms are studied is a key challenge for network reconstruction. In this study, without any assumptions, we develop a novel model-free framework to uncover only the direct relationships in networked systems from observations of their nonlinear dynamics. Our proposed methods are termed multiple-order Polynomial Conditional Granger Causality (PCGC) and sparse PCGC (SPCGC). PCGC mainly adopts polynomial functions to approximate the whole system model, which can be used to judge the interactions among nodes through subsequent nonlinear Granger causality analysis. For SPCGC, Lasso optimization is first used for dimension reduction, and then PCGC is executed to obtain the final network. Specifically, the conditional variables are fused in this general, model-free framework regardless of their formulations in the system model, which could effectively reconcile the inference of direct interactions with an indirect influence. Based on many classical dynamical systems, the performances of PCGC and SPCGC are analyzed and verified. Generally, the proposed framework could be quite promising for the provision of certain guidance for data-driven modeling with an unknown model.



Citation: Yang, G.; Lei, S.; Yang, G. Robust Model-Free Identification of the Causal Networks Underlying Complex Nonlinear Systems. *Entropy* **2024**, *26*, 1063. <https://doi.org/10.3390/e26121063>

Academic Editors: Jiang Zhang, Peng Cui and Hector Zenil

Received: 8 October 2024

Revised: 28 November 2024

Accepted: 30 November 2024

Published: 6 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: model-free; data-driven; causal inference; Granger causality; nonlinear dynamics

1. Introduction

Grasping the structure and dynamics of a networked system is crucial for understanding the functionality and influence mechanisms of these systems, which could lay the foundation for better controlling complex networked systems. To date, advances in data science and machine learning have promoted the prosperity and development of network science. As is well known, networks are represented not only by their structures, but also by the dynamical information taking place on the structures. Network reconstruction, as an important aspect of research on complex networks, has attracted the attention of many researchers from various fields, including biological, electrical, physical, financial, and ecological sciences, the study of social networks, and so on [1–5]. Multidisciplinary studies have led to the emergence of technological means based on the understanding and analysis of big data. However, the transient evolution of real complex systems is often accompanied by high complexity and nonlinearity. There is, in actuality, a lack of unified physical models for dynamical processes from complex data, where the dynamical laws underlying the real systems are usually elusive. Hence, in view of the large amount of data and complex laws, data-driven modeling and causal discoveries from system dynamics are expected to play an important part in the era of big data.

Over the last decade, a great deal of methods have been developed to infer networks from observational datasets, such as ordinary differential equations, nonlinear dynamic modeling methods, similarity-based methods, Granger causality, and so on. Many of these

methods suppose that a model of the system dynamics is known in advance. Based on predefined or known mathematical models, the identification of undetermined parameters or the inference of network topology is relatively easy [5–8]. The methods based on these models sometimes do not work well, as most related modeling assumptions and conditions may not be satisfied and may be unavailable in real experimental studies. In linear dynamic modeling, many methods for network reconstruction only adopt a linear model to fit the data. These methods only consider the dynamical evolution of complex systems under a linear ordinary differential equation [9,10]. If the dynamic trajectories of a system vary from transients toward steady states, some of the related methods could also involve a linear regression method based on the datasets of steady states [11]. In actuality, these linear modeling methods cannot satisfy the demands and characteristics of complex nonlinear networked systems. As such, an adaptive differentiation scheme and pre-filtration step were introduced to linear ordinary differential equations, in order to tackle the challenge of inferring large-scale complex nonlinear networks, and this method was able to outperform some current state-of-the-art methods [12]. However, there is still a great deal of work to be carried out on network inference for complex nonlinear systems.

For unknown complex dynamical systems, many nonlinear dynamic modeling methods have been presented, and the nonlinear forms are different for different applications. Specifically, some works—which intend to build spaces of all possible basis functions for the subsequent model selection with sparse-promoting methods, if such nonlinear system dynamics can be sparsely represented by the proposed basis functions—have developed a novel framework to identify a dynamical system governed by nonlinear equations [13–15]. Although these methods could resort to using some basic knowledge and information about the physical laws of specific systems, it is worth mentioning that the reasonable choice of the function basis is not always explicit. Another approach determines the topology of networks based on the assumption that the network dynamics are a combination of self-dynamics and coupling dynamics, the crucial idea of which is to approximate a node's self-dynamics and coupling dynamics with complete orthogonal bases (e.g., the Legendre or Chebyshev polynomials or the Fourier basis) [16]. Furthermore, a novel model-free inference method was posed as a multivariate regression problem by systematically decomposing each unit's dynamics into pairwise, three-point, and higher-order interactions with other units in the network, as well as by introducing an explicit dependency matrix in combination with a block-orthogonal regression algorithm [17]. However, it is inevitable that this method would also require the network dynamics to be additive among multiple-order interactions and, sometimes, the system's dynamical model will not be able to decompose combinations of multiple-order interacting items. Therefore, these types of assumptions are restricted to some real fields of science and engineering.

In addition, similarity-based indicators are the basis of some popular methods in statistical science, the basic theories and technologies of which are correlation coefficients, mutual information, transfer entropy, and so on. They are usually used to extrapolate the causality and correlation among datasets. For pairwise relationships, there exist correlation coefficients and mutual information for linear and nonlinear cases, respectively. However, these methods result in indirect links when the number of nodes increases. As such, partial mutual information and partial correlation coefficients, which are designed to silence indirect relationships through the influence of common external inputs induced by other nodes in the whole network, have been proposed for this purpose [18–20]. In addition, these correlation coefficient and mutual information methods are only designed based on the consideration of the correlation, which cannot be used for causal inference. Hence, transfer entropy based on information theory was developed to uncover causality using time series data. Transfer entropy was first proposed to find the nonlinear causal interactions between bivariate time series [21]. Then, conditional transfer entropy [22,23] was put forward to eliminate the indirect influence of other variables in the multivariate case. For information theory methods, when the dimension of the variable increases, the calculation of the probability density function becomes difficult and complicated. Moreover, when used

for the general judgment of statistical dependencies, the ability of these similarity-based methods should be further improved for specialized fields.

As an equivalent companion of transfer entropy for Gaussian variables [24], Granger causality is another popular and powerful causality detection method that is usually designed to find the temporal interactions among time-course variables in complex multi-variate systems [25–28]. As a heuristic statistical prediction concept based on the vector autoregressive (VAR) model, Granger causality was originally presented in econometrics with the aim of determining whether or not the past observations of a process help to predict the future values of another process [29]. Accordingly, conditional Granger causality was introduced to distinguish direct interactions from indirect ones [30]. Then, formulations of Granger causality in the frequency domain were proposed, such as directed transfer functions, partially directed coherence, and so on [31–33]. In order to deal with nonlinear relationships, some nonlinear Granger causality methods have also been proposed based on some kernel functions and copula theory, such as kernel Granger causality [34–36] and copula Granger causality [37].

In order to make up for the deficiency of current Granger causality methods, recent works on Granger causality are still at the forefront of scientific research on network inference [38–40]. Additionally, in order to deal with the problem of small sample sizes for large-scale networks, many sparse-promoting methods can be applied to traditional Granger causality methods, such as truncated Lasso Granger causality [41], grouped Lasso graphical Granger [42], and an improved backward conditional Granger causality (mBTSCGC) [43]. Direct comparisons of many penalized regression approaches have been extensively conducted to provide guidelines for further applications [44]. However, these sparse Granger causality methods are mainly proposed based on linear VAR models. For the consideration of nonlinear modeling, group Lasso nonlinear conditional Granger causality (GLasso-NCGC) was proposed for nonlinear causal inference [45], the nonlinear formulation of which is only for certain types of systems. As the nonlinear variant of mBTSCGC, rpCGCI was further proposed, where the mBTS scheme was applied to the polynomial VAR model for Granger causality inference [46]. However, rpCGCI only focuses on model formulation up to second-order polynomial products, and the second-order polynomial approximation is not sufficient for highly complex nonlinear systems. Furthermore, LASSO regression was used in a VAR framework and state-space model to compute the information dynamics [47]. Then, artificial neural networks (ANNs) were combined with state-space models for Granger causality analysis [48].

In this study, we put forward a new and unified framework for identifying causal networks and predicting nonlinear dynamics from noisy observational datasets in complex nonlinear systems. Rather than identifying parameters of complex systems governed by predefined and known models or taking some polynomial, rational, or other basis functions as prior information for the subsequent model selection, we first propose Granger causality based on the general nonlinear dynamical framework, which is termed Polynomial Conditional Granger Causality (PCGC) and Sparse Polynomial Conditional Granger Causality (SPCGC); these not only consider the nonlinearity and causality of mutual interactions, but also reconcile the inference of direct interactions with the influence of indirect links. In particular, in our PCGC model, the polynomial approximation can be extended to multiple-order products for the demands of highly complex nonlinear systems, which are not limited to the second order. The network dynamics in our nonlinear framework are not simply additive between the self-dynamics and the coupling dynamics (pairwise, three-point, and other higher-order interactions among nodes in the network). We investigate our approach in many different types of complex nonlinear systems and verify the performance of several standard classical dynamical models, the system trajectories of which contain dynamical behaviors from transients towards steady states and from transients to periodic dynamics, in addition to non-periodic dynamical trajectories and chaotic dynamical trajectories. All of the simulation results demonstrate the effectiveness and robustness of the proposed method when there is a specific polynomial order and a proper size of observations.

2. Materials and Methods

2.1. Polynomial Conditional Granger Causality Model

The dynamical behavior of complex networked systems is governed by the interactions among a population of units (nodes). For each state variable in a complex system, the dynamical response or evolution is determined by the driving variables on the right side of the equation.

In general, nonlinear dynamical systems can be described using the following general differential equations:

$$\begin{cases} \dot{x}_1 = f_1(x_1, x_2, \dots, x_n) \\ \dot{x}_2 = f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ \dot{x}_n = f_n(x_1, x_2, \dots, x_n) \end{cases}, \quad (1)$$

where x_i is the state variable and $i = 1, 2, \dots, n$; n is the dimension of the system; $\dot{x}_i$ describes the time derivative of x_i ; and the nonlinear function $f_i(\bullet)$ represents the dynamical law of the system.

In detail, for each target variable x_i , Equation (2) can be obtained as follows:

$$\dot{x}_i = f_i(x_1, x_2, \dots, x_n). \quad (2)$$

Obviously, this is a unified manner of establishing a nonlinear framework for all complex networked systems. In order to investigate the causal relationship between the set of driving variables $\{x_1, x_2, \dots, x_n\}$ and the target variable x_i , the exact identification of the nonlinear function $f_i(\bullet)$ is of crucial importance. Here, we use the r -th-order polynomial function to fit the nonlinear function $f_i(\bullet)$. Generally, the nonlinear function $f_i(x_1, x_2, \dots, x_n)$ can consist of polynomials from the driving variable set $\{x_1, x_2, \dots, x_n\}$. Then, $f_i(x_1, x_2, \dots, x_n)$ can be denoted as shown in Equation (3).

$$\begin{aligned} f_i(x_1, x_2, \dots, x_n) &= \Phi \mathbf{a}_i \\ &= \sum_{k_1, k_2, \dots, k_n} a_{k_1, k_2, \dots, k_n} x_1^{k_1} x_2^{k_2} \dots x_n^{k_n}, \end{aligned} \quad (3)$$

where Φ is a vector of polynomial basis; $\mathbf{a}_i$ is an unknown parameter; $\{k_1, k_2, \dots, k_n\} \subset \mathbb{N}$; and $k_1 + k_2 + \dots + k_n \leq r$. r is the order of the polynomial. In detail, the candidate function library Φ can be expressed as follows:

$$\Phi = [1 \quad \mathbf{x}^{p_1} \quad \mathbf{x}^{p_2} \quad \mathbf{x}^{p_3} \quad \dots \quad \mathbf{x}^{p_r}], \quad (4)$$

where $\mathbf{x}^{p_r}$ can be taken to obtain the r -th-order polynomial, which includes the set of r -th-order monomials relating to $\{x_1, x_2, \dots, x_n\}$. Here, $\mathbf{x}^{p_r}$ is defined as the r -th-order polynomial vector. For example, the first-order vector and the second-order vector can be described as follows:

$$\begin{aligned} \mathbf{x}^{p_1} &= [x_1 \quad x_2 \quad \dots \quad x_n], \\ \mathbf{x}^{p_2} &= [x_1^2 \quad x_1 x_2 \quad \dots \quad x_1 x_n \quad x_2^2 \quad x_2 x_3 \quad \dots \quad x_n^2]. \end{aligned}$$

Other higher-order polynomial vectors (e.g., $\mathbf{x}^{p_3}$, $\mathbf{x}^{p_4}$, etc.) can be expressed similarly. As shown in Equation (4), when the polynomial order r is given, Φ includes all related polynomial items whose order is less than or equal to r .

Next, based on the theory of Granger causality, our nonlinear causal index is proposed for a nonlinear dynamical system as follows. For Equation (2), the complete set of all variables in the system is denoted as $\Gamma = \{x_1, x_2, \dots, x_n\}$. For each target variable x_i , the unrestricted model can be described as shown in Equation (5).

$$\dot{x}_i = f_i(\Gamma) + \xi_i. \quad (5)$$

In order to judge which driving variable x_j in Γ causally influences the target variable x_i , the restricted model can be written as shown in Equation (6).

$$\dot{x}_i = f_i(\Gamma \setminus x_j) + \xi_j, \quad (6)$$

where $\Gamma \setminus x_j$ means removing the variable x_j in Γ , i.e., $\Gamma \setminus x_j = \{x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_n\}$. Accordingly, for the nonlinear function $f_i(\Gamma \setminus x_j)$, all relevant items of x_j in $f_i(\Gamma)$ are removed. Then, the residual ξ_i of the unrestricted model and the residual ξ_j of the restricted model are calculated. In this case, it is worth noting that the inference of the causal relationship between x_j and x_i is conditioned on the set of other nodes in the network $\Gamma \setminus x_j$, which is designed to silence the indirect relationships through the influence of common external inputs induced by other nodes in the entire network. The conditional variables are not just explicitly added to the driving variable, but they are latently integrated into the nonlinear function, which has already involved the driving variable.

Finally, the nonlinear causal index (PCGC) is calculated as follows:

$$\text{PCGC}_{x_j \rightarrow x_i} = \ln \frac{\sigma_j^2}{\sigma_i^2}, \quad (7)$$

where σ_i^2 and σ_j^2 are the variances in error ξ_i and ξ_j , respectively. Generally, if PCGC is greater than the given threshold value δ , this confirms that x_j causes x_i , i.e., there exists a link $x_j \rightarrow x_i$.

In theory, an unrestricted model with more parameters will always be able to fit the data at least as well as a restricted model with fewer parameters. So, an unrestricted model with lower error will result in a better fit to the data than a restricted model. Meanwhile, in order to determine whether the unrestricted model gives a significantly better fit to the data, the F-statistic hypothesis is taken when obtaining the final network. The F-statistic value should be smaller than the given significance level P_{val} . The formulation of the F-statistic hypothesis is defined as

$$F = \frac{(\|\xi_i\|^2 - \|\xi_j\|^2) / (n_i - n_j)}{\|\xi_i\|^2 / (n_i - M)}, \quad (8)$$

where $\|\xi_i\|^2$ and $\|\xi_j\|^2$ are the residual sums of squares in the unrestricted model and the restricted model, respectively. n_i and n_j are the total numbers of parameters in the unrestricted and restricted models, respectively. M is the number of observations.

In order to determine the interactions among variables in the system, all of the time histories of state variables x_i ($i = 1, 2, \dots, n$) are collected, and the derivatives $\dot{x}_i$ can be measured directly or approximated using several methods, such as the Euler and Runge–Kutta methods.

Furthermore, suppose that the data are sampled several times, and T time points are obtained at each measurement. The state matrix $\mathbf{X}$ and derivative matrix $\dot{\mathbf{X}}$ are collected as follows:

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{x}_{1,1} & \dot{x}_{2,1} & \cdots & \dot{x}_{n,1} \\ \dot{x}_{1,2} & \dot{x}_{2,2} & \cdots & \dot{x}_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_{1,T} & \dot{x}_{2,T} & \cdots & \dot{x}_{n,T} \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{n,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,T} & x_{2,T} & \cdots & x_{n,T} \end{bmatrix},$$

where $x_{i,t}$ denotes the data point of x_i that is sampled at moment t , and $\dot{x}_{i,t}$ denotes the value of the derivative $\dot{x}_i$ that is measured or estimated at moment t .

Sometimes, in order to ensure a variety of datasets, multiple measurements are carried out under different conditions (random initial values, adding perturbations, removing nodes, etc.). Here, the number of multiple measurements is set to m . Then, the total number of observations is set to $M = mT$. Then, data matrices ($\dot{\mathbf{X}}$ and $\mathbf{X}$) with $M \times n$ are collected, i.e.,

$$\dot{\mathbf{X}} = \begin{bmatrix} \dot{\mathbf{X}}^{[1]} \\ \dot{\mathbf{X}}^{[2]} \\ \vdots \\ \dot{\mathbf{X}}^{[m]} \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}^{[1]} \\ \mathbf{X}^{[2]} \\ \vdots \\ \mathbf{X}^{[m]} \end{bmatrix},$$

where, at the m -th measurement, $\dot{\mathbf{X}}^{[m]}$ and $\mathbf{X}^{[m]}$ are denoted as

$$\dot{\mathbf{X}}^{[m]} = \begin{bmatrix} \dot{x}_{1,1}^{[m]} & \dot{x}_{2,1}^{[m]} & \cdots & \dot{x}_{n,1}^{[m]} \\ \dot{x}_{1,2}^{[m]} & \dot{x}_{2,2}^{[m]} & \cdots & \dot{x}_{n,2}^{[m]} \\ \vdots & \vdots & \ddots & \vdots \\ \dot{x}_{1,T}^{[m]} & \dot{x}_{2,T}^{[m]} & \cdots & \dot{x}_{n,T}^{[m]} \end{bmatrix} \quad \text{and} \quad \mathbf{X}^{[m]} = \begin{bmatrix} x_{1,1}^{[m]} & x_{2,1}^{[m]} & \cdots & x_{n,1}^{[m]} \\ x_{1,2}^{[m]} & x_{2,2}^{[m]} & \cdots & x_{n,2}^{[m]} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,T}^{[m]} & x_{2,T}^{[m]} & \cdots & x_{n,T}^{[m]} \end{bmatrix}.$$

Next, for the target variable x_i , the matrix formulation of Equation (2) is shown in Equation (9).

$$\dot{\mathbf{X}}_i = f(\mathbf{X}) = \Theta \mathbf{a}_i + \Xi_i, \quad (9)$$

where $\dot{\mathbf{X}}_i = [\dot{x}_{i,1}^{[1]} \cdots \dot{x}_{i,T}^{[1]} \cdots \dot{x}_{i,1}^{[m]} \cdots \dot{x}_{i,T}^{[m]}]^T$ is the i -th column of $\dot{\mathbf{X}}$, which corresponds to the derivative data of variable i . Θ is a nonlinear data matrix under the polynomial basis Φ . $\mathbf{a}_i$ is the coefficient to be solved. Ξ_i is the vector of the model error.

Taking a Lorenz oscillator system as an example, a flowchart of Polynomial Conditional Granger Causality based on the results of a simulation of this classical chaotic system is illustrated in Figure 1. Firstly, a benchmark network is obtained from the equation of the Lorenz oscillator system for comparison. In order to obtain the inferred network, the identification of the whole network is decomposed into the selection of neighbors for each target node. For example, to investigate the relationship of x with all of the variables in the system (a self-loop is also considered), unrestricted and restricted regression models are jointly constructed. In detail, to explore the impact of the driving variable y on the target variable x , an unrestricted model of the complete space of polynomials in (x, y, z) and a restricted model of the subspace of polynomials in (x, z) are constructed. The corresponding dynamic trajectories of x predicted by the unrestricted model in (x, y, z) and the restricted model in (x, z) are illustrated. From the errors of regression (r_{xyz} and r_{xz}), we find that the accuracy of curve fitting decreases when the past information of y is excluded. Hence, the causal inference from y to x ($y \rightarrow x$) is found. Similarly, the corresponding dynamic trajectories of x predicted by the restricted model in (y, z) and (x, y) are also plotted. The accuracy sharply decreases when the past information of x is removed. In this case, the self-loop of x is inferred. However, high accuracy is also obtained without the past information of z , which means that z is not the cause of x . So, a subnetwork relating to target node x is formed. In the same way, the entire network is obtained, and the Granger causal strength is marked on each link by calculating the nonlinear causal index. Meanwhile, the P_{val} matrix is obtained based on the F-statistic test. The significance level P_{val} is found to be 10^{-10} . In Figure 1, we find that the inferred network perfectly matches the benchmark network, with high causal strength and high confidence.

2.2. Sparse Polynomial Conditional Granger Causality

In general, the classical least squares method can be used to deal with Equations (5) and (6). However, the solution procedure for these equations may be problematic and inexact when the number of nonlinear basis functions is larger than the number of available observations. In fact, typically, not all basis function terms can influence the target variable. Most equations of real systems are governed by only a few relevant terms (or variables) that define the dynamical behaviors of target variables, which makes the system equations sparse in a high-dimensional space of a nonlinear basis function, especially for large networked systems. Actually, the consideration of all of the basis functions (including other unrelated terms) may lead to the deterioration of prediction performance in the

whole system. In this case, Sparse Polynomial Conditional Granger Causality (SPCGC) is proposed, a flowchart of which is shown in Figure 2 based on the results of the simulation of a Lorenz 96 system with $N = 10$ nodes. A benchmark network is obtained from the system model. Both the area under the receiver operating characteristic curve (AUROC) and the area under the precision–recall curve (AUPR) are utilized as standard metrics for evaluation.

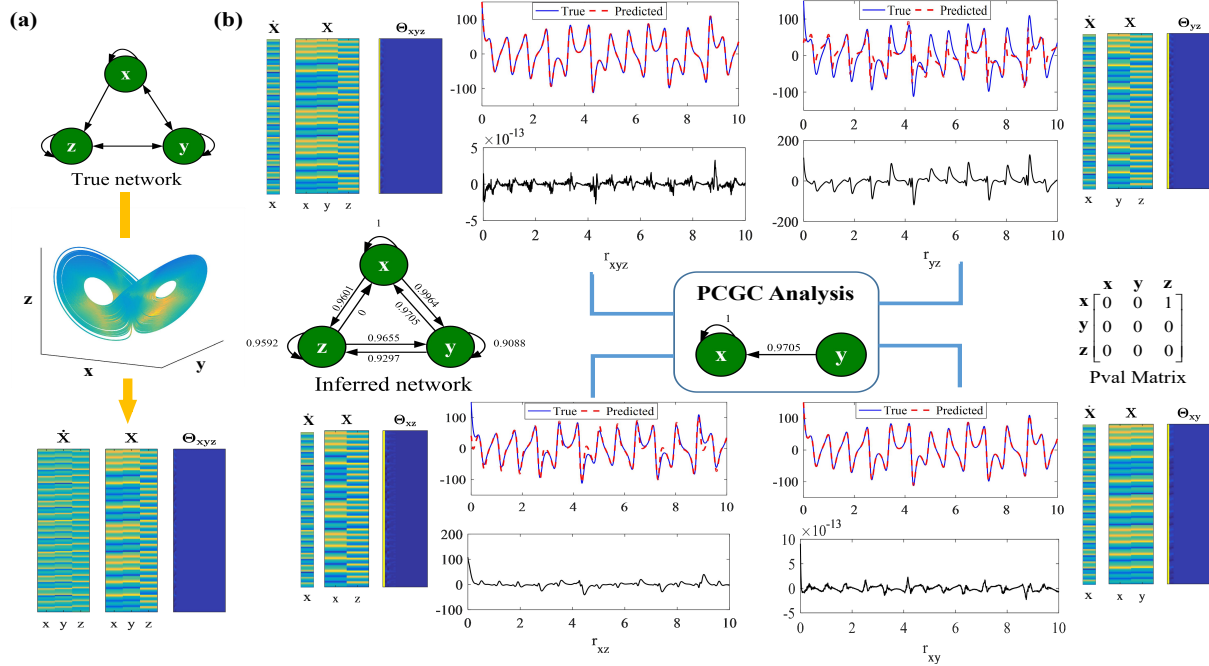


Figure 1. A flowchart of Polynomial Conditional Granger Causality based on the results of a simulation of a Lorenz oscillator system. (a) Datasets ($\dot{X}$ and X) are generated from the Lorenz system equations. The noise strength σ is set to zero. Here, $m = 1$ and $T = 2000$; that is, data are collected from the interval $t \in [0, 20]$ with the time step $\Delta t = 0.01$. Θ is a nonlinear data matrix with the polynomial basis Φ . The unrestricted model of the Lorenz system is initially constructed in the complete space of polynomials in (x, y, z) up to the fifth order, and the corresponding nonlinear data matrix is Θ_{xyz} . (b) For the target variable x , the derivative vector $\dot{x}$ is arranged to be the output variable of both the unrestricted model and the restricted model. Then, three restricted models are obtained to judge which driving variable causally influences the target variable. In detail, the restricted models of the system are constructed in the subspace of polynomials in (y, z) , (x, z) , and (x, y) up to the fifth order. The corresponding nonlinear data matrixes are Θ_{yz} , Θ_{xz} , and Θ_{xy} , respectively. The dynamic trajectories of x in different cases are plotted, and the respective errors of prediction are analyzed. The Lorenz oscillator system is shown in Equation (A1) in Appendix A.

The process of SPCGC can be divided into two steps. Datasets ($\dot{X}$ and X) are generated from the system model. Based on the datasets of X and the polynomial order r , a nonlinear data matrix Θ is obtained with the candidate basis function library Φ . For each target variable i , the regression model underlying the derivative $\dot{X}_i$ and nonlinear data matrix Θ is first formulated, and the coefficient $\mathbf{a}_i$ of this can be solved using the sparse optimization shown in Equation (10).

$$\hat{\mathbf{a}}_i = \arg \min_{\mathbf{a}_i} \left(\|\dot{X}_i - \Theta \mathbf{a}_i\|_2^2 + \lambda \|\mathbf{a}_i\|_1 \right), \quad (10)$$

where λ is a sparse penalty parameter. Based on the results of Equation (10), the basis function terms corresponding to the nonzero elements in the vector of coefficient $\mathbf{a}_i$ are fetched to form a new set of basis functions. Then, we obtain the candidate variable sets for each $\dot{X}_i$ according to the new set of basis functions. Next, we rearrange Θ with the

candidate variable sets, which are expressed as the new nonlinear data matrix Θ'_i , and reform the regression model of the PCGC model. Finally, we execute Granger causality analysis with the F-test in terms of the given significance level and confirm the causal variables of the target variable i . Another subnetwork can be inferred in a similar way. As a result, the whole network can be identified. The detailed procedure of the SPCGC algorithm can be illustrated as follows.

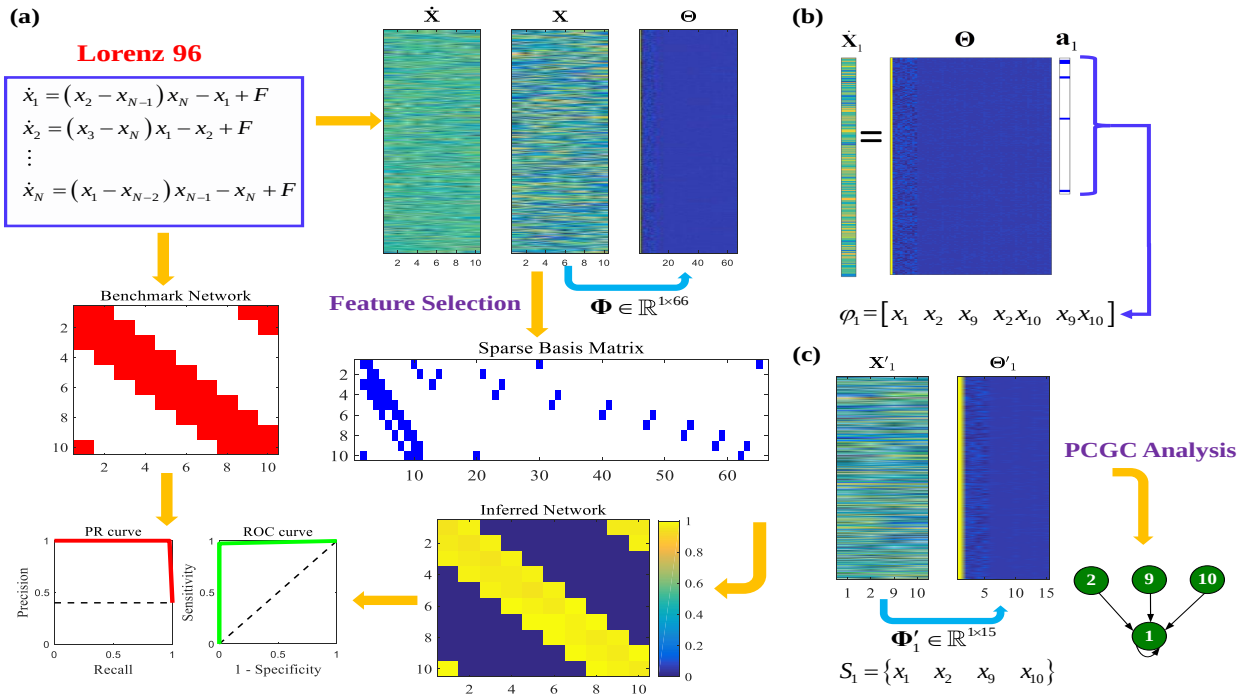


Figure 2. A flowchart of Sparse Polynomial Conditional Granger Causality based on the results of the simulation of a Lorenz 96 system with $N = 10$. Here, $m = 300$, $T = 3$, $r = 2$, $\lambda = 0.001$, and $P_{val} = 10^{-10}$. **(a)** Datasets ($\tilde{X}$ and X) are generated from the Lorenz 96 system. Based on X , Θ under $\Phi \in \mathbb{R}^{1 \times 66}$ is obtained. Then, feature selection based on sparse optimization is executed to select the contributing items in Φ , and a sparse basis matrix is acquired. Next, with the analysis of PCGC, the inferred network is compared with the benchmark network in terms of the PR and ROC curves. **(b)** Taking node no. 1 as an example, the regression equation underlying $\dot{X}_1$ and Θ is solved through sparse optimization. The nonzero elements in the vector of coefficient a_1 are fetched to form φ_1 . **(c)** Then, the corresponding set of nodes S_1 is refined. Based on S_1 , a nonlinear data matrix Θ'_1 is obtained from X'_1 with the new basis function library $\Phi'_1 \in \mathbb{R}^{1 \times 15}$. The new regression equation underlying $\dot{X}_1$ and Θ'_1 is constructed. As a result, the final subnetwork of node no. 1 is judged using the PCGC analysis.

Step 1: Based on Equation (2), a feature selection method is first used to select the columns of Θ for dimensionality reduction. Actually, this is equivalent to selecting the candidate set of polynomial items from Φ . Here, the feature selection method is formulated based on the Lasso optimization in Equation (10). The candidate set of polynomial items C corresponds to the nonzero elements in the vector of coefficient a_i .

Step 2: Based on the candidate set of polynomial items C , the corresponding driving variables are denoted as D_j , $j = 1, 2, \dots, K$, where K is the number of driving variables fetched from C . The polynomial items related to D_j in C are denoted as C^{D_j} .

Step 3: Based on the set of driving variables $D = \{D_1, D_2, \dots, D_K\}$, the corresponding unrestricted model can be reformulated. For each target variable x_i , let $y_i = \dot{X}_i$; the corresponding unrestricted model is shown in Equation (11).

$$y_i = \Theta_D a_i + \Xi_i, \quad (11)$$

where Θ_D is the nonlinear data matrix with the polynomial basis of driving variables D . Then, the residual Ξ_i can be calculated as follows:

$$\Xi_i = y_i - \Theta_D \Theta_D^\dagger y_i, \quad (12)$$

where the pseudo-inverse $\Theta_D^\dagger = (\Theta_D^T \Theta_D)^{-1} \Theta_D^T$.

Step 4: Then, the restricted model should be formulated as shown in Equation (13).

$$y_i = \Theta_{D \setminus D_j} a'_i + \Xi'_i, \quad (13)$$

where $D \setminus D_j = \{D_1, D_2, \dots, D_{j-1}, D_{j+1}, \dots, D_K\}$, and $D \setminus D_j$ means removing all of the related polynomial items C^{D_j} in the candidate set C . Then, $\Theta_{D \setminus D_j}$ is constructed based on the basis $C^{D \setminus D_j}$. So, for each D_j , the residual Ξ'_i can be calculated as follows:

$$\Xi'_i = y_i - \Theta_{D \setminus D_j} \Theta_{D \setminus D_j}^\dagger y_i, \quad (14)$$

where the pseudo-inverse $\Theta_{D \setminus D_j}^\dagger = (\Theta_{D \setminus D_j}^T \Theta_{D \setminus D_j})^{-1} \Theta_{D \setminus D_j}^T$.

Step 5: In order to judge the causal relationship between node D_j and node i , the PCGC index is calculated as follows:

$$PCGC_{D_j \rightarrow x_i} = \ln \frac{\sigma^2(\Xi'_i)}{\sigma^2(\Xi_i)}. \quad (15)$$

If $\sigma^2(\Xi'_i) > \sigma^2(\Xi_i)$, i.e., removing the effect of D_j will increase the error variance of the prediction of x_i , there exists a link $D_j \rightarrow x_i$. A positive threshold $0 < \delta < 1$ can be set for the extraction of causal links with larger Granger causal weights. In addition, F-statistic tests can be performed to confirm the significance level for each link.

2.3. Selection of the Polynomial Order r and Sparse Penalty Parameter λ

As is well known, for the determination of models that provide better predictions, overfitting and generalization are critical factors for judging which model performs better. Cross-validation and regularization are two classical ways of restraining the complexity of a model. In our work, the performance of our proposed methods (PCGC and SPCGC) is also based on the selection of the following parameters: the polynomial order r and the sparse penalty parameter λ , respectively. Although these parameters can be selected according to practical experience, an automatic selection scheme should be introduced and analyzed.

For PCGC, in order to ensure the fitting precision of nonlinear system equations and avoid overfitting, a proper polynomial order r is vital for dynamical prediction and Granger causality analysis. Here, K -fold cross-validation is adopted. Firstly, a cross-validation partition is created for data, and this can be used to define a specified size of training sets and test sets for validating the effectiveness of a statistical model based on cross-validation. In detail, all M observations are divided into K disjoint subsamples (folds) using a random partition with a roughly equal size. Then, through randomly repeated cross-validations, we minimize the cost function of the error to determine the best model.

In addition, for the simulation of SPCGC, the selection of the sparse penalty parameter λ in the regularization is important. Here, the Hannan–Quinn criterion (HQC) [49] can be used for the selection of the sparse penalty parameter. In detail, the HQC is adopted to balance the model complexity and model accuracy and to seek the best model's ability to describe datasets. For our case, based on the basic idea of the HQC, its detailed definition for the selection of λ is shown as follows:

$$HQC(\lambda) = \ln |\sigma_\lambda^2| + \frac{2n_\lambda \ln(\ln M)}{M}, \quad (16)$$

where M is the size of the data, n_λ is the number of variables, and σ_λ^2 is the variance in the error. n_λ and σ_λ^2 can be computed for each value of λ to search for the optimal solution.

3. Results

In order to verify the performance of our proposed methods, different nonlinear system models were collected and employed, including dynamical trajectories displaying transients towards steady states, dynamical trajectories exhibiting transients toward periodic dynamics, non-periodic dynamical trajectories, and chaotic dynamical trajectories [17].

These different nonlinear systems were categorized and verified according to their different types of dynamical trajectories. For dynamical trajectories from transients towards steady states, we employed a gene regulation network model with Michaelis Menten kinetics [50] and a metabolic network model (yeast glycolysis) [15]. For dynamical trajectories from transients to periodic dynamics, we simulated another two biological models: another metabolic network model (glycolytic oscillator) [13,17,51–53] and a biochemical reaction network model [14]. For non-periodic dynamical trajectories, we adopted two networked models: a Kuramoto oscillator and a phase-coupled oscillator [17]. For chaotic dynamical trajectories, we tested the Lorenz oscillator system and Rössler oscillator system. To verify the reconstruction performance of the larger networks, we also utilized the Lorenz 96 networked system. All of the system models are described in Appendix A.

3.1. Detailed Case Description of PCGC: Glycolytic Oscillator Network

In this section, we consider the inference of a metabolic network based on the glycolytic oscillator model. The complex formulations of the equations in this model are composed of constant, polynomial, and fractional terms. This is a classical benchmark problem for model prediction, automatic inference, and system identification. The glycolytic oscillator model is set to describe the dynamical behavior of metabolic networks. Accelerating the inference of metabolic networks would aid interventions in metabolic diseases.

First, to illustrate the effectiveness of our proposed PCGC method in the case of dynamical trajectories from transients to periodic dynamics, the following simulation is based on the nonlinear dynamical equations of a glycolytic oscillator with $n = 7$ nodes (variables). In Figure 3a, a benchmark adjacency matrix and a directed graph are drawn from the given equations of the system model. For the adjacency matrix, the red squares are existent links among nodes, and the white areas are nonexistent links.

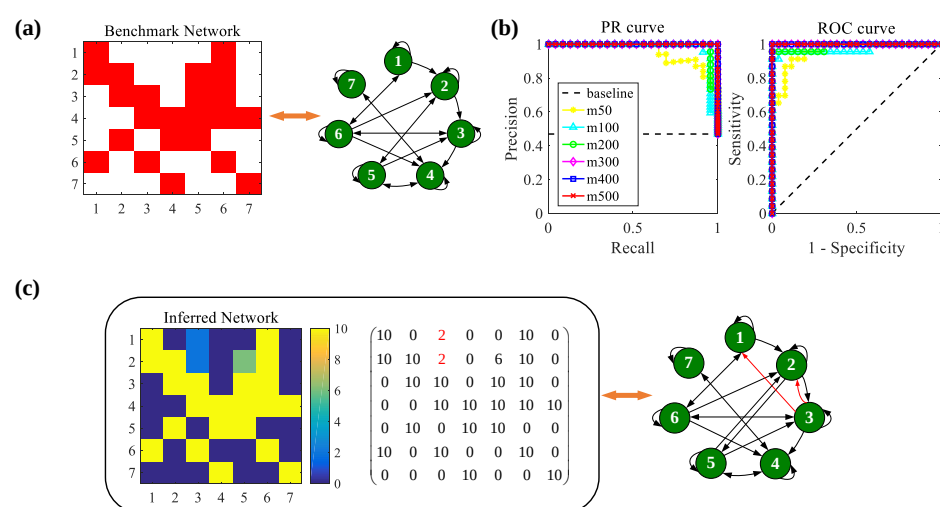


Figure 3. Simulation results for the glycolytic oscillator model. Here, $T = 2$, $r = 5$, $P_{val} = 10^{-10}$, and $\sigma = 0.01$. (a) Benchmark network together with the corresponding directed graph. (b) The curves of the ROC and PR with different numbers of samples. (c) Discovery rate matrix inferred from 10 realizations and the corresponding directed graph.

Then, synthetic datasets are derived, the related parameters of which in the simulation are shown in Figure 3. In Figure 3b, the curves of the ROC and PR are plotted with different numbers of observations. Here, the number of time points $T = 2$ is kept unchanged, and the number of measurements is taken from the set $m = \{50, 100, 200, 300, 400, 500\}$. We can see that the values of the AUROC and AUPR are relatively lower when $m = 50$. As m increases, the performance becomes better and better, and perfect reconstruction is realized when $m = \{300, 400, 500\}$. Next, the performance of robust estimation with noise is investigated. The Gaussian noise variance is set to $\sigma = 0.01$. In this case, multiple realizations are carried out, and the number of realizations is set to 10. Based on the PCGC analysis with $P_{val} = 10^{-10}$, the final network is inferred. In Figure 3c, the discovery rate matrix inferred from 10 realizations in the glycolytic oscillator model is demonstrated (the discovery rate means the total number of links that are rightly discovered over multiple realizations). It is obvious that the true positive links are always found, except for link $5 \rightarrow 2$, which has a 60% discovery rate. Meanwhile, there are two false positive links with a 20% discovery rate ($3 \rightarrow 1$ and $3 \rightarrow 2$). In Figure 3c the final directed graph is plotted from the discovery rate matrix, where the thickness of the line indicates the value of the discovery rate. In general, the results as a whole demonstrate the robust performance of the PCGC method.

3.2. Comparison of PCGC and SPCGC: Gene Regulatory Network

As mentioned above, PCGC is a method based on a nonlinear regression model. If the number of nonlinear basis functions is greater than the number of observations, the regression equation for Granger causality analysis is an underdetermined problem, and the exact solution will not be obtained. In order to verify the validity and necessity of SPCGC, a comparison between PCGC and SPCGC is made by executing a simulation of a gene regulatory dynamical system with Michaelis Menten kinetics. We first apply our proposed methods to reconstruct a gene regulatory network with $N = 10$ nodes. The generation of a benchmark network follows the mechanism using an approximate power-law output degree distribution with an average degree of $\langle k \rangle = 3$. Then, synthetic datasets with $M = mT$ observations are obtained. Here, the time points are kept unchanged at each measurement ($T = 2$). In order to investigate how the number of observations influences the performance of our method, the number of multiple measurements m is set to 400 and 700. Other related parameters are given as follows: Gaussian noise strength $\sigma = 0.01$, the polynomial order $r = 5$, and the sparse penalty parameter $\lambda = 0.001$.

In Figure 4a, the Granger causality matrixes of PCGC and SPCGC are demonstrated in the cases of $M = 1400$ and $M = 800$ observations, respectively. With adequate observations ($M = 1400$), the two methods could find all existent links, especially the existent links in the Granger causality matrix of SPCGC with a higher causal strength. For PCGC, compared with SPCGC, the causal strength of many existent links is relatively smaller. Meanwhile, all of the nonexistent links have a small causal strength, which can be eliminated by setting a specific threshold. In this case, both PCGC and SPCGC can achieve a perfect reconstruction. With insufficient observations ($M = 800$), PCGC obviously loses its function in reconstructing the network, as all of the links in the Granger causality matrix of PCGC have a high causal strength. However, although there are a few nonexistent links with slight causal intensity, SPCGC still maintains a good performance. In the following section, the performance of robust estimation with different numbers of samples is explored. Robust estimation is verified by running over 10 realizations. In particular, the elements in the Granger causality matrixes reconstructed using PCGC and SPCGC methods over 10 realizations are exhibited in Figure 4b. As shown in Figure 4b, both PCGC and SPCGC achieve perfect reconstruction when $M = 1400$. The existent links and nonexistent links can be divided using a given threshold. When $M = 800$, we find that PCGC largely mixes existent links with nonexistent links. This is short of a certain threshold for distinguishing existent links from nonexistent links. However, in this case, an almost perfect separation by SPCGC can also be seen when the number of samples is smaller than the number of nonlinear basis functions. The average values (standard deviations) of the AUROC and

AUPR over 10 realizations are computed and shown in Figure 4c. Overall, it is worth emphasizing that SPCGC demonstrates perfect performance with different numbers of observations, even when the number of samples is smaller than the number of nonlinear basis functions, i.e., the regression equation is underdetermined.

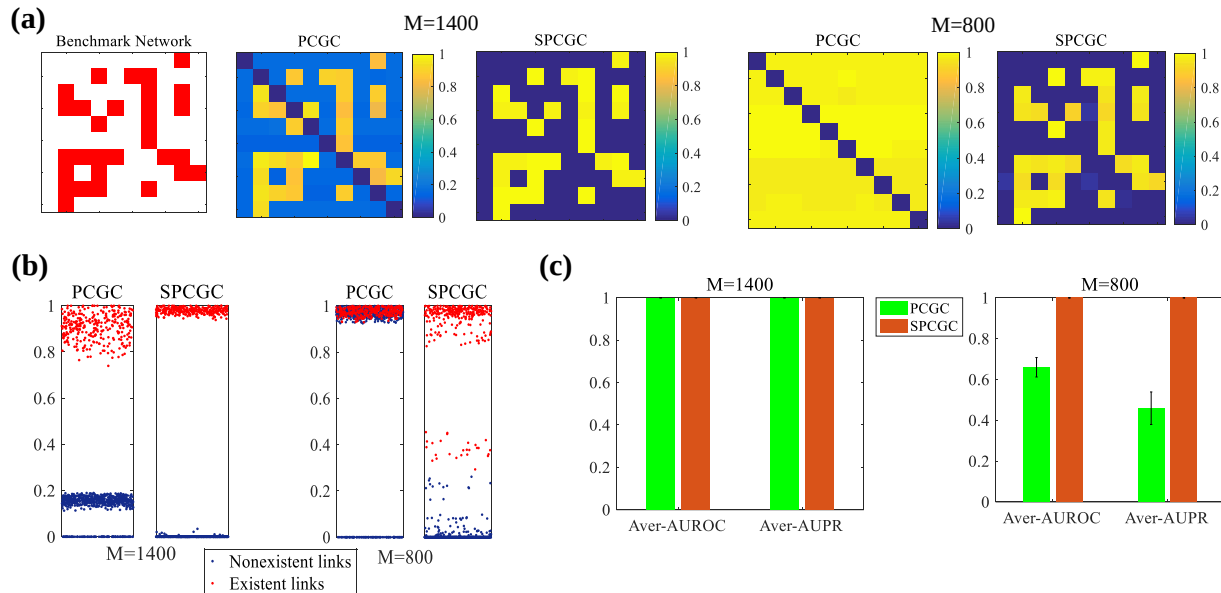


Figure 4. Simulation results for the comparison of PCGC and SPCGC in different cases by using $M = 1400$ and $M = 800$ observations. (a) The benchmark network of a gene regulation network with $N = 10$ nodes and the Granger causality index matrixes obtained by using $M = 1400$ and $M = 800$ observations. (b) Granger causality indexes of reconstructed elements in the matrixes recovered using PCGC and SPCGC over 10 realizations. (c) The average values with standard deviations of the AUROC and AUPR over 10 realizations.

3.3. Robust Results of Classical Models in Different Cases

To further assess the performance of our proposed method, both the AUROC and the AUPR were utilized to evaluate our proposed method in different classical models. These models include a biochemical reaction network, glycolytic oscillator network, yeast glycolysis network, gene regulatory network, Kuramoto oscillator network, and phase-coupled oscillator network, as well as the Lorenz 96, Lorenz, and Rössler networks.

Firstly, the simulation of different models with different numbers of observations was carried out. The detailed parameter settings of the simulations of these different models for different numbers of observations are shown in Table 1, and they include m , T , σ , and r . The values of the average AUROC and average AUPR were computed by averaging over 10 independent realizations. In detail, the inference results of different models under different conditions were explored. The different conditions involve only one parameter being changed while the others are kept the same. In Table 1, the “-” sign means that these parameters vary in the simulations. Here, a different number of observations (m or T) is used for detailed comparisons, and the specific ranges of values are shown in Figure 5. As is known, most of the benchmark network topologies can be derived from system models, except for the gene regulatory network, Kuramoto oscillator network, and phase-coupled oscillator network. The network topologies of these three models should be given in advance as the benchmark networks. Here, the preset topology follows the scale-free generative mechanism, in which the output degree of the network approximates a popular power-law degree distribution with $N = 10$ and $\langle k \rangle = 3$. For the PCGC method, the simulation results of different models with different numbers of observations are shown in Figure 5. As shown in Figure 5, the performance trend rises overall with the increase in the number of samples. As we have seen, with a certain number of samples, nearly perfect reconstruction is realized.

Table 1. Parametric settings of simulations of different models with different numbers of observations. Among these parameters, m is the number of measurements, T is the number of time points at each measurement, σ is the strength of Gaussian noise, and r is the polynomial order. Based on m and T , the total number of observations M is collected ($M = mT$). The “-” sign means that parameters vary in the simulations.

Model	m	T	σ	r
Lorenz oscillator	1	-	0.01	5
Rössler oscillator	1	-	0.01	5
Lorenz 96	-	2	0.01	4
Biochemical reaction network	-	2	0.01	4
Glycolytic oscillator	-	2	0.01	5
Yeast glycolysis	-	2	0.01	5
Gene regulatory network	-	2	0.01	4
Kuramoto oscillator	-	2	0.01	4
Phase-coupled oscillator	-	2	0.01	4

Next, in order to further address the robustness to uncertainties in data, a simulation with different noise intensities σ should be considered for the further verification of PCGC. The simulation results of different models at different noise intensities σ are shown in Figure 6. All of the values of the AUROC and AUPR are averaged over 10 realizations. The related parametric settings of the simulations in Figure 6 are listed in Table 2. The values of m , T , and r are fixed, while σ is taken from the set $\{0.01; 0.05; 0.1; 0.5; 1\}$. With the increase in σ , the performance of PCGC on these classical system models is generally stable, except for the gene regulation network and Rössler oscillator system. In the simulation of the gene regulation network and Rössler oscillator, the performance of PCGC goes down quite sharply with the increase in σ , especially for the gene regulation network. For the Kuramoto oscillator, phase-coupled oscillator, Lorenz oscillator, and Lorenz 96 model, both the average AUROC and average AUPR of PCGC over 10 realizations basically stay at the maximum value, which indicates that PCGC has strong robustness in the case of different values of σ . For the biochemical reaction network, glycolytic oscillator network, and yeast glycolysis network, the performance of PCGC slightly decreases when σ is high. In general, the robustness of PCGC is demonstrated in the case of different values of σ .

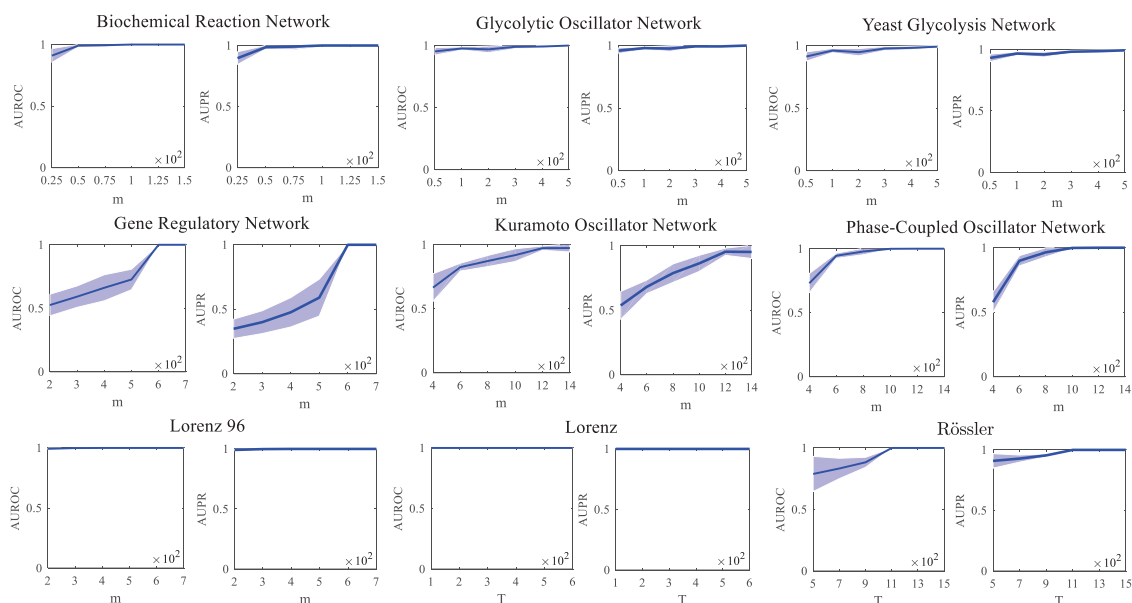
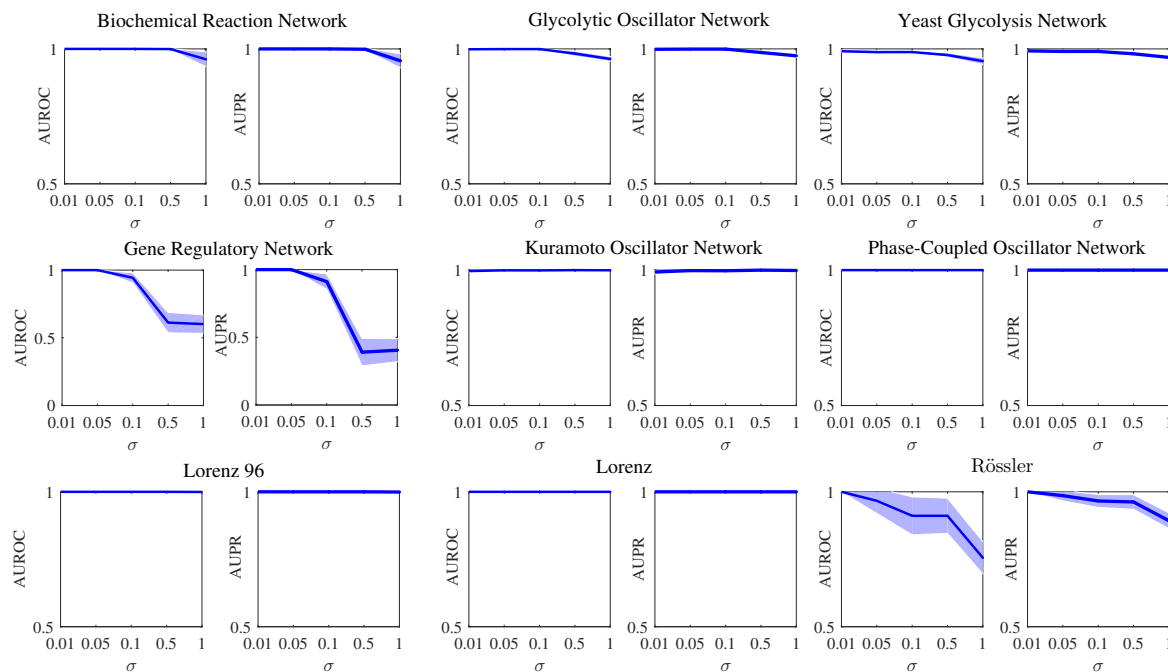


Figure 5. Simulation results of different models with different numbers of observations (m or T). All of the values of the AUROC and AUPR are averaged over 10 realizations. The blue-shaded area is the error bound of the AUROC or AUPR over 10 realizations.

Table 2. Parametric settings of simulations of different models with different noise intensities σ .

Model	m	T	σ	r
Lorenz oscillator	1	600	{0.01;0.05;0.1;0.5;1}	5
Rössler oscillator	1	1500	{0.01;0.05;0.1;0.5;1}	5
Lorenz 96	700	2	{0.01;0.05;0.1;0.5;1}	4
Biochemical reaction network	150	2	{0.01;0.05;0.1;0.5;1}	4
Glycolytic oscillator	500	2	{0.01;0.05;0.1;0.5;1}	5
Yeast glycolysis	500	2	{0.01;0.05;0.1;0.5;1}	5
Gene regulatory network	700	2	{0.01;0.05;0.1;0.5;1}	4
Kuramoto oscillator	1400	2	{0.01;0.05;0.1;0.5;1}	4
Phase-coupled oscillator	1400	2	{0.01;0.05;0.1;0.5;1}	4

Finally, in the case of a higher polynomial order with relatively insufficient samples, the nonlinear regression equation for PCGC will be underdetermined. Then, high performance will be not guaranteed. To some extent, it will even become very poor. Meanwhile, even if the number of variables (nodes) is small, the polynomial space might also be very large with a higher polynomial order. In this way, it is necessary to propose SPCGC for dimension reduction to improve the reconstruction performance. Simulations comparing the PCGC and SPCGC methods are demonstrated in Figure 7. Here, Lorenz 96, the gene regulatory network, the Kuramoto oscillator network, and the phase-coupled oscillator network are taken for further investigation; the network generation mechanism and network parameter settings are the same as those in the simulation in Figure 5. The other parameters for the simulations in Figure 7 are listed in Table 3. As shown in Figure 7, based on the average AUROC and average AUPR over multiple realizations, the performance of SPCGC is higher than that of PCGC as a whole. In particular, for the Lorenz 96 network, unlike PCGC, SPCGC has almost the same values for the AUROC and AUPR. However, the final reconstructed network of SPCGC has more exact and reliable causal weights with higher confidence levels. In this case, a similar comparison between SPCGC and PCGC can be performed by referring to the analysis in Figure 4.

**Figure 6.** Simulation results of different models at different noise intensities σ . All of the values of the AUROC and AUPR are averaged over 10 realizations. The blue-shaded area is the error bound of the AUROC or AUPR over 10 realizations.

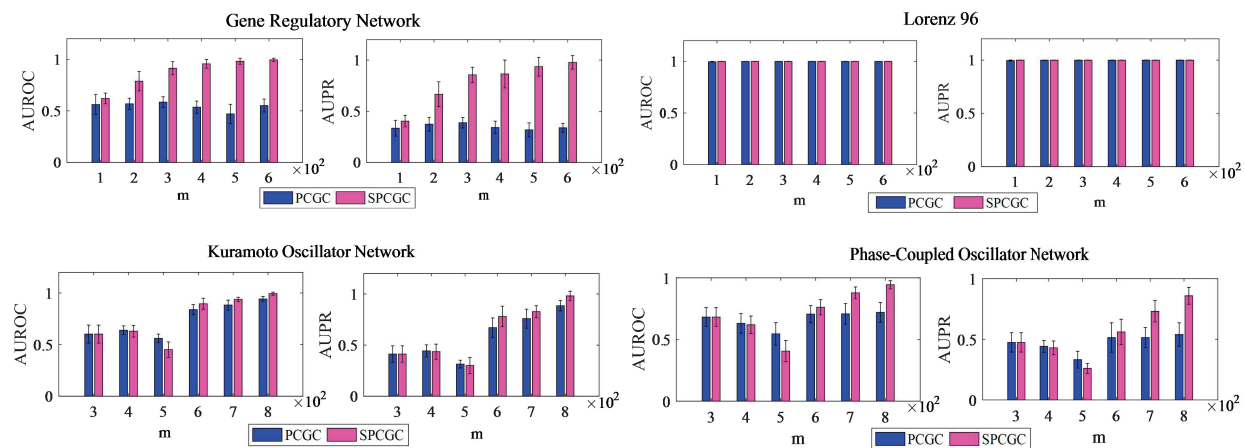


Figure 7. Simulation results of different models for the comparison of PCGC and SPCGC. All of the values of the AUROC and AUPR are averaged over 10 realizations.

Table 3. Parametric settings of simulations of different models for the comparison of PCGC and SPCGC. Here, m is the number of measurements, T is the number of time points at each measurement, σ is the strength of noise, r is the polynomial order, and λ is the sparse penalty parameter.

Model	m	T	σ	r	λ
Gene regulatory network	{100; 200; 300; 400; 500; 600}	2	0.01	5	0.002
Lorenz 96	{100; 200; 300; 400; 500; 600}	2	0.01	5	0.001
Kuramoto oscillator	{300; 400; 500; 600; 700; 800}	2	0.01	5	0.003
Phase-coupled oscillator	{300; 400; 500; 600; 700; 800}	2	0.01	5	0.003

4. Discussion

The traditional Granger causality method was proposed based on a linear VAR model, which often cannot achieve good results under the condition of an insufficient sample size. Based on the consideration of nonlinearity, sparsity, and causality, the model-free PCGC and SPCGC methods are presented in this study and, based on the simulation of different complex nonlinear networked systems, the robustness and effectiveness of our proposed method are verified. However, the following questions still need to be addressed and discussed.

(1) The heterogeneity and diversity of data can guarantee the performance and quality of causal inference methods. Therefore, different initial values under different conditions are used to collect data using a scheme with multiple measurements. In this case, our proposed methods are generally proven to be effective and robust in simulations of many system models. However, some issues should be addressed. In Figure 5, for the gene regulatory network and Kuramoto oscillator network, when the sample size is relatively small, the heterogeneity of the data is not adequate. This is also the reason why our proposed method PCGC obtains lower AUROC and AUPR values in the simulations of the gene regulatory network and Kuramoto oscillator network. However, as the sample size increases, the performance of PCGC is greatly improved. On the other hand, the Lorenz 96 and the Lorenz systems have chaotic dynamic trajectories. The data generated from them are heterogeneous enough. So, PCGC still achieves good performance even with a relatively smaller sample size in the simulation of Lorenz 96 and the Lorenz systems. The AUROC and AUPR of PCGC reach stable values as the sample size increases. In particular, we conducted a detailed comparison between PCGC and SPCGC in a gene regulatory network, as shown in Figure 4. The performance of methods varies greatly as the sample size increases. We attempted to analyze the reasons in detail to make this study more representative. Of course, other models can be used for the same analysis and comparison.

(2) The focus of our theoretical research is not to specifically study the impact of the presence of measurement noise on Granger causality inference, but to consider the

robustness of our method in the presence of certain measurement noise. For this reason, we also performed some simulation studies in the case of different noise intensities. The robustness to uncertainties in the data was addressed; see Figure 6. Figure 6 provides the simulation results of different models at different noise intensities. We discuss this problem of measurement noise on Granger causality analysis [54,55] and refer to some works that have already attempted to explore this aspect [56,57]. Based on the consideration of noise, we believe that a combination of state-space modeling and some filtering methods for our PCGC framework would work well.

(3) To improve the performance of our proposed method, the choice of method parameters is also a crucial issue. In general, for a complex system, the same polynomial order r can be set for all system equations. In particular, for most networked systems, due to the difference in the complexity of the system equation for each node, it is necessary to give a specific polynomial order to different system equations. In this way, the respective complexity of the system equations can be roughly estimated by determining different polynomial orders of the system equations. Similarly, for a complex system, the same penalty parameter λ can be set for all system equations. For different equations of a system with heterogeneous complexity, it is more reasonable to choose different values of λ for different target nodes (with the respective equations of the system model) rather than using the same value of λ .

(4) In our study, we specifically focus on the proposed methods to verify their performance from different perspectives in detail, such as different types of models, different types of data, different sample sizes, and different noise intensities. In the future, the proposed methods will be compared and analyzed with existing advanced causal network inference methods. For further perspectives, the Lasso selection in SPCGC can be replaced with other compressed sensing algorithms for the investigation of better performance. We can make a similar comparison of penalized regression methods, as shown in [44]. On the other hand, based on a summary of some existing Granger causality methods [38–40], including an analysis of causal methods based on information theory [58,59], we should obtain an idea of solving causal network inference. In particular, a similar method should be pointed out, as it was proposed based on software space intervention analysis, and it is named algorithmic information dynamics. It might be considered as an alternative to our causal approach, but it uses computer programs rather than polynomial differential equations [59], and this needs to be studied in our future work.

5. Conclusions

In brief, the exact identification and dynamic prediction of causal networks from data were demonstrated by our proposed methods, one of which is termed Polynomial Conditional Granger Causality (PCGC), while the other is its expanded sparse-promoting version (SPCGC). These methods do not require assumptions on the specific form of the nonlinear system equations. In particular, a general polynomial approximation is adopted to fit the relationships among nodes, which has the superiority of turning the nonlinearity of complex systems into some additive terms. Then, SPCGC combines nonlinear regression based on Granger causality with sparse optimization, which is fit for larger high-dimensional systems with a relatively smaller number of samples.

From the point of view of network science, all real systems can be represented by networks. Due to the complexity of real systems, the mathematical model of a real system is often indistinct or unavailable. Therefore, for networked systems, it is necessary to develop a data-driven causal inference method that can provide a foundation for subsequent network analysis. In general, as a simple and effective method, this model-free framework can, to some extent, provide new prospects, and represents an attempt to develop a basis for network reconstruction methods in the future.

Author Contributions: Conceptualization, G.Y. (Guanxue Yang), S.L. and G.Y. (Guanxiao Yang); investigation, G.Y. (Guanxue Yang) and S.L.; writing—original draft preparation, G.Y. (Guanxue Yang) and S.L.; writing—review and editing, G.Y. (Guanxue Yang), S.L. and G.Y. (Guanxiao Yang); funding acquisition, G.Y. (Guanxue Yang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grant number 61903161.

Data Availability Statement: The original contributions presented in the study are included in this article, further information can be available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Nonlinear System Models for Simulation

Appendix A.1. Lorenz Oscillator

The system model of the Lorenz oscillator is formulated as shown in Equation (A1).

$$\begin{aligned}\dot{x} &= \sigma(y - x) + \tilde{\xi}_x \\ \dot{y} &= x(\rho - z) - y + \tilde{\xi}_y \\ \dot{z} &= xy - \beta z + \tilde{\xi}_z\end{aligned}\tag{A1}$$

where the standard parameters are $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. $\tilde{\xi}_x$, $\tilde{\xi}_y$, and $\tilde{\xi}_z$ are zero-mean uncorrelated Gaussian noise terms with identical unit variance ($\tilde{\xi} \sim \mathcal{N}(0, \sigma^2)$).

Appendix A.2. Rössler Oscillator

The system model of the Rössler oscillator is formulated as shown in Equation (A2).

$$\begin{aligned}\dot{x} &= -y - z + \tilde{\xi}_x \\ \dot{y} &= x + ay + \tilde{\xi}_y \\ \dot{z} &= b + z(x - c) + \tilde{\xi}_z\end{aligned}\tag{A2}$$

where $a = 0.1$, $b = 0.1$, and $c = 14$. $\tilde{\xi}_x$, $\tilde{\xi}_y$, and $\tilde{\xi}_z$ are zero-mean uncorrelated Gaussian noise terms with identical unit variance ($\tilde{\xi} \sim \mathcal{N}(0, \sigma^2)$).

Appendix A.3. Lorenz 96

The Lorenz 96 model is a dynamical system formulated by Edward Lorenz in 1996, and it is defined as shown in Equation (A3).

$$\dot{x}_i = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F + \tilde{\xi}_i,\tag{A3}$$

where $i = 1, 2, \dots, N$; $x_{-1} = x_{N-1}$, $x_0 = x_N$, and $x_1 = x_{N+1}$. x_i is the state of the system, and F is a forcing constant. $F = 8$ is a common value known to cause chaotic behavior. $\tilde{\xi}_i$ is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\tilde{\xi}_i \sim \mathcal{N}(0, \sigma^2)$).

Appendix A.4. Biochemical Reaction Network

The system model of the biochemical reaction network is formulated as shown in Equation (A4).

$$\begin{aligned}
\dot{x}_1 &= -\gamma_1 x_1 + \frac{\alpha_1}{1 + x_6^{n_1}} + \xi_1 \\
\dot{x}_2 &= -\gamma_2 x_2 + \frac{\alpha_2}{1 + x_4^{n_2}} + \xi_2 \\
\dot{x}_3 &= -\gamma_3 x_3 + \frac{\alpha_3}{1 + x_5^{n_3}} + \xi_3 \\
\dot{x}_4 &= -\gamma_4 x_4 + \beta_1 x_1 + \xi_4 \\
\dot{x}_5 &= -\gamma_5 x_5 + \beta_2 x_2 + \xi_5 \\
\dot{x}_6 &= -\gamma_6 x_6 + \beta_3 x_3 + \xi_6
\end{aligned} \tag{A4}$$

where x_1 , x_2 , and x_3 are concentrations of the mRNA transcripts of genes 1, 2, and 3, respectively; x_4 , x_5 , and x_6 are concentrations of the proteins of genes 1, 2, and 3, respectively; α_1 , α_2 , and α_3 are the maximum promoter strengths for the corresponding genes; γ_1 , γ_2 , and γ_3 are mRNA decay rates; γ_4 , γ_5 , and γ_6 are protein decay rates; β_1 , β_2 , and β_3 are protein production rates; and n_1 , n_2 , and n_3 are hill coefficients. ξ_i is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\xi_i \sim \mathcal{N}(0, \sigma^2)$). The parameters of Equation (A4) are given in Table A1.

Table A1. Parameter settings in the biochemical reaction network.

$\alpha_1 = 4$	$\alpha_2 = 3$	$\alpha_3 = 5$			
$\gamma_1 = 0.3$	$\gamma_2 = 0.4$	$\gamma_3 = 0.5$	$\gamma_4 = 0.2$	$\gamma_5 = 0.4$	$\gamma_6 = 0.6$
$\beta_1 = 1.4$	$\beta_2 = 1.5$	$\beta_3 = 1.6$			
$n_1 = 4$	$n_2 = 4$	$n_3 = 4$			

Appendix A.5. Metabolic Network 1: Glycolytic Oscillator

The system model of the glycolytic oscillator is formulated as shown in Equation (A5).

$$\begin{aligned}
\dot{x}_1 &= J_0 - \frac{k_1 x_1 x_6}{1 + (x_6/K_1)^4} + \xi_1 \\
\dot{x}_2 &= \frac{2k_1 x_1 x_6}{1 + (x_6/K_1)^4} - k_2 x_2 (N - x_5) - k_6 x_2 x_5 + \xi_2 \\
\dot{x}_3 &= k_2 x_2 (N - x_5) - k_3 x_3 (A - x_6) + \xi_3 \\
\dot{x}_4 &= k_3 x_3 (A - x_6) - k_4 x_4 x_5 - \kappa(x_4 - x_7) + \xi_4 \\
\dot{x}_5 &= k_2 x_2 (N - x_5) - k_4 x_4 x_5 - k_6 x_2 x_5 + \xi_5 \\
\dot{x}_6 &= \frac{-2k_1 x_1 x_6}{1 + (x_6/K_1)^4} + 2k_3 x_3 (A - x_6) - k_5 x_6 + \xi_6 \\
\dot{x}_7 &= \psi \kappa(x_4 - x_7) - k x_7 + \xi_7
\end{aligned} \tag{A5}$$

where x_1 is the concentration of glucose; x_2 is a pool compounded of triose phosphates, glyceraldehyde 3-phosphate, and dihydroxyacetone phosphate; x_3 is the concentration of 1,3-bisphosphoglycerate; x_4 is a pooled compound of pyruvate and acetaldehyde; x_5 is the concentration of NADH; x_6 is the concentration of ATP; and x_7 is the concentration of extracellular coupling substances [51,52]. ξ_i is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\xi_i \sim \mathcal{N}(0, \sigma^2)$). The parameters of Equation (A5) are given in Table A2.

Table A2. Parameter settings in the glycolytic oscillator model.

$J_0 = 2.5;$	$k = 1.8;$	$\psi = 0.1;$	$\kappa = 13$	
$K_1 = 0.52;$	$N = 1;$	$A = 4;$	$k_1 = 100$	
$k_2 = 6;$	$k_3 = 16;$	$k_4 = 100;$	$k_5 = 1.28;$	$k_6 = 12$

Appendix A.6. Metabolic Network 2: Yeast Glycolysis

The system model of yeast glycolysis is formulated as shown in Equation (A6).

$$\begin{aligned}
 \dot{x}_1 &= a_1 + \frac{a_2 x_1 x_6}{1 + a_3 x_6^4} + \zeta_1 \\
 \dot{x}_2 &= \frac{b_1 x_1 x_6}{1 + b_2 x_6^4} + b_3 x_2 - b_4 x_2 x_7 + \zeta_2 \\
 \dot{x}_3 &= c_1 x_2 + c_2 x_3 + c_3 x_2 x_7 + c_4 x_3 x_6 + \zeta_3 \\
 \dot{x}_4 &= d_1 x_3 + d_2 x_4 + d_3 x_5 + d_4 x_3 x_6 + d_5 x_4 x_7 + \zeta_4 \\
 \dot{x}_5 &= e_1 x_4 + e_2 x_5 + \zeta_5 \\
 \dot{x}_6 &= \frac{f_1 x_1 x_6}{1 + f_2 x_6^4} + f_3 x_3 + f_4 x_3 x_7 + f_5 x_6 + \zeta_6 \\
 \dot{x}_7 &= g_1 x_2 + g_2 x_2 x_7 + g_3 x_4 x_7 + \zeta_7
 \end{aligned} \tag{A6}$$

where the description of x_i ($i = 1, 2, \dots, 7$) is the same as above. ζ_i is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\zeta_i \sim \mathcal{N}(0, \sigma^2)$). The parameters of Equation (A6) are given in Table A3.

Table A3. Parameter settings in the yeast glycolysis model.

$a_1 = 2.5;$	$a_2 = -100;$	$a_3 = 13.6769$		
$b_1 = 200;$	$b_2 = 13.6769;$	$b_3 = -6;$	$b_4 = -6$	
$c_1 = 6;$	$c_2 = -64;$	$c_3 = 6;$	$c_4 = 16$	
$d_1 = 64;$	$d_2 = -13;$	$d_3 = 13;$	$d_4 = -16;$	$d_5 = -100$
$e_1 = 1.3;$	$e_2 = -3.1$			
$f_1 = -200;$	$f_2 = 13.6769;$	$f_3 = 128;$	$f_4 = -1.28;$	$f_5 = -32$
$g_1 = 6;$	$g_2 = -18;$	$g_3 = -100$		

Appendix A.7. Gene Regulatory Network

Suppose that gene interactions are modeled by the Michaelis–Menten equation, as shown in Equation (A7). This is a popular system of equations for gene networks.

$$\dot{x}_i = -x_i^f + \sum_{j=1}^N b_{ij} \frac{x_j^h}{x_j^h + 1} + \zeta_i, \tag{A7}$$

where $f = 1$ and $h = 2$. Let $\mathbf{B} = [b_{ij}]_{N \times N}$ be the adjacency matrix, which characterizes the interactions among nodes. $b_{ij} \neq 0$ if the j -th variable connects to the i -th one; otherwise, $b_{ij} = 0$. ζ_i is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\zeta_i \sim \mathcal{N}(0, \sigma^2)$).

Appendix A.8. Kuramoto Oscillator

The system model of the Kuramoto oscillator is formulated as shown in Equation (A8).

$$\dot{x}_i = \omega_i + \sum_{j=1}^N b_{ij} \sin(x_j - x_i) + \zeta_i, \tag{A8}$$

where ω_i is a constant natural frequency. The second term describes the interactions between variable i and its interacting partners j . The nonlinear functions $\sin(\bullet)$ represent the dynamical laws that govern the variables of the system. $\mathbf{B} = [b_{ij}]_{N \times N}$ is an adjacency matrix that describes the interactions among variables. $b_{ij} \neq 0$ if the j -th variable connects to the i -th one; otherwise, $b_{ij} = 0$. ξ_i is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\xi_i \sim \mathcal{N}(0, \sigma^2)$).

Appendix A.9. Phase-Coupled Oscillator

The model of a phase-coupled oscillator is given with coupling functions that have two Fourier items in Equation (A9).

$$\dot{x}_i = \omega_i + \sum_{j=1}^N b_{ij} [\sin(x_j - x_i - 1.05) + 0.33 \sin(2(x_j - x_i))] + \xi_i, \quad (\text{A9})$$

where ω_i is a constant natural frequency. The two Fourier modes describe the interactions between variable i and its interacting partners j . The nonlinear functions $\sin(\bullet)$ represent the dynamical laws that govern the variables of the system. $\mathbf{B} = [b_{ij}]_{N \times N}$ is an adjacency matrix that describes the interactions among variables. $b_{ij} \neq 0$ if the j -th variable connects to the i -th one; otherwise, $b_{ij} = 0$. ξ_i is a zero-mean uncorrelated Gaussian noise term with identical unit variance ($\xi_i \sim \mathcal{N}(0, \sigma^2)$).

References

1. Marbach, D.; Costello, J.C.; Küffner, R.; Vega, N.M.; Prill, R.J.; Camacho, D.M.; Allison, K.R.; Kellis, M.; Collins, J.J.; Stolovitzky, G. Wisdom of crowds for robust gene network inference. *Nat. Methods* **2012**, *9*, 796–804. [\[CrossRef\]](#) [\[PubMed\]](#)
2. Ahmed, A.; Xing, E.P. Recovering time-varying networks of dependencies in social and biological studies. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 11878–11883. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Pratapa, A.; Jaliha, A.P.; Law, J.N.; Bharadwaj, A.; Murali, T. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **2020**, *17*, 147–154. [\[CrossRef\]](#)
4. Runge, J.; Petoukhov, V.; Donges, J.F.; Hlinka, J.; Jajcay, N.; Vejmelka, M.; Hartman, D.; Marwan, N.; Paluš, M.; Kurths, J. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Commun.* **2015**, *6*, 8502. [\[CrossRef\]](#)
5. Wang, W.X.; Lai, Y.C.; Grebogi, C. Data based identification and prediction of nonlinear and complex dynamical systems. *Phys. Rep.* **2016**, *644*, 1–76. [\[CrossRef\]](#)
6. Chang, Y.H.; Gray, J.W.; Tomlin, C.J. Exact reconstruction of gene regulatory networks using compressive sensing. *BMC Bioinform.* **2014**, *15*, 400. [\[CrossRef\]](#)
7. Shen, Z.; Wang, W.X.; Fan, Y.; Di, Z.; Lai, Y.C. Reconstructing propagation networks with natural diversity and identifying hidden sources. *Nat. Commun.* **2014**, *5*, 4323. [\[CrossRef\]](#)
8. Mei, G.; Wu, X.; Wang, Y.; Mi, H.; Lu, J.A.; Chen, G. Compressive-sensing-based structure identification for multilayer networks. *IEEE Trans. Cybern.* **2017**, *48*, 754–764. [\[CrossRef\]](#)
9. Yeung, M.S.; Tegnér, J.; Collins, J.J. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 6163–6168. [\[CrossRef\]](#)
10. Alter, O.; Brown, P.O.; Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10101–10106. [\[CrossRef\]](#)
11. Rogers, S.; Girolami, M. A Bayesian regression approach to the inference of regulatory networks from gene expression data. *Bioinformatics* **2005**, *21*, 3131–3137. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Deng, Y.; Zenil, H.; Tegnér, J.; Kiani, N.A. HiDi: An efficient reverse engineering schema for large-scale dynamic regulatory network reconstruction using adaptive differentiation. *Bioinformatics* **2017**, *33*, 3964–3972. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3932–3937. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Pan, W.; Yuan, Y.; Gonçalves, J.; Stan, G.B. A sparse Bayesian approach to the identification of nonlinear state-space systems. *IEEE Trans. Autom. Control* **2015**, *61*, 182–187. [\[CrossRef\]](#)
15. Mangan, N.M.; Brunton, S.L.; Proctor, J.L.; Kutz, J.N. Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Trans. Mol. Biol. Multi-Scale Commun.* **2016**, *2*, 52–63. [\[CrossRef\]](#)
16. Wang, S.; Herzog, E.D.; Kiss, I.Z.; Schwartz, W.J.; Bloch, G.; Sebek, M.; Granados-Fuentes, D.; Wang, L.; Li, J.S. Inferring dynamic topology for decoding spatiotemporal structures in complex heterogeneous networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9300–9305. [\[CrossRef\]](#)
17. Casadiego, J.; Nitzan, M.; Hallerberg, S.; Timme, M. Model-free inference of direct network interactions from nonlinear collective dynamics. *Nat. Commun.* **2017**, *8*, 2192. [\[CrossRef\]](#)

18. Rice, J.J.; Tu, Y.; Stolovitzky, G. Reconstructing biological networks using conditional correlation analysis. *Bioinformatics* **2004**, *21*, 765–773. [\[CrossRef\]](#)
19. De La Fuente, A.; Bing, N.; Hoeschele, I.; Mendes, P. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics* **2004**, *20*, 3565–3574. [\[CrossRef\]](#)
20. Frenzel, S.; Pompe, B. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.* **2007**, *99*, 204101. [\[CrossRef\]](#)
21. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Faes, L.; Nollo, G.; Porta, A. Information-based detection of nonlinear Granger causality in multivariate processes via a nonuniform embedding technique. *Phys. Rev. E* **2011**, *83*, 051112. [\[CrossRef\]](#)
23. Kugiumtzis, D. Partial transfer entropy on rank vectors. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 401–420. [\[CrossRef\]](#)
24. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Youssofzadeh, V.; Prasad, G.; Naeem, M.; Wong-Lin, K. Temporal information of directed causal connectivity in multi-trial ERP data using partial Granger causality. *Neuroinformatics* **2016**, *14*, 99–120. [\[CrossRef\]](#)
26. Zhou, D.; Xiao, Y.; Zhang, Y.; Xu, Z.; Cai, D. Causal and structural connectivity of pulse-coupled nonlinear networks. *Phys. Rev. Lett.* **2013**, *111*, 054102. [\[CrossRef\]](#)
27. Lusch, B.; Maia, P.D.; Kutz, J.N. Inferring connectivity in networked dynamical systems: Challenges using Granger causality. *Phys. Rev. E* **2016**, *94*, 032220. [\[CrossRef\]](#)
28. Wu, X.; Wang, W.; Zheng, W.X. Inferring topologies of complex networks with hidden variables. *Phys. Rev. E* **2012**, *86*, 046106. [\[CrossRef\]](#)
29. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econom. J. Econom. Soc.* **1969**, *37*, 424–438. [\[CrossRef\]](#)
30. Geweke, J.F. Measures of conditional linear dependence and feedback between time series. *J. Am. Stat. Assoc.* **1984**, *79*, 907–915. [\[CrossRef\]](#)
31. Dhamala, M.; Rangarajan, G.; Ding, M. Estimating Granger causality from fourier and wavelet transforms of time series data. *Phys. Rev. Lett.* **2008**, *100*, 018701. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Baccala, L.A.; Takahashi, D.Y.; Sameshima, K. Directed Transfer Function: Unified Asymptotic Theory and Some of its Implications. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 2450–2460. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Omidvarnia, A.; Azemi, G.; Boashash, B.; Colditz, P.B.; Vanhatalo, S. Orthogonalized Partial Directed Coherence: Measuring time-varying interactions within EEG channels. *IEEE Trans. Bio-Med Eng.* **2014**, *61*, 680–693. [\[CrossRef\]](#)
34. Tao, C.; Feng, J. Nonlinear association criterion, nonlinear Granger causality and related issues with applications to neuroimage studies. *J. Neurosci. Methods* **2016**, *262*, 110–132. [\[CrossRef\]](#)
35. Marinazzo, D.; Pellicoro, M.; Stramaglia, S. Kernel method for nonlinear Granger causality. *Phys. Rev. Lett.* **2008**, *100*, 144103. [\[CrossRef\]](#)
36. Ancona, N.; Marinazzo, D.; Stramaglia, S. Radial basis function approach to nonlinear Granger causality of time series. *Phys. Rev. E* **2004**, *70*, 056221. [\[CrossRef\]](#)
37. Hu, M.; Liang, H. A copula approach to assessing Granger causality. *Neuroimage* **2014**, *100*, 125–134. [\[CrossRef\]](#)
38. Finkle, J.D.; Wu, J.J.; Bagheri, N. Windowed granger causal inference strategy improves discovery of gene regulatory networks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 2252–2257. [\[CrossRef\]](#)
39. Stokes, P.A.; Purdon, P.L. A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E7063–E7072. [\[CrossRef\]](#)
40. Sheikhattar, A.; Miran, S.; Liu, J.; Fritz, J.B.; Shamma, S.A.; Kanold, P.O.; Babadi, B. Extracting neuronal functional network dynamics via adaptive Granger causality analysis. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E3869–E3878. [\[CrossRef\]](#)
41. Shojaie, A.; Michailidis, G. Discovering graphical Granger causality using the truncating lasso penalty. *Bioinformatics* **2010**, *26*, i517–i523. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Lozano, A.C.; Abe, N.; Liu, Y.; Rosset, S. Grouped graphical Granger modeling for gene expression regulatory networks discovery. *Bioinformatics* **2009**, *25*, i110–i118. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Siggiridou, E.; Kugiumtzis, D. Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model. *IEEE Trans. Signal Process.* **2015**, *64*, 1759–1773. [\[CrossRef\]](#)
44. Antonacci, Y.; Toppi, P.; Pietrabissa, A.; Anzolin, A.; Astolfi, L. Measuring connectivity in linear multivariate processes with penalized regression techniques. *IEEE Access* **2024**, *12*, 30638–30652. [\[CrossRef\]](#)
45. Yang, G.; Wang, L.; Wang, X. Reconstruction of complex directional networks with group lasso nonlinear conditional granger causality. *Sci. Rep.* **2017**, *7*, 2991. [\[CrossRef\]](#)
46. Siggiridou, E.; Kugiumtzis, D. Dimension reduction of polynomial regression models for the estimation of Granger causality in high-dimensional time series. *IEEE Trans. Signal Process.* **2021**, *69*, 5638–5650. [\[CrossRef\]](#)
47. Antonacci, Y.; Astolfi, L.; Nollo, G.; Faes, L. Information transfer in linear multivariate processes assessed through penalized regression techniques: Validation and application to physiological networks. *Entropy* **2020**, *22*, 732. [\[CrossRef\]](#)

48. Antonacci, Y.; Minati, L.; Faes, L.; Pernice, R.; Nollo, G.; Toppi, J.; Pietrabissa, A.; Astolfi, L. Estimation of Granger causality through Artificial Neural Networks: applications to physiological systems and chaotic electronic oscillators. *PeerJ Comput. Sci.* **2021**, *7*, e429. [[CrossRef](#)]
49. Dong, C.Y.; Shin, D.; Joo, S.; Nam, Y.; Cho, K.H. Identification of feedback loops in neural networks based on multi-step Granger causality. *Bioinformatics* **2012**, *28*, 2146–2153. [[CrossRef](#)]
50. Gao, J.; Barzel, B.; Barabási, A.L. Universal resilience patterns in complex networks. *Nature* **2016**, *530*, 307–312. [[CrossRef](#)]
51. Wolf, J.; Heinrich, R. Effect of cellular interaction on glycolytic oscillations in yeast: A theoretical investigation. *Biochem. J.* **2000**, *345*, 321–334. [[CrossRef](#)] [[PubMed](#)]
52. Ruoff, P.; Christensen, M.K.; Wolf, J.; Heinrich, R. Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations. *Biophys. Chem.* **2003**, *106*, 179–192. [[CrossRef](#)] [[PubMed](#)]
53. Daniels, B.C.; Nemenman, I. Efficient inference of parsimonious phenomenological models of cellular dynamics using S-systems and alternating regression. *PLoS ONE* **2015**, *10*, e0119821. [[CrossRef](#)] [[PubMed](#)]
54. Nalatore, H.; N, S.; Rangarajan, G. Effect of measurement noise on Granger causality. *Phys. Rev. E* **2014**, *90*, 062127. [[CrossRef](#)]
55. Vinck, M.; Huurdeman, L.; Bosman, C.A.; Fries, P.; Battaglia, F.P.; Pennartz, C.M.; Tiesinga, P.H. How to detect the Granger-causal flow direction in the presence of additive noise? *Neuroimage* **2015**, *108*, 301–318. [[CrossRef](#)]
56. Sommerlade, L.; Thiel, M.; Platt, B.; Plano, A.; Riedel, G.; Grebogi, C.; Timmer, J.; Schelter, B. Inference of Granger causal time-dependent influences in noisy multivariate time series. *J. Neurosci. Methods* **2012**, *203*, 173–185. [[CrossRef](#)]
57. Ahmadi, S.; Nair, G.N.; Weyer, E. Granger causality of Gaussian signals from noisy or filtered measurements. *IFAC-PapersOnLine* **2020**, *53*, 506–511. [[CrossRef](#)]
58. Zenil, H.; Kiani, N.A.; Tegnér, J. Low-algorithmic-complexity entropy-deceiving graphs. *Phys. Rev. E* **2017**, *96*, 012308. [[CrossRef](#)]
59. Zenil, H.; Kiani, N.A.; Marabita, F.; Deng, Y.; Elias, S.; Schmidt, A.; Ball, G.; Tegner, J. An algorithmic information calculus for causal discovery and reprogramming systems. *Iscience* **2019**, *19*, 1160–1172. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.