
Identifying Macro Causal Effects in C-DMGs over DMGs

Simon Ferreira

Sorbonne Université, INSERM,
Institut Pierre Louis d'Epidémiologie
et de Santé Publique,
F75012, Paris, France

`simon.ferreira@sorbonne-universite.fr`

Charles K. Assaad

Sorbonne Université, INSERM,
Institut Pierre Louis d'Epidémiologie
et de Santé Publique,
F75012, Paris, France

`charles.assaad@inserm.fr`

Abstract

The do-calculus is a sound and complete tool for identifying causal effects in acyclic directed mixed graphs (ADMGs) induced by structural causal models (SCMs). However, in many real-world applications, especially in high-dimensional settings, constructing a fully specified ADMG is often infeasible. This limitation has led to growing interest in partially specified causal representations, particularly through cluster-directed mixed graphs (C-DMGs), which group variables into clusters and offer a more abstract yet practical view of causal dependencies. While these representations can include cycles, recent work has shown that the do-calculus remains sound and complete for identifying macro-level causal effects in C-DMGs over ADMGs under the assumption that all clusters sizes are greater than 1. Nevertheless, real-world systems often exhibit cyclic causal dynamics at the structural level. To account for this, input-output structural causal models (ioSCMs) have been introduced as a generalization of SCMs that allow for cycles. ioSCMs induce another type of graph structure known as a directed mixed graph (DMG). Analogous to the ADMG setting, one can define C-DMGs over DMGs as high-level representations of causal relations among clusters of variables. In this paper, we prove that, unlike in the ADMG setting, the do-calculus is unconditionally sound and complete for identifying macro causal effects in C-DMGs over DMGs. Furthermore, we show that the graphical criteria for non-identifiability of macro causal effects previously established C-DMGs over ADMGs naturally extends to a subset of C-DMGs over DMGs.

1 Introduction

Understanding and identifying causal effects is a central goal in many scientific disciplines. In recent years, structural causal models (SCMs) have emerged as a foundational framework for reasoning about causality. These models encode causal assumptions through structural equations and are typically represented by acyclic directed mixed graphs (ADMGs), which capture both causal and confounding relationships. Within this framework, the do-calculus [Pearl, 1995]—based on the notion of d-separation [Pearl, 1988]—provides a complete and sound set of inference rules for identifying causal effects from observational data, assuming the causal structure is fully specified. However, SCMs do not fully capture systems with cyclic causal dependencies at the structural level, which are common in public health, biology, economics, and engineering systems. For example, there can be a cyclic relation between poor mental health (*e.g.*, depression or anxiety) and substance use (*e.g.*, alcohol, drugs). The worsening of mental health and increase in substance use can occur in tight time-frames (daily or even hourly), especially in high-risk populations. Over time, they may reach a cyclic equilibrium where both reinforce each other without a clear causal ordering. To address

this, the notion of input-output structural causal models (ioSCMs) has been proposed [Forré and Mooij, 2020]. These models generalize SCMs by allowing for cycles and induce a new class of graphs known as directed mixed graphs (DMGs) [Richardson, 1997, Forré and Mooij, 2017, 2018, Forré and Mooij, 2020, Boeken and Mooij, 2024], which provide a richer representation of causal structures. Furthermore, Forré and Mooij [2020] introduced an extension of d-separation to DMGs, called σ -separation, and showed that the do-calculus, when replacing d-separation by σ -separation becomes sound for identifying causal effects in DMGs [Forré and Mooij, 2020].

However, in many real-world applications—particularly those involving high-dimensional data or limited domain knowledge—it is often unrealistic to assume a complete specification of the underlying causal graph. This has motivated the development of partially specified graphical models [Maathuis and Colombo, 2013, Perkovic et al., 2016, Perkovic, 2020, Jaber et al., 2022, Wang et al., 2023, Assaad et al., 2023, Anand et al., 2023, Wahl et al., 2024, Reiter et al., 2024, Boeken and Mooij, 2024, Ferreira and Assaad, 2024, 2025a], and in particular cluster graphs. Cluster graphs abstract away some of the fine-grained details by grouping variables into clusters, thus offering a more flexible and scalable representation of complex systems. Importantly, cluster graphs allow for cycles, which can arise naturally in feedback systems or time-dependent processes, complicating the analysis compared to traditional ADMGs. In these graphs, causal effects can be separated into two types: a micro causal effect where the interest is the effect of variable within a cluster on another variable in another cluster; and the macro causal effect where the interest is the effect of a set of an entire cluster on another entire cluster. In this work, we focus on the latter. Anand et al. [2023], Tikka et al. [2023] have shown that do-calculus (the version using d-separation) remains both sound and complete for identifying macro causal effects when the cluster graph representing ADMGs is acyclic. Ferreira and Assaad [2025a,b] showed that the do-calculus (the version using d-separation) is also sound and complete for identifying macro-level causal effects when the cluster graph representing an ADMG is cyclic, denoted here as C-DMG over ADMGs, assuming either that the size of the clusters is unknown, or that each cluster contains more than one variable [Ferreira and Assaad, 2025b].

Motivated by these developments, we consider the problem of identifying macro-level causal effects in cluster graphs representing DMGs, denoted as C-DMGs over DMGs, a natural generalization of previous work. Our contributions are threefold:

- We prove that σ -separation [Forré and Mooij, 2018]—a fundamental tool in causal reasoning in DMGs—is sound and complete in C-DMGs over DMGs.
- We prove that do-calculus (the version using σ -separation) [Forré and Mooij, 2020] is sound and complete for identifying macro-level causal effects in C-DMGs over DMGs—unconditionally, and without the constraints needed in the case of C-DMGs over ADMGs [Ferreira and Assaad, 2025b, Yvernes, 2025].
- We show that the graphical characterization of non-identifiability previously developed for C-DMGs over ADMGs [Ferreira and Assaad, 2025a,b] also applies for C-DMGs over DMGs under an additional assumption.

The remainder of the paper is organized as follows: In Section 2, we formally presents C-DMGs over DMGs. In Section 3, we show that σ -separation and the do-calculus is sound and complete for macro causal effects in C-DMGs over DMGs and present a graphical characterization for the non-identifiability of these effects. Finally in Section 4, we conclude the paper while showing its limitations. All proofs are deferred to the appendix.

2 Preliminaries

To streamline the presentation and avoid repetitive explanations, we will adopt the unified notation $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$ to refer to any type of graph. This notation allows us to generalize results and discussions without redundancy across different graph types. In the remainder, for every vertex $V^* \in \mathbb{V}^*$ in a graph $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$, we will refer to its parents by $Pa(V^*, \mathcal{G}^*)$, its ancestors by $An(V^*, \mathcal{G}^*)$, and its descendants by $De(V^*, \mathcal{G}^*)$. We consider that a vertex counts as its own descendant and as its own ancestor. In addition, the strongly connected component of a vertex is defined as $Scc(V, \mathcal{G}^*) = An(V, \mathcal{G}^*) \cap De(V, \mathcal{G}^*)$.

In this section, we present the essential definitions and notations that will be used throughout the paper, ensuring clarity and consistency in the exposition of our results. In this work, we assume

causal relations are modeled using an input/output structural causal model (ioSCM) [Forré and Mooij, 2020]—which extends classical structural causal models (SCMs) [Pearl, 2009] by allowing for the presence of cycles. Unlike classical SCMs, ioSCMs allow structural equations to mutually depend on each other. For example, in the cyclic system:

$$X := f_X(Y, L_X) \quad ; \quad Y := f_Y(X, L_Y) \quad ; \quad (X, Y) := f_{(X,Y)}(L_X, L_Y),$$

X functionally depends on Y , and Y functionally depends on X , forming cycle. When cycles exist, instead of computing variables in a top-down order as in SCMs, ioSCMs rely on fixed-point solutions. That is, a joint assignment to the variables that simultaneously satisfies all equations. This is analogous to finding an equilibrium in dynamic systems. Firstly, let us properly define the notion of loops as it will be useful to guarantee the compatibility of the causal mechanisms in ioSCMs.

Definition 1 (Loops). *In a directed graph $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$, a loop is a set of vertices $\mathbb{S} \subseteq \mathbb{V}^*$ such that there exists a directed path between every pair of distinct vertices in the subgraph induced by $\mathbb{S}$ i.e., $\forall U \neq V \in \mathbb{S}, V \in \text{De}(U, \mathcal{G}^*|_{\mathbb{S}})$.*

Note that every singleton $\mathbb{S} \in \{\{V\} \mid V \in \mathbb{V}^\}$ and every strongly connected components $\mathbb{S} \in \{\text{Scc}(V, \mathcal{G}^*) \mid V \in \mathbb{V}^*\}$ are loops. We write the set of all loops of the graph $\mathcal{G}^*$ as $\mathcal{L}(\mathcal{G}^*)$. We call cycles the loops that are not singletons.*

Next, we recall the definition of ioSCMs from Forré and Mooij [2020] with the omission of the domains of the variables.

Definition 2 (input/output Structural Causal Model (ioSCM)). *An input/output structural causal model is a tuple $\mathcal{M} = (\mathbb{L}, \mathbb{V}_{obs}, \mathbb{J}, \mathcal{G}^+, \mathbb{F}, \text{Pr}(\mathbb{I}))$, where*

- $\mathbb{L}$ is a set of latent/exogenous variables, which cannot be observed but affect the rest of the model.
- $\mathbb{V}_{obs}$ is a set of observed/endogenous variables, which are observed and every $V \in \mathbb{V}_{obs}$ is functionally dependent on some subset of $(\mathbb{L} \cup \mathbb{V}_{obs} \cup \mathbb{J}) \setminus \{V\}$.
- $\mathbb{J}$ is a set of input/intervention variables which are not functionally dependent of any other variable but rather are fixed to specific values.
- $\mathcal{G}^+ = (\mathbb{V}^+, \mathbb{E}^+)$ is a graphical structure where:
 - $\mathbb{V}^+ = \mathbb{V}_{obs} \cup \mathbb{L} \cup \mathbb{J}$
 - $\mathbb{V}_{obs} = \text{Ch}(\mathbb{L} \cup \mathbb{J}, \mathcal{G}^+)$
 - $\text{Pa}(\mathbb{L} \cup \mathbb{J}, \mathcal{G}^+) = \emptyset$
- $\mathbb{F}$ is a set of functions such that for all $\mathbb{S} \in \mathcal{L}(\mathcal{G}^+|_{\mathbb{V}_{obs}})$, $f^{\mathbb{S}}$ is a function taking as input the values of $\text{Pa}(\mathbb{S}, \mathcal{G}^+) \setminus \mathbb{S}$ and outputting values for $\mathbb{S}$ and such that $\mathbb{F}$ satisfies the global compatibility condition:

$$\begin{aligned} \forall \mathbb{S}' \not\subseteq \mathbb{S} \in \mathcal{L}(\mathcal{G}^+|_{\mathbb{V}_{obs}}), \forall \mathbb{V} \text{ values of } \text{Pa}(\mathbb{S}, \mathcal{G}^+) \cup \mathbb{S}, \\ f^{\mathbb{S}}(\mathbb{V}|_{\text{Pa}(\mathbb{S}, \mathcal{G}^+) \setminus \mathbb{S}}) = \mathbb{V}|_{\mathbb{S}} \implies f^{\mathbb{S}'}(\mathbb{V}|_{\text{Pa}(\mathbb{S}', \mathcal{G}^+) \setminus \mathbb{S}'}) = \mathbb{V}|_{\mathbb{S}'} \end{aligned} \quad (1)$$

- $\text{Pr}(\mathbb{I})$ is a joint probability distribution over $\mathbb{L}$.

An ioSCM induces a directed graph, where every variable in $\mathbb{V}^+$ corresponds to a vertex in the graph. In this directed graph, a directed edge $\rightarrow$ is drawn from one variable to another if the former serves as an input to the function that determines the latter. For simplicity, instead of working directly with these directed graphs, we consider an alternative representation known as a directed mixed graph (DMG). In a DMG, only the observed and intervened variables (i.e., $\mathbb{V}_{obs} \cup \mathbb{J}$) correspond to vertices, while hidden variables in $\mathbb{L}$ that share common outputs are represented by bidirected edges $\leftrightarrow$ between the corresponding observed variables, thereby implicitly accounting for the hidden confounding. Formally, DMGs are defined as follows:

Definition 3 (Directed mixed graph (DMG)). *Consider an ioSCM $\mathcal{M}$. The directed mixed graph $\mathcal{G} = (\mathbb{V}, \mathbb{E} = \mathbb{E}_{\rightarrow} \cup \mathbb{E}_{\leftrightarrow})$ induced by $\mathcal{M}$ is the DMG where:*

- the vertices $\mathbb{V} = \mathbb{V}_{obs} \cup \mathbb{J}$ are the endogenous variables and the intervention variables of the ioSCM; and

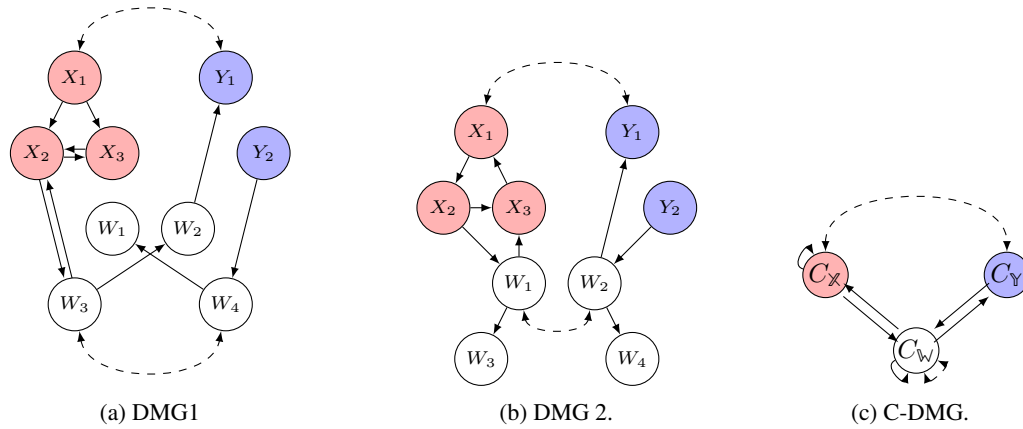


Figure 1: Two DMGs and their compatible C-DMG. Red vertices represent the exposures of interest in and blue vertices represent the outcome of interest.

- the directed edges in $\mathcal{G}$ are $\mathbb{E}_{\rightarrow} = \mathbb{E}^+|_{\mathbb{V}_{obs} \cup \mathbb{J}}$; and
- the bidirected edges in $\mathcal{G}$ are $\mathbb{E}_{\leftrightarrow} = \{X \leftrightarrow Y \mid X, Y \in \mathbb{V}, \exists L \in \mathbb{L} \text{ such that } L \rightarrow X, L \rightarrow Y \in \mathbb{E}^+\}$.

However, in many fields, constructing, analyzing, and validating a DMG remains a significant challenge for researchers due to the inherent difficulty in accurately determining causal relationships among individual variables. This complexity primarily stems from the uncertainty surrounding causal relations, making it challenging to specify the precise structure of the graph. Nevertheless, researchers can often provide a partially specified version of the DMG, which offers a more practical and compact representation of the underlying causal structure. These simplified representations, which we call Cluster-Directed Mixed Graphs over DMGs (C-DMGs over DMGs), group several variables into clusters, allowing for the representation of causal relationships at a higher level of abstraction while retaining essential structural properties of the system. In a C-DMG over DMGs, directed edges between clusters represent causal influences at the higher level, while bidirected edges capture hidden confounding effects that exist between clusters. Formally, C-DMGs over DMGs are defined as follows:

Definition 4 (Cluster directed mixed graph over DMGs (C-DMG over DMGs)). Let $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ be a DMG induced from an ioSCM $\mathcal{M}$ and $\mathbb{C} = \{C_1, \dots, C_k\}$ a partition of $\mathbb{V}$. A C-DMG over DMGs compatible with $\mathcal{G}$ according to $\mathbb{C}$ is a graph $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ where $\forall C_i, C_j \in \mathbb{C}$ the edge $C_i \rightarrow C_j$ (resp. $C_i \leftrightarrow C_j$) is in $\mathbb{E}^c$ if and only if there exists $V_i \in C_i$ and $V_j \in C_j$ such that $V_i \rightarrow V_j$ (resp. $V_i \leftrightarrow V_j$) is in $\mathbb{E}$.

Figure 1 presents a simple C-DMG over DMGs along with two of its compatible DMGs. Cycles in a C-DMG over DMGs can arise for two distinct reasons. First, unlike in a C-DMG over ADMGs, a C-DMG over DMGs can contain a cycle if there is a genuine cyclic relationship in the underlying DMG between nodes belonging to different clusters. For example, in Figure 1a the cycle between X_2 and W_3 in the DMG induces a cycle between clusters C_X and C_W in the corresponding C-DMG over DMGs in Figure 1c. Secondly, even in the absence of an actual cycle in the underlying DMG, cycles can appear in the C-DMG over DMGs due to its partial specification. This is illustrated in Figure 1b, where the edges $X_2 \rightarrow W_1$ and $W_1 \rightarrow X_3$ together create a cycle between clusters C_X and C_W in the C-DMG over DMGs in Figure 1c. Lastly, cycles that are contained in a single cluster do not appear in the C-DMG over DMGs. This is illustrated in Figure 1a with the cycle $X_2 \rightleftarrows X_3$ that does not show in the C-DMG over DMGs in Figure 1c.

We distinguish between two types of causal effects in the context of C-DMGs, the macro causal effect [Anand et al., 2023, Ferreira and Assaad, 2025a,b] and the micro causal effect [Assaad et al., 2024, Assaad, 2025]. In this paper we focus on the former and we formally define it below:

Definition 5 (Macro causal effect). Consider a DMG $\mathcal{G}$ over variables $\mathbb{V}$ induced from an ioSCM and let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a compatible C-DMG. A macro causal effect is a causal effect from a set of macro-variables $\mathbb{C}_X$ on another set of macro-variables $\mathbb{C}_Y$ where $\mathbb{C}_X$ and $\mathbb{C}_Y$ are disjoint subsets

of $\mathbb{C}$. It is written $\Pr(\mathbb{C}_Y = \mathbb{c}_Y \mid \text{do}(\mathbb{C}_X = \mathbb{c}_X))$, where the $\text{do}(\cdot)$ operator represents an external intervention.

The identification problem in causal inference aims to establish whether a causal effect of a set of variables on another set of variables can be expressed exclusively in terms of observed variables and standard probabilistic notions, such as conditional probabilities. Formally, the identification problem in the context of macro causal effects and C-DMGs over DMGs is defined as follows:

Definition 6 (Identifiability in C-DMGs over DMGs). *Let $\mathbb{C}_X$ and $\mathbb{C}_Y$ be disjoint sets of vertices in a C-DMGs over DMGs $\mathcal{G}^c$. The macro causal effect of $\mathbb{C}_X$ on $\mathbb{C}_Y$ is identifiable in $\mathcal{G}^c$ if $\Pr(\mathbb{C}_Y = \mathbb{c}_Y \mid \text{do}(\mathbb{C}_X = \mathbb{c}_X))$ is uniquely computable from any observational positive distribution compatible with $\mathcal{G}^c$.*

In the following, we will abuse the notation by writing $\Pr(\mathbb{c}_Y \mid \text{do}(\mathbb{c}_X))$ instead of $\Pr(\mathbb{C}_Y = \mathbb{c}_Y \mid \text{do}(\mathbb{C}_X = \mathbb{c}_X))$ when the setting is clear. In addition, whenever the context is clear, we will refer to C-DMGs over DMGs simply as C-DMGs.

3 Identification of Macro Causal Effects in C-DMGs over DMGs

In this section, we aim to establish that the do-calculus is both sound and complete for identifying macro-level causal effects in a C-DMG over DMGs. We begin by showing, in the first subsection, that σ -separation—originally developed for DMGs as a tool for identifying conditional independencies—remains sound and complete when extended to C-DMGs over DMGs for detecting macro-level conditional independencies. In the second subsection, we present the core theoretical contribution of this section: the soundness and completeness of do-calculus for macro causal effect identification in this setting. Finally, we provide a graphical characterization of non-identifiability, shedding light on cases where causal effects cannot be inferred from observational data alone.

3.1 The σ -separation in C-DMGs over DMGs

The standard notion of d-separation [Pearl, 1988] was originally introduced for acyclic directed mixed graphs (ADMGs). It was later shown to remain valid when extended to C-ADMGs over ADMGs [Anand et al., 2023] and C-DMGs over ADMGs [Ferreira and Assaad, 2025a,b]. However, d-separation does not apply to DMGs, which may contain cyclic causal relations in the SCM. To address this limitation, σ -separation was introduced as a generalization suitable for DMGs [Forré and Mooij, 2020]. In this subsection, we demonstrate that σ -separation can be naturally applied to C-DMGs over DMGs. We begin by formally defining σ -blocked walks and the concept of σ -separation in this generalized setting.

Definition 7 (σ -blocked walk [Forré and Mooij, 2020]). *In a graph $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$, a walk $\tilde{\pi} = \langle V_1^*, \dots, V_n^* \rangle$ is said to be σ -blocked by a set of vertices $\mathbb{W}^* \subseteq \mathbb{V}^*$ if:*

1. $V_1^* \in \mathbb{W}^*$ or $V_n^* \in \mathbb{W}^*$, or
2. $\exists 1 < i < n$ such that $\langle V_{i-1}^* \leftrightarrow V_i^* \leftrightarrow V_{i+1}^* \rangle \subseteq \tilde{\pi}$ and $V_i^* \notin \mathbb{W}^*$, or
3. $\exists 1 < i < n$ such that $\langle V_{i-1}^* \leftarrow V_i^* \leftrightarrow V_{i+1}^* \rangle \subseteq \tilde{\pi}$ and $V_i^* \in \mathbb{W}^* \setminus \text{Scc}(V_{i-1}^*, \mathcal{G}^*)$, or
4. $\exists 1 < i < n$ such that $\langle V_{i-1}^* \leftrightarrow V_i^* \rightarrow V_{i+1}^* \rangle \subseteq \tilde{\pi}$ and $V_i^* \in \mathbb{W}^* \setminus \text{Scc}(V_{i+1}^*, \mathcal{G}^*)$, or
5. $\exists 1 < i < n$ such that $\langle V_{i-1}^* \leftarrow V_i^* \rightarrow V_{i+1}^* \rangle \subseteq \tilde{\pi}$ and $V_i^* \in \mathbb{W}^* \setminus (\text{Scc}(V_{i-1}^*, \mathcal{G}^*) \cap \text{Scc}(V_{i+1}^*, \mathcal{G}^*))$.

where $\leftrightarrow$ represents $\rightarrow$ or $\leftarrow$, $\leftarrow$ represents $\leftarrow$ or $\leftrightarrow$, and $\leftrightarrow$ represents any of the three arrow type $\rightarrow$, $\leftarrow$ or $\leftrightarrow$. A walk which is not σ -blocked is said to be σ -active.

Definition 8 (σ -separation [Forré and Mooij, 2020]). *In a graph $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$, let $\mathbb{X}^*, \mathbb{Y}^*, \mathbb{W}^*$ be distinct subsets of $\mathbb{V}^*$. $\mathbb{W}^*$ is said to σ -separate $\mathbb{X}^*$ and $\mathbb{Y}^*$ if and only if $\mathbb{W}^*$ σ -blocks every walk from a vertex in $\mathbb{X}^*$ to a vertex in $\mathbb{Y}^*$. It is written $(\mathbb{X}^* \perp_{\sigma} \mathbb{Y}^* \mid \mathbb{W}^*)_{\mathcal{G}^*}$.*

The following theorem shows that σ -separation is applicable as is to C-DMGs over DMGs.

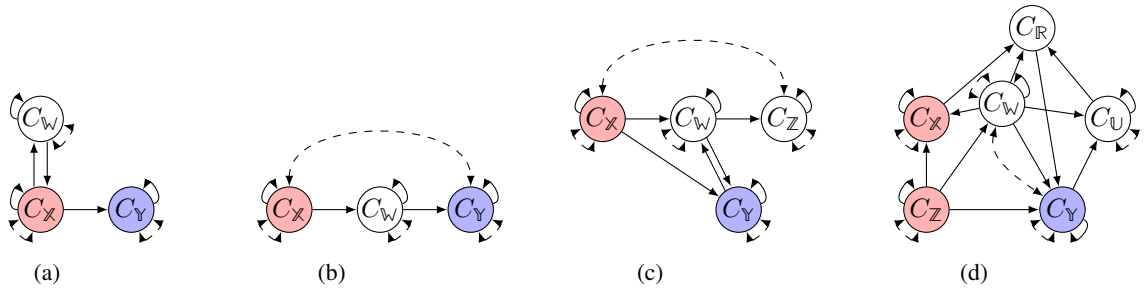


Figure 2: C-DMGs with identifiable macro causal effects. Each pair of red and blue vertices represents the causal effect we are interested in.

Theorem 1 (Soundness of σ -separation in C-DMGs over DMGs). *Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG and $\mathbb{C}_X, \mathbb{C}_Y, \mathbb{C}_W$ be disjoint subsets of $\mathbb{C}$. If $\mathbb{C}_X$ and $\mathbb{C}_Y$ are σ -separated by $\mathbb{C}_W$ in $\mathcal{G}^c$ then, in any compatible DMG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$, $\mathbb{X} = \bigcup_{C \in \mathbb{C}_X} C$ and $\mathbb{Y} = \bigcup_{C \in \mathbb{C}_Y} C$ are σ -separated by $\mathbb{W} = \bigcup_{C \in \mathbb{C}_W} C$.*

Theorem 1 establishes that σ -separation in C-DMGs over DMGs ensures the existence of a corresponding macro-level σ -separation across all compatible DMGs. According to [Forré and Mooij, 2020, Theorem 5.2], this implies that some conditional independencies in the underlying probability distribution can be inferred directly from the C-DMG. By extending the applicability of σ -separation to C-DMGs over DMGs, this result enables the identification of macro-level conditional independencies even when the underlying causal structure is only partially specified. To illustrate the practical value of this result, we now present two examples demonstrating the application of σ -separation in a C-DMG over DMGs.

Example 1. *Let $\mathcal{G}$ be the true unknown DMG and consider that its compatible C-DMG, denoted as $\mathcal{G}^c$ is one given in Figure 2a. Using Definition 8, we can directly deduce $(C_W \perp_{\sigma} C_Y \mid C_X)_{\mathcal{G}^c}$. Thus according to Theorem 1 and [Forré and Mooij, 2020, Theorem 5.2], C_W is conditionally independent of C_Y given C_X in every distribution compatible with the true ADMG.*

Example 2. *Let $\mathcal{G}$ be the true unknown DMG and consider that its compatible C-DMG, denoted as $\mathcal{G}^c$ is one given in Figure 2c. Using Definition 8, we can directly deduce that $(C_Z \perp_{\sigma} C_Y \mid C_X, C_W)_{\mathcal{G}^c}$. Thus according to Theorem 1 and [Forré and Mooij, 2020, Theorem 5.2], C_Z is conditionally independent of C_Y given C_X and C_W in every distribution compatible with the true ADMG.*

The following theorem shows that σ -separation is also complete in C-DMGs over DMGs.

Theorem 2 (Completeness of σ -separation in C-DMGs). *Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG, $\mathbb{C}_X, \mathbb{C}_Y, \mathbb{C}_W$ be disjoint subsets of $\mathbb{C}$, $\mathbb{X} = \bigcup_{C \in \mathbb{C}_X} C$, $\mathbb{Y} = \bigcup_{C \in \mathbb{C}_Y} C$ and $\mathbb{W} = \bigcup_{C \in \mathbb{C}_W} C$. If $\mathbb{C}_X$ and $\mathbb{C}_Y$ are not σ -separated by $\mathbb{C}_W$ in $\mathcal{G}^c$, then there exists a compatible DMG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ such that $\mathbb{X}$ and $\mathbb{Y}$ are not σ -separated by $\mathbb{W}$.*

The findings of Theorems 1 and 2 establish that identifying a σ -separation in C-DMGs over DMGs ensures the recovery of all common macro-level σ -separations across all compatible DMGs. This result is particularly valuable for constraint-based causal discovery methods, especially when the goal is to infer the structure of a C-DMG without needing to fully specify an underlying DMG. Most importantly, these insights lay the theoretical groundwork for the results developed in the next subsection.

3.2 The do-calculus in C-DMGs over DMGs

[Pearl, 1995] introduced an important tool in causal reasoning referred to as the do-calculus. This do-calculus consists of three rules each relying on some d-separation in the ADMG to guarantee an equality between different probabilities. Every one of these three rules can be interpreted differently: the first one allows the insertion or deletion of an observation, the second one allows for the exchange between actions and observations and the third one allows the insertion or deletion of actions. The do-calculus in ADMGs is complete and thus allows, whenever it is possible, to identify causal effects. In other words, it allows, whenever it is possible, to express a causal effect containing a $\text{do}(\cdot)$ operator as a probabilistic expression without any $\text{do}(\cdot)$ operator and thus allows to compute it from a positive observational distribution. Since the do-calculus was initially introduced for ADMGs, it is not easily

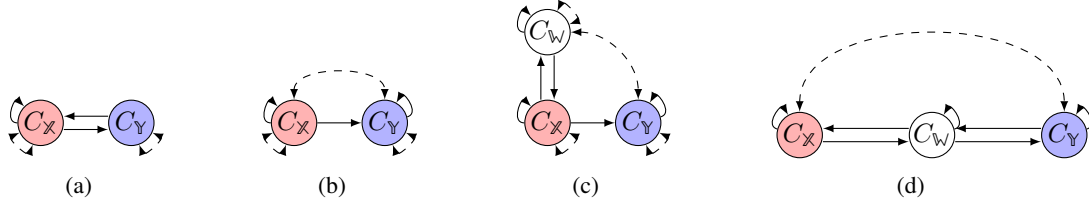


Figure 3: C-DMGs with not identifiable macro causal effects. Each pair of red and blue vertices represents the total effect we are interested in.

extendable to cyclic graphs. However, Forré and Mooij [2020] showed that replacing d-separation by σ -separation in the three rules allows the do-calculus to be applicable on DMGs induced by ioSCMs. In this subsection, we show that this version of the do-calculus is also readily applicable to C-DMGs over DMGs.

There exists multiple equivalent ways of writing the do-calculus rules, for example [Pearl, 2009] uses the notion of mutilated graph to define the three rules of do-calculus. In this paper, we will follow the notation used in Forré and Mooij [2020]. To do so, we will extend every graph $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$, by adding an intervention vertex I_X and the edge $I_X \rightarrow X$ for every vertex of the graph $X \in \mathbb{V}^*$. Moreover, we will use the σ -separation notation with a $\text{do}(\cdot)$ operator (e.g., $(A \perp_{\sigma} B \mid C, \text{do}(D))_{\mathcal{G}^*}$) to place ourselves in the intervened graph *i.e.*, where all arrows going in D are deleted and in which D is conditioned on¹. More formal definitions of extended graphs and intervened graphs can be found in the appendix. Using these newly defined notations and σ -separation we can now state the rules of do-calculus and show their applicability to C-DMGs over DMGs.

Theorem 3 (do-calculus for C-DMGs over DMGs and macro causal effects). *Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG over DMGs and $\mathbb{C}_X, \mathbb{C}_Y, \mathbb{C}_Z, \mathbb{C}_W$ be disjoint subsets of $\mathbb{C}$. The three following rules of the do-calculus are sound.*

$$\textbf{Rule 1: } \Pr(c_Y \mid \text{do}(c_Z), c_X, c_W) = \Pr(c_Y \mid \text{do}(c_Z), c_W) \\ \text{if } (\mathbb{C}_Y \perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$$

$$\textbf{Rule 2: } \Pr(c_Y \mid \text{do}(c_Z), \text{do}(c_X), c_W) = \Pr(c_Y \mid \text{do}(c_Z), c_X, c_W) \\ \text{if } (\mathbb{C}_Y \perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$$

$$\textbf{Rule 3: } \Pr(c_Y \mid \text{do}(c_Z), \text{do}(c_X), c_W) = \Pr(c_Y \mid \text{do}(c_Z), c_W) \\ \text{if } (\mathbb{C}_Y \perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$$

Now that the rules of do-calculus have been stated for C-DMGs over DMGs and their soundness proven, one can use them sequentially to identify the causal effect $\Pr(c_Y \mid \text{do}(c_X))$ in all C-DMGs over DMGs in Figure 2. That is to say, to express the causal effect as a probabilistic expression without any $\text{do}(\cdot)$ operator. This is done in the following examples.

Example 3. Both in Figure 2a and 2c, one can verify that $(\mathbb{C}_Y \perp_{\sigma} I_{C_X} \mid \mathbb{C}_X)_{\mathcal{G}^c}$, thus Rule 2 of the do-calculus is applicable and $\Pr(c_Y \mid \text{do}(c_X)) = \Pr(c_Y \mid c_X)$.

Example 4. Notice that Figure 2b does not contain any cycle other than self-loops and is very similar to Figure 1(b) of Anand et al. [2023] which corresponds to the well-known front-door criterion [Pearl, 2009]. [Forré and Mooij, 2018] have shown that in the acyclic case, σ -separation coincides with d-separation. Thus, using the corresponding sequence of classical rules of probability and rules of do-calculus as the one given in [Pearl, 2009, p.83], one obtains $\Pr(c_Y \mid \text{do}(c_X)) = \sum_{c_W} \Pr(c_W \mid c_X) \sum_{c_{X'}} \Pr(c_Y \mid c_W, c_{X'}) \Pr(c_{X'})$.

Example 5. Consider the C-DMG over DMGs in Figure 2d containing a cycle between $\mathbb{C}_Y, \mathbb{C}_R$, and $\mathbb{C}_U$ and a hidden confounding between $\mathbb{C}_Y$ and $\mathbb{C}_W$. Let $\Pr(c_Y \mid \text{do}(c_X, c_Z))$ be the causal effect of interest. Using the rule of total probability we can rewrite $\Pr(c_Y \mid \text{do}(c_X, c_Z))$ as

$$\sum_{c_W} \Pr(c_Y \mid \text{do}(c_X, c_Z), c_W) \Pr(c_W \mid \text{do}(c_X, c_Z)).$$

¹In other words we write $(A \perp_{\sigma} B \mid C, \text{do}(D))_{\mathcal{G}^*}$ to mean, in Pearl [2009]’s notation, $(A \perp_{\sigma} B \mid C, D)_{\mathcal{G}_D^*}$ where $\mathcal{G}_D^*$ is obtained from $\mathcal{G}^*$ by removing every edge going in D .

We first focus on $\Pr(c_y | do(c_x, c_z), c_w)$. Notice that $(C_{\mathbb{Y} \perp \sigma} I_{C_{\mathbb{Z}}} | C_{\mathbb{X}}, C_{\mathbb{W}}, do(C_{\mathbb{Z}}))_{\mathcal{G}^c}$ and that $(C_{\mathbb{Y} \perp \sigma} I_{C_{\mathbb{Z}}} | C_{\mathbb{Z}}, C_{\mathbb{X}}, C_{\mathbb{W}})_{\mathcal{G}^c}$ which means using two consecutive applications of Rule 2 we can rewrite $\Pr(c_y | do(c_x, c_z), c_w)$ as: $\Pr(c_y | c_z, c_x, c_w)$.

Now we focus on $\Pr(c_w | do(c_x, c_z))$. Notice that $(C_{\mathbb{W} \perp \sigma} I_{C_{\mathbb{X}}} | do(C_{\mathbb{Z}}))_{\mathcal{G}^c}$ which means by Rule 3 of the do-calculus we can completely remove $do(c_x)$ from the expression. Furthermore, we have $(C_{\mathbb{W} \perp \sigma} I_{C_{\mathbb{Z}}} | C_{\mathbb{Z}})_{\mathcal{G}^c}$ which means by Rule 2 we can replace $do(c_z)$ by c_z . So we can rewrite $\Pr(c_w | do(c_x, c_z))$ as $\Pr(c_w | c_z)$.

These three examples show how the rules of do-calculus in C-DMGs on DMGs can be used to write macro causal effects as an expression of observed probabilities. Therefore, the macro causal effects in these C-DMGs can be estimated from the observational data, provided there is no further issues in the data (e.g., positivity violations).

In the following theorem, we show that not only is the do-calculus applicable in C-DMGs over DMGs, but it is also complete.

Theorem 4 (Completeness of do-calculus for C-DMGs and macro causal effects). *If one of the do-calculus rules does not apply for a given C-DMG over DMGs, then there exists a compatible DMG for which the corresponding rule does not apply.*

The proof of Theorem 4 relies on a specific compatible DMG called the maximal compatible DMG which is properly defined in the appendix, therefore it can be reformulated as: if one of the do-calculus rules does not apply for a given C-DMG over DMGs, then this same rule does not apply for its maximal compatible DMG.

Note that this completeness result links C-DMGs to the underlying compatible DMGs, however it does not guarantee the absence of other rules to identify causal effects. While the do-calculus based on d-separation in ADMGs introduced in [Pearl, 1995] has been proven to be complete [Shpitser and Pearl, 2006, Huang and Valtorta, 2006], the do-calculus based on σ -separation in DMGs is not yet proven to be complete and thus our results suffer the same limitation. Using the completeness of the do-calculus in C-DMGs (Theorem 4), one can be convinced, by going through every possible sequence of rules of the do-calculus, that the causal effects of interest in every C-DMGs depicted in Figure 1 and 3 is not identifiable. Unfortunately, going through every possible sequence of rules of the do-calculus is time-consuming and can become impractical when considering larger graphs. In an effort to solve this issue, the following subsection introduces a sub-graphical structure which allows to recognize more efficiently when a causal effect is not identifiable using the do-calculus.

3.3 Non-Identifiability: a graphical characterization

In ADMGs, there exists a sub-graphical structure, called a hedge [Shpitser and Pearl, 2006], which is employed to graphically characterize non-identifiability as shown in Shpitser and Pearl 2006, Theorem 4. This graphical criterion is complete when considering ADMGs, that is to say, if the total effect of X on Y is not identifiable in an ADMG then there exists a hedge for (X, Y) [Shpitser and Pearl, 2006]. However, it has been shown that this characterization is too weak to characterize all non-identifiabilities in the case of C-DMGs over ADMGs [Ferreira and Assaad, 2025b]. Thus, a modified version of the hedge structure called the SC-hedge (strongly connected hedge) has been introduced in Ferreira and Assaad [2025a]. The presence of such SC-hedge in C-DMGs over ADMGs guarantees non-identifiability [Ferreira and Assaad, 2025a,b]. In this subsection, we will show that SC-hedge can also be used to characterize non-identifiability in C-DMGs over DMGs in specific conditions. First, let us recall some useful concepts to define a hedge.

Definition 9 (C-component, Tian and Pearl [2002]). *Let $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$ be a graph. A subset of vertices $\mathbb{V}_C^* \subseteq \mathbb{V}^*$ such that $\forall V_1^*, V_n^* \in \mathbb{V}_C^*, \exists V_1^*, \dots, V_n^* \in \mathbb{V}^*$ with $\forall 1 \leq i < n, V_i^* \leftrightarrow V_{i+1}^*$ is called a C-component.*

Definition 10 (C-forest, Shpitser and Pearl [2006]). *Let $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$ be a graph. If $\mathcal{G}^*$ is acyclic, $\mathcal{G}^*$ is a forest (i.e., every of its vertices has at most one child), and $\mathcal{G}^*$ is a C-component then $\mathcal{G}^*$ is called a C-forest. The vertices which have no children are called roots and we say a C-forest is $\mathbb{R}^*$ -rooted if it has roots $\mathbb{R}^* \subseteq \mathbb{V}^*$.*

Definition 11 (Hedge, Shpitser and Pearl [2006, 2008]). *Consider a graph $\mathcal{G}^* = (\mathbb{V}^*, \mathbb{E}^*)$ and two disjoint sets of vertices $\mathbb{X}^*, \mathbb{Y}^* \subseteq \mathbb{V}^*$. Let $\mathcal{F} = (\mathbb{V}_{\mathcal{F}}^*, \mathbb{E}_{\mathcal{F}}^*)$ and $\mathcal{F}' = (\mathbb{V}_{\mathcal{F}'}^*, \mathbb{E}_{\mathcal{F}'}^*)$ be two $\mathbb{R}^*$ -rooted*

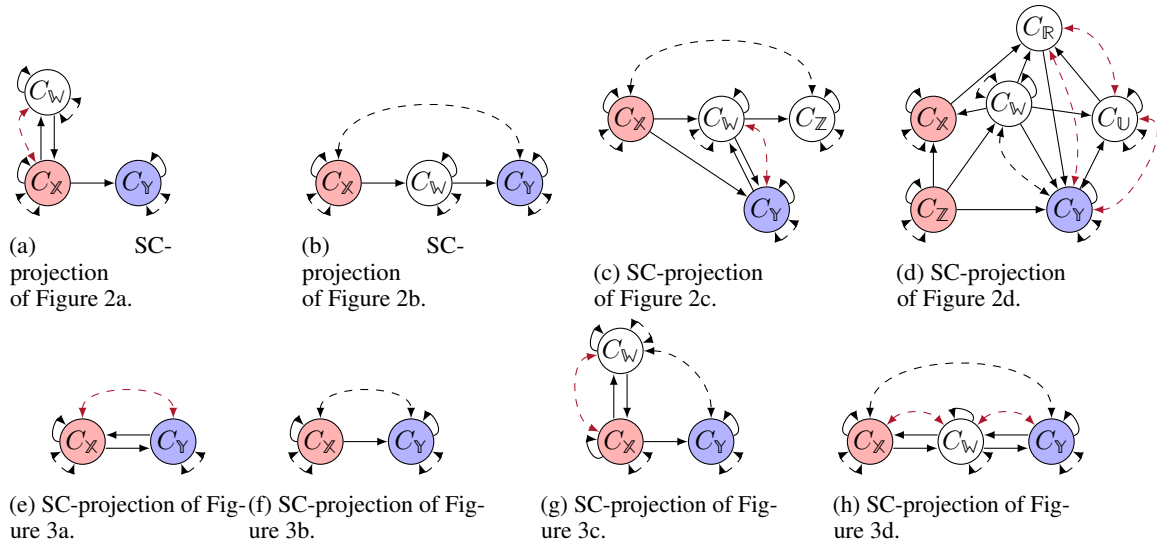


Figure 4: SC-projections of the C-DMGs in Figures 1, 2, and Figures 3. Each pair of red and blue vertices represents the total effect we are interested in, and the red edges indicate those added through the SC-projection.

C-forests subgraphs of $\mathcal{G}^*$ such that $\mathbb{X}^* \cap \mathbb{V}_{\mathcal{F}}^* \neq \emptyset$, $\mathbb{X}^* \cap \mathbb{V}_{\mathcal{F}'}^* = \emptyset$, $\mathbb{F}' \subseteq \mathbb{F}$, and $\mathbb{R}^* \subset \text{An}(\mathbb{Y}^*, \mathcal{G}^* \setminus \mathbb{X}^*)$. Then $\mathbb{F}$ and $\mathbb{F}'$ form a hedge for the pair $(\mathbb{X}^*, \mathbb{Y}^*)$ in $\mathcal{G}^*$.

As mentioned before, a hedge turned out to be too weak to cover non-identifiability in C-DMGs [Ferreira and Assaad, 2025b]. For example, the C-DMG in Figure 3a contains no hedge but the macro causal effect is not identifiable due to the cycle between $\mathbb{C}_X$ and $\mathbb{C}_Y$. In the following, we formally define SC-hedges in the context of C-DMGs over DMGs and demonstrate that this substructure serves as a sound criterion for detecting non-identifiable macro causal effects if every cluster in a cycle is of size strictly greater than 1.

Definition 12 (Strongly connected projection (SC-projection)). Consider a C-DMG $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$. The SC-projection $\mathcal{H}^c$ of $\mathcal{G}^c$ is the graph that includes all vertices and edges from $\mathcal{G}^c$, plus a dashed bidirected edge between each pair $\mathbb{C}_X, \mathbb{C}_Y \in \mathbb{C}$ such that $\text{Scc}(\mathbb{C}_X, \mathcal{G}^c) = \text{Scc}(\mathbb{C}_Y, \mathcal{G}^c)$ and $\mathbb{C}_X \neq \mathbb{C}_Y$.

Definition 13 (Strongly Connected Hedge (SC-Hedge)). Consider an C-DMG $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$, its SC-projection $\mathcal{H}^c$ and two disjoint sets of vertices $\mathbb{C}_X, \mathbb{C}_Y \subseteq \mathbb{C}$. A hedge for $(\mathbb{C}_X, \mathbb{C}_Y)$ in $\mathcal{H}^c$ is an SC-hedge for $(\mathbb{C}_X, \mathbb{C}_Y)$ in $\mathcal{G}^c$.

The following theorem guarantees the soundness of the SC-hedge criterion in C-DMGs over DMGs when every cluster in a cycle is of size strictly greater than 1. This assumption is useful as it allows the existence of compatible ADMGs even when the C-DMG over DMGs contains cycles. Thus, one can use Ferreira and Assaad 2025b, Theorem 5 in the presence of an SC-hedge to show that for every identifying sequence of do-calculus rules, there exists a compatible ADMG in which this sequence is not applicable.

Theorem 5. Consider an C-DMG $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ such that every cluster which is in a cycle is of size at least 2 and two disjoint sets of vertices $\mathbb{C}_X, \mathbb{C}_Y \subseteq \mathbb{C}$. If there exists an SC-hedge for $(\mathbb{C}_X, \mathbb{C}_Y)$ in $\mathcal{G}^c$ then $\Pr(\mathbb{C}_Y \mid \text{do}(\mathbb{C}_X))$ is not identifiable.

The SC-projections of the C-DMGs illustrated in Figures 2 and 3 are given in Figure 4. One can notice that the SC-projections of the C-DMGs in Figure 3 all contain a hedge, thus the C-DMGs in Figure 3 contain a SC-hedge and the causal effect of interest is therefore not identifiable according to Theorem 5. In contrast, the projections of the C-DMGs in Figure 2 do not contain a hedge, this highlights the usefulness of SC-hedges for causal effect identification.

4 Conclusion

In this paper, we established the soundness and completeness of σ -separation and the do-calculus using σ -separation for identifying macro causal effects in C-DMGs over DMGs. There are two main limitations to this work. The first limitation is that the completeness result in Theorem 4 does not take into account that there might exist different sequences of rules of the do-calculus in different DMGs that can give the same final identification of the causal effect. Moreover, the completeness results only link the cluster graphs to the compatible underlying graphs, our results say nothing on the completeness of σ -separation and do-calculus in DMGs. A second related limitation is that we provided a graphical characterization for the non-identifiability of macro causal effects, however this characterization is not proven to be complete, even though we did not find any counter-example of its completeness. Proving it to be complete remains an open problem.

Acknowledgments and Disclosure of Funding

This work was supported by the CIPHOD project (ANR-23-CPJ1-0212-01).

References

- Tara V. Anand, Adele H. Ribeiro, Jin Tian, and Elias Bareinboim. Causal effect identification in cluster dags. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(10):12172–12179, Jun. 2023. doi: 10.1609/aaai.v37i10.26435.
- Charles K. Assaad. Towards identifiability of micro total effects in summary causal graphs with latent confounding: extension of the front-door criterion. *Transactions on Machine Learning Research*, 2025. ISSN 2835-8856.
- Charles K. Assaad, Imad Ez-Zejjari, and Lei Zan. Root cause identification for collective anomalies in time series given an acyclic summary causal graph with loops. In Francisco Ruiz, Jennifer Dy, and Jan-Willem van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 8395–8404. PMLR, 25–27 Apr 2023.
- Charles K. Assaad, Emilie Devijver, Eric Gaussier, Gregor Goessler, and Anouar Meynaoui. Identifiability of total effects from abstractions of time series causal graphs. In *Proceedings of the Fourtieth Conference on Uncertainty in Artificial Intelligence*, Proceedings of Machine Learning Research. PMLR, 2024.
- Philip Boeken and Joris M. Mooij. Dynamic structural causal models, 2024. URL <https://arxiv.org/abs/2406.01161>.
- Simon Ferreira and Charles K. Assaad. Identifiability of direct effects from summary causal graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(18):20387–20394, Mar. 2024. doi: 10.1609/aaai.v38i18.30021.
- Simon Ferreira and Charles K. Assaad. Identifying macro conditional independencies and macro total effects in summary causal graphs with latent confounding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39, 2025a.
- Simon Ferreira and Charles K. Assaad. Identifying macro causal effects in C-DMGs over ADMGs. *Transactions on Machine Learning Research*, 2025b.
- Patrick Forré and Joris M. Mooij. Causal calculus in the presence of cycles, latent confounders and selection bias. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 71–80. PMLR, 22–25 Jul 2020.
- Patrick Forré and Joris M. Mooij. Markov properties for graphical models with cycles and latent variables. *arXiv: Statistics Theory*, 2017.
- Patrick Forré and Joris M. Mooij. Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders, 2018. URL <https://arxiv.org/abs/1807.03024>.

- Yimin Huang and Marco Valtorta. Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, page 217–224, Arlington, Virginia, USA, 2006. AUAI Press. ISBN 0974903922.
- Amin Jaber, Adele Ribeiro, Jiji Zhang, and Elias Bareinboim. Causal identification under markov equivalence: Calculus, algorithm, and completeness. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 3679–3690. Curran Associates, Inc., 2022.
- Marloes Maathuis and Diego Colombo. A generalized backdoor criterion. *The Annals of Statistics*, 43, 07 2013. doi: 10.1214/14-AOS1295.
- Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. ISBN 1558604790.
- Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 12 1995. ISSN 0006-3444. doi: 10.1093/biomet/82.4.669.
- Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition, 2009. ISBN 052189560X.
- Emilija Perkovic. Identifying causal effects in maximally oriented partially directed acyclic graphs. In Jonas Peters and David Sontag, editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 530–539. PMLR, 03–06 Aug 2020.
- Emilija Perkovic, Johannes Textor, Markus Kalisch, and Marloes H. Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *J. Mach. Learn. Res.*, 18:220:1–220:62, 2016.
- Nicolas-Domenic Reiter, Jonas Wahl, Andreas Gerhardus, and Jakob Runge. Causal inference on process graphs, part ii: Causal structure and effect identification, 2024.
- Thomas Richardson. A characterization of markov equivalence for directed cyclic graphs. *International Journal of Approximate Reasoning*, 17(2):107–162, 1997. ISSN 0888-613X. doi: [https://doi.org/10.1016/S0888-613X\(97\)00020-0](https://doi.org/10.1016/S0888-613X(97)00020-0). Uncertainty in AI (UAI'96) Conference.
- Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, page 1219–1226. AAAI Press, 2006. ISBN 9781577352815.
- Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, 9:1941–1979, 2008.
- Jin Tian and Judea Pearl. On the testable implications of causal models with hidden variables. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02, page 519–527, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc. ISBN 1558608974.
- Santtu Tikka, Jouni Helske, and Juha Karvanen. Clustering and structural robustness in causal diagrams. *J. Mach. Learn. Res.*, 24(1), January 2023. ISSN 1532-4435.
- Jonas Wahl, Urmi Ninad, and Jakob Runge. Foundations of causal discovery on groups of variables. *Journal of Causal Inference*, 12(1):20230041, 2024. doi: doi:10.1515/jci-2023-0041.
- Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Estimating possible causal effects with latent variables via adjustment. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 36308–36335. PMLR, 23–29 Jul 2023.
- Clement Yvernes. Note on the identification of total effect in cluster-dags with cycles. Unpublished note, 2025.

A Appendix

A.1 Additional Notations and Properties

Firstly, the proofs require the definitions of intervened graphs and extended graphs which are the graphs induced respectively by intervened ioSCMs and extended ioSCMs Forré and Mooij [2020].

Definition 14 (Intervened graph). *Let $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ be a DMG induced from an ioSCM $\mathcal{M}$ and $\mathbb{A} \subseteq \mathbb{V}$ a set of variables. The intervened graph $(\mathcal{G}_{|\text{do}(\mathbb{A})})$ is obtained by removing all edges going to $\mathbb{A}$ in $\mathcal{G}$.*

The intervened graph $(\mathcal{G}_{|\text{do}(\mathbb{A})})$ is also written as $\mathcal{G}_{\bar{\mathbb{A}}}$ in [Pearl, 2009].

Definition 15 (Extended graph). *Let $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ be a DMG induced from an ioSCM $\mathcal{M}$. The extended graph $[\mathcal{G}] = ([\mathbb{V}], [\mathbb{E}])$ is obtained by adding for each vertex $V \in \mathbb{V}$ an extended vertex $I_V \in [\mathbb{V}]$ and the edge $I_V \rightarrow V \in [\mathbb{E}]$. i.e., $[\mathbb{V}] = \mathbb{V} \cup \{I_V \mid V \in \mathbb{V}\}$ and $[\mathbb{E}] = \mathbb{E} \cup \{I_V \rightarrow V \mid V \in \mathbb{V}\}$.*

In order to map the vertices in a C-DMG $\mathcal{G}^c$ with the vertices in a compatible DMG $\mathcal{G}$, we will use the notion of corresponding cluster: $\forall C \in \mathbb{C}, \forall V \in C, Cl(V, \mathcal{G}^c) = C$. Additionally, we will write for every set of clusters $\mathbb{C}_{\mathbb{A}} \subseteq \mathbb{C}$, $\bar{\mathbb{A}} = \bigcup_{C \in \mathbb{C}_{\mathbb{A}}} C$.

Definition 16 (Maximal compatible DMG). *Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG. Let us define the following sets:*

$$\begin{aligned}\mathbb{V} &:= \bigcup_{C \in \mathbb{C}} C \\ \mathbb{E}^{\rightarrow} &:= \{V \rightarrow V' \mid \forall V \in C, V' \in C' \text{ such that } C \rightarrow C' \in \mathbb{E}^c\} \\ \mathbb{E}^{\leftrightarrow} &:= \{V \leftrightarrow V' \mid \forall V \in C, V' \in C' \text{ such that } C \leftrightarrow C' \in \mathbb{E}^c\} \\ \mathbb{E} &:= \mathbb{E}^{\rightarrow} \cup \mathbb{E}^{\leftrightarrow}\end{aligned}$$

The graph $\mathcal{G}_m^c = (\mathbb{V}, \mathbb{E})$ is compatible with $\mathcal{G}^c$ and for every compatible DMG $\mathcal{G}$, we have $\mathcal{G} \subseteq \mathcal{G}_m^c$, thus $\mathcal{G}_m^c$ is called the maximal compatible DMG of $\mathcal{G}^c$.

Property 1 (Compatibility of extended graphs). *Let $\mathcal{G}^c$ be a C-DMG and $\mathcal{G}$ be a compatible DMG. Let us consider $[\mathcal{G}^c]$ and $[\mathcal{G}]$ the corresponding extended graphs. Take $[\mathbb{V}] = \mathbb{V} \cup \{I_V \mid V \in \mathbb{V}\}$ the extended micro variables and $[\mathbb{C}] = \mathbb{C} \cup \{\{I_V \mid V \in C\} \mid C \in \mathbb{C}\}$ the extended partition. $[\mathcal{G}]$ is compatible with $[\mathcal{G}^c]$ according to partition $[\mathbb{C}]$.*

Proof. Firstly, $\forall V, V' \in \mathbb{V}, V \rightarrow V' \in [\mathcal{G}]$ (resp. $\leftrightarrow$) $\iff V \rightarrow V' \in \mathcal{G}$ (resp. $\leftrightarrow$) $\iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in \mathcal{G}^c$ (resp. $\leftrightarrow$) $\iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in [\mathcal{G}^c]$ (resp. $\leftrightarrow$). Secondly, $\forall V \in \mathbb{V}, I_V \rightarrow V \in [\mathcal{G}]$ and $I_{Cl(V, \mathcal{G}^c)} \rightarrow Cl(V, \mathcal{G}^c) \in [\mathcal{G}^c]$. Lastly, $\forall C \in \mathbb{C}, I_C \rightarrow C \in [\mathcal{G}^c]$ and $\exists V \in C, I_V \rightarrow V \in [\mathcal{G}]$ because $\mathbb{C}$ is a partition so $\forall C \in \mathbb{C}, C \neq \emptyset$. $\square$

Property 2 (Extended maximal compatible graphs). *Let $\mathcal{G}^c$ be a C-DMG, $\mathcal{G}_m^c$ be the maximal compatible graph of $\mathcal{G}^c$, $[\mathcal{G}_m^c]$ be the extended graph of $\mathcal{G}_m^c$ and $[\mathcal{G}^c]_m$ be the maximal compatible graph of $[\mathcal{G}^c]$. These graphs verify*

- $[\mathcal{G}_m^c] \subseteq [\mathcal{G}^c]_m$, and
- $\forall V \in [\mathbb{V}], Scc(V, [\mathcal{G}_m^c]) = Scc(V, [\mathcal{G}^c]_m)$.

Therefore, any macro-level σ -connection that holds in $[\mathcal{G}^c]_m$ also holds in $[\mathcal{G}_m^c]$.

Proof. Firstly, $\forall V, V' \in \mathbb{V}, V \rightarrow V' \in [\mathcal{G}_m^c]$ (resp. $\leftrightarrow$) $\iff V \rightarrow V' \in \mathcal{G}_m^c$ (resp. $\leftrightarrow$) $\iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in \mathcal{G}^c$ (resp. $\leftrightarrow$) $\iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in [\mathcal{G}^c]$ (resp. $\leftrightarrow$) $\iff V \rightarrow V' \in [\mathcal{G}^c]_m$ (resp. $\leftrightarrow$). Secondly, $\forall V \in \mathbb{V}, I_V \rightarrow V \in [\mathcal{G}_m^c]$ and $I_V \in Cl(I_V, [\mathcal{G}^c]), V \in Cl(V, [\mathcal{G}^c]), Cl(I_V, [\mathcal{G}^c]) \rightarrow Cl(V, [\mathcal{G}^c]) \in [\mathcal{G}^c]$ thus $I_V \rightarrow V \in [\mathcal{G}^c]_m$.

Regarding the strongly connected components, $\forall V \in \mathbb{V}, Scc(V, [\mathcal{G}_m^c]) = Scc(V, \mathcal{G}_m^c) = \{V\} \cup \left(\bigcup_{C \in Scc(Cl(V, \mathcal{G}^c), \mathcal{G}^c)} C \right)^2 = \{V\} \cup \left(\bigcup_{C \in Scc(Cl(V, [\mathcal{G}^c]), [\mathcal{G}^c])} C \right) = Scc(V, [\mathcal{G}^c]_m)$. Moreover, $\forall V \in$

²While the usual strongly connected component, $Scc(V^*, \mathcal{G}^*)$, always contains the vertex V^* itself, this tweaked version, $Scc(V^*, \mathcal{G}^*)$, contains the vertex V^* if and only if there exists a self-loop on it i.e., $V^* \rightarrow V^* \in \mathcal{G}^*$.

$\mathbb{V}$, the set of edges including I_V in $[\mathcal{G}_m^c]$ is $\{I_V \rightarrow V\}$ and the set of edges including I_V in $[\mathcal{G}^c]_m$ is $\{I_V \rightarrow V' \mid V' \in Cl(V, \mathcal{G}^c)\}$ thus $Scc(I_V, [\mathcal{G}_m^c]) = \{I_V\} = Scc(I_V, [\mathcal{G}^c]_m)$.

In conclusion, using the Definition 8, any macro-level σ -connection that holds in $[\mathcal{G}^c]_m$ also holds in $[\mathcal{G}_m^c]$. $\square$

Property 3 (Compatibility of intervened graphs). *Let $\mathcal{G}^c$ be a C-DMG and $\mathcal{G}$ be a compatible DMG. Let $\mathbb{C}_\mathbb{A} \subseteq \mathbb{C}$ be a set of clusters. Let us consider $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})$ and $(\mathcal{G}|_{do(\mathbb{A})})$ the graphs obtained respectively by intervening on $\mathbb{C}_\mathbb{A}$ in $\mathcal{G}^c$ and by intervening on $\mathbb{A}$ in $\mathcal{G}$. $(\mathcal{G}|_{do(\mathbb{A})})$ is compatible with $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})$.*

Moreover, let $\mathcal{G}_m^c$ be the maximal compatible graph of $\mathcal{G}^c$, $(\mathcal{G}_m^c|_{do(\mathbb{A})})$ its intervened graph and $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})_m$ be the maximal compatible graph of $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})$. $(\mathcal{G}_m^c|_{do(\mathbb{A})})$ and $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})_m$ are the same graph.

Proof. $\forall V, V' \in \mathbb{V}$, $V \rightarrow V' \in (\mathcal{G}|_{do(\mathbb{A})})$ (resp. $\leftrightarrow$) $\iff V \rightarrow V' \in \mathcal{G}$ (resp. $\leftrightarrow$) and $V' \notin \mathbb{A} \iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in \mathcal{G}^c$ (resp. $\leftrightarrow$) and $Cl(V', \mathcal{G}^c) \notin \mathbb{C}_\mathbb{A} \iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in (\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})$ (resp. $\leftrightarrow$). $\square$

Property 4 (Intervened maximal compatible graphs). *Let $\mathcal{G}^c$ be a C-DMG, $\mathcal{G}_m^c$ be the maximal compatible graph of $\mathcal{G}^c$, $(\mathcal{G}_m^c|_{do(\mathbb{A})})$ be the intervened graph of $\mathcal{G}_m^c$ and $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})_m$ be the maximal compatible graph of $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})$. $(\mathcal{G}_m^c|_{do(\mathbb{A})})$ and $(\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})_m$ are the same graph.*

Proof. $\forall V, V' \in \mathbb{V}$, $V \rightarrow V' \in (\mathcal{G}_m^c|_{do(\mathbb{A})})$ (resp. $\leftrightarrow$) $\iff V \rightarrow V' \in \mathcal{G}_m^c$ (resp. $\leftrightarrow$) and $V' \notin \mathbb{A} \iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in \mathcal{G}^c$ (resp. $\leftrightarrow$) and $Cl(V', \mathcal{G}^c) \notin \mathbb{C}_\mathbb{A} \iff Cl(V, \mathcal{G}^c) \rightarrow Cl(V', \mathcal{G}^c) \in (\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})$ (resp. $\leftrightarrow$) $\iff V \rightarrow V' \in (\mathcal{G}^c|_{do(\mathbb{C}_\mathbb{A})})_m$ (resp. $\leftrightarrow$). $\square$

A.2 Proof of Theorem 1

Proof. Suppose $\mathbb{C}_\mathbb{X}$ and $\mathbb{C}_\mathbb{Y}$ are σ -separated by $\mathbb{C}_\mathbb{W}$ in $\mathcal{G}^c$ and there exists a compatible DMG $\mathcal{G} = (\mathbb{V}, \mathbb{E})$ and a walk $\pi = \langle V_1, \dots, V_n \rangle$ in $\mathcal{G}$ from $V_1 \in \mathbb{X}$ to $V_n \in \mathbb{Y}$ which is not σ -blocked by $\mathbb{W}$. Consider the walk $\tilde{\pi} = \langle C_1, \dots, C_n \rangle$ with $\forall 1 \leq i \leq n$, $C_i = Cl(V_i, \mathcal{G}^c)$ and $\forall 1 \leq i < n$, $\langle C_i \rightarrow C_{i+1} \rangle \subseteq \tilde{\pi}$ (resp. $\leftarrow, \leftrightarrow$) $\iff \langle V_i \rightarrow V_{i+1} \rangle \subseteq \pi$ (resp. $\leftarrow, \leftrightarrow$). $\tilde{\pi}$ is a walk from $\mathbb{C}_\mathbb{X}$ to $\mathbb{C}_\mathbb{Y}$ in $\mathcal{G}^c$. Since $\mathbb{C}_\mathbb{X}$ and $\mathbb{C}_\mathbb{Y}$ are σ -separated by $\mathbb{C}_\mathbb{W}$, we know that $\mathbb{C}_\mathbb{W}$ σ -blocks $\tilde{\pi}$.

- If $C_1 \in \mathbb{W}$ or $C_n \in \mathbb{C}_\mathbb{W}$, then $V_1 \in \mathbb{W}$ or $V_n \in \mathbb{W}$ and thus π is σ -blocked by $\mathbb{W}$ which contradicts the initial assumption.

Otherwise, take $1 < i < n$ such that $\langle C_{i-1}, C_i, C_{i+1} \rangle$ is $\mathbb{C}_\mathbb{W}$ - σ -blocked.

- If $\langle C_{i-1} \leftrightarrow C_i \leftrightarrow C_{i+1} \rangle \subseteq \tilde{\pi}$ and $C_i \notin \mathbb{C}_\mathbb{W}$ then, $\langle V_{i-1} \leftrightarrow V_i \leftrightarrow V_{i+1} \rangle \subseteq \pi$ and $V_i \notin \mathbb{W}$. Thus π is σ -blocked by $\mathbb{W}$ which contradicts the initial assumption.
- If $\langle C_{i-1} \leftarrow C_i \leftrightarrow C_{i+1} \rangle \subseteq \tilde{\pi}$ and $C_i \in \mathbb{C}_\mathbb{W} \setminus Scc(C_{i-1}, \mathcal{G}^c)$ then, $\langle V_{i-1} \leftarrow V_i \leftrightarrow V_{i+1} \rangle \subseteq \pi$. Moreover, $Scc(V_{i-1}, \mathcal{G}) \subseteq \bigcup_{C \in Scc(C_{i-1}, \mathcal{G}^c)} C$ and thus $V_i \in \mathbb{W} \setminus Scc(V_{i-1}, \mathcal{G})$. Therefore, π is σ -blocked by $\mathbb{W}$ which contradicts the initial assumption.
- If $\langle C_{i-1} \leftrightarrow C_i \rightarrow C_{i+1} \rangle \subseteq \tilde{\pi}$ and $C_i \in \mathbb{C}_\mathbb{W} \setminus Scc(C_{i+1}, \mathcal{G}^c)$ then, $\langle V_{i-1} \leftrightarrow V_i \rightarrow V_{i+1} \rangle \subseteq \pi$. Moreover, $Scc(V_{i+1}, \mathcal{G}) \subseteq \bigcup_{C \in Scc(C_{i+1}, \mathcal{G}^c)} C$ and thus $V_i \in \mathbb{W} \setminus Scc(V_{i+1}, \mathcal{G})$. Therefore, π is σ -blocked by $\mathbb{W}$ which contradicts the initial assumption.
- If $\langle C_{i-1} \leftarrow C_i \rightarrow C_{i+1} \rangle \subseteq \tilde{\pi}$ and $C_i \in \mathbb{C}_\mathbb{W} \setminus (Scc(C_{i-1}, \mathcal{G}^c) \cap Scc(C_{i+1}, \mathcal{G}^c))$ then, $\langle V_{i-1} \leftarrow V_i \rightarrow V_{i+1} \rangle \subseteq \pi$. Moreover, $(Scc(V_{i-1}, \mathcal{G}) \cap Scc(V_{i+1}, \mathcal{G})) \subseteq \bigcup_{C \in (Scc(C_{i-1}, \mathcal{G}^c) \cap Scc(C_{i+1}, \mathcal{G}^c))} C$ and thus $V_i \in \mathbb{W} \setminus (Scc(V_{i-1}, \mathcal{G}) \cap Scc(V_{i+1}, \mathcal{G}))$. Therefore, π is σ -blocked by $\mathbb{W}$ which contradicts the initial assumption.

In conclusion, the σ -separation is sound in C-DMGs over DMGs. $\square$

A.3 Proof of Theorem 2

Proof. Suppose $\mathbb{C}_X$ and $\mathbb{C}_Y$ are not σ -separated by $\mathbb{C}_W$ in $\mathcal{G}^c$. There exists an $\mathbb{C}_W$ - σ -active path $\pi = \langle C_1, \dots, C_n \rangle$ with $C_1 \in \mathbb{C}_X$ and $C_n \in \mathbb{C}_Y$. Take $\mathcal{G}_m^c = (\mathbb{V}, \mathbb{E})$ the maximal compatible DMG of $\mathcal{G}^c$ as in Definition 16. Take for every cluster $C \in \mathbb{C}$ a representative of this cluster $V_C \in \mathbb{C}$. The maximal compatible graph $\mathcal{G}_m^c$ contains the path $\pi_m = \langle V_{C_1}, \dots, V_{C_n} \rangle$ and for every cluster $C \in \mathbb{C}$ and every variable in that cluster $V \in C$, $Scc(V, \mathcal{G}_m^c) = \bigcup_{C' \in Scc(C, \mathcal{G}^c)} C'$. Therefore, π being $\mathbb{C}_W$ - σ -active in $\mathcal{G}^c$ clearly implies that π_m is $\mathbb{W}$ - σ -active in $\mathcal{G}_m^c$. In conclusion, the σ -separation criterion in C-DMGs over DMGs is complete.

Notice, that not only did we prove Theorem 2—i.e., if a σ -separation does not hold in a C-DMG then there exists a compatible DMG in which the corresponding σ -separation does not hold—but we also explicitly exhibited this compatible DMG as being the maximal compatible DMG. $\square$

A.4 Proof of Theorem 3

Proof. Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG, $\mathcal{G}$ a compatible DMG and $\mathbb{C}_X, \mathbb{C}_Y, \mathbb{C}_Z, \mathbb{C}_W \subseteq \mathbb{C}$ be disjoint subsets of vertices. Suppose a rule of the do-calculus applies in $\mathcal{G}^c$ then Theorem 1, Property 1 and Property 3 guarantees that this rule applies in $\mathcal{G}$. More explicitly:

- If rule 1 applies i.e., $(\mathbb{C}_Y \perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$, then using Theorem 1 as well as Properties 1 and 3 one knows that $(\mathbb{Y} \perp_{\sigma} \mathbb{X} \mid \mathbb{W}, \text{do}(\mathbb{Z}))_{\mathcal{G}}$ and thus rule 1 applies in $\mathcal{G}$.
- If rule 2 applies i.e., $(\mathbb{C}_Y \perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_X, \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$, then using Theorem 1 as well as Properties 1 and 3 one knows that $(\mathbb{Y} \perp_{\sigma} \mathbb{X} \mid \mathbb{X}, \mathbb{W}, \text{do}(\mathbb{Z}))_{\mathcal{G}}$ and thus rule 2 applies in $\mathcal{G}$.
- If rule 3 applies i.e., $(\mathbb{C}_Y \perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$, then using Theorem 1 as well as Properties 1 and 3 one knows that $(\mathbb{Y} \perp_{\sigma} \mathbb{X} \mid \mathbb{W}, \text{do}(\mathbb{Z}))_{\mathcal{G}}$ and thus rule 3 applies in $\mathcal{G}$.

Notice that because $\mathbb{C}_X$ and $\mathbb{C}_Z$ are disjoint, the actions of taking the intervened graph and taking the extended graph can be done in any order without any repercussion in the σ -separations of interest.

In conclusion, the do-calculus using σ -separation is sound in C-DMG over DMGs. $\square$

A.5 Proof of Theorem 4

Proof. Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG, $\mathcal{G}_m^c$ be the maximal compatible DMG and $\mathbb{C}_X, \mathbb{C}_Y, \mathbb{C}_Z, \mathbb{C}_W \subseteq \mathbb{C}$ be disjoint subsets of vertices. Suppose a rule of the do-calculus does not apply in $\mathcal{G}^c$, then Theorem 2, Property 1 and Property 3 show that this rule does not apply in $\mathcal{G}_m^c$. More explicitly:

- If rule 1 does not apply i.e., $(\mathbb{C}_Y \not\perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$, then using Theorem 2 as well as Properties 2 and 4 one knows that $(\mathbb{Y} \not\perp_{\sigma} \mathbb{X} \mid \mathbb{W}, \text{do}(\mathbb{Z}))_{\mathcal{G}_m^c}$ and thus rule 1 does not apply in $\mathcal{G}_m^c$.
- If rule 2 does not apply i.e., $(\mathbb{C}_Y \not\perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_X, \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$, then using Theorem 2 as well as Properties 2 and 4 one knows that $(\mathbb{Y} \not\perp_{\sigma} \mathbb{X} \mid \mathbb{X}, \mathbb{W}, \text{do}(\mathbb{Z}))_{\mathcal{G}_m^c}$ and thus rule 2 does not apply in $\mathcal{G}_m^c$.
- If rule 3 does not apply i.e., $(\mathbb{C}_Y \not\perp_{\sigma} \mathbb{C}_X \mid \mathbb{C}_W, \text{do}(\mathbb{C}_Z))_{\mathcal{G}^c}$, then using Theorem 2 as well as Properties 2 and 4 one knows that $(\mathbb{Y} \not\perp_{\sigma} \mathbb{X} \mid \mathbb{W}, \text{do}(\mathbb{Z}))_{\mathcal{G}_m^c}$ and thus rule 3 does not apply in $\mathcal{G}_m^c$.

Notice that because $\mathbb{C}_X$ and $\mathbb{C}_Z$ are disjoint, considering the extended graph of the intervened graph or considering the intervened graph of the extended graph does not have any repercussion in the σ -connections of interest.

In conclusion, the do-calculus using σ -separation is complete in C-DMG over DMGs. $\square$

A.6 Proof of Theorem 5

Proof. Let $\mathcal{G}^c = (\mathbb{C}, \mathbb{E}^c)$ be a C-DMG and take disjoint subsets $\mathbb{C}_X, \mathbb{C}_Y \subseteq \mathbb{C}$. Additionally, suppose that every cluster which is in a cycle in $\mathcal{G}^c$ is of size at least 2. More formally, $\forall C \in \mathbb{C}, |\text{Scc}(C, \mathcal{G}^c)| > 1 \implies |C| > 1$. Thanks to this assumption, one can view $\mathcal{G}^c$ as a C-DMG over ADMGs and thus use prior work[Ferreira and Assaad, 2025b]. Suppose there exists a SC-hedge for the pair $(\mathbb{C}_X, \mathbb{C}_Y)$ in $\mathcal{G}^c$. Then, according to Theorem 5 of Ferreira and Assaad [2025b], the effect of $\mathbb{C}_X$ on $\mathbb{C}_Y$ is not identifiable.

In conclusion, the SC-hedge criterion is sound in C-DMG over DMGs under the additional assumption that every cluster which is in a cycle is of size at least 2. $\square$

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [Yes]

Justification: In the abstract and in the introduction we clearly state our contributions. Each of them is formalized in a theorems in the main body of the paper.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: Section 4 is dedicated to the discussion of the limitations of our work. The main limitation of this work lies in the completeness result of the do-calculus, and this limitation is exhaustively discussed at the end of Section 3.2. Moreover, the assumption necessary for Theorem 5 is clearly stated.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The main assumption lies in the problem setting i.e., the system we wish to study is modeled by an ioSCM. However, another additional assumption which is clearly stated is required for Theorem 5. Complete proofs are available in the supplementary material, unfortunately, due to a lack of writing space, we were not capable of adding every proof sketches in the main paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: These results are purely theoretical. We do not believe that any experiment is needed nor would any experiment strengthen the claims of this work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: These results are purely theoretical. We do not believe that any experiment is needed nor would any experiment strengthen the claims of this work.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: These results are purely theoretical. We do not believe that any experiment is needed nor would any experiment strengthen the claims of this work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: These results are purely theoretical. We do not believe that any experiment is needed nor would any experiment strengthen the claims of this work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: These results are purely theoretical. We do not believe that any experiment is needed nor would any experiment strengthen the claims of this work.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: This work does not violate the NeurIPS Code of Ethics in anyway.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: The work is very theoretical and does not have any clear societal impact.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This work does not pose such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [NA]

Justification: This work does not use any existing assets.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.

- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: This work does not release any new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: This work does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: This work does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.

- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This work does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.