A Graphical Criterion for Effect Identification in Equivalence Classes of Causal Diagrams

Amin Jaber[1]**, Jiji Zhang**[2]**, Elias Bareinboim**[1]
[1] Department of Computer Science, Purdue University
[2] Department of Philosophy, Lingnan University
jaber0@purdue.edu, jijizhang@ln.edu.hk, eb@purdue.edu

Abstract

Computing the effects of interventions from observational data is an important task encountered in many data-driven sciences. The problem is addressed by identifying the post-interventional distribution with an expression that involves only quantities estimable from the pre-interventional distribution over observed variables, given some knowledge about the causal structure. In this work, we relax the requirement of having a fully specified causal structure and study the identifiability of effects with a singleton intervention (X), supposing that the structure is known only up to an equivalence class of causal diagrams, which is the output of standard structural learning algorithms (e.g., FCI). We derive a necessary and sufficient graphical criterion for the identifiability of the effect of X on all observed variables. We further establish a sufficient graphical criterion to identify the effect of X on a subset of the observed variables, and prove that it is strictly more powerful than the current state-of-the-art result on this problem.

1 Introduction

Establishing cause-and-effect relations is one prominent task throughout data-intensive sciences and engineering. In medicine, for example, one commonly needs to evaluate the effectiveness of a new drug, trying to disentangle its healing effect from that due to other factors (perhaps diet and hygiene) that may somehow correlate with the administration of the drug. In Artificial Intelligence, one may need to learn the effect of a robot's actions, while not having control over what may motivate that robot to act in the way it is behaving. These are a few applications of causal inference. There is a growing number of methods and techniques that allow researchers to reason with cause-and-effect relationships in a principled and efficient manner [Pearl, 2000; Spirtes *et al.*, 2001; Bareinboim and Pearl, 2016].

One classical method to infer causal effects is to perform randomized experiments [Fisher, 1951], where the action variables are randomized (e.g., the drug is randomly assigned to patients) and the outcomes observed (e.g., recovery of the patients). In many situations, however, it is not feasible to carry out an experiment of this sort for ethical, technological, financial, or other reasons. An alternative strategy proposed to estimate the effect of interest is to combine non-experimental (observational) data with some information about the underlying causal model [Pearl, 1993]. The primary challenge here is the existence of unobserved (latent) variables, which generates spurious association between action and outcome. The difference between the association between X and Y and the effect of X on Y is known as *confounding bias*. For example, despite the strong correlation observed between ice-cream consumption and drowning on the beach, no one really believes that eating ice-cream causes drowning during the holiday season. Formally, deciding whether a causal distribution is computable from a combination of the observational distribution and a causal model is known as the problem of identification of causal effects (identification, for short) [Pearl, 2000].

The problem of identification has been extensively studied in the literature, and a number of criteria have been proposed [Pearl, 1993; Pearl and Robins, 1995; Galles and Pearl, 1995; Kuroki and Miyakawa, 1999; Halpern, 2000], including the celebrated back-door criterion and the do-calculus [Pearl, 1995]. In a series of results, a novel graphical decomposition strategy was developed to solve the problem of identification given a causal diagram, along with completeness results for both observational and interventional identification [Tian and Pearl, 2002; Huang and Valtorta, 2006; Shpitser and Pearl, 2006; Bareinboim and Pearl, 2012]. Despite the generality of such results, their applicability is contingent upon the explicit articulation of a causal model, which is, unfortunately, not always available in many practical, large-scale settings. In fact, if one attempts to learn the causal structure from observational data, allowing for the possibility of latent confounders, in general only an equivalence class of causal diagrams (with latent variables) can be consistently inferred. A useful graphical representation of such an equivalence class is known as a *partial ancestral graph (PAG)*.

Identification from an equivalence class of causal diagrams represented by a PAG is considerably more challenging than from a single causal diagram due to the structural uncertainty of both the direct causal relations among the variables and the presence of latent variables that confounds causal relations between observed variables. Still, there is a growing

interest in identifiability results in the context of equivalence classes. [Zhang, 2008a] extended the *do-calculus* to PAGs, but, in practice, it is computationally hard to decide whether there exists (and, if so, find) a sequence of applications of the rules of the generalized calculus to identify the causal distribution. Another line of work [van der Zander *et al.*, 2014; Maathuis and Colombo, 2015; Perković *et al.*, 2015] established a generalized back-door criterion for PAGs, and provided a sound and complete algorithm to find a back-door adjustment set, should such a set exist. However, the back-door criterion is not nearly as powerful as the do-calculus, since no adjustment set exists for many identifiable causal effects.[1]

In this paper, we generalize to PAGs the powerful identification strategy for singleton interventions developed in DAGs [Tian and Pearl, 2002]. This new approach is computationally more attractive than do-calculus as it provides an algorithm to identify a causal effect, if identifiable. It is also more powerful than the generalized back-door criterion, as we show later. Specifically, we make the following contributions:

1. We derive a complete (necessary and sufficient) graphical criterion to identify the effect of a single variable X on the set of all observed variables $\mathbf{V}$ (or, $P_x(\mathbf{v})$).

2. We derive a graphical criterion to identify the effect of X on a subset of observed variables $\mathbf{S}$ (i.e. $P_x(\mathbf{s})$) and show that it subsumes the state-of-the-art adjustment criterion.

2 Preliminaries

In this paper, bold capital letters denote sets of variables, while bold lowercase letters stand for particular assignments to those variables. Whenever it is clear from the context, we write $P(\mathbf{V} = \mathbf{v})$ as $P(\mathbf{v})$.

2.1 Structural Causal Models

We use the language of Structural Causal Models (SCM) ([Pearl, 2000], pp. 204-207) as our basic framework. Formally, an SCM M is a 4-tuple $\langle \mathbf{U}, \mathbf{V}, \mathbf{F}, P(\mathbf{u}) \rangle$, where $\mathbf{U}$ is a set of exogenous (latent) variables and $\mathbf{V}$ is a set of endogenous (measured) variables. $\mathbf{F}$ represents a collection of functions $\mathbf{F} = \{f_i\}$ such that each endogenous variable $V_i \in \mathbf{V}$ is determined by a function $f_i \in \mathbf{F}$, where f_i is a mapping from the respective domain of $U_i \cup Pa_i$ to V_i, $U_i \subseteq \mathbf{U}$, $Pa_i \subseteq \mathbf{V} \setminus V_i$, and the entire set $\mathbf{F}$ forms a mapping from $\mathbf{U}$ to $\mathbf{V}$. The uncertainty is encoded through a probability distribution over the exogenous variables, $P(\mathbf{u})$, which induces a probability distribution over the measured (observed) variables, $P(\mathbf{v})$. A causal diagram associated with an SCM encodes the structural relations among $\mathbf{V} \cup \mathbf{U}$, in which an arrow is drawn from each member of $U_i \cup Pa_i$ to V_i, where Pa_i denotes the endogenous parents of V_i in the causal diagram. We assume the underlying structural system is acyclic. The observational distribution, $P(\mathbf{V})$, is a marginal over $\mathbf{V}$ of the joint distribution of $\mathbf{V} \cup \mathbf{U}$, and it factorizes according to the causal diagram.

$$P(\mathbf{v}) = \sum_{\mathbf{u}} \prod_i P(v_i|pa_i, u_i)P(\mathbf{u}) \qquad (1)$$

Within the structural semantics, performing an action $X{=}x$ is represented through the do-operator, *do(X=x)*, which encodes the operation of replacing the original equation for X by the constant x and induces a submodel M_x. The resulting distribution is denoted by P_x, which is the main target for identification in this paper. For a more detailed discussion of structural causal models, we refer readers to [Pearl, 2000; Spirtes *et al.*, 2001; Bareinboim and Pearl, 2016].

2.2 Identification Given a Causal DAG

We will build on the notion of c-components and the graphical condition for the identification of $P_x(\mathbf{v})$ developed in [Tian and Pearl, 2002].

Definition 1 (C-Component). *In a causal DAG, two observed variables are said to be in the same confounded component (c-component) if and only if they are connected by a bi-directed path, i.e. a path composed solely of such bi-directed treks as $V_i \leftarrow U_{ij} \rightarrow V_j$, where U_{ij} is an exogenous variable.*

For convenience we will often refer to a bi-directed trek like $V_i \leftarrow U_{ij} \rightarrow V_j$ as a bi-directed edge between V_i and V_j.

Proposition 1 (Tian and Pearl). *Given a causal diagram $\mathcal{G}$, let $V_1 < V_2 < \cdots < V_n$ be a topological order over $\mathbf{V}$, and let $V^{(i)} = \{V_1, \ldots, V_i\}$ with $i = 1, \ldots, n$ and $V^{(0)} = \phi$. $P_x(\mathbf{v})$ is identifiable given $\mathcal{G}$ if and only if X is not in the same c-component with any of its children. If identifiable, the effect is given by*

$$P_x(\mathbf{v}) = \frac{P(\mathbf{v})}{\prod_{\{i|V_i \in S^X\}} P(v_i|v^{(i-1)})} \sum_{x'} \prod_{\{i|V_i \in S^X\}} P(v_i|v^{(i-1)})$$

where S^X is the c-component that X belongs to and x' ranges over all the values of X.

2.3 Ancestral Graphs

We now introduce a graphical representation of equivalence classes of causal diagrams. A *mixed* graph can contain directed ($\rightarrow$) and bi-directed edges ($\leftrightarrow$). A is a *spouse* of B if $A \leftrightarrow B$ is present. An *almost directed cycle* happens when A is both a spouse and an ancestor of B. An *inducing path relative to* $\mathbf{L}$ is a path on which every node $V \notin \mathbf{L}$ (except for the endpoints) is a collider on the path (i.e., both edges incident to V are into V) and every collider is an ancestor of an endpoint of the path. A mixed graph is *ancestral* if it doesn't contain a directed or almost directed cycle. It is *maximal* if there is no inducing path (relative to the empty set) between any two non-adjacent nodes. A *Maximal Ancestral Graph* (MAG) is a graph that is both ancestral and maximal. A MAG entails a conditional independence model by a generalization of d-separation called m-separation, and MAG models are closed under marginalization [Richardson and Spirtes, 2002]. Algorithm 1 is a procedure for marginalization of MAGs.

Given a causal DAG $\mathcal{G}(\mathbf{V}, \mathbf{L})$ where $\mathbf{V}$ and $\mathbf{L}$ are observed and latent variables, respectively, we can marginalize out $\mathbf{L}$ and obtain a MAG $\mathcal{M}_\mathcal{G}$ over $\mathbf{V}$ by Alg. 1. This MAG retains

---

[1]Another promising approach is based on SAT solvers [Hyttinen *et al.*, 2015]. Given its somewhat distinct nature, a closer comparison lies outside the scope of this paper.

Algorithm 1: MAG Marginalization

Input : MAG $\mathcal{G}$ over $\mathbf{V} \cup \mathbf{L}$

Output: MAG $\mathcal{M}_{\mathcal{G}}$ over $\mathbf{V}$

1- for pair $X, Y \in \mathbf{V}$: X and Y adjacent in $\mathcal{M}_{\mathcal{G}}$ iff there exist an inducing path between them relative to $\mathbf{L}$ in $\mathcal{G}$.

2- for each pair of adjacent nodes $X, Y \in \mathcal{M}_{\mathcal{G}}$:

   1. $X \rightarrow Y$ if X is ancestor of Y in $\mathcal{G}$.

   2. $X \leftarrow Y$ if Y is ancestor of X in $\mathcal{G}$.

   3. $X \leftrightarrow Y$ otherwise.

both the independence and the ancestral relations among variables in $\mathbf{V}$ that are entailed by the original DAG. In general, a causal MAG represents a set of causal DAGs with the same set of observed variables that entail the same independence and ancestral relations among the observed variables.

Different MAGs may be Markov equivalent in that they entail the exact same independence model. A partial ancestral graph (PAG) represents an equivalence class of MAGs $[\mathcal{M}]$, which shares the same adjacencies as every MAG in $[\mathcal{M}]$ and displays all and only the invariant edge marks.

Definition 2 (PAG). *Let $[\mathcal{M}]$ be the Markov equivalence class of an arbitrary MAG $\mathcal{M}$. The PAG for $[\mathcal{M}]$, $\mathcal{P}$, is a partial mixed graph such that:*

   i. $\mathcal{P}$ has the same adjacencies as $\mathcal{M}$ (and hence any member of $[\mathcal{M}]$) does.

   ii. An arrowhead is in $\mathcal{P}$ iff it is shared by all MAGs in $[\mathcal{M}]$.

   iii. A tail is in $\mathcal{P}$ iff it is shared by all MAGs in $[\mathcal{M}]$.

   iv. A mark that is neither an arrowhead nor a tail is recorded as a circle.

A PAG is learnable from the conditional independence and dependence relations among the observed variables [Zhang, 2008b], and represents an equivalence class of DAG models with the same observed variables.

Given a PAG, a path between X and Y is *potentially directed (causal)* from X to Y if there is no arrowhead on the path pointing towards X. Y is called a *possible descendant* of X and X a *possible ancestor* of Y if $X = Y$ or there is a potentially directed path from X to Y. We write $\text{An}(\mathbf{Y})$ to denote the set of possible ancestors of set $\mathbf{Y}$. A set $\mathbf{A}$ is *ancestral* if $\text{An}(\mathbf{A}) = \mathbf{A}$. Y is called a *possible child* of X and X a *possible parent* of Y if they are adjacent and the edge is not into X.

A directed edge $X \rightarrow Y$ in a MAG or PAG is *visible* if there exists no DAG $\mathcal{G}(\mathbf{V}, \mathbf{L})$ in the corresponding equivalence class where there is an inducing path between X and Y that is into X relative to $\mathbf{L}$. This implies that a visible edge is not confounded (i.e. $X \leftarrow U_i \rightarrow Y$ doesn't exist, for any $U_i \in \mathbf{L}$). Which directed edges are visible is easily decidable by a graphical condition [Zhang, 2008a], so we will simply mark visible edges by v. For brevity, we refer to any edge that is not a visible directed edge as *invisible*.

If the edge marks on a path between X and Y are all circles, we call the path a *circle path*. For convenience, We use an asterisk (*) to denote any of the possible marks of a PAG $(\circ, >, -)$ or a MAG $(>, -)$.
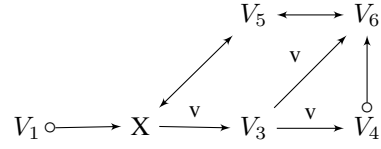
Figure 1: Causal query $P_x(\mathbf{v})$ over PAG $\mathcal{P}$.

3 Identification of $P_x(\mathbf{v})$

We first define the notion of identification in PAGs, which generalizes the model-specific notion [Pearl, 2000, pp. 70].

Definition 3. *Given a PAG $\mathcal{P}$ over $\mathbf{V}$ and a query $P_{\mathbf{t}}(\mathbf{s})$ where $\mathbf{T}, \mathbf{S} \subset \mathbf{V}$. $P_{\mathbf{t}}(\mathbf{s})$ is identifiable in PAG $\mathcal{P}$ if and only if $P_{\mathbf{t}}(\mathbf{s})$ is identifiable in all the Markov equivalent DAGs with the same expression.*

Let $\mathbf{V}$ denote the set of all nodes in a given PAG $\mathcal{P}$. In this section, we generalize Proposition 1 to PAGs and derive an identification criterion for the effect of X on all variables in $\mathbf{V} \setminus X$, denoted by $P_x(\mathbf{v})$. The following challenges are immediate when considering this more general setting:

1. Structural uncertainty regarding c-components.

2. Lack of a topological order over the variables with respect to a PAG.

To address the first challenge, we generalize the notion of c-component for MAGs and PAGs.

Definition 4 (PC-Component). *Given a MAG or a PAG, two nodes X and Y are in the same possible c-component (pc-component) if there is a path between the two variables such that (1) all non-endpoint nodes along the path are definite colliders, and (2) none of the edges, if directed, are visible.*

Given that all the non-endpoint nodes along the path, if any, are colliders by the first condition of Def. 4, the second condition is only concerned about the first and the last edges on the path, as all the other edges must be bi-directed. For instance, V_1 and V_4 in Fig. 1 are in the same pc-component due to the path $\langle V_1, X, V_5, V_6, V_4 \rangle$. The following proposition shows that being in the same pc-component in a MAG or PAG is necessary for being in the same c-component in some DAG in the corresponding equivalence class.

Proposition 2. *Let X and Y be two nodes in a MAG or PAG $\mathcal{P}$. If X and Y are not in the same pc-component in $\mathcal{P}$, then X and Y are not in the same c-component in any DAG in the equivalence class.*[2]

For example, X and V_3 in Figure 1 are not in the same pc-component, and consequently they are not in the same c-component in any DAG in the equivalence class.

The converse of Proposition 2 does not hold for PAGs (though it does for MAGs). However, the following weaker proposition holds and is sufficient for our purpose:

Proposition 3. *Let X and Y be two nodes in a MAG or PAG $\mathcal{P}$. If X and Y are in the same pc-component in $\mathcal{P}$, and either X and Y are adjacent or there is no circle path between them, then X and Y are in the same c-component in some DAG in the equivalence class.*

---

[2]See [Jaber *et al.*, 2018] for all the proofs.

Algorithm 2: PTO Algorithm

Input : PAG $\mathcal{G}$ over $\mathbf{V}$
Output: PTO over $\mathcal{G}$
1- Create singleton buckets $\mathbf{B}_i$ each containing $V_i \in \mathbf{V}$.
2- Merge buckets $\mathbf{B}_i$ and $\mathbf{B}_j$ if there is a circle edge
  between them ($\mathbf{B}_i \ni X \circ\!\!-\!\!\circ Y \in \mathbf{B}_j$).
3- **while** *set of buckets (*$\mathbf{B}$*) is not empty* **do**
    | (i) Extract $\mathbf{B}_i$ with only arrowheads incident on it.
    | (ii) Remove edges between $\mathbf{B}_i$ and other buckets.
end
4- The partial order is $\mathbf{B}_1 < \cdots < \mathbf{B}_m$ in reverse order
  of bucket extraction, i.e. $\mathbf{B}_1$ is the last extracted bucket.

For example, the condition in Proposition 3 is satisfied for X and V_4 in Figure 1. Hence, there exists a DAG in the equivalence class where X and V_4 are in the same c-component.

A special case of pc-component is the following:

Definition 5 (DC-Component). *In a MAG or PAG, two variables are in the same definite c-component, dc-component, if and only if they are connected with a bi-directed path, i.e. a path composed solely of bi-directed edges.*

For instance, nodes X and V_6 in Figure 1 are in the same dc-component, which implies that they are in the same c-component in every DAG in the equivalence class. It is important to consider when nodes are in the same c-component in some or all the DAGs in the equivalence class.

In the sequel, we address the second challenge by showing how to construct a partial topological order over a PAG that is valid for all DAGs in the equivalence class. It is easy to see that we can't always find a complete topological order over the nodes in a PAG that is valid for all DAGs (consider e.g., $X \circ\!\!-\!\!\circ Y$). Instead, Algorithm 2, called *PTO*, constructs a partial topological order. We refer to the output of the algorithm as *Bucketed PAG* (BPAG). Note that each BPAG's bucket is called in the literature as a *circle component*.

Consider the PAG in Fig. 1 where we intend to construct a topological order. All the buckets are singleton sets since there are no circle edges. Hence, a possible extraction order is $V_1 < V_5 < X < V_3 < V_4 < V_6$, which is valid for all DAGs in the equivalence class.

Lemma 1. *The* PTO *algorithm is sound, i.e. the partial order is valid for all the DAGs in the equivalence class.*

We are now ready to state the main theorem for the identification of $P_x(\mathbf{v})$.

Theorem 1. *Let a partial topological order over PAG $\mathcal{P}$ be $\mathbf{B}_1 < \mathbf{B}_2 < \cdots < \mathbf{B}_m$, and let $\mathbf{B}^{(i)} = \bigcup\{\mathbf{B}_1, \ldots, \mathbf{B}_i\}$, $i = 1 \ldots m$, and $\mathbf{B}^{(0)} = \emptyset$. $P_x(\mathbf{v})$ is identifiable if and only if X is not in the same pc-component with any of its possible children. When identifiable, the effect is given by*

$$P_x(\mathbf{v}) = \frac{P(\mathbf{v})}{\prod_{\{i|\mathbf{B}_i \subseteq \mathbf{S}^X\}} P(\mathbf{B}_i|\mathbf{B}^{(i-1)})} \sum_{x'} \prod_{\{i|\mathbf{B}_i \subseteq \mathbf{S}^X\}} P(\mathbf{B}_i|\mathbf{B}^{(i-1)})$$
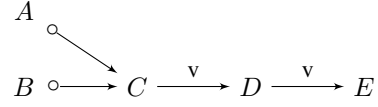
where $\mathbf{S}^X$ is the dc-component of X.

Figure 2: Sample PAG for the special cases.

Proof Sketch. Proposition 3 is sufficient to guarantee that if X is in the same pc-component with a possible child in $\mathcal{P}$, then there is a DAG in the equivalence class in which X is in the same c-component with a child, and so $P_x(\mathbf{v})$ is not identifiable. Conversely, Proposition 2 entails that if X is not in the same pc-component with any of its possible children in $\mathcal{P}$, then X satisfies the condition of Proposition 1 in every DAG in the equivalence class. We can then show that in every DAG, the identification expression is equivalent to the expression above. See Appendix for details. $\square$

Consider the PAG in Figure 1. Since X is not in the same pc-component with any of its possible children, namely V_3, $P_x(\mathbf{v})$ is identifiable by Theorem 1. The dc-component of X is $S^X = \{X, V_5, V_6\}$, and we use the topological order derived earlier (i.e. $V_1 < V_5 < X < V_3 < V_4 < V_6$). Hence, the expression for the causal effect is given by

$$P_x(\mathbf{v}) = \frac{P(\mathbf{v})}{P(v_5|v_1)P(x|v_1,v_5)P(v_6|v_1,v_5,x,v_3,v_4)} \times \quad (2)$$
$$\sum_{x'} P(v_5|v_1)P(x'|v_1,v_5)P(v_6|v_1,v_5,x',v_3,v_4)$$

This example also illustrates why the expression in Thm. 1 discards nodes that are in the pc-component but not in the dc-component of X. V_4 is in the pc-component of X (i.e. $X \leftrightarrow V_5 \leftrightarrow V_6 \leftarrow\!\circ V_4$) and satisfies the condition in Prop. 3, hence it is in the c-component of X in some DAG in the equivalence class. However, the fact that $V_4 \perp\!\!\!\perp X|(V_1, V_3, V_5)$ can be used to simplify the expression to that in Thm. 1.

3.1 Special Cases

In two special cases, analogous to those considered in the context of DAGs [Tian and Pearl, 2002], the expression in Theorem 1 can be considerably simplified. These simpler results are worth mentioning because the more compact identification expressions in these special cases entail a lower sample and computational complexity when evaluating them from data, and because the simpler graphical conditions allow the causal analyst to decide identifiability (yes/no) almost immediately by inspection. Let Pa_x and Ch_x denote the sets of possible parents and possible children of X, respectively.

Corollary 1. *If all the edges incident on X are visible, then $P_x(\mathbf{v})$ is identifiable and is given by*

$$P_x(\mathbf{v}) = P(\mathbf{v}|x, pa_x)P(pa_x)$$

Proof. The condition implies that X is not in the same pc-component with any node, including its possible children, and that $S^X = \{X\}$. Hence, the condition of Th. 1 is satisfied and the identification expression becomes $P(\mathbf{v})/P(x|x^{(i-1)})$.
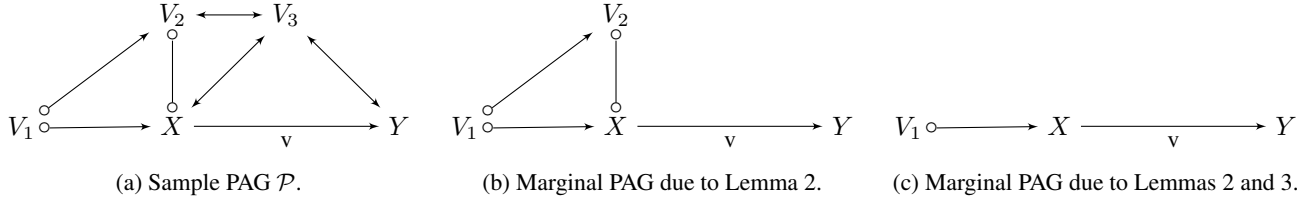
(a) Sample PAG $\mathcal{P}$.

(b) Marginal PAG due to Lemma 2.

(c) Marginal PAG due to Lemmas 2 and 3.

Figure 3: Example of Definition 7.

Note that $X \perp\!\!\!\perp X^{(i-1)} \setminus Pa_x | Pa_x$ since every path p between X and $X^{(i-1)} \setminus Pa_x$ includes either (1) an edge out of X, in which case there is a collider along p in $\mathbf{V} \setminus X^{(i)}$ that blocks p, or (2) a directed edge into X which is a definite non-collider along p ($pa_x \ni V_i \to X$) and so blocks p. Hence, $P(x|x^{(i-1)})$ simplifies to $P(x|pa_x)$. □

For example, consider the PAG in Figure 2 and the distribution $P_d(\mathbf{v})$. The effect is identifiable by Corollary 1 as

$$P_d(\mathbf{v}) = P(a,b,e|c,d)P(c)$$

The second case relaxes the previous condition on X but imposes a condition on the possible children of X.

Corollary 2. *If all the edges incident on the possible children of X are visible, then $P_x(\mathbf{v})$ is identifiable and is given by*

$$P_x(\mathbf{v}) = \Big(\prod_{\{i|V_i \in Ch_x\}} P(v_i|pa_i) \Big) \sum_{x'} \frac{P(\mathbf{v})}{\prod_{\{i|V_i \in Ch_x\}} P(v_i|pa_i)}$$

Proof Sketch. The identification criterion of Theorem 1 is satisfied as the nodes in Ch_x are not in the same pc-component with any other node, including X. The identification expression can be rewritten as

$$\Big(\prod_{\{i|\mathbf{B_i} \not\subseteq \mathbf{S^x}\}} P(\mathbf{B_i}|\mathbf{B^{(i-1)}}) \Big) \sum_{x'} \frac{P(\mathbf{v})}{\prod_{\{i|\mathbf{B_i} \not\subseteq \mathbf{S^x}\}} P(\mathbf{B_i}|\mathbf{B^{(i-1)}})}$$

We then simplify the expression using the independence relations among the variables to obtain the expression above. The detailed proof can be found in the Appendix. □

Consider, for example, the distribution $P_c(\mathbf{v})$ over the PAG in Figure 2. The effect is identifiable and given by

$$P_c(\mathbf{v}) = P(d|c)\sum_{c'} \frac{P(\mathbf{v})}{P(d|c')}$$

$$= P(d|c)\sum_{c'} P(a,b,e|c',d)P(c')$$

4 Identification of $P_x(\mathbf{s})$

So far, we have derived a complete criterion to identify $P_x(\mathbf{v})$ in a given PAG. Now, suppose that we are interested in estimating the effect of X on a subset of observed variables ($\mathbf{S} \subseteq \mathbf{V} \setminus X$). One may be tempted to surmise that Theorem 1 should be enough for this problem – namely, first identify $P_x(\mathbf{v})$, and then marginalize out unintended variables ($\mathbf{V} \setminus \mathbf{S} \cup X$). For instance, for the query $P_x(v_4)$ in

Fig. 1, $P_x(\mathbf{v})$ can be computed first (i.e. Eq. 2), and then $\mathbf{V} \setminus \{X, V_4\}$ can be marginalized out. While this solution is certainly sound when the conditions of the theorem are met, the strategy is not necessary. To see this, consider the query $P_x(y)$ over the PAG in Fig. 3a. As we will show later, $P_x(y)$ is identifiable despite the non-identifiability of $P_x(\mathbf{v})$. The aim of this section is to explain these subtleties and derive a stronger graphical criterion to identify $P_x(\mathbf{s})$ where $\mathbf{S} \subseteq \mathbf{V} \setminus \{X\}$ is the outcome set. We start by introducing a new construction called *marginal PAG* which allows us to systematically eliminate nodes that need not be considered.

4.1 Marginal PAG

Definition 6 (Marginal PAG). *Let $[\mathcal{M}]$ be a Markov equivalence class of MAGs over $\mathbf{V}$. For $\mathbf{A} \subseteq \mathbf{V}$, let $[\mathcal{M}]_\mathbf{A} = \{\mathcal{M}'_\mathbf{A} | \mathcal{M}' \in [\mathcal{M}]\}$ where $\mathcal{M}'_\mathbf{A}$ denotes the MAG over $\mathbf{A}$ that results from marginalizing out $\mathbf{V} \setminus \mathbf{A}$ in M' (Algorithm 1). A marginal PAG for $[\mathcal{M}]$ relative to $\mathbf{A}$ is the partial mixed graph that has the same adjacencies as every graph in $[\mathcal{M}]_\mathbf{A}$ and displays all and only the shared edge marks in $[\mathcal{M}]_\mathbf{A}$.*

Note that in this definition, $[\mathcal{M}]_\mathbf{A}$ is in general *not* a full equivalence class of MAGs over $\mathbf{A}$, but a subset of an equivalence class. For example, let $[\mathcal{M}]$ be the equivalence class represented by the PAG in Fig. 3a. Let $\mathbf{A} = \{V_1, X, Y\}$. Then, every MAG in $[\mathcal{M}]_\mathbf{A}$, according to the above definition, contains an edge $V_1 *\!\!\to X$ and an edge $X \to Y$. Consequently, the marginal PAG relative to $\mathbf{A}$ is the graph in Fig. 3c. In this case, $[\mathcal{M}]_\mathbf{A}$ is not a full equivalence class because, for example, $V_1 \leftarrow X \leftarrow Y$ is also Markov equivalent to graphs in $[\mathcal{M}]_\mathbf{A}$ but not contained therein. If we consider the full equivalence class of which $[\mathcal{M}]_\mathbf{A}$ is a subset, the corresponding PAG is $V_1 \circ\!-\!\circ X \circ\!-\!\circ Y$. Therefore, a marginal PAG according to our definition is not an ordinary PAG, and is in general more informative.

The following two lemmas describe two cases of constructing a marginal PAG that are relevant to our purpose.

Lemma 2. *Let $\mathbf{A}$ be an ancestral set in a PAG $\mathcal{P}$. The marginal PAG for the equivalence class represented by $\mathcal{P}$ relative to $\mathbf{A}$ is simply $\mathcal{P}_\mathbf{A}$, the induced subgraph of $\mathcal{P}$ over $\mathbf{A}$. Furthermore, a visible edge in $\mathcal{P}$ remains visible in the marginal PAG.*

Lemma 3. *Let $\mathcal{P}$ be a PAG over $\mathbf{V}$ and let $\mathbf{B}$ be a circle component in $\mathcal{P}$ that is partitioned into two nonempty sets $\mathbf{T}$ and $\mathbf{C}$, i.e. $\mathbf{T} \cup \mathbf{C} = \mathbf{B}$ and $\mathbf{T} \cap \mathbf{C} = \emptyset$. If every possible child of $\mathbf{C}$ is in $\mathbf{B}$, then the marginal PAG relative to $\mathbf{V} \setminus \mathbf{C}$ can be constructed from $\mathcal{P}$ as follows:*

1. Remove $\mathbf{C}$ and all the incident edges, and;

2. *Add a circle edge between two non-adjacent nodes in* **T** *if there exists a circle path between them where every node along the path is in* **C**.

Moreover, all the visible edges in $\mathcal{P}$ remain visible in the marginal PAG.

Although a marginal PAG is not necessarily an ordinary PAG, for the marginal PAGs constructed according to Lemmas 2 and 3, we can show that they retain a crucial graphical property of ordinary PAGs, namely:

Lemma 4. *The following property holds in a marginal PAG constructed according to Lemmas 2 and 3:*
for any three nodes A, B, C, if $A*\!\!\to B \circ\!\!-\!\!* C$, then there is an edge between A and C with an arrowhead at C, namely, $A*\!\!\to C$. Furthermore, if the edge between A and B is $A \to B$, then the edge between A and C is either $A \to C$ or $A\circ\!\!\to C$ (i.e., it is not $A \leftrightarrow C$).

Proof. The proof is trivial for the marginal PAG in Lemma 2, for a violation of it in the marginal PAG obviously implies its violation in the original PAG, which is not possible.

As for the marginal PAG in Lemma 3, it can only introduce new adjacencies between nodes in the form of circle edges within a circle component (bucket). Hence, a violation of the property in the marginal PAG also implies its violation in the original PAG which is not possible. $\square$

It follows from Lemma 4 that a marginal PAG constructed by Lemmas 2 and 3 preserves the main properties established for a PAG, and specifically all the properties needed for the derivation in Section 3. We can then define a *simplified PAG*:

Definition 7 ($\mathcal{P}_{\mathbf{Y}}^X$). *Given a PAG $\mathcal{P}$ over $\mathbf{V}$. $\mathcal{P}_{\mathbf{Y}}^X$, referred to as simplified PAG with respect to X and $\mathbf{Y}$, is the result of applying the marginalization in Lemma 2 relative to $An(\mathbf{Y})$, and then Lemma 3, if applicable, with respect to $\mathbf{T} = \mathcal{X}$, where $\mathcal{X} = \{X\} \cap An(\mathbf{Y})$ and $\mathbf{C} \cap \mathbf{Y} = \emptyset$.*

For example, the simplified PAG $\mathcal{P}_Y^X$ for the PAG in Fig. 3a is constructed as follows. The set $An(Y) = \{Y, X, V_1, V_2\}$ is ancestral in $\mathcal{P}$, hence Lemma 2 can be applied, and the marginal PAG over $An(Y)$, $\mathcal{P}'$, is given in Fig. 3b. In $\mathcal{P}'$, nodes X and V_2 correspond to the sets $\mathbf{T}$ and $\mathbf{C}$ in Lemma 3, respectively. X is the only possible child of V_2 and is contained in the corresponding circle component. So, Lemma 3 is applicable, which yields the simplified PAG in Fig. 3c.

4.2 A Sufficient Criterion

The relevance of marginal PAGs is due to the following result:

Lemma 5. *Given a PAG $\mathcal{P}$, $P_x(\mathbf{s})$ is identifiable in $\mathcal{P}$ if $P_x(\mathbf{s})$ is identifiable in $\mathcal{P}_{\mathbf{S}}^X$ where $\mathcal{X} = \{X\} \cap An(\mathbf{S})$.*

In other words, we can focus our attention on the component of X that persists in the simplified graph, and ignore all the variables that are marginalized out. For instance, given the query $P_x(y_1)$ over the PAG $Y_1 \circ\!\!\to X \leftarrow\!\!\circ Y_2$, Lemma 5 suggests that we drop X along with Y_2 from the simplified PAG and the corresponding marginal distribution, i.e. $P(y_1)$. Hence, the interventional distribution for this trivial query is $P(y_1)$. We use this observation to prove the main result of this section, an identification criterion for $P_x(\mathbf{s})$.
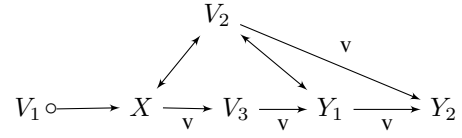

Figure 4: $P_x(\mathbf{y})$ is not identifiable using the adjustment criterion.

Theorem 2. *Given a PAG $\mathcal{P}$, $P_x(\mathbf{s})$ is identifiable if $\mathcal{X} = \{X\} \cap An(\mathbf{S})$ is not in the same pc-component with any of its possible children in $\mathcal{P}_{\mathbf{S}}^X$.*

Proof. Let $\mathbf{V}'$ be the set of variables in $\mathcal{P}_{\mathbf{S}}^X$. By Lemma 5, it is sufficient to consider the query $P_x(\mathbf{s})$ over $\mathcal{P}_{\mathbf{S}}^X$. As stated in Subsection 4.1, all the properties required for the correctness of the PTO algorithm and Theorem 1 remain valid in $\mathcal{P}_{\mathbf{S}}^X$. Since the condition here is just the condition of Thm. 1 over $\mathcal{P}_{\mathbf{S}}^X$, $P_x(\mathbf{v}')$ is identifiable using Theorem 1. We then marginalize out $\mathbf{V}' \setminus \mathbf{S} \cup \{\mathcal{X}\}$ to get $P_x(\mathbf{s})$. $\square$

Given a query $P_x(\mathbf{s})$, Thm. 2 provides a sufficient condition over the simplified PAG $\mathcal{P}_{\mathbf{S}}^X$ such that the causal distribution can be computed through the formula in Thm. 1. For non-trivial queries of the form $P_x(\mathbf{v})$, note that the simplified PAG remains $\mathcal{P}$ and Thm. 2 reduces to Thm. 1. Consider the example in Fig. 3a and the causal query $P_x(y)$. The corresponding simplified PAG $\mathcal{P}_Y^X$ is shown in Fig. 3c. We compute $P_x(y, v_1)$ by applying Thm. 1 over $\mathcal{P}_Y^X$, then we marginalize out variable V_1, obtaining

$$P_x(y) = \sum_{v_1} \left(\frac{P(v_1, x, y)}{p(x|v_1)} \sum_{x'} p(x'|v_1) \right)$$
$$= \sum_{v_1} P(v_1) P(y|x, v_1) = P(y|x)$$

4.3 Criterion Strength

The identification criterion for $P_x(\mathbf{s})$ in Theorem 2 is strictly more powerful than the generalized adjustment criterion proposed in [Perković *et al.*, 2016], which is proven to be complete for adjustment. For example, the causal query $P_x(y_1, y_2)$ over the PAG in Figure 4 is identifiable using the criterion in Theorem 2 while it is not identifiable using the adjustment method. On the other hand, the following theorem shows that there is no singleton intervention effect that can be identified using the adjustment method but not identifiable using Theorem 2.

Theorem 3. *Let $\mathcal{P}$ be a PAG over a set of nodes $\mathbf{V}$ and let $P_x(\mathbf{s})$ be a causal query where $X \in \mathbf{V}$, $\mathbf{S} \subseteq \mathbf{V} \setminus X$. If $P_x(\mathbf{s})$ is not identifiable using Theorem 2, then there exist no set $\mathbf{Z}$ that satisfies the generalized adjustment criterion.*

5 Conclusion

In this paper, we investigated the problem of identification of causal distributions with singleton interventions in a Markov equivalence class represented by a PAG. We proved three graphical criteria for the identification of $P_x(\mathbf{v})$, where $\mathbf{V}$ is the set of all variables, including a general criterion that

is necessary and sufficient for the identifiability of $P_x(\mathbf{v})$. These results can already be used to identify causal queries in challenging settings that backdoor-like methods cannot solve (e.g., given the PAG in Fig. 4, $P_x(y_1, y_2)$ is identifiable using Corollary 2, but not by adjustment). In addition, we introduced a new construction called marginal PAGs, with which we derived a sufficient graphical condition for the identification of $P_x(\mathbf{s})$, where $\mathbf{S}$ is a subset of the variables. Our criterion was shown to be strictly stronger than the state-of-the-art adjustment method found in the literature. We expect that our results will be helpful to causal analysts when studying complex, high-dimensional settings where learning the full causal model is often infeasible.

Acknowledgments

References

[Bareinboim and Pearl, 2012] Elias Bareinboim and Judea Pearl. Causal inference by surrogate experiments: z-identifiability. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 2012.

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

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

[Galles and Pearl, 1995] David Galles and Judea Pearl. Testing identifiability of causal effects. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, UAI'95, pages 185–195, 1995.

[Halpern, 2000] Joseph Y. Halpern. Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research*, 12(1):317–337, May 2000.

[Huang and Valtorta, 2006] Y. Huang and M. Valtorta. Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, pages 217–224, 2006.

[Hyttinen *et al.*, 2015] A. Hyttinen, F. Eberhardt, and M. Järvisalo. Do-calculus when the true graph is unknown. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 395–404, 2015.

[Jaber *et al.*, 2018] Amin Jaber, Jiji Zhang, and Elias Bareinboim. A graphical criterion for effect identification in equivalence classes of causal diagrams. Technical report, R-32, Purdue AI Lab, Department of Computer Science, Purdue University, 2018.

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

[Maathuis and Colombo, 2015] Marloes H Maathuis and Diego Colombo. A generalized back-door criterion. *The Annals of Statistics*, 43(3):1060–1088, 2015.

[Pearl and Robins, 1995] Judea Pearl and James Robins. Probabilistic evaluation of sequential plans from causal models with hidden variables. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, UAI'95, pages 444–453, 1995.

[Pearl, 1993] Judea Pearl. Aspects of graphical models connected with causality. In *Proceedings of the 49th Session of the International Statistical Institute*, pages 391–401, 1993.

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

[Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000. 2nd edition, 2009.

[Perković *et al.*, 2015] Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H Maathuis. A complete generalized adjustment criterion. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pages 682–691, 2015.

[Perković *et al.*, 2016] Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *arXiv preprint arXiv:1606.06903*, 2016.

[Richardson and Spirtes, 2002] Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *Annals of Statistics*, pages 962–1030, 2002.

[Shpitser and Pearl, 2006] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the Twenty-First Conference on Artificial Intelligence*, 2006.

[Spirtes *et al.*, 2001] Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, prediction, and search*, volume 81. MIT press, 2001.

[Tian and Pearl, 2002] Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Proceedings of the Eighteenth Conference on Artificial Intelligence*, AAAI'02, pages 567–573, 2002.

[van der Zander *et al.*, 2014] Benito van der Zander, Maciej Liskiewicz, and Johannes Textor. Constructing separators and adjustment sets in ancestral graphs. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pages 11–24, 2014.

[Zhang, 2008a] Jiji Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9(Jul):1437–1474, 2008.

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