

IFT - UNESP  
INSTITUTO DE FÍSICA TEÓRICA

---

---

DISSERTAÇÃO DE MESTRADO

IFT-D.006/2021

# Observational Equivalences Between Causal Structures

Marina Maciel Ansanelli

Orientador

*Robert Spekkens*

Dezembro de 2021

A617o Ansanelli, Marina Maciel  
Observational equivalences between causal structures / Marina Maciel  
Ansanelli. – São Paulo, 2021  
76 f.

Dissertação (mestrado) – Universidade Estadual Paulista (Unesp),  
Instituto de Física Teórica (IFT), São Paulo  
Orientador: Robert Spekkens

1. Teoria quântica. 2. Causalidade (Física). 3. Informação quântica. I.  
Título

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca  
do Instituto de Física Teórica (IFT), São Paulo. Dados fornecidos pelo  
autor(a).

*Dedico esta dissertação a  
Márcia e Donato,  
Léo  
e a dupla Chico e Frodo.*

# Acknowledgements

I would first like to thank my supervisor Robert Spekkens and my co-supervisor Elie Wolfe. Their guidance, always with enthusiasm and patience, made this research project exceptionally enjoyable. I would also like to thank TC Fraser, Robin Evans and Noam Finkelstein for discussions and insights on this project.

I wish to thank Leonardo Almeida Lessa for immeasurable help with Python, Mathematica, and life. I am also grateful to Bruno Arderucio and Hugo Marrochio for being crucial influences in my decision to take the path of physics.

I would like to thank my parents, for always being and encouraging me to be curious and inquisitive. Furthermore, for being my almost exclusive in-person human companion for the last two pandemic years.

This work was supported by the Brazilian National Council for Scientific and Technological Development (CNPq).

# Resumo

O objetivo da área de Inferência Causal é apontar explicações causais subjacentes para dados observáveis. A explicação causal muitas vezes não é única, uma vez que podemos ter duas estruturas causais que explicam o mesmo conjunto de distribuições. Como tal, podemos agrupar as estruturas causais que originam as mesmas correlações no que chamamos de classes de Equivalência Observacional.

Nesta tese, revisamos trabalhos anteriores sobre equivalências observacionais entre estruturas causais clássicas e apresentamos duas extensões de resultados existentes sobre equivalência. Além disso, aplicamos esses resultados aos casos de 3 e 4 variáveis, mostrando que a classificação é completa para o primeiro caso e ainda incompleta para o segundo. Finalmente, mostramos como essa classificação pode ajudar a filtrar exemplos de estruturas que podem exibir lacunas Quântico-Clássicas e discutimos a validade quântica dos resultados de equivalência.

**Palavras Chaves:** Inferência Causal; Equivalência Observacional; Quantum-Classical Gaps;

**Áreas do conhecimento:** Física; Fundamentos da Mecânica Quântica; Inferência Causal.

# Abstract

The goal of Causal Inference is to point out underlying causal explanations for observable data. The causal explanation is often not unique, since we can have two causal structures that explain the same set of distributions. As such, we can group the causal structures that originate the same correlations into what we call Observational Equivalence classes.

In this thesis, we review previous work on observational equivalences between classical causal structures and present two extensions to existent rules to prove equivalence. Furthermore, we apply these results to the cases of 3 and 4 variables, showing that the classification is complete for the former case and still incomplete for the latter. Finally, we show how this classification can help filter examples of structures that may exhibit Quantum-Classical gaps and we discuss the quantum validity of the equivalence results.

**Key Words:** Causal Inference; Observational Equivalences; Quantum-Classical Gaps;

**Areas:** Physics; Quantum Foundations; Causal Inference.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation for Physics: The Bell Causal Structure . . . . .	2
1.2	Observational Partial Order . . . . .	3
<b>2</b>	<b>Preliminary Notions</b>	<b>6</b>
2.1	The Causal Compatibility Problem . . . . .	6
2.2	The d-separation Criterion . . . . .	9
2.3	Marginal Models and mDAGs . . . . .	10
2.4	Dominance: Hypergraph and Directed Structure Inclusion . . . . .	12
2.5	Graph Patterns . . . . .	13
2.6	Computational Work: Constructing Metagraph . . . . .	16
<b>3</b>	<b>Ways to Show Equivalence</b>	<b>18</b>
3.1	HLP Proposition . . . . .	18
3.2	Evans’s Proposition and its Spin-Offs . . . . .	19
3.2.1	Weak Face Splitting . . . . .	20
3.2.2	Moderate Face Splitting . . . . .	23
3.2.3	Strong (Simultaneous) Face Splitting . . . . .	23
<b>4</b>	<b>Comparisons that Show Inequivalence</b>	<b>26</b>
4.1	Graphical Methods . . . . .	26
4.1.1	Skeletons . . . . .	26
4.1.2	CI Relations . . . . .	27
4.1.3	e-separation Relations . . . . .	28
4.1.4	Pairs of Densely Connected Nodes . . . . .	29
4.1.5	Only Hypergraphs . . . . .	31
4.2	Comparison of Infeasible Supports . . . . .	33
4.2.1	Infeasible Supports beyond e-separation . . . . .	34
4.2.2	Inequivalences Shown by Comparison of Infeasible Supports . . . . .	34
<b>5</b>	<b>3 and 4 Observable Nodes</b>	<b>36</b>
5.1	Completely Solved Classes . . . . .	37

5.2	Classes Compatible with Every Support . . . . .	38
<b>6</b>	<b>Quantum Causal Structures</b>	<b>41</b>
6.1	Quantum Validity of the Rules for Proving Equivalence and In- equivalence . . . . .	45
6.1.1	Ways to Prove Equivalence . . . . .	45
6.1.2	Comparisons that Show Inequivalence . . . . .	48
6.2	Searching for QC Gaps . . . . .	49
6.2.1	Fundamental mDAGs . . . . .	49
<b>7</b>	<b>Conclusion</b>	<b>53</b>
<b>A</b>	<b>Evans’s Rules</b>	<b>55</b>
<b>B</b>	<b>Proofs of Equivalence Results</b>	<b>58</b>
B.1	Proof of HLP Proposition . . . . .	58
B.2	Proof of Weak Face Splitting . . . . .	60
B.3	Proof of Strong (Simultaneous) Face Splitting . . . . .	63
<b>C</b>	<b>Proof of the Only Hypergraphs Rule (GPT valid)</b>	<b>64</b>
<b>D</b>	<b>TC Fraser’s Algorithm for Feasible Supports</b>	<b>68</b>
<b>E</b>	<b>Counter-Example of a quantum exogenization Lemma</b>	<b>70</b>
<b>F</b>	<b>District Factorization</b>	<b>72</b>
	<b>Bibliography</b>	<b>74</b>



# Chapter 1

## Introduction

Many times, it passes unnoticed that most scientific statements talk about *correlations* between observed events, not about causal relationships. For example, when we say that the barometer reading  $B$  is related to the atmospheric pressure  $P$  by  $B = kP$ , we are not specifying if  $P$  causes  $B$  or if  $B$  causes  $P$ <sup>1</sup>. Of course, we have a strong belief that the atmospheric pressure causes the barometer reading, but this belief is not imprinted in the equation.

The barometer example is very simple to figure out, but as we go to increasingly complicated scenarios, it can be dangerous not to have a way to express causal information. In particular, mixing correlation with causation can lead to wrong conclusions.

The area of Causal Inference aims to provide a formal way to talk about causality. It does so by introducing *causal structures*, directed acyclic graphs where the vertices are the studied events and the arrows mean direct causal influence. These causal structures provide potential underlying explanations to the observed statistics.

Causal structures impose certain constraints on the probability distributions that can be satisfied by the observed events. We will see the first example of these causal constraints with the graph of Figure 1.1, and they will be explained in more detail with the examples of Figure 2.1. Through causal constraints, we can relate our causal investigation to experimental data: a certain structure  $\mathcal{G}$  can only explain a phenomenon  $\mathcal{F}$  if the observed correlations of  $\mathcal{F}$  satisfy all the constraints of  $\mathcal{G}$ .

Causal Inference has applications in all areas of science, from machine learning, where it was born, to health studies[2][3][4] and Earth sciences[5]. Now, we will see a motivation for physicists working in Quantum Foundations to study Causal Inference.

---

<sup>1</sup>Example taken from [1].

## 1.1 Motivation for Physics: The Bell Causal Structure

Bell's Theorem[6] can be rephrased using Causal Inference, as was first shown in [7]: the deterministic (classical) causal structure that satisfies local causality and no super-determinism also yields an inequality constraint that is precisely Bell's inequality.

The Bell structure is presented in Figure 1.1. There,  $S$  and  $T$  represent the choices of measurement made by Alice and Bob, and  $A$  and  $B$  represent their measurement outcomes. The variable  $\lambda$  is the hidden variable that influences both  $A$  and  $B$ . Every node in this graph represents a *classical* variable; we will see precisely what this means in Definition 3.

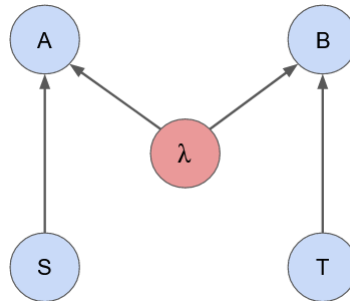


Figure 1.1: The Bell Structure

In Figure 1.1, we can see that  $S$  and  $T$  are independent, and also independent of  $\lambda$ . Furthermore, if we know  $\lambda$ , the two wings of the experiment become independent. Denoting independence by  $\perp$  and conditioning by  $|$ , this means:

$$S \perp T\lambda \text{ and } T \perp S\lambda \quad (\text{No Super-Determinism}) \quad (1.1)$$

$$A \perp BT|S\lambda \text{ and } B \perp AS|T\lambda \quad (\text{Local Causality}) \quad (1.2)$$

Equations (1.1) and (1.2), that are of the form "variable  $x$  is independent of variable  $y$  given variable  $z$ ", are called *Conditional Independence* (CI) relations. We can see that the Conditional Independence relations that the Bell causal structure yields are exactly the hypotheses of Bell's Theorem: no super-determinism and local causality.

When there are hidden (unmeasurable) variables such as  $\lambda$ , Conditional Independences like the ones above are in general not the only causal constraints we have. In the case of the Figure 1.1, if we assume binary variables, there is also the

following constraint:

$$\frac{1}{4} \sum_{a=b} P_{AB|ST}(ab|00) + \frac{1}{4} \sum_{a=b} P_{AB|ST}(ab|01) + \frac{1}{4} \sum_{a=b} P_{AB|ST}(ab|10) + \frac{1}{4} \sum_{a \neq b} P_{AB|ST}(ab|11) \leq 3/4, \quad (1.3)$$

This is Bell's inequality.

Thus, as stated in [7], the fact that quantum correlations violate this inequality means that they cannot be explained by the classical causal structure of Figure 1.1. In reality, [7] proves more than that: even if we change the causal structure so that it violates local causality or no super-determinism (but keeping it a *classical* causal structure), the quantum correlations can only be explained if we allow for fine-tuning of the parameters. As will be discussed in Section 2.1, we consider those as bad explanations.

If one changes the hidden variable  $\lambda$  in Figure 1.1 to a quantum system, as will be discussed in more detail in Chapter 6, one can violate Bell's inequality and explain quantum correlations. Therefore, we say that this causal structure yields a *Quantum-Classical gap* (QC gap).

To have a QC gap, a causal structure needs to have inequality constraints [8], although not every inequality constraint leads to a QC gap. In Section 6.2, we will discuss the task of finding other examples that can lead to QC gaps and Bell-like theorems.

## 1.2 Observational Partial Order

If we have a set of data, our goal in causal inference is to find its causal explanation. In general, the answer is not unique; the set of constraints of one causal structure can include the constraints of another, or even be the same.

There is a partial order of causal structures that have the same number of measured variables: if every probability distribution that can be explained by  $\mathcal{G}$  can also be explained by  $\mathcal{G}'$ , then  $\mathcal{G}'$  *dominates*  $\mathcal{G}$ . If the sets of explainable correlations are the same,  $\mathcal{G}$  and  $\mathcal{G}'$  are *equivalent*.

Note that if two causal structures are equivalent, it is impossible to make a passive observation experiment that decides which one governs the relationships between a set of events. This is why we call this an *Observational* equivalence.

There are cases when it is hard to explicitly find all the constraints that a

causal structure yields. Thus, finding other ways to classify causal structures into equivalence classes of Observational equivalence is an important step to identify causal explanations of observed data. This filtering can also help us spot causal structures that are good candidates for having a QC gap; for example, if we know that a structure does not yield inequality constraints, then all the structures equivalent to this one will also not have inequality constraints.

In this project, we organized causal structures into Observational equivalence classes. To do so, we used the equivalence rules of Chapter 3, which tell us when two structures are known to be equivalent, and the comparisons of Chapter 4, that tell us when two structures are known to be inequivalent. For 3 visible variables, the classification is solved. As we will see, the results we have are *not* enough to completely characterize the equivalence classes for causal structures of 4 visible variables.

The classification process is illustrated in Figure 1.2. There, the classical causal structures (red dots) are organized into true equivalence classes (green dashed bubbles). Our goal is to find these equivalence classes.

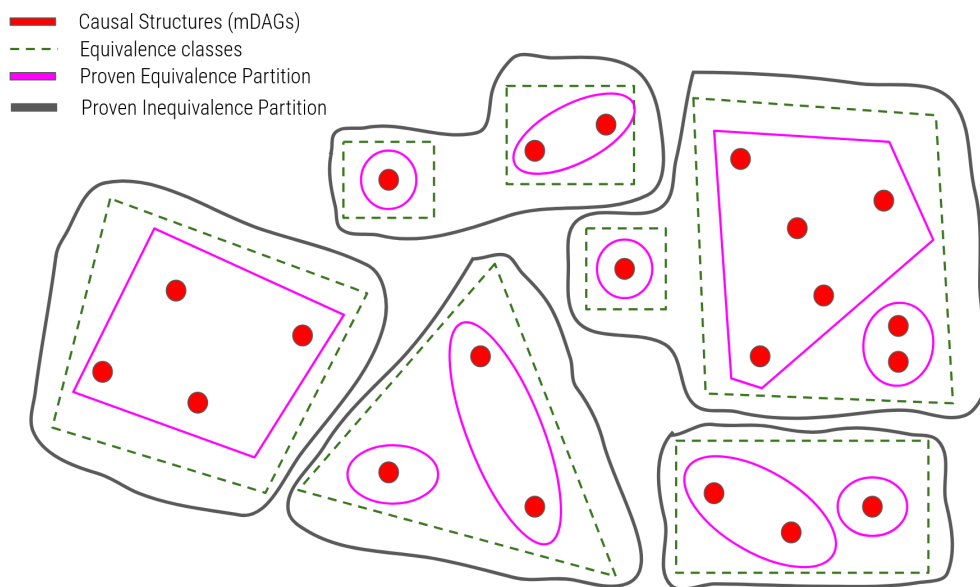


Figure 1.2: The proven-equivalence partition (pink bubbles) is found by applying the equivalence propositions of Chapter 3. The proven-inequivalence partition (gray bubbles) is found by the comparisons of Chapter 4.

Through the equivalence rules, we find the proven-equivalence partition, represented by pink bubbles in Figure 1.2. If two causal structures are in the same pink bubble, we know that they are observationally equivalent. The successive application of the equivalence results shows more and more equivalences, thus

merging pink bubbles. On the other hand, making the comparisons described in Chapter 4 proves inequivalences between the causal structures, thus establishing the gray bubbles of Figure 1.2; if two structures are *not* in the same gray bubble, we know that they are *not* observationally equivalent. Implementing more comparisons successively shows more and more inequivalences, thus splitting gray bubbles.

If a given set of causal structures is an element of both the proven-equivalence partition and the proven-inequivalence partition, then it describes an equivalence class of causal structures. The leftmost class of Figure 1.2 illustrates this circumstance.

This is the method we used to apply the existent equivalence results to classify the structures with 3 and 4 observable variables. After an introduction of the framework in Chapter 2, we will present the equivalence results in Chapters 3 and 4 (including two original theoretical results in Sections 3.2.2 and 3.2.3), and show the outcomes of their implementation in Chapter 5. Finally, a discussion about the quantum case and the search for QC gaps is done in Chapter 6. At the end of each chapter, we include a summary box with the more relevant ideas.

# Chapter 2

## Preliminary Notions

In this thesis (and in the majority of the discussions within the Causal Inference framework), we do not consider the possibility of closed cycles of causal influence. Thus, the causal structure that relates a given set of events is represented by a Directed Acyclic Graph (DAG), a directed graph that has no cycles.

These DAGs are composed of vertices (nodes) and directed edges, that are ordered pairs of nodes. Below, we define some terminology about DAGs.

**Definition 1** (Children, Parents, Descendants, Ancestors). *Let  $v$  be a node of a DAG  $\mathcal{G}$ . If  $w$  is a node in  $\mathcal{G}$  such that there is a directed edge  $v \rightarrow w$ , then  $w$  is called a child of  $v$ . Conversely, all the nodes  $u$  in  $\mathcal{G}$  that have  $v$  as a child are the parents of  $v$ . The set of all children of  $v$  is denoted  $ch_{\mathcal{G}}(v)$ , and the set of all parents of  $v$  is denoted  $pa_{\mathcal{G}}(v)$ .*

*The descendants of  $v$  are all the nodes in  $\mathcal{G}$  that can be reached from  $v$  by a sequence of nodes through the directed edges. The ancestors of  $v$  are the nodes that have  $v$  as a descendant.*

**Definition 2** (Subgraph). *Let  $\mathcal{G}$  be a DAG with vertices  $V$  and edges  $\mathcal{E}$ , and  $\mathcal{G}'$  be a DAG with vertices  $V'$  and edges  $\mathcal{E}'$ . If  $V' \subseteq V$  and  $\mathcal{E}' \subseteq \mathcal{E}$ , we say that  $\mathcal{G}'$  is a subgraph of  $\mathcal{G}$ .*

### 2.1 The Causal Compatibility Problem

When performing an experiment that assesses the occurrence of a set of events, we obtain probability distributions over these events. Using statistical methods, we can find correlations between these probability distributions. We can infer properties of the underlying causal structure by investigating which DAGs are or are not compatible with the observed correlations. This is the *Causal Compatibility* problem.

The DAGs in Figure 2.1 give a first example of this: in 2.1(a), the causal structure imposes that variables  $B$  and  $C$  become independent when we know the value of

A. This is called a *Conditional Independence* (CI) relation. It means that the joint probability distributions that are compatible with 2.1(a) need to factorize as:

$$P(ABC) = P(C|A)P(B|A)P(A) \quad (2.1)$$

In general, the probability distribution of a given node  $v$  conditioned on  $v$ 's parents is independent of all the non-descendants of  $v$ .

If a probability distribution does not satisfy the constraint of Equation (2.1), then it is not compatible with DAG 2.1(a). The DAG 2.1(b), on the other hand, is *saturated*: it does not yield any constraint, so it is compatible with any probability distribution over the visible variables.

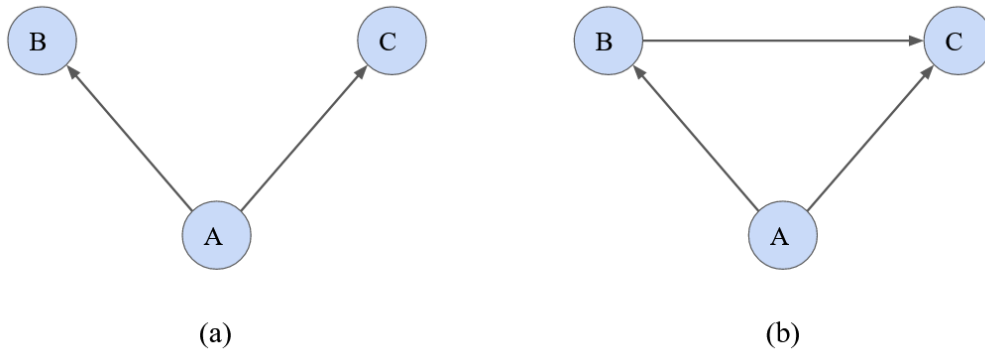


Figure 2.1: In (a), the nodes B and C are independent when we condition on A. In (b), no such Conditional Independence relation exists. In fact, the DAG in (b) is saturated: every probability distribution over 3 variables is compatible with it.

To formalize the causal compatibility problem, we identify every node  $v$  of a DAG  $\mathcal{G}$  with a random variable  $X_v : \Omega \rightarrow \mathcal{X}_v$ . Here,  $(\Omega, \mathcal{F}, P)$  is some probability space, where  $\mathcal{F}$  is a  $\sigma$ -algebra on the set  $\Omega$  and  $P$  is a probability measure. The set  $\Omega$  and the  $\sigma$ -algebra  $\mathcal{F}$  used to define the random variable are not important; we can record the probability of  $X$  taking each value in  $\mathcal{X}$ , and this will be our probability distribution  $P$ .  $\mathcal{X}_v$  is the set in which  $X_v$  takes values. At this point, we are not making any assumptions on  $\mathcal{X}_v$ .

The structure of the DAG is reflected in this identification by stating that  $X_v$  only "depends" on  $X_{\text{pa}_{\mathcal{G}}(v)}$ . This idea is made precise with the Structural Equation Property, taken from [9]:

**Definition 3** (Structural Equation Property - SEP). *Let  $\mathcal{G}$  be a DAG, and  $V$  is the set of vertices of  $\mathcal{G}$ . Let  $P$  be some joint probability distribution over the random variables  $X_v$ ,  $v \in V$ .*

We say that  $P$  obeys the Structural Equation Property for the DAG  $\mathcal{G}$  if there exists:

- independent random variables  $E_v : \Omega \rightarrow \mathcal{E}_v$  (called error variables)
- measurable functions  $f_v : \mathcal{X}_{pa_{\mathcal{G}}(v)} \times \mathcal{E}_v \rightarrow \mathcal{X}_v$

such that for all  $v \in V$ , the assignment

$$X_v = f_v(X_{pa_{\mathcal{G}}(v)}, E_v) \quad (2.2)$$

ensures that  $X_V$  has the joint probability distribution  $P$ .

Saying that  $X_v = f_v(X_{pa_{\mathcal{G}}(v)}, E_v)$  is equivalent to saying that  $X_v$  is  $\sigma$ - $(X_{pa_{\mathcal{G}}(v)}, E_v)$  measurable.

We denote by  $\mathcal{M}(\mathcal{G})$  the set of probability distributions that obey the SEP for the DAG  $\mathcal{G}$ .

Note that here enters the fact that our causal models are *classical*: every node is given by a deterministic function of its parents and an error function. This error function allows for a node to respond probabilistically to its parents; however, it is still in the scope of classical probabilities.

One could ask: why can't we assume that saturated DAGs explain every experiment? In other words, why do we look for DAGs with constraints?

This is justified with the Popperian principle of Falsifiability [10]: When two theories are compatible with the experiments we have access to, we consider that the best one is the theory that is more falsifiable. In the context of causal modeling, the more falsifiable explanation is the one that presents more constraints. Therefore, we disregard causal explanations that admit more general probability distributions than the ones we see in experiments; this is what happens for example with classical structures for Bell's scenario that allow for superluminal signaling.

Many causal structures involve variables that are not measurable, or that we choose not to measure. For example, in Figure 1.1, the variable  $\lambda$  is not measurable. We call those variables "latent" and denote their set by  $L$ . Similarly, we will call the measurable variables "visible", and denote their set by  $V$ . Furthermore, we say that a DAG that is composed only of visible variables is a *visible-variable* or *latent-free* DAG, as opposed to the more general *latent-variable* DAGs, that may include latent nodes.

In Theorem 1 of [11] it is shown that  $P$  admitting a recursive factorization (obeying SEP) for a visible-variable DAG  $\mathcal{G}$  is equivalent to  $P$  presenting CI relations that mirror  $\mathcal{G}$ 's CI relations (in [11], this is called Global Markov Property).



This means that, in the case where all variables are visible (called latent-free), the Causal Compatibility problem is just a problem of finding the CI Relations of a DAG. This is completely solved by the method of *d-separation*, which is outlined in the next section.

## 2.2 The d-separation Criterion

In [12], we can find a general algorithm for finding CI relations like the one observed in the DAG of Figure 2.1(a). It is called the *d-separation criterion*, and it is given by:

**Definition 4** (d-Separation). *An undirected path  $p$  in a DAG  $\mathcal{G}$  is said to be blocked by a set of nodes  $Z$  if:*

1.  *$p$  contains a chain or a fork (see Figure 2.2) such that the middle node is in  $Z$ , or*
2.  *$p$  contains a collider (see Figure 2.2) such that neither the middle node nor any of its descendants are in  $Z$*

*We say that nodes  $x$  and  $y$  in the DAG  $\mathcal{G}$  are d-separated by  $Z$  if all the paths that connect them are blocked by  $Z$ . This implies that  $x$  and  $y$  are independent conditioned on  $Z$ .*

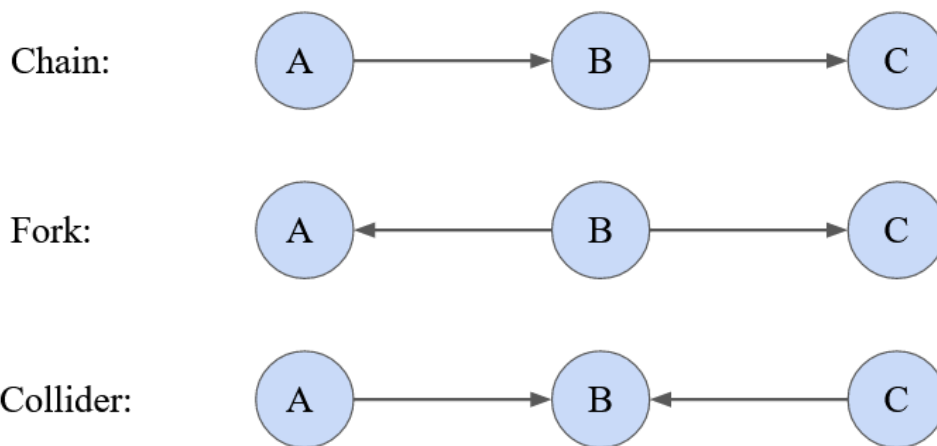


Figure 2.2: Definitions used in the d-separation Criterion. In the two first cases, the path  $A \rightarrow B \rightarrow C$  is blocked by  $Z$  if  $B \in Z$ . In the third case, this path is blocked by  $Z$  if  $B \notin Z$  and no descendant of  $B$  is in  $Z$ .

Looking at the cases of the chain and the fork in Figure 2.2, it is easy to convince oneself that  $A$  and  $C$  become independent when we know  $B$ . This is what happened in the case of Figure 2.1(a), where we had a fork.

In the case of the collider, the inverse happens: when we know  $B$ , the variables  $A$  and  $C$  become *dependent*. We can make sense of this by imagining the following scenario: let  $B$  represent some illness, and  $A$  and  $C$  represent two causes of the same illness; for example, unhealthy lifestyle and genetic predisposition. Say that these two causes are in principle uncorrelated. However, when a patient presents the illness ( $B$ ) and we learn that they have a healthy lifestyle ( $A$ ), we increase the likelihood of them having the genetic predisposition ( $C$ ). So,  $A$  and  $C$  become correlated under knowing  $B$ .

As already mentioned, in the latent-free case, the only constraints on the distributions are the CI relations. So, the d-separation criterion solves the entire problem. When there are latent variables, however, the problem becomes much more complicated. Now we will turn our attention to this case, introducing the concept of an mDAG.

## 2.3 Marginal Models and mDAGs

As we have seen in the example of the Bell Structure (Figure 1.1), when we have latent nodes we can obtain constraints other than just CI relations (like the inequality constraint of Equation (1.3)). These inequalities are not easy to find in general.

For a DAG with latent variables, the question we ask is not whether a probability distribution of  $(X_V, X_L)$  is compatible with the DAG, but whether a probability distribution of  $X_V$  is compatible with the DAG. Following [9], this is posed as:

**Definition 5.** Let  $\mathcal{G}$  be a DAG with visible variables  $V$  and latent variables  $L$ . The state space for the visible variables is  $\mathcal{X}_V$ . We say that a probability distribution  $P$  over  $\mathcal{X}_V$  is "compatible with the marginal DAG model of  $\mathcal{G}$ " if there exists:

- a state space  $\mathcal{X}_L$  for  $X_L$ ; and
- a probability measure  $Q \in \mathcal{M}(\mathcal{G})$  over  $\mathcal{X}_V \times \mathcal{X}_L$

such that  $P$  is the marginal of  $Q$  over  $X_V$ . We denote this by  $P \in \mathcal{M}(\mathcal{G}, V)$ .

For example, in Figure 1.1, we consider only probability distributions over the observables  $A, B, S$ , and  $T$ , and not the hidden variable  $\lambda$ .

In [9] it is shown that, for classical causal structures, it suffices to consider latent variables that have no parents (called *exogenized*) and whose children are not a subset of the children of another latent variable (called *non-redundant*). These

results, which we call Evans's rules, are presented in Appendix A as Lemmas 4 and 5. These rules suggest the introduction of a new structure for studying marginal distributions of classical causal models: the marginalized DAG, or mDAG (first introduced in [9]). This structure only makes sense for DAGs whose latent variables are all exogenized and non-redundant. Thus, mDAGs are *not* applicable in the case of quantum causal structures, because Lemma 4 is *not* valid in that case (as shown in [13]).

To define an mDAG, we first introduce the concept of an *Abstract Simplicial Complex*:

**Definition 6** (Abstract Simplicial Complex). *An abstract simplicial complex over a finite set  $V$  is a set  $\mathcal{B}$  of subsets of  $V$  such that*

- $\{v\} \in \mathcal{B}$  for all  $v \in V$ ;
- If  $A \subseteq B \subseteq V$  and  $B \in \mathcal{B}$ , then  $A \in \mathcal{B}$ .

The elements of  $\mathcal{B}$  are called *faces*. The inclusion-maximal elements of an abstract simplicial complex (the faces that are maximal in the order over faces induced by subset inclusion) are called *facets*.

An mDAG is a pair  $(\mathcal{D}, \mathcal{B})$ , where  $\mathcal{D}$  is a visible-variable DAG and  $\mathcal{B}$  is an abstract simplicial complex over the set of nodes. All the information about the latent variables is encoded in the simplicial complex. We say that  $\mathcal{D}$  is the *directed structure* of the mDAG  $\mathcal{G}$ .

The identification between latent-variable DAGs and associated mDAGs is exemplified in Figure 2.3: each latent variable  $U$  in the DAG corresponds to a facet in the mDAG whose elements are the children of  $U$ .

Because of Lemmas 4 and 5 (See Appendix A), two latent-variable DAGs,  $\mathcal{G}$  and  $\mathcal{G}'$ , whose associated mDAGs are the same also have the same model over the visible variables,  $\mathcal{M}(\mathcal{G}, V) = \mathcal{M}(\mathcal{G}', V)$ . Therefore, the mDAG structure contains all the information needed to study marginal distributions over visible variables. In all of this work except for Section 6 we discuss *classical* causal structures, thus we use the language of mDAGs.

The results and definitions that we presented for DAGs are trivially extended for mDAGs if we think of their associated DAGs. For example:

**Definition 7.** *Let  $\mathcal{G}$  be an mDAG with variables  $V$ . If a probability distribution  $P$  over  $\mathcal{X}_v$  obeys Definition 5 for the latent-variable DAG associated with  $\mathcal{G}$  ( $P \in \mathcal{M}(\mathcal{G}, V)$ ), we say that  $P$  is compatible with the mDAG  $\mathcal{G}$ .*

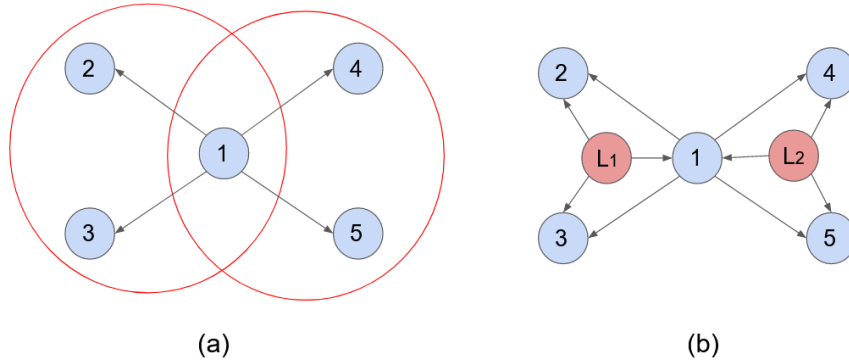


Figure 2.3: (a) An mDAG with abstract simplicial complex  $\mathcal{B} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 4\}, \{1, 5\}, \{4, 5\}, \{1, 2, 3\}, \{1, 4, 5\}\}$ . The inclusion-maximal elements of  $\mathcal{B}$  are indicated by the red bubbles. (b) The DAG associated with the mDAG (a).

**Definition 8** (Subgraph of an mDAG). Let  $\mathcal{G} = (D, \mathcal{B})$  and  $\mathcal{G}' = (D', \mathcal{B}')$  be two mDAGs. If  $D'$  is a subgraph of  $D$  and  $\mathcal{B}' \subseteq \mathcal{B}$ , we say that  $\mathcal{G}'$  is a subgraph of  $\mathcal{G}$ .

## 2.4 Dominance: Hypergraph and Directed Structure Inclusion

**Definition 9** (Observational dominance and equivalence). Let  $\mathcal{D}$  and  $\mathcal{D}'$  be latent-variable DAGs that have the same set  $V$  of visible variables. We say that  $\mathcal{D}$  dominates  $\mathcal{D}'$  when it accommodates at least all the probability distributions compatible with  $\mathcal{D}'$ , i.e. when  $\mathcal{M}(\mathcal{D}', V) \subseteq \mathcal{M}(\mathcal{D}, V)$ . In this case, we denote  $\mathcal{D} \succeq \mathcal{D}'$ .

If the sets of probability distributions compatible with  $\mathcal{D}$  and  $\mathcal{D}'$  are the same, we say that these DAGs are observationally equivalent.

If it is known that  $\mathcal{D} \succeq \mathcal{D}'$  but  $\mathcal{D}$  and  $\mathcal{D}'$  are not equivalent, we say that  $\mathcal{D}$  strictly dominates  $\mathcal{D}'$  and denote  $\mathcal{D} \succ \mathcal{D}'$ .

These are defined similarly for mDAGs, using Definition 7.

In this section, we will present a result that gives some circumstances under which we know that the latent-variable DAG  $\mathcal{D}$  dominates  $\mathcal{D}'$ , and we will show how this translates to the language of mDAGs. The result is presented as proposition 3.3(b) in [9]:

**Proposition 1** (Edge Dropping). Let  $\mathcal{D}$  and  $\mathcal{D}'$  be two latent-variable DAGs with the same sets of vertices. Let  $V$  be their set of visible variables and  $L$  be their set of latent

variables. If  $\mathcal{D}'$  is a subgraph of  $\mathcal{D}$  (as in Definition 2), then  $\mathcal{M}(\mathcal{D}', V) \subseteq \mathcal{M}(\mathcal{D}, V)$ .

*Proof.* Consider a probability distribution compatible with  $\mathcal{D}'$ . For  $v \in V$ , this means that  $X_v$  is  $\sigma$ - $(X_{\text{pa}_{\mathcal{D}'}(v)}, E_v)$  measurable. Since  $\mathcal{D}'$  is a subgraph of  $\mathcal{D}$ , we have that  $\text{pa}_{\mathcal{D}'}(v) \subseteq \text{pa}_{\mathcal{D}}(v)$ , for both visible and latent parents of  $v$ . Thus, the random variable  $X_v$  must also be  $\sigma$ - $(X_{\text{pa}_{\mathcal{D}}(v)}, E_v)$  measurable, and the probability distribution is also compatible with  $\mathcal{D}$ .  $\square$

This Proposition gives some dominance relations between latent-variable DAGs. From this result, we can derive two types of dominance relations between mDAGs.

**Proposition 2** (Directed Structure Dominance). *Let  $\mathcal{G} = (\mathcal{D}, \mathcal{B})$  and  $\mathcal{G}' = (\mathcal{D}', \mathcal{B})$  be two mDAGs that have the same simplicial complex  $\mathcal{B}$  but different directed structures. If  $\mathcal{D}'$  is a subgraph of  $\mathcal{D}$  (as in Definition 2), then  $\mathcal{M}(\mathcal{D}') \subseteq \mathcal{M}(\mathcal{D})$ .*

*Proof.* This follows from application of Proposition 1 to the visible-variable DAGs  $\mathcal{D}$  and  $\mathcal{D}'$ .  $\square$

**Proposition 3** (Simplicial Complex Dominance). *Let  $\mathcal{G} = (\mathcal{D}, \mathcal{B})$  and  $\mathcal{G}' = (\mathcal{D}, \mathcal{B}')$  be two mDAGs that have the same directed structure  $\mathcal{D}$  but different simplicial complexes. If  $\mathcal{B}' \subseteq \mathcal{B}$ , then  $\mathcal{G} \succeq \mathcal{G}'$ .*

*Proof.* This follows from application of Proposition 1 to the latent-variable DAGs associated to the mDAGs  $\mathcal{G}$  and  $\mathcal{G}'$  (substituting facets by latent nodes).

Starting from the latent-variable DAG associated with  $\mathcal{G}$ , we make use of Lemma 5 to add redundant latent variables correspondent to all the faces of  $\mathcal{B}$  (not just the facets). Removing edges from this new set of latent variables (that includes redundant ones), we can get to the latent-variable DAG associated with the simplicial complex  $\mathcal{B}'$ .  $\square$

Note that  $\mathcal{B}' \subseteq \mathcal{B}$ , as used in the definition of Proposition 3, means that every *face* of  $\mathcal{B}'$  is also a face of  $\mathcal{B}$ , not every *facet*. For example, in Figure 2.4, the facets (red bubbles) of the mDAG 2.4(b) are not a subset of the facets of 2.4(a). However, their *faces* are.

## 2.5 Graph Patterns

If we want to check the compatibility of a probability distribution with a causal structure, clearly we need to know which node of the mDAG refers to which

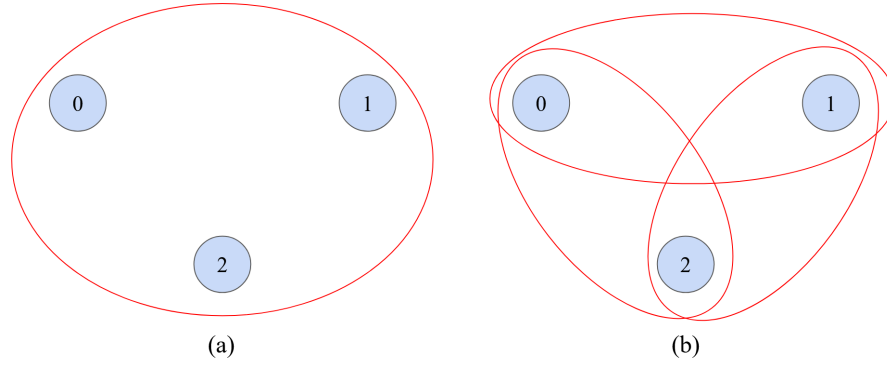


Figure 2.4: (a) An mDAG with abstract simplicial complex  $\mathcal{B} = \{\{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}, \{0,1,2\}\}$ . (b) An mDAG with abstract simplicial complex  $\mathcal{B}' = \{\{0\}, \{1\}, \{2\}, \{0,1\}, \{0,2\}, \{1,2\}\}$ . Since  $\mathcal{B}' \subseteq \mathcal{B}$ , Proposition 3 says that (a) dominates (b). The facets (inclusion-maximal elements of  $\mathcal{B}$  and  $\mathcal{B}'$ ) are indicated by the red bubbles.

variable. For example, the two causal structures of Figure 2.5 are the same up to assignments of variables to nodes, but they are *not* observationally equivalent structures.

We call the transformation of changing the assignment of variables to nodes a *relabelling* transformation. Furthermore, we call an unlabeled mDAG, that corresponds to an equivalence class under the relabelling transformation, a *graph pattern*.

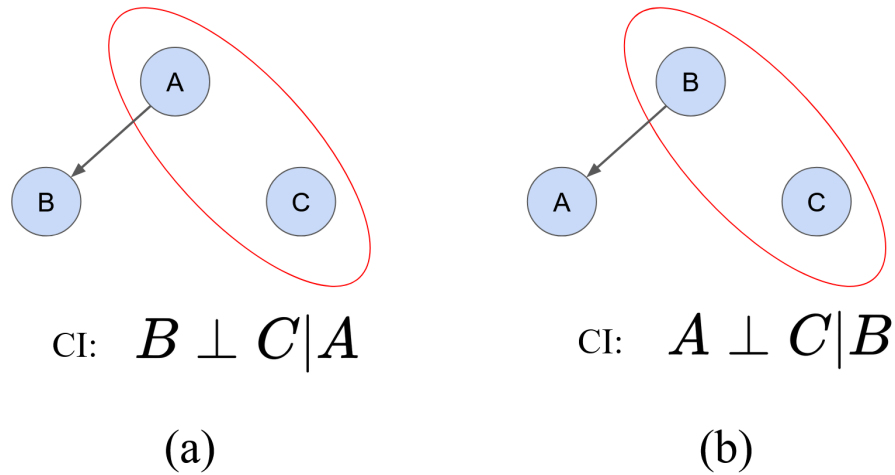


Figure 2.5: Two mDAGs on variables  $\{A, B, C\}$  that are equivalent under relabelling transformation, but are *not* observationally equivalent. We know that because they have different sets of CI Relations and, as we will see, this implies inequivalence.

In this work, we will talk about graph patterns instead of labeled mDAGs. This is so because our goal here is to study the general behavior of the causal

structures from a more abstract point of view, instead of a more practical one of explicitly finding which probability distributions are compatible with each mDAG. Since many mDAGs have symmetries, ignoring the labels makes our computational work much faster. For example, Table 2.1 compares the numbers of labeled mDAGs and graph patterns for fixed numbers of visible nodes.

Visible Nodes	2	3	4	5
Number of labeled mDAGs	6	225	61902	201863214
Number of graph patterns	4	46	2809	1718596

Table 2.1: Numbers of labeled mDAGs and graph patterns for fixed numbers of visible nodes. These numbers were found implementing Evans's Rules (Lemmas 4 and 5) in Python, as well as the relabelling transformation.

Therefore, we want to establish "unlabeled observational equivalences" and "unlabeled observational dominance relations" between graph patterns. We start by defining unlabeled observational equivalence:

**Definition 10** (Observational Equivalence of Graph Patterns). *We say that a relabelling class (graph pattern)  $\mathcal{G}$  is equivalent to another relabelling class  $\mathcal{G}'$  if some mDAG of  $\mathcal{G}$  is equivalent to an mDAG of  $\mathcal{G}'$ .*

Using the relabelling transformation, we see that if *some* mDAG of the first relabelling class is equivalent to an mDAG of the second relabelling class, then *every* mDAG of the first class is equivalent to an mDAG of the second class. Thus, Definition 10 is reasonable.

For properties like the Conditional Independence Relations, we will also define two graph patterns to have the same CI Relations if some element of the first relabelling class have the same CI Relations as an element of the second one.

To extend the definition of observational dominance to graph patterns, we do something similar:

**Definition 11** (Observational Dominance of Graph Patterns). *We say that a relabelling class (graph pattern)  $\mathcal{G}$  dominates another relabelling class  $\mathcal{G}'$  if any mDAG of  $\mathcal{G}$  dominates an mDAG of  $\mathcal{G}'$ .*

For this definition to be sensible we need to show that, if we have observational dominance both ways, then there is observational equivalence between the graph patterns. In other words, we need to show that the case presented in Figure 2.6 does not happen.

Indeed, suppose that mDAGs  $\mathcal{G}_1^A$  and  $\mathcal{G}_1^B$  belong to the graph pattern 1, while mDAG  $\mathcal{G}_2$  belongs to the graph pattern 2. Now, suppose that  $\mathcal{G}_1^A \succeq \mathcal{G}_2$ , while  $\mathcal{G}_2 \succeq \mathcal{G}_1^B$ . According to the first dominance,  $\mathcal{G}_1^A$  is compatible with at least the same volume of distributions as  $\mathcal{G}_2$ . According to the second dominance,  $\mathcal{G}_2$  is compatible with at least the same volume of distributions as  $\mathcal{G}_1^B$ . However,  $\mathcal{G}_1^A$  and  $\mathcal{G}_1^B$  are just a relabelling of each other; so, even if their compatible distributions are different, by symmetry we know that the volume of the compatible distributions must be the same. Therefore, since

$$\begin{aligned} \text{vol}(\mathcal{G}_1^A) &\geq \text{vol}(\mathcal{G}_2) \geq \text{vol}(\mathcal{G}_1^B) \\ \text{vol}(\mathcal{G}_1^A) &= \text{vol}(\mathcal{G}_1^B), \end{aligned}$$

we conclude that all volumes are equal, thus  $\mathcal{G}_2$  must be equivalent to both  $\mathcal{G}_1^A$  and  $\mathcal{G}_1^B$ . Therefore,  $\mathcal{G}_1^A$  and  $\mathcal{G}_1^B$  are the same mDAG.

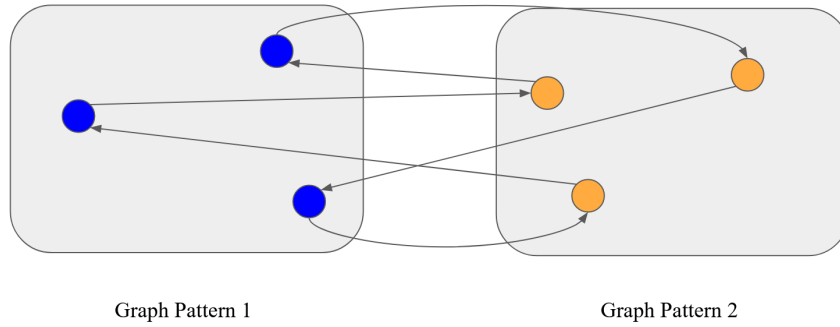


Figure 2.6: Here, blue and orange circles represent mDAGs, while arrows represent dominance relations between mDAGs. Two mDAGs inside the same graph pattern are related to each other by a relabelling of the visible nodes. In this hypothetical case, there is observational dominance both ways according to Definition 11, but there is no observational equivalence as in Definition 10. However, this case is impossible.

## 2.6 Computational Work: Constructing Metagraph

To deduce a proven-equivalence partition of causal structures (pink bubbles of Figure 1.2), we constructed a "metagraph", a graph where the nodes are the graph patterns. In this metagraph, we added undirected and directed edges, respectively meaning equivalences and dominance relations according to Definitions 10 and 11. The undirected edges were represented in the code by adding two arrows in



opposite directions, so dominance in both ways is the same as equivalence. Our code can be found in [14].

Propositions 2 and 3, that talk about dominance between mDAGs, give rise to directed edges. The propositions that will be presented in Section 3, that establish equivalence, give rise to undirected edges. The strongly connected components of the metagraph determine the proven-equivalence partition.

#### Summary of the chapter:

- Causal Structures are represented by Directed Acyclic Graphs (DAGs). A DAG can have visible (measured) and latent (unmeasured) nodes. In this work, we allow for unbounded cardinality of the latents.
- A probability distribution  $P$  over a set of variables  $V$  is *classically compatible* with a DAG  $G$  that has the visible nodes  $V$  if it obeys the constraints that are classically implied by  $G$ . This can be reformulated by saying that one can find measurable functions following the structure of the DAG such that the joint probability  $P$  is reproduced (Definition 3).
- We say that the DAG  $G$  *observationally dominates* the DAG  $G'$  if all the probability distributions compatible with  $G'$  are also compatible with  $G$ . We say that they are *observationally equivalent* if they have exactly the same set of compatible probability distributions.
- Constraints of the form of Conditional Independence (CI) relations are found by the method of d-separation. Inequality constraints, on the other hand, are harder to find.
- Lemmas 4 and 5 guarantee that, in the classical case, we can work with mDAGs (structures where all the latents in a DAG are switched for facets of a simplicial complex; see Figure 2.3).
- In this work, we look for the observational equivalence classes of graph patterns (unlabelled mDAGs).

# Chapter 3

## Ways to Show Equivalence

In this section, we are going to show some propositions that provide conditions under which two mDAGs are known to be equivalent. These results give us our proven-equivalence partition (pink bubbles of Figure 1.2), but we still do not know if they are enough to find the final equivalence classes when we have more than 3 observable nodes. The number of elements of the proven-equivalence partition is an upper bound on the number of equivalence classes (green dashed bubbles of Figure 1.2).

Sections 3.2.2 and 3.2.3 show original theoretical results.

The equivalences provided by the propositions of this section are going to give the undirected edges in our metagraph, and the proven-equivalence partition is given by the strongly connected components of the metagraph.

### 3.1 HLP Proposition

The first equivalence result that we reproduce here is Theorem 26.4 of [8]:

**Proposition 4** (HLP Proposition). *Let  $\mathcal{G}$  be a mDAG, and  $X$  and  $Y$  be two of its visible nodes. If we have:*

- $pa_{\mathcal{G}}(X) \subseteq pa_{\mathcal{G}}(Y)$ .
- $X$  is included in at least one facet.
- If  $X \in B$  for a facet  $B \in \mathcal{B}$ , then  $Y \in B$ .

*Then  $\mathcal{G}$  is equivalent to the mDAG  $\mathcal{G}'$  defined by adding an edge  $X \rightarrow Y$  to  $\mathcal{G}$ .*

*Proof.* Presented in Appendix B.1. □

For example, by the HLP proposition it follows that the two mDAGs of Figure 3.1 are observationally equivalent.

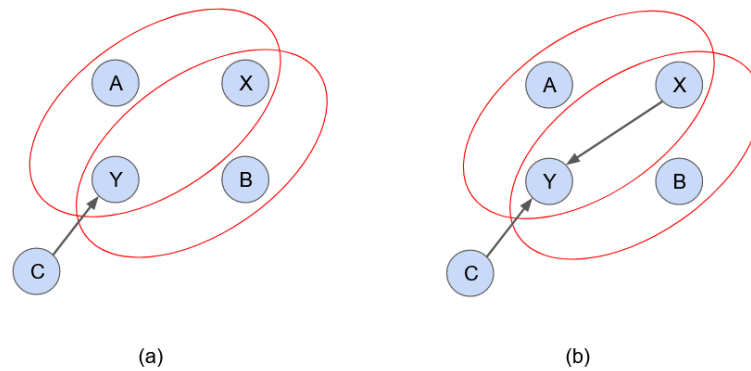


Figure 3.1: Two mDAGs that can be proven equivalent by the HLP proposition [8].

Results Applied	N° of Pink Bubbles 3 visible nodes	N° of Pink Bubbles 4 visible nodes
HLP	27	1597

Table 3.1: Number of elements of the proven-equivalence partition (number of pink bubbles of Figure 1.2) in which the 46 graph patterns at 3 visible nodes and the 2809 graph patterns at 4 visible nodes divide, when we use the HLP Proposition (Proposition 4). The dominance relations of Section 2.4 are also included.

Implementing this in the metagraph (together with the dominance relations of Section 2.4), the numbers of elements of the proven-equivalence partition that we get are presented in Table 3.1.

## 3.2 Evans's Proposition and its Spin-Offs

Proposition 6.1 of [9] is a very important equivalence result, that we are going to call "Evans' Proposition". We reproduce it below, and Figure 3.2 may help to understand the statement.

**Proposition 5** (Evans's Proposition). *Let  $\mathcal{G}$  be an mDAG containing a facet  $B = C \cup D$ ,  $C$  and  $D$  disjoint, such that:*

1.  $pa_{\mathcal{G}}(C) \subseteq pa_{\mathcal{G}}(d)$  for each  $d \in D$
2. For every  $c \in C$ ,  $B$  is the only facet that contains  $c$ .

*Then,  $\mathcal{G}$  is equivalent to the mDAG  $\mathcal{G}'$  defined by adding edges  $c \rightarrow d$  for every  $c \in C$ ,  $d \in D$  and removing  $B$  from  $\mathcal{B}$ , so that now  $C$  and  $D$  are inclusion-maximal.*

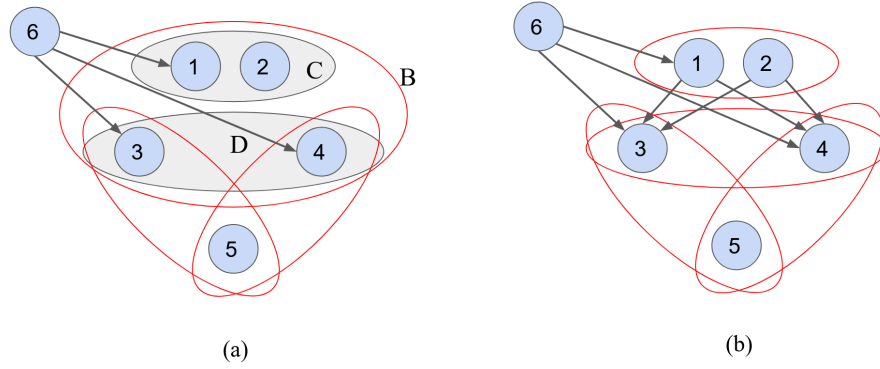


Figure 3.2: Two mDAGs that can be proven equivalent by Evans proposition [9].

As it turns out, Evans's Proposition can be subsumed by a result that only talks about the splitting of the facets, called Weak Face Splitting, while the addition of edges is taken care of by the HLP Proposition.

### 3.2.1 Weak Face Splitting

In this section, we formulate the Weak Face Splitting Proposition:

**Proposition 6** (Weak Face Splitting). *Let  $\mathcal{G}$  be an mDAG containing a facet  $B = C \cup D$ ,  $C$  and  $D$  disjoint, such that:*

1.  $pa_{\mathcal{G}}(C) \cup C \subseteq pa_{\mathcal{G}}(d)$  for each  $d \in D$
2. *If for a face  $B'$  we have  $c \in B'$  for some  $c \in C$ , then  $B' \subseteq B$ .*

*Then,  $\mathcal{G}$  is equivalent to the mDAG  $\mathcal{G}'$  defined by removing  $B$  from  $\mathcal{B}$ , so that now  $C$  and  $D$  are inclusion-maximal.*

*Proof.* Presented in Appendix B.2. □

It is easy to see that Proposition 6 is weaker than Proposition 5: in Proposition 6, we need the edges  $c \rightarrow d$  to be present in order to satisfy condition 1. However, as we will see below, when we associate Proposition 6 with the HLP Proposition we get a result that implies Proposition 5.

**Lemma 1** (Weak FS+HLP $\geq$ Evans). *The application of Weak Face Splitting (Proposition 6) together with the HLP Proposition (Proposition 4) subsumes Evans's Proposition (Proposition 5).*

*Proof.* Suppose that we have an mDAG  $\mathcal{G}$  that satisfies the conditions of Proposition 5. This means that  $\mathcal{G}$  contains a facet  $B = C \cup D$  with  $C$  and  $D$  disjoint and, for each  $c \in C$  and  $d \in D$ :

1.  $pa_{\mathcal{G}}(c) \subseteq pa_{\mathcal{G}}(d)$
2.  $c$  is included in the facet  $B$  and the only facet that includes  $c$  is  $B$ .
3.  $B$  also includes  $d$ .

This shows that the premises of the HLP Proposition (Proposition 4) are satisfied for  $X = c$  and  $Y = d$ . Thus, HLP implies that one can add edges  $c \rightarrow d$  for every  $c \in C$  and  $d \in D$ .

With this addition, now the mDAG satisfies all the conditions of Proposition 6, the Weak Face Splitting rule. This means that we can remove the facet  $B$  and add facets  $C$  and  $D$ , and get the final mDAG that could have been obtained by Proposition 5. Thus, Proposition 5 is a special case of the joint application of Propositions 6 and 4.  $\square$

Hence, we separated the part of Evans's Proposition that splits the facets, that is included in the Weak Face Splitting Proposition, from the part that adds the edges  $c \rightarrow d$ , that is taken care of by the HLP Proposition. This is made explicit in Figure 3.3, where we break the example of Figure 3.2 into two steps. However, this is not the only case in which the HLP Proposition is useful; we make that explicit in the Lemma below.

**Lemma 2** (Weak FS+HLP > Evans). *The application of Weak Face Splitting (Proposition 6) together with the HLP Proposition (Proposition 4) is strictly stronger than Evans's Proposition (Proposition 5).*

*Proof.* The equivalence between the mDAGs of Figure 3.1 cannot be shown by Evans's Proposition.  $\square$

Using HLP, Weak FS, and the dominance relations of Section 2.4, we get the numbers of elements of the proven-equivalence partitions (strongly connected components of the metagraph) for 3 and 4 observable nodes that are shown in Table 3.2.

When we have a set of equivalence classes obtained by one equivalence result and we implement a different non-redundant result, the final proven-equivalence partition is a coarse-graining of the original one. In other words, some of the pink

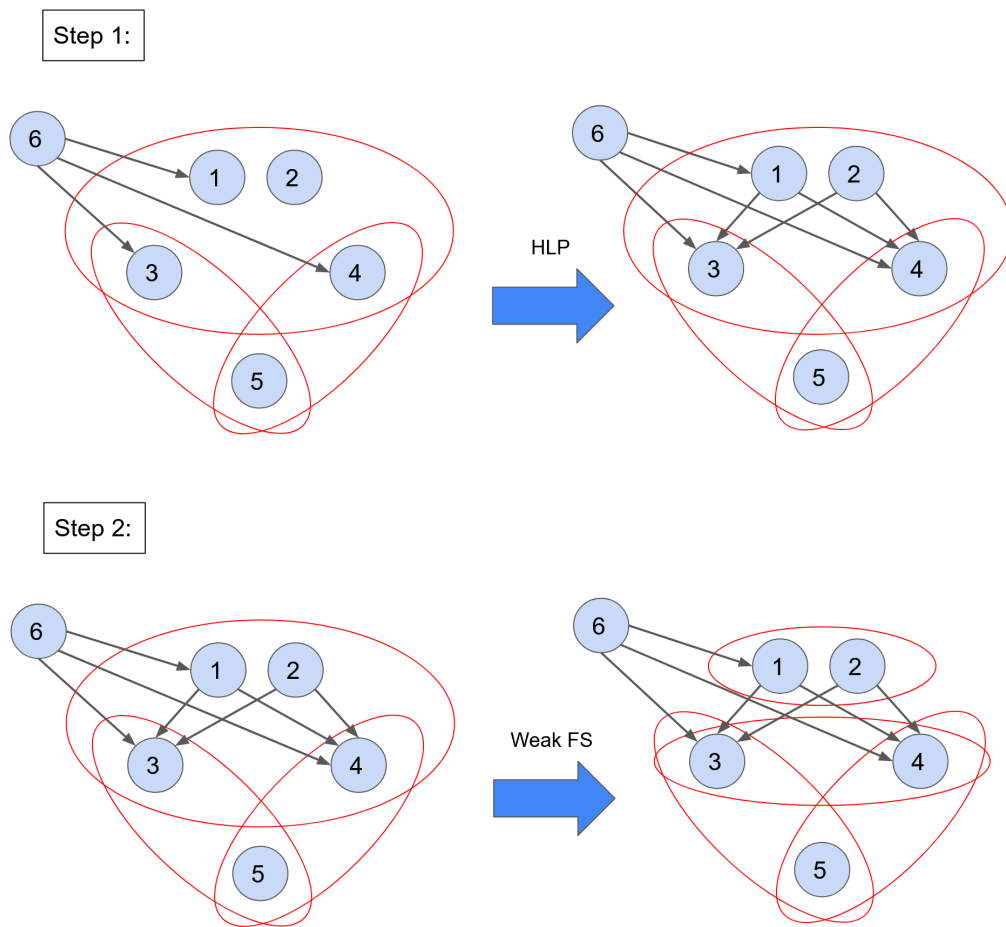


Figure 3.3: The equivalence of Figure 3.2 can be shown by HLP together with Weak FS (Propositions 4 and 6).

Results Applied	N° of Pink Bubbles 3 visible nodes	N° of Pink Bubbles 4 visible nodes
HLP+Weak FS	8	416

Table 3.2: Number of elements of the proven-equivalence partition (pink bubbles of Figure 1.2) determined by Weak Face Splitting (Proposition 6) together with the HLP Proposition (Proposition 4). The dominance relations of Section 2.4 are also included.

bubbles of Figure 1.2 merge into larger pink bubbles. Looking at Tables 3.1 and 3.2, we see that HLP+Weak FS is better than HLP alone. This means that Weak FS (and Evans’s Proposition) shows some equivalences that HLP cannot show.

### 3.2.2 Moderate Face Splitting

Our first original theoretical result is an extension of Weak Face Splitting, where we show that condition 2 can be relaxed.

**Proposition 7** (Moderate Face Splitting). *Let  $\mathcal{G}$  be an mDAG containing a face  $B = C \cup D$ ,  $C$  and  $D$  disjoint, such that:*

1.  $pa_{\mathcal{G}}(C) \cup C \subseteq pa_{\mathcal{G}}(d)$  for each  $d \in D$
2. *If for a face  $B'$  we have  $c \in B'$  for any  $c \in C$ , then  $D \subset B'$ .*

*Then,  $\mathcal{G}$  is equivalent to the mDAG  $\mathcal{G}'$  defined by removing  $B$  for  $\mathcal{B}$  (so  $C$  and  $D$  become inclusion-maximal).*

Note the difference in premise 2 between this proposition and Propositions 5 and 6: there, we had that  $B' \subseteq B$  and here, we have  $D \subset B'$ .

Proposition 7 follows from a simple application of Proposition 6 when treating all latent variables of the latent-variable DAG associated to  $\mathcal{G}$  (apart from the one associated to the facet that we want to split) as if they were visible. We can do that because measuring the latent variables without changing any of their other properties should not change the marginal probability distributions over the nodes that were already visible. The second hypothesis in Proposition 7 comes from the fact that all the parents of  $c \in C$ , visible *and* latent, need to be parents of every  $d \in D$ .

This is indeed more flexible than the second hypothesis of Proposition 6, which says that in  $\mathcal{G}$  the nodes  $c \in C$  are not in any facet other than  $B$ . The second hypothesis of Proposition 7 allows for  $c \in C$  to be in facets that include nodes outside of  $B$  if they also include all of  $D$  (see Figure 3.4).

The number of elements of the proven-equivalence partition we get when applying this result to the cases of 3 and 4 visible nodes are presented in Table 3.3.

### 3.2.3 Strong (Simultaneous) Face Splitting

In Figure 3.5, we have two mDAGs that cannot be shown equivalent by Proposition 5 nor by HLP together with our Propositions 6 or 7.

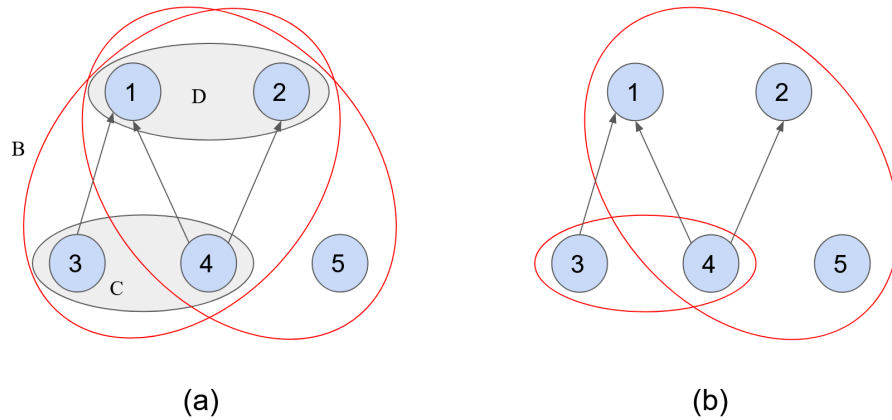


Figure 3.4: The facet  $B$  in mDAG (a) could not be split into  $C$  and  $D$  by Proposition 6 (because  $4 \in C$  is part of another facet), but it can be split by Proposition 7.

Results Applied	N° of Pink Bubbles 3 visible nodes	N° of Pink Bubbles 4 visible nodes
HLP + Moderate FS	8	408

Table 3.3: Number of elements of the proven-equivalence partition (pink bubbles) determined by HLP + Moderate Face Splitting (Propositions 4 and 7). The dominance relations of Section 2.4 are also included.

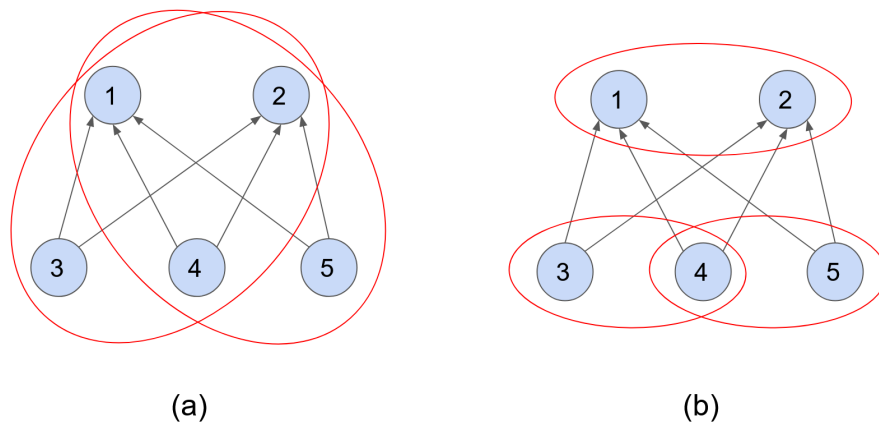


Figure 3.5: Simultaneous splitting example

As it turns out, however, these mDAGs are indeed observationally equivalent. This can be shown by the following proposition, which is our second original theoretical result:

**Proposition 8 (Strong Face Splitting).** *Let  $\mathcal{G}$  be an mDAG containing a sequence of facets  $B_1, \dots, B_n$  that share the subset  $D \in \cap_{i=1, \dots, n} B_i$ . Denote  $C_i \equiv B_i \setminus D$ . If the following conditions hold:*



1.  $pa_{\mathcal{G}}(C_i) \cup C_i \subseteq pa_{\mathcal{G}}(d)$  for each  $d \in D, i = 1, \dots, n$
2. If for a face  $B'$  we have  $c \in B'$  for some  $c \in \cup_{i=1, \dots, n} C_i$ , then  $D \subseteq B'$ .

Then,  $\mathcal{G}$  is equivalent to the mDAG  $\mathcal{G}'$  defined by removing the facets  $B_1, \dots, B_n$  and adding the facets  $C_1, \dots, C_n$  and  $D$ .

*Proof.* Presented in Appendix B.3. □

We also call this result "Simultaneous Face Splitting", because it shows that multiple facets can be split simultaneously. In the example of Figure 3.5, we have  $D = \{1, 2\}$ ,  $C_1 = \{3, 4\}$ ,  $C_2 = \{4, 5\}$ ,  $B_1 = \{1, 2, 3, 4\}$  and  $B_2 = \{1, 2, 4, 5\}$ .

Table 3.4 presents the number of elements of the proven-equivalence partition that we get when applying the dominances of Section 2.4, HLP Proposition, and Strong Face Splitting together. These are all the results that are presently known, so Table 3.4 shows the best upper bounds that we currently have on the number of equivalence classes.

Results Applied	N° of Pink Bubbles 3 visible nodes	N° of Pink Bubbles 4 visible nodes
HLP + Strong FS	8	396

Table 3.4: Number of elements of the proven-equivalence partition (pink bubbles of Figure 1.2) determined by HLP + Strong Face Splitting (Propositions 4 and 8). The dominance relations of Section 2.4 are also included.

#### Summary of the chapter:

- HLP [8] and Evans [9] presented different conditions under which two mDAGs are proven to be observationally equivalent.
- In Sections 3.2.2 and 3.2.3, we present two original extensions to Evans's result.
- These results allow us to obtain the proven-equivalence partition (pink bubbles of Figure 1.2). We found it for mDAGs with 3 and 4 visible variables.

# Chapter 4

## Comparisons that Show Inequivalence

In this section, we are going to present all the known methods to show that two mDAGs are *not* equivalent. The application of these methods defines the proven-inequivalence partition (gray bubbles of Figure 1.2). Two mDAGs are known to be inequivalent if they are **not** in the same element of the partition.

We will pick one representative of each element of the proven-equivalence partition (pink bubble of Figure 1.2) and make the comparisons between the representatives. In the numerical results for 3 and 4 visible variables that we will present, we picked representatives of the elements of the proven-equivalence partition found by HLP + Strong FS.

### 4.1 Graphical Methods

#### 4.1.1 Skeletons

We are going to start showing inequivalences by using a method that was introduced in [9]. For that, we first define the *Skeleton* of an mDAG:

**Definition 12** (Skeleton). *Let  $\mathcal{G} = (D, \mathcal{B})$  be an mDAG. We define the skeleton of  $\mathcal{G}$  by the undirected graph with the same nodes as  $D$  and with an edge between nodes  $u$  and  $w$  when there is an arrow between them in  $D$  or when  $u, w \in B$  for some  $B \in \mathcal{B}$ .*

You can see an example of this definition in Figure 4.1.

Below, we reproduce Proposition 6.5 of [9]:

**Proposition 9.** *(Skeleton Comparison) Let  $\mathcal{G}$  and  $\mathcal{G}'$  be two mDAGs, and suppose that the state space of the random variables  $\mathcal{X}_V$  is discrete. If  $\mathcal{G}$  and  $\mathcal{G}'$  have different skeletons, then they are inequivalent.*

Therefore, if we assume that  $\mathcal{X}_V$  is discrete, the comparison of Skeletons can give us some proven inequivalences. This allows us to organize the mDAGs into

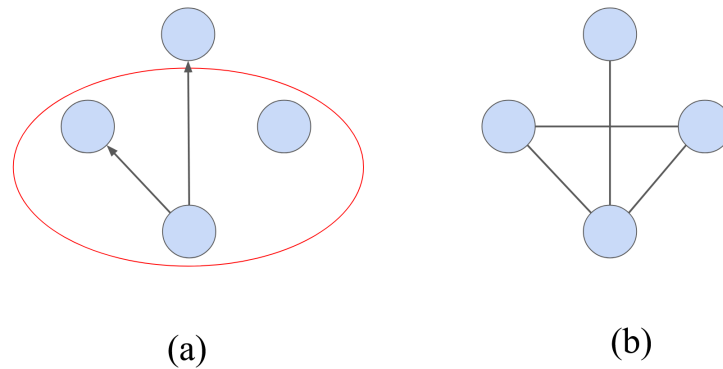


Figure 4.1: (a) An mDAG; and (b) Its skeleton.

Comparisons Made	N° of Gray Bubbles 3 visible nodes	N° of Gray Bubbles 4 visible nodes
Skeletons	4	11

Table 4.1: Numbers of Skeleton Classes (gray bubbles of Figure 1.2) in which the graph patterns divide. These numbers give lower bounds on the number of equivalence classes.

a first proven-inequivalence partition. We will call the elements of this proven-inequivalence partition "Skeleton classes". The number of Skeleton classes defines a first lower bound on the number of observational equivalence classes. Implementing that for mDAGs of 3 and 4 observable variables, we get the numbers in Table 4.1.

### 4.1.2 CI Relations

As explored in the example of Figure 1.1, a causal structure can give two types of constraints on the probability distributions compatible with it: equality and inequality constraints. The equality constraints are the Conditional Independence (CI) Relations, like the ones exemplified in equations (1.1) and (1.2).

If two mDAGs have different sets of CI Relations among the visible variables, there is at least one probability distribution that is realizable by one of the mDAGs but not by the other. This is a special case of the Proposition 12 of [15], as will be discussed in Section 4.1.3. Therefore, we can use the comparison of CI Relations to prove inequivalences between mDAGs, classifying them into another proven-inequivalence partition. The elements of this partition are called CI classes.

The problem of finding all the CI Relations of a given causal structure is solved by the method of d-separation, as presented in Section 2.2. Implementing this

Comparisons Made	N° of Gray Bubbles 3 visible nodes	N° of Gray Bubbles 4 visible nodes
CI	5	25
CI+Skeletons	6	41

Table 4.2: Numbers of CI and CI+Skeletons Classes (gray bubbles of Figure 1.2) in which the graph patterns divide. These numbers give lower bounds on the number of equivalence classes.

method for the case of 3 and 4 observable variables, we get the numbers of CI classes presented in Table 4.2.

As we can see here, the comparisons of Skeletons and CI Relations are not redundant to each other: in 4 observable variables, Skeleton+CI is better than each one separately, so each one of them shows inequivalences that the other one does not.

### 4.1.3 e-separation Relations

In [16], an extension of d-separation was introduced, called *e-separation*. While the d-separation method gives the Conditional Independences of a causal structure, the e-separation method gives inequality constraints, that are explicitly provided in [15] (note that not all inequality constraints come from e-separation).

**Definition 13** (e-separation). *Let  $\mathcal{G}$  be an mDAG with nodes  $V$ . Let  $A, B, C$  and  $D$  be four subsets of  $V$ . We say that  $A$  and  $B$  are e-separated by  $C$  after deletion of  $D$  if  $A \perp_d B | C$  in the subgraph  $\mathcal{G}_{V \setminus D}$ .*

Clearly, d-separation is a special case of e-separation, with  $D = \emptyset$ .

In the Proposition 12 of [15], it is shown that if an mDAG  $\mathcal{G}$  does not have a certain e-separation relation, then there exists some probability distribution that violates the inequality derived from that e-separation relation and is compatible with  $\mathcal{G}$ . Therefore, if two mDAGs exhibit different sets of e-separation relations, they are inequivalent. Thus, as we did with Skeleton and CI, we can also find the proven-inequivalences partitions by e-separation for 3 and 4 observable variables. The numbers of elements in these partitions are shown in Table 4.3.

The comparison of e-separation Relations shows itself as a better criterion than both comparisons of Skeletons and CI Relations. This happens because d-separation is a special case of e-separation, and different Skeletons also imply different e-separation Relations. The last statement is true because nodes  $u$  and  $v$

Comparisons Made	N° of Gray Bubbles 3 visible nodes	N° of Gray Bubbles 4 visible nodes
e-separation	6	55

Table 4.3: Numbers of elements in the e-separation proven-inequivalence partitions (gray bubbles of Figure 1.2) in which the graph patterns divide. These numbers give lower bounds on the number of equivalence classes. They are the same in both rows because the comparison of Skeletons is a special case of the comparison of e-separation Relations.

are connected by an edge in the skeleton if and only if  $u$  and  $v$  are not e-separated by the empty set upon deleting  $D$  for any set  $D \subseteq V$ .

#### 4.1.4 Pairs of Densely Connected Nodes

In this section, we show a test for inequivalences that is based on the main result of [17]. This test depends on the concept of *densely connected nodes*. To define this, we will first define *districts* of an mDAG and the *closure* of a set of nodes.

**Definition 14** (Districts). *Let  $\mathcal{G}$  be an mDAG with simplicial complex  $\mathcal{B}$ . Let  $D$  be a set of nodes of  $\mathcal{G}$  that are sequentially connected to each other by facets  $B \in \mathcal{B}$ ; i.e., such that if  $v, w \in D$ , then there is a sequence  $v = u_0, u_1, u_2, \dots, u_{n-1}, w = u_n$  such that for all  $i$ ,  $\{u_i, u_{i+1}\} \subseteq B_i \in \mathcal{B}$ . If  $D$  is inclusion maximal, we say that it is a district. We say that a district is a "common-cause district" if it contains more than one node.*

*If  $v$  is a node of  $\mathcal{G}$ , we are going to denote the district of  $\mathcal{G}$  in which  $v$  belongs by  $dis_{\mathcal{G}}(v)$ . Similarly, for a set of nodes  $S$ , we have  $dis_{\mathcal{G}}(S) = \cup_{v \in S} dis_{\mathcal{G}}(v)$ .*

In the mDAG of Figure 4.2, for example, we have three districts:  $\{A, B, C, H\}$ ,  $\{D\}$  and  $\{E, F, G\}$ . Out of these,  $\{A, B, C, H\}$  and  $\{E, F, G\}$  are common-cause districts.

**Definition 15** (Closure of a set of nodes  $B$ ). *Let  $\mathcal{G}$  be an mDAG with nodes  $V$ . Let  $B \subseteq V$ . Set  $B^{(0)} \equiv B$  and construct a sequence applying alternately:*

$$B^{(i+1)} = dis_{B^{(i)}}(B) \qquad B^{(i+2)} = an_{B^{(i+1)}}(B)$$

*Where  $dis_{\mathcal{G}}(S)$  is the district of  $S$  in  $\mathcal{G}$ , and  $an_{\mathcal{G}}(S)$  is the set of ancestors of the nodes of  $S$  together with the nodes of  $S$  themselves. Each one of these operations removes vertices from the set  $B^{(i)}$  or keeps it the same. When the process terminates, we call the resulting set the closure of  $B$  and denote it  $\langle B \rangle_{\mathcal{G}}$ .*

Now, we can say what it means for nodes  $v$  and  $w$  to be densely connected.

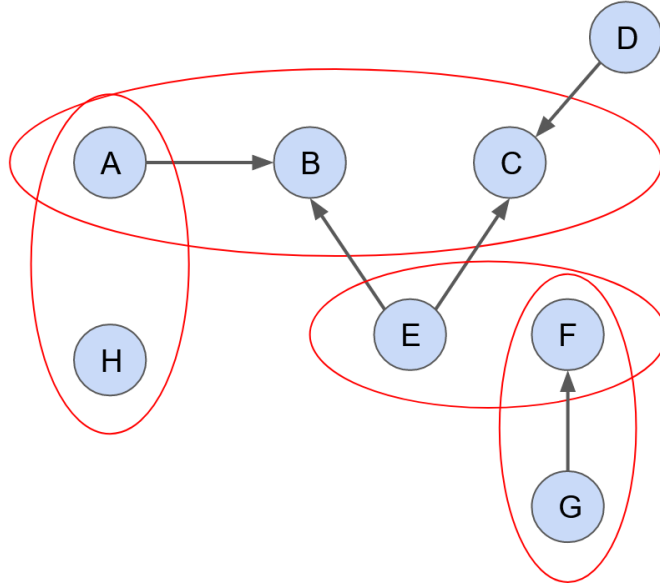


Figure 4.2: The districts of this mDAG are  $\{A, B, C, H\}$ ,  $\{D\}$  and  $\{E, F, G\}$ .

**Definition 16** (Densely Connected Nodes). *A pair of vertices  $v \neq w$  in an mDAG  $\mathcal{G}$  is said to be densely connected if any of the following conditions are satisfied:*

- $v \in pa_{\mathcal{G}}(\langle w \rangle_{\mathcal{G}})$
- $w \in pa_{\mathcal{G}}(\langle v \rangle_{\mathcal{G}})$
- $\langle \{v, w\} \rangle_{\mathcal{G}}$  is a bidirected-connected set

where we say that a set of nodes  $S$  is bidirected-connected if every vertex of  $S$  can be reached from every other using a path of nodes, all in  $S$ , that are connected by facets.

In a recent work [17], Evans showed that the specific probability distribution where  $v = w$  and all other variables are random is realizable by a causal structure  $\mathcal{G}$  if and only if the nodes  $v$  and  $w$  are densely connected in  $\mathcal{G}$ .

Therefore, if we have a structure  $\mathcal{G}_1$  where  $v$  and  $w$  are densely connected and a different structure  $\mathcal{G}_2$  where they are not, [17] explicitly shows one probability distribution that is realizable by  $\mathcal{G}_1$  but not by  $\mathcal{G}_2$ . This means that  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are surely not equivalent.

Implementing this comparison for structures of 3 and 4 observable vertices, we get proven-inequivalence partitions whose numbers of elements are presented in Table 4.4.

Comparisons Made	N <sup>o</sup> of Gray Bubbles 3 visible nodes	N <sup>o</sup> of Gray Bubbles 4 visible nodes
Dense Connectedness (DC)	4	11
DC+Skeletons	5	22
DC+CI	5	27
DC + e-separation	6	58

Table 4.4: Numbers of elements of the Dense Connectedness, Dense Connectedness+Skeletons, Dense Connectedness+CI and Dense Connectedness+ e-separation proven-inequivalence partitions (gray bubbles of Figure 1.2) in which the graph patterns divide. These numbers give lower bounds on the number of equivalence classes.

### 4.1.5 Only Hypergraphs

Another method of showing inequivalences is presented in Proposition 6.8 of [9]:

**Proposition 10.** (*Only Hypergraphs Rule*) Let  $\mathcal{G}$  and  $\mathcal{G}'$  be two mDAGs with visible nodes  $V$  that do not have any directed edge. If  $\mathcal{G} \neq \mathcal{G}'$ , then  $\mathcal{M}(\mathcal{G}, V) \neq \mathcal{M}(\mathcal{G}', V)$ .

*Proof.* Presented in Appendix C. □

Note that this method has a difference from the other graphical methods shown in this section: not all mDAGs can be compared by Proposition 10. For example, every mDAG has a Skeleton that can be compared to the Skeleton of any other mDAG. Proposition 10, on the other hand, only establishes comparisons between mDAGs that are equivalent to a Hypergraph-Only mDAG.

This is explained in Figure 4.3. There, we can split the first gray bubble into four gray bubbles, because all the elements of the proven-equivalence partition (pink bubbles) inside this gray bubble do contain a Hypergraph-Only mDAG, so Proposition 10 says that they are all inequivalent. However, the second gray bubble still cannot be split, because it has a pink bubble that does not have any Hypergraph-Only mDAG, thus cannot be compared to the others by Proposition 10. We know that this gray bubble will be split into at least five gray bubbles, but we do not know where the mentioned pink bubble will belong.

For cases like the second gray bubble of Figure 4.3, even if we cannot split it, we have an improvement in the lower bound on the number of equivalence classes: we know that the referred gray bubble will be split in at least five parts, even if we still do not know how. Hence, Table 4.5 below shows how this lower bound changes when apply Proposition 10 on top of the e-separation proven-

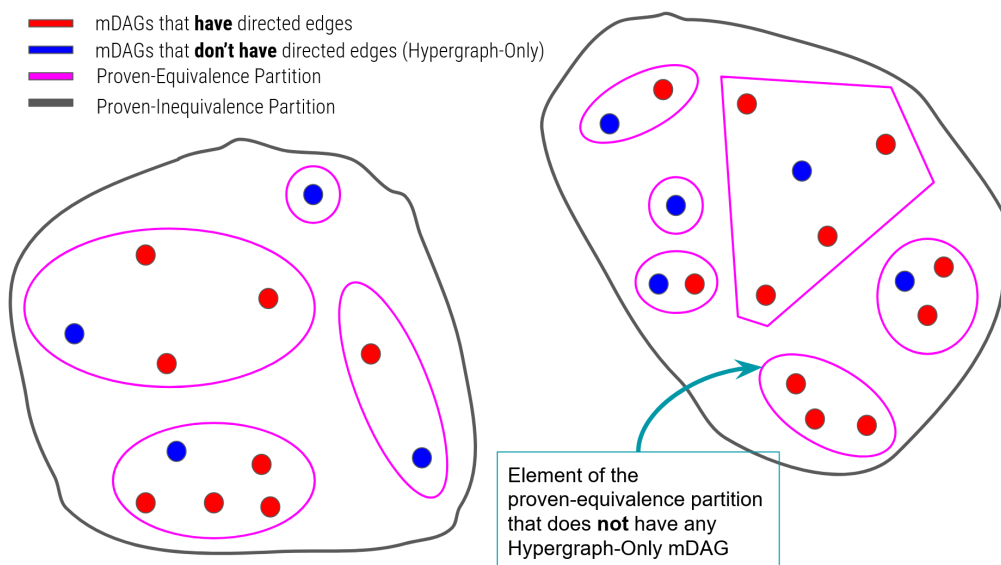


Figure 4.3: The mDAGs that do not have any directed edge (Hypergraph-Only mDAGs) are represented by blue dots, while the other mDAGs are red dots.

Comparisons Made	N° of Gray Bubbles	
	3 visible nodes	4 visible nodes
OH + e-separation	7	64
OH + e-separation + DC	7	67

Table 4.5: Numbers of elements of the Only Hypergraphs (OH) + e-separation and the Only Hypergraphs + e-separation + Dense Connectedness proven-inequivalence partitions (gray bubbles of Figure 1.2) in which the graph patterns divide. These numbers give lower bounds on the number of equivalence classes.

inequivalence partition and on top of the e-separation + Dense Connectedness proven-inequivalence partition.

Looking at the numbers of the Tables 4.3, 4.4 and 4.5, we see that the methods of comparing e-separation Relations and densely connected pairs and implementing the Only Hypergraphs Rule are not redundant to each other; each one of these three methods can show inequivalences that the other two cannot. The best lower bounds emerge when we apply these three together: 7 for 3 visible variables and 67 for 4 visible variables. Note that none of these cases is solved yet: these lower bounds still do not meet our upper bounds, which are 8 for 3 visible variables and 369 for 4 visible variables.



## 4.2 Comparison of Infeasible Supports

In this section, we present a stronger method to show inequivalence between two mDAGs, that explicitly assumes that the nodes respond to their parents by deterministic functions. This method is based on [18], and it deals with the notion of *possibilistic compatibility*, which we will define below.

An important *caveat* on this section is that it can only be applied when the visible variables have finite cardinalities; i.e., when the observables have a finite number of possible outcomes. Furthermore, the method is also only valid classically, because the algorithm presented in Appendix D consists in assigning deterministic outcomes to each valuation of the parents of a node.

As discussed before, the central problem in Causal Inference is the Probabilistic Causal Compatibility (Definitions 5 and 7) problem. In this section, we turn our attention to a different, strictly weaker, question: the *Possibilistic Causal Compatibility*. We will check if the events that are possible under our probability distribution, i.e. the events that have a probability different from zero, are the events classically allowed by the causal structure. To do so, let us define the support of a probability distribution:

**Definition 17** (Support). *Let  $P : \mathcal{X} \rightarrow [0, 1]$  be a probability distribution. The elements of  $\mathcal{X}$  are called “events”. The support of  $P$  is the set of possible events:*

$$S(P) = \{\omega \in \mathcal{X} | P(\omega) > 0\}$$

A probability distribution  $P$  is said to be *possibilistically compatible* with an mDAG  $\mathcal{G}$  if there exists some probability distribution  $\tilde{P}$  that is probabilistically compatible with  $\mathcal{G}$  and has the same support as  $P$ .

If a probability distribution is not possibilistically compatible with an mDAG, it will also not be probabilistically compatible with it. Thus, if two mDAGs have different sets of compatible supports, they cannot be equivalent. It is still an open question whether *possibilistic* equivalence implies *probabilistic* equivalence; in that case, the comparison of compatible supports would completely characterize the equivalence classes of mDAGs.

In Appendix D, we present the algorithm designed by TC Fraser in [18] to find the set of supports that are compatible with a given causal structure  $\mathcal{G}$ . Implementing this algorithm, we can compare different mDAGs in terms of the set of Infeasible Supports, thus getting a better fine graining of the proven-inequivalence

partition.

As justified at the end of Appendix D, here we consider all visible variables to be binary. Thus, the maximum number of events that we have is  $2^{|V|}$ , where  $|V|$  is the number of visible variables.

### 4.2.1 Infeasible Supports beyond e-separation

For a given mDAG, the e-separation relations themselves may already rule out some supports. As an example of this, imagine a causal structure that implies the CI relation  $A \perp B$ , where  $A$  and  $B$  are binary variables. Suppose that  $S$  is a support that does not contain any events involving  $\{A = 0, B = 1\}$  or  $\{A = 1, B = 0\}$ . This implies that in any probability distribution that has the support  $S$ , the variables  $A$  and  $B$  are perfectly correlated. This is incompatible with  $A \perp B$ , thus the support  $S$  is ruled out by  $A \perp B$ .

Since the algorithm presented in Appendix D is a brute force search, it is computationally expensive. Thus, to save time in the computation, we only apply this algorithm for supports that are not already ruled out by the e-separation relations of the mDAG in question.

### 4.2.2 Inequivalences Shown by Comparison of Infeasible Supports

To check if a support is compatible with an mDAG, we apply the algorithm of Appendix D with  $k = s$ , where  $k$  is the cardinality of the latents and  $s$  is the number of events of the given Support.

We did that for the cases of 3 and 4 visible variables, successively increasing the number of events of the Supports that we were checking and finding all the Infeasible Supports beyond e-separation of an mDAG at a given number of events. In Table 4.6, we present the number of elements of the proven-inequivalence partitions we get when applying all the available graphical methods together with the comparison of Infeasible Supports up to different numbers of events.

Maximum Number of Events	N° of Gray Bubbles 3 visible nodes	N° of Gray Bubbles 4 visible nodes
2	7	77
3	8	148
4	8	278
5	8	307
6	8	332
7	8	333

Table 4.6: Numbers of elements of the e-separation + Only Hypergraphs + Dense Connectedness + Support proven-inequivalence partitions (gray bubbles of Figure 1.2) in which the graph patterns divide. These numbers give lower bounds on the number of equivalence classes.

#### Summary of the chapter:

- The known methods to prove that two mDAGs are *not* observationally equivalent are: the comparison of skeletons, comparison of CI relations, comparison of e-separation relations, comparison of the pairs of densely connected nodes, the only hypergraphs rule and the comparison of Infeasible Supports.
- These methods give us the proven-inequivalence partition (grey bubbles of Figure 1.2). We found it for 3 and 4 visible nodes, checking supports up to 7 events.
- The comparisons of skeletons and CI relations are subsumed by the comparison of e-separation relations. On the other hand, the methods of comparison of e-separation relations, comparison of the pairs of densely connected nodes and the only hypergraphs rule are *not* redundant to each other.

# Chapter 5

## 3 and 4 Observable Nodes

In this section, we sum up and discuss the numerical results presented in Sections 3 and 4 for the classification of mDAGs with 3 and 4 observable nodes.

First, for 3 observable nodes, the number of elements of the proven-equivalence partition that we have (pink bubbles of Figure 1.2) for HLP+Weak Face Splitting exactly matches the number of elements of the proven-inequivalence partition (gray bubbles of Figure 1.2) for Supports at 3 events: Table 3.2 presents the upper bound of 8 and Table 4.6 presents the lower bound of 8. Thus, we confirmed that the proven-equivalence partition established by Evans's Proposition, which was already presented in Figure 13 of [9], indeed gives the correct equivalence classes. The fact that this classification is complete is an original result. In Figure 5.1, we display representatives of each one of these equivalence classes.

The two classes that are not shown inequivalent by the graphical methods, only by Supports at 3 events, are the ones represented by the Instrumental Scenario (Figure 5.1(e)) and Evans's Scenario (Figure 5.1(g)).

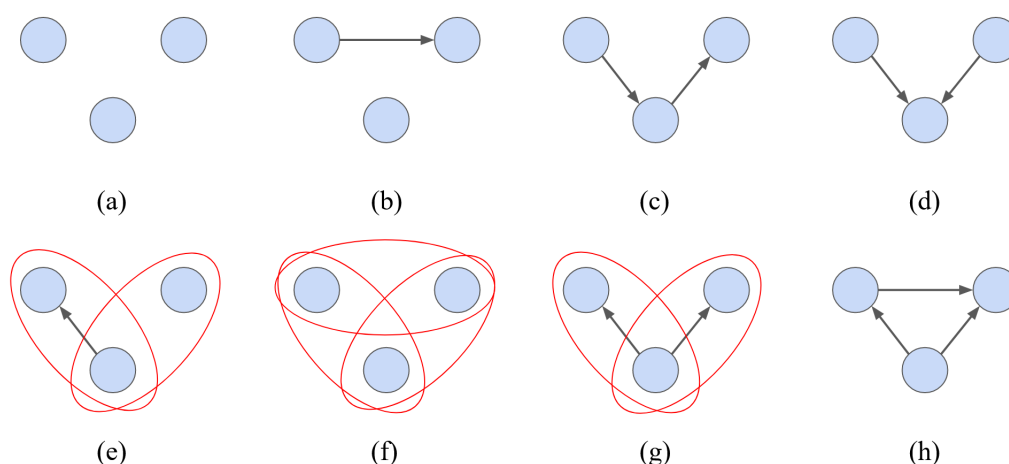


Figure 5.1: Representatives of the 8 equivalence classes of graph patterns with 3 nodes. The mDAGs (e), (f) and (g) have inequality constraints. (e) is called the Instrumental Scenario, (f) is the Triangle Scenario and (g) is the Unrelated Confounders Scenario.

If one looks only at the 3 visible variables case, they may think that HLP and

Equivalence Results	N° of Pink Bubbles
HLP	1597
HLP + Weak FS	416
HLP + Moderate FS	408
HLP + Strong FS	396

Table 5.1: Number of elements of the proven-equivalence partitions for successive application of equivalence results on the case of 4 visible variables. All these results also include the dominance relations of Section 2.4. These numbers give upper bounds on the number of equivalence classes.

Weak Face Splitting, the results that already existed, are enough to completely solve the classification of mDAGs. However, the case of 4 visible variables shows that this is not true: Table 5.1 summarizes the results showing that the upper bound gets smaller upon successive application of equivalence results. Thus, our two extensions indeed show some equivalences that the pre-existent results cannot show.

As we can see from Tables 3.4 and 4.6, for the case of 4 visible variables, the problem is not solved. The best upper bound on the number of equivalence classes that we have is 396, while the best lower bound is 333.

One could argue that the reason why we did not solve the problem is that we did not apply all the available methods, since we only compared supports up to 7 events. However, Section 5.2 reveals that this is not the case: there, we will study one specific element of the proven-inequivalence partition (gray bubble of Figure 1.2), and show that it is not solved even after doing every possible support comparison.

## 5.1 Completely Solved Classes

In the cases where an element of the proven-equivalence partition is also an element of the proven-inequivalence partition (a pink bubble coincides with a gray bubble in Figure 1.2), we get an equivalence class (green dashed line in Figure 1.2). In that case, we say that the class is "solved". If we successively apply the comparisons of Section 4, we can potentially solve more and more classes.

For the case of 3 visible nodes, as discussed, all the classes are solved when we apply the comparison of Supports up to 3 events. For the case of 4 visible nodes this is not true for all the classes, but it is true for some of them. Table 5.2 presents the number of such classes solved by each comparison.

Comparisons Made	N° of Solved Classes
CI	18
e-separation	19
OH + e-separation	23
OH + e-separation + DC	26
OH + e-separation + DC + Supps up to 2 events	39
OH + e-separation + DC + Supps up to 3 events	85
OH + e-separation + DC + Supps up to 4 events	219
OH + e-separation + DC + Supps up to 5 events	259
OH + e-separation + DC + Supps up to 6 events	302
OH + e-separation + DC + Supps up to 7 events	304

Table 5.2: Number of solved classes (subsets of graph patterns that are both elements of the respective proven-inequivalence partition and of the HLP + Strong FS proven-equivalence partition) under different comparisons.

In particular, here we reproduced one minor result of [7]: we found that the equivalence class of the Bell causal structure indeed has the 9 mDAGs shown in Figure 24 of [7]. The Bell equivalence class is solved by CI alone.

One could ask for example how many equivalence classes we can get right using only the comparison of CI Relations. If we only identify the CI Relations of our data (and not the inequality constraints), we can correctly guess the equivalence class of causal structures only if the CI Relations correspond to those of classes that are solved by CI alone, like the Bell class. In Table 5.2 we see that this is not very common: only 18 classes are solved by CI alone. Thus, relying only on CI Relations is in general a bad strategy for causal investigation.

Remember that the number of completely solved classes is *not* the same as the lower bound on the number of equivalence classes. In the case of 4 visible variables, the better lower bound we have is 333 while the best number of solved classes is 304.

## 5.2 Classes Compatible with Every Support

There is one element of the proven-inequivalence partition that is particularly interesting, the one whose mDAGs have no infeasible supports and no CI Relations. Even comparing supports up to  $2^{|V|} = 16$  events, this CI+Supports class still contains 8 different elements of the proven-equivalence partition (pink bubbles of Figure 1.2). Therefore, our current theoretical results are not powerful enough to show if these are equivalent or inequivalent to each other.

An interesting property of these 8 classes is that 7 of them have only one graph pattern each. The last one is the saturated class, which contains 1117 graph patterns. This is represented in the diagram of Figure 5.2.

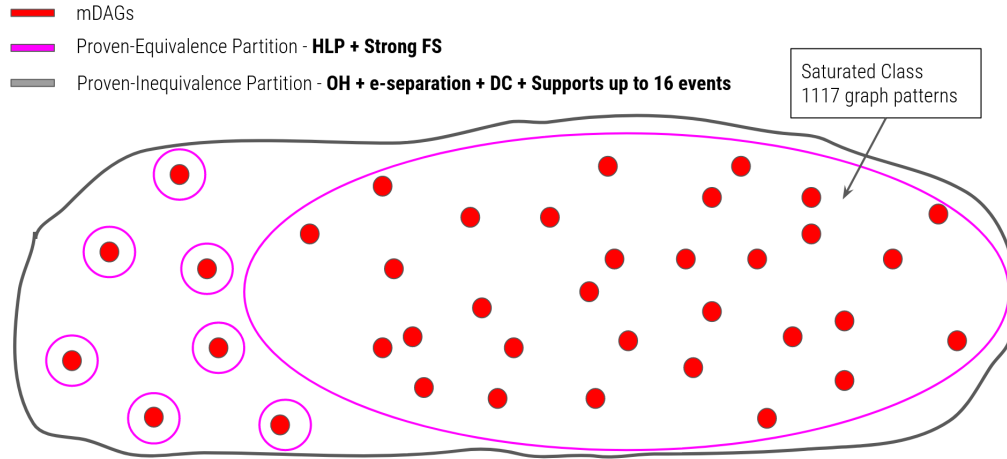


Figure 5.2: Element of the proven-inequivalence partition (gray bubble) corresponding to the mDAGs compatible with every support. It contains 8 elements of the proven-equivalence partition (pink bubbles), whose representatives are shown in Figure 5.3.

In Figure 5.3, we present representatives of each element of the proven-equivalence partition inside of this element of the proven-inequivalence partition.

There are two possibilities. The first one is that the mDAGs 5.3(a)-5.3(g) are all saturated (all the pink bubbles of Figure 5.2 ultimately merge), and these are examples of equivalent mDAGs whose equivalence cannot be shown by the known propositions. In this case, we would potentially be guided to new rules for proving equivalence. The second option is that at least one of the mDAGs 5.3(a)-5.3(g) is not saturated (the gray bubble of Figure 5.2 ultimately splits). This would prove that possibilistic compatibility does not imply probabilistic compatibility. Either possibility would be insightful if confirmed.

We still do not know which is the correct option, but it is worth noting that, by Proposition 3, the mDAG 5.3(a) is dominated by 5.3(b), 5.3(e), 5.3(f) and 5.3(g), while the mDAG 5.3(c) is dominated by 5.3(d), 5.3(e), 5.3(f) and 5.3(g). Therefore, we just need to check if 5.3(a) and 5.3(c) are saturated.

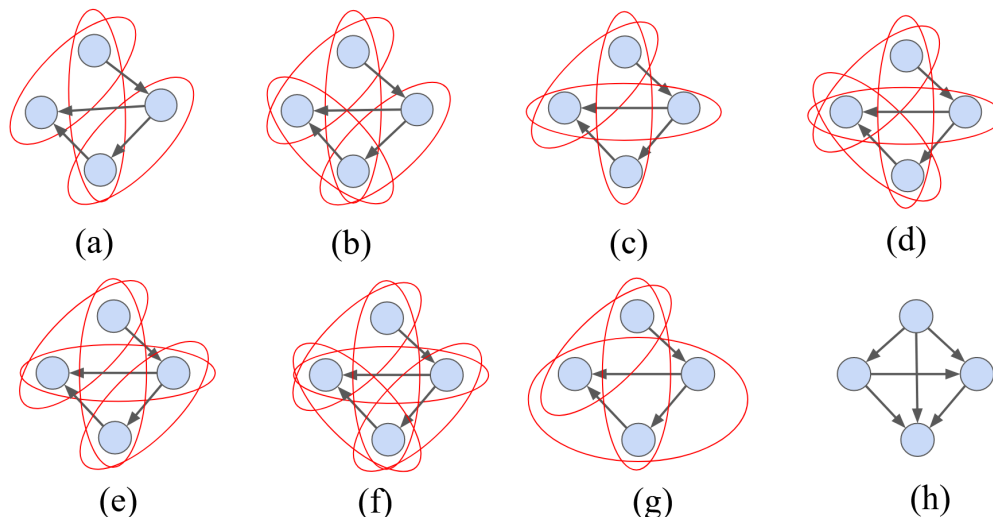


Figure 5.3: Representatives of the 8 pink bubbles whose mDAGs are compatible with any Support and do not have any CI relation. The classes (a)-(g) have only one graph pattern each. The class (h) is the saturated class, that has 1117 graph patterns.

#### Summary of the chapter:

- For 3 visible nodes, we find that there are 8 equivalence classes, whose representatives are given in Figure 5.1. This proven-equivalence partition was already presented in [9], but the fact that it coincides with the proven-inequivalence partition is an original result.
- For 4 visible nodes, the number of equivalence classes is between 333 and 396. The methods that we currently have are not sufficient to find the exact equivalence classes.
- There are some sets of mDAGs that we can already recognize as equivalence classes; we know that they will not split or merge with any other set in the final classification. We call these "solved" classes. Using all the known methods, we found 304 solved classes in the case of 4 visible variables. Using only comparison of CI Relations we find only 18 solved classes, what highlights that it is a relatively weak method.
- The mDAGs (a)-(g) of Figure 5.3 cannot be shown equivalent nor inequivalent to the saturated class by any of our methods. If one can show that they are saturated (or non-saturated), this can lead to hints for new methods to prove equivalence (or inequivalence).



# Chapter 6

## Quantum Causal Structures

In Figure 1.1 we saw an example of a causal structure that admits a Quantum-Classical Gap: the classical Bell causal structure gives an inequality constraint that can be violated if we consider its quantum version. In this section we will define a quantum causal structure and investigate whether the observational equivalence results presented in Sections 3 and 4 are still valid for quantum causal structures. Then, we will go back to our case studies of 3 and 4 visible variables and search for new examples of QC Gaps.

Following [13], here we define a quantum causal structure by a latent-variable DAG whose latent variables are not associated with classical random variables anymore, but now to quantum systems. All the visible variables are still associated with random variables.

The outgoing edges of a classical node are copies of the node variable, just like in the case of classical causal structures. The outgoing edges of a quantum node, on the other hand, are subsystems of the quantum system associated with the node. For example, if the Hilbert space  $\mathcal{H}_A \otimes \mathcal{H}_B$  is assigned to a latent node  $l$  that has two children, then  $l$  can send the subsystem  $\mathcal{H}_A$  to one of its children and  $\mathcal{H}_B$  to the other. See Figure 6.1.

In a DAG  $\mathcal{G}$ , we are going to denote the incoming classical edges (that come from visible parents) of a node  $x$  by  $ed_{\mathcal{G}}^{(c)}(x)$ , and the incoming quantum edges (that come from latent parents) by  $ed_{\mathcal{G}}^{(q)}(x)$ . Similarly, the set of all parents of node  $x$  can be divided into visible parents  $pa_{\mathcal{G}}^{(c)}(x)$  and latent parents  $pa_{\mathcal{G}}^{(q)}(x)$ .

To discuss the quantum version of the causal compatibility problem, we need to know how the nodes respond to their parents. While in the classical case we had the measurable functions of the SEP (Definition 3), here we define the *quantum realization* of a DAG:

**Definition 18** (Quantum Realization of  $\mathcal{G}$ ). *Let  $\mathcal{G}$  be a latent-variable DAG, with  $V$  being its set of visible nodes and  $L$  being its set of latent nodes.*

*The Quantum Realization of the DAG  $\mathcal{G}$  is a specification of the following quantum*

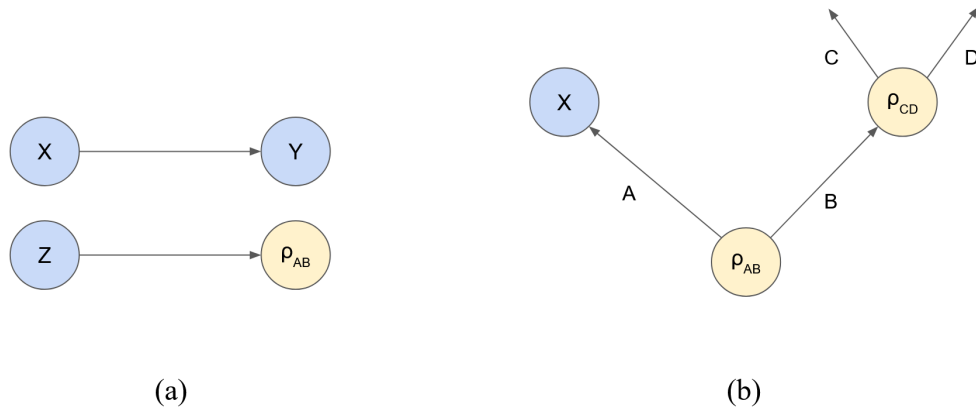


Figure 6.1: (a) Outgoing edges from visible (classical) nodes are copies of the node's random variable. (b) Outgoing edges from latent (quantum) nodes are subsystems of the node's quantum system. Here we will represent quantum latent nodes in yellow, while in the previous sections we represented classical latent nodes in red.

operations:

- With each latent node  $l \in L$ , we associate a density matrix (preparation)

$$\begin{aligned} \rho_l &\in \mathcal{B}(\mathcal{H}_l) \\ \rho_l &\text{ positive semi-definite} \\ \text{Tr}(\rho_l) &= 1 \end{aligned}$$

such that  $\mathcal{H}_l$  is the Hilbert space associated with  $l$ . Each outgoing edge of  $l$  carries a subsystem of  $\mathcal{H}_l$ .

- Let  $l \in L$  be a latent node. It responds to its incoming edges with a quantum channel (transformation)

$$\Omega_{ed_G^{(c)}(l)} : \mathcal{B}\left(\mathcal{H}_{ed_G^{(q)}(l)}\right) \rightarrow \mathcal{B}(\mathcal{H}_l)$$

Thus, the classical incoming edges of  $l$  determine what is the transformation that takes the density matrix of the quantum incoming edges of  $l$  and gives the density matrix of  $l$ . See an example in Figure 6.2(a).

If  $l$  does not have latent parents, it is given by a source of quantum states followed by a transformation determined by the visible parents of  $l$ .

- Let  $v \in V$  be a visible node. It responds to its incoming edges with a POVM

(measurement)

$$E_{v|ed_G^{(c)}(v)} \in \mathcal{B}\left(\mathcal{H}_{ed_G^{(Q)}(v)}\right)$$

$$E_{v|ed_G^{(c)}(v)} \text{ positive semi-definite}$$

$$\sum_v E_{v|ed_G^{(c)}(v)} = \mathbb{I}_{\mathcal{H}_{ed_G^{(Q)}(v)}}$$

Thus, the classical incoming edges of  $v$  determine what is the measurement that takes the density matrix of the quantum incoming edges of  $v$  and gives the classical random variable associated with  $v$ . See an example in Figure 6.2(b).

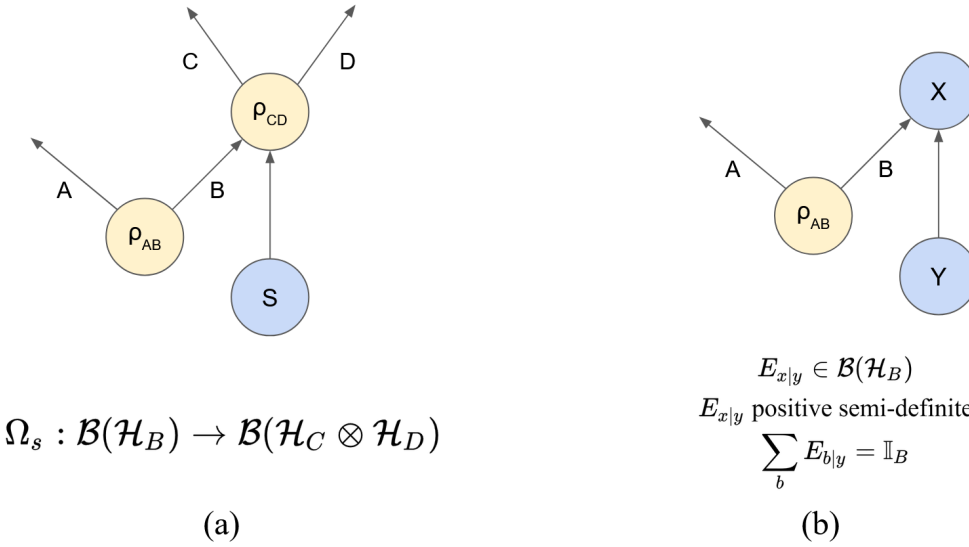


Figure 6.2: Examples of (a) how latent (quantum) nodes respond to its incoming edges and (b) how visible (classical) nodes respond to its incoming edges.

Now we can define the quantum version of the causal compatibility problem, that makes use of the Born Rule for probabilities of quantum measurements. The definition can be better understood following the example of Figure 6.3.

**Definition 19** (Quantum Causal Compatibility Problem). *Let  $\mathcal{G}$  be a latent-variable DAG, with  $V$  being its set of visible nodes and  $L$  being its set of latent nodes. Let  $V^{(no-l)} \subseteq V$  be the set of visible nodes that do not have any latent parent, and  $V^{(l)} = V \setminus V^{(no-l)}$  be the set of visible nodes that have some latent parents.*

*Say that  $P_V$  is some joint probability distribution over the random variables  $X_v$ ,  $v \in V$ . We say that  $P_V$  is quantum-compatible with  $\mathcal{G}$  if there exists a quantum realization of  $\mathcal{G}$*

such that

$$P_V(v) = \left[ \prod_{A \in V^{(n_0-l)}} P_A(a|pa_{\mathcal{G}}(A)) \right] \text{Tr} \left[ \left( \bigotimes_{B \in V^{(l)}} E_{b|pa_{\mathcal{G}}^{(c)}(B)} \right) \rho_S \right];$$

$$\rho_S \in \mathcal{B}(\mathcal{H}_S); \mathcal{H}_S = \bigotimes_{B \in V^{(l)}} \left[ \mathcal{H}_{ed_{\mathcal{G}}^{(Q)}(B)} \right]; E_{b|pa_{\mathcal{G}}^{(c)}(B)} \in \mathcal{B} \left( \mathcal{H}_{ed_{\mathcal{G}}^{(Q)}(B)} \right)$$

where  $v$  is an evaluation of all the visibles that corresponds to the evaluations  $a$  for  $A \in V^{(n_0-l)}$  and  $b$  for  $B \in V^{(l)}$  that appear in the right-hand side.

We denote by  $\mathcal{M}_Q(\mathcal{G}, V)$  the set of probability distributions that are quantum-compatible with the DAG  $\mathcal{G}$ .

The density matrix  $\rho_S$  that appears in Definition 19 corresponds to the quantum states that are sent as edges to the visible nodes. Therefore, it is an element of  $\mathcal{B}(\mathcal{H}_S)$ , as defined above. The quantum states that compose the subsystems of  $\mathcal{H}_S$  can come from a source of quantum states (if the latent node in question does not have latent parents) or can be the outcome of a quantum channel (if the latent node has latent parents). See Figure 6.3 for clarification.

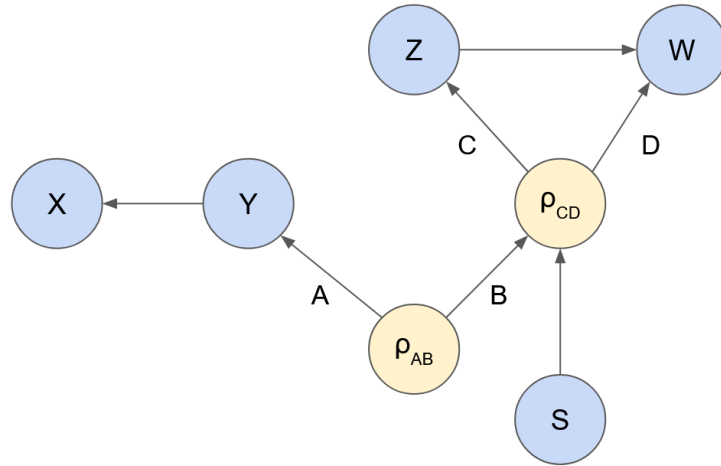


Figure 6.3: A probability distribution  $P_{X,Y,Z,W,S}$  is quantum-compatible with this DAG if there exists a quantum realization composed of the quantum state  $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$ , the quantum channel  $\Omega_S : \mathcal{B}(\mathcal{H}_B) \rightarrow \mathcal{B}(\mathcal{H}_C \otimes \mathcal{H}_D)$  and the POVMs  $E_y \in \mathcal{B}(\mathcal{H}_A)$ ,  $F_z \in \mathcal{B}(\mathcal{H}_C)$  and  $G_{w|z} \in \mathcal{B}(\mathcal{H}_D)$  such that  $P_{X,Y,Z,W,S}(x,y,z,w,s) = P_S(s)P_{X|Y}(x|y)\text{Tr}[(\mathbb{I}_A \otimes \Omega_s)\rho_{AB}(E_y \otimes F_z \otimes G_{w|z})]$ .

## 6.1 Quantum Validity of the Rules for Proving Equivalence and Inequivalence

Two latent-variable DAGs that are observationally equivalent when treated as classical causal structures are not necessarily observationally equivalent when treated as quantum causal structures. A pertinent question that arises is: which tools can we still use if we want to find the classification of quantum DAGs into Observational Equivalence classes?

Some of the theoretical results presented in Sections 2.4, 3 and 4 are proved assuming something about the classical nature of the nodes, for example using the measurable functions of the SEP. However, there is still the possibility that the proofs we have so far are too restrictive, and the results are still valid (but require more general proof). If that is not the case, one must be able to find a counter-example.

### 6.1.1 Ways to Prove Equivalence

In [13], it is shown that there is no quantum version for the exogenization rule (Lemma 4): an explicit counterexample is presented for the quantum case. This counterexample is reproduced in Appendix E. Therefore, it does not make sense to use the concept of mDAG to describe quantum causal structures.

Proposition 1, on the other hand, does have a quantum version:

**Proposition 11** (Edge Dropping - Quantum Version). *Let  $\mathcal{D}$  and  $\mathcal{D}'$  be two latent-variable DAGs with the same sets of vertices. Let  $V$  be their set of visible variables and  $L$  be their set of latent variables. If  $\mathcal{D}'$  is a subgraph of  $\mathcal{D}$  (as in Definition 2), then  $\mathcal{M}_{\mathcal{Q}}(\mathcal{D}', V) \subseteq \mathcal{M}_{\mathcal{Q}}(\mathcal{D}, V)$ .*

*Proof.* Let  $u \in V \cup L$  be a node of  $\mathcal{D}$  and  $\mathcal{D}'$ . Since  $\mathcal{D}'$  is a subgraph of  $\mathcal{D}$ , we have that  $\text{pa}_{\mathcal{D}'}(u) \subseteq \text{pa}_{\mathcal{D}}(u)$ , for both visible and latent parents of  $u$ . Say that  $\text{pa}_{\mathcal{D}}^{(\mathcal{C})^*}(u)$  are the extra visible parents of  $u$  that are in  $\mathcal{D}$  but not in  $\mathcal{D}'$ , and  $\text{pa}_{\mathcal{D}}^{(\mathcal{Q})^*}(u)$  are the extra latent parents of  $u$  that are in  $\mathcal{D}$  but not in  $\mathcal{D}'$ .

Consider a probability distribution quantum-compatible with  $\mathcal{D}'$ . This means that this probability distribution can be obtained from a quantum realization of  $\mathcal{D}'$  that involves:

$$\begin{aligned} \text{POVMs } E_{v|pa_{\mathcal{D}'}^{(c)}(v)} &\in \mathcal{B}\left(\mathcal{H}_{ed_{\mathcal{D}'}^{(Q)}(v)}\right), v \in V \\ \text{Transformations } \Omega_{pa_{\mathcal{D}'}^{(c)}(l)} &: \mathcal{B}\left(\mathcal{H}_{ed_{\mathcal{D}'}^{(Q)}(l)}\right) \rightarrow \mathcal{B}(\mathcal{H}_l), l \in L \end{aligned}$$

The probability distributions quantum-compatible with  $\mathcal{D}$ , on the other hand, come from quantum realizations of  $D$  that involve:

$$\begin{aligned} \text{POVMs } E_{v|(pa_{\mathcal{D}'}^{(c)}(v), pa_{\mathcal{D}}^{(c)*}(v))} &\in \mathcal{B}\left(\mathcal{H}_{ed_{\mathcal{D}'}^{(Q)}(v)} \otimes \mathcal{H}_{ed_{\mathcal{D}}^{(Q)*}(v)}\right), v \in V \\ \text{Transformations } \Omega_{(pa_{\mathcal{D}'}^{(c)}(l), pa_{\mathcal{D}}^{(c)*}(l))} &: \mathcal{B}\left(\mathcal{H}_{ed_{\mathcal{D}'}^{(Q)}(l)} \otimes \mathcal{H}_{ed_{\mathcal{D}}^{(Q)*}(l)}\right) \rightarrow \mathcal{B}(\mathcal{H}_l), l \in L \end{aligned}$$

We can always choose such a quantum realization of  $\mathcal{D}$  where every

$$E_{v|(pa_{\mathcal{D}'}^{(c)}(v), pa_{\mathcal{D}}^{(c)*}(v))}$$

does not depend on  $pa_{\mathcal{D}}^{(c)*}(v)$  and acts trivially on  $\mathcal{H}_{ed_{\mathcal{D}}^{(Q)*}(v)}$ , and every

$$\Omega_{(pa_{\mathcal{D}'}^{(c)}(l), pa_{\mathcal{D}}^{(c)*}(l))}$$

does not depend on  $pa_{\mathcal{D}}^{(c)*}(l)$  and maps operators on  $\mathcal{H}_{ed_{\mathcal{D}'}^{(Q)}(l)}$  to operators on  $\mathcal{H}_l$ , ignoring what is the operator on  $\mathcal{H}_{ed_{\mathcal{D}}^{(Q)*}(l)}$ . Therefore, we can reproduce exactly the same quantum realization of  $\mathcal{D}'$  described above, showing that every probability distribution that is quantum-compatible with  $\mathcal{D}'$  is also quantum-compatible with  $\mathcal{D}$ .  $\square$

If we admit arbitrary cardinalities for the latent nodes, Lemma 5 also has a quantum version:

**Lemma 3** (Remove Redundant Latent Variables - Quantum Version). *Let  $\mathcal{G}$  be a DAG with vertices  $V \cup \{u, w\}$ , where  $pa_{\mathcal{G}}(w) = pa_{\mathcal{G}}(u) = \emptyset$ . Let  $\mathcal{G}_{-w}$  be the DAG constructed by removing  $w$  from  $\mathcal{G}$ . If  $ch_{\mathcal{G}}(w) \subseteq ch_{\mathcal{G}}(u)$  and we admit any cardinality for the latent nodes, then  $\mathcal{M}_{\mathcal{Q}}(\mathcal{G}, V) = \mathcal{M}_{\mathcal{Q}}(\mathcal{G}_{-w}, V)$ .*

*Proof.* By Proposition 11, it is clear that  $\mathcal{G}$  dominates  $\mathcal{G}_{-w}$ . Now, suppose that a certain distribution over the visible nodes is obtained with a quantum realization

of  $\mathcal{G}$  where  $u$  is associated with  $\rho_u \in \mathcal{H}_u$  and  $w$  is associated with  $\rho_w \in \mathcal{H}_w$  (since  $u$  and  $w$  do not have parents, these density matrices come directly from a source of quantum states).

Since we admit latent nodes of any cardinality, we can have a quantum realization of  $\mathcal{G}_{-w}$  where  $u$  is associated with  $\rho_u \otimes \rho_w \in \mathcal{H}_u \otimes \mathcal{H}_w$ . Therefore,  $u$  in  $\mathcal{G}_{-w}$  can send to its children exactly the same states that  $u$  and  $w$  together send to their children in  $\mathcal{G}$  (see Figure 6.4). Furthermore, nothing else changes from  $\mathcal{G}$  to  $\mathcal{G}_{-w}$ , because nodes  $u$  and  $w$  have no parents. Hence,  $\mathcal{G}_{-w}$  dominates  $\mathcal{G}$ , showing that they are observationally equivalent.  $\square$

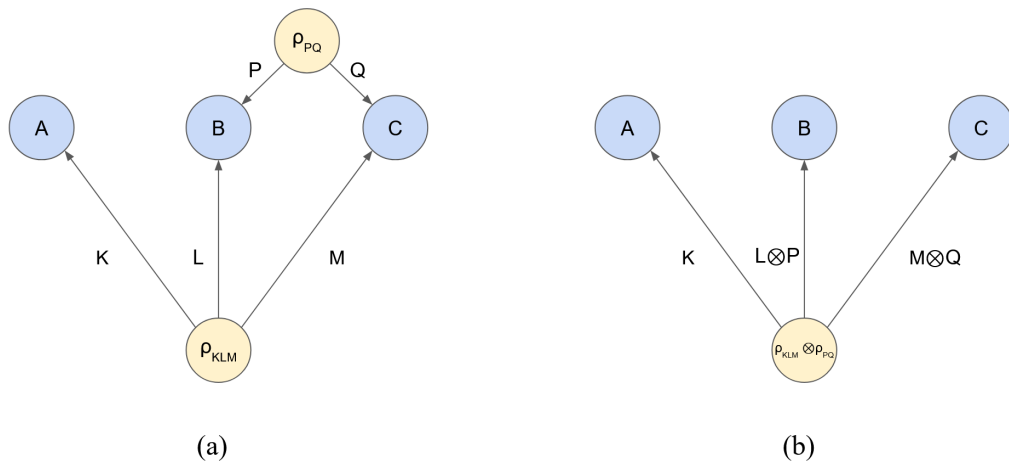


Figure 6.4: When we admit any cardinality of the latent nodes, the redundant ones can be absorbed by enlarging the Hilbert space associated with the node (Lemma 3). Thus, DAGs (a) and (b) in this Figure are quantum observationally equivalent.

We can also ask about the validity of the other equivalence rules, HLP and Face Splitting, when translated from mDAGs to their associated DAGs. We still do not have a definite answer to whether these have quantum versions or not.

We tried to find counter-examples to these results inspired by the counter-example of [13] reproduced in Appendix E. To find such counter-examples it is clear that we need causal structures that present QC gaps, such as the Bell-like structure of Figure E.1(a). More specifically, we need a DAG that can quantumly violate a classical inequality, but whose resulting DAG after the application of the proposition cannot.

However, both HLP and Evans require the parents of one node (or one set of nodes) to be a subset of the parents of another node. In Bell scenarios, the classical nodes  $A$  and  $B$  that give the outcomes of Alice's and Bob's measurements need to have at least one parent each that is not a parent of the other, to play the role of

the measurement settings  $S$  and  $T$ . Thus, there is no Bell-like structure where we can apply HLP or Evans to the nodes that are relevant for the QC Gap.

If we try to find counter-examples based not on Bell but on the Instrumental and the Triangle Scenarios (Figure 5.1(e) and 5.1(f)), that are also known to present QC Gaps, we encounter the same problem: we cannot find an Instrumental-like or Triangle-like DAG whose QC Gap is closed by applying HLP or Evans, because both of these structures require the relevant nodes to have (latent or visible) parents that are not shared with the other relevant nodes.

This is not definite proof, because there could still be another type of structure that yields a QC Gap whose gap is closed under the application of HLP or Evans. However, the fact that we cannot find such a counter-example looking at Bell, Instrumental, and Triangle scenarios is evidence in favor of the conjecture that HLP and Evans are valid in the quantum case because they are not even applicable for the portions of the structures that are responsible for the QC Gaps.

### 6.1.2 Comparisons that Show Inequivalence

As well as asking about the quantum validity of our methods to show equivalence, we can ask about our methods to show inequivalence.

The proof that different e-separation relations lead to inequivalence still holds in the quantum case, because entropic inequalities (such as the one involved in Proposition 12 of [15]) are quantum valid. As special cases, comparisons of skeletons and CI relations also show inequivalence in the quantum case.

Similarly, we can formulate a quantum version of Proposition 10:

**Proposition 12** (Only Hypergraphs Rule - Quantum Version). *Let  $\mathcal{G}$  and  $\mathcal{G}'$  be two latent-variable DAGs with visible nodes  $V$  that do not have any edges between the visible nodes. If  $\mathcal{G} \neq \mathcal{G}'$ , then  $\mathcal{M}_{\mathcal{Q}}(\mathcal{G}, V) \neq \mathcal{M}_{\mathcal{Q}}(\mathcal{G}', V)$ .*

*Proof.* The proof presented in Appendix C is quantum (and GPT) valid. □

Finally, it is evident that the technique we used to find feasible Supports does not apply to quantum causal structures, because it consists in enumerating all the possible deterministic response functions that a node can have to the valuations of its parents (Appendix D).

The quantum validity of the comparison of pairs of densely connected nodes as an inequivalence test is left for future research.



## 6.2 Searching for QC Gaps

The classification of mDAGs into observational equivalence classes can help us filter which structures may yield a QC Gap.

Since we still did not study the equivalence classes of *quantum* causal structures, there still exists the possibility that a structure  $\mathcal{G}_1$  does not present a QC Gap, but another structure  $\mathcal{G}_2$  that is classically equivalent to it does. This could happen if  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are observationally equivalent when treated as classical causal structures, but their quantum versions are inequivalent.

However, an instance where we *can* infer that none of the structures in a given classical equivalence class have a QC Gap is when inside the class there is a latent-free structure  $\mathcal{G}$ : it does not have inequality constraints, that are necessary for QC Gaps. Therefore, all structures that are dominated by  $\mathcal{G}$  also do not have inequality constraints nor QC Gaps. This observation is sometimes called the HLP Criterion, since it was first noted in [8].

For the case of 3 visible nodes, this consideration shows that only the mDAGs equivalent to 5.1(e), 5.1(f) and 5.1(g) may have a QC Gap. In fact, we know that 5.1(e) and 5.1(f) do have QC Gaps, while 5.1(g) is still an open problem.

Let us now present a concept that reduces the number of mDAGs that we need to check for QC Gaps.

### 6.2.1 Fundamental mDAGs

In Appendix F, there is a discussion about a result from [19] that teaches us a way of factorizing mDAGs into subgraphs. This is relevant because, if an mDAG presents a QC gap, then at least one of its factors presents the QC gap when looked at as an mDAG by itself. Therefore, to find new QC gaps it suffices to look at mDAGs that have only one of such factors, that we will call *fundamental*:

**Definition 20** (Fundamental mDAG). *Let  $\mathcal{G}$  be an mDAG. We say that  $\mathcal{G}$  is fundamental if it contains only one common-cause district  $D$ , and every node  $v \notin D$  points into  $D$ .*

Table 6.1 shows how many of the elements of the proven-inequivalence partition obtained by HLP+Strong Face Splitting are completely composed of fundamental mDAGs. These are the only ones that we need to check when searching for new QC Gaps.

Visible Nodes	3	4
All elements of partition	8	396
Elements that only have fundamental mDAGs	3	331

Table 6.1: Numbers of elements of the HLP+Strong FS proven-equivalence partition (pink bubbles of Figure 1.2) that are composed only of fundamental mDAGs.

Note that the HLP Criterion is already embraced when we ignore the classes that are not 100% fundamental; mDAGs that do not have any facet with more than one node (latent-free) also do not have any common-cause district. Indeed, the three classes of 3 visible nodes that are 100% fundamental correspond exactly to the classes of 5.1(e), 5.1(f) and 5.1(g).

Furthermore, the criterion of considering only 100% fundamental classes is strictly stronger than the HLP Criterion. Out of the 396 elements of the proven-equivalence partition for 4 visible nodes, only 20 include latent-free structures. On the other hand, 65 of them include non-fundamental structures, thus leaving us with 331 elements of the HLP+Strong FS proven-equivalence partition that may yield QC gaps.

This is still a quite big number of classes, so it is a hard task to single out the promising candidates for a QC Gap. Another possibility is that QC Gaps are very common, and many of these 331 classes have them.

Since at the moment we do not have any better criterion to filtrate QC Gaps, it is worth studying what happens when we apply the fundamentality criterion to some interesting subsets of the proven-equivalence partition. For example, when we apply this criterion to the 18 classes that are solved by CI alone (see table 5.2) we obtain only 4 equivalence classes, whose representatives are displayed in Figure 6.5.

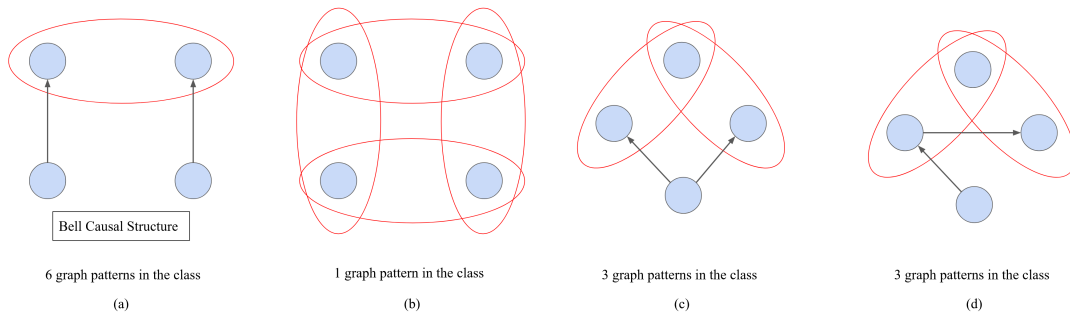


Figure 6.5: Representatives of the 4 elements of the HLP+Strong FS proven-equivalence partition that are 100% fundamental and solved by CI.

The first mDAG of Figure 6.5 is the Bell causal structure. The other three are also famous examples: 6.5(b) is the square scenario, studied in [20] and [21]; 6.5(c) and 6.5(d) are studied in [8] and [22] (Figures 15 and 17 of [22]). All of them yield inequality constraints, and 6.5(b) is known to have a QC Gap [21].

The fact that the Bell causal structure is solved by CI alone is interesting information for model selection because it says that this is indeed the only classical causal structure that guarantees the hypothesis of local causality and no superdeterminism. This same information may also help in the design of new scenarios related to the other three mDAGs of Figure 6.5.

**Summary of the chapter:**

- When we take the latent variables of a DAG to be quantum systems instead of classical random variables, some of the inequality constraints might not hold anymore. This constitutes a *Quantum-Classical Gap*. The first and more famous example of a Quantum-Classical Gap is Bell Scenario.
- In the quantum case, Lemma 4 is not valid. Thus, in general, we work with quantum DAGs instead of mDAGs.
- We know that the Edge Dropping Proposition, that gives conditions under which one DAG dominates another, is also valid in the quantum case.
- It is still unknown whether our methods to prove equivalence (presented in Chapter 3) are valid in the quantum case or not. However we have reasons to suspect that, in the DAGs where these methods are applicable, the relevant part of the DAG does not even yield a QC Gap in the first place.
- The comparison of e-separation relations and the Only Hypergraphs rule are also quantum valid. The comparison of Infeasible Supports is not quantum valid.
- When searching for QC Gaps, we only need to look only at the classes that are 100% composed of fundamental mDAGs (Definition 20). For 3 visible nodes, this filters the 3 equivalence classes that correspond to the Instrumental, Triangle and Evans Scenarios. For 4 visible nodes, this filters 331 out of the 396 elements of the proven-equivalence partition.

# Chapter 7

## Conclusion

The problem of causal discovery is interesting both from the point of view of classical statistics and from the point of view of foundations of quantum theory. Classifying the classical causal structures into Observational equivalence classes is important because, at the same time as it lowers the number of possible causal explanations to be considered for a set of data, it also helps filtering which ones might provide Quantum-Classical gaps. For example, if a causal structure is known to be observationally equivalent to another one that has no latent nodes, then by [8] we know that it cannot present non-trivial inequalities (thus it cannot have a QC Gap).

The aim of this work was to recapitulate all the currently known techniques that prove Observational equivalence or inequivalence between causal structures, extending the scope of these techniques when possible and applying them to causal structures with 3 and 4 visible variables.

Using techniques that already existed, we were able to show that the classification of unlabelled causal structures (graph patterns) with 3 visible variables into 8 equivalence classes is complete. Representatives of these classes can be found in Figure 5.1. These representatives were previously shown in Figure 13 of [9], but the fact that this classification is complete (none of these equivalence classes will merge) is an original result.

For the case of 4 visible variables, apart from the techniques that already existed we also used our two new techniques (Propositions 7 and 8), that are extensions of Proposition 6.8 of [9]. We obtained that the total number of equivalence classes for 4 visible variables is between 333 and 396. Furthermore, we found out that 304 of the classes are already "solved": we know that they will not be merged with any other.

The case of 4 visible variables makes it explicit that the classification problem is still not fully resolved; there are still causal structures whose equivalence we currently cannot either prove nor disprove. Even if we check all possible infeasible supports (up to  $2^{|V|}$  events), thus applying the support analysis to its maximum

power, the causal structures of Figure 5.3 cannot be proven to be inequivalent. At the same time, none of our current equivalence results can show their equivalence. This is explained in Figure 5.2.

This is related to an open problem that we leave for future research: does possibilistic equivalence imply probabilistic equivalence? i.e. if two causal structures have the same set of infeasible supports, does this mean that they are observationally equivalent? The question of whether 5.3(a) and 5.3(c) are saturated might help with this. If one can show that they are not saturated, one proves that possibilistic equivalence does not imply probabilistic equivalence. On the other hand, if they are saturated, analyzing them would be potentially insightful for finding new equivalence rules that capture their equivalence with the saturated class.

Our criterion when searching for examples that may lead to Quantum-Classical gaps was to look at 100% fundamental classes. In the definition used here, this involves only structures that cannot be decomposed into smaller structures and that are not latent-free (thus including the HLP criterion from [8]). This gave us 331 potential equivalence classes, therefore it is still hard to filter a few examples of interest. We found that only four classes are 100% fundamental and solved by CI (Figure 6.5). These four are famous and known to yield inequality constraints; two of them are also known to have a QC Gap [6] [21].

All the classification we have done in this work refers solely to *classical* causal structures; for quantum causal structures even the concept of mDAG (from Section 2.3) cannot be used in general. However, some of the equivalence results used here are still quantumly valid, as discussed in Section 6.1. The question of whether Evans' Proposition, HLP Proposition and the comparison of densely connected pairs as a test of inequivalence are quantum valid or not is left for future research.

# Appendix A

## Evans's Rules

In section 3.2 of [9], one can find redundancy Lemmas that justify the definition of mDAGs. Here we reproduce these Lemmas, and Figure A.1 shows an example of their application.

The first one is Lemma 3.7 in [9]:

**Lemma 4** (Exogenize Latent Variables). *Let  $\mathcal{G}$  be a DAG with vertices  $V \cup \{u\}$ . Construct the exogenized DAG  $\tau(\mathcal{G}, u)$  by adding edges from every parent of  $u$  to every child of  $u$ , and deleting every edge from parents of  $u$  to  $u$ . Note that now  $u$  is a node without parents, so we call it “exogenous”.*

*We have that  $\mathcal{M}(\mathcal{G}, V) = \mathcal{M}(\tau(\mathcal{G}, u), V)$ .*

*Proof.* Consider a probability distribution that obeys the SEP for the graph  $\mathcal{G}$ , before the exogenization. By definition, under that distribution the random variable  $X_u$  is  $\sigma(X_{\text{pa}_{\mathcal{G}}(u)}, E_u)$ -measurable. This means that

$$X_u^{-1}(2^{\mathcal{X}_u}) \subseteq \sigma \left( X_{\text{pa}_{\mathcal{G}}(u)}^{-1}(2^{\mathcal{X}_{\text{pa}_{\mathcal{G}}(u)}}) \cup E_u^{-1} \left( 2^{\mathcal{E}_u} \right) \right) \quad (\text{A.1})$$

Now, for  $c \in \text{ch}_{\mathcal{G}}(u)$ , we have that  $X_c$  is  $\sigma(X_u, X_{\text{pa}_{\mathcal{G}}(c) \setminus u}, E_c)$ -measurable:

$$X_c^{-1}(2^{\mathcal{X}_c}) \subseteq \sigma \left( X_u^{-1} \left( 2^{\mathcal{X}_u} \right) \cup X_{\text{pa}_{\mathcal{G}}(c) \setminus u}^{-1} \left( 2^{\mathcal{X}_{\text{pa}_{\mathcal{G}}(c) \setminus u}} \right) \cup E_c^{-1} \left( 2^{\mathcal{E}_c} \right) \right) \quad (\text{A.2})$$

Furthermore, using Equation (A.1), we get:

$$\begin{aligned} & \sigma \left( X_u^{-1} \left( 2^{\mathcal{X}_u} \right) \cup X_{\text{pa}_{\mathcal{G}}(c) \setminus u}^{-1} \left( 2^{\mathcal{X}_{\text{pa}_{\mathcal{G}}(c) \setminus u}} \right) \cup E_c^{-1} \left( 2^{\mathcal{E}_c} \right) \right) \\ & \subseteq \sigma \left( X_{\text{pa}_{\mathcal{G}}(u)}^{-1} \left( 2^{\mathcal{X}_{\text{pa}_{\mathcal{G}}(u)}} \right) \cup E_u^{-1} \left( 2^{\mathcal{E}_u} \right) \cup X_{\text{pa}_{\mathcal{G}}(c) \setminus u}^{-1} \left( 2^{\mathcal{X}_{\text{pa}_{\mathcal{G}}(c) \setminus u}} \right) \cup E_c^{-1} \left( 2^{\mathcal{E}_c} \right) \right) \end{aligned} \quad (\text{A.3})$$

From Equations (A.2) and (A.3), we conclude that the random variable  $X_c$  is  $\sigma(X_{\text{pa}_{\mathcal{G}}(u)}, E_u, X_{\text{pa}_{\mathcal{G}}(c) \setminus u}, E_c)$  measurable.

This shows that the probability distribution of the non-exogenized graph  $\mathcal{G}$  is also compatible with the exogenized graph  $\tilde{\mathcal{G}}$ . This is so because in the exogenized graph  $\tilde{\mathcal{G}}$ , the random variable of  $u$  is now simply given by  $\tilde{X}_u = E_u$ , since it has no parents. So,  $X_c$  is  $\sigma$ - $(X_{\text{pa}_{\tilde{\mathcal{G}}}(u)}, \tilde{X}_u, X_{\text{pa}_{\tilde{\mathcal{G}}}(c) \setminus u}, E_c)$  measurable, what shows that it obeys the SEP for  $\tilde{\mathcal{G}}$ . So,  $\mathcal{M}(\mathcal{G}, V) \subseteq \mathcal{M}(\tilde{\mathcal{G}}, V)$ .

Now, we will prove the converse. Start with a probability distribution that is compatible with  $\tilde{\mathcal{G}}$ . We have that  $X_c$  is  $\sigma$ - $(X_{\text{pa}_{\tilde{\mathcal{G}}}(u)}, \tilde{X}_u, X_{\text{pa}_{\tilde{\mathcal{G}}}(c) \setminus u}, E_c)$  measurable, for  $c \in \text{ch}_{\tilde{\mathcal{G}}}(u)$ . If we define  $X_u \equiv (X_{\text{pa}_{\tilde{\mathcal{G}}}(u)}, \tilde{X}_u)$ , we get that  $X_c$  is  $\sigma$ - $(X_u, X_{\text{pa}_{\tilde{\mathcal{G}}}(c) \setminus u}, E_c)$  measurable. This shows that our probability distribution obeys the SEP for a graph where  $c$  is a child of  $u$  and  $\text{pa}_{\tilde{\mathcal{G}}}(c) \setminus u$ , and  $u$  is a child of  $\text{pa}_{\tilde{\mathcal{G}}}(u)$ . Such graph is precisely  $\mathcal{G}$ , so  $\mathcal{M}(\tilde{\mathcal{G}}, V) \subseteq \mathcal{M}(\mathcal{G}, V)$ . Note that we defined the error function of  $u$  as  $E_u \equiv \tilde{X}_u$ . □

Thus, we can exogenize every latent node of a DAG keeping the same marginal distributions over the visible variables. The second relevant result is Lemma 3.8 of [9]:

**Lemma 5 (Remove Redundant Latent Variables).** *Let  $\mathcal{G}$  be a DAG with vertices  $V \cup \{u, w\}$ , where  $\text{pa}_{\mathcal{G}}(w) = \text{pa}_{\mathcal{G}}(u) = \emptyset$ . Let  $\mathcal{G}_{-w}$  be the DAG constructed by removing  $w$  from  $\mathcal{G}$ . If  $\text{ch}_{\mathcal{G}}(w) \subseteq \text{ch}_{\mathcal{G}}(u)$ , then  $\mathcal{M}(\mathcal{G}, V) = \mathcal{M}(\mathcal{G}_{-w}, V)$ .*

*Proof.* By Proposition 1, we can remove all the connections of  $w$  to its children, and conclude that  $\mathcal{M}(\mathcal{G}_{-w}, V) \subseteq \mathcal{M}(\mathcal{G}, V)$ .

To prove the inverse, take  $P \in \mathcal{M}(\mathcal{G}, V)$ . Every  $X_c$  for  $c \in \text{ch}_{\mathcal{G}}(w)$  is  $\sigma$ - $(X_u, X_w, X_{\text{pa}_{\mathcal{G}}(c) \setminus \{u, w\}}, E_c)$  measurable. If we define  $\tilde{X}_u \equiv (X_u, X_w)$ , we have that  $X_c$  is  $\sigma$ - $(\tilde{X}_u, X_{\text{pa}_{\mathcal{G}}(c) \setminus \{u, w\}}, E_c)$  measurable. Identifying  $\tilde{X}_u$  with the node  $u$  of  $\mathcal{G}_{-w}$ , this gives us  $\mathcal{M}(\mathcal{G}, V) \subseteq \mathcal{M}(\mathcal{G}_{-w}, V)$ . □

In other words, the marginal model over the visible variables is unchanged if we delete all latent variables whose children are a subset of the children of some other latent.



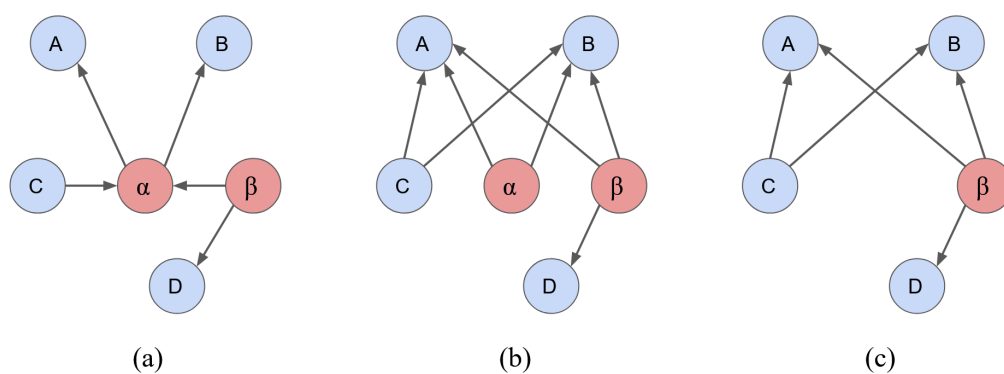


Figure A.1: (a) A DAG, with visible nodes in blue and latent nodes in red. (b) DAG obtained after exogenizing  $\alpha$  (Lemma 4). (c) DAG obtained after removing  $\alpha$ , since its children in (b) are a subset of the children of  $\beta$  (Lemma 5). These three DAGs are observationally equivalent.

# Appendix B

## Proofs of Equivalence Results

### B.1 Proof of HLP Proposition

Here, we will prove Theorem 26.4 from [8], that we named HLP Proposition (Proposition 4).

In this proof, we are looking at DAGs (and not mDAGs) and we include latent parents when writing  $pa_{\mathcal{G}}(v)$ . It may be helpful to follow Figure B.1 during the proof.

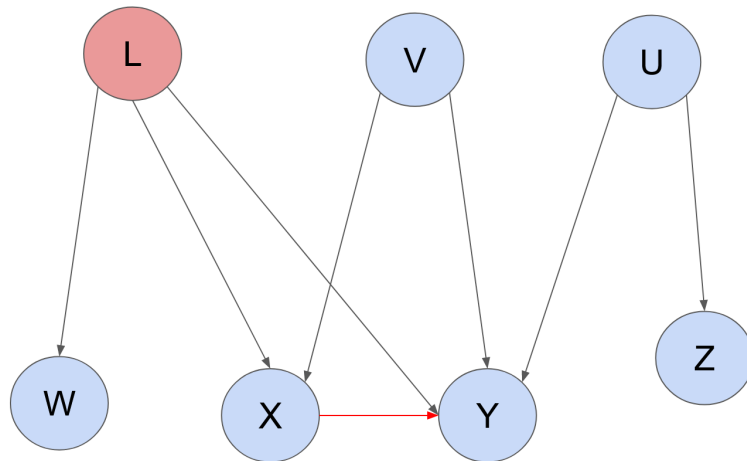


Figure B.1: The red arrow  $X \rightarrow Y$  in this DAG can be added by means of the HLP Proposition. L is a latent node, while U, V, W, X, Y, Z are visible nodes.

Let  $\mathcal{G}$  be a DAG (with latent and visible nodes) and  $X$  and  $Y$  be two of its visible nodes such that the hypotheses of Proposition 4 are satisfied, and that there is no edge  $X \rightarrow Y$ . Let  $\mathcal{G}'$  be the same DAG, but with the addition of the edge  $X \rightarrow Y$ .

By edge dropping (Proposition 1), we can see that every probability distribution compatible with  $\mathcal{G}$  is also compatible with  $\mathcal{G}'$ .

To show the other way around, we look at the Structural Equation Property of

$\mathcal{G}'$ . Since  $pa_{\mathcal{G}'}(Y) = pa_{\mathcal{G}}(Y) \cup X$ , we have that:

$$\begin{aligned} X_Y \text{ is } \sigma(X_{pa_{\mathcal{G}'}(Y)}, E_Y)\text{-measurable} \\ \Downarrow \\ X_Y \text{ is } \sigma(X_{pa_{\mathcal{G}}(Y)}, X_X, E_Y)\text{-measurable} \end{aligned}$$

Where  $X_v$  is the random variable and  $E_v$  is the error variable for node  $v$ .

If a probability distribution  $P$  is compatible with  $\mathcal{G}'$ , it has to be the probability distribution of one set of random variables  $X_v$  that obeys this.

From the Structural Equation Property of  $\mathcal{G}'$ , we also have:

$$X_X \text{ is } \sigma(X_{pa_{\mathcal{G}'}(X)}, E_X)\text{-measurable ,}$$

and we know that the parents of  $X$  are the same in both  $\mathcal{G}$  and  $\mathcal{G}'$ :

$$X_X \text{ is } \sigma(X_{pa_{\mathcal{G}}(X)}, E_X)\text{-measurable ,}$$

Since  $X$  has at least one latent parent, we can include the error variable  $E_X$  into one of these latent parents (if we view  $E_v$  as one latent variable with only  $X$  as a child, we can absorb it into other latent parent by the second reduction rule of [9] - Lemma 3.8 in the Arxiv version). So:

$$\begin{aligned} X_X \text{ is } \sigma(X_{pa_{\mathcal{G}}(X)})\text{-measurable} \\ \Downarrow \\ \sigma(X_{pa_{\mathcal{G}}(Y)}, X_X, E_Y) \subseteq \sigma(X_{pa_{\mathcal{G}}(Y)}, X_{pa_{\mathcal{G}}(X)}, E_Y) \end{aligned}$$

This means that, if

$$\begin{aligned} X_Y \text{ is } \sigma(X_{pa_{\mathcal{G}}(Y)}, X_X, E_Y)\text{-measurable,} \\ \text{then } X_Y \text{ is also } \sigma(X_{pa_{\mathcal{G}}(Y)}, X_{pa_{\mathcal{G}}(X)}, E_Y)\text{-measurable.} \end{aligned}$$

Since  $pa_{\mathcal{G}}(X) \subseteq pa_{\mathcal{G}}(Y)$ , we have

$$\sigma(X_{pa_{\mathcal{G}}(Y)}, X_{pa_{\mathcal{G}}(X)}, E_Y) = \sigma(X_{pa_{\mathcal{G}}(Y)}, E_Y)$$

Thus, our final conclusion is:

$$\begin{aligned} X_Y \text{ is } \sigma(X_{pa_{\mathcal{G}'}(Y)}, E_Y)\text{-measurable} \\ \Downarrow \\ X_Y \text{ is } \sigma(X_{pa_{\mathcal{G}}(Y)}, E_Y)\text{-measurable} \end{aligned}$$

What means that all probability distributions compatible with  $\mathcal{G}'$  are also compatible with  $\mathcal{G}$ . This proves the other way around and shows that  $\mathcal{G}$  and  $\mathcal{G}'$  are equivalent.

## B.2 Proof of Weak Face Splitting

Here, we prove Proposition 6. This proof is due to Noam Finkelstein. It is *not* the proof originally presented in [9].

Let  $pa_{\mathcal{G}}^*(D) = pa_{\mathcal{G}}(D) \setminus (pa_{\mathcal{G}}(C) \cup C)$  be the set of parents of  $D$  apart from  $C$  and  $pa_{\mathcal{G}}(C)$ . In this proof, we are going to treat all these extra parents of  $D$  as visible; if some of them are claimed to be latent, we can simply ignore their values at the end, and this will not change the marginal probability distribution over the other nodes.

Any joint probability distribution  $P(C, D, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D))$  can be written as:

$$P(C, D, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D)) = P(C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D)) P(D|C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D))$$

Now, we analyse the constraints that both mDAGs (pre-split and post-split) impose on this probability distribution.

Firstly, we see that both mDAGs impose the same constraints on the first term,  $P(C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D))$ . This is true because the subgraph over  $C, pa_{\mathcal{G}}(C)$  and  $pa_{\mathcal{G}}^*(D)$  is the same in both mDAGs, and marginalizing the descendants (in this case,  $D$ ) is just like removing them from the graph.

Thus, what is left to study is the second term,

$$P(D|C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D)).$$

This term can have constraints on specific evaluations of the parents of  $D$ , namely constraints that involve only one distribution

$$P(D|C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z),$$

and it also can have constraints that involve more than one evaluation, for example a constraint that involves

$$P(D|C = x, pa_G(C) = y, pa_G^*(D) = z)$$

and  $P(D|C = x', pa_G(C) = y', pa_G^*(D) = z')$ .

It is immediate that both mDAGs impose the same constraints of the first type, because conditioning on ancestors (fixing their values) is just like removing them from the graph, and the subgraph over  $D$  is the same in both mDAGs.

Therefore, we just need to check if there are changes on constraints that involve more than one evaluation of the parents of  $D$ , that we will call "cross-counterfactual constraints".

Below we show the system of equations that represents one possible evaluation of the parents of  $D$ :

$$\left\{ \begin{array}{l} P(D = X_D^{(1)} | C = x, pa_G(C) = y, pa_G^*(D) = z) = \\ = \int_u P(u | C = x, pa_G(C) = y, pa_G^*(D) = z) \\ \quad \prod_{d_i \in D} P(d_i = X_{d_i}^{(1)} | C = x, pa_G(C) = y, pa_G^*(d_i) = z_i, u) \\ \cdot \\ \cdot \\ \cdot \\ P(D = X_D^{(k)} | C = x, pa_G(C) = y, pa_G^*(D) = z) = \\ = \int_u P(u | C = x, pa_G(C) = y, pa_G^*(D) = z) \\ \quad \prod_{d_i \in D} P(d_i = X_{d_i}^{(k_i)} | C = x, pa_G(C) = y, pa_G^*(d_i) = z_i, u) \end{array} \right.$$

where  $D = \{d_1, \dots, d_n\}$ ,  $pa_G^*(D) = (pa_G^*(d_1), \dots, pa_G^*(d_n))$  and  $z = (z_1, \dots, z_n)$ , and  $k_i$  is the cardinality of  $d_i$ , such that  $k = \prod_{i=1, \dots, n} k_i$ . The variable  $u$  represents the latent that will be split.

Thus, for each evaluation of the parents of  $D$  we have a system of equations where each equation corresponds to the probability of one specific evaluation of

D. The parameters of such systems are

$$P(u|C = x, pa_G(C) = y, pa_G^*(D) = z) \text{ and} \\ P(d_i = X_{d_i}^{(m)}|C = x, pa_G(C) = y, pa_G^*(d_i) = z_i, u),$$

for  $i = 1, \dots, n$  and  $m = 1, \dots, k_i$ .

Cross-counterfactual constraints between two evaluations of the parents of  $D$  appear when we have a parameter that is present in equations of the two corresponding systems. For example, we could have two systems whose evaluations of the parents of  $D$  only differ by one of the  $z_i$ 's. Then, the parameter

$$P(d_j = X_{d_j}^{(m)}|C = x, pa_G(C) = y, pa_G^*(d_j) = z_j, u); i \neq j$$

is present in both systems of equations when the extra parents of  $d_j$  do not overlap with the extra parents of  $d_i$ :

$$P(d_j|C, pa_G(C), pa_G^*(d_j), pa_G^*(d_i) = z_i, u) = P(d_j|C, pa_G(C), pa_G^*(d_j), pa_G^*(d_i) = z'_i, u)$$

This equality constraint involving  $P(D|C, pa_G(C), pa_G^*(D), u)$  leads to an inequality constraint on  $P(D|C, pa_G(C), pa_G^*(D))$ , after marginalizing on  $u$ .

Constraints like the one in this example, that arise from the  $z_i$ 's, will be called "cross-Z-counterfactual" constraints. These are the same both before and after splitting, because the split does not affect  $pa_G^*(D)$ . The constraints that are more relevant to us are the "cross-C-counterfactual", that arise from changing the evaluation of nodes in  $C$ , and "cross-P-counterfactual", that arise from changing the evaluation of nodes in  $pa_G(C)$ .

Since all of  $C$  and  $pa_G(C)$  are parents of every  $d_i \in D$ , if two systems differ by the evaluation of some node of  $C$  or  $pa_G(C)$  then all of their parameters of the form  $P(d_i|C, pa_G(C), pa_G^*(d_i), u)$  are different. Thus, there are neither cross-C-counterfactual nor cross-P-counterfactual constraints on the parameter  $P(D|C, pa_G(C), pa_G^*(D), u)$ . This is true both before and after the split.

The only parameters that change with the split are the ones of the form  $P(u|C, pa_G(C), pa_G^*(D))$ , that appear in all equations. After the split, we have that  $u \perp C, u \perp pa_G(C)|C$  and  $u \perp pa_G^*(D)$ . Thus,  $P(u|C, pa_G(C), pa_G^*(D)) = P(u)$ . Before the split, on the other hand, we only have  $u \perp pa_G^*(D)$ .

However, this is not a problem: since  $u$  is integrated out, these parameters can only give rise to inequality constraints if we also have equality constraints on

$P(D|C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(d_j), u)$ . As we discussed,  $P(D|C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(d_j), u)$  does not have cross-C-counterfactual or cross-P-counterfactual equality constraints.

In conclusion, all the single-evaluation constraints and cross-system constraints that the distribution of  $D$  conditioned on its parents have are the same in both the pre-split and the post-split mDAGs. Thus, these mDAGs are observationally equivalent.

### B.3 Proof of Strong (Simultaneous) Face Splitting

The proof is analogous to the proof of Weak Face Splitting. Just like in that case, we only need to look at cross-C-counterfactual and cross-P-counterfactual constraints on  $P(D|C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(D))$ . The system of equations is:

$$\left\{ \begin{array}{l} P(D = X_D^{(1)} | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z) = \\ \int_u P(u_1 | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z) \dots P(u_n | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z) \\ \prod_{d_i \in D} P(d_i = X_{d_i}^{(1)} | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(d_i) = z_i, u_1, \dots, u_n) \\ \cdot \\ \cdot \\ \cdot \\ P(D = X_D^{(k)} | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z) = \\ \int_u P(u_1 | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z) \dots P(u_n | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(D) = z) \\ \prod_{d_i \in D} P(d_i = X_{d_i}^{(k_i)} | C = x, pa_{\mathcal{G}}(C) = y, pa_{\mathcal{G}}^*(d_i) = z_i, u_1, \dots, u_n) \end{array} \right.$$

where we defined  $C = \cup_{i=1, \dots, n} C_i$ .

Just like in Appendix B.2, here  $P(d_i = X_{d_i}^{(m)} | C, pa_{\mathcal{G}}(C), pa_{\mathcal{G}}^*(d_i), u_1, \dots, u_n)$  does not give any cross-C-counterfactual or cross-P-counterfactual constraint. Thus, the same arguments follow and the proof of Strong Face Splitting is just a simple extension to the proof of Weak Face Splitting.

# Appendix C

## Proof of the Only Hypergraphs Rule (GPT valid)

Here, we prove Propositions 10 and 12. To make this proof, we are going to use the *inflation technique*, developed in [22].

Let  $\mathcal{G}$  be a DAG and  $X$  be a subset of the nodes of  $\mathcal{G}$ . If  $\text{an}_{\mathcal{G}}(X)$  is the set of ancestors of  $X$  together with  $X$  itself, we say that the *ancestral subgraph*  $\text{AnSubDAG}_{\mathcal{G}}(X)$  of  $X$  is the subgraph of  $\mathcal{G}$  formed by the nodes in  $\text{an}_{\mathcal{G}}(X)$  and the edges that have both endpoints in  $\text{an}_{\mathcal{G}}(X)$ .

We can construct a different DAG  $\mathcal{G}'$  by creating copies of the nodes  $v$  of  $\mathcal{G}$ , that we will denote as  $v_i$ . The index  $i$  is called "copy-index", and it increases with the number of copies we make. If two subgraphs  $\mathcal{G}'_1$  and  $\mathcal{G}'_2$  of  $\mathcal{G}'$  are the same up to removal of the copy-indices, we denote  $\mathcal{G}'_1 \sim \mathcal{G}'_2$ . If we have a set  $X$  of nodes of  $\mathcal{G}$  and a set  $X'$  of nodes of  $\mathcal{G}'$  such that  $X'$  contains exactly one copy of every node of  $X$ , we also denote  $X' \sim X$ .

We define an inflation of a causal model as:

**Definition 21** (Inflation of a DAG). *Let  $\mathcal{G}$  be a DAG. Another DAG  $\mathcal{G}'$  is said to be an inflation of  $\mathcal{G}$  if it is constructed from copies of the nodes of  $\mathcal{G}$  as described above, and, for every observed node  $v_i$  of  $\mathcal{G}'$ , we have*

$$\text{AnSubDAG}_{\mathcal{G}'}(v_i) \sim \text{AnSubDAG}_{\mathcal{G}}(v), v_i \sim v$$

For example, Figure C.2 presents an inflation of the DAG of Figure C.1. There, we used the prime as the copy-index;  $A'$  and  $A$  are the same upon removal of the copy-index.

**Definition 22** (Injectable Set). *Let  $\mathcal{G}$  be a DAG and  $\mathcal{G}'$  a DAG constructed from copies of the nodes of  $\mathcal{G}$ . A set  $X'$  of observed nodes of  $\mathcal{G}'$  is said to be injectable, denoted by  $X' \in \text{InjectableSets}(\mathcal{G}')$ , if there is some set  $X$  of nodes of  $\mathcal{G}$  such that*

$$X' \sim X \text{ and } \text{AnSubDAG}_{\mathcal{G}'}(X') \sim \text{AnSubDAG}_{\mathcal{G}}(X).$$



In this case, we say that  $X$  is an element of  $\text{ImagesInjectableSets}(\mathcal{G})$ .

For example, in C.2,  $\{A, B\}$  is an injectable set but  $\{D, A\}$  is not.

Let  $\mathcal{G}$  be a DAG.  $\mathcal{G}'$  is an inflation of  $\mathcal{G}$ ,  $S' \subseteq \text{InjectableSets}(\mathcal{G}')$  is a collection of injectable sets and  $S \subseteq \text{ImagesInjectableSets}(\mathcal{G})$  are the images of  $S'$ . Lemma 4 of [22] proves that if a probability distribution  $\{P_V : V \in S\}$  is compatible with  $\mathcal{G}$ , then  $\{P_{V'} : V' \in S'\}$  defined via  $P_{V'} = P_V$  for  $V' \sim V$  must be compatible with  $\mathcal{G}'$ .

Ultimately what happens is that inequalities for the original graph can be derived from inequalities for the inflated graph. Now, let us see when this is also applicable to quantum causal structures.

**Definition 23** (Inflationary fan-out). *An inflationary fan-out is a latent node that points to two or more nodes that are copies of one another.*

As discussed in [22], when an inflation does not have any inflationary fan-outs, the inequalities that can be derived from it are GPT valid. In particular, they are quantumly valid.

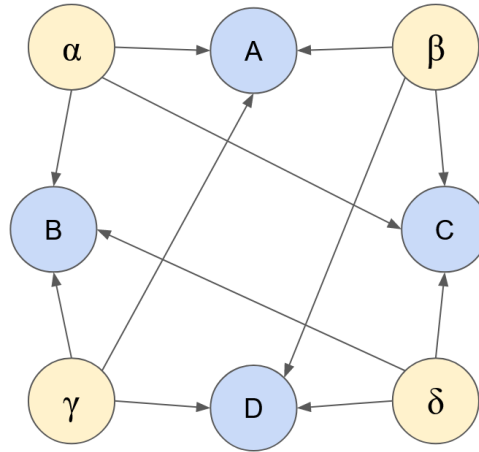


Figure C.1: DAG considered in the proof of Lemma 6: it has four visible nodes and a latent common cause between every set of three visible nodes. It does not have any edges going out of visible nodes.

Now, we can prove an useful Lemma:

**Lemma 6.** *Let  $\mathcal{G}$  be a latent-variable DAGs with  $N$  visible nodes that do not have any edges going out of visible nodes. If  $\mathcal{G}$  does not include a latent node that points to all the  $N$  visible nodes, then it is not saturated when treated as a quantum causal structure.*

*Proof.* We are going to show that such DAGs are not saturated by proving that they cannot explain a distribution where there is perfect correlation between all observed nodes.

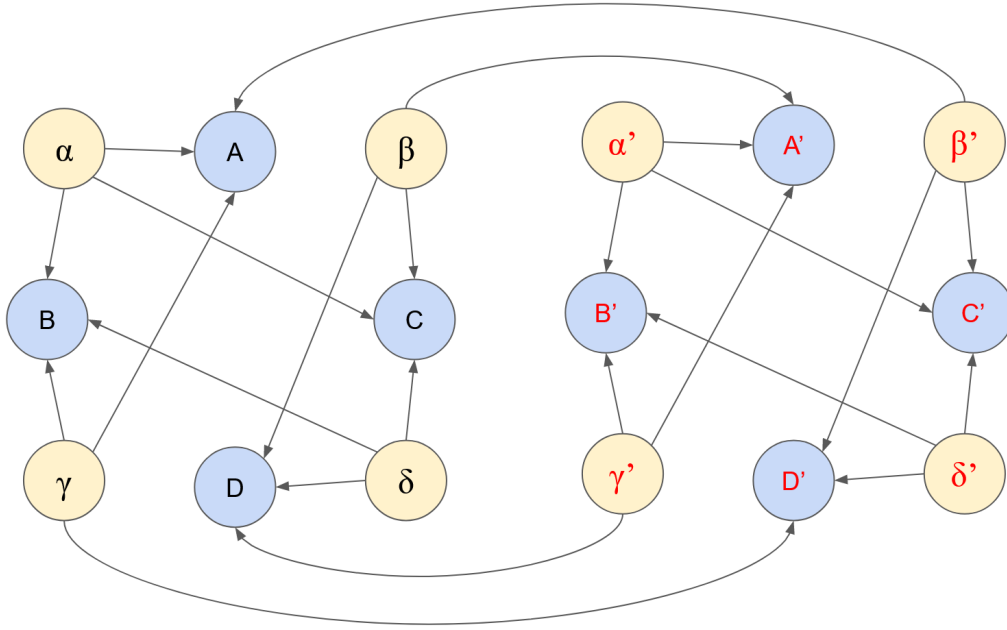


Figure C.2: Inflation of the DAG of Figure C.1.

All the DAGs that do not have edges between the visibles and do not have a  $N$ -way common cause are dominated by the one that has common causes between every subset of  $(N-1)$  visible variables. For the case  $N = 4$ , this DAG is shown in Figure C.1.

In Figure C.2, we present an inflation of the DAG C.1. This inflation does not contain any fan-outs: there is no latent node in C.2 that parents one of the original observed nodes and its primed version at the same time. Therefore, the inequalities obtained from this inflation are quantum (and GPT) valid.

In the inflation C.2, the sets  $\{A, B\}$ ,  $\{B, C\}$  and  $\{C, D\}$  are injectable. We can apply Lemma 4 of [22] to the the family of injectable sets

$$S' = \{\{A, B\}, \{B, C\}, \{C, D\}\}.$$

Thus, if the probability distribution where  $A = B = C = D$  was compatible with the original DAG, then necessarily the probability distribution  $\{P_{V'} : V' \in S'\}$  where  $A = B$ ,  $B = C$  and  $C = D$  should be compatible with the inflated DAG. However this would imply  $D = A$ , what is not allowed because  $A \perp D$  in C.2.

Therefore, the DAG C.1 cannot explain perfect correlation, what shows that it is not saturated. Thus, all the other DAGs with 4 observed variables considered by the Lemma are also not saturated.

The generalization of this construction to arbitrary  $N$  is given in Appendix B

of [23]. □

With this Lemma, we are ready to prove the proposition:

**Proposition 12** (Only Hypergraphs Rule - Quantum Version). *Let  $\mathcal{G}$  and  $\mathcal{G}'$  be two latent-variable DAGs with visible nodes  $V$  that do not have any edges between the visible nodes. If  $\mathcal{G} \neq \mathcal{G}'$ , then  $\mathcal{M}_{\mathcal{Q}}(\mathcal{G}, V) \neq \mathcal{M}_{\mathcal{Q}}(\mathcal{G}', V)$ .*

*Proof.* If  $\mathcal{G} \neq \mathcal{G}'$ , then there must be at least one facet in one of them that is not included in any facet of the other. Say that  $\mathcal{G}$  has the facet  $B$ , but  $\mathcal{G}'$  does not have  $B$  as a facet nor a face.

Then, if we look at the subgraphs defined by the nodes in  $B$ , Lemma 6 says that such a subgraph of  $\mathcal{G}$  can explain perfect correlation (and it is saturated), while the subgraph of  $\mathcal{G}'$  cannot. □

The Only Hypergraphs Rule is GPT valid, because Lemma 6 is GPT valid.

# Appendix D

## TC Fraser's Algorithm for Feasible Supports

The algorithm for finding the supports compatible with an mDAG is fundamentally a brute force search. It consists of enumerating all the possible responses that a visible variable can have to its parents, given that the cardinalities of the visible variables and the latent variables are fixed.

In [18], it is shown that whenever the visible variables have finite cardinalities, we can assume the latent variables to have finite cardinalities without loss of generality. Furthermore, for a given number of events in the support, one gets an upper bound on the cardinality of the latent variables.

Given an mDAG  $\mathcal{G}$ :

1. Fix the cardinalities of every latent variable to be  $k$  and the cardinality of each visible variable  $v$  to be  $c_v$ .
2. For each visible variable, enumerate all possible response functions it can have to the valuations of its parents.

For example, if a visible  $v$  has only one parent and it is latent, it has  $(c_v)^k$  possible response functions; we could have that  $v$  reacts with the each one of its  $c_v$  possible values for each one of the  $k$  possible valuations of its parent.

In general, if the total cardinality of the parents of  $v$  is  $K(v)$ , then we have  $(c_v)^{K(v)}$  possibilities of response function for  $v$ .

3. For each possibility of response functions of all the visible variables of  $\mathcal{G}$ , we compute the support: the set of visible events that occur under that response function, for some valuation of the latent variables.

The question that is left is how to choose  $k$ . First we note that, if a support containing  $s$  events is compatible with  $\mathcal{G}$ , it will certainly be found with the

algorithm for all  $k \geq s$ . Thus,  $s$  is an upper bound on the values of  $k$  (cardinality of latent variables) that we need to consider.

Assuming that all the  $|V|$  visible variables have the same cardinality  $c$ , the maximum number of events that we can have is  $c^{|V|}$ . So, to be certain that we found all compatible supports, we should choose  $k = c^{|V|}$ .

When implementing the algorithm in this work, we consider that all the visible variables are binary ( $c = 2$ ). This restriction is not a problem for our goal of checking that two mDAGs *do not* have the same set of Infeasible Supports. For mDAGs  $\mathcal{G}$  and  $\mathcal{G}'$  with higher  $c > 2$ , we can gather events together (for example joining events  $x$  and  $y$  in  $z = x \wedge y$ ) to go to the binary case. Thus, if the sets of Infeasible Supports are different for  $c = 2$ , they are also different for  $c > 2$ .

# Appendix E

## Counter-Example of a quantum exogenization Lemma

A counter-example for Lemma 4 in the quantum case is presented in [13]. It is given by the two DAGs of Figure E.1.

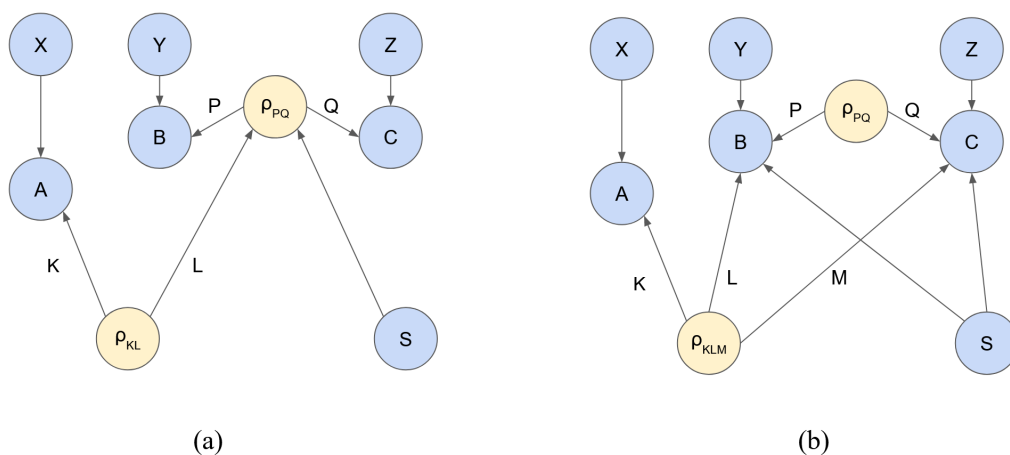


Figure E.1: In (a), it is possible for  $P(A, B|X, Y, S = 0)$  to maximally violate a Bell inequality between A and B while  $P(A, C|X, Z, S = 1)$  maximally violates a Bell inequality between A and C. In (b), this is not possible because of the monogamy of pure entanglement.

The DAG E.1(b) is the exogenized version of E.1(a). By the following reasoning, we can show that DAGs E.1(a) and E.1(b) are *not* quantumly equivalent.

In E.1(a), the classical variable S determines what is the quantum channel  $\Omega_s : \mathcal{B}(\mathcal{H}_L) \rightarrow \mathcal{B}(\mathcal{H}_{PQ})$ . We could have a quantum realization of this DAG such that when  $S = 0$  we have  $\rho_{PQ} = \rho_L \otimes \mathbb{I} \in \mathcal{B}(\mathcal{H}_{PQ})$  and when  $S = 1$  we have  $\rho_{PQ} = \mathbb{I} \otimes \rho_L \in \mathcal{B}(\mathcal{H}_{PQ})$ . In other words, S determines whether the state  $\rho_L$  goes to node B or to node C. Therefore, we can establish a Bell scenario between A and B when  $S = 0$  and a Bell scenario between A and C when  $S = 1$ . Thus,  $P(A, B|X, Y, S = 0)$  can maximally violate a Bell inequality between A and B, and  $P(A, C|X, Z, S = 1)$  can maximally violate a Bell inequality between A and C.

In E.1(b), the only quantum node that connects A with B and C is  $\rho_{KLM}$ , and

it is prepared independently of S. This means that we cannot use a strategy like the one implemented in E.1(a), where S determines if the state that goes to A is entangled with the state that goes to B or with the state that goes to C.

If A had access to S, we could create a quantum realization where  $\mathcal{H}_K = \mathcal{H}_{K'} \otimes \mathcal{H}_{K''}$ , and S tells A whether to measure  $\mathcal{H}_{K'}$  or  $\mathcal{H}_{K''}$ . Thus, setting  $\rho_{K'}$  to be purely entangled with  $\rho_L$  and  $\rho_{K''}$  to be purely entangled with  $\rho_M$ , we could have that S determines which of the pairs AB or AC maximally violate Bell's inequalities. However, this cannot be done because A does not have access to S in the DAG E.1(b).

Finally, the monogamy of pure entanglement says we cannot have a joint state  $\rho_{KLM}$  such that the state that goes to A is purely correlated to the state that goes to B at the same time that it is purely correlated to the state that goes to C (independently of S).

Therefore, we have a probability distribution that can be achieved by E.1(a) but not by E.1(b).

# Appendix F

## District Factorization

In this work, we have been talking only about *passive* observation of events. Thus, we use conditional probabilities, that say what is the updated knowledge you have about the remaining variables when you learn about an event. However, we could also design experiments that involve *active interventions*: fixing the value of some variable. For example, in a drug test the active intervention is giving the medicine for a randomized group of patients, while the passive observation is to analyze data from the population of patients that voluntarily took the medicine<sup>1</sup>. In the active case, the influence of external factors on the decision of taking the medicine is eliminated. Therefore, fixed nodes do not have parents in the causal structure.

To describe interventions, we need the definition of a *probability kernel*. It is taken from [24]:

**Definition 24** (Probability Kernel). *Let  $X_V$  and  $X_W$  be random variables that take values in  $\mathcal{X}_V$  and  $\mathcal{X}_W$ , respectively. A probability kernel is a non-negative map  $q_V : \mathcal{X}_V \times \mathcal{X}_W \rightarrow [0, 1]$ , satisfying:*

$$\sum_{x_V \in \mathcal{X}_V} q_V(x_V | x_W) = 1 \text{ for all } x_W \in \mathcal{X}_W$$

where we write the evaluation of  $q_V$  in  $x_V \in \mathcal{X}_V$  and  $x_W \in \mathcal{X}_W$  as  $q_V(x_V | x_W)$ .

Note that, in general,  $q_V$  cannot be formed by taking a probability distribution on  $X_V$  and conditioning on  $X_W$ . Such conditional probabilities are special cases of kernels.

We use kernels to express the joint distribution of a set of variables when there are interventions. The referred kernel is obtained by dividing the distribution over all variables by the distribution of the fixed one. See e.g. Equation (1) of [25].

Many times, the problem of finding the compatible distributions of an mDAG can be modularized. As shown in [19], there are conditions under which we can factorize this analysis over certain subgraphs of our mDAG.

---

<sup>1</sup>Example taken from [1].



We define the subgraph induced by the district  $D$  (see Definition 14) as the mDAG  $\mathcal{G}[D]$  formed by the nodes of  $D$  as random, the nodes of  $\text{pa}_{\mathcal{G}}(D) \setminus D$  as fixed, the edges that connect these nodes and the facets  $\{B \cap D : B \in \mathcal{B}(\mathcal{G})\}$ . An example of such subgraphs can be found in Figure F.1.

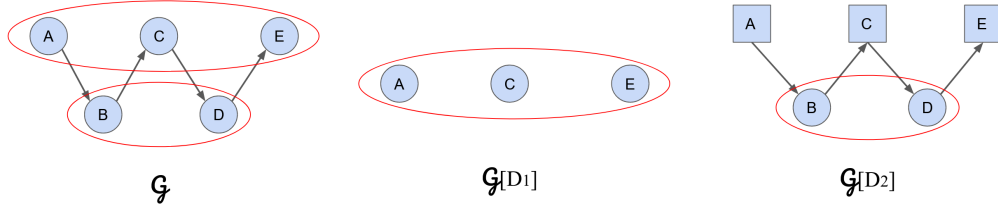


Figure F.1: An mDAG  $\mathcal{G}$  and its factorization into districts.  $\mathcal{G}$  has two districts:  $D_1 = \{A, C, E\}$  and  $D_2 = \{B, D\}$ . Squares represent fixed nodes. Example taken from [26].

The important result involving the subgraphs  $\mathcal{G}[D]$  is presented as Proposition 2.11 in [19]:

**Proposition 13** (Factorization into Districts). *Let  $\mathcal{G}$  be an mDAG with nodes  $V$  and districts  $D_1, \dots, D_k$ .*

*A probability distribution  $P$  over the variables  $V$  is compatible with  $\mathcal{G}$  if and only if:*

$$p(x_V) = \prod_{i=1}^k g_i(x_{D_i} | x_{\text{pa}_{D_i} \setminus D_i})$$

where each  $g_i$  is a probability kernel compatible with  $\mathcal{G}[D_i]$ .

For the example in Figure F.1, probability distributions  $p_{\mathcal{G}}$  that are compatible with  $\mathcal{G}$  take the form

$$p_{\mathcal{G}}(A, B, C, D, E) = p_{\mathcal{G}[D_1]}(A, C, E) g_{\mathcal{G}[D_2]}(B, D | A, C, E)$$

where  $p_{\mathcal{G}[D_1]}$  is compatible with  $\mathcal{G}[D_1]$  and  $g_{\mathcal{G}[D_2]}$  is compatible with  $\mathcal{G}[D_2]$ .

If an mDAG  $\mathcal{G}$  that can be decomposed in this way presents inequality constraints, these inequalities will already be present in subgraphs  $\mathcal{G}[D]$ . Therefore, to find examples that give QC gaps, it suffices to look at the *fundamental* mDAGs (Definition 20).

# Bibliography

- [1] Judea Pearl and Dana Mackenzie. *The Book of Why*. Basic Books, New York, NY, 2018.
  - [2] Jonathan G. Richens, Ciarán M. Lee, and Saurabh Johri. Improving the accuracy of medical diagnosis with causal machine learning. *Nature Communications*, 11(3923), 2020.
  - [3] Jessica C Bird, Robin Evans, Felicity Waite, Bao S Loe, and Daniel Freeman. Adolescent Paranoia: Prevalence, Structure, and Causal Mechanisms. *Schizophrenia Bulletin*, 45(5):1134–1142, 12 2018.
  - [4] Andres Babino and Marcelo O. Magnasco. Masks and distancing during COVID-19: a causal framework for imputing value to public-health interventions. *Scientific Reports*, 11(5183), 2021.
  - [5] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. Inferring causation from time series in Earth system sciences. *Nature Communications*, 10(2553), 2019.
  - [6] J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika*, 1:195–200, Nov 1964.
  - [7] Christopher Wood and Robert Spekkens. The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning. *New Journal of Physics*, 17, 08 2012.
  - [8] Joe Henson, Raymond Lal, and Matthew F. Pusey. Theory-independent limits on correlations from generalized Bayesian networks. 16(11):113043, 2014. Publisher: IOP Publishing.
  - [9] Robin J. Evans. Graphs for Margins of Bayesian Networks. *Scandinavian Journal of Statistics*, 43(3):625–648, 2016.
- Note: Whenever we cited Propositions and Lemmas from this reference, we

- mentioned the numbering of the Arxiv version, that is different from the numbering of the published version. This choice was made because the Arxiv version may be more accessible to the reader.
- [10] Stephen Thornton. "Karl Popper". In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition, 2021.
  - [11] Steffen L. Lauritzen, AP Dawid, BN Larsen, and HG Leimer. Independence properties of directed Markov fields. *Networks*, 20(5):491–505, 1990.
  - [12] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition, 2009.
  - [13] Elie Wolfe, Alejandro Pozas-Kerstjens, Matan Grinberg, Denis Rosset, Antonio Acín, and Miguel Navascués. Quantum Inflation: A General Approach to Quantum Causal Compatibility. *Phys. Rev. X*, 11:021043, May 2021.
  - [14] Elie Wolfe and Marina M. Ansanelli. mdag-analysis. <https://github.com/eliewolfe/mDAG-analysis>.
  - [15] Noam Finkelstein, Beata Zjawin, Elie Wolfe, Ilya Shpitser, and Robert W. Spekkens. Entropic Inequality Constraints from e-separation Relations in Directed Acyclic Graphs with Hidden Variables. Arxiv: 2107.07087, 2021.
  - [16] Robin J. Evans. Graphical Methods for Inequality Constraints in Marginalized DAGs. *2012 IEEE International Workshop on Machine Learning for Signal Processing*, pages 1–6, 2012.
  - [17] Robin J. Evans. Dependency in DAG models with Hidden Variables. 2021.
  - [18] Thomas Fraser. A Combinatorial Solution to Causal Compatibility. *Journal of Causal Inference*, 8:22–53, 07 2020.
  - [19] Robin Evans. Margins of discrete Bayesian networks. *Annals of Statistics*, 46, 01 2015.
  - [20] Tobias Fritz. Beyond Bell's theorem: correlation scenarios. *New Journal of Physics*, 14(10):103001, oct 2012.

- 
- [21] Rafael Chaves, George Moreno, Emanuele Polino, Davide Poderini, Iris Agresti, Alessia Suprano, Mariana R. Barros, Gonzalo Carvacho, Elie Wolfe, Askery Canabarro, Robert W. Spekkens, and Fabio Sciarrino. Causal Networks and Freedom of Choice in Bell's Theorem, 2021.
- [22] Elie Wolfe, Robert W. Spekkens, and Tobias Fritz. The Inflation Technique for Causal Inference with Latent Variables. *Journal of Causal Inference*, 7(2):20170020, 2019.
- [23] Miguel Navascués, Elie Wolfe, Denis Rosset, and Alejandro Pozas-Kerstjens. Genuine network multipartite entanglement. *Phys. Rev. Lett.*, 125:240505, Dec 2020.
- [24] Ilya Shpitser, Robin J. Evans, Thomas S. Richardson, and James M. Robins. Introduction to Nested Markov Models. *Behaviormetrika*, 41:3–39, 01 2014.
- [25] Thomas S. Richardson, Robin J. Evans, James M. Robins, and Ilya Shpitser. Nested Markov Properties for Acyclic Directed Mixed Graphs. *arXiv e-prints*, page arXiv:1701.06686, January 2017.
- [26] Jin Tian and Judea Pearl. A general identification condition for causal effects. page 567–573, 2002.