TECHNICAL REPORT
R-77
First version: Jun, 2021
Last version: May, 2022

Causal Effect Identification in Cluster DAGs

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 in the empirical sciences is to determine the effect of interventions from observational data. It is well-known that assumptions are necessary to perform causal inferences, which are commonly articulated through causal diagrams [13]. Despite the power of this approach, there are settings where the knowledge necessary to specify a causal diagram over all observed variables may not be 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 causal inferences over C-DAGs. In particular, we first prove the soundness and completeness of the d-separation rules extended to C-DAGs. Secondly, we prove the validity of Pearl's do-calculus inference rules over C-DAGs. Lastly, we show that a standard identification algorithm is sound and complete to systematically compute causal effects from observational data given a C-DAG.

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 [13, 19, 2, 15, 14].

Standard tools necessary for identifying the aforementioned do-distribution, such as d-separation, do-calculus [12], and the ID-algorithm [21, 18, 7, 10] take as input a combination of an observational distribution and a qualitative descripion of the underlying causal system, often articulated in the form of a causal diagram [13]. 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 use of the graphical approach to inferring causality [8]. In many cases, however, contextual or temporal information about variables is available, which may partially inform how these variables are situated

*These authors contributed equally to this work.

Preprint. Under review.

in a causal diagram relative to other key variables. For instance, 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 identification of causal effects 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 specify the full causal diagram or causal relationships over all variables. First, we formally define and characterize a novel class of graphs 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 causal inference, 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. These are fundamental first steps in terms of causal semantics and graphical conditions to perform causal 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 a set of clusters of variables where the relationships amongst the variables inside the clusters are left unspecified *(Definition 1)*. Semantically, a C-DAG represents an equivalence class of all compatible causal diagrams that share the relationships among the clusters.

2. We prove the soundness and completeness of Pearl's d-separation and do-calculus rules extended to the coarse graphical representation of C-DAGs *(Theorems 1, 2 and 3)*, despite C-DAGs' abstracted representation of possibly numerous underlying causal diagrams.

3. We prove that interventional distributions admit a convenient factorization following the C-DAG's structure *(Theorem 4)*. This factorization 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 5)*.

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 [17]. 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 [5, 4, 16, 3]. 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. Other related works include chain graphs [9] and ancestral causal graphs [22] 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 background knowledge is not sufficient to fully delineate a causal diagram [13, 19, 15]. However, in general, it is not possible to fully recover the causal diagram based solely on observational data, without making strong assumptions about the underlying causal model, including causal sufficiency (all relevant 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) [6]. Then, there are cases where a meaningful causal diagram cannot be learned and background 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 a set $\mathbf{Z}$ 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 $G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$ is the result of removing from G edges coming into variables in $\mathbf{X}$ and going out of variables in $\mathbf{Z}$.

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 (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 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) $G(\mathbf{V}, \mathbf{E})$ with bidirected edges, 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 $(V_j \to V_i)$ for every $V_i \in \mathbf{V}$ and $V_j \in Pa(V_i)$, and there is a dashed bidirected edge $(V_j \dashleftarrow\dashrightarrow V_i)$ 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 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 $P(\mathbf{v} \setminus \mathbf{x} | do(\mathbf{x}))$.

3 C-DAGs: Definition and Properties

Standard causal inference tools typically require causal assumptions articulated through causal diagrams. We investigate the situations where the knowledge necessary to specify a 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 $\mathbf{C}_1, \ldots, \mathbf{C}_k$ 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.

To this end, we formally introduce a graphical object called *cluster DAGs* (or C-DAGs) to capture our partial knowledge, which is a coarser representation of a causal diagram:

Definition 1 (Cluster DAG or C-DAG). Given a causal diagram $G(\mathbf{V}, \mathbf{E})$ and 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 edges $\mathbf{E}_{\mathbf{C}}$ defined as follows:

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

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 in a compatible causal diagram G.

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.

Interestingly, a causal diagram is a C-DAG where each variable forms its own cluster, which means that all clusters are of size one. In practice, however, the relationships among every pair of variables

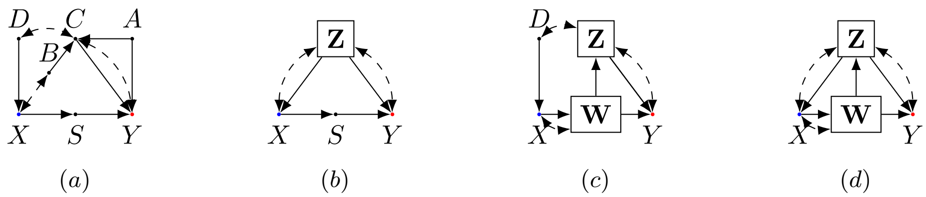


Figure 1: (a): a possible causal diagram over lisinopril (X), stroke (Y), age (A), blood pressure (B), comorbidities (C), medication history (D), and sleep quality (S). (b): a C-DAG of (a) with $\mathbf{Z} = \{A, B, C, D\}$. (c): a C-DAG of (a) with $\mathbf{W} = \{S, B\}$, $\mathbf{Z} = \{A, C\}$. (d):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

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, such that a C-DAG can be constructed.

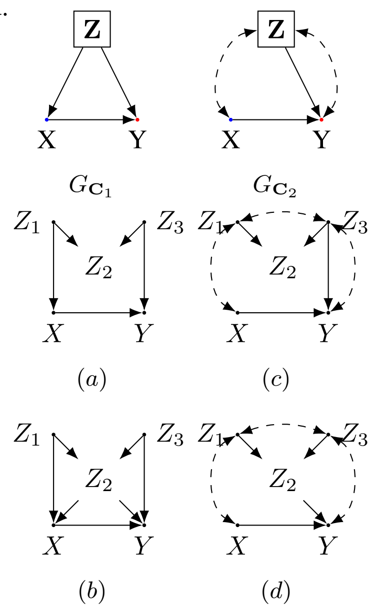


Figure 2: $G_{\mathbf{C}_1}$ is the C-DAG for diagrams (a) and (b) and $G_{\mathbf{C}_2}$ is the C-DAG for diagrams (c) and (d), where $\mathbf{Z} = \{Z_1, Z_2, Z_3\}$. $P(y|do(x))$ is identifiable in $G_{\mathbf{C}_1}$ by backdoor adjustment over $\mathbf{Z}$ and is not identifiable in $G_{\mathbf{C}_2}$.

Remark 2. Although a C-DAG is defined in terms of an underlying causal diagram G, in practice, one will construct a C-DAG when complete knowledge of the diagram G is unavailable. As an example of this construction, consider the diagram in Fig. 1(a) that illustrates a model of the effect of lisinopril (X) on the outcome of having a stroke (Y). 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 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 $\mathbf{Z} \leftarrow\!\!\dashrightarrow X$ and $\mathbf{Z} \leftarrow\!\!\dashrightarrow Y$. 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. Partial knowledge of this sort couldn't be formally articulated and used up to this point. 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 clarify that a C-DAG $G_{\mathbf{C}}$ as defined in Def. 1 is merely a graph over clusters of nodes $\mathbf{C}_1, ..., \mathbf{C}_k$, and does not have a priori the semantics and properties of a causal diagram over macro-variables $\mathbf{C}_i$. It's not clear, for example, whether the cluster nodes $\mathbf{C}_i$ satisfy the Markov properties with respect to the graph $G_{\mathbf{C}}$. Rather, a C-DAG can be seen as a graphical representation of an equivalence class of compatible causal diagrams, and represents a collection of causal diagrams 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 causal diagrams (a) and (b) can be represented by C-DAG $G_{\mathbf{C}_1}$ (top) and can therefore be thought of as being members of an equivalence class (EC, for short) represented by $G_{\mathbf{C}_1}$. The same can be concluded for the causal diagrams (c) and (d), both represented by C-DAG $G_{\mathbf{C}_2}$. The graphical representation of this ECs are shown in Fig. 3, where on the left we have the space of all possible DAGs, and on the right we have the space of C-DAGs.

Given the semantics of a C-DAG as representing an equivalence class of causal diagrams, what valid inference can one perform about the cluster variables given a C-DAG $G_{\mathbf{C}}$? In principle, we can only draw definite conclusions about properties shared by all members of the equivalence class. Going back to Fig. 3, we identify an effect in a C-DAG (e.g., the C-DAG $G_{\mathbf{C}_1}$ in Fig. 2) whenever this effect is identifiable in all members of the EC; e.g., diagrams (a), (b), and all other diagrams compatible with $G_{\mathbf{C}_1}$. Note that in this particular EC, all dots are marked with green, which means that the effect is identifiable in each of them. On the other hand, if there exists one causal diagram in the EC where the effect is not identifiable, this effect will not be identifiable in the corresponding C-DAG, e.g., due to diagram (d) in the figure, the effect is not identifiable in the C-DAG $G_{\mathbf{C}_2}$.

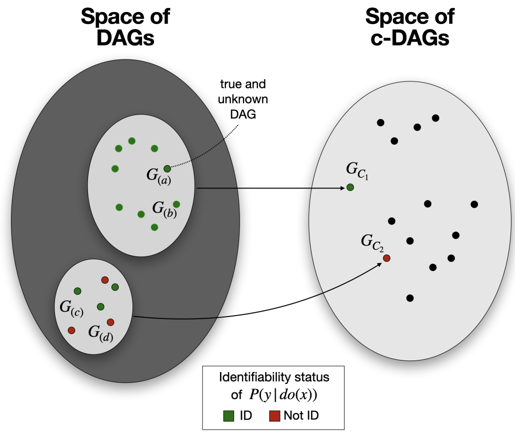


Figure 3: Identifying $P(y|do(x))$ in a C-DAG means identifying such an effect for the entire class of causal diagrams represented. In $G_{\mathbf{C}_1}$, the effect is identifiable (green) because it is identifiable in $G_{(a)}$, $G_{(b)}$, and all the other causal diagrams represented. In $G_{\mathbf{C}_2}$, 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., $G_{(d)}$).

Interestingly enough, it's certainly possible that the effect is identifiable in the true, underlying causal diagram, but identifiability will not be warranted if there exists another member of the EC such that this effect is not identifiable. In fact, this suggests a tradeoff of expressivity versus identifiability power. On the one hand, C-DAGs are more expressive than DAGs allowing just partial specification of knowledge to be articulated, while DAGs are more brittle, requiring full knowledge among all pairs of variables. On the other hand, the identification status of any query will be more likely to be positive in the case of DAGs than in C-DAGs, since they are much stricter. The modeling task requires one to find a balance between the amount of knowledge put in the model (C-DAGs) in a way such that identification may be achieved.

Once the semantics of C-DAGs is well-understood, now we turn our attention to computational issues. One naive approach to causal inference with cluster variables, e.g. identifying $Q = P(\mathbf{C}_i|do(\mathbf{C}_k))$, goes as follows – first enumerate all causal diagrams compatible with $G_{\mathbf{C}}$; then, evaluate the identifiability of Q in each causal diagram; finally, output $P(\mathbf{C}_i|do(\mathbf{C}_k))$ if all the causal diagrams give the same answer, otherwise output "non-identifiable". However, this naive approach will be time-consuming and impractical in high-dimensional settings - given a cluster $\mathbf{C}_i$ of size m, the number of possible DAG structures over the variables in $\mathbf{C}_i$ is super-exponential in m. Can we perform valid inferences about cluster variables using C-DAGs directly, without enumerating all the possible underlying causal diagrams? What properties of C-DAGs are shared by all the compatible causal diagrams? The next two sections present theoretical results to address these questions.

Finally, we note that not all properties of C-DAGs are shared across all the compatible causal 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 $X \leftarrow\!\!\!\dashrightarrow \mathbf{Z} \rightarrow Y$ in $G_{\mathbf{C}_2}$ 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 compatible causal diagrams in the equivalence class. On the other hand, we show in the next section a pleasant and surprising result (Theorem 1) that d-separations in a C-DAG do hold in all compatible causal diagrams. This powerful result is indeed critical to deriving causal inference rules and algorithms that are applicable to *all* the causal diagrams compatible with a given C-DAG (Theorems 2-5) regardless of the unknown relationships within each cluster.

4 D-Separation and do-Calculus in C-DAGs

D-separation [11] and the Pearl's celebrated *do-calculus* [12] are fundamental tools in causal inference from causal diagrams, which have been used extensively in the context of probabilistic and causal reasoning. In this section, we investigate extending these tools to the coarser representation of a C-DAG, when the underlying causal diagram is unknown.

As highlighted previously, a d-connecting path in a C-DAG not necessarily implies that the corresponding paths in a compatible causal diagram 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 causal diagrams.[2] 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 $G_{\mathbf{C}}$ 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. $\mathbf{C}_i *\!\!-\!\!* \mathbf{C}_m \rightarrow \mathbf{C}_j$ such that the non-collider cluster $\mathbf{C}_m$ is in $\mathbf{Z}$, or
2. $\mathbf{C}_i *\!\!\rightarrow \mathbf{C}_m \leftarrow\!\!* \mathbf{C}_j$ such that the collider cluster $\mathbf{C}_m$ 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 $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{G_{\mathbf{C}}}$, if and only if $\mathbf{Z}$ blocks every path from a cluster in $\mathbf{X}$ to a cluster in $\mathbf{Y}$.

Specifically, we show in Theorem 1 that the d-separation rules in C-DAGs are sound and complete in the following sense: whenever a d-separation holds in a C-DAG, it holds for all causal diagrams compatible with it; on the other hand, if a d-separation does not hold in a C-DAG, then there exists at least one causal diagram compatible with it for which the same d-separation statement does not hold.

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, that is,*

$$(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{G_{\mathbf{C}}} \implies (\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z})_{G}. \tag{1}$$

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.

Armed with the understanding coming from the d-separation rules in C-DAGs, next we extend the do-calculus rules to causal C-DAGs. An important lemma necessary to prove 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 all compatible causal diagrams. This result is shown in the next lemma:

Lemma 1. *If a C-DAG $G_{\mathbf{C}}$ is compatible with a causal diagram G, then, for $\mathbf{X}, \mathbf{Z} \subset \mathbf{C}$, the mutilated C-DAG $G_{\mathbf{C}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$ is compatible with the mutilated causal diagram $G_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}$.*

The soundness of do-calculus in C-DAGs as stated in Theorem 2 follows from Theorem 1, and Lemma 1.

Theorem 2. *(**Do-calculus in causal 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 $G_{\mathbf{C}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}$ is obtained from $G_{\mathbf{C}}$ 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 $G_{\mathbf{C}_{\overline{\mathbf{X}}}}$.

In Theorem 2, $\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{W}$ represent both the sets of clusters in $G_{\mathbf{C}}$ and the sets of variables contained in the corresponding sets of clusters.

We show next that the do-calculus rules in C-DAGs are also 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.*

5 Causal Identification in Causal C-DAGs

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 [2]. In this paper, we study the problem of identifying causal effects from observational data in C-DAGs.

[2] In Appendix **??**, we investigate in detail how path analysis is extended to C-DAGs.

ID-Algorithm

There exists a complete algorithm to determine whether $P(\mathbf{y}|do(\mathbf{x}))$ is identifiable from the combination of the causal diagram G and the observational distribution $P(\mathbf{V})$ [20, 18, 7]. This identification algorithm, or ID-algorithm for short, is based on the factorization of the interventional distributions according to the graphical structure, known as the truncated factorization product, i.e.:

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

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

We 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*

$$P(\mathbf{c} \setminus \mathbf{x}|do(\mathbf{x})) = \sum_{\mathbf{u}} P(\mathbf{u}) \prod_{k:\mathbf{C}_k \in \mathbf{C}\setminus\mathbf{X}} P(\mathbf{c}_k|pa_{\mathbf{C}_k}, \mathbf{u}'_k), \tag{3}$$

where $Pa_{\mathbf{C}_k}$ are the parents of the cluster $\mathbf{C}_k$ in $G_{\mathbf{C}}$ 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 ($\mathbf{C}_i \leftarrow\!\!\rightarrow \mathbf{C}_j$) between $\mathbf{C}_i$ and $\mathbf{C}_j$ in $G_{\mathbf{C}}$.

In Eq. (3), $\mathbf{X}, \mathbf{C}, \mathbf{C}_k, Pa_{\mathbf{C}_k}$ are the sets of variables contained in the corresponding sets of clusters. Theorem 4 essentially shows that C-DAGs can be treated as Causal Bayesian Networks (CBNs) [1, Def. 16] over macro-variables $\mathbf{C}$.

Since the ID-algorithm relies on the truncated factorization, Theorem 4 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 $G_{\mathbf{C}}$.

Theorem 5. *(**Soundness and Completeness of ID-algorithm**). The ID-algorithm is sound and complete 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 $G_{\mathbf{C}}$.*

The ID algorithm returns a formula for identifiable $P(\mathbf{y}|do(\mathbf{x}))$ that is valid in all causal diagrams compatible with the C-DAG $G_{\mathbf{C}}$. The completeness result ensures that if the ID-algorithm fails to identify $P(\mathbf{y}|do(\mathbf{x}))$ from $G_{\mathbf{C}}$, then there exists a causal diagram G compatible with $G_{\mathbf{C}}$ where the effect $P(\mathbf{y}|do(\mathbf{x}))$ is not identifiable.

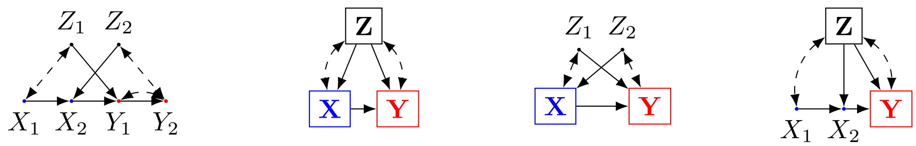
Appendix **??** contains an experimental study evaluating the ability of C-DAGs to accurately assess the identifiability of a causal effect while requiring less domain knowledge for their construction.

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 [13, 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 [13, 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.

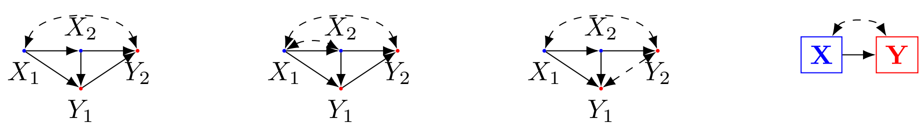
Identification in Fig. 4. In causal diagram (a), the effect of $\{X_1, X_2\}$ on $\{Y_1, Y_2\}$ is identifiable by backdoor adjustment over $\{Z_1, Z_2\}$ as follows: $P(y_1, y_2|do(x_1, x_2)) = \sum_{z_1,z_2} P(y_1, y_2|x_1, x_2, z_1, z_2)P(z_1, z_2)$. Note, however, that the backdoor path cannot be blocked



$(a)\ G$ $(b)\ G_{\mathbf{C}_1}$ $(c)\ G_{\mathbf{C}_2}$ $(d)\ G_{\mathbf{C}_3}$

Figure 4: (a) : causal diagram G where the effect $P(y_1, y_2|do(x_1, x_2))$ is identifiable. (b) : C-DAG $G_{\mathbf{C}_1}$ with clustering $\mathbf{X} = \{X_1, X_2\}$, $\mathbf{Y} = \{Y_1, Y_2\}$, and $\mathbf{Z} = \{Z_1, Z_2\}$. (c) : C-DAG $G_{\mathbf{C}_2}$ with clustering $\mathbf{X} = \{X_1, X_2\}$ and $\mathbf{Y} = \{Y_1, Y_2\}$. (d) : C-DAG $G_{\mathbf{C}_3}$ with clustering $\mathbf{Y} = \{Y_1, Y_2\}$ and $\mathbf{Z} = \{Z_1, Z_2\}$. The effect $P(\mathbf{y}|do(\mathbf{x}))$ is not identifiable in $G_{\mathbf{C}_1}$, but is identifiable in $G_{\mathbf{C}_2}$ and $P(\mathbf{y}|do(x_1, x_2))$ is identifiable in $G_{\mathbf{C}_3}$.

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 $P(\mathbf{y}|do(\mathbf{x}))$ is not identifiable. If the covariates Z_1 and Z_2 are not clustered together as shown in the C-DAG $G_{\mathbf{C}_2}$ (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 $P(\mathbf{y}|do(\mathbf{x}))$ is identifiable by backdoor adjustment over $\{Z_1, Z_2\}$ as follows: $P(\mathbf{y}|do(\mathbf{x})) = \sum_{z_1, z_2} P(\mathbf{y}|\mathbf{x}, z_1, z_2)P(z_1, z_2)$. If the treatments X_1 and X_2 are not clustered together as shown in the C-DAG $G_{\mathbf{C}_3}$ (d), then the joint effect of X_1 and X_2 on the cluster $\mathbf{Y}$ is identifiable and given by the following expression: $P(\mathbf{y}|do(x_1, x_2)) = \sum_{\mathbf{z}, x'_1} P(\mathbf{y}|x'_1, x_2, \mathbf{z})P(x'_1, \mathbf{z})$.



$(a)\ G_1$ $(b)\ G_2$ $(c)\ G_3$ $(d)\ G_{\mathbf{C}}$

Figure 5: $(a), (b)$, and (c) are causal diagrams compatible with the C-DAG $G_{\mathbf{C}}$ 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 $P(\mathbf{y}|do(\mathbf{x}))$ is not identifiable from the C-DAG $G_{\mathbf{C}}$.

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 $P(\mathbf{y}|do(\mathbf{x}))$ is also not identifiable.

6 Conclusions

Causal diagrams are a popular language for specifying the necessary assumptions for causal inference. Yet, despite their power, 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, including in the health and social sciences. This paper introduces a new class of graphical models which 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 as delineated by Def. 1, these variables can be clustered together. A causal diagram itself is an extreme case of a C-DAG where each cluster contains exactly one variable. We prove fundamental results to allow causal inferences within this equivalence class, which translate to statements about the set of causal diagrams compatible with the encoded constraints. Specifically, we show that d-separation is sound and complete even when clusters are considered (Theorem 1) and that Pearl's do-calculus is both sufficient and necessary for derivations in C-DAGs (Theorems 2 and 3). We then generalize the truncated factorization product to when knowledge is available only in C-DAG-form (Theorem 5). Building on these results, we prove that causal identification algorithms can take C-DAGs as input and are both sound and complete. These results are critical in establishing the foundations for

C-DAGs and enabling their use in ways comparable to causal diagrams. We hope the new machinery for C-DAGs 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.

Broader Societal Impact and Limitations

C-DAGs will have a positive societal impact by enabling researchers to perform causal inferences in practice, and to do so accurately through using graphs as representations of knowledge and assumptions, with observational data and the do-calculus. This has the potential to enable causal inference and discovery in fields where using the more restrictive DAGs is too challenging, leading to advancements in causal knowledge and understanding. No negative societal impacts of this work are foreseen since it is a refinement and generalization of known theory.

Challenges or limitations of applying the machinery of C-DAGs include that not all admissible clusterings are guaranteed to yield an identifiable effect for a corresponding causal diagram where the effect is identifiable. Obviously, if knowledge of the granularity of the causal diagram is available (which is a C-DAG where all clusters have size one), this would lead to the purported identification result. In addition, C-DAGs are assumed to be constructed in this work based on knowledge, albeit requiring less knowledge than DAGs. The development of tools to assist the construction of a C-DAG from observational data constitute a very promising research direction. An approach that extends current structure learning algorithms such as the FCI [23] to learn C-DAGs can be less prone to errors, computationally less expensive, and statistically more robust since fewer conditional independence tests will be required. However, because conditional independencies over clusters of variables do not necessarily correspond to d-separations in C-DAGs, this is a challenging open research problem.

References

[1] E. Bareinboim, J. D. Correa, D. Ibeling, and T. Icard. On Pearl's Hierarchy and the Foundations of Causal Inference. Technical Report R-60, Causal Artificial Intelligence Lab, Columbia University, 7 2020.

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

[3] S. Beckers and J. Y. Halpern. Abstracting causal models. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, AAAI'19, 2019.

[4] K. Chalupka, F. Eberhardt, and P. Perona. Multi-level cause-effect systems. In A. Gretton and C. C. Robert, editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 361–369, Cadiz, Spain, 09–11 May 2016. PMLR.

[5] K. Chalupka, P. Perona, and F. Eberhardt. Visual causal feature learning. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, UAI'15, page 181–190, Arlington, Virginia, USA, 2015. AUAI Press.

[6] C. Glymour, K. Zhang, and P. Spirtes. Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10, 2019.

[7] 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. AUAI Press, 2006.

[8] S. Kleinberg and G. Hripcsak. A review of causal inference for biomedical informatics. *Journal of Biomedical Informatics*, 44(6):1102–1112, 2011.

[9] S. L. Lauritzen and T. S. Richardson. Chain graph models and their causal interpretations. *Royal Statistical Society*, 64(Part 2):1–28, 2002.

[10] S. Lee, J. D. Correa, and E. Bareinboim. General Identifiability with Arbitrary Surrogate Experiments. In *Proceedings of the Thirty-Fifth Conference Annual Conference on Uncertainty in Artificial Intelligence*. AUAI Press, 2019.

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

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

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

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

[15] 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.

[16] P. K. Rubenstein, S. Weichwald, S. Bongers, J. M. Mooij, D. Janzing, M. Grosse-Wentrup, and B. Schölkopf. Causal consistency of structural equation models. In G. Elidan, K. Kersting, and A. T. Ihler, editors, *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence, UAI 2017, Sydney, Australia, August 11-15, 2017*. AUAI Press, 2017.

[17] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021.

[18] 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*, pages 1219–1226, 2006.

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

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

[21] 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/MIT Press.

[22] J. Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17):1873–1896, 11 2008.

[23] J. Zhang. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16):1873–1896, 2008.