

# Causal Effect Identification in Cluster DAGs

Tara V. Anand<sup>\*1</sup>, Adèle H. Ribeiro<sup>\*2</sup>, Jin Tian<sup>3</sup>, Elias Bareinboim<sup>2</sup>

<sup>1</sup>Department of Biomedical Informatics, Columbia University

<sup>2</sup>Department of Computer Science, Columbia University

<sup>3</sup>Department of Computer Science, Iowa State University

tara.v.anand@columbia.edu, adele@cs.columbia.edu, jtian@iastate.edu, eb@cs.columbia.edu

## Abstract

Reasoning about the effect of interventions and counterfactuals is a fundamental task found throughout the data sciences. A collection of principles, algorithms, and tools has been developed for performing such tasks in the last decades (Pearl 2000). One of the pervasive requirements found throughout this literature is the articulation of assumptions, which commonly appear in the form of causal diagrams. Despite the power of this approach, there are significant settings where the knowledge necessary to specify a causal diagram over all variables is not available, particularly in complex, high-dimensional domains. In this paper, we introduce a new graphical modeling tool called *cluster DAGs* (for short, C-DAGs) that allows for the partial specification of relationships among variables based on limited prior knowledge, alleviating the stringent requirement of specifying a full causal diagram. A C-DAG specifies relationships between clusters of variables, while the relationships between the variables within a cluster are left unspecified, and can be seen as a graphical representation of an equivalence class of causal diagrams that share the relationships among the clusters. We develop the foundations and machinery for valid inferences over C-DAGs about the clusters of variables at each layer of Pearl’s Causal Hierarchy (Pearl and Mackenzie 2018; Bareinboim et al. 2020) -  $\mathcal{L}_1$  (probabilistic),  $\mathcal{L}_2$  (interventional), and  $\mathcal{L}_3$  (counterfactual). In particular, we prove the soundness and completeness of d-separation for probabilistic inference in C-DAGs. Further, we demonstrate the validity of Pearl’s do-calculus rules over C-DAGs and show that the standard ID identification algorithm is sound and complete to systematically compute causal effects from observational data given a C-DAG. Finally, we show that C-DAGs are valid for performing counterfactual inferences about clusters of variables.

## 1 Introduction

One of the central tasks found in data-driven disciplines is to infer the effect of a treatment  $X$  on an outcome  $Y$ , which is formally written as the interventional distribution  $P(Y|do(X = x))$ , from observational (non-experimental) data collected from the phenomenon under investigation. These relations are considered essential in the construction of explanations and for making decisions about interventions that have never been implemented before (Pearl 2000;

Spirtes, Glymour, and Scheines 2000; Bareinboim and Pearl 2016; Peters, Janzing, and Schölkopf 2017).

Standard tools necessary for identifying the aforementioned do-distribution, such as d-separation, do-calculus (Pearl 1995), and the ID-algorithm (Tian and Pearl 2002a; Shpitser and Pearl 2006; Huang and Valtorta 2006; Lee, Correa, and Bareinboim 2019) take as input a combination of an observational distribution and a qualitative description of the underlying causal system, often articulated in the form of a causal diagram (Pearl 2000). However, specifying a causal diagram requires knowledge about the causal relationships among all pairs of observed variables, which is not always available in many real-world applications. This is especially true and acute in complex, high-dimensional settings, which curtails the applicability of causal inference theory and tools.

In the context of medicine, for example, electronic health records include data on lab tests, drugs, demographic information, and other clinical attributes, but medical knowledge is not yet advanced enough to lead to the construction of causal diagrams over all of these variables, limiting the use of the graphical approach to inferring causality (Kleinberg and Hripacsak 2011). In many cases, however, contextual or temporal information about variables is available, which may partially inform how these variables are situated in a causal diagram relative to other key variables. For instance, a data scientist may know that covariates occur temporally before a drug is prescribed or an outcome occurs. They may even suspect that some pre-treatment variables are causes of the treatment and the outcome variables. However, they may be uncertain about the relationships among each pair of covariates, or it may be burdensome to explicitly define them. Given that a misspecified causal diagram may lead to wrong causal conclusions, this issue raises the question of whether a coarser representation of the causal diagram, where no commitment is made to the relationship between certain variables, would still be sufficient to determine the causal effect of interest.

In this paper, our goal is to develop a framework for causal inferences in partially understood domains such as the medical domain discussed above. We will focus on formalizing the problem of causal effect identification considering that the data scientist does not have prior knowledge to fully specify a causal diagram over all pairs of variables. First, we formally define and characterize a novel class of graphs

<sup>\*</sup>These authors contributed equally.

called *cluster DAGs* (or C-DAG, for short), which will allow for encoding of partially understood causal relationships between variables in different abstracted clusters, representing a group of variables among which causal relationships are not understood or specified. Then, we develop the foundations and machinery for valid probabilistic and causal inferences, akin to Pearl’s d-separation and do-calculus for when such a coarser graphical representation of the system is provided based on the limited prior knowledge available. In particular, we follow Pearl’s Causal Hierarchy (Pearl and Mackenzie 2018; Bareinboim et al. 2020) and develop the machinery for inferences in C-DAGs at all three inferential layers –  $\mathcal{L}_1$  (associational),  $\mathcal{L}_2$  (interventional), and  $\mathcal{L}_3$  (counterfactual). The results are fundamental first steps in terms of semantics and graphical conditions to perform probabilistic, interventional, and counterfactual inferences over clusters of variables.

Specifically, we outline our technical contributions below.

1. We introduce a new graphical modelling tool called *cluster DAGs* (or C-DAGs) over macro-variables representing clusters of variables where the relationships among the variables inside the clusters are left unspecified (*Definition 1*). Semantically, a C-DAG represents an equivalence class of all underlying graphs over the original variables that share the relationships among the clusters.
2. We show that a C-DAG is a (probabilistic) Bayesian Network (BN) over macro-variables and Pearl’s d-separation is sound and complete for extracting conditional independencies over macro-variables if the underlying graph over the original variables is a BN (*Theorems 1 and 2*).
3. We show that a C-DAG is a Causal Bayesian Network (CBN) over macro-variables and Pearl’s do-calculus is sound and complete for causal inferences about macro-variables in C-DAGs if the underlying graph over the original variables is a CBN (*Theorems 3, 4, and 5*). The results can be used to show that the ID-algorithm is sound and complete to systematically infer causal effects from the observational distribution and partial domain knowledge encoded as a C-DAG (*Theorem 6*).
4. We show that, assuming the underlying graph  $G$  is induced by an SCM  $\mathcal{M}$ , then there exists an SCM  $\mathcal{M}_C$  over macro-variables  $\mathbf{C}$  such that its induced causal diagram is  $G_C$  and it is equivalent to  $\mathcal{M}$  on statements about the macro-variables (*Theorem 7*). Therefore, the CTFID algorithm (Correa, Lee, and Bareinboim 2021) for the identification of nested counterfactuals from an arbitrary combination of observational and experimental distributions can be extended to the C-DAGs.

## 1.1 Related work

Since a group of variables may constitute a semantically meaningful entity, causal models over abstracted clusters of variables have attracted increasing attention for the development of more interpretable tools (Schölkopf et al. 2021; Shen, Choi, and Darwiche 2018). (Parviainen and Kaski 2016) studied the problem of, given a DAG, under what assumptions a DAG over macro-variables can represent the same conditional independence relations between

the macro-variables. Recent developments on causal abstraction have focused on the distinct problem of investigating mappings of a cluster of (micro-)variables to a single (macro-)variable, while preserving some causal properties (Chalupka, Perona, and Eberhardt 2015; Chalupka, Eberhardt, and Perona 2016; Rubenstein et al. 2017; Beckers and Halpern 2019). The result is a new structural causal model defined on a higher level of abstraction, but with causal properties similar to those in the low-level model.<sup>1</sup> Other related works include chain graphs (Lauritzen and Richardson 2002) and ancestral causal graphs (Zhang 2008) developed to represent collections of causal diagrams equivalent under certain properties. By contrast, our work proposes a new graphical representation of a class of compatible causal diagrams, representing limited causal knowledge when the full structural causal model is unknown.

Causal discovery algorithms can be an alternative for when prior knowledge is insufficient to fully delineate a causal diagram (Pearl 2000; Spirtes, Glymour, and Scheines 2000; Peters, Janzing, and Schölkopf 2017). However, in general, it is impossible to fully recover the causal diagram based solely on observational data, without making strong assumptions about the underlying causal model, including causal sufficiency (all variables have been measured), the form of the functions (e.g., linearity, additive noise), and the distributions of the error terms (e.g. Gaussian, non-Gaussian, etc) (Glymour, Zhang, and Spirtes 2019). Then, there are cases where a meaningful causal diagram cannot be learned and prior knowledge is necessary for its construction. Our work focuses on establishing a language and corresponding machinery to encode partial knowledge and infer causal effects over clusters, alleviating some challenges in causal modeling in high-dimensional settings.

## 2 Preliminaries

**Notation.** A single 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. We use kinship relations, defined along the full edges in the graph, ignoring bidirected edges. We denote by  $Pa(\mathbf{X})_G$ ,  $An(\mathbf{X})_G$ , and  $De(\mathbf{X})_G$ , the sets of parents, ancestors, and descendants in  $G$ , respectively. A vertex  $V$  is said to be *active* on a path relative to  $\mathbf{Z}$  if 1)  $V$  is a collider and  $V$  or any of its descendants are in  $\mathbf{Z}$  or 2)  $V$  is a non-collider and is not in  $\mathbf{Z}$ . A path  $p$  is said to be *active* given (or conditioned on)  $\mathbf{Z}$  if every vertex on  $p$  is active relative to  $\mathbf{Z}$ . Otherwise,  $p$  is said to be *inactive*. Given a graph  $G$ ,  $\mathbf{X}$  and  $\mathbf{Y}$  are d-separated by  $\mathbf{Z}$  if every path between  $\mathbf{X}$  and  $\mathbf{Y}$  is inactive given  $\mathbf{Z}$ . We denote this d-separation by  $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G$ . The mutilated graph  $G_{\overline{\mathbf{XZ}}}$  is the result of removing from  $G$  edges with an arrowhead into  $\mathbf{X}$  (e.g.,  $A \rightarrow \mathbf{X}$ ,  $A \leftrightarrow \mathbf{X}$ ), and edges with a tail from  $\mathbf{Z}$  (e.g.,  $A \leftarrow \mathbf{Z}$ ).

**Structural Causal Models (SCMs)** 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 (mea-

<sup>1</sup>In (Beckers and Halpern 2019)’s notation, we investigate the case of a *constructive  $\tau$ -abstraction* where the mapping  $\tau$  only groups the low-level variables into high-level (cluster) variables.

	Level (Symbol)	Typical Activity	Typical Model	Typical Question
$\mathcal{L}_1$	Associational $P(y x)$	Seeing	BN	What is? How would seeing $X$ change my belief in $Y$ ?
$\mathcal{L}_2$	Interventional $P(y do(x), c)$	Doing	CBN	What if? What if I do $X$ ?
$\mathcal{L}_3$	Counterfactual $P(y_x x', y')$	Imagining	SCM	Why? What if I had acted differently?

Table 1: The Ladder of Causation / Pearl’s Causal Hierarchy

sured) 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 a function  $f_i \in \mathcal{F}$  of  $\mathbf{U}_i \cup Pa(V_i)$ , 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) with bidirected edges – or an acyclic directed mixed graph (ADMG) –  $G(\mathbf{V}, \mathbf{E})$ , known as a *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 edge ( $V_j \rightarrow V_i$ ) for every  $V_i \in \mathbf{V}$  and  $V_j \in Pa(V_i)$ , and there is a dashed bidirected edge ( $V_i \leftrightarrow V_j$ ) 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 intervention  $\mathbf{X} = \mathbf{x}$  is represented through the do-operator,  $do(\mathbf{X} = \mathbf{x})$ , which represents the operation of fixing a set  $\mathbf{X}$  to a constant  $\mathbf{x}$ , and 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  $P(\mathbf{v} \setminus \mathbf{x} | do(\mathbf{x}))$ . For any subset  $\mathbf{Y} \subseteq \mathbf{V}$ , the *potential response*  $\mathbf{Y}_{\mathbf{x}}(\mathbf{u})$  is defined as the solution of  $\mathbf{Y}$  in the submodel  $\mathcal{M}_{\mathbf{x}}$  given  $\mathbf{U} = \mathbf{u}$ .  $P(\mathbf{U})$  then induces a *counterfactual variable*  $\mathbf{Y}_{\mathbf{x}}$ .

**Pearl’s Causal Hierarchy (PCH) / The Ladder of Causation** (Pearl and Mackenzie 2018; Bareinboim et al. 2020) is a formal framework that divides inferential tasks into three different layers, namely, 1) associational, 2) interventional, and 3) counterfactual (see Table 1). An important result formalized under the rubric of the Causal Hierarchy Theorem (CHT) (Bareinboim et al. 2020, Thm. 1) states that inferences at any layer of the PCH almost never can be obtained by using solely information from lower layers.

### 3 C-DAGs: Definition and Properties

Standard causal inference tools typically require assumptions articulated through causal diagrams. We investigate the situations where the knowledge necessary to specify the underlying causal diagram  $G(\mathbf{V}, \mathbf{E})$  over the individual variables in  $\mathbf{V}$  may not be available. In particular, we assume that variables are grouped into a set of clusters of variables  $\mathbf{C}_1, \dots, \mathbf{C}_k$  that form a partition of  $\mathbf{V}$  (note that a variable may be grouped in a cluster by itself) such that we do not have knowledge about the relationships amongst the variables inside the clusters  $\mathbf{C}_i$  but we have some knowledge about the relationships between variables in different groups. We are interested in performing probabilistic and causal inferences about these clusters of variables; one may consider each cluster as defining a *macro-variable* and our aim is to reason about these macro-variables.

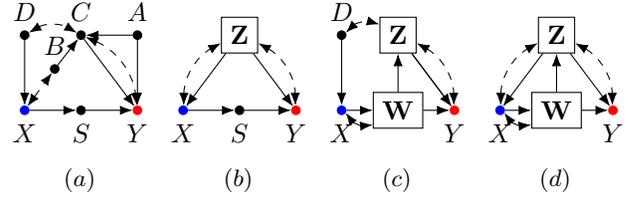


Figure 1: (a): a possible ADMG over lisinopril ( $X$ ), stroke ( $Y$ ), age ( $A$ ), blood pressure ( $B$ ), comorbidities ( $C$ ), medication history ( $D$ ), and sleep quality ( $S$ ). (b): a C-DAG of (a) with  $\mathbf{Z} = \{A, B, C, D\}$ . (c): a C-DAG of (a) with  $\mathbf{W} = \{S, B\}$ ,  $\mathbf{Z} = \{A, C\}$ . (d): an invalid C-DAG of (a), as the partition  $\{X, Y, \mathbf{W}, \mathbf{Z}\}$ , with  $\mathbf{W} = \{S, B\}$ ,  $\mathbf{Z} = \{A, C, D\}$  is inadmissible due to the cycle among  $(X, \mathbf{W}, \mathbf{Z})$ .

Formally, we address the following technical problem:

**Problem Statement:** Consider a causal diagram  $G^*$  over  $\mathbf{V}$  and a set of clusters of variables  $\mathbf{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_k\}$  forming a partition of  $\mathbf{V}$ . We aim to perform probabilistic, interventional, or counterfactual inferences about the macro-variables. Can we construct a causal diagram  $G_{\mathbf{C}}^*$  over the macro-variables in  $\mathbf{C}$  such that inferences by applying standard tools (d-separation, do-calculus, ID algorithm) on  $G_{\mathbf{C}}^*$  are valid in the sense they lead to the same conclusions as inferred on  $G^*$ ?

To this end, we propose a graphical object called *cluster DAGs* (or C-DAGs) to capture our partial knowledge about the underlying causal diagram over individual variables.

**Definition 1 (Cluster DAG or C-DAG).** Given an ADMG  $G(\mathbf{V}, \mathbf{E})$  (a DAG with bidirected edges) and a partition  $\mathbf{C} = \{\mathbf{C}_1, \dots, \mathbf{C}_k\}$  of  $\mathbf{V}$ , construct a graph  $G_{\mathbf{C}}(\mathbf{C}, \mathbf{E}_{\mathbf{C}})$  over  $\mathbf{C}$  with a set of edges  $\mathbf{E}_{\mathbf{C}}$  defined as follows:

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

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

Throughout the paper, we will use the same symbols (e.g.  $\mathbf{C}_i$ ) to represent both a cluster node in a C-DAG  $G_{\mathbf{C}}$  and the set of variables contained in the cluster.

**Remark 1.** The definition of C-DAGs does not allow for cycles in order to utilize standard graphical modeling tools that work only in DAGs. An inadmissible partition of  $\mathbf{V}$  means that the partial knowledge available for constructing  $G_{\mathbf{C}}$  is not enough for drawing conclusions using the tools developed in this paper.

**Remark 2.** Although a C-DAG is defined in terms of an underlying graph  $G$ , in practice, one will construct a C-DAG when complete knowledge of the graph  $G$  is unavailable. As an example of this construction, consider the model of the effect of lisinopril ( $X$ ) on the outcome of having a stroke ( $Y$ ) in Fig. 1(a). If not all the relationships specified in Fig. 1(a) are known, a data scientist cannot construct a full causal diagram, but may still have enough knowledge to create a C-DAG. For instance, they may have partial knowledge that the

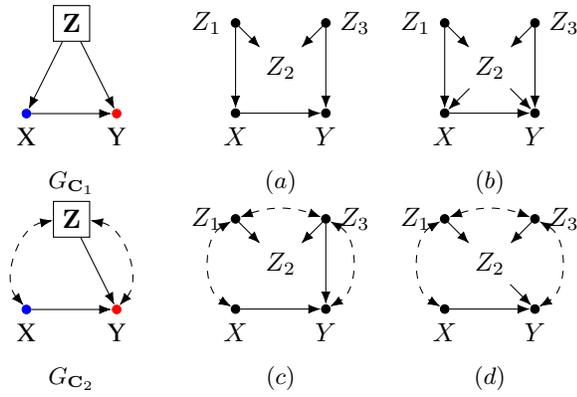


Figure 2:  $G_{C_1}$  is the C-DAG for diagrams (a) and (b) and  $G_{C_2}$  is the C-DAG for diagrams (c) and (d), where  $\mathbf{Z} = \{Z_1, Z_2, Z_3\}$ .  $P(y|do(x))$  is identifiable in  $G_{C_1}$  by backdoor adjustment over  $\mathbf{Z}$  and is not identifiable in  $G_{C_2}$ .

covariates occur temporally before lisinopril is prescribed, or that a stroke occurs and the suspicion that some of the pre-treatment variables are causes of  $X$  and  $Y$ . Specifically, they can create the cluster  $\mathbf{Z} = \{A, B, C, D\}$  with all the covariates, and then construct a C-DAG with edges  $\mathbf{Z} \rightarrow X$  and  $\mathbf{Z} \rightarrow Y$ . Further, the data scientist may also suspect that some of the variables in  $\mathbf{Z}$  are confounded with  $X$  and others with  $Y$ , an uncertainty that is encoded in the C-DAG through the bidirected edges  $\mathbf{Z} \leftrightarrow X$  and  $\mathbf{Z} \leftrightarrow Y$ . With the additional knowledge that sleep quality ( $S$ ) acts as a mediator between the treatment and outcome, the C-DAG in Fig. 1(b) can be constructed. Note that this C-DAG is consistent with the true underlying causal diagram in Fig. 1(a), but was constructed without knowing this diagram and using much less knowledge than what is encoded in it. Alternatively, if clusters  $\mathbf{W} = \{S, B\}$  and  $\mathbf{Z} = \{A, C\}$  are created, then the C-DAG shown in Fig. 1(c) would be constructed. Note that both (a) and (b) are considered valid C-DAGs because no cycles are created. Finally, if a clustering with  $\mathbf{W} = \{S, B\}$  and  $\mathbf{Z} = \{A, C, D\}$  is created, this would lead to the C-DAG shown in Fig. 1(d), which is invalid. The reason is that a cycle  $X \rightarrow \mathbf{W} \rightarrow \mathbf{Z} \rightarrow X$  is created due to the connections  $X \rightarrow S$ ,  $B \rightarrow C$ , and  $D \rightarrow X$  in the diagram (a).

**Remark 3.** It is important to note that a C-DAG  $G_C$  as defined in Def. 1 is merely a graph over clusters of nodes  $C_1, \dots, C_k$ , and does not have a priori the semantics and properties of a BN or CBN over macro-variables  $C_i$ . It's not clear, for example, whether the cluster nodes  $C_i$  satisfy the Markov properties w.r.t. the graph  $G_C$ . Rather, a C-DAG can be seen as a graphical representation of an equivalence class (EC, for short) of graphs that share the relationships among the clusters while allowing for any possible relationships among the variables within each cluster. For instance, in Fig. 2, the diagrams (a) and (b) can be represented by C-DAG  $G_{C_1}$  (top) and can, therefore, be thought of as being members of an EC represented by  $G_{C_1}$ . The same can be concluded for diagrams (c) and (d), both represented by C-DAG  $G_{C_2}$ . The graphical representation of this ECs are shown in Fig. 3, where on the left we have the space of all possible ADMGs, and on the right the space of C-DAGs.

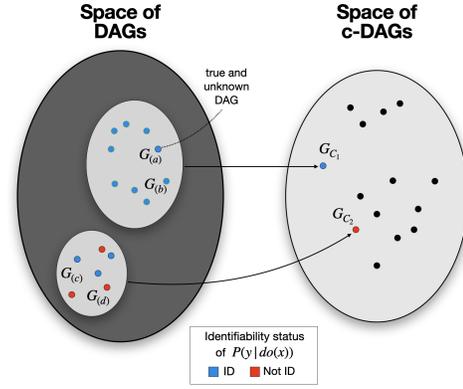


Figure 3: Identifying  $P(y|do(x))$  in a C-DAG means identifying such an effect for the entire equivalence class. In  $G_{C_1}$ , the effect is identifiable (blue) because it is identifiable in  $G_{(a)}$ ,  $G_{(b)}$ , and all the other causal diagrams represented. In  $G_{C_2}$ , the same effect is non-identifiable (red), as the encoded partial knowledge is compatible with some causal diagrams in which the effect is not identifiable (e.g.,  $G_{(d)}$ ).

Given the semantics of a C-DAG as an equivalence class of ADMGs, what valid inferences can one perform about the cluster variables given a C-DAG  $G_C$ ? What properties of C-DAGs are shared by all compatible ADMGs?

In principle, definite conclusions from C-DAGs can only be drawn about properties shared among all EC's members. Going back to Fig. 3, we identify an effect in a C-DAG (e.g.,  $G_{C_1}$  in Fig. 2) whenever this effect is identifiable in all members of the EC; e.g., causal diagrams (a), (b), and all other diagrams compatible with  $G_{C_1}$ . Note that in this particular EC, all dots are marked with blue, which means that the effect is identifiable in each one of them. On the other hand, if there exists one diagram in the EC where the effect is not identifiable, this effect will not be identifiable in the corresponding C-DAG. For instance, the effect is not identifiable in the C-DAG  $G_{C_2}$  due to diagram (d) in Fig. 2.

Once the semantics of C-DAGs is well-understood, we turn our attention to computational issues. One naive approach to causal inference with cluster variables, e.g. identifying  $Q = P(C_i|do(C_k))$ , goes as follows – first enumerate all causal diagrams compatible with  $G_C$ ; then, evaluate the identifiability of  $Q$  in each diagram; finally, output  $P(C_i|do(C_k))$  if all the diagrams entail the same answer, otherwise output “non-identifiable”. However, in practice, this approach is intractable in high-dimensional settings – given a cluster  $C_i$  of size  $m$ , the number of possible causal diagrams over the variables in  $C_i$  is super-exponential in  $m$ . Can valid inferences be performed about cluster variables using C-DAGs directly, without going through exhaustive enumeration? What properties of C-DAGs are shared by all the compatible causal diagrams? The next sections present theoretical results to address these questions.

Finally, note that not all properties of C-DAGs are shared across all compatible diagrams. To illustrate, consider the case of *backdoor paths*, i.e., paths between  $X$  and  $Y$  with an arrowhead into  $X$ , in Fig. 2. The path  $X \leftrightarrow \mathbf{Z} \rightarrow Y$

in  $G_{C_2}$  is active when not conditioning on  $\mathbf{Z}$ . However, the corresponding backdoor paths in diagram (c) are all inactive. Therefore, a d-connection in a C-DAG does not necessarily correspond to a d-connection in all diagrams in the EC.

#### 4 C-DAGs for $\mathcal{L}_1$ -Inferences

In this section, we study probabilistic inference with C-DAGs -  $\mathcal{L}_1$  inferences. We assume the underlying graph  $G$  over  $\mathbf{V}$  is a Bayesian Network (BN) *with no causal interpretation*.<sup>2</sup> We aim to perform probabilistic inferences about macro-variables with  $G_C$  that are valid in  $G$  regardless of the unknown relationships within each cluster.

First, we extend d-separation (Pearl 1988), a fundamental tool in probabilistic reasoning in BNs, to C-DAGs. As noticed earlier, a d-connecting path in a C-DAG does not necessarily imply that the corresponding paths in a compatible ADMG  $G$  are connecting. Such paths can be either active or inactive. However, d-separated paths in a C-DAG correspond to only d-separated paths in all compatible ADMGs.<sup>3</sup> These observations together, lead to the following definition where the symbol  $*$  represents either an arrow head or tail:

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

1.  $C_i * \rightarrow C_m \rightarrow C_j$  such that the non-collider cluster  $C_m$  is in  $\mathbf{Z}$ , or
2.  $C_i * \rightarrow C_m \leftarrow * C_j$  such that the collider cluster  $C_m$  and its descendants are not 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  $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{G_C}$ , if and only if  $\mathbf{Z}$  blocks every path from a cluster in  $\mathbf{X}$  to a cluster in  $\mathbf{Y}$ .

We show in the following proposition that the d-separation rules are sound and complete in C-DAGs in the following sense: whenever a d-separation holds in a C-DAG, it holds for all ADMGs compatible with it; on the other hand, if a d-separation does not hold in a C-DAG, then there exists at least one ADMG compatible with it for which the same d-separation statement does not hold.

**Theorem 1. (Soundness and completeness of d-separation in C-DAGs).** *In a C-DAG  $G_C$ , let  $\mathbf{X}, \mathbf{Z}, \mathbf{Y} \subset \mathbf{C}$ . If  $\mathbf{X}$  and  $\mathbf{Y}$  are d-separated by  $\mathbf{Z}$  in  $G_C$ , then, in any ADMG  $G$  compatible with  $G_C$ ,  $\mathbf{X}$  and  $\mathbf{Y}$  are d-separated by  $\mathbf{Z}$  in  $G$ , i.e.,*

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{G_C} \implies (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_G. \quad (1)$$

*If  $\mathbf{X}$  and  $\mathbf{Y}$  are not d-separated by  $\mathbf{Z}$  in  $G_C$ , then, there exists an ADMG  $G$  compatible with  $G_C$  where  $\mathbf{X}$  and  $\mathbf{Y}$  are not d-separated by  $\mathbf{Z}$  in  $G$ .*

Theorem 1 implies that  $G_C$  does not imply any conditional independence that is not implied by the underlying  $G$ .

ADMGs are commonly used to represent BNs with latent variables (which may imply Verma constraints on  $P(v)$ )

<sup>2</sup>For a more detailed discussion on the tension between layers  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , please refer to (Bareinboim et al. 2020, Sec. 1.4.1).

<sup>3</sup>In Appendix ??, we investigate in detail how path analysis is extended to C-DAGs. We note that d-separation in ADMGs has also been called m-separation (Richardson 2003).

not captured by independence relationships (Tian and Pearl 2002b)) where the observational distribution  $P(\mathbf{v})$  factorizes according to the ADMG  $G$  as follows

$$P(\mathbf{v}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k: V_k \in \mathbf{V}} P(v_k | pa_{v_k}, \mathbf{u}_k), \quad (2)$$

where  $Pa_{V_k}$  are the endogenous parents of  $V_k$  in  $G$  and  $\mathbf{U}_k \subseteq \mathbf{U}$  are the latent parents of  $V_k$ . We show next that the observational distribution  $P(\mathbf{v})$  factorizes according to the graphical structure of the C-DAG  $G_C$  as well.

**Theorem 2. (C-DAG as BN).** *Let  $G_C$  be a C-DAG compatible with an ADMG  $G$ . If the observational distribution  $P(\mathbf{v})$  factorizes according to  $G$  by Eq. (2), then the observational distribution  $P(\mathbf{v}) = P(\mathbf{c})$  factorizes according to  $G_C$ , i.e.,*

$$P(\mathbf{c}) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k: C_k \in \mathbf{C}} P(\mathbf{c}_k | pa_{C_k}, \mathbf{u}'_k), \quad (3)$$

where  $Pa_{C_k}$  are the parents of the cluster  $C_k$ , and  $\mathbf{U}'_k \subseteq \mathbf{U}$  such that, for any  $i, j$ ,  $\mathbf{U}'_i \cap \mathbf{U}'_j \neq \emptyset$  if and only if there is a bidirected edge  $(C_i \leftrightarrow C_j)$  between  $C_i$  and  $C_j$  in  $G_C$ .

Thm. 2 implies that if the underlying ADMG  $G$  represents a BN with latent variables over  $\mathbf{V}$ , then the C-DAG  $G_C$  represents a BN over micro-variables  $\mathbf{C}$ .

#### 5 C-DAGs for $\mathcal{L}_2$ -Inferences

We study now interventional ( $\mathcal{L}_2$ ) reasoning with C-DAGs. We assume the underlying graph  $G$  over  $\mathbf{V}$  is a CBN. Our goal is to perform causal reasoning about macro-variables with  $G_C$  that are guaranteed to be valid in each  $G$  of the underlying EC. We focus on extending to C-DAGs Pearl's celebrated *do-calculus* (Pearl 1995) and the ID-algorithm (Tian 2002; Shpitser and Pearl 2006; Huang and Valtorta 2006).

##### Do-Calculus in C-DAGs

Do-calculus is a fundamental tool in causal inference from causal diagrams and has been used extensively for solving a variety of identification tasks. We show that if the underlying ADMG  $G$  is a CBN on which do-calculus rules hold, then do-calculus rules are valid in the corresponding C-DAG  $G_C$ . We first present a key lemma for proving the soundness of do-calculus in C-DAGs that the mutilation operations in a C-DAG to create  $G_{C_{\underline{\mathbf{x}}}}$  and  $G_{C_{\overline{\mathbf{x}}}}$  carry over to all compatible underlying ADMGs. This result is shown in the following:

**Lemma 1.** *If a C-DAG  $G_C$  is compatible with an ADMG  $G$ , then, for  $\mathbf{X}, \mathbf{Z} \subset \mathbf{C}$ , the mutilated C-DAG  $G_{C_{\overline{\mathbf{x}}\mathbf{z}}}$  is compatible with the mutilated ADMG  $G_{\overline{\mathbf{x}}\mathbf{z}}$ .*

The soundness of do-calculus in C-DAGs as stated next follows from Theorem 1 and Lemma 1.

**Theorem 3. (Do-calculus in causal C-DAGs).** *Let  $G_C$  be a C-DAG compatible with an ADMG  $G$ . If  $G$  is a CBN encoding interventional distributions  $P(\cdot | do(\cdot))$ , then for any disjoint subsets of clusters  $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W} \subset \mathbf{C}$ , the following three rules hold:*

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

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

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

where  $G_{C_{\overline{\mathbf{XZ}}}}$  is obtained from  $G_C$  by removing the edges into  $\mathbf{X}$  and out of  $\mathbf{Z}$ , and  $\mathbf{Z}(\mathbf{W})$  is the set of  $\mathbf{Z}$ -clusters that are non-ancestors of any  $\mathbf{W}$ -cluster in  $G_{C_{\overline{\mathbf{XZ}}}}$ .

We also show next that the do-calculus rules in C-DAGs are complete in the following sense:

**Theorem 4. (Completeness of do-calculus).** *If in a C-DAG  $G_C$  a do-calculus rule does not apply, then there is a CBN  $G$  compatible with  $G_C$  for which it also does not apply.*

### Truncated Factorization

An ADMG  $G$  represents a CBN if the interventional distributions factorizes according to the graphical structure, known as the truncated factorization, i.e., for any  $\mathbf{X} \subseteq \mathbf{V}$

$$P(\mathbf{v} \setminus \mathbf{x} | do(\mathbf{x})) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k: V_k \in \mathbf{V} \setminus \mathbf{X}} P(v_k | pa_{V_k}, \mathbf{u}_k), \quad (4)$$

where  $Pa_{V_k}$  are the endogenous parents of  $V_k$  in  $G$  and  $\mathbf{U}_k \subseteq \mathbf{U}$  are the latent parents of  $V_k$ .

We show that the truncated factorization holds in C-DAGs as if the underlying ADMG is a CBN, in the following sense.

**Theorem 5. (C-DAG as CBN)** *Let  $G_C$  be a C-DAG compatible with an ADMG  $G$ . If  $G$  satisfies the truncated factorization (4) with respect to the interventional distributions, then, for any  $\mathbf{X} \subseteq \mathbf{C}$ , the interventional distribution  $P(\mathbf{c} \setminus \mathbf{x} | do(\mathbf{x}))$  factorizes according to  $G_C$ , i.e.,*

$$P(\mathbf{c} \setminus \mathbf{x} | do(\mathbf{x})) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k: C_k \in \mathbf{C} \setminus \mathbf{X}} P(c_k | pa_{C_k}, \mathbf{u}'_k), \quad (5)$$

where  $Pa_{C_k}$  are the parents of the cluster  $C_k$ , and  $\mathbf{U}'_k \subseteq \mathbf{U}$  such that, for any  $i, j$ ,  $\mathbf{U}'_i \cap \mathbf{U}'_j \neq \emptyset$  if and only if there is a bidirected edge  $(C_i \leftrightarrow C_j)$  between  $C_i$  and  $C_j$  in  $G_C$ .

Theorem 5 essentially shows that a C-DAG  $G_C$  can be treated as a CBN over the macro-variables  $\mathbf{C}$  if the underlying ADMG is a CBN.

### ID-Algorithm

Equipped with d-separation, do-calculus, and the truncated factorization in C-DAGs, causal inference algorithms developed for a variety of tasks that rely on a known causal diagram can be extended to C-DAGs (Bareinboim and Pearl 2016). In this paper, we consider the problem of identifying causal effects from observational data using C-DAGs.

There exists a complete algorithm to determine whether  $P(y|do(x))$  is identifiable from a causal diagram  $G$  and the observational distribution  $P(\mathbf{V})$  (Tian 2002; Shpitser and Pearl 2006; Huang and Valorta 2006). This algorithm, or ID-algorithm for short, is based on the truncated factorization, therefore, Theorem 5 allows us to prove that the ID-algorithm is sound and complete to systematically infer causal effects from the observational distribution  $P(\mathbf{V})$  and partial domain knowledge encoded as a C-DAG  $G_C$ .

**Theorem 6. (Soundness and Completeness of ID-algorithm).** *The ID-algorithm is sound and complete when applied to a C-DAG  $G_C$  for identifying causal effects of the form  $P(y|do(x))$  from the observational distribution  $P(\mathbf{V})$ , where  $\mathbf{X}$  and  $\mathbf{Y}$  are sets of clusters in  $G_C$ .*

The ID algorithm returns a formula for identifiable  $P(y|do(x))$  that is valid in all causal diagrams compatible with the C-DAG  $G_C$ . The completeness result ensures that if the ID-algorithm fails to identify  $P(y|do(x))$  from  $G_C$ , then there exists a causal diagram  $G$  compatible with  $G_C$  where the effect  $P(y|do(x))$  is not identifiable. Appendix ?? contains an experimental study evaluating the ability of C-DAGs to accurately assess the identifiability of effects while requiring less domain knowledge for their construction.

### Examples of Causal Identifiability in C-DAGs

We show examples of identification in C-DAGs in practice. Besides illustrating identification of causal effects in the coarser graphical representation of a C-DAG, these examples demonstrate that clustering variables may lead to diagrams where effects are not identifiable. Therefore, care should be taken when clustering variables, to ensure not so much information is lost in a resulting C-DAG, such that identifiability is maintained when possible.

**Identification in Fig. 1.** In diagram (a) the effect of  $X$  on  $Y$  is identifiable through backdoor adjustment (Pearl 2000, pp. 79-80) over the set of variables  $\{B, D\}$  In the C-DAG in Fig. 1(b), with cluster  $\mathbf{Z} = \{A, B, C, D\}$ , the effect of  $X$  on  $Y$  is identifiable through front-door adjustment (Pearl 2000, p. 83) over  $S$ , given by  $P(y|do(x)) = \sum_s P(s|x) \sum_{x'} P(y|x', s) P(x')$ . Because this front-door adjustment holds for the C-DAG in Fig. 1(b) with which diagram (a) is compatible, this front-door adjustment identification formula is equivalent to the adjustment in the case of diagram (a) and gives the correct causal effect in any other compatible causal diagram. In the C-DAG in (c), the loss of separations from the creation of clusters  $\mathbf{Z} = \{A, B, C, D\}$  and  $\mathbf{W} = \{B, S\}$  render the effect no longer identifiable, indicating that there exists another graph compatible with (c) for which the effect cannot be identified.

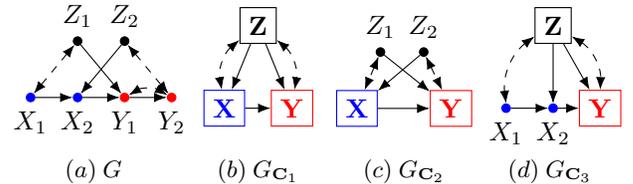


Figure 4: (a): causal diagram  $G$  where the effect  $P(y_1, y_2 | do(x_1, x_2))$  is identifiable. (b): C-DAG  $G_{C_1}$  with clustering  $\mathbf{X} = \{X_1, X_2\}$ ,  $\mathbf{Y} = \{Y_1, Y_2\}$ , and  $\mathbf{Z} = \{Z_1, Z_2\}$ . (c): C-DAG  $G_{C_2}$  with clustering  $\mathbf{X} = \{X_1, X_2\}$  and  $\mathbf{Y} = \{Y_1, Y_2\}$ . (d): C-DAG  $G_{C_3}$  with clustering  $\mathbf{Y} = \{Y_1, Y_2\}$  and  $\mathbf{Z} = \{Z_1, Z_2\}$ . The effect  $P(y|do(x))$  is not identifiable in  $G_{C_1}$ , but is identifiable in  $G_{C_2}$  and  $P(y|do(x_1, x_2))$  is identifiable in  $G_{C_3}$ .

**Identification in Fig. 4.** In causal diagram (a), the ef-

fect of  $\{X_1, X_2\}$  on  $\{Y_1, Y_2\}$  is identifiable by backdoor adjustment over  $\{Z_1, Z_2\}$  as follows:  $P(y_1, y_2|do(x_1, x_2)) = \sum_{z_1, z_2} P(y_1, y_2|x_1, x_2, z_1, z_2)P(z_1, z_2)$ . Note, however, that the backdoor path cannot be blocked in the C-DAG  $G_1$  (b) with clusters  $\mathbf{X} = \{X_1, X_2\}$ ,  $\mathbf{Y} = \{Y_1, Y_2\}$ , and  $\mathbf{Z} = \{Z_1, Z_2\}$ . In this case, the effect  $P(y|do(\mathbf{x}))$  is not identifiable. If the covariates  $Z_1$  and  $Z_2$  are not clustered together as shown in the C-DAG  $G_{C_2}$  (c), the backdoor paths relative to  $\mathbf{X}$  and  $\mathbf{Y}$  can still be blocked despite the unobserved confounders between  $Z_1$  and  $\mathbf{X}$  and between  $Z_2$  and  $\mathbf{Y}$ . So the effect  $P(y|do(\mathbf{x}))$  is identifiable by backdoor adjustment over  $\{Z_1, Z_2\}$  as follows:  $P(y|do(\mathbf{x})) = \sum_{z_1, z_2} P(y|\mathbf{x}, z_1, z_2)P(z_1, z_2)$ . If the treatments  $X_1$  and  $X_2$  are not clustered together as shown in the C-DAG  $G_{C_3}$  (d), then the joint effect of  $X_1$  and  $X_2$  on the cluster  $\mathbf{Y}$  is identifiable and given by the following expression:  $P(y|do(x_1, x_2)) = \sum_{z, x'_1} P(y|x'_1, x_2, z)P(x'_1, z)$ .

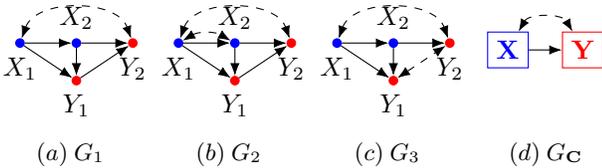


Figure 5: (a), (b), and (c) are causal diagrams compatible with the C-DAG  $G_C$  in (d) where  $\mathbf{X} = \{X_1, X_2\}$  and  $\mathbf{Y} = \{Y_1, Y_2\}$ . The causal effect  $P(y_1, y_2|do(x_1, x_2))$  is identifiable in (a) but not in (b) or (c). Consequently, the effect  $P(y|do(\mathbf{x}))$  is not identifiable from the C-DAG  $G_C$ .

**Identification in Fig. 5.** In the causal diagram (a), the effect of the joint intervention to  $\{X_1, X_2\}$  on both outcomes  $\{Y_1, Y_2\}$  is identifiable as follows:  $P(y_1, y_2|do(x_1, x_2)) = P(y_1|x_1, x_2) \sum_{x'_1} P(y_2|x'_1, x_2, y_1)P(x'_1)$ . By clustering the two treatments as  $\mathbf{X}$  and the two outcomes as  $\mathbf{Y}$ , we lose the information that  $X_2$  is not a confounded effect of  $X_1$  and that  $Y_1$  and  $Y_2$  are not confounded. If this is the case, as in causal diagrams  $G_2$  (b) and  $G_3$  (c), the effect would not be identifiable. Note that the C-DAG (d), representing causal diagrams (a), (b), and (c), is the bow graph, where the effect  $P(y|do(\mathbf{x}))$  is also not identifiable.

## 6 C-DAGs for $\mathcal{L}_3$ -Inferences

Now, we study counterfactual ( $\mathcal{L}_3$ ) inferences in C-DAGs. We assume that the underlying graph  $G$  over  $\mathbf{V}$  is induced by an SCM  $\mathcal{M}$ , and our goal is to perform counterfactual reasoning about macro-variables with  $G_C$  that are always valid in  $\mathcal{M}$  (while both  $G$  and the SCM  $\mathcal{M}$  are unknown).

We show that for any SCM over  $\mathbf{V}$  with causal diagram  $G$ , there is an equivalent SCM over macro-variables  $\mathbf{C}$  that induces C-DAG  $G_C$  and makes the same predictions about counterfactual distributions over the macro-variables.

**Theorem 7.** *Let  $G_C$  be a C-DAG compatible with an ADMG  $G$ . Assume  $G$  is induced by an SCM  $\mathcal{M}$ , then there exists an SCM  $\mathcal{M}_C$  over macro-variables  $\mathbf{C}$  such that its induced causal diagram is  $G_C$  and, for any set of counterfactual variables  $\mathbf{Y}_x, \dots, \mathbf{Z}_w$  where  $\mathbf{Y}, \mathbf{X}, \dots, \mathbf{Z}, \mathbf{W} \subseteq \mathbf{C}$ ,*

$$P_{\mathcal{M}}(\mathbf{y}_x, \dots, \mathbf{z}_w) = P_{\mathcal{M}_C}(\mathbf{y}_x, \dots, \mathbf{z}_w).$$

Following this result, algorithms developed for a variety of counterfactual inference tasks that rely on a known causal diagram, such as the CTFID algorithm (Correa, Lee, and Bareinboim 2021), can be used in the context of C-DAGs.

For example, consider the C-DAG  $G_{C_1}$  in Fig. 2, where  $X$  is a drug,  $Y$  is a disease, and  $\mathbf{Z}$  is a cluster of factors potentially affecting  $X$  and  $Y$ . Suppose that a patient who took the drug ( $X = 1$ ) would like to know what his chances of being cured ( $Y = 1$ ) would have been had he not taken the drug ( $X = 0$ ). This quantity is defined by  $P(Y_{X=0} = 1|X = 1)$ . The CTFID algorithm applied to  $G_{C_1}$  will conclude that  $P(Y_{X=0} = 1|X = 1)$  is identifiable and given by  $\sum_z P(Y = 1|X = 0, Z = z)P(Z = z|X = 1)$ . This formula is correct in all ADMGs compatible with  $G_{C_1}$ , regardless of the relationships within cluster  $Z$ .

After all, we note that inferences in the lower layers assume less knowledge than the higher layers. On the one hand, some results about the lower layers are implied by the higher layers. For instance, if  $G$  is induced by an SCM, then  $G$  represents a CBN and a BN, and Thm. 7 implies that if the underlying  $G$  is induced by an SCM, then  $G_C$  represents a CBN and a BN. If  $G$  is a CBN, then it is necessarily a BN, therefore Thm. 5 implies that if the underlying  $G$  represents a CBN, then  $G_C$  represents a BN. On the other hand, if one does not want to commit to the SCM generative process, but can only ascertain that the truncated factorization holds (e.g., for  $\mathcal{L}_2$ ), it's still possible to leverage the machinery developed without any loss of inferential power or making unnecessary assumptions about the upper layers.

## 7 Conclusions

Causal diagrams provide an intuitive language for specifying the necessary assumptions for causal inferences. Despite all their power and successes, the substantive knowledge required to construct a causal diagram – i.e., the causal and confounded relationships among all pairs of variables – is unattainable in some critical settings found across society, including in the health and social sciences. This paper introduces a new class of graphical models that allow for a more relaxed encoding of knowledge. In practice, when a researcher does not fully know the relationships among certain variables, under some mild assumptions delineated by Def. 1, these variables can be clustered together. (A causal diagram is an extreme case of a C-DAG where each cluster has exactly one variable.) We prove fundamental results to allow causal inferences within C-DAG's equivalence class, which translate to statements about all diagrams compatible with the encoded constraints. We develop the formal machinery for probabilistic, interventional, and counterfactual reasoning in C-DAGs following Pearl's hierarchy assuming the (unknown) underlying model over individual variables are BN ( $\mathcal{L}_1$ ), CBN ( $\mathcal{L}_2$ ), and SCM ( $\mathcal{L}_3$ ), respectively. These results are critical for enabling C-DAGs use in ways comparable to causal diagrams. We hope these new tools will allow researchers to represent complex systems in a simplified way, allowing for more relaxed causal inferences when substantive knowledge is largely unavailable and coarse.

## Acknowledgements

This work was done in part while Jin Tian was visiting the Simons Institute for the Theory of Computing. Jin Tian was partially supported by NSF grant IIS-2231797. This research was supported by the NSF, ONR, AFOSR, DoE, Amazon, JP Morgan, The Alfred P. Sloan Foundation and the United States NLM T15LM007079.

## References

- Bareinboim, E.; Correa, J. D.; Ibeling, D.; and Icard, T. 2020. On Pearl's Hierarchy and the Foundations of Causal Inference. Technical Report R-60, Causal Artificial Intelligence Lab, Columbia University.
- Bareinboim, E.; and Pearl, J. 2016. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27): 7345–7352.
- Beckers, S.; and Halpern, J. Y. 2019. Abstracting Causal Models. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, AAAI'19.
- Chalupka, K.; Eberhardt, F.; and Perona, P. 2016. Multi-Level Cause-Effect Systems. In Gretton, A.; and Robert, C. C., eds., *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, 361–369. Cadiz, Spain: PMLR.
- Chalupka, K.; Perona, P.; and Eberhardt, F. 2015. Visual Causal Feature Learning. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, UAI'15, 181–190. Arlington, Virginia, USA: AUAI Press.
- Correa, J.; Lee, S.; and Bareinboim, E. 2021. Nested Counterfactual Identification from Arbitrary Surrogate Experiments. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P. S.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 6856–6867. Curran Associates, Inc.
- Glymour, C.; Zhang, K.; and Spirtes, P. 2019. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics*, 10.
- Huang, Y.; and Valtorta, M. 2006. Pearl's Calculus of Intervention Is Complete. In T.S. Richardson, R. D. a., ed., *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 217–224. AUAI Press.
- Kleinberg, S.; and Hripcsak, G. 2011. A review of causal inference for biomedical informatics. *Journal of Biomedical Informatics*, 44(6): 1102–1112.
- Lauritzen, S. L.; and Richardson, T. S. 2002. Chain graph models and their causal interpretations. *Royal Statistical Society*, 64(Part 2): 1–28.
- Lee, S.; Correa, J. D.; and Bareinboim, E. 2019. General Identifiability with Arbitrary Surrogate Experiments. In *Proceedings of the Thirty-Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence*. AUAI Press.
- Parviainen, P.; and Kaski, S. 2016. Bayesian Networks for Variable Groups. In Antonucci, A.; Corani, G.; and Campos, C. P., eds., *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, volume 52 of *Proceedings of Machine Learning Research*, 380–391. PMLR.
- Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. San Mateo, CA: Morgan Kaufmann.
- Pearl, J. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4): 669–688.
- Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. NY, USA: Cambridge University Press, 2nd edition.
- Pearl, J.; and Mackenzie, D. 2018. *The Book of Why*. New York: Basic Books.
- Peters, J.; Janzing, D.; and Schölkopf, B. 2017. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning. Cambridge, MA: MIT Press.
- Richardson, T. 2003. Markov Properties for Acyclic Directed Mixed Graphs. *Scandinavian Journal of Statistics*, 30(1): 145–157.
- Rubenstein, P. K.; Weichwald, S.; Bongers, S.; Mooij, J. M.; Janzing, D.; Grosse-Wentrup, M.; and Schölkopf, B. 2017. Causal Consistency of Structural Equation Models. In Elidan, G.; Kersting, K.; and Ihler, A. T., eds., *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017*. AUAI Press.
- Schölkopf, B.; Locatello, F.; Bauer, S.; Ke, N. R.; Kalchbrenner, N.; Goyal, A.; and Bengio, Y. 2021. Toward Causal Representation Learning. *Proceedings of the IEEE*, 109(5): 612–634.
- Shen, Y.; Choi, A.; and Darwiche, A. 2018. Conditional PSDDs: Modeling and Learning With Modular Knowledge. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).
- Shpitser, I.; and Pearl, J. 2006. Identification of Joint Interventional Distributions in Recursive semi-Markovian Causal Models. In *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence*, 1219–1226.
- Spirtes, P.; Glymour, C. N.; and Scheines, R. 2000. *Causation, Prediction, and Search*. Cambridge, MA: MIT Press, 2nd edition.
- Tian, J. 2002. *Studies in Causal Reasoning and Learning*. Ph.D. thesis, Computer Science Department, University of California, Los Angeles, CA.
- Tian, J.; and Pearl, J. 2002a. A General Identification Condition for Causal Effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, 567–573. Menlo Park, CA: AAAI Press/MIT Press.
- Tian, J.; and Pearl, J. 2002b. On the Testable Implications of Causal Models with Hidden Variables. *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-02)*, 519–527.
- Zhang, J. 2008. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16): 1873–1896.