TECHNICAL REPORT
R-32-L
July 2018

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.

$$P(\mathbf{v}) = \sum_{\mathbf{u}} \prod_{i} P(v_i|pa_i, u_i)P(\mathbf{u}) \tag{1}$$

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. $X \to Y$ 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 $\mathtt{An}(\mathbf{Y})$ to denote the set of possible ancestors of set $\mathbf{Y}$. A set $\mathbf{A}$ is *ancestral* if $\mathtt{An}(\mathbf{A}) = \mathbf{A}$. 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 $X \to Y$ 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. $X \leftarrow U_i \to Y$ 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 ($>, -$).

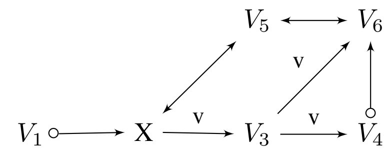


Figure 1: Causal query $P_x(\mathbf{v})$ over PAG $\mathcal{P}$.

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 $\mathbf{B}^{(i)} = \bigcup\{\mathbf{B}_1, \dots, \mathbf{B}_i\}$, $i = 1 \dots m$, 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.

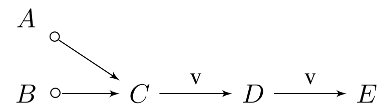


Figure 2: Sample PAG for the special cases.

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. □

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

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

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.

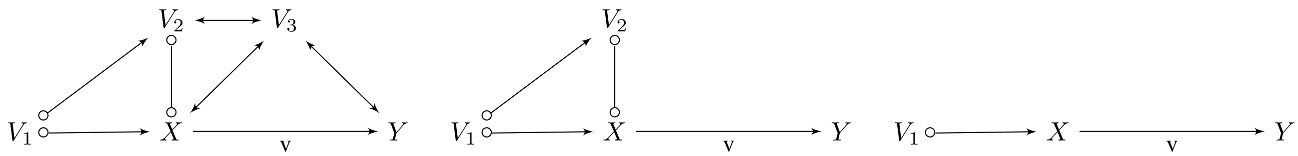
3.1 Special Cases

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

$$\begin{aligned} P_c(\mathbf{v}) &= P(d|c) \sum_{c'} \frac{P(\mathbf{v})}{P(d|c')} \\ &= P(d|c) \sum_{c'} P(a, b, e|c', d)P(c') \end{aligned}$$

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* $\mathbf{T}$ *if there exists a circle path between them where every node along the path is in* $\mathbf{C}$.

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* $A{*}\!\!\rightarrow B \circ\!\!-\!\!* C$, *then there is an edge between* A *and* C *with an arrowhead at* C, *namely,* $A{*}\!\!\rightarrow C$. *Furthermore, if the edge between* A *and* B *is* $A \rightarrow B$, *then the edge between* A *and* C *is either* $A \rightarrow C$ *or* $A\circ\!\!\rightarrow C$ *(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 ($\mathcal{P}^X_{\mathbf{Y}}$)**.** *Given a PAG* $\mathcal{P}$ *over* $\mathbf{V}$. $\mathcal{P}^X_{\mathbf{Y}}$, *referred to as simplified PAG with respect to* X *and* $\mathbf{Y}$, *is the result of applying the marginalization in Lemma 2 relative to* $\mathtt{An}(\mathbf{Y})$, *and then Lemma 3, if applicable, with respect to* $\mathbf{T} = \mathcal{X}$, *where* $\mathcal{X} = \{X\} \cap \mathtt{An}(\mathbf{Y})$ *and* $\mathbf{C} \cap \mathbf{Y} = \emptyset$.

For example, the simplified PAG $\mathcal{P}^X_Y$ for the PAG in Fig. 3a is constructed as follows. The set $\mathtt{An}(Y) = \{Y, X, V_1, V_2\}$ is ancestral in $\mathcal{P}$, hence Lemma 2 can be applied, and the marginal PAG over $\mathtt{An}(Y)$, $\mathcal{P}'$, 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* $P_{\mathcal{X}}(\mathbf{s})$ *is identifiable in* $\mathcal{P}^X_{\mathbf{S}}$ *where* $\mathcal{X} = \{X\} \cap \mathtt{An}(\mathbf{S})$.

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 $Y_1 \circ\!\!\rightarrow X \leftarrow\!\!\circ Y_2$, 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 $P_x(\mathbf{s})$.

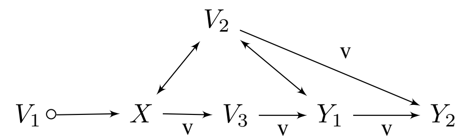


Figure 4: $P_x(\mathbf{y})$ is not identifiable using the adjustment criterion.

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

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

Given a query $P_x(\mathbf{s})$, Thm. 2 provides a sufficient condition over the simplified PAG $\mathcal{P}^X_{\mathbf{S}}$ 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 $\mathcal{P}^X_Y$ is shown in Fig. 3c. We compute $P_x(y, v_1)$ by applying Thm. 1 over $\mathcal{P}^X_Y$, then we marginalize out variable V_1, obtaining

$$\begin{aligned} P_x(y) &= \sum_{v_1} \Big(\frac{P(v_1, x, y)}{p(x|v_1)} \sum_{x'} p(x'|v_1) \Big) \\ &= \sum_{v_1} P(v_1) P(y|x, v_1) = P(y|x) \end{aligned}$$

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

We thank the reviewers for all the feedback provided. Bareinboim's research was supported in parts by grants from NSF IIS-1704352 and IIS-1750807 (CAREER). Zhang's research was supported in part by the Research Grants Council of Hong Kong under the General Research Fund LU13600715.

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.

“A Graphical Criterion for Effect Identification in Equivalence Classes of Causal Diagrams” Supplemental Material

A Full proofs

Proof of Proposition 2. This result follows from Definition 4 and the contrapositives of lemmas 6 and 7. □

Proof of Proposition 3. This result follows from Lemmas 8 and 9. □

Proof of Lemma 1. We refer to the second step of the algorithm as ”Bucketing” and to the third step as ”Extraction”. First, note that all the edges within a bucket in the BPAG, if any, are circle edges (∘–∘). Initially, the buckets contain single variables and there are no edges within the buckets so our claim is valid. In the Bucketing step, We merge two buckets $\mathbf{B_i}$ and $\mathbf{B_j}$ when they have a circle edge between them. This means that there is a circle path between any two variables $X \in \mathbf{B_i}$ and $Y \in \mathbf{B_j}$. By lemma 14, an edge between X and Y if it exists would only be a circle edge. Hence, all the edges in a bucket are circle edges only. Buckets $\mathbf{B_1}$ and $\mathbf{B_3}$ in the previous example illustrate this idea.

If a bucket $\mathbf{B_i}$ has no tails or circles incident on it, then none of its variables are possible ancestors of any variable in the remaining buckets. This implies that no node in bucket $\mathbf{B_i}$ is an ancestor of any node in the other buckets in all the DAGs in the equivalence class. Hence, it is valid to put bucket $\mathbf{B_i}$ in front of all the remaining buckets in the topological order. This partial order between the buckets is valid in all the DAGs in the equivalence class of PAG $\mathcal{P}$. A problem arises if we don’t have any bucket with no tail or circle incident on it. Next, we prove that such a case does not exist which concludes our proof.

In the BPAG, we don’t have circle edges (∘–∘) between the buckets. If there is an edge from variable X in bucket $\mathbf{B_i}$ into a variable Y in another bucket $\mathbf{B_j}$, then there is an edge from X into every variable in bucket $\mathbf{B_j}$. This follows from lemma 14 (2) since there is a circle path between Y and every other variable in bucket $\mathbf{B_j}$. This is an important observation for what follows.

Consider any iteration of the Extraction step of the algorithm. Assume every remaining bucket has one or more circles or tails incident on it, then there exist a possible directed cycle structure over the buckets as shown below with $2 \leq k \leq m$ where ? stands for a circle or tail. The existence of this structure in a PAG contradicts Lemma 20. Hence, such a possible directed cycle over the buckets can’t exist and the assumption is invalid. Consequently, we have at least one bucket with no tails or circles incident on it at every iteration of the Extraction step. This concludes the proof.

$$\begin{tikzcd} B_{i_1} \; ? \arrow[r] & B_{i_2} \; ? \arrow[r, dashed] & ? \; B_{i_k} \arrow[ll, bend right] \end{tikzcd}$$

□

Proof of Theorem 1. First we prove the (*if*):

In order to satisfy the condition of the Theorem, Proposition 2 necessitates that all edges incident on X are either into X or out of it. Also, all the edges out of X should be visible. Otherwise, there exist a path of size one between X and one of its possible children and the edge is not visible. This also implies that X is in a singleton bucket since it is not connected to any other node with a circle edge.

If the condition of the theorem is satisfied, then X is not in the same c-component with any of its children in all the DAGs in the equivalence class of $\mathcal{P}$. Hence, the causal effect $P_x(v)$ is identifiable in all the DAGs and is given by the expression in Proposition 1. The partial topological order represented by the BPAG is valid for all the DAGs, so it can be used to derive the conditional terms $P(\mathbf{B_i}|\mathbf{B^{(i-1)}})$, where $\mathbf{B_i} \subseteq \mathbf{S^x}$. In order to prove that the causal effect in all the DAGs is unanimous, we need to show that the expression for each DAG under the given partial order is simplified into the one given in Theorem 1.

If there is a bi-directed path between X and node $O_i \in \mathbf{B_i}$ ($\langle X, \ldots, O_{i-1}, O_i \rangle$) in a PAG, then there is a bi-directed path between X and every node in $\mathbf{B_i}$. By lemma 14, if there is bi-directed edge between O_{i-1} and O_i, then there is a bi-directed edge between O_{i-1} and every node in $\mathbf{B_i}$. Hence, $\mathbf{B_i} \subseteq \mathbf{S^X}$ and the conditional probability terms in the expression referring to the nodes in $\mathbf{B_i}$ can be jointly represented by $P(\mathbf{B_i}|\mathbf{B^{(i-1)}})$ since they are consecutive in any order over $\mathbf{B_i}$.

Suppose that a node $O_i \in \mathbf{B_i}$ is in the same pc-component as X, but not the same dc-component, then there is no bi-directed path between the two nodes in the PAG. This implies that there is no bi-directed path between X and any node in $\mathbf{B_i}$ as well. If bucket $\mathbf{B_i}$ comes before the bucket of X ($\mathbf{B_j}$) in the BPAG, then the conditional term referring to O_i in the identification solution of each DAG does not condition on X and will cancel out with the one in the denominator. If $\mathbf{B_i}$ comes after $\mathbf{B_j}$ in the BPAG, we show using lemma 11 that $\mathbf{B_i} \perp\!\!\!\perp \mathbf{B_j}|\mathbf{B^{(i-1)}} \setminus \mathbf{B_j}$ and hence $O_i \perp\!\!\!\perp X|\mathbf{B^{(i-1)}} \setminus X$ in the PAG. This implies that $O_i \perp\!\!\!\perp X|\mathbf{B^{(i-1)}} \setminus X$ in each DAG in the equivalence class and X can be removed from the conditional term of O_i in the identification expression. Hence, we can perform the same simplification as before. This will conclude the proof.

First, if X is in the same pc-component with $O_i \in \mathbf{B_i}$, then there is no path of size one between X and any node in $\mathbf{B_i}$. Suppose that there is such a path between X and $O_j \in \mathbf{B_i}$, then it should be one of the following:

1. $X \leftarrow\!\!* O_j$: if the edge is into O_j, then there is a bi-directed edge between X and each node in $\mathbf{B_i}$ according to lemma 14. This contradicts with X being in the same pc-component as O_i. If the edge is not into O_j, then X is a possible descendant of O_j and this contradicts with the partial topological order since $\mathbf{B_i}$ comes after $\mathbf{B_j}$.
2. $X \circ\!\!-\!\!\circ O_j$: This is not possible as X and O_j are in different buckets.
3. $X \rightarrow O_j$ or $X \circ\!\!\rightarrow O_j$: According to lemma 14, there must be an edge from X to each node in $\mathbf{B_i}$ including O_i. Thus, X is in the same pc-component with a possible child. This violates the condition of Theorem 1.

Next, we consider all the proper definite status paths between $\mathbf{B_i}$ and $\mathbf{B_j}$ in PAG $\mathcal{P}$. Since there are no bi-directed paths between nodes in $\mathbf{B_i}$ and $\mathbf{B_j}$, then each path has at least one of the following properties:

1. Contains at least one definite non-collider.
2. The path is not into $\mathbf{B_i}$.
3. The path is not into X ($\mathbf{B_j}$).

Cases one and two are the conditions for lemma 11 with the special case where $\boldsymbol{\mathcal{H}} = \mathbf{B_j}$ is a singleton bucket (X). We show that case three can't happen as it violates the condition of Theorem 1. Hence, $\mathbf{B_i} \perp\!\!\!\perp X|\mathbf{B^{(i-1)}} \setminus X$ using lemma 11.

If cases one and two are not applicable to a proper definite status path p between $\mathbf{B_i}$ and $\mathbf{B_j}$, then case three must apply. Hence, the path is as follows $X \rightarrow O_j \leftrightarrow \cdots \leftrightarrow O_k$ where all the non-endpoint nodes are colliders and the path is into $O_k \in \mathbf{B_i}$. The edge between X and O_j has to be directed and visible, otherwise it violates the condition of the Theorem. Recall that X is in the same pc-component with $O_i \in \mathbf{B_i}$. Now O_j, the child of X, is in the same dc-component as $O_k \in \mathbf{B_i}$, then O_j is in the same dc-component as O_i since there is a bi-directed edge between the last non-endpoint node along path p and every node in $\mathbf{B_i}$ including O_i (lemma 14). If X is in the same pc-component with O_i, then X is in the same c-component with O_i in some DAG $\mathcal{G}$ in the equivalence class according to Proposition 3. Also, O_j is in the same c-component with O_i in $\mathcal{G}$, hence X and O_j are in the same c-component in $\mathcal{G}$. Therefore, X is in the same pc-component with a possible child (O_j) by lemmas 6 and 7 which violates the condition of the Theorem.

Next we prove the *(only if))*:

If X is in the same pc-component with a possible child in $\mathcal{P}$, we can always construct a DAG in the equivalence class of $\mathcal{P}$ where X is in the same c-component with a child. $P_x(\mathbf{v})$ is not identifiable in such a DAG according to Theorem 1.

Let C be a possible child of X in $\mathcal{P}$ where X and C are in the same pc-component, then X and C are adjacent and the edge is not into X. If the edge has a circle incident on X or it is oriented but not visible ($X \rightarrow C$), lemma 22 shows that we can construct a MAG $\mathcal{M}$ such that (1) $\mathcal{M}$ is in the equivalence class of $\mathcal{P}$ and (2) edge $X \rightarrow C$ is not visible. Next, lemma 13 provides a construction of a DAG $\mathcal{G}$ from MAG $\mathcal{M}$ such that (1) $\mathcal{G}$ is in the equivalence class of $\mathcal{M}$ and (2) edge $X \rightarrow C$ in $\mathcal{G}$ is confounded. Hence, we are able to construct a DAG that is in the equivalence class of $\mathcal{P}$ where the causal effect $P_x(v)$ is not identifiable.

Another case is when the edge between X and C is visible. Then, there exist a collider path between X and C as follows $X \leftrightarrow O_1 \leftarrow\!\!\dashrightarrow O_n \leftarrow\!\!* C$ where $n \geq 1$. Note the edge between X and O_1 is bi-directed for otherwise O_1 would be a possible child of X and X and O_1 are in the same pc-component, hence we go back to the previous case. Then, we use lemma 22 to orient the PAG into a MAG such that edge $O_n \leftarrow\!\!* C$ remains invisible. Next, we use lemma 13 to construct a DAG in the equivalence class of the MAG. The construction introduces a latent variable in place of each bi-directed edge in the MAG. Hence, the DAG will have a bi-directed path between X and C while keeping C a child of X. □

Proof of Corollary 2. Under the condition of the corollary, all the edge marks at X are heads ($>$) and tails ($-$) and all the edges out of X are visible, otherwise X has a possible child with an invisible edge. Hence, the identification condition 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

$$P_x(\mathbf{v}) = \prod_{\{i|\mathbf{B_i} \not\subseteq \mathbf{S^x}\}} P(\mathbf{B_i}|\mathbf{B^{(i-1)}}) \sum_x \frac{P(\mathbf{v})}{\prod_{\{i|\mathbf{B_i} \not\subseteq \mathbf{S^x}\}} P(\mathbf{B_i}|\mathbf{B^{(i-1)}})}$$

where $\mathbf{S^X}$ is the dc-component of X.

If $\mathbf{B_i}$ comes before X in the BPAG, then the conditional term doesn't depend X and the ones in the denominator cancels out with the corresponding one outside the summation. Otherwise, consider all the proper definite status paths between X and $\mathbf{B_i}$. Since there are no bi-directed paths between X and the nodes in $\mathbf{B_i}$, then each path has at least one of the following properties:

1. Contains at least one definite non-collider.
2. The path is not into $\mathbf{B_i}$.

3. The path is not into X. This is impossible as it implies that one of the possible children of X has a bi-directed edge incident on it.

It follows using lemma 11 that $(\mathbf{B_i} \perp\!\!\!\perp X|\mathbf{B^{(i-1)}} \setminus X)$ for every $\mathbf{B_i}$ in the expression above except the ones referring to Ch_x. Thus, the term in the denominator cancels out with the corresponding term outside the summation and we have the following

$$P_x(\mathbf{v}) = \prod_{\{i|V_i \in Ch_x\}} P(v_i|\mathbf{v^{(i-1)}}) \sum_x \frac{P(\mathbf{v})}{\prod_{\{i|V_i \in Ch_x\}} P(\mathbf{v_i}|\mathbf{v^{(i-1)}})}$$

Moreover, let V_i, as shown in the expression above, be some node in Ch_x. Then, $(V_i \perp\!\!\!\perp V^{(i-1)} \setminus Pa_{V_i}|\ Pa_{V_i})$ since every path p between V_i and $V^{(i-1)} \setminus Pa_{V_i}$ includes either (1) an edge out of V_i, hence there is a definite collider along p in $V \setminus V^{(i)}$ which is not conditioned on, or (2) a directed edge into V_i which is a definite non-collider along p ($Pa_x \ni V_i \to X$). Thus, $P(v_i|v^{(i-1)})$ simplifies to $P(v_i|pa_i)$ This concludes the proof. □

Proof of Lemma 2. According to the work of [Richardson and Spirtes, 2002], ancestral graphs are closed under marginalization, hence each MAG in the equivalence class has a corresponding MAG which represents the marginal independence model over $\mathbf{A}$ when we marginalize out $\mathbf{V} \setminus \mathbf{A}$. Let $\mathcal{M}$ correspond to any MAG in the equivalence class of $\mathcal{P}$. The steps to marginalize a subset of nodes in an input MAG are given in Algorithm 1. It follows that all the adjacencies and the corresponding orientations in the sub-MAG of $\mathcal{M}$ over $\mathbf{A}$ remain valid. We show that the marginalization over $\mathbf{V} \setminus \mathbf{A}$ doesn't introduce any new inducing paths and hence we have no additional adjacencies than those found in $\mathcal{M}_{\mathbf{A}}$.

Suppose that there exist a new inducing path relative to $\mathbf{V} \setminus \mathbf{A}$ between the variables $X, Y \in \mathbf{A}$ and denote this path by p. Let L be the last node along p starting from X where $L \notin \mathbf{A}$. If L is a collider along the path, then it has to be an ancestor of X or Y, otherwise p is not an inducing path. But L is not an ancestor of X or Y since $\mathbf{A}$ is ancestral and $L \notin \mathbf{A}$. If L is a non-collider along p then at least one of its two adjacent nodes along p is not in $\mathbf{A}$ as well, otherwise L is an ancestor of $\mathbf{A}$ which is not possible. This contradicts the choice of L as the last node along p starting from X where $L \notin \mathbf{A}$. Therefore, there is no new inducing path relative to $\mathbf{V} \setminus \mathbf{A}$ between any two variables in $\mathbf{A}$ and the marginal MAG corresponding to $\mathcal{M}$ is the sub-MAG over $\mathbf{A}$. Given the above result, it follows that the sub-PAG over $\mathbf{A}$ represents the marginal PAG over $\mathbf{A}$.

Next we prove the second part of the lemma concerned with visible edges. Note that the marginalization over $\mathbf{V} \setminus \mathbf{A}$ doesn't introduce any new inducing path between the nodes in $\mathbf{A}$. Then, a directed edge between $X, Y \in \mathbf{A}$ that is visible in $\mathcal{P}$ remains visible in this marginal PAG. □

Proof of Lemma 3. According to [Richardson and Spirtes, 2002], ancestral graphs are closed under marginalization, hence each MAG in the equivalence class has a corresponding MAG which represents the marginal independence model over $\mathbf{V} \setminus \mathbf{C}$. We consider a MAG $\mathcal{M}^*$ in the equivalence class of $\mathcal{P}$ using the construction in lemma 19 and we use it to ease proving the lemma.

The steps to marginalize a subset of nodes in an input MAG are given in Algorithm 1. Marginalizing out subset $\mathbf{C}$ in a MAG doesn't remove any adjacency between the nodes in $\mathbf{V} \setminus \mathbf{C}$ but might introduce new adjacencies between them due to a new inducing path relative to $\mathbf{C}$. Let $\mathbf{D}$ denote the set of nodes in $\mathcal{P}$ that are possible descendants of $\mathbf{B}$ and let $\mathbf{A}$ denote $\mathbf{V} \setminus \{\mathbf{B} \cup \mathbf{D}\}$. Suppose there exist at least one inducing path relative to $\mathbf{C}$ between two non-adjacent nodes in $\mathcal{M}^*$. We have the following observations about any inducing path relative to $\mathbf{C}$ in $\mathcal{M}^*$:

1. All the proper possible children of $\mathbf{B}$ in $\mathcal{P}$ are due to $\mathbf{T}$ according to the condition in the lemma. Also, all edges $\circ\!\!\rightarrow$ are oriented as $\rightarrow$ in $\mathcal{M}^*$. Hence, any inducing path relative to $\mathbf{C}$ in $\mathcal{M}^*$ can't include any edge $B_i \to D_i$ where $B_i \in \mathbf{B}$ and $D_i \in \mathbf{D}$ since B_i would be an observable non-collider along the path.

2. Suppose there exist at least one edge $A_i *\!\!-\!\!* B_i$ along an inducing path relative to $\mathbf{C}$ such that $A_i \in \mathbf{A}$ and $B_i \in \mathbf{B}$. Then, We have the followings remarks. The corresponding edge in $\mathcal{P}$ must be into B_i as A_i is not a possible descendant of any node in $\mathbf{B}$. Also, if such an edge exists in $\mathcal{P}$, then there is an edge from A_i into every node in $\mathbf{B}$ by lemma 14. Hence, those edges exist in $\mathcal{M}^*$ as well and are into $\mathbf{B}$. Moreover, the construction of $\mathcal{M}^*$ doesn't introduce any new bi-directed edges. If $A_i \leftrightarrow B_i$ is in $\mathcal{M}^*$, then it is present in $\mathcal{P}$ and consequently we have $A_i \leftrightarrow B_j$ for all $B_j \in \mathbf{B}$ in $\mathcal{P}$ (lemma 14), hence those bi-directed edges are present in $\mathcal{M}^*$ too. We consider the following exhaustive options for an inducing path relative to $\mathbf{C}$ which includes at least one edge between $\mathbf{A}$ and $\mathbf{B}$:

   (a) Both endpoints of the path are in $\mathbf{B}$: A_i must be a collider along the path and hence we have $A_i \leftrightarrow B_i$ in $\mathcal{M}^*$. By the previous remarks, we have $A_i \leftrightarrow B_j$ for all $B_j \in \mathbf{B}$. This creates a contradiction as A_i shares a bi-directed edge with an endpoint of the path while being an ancestor of it. Hence, this case is not possible.

   (b) One endpoint B_k is in $\mathbf{B}$: Starting from the endpoint not in $\mathbf{B}$, let $A_i *\!\!\rightarrow B_i$ be the first edge where $A_i \in \mathbf{A}$ and $B_i \in \mathbf{B}$. Again, A_i must be a collider and we have $A_i \leftrightarrow B_j$ for all $B_j \in \mathbf{B}$. Then, we have $A_i \leftrightarrow B_k$ in $\mathcal{M}^*$ where B_k is an endpoint of the inducing path. Hence, we have an inducing path which doesn't go through any node in $\mathbf{C}$. We have a contradiction with our assumption that this is a new inducing path relative to $\mathbf{C}$.

(c) None of the endpoints is in $\mathbf{B}$: By observation (1), an inducing path relative to $\mathbf{C}$ can't include any proper edge out of $\mathbf{B}$, then we must have the following along the inducing path: $A_i{*\!\!\rightarrow} B_i \circ - - - \circ B_j \leftarrow\!\!* A_j$ where $B_i \circ - - - \circ B_j$ is a sequence of nodes in $\mathbf{B}$. There must exist at least one collider in $\mathcal{M}^*$ along $A_i{*\!\!\rightarrow} B_i \circ - - - \circ B_j \leftarrow\!\!* A_j$ and the collider must be an ancestor of one of the endpoints. The endpoint can't be in $\mathbf{A}$ as none of its nodes are possible descendants of $\mathbf{B}$ in $\mathcal{P}$. Hence, the endpoint (D_i) must be in $\mathbf{D}$. The directed path from the collider to D_i must go through some node $T_i \in \mathbf{T}$ by the condition of the lemma. Therefore, by Lemma 14 , there exist another inducing path that goes through $A_i{*\!\!\rightarrow} T_i \leftarrow\!\!* A_j$ and doesn't include any node in $\mathbf{C}$. Hence, the endpoints of the inducing path are already adjacent which is a contradiction.

Every option among (a), (b), and (c) reaches a contradiction. Hence, a new inducing path relative to $\mathbf{C}$ between two non-adjacent nodes in $\mathcal{M}^*$ doesn't go through any edge between $\mathbf{A}$ and $\mathbf{B}$.

By observations (1) and (2), an inducing path relative to $\mathbf{C}$ between two non-adjacent nodes doesn't include edges between $\mathbf{A}$ and $\mathbf{B}$ nor edges between $\mathbf{B}$ and $\mathbf{D}$. Hence, such a path is contained within $\mathbf{B}$ and consequently any new adjacencies in $\mathbf{M}^*$ are between the nodes in $\mathbf{T} \subset \mathbf{B}$. Consider $T_i, T_j \in \mathbf{T} \subset \mathbf{B}$ to be two non-adjacent nodes in $\mathcal{M}^*$:

- If there exist a circle path between T_i and T_j in $\mathcal{P}$ where every node along the path is in $\mathbf{C}$, then there exist an uncovered circle path p^* between T_i and T_j where every node along this path is in $\mathbf{C}$ (lemma 17). The corresponding path in $\mathcal{M}^*$ doesn't include any non-endpoint collider by construction. Hence, this path between T_i and T_j is an inducing path relative to $\mathbf{C}$ in $\mathcal{M}^*$ and T_i and T_j are adjacent when marginalizing out $\mathbf{C}$.
- Otherwise, there exist at least one observable node (in $\mathbf{T}$) along every circle path between T_i and T_j. Suppose one of those paths is an inducing path relative to $\mathbf{C}$ in $\mathcal{M}^*$ and let p^* be one such path with the fewest number of observable nodes. Let $\langle B_1, B_2, B_3\rangle$ be a triple along p^* where $B_2 \in \mathbf{T}$, B_1 is closer to T_i, and B_3 closer to T_j. B_2 is a collider along p^* and consequently B_1 and B_3 must be adjacent in $\mathcal{P}$ and $\mathcal{M}^*$. Recall that the construction of $\mathcal{M}^*$ doesn't introduce new bi-directed edges. Hence, we have $B_1 \rightarrow B_2 \leftarrow B_3$ with $B_1 \rightarrow B_3$ or $B_1 \leftarrow B_3$. Without loss of generality, suppose the edge is out of B_1 and consider path p' composed of $p_1 = \langle T_i, \ldots, B_1\rangle$, $B_1 \rightarrow B_3$, and $p_2 = \langle B_3, \ldots, T_j\rangle$ where p_1 and p_2 are the corresponding subpaths of p^*. We have the following remarks on p' compared to p^*:
  (a) B_1 is a non-collider along both p^* and p' and hence $B_1 \in \mathbf{C}$.
  (b) If B_3 has both edges out of it along p^*, then B_3 is a non-collider along both paths and is in $\mathbf{C}$. Otherwise, the first edge along p_2 is into B_3 and B_3 is a collider along p'. Also, B_3 is an ancestor of one of the endpoints since we have $B_3 \rightarrow B_2$ and B_2 is a collider along p^* and an ancestor of one of the endpoints.

  Based on remarks (a) and (b), p' is an inducing path relative to $\mathbf{C}$ with fewer number of observed variables than p^* which is a contradiction. Hence, no inducing path relative to $\mathbf{C}$ exists between T_i and T_j.

Next we show that the above adjacencies are common for all the MAGs in the equivalence class of $\mathcal{P}$. Let $\mathcal{M} \neq \mathcal{M}^*$ be a MAG in the equivalence class of $\mathcal{P}$ such that there exist an inducing path between V_1 and V_2 relative to $\mathbf{C}$ which is not present in $\mathcal{M}^*$. This implies that for some $\mathbf{Z} \subseteq \mathbf{V} \setminus \{\mathbf{C}, V_1, V_2\}$, $(V_1 \perp\!\!\!\perp V_2|\mathbf{Z})$ in $\mathcal{M}^*$ while $(V_1 \not\!\perp\!\!\!\perp V_2|\mathbf{Z})$ in $\mathcal{M}$. However, this contradicts with the fact that both MAGs share the same conditional independence model. So, such a MAG $\mathcal{M}$ does not exist in the equivalence class of $\mathcal{P}$. The same argument can be applied for an inducing path that is present in $\mathcal{M}^*$ but not in $\mathcal{M}$. Hence, the marginal MAG over $\mathbf{V} \setminus \mathbf{C}$ for every MAG $\mathcal{M}$ in the equivalence class of $\mathcal{P}$ is the sub-graph over $\mathbf{V} \setminus \mathbf{C}$ with additional adjacencies according to the criterion in the lemma.

Hence, the marginal PAG over $\mathbf{V} \setminus \mathbf{C}$ is the sub-graph of $\mathcal{P}$ over $\mathbf{V} \setminus \mathbf{C}$ with additional adjacencies within $\mathbf{T}$ according to the criterion in the lemma. One issue remains unaddressed is the orientation of those additional edges in the marginal PAG. Suppose there exist a new edge in the marginal PAG between $T_i, T_j \in \mathbf{T}$. Recall that if there exist a circle path between T_i and T_j in $\mathcal{P}$, then there exist an uncovered circle path p^* between T_i and T_j (lemma 17). If the edge incident on T_i along p^* is oriented out of T_i in some MAG $\mathcal{M}$ in the equivalence class, then p^* must be a directed path from T_i to T_j in $\mathcal{M}$ since all the non-endpoints nodes along p^* are definite non-colliders in $\mathcal{P}$ and consequently in $\mathcal{M}$. The opposite applies if we orient the edge incident on T_j along p^* out of T_j and we have a directed path from T_j to T_i. Hence, the complete orientation of all the additional edges in the marginal PAG is with circles (∘—∘).

Finally, we show that visible edges in $\mathcal{P}$ remain visible in the marginal PAG. Consider a visible edge $X \rightarrow Y$ and suppose there is a new inducing path relative to $\mathbf{C}$ between X and Y that is into X. Since X is an ancestor of Y, this creates another inducing path between Y and any node Z that is adjacent to X ($Z{*\!\!\rightarrow} X$). Moreover, Z can't be in the same bucket as Y due to $X \rightarrow Y$ (lemma 14). Such cases are not possible in the marginal PAG as the additional adjacencies are within the bucket $\mathbf{B}$. Hence, visible edges in $\mathcal{P}$ remain visible in the marginal PAG. □

Proof of Lemma 5. If $P_{\mathcal{X}}(\mathbf{s})$ is identifiable in $\mathcal{P}^X_{\mathbf{S}}$, then the effect is unanimously identifiable in every DAG in the equivalence class. Let $\mathcal{D}'$ be one such DAG in the equivalence class. By Lemma 12, $P_{\mathcal{X}}(\mathbf{s})$ is identifiable in DAG $\mathcal{D}$, the corresponding DAG of $\mathcal{D}'$ before marginalization. Recall that X is a singleton intervention, consequently $\mathcal{X}$ is singleton or empty. If $\mathcal{X}$ is empty, then $X \notin \mathtt{An}(\mathbf{S})$ in $\mathcal{P}$ and consequently in $\mathcal{D}$. Hence, $P_{\mathcal{X}}(\mathbf{s}) = P_x(\mathbf{s})$ in $\mathbf{D}$ by rule 3 of the do-calculus. It follows that $P_x(\mathbf{s})$ is unanimously identifiable in every DAG in the equivalence class of $\mathcal{P}$, hence the effect is identifiable in $\mathcal{P}$. □

Proof of Theorem 3. If $P_x(\mathbf{s})$ is not identifiable using Theorem 2, then X is in the same pc-component with at least one possible child in $\mathcal{P}^X_{\mathbf{S}}$. By definition, all the nodes in $\mathcal{P}^X_{\mathbf{S}}$ are possible ancestors of $\mathbf{S}$. If there is at least one edge between X and a possible child and the edge is not out of X or not visible, then by the construction of $\mathcal{P}^X_{\mathbf{S}}$ (Definition 7) there exist a possible child of X (C) which is the first node along a proper possibly directed path from X to $\mathbf{S}$. The corresponding path is present in $\mathcal{P}$ and $\mathcal{P}$ is not amenable relative to (X,$\mathbf{S}$) (Definition 8). Hence, no set $\mathbf{Z}$ satisfies the generalized adjustment criterion (Definition 10) relative to (X,$\mathbf{S}$).

The other option is that all the edges incident on X are oriented arrowheads and tails and all the edges out of X are visible in $\mathcal{P}^X_{\mathbf{S}}$. Hence, X is in the same pc-component with at least one of its children (C) and there is a collider path p between X and C: $(X)V_0 \leftrightarrow V_1 \leftarrow\rightarrow V_{n-1} \leftarrow\!* V_n(C)$ where $n \geq 2$. Using lemma 23, it is sufficient to show that the set Adjust(X,$\mathbf{S}$) (Definition 11) doesn't satisfy the generalized adjustment criterion in $\mathcal{P}$ to prove that there is no adjustment set relative to (X,$\mathbf{S}$) in $\mathcal{P}$.

All the nodes in $\mathcal{P}^X_{\mathbf{S}}$ are possible ancestors of $\mathbf{S}$ and hence they are all in Adjust(X,$\mathbf{S}$), Forb(X,$\mathbf{S}$,$\mathcal{P}$), $\mathbf{S}$, or X. Moreover, there exist at least one node along p that is in Forb(X,$\mathbf{S}$,$\mathcal{P}$). One such node is C since it is along a proper possibly directed path from X to $\mathbf{S}$.

Let V_j be the closest node to X along p that is in $\mathbf{S}$ or Forb(X,$\mathbf{S}$,$\mathcal{P}$). If $V_j = V_1$, then $V_1 \in \mathbf{S}$ as X can't be an ancestor of V_1 due to $X \leftrightarrow V_1$ and hence it can't be in Forb(X,$\mathbf{S}$,$\mathcal{P}$). In this case, there is no adjustment set that blocks the non-causal path $X \leftrightarrow V_1$ and it follows that Adjust(X,$\mathbf{S}$) doesn't satisfy the generalized adjustment criterion in $\mathcal{P}$. If $j > 1$, then we have a collider path between X and V_j along which every V_i for $1 \leq i < j$ is in Adjust(X,$\mathbf{S}$). If $V_j \in \mathbf{S}$, then we have an unblocked definite status non-causal path between X and $V_j = S_i \in \mathbf{S}$ in $\mathcal{P}^X_{\mathbf{S}}$ and the same path is unblocked in $\mathcal{P}$ by Adjust(X,$\mathbf{S}$). Hence, Adjust(X,$\mathbf{S}$) doesn't satisfy the generalized adjustment criterion relative to (X,$\mathbf{S}$) in $\mathcal{P}$. Otherwise, V_j is in Forb(X,$\mathbf{S}$,$\mathcal{P}$) due to being a descendant of a node W where W lies along a proper possibly directed path from X to some node in $\mathbf{S}$. But, V_j is in $\mathcal{P}^X_{\mathbf{S}}$ so it is a possible ancestor of $\mathbf{S}$ and hence V_j lies along a proper possibly directed path from X to $\mathbf{S}$ as well. We refer to the latter path as p'. So based on lemma 17, there are uncovered possibly directed paths from X to V_j and from V_j to $S_i \in \mathbf{S}$ which we refer to as p_1 and p_2, respectively. All edges incident on X are oriented, so using lemma 18 p_1 is a directed path from X to V_j in $\mathcal{P}^X_{\mathbf{S}}$ and consequently in $\mathcal{P}$. By lemma 21, there exist a MAG $\mathcal{M}$ in the equivalence class of $\mathcal{P}$ with no additional arrowheads into V_j thus p_2 is a directed path from V_j to S_i in $\mathcal{M}$.

Now, consider the concatenated path composed of the subpath of p $\langle X, \dots, V_j \rangle$ and p_2 in $\mathcal{M}$. Note that V_j is a non-collider along this path in $\mathcal{M}$ as p_2 is out of V_j. Also, V_j is along a proper causal path from X to S_i in $\mathcal{M}$ since p_1 and p_2 are both directed paths. Then, Adjust(X,$\mathbf{S}$) doesn't block this proper non-causal path between X and $\mathbf{S}$ in $\mathcal{M}$. It follows using lemma 24 that there is an m-connecting proper definite status non-causal path between X and $\mathbf{S}$ in $\mathcal{P}$ given Adjust(X,$\mathbf{S}$). Therefore, Adjust(X,$\mathbf{S}$) doesn't satisfy the generalized adjustment criterion in $\mathcal{P}$. This concludes our proof. □

Lemma 6. *Let $\mathcal{G}(\mathbf{V}, \mathbf{L})$ be a DAG, and $\mathcal{M}$ be the MAG over $\mathbf{V}$ that represents the DAG. For any $X, Y \in \mathbf{V}$, if there is a bi-directed path p between X and Y in $\mathcal{G}$, then there is a path p' between X and Y in $\mathcal{M}$ such that (1) all the non-endpoint nodes are colliders, and (2) all directed edges on p' are not visible.*

Proof. We prove this lemma by induction on the size of the bi-directed path between X and Y in $\mathcal{G}$. Before we start with the proof, note that the only possible directed edges along the path p' are the first and last edges since all non-endpoint nodes are colliders. Also, we will prove the following stronger statement in order to prove our main lemma.

Let $\mathcal{G}(\mathbf{V}, \mathbf{L})$ be a DAG, and $\mathcal{M}$ be the MAG over $\mathbf{V}$ that represents the DAG. For any $X, Y \in \mathbf{V}$, if there is a bi-directed path p between X and Y in $\mathcal{G}$, then there is a path p' between X and Y in $\mathcal{M}$ such that:

1. all the non-endpoint nodes are colliders; and
2. all directed edges $A \rightarrow B$ on p', have an inducing path between A and B into A in $\mathcal{G}$.

The above statement implies the main lemma as the only difference is the second point regarding the directed edges along path p' in $\mathcal{M}$. It is evident that a directed edge in MAG $\mathcal{M}$ is not visible if it has an inducing path into its source in DAG $\mathcal{G}$ where $\mathcal{G}$ is in the equivalence class of $\mathcal{M}$.

Consider as a base case path p of size 1 between X and Y in $\mathcal{G}$. The bi-directed edge between X and Y ($X \leftarrow L_{XY} \rightarrow Y$) is an inducing path between X and Y relative to the latent variable L_{XY} in $\mathcal{G}$. Hence, X and Y are adjacent in MAG $\mathcal{M}$. If neither X nor Y is an ancestor of the other in $\mathcal{G}$, then the edge between X and Y in $\mathcal{M}$ is bi-directed and the lemma holds. If X is an ancestor of Y or the opposite in $\mathcal{G}$, then the edge in $\mathcal{M}$ is directed ($X \rightarrow Y$ or $X \leftarrow Y$). In both cases, the latent variable L_{XY} in $\mathcal{G}$ provides an inducing path between X and Y that is into both ends. Hence, the lemma holds.

In the induction step, we assume that the lemma holds for bi-directed paths between X and Y in $\mathcal{G}$ of length $\leq n$ and prove it for bi-directed paths of length $n + 1$. Consider the bi-directed path $p = \langle X, O_1, \dots, O_n, Y \rangle$ of length n+1 in $\mathcal{G}$. We split the path into two paths $p_1 = \langle X, O_1 \rangle$ and $p_2 = \langle O_1, \dots, Y \rangle$. By the assumption of the induction step, the lemma holds for each path and there are paths p'_1 and p'_2 in MAG $\mathcal{M}$ corresponding to p_1 and p_2 in $\mathcal{G}$. Consider the options of path p'_1:

case 1: O_1 is not an ancestor of X in $\mathcal{G}$. Hence, the edge between X and O_1 in $\mathcal{M}$ is into O_1 ($X \rightarrow O_1$ or $X \leftrightarrow O_1$). If the edge is directed ($X \rightarrow O_1$), there is an inducing path into X in $\mathcal{G}$ ($X \leftarrow L_{XO_1} \rightarrow O_1$). Now consider the first edge in p'_2 which, according to the lemma, could be one of the following two options:

case 1-1: The edge is $O_1 \leftrightarrow O_i$. Then, edge $X \rightarrow O_1$ concatenated with path p'_2 create a path between X and Y in $\mathcal{M}$ which is consistent with the conditions in the lemma. Hence, the lemma holds.

case 1-2: The edge is $O_1 \rightarrow O_i$. By the induction step, there is an inducing path between O_1 and O_i into O_1 in $\mathcal{G}$ so the edge $O_1 \rightarrow O_i$ is not visible in $\mathcal{M}$. Hence, X and O_i must be adjacent in $\mathcal{M}$ for otherwise edge $O_1 \rightarrow O_i$ would be visible. Also, the edge between X and O_i is directed into O_i, otherwise the MAG is not ancestral. Moreover, there is an inducing path between X and O_1 that is into X and there is an inducing path between O_1 and O_i that is into O_1. Node O_1 is a collider between those two inducing paths and an ancestor of O_i. So, there is an inducing path between X and O_i that is into X in $\mathcal{G}$. Hence, there is a path between X and Y in $\mathcal{M}$ composed of edge $X \rightarrow O_i$ and the subpath of p'_2 $\langle O_i, \ldots, Y \rangle$ which is consistent with the conditions of the lemma.

case 2: O_1 is an ancestor of X in $\mathcal{G}$. Hence, the edge between X and O_1 in $\mathcal{M}$ is $X \leftarrow O_1$. Also, there is an inducing path between X and O_1 into O_1 in $\mathcal{G}$ ($X \leftarrow L_{XO_1} \rightarrow O_1$) so edge $X \leftarrow O_1$ in $\mathcal{M}$ is not visible. Next, we consider the first edge in path p'_2.

case 2-1: The edge is $O_1 \leftrightarrow O_i$. Then, path p'_2 is a collider path into O_1 in $\mathcal{M}$. We show, by induction, that X is a child of each node along p'_2 in $\mathcal{M}$ until there exist a bi-directed edge between X and some node along p'_2. For the base case, O_i is adjacent to X for otherwise edge $X \leftarrow O_1$ is visible. Also, the edge is into X; otherwise the MAG is not ancestral. If the edge is bi-directed, then we are done and the statement holds; otherwise the edge is out of O_i. In the induction step, assume all nodes along path p'_2 starting with O_i until O_j are parents of X. We have a collider path between O_{j+1}, which could be Y, and O_i that is into O_1 and each collider is a parent of X. Hence, node O_{j+1} must be adjacent to X for otherwise $X \leftarrow O_1$ is visible. Also, the edge between X and O_{j+1} is into X for otherwise $\mathcal{M}$ is not ancestral due to $O_j \rightarrow X \rightarrow O_{j+1}$ and $O_{j+1} {*\!\!\rightarrow} O_j$. Hence, the statement holds.

If there exist a bi-directed edge between X and some node O_k along p'_2, then there exist a path between X and Y composed of $X \leftrightarrow O_k$ and the subpath of p'_2 ($\langle O_k, \ldots, Y \rangle$) that is consistent with the lemma. Otherwise, X and Y are adjacent and Y is a parent of X by the above induction. In the latter case, consider the concatenated path of p'_1 and p'_2 ($X \leftarrow O_1 \leftrightarrow O_i \leftarrow\!\!-\!\!-\!\!-\!\!\rightarrow O_j \leftrightarrow\!\!* Y$). There exist an inducing path between every two consecutive nodes and every non-endpoint node is an ancestor of X. Also, by the induction assumption, the first and last edges along the path have inducing paths into the source node (O_1 and Y if $O_j \leftarrow Y$). Hence, there exist an inducing path between X and Y that is into Y in $\mathcal{G}$ and the lemma holds.

case 2-2: The edge is $O_1 \rightarrow O_i$. Both edges $X \leftarrow O_1$ and $O_1 \rightarrow O_i$ have inducing paths into O_1 in $\mathcal{G}$. Hence, X and O_i are adjacent in $\mathcal{M}$. We consider all the orientations of the edge between X and O_i in $\mathcal{M}$. If the edge is bi-directed, then there is a path between X and Y in $\mathcal{M}$ composed of $X \leftrightarrow O_i$ and the subpath of p'_2 $\langle O_i, \ldots, Y \rangle$ and it is consistent with the conditions of the lemma. If the edge is out of X, then we have the path $\langle X, O_i, \ldots, Y \rangle$. Moreover, the edge $X \rightarrow O_i$ has an inducing path into X in $\mathcal{G}$. The reason is that O_1 is a collider in $\mathcal{G}$ between the two inducing paths of edges $X \leftarrow O_1$ and $O_1 \rightarrow O_i$ and it is an ancestor of X and O_i. Hence, the lemma holds too. The last case is when the edge is out of O_i, hence we have the path $X \leftarrow O_i, \ldots, Y$. Similar to the earlier case, the edge $X \leftarrow O_i$ has an inducing path into O_i in $\mathcal{G}$. The argument for this case follows same as that of case 2-1.

This exhausts all the options in the induction step and the lemma holds. □

Lemma 7. *Let $\mathcal{M}$ be a MAG over $\mathbf{V}$, and $\mathcal{P}$ be the PAG that represents the equivalence class of $\mathcal{M}$. For any $X, Y \in \mathbf{V}$, if there is a path p between X and Y in $\mathcal{M}$ such that (1) all non-endpoint nodes are colliders and (2) all directed edges, if any, are not visible, then there is a path p^* between X and Y in $\mathcal{P}$ such that (1) all non-endpoint nodes along the path are definite colliders, and (2) none of the edges are visible.*

Proof. We will use the following lemma from the work of Zhang [Zhang, 2006] several times throughout the proof.

lemma I: If a path $\langle U, \ldots, X, Y, Z \rangle$ is a discriminating path for Y in $\mathcal{M}$, and the corresponding subpath between U and Y in $\mathcal{P}$ is (also) a collider path, then the path is also a discriminating path for Y in $\mathcal{P}$.

We denote the sequence of nodes along a path between X and Y by $\langle X = O_0, \ldots, O_n = Y \rangle$. In order to prove the main lemma, we start by proving a simpler lemma.

lemma II: For any $X, Y \in \mathbf{V}$, if X and Y are not adjacent and there is a collider path p between X and Y in $\mathcal{M}$ and p is the shortest collider path between X and Y over any subsequence of $O_1, \ldots, O_{n-1}$ in $\mathcal{M}$, then path p is of definite status in $\mathcal{P}$.

proof: We prove this lemma by contradiction. For that purpose, we establish the following claim.

claim$_{II}$**:** For every $1 \leq j \leq n-1$, if O_j is not of a definite status on p in $\mathcal{P}$, then O_{j+1} is a parent of O_{j-1} in $\mathcal{M}$.

The claim holds if all the non-endpoint nodes are of definite status. However, suppose there exist at least one non-endpoint node that is not of definite status. We prove the above claim by induction. In the base case, let O_j where $1 \leq j \leq n-1$ be the first non-definite status node along p that is closest to X. Then, nodes O_{j-1} and O_{j+1} must be adjacent for otherwise we are able to detect the collider at O_j. The edge between O_{j-1} and O_{j+1} can't be bi-directed in $\mathcal{M}$ since this creates a shorter collider path over a subsequence of $O_1, \ldots, O_{n-1}$ and violates our choice of path p in lemma II. Suppose O_{j-1} is a parent of O_{j+1} ($O_{j-1} \rightarrow O_{j+1}$) in $\mathcal{M}$. The collider path between X and O_j is of definite status in $\mathcal{P}$ as O_j is the first node that is not of definite status along the path. If O_{j-2} is not adjacent to O_{j+1} then $\langle O_{j-2}, O_{j-1}, O_j, O_{j+1}\rangle$ is a discriminating path for O_j in $\mathcal{M}$ and the subpath $\langle O_{j-2}, O_{j-1}, O_j\rangle$ is a collider path in $\mathcal{P}$. Hence, O_j is oriented as a collider along the path in $\mathcal{P}$ by lemma I and that contradicts our choice of O_j which is not of definite status. Hence, O_{j-2} is adjacent to O_{j+1} and the edge is out of O_{j-2}. If the edge is bi-directed, we violate our choice of p in lemma II as shortest over $O_1, \ldots, O_{n-1}$ and if the edge is out of O_{j+1}, we violate the ancestral property in $\mathcal{M}$. We apply the previous argument by induction on all the nodes starting with O_{j-2} back to O_0. Hence, each node O_k where $0 \leq k \leq j-1$ is adjacent to O_{j+1} and the edge is out of O_k in $\mathcal{M}$. Now, X has an edge out of it and into O_{j+1} which violates our choice of p in lemma II as the shortest path. Therefore, the initial assumption about the orientation of the edge between O_{j-1} and O_{j+1} is invalid and the edge is out of O_{j+1} in $\mathcal{M}$.

In the induction step, we assume that the lemma holds for some non-definite status node O_r where $1 \leq r \leq n-1$ and we prove that the lemma holds for the next non-definite status node O_{r+l} along the path. Similar to the base case argument, the nodes O_{r+l-1} and O_{r+l+1} must be adjacent and the edge can't be bi-directed in $\mathcal{M}$. Assume that the edge is out of O_{r+l-1} in $\mathcal{M}$, then we show by induction, similar to the base case, that each node O_j where $r \leq j \leq r+l-1$ is adjacent to O_{r+l+1} and the edge is out of O_j. Now, we consider node O_r which is not of definite status and O_{r+1} is a parent of O_{r-1} in $\mathcal{M}$ by the induction step. We use the same argument by induction as before to show that each node O_k where $r+1 \leq k \leq r+l$ is adjacent to O_{r-1} and the edge is out of O_k. This creates an inducing path between O_{r-1} and O_{r+l+1} hence those two nodes must be adjacent in $\mathcal{M}$, otherwise we violate the maximality property. The edge between O_{r-1} and O_{r+l+1} can't be bi-directed in $\mathcal{M}$ as it violates the choice of the path in lemma II. If the edge is out of O_{r-1}, then we violate the ancestral property of $\mathcal{M}$ due to the structure $O_{r+l} \rightarrow O_{r-1} \rightarrow O_{r+l+1}$ and $O_{r+l} \leftarrow\!\!* O_{r+l+1}$. Similarly, the edge can't be out of O_{r+l+1}. Hence, the initial assumption that the edge between O_{r+l-1} and O_{r+l+1} is out of O_{r+l-1} is invalid and the edge is out of O_{r+l+1} in $\mathcal{M}$. Hence , the claim holds for O_{r+l}.

With the previous claim proven, we finish our proof for lemma II by establishing the following claim.

claim$^*_{II}$**:** For every $1 \leq j \leq n-1$, if O_j is not of a definite status on p in $\mathcal{P}$, then O_{j-1} is a parent of O_{j+1} in $\mathcal{M}$.

claim$^*_{II}$ is symmetric to the previous claim (**claim**$_{II}$). Hence, its proof is carried out same as the proof of the previous claim with the difference of starting with the first non-definite status node that is closest to $Y(O_n)$ and not $X(O_0)$. However, both claims being valid is contradicting as long as there exist at least one non-definite status node O_j where $1 \leq j \leq n-1$. Therefore, all the nodes along path p are of definite status. This concludes our proof of lemma II.

Now that we proved lemma II , we are ready to prove our main lemma. We choose the shortest path p^* between X and Y in $\mathcal{M}$ such that (1) all none endpoint nodes along the path are colliders, and (2) none of the directed edges, possibly the first and last edges along p^*, are visible. We show that path p^* is of definite status and none of the directed edges along the path are visible in $\mathcal{P}$.

If p^* is a single edge between X and Y, then the edge is not visible in $\mathcal{M}$ and hence it is not visible in $\mathcal{P}$ and the lemma holds. Otherwise, path p^* is a collider path with at least one non-endpoint node. It is evident that if the first or last edge along p^* is directed and not visible in $\mathcal{M}$, then those edges will not be visible in $\mathcal{P}$ if they were directed ($\rightarrow$). We use a proof similar to that of lemma II to prove that p^* is of definite status in $\mathcal{P}$ and start with the following claim.

claim$_I$**:** For every $1 \leq j \leq n-1$, if O_j is not of a definite status on p^* in $\mathcal{P}$, then O_{j+1} is a parent of O_{j-1} in $\mathcal{M}$.

We prove this claim by induction. In the base case, we choose the first node O_j that is not of definite status along the path in $\mathcal{P}$ and closest to X. Then, O_{j-1} and O_{j+1} are adjacent in $\mathcal{M}$ and the edge can't be bi-directed because of the choice of p^*. First, we assume that the edge is out of O_{j-1}. Similar to the proof of **claim**$_{II}$, each node O_k where $0 \leq k \leq j-1$ must be adjacent to O_{j+1} and the edge is out of O_k in $\mathcal{M}$. Then, X and O_{j+1} are adjacent and the edge is out of X in $\mathcal{M}$. This edge has to be visible in $\mathcal{M}$ for otherwise we violate our choice of p^*. Consider the shortest path into X consistent with the graphical condition for visibility that makes edge $X \rightarrow O_{j+1}$ visible in $\mathcal{M}$. We refer to the path as p^v and denote the nodes along p^v as $\langle C_m, \ldots C_1\rangle$ where $m \geq 1$. Note that $m = 1$ when the visibility is due to a single edge into X ($C_1 *\!\!\rightarrow X \rightarrow O_{j+1}$) and C_1 is not adjacent to O_{j+1}.

Let p' denote the concatenated path of p^v and the subpath of p^* $\langle O_0 = X, \ldots, O_{j+1}\rangle$. The following two points hold:

1. There exist at least one collider path between C_m and O_{j+1} over a subsequence of the non-endpoint nodes along p'. Note that all the non-endpoint nodes along p' are colliders in $\mathcal{M}$ except for $\langle C_1, X, O_1\rangle$ where the edge between X and O_1 can be directed out of X. If the edge between X and O_1 is bi-directed, then we are done as p' is a collider path between C_m

and O_{j+1} in $\mathcal{M}$. If the edge is directed ($X \to O_1$) in $\mathcal{M}$, we prove by induction that O_1 is the child of every node along p^v starting with C_1 until O_1 is connected with a bi-directed edge to some C_i in $\mathcal{M}$. In the base case, the edge between C_1 and X is into X so C_1 and O_1 must be adjacent for otherwise $X \to O_1$ is visible and this violates our choice of p^* where none of the edges are visible. The edge between C_1 and O_1 can't be out of O_1 for this violates the ancestral property of the MAG $\mathcal{M}$ due to $X \to O_1 \to C_1 {*\!\!\to} X$. If the edge is bi-directed, then we are done. Otherwise, the edge is out of C_1 and we continue with the induction step. In the induction step, we consider node C_i and assume that all the nodes C_j where $1 \leq j < i \leq m$ are parents of O_1. This creates a collider path between C_i and X that is into X and each collider along the path is a parent of O_1 hence $X \to O_1$ is visible in $\mathcal{M}$. Thus, C_i must be adjacent to O_1 and the edge can't be out of O_1 as this violates the ancestral property due to $C_{i-1} \to O_1 \to C_i {*\!\!\to} C_{i-1}$. Hence, the edge has to be either bi-directed or out of C_i. This concludes the proof.

2. Every collider path between C_m and O_{j+1} over a subsequence of the non-endpoint nodes along p' must go through O_j. The reason is that O_{j+1} is a child of each non-endpoint node along p' except for O_j and it is not adjacent to C_m due to the visibility of edge $X \to O_{j+1}$. O_j is the only node among the nodes in p' such that O_{j+1} is connected to and the edge is into O_j in $\mathcal{M}$. Hence, the second point is valid as well.

Given the above two points above, we choose the shortest collider path in $\mathcal{M}$ between C_m and O_{j+1} over any subsequence of the nodes along p'. By lemma II, this path is of definite status in $\mathcal{P}$. Hence, the edge between O_{j+1} and O_j is into O_j in $\mathcal{P}$. In order to prove that the edge between O_{j-1} and O_j is into O_j in $\mathcal{P}$, we use the following lemma from the work of Zhang [Zhang, 2006].

lemma III: 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, $A {*\!\!\to} C$.

Using lemma III, if the edge between O_{j-1} and O_j has a circle incident on O_j in $\mathcal{P}$, then the edge between O_{j-1} and O_{j+1} has an arrowhead incident on O_{j-1} in $\mathcal{P}$. However, this is not possible as the edge between O_{j-1} and O_{j+1} is out of O_{j-1} in $\mathcal{M}$ according to the assumption in the base case of the proof of the **claim**$_I$. Since $\mathcal{P}$ is the PAG representing the equivalence class of MAG $\mathcal{M}$, then $\mathcal{P}$ can't have an arrowhead incident on O_{j-1} from O_{j+1} as this orientation is not invariant in all the MAGs in the equivalence class. Hence, the edge between O_{j-1} and O_j can't have a circle incident on O_j and the edge will be into O_j in $\mathcal{P}$. Having O_j a definite collider along $\langle O_{j-1}, O_j, O_{j+1}\rangle$ contradicts our assumption in the base case that O_j is the first node that is not of definite status along p^* in $\mathcal{P}$. Hence, the initial assumption that the edge between O_{j-1} and O_{j+1} is out of O_{j-1} in $\mathcal{M}$ is invalid and the edge must be out of O_{j+1}.

In the induction step, we apply exactly the same argument as that applied in the induction step of the proof of **claim**$_{II}$. We omit the proof to avoid redundancy. This concludes the proof of claim$_I$. Next, we prove the following claim which is symmetric to **claim**$_I$.

claim$_I^*$: For every $1 \leq j \leq n-1$, if O_j is not of a definite status on p^* in $\mathcal{P}$, then O_{j-1} is a parent of O_{j+1} in $\mathcal{M}$.

Again, the proof of **claim**$_I^*$ is exactly the same as the proof of **claim**$_I$ with the difference of starting the induction with the first node that is not of definite status along p^* and is closest to Y instead of X. However, having both claims valid is contradicting as long as there exist at least one non-definite status node along p^* in $\mathcal{P}$. Hence, path p^* is of definite status and any directed edge along the path of not visible. This concludes our proof for the main lemma. □

Lemma 8. *Let $\mathcal{M}$ be any MAG over a set of variables $\mathbf{V}$. If A and B are in the same pc-component in $\mathcal{M}$, then there is a DAG in the equivalence class of $\mathcal{M}$ where A and B are in the same c-component.*

Proof. If A and B are in the same pc-component in MAG $\mathcal{M}$, then one of the following must be true:

1. A and B are adjacent in $\mathcal{M}$ and the edge is not visible: A DAG is constructed according to lemma 13 where A and B share a latent parent L_{AB}.
2. A collider path exists between A and B where no directed edge is visible ($A {*\!\!\to} O_1 \leftarrow - \to O_n \leftarrow\!\!* B$): If $A {*\!\!\to} O_1$ or $O_n \leftarrow\!\!* B$ is bi-directed in $\mathcal{M}$, then the construction in lemma 13 creates a DAG where A and B share a bi-directed path and hence are in the same c-component.

   The other case is when both $A {*\!\!\to} O_1$ and $O_n \leftarrow\!\!* B$ are directed invisible edges. If A and B share a bi-directed path in $\mathcal{M}$, then the previous construction puts A and B in the same c-component in a DAG $\mathcal{G}$. Otherwise, the conditions of lemma 10 are satisfied and we construct a DAG $\mathcal{G}$ in the equivalence class of $\mathcal{M}$ where $A \to O_1$ and $O_n \leftarrow B$ are confounded simultaneously. Hence, A and B are in the same c-component in $\mathcal{G}$.

□

Lemma 9. *Let $\mathcal{P}$ be a PAG over $\mathbf{V}$, and let A and B be two nodes that are in the same pc-component in $\mathcal{P}$. There is a MAG in the equivalence class of $\mathcal{P}$ where A and B are in the same pc-component if A and B in $\mathcal{P}$ are adjacent or in different buckets (i.e. no circle path between them).*

Proof. Under the conditions of the lemma, we have the following options for A and B to be in the same pc-component:

1. A and B are adjacent and the edge not fully oriented, bi-directed, or directed and not visible in $\mathcal{P}$: If the edge is bi-directed in $\mathcal{P}$, then it is bi-directed in every MAG $\mathcal{M}$. Otherwise, the edge is oriented in some MAG into an invisible directed edge by lemma 22.
2. A and B are in different circle components of PAG $\mathcal{P}$ (different buckets), they share a collider path ($A{*}\!\!\rightarrow O_1 \leftarrow - - \rightarrow O_n \leftarrow\!\!* B$), and any directed edge is not visible. By lemma 21, we can construct a MAG $\mathcal{M}$ in the equivalence class of $\mathcal{P}$ such that we have no new edges into A nor into B. This extension to the lemma is possible since A and B correspond to two separate circle components in $\mathcal{P}$ according to the condition in the lemma. Hence, the orientations can be done independently in each circle component. Therefore, both edges $A \rightarrow O_1$ and $O_n \leftarrow B$ are invisible in $\mathcal{M}$ by lemma 22. The latter is correct as the same argument in the proof of lemma 22 can be made for both edges simultaneously. Thus, we have a collider path between A and B in MAG $\mathcal{M}$ ($A \rightarrow O_1 \leftarrow - - \rightarrow O_n \leftarrow B$) and the first and last edges are not visible.

□

Lemma 10. *Let $\mathcal{M}$ be any MAG over a set of variables $\mathbf{V}$, and $A \rightarrow B$ and $C \rightarrow D$ be any two invisible directed edges in $\mathcal{M}$ such that $A \neq C$. If there is no bi-directed path between A and C in $\mathcal{M}$, then there is a DAG whose MAG is $\mathcal{M}$ in which A and B share a latent parent and C and D share another latent parent simultaneously.*

Proof. Under the condition in the lemma, construct a DAG $\mathcal{G}$ from $\mathcal{M}$ using the following steps:

1. Leave every directed edge in $\mathcal{M}$ as it is.
2. Replace every bi-directed edge $U \leftrightarrow W$ in $\mathcal{M}$ with a latent variable L_{UW} between U and W.
3. Introduce two additional latent variables: L_{AB} between A and B and L_{CD} between C and D.

In order to prove the lemma, we need to show that $\mathcal{M}_{\mathcal{G}} = \mathcal{M}$ where $\mathcal{M}_{\mathcal{G}}$ is the MAG corresponding to DAG $\mathcal{G}$. Hence, we have the following exhaustive options given any pair of variables X and Y:

1. $X \rightarrow Y$ in $\mathcal{M}$: DAG $\mathcal{G}$ retains the same directed edges in $\mathcal{M}$, so the same edge is present in $\mathcal{M}_{\mathcal{G}}$.
2. $X \leftarrow Y$ in $\mathcal{M}$: Same as the previous case.
3. $X \leftrightarrow Y$ in $\mathcal{M}$: $\mathcal{G}$ replaces every bi-directed edge in $\mathcal{M}$ with a latent variable which in turn is an inducing path between X and Y in $\mathcal{G}$. Since the construction of $\mathcal{G}$ doesn't introduce any new ancestral relations between variables other than the ones already present in $\mathcal{M}$, then $X \leftrightarrow Y$ is present in $\mathcal{M}_{\mathcal{G}}$.
4. X and Y are not adjacent in $\mathcal{M}$: In this case, we need to show that there is no inducing path between X and Y in $\mathcal{G}$. In what follows, we prove this result using contradiction.

Suppose that there is an inducing path between X and Y in $\mathcal{G}$ while they are not adjacent in $\mathcal{M}$. Let p be such a path with the smallest number of observed variables in $\mathcal{G}$ and let $\langle X, O_1, \ldots, O_n, Y\rangle$ denote the sequence of observed nodes along the path. Since the construction of $\mathcal{G}$ doesn't introduce new ancestral relations than the ones in $\mathcal{M}$, then each node O_i must be an ancestor of X or Y in $\mathcal{M}$ as well as in $\mathcal{G}$. If a maximum of one of the edges $A \rightarrow B$ and $C \rightarrow D$ is along the inducing path p, then the proof by contradiction follows exactly as that of lemma 5.1.2 in [Zhang, 2006].

It remains to consider the case where both edges are along the inducing path p. The condition in the lemma states that $A \neq C$ and there is no collider path between A and C in $\mathcal{M}$. So by the construction of $\mathcal{G}$, we can't have $B \leftarrow A \leftarrow - - \rightarrow C \rightarrow D$ as a sub-path of p with zero or more intermediate observed variables. Hence, both edges ($A \rightarrow B$ and $C \rightarrow D$) can exist along p simultaneously if we have one of the following as sub-path of p:

1. $A \rightarrow B \leftarrow - - \rightarrow D \leftarrow C$: Without loss of generality, suppose that X is closest to A along the path and Y is closest to C. If $X = A$ and $Y = C$, then both $A \rightarrow B$ and $C \rightarrow D$ are into $O_1 = B$ and $O_n = D$ in $\mathcal{M}$, respectively. As for the intermediate edges along p, they correspond to latent variables $L_{O_iO_{i+1}}$ for $1 < i < n-1$. Those latent variables correspond to bi-directed edges in $\mathcal{M}$ by construction of $\mathcal{G}$. Therefore, an inducing path between X and Y is present in $\mathcal{M}$. This is not possible as it contradicts with the maximality property of a MAG.

   If $X \neq A$ or $Y \neq C$, at least one of the two edges is not the first along p. Without loss of generality, suppose that $A \rightarrow B$ is one such edge. The edge between X and O_1 in $\mathcal{M}$ must be into O_1 as it is into O_1 in $\mathcal{G}$. Also, every intermediate edge between O_1 and $O_i = A$ in $\mathcal{M}$ must be bi-directed as it corresponds to a latent variable in $\mathcal{G}$. We argue by induction that every node O_j $j < i$ is a parent of B in $\mathcal{M}$. In the base case, O_{i-1} is adjacent to B or else $A \rightarrow B$ is visible. The edge between O_{i-1} and B is not out of B as it violates the ancestral property of $\mathcal{M}$ due to $A \rightarrow B \rightarrow O_{i-1}$ and $O_{i-1} \leftrightarrow O_i$. Also, the edge is not bi-directed as this would lead to an inducing path in $\mathcal{G}$ shorter than p. In the inductive step, assume all the nodes starting with O_{i-1} till O_j are parents of B. Then, O_{j+1} must be adjacent to B, else $A \rightarrow B$ is visible. Also, the edge has to be out of O_{j+1} for the same reasons as before. This implies that X is adjacent to B and the edge is into B in $\mathcal{M}$. Again, we have an inducing path between X and Y in $\mathcal{G}$ that has fewer number of observed variables than p which is a contradiction.

2. $A \to B \leftarrow - \to C \to D$: Again, without loss of generality, suppose that X is closest to A along the path and Y is closest to D. If $X \neq A$, we can use the same proof by induction as before to show that every node along p starting with X until A is a parent of B in $\mathcal{M}$. Hence, we have an inducing path in $\mathcal{G}$ with fewer observed variables than p which is a contradiction.

   If $X = A$, then we have a collider path into C starting with $X = A$ in $\mathcal{M}$. Again, every node along this path is a parent of D by the induction earlier. Hence, X is a parent of D in $\mathcal{M}$ and we have a shorter inducing path than p which is a contradiction.

3. $B \leftarrow A \leftarrow - \to D \leftarrow C$: The argument for this case is the same as the previous case.

The latter three cases all arrive to contradictions, thus a pair of variables that is not adjacent in $\mathcal{M}$ is not adjacent in $\mathcal{M}_{\mathcal{G}}$ as well. Therefore, $\mathcal{M} = \mathcal{M}_{\mathcal{G}}$. □

Lemma 11. *Given a PAG $\mathcal{P}$ and a corresponding PTO $\mathbf{B_1} < \mathbf{B_2} < \cdots < \mathbf{B_m}$ over $\mathcal{P}$, $\mathbf{B_i}$ is independent of $\boldsymbol{\mathcal{H}} \subseteq \mathbf{B^{(i-1)}}$ given $\mathbf{B^{(i-1)}} \setminus \boldsymbol{\mathcal{H}}$, i.e. ($\mathbf{B_i} \perp\!\!\!\perp \boldsymbol{\mathcal{H}} | \mathbf{B^{(i-1)}} \setminus \boldsymbol{\mathcal{H}}$), if:*

1. *$\nexists(v_i \in \mathbf{B_i} \wedge v_j \in \boldsymbol{\mathcal{H}} \wedge v_i *\!\!-\!\!* v_j)$, i.e. there is no path of size one connecting $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$, and*
2. *Each proper definite status path between $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$:*
   (a) Contains at least one definite non-collider not in $\boldsymbol{\mathcal{H}}$, or
   (b) The path is not into $\mathbf{B_i}$.

Proof. Any subset of $\mathbf{B^{(i-1)}}$ refers to a subset of the buckets represented by $B^{(i-1)}$. The first condition is necessary as it is not possible to separate two sets of nodes in a PAG if there exist one node in $\mathbf{B_i}$ ($v_i \in \mathbf{B_i}$) and another in $\boldsymbol{\mathcal{H}}$ ($v_j \in \boldsymbol{\mathcal{H}}$) such that v_i and v_j are adjacent.

We need to block the definite status paths between any $v_i \in \mathbf{B_i}$ and $v_j \in \boldsymbol{\mathcal{H}}$ for $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$ to be independent. But, any definite status path between v_i and v_j includes one of the proper definite status paths between $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$ as a subpath. Hence, we only consider blocking the proper definite status paths between $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$ since this will be sufficient to block all the definite status paths between any $v_i \in \mathbf{B_i}$ and $v_j \in \boldsymbol{\mathcal{H}}$. For simplicity, we assume that the path starts in $\mathbf{B_i}$ and ends in $\boldsymbol{\mathcal{H}}$.

Consider any proper definite status path between $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$; $p = \langle O_1, \ldots, O_n \rangle$ where $n > 2$, $O_1 \in \mathbf{B_i}$, $O_n \in \boldsymbol{\mathcal{H}}$, and $O_2, \ldots, O_{n_1} \notin \mathbf{B_i} \cup \boldsymbol{\mathcal{H}}$.

case a: The path contains at least one definite non-collider not in $\boldsymbol{\mathcal{H}}$ referred to as $O_l \in \mathbf{B_k}$. If $k < i$, then the path is blocked as $O_l \in \mathbf{B^{(i-1)}} \setminus \boldsymbol{\mathcal{H}}$. If $k > i$, then O_l is in a bucket that comes after $\mathbf{B_i}$ in the BPAG. We show that at least one of the subpaths $p_1 = \langle O_1, \ldots, O_l \rangle$ and $p_2 = \langle O_l, \ldots, O_n \rangle$ must contain a collider and this collider is in a bucket $\mathbf{B_p}$ where $p > k > i$. Note that path p is of definite status so each non-endpoint node along the path should be a definite collider or a definite non-collider (DNC).

DNC 1 Consider one of the edges of O_l along path p to be out of O_l. Without loss of generality, we consider the target node of the edge to be O_{l+1} ($O_l \to O_{l+1}$). A bucket only contains circle edges, so O_{l+1} can't be in the same bucket as O_l. Hence, it must be in another bucket $\mathbf{B_p}$ where $p > k$ since the edge is directed.

DNC 2 The DNC is such that $O_{l-1} *\!\!-\!\!\circ O_l \circ\!\!-\!\!* O_{l+1}$ and O_{l-1} and O_{l+1} are not adjacent: If at least one of the two edges has an arrowhead on the other end, then the target node would be in a bucket $\mathbf{B_p}$ where $p > k$ for a similar argument as in **DNC 1**. If both edges have circles on the other ends ($O_{l-1} \circ\!\!-\!\!\circ O_l \circ\!\!-\!\!\circ O_{l+1}$), then the three nodes reside in the same bucket. Only a DNC with one of the edge $\to$ or $\circ\!\!\to$ would preserve the definite status property of path p and extend it to a node outside the current bucket. Again, that target node will be in a bucket $\mathbf{B_p}$ where $p > k$.

Therefore, a sequence of DNCs along path p starting with O_l keeps extending path p within $\mathbf{B_k}$ or to buckets ahead of it relative to the PTO. Only a definite collider would connect the path back to $\mathbf{B_i}$ or $\boldsymbol{\mathcal{H}}$. Note that the definite collider will be in a bucket $\mathbf{B_p}$ where $p > i$ and all the possible descendants of the collider will be in buckets including and after $\mathbf{B_p}$. Hence, this collider blocks the path since no possible descendants is in $\mathbf{B_i}$ or $\mathbf{B^{(i-1)}}$.

case b: If the path is not into $\mathbf{B_i}$, then the edge mark incident on O_1 is either a tail ($O_1 \to O_2$) or a circle ($O_1 \circ\!\!\to O_2$). If a circle is incident on O_1, then the other end of the edge has to be an arrowhead ($O_1 \circ\!\!\to O_2$) since a circle would put O_2 in the same bucket as O_1 and the path would not be proper anymore. In both cases, O_2 is in a bucket after $\mathbf{B_i}$. The argument for the existence of a collider after $\mathbf{B_i}$ along path p follows similar to case a. Hence, path p is blocked as well.

Therefore, the set $\mathbf{B^{(i-1)}} \setminus \boldsymbol{\mathcal{H}}$ blocks all the proper paths between $\mathbf{B_i}$ and $\boldsymbol{\mathcal{H}}$. This concludes the proof of the lemma. □

Lemma 12. *Let $G(\mathbf{V}, \mathbf{L})$ and $G'(\mathbf{V}', \mathbf{L}')$ be two DAGs with the same skeleton such that $\mathbf{V}' = \mathbf{V} \setminus \mathbf{Z}$, $\mathbf{L}' = \mathbf{L} \cup \mathbf{Z}$, $\mathbf{Z} \subset \mathbf{V}$. $P_{\mathbf{x}}(\mathbf{y})$ is identifiable in $\mathcal{G}$ if it is identifiable in $\mathcal{G}'$.*

Proof. Let P be a distribution that factorizes over $\mathcal{G}$. We marginalize out subset $\mathbf{Z}$ from distribution P ($P' = \sum_{\mathbf{Z}} P$). Now P' factorizes over DAG $\mathcal{G}'$ where the nodes in $\mathbf{Z}$ are latent. If $P'_{\mathbf{x}}(\mathbf{y})$ is identifiable in $\mathcal{G}'$, then it is computable from P and $P_{\mathbf{x}}(\mathbf{y})$ is identifiable in $\mathcal{G}$. □

B Background Results

In this section, we list some results from the literature that are used throughout the proofs.

Lemma 13 (Lemma 5.1.2 in [Zhang, 2006]). *Let $\mathcal{M}$ be any MAG over a set of variables O, and $A \to B$ be any directed edge in $\mathcal{M}$. If $A \to B$ is invisible in $\mathcal{M}$, then there is a DAG whose MAG is $\mathcal{M}$ in which A and B share a latent parent, i.e., there exists a latent variable L_{AB} in the DAG.*

Lemma 14 (Lemma 3.3.2 in [Zhang, 2006]). *In a PAG $\mathcal{P}$, for any two nodes A and B, if there is a circle path between A and B, i.e. a path consisting of ∘–∘ edges, then:*

1. *if there is an edge between A and B, then the edge is not into A or B, i.e. A ∘–∘ B in the absence of selection bias.*
2. *for any other node C, $C{*}\!\!\to A$ if and only if $C{*}\!\!\to B$. Furthermore, $C \leftrightarrow A$ if and only if $C \leftrightarrow B$.*

Lemma 15 (Lemma 20 in [Zhang, 2008a]). *Let G be any DAG over $\mathbf{O} \cup \mathbf{L}$, and M be the MAG of G over $\mathbf{O}$. For any $A, B \in \mathbf{O}$ and $\mathbf{C} \subseteq \mathbf{O}$ that does not contain A or B, there is a path d-connecting A and B given $\mathbf{C}$ in G if and only if there is a path m-connecting A and B given $\mathbf{C}$ in M.*

Lemma 16 (Lemma 26 in [Zhang, 2008a]). *Let M be a MAG over $\mathbf{O}$, and P be the PAG that represents the equivalence class of M. For any $A, B \in \mathbf{O}$ and $\mathbf{C} \subseteq \mathbf{O}$ that does not contain A or B, if there is a path m-connecting A and B given $\mathbf{C}$ in M, then there is a path definitely m-connecting A and B given $\mathbf{C}$ in P.*

Lemma 17 (Lemma B.1 in [Zhang, 2008b]). *If $p = \langle A, \dots, B \rangle$ is a possibly directed path from A to B in PAG $\mathcal{P}$, then some subsequence of p forms an uncovered possibly directed path from A to B in $\mathcal{P}$.*

Lemma 18 (Lemma B.2 in [Zhang, 2008b]). *If p is an uncovered possibly directed path from A to B in PAG $\mathcal{P}$, then*

1. *if there is an ∘→ edge on p, then any ∘–∘ edge on p is before that edge, and any → edge on p is after that edge; and*
2. *there is at most one ∘→ edge on p.*

Lemma 19 (Theorem 2 in [Zhang, 2008b]). *Let $\mathcal{M}$ be the MAG resulting from the following procedure applied to PAG $\mathcal{P}$:*

1. *orient the circles on ∘→ edges in $\mathcal{P}$ as tails; and*
2. *orient the circle components of $\mathcal{P}$ into a DAG with no unshielded colliders.*

Then $\mathcal{M}$ is in the equivalence class of $\mathcal{P}$.

Lemma 20 (Lemma 7.5 in [Maathuis and Colombo, 2015]). *Let X and Y be two distinct nodes in a PAG $\mathcal{P}$. Then $\mathcal{P}$ cannot have both a possibly directed path from X to Y and an edge of the form $Y{*}\!\!\to X$.*

Lemma 21 (Lemma 7.6 in [Maathuis and Colombo, 2015]). *Let X be a node in a PAG $\mathcal{P}$. Let $\mathcal{M}$ be the MAG resulting from the following procedure applied to a $\mathcal{P}$:*

1. *replace all partially directed edges (∘→) in $\mathcal{P}$ with directed edges (→), and*
2. *orient the subgraph of $\mathcal{P}$ consisting of all circle edges (∘–∘) into a DAG with no unshielded colliders and no new edges into X.*

Then, $\mathcal{M}$ is in the Markov equivalence class of $\mathcal{P}$.

Lemma 22 (lemma B.1 in [Perković *et al.*, 2016]). *Given a PAG $\mathcal{P}$ and X a node in $\mathcal{P}$. Let MAG $\mathcal{M}$ be in the equivalence class of $\mathcal{P}$ and satisfy the construction in lemma 21. Then, the edge X ∘–∘ Y, X∘→ Y, or invisible $X \to Y$ in $\mathcal{P}$ is invisible $X \to Y$ in $\mathcal{M}$.*

The following Definitions, Theorem, and lemma summarize the generalized adjustment criterion introduced in [Perković *et al.*, 2016] and introduced necessary results relevant to our work.

Definition 8 (Amenability). *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in a MAG or PAG $\mathcal{G}$. Then $\mathcal{G}$ is said to be amenable relative to ($\mathbf{X}$,$\mathbf{Y}$) if every possibly directed proper path from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$ starts with a visible edge out of $\mathbf{X}$.*

Definition 9 (Forbidden set; Forb($\mathbf{X}$,$\mathbf{Y}$,$\mathcal{G}$)). *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in a MAG or PAG $\mathcal{G}$. Then the forbidden set relative to ($\mathbf{X}$,$\mathbf{Y}$) is defined as the set of nodes that are possible descendants of nodes $W \notin \mathbf{X}$ that lie along proper possibly directed paths from $\mathbf{X}$ to $\mathbf{Y}$ in $\mathcal{G}$.*

Definition 10 (Generalized adjustment criterion). *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a MAG or PAG $\mathbf{G}$. Then $\mathbf{Z}$ satisfies the generalized adjustment criterion relative to ($\mathbf{X}$,$\mathbf{Y}$) in $\mathbf{G}$ if the following three conditions hold:*

*(**Amenability**) $\mathbf{G}$ is adjustment amenable relative to ($\mathbf{X}$,$\mathbf{Y}$), and*

*(**Forbidden set**) $\mathbf{Z} \cap$ Forb($\mathbf{X}$,$\mathbf{Y}$,$\mathcal{G}$) $= \phi$, and*

*(**Blocking**) all proper definite status non-causal paths from $\mathbf{X}$ to $\mathbf{Y}$ are blocked by $\mathbf{Z}$ in $\mathcal{G}$.*

Theorem 4. *Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint node sets in a MAG or PAG $\mathbf{G}$. Then $\mathbf{Z}$ is an adjustment set relative to ($\mathbf{X}$,$\mathbf{Y}$) in $\mathbf{G}$ if and only if $\mathbf{Z}$ satisfies the generalized adjustment criterion relative to ($\mathbf{X}$,$\mathbf{Y}$) in $\mathbf{G}$.*

Definition 11 (Adjust($\mathbf{X}$,$\mathbf{Y}$,$\mathcal{G}$))**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in a MAG or PAG $\mathcal{G}$. We define Adjust($\mathbf{X}$,$\mathbf{Y}$,$\mathcal{G}$) to be the set of possible ancestors of $\mathbf{X}$ and $\mathbf{Y}$ excluding $\mathbf{X}$, $\mathbf{Y}$, and Forb($\mathbf{X}$,$\mathbf{Y}$,$\mathcal{G}$).*

Lemma 23 (corollary 4.4 in [Perković *et al.*, 2016])**.** *Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint node sets in a MAG or PAG $\mathcal{G}$. There exists an adjustment set relative to ($\mathbf{X}$,$\mathbf{Y}$) in $\mathbf{G}$ if and only if Adjust($\mathbf{X}$,$\mathbf{Y}$,$\mathcal{G}$) satisfies the generalized adjustment criterion relative to ($\mathbf{X}$,$\mathbf{Y}$) in $\mathbf{G}$.*

Lemma 24 (Lemma B.6 in [Perković *et al.*, 2016])**.** *Let $\mathbf{X}$,$\mathbf{Y}$ and $\mathbf{Z}$ be pairwise disjoint node sets in a PAG $\mathcal{P}$ and let $\mathcal{M}$ be a MAG in the equivalence class of $\mathcal{P}$. Let $\mathbf{Z}$ satisfy the amenability condition and the forbidden set condition relative to ($\mathbf{X}$,$\mathbf{Y}$) in $\mathcal{P}$. If there is a proper non-causal paths from $\mathbf{X}$ to $\mathbf{Y}$ that is m-connecting given $\mathbf{Z}$ in $\mathcal{M}$, then there is a proper definite status non-causal path from $\mathbf{X}$ to $\mathbf{Y}$ that is m-connecting given $\mathbf{Z}$ in $\mathcal{P}$.*

References

[SuppZhang, 2006] Jiji Zhang. *Causal inference and reasoning in causally insufficient systems*. PhD thesis, Carnegie Mellon University, 2006.