

Graph-theoretic Identification of Dynamic Networks

By

Sina Jahandari

A thesis submitted to the Faculty of the Graduate School of the

University of Minnesota, Twin Cities

in partial fulfillment of the

requirement for the degree of

Doctor of Philosophy

Advisor: Donatello Materassi

May 2022

©2022

SINA JAHANDARI

ALL RIGHTS RESERVED

This dissertation for Doctor of Philosophy degree by

Sina Jahandari

has been approved for the

Department of Electrical and Computer Engineering

by

Nicola Elia,

Murti Salapaka,

Demoz Gebre-Egziabher,

Donatello Materassi ,

Maziar Hemati

May 27, 2022

Date

DEDICATION

To my mom

who gave me wings to fly,

and my dad

who taught me there is nothing in the sky.

ACKNOWLEDGMENTS

I thank Donatello Materassi for teaching me academic elegance.

ABSTRACT

In this dissertation we take a graph-theoretic approach to address different aspects of identification of dynamic networks. We consider a class of networks where each node is assumed to be a stochastic process whose output is influenced by an independent stochastic forcing input and outputs of other nodes. The links (or edges), which represent the influence of other nodes, are assumed to be causal transfer functions.

The first contribution of this work is to develop a technique to consistently identify a single transfer function in a network of dynamic systems using only observational data. It is assumed that the topology is partially known, the forcing inputs are not measured, and that only a subset of the nodes outputs is accessible. The developed technique is applicable to scenarios encompassing confounding variables and feedback loops, which are complicating factors potentially introducing bias in the estimate of the transfer function. The results are based on the prediction of the output node using the input node along with a set of additional auxiliary variables which are selected only from the observed nodes. As in other related prediction error methods, the role of the auxiliary variables is to guarantee that the transfer function from the input node to the output node is consistently identified. However, similar prediction error methods provide only sufficient conditions for the appropriate choice of auxiliary variables and assume a priori information about the location of strictly causal operators in the network. In this dissertation, such an a priori knowledge is not required.

Indeed, another contribution of this work is algorithms to determine if a transfer function in a dynamic network is strictly causal or not. A most remarkable feature of our approach is that the conditions for the selection of the auxiliary variables are purely graphical. Further-

more, within single-output prediction methods such conditions are proven to be necessary and sufficient to consistently identify all networks with a given topology. A fundamental consequence of this characterization is to enable the search of a set of auxiliary variables minimizing a suitable cost function for single-output prediction error identification.

In particular, assuming that the observations have positive additive costs, we develop a systematic algorithm to select a set of auxiliary measurements in order to consistently identify certain transfer functions while minimizing an appropriate cost function. It is shown that sufficient and necessary conditions for consistent identification of a single transfer function are equivalent to the notion of minimum cut in an augmented graph resulted from systematically manipulating the graphical representation of the network. Then, the optimal set of auxiliary measurements minimizing the cost could be found using different approaches such as algorithms from graph theory (i.e. Ford-Fulkerson), distributed algorithms (i.e. push-relabel algorithm), or purely optimization based procedures (i.e. linear programming). The results are also extended to the more challenging scenario where the objective is simultaneously identifying multiple transfer functions. It is shown that the optimal set of observations in this case could be determined via an optimal multi-commodity flow problem with additional commodity specific constraints.

Finally, we consider the problem of designing controllers for networked systems in presence of topological uncertainties. We show that, in some cases, our results in identification of networked systems can be used to model the system by a deterministic term and an uncertain term. Using properties of power spectral density, different bounds for the uncertain term of the model are found in different scenarios. Consequently, standard robust control

methods are directly applicable to design a stabilizer for the closed loop system. However, because of the discrete nature of uncertainties in the structure of the network, more specialized methods might lead to a larger set of controllers.

TABLE OF CONTENTS

CHAPTER

List of Tables	xii
List of Figures	xiii
I. Introduction	1
1.1 Motivation for dynamic networks and Observational data	1
1.2 Control of networks with graph uncertainties	5
1.3 Related works	7
1.4 Contributions of this dissertation	10
1.5 Structure of this dissertation	13
1.5.1 Publications	14
II. Preliminaries, Background and Assumptions	16
2.1 Representations of causal structure and causal relation	16
2.2 A model class for dynamic networks	20
2.3 j -pointing separation	28
2.4 Single door criterion	30
2.5 Flow networks and max-flow min-cut theorem	32
III. Consistent Identification in Dynamic Networks	33
3.1 MISO prediction error method	34
3.2 State of the art in literature	40

3.3	Sufficient conditions for consistent identification	44
3.4	Necessity of the graphical conditions for the selection of auxiliary variables	57
3.5	Proofs related to Chapter III	58
3.5.1	Proof of Proposition 1 and corollary 1.1	58
3.5.1.1	Proof of Proposition 1	58
3.5.1.2	Proof of Corollary 1.1	59
3.5.2	Proof of Theorem 2	59
3.5.2.1	Lemma 4	60
3.5.2.2	Lemma 5	61
3.5.2.3	Lemma 6	62
3.5.2.4	Lemma 8	63
3.5.2.5	Proof of Theorem 2	64
3.5.3	Proof of Theorem 3	72
IV.	Detecting Delays in Dynamic Networks	77
4.1	Detecting Delays	79
4.2	Detecting Feedthroughs	83
4.3	Proofs related to Chapter IV	88
4.3.1	Proof of Lemma 9	88
4.3.2	Proof of Theorem 10	91
4.3.2.1	Lemma 12	92
4.3.3	Proof of Theorem 11	95
V.	Optimal Observations for Consistent Identification in Dynamic Networks	99

5.1	Optimal Selection of Observations for Identification of a Single Module . . .	103
5.2	Optimal Selection of Observations for Identification of Multiple Modules . . .	115
5.3	Proofs related to Chapter V	125
5.3.1	Proof of Theorem 13	127
5.3.2	Proof of Theorem 16	128
5.3.3	Proof of Corollary 25.1	128
5.3.4	Proof of Theorem 17	128
5.3.5	Proof of Theorem 18	128
5.3.6	Proof of Proposition 19	131
5.3.7	Proof of Proposition 20	132
5.3.8	Proof of Proposition 21	132
5.3.9	Proof of Theorem 22	133
5.3.10	Proof of Theorem 23	133
VI.	How Can We Be Robust Against Graph Uncertainties?	134
6.1	A motivating example	135
6.2	Handling unmeasured Confounders	138
6.3	Robustness against graph uncertainties	141
6.3.1	Consistent Identification of $\bar{P}(z)$	143
6.3.2	Finding an upper bound for Δ	144
6.3.3	Modeling graph uncertainty by making ϵ approaching zero	148
6.4	A tighter bound for uncertainty	149
VII.	Conclusion and Discussion	156

	xi
APPENDICES	158
A.	158
1.1 Optimal selection of observations in acyclic networks	158
REFERENCES	166

LIST OF TABLES

TABLE

3.1	All possible predictor sets to consistently identify $H_{21}(z)$ in Example 10 for the given graphical representation	54
5.1	Costs of observing each node in Example 15	101
5.2	Costs of different predictor sets in Example 15	102
5.3	Costs of observing each node in Example 22	126
1.1	Costs of observing each node in Example 27	165

LIST OF FIGURES

FIGURE

1.1	Turing Award Laureate Judea Pearl, best known for championing the probabilistic approach to artificial intelligence and the development of Bayesian networks, commenting on one of our papers showing that causal graphs are beneficial in control theory. Pearl’s comment is available, as of today, at: https://twitter.com/yudapearl/status/1466701863676305408	11
2.1	(a) Equations governing a four-variate stochastic processes. (b) The repetitive causal graph. (c) The unit causal graph. (d) The graphical representation (collapsed causal graph).	17
2.2	Representation of a directed graph.	22
2.3	Block diagram of the network \mathcal{G} discussed in Example 1.	25
2.4	(a) The perfect graphical representation G^p of the dynamic network \mathcal{G} discussed in Example 1; (b) A recursive graphical representation G of \mathcal{G} ; (c) The graph of instantaneous propagations G^f associated with G	25
2.5	Path diagram G of a SEM discussed in Example 3.	31
2.6	The mutilated path diagram G' obtained after removing the path $2 \rightarrow 3$ from G discussed in Example 3.	31
3.1	(a) Block diagram of a two-node network; (b) Corresponding graphical representation.	34

3.2	(a) Equations governing a two-variate stochastic processes (b) The repetitive causal graph including hidden noise processes (c) The graphical representation (collapsed causal graph).	36
3.3	Block diagram of a simple three-node network containing a 1-pointing path ($2 \leftarrow 3 \rightarrow 1$) where the objective is the identification of transfer function $H_{12}(z)$	38
3.4	Graphical representation of the simple three-node network containing a 1-pointing path ($2 \leftarrow 3 \rightarrow 1$) where the objective is the identification of transfer function $H_{12}(z)$	38
3.5	Block diagram of a simple four-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$	39
3.6	Graphical representation of a simple four-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$	39
3.7	Block diagram of a simple three-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$	40
3.8	Graphical representation of a simple three-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$	40
3.9	A simple network highlighting the flaw in [1]	42
3.10	Block diagram of a network discussed in Example 6.	43
3.11	Graphical representation G of a network with a confounding variable node 4 discussed in Example 6.. . . .	43
3.12	The mutilated graph G'_1 obtained after removing the edge $1 \rightarrow 2$ from G in Figure 3.11.	44

3.13	The mutilated graph G'_2 obtained after removing the edge $2 \rightarrow 3$ from G in Figure 3.11.	44
3.14	The block diagram of the network discussed in Example 7.	47
3.15	The Graphical representations of the networks discussed in Example 7.	47
3.16	Revisiting the networks of Example 5.	50
3.17	Block diagram of the network discussed in Example 9. The nodes 3 and 4 act as confounders.	51
3.18	The graphical representation of the network discussed in Example 9. The nodes 3 and 4 act as confounders.	52
3.19	The graphical representation of the network discussed in Example 10.	53
3.20	The graphical representation of the benchmark network discussed in Example 11.	55
3.21	Identification performance of the predictor set $Z_1 = \{4, 5, 6, 7\}$ for different number of measurements.	56
3.22	Identification performance of the predictor set $Z_2 = \{3, 9\}$ for different number of measurements.	57
3.23	The new variable q in G'' is introduced in between node j and all its parents that are not the node i	65
3.24	Scenario one in the proof of Theorem 3.	72
3.25	Generic configurations of scenario two in the proof of Theorem 3.	74
4.1	(a) Block diagram of a network (b) A recursive graphical representation of the network containing two nodes in a feedback loop	78

4.2	The non-recursive graphical representation of the network discussed in Example 13. The nodes 3 and 4 are not measured and act as confounders. . . .	81
4.3	(a) The unknown recursive graphical representation of a network discussed in Example 14; (b) The given non-recursive graphical representation of the network discussed in Example 14; The application of Theorems 10 and 11 allows one to conclude that either (c) or (d) is a valid graphical representation of the network. The application of Theorem 2 with $Z = \{4, 6\}$ leads to the same sets $D^- = \{2, 6\}$ and $D^+ = \{1, 4\}$ for the identification of $H_{21}(z)$ via Procedure 1.	85
5.1	The graphical representation of the network discussed in Example 15 for the determination of the set of predictors with minimal cost.	101
5.2	Placement of new variables p and q when the objective is identification of transfer function $H_{ji}(z)$.(a) Graph G (b) Graph G' after adding variables p and q	104
5.3	A graphical representation G of a network discussed in Example 25. The objective is identification of the transfer function $H_{21}(z)$	105
5.4	Graph G' resulted from graph G of Figure 1.1 after adding fictitious nodes p and q , discussed in Example 25	106
5.5	The ancestor graph G^a resulted from limiting graph G' of Figure 5.4 to ancestors of nodes 1, 2, and p , discussed in Example 17. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set d -separating nodes 1 and 2 in G^a	107

- 5.6 The undirected graph G^{mor} resulted from moralizing the graph G^a of Figure 5.5. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set separating nodes $\{1,2\}$ from node p in G^{mor} 109
- 5.7 (a) A two-node undirected graph, (b) Corresponding augmented graph after splitting the nodes. Flow capacity of the edges of the form $k_{in} \rightarrow k_{out}$ is equal to c_k and flow capacities of all other edges are equal to infinity. 110
- 5.8 The augmented graph G^{aug} resulted from graph G^{mor} of Figure 1.4, reformulating a separation problem to a min-cut problem. 111
- 5.9 The graphical representation G of a network highlighting the differences between identification of a single transfer function and multiple transfer functions using Algorithm 1. To identify only $H_{j_1 i_1}(z)$ the set $\{2, 4\}$ is the optimal predictors set with cost 30 and to identify only $H_{j_2 i_2}(z)$ the set $\{3\}$ is the optimal predictors set with the cost 20. It is possible, however, to identify both $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$ using the optimal predictors set $Z^* = \{1, 4\}$ with the cost 40. Observe that the optimal predictor set to identify both transfer functions is not the union of the optimal predictors set to identify each transfer function independently. 117
- 5.10 The graph G' obtained after adding fictitious nodes $p_1, p_2, q_1,$ and q_2 to the graph G of Figure 5.9 for identification of $H_{j_1 i_1}(z)$ and $H_{j_1 i_2}(z)$ 119
- 5.11 The network discussed in Example 21. The objective is Identification of $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$ while minimizing the cost of observation. 123

5.12	Different regions of parameters α and β corresponding to different solutions for Z^*	124
5.13	Graphical representation of the network discussed in Example 22. The objective is Identification of $H_{j_1 i_1}(z)$, $H_{j_2 i_2}(z)$, and $H_{j_3 i_3}(z)$ simultaneously while minimizing the cost of observations.	125
5.14	(a) An illustrative graph G of the proof of Theorem 13 (b) Graph G' of the proof of Theorem 13	127
6.1	(a) Graphical representation of the network when node 4 is a confounder (b) Graphical representation of the network when node 4 is an intermediate node in a parallel path	135
6.2	Block diagram of the class of networked systems described by (VI.1)	137
6.3	Assumption (VI.5) is equivalent with assuming that either the switch S_{41} is off ($H_{41} = 0$) or the switch S_{14} is off ($H_{14} = 0$).	137
6.4	(a) The unknown true topology of the interconnections of the network (b) The known topology of the network	139
6.5	(a) The unknown true topology of the interconnections of the network (b) The known topology of the network	139
6.6	Designing a controller C when there is a feedback between 1 and 4	142
6.7	(a) The Nyquist plot of $P_a(z)$ in Example 24, (b) The Nyquist plot of $P_b(z)$ in Example 24, (c) The Nyquist plot of $\frac{P_a(z)+P_b(z)}{2}$ in Example 24	153
1.1	A graphical representation G of a network discussed in Example 25. The objective is identification of the transfer function $H_{21}(z)$ using (A.1) and minimizing the cost of observation (V.1).	161

- 1.2 The ancestor graph G^a resulted from limiting graph G of Figure 1.1 to ancestors of nodes 1 and 2 discussed in Example 25. 162
- 1.3 The mutilated graph G' resulted from removing the edge $1 \rightarrow 2$ from the graph G^a of Figure 1.2 discussed in Example 25. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set d -separating nodes 1 and 2 in G' 162
- 1.4 The undirected graph G^{mor} discussed in Example 26 resulted from moralizing the graph G' of Figure 1.3. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set separating nodes 1 and 2 in G^{mor} 163
- 1.5 The graphical representation G of the network discussed in Example 27 for the determination of the set of predictors with minimal cost. 164
- 1.6 The undirected graph G^{mor} in Example 27 resulted from moralizing graph G' which is obtained by removing the edge $1 \rightarrow 2$ from G^a which is the subgraph of G limited to ancestors of nodes $\{1, 2\}$. The optimal set separating nodes 1 and 2 in G^{mor} is the optimal set of predictors guaranteeing consistent identification of $H_{21}(z)$ using (A.1). 165

CHAPTER I

INTRODUCTION

1.1 Motivation for dynamic networks and Observational data

By digitalization and technological advancements in the past few decades, we are producing, collecting, and storing enormous amount of data from variety of real-world phenomena. Also, engineering systems have become increasingly complex and interconnected [2–6]. Dynamic networks are a widespread modeling tool to describe the interactions in such large scale systems. Such models have applications in numerous fields such as physics [7, 8], biology [9], chemistry [10], medicine [11], neuropsychology [12], ecology [13, 14], economics [15, 16], engineering [17] and social networks [18].

Identification just by making use of observational measurements [19,20] is of paramount importance for any large scale network fulfilling critical or uninterruptible functions or in situations where it is impractical or too expensive to inject known probing signals into the system .

By observational data, we mean that what is being observed is not the system response to known inputs that have been actively injected to probe or identify the network, but rather the available measurements are being acquired while the system is currently operating and forced by potentially unknown excitations.

Under this paradigm, it will be possible to shape the behavior of a networked system or infrastructure while it is performing uninterruptible or critical functions by synthesizing controllers that could be readily deployed in-line.

The capability of designing controllers just by making use of passive observations is of paramount importance for any large scale network fulfilling critical or uninterruptible functions (i.e., a power grid, a logistic system) or in situations where it is impractical or too expensive to inject known probing signals into the system (i.e., a gene network [20], a financial network [21]). Other relevant applications are in medicine (i.e., repeated drug testing [22], automatically assisted anesthesia [23], Deep Brain Stimulation for Parkinson disease [24]). Indeed, in these cases, for obvious safety and health concerns, it is not desirable to actively test the response of the patient to a different drug dosage or stimulation if comparably useful information could be inferred from passively obtained observations. The possibility of dealing with uncertainties in the network structure also offers the capability of creating self-adapting protocols that are robust with respect to potentially unknown side effects of the treatment.

Furthermore, controllers that can be implemented in medium to large networks can typically access and process only a limited amount of information. It is often not reasonable to assume that the controller has access to all the network measurements along with a full knowledge of the system structure (which might also be time-varying). Thus, a degree

of uncertainty about the network structure often has to be taken into account. In addition to this, in some cases controllers need to be deployed immediately and in-line, for example when activated to counteract sudden failures. Hence, there might not be enough time to identify the specific configuration of the networked system via standard system identification methods based on the active injections of a known signal. Instead, it is feasible to consider a situation where the networked system is continuously operating providing its critical functions, while its configuration is being passively tracked and monitored and such information is used to define the parameters of a controller to be deployed if required. An innovative design methodology allows one to create local, self-adapting, fast, ready to be deployed controllers capable of increasing the resilience of a critical large scale system. In other words it provides a concrete foundation to realize what is usually referred to as a self-healing network.

The possible application areas of the observational framework are numerous. However, we discuss three specific domains of application: design of control mechanisms to improve responses to faults in power grids; stock market analysis techniques; and safe self-tuning of deep brain stimulation for Parkinson's disease.

Power grid reconfiguration to counteract cascade failures: The power grid is an infrastructure providing a critical service that cannot be interrupted. The consequences of a large scale blackout cannot be quantified just in monetary terms, but also in discomfort, hardship and, unfortunately, loss of human lives [25]. It is not possible to have full information about the state of a power grid and its ever-changing configuration.

Monitoring a market, balancing a portfolio and an advanced form of pair trading: From the point of view of a small investor (an investor whose buying or selling transactions

have no significant effect on the market) or of a regulatory agency, a financial network can only be passively observed. The network inference techniques developed in this project can be applied to obtain a snapshot of dynamic relations among a set of stocks. Preliminary evidence for this kind of analysis has been provided in [26] and [27]. Small investors could use these techniques to determine in a quantitative way if a portfolio is balanced or use the information on the stock structure to develop novel trading strategies. More importantly, regulatory agencies could use these methods to detect potential manipulations in the market.

Deep Brain Stimulation for Parkinson's disease: Deep Brain Stimulation (DBS) is considered an effective treatment for several neurological disorders including Parkinson's disease [28]. It involves surgical implantation of a stimulating device in certain areas of the patient's brain. The mechanism of action of DBS for the treatment of Parkinson's disease is yet not completely explained [29]. The stimulating device typically delivers several orders of magnitude more current than what a cluster of neurons would normally produce, and at definitely higher frequencies [30]. In this project the PI intends to investigate what are optimal intensities and frequencies for the stimulation device to mitigate the tremors associated with Parkinson's disease while guaranteeing the safety of the patients. The techniques described in this proposal can be effectively specialized toward this goal. Indeed, the last generation of DBS devices can also measure brain activity at the neural level. By passively monitoring the brain activity, control algorithms for DBS devices can be devised to improve the outcome of the stimulation without compromising the health of the patient.

1.2 Control of networks with graph uncertainties

Designing controllers for unknown or uncertain systems is the goal of several branches of control theory, such as robust control and adaptive control [31–35]. Robust control typically assumes that a previous identification procedure has already provided a model for the system along with a description of its uncertainties. On the other hand, adaptive control tries to identify the plant while tuning the parameters of a controller in closed loop with the system. In both cases uncertainties are typically only on the dynamic operators of the network (i.e., transfer functions in a linear scenario). Indeed, neither standard robust control nor standard adaptive control methods are capable of dealing with the uncertainties of the structure of the system (e.g the uncertainty due to the action of a confounding signal that is not measured) when only observational data is available . Indeed, this kind of uncertainty is not represented via bounds on dynamic operators.

Furthermore, methods to reconstruct networks rely on node knockout techniques and the active injection of exciting signals [36–38]. This dissertation deals with the more challenging scenario where only observational data is available. Control problems with this kind of constraint require an identification and a control analysis/synthesis procedure both of which need to be robust with respect to the uncertainties in the network structure. Robust control and adaptive control are frameworks that were originally formulated for single individual systems. Only subsequently, because of their increasing relevance, researchers have started to consider extensions of these frameworks to distributed systems. Typically though, they limit themselves to simple scenarios that describe single localized failures

(such as in [39]). Other approaches try to cover more complex situations, but the methodology to the control design does not happen to be systematic [40,41].

On the other hand, there are also control methodologies explicitly oriented toward interconnected systems. Indeed, the design of distributed controllers for networks of dynamical systems is an extremely active area of research [42,43]. It is well-established that constraints on the amount of information that controller components can exchange with the plant and between each other pose formidable challenges to the design of optimal regulation strategies [44–46]. However, the communication structure of a decentralized controller is either chosen a priori [47,48] or determined using optimization techniques that promote sparsity of the communication structure (i.e., ℓ_1 regularization, Reweighted Least Squares, Orthogonal Matching Pursuit) [49]. These approaches are effective when both the structure and the dynamics of the networked system are fully known and not dynamically changing. Indeed, these methods tend to exploit the known structure and dynamics of the system along with some spatial/time invariant properties of its topology. Furthermore, many of these methodologies are not well-suited to be extended to a dynamically changing structure. For example, techniques based on sparse optimization would be required to solve a new optimization problem every time a change in the network structure is detected.

However, it is possible to leverage relevant results developed in the area of graphical models [50–52] and statistics, in particular Structural Equation Models (SEMs) [53,54], to address the problem of designing controllers for passively observed systems. The literature on graphical models and SEMs is extensive, but it is principally focused on the graphical description of a joint probability distribution of random variables. Such a distribution is assumed to admit a sparse factorization that can be aptly represented by a Directed Acyclic

Graph [55]. Many methodologies have been derived to infer information about graphs that are compatible with a valid factorization. Fundamental work in this area has been pioneered by Judea Pearl and his group (see [50, 56, 57]) and by many other researchers [51, 52, 58].

However, in the case that we plan to examine in this dissertation, namely a network of processes influencing each other according to a directed network, such a factorization has no meaning, especially if feedback loops have to be taken into account. The treatment of stochastic processes bears other several technical complications, such as the presence of noise terms potentially correlated in time and the necessity of causal (real-time) estimators. These complications are not present in graphical models since they simply represent random variables. Furthermore, as an additional limitation, standard graphical model approaches and results are usually derived considering a finite number of random variables. For all these reasons, these methodologies will need to be revisited, sometimes in a radical way, to be applied to the case of dynamical systems that is of interest for this dissertation.

1.3 Related works

Designing a suitable input to be injected into a system in order to identify its dynamics is a common strategy in identification theory [59]. Apart from injecting known inputs [60], networked systems offer other options of active intervention to facilitate the identification process. For example, standard approaches involve removing certain edges and/or knocking out specific nodes [36, 61].

In several recent results, it has been shown that, by appropriately introducing additional measured variables to a set of predictors, a consistent estimate of a certain trans-

fer function can be obtained using prediction error method or substantially equivalent tools [1, 60, 62–65]. This idea has been explored in the extension of closed-loop identification techniques to network identification [1, 62, 66] and by applying graphical model tools [64].

However, in several applications the network cannot be actively manipulated and data are merely observational. Namely, the network measurements are usually acquired while the system is responding to excitations that are not necessarily known [19, 20].

A wealth of methodologies has also been developed to deal with the problem of identifying a network of dynamic systems from observational data [67–74]. These methods rely on different a priori assumptions and have different identification goals. On one hand, some techniques have as primary goal the recovery of the unknown network graph [67, 68, 75], while the quantitative identification of the network dynamics (i.e. transfer functions) might only be a complementary outcome [69]. On the other hand, some techniques assume that the underlying graph is at least partially known and the very objective becomes to identify the transfer functions describing the dynamics coupling the nodes [70–74, 76]. When the structure of the network is fully known and all the nodes are observed, there exist methods which can identify the full network dynamics [71–74, 76]. Instead, if the structure of the network is only partially known or some nodes are not observed, there are also some techniques capable of identifying a single transfer function in the network. These last methods can be considered more general, since in situations where the network is fully observed and the required conditions are met for all the transfer functions, they could be globally applied to identify the full network dynamics [71–74, 76].

The results of this dissertation fall into the category of techniques aiming at identifying an individual transfer function in a network where the graph topology is partially known. A defining feature of our techniques is that they rely on conditions that are purely graphical and are inspired by the theory of probabilistic graphical models of random variables. The main advantage of graphical model techniques is that they tend to be particularly suited to deal with confounding variables. However, graphical models are typically defined on directed acyclic graphs. Thus, they might not be considered an adequate model to describe scenarios involving feedback loops which are instead central in the theory of automatic control. Conversely, the problem of determining a transfer function involved in a feedback loop within a network is an active topic of research in identification theory. In [70] classical closed-loop prediction error techniques such as direct, two-stage, and joint-input-output methods are extended to be applicable in local network identification settings. Improvements upon the same general ideas were presented in [62] using tools from graph theory, in [65] incorporating Bayesian/kernel methods and in [63] to deal with sensor noise. Similarly, a parametric identification strategy based on the instrumental variable method is proposed in [77]. In [63], the scenario where sensor noise affects the measurements is studied and [60] approaches the problem using informative experiments. Furthermore, the possibility of parametric identification strategies based on instrumental variable methods was explored in [77].

However, closed loop identification methods typically cannot tackle confounding variables as effectively as certain graphical model techniques do. Furthermore, prediction error methods typically require that some information about the locations of strictly causal transfer functions is available. This knowledge is typically formalized by requiring

that there is no algebraic loop for any value of the parameters in the full network parameterization [62].

1.4 Contributions of this dissertation

The first contribution of this dissertation can be interpreted as an attempt to extend certain graphical model tools to deal with the problem of closed loop identification combining the best of the two worlds: an effective way to take into account the unknown locations of strictly causal transfer functions while obtaining an unbiased closed loop identification in presence of confounding variables (see Figure 1.1). In particular, our identification procedure falls in the class of prediction error methods, while the selection of auxiliary variables borrows elements from the theory of graphical models.

The result is a technique which guarantees the consistent identification of a transfer function in a partially observed network by selecting auxiliary predictors using only graphical conditions. Unlike other works our technique is capable of detecting the location of strictly causal transfer functions directly from data. Remarkably, the derived graphical conditions are also proven to be necessary providing a complete characterization of the sets of auxiliary variables that lead to a consistent identification in single-output error prediction methods. This characterization is the basis for the search for a set of auxiliary variables minimizing a suitable cost function for the identification.

As another contribution, in this dissertation we present an algorithm that selects an optimal set of predictors that guarantees consistent identification of a single transfer func-



Judea Pearl ✓
@yudapearl



I taught "control theory" in 1970-1972 and little could I imagine that, one day, control theorists would find Causal Graphs to be beneficial, as this paper shows: [ieeexplore.ieee.org/abstract/docum...](https://ieeexplore.ieee.org/abstract/document/9788888)
The key feature they find useful is the "Single Door Criterion" (see ucla.in/2LPWt9j)



3:32 AM · Dec 3, 2021 · Twitter Web App

Figure 1.1: Turing Award Laureate Judea Pearl, best known for championing the probabilistic approach to artificial intelligence and the development of Bayesian networks, commenting on one of our papers showing that causal graphs are beneficial in control theory. Pearl's comment is available, as of today, at: <https://twitter.com/yudapearl/status/1466701863676305408>

tion in a generic dynamic network. The results are, then, extended to the challenging scenario where the objective is the optimal identification of multiple transfer functions.

For identification of a single transfer function, we establish an equivalency between the sufficient and necessary conditions derived in this dissertation and the notion of minimum-cut set in a flow network resulted from the manipulation of graphical representation of the network.

This is done by adding some meaningful fictitious variables to the network and systematically creating a new directed graph from the graphical representation of the network and reformulating the proposed graphical conditions to the notion of d -separation between

certain nodes in the new graph. Then, taking a few graph theoretical steps, d -separation is reformulated as separation in ancestor moral graphs [56] and eventually as finding the minimum-cut set in an augmented flow graph. Variety of algorithms could be used to find the desired min-cut in the max flow problem setting formulated in this manner. For example, algorithms from graph theory such as Ford-Fulkerson [78], distributed algorithms such as push-relabel algorithm [79], or purely optimization based procedures such as linear programming. The optimal set of predictors is, then, found using the identified min-cut.

The extension to the case when the objective is identifying multiple transfer functions simultaneously is not trivial. Indeed, it turns out that the optimal set of predictors when the objective is identification of multiple transfer functions is not the union of optimal sets of predictors resulted from identifying each single transfer function of interest separately.

We, however, propose a method based on a series of linear programs that given a graphical representation of a network and specifying the transfer functions of interest, finds the optimal set of predictors guaranteeing consistent identification of each transfer function of interest.

Since for a general case the proposed method finds the optimal set of predictors using linear programs, a defining feature of the results of this dissertation is that they enable optimal identification of as many desired transfer functions in a potentially complex network in a systematic way.

Finally, we introduce a notion of robustness with respect to uncertainties in the graph structure of the network. We suggest that, in some cases, it is possible to leverage our results developed in the area of identification of dynamic networks to tackle these forms of uncertainties. We show that while uncertainties in the topology of a network can still

be tackled in some ways under a standard robust control formulation, their discrete nature makes them amenable to be treated using more recent tools borrowed from the area of graphical models. In particular, in a motivating example where only a subset of the variables is being measured, a data driven procedure to design a controller can be successfully obtained even if the structure of the network is partially known. However, the application of methods from the theory of graphical models allows the determination of a larger set of stabilizing controllers.

1.5 Structure of this dissertation

The dissertation is organized as follows. Chapter II reviews preliminary definitions and concepts of causal structures, dynamic networks and their graphical representations, single door criterion, and flow networks.

In Chapter III, a MISO prediction error method and sufficient and necessary graphical conditions for a set of predictor inputs guaranteeing consistent identification in causal networks are presented.

In Chapter IV, two data-driven tests are proposed to detect strictly causal transfer functions and presence of feedthroughs in dynamic networks.

Chapter V casts a formal optimal identification framework for dynamic networks to minimize the cost of observations. An algorithm to find the optimal set of predictor inputs guaranteeing consistent identification of a certain transfer function while minimizing the cost of observations is presented. A simplified algorithm for optimal identification in acyclic networks is provided in the Appendix. Chapter V also extends the results of

general networks containing loops to the case where the objective is optimal identification of multiple transfer functions simultaneously.

Chapter VI investigates robustness against graph uncertainties when the objective is to design a controller for a networked system with structural uncertainties.

Concluding remarks and future research ideas are given in Chapter VII.

1.5.1 Publications

The material in this dissertation is based on the following articles [75, 80–86]:

- Jahandari, Sina, and Donatello Materassi. "Sufficient and Necessary Graphical Conditions for MISO Identification in Networks with Observational Data." *IEEE Transactions on Automatic Control* 2021. (Chapters II, III, IV)
- Jahandari, Sina, and Donatello Materassi. "Optimal Selection of Observations for Identification of Multiple Modules in Dynamic Networks." *IEEE Transactions on Automatic Control* 2022. (Chapters II, V)
- Jahandari, Sina, and Donatello Materassi. "How Can We Be Robust Against Graph Uncertainties?" (Chapters II, VI)
- Jahandari, Sina, and Donatello Materassi. "Optimal Observations for Identification of a Single Transfer Function in Acyclic Networks." *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021. (Chapters II, Appendix)
- Jahandari, Sina, and Donatello Materassi. "Identification of dynamical strictly causal networks." *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018. (Chapters II, VII)

- Jahandari, Sina, and Donatello Materassi. "Topology identification of dynamical networks via compressive sensing." *IFAC-PapersOnLine 51.15 (2018)*: 575-580. (Chapters II, VII)
- Jahandari, Sina, and Donatello Materassi. "Analysis and compensation of asynchronous stock time series." *2017 American Control Conference (ACC). IEEE, 2017*. (Chapters II, VII)
- Jahandari, Sina and D. Materassi, "Links between causal inference and system identification in control theory: optimal selection of adjusting variables in Dynamic Bayesian Networks", (Chapters II, III, VII)

CHAPTER II

PRELIMINARIES, BACKGROUND AND ASSUMPTIONS

In this chapter, we introduce the class of models that is going to be the object of our investigation along with some preliminary concepts and notions from the areas of graphical models, system identification, and flow networks.

2.1 Representations of causal structure and causal relation

In [87], the authors consider a model for a set of stochastic processes x_j , for j in a set of indexes V , governed by the relation

$$x_j(t) = \sum_{i \in V \setminus \{j\}} h_{j,i,0} x_i(t) + \sum_{i \in V} \sum_{d \geq 1} h_{j,i,d} (x_i(t-d)) + \epsilon_j(t) \quad (\text{II.1})$$

where, for integers $d \geq 1$, $h_{j,i,d}$ are smooth univariate functions, $h_{j,i,0}$ are multiplicative constants, and $\epsilon_j(t)$ are unknown independent identically distributed stochastic processes, which are also mutually independent. In model (II.1) the causal coupling among random variables at the same time t is linear while the coupling between past random variables

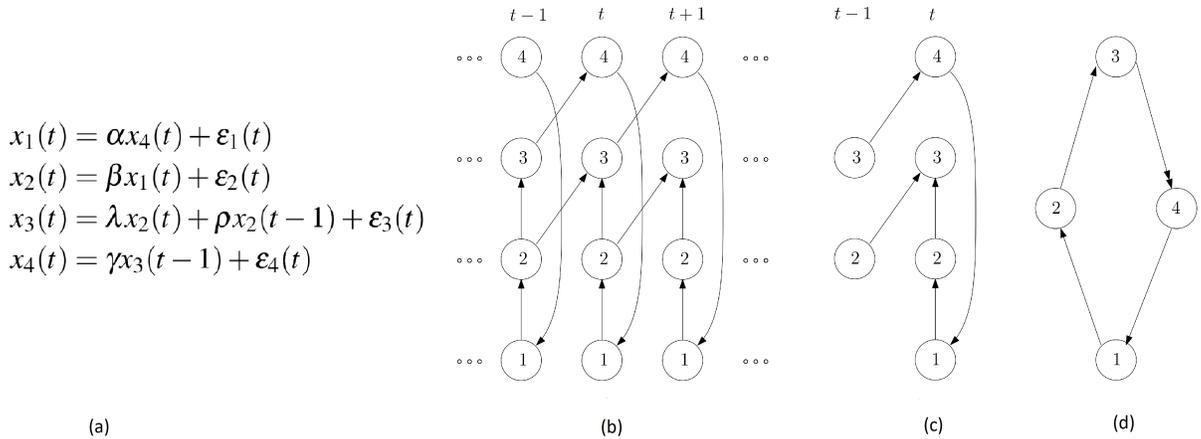


Figure 2.1: (a) Equations governing a four-variate stochastic processes. (b) The repetitive causal graph. (c) The unit causal graph. (d) The graphical representation (collapsed causal graph).

and present ones is potentially nonlinear. The main problem that the authors tackle is the reconstruction of the underlying causal graph using the FCI algorithm [88]. Specifically, the authors introduce two equivalent representations: the repetitive causal graph and the unit causal graph. The repetitive causal graph is a standard graphical model representation of the causal relations among the individual random variables, while the unit causal graph is just the plate representation of the repetitive causal graph [58, 89]. For example consider a four-variate stochastic processes described by the equations given in Figure 2.1 (a) where α , β , λ , ρ , and γ are constants. For such stochastic processes Figure 2.1 (b) shows the repetitive causal graph and Figure 2.1 (c) shows the unit causal graph. One fundamental assumption in (II.1) is that the underlying causal graphs (either the repetitive or equivalently the unit causal graph) is acyclic, that can be formalized by saying that there is no sequence of indices i_1, i_2, \dots, i_m such that $h_{i_1, i_2, 0}$, $h_{i_2, i_3, 0}$, \dots , $h_{i_{m-1}, i_m, 0}$, $h_{i_m, i_1, 0}$, are all nonzero (see C4 of Definition 1 in [87]).

A related problem is the determination of the constants $h_{j,i,0}$ and the functions $h_{j,i,d}$ when the underlying causal graph is known. In most applications, this problem is often formulated considering the case where all the relations are linear and the functions $h_{j,i,d}$ can be fully described by using a simple multiplicative coefficient. So, in the linear case, with a slight change of notation, model (II.1) reduces to

$$x_j(t) = \sum_{d \geq 1} h_{j,j}(d)x_j(t-d) + \sum_{i \in V \setminus \{j\}} \sum_{d \geq 0} h_{j,i}(d)x_i(t-d) + \epsilon_j(t). \quad (\text{II.2})$$

For a pair (i, j) , the causal influence from i to j is simply given by the coefficients $h_{j,i}(d)$, for $d \geq 0$. If $h_{j,i}(0)$ is known to be equal to zero we say that the causal influence is delayed, otherwise we say that it is (potentially) instantaneous. Since all relations in (II.2) are linear, given a pair (i, j) the coefficients $h_{j,i}(d)$ can be determined using the Single Door criterion, or similar tools from the area of graphical models.

Recently, in the area of control theory and system identification there has been growing interest in learning the causal interactions among a set of stochastic processes modeled in ways similar to (II.2). For example, in system identification, models of the following form are often considered [1, 70–72, 74]

$$x_j(t) = \sum_{d \geq 1} h_{j,j}(d)x_j(t-d) + \sum_{i \in V \setminus \{j\}} \sum_{d \geq 0} h_{j,i}(d)x_i(t-d) + \sum_{d \geq 0} f_j(d)\epsilon_j(t-d) \quad (\text{II.3})$$

where the goal of the learning procedure is the same, since we still want to quantitatively determine all the coefficients $h_{j,i}(d)$ describing the causal relation from i to j . The main difference in this class of models is that the unobserved noise components ϵ_j at different

times can affect the variable $x_j(t)$ via the multiplicative coefficients $f_j(d)$, with $d \geq 0$. Adopting some nomenclature from control theory, when the model has the form (II.2) we say that the unobserved components $\epsilon_j(t)$ act as a white noise on the system; instead, when we have generic unknown coefficients $f_j(d)$ (for $d \geq 0$) we say that the noise on the system is auto-correlated. For auto-correlated noise, to the best knowledge of the authors, graphical model methods do not provide a straightforward approach to learn the causal relations $h_{j,i}(d)$, while there are instead systematic techniques that can be borrowed from the area of control theory.

In control theory, the graphical representation of the causal interactions is slightly different. Models of the form (II.3) are typically associated with a graph where each node represents a time series (as opposed to a single random variable) and there is a direct edge from the process i to the process j if at least one of the coefficients $h_{j,i}(d)$ is potentially different from zero [1,66,90]. Observe, that this kind of representation can lead to a graph with directed loops even when the unit causal or the repetitive causal graphs are acyclic. Thus, in other words, this simpler graphical representation might lose the fundamental information that the underlying causal graph is acyclic.

In order to preserve the information of the delayed causal relations, we introduce an extension of the standard notion of graph where there are two sets of edges: E_1 is the set of single-headed edges representing causal relations which might have a potential instantaneous propagation, and E_2 is the set of double-headed edges representing delayed causal relations. We name this kind of graphical representation collapsed causal graph and will simply refer to it as graphical representation of the network. In Figure 2.1(d) we report the graphical representation (collapsed causal graph) of the repetitive causal graph

of Figure 2.1(b). We will formally define these concepts later on. The assumption that the repetitive causal graph or the unit causal graph are acyclic translates into the assumption that there is no loop in the collapsed causal graph involving all instantaneous causal influences.

For a fixed pair (i, j) , the control theoretical approach offers a compact way to represent all the coefficients $h_{j,i}(d)$ and $h_{j,j}(d)$ and learn them from data by making use of the so-called “transfer function” approach [91]. Specifically, if the variables x_i and x_j follow the linear difference equation

$$x_j(t) = \sum_{d \geq 0} h_{j,i}(d)x_i(t-d) - \sum_{d \geq 1} h_{j,j}(d)x_j(t-d),$$

we can define the transfer function from x_i to x_j as

$$H_{ji}(z) = \frac{\sum_{d \geq 0} h_{j,i}(d)z^{-d}}{1 + \sum_{d \geq 1} h_{j,j}(d)z^{-d}},$$

which is a function of the “frequency variable” z^{-1} . The knowledge of the transfer function $H_{ji}(z)$ is equivalent to the knowledge of the whole sequences $h_{j,i}(d)$ and $h_{j,j}(d)$ for all d [92].

2.2 A model class for dynamic networks

In the last decade several network models have emerged, such as Dynamical Structural Function (DSF) models [76], linear dynamic graphs [67] and the dynamic interconnection proposed in [90]. These models share very similar characteristics, the main one be-

ing that they can be interpreted via graphs in a way analogous to Signal Flow Graphs [93]. The class of dynamic networks considered in this dissertation is given by linear dynamic influence models [64] which can be equivalently interpreted as DSF models/signal flow graphs with unknown stochastic inputs, as dynamic Bayesian networks with noise components potentially correlated in time, or as the interconnection model in [90] without measured reference signals. Similar network models have been considered and investigated in [64, 76, 90, 93].

Definition II.1. *A network \mathcal{G} is a pair $(H(z), n)$ where $H(z)$ is a proper rational discrete-time $v \times v$ transfer matrix and n is a vector of v mutually independent stochastic processes with rational power spectral density. The output signals of the network are defined by the relation*

$$x_j(t) = n_j(t) + \sum_{i \in V} H_{ji}(z)x_i(t), \quad \text{for } j = 1, \dots, v \quad (\text{II.4})$$

Using a vector notation and defining $V = \{1, \dots, v\}$ we can represent the model in a more compact way as

$$x_V(t) = n_V(t) + H(z)x_V(t) \quad (\text{II.5})$$

Note that the noise process $x_j(t)$ could be colored. The dynamics of the matrix $H(z)$ in model (II.5) indicates how the process x_i directly affects the process x_j . If $H_{ji}(z) = 0$ there is no direct effect of x_i on x_j (even though x_i could still affect x_j indirectly through other processes). For this reason models described by (II.5) lend themselves to be represented via graphs. We assume that the reader is already familiar with basic notions of

graph theory [55] and in this section we just introduce our notation and nomenclature. For a directed graph G , defined by the pair (V, E) where $V = \{1, 2, \dots, v\}$ is the set of nodes and $E \subseteq V \times V$ is the set of edges, we denote an edge $(i, j) \in E$ as $i \rightarrow j$ or $j \leftarrow i$ and say that the edge is oriented from i to j . We also say that two distinct edges $i \rightarrow j$ and $k \rightarrow \ell$ are adjacent if they share at least a node, namely $\{i, j\} \cap \{k, \ell\} \neq \emptyset$. In a directed graph, a path between i and j is a sequence of distinct edges such that the first edge contains i , the last edge contains j and each two consecutive edges in the sequence are adjacent. A path can be suggestively denoted by using the arrow symbols (\rightarrow and \leftarrow) to separate the nodes involved in the path while at the same time representing the orientation of the edges.

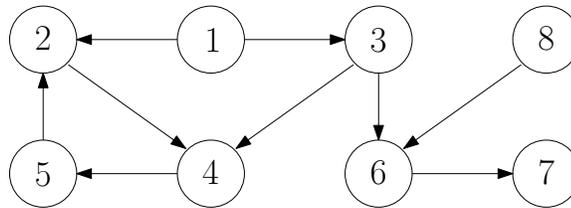


Figure 2.2: Representation of a directed graph.

For example, in the graph of Figure 2.2, there are four paths between nodes 3 and 5 which can be denoted as $\{3 \rightarrow 4 \rightarrow 5\}$, $\{3 \rightarrow 4 \leftarrow 2 \leftarrow 5\}$, $\{3 \leftarrow 1 \rightarrow 2 \rightarrow 4 \rightarrow 5\}$, $\{3 \leftarrow 1 \rightarrow 2 \leftarrow 5\}$. Furthermore, if the edges have all the same orientation (as in $\{1 \rightarrow 3 \rightarrow 4 \rightarrow 5\}$) the path is called a dipath or a chain. For a graph G , we also recall the following relations among its nodes

- node j is a child of node i if the edge $i \rightarrow j$ is present in the graph. We also say that i is a parent of j . We denote the set containing all children of node j by $\text{ch}_G(j) = \{v \in V \mid j \rightarrow v \in E\}$ and the set containing all parents of node i by

$\text{pa}_G(i) = \{v \in V \mid v \rightarrow i \in E\}$. Moreover for a set $A \subseteq V$ we define $\text{ch}_G(A) = \cup_{j \in A} \text{ch}_G(j)$ and $\text{pa}_G(A) = \cup_{i \in A} \text{pa}_G(i)$.

- node j is a descendant of node i if $j = i$ or if there is a dipath from i to j . Equivalently, we say that i is an ancestor of j . We denote the set containing all descendants of node i as $\text{de}_G(i)$ and the set containing all ancestors of node i as $\text{an}_G(i)$.

For the description of model (II.4) we are going to use a special instance of multi-typed graphs [94] which are an extension of standard directed graphs.

Definition II.2. *A multi-arrowed graph is a triple $G = (V, E_1, E_2)$ where E_1 , the set of single-headed edges, and E_2 , the set of double-headed edges, are disjoint subsets of $V \times V$.*

We represent a multi-arrowed graph in the same way we represent a standard graph but we draw a single-headed edge to represent $i \rightarrow j \in E_1$ and a double-headed edge to represent $i \rightarrow j \in E_2$. Note that multi-arrowed graphs generalize directed graphs and all the graphical notions extend naturally to multi-arrowed graphs, as well, by simply considering $E_1 \cup E_2$ as a set of standard edges. For example, vertex i is a parent of j whether there is a single-headed or double-headed edge from i to j . If $i \rightarrow j \in E_1$ we say that i is a single-headed parent of j , while if $i \rightarrow j \in E_2$ we say that i is a double-headed parent of j .

Definition II.3. *We say that the multi-arrowed graph $G = (V, E_1, E_2)$ is recursive if in every directed loop there is at least one double-headed edge.*

We can use multi-arrowed graphs to describe the sparsity pattern of $H(z)$ in model (II.4) along with some information about the strict causality of the entries in $H(z)$.

Definition II.4. *Let $\mathcal{G} = (H, n)$ be a network with output processes x_V , where $V := \{1, \dots, v\}$, and let E_1 and E_2 be two disjoint subsets of $V \times V$ such that*

(a) $i \rightarrow j \notin E_1 \cup E_2$ implies $H_{ji}(z) = 0$

(b) $i \rightarrow j \notin E_1$ implies $H_{ji}(z)$ is strictly causal.

We say that the multi-arrowed graph $G = (V, E_1, E_2)$ is a graphical representation of the network. Furthermore, if the implications (a) and (b) hold also in the opposite direction, we say that $G = (V, E_1, E_2)$ is a perfect graphical representation of the network.

In other words, the absence of the edge $i \rightarrow j$ in a graphical representation implies that $H_{ji}(z) = 0$ while the presence of a double-headed edge $i \rightarrow j$ implies that $H_{ji}(z)$ is strictly causal (potentially zero). Thus, a network can have different graphical representations each providing different degrees of information on its dynamics, as the following example illustrates.

Example 1. Consider a dynamic network $\mathcal{G} = (H(z), n)$ with four nodes governed by the following equations.

$$x_1(t) = n_1(t) + H_{12}(z)x_2(t) + H_{14}(z)x_4(t)$$

$$x_2(t) = n_2(t)$$

$$x_3(t) = n_3(t) + H_{31}(z)x_1(t)$$

$$x_4(t) = n_4(t) + H_{43}(z)x_3(t).$$

The nonzero entries of $H(z)$ in this network are

$$H_{12}(z) = \frac{z}{z + \frac{1}{2}}, \quad H_{14}(z) = \frac{1}{z + \frac{1}{2}}, \quad H_{31}(z) = \frac{1}{z^2}, \quad H_{43}(z) = \frac{1}{2}.$$

Figure 2.3 shows the block diagram of the network \mathcal{G} . Figure 2.4 (a) shows the perfect

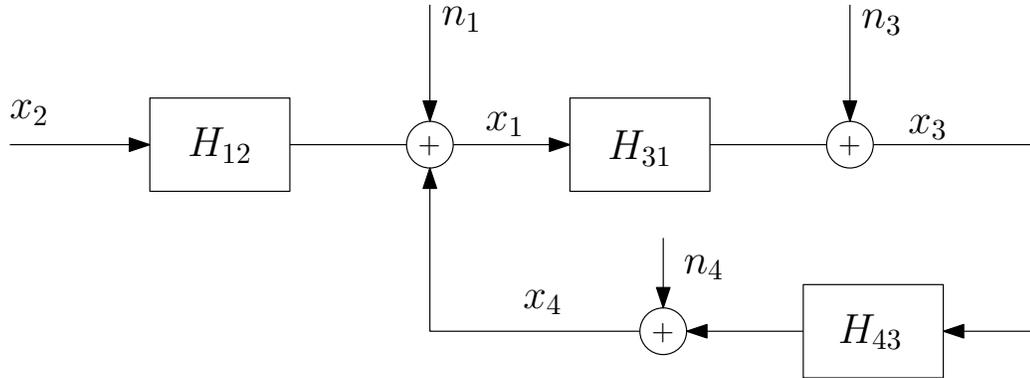


Figure 2.3: Block diagram of the network \mathcal{G} discussed in Example 1.

graphical representation G^p of the network \mathcal{G} . The information that transfer functions $H_{31}(z)$ and $H_{14}(z)$ are strictly causal and $H_{23}(z) = 0$ is available in G^p . Figure 2.4 (b) shows a recursive graphical representation G of the network \mathcal{G} . Unlike G^p , the information that transfer function $H_{14}(z)$ is strictly causal or that $H_{23}(z) = 0$ is not available from G .

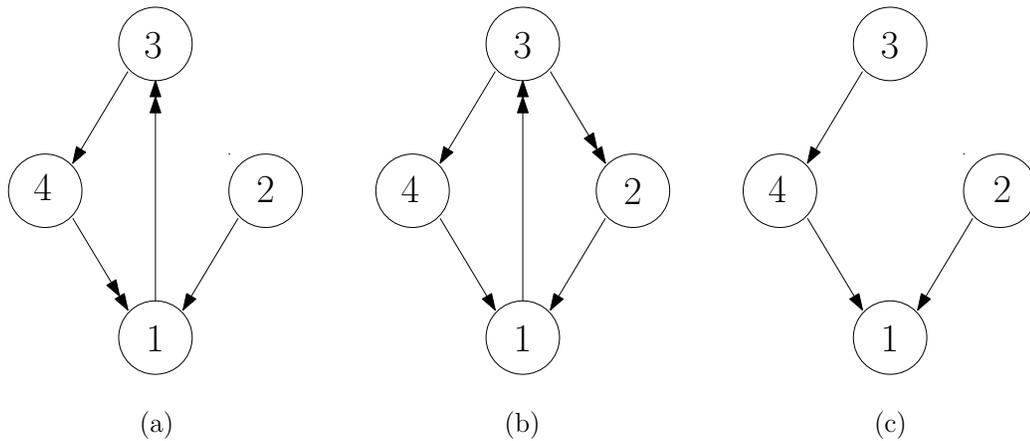


Figure 2.4: (a) The perfect graphical representation G^p of the dynamic network \mathcal{G} discussed in Example 1; (b) A recursive graphical representation G of \mathcal{G} ; (c) The graph of instantaneous propagations G^t associated with G .

Observe that a graphical representation provides partial information about the network's topology. Indeed, given a graphical representation of a network it is always possible to obtain another less informative graphical representation by introducing additional

single-headed or double-headed edges and/or replacing a double-headed edge with a single-headed one.

From its definition if a network has recursive graphical representation then it has no algebraic loops. The following definition of graph of instantaneous propagations is an important tool to deal with the presence of direct feed-throughs.

Definition II.5. *Consider a multi-headed graph $G = (V, E_1, E_2)$. Its associated graph of instantaneous propagations, denoted as G^ζ , is the standard directed graph (V, E_1) obtained from G by removing the double-headed edges.*

For example, Figure 2.4 (c) shows the graph of instantaneous propagations G^ζ associated with graph G depicted in Figure 2.4 (b). It is an immediate consequence of the definition that if G is recursive, G^ζ is a directed acyclic graph. We refer to G^ζ as the graph of instantaneous propagations because, if G is a graphical representation, new information entering a node k at time t can potentially propagate to all nodes in $\text{de}_{G^\zeta}(k)$ at the same time t with no delay.

Throughout the dissertation, we will sometimes refer to nodes, edges, paths and chains of a network even though, formally, we should refer to them as nodes, edges, paths and chains of its graphical representation or its perfect directed graph.

As shown in [64], there is a strong relationship between signal estimators and graphical representations in a network. Such a relationship will play a central role in the development of our results. For this reason we recall some fundamental notions from estimation theory and introduce our notation.

Definition II.6. *Given a probability space, for a set of stochastic processes x_A where $A \subseteq V$, we denote the natural filtration generated by the processes x_A up to time t as $I_A(t)$.*

In this dissertation we typically consider the estimate $\hat{x}_j(t)$ of $x_j(t)$ using the information of processes x_{D^+} up to time t and the information of processes x_{D^-} up to time $t - 1$. Using the notation introduced in Definition II.6 the least square estimate $\hat{x}_j(t)$ can be written as

$$\hat{x}_j(t) = \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)). \quad (\text{II.6})$$

In the linear Gaussian case this estimation problem can be solved via Wiener filters, reducing (II.6) to

$$\hat{x}_j = \sum_{k \in D^+} W_{jk}(z)x_k + \sum_{k \in D^-} W_{jk}(z)x_k \quad (\text{II.7})$$

where $W_{jk}(z)$ for $k \in D^+$ are causal transfer functions and for $k \in D^-$ are strictly causal transfer functions. So long as the power spectral density matrix of the signals x_j, x_{D^+} and x_{D^-} is the same, the expressions of the Wiener filter components are the same when considering a least square estimation even in the linear non-Gaussian case. In the following, we assume for simplicity that all the processes are jointly Gaussian even though the same results can be easily shown to hold in the linear non-Gaussian case, as well.

The sparsity properties of the Wiener filters $W_{jk}(z)$ are connected to the notion of conditional independence.

Definition II.7. *We say that $x_j(t)$ and the information of x_i up to time t are conditionally independent given $I_{D^+}(t)$ and $I_{D^-}(t-1)$ if*

$$\mathbb{E}(x_j(t) \mid I_{\{i\} \cup D^+}(t), I_{D^-}(t-1)) = \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)). \quad (\text{II.8})$$

We denote this by $x_j(t) \perp I_i(t) \mid I_{D^+}(t), I_{D^-}(t-1)$. Similarly, if $\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{\{i\} \cup D^-}(t-1)) = \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1))$ we say that $x_j(t)$ and the information of x_i up to time $t-1$ are conditionally independent which we denote by $x_j(t) \perp I_i(t-1) \mid I_{D^+}(t), I_{D^-}(t-1)$.

In the linear Gaussian case, using a Wiener filter formulation, the estimate $\hat{x}_j(t)$ of $x_j(t)$ from the processes x_{D^+} up to time t , the processes x_{D^-} up to time $t-1$ and the process x_i can be expressed as

$$\hat{x}_j = W_{ji}(z)x_i + \sum_{k \in D^+} W_{jk}(z)x_k + \sum_{k \in D^-} W_{jk}(z)x_k \quad (\text{II.9})$$

where $W_{jk}(z)$ are causal for $k \in D^+$, strictly causal for $k \in D^-$ and $W_{ji}(z)$ is causal if the information of x_i is used up to time t and strictly causal if the information of x_i is used up to time $t-1$. The relation of conditional independence between $x_j(t)$ with $I_i(t)$ (or analogously $I_i(t-1)$) translates into having $W_{ji}(z) = 0$ in Equation (II.9).

2.3 j -pointing separation

In the theory of graphical models the internal nodes of a path are classified as forks, colliders or chain links.

Definition II.8. Given a path π in a graph G we say that a node j is

- a fork, when there exist two consecutive edges in the path of the form $i \leftarrow j$ and $j \rightarrow k$
- a collider (or an inverted fork), when there exist two consecutive edges in the path of the form $i \rightarrow j$ and $j \leftarrow k$

- a chain link, when there exist two consecutive edges in the path of the form $i \rightarrow j$ and $j \rightarrow k$

Specifically, the notion of colliders allows one to define if a path π is blocked by a set Z .

Definition II.9. In a directed graph G , a path π between nodes i and j is blocked by a set of nodes Z if

- there is at least a non-collider on π that belongs to Z ; or
- there is at least a collider c on π such that $de_G(c) \cap Z = \emptyset$.

Otherwise, we say that the path π is activated by Z .

In the theory of graphical models, a fundamental concept defined over the nodes of a directed graph is d -separation [50].

Definition II.10. In a directed graph $G = (V, E)$ let A , B , and C be disjoint subsets of V . A and B are d -separated by C if for all nodes $a \in A$ and $b \in B$, all paths between a and b are blocked by C . If A and B are not d -separated by C in G , we say that they are d -connected by C in G .

Example 2. In the directed graph depicted in Fig. 2.2 $A = \{1\}$ and $B = \{8\}$ are d -separated by $C = \emptyset$ because 6 is a collider on a path from 1 to 8. For the same reason, $A = \{1\}$ and $B = \{8\}$ become d -connected if we choose $C = \{6\}$ and also if we choose $C = \{7\}$ or $C = \{6, 7\}$. Again, $A = \{2\}$ and $B = \{6\}$ are d -connected by $C = \emptyset$ because of the path $2 \leftarrow 1 \rightarrow 3 \rightarrow 6$. If we consider $C = \{1\}$ to “block” such a path, $A = \{2\}$ and $B = \{6\}$ are still d -connected because of the other path $2 \leftarrow 5 \leftarrow 4 \leftarrow 3 \rightarrow 6$. If we now consider $C = \{1, 4\}$ to “block” this other path, $A = \{2\}$ and $B = \{6\}$ are still d -connected because now 4 is a collider in C on the path $2 \rightarrow 4 \leftarrow 3 \rightarrow 6$. By choosing

$C = \{1, 3, 4\}$, we make $A = \{2\}$ and $B = \{6\}$ d -separated. Alternatively, $C = \{3\}$ would have been a smaller set making $A = \{2\}$ and $B = \{6\}$ d -separated.

In [64], some criteria for consistent identification are derived using the notion of d -separation. This dissertation obtains more powerful criteria by using a weaker notion of separation that involves only a subset of the paths between the nodes i and j .

Definition II.11. A path π between nodes i and j is called j -pointing if the last edge in the path π is of the form $k \rightarrow j$ for some node k . If all the j -pointing paths between nodes i and j , with the exception of the path constituted by only the edge $i \rightarrow j$, are blocked by a set of nodes Z , we say that i and j are j -pointing separated.

Note that a j -pointing path between i and j might or might not be i -pointing.

2.4 Single door criterion

In this section we explain the single-door criterion which helps to identify single path coefficients in structural equation models (SEMs) [56]. Consider a graph G corresponding to a standard linear structural equations of the form

$$x_j = \sum_{j \neq i} h_{ji} x_i + \epsilon_j \quad j = 1, \dots, v \quad (\text{II.10})$$

where h_{ji} are scalars and ϵ_j represent errors due to omitted factors normally distributed.

The single door criterion can be stated as follows. Let G be any path diagram in which h_{ji} is the path coefficient associated with link $i \rightarrow j$ and let G' denote the diagram that results when $i \rightarrow j$ is deleted from G . The coefficient h_{ji} is identifiable if there exists a set of variables Z such that (i) Z contains no descendant of j and (ii) Z d -separates i from

j in G' . If Z satisfies these two conditions, then h_{ji} is equal to the regression coefficient $r_{ji,Z}$ (the coefficient of x_i in the linear regression of x_j on x_i and x_Z). Conversely, if Z does not satisfy these conditions, then $r_{ji,Z}$ is not a consistent estimand of (except in rare instances of measure zero).

Example 3. Consider a linear SEM described by II.10 with a path diagram G depicted in Figure 2.5. Suppose the goal is to identify the coefficient h_{32} . The mutilated path diagram G'

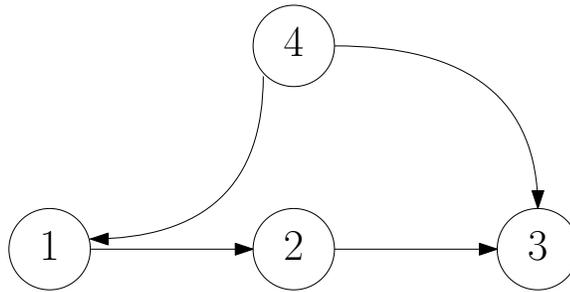


Figure 2.5: Path diagram G of a SEM discussed in Example 3.

obtained after removing the path $2 \rightarrow 3$ from G is depicted in Figure 2.6. The set $Z = \{1\}$

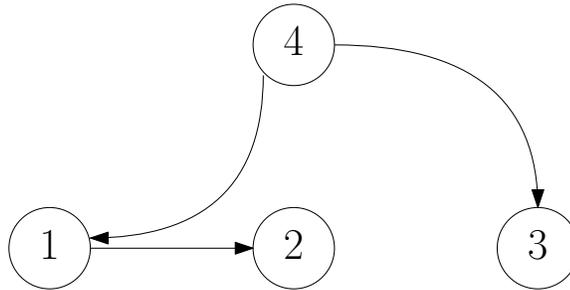


Figure 2.6: The mutilated path diagram G' obtained after removing the path $2 \rightarrow 3$ from G discussed in Example 3.

d -separates nodes 2 and 3 in G' . Therefore, by the single door criterion we have that $r_{32,1}$ is a consistent estimate of h_{32} . Similarly, set $Z = \{4\}$ also d -separates nodes 2 and 3 in G' . Therefore, by the single door criterion we have that $r_{32,4}$ is a consistent estimate of h_{32} .

In [64] an extension of single door criterion is used to identify individual transfer functions in a dynamic network which we will discuss in more detail later on.

2.5 Flow networks and max-flow min-cut theorem

In optimization theory, maximum flow problems involve finding a feasible flow through a flow network that obtains the maximum possible flow rate. Formally we define a flow network as follows.

Definition II.12. A flow network $(G = (V, E), s, t, c)$ is a directed graph G involving a source node $s \in V$ and a sink node $t \in V$ with a capacity function $c : E \rightarrow R^+$ which associates to each edge $u \rightarrow v \in E$ its maximum capacity $c_{u \rightarrow v} \in R^+$.

In many applications it is of interest to determine the cut in a flow network with minimum capacity as defined in the following.

Definition II.13. A cut in a flow network $(G = (V, E), s, t, c)$, is partition $(S'_m, V \setminus S'_m)$ of the set of vertices V into two sets, such that $s \in S'_m$ and $t \in V \setminus S'_m$. We will usually identify a cut with the set of nodes S'_m that contains s or the set of edges $S_m := \{u \rightarrow v \mid u \in S'_m, v \notin S'_m\}$. The capacity of a cut is the quantity

$$\text{capacity}(S_m) = \sum_{(u \rightarrow v) \in S_m} c_{u \rightarrow v} \quad (\text{II.11})$$

The problem of finding a cut with minimum capacity (min cut problem) can be formulated as a linear program. The dual problem is often referred to as max flow problem. See [95] for an extensive discussion about min cut and max flow problems.

CHAPTER III

CONSISTENT IDENTIFICATION IN DYNAMIC NETWORKS

A standard assumption when dealing with the problem of identifying a module in a dynamic network is the absence of algebraic loops. Furthermore, most identification methods also need to include among their assumptions some a priori information about the location of the strictly causal transfer functions in each loop [60, 63, 65, 70, 77]. In this dissertation we still keep the assumption that the network has no algebraic loops, but, as an important distinction from other methods, we also reduce the need of a priori information about the locations of strictly causal transfer functions. Indeed, as we later show in Chapter IV, we provide some methods to infer the locations of strictly causal transfer functions directly from data. In the derivation of the result for this section, we temporarily assume that such information is obtained and available in the form of a recursive graphical representation.

3.1 MISO prediction error method

Consider a simple two-node system with x_i as input, x_j as output, and n_j as additive output error independent of x_i , namely

$$x_j(t) = n_j(t) + H_{ji}(z)x_i(t),$$

for some causal transfer function $H_{ji}(z)$. The block diagram of this system is depicted in Figure 3.1 (a), while a graphical representation of the system is shown in Figure 3.1 (b).

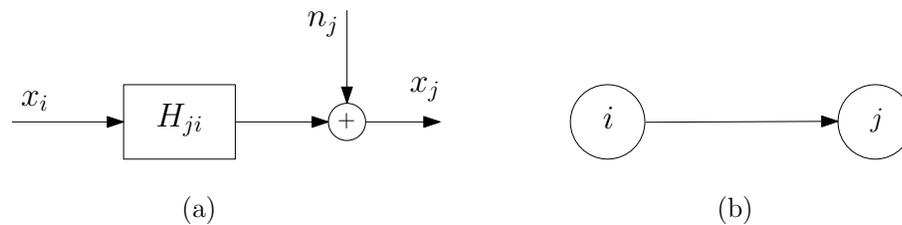


Figure 3.1: (a) Block diagram of a two-node network; (b) Corresponding graphical representation.

Following [96], a possible technique to identify $H_{ji}(z)$ is to compute a linear least square prediction $\hat{x}_j(t)$ for $x_j(t)$ by using the past information of x_j and the past and present information of x_i . Namely, we compute

$$\hat{x}_j(t) = \mathbb{E}[x_j(t) \mid I_i(t), I_j(t-1)] = W_{jj}(z)x_j(t) + W_{ji}(z)x_i(t),$$

where $W_{jj}(z)$ is strictly causal and $W_{ji}(z)$ is causal. After computing $\hat{x}_j(t)$, we calculate the quantity

$$\hat{H}_{ji}(z) = \frac{W_{ji}(z)}{1 - W_{jj}(z)}$$

which can be proven to be a consistent estimate for $H_{ji}(z)$. In other words, in this simple two-node system, the transfer function $H_{ji}(z)$ can be consistently identified via the following procedure by setting $D^- = \{j\}$ and $D^+ = \{i\}$.

Algorithm 1 Identification via prediction error

- 1: **Given:** Sets of nodes D^- , D^+ and $i, j \in D^- \cup D^+$
- 2: **Output:** $\hat{H}_{ji}(z)$
- 3:

$$\mathbb{E}(x_j(t) \mid I_{D^-}(t-1), I_{D^+}(t)) = \sum_{k \in D^- \cup D^+} W_{jk}(z)x_k(t)$$

- 4: $\hat{H}_{ji}(z) = \frac{W_{ji}(z)}{1 - W_{jj}(z)}$
-

We stress that Procedure 1 is not an algorithm but more precisely a meta-algorithm since the computation of $\hat{x}_j(t)$ could be obtained using a variety of methods either parametric or non-parametric [96]. In Procedure 1, the estimate of $x_j(t)$ is obtained using the information from the past for the variables in D^- and information from the past and present for the variables in D^+ . In other words, the transfer functions $W_{jk}(z)$ are strictly causal if $k \in D^-$ and causal if $k \in D^+$.

However, when dealing with more complex networks, applying Procedure 1 with $D^- = \{j\}$ and $D^+ = \{i\}$, leads, in general, to an estimate $\hat{H}_{ji}(z)$ for $H_{ji}(z)$ which is

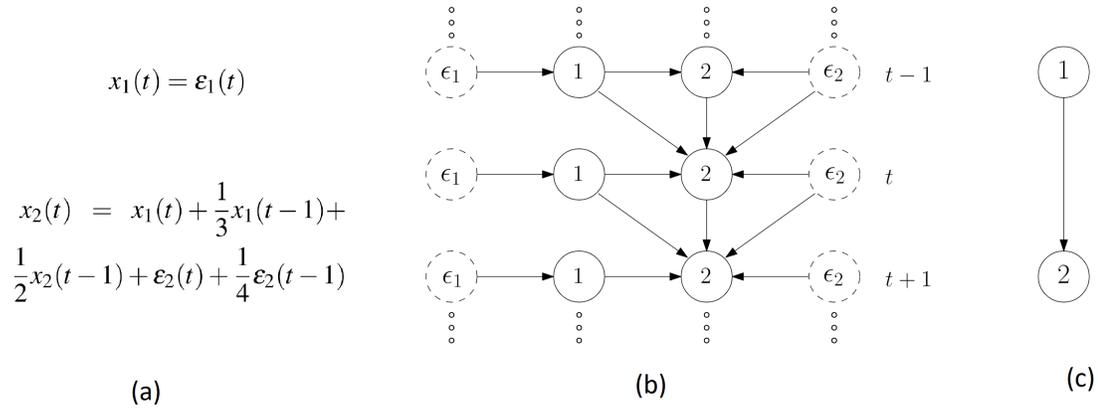


Figure 3.2: (a) Equations governing a two-variate stochastic processes (b) The repetitive causal graph including hidden noise processes (c) The graphical representation (collapsed causal graph).

not consistent because of the presence of feedback loops or because other variables in the network might act as confounders between i and j .

Example 4. *As a minimalistic example of the transfer function approach, consider the two-variable network governed by the equations given in Figure 3.2 (a). Note that $x_2(t)$ is a function of $x_2(t-1)$ and the noise affecting $x_2(t)$ is auto-correlated ($\epsilon_2(t-1)$ is acting as a confounder on $x_2(t)$ and $x_2(t-1)$). Suppose the objective is to estimate the unknown coefficients $h_{2,1}(0) = 1$, $h_{2,1}(1) = 1/3$, and $h_{2,2}(1) = 1/2$. Because of the auto-correlated noise component, criteria like the single door criterion do not provide a straightforward and systematic approach. Instead, following Algorithm 1 with $Z = \emptyset$, $i = 1$, $j = 2$, $D^+ = \{1\}$ and $D^- = \{2\}$, we compute, at Step 4, a linear least square estimate of $x_2(t)$ from the past and present of x_1 and past of x_2 . In the limit of infinite data, such linear least square estimators can be shown to have transfer functions, computed at Step 5 of Algorithm 1, that converge to*

$$W_{21}(z) = \frac{1 + \frac{1}{3}z^{-1}}{1 + \frac{1}{4}z^{-1}} \quad \text{and} \quad W_{22}(z) = \frac{\frac{3}{4}z^{-1}}{1 + \frac{1}{4}z^{-1}}. \quad (\text{III.1})$$

Consequently, Step 6 in Algorithm 1 gives

$$\hat{H}_{21}(z) = \frac{W_{21}(z)}{1 - W_{22}(z)} = \frac{\frac{1 + \frac{1}{3}z^{-1}}{1 + \frac{1}{4}z^{-1}}}{1 - \frac{\frac{3}{4}z^{-1}}{1 + \frac{1}{4}z^{-1}}} = \frac{1 + \frac{1}{3}z^{-1}}{1 - \frac{1}{2}z^{-1}}. \quad (\text{III.2})$$

Hence, we conclude that the following relation holds

$$x_2(t) - \sum_{d \geq 0} f_2(d) \epsilon_j(t-d) = x_1(t) + \frac{1}{3}x_1(t-1) + \frac{1}{2}x_2(t-1), \quad (\text{III.3})$$

for undetermined values of $f_2(d)$, which gives $\hat{h}_{2,1}(0) = 1$, $\hat{h}_{2,1}(1) = 1/3$, and $\hat{h}_{2,2}(1) = 1/2$ which are consistent with their true values.

To verify the estimations numerically, we generated two processes x_1 and x_2 governed by equations in Figure 3.2 (a) for $N = 1000$ number of measurements and used Algorithm 1 to learn $H_{21}(z)$ and consequently estimated coefficients $h_{2,1}(0)$, $h_{2,1}(1)$, and $h_{2,2}(1)$. We repeated this simulation 100 times. In our implementation, we used the System Identification toolbox of Matlab. For $h_{2,1}(0)$ with the unknown true value of 1, the mean of estimations over the 100 runs was 0.992 with the standard deviation of 0.031. For $h_{2,1}(1)$ with the unknown true value of $\frac{1}{3}$, the mean of estimations over the 100 runs was 0.3336 with the standard deviation of 0.359. For $h_{2,2}(1)$ with the unknown true value of $\frac{1}{2}$, the mean of estimations over the 100 runs was 0.5003 with the standard deviation of 0.027. File `secA1fig2` in supplementary material contains the Matlab code of these simulations. As we will explore in the next example, the quality of estimations increases with the number of measurements.

Example 5. Consider a simple network with a block diagram depicted in Figure 3.3 and a graphical representation depicted in Figure 3.4.

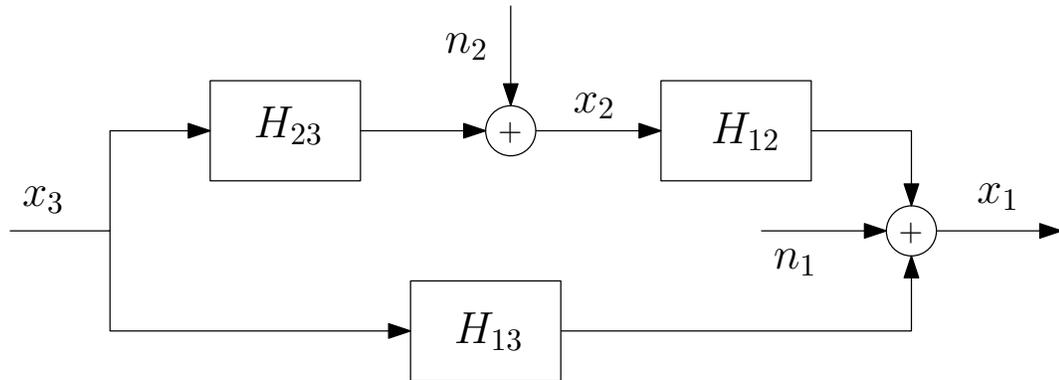


Figure 3.3: Block diagram of a simple three-node network containing a 1-pointing path ($2 \leftarrow 3 \rightarrow 1$) where the objective is the identification of transfer function $H_{12}(z)$.

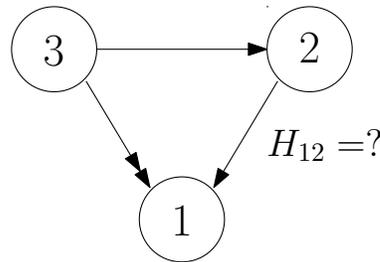


Figure 3.4: Graphical representation of the simple three-node network containing a 1-pointing path ($2 \leftarrow 3 \rightarrow 1$) where the objective is the identification of transfer function $H_{12}(z)$.

Suppose the objective is identification of $H_{12}(z)$. If we apply Algorithm 1 with $D^+ = \{2\}$ and $D^- = \{1\}$ the estimate $\hat{H}_{12}(z)$ computed in line 4 of Algorithm 1 will not be a consistent estimate of $H_{12}(z)$, in general. We will see later on that by including node 3 in D^- Algorithm 1 will result in a consistent estimate of $H_{12}(z)$.

Similarly, consider a network with a block diagram depicted in Figure 3.5 and a graphical representation depicted in Figure 3.6.

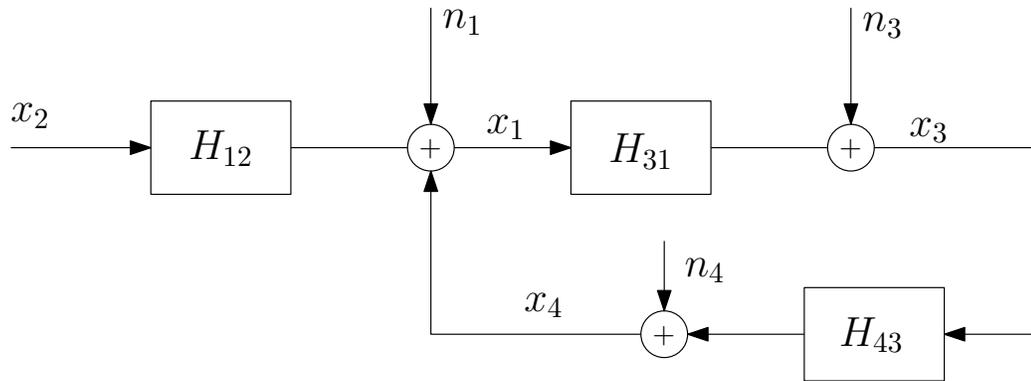


Figure 3.5: Block diagram of a simple four-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$.

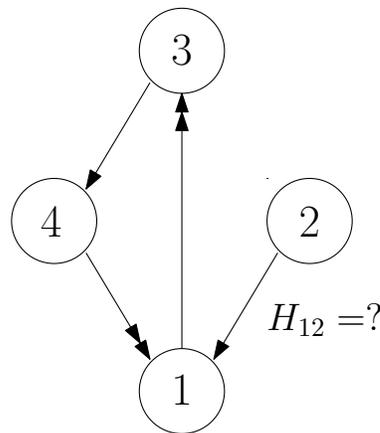


Figure 3.6: Graphical representation of a simple four-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$.

Suppose the objective is identification of $H_{12}(z)$. If we apply Algorithm 1 with $D^+ = \{2\}$ and $D^- = \{1\}$ the estimate $\hat{H}_{12}(z)$ computed in line 4 of Algorithm 1 will not be a consistent estimate of $H_{12}(z)$, in general. We will see later on that by including either node 3 or node 4 in D^- Algorithm 1 will result in a consistent estimate of $H_{12}(z)$.

Conversely, consider a network with a block diagram depicted in Figure 3.7 graphical representation depicted in Figure 3.8.

Suppose the objective is identification of $H_{12}(z)$. If we apply Algorithm 1 with $D^+ = \{2\}$ and $D^- = \{1\}$ the estimate $\hat{H}_{12}(z)$ computed in line 4 of Algorithm 1 will

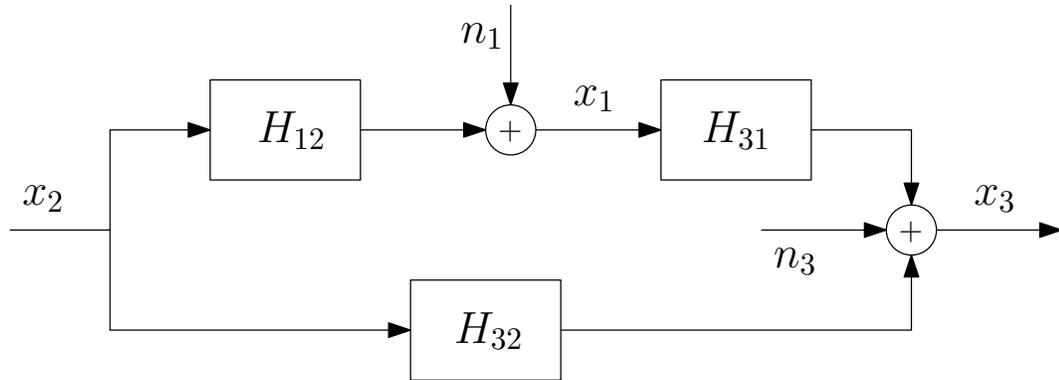


Figure 3.7: Block diagram of a simple three-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$.

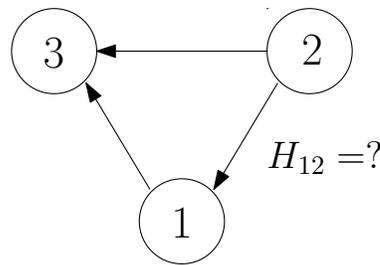


Figure 3.8: Graphical representation of a simple three-node network containing a feedback loop where the objective is the identification of transfer function $H_{12}(z)$.

be a consistent estimate of $H_{12}(z)$, in general. However by including node 3 in D^+ Algorithm 1 will result in a biased estimate of $H_{12}(z)$.

3.2 State of the art in literature

In several recent results, it has been shown that, by appropriately introducing additional measured variables to the sets of predictors D^- and D^+ , Procedure 1 (or substantially equivalent tools such as the methods in [77]) can still achieve a consistent estimate of $H_{ji}(z)$.

This idea has been explored in the extension of closed-loop identification techniques to network identification [1, 62] and by applying graphical model tools [64]. The drawback

of using graphical model techniques is that they tend to be limited to acyclic networks. In particular, the related results in [64] are not as powerful when the target node j is involved in a directed feedback loop. On the hand, techniques base on closed loop identification have difficulties handling confounders.

In particular, in [62] sufficient conditions for identification of the transfer function $H_{ji}(z)$ are provided under the restrictive assumption that there is no confounding variables. Namely, it is required that a set of estimating processes blocks all the parallel paths between nodes i and j and all the feedback loops involving node j .

The results are extended in [1] to handle presence of confounding variables. Since [1] is closely related to our work and we believe it presents the state of the art of network identification, we will discuss its results in more details and show its weaknesses. For identification of the transfer function $H_{ji}(z)$, a set of estimating processes \mathcal{A}_j is required to block all the parallel paths between nodes i and j and all the feedback loops involving node j . For a choice of \mathcal{A}_j , a set of confounding variables $\mathcal{C}(j, \mathcal{A}_j)$ is defined that contains the variables that act as confounders between j and elements of \mathcal{A}_j . Then, another set of estimating processes \mathcal{B}_j is required to block all the paths between any confounder c in $\mathcal{C}(j, \mathcal{A}_j)$ and j or all the paths between c and $k \in \mathcal{A}_j$ and not create a sequence of confounders.

We point out that from a theoretical standpoint, the graphical conditions that we will derive in this dissertation are necessary and sufficient while the conditions in [1] are stated to be sufficient only.

Indeed, we also believe that the conditions in [1] are not precise and found an example where they seem to incorrectly give a biased estimate. Consider a network with a graphical

representation shown in Figure 3.9 where the transfer functions are indicated next to the edges and the external noises are jointly independent with power spectral density equal to identity. Suppose the objective is identification of $H_{21}(z)$. According to our understanding

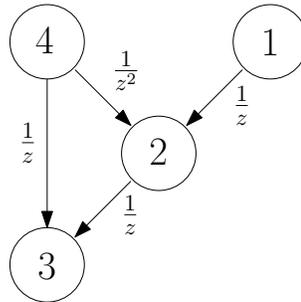


Figure 3.9: A simple network highlighting the flaw in [1]

of [1], $\mathcal{A}_2 = \{1\}$ and $\mathcal{B}_2 = \{3\}$ satisfy the required conditions proposed in [1]. Namely, $\mathcal{A}_2 = \{1\}$ satisfies Property 1 in [1] because there is no parallel path between $i = 1$ and the target node $j = 2$ and there is no feedback involving node 2. Also, $\mathcal{B}_2 = \{3\}$ satisfies Property 3 in [1] since $\mathcal{C}(2, \mathcal{A}_2) = \emptyset$. However, this choice of \mathcal{A}_2 and \mathcal{B}_2 leads to a biased estimation.

Another closely related work is [64] which provides an extension of single door criterion that is used to identify individual transfer functions in a dynamic network : Let G be a graphical representation of a network G described by (II.4) such that node j is not involved in a self-loop. Let G' be the mutilated graph obtained by removing the link $i \rightarrow j$ from the graph G . If i and j are d -separated by Z in G' and Z has no descendent of j , then W_{ji} in

$$\mathbb{E}(x_j(t) | I_{\{i\} \cup Z}(t)) = \sum_{r \in \{i\} \cup Z} W_{jr}(z)x_r(t). \quad (\text{III.4})$$

is a consistent estimate of $H_{ji}(z)$.

Example 6. Consider a network with a graphical representation depicted in Figure 3.11.

It is assumed that node 4 is not measured. The goal is to identify transfer functions $H_{21}(z)$

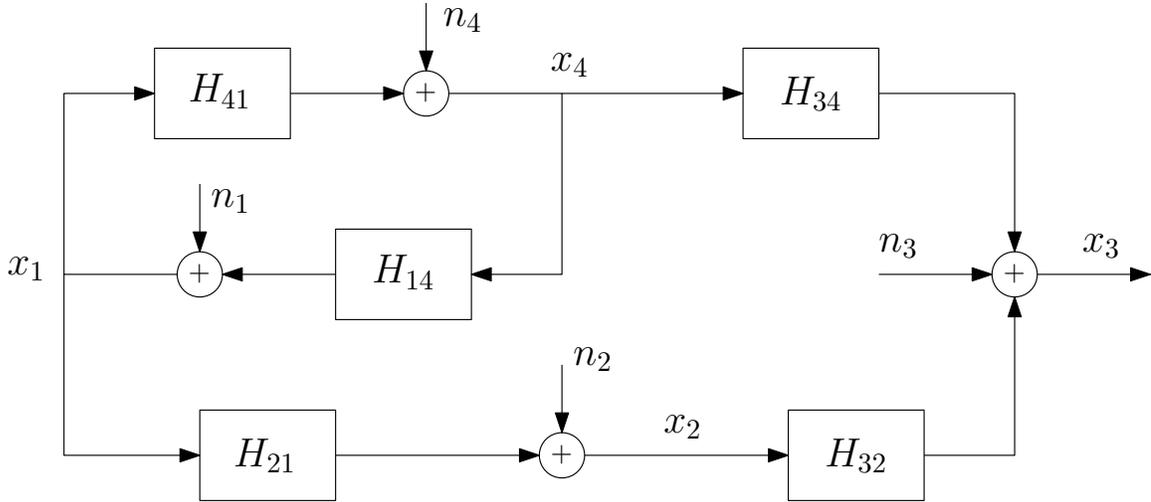


Figure 3.10: Block diagram of a network discussed in Example 6.

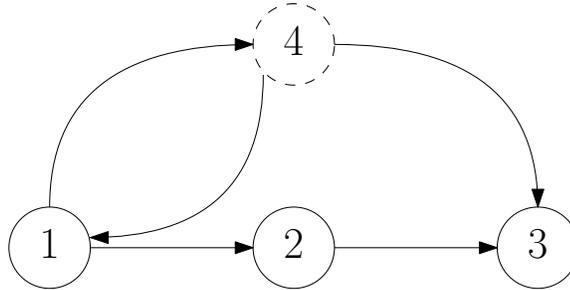


Figure 3.11: Graphical representation G of a network with a confounding variable node 4 discussed in Example 6..

and $H_{32}(z)$. Let G'_1 depicted in Figure 3.12 be the graph obtained by removing the edge $1 \rightarrow 2$ from G . Observe that nodes 1 and 2 are d -separated given the empty set in G'_2 . Thus $H_{21}(z)$ can be identified by the result above as the corresponding transfer function when estimating $x_2(t)$ using $x_1(t)$. Now let G'_2 depicted in Figure 3.13 be the graph obtained by removing the edge $1 \rightarrow 2$ from G . Observe that nodes 1 and 2 are d -separated given the

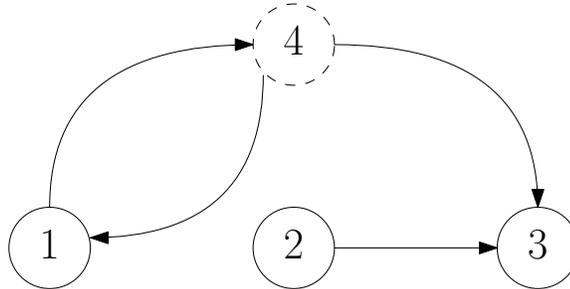


Figure 3.12: The mutilated graph G'_1 obtained after removing the edge $1 \rightarrow 2$ from G in Figure 3.11.

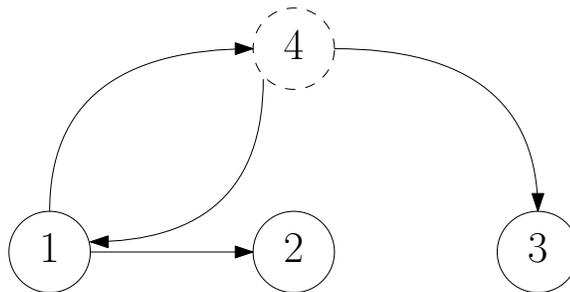


Figure 3.13: The mutilated graph G'_2 obtained after removing the edge $2 \rightarrow 3$ from G in Figure 3.11.

set $Z = \{3\}$ in G'_2 . Thus $H_{32}(z)$ can be identified by the result above as the corresponding transfer function when estimating $x_3(t)$ using $x_1(t)$ and $x_2(t)$.

As can be seen, while closed loop identification techniques have difficulties dealing with confounders, graphical models techniques are restricted to networks where the target node is not involved in feedback loops.

3.3 Sufficient conditions for consistent identification

All the methodologies discussed in the previous section basically try to solve or are applicable to the following problem [1, 60, 62–64].

Problem 1. Consider a network $\mathcal{G} = (H(z), n)$ with a known graphical representation $G = (V, E_1, E_2)$. Suppose that the forcing inputs n are unknown and that a subset $O \subseteq V$

of the node outputs is observable with $i, j \in O$. Find sets of predictors D^- and D^+ with $\{i, j\} \subseteq D^- \cup D^+ \subseteq O$ such that Procedure 1 guarantees a consistent identification of the transfer function $H_{ji}(z)$.

A unifying feature of most of these approaches is to exploit additional measurements (apart from i and j) as auxiliary predictors. More specifically, apart from the nodes i and j , these methods require an extra set of nodes Z to be observable and a way of partitioning $Z \cup \{i, j\}$ into the sets $D^- \cup D^+$ for Procedure 1 to consistently estimate $H_{ji}(z)$.

A first contribution of this dissertation is a solution to Problem 1 that can be interpreted as an attempt to combine graphical model and closed loop identification methods in order to effectively deal with confounding variables and feedback loops. Namely, we provide conditions, of purely graphical nature, to determine the set Z of auxiliary predictors along with a way to partition $Z \cup \{i, j\}$ into the sets D^- and D^+ in order to guarantee that Procedure 1 obtains a consistent identification.

Furthermore, we prove that such graphical conditions on Z are also necessary for consistency given the known graphical representation G . Having sufficient and necessary conditions for the set of auxiliary predictors enables the search for an optimal Z that provides a consistent identification while at the same time minimizing an assigned cost function to select the auxiliary variables.

As a first observation, if all the parents of the target node j are available, then it is possible to consistently identify the transfer function $H_{ji}(z)$ when some knowledge about the delays of the transfer functions $H_{jk}(z)$, $k \in \text{pa}_G(j)$ is available from a recursive graphical representation G of the network.

Proposition 1. Consider a network \mathcal{G} with no algebraic loops. Let $G = (V, E_1, E_2)$ be a recursive graphical representation of \mathcal{G} and P^+ and P^- be respectively the sets of single-headed parents and double-headed parents of the node j in G .

Then

$$\mathbb{E}(x_j(t) | I_{\{j\} \cup P^-}(t-1), I_{P^+}(t)) = W_{jj}(z)x_j(t) + \sum_{k \in P^- \cup P^+} [1 - W_{jj}(z)]H_{jk}(z)x_k(t). \quad (\text{III.5})$$

Proof. See the appendix. □

A consequence of Proposition 1 is the following corollary.

Corollary 1.1. Under the assumptions of Proposition 1, let the power spectral density matrix of x_i , x_j and $x_{P^+ \cup P^-}$ be non-singular. The application of Procedure 1 with $D^- = P^- \cup \{j\}$ and $D^+ = P^+$ leads to a consistent estimate of $H_{ji}(z)$.

Proof. See the appendix. □

Corollary 1.1 is quite intuitive and can be interpreted in terms of the results in [62]. We use Corollary 1.1 as a starting point for the derivation of our results and also to illustrate how simple manipulations on a graphical representation can lead to different selections of the sets D^+ and D^- in Procedure 1, as done in the following example.

Example 7. Consider a network with a block diagram depicted in Figure 3.14 and a graphical representation G depicted in Figure 3.15 (a). The objective is the identification of the transfer function $H_{21}(z)$. Since G is recursive and all the parents of node 2 are measured, we can apply Corollary 1.1. Node 1 is a single-headed parent of node 2, and node 4 is a

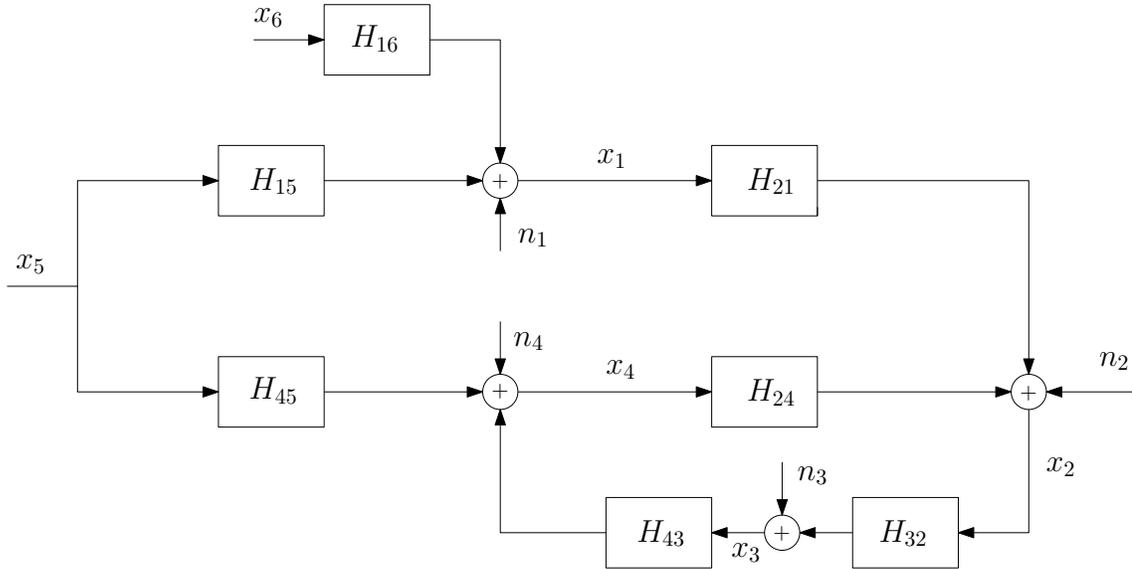


Figure 3.14: The block diagram of the network discussed in Example 7.

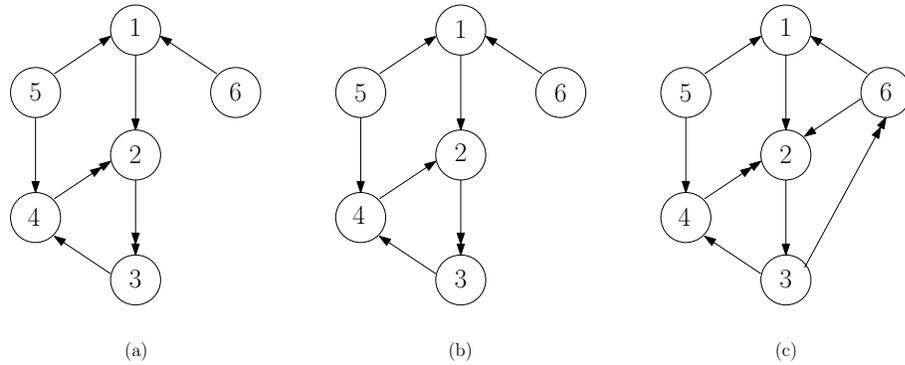


Figure 3.15: The Graphical representations of the networks discussed in Example 7.

double-headed parent of node 2. Therefore, the application of Corollary 1.1 for $j = 2$ and $i = 1$ in the graphical representation G , leads to the choice of $D^- = \{4, 2\}$ and $D^+ = \{1\}$ in Procedure 1 which gives a consistent estimate of $H_{21}(z)$. However, if the graph G of Figure 3.15 (a) is a graphical representation of the network under study, so must be the graph G' depicted in Figure 3.15(b). The difference between the two graphical representations is that the information that $H_{24}(z)$ is strictly causal is available in G , but is not available in G' . Since G' is also recursive, we can still apply Corollary 1.1 to it. When

applied to G' , Corollary 1.1 leads to a different choice for the sets D^- and D^+ (namely $D^- = \{2\}$ and $D^+ = \{1, 4\}$) which also provides a consistent estimate of $H_{21}(z)$ via Procedure 1. Furthermore, if the graph G of Figure 3.15 (a) is a graphical representation of the network under study, so must be the graph G'' depicted in Figure 3.15(c). There are a few differences between G and G'' . It can be seen from G that $H_{32}(z)$ is strictly causal, and $H_{26}(z) = H_{63}(z) = 0$. This information is not available in G'' . Since G'' is also recursive, we can still apply Corollary 1.1 to it. When applied to G'' , Corollary 1.1 leads to yet a different choice for the sets D^- and D^+ (namely $D^- = \{2, 4\}$ and $D^+ = \{1, 6\}$) which also provides a consistent estimate of $H_{21}(z)$ via Procedure 1.

This example shows that if a recursive graphical representation G is available, we can still apply Corollary 1.1 to a less informative graphical representation which can be obtained by introducing additional edges in G or by replacing double-headed edges in G with single-headed edges. The only requirement for the application of Corollary 1.1 is that such a less informative graphical representation has still to be recursive.

Observe that in Proposition 1 we have $D^- \cup D^+ = \text{pa}_G(j) \cup \{j\}$. Hence, Proposition 1 substantially states that Procedure 1 can consistently identify the transfer function $H_{ji}(z)$, but, in order to do so, it requires the observation of all parents of j in a given graphical representation. In some scenarios, assuming that all parents of j are being observed might be overly restrictive, since missing information from some parents of the target node j does not necessarily hinder the consistent identification of $H_{ji}(z)$. Indeed, information from other observed nodes can be exploited to compensate the missing information from the unmeasured parents, so that a consistent identification of $H_{ji}(z)$ can still be achieved. The first main contribution of this dissertation is the following result providing a criterion to

appropriately select the sets D^- and D^+ in Procedure 1 to guarantee an unbiased estimate of $H_{ji}(z)$ under significantly more general conditions than Corollary 1.1.

Theorem 2. Consider a network $\mathcal{G} = (H(z), n)$ with recursive graphical representation $G = (V, E_1, E_2)$. Let $Z \cap \{i, j\} = \emptyset$ be a set such that

- (i) Z is j -pointing separating the nodes i and j in G ; and
- (ii) $Z \cup \{i\}$ blocks all j -pointing paths from j to itself G .

Let G^t be the graph of instantaneous propagations associated to G and let D^- and D^+ be the following two disjoint sets partitioning $Z \cup \{i, j\}$

- $D^+ := an_{G^t}(j) \cap (Z \cup \{i\})$
- $D^- := (Z \cup \{i, j\}) \setminus D^+$

The application of Procedure 1 with D^- and D^+ leads to a consistent estimate of $H_{ji}(z)$ when the power spectral density matrix of (x_i, x_j, x_Z) is non-singular.

Proof. See the appendix. □

Theorem 2 presents a systematic procedure for selecting two sets of predictors D^- and D^+ to identify a specific transfer function $H_{ji}(z)$ via Procedure 1. Observe that the fact that the graphical representation is recursive allows one to determine D^+ and D^- in a unique way. The expressions for D^+ and D^- in Theorem III.2 state that we always have $j \in D^-$ and for any $k \in \{i\} \cup Z$ if there is no delay from k to j , given the information by the recursive graphical representation, we have $k \in D^+$ and D^- contains the remaining variables. It is possible to apply Theorem 2 also in presence of feedback loops and/or confounding variables affecting the nodes i and j . In the following example, we revisit

the networks of Example 5 and explain how Theorem 2 can be used to identify transfer functions of interest.

Example 8. Consider a network with a recursive graphical representation depicted in Figure 3.16 (a).

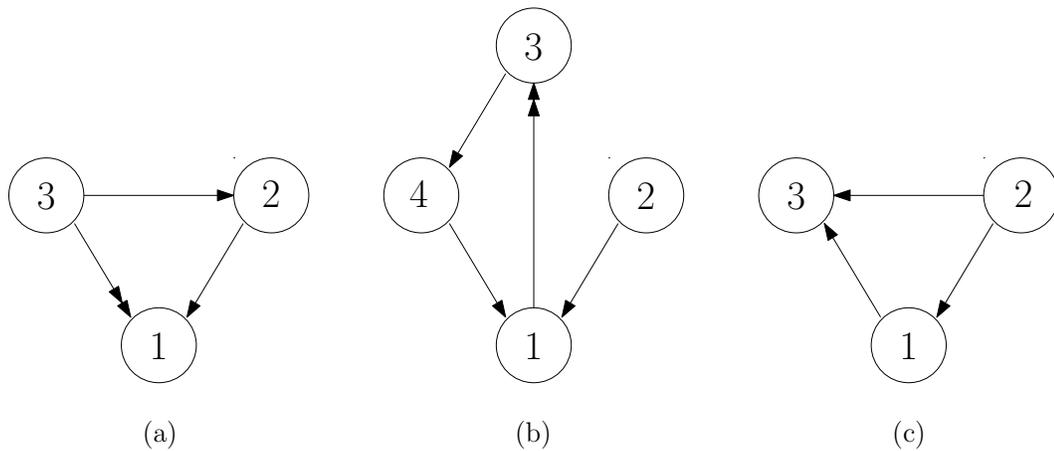


Figure 3.16: Revisiting the networks of Example 5.

Suppose the objective is identification of transfer function $H_{12}(z)$. Based on Theorem 2 the 1-pointing path $2 \leftarrow 3 \rightarrow 1$ needs to be blocked. Thus $Z = \{3\}$ satisfies conditions (i) and (ii) of Theorem 2 and results in a consistent estimate of $H_{12}(z)$ using Algorithm 1.

Similarly, consider a network with a recursive graphical representation depicted in Figure 3.16 (b). Suppose the objective is identification of transfer function $H_{12}(z)$. Based on Theorem 2 the 1-pointing path $2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ from 1 to itself needs to be blocked. Thus $Z = \{3\}$ or $Z = \{4\}$ or $Z = \{3, 4\}$ satisfy conditions (i) and (ii) of Theorem 2 and result in a consistent estimate of $H_{12}(z)$ using Algorithm 1. Conversely, consider a network with a recursive graphical representation depicted in Figure 3.16 (c). Suppose the objective is identification of transfer function $H_{12}(z)$. Based on Theorem 2 $Z = \{\emptyset\}$

satisfies conditions (i) and (ii) of Theorem 2 and results in a consistent estimate of $H_{12}(z)$ using Algorithm 1.

The following example illustrates how Theorem 2 can successfully deal with unobserved confounding variables.

Example 9. Consider the network with a block diagram depicted in Figure 3.17 and a recursive graphical representation shown in Figure 3.18. Suppose the objective is to

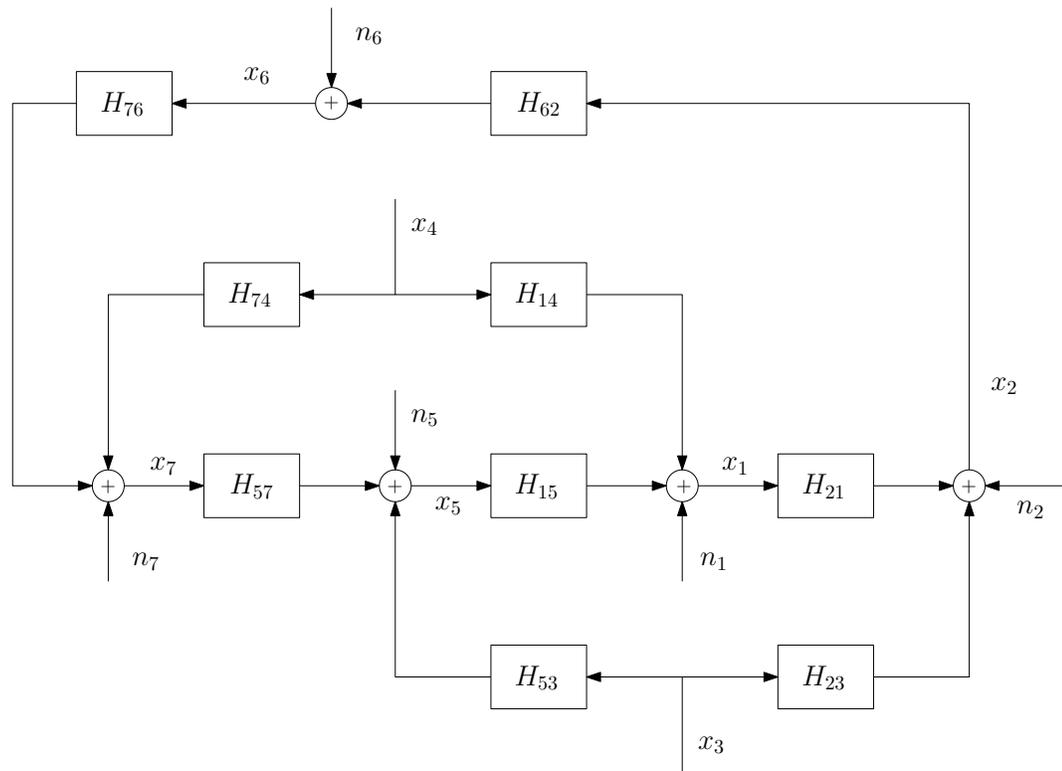


Figure 3.17: Block diagram of the network discussed in Example 9. The nodes 3 and 4 act as confounders.

identify the transfer function $H_{21}(z)$. Assume that nodes 3 and 4 which are depicted with a dashed line are not measured $3, 4 \notin O = \{1, 2, 5, 6, 7\}$. Note that Proposition 1 could not be applied because node 3 is a parent of the target node 2 but is not observable. Instead, we could search for a set $Z \subseteq O$ that satisfies conditions (i) and (ii) of Theorem 2. The

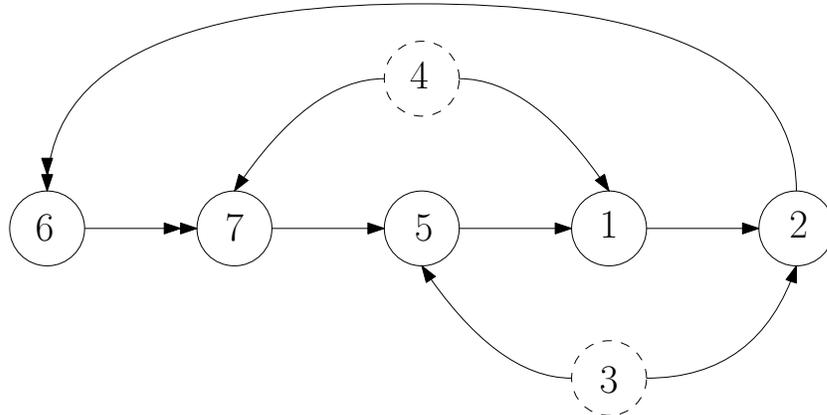


Figure 3.18: The graphical representation of the network discussed in Example 9. The nodes 3 and 4 act as confounders.

2-pointing path $\{1 \leftarrow 5 \leftarrow 3 \rightarrow 2\}$ needs to be blocked and since $3 \notin O$ we need to have $5 \in Z$. However, when node 5 is observed, it becomes an activated collider in the path $\{1 \leftarrow 4 \rightarrow 7 \rightarrow 5 \leftarrow 3 \rightarrow 2\}$. To block $\{1 \leftarrow 4 \rightarrow 7 \rightarrow 5 \leftarrow 3 \rightarrow 2\}$ we need to have $7 \in Z$. Since all the 2-pointing path from 2 to itself are also blocked by $\{5, 7\} \cup \{1\}$, we obtain that $Z = \{5, 7\}$ is a subset of $O = \{1, 2, 5, 6, 7\}$ that satisfies conditions (i) and (ii) of Theorem 2. Thus, applying Procedure 1 with $D^- = \{2\}$ and $D^+ = \{1, 5, 7\}$ leads to a consistent estimate of $H_{21}(z)$.

Observe that, by setting $Z = \text{pa}_G(j)$, Theorem 2 becomes equivalent to Corollary 1.1. One main advantage of Theorem 2 is that it can successfully deal with confounding variables in a way similar to the formulation of the Single Door Criterion for dynamic systems [56, 64]. However, contrary to the Single Door Criterion, Theorem 2 can be easily applied to networks where the node j is involved in feedback loops. Furthermore, the graphical conditions on the nodes i and j for the application of Single Door Criterion are stronger than the graphical condition required for Theorem 2. Indeed, Single Door Criterion needs all the paths between i and j to be blocked by a set Z that does not contain

descendants of j . Conversely, Theorem 2 only needs Z to block the j -pointing ones and Z also can contain descendants of j .

Note that the choice of predictors is not unique since multiple sets Z might satisfy the conditions of Theorem 2. In addition to this, for a fixed Z a further degree of flexibility can be obtained as follows: so long as another recursive graphical representation can be obtained from the graphical representation G of the network, different choices of the sets D^- and D^+ are also possible. The following example computes all the possible sets Z satisfying conditions (i) and (ii) of Theorem 2.

Example 10. Consider a network with a graphical representation G depicted in Figure 3.19. The objective is the identification of the transfer function $H_{21}(z)$. Node 4 is

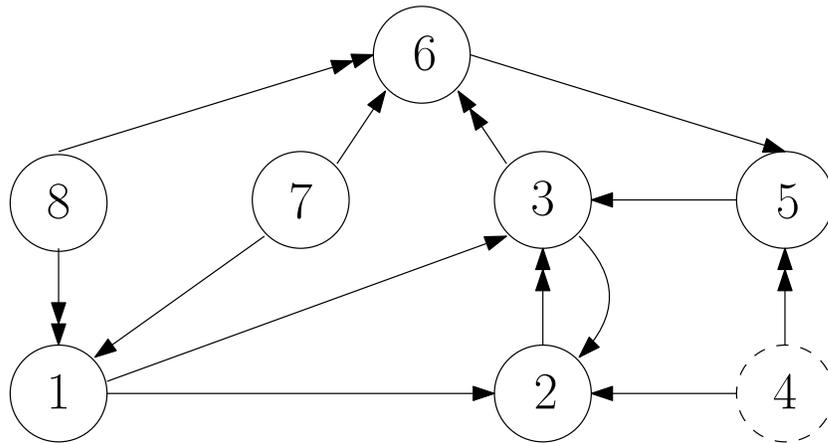


Figure 3.19: The graphical representation of the network discussed in Example 10.

not observable. Node 3 should be measured since it is the only choice to block the 2-loop $\{2 \rightarrow 3 \rightarrow 2\}$ and the 2-pointing path $1 \rightarrow 3 \rightarrow 2$. Now, since $3 \in Z$ node 3 will act as an activated collider on the 2-pointing path $\pi_1 = \{1 \rightarrow 3 \leftarrow 5 \leftarrow 4 \rightarrow 2\}$. Since 4 is not measured, the only choice for blocking π_1 is to have $5 \in Z$. Since $5 \in Z$ node 5 will act as an activated collider on the 2-pointing paths $\pi_2 = \{1 \leftarrow 7 \rightarrow 6 \rightarrow 5 \leftarrow 4 \rightarrow 2\}$ and

$\pi_3 = \{1 \leftarrow 8 \rightarrow 6 \rightarrow 5 \leftarrow 4 \rightarrow 2\}$. To block π_2 and π_3 we need to measure either node 6 or nodes 7 and 8 together. Table 10, lists all choices for Z that satisfy conditions (i) and (ii) of Theorem 2 and their corresponding D^- and D^+ .

Table 3.1: All possible predictor sets to consistently identify $H_{21}(z)$ in Example 10 for the given graphical representation

	Z	D^-	D^+
1	$\{3, 5, 6\}$	$\{2\}$	$\{1, 3, 5, 6\}$
2	$\{3, 5, 7, 8\}$	$\{2, 8\}$	$\{1, 3, 5, 7\}$
3	$\{3, 5, 6, 7\}$	$\{2\}$	$\{1, 3, 5, 6, 7\}$
4	$\{3, 5, 6, 8\}$	$\{2, 8\}$	$\{1, 3, 5, 6\}$
5	$\{3, 5, 6, 7, 8\}$	$\{2, 8\}$	$\{1, 3, 5, 6, 7\}$

In Section 3.4 we will also show that the choices listed in Table 3.1 are the only possible choices for Z guaranteeing a consistent identification of $H_{21}(z)$ using Procedure 1 for all networks with graphical representation G .

The following example explores the identification performance of our variable selection method in the case of finite data and also provides a numerical illustration of the consistency properties proven in the theoretical sections.

Example 11. Consider a network \mathcal{G} with a recursive graphical representation G shown in Figure 3.20.

Suppose the objective is the identification of the transfer function $H_{21}(z)$ using Procedure 1. To verify the consistency property of the identification when choosing a set of predictors satisfying the graphical conditions of Theorem 2, we numerically simulated the network \mathcal{G} and obtained time-series data. Using such generated time-series data, we con-

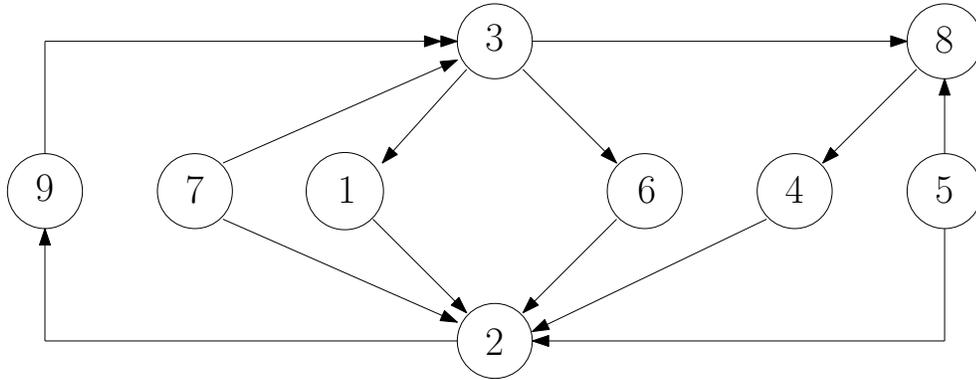


Figure 3.20: The graphical representation of the benchmark network discussed in Example 11.

sidered different sets of predictors and computed the bias and the variance of the estimated transfer function.

We consider a parameterization $H(z, \theta)$ of the network and we denote as θ_{21} the subset of parameters associated with the transfer function $H_{21}(z)$ and as $\hat{\theta}_{21}$ the estimated parameters. We chose two predictor sets $Z_1 = \{4, 5, 6, 7\}$ and $Z_2 = \{3, 9\}$ satisfying conditions of Theorem 2 and proceeded to the identification using time series of different lengths. For each set of predictors and for each time series length, we simulated the network \mathcal{G} and used a linear regression technique to obtain the estimate $\hat{\theta}_{21}$. We repeated this procedure 1000 times in order to estimate $\mathbb{E}(\hat{\theta}_{21})$ and the covariance matrix of $\hat{\theta}_{21}$.

In Figure 3.21 we have reported the results of our Monte Carlo simulations for the set Z_1 . On the horizontal axis we have the different time series lengths. The red squares represent the estimates of $\mathbb{E}\|\theta_{21} - \hat{\theta}_{21}\|_1$. We observe that for longer time series this quantity goes to zero numerically verifying that the bias of the estimated $\hat{\theta}_{21}$ asymptotically vanishes. The blue candle sticks define an interval the semi-amplitude of which is the square root of the trace of our estimate of the covariance matrix of $\hat{\theta}_{21}$. Since the amplitudes of these

intervals go to zero for longer time series we have numerically verified the consistency property of our estimate.

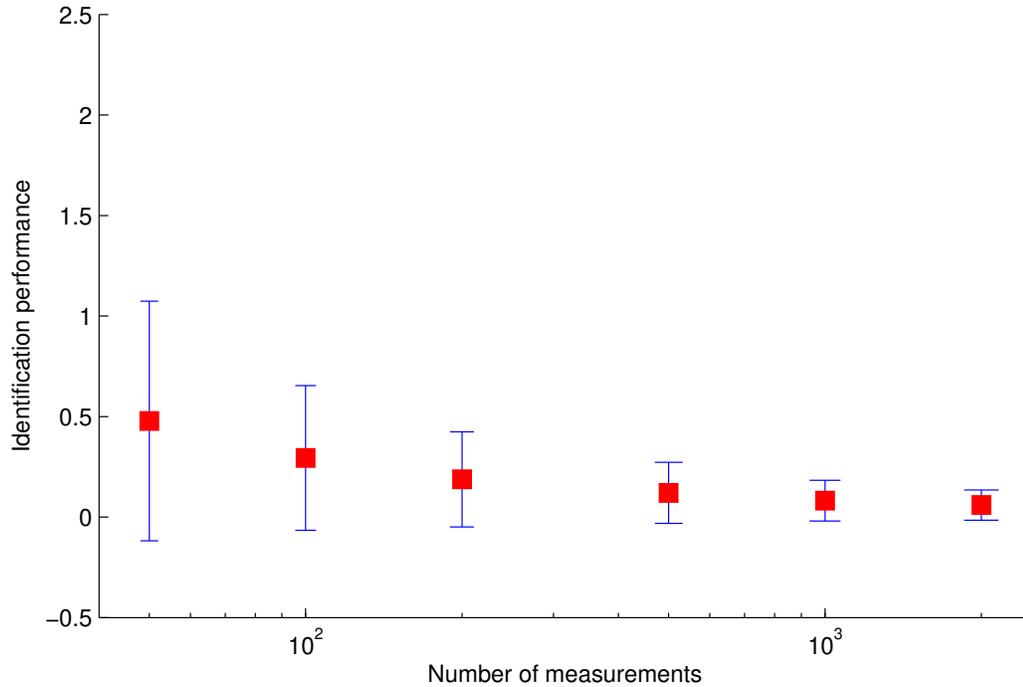


Figure 3.21: Identification performance of the predictor set $Z_1 = \{4, 5, 6, 7\}$ for different number of measurements.

We ran a similar set of simulations for Z_2 and the results are reported in Figure 3.22.

Notice that even though both sets Z_1 and Z_2 guarantee consistent identification, in the case of finite data they provide different performance in terms of bias and variance of estimated parameters.

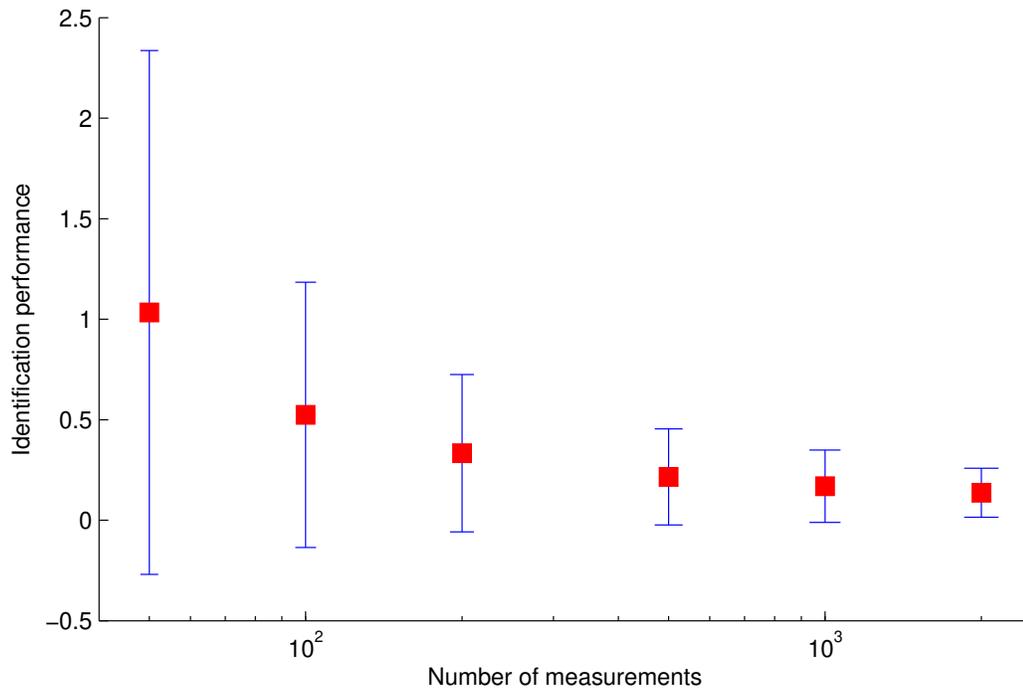


Figure 3.22: Identification performance of the predictor set $Z_2 = \{3, 9\}$ for different number of measurements.

3.4 Necessity of the graphical conditions for the selection of auxiliary variables

Theorem 2 provides sufficient conditions on how to select the set of auxiliary variables Z in order to consistently identify the transfer function $H_{ji}(z)$, namely, Z has to j -pointing separate i and j and block all the j -pointing paths from j to itself. In this section we show that these conditions are also necessary for the successful application of Procedure 1 when the only information about the network is given by a graphical representation.

Theorem 3. *For any recursive graph $G = (V, E_1, E_2)$, if the graphical conditions (i) and (ii) of Theorem 2 are not met by a set Z , there exists a network $\mathcal{G} = (H(z), n)$ with*

graphical representation G such that the estimate $\hat{H}_{ji}(z)$ of Procedure 1 will be inconsistent for all the possible choices of sets D^+ and D^- such that $D^+ \cup D^- = Z \cup \{i, j\}$ and $j \in D^-$.

Proof. See the appendix. □

Since the choice of the set Z is not unique, in some applications we might be interested in finding an optimal predictors set according to some cost function. The sufficiency of the conditions of Theorem 2, along with their necessity as proven in Theorem 3, are the basis to enable the search for an optimal predictors set which guarantees a consistent identification. In Chapter V we will present systematic algorithms to find an optimal predictors set for different types of identification problems in dynamic networks.

3.5 Proofs related to Chapter III

3.5.1 Proof of Proposition 1 and corollary 1.1

3.5.1.1 Proof of Proposition 1

Proof. Consider an estimate of $x_j(t)$ based on the information $I_{\{j\} \cup P^-}(t-1), I_{P^+}(t)$.

$$\begin{aligned} \mathbb{E}(x_j(t) \mid I_{\{j\} \cup P^-}(t-1), I_{P^+}(t)) &= \sum_{k \in \{j\} \cup P^-} H_{jk}(z)x_k(t) \\ &+ \sum_{k \in P^+} H_{jk}(z)x_k(t) + \mathbb{E}(n_j(t) \mid I_{\{j\} \cup P^-}(t-1), I_{P^+}(t)). \end{aligned} \quad (\text{III.6})$$

Define

$$\epsilon_j(t) = n_j(t) - \hat{n}_j(t) \quad (\text{III.7})$$

where $\hat{n}_j(t) = \mathbb{E}(n_j(t) | I_{\{n_j\}}(t-1)) = W_{jj}(z)n_j(t)$ with $W_{jj}(z)$ being strictly causal. Observe that $\mathbb{E}(n_j(t) | I_{\{n_j\}}(t-1))$ is $(I_{\{j\} \cup P^-}(t-1), I_{P^+}(t))$ -measurable since $n_j(t) = y_j(t) - \sum_{k \in P^-} H_{jk}(z)x_k(t) + \sum_{k \in P^+} H_{jk}(z)x_k(t)$. Since $\epsilon_j(t) \perp I_{\{j\} \cup P^-}(t-1), I_{P^+}(t)$, we get

$$\begin{aligned} \mathbb{E}(x_j(t) | I_{\{j\} \cup P^-}(t-1), I_{P^+}(t)) &= \\ & \sum_{k \in P^-} H_{jk}(z)x_k(t) + \sum_{l \in P^+} H_{jk}(z)x_k(t) + \hat{n}_j(t) = \\ & \sum_{k \in P^-} H_{jk}(z)x_k(t) + \sum_{k \in P^+} H_{jk}(z)x_k(t) + W_{jj}(z)[x_j(t) - \sum_{k \in P^-} H_{jk}(z)x_k(t) + \sum_{k \in P^+} H_{jk}(z)x_k(t)] = \\ & W_{jj}(z)x_j(t) + \sum_{k \in P^-} [1 - W_{jj}(z)]H_{jk}(z)x_k(t) + \sum_{k \in P^+} [1 - W_{jj}(z)]H_{jk}(z)x_k(t). \quad (\text{III.8}) \end{aligned}$$

□

3.5.1.2 Proof of Corollary 1.1

Proof. Since, the minor of the power spectral density matrix corresponding to $\text{pa}_G(j)$ is non-singular, the Wiener filter components estimating x_j from x_i and $x_{P^+ \cup P^-}$ is unique.

Hence, we have

$$W_{ji}(z) = [1 - W_{jj}(z)]H_{ji}(z). \quad (\text{III.9})$$

□

3.5.2 Proof of Theorem 2

To prove Theorem 2, we first need to provide a few lemmas.

3.5.2.1 Lemma 4

Lemma 4. Consider a network $\mathcal{G} = (H(z), n)$ with graphical representation $G = (V, E_1, E_2)$ and output processes x_V described by (II.5). Suppose $j \in V$ is a node in the network such that $ch_{G^c}(j) = \emptyset$. Define a network $\bar{\mathcal{G}} = (\bar{H}, \bar{n}_V)$ with output processes \bar{x}_V , where

$$\bar{H}_{jj}(z) = H_{jj}(z)$$

$$\bar{H}_{ab}(z) = H_{ab}(z) \quad \text{for } a, b \neq j$$

$$\bar{H}_{aj}(z) = zH_{aj}(z) \quad \text{for } a \neq j$$

$$\bar{H}_{ja}(z) = z^{-1}H_{ja}(z) \quad \text{for } a \neq j$$

$$\bar{n}_a(t) = n_a(t) \quad \text{for } a \neq j$$

$$\bar{n}_a(t) = n_a(t - 1) \quad \text{for } a = j.$$

Then, in $\bar{\mathcal{G}}$ we have that

$$\bar{x}_a(t) = x_a(t) \quad \text{for } a \neq j$$

$$\bar{x}_a(t) = x_a(t - 1) \quad \text{for } a = j.$$

Furthermore, $\bar{\mathcal{G}}$ has a graphical representation given by $\bar{G} = (V, \bar{E}_1, \bar{E}_2)$, where

$$\bar{E}_1 = (E_1 \cup_{c \in ch_G(j)} \{j \rightarrow c\}) \setminus \{\cup_{p \in pa_G(j)} p \rightarrow j\} \quad (\text{III.10})$$

$$\bar{E}_2 = (E_2 \cup_{p \in pa_G(j)} \{p \rightarrow j\}) \setminus \{\cup_{c \in ch_G(j)} j \rightarrow c\} \quad (\text{III.11})$$

and the relation $de_{\bar{G}^\ell}(k) = de_{G^\ell}(k) \setminus \{j\}$ holds for every node $k \neq j$.

Proof. Define the square matrix $M^j(z) = [m_{ab}(z)]$, $a, b, j \in V$ such that all off-diagonal entries $m_{ab}(z) = 0$ for $a \neq b$, $m_{ab}(z) = 1$ for $a = b \neq j$, and $m_{jj}(z) = z^{-1}$. Therefore, we have that $\bar{H}(z) = M^j(z)H(z)M^{j-1}(z)$ and $\bar{n} = M^j(z)n$. Thus, the output processes of \bar{G} could be calculated in terms of output processes of G as follows.

$$\begin{aligned}\bar{x}_V &= \bar{H}(z)\bar{x}_V + \bar{n} = M^j(z)H(z)M^{j-1}(z)\bar{x}_V + M^j(z)n = \\ &M^j(z)H(z)x_V + M^j(z)n = M^j(z)(H(z)x_V + n) = M^j(z)x_V,\end{aligned}$$

which verifies that $\bar{x}_a(t) = x_a(t)$ for $a \neq j$ and $\bar{x}_j(t) = x_j(t-1)$. Moreover, since all the transfer functions $\bar{H}_{jb}(z) = z^{-1}H_{jb}(z)$, for $b \neq j$, are strictly causal, all the edges $p \rightarrow j$ for $p \in \text{pa}_{\bar{G}}(j)$ can be double-headed. Therefore, \bar{G} is a graphical representation of \bar{G} . Finally, since all the edges $p \rightarrow j$ for $p \in \text{pa}_{\bar{G}}(j)$ can be double-headed, and $\text{ch}_{G^\ell}(j) = \emptyset = \text{de}_{G^\ell}(j) \setminus j$, we have that $de_{\bar{G}^\ell}(k) = de_{G^\ell}(k) \setminus \{j\}$ for any node $k \neq j$.

□

3.5.2.2 Lemma 5

Lemma 5. *Given a recursive graph G , for any node j there exists at least a node $d \in de_G(j)$ such that $ch_{G^\ell}(d) = \emptyset$.*

Proof. The result is an immediate consequence of the fact that for any recursive graph G , the graph of instantaneous propagations G^ℓ is a directed acyclic graph. □

3.5.2.3 Lemma 6

The following lemma provides a connection between the standard notion of d -separation and the notion of pointing separation adopted in this article.

Lemma 6. *Let $i, j \in V$, $Z \subset V$, and $\{i, j\} \cap Z = \emptyset$ in a network with graphical representation $G = (V, E)$. If all the j -pointing paths between nodes i and j are blocked by Z , then i and $\text{pa}(j) \setminus \{Z \cup \{j\}\}$ are d -separated by $Z \cup \{j\}$ in G .*

Proof. By contradiction suppose $w \in \text{pa}(j) \setminus \{Z \cup \{j\}\}$ such that there is a connected path $\tilde{\pi}$ between nodes i and w not blocked by $Z \cup \{j\}$. We can have two cases. Either j is in $\tilde{\pi}$ or not. Suppose j is not in $\tilde{\pi}$ and $\tilde{\pi}$ is not blocked by $Z \cup \{j\}$. If j is not a descendent of colliders in $\tilde{\pi}$, then $\pi = (\tilde{\pi}, w \rightarrow j)$ is a j -pointing path connecting nodes i and j not blocked by Z , which is a contradiction. If j is a descendent of some colliders in $\tilde{\pi}$, let c be the closest such collider to i . Then, $\pi = (i \cdots \rightarrow c \rightarrow \cdots \rightarrow j)$ is a j -pointing path connecting nodes i and j not blocked by Z , which is a contradiction. Now, suppose j is in $\tilde{\pi}$ and $\tilde{\pi}$ is not blocked by $Z \cup \{j\}$. Then, $\tilde{\pi}$ is either of the form $\tilde{\pi} = (\hat{\pi} \rightarrow j \cdots w)$ or $\tilde{\pi} = (\hat{\pi} \leftarrow j \cdots w)$. If $\tilde{\pi} = (\hat{\pi} \rightarrow j \cdots w)$, then $\pi = \hat{\pi} \rightarrow j$ is a j -pointing path connecting nodes i and j not blocked by Z , which is a contradiction. If $\tilde{\pi} = (\hat{\pi} \leftarrow j \cdots w)$, then π is blocked by $Z \cup \{j\}$, which is a contradiction. \square

We will use some of the basic properties of the conditional independence relation which are given by the following lemma.

Lemma 7. *let I_A, I_B, I_C , and I_D be sub- σ -algebras of \mathcal{F} . Then the following conditional orthogonality properties hold.*

1. *Symmetry:* $I_A \perp I_B \mid I_C \iff I_B \perp I_A \mid I_C$

2. *Composition/decomposition:* $I_A \perp I_{B \cup C} \mid I_D \iff I_A \perp I_B \mid I_D \text{ and } I_A \perp I_C \mid I_D$

3. *Weak union:* $I_A \perp I_{B \cup C} \mid I_D \implies I_A \perp I_B \mid I_{C \cup D}$

4. *Contraction:* $I_A \perp I_B \mid I_C \text{ and } I_A \perp I_D \mid I_{B \cup C} \implies I_A \perp I_{B \cup D} \mid I_C$

Moreover, if $I_{B \cup C}$ is separable and $(I_{B \cup D}) \cap (I_{C \cup D}) = I_D$, then we additionally have

5. *Intersection:* $I_a \perp I_b \mid I_{C \cup D} \text{ and } I_A \perp I_C \mid I_{B \cup D} \implies I_A \perp I_{B \cup C} \mid I_D$

In graph theory, a dependency model that is closed under the five axioms of lemma 7 is defined as a graphoid.

3.5.2.4 Lemma 8

The following lemma allows us to determine when two filtrations are equivalent.

Lemma 8. *Let $x_k, k \in A = \{i \cup B\}$ be a set of rationally related processes. Define the variable q as follows.*

$$q = \sum_{k \in A} H_k(z)x_k \quad (\text{III.12})$$

where $H_k(z)$ are causal transfer functions. Then,

$$I_{q \cup B}(t) = I_A(t) \quad (\text{III.13})$$

if and only if $H_i^{-1}(z)$ is causal.

Proof. We need to show two things. $I_{q \cup B}(t) \supseteq I_A(t)$ and $I_{q \cup B}(t) \subseteq I_A(t)$. Since $q = \sum_{k \in A} H_k(z)x_k$, we get $I_{q \cup B}(t) = I_{A \cup B}(t)$. Therefore, $I_{q \cup B}(t) \supseteq I_A(t)$. On the other

hand, $H_i(z)x_i = q - \sum_{k \in B} H_k(z)x_k$. For $H_i(z) \neq 0$ we get

$$x_i = H_i^{-1}(z)[q - \sum_{k \in B} H_k(z)x_k]. \quad (\text{III.14})$$

Therefore, $x_i \in I_{q \cup B}(t)$ if and only if $H_i^{-1}(z)$ is causal. \square

We are now ready to prove Theorem 2.

3.5.2.5 Proof of Theorem 2

Proof. Let $G' = (V, (E_1 \cup E_2) \setminus \{i \rightarrow j\})$ be the standard directed graph associated to G after removing the edge $i \rightarrow j$. Define $E = E_1 \cup E_2$. Also define a new process $x_q(t) = x_j(t) - H_{ji}(z)x_i(t)$. We are going to define a new network $\mathcal{G}'' = (H'', n'')$ with all the variables of the original network and the additional variable x_q . Let

$$H''_{jq}(z) = 1,$$

$$H''_{qr}(z) = H_{jr}(z) \text{ for } r \in \text{pa}_G(j) \setminus i,$$

$$H''_{k\ell}(z) = H_{k\ell}(z) \text{ in all other cases}$$

and $n''_q = n_j$, $n''_j = 0$ and $n''_k = n_k$ in all other cases. From G we can obtain a graphical representation for \mathcal{G}'' given by $G'' = (V'', E''_1, E''_2)$ in the following way. Let $K_1 := \{k | k \neq i \text{ and } k \rightarrow j \in E_1\}$ be the set of single-headed parents of j in G that are not node i and $K_2 := \{k | k \neq i \text{ and } k \rightarrow j \in E_2\}$ be the set of double-headed parents of j in G that are

not node i . Then

$$V'' := V \cup \{q\}$$

$$E_1'' := E_1 \cup \{q \rightarrow j\} \cup \{k \rightarrow q \mid k \in K_1\} \setminus \{k \rightarrow j \mid k \in K_1\}$$

$$E_2'' := E_2 \cup \{k \rightarrow q \mid k \in K_2\} \setminus \{k \rightarrow j \mid k \in K_2\}.$$

Namely, in $G'' = (V'', E_1'', E_2'')$, the additional node q is placed in between j and its original parents in G that are not the node i , see Figure 3.23. Also, notice that since G is recursive,

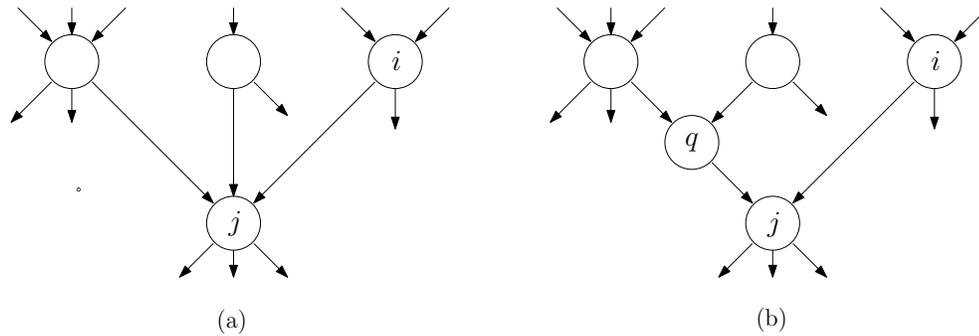


Figure 3.23: The new variable q in G'' is introduced in between node j and all its parents that are not the node i .

G'' is trivially recursive, as well. Define $P = \text{pa}_{G''}(q) \setminus Z$. Decompose P into P^- and P^+ where P^- contains all double headed parents of q that are not in Z and $P^+ = P \setminus P^-$ contains all the single-headed parents of q that are not in Z . Observe that $j \in D^-$, hence D^- is never empty. Conversely, the node i belongs to either D^- or D^+ . First, we consider the case $i \in D^+$. Let S_0 be the set containing all descendants of the nodes in D^- in G'' , $S_0 := \text{de}_{G''}(D^-)$. Since G'' is recursive and D^- is not empty, by Lemma 5 there exists a node $w_0 \in S_0$ such that $\text{ch}_{G''}(w_0) = \emptyset$. Apply Lemma 4 on w_0 obtaining a

new network with recursive graphical representation G_1 . Define $S_1 := \text{de}_{G_1^{\neq}}(D^-)$. From Lemma 4 it follows that $S_1 = S_0 \setminus \{w_0\}$. Again, by Lemma 5 there exists a node $w_1 \in S_1$ such that $\text{ch}_{G_1^{\neq}}(w_1) = \emptyset$. Apply Lemma 4 on w_1 , represent the resulting network with G_2 and let $S_2 := \text{de}_{G_2^{\neq}}(D^-)$. Again, from Lemma 4 it follows that $S_2 = S_1 \setminus \{w_1\}$. Repeat the procedure N times, till $S_N = \emptyset$, for $N \in \mathbb{Z}$, the number of elements in S_0 . Let (H_N, n_N) be the resulting network with recursive graphical representation G_N . Build a new graphical representation \underline{G} for (H_N, n_N) by adding single-headed edges from nodes in Z to q . Observe that in \underline{G} , the parents of q are now given by $P^- \cup P^+ \cup Z$. Let $Z^- := S_0 \cap Z$ and $Z^+ := Z \setminus Z^-$. Since S_0 contains all the elements in D^- we have that $D^- = Z^- \cup \{j\}$. Consequently, $D^+ = Z^+ \cup \{i\}$. Also, it follows from Lemma 4 that in (H_N, n_N) the output processes of nodes k in Z^- are now $x_k(t-1)$ while the output processes of nodes i (since $i \in D^+$) and ℓ in Z^+ remain unchanged, $x_\ell(t)$. Applying Proposition 1 on q in \underline{G} , which is recursive, gives

$$x_q(t) \perp I_i(t) \mid I_{P^- \cup Z^- \cup \{q\}}(t-1), I_{P^+ \cup Z^+}(t). \quad (\text{III.15})$$

Apply Lemma 4 one more time on q in G_N to get \bar{G} . Every q -pointing path $\bar{\pi}_1 = \{i \cdots k \rightarrow q\}$ between i and q in \bar{G} that does not pass through j could be mapped to a j -pointing path $\pi_1 = \{i \cdots k \rightarrow j\}$ between i and j in G' . Thus, from assumption (i) we have that Z also blocks all paths $\bar{\pi}_1$ in \bar{G} . Every q -pointing path $\bar{\pi}_2 = \{i \cdots j \rightarrow \hat{\pi}_3 \rightarrow q\}$, where $\hat{\pi}_3$ is a node or a path that might or might not be directed, between i and q in \bar{G} that passes through j corresponds to a j -pointing path $\pi_2 = \{j \rightarrow \hat{\pi}_3 \rightarrow j\}$ between j and itself in G . Thus, from assumption (ii) we have that Z also blocks all paths $\bar{\pi}_2$ in \bar{G} . Therefore, Z blocks all the q -pointing paths between i and q in \bar{G} . By Lemma 6, we have that $\text{pa}_{\bar{G}}(q) \setminus Z$

are d -separated from i by $Z \cup \{q\}$. Theorem 24 in [64] states that if two sets of nodes A and B are d -separated by C in a graphical representation of a network we have that $I_A(t) \perp I_B(t) \mid I_C(t)$. Therefore, applying Theorem 24 in [64] to \bar{G} we get

$$I_{P^-}(t-1), I_{P^+}(t) \perp I_i(t) \mid I_{\{q\} \cup Z^-}(t-1), I_{Z^+}(t). \quad (\text{III.16})$$

By Contraction property of conditional independence [56, Chapter 1, Page 11], combining (III.15) and (III.16), we obtain

$$I_{P^-}(t-1), I_{P^+ \cup \{q\}}(t) \perp I_i(t) \mid I_{Z^- \cup \{q\}}(t-1), I_{Z^+}(t). \quad (\text{III.17})$$

By Decomposition property of conditional independence [56, Chapter 1, Page 11], this yields

$$x_q(t) \perp I_i(t) \mid I_{Z^- \cup \{q\}}(t-1), I_{Z^+}(t). \quad (\text{III.18})$$

Therefore, when estimating $x_q(t)$ from $I_{\{i\} \cup Z^+}(t), I_{\{q\} \cup Z^-}(t-1)$, the transfer function corresponding x_i will be zero:

$$\begin{aligned} \mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{q\} \cup Z^-}(t-1)) &= \\ \mathbb{E}(x_q(t) \mid I_{Z^+}(t), I_{\{q\} \cup Z^-}(t-1)) &= \\ F_{qq}(z)x_q(t) + \sum_{k \in Z^-} F_{qk}(z)x_k(t) + \sum_{k \in Z^+} F_{qk}(z)x_k(t), & \quad (\text{III.19}) \end{aligned}$$

where $F_{qq}(z)$ and $F_{qk}(z), k \in Z^-$ are strictly causal transfer functions, and $F_{qk}(z), k \in Z^+$ are causal transfer functions. Since $H''_{jq}(z) = 1$ is causally invertible, it follows that the

filtration induced by random processes x_i and x_{Z^+} till time t and x_{Z^-} and x_q till time $t - 1$ is equal to the filtration induced by random processes x_i and x_{Z^+} till time t and x_{Z^-} and x_j till time $t - 1$. Since $x_j(t) = x_q(t) + H_{ji}(z)x_i(t)$, when $x_j(t)$ is projected on $I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t - 1)$, we get

$$\begin{aligned}
\mathbb{E}(x_j(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t - 1)) &= \\
\mathbb{E}(x_q(t) + H_{ji}(z)x_i(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t - 1)) &= \\
H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t - 1)) &= \\
H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{q\} \cup Z^-}(t - 1)) &= \\
H_{ji}(z)x_i(t) + F_{qq}(z)x_q(t) + \sum_{k \in Z^-} F_{qk}(z)x_k(t) + \sum_{k \in Z^+} F_{qk}(z)x_k(t) &= \\
H_{ji}(z)x_i(t) + F_{qq}(z)(x_j(t) - H_{ji}(z)x_i(t)) + \sum_{k \in Z^-} F_{qk}(z)x_k(t) + \sum_{k \in Z^+} F_{qk}(z)x_k(t) &= \\
F_{qq}(z)x_j(t) + [H_{ji}(z) - F_{qq}(z)H_{ji}(z)]x_i(t) + \sum_{k \in Z^-} F_{qk}(z)x_k(t) + \sum_{k \in Z^+} F_{qk}(z)x_k(t). &
\end{aligned} \tag{III.20}$$

Procedure 1 in step 3 computes

$$\mathbb{E}(x_j(t) \mid I_{D^-}(t - 1), I_{D^+}(t)) = \sum_{k \in D^- \cup D^+} W_{jk}(z)x_k(t). \tag{III.21}$$

Since the power spectral density matrix associated with (x_i, x_j, x_Z) is non-singular, comparing the two expressions for $\mathbb{E}(x_j(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t - 1))$ we can conclude

$$W_{jk}(z) = F_{qk}(z) \quad \text{for all } k \in Z^- \cup Z^+, \tag{III.22}$$

$$W_{jj}(z) = F_{qq}(z), \quad (\text{III.23})$$

and

$$W_{ji}(z) = H_{ji}(z) - F_{qq}(z)H_{ji}(z) = H_{ji}(z) - W_{jj}(z)H_{ji}(z). \quad (\text{III.24})$$

This verifies the assertion for the case where $i \in D^+$.

Now suppose $i \in D^-$. Taking similar steps we get

$$I_{P^-}(t-1) \perp I_i(t-1) \mid I_q(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t), \quad (\text{III.25})$$

and

$$I_{P^+}(t) \perp I_i(t-1) \mid I_q(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t). \quad (\text{III.26})$$

Combining (III.25) and (III.26) by the union property of lemma 7 we get

$$I_{P^-}(t-1) \cup I_{P^+}(t) \perp I_i(t-1) \mid I_q(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t). \quad (\text{III.27})$$

On the other hand, we have

$$x_q(t) \perp I_i(t-1) \mid I_{P^-}(t-1) \cup I_{P^+}(t) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1). \quad (\text{III.28})$$

By Contraction property of lemma 7, combining (III.28) and (III.27), we obtain

$$I_{P^-}(t-1) \cup I_{P^+}(t) \cup x_q(t) \perp I_i(t-1) \mid I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1). \quad (\text{III.29})$$

By Decomposition property of lemma 7, this yields

$$x_q(t) \perp I_i(t-1) \mid I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1). \quad (\text{III.30})$$

Therefore, when estimating $x_q(t)$ using $I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1)$, the transfer function corresponding x_i will be zero:

$$\begin{aligned} \mathbb{E}(x_q(t) \mid I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1)) = \\ W_{qq}(z)x_q(t) + \sum_{k \in Z^-} W_{qk}(z)x_k(t) + \sum_{l \in Z^+} W_{ql}(z)x_l(t), \quad (\text{III.31}) \end{aligned}$$

where $W_{qq}(z)$ and $W_{qk}(z)$, $k \in Z^-$ are strictly causal transfer functions, and $W_{ql}(z)$, $l \in Z^+$ are causal transfer functions. It follows from proposition 8 that $I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1) = I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_j(t-1)$. Since, $x_j(t) = x_q(t) + H_{ji}(z)x_i(t)$, when $x_j(t)$ is estimated using $I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_j(t-1)$,

we get

$$\begin{aligned}
& \mathbb{E}(x_j(t) \mid I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_j(t-1)) = \\
& \mathbb{E}(x_q(t) + H_{ji}(z)x_i(t) \mid I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_j(t-1)) = \\
& H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_j(t-1)) = \\
& H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_q(t-1)) = \\
& H_{ji}(z)x_i(t) + W_{qq}(z)x_q(t) + \sum_{k \in Z^-} W_{qk}(z)x_k(t) + \sum_{l \in Z^+} W_{ql}(z)x_l(t) = \\
& H_{ji}(z)x_i(t) + W_{qq}(z)(x_j(t) - H_{ji}(z)x_i(t)) + \sum_{k \in Z^-} W_{qk}(z)x_k(t) + \sum_{l \in Z^+} W_{ql}(z)x_l(t) = \\
& W_{qq}(z)x_j(t) + [H_{ji}(z) - W_{qq}(z)H_{ji}(z)]x_i(t) + \sum_{k \in Z^-} W_{qk}(z)x_k(t) + \sum_{l \in Z^+} W_{ql}(z)x_l(t),
\end{aligned} \tag{III.32}$$

where $H_{ji}(z)$ is strictly causal. Since the power spectral density matrix associated with (x_i, x_j, x_Z) is non-singular, comparing the two expressions for $\mathbb{E}(x_j(t) \mid I_i(t-1) \cup I_{Z^-}(t-1) \cup I_{Z^+}(t) \cup I_j(t-1))$ we can conclude:

$$W_{jk}(z) = F_{qk}(z) \quad \text{for all } k \in Z^- \cup Z^+, \tag{III.33}$$

$$W_{jj}(z) = F_{qq}(z), \tag{III.34}$$

and

$$W_{ji}(z) = H_{ji}(z) - F_{qq}(z)H_{ji}(z) = H_{ji}(z) - W_{jj}(z)H_{ji}(z). \tag{III.35}$$

where $H_{ji}(z)$ is strictly causal. This verifies the assertion for the case where $i \in D^-$ and completes the proof. \square

3.5.3 Proof of Theorem 3

Proof. Suppose the first condition of Theorem 2 is not met. That is, there is at least a j -pointing path between nodes i and j that is not blocked by Z and such a path is not the edge from i to j . Let π be the activated path with the least number of colliders. If π is collider-free, none of the nodes on π is a member of the separating set Z . Choose $H_{ji}(z)$ and all the other transfer functions outside of π as zero. Let the transfer function from the parent of j on π to j be equal to z^{-m} where m is the length of π . Let all the other transfer functions on π be equal to z^{-1} . This reduces the perfect graphical representation of the network to only path π as shown in Figure 3.24. Let the noise processes on all the chain

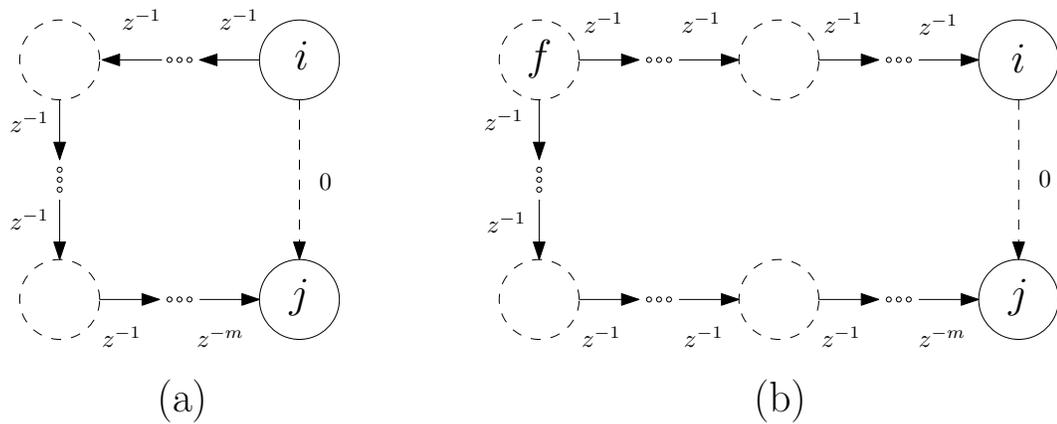


Figure 3.24: Scenario one in the proof of Theorem 3.

links and j be zero, and let the noise process on the fork, if there is one, and i be white with nonzero power spectral density. Since the transfer functions are all strictly causal the graph G is a graphical representation of this network. Notice that all the nodes in Z are independent of i and j . Also notice that the past of x_i is correlated with $x_j(t)$, and $x_i(t)$

and $x_j(t)$ are independent of each other. Thus, for every possible choice of D^+ and D^- we have

$$\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)) = \mathbb{E}(x_j(t) \mid I_{\{i,j\}}(t-1)) \neq 0$$

that is a biased estimate of $H_{ji}(z)$.

In the second scenario there is at least one collider on π . Since the path is activated by Z , each collider on π must be either in Z or have at least one of its descendants in Z according to graph G . Furthermore, none of the non-colliders on π is in Z . Let every transfer function on the dipath from each collider to its descendants be equal to one. Let the noise process on each collider have variance one and let the noise on its other descendants be equal to zero. Since π has the minimal number of colliders, distinct colliders need to have disjoint descendants. Thus, measuring each descendant of each collider is the same as measuring the collider itself. Hence, we can assume without any loss of generality, that Z is a set of colliders on π .

In the second scenario, suppose there exist r colliders $c_h \in V$, $h = 1, 2, 3, \dots, r$ on π . Choose the transfer function entering j to be z^{-2} , all the transfer functions entering all the colliders to be z^{-1} , and all other transfer functions on π to be one. Let all the noise processes on all the chain links and j be zero, and the noise processes on the forks and colliders be white with variance one. If π is not i -pointing (Figure 3.25 (a)), let the noise on i be one. Otherwise, if π is i -pointing (Figure 3.25 (b)), let the the noise on i be white with variance zero.

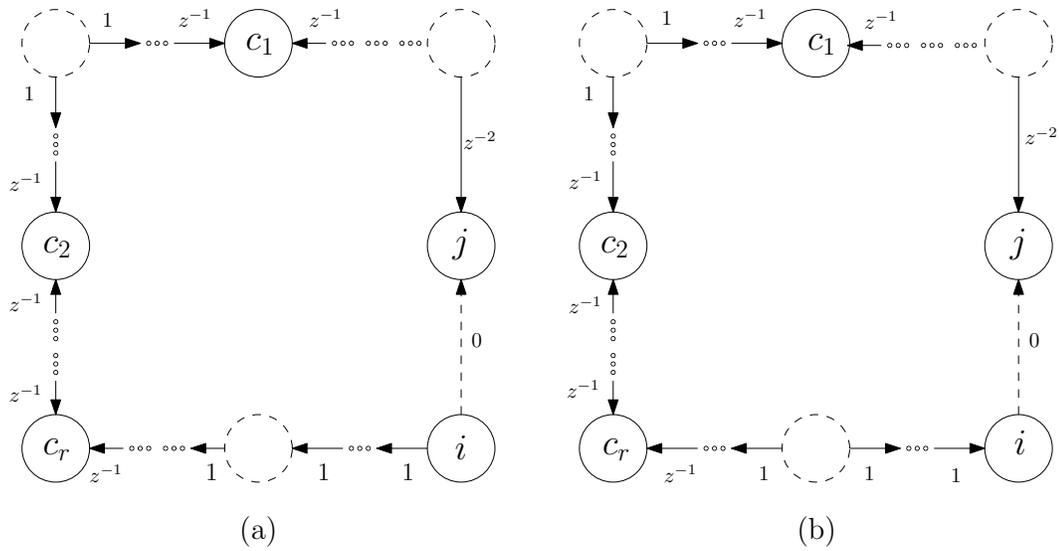


Figure 3.25: Generic configurations of scenario two in the proof of Theorem 3.

The power spectral density associated with random processes $x_i, x_{c_1}, x_{c_2}, \dots, x_{c_r}, x_j$ is given by the $(r + 2) \times (r + 2)$ matrix

$$\Sigma = \begin{bmatrix} 1 & z & 0 & 0 & 0 & \dots & 0 \\ z^{-1} & 3 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 3 & 1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & & \\ 0 & 0 & \dots & 1 & 3 & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 & 3 & z \\ 0 & 0 & \dots & 0 & 0 & z^{-1} & 1 \end{bmatrix}. \quad (\text{III.36})$$

From Σ and the inverse of tridiagonal matrices formula [97] it is possible to compute the non-causal Wiener filter estimating x_j from $x_i, x_{c_1}, x_{c_2}, \dots, x_{c_r}$. In particular, the component of the Wiener filter associated with x_i and x_{c_h} for $h = 1, 2, \dots, r$ are given by

$$W_{ji}(z) = \frac{z^{-2}}{\theta_{r+1}} \quad W_{jc_h}(z) = \frac{z^{-1}\theta_h}{\theta_{r+1}} \quad (\text{III.37})$$

where

$$\theta_k = 3\theta_{k-1} - \theta_{k-2} \quad (\text{III.38})$$

with $\theta_0 = 1$ and $\theta_1 = 1$ for $k = 2, \dots, r + 1$. Since the non-causal Wiener filter is a strictly causal transfer function, it matches the Wiener-Hopf filter. Also, the strict causality of the Wiener filter implies that the expression of $W_{ji}(z)$ does not change for all choices of D^+ and D^- such that $D^+ \cup D^- = Z \cup \{i, j\}$ proving the estimate of $H_{ji}(z)$ via Procedure 1 is biased.

Now suppose the second condition of Theorem 2 is not met. That is, there exists a j -pointing path between j and itself which is not blocked. There are two cases. Either the unblocked path is directed or not. First, we consider the case where there exists a directed feedback loop from the target node j to itself. Assume a network where all the transfer functions that are not involved in such a loop are zero. Instead, let all the transfer functions on the loop be $\frac{\alpha}{z}$ with $0 < |\alpha| < 1$. Let $W_{jj}(z)$ be the product of all the transfer functions on this directed feedback loop. Then the estimate of $H_{ji}(z)$ for all choices of D^- and D^+ will be given by

$$\hat{H}_{ji}(z) = \frac{H_{ji}(z)}{1 - W_{jj}(z)} \quad (\text{III.39})$$

which is biased. Now we consider the case where there is a j -pointing path ℓ between j and itself which is not directed. That is, there is at least one collider on ℓ . Since ℓ is activated by Z , each collider on ℓ must be either in Z or have at least one of its descendants

in Z according to graph G . Furthermore, none of the non-colliders on ℓ is in Z . Similar to above, we can assume without any loss of generality, that Z is a set of colliders on ℓ . Suppose there exist r colliders $c_h \in V$, $h = 1, 2, 3, \dots, r$ on ℓ . Choose $H_{ji}(z)$ to be z^{-1} and all the other transfer functions outside of ℓ as zero. Choose the transfer function entering j from its parent j to be z^{-2} , all the transfer functions entering all the colliders on ℓ to be z^{-1} , and all other transfer functions on ℓ to be one. Let all the noise processes on all the chain links and j be zero, and the noise processes on the forks, colliders, and i be white with variance one. We have that $D^+ = \emptyset$ and $D^- = \{i, j, Z\}$

$$\begin{aligned}
\mathbb{E}(x_j(t) \mid I_{\{i,j\} \cup Z}(t-1)) &= \\
\mathbb{E}(H_{ji}(z)x_i(t) + H_{jp}(z)x_p(t) + n_j(t) \mid I_{\{i,j\} \cup Z}(t-1)) &= \\
\mathbb{E}(z^{-1}x_i(t) + z^{-2}x_p(t) + n_j(t) \mid I_{\{i,j\} \cup Z}(t-1)) &= \\
z^{-1}x_i(t) + \mathbb{E}(z^{-2}x_p(t) + n_j(t) \mid I_{\{i,j\} \cup Z}(t-1)) &= \\
z^{-1}x_i(t) + \mathbb{E}(z^{-2}x_p(t) + n_j(t) \mid I_{\{j\} \cup Z}(t-1)) &= \\
z^{-1}x_i(t) + \sum_{k \in \{j\} \cup Z} W_{jk}(z)x_k(z) & \quad \text{(III.40)}
\end{aligned}$$

Computing the Wiener filters $W_{jk}(z)$, $k \in \{j\} \cup Z$, using the power spectral density matrix like above, it turns out that $W_{jj}(z)$ is nonzero. Since $W_{ji}(z) = H_{ji}(z) = z^{-1}$, the estimate of $H_{ji}(z)$ via Procedure 1 is $\hat{H}_{ji}(z) = \frac{W_{ji}(z)}{1-W_{jj}(z)} = \frac{H_{ji}(z)}{1-W_{jj}(z)}$ which is biased.

□

CHAPTER IV

DETECTING DELAYS IN DYNAMIC NETWORKS

While deriving the results of the previous chapters, it was assumed that some partial knowledge about causality or strict causality of the transfer functions of the network was available. This knowledge was embedded in the fact that a recursive graphical representation of the network had to be a priori available.

If such a recursive graphical representation is not available, the application of Procedure 1 to the sets D^- and D^+ given by Theorem 2 leads, in the general case, to an inconsistent estimate of $H_{ji}(z)$. Other network identification methods require analogous information about the locations of strictly causal transfer functions. This knowledge is typically formalized by requiring that there is no algebraic loop for any value of the parameters in the full network parameterization [62], which substantially implies the knowledge of a recursive graphical representation.

Example 12. Consider a simple two-node network with a block diagram depicted in Figure 6.6 (a) and a recursive graphical representation of the network is depicted in Figure 6.6 (b). Suppose the objective is identification of transfer function $H_{ji}(z)$. Note that nodes

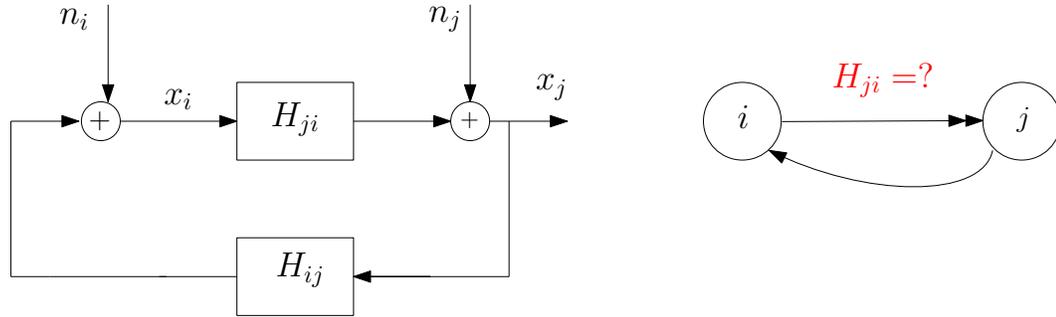


Figure 4.1: (a) Block diagram of a network (b) A recursive graphical representation of the network containing two nodes in a feedback loop

i and j are involved in a feedback loop. Since the edge from i to j is double-headed, we know that the transfer function $H_{ji}(z)$ is strictly causal. This information is crucial for the identification. Indeed, if we include i in D^+ , Algorithm 1 will result in a biased estimate of $H_{ji}(z)$ in general. Namely $\hat{H}_{ji} = \frac{W_{ji}(z)}{1-W_{jj}(z)}$ is a biased estimate of $H_{ji}(z)$ when

$$\mathbb{E}(x_j(t)|I_j(t-1), I_i(t)) = W_{jj}x_j + W_{ji}x_i.$$

However, if we include i in D^- , Algorithm 1 will result in a consistent estimate of $H_{ji}(z)$ in general. Namely $\hat{H}_{ji} = \frac{W_{ji}(z)}{1-W_{jj}(z)}$ is a consistent estimate of $H_{ji}(z)$ when

$$\mathbb{E}(x_j(t)|I_j(t-1), I_i(t-1)) = W_{jj}x_j + W_{ji}x_i.$$

In some cases a recursive graphical representation might not be a priori available. In this chapter, we provide sufficient criteria to obtain a recursive graphical representation in a network with no algebraic loops, from a known graphical representation which is not necessarily recursive. These results have the main advantage of widening not only

the applicability of Theorem 2 but also the applicability of other network identification methods such as the ones described in [62, 66].

The following Lemma will be the basis for the two main results of this chapter.

Lemma 9. *Consider a network with recursive graphical representation $G = (V, E_1, E_2)$ and $i \in pa_G(j)$. Let $Z \cap \{i, j\} = \emptyset$ be a set that j -pointing separates nodes i and j in G . Let G^t be the graph of instantaneous propagations associated to G and let D^- and D^+ be the following two disjoint sets partitioning $Z \cup \{i, j\}$*

- $D^+ := an_{G^t}(j) \cap (Z \cup \{i\})$
- $D^- := (Z \cup \{i, j\}) \setminus D^+$

Let

$$\mathbb{E}(x_j(t) \mid I_{D^-}(t-1), I_{D^+}(t)) = \sum_{r \in D^- \cup D^+} W_{jr}(z)x_r(t). \quad (\text{IV.1})$$

Then, $H_{ji}(z)$ is strictly causal if and only if $W_{ji}(z)$ is strictly causal.

Proof. See the appendix. □

4.1 Detecting Delays

First, we explain the intuition behind the results for delay detection. These results that we will present are a consequence of Pearl-Verma theorem from the theory of graphical models that holds for directed acyclic graphs. In a nutshell, this theorem allows for the identification of zero/non-zero edges in a graphical model with no directed loops. Since our dynamic network has no algebraic loops, it means that there is no sequence of direct feedthroughs from x_j to itself. The results that we will present are a modification of Pearl-Verma theorem applied to the graph of instantaneous propagations, which has no loops

if the underlying network has no algebraic loops. As a result our results are sufficient conditions to detect strictly causal and non-strictly causal transfer functions. Once this information is obtained our causal Wiener approach uses it to prove unbiased identification. Indeed we assume that there are no algebraic loops but we simply do not know which transfer functions are strictly causal in a loop. Instead we are using graphical model techniques to infer them from data when this information is not a priori known.

The first result gives a sufficient criterion to determine if a transfer function is strictly causal directly from observational data.

Theorem 10. *Consider a network with no algebraic loops and with (non-necessarily recursive) graphical representation $G = (V, E_1, E_2)$ and $i \in pa_G(j)$. Let $Z \cap \{i, j\} = \emptyset$ be a set that j -pointing separates nodes i and j in G . Let A, Z^-, Z^+ be a partition of Z such that*

- $Z^- := \{\ell \in Z : \ell \notin an_{G^t}(j)\}$
- $Z^+ := \{k \in Z : k \notin de_{G^t}(j)\} \setminus Z^-$
- $A := Z \setminus (Z^- \cup Z^+)$.

If

$$\lim_{z \rightarrow \infty} W_{ji}(z) = 0 \quad (\text{IV.2})$$

in

$$\mathbb{E}(x_j(t) \mid I_{\{j\} \cup Z^- \cup A^-}(t-1), I_{Z^+ \cup A^+ \cup \{i\}}(t)) = \sum_{r \in Z^- \cup Z^+ \cup A^- \cup A^+ \cup \{i, j\}} W_{jr}(z) x_r(t). \quad (\text{IV.3})$$

for all possible combinations of disjoint A^- and A^+ with $A^- \cup A^+ = A$, then the transfer function $H_{ji}(z)$ is strictly causal.

Proof. See the appendix. □

The following example revisits Example 9 showing that the consistent identification can be achieved even without knowing which transfer functions are strictly causal.

Example 13. Consider a network with a perfect graphical representation as given in Figure 3.18. Suppose though that the information about the strict causality of the transfer functions is not available. Hence, what is known about the network is given by the graphical representation of Figure 4.2. Suppose the objective is to identify the transfer func-

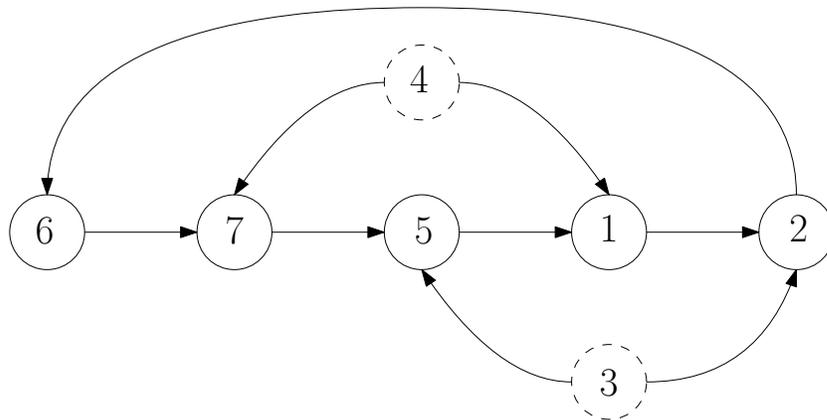


Figure 4.2: The non-recursive graphical representation of the network discussed in Example 13. The nodes 3 and 4 are not measured and act as confounders.

tion $H_{21}(z)$. Assume that nodes 3 and 4 which are depicted with a dashed line are not measured $3, 4 \notin O = \{1, 2, 5, 6, 7\}$. Similar to Example 9 $Z = \{5, 7\}$ is a subset of $O = \{1, 2, 5, 6, 7\}$ that satisfies conditions (i) and (ii) of Theorem 2. However, since the available graphical representation is not recursive, the sets D^+ and D^- cannot be determined. We will show, however, a recursive graphical representation can be obtained using

the results of this section. Applying Theorem 10 on the transfer function $H_{76}(z)$ we get that the set $\{2\}$ 7-pointing separates nodes 6 and 7 and $Z^- = Z^+ = \emptyset$. To consider all possible combinations of A^+ and A^- , we need to consider two cases. In the first case, we have $A^+ = \{2\}$ and $A^- = \{\emptyset\}$. It turns out that in

$$\mathbb{E}(x_7(t) \mid I_7(t-1), I_{2,6}(t)) = \sum_{r \in \{2,6,7\}} W_{7r}(z)x_r(t) \quad (\text{IV.4})$$

the transfer function $W_{76}(z)$ is strictly causal. In the second case, we have $A^+ = \{\emptyset\}$ and $A^- = \{2\}$. Similarly, it turns out that in

$$\mathbb{E}(x_7(t) \mid I_{2,7}(t-1), I_6(t)) = \sum_{r \in \{2,6,7\}} W_{7r}(z)x_r(t) \quad (\text{IV.5})$$

the transfer function $W_{76}(z)$ is strictly causal. Thus, according to Theorem 10 we can conclude that the transfer function $H_{76}(z)$ is strictly causal. This information enables us to obtain a recursive graphical representation for the network, determine the sets $D^+ = \{1, 5, 7\}$ and $D^- = \{2\}$ and consistently estimate the transfer function $H_{21}(z)$ using Procedure 1.

4.2 Detecting Feedthroughs

In the previous section we saw how we can infer if a transfer function is strictly causal. This second result instead gives a sufficient criterion to determine if a transfer function has a nonzero feedthrough.

Theorem 11. *Consider a network with no algebraic loops and with (non-necessarily recursive) graphical representation $G = (V, E_1, E_2)$ and $i \in pa_G(j)$. Let $Z \cap \{i, j\} = \emptyset$ be a set that*

- (i) *j -pointing separates i and j in G*
- (ii) *i -pointing separates i and j in G*

Let Z^-, Z^+, A be a partition of Z such that

- $Z^- := \{\ell \in Z : \ell \notin an_{G'}(j)\}$
- $Z^+ := \{k \in Z : k \notin de_{G'}(j)\} \setminus Z^-$
- $A = Z \setminus (Z^- \cup Z^+)$.

If

$$\lim_{z \rightarrow \infty} W_{ji}(z) \neq 0 \quad (\text{IV.6})$$

in

$$\mathbb{E}(x_j(t) \mid I_{\{j\} \cup Z^- \cup A^-}(t-1), I_{Z^+ \cup A^+ \cup \{i\}}(t)) = \sum_{r \in Z^- \cup Z^+ \cup A^- \cup A^+ \cup \{i, j\}} W_{jr}(z) x_r(t) \quad (\text{IV.7})$$

and

$$\lim_{z \rightarrow \infty} W_{ij}(z) \neq 0 \quad (\text{IV.8})$$

in

$$\mathbb{E}(x_i(t) \mid I_{\{i\} \cup Z^- \cup A^-}(t-1), I_{Z^+ \cup A^+ \cup \{j\}}(t)) = \sum_{\ell \in Z^- \cup Z^+ \cup A^- \cup A^+ \cup \{i,j\}} W_{ir}(z) x_\ell(t) \quad (\text{IV.9})$$

for all possible combinations of disjoint A^- and A^+ with $A^- \cup A^+ = A$, then either the transfer function $H_{ji}(z)$ is not strictly causal or the transfer function $H_{ij}(z)$ is not strictly causal.

Proof. See the appendix. □

Theorems 10 and 11 provide sufficient conditions to determine if a transfer function in the network is strictly causal or not, respectively. These conditions are only sufficient. Hence, there could be situations where their application would be inconclusive.

However, in several scenarios the information obtained from these two theorems, might be enough to determine a recursive graphical representation from a non-recursive one. The following example illustrates a situation when this occurs.

Example 14. Consider a network with an unknown recursive graphical representation \overline{G} depicted in Figure 4.3 (a). Assume, instead, that the less informative non-recursive graphical representation G of Figure 4.3 (b) is available, even though the network is known not to have any algebraic loops.

As can be seen in \overline{G} , transfer functions $H_{43}(z)$, $H_{32}(z)$ and $H_{76}(z)$ are strictly causal. This information, however, is not available from G . The objective is identification of transfer function $H_{21}(z)$ given the topology and outputs of nodes $O = \{1, 2, 3, 4, 6, 7\} \subset V = \{1, 2, 3, 4, 5, 6, 7\}$. Node 5 is not measured. The set $\{4, 6\}$ satisfies graphical conditions

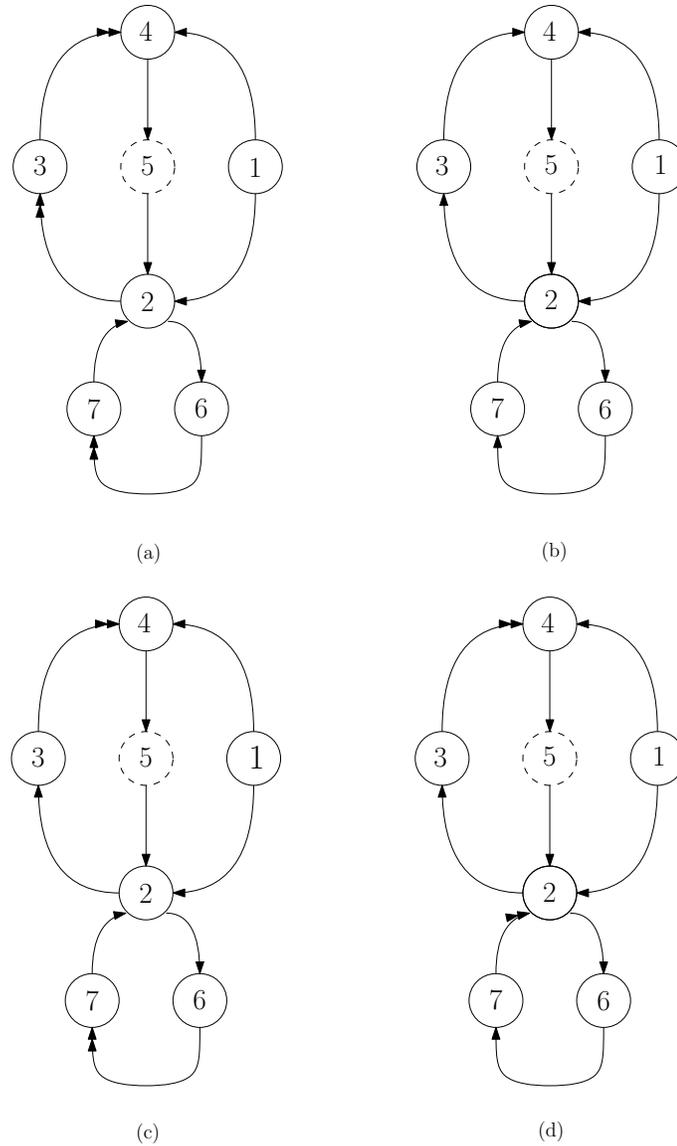


Figure 4.3: (a) The unknown recursive graphical representation of a network discussed in Example 14; (b) The given non-recursive graphical representation of the network discussed in Example 14; The application of Theorems 10 and 11 allows one to conclude that either (c) or (d) is a valid graphical representation of the network. The application of Theorem 2 with $Z = \{4, 6\}$ leads to the same sets $D^- = \{2, 6\}$ and $D^+ = \{1, 4\}$ for the identification of $H_{21}(z)$ via Procedure 1.

(i) and (ii) of Theorem 2. However, since there is no information available about the locations of strictly causal transfer functions, standard techniques cannot be applied to identify $H_{12}(z)$. For instance, in order to apply Procedure 1 to identify $H_{12}(z)$ we need to know a

recursive graphical representation to determine D^- and D^+ . Theorems 10 and 11 instead could be effectively applied in this case. If we consider the set $\{2\}$ we notice that such a set 4-pointing separates nodes 3 and 4. Hence, we can apply Theorem 10 on transfer function $H_{43}(z)$. Since $Z^- = Z^+ = \emptyset$, to consider all possible combinations of A^+ and A^- , we need to consider two cases. In the first case, we have $A^+ = \{2\}$ and $A^- = \{\emptyset\}$. In the second case, we have $A^+ = \{\emptyset\}$ and $A^- = \{2\}$. Taking similar steps as in Example 13, it turns out that the transfer function $H_{43}(z)$ is strictly causal.

On the other hand, if we consider the set $\{7\}$ we notice that such a set 2-pointing and 6-pointing separates nodes 2 and 6. Hence, we can apply Theorem 11 on transfer function $H_{62}(z)$. Since $Z^- = Z^+ = \emptyset$, to consider all possible combinations of A^+ and A^- , again we need to consider two cases. In the first case, we have $A^+ = \{7\}$ and $A^- = \{\emptyset\}$. It turns out that in

$$\mathbb{E}(x_6(t) \mid I_6(t-1), I_{2,7}(t)) = \sum_{r \in \{2,6,7\}} W_{6r}(z)x_r(t) \quad (\text{IV.10})$$

the transfer function $W_{62}(z)$ is not strictly causal and in

$$\mathbb{E}(x_2(t) \mid I_2(t-1), I_{6,7}(t)) = \sum_{r \in \{2,6,7\}} W_{2r}(z)x_r(t) \quad (\text{IV.11})$$

the transfer function $W_{26}(z)$ is not strictly causal. In the second case, we have $A^+ = \{\emptyset\}$ and $A^- = \{7\}$. It turns out that in

$$\mathbb{E}(x_6(t) \mid I_{6,7}(t-1), I_2(t)) = \sum_{r \in \{2,6,7\}} W_{6r}(z)x_r(t) \quad (\text{IV.12})$$

the transfer function $W_{62}(z)$ is not strictly causal and in

$$\mathbb{E}(x_2(t) \mid I_{2,7}(t-1), I_6(t)) = \sum_{r \in \{2,6,7\}} W_{2r}(z)x_r(t) \quad (\text{IV.13})$$

the transfer function $W_{26}(z)$ is not strictly causal. Thus, it follows Theorem 11 that the transfer function $H_{62}(z)$ is not strictly causal.

Since the network has no algebraic loops, we can conclude that either the transfer function $H_{76}(z)$ or $H_{27}(z)$ or both are strictly causal. As a consequence, the graph in Figure 4.3 (c) or the graph in Figure 4.3 (d) is a graphical representation of network and they are both recursive. Therefore, Theorem 2 can be applied to either graph leading to the same choice of $D^+ = \{1, 4\}$ and $D^- = \{2, 6\}$ for the consistent identification of $H_{21}(z)$. Note that Theorem 11 can also be applied on the transfer function $H_{27}(z)$ revealing that it is not strictly causal. As a consequence, since the network has no algebraic loops, it can be inferred that the transfer function $H_{76}(z)$ is strictly causal.

4.3 Proofs related to Chapter IV

4.3.1 Proof of Lemma 9

Proof. Let $G' = (V, (E_1 \cup E_2) \setminus \{i \rightarrow j\})$ be the standard directed graph associated to G after removing the edge $i \rightarrow j$. Define $E = E_1 \cup E_2$. Also define a new processes $x_q(t) = x_j(t) - H_{ji}(z)x_i(t)$ and $x_w(t) = x_i(t)$. We are going to define a new network $\mathcal{G}'' = (H'', n'')$ with all the variables of the original network and the additional variables x_q and x_w . Let

$$H''_{jq}(z) = 1,$$

$$H''_{jw}(z) = H_{ji}(z),$$

$$H''_{qr}(z) = H_{jr}(z) \text{ for } r \in \text{pa}_G(j) \setminus i,$$

$$H''_{wi}(z) = H_{ji}(z)$$

$$H''_{k\ell}(z) = H_{k\ell}(z) \text{ in all other cases}$$

and $n''_q = n_j$, $n''_w = 0$, $n''_j = 0$, and $n''_k = n_k$ in all other cases. From G we can obtain a graphical representation for \mathcal{G}'' given by $G'' = (V'', E''_1, E''_2)$ in the following way. Let $K_1 := \{k | k \neq i \text{ and } k \rightarrow j \in E_1\}$ be the set of single-headed parents of j in G that are not node i and $K_2 := \{k | k \neq i \text{ and } k \rightarrow j \in E_2\}$ be the set of double-headed parents of j in

G that are not node i . Then

$$V'' := V \cup \{q, w\}$$

$$E_1'' := E_1 \cup \{q \rightarrow j, w \rightarrow j, i \rightarrow w\} \cup \{k \rightarrow q | k \in K_1\} \setminus \{k \rightarrow j | k \in K_1\}$$

$$E_2'' := E_2 \cup \{k \rightarrow q | k \in K_2\} \setminus \{k \rightarrow j | k \in K_2\}.$$

Namely, in $G'' = (V'', E_1'', E_2'')$, the additional node q is placed in between j and its original parents in G that are not the node i , and node w is placed in between nodes i and j . Also, notice that since G is recursive, G'' is trivially recursive, as well. Observe that $i \in D^+$ and $j \in D^-$, hence D^- is never empty. Let S_0 be the set containing all descendants of the nodes in D^- in G'' , $S_0 := \text{de}_{G''}(D^-)$. Since G'' is recursive and D^- is not empty, by Lemma 5 there exists a node $w_0 \in S_0$ such that $\text{ch}_{G''}(w_0) = \emptyset$. Apply Lemma 4 on w_0 obtaining a new network with recursive graphical representation G_1 . Define $S_1 := \text{de}_{G_1}(D^-)$. From Lemma 4 it follows that $S_1 = S_0 \setminus \{w_0\}$. Again, by Lemma 5 there exists a node $w_1 \in S_1$ such that $\text{ch}_{G_1}(w_1) = \emptyset$. Apply Lemma 4 on w_1 , represent the resulting network with G_2 and let $S_2 := \text{de}_{G_2}(D^-)$. Again, from Lemma 4 it follows that $S_2 = S_1 \setminus \{w_1\}$. Repeat the procedure N times, till $S_N = \emptyset$, for $N \in \mathbb{Z}$, the number of elements in S_0 . Let (H_N, n_N) be the resulting network with recursive graphical representation G_N . Let $Z^- := S_0 \cap Z$ and $Z^+ := Z \setminus Z^-$. Since S_0 contains all the elements in D^- we have that $D^- = Z^- \cup \{j\}$. Consequently, $D^+ = Z^+ \cup \{i\}$. Also, it follows from Lemma 4 that in (H_N, n_N) the output processes of nodes k in Z^- are now $x_k(t-1)$ while the output processes of nodes i (since $i \in D^+$) and ℓ in Z^+ remain unchanged, $x_\ell(t)$. Apply Lemma 4 one more time on w in G_N to get \bar{G} . By Lemma 6, we have that node q is d -separated from node i given

$\{w, j\} \cup Z^- \cup Z^+$ in \bar{G} . Note that the output process of node w in \bar{G} is $x_i(t-1)$. Therefore, applying Theorem 24 in [64] yields

$$x_q(t) \perp x_i(t) \mid I_{\{i,j\} \cup Z^-}(t-1), I_{Z^+}(t). \quad (\text{IV.14})$$

Thus, we have that

$$\mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) = \mathbb{E}(x_q(t) \mid I_{Z^+}(t), I_{\{i,j\} \cup Z^-}(t-1)). \quad (\text{IV.15})$$

Therefore, we can write

$$\begin{aligned} \mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1)) &= \\ \mathbb{E}(x_j(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) &= \\ \mathbb{E}(x_q(t) + H_{ji}(z)x_i(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) &= \\ H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_{\{i\} \cup Z^+}(t), I_{\{j\} \cup Z^-}(t-1)) &= \\ H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_{Z^+}(t), I_{\{i,j\} \cup Z^-}(t-1)) &= \\ H_{ji}(z)x_i(t) + \sum_{r \in \{i,j\} \cup Z^- \cup Z^+} F_{qr}(z)x_r(t) &= \\ H_{ji}(z)x_i(t) + F_{qi}(z)x_i(t) + F_{qj}(z)x_j(t) + \sum_{r \in Z^- \cup Z^+} F_{qr}(z)x_r(t) &= \\ [H_{ji}(z) + F_{qi}(z)]x_i(t) + F_{qj}(z)x_j(t) + \sum_{r \in Z^- \cup Z^+} F_{qr}(z)x_r(t) & \quad (\text{IV.16}) \end{aligned}$$

where $F_{qi}(z)$, $F_{qj}(z)$, and $F_{qr}(z)$, $r \in Z^-$ are strictly causal transfer functions. Since the power spectral density matrix associated with (x_i, x_j, x_Z) is non-singular, comparing the

two expressions for $\mathbb{E}(x_j(t) \mid I_{D^+}(t), I_{D^-}(t-1))$ we can conclude $W_{jk}(z) = F_{qk}(z)$ for all $k \in Z^- \cup Z^+$, $W_{jj}(z) = F_{qj}(z)$ and $W_{ji}(z) = H_{ji}(z) + F_{qi}(z)$. Since $F_{qi}(z)$ is strictly causal, $W_{ji}(z) = H_{ji}(z) + F_{qi}(z)$ is strictly causal if and only if $H_{ji}(z)$ is strictly causal. Also, $W_{ji}(z)$ is not strictly causal if and only if $H_{ji}(z)$ is not strictly causal. \square

4.3.2 Proof of Theorem 10

We provide two independent proofs for Theorem 10. The first proof is based on Lemma 9.

Proof. Let $G^p = (V, E_1^p, E_2^p)$ be the perfect graphical representation of the network. Since the network has no algebraic loops, G^p is recursive. Build a new graphical representation $\bar{G} = (V, \bar{E}_1, \bar{E}_2)$ of the network by adding single-headed edges from all nodes $k \in Z^+$ to j in G^p . That is, $\bar{E}_2 = E_2^p$ and $\bar{E}_1 = E_1^p \cup_{k \in Z^+} \{k \rightarrow j\}$. This implies that $Z^+ \subseteq \text{an}_{\bar{G}^\dagger}(j)$. Note that \bar{G} is recursive because for all edges $k \rightarrow j$ that we added to E_1^p to obtain \bar{E}_1 , we have that $k \notin \text{de}_{\bar{G}^\dagger}(j)$. Assume, by contradiction that $H_{ji}(z)$ is not strictly causal. Since $Z^- := \{\ell \in Z : \ell \notin \text{an}_{\bar{G}^\dagger}(j)\}$, we have that $Z^- \cap \text{an}_{\bar{G}^\dagger}(j) = \emptyset$. Since Z^- does not contain any ancestor of Z^+ in \bar{G}^\dagger , it also follows that $Z^- \cap \text{an}_{\bar{G}^\dagger}(j) = \emptyset$. Hence, applying Lemma 9 on \bar{G} , we get that $Z^- \subset D^-$. On the other hand, since $Z^+ \subseteq \text{an}_{\bar{G}^\dagger}(j)$, we have that $Z^+ \subset D^+$. Since i is a parent of j in \bar{G}^\dagger , which is a recursive graph, there is one choice of A_1 and A_2 where $D^- = Z^- \cup A_1 \cup \{j\}$ and $D^+ = Z^+ \cup A_2 \cup \{i\}$ meeting the conditions of Lemma 9 on \bar{G} . For those A_1 and A_2 Lemma 9 gives that $H_{ji}(z)$ is strictly causal which is a contradiction. \square

4.3.2.1 Lemma 12

For the second proof of Theorem 10 we need the following lemma.

Lemma 12. *Consider a LDIM $(H(z), n)$ with no algebraic loops and with graphical representation $G = (V, E_1, E_2)$. Assume i and j are j -pointing separated by Z in G . Let Q be a set of nodes such that $Q \cap (Z \cup \{i, j\}) = \emptyset$. Let $G^r = (V, E_1^r, E_2^r)$ be a graph obtained from G in the following way.*

1. *if there is at least one dipath from k to $\ell \in V \setminus Q$ in G with all internal nodes in Q and all single-headed edges, then $k \rightarrow \ell \in V \setminus Q$ is in E_1^r*
2. *if all the dipaths from k to ℓ in G with all internal nodes in Q have at least one double-headed edge, then $k \rightarrow \ell$ is in E_2^r*
3. *if there is no dipath from k to ℓ in G with all internal nodes in Q , or $\ell \in Q$, then $k \rightarrow \ell \notin E_1^r \cup E_2^r$*

We have that

- G^r is a graphical representation of the network $(H^r(z), n)$ obtained by marginalizing Q , (see Lemma 15 in [64])
- $Z \cap de_{G^t}(j) = Z \cap de_{G^{rt}}(j)$
- $Z \cap an_{G^t}(j) = Z \cap an_{G^{rt}}(j)$
- $H_{ji}^r(z)$ is strictly causal if and only if $H_{ji}(z)$ is strictly causal
- i and j are j -pointing separated in G^r

Proof. Following the proof of property 2) of Lemma 15 in [64] (Node Marginalization Lemma), we find that the graph G^r is a graphical representation of the reduced model. In the graph G^r , the nodes in Q are not descendants of any node since they have no incoming

edges. Furthermore, we have that $\text{de}_{G^r}(k) = \text{de}_G(k) \setminus Q$ for all $k \notin Q$. Indeed, if there is a dipath π in G from k to $\ell \notin Q$, replace every sequence in π of the form $v \rightarrow a_1 \rightarrow \dots \rightarrow a_M \rightarrow w$ where $\{a_1, \dots, a_M\} \subseteq Q$ with $v \rightarrow w \in E^r$ to obtain the dipath π^r . We have that π^r is a dipath in G^r from k to ℓ .

To establish that $Z \cap \text{de}_{G^t}(j) = Z \cap \text{de}_{G^{rt}}(j)$ we show that $Z \cap \text{de}_{G^t}(j) \subseteq Z \cap \text{de}_{G^{rt}}(j)$ and $Z \cap \text{de}_{G^t}(j) \supseteq Z \cap \text{de}_{G^{rt}}(j)$. Suppose $y \in Z \cap \text{de}_{G^t}(j)$. That is, there exists a dipath π with all single-headed edges from j to y in G . If no internal node on π is in Q , then the very same path exists in G^r . If some of the nodes on π are also in Q , then π has the form $j \cdots k \rightarrow a_1 \rightarrow \dots a_m \rightarrow \ell \cdots \rightarrow y$, where k and ℓ are not in Q and $a_1 \cdots a_m \in Q$. Then by condition 1) of the Lemma, there exists a single-headed edge from k to ℓ in G^r that can be used to replace $k \rightarrow a_1 \rightarrow \dots a_m \rightarrow \ell$ to $k \rightarrow \ell$. We can iterate this procedure to eliminate all internal nodes in Q . Eventually, we find a path $\pi' = j \cdots k \rightarrow \ell \cdots \rightarrow y$ with all single-headed edges from j to y in G^r . Therefore, we have that $y \in Z \cap \text{de}_{G^{rt}}(j)$, giving $Z \cap \text{de}_{G^t}(j) \subseteq Z \cap \text{de}_{G^{rt}}(j)$. Now suppose, $y \in Z \cap \text{de}_{G^{rt}}(j)$. That is, there exists a dipath with all single-headed edges from j to y in G^r . Then, it follows from condition 1) of the Lemma that there exists a dipath with all single-headed edges from j to y in G . Therefore, we have that $y \in Z \cap \text{de}_{G^t}(j)$, giving $Z \cap \text{de}_{G^t}(j) \supseteq Z \cap \text{de}_{G^{rt}}(j)$. The assertion $Z \cap \text{an}_{G^t}(j) = Z \cap \text{an}_{G^{rt}}(j)$ can be established using analogous steps swapping the roles of j and y .

We prove that i and j are j -pointing separated in G^r by contradiction. By contradiction assume that there exists a path π^r between i and j that is not blocked by Z in G^r . Construct a new path π from π^r in the following way: for all edges $k \rightarrow \ell$ in π^r such that $k \rightarrow \ell \notin E$, replace $k \rightarrow \ell$ with sequence of edges $k \rightarrow a_1 \rightarrow \dots \rightarrow a_M \rightarrow \ell$ with

$\{a_1, \dots, a_M\} \subseteq Q$. Such a sequence of edges $k \rightarrow a_1 \rightarrow \dots \rightarrow a_M \rightarrow \ell$ is always a dipath in G because of the way G^r has been constructed.

Observe that π is a path in G . Furthermore, π and π^r have the same colliders. We now want to show that π is not blocked by Z in G . If π^r has no colliders, then π has no colliders either. Hence, since π^r is not blocked by Z , π is not blocked by Z leading to a contradiction. Consider, then, the case where π^r has all active colliders, and no non-colliders in Z . Because of the way π was obtained, π has no non-colliders in Z . Also, a collider c in π^r is a collider in π . Because $\text{de}_{G^r}(c) = \text{de}_G(c) \setminus Q$ and $Z \cap Q = \emptyset$ we have $\text{de}_{G^r}(c) \cap Z = \text{de}_G(c) \cap Z$. Hence a collider on π^r activated by Z in G^r is also a collider on π activated by Z in G . This again leads to a contradiction. \square

The following is the second proof of Theorem 10 based on Lemma 12.

Proof. First we consider the scenario where all j -loops are blocked by Z . Let $G^p = (V, E_1^p, E_2^p)$ be the perfect graphical representation of the network. Since the network has no algebraic loops, G^p is recursive. Build a new graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding single-headed edges from all nodes $k \in Z^+$ to j in G^p . That is, $\overline{E}_2 = E_2^p$ and $\overline{E}_1 = E_1^p \cup_{k \in Z^+} \{k \rightarrow j\}$. This implies that $Z^+ \subseteq \text{an}_{\overline{G}^\dagger}(j)$. Note that \overline{G} is recursive because for all edges $k \rightarrow j$ that we added to E_1^p to obtain \overline{E}_1 , we have that $k \notin \text{de}_{\overline{G}^\dagger}(j)$.

Assume, by contradiction that $H_{ji}(z)$ is not strictly causal. Since $Z^- := \{\ell \in Z : \ell \notin \text{an}_{\overline{G}^\dagger}(j)\}$, we have that $Z^- \cap \text{an}_{\overline{G}^\dagger}(j) = \emptyset$. Since Z^- does not contain any ancestor of Z^+ in G^\dagger , it also follows that $Z^- \cap \text{an}_{\overline{G}^\dagger}(j) = \emptyset$. Hence, applying Theorem 2 on \overline{G} , we get that $Z^- \subset D^-$. On the other hand, since $Z^+ \subseteq \text{an}_{\overline{G}^\dagger}(j)$, we have that $Z^+ \subset D^+$.

Since i is a parent of j in \overline{G}^f , which is a recursive graph, there is one choice of A_1 and A_2 where $D^- = Z^- \cup A_1 \cup \{j\}$ and $D^+ = Z^+ \cup A_2 \cup \{i\}$ meeting the conditions of Theorem 2 on \overline{G} . For those A_1 and A_2 Theorem 2 gives

$$W_{ji}(z) = (1 - W_{jj}(z))H_{ji}(z). \quad (\text{IV.17})$$

Since $W_{jj}(z)$ is strictly causal, we necessarily have that $H_{ji}(z)$ is strictly causal which is a contradiction.

If Z does not block all j -loops, then marginalize the network $(H(z), n)$ with respect to the nodes $A = V \setminus (Z \cup \{i, j\})$ and obtain the reduced network $(H^r(z), n^r)$ as in Lemma 12. Since the original network $(H(z), n)$ has no algebraic loops, the reduced network has no algebraic loops either. Again because of Lemma 12, all j -pointing paths between i and j that are not the edge $i \rightarrow j$ are blocked by Z in G^r . Furthermore, since the only nodes in G^r are $Z \cup \{i, j\}$, all j -loops are blocked by $Z \cup \{i\}$. Hence, we can apply the same argument to the reduced network (H^r, n^r) and conclude that $H_{ji}^r(z)$ is strictly causal. Again, because of Lemma 12, the transfer function $H_{ji}(z)$ is going to be strictly causal if and only if $H_{ji}^r(z)$ is strictly causal proving the assertion. \square

4.3.3 Proof of Theorem 11

We provide two independent proofs for Theorem 10. The first proof is based on Lemma 9.

Proof. Without any loss of generality, assume that $\{j \rightarrow i\} \in E_1$, otherwise we can redefine E_1 and E_2 respectively as $E_1 \cup \{j \rightarrow i\}$, and $E_2 \setminus \{j \rightarrow i\}$, since this would still

give us a (non-necessarily recursive) graphical representation of the network where the set Z still satisfies the Theorem's assumption. Let $G^p = (V, E_1^p, E_2^p)$ be the perfect graphical representation of the network. Since the network has no algebraic loops, G^p is recursive. Since G^p is recursive it holds that (i) every dipath from j to i has at least a double headed edge or (ii) every dipath from i to j has at least a double headed edge. Consider first case (i). Build a graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding the single headed edge $i \rightarrow j$ and single-headed edges from all nodes $k \in Z^+$ to j in G^p . That is, $\overline{E}_2 = E_2^p$ and $\overline{E}_1 = E_1^p \cup \{i \rightarrow j\} \cup_{k \in Z^+} \{k \rightarrow j\}$. This implies that $Z^+ \cup \{i\} \subseteq \text{an}_{\overline{G}}(j)$. Since Z^- does not contain any ancestor of Z^+ in G^f , it also follows that $Z^- \cap \text{an}_{\overline{G}}(j) = \emptyset$. Observe also that because of (i) and the definition of Z^+ , \overline{G} is recursive. Hence, by applying Lemma 9 on \overline{G} , we get that there exist disjoint A_1 and A_2 such that $D^- = A_1 \cup Z^- \cup \{j\}$ and $D^+ = A_2 \cup Z^+ \cup \{i\}$ giving a non-strictly causal estimate of the transfer function $H_{ji}(z)$. Since for all choices of A_1 and A_2 the transfer function estimate that would result from Equation (IV.6) and Equation (IV.7) has a non-zero feedthrough component, $H_{ji}(z)$ needs to be non-strictly causal under scenario (i). If instead scenario (ii) holds, we build a graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding the single headed edge $j \rightarrow i$ and single-headed edges from all nodes $k \in Z^+$ to i in G^p . By repeating steps similar to scenario (i) with reversed roles for the nodes i and j , we would find that, in scenario (ii), because of Equation (IV.8) and Equation (IV.9) the transfer function $H_{ij}(z)$ needs to be non-strictly causal. Now, we do not know if scenario (i) or scenario (ii) holds, thus, we can only conclude that either $H_{ji}(z)$ is strictly causal or $H_{ij}(z)$ is strictly causal.

□

The second proof of Theorem 11 is based on Lemma 12.

Proof. Without any loss of generality, assume that $\{j \rightarrow i\} \in E_1$, otherwise we can redefine E_1 and E_2 respectively as $E_1 \cup \{j \rightarrow i\}$, and $E_2 \setminus \{j \rightarrow i\}$, since this would still give us a (non-necessarily recursive) graphical representation of the network. Let $G^p = (V, E_1^p, E_2^p)$ be the perfect graphical representation of the network. Since the network has no algebraic loops, G^p is recursive. Since G^p is recursive it holds that (i) every dipath from j to i has at least a double headed edge or (ii) every dipath from i to j has at least a double headed edge. Consider first case (i). As in the proof of Theorem 10, we first assume that $Z \cup \{i\}$ blocks all j -loops. Then, build a graphical representation $\bar{G} = (V, \bar{E}_1, \bar{E}_2)$ of the network by adding the single headed edge $i \rightarrow j$ and single-headed edges from all nodes $k \in Z^+$ to j in G^p . That is, $\bar{E}_2 = E_2^p$ and $\bar{E}_1 = E_1^p \cup \{i \rightarrow j\} \cup_{k \in Z^+} \{k \rightarrow j\}$. This implies that $Z^+ \cup \{i\} \subseteq \text{an}_{\bar{G}^f}(j)$. Since Z^- does not contain any ancestor of Z^+ in G^f , it also follows that $Z^- \cap \text{an}_{\bar{G}^f}(j) = \emptyset$. Observe also that because of (i) and the definition of Z^+ , \bar{G} is recursive. Hence, by applying Theorem 2 on \bar{G} , we get that there exist disjoint A_1 and A_2 such that $D^- = A_1 \cup Z^- \cup \{j\}$ and $D^+ = A_2 \cup Z^+ \cup \{i\}$ giving a consistent estimate of the transfer function $H_{ji}(z)$. Since for all choices of A_1 and A_2 the transfer function estimate that would result from Equation (IV.6) and Equation (IV.7) has a non-zero feedthrough component, $H_{ji}(z)$ needs to be non-strictly causal under scenario (i) when all j -loops are blocked by $Z \cup \{i\}$. If $Z \cup \{i\}$ does not block all j -loops then marginalize the network $(H(z), n)$ with respect to the nodes $A = V \setminus (Z \cup \{i, j\})$ and obtain the reduced network $(H^r(z), n^r)$ as in Lemma 12. Since the original network $(H(z), n)$ has no algebraic loops, the reduced network has no algebraic loops either. Again because

of Lemma 12, all j -pointing paths between i and j that are not the edge $i \rightarrow j$ are blocked by Z in G^r . Furthermore, since the only nodes in G^r are $Z \cup \{i, j\}$, all j -loops are blocked by $Z \cup \{i\}$. Hence we can apply the same argument to the reduced network (H^r, n^r) and conclude that $H_{ji}^r(z)$ is not strictly causal. Again, because of Lemma 12, the transfer function $H_{ji}(z)$ is going to be strictly causal if and only if $H_{ji}^r(z)$ is strictly causal proving the assertion.

If instead scenario (ii) holds, we build a graphical representation $\overline{G} = (V, \overline{E}_1, \overline{E}_2)$ of the network by adding the single headed edge $j \rightarrow i$ and single-headed edges from all nodes $k \in Z^+$ to i in G^p . By repeating steps similar to scenario (i) with reversed roles for the nodes i and j , we would find that, in scenario (ii), because of Equation (IV.8) and Equation (IV.9) the transfer function $H_{ij}(z)$ needs to be non-strictly causal.

Now, we do not know if scenario (i) or scenario (ii) holds, thus, we can only conclude that either $H_{ji}(z)$ is strictly causal or $H_{ij}(z)$ is strictly causal.

□

CHAPTER V

OPTIMAL OBSERVATIONS FOR CONSISTENT IDENTIFICATION IN DYNAMIC NETWORKS

The choice of the predictors set Z in Theorem 2 is not unique and different choices can have different costs of observations. Since the conditions of Theorem 2 are sufficient and necessary, they make it possible to search for an optimal set Z^* guaranteeing consistent identification. That is, if $\mathcal{Z}(O)$ is a set of all sets $Z \subset O$ that satisfy conditions (i) and (ii) of Theorem 2 and if the objective is consistent identification of the transfer function $H_{ji}(z)$ while minimizing the cost function $C(Z)$, we can search for an optimal element Z^* in $\mathcal{Z}(O)$.

For a network with a recursive graphical representation G and no algebraic loops, Theorem 2 allows this problem to be cast as

$$Z^* = \arg \min_{Z \in \mathcal{Z}(O)} C(Z). \quad (\text{V.1})$$

The optimization problem (V.1) is equivalent to the search of an optimal element in the set $\mathcal{Z}(O)$. For an arbitrary cost $C(Z)$, Z^* can be found by an exhaustive search over $\mathcal{Z}(O)$. However, more efficient ways to find Z^* could be devised if $C(Z)$ has a more specific structure. For instance, the cost of measuring the variables in Z could be given by the sum of the individual costs for each variable:

$$C(Z) = \sum_{k \in Z} c_k. \quad (\text{V.2})$$

where the cost of observing node $k \in V$ is $c_k \geq 0$. In such a case, the following optimization problem could be formulated.

Problem 2. *Consider a network $\mathcal{G} = (H, n)$ where n are unknown mutually independent inputs and measuring node $k \in V$ has a cost $c_k \geq 0$. Given a graphical representation G of \mathcal{G} and a set $O \supseteq \{i, j\}$ of measurable outputs, find a set $Z^* \subset O$ of auxiliary variables such that Algorithm 1 guarantees a consistent identification of the transfer function $H_{ji}(z)$, minimizing the cost function (V.2).*

The following example illustrates how strategies analogous to "branch and bound" could be applied, to find the optimal set of predictors Z^* .

Example 15. *Consider a network with the recursive graphical representation depicted in Figure 1.5.*

The objective is the identification of the transfer function $H_{21}(z)$. Assuming that we are applying Procedure 1 for identification, we would like to find an optimal set of predictors minimizing the cost function (V.2). The cost J_r of observing each node $r \in V$ is reported in Table 1.1.

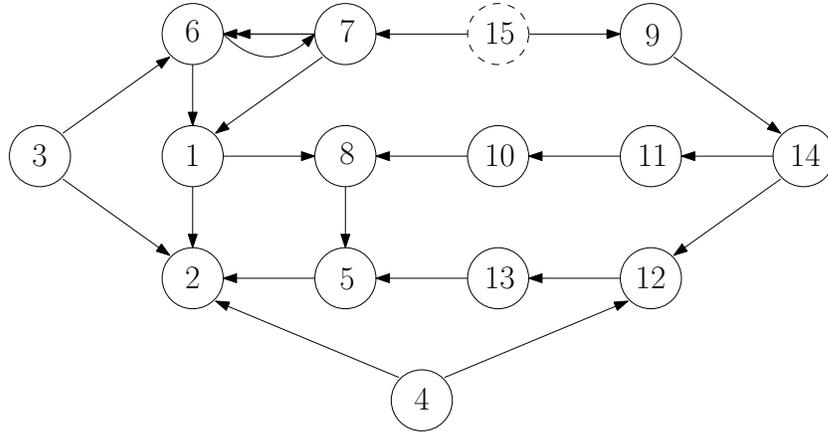


Figure 5.1: The graphical representation of the network discussed in Example 15 for the determination of the set of predictors with minimal cost.

Table 5.1: Costs of observing each node in Example 15

node r	J_r						
3	90	6	20	9	10	12	60
4	60	7	60	10	30	13	20
5	20	8	90	11	120	14	40

A trivial choice for the set Z is all the parents of node 2 except 1 (see Proposition 1), with the cost $C(\{3, 4, 5\}) = 90 + 60 + 20 = 170$.

In particular, node 3 blocks the 2-pointing paths containing the edge $3 \rightarrow 2$. The 2-pointing path $\pi_1 = \{1 \leftarrow 6 \leftarrow 3 \rightarrow 2\}$ blocked by node 3 can also be blocked by node 6 with the lower cost $J_6 = 20 < 90 = J_3$.

However, if selected, node 6 will act as an activated collider on the 2-pointing path $\pi_2 = \{1 \leftarrow 7 \rightarrow 6 \leftarrow 3 \rightarrow 2\}$. π_2 can be blocked by node 7 with the cost $J_7 = 60$.

Since node 6 blocks π_1 and nodes 6 and 7 are adjacent on every other 2-pointing path containing the edge $3 \rightarrow 2$ we can replace node 3 with nodes 6 and 7 and still satisfy the assumptions of Theorem 2. Since $J_6 + J_7 = 80 < 90 = J_3$, it is always more efficient to measure nodes 6 and 7 instead of node 3.

Hence, node 3 is in no optimal set of auxiliary variables. Since node 3 is not in Z^* , blocking π_1 and π_2 requires $6, 7 \in Z^*$.

Thus, the set $\{6, 7, 4, 5\}$ with the cost $C(\{4, 5, 6, 7\}) = J_4 + J_5 + J_6 + J_7 = 60 + 20 + 20 + 60 = 160$ also satisfies the conditions of Theorem 2. The cost of $\{6, 7, 4, 5\}$ can be reduced only by replacing nodes 4 and 5 with some other nodes.

Since $J_4 + J_5 = 60 + 20 = 80$, nodes 4 and 5 cannot be replaced by any node with a cost higher than 80 without increasing the cost of observations. Thus, using this "branch and bound" argument all sets containing nodes 8 and 11 can be ignored in the search for Z^* .

Since node 8 is not in Z^* and the 2-pointing path $1 \rightarrow 8 \rightarrow 5 \rightarrow 2$ needs to be blocked, node 5 is definitely in the optimal solution Z^* .

Note that node 5 is a descendant of nodes 8 and 12. Therefore, nodes 5, 8, and 12 act as activated colliders on the 2-pointing paths $\pi_3 = \{1 \rightarrow 8 \leftarrow 10 \leftarrow 11 \leftarrow 14 \rightarrow 12 \leftarrow 4 \rightarrow 2\}$ and $\pi_4 = \{1 \rightarrow 8 \rightarrow 5 \leftarrow 13 \leftarrow 12 \leftarrow 4 \rightarrow 2\}$. Node 4 blocks the 2-pointing paths, π_3 and π_4 . Since node 9 cannot block π_3 and π_4 it does not belong to Z^* .

Thus, the optimal set $Z^* \supset \{5, 6, 7\}$ can be found by searching for any possible subset of nodes $\{10, 12, 13, 14\}$ that blocks π_3 and π_4 and costs less than node 4. Table 5.2 lists such different choices.

Table 5.2: Costs of different predictor sets in Example 15

	Z	$C(Z)$
1	$\{5, 6, 7, 13, 14\}$	160
2	$\{5, 6, 7, 13, 10\}$	150
3	$\{5, 6, 7, 12, 14\}$	190
4	$\{5, 6, 7, 12, 10\}$	160

As can be seen, $Z^* = \{5, 6, 7, 10, 13\}$ since it has the minimum cost. Thus, applying Procedure 1 with $D^+ = \{1, 5, 6, 7, 10, 13\}$ and $D^- = \{2\}$ leads to a consistent estimate of $H_{21}(z)$ with minimum cost.

5.1 Optimal Selection of Observations for Identification of a Single Module

In this section we show that it is possible to design an algorithm to systematically find the solution Z^* of Problem 2. In particular, we reformulate conditions of Theorem 2 and optimality condition a few times to reach to a problem that we can systematically solve.

First, we show that for a network \mathcal{G} with graphical representation G , the graphical conditions (i) and (ii) of Theorem 2 can be reformulated to the notion of d -separation in a new network \mathcal{G}' built from \mathcal{G} .

Theorem 13. *Consider a directed graph $G = (V, E)$ where $i, j \in V$ and $i \in pa_G(j)$. Let $K := \{k | k \neq i \text{ and } k \rightarrow j \in E\}$ and define a graph $G' = (V', E')$ derived from G as follows.*

$$V' := V \cup \{p, q\} \tag{V.3}$$

$$E' := E \cup (q \rightarrow j, p \rightarrow q, \{k \rightarrow p | k \in K\}) \setminus \{k \rightarrow j | k \in K\} \tag{V.4}$$

Consider a set $Z \cap \{i, j\} = \emptyset$, then the set $Z \cup q$ d -separates $\{i, j\}$ and p in G' if and only if

- (i) Z is j -pointing separating the nodes i and j in G ; and
- (ii) $Z \cup \{i\}$ blocks all j -pointing paths from j to itself in G .

Proof. See the appendix. □

In a less formal way, Theorem 13 transforms the graph G into a new graph G' which contains two additional nodes p and q . In G' node q is going to be the only parent of node j that is not i , and node p is placed between q and all the former parents of j in G that were not i , see Figure 5.2.

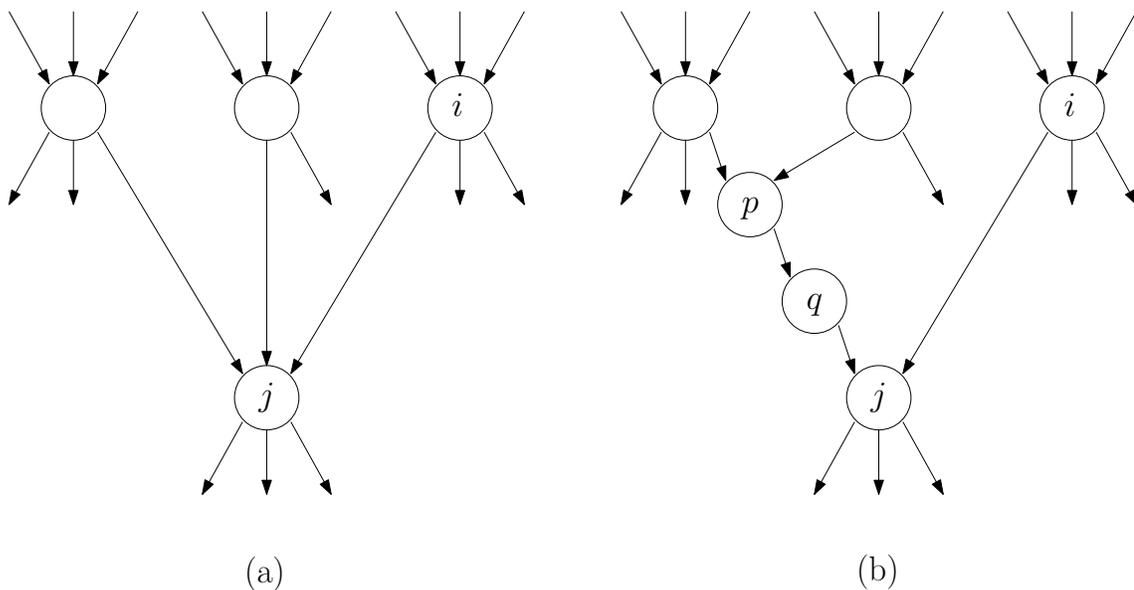


Figure 5.2: Placement of new variables p and q when the objective is identification of transfer function $H_{ji}(z)$. (a) Graph G (b) Graph G' after adding variables p and q

All the other edges are the same in G and G' . Theorem 13 states that if Z satisfies conditions (i) and (ii) in G if and only if Z d -separates $\{i, j\}$ and p in G' . The following example illustrates how Theorem 13 can be applied to reformulate pointing separation conditions of Theorem 2 to d -separation conditions by adding fictitious variables p and q .

Example 16. Consider a network with a graphical representation G shown in Figure 1.1. Suppose the objective is the identification of the transfer function $H_{21}(z)$ using Algorithm 1.

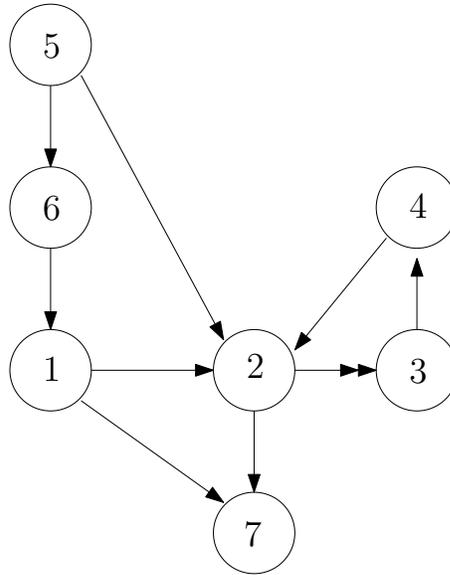


Figure 5.3: A graphical representation G of a network discussed in Example 25. The objective is identification of the transfer function $H_{21}(z)$.

Therefore, we need to find a set Z that 2-pointing separates nodes 1 and 2 and blocks all the 2-pointing paths from node 2 to itself. Theorem 13 states that $Z \cup q$ d -separates nodes $\{1, 2\}$ and p in the graph G' shown in Figure 5.4. For example, $Z = \{3, 5\}$ 2-pointing separates nodes 1 and 2 and blocks all the 2-pointing paths from node 2 to itself in G and at the same time d -separates $\{1, 2\}$ and p in the graph G' .

Theorem 13 stated that if we want to identify the transfer function $H_{ji}(z)$ we can look for a set Z such that $Z \cup q$ d -separates nodes $\{i, j\}$ and p in graph G' . The following two lemmas show that we can limit our search for the optimal set Z^* to the ancestors of i , j , and p .

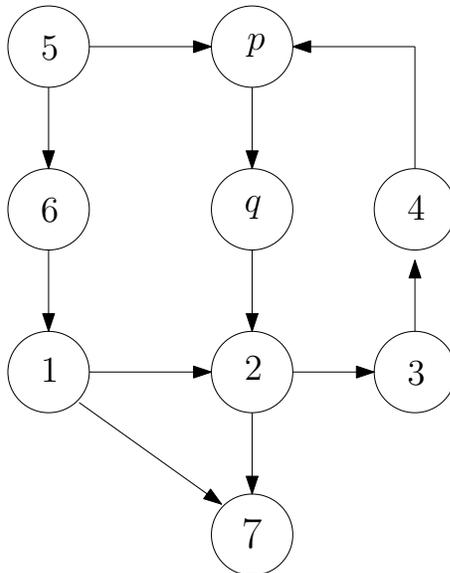


Figure 5.4: Graph G' resulted from graph G of Figure 1.1 after adding fictitious nodes p and q , discussed in Example 25

Lemma 14. Consider a directed graph $G = (V, E)$ and let G^a be the subgraph of G obtained from restricting G to $an_G(A \cup B)$. Let $C \subset an_G(A \cup B)$. If A and B are d -separated given C in G^a then A and B are d -separated given C in G .

Lemma 14 is a direct consequence of Lemma 4 in [54]. The following result from [98] states that if a set C d -separates A and B , we can get a smaller d -separator ($C \cap an(A \cup B)$) by removing from C all nodes that are not ancestors of A or B .

Lemma 15. Consider a directed graph $G = (V, E)$ where A , B , and C are disjoint subsets of V such that A and B are d -separated given C in G . Then A and B are d -separated given $C \cap an_G(A \cup B)$ in G .

Lemmas 14 and 15 enable us to limit our search for Z^* to the ancestors of nodes i , j , and p . However, since the cost function (V.2) is additive and $c_k \geq 0$, there is always going to be an optimal set Z^* that is a subset of $an_{G'}(\{i, j, p\})$.

The following example illustrates how using lemmas 14 and 15 we can manipulate graph G' to limit our search for the set Z^* .

Example 17. Suppose we are looking for an optimal set Z^* with respect to the cost function (V.2) to identify the transfer function $H_{21}(z)$ in a network with a graphical representation G shown in Figure 1.1. Lemmas 14 and 15 state that, to find Z^* , we can look for an optimal set Z^* that d -separates nodes 1 and p in graph G^a shown in Figure 5.5 which results from graph G' depicted in Figure 5.4 after removing node $7 \notin an_{G'}(\{1, 2, p\})$.

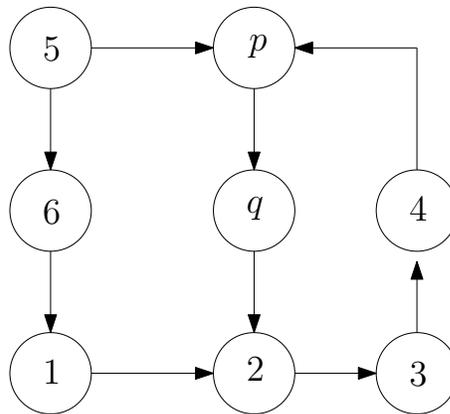


Figure 5.5: The ancestor graph G^a resulted from limiting graph G' of Figure 5.4 to ancestors of nodes 1, 2, and p , discussed in Example 17. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set d -separating nodes 1 and 2 in G^a .

The fact that we limited ourselves to the ancestor graph G^a in our search for Z^* allows us to reformulate the d -separation conditions in graphs G' and G^a to separation conditions in undirected graphs. The following result is an extension of a standard result for acyclic graphs [99] to general loopy graphs.

Theorem 16. Consider a directed graph $G' = (V', E')$ with $i, p \in V'$. Let G^a be the subgraph of G' restricted to $an_{G'}(\{i, j, p\})$. Let G^{mor} be the undirected moral graph of G^a .

The set Z separates nodes i and p in G^{mor} , if and only if Z d -separates nodes i and p in G^a .

Proof. See the appendix. □

As a corollary of Theorem 16 we can reformulate the sufficient and necessary conditions for consistent identification of $H_{ji}(z)$ using Algorithm 1 to the notion of separation in an undirected graph.

Corollary 16.1. Consider a directed graph $G = (V, E)$ where $i, j \in V$ and $i \in pa_G(j)$. Let $K := \{k | k \neq i \text{ and } k \rightarrow j \in E\}$ and define a graph $G' = (V', E')$ derived from G as follows.

$$V' := V \cup \{p, q\} \tag{V.5}$$

$$E' := E \cup (q \rightarrow j, p \rightarrow q, \{k \rightarrow p | k \in K\}) \setminus \{k \rightarrow j | k \in K\} \tag{V.6}$$

Let G^a be the ancestor graph resulted from limiting graph G' to $an_{G'}(i, j, p)$ and let G^{mor} be the moral graph of G^a . Then the set $Z \cup q$ where $Z \subset an_G(i, j)$ separates nodes $\{i, j\}$ and p in G^{mor} if and only if

- (i) Z is j -pointing separating the nodes i and j in G ; and
- (ii) $Z \cup \{i\}$ blocks all j -pointing paths from j to itself in G .

Proof. See the appendix. □

Corollary 25.1 says that if we want to consistently identify a transfer function $H_{ji}(z)$ using Algorithm 1 and minimize the cost of observations, we can look for a set Z^* that,

among all sets Z that separate nodes $\{i, j\}$ from node p in graph G^{mor} , has the minimum cost.

Example 18. Suppose we are looking for an optimal set Z^* with respect to the cost function (V.2) to identify the transfer function $H_{21}(z)$ in a network with a graphical representation G shown in Figure 1.1. Corollary 25.1 states that, to find Z^* , we can look for an optimal set Z^* that separates nodes $\{1, 2\}$ from node p in graph G^{mor} shown in Figure 1.4 which is resulted from moralizing graph G^a depicted in Figure 5.5.

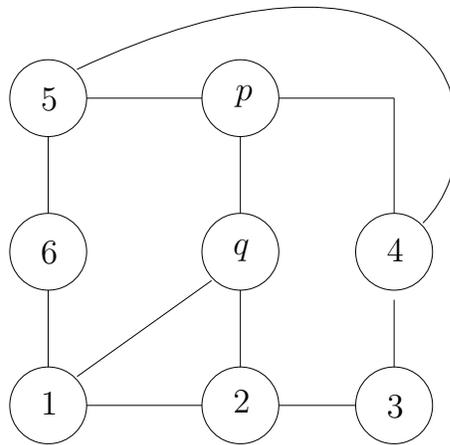


Figure 5.6: The undirected graph G^{mor} resulted from moralizing the graph G^a of Figure 5.5. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set separating nodes $\{1,2\}$ from node p in G^{mor} .

Note that node q is in every set that separates nodes $\{i, j\}$ from node p in the moral graph G^{mor} . Therefore, its observation cost does not play a role in minimizing the cost of observing additional variables and we can assign any cost to q .

To design an algorithm for finding the optimal set Z^* that separates nodes $\{i, j\}$ from node p in graph G^{mor} , we shall take advantage of the strong relationship that exists between problems of connectivity and flow problems in graphs. There is a standard procedure [100] that translates the problem of finding a set separating two sets of nodes in an undirected

graph to the problem of finding a minimum-cut set maximizing the flow between those two sets of nodes.

The procedure is as follows. We create a flow network by building an augmented graph $G^{aug} = (V^{aug}, E^{aug})$ from G^{mor} where every edge has a specific flow. Every node k in G^{mor} except node p is split into two nodes k_{in} and k_{out} . If there is an edge (k, ℓ) in G^{mor} then we will have the following edges in G^{aug} : $k_{out} \rightarrow \ell_{in}$ with flow capacity equal to infinity, $\ell_{out} \rightarrow k_{in}$ with flow capacity equal to infinity, $k_{in} \rightarrow k_{out}$ with flow capacity equal to c_k , and $\ell_{in} \rightarrow \ell_{out}$ with flow capacity equal to c_ℓ . Figure 5.7 shows an example of this process for a two-node network.

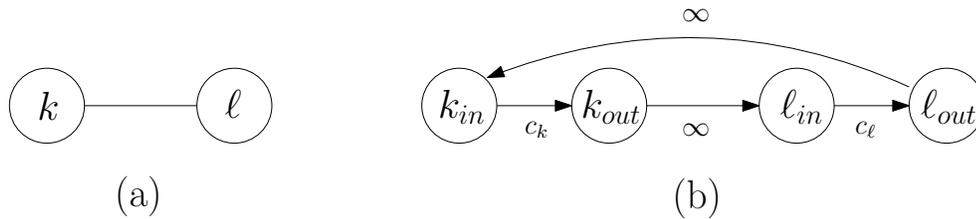


Figure 5.7: (a) A two-node undirected graph, (b) Corresponding augmented graph after splitting the nodes. Flow capacity of the edges of the form $k_{in} \rightarrow k_{out}$ is equal to c_k and flow capacities of all other edges are equal to infinity.

Example 19. Consider the undirected graph G^{mor} shown in Figure 1.4. Suppose the objective is to separate nodes $\{i, j\}$ from node p in G^{mor} . To reformulate the problem to a max-flow (min-cut) problem, the augmented graph G^{aug} corresponding to graph G^{mor} after splitting all the nodes except the sink p is shown in Figure 5.8.

In the flow network G^{aug} , nodes $\{i_{out}, j_{out}\}$ are the sources and node p is the sink. It is well known that a maximum flow problem is the dual of a minimum-cut problem. Because of the strong duality, the cost of the maximum flow from the sources $\{i_{out}, j_{out}\}$

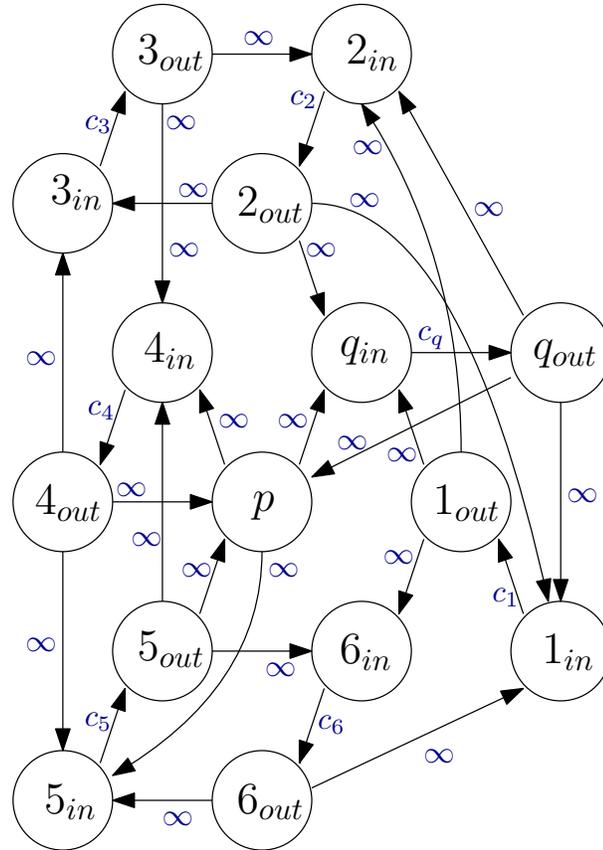


Figure 5.8: The augmented graph G^{aug} resulted from graph G^{mor} of Figure 1.4, reformulating a separation problem to a min-cut problem.

to the sink p is the same as the cost of the minimum cut separating the nodes $\{i_{out}, j_{out}\}$ from p . If the maximum flow is finite, the minimum cut set S separating $\{i_{out}, j_{out}\}$ from p in G^{aug} exists and contains only edges of the form $k_{in} \rightarrow k_{out}$.

Theorem 17. *Let S be the minimum cut separating $\{i_{out}, j_{out}\}$ from p in G^{aug} . Define $Z^* = \{k \mid k_{in} \rightarrow k_{out} \in S\} \setminus q$. $Z^* \cup q$ is the optimal set separating $\{i, j\}$ and p in G^{mor} minimizing the cost function (V.2).*

Proof. See the appendix. □

The minimum cut set S can be found using variety of well established algorithms. A standard approach is to solve its dual problem which is the maximization of the flow

from nodes $\{i_{out}, j_{out}\}$ to node p . For example, the Ford Fulkerson algorithm is an easily implementable way to compute the maximum flow [78]. Other algorithms such as implementations the push-relabel algorithm have the advantage of allowing a distributed implementation [79]. Then, the minimum-cut corresponding to the maximum flow can be found using, for example, Depth-first search algorithm [101]. Another approach is to follow an optimization based procedure such as linear programming. In such a case the minimum-cut can be found directly by solving a linear program formulated as the dual of the standard linear program that computes the max-flow from nodes $\{i_{out}, j_{out}\}$ to node p in G^{aug} .

In particular, for every edge $u \rightarrow v$ in G^{aug} consider a variable $y_{u \rightarrow v}$ and for every node u in G^{aug} , consider a variable y^u . The result is the following LP which can be used to find the minimum-cut separating nodes $\{i_{out}, j_{out}\}$ from node p in G^{aug} .

$$\min \sum_{(u \rightarrow v) \in E^{aug}} c_{u \rightarrow v} y_{u \rightarrow v} \quad (\text{V.7})$$

$$\text{subject to} \quad (\text{V.8})$$

$$y^v - y^u + y_{u \rightarrow v} \geq 0 \quad \forall u \rightarrow v \in E^{aug} \quad (\text{V.9})$$

$$y^u \geq 1 \text{ for } u \in \{i^{out}, j^{out}\} \quad (\text{V.10})$$

$$y^u = 0 \text{ for } u = p \quad (\text{V.11})$$

Since the matrices associated with the LP (V.7) are unimodular there is always an optimal solution where each of the variables $y_{u \rightarrow v}$ are either zero or one. Given any such optimal solution of (V.7), for any variable $y_{u \rightarrow v} = 1$ we have that $u = w^{in}$ and $v = w^{out}$ for some

$w \in V'$. Define, then, the set

$$\bar{S} := \{v | y_{v^{in} \rightarrow v^{out}} = 1\}. \quad (\text{V.12})$$

The set \bar{S} contains all the nodes separating $\{i, j\}$ from p in G^{mor} . Note that \bar{S} always contains the fictitious variable q . Thus, Z^* could be determined, following Corollary 25.1, by

$$Z^* = \bar{S} \setminus q. \quad (\text{V.13})$$

As a consequence of Theorem 13, Lemmas 14 and 15 and Theorem 16, the set Z^* satisfies the conditions (i) and (ii) of Theorem 2 and minimizes the cost of observations (V.2).

Algorithm 4 summarizes the steps that lead to the optimal set of predictors for identifying a certain transfer function $H_{ji}(z)$ using Algorithm 1.

Algorithm 2 Finding the optimal set of auxiliary variables (single module)

- 1: **Given:** topology $G = (V, E)$, target link $i \rightarrow j$, cost c_k for $k \in V$
 - 2: **Output:** optimal Z^*
 - 3: define G' by adding nodes p and q to G (as in Theorem 13)
 - 4: define G^a restricting G' to $\text{an}_G(\{i, j, p\})$ (as in Lemmas 14 and 15)
 - 5: define G^{mor} by moralizing G^a (as in Theorem 16)
 - 6: define G^{aug} by augmenting G^{mor} (as in Theorem 17)
 - 7: find the minimum cut S separating $\{i, j\}$ from p in G^{aug} (as in Theorem 17)
 - 8: $Z^* = \{k | k_{in} \rightarrow k_{out} \in S\} \setminus q$
-

Regarding the computational complexity of Algorithm 4, note that adding nodes p and q is computationally negligible and defining G^a (for example by using Dijkstra algorithm to find the ancestors) can be done in the worst case scenario in quadratic time. Defining G^{mor} can also be done in cubic time (multiplying the adjacency matrix by its transpose). Defining G^{aug} can be done in linear time by splitting the nodes. Therefore, for-

mulating LP (V.7) can be done in cubic time using standard procedures while specialized algorithms might perform even better. The LP (V.7) itself can be solved using, for example, Ford Fulkerson algorithm in quadratic time.

To show how the method presented in this section can be applied to general networks to find the optimal set of predictors for consistent identification of a certain transfer function, we reconsider Example 15. While in Example 15 the optimal set of predictors was found through an ad hoc procedure, in particular "branch and bound" strategy, the following example illustrates that the proposed method enables finding the optimal set of predictors systematically.

Example 20. *Consider a network with the recursive graphical representation depicted in Figure 1.5. The objective is the optimal identification of the transfer function $H_{21}(z)$ using Algorithm 1 by solving (V.1) where the cost $C(Z)$ is defined by (V.2). The cost c_r of observing each node $r \in V$ is reported in Table 1.1.*

In Example 15 a series of ad hoc arguments were made to show that $Z^ = \{5, 6, 7, 10, 13\}$ is the optimal set of predictors guaranteeing consistent identification of $H_{21}(z)$. However, Z^* could be systematically found using the ideas presented in this section. For example, Z^* could be found implementing a Ford-Fulkerson algorithm to find the maximum flow and then a depth-first algorithm to find the minimum cut (see step 7 of Algorithm 4). Alternatively, we could replace Ford-Fulkerson with Push-relabel and have a distributed version of Algorithm 4. In our numerical verification we have opted for a linear program as in (V.7). The solution of such LP leads to $Z^* = \{5, 6, 7, 10, 13\}$ which is equivalent to the the optimal set of predictors found in Example 15.*

5.2 Optimal Selection of Observations for Identification of Multiple Modules

In the previous section we developed a systematic procedure to select an optimal set of variables to identify a single transfer function in a network. In this section, we extend those results to the scenario where we want to select an optimal set of variables to identify multiple transfer functions simultaneously.

When trying to find the optimal estimating processes set for identifying multiple transfer functions, one might intuitively think that it is possible to solve smaller problems like (V.7) for each transfer function. This is, however, not true in general. Inspired by the results of the previous section, one might also speculate that it is possible to find the optimal estimating processes set by separating the terminal nodes of all transfer functions of interests all at once. This conjecture is also false and it turns out that finding an optimal set of processes to identify multiple transfer functions simultaneously is more challenging than the scenario of identifying a single transfer function.

For example, consider a network with a graphical representation depicted in Figure 5.9. Cost of observing each node is shown outside of the node. If we want to identify the transfer function $H_{j_1 i_1}(z)$ using Algorithm 1, the set of auxiliary variables $\{4, 2\}$ with the cost $10+20 = 30$ satisfies the conditions of Theorem 2 for identification of $H_{j_1 i_1}(z)$ and can be proven to be optimal. Similarly, if we want to identify the transfer function $H_{j_2 i_2}(z)$, the set $\{3\}$ with the cost 20 satisfies the conditions of Theorem 2 for identification of $H_{j_2 i_2}(z)$ and can be proven optimal. Therefore, we can identify both $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$ by

observing $\{4, 2\} \cup \{3\} = \{2, 3, 4\}$ with the cost $10 + 20 + 20 = 50$. However, if we want to simultaneously identify both transfer functions $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$, it suffices to observe the nodes in the set $Z^* = \{1, 4\}$ with the cost $10 + 30 = 40$. Then, $Z_{j_1 i_1} = \{1, 4\} \subseteq Z^*$ satisfies the conditions of Theorem 2 for identification of $H_{j_1 i_1}(z)$ and $Z_{j_2 i_2} = \{1\} \subset Z^*$ satisfies the conditions of Theorem 2 for identification of $H_{j_2 i_2}(z)$.

If we define G' , G^{mor} , and G^{aug} as was explained in the previous section, and find the minimum cut separating $\{i_1^{out}, i_2^{out}, j_1^{out}, j_2^{out}\}$ from $\{p_1, p_2\}$, we get $\{1, 4, 5\}$ as the predictors set which is different from Z^* . The reason why node 5 is mistakenly selected is as follows. Finding the minimum cut separating $\{i_1^{out}, i_2^{out}, j_1^{out}, j_2^{out}\}$ from $\{p_1, p_2\}$ in G^{aug} is equivalent to finding a set satisfying conditions (i) and (ii) of Theorem 2 in G for both transfer functions $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$. For identification of $H_{j_1 i_1}(z)$ the j_1 -pointing path $\pi_1 = \{i_1 \rightarrow 4 \rightarrow j_1\}$ in G needs to be blocked. Therefore, node 4 is definitely in the predictors set. When selected, node 4 acts as an activated collider on the j_2 -pointing path $\pi_2 = \{i_2 \rightarrow 4 \leftarrow 5 \rightarrow j_2\}$ in G . Node 5, then, needs to be selected to block π_2 for identification of $H_{j_2 i_2}(z)$ which leads to a non-optimal set of predictors.

The example above shows that, when identifying multiple transfer functions the extension of Algorithm 4 is not straightforward. In what follows, we formally cast and solve the more challenging problem of finding the optimal set of predictors for identifying an arbitrary number of transfer functions in a general network.

Consider a network with graphical representation $G = (V, E)$. Suppose we want to simultaneously identify M transfer functions $H_{j_1 i_1}(z), H_{j_2 i_2}(z), \dots, H_{j_m i_m}(z), \dots, H_{j_M i_M}(z)$ for $i_m, j_m \in V$ and $m \in \{1, 2, \dots, M\}$. We can formally cast this problem as follows.

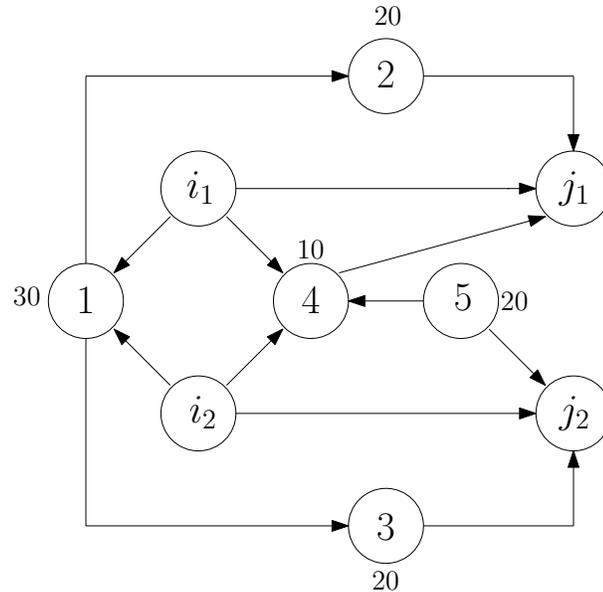


Figure 5.9: The graphical representation G of a network highlighting the differences between identification of a single transfer function and multiple transfer functions using Algorithm 1. To identify only $H_{j_1 i_1}(z)$ the set $\{2, 4\}$ is the optimal predictors set with cost 30 and to identify only $H_{j_2 i_2}(z)$ the set $\{3\}$ is the optimal predictors set with the cost 20. It is possible, however, to identify both $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$ using the optimal predictors set $Z^* = \{1, 4\}$ with the cost 40. Observe that the optimal predictor set to identify both transfer functions is not the union of the optimal predictors set to identify each transfer function independently.

Problem 3. Consider a network $\mathcal{G} = (H, n)$ where n are unknown mutually independent inputs and measuring node $u \in V$ has a cost $c_u \geq 0$. Given a graphical representation G of \mathcal{G} , find a set of predictors Z^* of minimum cost with respect to (V.2) such that for every $m \in \{1, 2, \dots, M\}$ there is a $Z_m \subseteq Z^*$ using which Algorithm 1 guarantees a consistent identification of the transfer function $H_{j_m i_m}(z)$.

Similar to the results of the previous section, we are going to use the sufficient and necessary graphical conditions for a set Z_m to guarantee consistent identification of the transfer function $H_{j_m i_m}(z)$ using Algorithm 1. Given $G = (V, E)$ and pairs $i_m, j_m \in V$ with $m \in \{1, 2, \dots, M\}$, construct the graph $G' = (V', E')$ by adding nodes p_m and q_m

between node j_m and $\text{pa}_G(j_m) \setminus i_m$, for all m . Namely,

$$\begin{aligned}
 V' &:= \bigcup_{m=1}^M \{p_m, q_m\} \cup V \\
 E' &:= \left(\bigcup_{m=1}^M \{q_m \rightarrow j_m, p_m \rightarrow q_m, w_m \rightarrow p_m \mid w_m \in W_m\} \cup E \right) \\
 &\quad \setminus \left(\bigcup_{m=1}^M \{w_m \rightarrow j_m \mid w_m \in W_m\} \right) \quad (\text{V.14})
 \end{aligned}$$

where $W_m := \{w_m \mid w_m \neq i_m \text{ and } w_m \rightarrow j_m \in E\}$. Graph G' has $|V| + 2M$ nodes because for every transfer function $H_{j_m i_m}(z)$ that we want to identify we are adding two nodes (p_m and q_m) to $|V|$ nodes of G .

For example, considering the network with a graphical representation G shown in Figure 5.9 where the objective is identification of transfer functions $H_{j_1 i_1}(z)$ and $H_{j_1 i_2}(z)$, the corresponding graph G' after adding fictitious nodes p_1, q_1, p_2 , and q_2 is shown in Figure 5.10.

Now, we can express the conditions that a set Z needs to satisfy to result in a consistent identification of transfer functions $H_{j_m i_m}(z)$ in terms of some graphical conditions in graph G' .

Theorem 18. Z_m satisfies conditions (i) and (ii) of Theorem 2 for identification of transfer function $H_{j_m i_m}(z)$ if and only if the set $\{i_m, j_m\}$ and the node p_m are d -separated by $Z_m \cup q_m$ in G' defined by (V.14).

Proof. See the appendix. □

It follows from Theorem 18 that using a $Z \supseteq Z_m$, for all m , we can consistently identify all transfer functions $H_{j_m i_m}(z)$. Among all such Z , we are looking for the one

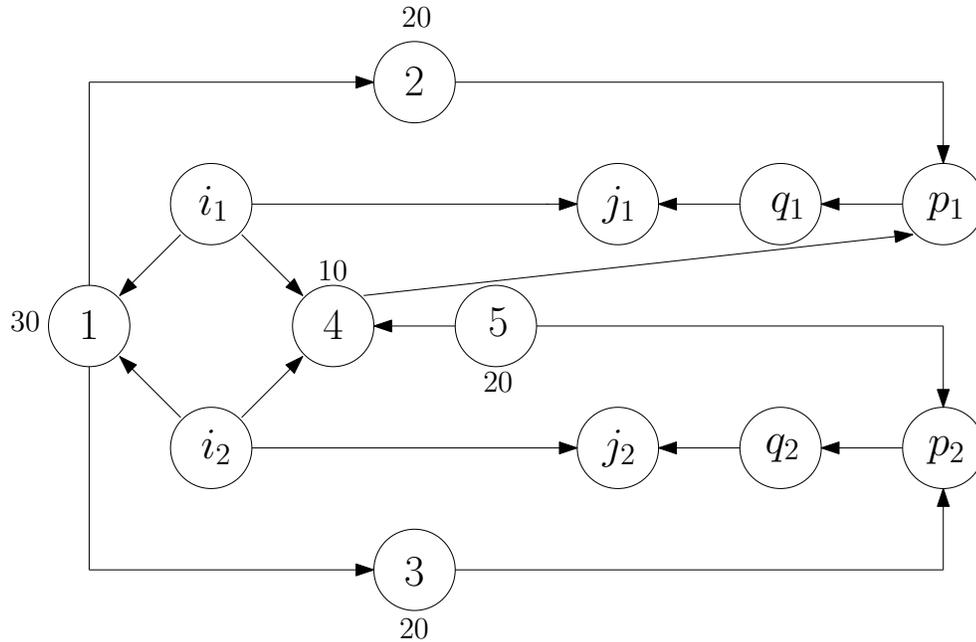


Figure 5.10: The graph G' obtained after adding fictitious nodes p_1 , p_2 , q_1 , and q_2 to the graph G of Figure 5.9 for identification of $H_{j_1 i_1}(z)$ and $H_{j_1 i_2}(z)$.

minimizing the cost function (V.2), namely Z^* . Reformulating the graphical conditions of Theorem 18, we will formulate a LP whose solution will lead to Z^* .

Construct $G^{mor} = (V^{mor}, E^{mor})$ by moralizing G' and construct the graph $G^{aug} = (V^{aug}, E^{aug})$ by splitting the nodes of G^{mor} . For every edge $u \rightarrow v \in E^{aug}$ assign a flow capacity $c_{u \rightarrow v}$ as explained in the previous section. For every m , define graph G'_m as the subgraph of G' limited to ancestors of i_m, j_m , and p_m . Similarly, G_m^{mor} is the moral graph of G'_m , and G_m^{aug} is resulted from splitting the nodes of G_m^{mor} . For every edge $u \rightarrow v$ in G_m^{aug} consider a variable $y_{u \rightarrow v}$ and for every node u in G_m^{aug} , consider M variables y_m^u , $m \in \{1, 2, \dots, M\}$. We will show that a solution of the following LP leads to an optimal set of predictors Z^* enabling consistent identification of M transfer functions $H_{j_m i_m}(z)$, for

$$m \in \{1, 2, \dots, M\}.$$

$$\min \sum_{(u \rightarrow v) \in E^{aug}} c_{u \rightarrow v} y_{u \rightarrow v} \quad (\text{V.15})$$

$$\text{subject to} \quad (\text{V.16})$$

$$y_m^v - y_m^u + y_{u \rightarrow v} \geq 0 \quad \forall u \rightarrow v \in E_m^{aug} \quad \forall m \quad (\text{V.17})$$

$$y_m^u \geq 1 \text{ for } u \in \{i_m^{out}, j_m^{out}\} \quad \forall m \quad (\text{V.18})$$

$$y_m^u = 0 \text{ for } u = p_m \quad \forall m \quad (\text{V.19})$$

First we show that for any base solution of (V.15), the variables $y_{u \rightarrow v}$ and y_m^u will turn out to be integers.

Proposition 19. *Any base solution of the LP described by (V.15) is integral.*

Proof. See the appendix. □

The LP described by (V.15) assigns a weight $y_{u \rightarrow v}$ to each edge $u \rightarrow v$, which we may think of as a "length". The constraints of LP described by (V.15) are specifying that, for a fixed m , along each possible path in G_m^{aug} , the sources $\{i_m^{out}, j_m^{out}\}$ and the sink p_m are at distance at least one. This means that the LP variables are expressing a way of "separating" $\{i_m^{out}, j_m^{out}\}$ from p_m in G_m^{aug} .

Proposition 20. *Let $S = \bigcup S_m$. If S_m is a feasible cut-set separating $\{i_m^{out}, j_m^{out}\}$ from p_m in G_m^{aug} , then there is a feasible solution to (V.15) such that*

$$\text{capacity}(S) = \sum_{(u \rightarrow v) \in E^{aug}} c_{u \rightarrow v} y_{u \rightarrow v} \quad (\text{V.20})$$

Proof. See the appendix. □

Among all S , let S^* be the one with minimum capacity. Proposition 20 implies that the optimum of (V.15) is smaller than or equal to the capacity of S^* . The next result shows that the optimum of (V.15) is actually equal to the capacity of S^* .

Proposition 21. *Given a feasible integer solution of (V.15), there is a set S^* with capacity equal to the cost of the solution such that for every m there is a cut-set $S_m \subseteq S^*$ for G_m^{aug} .*

Proof. See the appendix. □

From Proposition 19 there is no loss of generality if we only consider an optimal integer solution S^* for LP (V.15).

Given any feasible integer solution of (V.15) with finite cost, for any variable $y_{u \rightarrow v} = 1$ we have that $u = w^{in}$ and $v = w^{out}$ for some $w \in V'$. The following result provides a method for finding an optimal set Z^* using a solution of LP described by (V.15).

Theorem 22. *Consider an integer solution of the LP (V.15) and define the set*

$$\bar{S} := \{v | y_{v^{in} \rightarrow v^{out}} = 1\}. \quad (\text{V.21})$$

An optimal set of observations that leads to a consistent identification of $H_{j_m i_m}(z)$ for all m is given by

$$Z^* = \bar{S} \setminus \{q_1, q_2, \dots, q_M\}. \quad (\text{V.22})$$

Proof. See the appendix. □

Note that \bar{S} in Theorem 22 always contains the fictitious variables q_m . Algorithm 3 provides the steps leading to determination of the optimal set Z^* which contains a set Z_m

for consistent identification of $H_{j_m i_m}(z)$ for all m . Regarding the computational complexity

Algorithm 3 Finding an optimal observation set (multiple modules)

- 1: **Given:** topology $G = (V, E)$, target links $i_m \rightarrow j_m$, $m = 1, \dots, M$, cost c_k for $k \in V$
 - 2: **Output:** optimal Z^* ,
 - 3: **for** <all m > **do**
 - 4: define G' by adding nodes p_m and q_m to G (as in (V.14))
 - 5: **end for**
 - 6: define G^{mor} by moralizing G'
 - 7: define G^{aug} by augmenting G^{mor}
 - 8: **for** <all m > **do**
 - 9: < define G_m^a by restricting G' to $\text{an}_G(\{i_m, j_m, p_m\})$ >
 - 10: < define G_m^{mor} by moralizing G_m^a >
 - 11: < define G_m^{aug} by augmenting G_m^{mor} >
 - 12: **end for**
 - 13: < solve the LP (V.15) >
 - 14: < $\bar{S} := \{v | y_{v^{in} \rightarrow v^{out}} = 1\}$ >
 - 15: < $Z^* = \bar{S} \setminus \{q_1, q_2, \dots, q_M\}$ >
-

of Algorithm 3, note that similar to the single transfer function identification case, formulating LP (V.15) can be done in cubic time. Unlike the single transfer function identification case, where we could exploit the special form of the LP (V.7), to solve it in quadratic time using, for example, Ford Fulkerson algorithm, the LP (V.15) cannot be tackled using some already developed specialized algorithm. We conjecture that the specific structure of the LP (V.15) still lends itself to be solved efficiently but this will be investigated in future work.

Once Z^* is found, the following result presents a way to determine an appropriate set $Z_m \subseteq Z^*$ such that Algorithm 1 will output a consistent estimate for $H_{j_m i_m}(z)$.

Theorem 23. *For a fixed m , Algorithm 1 consistently estimates $H_{j_m i_m}(z)$ using Z_m given by*

$$Z_m = Z^* \cap V_m^{mor}. \quad (\text{V.23})$$

Proof. See the appendix. □

The following examples show how the proposed method can be applied to detect the optimal set of predictors for identification of multiple transfer functions.

Example 21. Consider a network with a graphical representation shown in Figure 5.11. Except for the nodes 1 and 2 which their costs of observations are parameters α and β ,

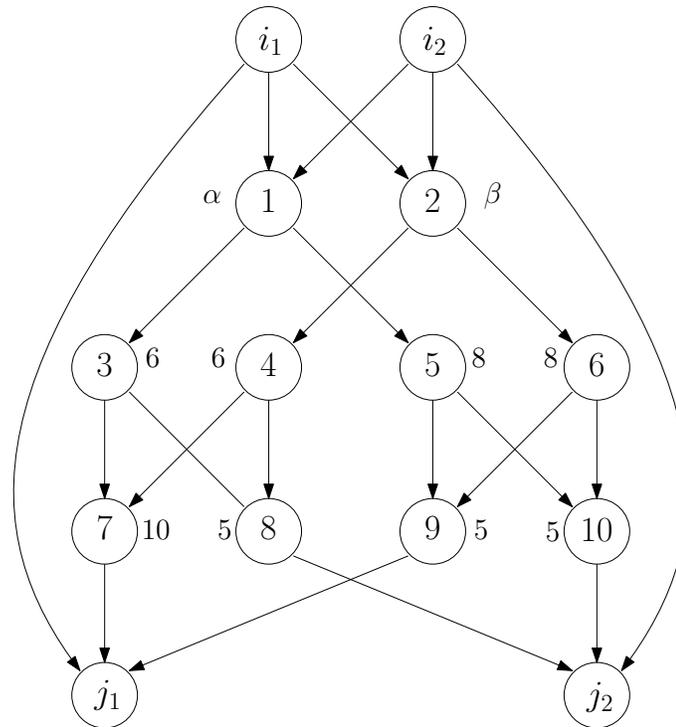


Figure 5.11: The network discussed in Example 21. The objective is Identification of $H_{j_1 i_1}(z)$ and $H_{j_2 i_2}(z)$ while minimizing the cost of observation.

respectively, the cost of observing any other node is fixed and reported outside of the node in bold. The objective is identification of transfer functions $H_{j_1 i_1}(z)$, and $H_{j_2 i_2}(z)$ while minimizing the cost of observations. Taking the steps explained in this section, we can write a LP similar to (V.15) to find the optimal set of predictors. For different values of α and β we get different solutions for the optimal sets of predictors. Figure 5.12 specifies different regions of parameters α and β corresponding to different solutions for Z^* . In region 1

the optimal set of predictors is $Z^* = \{3, 4, 9, 10\}$, in region 2 $Z^* = \{2, 3, 5\}$, in region 3 $Z^* = \{1, 2\}$ and in region 4 $Z^* = \{1, 4, 6\}$.

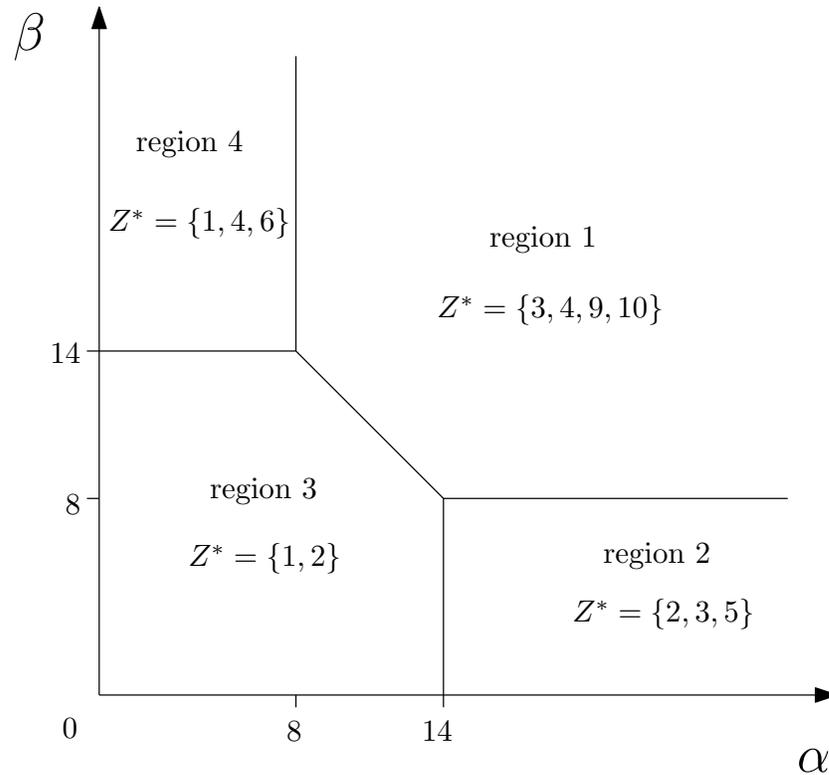


Figure 5.12: Different regions of parameters α and β corresponding to different solutions for Z^*

Example 22. Consider a network with a graphical representation shown in Figure 5.13. The costs of observations of the nodes are reported in Table 5.3.

The objective is identification of transfer functions $H_{j_1 i_1}(z)$, $H_{j_2 i_2}(z)$, and $H_{j_3 i_3}(z)$ using algorithm 1 while minimizing the cost of observations. Taking the steps explained in this section, we can write a LP to find the optimal set of predictors. If we wanted to identify $H_{j_1 i_1}(z)$, applying the method presented in Section 5.1, we would get $\{5\}$ as the optimal set of predictors. Similarly, if we wanted to identify $H_{j_2 i_2}(z)$ we would get $\{11, 14, 32\}$ as the optimal set of predictors and if we wanted to identify $H_{j_3 i_3}(z)$ we would get $\{42, 69\}$

Table 5.3: Costs of observing each node in Example 22

node r	c_r						
1	10	20	4	39	3	58	19
2	20	21	19	40	4	59	19
3	17	22	23	41	16	60	20
4	18	23	26	42	5	61	23
5	4	24	22	43	21	62	21
6	19	25	29	44	22	63	18
7	22	26	17	45	3	64	20
8	23	27	4	46	4	65	19
9	20	28	5	47	5	66	17
10	20	29	5	48	7	67	25
11	14	30	5	49	4	68	20
12	19	31	17	50	6	69	4
13	18	32	5	51	21	70	29
14	5	33	20	52	22	71	26
15	5	34	5	53	22	72	23
16	3	35	20	54	20	73	18
17	6	36	5	55	5	74	24
18	3	37	12	56	4	75	21
19	5	38	20	57	3	76	19

5.3.1 Proof of Theorem 13

Proof. We prove the sufficiency. Suppose that $Z \cup q$ d -separates $\{i, j\}$ and p in G' . By contradiction suppose there is a j -pointing path $\pi = \{i \cdots k \rightarrow j\}$ between i and j that is not blocked by Z in G see Figure 5.14 (a)). Thus, the path $\pi' = \{i \cdots k \rightarrow p\}$ is an active path between $\{i, j\}$ and p in G' which is a contradiction. Similarly, by contradiction suppose there is a j -pointing path $\ell = \{j \rightarrow \cdots k \rightarrow j\}$ from j to itself that is not blocked by $Z \cup i$ in G . Note that ℓ could be directed or not. Then, the path $\ell' = \{j \rightarrow \cdots k \rightarrow p\}$ is an active path between $\{i, j\}$ and p in G' which is a contradiction(see Figure 5.14 (b)).

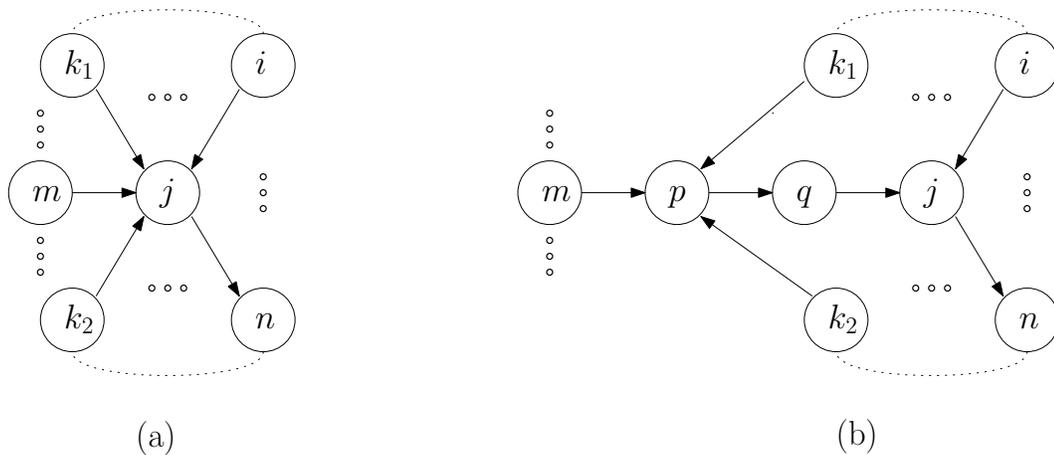


Figure 5.14: (a) An illustrative graph G of the proof of Theorem 13 (b) Graph G' of the proof of Theorem 13

Now we prove the necessity of the conditions. Suppose Z is j -pointing separating the nodes i and j in G ; and $Z \cup \{i\}$ blocks all j -pointing paths from j to itself in G . By contradiction suppose there is an active path $\pi' = \{i \cdots k \rightarrow p\}$ between i and p , not containing j , that is not blocked by $Z \cup q$ in G' . Then, the path $\pi = \{i \cdots k \rightarrow j\}$ is an active j -pointing path in G not blocked by Z which is a contradiction. Similarly, by contradiction suppose there is an active path $\ell' = \{j \rightarrow \cdots k \rightarrow p\}$ between j and p that

is not blocked by $Z \cup q$ in G' . Then, the path $\ell = \{j \rightarrow \cdots k \rightarrow j\}$ is an active j -pointing path from j to itself in G not blocked by Z which is a contradiction. It is also trivial that the path $j \leftarrow q \leftarrow p$ in G' is always blocked by q . \square

5.3.2 Proof of Theorem 16

Proof. The arguments in the proof of Proposition 3 in [99] follow also in the case of cyclic graphs. \square

5.3.3 Proof of Corollary 25.1

Proof. The results follow immediately from Theorem 16. \square

5.3.4 Proof of Theorem 17

Proof. That $Z^* \cup q$ is the optimal set separating $\{i, j\}$ and p in G^{mor} minimizing the cost function (V.2) follows from the well known max-flow min-cut theorem [102] (for more details see section 5 of [98]). \square

5.3.5 Proof of Theorem 18

Proof. First, we prove the sufficiency. Suppose that $Z_m \cup q_m$ d -separates $\{i_m, j_m\}$ and p_m in G' . By contradiction suppose there is a j_m -pointing path $\pi_1 = \{i_m \cdots k \rightarrow j_m\}$ between i_m and j_m that is not blocked by Z_m in G and π_1 does not contain any j_n , $n \neq m$. Thus the path $\pi'_1 = \{i_m \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. By contradiction suppose there is a j_m -pointing path $\pi_2 = \{i_m \cdots v_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ between i_m and j_m that is not blocked by Z_m in G where $v_n \neq i_n$. Thus the path

$\pi'_2 = \{i_m \cdots v_n \rightarrow p_n \rightarrow q_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. By contradiction suppose there is a j_m -pointing path $\pi_3 = \{i_m \cdots v_n \rightarrow j_n \leftarrow w_n \cdots k \rightarrow j_m\}$ between i_m and j_m that is not blocked by Z_m in G where $v_n \neq i_n$ and $w_n \neq i_n$. Thus the path $\pi'_3 = \{i_m \cdots v_n \rightarrow p_n \leftarrow w_n \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. By contradiction suppose there is a j_m -pointing path $\pi_4 = \{i_m \cdots i_n \rightarrow j_n \leftarrow w_n \cdots k \rightarrow j_m\}$ between i_m and j_m that is not blocked by Z_m in G . Thus the path $\pi'_4 = \{i_m \cdots i_n \rightarrow j_n \leftarrow q_n \leftarrow p_n \leftarrow w_n \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. By contradiction suppose there is a j_m -pointing path $\pi_5 = \{i_m \cdots i_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ between i_m and j_m that is not blocked by Z_m in G . Thus the path $\pi'_5 = \{i_m \cdots i_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. By contradiction suppose there is a j_m -pointing path $\pi_6 = \{i_m \cdots \leftarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ between i_m and j_m that is not blocked by Z_m in G . Thus the path $\pi'_6 = \{i_m \cdots \leftarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction.

Similarly, suppose by contradiction that there is a j_m -pointing path $\ell_1 = \{j_m \rightarrow \cdots k \rightarrow j_m\}$ from j_m to itself and not involving j_n , $n \neq m$, that is not blocked by $Z_m \cup i_m$ in G . Note that ℓ_1 could be directed or not. Then, the path $\ell'_1 = \{j_m \rightarrow \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. Similarly, suppose by contradiction that there is a j_m -pointing path $\ell_2 = \{j_m \rightarrow \cdots i_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ from j_m to itself that is not blocked by Z_m in G . Then, the path $\ell'_2 = \{j_m \rightarrow \cdots i_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. Similarly, suppose by contradiction that there is a j_m -pointing path $\ell_3 = \{j_m \rightarrow \cdots v_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$, $v_n \neq i_n$, that is not blocked by Z_m in G . Then, the path $\ell'_3 = \{j_m \rightarrow \cdots v_n \rightarrow p_n \rightarrow q_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ is an

active path in G' which is a contradiction. Similarly, suppose by contradiction that there is a j_m -pointing path $\ell_4 = \{j_m \rightarrow \cdots v_n \rightarrow j_n \leftarrow w_n \cdots k \rightarrow j_m\}$, $v_n \neq i_n$ and $w_n \neq i_n$, that is not blocked by Z_m in G . Then, the path $\ell'_4 = \{j_m \rightarrow \cdots v_n \rightarrow p_n \leftarrow w_n \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction. Similarly, suppose by contradiction that there is a j_m -pointing path $\ell_5 = \{j_m \rightarrow \cdots \leftarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ that is not blocked by Z_m in G . Then, the path $\ell'_5 = \{j_m \rightarrow \cdots \leftarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ is an active path in G' which is a contradiction.

Now we prove the necessity of the conditions. Suppose Z_m is j_m -pointing separating the nodes i_m and j_m in G ; and $Z_m \cup \{i_m\}$ blocks all directed loops involving j_m in G . By contradiction suppose there is an active path $\pi'_1 = \{i_m \cdots k \rightarrow p_m\}$ between i_m and p_m that is not blocked by $Z_m \cup q_m$ in G' and does not contain p_n , for all $n \neq m$. Then, the path $\pi_1 = \{i_m \cdots k \rightarrow j_m\}$ is an active j_m -pointing path in G not blocked by Z_m which is a contradiction. By contradiction suppose there is an active path $\pi'_2 = \{i_m \cdots v_n \rightarrow p_n \rightarrow q_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ between i_m and p_m that is not blocked by $Z_m \cup q_m$ in G' , for some $n \neq m$. Then, the path $\pi_2 = \{i_m \cdots v_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ is an active j_m -pointing path in G not blocked by Z_m which is a contradiction. By contradiction suppose there is an active path $\pi'_3 = \{i_m \cdots i_n \rightarrow j_n \leftarrow q_n \leftarrow p_n \leftarrow v_n \cdots k \rightarrow p_m\}$ between i_m and p_m that is not blocked by $Z_m \cup q_m$ in G' , for some $n \neq m$. Then, the path $\pi_3 = \{i_m \cdots i_n \rightarrow j_n \leftarrow v_n \cdots k \rightarrow j_m\}$ is an active j_m -pointing path in G not blocked by Z_m which is a contradiction. By contradiction suppose there is an active path $\pi'_4 = \{i_m \cdots v_n \rightarrow p_n \rightarrow q_n \rightarrow j_n \leftarrow i_n \cdots k \rightarrow p_m\}$ between i_m and p_m that is not blocked by $Z_m \cup q_m$ in G' , for some $n \neq m$. Then, the path $\pi_4 = \{i_m \cdots v_n \rightarrow j_n \leftarrow i_n \cdots k \rightarrow j_m\}$ is an active j_m -pointing path in G not blocked by Z_m which is a contradiction. By

contradiction suppose there is an active path $\pi'_5 = \{i_m \cdots v_n \rightarrow p_n \leftarrow w_n \cdots k \rightarrow p_m\}$ between i_m and p_m that is not blocked by $Z_m \cup q_m$ in G' , for some $n \neq m$. Then, the path $\pi_5 = \{i_m \cdots v_n \rightarrow j_n \leftarrow w_n \cdots k \rightarrow j_m\}$ is an active j_m -pointing path in G not blocked by Z_m which is a contradiction.

Similarly, by contradiction suppose there is an active path $\ell'_1 = \{j_m \rightarrow \cdots k \rightarrow p_m\}$ between j_m and p_m that does not contain j_n , for any $n \neq m$, and that is not blocked by $Z_m \cup q_m$ in G' . Then, the the path $\ell_1 = \{j_m \rightarrow \cdots k \rightarrow j_m\}$ is an active j -pointing path from j_m to itself in G not blocked by Z_m which is a contradiction. Similarly, by contradiction suppose there is an active path $\ell'_2 = \{j_m \rightarrow \cdots i_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$ between j_m and p_m that is not blocked by $Z_m \cup q_m$ in G' . Then, the path $\ell_2 = \{j_m \rightarrow \cdots i_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ is a j -pointing path from j_m to itself in G not blocked by Z_m which is a contradiction. Similarly, by contradiction suppose there is an active path $\ell'_3 = \{j_m \rightarrow \cdots v_n \rightarrow p_n \rightarrow q_n \rightarrow j_n \rightarrow \cdots k \rightarrow p_m\}$, $v_n \neq i_n$, between j_m and p_m that is not blocked by $Z_m \cup q_m$ in G' . Then, the directed loop $\ell_3 = \{j_m \rightarrow \cdots v_n \rightarrow j_n \rightarrow \cdots k \rightarrow j_m\}$ is an active directed loop involving j_m in G not blocked by Z_m which is a contradiction. It is also trivial that the path $j_m \leftarrow q_m \leftarrow p_m$ in G' is always blocked by q_m . □

5.3.6 Proof of Proposition 19

Proof. In a canonical (or standard) form of (V.15), the vector of coefficients are integral and it follows from [103] that the matrix of coefficients is unimodular. Thus, the LP described by (V.15) has integral optima. □

5.3.7 Proof of Proposition 20

Proof. Consider the division of the nodes of G_m^{aug} into two parts by S_m , with the sources $\{i_m^{out}, j_m^{out}\}$ in one part S'_m and the sink p_m in the other. Define $y_m^v = 1$ if $v \in S'_m$, and $y_m^v = 0$ otherwise. Define $y_{u \rightarrow v} = 1$ if $u \rightarrow v \in S$, and $y_{u \rightarrow v} = 0$ otherwise. By construction, S_m is a feasible cut-set in G_m^{aug} . By inspection it can be seen that the constraints of the LP (V.15) are met and this is a feasible solution. \square

5.3.8 Proof of Proposition 21

Proof. Let

$$S^* := \{u \rightarrow v \mid y_{u \rightarrow v} = 1\}. \quad (\text{V.24})$$

The capacity of S^*

$$\text{capacity}(S^*) = \sum_{(u \rightarrow v) \in S} c_{u \rightarrow v} = \sum_{(u \rightarrow v) \in E^{aug}} c_{u \rightarrow v} y_{u \rightarrow v} \quad (\text{V.25})$$

is equal to the cost of the solution. For every node $v \in V_m^{aug}$, let $v \in S'_m$ if $y_m^v = 1$, and $v \notin S'_m$ if $y_m^v = 0$. By construction, S'_m is a feasible cut in G_m^{aug} . Define $S_m := \{u \rightarrow v \mid (u \rightarrow v) \in E_m^{aug}, u \in S_m, v \notin S_m\}$. It remains to show that $S_m \subseteq S^*$. Consider any edge $(u \rightarrow v) \in S_m$. Since $y_m^u = 1$ and $y_m^v = 0$, it follows from the constraint $y_m^v - y_m^u + y_{u \rightarrow v} \geq 0$, $\forall u \rightarrow v \in E_m^{aug}$, that $y_{u \rightarrow v} \geq 1$. Variable $y_{u \rightarrow v}$ is either zero or one, thus, $y_{u \rightarrow v} = 1$. Since $y_{u \rightarrow v} = 1$, we have that $u \rightarrow v \in S^*$ and therefore $S_m \subseteq S^*$. \square

5.3.9 Proof of Theorem 22

Proof. Let $Z_m = Z^* \cap V_m^{mor}$. Algorithm 1 results in a consistent estimate of $H_{j_m i_m}(z)$ using Z_m (see Theorem 23). Thus, for every m , Z^* contains a set Z_m that leads to a consistent learning of $H_{j_m i_m}(z)$. It follows from Proposition 20 and 21 that Z^* has the minimum observation cost. \square

5.3.10 Proof of Theorem 23

Proof. Z_m satisfies the constraints of the LP (V.15). Thus, $Z_m \cup \{q_m\}$ separates $\{i_m, j_m\}$ from p_m in G_m^{mor} . Thus, $Z_m \cup \{q_m\}$ d -separate $\{i_m, j_m\}$ from p_m in G'_m . By Theorem 22, Z_m blocks all the j_m -pointing paths between i_m and j_m with the exception of $i_m \rightarrow j_m$ in G and blocks all the j_m -pointing paths from j_m to itself in G . By Theorem 2, Algorithm 1 results in a consistent estimate of $H_{j_m i_m}(z)$ using Z_m . \square

CHAPTER VI

HOW CAN WE BE ROBUST AGAINST GRAPH UNCERTAINTIES?

The goal of this chapter is to lay the foundation of a novel methodological paradigm to design controllers for networked systems when the interconnection structure of the system is uncertain and only observational data is available. We show that it is possible to leverage relevant results of previous chapters developed in the area of identification of dynamic networks to introduce a notion of robustness with respect to uncertainties in the graph structure of a distributed system. It is assumed that in an observational framework, only a subset of the variables of a networked system are measured and the topology of the interconnections between the variables is not fully known. When the objective is designing a controller for the overall system, the topological uncertainties impede the exact identification of the overall open-loop transfer function and consequently, the exact design of the controller. It is shown, however, that some of the transfer functions of the network could be consistently identified using some sufficient and necessary graphical conditions and, in some cases, the overall open loop transfer function can be modeled by a term that can be

consistently estimated and an uncertain term which is proven to be bounded. Consequently, this allows one to borrow control design techniques from the area of robust control. However, it is also shown that the discrete nature of uncertainties in the topology of the network might require more specialized techniques leading to less conservative results.

6.1 A motivating example

Consider a networked system where nodes 1, 2, and 3 are being measured. It is known that node 1 causally influences node 2 and node 2 causally influences node 3. Therefore node 1 indirectly influences node 3. However, it is not known if there are other unmeasured variables in the system which might play a role influencing the nodes 1, 2, and 3.

For example, there could be an unmeasured node that influences at the same time node 1 and node 3, creating additional correlation between the two nodes 1 and 3 as shown in Figure 6.1 (a). In a scenario like this we refer to node 4 as a confounding variable.

Alternatively, node 1 could affect node 3 through via another parallel path that does not go through node 2 as shown in Figure 6.1 (b).

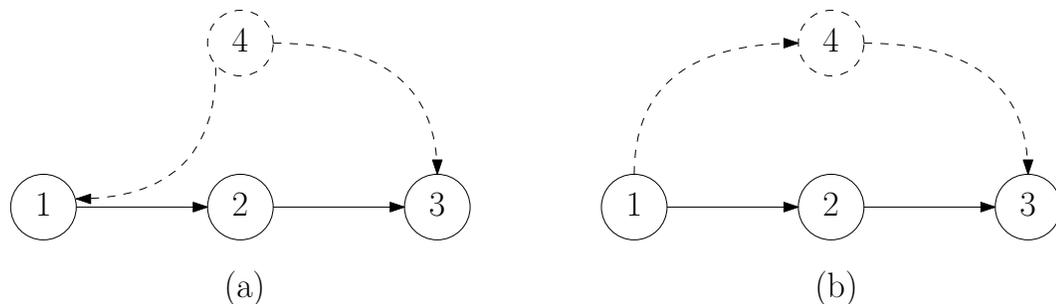


Figure 6.1: (a) Graphical representation of the network when node 4 is a confounder (b) Graphical representation of the network when node 4 is an intermediate node in a parallel path

Despite this uncertainty in the structure of the system, we would like to be able to design a controller to stabilize the closed-loop system via the input u that is injected to node 1 considering as output y the observation of node 3.

To design such controller we can consider a model of the system flexible enough to capture all the a priori information we have about its structural uncertainty. To this end, we assume that both edges $1 \rightarrow 4$ and $4 \rightarrow 1$ are simultaneously present in the graphical representation of the system with the assumption that one of the two transfer functions corresponding to these edges is zero. It is possible to describe such a system by the following set of equations.

$$x_1 = n_1 + H_{14}(z)x_4 + u \quad (\text{VI.1})$$

$$x_2 = n_2 + H_{21}(z)x_1 \quad (\text{VI.2})$$

$$x_3 = n_3 + H_{32}(z)x_2 + H_{34}(z)x_4 \quad (\text{VI.3})$$

$$x_4 = n_4 + H_{41}(z)x_1. \quad (\text{VI.4})$$

with the block diagram depicted in Figure 6.2.

To capture the uncertainty in the structure of the system we have the assumption

$$\|H_{14}(z)\| \|H_{41}(z)\| = 0. \quad (\text{VI.5})$$

Considering the graphical representation of system (VI.1) in Figure 6.3, assumption (VI.5) is equivalent with assuming that either the switch S_{41} is off ($H_{41} = 0$) or the switch

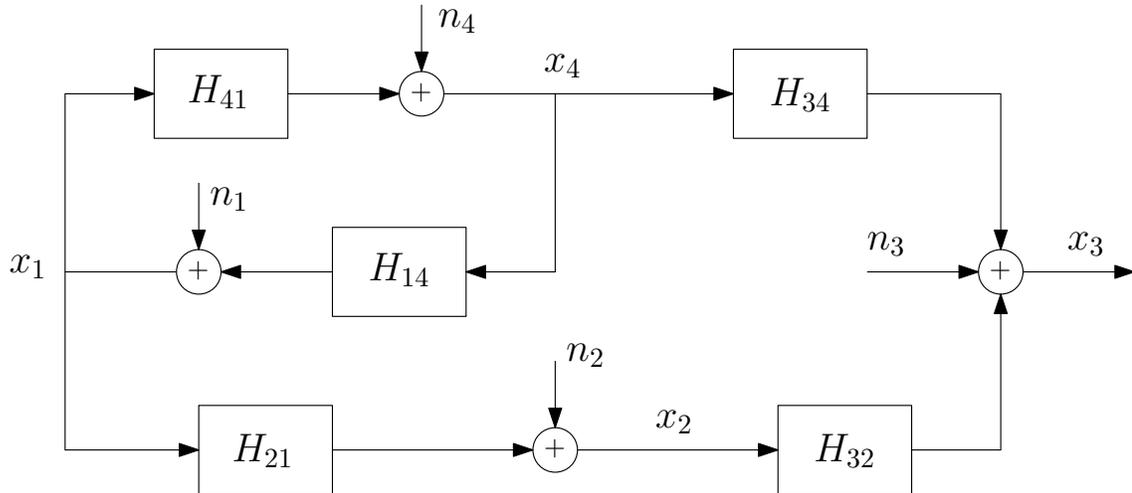


Figure 6.2: Block diagram of the class of networked systems described by (VI.1)

S_{14} is off ($H_{14} = 0$). This assumption can be interpreted as follows. By assumption

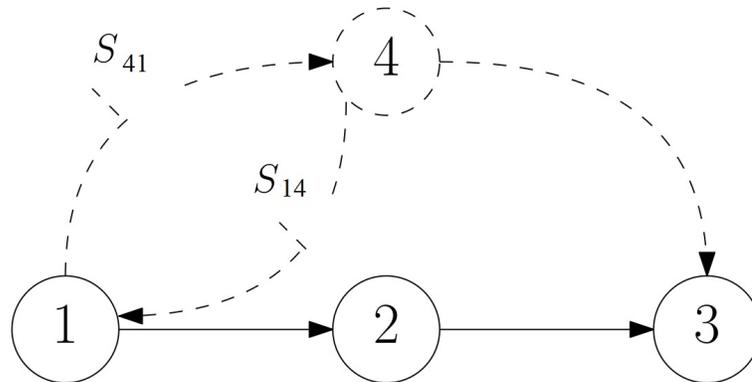


Figure 6.3: Assumption (VI.5) is equivalent with assuming that either the switch S_{41} is off ($H_{41} = 0$) or the switch S_{14} is off ($H_{14} = 0$).

(VI.5) we know that there is potentially a link between nodes 1 and 4 but we do not know its direction. There could be an edge from node 1 to node 4 or there could be an edge from node 4 to node 1 but not simultaneously (there is no feedback between nodes 1 and 4). Therefore, the system (VI.1) can be broken down into two cases depending on which switches are off (which transfer functions are zero).

If the switch S_{41} is off ($H_{41}(z) = 0$), and S_{14} is on, then the graphical representation of the system reduces to Figure 6.1 (a).

If the switch S_{41} is on and S_{14} is off ($H_{14}(z) = 0$) then the graphical representation of the system reduces to Figure 6.1 (b). We do not know which case we are in. That is, we know there is a link between nodes 1 and 4 but we do not know if its from 1 to 4 or from 4 to 1.

In presence of such uncertainty about the structure of the system, the objective is to design a robust controller C for the closed-loop system with reference input u that is injected to node 1 and output y which is the output of node 3.

6.2 Handling unmeasured Confounders

Consider a class of four-variate linear time-invariant discrete-time networked systems described by the following set of input-output equations.

$$x_1 = n_1 + H_{14}(z)x_4 \quad (\text{VI.6})$$

$$x_2 = n_2 + H_{21}(z)x_1 \quad (\text{VI.7})$$

$$x_3 = n_3 + H_{32}(z)x_2 + H_{34}(z)x_4 \quad (\text{VI.8})$$

$$x_4 = n_4 \quad (\text{VI.9})$$

It is assumed that variables x_1 , x_2 and x_3 are measured but x_4 is not. Figure 6.4 (a) shows the graphical representation of system (VI.6) where node 4 and its corresponding edges are hidden. That is, the known topology of the system is the graph depicted in Figure 6.4 (b).

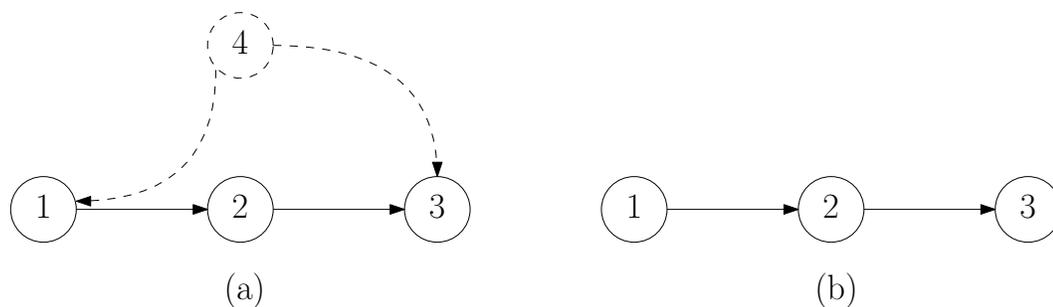


Figure 6.4: (a) The unknown true topology of the interconnections of the network (b) The known topology of the network

Suppose the objective is to design a robust controller C for the closed-loop system with reference input u that is injected to node 1 and output y which is the output of node 3.

Figure 6.6 shows the schematic of such scenario.

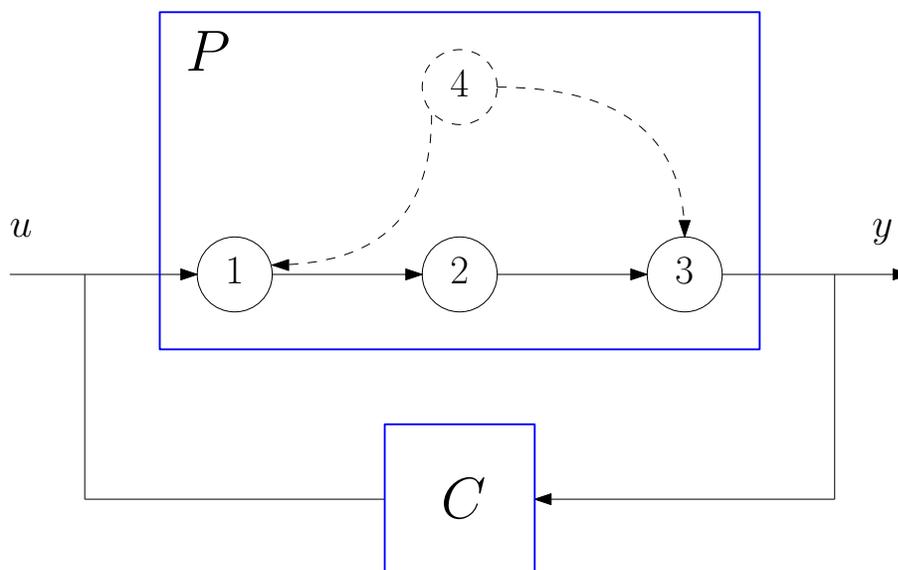


Figure 6.5: (a) The unknown true topology of the interconnections of the network (b) The known topology of the network

If node 4 did not exist, or either transfer functions $H_{14}(z)$ or $H_{34}(z)$ were zero, it would have been straightforward to identify the overall transfer function $P(z)$ from u to y . Indeed, in such a scenario, the transfer function $H_{21}(z)$ can be consistently identified as the

Wiener filter $W_{21}(z)$ corresponding to $x_1(t)$ when estimating $x_2(t)$ given $I_1(t)$. Namely,

$$\hat{H}_{21}(z) = W_{21}(z) \quad (\text{VI.10})$$

where

$$\mathbb{E}(x_2(t) | I_1(t)) = W_{21}(z)x_1(t). \quad (\text{VI.11})$$

Similarly, the transfer function $H_{32}(z)$ can be consistently identified as the Wiener filter $W_{32}(z)$ corresponding to $x_2(t)$ when estimating $x_3(t)$ given $I_2(t)$. Namely,

$$\hat{H}_{32}(z) = W_{32}(z) \quad (\text{VI.12})$$

where

$$\mathbb{E}(x_3(t) | I_2(t)) = W_{32}(z)x_2(t). \quad (\text{VI.13})$$

Consequently, the overall transfer function $P(z)$ from u to y can be consistently identified by

$$\hat{P}(z) = \hat{H}_{32}(z)\hat{H}_{21}(z) = W_{32}(z)W_{21}(z) \quad (\text{VI.14})$$

where $W_{21}(z)$ and $W_{32}(z)$ are computed via (VI.21) and (VI.22), respectively. When $P(z)$ is known, the controller C can be designed using variety of techniques.

However, if node 4 is present, $W_{32}(z)$ computed via (VI.22) is not going to be a consistent estimate of $H_{32}(z)$ and, consequently, (VI.14) is not going to be an accurate model for $P(z)$. Indeed, node 4 plays the role of a confounder between nodes 2 and 3 in estimation (VI.22) and potentially creates bias in $\hat{H}_{32}(z)$. In what follows, however,

we show that $H_{32}(z)$ can be consistently identified even in presence of an unmeasured confounder.

By Theorem 24, considering $j = 3$ and $Z = \{1\}$ we have that $W_{32}(z)$ in

$$\mathbb{E}(x_3(t) \mid I_{\{1,2\}}(t)) = \sum_{r \in \{1,2\}} W_{3r}(z)x_r(t). \quad (\text{VI.15})$$

is a consistent estimate of $H_{32}(z)$. On the other hand, $\hat{H}_{21}(z) = W_{21}(z)$ where $W_{21}(z)$ is computed by (VI.21) is still a consistent estimate of $H_{21}(z)$ even in presence of confounder 4.

Therefore, in presence of confounder 4, the overall transfer function $P(z)$ from u to y can be consistently identified by (VI.14) when $W_{21}(z)$ is computed by (VI.21) and $W_{32}(z)$ is computed by (VI.15).

Again, when $P(z)$ is consistently identified, the controller C can be designed using variety of techniques.

6.3 Robustness against graph uncertainties

In this section we consider a general scenario where both edges from node 1 to 4 and from 4 to 1 are simultaneously present. Consider a class of four-variate networked systems with a block diagram depicted in Figure 6.2 described by (VI.1) with the assumption that

$$\|H_{14}(z)H_{41}(z)\| \leq \epsilon < 1. \quad (\text{VI.16})$$

It is assumed that variables x_1 , x_2 and x_3 are measured but x_4 is not.

Suppose the objective is to design a robust controller C for the closed-loop system with reference input u that is injected to node 1 and output y which is the output of node 3. Figure 6.6 shows the schematic of such scenario.

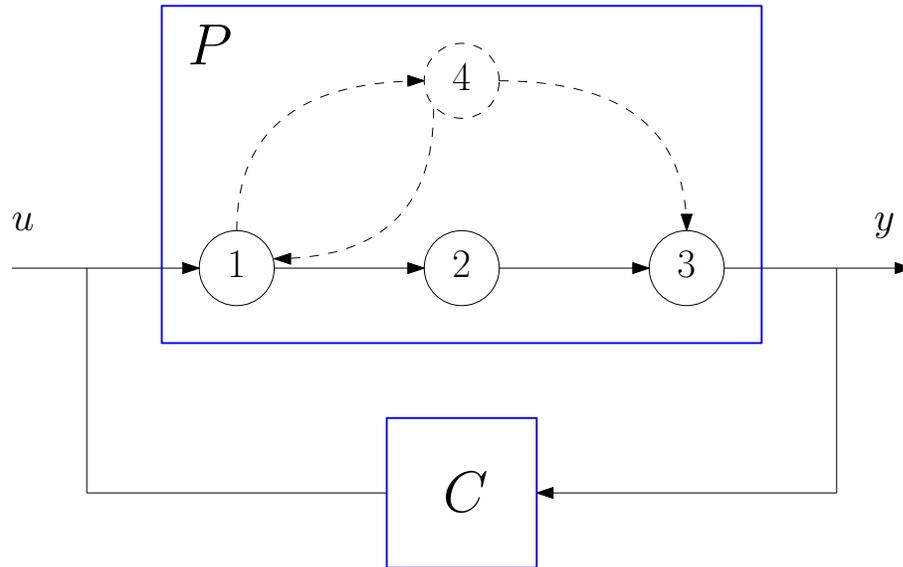


Figure 6.6: Designing a controller C when there is a feedback between 1 and 4

To design the controller C we need an estimate of the transfer function $P_{u \rightarrow y}(z)$. The unknown ground truth is that

$$P(z) = H_{21}(z)H_{32}(z) + \frac{H_{41}(z)H_{34}(z)}{1 - H_{14}(z)H_{41}(z)}. \quad (\text{VI.17})$$

Given only x_1 , x_2 , and x_3 , $P(z)$ cannot be consistently identified using conventional identification methods. We show, however, that it is possible to consistently identify

$$\bar{P}(z) = H_{21}(z)H_{32}(z) \quad (\text{VI.18})$$

and also find an upper bound for

$$\Delta(z) = \frac{H_{41}(z)H_{34}(z)}{1 - H_{14}(z)H_{41}(z)} \quad (\text{VI.19})$$

in all frequencies. This would enable us to model $P(z)$ as

$$P(z) = \bar{P}(z) + \Delta(z). \quad (\text{VI.20})$$

where $\bar{P}(z)$ is identifiable and $\Delta(z)$ can be seen as a bounded uncertainty.

6.3.1 Consistent Identification of $\bar{P}(z)$

First we show that $\bar{P}(z)$ is consistently identifiable. For the network of Figure 6.6, considering $i = 1$ and $j = 2$, we have that $Z = \{\emptyset\}$ satisfies the conditions of Theorem 2 for identification of $H_{21}(z)$. Thus, the transfer function $H_{21}(z)$ can be consistently identified as the Wiener filter $W_{21}(z)$ corresponding to $x_1(t)$ when estimating $x_2(t)$ using the information of $x_1(t)$. Namely,

$$\hat{H}_{21}(z) = W_{21}(z). \quad (\text{VI.21})$$

Similarly, considering $i = 2$, and $j = 3$ we have that $Z = \{1\}$ satisfies the conditions of Theorem 2 for identification of $H_{32}(z)$. Thus, $H_{32}(z)$ can be consistently identified as the Wiener filter $W_{32}(z)$ corresponding to $x_2(t)$ when estimating $x_3(t)$ using the information of $x_1(t)$ and $x_2(t)$. Namely,

$$\hat{H}_{32}(z) = W_{32}(z). \quad (\text{VI.22})$$

Therefore, the transfer function $\bar{P}(z)$ defined by (VI.18) can be consistently identified by

$$\hat{\bar{P}} = W_{32}(z)W_{21}(z) \quad (\text{VI.23})$$

when $W_{21}(z)$ is computed by (VI.21) and $W_{32}(z)$ is computed by (VI.22).

6.3.2 Finding an upper bound for Δ

Now we show that Δ defined in (VI.19) is bounded in all frequencies. Rewriting the equation governing x_1 we have

$$x_1 = \quad (\text{VI.24})$$

$$n_1 + H_{14}(z)x_4 = \quad (\text{VI.25})$$

$$n_1 + H_{14}(z)(n_4 + H_{41}(z)x_1) = \quad (\text{VI.26})$$

$$n_1 + H_{14}(z)n_4 + H_{14}(z)H_{41}(z)x_1. \quad (\text{VI.27})$$

Hence, we can write x_1 in terms of n_1 and n_4

$$x_1 = \frac{n_1 + H_{14}(z)n_4}{1 - H_{14}(z)H_{41}(z)}. \quad (\text{VI.28})$$

Define a new variable $x_q = x_3 - H_{32}(z)x_2$. Since a consistent estimate of $H_{32}(z)$ is determinable, we can assume that x_q is known. Simple algebraic manipulations of the

equation governing x_3 yields

$$x_q = \tag{VI.29}$$

$$x_3 - H_{32}(z)x_2 = \tag{VI.30}$$

$$n_3 + H_{34}(z)x_4 = \tag{VI.31}$$

$$n_3 + H_{34}(z)(n_4 + H_{41}(z)x_1) = \tag{VI.32}$$

$$n_3 + H_{34}(z)n_4 + H_{34}(z)H_{41}(z)x_1 \tag{VI.33}$$

Computing the power spectral densities of both sides of the equation we have

$$\begin{aligned} \Phi_{x_q}(z) = & \\ & \Phi_{n_3}(z) + |H_{34}(z)|^2\Phi_{n_4} + |H_{34}(z)H_{41}(z)|^2\Phi_{x_1}(z) \\ & + 2|H_{34}(z)|^2\left[\frac{H_{41}(z)H_{14}(z)}{1 - H_{41}(z)H_{14}(z)}\right]^*\Re(\Phi_{n_4}(z)) \end{aligned} \tag{VI.34}$$

where $\Phi_i(z)$ denotes the power spectral density of variable i and $H^*(z)$ denotes the complex conjugate of $H(z)$. Since $\Phi_{n_4}(z)$ is always real and even, we have that $\Re(\Phi_{n_4}(z)) =$

$\Phi_{n_4}(z)$. Thus, $\Phi_{x_q}(z)$ could be written in terms of $\Phi_{n_3}(z)$, $\Phi_{n_4}(z)$, and $\Phi_{x_1}(z)$ as follows.

$$\begin{aligned} \Phi_{x_q}(z) = & \\ & \Phi_{n_3}(z) + |H_{34}(z)H_{41}(z)|^2\Phi_{x_1}(z) + \\ & |H_{34}(z)|^2\Phi_{n_4}(z) \left(1 + 2\left[\frac{H_{41}(z)H_{14}(z)}{1 - H_{41}(z)H_{14}(z)}\right]^*\right) = \\ & \Phi_{n_3}(z) + |H_{34}(z)|^2\Phi_{n_4}(z) \left(\frac{1 + H_{14}^*(z)H_{41}^*(z)}{1 - H_{14}^*(z)H_{41}^*(z)}\right) \\ & + |H_{34}(z)H_{41}(z)|^2\Phi_{x_1}(z). \quad (\text{VI.35}) \end{aligned}$$

By assumption (VI.16) we have that $\|H_{14}(z)H_{41}(z)\| < 1$. Therefore, we get that the first two terms of the right hand side of the last equality are positive. Thus we get that

$$\|H_{34}(z)H_{41}(z)\|^2\Phi_{x_1}(z) \leq \Phi_{x_q}(z). \quad (\text{VI.36})$$

By Wiener–Khinchin theorem [104] we can compute the power spectral density of x_q in terms of power spectral densities of x_1 , x_2 , and x_3 (the processes that are measurable) as follows.

$$\begin{aligned}
\Phi_{x_q}(z) &= \begin{bmatrix} 0 \\ -H_{32}(z) \\ 1 \end{bmatrix}^T \begin{bmatrix} \Phi_{x_1}(z) & \Phi_{x_1,x_2}(z) & \Phi_{x_1,x_3}(z) \\ \Phi_{x_2,x_1}(z) & \Phi_{x_2}(z) & \Phi_{x_2,x_3}(z) \\ \Phi_{x_3,x_1}(z) & \Phi_{x_3,x_2}(z) & \Phi_{x_3}(z) \end{bmatrix} \begin{bmatrix} 0 \\ -H_{32}^*(z) \\ 1 \end{bmatrix} \\
&= H_{32}(z)\Phi_{x_2}(z)H_{32}^*(z) - H_{32}(z)\Phi_{x_2,x_3}(z) \\
&\quad - \Phi_{x_3,x_2}(z)H_{32}^*(z) + \Phi_{x_3}(z) \quad (\text{VI.37})
\end{aligned}$$

where $\Phi_{i,j}(z)$ denotes the cross power spectral density of variables i and j and $H_{32}(z)$ is given by

$$H_{32}(z) = \begin{bmatrix} \Phi_{x_3,x_1}(z) & \Phi_{x_3,x_2}(z) \end{bmatrix} \begin{bmatrix} \Phi_{x_1}(z) & \Phi_{x_1,x_2}(z) \\ \Phi_{x_2,x_1}(z) & \Phi_{x_2}(z) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (\text{VI.38})$$

Dividing both sides of (VI.36) by $\|1 - H_{14}(z)H_{41}(z)\|$ and replacing $\Phi_{x_q}(z)$ by its analytic expression derived in (VI.37) we get an upper bound for the uncertainty Δ :

$$\|\Delta(z)\| = \left\| \frac{H_{41}(z)H_{34}(z)}{1 - H_{14}(z)H_{41}(z)} \right\| \leq \sqrt{\frac{\Phi_{x_q}(z)}{(1 - \epsilon)^2 \Phi_{x_1}(z)}}. \quad (\text{VI.39})$$

The right hand side of the above equation is an upper bound for the uncertainty Δ .

Reconsidering (VI.20), we now have a consistent estimate of $\bar{P}(z)$ and have a bound for Δ . Therefore, (VI.20) is the same as standard models in robust control where the system (in this case $P(z)$) is modeled with a deterministic part (in this case $\bar{P}(z)$), and a bounded uncertain part (in this case Δ). There exist well-known methods to design a controller for such systems.

In the next section we show that by making ϵ to approach zero, we can reformulate the graph uncertainty as a robust control problem as was shown in this section.

6.3.3 Modeling graph uncertainty by making ϵ approaching zero

If we make ϵ in assumption (VI.16) to approach zero, we get the following assumption.

$$\|H_{14}(z)H_{41}(z)\| = 0. \quad (\text{VI.40})$$

In other words, for $\epsilon = 0$, assumption (VI.16) reduces to assumption (VI.40). Assumption (VI.40) requires that, for rational transfer functions, either $H_{14}(z)$ or $H_{41}(z)$ to be zero. If $H_{41}(z) = 0$ then the graphical representation of the system reduces to Figure 6.1 (a). If $H_{14}(z) = 0$ then the graphical representation of the system reduces to Figure 6.1 (b). Thus, by putting a limit ϵ on the modulus of the feedback loop between nodes 1 and 4 in assumption (VI.16) and making ϵ to approach zero to get the assumption (VI.40), we have established a way to model graph uncertainty where it is not known if the graphical representation of the system is the graph of Figure 6.1 (a) or the graph of Figure 6.1 (b).

Similar to the previous case, when the objective is to design a robust controller C for the closed-loop system with reference input u that is injected to node 1 and output y which

is the output of node 3, the unknown transfer function $P_{u \rightarrow y}(z)$ is given by (VI.17) and can be modeled as (VI.20) where $\bar{P}(z)$ is the same as (VI.18) and Δ is the same as (VI.19).

Taking similar steps, $\bar{P}(z)$ can be consistently identified as before and an upper bound for Δ can be found.

$$\|\Delta(z)\| = \left\| \frac{H_{41}(z)H_{34}(z)}{1 - H_{14}(z)H_{41}(z)} \right\| \leq \sqrt{\frac{\Phi_{x_q}(z)}{\Phi_{x_1}(z)}}. \quad (\text{VI.41})$$

Thus, for $\epsilon = 0$, similar to the previous case, we can consistently identify $\bar{P}(z)$ and have a bound for Δ . Therefore, we can use robust control techniques to stabilize the closed loop system.

6.4 A tighter bound for uncertainty

For a networked system described by (VI.1), we showed that it is possible to formulate the uncertainty in the graphical representation of the system in a robust control paradigm. This was done by putting a limit ϵ on the modulus of the feedback loop between nodes 1 and 4 in assumption (VI.16) and making ϵ to approach zero to get the assumption (VI.40). This approach led to an upper-bound for the uncertainty Δ given by (VI.41)

Assumption (VI.40), however, is slightly different from assumption (VI.5) which describes the uncertainty in the graphical representation of the system. In this section, we show that the discrete nature of uncertainty in graphical representation, and the fact that assumption (VI.5) leads to a binary situation for the graphical representation of the system (VI.1), enables us to find a tighter bound for Δ .

Consider the system (VI.1) with assumption (VI.5) which means that the graphical representation of the system is the graph of Figure 6.1 (a) when $H_{41}(z) = 0$, or the graph of Figure 6.1 (a) when $H_{14}(z) = 0$ (the case where both $H_{14}(z)$ and $H_{41}(z)$ are zero is trivial).

When $H_{41}(z) = 0$ the transfer function $H_{21}(z)$ can be computed by

$$\hat{H}_{21}(z) = \frac{\Phi_{x_2,x_1}(z)}{\Phi_{x_1}(z)} \quad (\text{VI.42})$$

and $H_{32}(z)$ can be computed by

$$\hat{H}_{32}(z) = \begin{bmatrix} \Phi_{x_3,x_1}(z) \\ \Phi_{x_3,x_2}(z) \end{bmatrix}^T \begin{bmatrix} \Phi_{x_1}(z) & \Phi_{x_1,x_2}(z) \\ \Phi_{x_2,x_1}(z) & \Phi_{x_2}(z) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (\text{VI.43})$$

Thus $P_{u \rightarrow y}(z)$ is equal to $P_a(z)$ given by

$$P_a(z) = \frac{\Phi_{x_2,x_1}(z)}{\Phi_{x_1}(z)} \begin{bmatrix} \Phi_{x_3,x_1}(z) \\ \Phi_{x_3,x_2}(z) \end{bmatrix}^T \begin{bmatrix} \Phi_{x_1}(z) & \Phi_{x_1,x_2}(z) \\ \Phi_{x_2,x_1}(z) & \Phi_{x_2}(z) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (\text{VI.44})$$

On the other hand, if $H_{14}(z) = 0$, then $P_{u \rightarrow y}(z)$ is equal to $P_b(z)$ given by

$$P_b(z) = \frac{\Phi_{x_3,x_1}(z)}{\Phi_{x_1}(z)} \quad (\text{VI.45})$$

Now that we have estimates for both cases ($H_{14}(z) = 0$ or $H_{41}(z) = 0$), we can model $P_{u \rightarrow y}(z)$ as the midpoint of $P_a(z)$ and $P_b(z)$ with uncertainty Δ :

$$\hat{P}_{u \rightarrow y}(z) = \frac{P_a(z) + P_b(z)}{2} + \Delta(z) \quad (\text{VI.46})$$

where Δ is bounded by

$$\|\Delta(z)\| \leq \left\| \frac{P_a(z) - P_b(z)}{2} \right\| \quad (\text{VI.47})$$

Equation (VI.47) provides a tighter bound for Δ compared to (VI.41) when we know either $H_{14}(z)$ or $H_{41}(z)$ is zero.

The following example shows that for some realizations of the network it is possible to apply the robust control paradigm with the upper bound (VI.47) while the bound (VI.41) will not lead to a solution.

Example 23. Consider a networked system described by (VI.1) with the assumption that there is no feedback between nodes 1 and 4. The transfer functions are not known, only nodes 1, 2, and 3 are measured, and we want to design a controller C to stabilize the closed loop system with input u and output $y = x_3$. Consider a realization of the network where $H_{14}(z) = H_{41}(z) = 0$, $\|H_{34}(z)\| \gg \|H_{32}(z)\| \gg \|H_{21}(z)\|$, and $\text{var}(n_4) \gg \text{var}(n_3) \gg \text{var}(n_2) \gg \text{var}(n_1)$, where var is the variance function. In such scenario $\Phi_{x_q}(z)$ will be relatively large and $\Phi_{x_1}(z)$ will be relatively small making the right hand side of equation (VI.41), namely $\sqrt{\frac{\Phi_{x_q}(z)}{\Phi_{x_1}(z)}}$, very large. Therefore, the technique where the bound for Δ is computed via (VI.41) will not be applicable. On the other hand, using the technique presented in this section, computing $P_a(z)$ given by (VI.44) and $P_b(z)$ given by (VI.45), we

will have that the upper bound for Δ given by the right hand side of (VI.47) will be close to zero because $P_a(z)$ and $P_b(z)$ are the same. Therefore, any controller C that stabilizes $P_a(z)$, will stabilize $P_b(z)$ and $P_{u \rightarrow v}$. Similarly, any controller C that stabilizes $P_b(z)$, will stabilize $P_a(z)$ and $P_{u \rightarrow v}$.

So far we saw how we can formulate the uncertainty in the graphical representation of the network in a robust control problem. It is important to mention that this approach may lead to the discardment of some controllers that would have shown satisfactory performance if applied in both cases of $H_{14}(z) = 0$ or $H_{41}(z) = 0$. This is due to the fact that graph uncertainty has a discrete nature.

The following example shows that modeling a graph uncertainty as a nominal plant plus a bounded perturbation might lead to more conservative results because the approach might miss some stabilizing controllers.

Example 24. Consider a network described by (VI.1) where the unknown random processes n_j are mutually independent zero mean white Gaussian processes with variance 0.01 and the unknown transfer functions are given by

$$H_{41}(z) = \frac{1}{z - 0.5} \quad (\text{VI.48})$$

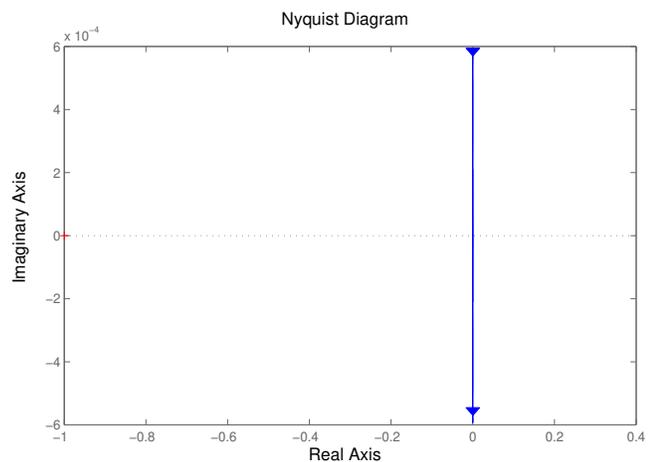
$$H_{21}(z) = \frac{0.01}{z - 0.4} \quad (\text{VI.49})$$

$$H_{32}(z) = \frac{0.03}{z - 0.3} \quad (\text{VI.50})$$

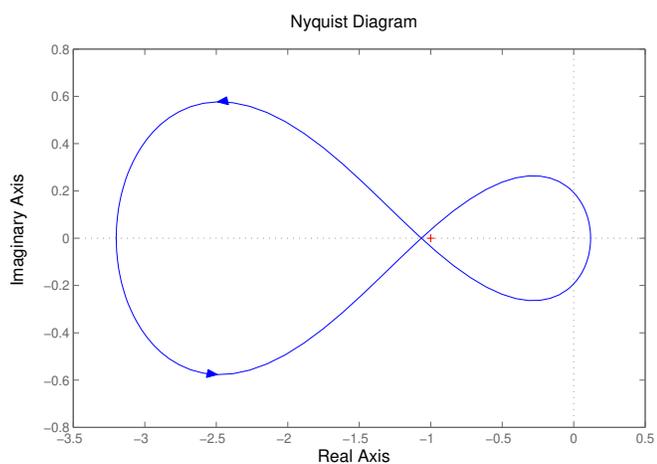
$$H_{34}(z) = \frac{0.4}{z - 1.25} \quad (\text{VI.51})$$

$$H_{14}(z) = 0 \quad (\text{VI.52})$$

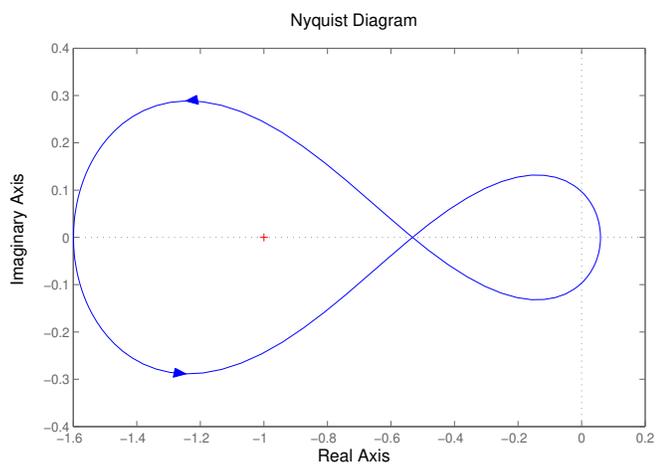
$$H_{13}(z) = C(z) = -\frac{2}{3} - k. \quad (\text{VI.53})$$



(a)



(b)



(c)

Figure 6.7: (a) The Nyquist plot of $P_a(z)$ in Example 24, (b) The Nyquist plot of $P_b(z)$ in Example 24, (c) The Nyquist plot of $\frac{P_a(z)+P_b(z)}{2}$ in Example 24

Nodes 1, 2, and 3 are measured and node 4 is hidden and $k \in R$ is initially equal to zero. The objective is to design a proportional controller $C(z) = -\frac{2}{3} - k \in R$ by manipulating k under assumption (VI.5) to stabilize the plant $P_{u \rightarrow y}(z)$ where u is a control signal injected into node 1 and y is the output of node 3 in such a way that all the poles of the closed loop system lie out of the circle with radius $\frac{1}{4}$ and inside the circle with radius 0.91 in z -plane. The unknown ground truth is that

$$P_{u \rightarrow y}(z) = \frac{0.003z^2 - 0.2805z + 0.04819}{z^4 - 2.45z^3 + 1.97z^2 - 0.6475z + 0.075}. \quad (\text{VI.54})$$

For $k = 0$, transfer functions $P_a(z)$ and $P_b(z)$ can be consistently identified by closed loop identification techniques [80]. Thus, considering a model of the form (VI.46) for the system, $P_{u \rightarrow y}(z)$ can be modeled by (VI.46) as

$$\hat{P}_{u \rightarrow y}(z) = \Delta(z) + \frac{0.4006z^4 - 0.5615z^3 + 0.2932z^2 - 0.6759z + 0.005805}{2z^6 - 6.3z^5 + 7.61z^4 - 4.641z^3 + 1.529z^2 - 0.2604z + 0.018}. \quad (\text{VI.55})$$

where an upper bound for Δ is given by (VI.47). The Nyquist plots of $P_a(z)$, $P_b(z)$, and $\frac{P_a(z)+P_b(z)}{2}$ are shown in Figure 6.7, respectively.

Consider a proportional controller

$$C(z) = -\frac{1}{2}, \quad (\text{VI.56})$$

by designing k to be $-\frac{1}{6}$. $P_a(z)$ in feedback with $C(z) = \frac{-1}{2}$ results in the closed loop system with poles at 0.3985 and 0.3015 which satisfy the design requirements. Similarly, $P_b(z)$ in feedback with $C(z) = \frac{-1}{2}$ results in the closed loop system

$$\frac{P_b(z)}{1 - P_b(z)C(z)} = \frac{0.4003z^2 - 0.2805z + 0.04819}{z^4 - 2.45z^3 + 2.17z^2 - 0.7878z + 0.09909} \quad (\text{VI.57})$$

with poles at $0.8749 \pm 0.2436i$, 0.3995, and 0.3007 which satisfy the design requirements. Therefore, even in presence of uncertainty about the structure of the system where we do not know whether the structure of the network is Figure 6.1 (a) or Figure 6.1 (b), we can be sure that the controller $C(z) = \frac{-1}{2}$ stabilizes the system meeting the performance requirements.

However, following the approach of modeling the system by (VI.46), $\frac{P_a(z)+P_b(z)}{2}$ (the second term in the right hand side of (VI.55)) in feedback with $C(z) = \frac{-1}{2}$ results in a closed loop system with one of its poles being at 1.0766. Thus, regardless of Δ , we can not conclude that the system could be stabilized with the controller $C(z) = \frac{-1}{2}$ when following (VI.46).

Example 24 shows that because of the discrete nature of uncertainties in the structure of the network, applying a model for the uncertainties that only considers a bound, may lead to missing some controllers that could have worked for all possible structures. Thus, more specialized methods might lead to a larger set of controllers.

CHAPTER VII

CONCLUSION AND DISCUSSION

In this dissertation we have attempted to lay a graph-theoretic foundation for identification in dynamic networks. We have shown that many physical phenomena can be modeled as dynamic networks and that there is a strong motivation to identify particular transfer functions embedded in the dynamic networks.

We approached the problem of identifying certain transfer functions in a dynamic network by connecting causal inference methods developed in the area of graphical models with more recent control theoretic results in closed loop system identification. We introduced graphical conditions on the set of auxiliary variables in order to consistently identify a certain transfer function in a partially observed causal dynamic network via a prediction error algorithm using only observational data. The results extend previous techniques borrowing elements from the theory of Structural Equation Models, Graphical Models and System Identification. One main advantage is that a consistent identification can be obtained for a network with no algebraic loops even when the class of parameterized models is allowed to contain algebraic loops. This is achieved by devising specific tests to detect

strictly causal transfer functions. Most importantly, the graphical conditions on the set of auxiliary variables are proven to be sufficient and necessary. This characterization allows one to formulate identification problems while at the same time optimizing a cost function to take into account the potential cost of the observations.

In particular, it was shown that sufficient and necessary conditions for consistent estimation could be reformulated as the notion of d -separation in a slightly modified graph of the network. An algorithm was presented to determine the optimal set of observations by finding the minimum cut set in a flow graph systematically created from the graphical representation of the network. Extending the results to the non-trivial case where multiple causal relations needed to be simultaneously estimated, it was shown that the optimal set of observations could be determined via an optimal multi-commodity flow problem with additional commodity specific constraints.

Finally we showed that it is possible to leverage our results of identification of dynamic networks to introduce a notion of robustness with respect to uncertainties in the graph structure of a distributed system. When the objective is designing a controller for the overall system, the topological uncertainties impede the exact identification of the overall open-loop transfer function and consequently, the exact design of the controller. It was shown, however, that some of the transfer functions of the network could be consistently identified using our identification results and, in some cases, the overall open loop transfer function can be modeled by a term that can be consistently estimated and an uncertain term which is proven to be bounded. Consequently, this allows one to borrow control design techniques from the area of robust control.

Appendix A

1.1 Optimal selection of observations in acyclic networks

If the target node j is not involved in a feedback loop the result of Chapter V could be simplified. In this chapter, we provide such simplified results for identification of a single transfer function in an acyclic network while minimizing the cost of observations.

Consider a class of dynamic networks where the node j is not involved in a feedback loop and suppose that the objective is the identification of a single transfer function $H_{ji}(z)$. Suppose that a subset O of the nodes of the network is measurable. The following result provides a criterion for the selection of a set of additional auxiliary variables Z that leads to an unbiased estimate of $H_{ji}(z)$.

Theorem 24. *Consider a network with a graphical representation G where node j is not involved in a feedback loop. Let G^a be the subgraph of G limited to the nodes in $an_G(i, j)$ and let G' be the mutilated graph [56] resulted from removing the edge $i \rightarrow j$ from G^a . If*

- (i) *a set $Z \cap de_G(j) = \emptyset$ d -separates nodes i and j in G' ,*

then $W_{ji}(z)$ in

$$\mathbb{E}(x_j(t) | I_{i \cup Z}(t)) = \sum_{r \in \{i\} \cup Z} W_{jr}(z) \quad (\text{A.1})$$

is a consistent estimate of $H_{ji}(z)$ if the power spectral density matrix of (x_i, x_j, x_Z) is non-singular.

Proof. First we show that $Z \cap \text{de}_G(j) = \emptyset$. By contradiction suppose that there is a node $k \in Z$ that is a descendent of node j in G . Since k is a descendent of node j in G , there is a dipath from node j to node k in G . Since Z is defined in G' and G' is resulted from G^a , every node in Z is an ancestor of j in G . Thus, there is a dipath from node k to node j in G . Since there is a dipath from node j to node k in G and there is a dipath from node k to node j in G , node j is involved in a feedback loop which is a contradiction. Thus, we have that Z does not contain any descendent of node j . Define a new variable $x_q(t)$ as follows

$$x_q(t) = x_j(t) - H_{ji}(z)x_i(t). \quad (\text{A.2})$$

Since we verified that $Z \cap \text{de}_G(j) = \emptyset$, it follows from Lemma 25 and Theorem 26 in [64] that

$$x_q(t) \perp I_i(t) | I_Z(t). \quad (\text{A.3})$$

That is, when estimating $x_q(t)$ from $I_{\{i\} \cup Z}(t)$ the transfer function corresponding $x_i(t)$ is zero (see Definition II.7):

$$\mathbb{E}(x_q(t) | I_{\{i\} \cup Z}(t)) = \mathbb{E}(x_q(t) | I_Z(t)) = \sum_{r \in Z} W_{jr}(z)x_r(t) \quad (\text{A.4})$$

Therefore, we have

$$\begin{aligned}
\mathbb{E}(x_j(t) \mid I_{i \cup Z}(t)) &= \\
\mathbb{E}(x_q(t) + H_{ji}(z)x_i(t) \mid I_{i \cup Z}(t)) &= \\
H_{ji}(z)x_i(t) + \mathbb{E}(x_q(t) \mid I_{i \cup Z}(t)) &= \\
H_{ji}(z)x_i(t) + \sum_{r \in Z} W_{jr}(z)x_r(t). &
\end{aligned}
\tag{A.5}$$

Since power spectral density matrix of (x_i, x_j, x_Z) is non-singular, comparing the expressions for $\mathbb{E}(x_j(t) \mid I_{i \cup Z}(t))$ we get that $W_{ji}(z) = H_{ji}(z)$ which verifies the assertion. \square

Theorem 24 provides sufficient conditions for consistent identification of a certain transfer function $H_{ji}(z)$ using equation (A.1) in a class of networks where node j is not involved in a feedback loop. It turns out that for such a class of networks condition (i) of Theorem 24 is actually necessary.

Theorem 25. *Consider all graphs G and graphs G' resulted from limiting G to the nodes in $an_G(i, j)$ and removing the edge $i \rightarrow j$. If the graphical condition (i) of Theorem 24 is not satisfied by a set Z in G' , there exists a network $\mathcal{G} = (H(z), n)$ with graphical representation G such that $W_{ji}(z)$ in (A.1) will be a biased estimate of $H_{ji}(z)$.*

Proof. For Gaussian input processes and scalar transfer functions the assertion follows from Lemma 11 of [105]. \square

Now we show that it is possible to design an algorithm to systematically find the solution Z^* of Problem 2. In particular, we reformulate conditions of Theorem 24 and optimality condition a few times to reach to a problem that we can systematically solve.

The following example illustrates how using lemmas 14 and 15 we can manipulate graph G to limit our search for the set Z^* .

Example 25. *Suppose we are looking for an optimal set Z^* with respect to the cost function (V.2) to identify the transfer function $H_{21}(z)$ using (A.1) in a network with a graphical representation G shown in Figure 1.1. Figure 1.2 shows the subgraph of G limited to*

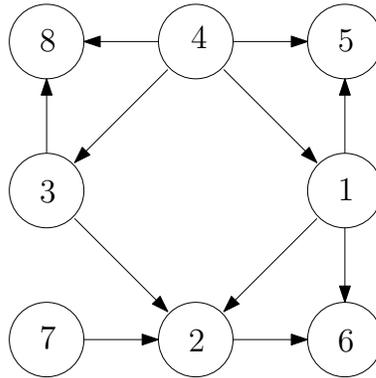


Figure 1.1: A graphical representation G of a network discussed in Example 25. The objective is identification of the transfer function $H_{21}(z)$ using (A.1) and minimizing the cost of observation (V.1).

the ancestors of nodes 1 and 2. Figure 1.3 shows the mutilated graph G' resulted from removing the edge $1 \rightarrow 2$ from G^a . Lemmas 14 and 15 state that to identify the transfer function $H_{21}(z)$ we can look for a set that d -separates nodes 1 and 2 in the graph G' in Figure 1.3.

The fact that we limited ourselves to the ancestor graph G^a in our search for Z^* allows us to reformulate the d -separation conditions in graph G' to separation conditions in an undirected graph.

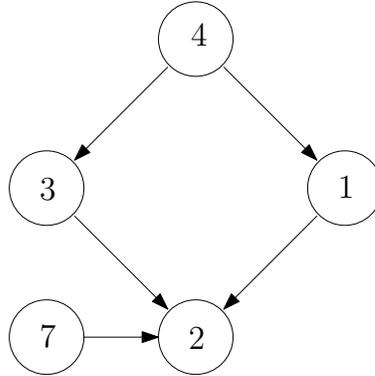


Figure 1.2: The ancestor graph G^a resulted from limiting graph G of Figure 1.1 to ancestors of nodes 1 and 2 discussed in Example 25.

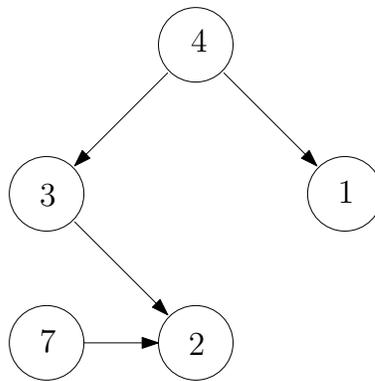


Figure 1.3: The mutilated graph G' resulted from removing the edge $1 \rightarrow 2$ from the graph G^a of Figure 1.2 discussed in Example 25. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set d -separating nodes 1 and 2 in G' .

As a corollary of Theorem 16 we can claim that we can reformulate the sufficient and necessary conditions for consistent identification of $H_{ji}(z)$ using (A.1) to the notion of separation in an undirected graph.

Corollary 25.1. *Consider all networks with graphical representation $G = (V, E)$ and $i \in pa_G(j)$. Let G^a be the subgraph of G limited to the nodes in $an_G(i \cup j)$ and G' be the graph obtained by removing the edge $i \rightarrow j$ from G^a . Let G^{mor} be the moral graph of G' . Then the set Z separates nodes i and j in G^{mor} if and only if the transfer function $W_{ji}(z)$*

in (A.1) is a consistent estimate of the transfer function from node i to node j in all the networks.

Proof. The result is an immediate consequence of Theorem 16. \square

Corollary 25.1 says that if we want to consistently identify a transfer function $H_{ji}(z)$ using (A.1) and minimize the cost of observations, we can look for a set Z^* that, among all sets Z that separate node $\{i\}$ from node $\{j\}$ in graph G^{mor} , has the minimum cost.

Example 26. Suppose we are looking for an optimal set Z^* with respect to the cost function (V.2) to identify the transfer function $H_{21}(z)$ in a network with a graphical representation G shown in Figure 1.1. Corollary 25.1 states that, to find Z^* , we can look for an optimal set Z^* that separates nodes 1 and 2 in graph G^{mor} shown in Figure 1.4 which is resulted from moralizing graph G' depicted in Figure 1.3.

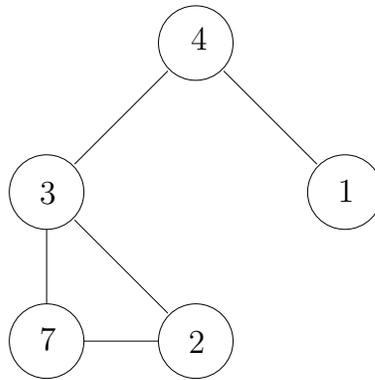


Figure 1.4: The undirected graph G^{mor} discussed in Example 26 resulted from moralizing the graph G' of Figure 1.3. To find the optimal predictors set Z^* to identify the transfer function $H_{21}(Z)$ we can look for an optimal set separating nodes 1 and 2 in G^{mor} .

Using the techniques explained in Section 5.1, we can find the optimal separator in G^{mor} and consequently Z^* . Algorithm 4 summarizes the steps that lead to the optimal set

of predictors with respect to (V.1) for identifying a certain transfer function $H_{ji}(z)$ using (A.1) in an acyclic network.

Algorithm 4 Finding the optimal set of auxiliary variables

- 1: **Given:** topology $G = (V, E)$, target link $i \rightarrow j$, cost c_k for $k \in V$
 - 2: **Output:** optimal Z^*
 - 3: $G^a = (V^a, E^a) \leftarrow$ restrict G to $\text{an}_G(\{i, j\})$
 - 4: $G' :=$ remove the edge $i \rightarrow j$ from G^a
 - 5: $G^{mor} \leftarrow$ moralize G'
 - 6: assign the costs
 - 7: $Z^* \leftarrow$ the optimal set separating $\{i\}$ from $\{j\}$ in G^{mor}
-

To show how the method presented in this section can be applied to acyclic networks to find the optimal set of predictors for consistent identification of a certain transfer function, we consider the following example.

Example 27. Consider a network with a graphical representation G depicted in Figure 1.5.

The objective is the optimal identification of the transfer function $H_{21}(z)$ using Theorem

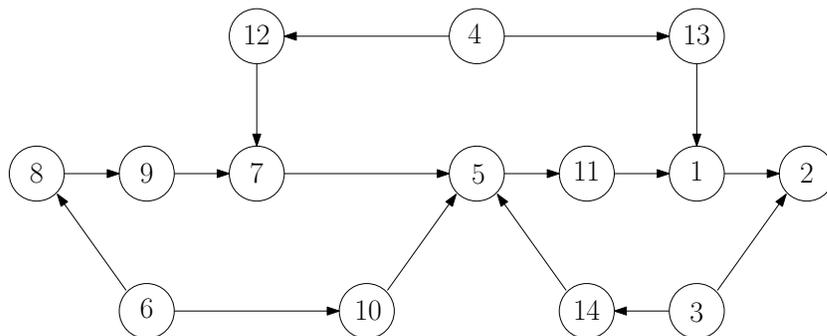


Figure 1.5: The graphical representation G of the network discussed in Example 27 for the determination of the set of predictors with minimal cost.

24 and by solving (V.1) where the cost $C(Z)$ is defined by (V.2). The cost c_r of observing each node $r \in V$ is reported in Table 1.1. Since all the nodes are ancestors of node j the ancestor graph G^a is the same as G and the mutilated graph G' is resulted after removing the edge $i \rightarrow j$. Figure 1.6 shows the undirected graph G^{mor} which is resulted from

Table 1.1: Costs of observing each node in Example 27

node r	c_r						
3	190	6	30	9	30	12	160
4	160	7	10	10	130	13	120
5	10	8	20	11	120	14	140

moralizing G' . Based on the results of this paper an optimal set of predictors guaranteeing

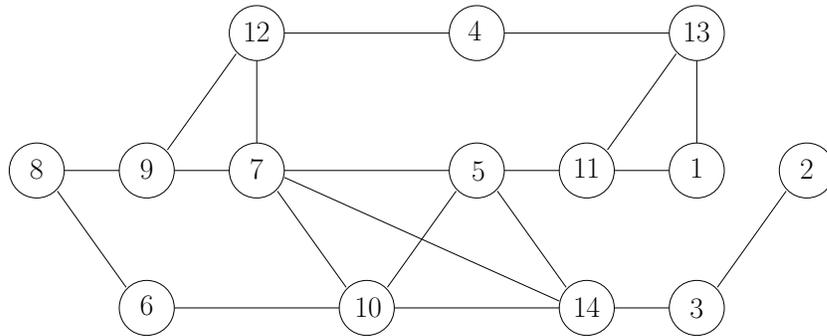


Figure 1.6: The undirected graph G^{mor} in Example 27 resulted from moralizing graph G' which is obtained by removing the edge $11 \rightarrow 2$ from G^a which is the subgraph of G limited to ancestors of nodes $\{1, 2\}$. The optimal set separating nodes 1 and 2 in G^{mor} is the optimal set of predictors guaranteeing consistent identification of $H_{21}(z)$ using (A.1).

consistent identification of $H_{21}(z)$ using the specific identification method (A.1) minimizing the cost (V.1) can be determined by finding the optimal separator set separating nodes 1 and 2 in G^{mor} . It turns out that the optimal set separating nodes 1 and 2 in G^{mor} is the set $Z^* = \{5, 7, 8\}$ with the cost

$$C(\{5, 7, 8\}) = \sum_{k \in \{5, 7, 8\}} c_k = 10 + 10 + 20 = 40. \quad (\text{A.6})$$

Thus, $W_{21}(z)$ in

$$\mathbb{E}(x_2(t) \mid I_{\{1, 5, 7, 8\}}(t)) = \sum_{r \in \{1, 5, 7, 8\}} W_{2r}(z) x_r(t) \quad (\text{A.7})$$

is a consistent estimate of $H_{21}(z)$ that requires minimal observation cost.

REFERENCES

- [1] A. Dankers, P. M. Van den Hof, D. Materassi, and H. H. Weerts, “Conditions for handling confounding variables in dynamic networks,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 3983–3988, 2017.
- [2] P. Kundur, “power system analysis and control,” 1994.
- [3] J. G. Proakis, M. Salehi, N. Zhou, and X. Li, *Communication systems engineering*, vol. 2. Prentice Hall New Jersey, 1994.
- [4] W. Ren and R. W. Beard, *Distributed consensus in multi-vehicle cooperative control*, vol. 27. Springer, 2008.
- [5] K. E. Johnson and N. Thomas, “Wind farm control: Addressing the aerodynamic interaction among wind turbines,” in *2009 American Control Conference*, pp. 2104–2109, IEEE, 2009.
- [6] M. Soleimanzadeh and R. Wisniewski, “Controller design for a wind farm, considering both power and load aspects,” *Mechatronics*, vol. 21, no. 4, pp. 720–727, 2011.
- [7] K.-P. Kim, D. Chang, S. K. Lim, S.-K. Lee, H.-K. Lyu, and D.-K. Hwang, “Thermal annealing effects on the dynamic photoresponse properties of al-doped zno nanowires network,” *Current Applied Physics*, vol. 11, no. 6, pp. 1311–1314, 2011.
- [8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, “Complex networks: Structure and dynamics,” *Physics reports*, vol. 424, no. 4-5, pp. 175–308, 2006.

- [9] N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S. A. Teichmann, and M. Gerstein, “Genomic analysis of regulatory network dynamics reveals large topological changes,” *Nature*, vol. 431, no. 7006, pp. 308–312, 2004.
- [10] J. Luo and L. Kuang, “A new method for predicting essential proteins based on dynamic network topology and complex information,” *Computational biology and chemistry*, vol. 52, pp. 34–42, 2014.
- [11] P. Petousis, S. X. Han, D. Aberle, and A. A. Bui, “Prediction of lung cancer incidence on the low-dose computed tomography arm of the national lung screening trial: A dynamic bayesian network,” *Artificial intelligence in medicine*, vol. 72, pp. 42–55, 2016.
- [12] S. Anzellotti, D. Kliemann, N. Jacoby, and R. Saxe, “Directed network discovery with dynamic network modelling,” *Neuropsychologia*, vol. 99, pp. 1–11, 2017.
- [13] C. N. Kaiser-Bunbury, J. Mougil, A. E. Whittington, T. Valentin, R. Gabriel, J. M. Olesen, and N. Blüthgen, “Ecosystem restoration strengthens pollination network resilience and function,” *Nature*, vol. 542, no. 7640, pp. 223–227, 2017.
- [14] P. Monestiez, J.-S. Bailly, P. Lagacherie, and M. Voltz, “Geostatistical modelling of spatial processes on directed trees: Application to fluvisol extent,” *Geoderma*, vol. 128, no. 3-4, pp. 179–191, 2005.
- [15] V. M. Carvalho, “From micro to macro via production networks,” *Journal of Economic Perspectives*, vol. 28, no. 4, pp. 23–48, 2014.
- [16] R. N. Mantegna and H. E. Stanley, *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press, 1999.

- [17] M. Nogal, A. O'Connor, B. Caulfield, and B. Martinez-Pastor, "Resilience of traffic networks: From perturbation to recovery via a dynamic restricted equilibrium model," *Reliability Engineering & System Safety*, vol. 156, pp. 84–96, 2016.
- [18] V. Chetty, N. Woodbury, J. Brewer, K. Lee, and S. Warnick, "Applying a passive network reconstruction technique to twitter data in order to identify trend setters," in *2017 IEEE Conference on Control Technology and Applications (CCTA)*, pp. 1344–1349, IEEE, 2017.
- [19] C. Alexander, *Market models: A guide to financial data analysis*. John Wiley & Sons, 2001.
- [20] M. B. Eisen, P. T. Spellman, P. O. Brown, and D. Botstein, "Cluster analysis and display of genome-wide expression patterns," *Proceedings of the National Academy of Sciences*, vol. 95, no. 25, pp. 14863–14868, 1998.
- [21] A. Carol, "Market models: a guide to financial data analysis," *John Willer & Sons*, 2001.
- [22] T. D. Cook, D. T. Campbell, and W. Shadish, *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston, MA, 2002.
- [23] K. Soltesz, K. Van Heusden, G. A. Dumont, T. Hägglund, C. L. Petersen, N. West, and J. M. Ansermino, "Closed-loop anesthesia in children using a pid controller: A pilot study," *IFAC Proceedings Volumes*, vol. 45, no. 3, pp. 317–322, 2012.
- [24] G. Deuschl, C. Schade-Brittinger, P. Krack, J. Volkmann, H. Schäfer, K. Bötzel, C. Daniels, A. Deutschländer, U. Dillmann, W. Eisner, *et al.*, "A randomized trial of deep-brain stimulation for parkinson's disease," *New England Journal of Medicine*, vol. 355, no. 9, pp. 896–908, 2006.

- [25] A. Muir and J. Lopatto, “Final report on the august 14, 2003 blackout in the united states and canada: causes and recommendations,” 2004.
- [26] D. Materassi and G. Innocenti, “Unveiling the connectivity structure of financial networks via high-frequency analysis,” *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 18, pp. 3866–3878, 2009.
- [27] D. Materassi and G. Innocenti, “Topological identification in networks of dynamical systems,” *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1860–1871, 2010.
- [28] R. Kumar, A. Lozano, Y. Kim, W. Hutchison, E. Sime, E. Halket, and A. Lang, “Double-blind evaluation of subthalamic nucleus deep brain stimulation in advanced parkinson’s disease,” *Neurology*, vol. 51, no. 3, pp. 850–855, 1998.
- [29] C. C. McIntyre, M. Savasta, L. Kerkerian-Le Goff, and J. L. Vitek, “Uncovering the mechanism (s) of action of deep brain stimulation: activation, inhibition, or both,” *Clinical neurophysiology*, vol. 115, no. 6, pp. 1239–1248, 2004.
- [30] L. Rocchi, L. Chiari, and F. Horak, “Effects of deep brain stimulation and levodopa on postural sway in parkinson’s disease,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 73, no. 3, pp. 267–274, 2002.
- [31] M. Safonov and K. Fan, “Special issue: Multivariable stability margin,” *Int. J. Robust & Nonlinear Control*, vol. 7, no. 2, pp. 97–226, 1997.
- [32] K. Zhou and J. C. Doyle, *Essentials of robust control*, vol. 104. Prentice hall Upper Saddle River, NJ, 1998.
- [33] E. Mosca, *Optimal, predictive, and adaptive control*. Prentice-Hall, Inc., 1994.
- [34] K. J. Åström and B. Wittenmark, *Adaptive control*. Courier Corporation, 2013.

- [35] I. D. Landau, R. Lozano, M. M'Saad, and A. Karimi, *Adaptive control: algorithms, analysis and applications*. Springer Science & Business Media, 2011.
- [36] M. Nabi-Abdolyousefi and M. Mesbahi, "Network identification via node knock-out," *IEEE Transactions on Automatic Control*, vol. 57, no. 12, pp. 3214–3219, 2012.
- [37] S. Shahrampour and V. M. Preciado, "Reconstruction of directed networks from consensus dynamics," in *2013 American Control Conference*, pp. 1685–1690, IEEE, 2013.
- [38] A. Dankers, P. M. Van den Hof, X. Bombois, and P. S. Heuberger, "Errors-in-variables identification in dynamic networks by an instrumental variable approach," *IFAC Proceedings Volumes*, vol. 47, no. 3, pp. 2335–2340, 2014.
- [39] J. C. Nacher and T. Akutsu, "Structurally robust control of complex networks," *Physical Review E*, vol. 91, no. 1, p. 012826, 2015.
- [40] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *nature*, vol. 473, no. 7346, pp. 167–173, 2011.
- [41] J. Gao, Y.-Y. Liu, R. M. D'souza, and A.-L. Barabási, "Target control of complex networks," *Nature communications*, vol. 5, no. 1, pp. 1–8, 2014.
- [42] N. Elia and S. K. Mitter, "Stabilization of linear systems with limited information," *IEEE transactions on Automatic Control*, vol. 46, no. 9, pp. 1384–1400, 2001.
- [43] F. Deroo, M. Ulbrich, S. Hirche, and B. D. Anderson, "Distributed controller design for a class of sparse singular systems with privacy constraints," *IFAC Proceedings Volumes*, vol. 46, no. 27, pp. 190–197, 2013.

- [44] H. S. Witsenhausen, “A counterexample in stochastic optimum control,” *SIAM Journal on Control*, vol. 6, no. 1, pp. 131–147, 1968.
- [45] V. D. Blondel and J. N. Tsitsiklis, “A survey of computational complexity results in systems and control,” *Automatica*, vol. 36, no. 9, pp. 1249–1274, 2000.
- [46] S. Tatikonda and S. Mitter, “Control under communication constraints,” *IEEE Transactions on automatic control*, vol. 49, no. 7, pp. 1056–1068, 2004.
- [47] R. D’Andrea and G. E. Dullerud, “Distributed control design for spatially interconnected systems,” *IEEE Transactions on automatic control*, vol. 48, no. 9, pp. 1478–1495, 2003.
- [48] M. R. Jovanovic and B. Bamieh, “Lyapunov-based distributed control of systems on lattices,” *IEEE Transactions on Automatic Control*, vol. 50, no. 4, pp. 422–433, 2005.
- [49] F. Lin, M. Fardad, and M. R. Jovanović, “Design of optimal sparse feedback gains via the alternating direction method of multipliers,” *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2426–2431, 2013.
- [50] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.
- [51] S. L. Lauritzen, *Graphical models*, vol. 17. Clarendon Press, 1996.
- [52] P. Spirtes, C. N. Glymour, R. Scheines, and D. Heckerman, *Causation, prediction, and search*. MIT press, 2000.
- [53] S. Wright, “Correlation and causation,” 1921.

- [54] J. T. Koster, “On the validity of the markov interpretation of path diagrams of gaussian structural equations systems with correlated errors,” *Scandinavian Journal of Statistics*, vol. 26, no. 3, pp. 413–431, 1999.
- [55] R. Diestel, “Graph theory 3rd ed,” *Graduate texts in mathematics*, vol. 173, 2005.
- [56] J. Pearl, *Causality: models, reasoning and inference*, vol. 29. Springer, 2000.
- [57] T. Verma and J. Pearl, “Causal networks: Semantics and expressiveness,” in *Machine intelligence and pattern recognition*, vol. 9, pp. 69–76, Elsevier, 1990.
- [58] D. Koller and N. Friedman, *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [59] G. C. Goodwin and R. L. Payne, *Dynamic system identification: experiment design and data analysis*. Academic press, 1977.
- [60] M. Gevers, A. S. Bazanella, and G. V. da Silva, “A practical method for the consistent identification of a module in a dynamical network,” *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 862–867, 2018.
- [61] S. Shahrampour and V. M. Preciado, “Topology identification of directed dynamical networks via power spectral analysis,” *IEEE Transactions on Automatic Control*, vol. 60, no. 8, pp. 2260–2265, 2014.
- [62] A. Dankers, P. M. Van den Hof, X. Bombois, and P. S. Heuberger, “Identification of dynamic models in complex networks with prediction error methods: Predictor input selection,” *IEEE Transactions on Automatic Control*, vol. 61, no. 4, pp. 937–952, 2015.
- [63] N. Everitt, G. Bottegal, and H. Hjalmarsson, “An empirical bayes approach to identification of modules in dynamic networks,” *Automatica*, vol. 91, pp. 144–151, 2018.

- [64] D. Materassi and M. V. Salapaka, “Signal selection for estimation and identification in networks of dynamic systems: a graphical model approach,” *IEEE Transactions on Automatic Control*, 2019.
- [65] K. R. Ramaswamy, G. Bottegal, and P. M. Van den Hof, “Local module identification in dynamic networks using regularized kernel-based methods,” in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 4713–4718, IEEE, 2018.
- [66] K. R. Ramaswamy and P. M. Vandenhof, “A local direct method for module identification in dynamic networks with correlated noise,” *IEEE Transactions on Automatic Control*, 2020.
- [67] D. Materassi and M. V. Salapaka, “On the problem of reconstructing an unknown topology via locality properties of the wiener filter,” *IEEE transactions on automatic control*, vol. 57, no. 7, pp. 1765–1777, 2012.
- [68] B. M. Sanandaji, T. L. Vincent, and M. B. Wakin, “Exact topology identification of large-scale interconnected dynamical systems from compressive observations,” in *Proceedings of the 2011 American Control Conference*, pp. 649–656, IEEE, 2011.
- [69] D. Hayden, Y. Yuan, and J. Gonçalves, “Network identifiability from intrinsic noise,” *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3717–3728, 2016.
- [70] P. M. Van den Hof, A. Dankers, P. S. Heuberger, and X. Bombois, “Identification of dynamic models in complex networks with prediction error methods—basic methods for consistent module estimates,” *Automatica*, vol. 49, no. 10, pp. 2994–3006, 2013.
- [71] Y. Yuan, G.-B. Stan, S. Warnick, and J. Goncalves, “Robust dynamical network structure reconstruction,” *Automatica*, vol. 47, no. 6, pp. 1230–1235, 2011.

- [72] H. H. Weerts, P. M. Van den Hof, and A. G. Dankers, “Identifiability of linear dynamic networks,” *Automatica*, vol. 89, pp. 247–258, 2018.
- [73] A. Haber and M. Verhaegen, “Subspace identification of large-scale interconnected systems,” *IEEE Transactions on Automatic Control*, vol. 59, no. 10, pp. 2754–2759, 2014.
- [74] A. S. Bazanella, M. Gevers, J. M. Hendrickx, and A. Parraga, “Identifiability of dynamical networks: which nodes need be measured?,” in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 5870–5875, IEEE, 2017.
- [75] S. Jahandari and D. Materassi, “Identification of dynamical strictly causal networks,” in *2018 IEEE Conference on Decision and Control (CDC)*, pp. 4739–4744, IEEE, 2018.
- [76] J. Gonçalves and S. Warnick, “Necessary and sufficient conditions for dynamical structure reconstruction of lti networks,” *IEEE Transactions on Automatic Control*, vol. 53, no. 7, pp. 1670–1674, 2008.
- [77] A. Dankers, P. M. Van den Hof, X. Bombois, and P. S. Heuberger, “Errors-in-variables identification in dynamic networks—consistency results for an instrumental variable approach,” *Automatica*, vol. 62, pp. 39–50, 2015.
- [78] L. R. Ford and D. R. Fulkerson, “Maximal flow through a network,” *Canadian journal of Mathematics*, vol. 8, pp. 399–404, 1956.
- [79] A. V. Goldberg, É. Tardos, and R. Tarjan, “Network flow algorithm,” tech. rep., Cornell University Operations Research and Industrial Engineering, 1989.

- [80] S. Jahandari and D. Materassi, “Sufficient and necessary graphical conditions for miso identification in networks with observational data,” *IEEE Transactions on Automatic Control*, 2021.
- [81] S. Jahandari and D. Materassi, “Optimal observations for identification of a single transfer function in acyclic networks,” in *2021 60th IEEE Conference on Decision and Control (CDC)*, pp. 852–857, IEEE, 2021.
- [82] S. Jahandari and D. Materassi, “Analysis and compensation of asynchronous stock time series,” in *2017 American Control Conference (ACC)*, pp. 1085–1090, IEEE, 2017.
- [83] S. Jahandari and D. Materassi, “Topology identification of dynamical networks via compressive sensing,” *IFAC-PapersOnLine*, vol. 51, no. 15, pp. 575–580, 2018.
- [84] S. Jahandari and D. Materassi, “Optimal selection of observations for identification of multiple modules in dynamic networks,” *IEEE Transactions on Automatic Control*, 2022.
- [85] S. Jahandari and D. Materassi, “How can we be robust against graph uncertainties?”
- [86] S. Jahandari and D. Materassi, “Links between causal inference and system identification in control theory: optimal selection of adjusting variables in dynamic bayesian networks,”
- [87] T. Chu, C. Glymour, and G. Ridgeway, “Search for additive nonlinear time series causal models.,” *Journal of Machine Learning Research*, vol. 9, no. 5, 2008.
- [88] P. Spirtes, C. Glymour, R. Scheines, S. Kauffman, V. Aimale, and F. Wimberly, “Constructing bayesian network models of gene expression networks from microarray data,” 2000.

- [89] W. L. Buntine, "Operations for learning with graphical models," *Journal of artificial intelligence research*, vol. 2, pp. 159–225, 1994.
- [90] A. G. Dankers, P. M. Van den Hof, P. S. Heuberger, and X. Bombois, "Dynamic network structure identification with prediction error methods-basic examples," *IFAC Proceedings Volumes*, vol. 45, no. 16, pp. 876–881, 2012.
- [91] M. Droste and W. Kuich, "Semirings and formal power series," in *Handbook of Weighted Automata*, pp. 3–28, Springer, 2009.
- [92] A. V. Oppenheim, *Discrete-time signal processing*. Pearson Education India, 1999.
- [93] C. E. Shannon, *The theory and design of linear differential equation machines*. 1942.
- [94] T. Shafie, "A multigraph approach to social network analysis.," *Journal of Social Structure*, vol. 16, 2015.
- [95] G. Dantzig and D. R. Fulkerson, "On the max flow min cut theorem of networks," *Linear inequalities and related systems*, vol. 38, pp. 225–231, 2003.
- [96] L. Ljung, "System identification," *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp. 1–19, 1999.
- [97] R. A. Usmani, "Inversion of a tridiagonal jacobi matrix," *Linear Algebra and its Applications*, vol. 212, no. 213, pp. 413–414, 1994.
- [98] J. Tian, A. Paz, and J. Pearl, *Finding minimal d-separators*. Citeseer, 1998.
- [99] S. L. Lauritzen, A. P. Dawid, B. N. Larsen, and H.-G. Leimer, "Independence properties of directed markov fields," *Networks*, vol. 20, no. 5, pp. 491–505, 1990.
- [100] S. Acid and L. M. De Campos, "An algorithm for finding minimum d-separating sets in belief networks," *arXiv preprint arXiv:1302.3549*, 2013.

- [101] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.
- [102] S. Even, *Graph algorithms*. Cambridge University Press, 2011.
- [103] A. Ghouila-Houri, “Caractérisation des matrices totalement unimodulaires,” *Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences (Paris)*, vol. 254, pp. 1192–1194, 1962.
- [104] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear estimation*. No. BOOK, Prentice Hall, 2000.
- [105] C. Meek, “Strong completeness and faithfulness in bayesian networks,” *arXiv preprint arXiv:1302.4973*, 2013.