Minimum Intervention Cover of a Causal Graph

Saravanan Kandasamy
Tata Institute of Fundamental Research
Mumbai, India

Arnab Bhattacharyya
Dept. of Computer Science
National Univ. of Singapore
Singapore
&
Dept. of Computer Science & Automation
Indian Institute of Science
Bangalore, India

Vasant G Honavar
Artificial Intelligence Research Laboratory
College of Information Sciences
& Technology, Pennsylvania State Univ.
University Park, PA, USA

Abstract

Eliciting causal effects from interventions and observations is one of the central concerns of science, and increasingly, artificial intelligence. We provide an algorithm that, given a causal graph G, determines $\mathsf{MIC}(G)$, a *minimum intervention cover* of G, i.e., a minimum set of interventions that suffices for identifying *every* causal effect that is identifiable in a causal model characterized by G. We establish the completeness of do-calculus for computing $\mathsf{MIC}(G)$. $\mathsf{MIC}(G)$ effectively offers an efficient compilation of all of the information obtainable from all possible interventions in a causal model characterized by G. Minimum intervention cover finds applications in a variety of contexts including counterfactual inference, and generalizing causal effects across experimental settings. We analyze the computational complexity of minimum intervention cover and identify some special cases of practical interest in which $\mathsf{MIC}(G)$ can be computed in time that is polynomial in the size of G.

Introduction

Determining causal effects from interventions and observations is one of the central concerns of science, and increasingly, of artificial intelligence (Pearl 2009; Spirtes, Glymour, and Scheines 2000; Imbens and Rubin 2015; Morgan and Winship 2014; Hernan and Robins 2010; Berzuini, Dawid, and Bernardinell 2012; Peters, Janzing, and Schölkopf 2017). In the framework pioneered by Pearl (Pearl 2009), the *structure* of a *causal model* is encoded using a *causal graph* G, defined over a set of observable variables (vertices) $\mathbf{V}$. In the resulting causal model, for any two variables X and Y, a directed edge $X \rightarrow Y$ denotes that X is a direct cause of Y; and a bi-directed edge $X \leftrightarrow Y$ indicates that X and Y are confounded by an *unobservable* variable (which is a common parent). The parameters of the causal model correspond to the probability distributions of the variables conditioned on their parents in G. In a causal model encoded using a graph G, for a given subset of observable variables $\mathbf{X}$, and an assignment $\mathbf{x}$ of $\mathbf{X}$, the *intervention* $do(\mathbf{x})$ refers to the action of fixing $\mathbf{X}$ to $\mathbf{x}$, irrespective of the values of the parents of $\mathbf{X}$. For any $\mathbf{Y} \subseteq \mathbf{V}$, the causal effect of $do(\mathbf{x})$ on $\mathbf{Y}$ will be the interventional distribution obtained on $\mathbf{Y}$ by the intervention $do(\mathbf{x})$ (Pearl 2009). Given a causal model, *Do-calculus* (Pearl 1995; 2009) offers a general machinery that can be used to identify causal effects from observations and interventions, answer counterfactual queries, etc., given a causal graph.

Related Work

The problem of identifying causal effects from data has been extensively studied in the literature. In the absence of unobservable variables, all causal effects are identifiable from the observational distribution (Robins 1986; Spirtes, Glymour, and Scheines 2000; Pearl 1995). When some of the variables are unobservable, it is not always possible to identify causal effects from the observational distribution alone. A series of papers established sufficient graphical conditions for solving this problem (Spirtes, Glymour, and Scheines 2000; Pearl 1995; Galles and Pearl 1995; Pearl and Robins 1995; Halpern 1998; Kuroki and Miyakawa 1999; Tian and Pearl 2002a), eventually leading to a sound and complete algorithm (Shpitser and Pearl 2006; Huang and Valtorta 2006). The resulting methods have been generalized to work in settings where the underlying causal graph is unknown (Hyttinen, Eberhardt, and Järvisalo 2015).

Recent work has considered the problem of generalizing a causal effect from one or more source domains (where observational and all interventional distributions are available) to a target domain (where only the observational distribution is available), provided some invariances in causal mechanisms hold across the source and target domains (Pearl and Bareinboim 2011; Lee and Honavar 2013b; Bareinboim and Pearl 2013b). Extensions of this problem consider the identification of causal effects in the target domain, but from the observational and interventional distributions on *subsets* of observable variables (that are amenable to intervention) of the source domains (Bareinboim and Pearl 2013a; Lee and Honavar 2013a; Bareinboim et al. 2013; Bareinboim and Pearl 2014). These results provide a sound theoretical foundation for integrative analyses of observational and experimental data (Tsamardinos, Triantafillou, and Lagani 2012; Bareinboim and Pearl 2016).

A related line of work (Shpitser and Pearl 2008) provides a sound and complete graphical characterization for the problem of answering counterfactual queries in the setting where the observational distribution and all the interventional distributions are available.

Motivation

The focus of the entire body of existing work (summarized above) on generalizing interventional data across multiple domains and on identifying counterfactual queries from interventional data is on whether the required quantity of interest can be determined when *all* of the interventional distributions on the observable variables (or a subset of observables that are amenable to experimental manipulation) are obtainable. However, obtaining an interventional distribution requires performing the corresponding intervention. Because interventions can incur significant cost and effort, it is important to minimize the number of interventions that need to be performed. Minimizing the number of interventions is especially useful in cases where a significant amount of interventional distributions are actually required by the existing algorithms for the identification of the quantities of interest.

Contributions

We address the following question: Given a causal graph G, find a smallest set of interventions that suffices to determine all interventional distributions. We call such a minimum set of interventions of G a *minimum intervention cover* of G ($\mathsf{MIC}(G)$). We treat the case where any observable variable may be manipulated by interventions. A similar analysis when the set of manipulable variables is restricted remains a challenging open problem.

The main contributions of this paper include:

1. A necessary and sufficient condition for a set of interventions (or, equivalently, interventional distributions) that form $\mathsf{MIC}(G)$, a minimum intervention cover of a causal graph G.
2. A sound and complete algorithm for finding the minimum intervention cover of a causal graph G.
3. Proof of completeness of *do*-calculus for the minimum intervention cover problem.
4. An analysis of the computational complexity of $\mathsf{MIC}(G)$, including a characterization of special cases of practical interest in which $\mathsf{MIC}(G)$ can be determined in time that is polynomial in the size of G. In particular, we provide an efficient algorithm to determine $\mathsf{MIC}(G)$ when G has bounded in-degree, out-degree, and C-components of bounded size[1].

Preliminaries

Probabilistic Causal Models and Causal Graphs

We follow the notational conventions of probabilistic causal models (PCM) (Pearl 2009), also known as structural causal models or data-generating models. As is customary, without loss of generality (see (Verma and Pearl 1990; Tian and Pearl 2002b)), we limit our attention to causal graphs with the unobservable variables as root nodes, each with exactly two observable children. The resulting causal graph is directed acyclic with respect to the observable variables, but contains bi-directed edges between observable variables to represent the common unobservable parent between those observable variables. A probabilistic causal model consists of a causal graph G, and a four-tuple $\langle \mathbf{V}, \mathbf{U}, F, \Pr(\mathbf{U})\rangle$, where $\mathbf{V}$ is the set of all observable variables, $\mathbf{U}$ is the set of all unobservable variables (connected by the bi-directed edges of G) distributed according to $\Pr[\mathbf{U}]$, and $F = \{f_1, \ldots, f_{|\mathbf{V}|}\}$ is a set of functions. The value of each variable $V \in \mathbf{V}$ is determined by the function f_V based on the values of the parents (both observable and unobservable) of V.

[1]See Definition 6.

Notation

We use uppercase letters to denote variables; lowercase to denote value assignments of variables; bold uppercase and bold lowercase letters to denote sets of variables and their value assignments respectively[2]; $G_{\overline{\mathbf{X}}}$ and $G_{\underline{\mathbf{X}}}$ to denote the graph obtained from G by removing the incoming edges (including bi-directed edges) to $\mathbf{X}$ and outgoing edges from $\mathbf{X}$ respectively; $\mathbf{V} = \{V_1, V_2, \ldots, V_n\}$ to denote the observable vertices of graph G, with the vertex indices being topologically ordered; $\mathsf{Pa}(\mathbf{X})$ and $\mathsf{An}(\mathbf{X})$ to denote the observable parents and observable ancestors of $\mathbf{X}$ (excluding $\mathbf{X}$) in G respectively; $\mathsf{pa}(\mathbf{X})$ to denote an assignment to $\mathsf{Pa}(\mathbf{X})$; $G[\mathbf{D}]$ to denote the induced subgraph of G on $\mathbf{D}$: whose vertex set is $\mathbf{D}$ and whose edge set contains all the edges (including bi-directed edges) of G that have both endpoints in $\mathbf{D}$, for any $\mathbf{D} \subseteq \mathbf{V}$; $\Pr^M[\cdot]$ to denote $\Pr[\cdot]$ in model M. Two assignments $\mathbf{x}, \mathbf{y}$ of $\mathbf{X}, \mathbf{Y}$ are said to be consistent if they agree on all of the vertices in $\mathbf{X} \cap \mathbf{Y}$. We use set operations to denote values of a set of variables. For example, $\mathbf{a} \setminus \mathsf{pa}(\mathbf{A})$ is used to represent the values of $\mathbf{A} \setminus \mathsf{Pa}(\mathbf{A})$. When the graph being referenced is not clear from context, we use $\mathsf{Pa}(\mathbf{X})_G$ to denote the observable parents of $\mathbf{X}$ (excluding $\mathbf{X}$) in graph G. For any non-negative integer k, we use $[k]$ to denote the set $\{1, 2, \ldots, k\}$; For a given a collection of binary values $\mathbf{s}$, we use $bp(\mathbf{s})$ to denote the bit parity of $\mathbf{s}$ and $\overline{bp}(\mathbf{s})$ to denote the complement of the bit parity of $\mathbf{s}$; We also use $\mathbf{0}$ and $\mathbf{1}$ to represent a set of 0 and 1 values respectively.

Interventions and Identifiability

We review some essential definitions:

Intervention. Given a causal graph G, a set of observable variables $\mathbf{X} \subseteq \mathbf{V}$, and an assignment $\mathbf{x}$ of $\mathbf{X}$, an *intervention* $do(\mathbf{x})$ is the process of fixing $\mathbf{X}$ to $\mathbf{x}$ irrespective of the values of parents of $\mathbf{X}$ (Pearl 2009), which induces a new graph $G_{\overline{\mathbf{X}}}$ obtained from G by removing all incoming edges of $\mathbf{X}$.

Interventional distribution. Given two disjoint subsets $\mathbf{X}, \mathbf{Y}$ of $\mathbf{V}$, and an intervention $do(\mathbf{x})$, the *interventional distribution* denoted by $\Pr[\mathbf{Y} \mid do(\mathbf{x})]$, is the *causal effect* of the intervention $do(\mathbf{x})$ on $\mathbf{Y}$, that is the distribution over $\mathbf{Y}$ obtained over the intervention $do(\mathbf{x})$.

Information set. An *information set* $\mathsf{IS}(G)$ denotes a set of interventional distributions over the causal graph G.

[2]Without loss of generality, the variables are assumed to take values from the domain Σ. The results presented in the paper generalize in a straightforward way to the setting where each variable X takes values from the corresponding domain Σ_X

Joint interventional distribution For a given intervention $do(\mathbf{x})$, the distribution observed over the rest of the variables $\Pr[\mathbf{V} \setminus \mathbf{X} \mid do(\mathbf{x})]$ is a *joint interventional distribution*.

Joint information set. A *joint information set* is an information set that contains only joint interventional distributions.

Note that interventions with different assignments represent distinct elements in the information set. For example, $do(\text{Smoking} = 0)$ and $do(\text{Smoking} = 1)$ are different.

Definition 1 (Identifiability)**.** *For a given causal graph G, let $\mathbf{X}, \mathbf{Y}$ be disjoint subsets of the observable vertices $\mathbf{V}$ and let $\mathsf{IS}(G)$ be a given information set for G. The causal effect of an action $do(\mathbf{x})$ on $\mathbf{Y}$, denoted by $\Pr[\mathbf{Y} \mid do(\mathbf{x})]$, is* identifiable from $\mathsf{IS}(G)$ *in G if $\Pr[\mathbf{Y} \mid do(\mathbf{x})]$ is uniquely determined by $\mathsf{IS}(G)$ in any causal model that is defined on G. Similarly, an information set $\mathsf{IS}'(G)$ is* identifiable from $\mathsf{IS}(G)$ *in G, if for each intervention $\mathrm{I} \in \mathsf{IS}'(G)$, I is identifiable from $\mathsf{IS}(G)$ in G.*

The following lemma directly follows from the definition of identifiability.

Lemma 1. *Given a causal graph G and an information set $\mathsf{IS}(G)$, the interventional distribution $\Pr[\mathbf{Y} \mid do(\mathbf{x})]$ is not identifiable from $\mathsf{IS}(G)$, if there exist two models M_1, M_2 that is defined on the same causal graph G such that (i) $\Pr^{M_1}[\mathbf{Y} \mid do(\mathbf{x})] \neq \Pr^{M_2}[\mathbf{Y} \mid do(\mathbf{x})]$; and (ii) $\Pr^{M_1}[\mathbf{T} \mid do(\mathbf{s})] = \Pr^{M_2}[\mathbf{T} \mid do(\mathbf{s})]$ for each intervention $\Pr[\mathbf{T} \mid do(\mathbf{s})] \in \mathsf{IS}(G)$.*

Definition 2 (All possible interventional distributions ($\mathsf{IS}^*(G)$))**.** *For a given causal graph G, let*

$$\mathsf{IS}^*(G) := \bigcup_{\mathbf{S} \subseteq \mathbf{V}} \bigcup_{\mathbf{s}} \{\Pr[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})]\}$$

represent the set of all possible interventional distributions*.*

Definition 3 (Size of Information Set)**.** *For a given information set $\mathsf{IS}(G)$, the* size of $\mathsf{IS}(G)$ *is the cardinality of $\mathsf{IS}(G)$, i.e., the number of interventions in the set $\mathsf{IS}(G)$.*

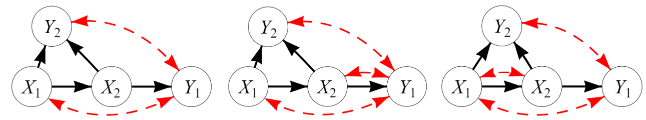
Definition 4 (Intervention Cover)**.** *For a given causal graph G, an information set $\mathsf{IS}(G)$ is an* intervention cover *of G if $\mathsf{IS}^*(G)$ is identifiable from $\mathsf{IS}(G)$ in G.*

Definition 5 (Minimum Intervention Cover ($\mathsf{MIC}(G)$))**.** *For a given causal graph G, a* minimum intervention cover of *G ($\mathsf{MIC}(G)$) is an information set $\mathsf{IS}^{\mathsf{min}}(G)$ such that (i) $\mathsf{IS}^{\mathsf{min}}(G)$ is an intervention cover of G; (ii) there exists no intervention cover of G of size smaller than $\mathsf{IS}^{\mathsf{min}}(G)$.*

Remark 1. *Note that because we are minimizing the number of interventions in the information set, not the number of variables involved in each intervention, we may take $\mathsf{MIC}(G)$ to be a joint information set without loss of generality.*

Remark 2. *Determining causal graphs from a small number of interventions has been studied in the literature under the faithfulness assumption (using Conditional Independence (CI) tests) (Shanmugam et al. 2015; Kocaoglu, Shanmugam, and Bareinboim 2017). $\mathsf{MIC}(G)$ can be applied to the causal graphs constructed using the above algorithms.*

We assume that the distributions are positive as in (Pearl 2009). We adopt the definitions and rules of do-calculus (Pearl 2009) (See Supplementary material). In what follows, we will use the graph H (shown in Figure 1a) to construct examples that illustrate the key definitions, arguments and results.



(a) Causal graph H (b) Causal Graph H' (c) Causal Graph H''

Figure 1: X_1 – Smoking, X_2 – Alcohol Consumption, Y_1 – Depression, Y_2–Sleep Disorder

C-Component Factorization

Definition 6 ((Tian and Pearl 2002a) *C-component*)**.** *Given a causal graph G, and a set of observable vertices $\mathbf{S} \subseteq \mathbf{V}$, $\mathbf{S}$ is a* C-component *of G if in the induced subgraph $G[\mathbf{S}]$ there is a path between any two vertices of $\mathbf{S}$ that consists of only bi-directed edges.*

For a given causal graph G, we use $C(\mathbf{V}) := \{\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_{k-1}, \mathbf{S}_k\}$ to represent the partition of $\mathbf{V}$ into the maximal C-components of G, *i.e.,* each $\mathbf{S}_i$ is a maximal C-component of G. Similarly, for any subset $\mathbf{W} \subseteq \mathbf{V}$, we use $C(\mathbf{W})$ to denote the set of all maximal C-components of the subgraph $G[\mathbf{W}]$ induced on $\mathbf{W}$.

Example 1. The C-components of the graph H are: $\{X_1\}, \{X_2\}, \{Y_1\}, \{Y_2\}, \{X_1, Y_1, Y_2\}, \{X_1, Y_1\}, \{Y_1, Y_2\}$ Hence, the maximal C-components of H are: $C(\{X_1, X_2, Y_1, Y_2\}) = \{\{X_1, Y_1, Y_2\}, \{X_2\}\}$.

As we will see below, the properties of C-components play a crucial role in the identification of causal effects. Let us first recall a fact that directly follows from the definition of probabilistic causal models:

Lemma 2. *Given a subset of observable variables $\mathbf{S} \subseteq \mathbf{V}$, an assignment $\mathbf{s}$ of $\mathbf{S}$, and an assignment $\mathbf{pa}(\mathbf{S})$ of the observable parents of $\mathbf{S}$, the interventional probability $\Pr[\mathbf{s} \mid do(\mathbf{pa}(\mathbf{S}))]$ can be expressed as*

$$\Pr[\mathbf{s} \mid do(\mathbf{pa}(\mathbf{S}))] = \Pr[\mathbf{s} \mid do(\mathbf{pa}(\mathbf{S}), \mathbf{o})]$$

for any assignment $\mathbf{o}$ of any $\mathbf{O} \subseteq \mathbf{V} \setminus (\mathbf{Pa}(\mathbf{S}) \cup \mathbf{S})$.

Proof. By the definition of probabilistic causal models, when all the observable parents of $\mathbf{S}$ are targeted by an intervention, the distribution on $\mathbf{S}$ remains unchanged regardless of whether the other vertices (i.e., $\mathbf{O}$) are also targeted by the intervention. □

Example 2. Consider the graph H. Let $\mathbf{S} = \{Y_1\}$ and $\mathbf{O} = \{X_1, Y_2\}$. Hence $\mathbf{Pa}(\mathbf{S}) = \{X_2\}$. By the definition of probabilistic causal model, when X_2 is set by an intervention (say to the value x_2), the probability distribution of Y_1 is unaffected by whether or not X_1 or Y_2 are are also set by intervention. Hence $\Pr[\mathbf{s} \mid do(\mathbf{pa}(\mathbf{S}))] = \Pr[\mathbf{s} \mid do(\mathbf{pa}(\mathbf{S}), \mathbf{o})]$.

The following C-component factorization Lemma (Tian and Pearl 2002b) highlights the role played by C-components in the identification of causal effects:

Lemma 3. *(Tian and Pearl 2002b) For a given subset* $\mathbf{X} \subseteq \mathbf{V}$*, let* $C(\mathbf{V} \setminus \mathbf{X}) = \{\mathbf{B}_1, \ldots, \mathbf{B}_k\}$*. Then, for any assignment* $\mathbf{v}$ *of the observable vertices* $\mathbf{V}$*, the interventional probability* $\Pr[\mathbf{v} \setminus \mathbf{x} \mid do(\mathbf{x})]$ *can be expressed as* $\Pr[\mathbf{v} \setminus \mathbf{x} \mid do(\mathbf{x})] = \prod_i \Pr[\mathbf{b}_i \mid do(\mathbf{v} \setminus \mathbf{b}_i)]$*. Hence:*

$$\begin{aligned}\Pr[\mathbf{v} \setminus \mathbf{x} \mid do(\mathbf{x})] &= \prod_i \Pr[\mathbf{b}_i \mid do(\mathbf{v} \setminus \mathbf{b}_i)] \\ &= \prod_i \Pr[\mathbf{b}_i \mid do(\mathbf{pa}(\mathbf{B}_i))] \, \textit{(by Lemma 2)}\end{aligned}$$

where the assignment $\mathbf{pa}(\mathbf{B}_i)$ *is consistent with* $\mathbf{v} \setminus \mathbf{b}_i$*.*

Example 3. Let x_1, x_2, y_1, y_2 be an arbitrary assignment of the variables X_1, X_2, Y_1, Y_2 respectively , and $do(x_1)$ an intervention in a causal model characterized by the graph H. It is easy to see that $\{X_2\}$ and $\{Y_1, Y_2\}$ are the the maximal C-components of the causal graph $H_{\overline{X_1}}$, resulting from the intervention, $do(x_1)$. Lemma 3 says that the interventional probability $\Pr[x_2, y_1, y_2 \mid do(x_1)]$ can be expressed as the product of $\Pr[x_2 \mid do(x_1)]$ and $\Pr[y_1, y_2 \mid do(x_1, x_2)]$.

Minimum Intervention Cover

Overview: We begin by defining an information set $\mathsf{LD}(G)$, a set of *local distributions*, such that $\mathsf{IS}^*(G)$ (the set of all possible interventional distributions of a causal model characterized by a causal graph G) is identifiable from $\mathsf{LD}(G)$. We then define $\mathsf{ILD}(G)$, an *informative* subset of $\mathsf{LD}(G)$, and show that $\mathsf{LD}(G)$, and hence $\mathsf{IS}^*(G)$, is identifiable from $\mathsf{ILD}(G)$. We proceed to introduce a sound and complete graphical criteria for identifying $\mathsf{ILD}(G)$ from a joint information set $\mathsf{IS}^{\mathsf{inp}}(G)$. We show that $\mathsf{IS}^{\mathsf{inp}}(G)$ is a minimum intervention cover of a causal model characterized by G if and only if $\mathsf{IS}^{\mathsf{inp}}(G)$ identifies $\mathsf{ILD}(G)$ and no other information set of a smaller size does so.

Definition 7. *For a given causal graph G, we define*

$$\mathsf{LD}(G) := \bigcup_{\mathbf{B}_i : \mathbf{B}_i \textit{ is a C-component of } G} \; \bigcup_{\mathbf{pa}(\mathbf{B}_i)} \Pr[\mathbf{B}_i | do(\mathbf{pa}(\mathbf{B}_i))].$$

Example 4. Consider the graph H shown in Figure 1a. For every C-component $\mathbf{B}_i$ of H (See Example 1), and for every assignment of values to their parents $\mathbf{pa}(\mathbf{B}_i)$, it is easy to see that the corresponding interventional distribution $\Pr[\mathbf{B}_\mathrm{j} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{j}))]$ is in $\mathsf{LD}(H)$.

The next claim, which directly follows from the C-component factorization of Lemma 3, shows that every intervention of G is identifiable from $\mathsf{LD}(G)$.

Claim 1. $\mathsf{IS}^*(G)$ *is identifiable from* $\mathsf{LD}(G)$*.*

Proof. Recall that each interventional distribution of $\mathsf{IS}^*(G)$ can be factorized in terms of its corresponding C-factors $\Pr[\mathbf{b}_\mathrm{i} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{i}))]$ (Lemma 3). All such C-factors are available in $\mathsf{LD}(G)$. □

Next we show that an "informative" subset of $\mathsf{LD}(G)$, which we call $\mathsf{ILD}(G)$, suffices to identify each distribution in $\mathsf{LD}(G)$. $\mathsf{ILD}(G)$ is the set of all $\Pr[\mathbf{B}_\mathrm{j} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{j}))] \in \mathsf{LD}(G)$ such that $\mathbf{B}_j$ is a maximal C-component of the induced subgraph $G[\mathbf{V} \setminus \mathbf{Pa}(\mathbf{B}_j)]$.

Definition 8. *For a given causal graph G, we define*

$$\mathsf{ILD}(G) := \bigcup_{\mathbf{B}_j : \mathbf{B}_j \in C(\mathbf{V} \setminus \mathbf{Pa}(\mathbf{B}_j))} \; \bigcup_{\mathbf{pa}(\mathbf{B}_j)} \Pr[\mathbf{B}_\mathrm{j} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{j}))] \, .$$

Example 5. Consider the graph H shown in shown in Figure 1a. It is easy to verify that $\mathsf{ILD}(H)$ contains only the interventional distributions $\Pr[\mathbf{B}_\mathrm{j} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{j}))]$ such that $\mathbf{B}_j \in \{\{X_2\}, \{Y_1, Y_2\}, \{X_1, Y_1, Y_2\}\}$.

Claim 2. $\mathsf{LD}(G)$ *is identifiable from* $\mathsf{ILD}(G)$*.*

Proof. Let $\mathsf{LD}(G) \ni \Pr[\mathbf{B}_\mathrm{i} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{i}))] \notin \mathsf{ILD}(G)$. We prove the claim by demonstrating the existence of an informative local distribution that identifies $\Pr[\mathbf{B}_\mathrm{i} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{i}))]$. Define

$$\mathbf{B}_j := \mathbf{B}_i \cup \mathbf{D}$$

where $\mathbf{D} \subseteq \mathbf{V} \setminus (\mathbf{B}_i \cup \mathbf{Pa}(\mathbf{B}_i))$ such that $\mathbf{B}_i \cup \mathbf{D} \in C(\mathbf{V} \setminus \mathbf{Pa}(\mathbf{B}_i))$ (See Figure 2).

Let $\mathbf{pa}(\mathbf{B}_j)$ be an assignment consistent with $\mathbf{pa}(\mathbf{B}_i)$.

By definition, $\Pr[\mathbf{B}_\mathrm{j} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}_\mathrm{j}))] \in \mathsf{ILD}(G)$, because $\mathbf{B}_j \in C(\mathbf{V} \setminus \mathbf{Pa}(\mathbf{B}_j))$. Hence:

$$\begin{aligned}\Pr[\mathbf{B}_i | do(\mathbf{pa}(\mathbf{B}_i))] &= \Pr[\mathbf{B}_i | do(\mathbf{pa}(\mathbf{B}_j))] \\ &= \sum_{\mathbf{b}_j \setminus \mathbf{b}_i} \Pr[\mathbf{B}_i, (\mathbf{b}_j \setminus \mathbf{b}_i) | do(\mathbf{pa}(\mathbf{B}_j))].\end{aligned}$$

The first equality follows from Lemma 2, since (i) $\mathbf{Pa}(\mathbf{B}_i) \subseteq \mathbf{Pa}(\mathbf{B}_j)$; (ii) $\mathbf{Pa}(\mathbf{B}_j)$ and $\mathbf{B}_i$ are disjoint; (iii) the assignments $\mathbf{pa}(\mathbf{B}_i)$ and $\mathbf{pa}(\mathbf{B}_j)$ are consistent. The second equality is obtained by marginalization. Hence the claim. □

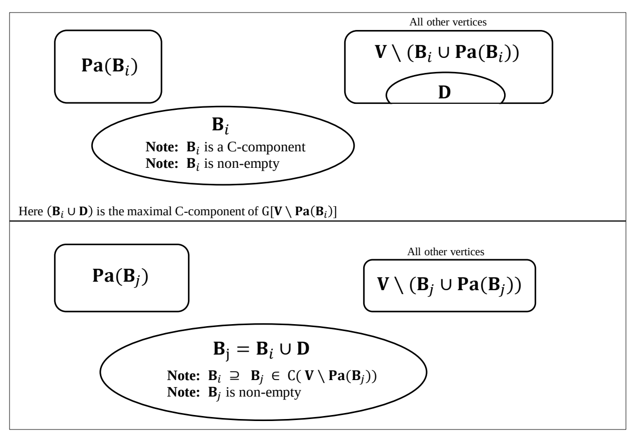


Figure 2: Illustration of Claim 2

Claims 1 and 2 imply that $\mathsf{ILD}(G)$ suffices to identify all interventional distributions $\mathsf{IS}^*(G)$. We now we proceed to specify a graphical criterion for the identifiability of $\mathsf{ILD}(G)$

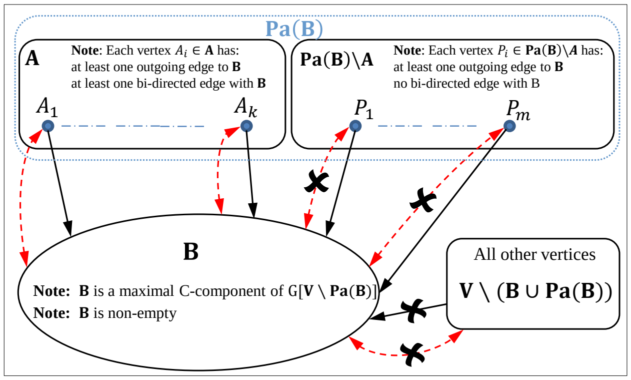


Figure 3: Bush $\mathbf{A}, \mathbf{B}$

from a given joint information set $\mathsf{IS}^{\mathsf{inp}}(G)$. Our analysis relies on the existence of a special structure, which we call a *bush*.

Definition 9 (Bush)**.** *For a given causal graph G, let $\mathbf{A}$ and $\mathbf{B}$ be disjoint subsets of the observable variables $\mathbf{V}$. Then $\mathbf{A}, \mathbf{B}$ form a* bush *in G, if*
(i) $|\mathbf{B}| \neq 0$
(ii) $\mathbf{B} \in C(\mathbf{V} \setminus \mathbf{Pa}(\mathbf{B}))$
(iii) For each $A_i \in \mathbf{A}$, (iii.a) $A_i \in \mathbf{Pa}(\mathbf{B})$ (and) (iii.b) $C(\{A_i\} \cup \mathbf{B}) = \{\{A_i\} \cup \mathbf{B}\}$
(iv) For each $P_i \in \mathbf{Pa}(\mathbf{B}) \setminus \mathbf{A}$, $C(\{P_i\} \cup \mathbf{B}) = \{\{P_i\}, \mathbf{B}\}$.

In other words, if $\mathbf{A}, \mathbf{B}$ form a bush, then (i) $\mathbf{B}$ is non-empty; (ii) $\mathbf{B}$ is a maximal C-component of $G[\mathbf{V} \setminus \mathbf{Pa}(\mathbf{B})]$; (iii) Each $A_i \in \mathbf{A}$ a) has at least one child in $\mathbf{B}$ (and) b) share a bi-directed edge with at least one vertex in $\mathbf{B}$; (iv) no vertex in $(\mathbf{Pa}(\mathbf{B}) \setminus \mathbf{A})$ share a bi-directed edge to a vertex in $\mathbf{B}$ (See Figure 3).

Example 6. There are three bushes in the graph H shown in Figure 1a: (i) $\mathbf{A_1} = \{\}, \mathbf{B_1} = \{X_1, Y_1, Y_2\}$; (ii) $\mathbf{A_2} = \{\}$, $\mathbf{B_2} = \{X_2\}$; and (iii) $\mathbf{A_3} = \{X_1\}, \mathbf{B_3} = \{Y_1, Y_2\}$.

Claim 3. *For a given causal graph G, there is a one-to-one correspondence between $\mathsf{ILD}(G)$, and the set of all bush and assignment pairs of G*

$$\textit{i.e.,} \quad \bigcup_{\textit{Bushes } \mathbf{A}, \mathbf{B}} \; \bigcup_{\mathbf{pa}(\mathbf{B})} \{((\mathbf{A}, \mathbf{B}), \mathbf{pa}(\mathbf{B}))\}.$$

Proof. Every $\mathbf{B} \subseteq \mathbf{V}$ that respects $\mathbf{B} \in C(\mathbf{B} \setminus \mathbf{Pa}(\mathbf{B}))$ maps to a unique bush $\mathbf{A}, \mathbf{B}$, where $\mathbf{A} = \{A_i \in \mathbf{Pa}(\mathbf{B}) : C(\{A_i\} \cup \mathbf{B}) = \{\{A_i\} \cup \mathbf{B}\}\}$. Hence, every informative local distribution $\Pr[\mathbf{B} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}))] \in \mathsf{ILD}(G)$ maps to a unique bush $\mathbf{A}, \mathbf{B}$ and assignment $\mathbf{pa}(\mathbf{B})$ pair.

Furthermore, every bush $\mathbf{A}, \mathbf{B}$ and assignment $\mathbf{pa}(\mathbf{B})$ pair maps to a unique informative local distribution $\Pr[\mathbf{B} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}))] \in \mathsf{ILD}(G)$. □

For a given bush and assignment pair for a causal graph G, we define what we call an *informative intervention set*, which plays an important role in identifying the corresponding informative local distribution from a given joint information set.

Definition 10 (Informative Intervention Set)**.** *Given a causal graph G, and a bush and assignment pair $((\mathbf{A}, \mathbf{B}), \mathbf{pa}(\mathbf{B}))$ of G, the* informative intervention set $\mathtt{IIS}(\mathbf{a}\,;\, \mathbf{pa}(\mathbf{B}) \setminus \mathbf{a}\,;\, \mathbf{B})$ *is a joint information set*[3] *that contains the set of all interventions* I *such that*

1. I *intervenes on all the vertices of* $\mathbf{A}$ *with assignment* $\mathbf{a}$.
2. I *does not intervene on any vertex in* $\mathbf{B}$.
3. I *and* $\mathbf{pa}(\mathbf{B})$ *are consistent on* $\mathbf{Pa}(\mathbf{B})$.

Example 7. Consider the bush $\mathbf{A_3} = \{X_1\}, \mathbf{B_3} = \{Y_1, Y_2\}$ of H. For a given assignment x_1, x_2, $\mathsf{IIS}(x_1, x_2, \{Y_1, Y_2\})$ is a joint information set that contains at least one of the following interventional distributions: a) $\Pr[Y_1, Y_2, X_2 \mid do(x_1)]$; b) $\Pr[Y_1, Y_2 \mid do(x_1, x_2)]$.

Theorem 4 shows that bushes and IISs can be used to characterize the identifiability of informative local distributions $\mathsf{ILD}(G)$ from $\mathsf{IS}^{\mathsf{inp}}(G)$, a given joint information set of a causal model characterized by G.

Theorem 4. *Given a causal graph G, a bush and assignment pair $((\mathbf{A}, \mathbf{B}), \mathbf{pa}(\mathbf{B}))$ of G, and a joint information set $\mathsf{IS}^{\mathsf{inp}}(G)$, the distribution $\Pr[\mathbf{B} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}))]$ is identifiable from $\mathsf{IS}^{\mathsf{inp}}(G)$, if and only if, $\mathsf{IS}^{\mathsf{inp}}(G) \cap \mathtt{IIS}(\mathbf{a}\,;\, \mathbf{pa}(\mathbf{B}) \setminus \mathbf{a}\,;\, \mathbf{B})$ is non-empty.*

The proof of Theorem 4 is given in the Appendix. Theorem 4 asserts that an informative local distribution $\Pr[\mathbf{B} \mid \mathrm{do}(\mathbf{pa}(\mathbf{B}))] \in \mathsf{ILD}(G)$ is uniquely determinable from $\mathsf{IS}^{\mathsf{inp}}(G)$, if and only if, $\mathsf{IS}^{\mathsf{inp}}(G)$ contains an intervention from the corresponding informative intervention set. We will now illustrate how the concept of bushes and Theorem 4 can be used to find minimum intervention covers. Consider the causal graphs shown in Figure 1. For simplicity, we assume that all the variables are boolean.

Example 8 ($\mathsf{MIC}(H)$)**.** Consider the graph H (shown in Figure 1a), where the only possible bushes are: (i) $\mathcal{B}_1 : \mathbf{A_1} = \{\}, \mathbf{B_1} = \{X_1, Y_1, Y_2\}$; (ii) $\mathcal{B}_2 : \mathbf{A_2} = \{\}, \mathbf{B_2} = \{X_2\}$; and (iii) $\mathcal{B}_3 : \mathbf{A_3} = \{X_1\}, \mathbf{B_3} = \{Y_1, Y_2\}$.

With respect to bush $\mathcal{B}_1$, for each assignment $\mathbf{pa}(\mathbf{B}_1)$ of $\mathbf{Pa}(\mathbf{B}_1)$, i.e., for each $x_2 \in \{0, 1\}$, identifying the informative local distribution $\Pr[\mathbf{B}_1 \mid \mathbf{pa}(\mathbf{B}_1)]$, i.e., $\Pr[X_1, Y_1, Y_2 \mid do(x_2)]$, requires that any minimum intervention cover include an intervention from $\mathtt{IIS}(\emptyset\,;\, x_2\,;\, \{X_1, Y_1, Y_2\})$. Similarly, with respect to bush $\mathcal{B}_2$, for each assignment $x_1 \in \{0, 1\}$, identifying $\Pr[X_2 \mid do(x_1)]$ requires that any minimum intervention cover include an intervention from $\mathtt{IIS}(\emptyset\,;\, x_1\,;\, \{X_2\})$; and for bush $\mathcal{B}_3$, identifying $\Pr[Y_1, Y_2 \mid do(x_1, x_2)]$ for each assignment of $(x_1, x_2) \in \{0, 1\}^2$ requires that $\mathsf{MIC}(H)$ include an intervention from $\mathtt{IIS}(x_1\,;\, x_2\,;\, \{Y_1, Y_2\})$.

Note that the observable distribution intersects the informative intervention sets corresponding to bushes $\mathcal{B}_1$ and $\mathcal{B}_2$. Similarly, for each $x_1 \in \{0, 1\}$, $do(x_1)$ intersects the informative intervention set $\mathtt{IIS}(x_1\,;\, x_2\,;\, \{Y_1, Y_2\})$ that correspond to bush $\mathcal{B}_3$ for both values of $x_2 \in \{0, 1\}$. By claims 1 and 2, we know that the informative local distributions $\mathsf{ILD}(H)$ are sufficient to identify the set of all interventional distributions $\mathsf{IS}^*(H)$. Hence, it is easy to see

[3] Note that $\mathbf{a}$ is the assignment of $\mathbf{A}$ consistent with $\mathbf{pa}(\mathbf{B})$.

that the observable distribution together with $do(X_1 = 0)$ and $do(X_1 = 1)$ form a minimum intervention cover of H, because any other information set with fewer than 3 interventions cannot intersect all the required IISs (a fact that can be verified by brute-force enumeration).

Example 9 ($\mathsf{MIC}(H')$)**.** Now consider the graph H' which contains an additional bi-directed edge $X_2 \leftrightarrow Y_1$. H' contains four different bushes: (i) $\mathcal{B}_1 : \mathbf{A}_1 = \{\}, \mathbf{B}_1 = \{X_1, X_2, Y_1, Y_2\}$; (ii) $\mathcal{B}_2 : \mathbf{A}_2 = \{X_1\}, \mathbf{B}_2 = \{X_2, Y_1, Y_2\}$; (iii) $\mathcal{B}_3 : \mathbf{A}_3 = \{X_2\}, \mathbf{B}_3 = \{X_1, Y_1, Y_2\}$; and (iv) $\mathcal{B}_4 : \mathbf{A}_4 = \{X_1, X_2\}, \mathbf{B}_4 = \{Y_1, Y_2\}$.

By Theorem 4, for bushes $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4$, identifying the respective local distributions, i.e., $\Pr[\mathbf{B}_\mathrm{i} \mid \mathrm{do}(\mathsf{pa}(\mathbf{B}_\mathrm{i}))]$s, requires (1) the observable distribution, (2) $do(x_1)$ for each $x_1 \in \{0,1\}$, (3) $do(x_2)$ for each $x_2 \in \{0,1\}$, and (4) $do(x_1, x_2)$ for each $(x_1, x_2) \in \{0,1\}^2$ respectively. Also note that no two IISs have any intervention in common. Hence, the size of $\mathsf{MIC}(H')$ is 9.

Based on the previous examples, one might be tempted to conclude that when the graph is a C-component, minimum intervention cover must include all the possible interventions (except the trivial interventions that target a leaf node). However, this is not true because *structure* of the C-component plays a crucial role in determining the minimum intervention cover.

Example 10 ($\mathsf{MIC}(H'')$)**.** Consider the graph H'' shown in Figure 1c which consists of the bushes: (i) $\mathcal{B}_1 : \mathbf{A}_1 = \{\}, \mathbf{B}_1 = \{X_1, X_2, Y_1, Y_2\}$; (ii) $\mathcal{B}_2 : \mathbf{A}_2 = \{X_1\}, \mathbf{B}_2 = \{X_2\}$; (iii) $\mathcal{B}_3 : \mathbf{A}_3 = \{X_1\}, \mathbf{B}_3 = \{Y_1, Y_2\}$; and (iv) $\mathcal{B}_4 : \mathbf{A}_4 = \{X_2\}, \mathbf{B}_4 = \{X_1, Y_1, Y_2\}$.

By Theorem 4, we know that $\mathcal{B}_1$ requires the observable distribution, and $\mathcal{B}_4$ requires the two interventions $do(X_2 = 0)$ and $do(X_2 = 1)$ to be included in $\mathsf{MIC}(H'')$. Also, the two interventions $do(X_1 = 0)$ and $do(X_1 = 1)$ intersect the informative intervention sets that arise from bushes $\mathcal{B}_2$ and $\mathcal{B}_3$. It is easy to verify by brute force that no information set of size less than 5 can intersect all the required IISs. Hence the size of $\mathsf{MIC}(H'')$ is 5.

An Efficient Algorithm for $\mathsf{MIC}(G)$

Based on the results presented in the previous section, it is possible to find $\mathsf{MIC}(G)$ by exhaustively enumerating the subsets of IS^*(G) in increasing order of size, and checking whether the subset being considered is an intervention cover of G. First, such an approach is impractical for causal graphs with more than a few variables. Second, it is unclear whether such a brute-force approach is optimal. Hence, we proceed to study the computational complexity of $\mathsf{MIC}(G)$. Specifically, we exploit the structural properties of bushes to reduce the problem of finding $\mathsf{MIC}(G)$ to that of finding a minimum vertex coloring of an *undirected* graph $\widehat{G}$ obtained from G. Minimum vertex coloring of a graph with vertices $\mathbf{W}$ and edges E is the well-known problem of minimizing the number of colors required to assign colors c_{W_i} to vertices $W_i \in \mathbf{W}$ such that $(W_i, W_j) \in E \implies c_{W_i} \neq c_{W_j}$. If we denote the subset of vertices that are assigned the color c_i by $\mathbf{W}_{c_i}$, then a vertex coloring of a graph corresponds to

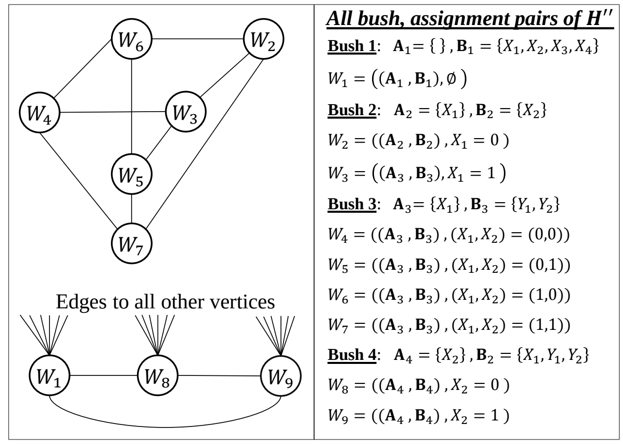


Figure 4: $\widehat{H''}$ obtained by the reduction from H''

a partition of the vertices into (disjoint) subsets where each subset is assigned a distinct color and no subset contains an edge. Minimum vertex coloring is NP-complete (Garey and Johnson 1990).

Reduction of $\mathsf{MIC}(G)$ to Minimum Vertex Coloring

We proceed to describe how to construct the graph $\widehat{G}$ from G and how to use the resulting graph $\widehat{G}$ to compute $\mathsf{MIC}(G)$.

We define the vertex set of $\widehat{G}$ as follows: With each bush and assignment pair $((\mathbf{A}, \mathbf{B}), \mathsf{pa}(\mathbf{B}))$ for a causal model characterized by G, we associate a vertex in $\widehat{G}$. We define the edge set of $\widehat{G}$ as follows: Two vertices $W_i = ((\mathbf{A}_i, \mathbf{B}_i), \mathsf{pa}(\mathbf{B}_i))$ and $W_j = ((\mathbf{A}_j, \mathbf{B}_j), \mathsf{pa}(\mathbf{B}_j))$ of $\widehat{G}$ share an edge if and only if there exists no intervention that identifies both $\Pr[\mathbf{B}_i \mid do(\mathsf{pa}(\mathbf{b}_i))]$ and $\Pr[\mathbf{B}_j \mid do(\mathsf{pa}(\mathbf{b}_j))]$. By Theorem 4, this translates to the following definition.

Definition 11 (Edge set of $\widehat{G}$)**.** *There exists an edge (conflict edge) between two distinct vertices* $W_i = ((\mathbf{A}_i, \mathbf{B}_i), \mathsf{pa}(\mathbf{B}_i))$ *and* $W_j = ((\mathbf{A}_j, \mathbf{B}_j), \mathsf{pa}(\mathbf{B}_j))$ *of* $\widehat{G}$, *if one of the following conditions hold:*

1. $\mathbf{A}_i \cap \mathbf{B}_j$ *is non-empty;*
2. $\mathbf{A}_j \cap \mathbf{B}_i$ *is non-empty;*
3. $\mathbf{a}_i$ *and* $\mathsf{pa}(\mathbf{b}_j)$ *are inconsistent;*
4. $\mathbf{a}_j$ *and* $\mathsf{pa}(\mathbf{b}_i)$ *are inconsistent.*

Given an undirected graph $\widehat{G}$ with the above specifications of vertices and edges, any minimum vertex coloring of $\widehat{G}$ corresponds to a minimum intervention cover of G.

Example 11 ($\mathsf{MIC}(H'')$)**.** Consider the graph H'' (Figure 1c). For each bush and assignment pair of H'', there exists a vertex in $\widehat{H''}$ (Figure 4). The edges are constructed according to Definition 11. Note that any vertex coloring would contain three distinct colors for W_1, W_8 and W_9. It is easy to see that no other vertex can use these colors. Hence, let $\mathbf{W}_{c_1} := \{W_1\}$, $\mathbf{W}_{c_2} := \{W_8\}$ and $\mathbf{W}_{c_3} := \{W_8\}$.

Also, it is easy to see that coloring the remaining vertices requires at least two colors. Suppose $\mathbf{W}_{c_4} := \{W_2, W_4, W_5\}$ and $\mathbf{W}_{c_5} := \{W_3, W_6, W_7\}$. From Theorem 4, it follows that the interventions $do(\emptyset)$, $do(X_2 = 0)$, $do(X_2 = 1)$, $do(X_1 = 0)$ and $do(X_1 = 1)$ will identify all the informative local distributions (bush assignment pairs) in $\mathbf{W}_{c_1}$, $\mathbf{W}_{c_2}$, $\mathbf{W}_{c_3}$, $\mathbf{W}_{c_4}$, $\mathbf{W}_{c_5}$ respectively.

The following lemma is useful in designing an efficient algorithm to construct the vertex set of $\widehat{G}$.

Lemma 5. *Let $\mathbf{A}', \mathbf{B}'$ form a bush for G. Then for all $\mathbf{A} \subset \mathbf{A}'$, there is a bush $\mathbf{A}, \mathbf{B}$ for G such that $\mathbf{B} \supseteq \mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A})$.*

Proof. Let $\mathbf{A} \subset \mathbf{A}'$. Since $\mathbf{A}', \mathbf{B}'$ form a bush for G,

- $C(\mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A})) = \{\mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A})\}$ is singleton, because every vertex of $\mathbf{A}'$ has bi-directed edge to a vertex in $\mathbf{B}'$.
- $\mathbf{A} \subseteq \mathsf{Pa}(\mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A}))$, because $\mathbf{A} \subseteq \mathsf{Pa}(\mathbf{B}')$ by bush definition.

Therefore, there exists a $\mathbf{B} \supseteq \mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A})$ such that $\mathbf{A}, \mathbf{B}$ form a bush for G (where $\mathbf{B}$ will be the maximal C-component of $G[\mathbf{V} \setminus \mathbf{A}]$ that contains $\mathbf{B}'$). $\square$

Algorithm 1 $\mathsf{MIC}(G)$

1: $\mathtt{IB} \leftarrow$ FINDALLBUSHES(G)
2: $\mathbf{W} \leftarrow$ VERTEXCONSTRUCTION($\mathtt{IB}, G$)
3: Let $\mathbf{W}$ be the vertex set of G^{co}.
4: Construct the edges of G^{co} using Definition 11.
5: Find a minimum vertex coloring of G^{co}.
Let $\mathbf{W}_c$ be the vertices colored using color c.
6: For each color c, let $\mathrm{I}_c = \Pr[\mathbf{V} \setminus \mathbf{A}_c \mid do(\mathbf{a}_c)]$
where, $\mathbf{A}_c = \left(\bigcup_{((\mathbf{A},\mathbf{B}),\mathsf{pa}(\mathbf{b}))\in\mathbf{W}_c} \mathbf{A}\right)$
and $\mathbf{a}_c = \left(\bigcup_{((\mathbf{A},\mathbf{B}),\mathsf{pa}(\mathbf{b}))\in\mathbf{W}_c} \mathbf{a}\right)$ [4].
7: **return** $\bigcup_c \mathrm{I}_c$.

Algorithm 2 FINDALLBUSHES(G)

Initialize $\mathtt{IB}_0 = \bigcup_{\mathbf{B}_j \in C(\mathbf{V})} \{(\{\}, \mathbf{B}_j)\}$
for $i \leftarrow 1$ *to* n **do**
  $\mathtt{IB}_i = \{\}$
  if $\mathtt{IB}_{i-1}$ is empty **then** break
  for all $(\mathbf{A}, \mathbf{B}) \in \mathtt{IB}_{i-1}$ **do**
    for all $W \in \mathbf{B}$ **do**
      $\mathtt{B}$=FINDPAIRS($\mathbf{A} \cup \{W\}, \mathbf{B} \setminus \{W\}, G$)
      $\mathtt{IB}_i \leftarrow \mathtt{B} \cup \mathtt{IB}_i$
return $\bigcup_{i=0}^{n} \mathtt{IB}_i$

[4]Here $\mathbf{a}$ is an assignment of $\mathbf{A}$ consistent with $\mathsf{pa}(\mathbf{B})$. Since no two vertices in $\mathbf{W}_c$ are connected by an edge, the union operation results an unambiguous assignment $\mathbf{a}_c$, due to conditions 3 and 4 of Definition 11.

Algorithm 3 FINDPAIRS($\mathbf{A}', \mathbf{B} \setminus \{W\}, G$)

Initialize $\mathtt{B} = \{\}$
for all $\mathbf{B}' \in C(\mathbf{B} \setminus \{W\})$ **do**
  if $\mathbf{A}', \mathbf{B}'$ is a bush **then**
    $\mathtt{B} \leftarrow \mathtt{B} \cup (\mathbf{A}', \mathbf{B}')$
return $\mathtt{B}$

Algorithm 4 VERTEXCONSTRUCTION($\mathtt{IB}$,G)

Initialize pairs $\mathcal{P} = \{\}$
for all $(\mathbf{A}, \mathbf{B}) \in \mathtt{IB}$ **do**
  for all ($\mathsf{pa}(\mathbf{B}) \in \Sigma^{|\mathsf{Pa}(\mathbf{B})|}$)[5] **do**
    Add a new vertex $((\mathbf{A}, \mathbf{B}), \mathsf{pa}(\mathbf{B}))$ to $\mathcal{P}$
return $\mathcal{P}$

Our procedure for finding $\mathsf{MIC}(G)$ is shown in Algorithm 1. In order to find the vertex set of $\widehat{G}$ from the given graph G, first the algorithm stores the set of all bushes in the set $\mathtt{IB}$. This step is efficiently executed by making use of Lemma 5. For a non-negative integer i, let $\mathtt{IB}_i$ represent the set of all bushes $\mathbf{A}, \mathbf{B}$ such that $|\mathbf{A}| = i$. Given $\mathtt{IB}_i$, the algorithm determines $\mathtt{IB}_{i+1}$ as follows: For every bush $(\mathbf{A}, \mathbf{B}) \in \mathtt{IB}_i$ and $W \in \mathbf{B}$, bushes of the form $(\mathbf{A}' = \mathbf{A} \cup \{W\}, \mathbf{B}')$ are added to $\mathtt{IB}_{i+1}$. By Lemma 5, we know that every bush $\mathbf{A}', \mathbf{B}'$ with $|\mathbf{A}'| = i+1$ will get added to $\mathtt{IB}_{i+1}$ from some bush $\mathbf{A}, \mathbf{B} \in \mathtt{IB}_i$ (where $\mathbf{A} = \mathbf{A}' \setminus \{W\}$ for some $W \in \mathbf{B}$) in this process. Hence, at the end of this process, $\mathtt{IB}$ will contain the set of all bushes of G. Next, for every bush $\mathbf{A}, \mathbf{B} \in \mathtt{IB}$ and assignment $\mathsf{pa}(\mathbf{B})$ pair, the algorithm creates a vertex $((\mathbf{A}, \mathbf{B}), \mathsf{pa}(\mathbf{B}))$ and adds it to $\mathbf{W}$. Thus $\mathbf{W}$ forms the vertex set of $\widehat{G}$. The edges are added according to Definition 11.

The algorithm then proceeds to find a minimum vertex coloring of $\widehat{G}$. For each color c, the algorithm defines an intervention I_c such that I_c identifies $\Pr[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{b}))]$ for every $((\mathbf{A}, \mathbf{B}), \mathsf{pa}(\mathbf{b})) \in \mathbf{W_c}$, which follows from Theorem 4, Definition 11, the definition of I_c as in line 6 of Algorithm 1, and the fact that $\mathbf{W}_c$ forms an independent set (i.e., no two vertices in $\mathbf{W}_c$ are connected by an edge). Hence, $\bigcup_c \mathrm{I}_c$ identifies ILD. From Claims 1 and 2, we know that $\mathsf{ILD}(G)$ identifies $\mathsf{LD}(G)$, and $\mathsf{LD}(G)$ identifies $\mathsf{IS}^*(G)$. **Hence, $\bigcup_c \mathrm{I}_c$ returned by the algorithm is an intervention cover of G.**

Next we prove (by contradiction) that $\bigcup_c \mathrm{I}_c$ is indeed a *minimum* intervention cover of G. Suppose there exists a joint information set $\mathtt{IS}'$ of size smaller than $\bigcup_c \mathrm{I}_c$ that identifies $\mathsf{ILD}(G)$. For each $\mathrm{I} \in \mathtt{IS}'$, suppose, using Theorem 4, we identify vertices $((\mathbf{A}, \mathbf{B}), \mathsf{pa}(\mathbf{B}))$ in $\widehat{G}$ such that $\Pr[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{b}))]$ is identifiable from I, *i.e.,* $\mathrm{I} \in \mathtt{IIS}(\mathbf{a}\,;\,\mathsf{pa}(\mathbf{B}) \setminus \mathbf{a}\,;\,\mathbf{B})$. By Definition 11, all such vertices form an independent set. Also, for each vertex $((\mathbf{A}, \mathbf{B}), \mathsf{pa}(\mathbf{B}))$ of $\widehat{G}$ there must exist an intervention $\mathrm{I} \in \mathtt{IS}'$ such that $\mathrm{I} \in \mathtt{IIS}(\mathbf{a}\,;\,\mathsf{pa}(\mathbf{B}) \setminus \mathbf{a}\,;\,\mathbf{B})$, by Theo-

[5]When the variables take values from different alphabets, $\mathsf{pa}(\mathbf{B})$ will loop over all possible assignments of $\mathsf{Pa}(\mathbf{B})$.

rem 4. It therefore follows that $\bigcup_c \mathrm{I}_c$ cannot be a minimum vertex coloring of $\widehat{G}$.

The proof of completeness of do-calculus follows from the proof of Theorem 4.

Theorem 6. *Do-calculus is complete for* $\mathsf{MIC}(G)$.

Proof. Follows from the fact that all of the results needed to prove Theorem 4 have been obtained using only the 3 rules of do-calculus along with basic manipulations. □

Complexity of Minimum Intervention Cover

The complexity of Algorithm 1 is a function of the complexity of the minimum vertex coloring for the class of graphs $\Gamma = \{\widehat{G} | G \text{ is a causal graph}\}$. While the minimum vertex coloring for general graphs is known to be NP-complete, it is unclear whether this hardness result necessarily holds for the class of graphs Γ. In what follows, we show that (i) the size of $\mathsf{MIC}(G)$ can be large when the size of the largest C-component is large; and (ii) when the degree of the causal graph G and the C-component sizes are bounded, the size of the minimum intervention is small, and the runtime of Algorithm 1 is polynomial in the size of G.

Lemma 7. *For a given causal graph G, suppose there exists a bush* $\mathbf{A}', \mathbf{B}'$ *such that* $|\mathbf{A}'| = k$. *Then the size of* $\mathsf{MIC}(G)$ *is at least* $(|\Sigma| + 1)^k$.

Proof. The proof directly follows from Lemma 5. We know that for any $\mathbf{A} \subseteq \mathbf{A}'$, there exists a bush $\mathbf{A}, \mathbf{B}$. Let $((\mathbf{A}, \mathbf{B}), \kappa)$ and $((\mathbf{A}, \mathbf{B}), \tau)$ be vertices of $\widehat{G}$ such that κ, and τ, are inconsistent on $\mathbf{A}$. Then there is an edge between the above vertices in $\widehat{G}$, because the assignments over $\mathbf{A}$ are inconsistent. Hence, for each such bush $\mathbf{A}, \mathbf{B}$, over all the possible assignments of $\mathbf{A}$, there are $|\Sigma|^{|\mathbf{A}|}$ vertices in $\widehat{G}$ that form a clique.

Let $\mathbf{A}_i$ and $\mathbf{A}_j$ be distinct subsets of $\mathbf{A}'$. From Lemma 5 we know the existence of bushes $\mathbf{A}_i, \mathbf{B}_i$ and $\mathbf{A}_j, \mathbf{B}_j$. Note that there is an edge in $\widehat{G}$ between $((\mathbf{A}_i, \mathbf{B}_i), \kappa)$ and $((\mathbf{A}_j, \mathbf{B}_j), \tau)$ for any assignments κ and τ, because[6] $\mathbf{A}_i \cap \mathbf{B}_j \neq \emptyset$ or $\mathbf{A}_j \cap \mathbf{B}_i \neq \emptyset$. The preceding claims together imply that the vertices described above must form a clique of size $(|\Sigma|+1)^k$ in $\widehat{G}$ $\left(\text{because } \sum_{i=0}^{k} |\Sigma|^i \binom{k}{i} = (|\Sigma| + 1)^k\right)$. Hence the coloring number of $\widehat{G}$ is at least $(|\Sigma| + 1)^k$. □

The preceding lemma provides a characterization of the class of graphs for which the size of $\mathsf{MIC}(G)$ is super-polynomial in the size of G.

Corollary 7.1. *For a given causal graph G, the size of* $\mathsf{MIC}(G)$ *is super-polynomial in* n *if there exists a bush* $\mathbf{A}, \mathbf{B}$ *such that* $|\mathbf{A}|$ *is* $\omega(\log n)$.

Next we show that when the C-component size and the sum of the in-degree (excluding the unobservable parents) and out-degree of each vertex in the causal graph G are bounded, then MIC is small.

[6] Recall that $\mathbf{B}_i \supseteq (\mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A}_i))$ and $\mathbf{B}_j \supseteq (\mathbf{B}' \cup (\mathbf{A}' \setminus \mathbf{A}_j))$.

Lemma 8. *For a given causal graph G, suppose the sum of the in-degree (excluding the unobserved parents) and out-degree of each vertex of G is bounded by d, and the size of each C-component is bounded by p. Then the degree of the undirected graph,* $\widehat{G}$, *constructed by Algorithm 1, is at most* $2^{2(p+pd^2)}|\Sigma|^{pd^3}$.

Proof. Fix a vertex $W = ((\mathbf{A}, \mathbf{B}), \mathbf{pa}(\mathbf{B})) \in \mathbf{W}$ of the undirected graph $\widehat{G}$ obtained using the reduction of Algorithm 1. Since the C-component size is at most p, the size of $\mathbf{A} \cup \mathbf{B}$ is at most p. For any vertex $W' = ((\mathbf{A}', \mathbf{B}'), \mathbf{pa}(\mathbf{B}'))$ that share an edge with W, we know that there does not exist an intervention I that satisfies $\mathrm{I} \in \mathtt{IIS}(\mathbf{a}\,;\, \mathbf{pa}(\mathbf{B}) \setminus \mathbf{a}\,;\, \mathbf{B})$ and $\mathrm{I} \in \mathtt{IIS}(\mathbf{a}'\,;\, \mathbf{pa}(\mathbf{B}') \setminus \mathbf{a}'\,;\, \mathbf{B}')$. That is, W' satisfies one of the following properties:

1. $\mathbf{A} \cap \mathbf{B}'$ is non-empty
2. $\mathbf{B} \cap \mathbf{A}'$ is non-empty
3. $\mathbf{a}$ and $\mathbf{pa}(\mathbf{B}')$ are inconsistent
4. $\mathbf{a}'$ and $\mathbf{pa}(\mathbf{B})$ are inconsistent.

Note that the size of $\mathbf{A}, \mathbf{A}', \mathbf{B}$ and $\mathbf{B}'$ is at most p. Hence for a fixed W, the total number of W's satisfying the first condition above is at most 2^p, and similarly, the number of W' satisfying the second condition is at most 2^p. Since the in-degree is bounded by d, $|\mathbf{Pa}(\mathbf{B})|$ is at most pd. Also, the out-degree of the graph is bounded by d, implying that there can be at most pd^2 children of $\mathbf{Pa}(\mathbf{B})$. Hence for a fixed W, the total number of W's satisfying the third condition is at most $2^{pd^2}|\Sigma|^{pd^3}$, and the fourth condition is at most $2^{pd^2}|\Sigma|^{pd^3}$. Thus, the degree of W is bounded by at most $2^{2p(1+d^2)}|\Sigma|^{2pd^3}$. □

As the chromatic number of any graph of degree r is at most $r + 1$, we obtain a characterization of a class of causal graphs where the size of MIC is constant.

Corollary 8.1. *When the size of the C-component and the degree of the causal graph G are bounded by constants, the size of* $\mathsf{MIC}(G)$ *is* $O(1)$.

Note that the number of vertices of $\hat{G}$ is $O(n)$ when the size of the C-component and the sum of the in-degree and out-degree of the causal graph G are bounded by constants, and hence $\mathsf{MIC}(G)$ can be computed in time that is polynomial in the number of variables. Arguably, causal models that are readily communicable to humans need to be "simple", and hence likely to have small in-degree and out-degree and C-components of bounded size.

Summary and Discussion

We have provided an algorithm that, given a causal graph G, computes $\mathsf{MIC}(G)$, a *minimum intervention cover* of G, i.e., a minimum set of interventions that suffice for identifying *every* causal effect of a causal model that is characterized by G. We have established the completeness of do-calculus for the minimum intervention cover problem. $\mathsf{MIC}(G)$ effectively offers an efficient compilation of all of the information that is obtainable from observations and interventions relative to a causal graph G in anticipation of all possible causal queries that are answerable by any causal model with structure specified by G. These results find applications in a variety of contexts, including in particular, counterfactual inference, and generalizing causal effects across experimental settings. Work in progress is aimed at generalizing the definition of $\mathsf{MIC}(G)$ relative to *an arbitrary subset of feasible interventions*, as opposed to all possible interventions on $\mathbf{V}$. This paper focused on minimizing the *number of interventions*, and *not* the *number of variables targeted by each intervention*. It would be interesting to consider variants of $\mathsf{MIC}(G)$ that minimize the sum of the numbers of variables targeted by all interventions, or minimize over both the number of interventions as well as the the number of variables targeted by the interventions.

Appendix: Proof of Theorem 4

Before proving Theorem 4, first we recall the hedge structure of (Shpitser and Pearl 2006) which determines the identifiability of causal effects from the observable distribution of a causal graph G.

Definition 12 (Hedge)**.** *For a given causal graph G, two disjoint variables $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, and assignments $\mathbf{x}, \mathbf{y}$, there exists a hedge for $\Pr[\mathbf{y} \mid do(\mathbf{x})]$ in G, if there exists two subgraphs[7] F, F' of G such that*

- *F and F' are C-components.*
- *F is a subset[8] of F'.*
- *The leaf vertices[9] of F and F' are common.*
- *F' includes at least one vertex from $\mathbf{X}$.*
- *F does not include any vertex from $\mathbf{X}$.*
- *Each observable variable has at most one outgoing edge in F and F'.*
- *The set of leaf nodes of F' is a subset of $\mathsf{An}(\mathbf{Y})_{G_{\overline{\mathbf{X}}}} \cup \mathbf{Y}$.*

Theorem 9 (Identifying causal effects from observable distribution (Shpitser and Pearl 2006))**.** *For a given causal graph G, $\Pr[\mathbf{y} \mid do(\mathbf{x})]$ is identifiable from the observable distribution $\Pr[\mathbf{V}]$, if and only if, there does not exist a hedge for $\Pr[\mathbf{y} \mid do(\mathbf{x})]$ in G.*

Now we are ready to prove Theorem 4.

Soundness proof, Theorem 4. The if part of Theorem 4 easily follows from the known identification algorithm of (Shpitser and Pearl 2006). Let M be the probabilistic causal model M_G defined over the causal graph G.

Suppose there exists an intervention $\Pr^M[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})] \in \mathsf{IS}^{\mathsf{inp}}$ such that $\Pr^M[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})] \in \mathtt{IIS}(\mathbf{a}\,;\, \mathsf{pa}(\mathbf{B})_G \setminus \mathbf{a}\,;\, \mathbf{B})$. Our goal is to show that $\Pr^M[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{B})_G)]$ is identifiable from $\Pr^M[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})]$.

Note that the intervention $do(\mathbf{s})$ induces a new model N on the graph $G' = G[\mathbf{V} \setminus \mathbf{S}]$, which essentially simulates the interventional distribution $\Pr^M[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})]$ as its observable distribution $\Pr^N[\mathbf{V} \setminus \mathbf{S}]$ by excluding the variables $\mathbf{S}$ in the causal graph G'. Formally, N is a probabilistic causal model, with observable variables $\mathbf{V} \setminus \mathbf{S}$, on the causal graph $G' = G[\mathbf{V} \setminus \mathbf{S}]$ such that:

- The unobservable variables $\mathbf{U}$, and the underlying distribution $\Pr[\mathbf{U}]$ in N remains the same as M.
- For each vertex $V \in \mathbf{V} \setminus \mathbf{S}$, the function f_V in the induced model N behaves the same as the corresponding function in M, when restricted to the assignment $\mathbf{s}$.

From the above definition,

$$\Pr^N[\mathbf{V} \setminus \mathbf{S}] = \Pr^M[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})]. \tag{1}$$

By the definition of IIS, we know that $\mathbf{s}$ and $\mathsf{pa}(\mathbf{B})_G$ are consistent, and that $\mathbf{B}$ and $\mathbf{S}$ are disjoint. Also, $\mathsf{Pa}(\mathbf{B})_{G'} = \mathsf{Pa}(\mathbf{B})_G \setminus \mathbf{S}$. Hence, it follows that:

$$\Pr^N[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{B})_{G'})] = \Pr^M[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{B})_G)]. \tag{2}$$

where $\mathsf{pa}(\mathbf{B})_{G'}$ is the assignment consistent with $\mathsf{pa}(\mathbf{B})_G$.

Also, since $\mathbf{A} \subseteq \mathbf{S}$, and $\mathbf{B}$ does not have a bi-directed edge to any vertex of $\mathbf{V} \setminus (\mathbf{B} \cup \mathbf{A})$ (refer Figure 3), we know that $\mathbf{B}$ is a maximal C-component in the graph G'. Note that there can not exist a subgraph of G', F', that (i) is a C-component; and (ii) contains a vertex from $\mathsf{Pa}(\mathbf{B})_{G'}$. Hence there can not exist a hedge (Shpitser and Pearl 2006) for $\Pr^N[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{B})_{G'})]$ in the graph G', which implies that $\Pr^N[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{B})_{G'})]$ is identifiable from the observable distribution $\Pr^N[\mathbf{V} \setminus \mathbf{S}]$ of N. This, combined with Equations (1) and (2), imply that the required informative local distribution $\Pr^M[\mathbf{B} \mid do(\mathsf{pa}(\mathbf{B})_G)]$ is identifiable form $\Pr^M[\mathbf{V} \setminus \mathbf{S} \mid do(\mathbf{s})]$. □

Completeness proof, Theorem 4. Omitted [10]. □

Acknowledgements

Saravanan Kandasamy was supported in part by the DRDO Frontiers Project DRDO0687, and by Ramanujan grant SB/S2/RJN-020/2017 of DST India at the Tata Institute of Fundamental Research. Arnab Bhattacharyya was supported at IISc by Ramanujan grant DSTO1358 and the DRDO Frontiers Project DRDO0687 and at NUS by an AcRF Tier 1 grant on "Inference and Testing of Sparse Models in High Dimensions". Vasant G Honavar was supported in part by the National Center for Advancing Translational Sciences, NIH through the grant UL1 TR000127 and TR002014, by the NSF, USA, through the grants 1518732, 1640834, and 1636795, the Pennsylvania State University's Institute for Cyberscience and the Center for Big Data Analytics and Discovery Informatics, the Edward Frymoyer Endowed Professorship in Information Sciences and Technology at Pennsylvania State University and the Sudha Murty Distinguished Visiting Chair in Neurocomputing and Data Science funded by the Pratiksha Trust at the Indian Institute of Science.The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

[7] A subgraph of G is a graph obtained from G by excluding some vertices and edges.

[8] F is a subset of F' if all the vertices and edges of F are also present in F'.

[9] A vertex with no child is called a leaf vertex.

[10] See supplementary material available at https://www.comp.nus.edu.sg/~arnab/aaai19-sup.pdf

References

Bareinboim, E., and Pearl, J. 2013a. Causal transportability with limited experiments. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 95–101.

Bareinboim, E., and Pearl, J. 2013b. Meta-transportability of causal effects: A formal approach. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, 135–143.

Bareinboim, E., and Pearl, J. 2014. Transportability from multiple environments with limited experiments: Completeness results. In *Advances in Neural Information Processing Systems 27*. 280–288.

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

Bareinboim, E.; Lee, S.; Honavar, V.; and Pearl, J. 2013. Transportability from multiple environments with limited experiments. In *Advances in Neural Information Processing Systems 26*. 136–144.

Berzuini, C.; Dawid, P.; and Bernardinell, L. 2012. *Causality: Statistical perspectives and applications*. John Wiley & Sons.

Galles, D., and Pearl, J. 1995. Testing identifiability of causal effects. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 185–195.

Garey, M. R., and Johnson, D. S. 1990. *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co.

Halpern, J. Y. 1998. Axiomatizing causal reasoning. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 202–210.

Hernan, M. A., and Robins, J. M. 2010. *Causal inference*. CRC Boca Raton, FL.

Huang, Y., and Valtorta, M. 2006. Identifiability in causal bayesian networks: a sound and complete algorithm. In *Proceedings of the 21st National Conference on Artificial intelligence*, 1149–1154. AAAI Press.

Hyttinen, A.; Eberhardt, F.; and Järvisalo, M. 2015. Do-calculus when the true graph is unknown. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence*, 395–404.

Imbens, G. W., and Rubin, D. B. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kocaoglu, M.; Shanmugam, K.; and Bareinboim, E. 2017. Experimental design for learning causal graphs with latent variables. In *Advances in Neural Information Processing Systems 30*, 7018–7028.

Kuroki, M., and Miyakawa, M. 1999. Identifiability criteria for causal effects of joint interventions. *Journal of the Japan Statistical Society* 29(2):105–117.

Lee, S., and Honavar, V. 2013a. Causal transportability of experiments on controllable subsets of variables: z-transportability. In *Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence*, 361–370.

Lee, S., and Honavar, V. 2013b. m-transportability: Transportability of a causal effect from multiple environments. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 583–590.

Morgan, S. L., and Winship, C. 2014. *Counterfactuals and causal inference*. Cambridge University Press.

Pearl, J., and Bareinboim, E. 2011. Transportability of causal and statistical relations: A formal approach. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 247–254.

Pearl, J., and Robins, J. 1995. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 444–453.

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

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. New York, USA: Cambridge University Press, 2nd edition.

Peters, J.; Janzing, D.; and Schölkopf, B. 2017. *Elements of causal inference: foundations and learning algorithms*. MIT press.

Robins, J. 1986. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling* 7:1393–1512.

Shanmugam, K.; Kocaoglu, M.; Dimakis, A. G.; and Vishwanath, S. 2015. Learning causal graphs with small interventions. In *Advances in Neural Information Processing Systems 28*, 3195–3203.

Shpitser, I., and Pearl, J. 2006. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, 1219–1226.

Shpitser, I., and Pearl, J. 2008. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research* 9:1941–1979.

Spirtes, P.; Glymour, C.; and Scheines, R. 2000. *Causation, Prediction, and Search*. MIT press, 2nd edition.

Tian, J., and Pearl, J. 2002a. A general identification condition for causal effects. In *Proceedings of the 18th National Conference on Artificial Intelligence*, 567–573.

Tian, J., and Pearl, J. 2002b. On the testable implications of causal models with hidden variables. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, 519–527.

Tsamardinos, I.; Triantafillou, S.; and Lagani, V. 2012. Towards integrative causal analysis of heterogeneous data sets and studies. *Journal of Machine Learning Research* 13:1097–1157.

Verma, T., and Pearl, J. 1990. Causal networks: Semantics and expressiveness. In *Proceedings of the 4th Conference on Uncertainty in Artificial Intelligence*, 69–78.