THE ROLE OF LOCAL BACKGROUND KNOWLEDGE IN MAXIMALLY ORIENTED PARTIALLY DIRECTED ACYCLIC GRAPHS

Anonymous authors
Paper under double-blind review

ABSTRACT

In observational causal inference, background knowledge about specific edge orientations can refine a Markov equivalence class (MEC), with the resulting restricted class represented by a maximally oriented partially directed acyclic graph (MPDAG). Focusing on the causal effect of a treatment X on an outcome Y, this paper addresses the question of how and when local background knowledge about edges incident to X improves the identifiability of this effect. To answer this, we establish a suite of local graphical criteria. First, we derive a necessary and sufficient condition for the validity of proposed orientations of edges incident to X. Second, we provide local orientation rules that determine whether an adjacent undirected edge of X can be directed by the given background knowledge, without enumerating all DAGs in the MEC. Finally, via a notion of possible mediators, we give a necessary and sufficient, polynomial-time criterion for determining when background knowledge improves identifiability, validated on benchmark networks.

1 INTRODUCTION

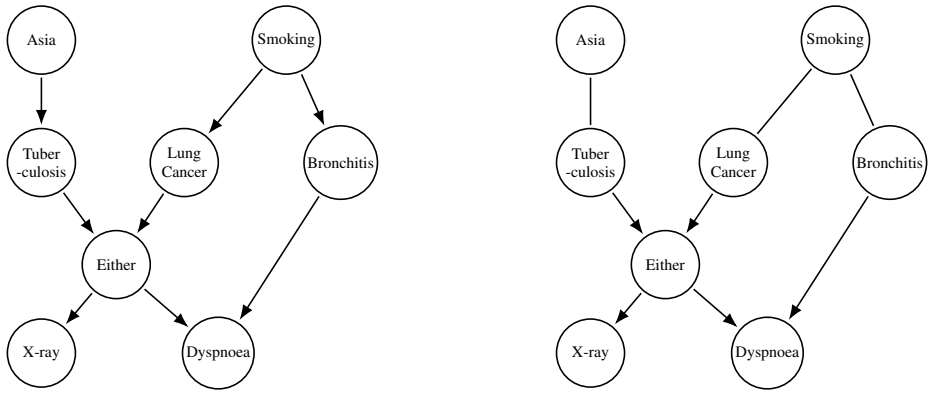
The gold standard for establishing causal relationships is randomized controlled experimentation. In many settings, however, experiments are infeasible and one must rely on observational data. Classic observational methods such as the backdoor adjustment typically require precise knowledge of the causal graph; when the graph is unknown, causal effects become unidentifiable from data alone.

A popular two-stage approach first learns the causal structure from data and then estimates effects from the recovered graph (Gradu et al., 2025). Because causal discovery can only recover the graph up to a Markov equivalence class (MEC), the precision of the subsequent effect estimation depends heavily on the size of this class. Background knowledge, in the form of known causal directions, can narrow the MEC, and the reduced equivalence class is compactly represented by a maximally oriented partially directed acyclic graph (MPDAG).

When the goal is to estimate the causal effect of an exposure X on an outcome Y, Liu et al. (2020) showed that inference can be performed using only the local structure around X. Fang & He (2020) extended this line of work to MPDAGs. Our Theorem 3.2 generalizes their validity condition for background knowledge to the case where only a *subset* of the siblings around X is specified. Despite these advances, it remains unclear whether and how additional background knowledge can *improve* the identifiability of causal effects without enumerating all DAGs in the MEC.

The well-known ASIA network (also known as the Chest clinic example) has become a canonical benchmark in causal discovery and probabilistic inference since it was introduced by Lauritzen & Spiegelhalter (1988). In its original formulation, the entire causal structure—every directed edge—was dictated by expert medical knowledge, resulting in a fully specified directed acyclic graph (DAG). If we do not have prior expert knowledge or we do not trust it, we instead begin by learning a completed partially directed acyclic graph (CPDAG) from purely observational data, representing an MEC of indistinguishable causal structures. Within such a CPDAG, many causal effects of practical interest remain unidentified because of undirected edges that admit multiple confounding or causal pathway configurations.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) Original DAG (b) CPDAG (Markov equivalence class)

Figure 1: The ASIA network: (a) the original causal DAG; (b) the corresponding CPDAG.

If the epidemiologist now receives a piece of domain knowledge—say, *Smoking causes Bronchitis*—several natural questions arise. First, is this piece of knowledge compatible with the CPDAG at all? Second, does it force the orientation of the remaining undirected edge between Smoking and LungCancer? Third, and most importantly, does it improve our ability to estimate the causal effect of Smoking on Dyspnoea? Motivated by such practical questions, this paper studies the role of *local* background knowledge about a target node X in an MPDAG. We assume that the knowledge takes the form of directed edges incident to X: a set $\mathbf{S}_1$ pointing into X and a set $\mathbf{S}_2$ pointing out of X. Our contribution is a set of graphical criteria that answer the three questions above *locally*, without enumerating all DAGs in the equivalence class. Concretely, we proceed as follows.

1. **Compatibility.** We give a necessary and sufficient condition for such background knowledge to be valid in the current MPDAG (Theorem 3.2). This condition generalizes earlier results for complete specifications of siblings around X.
2. **Local orientation.** We show that the orientation of any undirected edge $X - Z$ after incorporating the background knowledge depends solely on a small induced subgraph around X (Theorem 3.4), and we provide explicit orientation rules.
3. **Identifiability improvement.** By introducing the concept of possible mediators, we prove that the set of possible causal effects shrinks if and only if the possible mediator set shrinks, not merely because the number of b-possibly causal paths decreases (Theorems 4.3 and 4.4). This is illustrated by a concrete example in Section 4 and enables a purely local, polynomial-time decision procedure (Theorem 4.9 and Algorithm 1).

We answer the first two questions in Section 3 by establishing a necessary and sufficient compatibility condition and a purely local orientation rule. Section 4 addresses the third question by introducing the notion of possible mediators and characterizing how background knowledge can eliminate them, thereby enhancing identifiability. These local rules can be directly integrated into algorithms like IDA to avoid exhaustive enumeration, boosting both efficiency and, potentially, identifiability.

2 Setup

2.1 Notations

Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ denote a graph with node set $\mathbf{V}$ and edge set $\mathbf{E}$. A path $p = \langle V_1, \dots, V_k \rangle$ is a sequence of adjacent nodes in $\mathcal{G}$. The path p is said to be causal if $V_i \to V_{i+1}$ for every i, a cycle if $V_1 = V_k$, and partially directed if no $V_i \leftarrow V_{i+1}$. A directed acyclic graph (DAG) contains only directed edges and no partially directed cycles. For two nodes X and Y, X is an ancestor of Y (denoted by $X \in an(Y, \mathcal{G})$) and Y is a descendant of X ($Y \in de(X, \mathcal{G})$) if there exists a causal path from X to Y or $X = Y$. A non-endpoint V_i on a path p is a collider if $V_{i-1} \to V_i \leftarrow V_{i+1}$; otherwise, it is a non-collider on p. The path p is d-connected (or open) given $\mathbf{Z}$ if $\mathbf{Z}$ contains no non-colliders and every collider V_i satisfies $de(V_i, \mathcal{G}) \cap \mathbf{Z} \neq \emptyset$; otherwise, p is blocked by $\mathbf{Z}$. For

distinct node sets $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, the sets $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ if every path between any node $X \in \mathbf{X}$ and $Y \in \mathbf{Y}$ is blocked by $\mathbf{Z}$. We say that a distribution P is *consistent* with a graph $\mathcal{G}$, or *Markovian* with respect to $\mathcal{G}$, if every d-separation in $\mathcal{G}$ corresponds to a conditional independence in P. The notation $\mathbf{X} \perp \mathbf{Y} \mid \mathbf{Z}$ is used for d-separation in $\mathcal{G}$ and for conditional independence in P; the meaning is determined by context.

The skeleton of a graph $\mathcal{G}$ is the undirected graph obtained from $\mathcal{G}$ by removing all edge directions. A v-structure is a triple (X_{i-1}, X_i, X_{i+1}) such that $X_{i-1} \to X_i \leftarrow X_{i+1}$ and X_{i-1} and X_{i+1} are not adjacent. Two DAGs are Markov equivalent if they share the same d-separations, which holds if and only if they have the same skeleton and v-structures (Verma & Pearl, 1990). The Markov equivalence class (MEC) $[\mathcal{G}]$ is the set of all DAGs Markov equivalent to $\mathcal{G}$. Each MEC $[\mathcal{G}]$ can be uniquely represented by a complete partially directed acyclic graph (CPDAG) $\mathcal{C}$ (Verma & Pearl, 1990) which has the same skeleton as all DAGs in $[\mathcal{G}]$, and each edge is directed if and only if it has the same orientation in all DAGs in $[\mathcal{G}]$. We also denote this MEC by $[\mathcal{C}]$.

Let $\mathcal{B}$ be a set of background knowledge comprising only direct causal information (i.e., entries of the form $X \to Y$). Applying $\mathcal{B}$ to a CPDAG $\mathcal{C}$ via Meek's rules (Meek, 1995) yields a MPDAG $\mathcal{G}^*$, which we refer to as the MPDAG under background knowledge $\mathcal{B}$. Let $[\mathcal{G}^*]$ be the set of DAGs in $[\mathcal{C}]$ consistent with $\mathcal{B}$; then an edge in $\mathcal{G}^*$ is directed if and only if it is oriented identically in every DAG in $[\mathcal{G}^*]$ (Meek, 1995). A path $p = \langle V_1, \ldots, V_k \rangle$ is said to be a b-possibly causal path in $\mathcal{G}^*$ if no edge $V_i \leftarrow V_j$ exists in $\mathcal{G}^*$ for $1 \leq i < j \leq k$. A path p is chordless if no two non-consecutive nodes on p are adjacent.

In a partially directed graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ (which may be a DAG, CPDAG, or MPDAG), the children, parents, and siblings of a node X are defined as follows: $ch(X, \mathcal{G}) = \{Y \in \mathbf{V} : X \to Y \in \mathcal{G}\}$, $pa(X, \mathcal{G}) = \{Y \in \mathbf{V} : Y \to X \in \mathcal{G}\}$, $sib(X, \mathcal{G}) = \{Y \in \mathbf{V} : X - Y \in \mathcal{G}\}$. Let $adj(X, \mathcal{G}) = pa(X, \mathcal{G}) \cup ch(X, \mathcal{G}) \cup sib(X, \mathcal{G})$ be the adjacent nodes of X in $\mathcal{G}$. A clique $\mathbf{Q} \subseteq \mathbf{V}$ is a set where every pair of nodes is adjacent, and we also say $\mathbf{Q}$ forms a complete graph.

2.2 Interventional distribution and total causal effect

Let $\mathbf{X}$ and $\mathbf{Y}$ be disjoint subsets of variables in a causal DAG $\mathcal{G}$. To define the causal effect of an intervention on a response variable, we consider the post-interventional distribution $f(\mathbf{v} \mid do(\mathbf{x}))$. Under the assumption of independent causal mechanisms, the joint density after fixing $\mathbf{X}$ to $\mathbf{x}$ by an external intervention factorises as

$$f\big(\mathbf{v} \mid do(\mathbf{x})\big) \;=\; \Big(\prod_{X_i \notin \mathbf{X}} f\big(x_i \mid pa(X_i, \mathcal{G})\big) \Big)\Big|_{\mathbf{X}=\mathbf{x}}, \tag{1}$$

where the truncation indicates that variables in $\mathbf{X}$ are set to the fixed values $\mathbf{x}$. When this truncated factorisation is expressible solely in terms of the observational distribution, the causal effect is said to be *identifiable*. A well-known sufficient condition is the back-door criterion (Pearl, 2009): if a set $\mathbf{Z}$ satisfies the back-door criterion relative to $(\mathbf{X}, \mathbf{Y})$, then $f(\mathbf{y} \mid do(\mathbf{x})) = \int f(\mathbf{y} \mid \mathbf{x}, \mathbf{z})\, f(\mathbf{z})\, d\mathbf{z}$. Following Pearl (2009), the *total causal effect* of $\mathbf{X}$ on $\mathbf{Y}$ is defined as the change in the distribution of $\mathbf{Y}$ induced by an external intervention on $\mathbf{X}$. For a continuous treatment variable X, the total causal effect is often quantified by $\frac{\partial}{\partial x}\mathbb{E}[Y \mid do(X = x)]$; for a discrete X, it is typically the contrast $f(\mathbf{y} \mid do(x_1)) - f(\mathbf{y} \mid do(x_0))$ or a similar function of the interventional distributions. In this paper we simply use "the causal effect of $\mathbf{X}$ on $\mathbf{Y}$" to refer to the total causal effect encoded by $f(\mathbf{y} \mid do(\mathbf{x}))$, and we say that $\mathbf{X}$ has *no* causal effect on $\mathbf{Y}$ if $f(\mathbf{y} \mid do(\mathbf{x}))$ does not vary with $\mathbf{x}$. For ease of presentation we assume X is binary; all results extend to other variable types.

When the underlying DAG is known only up to a Markov equivalence class (MEC) represented by an MPDAG $\mathcal{G}^*$, the causal effect of X on Y is identifiable if it is the same across all DAGs in the MEC; Perkovic (2020) proved this holds if and only if every b-possibly causal path from X to Y starts with a directed edge in the MPDAG. Throughout, we assume a fixed distribution f^* consistent with the graph model is known, and all causal effects of X on Y are defined with respect to f^*.

We use $\mathcal{E}(X, Y; \mathcal{G}^*)$ to denote the set of all possible causal effects of X on Y across all DAGs in the MEC. Applying background knowledge $\mathcal{B}$ to $\mathcal{G}^*$ yields a new MPDAG $\mathcal{G}^{**}$. If

$$\mathcal{E}(X, Y; \mathcal{G}^{**}) \subsetneq \mathcal{E}(X, Y; \mathcal{G}^*),$$

we say that background knowledge $\mathcal{B}$ improves the identifiability of the causal effect of X on Y in $\mathcal{G}^*$.

In this paper, however, we are concerned with a more nuanced question: whether local background knowledge can *shrink the set of possible causal effects* even when the effect remains non-identified. The possible causal effect set is the collection of causal effects obtained from all DAGs in the current equivalence class; shrinking this set is a weaker but practically important form of improvement compared with point identification.

2.3 Related Work

Causal discovery from observational data typically recovers only a Markov equivalence class, represented by a CPDAG (Andersson et al., 1997; Spirtes et al., 2000; Chickering, 2002). Incorporating background knowledge about edge directions can refine the equivalence class, leading to a restricted class encoded by a MPDAG (Meek, 1995; Perković et al., 2017).

When the target parameter is the causal effect of a single treatment X on an outcome Y, local causal inference methods obviate the need to learn the full graph. The IDA algorithm (Maathuis et al., 2009; 2010; Nandy et al., 2017) estimates intervention effects by enumerating local structures around X in the CPDAG. In the context of MPDAGs, Perkovic (2020) established necessary and sufficient conditions for identifying the causal effect of $\mathbf{X}$ on $\mathbf{Y}$. Wang et al. (2023b) extended this local causal inference framework to settings with latent variables, broadening its applicability to more realistic scenarios with unobserved confounders. Fang & He (2020) investigated the validity of incorporating background knowledge locally and derived a necessary and sufficient condition for a set of directed edges around X to be compatible with the current MPDAG. Notably, their criterion assumes that the orientations of *all* siblings of X are specified, a requirement that is often unrealistic in practice.

Despite these advances, existing work leaves open a fundamental question: how to determine, using only local information around X, whether a given piece of background knowledge actually improves the identifiability of the causal effect. No local criterion currently exists for this task, and the naive approach requires enumerating all DAGs in the MEC. Our work fills this gap by introducing the concept of possible mediators and developing a local search procedure that yields a sound and complete condition for identifiability improvement. We also relax the full-sibling assumption of Fang & He (2020), allowing only a partial specification of edge directions around X.

3 Validity and Local Orientation Rules of Background Knowledge in MPDAGs

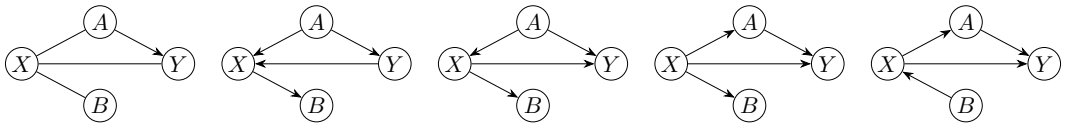
Given an MPDAG $\mathcal{G}^*$ and a target node X, consider a set of background knowledge $\mathcal{B}(X)$ specifying the orientations of some edges incident to X: $\mathcal{B}(X) = \{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$, where $\mathbf{S}_1, \mathbf{S}_2$ are disjoint subsets of $sib(X, \mathcal{G}^*)$. Our goals are to determine (1) whether $\mathcal{B}(X)$ is compatible with $\mathcal{G}^*$, and (2) if so, how it locally orients other undirected edges incident to X in the resulting MPDAG.

3.1 Validity of Local Background Knowledge

We first formally define the validity of background knowledge. A set of background knowledge $\mathcal{B}$ is said to be *valid* for an MPDAG $\mathcal{G}^*$ if there exists at least one DAG in the Markov equivalence class represented by $\mathcal{G}^*$ (denoted by $[\mathcal{G}^*]$) that is consistent with all edge directions in $\mathcal{B}$. Conversely, $\mathcal{B}$ is invalid if no such DAG exists. The following example illustrates a scenario of invalid background knowledge.

Example 3.1 *In Figure 2, $\mathcal{G}^*$ is an MPDAG with its corresponding MEC containing 4 different DAGs ($\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4$). Then the background knowledge $\{X \to A, X \to B\}$ is valid, but the background knowledge $\{A \to X, B \to X\}$ is not valid, because no DAG in this MEC has such local structure.*

A naive approach to check validity is to enumerate every DAG in $[\mathcal{G}^*]$, which is computationally prohibitive for large graphs. Our first main result provides a purely local, graphical criterion.



(a) $\mathcal{G}^*$ (b) $\mathcal{G}_1$ (c) $\mathcal{G}_2$ (d) $\mathcal{G}_3$ (e) $\mathcal{G}_4$

Figure 2: An MPDAG $\mathcal{G}^*$ and the four DAGs of its Markov equivalence class.

Theorem 3.2 *Let $\mathcal{G}^*$ be an MPDAG. For any vertex X, let $\mathbf{S}_1, \mathbf{S}_2 \subseteq sib(X, \mathcal{G}^*)$ and $\mathbf{S}_1 \cap \mathbf{S}_2 = \emptyset$. Then the background knowledge $\mathbf{S}_1 \to X$ and $X \to \mathbf{S}_2$ is valid if and only if $\mathbf{S}_1$ forms a complete graph, and $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$.*

The first condition prevents the creation of new v-structures colliding at X, while the second prevents directed cycles involving X. Revisiting Example 3.1, the invalid set $\{A \to X,\ B \to X\}$ violates the first condition because $\{A, B\}$ is not a clique in $\mathcal{G}^*$, immediately confirming its invalidity.

When the background knowledge completely specifies the orientations of *all* siblings of X, i.e., $\mathbf{S}_1 \cup \mathbf{S}_2 = sib(X, \mathcal{G}^*)$, Theorem 3.2 reduces to Theorem 1 in Fang & He (2020). For the special case where $\mathcal{G}^*$ is a CPDAG, the absence of semi-directed cycles simplifies the criterion further.

Corollary 3.3 *Let $\mathcal{G}^*$ be a CPDAG. For any vertex X, let $\mathbf{S}_1, \mathbf{S}_2 \subseteq sib(X, \mathcal{G}^*)$ and $\mathbf{S}_1 \cap \mathbf{S}_2 = \emptyset$. Then the background knowledge $\mathbf{S}_1 \to X$ and $X \to \mathbf{S}_2$ is valid if and only if $\mathbf{S}_1$ forms a complete graph.*

3.2 Local Orientation Rule

Once validity is established, the background knowledge $\{\mathbf{S}_1 \to X,\ X \to \mathbf{S}_2\}$ can be incorporated into $\mathcal{G}^*$, yielding a refined MPDAG $\mathcal{G}'$ that represents the restricted MEC. A global way to obtain $\mathcal{G}'$ is to enumerate all DAGs in $[\mathcal{G}^*]$, keep those consistent with the background knowledge, and then construct the MPDAG from the remaining set. This is again computationally heavy. Alternatively, one can simply apply Meek's orientation rules (Meek, 1995) starting from the forced orientations (also illustrated in Appendix). Our next theorem shows that the resulting neighbourhood of X in $\mathcal{G}'$ can be determined by a *local* calculation, without any global propagation.

Theorem 3.4 *Let $\mathcal{G}^*$ be an MPDAG and $\mathcal{B}$ the valid background knowledge $\{\mathbf{S}_1 \to X,\ X \to \mathbf{S}_2\}$. Denote by $\mathcal{G}'$ the new MPDAG representing the restricted MEC attained by incorporating $\mathcal{B}$ into $\mathcal{G}^*$. Define the following subsets of $sib(X, \mathcal{G}^*)$:*

$$\mathbf{S}_3 = sib(X, \mathcal{G}^*) \setminus \mathbf{S}_1 \setminus \Big(\bigcap_{A \in \mathbf{S}_1} adj(A, \mathcal{G}^*) \Big),$$

$$\mathbf{S}_4 = sib(X, \mathcal{G}^*) \cap \left(\bigcup_{A \in \mathbf{S}_2} ch(A, \mathcal{G}^*) \right),$$

$$\mathbf{S}_5 = sib(X, \mathcal{G}^*) \cap \left(\bigcup_{A \in \mathbf{S}_1} pa(A, \mathcal{G}^*) \right).$$

Then

$$pa(X, \mathcal{G}') = pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \cup \mathbf{S}_5, \qquad ch(X, \mathcal{G}') = ch(X, \mathcal{G}^*) \cup \mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4.$$

Interpretation. $\mathbf{S}_3$ contains siblings of X that are not adjacent to *every* node in $\mathbf{S}_1$; they become children by Meek's rule R1. $\mathbf{S}_4$ are siblings that are children of $\mathbf{S}_2$; they become children by rule R2. $\mathbf{S}_5$ are siblings that are parents of $\mathbf{S}_1$; they become parents by rule R2. The remaining siblings of X (those not in $\mathbf{S}_1 \cup \mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4 \cup \mathbf{S}_5$) stay undirected.

A particularly useful consequence of this theorem is that the orientation of a single undirected edge $X - Z$ depends only on a very small induced subgraph.

Proposition 3.5 *Under the conditions of Theorem 3.4, for any node $Z \in sib(X, \mathcal{G}^*)$, the direction of the edge between X and Z in $\mathcal{G}'$ is determined solely by the induced subgraph of $\mathcal{G}^*$ over $\mathbf{S}_1 \cup \mathbf{S}_2 \cup \{X, Z\}$.*

Together, Theorem 3.2 and 3.4 give a complete, local characterization of whether background knowledge around X is compatible with an MPDAG and how it orients the undirected edges incident to X, forming the basis for our subsequent analysis on causal effect identification improvement.

4 The Role of Background Knowledge in Causal Inference

4.1 Possible Mediator and Impossible Mediator

To study how the causal effect of X on Y is affected by background knowledge, we partition the siblings of X into two sets: *possible mediators* and *impossible mediators*.

Definition 4.1 (Possible and impossible mediators) *Let $\mathcal{G}^*$ be an MPDAG and let X, Y be distinct nodes in $\mathcal{G}^*$ with $Y \notin pa(X, \mathcal{G}^*)$. For any $Z \in sib(X, \mathcal{G}^*)$,*

- *Z is called an **impossible mediator** from X to Y in $\mathcal{G}^*$ if there is* no *b-possibly causal path from X to Y that starts with the undirected edge $X - Z$. The set of all impossible mediators is denoted by $IM(X, Y, \mathcal{G}^*)$.*
- *Z is called a **possible mediator** from X to Y in $\mathcal{G}^*$ if it is not an impossible mediator. The set of all possible mediators is denoted by $PM(X, Y, \mathcal{G}^*)$.*

Example 4.2 *In Figure 2, the only possible mediator is A, i.e., $PM(X, Y, \mathcal{G}^*) = \{A\}$. Indeed, A acts as a mediator from X to Y in $\mathcal{G}_3$ and $\mathcal{G}_4$, but not in $\mathcal{G}_1$ or $\mathcal{G}_2$; hence it is a "possible" mediator.*

The next result establishes the connection between possible mediators and the identifiability of the total causal effect.

Theorem 4.3 *Let $\mathcal{G}^*$ be an MPDAG and let X, Y be distinct nodes with $Y \notin pa(X, \mathcal{G}^*)$. Suppose we add valid background knowledge $\mathbf{S}_1 \to X$ and $X \to \mathbf{S}_2$ to obtain a new MPDAG $\mathcal{G}'$. If $PM(X, Y, \mathcal{G}^*) = PM(X, Y, \mathcal{G}')$, then*

$$\mathcal{E}(X, Y; \mathcal{G}^*) = \mathcal{E}(X, Y; \mathcal{G}').$$

A direct corollary of Theorem 4.3 is the identifiability condition of Perkovic (2020), see also Theorem 3.1 in Guo & Perkovic (2021), namely, the absence of possible mediators. Indeed, the absence of possible mediators is equivalent to the absence of b-possible causal paths. Conversely, a change in the possible mediator set implies that the background knowledge can actually affect the causal effect.

Theorem 4.4 *Let $\mathcal{G}^*$ and $\mathcal{G}'$ be as above, with $Y \notin pa(X, \mathcal{G}^*)$. If $PM(X, Y, \mathcal{G}^*) \neq PM(X, Y, \mathcal{G}')$, then there exists a distribution f consistent to $\mathcal{G}^*$. For this distribution, then*

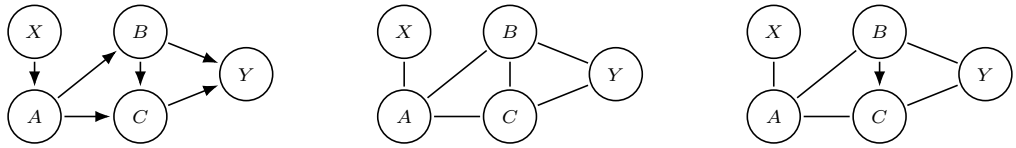
$$\mathcal{E}(X, Y; \mathcal{G}') \subsetneq \mathcal{E}(X, Y; \mathcal{G}^*).$$

The next example illustrates a complementary phenomenon: the number of b-possibly causal paths can decrease while the possible mediator set, and therefore the set of possible causal effects, remains unchanged. Together, these cases highlight that although possible mediators and b-possibly causal paths are closely related, they play different roles from the perspective of causal effects—only changes in the possible mediator set matter for identifiability.

Example 4.5 *Assume that Figure 3(a) represents a DAG and Figure 3(b) represents its corresponding CPDAG. When the background knowledge $B \to C$ is incorporated, the number of b-possible causal paths from X to Y changes. For instance, the path $X - A - C - B - Y$ is no longer a b-possible causal path, because traversing the edge between B and C from C to B conflicts with the known direction $B \to C$. Nevertheless, we find that the size of the set of all possible causal effects from X to Y remains unchanged. This observation suggests that the size of the possible causal effect set is not determined by the number of b-possible causal paths.*

Thus, to decide whether a given piece of background knowledge improves identifiability, we only need to check whether it shrinks the possible mediator set. Interestingly, even if $\mathbf{S}_1$ and $\mathbf{S}_2$ contain no nodes from $PM(X, Y, \mathcal{G}^*)$, the possible mediator set may still change.

Example 4.6 *Revisit Figure 2. Initially $PM(X, Y, \mathcal{G}^*) = \{A\}$ and $B \notin PM(X, Y, \mathcal{G}^*)$. Adding the background knowledge $B \to X$ forces, by the local orientation rules, $X \to A$. Then A is no longer a possible mediator in the new MPDAG, so the set changes.*



(a) DAG (b) CPDAG (c) Adding $B \to C$

Figure 3: Three graphs on the same node set: (a) a complete DAG, (b) the corresponding CPDAG with all edges undirected, and (c) the graph after adding the background knowledge $B \to C$.

4.2 Essential b-Possibly Causal Paths

Because the removal of a possible mediator depends on the underlying b-possibly causal paths, we introduce a notion capturing minimal such paths.

Definition 4.7 (Essential b-possibly causal path) *Let $\mathcal{G}^*$ be an MPDAG and X, Y distinct nodes with $Y \notin pa(X, \mathcal{G}^*)$. For $Z \in PM(X, Y, \mathcal{G}^*)$, an **essential b-possibly causal path** from X to Y with Z is a b-possibly causal path from X to Y that starts with the undirected edge $X - Z$ and whose subpath from Z to Y is chordless.*

Every possible mediator possesses at least one essential b-possibly causal path, and adding background knowledge cannot create new ones.

Theorem 4.8 *If $Z \in PM(X, Y, \mathcal{G}^*)$, then there exists at least one essential b-possibly causal path from X to Y with Z in $\mathcal{G}^*$.*

The survival of an essential path under new background knowledge can be decided by a purely local criterion.

Theorem 4.9 *Let $\mathcal{G}^*$ be an MPDAG and add valid background knowledge $\mathbf{S}_1 \to X$, $X \to \mathbf{S}_2$ to obtain $\mathcal{G}'$. Suppose p is an essential b-possibly causal path from X to Y starting with $X - Z$ in $\mathcal{G}^*$. Then p is* no longer *an essential b-possibly causal path in $\mathcal{G}'$ if and only if at least one of the following conditions holds:*

(1) some node on p (other than X) belongs to $pa(X, \mathcal{G}')$;

(2) $X \to Z$ in $\mathcal{G}'$.

By Theorem 4.9 and the local orientation rules (Theorem 3.4), changes in the PM set—and hence whether background knowledge improves identifiability—can be detected locally.

Example 4.10 (Possible mediator reduction via background knowledge) *Consider the MPDAG in Figure 1(b) derived from the ASIA network, where the edges $X = Smoking - LungCancer$ and $X - Bronchitis$ are undirected due to Markov equivalence. We are interested in the causal effect of Smoking (X) on Dyspnoea (Y). One can verify that $sib(X, \mathcal{G}^*) = \{LungCancer, Bronchitis\}$ and that both siblings lie on some b-possibly causal path from X to Y, therefore $PM(X, Y, \mathcal{G}^*) = \{LungCancer, Bronchitis\}$.*

Now suppose an expert states that Smoking is a direct cause of LungCancer, i.e. Smoking $\to$ LungCancer. By Theorem 3.2, this expert knowledge is valid. Incorporating it according to Theorem 3.4, Bronchitis remains a sibling of X in the new MEC. Consequently, the only essential

Algorithm 1 Decision Procedure for Identifiability Improvement

Require: An MPDAG $\mathcal{G}^*$, distinct nodes X, Y with $Y \notin pa(X, \mathcal{G}^*)$, and valid background knowledge $\mathcal{B} = \{\mathbf{S}_1 \to X,\ X \to \mathbf{S}_2\}$.
Ensure: **true** if $\mathcal{B}$ can improve the identifiability of the causal effect of X on Y for some distribution consistent with $\mathcal{G}^*$; **false** otherwise.

```
if B is not valid for G* (Theorem 3.2) then
    return false
end if
Compute PM = PM(X, Y, G*)(see algorithm 2 in appendix).
Compute S_3, S_4, S_5 according to Theorem 3.4.
for each Z ∈ PM do
    If Z is not a possible mediator after adding B (use algorithm 3 in appendix), mark Z as removed.
end for
if at least one Z is marked as removed then
    return true
else
    return false
end if
```

b-possibly causal path starting with an undirected edge $X - Z$ and reaching Y is the one through Bronchitis. Hence the new possible mediator set becomes $PM(X, Y, \mathcal{G}') = \{$Bronchitis$\}$, which is strictly smaller than the original one. The reduction of the possible mediator set may improve the identifiability of the causal effect.

Algorithm 1 summarizes the resulting local procedure for deciding whether candidate background knowledge improves identifiability of any target quantity. The correctness of Algorithm 1 follows directly from Theorems 4.3, 4.4, and 4.9. A detailed complexity analysis is deferred to Appendix; in summary, the algorithm runs in polynomial time (specifically $\mathrm{poly}(n, m)$), where n and m denote the number of nodes and edges of the MPDAG, respectively. Importantly, it avoids enumerating the entire Markov equivalence class.

5 Empirical Evaluation on ASIA Network

In this section, we evaluate the proposed local criteria on the ASIA network shown in Figure 1. Throughout this section, we use the shorthand $A =$ Asia, $S =$ Smoking, $T =$ Tuberculosis, $L =$ LungCancer, $B =$ Bronchitis, $D =$ Dyspnoea.

In the first part, we assume that the true CPDAG $\mathcal{G}^*$ is known and focus on the effect of background knowledge without structure learning uncertainty. The initial MPDAG is taken to be the CPDAG in Figure 1(b). Its Markov equivalence class contains only six DAGs because the reversible edges are $A - T$, $S - L$, and $S - B$, and the orientation $L \to S \leftarrow B$ would create a new v-structure and is therefore not allowed. We calculate the set of possible causal effects from the standard ASIA conditional probability tables (Lauritzen & Spiegelhalter, 1988). Table 1 summarises the six cases as follows: For the causal effect of T on D, the possible mediator set is already empty in the initial CPDAG, so the background knowledge about $A - T$ does not improve identifiability. For the causal effect of S on D, the initial possible mediator set is $\{L, B\}$, and the possible effect set across the full Markov equivalence class is $\{0.310, 0.283, 0.043\}$. For the causal effect of S on D, adding $S \to L$ or $S \to B$ removes one possible mediator and strictly shrinks the possible effect set. Adding $L \to S$ or $B \to S$ forces the remaining undirected edge by Meek's rule, removes all possible mediators, and therefore yields a point-identified causal effect. These results agree exactly with Theorems 4.3, 4.4, and 4.9.

The results above assume the true CPDAG is known. We now relax this and examine the procedure when the CPDAG is estimated by the MMHC algorithm (Tsamardinos et al., 2006) from finite samples of sizes $n \in \{200, 300, 500, 1000, 2000\}$ drawn from the true ASIA distribution. For each dataset we record whether the local decision procedure correctly identifies whether the background knowledge improves identifiability; the proportion over 1000 replications is shown in Table 2. In

Table 1: Simulation results for the ASIA network. For the background knowledge $\mathcal{B}$ in each case, we denote the resulting MPDAG by $\mathcal{G}'$.

Case	$X \to Y$	Background $\mathcal{B}$	PM_0	PM_1	$\mathcal{E}(X, Y; \mathcal{G}^*)$	$\mathcal{E}(X, Y; \mathcal{G}')$
1	$T \to D$	$A \to T$	$\emptyset$	$\emptyset$	$\{0.328\}$	$\{0.328\}$
2	$T \to D$	$T \to A$	$\emptyset$	$\emptyset$	$\{0.328\}$	$\{0.328\}$
3	$S \to D$	$S \to L$	$\{L, B\}$	$\{B\}$	$\{0.310, 0.283, 0.043\}$	$\{0.310, 0.043\}$
4	$S \to D$	$L \to S$	$\{L, B\}$	$\emptyset$	$\{0.310, 0.283, 0.043\}$	$\{0.283\}$
5	$S \to D$	$S \to B$	$\{L, B\}$	$\{L\}$	$\{0.310, 0.283, 0.043\}$	$\{0.310, 0.283\}$
6	$S \to D$	$B \to S$	$\{L, B\}$	$\emptyset$	$\{0.310, 0.283, 0.043\}$	$\{0.043\}$

addition, we evaluate the proposed local criteria on the Gaussian ECOLI70 network Schäfer & Strimmer (2005) under finite-sample structure learning; the detailed setup and results are deferred to Appendix B.

Table 2: Proportion of correct identifications for different sample sizes in the ASIA network. Cases 1–6 correspond to the background knowledge scenarios in Table 1.

			Sample size n				
Case	$X \to Y$	Background $\mathcal{B}$	200	300	500	1000	2000
1	$T \to D$	$A \to T$	99.9%	99.9%	99.4 %	99.3%	99.7 %
2	$T \to D$	$T \to A$	99.9%	99.9%	99.4 %	99.3 %	99.7%
3	$S \to D$	$S \to L$	77.4%	86.4%	98.5 %	100.0 %	100.0 %
4	$S \to D$	$L \to S$	77.4%	86.4%	98.5 %	100.0 %	100.0 %
5	$S \to D$	$S \to B$	77.4%	86.4 %	98.5 %	100.0 %	100.0 %
6	$S \to D$	$B \to S$	77.4%	86.4 %	98.5%	100.0 %	100.0%

6 Conclusion

In this paper, we studied the role of *local* background knowledge about edges incident to a treatment X in improving the identifiability of the causal effect of X on an outcome Y within an MPDAG, addressing the three questions posed in the introduction: when is such knowledge compatible with the current MPDAG, how does it orient other undirected edges incident to X, and does it improve identifiability of the target causal effect.

Our contributions are purely local graphical criteria answering each question *without* enumerating the MEC: Theorem 3.2 gives a necessary and sufficient condition for validity of a partial orientation specification around X; Theorem 3.4 orients undirected edges incident to X from a small induced subgraph; and, via *possible mediators* (PM), we proved that background knowledge improves identifiability iff the PM set between X and Y shrinks (Theorems 4.3, 4.4), with Theorem 4.9 deciding disappearance of an essential b-possibly causal path by two local conditions. These results yield a polynomial-time local procedure for identifiability improvement.

The proposed local criteria can be directly integrated into existing causal inference algorithms such as IDA for MPDAGs, avoiding the computationally prohibitive enumeration of all DAGs in the equivalence class. They are also applicable in interactive settings where a domain expert iteratively supplies small amounts of background knowledge and expects immediate feedback on whether the new knowledge helps to pinpoint the causal effect of interest.

Several directions remain open. Extending the possible-mediator framework to joint interventions, path-specific effects, non-linear models, or settings with latent confounders would broaden its applicability; in particular, accommodating hidden variables (Wang et al., 2023a) is important, as unobserved common causes are ubiquitous and may substantially affect identifiability. An empirical evaluation on large-scale benchmarks would also quantify the computational savings and success rate of the local approach under realistic background-knowledge patterns.

AI USE STATEMENT

In this work, we used generative AI tools to edit and polish the writing of this paper to improve readability. We have not used generative AI tools for retrieval and discovery (e.g., finding related work), research ideation or execution, drafting sections of the paper, generating synthetic datasets, or proving mathematical claims. We have reviewed all AI-assisted work, and all AI-assisted text was verified by the authors. We take responsibility for the final content of this work, including all text, claims, and artifacts.

REFERENCES

Steen A. Andersson, David Madigan, and Michael D. Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 04 1997.

David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of machine learning research*, 3(Nov):507–554, 2002.

Zhuangyan Fang and Yangbo He. Ida with background knowledge. In *Conference on Uncertainty in Artificial Intelligence*, pp. 270–279. PMLR, 2020.

Paula Gradu, Tijana Zrnic, Yixin Wang, and Michael I Jordan. Valid inference after causal discovery. *Journal of the American Statistical Association*, 120(550):1127–1138, 2025.

Richard Guo and Emilija Perkovic. Minimal enumeration of all possible total effects in a markov equivalence class. In *International Conference on Artificial Intelligence and Statistics*, pp. 2395–2403. PMLR, 2021.

Steffen L. Lauritzen and David J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 50(2):157–224, 1988.

Yue Liu, Zhuangyan Fang, Yangbo He, Zhi Geng, and Chunchen Liu. Local causal network learning for finding pairs of total and direct effects. *Journal of Machine Learning Research*, 21(148):1–37, 2020. URL `http://jmlr.org/papers/v21/18-083.html`.

M. H. Maathuis, M. Kalischand, and P. Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A):3133–3164, 2009. ISSN 0090-5364.

Marloes H Maathuis, Diego Colombo, Markus Kalisch, and Peter Bühlmann. Predicting causal effects in large-scale systems from observational data. *Nature methods*, 7(4):247–248, 2010.

Christopher Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 403–410. Morgan Kaufmann Publishers Inc., 1995.

Preetam Nandy, Marloes H Maathuis, and Thomas S Richardson. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *The Annals of Statistics*, pp. 647–674, 2017.

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

Emilija Perkovic. Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Conference on Uncertainty in Artificial Intelligence*, pp. 530–539. PMLR, 2020.

Emilija Perković, Markus Kalisch, and Marloes H Maathuis. Interpreting and using CPDAGs with background knowledge. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*. AUAI press, 2017.

Juliane Schäfer and Korbinian Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1), 2005.

David J Spiegelhalter and Steffen L Lauritzen. Techniques for bayesian analysis in expert systems. *Annals of Mathematics and Artificial Intelligence*, 2(1):353–366, 1990.

Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, second edition, 2000.

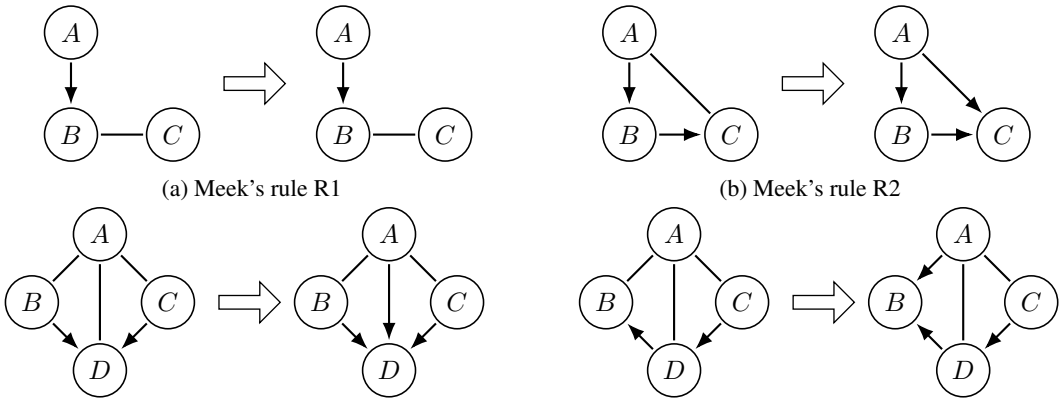
Ioannis Tsamardinos, Laura E. Brown, and Constantin F. Aliferis. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*, 65(1):31–78, Oct 2006.

T. Verma and J. Pearl. An algorithm for deciding if a set of observed independencies has a causal explanation. In *Proceedings of the Eighth international conference on uncertainty in artificial intelligence*, pp. 323–330. Morgan Kaufmann Publishers Inc., 1992.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pp. 220–227. Elsevier Science Inc., 1990.

Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Estimating possible causal effects with latent variables via adjustment. In *International Conference on Machine Learning*, pp. 36308–36335. PMLR, 2023a.

Tian-Zuo Wang, Tian Qin, and Zhi-Hua Zhou. Sound and complete causal identification with latent variables given local background knowledge. *Artificial Intelligence*, 322:103964, 2023b.



(a) Meek's rule R1

(b) Meek's rule R2

(c) Meek's rule R3

(d) Meek's rule R4

Figure 4: Illustration of Meek's rules. For each rule, if an induced subgraph of the PDAG matches the pattern on the left, the specified edge is oriented as shown on the right.

A Appendix

A.1 Illustration of Meek's rules

Some works (Verma & Pearl, 1992; Meek, 1995) proposed that a CPDAG can be obtained from the pattern of a DAG by iteratively applying a set of orientation rules known as the Meek's rules. These rules repeatedly search for the following structures in a partially directed graph and orient edges accordingly until no further changes occur:

R1. For each $A \to B - C$ where A and C are not adjacent, orient $B \to C$;

R2. For each $A \to B \to C$ where $A - C$, orient $A \to C$;

R3. For each $A - B \to D$, $A - C \to D$, $A - D$, and B, C not adjacent, orient $A \to D$;

R4. For each $C \to D \to B$, $C - A - B$, $A - D$, and B, C not adjacent, orient $A \to D$.

Figure 4 illustrates these four rules.

A.2 Proof of Theorem 3.2

First, we introduce three lemmas:

Lemma A.1 (Theorem 1 in Fang & He (2020)) *Let $\mathcal{G}^*$ be an MPDAG. For any vertex X and $\boldsymbol{S} \subset sib(X, \mathcal{G}^*)$, the following two statements are equivalent:*

(1) There is a DAG $\mathcal{G}$ in the restricted Markov equivalence class represented by $\mathcal{G}^*$ such that $pa(X, \mathcal{G}) = \boldsymbol{S} \cup pa(X, \mathcal{G}^*)$ and $ch(X, \mathcal{G}) = sib(X, \mathcal{G}^*) \cup ch(X, \mathcal{G}^*) \setminus \boldsymbol{S}$.

(2) $\mathcal{G}^*_{\boldsymbol{S} \to X}$ does not have any new v-structure collided on X or any directed triangle containing X.

Lemma A.2 *Let $\mathcal{G}^*$ be an MPDAG. For any vertex X, and $A, B \in sib(X, \mathcal{G}^*)$, if $A \in an(B, \mathcal{G}^*)$, then $A \in pa(B, \mathcal{G}^*)$.*

Proof. For the sake of contradiction, assume that $A \notin pa(B, \mathcal{G}^*)$. Since $A \in an(B, \mathcal{G}^*)$, let $p = \langle A = A_0, A_1, \ldots, A_k = B \rangle$ be the shortest directed path from A to B in $\mathcal{G}^*$. First we know that A and B are not adjacent. Otherwise, the edge between A and B must be $A \to B$ in every $\mathcal{G} \in [\mathcal{G}^*]$ to avoid cycle, so it is also $A \to B$ in $\mathcal{G}^*$, which contradicts with our assumption. So $k \geq 2$.

Then we show that every node in p is in $sib(X, \mathcal{G}^*)$. Otherwise, let j be the largest index such that $A_j \notin sib(X, \mathcal{G}^*)$, then $0 < j < k$. If A_j and X are not adjacent, then $A_j \to A_{j+1}$ and $X - A_{j+1}$

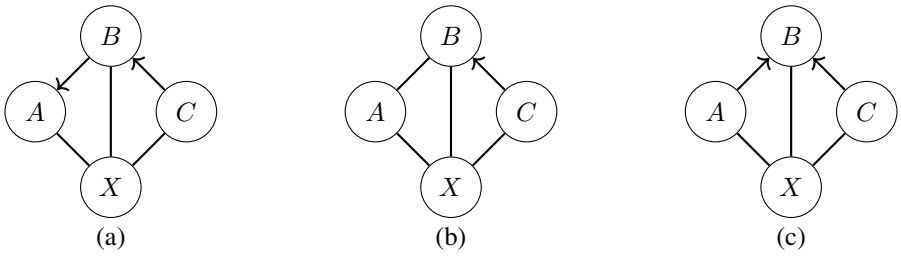


Figure 5: Illustration for contradiction in the proof of Lemma A.3.

contradicts with Meek's rule R1. If $X \to A_j \in \mathcal{G}^*$, then $A_j \to A_{j+1}$ and $X - A_{j+1}$ contradicts with Meek's rule R2. Finally assume $A_j \to X \in \mathcal{G}^*$, then A_j and A are adjacent, otherwise $A - X$ contradicts with Meek's rule R1. Consider the edge between A_j and A. This edge can only be $A \to A_j$, since for every $\mathcal{G} \in [\mathcal{G}^*]$ we have $A \to A_j \in \mathcal{G}$ to avoid cycle. However, the edges $A \to A_j, A_j \to X, A - X$ contradicts with Meek's rule R2. Since all other assumption leads to contradiction, every node in p is in $sib(X, \mathcal{G}^*)$.

Finally we construct a contradiction by showing that $A_{k-2} \to B$. If A_{k-2} and B are not adjacent, by Meek's rule R4 we have $X \to B$, leading to a contradiction. So A_{k-2} and B are adjacent. Since $A_{k-2} \in an(B, \mathcal{G}^*)$, for every DAG $\mathcal{G} \in [\mathcal{G}^*]$ we have $A_{k-2} \to B \in \mathcal{G}$, so $A_{k-2} \to B \in \mathcal{G}^*$. This contradicts with the fact that p is the shortest directed path from A to B in $\mathcal{G}^*$. ■

Lemma A.3 *Let $\mathcal{G}^*$ be an MPDAG, X be a node in $\mathcal{G}^*$. let $A, B, C \in sib(X, \mathcal{G}^*)$ be three distinct nodes. If $C \to B \in \mathcal{G}^*$, A and B are adjacent, then A and C are adjacent.*

Proof. For the sake of contradiction, assume that A and C are not adjacent. Then as illustrated in Figure 5, all cases for the edge between A and B leads to contradiction with Meek's rules:

1. If $A \leftarrow B \in \mathcal{G}^*$, as illustrated in Figure 5(a), by Meek's rule R4 we have $X \to A \in \mathcal{G}^*$, contradiction.
2. If $A - B \in \mathcal{G}^*$, as illustrated in Figure 5(b), by Meek's rule R1 we have $A \leftarrow B$.
3. If $A \to B \in \mathcal{G}^*$, as illustrated in Figure 5(c), by Meek's rule R3 we have $X \to B$, contradiction.

Therefore, A and C are adjacent. ■

Proof. (Necessity.) Suppose that the background knowledge $\mathcal{B} = \{\mathbf{S}_1 \to X,\ X \to \mathbf{S}_2\}$ is valid. By definition, there exists a DAG $\mathcal{G} \in [\mathcal{G}^*]$ such that $\mathbf{S}_1 \subseteq pa(X, \mathcal{G})$ and $\mathbf{S}_2 \subseteq ch(X, \mathcal{G})$. Let $S \coloneqq pa(X, \mathcal{G}) \cap sib(X, \mathcal{G}^*)$. Since every directed edge of $\mathcal{G}^*$ appears in every DAG of $[\mathcal{G}^*]$, we have $pa(X, \mathcal{G}^*) \subseteq pa(X, \mathcal{G})$, and hence

$$pa(X, \mathcal{G}) = pa(X, \mathcal{G}^*) \cup S,$$
$$ch(X, \mathcal{G}) = ch(X, \mathcal{G}^*) \cup \big(sib(X, \mathcal{G}^*) \setminus S\big).$$

Thus $\mathcal{G}$ satisfies condition (1) of Lemma A.1, and Lemma A.1 implies that $\mathcal{G}^*_{S \to X}$ has no new v-structure collided on X and no directed triangle containing X. Then we prove:

1. $\mathbf{S}_1$ forms a complete graph. If two nodes $A, B \in S$ were non-adjacent in $\mathcal{G}^*$, then $A \to X \leftarrow B$ would be a new v-structure collided on X in $\mathcal{G}^*_{S \to X}$, a contradiction. Hence S is a clique in $\mathcal{G}^*$. Since $\mathbf{S}_1 \subseteq pa(X, \mathcal{G})$ and $\mathbf{S}_1 \subseteq sib(X, \mathcal{G}^*)$, we have $\mathbf{S}_1 \subseteq S$, so $\mathbf{S}_1$ also forms a complete graph.
2. $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$. Suppose, for a contradiction, that $A \in \mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*)$. Then there is $B \in \mathbf{S}_2$ with $B \to A \in \mathcal{G}^*$. In the DAG $\mathcal{G}$ we have $A \to X$ and $X \to B$; moreover every directed edge of $\mathcal{G}^*$ is an edge of $\mathcal{G}$, so $B \to A \in \mathcal{G}$. The three edges $X \to B$, $B \to A$, and $A \to X$ form the directed cycle $X \to B \to A \to X$ in $\mathcal{G}$, contradicting that $\mathcal{G}$ is a DAG. Hence $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$.

(Sufficiency.) We only need to prove that, for background knowledge $\{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$, where $\mathbf{S}_1$ forms a complete graph and $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$, there is a DAG $\mathcal{G}$ in the MEC $[\mathcal{G}^*]$ such that $\mathbf{S}_1 \subseteq pa(X, \mathcal{G})$ and $\mathbf{S}_2 \subseteq ch(X, \mathcal{G})$. Let $\mathbf{S} = an(\mathbf{S}_1, \mathcal{G}^*) \cap sib(X, \mathcal{G}^*)$. We will show that $\mathcal{G}^*_{\mathbf{S}\to X}$ does not have any v-structure collided on X or any directed triangle containing X.

• We first show that $\mathbf{S}$ forms a complete graph. For any node $X_1, X_2 \in S$, if $X_1 \in an(X_2, \mathcal{G}^*)$ or $X_2 \in an(X_1, \mathcal{G}^*)$, then by Lemma A.2, X_1 and X_2 are adjacent. Now assume that $X_1 \notin an(X_2, \mathcal{G}^*)$ and $X_2 \notin an(X_1, \mathcal{G}^*)$. Since $\mathbf{S} \subseteq an(\mathbf{S}_1, \mathcal{G}^*)$, there exists $A, B \in \mathbf{S}_1$ such that $X_1 \in an(A, \mathcal{G}^*)$ and $X_2 \in an(B, \mathcal{G}^*)$. Since $\mathbf{S}_1$ forms a complete graph, A and B are adjacent.

If $A = X_1$ or $B = X_2$, without loss of generality assume $A = X_1$. If $X_2 = B$ then X_1 and X_2 are adjacent. Otherwise by Lemma A.2 we have $X_2 \to B \in \mathcal{G}^*$. Then by Lemma A.3, X_2 and X_1 are adjacent.

Now assume that $A \neq X_1$ and $B \neq X_2$. By Lemma A.2, $X_1 \to A \in \mathcal{G}^*$ and $X_2 \to B \in \mathcal{G}^*$. If $A = X_2$ or $B = X_1$, then X_1 and X_2 are obviously adjacent, so assume $A \neq X_2$ and $B \neq X_1$. If $A = B$, then the assumption that X_1 and X_2 are not adjacent contradicts with Meek's rule R3. So assume $A \neq B$, i.e. X_1, X_2, A, B are four distinct nodes. Then by Lemma A.3, X_1 and B are adjacent and X_2 and A are adjacent. Moreover, by Lemma A.3 again, X_1 and X_2 are adjacent.

After all, for any two different node X_1 and X_2 in $\mathbf{S}$, they are adjacent, so $\mathbf{S}$ forms a complete graph. By Meek's rule R1, any node in $pa(X, \mathcal{G}^*)$ is also adjacent to $\mathbf{S}$. Therefore, $\mathcal{G}^*_{\mathbf{S}\to X}$ does not have any new v-structure collided on X comparing to $\mathcal{G}^*$.

• For any node $X_1 \in S$ and $X_2 \in sib(X, \mathcal{G}^*) \setminus \mathbf{S}$, if $X_2 \to X_1 \in \mathcal{G}^*$, then X_2 is an ancestor of S_1 in $\mathcal{G}^*$, so $X_2 \in \mathbf{S}$, leading to a contradiction. So $\mathcal{G}^*_{\mathbf{S}\to X}$ does not have any directed triangle containing X.

Therefore, by Lemma A.1, there exists an DAG $\mathcal{G} \in [\mathcal{G}^*]$ such that $\mathbf{S} \subseteq pa(X, \mathcal{G})$ and $sib(X, \mathcal{G}^*) \setminus \mathbf{S} \subseteq ch(X, \mathcal{G})$. Since $\mathbf{S}_1 \subseteq sib(X, \mathcal{G}^*)$ and $\mathbf{S} = an(\mathbf{S}_1, \mathcal{G}^*) \cap sib(X, \mathcal{G}^*)$, we have $\mathbf{S}_1 \subseteq \mathbf{S}$. Since $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$, by Lemma A.2 we have $\mathbf{S}_2 \cap \mathbf{S} = \emptyset$, so $\mathbf{S}_2 \subseteq sib(X, \mathcal{G}^*) \setminus \mathbf{S}$. Therefore, $\mathcal{G}$ is a DAG that is consistent with background knowledge $\{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$ and this background knowledge is valid. ■

A.3 Proof of Corollary 3.3

Proof. Since a CPDAG is also an MPDAG, by Theorem 3.2, we only need to show that $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$. For any $A \in \mathbf{S}_1, B \in \mathbf{S}_2$, we also have $A, B \in sib(X, \mathcal{G}^*)$. For the sake of contradiction assume that $B \to A \in \mathcal{G}^*$. By Lemma 1 in Meek (1995), with $B \to A$ and $A - X$ we have $B \to X$, which contradicts with $B \in sib(X, \mathcal{G}^*)$. ■

A.4 Proof of Theorem 3.4

Proof. We first show that $pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \cup \mathbf{S}_5 \subseteq pa(X, \mathcal{G}')$. Since $\mathcal{G}'$ is obtained by adding background knowledge $\{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$ to $\mathcal{G}^*$, we have $pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \subseteq pa(X, \mathcal{G}')$. Recall that $\mathbf{S}_5 = sib(X, \mathcal{G}^*) \cap pa(\mathbf{S}_1, \mathcal{G}^*)$. For every $A \in \mathbf{S}_5$ we have $A - X \in \mathcal{G}^*$ and there exists $B \in \mathbf{S}_1$ such that $A \to B \in \mathcal{G}^*$. After adding background knowledge $B \to X$, by Meek's rule R2 we have $A \to X \in \mathcal{G}'$. So $\mathbf{S}_5 \subseteq pa(X, \mathcal{G}')$.

Then we show that $ch(X, \mathcal{G}^*) \cup \mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4 \subseteq ch(X, \mathcal{G}')$. Since $\mathcal{G}'$ is obtained by adding background knowledge $\{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$ to $\mathcal{G}^*$, we have $ch(X, \mathcal{G}^*) \cup \mathbf{S}_2 \subseteq ch(X, \mathcal{G}')$. Recall that $\mathbf{S}_3 = sib(X, \mathcal{G}^*) \setminus \bigcap_{A \in \mathbf{S}_1} adj(A, \mathcal{G}^*)$. For every $B \in \mathbf{S}_3$, we have $B \in sib(X, \mathcal{G}^*)$ and there exists $A \in \mathbf{S}_1$ such that $B \notin adj(A, \mathcal{G}^*)$. After adding background knowledge $A \to X$, by Meek's rule R1 we have $X \to B \in \mathcal{G}'$. So $\mathbf{S}_3 \subseteq ch(X, \mathcal{G}')$. Recall that $\mathbf{S}_4 = sib(X, \mathcal{G}^*) \cap ch(\mathbf{S}_2, \mathcal{G}^*)$. For every $B \in \mathbf{S}_4$, we have $X - B \in \mathcal{G}^*$ and there exists $A \in \mathbf{S}_2$ such that $A \to B \in \mathcal{G}^*$. After adding background knowledge $X \to A$, by Meek's rule R2 we have $X \to B \in \mathcal{G}'$. So $\mathbf{S}_4 \subseteq ch(X, \mathcal{G}')$.

Finally we show that $pa(X, \mathcal{G}') \subseteq pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \cup \mathbf{S}_5$ and $ch(X, \mathcal{G}') \subseteq ch(X, \mathcal{G}^*) \cup \mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4$. In fact, we only need to show that $\mathbf{Z} := sib(X, \mathcal{G}^*) \setminus (\mathbf{S}_1 \cup \mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4 \cup \mathbf{S}_5) \subseteq sib(X, \mathcal{G}')$. Recall that $\mathcal{B} = \{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$ denotes the background knowledge. For every $Z \in \mathbf{Z}$, we only need to show that $\mathcal{B} \cup \{X \to Z\}$ and $\mathcal{B} \cup \{X \leftarrow Z\}$ are both valid for $\mathcal{G}^*$.

First consider $\mathcal{B} \cup \{X \to Z\}$. Let $\mathbf{S}_2' = \mathbf{S}_2 \cup \{Z\}$. Since $\mathcal{B}$ is valid, $\mathbf{S}_1$ forms a complete graph and $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$. Recall that $\mathbf{S}_5 = sib(X, \mathcal{G}^*) \cap pa(\mathbf{S}_1, \mathcal{G}^*)$ and $Z \in sib(X, \mathcal{G}^*) \setminus \mathbf{S}_5$, so $\mathbf{S}_1 \cap ch(Z, \mathcal{G}^*) = \emptyset$. That is, $\mathbf{S}_1 \cap ch(\mathbf{S}_2', \mathcal{G}^*) = \emptyset$. By Theorem 3.2, $\mathcal{B} \cup \{X \to Z\}$ is valid.

Then consider $\mathcal{B} \cup \{X \leftarrow Z\}$. Let $\mathbf{S}_1' = \mathbf{S}_1 \cup \{Z\}$. Recall that $\mathbf{S}_3 = sib(X, \mathcal{G}^*) \setminus \bigcap_{A \in \mathbf{S}_1} adj(A, \mathcal{G}^*)$. Since $Z \in sib(X, \mathcal{G}^*) \setminus \mathbf{S}_3$, we have $Z \in \bigcap_{A \in \mathbf{S}_1} adj(A, \mathcal{G}^*)$, so Z is adjacent to every node in $\mathbf{S}_1$. Since $\mathcal{B}$ is valid, $\mathbf{S}_1$ forms a complete graph and $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$, so $\mathbf{S}_1'$ also forms a complete graph. Recall that $\mathbf{S}_4 = sib(X, \mathcal{G}^*) \cap ch(\mathbf{S}_2, \mathcal{G}^*)$. Since $Z \in sib(X, \mathcal{G}^*) \setminus \mathbf{S}_4$, we have $Z \notin ch(\mathbf{S}_2, \mathcal{G}^*)$, so $\mathbf{S}_1' \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \emptyset$. By Theorem 3.2, $\mathcal{B} \cup \{X \leftarrow Z\}$ is valid.

Therefore, there exist DAGs $\mathcal{G}_1, \mathcal{G}_2 \in [\mathcal{G}^*]$ such that $\mathcal{G}_1$ is consistent with background knowledge $\mathcal{B} \cup \{X \to Z\}$, and $\mathcal{G}_2$ is consistent with background knowledge $\mathcal{B} \cup \{X \leftarrow Z\}$. Since both $\mathcal{G}_1$ and $\mathcal{G}_2$ are consistent with $\mathcal{B}$, they both belong to the restricted MEC of $\mathcal{G}^*$ under $\mathcal{B}$, i.e., the class represented by $\mathcal{G}'$. Thus the two orientations $X \to Z$ and $X \leftarrow Z$ both occur among the DAGs of this class, so by the defining property of an MPDAG (an edge is directed if and only if it is oriented identically in every DAG of the class, cf. Section 2), the edge between X and Z is undirected in $\mathcal{G}'$; that is, $Z \in sib(X, \mathcal{G}')$. This already rules out any global Meek cascade: such a cascade could orient $X - Z$ only if $X \to Z$ or $X \leftarrow Z$ is oriented identically in every DAG of this class.

■

A.5 Proof of Proposition 3.5

Proof. From Theorem 3.4, Z enters $pa(X, \mathcal{G}')$ only through $\mathbf{S}_5$, and $ch(X, \mathcal{G}')$ only through $\mathbf{S}_3$ or $\mathbf{S}_4$. All these sets are defined using only adjacency relations among $\mathbf{S}_1$, $\mathbf{S}_2$, X, and Z. Hence membership can be decided locally. ■

A.6 Proof of Theorem 4.3

We first need the following lemma:

Lemma A.4 *For a MPDAG $\mathcal{G}^*$, the causal effect from node X on Y is identified if and only if $PM(X, Y, \mathcal{G}^*)$ is an empty set.*

Proof. Directly follows Theorem 1 in Guo & Perkovic (2021). ■

Proof. Let $\mathcal{B}_1 = \{\mathbf{S}_1 \to X, X \to \mathbf{S}_2\}$ denote the valid background knowledge added to $\mathcal{G}^*$ to get $\mathcal{G}'$. We just need to prove that, for any DAG $\mathcal{G}_1 \in [\mathcal{G}^*]$, there exists an DAG $\mathcal{G}_2 \in [\mathcal{G}']$ such that the causal effects of X on Y are same in $\mathcal{G}_1$ and $\mathcal{G}_2$.

Let $\mathbf{T}_1 = PM(X, Y, \mathcal{G}^*) \cap pa(X, \mathcal{G}_1)$ and $\mathbf{T}_2 = PM(X, Y, \mathcal{G}^*) \cap ch(X, \mathcal{G}_1)$. Then we have $\mathbf{T}_1 \to X \in \mathcal{G}_1$ and $X \to \mathbf{T}_2 \in \mathcal{G}_1$. Denote $\mathcal{B}_2 = \{\mathbf{T}_1 \to X, X \to \mathbf{T}_2\}$. Then $\mathcal{B}_2$ is valid with respect to $\mathcal{G}^*$.

Then we show that background knowledge $\mathcal{B}_1 \cup \mathcal{B}_2$ is valid with respect to $\mathcal{G}^*$. Since $PM(X, Y, \mathcal{G}^*) = PM(X, Y, \mathcal{G}')$, we have $\mathbf{T}_1 \cup \mathbf{T}_2 \subseteq PM(X, Y, \mathcal{G}')$, and the edges between X and $\mathbf{T}_1 \cup \mathbf{T}_2$ are undirected in $\mathcal{G}'$ by the definition of PM. So we also have $(\mathbf{T}_1 \cup \mathbf{T}_2) \cap (\mathbf{S}_1 \cup \mathbf{S}_2) = \emptyset$.

Since $\mathcal{B}_2$ is valid with respect to $\mathcal{G}^*$, $\mathbf{T}_1$ form a complete graph and $\mathbf{T}_1 \cap ch(\mathbf{T}_2, \mathcal{G}^*) = \emptyset$. For every $S \in \mathbf{S}_1, T \in \mathbf{T}_1$, if S and T are not adjacent, then $S \to X, X - T \in \mathcal{G}'$ contradicts with Meek's rule R1, so S and T are adjacent. Therefore, $\mathbf{S}_1 \cup \mathbf{T}_1$ forms a complete graph.

If there exists $S \in \mathbf{S}_1$ and $T \in \mathbf{T}_2$ such that $S \in ch(T, \mathcal{G}^*)$, we also have $T \to S \in \mathcal{G}'$, but $T \to S, S \to X, X - T \in \mathcal{G}'$ contradicts with Meek's rule R2. If there exists $S \in \mathbf{S}_2$ and $T \in \mathbf{T}_1$ such that $T \in ch(S, \mathcal{G}^*)$, we also have $S \to T \in \mathcal{G}'$, but $X \to S, S \to T, T - X$ contradicts with Meek's rule R2. Therefore, $(\mathbf{S}_1 \cup \mathbf{T}_1) \cap ch(\mathbf{S}_2 \cup \mathbf{T}_2, \mathcal{G}^*) = \emptyset$. By Theorem 3.2, $\mathcal{B}_1 \cup \mathcal{B}_2$ is valid with respect to $\mathcal{G}^*$.

Let $\mathcal{G}''$ be the MPDAG obtained by applying $\mathcal{B}_2$ to $\mathcal{G}^*$. Since applying background knowledge only orient undirected edges into directed edges, any b-possibly causal path in $\mathcal{G}''$ is also a b-possibly causal path in $\mathcal{G}^*$, and any undirected edge in $\mathcal{G}''$ is also undirected in $\mathcal{G}^*$. Therefore $PM(X, Y, \mathcal{G}'') \subseteq PM(X, Y, \mathcal{G}^*) = \mathbf{T}_1 \cup \mathbf{T}_2$. Since the edge between X and $\mathbf{T}_1 \cup \mathbf{T}_2$ are all directed in $\mathcal{G}''$, we have $PM(X, Y, \mathcal{G}'') = \emptyset$. By Lemma A.4, the causal effect from X on Y is

identified in $\mathcal{G}''$. Let $\mathcal{G}_2 \in [\mathcal{G}'']$ be any DAG represented by $\mathcal{G}''$ such that $\mathcal{G}_2$ is also consistent with $\mathcal{B}_1$ (such $\mathcal{G}_2$ exists since $\mathcal{B}_1 \cup \mathcal{B}_2$ is valid with respect to $\mathcal{G}^*$), we also have $\mathcal{G}_2 \in [\mathcal{G}']$. Since $\mathcal{G}_1$ is also consistent with $\mathcal{B}_2$, we have $\mathcal{G}_1 \in [\mathcal{G}'']$, so the causal effects from X on Y are same in $\mathcal{G}_1$ and $\mathcal{G}_2$. ■

A.7 Proof of Theorem 4.4

Proof. In the following proof, we use $p \oplus q$ to denote the connection of two paths p and q, $p(A, B)$ to denote the sub-path of a path p from node A to B. In this proof we will use Meek's Rule 1 (Meek, 1995), which states that if $A \to B$ in an MPDAG, B and C are adjacent, A and C are not adjacent, and (A, B, C) does not form a v-structure, then $B \to C$ is in that MPDAG. This proof also relies on Theorem 4.9, whose proof does not rely on this theorem.

First of all, consider the case that $X - Y \in \mathcal{G}^*$ and $X \leftarrow Y \in \mathcal{G}'$. In this case, the causal effect of X on Y is always zero for every DAG in $[\mathcal{G}']$, and there exists a DAG $\mathcal{G} \in [\mathcal{G}^*] \setminus [\mathcal{G}']$ such that $X \to Y \in \mathcal{G}$. For a simple linear gaussian model on $\mathcal{G}$, the causal effect of X on Y can be non-zero. So the statement of the theorem holds.

After adding background knowledge $\mathbf{S}_1 \to X$ and $X \to \mathbf{S}_2$, only some undirected edges in $\mathcal{G}^*$ are changed to directed edges in $\mathcal{G}'$. Therefore, all undirected edges in $\mathcal{G}'$ are still undirected in $\mathcal{G}^*$, and all possibly causal path in $\mathcal{G}'$ is a possibly causal path in $\mathcal{G}^*$. By the definition of PM, we have $PM(X, Y, \mathcal{G}') \subseteq PM(X, Y, \mathcal{G}^*)$. Since $PM(X, Y, \mathcal{G}^*)$ and $PM(X, Y, \mathcal{G}')$ are different, there exists a node $Z_1 \in PM(X, Y, \mathcal{G}^*) \setminus PM(X, Y, \mathcal{G}')$.

If $Z_1 = Y$, then $X - Y \in \mathcal{G}^*$ and not in $\mathcal{G}'$. The case $X \leftarrow Y \in \mathcal{G}'$ is already considered. If $X \to Y \in \mathcal{G}'$, there exists an DAG $\mathcal{G} \in [\mathcal{G}^*] \setminus [\mathcal{G}']$ such that $X \leftarrow Y \in \mathcal{G}$. For a simple linear gaussian model consistent to a DAG with $X \to Y$ in $\mathcal{G}'$, the causal effect is not zero for every DAG in $[\mathcal{G}']$ but zero for $\mathcal{G}$, so the statement of the theorem holds. In the following proof, we can assume $Z_1 \neq Y$.

Let $p = \langle X, Z_1, Z_2, \ldots, Z_n, Y \rangle$, $n \geq 1$ be a path which satisfying the following conditions: (1) p is a possibly causal path beginning with $X - Z_1$ in $\mathcal{G}^*$ but not a possibly causal path beginning with $X - Z_1$ in $\mathcal{G}'$; (2) $X - Z_1 \in \mathcal{G}^*$; (3) For every $i = 2, 3, \ldots, n$, $Z_i \notin PM(X, Y, \mathcal{G}^*) \setminus PM(X, Y, \mathcal{G}')$. (4) p is the shortest path satisfying (1)-(3). We first show that such path p exists. Since $Z_1 \in PM(X, Y, \mathcal{G}^*) \setminus PM(X, Y, \mathcal{G}')$, every possibly causal path beginning with $X - Z_1$ in $\mathcal{G}^*$ is not a possibly causal path beginning with $X - Z_1$ in $\mathcal{G}'$, and there exists a possibly causal path beginning with $X - Z_1$ in $\mathcal{G}^*$, which we denote as $p_0 = \langle X, Z_1, Z_{01}, Z_{02}, \ldots, Z_{0s}, Y \rangle$. Then p_0 satisfies (1) and (2). If there exists $1 \leq i \leq s$ such that $Z_{0i} \in PM(X, Y, \mathcal{G}^*) \setminus PM(X, Y, \mathcal{G}')$, then we have $X - Z_{0i} \in \mathcal{G}^*$, so $p_1 \coloneqq \langle X, Z_{0i} \rangle \oplus p(Z_{0i}, Y)$ is a path satisfying (1) and (2), and is shorter than p_0. Since the length of path is always a positive integer, repeating this process and we will get a path satisfying (1)-(3). Then the shortest one satisfying (1)-(3), which we denote as p, satisfies (1)-(4).

By the selection of path p, rewriting $Z_0 = X$ and $Z_{n+1} = Y$ we have the following properties:

1. $p(Z_1, Y)$ is chordless. Otherwise, suppose that Z_i and Z_j are adjacent for $1 \leq i < j \leq n + 1, i + 1 < j$. Since p is possibly causal we have $Z_i \to Z_j \in \mathcal{G}^*$ or $Z_i - Z_j \in \mathcal{G}^*$. Therefore, $p(X, Z_i) \oplus \langle Z_i, Z_j \rangle \oplus p(Z_j, Y)$ is a path satisfying conditions (1)-(3) and shorter than p, which contradicts with condition (4).

2. Either $X \to Z_1 \in \mathcal{G}'$ or $X \leftarrow Z_1 \in \mathcal{G}'$. Otherwise, by Theorem 4.9, there exists $2 \leq i \leq n + 1$ such that $X \leftarrow Z_i \in \mathcal{G}'$. Since p is a possibly causal path in $\mathcal{G}^*$, we have $X \leftarrow Z_i \notin \mathcal{G}^*$, so $X - Z_i \in \mathcal{G}^*$. Then $\langle X, Z_i \rangle \oplus p(Z_i, Y)$ is a possibly causal path beginning with $X - Z_i$ in $\mathcal{G}^*$, so $Z_i \in PM(X, Y, \mathcal{G}^*)$. But $X \leftarrow Z_i \in \mathcal{G}'$ implies $Z_i \notin PM(X, Y, \mathcal{G}')$, so $Z_i \in PM(X, Y, \mathcal{G}^*) \setminus PM(X, Y, \mathcal{G}')$. The case $Z_i = Y$ has been discussed above, and the case $Z_i \neq Y$ contradicts with condition (3) while constructing the path p.

We first assume that $X \to Z_1 \in \mathcal{G}'$. Let $I_Z \coloneqq \{i = 2, 3, \ldots, n + 1 : Z_i \in adj(X, \mathcal{G}^*)\}$. Since p is possibly causal in $\mathcal{G}^*$, there does not exist $Z_1 \leftarrow Z_i$ for any $i = 2, 3, \ldots, n + 1$. By the local orientation rule in MPDAG (Theorem 3.2), there exists an DAG $\mathcal{G}_2 \in [\mathcal{G}^*]$ such that $X \leftarrow Z_1 \in \mathcal{G}_2$ and $X \to Z_i \in \mathcal{G}_2$ for every $i \in I_Z$.

If $I_Z = \emptyset$, by Meek's Rule 1, for every $\mathcal{G} \in [\mathcal{G}']$, p is a causal path in $\mathcal{G}$. Let f be a distribution generated by a linear gaussian model with $\mathcal{G}$ as the true DAG, where only edges on p have strength 1:

$$Z_i = Z_{i-1} + \varepsilon_{Z_i}, \quad i = 1, 2, \ldots, n+1.$$

Then f is consistent with every $\mathcal{G} \in [\mathcal{G}']$, and also consistent with $\mathcal{G}_2$. However, the causal effect of X on Y is 1 in every $\mathcal{G} \in [\mathcal{G}']$, while the causal effect of X on Y is 0 in $\mathcal{G}_2$. So the statement of the theorem holds. Therefore, in the following proof we assume $I_Z \neq \emptyset$.

Let $m = \max\{i : i \in I_Z\}$. We first show that $Z_n \to Y \in \mathcal{G}_2$. Otherwise if $Z_n \leftarrow Y \in \mathcal{G}_2$, since $p(Z_1, Y)$ is chordless and has no v-structure, $q \coloneqq \langle Y, Z_n, Z_{n-1}, \ldots, Z_1 \rangle$ is a causal path in $\mathcal{G}_2$ by Meek's Rule 1. And since $X \leftarrow Z_1 \in \mathcal{G}_2$, $q \oplus \langle Z_1, X \rangle$ is a causal path in $\mathcal{G}_2$. However, in this way, $\langle X, Z_m \rangle \oplus q(Z_m, Z_1) \oplus \langle Z_1, X \rangle$ constructs an directed cycle in $\mathcal{G}_2$, leading to a contradiction. Therefore, we have $Z_n \to Y \in \mathcal{G}_2$.

Let $t = \min\{1 \leq i \leq n : Z_i \to Z_{i+1} \in \mathcal{G}_2\}$. Since $Z_n \to Y \in \mathcal{G}_2$, such t exists. Since $p(Z_1, Y)$ is chordless and has no v-structure, $p(Z_t, Y)$ is a causal path in $\mathcal{G}_2$. Therefore, by the selection of t, the path p is of the form $X \leftarrow Z_1 \leftarrow \cdots \leftarrow Z_t \to Z_{t+1} \to \cdots \to Y$ in $\mathcal{G}_2$. Moreover, we have $t < i_Z$ for every $i_Z \in I_Z$, otherwise the path $\langle X, Z_{i_Z}, Z_{i_Z - 1}, \ldots, Z_1, X \rangle$ is a directed cycle in $\mathcal{G}_2$.

Let $\mathbf{V}' \coloneqq \{X, Z_1, \ldots, Z_n, Y\}$. Let f be a distribution over $\mathbf{V}$, where variables in $\mathbf{V} \setminus \mathbf{V}'$ are independent standard normal and are independent with $\mathbf{V}'$, and the variables in $\mathbf{V}'$ are generated with respect to the following linear gaussian model:

$$Z_i = Z_{i+1} + \varepsilon_{Z_i}, \quad i = t-1, t-2, \ldots, 0; \tag{2}$$

$$Z_j = Z_{j-1} + \mathbb{1}\{j \in I_Z\} X + \varepsilon_{Z_j}, \quad j = t+1, t+2, \ldots, n, n+1. \tag{3}$$

Moreover, let $\mathrm{Var}(\varepsilon_X) = \mathrm{Var}(Z_t)$, and for $i = 1, 2, \ldots, t-1, t+1, \ldots, n$, let $\mathrm{Var}(\varepsilon_{Z_i}) = \delta > 0$ be much smaller than $\mathrm{Var}(Z_t)$. All noise terms ε_i are independent.

By the construction of f, we know that f is consistent to $\mathcal{G}_2(\mathbf{V}')$, the induced subgraph of $\mathcal{G}_2$ over $\mathbf{V}'$, and is also consistent to $\mathcal{G}_2$. Since there are total $|I_Z|$ causal paths from X to Y in $\mathcal{G}_2(\mathbf{V}')$, and the strength of each causal path is 1, the causal effect of X on Y under distribution f with underlying DAG $\mathcal{G}_2$ is $|I_Z|$.

For every $\mathcal{G} \in [\mathcal{G}']$, let $\tau_{\mathcal{G}}$ denote the causal effect of X on Y under distribution f with underlying DAG $\mathcal{G}$. It suffices to show that $\tau_{\mathcal{G}} \neq |I_Z|$ for every $\mathcal{G} \in [\mathcal{G}']$. Let $\mathbf{S}(\mathcal{G}) = pa(X, \mathcal{G}) \cap \mathbf{V}'$.

If $\mathbf{S}(\mathcal{G}) = \emptyset$, we have

$$\tau_{\mathcal{G}} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}. \tag{4}$$

From the construction of f, i.e., (2) and (3), we have

$$Y = Z_t + |I_Z| X + \varepsilon_{t+1:n} + \varepsilon_Y, \quad X = Z_t + \varepsilon_{1:t-1} + \varepsilon_X, \tag{5}$$

where $\varepsilon_{t+1:n} = \sum_{i=t+1}^{n} \alpha_i \varepsilon_{Z_i}$, $\varepsilon_{1:t-1} = \sum_{i=1}^{t-1} \alpha_i \varepsilon_{Z_i}$ for some positive integers α_i, $i = 1, 2, \ldots, t-1, t+1, \ldots, n$. Then we have

$$\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{Cov}(X, Z_t + |I_Z| X + \varepsilon_{t+1:n} + \varepsilon_Y) && (6)\\
&= |I_Z| \mathrm{Var}(X) + \mathrm{Cov}(Z_t + \varepsilon_{1:t-1} + \varepsilon_X, Z_t + \varepsilon_{t+1:n} + \varepsilon_Y) && (7)\\
&= |I_Z| \mathrm{Var}(X) + \mathrm{Var}(Z_t). && (8)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\tau_{\mathcal{G}} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)} &= \frac{|I_Z| \mathrm{Var}(X) + \mathrm{Var}(Z_t)}{\mathrm{Var}(X)} = |I_Z| + \frac{\mathrm{Var}(Z_t)}{\mathrm{Var}(X)} && (9)\\
&= |I_Z| + \frac{\mathrm{Var}(Z_t)}{\mathrm{Var}(Z_t) + \mathrm{Var}(\varepsilon_{1:t-1}) + \mathrm{Var}(\varepsilon_X)} > |I_Z|, && (10)
\end{aligned}$$

since $\mathrm{Var}(Z_t) = \mathrm{Var}(\varepsilon_X)$ and $\mathrm{Var}(\varepsilon_{1:t-1}) \geq 0$. Therefore, $\tau_{\mathcal{G}} \neq |I_Z|$.

If $\mathbf{S}(\mathcal{G}) \neq \emptyset$, recall that $X \to Z_1 \in \mathcal{G}'$ so $X \to Z_1 \in \mathcal{G}$, so $\mathbf{S}(\mathcal{G}) \subseteq \{Z_i : i \in I_Z\}$. Let $k = \max\{i : Z_i \in \mathbf{S}(\mathcal{G})\}$. Since $k \in I_Z$ and $I_Z \subseteq \{t+1, \ldots, n+1\}$ by the acyclic argument, we have $k \geq t+1$; hence $|I_Z \cap \{t+1, \ldots, k\}| \geq 1$. Iterating (3) from $j = t+1$ up to $j = k$ gives

$Z_k = Z_t + |I_Z \cap \{t+1, \ldots, k\}|\, X + \sum_{i=t+1}^{k} \varepsilon_{Z_i}$; together with $Y = Z_t + |I_Z|\, X + \varepsilon_{t+1:n} + \varepsilon_Y$ from (2)(3), we obtain

$$Y = Z_k + (|I_Z| - |I_Z \cap \{t+1, \ldots, k\}|)\, X + \varepsilon_{k+1:n} + \varepsilon_Y, \tag{11}$$

where $\varepsilon_{k+1:n} = \sum_{i=k+1}^{n} \beta_i \varepsilon_{Z_i}$ with positive integers β_i for $i = k+1, \ldots, n$. Since f is consistent to $\mathcal{G}_2$, and $\mathcal{G}_2$ and $\mathcal{G}$ are Markov equivalent (both lie in $[\mathcal{G}^*]$), f is also consistent to $\mathcal{G}$. As f is linear Gaussian, $\tau_{\mathcal{G}}$ is the coefficient of X in $Y \sim X, pa(X, \mathcal{G})$. Every node in $pa(X, \mathcal{G}) \setminus \mathbf{V}'$ is independent of $\mathbf{V}'$ by construction, hence of X and Y; such nodes do not affect this coefficient, so $\tau_{\mathcal{G}}$ is the coefficient of X in $Y \sim X, \mathbf{S}(\mathcal{G})$ with $\mathbf{S}(\mathcal{G}) = pa(X, \mathcal{G}) \cap \mathbf{V}'$.

If $Z_k = Y$, then $Y \in \mathbf{S}(\mathcal{G})$, so $Y \to X \in \mathcal{G}$; since a directed path from X to Y would then form a cycle, $\tau_{\mathcal{G}} = 0 \neq |I_Z|$ (as $I_Z \neq \emptyset$), and we are done. Hence assume $k \leq n$, so $Z_k \neq Y$. For every $Z_i \in \mathbf{S}(\mathcal{G}) \setminus \{Z_k\}$, iterating (3) from $j = t+1$ to $j = i$ gives $Z_i = Z_t + |I_Z \cap \{t+1, \ldots, i\}|\, X + \sum_{j=t+1}^{i} \varepsilon_{Z_j}$, so X and every $Z_i \in \mathbf{S}(\mathcal{G})$ are functions of $\{Z_t, \varepsilon_X, \varepsilon_{Z_j} : 1 \leq j \leq k\}$. By (11), $Y = Z_k + \big(|I_Z| - |I_Z \cap \{t+1, \ldots, k\}|\big)X + \varepsilon_{k+1:n} + \varepsilon_Y$. The residual $\varepsilon_{k+1:n} + \varepsilon_Y$ depends only on $\{\varepsilon_{Z_j} : k < j \leq n+1\}$; since Z_t is exogenous (independent of all noise terms, cf. (2)) and all noise terms are independent, this residual is independent of X, of Z_k, and of every $Z_i \in \mathbf{S}(\mathcal{G}) \setminus \{Z_k\}$. Therefore $\mathbb{E}[Y \mid X, \mathbf{S}(\mathcal{G})] = Z_k + \big(|I_Z| - |I_Z \cap \{t+1, \ldots, k\}|\big)X$, so the coefficient of X in $Y \sim X, \mathbf{S}(\mathcal{G})$ is $|I_Z| - |I_Z \cap \{t+1, \ldots, k\}|$. Since $k \in I_Z$ and $k \geq t+1$, we have $|I_Z \cap \{t+1, \ldots, k\}| \geq 1$, and therefore $\tau_{\mathcal{G}} = |I_Z| - |I_Z \cap \{t+1, \ldots, k\}| \neq |I_Z|$.

We have completed the proof assuming $X \to Z_1 \in \mathcal{G}'$, so it suffices to show that the statement holds when $X \leftarrow Z_1 \in \mathcal{G}'$. Let $\mathcal{G}''$ be the MPDAG obtained by adding background knowledge $X \to Z_1$ in $\mathcal{G}^*$.

We first show that there exists an DAG $\mathcal{G}_3 \in [\mathcal{G}''] \subseteq [\mathcal{G}^*] \setminus [\mathcal{G}']$ such that p is a causal path in $\mathcal{G}_3$. If X and Z_2 are not adjacent, we have $Z_1 \to Y \in \mathcal{G}''$ by Meek's Rule 1, and $p(Z_1, Y)$ is causal in $\mathcal{G}''$ by Meek's Rule 1 since $p(Z_1, Y)$ is chordless. So p is causal in $\mathcal{G}''$, also in $\mathcal{G}_3 \in [\mathcal{G}'']$. If X and Z_2 are adjacent, we have $X \to Z_2 \in \mathcal{G}^*$ or $X - Z_2 \in \mathcal{G}^*$ since p is possibly causal. For both cases, by applying Theorem 3.2 on Z_1, adding background knowledge $X \to Z_1$ and $Z_1 \to Z_2$ to $\mathcal{G}^*$ is valid. Therefore, there exists $\mathcal{G}_3 \in [\mathcal{G}'']$ such that $Z_1 \to Z_2 \in \mathcal{G}_3$. By Meek's Rule 1, $p(Z_1, Y)$ is a causal path in $\mathcal{G}_3$, so p is a causal path in $\mathcal{G}_3$.

Let f be a distribution over $\mathbf{V}$, where variables in $\mathbf{V} \setminus \mathbf{V}'$ are independent standard normal variables and are independent with $\mathbf{V}'$, and the variables in $\mathbf{V}'$ are generated with respect to the following linear gaussian model:

$$Z_i = Z_{i-1} + \varepsilon_{Z_i}, \quad i = 1, 2, \ldots, n+1. \tag{12}$$

Then f is consistent with $\mathcal{G}_3$, and consistent with every DAG $\mathcal{G} \in [\mathcal{G}']$. By the construction of f, the causal effect of X on Y under f with underlying DAG $\mathcal{G}_3$ is 1. However, for every $\mathcal{G} \in [\mathcal{G}']$, let $\mathbf{S}(\mathcal{G}) = pa(X, \mathcal{G}) \cap \mathbf{V}'$, the causal effect of X on Y under f with underlying DAG is the coefficient of X for regression $Y \sim X, \mathbf{S}(\mathcal{G})$, which is zero since $Z_1 \in \mathbf{S}(\mathcal{G})$ and $Y \perp X \mid Z_1$. Therefore, the statement of the theorem holds. ■

A.8 Proof of Theorem 4.8

Proof. By definition, there is a b-possibly causal path p from X to Y starting with $X - Z$. Choose p to be a shortest such path. If the segment from Z to Y were not chordless, we could shortcut it to obtain a shorter b-possibly causal path still starting with $X - Z$, contradicting minimality. Hence p is essential. ■

A.9 Proof of Theorem 4.9

To prove Theorem 4.9, we need the following lemma:

Lemma A.5 *Let $\mathcal{G}'$ be the MPDAG obtained by incorporating background knowledge $\mathcal{B}_0 = \{A_r \to B_r : r = 1, \ldots, q\}$ into $\mathcal{G}^*$. For any directed edge $C \to D$ in $\mathcal{G}'$ which is undirected in $\mathcal{G}^*$, there exists $r \in \{1, \ldots, q\}$ such that $B_r \in an(D, \mathcal{G}')$.*

Proof. Since $\mathcal{G}'$ is obtained by applying $\mathcal{B}_0 = \{A_r \to B_r : r = 1, \ldots, q\}$ and running Meek's rules

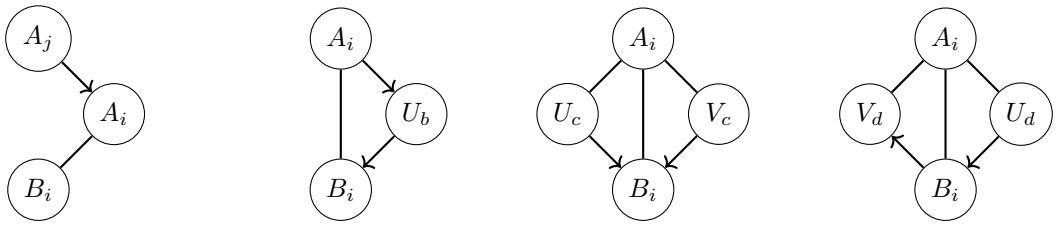


Figure 6: Illustration for the proof of Lemma A.5.

in $\mathcal{G}^*$, every directed edge of $\mathcal{G}'$ that is undirected in $\mathcal{G}^*$ is either one of the initial background edges $A_r \to B_r$ or is added by one of the four Meek's rules. We prove by induction that for every such edge $C \to D$ in $\mathcal{G}'$, there exists $r \in \{1, \dots, q\}$ with $B_r \in an(D, \mathcal{G}')$; the claim is trivial for the initial edges ($B_r \in an(B_r, \mathcal{G}')$). Denote the initial edges by $e_0^{(r)} = A_r \to B_r$ ($r = 1, \dots, q$) and let $e_i = A_i \to B_i$ denote the directed edge added by the i-th Meek-rule application.

For every $i \geq 1$, assume the claim holds for every edge added before the i-th Meek-rule application, and let $e_i = A_i \to B_i$ be the edge added at step i. Since e_i is inferred by exactly one of the four Meek rules from earlier edges (initial or added), exactly one of the following four cases occurs, illustrated by Figure 6; in each we show that a background head lies in $an(B_i, \mathcal{G}')$. We use one observation throughout: every directed edge appearing in the triggering configuration of a Meek rule is already directed before step i (each rule fires only on already-directed edges), so each such edge is either a background edge $A_r \to B_r$ or an added edge e_j with $j < i$; in the former case its head B_r is itself a background head, and in the latter case the induction hypothesis applies to it.

1. **(R1)** Figure 6(a). Meek's rule R1 is applied to the directed edge $e_j = A_j \to A_i$ (so $B_j = A_i$) and the undirected edge $A_i - B_i$, with A_j and B_i non-adjacent, and orients $A_i \to B_i$. The earlier edge e_j has head $A_i \in an(B_i, \mathcal{G}')$. If e_j is a background edge, then A_i is a background head in $an(B_i, \mathcal{G}')$; otherwise, by the induction hypothesis applied to e_j, a background head lies in $an(A_i, \mathcal{G}') \subseteq an(B_i, \mathcal{G}')$.
2. **(R2)** Figure 6(b). Meek's rule R2 is applied to the directed path $A_i \to U_b \to B_i$ (both directed before step i, by R2's premise) with the undirected edge $A_i - B_i$, and orients $A_i \to B_i$. Either directed edge is a valid choice for e_j with $j < i$; we choose $e_j = U_b \to B_i$, whose head is B_i. If e_j is a background edge, B_i is itself a background head; otherwise the induction hypothesis applied to e_j gives a background head in $an(B_i, \mathcal{G}')$.
3. **(R3)** Figure 6(c). Meek's rule R3 is applied to $A_i - U_c \to B_i$ and $A_i - V_c \to B_i$ with the undirected edge $A_i - B_i$ and U_c, V_c non-adjacent, and orients $A_i \to B_i$. Both $U_c \to B_i$ and $V_c \to B_i$ are directed before step i (by R3's premise) and have head B_i; either is a valid choice for e_j with $j < i$. If e_j is a background edge, B_i is itself a background head; otherwise the induction hypothesis applied to e_j gives a background head in $an(B_i, \mathcal{G}')$.
4. **(R4)** Figure 6(d). Meek's rule R4 is applied to the directed chain $U_d \to B_i \to V_d$ (both directed before step i, by R4's premise) with the undirected edges $A_i - U_d$, $A_i - V_d$, $A_i - B_i$ and U_d, V_d non-adjacent, and orients $A_i \to B_i$. Either directed edge is a valid choice for e_j with $j < i$; we choose $e_j = U_d \to B_i$, whose head is B_i. If e_j is a background edge, B_i is itself a background head; otherwise the induction hypothesis applied to e_j gives a background head in $an(B_i, \mathcal{G}')$.

Therefore, by induction, for every edge $C \to D$ of $\mathcal{G}'$ undirected in $\mathcal{G}^*$ there exists $r \in \{1, \dots, q\}$ with $B_r \in an(D, \mathcal{G}')$.

■

Proof. Let $p = \langle X, Z = Z_0, Z_1, \dots, Z_n = Y \rangle$ be an essential-b-possibly causal path in $\mathcal{G}^*$ from X to Y with Z and that p is not an essential-b-possibly causal path in $\mathcal{G}'$. Since the subpath of p from Z to Y is also chordless in $\mathcal{G}'$, one of the three following cases holds, and we will show that in each cases one of the two conditions in Theorem 4.9 holds.

1. There exists $0 \leq i \leq n$ such that $Z_i \to X \in \mathcal{G}'$, so p is not a b-possibly causal path in $\mathcal{G}'$. Then condition (1) holds.

2. There exists $0 < i \leq n$ such that $Z_i \to Z_{i-1} \in \mathcal{G}'$, so p is not a b-possibly causal path in $\mathcal{G}'$. Since the subpath of p from Z_0 to Z_n is chordless, we have $Z_1 \to Z \in \mathcal{G}'$, otherwise it will contradicts with Meek's rule R1 or generate new v-structure, which should originally exists in $\mathcal{G}^*$ and causes p not to be a b-possibly causal path in $\mathcal{G}^*$. Since p is a b-possibly causal path in $\mathcal{G}^*$ and $[\mathcal{G}'] \subseteq [\mathcal{G}^*]$, we have $Z_1 - Z \in \mathcal{G}^*$. Since $\mathcal{G}'$ is obtained by adding background knowledge $\mathbf{S}_1 \to X$ and $X \to \mathbf{S}_2$, and $Z_1 \to Z$ is undirected in $\mathcal{G}^*$, Lemma A.5 implies that some background edge $A_r \to B_r$ has $B_r \in an(Z, \mathcal{G}')$. The background heads are exactly X (from $\mathbf{S}_1 \to X$) and the nodes of $\mathbf{S}_2$ (from $X \to \mathbf{S}_2$). If $B_r = X$, then $X \in an(Z, \mathcal{G}')$ directly. If $B_r \in \mathbf{S}_2$, then $X \to B_r \in \mathcal{G}'$ is itself a background edge, so $X \in an(B_r, \mathcal{G}') \subseteq an(Z, \mathcal{G}')$. In either case $X \in an(Z, \mathcal{G}')$. For every DAG $\mathcal{G} \in [\mathcal{G}']$, if $Z \to X \in \mathcal{G}$, then the directed path $X \to \cdots \to Z$ (directed in $\mathcal{G}'$, hence in $\mathcal{G}$) would form a directed cycle in $\mathcal{G}$; therefore $X \to Z \in \mathcal{G}$ for every $\mathcal{G} \in [\mathcal{G}']$, so $X \to Z \in \mathcal{G}'$ and condition (2) holds.

3. $X \to Z \in \mathcal{G}'$ or $X \leftarrow Z \in \mathcal{G}'$, so Z is no longer in $PM(X, Y, \mathcal{G}')$. Then condition (1) or (2) holds.

Conversely, suppose that one of the two conditions holds. If condition (2) holds, i.e., $X \to Z \in \mathcal{G}'$, then the first edge of p is no longer undirected in $\mathcal{G}'$, so by Definition 4.7 p is not an essential b-possibly causal path in $\mathcal{G}'$. If condition (1) holds, then $Z_i \to X \in \mathcal{G}'$ for some $0 \leq i \leq n$; since X is the first vertex of p and Z_i appears after it, $Z_i \to X$ is a backward edge of p in $\mathcal{G}'$, so p is not a b-possibly causal path in $\mathcal{G}'$ and hence not an essential b-possibly causal path. In either case p is no longer an essential b-possibly causal path in $\mathcal{G}'$, which, together with the necessity proved above, yields the desired equivalence. ■

B Empirical Evaluation on ECOLI70 network

We now examine the behaviour of the proposed local criteria when the CPDAG is estimated from finite samples. We consider the ECOLI70 network, with data generated from the true ECOLI70 distribution. The original DAG $\mathcal{G}$ and its corresponding CPDAG $\mathcal{G}^*$ (the MEC without background knowledge) are shown in Figure 7. It is worth noting that the true distribution of the ECOLI70 network is joint Gaussian, which differs from the discrete ASIA network data considered earlier. For each of the six scenarios described in Table 3, we generate datasets of sample sizes $n \in \{1000, 2000, 3000, 5000\}$. For each dataset, we learn a CPDAG using the MMHC algorithm (Tsamardinos et al., 2006). We then apply the same local decision procedure to the learned CPDAG and record whether it correctly identifies whether the background knowledge improves identifiability. The proportion of correct identifications over 1000 replications is shown in Table 4.

Table 3: Summary of the six cases in the ECOLI70 network. For the background knowledge $\mathcal{B}$ in each case, we denote the resulting MPDAG by $\mathcal{G}'$.

Case	$X \to Y$	Background $\mathcal{B}$	PM_0	PM_1	$\mathcal{E}(X, Y; \mathcal{G}^*)$	$\mathcal{E}(X, Y; \mathcal{G}')$
1	$b1191 \to b1583$	$fixC \to b1191$	$\{fixC\}$	$\emptyset$	$\{-0.010, 0.067\}$	$\{-0.010\}$
2	$b1191 \to b1583$	$b1191 \to fixC$	$\{fixC\}$	$\emptyset$	$\{-0.010, 0.067\}$	$\{-0.010\}$
3	$yedE \to nmpC$	$yedE \to lpdA$	$\{lpdA, pspB\}$	$\{pspB\}$	$\{0.786, 0.617\}$	$\{0.786\}$
4	$yedE \to nmpC$	$lpdA \to yedE$	$\{lpdA, pspB\}$	$\emptyset$	$\{0.786, 0.617\}$	$\{0.617\}$
5	$b1191 \to asnA$	$fixC \to b1191$	$\emptyset$	$\emptyset$	$\{1.151\}$	$\{1.151\}$
6	$b1191 \to asnA$	$b1191 \to fixC$	$\emptyset$	$\emptyset$	$\{1.151\}$	$\{1.151\}$

The first four cases correspond to scenarios where the background knowledge does improve identifiability under the true CPDAG (i.e., the possible mediator set becomes smaller or empty), while the last two cases correspond to scenarios where it does not (the possible mediator set remains empty). Consequently, the table reports the proportion of replications in which the procedure correctly distinguishes these two situations under finite-sample structure learning with MMHC.

Table 4: Proportion of correct identifications over 1000 replications for the ECOLI70 network. Cases 1–6 correspond to the background knowledge scenarios in Table 3.

Case	$X \to Y$	Background $\mathcal{B}$	$n = 1000$	$n = 2000$	$n = 3000$	$n = 5000$
1	$b1191 \to b1583$	$fixC \to b1191$	71.1%	96.9%	99.9%	100.0%
2	$b1191 \to b1583$	$b1191 \to fixC$	71.0%	96.9%	99.9%	100.0%
3	$yedE \to nmpC$	$yedE \to lpdA$	93.0%	97.3%	98.9%	99.8%
4	$yedE \to nmpC$	$lpdA \to yedE$	93.0%	97.3%	98.9%	99.8%
5	$b1191 \to asnA$	$fixC \to b1191$	97.9%	98.0%	100.0%	100.0%
6	$b1191 \to asnA$	$b1191 \to fixC$	98.0%	98.0%	100.0%	100.0%

C Example: Causal Effect of Disease on LowBodyO2 in the Child Network

We illustrate the local criteria developed in Sections 3 and 4 using the well-known Child network (Spiegelhalter & Lauritzen, 1990). The original DAG is shown in Figure 7a and its corresponding CPDAG (the MEC without background knowledge) in Figure 7b. We are interested in the total causal effect of *Disease* (X) on *LowBodyO2* (Y).

C.1 Initial State and Possible Mediators

From the CPDAG $\mathcal{G}^*$ in Figure 7b, the siblings of X are

$$sib(X, \mathcal{G}^*) = \{\text{BirthAsph, DuctFlow, CardMix, LungParen, LungFlow, Sick, LVH, Age}\}.$$

According to Definitions 4.1 and 4.7, a node $Z \in sib(X, \mathcal{G}^*)$ is a *possible mediator* from X to Y precisely when there exists a b-possibly causal path that starts with the undirected edge $X - Z$ and reaches Y. Inspecting the CPDAG:

- DuctFlow $\to$ HypDistrib $\to$ LowBodyO2, hence DuctFlow $\in PM(X, Y, \mathcal{G}^*)$;
- CardMix $\to$ HypDistrib $\to$ LowBodyO2 and CardMix $\to$ HypoxInO2 $\to$ LowBodyO2, so CardMix $\in PM(X, Y, \mathcal{G}^*)$;
- LungParen $\to$ HypoxInO2 $\to$ LowBodyO2, so LungParen $\in PM(X, Y, \mathcal{G}^*)$.

No other sibling can reach Y via a b-possibly causal path without reversing an edge. Therefore

$$PM(X, Y, \mathcal{G}^*) = \{\text{DuctFlow, CardMix, LungParen}\}.$$

Because PM is non-empty, the total causal effect of X on Y is not identifiable in the CPDAG (Perkovic, 2020).

C.2 Stepwise Addition of Local Background Knowledge

We now demonstrate how small pieces of local background knowledge progressively shrink the possible mediator set, eventually rendering the causal effect identifiable. All the operations below respect the purely local criteria of Sections 3 and 4.

First step: $X \to$ DuctFlow

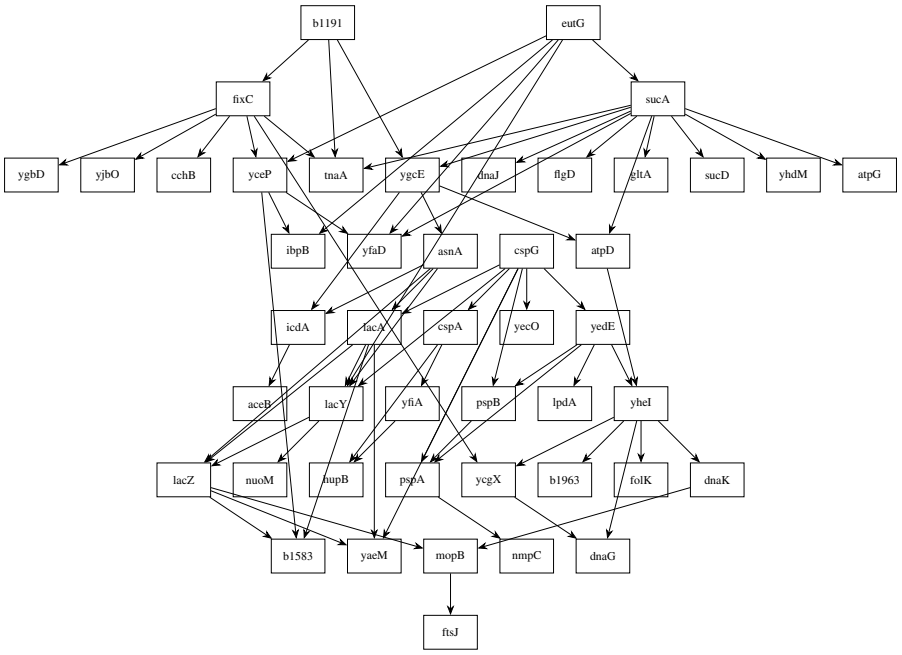
Suppose a clinician asserts that "Disease is a direct cause of abnormal duct flow". This corresponds to background knowledge $\mathcal{B}_1 = \{X \to \text{DuctFlow}\}$, i.e., $\mathbf{S}_1 = \emptyset$, $\mathbf{S}_2 = \{\text{DuctFlow}\}$. By Corollary 3.3, any background knowledge with $\mathbf{S}_1 = \emptyset$ is immediately valid. Theorem 3.4 gives

$$\mathbf{S}_3 = \emptyset, \quad \mathbf{S}_4 = sib(X, \mathcal{G}^*) \cap ch(\text{DuctFlow}, \mathcal{G}^*) = \emptyset, \quad \mathbf{S}_5 = \emptyset,$$

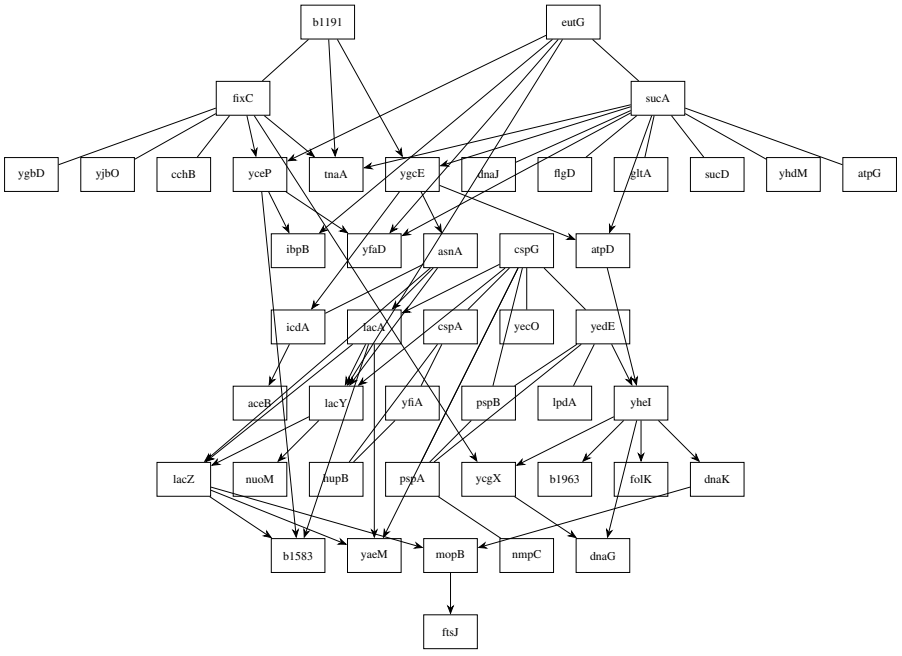
so the new MPDAG $\mathcal{G}_1'$ simply orients $X \to$ DuctFlow; all other undirected edges incident to X stay unchanged. Consequently DuctFlow leaves the possible mediator set, while CardMix and LungParen remain. Thus

$$PM(X, Y, \mathcal{G}_1') = \{\text{CardMix, LungParen}\} \subsetneq PM(X, Y, \mathcal{G}^*).$$

According to Theorems 4.3 and 4.4, the set of possible causal effects is strictly reduced: identifiability improves, even though the effect is not yet point-identified.



(a) Original DAG



(b) CPDAG (Markov equivalence class)

Figure 7: DAG and CPDAG in the ECOLI 70 network.

SECOND STEP: $X \to$ CARDMIX

Next we incorporate the knowledge "Disease directly influences cardiac mixing". Adding $\mathcal{B}_2 = \{X \to \text{CardMix}\}$ to $\mathcal{G}_1'$ (i.e., $\mathbf{S}_1 = \emptyset,\ \mathbf{S}_2 = \{\text{CardMix}\}$) again satisfies Corollary 3.3. Because all siblings of X except the newly oriented ones are still undirected, Theorem 3.4 again yields $\mathbf{S}_3 = \mathbf{S}_4 = \mathbf{S}_5 = \emptyset$. The new MPDAG $\mathcal{G}_2'$ has both $X \to$ DuctFlow and $X \to$ CardMix. Now

$$PM(X, Y, \mathcal{G}_2') = \{\text{LungParen}\}.$$

The possible mediator set has shrunk further; the causal effect is now restricted to a single possible pathway.

THIRD STEP: $X \to$ LUNGPAREN

Finally, the expert adds "Disease causes alterations in lung parenchyma". With $\mathcal{B}_3 = \{X \to \text{LungParen}\}$ added to $\mathcal{G}_2'$, the three former possible mediators all become children of X. Once again the conditions of Corollary 3.3 and Theorem 3.4 are trivially met, and the new MPDAG $\mathcal{G}_3'$ satisfies

$$PM(X, Y, \mathcal{G}_3') = \emptyset.$$

By Lemma A.4, the total causal effect of X on Y is now *identified*.

C.3 DISCUSSION

This example illustrates the three questions posed in the Introduction.

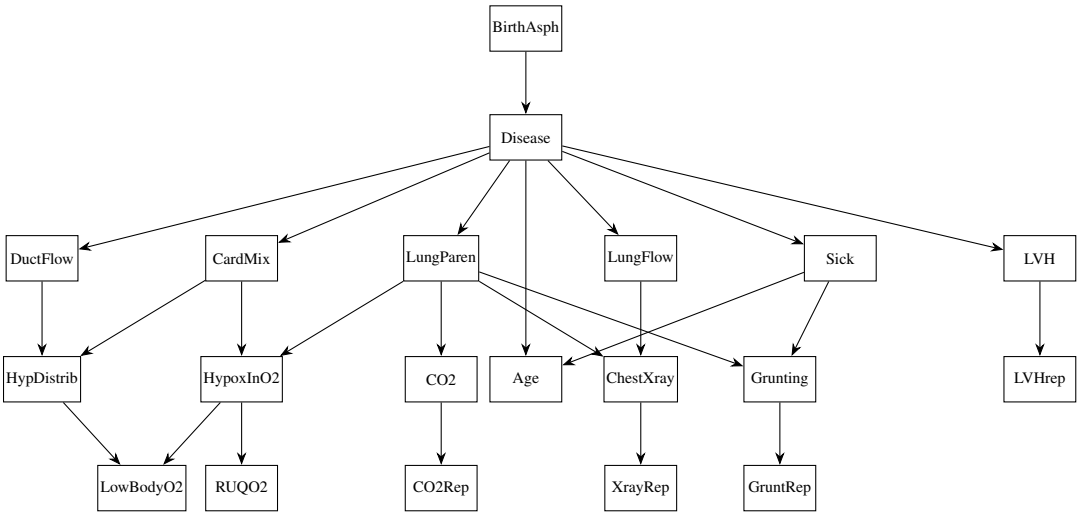
1. *Compatibility*: Each piece of background knowledge ($X \to$ DuctFlow, etc.) was verified to be compatible with the current MPDAG in a single glance using Corollary 3.3 ($\mathbf{S}_1$ is empty, hence trivially a clique).
2. *Local orientation*: Theorem 3.4 confirmed that orienting one sibling of X did not force any extra orientations among the other siblings because the structural conditions for $\mathbf{S}_3$, $\mathbf{S}_4$, and $\mathbf{S}_5$ were absent. Thus the local neighbourhood of X could be updated without global propagation.
3. *Identifiability improvement*: The possible mediator set shrank from $\{\text{DuctFlow}, \text{CardMix}, \text{LungParen}\}$ to $\emptyset$ in three steps, each time strictly reducing the set of possible causal effects. Theorem 4.9 explains this behaviour: an essential b-possibly causal path starting with an undirected edge $X - Z$ is destroyed exactly when Z becomes a child of X (condition (2) of the theorem).

Importantly, the full Markov equivalence class was never enumerated. This demonstrates how our local criteria can be integrated into interactive causal discovery workflows, allowing a domain expert to iteratively supply small pieces of background knowledge and immediately receive feedback on whether the target causal effect becomes identifiable.

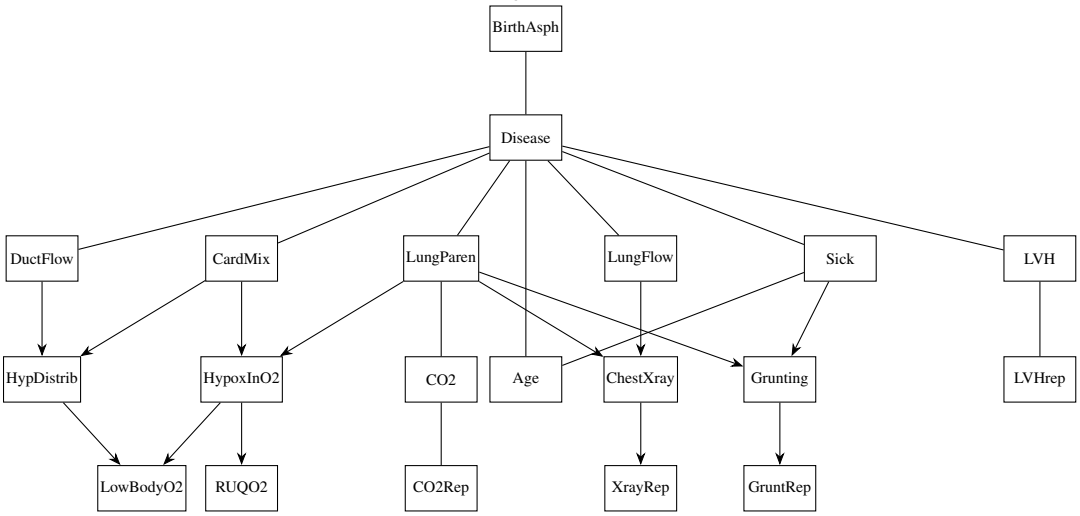
Table 5 summarises, for every sibling Z of X in the initial CPDAG, whether orienting the edge as $Z \to X$ or $X \to Z$ improves identifiability of the causal effect of Disease on LowBodyO2. A ✓ indicates a strict reduction of the possible mediator set, while '—' denotes no change. As can be seen, adding any single parent orientation $Z \to X$ suffices to make the effect point-identifiable, whereas adding a child orientation $X \to Z$ helps only when Z is itself a possible mediator.

Table 5: Does local background knowledge $Z \to X$ or $X \to Z$ improve identifiability of the causal effect of Disease (X) on LowBodyO2 (Y)? The initial CPDAG is Figure 8b; ✓ indicates improvement (smaller or empty possible mediator set), — indicates no change.

Sibling Z	$Z \to X$	$X \to Z$
BirthAsph	✓	—
DuctFlow	✓	✓
CardMix	✓	✓
LungParen	✓	✓
LungFlow	✓	—
Sick	✓	—
LVH	✓	—
Age	✓	—



(a) Original DAG



(b) CPDAG (Markov equivalence class)

Figure 8: DAG and CPDAG in the Child network.

C.4 Complexity Analysis

Let n be the number of nodes and m the number of edges in the MPDAG $\mathcal{G}^*$, $\mathcal{G}'$ be the MPDAG obtained from $\mathcal{G}^*$ by adding the background knowledge $\mathcal{B}$ The algorithm proceeds in three stages, each of which can be performed in polynomial time.

Stage 1: Validity check. By Theorem 3.2, checking the validity of the background knowledge $\mathcal{B} = \{\mathbf{S}_1 \to X,\ X \to \mathbf{S}_2\}$ requires two tests:

1. $\mathbf{S}_1$ is a clique in $\mathcal{G}^*$, i.e., every pair of nodes in $\mathbf{S}_1$ is adjacent. This can be checked by inspecting all $\binom{|\mathbf{S}_1|}{2}$ node pairs in $O(|\mathbf{S}_1|^2)$ time.
2. $\mathbf{S}_1 \cap ch(\mathbf{S}_2, \mathcal{G}^*) = \varnothing$. For each $A \in \mathbf{S}_1$ and each $B \in \mathbf{S}_2$, we test whether the edge $B \to A$ exists in $\mathcal{G}^*$, requiring $O(|\mathbf{S}_1| \cdot |\mathbf{S}_2|)$ time.

Thus the validity check runs in $O(n^2)$ worst-case time.

Stage 2: Computing $\mathbf{S}_3, \mathbf{S}_4, \mathbf{S}_5$. According to Theorem 3.4, the sets $\mathbf{S}_3$, $\mathbf{S}_4$, and $\mathbf{S}_5$ are determined solely by the adjacency relations among X, its siblings, and the nodes in $\mathbf{S}_1 \cup \mathbf{S}_2$:

- $\mathbf{S}_3$ is obtained by checking, for each $Z \in sib(X, \mathcal{G}^*)$, whether Z is adjacent to every node in $\mathbf{S}_1$.
- $\mathbf{S}_4$ is obtained by checking, for each $Z \in sib(X, \mathcal{G}^*)$, whether there exists $A \in \mathbf{S}_2$ such that $A \to Z$ is in $\mathcal{G}^*$.
- $\mathbf{S}_5$ is obtained by checking, for each $Z \in sib(X, \mathcal{G}^*)$, whether there exists $A \in \mathbf{S}_1$ such that $Z \to A$ is in $\mathcal{G}^*$.

Each check takes $O(|sib(X, \mathcal{G}^*)| \cdot (|\mathbf{S}_1| + |\mathbf{S}_2|))$ time, which is $O(n^2)$ in the worst case.

Stage 3:Computing $PM(X, Y, \mathcal{G}^*)$. We first compute the set $PM = PM(X, Y, \mathcal{G}^*)$. By Definition 4.1 and Theorem 4.8, a node $Z \in sib(X, \mathcal{G}^*)$ belongs to $PM(X, Y, \mathcal{G}^*)$ if and only if there exists an *essential b-possibly causal path* from X to Y that starts with the undirected edge $X - Z$. To detect such paths efficiently, we construct an auxiliary graph $\mathcal{H}_Z$ for each $Z \in sib(X, \mathcal{G}^*)$ as follows: start from $\mathcal{G}^*$, delete the $X \cup pa(X, \mathcal{G}^*)$ (and all edges incident to them). For every sibling W of Z in the remaining graph (i.e., every node W adjacent to Z by an undirected edge, excluding X), orient the edge as $Z \to W$. Then we repeatedly apply only Meek's rule R1 (see Appendix A.1) until no further orientation can be inferred. We now prove that this propagation is both sound and complete for detecting essential b-possibly causal paths.

Lemma C.1 *Let $p = \langle X, Z, Z_1, \ldots, Z_k = Y \rangle$ be an essential b-possibly causal path in $\mathcal{G}^*$ starting with the undirected edge $X - Z$. If we force the orientation $Z \to Z_1$ (which is implied by our auxiliary construction because $Z_1 \in sib(Z, \mathcal{G}^*)$), then applying Meek's rule R1 repeatedly will orient every edge along the subpath $Z - Z_1 - \cdots - Z_k$ in the forward direction. Consequently, p becomes a directed path from Z to Y in $\mathcal{H}_Z$.*

Proof. By Definition 4.7 of an essential b-possibly causal path, the subpath $Z - Z_1 - \cdots - Z_k$ is chordless. Hence, for any $i = 1, \ldots, k-1$, the nodes Z_{i-1} and Z_{i+1} are not adjacent (with $Z_0 = Z$). After orienting $Z \to Z_1$, consider the first un-oriented triple $Z_{i-1} \to Z_i - Z_{i+1}$ (where Z_{i+1} is not yet oriented). Since Z_{i-1} and Z_{i+1} are non-adjacent by chordlessness, Meek's rule R1 applies and forces $Z_i \to Z_{i+1}$. By induction on i, we eventually orient all edges up to $Z_{k-1} \to Z_k = Y$, yielding a directed path from Z to Y. ■

Lemma C.2 *If applying Meek's rule R1 (after forcing all edges incident to Z to be outgoing) produces a directed path from Z to Y in $\mathcal{H}_Z$, then there exists an essential b-possibly causal path from X to Y starting with $X - Z$ in $\mathcal{G}^*$.*

Proof. Let q be a directed path from Z to Y in $\mathcal{H}_Z$ that avoids X. Because $\mathcal{H}_Z$ differs from $\mathcal{G}^*$ only in edge orientations by Meek's rule R1, q corresponds to a path in $\mathcal{G}^*$. Since all edges of q are oriented forward, q is a b-possibly causal path or directed path in $\mathcal{G}^*$. The concatenation $X - Z \oplus q$ gives a b-possibly causal path from X to Y starting with $X - Z$. By the same chordless-shortcut argument as in Theorem 4.8, if the subpath from Z to Y is not chordless, we can shortcut it to obtain an essential b-possibly causal path still starting with $X - Z$. Hence $Z \in PM(X, Y, \mathcal{G}^*)$. ■

By Lemmas C.1 and C.2, checking whether $Z \in PM(X, Y, \mathcal{G}^*)$ is equivalent to checking whether a directed path from Z to Y avoiding X exists in $\mathcal{H}_Z$ after R1 propagation. The construction of $\mathcal{H}_Z$ and the repeated application of R1 take polynomial time.

Stage 4:Computing $PM(X, Y, \mathcal{G}^{'})$. By Theorem 4.9, a node $Z \in sib(X, \mathcal{G}^*)$ belongs to $PM(X, Y, \mathcal{G}')$ if and only if

1. $Z \notin \mathbf{S}_1 \cup \mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4 \cup \mathbf{S}_5$, and
2. there exists an essential b-possibly causal path from X to Y starting with the undirected edge $X - Z$ that avoids every node in $pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \cup \mathbf{S}_5$.

Indeed, the nodes in $\mathbf{S}_1 \cup \mathbf{S}_5$ are precisely those that become parents of X in $\mathcal{G}'$, and any path passing through such a node is destroyed. To check this directly, for each $Z \in sib(X, \mathcal{G}^*) \setminus (\mathbf{S}_1 \cup$

Algorithm 2 Compute Initial Possible Mediators

Require: An MPDAG $\mathcal{G}^*$, distinct nodes X, Y with $Y \notin pa(X, \mathcal{G}^*)$.
Ensure: The set $PM = PM(X, Y, \mathcal{G}^*)$.

$PM \leftarrow \emptyset$
for each $Z \in sib(X, \mathcal{G}^*)$ **do**
  Construct an auxiliary graph $\mathcal{H}_Z$ from $\mathcal{G}^*$ as follows:
    Delete the node set $\{X\} \cup pa(X, \mathcal{G}^*)$ and all incident edges.
    For every undirected neighbor W of Z in the remaining graph, orient $Z \to W$.
    Repeatedly apply Meek's rule R1 until no further orientation can be inferred.
  if there exists a directed path from Z to Y in $\mathcal{H}_Z$ **then**
    $PM \leftarrow PM \cup \{Z\}$
  end if
end for
return PM

Algorithm 3 Compute Updated Possible Mediators after Background Knowledge

Require: MPDAG $\mathcal{G}^*$; distinct nodes X, Y with $Y \notin pa(X, \mathcal{G}^*)$; valid background knowledge $\mathcal{B} = \{\mathbf{S}_1 \to X,\ X \to \mathbf{S}_2\}$; node $Z \in PM(X, Y, \mathcal{G}^*)$; sets S_3, S_4, S_5 computed from Theorem 3.4.
Ensure: **true** if $Z \in PM(X, Y, \mathcal{G}')$, **false** otherwise.

$BlockedParents \leftarrow pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \cup S_5$
$Candidates \leftarrow sib(X, \mathcal{G}^*) \setminus (\mathbf{S}_1 \cup \mathbf{S}_2 \cup S_3 \cup S_4 \cup S_5)$
if $Z \notin Candidates$ **then**
  return false
end if
Construct an auxiliary graph $\mathcal{H}_Z$ from $\mathcal{G}^*$ as follows:
  Delete the node set $\{X\} \cup BlockedParents$ and all incident edges.
  For every undirected neighbor W of Z in the remaining graph, orient $Z \to W$.
  Repeatedly apply Meek's rule R1 until no further orientation can be inferred.
if there exists a directed path from Z to Y in $\mathcal{H}_Z$ **then**
  return true
else
  return false
end if

$\mathbf{S}_2 \cup \mathbf{S}_3 \cup \mathbf{S}_4 \cup \mathbf{S}_5$), construct an auxiliary graph $\mathcal{H}_Z$ as follows: start from $\mathcal{G}^*$, delete the node set $\{X\} \cup pa(X, \mathcal{G}^*) \cup \mathbf{S}_1 \cup \mathbf{S}_5$ together with all incident edges, and orient every undirected edge incident to Z (except the one to X, which is already deleted) as outgoing from Z. Then repeatedly apply only Meek's rule R1 until no further orientation can be inferred. By Lemmas C.1 and C.2, applied to the reduced graph, a directed path from Z to Y exists in $\mathcal{H}_Z$ if and only if $Z \in PM(X, Y, \mathcal{G}')$. Since the construction of $\mathcal{H}_Z$ and the repeated application of R1 take polynomial time, the overall complexity remains polynomial in n and m.

Comparison with MIDA. As an alternative, one could apply the MIDA algorithm to the original MPDAG $\mathcal{G}^*$, record the size of the causal effect set, then add the background knowledge, apply Meek's rules to obtain a new MPDAG, and run MIDA again to compare the sizes. However, MIDA has exponential complexity $O(2^{M(G)} \cdot \mathrm{poly}(m))$, where $M(G)$ denotes the number of undirected edges incident to X on a proper possibly causal path from X to Y. In contrast, our local decision procedure runs in polynomial time and avoids both the enumeration of the Markov equivalence class and the repeated application of Meek's rules to learn a new MPDAG. This yields a substantial efficiency improvement, especially when the number of siblings of X is large.