CSIAM Trans. Appl. Math.
doi: 10.4208/csiam-am.2020-0184

Vol. **2**, No. 4, pp. 680-696
December 2021

Detecting High-Dimensional Causal Networks using Randomly Conditioned Granger Causality

Huanfei Ma[1,*], Siyang Leng[2], and Luonan Chen[3,4,*]

[1] *School of Mathematical Sciences, Soochow University, Suzhou 215006, China.*
[2] *Institute of AI and Robotics, Academy for Engineering and Technology, Fudan University, Shanghai 200433, China.*
[3] *State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China.*
[4] *School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China.*

Received 20 August 2020; Accepted 29 April 2021

Abstract. Reconstructing faithfully causal networks from observed time series data is fundamental to revealing the intrinsic nature of complex systems. With the increase of the network scale, indirect causal relations will arise due to causation transitivity but existing methods suffer from dimension curse in eliminating such indirect influences. In this paper, we propose a novel technique to overcome the difficulties by integrating the idea of randomly distributed embedding into conditional Granger causality. Validated by both benchmark and synthetic data sets, our method demonstrates potential applicability in reconstructing high-dimensional causal networks based only on a short-term time series.

AMS subject classifications: 62D20, 37M10

Key words: Network reconstruction, Granger causality, conditional causality, randomly distributed embedding.

1 Introduction

Interactions between variables in a complex system constitute the function basis of the system. Thus correctly identifying such interactions are fundamental underpinnings in natural and engineering systems, particularly when systems details are typically unknown but only time series are observed. A straightforward and widely used way to infer linear dependence between variables is computing correlation between time series

*Corresponding author. *Email addresses:* `hfma@suda.edu.cn` (H. Ma), `lnchen@sibs.ac.cn` (L. Chen), `syleng@fudan.edu.cn` (S. Leng)

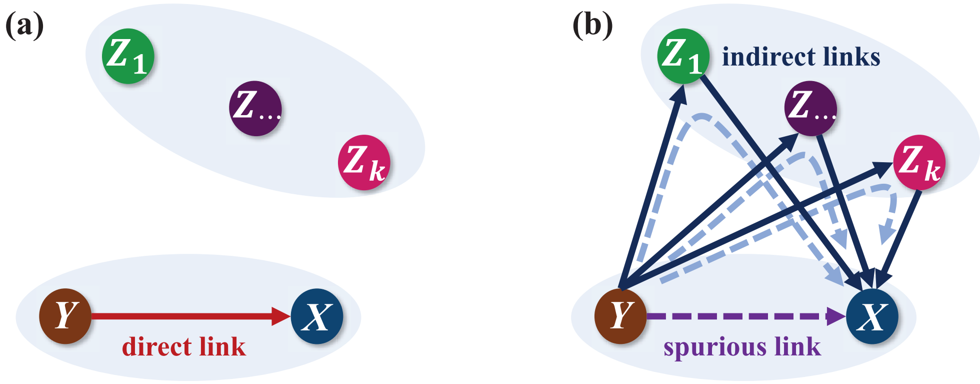

Figure 1: Direct causation versus indirect causation. Confounding variables $(Z_1,\cdots,Z_k)$ induce spurious causal links between Y and X (see (b)), which should be distinguished from true direct links (see (a)).

data, but the relationship revealed by correlation or association has neither direction nor mechanism [1]. On the other hand, the causal interactions, directional and cause-effect mechanism revealing, have attracted great attentions in the recent decades, and a number of data-driven methods have been fruitfully developed to identify such interactions. The representative methods include the celebrated Granger causality [2, 3], transfer entropy [4–9], and the mutual cross mapping [10–15].

If the system contains only two variables, the mutual causal relation between them could be straightforwardly detected by applying these detection methods. However, the situation becomes much more complicated when a large number of variables are interacting with each other in a networked fashion, where variables become connected nodes in a network. Due to the generic phenomenon of causation transitivity, the indirect causal relation will arise even for two nodes that are not directly connected, as illustrated in Fig. 1, and simply detecting the causal relation for two nodes in a pair-wise manner may yield false and redundant connections. Thus correctly distinguishing direct causal links from indirect causal influences is of paramount importance to reconstruct the underlying true causal network. To cope with this issue, several conditional derivatives of causality detection methods, such as the concept of the conditional Granger causality (cGC) [16, 17], the partial transfer entropy [18–20], and the recently proposed partial cross mapping [21], have been developed.

The basic idea of conditional causality lies in the fact that all the other potentially influential variables, or known as the confounding variables, should be taken into consideration while evaluating the direct causal relation between two underlying variables. Though the idea is straightforward and various successful applications have been reported [22–25], the application of conditional causal relation detection methods are mainly restricted to systems with small number of variables. When the number of variables increases, the computational complexity increases rapidly and several problems such as overfitting and requirement of impractical amount of training data, or simply known