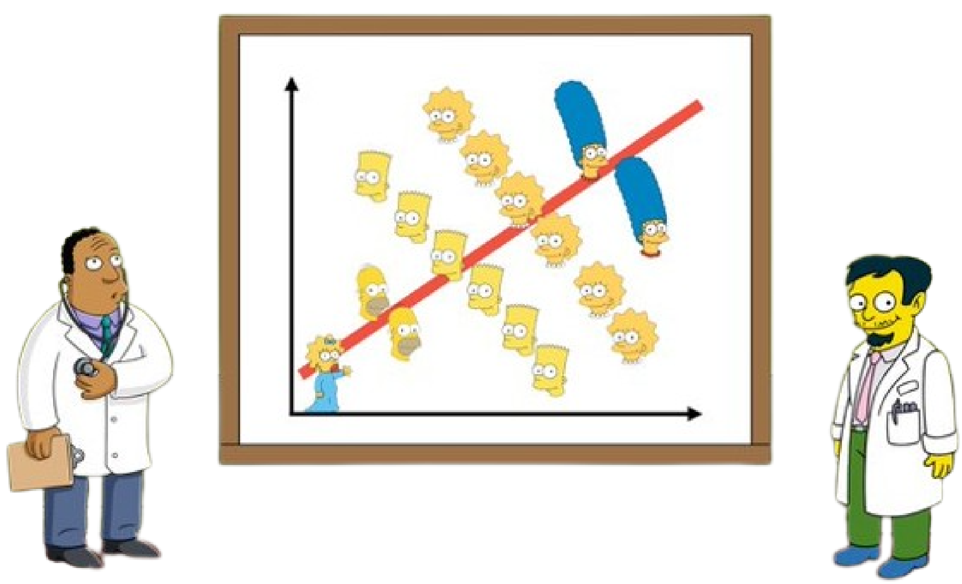
Master of Computer Science - Concepts & Applications Internship 2023
École Normale Supérieure de Lyon

Root Cause Analysis given a Cyclic Summary Causal Graph and Time Series

Simon Ferreira

Key words : *Causality, time series, root cause analysis, causal inference, summary causal graphs, cyclic graphs*

Internship supervisor :
Charles K. Assaad
cassaad@easyvista.com
COSERVIT, 34 av Europe, 38100 Grenoble, FRANCE
https://www.coservit.fr/



The Simpson's pair of docs

July 1, 2023

Acknowledgement

Firstly, I would like to express my gratitude to my supervisor Charles K. Assaad for his time, his thoughtful advice and all our interesting discussions. I look forward to continuing this collaboration. I have also enjoyed working with the rest of EasyVista's Lab team, Ali Ait-Bachir, Hossein Mohanna, Lei Zan and Mariia Liapustina and discussing with collaborators from the LIG, Eric Gaussier and Emilie Devijver. Secondly, thanks to EasyVista, I had the privilege of attending to and presenting a poster at two conferences, *When Causal Inference meets Statistical Analysis* in Paris and *Fundamental Challenges in Causality* in Grenoble. Lastly, my appreciations also go to my other colleagues for their warm-hearted welcome.

Contents

1 Introduction

Structural causal models (SCMs) are a powerful framework for representing and reasoning about causal relations between variables in dynamic systems with a long history in many fields such as genetics (Wright, 1920, 1921), econometrics (Haavelmo, 1943), social sciences (Duncan, 1975; Goldberger, 1972), and artificial intelligence (Pearl, 2000). Such causal models are often represented as graphs in which vertices are variables and edges are causal relations. This visual representation is easier to understand and validate by domain experts. An active field of research called causal discovery (Pearl, 2000; Spirtes et al., 2000) attempts to find the graph associated with the underlying SCM, given some data from a dynamic system. However, these causal discovery algorithms struggle to perform on real data.

Another approach is to assume that the causal graph to be known and to ask some causal questions regarding the system, such as estimating the direct effect of one variable on another. This can be done using respectively single-door criterion (Pearl, 2000). For example, epidemiologists are interested in measuring how smoking affects lung cancer risk without being mediated by genetic susceptibility (Zhou et al., 2021); ecologists are usually focus on understanding direct effects such as competition, herbivory, and predation (Connell, 1961); IT monitoring experts may try to identify the root cause of a system failure or a performance issue by comparing the direct causal impact of different components on each other before and after the failure (Assaad et al., 2023b).

In many areas, the data collected is a multivariate time series, thus the representation of systems in causal graphs has been extended to this case (Assaad et al., 2022a). While working with time series can be easier because it is assumed that causes cannot appear after effects, it also creates new challenges as experts often only have little knowledge of the temporal lags of causal phenomena and struggle to provide some insight concerning temporal information. Thus, there exists different types of graphs representing a system, each containing a different level of temporal information (Assaad et al., 2022a).

EasyVista, develops IT monitoring software and is looking for new innovative ways to help its customers manage their IT systems. Currently part of the research team is working on ways to automatize root cause analysis. Indeed, since IT systems are highly interconnected, if there is an issue, a large part of the system is likely to fail. As a result, it can be very challenging for experts to determine the root cause of the issue and fix it. The aim is that when some measured variables are considered to be anomalous, a root cause analysis algorithm would be able to provide experts with the set of root causes to enable a quick and lasting resolution of the issue.

This report is divided into three main sections. Section 2 states some useful definitions in causality and adapts them to time series. It also justifies the work presented in the other sections. During this internship, the underlying goal was to perform root cause analysis. The state of the art methods of root cause analysis mostly use causal discovery. Therefore, at the beginning of my internship I have participated in a collaborative work that consisted on testing different causal discovery algorithms on real IT monitoring data. The findings of this work are described in Section 3.1 which shows that causal discovery algorithms are still unsatisfactory for real-world applications. Note that this work has led to a scientific paper which was recently submitted to an international workshop:

> Ali Aït-Bachir, Charles K. Assaad, Christophe de Bignicourt, Emilie Devijver, Simon Ferreira, Eric Gaussier, Hosein Mohanna, and Lei Zan. Case studies of causal discovery from it monitoring time series. submitted, 2023

Section 3.2 presents a different framework for root cause analysis that does not rely primarily on causal discovery. More specifically, Section 3.2 introduces the EasyRCA algorithm (Assaad et al., 2023b) which makes a restrictive assumption of acyclicity of an abstraction of the system. Then in the same section I investigated how relaxing the acyclicity assumption affects the EasyRCA algorithm and I proposed necessary modifications to the EasyRCA algorithm to make it correct without the acyclicity assumption. This work resulted in a small paper and a poster which were presented at two national colloquia:

> Simon Ferreira and Charles Assaad. Challenges of root cause identification for collective anomalies in time series given a summary causal graph. *When Causal Inference meets Statistical Analysis*, 2023a

The main conclusion of Section 3.2 is that it is necessary to find a new way to estimate direct effects given a cyclic abstraction of the system. Therefore, Section 4, which constitutes my most valuable contribution of the internship, presents a new theoretical result for identifying direct effects from summary causal graphs and proposes two adjustment sets that can be used to estimate direct effect from data, in the case of identifiability This work led to writing a scientific theoretical paper which is available on arXiv and will soon be submitted to one of the top AI conferences:

Simon Ferreira and Charles K. Assaad. Identifiability of direct effects from summary causal graphs. *arXiv preprint arXiv*, 2023b

2 Problem setup

This section will go through the main definitions and concepts of causality along with some well-known results. Firstly, we will see how the graphical approach has been used in Pearl (2000) to answer some causal questions. Then, we will discuss on how this framework is extended to multivariate time series (Assaad et al., 2022a). The basic graphical notions are defined in Section A.1 of the Appendix so that the reader may refer to them at any moment.

2.1 Graphical criteria

In Pearl's causal framework (Pearl, 2000) a system is firstly represented by a SCM in which some variables are observed and others are not.

Definition 1 (Linear Structural Causal Model)**.** *Considering a finite set of random variables $\mathcal{V}$, a* linear structural causal model (SCM) *is a set of equations in which each variable (e.g., $Y \in \mathcal{V}$) is defined as a linear function of other variables and of some noise (e.g., ξ_Y). The noise variables are assumed to be jointly independent i.e., $\forall X, Y,\ X \neq Y \implies \xi_X \perp\!\!\!\perp \xi_Y$.*

$$Y := \sum_{X \neq Y \in \mathcal{V}} \alpha_{Y,X} * X + \xi_Y \tag{1}$$

Any coefficient α in Equation (1) *can be zero. One could also define non-linear SCMs, however in our case we will always assume linearity.*

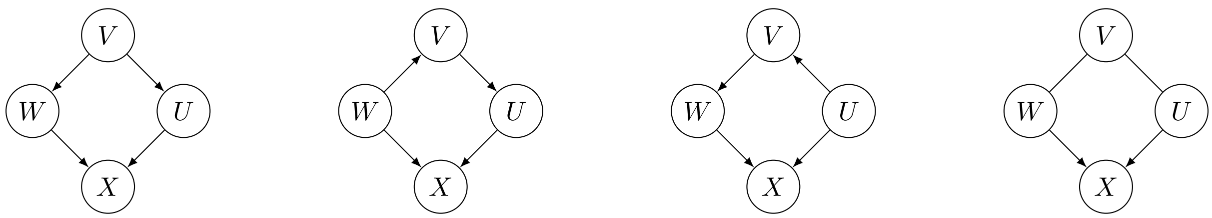
Intuitively, if $\alpha_{Y,X} \neq 0$ then X is said to cause Y. A SCM is represented as a causal graph in which every vertex corresponds to a random variable in the SCM and every edge corresponds to a causal effect *i.e.*, $X \rightarrow Y \in \mathcal{E} \iff \alpha_{Y,X} \neq 0$. One could also imagine having some weighted edges which, if the distributions of the noise variables are omitted, allow a bijection between causal graphs and linear SCMs. However, the weights are often not known and thus, in the following, we will only distinguish between $\alpha = 0$, which is the presence of an edge, and $\alpha \neq 0$, which is the absence of an edge. Note that the remainder of this section as well as Section 4 gives ways to recover the exact values of those coefficients.

Definition 2 (Intervention)**.** *Considering a SCM,* intervening *on a variable Y is modifying the equation defining Y in the SCM without modifying the rest of the SCM. We can distinguish two types of interventions:*

- *Structural interventions: the new equation is of the form $Y := y$ and thus Y looses its parents. This is written $do(y)$.*
- *Parametric interventions: the new equation is of the form $\sum_{X \neq Y \in \mathcal{V}} \tilde{\alpha}_{Y,X} * X$ and thus Y does not loose its parents. In this case we assume $\alpha_{Y,X} = 0 \implies \tilde{\alpha}_{Y,X} = 0$ and thus Y does not gain any new parent.*

Assumption 1 (Causal Sufficiency, Spirtes et al. (2000))**.** *We will assume that every variable of the system is observed.*

Given Definition 1, Property 1 shows how conditional independences can be helpful in recovering a causal graph from statistical data. Moreover, some additional rules can allow to learn the orientation of some arrows. Unfortunately, there exists multiple graphs which can be obtained from the same conditional independences, this gives rise to the Markov properties and equivalence relation.



(a) DAG 1 (b) DAG 2 (c) DAG 3 (d) CPDAG

Figure 1: An example of a MEC.

Property 1. *Consider a SCM, its corresponding causal graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and its skeleton $\mathcal{G}_{sk} = (\mathcal{V}_{sk}, \mathcal{E}_{sk})$ (Definition 24 in Appendix). For every pair of vertices X, Y we have:*

$$X - Y \in \mathcal{E}_{sk} \iff \nexists \mathcal{Z} \subseteq \mathcal{V} \text{ such that } X \perp\!\!\!\perp Y | \mathcal{Z}$$

where $\perp\!\!\!\perp$ represents statistical independence.

Property 2 (Local Directed Markov Property, (Pearl, 1988))**.** *A probability distribution verifies the* local directed Markov property *for a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if:*

$$\forall V \in \mathcal{V},\ V \perp\!\!\!\perp \mathcal{V} \backslash (Parents(V, \mathcal{G}) \cup Descendants(V, \mathcal{G})) | Parents(V, \mathcal{G})$$

Property 3 (Global Directed Markov Property, (Pearl, 1988))**.** *A probability distribution verifies the* global directed Markov property *for a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if $\forall \mathcal{X}, \mathcal{Y}, \mathcal{Z} \subseteq \mathcal{V}$ with $\mathcal{X} \cap \mathcal{Y} = \mathcal{X} \cap \mathcal{Z} = \mathcal{Y} \cap \mathcal{Z} = \emptyset$:*

$$\mathcal{X}, \mathcal{Y} \text{ are d-separated by } \mathcal{Z} \implies \mathcal{X} \perp\!\!\!\perp \mathcal{Y} | \mathcal{Z}$$

Properties 2 and 3 are equivalent when considering DAGs (Pearl, 1988).

Definition 3 (Markov Equivalence Relation)**.** *Two DAGs are said to be* Markov equivalent *if they represent the same conditional independences. The equivalence classes of this relation are called* Markov equivalent classes (MEC)*. An example of a MEC is given in Figure 1.*

Property 4 ((Verma and Pearl, 1990))**.** *Two DAGs $\mathcal{G}, \mathcal{G}'$ are Markov equivalent if and only if they have the same skeleton (Definition 24 in Appendix) and the same v-structures (Definition 31 in Appendix) i.e., $\mathcal{G}_{sk} = \mathcal{G}'_{sk}$ and $V(\mathcal{G}) = V(\mathcal{G}')$.*

Definition 4 (Completed Partially Directed Acyclic Graph (CPDAG))**.** *A partially directed graph is a graph with three type of edges: $\rightarrow, \leftarrow$ and $-$. Property 4 shows that a MEC can be uniquely represented by a so-called* completed partially directed acyclic graph*. The CPDAG $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represents the MEC $\{\mathcal{G}^i = (\mathcal{V}^i, \mathcal{E}^i)\}_{1 \leq i \leq n}$ if:*

- $\forall 1 \leq i \leq n,\ \mathcal{V} = \mathcal{V}^i,$ *and*
- $U - V \in \mathcal{E} \iff$ *:*
  - $\forall 1 \leq i \leq n,\ U - V \in \mathcal{E}^i_{sk},$ *and*
  - $\exists 1 \leq i \leq n,\ U \rightarrow V \in \mathcal{E}^i,$ *and*
  - $\exists 1 \leq i \leq n,\ U \leftarrow V \in \mathcal{E}^i,$ *and*
- $U \rightarrow V \in \mathcal{E} \iff \forall 1 \leq i \leq n,\ U \rightarrow V \in \mathcal{E}^i$

This notion of Markov equivalence and the associated graphical criterion helps gain a better understanding of the field and how statistical information can be represented as a graph. In the remainder of this section, we will see how one can use graphical knowledge to learn some causal coefficients and answer causal questions such as "To what extend does this one variable cause that second one".

Definition 5 (Blocked Walks). *In a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a walk $\pi =< V^1, \dots, V^n >$ is said to be* blocked *by a set of vertices $\mathcal{Z} \subseteq \mathcal{V}$ (or $\mathcal{Z}$-blocked) if:*

1. *$V^1 \in \mathcal{Z}$ or $V^n \in \mathcal{Z}$, or*
2. *$\exists 1 < i < n$ such that $V^i \in \mathcal{Z}$ and V^i is not a collider (Definition 30 in Appendix) i.e., $V^{i-1} \leftarrow V^i$ or $V^i \rightarrow V^{i+1}$, or*
3. *$\exists 1 < i < n$ such that $Descendants(V^i, \mathcal{G}) \cap \mathcal{Z} = \emptyset$ and V^i is a collider (Definition 30 in Appendix) i.e., $\rightarrow V^{i-1} \rightarrow V^i \leftarrow V^{i+1}$.*

A walk which is not blocked is said to be active*. When the set $\mathcal{Z}$ is not specified, it is implicit that we consider $\mathcal{Z} = \emptyset$. In the case of condition 2, we say that π is manually $\mathcal{Z}$-blocked by V^i and in the case of condition 3 we say that π is passively $\mathcal{Z}$-blocked by V^i.*

Definition 6 (d-Separation). *In a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, two distinct variables $X \neq Y \in \mathcal{V}$ are said to be* d-separated *by $\mathcal{Z} \subseteq \mathcal{V}$ if every path from X to Y is blocked by $\mathcal{Z}$.*

Definition 7 (Direct Effect, Pearl (2000)). *Considering a SCM, the* direct effect *of X on Y is given by*

$$P(y|do(x), do(v_{X,Y})) - P(y|do(x'), do(v_{X,Y}))$$

where $V_{X,Y} = \mathcal{V} \backslash \{X, Y\}$ and $P(y|do(x), do(v_{X,Y}))$ is the distribution of Y while X and $V_{X,Y}$ are held constant respectively at values x and $v_{X,Y}$. In the linear case, the direct effect corresponds to the structural coefficient[1] *$\alpha_{Y,X}$.*

The direct effect corresponds to how the distribution of a variable changes when one forces another variable to take a arbitrary value and maintains the rest of the system to a fix value. It helps answer questions such as "What portion of patients have been healed thanks to the medication?", "What portion of patients have been healed thanks to the placebo effect?".

This direct effect is very useful to answer causal questions and thus has been accorded a lot of attention from researchers. This focus has given rise to a criterion which explains how to compute the direct effect.

Definition 8 (Graphical Identifiability). *Considering a linear SCM, the direct effect of a variable X on another variable Y is said to be* identifiable *from a SCG if it can be computed uniquely from the observed distribution without any further assumption on the distribution i.e., if it is possible to eliminate the $do()$ from Definition 7.*

Definition 9 (Single-door Adjustment Set (Pearl, 2000)). *Considering a SCM and its compatible graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $X \neq Y \in \mathcal{V}$ and $\mathcal{Z} \subseteq \mathcal{V} \backslash \{X, Y\}$. $\mathcal{Z}$ is a* single-door adjustment set *if:*

- *$\mathcal{Z} \cap Descendants(Y, \mathcal{G}) = \emptyset$, and*
- *$\mathcal{Z}$ blocks every non-direct path from X to Y.*

Property 5 (Single-door Criterion (Pearl, 2000)). *Considering a SCM and its compatible graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, let $X \neq Y \in \mathcal{V}$ and $\mathcal{Z}$ a single-door adjustment set. The direct effect $\alpha_{Y,X}$ is computable with a simple regression.*

Property 5 allows to compute the direct effect.

2.2 Time series and their representations

Now that all these classical concepts have been formally defined this section will discuss of a particular type of data and how to adapt the previous definitions for it. Multivariate time series are a set of variables measured regularly. It is assumed that each variable is measured at the same instants. Therefore, we obtain one point of data per variable per time instant. This causes some complications, for example, since there is a potentially infinite amount of time instants, it can be difficult to represent such systems. However, it also has some advantages as temporal information can be quite instructive because it is assumed that a cause always comes before its effects.

Let us adapt the previous framework to time series.

[1] In Sewall Wright's terminology the structural coefficient is called path coefficient (Wright, 1920, 1921).

Definition 10 (Linear Dynamic SCM). *Considering a finite set of variables $\mathcal{V}$, a* linear dynamic structural causal model *is a set of equations in which each instant $t \in \mathbb{Z}$ of each variable (e.g., Y_t) is defined as a linear function of past instants of itself (e.g., $Y_{t-\gamma},\ \gamma > 0$), past or present instants of other variables (e.g., $X_{t-\gamma},\ X \neq Y,\ \gamma \geq 0$) and of some noise (e.g., ξ_{Y_t}). The noise variables are assumed to be jointly independent (i.e., $\forall X, Y\ \forall t^X, t^Y,\ X_{t^X} \neq Y_{t^Y} \implies \xi_{X_{t^X}} \perp\!\!\!\perp \xi_{Y_{t^Y}}$).*

$$Y_t := \left(\sum_{\gamma > 0} \alpha_{Y_t, Y_{t-\gamma}} * Y_{t-\gamma} \right) + \left(\sum_{X \neq Y,\ \gamma \geq 0} \alpha_{Y_t, X_{t-\gamma}} * X_{t-\gamma} \right) + \xi_{Y_t} \tag{2}$$

Any coefficient α in Equation (2) *can be zero.*

As for non-dynamic linear SCMs, in the case of a linear dynamic SCM, the direct effect of $X_{t-\gamma}$ on Y_t is $\alpha_{Y_t, X_{t-\gamma}}$. Just as before, one can represent a linear dynamic SCM with a causal graph. This graph is called a full-time causal graph.

Definition 11 (Full-Time Causal Graph (FTCG)). *Considering a finite set of variables and a SCM, one can define the* full-time causal graph (FTCG) $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ *associated to the SCM in the following way:*

$$\begin{aligned}
\mathcal{V}_f &:= \{Y_t & &|\forall Y \in \mathcal{V},\ \forall t \in \mathbb{Z}\} \\
\mathcal{E}_f &:= \{X_{t-\gamma} \rightarrow Y_t & &|\forall Y_t \in \mathcal{V}_f,\ \forall X_{t-\gamma} \text{ such that } \alpha_{Y_t, X_{t-\gamma}} \neq 0\}
\end{aligned}$$

As in many other works, we assume here that the FTCG is acyclic but we think that this assumption can be relaxed for the main contribution given in Sections 4.

Assumption 2 (Acyclicity of the FTCG). *Every FTCG is acyclic.*

The FTCG is the most natural way to represent a SCM but it is unpractical as it is infinite. However, given Assumption 3, it is possible to represent a FTCG in a finite graph which is called a window causal graph (Assaad et al., 2022a).

Assumption 3 (Stationarity of direct effects). *The causal mechanisms of the system considered do not change and therefore $\forall X, Y \in \mathcal{V},\ \forall t - \gamma \leq t \in \mathbb{Z},\ \alpha_{Y_t, X_{t-\gamma}} = \alpha_{Y_{t+1}, X_{t-\gamma+1}}$. There exists a maximum lag γ_{max} of a SCM as $\gamma_{max} := max\{\gamma \in \mathbb{N} | \exists X, Y \in \mathcal{V},\ \alpha_{Y_t, X_{t-\gamma}} \neq 0\}$*

Definition 12 (Window Causal Graph). *Considering a finite set of variables and a linear SCM and assuming Assumption 3, one can define the* window causal graph (WCG) $\mathcal{G}_w = (\mathcal{V}_w, \mathcal{E}_w)$ *associated to the SCM in the following way:*

$$\begin{aligned}
\mathcal{V}_w &:= \{Y_{t-\gamma} & &|\forall Y \in \mathcal{V},\ \forall 0 \leq \gamma \leq \gamma_{max}\} \\
\mathcal{E}_w &:= \{X_{t'-\gamma} \rightarrow Y_{t'} & &|\forall X_{t'-\gamma}, Y_{t'} \in \mathcal{V}_w \text{ such that } \alpha_{Y_{t'}, X_{t'-\gamma}} \neq 0\}
\end{aligned}$$

Notice that under Assumption 3, there exists a bijection between FTCGs and WCGs. Indeed, it is easy to obtain a WCG compatible with a FTCG by restricting to a subset of $\gamma_{max} + 1$ consecutive instants. It is also possible to obtain a FTCG compatible with a WCG by extending it to infinity to account for all instants and adding all the edges using equivalence $X_{t-\gamma} \rightarrow Y_t \in \mathcal{E}_w \iff \forall n \in \mathbb{Z},\ X_{t-\gamma+n} \rightarrow Y_{t+n} \in \mathcal{E}_f$.

While WCGs have the advantage of representing qualitatively a given linear SCM under Assumption 3, it is in practice difficult to obtain the temporal lags. Causal discovery methods are not always efficient (Aït-Bachir et al., 2023) due to the strong assumptions they require that are not always satisfied in real applications. So we often rely on experts to obtain a causal graph. However, experts struggle to have any insight on the temporal lags between causal relations. Therefore, it is much more reliable to ask for an abstraction of this graph which contains little to no temporal information. There exists two such abstractions, the summary causal graph (Assaad et al., 2022a) which has no temporal information and the extended summary causal graph which only distinguishes lagged and instantaneous relations.

Definition 13 (Extended Summary Causal Graph). *Consider a FTCG $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ (or equivalently a WCG). One can define the* extended summary causal graph (ESCG) $\mathcal{G}_e = (\mathcal{V}_e, \mathcal{E}_e)$ *in the following way:*

$$\begin{aligned}
\mathcal{V}_e &:= \{V_t, V_- | V \in \mathcal{V}\} \\
\mathcal{E}_e &:= \{X_t \rightarrow Y_t & &|\forall X, Y \in \mathcal{V} \text{ such that } X_t \rightarrow Y_t \in \mathcal{E}_f\} \\
&\quad \cup \{X_- \rightarrow Y_t & &|\forall X, Y \in \mathcal{V},\ \exists t' < t \in \mathbb{Z} \text{ such that } X_{t'} \rightarrow Y_t \in \mathcal{E}_f\}
\end{aligned}$$

Definition 14 (Summary Causal Graph)**.** *Consider a FTCG $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ (or equivalently a WCG). One can define the* summary causal graph (SCG) $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ *in the following way:*

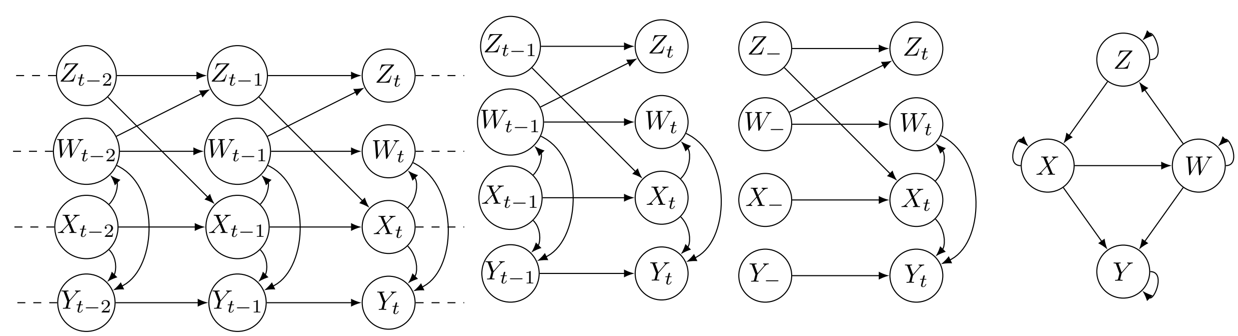
$$\begin{aligned}\mathcal{V}_s \;&:= \mathcal{V}\\ \mathcal{E}_s \;&:= \{X \to Y \quad | \forall X, Y \in \mathcal{V},\ \exists t' \leq t \in \mathbb{Z} \text{ such that } X_{t'} \to Y_t \in \mathcal{E}_f\}\end{aligned}$$

Notice that a SCG may have cycles and in particular two arrows in opposite directions, i.e., if in the FTCG we have $X_{t'} \to Y_t$ and $Y_{t''} \to X_t$ then in the SCG we have $X \rightleftarrows Y$.

The abstraction of summary causal graphs entails that, even though there is exactly one SCG compatible with a given FTCG, WCG or ESCG, there are in general several FTCGs, WCGs and ESCGs compatible with a given SCG.

For example, the FTCG, the WCG, the ESCG and the SCG of the following linear dynamic SCM, in which the coefficients α equal to zero are omitted, are respectively given in Figures 2a, 2b, 2c and 2d.

$$\begin{aligned}W_t &:= \alpha_{W_t,W_{t-1}} * W_{t-1} + \alpha_{W_t,X_t} * X_t + \xi_{W_t}\\ X_t &:= \alpha_{X_t,X_{t-1}} * X_{t-1} + \alpha_{X_t,Z_{t-1}} * Z_{t-1} + \xi_{X_t}\\ Y_t &:= \alpha_{Y_t,W_t} * W_t + \alpha_{Y_t,X_t} * X_t + \alpha_{Y_t,Y_{t-1}} * Y_{t-1} + \xi_{Y_t}\\ Z_t &:= \alpha_{Z_t,W_{t-1}} * W_{t-1} + \alpha_{Z_t,Z_{t-1}} * Z_{t-1} + \xi_{Z_t},\end{aligned}$$



(a) Full time causal graph (b) Window causal graph (c) Extended summary causal graph (d) Summary causal graph

Figure 2: Different causal graphs that one can infer from three time series (Assaad et al., 2022a): full time causal graph (2a), window causal graph (2b), extended summary causal graph (2c) and summary causal graph (2d). Note that the first one gives more information but cannot be inferred in practice, the second one is a schematic viewpoint of the full behavior, the third one only distinguishes lagged and instantaneous relations whereas the last one is an abstraction and can be deduced from any other one.

As SCGs represent FTCGs, the walks in a SCG can represent the paths of compatible FTCGs. This is gives rise to the notion of compatible walk.

Definition 15 (Compatible Walk)**.** *Let $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ be a FTCG and $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ the compatible SCG. A path $\pi_f =< V^1_{t_1}, \ldots, V^n_{t_n} >$ in $\mathcal{G}_f$ can be uniquely represented as a walk $\pi_s =< V^1, \ldots, V^n >$ in $\mathcal{G}_s$ in which the temporal information has been removed. We refer to π_s as π_f's* compatible walk *and we write $\pi_s = \phi(\pi_f)$. e.g., $\phi(< X_{t-1}, X_t, Y_t, Z_t, Z_{t+1}, X_{t+1} >) =< X, X, Y, Z, Z, X >$.*

An important graphical notion used in causal reasoning is the notion of blocked path (Pearl, 1998). The classical definition was introduced for directed acyclic graphs (DAGs) and thus can be directly used for FTCGs under Assumption 2. However, since the SCG compatible with a FTCG can be cyclic, one needs to adapt it. Spirtes (1993) explains that under the linearity assumption the notion of blocked path is readily extended. Moreover, Forré and Mooij (2017) introduced a more recent and general (non-parametric and allow for hidden confounding) adaptation called σ-blocked path. Forré and Mooij (2017)'s method relies on strongly connected components and segments, it is briefly explained in the appendix. However, no method has been adapted to summary causal graphs. In the following, we adapt Spirtes (1993)'s notion to time series, we consider walks and we reformulate it to explicitly consider pairs of vertices with two opposite arrows between them (*e.g.*, $X \rightleftarrows Y$).

Definition 16 (Blocked Walk in SCGs)**.** *In a SCG $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$, a walk $\pi_s =< V^1, \ldots, V^n >$ is said to be* blocked *by a set of vertices $\mathcal{Z}_s \subseteq \mathcal{V}_s$ if:*

1. *$\exists 1 < i < n$ such that $V^{i-1} \leftarrow V^i$ or $V^i \rightarrow V^{i+1}$ and $V^i \in \mathcal{Z}_s$, or*
2. *$\exists 1 < i \leq j < n$ such that $V^{i-1} \rightarrow V^i \rightleftarrows \cdots \rightleftarrows V^j \leftarrow V^{j+1}$ and $Descendants(V^i, \mathcal{G}_s) \cap \mathcal{Z}_s = \emptyset$*

A walk which is not blocked is said to be active*. When the set $\mathcal{Z}_s$ is not specified, it is implicit that we consider $\mathcal{Z}_s = \emptyset$. In the case of condition 1, we say that π_s is manually $\mathcal{Z}_s$-blocked by V^i and in the case of condition 2 we say that π_s is passively $\mathcal{Z}_s$-blocked by $\{V^k | i \leq k \leq j\}$.*

Condition 1 in Definition 16 is a direct adaptation of condition 2 in Definition 5. Condition 2 of Definition 16 is explained by the fact that for a walk $\pi_s =< V^1, \ldots, V^n >$ in a SCG $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ and a set of vertices $\mathcal{Z}_s \subseteq \mathcal{V}_s$, if $\exists 1 < i \leq j < n$ such that $V^{i-1} \rightarrow V^i \rightleftarrows \cdots \rightleftarrows V^j \leftarrow V^{j+1}$ and $Descendants(V^i, \mathcal{G}_s) \cap \mathcal{Z}_s = \emptyset$ then $\forall \pi_f =< V^1_{t^1}, \ldots, V^n_{t^n} >\in \phi^{-1}(\pi_s)$, $\exists 1 < i \leq k \leq j < n$ such that $V^{k-1}_{t^{k-1}} \rightarrow V^k_{t^k} \leftarrow V^{k+1}_{t^{k+1}}$ and $Descendants(V^k, \mathcal{G}_s) \cap \mathcal{Z}_s = \emptyset$ so $Descendants(V^k_{t^k}, \mathcal{G}_f) \cap \mathcal{Z}_f = \emptyset$ where $\mathcal{Z}_f \subseteq \{V_{t'} | V \in \mathcal{Z}_s,\ t' \in \mathbb{Z}\}$. Notice that there is no adaptation of condition 1 as having $V^1 \in \mathcal{Z}_s$ or $V^n \in \mathcal{Z}_s$ does not mean that the instants of interests $V^1_{t^1}$ and $V^n_{t^n}$ which are endpoints of compatible paths of interests, are in $\mathcal{Z}_f \subseteq \{V_{t'} | V \in \mathcal{Z}_s,\ t' \in \mathbb{Z}\}$. Moreover, in the remainder of this report we are interested in the direct effect and therefore we will always have $V^1_{t^1}, V^n_{t^n} \notin \mathcal{Z}_f$. However, having $V^1 \in \mathcal{Z}_s$ or $V^n \in \mathcal{Z}_s$ can activate or block walks *e.g.*, $< X \leftarrow W \rightarrow X \leftarrow Z \rightarrow Y >$ and $< X \leftarrow W \rightarrow X \rightarrow Z \rightarrow Y >$.

The purpose of this adaptation is to have some link between the sets $\mathcal{Z}_s \subseteq \mathcal{V}_s$ that blocks a walk in a SCG and the sets $\mathcal{Z}_f \subseteq \mathcal{V}_f$ that block the corresponding paths in corresponding FTCGs. This link will be discussed in Section 4.

3 Root cause analysis

As explained in Section 1, the main objective of this internship was to study methods of root cause analysis. EasyVista works on IT monitoring, therefore they measure regularly many components of IT systems and thus work with multivariate time series. The goal is, when a problem occurs and anomalies are detected, to find quickly and automatically the root cause of the issue.

An anomaly is defined as a variable measured at some value which differs significantly from what was expected. A basic example are outliers. Experts define a normal interval of values for each variable *e.g.*, one could consider that any IT system should have at least 1GO of free memory space in order to function properly. Any variable with a measured value outside of this interval is an outlier *e.g.*, a measure of free memory space under 1GO would then be considered as an outlier.

Assumption 4 (Anomalies (Assaad et al., 2023b))**.** *We make many assumptions about anomalies in order to facilitate the root cause analysis problem.*

1. *The anomalies are propagated through the SCM. In a FTCG, every children of an anomalous vertex is anomalous.*
2. *Every anomaly is either due to a another anomaly or due to an intervention. In a FTCG, every anomalous vertex has been intervened on or has an anomalous parent.* Root causes *are vertices which have been intervened on.*
3. *The anomalies are considered to be collective i.e., to last in time (Chandola et al., 2009). Indeed, if an anomaly does not last one most probably does not have enough data to do any statistical analysis. Moreover, one could claim that an anomaly that does not last does not need to be solved.*

3.1 Causal discovery methods and limitations

In the state of the art, one can find multiple recent approaches to identify the root causes of anomalies in multivariate time series. Most of these methods consist in discovering the anomalous causal graph (Pearl, 2000; Spirtes et al., 2000) (*i.e.*, the causal graph of the anomalous regime of the dynamic system (Wang et al., 2018, 2021)) before comparing it to the previously known normal causal graph. For example, one of the most widely used method is MicroCause (Meng et al., 2020) and uses either PCMCI and PCMCI$^+$ to discover the anomalous causal graph. However, Assaad et al. (2023b) shows that this method does not have satisfactory performances on real IT monitoring data collected by Easyvista. These low performances were expected as causal discovery methods are known to rely on strong assumptions and need large data sizes (Malinsky and Danks, 2018; Glymour et al., 2019; Assaad et al., 2022a). Thus, during this internship, I helped writing Aït-Bachir et al. (2023) which tests different causal discovery methods on real IT monitoring datasets collected by Easyvista to estimate their performance. This participation allowed me to better understand the difficulties of using causal discovery methods on real data which I will describe in Section 3.1.

Firstly, (Chickering et al., 2004) showed that causal discovery was NP-hard. Moreover, some current algorithms such as LiNGAM assume non-Gaussian noise which is an assumption rarely verified. Other algorithms which do not make this assumption only discover the CPDAG and not the full graph. In addition, real data is often not of the best quality: there may be some missing data, the sampling rate may be different or not synchronized for different variables, there may be some mixed data types, etc.

To verify the performance of causal discovery methods, Aït-Bachir et al. (2023) tried to use 10 different causal discovery algorithms to recover 4 different causal graphs which were previously known, either thanks to experts or because of the architecture of the system.

3.1.1 The causal discovery algorithms

The causal discovery algorithms chosen to be tested on our real data have different theoretical foundations. This is done in an effort to test a wide range of methods representative of the state of the art. In Table 1, we classify causal discovery algorithms with respect to the assumptions they rely on in addition to different characteristics.

GCMVL (Arnold et al., 2007) is multivariate Granger (Granger, 1969, 2004; Arnold et al., 2007) algorithm that use a lasso-based technique for variable selection. Dynotears (Pamfil et al., 2020) is a score-based method (Chickering, 2002) and was presented to infer a WCG from time series. Among the constraint-based approaches (Spirtes et al., 2000), PCMCI (Runge et al., 2019) which infers a WCG was initially not able to take into account instantaneous relations but this limitation was recently surmounted with the introduction of PCMCI$^+$ (Runge, 2020). Another algorithm in this family is PCGCE (Assaad et al., 2022b) which infers an ESCG. In a different line, VarLiNGAM (Hyvärinen et al., 2008; Hyvärinen et al., 2010), which is an extension of LiNGAM (Shimizu et al., 2011), and TiMINo (Peters et al., 2013), infer respectively a WCG and a SCG by looking at independences between the noise and the potential causes. There exist also hybrid algorithms which combine constraint-based with semi-parametric algorithms. Among hybrid methods, NBCB (Assaad et al., 2021, 2023a) starts by discovering the causal order between time series through a semi-parametric strategy, here VarLiNGAM, and then prunes unnecessary edges using a constraint-based strategy, either PCMCI$^+$ for NBCB-w or PCGCE for NBCB-e. CBNB-w and CBNB-e from the CBNB framework (Assaad et al., 2023a) can be considered respectively as the backward versions of NBCB-w and NBCB-e. Links for the source codes of these algorithms are in Section A.3.

3.1.2 The real datasets

EasyVista provided us with 7 real datasets coming from 4 different IT systems. Moreover, for each IT system, a real summary causal graph was constructed either by IT monitoring experts or directly using the system topology. In Table 1, we also classify datasets with respect to the different assumptions needed by causal discovery algorithms and to other different characteristics.

A more thorough description of the datasets as well as their real summary causal graphs is available in Aït-Bachir et al. (2023) and in Section A.3 of the appendix.

		Causal graph	Causal Markov Condition	Causal sufficiency	Faithfulness / Minimality	Semi-parametric model	Linear model	Consistency throughout time	Stationarity	Instantaneous relations	Misaligned time series	Sleeping time series	Timestamp errors	Missing values	Different sampling rate
Algorithms	GCMVL	S/E	✓	✓		✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
	Dynotears	W	✓	✓		✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
	PCMCI$^+$	W	✓	✓	F	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗
	PCGCE	E	✓	✓	F	✗	✗	✓	✓	✓	✗	✗	✗	✗	✗
	VarLiNGAM	W	✓	✓	M	✓	✓	✓	✓	✓	✗	✗	✗	✗	✗
	TiMINo	S	✓	✓	M	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗
	NBCB-w	S/W	✓	✓	M	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗
	NBCB-e	S/W	✓	✓	M	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗
	CBNB-w	S/W	✓	✓	M	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗
	CBNB-e	S/W	✓	✓	M	✓	✗	✓	✓	✓	✗	✗	✗	✗	✗
Datasets	MoM	S	✓	✓	?	?	?	✗	✗	✓	✗	✓	✓	✗	✗
	Ingestion	S	✓	✓	?	?	?	✗	✗	✓	✗	✓	✓	✗	✗
	Web activity	S	✓	?	?	?	?	✗	✗	✓	✓	✓	✓	✗	✗
	Antivirus	S	✓	?	?	?	?	✗	✗	✓	✓	✓	✓	✗	✓

Table 1: Summary of the main characteristics of algorithms and different IT monitoring datasets considered in the paper. For causal graphs, S means that the algorithm provides a summary causal graph, E means that the algorithm provides an extended summary causal graph and W means that the algorithm provides a window causal graph; F corresponds to faithfulness and M to minimality. An empty cell mean that the information given in the corresponding column was not discussed by the authors of the corresponding algorithm. A question mark means that the expert of the IT system do not know if the information given in the corresponding column is satisfied for the given dataset.

3.1.3 Results

For each algorithm and dataset, the obtained graph was compared to the real graph and attributed a score. This score is called the F_1-score and is defined as the harmonic mean of the precision and recall *i.e.*, F_1-score $= \frac{2}{\frac{1}{P}+\frac{1}{R}}$ where $P = \frac{tp}{tp+fp}$ and $R = \frac{tp}{tp+fn}$ and tp, fp, fn correspond to number of edges and are respectively the true positives, false positives and false negatives. The results of this study are given in Table 2.

One can see that the performance results of the causal discovery methods are not satisfactory. Moreover, such algorithms can only detect structural interventions. Thus, we have tried a different approach which does not rely on the discovery of the anomalous causal graph. This is described in Section 3.2.

	MoM 1	MoM 2	Ingestion	Web 1	Web 2	Antivirus 1	Antivirus 2
GCMVL	0.0	0.0	0.2	0.2	0.0	0.08	0.0
Dynotears	0.26	0.2	0.14	0.23	0.3	0.18	0.19
PCMCI$^+$	**0.4**	0.0	0.0	0.23	0.3	0.04	0.11
PCGCE	0.0	0.12	0.12	0.22	0.15	0.3	**0.45**
VLiNGAM	0.0	0.0	0.19	**0.29**	0.18	0.15	0.22
TiMINo	0.0	0.17	0.18	0.0	0.0	0.0	0.0
NBCB-w	**0.4**	0.0	0.13	0.23	0.3	0.14	0.24
NBCB-e	0.13	**0.29**	**0.27**	0.19	**0.42**	**0.31**	**0.45**
CBNB-w	**0.4**	0.0	0.15	0.23	0.3	0.17	0.16
CBNB-e	0	0.24	0.13	0.22	0.29	**0.31**	0.38

Table 2: Results for real IT monitoring datasets where γ_{max} is set according to the 15 seconds delay rule for MoM datasets and to the 15 minutes delay rule for Ingestion, Web and Antivirus datasets. We report the F_1-score.

3.2 EasyRCA

The EasyRCA algorithm (Assaad et al., 2023b) is capable to detect the root causes of an anomalous system which is represented as an acyclic SCG with loops (Assumption 5). It takes as input the normal SCG, the set of anomalous vertices, the instant at which each anomalous vertex has become anomalous and the anomalous data. Figure 3 shows the framework around EasyRCA.

Assumption 5 (Acyclic SCG with Loops (ASCGL))**.** *In the EasyRCA algorithm it is assumed that the normal SCG is acyclic with loops. Loops are cycles of the form* $Y \rightleftarrows Y$ *and thus (Assaad et al., 2023b) only considers SCGs* $\mathcal{G}_s$ *with* $Cycles(\mathcal{G}_s) = \{< V \rightleftarrows V > | V \in \mathcal{V}\}$.

The EasyRCA algorithm is decomposed in three distinct steps: the first one decomposes the ASCGL into *linked anomalous graph*, the second step finds root causes from graphical information only, lastly the third step finds every remaining root cause using the data. The following sections explain each step of the EasyRCA algorithm in the classical case of an ASCGL and discuss its applicability in the case of a general SCG.

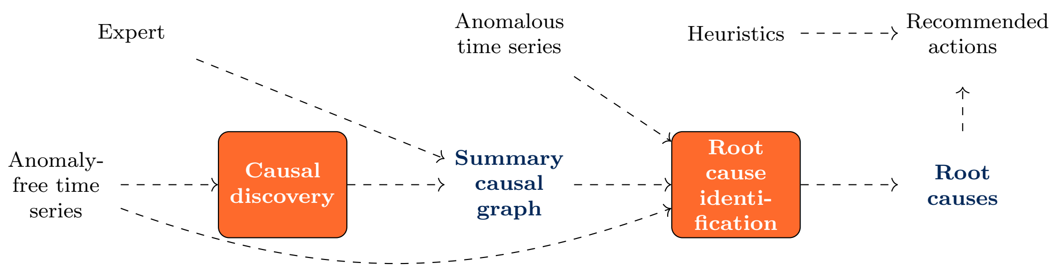


Figure 3: EasyVista's root cause analysis framework.

3.2.1 Step 1: Linked anomalous graphs

In the first step, EasyRCA divides the problem of root cause identification into sub-problems by dividing the main acyclic summary causal graph into many sub-graphs called linked anomalous graphs:

Definition 17 (Linked Anomalous Graph)**.** *Given an SCG* $\mathcal{G}_s = \{\mathcal{V}_s, \mathcal{E}_s\}$ *and a set of anomalous vertices* $\mathcal{A} \subset \mathcal{V}_s$, $\mathcal{L} = \{\mathcal{L}^1, \cdots, \mathcal{L}^m\}$ *is a set of linked anomalous graphs if* $\forall i \in \{1, \cdots, m\}$ $\mathcal{L}^i = (\mathcal{A}^i, \mathcal{E}^i)$ *is a sub-graph of* $\mathcal{G}_s$ *such that* $\mathcal{A}^i \subset \mathcal{A}$ *and there exists a set of vertices* $\mathcal{S} \subseteq \mathcal{V}_s \backslash \mathcal{A}$ *such that* $\mathcal{A}^i$ *and* $\mathcal{A} \backslash \mathcal{A}^i$ *are d-separated given* $\mathcal{S}$*. In other words the set of linked anomalous graphs is the set of anomalous connected components.*

Propositions 2 and 3 in (Assaad et al., 2023b) suggest that linked anomalous graphs are modular with respect to each other, which implies that the set of root causes of each linked anomalous graph can be identified independently of the rest of the anomalies in the graph. This step is described in Figure 4.

Considering a SCG $\mathcal{G}_s$ with a cycle $\mathcal{C} \in Cycles(\mathcal{G}_s)$, Assumption 4 states that either every vertex of the cycle $\mathcal{C}$ is anomalous or none of them are. Thus, a cycle is necessarily in a unique linked anomalous graph. This first step can remain unchanged in the case of general SCGs but linked anomalous graphs may themselves be cyclic.

3.2.2 Step 2: Finding external intervention from graph

In the second step, EasyRCA searches for root causes that can be identified purely from the graph and from the appearance times of anomalies. If a node of the graph is anomalous, then either it is a root cause or the anomaly has propagated from some of its parents and such parents are necessarily anomalous. Therefore, while considering the full time causal graph, an anomalous node with no anomalous parent is a root cause. In the summary graph, these nodes will be either anomalous nodes with no anomalous parents or nodes which have become anomalous before every of their parents.

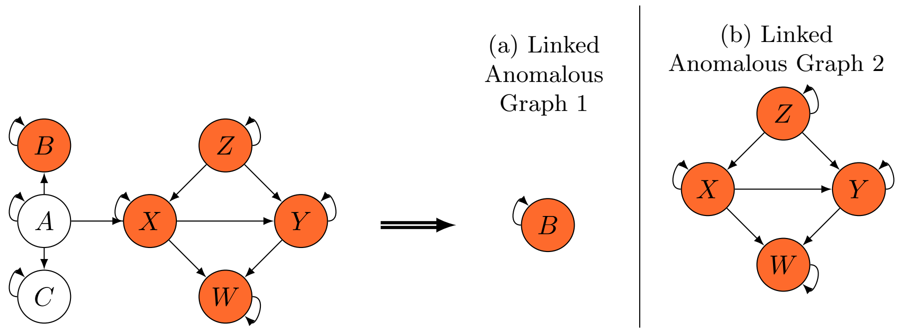


Figure 4: Step 1 of EasyRCA algorithm. The orange vertices are anomalous.

Definition 18 (Sub-root vertex)**.** *A sub-root vertex is root vertex in a linked anomalous graph.*

Definition 19 (Time defying vertex)**.** *Consider a linked anomalous graph $\mathcal{L}^i = (\mathcal{V}^i, \mathcal{E}^i)$. Y is a time defying vertex if and only if $\forall X \in Parents(Y, \mathcal{L}^i)$ the appearance time of the anomaly on Y precedes the appearance time of the anomaly on X.*

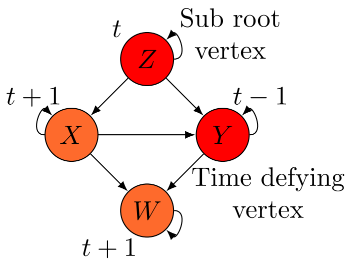


Figure 5: Step 2 of EasyRCA algorithm. The orange vertices are anomalous and the red ones are root causes.

Proposition 4 in (Assaad et al., 2023b) states that if a vertex is a sub-root or a time-defying vertex in a linked anomalous graph then it belongs to the set of root causes. However, this does not mean that every element in the set of root causes is necessarily a sub-root or a time-defying vertex.

In an ASCGL, there is at least one sub-root per linked anomalous graph. Therefore, it would be sufficient to eliminate the anomalies detected as sub-root vertices, to wait for some more data and to repeat computing new linked anomalous graphs and new sub-root vertices in order to detect iteratively every root cause. However, this method requires to wait and collect some data while some anomalies are not eliminated and thus is not fully efficient. This step is described in Figure 5.

In the case of general SCGs, sub-root vertices and time defying vertices are still root causes. Therefore, it remains interesting to detect them. However, unlike in the case of ASCGLs, there may exist cyclic linked anomalous graphs with no sub-root vertices as cycles do not have roots. This remark does not imply a modification of step 2 but it motivates the adaptation of the third step.

3.2.3 Step 3: Finding external intervention from data

Therefore, in step 3, EasyRCA proceeds to find the rest of the root causes by estimating the direct causal effect (Pearl, 2000) of X on Y in the normal regime, $\alpha_{Y_t, X_{t-\gamma_{xy}}}$, and the direct causal effect of X on Y in the anomalous regime, $\alpha_{Y_t, \tilde{X}_{t-\gamma_{xy}}}$, and comparing them. Indeed, if the estimation of the direct effect in the anomalous regime and in the normal regime are equal up to some estimation error, then the SCM has not been modified. Otherwise, if the direct effects are different, then the SCM has been modified which shows that the children Y has been intervened on and thus is a root cause.

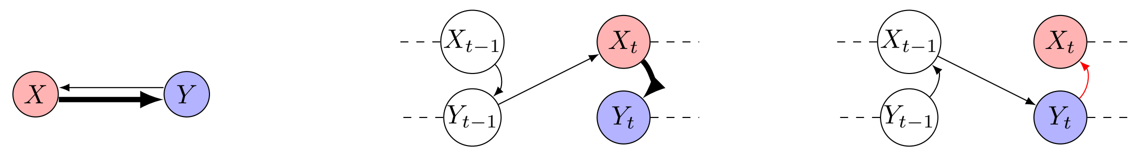
To estimate the direct causal effect expression, EasyRCA finds an adequate single-door adjustment set. Since the SCG is assumed to have no cycles of length greater than two in Assaad et al. (2023b), $Ancestors(Y, \mathcal{G}_f) \cap Descendants(Y, \mathcal{G}_s) = \{Y\}$. Thus, using only parents of Y allows to block every path between X and Y without risking to condition on a descendant of Y.

Note that using the direct causal effect, EasyRCA also distinguishes between parametric intervention and structural intervention. Given that $\alpha_{Y_t, X_{t-\gamma_{xy}}} \neq \alpha_{Y_t, \tilde{X}_{t-\gamma_{xy}}}$, if $\alpha_{Y_t, \tilde{X}_{t-\gamma_{xy}}} = 0$, one can conclude that there is a structural intervention on Y, otherwise there is a parametric intervention on Y.

The single-door adjustment set found by Assaad et al. (2023b) works in the context of Assumption 5 but it does not transpose easily to general SCGs. Therefore, in the next section, we study the identifiability of direct effects in general SCGs and we provide two adequate adjustment sets.

4 Direct effect in a general SCG

In Section 3.2.3 we discussed the need to compute the direct effect in the case of a *cyclic* SCG. However, it is easy to see with the example of Figure 6 that a direct effect is not always identifiable in the case of a cyclic SCG.



(a) A SCG. How to compute the direct effect of X_t to Y_t?

(b) A compatible FTCG.

(c) A compatible FTCG.

Figure 6: An example of an SCG for which the direct effect is not identifiable.

4.1 Identifiability result

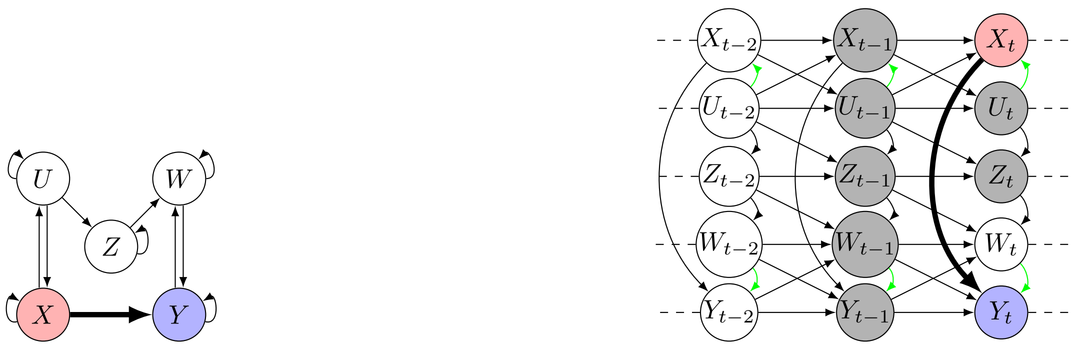
In this section, we present the identifiability result stating in which case a direct effect is or not identifiable. To do so, we first state some intermediate lemmas which are proven in Section A.5 of the Appendix. We start with a trivial lemma that will be needed for the the general identifiability result.

Lemma 1. *Let $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ be a SCG. Consider the direct effect of $X_{t-\gamma_{xy}}$ on Y_t given by $\alpha_{Y_t, X_{t-\gamma_{xy}}}$. If $X \notin Parents(Y, \mathcal{G}_s)$ then $\alpha_{Y_t, X_{t-\gamma_{xy}}}$ is identifiable.*

Lemma 2. *Let $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ be a FTCG, $X_{t-\gamma_{xy}} \neq Y_t \in \mathcal{V}_f$ and $\pi_f =< V^1_{t^1}, \dots, V^n_{t^n} >$ a path from $X_{t-\gamma_{xy}}$ to Y_t in $\mathcal{G}_f$. If $t_{max}(\pi_f) > t$ then π_f is passively blocked by any $\mathcal{Z}_f \subseteq \mathcal{V}_f$ such that $\mathcal{Z}_f \cap \{V_{t'} \in \mathcal{V}_f | t' > t\} = \emptyset$. If $t_{min}(\pi_f) < t - \gamma_{xy}$ then π_f is manually blocked by any $\mathcal{Z}_f \subseteq \mathcal{V}_f$ such that $\{V_{t'} \in \mathcal{V}_f | t - \gamma_{max} \leq t' < t\} \subseteq \mathcal{Z}_f$.*

Lemma 1 and Lemma 2 respectively show that the case where $X \notin Parents(Y, \mathcal{G}_s)$ is trivial and that the cases where $t_{min}(\pi_f) < t - \gamma_{xy}$ and $t_{max}(\pi_f) > t$ are trivial. Thus we will consider $X \in Parents(Y, \mathcal{G}_s)$ and $t - \gamma_{xy} \leq t_{min}(\pi_f) \leq t_{max}(\pi_f) \leq t$ in the following Lemmas.

In these cases, one might think that to block all activated non-direct paths between $X_{t-\gamma_{xy}}$ and Y_t, it is simply sufficient to adjust on all vertices in the FTCG which do not temporally succeed the effect Y_t and have compatible vertices on an activated path between X and Y in the SCGs. In Figure 7 we give an example where this is not true. In particular, in the FTCG in Figure 7b which is compatible with the SCG in Figure 7a, one can visually check that all gray vertices block all activated non-direct paths between X_t and Y_t and do not create any artificial bias since none of the gray vertices is a descendant of Y_t. Notice that this adjustment set is still valid if any of the black edges is omitted and if the orientation of the green edges is inverted and the new FTCG is still compatible with the SCG. One might also notice that W_t do not create any artificial bias in the FTCG in Figure 7b so it can be added to the adjustment set. However the general validity does not hold if we add W_t to the adjustment set because there exist another FTCG compatible with the same SCG in which the green edge between W and Y has been inverted such that W_t is a descendant on Y_t. Therefore, if only the SCG is known, W_t should not be used in the adjustment set. It turns out that this observation can be generalized. In the following we start, by formally defining an adjustment set for any SCGs that share the same characteristics as the adjustment set of Figure 7 and then we give a lemma that shows the usefulness of such a set.



(a) A SCG.

(b) A compatible FTCG.

Figure 7: An example of SCG and its compatible FTCG where red and blue vertices respectively represent the cause and the effect we are interested in, the thick edge corresponds to the the edge between them, and the gray vertices in the FTCG correspond to an adjustment set. In the given FTCG, one could choose to add W_t to the adjustment set to block every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t. However, there exist another FTCG compatible with the SCG in (a) such that W_t is a descendant on Y_t. Therefore, if only the SCG is known, W_t should not be used in the adjustment set.

Definition 20 (A first finite adjustment set)**.** *Consider a SCG $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$, a maximal lag γ_{max}, two vertices X and Y with $X \in Parents(Y, \mathcal{G}_s)$ and a lag γ_{xy}. $\mathcal{A}_{\leq t}$ is defined as the set of instants of non-descendants of Y smaller or equal to t and greater or equal to $t - \gamma_{max}$ except $X_{t-\gamma_{xy}}$, i.e.*

$$\mathcal{A}_{\leq t} = \{V_{t'} | V \in \mathcal{V}_s \backslash Descendants(Y, \mathcal{G}_s), t - \gamma_{max} \leq t' \leq t\} \backslash \{X_{t-\gamma_{xy}}\},$$

and $\mathcal{D}_{<t}$ is defined as the set of instants of descendants of Y strictly smaller than t and greater or equal to $t - \gamma_{max}$ except $X_{t-\gamma_{xy}}$, i.e.

$$\mathcal{D}_{<t} = \{V_{t'} | V \in Descendants(Y, \mathcal{G}_s), t - \gamma_{max} \leq t' < t\} \backslash \{X_{t-\gamma_{xy}}\}.$$

$\mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$ is called an adjustment set relative to $(X_{t-\gamma_{xy}}, Y_t)$.

Lemma 3. *Let $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ be a SCG and γ_{max} a maximal lag. For every non-direct walk $\pi_s =< V^1, \dots, V^n >$ between X and Y such that $< V^2, \dots, V^{n-1} > \not\subseteq Descendants(Y, \mathcal{G}_s)$, every compatible path π_f (i.e., $\pi_f \in \phi^{-1}(\pi_s)$) from $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} \geq 0$ can be blocked by the adjustment set $\mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$ defined in Definition 20.*

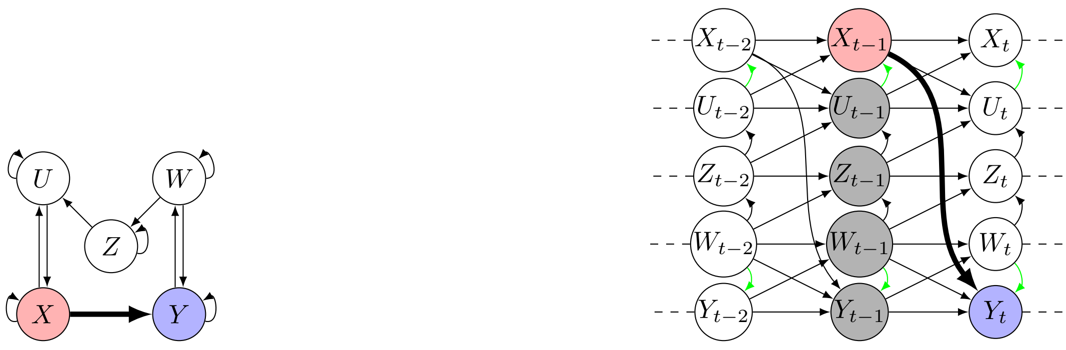
For the same SCG, the direct effect between X and Y can depend on the lag between X and Y. Therefore, for a same SCG and X, Y the direct effect from $X_{t-\gamma_{xy}}$ to Y_t can be identifiable if γ_{xy} even though it is not identifiable if $\gamma_{xy} = 0$.

Furthermore, the only graph with a path made of descendants of Y in which the direct effect $\alpha_{Y_t, X_{t-\gamma_{x,y}}}$ is always identifiable is Figure 19c. This can be explained by the fact that the path is initially passively blocked as Z is a collider and this is generalized in Lemma 7.

Among the remaining figures, the ones in which $\alpha_{Y_t, X_{t-\gamma_{x,y}}}$ is identifiable if and only if $\gamma_{xy} > 0$ are those in which the path has an edge strictly pointing to the left, towards X (*e.g.*, $X \leftarrow Z$ in Figures 19e and 19f, $Z \leftarrow Y$ in Figure 19g, $U \leftarrow Z$ in Figures 20f and 20g and $Z \leftarrow W$ in Figures 20f and 20h). This observation is generalized in Lemmas 4 and 6. Lastly, Lemma 5 focuses on the case $X \rightleftarrows Y$.

Lemma 4. *Let $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ be a SCG and γ_{max} a maximal lag. For every non-direct walk $\pi_s =< V^1, \dots, V^n >$ from X to Y such that $\exists 1 \leq i < n$, $V^i \leftarrow V^{i+1}$ (i.e., not $V^i \rightarrow V^{i+1}$ and not $V^i \rightleftarrows V^{i+1}$), every compatible path π_f (i.e., $\pi_f \in \phi^{-1}(\pi_s)$) from $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} > 0$ can be blocked by $\mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$ because $\mathcal{Z}_f \cap \mathcal{D}_{\geq t} = \emptyset$ where:*

- *$\mathcal{A}_{\leq t}$, $\mathcal{D}_{<t}$ are defined in Definition 20 and*



(a) A SCG.

(b) A compatible FTCG.

Figure 8: An example of SCG and its compatible FTCG where red and blue vertices respectively represent the cause and the effect we are interested in, the thick edge corresponds to the the edge between them, and the gray vertices in the FTCG correspond to an adjustment set. In the given FTCG, one could choose to add X_t, U_t, Z_t and W_t to the adjustment set to block every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t. However, there exist another FTCG compatible with the SCG in (a) such that X_t, U_t, Z_t and W_t are a descendant on Y_t. Therefore, if only the SCG is known, X_t, U_t, Z_t and W_t should not be used in the adjustment set.

- $\mathcal{D}_{\geq t}$ *is the set of instants of descendants of* Y *greater or equal to* t*, i.e.* $\mathcal{D}_{\geq t} = \{V_{t'} | V \in Descendants(Y, \mathcal{G}_s),\ t' \geq t\}$.

Note that $\pi_s =< V^1, \ldots, V^n >$ *from* X *to* Y *is non-direct,* $X \in Parents(Y, \mathcal{G}_s)$ *by Lemma 1 and* $\exists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$ *implies that* $n \geq 3$.

Now we give the complementary of Lemma 4 is the case where $X \rightleftarrows Y$.

Lemma 5. *Let* $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ *be a SCG and* γ_{max} *a maximal lag. For every non-direct walk* $\pi_s =< V^1, \ldots, V^n >$ *from* X *to* Y *where* $\exists 1 < i < n,\ V^i = Y$*, every compatible path* π_f *(i.e.,* $\pi_f \in \phi^{-1}(\pi_s)$*) from* $X_{t-\gamma_{xy}}$ *to* Y_t *for* $\gamma_{xy} > 0$ *can be blocked by* $\mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$ *(notice* $\mathcal{Z}_f \cap \mathcal{D}_{\geq t} = \emptyset$*) where* $\mathcal{A}_{\leq t},\ \mathcal{D}_{<t}$ *are defined in Definition 20 and* $\mathcal{D}_{\geq t}$ *is defined in Lemma 4.*

In the following two lemmas, we show that the conditions given in Lemma 3 and 4 are not only sufficient for blocking all non-direct active paths but they are also necessary.

Lemma 6. *Let* $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ *be a SCG. Consider the direct effect of* $X_{t-\gamma_{xy}}$ *on* Y_t *given by* $\alpha_{Y_t, X_{t-\gamma_{xy}}}$*. If there exists a path* $\pi_s =< V^1, \ldots, V^n >$ *from* X *to* Y *with* $< V^2, \ldots, V^{n-1} > \subseteq Descendants(Y, \mathcal{G}_s)$ *and* $\nexists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$ *and either* $n \geq 3$ *or* $X \in Descendants(Y, \mathcal{G}_s)$ *and* $\exists C \in Cycles(X, \mathcal{G}_s)$ *with* $Y \notin C$ *then* $\alpha_{Y_t, X_{t-\gamma_{xy}}}$ *is not identifiable.*

Lemma 7. *Let* $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ *be a SCG. Consider the direct effect of* $X_{t-\gamma_{xy}}$ *on* Y_t *given by* $\alpha_{Y_t, X_{t-\gamma_{xy}}}$*. If* $\gamma_{xy} = 0$ *and there exists an active non-direct path* $\pi_s =< V^1, \ldots, V^n >$ *from* X *to* Y *in* $\mathcal{G}_s$ *with* $< V^2, \ldots, V^{n-1} > \subseteq Descendants(Y, \mathcal{G}_s)$ *then* $\alpha_{Y_t, X_{t-\gamma_{xy}}}$ *is not identifiable.*

Since Lemmas 3,4 and 5 consider walks in SCGs while Lemma 6 and 7 consider paths in SCGs, in the following we provide a list of properties to reconcile these two notions. Let $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ be a SCG and $\pi_s =< V^1, \ldots, V^n >$ a walk from X to Y. $\pi'_s =< U^1, \ldots, U^m >$ such that $U^1 = V^1$ and $U^{k+1} = V^{max\{i | V^i = U^k\}+1}$ is called the subpath of π_s. It verifies the following properties.

Property 1 If π'_s is passively blocked by U^i then π_s is passively blocked by at least a descendant of U^i.

Property 2 If π'_s is direct then either π_s is direct or $< V^1, \ldots, V^{n-1} > \in Cycles(X, \mathcal{G}_s)$.

Property 3 If $< U^2, \ldots, U^{m-1} > \not\subseteq Descendants(Y, \mathcal{G}_s)$ then $< V^2, \ldots, V^{n-1} > \not\subseteq Descendants(Y, \mathcal{G}_s)$.

Property 4 If $m \geq 3$ and $\exists 1 \leq i < m,\ U^i \leftarrow U^{i+1}$ then $n \geq 3$ and $\exists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$.

Lemma 1 deals with the trivial case of identifiability, Lemma 3 together with Property 3 states that we will always be able to block paths in which some vertices are not descendants of Y, Lemmas 4 and 5 together with Property 2 and Property 4 show that in the case of positive lag (*i.e.*, $\gamma_{xy} > 0$) we can use this temporal information to block other specific paths, and lastly, Lemmas 6 and 7 prove that these identifiability criteria are necessary. Together, these lemmas give a set of necessary and sufficient conditions for a direct effect to be identifiable. This is summarized in the following theorem.

Theorem 1. *Let $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$ be a SCG, $X \neq Y \in \mathcal{V}_s$ and γ_{max} a maximal lag. Consider the direct effect of $X_{t-\gamma_{xy}}$ on Y_t given by $\alpha_{Y_t, X_{t-\gamma_{xy}}}$. $X \in Parents(Y, \mathcal{G}_s)$ and there exists an active path $\pi_s =< V^1, \dots, V^n >$ from X to Y in $\mathcal{G}_s$ such that $< V^2, \dots, V^{n-1} >\subseteq Descendants(Y, \mathcal{G}_s)$, and*

- *$\gamma_{xy} = 0$ and π_s is non-direct or*
- *$\gamma_{xy} > 0$, and*
  - *$n \geq 3$ and $\nexists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$ (i.e., $\forall 1 \leq i < n,\ V^i \rightarrow V^{i+1}$ or $V^i \rightleftarrows V^{i+1}$), or*
  - *$n = 2$ and $X \in Descendants(Y, \mathcal{G}_s)$ and $\exists C \in Cycles(X, \mathcal{G}_s)$ with $Y \notin C$.*

if and only if $\alpha_{Y_t, X_{t-\gamma_{xy}}}$ is not identifiable.

The proof is given in Section A.5 of the Appendix.

We give some examples of SCGs of with 3 vertices in Figure 19 and with 5 vertices in Figure 20 that are categorized by whether the direct effect $\alpha_{Y_t, X_{t-\gamma_{x,y}}}$ is always identifiable, identifiable if and only if $\gamma_{xy} > 0$ or never identifiable. Note that in Figure 20, for any SCG changing the edge type of the edges $U \rightleftarrows X$ or $W \rightleftarrows Y$ would either keep the same identifiability result or make the causal effect identifiable or identifiable if and only if $\gamma_{xy} > 0$. Which means that at least for SCGs with the same skeleton as the skeletons of Figures 19 and 20 and such that $X \rightarrow Y$ (with a strict right arrow) the number of cases of non identifiability is fewer than the number of cases of identifiability.

4.2 Finding adjustment sets

4.2.1 Soundness results

Theorem 1 gives a graphical criterion to determine whether a direct effect is identifiable from a SCG $\mathcal{G}_s$. For a FTCG $\mathcal{G}_f$ and two vertices $X_{t-\gamma xy}$ and Y_t it is necessary and sufficient to know a finite set of vertices $\mathcal{Z}_f$ such that $\mathcal{Z}_f \cap (Descendants(Y_t, \mathcal{G}_f) \cup \{X_{t-\gamma_{xy}}, Y_t\}) = \emptyset$ which blocks every non-direct path from $X_{t-\gamma xy}$ to Y_t in order to identify the direct effect of $X_{t-\gamma_{xy}}$ on Y_t. Thus, for a SCG $\mathcal{G}_s$, a maximal lag γ_{max}, two vertices X and Y and a lag γ_{xy} one needs to find a set of vertices $\mathcal{Z}_f$ such that $\mathcal{Z}_f \cap (Descendants(Y_t, \mathcal{G}_f) \cup \{X_{t-\gamma_{xy}}, Y_t\}) = \emptyset$ which blocks every non-direct path from $X_{t-\gamma xy}$ to Y_t in every FTCG $\mathcal{G}_f$ compatible with $\mathcal{G}_s$ of maximal lag at most γ_{max}.

Corollary 1. *Consider a SCG $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$, a maximal lag γ_{max}, two vertices X and Y with $X \in Parents(Y, \mathcal{G}_s)$ and a lag γ_{xy}. Suppose the direct effect of $X_{t-\gamma_{xy}}$ on Y_t is identifiable following Theorem 1. $\mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$ defined in Definition 20 and used in the proof of Theorem 1 verifies:*

- *$\mathcal{Z}_f \cap (Descendants(Y_t, \mathcal{G}_f) \cup \{X_{t-\gamma_{xy}}\}) = \emptyset$, and*
- *$\mathcal{Z}_f$ blocks every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t in every compatible FTCG $\mathcal{G}_f$ of maximal lag at most γ_{max}.*

which allows to estimate the direct effect $\alpha_{Y_t, X_{t-\gamma_{xy}}}$.

Proof. The proof of the backward implication of Theorem 1 proves this corollary. □

However, there may exist multiple such sets and in practice each one does not induce the same estimation error. Therefore, it is interesting to search for other such set in order to use the best one to optimize the identification of the direct effect. In the following, we give another (smaller) adjustment set that is sufficient for estimating direct effects when they are identifiable.

Definition 21 (A second finite adjustment set). *Consider a SCG $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$, a maximal lag γ_{max}, two vertices X and Y with $X \in Parents(Y, \mathcal{G}_s)$ and a lag γ_{xy}. Consider the following sets:*

- $\mathcal{D}_{t'}^{Anc(Y)} = \{V_{t'} | V \in Ancestors(Y, \mathcal{G}_s) \cap Descendants(Y, \mathcal{G}_s),\ t' \in [t - \gamma_{max}, t[\} \backslash \{X_{t-\gamma_{xy}}\}$
- $\mathcal{A}_{t'}^{Anc(Y)} = \{V_{t'} | V \in Ancestors(Y, \mathcal{G}_s) \backslash Descendants(Y, \mathcal{G}_s),\ t' \in [t - \gamma_{max}, t]\} \backslash \{X_{t-\gamma_{xy}}\}$
- $\mathcal{Z}_f = \mathcal{D}_{t'}^{Anc(Y)} \cup \mathcal{A}_{t'}^{Anc(Y)}$

$\mathcal{Z}_f$ is called a adjustment set relative to $(X_{t-\gamma_{xy}}, Y_t)$.

Proposition 1. *Consider a SCG $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$, a maximal lag γ_{max}, two vertices X and Y with $X \in Parents(Y, \mathcal{G}_s)$ and a lag γ_{xy}. Suppose the direct effect of $X_{t-\gamma_{xy}}$ on Y_t is identifiable following Theorem 1. The adjustment set $\mathcal{Z}_f$ relative to $(X_{t-\gamma_{xy}}, Y_t)$ defined in Definition 21 verifies:*

- *$\mathcal{Z}_f \cap (Descendants(Y_t, \mathcal{G}_f) \cup \{X_{t-\gamma_{xy}}\}) = \emptyset$, and*
- *$\mathcal{Z}_f$ blocks every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t in every compatible FTCG $\mathcal{G}_f$ of maximal lag at most γ_{max}.*

which allows to estimate the direct effect $\alpha_{Y_t, X_{t-\gamma_{xy}}}$.

The proof is given in Section A.5 of the Appendix.

4.2.2 Discussion towards completeness

The adjustment sets described in Section 4.2.1 are sound, which means that they allow to estimate every identifiable direct effect The ultimate goal would be to obtain a characterization of every such finite set $\mathcal{Z}_f \subseteq \mathcal{V}_f$ which allow for the identification of the direct effect of $X_{t-\gamma_{xy}}$ on Y_t. This would allow to try and find an optimal one which induces the smallest estimation error. While this would be the ideal result, since the SCGs offer very little information on lags we decide to firstly search for sets of the form $\mathcal{Z}_f = (\{V_{t'} | V \in \mathcal{A}, t' \in \tau_{\mathcal{A}}\} \cup \{V_{t'} | V \in \mathcal{D}, t' \in \tau_{\mathcal{D}}\}) \backslash \{X_{t-\gamma_{xy}}\}$ with

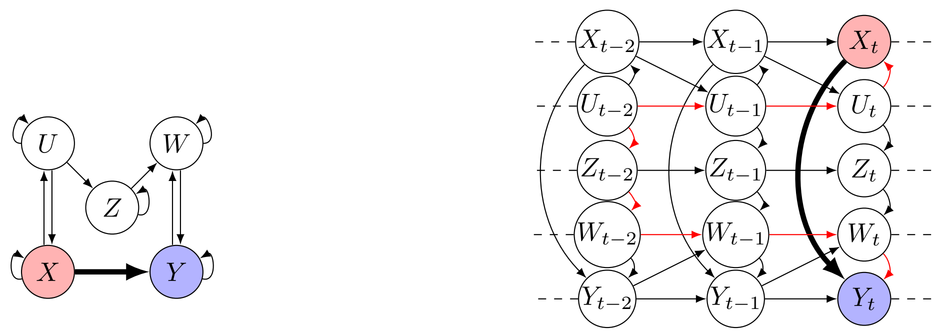
- $\mathcal{A} \subseteq \mathcal{V}_s \backslash Descendants(Y, \mathcal{G}_s)$, $\tau_{\mathcal{A}} = [t_{inf}, t]$.
- $\mathcal{D} \subseteq Descendants(Y, \mathcal{G}_s)$, $\tau_{\mathcal{D}} = [t_{inf}, t[$.

Indeed, one can notice that no $V_{t'}$ with $t' > t$ is necessary to block a path from $X_{t-\gamma_{xy}}$ to Y_t as Lemma 2 shows that every path π_f in a compatible FTCG with $t_{max}(\pi_f) > t$ has a collider $\rightarrow V_{t_{max}(\pi_f)} \leftarrow \in \pi_f$ which will not be activated if $\mathcal{Z}_f \cap \{V_{t'} | V \in \mathcal{V}_s,\ t' > t\} = \emptyset$.

Moreover, Figure 9 gives the intuition that paths in FTCGs can go back infinitely back in time through unblocked cycles. Therefore, when manually blocking a path $\pi_s =< V^1, \ldots, V^n >$ in a SCG $\mathcal{G}_s$ with V^i, one should either take $t_{inf} = t - \infty$ or one should also be careful to block every cycle intersecting a border of the path (*i.e.*, $\bigcup_{1 \leq j < i} Cycle(V^j, \mathcal{G}_s)$ or $\bigcup_{i < j \leq n} Cycle(V^j, \mathcal{G}_s)$). Since we are looking for the finite sets $\mathcal{Z}_f$ the second solution is best and one should always have give a special care to borders.

However, even with this in mind, one can notice that in Figure 10 conditioning on relevant instant of $\{W\}$ can be sufficient to block the red path as the left border $< X, U, Z >$ intersects no cycle but it is necessary to take $t_{inf} = t - \gamma_{xy} - (|\mathcal{V}_s| - 1) * \gamma_{max}$. Unfortunately, this forces $\mathcal{Z}_f$ to grow linearly in the size of $\mathcal{V}_s$ even if most vertices are in no path between X and Y which is not satisfactory. Therefore, we have decided to search for sets with $t_{inf} \geq t - \gamma_{xy} - \gamma_{max}$ despite the fact that this potentially forces $\mathcal{A} \cup \mathcal{D} \cup \mathcal{X} \cup \mathcal{Y}$ to be bigger.

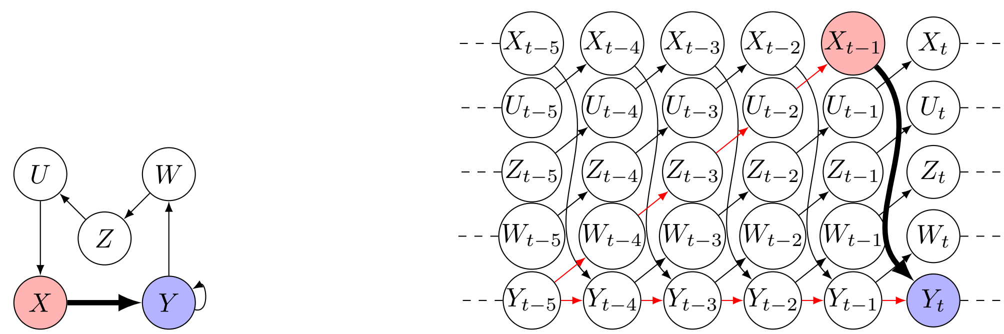
The remaining figures show some special cases in a hope to give some intuition on what characteristics a complete single-door criterion should have. Figure 11 shows that if $X \rightarrow X$ then it is mandatory to condition on relevant time instant of X and Figure 12 shows that if $Y \rightarrow Y$ and $\gamma_{xy} > 0$ then it is mandatory to condition on relevant time instant of Y. Figure 13 and 14 give some insights on the facts that $X \in Descendants(Y, \mathcal{G}_s) \implies Parents(X, \mathcal{G}_s) \subseteq \{V | V_{t'} \in \mathcal{Z}_f\}$ and $\gamma_{xy} > 0\ \ \&\ \ Cycle(Y, \mathcal{G}_s) \neq \emptyset \implies \{Y\} \cup Parents(Y, \mathcal{G}_s) \subseteq \{V | V_{t'} \in \mathcal{Z}_f\}$. Figure 15 shows that in the case of $\gamma_{xy} > 0$ and a path made of descendants of Y with a strict left arrow, one should surely block this path by conditioning on a vertex with



(a) A SCG (see Figure 20a). The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} \geq 0$ is identifiable by Theorem 1.

(b) A compatible FTCG. The path in red gives the intuition that a path can go infinitely back in time.

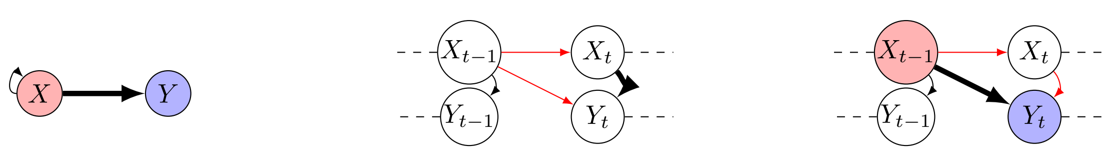
Figure 9: In this example, considering the SCG, one could choose to condition on relevant time instants of Z to block every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t. However, in this case and many others, the set of relevant time instants of Z is infinite. Therefore, one should also condition on a border $i.e. < X, U, Z >$ or $< Z, W, Y >$.



(a) A SCG. The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} > 0$ is identifiable by Theorem 1.

(b) A compatible FTCG. The path in red gives the intuition that even with a cycle-free border (here $< X, U, Z >$), a path can go back in time up to the instant $t - (|\mathcal{V}_s| - 1) * \gamma_{max}$.

Figure 10: In this example, considering the SCG, one could choose to condition on relevant time instants of W to block every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t. This does not cause the same problem as in Figure 9 as $< X, U, Z >$ intersects no cycle. In this extreme case, the set of relevant time instants of W is $[t - (|\mathcal{V}_s| - 1) * \gamma_{max}, t]$. It is possible to see that $t - (|\mathcal{V}_s| - 1) * \gamma_{max}$ is the smallest relevant time instant possible if every cycle of a border is blocked.



(a) A SCG. The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} \geq 0$ is identifiable by Theorem 1.

(b) A compatible FTCG with $\gamma_{xy} = 0$

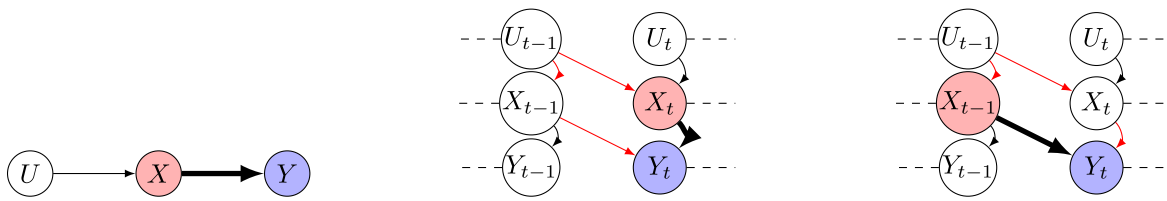
(c) A compatible FTCG with $\gamma_{xy} > 0$

Figure 11: Example of a self-loop on X $i.e.$, $X \to X$. In order to estimate the direct effect (in bold), it is necessary to block the secondary path (in red) which can only be done by conditioning on other time instants of X.

(a) A SCG. The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} \geq 0$ is identifiable by Theorem 1.

(b) A compatible FTCG with $\gamma_{xy} > 0$

Figure 12: Example of a self-loop on Y *i.e.*, $Y \to Y$. In order to estimate the direct effect (in bold), it is necessary to block the secondary path (in red) which can only be done by conditioning on other time instants of Y.

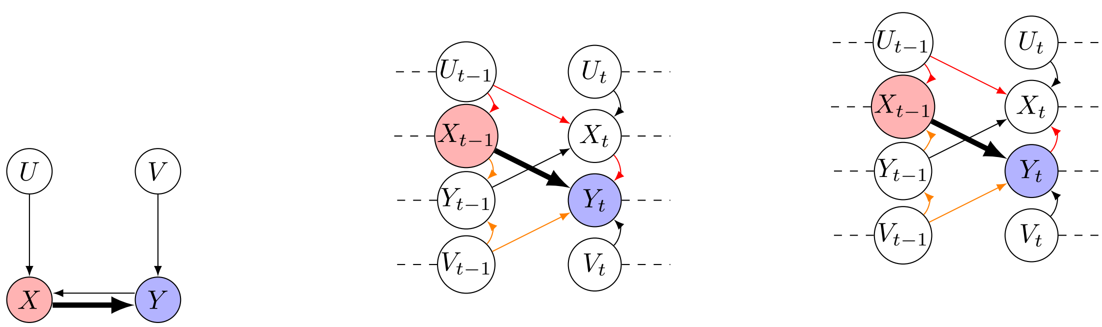


(a) A SCG. The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} \geq 0$ is identifiable by Theorem 1.

(b) A compatible FTCG with $\gamma_{xy} = 0$. The red secondary path can be blocked only by U_{t-1} or X_{t-1}.

(c) A compatible FTCG with $\gamma_{xy} > 0$. The red secondary path can be blocked only by U_{t-1} or X_t.

Figure 13: Example in the presence of a parent of X. In order to estimate the direct effect (in bold), it is necessary to block the secondary path (in red) which is only possible by conditioning on X or on every one of its parents.

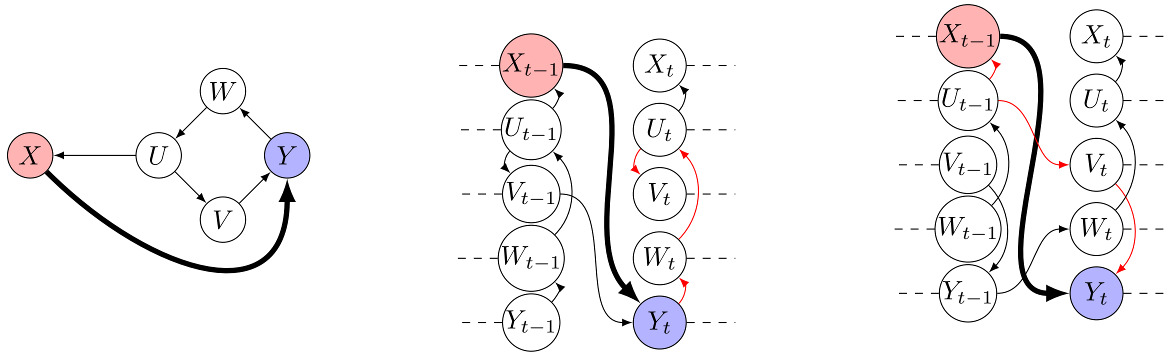


(a) A SCG. The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} > 0$ is identifiable by Theorem 1.

(b) A compatible FTCG. The red secondary path can be blocked only by U_{t-1}. The orange one can be blocked by Y_{t-1} and V_{t-1}.

(c) A compatible FTCG. The red secondary path can be blocked by U_{t-1} *and* X_t. The orange secondary path can be blocked by Y_{t-1} *or* V_{t-1}.

Figure 14: Example of parents of X and parents of Y in the presence of two opposite arrows between X and Y. In order to estimate the direct effect (in bold), it is necessary to block the secondary paths (in red and orange) which is only possible by conditioning on the parents of X and Y.



(a) A SCG. The direct effect of $X_{t-\gamma_{xy}}$ to Y_t for $\gamma_{xy} > 0$ is identifiable by Theorem 1.

(b) A compatible FTCG. The path in red shows $V_t \in Descendants(Y_t, \mathcal{G}_f)$.

(c) A compatible FTCG. The path in red can only be blocked by conditioning on U_{t-1} or V_t.

Figure 15: Example of a path made of descendants of Y and with a strict left arrow. Figure 15b show that every vertex of the path at instant t can be a descendant of Y_t. Figure 15c on the other hand shows compatible paths can be blocked by past instants of the vertex with an outgoing strict left arrow.

an outgoing strict left arrow. However, this example does not give any intuition on which such vertex to choose if there exists multiple ones. Lastly, a complex task is to block every path at once. Some paths as in Figure 19c can only be passively blocked so one should make sure that the vertices chosen to block a path or its border does not open another path.

5 Conclusion

During this internship, I have started to adapt the root cause analysis algorithm EasyRCA to remove the Assumption 5 which states that the system is represented by a ASCGL. In order to do so, the main work has been developing a new graphical criterion for the identifiability of direct effects in linear dynamic structural causal models given any general summary causal graph. Theorem 1 has important ramifications to the theory and practice of observational studies in dynamic systems. It implies that the key to graphical identifiability of the direct effect of $X_{t-\gamma_{xy}}$ on Y_t from summary causal graphs lies not only in finding a set non-descendant of Y in the summary causal graph that are capable of blocking paths between X and Y but also some descendant of Y in case $\gamma_{xy} > 0$. Furthermore, Proposition 1 gives a possible adjustment set that satisfies the identifiability result. Section 4.2.2 gives some insight on what to consider to find more adjustment sets.

These findings should be useful for many applications such as root cause identification in dynamic systems and it should open new research questions. Namely, for future works, it would be interesting to have a single-door criterion along with a completeness result describing every possible adjustment set. In addition, in this work, we considered only linear dynamic SCMs, however, in many real world applications, causal relations can be nonlinear so it would be interesting to extend this work to nonlinear SCMs and consider non parametric direct effects (Robins and Greenland, 1992; Pearl, 2001). Finally, as many other works, we assumed that the FTCG is acyclic but we think that this assumption can be relaxed, so it would be interesting to formally check the validity of our results for cyclic FTCGs.

In parallel, some implementing work is still needed to add these findings to EasyRCA. Moreover, this work allows EasyRCA to function on more systems but there is still some direct effects which are not identifiable. Thus an important future work would be to fill this gap and allow for root cause analysis even when some direct effects are not identifiable. This could be done by learning small subsets of the FTCG in advance using different causal discovery algorithms which have the most coherent assumptions regarding the small subset of the data of interest.

Bibliography

Andrew Arnold, Yan Liu, and Naoki Abe. Temporal causal modeling with graphical granger methods. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '07, page 66–75, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936097. doi: 10.1145/1281192.1281203.

Charles K. Assaad, Emilie Devijver, Eric Gaussier, and Ali Ait-Bachir. A mixed noise and constraint-based approach to causal inference in time series. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 453–468, Cham, 2021. Springer International Publishing. ISBN 978-3-030-86486-6.

Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *J. Artif. Int. Res.*, 73, apr 2022a. doi: 10.1613/jair.1.13428.

Charles K. Assaad, Emilie Devijver, and Eric Gaussier. Discovery of extended summary graphs in time series. In James Cussens and Kun Zhang, editors, *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pages 96–106. PMLR, 01–05 Aug 2022b.

Charles K. Assaad, Daria Bystrova, Julyan Arbel, Emilie Devijver, Eric Gaussier, and Wilfried Thuiller. Hybrids of constraint-based and noise-based algorithms for causal discovery from time series. *arXiv preprint arXiv:2306.08765*, 2023a.

Charles K. Assaad, Imad Ez-Zejjari, and Lei Zan. Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 8395–8404. PMLR, 25–27 Apr 2023b.

Ali Aït-Bachir, Charles K. Assaad, Christophe de Bignicourt, Emilie Devijver, Simon Ferreira, Eric Gaussier, Hosein Mohanna, and Lei Zan. Case studies of causal discovery from it monitoring time series. submitted, 2023.

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41:15:1–15:58, 2009.

David Maxwell Chickering. Learning equivalence classes of bayesian-network structures. *Journal of Machine Learning Research*, 2:445–498, 2002. ISSN 1532-4435. doi: 10.1162/153244302760200696.

Max Chickering, David Heckerman, and Chris Meek. Large-sample learning of bayesian networks is np-hard. *Journal of Machine Learning Research*, 5:1287–1330, 2004.

Joseph H. Connell. The influence of interspecific competition and other factors on the distribution of the barnacle chthamalus stellatus. *Ecology*, 42(4):710–723, 1961. doi: https://doi.org/10.2307/1933500.

O.D. Duncan. *Introduction to Structural Equation Models*. Studies in population. Academic Press, 1975. ISBN 9780122241505.

Simon Ferreira and Charles Assaad. Challenges of root cause identification for collective anomalies in time series given a summary causal graph. *When Causal Inference meets Statistical Analysis*, 2023a.

Simon Ferreira and Charles K. Assaad. Identifiability of direct effects from summary causal graphs. *arXiv preprint arXiv*, 2023b.

Patrick Forré and Joris Mooij. Markov properties for graphical models with cycles and latent variables. 10 2017.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10, 2019. ISSN 1664-8021. doi: 10.3389/fgene.2019.00524.

Arthur S. Goldberger. Structural equation methods in the social sciences. *Econometrica*, 40(6):979–1001, 1972. ISSN 00129682, 14680262.

Clive Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–38, 1969. URL `https://EconPapers.repec.org/RePEc:ecm:emetrp:v:37:y:1969:i:3:p:424-38`.

Clive W. J. Granger. Time series analysis, cointegration, and applications. *The American Economic Review*, 94(3):421–425, 2004. ISSN 00028282. URL `http://www.jstor.org/stable/3592936`.

Trygve Haavelmo. The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1): 1–12, 1943. ISSN 00129682, 14680262.

Kilian Holzinger, Henning Stubbe, Franz Biersack, Angela Gonzalez Mariño, Abdoul Kane, Francisco Fons Lluis, Zhang Haigang, Thomas Wild, Andreas Herkersdorf, and Georg Carle. Precise real-time monitoring of time-critical flows. CoNEXT '21, page 489–490, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450390989. doi: 10.1145/3485983.3493356.

Aapo Hyvärinen, Shohei Shimizu, and Patrik O. Hoyer. Causal modelling combining instantaneous and lagged effects: An identifiable model based on non-gaussianity. In *Proceedings of the 25th International Conference on Machine Learning*, ICML '08, pages 424–431, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-205-4. doi: 10.1145/1390156.1390210. URL `http://doi.acm.org/10.1145/1390156.1390210`.

Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O. Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(56):1709–1731, 2010.

Daniel Malinsky and David Danks. Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(1):e12470, 2018. doi: https://doi.org/10.1111/phc3.12470.

Yuan Meng, Shenglin Zhang, Yongqian Sun, Ruru Zhang, Zhilong Hu, Yiyin Zhang, Chenyang Jia, Zhaogang Wang, and Dan Pei. Localizing failure root causes in a microservice through causality inference. In *2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS)*, pages 1–10, 2020. doi: 10.1109/IWQoS49365.2020.9213058.

Roxana Pamfil, Nisara Sriwattanaworachai, Shaan Desai, Philip Pilgerstorfer, Konstantinos Georgatzis, Paul Beaumont, and Bryon Aragam. Dynotears: Structure learning from time-series data. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1595–1605. PMLR, 26–28 Aug 2020.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. ISBN 1558604790.

Judea Pearl. Graphs, causality, and structural equation models. *Sociological Methods & Research*, 27(2): 226–284, 1998. doi: 10.1177/0049124198027002004.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2000. ISBN 0-521-77362-8.

Judea Pearl. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, UAI'01, page 411–420, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc. ISBN 1558608001.

Jonas Peters, D. Janzing, and B. Schölkopf. Causal inference on time series using restricted structural equation models. In *Advances in Neural Information Processing Systems 26*, pages 154–162, 2013.

James M. Robins and Sander Greenland. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 3(2):143–155, 1992. ISSN 10443983.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 1388–1397. PMLR, 03–06 Aug 2020.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5(11), 2019. doi: 10.1126/sciadv.aau4996.

Shohei Shimizu, Takanori Inazumi, Yasuhiro Sogawa, Aapo Hyvärinen, Yoshinobu Kawahara, Takashi Washio, Patrik O. Hoyer, and Kenneth Bollen. Directlingam: A direct method for learning a linear non-gaussian structural equation model. *Journal of Machine Learning Research*, 12:1225–1248, 2011. ISSN 1532-4435. URL `http://dl.acm.org/citation.cfm?id=1953048.2021040`.

Peter Spirtes. Directed cyclic graphs, conditional independence, and non-recursive linear structural equation models. 1 1993. doi: 10.1184/R1/6491399.v1.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search.* MIT press, 2nd edition, 2000.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. *Probabilistic and Causal Inference*, 1990.

Hanzhang Wang, Zhengkai Wu, Huai Jiang, Yichao Huang, Jiamu Wang, Selcuk Kopru, and Tao Xie. Groot: An event-graph-based approach for root cause analysis in industrial settings. 2021. doi: 10.48550/ARXIV.2108.00344.

Ping Wang, Jingmin Xu, Meng Ma, Weilan Lin, Disheng Pan, Yuan Wang, and Pengfei Chen. Cloudranger: Root cause identification for cloud native systems. In *2018 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 492–502, 2018. doi: 10.1109/CCGRID.2018.00076.

Sewall Wright. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences of the United States of America*, 6(6): 320–332, 1920. ISSN 00278424.

Sewall Wright. Correlation and causation. *Journal of agricultural research*, 20(7):557–585, 1921.

Wen Zhou, Geoffrey Liu, Rayjean J. Hung, Philip C. Haycock, Melinda C. Aldrich, Angeline S. Andrew, Susanne M. Arnold, Heike Bickeböller, Stig E. Bojesen, Paul Brennan, Hans Brunnström, Olle Melander, Neil E. Caporaso, Maria Teresa Landi, Chu Chen, Gary E. Goodman, David C. Christiani, Angela Cox, John K. Field, Mikael Johansson, Lambertus A. Kiemeney, Stephen Lam, Philip Lazarus, Loïc Le Marchand, Gad Rennert, Angela Risch, Matthew B. Schabath, Sanjay S. Shete, Adonina Tardón, Shanbeh Zienolddiny, Hongbing Shen, and Christopher I. Amos. Causal relationships between body mass index, smoking and lung cancer: Univariable and multivariable mendelian randomization. *International Journal of Cancer*, 148(5):1077–1086, 2021. doi: https://doi.org/10.1002/ijc.33292.

A Appendix

A.1 Graphical notions

Definition 22 (Graph). *A (directed)* graph $\mathcal{G}$ *is a pair of sets* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *where* $\mathcal{V}$ *is a set of* vertices *and* $\mathcal{E} \subseteq \mathcal{V}^2$ *is a set of* edges.

An example of a graph is given in Figure 16a. We will often write $U \to V$ or $V \leftarrow U$ to represent $(U, V) \in \mathcal{E}$ and $(V, U) \notin \mathcal{E}$ as well as $U \rightleftarrows V$ to represent $(U, V), (V, U) \in \mathcal{E}$.

Definition 23 (Parents, Children, Adjacents, Ancestors, Descendants). *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *and for a vertex* $Y \in \mathcal{V}$ *we define the following sets:*

- $Parents(Y, \mathcal{G}) = \{U \in \mathcal{V} | U \to Y \in \mathcal{E} \text{ or } U \rightleftarrows Y \in \mathcal{E}\}$
- $Children(Y, \mathcal{G}) = \{U \in \mathcal{V} | Y \to U \in \mathcal{E} \text{ or } U \rightleftarrows Y \in \mathcal{E}\}$
- $Adjacents(Y, \mathcal{G}) = Parents(Y, \mathcal{G}) \cup Children(Y, \mathcal{G})$
- $Ancestors(Y, \mathcal{G}) = \bigcup_{n \in \mathbb{N}} P_n$ *where* $P_0 = \{Y\}$ *and* $P_{k+1} = \bigcup_{U \in P_k} Parents(U, \mathcal{G})$
- $Descendants(Y, \mathcal{G}) = \bigcup_{n \in \mathbb{N}} C_n$ *where* $C_0 = \{Y\}$ *and* $C_{k+1} = \bigcup_{U \in C_k} Children(U, \mathcal{G})$

Definition 24 (Skeleton). *To each graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ *can be associated a unique undirected graph* $\mathcal{G}_{sk} = (\mathcal{V}_{sk}, \mathcal{E}_{sk})$ *called* skeleton *of* $\mathcal{G}$ *such that:*

- $\mathcal{V}_{sk} = \mathcal{V}$, *and*
- $\mathcal{E}_{sk} = \{(U, V), (V, U) | (U, V) \in \mathcal{E}\}$

Edges of a skeleton are often noted without arrow heads i.e., $U - V$ *to represent* $(U, V), (V, U) \in \mathcal{E}_{sk}$

An example of a graph and its skeleton is given in Figure 16.

(a) Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{X, Y, Z\}$ and $\mathcal{E} = \{(X, Y), (Y, X), (Z, X), (Z, Y)\}$.

(b) Skeleton of $\mathcal{G}$.

Figure 16: An example of a graph and its skeleton.

Definition 25 (Walk). *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a* walk *between two vertices* X *to* Y *is an ordered sequence of vertices denoted as* $\pi = < V^1, \ldots, V^n >$ *where the first vertex is* X*, the last vertex is* Y *and each two consecutive vertices are adjacent, i.e.:*

- $V^1 = X$, *and*
- $V^n = Y$, *and*
- $\forall 1 \leq i < n$, $(V^i, V^{i+1}) \in \mathcal{E}$ *or* $(V^{i+1}, V^i) \in \mathcal{E}$

Definition 26 (Path). *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a* path *between two vertices* X *to* Y *is a walk* $\pi = < V^1, \ldots, V^n >$ *with no repeated vertices, i.e.:*

- $\forall 1 \leq i < j \leq n$, $V^i \neq V^j$

Definition 27 (Non-Direct Walk). *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a walk* π *from* X *to* Y *is said to be* non-direct *if:*

- $\pi \neq < X, Y >$, *or*

- $(Y, X) \in \mathcal{E}$

e.g., $< X \rightarrow Z \rightarrow Y >$, $< X \rightleftarrows Y >$ *and* $< X \leftarrow Y >$ *are non-direct.*

Definition 28 (Directed Walk)**.** *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a walk* $\pi =< V^1, \dots, V^n >$ *is said to be* directed *if:*

- $\forall 1 \leq i < n,\ (V^i, V^{i+1}) \in \mathcal{E}$ *(i.e.,* $V^i \rightarrow V^{i+1}$ *or* $V^i \rightleftarrows V^{i+1}$*)*

Definition 29 (Cycle)**.** *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a* cycle *is a directed walk* $\pi_s =< V^1, \dots, V^n >$ *from a vertex to itself with no repeated vertices except the endpoints, i.e.*

- $V^1 = V^n$*, and*
- $\forall 1 \leq i < j \leq n,\ V^i = V^j \implies i = 1$ *and* $j = n$*.*

The set of cycles with endpoints $Y \in \mathcal{V}$ *is written* $Cycles(Y, \mathcal{G})$*. A* directed acyclic graph *is called a DAG.*

Definition 30 (Collider)**.** *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a triple of variable* $(U, V, W) \in \mathcal{V}^3$ *is a* collider *if:*

- $U \rightarrow V \in \mathcal{E}$*, and*
- $V \leftarrow W \in \mathcal{E}$

An example of a collider is given in Figure 17.

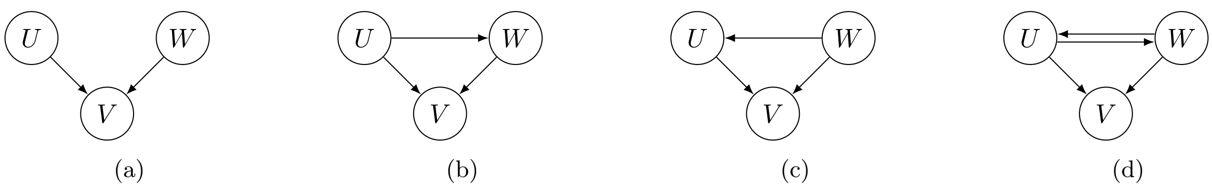


Figure 17: Examples of colliders.

Definition 31 (V-Structure)**.** *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a collider* $(U, V, W) \in \mathcal{V}^3$ *is a* v-structure *if* $U - W \notin \mathcal{E}_{sk}$*. The set of v-structures of a graph* $\mathcal{G}$ *is written* $V(\mathcal{G})$*.*

An example of a v-structure is given in Figure 17a.

A.2 σ-separation (Forré and Mooij, 2017)

Forré and Mooij (2017) introduced the concept of segments and showed how to use it to block paths in cyclic graphs. Even though I eventually decided not to use this work in my internship, it has been of great help to better understand cyclic graphs and it would be very interesting to express the results of Section 4 in term of segments.

Definition 32 (Strongly Connected Component)**.** *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, two vertices* X, Y *are strongly connected if there exists a directed path from* X *to* Y *and there exists a directed path from* Y *to* X*. The relation of strong connection is an equivalence relation. The equivalence class of a vertex* X *is called its strongly connected component and is written* $Sc(X)$*.*

Definition 33 (Path of Segments)**.** *In a graph* $\mathcal{G} = (\mathcal{V}, \mathcal{E})$*, a path* $\pi =< V^1, \dots, V^n >$ *can be uniquely written as a partition of sub-paths* $\pi =< \sigma^1, \dots, \sigma^m >$ *where* $\forall 1 \leq i \leq m,\ \exists p_i \leq q_i,\ \sigma_i =< V^{p_i}, \dots, V^{q_i} >$ *such that :*

- $\forall p_i \leq j \leq q_i,\ V^j \in Sc(V^{p_i})$*, and*
- $V^{p_i - 1} \notin Sc(V^{p_i})$ *and* $V^{q_i + 1} \notin Sc(V^{p_i})$

Such sub-paths are called segments and $\forall 1 \leq i \leq m$ *the left and right endpoints are written respectively* $\sigma_{i,l}$ *and* $\sigma_{i,r}$ *(i.e.,*$\sigma_i =< \sigma_{i,l}, \dots, \sigma_{i,r} >$*).*

Definition 34 (σ-Blocked Path, Forré and Mooij (2017))**.** *In a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a path of segments $\pi =< \sigma^1, \dots, \sigma^m >$ is said to be σ-blocked by a set of vertices $\mathcal{Z} \subseteq \mathcal{V}$ if:*

- $\sigma^{1,l} \in \mathcal{Z}$ *or* $\sigma^{m,r} \in \mathcal{Z}$*, or*
- $\exists 1 \leq i \leq m$ *such that* $\sigma^{i-1} \leftarrow \sigma^i$ *(or resp.* $\sigma^i \rightarrow \sigma^{i+1}$*) and* $\sigma^{i,l} \in \mathcal{Z}$ *(or resp.* $\sigma^{i,r} \in \mathcal{Z}$*), or*
- $\exists 1 \leq i \leq m$ *such that* $\sigma^{i-1} \rightarrow \sigma^i \leftarrow \sigma^{i+1}$ *and* $Descendants(\sigma^i, \mathcal{G}) \cap \mathcal{Z} = \emptyset$*.*

A path which is not σ-blocked is said to be σ-active. When the set $\mathcal{Z}$ is not specified, it is implicit that we consider $\mathcal{Z} = \emptyset$.

A.3 Causal discovery algorithms and datasets

Most source code of the tested algorithms are available online.

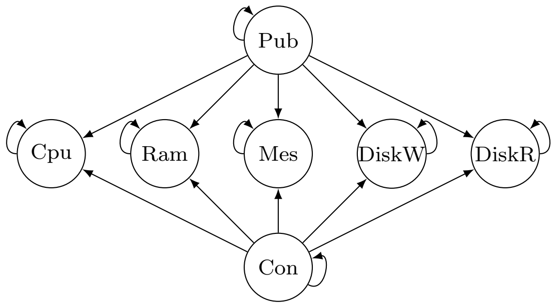
GCMVL	(Arnold et al., 2007)	?
Dynotears	(Pamfil et al., 2020)	`https://github.com/quantumblacklabs/causalnex/`
PCMCI$^+$	(Runge et al., 2019; Runge, 2020)	`https://github.com/jakobrunge/tigramite`
PCGCE	(Assaad et al., 2022b)	`https://github.com/ckassaad/PCGCE`
VarLiNGAM	(Hyvärinen et al., 2008; Hyvärinen et al., 2010)	`https://github.com/cdt15/lingam`
TiMINo	(Peters et al., 2013)	`http://web.math.ku.dk/~peters/code.html`
NBCB & CBNB	(Assaad et al., 2021, 2023a)	`https://github.com/ckassaad/NBCB`

In this section, we present the summary causal graph (the window causal graph and the extended summary causal graph are not available) for each case study. Note that all data points are collected using Nagios[2], an open-source software that monitors systems, networks and infrastructure, and which gives the timestamp according to the collection time which does not necessarily correspond to the real time of the value. In addition, on some of the case studies, the alignment between time series is not guaranteed as data collection is performed by different plugins with different starting times and different sampling rates (Holzinger et al., 2021).

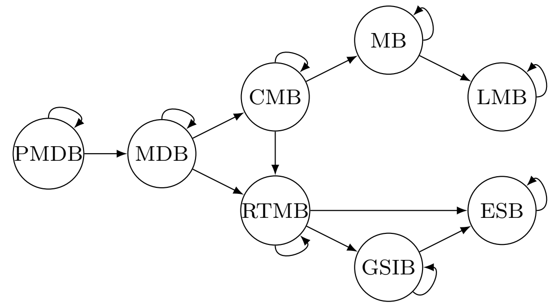
A.4 Examples

The following examples can help the reader gain more intuition of Theorem 1. It also shows that many cyclic summary causal graphs have an identifiable direct effect. Therefore, the work done during this internship can have some real implication in many cases.

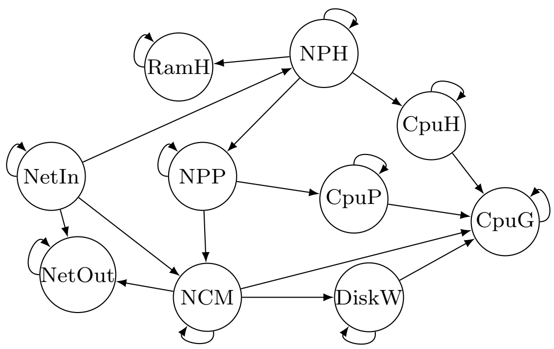
[2]`https://www.nagios.org/`



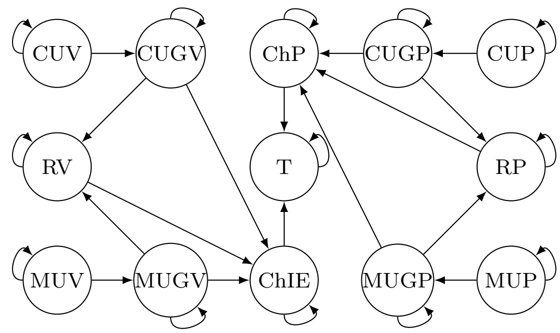
(a) MoM system.



(b) Ingestion IT monitoring system.



(c) Web-Activity.



(d) Antivirus-Activity.

Figure 18: Summary causal graphs for different datasets: MoM system based on Publish/Subscribe architecture (a), Ingestion IT monitoring system (b), Web-Activity (c) and Antivirus-Activity (d). Those summary causal graphs are constructed by EasyVista's system experts.

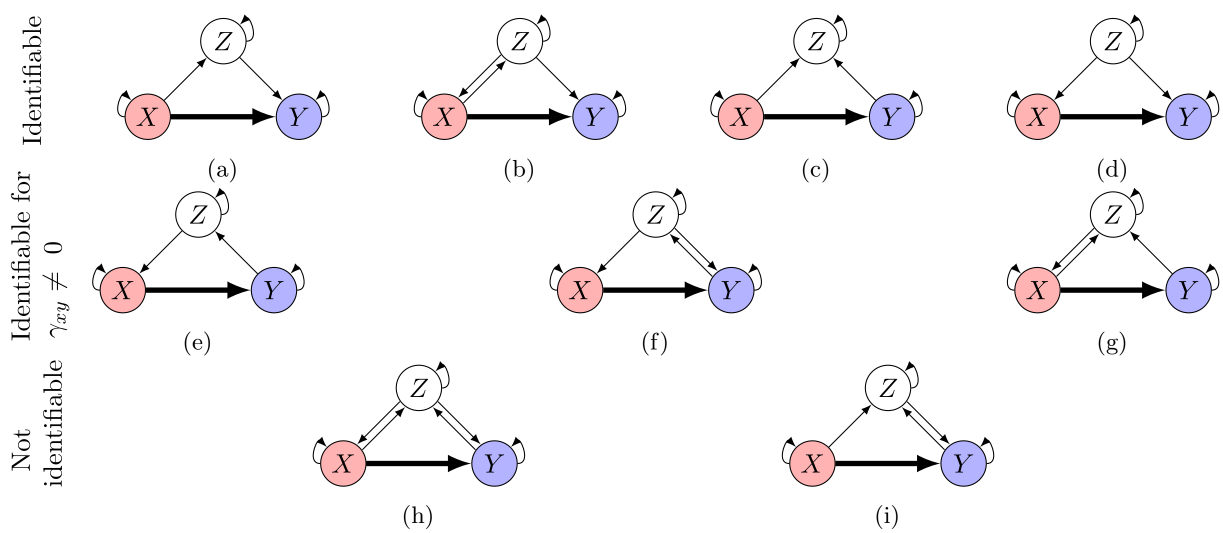


Figure 19: Examples of summary causal graphs with 3 vertices where in (a),(b),(c),(d) $\alpha_{Y_t,X_{t-\gamma_{x,y}}}$ is identifiable for all $\gamma_{x,y}$, in (e),(f),(g) $\alpha_{Y_t,X_{t-\gamma_{x,y}}}$ is identifiable only for $\gamma_{x,y} \neq 0$, and in (h),(i) $\alpha_{Y_t,X_{t-\gamma_{x,y}}}$ is non identifiable only all $\gamma_{x,y}$.

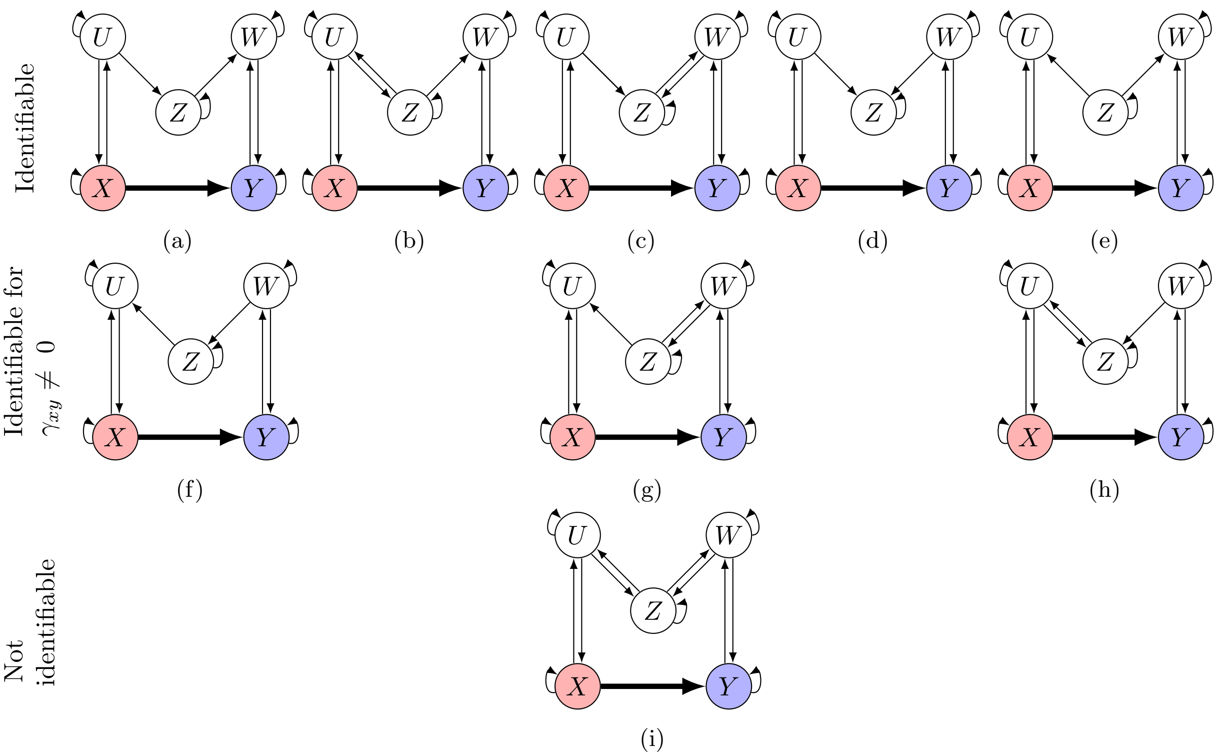


Figure 20: Examples of summary causal graphs with 5 vertices where in (a),(b),(c),(d),(e) $\alpha_{Y_t,X_{t-\gamma_{x,y}}}$ is identifiable for all $\gamma_{x,y}$, in (f),(g),(h) $\alpha_{Y_t,X_{t-\gamma_{x,y}}}$ is identifiable only for $\gamma_{x,y} \neq 0$, and in (i) $\alpha_{Y_t,X_{t-\gamma_{x,y}}}$ is non identifiable only all $\gamma_{x,y}$.

A.5 Proofs

Lemma 1. Suppose $X \notin Parents(Y, \mathcal{G}_s)$. Then $X_{t-\gamma_{xy}} \notin Parents(Y_t, \mathcal{G}_f)$ and the direct effect is equal to zero (*i.e.*, $\alpha_{Y_t,X_{t-\gamma_{xy}}} = 0$). □

Lemma 2. Let $\mathcal{G}_f = (\mathcal{V}_f, \mathcal{E}_f)$ be a FTCG, $X_{t-\gamma_{xy}} \neq Y_t \in \mathcal{V}_f$ and $\pi_f =< V^1_{t^1}, \dots, V^n_{t^n} >$ a path from $X_{t-\gamma_{xy}}$ to Y_t in $\mathcal{G}_f$.

- Suppose $t_{max}(\pi_f) > t$. Since $t - \gamma_{xy} \leq t < t_{max}(\pi_f)$ there exists $1 < i \leq j < n$ such that $t^{i-1} < t^i = t_{max}(\pi_f) = t^j > t^{j+1}$. Therefore, $V^{i-1}_{t^{i-1}} \rightarrow V^i_{t^i}$ and $V^j_{t^j} \leftarrow V^{j+1}_{t^{j+1}}$ in π_f. Thus, $\exists i \leq k \leq j$ such that $V^{k-1}_{t^{k-1}} \rightarrow V^k_{t^k} \leftarrow V^{k+1}_{t^{k+1}}$ in π_f and $t^k = t_{max}(\pi_f) > t$. In conclusion, π_f is passively blocked by any $\mathcal{Z}_f \subseteq \mathcal{V}_f$ such that $Descendants(V^k_{t^k}, \mathcal{G}_f) \cap \mathcal{Z}_f = \emptyset$ so by any $\mathcal{Z}_f$ such that $\mathcal{Z}_f \cap \{V_{t'} \in \mathcal{V}_f | t' > t\} = \emptyset$
- Suppose $t_{min}(\pi_f) < t - \gamma_{xy}$. Since $t_{min}(\pi_f) < t - \gamma_{xy} < t$ and $t^n = t$ there exists $1 < i < n$ such that $t^i < t \leq t^{i+1}$. Therefore, $V^i_{t^i} \rightarrow V^{i+1}_{t^{i+1}}$ in π_f and $t - \gamma_{max} \leq t^i < t \leq t^{i+1}$ so $V^i_{t^i} \in \{V_{t'} \in \mathcal{V}_f | t - \gamma_{max} \leq t' < t\}$. In conclusion, π_f is manually blocked by any $\mathcal{Z}_f \subseteq \mathcal{V}_f$ such that $V^i_{t^i} \in \mathcal{Z}_f$ so by any $\mathcal{Z}_f$ such that $\{V_{t'} \in \mathcal{V}_f | t - \gamma_{max} \leq t' < t\} \subseteq \mathcal{Z}_f$.

□

Lemma 3. Suppose there exists a non-direct walk $\pi_s =< V^1, \dots, V^n >$ between X and Y such that $< V^2, \dots, V^{n-1} > \not\subseteq Descendants(Y, \mathcal{G}_s)$. Then, take $\pi_f =< V^1_{t^1}, \dots, V^n_{t^n} >$ from $X_{t-\gamma_{xy}}$ to Y_t compatible with π_s (*i.e.*, $\pi_f \in \phi^{-1}(\pi_s)$). Take $j = max\{1 < i < n | V^i \notin Descendants(Y, \mathcal{G}_s)\}$. Notice $V^j_{t^j} \notin Descendants(Y, \mathcal{G}_s)$ and $V^{j+1}_{t^{j+1}} \in Descendants(Y, \mathcal{G}_s)$ so $V^j_{t^j} \rightarrow V^{j+1}_{t^{j+1}} \in \pi_f$. Therefore, since $t - \gamma_{xy} \leq t_{min}(\pi_f) \leq t_{max}(\pi_f) \leq t$ by Lemma 2, $t - \gamma_{xy} \leq t^j \leq t$ and thus π_f is manually blocked by $V^j_{t^j} \in \mathcal{A}_{\leq t} \subseteq \mathcal{Z}_f$. □

Lemma 4. Let $\pi_s =< V^1, \dots, V^n >$ be a non-direct walk between X and Y such that $< V^2, \dots, V^{n-1} >\subseteq Descendants(Y, \mathcal{G}_s)$ and $\gamma_{xy} > 0$ and $\exists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$. Then, take $\pi_f =< V^1_{t^1}, \dots, V^n_{t^n} >$ from $X_{t-\gamma_{xy}}$ to Y_t compatible with π_s (*i.e.*, $\pi_f \in \phi^{-1}(\pi_s)$) and $1 \leq i < n$ such that $V^i_{t^i} \leftarrow V^{i+1}_{t^{i+1}}$. $t_{min}(\pi_f) \geq t - \gamma_{xy}$ and $t_{max}(\pi_f) \leq t$ by Lemma 2.

If $V^{i+1}_{t^{i+1}} \notin \mathcal{D}_{\geq t}$ then since $t_{min}(\pi_f) \geq t - \gamma_{xy}$ and $t_{max}(\pi_f) \leq t$, π_f is manually blocked by $V^{i+1}_{t^{i+1}} \in \mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$.

Else, $V^{i+1}_{t^{i+1}} \in \mathcal{D}_{\geq t}$, so since $\gamma_{xy} > 0$ and $t_{min}(\pi_f) \geq t - \gamma_{xy}$ and $t_{max}(\pi_f) \leq t$, there exists $j = max\{1 < j \leq i | V^{j-1}_{t^{j-1}} \rightarrow V^j_{t^j}\}$. Thus V^j is a collider and $V^j \in \mathcal{D}_{\geq t}$ so $Descendants(V^j_{t^j}, \mathcal{G}_s) \subseteq \mathcal{D}_{\geq t}$. Therefore, π_f is passively blocked by $V^j_{t^j} \in \mathcal{Z}_f = \mathcal{A}_{\leq t} \cup \mathcal{D}_{<t}$.

In conclusion, π_f is blocked by $\mathcal{Z}_f$. $\square$

Lemma 5. Let $\pi_s =< V^1, \dots, V^n >$ be a non-direct walk between X and Y as described. Then, take $\pi_f =< V^1_{t^1}, \dots, V^n_{t^n} >$ non-direct from $X_{t-\gamma_{xy}}$ to Y_t compatible with π_s (*i.e.*, $\pi_f \in \phi^{-1}(\pi_s)$).

Since $Y \exists 1 < i_y < n$ such that $V^{i_y} = Y$ and because π_f is a path, $t^{i_y} \neq t$ so by Lemma 2 $t - \gamma_{xy} \leq t^{i_y} < t$.

- If $\leftarrow V^{i_y}_{t^{i_y}}$ or $V^{i_y}_{t^{i_y}} \rightarrow$ then π_f is manually blocked by $V^{i_y}_{t^{i_y}}$. Since $t - \gamma_{xy} \leq t^{i_y} < t$, $V^{i_y}_{t^{i_y}} \in \mathcal{Z}_f$ and π_f is manually blocked by $\mathcal{Z}_f$.
- If $\rightarrow V^{i_y}_{t^{i_y}} \leftarrow V^{i_y+1}_{t^{i_y+1}}$ then π_f is manually blocked by $V^{i_y+1}_{t^{i_y+1}}$. Using Lemma 2, $t - \gamma_{xy} \leq t^{i_y+1} \leq t$. Moreover, since $t - \gamma_{xy} \leq t^{i_y} < t$ and $V^{i_y}_{t^{i_y}} \leftarrow V^{i_y+1}_{t^{i_y+1}}$, one can see that $t - \gamma_{xy} \leq t^{i_y+1} < t$. Lastly, π_f is a path so $V^{i_y+1}_{t^{i_y+1}} \neq X_{t-\gamma_{xy}}$. Thus π_f is manually blocked by $V^{i_y+1}_{t^{i_y+1}} \in \mathcal{Z}_f$.

$\square$

Lemma 6. Let $\pi_s =< V^1, \dots, V^n >$ be a non-direct path from X to Y with $< V^2, \dots, V^{n-1} >\subseteq Descendants(Y, \mathcal{G}_s)$ and $\nexists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$ and either $n \geq 3$ or $X \in Descendants(Y, \mathcal{G}_s)$ and $\exists C \in Cycles(X, \mathcal{G}_s)$ with $Y \notin C$. If $n \geq 3$ take $\pi'_s =< U^1, \dots, U^{m+1} >= \pi_s$. If $n = 2$ and $\exists C =< U^1, \dots, U^m >\in Cycles(X, \mathcal{G}_s)$ with $Y \notin C$, then take $\pi'_s = C + \pi_s$ in order to have a walk $\pi'_s =< U^1, \dots, U^{m+1} >$.

- Firstly,
  - if $n \geq 3$ then $< V^2, \dots, V^{n-1} >\subseteq Descendants(Y, \mathcal{G}_s)$ and $\pi'_s =< U^1, \dots, U^{m+1} >= \pi_s$ give $< U^2, \dots, U^m >\subseteq Descendants(Y, \mathcal{G}_s)$, and
  - if $n = 2$ since $X \in Descendants(Y, \mathcal{G}_s)$, $C \subseteq Descendants(X, \mathcal{G}_s) \subseteq Descendants(Y, \mathcal{G}_s)$ so $\pi'_s = C + \pi_s =< U^2, \dots, U^m >\subseteq Descendants(Y, \mathcal{G}_s)$.

  Therefore, there exists a FTCG $\mathcal{G}_f$ compatible with $\mathcal{G}_s$ in which $< U^2_t, \dots, U^{m+1}_t >\subseteq Descendants(Y_t, \mathcal{G}_f)$.
- Secondly, since $Y \notin C$, X and Y are not repeated in $< U^2, \dots, U^{m+1} >$ and because $\nexists 1 \leq i \leq n,\ V^i \leftarrow V^{i+1}$ and $U^m = X \rightleftarrows Y = U^{m+1}$, there exists a FTCG $\mathcal{G}'_f$ compatible with $\mathcal{G}_s$ in which $\pi_f =< U^1_{t-\gamma_{xy}} \rightarrow U^2_t \rightarrow \cdots \rightarrow U^{m+1}_t >$ is a path. In this case, π_f is active and can only be blocked by a $\mathcal{Z}_f$ such that $\mathcal{Z}_f \cap < U^1_{t-\gamma_{xy}}, U^2_t, \dots, U^{m+1}_t >\neq \emptyset$.

Since adjusting on $X_{t-\gamma_{xy}}$ or descendants of Y_t induces a bias, these two cases are irreconcilable and therefore $\alpha_{Y_t, X_{t-\gamma_{xy}}}$ is not identifiable. $\square$

Lemma 7. Let $\pi_s =< V^1, \dots, V^n >$ be an active non-direct path from X to Y in $\mathcal{G}_s$ with $< V^2, \dots, V^{n-1} >\subseteq Descendants(Y, \mathcal{G}_s)$ and $\gamma_{xy} = 0$. Since $< V^2, \dots, V^{n-1} >\subseteq Descendants(Y, \mathcal{G}_s)$ there exists a FTCG $\mathcal{G}_f$ compatible with $\mathcal{G}_s$ in which $< V^2_t, \dots, V^{n-1}_t >\subseteq Descendants(Y_t, \mathcal{G}_f)$. Moreover, since π_s is active there exists a FTCG $\mathcal{G}'_f$ compatible with $\mathcal{G}_s$ in which $\pi_f =< V^1_t, \dots, V^n_t >$ is an active path and can only be blocked by a $\mathcal{Z}_f$ such that $\mathcal{Z}_f \cap < V^1_t, \dots, V^n_t >\neq \emptyset$. Since conditioning on $X_{t-\gamma_{xy}}$ or descendants of Y_t induces a bias, these two cases are irreconcilable and therefore $\alpha_{Y_t, X_{t-\gamma_{xy}}}$ is not identifiable. $\square$

Theorem 1. The previous lemmas contain most of the proof, we will show here how to combine them to obtain Theorem 1. Lemma 1 give the first trivial condition $\alpha_{Y_t,X_{t-\gamma_{xy}}}$ not identifiable $\implies X \in Parents(Y,\mathcal{G}_s)$ so we assume in the remaining of the proof $X \in Parents(Y,\mathcal{G}_s)$. Lemma 7 gives the forward implication when $\gamma_{xy} = 0$ and Lemma 6 gives the forward implication when $\gamma_{xy} > 0$. All what is left is to prove to backward implication.

To prove the backward implication it suffices to prove that if we suppose that $\mathcal{G}_s$ does not contain any path as described in Theorem 1, then there exists an adjustment set that:

1. does not contain any descendant of Y_t in any FTCG that is compatible with $\mathcal{G}_s$, and
2. blocks every non-direct path from $X_{t-\gamma_{xy}}$ to Y_t in every FTCG that is compatible with $\mathcal{G}_s$.

Consider $\mathcal{Z}_f$ as defined in Definition 20. Condition 1 is satisfied since by construction $\mathcal{Z}_f$ does not contain any descendant of Y_t. To prove condition 2, let us consider π_f to be a path from $X_{t-\gamma_{xy}}$ to Y_t in a FTCG $\mathcal{G}_f$ compatible with $\mathcal{G}_s$. Let $\phi(\pi_f) = \pi_s =< V^1,\dots,V^n >$ its compatible walk in $\mathcal{G}_s$ and $\pi'_s =< U^1,\dots,U^m >$ the subpath of π_s. In the following, we consider all cases where π'_s violates one of the conditions of Theorem 1 :

- if $< U^2,\dots,U^{m-1} >\not\subseteq Descendants(Y,\mathcal{G}_s)$ then $< V^2,\dots,V^{n-1} >\not\subseteq Descendants(Y,\mathcal{G}_s)$ by Property 3. Thus Lemma 3 shows that $\mathcal{Z}_f$ blocks π_f, or
- if $< U^2,\dots,U^{m-1} >\subseteq Descendants(Y,\mathcal{G}_s)$, $\gamma_{xy} = 0$ and π'_s is passively blocked by U^i then π_s is passively blocked by at least a descendant of U^i by Property 1. Then consider $t_{min}(\pi_f)$ and $t_{max}(\pi_f)$. If $t_{min}(\pi_f) < t$ or $t_{max}(\pi_f) > t$ then Lemma 2 shows that $\mathcal{Z}_f$ blocks π_f. Else $t_{min}(\pi_f) = t_{max}(\pi_f) = t$, since π_s is passively blocked by at least a descendant of U^i and $Descendants(U^i,\mathcal{G}) \subseteq Descendants(Y,\mathcal{G})$, there exists $V^j_t \in \mathcal{D}_{\geq t}$ such that π_f is passively blocked by V^j_t. Therefore, $\mathcal{Z}_f$ blocks π_f, or
- if $\gamma_{xy} = 0$ and π'_s is direct then either π_s is direct and π_f is direct, or $< V^1,\dots,V^{n-1} >\in Cycles(X,\mathcal{G}_s)$ by Property 2. Then $V^1 = V^{n-1} = X$ so since π_f is a path, $t^{n-1} \neq t$. Thus $t_{min}(\pi_f) < t$ or $t_{max}(\pi_f) > t$ and Lemma 2 shows that $\mathcal{Z}_f$ blocks π_f.
- if $\gamma_{xy} > 0$, $m \geq 3$ and $\exists 1 \leq i < m,\ U^i \leftarrow U^{i+1}$ then $n \geq 3$ and $\exists 1 \leq i < n,\ V^i \leftarrow V^{i+1}$ by Property 4. Thus Lemma 4 shows that $\mathcal{Z}_f$ blocks π_f, or
- if $\gamma_{xy} > 0$, $m = 2$ and $X \notin Descendants(Y,\mathcal{G}_s)$ or $\forall C \in Cycles(X,\mathcal{G}_s),\ Y \in C$, either π_s is direct and π_f is direct, or $\pi_s =< X \rightleftarrows Y >$ and $\pi_f =< X_{t-\gamma_{xy}} \rightarrow Y_t >$ as $\gamma_{xy} > 0$ and π_f is direct or $< V^1,\dots,V^{n-1} >\in Cycles(X,\mathcal{G}_s)$ by Property 2. If $t_{min}(\pi_f) < t$ or $t_{max}(\pi_f) > t$ then Lemma 2 shows that $\mathcal{Z}_f$ blocks π_f. If $X \notin Descendants(Y,\mathcal{G}_s)$ then since $V^{n-1} = X$ and π_f is a path, $V^{n-1}_{t^{n-1}} \in \mathcal{Z}_f$ and thus $\mathcal{Z}_f$ blocks π_f. Else, $\forall C \in Cycles(X,\mathcal{G}_s),\ Y \in C$ so $Y \in< V^2,\dots,V^{n-1} >$ and Lemma 5 shows that $\mathcal{Z}_f$ blocks π_f.

Thus, any violation of any condition Theorem 1 leads to π_f becoming $\mathcal{Z}_f$-blocked. In conclusion, if the conditions of Theorem 1 are verified then $\mathcal{Z}_f$ defined in Definition 20 allows to identify $\alpha_{Y_t,X_{t-\gamma_{xy}}}$. □

Property 1. Let $\mathcal{G}_s = (\mathcal{V}_s,\mathcal{E}_s)$ be a SCG, $X, Y \in \mathcal{V}_s$ with $X \in Parents(Y,\mathcal{G}_s)$ and γ_{xy} be a lag. Suppose the direct effect of $X_{t-\gamma_{xy}}$ on Y_t is identifiable following Theorem 1. Let $\mathcal{Z}_f$ be the single-door set relative to $(X_{t-\gamma_{xy}}, Y_t)$ defined in Definition 21. Let $\mathcal{G}_f = (\mathcal{V}_f,\mathcal{E}_f)$ be a compatible FTCG of maximal lag at most γ_{max} compatible with $\mathcal{G}_s$.

- Firstly, using the decomposition $\mathcal{Z}_f = \mathcal{D}^{Anc(Y)}_{t'} \cup \mathcal{A}^{Anc(Y)}_{t'}$ as in Definition 21, and because $Descendants(Y_t,\mathcal{G}_f) \subseteq \{V_{t'} | V \in Descendants(Y,\mathcal{G}_s),\ t' \geq t\}$ it is clear that $\mathcal{Z}_f \cap (Descendants(Y_t,\mathcal{G}_f) \cup \{X_{t-\gamma_{xy}}\}) = \emptyset$.
- Secondly, let $\pi_f =< V^1_{t^1},\dots,V^n_{t^n} >$ be a non-direct path from $X_{t-\gamma_{xy}}$ to Y_t in FTCG $\mathcal{G}_f$ and $\pi_s =< V^1,\dots,V^n >= \phi(\pi_f)$ its compatible walk. If $t_{max}(\pi_f) > t$ then $\exists 1 < k < n$ such that $\rightarrow V^k_{t_{max}(\pi_f)} \leftarrow$ in π_f with $Descendants(V^k_{t_{max}(\pi_f)},\mathcal{G}_f) \cap \mathcal{Z}_f = \emptyset$ and thus π_f is passively blocked by $\mathcal{Z}_f$. Therefore, for the following we can suppose $t_{max}(\pi_f) \leq t$.

Because the direct effect of $X_{t-\gamma_{xy}}$ on Y_t is identifiable following Theorem 1 we know that $\pi_f \neq < V^1_{t^1} \leftarrow \cdots \leftarrow V^n_{t^n} >$. Therefore, $\exists 1 < k \leq n$ such that $\rightarrow V^k_{t^k} \leftarrow \cdots \leftarrow V^n_{t^n}$ with $t^k \geq t$ and $V^k \in Descendants(Y, \mathcal{G}_s)$ and since π_f is non-direct this forces $n > 2$.

- If $k < n$ then $Descendants(V^k_{t^k}, \mathcal{G}_f) \cap \mathcal{Z}_f = \emptyset$ and thus π_f is passively blocked by $\mathcal{Z}_f$.
- If $k = n$ (*i.e.*, $V^{n-1}_{t^{n-1}} \rightarrow V^n_{t^n}$) and $\pi_f =< V^1_{t^1} \rightarrow \cdots \rightarrow V^n_{t^n} >$ then because the direct effect of $X_{t-\gamma_{xy}}$ on Y_t is identifiable following Theorem 1 we know that $\exists d_{max} = \max\{1 < d < n$ such that $V^d \notin Descendants(Y, \mathcal{G}_s)\}$. This forces either $t - \gamma_{max} \leq t^{d_{max}} \leq t$ or $\exists d_{max} < d < n$ such that $t - \gamma_{max} \leq t^d < t$ so since $V^{d_{max}}, V^d \in Ancestors(Y, \mathcal{G}_s)$, we have either $V^{d_{max}}_{t^{d_{max}}} \in \mathcal{A}^{Anc(Y)}_{t'}$ or $V^d_{t^d} \in \mathcal{D}^{Anc(Y)}_{t'}$ and thus π_f is manually blocked by $\mathcal{Z}_f$.
- If $k = n$ (*i.e.*, $V^{n-1}_{t^{n-1}} \rightarrow V^n_{t^n}$) and $\exists l_{max} = max\{1 < l < n | V^{l-1}_{t^{l-1}} \leftarrow V^l_{t^l}\}$ then $V^{l_{max}} \in Ancestors(Y, \mathcal{G}_s)$
  * If $t^{l_{max}} < t$ then $\exists l_{max} \leq i$ such that $V^i_{t^i} \rightarrow V^{i+1}_{t^{i+1}}$ and $t - \gamma_{max} \leq t^i < t^{i+1} = t$ and since $V^i \in Ancestors(Y, \mathcal{G}_s)$, $V^i_{t^i} \in \mathcal{Z}_f$ and thus π_f is manually blocked by $\mathcal{Z}_f$.
  * If $V^{l_{max}} \notin Descendants(Y, \mathcal{G}_s)$ and $t^{l_{max}} = t$ then, since $V^{l_{max}} \in Ancestors(Y, \mathcal{G}_s)$, $V^{l_{max}}_{t^{l_{max}}} \in \mathcal{A}^{Anc(Y)}_{t'}$ and thus π_f is manually blocked by $\mathcal{Z}_f$.
  * If $V^{l_{max}} \in Descendants(Y, \mathcal{G}_s)$ and $t^{l_{max}} = t$ then we can distinguish the cases $\gamma_{xy} = 0$ and $\gamma_{xy} > 0$:
    · If $\gamma_{xy} = 0$ then, because the direct effect of $X_{t-\gamma_{xy}}$ on Y_t is identifiable following Theorem 1 we know that $\phi(\pi_f)$ is blocked so $\exists r_{max} = max\{1 < r < l_{max} | V^{r-1}_{t^{r-1}} \rightarrow V^r_{t^r}\}$.
    · Similarly, if $\gamma_{xy} > 0$ since $V^{l_{max}-1}_{t^{l_{max}-1}} \leftarrow V^{l_{max}}_{t^{l_{max}}}$, then $t = t^{l_{max}} \leq t^{l_{max}-1} \leq t$ and there must exists $1 < r < l_{max}$ such that $V^{r-1}_{t^{r-1}} \rightarrow V^r_{t^r}$ with $t^r < t^{r+1} = t$ so $\exists r_{max} = max\{1 < r < l_{max} | V^{r-1}_{t^{r-1}} \rightarrow V^r_{t^r}\}$.

    Therefore, we have $\rightarrow V^r_{t^r} \leftarrow$ and since $V^{l_{max}} \in Descendants(Y, \mathcal{G}_s)$ and $V^{r_{max}} \in Descendants(V^{l_{max}}, \mathcal{G}_s)$ we have $V^{r_{max}} \in Descendants(Y, \mathcal{G}_s)$ so $\mathcal{Z}_f \cap Descendants(V^{r_{max}}_{t^{r_{max}}}, \mathcal{G}_f) = \emptyset$ and therefore π_f is passively blocked by $\mathcal{Z}_f$.

In conclusion, π_f is blocked by $\mathcal{Z}_f$.

□