TECHNICAL REPORT
R-46
First version: May 2019
Last version: Jul 2022

General Identifiability with Arbitrary Surrogate Experiments

Sanghack Lee[*] and **Juan D. Correa** and **Elias Bareinboim**
Causal Artificial Intelligence Laboratory
Department of Computer Science
Columbia University, USA

Abstract

We study the problem of causal identification from an arbitrary collection of observational and experimental distributions, and substantive knowledge about the phenomenon under investigation, which usually comes in the form of a causal graph. We call this problem *g-identifiability*, or gID for short. The gID setting encompasses two well-known problems in causal inference, namely, identifiability [Pearl, 1995] and z-identifiability [Bareinboim and Pearl, 2012] — the former assumes that an observational distribution is necessarily available, and no experiments can be performed, conditions that are both relaxed in the gID setting; the latter assumes that *all* combinations of experiments are available, i.e., the power set of the experimental set $\mathbf{Z}$, which gID does not require a priori. In this paper, we introduce a general strategy to prove non-gID based on *hedgelets* and *thickets*, which leads to a necessary and sufficient graphical condition for the corresponding decision problem. We further develop a procedure for systematically computing the target effect, and prove that it is sound and complete for gID instances. In other words, failure of the algorithm in returning an expression implies that the target effect is not computable from the available distributions. Finally, as a corollary of these results, we show that do-calculus is complete for the task of g-identifiability.

1 INTRODUCTION

One of the main tasks in the empirical sciences and data-driven disciplines is to infer cause and effect relationships from a combination of observations, experiments, and substantive knowledge about the phenomenon under investigation. Causal relations are deemed desirable and valuable for constructing explanations and for contemplating novel interventions that were never experienced before [Pearl, 2000, Spirtes et al., 2001, Bareinboim and Pearl, 2016, Pearl and Mackenzie, 2018].

In one line of investigation, this task is formalized through the question of whether the effect that an intervention on a set of variables $\mathbf{X}$ will have on another set of outcome variables $\mathbf{Y}$ (denoted $P_{\mathbf{x}}(\mathbf{y})$) can be uniquely computed from the probability distribution P over the observed variables $\mathbf{V}$ and a causal diagram $\mathcal{G}$. This is known as the problem of *identification* [Pearl, 1995, 2000, Bareinboim and Pearl, 2016], and has received great attention in the literature, starting with a number of sufficient conditions [Spirtes et al., 2001, Galles and Pearl, 1995, Pearl and Robins, 1995], and culminating in a complete graphical and algorithmic characterization [Tian and Pearl, 2002, Shpitser and Pearl, 2006, Huang and Valtorta, 2006]. Despite the generality of such results, it's the case that in some real-world applications the quantity $P_{\mathbf{x}}(\mathbf{y})$ is not identifiable (i.e., not uniquely computable) from the observational data and the causal diagram.

On an alternative thread in the literature, causal effects ($P_{\mathbf{x}}(\mathbf{y})$) are obtained directly through controlled experimentation [Fisher, 1951]. In the biomedical sciences, for instance, considerable resources are spent every year by the FDA, the NIH, and others, in supporting large-scale, systematic, and controlled experimentation, which comes under the rubric of *Randomized Controlled Trials* (RCTs). The same method is also leveraged in the context of reinforcement learning (RL), for example, when an autonomous agent is deployed in an environment and is given the capability of performing interventions and observing how they unfold in time. Through this process, experimental data is gathered, and used in the construction of a strategy, also known as policy, with the goal of optimizing the agent's cumulative reward (e.g., survival,

[*]This work was done while the authors were at Purdue University. Corresponding author's email: sl4712@columbia.edu.

profitability, happiness). Despite all the inferential power entailed by this approach, there are real-world settings where controlling the variables in $\mathbf{X}$ is not feasible, possibly due to economical, technical, or ethical constraints.

In this paper, we note that these two approaches can be seen as extremes in a spectrum of possible research designs, which can be combined to solve very natural, albeit non-trivial, causal inference problems. In fact, this generalized setting has been investigated in the literature under the rubric of *z-identifiability* (zID, for short) [Bareinboim and Pearl, 2012].[1] Formally, zID asks whether $P_{\mathbf{x}}(\mathbf{y})$ can be uniquely computed from the combination of the observational distribution $P(\mathbf{V})$ and the experimental distributions $P_{\mathbf{z}'}(\mathbf{V})$, for all $\mathbf{Z}' \subseteq \mathbf{Z}$, for some $\mathbf{Z} \subseteq \mathbf{V}$. We highlight two critical assumptions underlying this setting. First, note that it may be challenging to intervene on some of the subsets of the variables, which means that the original zID assumption that experiments over *all* subsets of $\mathbf{Z}$ are available may not be attainable in some real-world applications; e.g., for $\mathbf{Z}=\{Z_1, Z_2\}$, zID assumes that experimental data over $\{\{\}, \{Z_1\}, \{Z_2\}, \{Z_1, Z_2\}\}$ are available. Second, zID assumes that the observational distribution (null intervention) is always available, which, while attainable in many settings, may be hard to measure in others. For instance, when an RL agent has to act in an environment where no behavior policy exists. For concreteness, we discuss below two applications where these assumptions are shown to be too stringent.

Example 1. (Drug-drug interactions) Consider the causal graphs in Fig. 1, where Y represents *cardiovascular disease*, W *blood pressure*, X_1 taking an *antihypertensive* drug, and X_2 the use of an *anti-diabetic* drug. While it's currently understood that diabetes and hypertension do not affect each other (no direct link between them), it's common for patients with type 2 diabetes to be susceptible to hypertension, since both conditions share important confounding factors (graphically encoded through the bidirected arrows) [Ferrannini and Cushman, 2012]. The goal of the analysis is to assess the effect of prescribing a treatment for both conditions on the risk of developing cardiovascular diseases, $P_{x_1,x_2}(y)$. There are two RCTs that separately control for X_1 and X_2, which means that $P_{x_1}(\mathbf{V})$ and $P_{x_2}(\mathbf{V})$ are available. These distributions do not satisfy the requirements of zID, which expects all combinations of experiments, including $P_{x_1,x_2}(\mathbf{V})$ itself, the very target of the analysis. It turns out that for the models in Figs. 1a, b, $P_{x_1,x_2}(y) = \sum_w P_{x_2}(y|w)P_{x_1}(w)$, which means that the experimental studies suffice to identify the joint effect. The same effect is not identifiable in Figs. 1c, d (for further details, see Appendix A.1). □

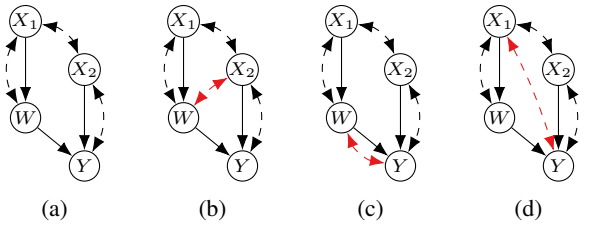


Figure 1: $P_{x_1,x_2}(y)$ can be identified from $P_{x_1}(\mathbf{V})$ and $P_{x_2}(\mathbf{V})$ in (a) and (b), but not in (c) and (d). Differences among the causal diagrams are highlighted in red.

Example 2. (Multivariate testing (MVT)) The causal graphs in Fig. 2 represent simplified models of a large-scale experimentation platform of a hypothetical Internet company. There, X_1, X_2 represent a set of features, and Y_1, Y_2 different outcome variables (e.g., click-through rates, users' happiness). The various teams perform online experiments varying a diverse set of features at the same time, and then track the changes in the different outcome variables. This procedure is known as multivariate testing (MVT), which generalizes A/B testing. In practice, MVT allows the company to estimate the joint experimental distribution $P_{x_1,x_2}(\mathbf{V})$. The goal is to identify the impact of changes in individual features, say $P_{x_1}(y_1)$ and $P_{x_2}(y_2)$, so that the teams can be rewarded based on their specific contributions. In Fig. 2a, given experiments performed simultaneously, $P_{x_1,x_2}(\mathbf{y})$, each of the team's announced outcomes can be obtained as $P_{x_1}(y_1) = P_{x_1,x_2}(y_1)$ and $P_{x_2}(y_2) = P_{x_1,x_2}(y_2)$. On the other hand, the individual effects are not identifiable from $P_{x_1,x_2}(\mathbf{y})$ in Fig. 2b and c (for more details, see Appendix A.1). □

Our goal in this paper is to explicate the conditions under which inferences such as the ones discussed above are licensed from first principles. More broadly, we investigate the problem of general identification of causal effects from an arbitrary combination of observational and experimental distributions, and substantive knowledge specified in the form of a causal graph, which we call g-identification (for short, gID). Specifically, our contributions are as follows: 1. We prove a necessary and sufficient graphical condition for gID, which follows from two new graphical constructions called *hedgelets* and *thickets*. These structures constitute flexible and general building blocks that are helpful to understand and characterize general identification problems (Sec. 3); 2. Leveraging these results, we develop a sound and complete algorithm that returns any expression derivable from an arbitrary collection of observations and experiments. As a corollary, we prove that do-calculus is complete for g-identification (Sec. 4).

[1] This problem can be seen as closely related to the non-parametric version of instrumental variables (IVs), but for when the combination of both observational and experimental data is available. By and large, but not exclusively, the literature on IVs is mostly focused on some parametric settings.

2 PRELIMINARIES

We denote variables by capital letters, X, and values by small letters, x. Bold letters, $\mathbf{X}$ or $\mathbf{x}$, represent sets of variables or values. The domain of a variable X is denoted by $\mathfrak{X}_X$. Two values $\mathbf{x}$ and $\mathbf{z}$ are said to be consistent if they share the common values for $\mathbf{X} \cap \mathbf{Z}$. We also denote by $\mathbf{x} \setminus \mathbf{Z}$ the value of $\mathbf{X} \setminus \mathbf{Z}$ consistent with $\mathbf{x}$. We assume that the domain of every variable is finite.

Our analysis heavily relies on causal graphs, to which we often assign a calligraphic letter, e.g., $\mathcal{G}$, $\mathcal{F}$, or $\mathcal{H}$. We denote by $\mathbf{V}(\mathcal{H})$ the set of vertices (i.e., variables) in a graph $\mathcal{H}$. A vertex-induced subgraph is denoted by brackets, e.g., $\mathcal{G}[\mathbf{W}]$, which includes $\mathbf{W}$ and the edges among its elements. We define $\mathcal{G} \setminus \mathbf{X}$ as $\mathcal{G}[\mathbf{V}(\mathcal{G}) \setminus \mathbf{X}]$. A root set of a graph is a set of variables that does not have outgoing edges. We use kinship notation for graphical relationships such as parents, children, descendants, and ancestors of a set of variables. For example, the set of parents of $\mathbf{X}$ in $\mathcal{G}$ is denoted by $pa(\mathbf{X})_{\mathcal{G}} := \bigcup_{X \in \mathbf{X}} pa(X)_{\mathcal{G}}$. Similarly, we define ch, de, and an. Written as Pa, Ch, De, and An (i.e., capitalized), the argument is included as well, e.g., $De(\mathbf{X})_{\mathcal{G}} := de(\mathbf{X})_{\mathcal{G}} \cup \mathbf{X}$. We denote by $\boldsymbol{\pi}$ a topological ordering of vertices in $\mathcal{G}$, and $\mathbf{V}_{\boldsymbol{\pi}}^{(i-1)}$ to be the set of observable variables preceding V_i in $\boldsymbol{\pi}$. A path consisting of only bidirected edges is called a bidirected path.

We use Structural Causal Models (SCMs) [Pearl, 2000] as our basic semantical framework. An SCM $\mathcal{M}$ is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{U}) \rangle$, where $\mathbf{U}$ is a set of exogenous variables; $\mathbf{V}$ is a set of endogenous variables; $\mathbf{F}$ is a set of functions $\{f_V\}_{V \in \mathbf{V}}$, which determines the value of a variable, e.g., $v \leftarrow f_V(\mathbf{pa}_V, \mathbf{u}^V)$ is a function with $\mathbf{PA}^V \subseteq \mathbf{V} \setminus \{V\}$ and $\mathbf{U}^V \subseteq \mathbf{U}$; and $P(\mathbf{U})$ is a joint probability distribution over $\mathbf{U}$. An SCM $\mathcal{M}$ induces a causal graph $\mathcal{G}$ (also called a semi-Markovian graph) where $\mathbf{V}$ is a set of vertices, directed edges are formed satisfying $\mathbf{PA}^V = pa(V)_{\mathcal{G}}$, and each bidirected edge corresponds to an unobserved confounder between two variables, that is, $V_i \leftrightarrow V_j$ if $\mathbf{U}^i$ and $\mathbf{U}^j$ correlate. Interventions are defined through an operator called $do(\mathbf{X} = \mathbf{x})$, which sets the intervened variables $\mathbf{X}$ to specific values $\mathbf{x} \in \mathfrak{X}_{\mathbf{X}}$. Given a model $\mathcal{M}$, an intervention $do(\mathbf{X} = \mathbf{x})$ induces a submodel $\mathcal{M}_{\mathbf{x}}$, where f_X of $\mathbf{F}$ is replaced by $f_X = x$ for every $X \in \mathbf{X}$ where x is consistent with $\mathbf{x}$. This submodel $\mathcal{M}_{\mathbf{x}}$ induces a causal graph $\mathcal{G}_{\overline{\mathbf{X}}}$, which reads as $\mathcal{G}$ with edges onto any of $\mathbf{X}$ removed.

We now revisit some key notions for deciding identifiability developed in the context of non-experimental settings. First, we define a special type of cluster of variables called *confounded components* [Tian and Pearl, 2002].

Definition 1 (C-component)**.** Let $\mathcal{G}$ be a semi-Markovian graph such that a subset of its bidirected arcs forms a spanning tree over all vertices in $\mathcal{G}$. Then $\mathcal{G}$ is a c-component.

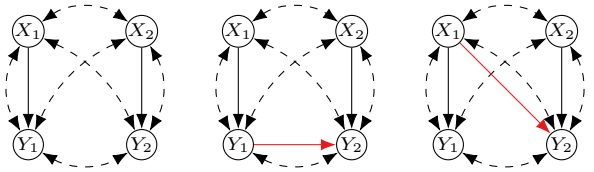


Figure 2: Given experiments available on $\{X_1, X_2\}$, both causal distributions $P_{x_1}(y_1)$ and $P_{x_2}(y_2)$ are identifiable in (a), but $P_{x_2}(y_2)$ is not identifiable in (b) and (c). In all cases, $P_{x_1}(y_2)$ and $P_{x_2}(y_1)$ are not identifiable.

Given a semi-Markovian graph $\mathcal{G}$ over a set of variables $\mathbf{V}$, there exists a unique partition such that each subgraph is a maximal c-component. We denote by $\mathcal{C}(\mathcal{G})$ the set of c-components that partitions the vertices in $\mathcal{G}$, so that $\mathcal{C}(\mathcal{G}) = \{\mathbf{W}_i\}_{i=1}^k$ implies that $\mathcal{G}[\mathbf{W}_i]$ is a c-component, for each $\mathbf{W}_i \subseteq \mathbf{V}$, and there is no bidirected edge between $\mathbf{W}_i$ and $\mathbf{W}_j$ in $\mathcal{G}$ for $i \neq j$. Armed with this definition, we build towards the hedge with the following notion adapted from [Shpitser and Pearl, 2006].

Definition 2 (C-forest)**.** A semi-Markovian graph $\mathcal{G}$ with root set $\mathbf{R}$ is said to be an $\mathbf{R}$-rooted c-forest if $\mathcal{G}$ is a c-component with a minimal number of edges.

The minimality with respect to the number of edges guarantees that every vertex not in the root set of a c-forest has one child and its bidirected edges form exactly a spanning tree. We are now ready to define a hedge as follows.

Definition 3 (Hedge)**.** A hedge is a pair of $\mathbf{R}$-rooted c-forests $\langle \mathcal{F}, \mathcal{F}' \rangle$ such that $\mathcal{F}' \subseteq \mathcal{F}$.

To realize the connection between definitions, note that given disjoint sets $\mathbf{X}, \mathbf{Y} \subset \mathbf{V}$, if $\mathbf{R} \subseteq An(\mathbf{Y})_{\mathcal{G}_{\overline{\mathbf{X}}}}$, $\mathcal{F} \cap \mathbf{X} \neq \emptyset$, and $\mathcal{F}' \cap \mathbf{X} = \emptyset$, Def. 3 reduces to the original definition. The existence of such structure precludes the identifiability of $P_{\mathbf{x}}(\mathbf{y})$ from $P(\mathbf{V})$ [Shpitser and Pearl, 2006]. Moreover, as it will become evident throughout this paper, tying a hedge to a particular effect constrains its use in tasks other than classic identification. In the new treatment pursued in this paper, we separate a hedge as a graphical structure itself from its use as a witness of the non-identifiability of a specific causal distribution. We say that the new hedge structure $\langle \mathcal{F}, \mathcal{F}' \rangle$ is *formed for* $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ (i.e., $\mathcal{G}$ has the hedge structure as a subgraph relative to $\mathbf{X}$ and $\mathbf{Y}$) whenever referring to the original semantics, i.e., regarding the non-identifiability of $P_{\mathbf{x}}(\mathbf{y})$.

Further, we will distinguish two parts of a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$: the 'top' part, denoted by $\mathcal{F}'' = \mathcal{F} \setminus \mathbf{V}(\mathcal{F}')$, and the 'bottom' part, which is $\mathcal{F}'$. When the top is empty (i.e., $\mathcal{F} = \mathcal{F}'$), we will call this hedge *degenerate*. The UCs

between the top and bottom parts are called 'crossing UCs'. And the variables transmitting their values through the crossing directed edges will be called 'frontiers'.

3 G-IDENTIFIABILITY

We first introduce a new task that formalizes and generalizes the identifiability and z-identifiability settings by allowing a more flexible input consisting of any combination of observational and experimental distributions.

Definition 4 (g-Identifiability)**.** Let $\mathbf{X}$, $\mathbf{Y}$ be disjoint sets of variables, $\mathbb{Z} = \{\mathbf{Z}_i\}_{i=1}^m$ be a collection of sets of variables, and let $\mathcal{G}$ be a causal diagram. $P_{\mathbf{x}}(\mathbf{y})$ is said to be g-identifiable from $\mathbb{Z}$ in $\mathcal{G}$, if $P_{\mathbf{x}}(\mathbf{y})$ is uniquely computable from positive distributions $\{P(\mathbf{V} \setminus \mathbf{Z} \mid do(\mathbf{z}))\}_{\mathbf{Z}\in\mathbb{Z},\mathbf{z}\in\mathfrak{X}_{\mathbf{Z}}}$ in any causal model which induces $\mathcal{G}$.

A traditional and pervasive assumption made throughout the identification literature is that a probability distribution describing the natural state of the system is available, that is, $P(\mathbf{V})$. In the setting defined above, such distribution is not a priori required unless the empty set is explicitly included in $\mathbb{Z}$. The following statement can be shown based on the definition of g-identifiability:

Lemma 1. *Let $\mathbf{X}$, $\mathbf{Y}$ be disjoint sets of variables, $\mathbb{Z} = \{\mathbf{Z}_i\}_{i=1}^m$ be a collection of sets of variables, and let $\mathcal{G}$ be a causal diagram. $P_{\mathbf{x}}(\mathbf{y})$ is not g-identifiable from $\mathbb{Z}$ in $\mathcal{G}$ if there exist two causal models $\mathcal{M}_1$ and $\mathcal{M}_2$ compatible with $\mathcal{G}$ such that $P^1_{\mathbf{z}}(\mathbf{v}) = P^2_{\mathbf{z}}(\mathbf{v})$ for all $\mathbf{Z} \in \mathbb{Z}$, $\mathbf{z} \in \mathfrak{X}_{\mathbf{Z}}$, but $P^1_{\mathbf{x}}(\mathbf{y}) \neq P^2_{\mathbf{x}}(\mathbf{y})$.*

Proof. The inequality eliminates the possibility of the existence of a function from available experimental distributions to $P_{\mathbf{x}}(\mathbf{y})$ given $\mathcal{G}$. □

Even though this statement formally characterizes non-g-identifiability of a certain data collection, it does not provide any insight on how to determine if such pair of models exist, or how to construct them when a given instance is not g-identifiable. If not ambiguous, we omit the prefix g- and use the term identifiability to convey its non-technical generic meaning.

HEDGELETS AND THICKETS

When considering multiple experimental distributions as inputs, a graphical structure that might be able to witness the non-g-identifiability has to account for all available experiments. To deal with the complexity added by a broader input, we introduce *hedgelets*, a unique decomposition of a hedge. Based on this decomposition, we will demonstrate a new way of proving non-identifiability, in the context of the more general task of g-identifiability.

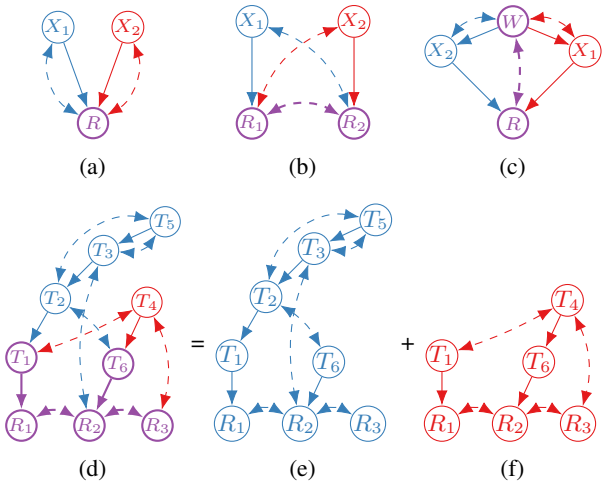


Figure 3: Hedgelet decomposition of hedges and a thicket (color-coded in blue and red with purple for shared elements). Each of (a) and (b) is a hedge formed for $P_{\mathbf{r}}(\mathbf{x})$ or a thicket with respect to $\mathbb{Z} = \{\{X_1\}, \{X_2\}\}$ while (c) is not a hedge but a thicket. The hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ in (d) is decomposed into (e) $\mathcal{F}(\{T_2, T_3, T_5, T_6\})$ (f) $\mathcal{F}(\{T_1, T_4\})$.

We define how to obtain the set of hedgelets associated with any given hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$.

Definition 5 (hedgelet decomposition)**.** The hedgelet decomposition of a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ is the collection of hedgelets $\{\mathcal{F}(\mathbf{W})\}_{\mathbf{W}\in\mathcal{C}(\mathcal{F}'')}$ where each hedgelet $\mathcal{F}(\mathbf{W})$ is a subgraph of $\mathcal{F}$ made of (i) $\mathcal{F}[\mathbf{V}(\mathcal{F}') \cup \mathbf{W}]$ and (ii) $\mathcal{F}[De(\mathbf{W})_{\mathcal{F}}]$ without bidirected edges.

Analogous to a hedge, a hedgelet $\mathcal{H}$ has a top section and a bottom section.

Let $\mathbb{H}_{\mathcal{F}} = \{\mathcal{F}(\mathbf{W})\}_{\mathbf{W}\in\mathcal{C}(\mathcal{F}'')}$ be the set of *hedgelets* of $\langle \mathcal{F}, \mathcal{F}' \rangle$. For a degenerate hedge, $\mathbb{H}_{\mathcal{F}}$ contains a single hedgelet $\mathcal{F}(\emptyset) = \mathcal{F}$, which we call a *degenerate* hedgelet. Given a non-degenerate hedge, for every hedgelet $\mathcal{H}$ in it, there exists at least one directed edge, and exactly one bidirected edge (i.e., a crossing UC) between the top and bottom sections by definition.[2]

For a simple example, see Fig. 3a, a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ for $P_{\mathbf{x}}(\mathbf{r})$. This hedge can be decomposed into two hedgelets $\mathcal{F}(\{X_1\})$ in blue (i.e., $\mathcal{G}[\{X_1, R\}]$) and $\mathcal{F}(\{X_2\})$ in red (i.e., $\mathcal{G}[\{X_2, R\}]$). Fig. 3b is a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ for $P_{\mathbf{x}}(\mathbf{r})$, which can be similarly decomposed into two hedgelets $\mathcal{F}(\{X_1\})$ and $\mathcal{F}(\{X_2\})$. For another example, consider a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ in Fig. 3d. The top $\mathcal{F}'' = \mathbf{T}$ decomposes

[2]Since directed edges of $\mathcal{F}$ form a forest with all its roots in $\mathcal{F}'$, there must be a directed edge between them. If there exists no bidirected edge (or more than one bidirected edge) between them, it contradicts the fact that $\mathcal{F}$ is a c-component (or $\mathcal{F}'$ is a c-component and $\mathcal{F}$ is a c-component with minimal edges due to Defs 2 and 3).

into $\mathbf{W}_1 = \{T_2, T_3, T_5, T_6\}$ and $\mathbf{W}_2 = \{T_1, T_4\}$.

For $\mathcal{H}_1 = \mathcal{F}(\mathbf{W}_1)$, shown in Fig. 3e, we first take $\mathcal{F}[\mathbf{W}_1 \cup \mathbf{R}]$, which is equivalent to $\mathcal{H}_1 \setminus \{T_1, T_4\}$. Then, $\mathcal{F}[De(\mathbf{W}_1)_{\mathcal{F}}]$ without bidirected edges is added, which is responsible for $T_2 \to T_1 \to R_1$, so that $\mathbf{W}_1 \subseteq an(\mathbf{R})_{\mathcal{H}_1}$. The same procedure is applied to obtain $\mathcal{H}_2 = \mathcal{F}(\mathbf{W}_2)$, shown in Fig. 3f. In this example, both hedgelets share common frontiers (i.e., $\{T_1, T_6\}$).

Now, we will describe a graphical structure relative to the available input distributions entailed by $\mathbb{Z}$, that precludes the g-identifiability of a causal effect $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$. That is, whenever $\mathcal{G}$ contains such structure, $P_{\mathbf{x}}(\mathbf{y})$ is not g-identifiable from $\{P_{\mathbf{z}}(\mathbf{V})\}_{\mathbf{Z}\in\mathbb{Z},\mathbf{z}\in\mathfrak{x}_{\mathbf{Z}}}$ in $\mathcal{G}$.

Definition 6 (Thicket)**.** Let $\mathbf{R}$ be a non-empty set of variables and $\mathbb{Z}$ be a collection of sets of variables in $\mathcal{G}$. A thicket $\mathcal{J} \subseteq \mathcal{G}$ is an $\mathbf{R}$-rooted c-component consisting of a minimal c-component over $\mathbf{R}$ and hedges

$$\mathbb{F}_{\mathcal{J}} = \{\langle \mathcal{F}_{\mathbf{Z}}, \mathcal{J}[\mathbf{R}]\rangle \mid \mathcal{F}_{\mathbf{Z}} \subseteq \mathcal{G} \setminus \mathbf{Z}, \mathbf{Z} \cap \mathbf{R} = \emptyset\}_{\mathbf{Z}\in\mathbb{Z}}.$$

Let $\mathbf{X}$, $\mathbf{Y}$ be disjoint sets of variables in $\mathcal{G}$. A thicket $\mathcal{J}$ is said to be formed for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ with respect to $\mathbb{Z}$ if $\mathbf{R} \subseteq An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}}$ and every hedgelet of each hedge $\langle \mathcal{F}_{\mathbf{Z}}, \mathcal{J}[\mathbf{R}]\rangle$ intersects with $\mathbf{X}$.

If $\mathbf{Z} \cap \mathbf{R} = \emptyset$ for some $\mathbf{Z} \in \mathbb{Z}$, a thicket can be viewed as a superimposition of hedges where each of them comes from a subgraph of the thicket obtained by excluding an available experiment that was not performed on any of $\mathbf{R}$. Otherwise if $\mathbf{Z} \cap \mathbf{R} \neq \emptyset$ for every $\mathbf{Z} \in \mathbb{Z}$, that is, every experiment disrupts $\mathbf{R}$, $\mathcal{J}$ will simply be a spanning tree over $\mathbf{R}$ with bidirected arcs. Whenever this is the case, we call this thicket *degenerate*, which consists of a degenerate hedge with a single degenerate hedgelet.

To illustrate see Figs. 3a to 3c. Each causal diagram is a thicket for $P_{\mathbf{x}}(\mathbf{r})$ with respect to $\mathbb{Z} = \{\{X_1\}, \{X_2\}\}$ with two hedges in red and blue where each hedge itself is a hedgelet. Fig. 4 illustrates a more involved thicket, which can be viewed as formed for a query $P_{a,f,g}(\mathbf{r})$ with experiments $\mathbb{Z} = \{\{A\}, \{D, F\}\}$.

Thicket, hedge, and hedgelet form a hierarchical structure where the former can be decomposed into the latter. These structures will be instrumental to our analysis of g-identifiability in the next sections.

3.1 NON-IDENTIFIABILITY WITH HEDGELET DECOMPOSITION

In this section, we focus on constructing two models demonstrating the non-identifiability of a query using a thicket $\mathcal{J}$ whose root set is denoted by $\mathbf{R}$ and top variables by $\mathbf{T}$ (i.e., $\mathbf{T} = \mathbf{V}(\mathcal{J}) \setminus \mathbf{R}$). Moreover, we allow a query $P_{\mathbf{x}}(\mathbf{r})$ to have $\mathbf{X} = \emptyset$, corresponding to an observational quantity, which is trivially identifiable in previous identifiability problems where an observational distribution $P(\mathbf{v})$ was always considered as one of the available distributions, but this is not obvious if only non-observational data is available.

Let $\mathbb{H} = \bigcup_{\langle \mathcal{F}, \mathcal{F}'\rangle \in \mathbb{F}_{\mathcal{J}}} \mathbb{H}_{\mathcal{F}}$, that is, the aggregation of all hedgelets induced by the hedges of $\mathcal{J}$. Let $\mathbb{H}(V)$ be the subset of $\mathbb{H}$ where $V \in \mathbf{V}$ appears. For a set of variables $\mathbf{V}'$, let $\mathbb{H}(\mathbf{V}') = \bigcup_{V\in\mathbf{V}'} \mathbb{H}(V)$.

Non-identifiability of a Causal Effect for a Non-degenerate Thicket We consider constructing two models $\mathcal{M}_1$ and $\mathcal{M}_2$ agreeing in the available distributions but yielding a different result for the causal effect. This section only considers non-degenerate thickets, and, hence, non-degenerate hedgelets.

To begin with, each bidirected edge in the thicket will be mapped to an unobserved confounder (random variable). We partition UCs in the thicket and denote by $\mathbf{U}'$, $\mathbf{U}''$, and $\mathbf{U}^{\times}$ the UCs among $\mathcal{J}[\mathbf{R}]$, $\mathcal{J} \setminus \mathbf{R}$, and the crossing UCs, respectively. The observable variables in the thicket is divided into $\mathbf{T} \triangleq \mathbf{V}(\mathcal{J}) \setminus \mathbf{R}$ and $\mathbf{R}$. $\mathbf{R}$ and $\mathbf{U}'$ are binary (i.e., a single bit) and other variables consist of k-bits where k is the number of hedges wherein the variable appears. Every bit of UCs is independent and uniformly distributed. We use $\wedge$, $\bigoplus$, and $\overline{v}$ to denote *and*, *exclusive-or*, and *bitwise-complement* operation, respectively.

We start by modeling a hedge-specific bit for $T \in \mathcal{F}_i \setminus \mathbf{R} \subseteq \mathbf{T}$ for each hedge $\langle \mathcal{F}_i, \mathcal{J}[\mathbf{R}]\rangle$ in $\mathbb{F}_{\mathcal{J}}$. Let $\mathbf{U}_i^{\times} \subseteq \mathbf{U}^{\times}$ and $\mathbf{V}_i^{\downarrow} \subseteq \mathbf{T}$ be the set of crossing UCs and frontiers, respectively, for hedge $\mathcal{F}_i$. If $|\mathbf{U}_i^{\times}| \equiv |\mathbf{V}_i^{\downarrow}| \pmod 2$, pick one $T_i^* \in \mathbf{V}_i^{\downarrow}$. If T_i^* is undefined or T is not T_i^*, define

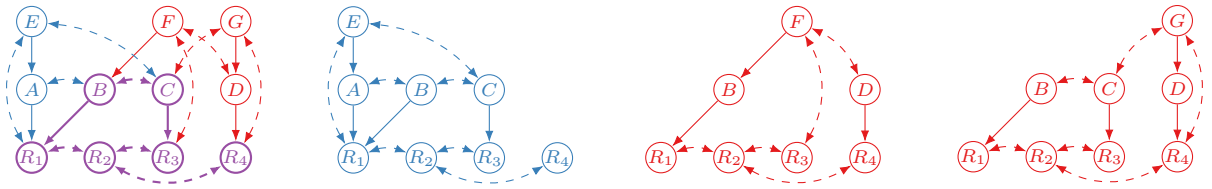
$$t_i \leftarrow \bigoplus\nolimits_{V\in pa(T)_{\mathcal{F}_i}} v_i \oplus \bigoplus\nolimits_{U\in\mathbf{U}^T_{\mathcal{F}_i}} u_i, \qquad (1)$$

where t_i, v_i, and u_i are the bits of T, V, and U corresponding to the hedge $\mathcal{F}_i$, and $\mathbf{U}^T_{\mathcal{F}_i}$ are UCs connected to T in $\mathcal{F}_i$. For T_i^*, we use the negation of the outcome of the above parametrization.

Consider $R \in \mathbf{R}$, then let $\mathbf{U}'^R \subseteq \mathbf{U}'$ be the UCs connected to R in $\mathcal{J}[\mathbf{R}]$, and let $\mathbf{U}^R \subseteq \mathbf{U}' \cup \mathbf{U}^{\times}$ be those connected to R in $\mathcal{J}$. Next, pick an arbitrary $R^* \in \mathbf{R}$ and define a function for $R \in \mathbf{R}$ in both models, except for R^* in $\mathcal{M}_2$ as follows.[3]

$$r \leftarrow \left(\bigwedge_{T\in pa(R)_{\mathcal{J}}} \mathbf{1}_{\bar{t}=0} \wedge \bigwedge_{U\in\mathbf{U}^R\setminus\mathbf{U}'^R} \mathbf{1}_{\bar{u}=0} \right) \wedge \bigoplus \mathbf{u}'^R. \quad (2)$$

[3]The $\bigwedge$ operator works as a universal quantifier and outputs 1 if its argument is empty, e.g., $pa(R)_{\mathcal{F}} = \emptyset$ or $\mathbf{U}^R \setminus \mathbf{U}'^R = \emptyset$.



(a) thicket (b) only hedgelet of the blue (c) 1st hedgelet of the red (d) 2nd hedgelet of the red

Figure 4: (a) Thicket with two hedges in red and blue with the shared parts in purple, and (b, c, d), their three hedgelets.

As for $R = R^*$ for $\mathcal{M}_2$ define:

$$r \leftarrow \left(\bigwedge_{T \in pa(R)_{\mathcal{J}}} \mathbf{1}_{\bar{t}=0} \wedge \bigwedge_{U \in \mathbf{U}^R \setminus \mathbf{U}'^R} \mathbf{1}_{\bar{u}=0} \right) \wedge \overline{\bigoplus \mathbf{u}'^R}, \quad (3)$$

where $\mathbf{1}_{\bar{u}=0}$ is 1 if every bit of u is 1, and 0 otherwise. To ensure the positivity, we can independently, randomly flip R with a small probability. For the sake of readability, we consider it as a separate post-process.

Now, we characterize this parametrization with respect to the distributions these two models generate. For simplicity, let $\mathbf{v}_i^{\downarrow}$ be the bits corresponding to the hedge $\mathcal{F}_i$ of the values of $\mathbf{V}_i^{\downarrow}$. Similarly we define $\mathbf{u}_i^{\times}$ for $\mathbf{U}_i^{\times}$.

Lemma 2. *Let $\langle \mathcal{F}_i, \mathcal{F}' \rangle$ be a hedge of a thicket, then the above parametrization before negation satisfies*

$$\bigoplus \mathbf{v}_i^{\downarrow} = \bigoplus \mathbf{u}_i^{\times}.$$

Proof. $\mathcal{F}'' \triangleq \mathcal{F}_i \setminus \mathcal{F}'$ is a $\mathbf{V}_i^{\downarrow}$-rooted forest where each $T \in \mathbf{V}_i^{\downarrow}$ is a root of a tree in the forest. Restricting our attention to the bits relevant to the hedge, by the parametrization, bits of $\mathcal{F}''$ carry the bit-parity of preceding UCs in $\mathcal{F}''$. Due to the forestness of the directed edges in $\mathcal{F}''$, taking the XOR of all $\mathbf{v}_i^{\downarrow}$ is equivalent to computing the XOR of all unobservable parents of variables in $\mathcal{F}''$. Since each one of such UCs is a parent of two variables in $\mathcal{F}''$, except for $\mathbf{u}_i^{\times}$, all but $\mathbf{u}_i^{\times}$ are counted twice. Due to the nature of XOR, repeated values cancel out and all that is left is the bit parity of $\mathbf{u}_i^{\times}$. □

Lemma 3. *Let $\mathbf{T}' \subsetneq \mathbf{T}$ such that there exists a hedgelet $\mathcal{H} \in \mathbb{H} \setminus \mathbb{H}(\mathbf{T}')$. Then, under the intervention $do(\mathbf{t}')$, there exists $R \in \mathbf{R}$, for any instantiation of $\mathbf{U}$, such that $r = 0$ in both models.*

Proof. Let $\langle \mathcal{F}_i, \mathcal{F}' \rangle$ be a hedge containing hedgelet $\mathcal{H}$. If $u_i = 0$ for some $U \in \mathbf{U}_i^{\times}$, then $\bar{u} \neq 0$ and the $R \in \mathbf{R}$ pointed by U in $\mathcal{J}$ will be 0.

Otherwise if $\bar{u} = 0$ for every $U \in \mathbf{U}_i^{\times}$, we will show that at least one $T \in \mathbf{V}_i^{\downarrow}$ satisfies $t_i = 0$, which leads to $\bar{t} \neq 0$. We investigate $\mathbf{v}_i^{\downarrow}$ before the negation utilizing Lemma 2. Let $\phi_i = |\mathbf{U}_i^{\times}| \bmod 2$ and $\psi_i = |\mathbf{V}_i^{\downarrow}| \bmod 2$.

- $\phi_i = 0, \psi_i = 0$: an even number of 1's and 0's exist.
- $\phi_i = 0, \psi_i = 1$: an odd number of 1's and 0's exist.
- $\phi_i = 1, \psi_i = 0$: an even number of 1's and an odd number of 0's exist.
- $\phi_i = 1, \psi_i = 1$: an odd number of 1's and an even number of 0's exist.

When $\phi_i \neq \psi_i$, an odd number of 0-value exists in $\mathbf{v}_i^{\downarrow}$. Then, done. Otherwise if $\phi_i = \psi_i$, an even number (including zero) of 0-value exists in $\mathbf{v}_i^{\downarrow}$. Negating exactly one frontier's bit, which either flips one of 1-value or one of two or more 0-values, ensures that there exists $t_i = 0$ for some $T \in \mathbf{V}_i^{\downarrow}$. As a consequence, $\bar{t} \neq 0$, and $r = 0$ for $R \in \mathbf{R}$ pointed by T for both models. □

Lemma 4. *Both models agree on $P_{\mathbf{t}'}(\mathbf{v})$ if there exists a hedgelet $\mathcal{H} \in \mathbb{H} \setminus \mathbb{H}(\mathbf{T}')$.*

Proof. Let us denote by superscript ^{1} and ^{2} for values for $\mathcal{M}_1$ and $\mathcal{M}_2$. First, given $\mathbf{u}^1 \setminus \mathbf{U}' = \mathbf{u}^2 \setminus \mathbf{U}'$, both models yield the same outcome for $\mathbf{T}$, i.e., $\mathbf{t}^1 = \mathbf{t}^2$. With at least one hedgelet intact, there must exist $R \in \mathbf{R}$ whose value must be 0 in both models regardless of the value of $\mathbf{U}'$ (Lemma 3). For readability, let us call such R as 'blacked-out' since its value is suppressed to 0 regardless of $\mathbf{U}'^R$. Then, the value of each non-blacked-out $R \in \mathbf{R}$ will be determined by $\mathbf{U}'^R$.

Now, we will prove that there exists an injective function from $\mathbf{u}'^1$ to $\mathbf{u}'^2$, which guarantees $\mathbf{v}^1 = \mathbf{v}^2$. Since $\mathcal{J}[\mathbf{R}]$ forms a spanning tree, there exists a bidirected path (including zero (edge) length) between any two vertices in $\mathbf{R}$. Consider a bidirected path $\mathbf{p}$ from the smallest (as defined by a topological ordering $\boldsymbol{\pi}$) blacked-out vertex to R^*.[4] If $\mathbf{u}^2$ is equivalent to $\mathbf{u}^1$ with the UCs in $\mathbf{p}$ negated, then, $\mathbf{r}^1 = \mathbf{r}^2$ since (see Fig. 5 for an example):

1. Each non-end vertex in the path is connected to exactly two negated UCs ensuring the bit-parity of the vertex is not changed; and

[4]The only required condition is to be consistent about the choice of a path given a set of blacked-out nodes.

2. The value of U at the end (other than R^*) does not affect the value of the blacked-out node.

Given the fact that $P(\mathbf{u}')$ is a uniform distribution for both models, they agree on $P_{\mathbf{t}'}(\mathbf{v})$. □

The proof for Lemma 4 can be directly applied to a more general case.

Corollary 1. *Both models agree on $P_{\mathbf{v}'}(\mathbf{v})$ for $\mathbf{V}' \subsetneq \mathbf{V}$ if there exists a hedgelet $\mathcal{H} \in \mathbb{H} \setminus \mathbb{H}(\mathbf{V}' \cap \mathbf{T})$.*

Proof. A bidirected path to negate $\mathbf{U}'$ can be found between R^* to either a blacked-out vertex or an intervened variable in $\mathbf{R}$. □

The agreement of the two models on the available distributions is the first piece to prove the non-identifiability of the causal effect. Now, we examine conditions under which the two constructed models disagree on a causal effect.

Lemma 5. *For a nonempty set $\mathbf{T}' \subseteq \mathbf{T}$ such that $\mathbb{H} = \mathbb{H}(\mathbf{T}')$, the two models disagree on $P_{\mathbf{t}'=\mathbf{1}}(\mathbf{r} = \mathbf{0})$.*

Proof. We first show that there exists a non-zero probability for none of $\mathbf{R}$ being blacked-out. By the condition, every hedgelet is intervened, that is, every hedge is intervened. In words, there exists an instantiation of $\mathbf{u}$ such that the term in parentheses of Eq. (2) is 1 for every $R \in \mathbf{R}$, that is, $P_{\mathbf{t}'=1}(\bar{\mathbf{v}}^{\downarrow} = \mathbf{0}, \bar{\mathbf{u}}^{\times} = \mathbf{0}) > 0$ for both models. Since the bits of top variables are independently modeled for hedges, we can simply show $P_{\mathbf{t}'=1}(\bar{\mathbf{v}}_i^{\downarrow} = \mathbf{0}, \bar{\mathbf{u}}_i^{\times} = \mathbf{0}) > 0$ for some hedge i.

We start by fixing $\mathbf{u}_i^{\times} = \mathbf{1}$ and setting the bits of $\mathbf{U}''_{\mathcal{F}_i}$ for the hedge to an arbitrary value. Select a frontier $W \in \mathbf{V}_i^{\downarrow}$ such that $w_i = 0$, which implies $W \notin \mathbf{T}'$ since $T \in \mathbf{T}'$ is fixed to $\mathbf{1}$. By the definition of hedgelet and the given condition that every hedgelet is intervened, W is connected via a bidirected path (non-zero length) to an ancestor (inclusive) of $T \in \mathbf{T}'$. Then, flipping the values (bits corresponding to the hedge) of the unobserved confounders along the bidirected path changes w_i to 1 but does not affect other frontiers' bits since either intermediate nodes are flipped twice or flipped once but the change is blocked by the intervened T. We can repeatedly apply the procedure for every hedge and frontier to ensure that every frontier is $\mathbf{1}$.

In such an event where there is no blacked-out node, $\mathcal{M}_1$ and $\mathcal{M}_2$ yield $\mathbf{r}^1$ and $\mathbf{r}^2$ such that $\bigoplus \mathbf{r}^1 = 0$ and $\bigoplus \mathbf{r}^2 = 1$. Combined with the fact that two models yield the same probability distributions when there exists a blacked-out node, two models disagree on $P_{\mathbf{t}'}(\mathbf{r})$ resulting $P^1_{\mathbf{t}'=1}(\mathbf{r} = \mathbf{0}) > P^2_{\mathbf{t}'=1}(\mathbf{r} = \mathbf{0})$. Independently flipping $\mathbf{R}$ with a small probability to ensure the positivity does not affect the disagreement. □

For example, consider Fig. 3d, which is also a thicket. Among $\mathbf{T}$, a pair of variables T_1 and T_6 are shared across hedgelets, while $\mathbf{T} \setminus \{T_1, T_6\}$ appear in only one of them. This implies that under the proposed parametrization, the two constructed models agree on, for example, $P_{t_2}(\mathbf{v})$, $P_{t_3,t_5}(\mathbf{v})$, $P_{t_2,t_3,r_2}(\mathbf{v})$, or $P_{t_4}(\mathbf{v})$. However, they will disagree on, distributions such as $P_{t_1}(\mathbf{v})$, $P_{t_6}(\mathbf{v})$, or $P_{t_2,t_4}(\mathbf{v})$. More formally, they agree on $P_{\mathbf{v}'}(\mathbf{v})$ where[5] $\mathbf{V}' \in 2^{\{T_2,T_3,T_5\} \cup \mathbf{R}} \cup 2^{\{T_4\} \cup \mathbf{R}}$ and they disagree on $P_{\mathbf{t}'}(\mathbf{r})$ for $\mathbf{T}' \subset 2^{\mathbf{T}}$, except for the aforementioned sets.

Non-identifiability of an Observational Probability for a Degenerate Thicket In g-identifiability, we also seek whether an observational probability, which was trivially identifiable in previous literature, can be uniquely determined by available experimental data. In this section, our focus is a degenerate thicket $\mathcal{J}$, which itself is a degenerate hedgelet $\mathcal{H}$, that is, $\mathcal{J} = \mathcal{J}[\mathbf{R}] = \mathcal{H}$. Consider identifying $P(\mathbf{r})$ given such $\mathcal{J}$.

We construct two models where $P(\mathbf{r})$ is not identifiable given experiments on *every* non-empty subset of $\mathcal{J}$. As in the previous section, R^* is an arbitrary variable in $\mathbf{R}$. For $R \in \mathbf{R}$ for $\mathcal{M}_1$ and $\mathcal{M}_2$, except R^* in $\mathcal{M}_2$, $r \leftarrow \bigoplus \mathbf{u}'^R$. For $R = R^*$ in $\mathcal{M}_2$, $r \leftarrow \overline{\bigoplus \mathbf{u}'^R}$. You may notice that this is exactly the same as Eqs. (2) and (3) in the previous section with the terms in parentheses explicitly removed — there is no input from the top.

Lemma 6. *Two models agree on $P_{\mathbf{r}'}(\mathbf{r})$ for $\emptyset \neq \mathbf{R}' \subseteq \mathbf{R}$.*

Proof. As in Lemma 4, we will show the existence of an injective function between $\mathbf{u}^1$ and $\mathbf{u}^2$ (note that $\mathbf{U} = \mathbf{U}'$ for a degenerate thicket). Find a bidirected path from one end at R^* and the other end at the smallest intervened variable. By using different values for the UCs in the path for $\mathbf{u}^1$ and $\mathbf{u}^2$, they will agree on $P_{\mathbf{r}'}(\mathbf{r})$. □

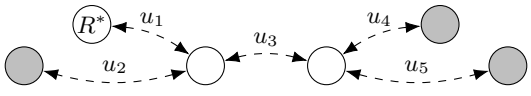
Lemma 7. *Two models disagree on $P(\mathbf{r})$.*

Proof. $\bigoplus \mathbf{r}^1 = 0$ while $\bigoplus \mathbf{r}^2 = 1$ under observation. □

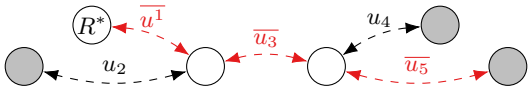
Again, the positivity can be guaranteed by randomly flipping R.

We investigated the non-identifiability of an *arbitrary* query $P_{\mathbf{x}}(\mathbf{r})$ given *arbitrary* experiments $\mathbb{Z}$ in an *arbitrary* thicket structure $\mathcal{J}$ rooted on $\mathbf{R}$, based on its unique hedgelet decomposition and the relationships among the hedgelets, query, and available experiments with a novel

[5] $2^{\mathbf{X}}$ represents a power set of $\mathbf{X}$, i.e., all subsets of $\mathbf{X}$ including an empty set.



(a) $\mathbf{u}'^1 = \{u_1, u_2, u_3, u_4, u_5\}$ for $\mathcal{M}_1$



(b) $\mathbf{u}'^2 = \{\overline{u_1}, u_2, \overline{u_3}, u_4, \overline{u_5}\}$ for $\mathcal{M}_2$

Figure 5: The bottom of a thicket is shown for two models where 'blacked-out' vertices are colored in gray. Instances $\mathbf{u}'^1$ and $\mathbf{u}'^2$ yield the same $\mathbf{r}$ given $\mathbf{u}^1 \setminus \mathbf{U}' = \mathbf{u}^2 \setminus \mathbf{U}'$.

parametrization. The next section extends this result to a general characterization of non-g-identifiability.

3.2 A GRAPHICAL CONDITION FOR NON-G-IDENTIFIABILITY

The following result ties the presence of a thicket to the non-g-identifiability of a causal effect.

Theorem 1. *If there exists a thicket $\mathcal{J}$ for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ with respect to $\mathbb{Z}$, then, $P_{\mathbf{x}}(\mathbf{y})$ is not g-identifiable in $\mathcal{G}$.*

Proof. Let $\mathbf{R}$ be the root set of $\mathcal{J}$. We construct two models for $\mathcal{J}$ demonstrating non-g-identifiability.

(Case: non-degenerate $\mathcal{J}$) Let $\mathbf{X}' = \mathbf{X} \cap \mathcal{J}$. Each hedgelet in $\mathcal{J}$ intersects with $\mathbf{X}$ ensuring that every hedgelet is intervened on, given $do(\mathbf{x})$. Hence, $P_{\mathbf{x}'}(\mathbf{r})$ is not g-identifiable following Corollary 1 and Lemma 5. We can map to $P_{\mathbf{x}'}(\mathbf{y}')$ where $\mathbf{Y}' \subseteq \mathbf{Y} \cap De(\mathbf{R})_{\mathcal{G}_{\overline{\mathbf{X}}}}$ (see Lemma 9 in Appendix A.2).

(Case: degenerate $\mathcal{J}$) By the definition, $\mathbf{X} \cap \mathcal{J} = \emptyset$, $P(\mathbf{r})$ is not g-identifiable following Lemmas 6 and 7. In the same way as the above, we map the result to $P(\mathbf{y}')$. □

Armed with a characterization of when the identification of a causal effect is not possible from the given input, in the next section, we provide a procedure that yields an expression for the target effect in terms of the input in all cases where there exists such mapping.

4 A COMPLETE ALGORITHM FOR G-IDENTIFIABILITY

Building on the graphical characterization of non-gID, in this section, we develop an algorithm for g-identifiability called GID (Alg. 1). For a given causal query, GID determines whether it's g-identifiable, and if so, it outputs a formula expressing the target effect in terms of the available distributions. The design of GID shares the same principles established by previous identifiability algorithms (e.g., IDENTIFY [Tian, 2002], ID [Shpitser and Pearl, 2006], zID [Bareinboim and Pearl, 2012]). Still, in our case, the identification process is decomposed into two parts: pre- and post-activation of an available distribution, where SUB-ID takes care of a (classic) identification task for each factored query with a fixed distribution treated as observational, relative to the call-specific graph.

The algorithm takes a query $P_{\mathbf{x}}(\mathbf{y})$, the causal graph $\mathcal{G}$, and available experiments $\mathbb{Z}$ as inputs. Without loss of generality, we assume that $\mathbf{Y} \neq \emptyset$ since otherwise the result is trivially 1. During the process, the query and the causal graph may be transformed when necessary, and broken down into smaller sub-problems. Accordingly, the parameters $\mathbf{y}$, $\mathbf{x}$, and $\mathcal{G}$ are local to each call, while $\mathbb{Z}$ is preserved throughout recursive calls.

The given $\mathcal{G}$ is modified only through Line 3, since experiments on variables that are not ancestors of $\mathbf{Y}$ have no effect on it, we only need to pay attention to experiments on ancestors of $\mathbf{Y}$. Line 2 utilizes any matching experiment whenever possible. As mentioned above, $\mathbf{Z}$ outside the current scope can be of any value. Lines 4 and 6 modify and factorize the given query, respectively. In Line 7, given a factorized query, the algorithm examines whether an available distribution might be helpful to estimate it, and delegates the identification to a subroutine, SUB-ID, which works as ID except that it uses one of the available distributions not necessarily $P(\mathbf{v})$.

The algorithm runs in $O(mn^4)$ where $m = |\mathbb{Z}|$ and $n = |\mathbf{V}|$. GID can be called subsequently $O(n)$ times due to the factorization at Line 6 where each GID may call SUB-ID up to m times, thus, totaling $O(nm)$ SUB-ID invocations, which may trigger, recursively, n times. Given that set or graphical operations take $O(n^2)$, it runs in $O(mn^4)$.

As for a running example, we revisit Fig. 1a where the query is $P_{\mathbf{x}}(y)$ and $\mathbb{Z} = \{\{X_1\}, \{X_2\}\}$. All variables are ancestors of Y, and no variable needs to be added as an intervention (Lines 3 and 4). Since W and Y are not confounded in $\mathcal{G} \setminus \mathbf{X}$, the query is factorized into $P_{\mathbf{x},w}(y)$ and $P_{\mathbf{x},y}(w)$ (Line 6). The first query $P_{\mathbf{x},w}(y)$ will pass through all conditions and SUB-ID will be called for experiments on both $\{X_1\}$ and $\{X_2\}$. Focusing on the latter, with $Q = P_{x_2}$ in $\mathcal{G} \setminus \{X_2\}$, $Q_{x_1,w}(y)$ will be identified as $Q(y|x_1, w) = P_{x_2}(y|x_1, w)$, which can be simplified into $P_{x_2}(y|w)$. GID will try both experiments for the second query $P_{\mathbf{x},y}(w)$. With experiment on $\{X_1\}$, $Q = P_{x_1}$, $P_{\mathbf{x},y}(w) = Q_{x_2,y}(w)$ will be refined to $Q(w)$ (Line 12), and will be trivially identified as $Q(w) = P_{x_1}(w)$ (Line 11). Therefore, the final formula becomes $P_{\mathbf{x}}(y) = \sum_w P_{x_2}(y|w)P_{x_1}(w)$.

Lemma 8. *Whenever* SUB-ID *returns an expression for $Q_{\mathbf{x}}(\mathbf{y})$, it is correct.*

Algorithm 1 GID: a complete identification algorithm for g-identifiability

1: **function** GID($\mathbf{y}, \mathbf{x}, \mathcal{G}, \mathbb{Z}$)
Input: $\mathbf{y}$, $\mathbf{x}$: value assignments, $\mathcal{G}$: causal diagram, $\mathbb{Z}$: a collection of available experiments
Output: an estimand computing $P_{\mathbf{x}}(\mathbf{y})$ with $\{P_{\mathbf{z}}\}_{\mathbf{Z}\in\mathbb{Z},\mathbf{z}\in\mathfrak{x}_{\mathbf{Z}}}$.
2: **if** $\exists_{\mathbf{Z}\in\mathbb{Z}}\mathbf{X}=\mathbf{Z}\cap\mathbf{V}$ **then return** $P_{\mathbf{z}\setminus\mathbf{V},\mathbf{x}}(\mathbf{y})$
3: **if** $\mathbf{V}\neq An(\mathbf{Y})_{\mathcal{G}}$ **then return** GID$(\mathbf{y},\mathbf{x}\cap An(\mathbf{Y})_{\mathcal{G}},\mathcal{G}[An(\mathbf{Y})_{\mathcal{G}}],\mathbb{Z})$
4: **if** $(\mathbf{W}\leftarrow(\mathbf{V}\setminus\mathbf{X})\setminus An(\mathbf{Y})_{\mathcal{G}_{\overline{\mathbf{X}}}})\neq\emptyset$ **then return** GID$(\mathbf{y},\mathbf{x}\cup\mathbf{w},\mathcal{G},\mathbb{Z})$
5: $\mathbb{S}\leftarrow\mathcal{C}(\mathcal{G}\setminus\mathbf{X})$
6: **if** $|\mathbb{S}|>1$ **then return** $\sum_{\mathbf{v}\setminus(\mathbf{y}\cup\mathbf{x})}\prod_{\mathbf{S}\in\mathbb{S}}$ GID$(\mathbf{s},\mathbf{v}\setminus\mathbf{s},\mathcal{G},\mathbb{Z})$
7: **for** $\mathbf{Z}\in\mathbb{Z}$ such that $\mathbf{Z}\cap\mathbf{V}\subseteq\mathbf{X}$ **do return** SUB-ID$(\mathbf{y},\mathbf{x}\setminus\mathbf{Z},P_{(\mathbf{z}\setminus\mathbf{V}),\mathbf{x}\cap\mathbf{Z}},\mathcal{G}\setminus(\mathbf{Z}\cap\mathbf{X}))$ **if** not NONE
8: **throw** FAIL
9: **function** SUB-ID($\mathbf{y}$, $\mathbf{x}$, Q, $\mathcal{G}$)
10: $\{\mathbf{S}\}\leftarrow\mathcal{C}(\mathcal{G}\setminus\mathbf{X})$
11: **if** $\mathbf{X}=\emptyset$ **then return** $\sum_{\mathbf{v}\setminus\mathbf{y}}Q(\mathbf{v})$
12: **if** $\mathbf{V}\neq An(\mathbf{Y})_{\mathcal{G}}$ **then return** SUB-ID$(\mathbf{y},\mathbf{x}\cap An(\mathbf{Y})_{\mathcal{G}},\sum_{\mathbf{v}\setminus An(\mathbf{Y})_{\mathcal{G}}}Q,\mathcal{G}[An(\mathbf{Y})_{\mathcal{G}}])$
13: **if** $\mathcal{C}(\mathcal{G})=\mathbf{V}$ **then return** NONE
14: **if** $\mathbf{S}\in\mathcal{C}(\mathcal{G})$ **then return** $\sum_{\mathbf{s}\setminus\mathbf{y}}\prod_{V_i\in\mathbf{Y}}Q(v_i|\mathbf{v}_\pi^{(i-1)})$.
15: **if** $\mathbf{S}\subsetneq\mathbf{S}'\in\mathcal{C}(\mathcal{G})$ **then**, **return** SUB-ID$(\mathbf{y},\mathbf{x}\cap\mathbf{S}',\prod_{V_i\in\mathbf{S}'}Q(V_i|\mathbf{V}_\pi^{(i-1)}\cap\mathbf{S}',\mathbf{v}_\pi^{(i-1)}\setminus\mathbf{S}'),\mathbf{S}')$

Proof. SUB-ID performs classic identifiability of $Q_{\mathbf{x}}(\mathbf{y})$ with Q. The SUB-ID is an excerpt of ID algorithm where unnecessary statements (related to Lines 4 and 6) are removed because its parameters $\mathbf{y}$, $\mathbf{x}$, and $\mathcal{G}$ throughout its procedure satisfy i) $(\mathbf{V}\setminus\mathbf{X})\setminus An(\mathbf{Y})_{\mathcal{G}_{\overline{\mathbf{X}}}}=\emptyset$, and ii) $\mathcal{G}\setminus\mathbf{X}$ forms a c-component. □

Theorem 2 (Soundness)**.** *Whenever* GID *returns an expression for* $P_{\mathbf{x}}(\mathbf{y})$*, it is correct.*

Proof. Let $\mathbf{x}$ and $\mathbf{y}$ be local to the arguments of GID. GID correctly transforms the given query $P_{\mathbf{x}}(\mathbf{y})$ together with $\mathcal{G}$, which is proved by [Shpitser and Pearl, 2006, Lemma 4–6]. The difference of GID compared to ID is i) returning an expression at Line 2, and ii) delegating identification with an available experiment at Line 7.

(i) Each experiment $\mathbf{Z}\in\mathbb{Z}$ outside the scope of $An(\mathbf{Y})_{\mathcal{G}}$ can be ignored by Rule 3 of do-calculus (Line 3). Then, $\mathbf{X}=\mathbf{Z}\cap\mathbf{V}$ implies that $P_{\mathbf{x}}(\mathbf{y})=P_{\mathbf{x},\mathbf{z}\setminus\mathbf{V}}(\mathbf{y})=P_{\mathbf{z}\cap\mathbf{V},\mathbf{z}\setminus\mathbf{V}}(\mathbf{y})=P_{\mathbf{z}}(\mathbf{y})$ with $\mathbf{z}$ consistent with $\mathbf{x}$.

(ii) First, the use of $P_{\mathbf{z}}$ is valid when $\mathbf{Z}\cap\mathbf{V}\subseteq\mathbf{X}$ since $P_{\mathbf{x}}(\mathbf{y})=P_{\mathbf{z},\mathbf{x}\setminus\mathbf{Z}}(\mathbf{y})$ where $\mathbf{z}$ is consistent with $\mathbf{x}$. Then, this is identifying $Q_{\mathbf{x}\setminus\mathbf{Z}}(\mathbf{y})$ with $Q=P_{\mathbf{z}}$, which is a classic identifiability instance assignable to SUB-ID.

Combined with Lemma 8, GID is sound. □

Theorem 3 (Completeness)**.** GID *is complete.*

Proof. We show that there exists a thicket for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ with respect to $\mathbb{Z}$ whenever GID fails (Line 8). Let the arguments of GID be $\mathbf{y}'$, $\mathbf{x}'$, $\mathcal{G}'$, and $\mathbb{Z}$ when it failed.

We first consider a case where $\mathbf{Z}\cap\mathbf{V}'\not\subseteq\mathbf{X}'$ for every $\mathbf{Z}\in\mathbb{Z}$. We construct a degenerate thicket $\mathcal{J}$ as an $\mathbf{R}$-rooted minimal c-component in $\mathcal{G}'[\mathbf{R}]$ where $\mathbf{R}=\mathbf{V}'\setminus\mathbf{X}'$. $\mathcal{J}$ is a valid thicket for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$ given $\mathbb{Z}$ because: (i) $\mathbf{R}\subseteq An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}}$ (Lines 3–6); (ii) $\mathcal{G}'[\mathbf{R}]$ is a c-component (Lines 5, 6); and $\mathbf{Z}\cap\mathbf{R}\neq\emptyset$ for every $\mathbf{Z}\in\mathbb{Z}$.

We now construct a non-degenerate thicket with hedges associated with the failed queries via SUB-ID. Consider a hedge for $P_{\mathbf{x}'\setminus\mathbf{Z}}(\mathbf{y}')$ in $\mathcal{G}'\setminus\mathbf{Z}$ for some $\mathbf{Z}\in\mathbb{Z}$ such that $\mathbf{Z}\cap\mathbf{V}'\subseteq\mathbf{X}'$. Replacing its bottom with $\mathbf{R}$ forming a minimal c-component, which is the same as the degenerate thicket above, results in a valid hedge for $P_{\mathbf{x}'\setminus\mathbf{Z}}(\mathbf{y}')$ in $\mathcal{G}'\setminus\mathbf{Z}$ since $\mathbf{R}=An(\mathbf{Y}')_{\mathcal{G}'_{\underline{\mathbf{X}'}}}$. Hence, a thicket formed by the union of the modified hedges will satisfy the characteristics of its root set as described in Def. 6.

We then show that each hedgelet of the hedges composing the thicket intersects with $\mathbf{X}$. We start by decomposing $\mathbf{V}(\mathcal{F}'')$ into three parts: $\mathbf{X}'_1=\mathbf{V}(\mathcal{F}'')\cap\mathbf{X}$; $\mathbf{X}'_2=\mathbf{V}(\mathcal{F}'')\cap\mathbf{W}$; and $\mathbf{X}'_3=\mathbf{V}(\mathcal{F}'')\setminus(\mathbf{X}\cup\mathbf{W})$ where $\mathbf{W}$ is the set of variables that was combined with $\mathbf{X}$ at Line 4, which occurs at most once. Then, there exists no directed edge from $\mathbf{X}'_2$ to $\mathbf{X}'_3$ (Line 4), and no bidirected edge between $\mathbf{X}'_3$ and $\mathbf{R}$.[6] For the sake of contradiction assume that $\mathbf{X}'_1=\emptyset$. The cross UC of the hedgelet should point towards $\mathbf{X}'_2$, which can only be connected to $\mathbf{R}$ via directed paths only through $\mathbf{X}'_1$ (Line 3). This contradicts the definition of hedgelet, which must be a forest. Consequently, the superimposition of the modified hedges is a

[6] Consider the first encounter with Line 5. If $|\mathbb{S}|=1$, then $\mathbf{X}'_3=\emptyset$. Otherwise if $|\mathbb{S}|>1$, $\mathbf{X}'_3$ corresponds to those variables in "$\mathcal{G}\setminus\mathbf{X}$" (that is, $\mathcal{G}[An(\mathbf{Y})_{\mathcal{G}}]\setminus((\mathbf{X}\cap An(\mathbf{Y})_{\mathcal{G}})\cup\mathbf{W})$ after Lines 3 and 4) but not connected to $\mathbf{Y}'=\mathbf{R}$ via bidirected edges.

thicket formed for $P_{\mathbf{x}}(\mathbf{y})$.

Whenever GID fails, there exists a thicket for $P_{\mathbf{x}}(\mathbf{y})$ with respect to $\mathbb{Z}$. Hence, the result follows from Thm. 1. □

Corollary 2 (Do-calculus Completeness)**.** *The rules of do-calculus together with standard probability manipulations are complete for determining g-identifiability of all causal effects of the form* $P_{\mathbf{x}}(\mathbf{y})$.

Proof. GID and SUB-ID reuse steps employed in ID except for Lines 2 and 7, which correspond to Rule 3 of do-calculus. Since all steps in ID can be mapped to applications of do-calculus and probability axioms ([Shpitser and Pearl, 2006, Thm. 7]), the result follows. □

5 CONCLUSIONS

We studied the identification of causal effects from arbitrary combinations of observational and experimental distributions, which generalizes two canonical settings in which no interventions [Pearl, 1995] and all interventions over a set of variables [Bareinboim and Pearl, 2012] are available. This problem has been called g-identifiability, or gID for short. We developed a general algorithm for solving gID and proved its completeness. We introduced new machinery to better understand and more precisely characterize non-trivial forbidden structures that preclude gID, which can be seen as instances of hedgelets and thickets. Finally, as a corollary of these results, we proved that do-calculus is complete for the task of g-identifiability.

Acknowledgements

This research is supported in parts by grants from NSF IIS-1704352, IIS-1750807 (CAREER), IBM Research, and Adobe Research.

References

E. Bareinboim and J. Pearl. Causal inference by surrogate experiments: z-identifiability. In N. de Freitas and K. Murphy, editors, *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pages 113–120, Corvallis, OR, 2012. AUAI Press.

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

E. Ferrannini and W. C. Cushman. Diabetes and hypertension: the bad companions. *The Lancet*, 380(9841): 601–610, 2012.

R. Fisher. *The Design of Experiments*. Oliver and Boyd, Edinburgh, 6th edition, 1951.

D. Galles and J. Pearl. Testing identifiability of causal effects. In P. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI 1995)*, pages 185–195. Morgan Kaufmann, San Francisco, 1995.

Y. Huang and M. Valtorta. Identifiability in causal bayesian networks: A sound and complete algorithm. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006)*, pages 1149–1156. AAAI Press, Menlo Park, CA, 2006.

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

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000.

J. Pearl and D. Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

J. Pearl and J. Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In P. Besnard and S. Hanks, editors, *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence (UAI 1995)*, pages 444–453. Morgan Kaufmann, San Francisco, 1995.

I. Shpitser and J. Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of The Twenty-First National Conference on Artificial Intelligence*, pages 1219–1226. AAAI Press, 2006.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition, 2001.

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

J. Tian and J. Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence (AAAI 2002)*, pages 567–573, Menlo Park, CA, 2002. AAAI Press/The MIT Press.

A A SUPPLEMENTARY MATERIAL TO GENERAL IDENTIFIABILITY WITH ARBITRARY SURROGATE EXPERIMENTS

A.1 DERIVATION

We derive an expression for Fig. 1a as follows

$$\begin{aligned} P_{x_1,x_2}(y) &= \sum_w P_{x_1,x_2}(y,w) \\ &= \sum_w P_{w,x_1,x_2}(y) P_{y,x_1,x_2}(w) \\ &= \sum_w P_{x_2,w,x_1}(y) P_{x_1}(w) \\ &= \sum_w P_{x_2,w}(y) P_{x_1}(w) \\ &= \sum_w P_{x_2}(y|w) P_{x_1}(w) \end{aligned}$$

The query $P_{x_1,x_2}(y)$ is rewritten as $\sum_w P_{x_1,x_2}(w,y)$ and factorized $\sum_w P_{w,x_1,x_2}(y) P_{y,x_1,x_2}(w)$ based on c-component form. For the first term, by Rule 3 and 2 of do-calculus, $P_{x_2,w,x_1}(y) = P_{x_2,w}(y) = P_{x_2}(y|w)$. For the second term, $P_{y,x_1,x_2}(w) = P_{x_1}(w)$ by Rule 3 of do-calculus. Hence, $P_{x_1,x_2}(y) = \sum_w P_{x_2}(y|w) P_{x_1}(w)$.

For Fig. 2a, it only requires a single application of Rule 3 of do-calculus. Simply put, intervened variables outside the ancestors of an outcome variable have no effect on the outcome variable. Hence, $P_{x_1,x_2}(y_1) = P_{x_1}(y_1)$ and $P_{x_1,x_2}(y_2) = P_{x_2}(y_2)$.

A.2 NON-IDENTIFIABILITY MAPPING

Lemma 9. *Let* $\mathbf{X}$, $\mathbf{Y}$ *be disjoint sets of variables in* $\mathcal{G}$. *Let* $\mathcal{J}$ *be a nonempty subgraph of* $\mathcal{G}$ *with root set* $\mathbf{R}$, *where* $\mathbf{R} \subseteq An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}}$. *Let* $\mathcal{M}_1$ *and* $\mathcal{M}_2$, *which are compatible with* $\mathcal{J}$, *satisfy*

$$\sum_{\mathbf{r}|\bigoplus \mathbf{r}=1} P^1_{\mathbf{x}\cap\mathcal{J}}(\mathbf{r}) \neq \sum_{\mathbf{r}|\bigoplus \mathbf{r}=1} P^2_{\mathbf{x}\cap\mathcal{J}}(\mathbf{r})$$

for some $\mathbf{x}$ *where all variables in* $\mathbf{R}$ *are binary. Then, there are two models* $\mathcal{M}'_1$ *and* $\mathcal{M}'_2$ *compatible with* $\mathcal{G}$ *such that* $P'^1_{\mathbf{x}}(\mathbf{y}) \neq P'^2_{\mathbf{x}}(\mathbf{y})$ *for some* $\mathbf{y}$.

Proof. Similar results appear in identifiability literature, e.g., [Shpitser and Pearl, 2006, Thm. 4]. We first employ their strategies in the proof and discuss some theoretical oversight. By the condition $An(\mathbf{Y})_{\mathcal{G}_{\underline{\mathbf{X}}}}$, there exist directed downward paths from $\mathbf{R}$ to $\mathbf{Y}$ where no $\mathbf{X}$ appear in-between and each node has at most one child. That is, one can parametrize each node (which is binary) in the paths as an exclusive-or of its observable parents. Then, the discrepancy in bit-parity for $\mathbf{R}$ in $\mathcal{M}_1$ and $\mathcal{M}_2$ will also be happened at $\mathbf{Y}$ in $\mathcal{M}'_1$ and $\mathcal{M}'_2$ under $do(\mathbf{x})$ (n.b. values of $\mathbf{x}$ outside $\mathcal{J}$ are irrelevant to $\mathbf{Y}$).

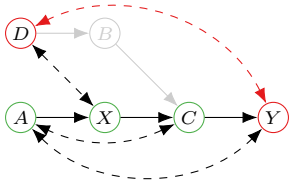


Figure 6: A causal graph $\mathcal{G}$ with a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ for $P_x(y)$ where $\mathcal{F} = \mathcal{G} \setminus \{B\}$ with $\mathcal{F}'$ shown in red and variables in $\mathcal{F}''$ shown in green. The bit parity of D and Y should be mapped to Y through B and C where C is at the top of the hedge.

A possible oversight is that the downward paths might cross $\mathcal{J}$ without passing $\mathbf{X}$ (see Fig. 6 for an example). The remedy is simple. For nodes appearing in the directed downward paths from $\mathbf{R}$ to $\mathbf{Y}$, we can assign an additional bit to pass bit parity information from $\mathbf{R}$ to $\mathbf{Y}$. Further, given a probability distribution $P_{\mathbf{w}}(\mathbf{z})$ on which $\mathcal{M}_1$ and $\mathcal{M}_2$ agree ($\mathbf{W}, \mathbf{Z} \subseteq \mathbf{V}(\mathcal{J})$), $\mathcal{M}'_1$ and $\mathcal{M}'_2$ will also agree on $P_{\mathbf{w}\cup\mathbf{b}}(\mathbf{z})$ for any $\mathbf{b} \in \mathfrak{X}_{\mathbf{B}}$ where $\mathbf{B} \subseteq \mathbf{V}(\mathcal{G}) \setminus \mathbf{V}(\mathcal{J})$ for two reasons: Variables outside the paths from $\mathbf{R}$ to $\mathbf{Y}$ and $\mathcal{J}$ are ignored. Both models $\mathcal{M}'_1$ and $\mathcal{M}'_2$ behave the same for nodes between $\mathbf{R}$ to $\mathbf{Y}$. □