Effect Identification in Causal Diagrams with Clustered Variables

Tara V. Anand* Columbia University tara.v.anand@columbia.edu

Adele H. Ribeiro* Columbia University adele@cs.columbia.edu

Jin Tian Iowa State University jtian@iastate.edu

Elias Bareinboim Columbia University eb@cs.columbia.edu

Abstract

One pervasive task found throughout the empirical sciences is to determine the effect of interventions from non-experimental data. It is well-understood that assumptions are necessary to perform such causal inferences, an idea popularized through Cartwright's motto: "no causes-in, no causes-out." One way of articulating these assumptions is through the use of causal diagrams, which are a special type of graphical model with causal semantics [Pearl, 2000]. The graphical approach has been applied successfully in many settings, but there are still challenges to its use, particularly in complex, high-dimensional domains. In medicine, for example, background knowledge may exist about the relationships among a subset of variables, but usually not all of them. In this paper, we introduce *cluster causal diagrams* (for short, C-DAGs) to allow for the representation of partial understanding of the relationships among variables, which has the potential of alleviating DAGs' somewhat stringent requirements. C-DAGs provide a simple yet effective way to partially abstract a grouping of variables among which causal relationships are not fully understood. Our goal is to develop machinery to reason on top of C-DAG's new representation. In particular, we first define a new version of the d-separation criterion, and prove its soundness and completeness. Second, we extend these new separation rules and prove the validity of the corresponding do-calculus. Lastly, we show that a standard identification algorithm can systematically compute causal effects from observational data with cluster causal diagrams.

1 Introduction

One of the main tasks in the empirical sciences and data-driven disciplines is to infer cause and effect relationships using observational (non-experimental) data collected from the phenomenon under investigation. These relations are considered central in the construction of explanations and for making decisions about interventions that were never implemented before [8, 13, 1, 10, 9].

Assumptions about the underlying generating processes are needed to perform these causal inferences [8]. One popular way of articulating such assumptions is through the language of graphical models, in particular in what is known as a *causal diagram*. Intuitively, a causal diagram G has an arrow from X to Y (e.g., see Fig. 1(a)) if Y "listens" to the value of X; in terms of the underlying causal system, X appears as an argument of the mechanism of Y. The importance of this notion has been emphasized in the literature by Pearl: "This listening metaphor encapsulates the entire knowledge that a causal network conveys; the rest can be derived, sometimes by leveraging data" [9, pp. 129].

Once a diagram is fully articulated, to infer the effect of a treatment X on an outcome Y, – written as the do-distribution $P(Y|do(X = x)$ – one engages in the identifiability task to determine whether $P(Y|do(X = x)$ is derivable from the combination of the diagram, G, and the observational distribution, $P(\mathbf{V})$, where $\mathbf{V}$ represents the set of all observed variables. The identifiability (decision)

*These authors contributed equally to this work.

Preprint. Under review.

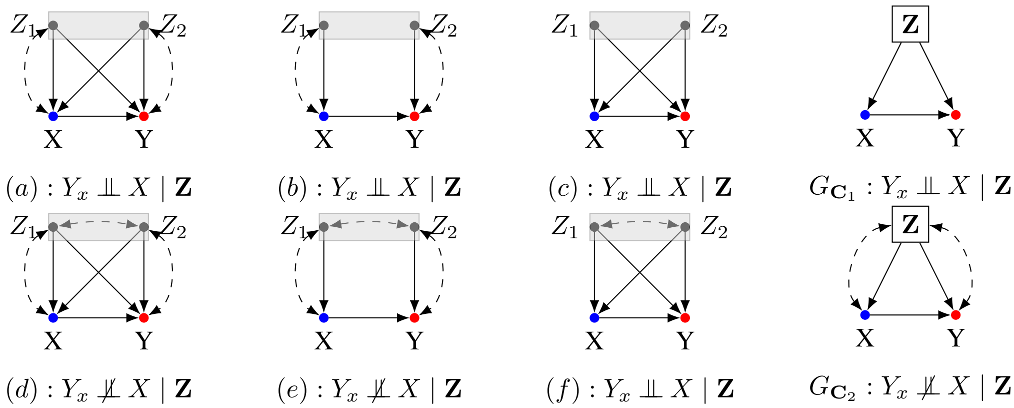

Figure 1: The difference between each model in the top and bottom rows is the relationship between the Z-nodes. Note that ignorability holds (and $P(Y|do(X))$ is identifiable from $P(\mathbf{V})$) in diagrams (a), (b), (c), and (f) but not in diagrams (d) and (e). $G_{\mathbf{C}_1}$ is a cluster causal diagram for (c) and (f). $G_{\mathbf{C}_2}$ is a cluster causal diagram for (a), (b), (d), and (e).

problem has been extensively studied in the literature and a number of conditions have been proposed. These include symbolic methods such as Pearl's do-calculus [7] as well as algorithmic strategies [2, 15, 12, 3, 5], to cite a few. Despite the power of such approaches, there are settings where the knowledge necessary to articulate the required assumptions may not be available. Obtaining the graph over the observed variables, $\mathbf{V}$, requires knowledge about the relationships among all pairs in $\mathbf{V}$, which may not be fully understood in some domains.

Another popular way of articulating assumptions about the underlying mechanisms is through a condition known as *conditional ignorability*, as introduced by Rubin [11, 4]. A new *potential outcome* variable Y_x is defined to represent the hypothetical value of Y had X been x. Ignorability is written as $Y_x \perp\!\!\!\perp X|\mathbf{Z}$, which represents the independence of the potential outcome Y_x and the treatment X conditional on a set of covariates $\mathbf{Z} \subset \mathbf{V}$, where $\mathbf{Z}$ is assumed pre-treatment. If conditional ignorability holds, the distribution $P(Y|do(X))$ is identifiable and given by the adjustment formula.

In some sense, this approach does not require the relationships among all pairs of variables to be elicited; X, Y, and $\mathbf{Z}$ are treated as blocks. This has the potential to alleviate the pressure required by the graphical approach of eliciting the relationships among all pairs of variables. Despite the power of such an approach, leaving the relationships among all covariates, $\mathbf{Z}$, unspecified may incur some possibly consequential loss of information. To understand this subtlety, consider the diagrams in Figs. 1(a) and (d), which we call G and G'. Note that they are identical except for the relationship within the $\mathbf{Z} = \{Z_1, Z_2\}$ cluster; Z_1 and Z_2 are independent in G (and $Y_x \perp\!\!\!\perp X|\mathbf{Z}$ holds), while they are correlated through an unobserved common cause (encoded by the bidirected arrow) in G' (and $Y_x \perp\!\!\!\perp X|\mathbf{Z}$ does not hold). The same phenomenon occurs with the pairs in the second column (b, e). If one merely has the qualitative knowledge of how the cluster $\mathbf{Z}$ interacts with X and Y, it's not possible in these cases to judge whether ignorability holds (and identifiability is entailed) while being oblivious to the relationships among the $\mathbf{Z}$-variables.

After all, it seems that, on the one hand, it may be hard to elicit the relationships among all variables in some practical, real world scenarios; on the other, clustering some variables while ignoring what is within each cluster may lead to inadvertent inferences. Even though the situation seems somewhat knotty, we note there is some way of striking a middle ground between these strategies. To witness, consider the graphs in Figs. 1(c) and (f), and note that $P(Y|do(X))$ is identifiable in both cases. In words, identifiability follows *regardless* of the relationship among the $\mathbf{Z}$ nodes.

In this paper, we aim to find such cases where the validity of the inferences are preserved while the relationships within a cluster can be left unspecified. More broadly, our goal is to bridge the gap between requiring a full specification of the system and clustering all covariates together, while ignoring what is within the cluster. In particular, to allow for more refined groupings of variables, we introduce a new class of graphical models over a set of clusters of variables, where the relationships amongst variables within a cluster are left unspecified. We develop machinery for valid causal inference in such models. Specifically, our contributions are as follows:

1. We introduce a new class of graphs called cluster causal diagrams or C-DAGs (for cluster DAGs) over a set of clusters of variables where the relationships amongst the variables inside the clusters are left unspecified *(Definition 1)*.
2. We show that despite C-DAGs' abstracted representation of numerous possible underlying paths, path separation rules can be augmented to include clusters. We then show completeness of the d-separation rules extended to C-DAGs *(Theorem 1)*.
3. We prove that the inference rules known as Pearl's do-calculus are sound and complete for the coarse representation of C-DAGs *(Theorems 2 and 3)*.
4. We prove that interventional distributions factorize according to C-DAGs *(Theorem 4)*. This ensures the ID-algorithm can systematically infer causal effects from the combination of an observational distribution and partial domain knowledge encoded as a C-DAG *(Theorem 5)*.

2 Preliminaries

We introduce in this section the necessary concepts and notation used throughout the paper.

Notation. A single random variable is denoted by a (non-boldface) uppercase letter X and its realized value by a small letter x. A boldfaced uppercase letter $\mathbf{X}$ denotes a set (or a cluster) of variables. For $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, we denote the conditional independence of $\mathbf{X}$ and $\mathbf{Y}$ conditioned on $\mathbf{Z}$ by $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$. Given a graph G, we denote the d-separation between $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ by $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G$. The mutilated graph $G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ is the result of removing from G arrows coming into variables in $\mathbf{X}$ and going out of variables in $\mathbf{Z}$. $G_{\mathbf{C}}$ denotes the subgraph of G over $\mathbf{C}$. We use kinship terminology (e.g., parents, children, descendants, ancestors) to denote various relationships in a graph. These kinship relations are defined along the full arrows in the graph, ignoring bidirected arrows. We use $Pa(\mathbf{X})_G$, $An(\mathbf{X})_G$, and $De(\mathbf{X})_G$ to represent the sets of parents, ancestors, and descendants in G, respectively.

Structural Causal Models. We use the language of Structural Causal Models (SCMs) [8, pp. 204-207] as our semantical framework. Formally, an SCM $\mathcal{M}$ is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U}) \rangle$, where $\mathbf{U}$ is a set of exogenous (latent) variables and $\mathbf{V}$ is a set of endogenous (measured) variables. $\mathcal{F}$ is a collection of functions $\{f_i\}_{i=1}^{|\mathbf{V}|}$ such that each endogenous variable $V_i \in \mathbf{V}$ is determined by a function $f_i \in \mathcal{F}$, where f_i is a mapping from the respective domain of $\mathbf{U}_i \cup Pa(V_i)$ to V_i, where $\mathbf{U}_i \subseteq \mathbf{U}$ and $Pa(V_i) \subseteq \mathbf{V} \setminus V_i$. The uncertainty is encoded through a probability distribution over the exogenous variables, $P(\mathbf{U})$. Each SCM $\mathcal{M}$ induces a directed acyclic graph (DAG) $G(\mathbf{V}, \mathbf{E})$, known as causal diagram, that encodes the structural relations among $\mathbf{V} \cup \mathbf{U}$, where every $V_i \in \mathbf{V}$ is a vertex, there is a directed arrow $(V_j \rightarrow V_i)$ for every $V_i \in \mathbf{V}$ and $V_j \in Pa(V_i)$, and there is a dashed bidirected arrow $(V_j \dashleftarrow\dashrightarrow V_i)$ for every pair $V_i, V_j \in \mathbf{V}$ such that $\mathbf{U}_i \cap \mathbf{U}_j \neq \emptyset$ (V_i and V_j have a common exogenous parent). Performing an action $\mathbf{X}{=}\mathbf{x}$ is represented through the do-operator, $do(\mathbf{X}{=}\mathbf{x})$, which represents the operation of fixing a set $\mathbf{X}$ to a constant $\mathbf{x}$ regardless of their original mechanisms. Such an intervention induces a submodel $\mathcal{M}_{\mathbf{x}}$, which is $\mathcal{M}$ with f_X replaced to x for every $X \in \mathbf{X}$. The post-interventional distribution induced by $\mathcal{M}_{\mathbf{x}}$ is denoted by $P(\mathbf{v} \setminus \mathbf{x} | do(\mathbf{x}))$.

3 Cluster Causal Diagrams: Definitions and Properties

We now provide a definition for what we will call a cluster causal diagram, or C-DAG, which defines a coarser graphical representation of an SCM where variables are grouped as entities called clusters. In this definition, every variable is a part of exactly one cluster, where a variable may be "grouped" in a cluster by itself and all vertices in a C-DAG are clusters.

Definition 1 (Cluster Causal Diagram or C-DAG). Consider an SCM $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathcal{F}, P(\mathbf{u}) \rangle$ and the corresponding causal diagram $G(\mathbf{V}, \mathbf{E})$. Given a partition $\mathbf{C} = \{\mathbf{C}_1, \ldots, \mathbf{C}_k\}$ of $\mathbf{V}$, construct a graph $G_{\mathbf{C}}(\mathbf{C}, \mathbf{E}_{\mathbf{C}})$ over $\mathbf{C}$ with a set of arrows $\mathbf{E}_{\mathbf{C}}$ defined as follows:

1. An arrow $\mathbf{C}_i \rightarrow \mathbf{C}_j$ is in $\mathbf{E}_{\mathbf{C}}$ if there exists some $V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that $V_i \in Pa(V_j)$;
2. A dashed bidirected arrow $\mathbf{C}_i \dashleftarrow\dashrightarrow \mathbf{C}_j$ is in $\mathbf{E}_{\mathbf{C}}$ if there exists some $V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that there is a bidirected arrow $V_i \dashleftarrow\dashrightarrow V_j$.

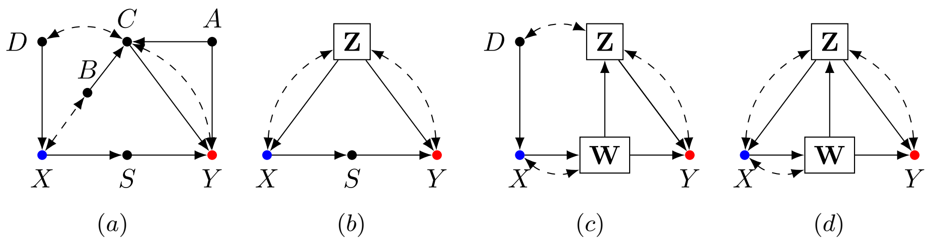

Figure 2: (a): a possible causal diagram representing the effect of taking lisinopril on having a stroke where X =lisinopril, Y = stroke, A =age, B = blood pressure, C = comorbidities, D = medication history, and S =sleep quality. (b): a C-DAG of (a) with $\mathbf{Z} = \{A, B, C, D\}$ where $P(y \mid do(x))$ is identifiable by front-door adjustment. (c): a C-DAG of (a) with $\mathbf{W} = \{S, B\}, \mathbf{Z} = \{A, C\}$ where $P(y \mid do(x))$ is not identifiable (d): an invalid C-DAG of (a) with $\mathbf{W} = \{S, B\}, \mathbf{Z} = \{A, C, D\}$; note the cycle among $(X, \mathbf{W}, \mathbf{Z})$.

If $G_{\mathbf{C}}(\mathbf{C}, \mathbf{E}_{\mathbf{C}})$ contains no cycles, then we say that $\mathbf{C}$ is an *admissible partition* of $\mathbf{V}$. Further, we then call $G_{\mathbf{C}}$ a *cluster causal diagram*, or *C-DAG* compatible with G.

Note: throughout the paper, we will use the same symbol (e.g. $\mathbf{C}$) to represent a set of clusters in a C-DAG $G_{\mathbf{C}}$ and the set of variables contained across these clusters in a compatible causal diagram G.

In practice, all causal relationships among variables may not be known or it may be burdensome to explicitly define them, as is common in fields like medicine and the social sciences. However, contextual or temporal knowledge may partially inform how variables are related in a causal diagram, such that a C-DAG can be constructed. Fig. 2(a) illustrates a medical example with a possible causal diagram for the effect of lisinopril (X), a treatment for hypertension, on the outcome of having a stroke (Y). A researcher can be certain that the variables of age (A), pre-treatment blood pressure (B), comorbidities (C), and medication history (D) occur temporally before lisinopril is prescribed or a stroke occurs, even if the relationships among such covariates are unknown, such that these variables can be grouped together as in 2(b). Similarly temporal information could allow a researcher to group post-treatment variables, such as sleep quality (S), as a mediator between the treatment and outcome. Knowledge of potential unmeasured confounders can be used in constructing a C-DAG, as indicated, for example, by the bidirected arrow between C and Y in 2(a), illustrating some latent connection (e.g. environmental factor) between prior comorbidites and some health outcome.

Not only are these clusterings helpful in visually communicating a complex diagram, but this coarser representation as a C-DAG is less prone to model misspecification than strict causal diagrams, which require more knowledge to be constructed. Following the analogy of an arrow being drawn from one variable X to another Y if Y "listens" to the value of X, intuitively we can have an arrow drawn from one cluster, $\mathbf{X}$ to another $\mathbf{Y}$ if at least some variables in $\mathbf{Y}$ "listen" to at least some variables in $\mathbf{X}$. In the rest of the paper, we will develop formal machinery to support causal analysis in C-DAGs.

Let Fig. 2(a) be a causal diagram G over $\mathbf{V} = \{X, Y, A, B, C, D, S\}$, Fig. 2(b) shows a graph over the partition $\mathbf{C}' = \{\{X\}, \{Y\}, \{A, B, C, D\}, \{S\}\}$ where $\mathbf{Z} = \{A, B, C, D\}$, Fig. 2(c) shows a graph over the partition $\mathbf{C} = \{\{X\}, \{Y\}, \{A, C\}, \{D\}, \{S, B\}\}$ with $\mathbf{W} = \{S, B\}$ and $\mathbf{Z} = \{A, C\}$, and Fig. 2(d) shows a graph over the partition $\mathbf{C}'' = \{\{X\}, \{Y\}, \{A, C, D\}, \{S, B\}\}$ where $\mathbf{W} = \{S, B\}$ and $\mathbf{Z} = \{A, C, D\}$. For simplicity, vertices for singleton clusters are denoted by the name of their only variable and vertices for clusters with more than one variable are denoted by the cluster name with a square drawn around them. Note that $G_{\mathbf{C}}$ and $G_{\mathbf{C}'}$ in 2(b) and (c) are C-DAGs compatible with G, since no cycles are created. In 2(d), the partition $\mathbf{C}''$ is not admissible for creating a C-DAG because $G_{\mathbf{C}''}$ contains a cycle. The arrow $\{X\} \to \mathbf{W}$ is needed because $S \in \mathbf{W}$ "listens" to X. The arrow $\mathbf{W} \to \mathbf{Z}$ is needed because $C \in \mathbf{Z}$ "listens" to $B \in \mathbf{W}$. The arrow $\mathbf{Z} \to X$ is needed because X "listens" to $D \in \mathbf{Z}$. This creates a cycle, so $G_{\mathbf{C}''}$ is not a C-DAG.

Cluster causal diagrams can be seen as an equivalence class of causal diagrams in which all the possible combinations of relationships among the variables within each cluster are allowed. For instance, Fig. 3 illustrates two causal diagrams, G_1 and G_2, that are both represented by the C-DAG $G_{\mathbf{C}}$. Note that the backdoor path $X \leftarrow Z_1 \to Z_2 \leftarrow Z_3 \to Y$ in G_1 is inactive due to the collider Z_2, while all backdoor paths in G_2 are active. In the next section, we will categorize the status

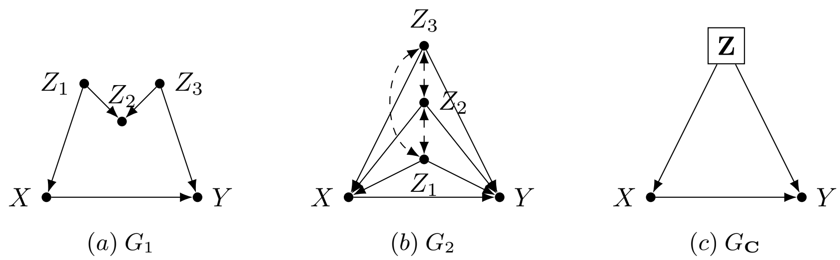

Figure 3: (a) and (b): causal diagrams G_1 and G_2, (c) represented by (c) C-DAG $G_{\mathbf{C}}$, where $\mathbf{Z} = \{Z_1, Z_2, Z_3\}$.

of different paths in a C-DAG and will see that seemingly active paths (e.g., $X \leftarrow \mathbf{Z} \rightarrow Y$) may represent an inactive path in some compatible causal diagram. To properly identify causal effects when knowledge about the relationships among some variables at play is only partially available, we will seek to derive causal inference rules and algorithms that are applicable to *all* the causal diagrams compatible with a given C-DAG regardless of the unknown relationships within clusters.

The criteria set forth in Definition 1 ensure that a C-DAG $G_{\mathbf{C}}$ has two properties that will prove critical later on. Namely, they preserve the connections and they preserve the ancestral relationships in any compatible causal diagram G, as formally stated below.

Proposition 1. ***(Preservation of connections)*** *Let $G_{\mathbf{C}}(\mathbf{C}, \mathbf{E}_{\mathbf{C}})$ be a C-DAG compatible with a causal diagram $G(\mathbf{V}, \mathbf{E})$. Consider distinct clusters $\mathbf{C}_i, \mathbf{C}_j \in \mathbf{C}$. If $V_i, V_j \in \mathbf{V}$ are connected in G and belong to $\mathbf{C}_i, \mathbf{C}_j$ respectively, then $\mathbf{C}_i$ and $\mathbf{C}_j$ are connected in $G_{\mathbf{C}}$. Further, if $\mathbf{C}_i$ and $\mathbf{C}_j$ are connected in $G_{\mathbf{C}}$, then there exists $V_i \in \mathbf{C}_i$ and $V_j \in \mathbf{C}_j$ such that V_i and V_j are connected in G.*

Proposition 2. ***(Preservation of ancestral relationships)*** *Let $G_{\mathbf{C}}(\mathbf{C}, \mathbf{E}_{\mathbf{C}})$ be a C-DAG compatible with causal diagram $G(\mathbf{V}, \mathbf{E})$. For any $\mathbf{C}_i \in \mathbf{C}$, $\forall V \in \mathbf{C}_i$, $An(V)_G \subseteq An(\mathbf{C}_i)_{G_{\mathbf{C}}}$ and $De(V)_G \subseteq De(\mathbf{C}_i)_{G_{\mathbf{C}}}$.*

Proposition 1 states that a C-DAG encodes any connection between variables is preserved in a compatible causal diagram. Proposition 2 states that the set of ancestor (descendant) clusters of $\mathbf{C}_i$ contain all ancestors (descendants) of any variable $V \in \mathbf{C}_i$. These results will prove helpful to extend standard causal inference tools to C-DAGs, including d-separation and do-calculus.

4 D-Separation in C-DAGs

D-separation is a central mechanism for inferences with causal diagrams, which has been discussed in the context of probabilistic [6] and causal reasoning [7]. In this section, we investigate d-separation rules in C-DAGs when conditioning on a mediator, a common cause, a collider, and a collider descendant.

Cluster in a Causal Chain

Consider a C-DAG $G_{\mathbf{C}}$ where $\mathbf{Z}$ is a mediator between $\mathbf{X}$ and $\mathbf{Y}$ in the chain as shown in Fig. 4(a).

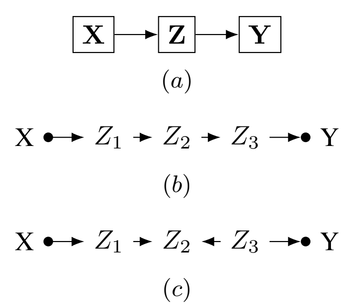

Figure 4: Causal diagrams (b) and (c) are both compatible with the cluster causal diagram (a), where $\mathbf{Z}$ is a mediator between $\mathbf{X}$ and $\mathbf{Y}$. The path is active in (b) and inactive in (c).

The causal diagrams (b) and (c) in Fig. 4, over $\mathbf{V} = \{X, Z_1, Z_2, Z_3, Y\}$, are compatible with the C-DAG $G_{\mathbf{C}}$, where $\mathbf{X} = \{X\}$, $\mathbf{Z} = \{Z_1, Z_2, Z_3\}$, and $\mathbf{Y} = \{Y\}$. Note that in diagram (b) the path between X and Y is active. However, in diagram (c), the path between X and Y is inactive, even though the abstracted cluster $\mathbf{Z}$ is a mediator. From these two examples, it is clear that when a cluster $\mathbf{Z}$ acts as a mediator between two other clusters $\mathbf{X}$ and $\mathbf{Y}$, the path between $\mathbf{X}$ and $\mathbf{Y}$ through $\mathbf{Z}$ may be either active or inactive.

Remark 1. *In a C-DAG, the path $\mathbf{X} \rightarrow \mathbf{Z} \rightarrow \mathbf{Y}$ may be either active or inactive.*

Conditioning on a cluster will be considered as equivalent to conditioning on all the variables in it. In both diagrams (b) and (c), the path between X and Y is inactive when conditioning on $\{Z_1, Z_2, Z_3\}$ and we show that this property will hold regardless of the connections within a cluster mediator as stated in the following lemma.

Lemma 1. *In a C-DAG, the path* $\mathbf{X} \to \mathbf{Z} \to \mathbf{Y}$ *is inactive when mediator* $\mathbf{Z}$ *is conditioned on.*

Cluster as a Common Cause

Consider a C-DAG $G_{\mathbf{C}}$ where $\mathbf{Z}$ is a common cause of $\mathbf{X}$ and $\mathbf{Y}$ as shown in Fig. 5(a).

The causal diagrams (b) and (c) in Fig. 5, over $\mathbf{V} = \{X, Z_1, Z_2, Z_3, Y\}$, are compatible with the C-DAG $G_{\mathbf{C}}$, where $\mathbf{X} = \{X\}$, $\mathbf{Z} = \{Z_1, Z_2, Z_3\}$, and $\mathbf{Y} = \{Y\}$. In diagram (b), the path is active while in diagram (c) the path is inactive. These examples illustrate that when a cluster $\mathbf{Z}$ acts as a common cause of two other clusters $\mathbf{X}$ and $\mathbf{Y}$, the path between $\mathbf{X}$ and $\mathbf{Y}$ through $\mathbf{Z}$ may be either active or inactive.

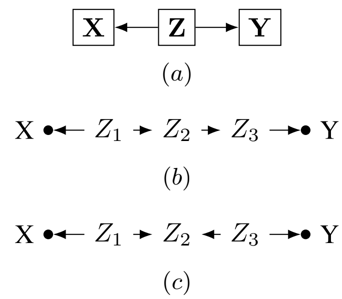

Figure 5: Causal diagrams (b) and (c) are both compatible with the cluster causal diagram (a), where $\mathbf{Z}$ is a common cause of $\mathbf{X}$ and $\mathbf{Y}$. The path is active in (b) and inactive in (c).

Remark 2. *In a C-DAG, the path* $\mathbf{X} \leftarrow \mathbf{Z} \to \mathbf{Y}$ *may be either active or inactive.*

When we condition on $\{Z_1, Z_2, Z_3\}$ in either diagram, we block the overall path from X to Y. Similar to the behavior discussed above for a cluster in a causal chain, we have the general result in the following lemma.

Lemma 2. *In a C-DAG, the path* $\mathbf{X} \leftarrow \mathbf{Z} \to \mathbf{Y}$ *is inactive when* $\mathbf{Z}$ *is conditioned on.*

Cluster as a Collider

Consider a C-DAG $G_{\mathbf{C}}$ where $\mathbf{Z}$ is a collider between $\mathbf{X}$ and $\mathbf{Y}$ as shown in Fig. 6(a).

The causal diagrams (b) and (c) in Fig. 6, over $\mathbf{V} = \{X, Z_1, Z_2, Z_3, Y\}$, are compatible with the C-DAG $G_{\mathbf{C}}$, where $\mathbf{X} = \{X\}$, $\mathbf{Z} = \{Z_1, Z_2, Z_3\}$, and $\mathbf{Y} = \{Y\}$. In diagram (b), the path between X and Y is inactive when conditioning on $\{Z_1, Z_2, Z_3\}$. However, in diagram (c), the path between X and Y is active when conditioning on $\{Z_1, Z_2, Z_3\}$. These examples illustrate that when a cluster $\mathbf{Z}$ acts as a collider between two other clusters $\mathbf{X}$ and $\mathbf{Y}$, the path between $\mathbf{X}$ and $\mathbf{Y}$ through $\mathbf{Z}$ may be either active or inactive when conditioning on $\mathbf{Z}$.

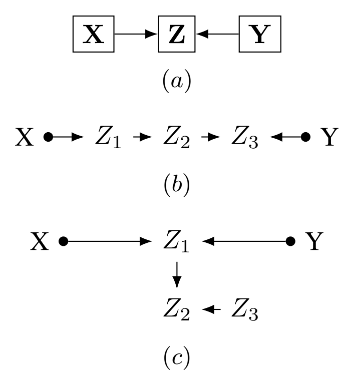

Figure 6: Causal diagrams (b) and (c) are both compatible with the C-DAG (a), where $\mathbf{Z}$ is a collider between $\mathbf{X}$ and $\mathbf{Y}$. When $\mathbf{Z}$ is conditioned on, the path is inactive in (b) and active in (c).

Remark 3. *In a C-DAG, the path* $\mathbf{X} \to \mathbf{Z} \leftarrow \mathbf{Y}$ *may be either active or inactive when* $\mathbf{Z}$ *is conditioned on.*

In both diagrams (b) and (c), the path between X and Y is inactive. We show this property holds in general regardless of the connections within a cluster collider as stated in the following lemma.

Lemma 3. *In a C-DAG, the path* $\mathbf{X} \to \mathbf{Z} \leftarrow \mathbf{Y}$ *is inactive.*

Cluster as a Descendent of a Collider

Consider a C-DAG $G_{\mathbf{C}}$ where $\mathbf{Z}$ is a descendant of collider $\mathbf{W}$ as shown in Fig. 7(a).

The causal diagram (b) in Fig. 7 over $\mathbf{V}_1 = \{X, W, Z_1, Z_2, Y\}$ and the causal diagram (c) over $\mathbf{V}_2 = \{X, W_1, W_2, Z_1, Z_2, Y\}$ are compatible with the C-DAG $G_{\mathbf{C}}$, where $\mathbf{X} = \{X\}$, $\mathbf{Z} = \{Z_1, Z_2\}$, and $\mathbf{Y} = \{Y\}$ in both diagrams, $\mathbf{W} = \{W\}$ for (b), and $\mathbf{W} = \{W_1, W_2\}$ for (c). In diagram (b), the path from X and Y is active when conditioning on $\{Z_1, Z_2\}$ while in diagram (c) the path is inactive when conditioning on $\{Z_1, Z_2\}$. We therefore have the following conclusion.

Remark 4. *In a C-DAG, the path* $\mathbf{X} \to \mathbf{W} \leftarrow \mathbf{Y}$ *may be either active or inactive when* $\mathbf{Z}$*, a descendant of* $\mathbf{W}$*, is conditioned on.*

In both diagrams (b) and (c), the path between X and Y is inactive. We have the following result.

Lemma 4. *In a C-DAG, the path* $\mathbf{X} \to \mathbf{W} \leftarrow \mathbf{Y}$ *is inactive when none of the descendants of* $\mathbf{W}$ *(nor* $\mathbf{W}$*) are conditioned on.*

D-separation rules in C-DAGs

The above discussion and lemmas illustrate how the original d-separation rules relate to C-DAGs. In summary, the rules for when a path is d-separated carry to C-DAGs, but when a path between two clusters is not d-separated in C-DAGs, the path between corresponding variables may or may not be active in a compatible causal diagram. These observations together, lead to the following definition:

Definition 2 (**d-Separation in C-DAGs**)**.** A path p in a C-DAG $G_{\mathbf{C}}$ is said to be d-separated (or blocked) by a set of clusters $\mathbf{Z} \subset \mathbf{C}$ if and only if

1. p contains a mediator $\mathbf{C}_i \to \mathbf{C}_m \to \mathbf{C}_j$ or a common cause $\mathbf{C}_i \leftarrow \mathbf{C}_m \to \mathbf{C}_j$ such that the middle cluster $\mathbf{C}_m$ is in $\mathbf{Z}$, or
2. p contains a collider $\mathbf{C}_i \to \mathbf{C}_m \leftarrow \mathbf{C}_j$ such that the middle cluster $\mathbf{C}_m$ is not in $\mathbf{Z}$ and no descendant of $\mathbf{C}_m$ is in $\mathbf{Z}$.

A set of clusters $\mathbf{Z}$ is said to d-separate two sets of clusters $\mathbf{X}, \mathbf{Y} \subset \mathbf{C}$, denoted by $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{G_{\mathbf{C}}}$, if and only if $\mathbf{Z}$ blocks every path from a cluster in $\mathbf{X}$ to a cluster in $\mathbf{Y}$.

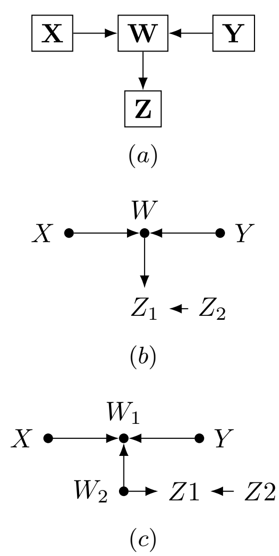

Figure 7: Causal diagrams (b) and (c) are both compatible with the C-DAG (a), where $\mathbf{W}$ is a collider between $\mathbf{X}$ and $\mathbf{Y}$ and $\mathbf{Z}$ is a descendant of $\mathbf{W}$. When $\mathbf{Z}$ is conditioned on, the path is active in (b) and inactive in (c).

The following result states the soundness and completeness of the d-separation rules in C-DAGs.

Theorem 1. *(**Soundness and completeness of d-separation**). Consider a C-DAG* $G_{\mathbf{C}}$*, and let* $\mathbf{X}, \mathbf{Z}, \mathbf{Y} \subset \mathbf{C}$*. If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are d-separated by* $\mathbf{Z}$ *in* $G_{\mathbf{C}}$*, then, in any causal diagram* G *compatible with* $G_{\mathbf{C}}$*,* $\mathbf{X}$ *and* $\mathbf{Y}$ *are d-separated by* $\mathbf{Z}$ *in* G*. If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are not d-separated by* $\mathbf{Z}$ *in* $G_{\mathbf{C}}$*, then, there exists a causal diagram* G *compatible with* $G_{\mathbf{C}}$ *where* $\mathbf{X}$ *and* $\mathbf{Y}$ *are not d-separated by* $\mathbf{Z}$ *in* G*.*

While C-DAGs inevitably lead to a loss of information, as evidenced by paths in a C-DAG which may be either active or inactive in a compatible DAG, we can still make strong conclusions about conditional independencies entailed by compatible DAGs. From the understanding of d-separation in DAGs, we can see that the d-separation rules for clusters allow us to read off conditional independencies from an even broader set of distributions. A d-separation between sets of variables $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ in a DAG G implies that $\mathbf{X}$ is independent of $\mathbf{Y}$ conditional on $\mathbf{Z}$ in every distribution induced by G. A d-separation between clusters $\mathbf{X}$ and $\mathbf{Y}$ given $\mathbf{Z}$ in a C-DAG $G_{\mathbf{C}}$ implies that $\mathbf{X}$ is independent of $\mathbf{Y}$ conditional on $\mathbf{Z}$ in every distribution compatible with any G compatible with $G_{\mathbf{C}}$.

5 Causal Identification in C-DAGs

The aforementioned d-separation criterion allows us to decide from a C-DAG $G_{\mathbf{C}}$ whether a set of clusters $\mathbf{X}$ is independent of another set $\mathbf{Y}$ when a third set $\mathbf{Z}$ is conditioned on in any compatible causal diagram G. This will be essential to establish conditions for causal identifiability in C-DAGs.

5.1 Do-Calculus in C-DAGs

One of the fundamental tools used in causal inference is Pearl's celebrated *do-calculus* [7]. The do-calculus has been used extensively for solving a variety of causal effect identification tasks.

Armed with the understanding coming from the d-separation rules in C-DAGs, we show next that the do-calculus rules are also sound in C-DAGs.

Theorem 2. *(**Do-calculus in C-DAGs**). Let $G_{\mathbf{C}}$ be a C-DAG compatible with a causal diagram G associated with an SCM $\mathcal{M}$. For any disjoint subsets of clusters $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \subseteq \mathbf{C}$, the following three rules hold:*

Rule 1: $P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{w})$ *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\mathbf{C}_{\overline{\mathbf{X}}}}}$

Rule 2: $P(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w})$ *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\mathbf{C}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}}$

Rule 3: $P(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{w})$ *if* $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\mathbf{C}_{\overline{\mathbf{X}\mathbf{Z}(\mathbf{W})}}}}$

where the graph $G_{\mathbf{C}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$ is obtained from $G_{\mathbf{C}}$ by removing the arrows incoming to $\mathbf{X}$ and outgoing from $\mathbf{Z}$, and $\mathbf{Z}(\mathbf{W})$ is the set of $\mathbf{Z}$-clusters that are non-ancestors of any $\mathbf{W}$-cluster in $G_{\mathbf{C}_{\overline{\mathbf{X}}}}$.

Note that in the above do-calculus rules, $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ represent sets of cluster nodes in $G_{\mathbf{C}}$ in the d-separation tests $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}, \mathbf{W})_{G_{\mathbf{C}}}$, while in the distribution $P(.)$ statements, $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ represent the sets of variables contained in the corresponding sets of clusters.

Given the soundness and completeness of the d-separation rules in C-DAGs, the main question to be solved in order to show the soundness of do-calculus in C-DAGs is whether the mutilation operations in a C-DAG to create $G_{\mathbf{C}_{\underline{\mathbf{X}}}}$ and $G_{\mathbf{C}_{\overline{\mathbf{X}}}}$ carry over to the compatible causal graphs. We obtain the following result based on Proposition 1.

Lemma 5. *If a C-DAG $G_{\mathbf{C}}$ is compatible with a causal diagram G, then, for $\mathbf{X}, \mathbf{Z} \subset \mathbf{C}$, $G_{\mathbf{C}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$ is compatible with $G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$.*

Theorem 2 then follows from Propositions 1 and 2, Theorem 1, and Lemma 5. In addition, we show that the do-calculus rules in C-DAGs are complete as follows:

Theorem 3. *(**Completeness of do-calculus**). If a do-calculus rule does not apply in a C-DAG $G_{\mathbf{C}}$, then there exists a causal diagram G compatible with $G_{\mathbf{C}}$ for which it also does not apply.*

Equipped with d-separation and do-calculus in C-DAGs, causal inference algorithms developed for a variety of tasks that rely on a known causal diagram can potentially be extended to C-DAGs [1]. In this paper, we study the problem of identifying causal effects from observational data in C-DAGs.

5.2 ID-Algorithm

A complete algorithm has been developed to determine whether $P(\mathbf{y}|do(\mathbf{x})$ is identifiable from the combination of the causal diagram G and the observational distribution $P(\mathbf{V})$ [14, 15]. This identification algorithm, or ID-algorithm for short, is based on the factorization of the post-interventional distributions according to the graphical structure, known as the truncated factorization. Specifically,

$$P(\mathbf{v} \setminus \mathbf{x}|do(\mathbf{x})) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k:V_k \in \mathbf{V}\setminus\mathbf{X}} P(v_k|pa_{v_k}, \mathbf{u}_k). \tag{1}$$

We will show that the truncated factorization holds in C-DAGs as well, in the following sense.

Theorem 4. *(**Truncated factorization in C-DAGs.**) Let $G_{\mathbf{C}}$ be a C-DAG compatible with a causal diagram G associated with an SCM $\mathcal{M} = \langle \mathbf{U}, \mathbf{V}, \mathcal{F}, P(\mathbf{U})\rangle$. For any $\mathbf{X} \subseteq \mathbf{C}$, the following holds*

$$P(\mathbf{c} \setminus \mathbf{x}|do(\mathbf{x})) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k:\mathbf{C}_k \in \mathbf{C}\setminus\mathbf{X}} P(\mathbf{c}_k|pa_{\mathbf{C}_k}, \mathbf{u}'_k), \tag{2}$$

where $Pa_{\mathbf{C}_k}$ are the parents of $\mathbf{C}_k$ in $G_{\mathbf{C}}$ and $\mathbf{U}'_k \subseteq \mathbf{U}$ such that $\mathbf{U}'_i \cap \mathbf{U}'_j \neq \emptyset$ if and only if there is a bidirected arrow ($\mathbf{C}_i \dashleftarrow\!\!\!\!\dashrightarrow \mathbf{C}_j$) between $\mathbf{C}_i$ and $\mathbf{C}_j$ in $G_{\mathbf{C}}$.

In Eq. (2), $\mathbf{X}, \mathbf{C}, \mathbf{C}_k, Pa_{\mathbf{C}_k}$ are the sets of variables contained in the corresponding sets of clusters.

Theorem 4 shows that C-DAGs are Causal Bayesian Networks where a cluster with N variables is treated as an N-dimensional random variable. This result allows us to prove that the ID-algorithm can systematically infer causal effects from a combination of an observational distribution $P(\mathbf{V})$ and partial domain knowledge encoded as a C-DAG $G_{\mathbf{C}}$ (since ID relies on the truncated factorization).

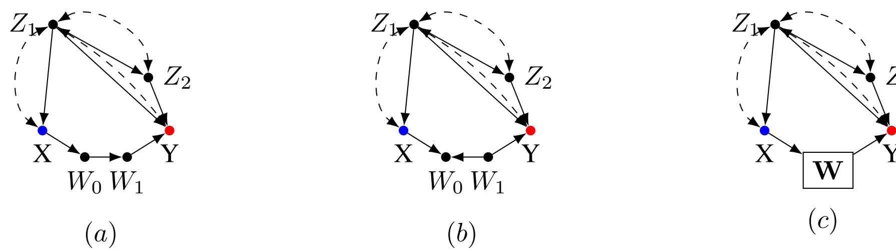

Figure 8: (a): a causal diagram where $P(y|do(x))$ is identifiable by front-door adjustment. (b): a modification of (a) where $W_0 \to W_1$ has been changed to $W_0 \leftarrow W_1$. $P(y|do(x))$ is identifiable. (c): a cluster causal diagram for both (a) and (b).

Theorem 5. *(**Soundness of ID-algorithm**). The ID-algorithm is sound when applied to a C-DAG $G_{\mathbf{C}}$ for identifying causal effects of the form $P(\mathbf{y}|do(\mathbf{x}))$ from the observational distribution $P(\mathbf{V})$, where $\mathbf{X}$ and $\mathbf{Y}$ are sets of clusters in $G_{\mathbf{C}}$.*

Furthermore, we show in Appendix A.4 (Theorem 6) that the ID-algorithm is not only sufficient but also necessary for identification from observational data (i.e., complete).

5.3 Effect Identification in C-DAGs

We now use the aforementioned results and show examples of identification in C-DAGs in practice. Due to the coarsening of the diagram by clusters, it is possible an effect will by identifiable in some causal graph G, but will not be identifiable by the clustering yielding $G_{\mathbf{C}}$.

Identification in Fig. 1. In diagrams (a), (c) and (f), the effect is determined by the backdoor adjustment over $\{Z_1, Z_2\}$, given by $P(y|do(x)) = \sum_{z_1,z_2} P(y|x, z_1, z_2)P(z_1, z_2)$. In (b), there is no backdoor path relative to (X, Y), so $P(y|do(x)) = P(y|x)$. By considering the clustering $\mathbf{Z} = \{Z_1, Z_2\}$, any separation between Z_1 and Z_2 is lost. However, under the assumption of no unobserved confounders relative to (Z_1, X) and (Z_2, X), as in diagrams (c) and (f), the effect of X on Y is identifiable regardless of the connections between Z_1 and Z_2. From the C-DAG $G_{\mathbf{C}_1}$ of (c) and (f), the effect is given by $P(y|do(x)) = \sum_{\mathbf{z}} P(y|x, \mathbf{z})P(\mathbf{z})$. If domain knowledge does not justify the assumption of no unobserved confounders relative to (Z_1, X) or (Z_2, X), then, from the C-DAG $G_{\mathbf{C}_2}$ of diagrams (a), (b), (d), and (e), the effect is not identifiable. This results from the existence of the graphs compatible with $G_{\mathbf{C}_2}$ (e.g. (d) and (e)), where the effect is not identifiable.

Identification in Fig. 2. In diagram (a) the effect of X on Y is identifiable through backdoor adjustment over the set of variables $\{B, D\}$. Because of the loss of separations during clustering in the C-DAG in (b) with $\mathbf{Z} = \{A, C\}$ and $\mathbf{W} = \{B, S\}$, the effect is no longer identifiable. Note that it is not possible to block all backdoor paths relative to (X, Y), and $\mathbf{W}$, a direct effect of X, is confounded. Fig. 2(c) is another valid C-DAG of G with $\mathbf{Z} = \{A, B, C, D\}$. With this cluster representation, the effect of X on Y is identifiable through front-door adjustment over S, given by $P(y|do(x)) = \sum_s P(s|x) \sum_{x'} P(y|x', s)P(x')$.

Identification in Fig. 8. Both diagrams (a) and (b) are compatible with the C-DAG shown in (c) when variables W_0 and W_1 are grouped in the cluster $\mathbf{W}$. In (a), we can identify the effect of X on Y using the front door: $P(y|do(x)) = \sum_{w_0,w_1} P(w_0, w_1|x) \sum_{x'} P(y|x', w_0, w_1)P(x')$. From the C-DAG (c), we can identify the effect similarly: $P(y|do(x)) = \sum_{\mathbf{w}} P(\mathbf{w}|x) \sum_{x'} P(y|x', \mathbf{w})P(x')$. When (b) is the underlying causal diagram, the front-door formula can be simplified to $P(y|do(x)) = P(y)$ using the marginal independence between W_1 and X. This information is lost during the clustering and the directed path from X to Y through $\mathbf{W}$ in the C-DAG is considered possibly active. However, the front-door formula ultimately provides an equivalent expression to the true causal effect.

6 Conclusion

We introduced in this paper a definition for cluster causal diagrams which can be used to group together variables in a causal graph for which the internal relationships may not be known a priori. We have provided a new d-separation criterion, and shown how it leads to the soundness and completeness of the do-calculus and the validity of the identification algorithm for C-DAGs. We hope this new machinery can enable researchers to represent complex systems with numerous variables in simplified ways, that allow for more nuance than previous approaches to abstraction, and maintain compatibility with the gold standard approaches to causal reasoning, such that these methods can be more easily used in practical applications.

References

[1] E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[2] D. Galles and J. Pearl. Testing identifiability of causal effects. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*, pages 185–195. Morgan Kaufmann, San Francisco, 1995.

[3] Y. Huang and M. Valtorta. Pearl's Calculus of Intervention Is Complete. In R. D. a. T.S.˜Richardson, editor, *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 217–224, Corvallis, OR, 2006. AUAI Press.

[4] G. W. Imbens and D. B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

[5] S. Lee and E. Bareinboim. Causal effect identifiability under partial-observability. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119. PMLR, 2020.

[6] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.

[7] J. Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

[8] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition, 2000.

[9] J. Pearl and D. Mackenzie. *The Book of Why*. Basic Books, New York, 2018.

[10] J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2017.

[11] D. B. Rubin. Bayesian inference for causal effects: The role of randomization. *Annals of Statistics*, 7(1):34–58, 1978.

[12] I. Shpitser and J. Pearl. Identification of Joint Interventional Distributions in Recursive semi-Markovian Causal Models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, volume 2, pages 1219–1226, 2006.

[13] P. Spirtes, C. N. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2000.

[14] J. Tian. *Studies in Causal Reasoning and Learning*. PhD thesis, Computer Science Department, University of California, Los Angeles, CA, 11 2002.

[15] J. Tian and J. Pearl. A General Identification Condition for Causal Effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, pages 567–573, Menlo Park, CA, 2002. AAAI Press/The MIT Press.