Applications of Granger Causality to Biological Data

A dissertation presented

By

Cunlu Zou

to

The Department of Computer Science

in partial fulfilment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Computational Biology

University of Warwick

Coventry,UK

December 2010

Thesis supervisor

Prof. Jianfeng Feng

Author

Cunlu Zou

Applications of Granger Causality to Biological Data

Abstract

In computational biology, one often faces the problem of deriving the causal relationship among different elements such as genes, proteins, metabolites, neurons and so on, based upon multi-dimensional temporal data. In literature, there are several well-established reverse-engineering approaches to explore causal relationships in a dynamic network, such as ordinary differential equations (ODE), Bayesian networks, information theory and Granger Causality. To apply the four different approaches to the same problem, a key issue is to choose which approach is used to tackle the data, in particular when they give rise to contradictory results.

In this thesis, I provided an answer by focusing on a systematic and computationally intensive comparison between the two common approaches which are dynamic Bayesian network inference and Granger causality. The comparison was carried out on both synthesized and experimental data. It is concluded that the dynamic Bayesian network inference performs better than

the Granger causality approach, when the data size is short; otherwise the Granger causality approach is better.

Since the Granger causality approach is able to detect weak interactions when the time series are long enough, I then focused on applying Granger causality approach on real experimental data both in the time and frequency domain and in local and global networks. For a small gene network, Granger causality outperformed all the other three approaches mentioned above. A global protein network of 812 proteins was reconstructed, using a novel approach. The obtained results fitted well with known experimental findings and predicted many experimentally testable results. In addition to interactions in the time domain, interactions in the frequency domain were also recovered.

In addition to gene and protein data, Granger causality approach was also applied on Local Field Potential (LFP) data. Here we have combined multiarray electrophysiological recordings of local field potentials in both right inferior temporal (rIT) and left IT (lIT) and right anterior cingulate (rAC) cortices in sheep with Granger causality to investigate how anaesthesia alters processing during resting state and exposure to pictures of faces. Results from both the time and frequency domain analyses show that loss of consciousness during anaesthesia is associated with a reduction/disruption of feed forward open-loop cortico-cortical connections and a corresponding increase in shorter-distance closed loop ones.

Contents

Citations to Previously Published Work

Large portions of Chapters 3, 4 and 5 appeared in the following four papers:

Zou CL, Denby K.J., Feng J.F. (2009) **Granger causality vs. Dynamic Bayesian network inference: A Comparative Study** BMC Bioinformatics vol. 10:122 doi:10.1186/1471-2105-10-122. (most viewed paper in past 30 days in the journal, flagged as 'highly accessed paper', IF=3.8)

Zou CL, Kendrick KM, Feng JF (2009) **The Fourth Way: Granger Causality is better than the three other Reverse-engineering Approaches** COMMENTS ON Cell (http://www.cell.com/comments/S0092-8674(09)00156-1.)

Zou CL, Ladroue C, Gou SX J.F. Feng,, (2010). **Indentifying interactions in the time and frequency domains in local and global networks** BMC Bioinformatics vol. 11: 337, (IF =3.4).

Zou CL, Nicol A, Kendrick KM, Feng J.F. (under revision). **Loss of consciousness due to anaesthesia is associated with altered long distance functional connectivity in the cortex**.

Acknowledgments

The past three years' PhD study was a really magnificent and also challenging experience to me. In all these years, I kept going forward for my target, step by step, accompanied with bitterness, frustration but also a sense of achievement after successes of novel discovery in my research area. What I have learned is not just the academic knowledge, but also the valuable experience of working in an advanced research group of the University of Warwick. Of course, there are many people who gave me the supports and kindly help during the study time. It was hardly possible for me to complete my doctoral work without their precious support.

First of all, it is a pleasure to give my sincere thanks to my supervisor Professor Jianfeng Feng, who accepted me as a PhD student and provided his valuable guidance., cheerful enthusiasm and great encouragement that I was able to complete my research work in a respectable manner. His keen and vigorous academic observation and wide knowledge enlightens me not only in this thesis but also in my future study. He has also organized the most communicative, friendly and collaborative research group, which is comprised by his students and post-docs. I also received a lot benefits from these people. They are not only my colleagues but also good friends in my life. There are many memorable moments with them. We selflessly shared our thoughts, findings and

novel ideas in the regular organised group seminars. I really enjoyed the three years' study and also felt privileged to collaborate with them.

I would also like to express my deep gratitude to Computer Science Department, University of Warwick. I have been studying in the Computer Science Department since I was an undergraduate student. During my undergraduate study, I learned not only the fundamental knowledge and skills in mathematics, statistics, programming and computing, but also the ability of working as a scientist. My undergraduate study helped and provided the direction to my following PhD study and also my life. For my PhD study, I received three years' financial support by Computer Science Department, through the generosity of the university and my supervisor.

It is also a pleasure to thank my parents for their selfless loving support and great confidence in me all the time. They provided the financial support for my long overseas study since 2001 so that I have the opportunity to be educated in one of the top universities around the world.

Dedicated to my mother, Shuhua, Zhang

my father, Deyong, Zou

Chapter 1

Introduction

Currently, biological measurement techniques produce massive quantities of data from genes, proteins, metabolites, neurons, brain areas and so on. It is a key issue to analyse the pattern of element interactions based upon multi-dimensional spatial and temporal data. It has been recognized that most interactions in biology are directional. To be able to assess the directionality of interactions, there are several widely used reverse-engineering approaches such as Granger causality, Bayesian network inference and so on. Firstly, this thesis seeks to find the best reverse-engineering approaches, which is shown to be Granger causality, by focusing on a systematic and computationally intensive comparison. Secondly, a novel approach called global Granger causality is introduced for deriving large network and its applications to gene and protein data. Finally, some interesting results are derived by using a Granger causality approach for local field potential data recorded from sheep brain.

1.1 Review

1.1.1 Networks describe biological principles and mechanisms.

In a topological sense, a network is a set of nodes and a set of directed or undirected edges between the nodes. There exist many complex and large biological networks, such as protein-protein interactions, transcription factor-binding networks, metabolic networks, neural connections. The study of these biological networks has been rapidly expanding over the last decade. More and more large-scale networks have been identified. Some unexpected functions of individual components have also been revealed. Biological networks are widely found not only as descriptions of complex interactions, but as key determinants of function, mechanisms and principles in systems biology. The relationships can be observed at different scales: global scale and local scale. For example, the scale-free network (degree distribution follows a power law [Mason, 2007]) and small-world structure (most nodes can be reached from every other by small steps [Kleinberg, 2000]) are some new paradigms of networks on the global scale. On the other hand, the importance of modularity, motifs (frequently occurred patterns [Alon, 2007]) and hubs (high-degree node [Elena, 2008]) is observed at the local scale. It is still a key issue in systems biology to

find out how biological function is related to the structure and dynamics of biological networks.

These biological networks are significantly different from random networks and commonly observed to exhibit different properties in terms of their structure and architecture. In the literature, it is often reported that the placement of an element is related to its biological characteristics. For example, hubs or high-degree nodes in a protein interaction network which usually has the scale-free or small-world property have high probability to be essential [Elena, 2008; Nizar 2006]. In other words, most of the hubs are related to the cell fate: knockout of the corresponding hubs would cause the organism death. It has also been found that the network motifs, which recur within a network much more often than expected in a random network, can be considered as simple building blocks from the composed network. The functions associated with common network motifs in transcription networks have been well studied both theoretically and experimentally [Alon, 2006].

1.1.2 Biological network inference

With the development of biological measurement techniques and equipments, the great bulk of high-throughput biological data, such as microar-

ray data, can be collected from various biology sources in laboratories. It is increasingly recognized that theoretical methods are required to understand and make biological predictions of these multi-dimensional spatial and temporal biological data [Albert, 2007]. These theoretical methods can be classified into three systems biology topics:

- Statistical network inference (*i.e.* reconstructing the network of interactions among a set of biology entities)
- Network analysis (*i.e.* mining the information content of the network)
- Dynamic modelling (connecting the interactive network to the dynamic behaviour of the system).

In this thesis, one focused on the network inference and the network analysis was also carried out in **Chapter 4** and **Chapter 5**. For the network inference topic, although there are many types of biological networks, few of them are known in anything approaching their complete structures, even for the simplest cells. Therefore it is still a big challenge for scientists and biologists to reliably and accurately reconstruct biological network structures.

Correlation networks

As of 2007, correlation-based algorithms were the most fruitful network inference methods used in biological application for microarray data [Friston, 1994; Basso et al., 2005; Faith et al., 2007]. Clustering or some form of statistical classification is typically employed to perform an initial organization of the high-throughput data. Defining the interactions between each node more precisely can be done by using background literature or information in public databases, combined with clustering results. It can also be done by using correlation-based algorithms. In the analysis of neuroimaging time-series data, temporal correlations between spatially remote neurophysiological events are commonly calculated for deriving functional connectivity and checking if two elements are co-expressed across the data set. Co-expression can also be measured by mutual information [Bansal et al., 2007]. Normally, such methods can be only applied to derive undirected networks; it cannot be used to reveal a causal (or directed) relationship between entities. Another problem is that basic correlations can result from the confounding effects of stimulus-locked transients evoked by common afferent inputs [Friston, 1994] although the partial Granger causality could deal with this problem in some cases. To precisely understand the relationships between the nodes (*e.g.* genes), that is often required to find the underlying directed influence of each node on the others.

Causal networks

It has been the philosophers' dream to learn causal relations from raw data since the time of Hume (1711-1776), but the big problem is what the definition of causal relationships is. In neuroimaging, the causal (or effective) connectivity can be defined as the influence one neural system exerts over another [Friston, 1994]. Unfortunately, unlike functional connectivity, the definition of effective connectivity is not operational since the definition of 'influence' is hard to present in mathematics. The possible solution arose when the mathematical relationships between graphs and probabilistic dependencies came into light. Probabilities encode degrees of belief about events in the world and data are used to strengthen, update, or weaken those degrees of belief. The interpretation of causation can then be translated by using probability theory: A causes B if the occurrence of A increases the probability of B. This probability theory is currently the most common mathematical language of most disciplines that use causal modelling. In addition, scientists are also concerned with the relative strength of those connections and with ways of inferring those connections from noisy observations [Pearl, 2000]. To derive directional relationships, several well defined network inference approaches have been successfully applied in the literature. These approaches include ordinary differential equations (ODE), Bayesian network inference and Granger causality approach.

ODEs relate changes in nodes to each other and also to an external perturbation. The model consists of a differential equation for each of the node in the network, describing the changing rate of the node as a function of the other nodes and of the perturbation. The parameters of the equations have to be inferred from the expression data. [Della Gatta et al., 2008; Gardner et al., 2003]

A **Bayesian network** is a special case of a diagrammatic representation of probability distributions, called probabilistic graphical models. The Bayesian network graph model comprises nodes (also called vertices) connected by directed links (also called edges or arcs) without cycles (*i.e.* a node cannot directly or indirectly regulate itself). To learn the structure and the parameters for the Bayesian networks from a set of data, we should search the space of all possible graph representations, and find out which structure is most likely to have produced our data. [Jensen, 1996; Bach et al., 2004; Buntine, 1994; Friedman, 2004; di Bernardo et al., 2005]

The **Granger causality** concept was firstly derived in econometrics [Granger, 1980 & 1969] and is slowly moving into other fields, like systems biology [Guo 2008]. Granger causality is focused on measurement taken over time, and how they may influence one another. The directional concept of Granger causality comes from the relationship between past, present and future: the present and past influence future developments but not the other way around.

Thus an event taking place in the future cannot cause another event in the past or present. According to Granger causality, we can determine a causal influence if the past values of one time series contain information that helps predict another one above and beyond the information already contained in its past values. This idea was applied in a mathematical formulation based on linear regression modelling of stochastic processes [Granger, 1969]. The improvement of the prediction is measured by the variance of the error term. Due to the temporal ordering idea applied in Granger causality, it is obvious that Granger causality can only predict functional causal relationships for which cause and influence are sufficiently separated in time [Schelter, 2006]. In general, it is impossible to derive influences in a unique direction between variables at the same time point. Since the concept does not rely on a specification of a causal model, although Auto Regressive (AR) models are commonly used in literature it is particularly suited for empirical investigations of causal interactions.

1.2 Research topics of interest

1.2.1 Reverse-engineering approaches

Generally, methodologies based on ODEs reconstruct networks by estimating the parameters in a differential equation model. In contrast, Bayesian

network inference methods use the idea of probabilistic graphical models to uncover the network structure. Information theory approaches (e.g. mutual information) extract the network structure based on the probability that a pair of elements are co-expressed across a data set. Granger causality approaches derive the connective networks based on the improvement of prediction by incorporating additional past knowledge. Each of these approaches has its own benefits and disadvantages. For instance, ODE-based methods have advantages in cases where known biological perturbations are used in the experiments and information theoretic approaches are quite effective for learning large networks by using a relatively small amount of data. Given various types of methods for network reconstruction, it is quite important to evaluate the relative strengths, weaknesses and reliabilities of these methods through some comparative studies. These are described in the extensive literature on network inference [e.g. see Bonneau, 2008; Smet, 2010]. In addition, workshops and special events, such as DREAM (Dialogue for Reverse Engineering Assessments and Methods), have been organized to reveal the relative strengths of different approaches by presenting reverse-engineering challenges [Prill, 2001]. Currently, we still lack accurate and fair benchmarks or comparison standards to access and validate the diversity of reverse engineering approaches for both *in silico* and *in vivo* networks.

During the first year of my PhD, I performed a comparative study between the two most common approaches: Granger causality *vs.* dynamic Bayesian network inference. The comparative study was carried out by focusing on a systematic and computationally intensive comparison (more than 100 computers over a few weeks) between them on both synthesized and experimental data in the linear and nonlinear model (as described in **Chapter 3.2.1**). The results indicated that for a data set with a long enough sampling length the Granger causality approach produced more accurate results. This result was further confirmed by applying Granger causality in an *in vivo* gold standard (a known small synthetic gene network) proposed in a recent *Cell* paper [Camacho, 2009; Cantone, 2009]. Hence we could reasonably expect that the Granger causality approach is the best among the four approaches. These comparative studies provided us useful information on systematically understanding these reverse-engineering approaches and thus help us choose the proper one to apply to the experimental data in practice.

The conventional Granger causality is general, simple and can be easily applied to other types of temporal data. Furthermore, the advantage of the Granger causality over the other three approaches is the frequency decomposition, which is usually informative when we deal with temporal data. For example, in neurophysiology data, the brain employs different frequency bands to communicate between neurons and brain areas [Wu, 2007; Zhan, 2006]. How-

ever, there are still some limitations for the biological applications of Granger causality. For example, Granger causality is generally applied in a linear model. Although extensions to nonlinear cases now exist, these extensions can be more difficult to use in practice and their statistical properties are less known. In fact, the conventional Granger concept has been rapidly extended and improved in the last few years. The initial pair-wised Granger causality [Geweke, 1982] has now been extended to conditional Granger causality [Geweke, 1984], partial Granger causality [Guo, 2008], complex Granger causality [Ladroue, 2009] and global Granger causality (Described in **Chapter 4.2.2**) for dealing with different problems of network reconstruction.

1.2.2 Development of networks reconstruction methods

Hidden exogenous inputs

Although the number of elements that can be simultaneously recorded is rapidly increasing with the progress of biological measurement equipment, only a subset of all the relevant variables is able to be recorded. Moreover, due to the limitation of computational technology, normally only a subset of elements is selected for detailed analysis according to the classification results derived by some statistical clustering methods. Such a recorded subset of all the related

multi-variable time series normally contains some common environmental (or external) inputs. Attempting to identify causal interaction in such multi-variable biological time series can be under-determined and misleading by using conventional reverse-engineering such as conditional Granger causality when the confounding influence of exogenous input is strong enough. Hence eliminating the exogenous inputs is a critical issue while applying reverse-engineering approaches to any experimental data.

Some novel approaches have been developed for dealing with such exogenous inputs and latent variable problem, such as the partial Granger causality [Guo, 2008]. The idea of partial Granger causality was inspired by the definition of partial correlation. Assuming a small subset of variables receiving common inputs, the partial Granger causality allows us to reveal the underlying interactions among element in a network by eliminating the effect of exogenous inputs. During my three years' PhD research, partial Granger causality approach was tested in various types of experimental data (including genes, proteins, neurons etc.), my results further proved that Partial Granger causality performed better than conventional conditional Granger causality in most cases.

Large network reconstruction

Due to the complexity of biological processes, in order to capture the dynamics of complex systems and investigate the functions of genes and neurons in detail, it is much better to treat the network as a whole instead of analyzing a very limited portion of it [Basso, 2005]. Until now, most of the analysis tools currently used for the whole network are based on clustering algorithms. These algorithms attempt to locate groups of genes that have similar expression patterns over a set of experiments. Such analysis has proven to be useful in discovering genes that are co-regulated and/or have similar function [Cantone, 2009; Gatta 2008; Smet, 2010]. A more ambitious goal for analysis is revealing the structure of the transcriptional regulation process, for example, for a given transcription factor, could we find all its upstream and downstream transcription factors? This is clearly a challenging and fascinating problem. In this thesis, one proposed a novel approach to solve such problem (as described in **Chapter 4.2.2**).

Most popular approaches, such as Granger causality, are powerful in cases where the length of the time series is much larger than the number of variables, which is exactly the reverse of the situation commonly found in microarray experiments, for which relatively short time series are measured over tens of thousands of genes or proteins. The real difficulty comes from the fact

that when the dimension is larger than the length of time series, the design matrix of predictors is rectangular, having more columns than rows; in such case, the model is under-determined and cannot be uniquely fitted. Bayesian network is a graph-based model of joint multivariate probability distributions that captures properties of conditional independence between variables, but as it requires a large number of parameters and assumptions upon the variable distribution, it also quickly becomes intractable for large networks. Keeping these limitations in mind, it is still an important task to develop methodologies that are both statistically sound and computationally tractable to make a full use of the wealth of data now at our disposal. In this thesis, we proposed a novel approach called Global Granger Causality (GGC) to solve this problem. The advantage of such an approach is that it provides a less biased structure of the network by explicitly taking more sources into account.

1.2.3 Biological network analysis

The question of how biological function is related to the structure and dynamics of biological regulatory networks is central to computational biology research. Given biological interactions, networks may be analysed with respect to their structure and dynamical pattern, which are associated with phenotypes of interest. More and more biological circuits and their underlying biological

functions have been well studied. For example, Alon and his group proposed the new idea of *motif*: small circuits can be considered as simple building blocks from which the network is composed [Alon, 2006]. Both statistical and experimental works have been devoted to understanding various types of network motifs in gene regulatory networks and also other types of biological networks such as neuronal networks and protein interaction networks. In plant biology, the first large-scale *Arabidopsis* protein interaction network was derived from the knowledge of interacting *Arabidopsis* protein orthologs in *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Homo sapiens* [Geisler-Lee, 2007]. However, it is still not very clear how biological phenomena relate to the interactions between molecules in most organisms.

During my PhD research, some interesting circuits and networks associated with the corresponding biological phenomena have been recovered. For example, a well-known circadian circuit of 7 genes in Arabidopsis Leaf was reconstructed. Interestingly, we found the plant rewired its circadian circuit after infection. By assigning a dynamics to the network and trying to decipher the implications of the rewiring, one found that a critical gene was recruited to save the whole network (as described in **Chapter 3.3.2** and **Chapter 3.4.5**). All these works revealed the correlation between molecule interactions and the surprisingly complex biological functions they created.

1.3 Original work

1.3.1 Granger causality vs. dynamic Bayesian network inference: a comparative study

Although comparative studies have been described in the extensive literature on network inference methods, until now, no one tried to compare between two common network inference approaches: dynamic Bayesian network inference and Granger causality. In this thesis, I carried out a comparative study between the two approaches using both synthesized and experimental data. The aim was to find the performance differences, reliabilities and sensitivities between these two approaches. For synthesized data, a critical point of the data length was found: the Granger causality approach outperforms the dynamic Bayesian network inference when the data length is long enough, and vice versa. The Granger causality approach was more reliable for detecting weak interactions, but it also had over-fitting problems when the data length was too short.

1.3.2 Identifying interactions in the time and frequency domains in local and global networks – A Granger causality approach

Reverse-engineering approaches such as Bayesian network inference, ordinary differential equations (ODEs) and information theory are widely applied to deriving causal relationships among different molecules. These approaches have difficulty in cases where the length of the time series is smaller than the number variables, which is a common situation found in microarray experiments. In this thesis, we proposed a novel approach called Global Granger Causality for dealing with large network re-construction problem. By taking iterative steps, all indirect links will be removed from the initial network (including both direct and indirect links) derived by bivariate pair-wise Granger causality. It provides a less biased structure of the network due to hidden variables than in a small network by explicitly taking more sources. It also provides information on the ancestors and descendents of key elements. The results can then guide experimentalists to investigate the properties of a small subset of specific proteins.

1.3.3 Loss of consciousness due to anaesthesia is associated with decreased long and increased short-distance functional connectivity in the cortex

In the literature, it has been reported that loss of consciousness with anaesthesia, sleep or vegetative states may involve reduced functional cortical connectivity by disruption of long feed forward connections [Mashour, 2006; Akire et al., 2008]. Here we have combined multi-array array recordings of local field potentials in three different cortices in sheep with Granger causality to investigate how anaesthesia alters neural processing during resting state and visual stimulation. Some interesting phenomena for connections changing due to anaesthesia were observed. These results showed good agreements with previous results and reported in the literature [Ge, 2009; Ladroue, 2009]. In additional to the previously reported findings, our results illustrated that loss of consciousness during anaesthesia is associated with reductions in extrinsic long-distance open-loop cortico-cortical connections, and loss of their unidirectional flow, coupled with an increase in the strength of shorter-distance intrinsic closed loop connections.

1.4 Summary

The major contributions of this PhD study are the investigation and application of reverse-engineering approaches especially the Granger causality approach. The work includes a comparative study between Granger causality and dynamic Bayesian network inference, the proposal of a novel approach for dealing with large network re-construction problem, and the applications of Granger causality to sheep data to investigate the effect of anaesthesia.

Chapter 2 presents detailed background knowledge of the reverse-engineering approaches, Granger causality and Bayesian network inference. Chapters 3, 4 and 5 provide the main results from my three year PhD study. Chapter 3 shows a comparative study between Granger causality and Bayesian network inference, including results derived from both synthesized and experimental data. Chapter 4 proposes a novel approach based on the Granger causality approach for identifying interactions in the time and frequency domains in local and global networks. Chapter 5 studies the effect of anaesthesia to the connectivity network among cortices in sheep by using a Granger causality approach. The conclusion and further extensions are discussed in the last chapter.

Chapter 2

Background

In this chapter, one focus on two common approaches used for reverse-engineering task. One is the Granger causality approach, and the other is the dynamic Bayesian network inference approach. In this chapter, I would like to give detailed and systematic descriptions of these two common reverse-engineering approaches. For Granger causality approach, the development and evolution of Granger causality will be introduced. It includes methods of pair-wise Granger causality, conditional Granger causality, partial Granger causality and also the corresponding frequency domain analysis. A background description and implementation of Bayesian network inference will also be expressed in **Chapter 2.2**.

2.1 Granger causality

In order to evaluate the statistical interdependence among signals, we normally calculate the cross-correlation functions in the time domain and ordinary coherence functions in the frequency domain. However, in many situations, symmetric measures are not completely satisfactory, and further dissection of the directed interaction patterns among the recorded signals is often required. Recent work has begun to consider the causal influence from one element to another. The basic idea can be traced back to Wiener who conceived the notion that, if the prediction of one time series could be improved by incorporating the knowledge of a second one, then the second series is said to have a causal influence on the first [Granger, 1969]. This idea lacks the machinery for practical implementation. Granger later formalized the prediction idea in the context of linear autoregression (AR) models. The AR model is one of a group of linear prediction formulas that attempt to predict an output of a system based on the previous outputs [Ding, 1999]. In the AR model, the variance of the prediction error is used to test the prediction improvement.

2.1.1 Pair-wise Granger causality

For simplicity, we introduced the pair-wise Granger causality analysis for two time series. This framework can also be generalized to two sets of time series.

Time Domain Analysis

Assuming two stochastic processes X_t and Y_t, that are jointly stationary (i.e. the joint probability distribution does not change when shifted in time). Each process can be auto-regressively represented by using their past knowledge separately.

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{1i} X_{t-i} + \varepsilon_{1t} \ , & \text{var}(\varepsilon_{1t}) = \Sigma_1 \\ Y_t = \sum_{i=1}^{\infty} b_{1i} Y_{t-i} + \eta_{1t} \ , & \text{var}(\eta_{1t}) = \Gamma_1 \end{cases} \tag{2.1.1}$$

Jointly, they can be represented as

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{2i} X_{t-i} + \sum_{i=1}^{\infty} b_{2i} Y_{t-i} + \varepsilon_{2t} \\ Y_t = \sum_{i=1}^{\infty} c_{2i} X_{t-i} + \sum_{i=1}^{\infty} d_{2i} Y_{t-i} + \eta_{2t} \end{cases} \tag{2.1.2}$$

And the noise covariance matrix for the system can be represented as

$$\mathbf{S}_2 = \begin{bmatrix} \text{var}(\varepsilon_{2t}) & \text{cov}(\varepsilon_{2t}, \eta_{2t}) \\ \text{cov}(\eta_{2t}, \varepsilon_{2t}) & \text{var}(\eta_{2t}) \end{bmatrix} = \begin{bmatrix} \Sigma_2 & \Upsilon_2 \\ \Upsilon_2 & \Gamma_2 \end{bmatrix} \tag{2.1.3}$$

In Equations (2.1.2) and (2.1.3), the value of Σ_1 measures the accuracy of the autoregressive prediction of X_t based on its previous values, whereas the value of Σ_2 represents the accuracy of prediction of X_t based on the previous values of both X_t and Y_t. According to Wiener [Wiener, 1956] and Granger [Granger, 1969], if Σ_2 is less than Σ_1, then Y_t is said to have a causal influence on X_t. We quantify this causal influence by

$$F_{Y \to X} = \ln \frac{\Sigma_1}{\Sigma_2} \tag{2.1.4}$$

It is clear that the coefficients b_{2i} are uniformly zero if there is no causal influence from Y to X, thus we can get $\Sigma_1 = \Sigma_2$. We then can deduce $F_{Y \to X} = 0$. For $\Sigma_1 > \Sigma_2$ and $F_{Y \to X} > 0$, we can say that there is a direct influence from Y to X.

Frequency domain analysis

We firstly define the lag operator L to be $LX_t = X_{t-1}$. Rewrite Equation (2.1.2) in terms of the lag operator

$$\begin{pmatrix} a_2(L) & b_2(L) \\ c_2(L) & d_2(L) \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{2t} \\ \eta_{2t} \end{pmatrix} \tag{2.1.5}$$

By taking Fourier transform on both sides of Equation (2.1.5), it leads to

$$\begin{pmatrix} a_2(\omega) & b_2(\omega) \\ c_2(\omega) & d_2(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} E_x(\omega) \\ E_y(\omega) \end{pmatrix} \tag{2.1.6}$$

Where the components of the coefficient matrix $\mathbf{A}(\omega)$ are

$$\begin{aligned} a_2(\omega) &= 1 - \sum_{j=1}^{\infty} a_{2j} e^{-iwj}, \quad b_2(\omega) = -\sum_{j=1}^{\infty} b_{2j} e^{-iwj} \\ c_2(\omega) &= -\sum_{j=1}^{\infty} c_{2j} e^{-iwj}, \qquad d_2(\omega) = 1 - \sum_{j=1}^{\infty} d_{2j} e^{-iwj} \end{aligned} \tag{2.1.7}$$

For normalization, we left multiply

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ -\dfrac{\Upsilon_2}{\Sigma_2} & 1 \end{pmatrix} \tag{2.1.8}$$

On both sides of Equation (2.1.6). The result is

$$\begin{pmatrix} a_2(\omega) & b_2(\omega) \\ c_3(\omega) & d_3(\omega) \end{pmatrix} \begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} E_x(\omega) \\ \tilde{E}_y(\omega) \end{pmatrix} \tag{2.1.9}$$

Where $c_3(\omega) = c_2(\omega) - \frac{\Upsilon_2}{\Sigma_2} a_2(\omega)$, $d_3(\omega) = d_2(\omega) - \frac{\Upsilon_2}{\Sigma_2} b_2(\omega)$, $\tilde{E}_y(\omega) = E_y(\omega) - \frac{\Upsilon_2}{\Sigma_2} E_x(\omega)$. Reformat Equation (2.1.9) into the transfer function format, we obtain

$$\begin{pmatrix} X(\omega) \\ Y(\omega) \end{pmatrix} = \begin{pmatrix} H_{xx}(\omega) & H_{xy}(\omega) \\ H_{yx}(\omega) & H_{yy}(\omega) \end{pmatrix} \begin{pmatrix} E_x(\omega) \\ \tilde{E}_y(\omega) \end{pmatrix} \tag{2.1.10}$$

Where the transfer function is $\mathbf{H}(\omega) = \mathbf{A}^{-1}(\omega)$ (-1 represents matrix inverse) whose components are

$$\begin{aligned} H_{xx} &= \frac{1}{\det \mathbf{A}} d_2(\omega) + \frac{\Upsilon_2}{\Sigma_2} \left(-\frac{1}{\det \mathbf{A}} b_2(\omega) \right), \quad H_{xy} = -\frac{1}{\det \mathbf{A}} b_2(\omega) \\ H_{yx} &= -\frac{1}{\det \mathbf{A}} c_2(\omega) + \frac{\Upsilon_2}{\Sigma_2} \left(\frac{1}{\det \mathbf{A}} d_2(\omega) \right), \quad H_{yy} = \frac{1}{\det \mathbf{A}} a_2(\omega) \end{aligned} \tag{2.1.11}$$

From Equation (2.1.10), the spectrum of X_t is found to be

$$S_{xx}(\omega) = H_{xx}(\omega) \Sigma_2 H_{xx}^*(\omega) + H_{xy}(\omega) \tilde{\Gamma}_2 H_{xy}^*(\omega) \tag{2.1.12}$$

Here the first term is interpreted as the intrinsic power and the second term as the causal power of X_t due to Y_t. Based on this interpretation we define the causal influence from Y_t to X_t at frequency ω as

$$f_{Y\to X}(\omega) = \ln \frac{S_{xx}(\omega)}{H_{xx}(\omega)\Sigma_2 H^*_{xx}(\omega)} \tag{2.1.13}$$

This definition of causal influence is expressed in terms of the intrinsic power rather than the causal power. It is expressed in this way so that the causal influence is zero when the intrinsic power equals the total power.

2.1.2 Conditional Granger causality

For three or more time series, one can perform a pairwise analysis and thus reduce the problem to a bivariate problem. This approach has some inherent limitations and could induce misleading results of indirect edges. Here we define conditional Granger causality which has the ability to resolve if the interaction between two time series is direct or is mediated by another recorded time series. The method is introduced for three time series [Ding, 1999]. The framework can be generalized to three sets of time series.

Time Domain Analysis

Consider three time series X_t, Y_t and Z_t. The joint autoregressive representation of X_t and Z_t is formalised as

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{4i} X_{t-i} + \sum_{i=1}^{\infty} b_{4i} Z_{t-i} + \varepsilon_{4t} \\ Z_t = \sum_{i=1}^{\infty} c_{4i} X_{t-i} + \sum_{i=1}^{\infty} d_{4i} Z_{t-i} + \gamma_{4t} \end{cases} \tag{2.1.14}$$

Where the covariance matrix of the noise terms is

$$\mathbf{S}_4 = \begin{bmatrix} \mathrm{var}(\varepsilon_{4t}) & \mathrm{var}(\varepsilon_{4t}, \gamma_{4t}) \\ \mathrm{var}(\gamma_{4t}, \varepsilon_{4t}) & \mathrm{var}(\gamma_{4t}) \end{bmatrix} = \begin{bmatrix} S_{xx} & S_{xz} \\ S_{zx} & S_{zz} \end{bmatrix} \tag{2.1.15}$$

The joint autoregressive representation of all the three time series is

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{5i} X_{t-i} + \sum_{i=1}^{\infty} b_{5i} Y_{t-i} + \sum_{i=1}^{\infty} c_{5i} Z_{t-i} + \varepsilon_{5t} \\ Y_t = \sum_{i=1}^{\infty} e_{5i} X_{t-i} + \sum_{i=1}^{\infty} f_{5i} Y_{t-i} + \sum_{i=1}^{\infty} g_{5i} Z_{t-i} + \eta_{5t} \\ Z_t = \sum_{i=1}^{\infty} u_{5i} X_{t-i} + \sum_{i=1}^{\infty} v_{5i} Y_{t-i} + \sum_{i=1}^{\infty} w_{5i} Z_{t-i} + \gamma_{5t} \end{cases} \tag{2.1.16}$$

Where the covariance matrix of the noise term is

$$\begin{aligned}\boldsymbol{\Sigma}_5 &= \begin{bmatrix} \mathrm{var}(\varepsilon_{5t}) & \mathrm{cov}(\varepsilon_{5t},\eta_{5t}) & \mathrm{cov}(\varepsilon_{5t},\gamma_{5t}) \\ \mathrm{cov}(\eta_{5t},\varepsilon_{5t}) & \mathrm{var}(\eta_{5t}) & \mathrm{cov}(\eta_{5t},\gamma_{5t}) \\ \mathrm{cov}(\gamma_{5t},\varepsilon_{5t}) & \mathrm{cov}(\gamma_{5t},\eta_{5t}) & \mathrm{var}(\gamma_{5t}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} \end{bmatrix}\end{aligned} \tag{2.1.17}$$

From above two sets of equations, the conditional Granger causality from Y to X conditional on Z can be defined as

$$F_{Y\to X|Z} = \ln\left(\frac{S_{xx}}{\Sigma_{xx}}\right) \tag{2.1.18}$$

When the causal influence from Y to X is entirely mediated by Z, the coefficients of b_{5i} are zero, and the two autoregression models for two time series and three time series will be exactly same, thus we can get $S_{xx} = \Sigma_{xx}$. We then deduce $F_{Y\to X|Z} = 0$, which means Y cannot further improve the prediction of X including past measurements of Y conditional on Z.

Frequency domain analysis

To derive the spectral decomposition of the time domain conditional Granger causality we carry out a normalization procedure like that for the bivariate case. For Equation (2.1.14), the normalized equations are

$$\begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} X_t \\ Z_t \end{pmatrix} = \begin{pmatrix} X_t^* \\ Z_t^* \end{pmatrix} \tag{2.1.19}$$

Where $\text{cov}(X_t^*, Z_t^*) = 0$ and $\text{var}(X_t^*) = S_{xx}$. For Equation (2.1.16), the normalization process involves left-multiplying both sides by the matrix

$$P = P_2 \times P_1 \tag{2.1.20}$$

where

$$P_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\Sigma_{yx}\Sigma_{xx}^{-1} & 1 & 0 \\ -\Sigma_{zx}\Sigma_{xx}^{-1} & 0 & 1 \end{pmatrix} \tag{2.1.21}$$

and

$$P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -(\Sigma_{zy} - \Sigma_{zx}\Sigma_{xx}^{-1}\Sigma_{xy})(\Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})^{-1} & 1 \end{pmatrix} \tag{2.1.22}$$

The normalised equation can be represented as

$$\begin{pmatrix} B_{11}(L) & B_{12}(L) & B_{13}(L) \\ B_{21}(L) & B_{22}(L) & B_{23}(L) \\ B_{31}(L) & B_{32}(L) & B_{33}(L) \end{pmatrix} \begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{xt} \\ \varepsilon_{yt} \\ \varepsilon_{zt} \end{pmatrix} \tag{2.1.23}$$

The Fourier transform of Equations (2.1.19) and (2.1.23) gives

$$\begin{pmatrix} X(\omega) \\ Z(\omega) \end{pmatrix} = \begin{pmatrix} G_{xx}(\omega) & G_{xz}(\omega) \\ G_{zx}(\omega) & G_{zz}(\omega) \end{pmatrix} \begin{pmatrix} X^*(\omega) \\ Z^*(\omega) \end{pmatrix} \tag{2.1.24}$$

And

$$\begin{pmatrix} X(\omega) \\ Y(\omega) \\ Z(\omega) \end{pmatrix} = \begin{pmatrix} H_{xx}(\omega) & H_{xy}(\omega) & H_{xz}(\omega) \\ H_{yx}(\omega) & H_{yy}(\omega) & H_{yz}(\omega) \\ H_{zx}(\omega) & H_{zy}(\omega) & H_{zz}(\omega) \end{pmatrix} \begin{pmatrix} E_x(\omega) \\ E_y(\omega) \\ E_z(\omega) \end{pmatrix} \tag{2.1.25}$$

As [Geweke, 1984] demonstrated the important relation:

$$F_{Y \to X|Z}(\omega) = F_{YZ^* \to X^*}(\omega) \tag{2.1.26}$$

We combined Equations (2.1.24) and (2.1.25) to get the spectrum of X_t^*

$$\begin{aligned}\begin{pmatrix} X^*(\omega) \\ Y(\omega) \\ Z^*(\omega) \end{pmatrix} &= \begin{pmatrix} G_{xx}(\omega) & 0 & G_{xz}(\omega) \\ 0 & I & 0 \\ G_{zx}(\omega) & 0 & G_{zz}(\omega) \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} H_{xx}(\omega) & H_{xy}(\omega) & H_{xz}(\omega) \\ H_{yx}(\omega) & H_{yy}(\omega) & H_{yz}(\omega) \\ H_{zx}(\omega) & H_{zy}(\omega) & H_{zz}(\omega) \end{pmatrix} \begin{pmatrix} E_x(\omega) \\ E_y(\omega) \\ E_z(\omega) \end{pmatrix} \\ &= \begin{pmatrix} Q_{xx}(\omega) & Q_{xy}(\omega) & Q_{xz}(\omega) \\ Q_{yx}(\omega) & Q_{yy}(\omega) & Q_{yz}(\omega) \\ Q_{zx}(\omega) & Q_{zy}(\omega) & Q_{zz}(\omega) \end{pmatrix} \begin{pmatrix} E_x(\omega) \\ E_y(\omega) \\ E_z(\omega) \end{pmatrix}\end{aligned} \tag{2.1.27}$$

The power spectrum of X_t^* is then obtained as

$$\begin{aligned}S_{x^*x^*} &= Q_{xx}(\omega)\hat{\Sigma}_{xx}Q_{xx}^*(\omega) + Q_{xy}(\omega)\hat{\Sigma}_{yy}Q_{yx}^*(\omega) \\ &+ Q_{xz}(\omega)\hat{\Sigma}_{zz}Q_{zx}^*(\omega)\end{aligned} \tag{2.1.28}$$

Then the Granger causality is obtained as

$$F_{Y\to X|Z}(\omega) = \ln\left(\frac{\left|S_{x^*x^*}\right|}{\left|Q_{xx}(\omega)\hat{\Sigma}_{xx}Q_{xx}^*(\omega)\right|}\right) \tag{2.1.29}$$

2.1.3 Partial Granger causality

There is one problem for the conditional Granger causality: it cannot deal with data sets which contain an unobserved common input. In order to

solve this common issue in computational biology, partial Granger causality approach is introduced [Guo, 2008].

Time domain analysis

Suppose we have two time series which could be generated as

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{6i} X_{t-i} + \sum_{i=1}^{\infty} b_{6i} Z_{t-i} + \varepsilon_{6t} + C_{6t} \\ Z_t = \sum_{i=1}^{\infty} c_{6i} X_{t-i} + \sum_{i=1}^{\infty} d_{6i} Z_{t-i} + \gamma_{6t} + C_{6t} \end{cases} \tag{2.1.30}$$

Where the C_{6t} is an unobserved external input. The noise covariance matrix for this model can be represented as

$$\mathbf{S}_6 = \begin{bmatrix} \mathrm{var}(\varepsilon_{6t}) & \mathrm{var}(\varepsilon_{6t}, \gamma_{6t}) \\ \mathrm{var}(\gamma_{6t}, \varepsilon_{6t}) & \mathrm{var}(\gamma_{6t}) \end{bmatrix} = \begin{bmatrix} S_{xx}^6 & S_{xz}^6 \\ S_{zx}^6 & S_{zz}^6 \end{bmatrix} \tag{2.1.31}$$

The vector autoregressive representation for a system involving three time series can be generated as

$$\begin{cases} X_t = \sum_{i=1}^{\infty} a_{7i} X_{t-i} + \sum_{i=1}^{\infty} b_{7i} Y_{t-i} + \sum_{i=1}^{\infty} c_{7i} Z_{t-i} + \varepsilon_{7t} + C_{7t} \\ Y_t = \sum_{i=1}^{\infty} e_{7i} X_{t-i} + \sum_{i=1}^{\infty} f_{7i} Y_{t-i} + \sum_{i=1}^{\infty} g_{7i} Z_{t-i} + \eta_{7t} + C_{7t} \\ Z_t = \sum_{i=1}^{\infty} u_{7i} X_{t-i} + \sum_{i=1}^{\infty} v_{7i} Y_{t-i} + \sum_{i=1}^{\infty} w_{7i} Z_{t-i} + \gamma_{7t} + C_{7t} \end{cases} \quad (2.1.32)$$

The noise covariance matrix for the model can be represented as

$$\begin{aligned} \mathbf{\Sigma}_7 &= \begin{bmatrix} \mathrm{var}(\varepsilon_{7t}) & \mathrm{cov}(\varepsilon_{7t}, \eta_{7t}) & \mathrm{cov}(\varepsilon_{7t}, \gamma_{7t}) \\ \mathrm{cov}(\eta_{7t}, \varepsilon_{7t}) & \mathrm{var}(\eta_{7t}) & \mathrm{cov}(\eta_{7t}, \gamma_{7t}) \\ \mathrm{cov}(\gamma_{7t}, \varepsilon_{7t}) & \mathrm{cov}(\gamma_{7t}, \eta_{7t}) & \mathrm{var}(\gamma_{7t}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{xx}^7 & \Sigma_{xy}^7 & \Sigma_{xz}^7 \\ \Sigma_{yx}^7 & \Sigma_{yy}^7 & \Sigma_{yz}^7 \\ \Sigma_{zx}^7 & \Sigma_{zy}^7 & \Sigma_{zz}^7 \end{bmatrix} \end{aligned} \quad (2.1.33)$$

In order to measure the accuracy of the autoregressive prediction without the influence of the common exogenous inputs, the notion of partial correlation is implied. Then the partial Granger causality equation can be obtained as

$$F_{Y \to X|Z}^{P} = \ln \left(\frac{\left| S_{xx}^6 - S_{xz}^6 (S_{zz}^6)^{-1} S_{zx}^6 \right|}{\left| \Sigma_{xx}^7 - \Sigma_{xz}^7 (\Sigma_{zz}^7)^{-1} \Sigma_{zx}^7 \right|} \right) \quad (2.1.34)$$

Frequency domain analysis

The spectral decomposition of the partial Granger causality is very similar to conditional Granger causality. To eliminate the effect of common inputs, we only need to change the normalization multiplier in Equations (2.1.8) as

$$\mathbf{P} = \begin{pmatrix} 1 & -S_{xz}^{6}(S_{zz}^{6})^{-1} \\ 0 & 1 \end{pmatrix} \tag{2.1.35}$$

Change Equations (2.1.21) and (2.1.22) as following

$$P_1 = \begin{pmatrix} 1 & 0 & -\Sigma_{xz}^{7}(\Sigma_{zz}^{7})^{-1} \\ 0 & 1 & -\Sigma_{yz}^{7}(\Sigma_{zz}^{7})^{-1} \\ 0 & 0 & 1 \end{pmatrix} \tag{2.1.36}$$

And

$$P_2 = \begin{pmatrix} 1 & 0 & 0 \\ -(\Sigma_{xy}^{7}\text{-}\Sigma_{xz}^{7}(\Sigma_{zz}^{7})^{-1}\Sigma_{zy}^{7})(\Sigma_{xx}^{7} - \Sigma_{xz}^{7}(\Sigma_{zz}^{7})^{-1}\Sigma_{zx}^{7})^{-1} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.1.37}$$

2.2 Bayesian network inference

Bayesian networks are probabilistic graphical models initially introduced by [Kim & Pearl, 1987]. A Bayesian network is a specific type of graphi-

cal model, which are called directed acyclic graph [Bishop, 1995 & 2006]. Each arc in the model is directed and there is no way to start from any node and travel along a set of directed edges and get back to the initial node. The set of nodes represent a set of random variables $\mathbf{X} = [X_1, X_2, \cdots, X_n]$, and the arcs express statistical dependence between the downstream variables and the upstream variables. The upstream variables are also called the parent variables of the downstream variables. Bayesian network inference yields the most concise model, automatically excluding arcs based on dependencies already explained by the model, which means the arcs in the network can be interpreted as conditional causality. The edges in the Bayesian network encode a particular factorization of the joint distribution. The joint probability distribution can be decomposed as following:

$$\mathrm{P}(X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} \mathrm{P}(X_i \mid parents(X_i)) \qquad (2.2.1)$$

That is, the joint probability distribution is the product of the local distributions of each node and its parents. If a node has no parents, its local probability distribution is said to be unconditional, otherwise it is conditional. This decomposition is useful for the Bayesian networks inference algorithm to deal with the uncertain situation and incomplete data.

To learn the parameter of the Bayesian network is to essentially estimate two kinds of probability distributions: the probability $\mathrm{P}(\mathbf{X})$ and the conditional probability $\mathrm{P}(\mathbf{X}\,|\,\mathbf{Y})$. There are two kinds of approaches to density estimation: the nonparametric method and the parametric method. The easiest estimation for the nonparametric method is to use the histogram approach. The distribution can then be a tabular conditional probability distribution, which is represented as a table. However this approach requires a much larger sample size to give an accurate estimation, which is not suitable for general experimental data. For the parametric method, one needs to make some assumptions about the form of the probability distribution, such as the widely used Gaussian distribution. For a D-dimensional vector $\mathbf{X}$, the multi-variate Gaussian distribution is in the form

$$\mathrm{N}(\mathbf{X}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\} \quad (2.2.2)$$

where $\boldsymbol{\mu}$ is a D-dimensional mean vector, $\boldsymbol{\Sigma}$ is a $D \times D$ covariance matrix, and $|\boldsymbol{\Sigma}|$ denotes the determinant of $\boldsymbol{\Sigma}$. The distribution $\mathbf{X}$ on a node can be defined as following:

$$-\textit{without parents}: \mathrm{P}(\mathbf{X}) \sim \mathrm{N}(\mathbf{X}\,|\,\boldsymbol{\mu},\boldsymbol{\sigma}) \quad (2.2.3)$$

$$\textit{-with parents}: \mathrm{P}(\mathbf{X} \mid \mathbf{Y} = \mathbf{y}) \sim \mathrm{N}(\mathbf{X} \mid \boldsymbol{\mu}_{x|y} + \mathbf{W}^{\mathrm{T}}\mathbf{y}, \boldsymbol{\sigma}_{x|y}) \quad (2.2.4)$$

where T is matrix transposition. $\mathbf{W}$ is the connection weight vector between node $\mathbf{X}$ and its parents $\mathbf{Y}$.

2.2.1 Parameter learning for linear Gaussian model

To estimate the parameters $\boldsymbol{\mu}_{x|y}$, $\boldsymbol{\sigma}_{x|y}$ and $\mathbf{W}$ in the equation (2.2.4), one partitions $\mathbf{X}$ into two disjoint subsets $\mathbf{X}_a$ and $\mathbf{X}_b$ with dimensions p and q, and we have $p + q = D$, such that

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_a \\ \mathbf{X}_b \end{pmatrix} \quad (2.2.5)$$

We also define mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ given by

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix} \quad (2.2.6)$$

Considering the quadratic form in the exponent of the Gaussian distribution, we can get the following equation by a transformation [Bishop, 2006]

$$
\begin{aligned}
&-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu}) \\
&=-\frac{1}{2}[(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}},(\mathbf{X}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}]\begin{bmatrix}\boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab}\\ \boldsymbol{\Sigma}_{ab} & \boldsymbol{\Sigma}_{bb}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{X}_a-\boldsymbol{\mu}_a\\ \mathbf{X}_b-\boldsymbol{\mu}_b\end{bmatrix} \\
&=-\frac{1}{2}\{(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}[\boldsymbol{\Sigma}_{aa}^{-1}+\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})^{-1}\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}] \\
&\quad\times(\mathbf{X}_a-\boldsymbol{\mu}_a)-2(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}[\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})^{-1}] \\
&\quad\times(\mathbf{X}_b-\boldsymbol{\mu}_b)+(\mathbf{X}_b-\boldsymbol{\mu}_b)^{\mathrm{T}}[(\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})^{-1}](\mathbf{X}_b-\boldsymbol{\mu}_b)\} \\
&=-\frac{1}{2}\{(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}} \\
&\quad+[(\mathbf{X}_b-\boldsymbol{\mu}_b)-\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\mathbf{X}_a-\boldsymbol{\mu}_a)]^{\mathrm{T}} \\
&\quad\times(\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})^{-1}[(\mathbf{X}_b-\boldsymbol{\mu}_b)-\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\mathbf{X}_a-\boldsymbol{\mu}_a)]^{\mathrm{T}}
\end{aligned}
\tag{2.2.7}
$$

The last equal sign is due to the following equations for any vectors u, v and a symmetric matrix $A=A^{\mathrm{T}}$

$$
\begin{aligned}
&u^{\mathrm{T}}Au-2u^{\mathrm{T}}Av+v^{\mathrm{T}}Av=u^{\mathrm{T}}Au-u^{\mathrm{T}}Av-u^{\mathrm{T}}Av+v^{\mathrm{T}}Av \\
&=u^{\mathrm{T}}A(u-v)-(u-v)^{\mathrm{T}}Av=u^{\mathrm{T}}A(u-v)-v^{\mathrm{T}}A(u-v) \\
&=(u-v)^{\mathrm{T}}A(u-v)
\end{aligned}
\tag{2.2.8}
$$

Now the joint distribution can be written as:

$$\mathrm{N}(\mathbf{X}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right\}$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}(2\pi)^{\frac{q}{2}}|\boldsymbol{\Sigma}_{aa}|^{\frac{1}{2}}\left|(\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})\right|^{\frac{1}{2}}}$$

$$\times\exp\left[-\frac{1}{2}(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}(\mathbf{X}_a-\boldsymbol{\mu}_a)^{\mathrm{T}}\right] \tag{2.2.9}$$

$$\times\exp\left\{\begin{array}{l}[(\mathbf{X}_b-\boldsymbol{\mu}_b)-\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\mathbf{X}_a-\boldsymbol{\mu}_a)]^{\mathrm{T}}\\ \times(\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})^{-1}[(\mathbf{X}_b-\boldsymbol{\mu}_b)-\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\mathbf{X}_a-\boldsymbol{\mu}_a)]^{\mathrm{T}}\end{array}\right\}$$

$$= \mathrm{N}(\mathbf{X}_a\,|\,\boldsymbol{\mu}_a,\boldsymbol{\Sigma}_{aa})\mathrm{N}(\mathbf{X}_b,\boldsymbol{\mu}_b+\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\mathbf{X}_a-\boldsymbol{\mu}_a),\boldsymbol{\Sigma}_{bb}-\boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab})$$

Then we can get the conditional distribution of $\mathbf{X}_b$ given $\mathbf{X}_a$ is

$$\mathrm{P}(\mathbf{X}_b\,|\,\mathbf{X}_a) = \frac{\mathrm{P}(\mathbf{X}_a,\mathbf{X}_b)}{\mathrm{P}(\mathbf{X}_a)}$$

$$= \frac{1}{(2\pi)^{\frac{q}{2}}\left|\,|\boldsymbol{\Sigma}_{b|a}|^{\frac{1}{2}}\right|}\exp\left[-\frac{1}{2}(\mathbf{X}_b-\boldsymbol{\mu}_{b|a})^{\mathrm{T}}\left|\boldsymbol{\Sigma}_{b|a}\right|^{-1}(\mathbf{X}_b-\boldsymbol{\mu}_{b|a})\right] \tag{2.2.10}$$

in which

$$\boldsymbol{\mu}_{b|a} = \boldsymbol{\mu}_b + \boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab}(\mathbf{X}_a-\boldsymbol{\mu}_a) \tag{2.2.11}$$

$$\boldsymbol{\Sigma}_{b|a} = \boldsymbol{\Sigma}_{bb} - \boldsymbol{\Sigma}_{ab}^{\mathrm{T}}\boldsymbol{\Sigma}_{aa}^{-1}\boldsymbol{\Sigma}_{ab} \tag{2.2.12}$$

Thus the conditional probability parameters in the Bayesian network can be learned from the above two equations by using joint probability distribution.

2.2.2 Structure learning

There are two very different approaches for structure learning: one is constraint-based and the other is search and score algorithm. For the constraint-based algorithm, we start with a fully connected network and then remove the arcs, which the connected nodes are conditional independent to each other. This has the disadvantage that repeated independence tests lose statistical power. For the latter algorithm, we perform a search on all possible graphs and select one graph which best describes the statistical dependence relationship in the observed data [Friedman, 2004].

Unfortunately, the number of possible graphs increases exponentially with the number of nodes, so some search algorithms are required for overcoming this kind of complex problem rather than doing an exhaustive search in the space. There are several search algorithms that can be applied; such as annealing search [Kirkpatrick, 1983], genetic algorithm search [Goldberg, 1989]. The question could become easier if we know the total order to the nodes. The K2 algorithm allows us to find the best structure by selecting the best set of parents

for each node independently [Murphy, 2001]. In the dynamic Bayesian networks, the order of nodes can be interpreted as the sequence of time lags represented for each node. The K2 algorithm tests parent insertion according to the order. The first node cannot have any parent, for other nodes, we can only choose the parent nodes which are behind it in this order. Then the scoring function can be applied to determine the best parent set, i.e. the one which gives the highest score.

Initially, each variable can be interpreted as a sequence of nodes which represents as the different time lags. Suppose we observed a set of independent and identically distributed time series data $\mathbf{Y} = [Y^1, Y^2, \cdots, Y^N]$, which has N dimensions. Every node in dynamic Bayesian network represents a specific time lag for a specific variable. For instance, time series $\{Y^j \mid j \in [1,2,\cdots,N]\}$ can be represented by using a series of nodes in the order of $[\cdots, Y_{t-3}^j, Y_{t-2}^j, Y_{t-1}^j, Y_t^j]$. Hence, the total order of the nodes can then be $[\cdots, \mathbf{Y}_{t-3}, \mathbf{Y}_{t-2}, \mathbf{Y}_{t-1}, \mathbf{Y}_t]$. Since we are only concerned with the causal relation between different time lags, the order of various variables for the same time lags can be randomly selected. Then the potential parent set for every node can then be determined according to the total order, which contains all the nodes before it. Finally, we can apply the K2 algorithm to select the set of best parents from the set of potential parents for every node independently.

In addition to the search algorithm, a scoring function must be defined in order to decide which structure is the best (a high scoring network). There are two popular choices. One is the Bayesian score metric which is the marginal likelihood of the model, and the other is BIC (Bayesian Information Criterion) which can be defined as [Murphy, 2001]:

$$LogP(Data \mid \theta) - \frac{d}{2} Log(N) \tag{2.2.13}$$

Where $Data$ is the observed data, θ is the estimated value of the parameters, d is the number of parameters and N is the number of data cases. The term of $\frac{d}{2} Log(N)$ is regarded as a penalty term in order to balance both simple and accurate structure representation.

Suppose we observed a set of independent and identically distributed $Data = \{Y^1, Y^2, \ldots, Y^N\}$, each of which represented a time series (can also be a case of vector). Then the log likelihood of the data set can be defined as:

$$\begin{aligned} LogP(Data \mid \theta) &= \sum_{i=1}^{N} \log P(Y^i \mid \theta) \\ &= \sum_{i=1}^{N} \log \prod_j P(Y_{t-j}^i \mid Pa(Y_{t-j}^i), \theta_j^i) \\ &= \sum_{i=1}^{N} \sum_j \log P(Y_{t-j}^i \mid Pa(Y_{t-j}^i), \theta_j^i) \end{aligned} \tag{2.2.14}$$

Where i, j represent the different time lags of different nodes or variables in the Bayesian network, $pa(Y_{t-j}^{i})$ is the set of parents of node Y_{t-j}^{i}, and θ_{j}^{i} are the parameters that define the conditional probability of Y_{t-j}^{i} given its parents.

Generally, the Bayesian networks inference can then be approached by following procedure: initially, K2 algorithm (as described in **Chapter 2.2**) is applied to search the space of possible graphs. For each possible structure, we can use the parameter learning algorithm (as described in **Chapter 2.1**) to estimate the parameters of the networks. The BIC scoring function assigns a corresponding score through the estimated parameters and observed data set. The best network we can get is the highest score structures among all the possible graphs. This procedure is described as following pseudo code for a specific node Y_{t}^{j}.

Step 1. Calculate **initialScore**: the initial BIC score for node Y_{t}^{j} (initially no parents are selected) by using Equation (2.2.13).

Step 2. Test each node in the set of potential parents **S** to be the parent node of Y_{t}^{j}, for each node, calculate the BIC score.

Step 3. Select the best parent node (**bestNode**) which gives the highest score: **newScore**.

Step 4. Compare **newScore** to the **initialScore**, if smaller then stop, else add a arc from **bestNode** to the node Y_t^j.

Step 5. Change **initialScore** to **newScore**.

Step 6. Remove **bestNode** from set **S**.

Step 7. Go back to step 2.

2.3 Summary

This chapter described the network inference methods for both Granger causality and dynamic Bayesian network inference, which are widely used in literature. In the following chapter, In order to test the performance, reliability and sensitivity, one is going to carry out a comparative study between these two methods. The application of Granger causality is also described in **Chapter 4** and **Chapter 5**.

Chapter 3

A comparative study between Granger causality and dynamic Bayesian network inference

In this chapter, we carried out a comparative study between the two commonly used reverse-engineering approaches: Granger causality and dynamic Bayesian network inference. We focused on a systematic and computationally intensive comparison on both synthesized and experimental data. For synthesized data, a critical point of the data length was found: the dynamic Bayesian network outperformed the Granger causality approach when the data length was short, and vice versa. We then tested our results in experimental data of short length which was a common scenario in current biological experiments: it was again confirmed that the dynamic Bayesian network worked better. In

conclusion, when the data size is short, the dynamic Bayesian network inference performs better than the Granger causality approach; otherwise the Granger causality approach is better for detecting weak interactions if the data length is long enough.

3.1 Introduction

Based upon high throughput data, to reliably and accurately explore the network structure of elements (genes, proteins, metabolites, neurons etc.) is one of the most important issues in computational biology [Klipp, et al., 2005; Feng, et al., 2007; Alon, 2007; Tong, et al., 2004; Tsai, et al., 2008; Lee, et al., 2002]. Currently, there are two main approaches which are often used to infer causal relationships [Pearl, 2000] or interactions among a set of elements [Albo, 2004; Horton, 2005]. One is the Granger causality approach [Guo, 2008; Wu, 2008], and the other is the Bayesian network inference approach [Jansen, 2003; Saches, 2005]. The latter is often applied to static data. However, one can employ dynamic Bayesian networks to deal with time series data for which the Granger causality has been solely developed. Granger causality has the advantage of having a corresponding frequency domain decomposition so that one can clearly find at which frequencies two elements interact with each other.

Given a multi-variable time series dataset, the Granger causality and dynamic Bayesian networks [Ghahramani, 2004] can both be applied. The Granger causality notation, which was firstly introduced by Wiener and Granger [Geweke, 1982 & 1984], proposed that we can determine a causal influence of one time series on another: the prediction of one time series can be improved by incorporating the knowledge of the second one. On the other hand,

the Bayesian network [Jensen, 1996] is a special case of a diagrammatic representation of probability distributions, called probabilistic graphical models [Bache, 2004; Buntine, 1994; Friedman, 2004]. The Bayesian network graph model comprises nodes (also called vertices) connected by directed links (also called edges or arcs) and there is no cycle in the graph. To learn the structure and the parameters for the Bayesian networks from a set of data, we should search the space(s) of all possible graph representations, and find out which structure is most likely to produce our data. If we have a scoring function (or likelihood function) which can determine the structure and parameter likelihood from the data, then the problem is to find the highest score (maximum likelihood) structure among all the possible representations.

The causal relationship derived from these two approaches could be different, in particular when we face the data obtained from experiments. Therefore it is of vital importance to compare these two causal inferring approaches before we can confidently apply them to biological data. By doing the comparison, one expects to find the advantages, performances and stabilities for each technique.

Adopting the most common existing methods to find the coefficients of the time series in both approaches in the literature, we compare the dynamic Bayesian network with the Granger causality both in the linear (as described in

Chapter 2) and nonlinear model (described in **Chapter 3.2**). Interestingly, a critical point of the data length is found. When the data length is shorter than the critical point, the dynamic Bayesian network approach outperforms the Granger causality approach. But when the data length is longer, the Granger causality is more reliable. The conclusion is obtained via intensive computations (more than 100 computers over a few weeks). A biological data set of gene microarray is analyzed using both approaches, which indicates that for a data set with a short sampling length the dynamic Bayesian network produces more reliable results. In summary, we would argue that the dynamic Bayesian network is more suitable for dealing with experimental data.

3.2 Methods

3.2.1 Granger causality

As described in the previous chapter, we can determine a causal influence of one time series on another, if the predication of one time series can be improved by incorporating the knowledge of the second one. Granger applied this notation by using the context of linear vector auto-regression (VAR) model of stochastic processes [Beamish, 1981; Morettin, 1984; Morf, 1978; Ancona, 2004]. A VAR model describes the evolution of a set of *k* variables (comparing

to univariate AR models) as a linear function of only their past evolution. In the AR model, the variance of the prediction error is used to test the prediction improvement. For instance, assume two time series; if the variance of the autoregressive prediction error of the first time series at the present time is reduced by inclusion of past measurements from the second time series, then one can conclude that the second time series has a causal influence on the first one. Geweke [Geweke, 1982 & 1984] decomposed the VAR process into the frequency domain, it converted the causality measurement into a spectral representation and made the interpretation more appealing. The detailed description of pair-wise Granger causality was introduced in **Chapter 2**.

The pairwise analysis can only be applied to bivairate time series. For more than two time series, a time series can have a direct or indirect causal influence to other time series. In this case, pairwise analysis is not sufficient or misleading for revealing whether the causal interaction between a pair is direct or indirect. In order to distinguish the direct and indirect causal affect, one introduces the conditional causality which takes account of the other time series' effect in a multivariate time series. In this chapter, we used conditional causality to compare with the Bayesian network inference introduced before. For the linear conditional Granger causality, its detailed description was introduced in the **Chapter 2**. Here, I extended the linear conditional Granger causality to a non-linear model.

Non-linear conditional Granger causality

We extended Granger causality to a non-linear model by using a series kernel functions [Chen, 2004; Ancona, 2004] . Let X , Y and Z be three time series, which are assumed to be stationary. We are supposed to quantify how much Y causes X conditional on Z . This model can be generalized to three sets of time series. The expression for the nonlinear model of two time series X and Z is:

$$\begin{cases} X_t = \sum_j w_{1j}\Phi_j(X_{t-j}) + \sum_j w_{2j}\Phi_j(Z_{t-j}) + \varepsilon_{8t} \\ Z_t = \sum_j w_{3j}\Phi_j(X_{t-j}) + \sum_j w_{4j}\Phi_j(Z_{t-j}) + \gamma_{8t} \end{cases} \tag{3.2.1}$$

Function Φ can be selected as the kernel function of X and Z which has the following expression:

$$\Phi_j(X) = \exp(-\left\| X - \bar{X}_j \right\|^2 / 2\sigma_X^2) \tag{3.2.2}$$

$$\Phi_j(Z) = \exp(-\left\| Z - \bar{Z}_j \right\|^2 / 2\sigma_Z^2) \tag{3.2.3}$$

Where $\bar{X}$, $\bar{Z}$ are centers (or means) of X and Z , σ_X^2 , σ_Z^2 are variances of X and Z . The covariance matrix of prediction error can be expressed as

$$\mathbf{S}_8 = \begin{bmatrix} \mathrm{var}(\varepsilon_{8t}) & \mathrm{cov}(\varepsilon_{8t}, \gamma_{8t}) \\ \mathrm{cov}(\gamma_{8t}, \varepsilon_{8t}) & \mathrm{var}(\gamma_{8t}) \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{XX}^8 & \mathbf{S}_{XZ}^8 \\ \mathbf{S}_{ZX}^8 & \mathbf{S}_{ZZ}^8 \end{bmatrix} \tag{3.2.4}$$

A joint autoregressive representation for three time series has the following expression:

$$\begin{cases} X_t = \sum_j w_{5j} \Phi_j(X_{t-j}) + \sum_j w_{6j} \Phi_j(Y_{t-j}) + \sum_j w_{7j} \Phi_j(Z_{t-j}) + \varepsilon_{9t} \\ Y_t = \sum_j w_{8j} \Phi_j(X_{t-j}) + \sum_j w_{9j} \Phi_j(Y_{t-j}) + \sum_j w_{10j} \Phi_j(Z_{t-j}) + \eta_{9t} \\ Z_t = \sum_j w_{11j} \Phi_j(X_{t-j}) + \sum_j w_{12j} \Phi_j(Y_{t-j}) + \sum_j w_{13j} \Phi_j(Z_{t-j}) + \gamma_{9t} \end{cases} \tag{3.2.5}$$

The covariance matrix of prediction error can be expressed as

$$\begin{aligned} \mathbf{\Sigma}_9 &= \begin{bmatrix} \mathrm{var}(\varepsilon_{9t}) & \mathrm{cov}(\varepsilon_{9t}, \eta_{9t}) & \mathrm{cov}(\varepsilon_{9t}, \gamma_{9t}) \\ \mathrm{cov}(\eta_{9t}, \varepsilon_{9t}) & \mathrm{var}(\eta_{9t}) & \mathrm{cov}(\eta_{9t}, \gamma_{9t}) \\ \mathrm{cov}(\gamma_{9t}, \varepsilon_{9t}) & \mathrm{cov}(\gamma_{9t}, \eta_{9t}) & \mathrm{var}(\gamma_{9t}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{XX}^9 & \Sigma_{XY}^9 & \Sigma_{XZ}^9 \\ \Sigma_{YX}^9 & \Sigma_{YY}^9 & \Sigma_{YZ}^9 \\ \Sigma_{ZX}^9 & \Sigma_{ZY}^9 & \Sigma_{ZZ}^9 \end{bmatrix} \end{aligned} \tag{3.2.6}$$

Similarly, we can define the conditional causality for non-linear model as

$$F_{Y \to X|Z} = \ln\left(\frac{S_{XX}^8}{\Sigma_{XX}^9}\right) \tag{3.2.7}$$

3.2.2 Dynamic Bayesian network

A detailed method description of parameter learning and structure learning for linear dynamic Bayesian network inference in time domain was introduced in **Chapter 2.2**. Here, an extended dynamic Bayesian network inference for non-linear model will be explained in this section. As introduced in the **Chapter 2.1**, Granger causality approach has advantages in deriving the causality in frequency domain by using a frequency decomposition method. Here, I also extended the dynamic Bayesian network inference in frequency domain by using the similar approach for Granger causality.

Parameter learning for non-linear Gaussian model

We also extended our linear model to a non-linear model as in the Granger causality case. Suppose we have two variables which can be expressed as in Equation (3.2.1). The kernel function was also chosen as described in Equations (3.2.2) and (3.2.3).

In the non-linear model, the probability distribution for X_t was no longer a Gaussian distribution. From the expression in Equation (3.2.1), we found that the probability distribution for X_t was a combined distribution of kernel function distribution for the past measured values of X and Z, and a

Gaussian distribution for the noise term. The kernel distribution was very difficult to derive, so one used a mixture of Gaussian models to approximate the real distribution. The mixture Gaussian model is in the form:

$$\mathrm{P}(X) = \sum_{k=1}^{K} \pi_k \, \mathrm{N}(X \mid \mu_k, \Sigma_k) \tag{3.2.8}$$

Each Gaussian density $\mathrm{N}(X \mid \mu_k, \Sigma_k)$ is called a component of the mixture and has its own mean μ_k and covariance Σ_k. The parameter π_k are called mixing coefficients which satisfy:

$$\sum_{k=1}^{K} \pi_k = 1 \tag{3.2.9}$$

The conditional probability distribution for X_t conditional on the past observation of X and Z in the nonlinear model is still a Gaussian distribution which can be easily obtained as following:

$$\mathrm{P}(X \mid \mathbf{Z} = \mathbf{z}) = \mathrm{N}(X \mid \mu + \sum_{i} w_i \Phi(z_i), \sigma) \tag{3.2.10}$$

where w_i are the connection weights between node X and its parents. They can be estimated by using the simple linear regression method. The structure learning for non-linear Bayesian network inference is same as linear one, which has been introduced in **Chapter 2.2**.

Dynamic Bayesian network inference in frequency domain

To extend our Bayesian network inference approach to a frequency domain, we applied a similar spectral decomposition procedure like that for the conditional Granger causality case. Suppose we have three variables X_t, Y_t and Z_t, then we can transfer our learned Bayesian Network parameters and structure to the polynomial equations. First, the joint autoregressive representation of X_t and Z_t can be represented as Equation (2.1.14), where a_{4i}, b_{4i}, c_{4i} and d_{4i} are the connection weights between two nodes in the dynamic Bayesian network. These connection weights can be estimated by using Equation (2.2.11) in the parameter learning method section. If there is no arc between two nodes, then the corresponding connection weight equals zero. The covariance matrix of the noise term is shown in Equation (2.1.15). This covariance matrix can also be estimated in our parameter learning method by using Equation (2.2.12). Next we consider the joint autoregressive representation by adding the variable Y_t into the system shown in Equation (2.1.16), where the covariance matrix of the noise term is represented in Equation (2.1.17). To derive the spectral decomposition of the time domain Bayesian network, we can apply the same procedure as the conditional Granger causality described by using Equations (2.1.19) to (2.1.29).

3.3 Results

To illustrate and compare the differences between the dynamic Bayesian network inference and the conditional Granger causality, a simple multivariate model with fixed coefficients, which has been discussed in many papers to test the Granger causality, was tested first. We then extended our comparisons to the more general case of the model with random coefficients, which required considerable computational resources. More than 100 networked computers were used to perform the comparisons for more than a week. Both the Granger causality and the dynamic Bayesian network were applied to nonlinear models. Finally, we tested our approach on a set of microarray data recently acquired from a comparison of mock and infected Arabidopsis leaf.

3.3.1 Synthesized data: linear case

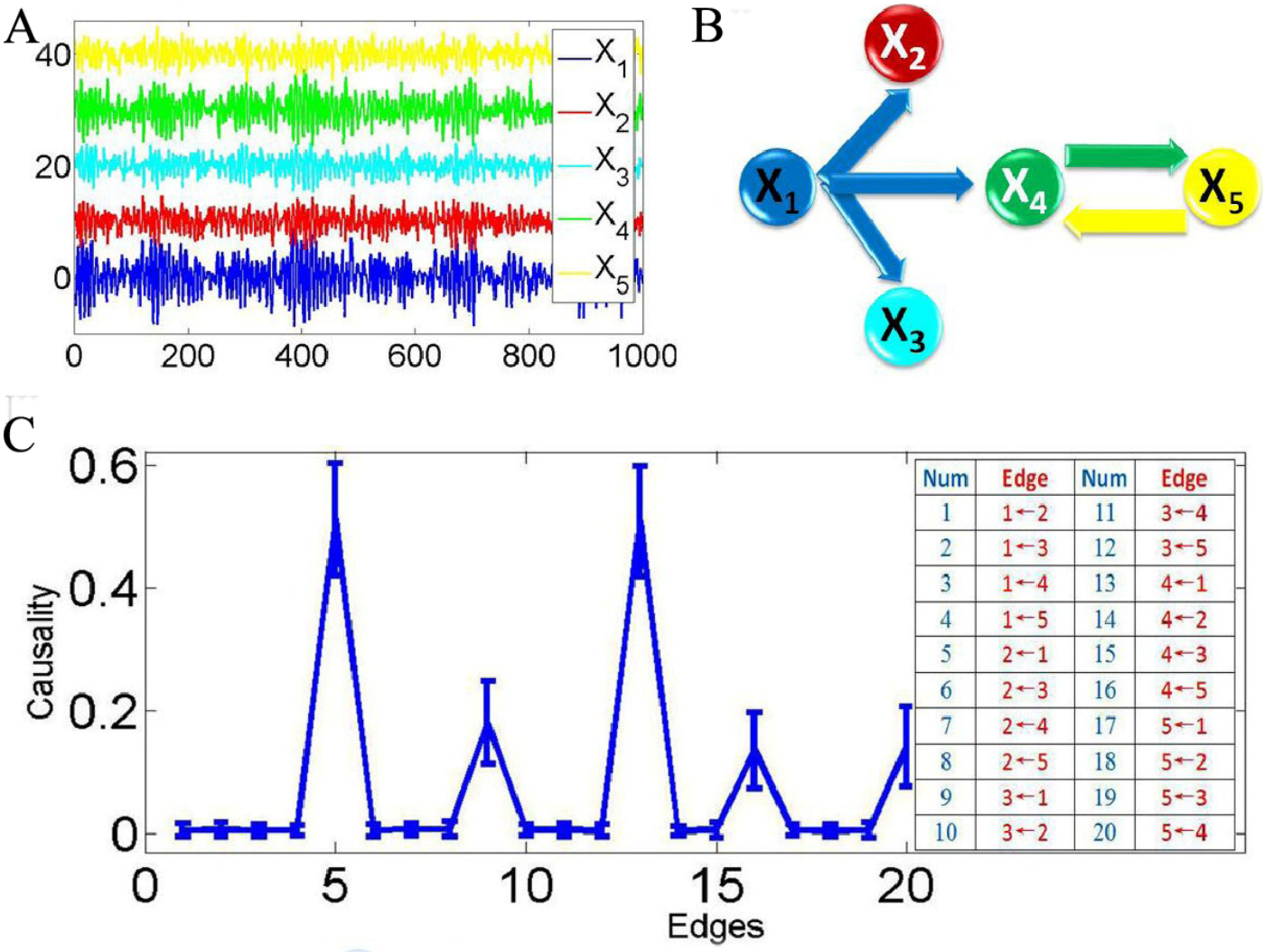
Example 1 Suppose we have 5 simultaneously recorded time series generated according to the equations (one popular toy model for test causal relation, also see in [Ding, 2006]):

$$\begin{cases} X_1(n) = 0.95\sqrt{2}X_1(n-1) - 0.9025X_1(n-2) + \varepsilon_1 \\ X_2(n) = 0.5X_1(n-2) + \varepsilon_2 \\ X_3(n) = -0.4X_1(n-3) + \varepsilon_3 \\ X_4(n) = -0.5X_1(n-1) + 0.25\sqrt{2}X_4(n-1) + 0.25\sqrt{2}X_5(n-1) + \varepsilon_4 \\ X_5(n) = -0.25\sqrt{2}X_4(n-1) + 0.25\sqrt{2}X_5(n-1) + \varepsilon_5 \end{cases} \quad (3.3.1)$$

where n is the time, and $[\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5]$ are independent Gaussian white noise processes with zero means and unit variances. From the equations, we see that $X_1(n)$ is a cause of $X_2(n)$, $X_3(n)$ and $X_4(n)$, and $X_4(n)$ and $X_5(n)$ share a feedback loop with each other, as depicted in **Figure 3.1_B**. **Figure 3.1_A** shows an example of the time trace of 5 time series. For the Granger causality approach, we simulated the fitted vector autoregressive (VAR) model to generate a data set of 100 realizations of 1000 time points, and applied the bootstrap approach to construct the 95% confidence intervals (**Figure 3.1_C**). For Granger causality, we assume the causality value is Gaussian distributed. Then the confidence intervals can be obtained by calculating the mean and standard derivation values [Guo, 2008; Chen, 2004; Wehrens, 2006]. According to the confidence intervals, one derived the network structure as shown in **Figure 3.1_B** which correctly recovered the pattern of the connectivity in our toy model. For the dynamic Bayesian network inference approach, we inferred a network structure (**Figure 3.2_A**) for each realization of 1000 time points. The final resulting causal network model was inferred with high-confidence

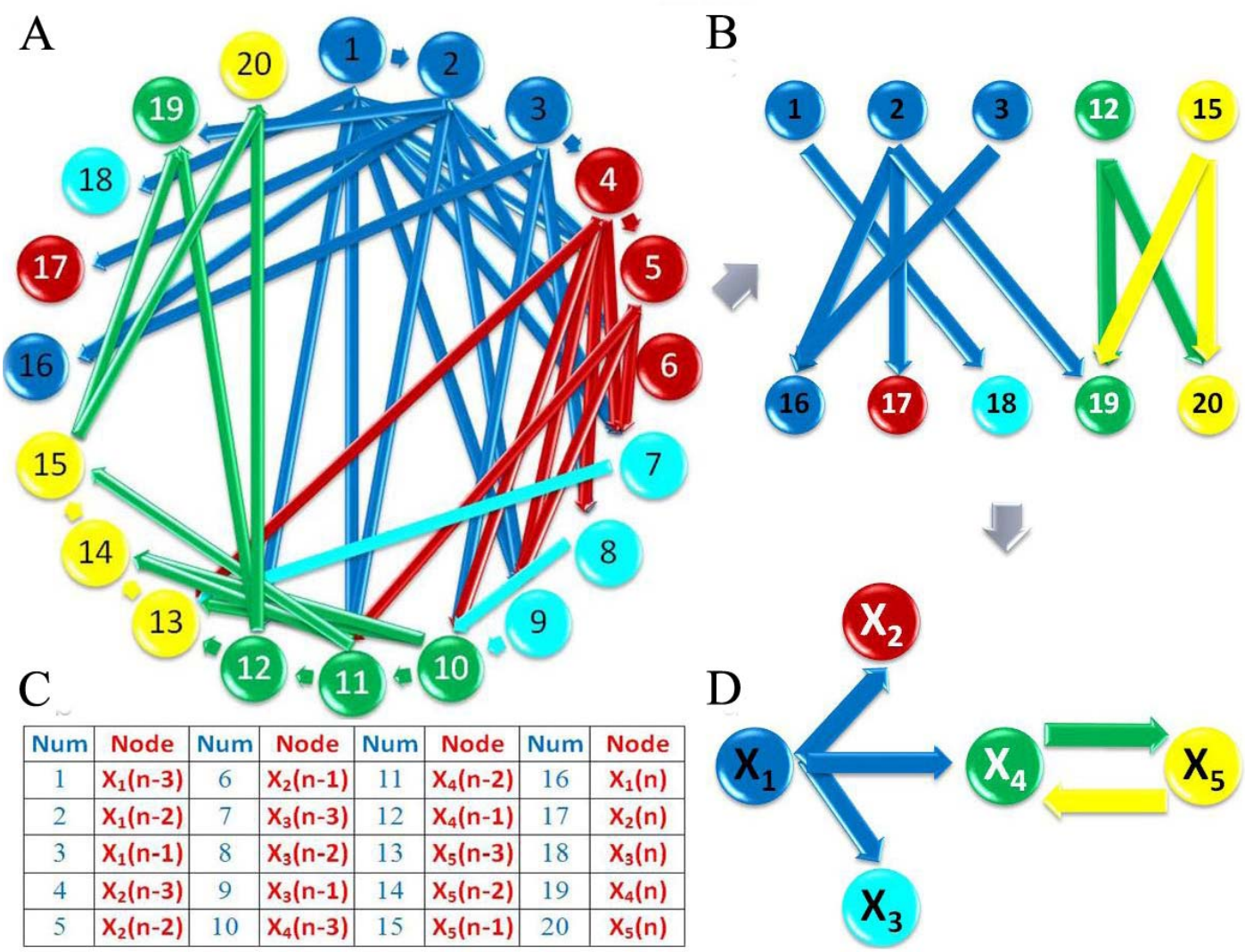
causal arcs (the arcs occur more than 95% of the time in the whole population) between various variables [Sachs, 2005]. This network contained the information of different time-lags for each variable. It fitted exactly the pattern of connectivity in our VAR model. In order to compare it with the Granger causality approach, we further simplified the network by only keeping the current status of variables and their parents, and hiding the information of time-lags. Thus we inferred exactly the same structure as the Granger causality approach (**Figure 3.2_D**). From this simple example, we found that both approaches could reveal correct network structures for the data with a large sample size (1000 here).

Most, if not all, experimental data has a very limited time step due to various experimental restrictions. Hence one of the key quantities to test the reliability of an approach is the data length (sample size). In the next setup, we reduced the sample size to a smaller value and checked its impact. The minimum sample size 20 was chosen since both approaches failed to detect any links under sample size 20. In order to find out the reliability changes with different sample size, we simply increased 20 time points for each step from



Num	Edge	Num	Edge
1	1←2	11	3←4
2	1←3	12	3←5
3	1←4	13	4←1
4	1←5	14	4←2
5	2←1	15	4←3
6	2←3	16	4←5
7	2←4	17	5←1
8	2←5	18	5←2
9	3←1	19	5←3
10	3←2	20	5←4

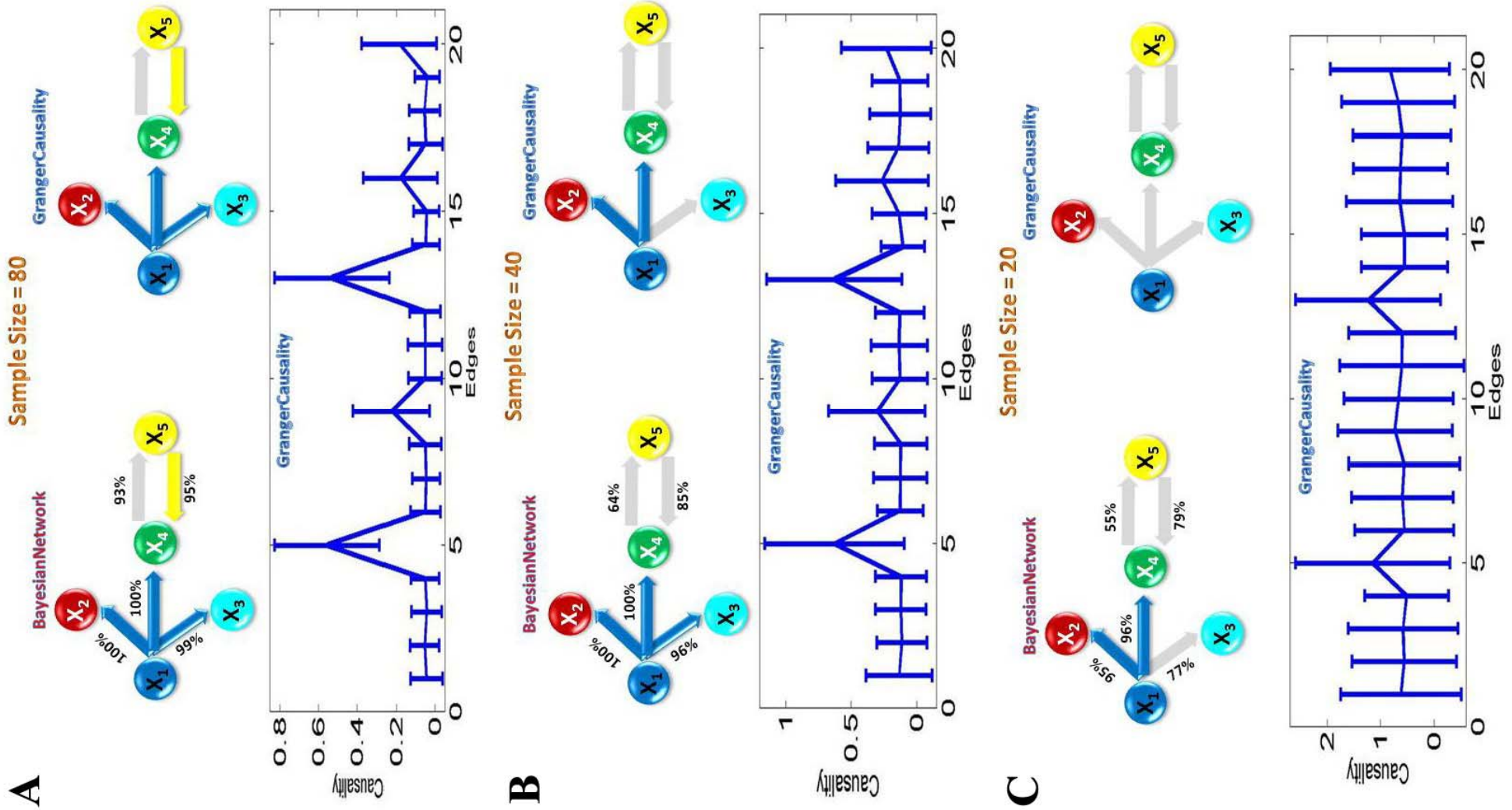
Figure 3.1 Granger causality applied on a simple linear toy model. (**A**) Five time series are simultaneously generated, and the length of each time series is 1000. X_2, X_3, X_4 and X_5 are shifted upward for visualization purpose. (**B**) The network structure inferred from Granger causality approach. (**C**) The 95% confidence intervals graph for all the possible directed connections. For visualization purpose, all directed edges (causalities) are sorted and enumerated into the table. The total number of edges is 20 for this case of 5 variables.



Num	Node	Num	Node	Num	Node	Num	Node
1	$X_1(n-3)$	6	$X_2(n-1)$	11	$X_4(n-2)$	16	$X_1(n)$
2	$X_1(n-2)$	7	$X_3(n-3)$	12	$X_4(n-1)$	17	$X_2(n)$
3	$X_1(n-1)$	8	$X_3(n-2)$	13	$X_5(n-3)$	18	$X_3(n)$
4	$X_2(n-3)$	9	$X_3(n-1)$	14	$X_5(n-2)$	19	$X_4(n)$
5	$X_2(n-2)$	10	$X_4(n-3)$	15	$X_5(n-1)$	20	$X_5(n)$

Figure 3.2 Dynamic Bayesian network inference results for the linear synthesized data. (**A**) The complete causal network structure derived by using dynamic Bayesian network inference method. Each variable is represented by 4 nodes (determined by the order of the model we chose) with different time-lags, thus we have a total of 20 nodes. (**B**) The simplified network structure: since we are only concerned the causality to the current time status, we can remove all the other edges and nodes that have no connection to the nodes 16 to 20 (five variables with current time status). Thus only nodes 16 to 20 and their parents are left. (**C**) For visualization purpose, all nodes are sorted and enumerated into the table. (**D**) We can further simplify the network structure derived by dynamic Bayesian network by removing the information of time lags.

Figure 3.3 Granger causality and dynamic Bayesian network inference approaches applied on data points of various sample sizes. They grey edges in the derived network structures indicate undetected causalities (false negative) in the toy model. For each sample size n, we simulated a data set of 100 realizations of n time points. The high-confidence arcs derived by dynamic Bayesian inference are shown. High-confidence arcs are the ones appearing in more than 95% of the derived networks for each realizations. The network structure derived by Granger causality is according to the 95% confidence interval constructed by using the bootstrap method. (**A**) The sample size is 80 (**B**) The sample size is 60 (**C**) The sample size is 20.



minimum sample size 20. **Figure 3.3_A** shows the case of the sample size of 80: we found both approaches start failing to detect some interactions (false negative). By reducing the sample size to 20 (**Figure 3.3_C**), we found that the Bayesian network inference could derive more true positive connections than the Granger causality. This is certainly an interesting phenomenon and we intend to explore whether it is true for a more general case.

Example 2 we considered a more general toy model; the coefficients in the equations (3.3.1) of Example 1 were randomly generated. This toy model aimed to test the causality sensitivity for the two approaches. Suppose 5 simultaneously generated time series according to the equations:

$$\begin{cases} X_1(n) = w_1 X_1(n-1) + w_2 X_1(n-2) + \varepsilon_1 \\ X_2(n) = w_3 X_1(n-2) + \varepsilon_2 \\ X_3(n) = w_4 X_1(n-3) + \varepsilon_3 \\ X_4(n) = w_5 X_1(n-1) + w_6 X_4(n-1) + w_7 X_5(n-1) + \varepsilon_4 \\ X_5(n) = w_8 X_4(n-1) + w_9 X_5(n-1) + \varepsilon_5 \end{cases} \tag{3.3.2}$$

where $[w_1, w_2, \cdots, w_9]$ are uniformly distributed random variables in the interval [-1,1]. The randomly generated coefficients are also required to make the system stable. The stability can be tested by using the z-plane pole-zero method, which states if the outermost poles of the z-transfer function describing the time

series are inside the unit circle on the z-plane pole-zero plots, then the system is stable [Michael, 2005].

The above toy model was then used to test the two different causality approaches: Bayesian network inference and Granger causality. They were applied with different sample sizes. For each sample size, we randomly generated 100 different coefficient vectors $[w_1, w_2, \cdots, w_9]$, which corresponded to 100 different toy models in **Example 1**. For each different coefficient vector model, we applied the same approach as in **Example 1**, using Monte Carlo method to construct 95% confidence intervals for the Granger causality approach and choosing high-confidence arcs (appearing in at least 95% of all samplings) for the Bayesian network inference approach. The total number of arcs (or causalities) was 500 (5 interactions for each realization) for each sample size. However we could not expect to detect the maximum number of arcs in our system, since the coefficients were randomly generated and could be relatively small.

Figure 3.4_A shows the comparison result of the percentage of true positive connections derived from these two methods. In general, the Granger causality approach can infer slightly more true positive causalities compared to the Bayesian network inference approach when the data length is long. It is interesting to see that there is a critical point at around 30 in **Figure 3.4_A**. If the

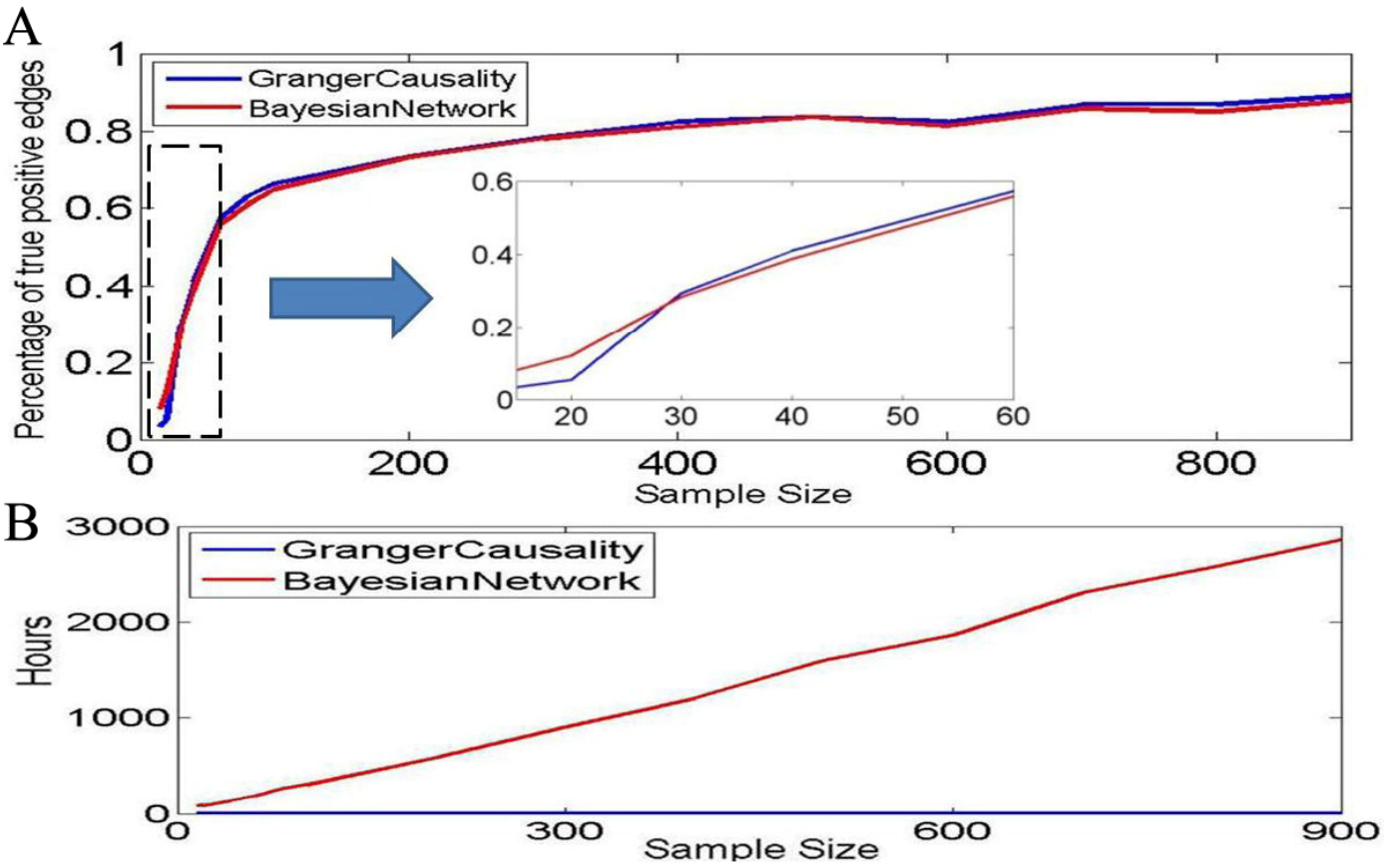


Figure 3.4 Granger causality and dynamic Bayesian network inference applied on a stochastic coefficients toy model. The parameters in polynomial equation are randomly generated in the interval [-1, 1]. For each randomly generated coefficient vector, we applied the same approach as **example 1**: bootstrapping method and 95% confidence interval were chosen for Granger causality approach; 95% high-confidence arcs are chosen from Bayesian network inference. (**A**) Both approaches were applied on different sample size (from 20 to 900). For each sample size, we generated 100 realizations with different coefficient vectors. The percentage of detected true positive causalities for both approaches is plot. (**B**) The total execution time cost for Granger causality and dynamic Bayesian network.

sample size is larger than 30, then the Bayesian network recovers less positive connections. However, if the sample size is smaller than 30, the Bayesian network performs better. From **Figure 3.4_B**, we see that computing time for the Bayesian network inference is much larger than the Granger causality.

Now we are in the position to find out why the dynamic Bayesian network is better than Granger causality when the data length is short, and vice versa. In **Figure 3.5_A**, we compared the performances on different coefficients (strength of interaction) for a fixed sample size of 900 (super-critical case). The x-axis shows the absolute value of coefficients, and y shows the corresponding causality (1 indicates positive causality and 0 indicates no causality). For visualization purposes, the figure for the Granger causality is shifted upward. From the five graphs, we can see that there is no difference between these two approaches if the coefficients are significant large (strong interactions with an absolute value of coefficients being greater than 0.15): both approaches can infer the correct connections. For most cases, the Granger causality approach performs with more stability when the coefficients are larger than 0.15, and the Bayesian network inference approach shows slightly more oscillations around this point. Hence we conclude that the Granger causality is less sensitive to the small value of the connection when the data length is large (see also the nonlinear case below).

We then compared the fitting accuracy of the two approaches, as shown in **Figure 3.5_B**. We used the average mean-square error as a measurement of the fitting. Not surprisingly, the dynamic Bayesian network approach considerably outperformed the simple fitting algorithm in the Granger approach [Geweke, 1982 & 1984], in particular when the data length was short.

In conclusion, when the data is a reasonable fit to the original model, Granger causality works better. This is due to the fact that the Granger causality approach is more sensitive to a small value of the interactions. When the data length is short, the Bayesian approach can fit the data much more reliably and it outperforms the Granger approach.

3.3.2 Synthesized data: non-linear case

In real situations, all data should be nonlinear and a linear relationship as described above is only an approximation. To address the nonlinear issue, we turn our attention to kernel models. As we know, any nonlinear relationship can be approximated by a series of kernel functions [Marinazzo, 2008].

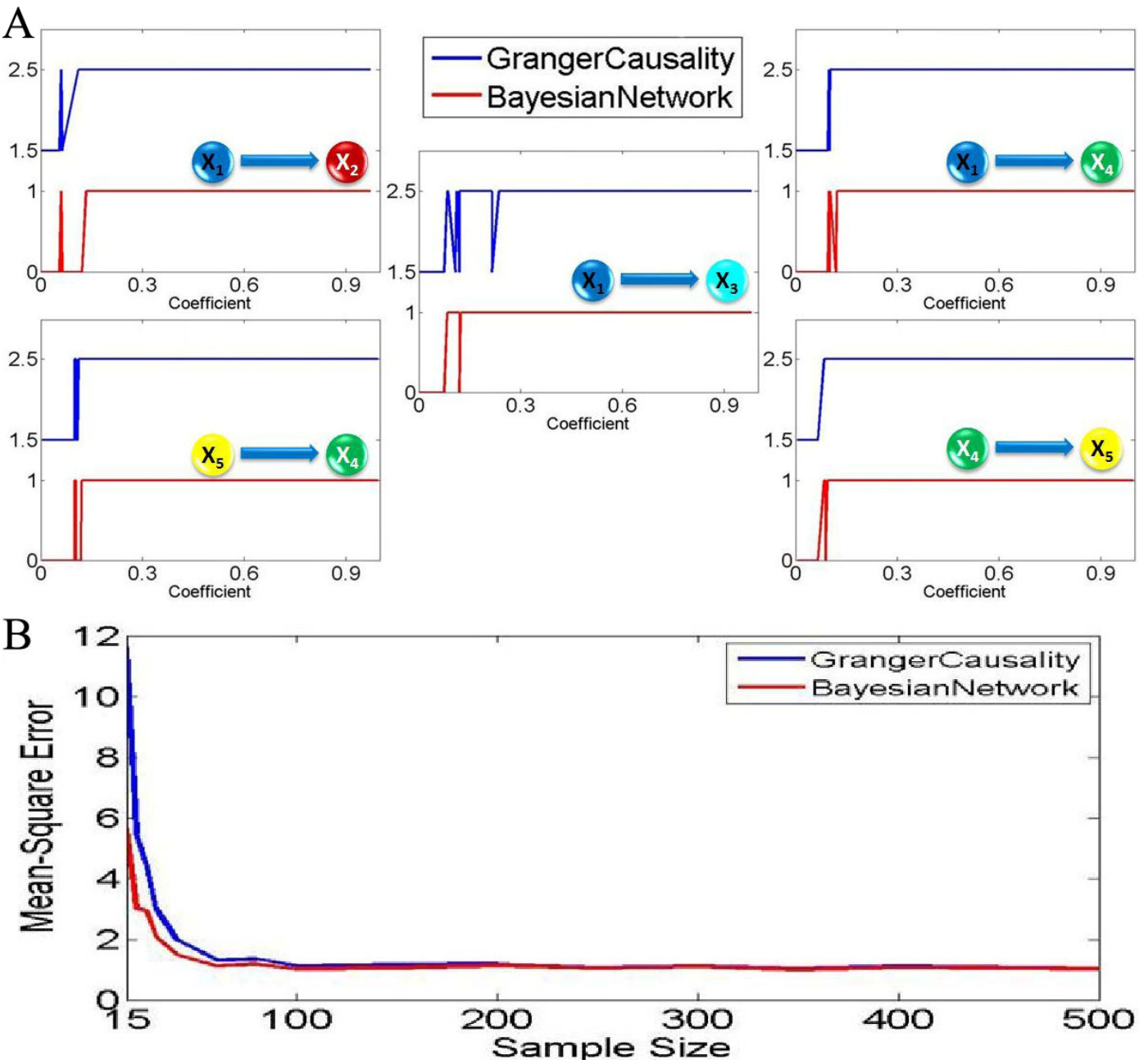


Figure 3.5 Sensitivity test for Granger causality and dynamic Bayesian inference applied on the stochastic coefficients toy model. (A) For sample size 900, the derived causality (1 represents positive causality and 0 represents negative) is plotted with the absolute value of corresponding coefficients. For visualization purpose, the figure for Granger causality is shifted upward. (B) Linear model fitting comparison for both Granger causality and dynamic Bayesian network inference. Using a number of training data points to fit both linear models, one can calculate the corresponding predicted mean-square error by appling a set of test data. It demonstrated that the Bayesian network inference method worked much better than the Granger causality approach when the sample size is significant small (around 100). When the sample size is significant large, both approach converge the standard error which exactly fits the noise term in the toy model.

Example 3 we modify the model in example 1 to a series of nonlinear equations as follows:

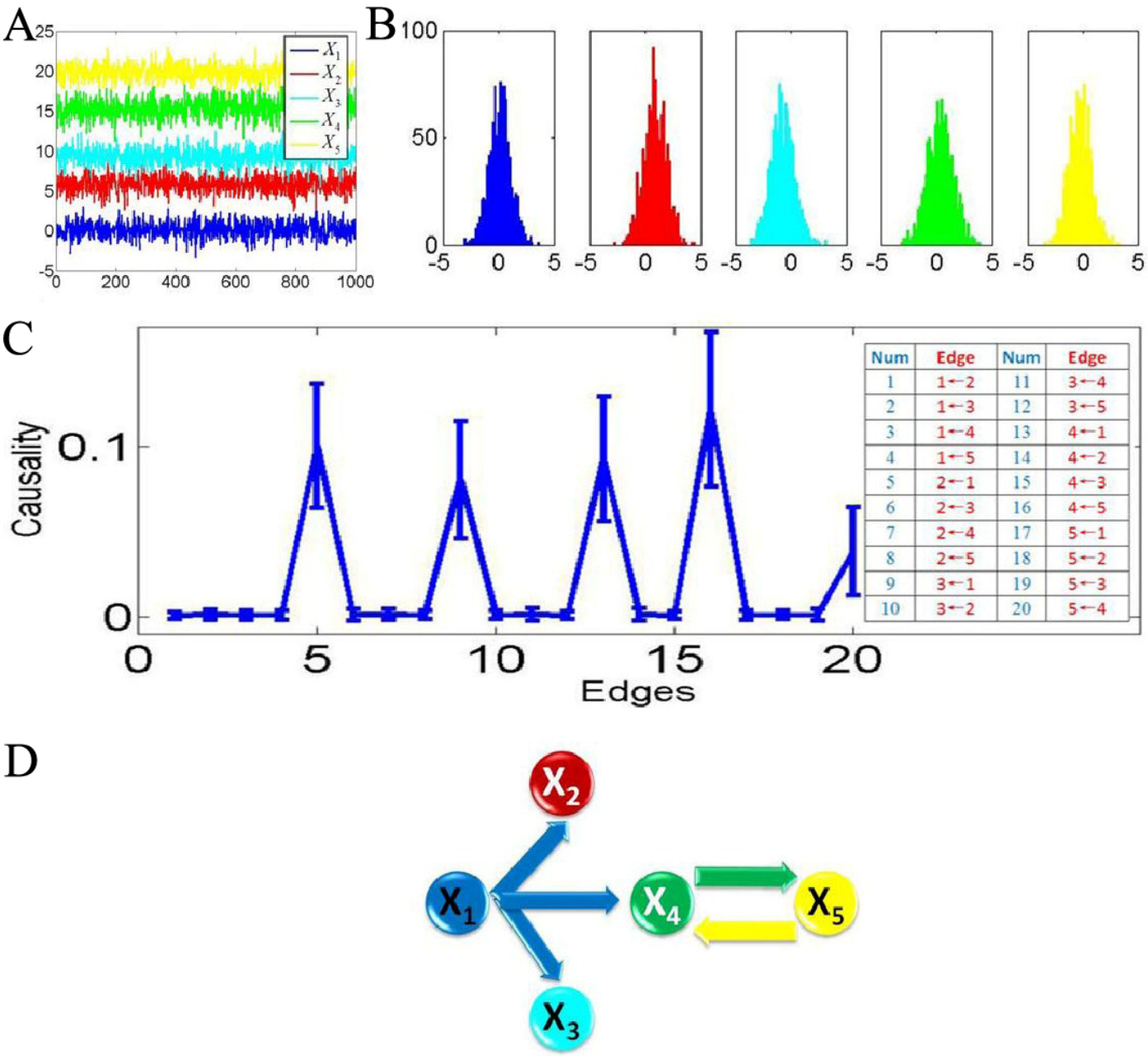
$$\begin{cases} X_1(n) = 0.125\sqrt{2}\exp(-\dfrac{X_1(n-1)^2}{2}) + \varepsilon_1 \\ X_2(n) = 1.2\exp(-\dfrac{X_1(n-1)^2}{2}) + \varepsilon_2 \\ X_3(n) = -1.05\exp(-\dfrac{X_1(n-1)^2}{2}) + \varepsilon_3 \\ X_4(n) = -1.15\exp(-\dfrac{X_1(n-1)^2}{2}) + 0.2\sqrt{2}\exp(-\dfrac{X_4(n-1)^2}{2}) \\ \qquad\qquad + 1.35\exp(-\dfrac{X_5(n-1)^2}{2}) + \varepsilon_4 \\ X_5(n) = -0.5\sqrt{2}\exp(-\dfrac{X_4(n-1)^2}{2}) + 0.25\sqrt{2}\exp(-\dfrac{X_5(n-1)^2}{2}) + \varepsilon_5 \end{cases} \tag{3.3.3}$$

In this example, the centre and variance of each time series was chosen as the centre and variance in the kernel function. We used the fuzzy c-mean method [Chuai-aree, 2001] to find the centre of each time series and then applied the same approach as in Example 1. In Fuzzy clustering, each point has a degree of belonging to clusters rather than completely belonging to just one cluster. For the Granger causality approach, we simulated the fitted VAR model to generate a data set of 100 realizations of 1000 time points (**Figure 3.6_A**), and applied the bootstrap approach to construct the 95% confidence intervals (**Figure 3.6_C**). According to the confidence interval, one derived the network structure (**Figure 3.6_D**) which correctly recovered the pattern of connectivity in our

non-linear model. For the Bayesian network inference approach, we inferred a network structure (**Figure 3.7_A**) for each realization of 1000 time points. We then obtained a simplified network structure (**Figure 3.7_C** and **Figure 3.7_D**).

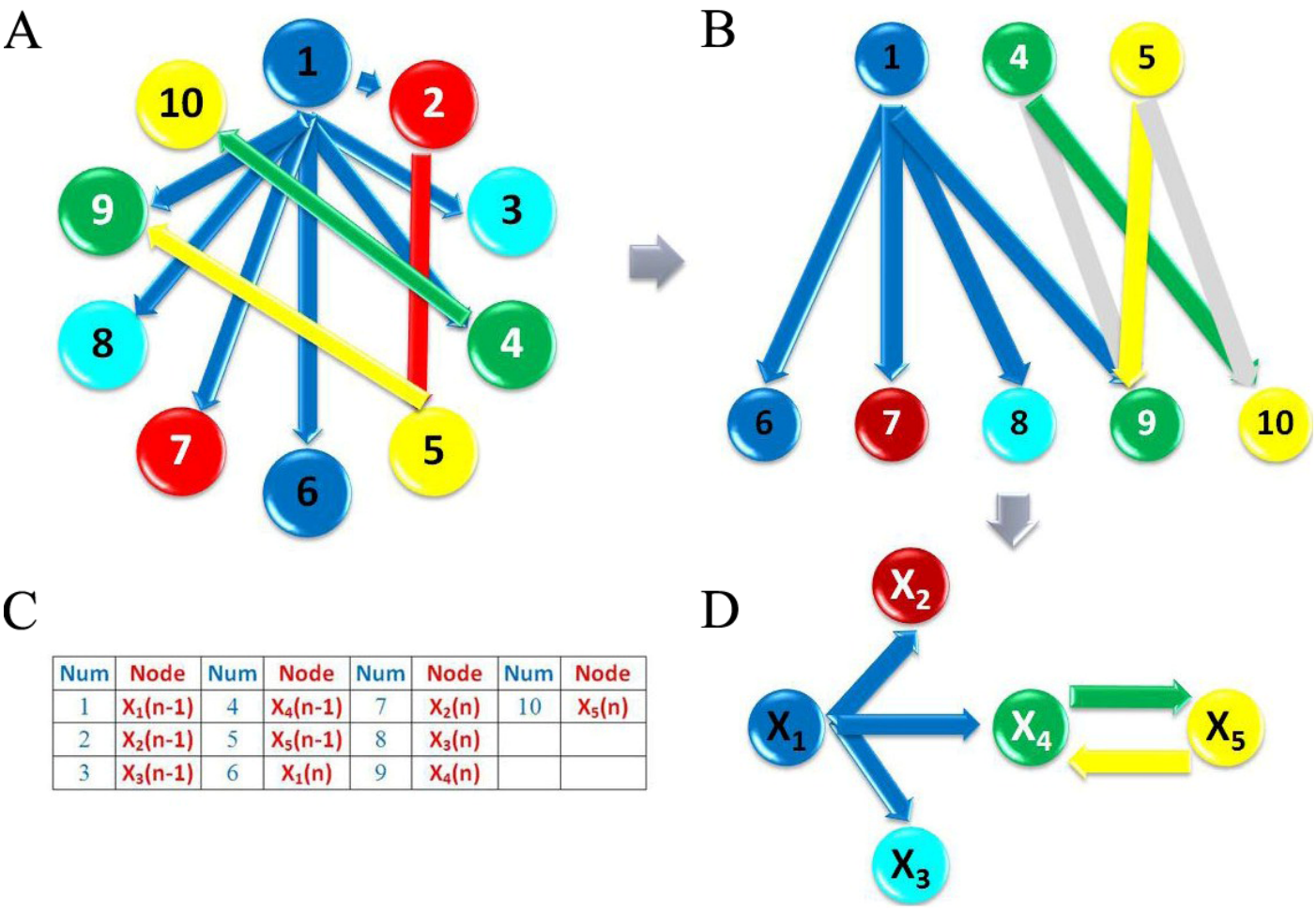
For a small sample size (see **Figure 3.8**), worse results were obtained for both approaches comparing to the previous linear model. Both approaches started to miss interactions when the sample size was smaller than 300. When the sample size was 150, the Bayesian network inference approach could detect one more true positive interaction than the Granger causality. However, when the sample size was 50, both approaches failed to detect all the interactions.

In the next step, we extended our non-linear model to a more general setting in which the coefficients in the equations were randomly generated. **Figure 3.9_A** shows the comparison result of the percentage of true positive connections derived from these two methods. It is very interesting to see that a critical point around 500 exists in the non-linear model, similar to the linear model before. From **Figure 3.9_B**, the computing time required for the Bayesian network inference is still much larger than the Granger causality. In **Figure 3.9_C**, we compared the performances on different coefficients (strength of interaction) for a fixed sample size of 900. From the five graphs, we can see that in general the Granger approach is more sensitive to a small value of the coefficients for non-linear case (see **Figure 3.9_C**. $X_5 \to X_4$ and $X_4 \to X_5$).



Num	Edge	Num	Edge
1	1←2	11	3←4
2	1←3	12	3←5
3	1←4	13	4←1
4	1←5	14	4←2
5	2←1	15	4←3
6	2←3	16	4←5
7	2←4	17	5←1
8	2←5	18	5←2
9	3←1	19	5←3
10	3←2	20	5←4

Figure 3.6 Granger causality applied on the simple non-linear toy model. (**A**) Five time series are simultaneously generated, and the length of each time series is 1000. They are assumed to be stationary. (**B**) Five histogram graphs show the probability distribution for these five time series. (**C**) Assuming no knowledge of the toy model we generated, Granger causality approach was applied. Bootstrapping approach was used to construct the 95% confidence intervals. The fitted MVAR model was then used to simulate a data set of 100 realizations of 1000 time points. For visualization purpose, all directed edges (causalities) are sorted and enumerated into the table. (**D**) The network structure derived from the Granger causality method correctly recovers the pattern of connectivity in the toy model.



Num	Node	Num	Node	Num	Node	Num	Node
1	$X_1(n-1)$	4	$X_4(n-1)$	7	$X_2(n)$	10	$X_5(n)$
2	$X_2(n-1)$	5	$X_5(n-1)$	8	$X_3(n)$		
3	$X_3(n-1)$	6	$X_1(n)$	9	$X_4(n)$		

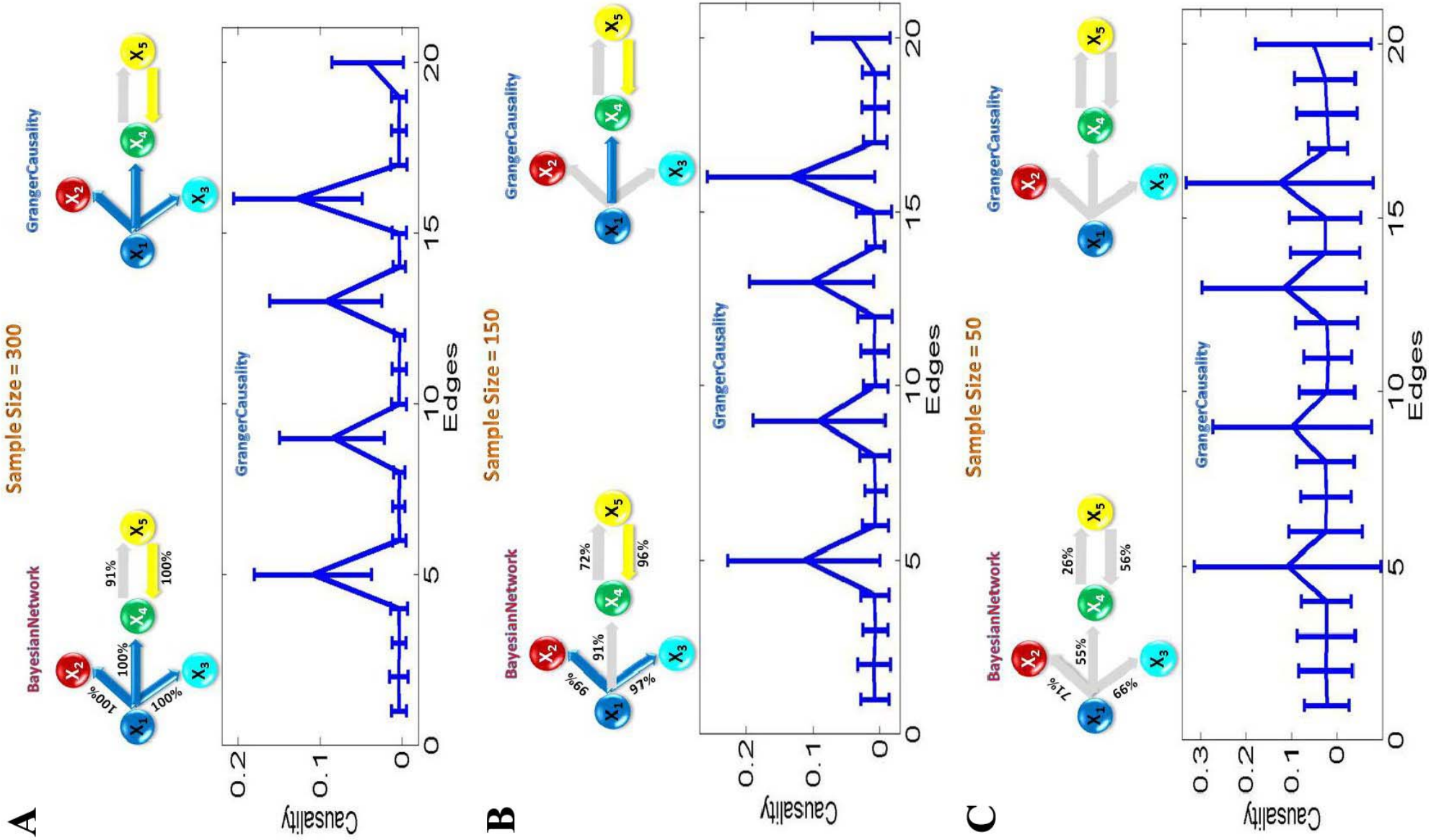
Figure 3.7 Dynamic Bayesian network inference applied on a simple non-linear toy model. (**A**) Assuming no knowledge of the MVAR toy model used for generating the sample data, the complete directed network structure (including the time lag information) inferred by dynamic Bayesian network method for one realization of 1000 time points. (**B**) A simplified network structure by removing all the edges and nodes that have no connection to the current status (nodes 6 to 10). (**C**) For visualization purpose, all nodes are sorted and enumerated into the table. (**D**) A further simplified network by hiding the information of time lags.

Therefore, all conclusions in the linear case were confirmed in the nonlinear model. In the literature [Marinazzo, 2008], the results they obtained are similar as we did here, which found that the Granger causality performed better than the dynamic Bayesian network inference concerning a nonlinear kernel model of genetic regulatory pathways and for a sufficiently large sample size (2000 data points).

3.3.2 Experimental data

Finally we carried out a study on experimental data of microarray experiments. The gene data were collected from two cases of *Arabidopsis* leaf: the mock (normal) case and the infected case with the plant pathogen *Botrytis cinerea*. A total of 31,000 genes were measured with a time interval of two hours, with a total of 24 sampling points (two days) and four replicates. We tested the Granger causality approach and dynamic Bayesian network inference approach on a well-known circadian circuit. This circuit contains 7 genes: PRR7, GI, PRR9, ELF4, LHY, CCA1 and TOC1. **Figure 3.10_A** shows the time traces of the 7 genes. From the time traces figure, it is clear to see that they exhibit a 24 hour rhythm. Note that the total number of time points is only 24. Compared to our previous toy model case, this sample size is quite small. We therefore expected the Bayesian network inference to be more reliable.

Figure 3.8 Granger causality and dynamic Bayesian network inference applied on insufficient number of data points for non-linear model. They grey edges in the re-constructed network indicate undetected causalities in our pre-defined toy model. (A) Sample size is 300. (B) Sample size is 150 (C) Sample size is 50.



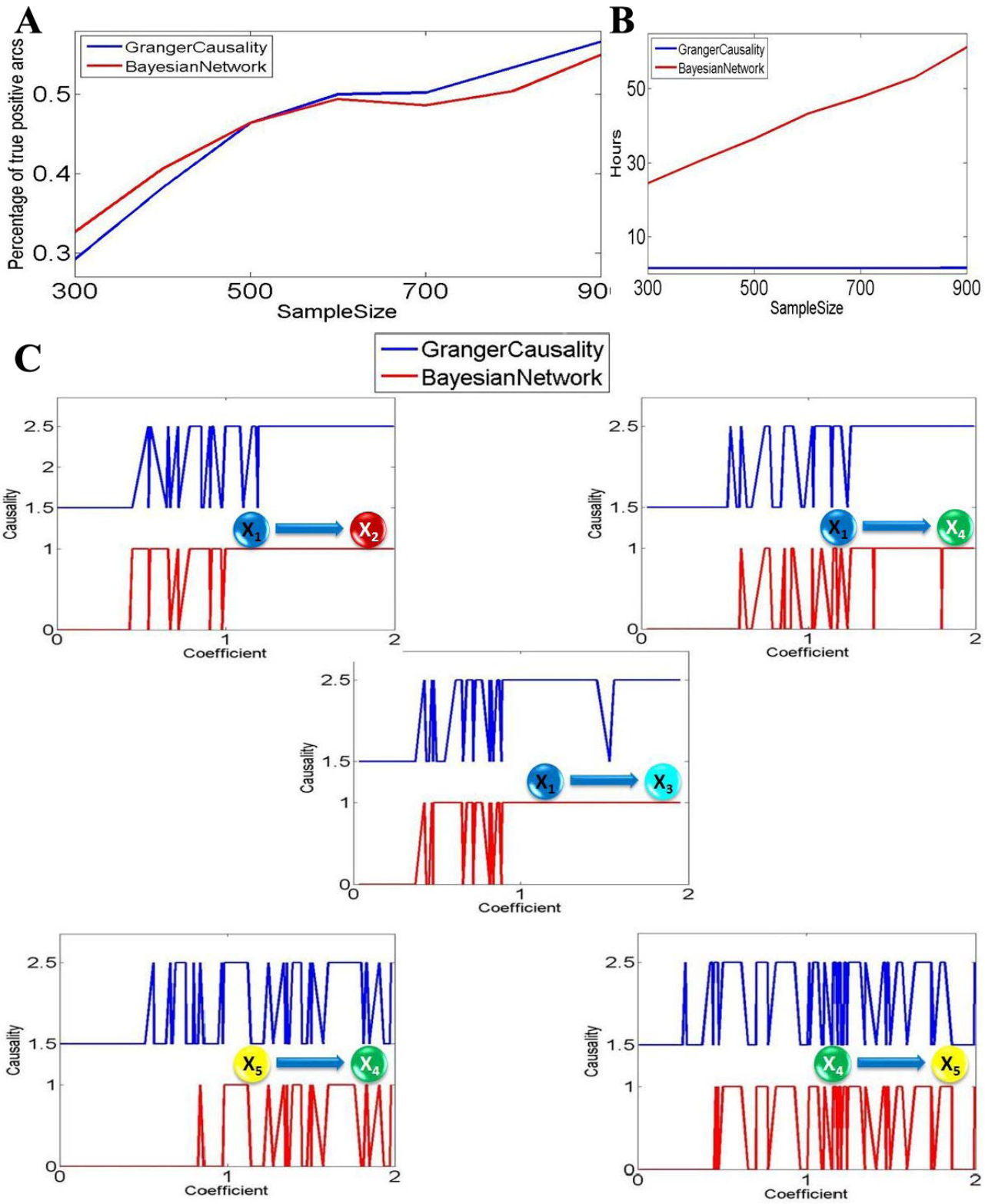


Figure 3.9 Granger causality and dynamic Bayesian network inference applied on a stochastic coefficients non-linear model. The parameters in polynomial equations are randomly generated in the interval [-2,2]. Both approaches were applied on different sample size (from 300 to 900). (A) The percentage of detected true positive causalities for both approaches. (B) Time cost comparison between both approaches. (C) Sensitive test for sample size 900, the derived causalities (1 represents positive causality and 0 represents negative) is plotted with the absolute value of corresponding coefficients. For visualization purpose, the figure for Granger causality was shifted upward.

We first applied the dynamic Bayesian network inference approach on these two data sets. The two network structures for the two cases are shown in **Figure 3.10_B** and **Figure 3.10_C**. In the next step, the conditional Granger causality approach was applied. By using the bootstrapping method, we constructed 95% confidence intervals as shown in **Figure 3.11_C**. Finally, we could also obtain two network structures for two different cases shown in **Figure 3.11_A** and **Figure 3.11_B**. It is clearly seen that the global patterns for the mock case and the infected case are different.

From the literature, there are three well known connections among the whole structure for the mock case. (1) It is known that GI alone is independent of the remaining six genes in the circuit. There should be no connection to and from the GI node (node 2 in the **Figure 3.10** and **Figure 3.11**) in our derived network. From **Figure 3.10_B** and **Figure 3.11_A**, we found that the dynamic Bayesian network inference method clearly picked this up, but the conditional Granger causality approach failed to detect this property. The Granger causality approach derived two false positive arcs which were connected to a GI node as shown in **Figure 3.11_A**. (2) It is known that PRR7 and LHY share a feedback loop. In other words, there should be two directed arcs connected from node 1 (PRR7) to node 5 (LHY) and from node 5 to node 1. The network structures derived from both approaches are in agreement with this known relationship. (3) It is known that ELF4 has some interactions with both LHY and CCA1. There

should be some connections between node 4 (ELF4) to node 5 (LHY), and between node 4 (ELF4) and node 6 (CCA1). From our derived structures, both approaches can detect these connections, which are in agreement with the known structure in the literature [Locke, 2006; Ueda, 2006].

According these three known relationships in the structure, we found that the Bayesian network structure was in agreement with all three rules, but the network structure derived from the conditional Granger causality was not: two more false positive interactions were found. Again for a small sample size, the Bayesian network inference approach could be more reliable than the conditional Granger causality approach.

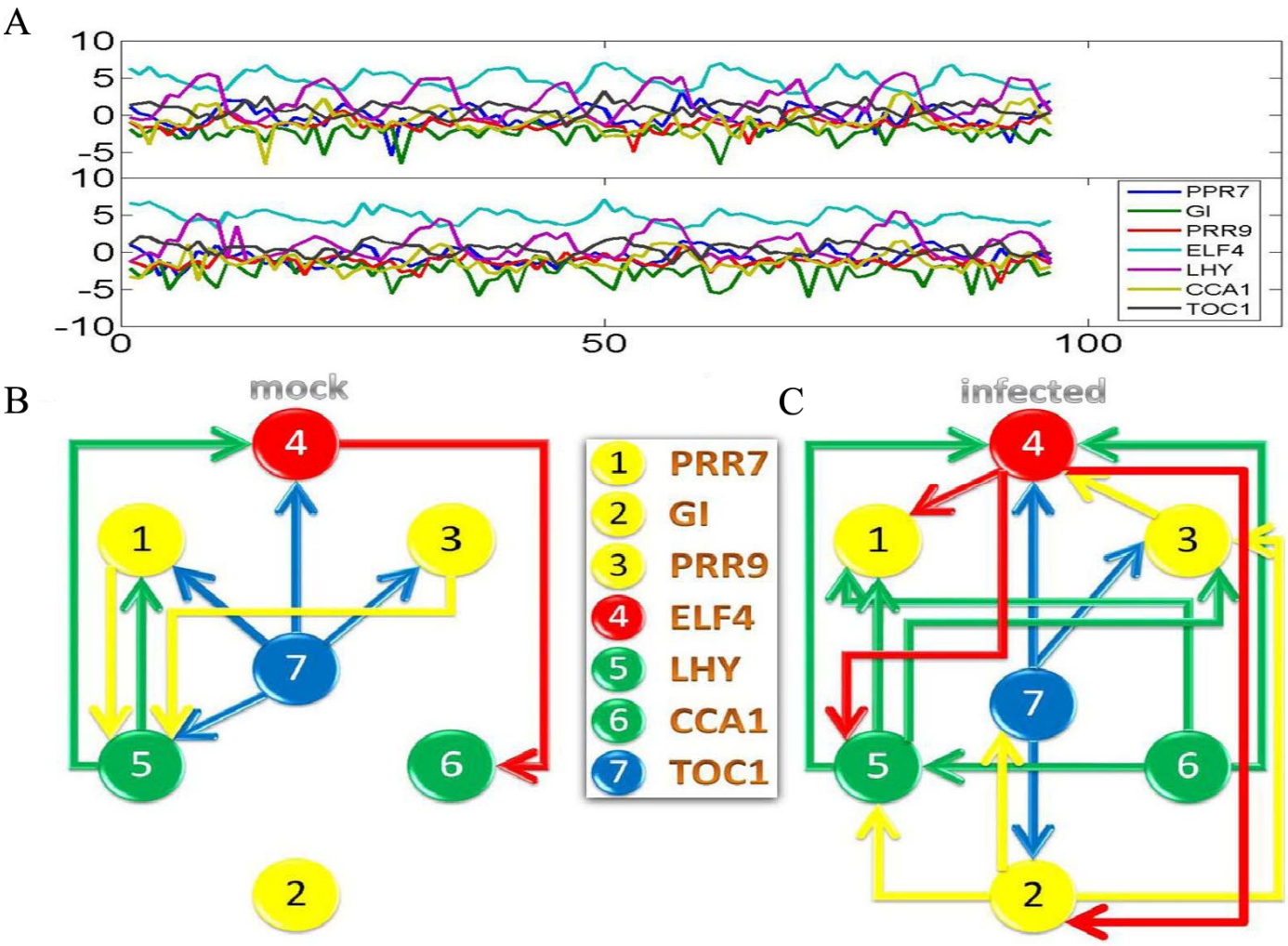
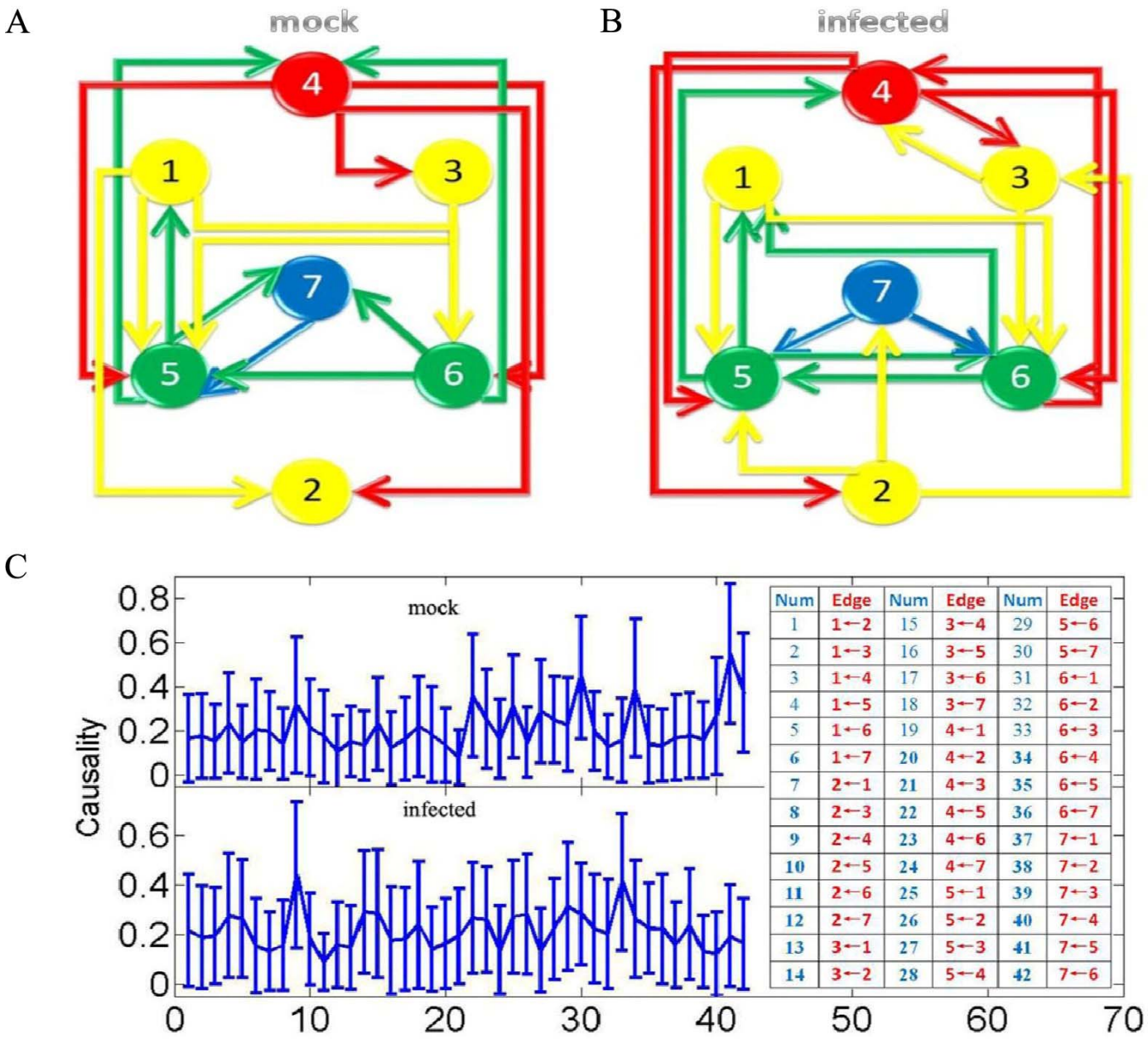


Figure 3.10 Dynamic Bayesian network inference method applied on experimental data (small sample size). The experiment measures the intensity of 7 genes in two causes of Arabidopsis leaf: mock (normal) and infected. (**A**) The time traces of 7 genes are plotted. There are 4 repeats of 24 time points. The time interval is 2 hours. (**B**) The network structures derived for mock case by using dynamic Bayesian network inference. (**C**) The network structures derived for infected case. All the genes are numbered into a table. The results illustrated that the network structure after infection was changed.



Num	Edge	Num	Edge	Num	Edge
1	1←2	15	3←4	29	5←6
2	1←3	16	3←5	30	5←7
3	1←4	17	3←6	31	6←1
4	1←5	18	3←7	32	6←2
5	1←6	19	4←1	33	6←3
6	1←7	20	4←2	34	6←4
7	2←1	21	4←3	35	6←5
8	2←3	22	4←5	36	6←7
9	2←4	23	4←6	37	7←1
10	2←5	24	4←7	38	7←2
11	2←6	25	5←1	39	7←3
12	2←7	26	5←2	40	7←4
13	3←1	27	5←3	41	7←5
14	3←2	28	5←4	42	7←6

Figure 3.11 Granger causality approach applied on experimental data of 7 genes in two cases of Arabidopsis Leaf. (**A**) The network structures derived for mock case by using Granger causality approach. (**B**) The network structures derived for infected case. (**C**) Using bootstrapping method to construct 95% confidence intervals. For visualization purpose, all the directed edges are numbered and enumerate them into the table.

3.4 Discussion

3.4.1 A fair comparison

In our results presented here, one of the key issues which is the cause of the critical point of the sampling size between the dynamic Bayesian approach and the Granger causality lies in the fact that a batch fitting approach is used in the Granger causality approach. One might argue that we could use the sequential fitting approach as in the Bayesian network to improve the performance of the Granger causality approach. This is certainly the case. However, due to the many publications in both topics [ISI Web of Knowledge, 2008], we simply adopted the most common approaches in the dynamic Bayesian network approach and the Granger causality. Developing one of the approaches, for example the Granger causality, so that it could always outperform the other is an interesting future research topic.

3.4.2 How long is long enough?

Although we have found the critical point of the two approaches, in practical applications, we have no certainty where the critical point is. Hence, we still have to choose one of them to tackle the data. In molecular biology, we

have to deal with a very limited data size; but in physiology, for example, neurophysiology, the data we record is usually very long. Hence one could argue that we use the dynamic Bayesian network in gene, protein or metabolite data, and apply the Granger causality to physiology data. The dynamic Bayesian network is more often reported in molecular biology, but the Granger causality has been very successfully applied in neurophysiological data [Wang, 2007] and fMRI. The result we chose to use was always chosen through experimental validation, as we did here for the plant data.

3.4.3 Frequency decomposition

As we emphasized in **Chapter 2.1**, the advantage of the Granger causality over the dynamic Bayesian network is the frequency decomposition, which is usually informative when we deal with temporal data. For example, in neurophysiology data, we know the brain employs different frequency bands to communicate between neurons and brain areas [Wu, 2007; Zhan, 2006]. We would expect a similar situation to arise in genes, proteins and metabolites, although we lack a detailed analysis due to the limited data length. To this end, we have also presented frequency decomposition results for the dynamic Bayesian network.

3.4.4 False positive

In our synthesized data, for both approaches, we did not find any false positive links in our experiments. However, there were a few false positive links found when we applied the conditional Granger causality and also partial Granger causality [Guo, 2008] on the gene data. One might ask why this is the case; there are several different reasons. Firstly, the experimental data is not strictly stationary: it is a natural process and evolves with time. As a first approximation, we treat it as stationary. Of course, we could use ARIMA rather than ARMA model to fit the data in the Granger causality. Secondly, the seven gene network is only a small network embedded in complete and large network, so there are latent variables. Using the partial Granger causality [Guo, 2008] which was originally developed for eliminating latent variables, gene GI still has links with the other six genes. Whether the dynamic Bayesian network could do a better job in the presence of latent variables is another research topic.

3.4.5 The meaning of the found motifs

Two circuits are found: one with the mock plant and one with the infected plant. The plant rewires its circadian circuit after infection. Ignoring the issue of identifying the molecular mechanisms which control circuit rewiring,

which is itself an interesting and challenging problem, we intend to discuss the functional meaning of the two circuits. To this end, we could assign a dynamics to the network and try to decipher the implications of the rewiring. Interestingly, we found that GI is recruited to save the network: if we leave GI as it is in the mock case, the whole network will rapidly converge to a fixed point state (a dead state).

3.4.6 Reasons for short size data

In our synthesized data, we test both short and long data samples and come to the conclusion that there is a critical size, at which the two approaches behave differently. In our experimental data, we only tested it for the short data set. Of course, as we mentioned above, in neurophysiological data, we have recordings of long time traces and the Granger causality is widely used there. However, we have to realize that all in vivo recordings are very dynamic and stationarity of data will become a key issue once we apply both approaches to a long dataset. Furthermore, when the dataset is long, both approaches could do well and it is more difficult to find the difference between the two. Hence we have only compared the results for short data length in the experimental setup.

3.4.7 Reasons for small size of variables

In our synthesized data, we only used 5 variables to simulate a small interacting network; the number of variables could affect the result we derived. As expected, see also [Marinazzo, 2008], the estimation of the Granger causality becomes unfeasible when the number of variables is large and the amount of the data sets is small. Hence, all results in the literature on estimating Granger causality are exclusive for small networks (around the order of 10), as we considered here. This is more or less true for dynamic Bayesian network inference as well. Extending the Granger causality and the dynamic Bayesian network inference to large networks is a challenging problem, even before we carry out the same comparison study on these two approaches as we did here.

3.5 Final Remark

In this chapter, we carried out a systematic and computationally intensive comparison between the two network structures derived from two common approaches: the dynamic Bayesian network inference and the Granger causality. These two approaches are applied on both synthesized and experimental data. For synthesized data (both linear model and non-linear model), a critical point of the data length was found: the Granger causality approach performed better

than the dynamic Bayesian network inference approach, when the data length was long enough, and vice versa. And the result was further confirmed in experimental data. This result leads us to focus on Granger causality on real experimental data when data length is relative large. However, there is a limitation for Granger causality as mentioned in this chapter. We have parameters estimation problem if the number of variables is larger than the length of the data, which is a common situation for deriving global (large) networks with thousands of genes, proteins and so on. This leads to our study in the next chapter-Identifying interactions in the time and frequency domains in local and global networks.

Chapter 4

Identifying interactions in the time and frequency domains in local and global networks

There are several well-established reverse-engineering approaches to explore causal relationships in a dynamic network, such as ordinary differential equations (ODE), Bayesian networks, information theory and Granger Causality, based upon multi-dimensional spatial and temporal data. In this chapter, one focus on Granger causality both in the time and frequency domain and in local and global networks, and apply our approach to experimental data (genes and proteins). For a small gene network, Granger causality outperformed all the other three approaches mentioned above. For large network reconstruction, a global protein network of 812 proteins was reconstructed, using a novel approach. The obtained results fitted well with known experimental findings and

predicted many experimentally testable results. In addition to interactions in the time domain, interactions in the frequency domain were also recovered. The results on the proteomic data and gene data confirm that Granger causality is a simple and accurate approach to recover the network structure. Our approach is general and can be easily applied to other types of temporal data.

4.1 Introduction

One of the most fundamental issues in computational biology is to reliably and accurately uncover the network structure of elements (genes, proteins, metabolites, neurons and brain areas etc.), based upon high throughput data [Alon, 2007; Klip, 2005]. There are several well-established reverse-engineering approaches to explore causal relationships in a dynamic network, such as ordinary differential equations (ODE), Bayesian networks, information theory and Granger Causality.

The notion of Granger causality, which was first introduced by Wiener and Granger [Wiener, 1956; Granger, 1969; Granger 1980], proposed that there is a causal influence from one time series to another if the prediction of one time series is improved with the knowledge of the second one. The prediction is made in terms of an auto-regressive model. In addition, Granger causality has the advantage of having a corresponding frequency domain decomposition so that one can clearly find at which frequencies two elements interact with each other. Granger's conception of causality has been widely and successfully applied in the econometrics literature and recently in the biological literature [Chen, 2004 & 2006; Feng, 2009; Guo, 2008; Ge, 2009; Marinazzo, 2008].

Considering the four different approaches to the same problem, a natural question is to investigate which should be preferred. In the previous chapter, we presented a comparison study of Granger causality and dynamic Bayesian network inference approaches. The result showed that Granger causality outperformed the dynamic Bayesian network inference when the time series were long enough because the Granger causality was then able to detect weak interactions. In a recent *Cell* paper [Camacho, 2009; Cantone, 2009], the authors carried out a systematic comparison between the ODE, Bayesian and information theoretic approaches for a small synthesized gene network in the *yeast* (*Saccharomyces cerevisiae*). The authors concluded that the ODE was the best approaches amongst those three. We have applied our conventional Granger causality approach on the same recorded time-series and found that the results derived by it were better than all the other three approaches' in the original paper. A small network of seven previously investigated proteins [Cohen, 2008] was also re-constructed. Interestingly, the two important proteins DDX5 and RFC1 found in experiments were at the top of the re-constructed network. Frequency domain results were analyzed and they indicated that DDX5 and BAG2 interacted at a frequency of one cycle per three hours.

In order to tackle the problem of large network reconstruction mentioned in **Chapter 1**, we propose a new framework: Global Granger Causality (GGC) This framework builds on the use of partial Granger causality which

was illustrated in [Guo, 2008]. We first construct an initial sparse network by considering all possible links by computing bivariate pair-wise Granger causality (described in **Chapter 2.1.1**). Once we identify such a network structure, there is uncertainty about the true causal structure; we need to check whether the links appearing in pairwise causality are direct or indirect. We do so by computing GGC step by step. If a link is found to be an indirect relationship in the sense of GGC, we delete such a link from the initial network. Theoretically, iterating the procedure will remove all indirect links and only direct connections will remain. The advantage of such an approach is obvious. By explicitly taking more sources into account, it provides a less biased structure of the network due to latent variables than in a small network as described above. It also provides information on the ancestors and descendents of key elements such as DDX and RFC1 in our network. The results can then guide experimentalists to investigate the properties of a small subset of specific proteins.

The rest of this chapter is divided in two sections. First, in the **Chapter 4.2**, we extended our Granger causality based on Autoregressive moving average) ARMA model to an Autoregressive integrated moving average (ARIMA) in details, as well as its formulation in the frequency domain. We also describe *global Granger causality*, the new procedure for applying Granger causality to large networks. Next, in the result section, we apply our method on small (local) and large (global) networks. In both cases, simulations and actual biological

data (gene and protein time-series) are used and results discussed. And we also provide a theoretical proof of its reliability.

4.2 Methods

In this section, we presented an analysis on how to define the conditional Granger causality on an ARIMA (autoregressive integrated moving average) model [Mills, 1990]. ARIMA is a generalization of an ARMA model specially used for dealing with non-stationary data, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied to remove the non-stationarity. The model is generally referred to as an ARIMA(p,d,q) model where p, d, and q natural numbers and refer to the order of the model. Given a time series of data X_t, an ARIMA(p,d,q) model is given by:

$$(1-\sum_{i=1}^{p}\alpha_i L^i)(1-L)^d X_t = (1+\sum_{i=1}^{q}\theta_i L^i)\varepsilon_t \qquad (4.2.1)$$

Where L is the lag operator, and the error term ε_t has normal distribution with 0 mean.

4.2.1 Conditional Granger Causality (ARIMA model)

Time domain Analysis

Giving two time series $\mathbf{X}_t$ and $\mathbf{Z}_t$ (can be represented vectors here) and their k[th] and m[th] order differences $\Delta^k\mathbf{X}_t$ and $\Delta^m\mathbf{Z}_t$ (without loss of generality, we assume that m=k from now on), the joint autoregressive representation for $\Delta^k\mathbf{X}_t$ and $\Delta^m\mathbf{Z}_t$ (as described in Equation 4.2.1) by using the knowledge of their past measurement can be expressed as

$$\begin{cases} \Delta^k\mathbf{X}_t = \sum_{i=1}^{\infty} a_{1i}\Delta^k\mathbf{X}_{t-i} + \sum_{i=1}^{\infty} c_{1i}\Delta^k\mathbf{Z}_{t-i} + \Delta\boldsymbol{\varepsilon}_{1t} \\ \Delta^k\mathbf{Z}_t = \sum_{i=1}^{\infty} b_{1i}\Delta^k\mathbf{Z}_{t-i} + \sum_{i=1}^{\infty} d_{1i}\Delta^k\mathbf{X}_{t-i} + \Delta\boldsymbol{\gamma}_{2t} \end{cases} \tag{4.2.2}$$

The noise covariance matrix for the system can be represented as

$$\Delta\mathbf{S} = \begin{bmatrix} \mathrm{var}(\Delta\boldsymbol{\varepsilon}_{1t}) & \mathrm{cov}(\Delta\boldsymbol{\varepsilon}_{1t}, \Delta\boldsymbol{\gamma}_{2t}) \\ \mathrm{cov}(\Delta\boldsymbol{\gamma}_{1t}, \Delta\boldsymbol{\varepsilon}_{1t}) & \mathrm{var}(\Delta\boldsymbol{\gamma}_{1t}) \end{bmatrix} = \begin{bmatrix} \Delta\mathbf{S}_{\mathbf{xx}} & \Delta\mathbf{S}_{\mathbf{xz}} \\ \Delta\mathbf{S}_{\mathbf{zx}} & \Delta\mathbf{S}_{\mathbf{zz}} \end{bmatrix} \tag{4.2.3}$$

where var and cov represent variance and co-variance respectively. Incorporating the knowledge of the third time series, the vector autoregressive mode involving the three time series $\Delta^k\mathbf{X}_t$, $\Delta^k\mathbf{Y}_t$ and $\Delta^k\mathbf{Z}_t$ can be represented as

$$\begin{cases} \Delta^k \mathbf{X}_t = \sum_{i=1}^{\infty} a_{2i} \Delta^k \mathbf{X}_{t-i} + \sum_{i=1}^{\infty} b_{2i} \Delta^k \mathbf{Y}_{t-i} + \sum_{i=1}^{\infty} c_{2i} \Delta^k \mathbf{Z}_{t-i} + \Delta \boldsymbol{\varepsilon}_{2t} \\ \Delta^k \mathbf{Y}_t = \sum_{i=1}^{\infty} d_{2i} \Delta^k \mathbf{X}_{t-i} + \sum_{i=1}^{\infty} e_{2i} \Delta^k \mathbf{Y}_{t-i} + \sum_{i=1}^{\infty} f_{2i} \Delta^k \mathbf{Z}_{t-i} + \Delta \boldsymbol{\eta}_{2t} \\ \Delta^k \mathbf{Z}_t = \sum_{i=1}^{\infty} g_{2i} \Delta^k \mathbf{X}_{t-i} + \sum_{i=1}^{\infty} h_{2i} \Delta^k \mathbf{Y}_{t-i} + \sum_{i=1}^{\infty} k_{2i} \Delta^k \mathbf{Z}_{t-i} + \Delta \boldsymbol{\gamma}_{2t} \end{cases} \quad (4.2.4)$$

And the noise covariance matrix for the above system is

$$\begin{aligned} \Delta\boldsymbol{\Sigma} &= \begin{bmatrix} \operatorname{var}(\boldsymbol{\varepsilon}_{2t}) & \operatorname{cov}(\boldsymbol{\varepsilon}_{2t}, \boldsymbol{\eta}_{2t}) & \operatorname{cov}(\boldsymbol{\varepsilon}_{2t}, \boldsymbol{\gamma}_{2t}) \\ \operatorname{cov}(\boldsymbol{\eta}_{2t}, \boldsymbol{\varepsilon}_{2t}) & \operatorname{var}(\boldsymbol{\eta}_{2t}) & \operatorname{cov}(\boldsymbol{\eta}_{2t}, \boldsymbol{\gamma}_{2t}) \\ \operatorname{cov}(\boldsymbol{\gamma}_{2t}, \boldsymbol{\varepsilon}_{2t}) & \operatorname{cov}(\boldsymbol{\gamma}_{2t}, \boldsymbol{\eta}_{2t}) & \operatorname{var}(\boldsymbol{\gamma}_{5t}) \end{bmatrix} \\ &= \begin{bmatrix} \Delta\boldsymbol{\Sigma}_{\mathbf{xx}} & \Delta\boldsymbol{\Sigma}_{\mathbf{xy}} & \Delta\boldsymbol{\Sigma}_{\mathbf{xz}} \\ \Delta\boldsymbol{\Sigma}_{\mathbf{yx}} & \Delta\boldsymbol{\Sigma}_{\mathbf{yy}} & \Delta\boldsymbol{\Sigma}_{\mathbf{yz}} \\ \Delta\boldsymbol{\Sigma}_{\mathbf{zx}} & \Delta\boldsymbol{\Sigma}_{\mathbf{zy}} & \Delta\boldsymbol{\Sigma}_{\mathbf{zz}} \end{bmatrix} \end{aligned} \quad (4.2.5)$$

where $\boldsymbol{\varepsilon}_{it}, i = 1,2,\cdots,5$ are the prediction errors, which are uncorrelated over time. If we rewrite Equations (4.2.2) and (4.2.4) in terms of $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ themselves, we see that whether a coefficient vanishes or not is almost unchanged. Hence it is safe to say that the conditional Granger causality form $\mathbf{Y}$ to $\mathbf{X}$ conditional on $\mathbf{Z}$ can be defined as (see [Ding, 2006] for the classical definition)

$$F_{\mathbf{Y}\to\mathbf{X}|\mathbf{Z}} = \ln\left(\frac{\left|\Delta\mathbf{S}_{\mathbf{xx}}\right|}{\left|\Delta\boldsymbol{\Sigma}_{\mathbf{xx}}\right|}\right) \quad (4.2.6)$$

When the causal influence from $\mathbf{Y}$ to $\mathbf{X}$ is entirely mediated by $\mathbf{Z}$, the coefficient b_{2i} is uniformly zero, and the two auto-regressive models for two or three time series will be exactly same, thus we can get $\text{var}(\boldsymbol{\varepsilon}_{1t}) = \text{var}(\boldsymbol{\varepsilon}_{3t})$. We then have $F_{\mathbf{Y}\to\mathbf{X}|\mathbf{Z}} = 0$, which means $\mathbf{Y}$ cannot further improve the prediction of $\mathbf{X}$ including past measurements of $\mathbf{Y}$ conditional on $\mathbf{Z}$. In other words, $\mathbf{Y}$ don't have an influence on $\mathbf{X}$. For $\Delta\mathbf{S}_{\mathbf{xx}} > \Delta\boldsymbol{\Sigma}_{\mathbf{xx}}$, $F_{\mathbf{Y}\to\mathbf{X}|\mathbf{Z}} > 0$ and therefore there is a direct influence from $\mathbf{Y}$ to $\mathbf{X}$, conditional on the past measurements of $\mathbf{Z}$.

Frequency domain analysis

To derive the spectral decomposition of the time domain conditional Granger causality, we multiply the normalization matrix

$$P_1 = \begin{pmatrix} I & 0 \\ -\Delta\mathbf{S}_{\mathbf{YX}}\Delta\mathbf{S}_{\mathbf{XX}}^{-1} & I \end{pmatrix} \tag{4.2.7}$$

to both side of Equation (4.2.2) and rewrite it in terms of the lag operator L. I is the identity matrix. The normalized equations are represented as:

$$\begin{pmatrix} D_{11}(L) & D_{12}(L) \\ D_{21}(L) & D_{22}(L) \end{pmatrix} \begin{pmatrix} \Delta^k \mathbf{X}_t \\ \Delta^k \mathbf{Z}_t \end{pmatrix} = \begin{pmatrix} \Delta^k \mathbf{X}_t^* \\ \Delta^k \mathbf{Z}_t^* \end{pmatrix} \tag{4.2.8}$$

Then we can apply the same normalization procedure to Equation (4.2.4) by multiplying with the matrix

$$P = P_3 \cdot P_2 \tag{4.2.9}$$

Where

$$P_2 = \begin{pmatrix} I & 0 & 0 \\ -\Delta\boldsymbol{\Sigma}_{\mathbf{yx}}\Delta\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} & I & 0 \\ -\Delta\boldsymbol{\Sigma}_{\mathbf{zx}}\Delta\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} & 0 & I \end{pmatrix} \tag{4.2.10}$$

And

$$P_3 = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -(\Delta\boldsymbol{\Sigma}_{\mathbf{zy}} - \Delta\boldsymbol{\Sigma}_{\mathbf{zx}}\Delta\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\Delta\boldsymbol{\Sigma}_{\mathbf{xy}})(\Delta\boldsymbol{\Sigma}_{\mathbf{yy}} - \Delta\boldsymbol{\Sigma}_{\mathbf{yx}}\Delta\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\Delta\boldsymbol{\Sigma}_{\mathbf{xy}})^{-1} & I \end{pmatrix} \tag{4.2.11}$$

to both sides of Equation (4.2.4) and rewrite it in terms of the lag operator

$$\begin{pmatrix} B_{11}(L) & B_{12}(L) & B_{13}(L) \\ B_{21}(L) & B_{22}(L) & B_{23}(L) \\ B_{31}(L) & B_{32}(L) & B_{33}(L) \end{pmatrix} \begin{pmatrix} \Delta^k \mathbf{X}_t \\ \Delta^k \mathbf{Y}_t \\ \Delta^k \mathbf{Z}_t \end{pmatrix} = \begin{pmatrix} \boldsymbol{\varepsilon}_{xt} \\ \boldsymbol{\varepsilon}_{yt} \\ \boldsymbol{\varepsilon}_{zt} \end{pmatrix} \tag{4.2.12}$$

After Fourier transforming Equation (4.2.8) and (4.2.12), we can rewrite them in the following representations

$$\begin{pmatrix}\mathbf{X}(\lambda)\\ \mathbf{Z}(\lambda)\end{pmatrix}\Delta(\lambda,k)=\begin{pmatrix}G_{\mathbf{xx}}(\lambda) & G_{\mathbf{xz}}(\lambda)\\ G_{\mathbf{zx}}(\lambda) & G_{\mathbf{zz}}(\lambda)\end{pmatrix}\begin{pmatrix}\mathbf{X}^{*}(\lambda)\\ \mathbf{Z}^{*}(\lambda)\end{pmatrix}\Delta(\lambda,k) \quad (4.2.13)$$

$$\begin{pmatrix}\mathbf{X}(\lambda)\\ \mathbf{Y}(\lambda)\\ \mathbf{Z}(\lambda)\end{pmatrix}\Delta(\lambda,k)=\begin{pmatrix}H_{\mathbf{xx}}(\lambda) & H_{\mathbf{xy}}(\lambda) & H_{\mathbf{xz}}(\lambda)\\ H_{\mathbf{yx}}(\lambda) & H_{\mathbf{yy}}(\lambda) & H_{\mathbf{yz}}(\lambda)\\ H_{\mathbf{zx}}(\lambda) & H_{\mathbf{zy}}(\lambda) & H_{\mathbf{zz}}(\lambda)\end{pmatrix}\begin{pmatrix}\mathbf{E}_{\mathbf{x}}(\lambda)\\ \mathbf{E}_{\mathbf{y}}(\lambda)\\ \mathbf{E}_{\mathbf{z}}(\lambda)\end{pmatrix}\Delta(\lambda,k) \quad (4.2.14)$$

where $\Delta(\lambda,k)$ is the Fourier transform of the difference operator Δ^{k}. Therefore, for ARIMA and ARMA models in the frequency domain, their causality is identical. This is in agreement with our conclusions in the time domain causality and in general the Kolmogorov identity holds true, that is: integrating the frequency-domain Granger causality over all frequencies yields the time domain Granger causality. Thus we can calculate the Granger causality in the frequency domain by using the similar Equations from (2.1.26) to (2.1.29).

4.2.2 Global Granger Causality

Partial Granger causality (PGC) provides an accurate description of the internal dynamics of the system when the number of nodes is much smaller than the length of recorded time series. However, when the number of nodes

increases, especially when they are larger than the length of time series, a problem of parameter fitting immediately arises, it is a situation for which usual methods break down.

Here we propose the following Global Granger Causality (GGC) algorithm to tackle this problem. The general idea is as follows: if we could find all ancestors of a given target T, the whole network could be reconstructed. Hence for a given target T, we want to find all directed ancestors (parents of target T). For illustration, a small subset of the whole network, which contains target T and all its ancestors, is shown in the **Figure 4.1_A**. We assume that each nodes from $\{X_1, \ldots, X_n\}$ has only a single pathway to target T, and each nodes from $\{Y_1, \ldots, Y_n\}$ has two distinct pathways to target T. From **Figure 4.1_A**, we can find the parents of target T are T_1, T_2, T_3.

The detailed algorithm is illustrated as follows:

First, apply the bivariate pair-wise Granger causality to find all of the ancestors of the target T. This set is denoted $A_0(T)$. In theory, we can detect all possible Granger-causal links in this procedure, both direct and indirect. In **Figure 4.1_A**, $A_0(T)=\{T_1, T_2, T_3, X_1, \ldots, X_n , Y_1, \ldots, Y_n\}$.

Secondly, we identify whether the links detected in step 1 are direct or indirect. For such a purpose, we carry out the following iterative procedures.

(Ⅰ) For each node in $A_0(T)$, compute the partial Granger causalities conditioned on all other single nodes in the $A_0(T)$. If the relationship vanishes, delete this node from the initial network and obtain the 1-stage network. After this procedure, all indirect links conditioned on one single node have been removed. In **Figure 4.1_A**, $\{X_1, \ldots, X_n\}$ are deleted from $A_0(T)$, denoting the remaining set as $A_1(T) = \{T_1, T_2, T_3, Y_1, \ldots, Y_n\}$. This is proved in **Lemma 1** of Discussion section.

(Ⅱ) For each node in $A_1(T)$, compute the partial Granger causalities conditioned on all possible pairs in $A_1(T)$. We obtain the 2-stage network in which all indirect links conditioned on a pair of nodes have been removed. In **Figure 4.1_B**, $\{Y_1, \ldots, Y_n\}$ is further deleted from $A_1(T)$, denoting the remaining set as $A_2(T) = \{T_1, T_2, T_3\}$.

(Ⅲ) Continue the procedure above until we can not remove any nodes from $A_k(T)$. The effect of choosing a different parameter k will be discussed in **Chapter 4.4**

The rationale is as follows: if the usual Granger causality from $Y \rightarrow X$ is large but significantly decreases to 0 when conditioned on a third signal Z ($F_{Y \rightarrow X|Z} = 0$), then the connection $Y \rightarrow X$ is only indirect and should be discarded. We use this principle to find the direct ancestors (signals acting on a target X

with no intermediate) of each nodes. At step 0, we search for all signals Y such that $F_{Y\to X}$ is large (a threshold is chosen, e.g. significantly large than 0). We call A_0 this collection of candidate ancestors. At step 1, we filter this set further by keeping the signals $Y \in A_0$ such that $F_{Y\to X|Z}$ is still large for all $Z \in A_0$. We call A_1 this new set and carry on the procedure by conditioning on groups of 2, then 3 etc. signals from the previous set until such an operation is not possible (the size of A_i decreases or stabilizes at each iteration). The result is a list of direct ancestors for each node, which we aggregate to produce the global network.

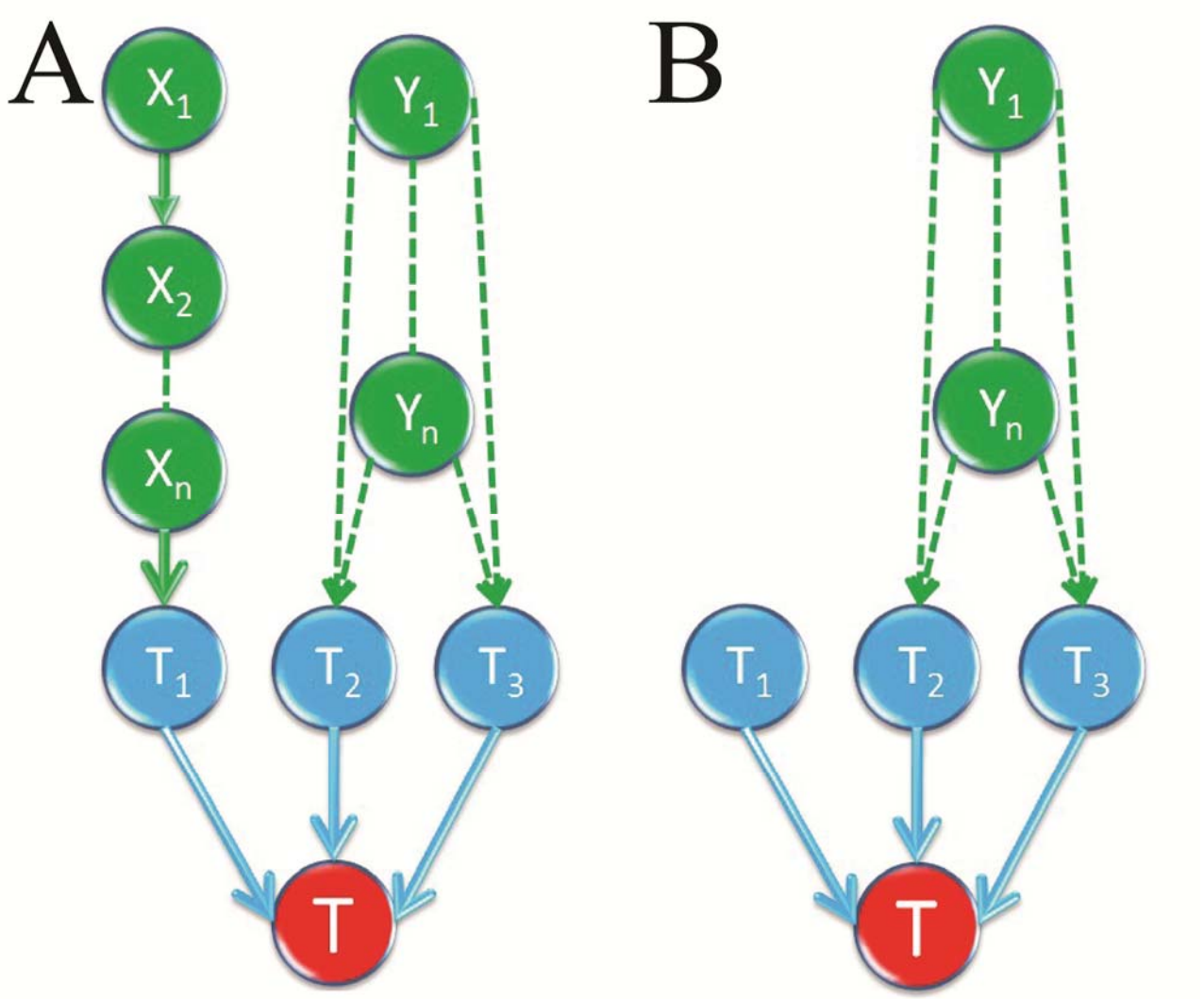


Figure 4.1 Global Granger causality approach. (**A**) Ancestors of target node T, $A_0(T)$ = {T_1, T_2, T_3, X_1, …, X_n, Y_1, …, Y_n }. T_1,T_2,T_3 are direct ancestors to target T. {X_1, …, X_n} connect to T through a single pathway, thus, {X_1, …, X_n} are not direct ancestors to target T. {Y_1, …, Y_n} connect to T through two distinctive pathways (**B**) {X_1, …, X_n} can be removed by Granger-conditioning on a single node, $A_1(T)$={ T_1, T_2, T_3, Y_1, …, Y_n}.

4.3 Results

4.3.1 Local Network: Synthesized Data

The results for the conditional Granger causality approach in time domain has been given in the previous chapter. To illustrate the conditional Granger causality approach in frequency domains, a simple multivariate model with fixed coefficients which has been discussed in previous chapter is tested first. Suppose we have 5 simultaneously recorded time series generated according to the Equations (3.3.1). We applied the conditional Granger causality approach on frequency domain as shown in **Figure 4.2**. The causal relationships from X_1 to X_2, X_3 and X_4 show strong interactions at around 25 Hz.

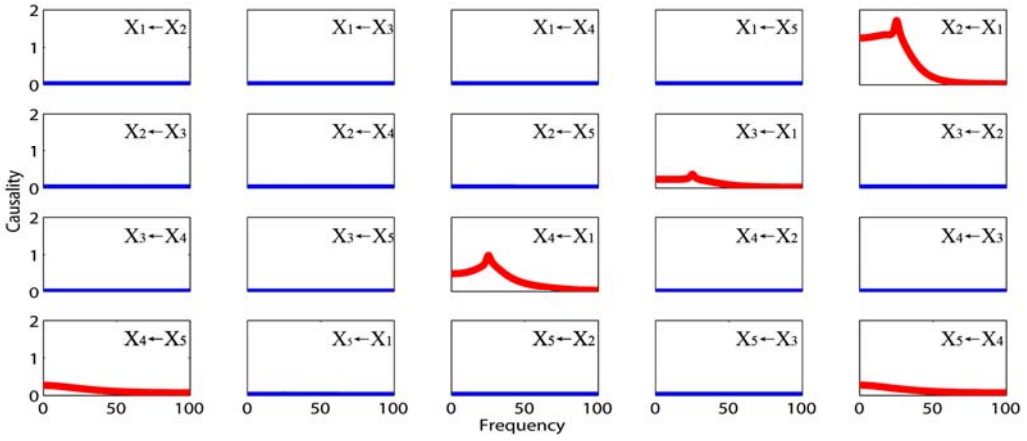


Figure 4.2 Conditional Granger causality approach applied on a simple linear toy model in frequency domain. The red line indicated the significant Granger causality derived by our method.

4.3.2 Local Network: A yeast synthetic network of five genes

A recent *Cell* paper [Cantone 2009] assessed systems biology approaches for reverse-engineering and modeling (see also [Camacho, 2009]). To recover a regulatory interaction network, the authors used three well-established reverse-engineering approaches: ordinary differential equations (ODEs), Bayesian networks and information theory. A gene synthetic network in the yeast consisting of 5 genes with 8 known interactions was investigated. From the results, the authors found ODEs and Bayesian networks could correctly infer most regulatory interactions from the experimental data with best values of PPV=0.75 [Positive Predictive Value = TP/(TP+FP)] and Se=0.5 [Sensitivity = TP/(TP+FN)]. In order to validate our approach, we applied conditional Granger causality (as described in **Chapter 2.1.2**) to the same experimental data. From our results, we found that the conditional Granger causality approach could also correctly infer most regulatory interactions and outperformed all the other three approaches reported in [Cantone, 2009] with the best values of PPV=0.83 and Se=0.83. Hence the Granger causality approach, although simple, can be successfully applied to recover the network structure from temporal data and it could play a significant role in systems biology.

Initially, we applied conditional Granger causality to the switch-off time series (culturing cells in galactose) which contained more time points than

switch-on time series (culturing cells in glucose). The switch-off experiment data consisted of 4 replicates. Since a shift from galactose-raffinose- to glucose-containing medium caused a large initial decay, we simply removed the first two time points for 2 replicates. The time series were not stationary as shown in the [Cantone, 2009]. The gene expression level decreased with time because of the inhibition effect of galactose-raffinose-containing medium. In order to apply conditional Granger causality, we were required to use ARIMA (Auto-Regressive Integrated Moving Average) rather than ARMA model to fit the non-stationary data. The level of difference for ARIMA was chosen to be 1 to avoid losing too many data points.

Firstly, we used the conditional Granger causality approach to infer regulatory interactions for 5 genes. By using the bootstrapping method, we constructed 95% confidence intervals as shown in **Figure 4.3_C**. From this figure, we then constructed the causal network, which is displayed in **Figure 4.3_A**. Only the 5 most significant edges are shown in this graph. From this causal network, there are 4 true-positive edges and 1 false-positive edge. Our approach performs better: the PPV is 0.8, instead of 0.6 and the Se is 0.5, instead of 0.38.

We then grouped Gal4 and Gal80 as a single node as they form a complex [Cantone, 2009], and then applied conditional Granger causality approach. **Figure 4.3_D**. shows 95% confidence intervals for the causality. From this fig-

ure, we can then recover a simplified causal network as shown in **Figure 4.3_B**. It shows the 6 most significant edges. There are 4 true-positive edges and 1 false-positive edge. By comparing our PPV (0.83) and Se (0.83) values with the original paper (PPV=0.75, Se=0.5), it is further confirmed that the performance of our algorithm is much better. The reason why Granger causality outperforms the other approaches is clear from the detailed analysis in the previous chapter where we have reported that the Granger causality is sensitive to detect weak interactions (this experimental data recorded 80 time points). The Bayesian approach is similar to the ODE approach as claimed in previous chapter as well. Hence we could reasonably expect that the Granger approach is the best among the four approaches.

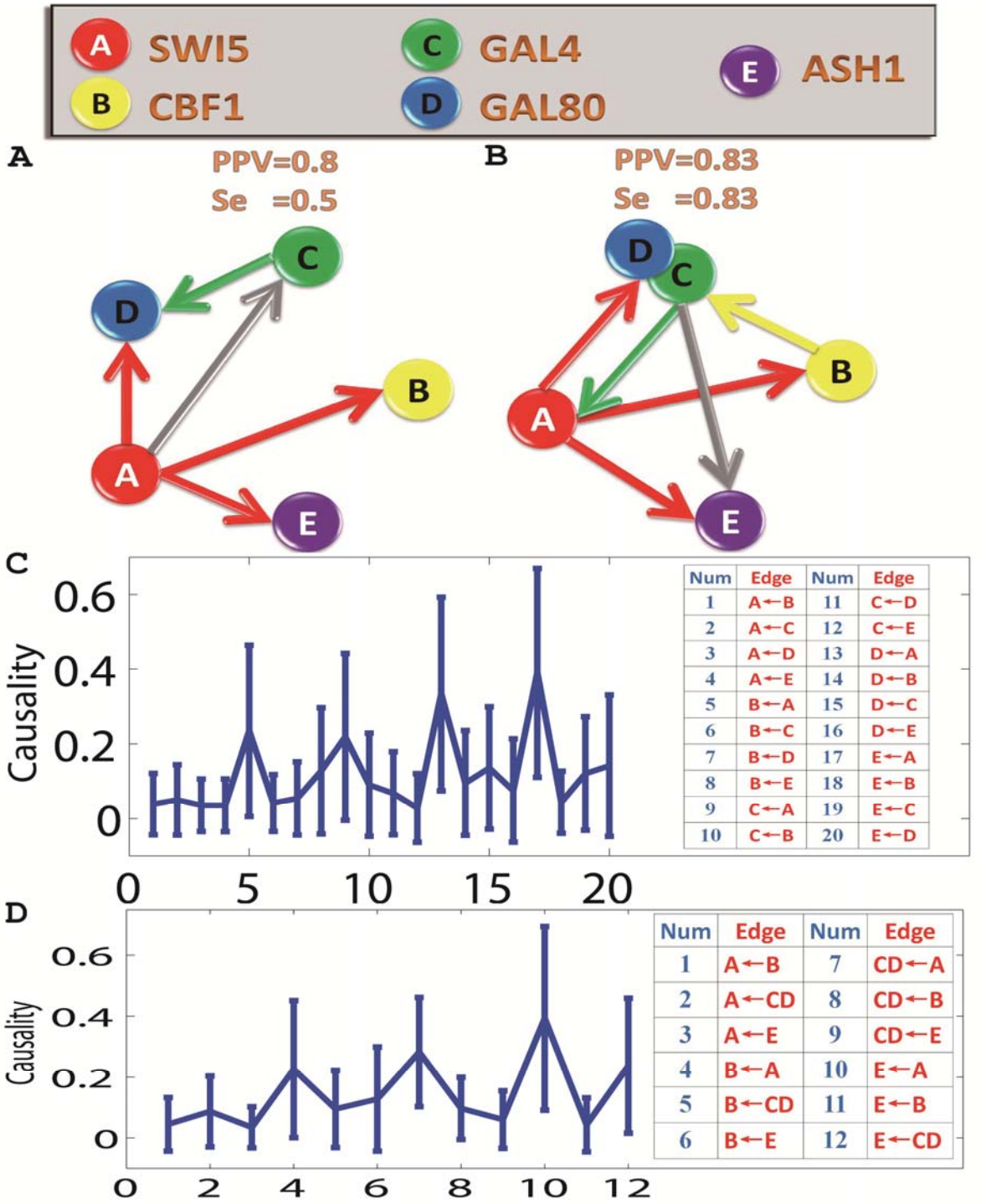


Figure 4.3 Conditional Granger causality approach applied to experimental gene data. The experiment measured the expression level of 5 genes after a shift from galactose-raffinose- to glucose-containing medium. The regulatory network was re-constructed. Solid gray lines represent inferred interactions that are not present in the real network, or that have the wrong direction (FP false positive). PPV [Positive Predictive Value = TP/(TP+FP)] and Se [Sensitivity = TP/(TP+FN)] values show the performance of the algorithm for an unsigned directed graph. TP, true positive; FN, false negative. (**A**) The network structure of 5 genes derived by conditional Granger causality. (**B**) Gal4 and Gal80 were grouped as a single node, so that only transcriptional regulation interactions are represented. (**C**) The 95% confidence intervals of conditional Granger causality results for 5 genes. (**D**) Conditional Granger causality results for a grouped genes (Gal4 and Gal80 are grouped). The 95% confidence intervals graph is plotted.

4.3.3 Local Network: A Local Circuit of Seven Proteins

After testing our approach in the gene circuit, we applied conditional Granger causality approach on dynamic proteomics of individual cancer cells in response to a drug treatment [Cohen 2008; Sigal, 2006]. In the experiment, an anticancer drug, camptothecin (CPT), with a well-characterized target and mechanism of action was used to affect the cell state. The drug is a topoisomerase-1 (TOP1) poison with no other target, which can eventually cause cell death. To follow the response to the drug, 812 different proteins in individual living cells were measured with a time interval of 20 minutes. A total number of 141 sample points (more than 40 hours) were collected. This dataset, much larger than the gene data reported above, gives us the opportunity to construct both local and global networks. In [Cohen, 2008], seven proteins were investigated in more details, including two proteins (DDX5 and RFC1) that were reported to be essential. **Figure 4.4_A** shows the time traces of the seven proteins, denoted as X. They clearly are not stationary, a property that is required for Granger Causality. To overcome this, the model used to fit the time series is changed from ARMA (Autoregressive moving average model) to ARIMA (Autoregressive integrated moving average). Crucially, this transformation does not impact on the true connections between elements. **Figure 4.4_B** shows the transformed data, obtained after differencing the original data term by term 3

times. The reason to choose order 3 is because the time series is significantly decrease with time, a large order 3 (comparing to order 1 used in the *yeast* data) is applied here to make the time series stationary.

Figure 4.5_C shows the Granger causality found for all possible pairs of proteins, together with their 95% confidence intervals calculated though a bootstrap. From the figure, we can then construct the causal protein-interaction network, which is displayed in **Figure 4.5_A**. Only the 12 most significant edges (according to the strength ranking derived by Granger causality) are shown in this graph. In the literature, it has been reported that the protein DDX5 was significantly correlated with the cell fate (with a p-value $p < 10^{-13}$). It has been further proved that it plays a functional role in the response to the drug: a doubling in the death rate was observed during the first 40 hours when DDX5 was knocked-down [Cohen, 2008]. Protein RFC1 also showed a significant correlation with cell fate (with a p-value $p < 10^{-6}$). Our derived network is in good agreement with these two biological characteristics. Protein DDX5, which is the most significantly correlated with the cell fate, is on the top level of the network. Protein RFC1 is in a lower level comparing to DDX5, since the causal relation is from DDX5 to RFC1. Therefore, the results on the proteomic data and gene data confirm that Granger causality is a simple and accurate algorithm to recover the network structure.

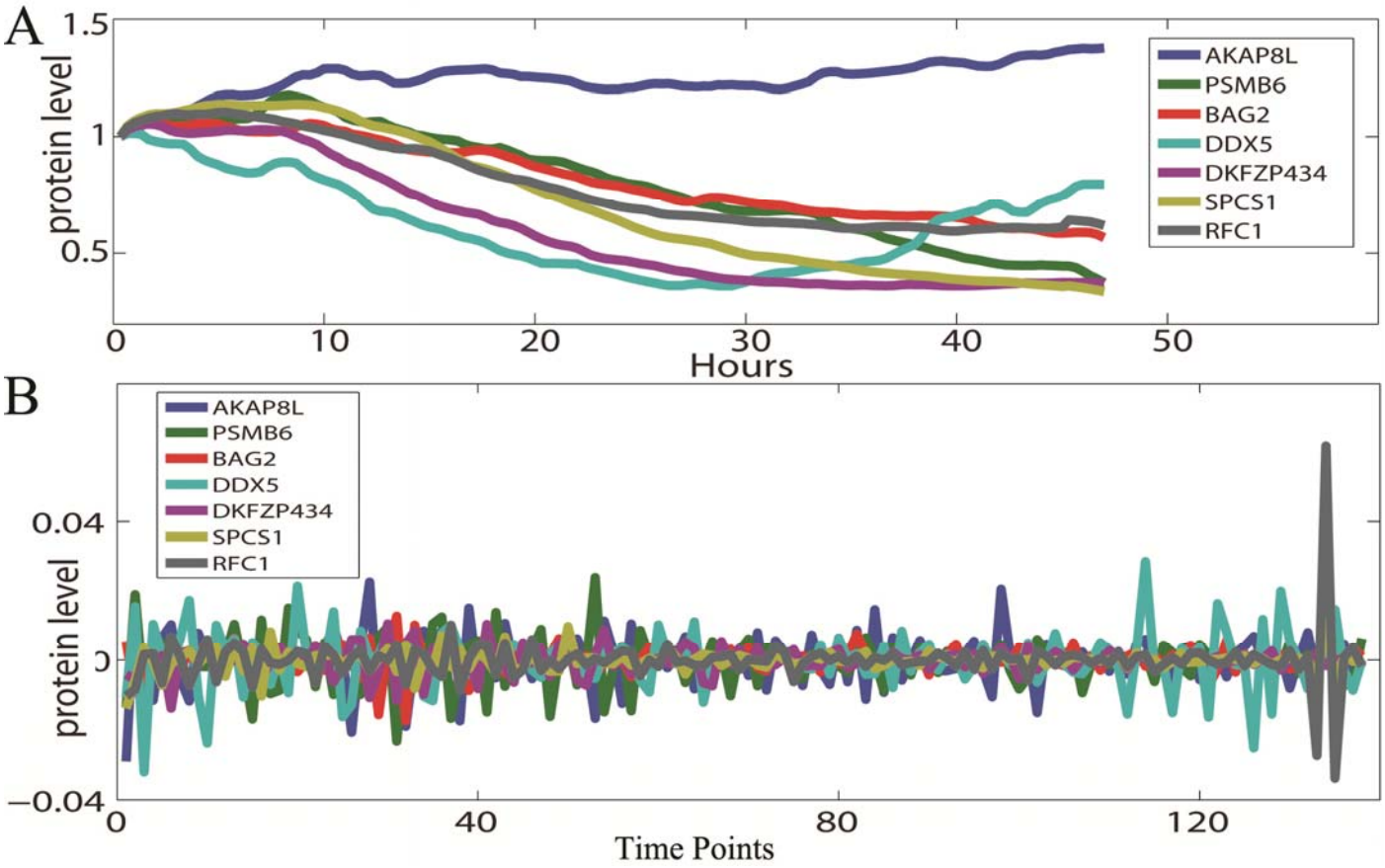
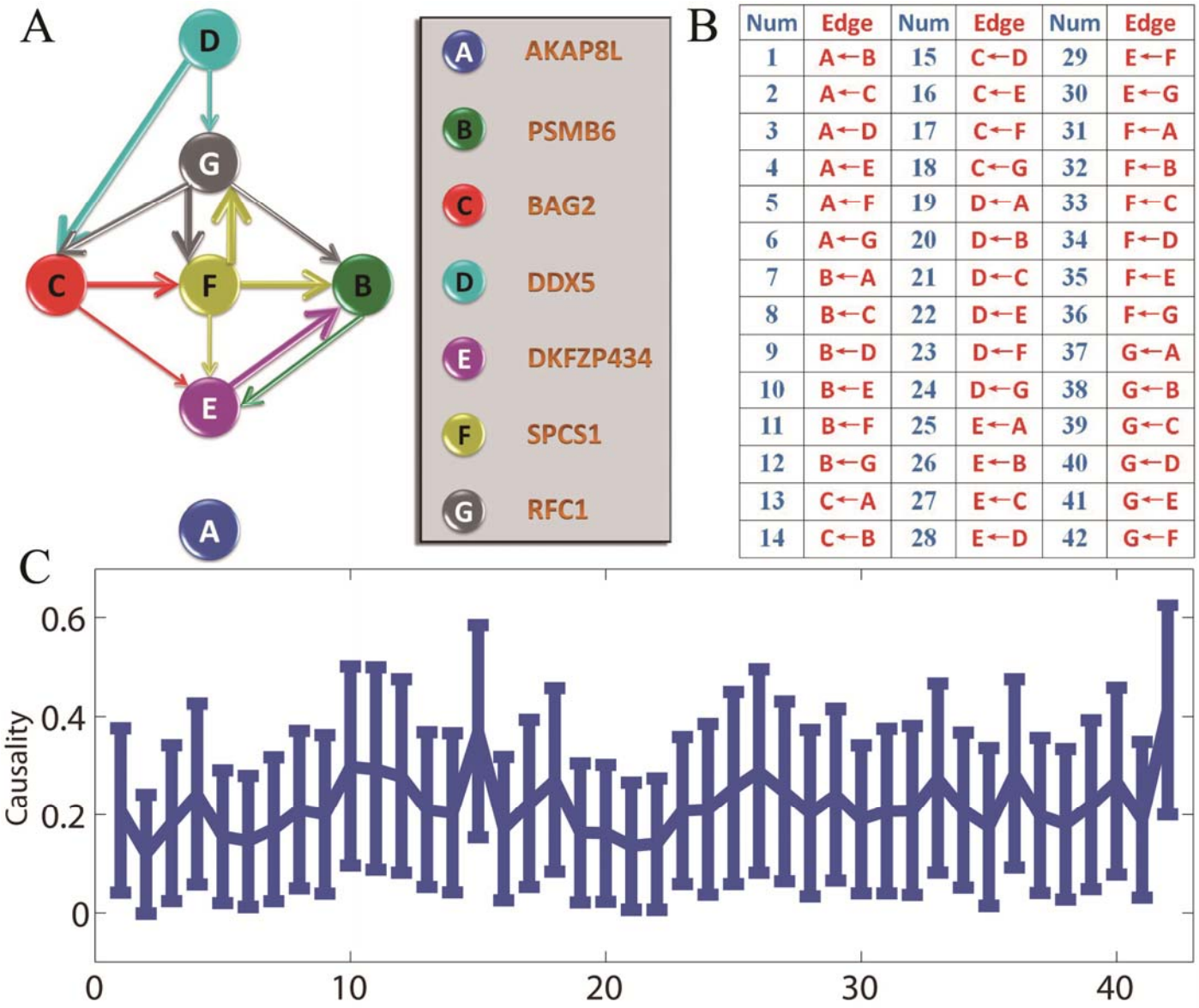


Figure 4.4 Conditional Granger causality approach applied on experimental protein data by using ARIMA model. The experiment measured the levels of 7 endogenously tagged proteins in individual living cells in response to a drug. (**A**) The time traces of 7 proteins are plotted. There are 141 time points. The time interval is 20 minutes. (**B**) ARIMA model is used to fit the data. We applied term-by-term differencing 3 times to the data.



Num	Edge	Num	Edge	Num	Edge
1	A←B	15	C←D	29	E←F
2	A←C	16	C←E	30	E←G
3	A←D	17	C←F	31	F←A
4	A←E	18	C←G	32	F←B
5	A←F	19	D←A	33	F←C
6	A←G	20	D←B	34	F←D
7	B←A	21	D←C	35	F←E
8	B←C	22	D←E	36	F←G
9	B←D	23	D←F	37	G←A
10	B←E	24	D←G	38	G←B
11	B←F	25	E←A	39	G←C
12	B←G	26	E←B	40	G←D
13	C←A	27	E←C	41	G←E
14	C←B	28	E←D	42	G←F

Figure 4.5 Conditional Granger causality results derived by using experimental protein data. (**A**) The network structure for 7 proteins derived by using conditional Granger causality approach. (**B**) For visualization purpose, all directed edges (causalities) are sorted and enumerated into the table. (**C**) Conditional Granger causality results. The 95% confidence intervals graph, which is constructed by using method bootstrap, is plotted.

In **Figure 4.6**, it shows the same analysis in the frequency domain. From the result, we find that there are strong interactions from D (DDX5) to C (BAG2) at around 0.006 cycle/min or one cycle every three hours. From the power spectrum result for D and C, we can also find an energy peak at this frequency. In addition, there is a strong chain interaction from D to G (RFC1) via C and F (SPCS1). This chain contains the 3 strongest interactions. Each element in the chain affects its downstream element at a similar frequency.

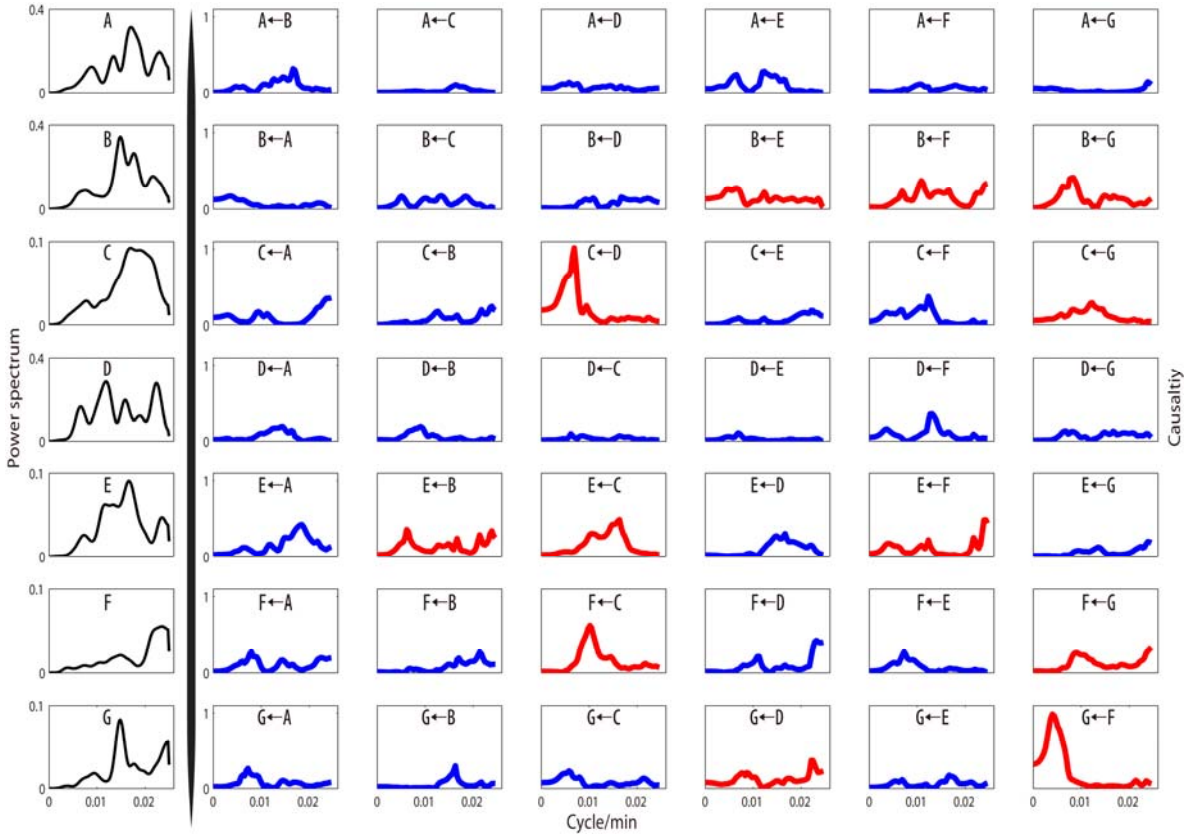


Figure 4.6 Conditional Granger causality analysis in frequency domain for experimental data. Conditional Granger causality was applied to experimental data in the frequency domain and power spectrum density analysis for 7 proteins (the most left column in black line). The significant causalities are shown in red lines in the figure.

4.3.4 Global Network: Synthesized Data

To measure the performance of the Global Granger Causality (GGC) algorithm introduced in this thesis, we first consider some toy models. The first toy model is a high-dimensional time series. We also compare the result of GGC with that of Partial Granger Causality (PGC).

Example 1 Suppose that 12 simultaneously generated time series were generated by the equations (The coefficients are randomly generated):

$$\begin{cases} X_1(t) = 0.95\sqrt{2}X_1(t-1) - 0.9025X_1(t-2) + \omega_1(t) \\ X_2(t) = -0.5X_1(t-2) + \omega_2(t) \\ X_3(t) = 0.8X_1(t-4) - 0.5X_2(t-2) + \omega_3(t) \\ X_4(t) = -1.2X_3(t-1) + 0.25\sqrt{2}X_4(t-1) \\ \qquad + 0.65\sqrt{2}X_1(t-1) + \omega_4(t) \\ X_5(t) = -0.25\sqrt{2}X_4(t-1) + 0.5\sqrt{6}X_6(t-1) + \omega_5(t) \\ X_6(t) = -0.6\sqrt{2}X_4(t-2) + 0.8X_5(t-1) + \omega_6(t) \\ X_7(t) = 0.8X_5(t-3) + 0.7X_6(t-1) + 0.8X_{10}(t-2) \\ \qquad + 0.7X_1(t-2) + \omega_7(t) \\ X_8(t) = 0.85X_6(t-2) + \omega_8(t) \\ X_9(t) = 1.15X_1(t-1) - 0.9025X_6(t-2) \\ \qquad + 0.7X_7(t-2) + \omega_9(t) \\ X_{10}(t) = 0.5X_7(t-2) + \omega_{10} \\ X_{11}(t) = 0.8X_6(t-2) + 0.6X_9(t-3) + \omega_{11}(t) \\ X_{12}(t) = -0.5X_7(t-3) + \omega_{12}(t) \end{cases} \tag{4.3.1}$$

where $\omega_1,\cdots,\omega_{12}$ are zero-mean uncorrelated process with identical variance. We generated time series of 80 points. The true network structure is depicted in **Figure 4.7_A**, there are 21 actual links. We first applied PGC to the data directly and used a bootstrap method to construct the network structure. More specifically, we simulated the fitted VAR model to generate a dataset of 1000 realizations of 80 time point, and used 3σ (around 99% confidence intervals by using same bootstrapping method described in **Chapter 3.3**) as the confidence interval. If the lower limit of the confidence interval was greater than zero, we considered there was a relationship between two units. The network structure is depicted in **Figure 4.7_B**. The network structure we obtained from PGC was misleading. The reason is that since the order of the model is 4, the number of total parameters we should estimate in this model is $12\times12\times4$, the estimator is unreliable with such little data.

Secondly, we used GGC to investigate the network structure. **Figure 4.7C** shows the results we obtained after applying pairwise Granger causality. There are 33 links in total. We computed partial Granger causality conditioned on any intermediate node to identify whether the links appearing in **Figure 4.7_C** are direct or indirect. If the lower limit of the confidence interval of partial Granger causality is less than zero, then the link is regarded to be indirect and is deleted from **Figure 4.7_C** (dashed arrows). **Figure 4.7D** is the final

structure we get from GGC; it is consistent with the actual structure **Figure 4.7_A**.

Example 2: Random network

Next we present a validation of our method with a series of experiments on random networks for which the true structure is known. We built an Erdös-Rényi random graph [Bollobás, 2001] with N = 200 nodes and $M = N \times \log(N) = 1060$ edges. From the network structure, we generated N time series with an auto-regressive model of order 1 whose transition matrix was the transpose of the adjacency matrix of the network, with its largest eigenvalue normalized to 0.99 to obtain a stable system. Each time series was 200 time-points long and normal noise of unit variance was added throughout. The algorithm was applied to each single node to get a list of their guessed ancestors. We then compared the true network with our derived ancestors. One should expect that the connection between two nodes is difficult to uncover if the corresponding coefficient in the linear model is small. To factor this out, we first considered the case where the non-zero coefficients of the transition matrix were all equal and maximized (**Constant coefficients**). We then applied the method on the case where the non-zero coefficients were randomly distributed (**Random coefficients**).

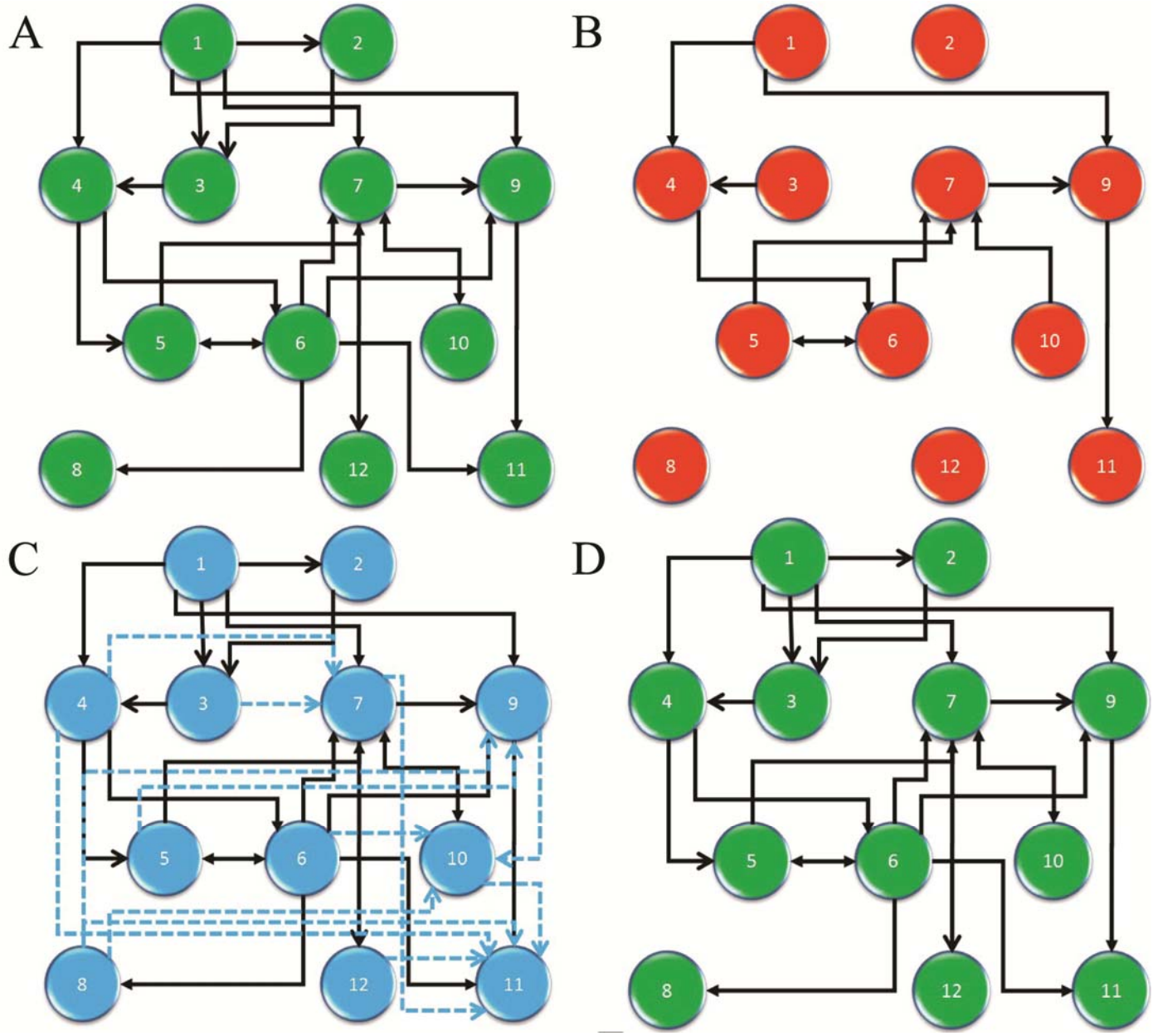


Figure 4.7 Global Granger Causality (GGC) algorithm applied on a simple toy model. (**A**) The actual network structure used in toy model of global network. (**B**) Network structure inferred from PGC. (**C**) Network structure inferred from pair-wise Granger causality (solid and dashed links). By using partial Granger causality among three units, we can delete some of them (dashed links). (**D**) The final network structure from GGC, it is consistent with the actual relationship.

Random network with constant coefficients

The data were generated by an auto-regressive model described above with transition matrix A (200×200). A is a scaled version of the transpose of the true adjacency matrix. The scaling factor was chosen so as to be maximal while leading to a largest eigenvalue for A of less than 1 (or the model degenerates). In this particular case, it was found to be 0.1736 (the corresponding coefficients for direct links and coefficients = 0 for disconnected nodes). The procedure has one parameter τ, the threshold at which a Granger-causality is deemed significant. By varying this parameter from 0 to 0.1, we obtained different large networks which we compared to the truth. The accuracy of each network was summarized by its true positive and false positive rates. **Figure 4.8_A** shows the resulting receiver operating characteristic (ROC) curve that is the graph obtained by plotting the false positive rate against the true positive rate. The performance of the method was extremely good, with an area under the curve close to 1. Crucially for biological applications, the false positive rate is always very small.

Random network with random coefficients

In this setup, the non-zero coefficients of the transition matrix were randomly distributed (normally distributed with mean 3 and multiplied by -1 with probability 1/2). The matrix was then scaled in the same manner as before. **Figure 4.8_B** shows the performance of the method on this harder problem. The method is not as accurate as before, with a maximum true positive rate just over 0.5. However, the false positive rate is still very low: the method doesn't guess as many ancestors as before but its guesses are rarely wrong. The fact that more connections are now missed out is not surprising: the non-zeros coefficients are randomly distributed and can be very small. **Figure 4.8_C** shows how the true positive rate varies with the magnitude of the coefficients; the true positive rate goes to zero with small coefficients.

4.3.5 Global Network: A Global Circuit of 812 Proteins

We then applied our GGC approach on the whole dataset of 812 proteins on dynamic proteomics of individual cancer cells in response to a drug treatment [Cohen 2008; Sigal, 2006]. **Figure 4.9_A** shows the direct ancestors of protein DDX5, known to be at the top level of the circuit, as shown in the previous section. Our result suggests that controlling for either BC037836,

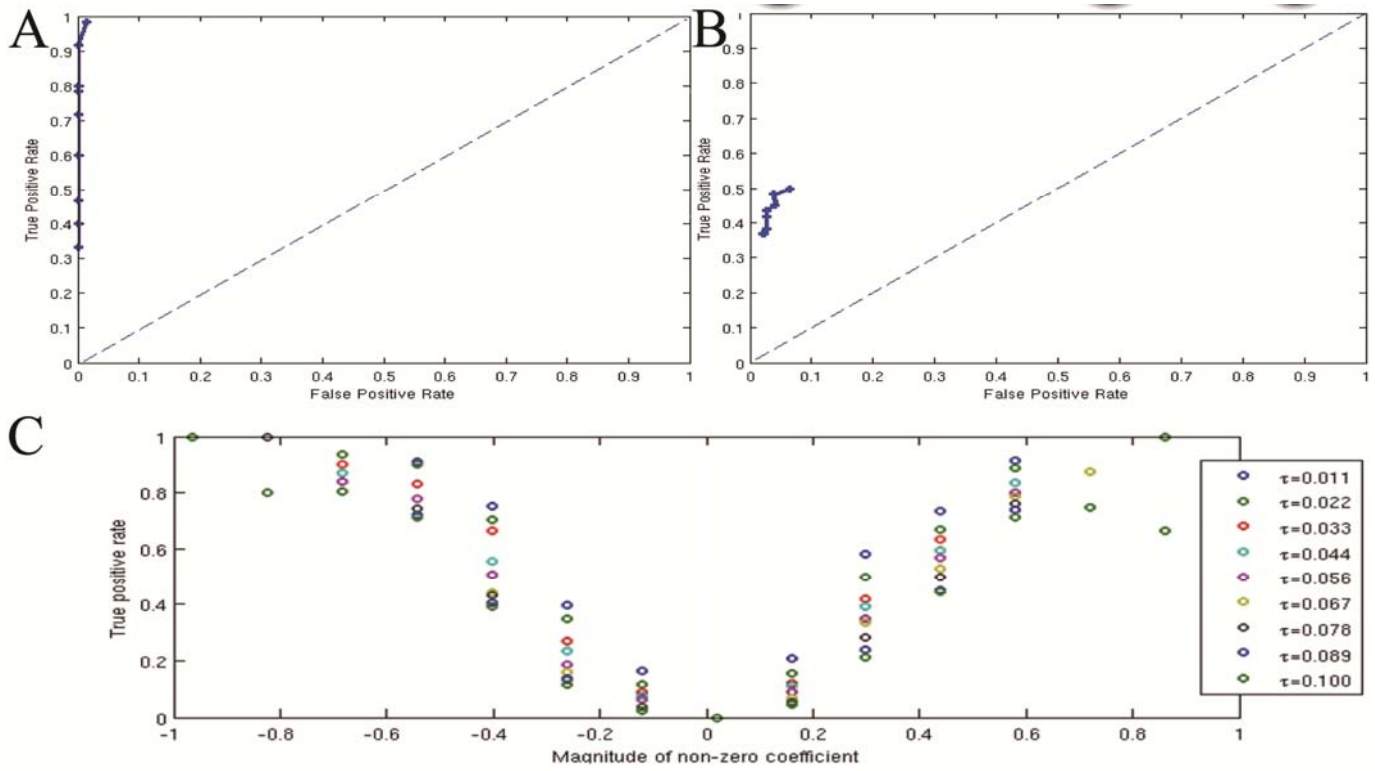


Figure 4.8 The performance analysis for Global Granger Causality (GGC) algorithm applied on toy models for both constant and random coefficients. (**A**) ROC curve summarizing the performance of the procedure on a random network with maximum non-zero coefficients. (**B**) ROC curve summarizing the performance of the algorithm on a random network with random non-zero coefficients. (**C**) True positive rate as a function of the magnitude of the non-zero coefficient.

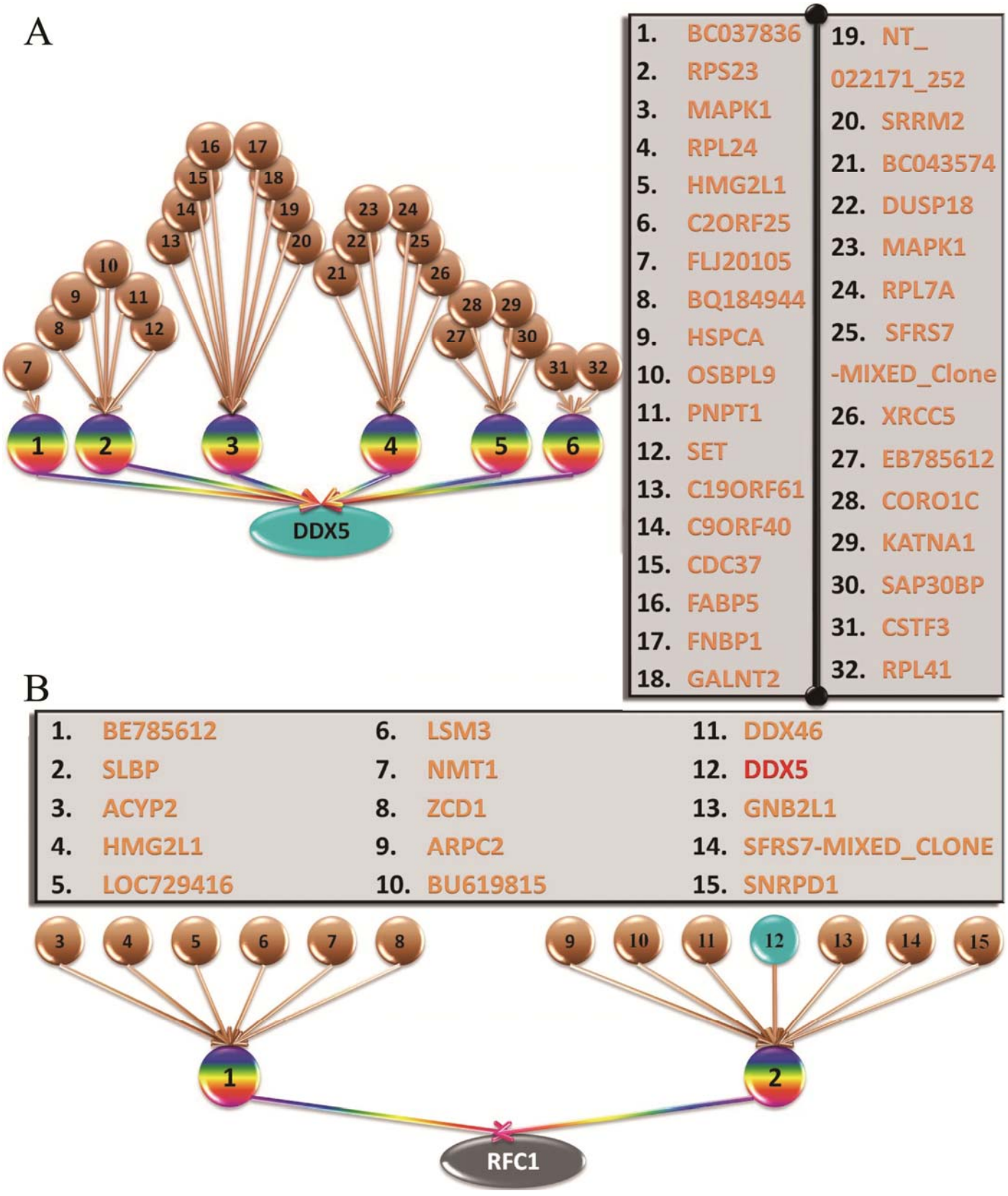


Figure 4.9 Global Granger Causality algorithm applied on experimental data of 812 recorded proteins for global network re-construction. (**A**) Direct ancestors of the protein DDX5: BC037836, C2ORF25, HMG2L1, MAPK1, RPL24 and RPS23. (**B**) Direct ancestors of RFC1, as well as their own direct ancestors. The causal link from DDX5 to RFC1 is now completely identified: an intermediate protein (SLBP) connects them. For visualization purpose, the proteins are enumerated into the table.

C2ORF25, HMG2L1, MAPK1, RPL24 or RPS23 will have an effect on DDX5 and thus on the whole circuit. These ancestors of DDX 5 prediced by GGC can be experimentally tested in the future by using knockout techniques. A similar figure for RFC1 is shown in the **Figure 4.9_B**.

Setting the same threshold (lower bound > 0.1) as the one used to obtain the small circuit, a large, sparse network is obtained: 768 nodes remain (discarding those with no connections) and 2972 edges were found, which represented only 0.5% of all the possible edges. The complete structure is displayed in **Figure 4.10_A**. The overall mean clustering coefficient is shown in **Figure 4.10_B**. It is an order of magnitude larger than the one of a random network (0.022 instead of 1/768=0.0013). But the network is not modular: the mean clustering coefficient with respect to degree is more or less constant. **Figure 4.10_C** shows the distributions of in-, out- and total degree of the nodes. All three distributions are exponential, precluding the possibility of a scale-free network. Each node has an average in-degree and out-degree of 3.8, indicating a well-connected network. This is confirmed by the characteristic path length (average of the shortest path between all pairs of nodes). Considered undirected, the graph has a characteristic path length of 3.8, in line with those of previously reported biological networks (see [Mason, 2007] and references within), including protein-protein interaction networks, although it should be noted that the present study is concerned with the dynamics of the proteins (i.e. [Sachs, 2005])

and not their physical interactions (in which case the network is undirected by construction). The directed graph also has a small characteristic path length of 5.7 nodes and a small diameter (largest shortest path) of 12 nodes. Such connectedness indicates that the network is a small world [Kleinberg, 2000; Wang, 2003]. However, it is not particularly modular: while its mean clustering coefficient is an order of magnitude (17 times) higher than one of a random network, the clustering coefficient is almost constant with respect to the node's degree. In other words, the same level of clustering is found everywhere regardless of the node's degree.

The previous small network in **Figure 4.5** was obtained by using the conditional Granger causality. It can be misled by common influence: if both nodes are subjected to an unknown common source, it can have an effect on their connections. Partial Granger causality – another extension of Granger causality [Guo, 2008; Ladroue, 2009;] – can address this issue by considering an unseen external input in the linear model and working out its effect on the connection. For example, the partial Granger causality between DDX5 and RFC1 is very small, even though the conditional Granger causality between them is high (**Figure 4.5**) and there exists a short path (1 intermediate) from DDX5 to RFC1 in the large network. This suggests the connection is affected by a common unseen source.

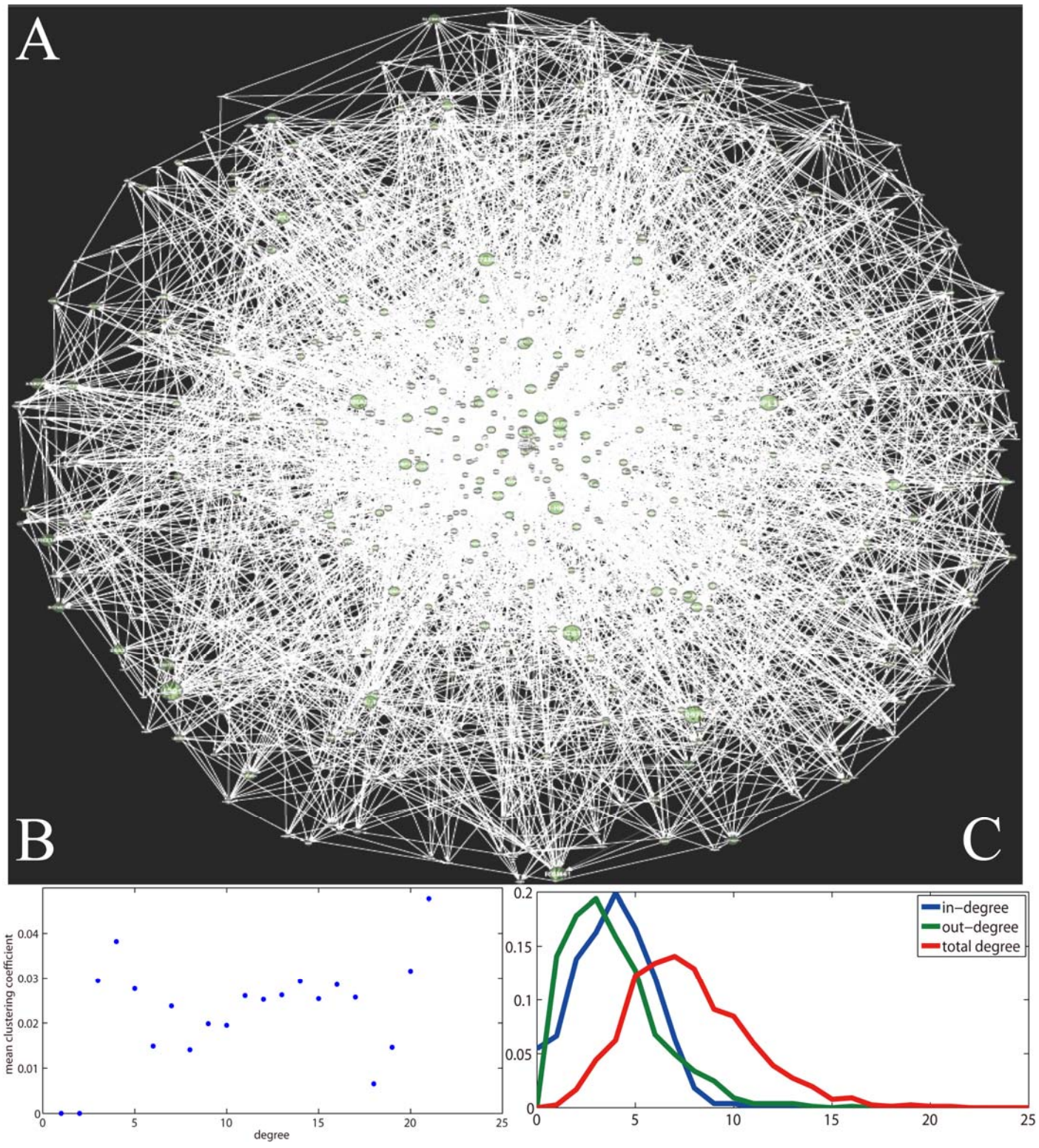


Figure 4.10 Global Granger Causality algorithm applied on experimental data for global network reconstruction. (**A**) The whole re-contructed network of 812 proteins. (**B**) The overall mean clustering coefficient (the probability of neighbours being inter-connected) (**C**) In-, out- and total degree distributions of the large network calculated from the whole dataset.

In order to identify which proteins have an influence on the connections between the 7 proteins of interest (AKAP8L, PSMB6, BAG2, DDX5, DKFZP434, SPCS1, and RFC1), we first extracted them as well as the proteins belonging to the shortest paths between them, resulting in a subset of 118 proteins. We then applied a filtering process on each of the 12 connections uncovered in the previous section. The rationale of the algorithm is that if removing the (explicit) influence of a protein makes the connection between two nodes change, then this protein should be kept as a potential influence on the connections – if Z is independent of X and Y, then z does not affect the Granger causality and $F_{X \to Y|Z} = F_{X \to Y}$. After filtering for those that have an influence, we then considered their pairs and build a new subset, then triplets etc.. The end-result is a set of proteins which have a substantial influence on the connection.

Figure 4.11 shows the small network of 7 proteins with the now-identified external influences. Note that those proteins do not necessarily belong to the path from one node to the other, but rather they have some substantial influence on the connection as a whole, for example on some members of the path.

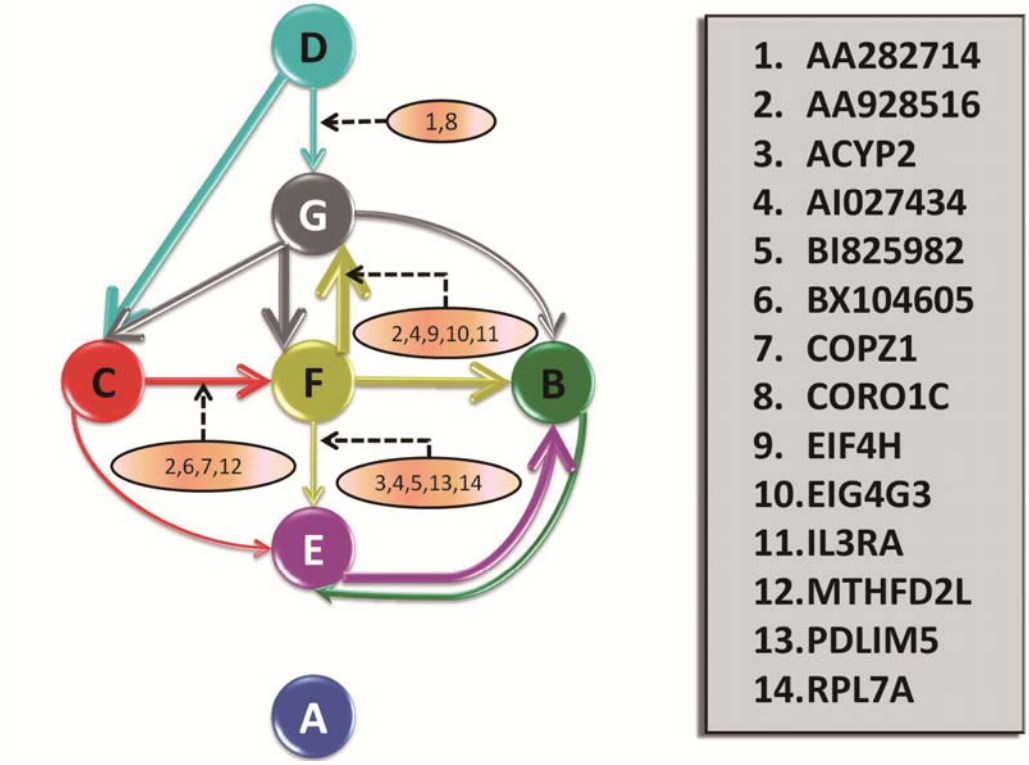


Figure 4.11 External influences identified by the second iterative procedure, in ovals. For visualization purpose, the proteins are enumerated into the table. The proteins A-G are enumerated into the table of **Figure 4.5_A**.

4.4 Discussion

How reliable is Global Granger Causality?

In theory, we can recover all possible links from the pairwise Granger causality procedure and have to apply conditional Granger causality on all combinations of the nodes in the system to remove an indirect connection. However, it is an NP-hard problem and we will stop at a stage k, i.e., we only need to apply conditional Granger causality on the combinations of up to k nodes. Therefore, the analysis on how to choose k and the probability of correctly uncovering the true relationship of the whole network when we stop at stage k is of vital importance. In this section, we will provide some simulation and theoretic results on these questions.

Consider a network with N nodes $\{X_1,\cdots,X_N\}$ with a connection probability p. There are $N\times(N-1)\times p$ direct links on average in the whole system. We intend to estimate how many indirect connections are left when we stop at stage k. Here we focus on a pair X to T, where X, T are in $\{X_1,\cdots,X_N\}$. If there exist only one single path from X to T, this link can be discarded by applying conditional Granger causality on a single intermediate node in the path. If there are more than one path from X to T, in theory, this link should be discarded by Granger-conditioning on all the other nodes.

***Definition 1 (bottleneck)**. Assume that there are m distinctive directed paths from $S \in \{X_1, \cdots, X_n\}$ to T and p(S,T) be the set of all distinctive directed paths from S to T. A set of nodes $\{Z_1, \cdots, Z_m\}$ is called a section from S to T if there is no directed path from S to T in the graph $\{X_1, \cdots, X_N\} - \{Z_1, \cdots, Z_m\}$. A section which minimizes its total number of elements of the section is called a bottleneck.*

For example, in **Figure 4.12_A** both $\{B_1, B_2\}$ and $\{B_3\}$ are sections from S to T, but $\{B_3\}$ is the bottleneck..

***Lemma 1.** Assume that the set $\{B_1, \cdots, B_m\}$ is the bottleneck from S to T, we have*

$$F_{S \to T | \{B_1 \cdots B_m\}} = 0 \tag{4.4.1}$$

Proof. We only check two cases here. The first case is that there is a single serial connection from S to T. For example, we have $S \to B_1 \to B_2 \to \cdots B_n \to T$ where every single node $\{B_i\}$ is a bottleneck of the path. If we condition on one of the single node B_i in the path, we need to show

$$F_{S \to T | \{B_i\}} = 0 \tag{4.4.2}$$

According to the definition, we need to find the autoregression expression:

$$T = C(\Gamma)T + D(\Gamma)B_i + \xi \tag{4.4.3}$$

where Γ is the delay operator and C, D are polynomials, ξ is the noise term. From the assumption of the path structure, we conclude

$$\begin{cases} y_1 = C_1(\Gamma)y_1 + D_1(\Gamma)X + \xi_1 \\ y_2 = C_2(\Gamma)y_2 + D_2(\Gamma)y_1 + \xi_2 \\ \vdots \\ y_{i+1} = C_{i+1}(\Gamma)y_{i+1} + D_i(\Gamma)y_i + \xi_{i+1} \\ \vdots \\ T = C_{n+1}(\Gamma)T + D_{n+1}(\Gamma)y_n + \xi_{n+1} \end{cases} \tag{4.4.4}$$

Therefore

$$\begin{aligned} T &= C_{n+1}(\Gamma)T + D_{n+1}(\Gamma)y_n + \xi_{n+1} \\ &= C_{n+1}(\Gamma)T + E(\Gamma)y_{i+1} + F(\Gamma)y_i + \xi' \\ &= C_{n+1}(\Gamma)T + G(\Gamma)y_i + \xi'' \end{aligned} \tag{4.4.5}$$

where E,F,G are polynomials, and ξ', ξ'' are system noises. From the equation above, we see that for any node B_i, we have $F_{S\to T|\{B_i\}} = 0$. Intuitively, in a serial path $S \to B_1 \to B_2 \to \cdots B_n \to T$, the information cannot be transmitted from S to T if B_i is removed. In conclusion, for a single path, the Granger causality is zero whenever we condition on one of its nodes in the path. It is not necessary to condition on the whole path to remove the causality.

The second case is as depicted in **Figure 4.12_A**. There are two different paths from S to T, B_1 and B_2 converge to a common bottleneck B_3. It is easy to see that information can not be transmitted from S to T if B_3 is removed, then we can easily see that

$$F_{S\to T|\{B_3\}} = 0 \tag{4.4.6}$$

Combining the above two cases completes the proof of the lemma.

Lemma 1 tells us that if there are m distinctive paths from S to T, i.e., the number of the bottleneck is m, then the causality between S and T will vanish when we take into account the partial Granger causality on $\{X_1,...,X_m\}$. There may exist other common drives to the observed nodes S and T such as **Figure 4.12_B**. We assume the partial Granger causality can delete the influence of such drive and exclude such case in our analysis.

The exact formula of the number of bottlenecks seems to be fairly complicated but we can have a first look at the empirical distribution of it. For a variety of connection probability p, we generate 500 random networks when N=100. For each network, we randomly select two nodes and compute the number of the bottleneck between them. **Figure 4.12_C** shows the histograms when p=0.015, 0.02, 0.03 and 0.05, respectively. From these figures, it can be easily seen that the sparser the network is, the quicker we can detect the true

structure from global Granger causality. When p=0.015, it is very likely for any two nodes to be unconnected or directly connected, then almost all the true relationships can be uncovered at stage 1. When p=0.02, all the true relationships can be uncovered at stage 2. When p=0.03, the probability of uncovering the true relationship is 90.8% at stage 2 and 98.6% at stage 3. When p=0.05, the probability of uncovering the true relationship is 82.2% at stage 4 and 97.8% at stage 6. It is not until stage 9 that all indirect links can be discarded.

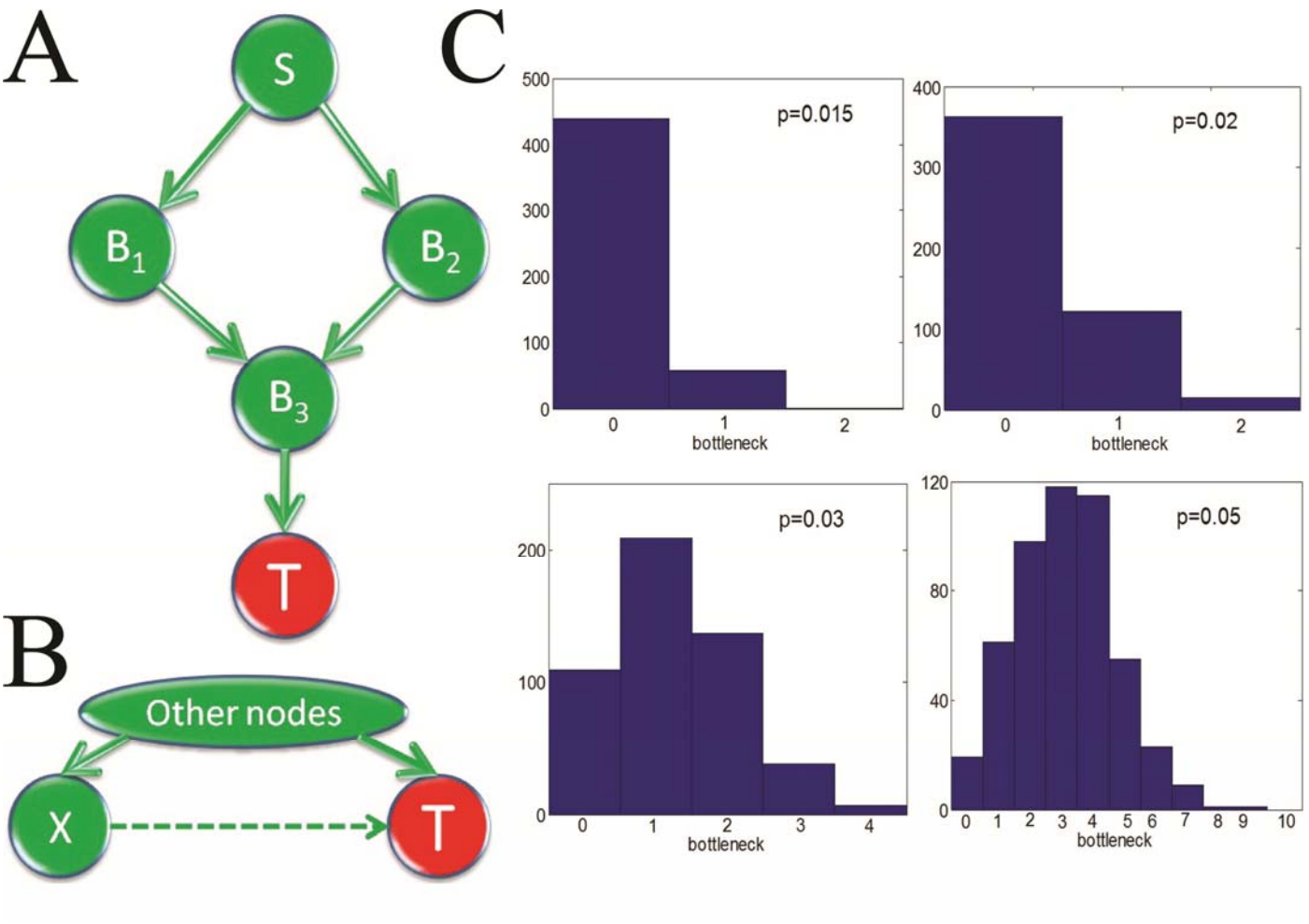


Figure 4.12 reliability of Global Granger causality approach. (**A**) S is connected to T through two different paths, both $\{B_1,B_2\}$ and $\{B_3\}$ are sections from S to T, but $\{B_3\}$ is the bottleneck. (**B**) There may exist other common drives to the observed nodes X and T, we assume the partial Granger causality can delete the influence of such drive and exclude such case in our analysis. (**C**) Histograms of the number of bottleneck for a variety of connection probability p for N=100 and 500 simulations.

4.5 Final Remarks

In this chapter, I focused on the Granger causality approach in both the time and frequency domains in local and global networks. For a local gene circuit, a recent *Cell* paper [Cantone, 2009] assessed systems biology approaches for reverse-engineering and modeling by investigating a gene synthetic network in the yeast consisting of 5 genes with 8 interactions. From our results, we found that our conditional Granger approach could also correctly infer most regulatory interactions and outperform the three approaches reported in the [Cantone, 2009]. For a local protein-interaction network, our derived network is in good agreement with biological characteristics. Therefore, the results on the proteomic data and gene data confirm that the Granger causality is a simple and accurate approach to recover the network structure. For a global network, our novel approach called global Granger causality was successfully used to build a large network from all the recorded 812 proteins. Since Granger causality is a very useful tool to investigate the connectivity among elements, we apply Granger causality to local field potential data and study the connectivity change due to the effect of anaesthesia. Some interesting results for brain connectivity are described in the next chapter.

Chapter 5

Loss of consciousness due to anaesthesia is associated with altered functional cortical connectivity

Loss of consciousness with anaesthesia, sleep or vegetative states may involve reduced functional cortical connectivity in a form of "cognitive unbinding" caused by disruption of long feed forward connections [Mashour,2006; Akire, 2008] and synchronization [Imas, 2005; John, 2005]. In this chapter, we have combined multiarray array recordings of local field potentials (LFP) in right (rIT) and left (lIT) inferotemporal and right anterior cingulate (rAC) cortices in sheep with Granger causality to investigate how anaesthesia alters neural processing during resting state and visual stimulation.

Results from both time and frequency (1-70Hz) domain analyses show that when animals view face images a predominantly unidirectional flow of direct (between ITs) (which was also shown in previous results [Ge 2009; Ladroue 2009]) and indirect (via rAC) causal connections from left to right IT occurs. Under resting state conditions information flow is in the opposite direction. Anaesthesia abolishes direct inter-hemispheric connections and indirect ones either change direction or became bidirectional. Patterns and synchronicity of visual evoked potentials are also altered. By contrast, short-distance causal connections within brain regions actually increase their frequency under anaesthesia. Loss of consciousness during anaesthesia is therefore associated with reductions in extrinsic long-distance open-loop cortico-cortical connections, and loss of their unidirectional flow, coupled with an increase in the strength of shorter-distance intrinsic closed loop connections.

In addition, we have also tested conditional Granger causality for synthesized data of neurodynamical model. The task of this synthesized network was designed to make a decision between two possible alternatives, according to the characteristics of a sensory input. Assuming there was no knowledge of the network structure, we investigated conditional Granger causality by using three types (LFP, firing rate and spiking intervals) of data generated by the model. Our results indicated that conditional Granger causality could correctly

deriving most direct connections and LFP data provided most reliable results comparing to the other two types of data.

5.1 Introduction

Based upon simultaneously multi-site physiological recording data (spikes, firing rates and local field potentials, EEG and MEG etc.), to reliably and accurately explore the directed network structure among different areas, is one of the most import tasks in computational biology [Klipp, 2005; Feng, 2007; Tong, 2004; Guo, 2008]. In this chapter, we applied Granger causality approach to both synthesized and experimental data.

In the previous chapters, we have investigated and compared Granger causality with other reverse-engineering approaches. Our results concluded Granger causality outperformed all other methods in some specific condition (*e.g.* long data length). In computational biology, various types of time series data can be collected in advanced laboratories for the same organisms. A natural question is which type of data is more reliable to use for reverse-engineering task by applying Granger causality approach. Here, we provided an answer by using synthesized data generated by a well-known neural model. We tested conditional Granger causality approach for three types of data (LFP, Firing Rate and Spiking Intervals data) generated by using a neurophysiological model. This model was firstly proposed by [Wang, 2002] and extensively explored by [Deco, 2006]. It is an integrate-and-fire attractor model of the decision-related activity of ventral premotor cortex (VPC) neurons during a vibrotactile fre-

quency comparison task. Our results illustrated that Granger causality could reliably recover most interactions correctly, especially by using the LFP data.

For experimental data, we have combined multiarray array recordings of local field potentials in three different cortices in sheep with Granger causality to investigate how anaesthesia alters neural connectivity during resting state and visual stimulation. In 21st century, one of the most important scientific and philosophical questions is the nature and mechanism of human consciousness. To investigate this question, the mechanisms of general anaesthesia have also been studied together with the mechanism of consciousness. In the literature, it is reported that loss of consciousness with anaesthesia may involve reduced functional cortical connectivity caused by disruption of long feed forward connections [Mashour, 2006; Akire, 2008] and synchronization [Imas, 2005; John, 2005].

In conscious state, our results show that lIT has either directly or indirectly causal influence via the rAC on the rIT with face picture stimuli (see also in [Ladroue, 2009]). It is consistent with the hypothesis that in the left brain hemisphere, which is thought to employ more localised processing in the context of analysing detail and organisation and control of action. By contrast, during resting state conditions the flow of causal connections was reversed from rIT to lIT both directly and indirectly via the rAC. It also showed a very good

agreement with the hypothesis that the right brain hemisphere is thought to play a key role in spatial information processing involving a more global type of processing.

With anaesthesia applied in sheep, it abolished direct causal connection between the hemispheres in both conditions. Numerous causal connections within brain regions were found in both conditions and their frequency was significantly increased by anaesthesia. In addition, patterns and synchronicity of visual evoked potentials were also altered. Thus, our results conclude that loss of consciousness during anaesthesia is therefore associated with a reduction/disruption of long-distance open-loop cortico-cortical connections and a corresponding increase in shorter-distance closed loop ones.

5.2 Methods

The detailed description of conditional Granger causality in time and frequency domain was introduced in Chapter 2. In this section, we introduce the architecture of neurodynamical model for generating the synthesized data. Since the network structure of this neural model is known, we can then test the reliability of our conditional Granger causality approach based on different types of data (LFP, firing rate and spiking intervals). For experimental data, we

give the detailed information and procedure for measuring the LFP data from sheep under different conditions.

5.2.1 Neurodynamical model used for synthesized data

The basic architecture of neurodynamical model was proposed by [Wang, 2002] and extensively explored by [Deco, 2006]. For our model, we used different parameters to describe the inhibitory neurons. The task of this network is to take a decision between two possible alternatives, according to the characteristics of a sensory input, by reaching one of two predetermined firing states. A typical task is to compare two different stimuli, for instance, vibrotactile stimulation frequency.

A single neuron in our neuron model is the leaky integrate and fire model, defined by the equation

$$C_{\mathrm{m}} \frac{\mathrm{d}V(t)}{\mathrm{d}t} = -\mathrm{g}_{\mathrm{L}}(V(t) - V_{\mathrm{L}}) - I_{\mathrm{syn}}(t) \qquad (5.2.1)$$

where $V(t)$ is the membrane potential of the neuron, C_{m} is the membrane capacitance, $I_{\mathrm{syn}}(t)$ is the synaptic input received by the neuron, and g_{L} is the leak current conductance. When no input is present, the membrane potential

drifts to the rest value V_{L} (-70 mV); if the membrane potential reaches a threshold V_{thr} (-50 mV), the neuron is said to have spiked and the potential is reset to a reset value V_{reset} (-55 mV). The potential is then clamped to the reset value for a brief refractory time τ_{ref}.

When a neuron spikes, it generates excitatory (driving towards 0 mV) or inhibitory (driving towards -70 mV) outputs to all the other neurons. Excitatory neurons generate fast AMPA-mediated current spikes and slower NMDA-mediated spikes, while inhibitory neurons generate GABA-mediated currents, and the synaptic input for a single neuron is the sum of these currents:

$$I_{\mathrm{syn}}(t) = I_{\mathrm{AMPA,ext}}(t) + I_{\mathrm{AMPA,rec}}(t) + I_{\mathrm{NMDA,rec}}(t) + I_{\mathrm{GABA}}(t) \qquad (5.2.2)$$

The currents are defined by:

$$I_{\mathrm{AMPA,ext}}(t) = \mathrm{g}_{\mathrm{AMPA,ext}}(V(t) - V_{\mathrm{E}}) \sum_{j=1}^{N_{\mathrm{ext}}} s_j^{\mathrm{AMPA,ext}}(t) \qquad (5.2.3)$$

$$I_{\mathrm{AMPA,rec}}(t) = \mathrm{g}_{\mathrm{AMPA,rec}}(V(t) - V_{\mathrm{E}}) \sum_{j=1}^{N_{\mathrm{E}}} w_j s_j^{\mathrm{AMPA,rec}}(t) \qquad (5.2.4)$$

$$I_{\mathrm{NMDA,rec}}(t) = \frac{g_{\mathrm{NMDA,rec}}(V(t) - V_{\mathrm{E}})}{1 + \exp(-0.062V(t))/3.57} \times \sum_{j=1}^{N_{\mathrm{E}}} w_j s_j^{\mathrm{NMDA,rec}}(t) \qquad (5.2.5)$$

$$I_{\mathrm{GABA}}(t) = \mathrm{g}_{\mathrm{GABA}}(V(t) - V_{\mathrm{I}}) \sum_{j=1}^{N_{\mathrm{I}}} s_j^{\mathrm{GABA}}(t) \tag{5.2.6}$$

Where V_{E} and V_{I} are the reversal potentials for AMPA and GABA currents. w_j are the synaptic weights, each receptor has its own fraction s_j of open channels, and its own synaptic conductance g. Every neuron receives inputs in the form of excitatory AMPA current spikes, following a Poisson distribution, from N_{ext} independent external connections. In absence of any input, the frequency is 3Hz for all the neurons in the network. The fractions of open channels are described by:

$$\frac{\mathrm{d}s_j^{\mathrm{AMPA,ext}}(t)}{\mathrm{d}t} = -\frac{s_j^{\mathrm{AMPA,ext}}(t)}{\tau_{\mathrm{AMPA}}} + \sum_k \delta(t - t_j^k) \tag{5.2.7}$$

$$\frac{\mathrm{d}s_j^{\mathrm{AMPA,rec}}(t)}{\mathrm{d}t} = -\frac{s_j^{\mathrm{AMPA,rec}}(t)}{\tau_{\mathrm{AMPA}}} + \sum_k \delta(t - t_j^k) \tag{5.2.8}$$

$$\frac{\mathrm{d}s_j^{\mathrm{NMDA}}(t)}{\mathrm{d}t} = -\frac{s_j^{\mathrm{NMDA}}(t)}{\tau_{\mathrm{NMDA,decay}}} + \alpha x_j(t)(1 - s_j^{\mathrm{NMDA}}(t)) \tag{5.2.9}$$

$$\frac{\mathrm{d}x_j(t)}{\mathrm{d}t} = -\frac{x_j(t)}{\tau_{\mathrm{NMDA,rise}}} + \sum_k \delta(t - t_j^k) \tag{5.2.10}$$

$$\frac{\mathrm{d}s_j^{\mathrm{GABA}}(t)}{\mathrm{d}t} = -\frac{s_j^{\mathrm{GABA}}(t)}{\tau_{\mathrm{GABA}}} + \sum_k \delta(t - t_j^k) \tag{5.2.11}$$

The values of all these parameters are displayed in **Table 1** and were taken from [Broadbent, 1975], in which they were calculated with a mean field analysis in order to obtain a network with stable decision states. The sums over k represent a sum over spikes formulated as δ-Peaks$[\delta(t)]$ emitted by pre-synaptic neuron j at time t_j^k. In our modified version of the network, we have two different populations of inhibitory neurons: a fraction S of the inhibitory neurons has a GABA spike decay time constant of $\tau_{\mathrm{GABA,slow}} = 100$ ms. To keep the constant amount of inhibition in the network we scale the GABA conductance for all neurons by a compensating factor f. To keep constant the average amount of inhibition in the network means to choose f so that the average charge transferred by a GABA spike remains constant, i.e.

$$N_{\mathrm{I}} w_j g_{\mathrm{GABA}} \tau_{\mathrm{GABA}} = S N_{\mathrm{I}} w_j f g_{\mathrm{GABA}} \tau_{\mathrm{GABA,slow}} + (1-S) N_{\mathrm{I}} w_j f g_{\mathrm{GABA}} \tau_{\mathrm{GABA}} \qquad (5.2.12)$$

Solving the equation, we can get f

$$f = \frac{\tau_{\mathrm{GABA}}}{S\tau_{\mathrm{GABA,slow}} + (1-S)\tau_{\mathrm{GABA}}} \qquad (5.2.13)$$

which can be read as f being the ratio between the original τ_{GABA} value and the new average time constant of the inhibitory neurons.

Table 1: Model Parameters [Broadbent, 1975**]**

<table>
<tr><th>All Neurons</th><th>Excitatory neurons</th><th>Network Parameters</th></tr>
<tr><td rowspan="13">V_L = -70 mV
V_{thr} = -50 mV
V_{reset} = -55 mV
V_E = 0 mV
V_I = -70 mV
τ_{ref} = 2 ms
τ_{AMPA} = 2 ms
$\tau_{NMDA,rise}$ = 2 ms
$\tau_{NMDA,decay}$ = 100 ms
τ_{GABA} = 10 ms
α = 0.5 ms^{-1}</td><td>C_m = 0.5 nF
g_m =25 nS
$g_{AMPA,ext}$ = 2.08 nS
$g_{AMPA,rec}$ = 0.104 nS
g_{NMDA} = 0.327 nS
g_{GABA} =1.287 nS</td><td>w_+ = 2.2
w_- = -0.8444
w_i =1.015
w_{medium} = 1
N_{ext} = 800</td></tr>
<tr><td>‘2 pools’ model only</td></tr>
<tr><td>Inhibitory neurons</td><td>$\tau_{GABA,slow}$ = 100 ms
S = 0.25
f = 0.3077</td></tr>
<tr><td rowspan="4">C_m = 0.2 nF
g_m = 20 nS
$g_{AMPA,ext}$ = 1.62 nS
$g_{AMPA,rec}$ = 0.081 nS
g_{NMDA} = 0.258 nS
g_{GABA} = 1.002 nS</td><td>‘Slow’ model only</td></tr>
<tr><td>τ_{GABA} = 32.5 ms
f = 0.3077</td></tr>
</table>

5.2.2 Experimental procedure

All animal experiments were performed in strict accordance with the UK 1986 Animals Scientific Procedures Act (including approval by the Babraham Institute Animal Welfare and Ethics Committee) and during the animals were housed inside individual pens and able to see and communicate with each other. Food and water were available *ad libitum*. Post-surgery all animals received both post-operative analgesia treatment to minimise discomfort and antibiotic treatment to prevent any possibility of infection.

Following a 2 week period of initial behavioral training, sheep were habituated to being placed in a trolley and viewing life-sized images of sheep faces back projected on a screen 1 metre in front of them. They were then surgically implanted under general anesthesia (30-35 ml propafol i.v. and closed-circuit isoflurane 3-5%) and full aseptic conditions with three planar 32-electrode arrays (epoxylite coated, etched, tungsten wires with 250μm spacing – total array area ~1mm x 1mm, electrode impedance ~0.2MΩ, tip diameter ~ 1μm, tip exposure length ~100um) aimed at the lIT, rIT and rAC. Animals were also given an anti-inflammatory (1.75ml carprofen i.m. just before surgery to reduced help reduce post-operative swelling, and immediately following surgery received a broad spectrum antibiotic (7ml Terramycin i.m.) and analgesic (1 ml Vetergesic, i.m.). The electrode lengths varied by ~1mm and so this com-

bined with the tip exposure electrodes would have been recording activity across all cortical layers. Holes (0.7cm diameter) were trephined in the skull and the *dura* beneath cut and reflected. IT electrode arrays were placed 18-20mm lateral to the midline, 35mm posterior to the tip of the frontal pole and at a depth of 20-22 mm from the brain surface using a stereotaxic micromanipulator. Electrode depths and placements were calculated with reference to X-rays, as previously described [Kendrick, 1991]. For the rAC electrode arrays were placed 20 mm posterior to the tip of the frontal pole, 5mm lateral to the midline and 12 mm from the brain surface. They were fixed in place with dental acrylic and stainless-steel screws attached to the skull. Two of these screws acted as reference electrodes, one for each array. Electrodes were connected to 34 pin female plugs also cemented in place on top of the skull.

Starting 3 weeks after surgery the electrodes were connected via male plugs and ribbon cables to a 128 channel electrophysiological recording system (Cerebrus 128 Data Acquisition System – Blackrock Microsystems, USA). Recordings were made in the same experimental setting when the animals were conscious and either exposed to a series of projected sheep face images (8 different frontal views of unfamiliar sheep faces on a black background repeated 3-times in a random sequence – i.e. 24 images in total. During each recording session the picture series was repeated 3-4 times. Images were displayed for a 1 s duration and a white fixation spot was shown in between. An experimenter

controlled stimulus presentation and made sure the sheep was looking at the fixation spot on each occasion using a CCTV camera. This image set was shown to each animal on three different occasions. For the resting state condition the animals were recorded on 1-2 occasions in the same environment but with no images projected on the screen in front of them. Subsequently the animals were given general anesthesia (30-35ml propafol and 4-5% isoflurane) and recordings were made again while they were exposed to the same two stimulus conditions. During exposure to visual images the eyes of the anaesthetised animals were held open with fine suture and clips and irrigated regularly with sterile saline. At the end of this final anesthesia recording session animals were euthanized with an intravenous injection of sodium pentobarbitone and the brains removed for subsequent histological confirmation of X-rays that array placements were within the IT cortex region and in the anterior dorsal cingulate cortex just above and behind the *genu* of the corpus *callosum* (i.e. broadly equivalent to Brodman area 24 in the human brain)(displayed in **Figure 5.1**). The experiments were carried out in Babraham institute and general region where electrodes were located within the IT is also shown in [Ge, 2009].

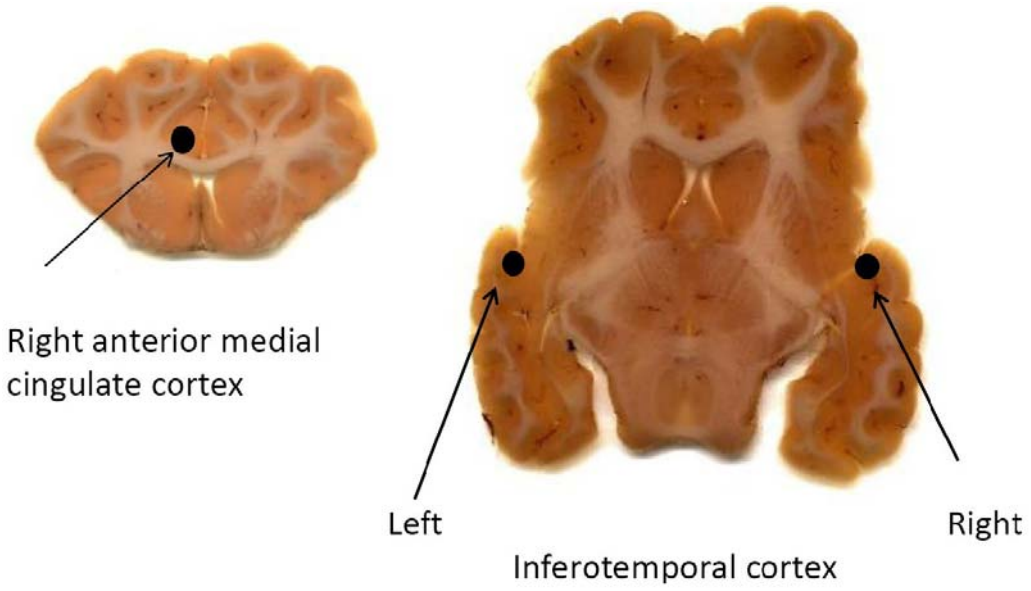


Figure 5.1 General location of electrode arrays in anterior cingulated and inferotemporal cortices. A coronal and horizontal section of the sheep brain showing the general location of the tips of recording electrode arrays (black circles) in the right anterior medial cingulated (approximately equivalent to Brodman 24) and in the left and right inferotemporal cortices.

The LFPs were sampled at 2 kHz and digitized for storage. We used custom Spike 2 (Cambridge Electronic Design, Cambridge, UK) scripts to translate these into text files arranged either by trial or electrode prior to further analysis. For the face stimulus sessions' data from 1 second before and 1 second during each face picture was used for analysis. For the resting condition a total of 3-5 min of continuous recording data was used and taken from a period during the recording session when the animal was calm and showing minimal interest in its surroundings. Any LFP data contaminated with noise artefacts, such as from animal chewing food, were excluded as were LFPs with unexpectedly high power. For LFPs, offline filtering was applied in the range of 1-200 Hz (third-order Chebyshev type I filter, with 0.5 dB of ripple in the pass band). Trend was removed before spectral analysis. All analyses were carried out using custom written routines in Matlab (The Mathworks Inc, Natick, MA).

5.3 Results and Discussion

5.3.1 Synthesized Data

Our neural network structure can be described as **Figure 5.2** by using the Equations (5.2.1) to (5.2.13). From the figure, we can easily find that it is a

fully connected network. The neuron model consistes of 5 different neuron pools which are connected to each other. And neurons in each pool are also inter-connected. However, the weights of the synaptic connections are different and assumed to be generated by a Hebbian learning process. Neurons are strongly connected to each other in the same specialized pool A and B, represented by a weight w_+. A weak connection w_- between neurons from different subsets or between specialized and non-specialized neurons was used; a dedicated weight w_i mediates the interaction from the inhibitory neurons to the excitatory neurons and a 'medium' weight w_{medium} covers all the other cases. All parameters are shown in **Table 1**.

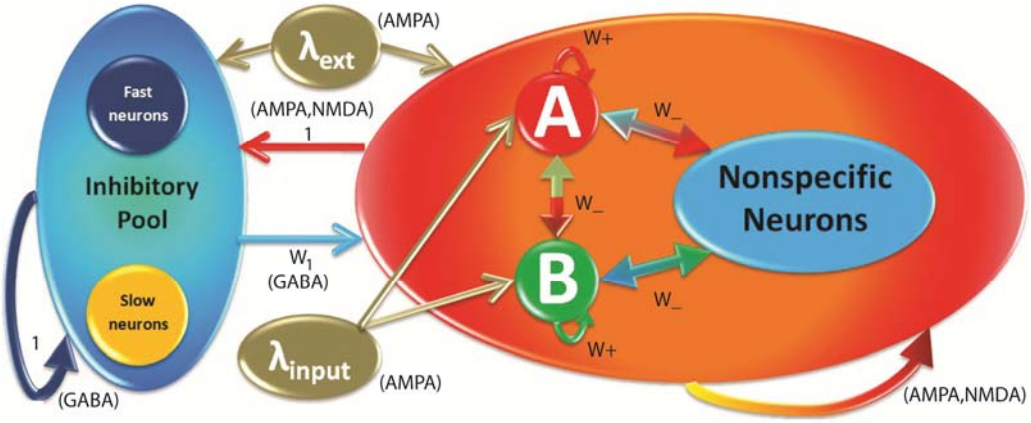


Figure 5.2 The neurodynamical model used for the simulations. Diagram of the 2 pools network model. The network architecture can be divided into two blocks: inhibition pool (200 neurons) and excitatory pool (800 neurons). Two specialized pools (A and B) which have 80 neurons each receive additional inputs.

The network is composed of 1000 neurons, 800 of which are excitatory and the rest are inhibitory neurons including fast and slow model. Two specialized pools (A and B) of 80 neurons each, taken from the 800 excitatory neurons, which receive additional inputs; the remaining 640 neurons construct the non-specialized pool. Three types of data are analysed by applying Granger causality analysis: LFP data (**Figure 5.3_A**), firing rate data (**Figure 5.3_B**) and spiking interval data (**Figure 5.3_C**). There is a stimulation at second 4, which causes a sharp increase for all signals.

For stationary purpose, the data is truncated and only the time points before stimulation are used for Granger causality approach. For spiking intervals data, Gaussian blur filter with standard deviation as parameter is applied firstly to obtain a continuous data set. A more detailed description for applying Granger causality to spiking train data by using Gaussian filter is shown in [Stevenson, 2010]. The 95% confidence intervals for all possible connection are plotted in **Figure 5.4**.

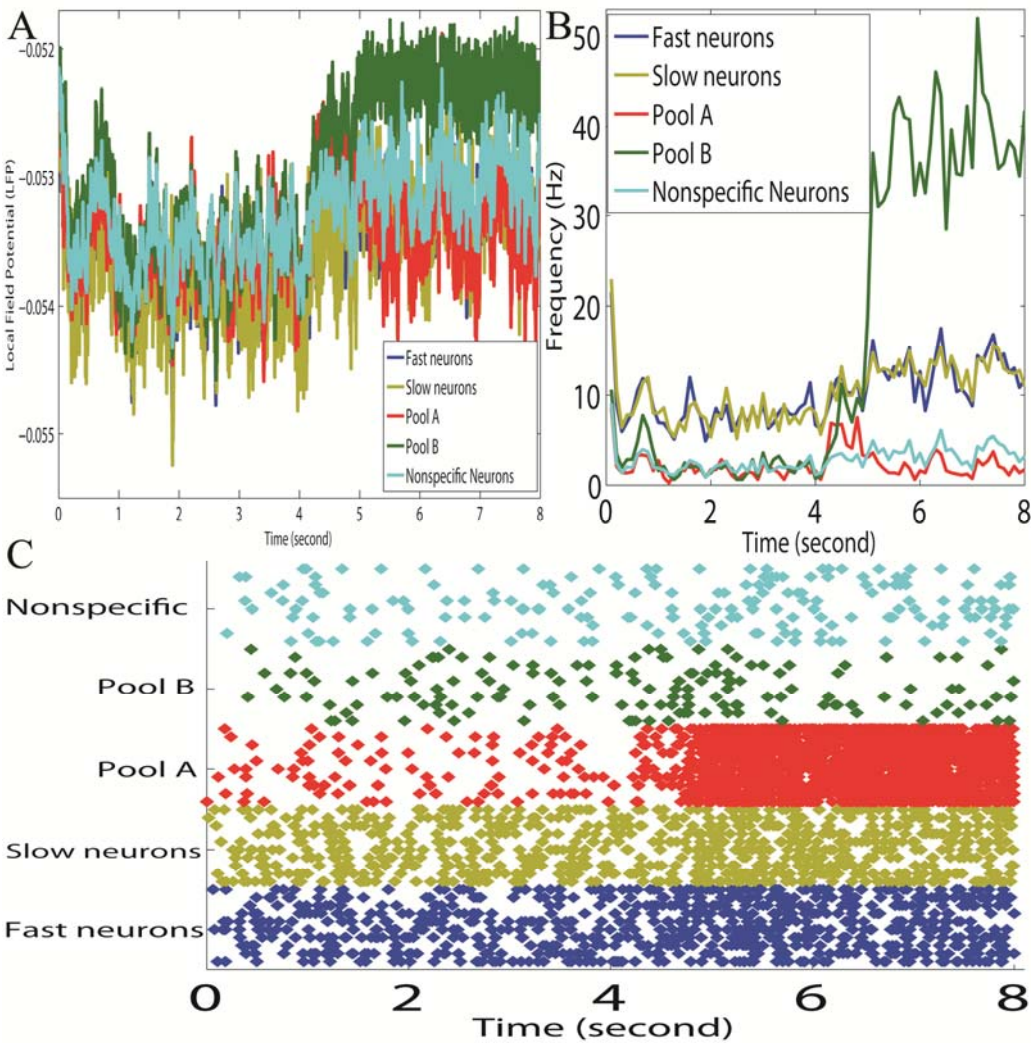
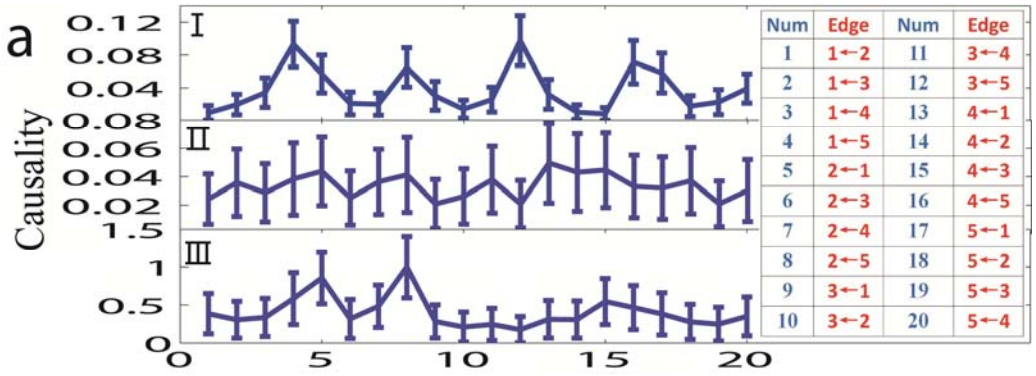


Figure 5.3 The data generated by using neurodynamical model. (**A**) The Local Field Potential (LFP) data generated by this neural model. (**B**) Firing rate data generated by this neural model. (**C**) Plot of spiking events versus time for 10 random neurons taken from each population.



Num	Edge	Num	Edge
1	1←2	11	3←4
2	1←3	12	3←5
3	1←4	13	4←1
4	1←5	14	4←2
5	2←1	15	4←3
6	2←3	16	4←5
7	2←4	17	5←1
8	2←5	18	5←2
9	3←1	19	5←3
10	3←2	20	5←4

Figure 5.4 Conditional Granger causality is applied on three types of data generated from neural network. Plot of the 95% confidence intervals for all the possible directed connections is displayed. For visualization purpose, all directed edges (causalities) are sorted and enumerated into the table. 1-Fast neuron pool. 2-Slow neuron pool. 3-Pool A. 4-Pool B. 5-Nonspecific Pool.

According to the confidence intervals we derived, **Figure 5.5** shows the network structures inferred from conditional Granger causality approach for LFP data, firing data and spiking intervals data in order (The 10 most significant causalities are shown). The ranking of causality strength is proportional to the width of arcs. From **Figure 5.5_A** and **Figure 5.5_B**, we find that the directed networks are almost identical with 8 same arcs. They are very consistent with the preset connection weights and number of neurons used in our original designed model. Since there are more than 600 neurons in the nonspecific pool, it thus has very strong influence on the other pools. For spiking interval data, since only 10 random neurons are selected from each population, it shows very weak relations among populations. **Figure 5.6** shows the Granger causality result in frequency domain. It shows a very strong interaction from the nonspecific pool to slow neurons at theta frequency (4-8Hz).

5.3.2 Experimental Data

Local field potential (LFP) recordings were made from the 32-electrode arrays implanted in the rIT, lIT and rAC of two adult sheep (some data examples can be viewed in [Ladoure, 2009]) while they viewed a series of sheep face images back-projected on a screen in front of them, or while they were in a

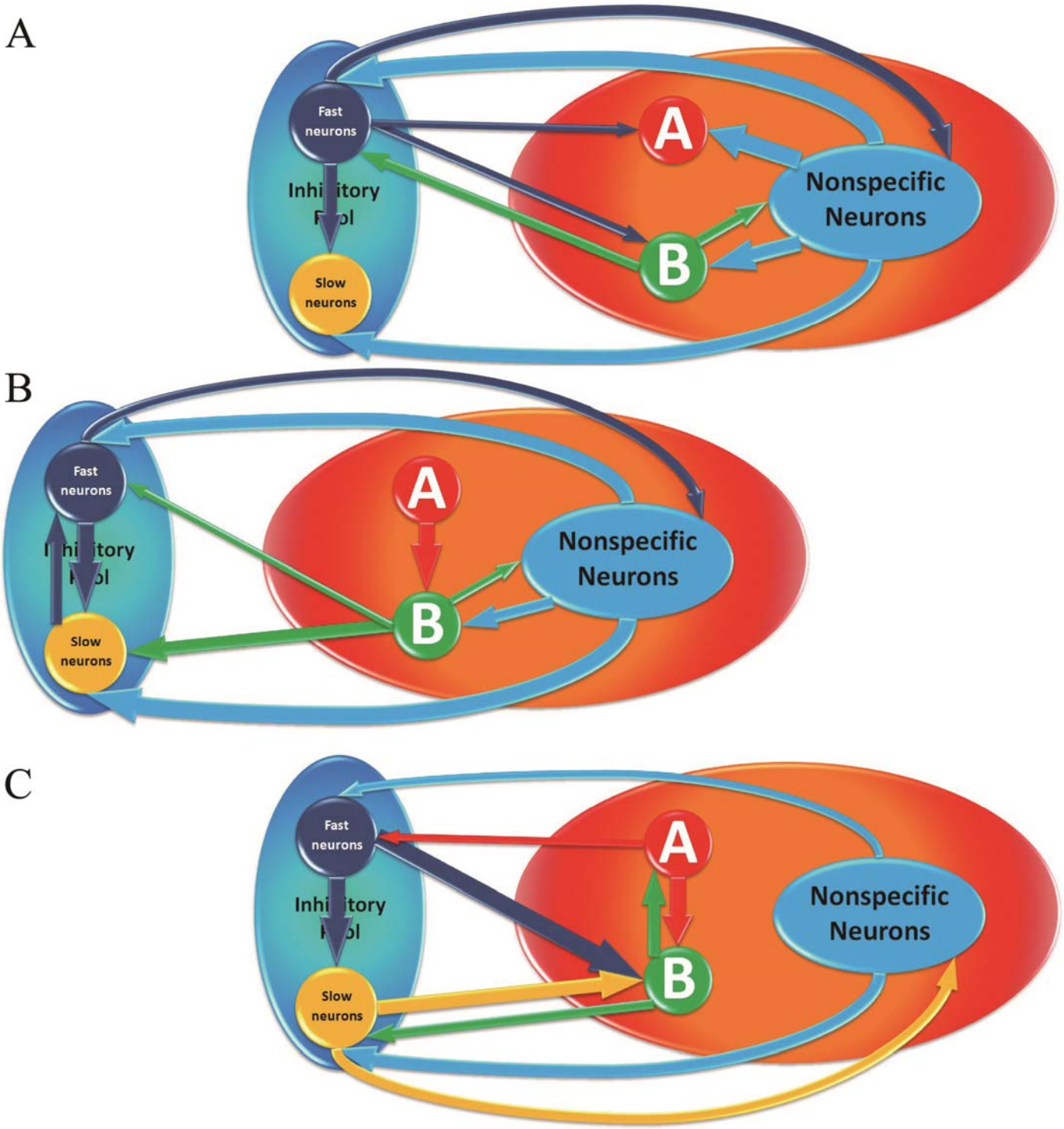


Figure 5.5 Granger causality approach applied on synthesized data. The 10 most significant connections are given. The strength of the connections are proportional to the thickness of the arcs (**A**) The directed network structure derived by using LFP data. (**B**) The directed network structure derived by using firing rate data. (**C**) The directed network structure derived by using spiking interval data.

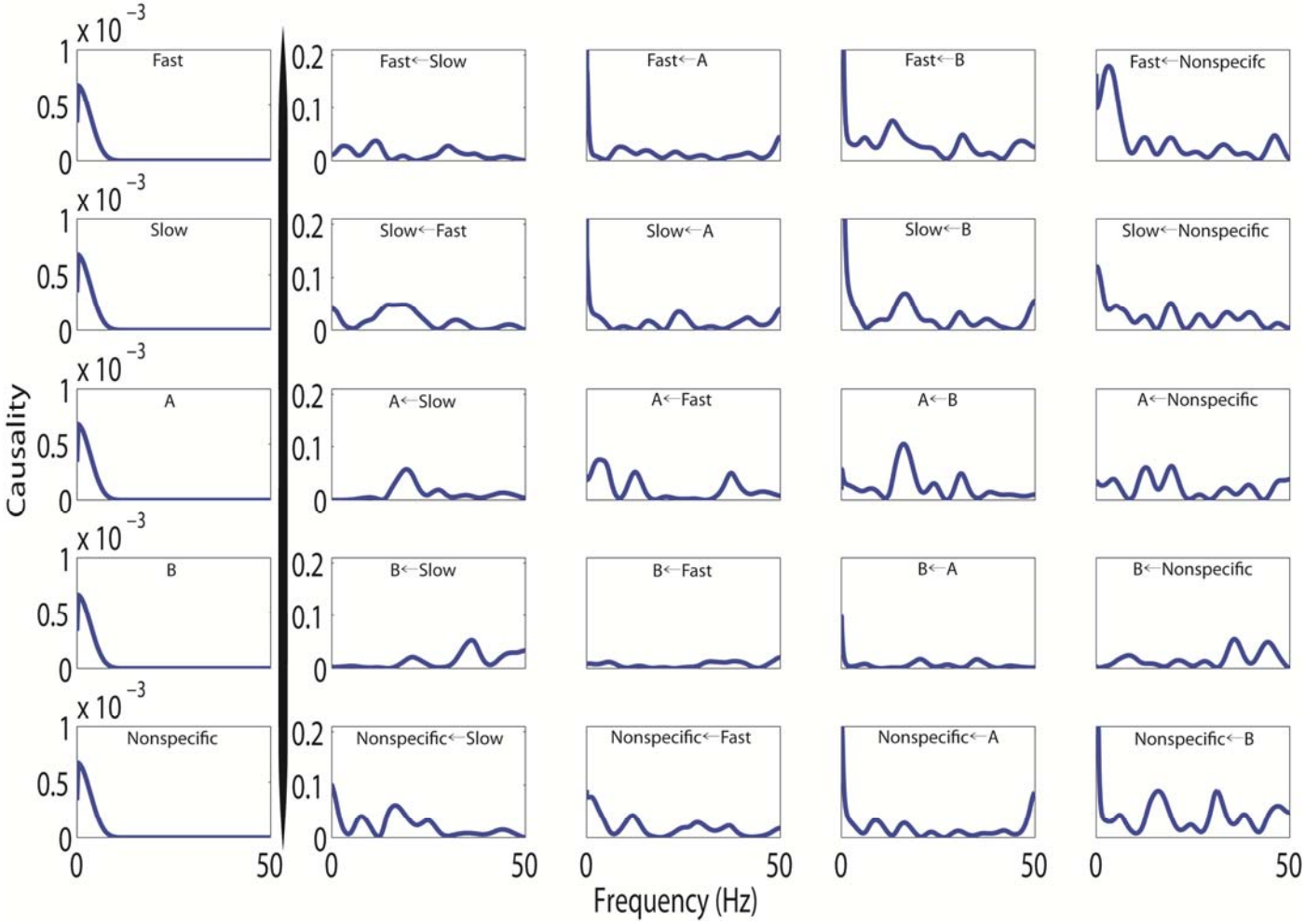


Figure 5.6 Conditional Granger causality is approached in frequency domain. The left-most column shows the power spectrum for each population. The other columns show the conditional Granger causality results in frequency domain. The frequency domain result shows a very strong interaction from nonspecific pool to inhibitory pool (slow neurons and fast neurons) at theta frequency (4-8Hz).

neutral resting state situation (i.e. the same familiar general visual environment but with no face pictures shown or other task requirements).

Effects of resting state, face stimuli and anaesthesia on cortical connections in the time domain

When the animals were recorded in a conscious state, a Granger causality analysis of the LFP data in the time domain showed that exposure to face stimuli resulted in causal connections from the lIT to the rIT either directly or indirectly via the rAC. There was also a reciprocal connection from the rAC to the lIT (**Figure 5.7_B, 5.8_A).** By contrast, during resting state conditions the flow of causal connections was in the reverse direction from rIT to lIT both directly and indirectly via the rAC. There was still however a strong reciprocal connection from the rAC to the lIT (**Figure 5.7_A, 5.8_A**). Therefore, while the direction of information flow and hemispheric dominance were different in the two conditions a common motif was for there to be strong feedback connections between the rAC and lIT.

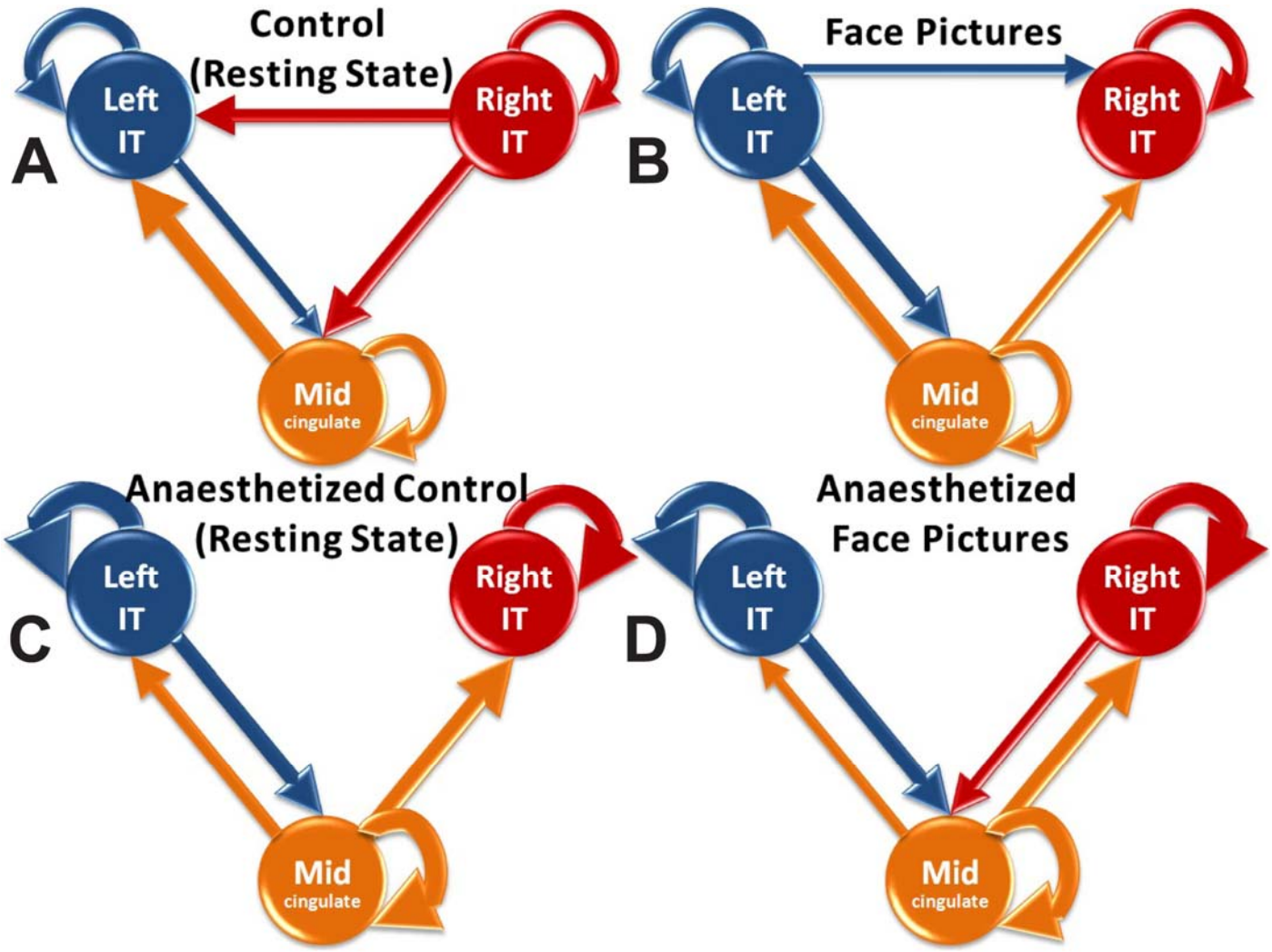


Figure 5.7 Conditional Granger causality is applied to LFP data from the lIT, rIT and rAC in two sheep to derive the direction and strength of both between and within structure connections. **(A)** when the sheep are conscious and resting and not exposed to visual stimuli **(B)** they are conscious and watching face pictures. **(C)** they are anaesthetized and not exposed to face stimuli and **(D)** they are anaesthetized and shown the same face images as in **(B).**

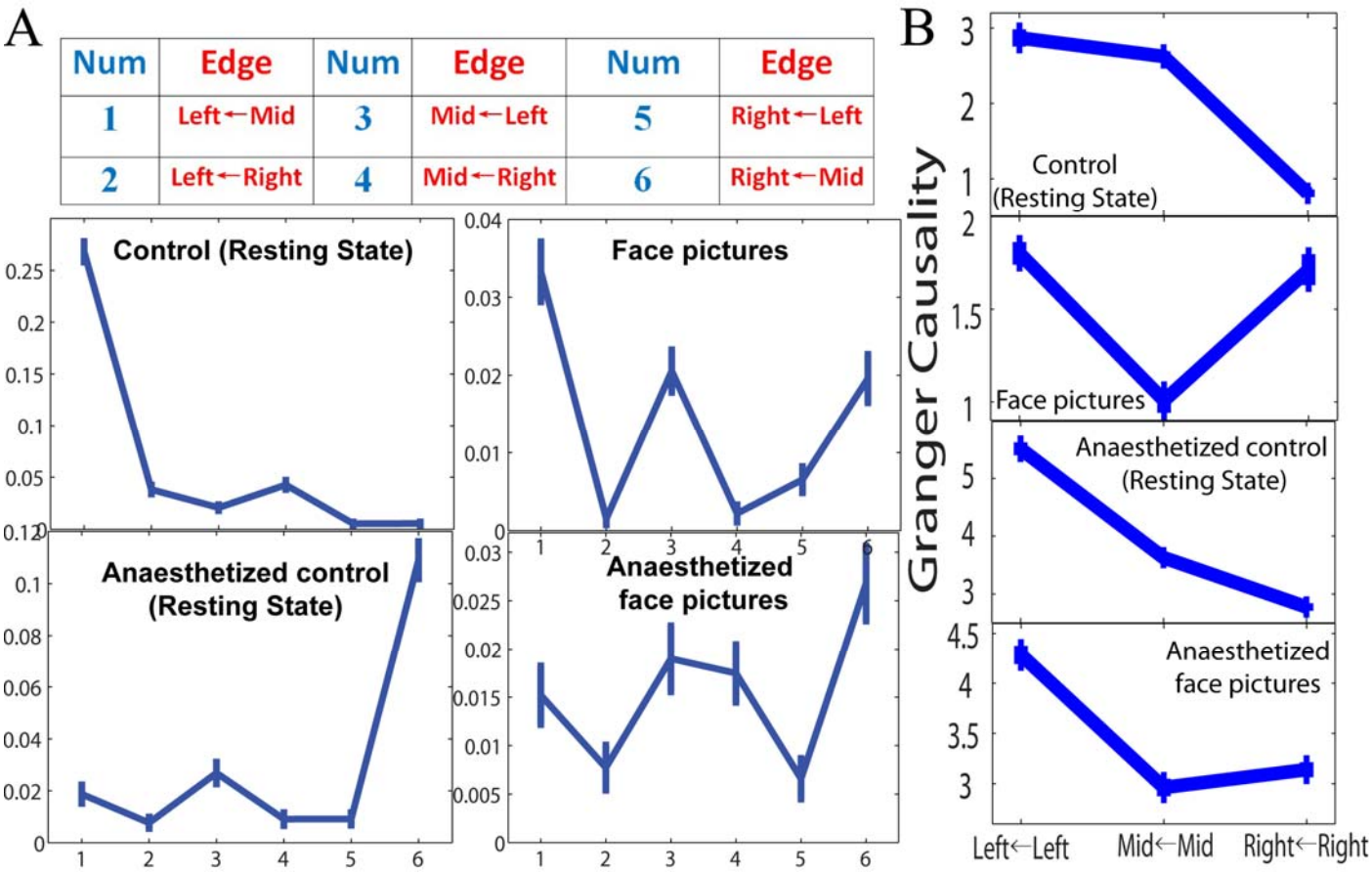


Figure 5.8 Conditional Granger causality is applied to experimental data. (**A**) Plots of the 95% confidence intervals of Granger causality calculated by using a bootstrapping method for the four conditions. (**B**) Plots of the 95% confidence intervals for the inner structure connections

When LFPs were recorded from the same animals exposed to the two conditions under general anesthesia (propafol and isoflurane), there was a considerable impact on the strength of causal connections and their direction (**Figure 5.7C, 5.8A)**. In both cases direct causal connections between the ITs were abolished. Indirect causal connections via the rAC emerged in the reverse direction to that seen when the animals were conscious, whereas those in the same direction were either absent (resting) or unchanged (face stimuli). In both situations a clear motif associated with anaesthesia was for the strong causal connection from the rAC to the lIT to be weakened and for the appearance, or strengthening, of causal connections from the rAC to the rIT. The anterior cingulate has been implicated in the emergence of conscious awareness [Liotti, 2001; Egan, 2003] and so perhaps this particularly involves its influence on processing in the left hemisphere. Indeed, in a human PET study it was the left cingulate gyrus that was activated during the emergence of a consciousness awareness of thirst [Egan, 2003].

Effects of resting state, face stimuli and anaesthesia on cortical connectivity in the frequency domain

An analysis of all the main brain oscillation frequencies revealed that during the resting state there was significantly ($p < 0.05$) greater mean causality

per Hz for delta (1-4Hz), low (4-6Hz) and high (6-8Hz) theta, alpha (8-13Hz) and beta (13-30Hz) frequencies compared with the face stimulus condition. However, gamma (30-70Hz) mean causality was higher during the face pictures (**Figure 5.10_A**). A detailed Granger causality analysis of the different frequencies for each of the six individual connections is shown in **Figure 5.11.** Thus during the resting state situation low frequency oscillations are functionally more dominant than higher ones, whereas following visual stimulation with faces there is an approximately equivalent contribution from all frequencies. Overall power levels of brain oscillations were similar in the two conditions (see **Figure 5.12**) and are therefore not predictive of functional connection strength.

In the frequency domain, when the animals were conscious both low (1-8Hz main contributed by theta and delta) and high (8-70Hz mainly contributed by Alpha, Beta and Gamma) oscillation-based causal connections showed the same unidirectional flow patterns as seen with the time-domain analysis, although the reciprocal connection between the rAC and lIT was only seen in the face picture condition (**Figure 5.10_B and 5.11)**.

During anaesthesia in both conditions there was a reduction in power in the higher frequencies (>10Hz), particularly gamma, and an increase in that for low frequencies (<10Hz), particularly in the delta range. The visual evoked po-

tential at each site following exposure to face pictures also showed a considerable simplification during anaesthesia, with less positive peaks in particular, and appeared less synchronised across structures. This is similar to previous reports [John, 2005] (see **Figure 5.9 and 5.12**). A frequency domain analysis revealed that direct connections between the ITs were lost for both the low and high frequencies and there was emergence of a reversed flow in the indirect connections via rAC for the higher frequencies under anaesthesia state. For the low frequencies this reversed flow only occurred in the resting state since for the face pictures there were no longer any indirect connections between the rIT and lIT, although reciprocal ones between the rIT and rAC emerged as in the time domain analysis. Loss of high frequency (gamma) coherence between the left and right frontal cortices has also been reported during anaesthesia in humans [John, 2005], and similarly alpha has been shown to be particularly involved in interhemispheric communication [Vecchio, 2007].

In the resting state averaged causality between all the regions was significantly decreased during anaesthesia for theta, alpha, beta, whereas delta and gamma were unchanged. In the face picture condition, causality in all frequencies other than delta was decreased (see **Fig 5.10_A**). The stronger contribution from lower frequency oscillations during the resting state, together with the direction of causal information flow from the right to the left hemisphere, is consistent with the hypothesis that low frequencies are particularly dominant in the

right brain hemisphere which is thought to play a key role in spatial information processing involving a more global type of processing [MacNeilage, 2009]. Conversely the greater contribution of gamma together with a left to right direction of information flow is consistent with the hypothesis that in the left brain hemisphere, which is thought to employ more localised processing in the context of analysing detail and organisation and control of action [MacNeilage, 2009], high frequency oscillations are more dominant. In the resting condition it therefore seems likely that anaesthesia reversed the flow of information to left to right due to a greater reduction of causal connections based on the low and middle frequency (theta, alpha and beta) oscillations relative to the highest ones (gamma). In the face picture condition on the other hand there is a relatively weaker contribution from low and middle frequencies and stronger one at the highest frequency, gamma, leading to left hemisphere dominance. In this case the reversed flow under anaesthesia may primarily reflect the reduction in gamma leading to a greater influence from the lower frequencies, even though causal connections in the latter are also reduced.

While for the most part low and high frequencies appeared to have similar causal influences on connectivity in this network, a notable exception was the connections between rAC and lIT where these were only weakened or abolished by anaesthesia in the low frequency domain. This again supports the influence of the right hemisphere over the left being low frequency dominated.

therefore our results provide clear evidence that in both resting state and visual stimulation connections loss of consciousness due to anaesthesia considerably alters information flow between temporal and cingulate cortices. The general motif in both cases is for direct inter-hemispheric connectivity to be lost and for the direction of indirect connectivity between ITs via the cingulate to be reversed. Right hemisphere dominance in the flow of connectivity tends to be maintained by strong causal connections in the lower to middle oscillation frequencies whereas left hemisphere dominance is associated with higher ones in the gamma range. Reciprocal functional connections within all three regions, and between the temporal and cingulate cortices in the right hemisphere, are actually increased by anaesthesia. This could perhaps simply be viewed as a general breakdown in the organisation of cortico-cortical connections under anaesthesia leading to lack of conscious awareness. However, it could also be viewed as a shift within these networks away from open-loop, long-distance feed forward connections which support a form of conscious meta representation of information towards a closed-loop recurrent network where information is still processed but in a more limited, localised and automated fashion with no conscious meta representation.

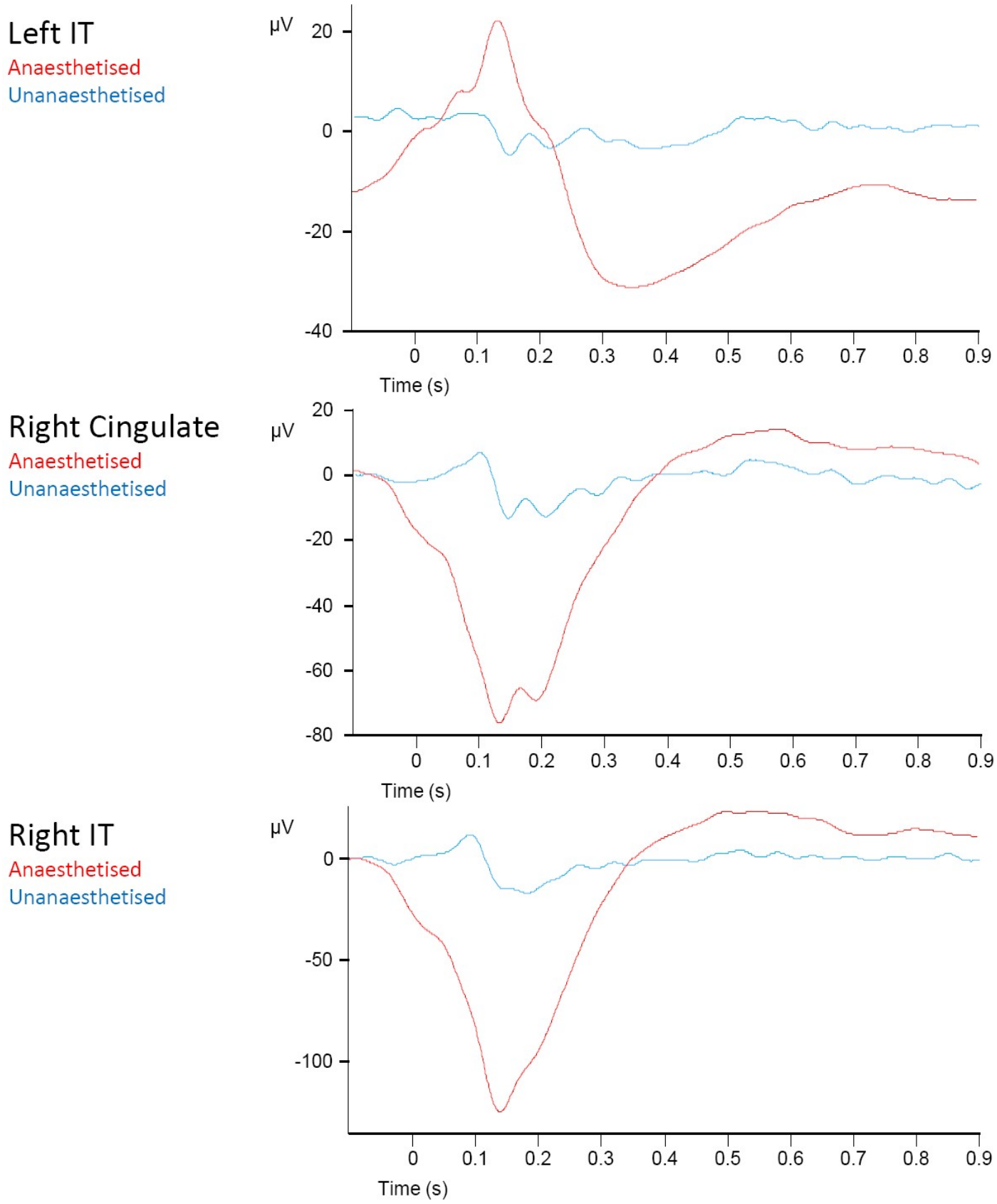


Figure 5.9 Effects of anaesthesia on visual evoked potentials. Averaged (across 32 electrodes and face presentations) visual evoked potentials (VEP) from the right anterior medial cingulate and the right and left inferotemporal cortices for one sheep. During anaesthesia it can be seen that many of the detailed components of the VEP are reduced or lost, particularly the positive ones. There is also more variation in VEPs between the three recording sites.

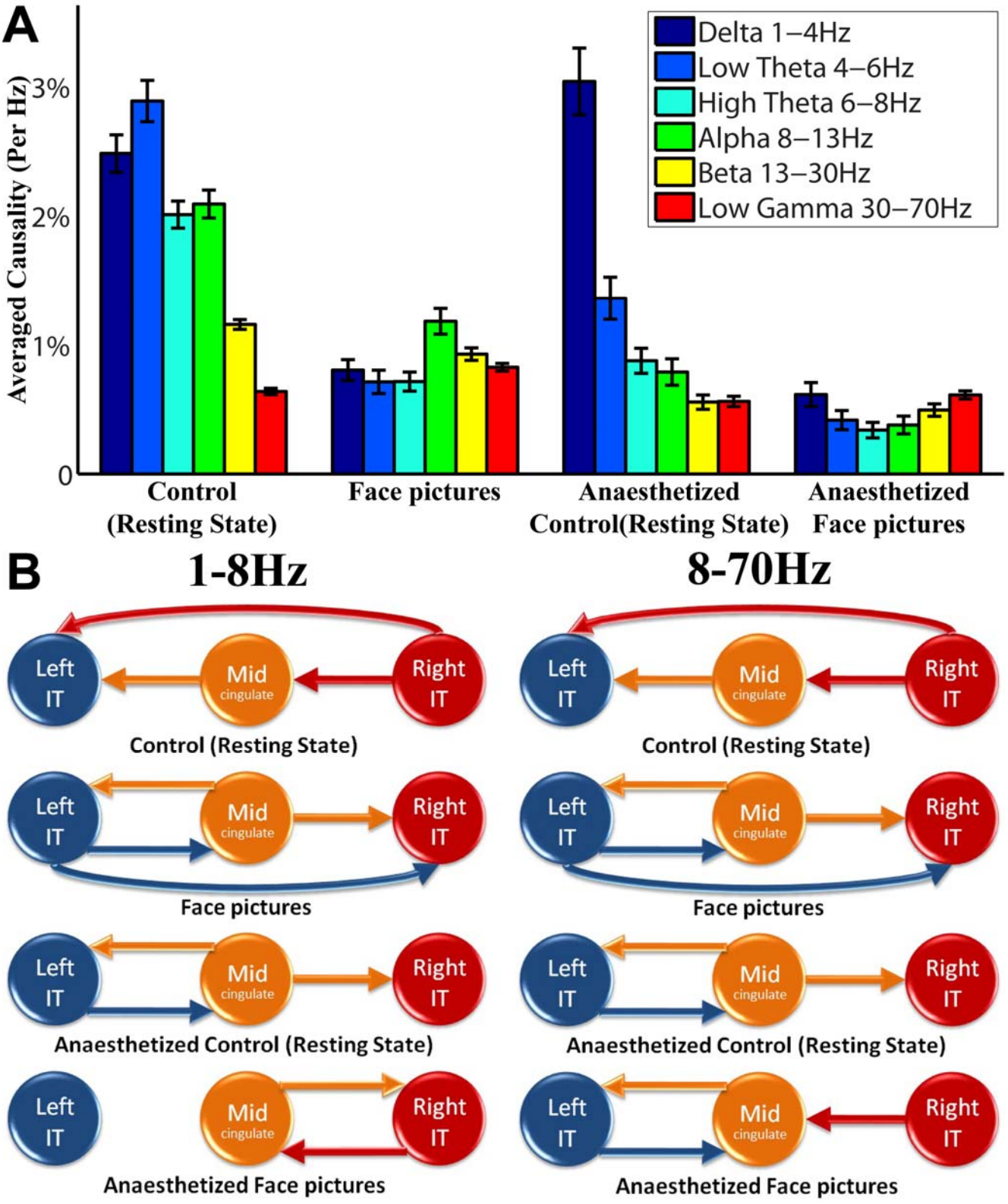


Figure 5.10 Conditional Granger causality applied to LFP recordings in the frequency domain. **(A)** Histograms show mean and 95% confidence levels for averaged percentage causality per Hz in different frequencies. The percentage of averaged causality per Hz for the range f_1 to f_2 can be calculated as $\left(\sum_{\omega=f_1}^{f_2} F_{i\leftarrow j}(\omega)\right) / \left[\left(\sum_{\omega=0}^{100} F_{i\leftarrow j}(\omega)\right)(f_2 - f_1)\right]$ for all $i, j \in \{Left, Mid, Right\}$.(**B**) Strength and direction of causal connections between lIT, rIT and rAC in the different conditions based either on averaged low frequencies (1-8Hz) or higher frequencies (8-70Hz).

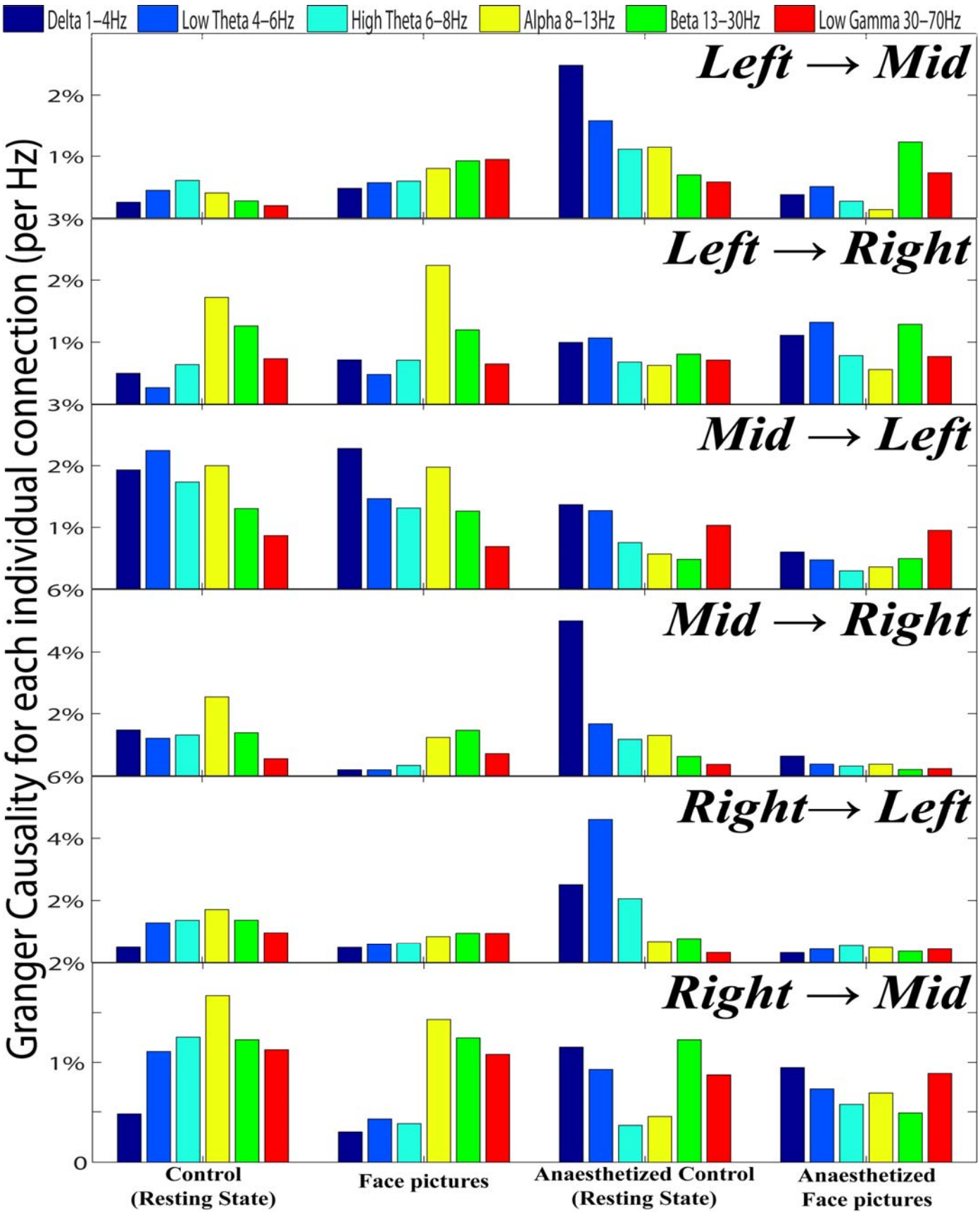


Figure 5.11 Granger causality per Hz calculated from local field potential data for each connection in the different experimental conditions. The delta, low theta, high theta, alpha, beta and low gamma band causalities (as a percentage) of each individual connection between the right anterior cingulate and the right and left inferotemporal cortices were calculated in the resting state and face picture stimuli conditions both when the animals were conscious and deeply anaesthetised.

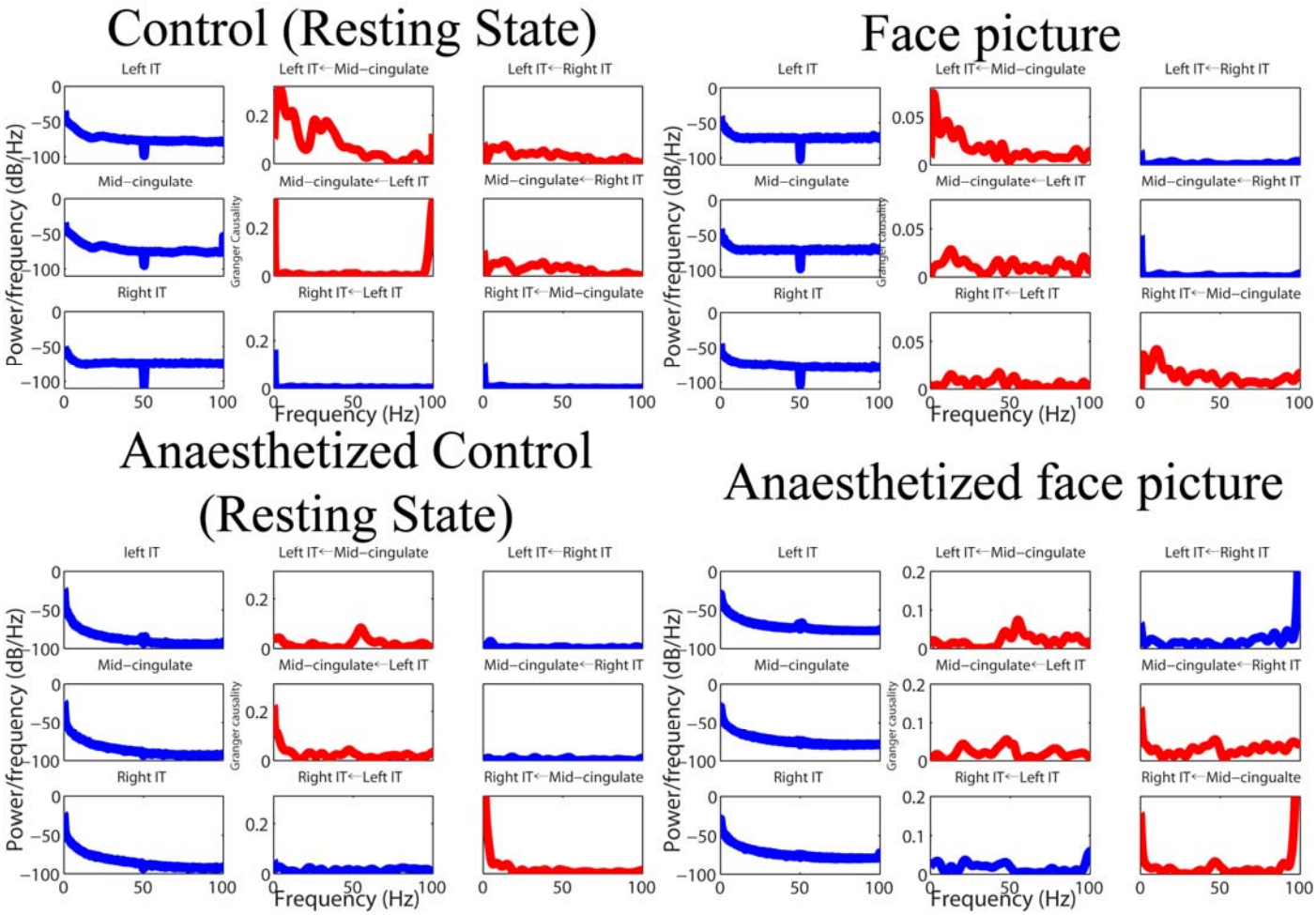


Figure 5.12 Power spectra and conditional Granger causality calculated in frequency domain for sheep data. Granger causality calculated in frequency domain when sheep were either in a resting state, or exposed to face picture stimuli and either conscious or deeply anaesthetised. The figure is divided into 4 panels corresponding to the four different conditions. The left-most column in each panel shows the averaged power spectra from local field potential recordings in the right anterior medial cingulate and right and left inferotemporal cortices in two sheep. There is no clear difference between the resting and face picture conditions but under anaesthesia the power in low frequencies (<10Hz) is increased and that in higher frequencies decreased. The rest two columns in each panel show the Granger causality results calculated in frequency domain refer to the six possible direct edges.

5.4 Summary

In this chapter, the Granger causality is applied on a complex neural dynamical model for three different types of data (LFP, firing rate and spiking train). From the results, we found Granger causality is still reliable for network inference especially for the LFP data. For experimental data, the connectivity in sheep's brain together with its biological meaning was well studied under different condition (consciousness, anaesthesia and visual stimulation). Our derived results show strong evidences that loss of consciousness during anaesthesia is therefore associated with reductions in long-distance open-loop connection and a corresponding increase in the power of short-distance closed loop ones. In the next chapter, the further extension for Granger causality would be discussed.

Chapter 6

Conclusion and Future Work

6.1 Conclusion and contribution

In this thesis, we have systematically investigated the reverse-engineering approaches commonly used in literature (especially the Granger causality approach) for the network re-construction problem in biology. Since there are several well established methods for the reverse-engineering problem, we first compared two commonly used approaches: dynamic Bayesian network inference and Granger causality approach. The comparing study was investigated by using both synthesized data (linear and non-linear model) and experimental data. A critical point of data length was found for synthesized data via intensive computations (more than 100 computers over a few weeks). Our

results illustrated that Granger causality approach outperformed the dynamic Bayesian network inference when the data length is long enough. In addition to the comparative study of the Bayesian network inference, we also tested our Granger causality approach on the biological data published in the *Cell* paper [Cantone, 2009]. In the paper, the authors tested three most commonly used reverse-engineering approaches (ODEs, Bayesian network inference and information theory introduced in **Chapter 1**.**1.2**) on a well known circuit consisting of 5 genes. They concluded that the ODEs worked better than the other two methods. By using the same experimental data, our results illustrated that Granger causality approach was the best one despite being simple. These conclusions directed our future research focus on the application of Granger causality on biological data.

In practice, we often face the problem of over-fit in large network reconstructions using the Granger causality approach when the length data is shorter than the dimensions of variables [Tetko, 1995]. Since the number of time points measured in biology labs is usually short (i.e. 100 time points) and the number of measured genes is very large (more than thousands genes in general), thus it is impossible to fit the AR model for all elements by using such kind of data. To solve this problem, we proposed a novel approach called Global Granger Causality (GGC) approach which is used to reconstruct large networks by using iterative steps.

Furthermore, we applied our Granger causality approach on the LFP data recorded in three areas of sheep brain under different conditions (Rest, Picture Stimulus and Anaesthesia). In cognitive study, the functionality and mechanism of consciousness system is emerging as one of the most important scientific questions. In order to understand the nature of consciousness, the mechanism of general anaesthesia has begun to be investigated and the field of anaesthesiology is playing an important role in understanding consciousness [Mashour, 2006; Akire, 2008]. Our Granger causality results derived from both time and frequency domain show that loss of consciousness due to anaesthesia is associated with an increase in short-distance intrinsic closed loop connections, coupled with reductions in long-distance cortico-cortical connections.

The main contributions of this thesis are:

1. **A comparative study of Bayesian network inference and Granger causality approach**. In computational biology, one often faces the problem of deriving the causal relationship among different elements. Two common approaches can be used to solve such problems, namely, the Bayesian network inference and the Granger causality approach. One natural question is to choose which method to tackle the data, in particular when different causal networks are derived from them. Although there are thousands of publications on each approach, no one

tried to find the advantages, performances and stabilities for both approaches by doing a comparative study before we did. Our results illustrated that Granger causality approach outperformed dynamic Bayesian network inference approach when the data length is long enough on both synthesized and experimental data. This conclusion helped us on choosing causal inferring approaches on biological data in our future research.

2. **Global Granger causality approach for large network reconstructions**. In order to capture the dynamics of complex systems and to investigate the functionalities and mechanisms of genes, proteins and neurons in detail, one often faces the problem of deriving the whole network rather than a very limited portion of it. Generally, most of the analysis tools currently used for a whole network are based on clustering algorithms. The idea of such algorithm is to locate groups of genes that have similar expression patterns over a set of experiments. Reverse-engineering approaches such as Granger causality are then applied to each small group for network reconstructions. However, such analysis could be grossly inaccurate if the group has strong effect from external inputs and hidden variables. To reveal the structure of the whole network, Granger causality has the difficulty come from the fact that the number of variables is larger than the length of the time series. As such,

it causes the model to be under-determined and therefore cannot be uniquely fitted. To solve this problem, we proposed a novel framework called global Granger causality. This novel approach reconstructs the network based on the idea of partial Granger causality by taking iterative steps. This method has been successfully applied on experimental data to derive a network which consisted of 812 proteins

3. **Identifying the network changing in sheep brain due to the effect of anaesthesia**. In literature, it has been reported that loss of consciousness under anaesthesia, sleep or vegetative states may involve reduced functional cortical connectivity caused by disruption of long feed forward connections and synchronization [Mashour, 2006; Akire, 2008; Imas, 2005; John, 2005]. In this thesis, we have combined multi-array recordings of local field potentials in right and left inferotemporal and right anterior cingulated cortices in sheep with Granger causality approach to investigate how anaesthesia alters neural processing during resting state and visual stimulation. Several interesting phenomena were observed from the results derived by our Granger causality approach. When animals view face images a predominatly unidirectional flow direct (between ITs) and indirect (via rAc) causal connections from left to right IT occurs. Under resting state information flow is in the opposite direction. These phenomena agree with the hypothesis that right brain

hemisphere (low frequency) plays a key role in a more global type of spatial information processing, and the left brain hemisphere (high frequency) employs more detailed analysis and action control. Under anaesthesia condition, it results in the abolishment of direct inter-hemispheric connections and increased the intrinsic closed loop connections.

6.2 Further extensions

Granger causality approach has not been widely used yet but it has slowly gathered some interest over the last few years. In its original application, Granger causality is limited to the investigation of pairs of time series. However, indirect connections may produce spurious relations between distant nodes. Conditional Granger causality [Chen, 2006; Ding, 2006; Barnett, 2009] is developed to deal with this restriction by removing the influence of an external node. However, it requires the explicit knowledge of the influencing node, which is usually not possible in practice. To overcome this, partial Granger causality [Guo, 2008] was developed for dealing with external and hidden variables. For large network reconstruction, our global Granger causality approach can be applied. In the complex systems of genes, proteins or neurons, it is often impor-

tant to study the interactions among groups of nodes since elements often work cooperatively or competitively to achieve a task. To tackle this problem, traditional Granger causality has been extended to complex Granger causality [Ladroue, 2009]. In contrast to many similar frameworks, Granger causality relies on dependence over time to define causality instead of using the concept of perturbation. Unified Causal Model [Ge, 2009] was developed to include the notion of stimuli and modifying coupling to traditional Granger causality. The corresponding matlab toolbox was also developed [Seth, 2010].

These recent extensions to Granger causality provided strong supports for its future development and applications in computational biology. It has enormous potentials for further improvement in performance, stability and implementation. Thus it is possible to be a powerful reverse-engineering tool and more widely used in the future.

6.2.1 Performance improvement

We have compared Granger causality approach and Bayesian network inference in the Chapter 3. The corresponding results demonstrated that the dynamic Bayesian network inference outperformed the Granger causality approach when the data length was short due to the over-fit problem that Granger

causality had. The over-fit problem is caused by the batch fitting approach used in the Granger causality method, and the dynamic Bayesian network inference used sequential fitting to overcome this problem. It is a very interesting future research topic for developing Granger causality approach so that it could always outperform the Bayesian network inference. To achieve this aim, one can improve the parameter estimation techniques for Granger causality by appling the similar sequential fitting techniques as the Bayesian network inference used.

6.2.2 Time-varying networks

A Granger causality derived network exposes the strength of the couplings between signals or groups of signals. While elaborate dynamics can efficiently be summarized by this representation, the underlying assumption is that the relations between signals (how they are effectively coupled) are constant during the observations. This issue has only recently been started to be addressed, as mostly for technical reasons reverse-engineering a fixed network is already a difficult task in itself. For the future work, a method for enabling Granger causality networks to cope with time-varying networks is required to be developed. To solve this issue, one could constrain the network not to change too dramatically on a small time scale and thus reduce the parameter

space. By using the sliding window technique, we can re-construct the time-varying networks and find how networks are changing with time.

6.2.3 Impact of signal processing on Granger causality

Granger causality is not often applied to raw data (the data without pre-processing such as noise reductions and stationary processing) in practice. For example, we changed the autoregressive moving average model to autoregressive integrated moving average model in **Chapter 4.2.1** for Granger causality approach. In that chapter, we have proved that this transformation did not impact on the true connection between elements in time and frequency domain. In biological experiments, signals are typically sampled, convolved and filtered before any analysis, and thus involving more kinds of transformation. For Granger causality to be more broadly used and trusted, the effects of these transformations must be elucidated in the future research.

6.2.4 Dissemination

A number of extensions for Granger causality are past the stage of development and reached full maturity. In order to facilitate the uptake of Granger

causality in the scientific community and make it an easily available option, it is necessary to write a publicly available software for computing causalities in various platforms. We have already developed a demo version of Granger causality GUI programmed in Matlab [Zou, 2009]. This software includes a step by step guide of using Granger causality such as downsampling, band-stop filtering, bootstrapping and so on. There are still some limitations for this demo version software. For the future work, a friendly and easy used interface, with a more effective programming platform (i.e. C language) can be developed. More and more extensions of Granger causality can be integrated into this software for various tasks in computational biology.

Bibliography

Akaike, H. (1969). Fitting Autoregressive Models for Prediction. Annals of the Institute of Statistcal Mathmatics, 21, 243-247.

Akire, M.T., Hudetz, A.G., Tononi, G. (2008). Consciousness and anesthesia. Science 322, 876-880.

Albert R. (2007). Network Inference, Analysis, and Modeling in Systems Biology. Plant Cell 19: 3327-3338.

Albo, Z., Di Prisco, G.V., Chen, Y., Rangarajan, G., Truccolo, W., Feng, J., Vertes, R.P., Ding, M. (2004) Is partial coherence a viable technique for identifying generators of neural oscillations. Biological Cybernetics, 90,318-326.

Alon, U. (2006). An introduction to systems biology: design principles of biological circuits. Boca Raton: CRC.

Alon, U. (2007). Network motifs: theory and experimental approaches. Nature, 8, 450-461.

Ancona, N., Marinazzo, D., Stramaglia, S. (2004). Radial Basis Function Approach to Nonlinear Granger Causality of Time Series. Physical Review. E, 70, 056221.

Bach, F.R., Jordan, M.I. (2004). Learning Graphical Models for Stationary Time Series. IEEE transactions on signal processing, 52 (8), 2189-2199.Bansal, M., Belcastro V., Ambesi-Impiombato, A., Bernardo D. (2007). How to infer gene networks from expression profiles. Molecular Systems Biology. 3,78

Barnett, L., Barrett, A.B., and Seth, A.K. (2009). Granger causality and transfer entropy are equivalent for Gaussian variables. Physical Review Letters. 103:238701

Basso, K., Margolin, A.A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat. Genet. 37, 382-390.

Beamish, N., Priestley, M.B. (1981). A Study of Autoregressive and Window Spectral Estimation. Applied Statistics 30 (1), 41-58.

Bishop, C.M. (1995). Neural networks for pattern recognition oxford. Oxford Univ. Press.

Bishop, C.M. (2006). Pattern recognition and machine learning. New York Springer Press.

Bollobás, B. (2001). *Random Graphs* (2nd ed.). Cambridge University Press.

Bonneau R. (2008). Learning biological networks: from modules to dynamics. Nature Chemical Biology 4, 658 - 664 (2008)

Broadbent, D.E, editor. (1975). The magic number seven after fifteen years. :Wiley.

Buntine, W.L. (1994). Operations for Learning with Graphical Models. Journal of Artificial Intelligence Research, 2, 159-225.

Camacho, D.M., Collins, J.J. (2009). Systems biology strikes gold. Cell, 137,24-26.

Cantone, I., Marucci, L., Iorio, F., Ricci, M., Belcastro, V., Bansal, M., Santini, S., di Bernardo, M., di Bernardo, D., and Cosma, M.P. (2009). A Yeast Synthetic Network for In Vivo Assessment of Reverse-Engineering and Modeling Approaches. Cell, 137, 172-181.

Chen, Y., Rangarajan, G., Feng, J., Ding, M. (2004). Analyzing multiple nonlinear time series with extended Granger causality. Physics Letters A, 324, 26-35.

Chen, Y.H., Bressler, S.L., and Ding, M.Z. (2006). Frequency Decomposition of Conditional Granger Causality and Application to Multivariate Neural Field Potential Data. J Neurosci Methods, 150, 228-237.

Chuai-aree, S., Lursinsap, C., Sophatsathit, P., Siripant, S. (2001). Fuzzy C-Mean: A statistical feature classification of text and image segmentation

method. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 9(6), 661-671.

Cohen, A.A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z., Cohen, L., Danon, T., Perzov, N., Alon, U. (2008). Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. Science, 322(5907), 1511-6.

Deco, G., Rolls, E.T. (2006). Decision-making and Weber's law: a neurophysiological model. European Jounal of Neuroscience 24:901-916

Della Gatta, G., Bansal, M., Ambesi-Impiombato, A., Antonini, D., Missero, C., and di Bernardo, D. (2008). Direct targets of the TRP63 transcription factor revealed by a combination of gene expression profiling and reverse engineering. Genome Res. 18, 939-948.

di Bernardo, D., Thompson, M.J., Gardner, T.S., Chobot, S.E., Eastwood, E.L., Wojtovich, A.P., Elliott, S.J., Schaus, S.E., and Collins, J.J. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. Nat.Biotechnol. 23, 377-383.

Ding, M., Chen, Y. and Bressler, S.L. (2006). Granger causality: basic theory and application to neuroscience. Arxiv preprint q-bio/0608035;

Egan, G., Silk, T., Zamarripa, F., Williams, J., Federico, P., Cunnington, R., Carabott, L., Blair-West, J., Shade, R., McKinley, M., Farrell, M., Lancaster, J., Jackson, G., Fox, P., Denton, D. (2003). Neural correlates of the emergence of consciousness or thirst. Proc. Natl. Acad. Sci. USA 100, 15241-15246.

Elena, Z., Julian, M., Dianne, P.O., Teresa, M.P. (2008) Why do hubs in the yeast protein interaction network tend to be essential:reexaminning the connection between the network topology and essentiality. PLoS Computational Biology 4(8):e1000140. Doi: 10.1371/journal.pcbi. 1000140.

Faith, J.J., Hayete, B., Thaden, J.T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J.J., and Gardner, T.S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PloS Biol. 5, e8.

Feng, J., Jost, J., Qian, M. (2007). Networks: From Biology to Theory London: Springer Press.

Feng, J.F., Yi, D.Y., Krishna, R., Guo, S.X., Buchanan-Wollaston, V. (2009) Listen to genes: dealing with microarray data in the frequency domain. PLoS One, 4(4), e5098.

Friedman, N. (2004). Inferring Cellular Networks Using Probabilistic Graphical Models. Science, 303, 799-805.

Friston K.J. (1994). Functional and effective connectivity in neuroimaging: a synthesis. Human brain Mapping 2, 56-78

Gardner,T.S., di Bernardo, D., Lorenz, D., and Collins, J.J. (2003). Inferring genetic networks and identifying compound mode of action via expression profiling. Science 301, 102-105.

Ge, T., Kendrick, K., Feng, J.F. (2009). A novel extended Granger causal model approach demonstrates brain hemispheric differences during face recognition learning. PLoS Comp. Biol. 5(11), e1000570. doi:10.1371/journal.pcbi.1000570

Ge, T., Kendrick, K., Feng, J.F. (2009). A Unified Dynamic and Granger Causal Model Approach Demonstrates Brain Hemispheric Differences During Face Recognition Learning. PLoS Comp. Biol. 5(11), e1000570.

Geisler-Lee, J., O'Toole, N., Ammar, R., Provart, N.J., Millar, A.H., and Geisler, M. (2007). A predicted interactome for Arabidopsis. Plant Physiol. 145: 317–329.

Geweke, J. (1982). Measurement of Linear Dependence and Feedback Between Multiple Time Series. Journal of the American Statistical Association, 77 (378), 304-313.

Geweke, J. (1984). Measurement of Conditional Linear Dependence and Feedback Between Time Series. Journal of the American Statistical Association, 79 (388), 907-915.

Ghahramani, Z. (2004). Learning Dynamic Bayesian Networks Berlin: Springer Press.

Granger, C. (1980). Testing for Causality: A Personal Viewpoint. Journal of Economic Dynamics and Control, 2, 329-352.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37, 424-438

Goldberg, David E. (1989). Genetic Algorithms in Search Optimization and Machine Learning. Addison Wesley 41.

Guo, S., Seth, A.K., Kendrick, K.M., Zhou, C., Feng, J. (2008). Partial Granger Causality-Eliminating Exogenous Inputs and latent Variables. Journal of neuroscience methods, 172 (1), 79-93.

Guo, S., Wu, J., Ding, M., Feng, J. (2008) Uncovering interactions in the frequence domain. PLoS Computational Biology, 4 (5), e1000087.

Horton, P.M., Bonny, L., Nicol, A.U., Kendrick, K.M., Feng, J.F. (2005). Applications of multi-variate analysis of variances (MANOVA) to multi-electrode array data. Journal of Neuroscience Methods, 146, 22-41.

Imas, O.A., Ropella, K.M., Ward, B.D., Wood, J.D., Hudetz, A.G. (2005). Volatile anaesthetics disrupt frontal-posterior recurrent information transfer at gamma frequencies in rat. Neurosci. Lett. 387, 145-150.

ISI Web of knowledge: http://www.isiknowledge.com. On 24th July, 2008, a search on ISI tells us that there are 3531 papers on Bayesian network in the area of Comp. Sci., Math., Math + Comp. Biol. and Business + Economics, and 1125 papers on Granger causality.

Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., Krogan, N.J., Chung, S., Emili, A., Snyder, M., Greenblatt, J.F., Gerstein, M. (2003). A Bayesian Networks Approach for Predicting Protein-Protein Interactions from Genomic Data. Science, 302, 449-453.

Jensen, F.V. (1996). An introduction to Bayesian networks London: UCL Press.

John, E.R., Prichep, L.S. (2005). The Anesthetic Cascade: A theory of how anesthesia suppresses consciousness. Anesthesiology 102, 447-471.

Kendrick, K.M., Baldwin, B.A. (1991) Single unit recording in the conscious sheep. In: Conn PM, editor. Methods in Neuroscience. New York: Academic Press. pp. 3-15.

Kim, J.H., Pearl, J. (1987). Convince: A conversational inference consolidation engine. IEEE Systems, Man, and Cybernetic Society 17, 120-132.

Kirkpatrick, S.; C. D. Gelatt, M. P. Vecchi (1983). Optimization by Simulated Annealing. Science New Series 220 (4598): 671–680.

Kleinberg, J.M. (2000). Navigation in a small world. Nature, 406, 845.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., Lehrach, H. (2005). Systems Biology in Practice: Concepts, Implementation and Application. Weinheim: Wiley-VCH Press.

Ladroue, C., Guo, S.X., Kendrick, K.M., Feng, J.F. (2009). Beyond element-wise interactions: Defining group-to-group interactions for biological processes. PLoS One, 4(9), e6899.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., Zeitlinger, J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J.B., Volkert, T.L., Fraenkel, E., Gifford, D.K., Young, R.A. (2002). Transcriptional Regulatory Networks in Saccharomyces cerevisiae. Science, 298,799-804.

Liotti M, Brannan S, Egan G, Shade R, Madden L, Abplanalp B, Robillard R, Lancaster J, Zamarripa FE, Fox PT, Denton D (2001) Brain responses associated with consciousness of breathlessness (air hunger). Proc Natl Acad Sci USA 98, 2035-2040.

Locke, J.C., Kozma-Bognár, L., Gould, P.D., Fehér, B., Kevei, E., Nagy, F., Turner, M.S., Hall, A., Millar, A.J. (2006). Experimental Validation of a predicted feedback loop in the multi-oscillator clock of Arabidopsis thaliana. Molecular Systems Biology, 2, 59.

MacNeilage, P.F., Rogers, L.J., Vallortigara, G. (2009) Evolutionary origins of our right and left brain. Scientific American. 6/24/09.

Marinazzo, D., Pellicoro, M., Stramaglia, S. (2008). Kernel-Granger causality and the analysis of dynamic networks. Physical review E, 77, 056215.

Mashour, G.A. (2006). Integrating the science of consciousness and anesthesia. Anesthesia and Analgesia 103, 975-982.

Mason, O., Verwoerd, M. (2007). Graph theory and networks in biology. IET Systems Biology, 1, 89-119.

Michael, H. (2005). Understanding Pole/Zero Plots on the Z-Plane. Connexions. http://cnx.rice.edu/content/m10556/2.8/.

Mills, T.C. (1990) Time Series Techniques for Economists. Cambridge University Press.

Morettin, P.A. (1984). Levinson Algorithm and Its Applications in Time Series Analysis. International Statistical Review. 52 (1), 83-92.

Morf, M., Vieira, A., Lee, D.T.L., Kailath, T. (1978). Recursive Multichannel Maximum Entropy Spectral Estimation. IEEE transactions on geosciences electronics, GE-16 (2), 85-94.

Murphy, K. Bayes Net Toolbox for Matlab. (2001). It can be downloaded from the website: http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html.

Nizar, N.B., Laurence, D., Hurst, Mike T. (2006) Evolutionary and physiological importance of hub proteins. PLoS Computational Biology 2(7):e88/ DOI: 10.1371/journal.pcbi.0020088

Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge: Cambridge Univ. Press.

Prill, R.J., Marbach, D., Saez-Rodriguez, J., Sorger, P.K., Alexopoulos, L.G., Xue, X., Clarke, N.D., Altan-Bonnet, G., and Stolovitzky, G. (2001). Towards a rigorous assessment of systems biology models: the DREAM3 challenges. PLoS One, 5(2):e9202.

Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D.A., Nolan, G.P. (2005). Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data. Science, 308, 523-529.

Schelter, B., Winterhalder, M., Timmer, J. (2006). Handbook of time series analysis: recent theoretical developments and applications. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., Alon, U. (2006). Variability and memory of protein levels in human cells. Nature, 444, 643-646.

Smet D.R., Marchal K. (2010). Advantages and limitations of current network inference methods. Nature Reviews Microbiology 8, 717-729

Stevenson, I.H., Körding, K.P. (2010). On the Similarity of Functional Connectivity between Neurons Estimated across Timescales. PLoS ONE 5(2): e9206. doi:10.1371/journal.pone.0009206

Tetko, I.V.; Livingstone, D.J.; Luik, A.I. (1995). Neural network studies. 1. Comparison of Overfitting and Overtraining, J. Chem. Inf. Comput. Sci., 35, 826-833

Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H.,

Goldberg, D.S., Haynes, J., Humphries, C., He, G., Hussein, S., Ke, L., Krogan, N., Li, Z., Levinson, J.N., Lu, H., Ménard, P., Munyana, C., Parsons, A.B., Ryan, O., Tonikian, R., Roberts, T., Sdicu, A.M., Shapiro, J., Sheikh, B., Suter, B., Wong, S.L., Zhang, L.V., Zhu, H., Burd, C.G., Munro, S., Sander, C., Rine, J., Greenblatt, J., Peter, M., Bretscher, A., Bell, G., Roth, F.P., Brown, G.W., Andrews, B., Bussey, H., Boone, C. (2004). Global Mapping of the Yeast Genetic Interaction Network. Science. 303, 808-813.

Seth, A.K. (2010). A MATLAB toolbox for Granger causal connectivity analysis. Journal of Neuroscience Methods. 186:262-273

Tsai, T.Y., Choi, Y.S., Ma, W., Pomerening, J.R., Tang, C., Ferrell, J.E. (2008). Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops. Science. 321,126-129.

Ueda, H.R. (2006). Systems biology flowering in the plant clock field. Molecular Systems Biology, 2, 60.

Vecchio F., Babiloni C., Ferreri F., Curcio G., Fini R., Del Percio C., Rossini P.M. (2007). Mobile phone emission modulates interhemipsheric functional coupling of EEG alpha rhythms. Eur J Neurosci 25:1908-1913.

Wang, S., Chen, Y., Ding, M., Feng, J., Stein, J.F., Aziz, T.Z., Liu, X. (2007). Revealing the dynamic causal interdenpendence between neural and muscular

signals in Parkinsonian tremor. Journal of Franklin Institute-Engineering and Applied Mathematics, 344 (3-4), 180-195.

Wang, X.F., Chen, G. (2003). Complex networks: small-world, scale-free and beyond. Circuits and Systems Magazine, IEEE, 3, 6-20

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron 36:955-968

Wehrens, R., Putter, H., Buydens, L.M. (2000). The bootstrap: a tutorial. Chemometrics and Intelligent Laboratory Systems, 54, 35-52.

Wiener, N. (1956). The theory of prediction. In E. F. Beckenbach, editor, Modern Mathematics for Engineers, chap. 8. McGraw-Hill, New York.

Wu, J., Kendrick, K., Feng, J. (2007). Detecting Hot-Spots in Multivariates Biological Data. BMC Bioinformatics, 8, 331.

Wu, J., Liu, X., Feng, J. (2008). Detecting causality between different frequencies. Journal of Neuroscience Method, 167, 367-375.

Zhan, Y., Halliday, D., Jiang, P., Liu, X., Feng, J. (2006). Detecting the time-dependent coherence between non-stationary electrophysiological signals-A combined statistical and time-frequency approach. Journal of Neuroscience Methods, 156, 322-332.

Zou, C. (2009). Additional materials including a Matlab program and more references on causality analysis are available at [http://www.dcs.warwick.ac.uk/~feng/cell_09.html]