Estimating Possible Causal Effects with Latent Variables via Adjustment

Tian-Zuo Wang [1] **Tian Qin** [1] **Zhi-Hua Zhou** [1]

Abstract

Causal effect identification is a fundamental task in artificial intelligence. A most ideal scenario for causal effect identification is that there is a *directed acyclic graph* as a prior causal graph encoding the causal relations of all relevant variables. In real tasks, however, the prior causal graph is usually not available, and some relevant variables may be latent as well. With observational data, we can only learn a partial ancestral graph (PAG), which contains some indeterminate causal relations. Since many causal graphs can correspond to one PAG, they are possibly associated with different causal effects. The aim of this paper is to estimate these possible causal effects via covariate adjustment given a PAG. This task is challenging because the number of causal graphs corresponding to a PAG grows super-exponentially with the number of variables. We propose a new graphical characterization for possible adjustment sets, and based on this, we develop the first method to determine the set of possible causal effects that are consistent with the given PAG *without* enumerating any causal graphs. Our method can output the same set as the enumeration method with *super-exponentially less* complexity. Experiments validate the effectiveness and tremendous efficiency improvement of the proposed method.

1. Introduction

Decision is a fundamental task throughout artificial intelligence, economics, and social sciences. Causal effect identification is inherently linked to decision-making, as it answers 'what will happen to outcome Y if X is set to x'. There exists a growing literature studying causal effect identification given a causal graph (Tian & Pearl, 2002; Shpitser & Pearl, 2006; Pearl, 2009; Shpitser et al., 2010; Jung et al., 2021). A practical method to estimate causal effect is via *(covariate) adjustment*, *i.e.,* estimating it by adjusting for some observed variables called *adjustment set* (van der Zander et al., 2014; Maathuis et al., 2015; Perkovic et al., 2017; van der Zander et al., 2019).

A most ideal scenario for causal effect identification is that there is a *directed acyclic graph (DAG)* as a prior causal graph which encodes the causal relations among *all the relevant variables*. In real tasks, however, such knowledge is often not available. Moreover, some relevant variables in the DAGs may be latent as well. For example, macroeconomic policy influences many factors in business, but it is hard to quantify. When there is not a prior causal graph, a common practice is to learn the causal graph at first.

In the presence of latent variables, a *partial ancestral graph (PAG)* can be learned with observational data (Spirtes et al., 2000; Ali et al., 2005; Zhang, 2008a). A PAG represents a Markov equivalence class (MEC) of *maximal ancestral graphs (MAG)* which encode the causal relations among *the observed variables*. Roughly speaking, a MAG is a "projection graph" of an underlying DAG containing *all relevant variables*[1]. Notably, many DAGs, even infinite ones, can be projected to the same MAG. Henceforth, any DAG that can be projected to a MAG in the MEC represented by a given PAG is considered *a DAG corresponding to the PAG*. An example of the three types of graphs is given in Fig. 1.

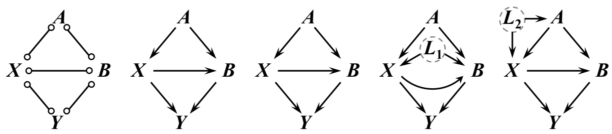
There are some studies (Entner et al., 2013; Hyttinen et al., 2015; Jaber et al., 2022) on the causal effect identifiability in a PAG, *i.e.*, whether all the DAGs corresponding to the PAG are associated with the *same causal effect* and can be estimated with observational data. Generally, due to the limited information of PAG, the causal effect is unidentifiable, in which case there can be many possible causal effects.

In light of the fact that the causal effect of some variable X on variable Y is not always identifiable in a PAG, this paper aims to determine the set of possible causal effects of X on Y obtainable via covariate adjustment, *i.e.*, the set includes the causal effects that can be identified via covariate adjustment in any DAGs corresponding to the PAG. When the causal effect is not identifiable, the set can provide some valuable information as well. Moreover, it requires neither

[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, 210023, China. Correspondence to: Zhi-Hua Zhou <zhouzh@lamda.nju.edu.cn>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

[1]The "projection" is elaborated in Appendix A.2



(a) PAG (b) MAG (c) DAG $\mathcal{D}_1$ (d) $\mathcal{D}_2$ (e) $\mathcal{D}_3$

Figure 1: Fig. 1(a): PAG $\mathcal{P}$. Fig. 1(b): MAG $\mathcal{M}$ in the MEC represented by $\mathcal{P}$. Fig. 1(c)/ 1(d)/ 1(e): DAG $\mathcal{D}_1$/$\mathcal{D}_2$/$\mathcal{D}_3$ that is projected to $\mathcal{M}$ over the observed variables $\{A, B, X, Y\}$. L_1, L_2: latent variables. As the number of latent variables is arbitrary, there can be infinite DAGs projected to $\mathcal{M}$. According to previous study (Appendix A.4, Def. 8), $f(Y|do(X))$ is identifiable via adjustment in $\mathcal{D}_1$ and $\mathcal{D}_3$ but not in $\mathcal{D}_2$, thus $f(Y|do(X))$ is not identifiable in $\mathcal{M}$ and $\mathcal{P}$.

prior structural knowledge nor experimental data, thus is *always feasible in practice*. Note that there are possibly some DAGs corresponding to the PAG such that the causal effect in the DAGs cannot be identified with observational data via covariate adjustment. We do not try to return such causal effects since it is usually beyond the ability of current observations when latent variables exist. To consistently identify them, it is necessary to observe more variables.

The primary challenge in set determination lies in the large number of DAGs corresponding to a PAG. With observational data, we can only learn a PAG, which can represent a super-exponential number of MAGs and each MAG can be a projection graph of infinite DAGs. The enumerations of either MAGs or DAGs are computationally prohibitive. In this paper, we propose the first method to determine the **PAG**-consistent **caus**al **e**ffect **s**et without any enumerations of DAGs or MAGs, called PAGcauses for short. The method relies on two main theoretical results. One is a graphical characterization for adjustment sets comprised of observed variables in all the DAGs which can be projected to a given MAG, through which we can circumvent the enumerations of DAGs and find adjustment sets based on mere MAGs. The other is a graphical characterization for adjustment sets in all the MAGs, through which we can determine whether each set of vertices could be an adjustment set without the enumerations of MAGs. The complexity of our method is super-exponentially less than the method based on (local) enumerations (Malinsky & Spirtes, 2016). Further, we prove that PAGcauses can output the set of possible causal effects equal to the direct method to enumerate all the DAGs where the causal effect can be identified via covariate adjustment with observational data. Experiments verify the effectiveness and tremendous efficiency improvement.

2. Related Work

In the literature, there are some classical methods to determine the set of possible causal effects. Maathuis et al. (2009) made significant contributions by presenting the first related method based on CPDAG under the assumption of no latent variables. Some further studies are also conducted by introducing background knowledge or reducing the variance (Fang et al., 2022; Henckel et al., 2022). Guo & Perkovic (2021) proposed an algorithm to list all the possible causal effects that take polynomial time per output instance. To relax the assumption of no latent variables, Malinsky & Spirtes (2016) proposed the first method to take latent variables into account by enumerating MAGs in a subspace of MEC. However, only considering MAGs may ignore some causal effects, which is detailed in Sec. 4.1, and is with a super-exponential complexity. Cheng et al. (2020) presented an efficient method assuming that all the variables except for $\{X, Y\}$ are not descendants of X. Due to the page limit, we give a detailed discussion about the differences between our paper and related studies in Appendix B.

Decision-making is a crucial task in real-world applications, and identifying causal effects is often considered as a vital aspect (Entner et al., 2013; Perkovic et al., 2017; Jaber et al., 2018; Chakrabortty et al., 2018; Jaber et al., 2019; Wang et al., 2020; Qin et al., 2021; Jaber et al., 2022; Pfister & Peters, 2022; Yata, 2021). However, causal effect identification requires accurate causal relations. Causal discovery, which involves learning causal relations, has attracted significant attention in this regard (Peters et al., 2014; Rothenhäusler et al., 2015; Zhang et al., 2017; Cai et al., 2018; Yu et al., 2019). Apart from methods based on causality, there are also approaches to decision-making based on bandit or MDP (Lee & Bareinboim, 2018; Zhao et al., 2021; 2022).

3. Preliminary

A graph $G = (\mathbf{V}, \mathbf{E})$ consists of a set of vertices $\mathbf{V} = \{V_1, \cdots, V_p\}$ and a set of edges $\mathbf{E}$. For any subset $\mathbf{V}' \subseteq \mathbf{V}$, the *subgraph induced by* $\mathbf{V}'$ is $G[\mathbf{V}'] = (\mathbf{V}', \mathbf{E}_{\mathbf{V}'})$, where $\mathbf{E}_{\mathbf{V}'}$ is the set of edges in $\mathbf{E}$ whose both endpoints are in $\mathbf{V}'$. For a graph G, $\mathbf{V}(G)$ denotes the set of vertices in G. G is a complete graph if there is an edge between any two vertices. The subgraph induced by an empty set is trivially a complete graph. $G[-\mathbf{V}']$ denotes the subgraph induced by $\mathbf{V}(G)\backslash\mathbf{V}'$. Usually, bold letter (*e.g.*, $\mathbf{V}$) denotes a set of vertices and normal letter (*e.g.*, V) denotes a vertex.

Due to the page limit, we show some graph-related definitions in Appendix A.1, including mixed, partial mixed graph (PMG), adjacent, parent, child, path, directed path, possible directed path, ancestor, possible ancestor, descendant, possible descendant, directed cycle, almost directed cycle, collider, collider path, minimal path, minimal collider path, unshielded, uncovered path, minimal possible directed path, DAG (denoted by $\mathcal{D}$), active path, m-separated, ancestral graph, maximal graph, MAG (denoted by $\mathcal{M}$), inducing path, discriminating path, visible, $\mathcal{M}_{\underline{X}}$, and $\mathcal{P}_{\underline{X}}$.

We denote the set of *parents/children/ancestors/possible ancestors/descendants/possible descendants* of V_i in G by Pa(V_i, G)/Chd(V_i, G)/Anc(V_i, G)/PossAn(V_i, G)/De(V_i, G)/ PossDe(V_i, G). $\bigoplus$ denotes concatenation of paths. For a path p, let $p[V_i, V_j]$ denote the sub-path of p from V_i to V_j.

The two ends of an edge are called *marks* and have two types *arrowhead* or *tail*. The *circle* ($\circ$) in a graph implies that the mark here could be either arrowhead or tail but is indeterminate. $*$ is a wildcard that denotes any of arrowhead, tail, or circle. We make a convention that if an edge is $\circ\!\!-\!\!*$, the $*$ here cannot be a tail for otherwise the circle can be replaced by an arrowhead due to the assumption of no selection bias (there is not an edge with two tails). An edge $V_i \circ\!\!-\!\!\circ V_j$ is a *circle edge*. The *circle component* in G is the subgraph consisting of all the $\circ\!\!-\!\!\circ$ edges in G. Two vertices V_i and V_j are in a *connected circle component* in G if there is a path comprised of circle edges from V_i to V_j in G. Two MAGs are *Markov equivalent* if they have the same m-separations. A class comprised of all the Markov equivalent MAGs is a *Markov equivalence class (MEC)*. A *partial ancestral graph (PAG, denoted by $\mathcal{P}$)* represents an MEC, where a tail/arrowhead occurs if the corresponding mark is tail/arrowhead in all the Markov equivalent MAGs, and a circle occurs otherwise. A MAG $\mathcal{M}$ is *consistent with a PAG $\mathcal{P}$* if $\mathcal{M}$ belongs to the MEC represented by $\mathcal{P}$. A MAG represents the conditional independences and causal relations over the observed variables in a DAG. The construction of a MAG based on a DAG is shown in Appendix A.2. We say $\mathcal{D}$ *is represented by (or can be projected to)* $\mathcal{M}$ if $\mathcal{M}$ can be obtained from $\mathcal{D}$ by the construction process. And $\mathcal{D}$ *is corresponding to* $\mathcal{P}$ if $\mathcal{D}$ is represented by a MAG $\mathcal{M}$ consistent with $\mathcal{P}$. Let $\mathcal{M}_{\underset{\sim}{X}}$ denote the graph obtained from $\mathcal{M}$ by deleting the directed edges out of X.

Zhang (2008a) proposed complete rules for obtaining a PAG with observational data. Further, Wang et al. (2022a;b) present complete rules for incorporating local background knowledge into a PAG. See Appendix A.3 for the rules.

Definition 1 (Adjustment set; Pearl (2009); van der Zander et al. (2014))**.** Given a DAG, MAG, or PAG G, $\mathbf{Z}$ is called an adjustment set relative to (X, Y) if for any density f compatible with G, the causal effect of X on Y $f(Y|do(X)) =$

$$\begin{cases} f(Y|X), & \text{if } \mathbf{Z} = \varnothing, \\ \int_{\mathbf{Z}} f(Y|\mathbf{Z}, X) f(\mathbf{Z})\,\mathrm{d}\mathbf{Z}, & \text{otherwise.} \end{cases} \tag{1}$$

A common method to estimate causal effect is *covariate adjustment*, *i.e.*, estimating by (1) with *adjustment sets* in Def. 1. *Adjustment criterion* (Appendix A.4, Def. 8) characterizes adjustment sets. A set $\mathbf{Z}$ satisfies adjustment criterion if and only if $\mathbf{Z}$ is an adjustment set (Shpitser et al., 2010; VanderWeele & Shpitser, 2011; van der Zander et al., 2014).

Definition 2 (D-SEP(X, Y, G); Spirtes et al. (2000); Colombo et al. (2012); Maathuis et al. (2015))**.** Let X and Y be two distinct vertices in a mixed graph G. We say that $V \in$ D-SEP(X, Y, G) if $V \neq X$, and there is a collider path between X and V in G, such that every vertex on this path (including V) is an ancestor of X or Y in G.

Maathuis et al. (2015) proposed *generalized back-door criterion* to identify causal effect of $\mathbf{X}$ on $\mathbf{Y}$ in a MAG/PAG by adjusting for a *generalized back-door set*. When X is singleton, there is a generalized back-door set relative to (X, Y) if and only if there is an adjustment set relative to (X, Y) (Perkovic et al., 2017). We combine them in Prop. 1.

Proposition 1 (Maathuis et al. (2015); Perkovic et al. (2017))**.** *Let X and Y be two distinct vertices in G, where G is a MAG or PAG. There exists an adjustment set relative to (X, Y) in G if and only if $Y \notin$* Adj$(X, G_{\underline{X}})$ *and* D-SEP$(X, Y, G_{\underline{X}}) \cap$ PossDe$(X, G) = \emptyset$. *Moreover, if an adjustment set exists, then* D-SEP$(X, Y, G_{\underline{X}})$ *is such a set. Denote* D-SEP$(X, Y, G_{\underline{X}})$ *by* $\mathbf{D}$*, then*

$$f(Y|do(X = x)) = \int_{\mathbf{D}} f(\mathbf{D}) f(Y|\mathbf{D}, X = x)\,\mathrm{d}\mathbf{D}. \tag{2}$$

4. Proposed Method

In this section, we present the method to determine the set of possible causal effects of a variable X on outcome Y via adjustment in a PAG $\mathcal{P}$. $\mathcal{P}$ can be learned by FCI algorithm from observational data (Spirtes et al., 2000). Let d denote the number of vertices in $\mathcal{P}$. We assume the absence of selection bias. If the causal effect is identifiable in $\mathcal{P}$ by Prop. 1, an unbiased estimate can be directly returned, thus there is no need to determine a set. Hence we focus on the case when it is unidentifiable. And our attention is only on finding all *valid adjustment sets* given $\mathcal{P}$, without touching upon practical calculation of causal effects by (1). Here, a set of vertices is a *valid adjustment set* if there exists a DAG corresponding to $\mathcal{P}$ such that in the DAG the causal effect of X on Y is identifiable by adjusting for this set.

A direct method is first enumerating all the MAGs in the MEC represented by $\mathcal{P}$, then enumerating all the DAGs for each above MAG. However, the enumerations of either DAGs or MAGs are computationally prohibitive. In Sec. 4.1, we present the theoretical result for circumventing the enumerations of DAGs. We provide a graphical characterization for the adjustment set comprised of observable variables in DAGs represented by a given MAG, through which we can find all valid adjustment sets based on mere MAGs instead of DAGs. In Sec. 4.2, we find all the valid adjustment sets without enumerating MAGs. The whole process includes two steps. First, we determine all possible definite local structures at X. The definite local structure is not sufficient for determining a unique adjustment set in the presence of latent variables. Hence, we further provide a graphical characterization of valid adjustment sets relative to (X, Y) under each definite local structure at X, which can be evaluated in $\mathcal{O}(d^3)$ for each set. The algorithm to determine the set of causal effects is presented in Sec. 4.3, associated with a worst-case complexity analysis.

4.1. Adjustment sets in DAGs represented by a MAG

In this part, we provide a graphical characterization for adjustment sets (relative to (X, Y)) in *all* DAGs represented by a given MAG. For a MAG $\mathcal{M}$, the causal effect of X on Y is *identifiable in $\mathcal{M}$ via covariate adjustment* if and only if all the DAGs represented by $\mathcal{M}$ are associated with the same causal effect and can be estimated with observational data via covariate adjustment (van der Zander et al., 2014, Def. 5.3). Prop. 1 presents the sufficient and necessary condition for the identifiability. Hence, if the graphical conditions of Prop. 1 are satisfied for a MAG $\mathcal{M}$, it is direct that all the DAGs represented by $\mathcal{M}$ are associated with the same causal effect, and we can obtain the adjustment set according to Prop. 1. However, when it is unidentifiable in $\mathcal{M}$, different DAGs represented by $\mathcal{M}$ are possibly associated with different causal effects. Perhaps surprisingly, we find that for each DAG represented by $\mathcal{M}$, it is either associated with a *common* causal effect that can be estimated, or with a causal effect that cannot be estimated with observational data by covariate adjustment. Concretely, we provide Thm. 1, establishing a graphical characterization for the adjustment sets comprised of $\mathbf{V}(\mathcal{M})$ in the DAGs.

Theorem 1. Suppose a MAG $\mathcal{M}$ where $X \in \text{Anc}(Y, \mathcal{M})$. There exists a DAG $\mathcal{D}$ represented by $\mathcal{M}$ such that the causal effect of X on Y in $\mathcal{D}$ can be identified by adjusting for a set comprised of $\mathbf{V}(\mathcal{M})$ if and only if D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{M}) = \emptyset$. Furthermore, if such a set exists, D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ is an adjustment set.

Remark 1. X is restricted to be an ancestor of Y in Thm. 1. If $X \notin \text{Anc}(Y, \mathcal{M})$, it is a trivial case that X has no causal effect on Y in any DAG $\mathcal{D}$ represented by the MAG $\mathcal{M}$.

Thm. 1 is similar to Prop. 1 in form, but they are quite different in both the implications and proofs. Prop. 1 implies an adjustment set in a MAG $\mathcal{M}$ when all the DAGs represented by $\mathcal{M}$ are associated with the same causal effect. Here, since we aim to determine the set of causal effects, *regardless of whether* all the DAGs represented by $\mathcal{M}$ are associated with the same causal effect, if there are some DAGs represented by the MAG associated with the causal effect that is identifiable by adjusting for observed variables, then we want to include the causal effect in the set. See the MAG $\mathcal{M}$ in Fig. 1(b) for an example, where D-SEP$(X, Y, \mathcal{M}_{\underline{X}}) \cap \text{PossDe}(X, \mathcal{M}) = \{B\}$ and D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{PossDe}(X, \mathcal{M}) = \emptyset$. According to Prop. 1, the causal effect of X on Y in $\mathcal{M}$ is not identifiable by covariate adjustment. It is true since the DAGs in Fig. 1(c), 1(d), and 1(e) represented by $\mathcal{M}$ are associated with different causal effects. Nevertheless, according to Thm. 1, there exists some DAG represented by $\mathcal{M}$ where the causal effect can be identified and $\{A\}$ is an adjustment set. It is true because it is the case in DAGs as Fig. 1(c), 1(e).

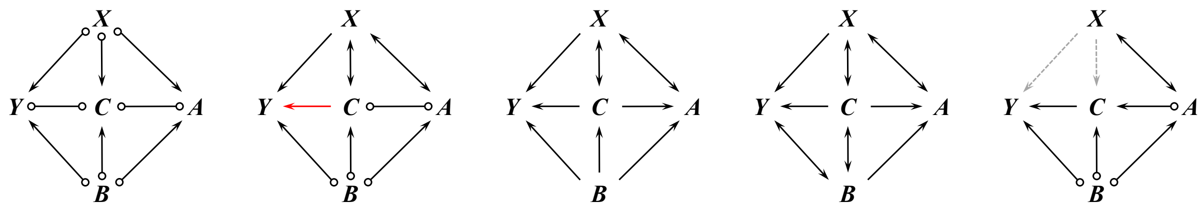
Thm. 1 has two implications. One is that it provides a graphical condition based on a mere MAG for the existence of adjustment sets comprised of observed variables in the DAGs represented by the MAG, thus we can find adjustment sets on the level of MAGs. The other is that it implies that the causal effects are *identical* in the DAGs where the causal effect is identifiable by adjusting for a set of observed variables. Hence, if D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{M}) = \emptyset$, D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ is an adjustment set in *all the DAGs* where the causal effect is identifiable by adjusting for a set comprised of $\mathbf{V}(\mathcal{M})$, thus finding D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ as the adjustment set is sufficient. These two aspects ensure that we can obtain the adjustment set without enumerating DAGs. Henceforth, we make a convention that when we say *an adjustment set in $\mathcal{M}$*, it implies an adjustment set comprised of $\mathbf{V}(\mathcal{M})$ in some DAG represented by $\mathcal{M}$ as Thm. 1, without restricting that all the DAGs represented by $\mathcal{M}$ are associated with the same causal effect as Prop. 1.

4.2. Adjustment sets in MAGs consistent with a PAG

In this part, we aim to find all the sets that can be adjustment sets relative to (X, Y) in some MAG consistent with a given PAG $\mathcal{P}$. This task is hard to achieve with only $\mathcal{P}$ due to the considerable uncertainty of the structure. Hence, we first determine all the *valid local structures at X*, *i.e.*, what orientations of the circles at X in $\mathcal{P}$ can be in the MAGs consistent with $\mathcal{P}$. To determine which local structure is valid, we introduce a result of Wang et al. (2022b) in Prop. 2, which presents a sufficient and necessary condition for the existence of MAGs with any given definite local structure at X. Before that, we introduce *bridged* in Def. 3.

Definition 3 (Bridged relative to $\mathbf{V}'$ in H, Wang et al. (2022b))**.** Let H be a partial mixed graph. Denote G a subgraph of H induced by a set of vertices $\mathbf{V}$. Given a set of vertices $\mathbf{V}'$ in H that is disjoint of $\mathbf{V}$, two vertices A and B in a connected circle component of G are *bridged relative to* $\mathbf{V}'$ if either $A = B$ or in each minimal circle path $A(= V_0) \circ\!\!-\!\!\circ V_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_n \circ\!\!-\!\!\circ B(= V_{n+1})$ from A to B in G, there exists one vertex $V_s, 0 \leq s \leq n+1$, such that $\mathcal{F}_{V_i} \subseteq \mathcal{F}_{V_{i+1}}, 0 \leq i \leq s-1$ and $\mathcal{F}_{V_{i+1}} \subseteq \mathcal{F}_{V_i}, s \leq i \leq n$, where $\mathcal{F}_i = \{V \in \mathbf{V}' \mid V \ast\!\!-\!\!\circ V_i \text{ in } H\}$. Further, G is *bridged relative to* $\mathbf{V}'$ *in* H if any two vertices in a connected circle component of G are bridged relative to $\mathbf{V}'$.

Proposition 2 (Wang et al. (2022b))**.** *Given a PAG $\mathcal{P}$, for any set $\mathbf{C} \subseteq \{V \mid X \circ\!\!\rightarrow\!\!\ast\, V \text{ in } \mathcal{P}\}$, there exists a MAG $\mathcal{M}$ consistent with $\mathcal{P}$ with $X \leftrightarrow\!\!\ast V$ for $\forall V \in \mathbf{C}$ and $X \rightarrow V$ for $\forall V \in \{V \mid X \circ\!\!-\!\!\ast\, V \text{ in } \mathcal{P}\} \backslash \mathbf{C}$ if and only if*



(a) PAG $\mathcal{P}$ (b) $\mathbb{M}_1$ (c) $\mathcal{M}_1$ (d) $\mathcal{M}_2$ (e) $\mathbb{M}_2$

Figure 2: Fig. 2(a) depicts a PAG $\mathcal{P}$. We first consider all the valid local structures at X. For the local structure dictated by $\mathbf{C} = \{A, C\}$, we could obtain a maximal local MAG $\mathbb{M}_1$ as Fig. 2(b), where the edge $C \to Y$ colored by red denotes the edges oriented by the complete orientation rules. Fig. 2(c) and 2(d) depict two MAGs valid to $\mathbb{M}$. We note the adjustment sets in them are $\{B, C\}$ and $\{C\}$, respectively. Fig. 2(e) depicts another maximal local MAG $\mathbb{M}_2$ obtained from the local structure dictated by $\mathbf{C} = \{A\}$. The edges with solid line are those remained after deleting the edges out of X.

(1) $\mathrm{PossDe}(X, \mathcal{P}[-\mathbf{C}]) \cap \mathrm{Pa}(\mathbf{C}, \mathcal{P}) = \emptyset$*;*

(2) the subgraph $\mathcal{P}[\mathbf{C}]$ *of* $\mathcal{P}$ *induced by* $\mathbf{C}$ *is a complete graph;*

(3) $\mathcal{P}[\mathrm{PossDe}(X, \mathcal{P}[-\mathbf{C}]) \backslash \{X\}]$ *is bridged relative to* $\mathbf{C} \cup \{X\}$ *in* $\mathcal{P}$*.*

Remark 2. Given any set $\mathbf{C} \subseteq \{V \mid X \circ\!\!\rightarrow\!\!* V \text{ in } \mathcal{P}\}$, we transform $X \circ\!\!\rightarrow\!\!* V$ to $X \leftarrow\!\!* V$ for $\forall V \in \mathbf{C}$ and transform $X \circ\!\!\rightarrow\!\!* V$ to $X -\!\!* V$ for others, the marks at X are definite. Hence, each set of vertices $\mathbf{C}$ dictates a local structure at X.

Hence, for each definite local structure dictated by $\mathbf{C}$, we can determine whether it is valid by Prop. 2. After enumerating each set $\mathbf{C}$ and using Prop. 2, we can obtain all valid local structures (at X). Given each valid local structure, we introduce the sound and complete orientation rules (Wang et al., 2022b) to further orient the PAG with the local structures, which are detailed in Appendix A.3. We give Fig. 2(a) as an example. The local structure dictated by $\mathbf{C} = \{A, C\}$ is valid according to Prop. 2. When we introduce this local structure, it can be seen as background knowledge, and thus we further orient the partial graph as Fig. 2(b) with the complete orientation rules of Wang et al. (2022b) until no rules are triggered. In the following, we call the obtained graph from $\mathcal{P}$ with a valid local structure dictated by $\mathbf{C}$ and the complete orientation rules by *maximal local MAG based on* $\mathcal{P}$ *and* $\mathbf{C}$, and denote it by $\mathbb{M}$. When it is clear from the context, we call it *maximal local MAG* for short. A MAG $\mathcal{M}$ is *valid to* $\mathbb{M}$ if $\mathcal{M}$ is consistent with $\mathcal{P}$ and has the non-circle marks in $\mathbb{M}$. For each valid local structure at X, we can obtain a maximal local MAG. The graph in Fig. 2(b) is an example of $\mathbb{M}$. By obtaining all valid local structures of X and using the complete orientation rules, we can obtain all maximal local MAGs.

For a maximal local MAG $\mathbb{M}$, we propose the method to find the adjustment sets in the MAGs valid to $\mathbb{M}$ in the following. We only consider the non-trivial case $Y \in \mathrm{PossDe}(X, \mathbb{M})$, otherwise X has no causal effect on Y according to Lemma 8 in Appendix D.2. A question here is, could we determine a common adjustment set for all the MAGs valid to $\mathbb{M}$? Unfortunately, it is not the case in the presence of latent variables. See $\mathbb{M}$ in Fig. 2(b) for an example, there are MAGs valid to $\mathbb{M}$ in Fig. 2(c) and 2(d) with distinct adjustment sets $\{B, C\}$ and $\{C\}$, respectively. The definite local structure at X is not sufficient for determining a common adjustment set. We need to consider further what adjustment sets can be in the MAGs valid to $\mathbb{M}$.

A trivial method is to enumerate each MAG $\mathcal{M}$ valid to $\mathbb{M}$, and obtain D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ as the adjustment set if the conditions of Thm. 1 are fulfilled. However, the space of MAGs is extremely large. In the worst case, the size is $\mathcal{O}(3^{d^2/2})$ ($d(d-1)/2$) edges with 3 types), which makes enumeration computationally prohibitive. To circumvent the super-exponential computation, we convert the problem of finding D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in each enumerated MAG to the problem of determining for any given subset $\mathbf{W}$ of the observed variables, whether there is a MAG $\mathcal{M}$ valid to $\mathbb{M}$ such that $\mathbf{W}$ can be adjustment set in $\mathcal{M}$ as Thm. 1, *i.e.*, $\mathbf{W} =$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ and D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \mathrm{De}(X, \mathcal{M}) = \emptyset$. The benefit of the conversion is apparent. The size of the space of subset $\mathbf{W}$ is $\mathcal{O}(2^d)$, which is super-exponentially less than that of the space of MAGs $\mathcal{O}(3^{d^2/2})$. Note when $\mathbb{M}$ is a complete graph and all the marks at X are arrowheads in $\mathbb{M}$, there are exactly 2^{d-2} adjustment sets that result in different causal effects. Hence 2^{d-2} is a lower bound of the time complexity of finding adjustment sets in MAGs valid to $\mathbb{M}$.

The key here is, for any maximal local MAG $\mathbb{M}$ and set $\mathbf{W}$, we need to determine the existence of MAGs $\mathcal{M}$ valid to $\mathbb{M}$ fulfilling the two conditions (1) $\mathbf{W} =$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ and (2) D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \mathrm{De}(X, \mathcal{M}) = \emptyset$. To address this issue, we propose a graphical characterization. The idea is that for any given $\mathbb{M}$ and $\mathbf{W}$, we determine *whether* *we can construct* a MAG $\mathcal{M}$ fulfilling the two conditions. If the construction succeeds, then there exists the MAGs aforementioned, otherwise the MAGs do not exist.

For the construction process, we introduce $\bar{\mathbf{W}}$ and *block set* $\mathbf{S}$ in Def. 4. Intuitively, $\bar{\mathbf{W}}$ denotes the set of vertices disjoint of $\mathbf{W}$ that are not allowed to be ancestors of Y in the constructed $\mathcal{M}$, otherwise D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})\backslash\mathbf{W} \neq \emptyset$; and $\mathbf{S}$ denotes a set of vertices located at the paths from $\bar{\mathbf{W}}$ to Y. To prevent $\bar{\mathbf{W}}$ to be ancestors of Y in the constructed $\mathcal{M}$, we need to orient the edges of $\mathbf{S}$ in the paths mentioned above. See Fig. 2(b) for an example. Given $\mathbf{W} = \{A, C\}$, to construct a MAG $\mathcal{M}$ such that D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) = \mathbf{W}$, it is necessary to restrict that $\{B\}$ is not an ancestor of Y in $\mathcal{M}$, otherwise $B \in$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})\backslash\mathbf{W}$. To achieve it, $B \circ\!\!\rightarrow Y/C$ must be oriented as bi-directed. Hence we define $\bar{\mathbf{W}} = \{B\}$ and $\mathbf{S} \supseteq \{Y, C\}$ in Def. 4.

Definition 4. Given a set of vertices $\mathbf{W}$ in a maximal local MAG $\mathbb{M}$, we define a set of vertices $\bar{\mathbf{W}}$ as $V \in \bar{\mathbf{W}}$ if and only if $V \in$ PossAn$(Y, \mathbb{M})\backslash\mathbf{W}$ and there exists a collider path beginning with an arrowhead from X to V where each non-endpoint vertex belongs to $\mathbf{W}$. We say $\mathbf{S}$ is a *block set* if Anc$(Y \cup \mathbf{W}, \mathbb{M}) \cap [$PossDe$(\bar{\mathbf{W}}, \mathbb{M})\backslash\bar{\mathbf{W}}] \subseteq \mathbf{S} \subseteq$ PossAn$(Y \cup \mathbf{W}, \mathbb{M}) \cap [$PossDe$(\bar{\mathbf{W}}, \mathbb{M})\backslash\bar{\mathbf{W}}]$.

We then present *potential adjustment set* in Def. 5. According to Def. 2 and Thm. 1, there is $\mathbf{W} =$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ and D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap$ De$(X, \mathcal{M}) = \emptyset$ in some MAG $\mathcal{M}$ valid to $\mathbb{M}$ *only if* $\mathbf{W}$ is a potential adjustment set[2].

Definition 5. In a maximal local MAG $\mathbb{M}$, $\mathbf{W}$ is a *potential adjustment set* if

(1) $\forall V \in \mathbf{W}$, there is a collider path $X \leftrightarrow \cdots \leftarrow\!\!* V$ such that each non-endpoint belongs to $\mathbf{W}$, and there is a possible directed path from V to Y that does not go through the vertices in $\bar{\mathbf{W}}$;

(2) $\mathbf{W} \cap$ PossDe$(X, \mathbb{M}) = \emptyset$;

(3) $\bar{\mathbf{W}} \cap$ Anc$(Y \cup \mathbf{W}, \mathbb{M}) = \emptyset$.

We present the graphical condition in Thm. 2. To prove the existence of MAGs, we give the construction method, shown by Alg. 2 in Appendix D.2. The complexity of judging the three conditions in Thm. 2 for a block set $\mathbf{S}$ is $\mathcal{O}(d^3)$.

Theorem 2. Given a maximal local MAG $\mathbb{M}$, for any potential adjustment set $\mathbf{W}$, there exists a MAG $\mathcal{M}$ valid to $\mathbb{M}$ such that $\mathbf{W}$ is an adjustment set in $\mathcal{M}$ if there exists a block set $\mathbf{S}$ such that

(1) PossDe$(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \cap$ Pa$(\mathbf{S}, \mathbb{M}) = \emptyset$;

(2) $\mathbb{M}[\mathbf{S}_V]$ is a complete graph for any $V \in \bar{\mathbf{W}}$, where $\mathbf{S}_V = \{V' \in \mathbf{S} | V \circ\!\!-\!\!* V' \text{ in } \mathbb{M}\}$;

(3) $\mathbb{M}[$PossDe$(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is bridged relative to $\mathbf{S}$ in $\mathbb{M}$.

Remark 3. The initial target of Thm. 2 is to present a sufficient condition for the existence of MAG $\mathcal{M}$ such that D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) = \mathbf{W}$, where $\mathbf{W}$ is a given potential adjustment set. However, following our MAG construction method as Alg. 2 in Appendix D.2, we can only construct a MAG $\mathcal{M}$ where D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \subseteq \mathbf{W}$ and $\mathbf{W} \cap$ De$(X, \mathcal{M}) = \emptyset$, as Alg. 2 cannot ensure that each vertex V in $\mathbf{W}$ is an ancestor of Y in the construction MAG. Nevertheless, perhaps surprisingly, there is always a good property that $\mathbf{W}\backslash$D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \perp Y \mid \{X,$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})\}$ if $\mathbf{W}\backslash$D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ is non-empty in the constructed $\mathcal{M}$. With this property, it is direct that using $\mathbf{W}$ and D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ as the adjustment sets in (1) can lead to the *same* causal effect, *i.e.*, $\mathbf{W}$ is equivalent to D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in the sense of estimating the causal effect. See the proof of Thm. 2 in Appendix D.2 for details.

Hence, for each subset $\mathbf{W}$ of $\mathbf{V}(\mathcal{P})\backslash\{X, Y\}$, we can evaluate whether it can be an adjustment set in some MAG valid to $\mathbb{M}$ by determining whether it is a potential adjustment set and whether the three conditions in Thm. 2 are satisfied for a block set $\mathbf{S}$ if so. By considering all the subsets, we can find all the valid adjustment sets in MAGs valid to $\mathbb{M}$.

Finally, there is one issue remaining to address: for the adjustment set D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in any MAG $\mathcal{M}$ valid to $\mathbb{M}$, can we always find it by the process above? We present Thm. 3 to show that the adjustment set D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in any MAG $\mathcal{M}$ valid to $\mathbb{M}$ is a potential adjustment set and satisfies the three conditions in Thm 2 for a block set $\mathbf{S}$.

Theorem 3. Given a maximal local MAG $\mathbb{M}$, suppose a MAG $\mathcal{M}$ valid to $\mathbb{M}$ such that there exists an adjustment set relative to (X, Y). Let $\mathbf{W}$ be D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. Then $\mathbf{W}$ is a potential adjustment set in $\mathbb{M}$ and there exists a block set $\mathbf{S}$ such that

(1) PossDe$(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \cap$ Pa$(\mathbf{S}, \mathbb{M}) = \emptyset$;

(2) $\mathbb{M}[\mathbf{S}_V]$ is a complete graph for any $V \in \bar{\mathbf{W}}$, where $\mathbf{S}_V = \{V' \in \mathbf{S} | V \circ\!\!-\!\!* V' \text{ in } \mathbb{M}\}$;

(3) $\mathbb{M}[$PossDe$(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is bridged relative to $\mathbf{S}$ in $\mathbb{M}$.

4.3. The algorithm to determine the set of causal effects

In this part, based on the theoretical results before, we present the method in Alg. 1 to find all the valid adjustment sets given a PAG $\mathcal{P}$ and then determine the set of causal effects by (1). If X is not a possible ancestor of Y

[2]The detailed proof will be given later in Thm. 3.

in $\mathcal{P}$, X has no effect on Y in every MAG consistent with $\mathcal{P}$, thereby returning no causal effect. If the causal effect is identifiable in $\mathcal{P}$ by Prop. 1, we obtain the adjustment set directly according to Prop. 1. When it is not identifiable, we first find all valid local structures at X based on Prop. 2, and obtain the corresponding maximal local MAGs. In each maximal local MAG $\mathbb{M}$, for any set $\mathbf{W} \subseteq \mathbf{V}(\mathcal{P})\backslash\{X, Y\}$, if $\mathbf{W}$ is a potential adjustment set as in Def. 5, we determine whether it is an adjustment set in some $\mathcal{M}$ valid to $\mathbb{M}$ by Thm. 2. Note the sufficient condition in Thm. 2 is that *there exists a block set* such that the three conditions are fulfilled, hence we search for each set $\mathbf{S}$ on Line 10. Thm. 4 indicates that PAGcauses can output the complete set of causal effects obtainable via covariate adjustment, *i.e.*, it returns the set equivalent to the direct method to enumerate all the DAGs corresponding to $\mathcal{P}$ and use covariate adjustment.

Algorithm 1: PAGcauses

Input: PAG $\mathcal{P}$, X, Y

1 $\widehat{\text{AS}}(\mathcal{P}) = \emptyset$ // Record all the valid adjustment sets;
2 **if** $X \notin \text{PossAn}(Y, \mathcal{P})$ **then return** No causal effects;
3 **if** *the conditions in Prop. 1 are satisfied for* $\mathcal{P}$ **then**
4   **return** $\widehat{\text{AS}}(\mathcal{P}) \leftarrow \{\text{D-SEP}(X, Y, \mathcal{P}_{\underline{X}})\}$ // Prop. 1
5 **for** *each set* $\mathbf{C} \subseteq \{V \mid V \ast\!\!-\!\!\circ X \text{ in } \mathcal{P}\}$ **do**
6   **if** *the three conditions in Prop. 2 are satisfied* **then**
7   Obtain a maximal local MAG $\mathbb{M}$ based on $\mathcal{P}$ and $\mathbf{C}$;
8   Find all potential adjustment sets $\mathbf{W}_1, \mathbf{W}_2, \cdots$ given $\mathbb{M}$ according to Def. 5;
9   **for** *each set potential adjustment set* $\mathbf{W}_i$ **do**
10   **for** *each block set* $\mathbf{S}$ **do**
11   **if** *the three conditions in Thm. 2 are satisfied given* $\mathbf{S}$ **then**
12                     $\widehat{\text{AS}}(\mathcal{P}) \leftarrow \widehat{\text{AS}}(\mathcal{P}) \cup \{\mathbf{W}_i\}$;
13   **break** // Break the loop of $\mathbf{S}$;

Output: Set of causal effects via adjustment in the given PAG $\mathcal{P}$ identified with $\widehat{\text{AS}}(\mathcal{P})$ by (1)

Theorem 4. Given a PAG $\mathcal{P}$, denote the set of (possible) causal effects in the DAGs corresponding to $\mathcal{P}$ which can be estimated with observational data by covariate adjustment and the set of (possible) causal effects obtained from Alg. 1 by $\text{CE}(\mathcal{P})$ and $\widehat{\text{CE}}(\mathcal{P})$. There is $\text{CE}(\mathcal{P}) \overset{set}{=} \widehat{\text{CE}}(\mathcal{P})$.

Proof. The result can be directly concluded according to Thm. 1, 2, and 3. As we consider all the valid local structures at X in Alg. 1, it suffices to show that the set of causal effects $\widehat{\text{CE}}(\mathbb{M})$ identified with $\widehat{\text{AS}}(\mathbb{M})$ by (1) is equal to the set $\text{CE}(\mathbb{M})$, where $\widehat{\text{AS}}(\mathbb{M})$ denotes the obtained adjustment sets on Line 8-13 of Alg. 1, and $\text{CE}(\mathbb{M})$ denotes the set of causal effects that can be identified with observational data in the DAGs represented by any MAGs valid to $\mathbb{M}$.

As shown by Thm. 1, the set of causal effects $\text{CE}(\mathbb{M})$ in the DAGs represented by the MAG $\mathcal{M}$ valid to $\mathbb{M}$ is equal to that identified with the set of $\text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ for each $\mathcal{M}$ valid to $\mathbb{M}$ such that $\text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{M}) = \emptyset$. According to Thm. 3, there is evidently $\text{CE}(\mathbb{M}) \subseteq \widehat{\text{CE}}(\mathbb{M})$. And Thm. 2 implies that for each set $\mathbf{V}'$ in $\widehat{\text{AS}}(\mathbb{M})$, there is some MAG $\mathcal{M}$ valid to $\mathbb{M}$ such that the set $\mathbf{V}'$ is an adjustment set in $\mathcal{M}$, *i.e.*, the adjustment set $\mathbf{V}'$ implies the same causal effect by (1) as that of $\text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in some MAG $\mathcal{M}$ valid to $\mathbb{M}$. Hence $\widehat{\text{CE}}(\mathbb{M}) \subseteq \text{CE}(\mathbb{M})$. We conclude $\widehat{\text{CE}}(\mathbb{M}) \overset{set}{=} \text{CE}(\mathbb{M})$, thus $\widehat{\text{CE}}(\mathcal{P}) \overset{set}{=} \text{CE}(\mathcal{P})$. $\square$

The worst-case complexity of Alg. 1 is $\mathcal{O}(5^d d^6)$, which is super-exponentially less than the complexity $\Omega(3^{d(d-1)/2} d^3)$ of LV-IDA (Malinsky & Spirtes, 2016). Due to the page limit, we show the details in Appendix C. Note the number of causal effects in the worst case grows exponentially with respect to the number of variables. Hence the task of set determination cannot be finished within polynomial time. To evaluate the complexity in general cases, we conduct empirical analysis in Sec. 5.

5. Experiments

In this part, we evaluate the effectiveness and efficiency of PAGcauses to determine the set of causal effects. We take LV-IDA (Malinsky & Spirtes, 2016) as a baseline, which determines the set by enumerating the MAGs in a subspace of MEC. As detailed in Sec. B.1, LV-IDA returns the causal effect in the MAG $\mathcal{M}$ where the causal effect is *identifiable* via adjustment, *i.e.*, $\text{D-SEP}(X, Y, \mathcal{M}_{\underline{X}}) \cap \text{De}(X, \mathcal{M}) = \emptyset$, while PAGcauses also considers the MAG $\mathcal{M}$ where the causal effect is not identifiable but there is some DAG $\mathcal{D}$ represented by $\mathcal{M}$ such that the causal effect is identifiable by some observed variables, *i.e.*, $\text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{M}) = \emptyset$. Hence, to compare the two methods fairly, we modify LV-IDA by finding $\text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ instead of $\text{D-SEP}(X, Y, \mathcal{M}_{\underline{X}})$. In this way, the two methods tackle with the totally same task. Note this modification does not take *any additional computation* to LV-IDA.

We generate random DAGs with vertex number $d = 8, 10, 12, 14, 16$ and each edge between two vertices occurs with probability $\rho = 0.2, 0.3, 0.4, 0.5$, which is called *graph density*. The DAG is parameterized as a linear Gaussian structural equation model. The weight of each edge is drawn from Uniform($[1, 2]$). For each graph, we randomly select three vertices as latent variables and the others as observed variables. To prevent degeneration to a trivial case where X has no causal effect on Y, we select the last vertex in the causal order as Y. We randomly select a decision variable X. For each set of parameters, we generate 100 causal graphs and obtain the output set in the time limit of 3000

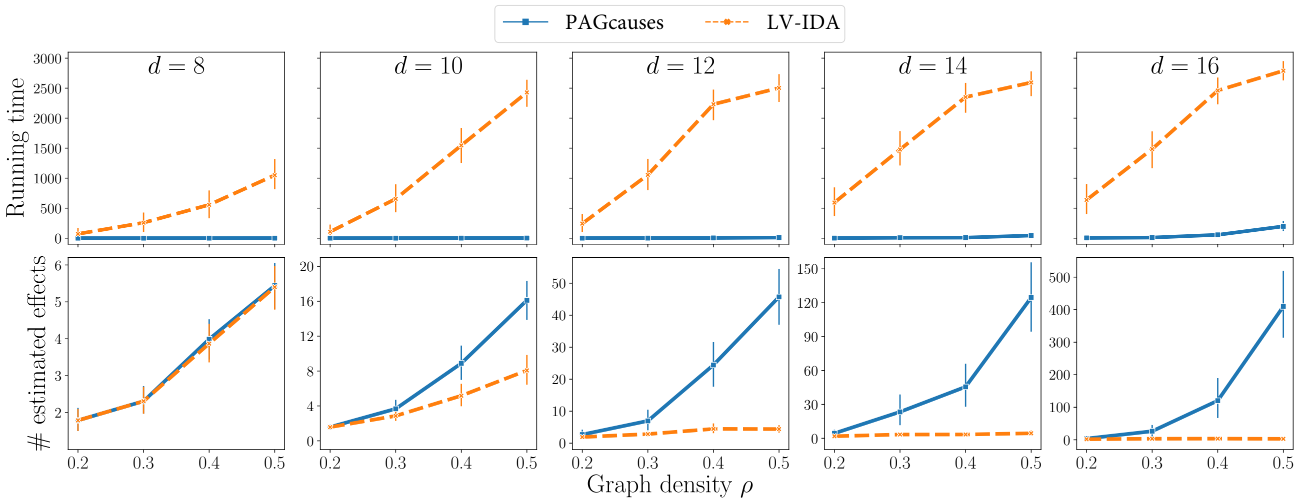


Figure 3: Results of the number of returned causal effects and running time over 100 simulations for each vertice number(including 3 latent ones)/graph density. The vertical line represents the 95% confidence interval generated by bootstrap sampling. The maximum running time for each simulation is 3000s.

seconds for each graph. For LV-IDA, when the running time achieves the limit, it stops and returns the causal effects in the enumerated MAGs. For PAGcauses, it returns no effects if the time is used up. Since the main focus is on the set determination based on a PAG, we obtain PAG directly with the true covariance matrix of the observable variables.

We show the average number of the output set of causal effects and running time in 100 graphs under each parameter in Fig. 3. The results demonstrate the effectiveness of our method and that the efficiency of determining the set of causal effects is improved tremendously by exploiting the proposed graphical characterization. When d and ρ is small, PAGcauses and LV-IDA obtain the same set. When $d \geq 10$, the sets are different because usually LV-IDA cannot enumerate all the MAGs within the limited time. When d grows, the number of returned causal effects by LV-IDA tends to 0. We give a rough analysis for this phenomenon. Suppose within the limited time we can enumerate N MAGs at most. 2^d is the rough number of causal effects and $3^{d^2/2}$ is the rough number of MAGs. As d grows, the expected number of returned causal effects $N \times 2^d/3^{d^2/2}$ tends to 0.

6. Conclusion

In this paper, we present the method to determine the set of possible causal effects obtainable via covariate adjustment in all the DAGs corresponding to a given PAG. We show that even if the causal effect is not identifiable in a MAG via covariate adjustment, there could exist DAGs represented by the MAG associated with an identifiable causal effect. Hence DAG needs to be considered in the task of determining the set of possible causal effects. By introducing new graphical conditions, our method can circumvent the enumerations of DAGs and MAGs, and output the same set as the enumerations methods with super-exponentially less complexity. Experiments validate the effectiveness and the significant improvement in efficiency of our approach.

Despite the efficient set determination of causal effects via covariate adjustment in a PAG achieved, it is important to note that accurate causal relations learned with observational data are a prerequisite for causal effect identification methods. The task of learning causal relations, however, is quite challenging in practice. Misinterpreted relations can result in inaccurate estimations, thereby highlighting the fragility of decision methods based on causal effect identification. This raises the question of whether *learning causal relations* is actually more challenging than *decisions*.

Recently, Zhou (2022) proposed to rethink the relation between *prediction*, *decision*, and *causality*, and indicated that *correlation* is crucial for prediction, *causation* is fundamental for scientific discovery, while decision-making requires something in-between, which is called *rehearsation*. This new perspective offers a fresh view for future study of decision methods.

Acknowledgements

This research was supported by NSFC (61921006), the Collaborative Innovation Center of Novel Software Technology and Industrialization, and the program A for Outstanding PhD candidate of Nanjing University. We thank Yangbo He, Peng Zhao, Wen-Bo Du, and Xin-Yu Zhang for insightful discussions, and thank reviewers for their valuable feedback.

References

Ali, R. A., Richardson, T., Spirtes, P., and Zhang, J. Orientation rules for constructing markov equivalence classes of maximal ancestral graphs. Technical report, 2005.

Cai, R., Qiao, J., Zhang, K., Zhang, Z., and Hao, Z. Causal discovery from discrete data using hidden compact representation. In *Advances in Neural Information Processing Systems*, pp. 2666–2674, 2018.

Chakrabortty, A., Nandy, P., and Li, H. Inference for individual mediation effects and interventional effects in sparse high-dimensional causal graphical models. *arXiv preprint arXiv:1809.10652*, 2018.

Cheng, D., Li, J., Liu, L., Liu, J., Yu, K., and Le, T. D. Causal query in observational data with hidden variables. In *24th European Conference on Artificial Intelligence*, volume 325, pp. 2551–2558. IOS Press, 2020.

Colombo, D., Maathuis, M. H., Kalisch, M., and Richardson, T. S. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pp. 294–321, 2012.

Dirac, G. A. On rigid circuit graphs. In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, volume 25, pp. 71–76. Springer, 1961.

Entner, D., Hoyer, P. O., and Spirtes, P. Data-driven covariate selection for nonparametric estimation of causal effects. In *the 16th International Conference on Artificial Intelligence and Statistics*, pp. 256–264, 2013.

Fang, Z., Zhao, R., Liu, Y., and He, Y. On the representation of causal background knowledge and its applications in causal inference. *CoRR*, abs/2207.05067, 2022.

Golumbic, M. C. *Algorithmic graph theory and perfect graphs*. Elsevier, 2004.

Guo, R. and Perkovic, E. Minimal enumeration of all possible total effects in a markov equivalence class. In *The 24th International Conference on Artificial Intelligence and Statistics*, pp. 2395–2403, 2021.

Henckel, L., Perković, E., and Maathuis, M. H. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society Series B*, 84(2):579–599, 2022.

Hu, Z. and Evans, R. Faster algorithms for markov equivalence. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence*, pp. 739–748, 2020.

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

Jaber, A., Zhang, J., and Bareinboim, E. Causal identification under markov equivalence. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence*, pp. 978–987, 2018.

Jaber, A., Zhang, J., and Bareinboim, E. Causal identification under markov equivalence: Completeness results. In *Proceedings of the 36th International Conference on Machine Learning*, pp. 2981–2989, 2019.

Jaber, A., Ribeiro, A., Zhang, J., and Bareinboim, E. Causal identification under markov equivalence: Calculus, algorithm, and completeness. In *Advances in Neural Information Processing Systems*, 2022.

Jung, Y., Tian, J., and Bareinboim, E. Estimating identifiable causal effects on markov equivalence class through double machine learning. In *Proceedings of the 38th International Conference on Machine Learning*, pp. 5168–5179, 2021.

Lee, S. and Bareinboim, E. Structural causal bandits: Where to intervene? In *Advances in Neural Information Processing Systems*, pp. 2573–2583, 2018.

Maathuis, M. H., Kalisch, M., Bühlmann, P., et al. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A):3133–3164, 2009.

Maathuis, M. H., Colombo, D., et al. A generalized backdoor criterion. *The Annals of Statistics*, 43(3):1060–1088, 2015.

Malinsky, D. and Spirtes, P. Estimating causal effects with ancestral graph markov models. In *Conference on Probabilistic Graphical Models*, pp. 299–309, 2016.

Pearl, J. *Causality*. Cambridge University Press, 2009.

Perkovic, E., Textor, J., Kalisch, M., and Maathuis, M. H. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *Journal of Machine Learning Research*, 18:220:1–220:62, 2017.

Peters, J., Mooij, J. M., Janzing, D., and Schölkopf, B. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15:2009–2053, 2014.

Pfister, N. and Peters, J. Identifiability of sparse causal effects using instrumental variables. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 1613–1622, 2022.

Qin, T., Wang, T.-Z., and Zhou, Z.-H. Budgeted heterogeneous treatment effect estimation. In *Proceedings of the*

38th International Conference on Machine Learning, pp. 8693–8702, 2021.

Richardson, T., Spirtes, P., et al. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

Rothenhäusler, D., Heinze, C., Peters, J., and Meinshausen, N. BACKSHIFT: learning causal cyclic graphs from unknown shift interventions. In *Advances in Neural Information Processing Systems*, pp. 1513–1521, 2015.

Shpitser, I. and Pearl, J. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference*, pp. 1219–1226, 2006.

Shpitser, I., VanderWeele, T. J., and Robins, J. M. On the validity of covariate adjustment for estimating causal effects. In *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence*, pp. 527–536, 2010.

Spirtes, P., Glymour, C. N., and Scheines, R. *Causation, prediction, and search.* MIT Press, 2000.

Tian, J. Generating markov equivalent maximal ancestral graphs by single edge replacement. In *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, pp. 591–598, 2005.

Tian, J. and Pearl, J. A general identification condition for causal effects. In *Proceedings of the 18th National Conference on Artificial Intelligence and 14th Conference on Innovative Applications of Artificial Intelligence*, pp. 567–573, 2002.

van der Zander, B., Liskiewicz, M., and Textor, J. Constructing separators and adjustment sets in ancestral graphs. In *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*, pp. 907–916, 2014.

van der Zander, B., Liskiewicz, M., and Textor, J. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artificial Intelligence*, 270:1–40, 2019.

VanderWeele, T. J. and Shpitser, I. A new criterion for confounder selection. *Biometrics*, 67(4):1406–1413, 2011.

Wang, T.-Z. and Zhou, Z.-H. Actively identifying causal effects with latent variables given only response variable observable. In *Advances in Neural Information Processing Systems*, pp. 15007–15018, 2021.

Wang, T.-Z., Wu, X.-Z., Huang, S.-J., and Zhou, Z.-H. Cost-effectively identifying causal effects when only response variable is observable. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 10060–10069, 2020.

Wang, T.-Z., Qin, T., and Zhou, Z.-H. Sound and complete causal identification with latent variables given local background knowledge. In *Advances in Neural Information Processing Systems*, 2022a.

Wang, T.-Z., Qin, T., and Zhou, Z.-H. Sound and complete causal identification with latent variables given local background knowledge. *Under Review*, 2022b.

Wienöbst, M., Bannach, M., and Liskiewicz, M. A new constructive criterion for markov equivalence of mags. In *Uncertainty in Artificial Intelligence, Proceedings of the 38th Conference on Uncertainty in Artificial Intelligence*, pp. 2107–2116, 2022.

Yata, K. Optimal decision rules under partial identification. *arXiv preprint arXiv:2111.04926*, 2021.

Yu, K., Liu, L., and Li, J. Learning markov blankets from multiple interventional data sets. *IEEE transactions on neural networks and learning systems*, 31(6):2005–2019, 2019.

Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence*, 172(16-17): 1873–1896, 2008a.

Zhang, J. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008b.

Zhang, J. and Spirtes, P. A transformational characterization of markov equivalence for directed acyclic graphs with latent variables. In *Proceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, pp. 667–674, 2005.

Zhang, K., Huang, B., Zhang, J., Glymour, C., and Schölkopf, B. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pp. 1347–1353, 2017.

Zhao, H., Zheng, Z., and Liu, B. On the markov equivalence of maximal ancestral graphs. *Science in China Series A: Mathematics*, 48(4):548–562, 2005.

Zhao, P., Wang, G., Zhang, L., and Zhou, Z.-H. Bandit convex optimization in non-stationary environments. *Journal of Machine Learning Research*, 22:125:1–125:45, 2021.

Zhao, P., Li, L.-F., and Zhou, Z.-H. Dynamic regret of online markov decision processes. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 26865–26894, 2022.

Zhou, Z.-H. Rehearsal: learning from prediction to decision. *Frontiers of Computer Science*, 16(4):164352, 2022.

A. Detailed Preliminary

A.1. Some definitions about graph

A graph G is *mixed* if the edges in the graph are either directed $\rightarrow$ or bi-directed $\leftrightarrow$. The two ends of an edge are called *marks* and have two types *arrowhead* or *tail*. A graph is a *partial mixed graph* (*PMG*) if it contains directed edges, bi-directed edges, and edges with *circles* ($\circ$). The circle implies that the mark here could be either arrowhead or tail but is indeterminate. An edge $V_i \circ\!\!-\!\!\circ V_j$ is a *circle edge*. The circle component in G is the subgraph consisting of all the $\circ\!\!-\!\!\circ$ edges in G. V_i is *adjacent* to V_j in G if there is an edge between V_i and V_j. Denote the set of vertices adjacent to V_i in G by $\text{Adj}(V_i, G)$. A vertex V_i is a *parent/child* of a vertex V_j if there is $V_i \rightarrow V_j/V_i \leftarrow V_j$. A path in a graph G is a sequence of distinct vertices $\langle V_0, \cdots, V_n \rangle$ such that for $0 \leq i \leq n-1$, V_i and V_{i+1} are adjacent in G. A *directed path* from V_i to V_j is a path comprised of directed edges pointing to the direction of V_j. A *possible directed path* from V_i to V_j is a path without arrowhead at the mark near V_i on every edge. V_i is an *ancestor/possible ancestor* of V_j if there is a *directed path/possible directed path* from V_i to V_j or $V_i = V_j$. V_i is a *descendant/possible descendant* of V_j if there is a *directed path/possible directed path* from V_j to V_i or $V_j = V_i$. Denote the set of *parents/children/ancestors/possible ancestors/descendants/possible descendants* of V_i in G by $\text{Pa}(V_i, G)$/$\text{Chd}(V_i, G)$/$\text{Anc}(V_i, G)$/$\text{PossAn}(V_i, G)$/$\text{De}(V_i, G)$/$\text{PossDe}(V_i, G)$. If $V_i \in \text{Anc}(V_j, G)$ and $V_i \leftarrow V_j$/$V_i \leftrightarrow V_j$, it forms a *directed cycle/almost directed cycle*. $*$ is a wildcard that denotes any of arrowhead, tail, or circle. We make a convention that if an edge is $\circ\!\!-\!\!*$, the $*$ here cannot be a tail for otherwise the circle can be replaced by an arrowhead due to the assumption of no selection bias (there is not an edge with two tails).

A non-endpoint vertex V_i is a collider on a path if the path contains $*\!\!\rightarrow V_i \leftarrow\!\!*$. A path p from V_i to V_j is a *collider path* if V_i and V_j are adjacent or all the passing vertices are colliders on p. p is a *minimal path* if there are no edges between any two non-consecutive vertices. A path p from V_i to V_j is a *minimal collider path* if p is a collider path and there is not a proper subset $\mathbf{V}'$ of the vertices in p such that there is a collider path from V_i to V_j comprised of $\mathbf{V}'$. A triple $\langle V_i, V_j, V_k \rangle$ on a path is *unshielded* if V_i and V_k are not adjacent. p is an *uncovered path* if every consecutive triple on p is unshielded. A path p is a *minimal possible directed path* if p is minimal and a possible directed path. A graph is a *directed acyclic graph (DAG)* if it contains only directed edges and has no directed cycles, denoted by $\mathcal{D}$.

Definition 6 (Active path; (Richardson et al., 2002; Zhang, 2008a))**.** In a mixed graph, a path p between vertices X and Y is *active* (*m-connecting*) relative to a (possibly empty) set of vertices $\mathbf{Z}$ ($X, Y \notin \mathbf{Z}$) if (1) every non-collider on p is not a member of $\mathbf{Z}$; (2) every collider on p has a descendant in $\mathbf{Z}$.

$\mathbf{X}$ and $\mathbf{Y}$ are *m-separated* by $\mathbf{Z}$ if there is no active path between any vertex in $\mathbf{X}$ and any vertex in $\mathbf{Y}$ relative to $\mathbf{Z}$. A mixed graph is an *ancestral graph* if there is no directed or almost directed cycle (since we assume no selection bias, we do not consider undirected edges in this paper). An ancestral graph is a *maximal ancestral graph (MAG, denoted by $\mathcal{M}$)* if it is *maximal*, i.e., for any two non-adjacent vertices, there is a set of vertices that m-separates them. A path p between X and Y in an ancestral graph G is an *inducing path* if every non-endpoint vertex on p is a collider and meanwhile an ancestor of either X or Y. An ancestral graph is maximal if and only if there is no inducing path between any two non-adjacent vertices (Richardson et al., 2002).

In a MAG, a path $p = \langle X, \cdots, W, V, Y \rangle$ is a *discriminating path for V* if (1) X and Y are not adjacent, and (2) every vertex between X and V on the path is a collider on p and a parent of Y. Two MAGs are *Markov equivalent* if they share the same m-separations. A class comprised of all the Markov equivalent MAGs is a *Markov equivalence class (MEC)*. A *partial ancestral graph (PAG, denoted by $\mathcal{P}$)* represents an MEC, where a tail/arrowhead occurs if the corresponding mark is tail/arrowhead in all the Markov equivalent MAGs, and a circle occurs otherwise. We say a MAG $\mathcal{M}$ is *consistent with a PAG $\mathcal{P}$* if $\mathcal{P}$ is a PAG that denotes the Markov equivalence class of $\mathcal{M}$. In a mixed graph G, a directed edge $A \rightarrow B$ is *visible* if there a vertex C not adjacent to B, such that there is an edge $C \rightarrow A$, or there is a collider path between C and A that is into A and every non-endpoint is a parent of B (Zhang, 2008b). Otherwise $A \rightarrow B$ is said to be *invisible*.

We introduce two definitions of Zhang (2008b); Maathuis et al. (2009) with slight modifications. For a MAG $\mathcal{M}$, let $\mathcal{M}_{\underline{X}}$ denote the graph obtained from $\mathcal{M}$ by removing all visible directed edges out of X in $\mathcal{M}$. For a PAG $\mathcal{P}$, let $\mathcal{M}$ be any MAG consistent with $\mathcal{P}$ that has the same number of edges into X as $\mathcal{P}$, and let $\mathcal{P}_{\underline{X}}$ denote the graph obtained from $\mathcal{M}$ by removing all directed edges out of X that are visible in $\mathcal{M}$ (it is not required to be unique). $\mathcal{M}_{\underset{\sim}{X}}$ denotes the graph obtained from $\mathcal{M}$ by deleting the directed edges out of X.

We also introduce some graph theory and terminology referring to Maathuis et al. (2009). A graph is *chordal* if any cycle of length four or more has a chord, which is an edge joining two vertices that are not adjacent in the cycle. If $G = (\mathbf{V}, \mathbf{E})$ is chordal, then all subgraphs of G induced by $\mathbf{V}' \subseteq \mathbf{V}$ are chordal.

A vertex A of G is called *simplicial* if its adjacency set Adj(A, G) induces a complete subgraph of G. As shown by Dirac (1961); Golumbic (2004), there are at least two non-adjacent simplicial vertices in any a non-complete chordal graph with more than one vertex. A *perfect elimination order* of a graph G is an ordering $\sigma = (V_1, \cdots, V_n)$ of its vertices, so that each vertex V_i is a simplicial vertex in the induced subgraph $G_{V_i,\cdots,V_n}$.

A.2. The algorithm to obtain a MAG based on a DAG

Next, we present the algorithm to obtain a MAG comprised of observable vertices $\mathbf{O}$ with a DAG comprised of both observable vertices $\mathbf{O}$ and latent vertices $\mathbf{L}$. Due to the assumption of the absence of selection bias, we do not consider the selection variable. We first present a more detailed definition of inducing path involving latent variables in Def. 7.

Definition 7 (Inducing path; Spirtes et al. (2000))**.** In an ancestral graph, let X, Y be any two vertices. And $\mathbf{L}, \mathbf{S}$ be two disjoint sets of vertices not containing X, Y. A path p between X and Y is called an *inducing path relative to* $\langle \mathbf{L}, \mathbf{S}\rangle$ if every non-endpoint vertex on p is either in $\mathbf{L}$ or a collider, and every collider on p is an ancestor of either X, Y, or a member of $\mathbf{S}$.

The construction of MAG comprised of observable vertices $\mathbf{O}$ based on a DAG $\mathcal{D}$ comprised of vertices $\mathbf{O} \cup \mathbf{L}$ is as follows:
Input: a DAG $\mathcal{D}$ over $\mathbf{V} = \mathbf{O} \cup \mathbf{L}$;
Output: a MAG $\mathcal{M}$ over $\mathbf{O}$.
(1) for each pair of variables $A, B \in \mathbf{O}$, A and B are adjacent in $\mathcal{M}$ if and only if there is an inducing path relative to $\langle \mathbf{L}, \emptyset\rangle$ between them in $\mathcal{D}$;
(2) for each pair of adjacent vertices A, B in $\mathcal{M}$, orient the edge between them as follows:

(a) orient it as $A \rightarrow B$ in $\mathcal{M}$ if $A \in \text{Anc}(B, G)$ and $B \notin \text{Anc}(A, G)$;

(b) orient it as $A \leftarrow B$ in $\mathcal{M}$ if $B \in \text{Anc}(A, G)$ and $A \notin \text{Anc}(B, G)$;

(c) orient it as $A \leftrightarrow B$ in $\mathcal{M}$ if $A \notin \text{Anc}(B, G)$ and $B \notin \text{Anc}(A, G)$.

A.3. Some proposed orientation rules

Zhang (2008a) proposed the sound and complete orientation rules for learning a PAG. We show the rules as follows. There are eleven rules $\mathcal{R}_0 - \mathcal{R}_{10}$. Since selection bias is not considered in this paper, we do not show the cases ($\mathcal{R}_5 - \mathcal{R}_7$) that happen only when there is selection bias. $\mathcal{R}_0$ is triggered according to the conditional independence relationship at the beginning of learning a PAG. It is evidently not triggered after, hence we do not show it as well.

$\mathcal{R}_1$: If $A{*\!\!\rightarrow} B \circ\!\!\!-\!\!* R$, and A and R are not adjacent, then orient the triple as $A{*\!\!\rightarrow} B \rightarrow R$.

$\mathcal{R}_2$: If $A \rightarrow B{*\!\!\rightarrow} R$ or $A{*\!\!\rightarrow} B \rightarrow R$, and $A *\!\!-\!\!\circ R$, then orient $A *\!\!-\!\!\circ R$ as $A{*\!\!\rightarrow} R$.

$\mathcal{R}_3$: If $A{*\!\!\rightarrow} B \leftarrow\!\!* R$, $A *\!\!-\!\!\circ D \circ\!\!-\!\!* R$, A and R are not adjacent, and $D *\!\!-\!\!\circ B$, then orient $D *\!\!-\!\!\circ B$ as $D{*\!\!\rightarrow} B$.

$\mathcal{R}_4$: If $\langle K, \ldots, A, B, R\rangle$ is a discriminating path between K and R for B, and $B \circ\!\!-\!\!* R$; then if $B \in$ Sepset(K, R), orient $B \circ\!\!-\!\!* R$ as $B \rightarrow R$; otherwise orient the triple $\langle A, B, R\rangle$ as $A \leftrightarrow B \leftrightarrow R$.

$\mathcal{R}_8$: If $A \rightarrow B \rightarrow R$, and $A {\circ\!\!\rightarrow} R$, orient $A{\circ\!\!\rightarrow} R$ as $A \rightarrow R$.

$\mathcal{R}_9$: If $A{\circ\!\!\rightarrow} R$, and $p = \langle A, B, D, \ldots, R\rangle$ is an uncovered possible directed path from A to R such that R and B are not adjacent, then orient $A{\circ\!\!\rightarrow} R$ as $A \rightarrow R$.

$\mathcal{R}_{10}$: Suppose $A{\circ\!\!\rightarrow} R$, $B \rightarrow R \leftarrow D$, p_1 is an uncovered possible directed path from A to B, and p_2 is an uncovered possible directed path from A to D. Let U be the vertex adjacent to A on p_1 (U could be B), and W be the vertex adjacent to A on p_2 (W could be D). If U and W are distinct, and are not adjacent, then orient $A{\circ\!\!\rightarrow} R$ as $A \rightarrow R$.

Further, Wang et al. (2022b) presented the sound and complete orientation rules when local background knowledge is incorporated. We call the background knowledge by *BK* for short. *BK is local*, if when the BK contains the causal information with respect to a variable X, for each variable adjacent to X in the PAG, the BK implies whether X causes it or not. When the BK is local, the additional orientation rules are as follows:

$\mathcal{R}'_4$: If $\langle K, \cdots, A, B, R\rangle$ is a discriminating path between K and R for B, and $B \circ\!\!\rightarrow\!* R$, then orient $B \circ\!\!\rightarrow\!* R$ as $B \rightarrow R$.

$\mathcal{R}_{11}$: If $A \twoheadrightarrow\!* B$, then $A \rightarrow B$.

Note that $\mathcal{R}'_4$ is triggered only when we incorporate the local BK into a PAG, while $\mathcal{R}_4$ is triggered in the process of obtaining the PAG $\mathcal{P}$.

A.4. Adjustment criterion

Definition 8 (Adjustment criterion; Shpitser et al. (2010); VanderWeele & Shpitser (2011); van der Zander et al. (2014))**.** Let $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ be pairwise disjoint sets of vertices in a DAG $\mathcal{D}$. Let $\text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D})$ denote the set of all descendants in $\mathcal{D}$ of any $W \notin \mathbf{X}$ which lies on a proper causal path from X to Y, *i.e.*, only the first node is in $\mathbf{X}$, in $\mathcal{D}$. Then $\mathbf{Z}$ satisfies adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{D}$ if the following two conditions hold:

(Forbidden set) $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{D}) = \emptyset$, and

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

B. Connection with Previous Works

In this part, we clarify the difference between our method PAGcauses and many previous related studies.

B.1. Malinsky & Spirtes (2016)

Both the input of LV-IDA and PAGcauses are a PAG $\mathcal{P}$. And both the methods aim to output the set of possible causal effects of some decision variable X on outcome Y. Compared to LV-IDA, we improve the efficiency super-exponentially, by building and exploiting the corresponding graphical properties. In addition, there is a bit difference for the output set of the two methods. In the method of LV-IDA, they enumerate all the MAGs consistent with $\mathcal{P}$ in a subspace of Markov equivalence represented by the PAG. For a MAG consistent with $\mathcal{P}$, they return the causal effect *when the causal effect is identifiable in* $\mathcal{M}$, while return NA when it is unidentifiable. For our method, PAGcauses returns the causal effect if there exists a DAG represented by the MAG $\mathcal{M}$ which is associated with the causal effect. The difference is not due to technical reasons. We argue that our definition of set of possible causal effects is more reasonable. When the causal effect cannot be uniquely identifiable in a PAG $\mathcal{P}$, we need to determine the set of possible causal effects. Our target of finding the set is to output what causal effects are possible, which can help decisions. MAG is a tool to represent a series of causal graphs (DAGs) in the presence of latent variables (Richardson et al., 2002). When the causal effect is not identifiable in a MAG, it does not mean that the causal effect is not identifiable in any DAGs represented by MAG. If we only return the causal effect when it is identifiable in $\mathcal{M}$, there are some possible causal effects that are left out, which could be the causal effect in the true underlying causal graph. This phenomenon is detailed in Sec. 4.1. For example, the causal effect of X on Y is not identifiable in the MAG in Fig. 1(b) according to Prop. 1. In this case, if we only return an NA, then we will left out the causal effect identified by adjusting for $\{A\}$, since there are DAGs as Fig. 1(c) and Fig. 1(e) where the causal effect can be identified by adjusting for $\{A\}$. Compared to LV-IDA, our method can output more possible causal effects.

B.2. Cheng et al. (2020)

The method of Cheng et al. (2020) introduces some additional assumptions, such that amenable assumption and pretreatment assumption. The pretreatment assumption says that any other variables $V \in \mathbf{V}(\mathcal{P})\backslash\{X, Y\}$ are not descendants of X. In our method, we do not introduce these assumptions.

B.3. Maathuis et al. (2009)

Maathuis et al. (2009) determines the set of possible causal effects of X on Y under the assumption of no latent variables. Under this assumption, the possible causal effects can be determined by considering all possible local structures at X. In our paper, we consider the set determination in the presence of latent variables. The main difference between PAGcauses and Maathuis et al. (2009) in techniques is that, when there exists latent variables, determining the local structures at X is not sufficient for determining the only causal effect, which is shown by the examples in Fig. 2(c) and Fig. 2(d). Hence after determining the local structures at X, we also need to further present a graphical characterizations for the adjustment sets in Thm. 2 and Thm. 3.

B.4. Entner et al. (2013); Hyttinen et al. (2015); Maathuis et al. (2015); Perkovic et al. (2017); Jaber et al. (2022)

Maathuis et al. (2015); Perkovic et al. (2017); Jaber et al. (2022) present some graphical conditions for the causal effect identifiability given a MAG or a PAG, and they return the causal effect *when the causal effect is identifiable* and return unidentifiable when it is not. In our paper, if the causal effect is identifiable in $\mathcal{P}$, then we return the causal effect according to their results. However, the main focus of our method is the case when it is unidentifiable. Different from the methods above which return unidentifiable, we need to return a set of possible causal effect which includes the causal effect if there exists some DAG represented by a MAG consistent with $\mathcal{P}$ associated with it. Hence the focus of our method and Maathuis et al. (2015); Perkovic et al. (2017); Jaber et al. (2022) are totally different.

Entner et al. (2013); Hyttinen et al. (2015) present the data-driven method to identify the causal effect *without learning a PAG first* which takes a huge cost. Similar to the methods ofMaathuis et al. (2015); Perkovic et al. (2017); Jaber et al. (2022), they return the causal effect when it is identifiable.

The causal effect identification method of Hyttinen et al. (2015); Jaber et al. (2022) is by do-calculus, which possibly identify some causal effects which are not identifiable via covariate adjustment. We leave the set determination of possible causal effects by do-calculus for future work.

C. Worst-case Complexity Analysis

We present a worst-case complexity analysis of our method and the baseline method LV-IDA (Malinsky & Spirtes, 2016) to determine the set of possible causal effects of X on Y in a complete graph including d vertices (in this case the PAG is a complete circle component without any arrowheads or tails). For LV-IDA, there are $3^{(d^2-d)/2}$ MAGs enumerated. For each MAG, we need to first judge whether the enumerated MAG is consistent with $\mathcal{P}$, it spends $\mathcal{O}(d^3)$ at least (Hu & Evans, 2020; Wienöbst et al., 2022). Then if the MAG is consistent with $\mathcal{P}$, we find D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in $\mathcal{O}(d)$ at least. Hence, the total complexity is $\mathcal{O}(d^4 3^{(d^2-d)/2})$. Another method of LV-IDA is obtaining the MAG consistent with $\mathcal{P}$ by valid transformation which guarantees that the obtained MAG is always consistent with $\mathcal{P}$ (Zhang & Spirtes, 2005; Tian, 2005). If we adopt this method, it is hard to analyze the complexity accurately as it is unavoidable that many repeatable MAGs are obtained in the process and it is hard to know when the transformation should stop. Anyway, even if we suppose that we will *never* obtain a repeatable MAG in the transformation process and there is an oracle tell us the point of time that we obtain all the MAGs consistent with $\mathcal{P}$, we could obtain a lower bound of computational complexity as $\Omega(d3^{(d^2-d)/2}) \times d^2 = \Omega(d^3 3^{(d^2-d)/2})$, where the extra d^2 is the complexity of judging the transformation characterization, which is shown by Lemma 1 of Zhang & Spirtes (2005).

Next we consider the complexity of our method. Note the first and second conditions of Thm. 2 can be determined in $\mathcal{O}(d^3)$ for a given block set $\mathbf{S}$. For the determination of the third condition, we exploit Lemma 5 to achieve it. The orientation in Alg. 2 and the testing of whether there are edges oriented with different directions or new unshielded colliders can be achieved in $\mathcal{O}(d^3)$. Hence the complexity of judging the three conditions of Thm. 2 given a maximal local MAG and a block set $\mathbf{S}$ is $\mathcal{O}(d^3)$.

As our method first determines all valid local structures at X, then finds adjustment sets in the MAGs valid to each maximal local MAG, we analyze the complexity of finding adjustment sets in the MAGs valid to a given maximal local MAG at first. Suppose in a maximal local MAG $\mathbb{M}$ which is obtained from $\mathcal{P}$ and $\mathbf{C}$ which dictates a local structure of X, there are i vertices with edges $*\!\!\rightarrow X$ and $d-i-1$ vertices with edges $*\!\!- X$. In this case, the complexity $T(i)$ of finding all adjustment sets in $\mathbb{M}$ is

$$T(i) \leq \sum_{k=0}^{i} C_i^k 2^{k+d-i-1}\mathcal{O}(d^3) = 2^{d-i-1}3^i\mathcal{O}(d^3),$$

where k denotes the number of vertices in $\mathbf{W}$ and C_i^k is because there are C_i^k sets $\mathbf{W}$ such that there are k vertices in $\mathbf{W}$. When there is k number of vertices in $\mathbf{W}$, since every vertex is adjacent to X, there must be $i-k$ vertex in $\bar{\mathbf{W}}$. Hence there are at most $d-1-i+k$ vertices in each block set $\mathbf{S}$ and thus there are at most $2^{k+d-i-1}$ sets of $\mathbf{S}$. $\mathcal{O}(d^3)$ is the complexity of judging the conditions in Thm. 2.

Next we consider the complexity $C(d)$ of finding adjustment sets in the MAGs consistent with $\mathcal{P}$ by our method. According to the result above, it directly conclude that the complexity is

$$C(d) \leq \sum_{i=0}^{d-1} C_{d-1}^{i} \mathcal{O}(d^3)(2^{d-i-1}3^{i}\mathcal{O}(d^3)) = 2^{d-1}\sum_{i=0}^{d-1}(3/2)^{i}\mathcal{O}(d^6) = \mathcal{O}(5^d d^6),$$

where C_{d-1}^{i} is because there are C_{d-1}^{i} sets $\mathbf{C}$ which dictates a local structure at X such that there are i vertices with edges $\ast\!\!\rightarrow X$, the first $\mathcal{O}(d^3)$ is the complexity of judging Prop. 2.

D. Proofs

D.1. Proof of Theorem 1

Proof. The proof of "if" statement is not hard. We could construct a DAG $\mathcal{D}$ by remaining the directed edges in $\mathcal{M}$ and adding a sub-structure $V_i \leftarrow L_{ij} \rightarrow V_j$ with a latent variable L_{ij} if there is $V_i \leftrightarrow V_j$ in $\mathcal{M}$. The paths from X to Y in $\mathcal{M}_{\underset{\sim}{X}}$ are *totally same* as the back-door paths from X to Y in $\mathcal{D}$, *i.e.*, the paths from X to Y in $\mathcal{D}_{\underline{X}}$. Since X is an ancestor of Y, it is evident that X and Y are not adjacent in $\mathcal{M}_{\underset{\sim}{X}}$. And $\mathcal{M}_{\underset{\sim}{X}}$ is also a MAG according to Prop. 3.5 and Corollary 4.6 of Richardson et al. (2002). By Lemma 4.1 of Maathuis et al. (2015), X and Y are m-separated by D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in $\mathcal{M}_{\underset{\sim}{X}}$. It is direct that X and Y are d-separated by D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in $\mathcal{D}_{\underline{X}}$. The reason is that the only difference between $\mathcal{M}_{\underset{\sim}{X}}$ and $\mathcal{D}_{\underline{X}}$ is that the bi-directed edges $V_i \leftrightarrow V_j$ in $\mathcal{M}_{\underset{\sim}{X}}$ are $V_i \leftarrow L_{ij} \rightarrow V_j$ in $\mathcal{D}_{\underline{X}}$, thus all the paths from X to Y in $\mathcal{M}_{\underset{\sim}{X}}$ and $\mathcal{D}_{\underline{X}}$ have the same colliders and non-colliders, and each latent variable $L_{i,j}$ in $\mathcal{D}_{\underline{X}}$ is a non-collider so that it does not influence whether a path is active or d-separated by D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ in $\mathcal{D}_{\underline{X}}$. Hence the Blocking condition of adjustment criterion in Def. 8 is fulfilled in $\mathcal{D}$ by D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. Due to D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \mathrm{De}(X, \mathcal{M}) = \emptyset$, the Forbidden set condition is evidently fulfilled. Hence D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ is an adjustment set in $\mathcal{D}$.

For the "only if" statement, we will first prove that D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ can block all non-causal paths from X to Y in DAG $\mathcal{D}$. And then we show D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \mathrm{De}(X, \mathcal{M}) = \emptyset$.

Denote D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ by $\mathbf{D}$. For contradiction, suppose in $\mathcal{D}$ there is a non-causal active path p as $X \leftarrow \cdots S_1, \cdots, S_2, \cdots, S_i, \cdots, S_n, \cdots, Y$ relative to $\mathbf{D}$, where only $S_0(=X), S_1, S_2, S_3, \cdots, S_i, \cdots, S_n, S_{n+1}(=Y)$ denotes the vertices in $\mathbf{V}(\mathcal{M})$ and there could be vertices except for $S_0(=X), S_1, S_2, S_3, \cdots, S_i, \cdots, S_n, S_{n+1}(=Y)$ in the path that do not belong to $\mathbf{V}(\mathcal{M})$. $\bigoplus$ denotes the concatenation of paths. For a path p, let $p[V_i, V_j]$ denote the sub-path of p from V_i to V_j.

We first present some facts given the conditions above.

Fact 1. All the colliders in p belongs to $\mathbf{D}$, and non-colliders belong to $\mathbf{V}(\mathcal{M})\backslash\mathbf{D}$.

Proof of Fact 1. It is directly concluded according to the definition of active path.

Fact 2. All the vertices in $\mathbf{D}$ are ancestors of either X or Y in $\mathcal{M}_{\underset{\sim}{X}}$.

Proof of Fact 2. It is concluded according to the definition of D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$.

Fact 3. If in $\mathcal{D}$ there exists a directed path from V to V' where $V \in \mathbf{V}(\mathcal{M})$, $V' \in \mathbf{D}$, then V is ancestor of either X or Y in $\mathcal{M}_{\underset{\sim}{X}}$.

Proof of Fact 3. It is concluded directly by Fact 2 and the algorithm to obtain a MAG based on $\mathcal{D}$ in Appendix A.2.

Fact 4. Suppose in p there is a sub-path $X \leftarrow \cdots \rightarrow S_1 \leftarrow \cdots \rightarrow S_2 \leftarrow \cdots\cdots \rightarrow S_{k-1} \leftarrow \cdots S_k$ where $S_1, S_2, \cdots, S_{k-1} \in \mathbf{D}$, if S_k is an ancestor of X or Y in $\mathcal{M}_{\underset{\sim}{X}}$, then X cannot be an ancestor of S_k in $\mathcal{D}$.

Proof of Fact 4. Suppose X is an ancestor of S_k in $\mathcal{D}$. According to the condition, S_k is an ancestor of Y in $\mathcal{M}_{\underset{\sim}{X}}$. Hence there is a directed path from X to Y across S_k in $\mathcal{D}$. Denote the directed path from S_k to Y in $\mathcal{D}$ by p_1. According the adjustment criterion, if there exists an adjustment set $\mathbf{W}$ comprised of some vertices in $\mathbf{V}(\mathcal{M})$, $\mathbf{W}$ cannot contain the vertices in p_1. We consider the path $p[X, S_k] \bigoplus p_1$, where $S_1, S_2, \cdots, S_{k-1}$ in $p[X, S_k]$ are colliders and S_k is non-colliders due to the directed path from S_k to Y. $\mathbf{W}$ is required to block this path, hence there exists some vertex V in $S_1, S_2, \cdots, S_{k-1}$ such that all the descendants of V does not belong to $\mathbf{W}$. Suppose the nearest vertex to X in $p[X, S_k]$ whose descendants do not belong to $\mathbf{W}$ is S_j. If there is a directed path p_2 from S_j to Y in $\mathcal{D}$ that does not go through

X, then $p[X, S_j] \bigoplus p_2$ is active relative to $\mathbf{W}$, contradiction with the blocking condition of adjustment criterion; if there is not a directed path from S_j to Y in $\mathcal{D}$ that does not go through X, there must be a directed path p_3 from S_j to X in $\mathcal{D}$ since S_j is an ancestor of either X or Y in $\mathcal{D}$. In this case we consider a new path $p_3 \bigoplus p[S_j, S_k] \bigoplus p_2$. To block this path, there is another vertex V in $S_{j+1}, S_{j+2}, \cdots, S_{k-1}$ such that all the descendants of V does not belong to $\mathbf{W}$. Suppose the nearest vertex to S_j whose descendants do not belong to $\mathbf{W}$ in $p[S_j, S_k]$ is S_t. If there is a directed path p_4 from S_t to Y that does not go through X, there is an active path $p_3 \bigoplus p[S_j, S_t] \bigoplus p_4$ relative to $\mathbf{W}$, contradicting with the blocking condition of adjustment criterion, hence there is directed path p_5 from S_t to X where each vertex does not belong to $\mathbf{W}$. As the process above, we consider $p_5 \bigoplus p[S_t, S_k] \bigoplus p_2$ instead. Repeat the process above, if there is not a contradiction, there must be a vertex S_m such that there is a directed path p_6 from S_m to X where each vertex does not belong to $\mathbf{W}$, and for each non-endpoint vertex V in $p[S_m, S_k]$ there is at least one descendant of V belonging to $\mathbf{W}$. In this case we have an active non-causal path $p_6 \bigoplus p[S_m, S_k] \bigoplus p_2$, contradicting with the blocking condition of adjustment criterion. Hence, X cannot be an ancestor of S_k.

Fact 5. If $S_1, S_2, \cdots, S_k$ are colliders in p, then there is a collider path from X to S_{k+1} beginning with an arrowhead at X in $\mathcal{M}$, *i.e.*, in the form of $X \leftrightarrow \cdots \leftrightarrow \leftarrow\!\!* S_{k+1}$.

Proof of Fact 5. According to fact 1, $S_1, \cdots, S_k \in \mathbf{D}$. We consider the sub-path $p[X, S_{k+1}]$. Note the sub-path $p[S_i, S_{i+1}], \forall 0 \le i \le k$ is an inducing path relative to $\langle \mathbf{V}(\mathcal{D}) \backslash \mathbf{V}(\mathcal{M}), \emptyset \rangle$. Hence we can always find a longest inducing path relative to $\langle \mathbf{V}(\mathcal{D}) \backslash \mathbf{V}(\mathcal{M}), \emptyset \rangle$ which is a sub-path of p starting by X. Denote the longest inducing path by $p[X, S_i]$. If $i = k+1$, there is evidently $X \leftarrow\!\!* S_{k+1}$ because $p[X, S_{k+1}]$ is an inducing path and fact 4 (there cannot be an edge $X \to S_{k+1}$ in $\mathcal{M}$ according to fact 4), the result holds. Hence we only consider $i \le k$ below. We will prove there is (a) $X \leftrightarrow S_i$; (b) $S_{i+1} *\!\!\rightarrow S_i$ in $\mathcal{M}$.

For the proof of (a), $X \leftarrow\!\!* S_i$ in $\mathcal{M}$ can be directly concluded because $p[X, S_i]$ is an inducing path and fact 4. If $S_i \to X$ in $\mathcal{M}$, S_i is an ancestor of X in $\mathcal{D}$, thus $p[X, S_i] \bigoplus p[S_i, S_{i+1}]$ is also an inducing path because S_i is a collider in p, contradicting with the premise that $p[X, S_i]$ is the longest inducing path staring by X. Hence there can only be $X \leftrightarrow S_i$ in $\mathcal{M}$. For the proof of (b), suppose $S_{i+1} \leftarrow S_i$ in $\mathcal{M}$, hence S_{i+1} is a descendant of S_i in $\mathcal{D}$. Considering the inducing path $p[X, S_i]$, each collider is an ancestor of either X or S_i, thus each collider is an ancestor of either X or S_{i+1}. Hence $p[X, S_i] \bigoplus p[S_i, S_{i+1}]$ is also an inducing path because S_i is a collider in p, contradicting with the premise that $p[X, S_i]$ is the longest inducing path staring by X. Hence S_{i+1} is not a descendant of S_i in $\mathcal{D}$, thus in $\mathcal{M}$ there is $S_i \leftarrow\!\!* S_{i+1}$.

According to the result above, if $i = k$, there is $X \leftrightarrow S_k \leftarrow\!\!* S_{k+1}$ in $\mathcal{M}$, we get the desired result. Hence we consider $i < k$ below.

Then, we could find the longest inducing path relative to $\langle \mathbf{V}(\mathcal{D}) \backslash \mathbf{V}(\mathcal{M}), \emptyset \rangle$ which is a sub-path of $p[S_i, S_{k+1}]$ starting by S_i. Suppose the path is $p[S_i, S_j]$. S_i and S_j is adjacent in $\mathcal{M}$ due to the inducing path. We first prove there is not $S_i \to S_j$ in $\mathcal{M}$. Otherwise, all the vertices in $p[S_i, S_j]$ are ancestors of S_j according to the definition of inducing path. Since all the vertices in $p[X, S_i]$ are ancestors of X or S_i, we have a new inducing path $p[X, S_j]$ since each vertex is an ancestor of X or S_j and meanwhile a collider because $S_1, S_2, \cdots, S_k$ are ancestors in p. It contradicts the fact that $p[X, S_i]$ is the longest inducing path starting by X. Hence there is $S_i \leftarrow\!\!* S_j$ in $\mathcal{M}$. If $j = k+1$, we get the desired result. If $j < k+1$, we will prove there is $S_i \leftrightarrow S_j$ in $\mathcal{M}$. Otherwise, if it is $S_j \to S_i$ in $\mathcal{M}$, S_j is an ancestor of S_i in $\mathcal{D}$, hence $p[S_i, S_j] \bigoplus p[S_j, S_{j+1}]$ is an inducing path relative to $\langle \mathbf{V}(\mathcal{D}) \backslash \mathbf{V}(\mathcal{M}), \emptyset \rangle$, contradicting with the fact that $p[S_i, S_j]$ is the longest inducing path that is a sub-path of $p[S_i, S_{k+1}]$ starting by S_i. Hence, there is $X \leftrightarrow S_i \leftrightarrow S_j$ in $\mathcal{M}$. Similarly, we can find a longest inducing path $p[S_j, S_t]$ starting by S_j such that $S_j \leftarrow\!\!* S_t$ (otherwise $p[S_i, S_j] \bigoplus p[S_j, S_t]$ is an inducing path), and in $\mathcal{M}$ there is $X \leftrightarrow S_i \leftrightarrow S_j \leftarrow\!\!* S_{k+1}$ if $t = k+1$ and there is $X \leftrightarrow S_i \leftrightarrow S_j \leftrightarrow S_t$ if $t < k+1$. Repeat the process above, we could always find a collider path $X \leftrightarrow \cdots \leftarrow\!\!* S_{k+1}$ in $\mathcal{M}$.

With the five facts above, we first prove that all of $S_0, S_1, S_2, \cdots, S_n$ are colliders in p by mathematical induction. Since p is active relative to $\mathbf{W}$, where only $S_0, S_1, S_2, \cdots, S_n, S_{n+1}$ are observed variables in $\mathcal{M}$. It is evident that there is an edge between S_i and S_{i+1} in $\mathcal{M}$, for $\forall 0 \le i \le n$.

We first prove that S_1 is a collider on the path. If not, the path is either $X \leftarrow \cdots \leftarrow S_1 \leftarrow \cdots Y$ or $X \leftarrow \cdots \to S_1 \to \cdots Y$. For the first case, it is evident that S_1 is an ancestor of X in $\mathcal{M}_{\underset{\sim}{X}}$; for the second case, if the path from S_1 to Y is directed, S_1 is ancestor of Y, otherwise there is a directed path from S_1 to a collider in p, according to fact 1 and fact 2 the collider is an ancestor of either X or Y in $\mathcal{M}_{\underset{\sim}{X}}$, hence S_1 is an ancestor of either X or Y in $\mathcal{M}_{\underset{\sim}{X}}$. That is, S_1 is always an ancestor of X or Y in $\mathcal{M}_{\underset{\sim}{X}}$. Note X is adjacent to S_1 in $\mathcal{M}$. According to fact 4, S_1 cannot be a descendant of X in $\mathcal{D}$. Hence there is $X \leftarrow\!\!* S_1$ in $\mathcal{M}$. According to the definition of $\mathbf{D} = \text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}})$, we conclude $S_1 \in \mathbf{D}$. In this case if S_1 is not a collider in p, it does not belong to $\mathbf{D}$ by fact 1, contradicting with $S_1 \in \mathbf{D}$. Hence S_1 is a collider in p. Since the path p is active relative to $\mathbf{D}$, there is $S_1 \in \mathbf{D}$.

Then we suppose any vertex S_i in $S_1, S_2, \cdots, S_k, k \geq 2$ is a collider in p and $S_i \in \mathbf{D}$. We could prove S_{k+1} is a collider in p and $S_{k+1} \in \mathbf{D}$. If not, the sub-path from S_k to Y in p is either $S_k \leftarrow \cdots \leftarrow S_{k+1} \leftarrow \cdots Y$ or $S_k \leftarrow \cdots \rightarrow S_{k+1} \rightarrow \cdots Y$. For the first case, S_{k+1} is an ancestor of X or Y in $\mathcal{M}_{\underset{\sim}{X}}$ according to fact 3; for the second case, if the path from S_{k+1} to Y is directed, S_{k+1} is ancestor of Y, otherwise there is a directed path from S_{k+1} to a collider in p, according to fact 1 and fact 2 the collider is an ancestor of either X or Y in $\mathcal{M}_{\underset{\sim}{X}}$, hence S_{k+1} is an ancestor of either X or Y in $\mathcal{M}_{\underset{\sim}{X}}$. Hence in both cases, S_{k+1} is an ancestor of X or Y in $\mathcal{M}_{\underset{\sim}{X}}$. By fact 5, there is a collider path $X \leftrightarrow \cdots \leftarrow\!\!* S_{k+1}$ in $\mathcal{M}$. Hence $S_{k+1} \in \mathbf{D}$ according to the definition of $\mathbf{D} = \text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. In this case if S_{k+1} is not a collider in p, it does not belong to $\mathbf{D}$ by fact 1, contradicting with $S_{k+1} \in \mathbf{D}$. Hence S_{k+1} is a collider in p. Since the path p is active relative to $\mathbf{D}$, there is $S_{k+1} \in \mathbf{D}$. The induction step completes.

By induction, we conclude that $S_1, S_2, \cdots, S_n$ are colliders in p. With fact 4, X cannot be an ancestor of $S_{n+1} (= Y)$ in $\mathcal{D}$, thus X is not an ancestor of Y in $\mathcal{M}$, contradicting with the condition $X \in \text{Anc}(Y, \mathcal{M})$. Hence there is always a contradiction if there is an active non-causal path p relative to $\mathbf{D}$ from X to Y. Hence, we conclude that there is not an active non-causal path relative to $\mathbf{D}$ from X to Y in DAG $\mathcal{D}$.

Next we prove D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ does not contain a vertex in De$(X, \mathcal{D})$. Suppose $V \in \text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{D})$, suppose there is $X (= V_0) \leftrightarrow V_1 \leftrightarrow \cdots \leftrightarrow V_{t-1} \leftarrow\!\!* V (= V_t)$ in $\mathcal{M}_{\underset{\sim}{X}}$ where each non-endpoint is an ancestor of X or Y in $\mathcal{M}_{\underset{\sim}{X}}$. Evidently each non-endpoint is an ancestor of either X or Y in $\mathcal{D}$, and there is a path p_1 in the form of $X \leftarrow \cdots \rightarrow V_1 \leftarrow \cdots \rightarrow V_{t-1} \leftarrow \cdots V$ in $\mathcal{D}$ where each sub-path $p_1[V_i, V_{i+1}], 0 \leq i \leq t-1$ is an inducing path relative to $\langle \mathbf{V}(\mathcal{D}) \backslash \mathbf{V}(\mathcal{M}), \emptyset \rangle$ and p_2 is a directed path from V to Y. If there exists an adjustment set by adjustment criterion, the set $\mathbf{W}$ does not contain the vertex in p_2. To block the path $p_1 \bigoplus p_2$, there exists some vertex in p_1 that does not belong to $\mathbf{W}$. Similar to the part of fact 4, we could prove there is always a path unblocked by $\mathbf{W}$, contradicting the Blocking condition of adjustment criterion. Hence we prove D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{D}) = \emptyset$, the Forbidden set condition in adjustment criterion is evidently fulfilled. Hence D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ fulfills adjustment criterion in $\mathcal{D}$. We get the desired result. ☐

D.2. Proof of Theorem 2

We first show some facts before presenting the proof for Thm. 2.

Lemma 1. *Given a maximal local MAG $\mathbb{M}$ of X, the following properties are satisfied:*

(Closed) *$\mathbb{M}$ is closed under the orientation rules;*

(Invariant) *The arrowheads and tails in $\mathbb{M}$ are invariant in all the MAGs consistent with $\mathcal{P}$ with the local marks at X.*

(Chordal) *the circle component in $\mathbb{M}$ is chordal;*

(Balanced) *for any three vertices 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 $A \leftrightarrow C$);*

(Complete) *For each circle at vertex A on any edge $A \circ\!\!-\!\!* B$ in $\mathbb{M}_s$, there exist MAGs $\mathcal{M}_1$ and $\mathcal{M}_2$ consistent with $\mathcal{P}$ and BK regarding $V_1, \ldots, V_s$ with $A \leftarrow\!\!* B \in \mathbf{E}(\mathcal{M}_1)$ and $A \rightarrow B \in \mathbf{E}(\mathcal{M}_2)$;*

(P6) *We can always obtain a MAG consistent with $\mathbb{M}$ by transforming the circle component into a DAG without unshielded colliders and transforming $A \circ\!\!\rightarrow B$ as $A \rightarrow B$.*

Proof. The first five properties directly follow Thm.1 of Wang et al. (2022b). The property $P6$ follows Lemma 16.1 of Wang et al. (2022b). ☐

Since the proof of existence of MAGs is involved in Thm. 2, we first present an algorithm to obtain a MAG valid to $\mathbb{M}$ in Alg. 2, with the proof for the validity of MAG construction in Lemma 7.

At first, we present Lemma 2, 3 and 4 following Wang et al. (2022b).

Algorithm 2: Orient a maximal local MAG of X as a MAG

input: Maximal local MAG $\mathbb{M}$, potential adjustment set $\mathbf{W}$ and corresponding $\bar{\mathbf{W}}$ according to Def. 4, block set $\mathbf{S}$

1: for $\forall K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $\forall T \in \mathbf{S}$ such that $K \circ\!\!\rightarrow\!* T$ in $\mathbb{M}$, orient it as $K \leftrightarrow\!* T$ (the mark at T remains);

2: update the subgraph $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ as follows until no feasible updates: for any two vertices V_i and V_j such that $V_i \circ\!\!-\!\!\circ V_j$, orient it as $V_i \rightarrow V_j$ if (1) $\mathcal{F}_{V_i} \backslash \mathcal{F}_{V_j} \neq \emptyset$ or (2) $\mathcal{F}_{V_i} = \mathcal{F}_{V_j}$ as well as there is a vertex $V_k \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ not adjacent to V_j such that $V_k \rightarrow V_i \circ\!\!-\!\!\circ V_j$, where $\mathcal{F}_{V_i} = \{V \in \mathbf{S} \mid V *\!\!-\!\!\circ V_i \text{ in } \mathbb{M}\}$;

3: orient the circles on the remaining $\circ\!\!\rightarrow$ edges as tails;

4: in subgraph $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$, orient the circle component into a DAG without new unshielded colliders;

5: in subgraph $\mathbb{M}[-\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$, orient the circle component into a DAG without new unshielded colliders.

output: A MAG $\mathcal{M}$

Lemma 2. *Consider a maximal local MAG $\mathbb{M}$. If there is a possible directed path from A to B in $\mathbb{M}$, then there is a minimal possible directed path from A to B in $\mathbb{M}$.*

Proof. Suppose the possible directed path $p = \langle V_0(= A), V_1, \dots, V_m(= B)\rangle$. If p is minimal, the result trivially holds. If not, we can always find a subpath $\langle V_i, V_{i+1}, \dots, V_j\rangle, j - i \geq 2$ such that any non-consecutive vertices are not adjacent except for an edge between V_i and V_j. We will show the impossibility of $V_i \leftrightarrow\!* V_j$ in $\mathbb{M}$. Suppose $V_i \leftrightarrow\!* V_j$ in $\mathbb{M}$. Note there is a circle/tail at V_i on the edge between V_i and V_{i+1} due to the possible directed path p. If $j - i = 2$, there is always an edge $V_{i+1} \leftrightarrow\!* V_{i+2}(= V_j)$ due to the balance/closed property of $\mathbb{M}$, contradicting the possible directed path p. If $j - i > 2$, due to the non-adjacency of the V_j and V_{i+1}, there is either $V_i \rightarrow V_{i+1} \rightarrow \dots V_j$ or $V_i \leftrightarrow\!* V_{i+1}$ identified in $\mathcal{P}$. The latter case is impossible due to the possible directed path p. For the former case, there is an almost directed or directed cycles, contradiction. Hence, the edge between V_i and V_j is either $V_i \rightarrow V_j$ or $V_i \circ\!\!\rightarrow\!* V_j$, we thus find a shorter possible directed path $\langle V_0, V_1, \dots, V_i, V_j, V_{j+1}, \dots, V_m\rangle$ in $\mathbb{M}$. Repeat this process until obtaining a possible directed path such that there is not a proper sub-structure where any non-consecutive vertices are not adjacent except for an edge between endpoints. This path is a minimal possible directed path. □

Lemma 3. *Consider a maximal local MAG $\mathbb{M}$. If there is $A *\!\!\rightarrow B$ in $\mathbb{M}$, then there is an edge as $A *\!\!\rightarrow V$ for any V in a connected circle component with B in $\mathbb{M}$, and A and B are not in a connected circle component.*

Proof. It is a direct conclusion of the balanced property of $\mathbb{M}$. We first consider any vertex V_1 that has a circle edge with B, there is $A *\!\!\rightarrow B \circ\!\!-\!\!\circ V_1$ in $\mathbb{M}$. According to the balanced property, there is $A *\!\!\rightarrow V_1$. Similarly, we can conclude that the result holds for all the vertices in a circle component with B. Hence A and B cannot be in a connected circle component. □

Lemma 4. *Consider a maximal local MAG $\mathbb{M}$ of X and a block set $\mathbf{S}$. Denote $\mathcal{F}_{V_i} = \{V \in \mathbf{S} \mid V *\!\!-\!\!\circ V_i \text{ in } \mathbb{M}\}$ for $\forall V_i \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. For an edge $J \circ\!\!-\!\!\circ K$ in $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$, if it is oriented as $J \rightarrow K$ in the second step of Alg. 2, then there is a vertex $V_m \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ such that there is a minimal path $V_m \circ\!\!-\!\!\circ \dots \circ\!\!-\!\!\circ V_1(= J) \circ\!\!-\!\!\circ V_0(= K), m \geq 1$ in $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ where $\mathcal{F}_{V_m} \supset \mathcal{F}_{V_{m-1}} = \cdots = \mathcal{F}_{V_0}$.*

Proof. A directed edge $J \rightarrow K$ is oriented in the second step only if in two situations: (1) $\mathcal{F}_K \subset \mathcal{F}_J$; (2) $\mathcal{F}_K = \mathcal{F}_J$ and there is another vertex $L \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ that is not adjacent to K and there is $L \rightarrow J$. Note $L \rightarrow J$ can only be oriented in the second step since (1) the edges connecting $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $\mathbf{S}$ is not oriented in the first step; (2) $L \rightarrow J$ cannot appear in $\mathbb{M}$ for otherwise either $J \rightarrow K$ or $J \leftrightarrow\!* K$ is identified in $\mathbb{M}$ due to the closed property of $\mathbb{M}$.

If $\mathcal{F}_{V_0} \subset \mathcal{F}_{V_1}$, there is a desired path where $m = 1$. If $\mathcal{F}_{V_0} = \mathcal{F}_{V_1}$, we could find $V_2 \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ that is not adjacent to V_0 and there is $V_2 \rightarrow V_1$ oriented in the second step. Similarly, we conclude either $\mathcal{F}_{V_1} \subset \mathcal{F}_{V_2}$, in which case there is a desired path where $m = 2$; or $\mathcal{F}_{V_1} = \mathcal{F}_{V_2}$, in which case there is $V_3 \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ that is not adjacent to V_1 and there is $V_3 \rightarrow V_2$ oriented in the second step. Repeat the process and we can always find an uncovered path $V_m \circ\!\!-\!\!\circ \dots \circ\!\!-\!\!\circ V_1(= J) \circ\!\!-\!\!\circ V_0(= K), m \geq 1$ in $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ where $\mathcal{F}_{V_0} = \cdots = \mathcal{F}_{V_{m-1}} \subset \mathcal{F}_{V_m}$. Finally, it suffices to prove that the path is minimal. If not, there exists a sub-structure $V_i \circ\!\!-\!\!\circ V_{i+1} \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_j, j > i + 2$ where any two non-consecutive vertices are not adjacent except for an edge between V_i and V_j. Since only the edges containing vertices in $\mathbf{S}$ are transformed in the first step, if there is a non-circle edge between V_i and V_j before the second step, the edge is non-circle in $\mathbb{M}$, in which case V_i and V_j cannot be in a circle component according to Lemma 3, contradicting with the circle path comprised of $V_i, V_{i+1}, \ldots, V_j$. Hence there is $V_i \circ\!\!-\!\!\circ V_j$ in $\mathbb{M}$, in which case the chordal property of $\mathbb{M}$ is not fulfilled. Thus the path can only be minimal. $\square$

Lemma 5. *Given the first two conditions of Thm. 2 fulfilled, there are neither edges oriented with different directions nor new unshielded colliders generated in the second step of Alg. 2 if and only if* $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ *is bridged relative to* $\mathbf{S}$ *in* $\mathbb{M}$.

Proof. We first prove the "if" statement. We prove that there are neither edges oriented with different directions nor new unshielded colliders generated in the second step of Alg. 2, respectively. For simplicity, denote $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ by $\mathbb{M}_1$.

Suppose we orient both $J \rightarrow K$ and $J \leftarrow K$ in the second step. According to Lemma 4, if we orient $J \rightarrow K$ in the second step, there is a minimal circle path $V_0 \circ\!\!-\!\!\circ V_1 \circ\!\!-\!\!\circ \cdots V_{m-1}(=J) \circ\!\!-\!\!\circ V_m(=K)$ where $\mathcal{F}_{V_0} \supset \mathcal{F}_{V_1} = \cdots = \mathcal{F}_{V_m}$. If we also orient $J \leftarrow K$ in the second step, there is another minimal circle path $V_{m-1}(=J) \circ\!\!-\!\!\circ V_m(=K) \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_n, n > m$ in $\mathbb{M}_1$ where $\mathcal{F}_{V_{m-1}} = \mathcal{F}_{V_m} = \cdots = \mathcal{F}_{V_{n-1}} \subset \mathcal{F}_n$. Note V_{m+1} is adjacent to V_m but not adjacent to V_{m-1}, while V_{m-2} is adjacent to V_{m-1} but not adjacent to V_m, hence $V_{m-2}, V_{m-1}, V_m, V_{m+1}$ are distinct vertices. According to Lemma 3, there cannot be non-circle edge between the variables in the circle path. Also note no circle edges in $\mathbb{M}_1$ are oriented in the first step. Hence the circle component in $\mathbb{M}_1$ after the first step is still chordal. Hence $V_0 \circ\!\!-\!\!\circ V_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_n$ is also a minimal circle path, otherwise there is a circle cycle whose length is larger than 3 without a chord because this cycle must contain $V_{m-2}, V_{m-1}, V_m, V_{m+1}$ where V_{m-2} is not adjacent to V_m and V_{m-1} is not adjacent to V_{m+1}. Since $V_0, \cdots, V_n \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, V_0 and V_n are not bridged relative to $\mathbf{S}$, contradicting with that $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is bridged relative to $\mathbf{S}$ in $\mathbb{M}$.

Suppose there is a new unshielded collider $A \rightarrow B \leftarrow C$ generated in the second step. According to Lemma 4 there is a minimal path $F_1 \rightarrow \cdots, \rightarrow F_m(=A) \rightarrow B, m \geq 2$ and a minimal path $V_1 \rightarrow \cdots V_n(=C) \rightarrow B, n \geq 2$ such that $\mathcal{F}_{F_1} \supset \mathcal{F}_{F_2} = \cdots = \mathcal{F}_B$ and $\mathcal{F}_{V_1} \supseteq \mathcal{F}_{V_2} = \cdots = \mathcal{F}_B$. A and C are evidently different vertices that are not adjacent. In this case there is a circle path $p: F_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ F_m(=A) \circ\!\!-\!\!\circ B \circ\!\!-\!\!\circ V_n(=C) \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_1$ in $\mathbb{M}$ such that $\mathcal{F}_{F_1} \supset \mathcal{F}_{F_2} = \cdots = \mathcal{F}_B = \cdots = \mathcal{F}_{V_2} \subset \mathcal{F}_{V_1}$. According to Lemma 3, there are no non-circle edges between the variables in p. In this case, there is always a minimal circle path from F_1 to V_1 such that F_1 and V_1 are not bridged relative to $\mathbf{S}$ in $\mathbb{M}$, contradiction.

We then prove the "only if" statement. Suppose $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is not bridged relative to $\mathbf{S}$ in $\mathbb{M}$.

If $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is not bridged relative to $\mathbf{S}$ in $\mathbb{M}$, we will prove the result by showing that there are either edges oriented with difference direction or new unshielded colliders generated in the second step of Alg. 2.

Suppose two vertices J, K in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ are not bridged relative to $\mathbf{S}$ due to the minimal circle path $J(=V_0) \circ\!\!-\!\!\circ V_1 \cdots V_n \circ\!\!-\!\!\circ K(=V_{n+1})$ in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$. There are two possible cases (they possibly happen simultaneously). One is that there exists $0 \leq s \leq n$ such that $\mathcal{F}_{V_s} \not\subseteq \mathcal{F}_{V_{s+1}}$ and $\mathcal{F}_{V_{s+1}} \not\subseteq \mathcal{F}_{V_s}$. The other is that there exists $1 \leq s \leq n$ such that $\mathcal{F}_{V_s} \subset \mathcal{F}_{V_{s-1}}$ and $\mathcal{F}_{V_s} \subset \mathcal{F}_{V_{s+1}}$.

For the first case, suppose there are two vertices $T_1, T_2 \in \mathbf{S}$ such that $T_1 \in \mathcal{F}_{V_s} \backslash \mathcal{F}_{V_{s+1}}$ and $T_2 \in \mathcal{F}_{V_{s+1}} \backslash \mathcal{F}_{V_s}$. In the second step of Alg. 2, we will orient both $V_i \rightarrow V_j$ and $V_i \leftarrow V_j$. For the second case, suppose a vertex $T_1 \in \mathcal{F}_{V_{s-1}} \backslash \mathcal{F}_{V_s}$ and a vertex $T_2 \in \mathcal{F}_{V_{s+1}} \backslash \mathcal{F}_{V_s}$. In the second step of Alg. 2, there is $V_{s-1} \rightarrow V_s \leftarrow V_{s+1}$ oriented. As the path is minimal, a new unshielded collider is generated. $\square$

Lemma 6. *Given the three conditions in Thm. 2 fulfilled with* $\mathbf{S}$*, for any* $K \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ *and any* $T_1, T_2 \in \mathbf{S}$ *such that there is* $T_1 *\!\!-\!\!\circ K \circ\!\!-\!\!* T_2$ *in* $\mathbb{M}$*,* T_1 *is adjacent to* T_2*.*

Proof. If $K \in \bar{\mathbf{W}}$, according to Def. 4, $T_1, T_2 \in \mathcal{S}_K$. According to the second condition of Thm. 2, T_1 is adjacent to T_2. It suffices to prove the result for $K \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \backslash \bar{\mathbf{W}}$.

According to the definition of $\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \backslash \bar{\mathbf{W}}$, we can always find a minimal possible directed path p_1 from a vertex in $\bar{\mathbf{W}}$ to K in $\mathbb{M}[-\mathbf{S}]$ where each non-endpoint vertex does not belong to $\bar{\mathbf{W}}$. We suppose p_1 is from $F_t \in \bar{\mathbf{W}}$ to K comprised of $V_0(=F_t), V_1, \cdots, V_s(=K)$. And according to the definition of $\bar{\mathbf{W}}$, there is a collider path as $X(=F_0) \leftrightarrow \cdots F_{t-1} \leftarrow\!\!\circ (\text{ or } \leftrightarrow) F_t$, where each non-endpoint belongs to $\mathbf{W}$. Next we will prove that there is always an edge $V_0(=F_t) \circ\!\!\rightarrow\!\!* T_1$. At first, we name a fact. For any vertex $V_i, 0 \leq i \leq s$, there is not an edge as $V_i \rightarrow T_1$ in $\mathbb{M}$ due to the first condition of Thm. 2. We discuss the possible edge between F_t and V_1: **(1).** $F_t \rightarrow V_1$; **(2).** $F_t \circ\!\!-\!\!\circ V_1$; **(3).** $F_t \circ\!\!\rightarrow V_1$. We will prove that for any case, there is $T_1 \ast\!\!-\!\!\circ V_0(= F_t)$.

(1). If there is $F_t \rightarrow V_1$ in $\mathbb{M}$, there is $F_t \rightarrow V_1 \rightarrow \cdots \rightarrow V_s(= K)$ in $\mathbb{M}$ since p_1 is a minimal possible directed path in $\mathbb{M}$ and the closed property of $\mathbb{M}$. Note there is $V_s(= K) \circ\!\!\rightarrow\!\!\ast\, T_1$ in $\mathbb{M}$, T_1 is adjacent to V_{s-1} because there is not a structure as $A \rightarrow B \circ\!\!-\!\!\ast\, C$ in $\mathbb{M}$ where A and C are not adjacent as a result of closed property of $\mathbb{M}$. Since there is $T_1 \ast\!\!-\!\!\circ V_s(= K)$ in $\mathbb{M}$ and due to $P2$ of $\mathbb{M}$ in Lemma 1, there must be $T_1 \leftarrow V_{s-1}$ or $T_1 \ast\!\!-\!\!\circ V_{s-1}$. Since for any vertex $V_i, 0 \leq i \leq s$ there is not an edge as $T_1 \leftarrow V_i$ in $\mathbb{M}$, the edge can only be $T_1 \ast\!\!-\!\!\circ V_{s-1}$. Repeat this process for $V_{s-1}, V_{s-2}, \cdots, F_t(= V_0)$, we can prove that there is $T_1 \ast\!\!-\!\!\circ V_0(= F_t)$.

(2). If there is $F_t \circ\!\!-\!\!\circ V_1$ in $\mathbb{M}$, p_1 must be in the form of $V_0(F_t) \circ\!\!-\!\!\circ V_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_j \circ\!\!\rightarrow V_{j+1} \rightarrow \cdots \rightarrow V_s, 1 \leq j \leq s$. Note there is a possible directed path $p_1[V_1, V_s] \bigoplus V_s \circ\!\!-\!\!\ast\, T_1$ from V_1 to T_1, V_1 is a possible ancestor of T_1 in $\mathbb{M}$. And according to the definition of $\mathbf{S}$, V_1 is a possible ancestor of Y in $\mathbb{M}$. At first, V_1 cannot be adjacent to F_{t-1}. Otherwise, there is $F_{t-1} \leftarrow\!\!\ast V_1$ due to the balanced property of $\mathbb{M}$ and $F_{t-1} \leftarrow\!\!\ast F_t \circ\!\!-\!\!\circ V_1$. And since V_1 is a possible ancestor of Y in $\mathbb{M}$, $V_1 \in \mathbf{W} \cup \bar{\mathbf{W}}$. However, V_1 cannot belong to $\mathbf{W}$ for otherwise $V_1 \in \mathbf{S}$ according to Def. 4 and thus V_1 cannot be in $\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])\backslash\bar{\mathbf{W}}$. V_1 cannot belong to $\bar{\mathbf{W}}$ since p_1 does not go through any vertex in $\bar{\mathbf{W}}$ except for the endpoints. Hence V_1 cannot belong to $\mathbf{W} \cup \bar{\mathbf{W}}$, contradiction. Hence V_1 is not adjacent to F_{t-1}. In this case, according to the third condition of Thm. 2, there is $F_t \circ\!\!-\!\!\ast\, T_1$ in $\mathbb{M}$, otherwise F_t and K are not bridged relative to $\mathbf{S}$ in light of the facts that (1) $p_1[F_t, V_s]$ is a minimal circle path; (2) V_1 is not adjacent to $F_{t-1} \in \mathbf{S}$; (3) there is $V_s \circ\!\!-\!\!\ast\, T_1 \in \mathbf{S}$ in $\mathbb{M}$.

(3). If there is $F_t \circ\!\!\rightarrow V_1$, p_1 is as $F_t \circ\!\!\rightarrow V_1 \rightarrow \cdots V_s$ due to the closed property of $\mathbb{M}$. We can conclude that there is $V_0(= F_t) \circ\!\!-\!\!\ast\, T_1$ with the same proof for the case **(1)**.

Hence, we conclude that there is always $V_0(= F_t) \circ\!\!-\!\!\ast\, T_1$. Similarly, there is always $V_0(= F_t) \ast\!\!-\!\!\circ T_2$. In this case, according to Def. 4, there is $T_1, T_2 \in \mathcal{S}_{F_t}$. Hence that T_1 is adjacent to T_2 directly follows the second condition of Thm. 2. $\square$

Lemma 7. *Given the three conditions in Thm. 2 fulfilled with* $\mathbf{S}$*, we can obtain a MAG consistent with* $\mathbb{M}$ *by Alg. 2.*

Proof. The whole proof in this part is comprised of two parts. **A.** we can obtain a unique graph $\mathcal{H}$ without circles by Alg. 2. **B.** $\mathcal{H}$ is a MAG consistent with $\mathbb{M}$. The unique graph means that we will not orient an edge with two directions.

A. We can obtain a unique graph $\mathcal{H}$ without circles by Alg. 2.

We have shown in Lemma 5 that there are neither edges oriented with different directions nor new unshielded colliders generated in the second step of Alg. 2. In the following, we will first prove that **A.1.** there are not circle edges connecting $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ and $\mathbb{M}[-\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$. Then, we prove that **A.2.** the circle component in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ and **A.3.** the circle component in $\mathbb{M}[-\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ are chordal, respectively. Given these results, the fourth and fifth steps can be executed since every chordal graph has a perfect elimination order, through which we can orient the chordal graph as a DAG without unshielded colliders. We conclude that we can obtain a unique graph $\mathcal{H}$ without circles by Alg. 2.

A.1. Suppose there is a circle edge $A \circ\!\!-\!\!\circ B$, where $A \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $B \notin \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, after the first third steps of Alg. 2. There must be $B \in \mathbf{S}$ for otherwise there is $B \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ due to $A \circ\!\!-\!\!\circ B$. However, for such a case, the circle edge should have been transformed to $A \leftarrow\!\!\circ B$ in the first step of Alg. 2, contradiction.

A.2. Denote $\bar{\mathbb{M}}$ the obtained graph after the first three steps of Alg. 2. We will prove that the circle component in $\bar{\mathbb{M}}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is chordal.

Suppose the circle component in $\bar{\mathbb{M}}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is not chordal, there is $V_0 \circ\!\!-\!\!\circ V_1 \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_n \circ\!\!-\!\!\circ V_0, n \geq 3$, where there is not a circle edge between every two unconsecutive vertices. There must exist non-circle edges between the unconsecutive vertices in this cycle, otherwise it is a cycle of length four or more without a chord in $\mathbb{M}$, contradicting with the chordal property of $\mathbb{M}$. Hence, we can always find a minimal sub-structure $V_k \circ\!\!-\!\!\circ V_{k+1} \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_m \leftarrow V_k, 0 \leq k < m \leq n$ without other directed edges between any two vertices among $V_k, \cdots, V_m$ except for a directed edge between V_m and V_k (we suppose it $V_m \leftarrow V_k$ without loss of generality), and there is not a proper sub-structure satisfying the conditions above. According to Lemma 3, $V_k \rightarrow V_m$ can only be a circle edge in $\mathbb{M}$. Hence in $\mathbb{M}$ there is $V_k \circ\!\!-\!\!\circ V_{k+1} \circ\!\!-\!\!\circ \cdots \circ\!\!-\!\!\circ V_m \circ\!\!-\!\!\circ V_k$. Since the circle component in $\mathbb{M}$ in chordal, the length of the sub-structure can only be three. Hence it holds $m = k + 2$ and there is $V_k \circ\!\!-\!\!\circ V_{k+1} \circ\!\!-\!\!\circ V_{k+2} \leftarrow V_k$ in $\bar{\mathbb{M}}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$. Next, we will prove its impossibility.

We first prove that the edge $V_{k+2} \leftarrow V_k$ cannot be oriented by the first step of Alg. 2. If it is, then there is $V_{k+2} \in$

$\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $V_k \in \mathbf{S}$. Note there is $V_{k+1} \circ\!\!-\!\!\circ V_{k+2}$ in $\mathbb{M}$, there must be $V_{k+1} \in \mathbf{S}$ for otherwise there is $V_{k+1} \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and thus $V_{k+1} \leftarrow V_k$ is oriented in the first step of Alg. 2. However, when $V_{k+1} \in \mathbf{S}$, there is $V_{k+2} \leftarrow V_{k+1}$ oriented in the first step of Alg. 2. Hence that $V_{k+2} \leftarrow V_k$ is oriented in the first step of Alg. 2 is impossible to obtain a sub-structure $V_k \circ\!\!-\!\!\circ V_{k+1} \circ\!\!-\!\!\circ V_{k+2} \leftarrow V_k$ in $\bar{\mathbb{M}}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$.

Thus $V_{k+2} \leftarrow V_k$ is only possible oriented in the second step. Due to $V_k \circ\!\!-\!\!\circ V_{k+1} \circ\!\!-\!\!\circ V_{k+2}$ in $\bar{\mathbb{M}}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$, there is $\mathcal{F}_{V_k} = \mathcal{F}_{V_{k+1}} = \mathcal{F}_{V_{k+2}}$. As $V_k \to V_{k+2}$ is oriented in the second step, according to Lemma 4, there exists a minimal path $F_t \circ\!\!-\!\!\circ \cdots F_1 (= V_k) \circ\!\!-\!\!\circ F_0 (= V_{k+2})$ in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ such that $\mathcal{F}_{F_t} \supset \mathcal{F}_{F_{t-1}} = \cdots = \mathcal{F}_{F_1} = \mathcal{F}_{F_0}$. Evidently F_2 is adjacent to V_{k+1}, otherwise $V_k \to V_{k+1}$ is also oriented. Since (1) $\mathcal{F}_{F_1} (\mathcal{F}_{V_k}) = \mathcal{F}_{F_2} = \mathcal{F}_{V_{k+1}}$, (2) there is not an edge oriented as different directions in the second step of Alg. 2 according to Lemma 5, and (3) F_2 is not adjacent to V_{k+2}, there can only be $F_2 \circ\!\!-\!\!\circ V_{k+1}$ in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$. We find a sub-structure $F_2 \circ\!\!-\!\!\circ V_{k+1} \circ\!\!-\!\!\circ V_k \leftarrow F_2$ such that $\mathcal{F}_{F_2} = \mathcal{F}_{V_{k+1}} = \mathcal{F}_{V_k}$. Similar to the previous proof, there is $F_3 \circ\!\!-\!\!\circ V_{k+1}$ in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ where $\mathcal{F}_{F_3} = \mathcal{F}_{V_{k+1}}$. Repeat this process until F_t and we conclude that there must be $F_t \circ\!\!-\!\!\circ V_{k+1}$. Since $\mathcal{F}_{F_t} \supset \mathcal{F}_{F_{t-1}} = \cdots = \mathcal{F}_{F_1} = \mathcal{F}_{F_0}$, it is oriented as $F_t \to V_{k+1}$ in the second step of Alg. 2, in which case $V_{k+1} \to V_{k+2}$ is also oriented in the second step of Alg. 2 due to the non-adjacency of F_t and V_{k+2}, contradiction.

A.3. In the first four steps of Alg. 2, the circle edges in $\mathbb{M}[-\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is never transformed. Due to the chordal property of $\mathbb{M}$ and the fact that the subgraph of a chordal graph is also chordal, the circle component in $\mathbb{M}[-\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is chordal.

According the three parts above, it is guaranteed that a graph could be output without different orientations on one edge by Alg. 2.

B. $\mathcal{H}$ is a MAG consistent with $\mathbb{M}$.

It evidently follows that $\mathcal{H}$ has the non-circle marks in $\mathbb{M}$. Hence, it suffices to prove that $\mathcal{H}$ is a MAG consistent with $\mathcal{P}$. To prove it, we construct an auxiliary graph $\mathcal{H}_0$ by transforming all the bi-directed edges $K \leftrightarrow T$ in $\mathcal{H}$ which are $K \circ\!\!\to T$ in $\mathbb{M}$ to $K \to T$. According to Alg. 2 to obtain $\mathcal{H}$ from $\mathbb{M}$, there is $K \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{S}$. We first show that **(B.1.)** $\mathcal{H}_0$ is a MAG consistent with $\mathcal{P}$. Then we show that **(B.2.)** $\mathcal{H}$ can be obtained from $\mathcal{H}_0$ by transformation that preserves the property being a MAG consistent with $\mathcal{P}$.

B.1. We will prove $\mathcal{H}_0$ is a MAG consistent with $\mathcal{P}$ by showing that $\mathcal{H}_0$ can be seen as a graph obtained from $\mathbb{M}$ by transforming $\circ\!\!\to$ to $\to$ and transforming the circle component into a DAG without unshielded colliders, through which we can get the desired result by Property $(P6)$ of $\mathbb{M}$ according to Lemma 1.

It is direct that $\mathcal{H}_0$ has the non-circle marks in $\mathbb{M}$ and there are no new bi-directed edges in $\mathcal{H}_0$ relative to $\mathbb{M}$ since all additional bi-directed edges in $\mathcal{H}$ relative to $\mathbb{M}$ are possibly introduced in only the first step of orientation process of $\mathcal{H}$, which have been transformed to directed edges in $\mathcal{H}_0$. Besides, all the circles on $\circ\!\!\to$ edges in $\mathbb{M}$ are oriented as tails in $\mathcal{H}_0$. In the following it suffices to show that $\mathcal{H}_0$ is also a graph oriented from $\mathbb{M}$ by orienting the circle component in $\mathbb{M}$ into a DAG without unshielded colliders.

Hence, we only consider the circle component in $\mathbb{M}$. We divide it into two parts, one is the circle component in $\mathbb{M}[\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$, denoted by CC_1; and the other is the circle component in $\mathbb{M}[-\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$, denoted by CC_2. Evidently, the set of vertices in CC_2 is $\mathbf{V}(\mathbb{M}) \backslash \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. In the following we will prove that in $\mathcal{H}_0$ both CC_1 and CC_2 are oriented to DAGs without new unshielded colliders and there are no new unshielded colliders or directed or almost directed cycles comprised of the vertices in both CC_1 and CC_2.

We first consider the orientation in $\mathcal{H}_0$ of the edges in CC_1 in the process of obtaining $\mathcal{H}$ from $\mathbb{M}$ and transforming $\mathcal{H}$ to $\mathcal{H}_0$. In Step 3, if there is an edge $\circ\!\!\to$, then the two vertices cannot be connected in the circle component, hence the edges in CC_1 cannot be oriented in Step 3. Hence, according to Alg. 2, the edges in CC_1 can only be oriented by either Step 2 or Step 4. There are no new unshielded colliders or directed or almost directed cycles oriented in the edges of CC_1 by the three following facts. **(1).** There are no new unshielded colliders or directed or almost directed cycles in the edges of CC_1 oriented by Step 2 according to Lemma 5. **(2).** There are no unshielded colliders or directed or almost directed cycles in the edges of CC_1 oriented in Step 4. **(3).** There are no new unshielded colliders or directed or almost directed cycles in edges of CC_1 oriented by both Step 2 and Step 4 due to the balanced property of $\mathbb{M}$ and the impossibility of the transformation of circle edges to bi-directed edges.

Then we consider the orientation in $\mathcal{H}_0$ of the edges in CC_2. The edges in CC_2 totally follows Step 5 of Alg. 2, which evidently does not introduce new unshielded colliders or directed or almost directed cycles.

Finally, we consider the circle edges in $\mathbb{M}$ connecting $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{V}(\mathbb{M})\backslash\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. Suppose $K \circ\!\!-\!\!\circ T$ where $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{V}(\mathbb{M})\backslash\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. In this case there can only be $T \in \mathbf{S}$, otherwise there is also $T \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ due to $K \circ\!\!-\!\!\circ T$ where $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, contradicting with $T \in \mathbf{V}(\mathbb{M})\backslash\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. Thus for all the circle edges in $\mathbb{M}$ connecting $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{V}(\mathbb{M})\backslash\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, in Step 1 of Alg. 2 we will orient it as $K \leftarrow T$ due to $T \in \mathbf{S}$, and we remain this directed edge when obtaining $\mathcal{H}_0$ from $\mathcal{H}$. Hence there cannot be a directed or almost directed cycle containing the vertices in both $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $\mathbf{V}(\mathbb{M})\backslash\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ in $\mathcal{H}_0$. Next we prove that there is not a new unshielded collider containing the vertices in both $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $\mathbf{V}(\mathbb{M})\backslash\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. According to the construction process, for any circle edge $K \circ\!\!-\!\!\circ T$ where $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{S}$, there is $K \leftarrow T$ oriented. Hence, it there is a new unshielded collider, it is as $T \rightarrow K_1 \leftarrow K_2$ or $T_1 \rightarrow K \leftarrow T_2$. We will prove the impossibility of both the cases. $T \rightarrow K_1 \leftarrow K_2$ is evidently impossible for otherwise there is $K_1 \rightarrow K_2$ oriented in Step 2, which contradicts with Lemma 5 that one edge cannot be oriented as different direction. The impossibility of $T_1 \rightarrow K \leftarrow T_2$ is due to Lemma 6. Hence there are not new unshielded colliders or directed or almost directed cycles comprised of the verttices in both CC_1 and CC_2.

Hence, we prove that the graph $\mathcal{H}_0$ constructed based on $\mathcal{H}$ can also be seen as a graph obtained from $\mathbb{M}$ by transforming all edges $\circ\!\!\rightarrow$ to $\rightarrow$ and orienting the circle component into a DAG without unshielded colliders. By Property $P6$ of $\mathbb{M}$ according to Lemma 1, $\mathcal{H}_0$ is a MAG consistent with $\mathcal{P}$.

B.2. We will prove that $\mathcal{H}$ can be obtained from $\mathcal{H}_0$ by transformation that preserves the property being a MAG consistent with $\mathcal{P}$.

Note that the only difference between $\mathcal{H}$ and $\mathcal{H}_0$ is that for $\forall K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $\forall T \in \mathbf{S}$ such that $K \circ\!\!\rightarrow T$ in $\mathbb{M}$, there is $K \rightarrow T$ in $\mathcal{H}_0$ but $K \leftrightarrow T$ in $\mathcal{H}$. Denote the set of different edges in $\mathcal{H}_0$ by $Edge(\mathcal{H}_0) = \{K \rightarrow T \text{ in } \mathcal{H} \mid K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]), T \in \mathbf{S}, K \circ\!\!\rightarrow T \text{ in } \mathbb{M}\}$. We could obtain $\mathcal{H}$ from $\mathcal{H}_0$ by transforming these edges to bi-directed edges. We transform one edge one time. At first, we select the edge $K \rightarrow T$ in $Edge(\mathcal{H}_0)$ according to the selection criterion that (1) we select K that is not an ancestor of any other V_1 such that there is an edge $V_1 \rightarrow V_2$ in $Edge(\mathcal{H}_0)$; and (2) given K selected in the first step, we select T that is not a descendant of any other V_2 such that there is an edge $K \rightarrow V_2$ in $Edge(\mathcal{H}_0)$. Then we obtain $Edge(\mathcal{H}_1)$ by deleting $K \rightarrow T$ from $Edge(\mathcal{H}_0)$. By such operation, we obtain a new graph $\mathcal{H}_1$ and $Edge(\mathcal{H}_1)$. Repeat the process above and we could obtain a series of graphs $\mathcal{H}_0, \mathcal{H}_1, \cdots, \mathcal{H}_m, \mathcal{H}_{m+1}(= \mathcal{H})$. We prove the desired result by induction. Given $\mathcal{H}_0$ is a MAG consistent with $\mathcal{P}$, we will show that for any $\mathcal{H}_i$ and $\mathcal{H}_{i+1}$, where $0 \leq i \leq m$, if $\mathcal{H}_i$ is a MAG, then $\mathcal{H}_{i+1}$ is a MAG Markov equivalent to $\mathcal{H}_i$. Suppose the edge that will be transformed in $\mathcal{H}_i$ is $K \rightarrow T$. According to Lemma 1 of Zhang & Spirtes (2005), given $\mathcal{H}_i$ is a MAG, it suffices to show that (1) there is no directed path from K to T in $\mathcal{H}_i$ other than $K \rightarrow T$; (2) for any $A \rightarrow K$ in $\mathcal{H}_i$, $A \rightarrow T$ is also in $\mathcal{H}_i$; and for any $B \leftrightarrow K$ in $\mathcal{H}_i$, either $B \rightarrow T$ or $B \leftrightarrow T$ is in $\mathcal{H}_i$; (3) there is no discriminating path for K on which T is the endpoint adjacent to K in $\mathcal{H}_i$.

(1) For the sake of contradiction, suppose there is a directed path from K to T in $\mathcal{H}_i$ other that $K \rightarrow T$, we suppose the minimal directed path of this path is $K(= F_0) \rightarrow F_1 \rightarrow \cdots \rightarrow F_m \rightarrow T(= F_{m+1})$. Since we only transform directed edges to bi-directed edges in the whole process, the directed path is also in $\mathcal{H}_0$. We first prove that there must be a vertex $F_n, 1 \leq n \leq m$ such that $F_n \in \mathbf{S}$. Otherwise, all of $F_1, \cdots, F_m$ belong to $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ since $F_0 \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and there is a possible directed path comprised of $F_0, F_1, \cdots, F_m$ in $\mathbb{M}$. **(i.)** If there is $F_m \rightarrow T$ in $\mathbb{M}$, it contradicts with the first condition of Thm. 2. **(ii.)** If there is $F_m \circ\!\!-\!\!\circ T$ in $\mathbb{M}$, there is $F_m \leftarrow T$ oriented in the first step of Alg. 2. Since we never reverse an edge in the process from $\mathcal{H}_0$ to $\mathcal{H}$, there cannot be an edge $F_m \rightarrow T$ in $\mathcal{H}_i$. **(iii.)** If there is $F_m \circ\!\!\rightarrow T$ in $\mathbb{M}$, there is $F_m \rightarrow T$ in $\mathcal{H}_0$ and $\mathcal{H}_i$. According to the edge selection criterion, when there is both $F_m \rightarrow T$ and $K \rightarrow T$ in $\mathcal{H}_i$, we transform $F_m \rightarrow T$ ahead of $K \rightarrow T$ due to $K \rightarrow F_1 \rightarrow \cdots \rightarrow F_m$, contradiction. For the other situations for the edge between F_m and T in $\mathbb{M}$, there cannot form an edge $F_m \rightarrow T$ in $\mathcal{H}_i$. Hence we conclude there is a vertex $F_n, 1 \leq n \leq m$ such that $F_n \in \mathbf{S}$.

Without loss of generality, we suppose $F_n \in \mathbf{S}$ and $F_l \notin \mathbf{S}, \forall 1 \leq l \leq n-1$. We first prove there is not a vertex $F_l, 1 \leq l \leq n-1$ adjacent to T. If there is, since $F_l \rightarrow \cdots \rightarrow F_m \rightarrow T$ in $\mathcal{H}_0$, there is $F_l \rightarrow T$ in $\mathcal{H}_0$ due to the ancestral property. In this case there is a directed path $F_1 \rightarrow \cdots F_l \rightarrow T$ without vertices in $\mathbf{S}$ in $\mathcal{H}_0$, which implies that there is a possible directed path where the sub-path from F_1 to F_l is minimal and any variables on the path do not belong to $\mathbf{S}$, contradicting the result we prove above. Hence F_l cannot be adjacent to T for $\forall 1 \leq l \leq n-1$. **(i.)** If $n \geq 2$, **(i.a.)** if there $F_n \circ\!\!\rightarrow\!\!* T$ or $F_n \rightarrow T$ in $\mathbb{M}$, there is an uncovered possible directed path comprised of $K, F_1, \cdots, F_n, T$ in $\mathbb{M}$ where F_1 is not adjacent to T. In this case $K \circ\!\!\rightarrow T$ has been oriented as $K \rightarrow T$ in $\mathbb{M}$ by $\mathcal{R}_9$ since $\mathbb{M}$ is closed under the orientation rules, contradiction. **(i.b.)** If there is $F_n \leftarrow\!\!* T$ in $\mathbb{M}$, note the non-adjacency of T and F_{n-1}. Due to the edge $T \!*\!\!\rightarrow F_n$ and the complete property of $\mathbb{M}$, the mark at F_n on the edge between F_{n-1} and F_n is identifiable in $\mathbb{M}$. And due to the possible directed path, there is $F_{n-1} \rightarrow F_n$ in $\mathcal{H}_0$, there can only be $F_{n-1} \rightarrow F_n$ or $F_{n-1} \circ\!\!\rightarrow F_n$ in $\mathbb{M}$. The former case contradicts with the first condition of Thm. 2 due to $F_{n-1} \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $F_n \in \mathbf{S}$. For the latter case, the edge $F_{n-1} \rightarrow F_n$ should be transformed to bi-directed edge ahead of $K \rightarrow T$, hence there cannot be an edge $F_{n-1} \rightarrow F_n$ in $\mathcal{H}_i$, contradiction. **(ii.)** If $n = 1$, there is $K \rightarrow T' \rightarrow T$ in $\mathcal{H}$, where $T' \in \mathbf{S}$. In this case if there is not $K \circ\!\!\rightarrow T'$ in $\mathbb{M}$, there cannot be an edge $K \rightarrow T'$ in $\mathcal{H}_i$; if there is $K \circ\!\!\rightarrow T'$ in $\mathbb{M}$, there is thus both $K \rightarrow T'$ and $K \rightarrow T$ in $\mathcal{H}_0$, $K \circ\!\!\rightarrow T'$ is transformed to bi-directed ahead of $K \rightarrow T$ due to $T' \rightarrow T$, thereby there is not an edge $K \rightarrow T'$ in $\mathcal{H}_i$. Hence there cannot be a sub-structure $K \rightarrow T' \rightarrow T$ in $\mathcal{H}_i$, contradiction. Hence, there is always a contradiction if there is a directed path from K to T in $\mathcal{H}_i$.

(2) In this part, we prove that if there is an edge $A \rightarrow K$ in $\mathcal{H}_i$, there is $A \rightarrow T$ in $\mathcal{H}_i$; if there is $B \leftrightarrow K$ in $\mathcal{H}_i$, either $B \rightarrow T$ or $B \leftrightarrow T$ is in $\mathcal{H}_i$. Note there is $K \circ\!\!\rightarrow T$ in $\mathbb{M}$, where $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{S}$.

It suffices to show that for A such that $A \rightarrow K$ or $A \leftrightarrow K$ in $\mathcal{H}_i$, A is adjacent to T. According to the ancestral property of $\mathcal{H}_i$, we get the desired result due to $K \rightarrow T$ in $\mathcal{H}_i$.

We discuss the possible cases of the edge between A and K in $\mathbb{M}$. If there is $A \!*\!\!\rightarrow K \circ\!\!\rightarrow T$ in $\mathbb{M}$, due to the closed property of $\mathbb{M}$, A is adjacent to T. Hence the result evidently holds.

If there is $A \circ\!\!-\!\!\circ K$ in $\mathbb{M}$, we discuss whether $A \in \mathbf{S}$. If not, then $A \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ due to $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. Suppose T is not adjacent to A for contradiction. In this case, we orient $K \rightarrow A$ in the second step due to $T \in \mathcal{F}_K \backslash \mathcal{F}_A$, there is thus $K \rightarrow A$ in $\mathcal{H}_0$. Considering we never reverse a directed edge in the whole procedure, there is not $A \rightarrow K$ in $\mathcal{H}_i$. And since only the directed edge connecting a vertex in $\mathbf{S}$ and a vertex in $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ is possibly converted to a bi-directed edge in the process from $\mathcal{H}_0$ to $\mathcal{H}_i$, $A \leftarrow K$ cannot be transformed to $A \leftrightarrow K$ due to $A, K \in \text{PossDe}(X, \mathbb{M}_i[-\mathbf{S}])$, so that $A \leftrightarrow K$ is not in $\mathcal{H}_i$. Hence when $A \circ\!\!-\!\!\circ K$ in $\mathbb{M}_i$ and $A \notin \mathbf{S}$, there is not an edge $A \rightarrow K$ or $A \leftrightarrow K$ in $\mathcal{H}_i$. If $A \in \mathbf{S}$, according to Alg. 2 to obtain $\mathcal{H}$ from $\mathbb{M}$, there is $A \leftarrow K$ in $\mathcal{H}$. And since in the process of transforming $\mathcal{H}_0$ to $\mathcal{H}$ we only transform directed edges to bi-directed edges, there is $A \leftarrow K$ in $\mathcal{H}_i$, in which case there is not $A \rightarrow K$ or $A \leftrightarrow K$ in $\mathcal{H}_i$.

If there is $A \leftarrow\!\!\circ K$ in $\mathbb{M}$, there is $A \leftarrow K$ in $\mathcal{H}_0$. Since we never reverse a directed edge in the whole process, and only the directed edge connecting a vertex in $\mathbf{S}$ and a vertex in $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ is possibly converted to a bi-directed edge in the process from $\mathcal{H}_0$ to $\mathcal{H}$, we only need to consider there is $A \leftrightarrow K$ in $\mathcal{H}_i$, where $A \in \mathbf{S}$. In this case, there is $A, T \in \mathcal{S}_K$. According to Lemma 6, A is adjacent to T. The result holds.

For the other cases for the edge between A and K in $\mathbb{M}$ except for $A \!*\!\!\rightarrow K$, $A \circ\!\!-\!\!\circ K$, and $A \leftarrow\!\!\circ K$, there cannot be an edge $A \rightarrow K$ or $A \leftrightarrow K$ in $\mathcal{H}_i$. We thus have considered all the possible cases and conclude that if there is $A \rightarrow K$ in $\mathcal{H}_i$, there is $A \rightarrow T$ in $\mathcal{H}_i$; if there is $A \leftrightarrow K$ in $\mathcal{H}_i$, either $A \rightarrow T$ or $A \leftrightarrow T$ is in $\mathcal{H}_i$ according to the balanced property.

(3) In this part, we prove that there is no discriminating path for K on which T is the endpoint adjacent to K in $\mathcal{H}_i$. The proof of this part refers to the proof of (T3) of Theorem 3 by Zhang (2008a) with some modifications.

Suppose a path $p = (V_0, V_1, \cdots, V_n = K, T)$ in $\mathcal{H}_i$ which is a discriminating path for K. Without loss of generality, suppose p is the shortest path. According to the construction of $\text{Edge}(\mathcal{H}_i)$, there is $K \circ\!\!\rightarrow T$ in $\mathbb{M}$. We derive a contradiction by showing that p is already a discriminating path in $\mathbb{M}$. Hence there cannot be an edge $K \circ\!\!\rightarrow T$ in $\mathbb{M}$, otherwise if $i \geq 1$ (there is local BK) it will be oriented as $K \rightarrow T$ by $\mathcal{R}_4'$ due to the closed property of $\mathbb{M}$. There is $V_{n-1} \leftrightarrow K$ in $\mathcal{H}_i$, for otherwise there would be a directed path $K \rightarrow V_{n-1} \rightarrow T$ from K to T other than the edge $K \rightarrow T$ in $\mathcal{H}_i$, contradiction. It follows that every edge on the subpath from V_1 to K is bi-directed in $\mathcal{H}_i$.

Next we will prove that there is an edge $V_0 \!*\!\!\rightarrow V_1$ in $\mathbb{M}$. Suppose for contradiction, the edge is either $V_0 \circ\!\!-\!\!\circ V_1$ or $V_0 \leftarrow\!\!\circ V_1$.

(i). Suppose $V_0 \circ\!\!-\!\!\circ V_1$ in $\mathbb{M}$. There cannot be an edge $V_1 \leftrightarrow V_2$ in $\mathbb{M}$, for otherwise there is $V_0 \leftrightarrow V_2$ in $\mathbb{M}$ due to the balanced property of $\mathbb{M}$, which contradicts with the shortest discriminating path p. Since we do not transform a circle edge in $\mathbb{M}$ to a bi-directed edge, the edge between V_1 and V_2 are either $V_1 \circ\!\!\rightarrow V_2$ or $V_1 \leftarrow\!\!\circ V_2$. For the former case, V_0 is adjacent to V_2, for otherwise $V_0 \!*\!\!\rightarrow V_1 \leftarrow\!\!* V_2$ is identifiable in $\mathcal{P}$ and $\mathbb{M}$ since $V_0 \!*\!\!\rightarrow V_1 \leftrightarrow V_2$ in $\mathcal{H}_i$ and $\mathcal{H}_i$ is a MAG Markov equivalent to $\mathcal{H}_0$ which is consistent with $\mathcal{P}$, contradicting with $V_0 \circ\!\!-\!\!\circ V_1$ in $\mathbb{M}$. According to the balanced property of $\mathbb{M}$, there is $V_0 \!*\!\!\rightarrow V_2$ in $\mathbb{M}$ thus there is $V_0 \!*\!\!\rightarrow V_2$ in $\mathcal{H}_i$, in which case there is a shorter discriminating path without V_1, contradiction. For the latter case, there is $V_0 \circ\!\!-\!\!\circ V_1 \leftarrow\!\!\circ V_2$ in $\mathbb{M}$. As shown by the orientation procedure, we only add an arrowhead at the vertex in $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, and we never orient an edge connecting two vertices from

$\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ as bi-directed, hence $V_0 {*\!\!\rightarrow} V_1$ and $V_1 \leftrightarrow V_2$ cannot be oriented at the same time in the process of obtaining $\mathcal{H}$ from $\mathcal{H}_0$.

(ii). Suppose $V_0 \leftarrow\!\circ V_1$. Due to the fact that a bi-directed edge is oriented in $\mathcal{H}_i$ compared to $\mathbb{M}$ only if the edge connects a vertex in $\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and a vertex in $\mathbf{S}$, and the fact that an arrowhead is added only at the vertex in $\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, there is $V_0 \in \mathbf{S}$ and $V_1 \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. According to the first condition of Thm. 2, there is not $V_1 \rightarrow T$ in $\mathbb{M}(=\mathcal{H})$. In light of the fact that we never transform an edge to a directed edge in the process from $\mathcal{H}_0$ to $\mathcal{H}$, there can only be $V_1 \circ\!\!\rightarrow T$ in $\mathbb{M}$. According to Def. 4, $V_0, T \in \mathcal{S}_{V_1}$. According to the definition of discriminating path, V_0 is not adjacent to T, which contradicts with Lemma 6. Hence $V_0 \leftarrow\!\circ V_1$ in $\mathbb{M}$ is impossible.

We conclude that there is $V_0 {*\!\!\rightarrow} V_1$ in $\mathbb{M}$. The remaining part is to prove by induction that for every $1 \leq i \leq n-1$, V_i is a collider and a parent of T in $\mathbb{M}$. $V_1 \rightarrow T$ is evident due to the non-adjacency of V_0 and T. Note $T \in \mathbf{C}$ and $V_1 \rightarrow T$ in $\mathbb{M}$, thus $V_1 \notin \mathrm{PossDe}(X, \mathbb{M}[-\mathbf{C}])$ due to $\mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \cap \mathrm{Pa}(\mathbf{S}, \mathbb{M}) = \emptyset$ according to the first condition of Thm. 2. Hence, there cannot be an edge $V_1 \rightarrow V_2$ in $\mathbb{M}$ since the edge cannot be oriented as $V_1 \leftrightarrow V_2$ in $\mathcal{H}_i$. If there is not a collider at V_1 in $\mathbb{M}$, there is $V_1 \circ\!\!\rightarrow V_2$. It is impossible because we never transform it to bi-directed in the process from $\mathbb{M}$ to $\mathcal{H}_i^0$ as $V_1 \notin \mathrm{PossDe}(X, \mathbb{M}[-\mathbf{C}])$. Hence the collider is identifiable in $\mathbb{M}$. Similarly, we could prove $V_2 \rightarrow T$ in $\mathbb{M}$. Then we prove there is $V_2 \leftarrow\!\!* V_3$ in $\mathbb{M}$. If the edge is a circle edge, then there must be $V_1 \circ\!\!-\!\!\circ V_3$ according to the balance property, in which case there is a shorter discriminating path, contradiction. Then we consider the edge is $V_2 {*\!\!\rightarrow} V_3$. Due to $T \in \mathbf{C}$ and $\mathrm{PossDe}(X, \mathbb{M}[-\mathbf{C}]) \cap \mathrm{Pa}(\mathbf{C}, \mathbb{M}) = \emptyset$, $V_2 \notin \mathrm{PossDe}(X, \mathbb{M}[-\mathbf{C}])$. Hence $V_2 {*\!\!\rightarrow} V_3$ in $\mathbb{M}$ can never be transformed to bi-directed since arrowhead is added at only the vertex in $\mathrm{PossDe}(X, \mathbb{M}[-\mathbf{C}])$. Thus $V_1 \leftrightarrow V_2 \leftarrow\!\!* V_3$ is identifiable in $\mathbb{M}$. By such way, we prove that the path is a discriminating path for K in $\mathbb{M}$. Thus there cannot be an edge $K \circ\!\!\rightarrow T$ in $\mathbb{M}$, otherwise it will be oriented as $K \rightarrow T$ by $\mathcal{R}_4'$ if $i \geq 1$ and oriented as $K \rightarrow T$ or $K \leftrightarrow T$ if $i = 0$ since $\mathbb{M}$ is closed under the orientation rules, contradicting with $K \circ\!\!\rightarrow T$ in $\mathbb{M}_i$.

Hence, we conclude that $\mathcal{H}$ is a MAG Markov equivalent to $\mathcal{H}_0$. Since we have proven in **B.1.** that $\mathcal{H}_0$ is a MAG consistent with $\mathcal{P}$, $\mathcal{H}$ is a MAG consistent with $\mathcal{P}$. And according to Alg. 2 to obtain $\mathcal{H}$, $\mathcal{H}$ has the non-circle marks in $\mathbb{M}$. Hence $\mathcal{H}$ is a MAG consistent with $\mathbb{M}$. ☐

We then show two results of Wang & Zhou (2021), which are used in the main proof.

Definition 9 (Critical vertex for (X, Y); Wang & Zhou (2021))**.** In a maximal local MAG $\mathbb{M}$, F_t is called a *critical vertex* for (X, Y) if there is a path $X \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1} \leftrightarrow F_t$ or $X \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1} \leftarrow\!\circ F_t$, $t \geq 1$, where each non-endpoint variable is an ancestor of X or Y in $\mathbb{M}$, and there is a non-empty variable set $\mathbf{S}$ relative to F_t defined as follows: $S \in \mathbf{S}$ if and only if in M (1) S is a child of $X, F_1, \cdots, F_{t-1}$, (2) there is one minimal possible directed path from F_t to Y in the form of $F_t \circ\!\!-\!\!* S \cdots Y$.

Proposition 3 (Wang & Zhou (2021))**.** *Let $\mathbb{M}$ be a maximal local MAG of X based on a PAG $\mathcal{P}$, suppose $\mathcal{M}$ a MAG valid to $\mathbb{M}$. Denote* D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ *by $\mathbf{D}$. If F_t is not a critical vertex in $\mathbb{M}$ and $F_t \notin \mathrm{Anc}(Y, \mathcal{M})$, then $F_t \perp Y \mid \mathbf{D}, X$ in $\mathcal{M}$.*

Note Proposition 3 is the part 1 of Theorem 3 of Wang & Zhou (2021). There are two modifications. One is in their paper they consider local MAG. Here the maximal local MAG is a special case of local MAG, hence the result holds also. The other is that a new notion PD-SEP(X, Y, M) is used in their paper for their settings. Such a notion is redundant here. Hence we do not mention that.

Lemma 8. *Suppose a maximal local MAG $\mathbb{M}$ and a MAG $\mathcal{M}$ consistent with $\mathbb{M}$. There is $\mathrm{PossDe}(X, \mathbb{M}) = \mathrm{De}(X, \mathcal{M})$.*

Proof. It is evident that $\mathrm{PossDe}(X, \mathbb{M}) \supseteq \mathrm{De}(X, \mathcal{M})$. It suffices to show $\mathrm{PossDe}(X, \mathbb{M}) \subseteq \mathrm{De}(X, \mathcal{M})$. Suppose a vertex $V \in \mathrm{PossDe}(X, \mathbb{M})$. Then there is a possible directed path from X to V in $\mathbb{M}$. According to Lemma 2, there is a minimal possible directed path p_1 from X to V in $\mathbb{M}$. Since the mark at X is definite in $\mathbb{M}$, p_1 starts with a directed edge out of X. According to the completeness property of $\mathbb{M}$, p_1 can only be a directed path in $\mathbb{M}$. Hence $V \in \mathrm{De}(X, \mathcal{M})$. We thus conclude $\mathrm{PossDe}(X, \mathbb{M}) \subseteq \mathrm{De}(X, \mathcal{M})$. The proof completes. ☐

Lemma 9. *Given a maximal local MAG $\mathbb{M}$ where $X \in \mathrm{Anc}(Y, \mathbb{M})$ and a potential adjustment set $\mathbf{W}$. Suppose the three conditions in Theorem 2 are satisfied and we obtain a MAG $\mathcal{M}$ according to Algorithm 2. If there is some vertex $V \in \mathbf{W} \cup \bar{\mathbf{W}}$, there always exists a collider path from X to V beginning with an arrowhead at X where each non-endpoint belongs to* D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$.

Proof. Denote the constructed MAG by $\mathcal{M}$. According to the definition of potential adjustment set $\mathbf{W}$ and $\bar{\mathbf{W}}$, there is a collider path as $X(=F_0) \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1} \leftarrow\!\!* V$, where $F_1, F_2, \cdots, F_{t-1} \in \mathbf{W}$. If each vertex in this path is an ancestor of Y in $\mathcal{M}$, the result evidently holds. Hence we consider if there exists a vertex that is not an ancestor of Y in the collider path $X \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1} \leftarrow\!\!* V$. We prove that in this case, we could *always* find another collider path from X to Y beginning with an arrowhead at X where each non-endpoint is an ancestor of Y.

Without loss of generality, suppose $F_i, 1 \le i \le t-1$ be the first vertex from X to V that is not an ancestor of Y in $\mathcal{M}$. According to Def. 5, (1) there cannot be an edge as $X \to F_s, 1 \le s \le i-1$, (2) there exists one minimal possible directed path p from F_i to Y that do not go through the vertex in $\bar{\mathbf{W}}$. Suppose $p = \langle F_i, V_1, \cdots, Y\rangle$. We first show that there is $F_i \circ\!\!\!-\!\!* V_1$ in $\mathbb{M}$. It is evidently not in the form of $F_i \leftarrow\!\!* V_1$ since p is a minimal possible directed path. If it is $F_i \to V_1$ in $\mathbb{M}$, since p is minimal possible directed and $\mathbb{M}$ is closed under orientation rules, there must be $F_i \to V_1 \to \cdots \to Y$ in $\mathbb{M}$, thus F_i is an ancestor of Y in $\mathcal{M}$, contradiction. Hence the edge can only be $F_i \circ\!\!\!-\!\!* V_1$ in $\mathbb{M}$.

Considering $F_{i-1} \leftrightarrow F_i \circ\!\!\!-\!\!* V_1$ and $F_{i+1} *\!\!\to F_i \circ\!\!\!-\!\!* V_1$ in $\mathbb{M}$, there is $F_{i-1} *\!\!\to V_1$ and $F_{i+1} *\!\!\to V_1$ in $\mathbb{M}$ according to the balance property of $\mathbb{M}$. At first, we prove that F_i is not a critical vertex here. Otherwise, suppose a minimal possible directed path from F_i to Y comprised of $F_i, V_1, \cdots, V_{k-1}, Y$ where $V_1 \in \text{Chd}(X, \mathcal{M})$. By Alg. 2, we only transform $K \circ\!\!\to T$ as $K \leftrightarrow T$ when $K \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $T \in \mathbf{S}$. Since there is $F_i \circ\!\!\to V_1$ in $\mathbb{M}$ and $F_i \in \mathbf{W}$, $F_i \circ\!\!\to V_1$ is transformed to $F_i \to V_1$ in $\mathcal{M}$. In this case, there must be $F_i \to V_1 \to \cdots \to Y$ in $\mathcal{M}$ since the path is a minimal path, otherwise there will be new unshielded colliders in $\mathcal{M}$ relative to $\mathbb{M}$, impossibility. There is thus $F_i \in \text{Anc}(Y, \mathcal{M})$, contradiction. F_i cannot be a critical vertex here. It is easy to prove that there exists $F_j, 0 \le j \le i-1$ such that there is an edge $F_j \leftrightarrow V_1$, and a vertex $F_k, i < k \le t-1$ such that there is an edge $V_1 \leftrightarrow F_k$ or an edge $F_t *\!\!\to V_1$ (otherwise there is a discriminating path for F_i, hence the circle at $F_i \circ\!\!\!-\!\!* V_1$ should be identified in $\mathcal{P}$, we do not show the details). In this case, if $V_1 \notin \mathbf{W}$, then $V_1 \in \bar{\mathbf{W}}$, contradiction. Hence V_1 could only belong to $\mathbf{W}$. Then we consider whether $V_1 \in \text{Anc}(Y, \mathbb{M})$, in which case $V_1 \in \text{D-SEP}(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. If it is, we find a new collider path $X \leftrightarrow \cdots \leftrightarrow F_j \leftrightarrow V_1 \leftrightarrow F_k \leftrightarrow \cdots F_{t-1} \leftarrow\!\!* F_t$ or $X \leftrightarrow \cdots \leftrightarrow F_j \leftrightarrow V_1 \leftarrow\!\!* F_t$. If $V_1 \notin \text{Anc}(Y, \mathbb{M})$, we conclude that V_1 is not a critical vertex as the process above. And similarly, we conclude $V_2 \in \mathbf{W}$ and further consider V_2. By such way, we could always find a vertex $V_t, 1 \le t \le k-1$ such that there is a collider path $X \leftrightarrow \cdots \leftrightarrow V_t \leftrightarrow \cdots \leftarrow\!\!* F_t$ and the first possible vertex that is not ancestor of Y in $\mathcal{M}$ is only possible in the sub-path from F_{i+1} to F_t. If such a vertex exists, $F_u, i+1 \le u \le t$ for example, repeat the process above and we could find a new collider path. Repeat the process above and we could always find a collider path beginning with arrowhead at X satisfying that each non-endpoint is an ancestor of Y in $\mathcal{M}$. It is evident that each non-endpoint in this path belongs to $\mathbf{D}$. Hence there is a collider path from X to V beginning with an arrowhead at X where each non-endpoint belongs to $\mathbf{D}$. □

Lemma 10. *If there exists a minimal collider path in a MAG $\mathcal{M}$ consistent with a PAG $\mathcal{P}$, then it is also a collider path in $\mathcal{P}$.*

Proof. Suppose a minimal collider path p in $\mathcal{M}$, we consider its corresponding path in $\mathcal{P}$. If there exists a circle or tail at the non-endpoint vertex on this path, according to the completeness of FCI (Zhang, 2008a), there exists a MAG Markov equivalent to $\mathcal{M}$ that has a tail there, which contradicts Theorem 2.1 of Zhao et al. (2005) that Markov equivalent MAGs have the same minimal collider paths. Hence the corresponding path of p in $\mathcal{P}$ is also a collider path. □

Theorem 2. Given a maximal local MAG $\mathbb{M}$, for any potential adjustment set $\mathbf{W}$, there exists a MAG $\mathcal{M}$ valid to $\mathbb{M}$ such that $\mathbf{W}$ is an adjustment set in $\mathcal{M}$ if there exists a block set $\mathbf{S}$ such that

(1) $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \cap \text{Pa}(\mathbf{S}, \mathbb{M}) = \emptyset$;

(2) $\mathbb{M}[\mathbf{S}_V]$ is a complete graph for any $V \in \bar{\mathbf{W}}$, where $\mathbf{S}_V = \{V' \in \mathbf{S} | V \circ\!\!\!-\!\!* V' \text{ in } \mathbb{M}\}$;

(3) $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is bridged relative to $\mathbf{S}$ in $\mathbb{M}$.

Proof. We will prove that $\mathbf{W}$ is an adjustment set relative to (X, Y) in the MAG $\mathcal{M}$ obtained by Alg. 2.

The whole proof process is comprised of three parts: (1) for $\forall V \in \mathbf{W}$, there is $V \in \mathbf{D}$ or $V \perp Y \mid \mathbf{D}, X$; (2) for $\forall V \in \bar{\mathbf{W}}$, $V \notin \text{Anc}(Y, \mathcal{M})$; (3) If $V \in \mathbf{D}$, then $V \in \mathbf{W}$. Then by the first and third parts, we could get the desired results.

(1) For $\forall V \in \mathbf{W}$, there is $V \in \mathbf{D}$ or $V \perp Y \mid \mathbf{D}, X$.

In this part, we first prove that if $V \in \mathbf{W}$, there is either $V \in \mathrm{Anc}(Y, \mathcal{M})$ or $V \perp Y \mid \mathbf{D}, X$. According to Def. 5 and Lemma 8, $V \notin \mathrm{De}(X, \mathcal{M})$.

If $V \in \mathrm{Anc}(Y, \mathbb{M})$, it is evident that $V \in \mathrm{Anc}(Y, \mathcal{M})$. Hence we only consider $V \notin \mathrm{Anc}(Y, \mathbb{M})$ in the following. According to the condition of $\mathbf{W}$, by Lemma 9, there is a collider path as $X(= F_0) \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1} \leftarrow\!\!* V$ where each non-endpoint belongs to $\mathbf{D}$.

If the edge between F_{t-1} and V is as $F_{t-1} \leftarrow\!\circ V$ in $\mathbb{M}$, by Alg. 2, there is $F_{t-1} \leftarrow V$ in $\mathcal{M}$, thereby $V \in \mathrm{Anc}(Y, \mathcal{M})$. If the edge is $F_{t-1} \leftrightarrow V$ in $\mathbb{M}$, according to Def. 5, there exist some minimal possible directed paths from V to Y in $\mathbb{M}$. **(i).** If V is a critical vertex, suppose a minimal possible directed path from V to Y comprised of $V, V_1, \cdots, V_{k-1}, Y$ where $V_1 \in \mathrm{Chd}(X, \mathcal{M})$. In this case there must be $V \circ\!\!\rightarrow V_1$ in $\mathbb{M}$ due to the balanced as well as closed property of $\mathbb{M}$ and the definition of critical vertex. By Alg. 2, we orient $V \circ\!\!\rightarrow V_1$ as $V \rightarrow V_1$, in this case there is $V \in \mathrm{Anc}(Y, \mathbb{M})$ (a new bi-directed edge is additionally introduced only if $V \in \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, $V_1 \in \mathbf{S}$, and there is $V \circ\!\!\rightarrow V_1$ in $\mathbb{M}$. However, when $V \in \mathbf{W}$, $V \notin \mathrm{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$, hence the edge can only be oriented as $V \rightarrow V_1$.). **(ii).** If V is not a critical vertex, if $V \in \mathrm{Anc}(Y, \mathcal{M})$, there is $V \in \mathbf{D}$. If $V \notin \mathrm{Anc}(Y, \mathcal{M})$, by Proposition 3, it holds $V \perp Y \mid \mathbf{D}, X$.

(2) for $\forall V \in \bar{\mathbf{W}}$, $V \notin \mathrm{Anc}(Y, \mathcal{M})$.

Suppose there is a vertex $V \in \bar{\mathbf{W}} \cap \mathrm{Anc}(Y, \mathcal{M})$ such that $V \in \mathrm{Anc}(Y, \mathcal{M})$ for contradiction. According to the definition of $\bar{\mathbf{W}}$, for $\forall V \in \bar{\mathbf{W}}$, there exists a collider path $X(= F_0) \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1} \leftrightarrow\!\!* V$ where each non-endpoint belongs to $\mathbf{W}$. If there exists one directed path from V to Y in $\mathcal{M}$, thus there is a minimal directed path from V to Y in $\mathcal{M}$. According to the condition $\bar{\mathbf{W}} \cap \mathrm{Anc}(Y, \mathbb{M}) = \emptyset$, there does not exist one directed path from V to Y in $\mathbb{M}$, hence the minimal directed path in $\mathcal{M}$ is not a directed path in $\mathbb{M}$.

If there exists some vertex $V' \in \bar{\mathbf{W}}$ that does not go through the vertex in $\bar{\mathbf{W}}$ except for V' in the path, there is $V' \in \mathrm{Anc}(Y, \mathcal{M})$. We consider V' instead of V. By such operations we could always find a vertex in $\bar{\mathbf{W}}$ that has a minimal possible directed path to Y in $\mathcal{M}$ where each non-endpoint does not belong to $\bar{\mathbf{W}}$. Hence, we suppose there does not exist a vertex in $\bar{\mathbf{W}}$ in the minimal possible directed path from V to Y without loss of generality. That is, each non-endpoint in the path does not belong to $\bar{\mathbf{W}}$.

Then there is a minimal directed path from V to Y in $\mathcal{M}$. We consider its corresponding path p in $\mathbb{M}$. Suppose $p = \langle V, V_1, \cdots, V_k(= Y)\rangle, k \geq 1$. p must be a minimal possible directed path from V to Y. Note p is not a directed path in $\mathbb{M}$, for otherwise it contradicts with the third condition of Def. 5. In light of the closed property of $\mathbb{M}$, there is a vertex $V_s, 1 \leq s \leq k$ in p which is an ancestor of Y in $\mathbb{M}$ and any vertex in $V, V_1, \cdots, V_{s-1}$ is not an ancestor of Y in $\mathbb{M}$. If for any V_i in $V_1, \cdots, V_{s-1}$ there is $V_i \notin \mathbf{S}$, then there is $V_s \in \mathbf{S}$ according to Def. 4. According to the first step of Alg. 2, there is $V_{s-1} \leftrightarrow\!\!* V_s$ oriented in $\mathcal{M}$, which contradicts with the fact that the corresponding path of p in $\mathcal{M}$ is a directed path. If there is some V_i in $V_1, \cdots, V_{s-1}$ such that $V_i \in \mathbf{S}$. Without loss of generality, suppose $V_1, V_2, \cdots, V_{j-1} \notin \mathbf{S}$. In this case there is $V_{j-1} \leftarrow V_j$ oriented in the first step of Alg. 2 thus there is $V_{j-1} \leftarrow V_j$ in $\mathcal{M}$, which contradicts with the fact that the corresponding path of p in $\mathcal{M}$ is a directed path. Hence both of the cases are impossible. Hence for any $V \in \bar{\mathbf{W}}$, there is $V \notin \mathrm{Anc}(Y, \mathcal{M})$.

(3) $\mathbf{D} \subseteq \mathbf{W}$. Suppose $V \in \mathbf{D} \backslash \mathbf{W}$, then there exists a minimal collider path $p = X \leftrightarrow F_1 \cdots \leftrightarrow F_{t-1} \leftrightarrow\!\!* V(= F_t)$ where each vertex is an ancestor of Y in $\mathcal{M}$. Without loss of generality, we suppose $F_1, \cdots, F_{t-1} \in \mathbf{W}$ and $V \in \mathbf{D} \backslash \mathbf{W}$ (if $F_i \notin \mathbf{W}$, we consider F_i instead of V), we will prove the result by showing $V \in \mathbf{W}$, which contradicts with $V \in \mathbf{D} \backslash \mathbf{W}$. If there is not an edge as $X \rightarrow V$, due to the minimal collider path, the collider path p is identifiable in $\mathcal{P}$ by Lemma 10. According to $V \in \mathbf{D}$, it follows $V \in \mathrm{PossAn}(Y, \mathbb{M})$. Since $F_1, \cdots, F_{t-1} \in \mathbf{W}$, there is $V \in \mathbf{W} \cup \bar{\mathbf{W}}$. Hence if $V \notin \mathbf{W}$, there is $V \in \bar{\mathbf{W}}$. However, as we have shown in the previous part, for $V \in \bar{\mathbf{W}}$, $V \notin \mathrm{Anc}(Y, \mathcal{M})$, contradiction.

If there is $X \rightarrow V$, due to $F_1, \cdots, F_{t-1} \in \mathbf{W}$ and $\mathbf{W} \cap \mathrm{PossDe}(X, \mathbb{M}) = \emptyset$, there is not an edge as $X \rightarrow F_i, 1 \leq i \leq t-1$. Note in Alg. 2 we never orient an edge $A \circ\!\!\rightarrow B$ where $A, B \in \mathbf{W} \cup \{X\}$ as bi-directed, thus if $X \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1}$ in $\mathcal{M}$ is obtained by Alg. 2, there is $X \leftrightarrow F_1 \leftrightarrow \cdots \leftrightarrow F_{t-1}$ in $\mathbb{M}$. Also note we never add an arrowhead at a vertex in $\mathbf{W}$ in a non-circle edge connecting a vertex in $\mathbf{W}$ and a vertex not in $\mathbf{W}$. Hence if $F_{t-1} \leftrightarrow\!\!* V$ in $\mathcal{M}$, there is either $F_{t-1} \leftrightarrow\!\!* V$ in $\mathbb{M}$, or $F_{t-1} \circ\!\!-\!\!\circ V$ in $\mathbb{M}$. For the former case, since $X \leftrightarrow F_1 \cdots \leftrightarrow F_{t-2} \leftrightarrow F_{t-1} \leftrightarrow\!\!* V$ is in $\mathbb{M}$, there is $V \in \mathbf{W} \cup \bar{\mathbf{W}}$. And since $V \in \mathbf{D} \backslash \mathbf{W}$, there is $V \in \bar{\mathbf{W}} \cap \mathbf{D}$. However, as we have shown in the previous part, for $V \in \bar{\mathbf{W}}$, $V \notin \mathrm{Anc}(Y, \mathcal{M})$, contradiction. For the latter case, since circle edge will only be transformed to directed edge by Alg. 2, and there is $V_{t-1} \leftrightarrow\!\!* V_t$ in $\mathcal{M}$, there is $F_{t-1} \leftarrow V$ in $\mathcal{M}$. Due to $X \rightarrow V$ in $\mathbb{M}$, there is $F_{t-1} \in \mathrm{PossDe}(X, \mathbb{M})$. Hence $F_{t-1} \in \mathbf{W} \cap \mathrm{PossDe}(X, \mathbb{M})$, contradiction with the second condition of Def. 5. Hence both of the cases are impossible. We thus conclude $\mathbf{D} \subseteq \mathbf{W}$.

Combining the first the third results, $\mathbf{D} \subseteq \mathbf{W}$; and for $V \in \mathbf{W}\backslash\mathbf{D}$, there is $V \perp Y \mid \mathbf{D}, X$. Since $\mathbf{W} \cap \text{De}(X, \mathbb{M}) = \emptyset$, there is $\mathbf{D} \cap \text{De}(X, \mathcal{M}) = \emptyset$, hence in this case $\mathbf{D}$ is an adjustment set relative to (X, Y) in $\mathcal{M}$. And since $V \perp Y \mid \mathbf{D}, X$ for $V \in \mathbf{W}\backslash\mathbf{D}$, $\mathbf{W}$ is also an adjustment set relative to (X, Y) in $\mathcal{M}$ by the following equations

$$
\begin{aligned}
\int_{\mathbf{W}} f(\mathbf{W}) f(Y \mid \mathbf{W}, X) \,\mathrm{d}\mathbf{W} &= \int_{\mathbf{W}} f(\mathbf{W}) f(Y \mid \mathbf{D}, X) \,\mathrm{d}\mathbf{W} \\
&= \int_{\mathbf{D}} f(\mathbf{D}) f(Y \mid \mathbf{D}, X) \,\mathrm{d}\mathbf{D} \;\; (\because \mathbf{W}\backslash\mathbf{D} \perp Y \mid \mathbf{D}, X) \\
&= f(Y \mid do(X)). \;\; (\because \text{Thm. 1 and (1)})
\end{aligned}
$$

□

D.3. Proof of Theorem 3

Theorem 3. Given a maximal local MAG $\mathbb{M}$, suppose a MAG $\mathcal{M}$ valid to $\mathbb{M}$ such that there exists an adjustment set relative to (X, Y). Let $\mathbf{W}$ be D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. Then $\mathbf{W}$ is a potential adjustment set in $\mathbb{M}$ and there exists a block set $\mathbf{S}$ such that

(1) $\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \cap \text{Pa}(\mathbf{S}, \mathbb{M}) = \emptyset$;

(2) $\mathbb{M}[\mathbf{S}_V]$ is a complete graph for any $V \in \bar{\mathbf{W}}$, where $\mathbf{S}_V = \{V' \in \mathbf{S} | V \circ\!\!\rightarrow\!\!* \, V' \text{ in } \mathbb{M}\}$;

(3) $\mathbb{M}[\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])]$ is bridged relative to $\mathbf{S}$ in $\mathbb{M}$.

Proof. Since there exists an adjustment set relative to (X, Y) in $\mathcal{M}$, there is D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}}) \cap \text{De}(X, \mathcal{M}) = \emptyset$ according to Thm. 1.

At first, we show that $\mathbf{W} =$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ is a potential adjustment set as Definition 5 in $\mathbb{M}$, by proving $\mathbf{W} =$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$ satisfied the three conditions in Def. 5. Note $\mathbf{W} \cap \bar{\mathbf{W}}$ cannot be a non-empty set according to the definition of $\mathbf{W}$ and $\bar{\mathbf{W}}$.

For the first condition, consider $V \in \mathbf{W} =$ D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. According to the definition of D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$, there is a collider path from X to V where each non-endpoint belongs to D-SEP$(X, Y, \mathcal{M}_{\underset{\sim}{X}})$. Since $V \in \mathbf{W}$, there is a directed path from V to Y. Thus there is a possible directed path from V to Y in $\mathbb{M}$. If there exists a minimal possible directed path from V to Y in $\mathbb{M}$ which does not go through the vertex in $\bar{\mathbf{W}}$, then there is $V \in \mathbf{W}$. If not, it follows that for each possible directed path from V to Y in $\mathbb{M}$, the path goes through the vertex in $\bar{\mathbf{W}}$. Since there is at least one minimal possible directed path in $\mathbb{M}$ is directed in $\mathcal{M}$, there exists some vertex $V' \in \bar{\mathbf{W}}$ which is an ancestor of Y. According to Def. 4, there is $V' \in \mathbf{W}$, thus it holds that $V' \in \mathbf{W} \cap \bar{\mathbf{W}}$, contradiction.

For the second condition, suppose $V \in \mathbf{W} \cap \text{PossDe}(X, \mathbb{M})$. According to Lemma 8, there is $V = \mathbf{W} \cap \text{De}(X, \mathcal{M}) \neq \emptyset$, contradicting with Thm. 1.

For the third condition, suppose $V \in \bar{\mathbf{W}} \cap \text{Anc}(Y, \mathbb{M})$. There is $V \in \text{Anc}(Y, \mathcal{M})$. According to the definition of $\bar{\mathbf{W}}$, there is a collider path in the form of $X \leftrightarrow \cdots \leftarrow\!\!* V$ in $\mathbb{M}$, hence there is $V \in \mathbf{D}(= \mathbf{W})$ due to $V \in \text{Anc}(Y, \mathbb{M})$, thus it holds $\mathbf{W} \cap \bar{\mathbf{W}} \neq \emptyset$, contradiction.

Hence $\mathbf{W}$ is a potential adjustment set. Then we set $\mathbf{S} = \text{Anc}(Y \cup \mathbf{W}, \mathcal{M}) \cap [\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M})\backslash\bar{\mathbf{W}}]$ which evidently fulfills that $\text{Anc}(Y \cup \mathbf{W}, \mathbb{M}) \cap [\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M})\backslash\bar{\mathbf{W}}] \subseteq \mathbf{S} \subseteq \text{PossAn}(Y \cup \mathbf{W}, \mathbb{M}) \cap [\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M})\backslash\bar{\mathbf{W}}]$. It suffices to show that the three conditions in Thm. 3 are satisfied for such $\mathbf{S}$.

For the first condition, suppose $V \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}]) \cap \text{Pa}(\mathbf{S}, \mathbb{M})$. Suppose $V \to S$ where $S \in \mathbf{S}$. According to the selected $\mathbf{S}$ and $\mathbf{D} = \mathbf{W}$, there is $V \in \text{Anc}(Y, \mathcal{M})$. In this case, there can only be $V \in \bar{\mathbf{W}}$, for otherwise there is $V \in \mathbf{S}$, which contradicts with $V \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. Note V is an ancestor of Y in $\mathcal{M}$, hence $V \in \mathbf{W}$. Thus $V \in \mathbf{W} \cap \bar{\mathbf{W}}$, contradiction.

For the second condition, if for $V \in \bar{\mathbf{W}}$, $\mathbb{M}[\mathbf{S}_V]$ is not a complete graph, to generate no new unshielded colliders in $\mathcal{M}$ relative to $\mathbb{M}$, there is a directed edge $V \to V'$ where $V' \in \mathbf{S}_V$, in this case there must be a directed path from V to Y

in $\mathcal{M}$ and thus $V \in \text{Anc}(Y, \mathcal{M})$. According to the definition of $\bar{\mathbf{W}}$, there is $V \in \mathbf{W}$, thus it holds that $V \in \mathbf{W} \cap \bar{\mathbf{W}}$, contradiction. Hence the second condition is satisfied.

For the third condition, if it is not bridged, then there must be vertice $V \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$ and $S \in \mathbf{S}$ such that there is $V \to S$ in $\mathcal{M}$. Note $\mathbf{S} = \text{Anc}(Y \cup \mathbf{W}, \mathcal{M}) \cap [\text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}) \backslash \bar{\mathbf{W}}]$. Hence there can only be $V \in \bar{\mathbf{W}}$, for otherwise there is $V \in \mathbf{S}$, which contradicts with $V \in \text{PossDe}(\bar{\mathbf{W}}, \mathbb{M}[-\mathbf{S}])$. Similar to the proof of the first condition, in this case there is $V \in \mathbf{W} \cap \bar{\mathbf{W}}$, contradiction.

The proof completes. $\square$