

# SCIENTIFIC REPORTS



OPEN

## Reconstruction of Complex Directional Networks with Group Lasso Nonlinear Conditional Granger Causality

Guanxue Yang, Lin Wang & Xiaofan Wang

Reconstruction of networks underlying complex systems is one of the most crucial problems in many areas of engineering and science. In this paper, rather than identifying parameters of complex systems governed by pre-defined models or taking some polynomial and rational functions as a prior information for subsequent model selection, we put forward a general framework for nonlinear causal network reconstruction from time-series with limited observations. With obtaining multi-source datasets based on the data-fusion strategy, we propose a novel method to handle nonlinearity and directionality of complex networked systems, namely group lasso nonlinear conditional granger causality. Specially, our method can exploit different sets of radial basis functions to approximate the nonlinear interactions between each pair of nodes and integrate sparsity into grouped variables selection. The performance characteristic of our approach is firstly assessed with two types of simulated datasets from nonlinear vector autoregressive model and nonlinear dynamic models, and then verified based on the benchmark datasets from DREAM3 Challenge4. Effects of data size and noise intensity are also discussed. All of the results demonstrate that the proposed method performs better in terms of higher area under precision-recall curve.

In recent years, there has been an explosion of various datasets, which are collected in scientific, engineering, medical and social applications<sup>1,2</sup>. They often contain information that represents a combination of different properties of the real world. Identifying causality or correlation among datasets<sup>3-7</sup> is increasingly vital for effective policy and management recommendations on climate, epidemiology, neuroscience, economics and much else. With these rapid advances in the studies of causality and correlation, complex network reconstruction has become an outstanding and significant problem in interdisciplinary science<sup>8-12</sup>. As we know, numerous real networked systems could be represented as network of interconnected nodes. But in lots of situations, network topology is fully unknown, which is hidden in the observations acquired from experiments. For complex systems, accompanied by the complexity of system dynamics, the limited observations with noisy measurements make the problem of network reconstruction even more challenging. An increased attention for network reconstruction is being attracted in the past few years<sup>13-19</sup>.

Among the developed methods, vector autoregressive model (VAR) is able to estimate the temporal dependencies of variables in multivariate model, which gains growing interest in recent years<sup>20-22</sup>. As one of the most prevalent VAR methods, Granger Causality (GC) can be efficiently applied in causal discovery<sup>23</sup>. Conditional Granger Causality (CGC) is put forward to differentiate direct interactions from indirect ones<sup>24</sup>. To extend the application of GC limited by linear dynamics, nonlinear GC is developed<sup>25-28</sup>, which is relatively less considered until now. Generally speaking, the application of these GC methods might get in trouble, especially when the size of samples is small and the number of variables is large. To conquer such problem, some composed methods are presented by integrating variable selection into CGC model, such as Lasso-CGC<sup>29,30</sup> and grouped lasso graphical granger<sup>31</sup>. Group lasso is also used in multivariate regression and multi-task learning<sup>32-36</sup>. Compressive sensing or sparse regression is popularly applied in the network reconstruction and system identification. However, most of these methods are confined to identify parameters of complex systems governed by pre-defined models<sup>18,37,38</sup>.

Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai, 200240, P. R. China. Correspondence and requests for materials should be addressed to X.W. (email: [xfwang@sjtu.edu.cn](mailto:xfwang@sjtu.edu.cn))

Moreover, some methods consider taking some polynomial and rational functions as a prior information which is usually difficult to obtain in advance, for subsequent model selection<sup>39,40</sup>.

In this paper, we concern over inferring complex nonlinear causal networks from time-series and propose a new method termed as group lasso nonlinear conditional granger causality (GLasso-NCGC). Particularly, we establish a general data-driven framework for nonlinear network reconstruction from time-series with limited observations, which doesn't require the pre-defined model or some polynomial and rational functions as a prior information. With obtaining multi-source datasets based on the data-fusion strategy introduced in recent high quality paper<sup>3</sup>, we first introduce the formulation of multivariate nonlinear conditional granger causality model (NCGC), and exploit different groups of radial basis functions (RBF) to approximate the nonlinear interactions between each pair of nodes respectively. Then we decompose the task of inferring the whole network into local neighborhood selections centered at each target node. Together with the natural sparsity of real complex networks, group lasso regression<sup>41</sup> is utilized for each local structure recovery, where different RBF variables from different centers in each node should be either eliminated or selected as a group. Next, we obtain the candidate structure of the network by resolving the problem of group lasso regression. As a result, the final network can be judged by significance levels with Fisher statistic test for removing false existent links.

For the purpose of performance evaluations for our proposed method, simulated datasets from nonlinear vector autoregressive model are firstly generated. Compared with CGC, Lasso-CGC and NCGC, we find GLasso-NCGC outperforms other methods in terms of Area Under Precision-Recall curve (AUPR) and Receiver-Operating-Characteristic curve (AUROC). Effects of data size and noise intensity are also investigated. Besides, the simulation based on nonlinear dynamic models with network topology given in advance is carried out. We consider two classical nonlinear dynamic models which are used for modelling biochemical reaction networks<sup>40</sup> and gene regulatory networks<sup>42</sup> respectively. Especially for gene regulatory networks with Michaelis-Menten dynamics, we simulate gene regulatory model on random, small-world and scale-free networks. Then we explore the performance of these methods influenced by different average degrees, noise intensities and amounts of data simultaneously for these three types of networks. Based on the sub-challenge of Dialogue on Reverse Engineering Assessment and Methods (DREAM), we finally use the benchmark datasets of DREAM3 Challenge<sup>4</sup> for investigation. All of the results demonstrate the proposed method GLasso-NCGC executes best on the previous mentioned datasets, which is fully certified to be optimal and robust to noise.

## Models and Methods

**Multivariate conditional granger causality.** Consider  $N$  time-course variables  $\{X, Y, Z_1, Z_2, \dots, Z_{N-2}\}$  in multivariate conditional granger causality model, the current value of  $Y_t$  can be expressed by the past values of  $Y$  and  $Z_1, Z_2, \dots, Z_{N-2}$  in Equation (1). Meanwhile,  $Y_t$  can be also written as the joint representation of the past values of  $Y, X$  and  $Z_1, Z_2, \dots, Z_{N-2}$  in Equation (2).

$$Y_t = \sum_{i=1}^p a_i Y_{t-i} + \sum_{k=1}^{N-2} \sum_{i=1}^p c_{k,i} Z_{k,t-i} + \varepsilon_1 \quad (1)$$

$$Y_t = \sum_{i=1}^p a'_i Y_{t-i} + \sum_{i=1}^p b'_i X_{t-i} + \sum_{k=1}^{N-2} \sum_{i=1}^p c'_{k,i} Z_{k,t-i} + \varepsilon_2 \quad (2)$$

where,  $a_i, c_{k,i}, a'_i, b'_i, c'_{k,i}$  are the coefficients in VAR model,  $p$  is the model order,  $\varepsilon_1$  and  $\varepsilon_2$  are the noise terms. Let  $Z = \{Z_1, Z_2, \dots, Z_{N-2}\}$ , then the CGC index can be shown as  $CGCI_{X \rightarrow Y|Z} = \ln(\text{var}(\varepsilon_1)/\text{var}(\varepsilon_2))$ , which is used to analyze the conditional causal interaction between two temporal variables. If the variance  $\text{var}(\varepsilon_1)$  is larger than the variance  $\text{var}(\varepsilon_2)$ , i.e.  $CGCI_{X \rightarrow Y|Z} > 0$ ,  $X$  causes  $Y$  conditioned on the other sets of  $N-2$  variables  $Z$ . Generally, the statistic judgement of terminal results can be executed by significant levels with F-test.

**Group lasso nonlinear conditional granger causality.** The unified framework of multivariate nonlinear conditional granger causality (NCGC) model is proposed as shown in Equation (3).

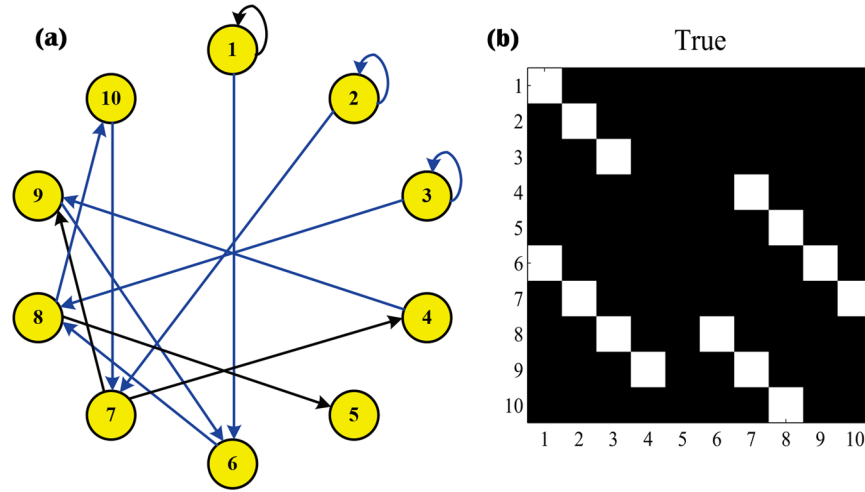
$$\mathbf{Y} = \Phi(\mathbf{X})\mathbf{A} + \mathbf{E} \quad (3)$$

where, target variables  $\mathbf{Y} = [y_1 \ y_2 \ \dots \ y_N]$ , coefficient matrix  $\mathbf{A} = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_N]$ , noise terms  $\mathbf{E} = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_N]$ ,  $\Phi(\mathbf{X})$  is data matrix of nonlinear kernel functions. The detailed descriptions for all the elements above are given in Equation (4).

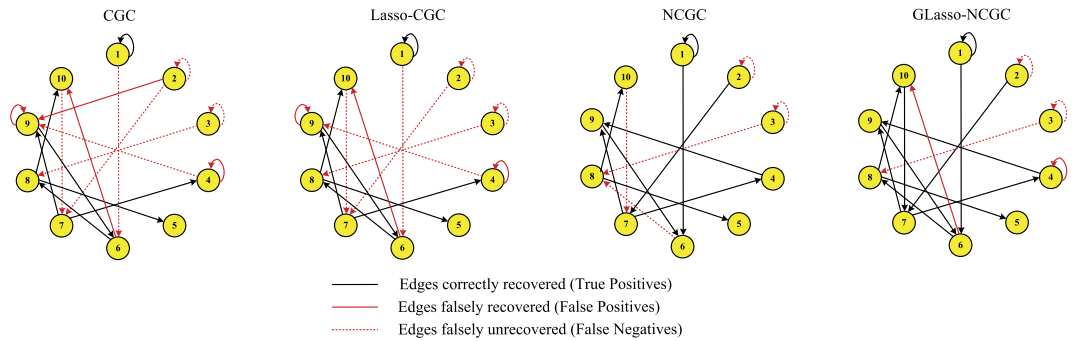
For each target variable  $y_i \in \mathbf{Y}$ , we can obtain  $N$  independent equations as follows.

$$y_i = \Phi(\mathbf{X})\alpha_i + \varepsilon_i, \quad i = 1, 2, \dots, N \quad (4)$$

where,



**Figure 1.** Directed graph and true adjacency matrix of nonlinear VAR model.



**Figure 2.** Networks inferred by CGC, Lasso-CGC, NCGC, GLasso-NCGC in nonlinear VAR simulation.

$$\begin{aligned}
 \mathbf{y}_i &= [x_i(p+1) \ x_i(p+2) \ \dots \ x_i(T)]^T \\
 \Phi(\mathbf{X}) &= [\Phi(\mathbf{X}_1) \ \Phi(\mathbf{X}_2) \ \dots \ \Phi(\mathbf{X}_N)] \\
 \mathbf{X}_j &= \begin{bmatrix} \mathbf{X}_j^1 \\ \mathbf{X}_j^2 \\ \vdots \\ \mathbf{X}_j^{T-p} \end{bmatrix} = \begin{bmatrix} x_j(p) & x_j(p-1) & \dots & x_j(1) \\ x_j(p+1) & x_j(p) & \dots & x_j(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_j(T-1) & x_j(T-2) & \dots & x_j(T-p) \end{bmatrix} \\
 \Phi(\mathbf{X}_j^k) &= [\varphi_1(\mathbf{X}_j^k) \ \varphi_2(\mathbf{X}_j^k) \ \dots \ \varphi_n(\mathbf{X}_j^k)], \\
 k &= 1, 2, \dots, t-p \\
 \varphi_\rho(\mathbf{X}_j^k) &= \exp(-\|\mathbf{X}_j^k - \tilde{\mathbf{X}}_j^\rho\|^2 / 2\sigma^2), \\
 \rho &= 1, 2, \dots, n \\
 \alpha_i &= [\alpha_{i1} \ \alpha_{i2} \ \dots \ \alpha_{iN}]^T \\
 \alpha_{ij} &= [a_{ij}(1) \ a_{ij}(2) \ \dots \ a_{ij}(n)]^T, \\
 j &= 1, 2, \dots, N
 \end{aligned}$$

$N$  is the number of variables (time-series),  $T$  is the length of each time-series. The value of  $i$ -th variable  $x_i(t)$  can be observed at time  $t$ . Here, the form of  $\varphi_\rho(\mathbf{X}_j^k)$  is taken as radial basis function.  $\{\tilde{\mathbf{X}}_j^\rho\}_{\rho=1}^n$  is the set of  $n$  centers in the space of  $\mathbf{X}_j$ , which can be acquired by  $k$ -means clustering methods.  $\alpha_i$  is the coefficient between target variable  $\mathbf{y}_i$  and  $\Phi(\mathbf{X})$ .  $a_{ij}(n)$  is the coefficient corresponding to the function  $\varphi_n(\mathbf{X}_j^k)$ .

Generally, we can use least square method to deal with Equation (4). But the solution procedure of Equation (4) may be problematic when the number of variables is relatively larger than the number of available samples. With the knowledge of sparsity in  $\mathbf{A}$ , we can turn to regularization-based methods for help. In this case, it's noteworthy that apparent groupings exist among variables. So variables belonging to the same group should be regarded as

a whole. Here, a series of RBF variables from the different centers in the same time-series should be either eliminated or selected as a group. Then group lasso is adopted to solve this problem of sparse regression as follows.

$$\hat{\alpha}_i = \underset{\alpha_i}{\operatorname{argmin}}(\|y_i - \Phi(\mathbf{X})\alpha_i\|_2^2 + \lambda_i \sum_{j=1}^N \|\alpha_{ij}\|_2) \tag{5}$$

In Equation (5), we take  $l_2$  norm as the intra-group penalty.

The detailed algorithm of Group Lasso Nonlinear Conditional Granger Causality is shown in Algorithm 1.

With small samples at hand, first-order vector autoregressive model is usually considered<sup>20,21</sup>. Besides, in our consideration, model order is also set as  $p = 1$ , which is in accordance with the characteristics of dynamic systems governed by state-space equations. For the time-delayed dynamic systems, the implementation of higher-order vector autoregressive model can be straightly extended. In addition, we can use cross validation criterion for the selection of optimal parameters, such as the number of RBF centers  $n$  and the coefficient of penalty terms  $\lambda_i$ . Then we mainly describe the selection of  $\lambda_i$  as follows. We use two stages of refined selection which are similarly adopted by Khan *et al.*<sup>43</sup>. In the first stage, we set the coarse values of the search space  $\lambda \in \{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$ , which determines the neighborhood of the optimal  $\lambda_i$ . In the second stage, we obtain the neighborhood of the optimal  $\lambda_i$  for refined search, i.e.  $\lambda \in [0.5\lambda_i, 1.5\lambda_i]$ , the interval  $\Delta\lambda = k\lambda_i$  ( $0 < k < 1$ ). Here,  $\lambda_1 = 10^{-4}$ ,  $\lambda_2 = 10^{-3}$ ,  $\lambda_3 = 10^{-2}$ ,  $\lambda_4 = 10^{-1}$ ,  $\lambda_5 = 1$ ,  $k = 0.1$ . For example, if we choose  $\lambda_3 = 10^{-2}$  in the first stage, we next confine the refined search in the range of  $\lambda \in [0.005, 0.015]$ , where  $\Delta\lambda = 10^{-3}$ . For large-scale networks, we can just take relatively larger  $k$  to reduce the range of search space and ensure the low computational cost.

Meanwhile, in order to ensure the variety of datasets, multiple measurements are often carried out under different conditions (adding perturbations or knocking out nodes *et al.*). As a result, the extension of Equation (3) can be written as the following formula.

$$\begin{bmatrix} \tilde{\mathbf{Y}}_1 \\ \vdots \\ \tilde{\mathbf{Y}}_m \end{bmatrix} = \Phi \left( \begin{bmatrix} \tilde{\mathbf{X}}_1 \\ \vdots \\ \tilde{\mathbf{X}}_m \end{bmatrix} \right) \mathbf{A} + \begin{bmatrix} \tilde{\mathbf{E}}_1 \\ \vdots \\ \tilde{\mathbf{E}}_m \end{bmatrix} \tag{6}$$

In Equation (6),  $m$  denotes the number of measurements. At  $m$ -th measurement,  $\tilde{\mathbf{Y}}_m$  and  $\tilde{\mathbf{X}}_m$  are acquired for integrating them in Equation (6). Then Equation (6) can be also divided into  $N$  independent equations that would be solved by group lasso optimization with Equation (5). Finally, we execute nonlinear conditional granger causality with F-test in terms of the given significant level  $P_{val}$ . In our paper, we set  $P_{val} = 0.01$  for all the statistic analysis.

---

**Algorithm 1** Group Lasso Nonlinear Conditional Granger Causality

---

1: **Input:** Time-series data  $\{x_i(t)\}$ ,  $i = 1, 2, \dots, N$ ;  $t = 1, 2, \dots, T$ .  $N$  is the number of variables.  $T$  is the length of each time-series.

2: According to model order  $p$ , form the data matrix  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N]$  and target matrix  $\mathbf{Y} = [y_1, y_2, \dots, y_N]$ .

3: Formulize the data matrix  $\mathbf{X}$  into  $\Phi(\mathbf{X})$  based on radial basis functions as shown in Equation (4).

4: **For** each target variable  $y_i \in \mathbf{Y}$

- Execute group lasso:

$$\hat{\alpha}_i = \underset{\alpha_i}{\operatorname{argmin}}(\|y_i - \Phi(\mathbf{X})\alpha_i\|_2^2 + \lambda_i \sum_{j=1}^N \|\alpha_{ij}\|_2).$$

- Obtain<sup>opt</sup> candidate variable sets  $S_i$  for each  $y_i$  according to  $\hat{\alpha}_i$ . Rearrange data matrix  $\Phi(\mathbf{X})$  with  $S_i$  expressed as  $\Phi_{S_i}$  and reform the expression of nonlinear conditional granger causality model.

- Execute nonlinear conditional granger causality with F-test in terms of the given significant level  $P_{val}$ . Confirm the causal variables of  $y_i$ .

**end**

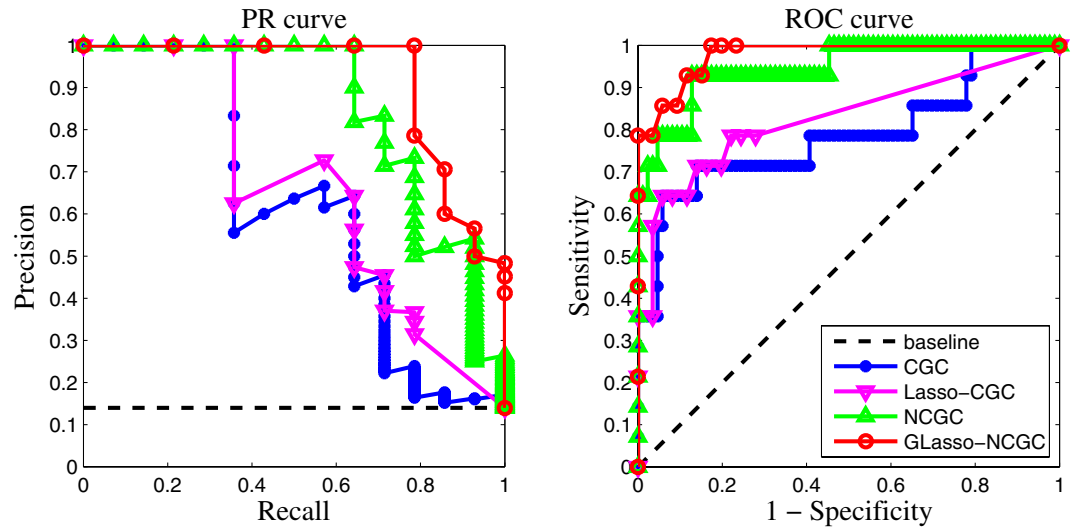
5: **Output:** Causal interactions among  $N$  variables, i.e. adjacency matrix  $\mathbb{R}^{N \times N}$ .

---

## Results

**Nonlinear vector autoregressive model.** The first simulation is based on a nonlinear vector autoregressive model with  $N = 10$  nodes (variables), see Equation (7).

$$\begin{aligned} x_{1,t} &= 0.5x_{1,t-1} + \varepsilon_{1,t} \\ x_{2,t} &= 0.6 \cos(x_{2,t-1}) + \varepsilon_{2,t} \\ x_{3,t} &= 0.7e^{-x_{3,t-1}^2} + \varepsilon_{3,t} \\ x_{4,t} &= 0.8x_{7,t-1} + \varepsilon_{4,t} \\ x_{5,t} &= 0.9x_{8,t-1} + \varepsilon_{5,t} \\ x_{6,t} &= \sin(x_{1,t-1}) - 0.6x_{9,t-1}^2 + \varepsilon_{6,t} \\ x_{7,t} &= 2 \cos(x_{2,t-1}) + 0.6 \sin(x_{10,t-1}) + \varepsilon_{7,t} \\ x_{8,t} &= 0.8 \cos(x_{3,t-1}) + \cos(x_{6,t-1}) + 1 + \varepsilon_{8,t} \\ x_{9,t} &= \sin(x_{4,t-1}) - 0.8x_{7,t-1} + \varepsilon_{9,t} \\ x_{10,t} &= e^{-x_{8,t-1}^2} + \varepsilon_{10,t} \end{aligned} \tag{7}$$



**Figure 3.** PR curves and ROC curves of CGC, Lasso-CGC, NCGC, GLasso-NCGC in nonlinear VAR simulation.

The directed graph from Equation (7) together with true adjacency matrix are shown in Fig. 1. In Fig. 1(a), black lines are linear edges and blue lines are nonlinear edges. In Fig. 1(b), white points are existent edges among nodes, while black areas stand for nonexistent edges. Next, the synthetic datasets ( $M = 100$  samples) are obtained based on the model of Equation (7), where  $\varepsilon$  is zero-mean uncorrelated gaussian noise terms with identical unit variance.

Networks inferred by CGC, Lasso-CGC, NCGC and GLasso-NCGC are shown in Fig. 2. According to the number of edges correctly recovered (True Positives), we can find that GLasso-NCGC almost captures all the existing edges except for the edges  $2 \rightarrow 2$ ,  $3 \rightarrow 3$ ,  $3 \rightarrow 8$ . Compared with GLasso-NCGC, NCGC additionally fails to recover two existing edges  $6 \rightarrow 8$ ,  $10 \rightarrow 7$ . Due to neglect the nonlinear influences among nodes, both CGC and Lasso-CGC fail to find many existing edges, which leads to many edges falsely unrecovered (False Negatives).

For further measuring performance we plot ROC curves and PR curves. Figure 3 is PR and ROC curves generated by CGC, Lasso-CGC, NCGC, GLasso-NCGC respectively. Obviously, it can be seen from Fig. 3 that GLasso-NCGC outperforms its competitors with the highest score (AUROC and AUPR).

In the following section, the performance of robust estimation with different methods is explored. Firstly, multi-realizations are carried out, here the number of realizations is set as 100. Figure 4 demonstrates discovery rate matrixes from multi-realizations. Table 1 shows the comparison of discovery rate of true positives inferred by these methods. Discovery rate means the total number of each edge rightly discovered over multi-realizations. Compared with CGC, Table 1 summarizes Lasso-CGC improves the performance with relatively larger discovery rate in true edges. However, due to neglecting the nonlinearity of modeling, they can't manifest better performance than NCGC and GLasso-NCGC. In general, for nonlinear edges, we can discover GLasso-NCGC nearly outperforms other methods. Specially, for edges  $1 \rightarrow 6$ ,  $2 \rightarrow 7$ ,  $3 \rightarrow 8$  and  $4 \rightarrow 9$ , GLasso-NCGC and NCGC greatly identify these true causal edges at large percentages of 100 independent realizations. For linear edges, both GLasso-NCGC and NCGC also maintain high discovery rate.

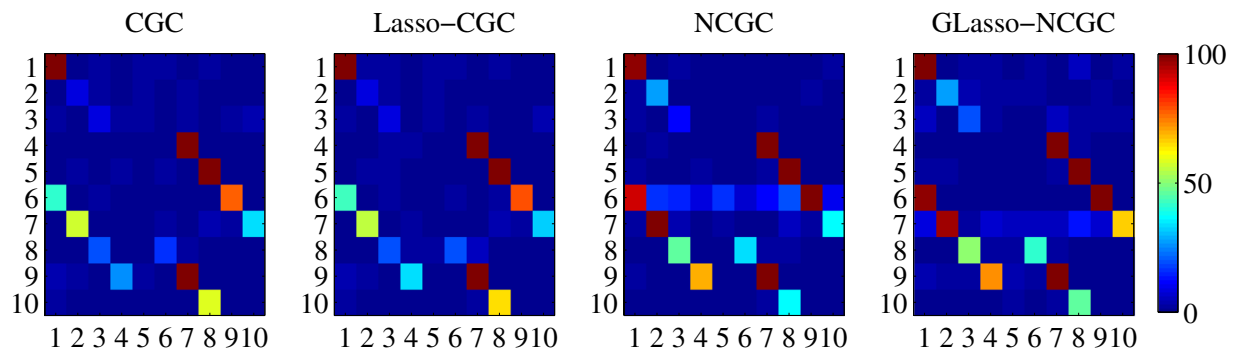
Relatively, GLasso-NCGC utilizes group lasso to promote the performance of NCGC by silencing many false positives and realize the best reconstruction of adjacency matrix at last. Average AUROC and Average AUPR of different methods over 100 realizations are calculated in Table 2.

Then, we explore the performance of CGC, Lasso-CGC, NCGC and GLasso-NCGC influenced by different number of samples. Given that the number of samples varies from 100 to 300, at each point, we get the Average AUROC and Average AUPR over 100 independent realizations respectively. From Fig. 5(a) and (b), we find the performance of these four methods improves as the number of samples increases. GLasso-NCGC scores highest in both Average AUROC and Average AUPR.

Next, given samples with  $M = 100$ , the performance of CGC, Lasso-CGC, NCGC and GLasso-NCGC influenced by different noise intensities  $\sigma$  (0.1~1) is also compared as shown in Fig. 5(c) and (d). At each intensity of noise, the Average AUROC and Average AUPR over 100 independent realizations are computed respectively. GLasso-NCGC also wins the best score in both Average AUROC and Average AUPR.

**Nonlinear dynamic model.** Consider the complex dynamic system of  $N$  variables (nodes) with the following coupled nonlinear equations.

$$\dot{x}_i = F(x_i) + \sum_{j=1, j \neq i}^N b_{ij} H(x_i, x_j) + \varepsilon_i \quad (8)$$



**Figure 4.** Discovery rate matrixes inferred from 100 realizations in nonlinear VAR simulation.

		CGC	Lasso-CGC	NCGC	GLasso-NCGC
Nonlinear	1 → 6	41%	43%	92%	97%
	2 → 2	8%	9%	28%	28%
	2 → 7	57%	56%	100%	96%
	3 → 3	9%	9%	11%	19%
	3 → 8	20%	20%	46%	50%
	4 → 9	26%	34%	69%	73%
	6 → 8	16%	19%	33%	42%
	8 → 10	58%	65%	37%	46%
	9 → 6	78%	79%	100%	100%
	10 → 7	33%	32%	36%	66%
Linear	1 → 1	99%	99%	97%	99%
	7 → 4	100%	100%	100%	100%
	7 → 9	100%	100%	100%	100%
	8 → 5	100%	100%	100%	100%

**Table 1.** Comparison of discovery rate of true positives extracted from true adjacency matrix of nonlinear VAR model.

Methods	Aver-AUROC (SD)	Aver-AUPR (SD)
CGC	0.8830 (0.0477)	0.6810 (0.0751)
Lasso-CGC	0.8837 (0.0448)	0.6967 (0.0673)
NCGC	0.9301 (0.0398)	0.7589 (0.0621)
GLasso-NCGC	<b>0.9519</b> (0.0356)	<b>0.8044</b> (0.0587)

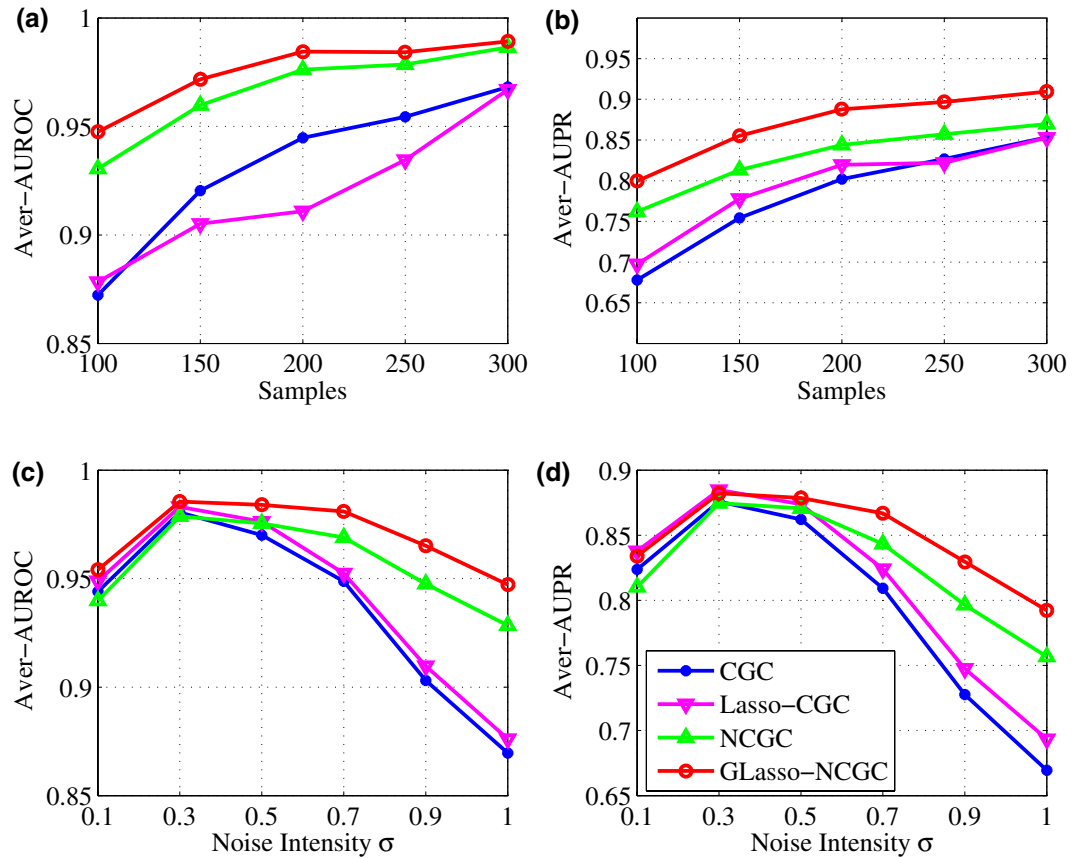
**Table 2.** Average AUPR and Average AUROC of different methods in nonlinear VAR simulation (standard deviation (SD) in parentheses). The values are obtained by averaging over 100 independent realizations.

where state vector  $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ , the first term on the right-hand side of Equation (8) describes the self-dynamics of the  $i$ -th variable, while the second term describes the interactions between variable  $i$  and its interacting partners  $j$ . The nonlinear functions  $F(x_i)$  and  $H(x_i, x_j)$  represent the dynamical laws that govern the variables of system. Let  $\mathbf{B}$  be the adjacent matrix which describes the interactions among variables.  $b_{ij} \neq 0$  if the  $j$ -th variable connects to the  $i$ -th one; otherwise,  $b_{ij} = 0$ .  $\varepsilon_i$  is zero-mean uncorrelated gaussian noise with variance  $\sigma^2$ . Indeed, there is no unified manner to establish the nonlinear framework for all the complex networked systems. But Equation (8) has the broad applications in many science domains. With an appropriate choice of  $F(x_i)$  and  $H(x_i, x_j)$ , Equation (8) is used to model various known systems, ranging from ecological systems, social systems and physical systems<sup>11, 18, 40, 42</sup>. Specifically, Equation (8) can be transformed into discrete-time expressions.

$$x_i(t_{k+1}) = x_i(t_k) + F(x_i(t_k)) + \sum_{j=1, j \neq i}^N b_{ij} H(x_i(t_k), x_j(t_k)) + \varepsilon_i(t_k)$$

where,  $t_{k+1} - t_k = 1$ . Here, the simulation systems are as follows.

**S1:** Biochemical reaction network (BRN).



**Figure 5.** Simulation of nonlinear VAR model. Average AUROC and Average AUPR of different methods acrossing samples (100~300), noise intensities  $\sigma$  (0.1~1) respectively. The values of points in these figures are computed by averaging over 100 independent realizations. **(a,b)** with noise  $\mathcal{N}(0,1)$ , while **(c,d)** with samples  $M=100$ .

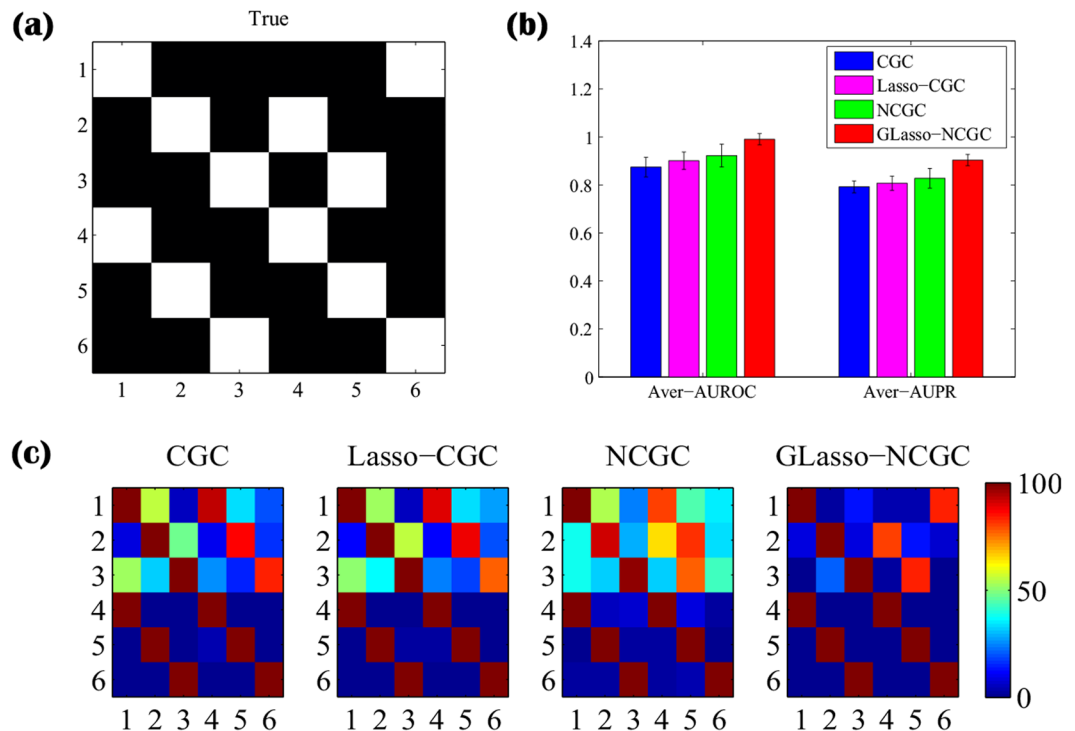
$$\begin{aligned}
 \dot{x}_1 &= -\gamma_1 x_1 + \frac{\alpha_1}{1 + x_6^{n_1}} + \varepsilon_1 \\
 \dot{x}_2 &= -\gamma_2 x_2 + \frac{\alpha_2}{1 + x_4^{n_2}} + \varepsilon_2 \\
 \dot{x}_3 &= -\gamma_3 x_3 + \frac{\alpha_3}{1 + x_5^{n_3}} + \varepsilon_3 \\
 \dot{x}_4 &= -\gamma_4 x_4 + \beta_1 x_1 + \varepsilon_4 \\
 \dot{x}_5 &= -\gamma_5 x_5 + \beta_2 x_2 + \varepsilon_5 \\
 \dot{x}_6 &= -\gamma_6 x_6 + \beta_3 x_3 + \varepsilon_6
 \end{aligned} \tag{9}$$

Here,  $x_1, x_2$  and  $x_3$  are concentrations of the mRNA transcripts of genes 1,2,3 respectively;  $x_4, x_5$  and  $x_6$  are concentrations of the proteins of genes 1,2,3 respectively;  $\alpha_1, \alpha_2, \alpha_3$  are maximum promoter strength for the corresponding gene;  $\gamma_1, \gamma_2, \gamma_3$  are mRNA decay rates;  $\gamma_4, \gamma_5, \gamma_6$  are protein decay rates;  $\beta_1, \beta_2, \beta_3$  are protein production rates;  $n_1, n_2, n_3$  are hill coefficients.

The parameters of Equation (9) are given as follows.  $\alpha_1 = 4, \alpha_2 = 3, \alpha_3 = 5; \gamma_1 = 0.3, \gamma_2 = 0.4, \gamma_3 = 0.5; \gamma_4 = 0.2, \gamma_5 = 0.4, \gamma_6 = 0.6; \beta_1 = 1.4, \beta_2 = 1.5, \beta_3 = 1.6; n_1 = n_2 = n_3 = 4$ . Meanwhile, we set the number of samples  $M = 50$  and noise  $\varepsilon \sim \mathcal{N}(0, 0.1^2)$ .

Based on the true network [Fig. 6(a)] derived from Equation (9), we can find GLasso-NCGC almost recover all of the real edges over 100 independent realizations as shown in Fig. 6(c). From Fig. 6(c), CGC, Lasso-CGC and NCGC can't discover the entire true positives but result in lots of false positives. In detail, we next choose some representative edges for further analysis. The discovery rate of these representative edges are calculated in Table 3. Compared with CGC, Lasso-CGC and NCGC, GLasso-NCGC demonstrates a considerable advantage for both these true positives and false positives.

With the quantitative comparison of these methods in Fig. 6(b), we can observe GLasso-NCGC get the highest Average AUROC and Average AUPR with the smallest standard deviations (resp. 0.9905 (0.0234) and 0.9037 (0.0235)).



**Figure 6.** Simulation of BRN. The number of multi-realizations is set as 100. (a) True network. (b) Average AUROC and Average AUPR with standard deviation over multi-realizations. (c) Discovery rate matrixes inferred from multi-realizations.

		CGC	Lasso-CGC	NCGC	GLasso-NCGC
True Positives	6 → 1	20%	27%	36%	83%
	4 → 2	10%	12%	65%	80%
	5 → 3	15%	18%	78%	83%
False Positives	4 → 1	93%	90%	81%	4%
	5 → 2	86%	88%	82%	13%
	6 → 3	84%	78%	44%	1%

**Table 3.** Comparison of discovery rate of representative edges extracted from true adjacency matrix in BRN simulation.

Here, we also explore the performance of robust estimation with different Granger Causality methods. Figure 7 demonstrates the curves of Average AUROC and Average AUPR of different methods acrossing samples (20~80) and noise intensities  $\sigma$  (0.1~1) respectively. The values are obtained by averaging over 100 independent realizations.

#### S2: Gene Regulatory Networks (GRN).

Suppose that gene interactions can be modeled by Michaelis-Menten equation as follows.

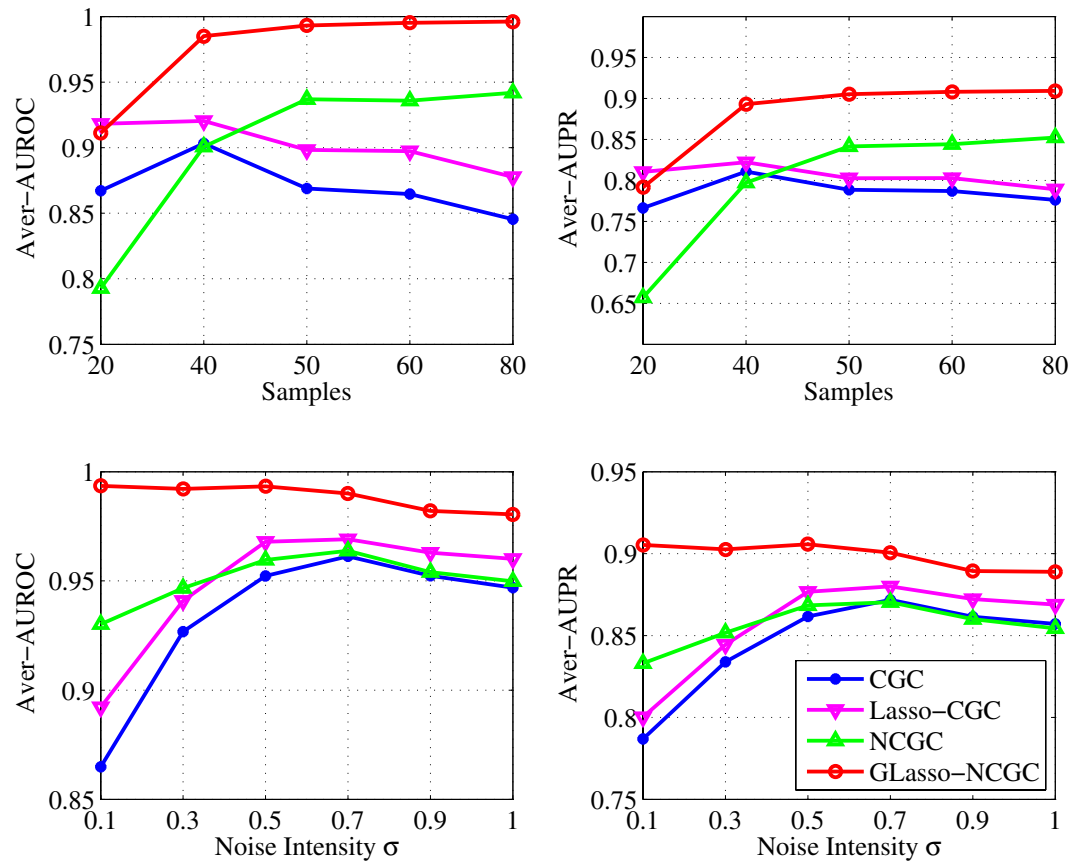
$$\dot{x}_i = -x_i^f + \sum_{j=1, j \neq i}^N b_{ij} \frac{x_j^h}{x_j^h + 1} + \varepsilon_i \quad (10)$$

where,  $f=1, h=2$ .

We first construct adjacent matrix  $\mathbf{B}$  with  $N$  nodes and simulate gene regulatory model with three classical types of networks, including random, small-world and scale-free networks (resp. RN, SW, SF). Then synthetic datasets of  $M$  samples are generated from the gene regulatory model of Equation (10). In order to reconstruct large-scale complex networks, multiple measurements from different conditions should be executed by adding perturbations in some different ways or with random initial values. Meanwhile, the outputs of model are supposed to be contaminated by gaussian noise. The number of multiple measurements is set as  $m$ .  $T$  time points are obtained at each measurement. As a result, data matrix with  $M \times N$  is collected ( $M = mT$ ).

Here, we take size 100 SW network with average degree  $\langle k \rangle = 5$  for detailed analysis. We set gaussian noise intensity as 0.1 and generated 400 samples with 100 multi-measurements (each measurement with 4 time points).





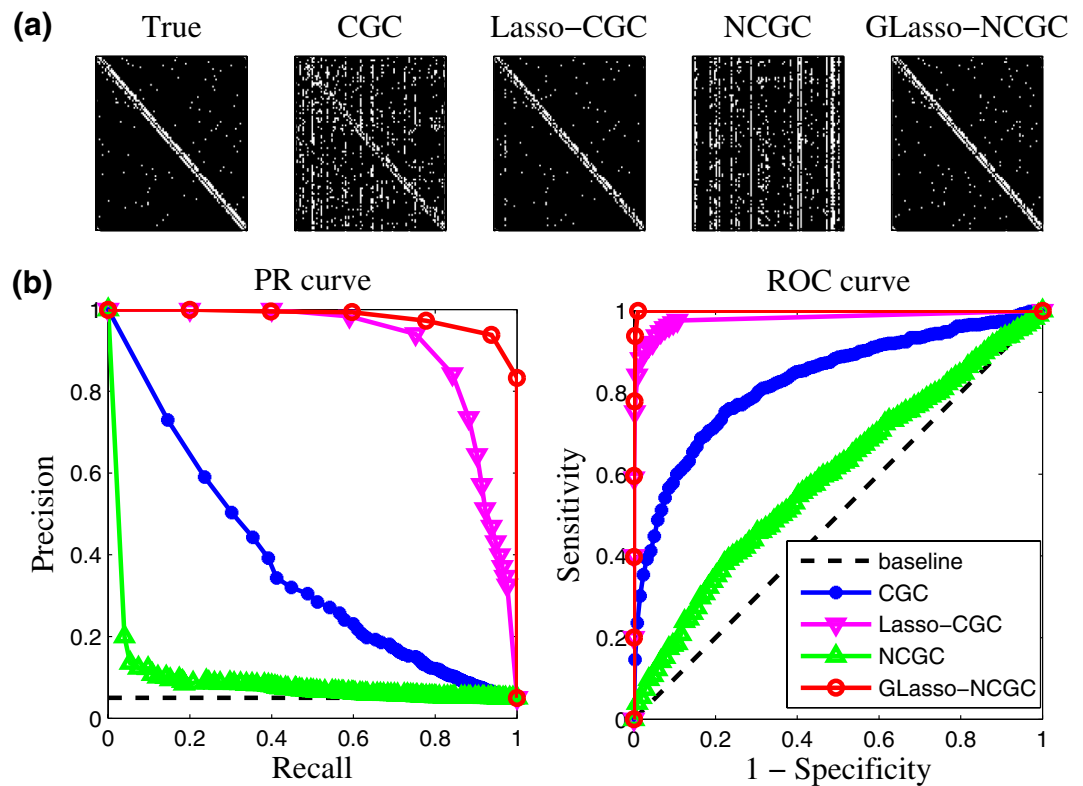
**Figure 7.** Simulation of BRN. Average AUROC and Average AUPR acrossing samples (20~80), noise intensities  $\sigma$  (0.1~1) respectively. The values of points in these figures are computed by averaging over 100 independent realizations.

True SW network are given in Fig. 8(a) together with reconstructed networks which are inferred by different methods. From Fig. 8(a), we can observe that the network inferred by GLasso-NCGC is most similar to the true network. To some extent, both CGC and Lasso-CGC have similar structures compared with the true network. However, they generated lots of false positives. Meanwhile, it can be seen that NCGC almost failed in this case. Because there are more parameters involved in NCGC model and the ratio of the number of samples to the number of parameters is relatively smaller than that of CGC model. Actually, the best performance of GLasso-NCGC proves that regularization-based methods have superiority in the case of more parameters and relatively smaller samples. In Fig. 8(b), we plot PR curves and ROC curves of CGC, Lasso-CGC, NCGC and GLasso-NCGC. The almost perfect reconstruction is ensured by GLasso-NCGC. Furthermore, reconstructed values of elements in different inferred matrixes for size 100 SW network can be shown in Fig. 9(a). Red points are existent edges and green points are nonexistent edges. CGC, Lasso-CGC and NCGC cannot make a distinction between existent and nonexistent edges, because there are so many overlaps between red and green points without a certain threshold ( $P_{val}$ ) for separation. However, GLasso-NCGC shows a vast and clear gap between existent and nonexistent edges. Based on the consideration of robustness, we next apply our method with respect to different intensities of noise  $\mathcal{N}(0, 0.3^2)$  and  $\mathcal{N}(0, 0.5^2)$  respectively. From Fig. 9(b), GLasso-NCGC also maintains a relatively good performance with strong measurement noise.

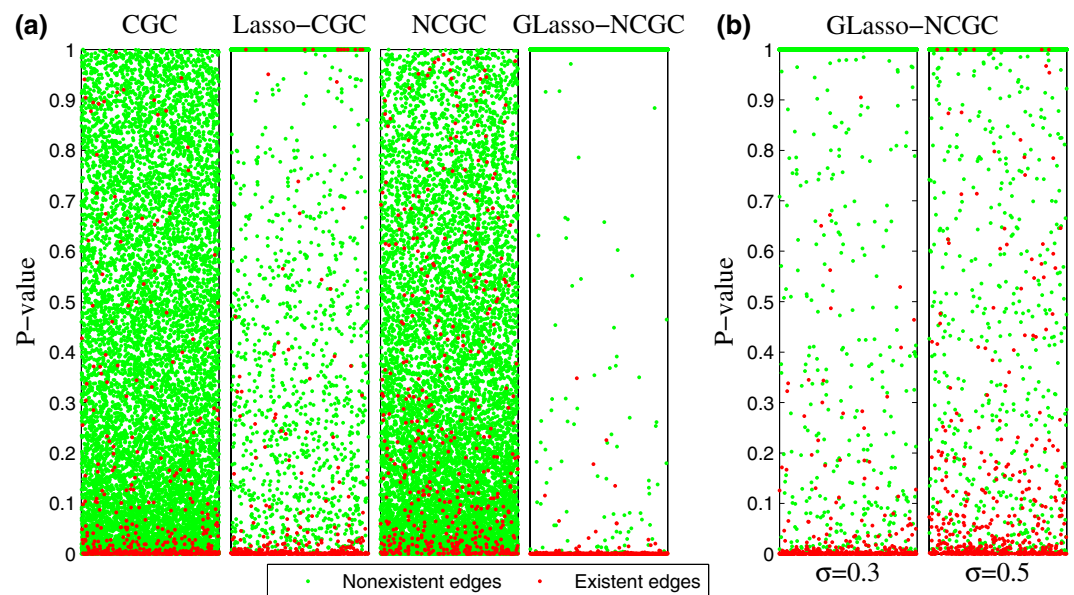
Then we explore the performance of these methods influenced by different average degrees, noise intensities and amounts of data simultaneously for different types of networks. The detailed results are shown in Table 4. In general, all of the results demonstrate the optimality and robustness of GLasso-NCGC.

In order to compare the computational requirement by these methods, we simulate the SW and SF networks for comparison ( $N = 100$ ,  $\langle k \rangle = 5$ ,  $M = 400$ ). And we calculate the average computational time over 10 independent realizations as shown in Table 5. The specifications of the computer used to run the simulations are as follows. Matlab version: R2013a (64 bit); Operating system: Windows 10 (64 bit); Processor: Intel(R) Core(TM) i7-4770 CPU @ 3.40GHZ 3.40GHZ; RAM: 16GB.

**DREAM3 Challenge4.** Dialogue on Reverse Engineering Assessment and Methods (DREAM) projects establish the general framework for the verification of various algorithms, which have the broad applications in many areas of research (<http://dreamchallenges.org/>). There are so many challenges in DREAM projects. Here, DREAM3 Challenge4 is used for the further assessment. The aim of DREAM3 Challenge4 is to infer gene regulatory networks with multiple types of datasets.



**Figure 8.** Simulation of GRN model on SW network ( $N=100$ ,  $\langle k \rangle=5$ ,  $M=400$ ). (a) True network and reconstructed network inferred by different methods with noise  $\mathcal{N}(0,0.1^2)$ . (b) PR curves and ROC curves with noise  $\mathcal{N}(0,0.1^2)$ .



**Figure 9.** Simulation of GRN model on SW network ( $N=100$ ,  $\langle k \rangle=5$ ,  $M=400$ ). (a) Reconstructed values of elements in different inferred matrixes with noise  $\mathcal{N}(0,0.1^2)$ . (b) Reconstructed values of elements in recovered matrixes by GLasso-NCGC with noise  $\mathcal{N}(0,0.3^2)$  and  $\mathcal{N}(0,0.5^2)$  respectively.

DREAM3 Challenge4 has three sub-challenges corresponding to gene regulatory networks with size 10, size 50 and size 100 nodes respectively. In our work, time-series of size 50 and size 100 are used, together with their gold standard networks. There are five sub-networks of Yeast or *E. coli*, all of which are with size 50 and size 100 respectively. Meanwhile, time-series datasets of multiple measurements are acquired under some different

Type	N	$\langle k \rangle$	$\sigma$	M	CGC	Lasso-CGC	NCGC	GLasso-NCGC
					Aver-AUROC/Aver-AUPR (SD/SD)			
RN	100	5	0.1	400	0.6367/0.2257	0.8118/0.6441	0.5472/0.1369	<b>0.9315/0.7719</b>
					(0.0290/0.0395)	(0.0072/0.0181)	(0.0112/0.0075)	(0.0083/0.0227)
		5	0.1	600	0.8022/0.4947	0.8989/0.7981	0.6321/0.2322	<b>0.9755/0.9072</b>
					(0.0234/0.0405)	(0.0102/0.0213)	(0.0256/0.0317)	(0.0036/0.0128)
		5	0.3	400	0.6518/0.2563	0.7963/0.5994	0.5530/0.1389	<b>0.9186/0.7476</b>
					(0.0180/0.0114)	(0.0129/0.0320)	(0.0132/0.0042)	(0.0052/0.0102)
8	0.1	400	0.7394/0.3308	0.9281/0.8382	0.5966/0.1341	<b>0.9714/0.8967</b>		
			(0.0233/0.0319)	(0.0104/0.0162)	(0.0192/0.0188)	(0.0034/0.0150)		
SW	100	5	0.1	400	0.8648/0.4658	0.9769/0.8823	0.6065/0.0929	<b>0.9996/0.9895</b>
					(0.0213/0.0503)	(0.0044/0.0142)	(0.0185/0.0140)	(0.0001/0.0034)
		5	0.1	600	0.8895/0.5384	0.9716/0.8826	0.6949/0.1684	<b>0.9997/0.9919</b>
					(0.0311/0.0733)	(0.0070/0.0275)	(0.0279/0.0255)	(0.0001/0.0027)
		5	0.3	400	0.8527/0.4206	0.9643/0.8342	0.6072/0.0894	<b>0.9975/0.9570</b>
					(0.0145/0.0312)	(0.0073/0.0243)	(0.0113/0.0093)	(0.0004/0.0085)
8	0.1	400	0.6500/0.1970	0.9538/0.8791	0.5476/0.1030	<b>0.9950/0.9312</b>		
			(0.0325/0.0309)	(0.0082/0.0196)	(0.0180/0.0081)	(0.0012/0.0118)		
SF	100	5	0.1	400	0.8304/0.4703	0.9358/0.8377	0.7680/0.2798	<b>0.9492/0.8812</b>
					(0.0128/0.0278)	(0.0055/0.0082)	(0.0217/0.0320)	(0.0077/0.0174)
		5	0.1	600	0.9216/0.6986	0.9498/0.8854	0.8939/0.5795	<b>0.9638/0.9056</b>
					(0.0098/0.0183)	(0.0025/0.0047)	(0.0110/0.0515)	(0.0050/0.0112)
		5	0.3	400	0.8464/0.4595	0.9233/0.8026	0.6929/0.1941	<b>0.9443/0.8415</b>
					(0.0198/0.0270)	(0.0057/0.0099)	(0.0166/0.0150)	(0.0078/0.0100)
8	0.1	400	0.7144/0.3078	0.8322/0.6410	0.6409/0.1786	<b>0.9095/0.7848</b>		
			(0.0170/0.0206)	(0.0086/0.0211)	(0.0266/0.0279)	(0.0072/0.0230)		

**Table 4.** The Average AUROC and Average AUPR with standard deviation in parentheses of different methods based on three types of networks, random (RN), small-world (SW) and scare-free (SF). The values are computed by averaging over 10 independent realizations. The results of different methods at different conditions are explored (type,  $N$ ,  $\langle k \rangle$ ,  $\sigma$  and  $M$ ). Here,  $N$  is the size of network,  $\langle k \rangle$  is average degree of network,  $\sigma$  is gaussian noise intensity,  $M$  is the number of samples. The highest scores of the Average AUROC and Average AUPR are highlighted.

Types	CGC	Lasso-CGC	NCGC	GLasso-NCGC
SW	12.7196	4.9657	30.0955	5.4432
SF	12.6883	4.0068	29.0900	6.0324

**Table 5.** The average computational time (in sec.) of different methods over 10 independent realizations in the simulation of GRN model on SW and SF networks.

conditions. For the networks of size 50 and size 100, the number of measurements are set as 23 and 46 respectively (each measurement with 21 time points).

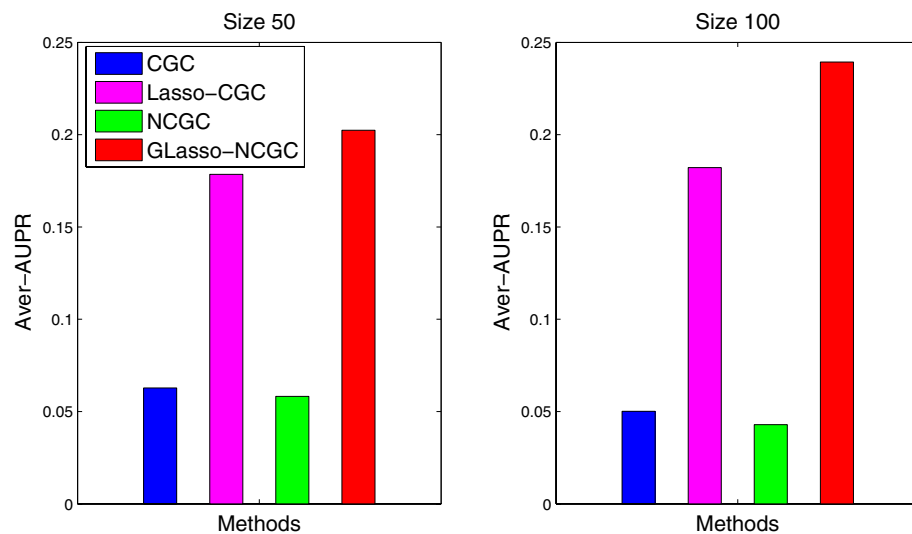
To assess these results of different methods, we also calculate the AUROC and AUPR. We can discover that some methods get the highest AUROCs, while their AUPRs are pretty small. In many cases, good AUROC might accompany by a low precision because of a large ratio of FP/TP. Thus, AUPR is taken as the final evaluation metric. In Table 6, GLasso-NCGC gets the best AUPR in all the datasets only except for size 50 network of Yeast1. Furthermore, the average AUPRs of these methods are subsequently computed and plotted as shown in Fig. 10. Finally, GLasso-NCGC is also found to execute optimally with the highest average AUPR.

## Conclusions

Reconstructing complex network is greatly useful for us to analyze and master the collective dynamics of interacting nodes. In our work, with getting multi-source datasets based on the data-fusion strategy, the new method namely group lasso nonlinear conditional granger causality (GLasso-NCGC) is proposed for network recovery with time-series. The evaluations of performance address that GLasso-NCGC is superior to other mentioned methods. Effects of data size and noise intensity are also discussed. Although the models or applications we mainly focus on here are biochemical reaction network and gene regulatory network, our method can be also used to other complex networked systems, such as kuramoto oscillator network and mutualistic network<sup>40,42</sup>. Here, it is also important to remember that we just adopt the model of first-order nonlinear conditional granger

	Methods	Ecoil1	Ecoil2	Yeast1	Yeast2	Yeast3
	AUROC/AUPR					
Size 50	CGC	0.5180/0.0290	0.5249/0.0388	0.6208/0.0645	0.5285/0.0701	0.5759/0.1115
	Lasso-CGC	0.5643/0.1564	0.5394/0.1060	0.5684/ <b>0.2084</b>	0.5286/0.1654	0.5470/0.2565
	NCGC	0.5492/0.0275	0.5590/0.0437	0.5868/0.0519	0.5227/0.0720	0.5357/0.0961
	GLasso-NCGC	0.5665/ <b>0.1686</b>	0.5316/ <b>0.1351</b>	0.5766/0.2057	0.5272/ <b>0.2299</b>	0.5337/ <b>0.2728</b>
Size 100	CGC	0.5864/0.0335	0.5730/0.0242	0.6196/0.0652	0.5234/0.0615	0.5474/0.0664
	Lasso-CGC	0.5936/0.1660	0.5841/0.1920	0.5807/0.2283	0.5270/0.1757	0.5162/0.1488
	NCGC	0.5629/0.0197	0.5266/0.0179	0.6132/0.0537	0.5348/0.0553	0.5239/0.0681
	GLasso-NCGC	0.5732/ <b>0.2016</b>	0.5696/ <b>0.2192</b>	0.5532/ <b>0.2953</b>	0.5254/ <b>0.2438</b>	0.5116/ <b>0.2370</b>

**Table 6.** Comparison of different methods on DREAM3 Challenge4 In-Silico networks with Size 50 and Size 100.



**Figure 10.** Average AUPR of different methods based on DREAM3 Challenge4 In-Silico networks with Size 50 and Size 100.

causality (NCGC), which is in accordance with the characteristics of dynamic systems governed by state-space equations. For the time-delayed dynamic systems, such as coupled Mackey-Glass system<sup>28</sup>, the framework of our method can be flexibly extended to higher-order NCGC model with group lasso regression, which can be waited for the prospective researches. In the information era with explosive growth of data, our proposed method provides a general and effective data-driven framework for nonlinear network reconstruction, especially for the complex networked systems that can be turned into the form of  $Y = \Phi(X)A$ .

## References

- Reichman, O. J., Jones, M. B. & Schildhauer, M. P. Challenges and opportunities of open data in ecology. *Science* **331**, 703–705, doi:10.1126/science.1197962 (2011).
- Marx, V. Biology: The big challenges of big data. *Nature* **498**, 255–260, doi:10.1038/498255a (2013).
- Bareinboim, E. & Pearl, J. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* **113**, 7345–7352, doi:10.1073/pnas.1510507113 (2016).
- Reshef, D. N. *et al.* Detecting novel associations in large data sets. *Science* **334**, 1518–24, doi:10.1126/science.1205438 (2011).
- Sugihara, G. & Munch, S. Detecting causality in complex ecosystems. *Science* **338**, 496–500, doi:10.1126/science.1227079 (2012).
- Zhao, J., Zhou, Y., Zhang, X. & Chen, L. Part mutual information for quantifying direct associations in networks. *Proceedings of the National Academy of Sciences* **113**, 5130–5135, doi:10.1073/pnas.1522586113 (2016).
- Yin, Y. & Yao, D. Causal inference based on the analysis of events of relations for non-stationary variables. *Scientific Reports* **6**, 29192, doi:10.1038/srep29192 (2016).
- Margolin, A. A. *et al.* Reverse engineering cellular networks. *Nature protocols* **1**, 662–671, doi:10.1038/nprot.2006.106 (2006).
- de Juan, D., Pazos, F. & Valencia, A. Emerging methods in protein co-evolution. *Nature Reviews Genetics* **14**, 249–261, doi:10.1038/nrg3414 (2013).
- Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nature methods* **9**, 796–804, doi:10.1038/nmeth.2016 (2012).
- Wang, W.-X., Lai, Y.-C. & Grebogi, C. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports* **644**, 1–76, doi:10.1016/j.physrep.2016.06.004 (2016).
- Chai, L. E. *et al.* A review on the computational approaches for gene regulatory network construction. *Computers in biology and medicine* **48**, 55–65, doi:10.1016/j.combiomed.2014.02.011 (2014).

13. Villaverde, A. F., Ross, J. & Banga, J. R. Reverse engineering cellular networks with information theoretic methods. *Cells* **2**, 306–329, doi:10.3390/cells2020306 (2013).
14. Zou, C. & Feng, J. Granger causality vs. dynamic bayesian network inference: a comparative study. *BMC bioinformatics* **10**, 122, doi:10.1186/1471-2105-10-122 (2009).
15. Wang, W.-X., Yang, R., Lai, Y.-C., Kovanis, V. & Harrison, M. A. F. Time-series-based prediction of complex oscillator networks via compressive sensing. *EPL* **94**, 48006, doi:10.1209/0295-5075/94/48006 (2011).
16. Shandilya, S. G. T. M. Inferring network topology from complex dynamics. *New Journal of Physics* **13**, 013004, doi:10.1088/1367-2630/13/1/013004 (2011).
17. Han, X., Shen, Z., Wang, W. X., Lai, Y. C. & Grebogi, C. Reconstructing direct and indirect interactions in networked public goods game. *Scientific Reports* **6**, 30241, doi:10.1038/srep30241 (2016).
18. Mei, G., Wu, X., Chen, G. & Lu, J. A. Identifying structures of continuously-varying weighted networks. *Scientific Reports* **6**, 26649, doi:10.1038/srep26649 (2016).
19. Iglesiasmartinez, L. F., Kolch, W. & Santra, T. Bgrmi: A method for inferring gene regulatory networks from time-course gene expression data and its application in breast cancer research. *Scientific Reports* **6**, 37140, doi:10.1038/srep37140 (2016).
20. Fujita, A. *et al.* Modeling gene expression regulatory networks with the sparse vector autoregressive model. *BMC Systems Biology* **1**, 39, doi:10.1186/1752-0509-1-39 (2007).
21. Opgen-Rhein, R. & Strimmer, K. Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC bioinformatics* **8**, S3, doi:10.1186/1471-2105-8-S2-S3 (2007).
22. Michailidis, G. & D'Alché-Buc, F. Autoregressive models for gene regulatory network inference: Sparsity, stability and causality issues. *Mathematical biosciences* **246**, 326–334, doi:10.1016/j.mbs.2013.10.003 (2013).
23. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society* 424–438 (1969).
24. Siggiridou, E. & Kugiumtzis, D. Granger causality in multivariate time series using a time-ordered restricted vector autoregressive model. *IEEE Transactions on Signal Processing* **64**, 1759–1773, doi:10.1109/TSP.2015.2500893 (2016).
25. Marinazzo, D., Pellicoro, M. & Stramaglia, S. Kernel method for nonlinear granger causality. *Physical Review Letters* **100**, 144103, doi:10.1103/PhysRevLett.100.144103 (2008).
26. Ancona, N., Marinazzo, D. & Stramaglia, S. Radial basis function approach to nonlinear granger causality of time series. *Physical Review E* **70**, 056221, doi:10.1103/PhysRevE.70.056221 (2004).
27. Fujita, A. *et al.* Modeling nonlinear gene regulatory networks from time series gene expression data. *Journal of bioinformatics and computational biology* **6**, 961–979, doi:10.1142/S0219720008003746 (2008).
28. Kugiumtzis, D. Direct-coupling information measure from nonuniform embedding. *Physical Review E* **87**, 062918, doi:10.1103/PhysRevE.87.062918 (2013).
29. Shojaie, A. & Michailidis, G. Discovering graphical granger causality using the truncating lasso penalty. *Bioinformatics* **26**, i517–i523, doi:10.1093/bioinformatics/btq377 (2010).
30. Arnold, A., Liu, Y. & Abe, N. Temporal causal modeling with graphical granger methods. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, 66–75 (2007).
31. Lozano, A. C., Abe, N., Liu, Y. & Rosset, S. Grouped graphical granger modeling for gene expression regulatory networks discovery. *Bioinformatics* **25**, i110–i118, doi:10.1093/bioinformatics/btp199 (2009).
32. Yuan, M. & Lin, Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society* **68**, 49–67, doi:10.1111/rssb.2006.68.issue-1 (2006).
33. Bolstad, A., Van Veen, B. D. & Nowak, R. Causal network inference via group sparse regularization. *IEEE Transactions on Signal Processing A Publication of the IEEE Signal Processing Society* **59**, 2628–2641, doi:10.1109/TSP.2011.2129515 (2011).
34. He, D., Kuhn, D. & Parida, L. Novel applications of multitask learning and multiple output regression to multiple genetic trait prediction. *Bioinformatics* **32**, i37–i43, doi:10.1093/bioinformatics/btw249 (2016).
35. Puniyani, K., Kim, S. & Xing, E. P. Multi-population gwa mapping via multi-task regularized regression. *Bioinformatics* **26**, i208–16, doi:10.1093/bioinformatics/btq191 (2010).
36. Argyriou, A., Evgeniou, T. & Pontil, M. Convex multi-task feature learning. *Machine Learning* **73**, 243–272, doi:10.1007/s10994-007-5040-8 (2008).
37. Chang, Y. H., Gray, J. W. & Tomlin, C. J. Exact reconstruction of gene regulatory networks using compressive sensing. *BMC Bioinformatics* **15**, 400, doi:10.1186/s12859-014-0400-4 (2014).
38. Han, X., Shen, Z., Wang, W.-X. & Di, Z. Robust reconstruction of complex networks from sparse data. *Physical review letters* **114**, 028701, doi:10.1103/PhysRevLett.114.028701 (2015).
39. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **113**, 3932–3937, doi:10.1073/pnas.1517384113 (2016).
40. Pan, W. & Yuan, Y. A sparse bayesian approach to the identification of nonlinear state-space systems. *IEEE Transactions on Automatic Control* **61**, 182–187, doi:10.1109/TAC.2015.2426291 (2016).
41. Yang, G., Wang, L. & Wang, X. Network reconstruction based on grouped sparse nonlinear graphical granger causality. *35th Chinese Control Conference*, 2229–2234 (2016).
42. Gao, J., Barzel, B. & Barabási, A. Universal resilience patterns in complex networks. *Nature* **530**, 307–312, doi:10.1038/nature16948 (2016).
43. Khan, J., Bouaynaya, N. & Fathallah-Shaykh, H. M. Tracking of time-varying genomic regulatory networks with a lasso-kalman smoother. *Eurasip Journal on Bioinformatics and Systems Biology* **2014**, 3, doi:10.1186/1687-4153-2014-3 (2014).

## Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61473189 and 61374176, the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (No. 61521063).

## Author Contributions

G.Y. designed and conducted the research. G.Y., L.W. and X.W. discussed the results and wrote the manuscript. All authors reviewed the manuscript.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017