General Identifiability with Arbitrary Surrogate Experiments

Sanghack Lee and **Juan D. Correa** and **Elias Bareinboim**
Causal Artificial Intelligence Laboratory,
Department of Computer Science,
Purdue University, West Lafayette, IN 47907
{lee2995,correagr,eb}@purdue.edu

Abstract

We study the problem of identification of causal effects from experimental data corresponding to a collection of regimes where different sets of variables have been intervened on. In particular, we relax one key assumption of this setting, namely, that all combinations of experiments are available, or the power set of the experimental set $\mathbf{Z}$, as required in z-identifiability. One implication of this relaxation is that the observational distribution itself doesn't need to be available, which is expected in both identifiability and z-identifiability solutions. In this paper, we call this relaxation of the problem g-identifiability (gID, for short), and provide a general solution for it. Specifically, we introduce a novel 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 the same is sound and complete in gID settings, i.e., failure of returning an effect by the algorithm implies that such an 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-identification.

1 INTRODUCTION

One of the main goals in data-driven sciences is to infer cause and effect relationships from a combination of observational and experimental data with 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, Pearl and Mackenzie, 2018].

In one line of investigation, this task is materialized 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 unambiguously established from a joint probability distribution P over the observed variables $\mathbf{V}$, and substantive knowledge encoded in the form of a causal diagram $\mathcal{G}$. This is known as the problem of *identification*, 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 is conceivable that in some real-world applications the quantity $P_{\mathbf{x}}(\mathbf{y})$ is not identifiable from the observational data and the causal graph.

On an alternative thread, causal effects, $P_{\mathbf{x}}(\mathbf{y})$, can be obtained by directly carrying out controlled experiments [Fisher, 1951]. For instance, this procedure appears under the rubric of *Randomized Controlled Trials* (RCTs) in the biomedical sciences, and, in fact, there is a vast amount of resources spent every year in supporting controlled, large-scale, and systematic experimentation by the FDA, the NIH, among others. This method is also leveraged in the context of AI, for example, when autonomous agents are deployed in an environment and have 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, 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.

These two approaches can be seen as extremes in a spectrum of possible research designs, which can be combined to solve non-trivial instances. Such approach 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 unambiguously computed from the combination of the observational distribution $P(\mathbf{V})$ and the experimental distributions $P_{\mathbf{z}'}(\mathbf{V})$, for all $\mathbf{Z}' \subseteq \mathbf{Z}$. We make two critical observations regarding the assumptions underlying this setting. First, that experiments over all subsets of $\mathbf{Z}$ are possibly available (e.g., if $\mathbf{Z}=\{Z_1, Z_2\}$, this means $\{\{\}, \{Z_1\}, \{Z_2\}, \{Z_1, Z_2\}\}$). In many practical situations, however, it may be unfeasible to intervene simultaneously on some set of variables, while other set may not be intervened on separately (see examples below). Second, this constraint includes the observational distribution (null intervention), which, while available in many settings, may be hard to measure in others. For instance, when an agent has to act in a reinforcement learning environment where no natural regime exists. For concreteness, we discuss in the sequel examples where these assumptions are shown to be somewhat 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 may be the case, following the current beliefs, that diabetes and hypertension do not affect each other (no direct link between them), it's common for patients with diabetes type 2 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 treatment for both conditions, simultaneously, on the risk of developing cardiovascular diseases, i.e., $P_{x_1,x_2}(y)$. 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. Note that this input does not match what is expected by zID since the combinations of all experiments are required, 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 and 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 and d (see Appendix A.1). □

Example 2. (Multivariate testing) The causal graphs in Fig. 2 represent simplified models of a large 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. The various teams perform experiments, called multivariate testing, on a diverse set of features at the same time and track the different outcome variables. In words, this means that the joint experimental distribution $P_{x_1,x_2}(\mathbf{V})$ is available. The goal is to identify the individual effects of the 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. 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 $P_{x_2}(y_2) = P_{x_1,x_2}(y_2)$ in Fig. 2a; on the other hand, these effects are not identifiable in Fig. 2b and c (for details, see Appendix A.1). □

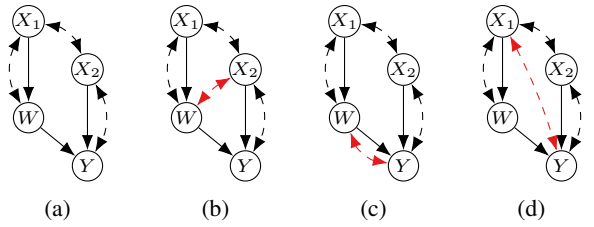


Figure 1: $P_{x_1,x_2}(y)$ can be identified from $P_{x_1}(\mathbf{V})$ and $P_{x_2}(\mathbf{V})$ in (a) and (b), but not in (c) and (d). Differences among the causal diagrams are highlighted in red.

In this paper, our goal is to explicate the conditions under which inferences of this type are allowed or not. More generally, we study the problem of identification of experimental distributions from an arbitrary combination of experiments and a causal model. Specifically, our contributions are as follows: 1. We prove a necessary and sufficient graphical condition for g-identifiability. This result follows from two new graphical structures called *hedgelet* and *thicket*, which are flexible and general building blocks to understand causal effect identification problems (Sec. 3); 2. Leveraging these results, we develop a sound and complete algorithm that returns any expression derivable from an arbitrary set of experiments. As a corollary, we prove that do-calculus is complete for g-identification (Sec. 4).

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 domain of every variable is finite.

[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.

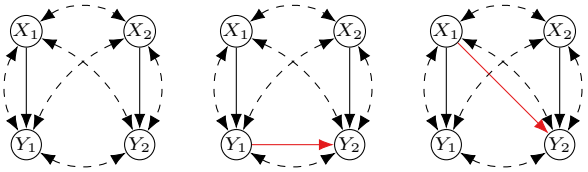


Figure 2: Given experiments available on $\{X_1, X_2\}$, both $P_{x_1}(y_1)$ and $P_{x_2}(y_2)$ are identifiable in (a), but $P_{x_2}(y_2)$ is not identifiable in (b) and (c). In all cases, $P_{x_1}(y_2)$ and $P_{x_2}(y_1)$ are not identifiable.

Our analysis heavily relies on graphs, 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., $\mathcal{G}[\mathbf{W}]$ consists of $\mathbf{W}$ and edges among them. 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 $pa(\mathbf{X})_{\mathcal{G}} := \bigcup_{X \in \mathbf{X}} pa(X)_{\mathcal{G}}$. We similarly define ch, de, and an. Written as Pa, Ch, De, and An include the argument 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 the ordering $\boldsymbol{\pi}$. 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 for causal analysis. 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 outside the model; $\mathbf{V}$ 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 $\mathbf{U}$. A 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 $\mathbf{U}^i \cap \mathbf{U}^j \neq \emptyset$. An intervention, defined via an operator called $do(\mathbf{x})$, is setting the intervened variables $\mathbf{X}$ to a specific value $\mathbf{x} \in \mathfrak{X}_{\mathbf{X}}$. Given a model $\mathcal{M}$, intervention $do(\mathbf{x})$ leads to a new model (also called, 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, since variables in $\mathbf{X}$ are not affected by any variable in the system.

2.1 A NEW LOOK AT HEDGE

We revisit a graphical structure called *hedge* [Shpitser and Pearl, 2006]. It was shown that the presence of a hedge indicates the inability to uniquely identify a target causal effect from an observational distribution. This understanding will be the starting point of our treatment of the more general g-identifiability problem.

Definition 1 (C-component [Tian and Pearl, 2002])**.** 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 (confounded component).

Given a semi-Markovian graph $\mathcal{G}$ over a set of variables $\mathbf{V}$, there exists a unique partition into subgraphs, such that each one is a 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 $i \neq j$. Based on this notion, we can build towards the definition of hedge with the following adopted 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 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 $\mathcal{F}' \subseteq \mathcal{F}$. Given $\mathbf{X}$ and $\mathbf{Y}$ two disjoint sets of variables in $\mathcal{G}$, if $\mathbf{R} \subseteq An(\mathbf{Y})_{\mathcal{G}_{\overline{\mathbf{X}}}}$, $\mathcal{F} \cap \mathbf{X} \neq \emptyset$, and $\mathcal{F}' \cap \mathbf{X} = \emptyset$, then a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ is said to be formed for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$.

The existence of a hedge for $P_{\mathbf{x}}(\mathbf{y})$ in $\mathcal{G}$, that is, $\mathcal{G}$ having the hedge as a subgraph implies that $P_{\mathbf{x}}(\mathbf{y})$ is not uniquely identifiable from $P(\mathbf{V})$ [Shpitser and Pearl, 2006]. The original definition of hedge intertwines a graphical structure with identifiability. This precludes the use of the underlying graphical structure itself without involving identifiability of $P_{\mathbf{x}}(\mathbf{y})$. Hence, we separate the definition of hedge itself from its use in the non-identifiability of a query (i.e., a hedge is *formed for* the query).

A hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ can be dissected into two sections: the 'top', denoted by $\mathcal{F}'' = \mathcal{F} \setminus \mathbf{V}(\mathcal{F}')$, and the 'bottom', which is $\mathcal{F}'$. When the top is empty, we call the hedge *degenerate*.

3 G-IDENTIFIABILITY

In contrast to the problems of identifiability and z-identifiability previously discussed, we propose a new

task that generalizes the aforementioned pair 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 distributions $\{P(\mathbf{V} \mid do(\mathbf{z}))\}_{\mathbf{Z}\in\mathbb{Z},\mathbf{z}\in\mathfrak{X}_{\mathbf{Z}}}$ in any causal model which induces $\mathcal{G}$.

A traditional assumption made in previous identification treatments is that a probability distribution describing the natural state of the system is available, that is, $P(\mathbf{V})$. Notice that the set defined above does not necessarily contain such distribution unless we explicitly include $\emptyset$ in $\mathbb{Z}$. Based on the definition of g-identifiability, we can formulate the following.

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 $\mathcal{G}$. □

This statement formally characterizes the situation that witnesses the non-g-identifiability of a certain instance, but does not provide an insight on how to determine if such pair of models exists, or how to construct them when a given instance is not g-identifiable. If not ambiguous, we omit g- prefix 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. 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.

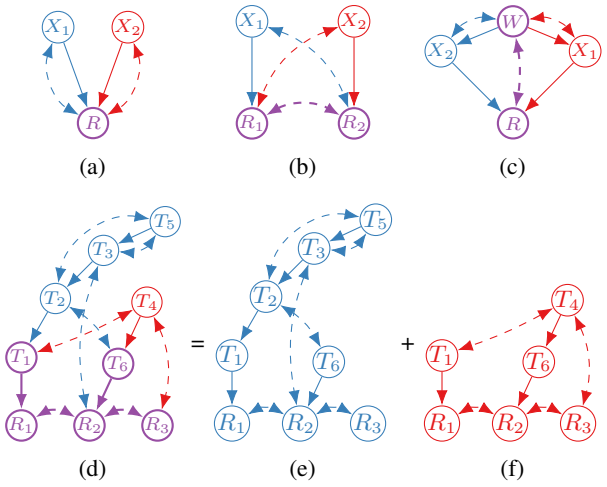


Figure 3: Hedgelet decomposition of hedges and a thicket (color coded in blue and red with purple for shared elements). Each of (a) and (b) is a hedge formed for $P_{\mathbf{r}}(\mathbf{x})$ or a thicket with respect to $\mathbb{Z} = \{\{X_1\}, \{X_2\}\}$ while (c) is not a hedge but a thicket. The hedge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

Analogous to a hedge, a hedgelet $\mathcal{H}$ has a top section $\mathcal{H}''$ and a bottom section $\mathcal{H}'$. The UCs between the top and bottom parts are called 'crossing UCs', and are denote by $\mathbf{U}^{\times}_{\mathcal{H}}$. And the variables transmitting their values through the crossing directed edges, will be called 'frontiers' and denoted by $\mathbf{V}^{\downarrow}_{\mathcal{H}}$.

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 $\mathcal{H}'$ and $\mathcal{H}''$ by definition.[2] We denote by $U^{\times}_{\mathcal{H}}$, the only crossing UC of a non-degenerate hedgelet $\mathcal{H}$, i.e., $\mathbf{U}^{\times}_{\mathcal{H}} = \{U^{\times}_{\mathcal{H}}\}$.

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 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,

[2] Since directed edges of $\mathcal{F}$ form a forest with all its roots in $\mathcal{F}'$, there must exists 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).

$\mathcal{F}[De(\mathbf{W}_1)_{\mathcal{F}}]$ 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., $\mathbf{V}^{\downarrow}_{\mathcal{H}_1} = \mathbf{V}^{\downarrow}_{\mathcal{H}_2} = \{T_1, T_6\}$).

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 some combination of the following queries and experimental specifications — $P_b(\mathbf{r})$, $P_{a,d}(\mathbf{r})$, $P_{e,f,g}(\mathbf{r})$, or $P_{c,f}(\mathbf{r})$, and $\{\{A, E\}, \{D, F, G\}\}$ or $\{\{A\}, \{D\}\}$.

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 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 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 $\mathcal{J}$.[3] 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.

We denote by $\mathbf{U}'$ and $\mathbf{U}''$ the UCs among $\mathcal{J}[\mathbf{R}]$ and $\mathcal{J} \setminus \mathbf{R}$, respectively. Let $U \in \mathbf{U}'$ be a binary variable with $\mathfrak{x}_U = \{0, 1\}$, and $P(\mathbf{u}') = 2^{-|\mathbf{U}'|}$. Let every $U \in \mathbf{U}\setminus\mathbf{U}'$ be a k-bit variable where k is the number of hedgelets wherein U appears, and $P(u) = 2^{-k}$. For both models, every $T \in \mathbf{T}$ is a ℓ-bit integer where $\ell = |\mathbb{H}(T)|$, so that each bit is responsible for its corresponding hedgelet in $\mathbb{H}(T)$. We use $\wedge$, $\bigoplus$, and $\overline{v}$ to denote *and*, *exclusive-or*, and *bitwise-complement* operation, respectively.

For each hedgelet $\mathcal{H} \in \mathbb{H}(T)$, define

$$t_{\mathcal{H}} \leftarrow \bigoplus\nolimits_{V\in pa(T)_{\mathcal{H}}} v_{\mathcal{H}} \oplus \bigoplus\nolimits_{U\in\mathbf{U}^T_{\mathcal{H}}} u_{\mathcal{H}} \qquad (1)$$

where $\mathbf{U}^T_{\mathcal{H}}$ is the set of UCs pointing towards T in $\mathcal{H}$, and $v_{\mathcal{H}}$ (or $u_{\mathcal{H}}$) is the bit of v (or u) corresponding to $\mathcal{H}$. Then t is defined as an integer made of bits $\langle t_{\mathcal{H}}\rangle_{\mathcal{H}\in\mathbb{H}(T)}$.[4]

Consider $R \in \mathbf{R}$, then let $\mathbf{U}'^R$ be the UCs connected to R in $\mathcal{J}[\mathbf{R}]$, and let $\mathbf{U}^R$ be those connected to R in $\mathcal{H}$. 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.[5]

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

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

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

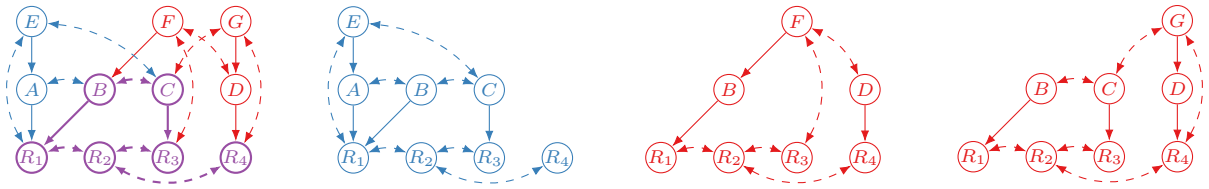
where $\mathbf{1}_{\bar{u}=0}$ is 1 if every bit of u is 1, and 0 otherwise.

Now, we characterize this parametrization with respect to the distributions these two models generate.

[3] There may be identical hedgelets coming from different hedges.

[4] The order of the bits does not matter.

[5] 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.

Lemma 2. *Let $\mathcal{H}$ be a hedgelet in $\mathbb{H}$ and U be $U^{\times}_{\mathcal{H}}$, then the above parametrization satisfies $\bigoplus_{T\in\mathbf{V}^{\downarrow}_{\mathcal{H}}} t_{\mathcal{H}} = u_{\mathcal{H}}$.*

Proof. Let $\langle \mathcal{F}, \mathcal{F}' \rangle$ be the hedge associated with $\mathcal{H}$, and let $\mathbf{W}\in\mathcal{C}(\mathcal{F}'')$ such that $\mathcal{H}=\mathcal{F}(\mathbf{W})$. $\mathcal{H}''$ is a $\mathbf{V}^{\downarrow}_{\mathcal{H}}$-rooted forest where each $T \in \mathbf{V}^{\downarrow}_{\mathcal{H}}$ is a root of a tree in the forest. By the parametrization, bits of $\mathbf{W}$ with respect to $\mathcal{H}$ carry the bit-parity of preceding UCs in $\mathcal{H}$. Due to the forestness of the directed edges in $\mathcal{H}''$, taking the XOR of all $\mathbf{V}^{\downarrow}_{\mathcal{H}}$ is equivalent to computing the XOR of all unobservable parents of variables in $\mathbf{W}$. Since each one of such UCs is a parent of two variables in $\mathbf{W}$, except for $U^{\times}_{\mathcal{H}}$, all but $U^{\times}_{\mathcal{H}}$ are counted twice. Due to the nature of XOR, repeated values cancel out and all that is left is $U^{\times}_{\mathcal{H}}$. □

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 $\mathbf{W} = \mathbf{V}^{\downarrow}_{\mathcal{H}}$ and $U = U^{\times}_{\mathcal{H}}$ be the frontiers and the crossing UC of $\mathcal{H}$, respectively. If $u_{\mathcal{H}} = 0$, then the $R \in \mathbf{R}$ pointed by U in $\mathcal{J}$, will be 0 since $U \in \mathbf{U}^{R} \setminus \mathbf{U}'^{R}$. Otherwise $u_{\mathcal{H}} = 1$, and by Lemma 2 we have $\bigoplus \mathbf{w}_{\mathcal{H}} = 1$. Therefore, there must be at least one W where $w_{\mathcal{H}} = 1$. As a consequence, $\overline{w_{\mathcal{H}}} = 0$ ensures that $r = 0$ for $R \in \mathbf{R}$ pointed by W. □

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^*.[6] 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
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 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 $P_{\mathbf{t}'=0}(\mathbf{r} = 0)$.*

Proof. By the condition, every hedgelet is intervened. We first show that there exists a non-zero probability for every $R \in \mathbf{R}$ being not blacked-out. In other words, there exists an instantiation of $\mathbf{u}$ such that the term in parentheses of Eq. (2) is 1 for every $R \in \mathbf{R}$. Let $\mathbf{u}^{\times} = \langle u_{\mathcal{H}} \mid U = U^{\times}_{\mathcal{H}} \rangle_{\mathcal{H}\in\mathbb{H}}$, that is, the concatenation of all bits from the crossing UCs of the hedgelets in $\mathbb{H}$. Let $\mathbf{u}^{\times} = 1$ (i.e., every bit is 1) and $\mathbf{u}'' = 0$. Let $\mathbf{W}$ be the frontiers of the hedge.

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

If $\mathbf{w} = 0$, then, together with $\mathbf{u}^{\times} = 1$ and $\mathbf{u}'' = 0$, all switches will be turned on. Otherwise, there exists $W \in \mathbf{W}$ such that $w \neq 0$. In other words, there exists a hedgelet $\mathcal{H}$ such that $w_{\mathcal{H}} = 1$. Since $\mathbf{u}'' = 0$ and $\mathbf{t}' = 0$, the value $w_{\mathcal{H}}$ being 1 is due to $u^{\times}_{\mathcal{H}} = 1$. Let S be the variable in $\mathcal{H}''$ pointed with the bidirected edge corresponding to $U^{\times}_{\mathcal{H}}$, the only element in $\mathbf{U}^{\times}_{\mathcal{H}}$. Find a bidirected path in $\mathcal{H}$ from S to $An(\mathbf{T}' \cap \mathcal{H})_{\mathcal{H}}$ (those variables that, under the intervention $\mathbf{T}'$, cannot influence frontiers of $\mathcal{H}$ when UCs connecting to them are changed). It must exist otherwise it conflicts the definition of hedgelet, which consists of a c-component and its descendants. Negating the UCs in the path only affects $\mathcal{H}$-specific bits of variables, and yields $w_{\mathcal{H}} = 0$. Since UCs on the top are partitioned by hedgelets, we can independently apply the above procedure for every $W \in \mathbf{W}$ such that $w \neq 0$ so as to obtain $\mathbf{u}''$ that leads to $P(\mathbf{w} = 0, \mathbf{u}^{\times} = 1) > 0$ for both models.

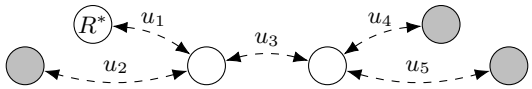
In such an event where there is no blacked-out node, $\mathcal{M}_1$ and $\mathcal{M}_2$ yield $\mathbf{r}^1$ and $\mathbf{r}^2$ such that $\bigoplus \mathbf{r}^1 = 0$ and $\bigoplus \mathbf{r}^2 = 1$. Combined with the fact that two models yield the same probability distributions when there exists a blacked-out node (partly from Lemma 4), two models disagree on $P_{\mathbf{t}'}(\mathbf{r})$ resulting $P^1_{\mathbf{t}'=0}(\mathbf{r} = 0) > P^2_{\mathbf{t}'=0}(\mathbf{r} = 0)$. □

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[7] $\mathbf{V}' \in 2^{\{T_2,T_3,T_5\} \cup \mathbf{R}} \cup 2^{\{T_4\} \cup \mathbf{R}}$ and they disagree on $P_{\mathbf{t}'}(\mathbf{r})$ for $\mathbf{T}' \subset 2^{\mathbf{T}}$, except for the aforementioned sets.

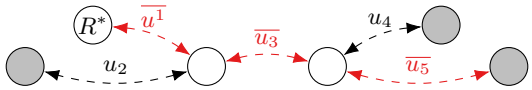
Non-identifiability of an Observational Probability for a Degenerate Thicket In g-identifiability, we also seek whether an observational probability, which was trivially identifiable in previous literature, can be uniquely determined by available experimental data. In this section, our focus is a degenerate 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

[7] $2^{\mathbf{X}}$ 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: $\mathcal{F}'$ with blacked-out vertices in gray (annotated with matching $\mathbf{u}'^1$ and $\mathbf{u}'^2$ that yield the same $\mathbf{r}$ given $\mathbf{u}^1 \setminus \mathbf{U}' = \mathbf{u}^2 \setminus \mathbf{U}'$)

— 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. □

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 parametrization. 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

Given the graphical characterization of non-g-identifiable queries, we devise GID (Alg. 1), a complete identification algorithm for g-identifiability, which outputs, for a given query, a formula expressed in terms of available probability distributions. The design of the algorithm shares the same principles established by previous identifiability algorithms (e.g., IDENTIFY [Tian, 2002], ID [Shpitser and Pearl, 2006], zID [Bareinboim and Pearl, 2012]). We decompose the identification process into two parts — pre- and post-activation of an available distribution, where SUB-ID is a subroutine taking care of (classic) identification task for each factored query with a fixed distribution treated as an observational one for the graph local to the recursive call.

The algorithm takes a query $P_{\mathbf{x}}(\mathbf{y})$, the causal graph $\mathcal{G}$, and available experiments $\mathbb{Z}$ as inputs. 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 the 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. At Line 7, given a factorized query, the algorithm examines whether an available distribution might be useful 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.*

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

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($\mathbf{y}, \mathbf{x}\cap\mathbf{S}', \prod_{V_i\in\mathbf{S}'} Q(V_i|\mathbf{V}_\pi^{(i-1)}\cap\mathbf{S}', \mathbf{v}_\pi^{(i-1)}\setminus\mathbf{S}'), \mathbf{S}'$)

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 $\mathbf{X}'$ into three parts: $\mathbf{X}'_1 = \mathbf{X}'\cap\mathbf{X}$; $\mathbf{X}'_2 = \mathbf{X}'\cap\mathbf{W}$; and $\mathbf{X}'_3 = \mathbf{X}'\setminus(\mathbf{X}\cup\mathbf{W})$ where $\mathbf{W}$ is the set of variables which was combined with $\mathbf{X}$ at Line 4, which occurs at most once. Then, there exists no directed edge from $\mathbf{X}'_1$ to $\mathbf{X}'_3$ (Line 4), and no bidirected edge between $\mathbf{X}'_3$ and $\mathbf{R}$.[8] 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 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. □

[8]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.

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 the 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 introduced a new graphical characterization for the g-identifiability problem, namely, uniquely computing a causal effect from a combination of observational and experimental distributions (including the case when the observational distribution is not available and an arbitrary set of experiments is available). We developed a general algorithm for g-identifiability and proved its completeness. This means that whenever it fails to g-identify an effect, this effect is provably not computable from the available data. These results were shown by leveraging two new graphical constructs, hedgelets and thickets, which lead to a better and more precise understanding of more involved forbidden structures. As a corollary, we proved the do-calculus is complete for the g-identifiability task.

Acknowledgements

This research is supported in parts by grants from NSF IIS-1704352, and 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. 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. *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 about 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 $\mathbf{Y}$).

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 exactly the same for nodes between $\mathbf{R}$ to $\mathbf{Y}$. □

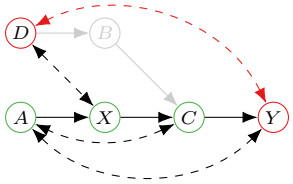


Figure 6: A causal graph $\mathcal{G}$ with a hedge $\langle \mathcal{F}, \mathcal{F}' \rangle$ for $P_x(y)$ where $\mathcal{F} = \mathcal{G} \setminus \{B\}$ with $\mathcal{F}'$ shown in red and variables in $\mathcal{F}''$ shown in green. Bit-parity of D and Y should be mapped to Y through B and C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