# Causal discovery limitations in learning molecular interaction networks: An empirical study using linear mechanistic models

*Myrto Krana*

Thesis submitted in partial fulfillment of the requirements for the

*Masters' of Science degree in Computer Science and Engineering*

University of Crete
School of Sciences and Engineering
Computer Science Department
Voutes University Campus, 700 13 Heraklion, Crete, Greece

Thesis Advisor: Prof. *Ioannis Tsamardinos*

This work has been performed at the University of Crete, School of Sciences and Engineering, Computer Science Department.

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 617393, CAUSALPATH - Next Generation Causal Analysis project.

University of Crete
Computer Science Department

**Causal discovery limitations in learning molecular interaction networks: An empirical study using linear mechanistic models**

Thesis submitted by
**Myrto Krana**
in partial fulfillment of the requirements for the
Masters' of Science degree in Computer Science

THESIS APPROVAL

Author: _____
Myrto Krana

Committee approvals: _____
Ioannis Tsamardinos
Professor, Thesis Supervisor

_____
Panagiotis Tsakalides
Professor, Committee Member

_____
Yannis Pantazis
Researcher, Committee Member

Departmental approval: _____
Antonios Argyros
Professor, Director of Graduate Studies

Heraklion, October 2018

# Causal discovery limitations in learning molecular interaction networks: An empirical study using linear mechanistic models


## Abstract

Mechanistic models have been traditionally successful in describing biological systems. Their accuracy, however, depends more on expert knowledge about the structure of causal interactions between system components than the abundance of experimental data. At the same time, several algorithms that can learn causal structures *de novo* from observational and experimental data have been developed over the past decades. Despite the attracted attention, their applicability in learning biological systems has been relatively poor.

In this work we systematically study the effect of violations of causal assumptions to basic structure learning algorithms using mechanistic models of protein signaling networks as testbed biological systems. Because the same network of causal interactions can be described using several mechanistic models we study several combinations of system topologies and model specifications. We calculate the analytical solution of each model at steady-state and juxtapose the solution with the fundamental causal discovery principles. We prove mathematically the conditions under which a causal learning algorithm is guaranteed to discover the structure of the system that a mechanistic model describes. Whenever there is no tractable analytic solution, a simulated one is employed. We show that the structure of interactions estimated using data from mechanistic models under steady-state conditions is, in general, inconsistent with the expected causal structure. Accordingly, we reveal that only under very specific conditions is the discovery of the structure guaranteed using a constrained-based causal discovery algorithm.

# Περιορισμοί αιτιακής ανακάλυψης στην εκμάθηση δικτύων μοριακής αλληλεπίδρασης: Μια εμπειρική μελέτη βασισμένη σε γραμμικά μηχανιστικά μοντέλα

## Περίληψη

Τα μηχανιστικά μοντέλα έχουν χρησιμοποιηθεί με μεγάλη επιτυχία στην περιγραφή των βιολογικών συστημάτων. Η ακρίβειά τους, ωστόσο, εξαρτάται περισσότερο από τις γνώσεις των ειδικών σχετικά με τη δομή των αιτιακών αλληλεπιδράσεων μεταξύ των συνιστωσών ενός συστήματος παρά από την αφθονία των πειραματικών δεδομένων. Ταυτόχρονα, αρκετοί αλγόριθμοι που μπορούν να μάθουν τις αιτιακές δομές *εκ νέου* από παρατηρητικά και πειραματικά δεδομένα αναπτύσσονται τις τελευταίες δεκαετίες. Παρά την ελκυστική προσοχή που συγκεντρώνουν, η εφαρμογή τους στην εκμάθηση βιολογικών συστημάτων είναι σχετικά ανεπαρκής.

Στη παρούσα εργασία μελετάμε συστηματικά την επίδραση των παραβιάσεων των αιτιακών υποθέσεων στους αλγόριθμους μάθησης βασικών δομών χρησιμοποιώντας μηχανιστικά μοντέλα και επιλέγοντας ως υπο μελέτη συστήματα σηματοδοτικά δίκτυα πρωτεϊνών. Επειδή το ίδιο δίκτυο αιτιακών αλληλεπιδράσεων μπορεί να περιγραφεί με τη χρήση αρκετών μηχανιστικών μοντέλων μελετάμε διάφορες τοπολογίες συστημάτων και διαφορετικούς συνδυασμούς στις προδιαγραφές των μοντέλων. Ακόμα, υπολογίζουμε την αναλυτική λύση κάθε μοντέλου σε κατάσταση ισορροπίας και αντιπαραβάλλουμε τη λύση με τις θεμελιώδεις αρχές της αιτιακής ανακάλυψης. Αποδεικνύουμε μαθηματικά τις συνθήκες κάτω από τις οποίες ένας αλγόριθμος αιτιακής εκμάθησης θα ανακαλύψει εγγυημένα τη δομή ενός συστήματος που περιγράφεται από ένα μηχανιστικό μοντέλο. Στις περιπτώσεις που δεν καθίσταται δυνατή η αναλυτική λύση, χρησιμοποιούμε προσομοιώσεις. Δείχνουμε ότι, υπό συνθήκες σταθερής κατάστασης, η δομή των αλληλεπιδράσεων που έχει υπολογιστεί από δεδομένα μηχανιστικών μοντέλων, είναι, γενικά, ασυνεπή με την αναμενόμενη αιτιακή δομή. Συνεπώς, αποκαλύπτουμε ότι μόνο υπό αυστηρές συνθήκες είναι δυνατή η ανακάλυψη της αιτιακής δομής, χρησιμοποιώντας έναν αλγόριθμο αιτιακής ανακάλυψης βασιζόμενο σε περιορισμούς.

## Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή κ. Ιωάννη Τσαμαρδίνο για την εμπιστοσύνη και τη στήριξή του καθ'ολη τη διάρκεια εκπόνησης της μεταπτυχιακής μου εργασίας. Ιδιαίτερη αναφορά θα ήθελα να κάνω στους Διδάκτορες κ. Γεώργιο Παπουτσόγλου και κ. Ιωάννη Πανταζή οι οποίοι με την πολύτιμη βοήθειά και στήριξή τους συνέβαλλαν ουσιαστικά στην ολοκλήρωσή της. Ακόμα, θα ήθελα να απευθύνω τις ευχαριστίες μου στην Διδάκτορα κα. Σοφία Τριανταφύλλου και στον Υποψήφιο Διδάκτορα κ. Γεώργιο Μπορμπουδάκη για τις συζητήσεις σχετικά με τις θεωρητικές πτυχές αυτού του έργου. Τέλος, δε θα μπορούσα να μην επισημάνω την υποστήριξη που μου προσέφερε όλη η ομάδα από το ερευνητικό εργαστήριο MensXMachina.

*στους γονείς μου*

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

The dissection of biological systems into parts has been effective in elucidating and characterizing biological processes for decades. In living organisms, however, these parts do not operate as autonomous entities. They rather organize into broad interaction networks giving rise to emergent properties. Signalling pathways, for example, consist of sets of proteins that convey information from the environment to the cell through a stimuli-initiated cascade of molecular interactions that leads to the execution of a response. Because a single protein may contain several functional subunits it is highly possible to be involved in a number of pathways each of which can trigger a different response. Therefore, although many of the low-level properties of biological processes may have been defined, the characterization of their system-level properties is still an open field of research.

In view of this complexity, the doctrine of systems biology is to explain, through powerful computational models, how biological components work together to carry out intended operations [8]. These models are usually mechanistic, being formulated as sets of coupled differential equations expressing reaction kinetics [1], [7], [22]. Differential equation coupling is principally assumed on the basis of some explicit hypotheses about the causal interactions between system components. The conventional method for determining these interactions and, in sequence, the causal structure, that is a structure that encodes the relation cause-effect on its participating components, of the system involves experimental testing. Two major problems exist in this approach. First, the space of hypothetical structures is so large that it renders its complete experimental exploration infeasible and, second, different experimental conditions may potentially suggest several plausible causal structures. As a result, it takes years of experimentation, extensive literature surveys and expert knowledge in order to posit a valid causal connectivity between the parts of any biological system. A key research topic, therefore, is to reverse engineer the network of causal interactions directly from the data using automated computational procedures.

Reverse engineering of interaction networks using data collected in the biological context of interest has been a subject of fast increasing interest, [23], [9], [5]. Existing approaches applied for this purpose can be loosely separated into competing-model approaches and *de novo*-learning approaches. Competing-model approaches consider a set of preconceived, deterministic or stochastic, equation models of different plausible structures that are ranked objectively using experimental data under a Bayesian inference framework [30], [10], [14], [29]. Although conceptually simple, the feasibility of these models relies heavily on the existence of prior knowledge about the connectivity of system components and on the dimensionality of the model parameter space. In contrast, approaches of the second category, employ statistical learning formalisms in order to reconstruct the structure directly from data. These involve graph models such as Causal Bayesian Networks (CBNs) or differential equation models [20], [28], [18], [15].

Out of all these approaches graph models have, perhaps, the most prominent role. Still, their success is only guaranteed when the data satisfy a set of rigid assumptions [17], [25] that are almost certainly violated in the biological setting [5], [28]. This is due to many, perhaps fundamental, limitations some of which concern: the intrinsic variability of molecular networks related to the stochastic change of system component concentrations [3]; our inability to observe biological systems with no measurement error [31], [21] and our inability to make precise interventions [11]. Despite not included in our study it is essential to note that among the previous limitations, biological systems are not always measured at the equilibrium state and the relationships between biological entities are usually non-linear.

Because of these limitations, our goal, in this work, is to study the reasons why causal assumptions are violated in the biological setting and result in the failure of graph models to learn molecular network structures. To this end, we conduct an empirical study where we investigate the capacity of CBNs to learn the structure of signaling pathway models that we generate in the form of differential equations. For simplicity, we consider only linear differential equation systems, and leave the extension to nonlinear systems as future work. As minimal examples of network structures, we employ three common topologies. To reproduce the limitations encountered in the biological setting we assume the existence or absence of intrinsic stochasticity, measurement noise and combinations thereof in the equation system of each pathway. Because Graph models encode the Conditional Independences (CIs) between variables in the data into graphs, we find, for each pathway topology, the CIs between system components from their distribution at steady-state. Finally, we assess the range of causal assumptions violation from comparing the structure recovered from the CIs at steady-state with the structure of the underlying pathway topology. Our results designate that only under certain conditions is the reconstruction of the underlying network guaranteed.

The present work is structured as follows. In the next Section 1.1 we propose the related work to the present one while in Section 1.2 we define our motivation. In Chapter 2, we first, give an overview of mechanistic model formalisms in terms

of linear differential equations. Then, we briefly sketch the theory behind CBNs and describe the common comparison space between the two modeling approaches. In Chapter 3, we illustrate our analysis approach while, in Chapter 4, we present the results along with a discussion on key findings. Also, in Chapter 5 we mathematically prove a property propagation in reaction networks and we end up in 6 drawing the conclusion of this work.

## 1.1 Related Work

Different, but related, approaches have been developed for causal inference and modeling. These approaches are based on different starting points. One approach starts by using causal Bayesian networks to link the joint distribution of the variables to causal connections. Another approach employs a structural causal model (or structural equation model) where each effect is expressed as a function of its direct causes and an unmeasured noise variable. The third approach uses the causal Markov condition and links observed independences to the causal graph [13].

Authors in [26], based on the first approach, made a theoretical study on the applicability of the CBNs in microarray data and recorded all the difficulties (averaging, sample size, measurement error etc.) of reliable data analysis. They stressed that if sample size of microarray data could be increased with low measurement noise, with each sample either gathered from a single cell or from a collection of cells (averaging) with low variance it would improve the performance of CBN algorithms. In [20], eliminating population-averaging effects and applying CBNs approach, they have been the only ones who almost reconstruct the basic structure of a protein signaling pathway network of real single-cell data, with no a prior knowledge of pathway connectivity. Nevertheless, we have to note that several factors aided this success including absence of latent confounders and mixture of observations and interventions. On the other hand, in [28] authors examined the performance of two state-of-the-art causal discovery methods (CLCD and BACKSHIFT) on single-cell real data and studied their response to several perturbations. They observed that the methods often disagree with each other and with the background knowledge and found out that the results shown to be reproducible in independent data sets showing the consistency between statistical patterns with certain causal models across different studies. It is important to point out that they used multiple data sets, including data of [20], without suffering from population averaging and/or different experimental conditions.

Out of the CBN approach, in [13] it is introduced, for first time, an alternative interpretation of structural causal models arises naturally when considering systems of Ordinary differential equations (ODEs). By considering how these differential equations behave in an equilibrium state, under certain stability assumptions on an ODE, authors arrive at a structural causal model (SCM) that is time independent, yet where the causal semantics pertaining to interventions is still valid. Also, the paper addresses that, although information is lost when

going from a dynamical system representation to an equilibrium representation, if equilibrium data are available, the SCM representation is preferable than the complicated dynamical system one.

In their next work [2] authors presented the Random differential equations (RDEs) that are the most natural extension of ordinary differential equations to the stochastic setting and have been widely accepted as an important mathematical tool in modeling and analysis of various processes. Since uncertainty and random fluctuations are a very common feature of real dynamical systems, they added disturbances (randomness in the initial condition or random coefficients) in ODEs and formulate RDEs. These disturbances are not only of stochastic nature, but they are also of causal nature in the sense that the disturbance processes are affecting other processes of the system. They show that when time tends to infinity the RDEs converge exactly to the structural equations of the SCM and that this construction is compatible with interventions under similar convergence assumptions. Similarly as in [13], they supported that SCMs representation is more compact for learning and prediction purposes of random systems that have reached equilibrium while when the system deals with confounders, there is no need to model their dynamics, but only their equilibrium distribution.

Extension of RDEs (disturbances such as Gaussian white noise) is the Stochastic differential equations (SDEs). In [18], they implemented a method for causal prediction, called INDUCE, that learns plausible connectivities from the covariance matrix of a steady-state SDE system. They applied their method to small networks on real data and were able to reconstruct a two-node network model.

Another persperctive, in [4], notes that the general stochastic process framework incorporates a major feature of causal relationship which is time. In [24] they gave a causal interpretation of SDEs. Their work focus more on postintervention SDE, resulting from an intervention in an SDE, which contrasts with the classical DAG-based case. Furthermore, they remarked that SDE reflects a data-generating mechanism in which the variables at a given timepoint are obtained as a function of the previous timepoints and discussed on a recent framework for causality based on the concept of conditional local independence.

Local independence has been suggested as a useful independence concept for stochastic dynamical systems. However, there is no well-developed theoretical framework for causal learning based on this notion of independence. [12] clarified that the dynamic evolution of a stochastic system cannot be modeled using SCM related to finite number of variables, and the infinitesimal conditional independence relationships between processes cannot be efficiently represented by the standard probabilistic independence. The paper defines the concept of local independence via asymmetry relation (the present of one process may be independent of the past of another process, or the reverse, or both) and tries to build a framework by abstractly proving the equivalence of pairwise and global Markov properties and by giving an oracle learning algorithm.

## 1.2 Motivation

In Section 1.1 we see that the applicability of CBN approach has been mainly doubted by the argument of averaging in cell populations, suggesting that the obstacles are not fundamental [26]. A reconstruction of the basic structure of a protein signaling pathway in real data [20], free from population averaging, was a break though in causal discovery community and strengthened the advantages over the CBN framework. Since then, CBNs have been used in a plethora of papers in order to learn biological systems. However, their applicability could not be identified as satisfactory.

It consequently rises the question why there is no continuity of the first success of the CBNs. Inspired by this question, we came up with the idea to simulate simple cases of molecular interaction systems (including several limitations of biological systems) using linear mechanistic models, that effectively capture the microcosm, so as to explore their background with mathematical formalism. Beyond the mathematical modeling, we generate observational single-cell data in equilibrium, without interventions or suffering from population averaging and we try to assess when and why the widely used CBN method succeed or fail to learn their structure.

# Chapter 2

# Methods

We divide this Chapter into three parts. In Section 2.1, we define the mathematical background required for modeling biosystems using mechanistic analogues. Traditionally, biosystems have been described with ordinary differential equations (ODEs) [29]. However, it has been widely acknowledged in the recent years that molecular kinetics are intrinsically stochastic and Stochastic Differential Equation (SDEs) models are potentially better suited for capturing a biosystem's behavior in a realistic way [29]. Consequently, we begin our analysis by citing the mathematical formalisms for both linear ODE and SDE models as derived from simple mass-action kinetics [27]. Next, in Section 2.2, we describe the theory behind graph-based reconstruction methods. Because the produced graphs imply exactly the conditional independences between variables in the data, we show, in Section 2.3, that independence assertions can be the basis for comparing the structure of mechanistic models, which we assume to be the ground truth, with the reconstructed structure. We close this Chapter by presenting in Section 2.4 the ODE and SDE solution at equilibrium.

Our analysis also deserves a special note in that the theory of reaction kinetics we employ below is generally applicable to any type of molecular interaction networks, either gene regulatory networks or signaling and metabolic pathways or others. By the same reasoning, our approach is not necessarily limited to reaction kinetics but can be employed on all network systems that can be described by the modeling formalism we show next.

## 2.1 Mechanistic Modeling of Biosystems

In general, a model is a description of a system designed to help an observer to understand how it works and to predict its behavior. Models are typically conceptual, existing as an idea, a computer program or a set of mathematical formulas.

Mechanistic models, however, have been successfully used to provide explanations and promote our understanding of complex biological systems. To conceive such model typically requires specifying the molecular components that are causally relevant to the system and how these are organized so that they compose the behavior in question. The structure of mechanistic models, henceforth called *mechanistic structure*, stems from the activities that each molecular component is engaged in. These activities usually involve biochemical reactions that affect other system components by changing their production or consumption rate. In other words, a mechanistic structure prescribes the landscape of chemical interactions that take place between system components and causally affect their molecular quantities.

Accordinlgy, mechanistic structures are usually depicted as *reaction diagrams* [6]. As seen in Fig. 2.1 (a)-(c), these diagrams consist of nodes and edges where each node corresponds to a molecular component and the edges express the chemical interaction that causally affects a pair of components. An edge from component $V_1$ to component $V_2$ denotes that $V_1$ causally affects the production of $V_2$ at a rate that is defined by the kinetic parameter (usually noted by the letter $k$) on top of that edge, while $V_1$ is degradated with the same rate. In mathematical terms, an edge of the reaction diagram corresponds to a rate equation derived from some kinetic law (e.g. mass-action kinetics, Michaelis-Menten kinetics, etc.). We have to clarify that a rate equation is an equation that links the reaction rate with the concentrations of the reactants and the constant parameters (e.g rate = $k_2$ $[R]$ in Fig 2.1(a)). On the other hand, the rate of change in the amount of a molecular component is related to several edges (including the $X_0$ element) and can be described as the sum of the incoming and outgoing edges that the diagram illustrates. Finally, depending on whether the state of the system is designed to change deterministically or stochastically, the corresponding mathematical formalism is assumed i.e. by using ODEs or SDEs.

In what follows, the rate equations we assume are described by simple mass action kinetics whereby the reaction rate is proportional to the amount of the system component undergoing the reaction. In addition, we restrict the analysis to linear reaction networks [1],[29]. The advantage of using a linear model is that the steady-state analytical solutions exist for most of the studied models making it straightforward to obtain theoretical results. We leave the study of non-linear, out-of-equilibrium reaction networks as future work.

### 2.1.1 Linear ODEs

Ordinary Differential Equations (ODEs) are deterministic models that describe the evolution of dynamically-changing phenomena. A first-order linear ODE system with constant coefficients is the simplest mechanistic model. Its general form is [29]:

$$\frac{dS(t)}{dt} = \mu_0 - AS(t) \ , \quad S(0) = S_0 \ , \tag{2.1}$$

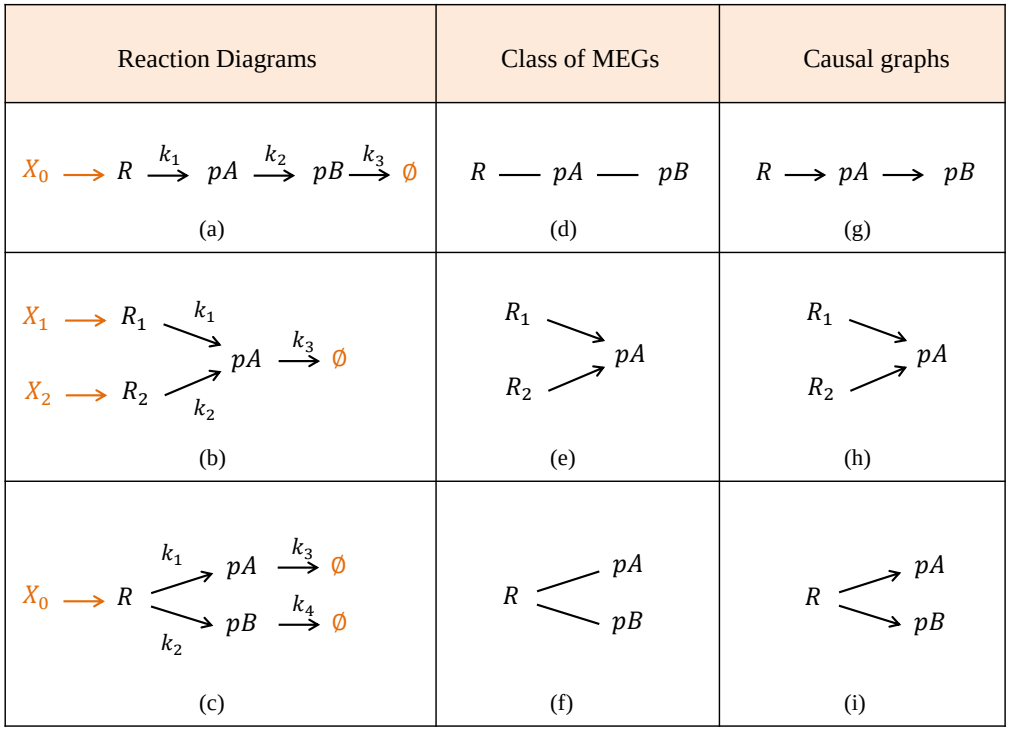



Figure 2.1: Minimal examples of reaction diagrams (a-c), respective class of MEGs (d-f), and, respective causal graphs (g-i) between three signaling proteins: $R$, $pA$ and $pB$. In (a-c) signaling cascade topologies are presented. These are some of the elementary building blocks of any reaction network in all biological systems. Edges refer to chemical reactions between respective protein pairs. Edge orientation denotes the causal influence i.e. the direction whereby in a pair of proteins, one is degradated at a rate k while the other is produced at the same rate. With black color we denote the measured quantities and with orange color the unmeasured ones. $X_0$ (or $X_1$, $X_2$) is the initial condition of reaction network while the empty set ($\varnothing$) the end of it. In (d-f) the respective class of MEGs are laid while in (g-i) the respective cascade-structure, v-structure, and $\Lambda$-structure causal graphs are shown. Edges denote the concept of probabilistic cause and effect. For more details see examples in text.

where $S(t) \in \mathbb{R}^n$ is the vector of time-varying system components and $S_0$ a constant vector of their corresponding initial values. The connectivity matrix, denoted $A$, is a $n \times n$ matrix that represents the true causal relationships between system components while $\mu_0$ is a vector that corresponds to the external forces e.g., it may represent the constant or the random inflow of stimulus in a signaling network.

As an example, consider a fictional molecular species that stimulates the following artificial system of a linear signaling cascade of three fictional proteins, $R$:

the receptor protein that is stimulated by a constant external force, $pA$: the protein that is activated, in turn, by $R$ at a rate $k_1$ and $pB$ that is activated by $pA$ at a rate $k_2$ and independently degraded at a constant rate $k_3$. Let also that, initially, the quantity of the stimulus be $X_0$ and no activated protein molecules exist in the system. Furthermore, we assume that the amount of cause entity is consumed equal to the amount of the effect entity production for consistency between the reaction diagram and the corresponding causal graph. The reaction diagram of this model is shown in Fig. 2.1 (a) and in matrix notation the respective first-order mass-action kinetics are written as:

$$\frac{d}{dt}\begin{bmatrix} R \\ p_A \\ p_B \end{bmatrix} = \begin{bmatrix} X_0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} k_1 & 0 & 0 \\ -k_1 & k_2 & 0 \\ 0 & -k_2 & k_3 \end{bmatrix} \begin{bmatrix} R \\ p_A \\ p_B \end{bmatrix}, \tag{2.2}$$

where $S = [R\ pA\ pB]^T$ represents a vector of quantities of system components. The initial conditions here are zero i.e. $S_0 = [X_0\ 0\ 0]^T$. In Fig. 2.2 we present how the above mentioned quantities change over time based on eq.(2.2).

### 2.1.2 Linear SDEs

SDEs are extensions to ODEs in that intrinsic stochasticity is added to the evolution of the process. Following eq.(2.1) the general form of a linear SDE with additive stochasticity is:

$$dS(t) = (\mu_0 - AS(t))dt + \sigma dW(t) \tag{2.3}$$

where $S$, $A$ and $\mu_0$ have the same meaning as in the case of ODEs, $\sigma > 0$ is the strength or average magnitude of the random fluctuations and $W(t)$ is the standard multidimensional Brownian motion in $\mathbb{R}^n$. A description for Brownian motion is a continuous stochastic process which takes the initial value $W(0) = 0$ with probability one and it is such that the increment $[W(t) - W(s)]$ from $W(s)$ to $W(t)$ $(t > s)$ is an independent Gaussian process with mean value zero and variance $(t - s)$ and its time derivative has the characteristics of white noise [16].

Compared to (2.1), eq.(2.3) differs in the additional stochastic term which manifests the degree of apparent uncertainty of the system. Accordingly, in the preceding example, the difference in the matrix notations between ODEs and SDEs will be the addition of the vector $\sigma \cdot [w_1, w_2, w_3]^T$ in eq.(2.2), where each $w$ denotes independent white noise. Particularly,

$$d\begin{bmatrix} R \\ p_A \\ p_B \end{bmatrix} = \left( \begin{bmatrix} X_0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} k_1 & 0 & 0 \\ -k_1 & k_2 & 0 \\ 0 & -k_2 & k_3 \end{bmatrix} \begin{bmatrix} R \\ p_A \\ p_B \end{bmatrix} \right) dt + \sigma \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}, \tag{2.4}$$

Because the connectivity matrix is not influenced by the additional stochasticity, the reaction diagram will remain the same as previously. This is expected

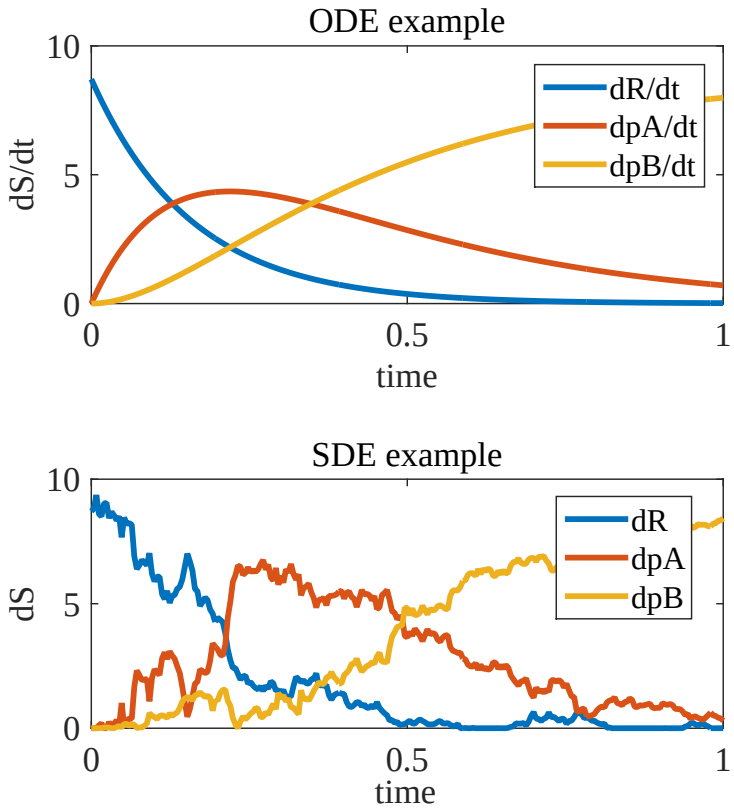



Figure 2.2: ODE and SDE example of signaling cascade-structure of three fictional proteins, as seen in Fig. 2.1(a), with constant initial condition. Up plot, the solution of eq.(2.2) with the receptor protein $R$ is degradated over time, with $pA$ protein initially increased and then degradated, and with $pB$ being steadily increasing over time. Down plot, solution of eq.(2.4), similar to ODE solution, with apparent effect of Brownian motion.

since both mechanistic modeling approaches describe the same chemical interactions between signaling components. Fig. 2.2 show the change of the above mentioned quantities over time based on eq.(2.4) and the effect of stochasticity in the system.

## 2.2 Causal Network Reconstruction

The induction of the causal connectivity of a biological system's components without performing extensive experimental testing is profoundly challenging. That is why a key goal for computational biology is to devise methods able to perform such a task directly from data. To this end, Causal Bayesian Networks (CBNs) provide

the appropriate framework for building models connecting observable properties of the joint distribution of system components to their underlying causal connections.

In CBNs, the interactions among system components are described with a directed acyclic graph (DAG) (see Figs. 2.1(g-i)) [17]. Similar to the reaction diagrams mentioned earlier, a causal graph, in its simplest form, consists of a set of nodes and a set of directed edges connecting some pairs of nodes. Again, a node denotes a system component. However, here, a directed edge from one component to another represents a *direct* (no component in the system mediates the relationship), *probabilistic* (a change in the distribution of the cause will result in a different distribution of the effect) *causal relationship* between the corresponding pair. The DAG is then connected to the joint probability distribution of the system components via the *Causal Markov Condition (CMC)*:
*Let $G$ be a causal graph with vertex set $V$ and $P$ be a probability distribution over the vertices in $V$ generated by the causal structure represented by $G$. $G$ and $P$ satisfy the Causal Markov Condition if and only if for every component $V_i$ in $V$, $V_i$ is independent of $V - \{(Descendants(V_i) \cup Parents(V_i))|Parents(V_i)\}$* [25].

Knowing the causal relationship between two system components allows the prediction of their response to an external intervention. For example, if component $V_1$ causally affects the amount of component $V_2$, then the perturbation of $V_1$ is expected to affect the levels of $V_2$, while the perturbation of $V_2$ will have no impact on $V_1$.

CBN reconstruction, also known as Causal Discovery, belongs to a special class of statistical analysis that identifies causal networks models, *de novo*, from observational data. To learn the structure of a causal model, henceforth called *causal structure*, commonly requires treating system components as a set of random variables $V = \{V_1, ..., V_n\}$ and, using appropriate tests of statistical independence, to detect all conditional independences (CIs) that hold over any subset of variables in $V$. We define that two events $R$ and $pB$ are conditionally independent given a third event $pA$ precisely if the occurrence of $R$ and the occurrence of $pB$ are independent events in their conditional probability distribution given $pA$ (see Fig. 2.1(g)). One can then design algorithms that attempt to identify causal structures that entail the observed CIs according to the *CMC*.

To do so efficiently, most algorithms also rely on the *faithfulness assumption*:
*Let $G$ be a causal graph and $P$ a probability distribution generated by $G$. $\langle G, P\rangle$ satisfies the Faithfulness Condition if and only if every conditional independence relation true in $P$ is entailed by the Causal Markov Condition applied to $G$.* [25]
In other words, all observed CIs in the data must be a product of the underlying causal structure, rather than being "accidental" or in virtue of specific model parameters [17]. Thus, for a given causal graph, CMC defines a unique set of CIs. A distribution that abides all and only these CIs is called faithful to the graph.

Under CMC and faithfulness, the graphical criterion of *d-separation* can be used on a causal graph to identify the CIs that hold in any faithful distribution. We say that *for a graph $G$, if $V_1$ and $V_2$ are vertices in $G$, $V_1 \neq V_2$, and $V$ is a set of vertices in $G$ not containing $V_1$ or $V_2$, then $V_1$ and $V_2$ are d-separated given*

*V in G if and only if there exists no undirected path U between $V_1$ and $V_2$, such that (i) every collider*[1] *on U has a descendent in V and (ii) no other vertex on U is in V* [17].
A schematic example of this is shown in the graph of Fig. 2.1(g) where, according to causal nomenclature it indicates that $R$ and $pB$ are *d-separated* given $pA$. That is, given the way information flows from $R$ to $pB$, providing knowledge for $R$ gives no further insight about the quantity of $pB$ if we already know $pA$. On the other hand, in the graph of Fig. 2.1(h) $R_1$ and $R_2$ are *not d-separated* given $pA$ since information about $R_1$ flows to $R_2$ through $pA$.

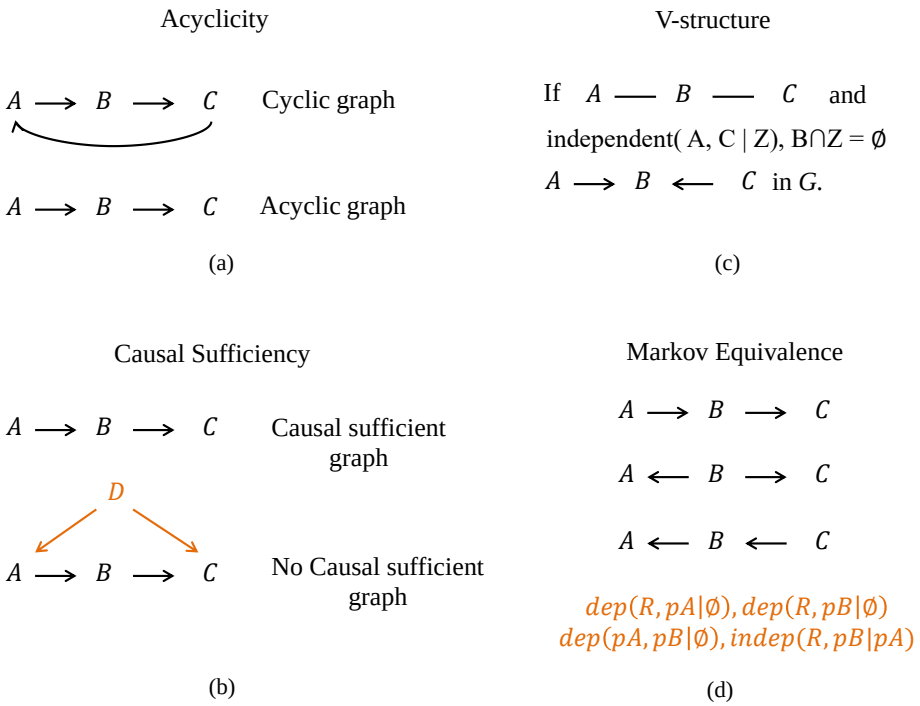



Figure 2.3: Graph explanation of (a) Acyclity. Up, graph with feedback look creating a cycle and down acyclic graph (b) Causal sufficiency. With black color we represent the measured variables while with orange the unmeasured ones. Up, all variables are measured and causal sufficiency is satisfied while down, $D$ is an unmeasured variable which affects two measured quantities $A$ and $C$. Then, causal sufficiency is violated. (c) Explanation of V-structure. (d) Markov Equivalent networks. They have the same edges, the same v-structures (here there are no v-structures) and the same conditional independences. We denote dependency using "*dep*" and independency using "*indep*".

[1] *In graph G a vertex V is a collider on undirected path U if and only if there are two distinct edges on U containing V as an endpoint and both are into V (see Fig. 2.1(h)).*

Another assumptions in CBN reconstruction are *acyclicity* (no feedback loops- see Fig. 2.3(a)) and *causal sufficiency*:
*A set $V$ of variables is causally sufficient for a population if and only if in the population every common cause of any two or more variables in $V$ is in $V$ or has the same value for all units in the population* [25].
By a way of explanation, the unmeasured variable that is a common cause to two or more variables in $V$ violates causal sufficiency. In Fig. 2.3(b) we present an example of causal sufficiency in a graph.

### 2.2.1 PC algorithm

Constraint-based algorithms for causal discovery try to reverse-engineer the causal graph from a data set of observations. They test the data for statistical independences and try to identify the causal graph that entails (only) them according to the CMC and faithfulness. Additionally, causal interpretation of the returned models requires the assumptions of *acyclicity* and *causal sufficiency*.

Identifying the optimal causal model that the data support from CI tests is very challenging, because the number of statistical tests required grows exponentially with the number of variables [25]. Moreover, the problem does not admit a unique solution: A set of graphs (called Markov equivalent) entail the same CIs. However, Markov equivalent networks share the same skeleton and some orientations (i.e. causal relationships). In Fig. 2.3(d) there is an example of Markov Equivalent networks.

PC is probably the most widely employed algorithm that schedules the tests in a way such that the whole space of models is efficiently explored [26]. The algorithm starts by forming the complete undirected graph, then reduce that graph by removing edges with zero order conditional independence relations, reduce again with first order conditional independence relations, and so on. Then, it learns v-structures (see Fig. 2.3(c)) and perform all orientations entailed by acyclicity and the set of v-structures found. To be specific, PC algorithm identifies a graph that summarizes the features (oriented edges) of all Markov-equivalent graphs that can encode the CI relationships that hold in the data. The algorithm heavily employs theoretical results on properties of Markov equivalent graphs and properties of CI to limit the number of statistical tests required for soundness and completeness in the sample limit. The resulting graph, which we will call Markov-Equivalent Graph (MEG), includes both directed and undirected edges. If an edge is directed, then all graphs in the class agree on the orientation of that edge, while, if the edge is undirected, there are at least two graphs that disagree.

## 2.3 Common Comparison Space

Our goal in this study is to employ toy models of molecular interaction systems where the underlying topology is known in advance and try to assess the reasons for which causal network reconstruction methods fail to learn their structure. To

this end, we already discussed in Section 2.1 that reaction diagrams correspond to mechanistic structures that integrate the chemical interactions that causally affect system components. This means that mechanistic structures can be considered as causal structures and, accordingly, reaction diagrams as informal causal graphs. Clearly though the semantics of reaction diagrams and causal graphs are incompatible. That is, a node in a reaction diagram represents a protein while the same node in a causal graph depicts the cause\effect.

To bridge this gap, we make some assumptions for our system. We consider that the system has reached a steady-state of equilibrium and our data are samples from that state. At steady-state a reaction network and its related dynamical system, in the form of either ODEs or SDEs, can be associated with a probability density function that directly corresponds to the measured data. From this density function the CIs between the system components can be calculated. As just discussed, the PC algorithm employs CI assertions to form a MEG. Hence, we propose that the class of MEGs to be the common space where both reaction diagrams and causal graphs are projected and compared.

## 2.4 Steady-State solution of ODE and SDE systems

In Section 2.1 we introduced the general form of linear ODE and SDE systems. In Section 2.3 we saw that equilibrium is the state that we examine our system. Thus, in this Section we present the steady-state solution of ODE and SDE systems. We divide this Section into two parts. In the first part we present the steady-state solution of an ODE system while in the second one the steady-state solution of an SDE system. In both systems we take into consideration two options for the initial condition. One is the constant case-consider in Fig. 2.1 $X_0$ has a constant value − and the other is the random case − consider in Fig. 2.1 $X_0$ is a random variable $\sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$.

***Steady-State***

*A system or a process is at* steady-state *if some properties that define the behavior of the system or the process remain unchanged over time.*

### 2.4.1 ODE system at *s.s*

ODEs in equilibrium means that the variables of the differential equation are constant with respect to time. Then, the steady-state solution ($S_{ss}$) for eq.(2.1), satisfies [29]

$$\frac{dS_{ss}}{dt} = 0 \implies AS_{ss} = \mu_0 \implies S_{ss} = A^{-1}\mu_0,$$

i.e. $S_{ss}$ is completely determined by the connectivity matrix $A$, which is invertible since it is a square matrix and its determinant is different from zero, and the

initial condition $\mu_0$. Either $\mu_0$ is constant or random the solution of an ODE system remains the same.

### 2.4.2 SDE system at *s.s*

In SDEs, the process does not rest at a unique equilibrium point rather it approaches a probability distribution that is time invariant. A formal solution of eq. (2.3) can be written as [16]:

$$S(t) = e^{-At}\mu_0 + \int_0^t e^{-A(t-a)}\sigma dW(a) \tag{2.5}$$

Using eq. (2.5) we can write the autocorrelation[2] matrix of the process S(t) is

$$C(t,a) = e^{-At}\left(C_0 + \int_0^{min(t,a)} e^{Ap}\Sigma e^{A^T p}dp\right)e^{-A^T a} \tag{2.6}$$

where $C_0$ the covariance matrix of the initial condition $\mu_0$.
Furthermore, the variance at time t, $\Sigma := C(t,t)$ , satisfies the differential equation

$$\frac{d\Sigma(t)}{dt} = -A\Sigma(t) - \Sigma(t)A^T + \sigma^2 I \tag{2.7}$$

At Steady State the distribution is $p(S) \sim N(\mu_S, \Sigma_{Lyap})$ with $\mu_S = A^{-1}\mu_0$ and $\Sigma_{Lyap}$ satisfying

$$A\Sigma_{Lyap} + \Sigma_{Lyap}A^T = \sigma^2 I \tag{2.8}$$

where $\sigma \in \mathbb{R}$ and $\Sigma_{Lyap}$ the steady state variance.
Eq. (5.5) is an example of Lyapunov equation which is widely used in control theory for the stability analysis of discrete-time systems. To calculate $\Sigma_{Lyap}$ we augment the Lyapunov equation in $n \times n$ system with the symmetry constraints for the covariance matrix (i.e., $\Sigma_{ij} = \Sigma_{ji}$). Therefore, we have

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad \text{and} \qquad \Sigma_{Lyap} = \left[\ \sigma_1 \ \middle|\ \sigma_2 \ \middle|\ \cdots \ \middle|\ \sigma_n\ \right]$$

where $a_i$ is a row vector in the $i_{th}$ row of $A$ and $\sigma_i$ is a column vector in the $i_{th}$ column of $\Sigma_{Lyap}$. The $A$ and $\Sigma_{Lyap}$ matrices consist of $n \times n$ elements or measured variables. The solution is given by:

$$\left[\ \sigma_1 \quad \sigma_2 \quad \cdots \quad \sigma_n\ \right]^T = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}^{-1} \left[\ \underbrace{\sigma^2 \quad \cdots \quad \sigma^2}_{n} \quad \underbrace{0 \quad \cdots \quad 0}_{n(n-1)}\ \right]^T \tag{2.9}$$

[2]The autocorrelation of a random process is the Pearson correlation between values of the process at different times, as a function of the two times or of the time lag.

where

$$A_1 = \begin{bmatrix} 2a_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 2a_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2a_3 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & 2a_n \\ a_2 & a_1 & 0 & 0 & \cdots & 0 \\ a_3 & 0 & a_1 & 0 & \cdots & 0 \\ 0 & a_3 & a_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & a_n & a_{n-1} \end{bmatrix}$$

is a $(\frac{n(n+1)}{2} \times n^2)$ matrix that gives constraint equations stem from Lyapunov equation and

$$A_2 = \begin{bmatrix} 0 & 1 & 0 & \cdots & -1 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & -1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & -1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

is a $(\frac{n(n-1)}{2} \times n^2)$ matrix that gives constraints stem from the symmetry of $\Sigma_{Lyap}$ ($\sigma_{ij} = \sigma_{ji}$, where i, j denote the rows and the columns in $\Sigma_{Lyap}$, respectively). The solution of Lyapunov equation, as it is described earlier, can be applied in any case, for any topology with different quantities and reactions as far as it remains linear.

However, based on the initial condition of the system the SDE solution will have same changes. Let $p(S) = \mathcal{N}(\mu_S, \Sigma)$ the probability distribution of the system components S at steady state with

$$\mu_S = \mathbb{E}(A^{-1}\mu_0) = A^{-1}\mathbb{E}(\mu_0)$$

where $A^{-1}$ is the inverse of $n$ $x$ $n$ connectivity matrix and $\mu_0$ a vector that correspond to the initial condition.

In the **constant case** the value of $\mu_0$ is fixed and equal to $X_0$. Thus, $\mu_S = A^{-1}[X_0 \cdots X_0]^T$ and $\Sigma = \Sigma_{Lyapunov}$ . On the other hand, when we add uncertainty on $\mu_0$ it means that $\mu_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0})$ and therefore

$$\mathbb{E}(\mu_0) = \begin{bmatrix} \mu_{X_0} & 0 & 0 \end{bmatrix}^T$$

and

$$cov(\mu_0) = \begin{bmatrix} \sigma^2_{X_0} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Also, based on the covariance of linear transformation [3] it holds that

$$\Sigma = cov(A^{-1}\mu_0) + \Sigma_{Lyap} = A^{-1}cov(\mu_0)(A^{-1})^T + \Sigma_{Lyap}$$

To sum up for the **random case** it holds

$$p(S) = \mathcal{N}(A^{-1}\mathbb{E}(\mu_0), cov(A^{-1}\mu_0) + \Sigma_{Lyap}).$$

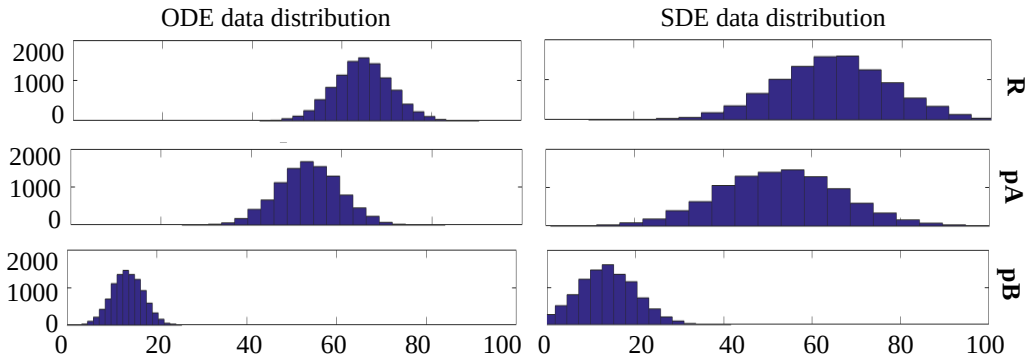



Figure 2.4: ODE and SDE data distribution at steady-state when there is measurement noise in the system. Consider that the system is similar to that of Fig. 2.2 by adding measurement noise.

We studied above the equilibrium solution of the ODE and SDE systems. Nevertheless, the question remains "how steady-state is the common comparison space between mechanistic models and causal network reconstruction". The answer is given by the steady-state distribution (see Fig. 2.4). From the steady-state distribution we are able to find all associated independence assertions that the data support. Particularly, the covariance matrix $\Sigma$ encodes all CIs given the empty set while its inverse, $\Theta = \Sigma^{-1}$, encodes all pairwise independences given all the other variables of the system. Matrix $\Theta$ is also known as the *concentration* matrix and since we restricted ourselves to systems with three variables, $\Theta$ encodes all other CIs associated with the steady-state distribution. Therefore we can construct the MEG that corresponds to the steady-state data using the information from $\Sigma$ and $\Theta$ supplemented with the causal discovery rules.

Fig. 2.5 shows how from mechanistic models we can recover the causal structure of a system. Explicitly, in ODE level we generate observational data and we validate them using PC algorithm. The same approach is applied on the SDE level, with the difference being that in some cases we can theoretically investigate $\Sigma$ and $\Theta$ matrices and consequently extract de-novo the underlying MEG.

[3] $cov(Ax) = A^{-1}cov(x)A^T$

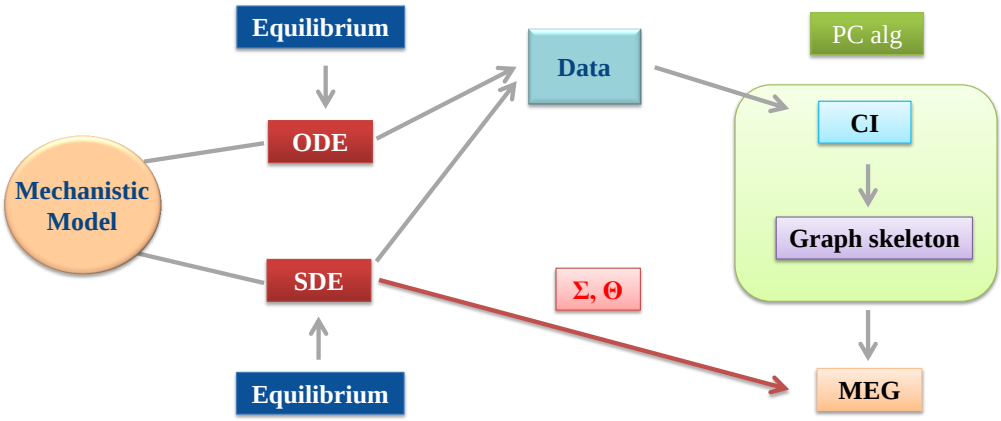



Figure 2.5: Common comparison space between mechanistic models and CBNs. Data are generated from the ODE and SDE steady-state solution and tested on PC algorithm which in turn returns a MEG. However, from the SDE steady-state solution, in some cases, it is tractable to derived find the MEG based on information given by both covariance $\Sigma$ and concentration $\Theta$ matrices. For more please see Section 2.4.

# Chapter 3

# Study Setup

We study three reaction network topologies that constitute some of the basic building blocks of any biological system. These topologies are designed to be acyclic so that the corresponding CBNs are well defined. We use a factorial design of virtual experiments to systematically test several combinations of model formalisms, model specifications and parameter settings. Our goal is to determine when the true causal graph, as defined by a CBN, can be recovered from data produced by each reaction network. For this, we make comparisons both at the level of probability density functions as they are quantified through a MEG and at the performance level of causal discovery algorithms. More specifically (see Fig 2.5), we compare between the MEG derived from the steady-state solution of mechanistic models and those that were reconstructed from the PC algorithm applied on data that are generated from simulations of the same mechanistic models. It is important to clarify that, to be consistent with the assumptions of causal modelling and discovery, we solely study observational data sampled from the steady-state distribution (all quantities are measured in equilibrium).

## 3.1 Network Topologies

We study three distinct signaling pathway topologies that comprise of three interacting molecular components. We restrict our presentation to triplets with two edges because they are the minimal examples where causal structure inferences can be made based on both conditional and unconditional independences. Three−node topologies with zero or one edge are subgraphs of the topologies that we study and they can be classified at those they are examined with unconditional independence tests. We leave the study of triplet topologies with three edges as a future work.

The first topology denotes a cascade as found in all signaling pathways. The second is a merging structure, that in causality nomenclature is called v-structure, where two components, together, produce (or activate) a third one. The third topology is a branching structure, the reversed of a merging structure, where one

component produces (or activates) two other components. In causality nomenclature this latter structure is called Λ-shaped or fork structure. Figure 2.1 illustrates the three structures.

Under the assumptions of causal discovery all conditional dependencies and independences of the structures are encoded by MEGs. In Figure 2.1(d and f), the cascade and the fork structure both entail the conditional independence of the endpoint given the mediator (i.e. $R$ is independent of $pB$ given $pA$). The v-structure (see Fig. 2.1(e)), on the other hand, entails that the endpoints are independent, but become dependent conditional on their common effect; no other structure entails the same CI pattern, and therefore the causal structure is identifiable from observational data.

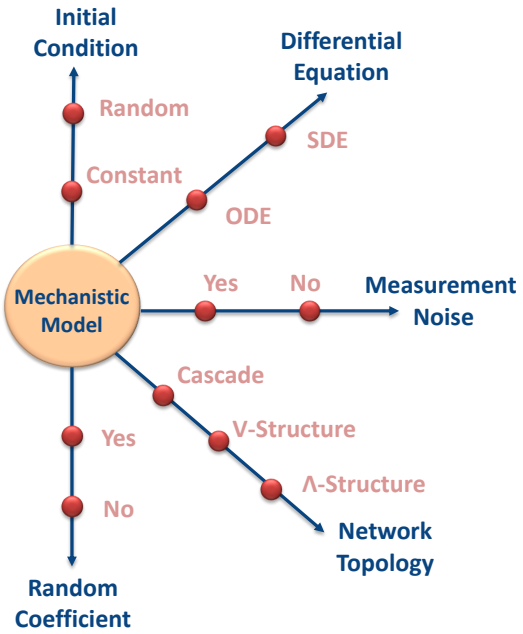
## 3.2 Factorial Design of Experiments

To reproduce the limitations of a biological setting for which causal discovery algorithms may fail to recover the true causal graph we devised an experimental design where we test several combinations of model formalisms, model specifications and parameter settings based on known conditions encountered in biological systems. We would say that all these factors are different forms of noise either measurement or biological.

It would be essential to elucidate the differentiation between measurement noise and biological noise. The former regards instrumental noise, errors in cell handling, etc. The latter regards the variability that stems from the biological system itself. There are many sources of biological noise one can incorporate while designing a mechanistic model of a molecular reaction network. These include, variation emanating from noisy gene transcription and/or expression, molecular diffusion, the microenvironment and others. So, in this work, we test for all these sources of variation by creating an in-silico experimental design where we employ several factors. Each factor accounts for the different source of noise.

In terms of model formalism the Differential Equations factor accounts for the variability in the chemical reactions. ODEs are usually utilized when the number of molecules in a well-mixed reaction network is large and a deterministic model can sufficiently describe the molecular interactions between system components. On the other hand, SDEs are preferred if the intrinsic randomness or stochasticity of the system is created from a low number of molecules and/or a molecular diffusion processes may critically affect the outcomes [29]. In terms of specifications we consider the effect of *Measurement Noise* factor because we want to take into account the possible inability to perfectly measure the amounts of each component. The noise we employ is identically and independently distributed white noise. In terms of parameter settings, we consider the kinetic rate constants of component reactions as *Random Coefficients* and, so, we model them as random variables. Biologically, random kinetic rates are a source of heterogeneity in cell populations [29]. Furthermore, we consider the stimulus of the reaction network

as *Initial Condition* making the assumption that sometimes all cells have the same *(constant)* or different *(random)* number of receptors [29]. We remind that in Section 2.4 we have followed the ODE and SDE steady-state solution for both initial conditions. Fig. 3.1 depicts the factorial design and the eight differential equation models that we investigate.




(a)

| Model | Diff.Eq. | Rand.Coef | Meas.Noise |
|---|---|---|---|
| 1 | ODE | No | No |
| 2 | ODE | No | Yes |
| 3 | ODE | Yes | No |
| 4 | ODE | Yes | Yes |
| 5 | SDE | No | No |
| 6 | SDE | No | Yes |
| 7 | SDE | Yes | No |
| 8 | SDE | Yes | Yes |

(b)

Figure 3.1: The factorial design. (a) In the graph-like diagram the orange sphere represents the mechanistic model. The direction of the rays ends up on the parameters or specifications of the model or else in the factors of the design. Every ray consists of some red nodes that depict the different ways of these parameters or specifications. Red nodes of different rays can be combined for building mechanistic models. For example, we can build a mechanistic model that corresponds to cascade topology using ODEs, with random coefficients, without measurement noise and constant initial condition or a V-structure topology using SDEs, without random coefficient, with measurement noise and random initial condition. (b) List of mechanistic models that represents the differential equation models that are extracted from the graph-like diagram. We do not take into consideration the initial condition and the network topology.

## 3.3 Observational Data

In our work we examine only single-cell observational data of molecular interaction networks at steady-state without performing any kind of interventions. The

objective of an observational study is to elucidate cause-and-effect relationships in the case-study system and to find procedures whose effects it is desired to discover, or to assign subjects at random to different procedures [19]. In an observational study, there is an effort to measure as many variables as possible (observational data) and after to test hypotheses about what changes in a set of those variables are associated with changes in other sets of variables drawing conclusions about causality in these associations.

On the other hand, it doesn't mean that observational data can not lead to errors when we discuss about causal inference. The issue that rises is that the observational data may not contain variables that are the real causes. These unmeasured variables are called "latent" variables and can make someone falsely assume that the measured variables are causing something, whereas in truth it is one of the latent variables.

However, it would be reasonable if someone asked why observational data and not real data? Indeed, the use of real data will greatly benefit the purposes of this work. However, to the best of our knowledge there are no such data publicly available. Even designing single-cell biological experiments that can reproduce the issues mentioned in this work is still very difficult [5].

# Chapter 4

# Results

In this Chapter we present results for each of the three network topologies and each of the eight mechanistic models (see Figs. 2.1 and 3.1 for their description) with both constant and random initial condition. For each case in the factorial design, which is described in Section 3.2, we discuss the consistency between the marginal causal graph and the graph that we recover from the joint distribution at steady-state. With the term marginal causal graph we call the causal graph which consists of nodes (or variables) in subset of nodes (or variables) being retained, regardless of the parameter settings or specifications of the model. For example, in Fig. 2.1(g-i) the variables $R$ (or $R_1, R_2$), $pA$ and $pB$ are retained in the causal graph of the different models derived from the factorial design that we will examine next.

Whenever the theoretical steady-state solution of a differential equation system is intractable numerical results are shown from its simulation. We also discuss how the causal assumptions are violated and provide an empirical validation of our theoretical considerations, again, from simulated data.

For ease of exposition, we will use the term $dep(V_i, V_j|V_k)$ to denote the dependence between variables $V_i$ and $V_j$ given variable $V_k$. In contrast, we will use the term $indep(V_i, V_j|V_k)$ to denote the CI or, equivalently, the *d-separation* between variables $V_i$ and $V_j$ given variable $V_k$. Furthermore, for the models where we suppose that each protein is measured with some experimental noise, their variables will no longer be able to represent the measured quantities. Instead, their noisy counterparts which we will symbolize using the prime character e.g. $R' \equiv R + \varepsilon$, where $R$ is a protein measured without experimental noise and $\varepsilon$ is an independent noise term $\sim \mathcal{N}(0, \sigma_\varepsilon^2)$.

## 4.1 Cascade topology

The marginal causal graph for the cascade topology is shown in Fig. 4.1(b). From this graph the following relations between system variables are induced: $dep(R, pA|\emptyset)$, $dep(R, pB|\emptyset)$, $dep(pA, pB|\emptyset)$, $dep(R, pA|pB)$, $dep(pA, pB|R)$ and

$indep(R, pB|pA)$.

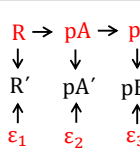
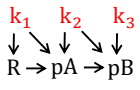
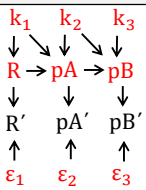
| Topology | Class of MEG | Expected DAG with | | | |
|---|---|---|---|---|---|
| | | No parameters (model 1&5) | Meas. Noise (model 2&6) | Rand. Coef. (model 3&7) | Meas. Noise & Rand.Coef. (model 4&8) |
| Cascade | R — pA — pB (a) | R → pA → pB (b) |  (c) |  (d) |  (e) |

Figure 4.1: **Cascade topology**. In (a) the desired MEG is presented (b) the marginal causal graph and the ground-truth causal graph for Models 1, 5 is shown, while, in (c) the respective graph for Models 2, 6 , in (d) for Models 3, 7, and in (e) for Models 4,8 based on the list from Fig. 3.1. Note that we do not show the Brownian motion variables for the SDE models (i.e., Models 5–8) because they do not alter the causal properties of the system since they are independent to each other and each one affect only one variable. We also remark that the unmeasured (latent) quantities are denoted with red color while the observed ones are denoted with black. For the corresponding mechanistic model see Fig. 2.1(a).

## Model 1

In Model 1, in which there is neither random coefficient nor measurement error, the variables are linearly dependent and deterministically related to each other. Because of this, the variance of the steady-state distribution is zero and the faithfulness assumption is violated. As a result, the application of any causal discovery algorithm is rendered meaningless [25].

## Model 2

In this model, we suppose that each protein is measured with some experimental noise. The resulting causal graph is shown in Fig. 4.1(c). As seen, the noise alters the underlying causal structure in that, introducing new variables to the system, all of which are latent. Here, the faithfulness condition is violated, because the addition of measurement noise will not affect the steady-state solution of the ODE model and, therefore, the deterministic relations between variables $R$, $pA$ and $pB$ still hold as in Model 1. Moreover, we notice from the graph, that between the actual measured quantities the relation $indep(R', pB'|pA')$ cannot hold because there exists a path through $\langle R, pA, pB\rangle$. In other words, the noise-contaminated variables cannot be d-separated since information can flow between $R'$ and $pB'$

even if we condition on $pA'$. Thus, theoretically, the underlying structure cannot be captured.

We also empirically validated these observations using PC algorithm. Assuming two levels of additive i.i.d Gaussian noise ($\sigma_\epsilon = 1$ or 3), two cases for initial condition (let $X_0$ be the initial condition then $X_0 = 0.2$ for the constant case and $X_0 \sim \mathcal{N}(0.4, \sigma^2_{X_0})$ with $\sigma_{X_0} = 0.05$ or 0.2 for the random case), we calculated for various sample sizes, for $k_1$ in the range $[10^{-3}, 10^{-1}]$ and $k_2 = 10^{-1.5}$ where PC algorithm returns a MEG that reenacts the underlying causal structure. Table 4.1 presents the simulation results. We can observe that when $X_0$ is fixed the obtained MEG is fully-unconnected in most of the cases while it is irrelevant in the others. The above results differ when the initial condition is random. In that case PC algorithm returns fully-connected MEGs.

The independence between the measured variables, in case of constant initial condition, is not consistent with the d-separation argument. Due to the deterministic relations between the variables $R$, $pA$ and $pB$, the statistics of the measured variables $R'$, $pA'$ and $pB'$ are actually the statistics of the noise which is independent and identically distributed. On the other hand, when $X_0$ is a random variable, the same source of variation is passed on from one measured variable to the other. For this reason all the measured variables show correlated and PC algorithm discovers fully-connected MEGs.

| Samples | Exp. Des | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_\varepsilon$ | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | - | - | 85.7 | 14.3 | 0.3 | 49.7 | - | 50 | - | 51 | - | 49 |
| | - | ✓ | - | - | 85.9 | 14.1 | 4 | 28 | - | 68 | - | 50.3 | - | 49.7 |
| 5000 | ✓ | - | - | - | 85.8 | 14.2 | - | 59 | - | 41 | - | 59 | - | 41 |
| | - | ✓ | - | - | 85.7 | 14.3 | 1 | 55 | - | 44 | - | 59 | - | 41 |
| 10000 | ✓ | - | - | - | 85.9 | 14.1 | - | 63 | - | 37 | - | 63 | - | 37 |
| | - | ✓ | - | - | 85.8 | 14.2 | 0.2 | 61 | - | 38.8 | - | 63 | - | 37 |

Table 4.1: **Cascade simulation results for Model 2** for constant initial condition $X_0$ and for random initial condition when $\sigma_{X_0} = 0.05$ and 0.2. "Check" symbol determines the actual value for the variance parameters for various sample sizes and 1000 repetitions. The columns Desired, Fully-Connected, Fully-Unconnected and Irrelevant represent which Markov equivalent graph is returned by the PC algorithm and report how many times is found.

## Model 3

Now, we assume that the kinetic rate constants $k_i$ are random coefficients (i.e., random variable). Fig 4.1(d) shows that the random coefficients variables $k_1$ and $k_2$ affect simultaneously two measured variables in the system. Such variables are called latent confounders because (i) they affect more than one quantity and (ii) they are not measured. Thus, the presence of the confounders violates causal sufficiency assumption. Furthermore, we notice that the latent variables $k_1$ and $k_2$ allow information to flow from $R$ to $pB$ and as a result the relation $indep(R, pB|pA)$

doesn't hold.

To evaluate the ability of PC algorithm to infer the desired MEG, we consider that the rate constants follow a log-normal distribution with parameters $\mu_k = -2$ and $\sigma_k = 0.1$ and 0.5 ($10^{\mathcal{N}(-2,\sigma_k^2)}$) and for both cases of initial condition as we have seen in Model 2. The simulation results, which are presented in Table 4.2, indicate that when the initial condition is constant PC algorithm returns fully-unconnected MEG. The same applies when the uncertainty of the initial condition is low and the uncertainty of the rate constants is high. On the other hand, if we keep the uncertainty of both $X_0$ and $k_i$ low or we increase the standard deviation of $\sigma_{X_0}$ at 0.2, the obtained MEG is consistent with the causal theory being fully-connected.

The independence that arises in the case of the fixed $X_0$ is due to the deterministic relations between the variables and, more specifically, to the cancellation of the effect of $k_i$ variables as latent confounders. For example, we know that at steady-state holds:

$$R = \frac{X_0}{k_1} \tag{4.1}$$

and

$$pA = \frac{k_1 R}{k_2} \tag{4.2}$$

however from eq. (4.1), (4.2)

$$pA = \frac{\cancel{k_1}\frac{X_0}{\cancel{k_1}}}{k_2} = \frac{X_0}{k_2} \tag{4.3}$$

that is $pA$ depends on $k_2$ while $k_1$ is eliminated. Hence, $pA$ depends only on $X_0$ and $k_2$, maintaining the deterministic relations of the variables. On the other hand, the randomness in the initial condition eliminates the deterministic relationships and for that the obtained MEGs are the expected ones. The fully-unconnected MEGs when the uncertainty is high, is due to the independence characteristics of $k_i$ variables that dictates the independence between the measured variables.

### Model 4

Model 4 is a combination of the presence of the measurement noise (Model 2) and the uncertainty of the kinetic rates (Model 3) as it is presented in Fig 4.1(e). As before, causal sufficiency and faithfulness assumptions are violated rendering the inference of the desired MEG extremely improbable to obtain and making the relation $dep(R, pB|pA)$ to be in effect.

Table 4.2 shows the simulation results. For the constant initial condition, as we have mention in Model 2 and 3, the variables $R$, $pA$ and $pB$ are deterministically related to each other and hence the statistics of the measured variables $R'$, $pA'$ and $pB'$ are the statistics of the noise. Therefore, PC algorithm returns fully-unconnected MEGs. In the random $X_0$ case, the outcome of the PC algorithm is the expected one from the ground-truth causal graph, that is fully-connected MEG.

| Samples | Exp. Des | | | | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_k$ | | $\sigma_\varepsilon$ | | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 0.1 | 0.5 | 0 | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | ✓ | - | - | - | - | 86.6 | 13.4 | 0.1 | 99.9 | - | - | - | 100 | - | - |
| | - | ✓ | ✓ | - | - | - | - | 88 | 12 | - | - | 86.4 | 13.6 | 2.7 | 11 | 14.8 | 71.5 |
| 5000 | ✓ | - | ✓ | - | - | - | - | 85.4 | 14.6 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | ✓ | - | - | - | - | 85.6 | 14.4 | - | - | 78.5 | 21.5 | 1.2 | 94.8 | - | 4 |
| 10000 | ✓ | - | ✓ | - | - | - | 0.1 | 87.3 | 12.6 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | ✓ | - | - | - | - | 84.7 | 15.3 | - | 0.1 | 73.5 | 26.4 | - | 100 | - | - |
| 1000 | ✓ | - | - | ✓ | - | - | - | 85.8 | 14.2 | - | 100 | - | - | - | 100 | - | - |
| | ✓ | - | - | - | ✓ | - | - | 87.5 | 12.5 | 0.1 | 99.9 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | - | - | 88.4 | 11.6 | - | - | 84.6 | 15.4 | 2.9 | 9.3 | 15.5 | 72.3 |
| | - | ✓ | - | - | ✓ | - | - | 87.8 | 12.2 | - | - | 84.2 | 15.8 | 2.5 | 9 | 14 | 74.5 |
| 5000 | ✓ | - | - | ✓ | - | - | - | 85.7 | 14.3 | - | 100 | - | - | - | 100 | - | - |
| | ✓ | - | - | - | ✓ | - | - | 85.4 | 14.6 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | - | - | 86.9 | 13.1 | | - | 78.9 | 21.1 | 0.6 | 94.3 | - | 5.1 |
| | - | ✓ | - | - | ✓ | - | - | 86 | 14 | - | - | 80.2 | 19.8 | 1.2 | 94.3 | - | 4.5 |
| 10000 | ✓ | - | - | ✓ | - | - | - | 86.8 | 13.2 | - | 100 | - | - | - | 100 | - | - |
| | ✓ | - | - | - | ✓ | - | - | 86.5 | 13.5 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | - | - | 86.7 | 13.3 | 0.1 | - | 73.7 | 26.2 | - | 99.9 | - | 0.1 |
| | - | ✓ | - | - | ✓ | - | - | 86 | 14 | - | 0.1 | 74.7 | 25.2 | - | 99.9 | - | 0.1 |

Table 4.2: **Cascade simulation results for Model 3** (first six rows) and **4** (remaining rows) for various sample sizes and 1000 repetitions for constant initial condition $X_0$ and for random initial condition when $\sigma_{X_0} = 0.05$ and 0.2. "Check" symbol determines the actual value for the variance parameters. The columns Desired, Fully-Connected, Fully-Unconnected and Irrelevant represent which Markov equivalent graph is returned by the PC algorithm and report how many times is found.

## Model 5

Due to the intrinsic stochasticity from the Brownian motion of the SDE, all causal assumptions are satisfied in Model 5. Thus, the question becomes whether or not the network we can recover from the steady-state observational data coincides with the structure of the true MEG. If not, the causal discovery algorithm will never be able to infer the underlying, true causal structure. Moving forward, we explore all the elements of the covariance matrix $\Sigma$ - it reveals the dependence of pairwise variables given the empty set - and the concentration matrix $\Theta$ - it implies the dependence of pairwise variables given the other variables of the system. It is important to refer that we denote with $\Sigma_{5_{const}}$ and $\Theta_{5_{const}}$ the covariance and concentration matrix for constant $X_0$ of Model 5 while with $\bar{\Sigma}_{5_{rand}}$ and $\bar{\Theta}_{5_{rand}}$ the covariance and concentration matrix of the given model for random $X_0$.

The study of $\Sigma_{5_{const}}$ and $\bar{\Sigma}_{5_{rand}}$ matrices shows that all the variable pairs are dependent given the empty set. Moreover, we find out that both $\Theta_{5_{const}}(1,2)$ / $\bar{\Theta}_{5_{rand}}(1,2)$ and $\Theta_{5_{const}}(2,3)$ / $\bar{\Theta}_{5_{rand}}(2,3)$ elements which correspond to the $indep(R, pA|pB)$ and $indep(pA, pB|R)$, respectively, are always positive denoting that the given pairs are not d-separated. However, the element $\Theta_{5_{const}}(1,3)$ / $\bar{\Theta}_{5_{rand}}(1,3)$ which correspond to the $indep(R, pB|pA)$ should be equal to zero in order to reconstruct the desired MEG of the marginal causal graph. Analytically,

$$\Theta_{5_{const}}(1,3) = \frac{2k_1k_2k_3(k_1+k_3)(k_2+k_3)(2k_1^2-k_2^2)}{\sigma^2(6k_1^4k_2^2+4k_1^4k_2k_3+\cdots+2k_2^3k_3^3+k_2^3k_3^4}$$

and

$$\bar{\Theta}_{5_{rand}}(1,3) = (\cdots)\Bigg[\frac{\big(k_1k_2k_3(-2k_1^2+k_2^2)\sigma^2\big)}{\sigma^2(6k_1^5k_2^3k_3\sigma^2+4k_1^5k_2^3\sigma_{X_0}^2+\cdots+4k_2^4k_3^4\sigma_{X_0}^2+2k_2^3k_3^5\sigma_{X_0}^2)} + \frac{\big(2k_1k_2(k_1+k_2)^2+2k_3(-k_1^3+2k_1k_2^2+k_2^3)\big)\sigma_{X_0}^2}{\sigma^2(6k_1^5k_2^3k_3\sigma^2+4k_1^5k_2^3\sigma_{X_0}^2+\cdots+4k_2^4k_3^4\sigma_{X_0}^2+2k_2^3k_3^5\sigma_{X_0}^2)}\Bigg]$$

Apparently, the above relations shows that $\Theta_{5_{const}}(1,3)$ and $\bar{\Theta}_{5_{rand}}(1,3)$ take non zero values. This is a rather surprising result since we expected that this relationship would always be zero. An explanation for this outcome is given by the discretization of the corresponding SDE. Explicitly, based on Fig. 2.1(a) and eq. (2.3), we define the Euler-Marayuma[1] structural equation model (SEM) on some interval of time [0, T], which consists of the following:

1. $\Delta t = T/N$, where $N$ equal subintervals of width $\Delta t > 0$
2. The initial or primary variables $R_0$, $pA_0$ and $pB_0$
3. The discretized Brownian motion $W_{N+1} - W_N = \Delta W_N$, where $\Delta W_N = \sqrt{\Delta t}$ $\mathcal{N}(0,1)$
4. The functional relationships are given by:

$$R_{N+1} = R_N - (k_1R_N)\Delta t + \sigma\Delta W_N$$

$$pA_{N+1} = pA_N + (k_1R_N - k_2pA_N)\Delta t + \sigma\Delta W'_N$$

$$pB_{N+1} = pB_N + (k_2pA_N - k_3pB_N)\Delta t + \sigma\Delta W''_N$$

A visualization of the unraveled Bayesian network is shown in Fig. 4.2 (a). We observe that the relationships $dep(R_{N+1}, pB_{N+1}|\emptyset)$ and $dep(R_{N+1}, pB_{N+1}|pA_{N+1})$ always hold since there are paths (presented in Fig. 4.2(b)-(c), respectively) that allow the flow of information from $R_{N+1}$ to $pB_{N+1}$ variable. We intuitively understand that the present of variable $pB$ depends on the past of variable $R$ given the past of variable $pA$.

Nevertheless, the relation $indep(R, pB|pA)$ is true, with probability equal to zero, if and only if $k_2 = \sqrt{2}\ k_1$, for the constant initial condition case, and $k_2 = 0.6194\ k_1$, for the random one. For that reason, we validate the ability of causal discovery algorithms to infer *de-novo* the true reaction diagram at least from steady-state data. In Fig. 4.3(a) we provide the contour diagrams for $\Theta_{5_{const}}(1,3)$ and $\bar{\Theta}_{5_{rand}}(1,3)$, respectively, as a function of $k_1$ and $k_2$ at infinite sample size, with the remaining parameters sets to 1. The yellow line highlights the relation for each case where the desired MEG can be discovered.

[1]Euler-Maruyama method (also called the Euler method) is a method for the approximate numerical solution of an SDE

For the more practical case of finite sample size we investigated the performance of the PC algorithm assuming that the variables are linearly correlated. Due to finite sampling, the estimation of the partial correlation that PC employs as a test of statistical independence may be biased, resulting in statistical errors. Such errors might broaden the area where $\Theta_{5_{const}}(1,3)$ and $\bar{\Theta}_{5_{rand}}(1,3)$ are statistically indistinguishable from zero, especially at the purple areas of Fig. 4.3(a) where the values of $\Theta_{5_{const}}(1,3)$ and $\bar{\Theta}_{5_{rand}}(1,3)$ are smaller. In Fig 4.3(b) we present eight statistical discovery diagrams at several samples sizes with $k_1$ being in the range $[10^{-3}10^{-1}]$, $k_2 = 10^{-1.5}$ and $10^{-2}$, for both fixed and random $X_0$ and average magnitude of the random fluctuations $\sigma = 0.5$. For the random case we examine several values for the standard deviation of $X_0$ ($\sigma_{X_0} = [0.05\ 0.2\ 0.5]$) while we keep the mean value constant at $\mu_{X_0} = 0.4$ .

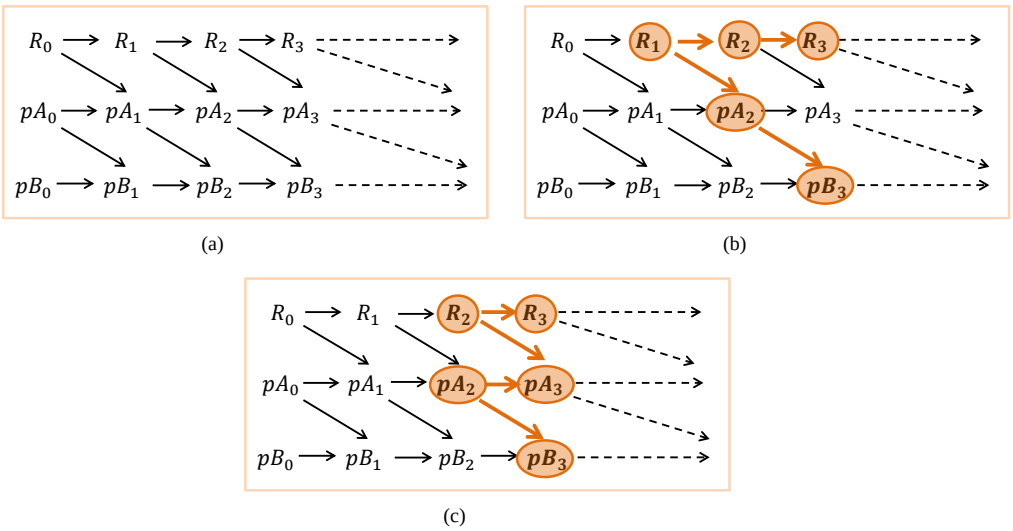



Figure 4.2: (a) Unraveled Bayesian Network of the Euler-Marayuma SEM of Model 5 for cascade topology. Orange nodes and arcs show that information flows via a path from variable $R_3$ to $pB_3$ condition on (b) empty set and (c) $pA_3$, respectively. We note that these paths may be not unique for each condition set; each state is an observation in equilibrium while the variables of Brownian motion $W_{N+1} - W_N$ are not depicted on the graphs since they are independent and identical distributed. The conditional dependency relations are generalized for $N+1$ variables.

We observe that in all cases PC algorithm finds an area (purple area) where the obtained MEG is the desired one. For fixed $X_0$ (Fig 4.3(b)) this area tends to increase as the sample size is decreased. However, as the sample size increases the statistical error between the numerical and the theoretical values for the concentration matrix becomes smaller and the numerical outcomes converge to the theoretical ones.

On the other hand, for the random initial condition we observe that as the

standard deviation of $X_0$ increases the area where we find the desired MEG shifts to right (see Fig 4.3(b)) and as the sample size increases the numerical outcomes diverge from $k_2 = 0.6194\ k_1$ relation. This transposition is due to the ratio of $\frac{\sigma}{\sigma_{X_0}}$. Particularly, when $\sigma \gg \sigma_{X_0}$, the term in $\bar{\Theta}_{5_{rand}}(1,3)$ which is multiplied by $\sigma$, determines the position of the area where PC returns the desired MEG. If $\sigma_{X_0}$ is close or greater of $\sigma$ (in any case when it has significant value) it shifts the desired area to right. However, we notice that in Fig 4.3(b) where $k_2 = 10^{-2}$ the area where we can find the desired MEG is greater for small sample size. In Fig 4.3 (a-right) we can observe that the area that corresponds to $k_2 = 10^{-2}$ is the blue one in which $\bar{\Theta}_{5_{rand}}(1,3)$ is close to zero. Thus PC algorithm can reenact the desired MEG.

### Model 6

The addition of measurement noise to the SDE of the previous model changes the distributional properties of the measured quantities. Both the steady-state and the noise distributions are independent Gaussians, thus the probability density here is also Gaussian with the covariance matrix being the sum of the covariance matrices of the respective distributions. Hence, the new concentration matrix will be $\Theta_{6_{const}} = (\Sigma_{5_{const}} + \sigma_\varepsilon^2 I)^{-1}$ and $\bar{\Theta}_{6_{rand}} = (\bar{\Sigma}_{5_{rand}} + \sigma_\varepsilon^2 I)^{-1}$ for constant and random $X_0$, respectively, with

$$\Theta_{6_{const}}(1,3) \propto \Big(k_2^2 - 2k_1^2 + \frac{\sigma_\varepsilon^2}{\sigma^2}(2k_1k_2^2 + 2k_3k_1k_2 + 2k_2^3 + 2k_3k_2^2)\Big)$$

and

$$\begin{aligned}\bar{\Theta}_{6_{rand}}(1,3) \propto\ & k_1k_2k_3(-2k_1^2 + k_2^2)\sigma^4 \\ &+\Big(2k_1k_2(k_1+k_2)^2 + 2k_3(-k_1^3 + 2k_1k_2^2 + k_2^3)\Big)\sigma^2\sigma_{X_0}^2 \\ &+\Big[4\Big(k_1k_2(k_1^2k_2 + 2k_1k_2^2 + k_2^3) + k_1k_2k_3(k_1^2 + 3k_1k_2 + k_1k_3 + 3k_2^2 + 2k_2k_3) \\ &+k_2^3k_3(k_2+k_3)\Big)\sigma_{X_0}^2 + 2k_1k_2k_3\Big(k_1k_2^2 + k_1k_2k_3 + k_2^3 + k_2^2k_3\Big)\sigma^2\Big]\sigma_\varepsilon^2\end{aligned}$$

Despite the fact that the measured variables are not d-separated, for the reasons we discussed in Model 2, the above formulas indicate that there is a combination of parameter values that makes $\Theta_{6_{const}}$ and $\bar{\Theta}_{6_{rand}}$ equal to zero and, hence, for that specific combination the observational data have the desired causal structure. A comparison with Model 5, reveals that the terms become zero are present in both models' $\Theta_{x_{const}}(1,3)$ and $\bar{\Theta}_{x_{rand}}(1,3)$, where $x = [5,6]$ denoting the number of the model, but there is an additional term when measurement noise is added.

Again, we validate our theoretical findings with numerical evidence but now for two levels of measurement noise $\sigma_\varepsilon = 1$ and 3, $k_1$ in the range $[10^{-3}, 10^{-1}]$ and various values for the sample size. The statistical discovery diagrams in Fig. 4.4 show that when the standard deviation of measurement noise equals to 1, the area where PC algorithm discovers the desired MEG has been shifted to right , for both

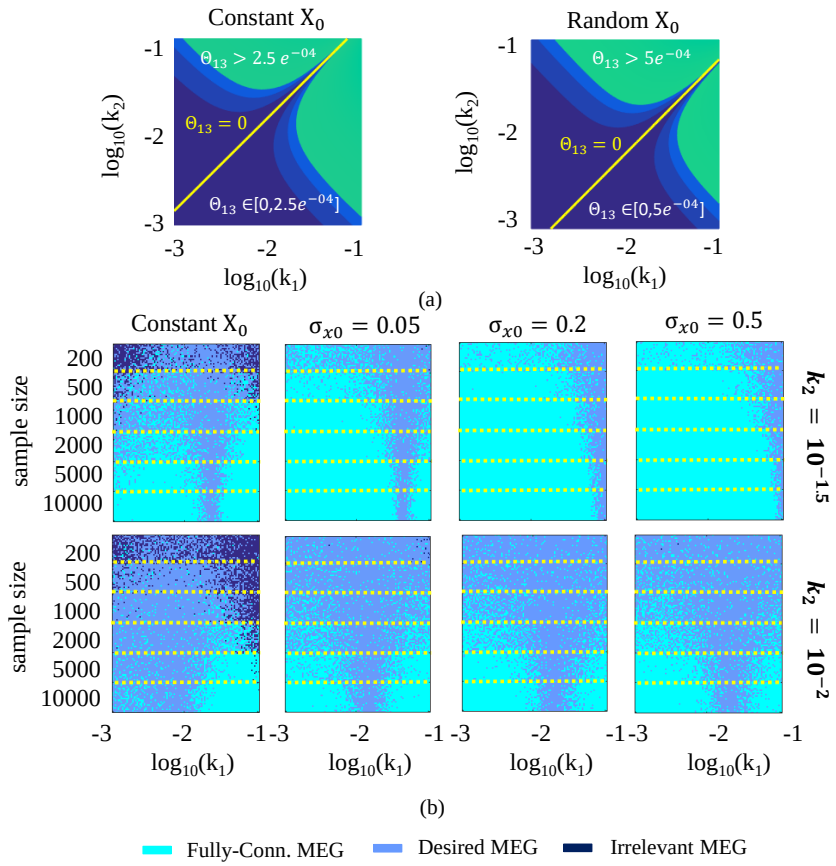



Figure 4.3: Cascade's (a) contour diagram of Model 5 at infinite sample size for constant initial condition $X_0$ (left) and random initial condition $X_0$ (right). Yellow line indicates the relation $k_2 = \sqrt{2}\ k_1$ (left) and $k_2 = 0.6194\ k_1$ (right) in which the element $\Theta_{5_{const}}(1,3)$ / $\bar{\Theta}_{5_{rand}}(1,3)$ is equal to zero and PC algorithm theoretically finds the desired MEG; blue color represents the area where $\Theta_{5_{const}}(1,3)$ / $\bar{\Theta}_{5_{rand}}(1,3)$ is close to zero while the green color represents the area where $\Theta_{5_{const}}(1,3)$ / $\bar{\Theta}_{5_{rand}}(1,3)$ is greater than zero and PC algorithm cannot discover the desired MEG. (b) statistical discovery diagrams of Model 5 when $k_2 = 10^{-1.5}$ (up diagrams) and $k_2 = 10^{-2}$ (down diagrams), for both fixed $X_0$ and random $X_0$ with $\sigma_{X_0} = [0.05,\ 0.2,\ 0.5]$. For each case we examine several finite sample sizes by running the PC algorithm twenty times. Blue color denotes the case where PC finds fully-connected MEG, cyan color the case where PC algorithm finds irrelevant to the ground-truth MEGs and purple color the case where PC algorithm finds the desired MEG. In the purple area the value of $k_1$ is close to $\frac{k_2}{\sqrt{2}}$ and $\frac{k_2}{0.6194}$ for constant and random initial condition, respectively.

cases of $X_0$, as expected by the theoretical value of $\Theta_{6_{const}}$ and $\bar{\Theta}_{6_{rand}}$, respectively. No other difference is visually observed when compared to the noiseless case (i.e., Model 5). In contrast, when the level of measurement noise increases to $\sigma_\varepsilon = 3$, for the fixed $X_0$ case there is no clear region where PC algorithm returns the desired MEG since noise dominates and thus, the statistical error between the numerical and theoretical concentration matrix becomes greater requiring larger sample sizes for robust discovery. In the random initial condition, as the uncertainty of $X_0$ increases, the area where the desired MEG is obtained is shifted to the right as well as exceeds the range of $k_1$ values that we test, resulting in fully-connected MEG.

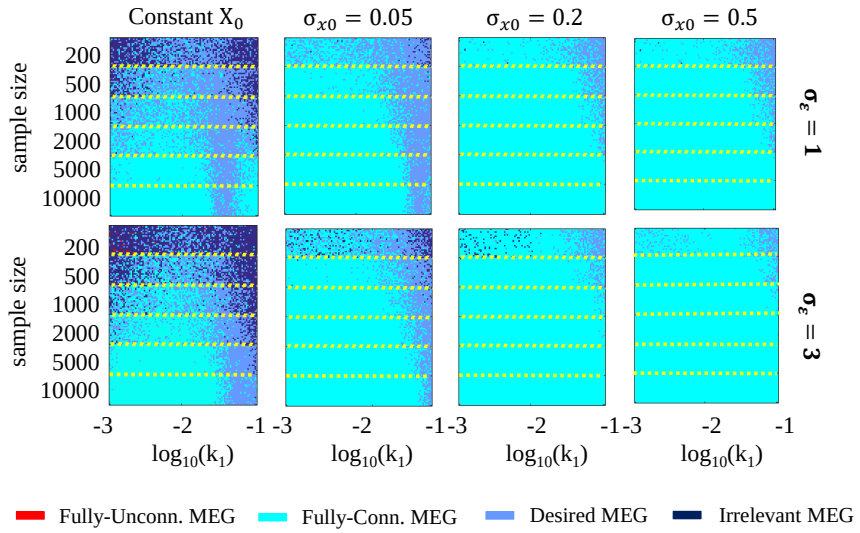



Figure 4.4: Cascade's statistical discovery diagrams of Model 6 for both for constant and random $X_0$ similar to Fig. 4.3(b) when $k_2 = 10^{-1.5}$ and $\sigma_\varepsilon = 1$ (up) and $\sigma_\varepsilon = 3$ (down), respectively. Blue color indicates the case where PC finds fully-connected MEG, cyan color the case where PC algorithm finds irrelevant to the ground-truth MEGs;red color the case where the measured variables are fully-independent and purple color the case where PC algorithm finds the desired MEG.

## Model 7

Similar to Model 3, the $\{k_i\}$'s here are random variables that act as latent confounders for the measured quantities. For this case there is no analytical expression for the distribution of the produced observational data therefore we resort to numerical experiments. For this, we test PC algorithm 1000 times on simulated data with different sample sizes, a log-normal distribution for the rate constants with parameters $\mu_k = -2$ and $\sigma_k = 0.1$ and 0.5 and both cases for initial condition.

| Samples | Exp. Des | | | | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_k$ | | $\sigma_\varepsilon$ | | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 0.1 | 0.5 | 0 | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | ✓ | - | - | 86.1 | 13.9 | - | - | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | ✓ | - | - | 29.3 | 5.5 | 0.4 | 64.8 | - | - | 83.4 | 16.6 | 3.1 | 10.4 | 9.5 | 77 |
| 5000 | ✓ | - | ✓ | - | - | 55.7 | 44.3 | - | - | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | ✓ | - | - | 56.1 | 37.5 | - | 6.4 | - | - | 77.7 | 22.3 | 0.7 | 97.5 | - | 1.8 |
| 10000 | ✓ | - | ✓ | - | - | 23 | 77 | - | 0.2 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | ✓ | - | - | 38.6 | 59.3 | - | 2.1 | - | - | 74.3 | 25.7 | - | 100 | - | - |
| 1000 | ✓ | - | - | ✓ | - | 93.5 | 6.3 | - | 0.2 | - | 100 | - | - | - | 100 | - | - |
| | ✓ | - | - | - | ✓ | 78.6 | 17.1 | - | 4.3 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | 29.6 | 6.6 | - | 63.8 | 5.8 | 24.3 | 3.8 | 66.1 | 5.1 | 77.6 | - | 17.3 |
| | - | ✓ | - | - | ✓ | 24.2 | 6.2 | 0.3 | 69.3 | 5.3 | 24.4 | 2.6 | 67.7 | 4.8 | 78.6 | - | 16.6 |
| 5000 | ✓ | - | - | ✓ | - | 84.1 | 15.9 | - | - | - | 100 | - | - | - | 100 | - | - |
| | ✓ | - | - | - | ✓ | 44 | 56 | - | - | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | 54 | 38.1 | - | 7.9 | 0.2 | 99.4 | - | 0.4 | - | 100 | - | - |
| | - | ✓ | - | - | ✓ | 54.4 | 34 | - | 11.6 | 0.1 | 99.5 | - | 0.4 | 0.1 | 99.9 | - | - |
| 10000 | ✓ | - | - | ✓ | - | 74.9 | 25.1 | - | - | - | 100 | - | - | - | 100 | - | - |
| | ✓ | - | - | - | ✓ | 13.1 | 86.9 | - | - | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | 35.8 | 62.5 | - | 1.7 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | - | ✓ | 39.7 | 57.8 | - | 2.5 | - | 100 | - | - | - | 100 | - | - |

Table 4.3: **Cascade simulation results for Model 7** and **8** similar to Table 4.2.

Table 4.3 presents the simulation results for both values of $\sigma_k$ (first six rows). For the constant $X_0$ case, the inferred MEG tends to be fully-connected as the sample size increases. This finding is in accordance to theory. Also, we observe that for the random initial condition the obtained MEGs are fully-connected with the only exception being when the standard deviation of $X_0$ is low and the uncertainty of the kinetic rates is high. Then PC algorithm discovers fully-unconnected MEGs. The reason for this discrepancy is the domination of $k_i$'s uncertainty.

### Model 8

In Model 8 we include measurement noise and uncertainty of rate constants into the basic SDE model (i.e., Model 5) combining both Models 6 and 7. In Table 4.3 (remaining rows) the simulation results indicate that for both cases of the initial condition PC algorithm finds fully-connected MEGs according to the ground-truth causal graph (see Fig 4.1). It is important to note that for the constant $X_0$ case, for low uncertainty in both kinetic rates and measurement noise, the desired MEG is recovered. This behavior is probably random due to the utilized sample size.

## 4.2 V-Structure topology

In the V-structure topology, the desired marginal causal graph is shown in Fig. 4.5(b). The respective MEG is exactly the same. Note that the defining relations for this topology are: $dep(R_1, pA|\emptyset)$, $dep(R_2, pA|\emptyset)$, $indep(R_1, R_2|\emptyset)$, $dep(R_1, pA|R_2)$, $dep(R_2, pA|R_1)$ and $dep(R_1, R_2|pA)$.

### Model 1

Due to deterministic relationship between the steady-state variables (violation of faithfulness assumption) and their linear dependence, the application of any causal discovery algorithm is trivial.

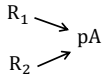
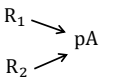
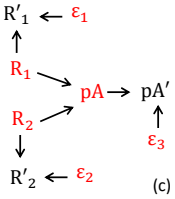
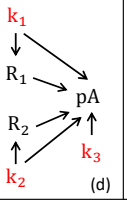
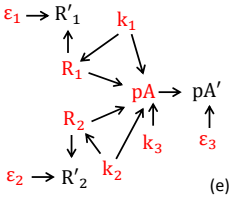
| Topology | Class of MEG | Expected DAG with | | | |
|---|---|---|---|---|---|
| | | No parameters (model 1&5) | Meas. Noise (model 2&6) | Rand. Coef. (model 3&7) | Meas. Noise &Rand. Coef. (model 4&8) |
| V-structure |  (a) |  (b) |  (c) |  (d) |  (e) |

Figure 4.5: **V-structure topology** similar to Fig. 4.1.

## Model 2

In Fig 4.5(c) we present the ground-truth causal graph when measurement noise is added in the system. We observe that the measured variables $R'_1$ and $R'_2$ are not d-separated conditioned on $pA'$ (information can flow from $R'_1$ to $R'_2$ through $\langle R_1, pA, R_2 \rangle$) and as a result the only violation is that of the faithfulness assumption due to deterministic relations between $R_1$, $pA$, $R_2$.

We simulate on two levels of measurement noise ($\sigma_\epsilon = 1$ or 3) both for constant and random initial condition $X_{1,2}$ (with $X_1 = X_2$ and $\sigma_{X1} = \sigma_{X2}$, respectively). Table 4.4 presents the simulation results. We notice that when the initial condition is fixed the obtained MEG is fully-unconnected in most of the cases while it is irrelevant in the others. On the other hand, when the initial condition is random PC algorithm returns the desired MEG as it is expected from the ground-truth graph. Similar to Model 2 of cascade topology (Section 4.1), for constant $X_{1,2}$ the second-order statistics of the measured variables are the second-order statistics of the measurement noise and thus the variables are mostly independent to each other.

It is important to stress that applying different values or standard deviations on $X_{1,2}$ there are no changes on the results.

| Samples | Exp. Des | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_\varepsilon$ | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | 0.2 | - | 85.8 | 14 | 95.5 | 5 | - | - | 95 | 5 | - | - |
| | - | ✓ | 0.1 | - | 85.9 | 14 | 83 | 2 | - | 15 | 95.5 | 5 | - | - |
| 5000 | ✓ | - | 0.3 | - | 85.7 | 14 | 95 | 5 | - | - | 95 | 5 | - | - |
| | - | ✓ | 0.2 | - | 85.8 | 14 | 94 | 3 | - | 3 | 95 | 5 | - | - |
| 10000 | ✓ | - | 0.3 | - | 85.7 | 14 | 95 | 5 | - | - | 95 | 5 | - | - |
| | - | ✓ | 0.2 | - | 85.8 | 14 | 95 | 2.5 | - | 2.5 | 95 | 5 | - | - |

Table 4.4: **V-structure simulation results for Model 2**, similar to Table 4.1

## Model 3

The rate constants $k_1$ and $k_2$ are latent confounders since they both affect two measured variables as it is shown in Fig. 4.5(d). Although causal sufficiency is violated, the defining relations of v-structure still hold in this model and hence, the desired MEG have to be recovered. Validating the results of Table 4.5 for two different values of $\sigma_k$, we see that these results are close to the results of Model 3 in cascade topology (Section 4.1). That is, when $X_{1,2}$ is fixed the obtained MEG, in most of the cases, is fully-unconnected because of the deterministic relations between the measured variables and the cancellation of $k_i$ as latent confounder. In all the other cases PC algorithm correctly infers the desired MEG, which is consistent to the causal theory, especially when the sample size is large and the statistical error is hence small.

We have to note that when $X_1 \neq X_2$ there are no changes on the results. When $\sigma_{X_1} \neq \sigma_{X_2}$ and $\sigma_k = 0.1$ most of the times (95 %) PC algorithm returns the desired MEG. On the other hand, when $\sigma_k = 0.5$ the obtained MEG is irrelevant to the desired one (86 %) showing the sensitivity of the causal discovery algorithms to the addition or lack of intrinsic stochasticity.

| **Samples** | Exp. Des | | | | | **Constant $X_0$** | | | | **Random $X_0$ with $\sigma_{X_0} = 0.05$** | | | | **Random $X_0$ with $\sigma_{X_0} = 0.2$** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_k$ | | $\sigma_\varepsilon$ | | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 0.1 | 0.5 | 0 | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | ✓ | - | - | 0.5 | - | 84.4 | 15.1 | 86.2 | 2.7 | 0.1 | 11 | 94.9 | 5.1 | - | |
| | - | ✓ | ✓ | - | - | 0.2 | - | 88.9 | 10.9 | - | - | 86.9 | 13.1 | 3.3 | 0.1 | 60.7 | 35.9 |
| 5000 | ✓ | - | ✓ | - | - | 0.2 | - | 85 | 14.8 | 93.3 | 3.9 | - | 2.8 | 94.8 | 5.2 | - | - |
| | - | ✓ | ✓ | - | - | 0.1 | - | 85 | 14.9 | 0.3 | - | 85.4 | 14.3 | 46.9 | 2 | 8.8 | 42.3 |
| 10000 | ✓ | - | ✓ | - | - | 0.5 | - | 84.8 | 14.7 | 95.5 | 1.8 | - | 2.7 | 93.9 | 6.1 | - | - |
| | - | ✓ | ✓ | - | - | 0.1 | - | 86.2 | 13.7 | 0.4 | - | 83 | 16.6 | 85.7 | 3 | 0.4 | 10.9 |
| 1000 | ✓ | - | - | ✓ | - | 0.3 | - | 83.8 | 15.9 | 87.9 | 2 | - | 10.1 | 94.7 | 5.3 | - | - |
| | ✓ | - | - | - | ✓ | 0.2 | - | 88.6 | 11.2 | 84.8 | 3.6 | 0.3 | 11.3 | 95.1 | 4.9 | - | - |
| | - | ✓ | - | ✓ | - | 0.1 | - | 88.6 | 11.3 | - | - | 85.3 | 14.7 | 3.6 | 0.2 | 60.4 | 35.8 |
| | - | ✓ | - | - | ✓ | 0.3 | - | 88.5 | 11.2 | - | - | 87.5 | 12.5 | 3.1 | 0.1 | 60.7 | 36.1 |
| 5000 | ✓ | - | - | ✓ | - | 0.2 | - | 87.7 | 12.1 | 95.1 | 2.7 | - | 2.2 | 95.9 | 4.1 | - | - |
| | ✓ | - | - | - | ✓ | 0.1 | - | 84.1 | 15.8 | 95.2 | 2.7 | - | 2.1 | 94.7 | 5.3 | - | - |
| | - | ✓ | - | ✓ | - | 0.3 | 0.1 | 85.3 | 14.3 | 0.5 | - | 83.9 | 15.6 | 38.6 | 1.1 | 12.4 | 47.9 |
| | - | ✓ | - | - | ✓ | 0.2 | - | 86.6 | 13.2 | 0.1 | - | 84.9 | 15 | 95.9 | 2.2 | - | 1.9 |
| 10000 | ✓ | - | - | ✓ | - | 0.2 | - | 88 | 11.8 | 95 | 2.2 | - | 2.8 | 94.6 | 5.4 | - | - |
| | ✓ | - | - | - | ✓ | 0.5 | 0.1 | 85.5 | 13.9 | 95 | 2.2 | - | 2.8 | 95.3 | 4.7 | - | - |
| | - | ✓ | - | ✓ | - | 0.2 | - | 86.3 | 13.5 | 0.4 | - | 82.6 | 17 | 80.6 | 3.2 | 0.7 | 15.5 |
| | - | ✓ | - | - | ✓ | - | - | 88 | 12 | 0.4 | - | 83.2 | 16.4 | 77.9 | 3.9 | 0.5 | 17.7 |

Table 4.5: **V-structure simulation results for Model 3** and **4** similar to Table 4.2.

## Model 4

Merging Models 2 and 3, the ground-truth causal graph is shown in Fig 4.5(e). According to causal discovery theory the results should be similar to Model 3. Evidently, in Table 4.5 we observe that the results are same as in previous model. Moreover, the results for different initial condition both for constant and random $X_{1,2}$ of Model 3 applies similarly on Model 4.

### Model 5

Examining the concentration matrices $\Theta_{5_{const}}$ and $\bar{\Theta}_{5_{rand}}$, we discover that all their elements take non-zero values. That means $dep(R_1, pA|R_2)$, $dep(R_2, pA|R_1)$, and $dep(R_1, R_2|pA)$. Additionally, we observe that $\Sigma_{5_{const}}(1,2)$ and $\bar{\Sigma}_{5_{rand}}(1,2)$ is exactly zero. That is $indep(R_1, R_2|\emptyset)$ and the probability density of the steady-state data for the given topology at infinite sample size is consistent with the desired MEG. In order to elucidate the above findings, we define in equilibrium, rely on Fig. 2.1(b) and eq. (2.3), the Euler-Marayuma SEM which consists of:

1. $\Delta t$ and $\Delta W_N$ similar to Model 5 in Section 4.1
2. The initial or primary variables $R_{1_0}$, $R_{2_0}$ and $pA_0$
3. The functional relationships are given by:

$$R_{1_{N+1}} = R_{1_N} - (k_1 R_{1_N})\Delta t + \sigma \Delta W_N$$
$$R_{2_{N+1}} = R_{2_N} - (k_2 R_{2_N})\Delta t + \sigma \Delta W'_N$$
$$pA_{N+1} = pA_N + (k_1 R_{1_N} + k_2 R_{2_N} - k_3 pA_N)\Delta t + \sigma \Delta W''_N$$

In Fig. 4.6 we depict a visualization of the unraveled Bayesian network based on the above. We observe that the independence relationships between $R_{1,N+1}$ and $R_{2,N+1}$ are always true verifying the theoretical conclusion. Furthermore, an numerical investigation has been carried out presented by statistical discovery diagrams (Fig. 4.7 for both constant and random initial condition, for $k_1 = 10^{-1.5}$ (up diagrams) and $10^{-2}$ (down diagrams)) for various $k_2$ values and several sample sizes. According to statistics, the area where the PC algorithm finds the desired MEG increases as the sample size is increased.

In case of different $X_{1,2}$ there are no differences on the simulation results. However, in the random initial case when $\sigma_{X1}$ is much greater than $\sigma_{X2}$ and vice versa (e.g $\sigma_{X1} = 0.5$ and $\sigma_{X2} = 0.05$) PC algorithm tends to find irrelevant to the desired MEGs, drawing again the responsiveness of the uncertainty.

### Model 6

As in Model 2, the addition of the measurement noise does not alter in the causal graph the defining marginal (in)-dependencies between the measured variables. Furthermore, the stochasticity of the SDE formalism eliminates the determinism in the variables and thus the recovery of the desired MEG becomes feasible given sufficiently large sample size. We investigate numerically, applying two different noise levels, the ability of PC algorithm to infer the desired MEG using statistical discovery diagrams as it is shown in Fig. 4.8. We observe that in both cases of the initial condition the desired MEG is obtained as the sample size increases, since the statistical error between the numerical and the theoretical concentration and covariance matrices tend to be low.

If we change the values of $X_{1,2}$ or $\sigma_{X12}$ there are no significant differences on the discovery diagrams.

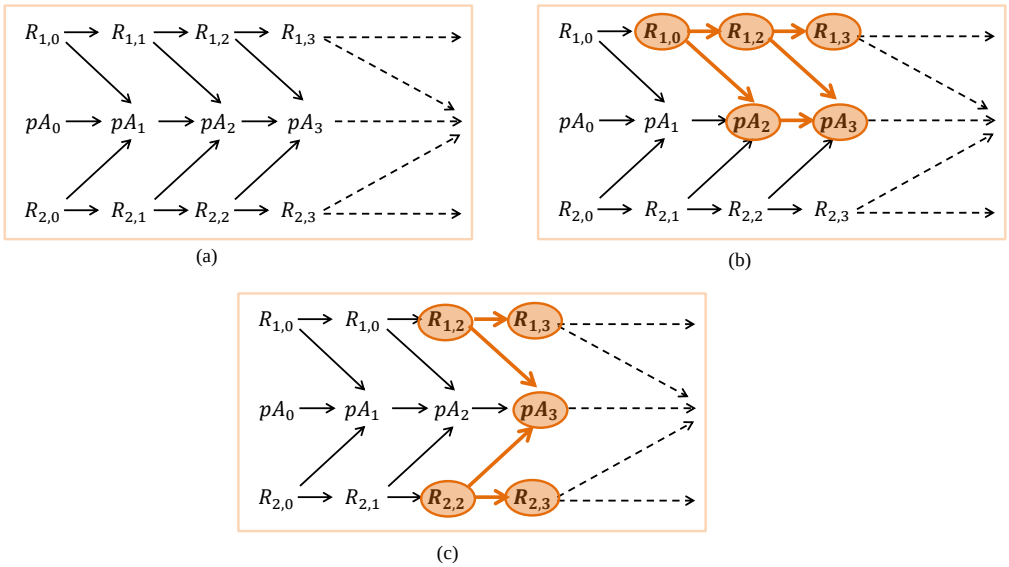



Figure 4.6: (a) Unraveled Bayesian Network of the Euler-Marayuma SEM of Model 5 for V-structure topology. Orange nodes and arcs show that information from variable $R_{1,3}$ to $R_{2,3}$ condition on (b) empty set cannot flow (b) $pA_3$ can flow, respectively. We note that these paths may be not unique for each condition set; each state is an observation in equilibrium while the variables of Brownian motion $W_{N+1} - W_N$ are not depicted on the graphs since they are independent and identical distributed. The conditional dependency and independency relations are generalized for $N + 1$ variables.

## Model 7

We apply the same set up as in the respective cascade model. Using the same arguments as in Model 3 we expect the desired MEG should be correctly inferred in all cases. Table 4.6 show the simulation results for both values of $\sigma_k$. We observe that in most of the times, for both fixed and random $X_{1,2}$, the obtained MEG coincides to the desired one and the success rate increases with the sample size. On the other hand, for low standard deviation of $X_{1,2}$ and high uncertainty of $k_i$ the uncertainty overshadows making the PC algorithm to find MEGs with all its variables fully-independent.

For different values for the initial condition the behavior of Model 7 is similar to Model 3.

## Model 8

Similar to Model 4, we supplement to the basic SDE model (i.e., Model 5) both measurement noise and uncertainty in the rate constants. The results, which are shown in Table 4.6 are expected to be similar to Model 7 according to causal

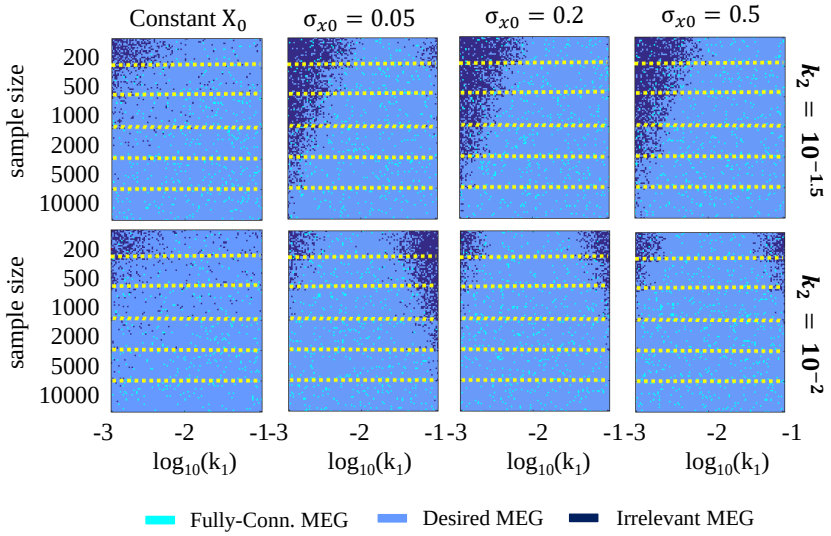



Figure 4.7: V-structure's statistical discovery diagrams of Model 5 when $k_2 = 10^{-1.5}$ (up diagrams) and $k_2 = 10^{-2}$ (down diagrams), for both fixed $X_0$ and random $X_0$ with $\sigma_{X_0} = [0.05, 0.2, 0.5]$. For each case we examine several finite sample sizes by running the PC algorithm twenty times. Blue color denotes the case where PC finds fully-connected MEG, cyan color the case where PC algorithm finds irrelevant to the ground-truth MEGs and purple color the case where PC algorithm finds the desired MEG.

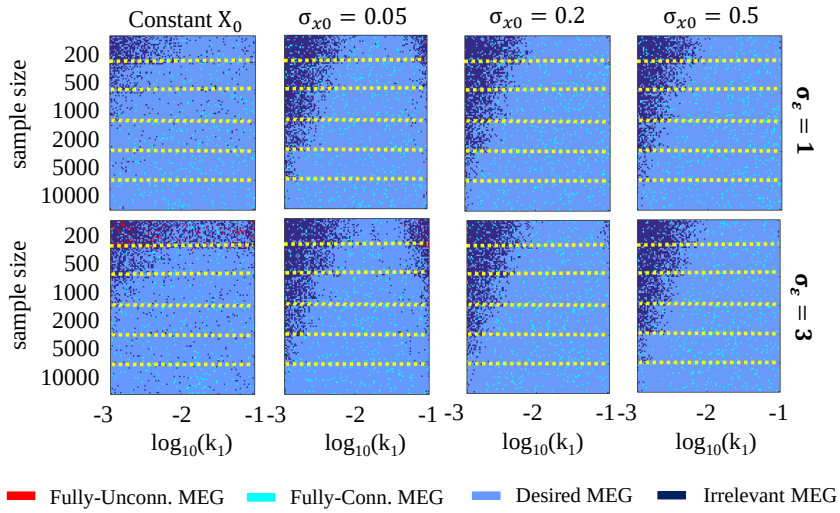



Figure 4.8: V-structure's statistical discovery diagrams of Model 6 for both for constant and random $X_0$ similar to Fig. 4.4.

| Samples | Exp. Des | | | | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_k$ | | $\sigma_\varepsilon$ | | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 0.1 | 0.5 | 0 | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | ✓ | - | - | 94.6 | 4.7 | - | 0.7 | 92.4 | 3.4 | - | 4.2 | 95.2 | 4.8 | - | - |
| | - | ✓ | ✓ | - | - | 62 | 1.2 | - | 36.8 | 0.4 | - | 86.8 | 12.8 | 4.9 | - | 58.7 | 36.4 |
| 5000 | ✓ | - | ✓ | - | - | 94.3 | 5.7 | - | - | 93.8 | 3.3 | - | 2.9 | 96.4 | 3.6 | - | - |
| | - | ✓ | ✓ | - | - | 85.4 | 2.6 | - | 12 | 0.6 | - | 83.2 | 16.2 | 45.5 | 2.1 | 7.5 | 44.9 |
| 10000 | ✓ | - | ✓ | - | - | 94.9 | 5.1 | - | - | 95.2 | 2.4 | - | 2.4 | 95 | 5 | - | - |
| | - | ✓ | ✓ | - | - | 90.6 | 2.2 | - | 7.2 | 0.5 | - | 86.1 | 13.4 | 84 | 3.8 | 0.5 | 11.7 |
| 1000 | ✓ | - | - | ✓ | - | 95.6 | 2.3 | - | 2.1 | 92.9 | 2.9 | - | 4.2 | 94.7 | 5.3 | - | - |
| | ✓ | - | - | - | ✓ | 94.9 | 2.8 | - | 2.3 | 90.6 | 3.3 | - | 6.1 | 94.3 | 5.7 | - | - |
| | - | ✓ | - | ✓ | - | 64.4 | 2.6 | - | 33 | 0.1 | - | 85.3 | 14.6 | 5.2 | 0.2 | 58.8 | 35.8 |
| | - | ✓ | - | - | ✓ | 59.7 | 1.9 | 0.1 | 38.3 | 0.2 | - | 86.1 | 13.7 | 3.4 | 0.3 | 61.5 | 34.8 |
| 5000 | ✓ | - | - | ✓ | - | 95.3 | 4.7 | - | - | 94.8 | 2.4 | - | 2.8 | 94.2 | 5.8 | - | - |
| | ✓ | - | - | - | ✓ | 94.7 | 3.5 | - | 1.8 | 94.9 | 3 | - | 2.1 | 95.1 | 4.9 | - | - |
| | - | ✓ | - | ✓ | - | 83.9 | 2.6 | - | 13.5 | 0.6 | - | 84 | 15.4 | 49.3 | 1.3 | 8.1 | 41.3 |
| | - | ✓ | - | - | ✓ | 81.7 | 2.6 | - | 15.7 | 0.3 | - | 85 | 14.7 | 46.6 | 1.6 | 9.8 | 42 |
| 10000 | ✓ | - | - | ✓ | - | 95.8 | 4.2 | - | - | 95.1 | 2.3 | - | 2.6 | 94.9 | 5.1 | - | - |
| | ✓ | - | - | - | ✓ | 95.5 | 4.5 | - | - | 94.5 | 2.9 | - | 2.6 | 93.3 | 6.7 | - | - |
| | - | ✓ | - | ✓ | - | 90.8 | 2.4 | - | 6.8 | 0.8 | - | 82.9 | 16.3 | 85.7 | 4.1 | 0.1 | 10.1 |
| | - | ✓ | - | - | ✓ | 89.3 | 3.1 | - | 7.6 | 0.3 | - | 84.9 | 14.8 | 84 | 3.6 | 0.3 | 12.1 |

Table 4.6: **V-structure simulation results for Model 7** and **8** similar to Table 4.6.

theory. Indeed, the behavior of the model 8 is similar to Model 7.

If the initial conditions have not the same values or the same standard deviation, Model 8 behaves as Model 4.

## 4.3 Λ-Structure topology

In Fig 4.9 the ground-truth causal graph defines the following relations: $dep(R, pA|\emptyset)$, $dep(R, pB|\emptyset)$, $dep(pA, pB|\emptyset)$, $dep(R, pA|pB)$, $dep(R, pB|pA)$ and $indep(pA, pB|R)$. We notice that Λ-structure has the same causal characteristics as the cascade topology and hence it holds the same causal arguments.

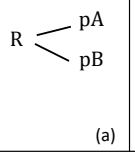
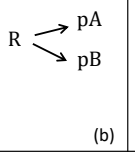
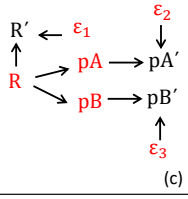
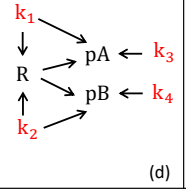
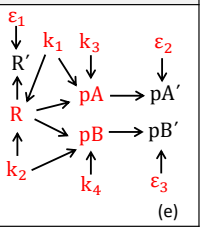
| Topology | Class of MEG | Expected DAG with | | | |
|---|---|---|---|---|---|
| | | No parameters (model 1&5) | Meas. Noise (model 2&6) | Rand. Coef. (model 3&7) | Meas. Noise &Rand. Coef. (model 4&8) |
| Λ-structure |   |   |   |   |   |

Figure 4.9: **Λ-structure topology** similar to Fig. 4.1.

### Model 1

As in previous topologies, causal discovery algorithms are not applicable for this model due to the lack of stochasticity.

## Model 2

The ground-truth causal graph of Model 2 is presented in Fig. 4.9(c). As in the cascade topology (see Section 4.1), we observe that while $pA$',$pB'$ are indeed d-separated by $R$ (a structural quality) the dependency relations are defined statistically, and the relation $indep(pA, pB|R)$ does not hold in this model due to violation of faithfulness assumption. Table 4.7 shows the simulation results which are similar to cascade topology. For the constant case PC algorithm returns fully-unconnected MEGs while for the random case in most of the times fully-connected.

| Samples | Exp. Des | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_\varepsilon$ | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | - | - | 85.7 | 14.3 | 4.5 | 44.6 | - | 50.9 | 5.3 | - | 51.4 | 43.3 |
| | - | ✓ | - | - | 85.6 | 14.4 | 0.5 | 0.9 | 11 | 87.6 | 4.8 | 47.5 | - | 47.6 |
| 5000 | ✓ | - | - | - | 85.8 | 14.2 | 4.2 | 64.3 | - | 31.5 | 4.4 | 68.7 | - | 26.9 |
| | - | ✓ | - | - | 85.8 | 14.2 | 2.5 | 18.3 | 0.7 | 78.5 | 4.3 | 66.2 | - | 29.5 |
| 10000 | ✓ | - | - | - | 85.8 | 14.2 | 3.9 | 72.1 | - | 24 | 3.9 | 76.1 | - | 20 |
| | - | ✓ | - | - | 85.7 | 14.3 | 2.4 | 38.9 | - | 58.7 | 4 | 73.7 | - | 22.3 |

Table 4.7: **Λ-structure simulation results for Model 2** similar to Table 4.1

## Model 3

Fig. 4.9(d) represents the underlying causal graph for Model 3, in which the rate constants $k_1$ and $k_2$ are latent confounders violating causal sufficiency assumption. Similar to the cascade topology, $indep(pA, pB|R)$ cannot hold because there is an inducing path from variable $pA$ to variable $pB$ created from the rate constants $k_1$ & $k_2$ and the measured variable $R$ when we condition on $R$. Following the same approach as in Model 3 of other topologies, we provide the simulation results in Table 4.8. For fixed $X_0$ and both levels of uncertainty ($\sigma_k = 0.1$ or $0.5$) in rate constants PC algorithm returns irrelevant to the desired MEG. In contrast with the other topologies, here, there is no cancellation of the effect of latent confounders since variables $pA$ and $pB$ depend on the sum of $k_1$ and $k_2$. That is, at steady state holds:

$$R = \frac{X_0}{k_1 + k_2}$$

then

$$pA = \frac{k_1 R}{k_3}$$

However,

$$pA = \frac{k_1 \frac{X_0}{k_1+k_2}}{k_2}$$

, that is $pA$ depends both on $k_1$ and $k_2$. We believe that the lack of stochasticity results on the incapacity of desired MEGs reconstruction.

Also, when the initial condition is random the simulation results are similar to cascade topology. The only exception is the reconstruction of the desired MEG

when the uncertainty in $X_0$ is high and low in the kinetic rates. Probably, this is a coincidence since for larger sample size fully-connected MEG will be inferred.

| Samples | Exp. Des | | | | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_k$ | | $\sigma_\varepsilon$ | | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 0.1 | 0.5 | 0 | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | ✓ | - | - | - | - | - | 100 | - | 91 | - | 9 | 86 | 14 | - | - |
| | - | ✓ | ✓ | - | - | - | 0.1 | 10.3 | 89.6 | - | 0.2 | 13.1 | 86.7 | 0.1 | 8.4 | 5.9 | 85.6 |
| 5000 | ✓ | - | ✓ | - | - | - | - | - | 100 | - | 100 | - | - | 77.2 | 22.8 | - | - |
| | - | ✓ | ✓ | - | - | - | - | 0.1 | 99.9 | - | 0.5 | 0.1 | 99.4 | 0.1 | 75.2 | - | 24.7 |
| 10000 | ✓ | - | ✓ | - | - | - | - | - | 100 | - | 100 | - | - | 65.7 | 34.3 | - | - |
| | - | ✓ | ✓ | - | - | - | 0.1 | - | 99.9 | - | 0.8 | - | 99.2 | - | 94.5 | - | 5.5 |
| 1000 | ✓ | - | - | ✓ | - | - | 0.1 | - | 99.9 | - | 89 | - | 11 | 89.8 | 10.2 | - | - |
| | ✓ | - | - | - | ✓ | - | 0.1 | 0.2 | 99.7 | - | 77.7 | - | 22.3 | 27.5 | 72.5 | - | - |
| | - | ✓ | - | ✓ | - | - | 0.1 | 1.8 | 98.1 | - | 0.2 | 13.6 | 86.2 | - | 6.8 | 4.3 | 88.9 |
| | - | ✓ | - | - | ✓ | - | 0.1 | 12.1 | 87.8 | - | 0.2 | 14.7 | 85.1 | - | 6.8 | 5.5 | 87.7 |
| 5000 | ✓ | - | - | ✓ | - | - | - | - | 100 | - | 100 | - | - | 90.1 | 9.9 | - | - |
| | ✓ | - | - | - | ✓ | - | 0.1 | - | 99.9 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | - | 0.2 | - | 99.8 | - | 0.4 | 0.2 | 99.4 | - | 75 | - | 25 |
| | - | ✓ | - | - | ✓ | - | - | 0.1 | 99.9 | - | 0.3 | - | 99.7 | 0.1 | 74.1 | - | 25.8 |
| 10000 | ✓ | - | - | ✓ | - | - | - | - | 100 | - | 100 | - | - | 89.7 | 10.3 | - | - |
| | ✓ | - | - | - | ✓ | - | 0.1 | - | 99.9 | - | 100 | - | - | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | - | - | - | 100 | - | 0.8 | - | 99.2 | - | 95.3 | - | 4.7 |
| | - | ✓ | - | - | ✓ | - | - | - | 100 | - | 1.3 | - | 98.7 | - | 95.2 | - | 4.8 |

Table 4.8: **Λ-structure simulation results for Model 3** and **4** similar to Table 4.2

## Model 4

For Model 4, the ground-truth causal graph is shown in Fig. 4.9(e). We expect that the simulation results will be similar to Model 3 according to causal theory. Table 4.8 verifies the above discussion.

## Model 5

The concentration matrix of both constant and random initial condition reveals that all its elements take always non-zero values. That means : $dep(R, pA|pB)$, $dep(R, pB|pA)$ and $dep(pA, pB|R)$. The elements $\Theta 5(2,3)$ and $\bar{\Theta}_5(2,3)$ are respectively:

$$\Theta_5 = \frac{-(4k_1k_2k_3k_4)(k_3+k_4)(k_1+k_2+k_3)(k_1+k_2+k_4)}{\sigma^2(2k_1^4k_3^2+\cdots+k_3^2k_4^4)}$$

and

$$\bar{\Theta}_5 = -(\sigma_{X0}^2(k_1k_3+k_1k_4+k_2k_3+k_2k_4+2k_3k_4)+\sigma^2(k_1k_3k_4+k_2k_3k_4))\cdot \frac{(4k_1k_2k_3k_4)(k_3+k_4)(k_1+k_2+k_3)(k_1+k_2+k_4)}{k_1k_3\sigma_{X0}^2+\cdots+2k_3^3k_4^5\sigma_{X0}^2}$$

Similar to Model 5 of Sections 4.1& 4.2 we define, again, Euler-Marayuma SEM with the following:

1. The initial or primary variables $R_0$, $pA_0$ and $pB_0$, and,
2. The functional relationships are given by:

$$R_{N+1} = R_N - ((k_1+k_2)R_N)\Delta t + \sigma\Delta W_N$$

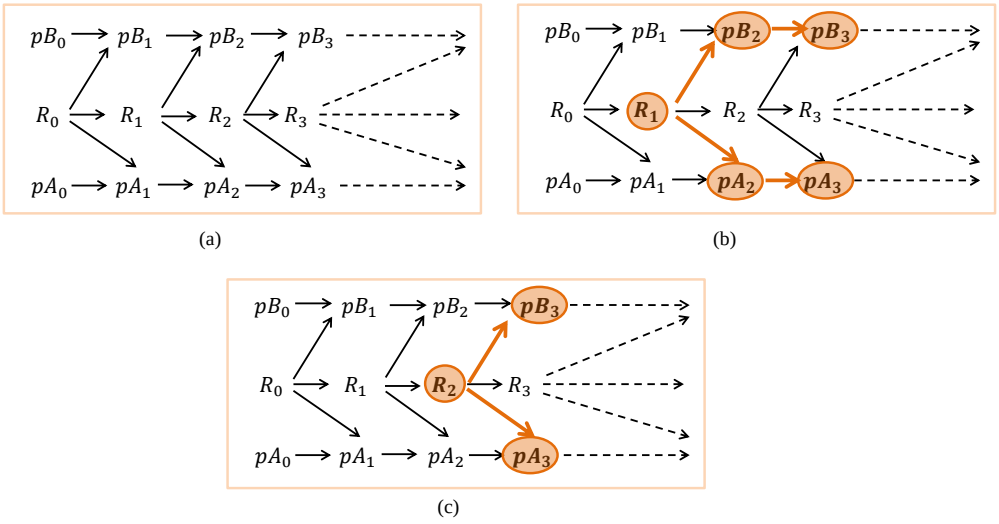



Figure 4.10: (a)Unraveled Bayesian Network of the Euler-Marayuma SEM of Model 5 for Λ-structure topology. Orange nodes and arcs show that information flows via a path from variable $pA_3$ to $pB_3$ condition on (b) empty set and (c) $R_3$, respectively. We note that these paths may be not unique for each condition set; each state is an observation in equilibrium while the variables of Brownian motion $W_{N+1} - W_N$ are not depicted on the graphs since they are independent and identical distributed. The conditional dependency relations are generalized for $N + 1$ variables.

$$pA_{N+1} = pA_N + (k_1 R_N - k_3 pA_N)\Delta t + \sigma \Delta W'_N$$
$$pB_{N+1} = pB_N + (k_2 R_N - k_4 pB_N)\Delta t + \sigma \Delta W''_N$$

We expect that the unraveled Bayesian network will show the same dependencies as in Model 5 of cascade topology. Fig. 4.10 indicates that the relationships $dep(pA_{N+1}, pB_{N+1}|\emptyset)$ and $dep(pA_{N+1}, pB_{N+1}|R_{N+1})$ indeed hold since there are paths that allows the information to flow. Therefore, any constraint-based algorithm is not able to reenact the desired MEG ,at least for infinite sample size and observational data. To ascertain if these theoretical findings are also well grounded, we numerically test, as before, the PC algorithm on generated data for several sample sizes. We present the statistical discovery diagrams in Fig. 4.11. We observe that, for both cases of initial condition, when the sample size is small the statistical error is increased and the obtained MEG is irrelevant to the desired one. For larger sample size PC algorithm tends to find fully-connected MEGs.

It is interestingly that when $k_2 = 10^{-2}$ there is a region where the desired MEG can be discovered. However, we can see that as the sample size increases this region tends to be smaller. We believe that that this behavior is probably an artifact of the utilized sample size.

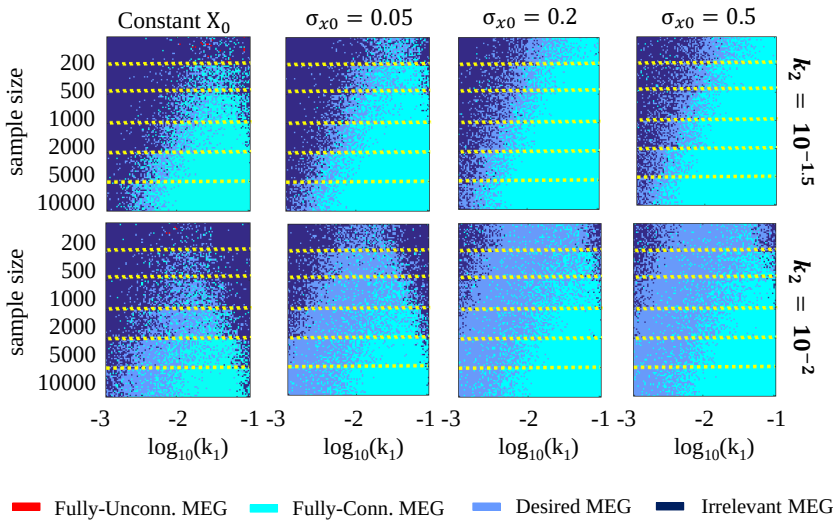



Figure 4.11: Λ-structure's statistical discovery diagrams of Model 5, similar to Fig. 4.7.

### Model 6

In Model 6, the addition of measurement noise does not change the theoretical findings of Model 5. For the shake of completeness, we perform empirical evaluation which is shown in Figs. 4.12 for two different values for $\sigma_\varepsilon$, for both initial conditions and various values for the sample size. Comparing with the respective results of Model 5 (Section 4.3) we observe that the effect of measurement noise results in reducing the blue area where the fully-connected MEG is found by PC algorithm. In case of constant initial condition, in small sample size, the statistical error is more substantial since PC returns fully-independent MEGs showing that the characteristics of the independent measurement noise dominate.

### Model 7

Similar to Model 7 of the Cascade and V-structure topologies, we validate PC algorithm on generated data for different sample sizes. Table 4.9 presents the simulation results for two different value of the standard deviation of $k_i$. In all cases of initial condition the obtained MEG is fully-connected consistent with the causal theory. Only for $\sigma_{x0} = 0.05$ and high uncertainty in the kinetic rates PC algorithm finds irrelevant to the desired MEGs.

### Model 8

Similar to Model 8 of the previous topologies, we examine the performance of PC algorithm on generated data. The simulation results in Table 4.9 are consistent

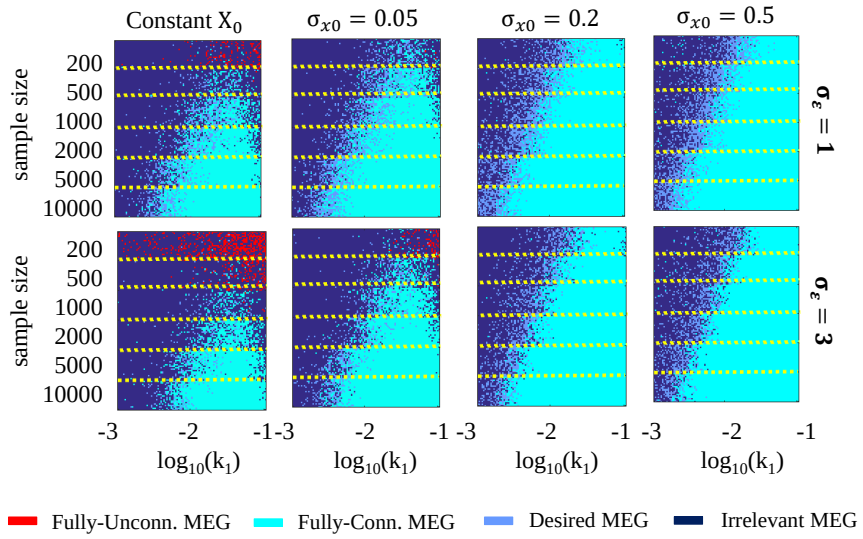



Figure 4.12: Λ-structure's statistical discovery diagrams of Model 6 for both for constant and random $X_0$ similar to Fig. 4.4.

| Samples | Exp. Des | | | | | Constant $X_0$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.05$ | | | | Random $X_0$ with $\sigma_{X_0} = 0.2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\sigma_k$ | | $\sigma_\varepsilon$ | | | Desired | Conn. | Uncon. | Irrelevant | Desired | Conn. | Uncon. | Irrelevant | Desired | Con. | Uncon. | Irrelevant |
| | 0.1 | 0.5 | 0 | 1 | 3 | | | | | | | | | | | | |
| 1000 | ✓ | - | ✓ | - | - | 13.3 | 85.8 | - | 0.9 | - | 23.5 | - | 76.5 | 76 | 24 | - | - |
| | - | ✓ | ✓ | - | - | 11 | 22.4 | - | 66.6 | - | - | 14.4 | 85.6 | - | 4.7 | 3.6 | 91.7 |
| 5000 | ✓ | - | ✓ | - | - | - | 100 | - | - | - | 78.4 | - | 21.6 | 26.2 | 73.8 | - | - |
| | - | ✓ | ✓ | - | - | 23.8 | 43.7 | - | 32.5 | - | 0.2 | - | 99.8 | - | 60.3 | - | 39.7 |
| 10000 | ✓ | - | ✓ | - | - | - | 100 | - | - | - | 97.4 | - | 2.6 | 6.9 | 93.1 | - | - |
| | - | ✓ | ✓ | - | - | 29.6 | 50.6 | - | 19.8 | - | 1.9 | - | 98.1 | - | 88.6 | - | 11.4 |
| 1000 | ✓ | - | - | ✓ | - | 13.9 | 84 | - | 2.1 | 0.1 | 19.2 | - | 80.7 | 66.8 | 33.2 | - | - |
| | ✓ | - | - | - | ✓ | 10.5 | 50.4 | 0.1 | 39 | 0.1 | 18 | - | 81.9 | 11.5 | 88.5 | - | - |
| | - | ✓ | - | ✓ | - | 9.9 | 20.4 | - | 69.7 | - | 0.2 | 15.8 | 84 | 0.1 | 4.5 | 4.6 | 90.8 |
| | - | ✓ | - | - | ✓ | 3.5 | 11.1 | 1 | 84.4 | - | 0.2 | 15.2 | 84.6 | 0.1 | 5.8 | 3.9 | 90.2 |
| 5000 | ✓ | - | - | ✓ | - | - | 100 | - | - | - | 73.7 | - | 26.3 | 10 | 90 | - | - |
| | ✓ | - | - | - | ✓ | 0.1 | 99.9 | - | - | - | 66.6 | - | 33.4 | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | 24.5 | 41.5 | - | 30.5 | - | 0.6 | 0.3 | 99.1 | - | 63.1 | - | 36.9 |
| | - | ✓ | - | - | ✓ | 13.9 | 45.3 | - | 40.8 | - | 0.7 | 0.1 | 99.2 | - | 63.1 | - | 36.9 |
| 10000 | ✓ | - | - | ✓ | - | - | 100 | - | - | - | 97.5 | - | 2.5 | 0.5 | 99.5 | - | - |
| | ✓ | - | - | - | ✓ | - | 100 | - | - | - | 91.2 | - | 8.8 | - | 100 | - | - |
| | - | ✓ | - | ✓ | - | 29.2 | 51.7 | - | 19.1 | - | 1.3 | - | 98.7 | 0.1 | 89.4 | - | 10.5 |
| | - | ✓ | - | - | ✓ | 17.3 | 57 | - | 25.7 | - | 1.3 | - | 98.7 | - | 92 | - | 8 |

Table 4.9: **Λ-structure simulation results for Model 7** and **8** similar to Table 4.3.

with Model 7. The obtained MEG most of the times is fully-connected, for both fixed and random $X_0$. In any case the desired MEG is never inferred.

# Chapter 5

# Generalization to larger reaction networks

In Section 4 we used symbolic calculations and compute the covariance and the concentration matrices of a reaction network with three species. The same calculations can be conducted with four, five or more species and check whether or not the properties found for instance in the three species network generalize to systems with similar structure but more species. Unfortunately, the required computational time increases exponentially with the size of the system making it impossible to obtain a timely outcome. Therefore, it is important to develop a mathematical theory that provides rigorous insights on how the properties of a reaction network propagate as we increase its size. The following theorem[1] for the cascade topology is shown in Fig. 5.1 which suggests that $\Theta_3(1,3) \propto (2k_2^2 - k_3^2)$ generalizes to $\Theta_n(1,j) \propto (2k_2^2 - k_3^2)$ for $n > 3$ and $j = 3, ..., n$ under the assumption that the reaction constants $k_2$ and $k_3$ remain the same across the reaction networks.

**Implications.** An implication of this theorem is that given an (in-)dependence relation between some species, similar relations can be extracted for larger or smaller reaction networks that admit the same topology structure. For instance, assuming that for a three species cascade network the property $X_1 \perp X_3|X_2$ holds then for a five species cascade network it holds that $X_1 \perp X_3|X_2, X_4, X_5$ as well as $X_1 \perp X_4|X_2, X_3, X_5$ and $X_1 \perp X_5|X_2, X_3, X_4$. It is also note worthy that the constructive proof provides an iterative algorithm on how to compute both the covariance and the concentration matrices of size $n$ from the respective matrices of size $n-1$. This algorithm has polynomial-time complexity and it is applicable not only for cascade topology but also for general topologies.

**Theorem 1** *Let $\Theta_n$ be the concentration matrix of an $n$-species cascade topology shown in Figure 5.1. Then, it holds for any $n \geq 3$ that*

$$\Theta_n(1,j) \propto (2k_2^2 - k_3^2) \quad \textit{for all} \quad j = 3, ..., n \ . \tag{5.1}$$

[1]The theorem and the proof is a result of Dr. Y.Pantazis work

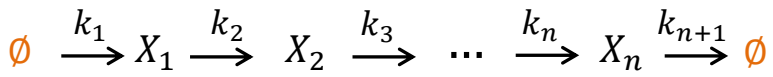



Figure 5.1: Cascade reaction network topology with $n$ species.

The proof is based on *induction*. First, we will prove that $\Theta_3(1,3) \propto (2k_2^2 - k_3^2)$. Then, we will assume that $\Theta_i(1,j) \propto (2k_2^2 - k_3^2)$ for $j = 3, \cdots, i$ and $i = 3, \cdots, n-1$ and prove that $\Theta_n(1,j) \propto (2k_2^2 - k_3^2)$ for $j = 3, \cdots, n$.
The connectivity matrix of the $n$-species cascade network is given by

$$A_n = \begin{bmatrix} -k_2 & 0 & 0 & \cdots & 0 & 0 \\ k_2 & -k_3 & 0 & \cdots & 0 & 0 \\ 0 & k_3 & -k_4 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & k_n & -k_{n+1} \end{bmatrix} = \begin{bmatrix} A_{n-1} & 0_{n-1} \\ c_{n-1}^T & -k_{n+1} \end{bmatrix} \tag{5.2}$$

where $0_n$ is a column vector with $n$ zero elements while

$$c_{n-1} = \begin{bmatrix} 0_{n-2} & k_n \end{bmatrix}^T \tag{5.3}$$

Notice that the breaking of the connectivity matrix into a $2\times2$ block with the upper left block being the connectivity matrix of the $(n-1)$-species cascade network is critical for the proof. Consequently, the covariance matrix can be written as

$$\Sigma_n = \begin{bmatrix} \bar{\Sigma}_{n-1} & \sigma_{.n} \\ \sigma_{.n}^T & \sigma_n^2 \end{bmatrix} \tag{5.4}$$

The Lyapunov equation states that

$$A_n \Sigma_n + \Sigma_n A_n^T = -\sigma^2 I_n \tag{5.5}$$

therefore, substituting the respective $2 \times 2$ block matrices, we obtain

$$\begin{bmatrix} A_{n-1} & 0_{n-1} \\ c_{n-1}^T & -k_{n+1} \end{bmatrix} \begin{bmatrix} \bar{\Sigma}_{n-1} & \sigma_{.n} \\ \sigma_{.n}^T & \sigma_n^2 \end{bmatrix} + \begin{bmatrix} \bar{\Sigma}_{n-1} & \sigma_{.n} \\ \sigma_{.n}^T & \sigma_n^2 \end{bmatrix} \begin{bmatrix} A_{n-1} & 0_{n-1} \\ c_{n-1}^T & -k_{n+1} \end{bmatrix} = -\sigma^2 I_n \tag{5.6}$$

Hence, we get that

$$\begin{cases} \bar{\Sigma}_{n-1} = \Sigma_{n-1} \\ \sigma_{.n} = -(A_{n-1} - k_{n+1} I_{n-1})^{-1} \Sigma_{n-1} c_{n-1} \\ \sigma_n^2 = \dfrac{\sigma^2 + 2 c_{n-1}^T \sigma_{.n}}{2k_{n+1}} \end{cases} \tag{5.7}$$

Using the inversion formula for a $2\times 2$ block matrix, we compute the concentration matrix as

$$\Sigma_n^{-1} = \Theta_n = \begin{bmatrix} \Theta_{n-1} + \frac{1}{w_n} e_{n-1} e_{n-1}^T & -\frac{1}{w_n} e_{n-1} \\ -\frac{1}{w_n} e_{n-1} & \frac{1}{w_n} \end{bmatrix} \tag{5.8}$$

where the column vector $e_{n-1}$ is given by

$$e_{n-1} = \Theta_{n-1}\sigma_{.n} = \left(A_{n-1} + \sigma^2\Theta_{n-1} + k_{n+1}I_{n-1}\right)^{-1} c_{n-1} \tag{5.9}$$

while

$$w_n = \sigma_n^2 - \sigma_{.n}^T e_{n-1} \tag{5.10}$$

is a real number. We remark here that we could devise an iterative algorithm using the equations (5.7)–(5.10) to compute the covariance and concentration matrices that contains one additional species in the reaction chain. Additionally, the following representation of $e_{n-1}$ will be proven useful in the induction step. Utilizing again the Lyapunov equation it holds that

$$\begin{aligned} e_{n-1} &= \Theta_{n-1}(-A_{n-1} + k_{n+1}I_{n-1})^{-1}\Sigma_{n-1}c_{n-1} \\ &= (\Theta_{n-1}(-A_{n-1})\Sigma_{n-1} + k_{n+1}\Theta_{n-1}\Sigma_{n-1})^{-1}c_{n-1} \\ &= \left(A_{n-1} + \sigma^2\Theta_{n-1} + k_{n+1}I_{n-1}\right)^{-1} c_{n-1} \end{aligned} \tag{5.11}$$

**'n = 3' case.** In order to show that $\Theta_3(1,3) \propto (2k_2^2 - k_3^2)$, it suffices from (5.8) to show that $e_2 \equiv e_2(1) \propto (2k_2^2 - k_3^2)$. Proceeding, a straightforward computation of $\Sigma_2$ is obtained by solving the Lyapunov equation. It yields that

$$\Sigma_2 = \begin{bmatrix} \frac{\sigma^2}{2k_2} & \frac{\sigma^2}{2(k_2+k_3)} \\ \frac{\sigma^2}{2(k_2+k_3)} & \frac{\sigma^2(2k_2+k_3)}{2k_2(k_2+k_3)} \end{bmatrix} \tag{5.12}$$

Then, the respective concentration matrix is

$$\Theta_2 = \frac{2(k_2+k_3)}{\sigma^2(2k_2^2+2k_2k_3+k_3^2)} \begin{bmatrix} k_2(2k_2+k_3) & -k_2k_3 \\ -k_2k_3 & k_3(k_2+k_3) \end{bmatrix} \tag{5.13}$$

From eq. (5.11), we compute

$$\begin{aligned} e_2(1) &= \frac{-k_3}{det(A_2^T + \sigma^2\Theta_2 + k_4I_2)} \left(A_{n-1}(2,1) + \sigma^2\Theta_2(1,2)\right) \\ &\propto \left(k_2 - \frac{2k_2k_3(k_2+k_3)}{2k_2^2+2k_2k_3+k_3^2}\right) \\ &\propto (2k_2^2 - k_3^2) \end{aligned} \tag{5.14}$$

**Induction step.** First, we observe that

$$\Theta_n(1,n) = -\frac{1}{w_n} e_{n-1}(1) \quad \propto \quad e_{n-1}(1)$$

as well as that

$$\Theta_n(1,j) = \Theta_{n-1}(1,j) + \frac{1}{w_n} e_{n-1}(1) e_{n-1}(j), \quad \text{for all } j = 1, \cdots, n-1.$$

Then, noting from the induction assumption that $\Theta_{n-1}(1,j) \propto (2k_2^2 - k_3^2)$, for $j = 3, \cdots, n-1$, it suffices to prove that

$$e_{n-1}(1) \propto (2k_2^2 - k_3^2)$$

and the proof is completed.
We proceed with eq. (5.11) and define the auxiliary matrices

$$B_{n-1} = A_{n-1}^T + \sigma^2 \Theta_{n-1} + k_{n+1} I_{n-1}$$

as well its inverse

$$D_{n-1} = B_{n-1}^{-1} \ .$$

Due to the special structure of $c_{n-1}$ (see eq. (5.3)), it holds that

$$e_{n-1}(1) = k_n D_{n-1}(1, n-1)$$

Thus, it suffices to prove that

$$D_{n-1}(1, n-1) \propto (2k_2^2 - k_3^2) \ .$$

Using minors and the cofactor matrix of a matrix, we have

$$D_{n-1}(1, n-1) = (-1)^n M_{n-1}(n-1, 1) \tag{5.15}$$

where $M_{n-1}(n-1,1)$ is the $(n-1,\ 1)$ minor of $B_{n-1}$ which is the determinant of a $(n-2) \times (n-2)$ matrix that results from deleting the $(n-1)$-th row and the first column of $B_{n-1}$. The $(n-1,\ 1)$ minor of $B_{n-1}$ is rewritten as

$$\begin{aligned} M_{n-1}(n-1,1) &= \begin{vmatrix} B_{n-1}(1,2) & \cdots & B_{n-1}(1,n-1) \\ \vdots & \vdots & \vdots \\ B_{n-1}(1,n-2) & \cdots & B_{n-1}(n-2,n-1) \end{vmatrix} \\ &= \sum_{j=2}^{n-1} (-1)^j B_{n-1}(1,j) M'_{n-1}(1,j) \end{aligned}$$

where $M'_{n-1}(1,j)$ are the respective minors. If we prove that

$$B_{n-1}(1,j) \propto \ (2k_2^2 - k_3^2), \quad \text{for all } j = 2, \cdots, n-1$$

then

$$M_{n-1}(1) \propto \ (2k_2^2 - k_3^2)$$

and thus the proof is completed.
First observe that for $j = 3, \cdots\ n-1$ it holds that

$$B_{n-1}(1,j) = \sigma^2 \Theta_{n-1}(1,j) \ \propto (2k_2^2 - k_3^2)$$

by the induction assumption. It only remains the element

$$B_{n-1}(1,2) = A_{n-1}(2,1) + \sigma^2\Theta_{n-1}(1,2) = k_2 + \sigma^2\Theta_{n-1}(1,2)$$

By iteratively solving for $\Theta_{n-1}(1,2)$, we get that

$$\begin{aligned}\Theta_{n-1}(1,2) &= \Theta_{n-2}(2,1) + \frac{1}{w_{n-1}} e_{n-2}(1)e_{n-2}(2) \\ &= \Theta_2(1,2) + \sum_{j=3}^{n-1} \frac{1}{w_j} e_{j-1}(1)e_{j-1}(2) \\ &= \Theta_2(1,2) - \sum_{j=3}^{n-1} \Theta_j(1,j)e_{j-1}(2)\end{aligned}$$

Overall,

$$B_{n-1}(1,2) = k_2 + \sigma^2\Theta_2(1,2) - \sigma^2 \sum_{j=3}^{n-1} \Theta_j(1,j)e_{j-1}(2)$$

From the $n = 3$ case we know that $k_2 + \sigma^2\Theta_2(1,2)$ is proportional to $(2k_2^2 - k_3^2)$. Additionally, the sum series is also proportional to $(2k_2^2 - k_3^2)$ by the induction assumption. Therefore,

$$B_{n-1}(1,2) \propto (2k_2^2 - k_3^2)$$

and the proof is completed.

# Chapter 6

# Conclusion

The huge benefit from developing algorithms that are able to induce the causal structure of complex biological systems is undoubted. However, the applicability of already developed causal discovery algorithms in biology data has been relatively poor. This is mainly because of the conditions in the biological setting that generates data that violate the stringent causal assumptions rendering the extraction of cause and effect from statistical models notoriously difficult. In the present we make an effort to explore a wide range of linear mechanistic models of signaling pathways and assess, both theoretically and numerically, the ability of learning algorithms to identify the correct causal structure of several topologies. For two different cases of the initial condition of the system, we examine different ODE and SDE models, at equilibrium, without interventions and with various parameters and model specifications.

A distinction between ODE and SDE models draws a conclusion that any form of stochasticity either from the initial condition or from the mathematical formalism itself eliminates the determinism and the obtained structures tend to be the expected according to causal theory.

Observing the results globally, we realize that in most models that we studied, the causal assumptions are violated and the discovery of the desired MEG using a constraint-based algorithm (in our case PC algorithm) is a difficult task. Although in some cases the causal assumptions are not satisfied, we notice that if the structural properties of the topology (e.g v-structure) hold, the desired MEG can be reenacted.

Also, exploring the analytical solution of the mathematical formalism of some models (SDE models) we revealed the relations under which the desired structure will be confidently discovered. Specifically, for the cascade topology, there are limited conditions that meet the expectations for causal reconstruction (these conditions can be propagated in larger identical topology structures) while for $\Lambda$-structure there is no condition for the reenactment of the desired MEG. In contrast, for the v-structure topology, it is guaranteed that a constraint-based algorithm will learn the underlying causal structure. In any case the reconstruction of a causal

graph have to fulfill rigorous and specific conditions.

These results illustrate the need for the systematic study of the effect of violations of causal assumptions to existing algorithms and novel causal discovery methods that focus on learning latent variables or confounders based on the proposed mathematical formalism. Furthermore, these results indicate the inability of learning a causal structure de-novo from observational data without interventions. Particular importance should be attached to the concept of local independence. Intuitively, we understand that the notion of this independence is fundamental since the relationships between the stochastic processes is asymmetric making the CBNs perhaps not the suitable approach of causal inference. However, there is no well-defined framework for causal learning based on the local independence. Hence, the need for design and development of local independence tests and learning algorithms is imperative.

In the future we plan to extend this work to a wider range of topologies (e.g. mixed, cyclic) where nonlinear interactions may occur and include the notion of local independence. Our ultimate goal is to provide detailed guidelines for biologists that will allow them to employ powerful structure learning algorithms in pursue of recovering the landscape of interactions in any biological system with minimal experimental efforts.

# Bibliography


[1] B. Aldridge, J. Burke, D. Lauffenburger, and P. Sorger. Physicochemical modelling of cell signalling pathways. *Nature Cell Biology*, 8:1195 – 1203, 2006.

[2] S. Bongers and J. Mooij. From random differential equations to structural causal models: the stochastic case. *arXiv:1803.08784*, 2018.

[3] M. Elowitz, A. Levine, E. Siggia, and P. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183 – 1186, 2002.

[4] Gegout-Petit.A and Commenges.D. A general definition of influence between stochastic processes. *Lifetime Data Analysis*, 16:33 – 44, 2010.

[5] S. Hill, L. Heiser, T. Cokelaer, and et al. Stochastic gene expression in a single cell. *Nature Methods*, 13:310 – 318, 2016.

[6] M. Kanehisa and S. Goto. Kegg: Kyoto encyclopaedia of genes and genomes. *Nucleic Acids Res*, 28(1):27 – 30, 2000.

[7] BN. Kholodenko. Cell-signalling dynamics in time and space. *Nature Reviews Molecular Cell Biology*, 7:165 – 176., 2006.

[8] H. Kitano. Systems biology: A brief overview. *Science*, 295:1662 – 1664., 2002.

[9] V. Lagani, S. Triantafillou, G. Ball, J. Tegnér, and I. Tsamardinos. *Probabilistic Computational Causal Discovery for Systems Biology*, pages 33–73. Springer International Publishing, 2016.

[10] Z. Li, P. Li, and A. Krishnan J. Liu. Large-scale dynamic gene regulatory network inference combining differential equation models with local dynamic bayesian network analysis. *Bioinformatics*, 27:2686 – 91, 2011.

[11] N. Meinshausen, A. Hauser, J. Mooij, J. Peters, P. Versteeg, and P. Bohlmann. Methods for causal inference from gene perturbation experiments and validation. *Proceedings of the National Academy of Sciences*, 113:7361 – 7368, 2016.

[12] Mogensen.S, , Malinsky.D, and Hansen.N. Causal learning for partially observed stochastic dynamical systems. *UAI*, 2018.

[13] J. Mooij, D. Janzing, and B. Schoelkopf. From ordinary differential equations to structural causal models: the deterministic case. *arXiv:1408.2063*, 2014.

[14] C. Oates, F. Dondelinger, N.Bayaniand J. Korkola, J. Gray, and S. Mukherjee. Causal network inference using biochemical kinetics. *Bioinformatics*, 30:i468–i474, 2014.

[15] Y. Pantazis and I. Tsamardinos. A unified approach for sparse dynamical system inference from temporal measurements. *arXiv preprint arXiv:1710.00718*, 2017.

[16] G. Pavliotis. *Stochastic Processes and Applications.* Springer, New York, 2014.

[17] J. Pearl. *Causality: models, reasoning and inference.* Cambridge University Press, New York, 2000.

[18] R. Prill, R. Vogel, G. Cecchi, G. Altan-Bonnet, and G. Stolovitzky. Noise-driven causal inference in biomolecular networks. *PLoS One*, 10:1 – 16, 2015.

[19] P. Rosenbaum. *Observational study*, pages 1451–1462. Wiley, Hoboken, N.J., 2005.

[20] K. Sachs, O. Perez, D. Pe'er, D. Lauffenburger, and G. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308:523 – 529, 2005.

[21] R. Scheines and J. Ramsey. Measurement error and causal discovery. *CEUR Workshop Proc.*, 1792:1 – 7, 2016.

[22] B. Schoeberl, C. Eichler-Jonsson, E. Gilles, and G. Muller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nature Biotechnology*, 20:370 – 5, 2002.

[23] M. Scutari and K. Strimmer. *Introduction to graphical modelling. In, Balding*, pages 1 – 21. Wiley, Chichester, UK, 2011.

[24] Sokol.A and Hansen.N. Causal interpretation of stochastic differential equations. *Electron. J. Probab.*, 19:34, 2014.

[25] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, New York, 2001.

[26] Spirtes.P, Glymour.C, and Scheines.R. Constructing bayesian network models of gene expression networks from microarray data. *In Proc. of the Atlantic Symposium on Computational Biology, Genome Information Systems & Technology*, 2000.

[27] M. Stumpf, D. Balding, and M. Girolami. *Handbook of Statistical Systems Biology*. Wiley, Chichester, UK, 2011.

[28] S. Triantafillou, V. Lagani, C. Heinze-Deml, A. Schmidt, J. Tegner, and I. Tsamardinos. Predicting causal relationships from biological data: Applying automated causal discovery on mass cytometry data of human immune cells. *Scientific Reports*, 7:12724, 2017.

[29] D. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10:122 – 33, 2009.

[30] TR. Xu, V. Vyshemirsky, A. Gormand, A. von Kriegsheim, M. Girolami, GS. Baillie, D. Ketley, AJ. Dunlop, G. Milligan, MD. Houslay, and W. Kolch. Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci Signal*, 3:ra20, 2010.

[31] K. Zhang, M. Gong, J. Ramsey, K. Batmanghelich, P. Spirtes, and C. Glymour. ausal discovery in the presence of measurement error: Identifiability conditions. *arXiv:1706.03768*, 2017.