Partial Structure Discovery is Sufficient for No-regret Learning in Causal Bandits

Muhammad Qasim Elahi
Electrical and Computer Engineering
Purdue University
elahi0@purdue.edu

Mahsa Ghasemi
Electrical and Computer Engineering
Purdue University
mahsa@purdue.edu

Murat Kocaoglu
Electrical and Computer Engineering
Purdue University
mkocaoglu@purdue.edu

Abstract

Causal knowledge about the relationships among the decision variables and a reward variable in a bandit setting can accelerate the learning of an optimal decision, as substantiated by studies on causal bandits. Current works often assume the causal graph is known; nevertheless, such knowledge may not be available a priori. Motivated by this challenge, we focus on the causal bandit problem in scenarios where the underlying causal graph is unknown and further, it may include latent confounders. While intervention on the parents of the reward node is optimal in the absence of latent confounders, that is not necessarily the case in general. Instead, one has to consider a set of possibly optimal arms/interventions, each being a special subset of ancestors of the reward node— deeming causal discovery beyond parents of the reward node imperative. However, for regret minimization, we identify that discovering the full causal structure is not necessary; yet there is no work that provides the necessary and sufficient components of the causal graph. We formally characterize the set of necessary and sufficient latent confounders one needs to detect/learn to ensure all the possibly optimal arms are learned correctly. We also propose a randomized algorithm for learning the causal graph with a limited number of samples. For the causal bandit setup, we propose a two-stage approach. In the first stage, we learn the induced subgraph on ancestors of the reward along with a necessary and sufficient subset of latent confounders to construct the set of possibly optimal arms. The next phase involves the application of a standard bandit algorithm, such as the UCB algorithm. We also establish regret bound for our two-phase approach, which is sublinear in the number of rounds and has polynomial scaling with respect to the number of nodes in the graph.

1 Introduction

Causal bandits have been a topic of interest since their inception and have been studied in various contexts [1]. The authors assumed precise knowledge of the causal graph and the impact of interventions or actions on the parents of the reward node. Subsequently, there has been a flurry of research on causal bandits [2, 3, 4]. The primary limitation of the majority of existing works on causal bandits is their assumption of full knowledge of the causal graph, which is often impractical for many real-world applications [1, 5, 6]. Recently, efforts have been made to overcome this limitation. In [7], the authors propose a sample efficient algorithm for cases where the causal graph can be represented

as a directed tree or a causal forest and later extend the algorithm to encompass a broader class of general chordal graphs. However, the proposed algorithm is only applicable to scenarios where the Markov equivalence class (MEC) of the causal graph is known and does not have confounders. The paper by Konobeev et al. [8] also deals with causal bandits with an unknown graph and proposes a two-phase approach. The first phase uses a randomized parent search algorithm to learn the parents of the reward node, and the second phase employs UCB to identify the optimal intervention over the parents of the reward node. However, similar to [7], they assume causal sufficiency, i.e., no latent confounders are present.In another related paper, [9], the authors initially emphasize the challenge of dealing with exponentially many arms when addressing causal bandits with an unknown graph. To tackle this issue, the authors assume that the reward is a noisy additive function of its parents. This assumption enables them to reframe the problem as an additive combinatorial linear bandit problem.

We also focus on the causal bandit setup where the causal graph is unknown, but we allow the presence of latent confounders and make no parametric assumptions. The optimal intervention in this case is not limited to parents of the reward node; instead, we have a candidate set of optimal interventions, called possibly optimal minimum intervention sets (POMISs), each being a special subset of the ancestors of the reward node [5]. Thus, learning only the parents of the reward, similar to [8], is insufficient. This implies that causal discovery beyond parents of the reward is imperative. However, for regret minimization, discovering the full causal structure is not necessary. Instead, we characterize the set of necessary and sufficient latent confounders one needs to detect/learn to ensure all the possibly optimal arms are learned correctly.

Causal discovery is a well-studied problem and can be applied to our setup [10, 11, 12]. However, the majority of the existing causal discovery algorithms rely on the availability of an infinite amount of interventional data [13, 14, 15]. Some prior work shows that discovery is possible with limited interventional data, with theoretical guarantees when the underlying causal graph is a tree and contains no latent confounders [16]. Also, the paper [17] proposes a sample-efficient active learning algorithm for causal graphs without latent confounders, given that the MEC for the underlying causal graph is known. Sample-efficient learning of causal graphs with latent confounders, without any parametric or graphical assumptions, with theoretical guarantees, remains an open problem.

We propose a randomized algorithm for sample-efficient learning of causal graphs with confounders. We analyze the algorithm and bound the maximum number of interventional samples required to learn the causal graph with all the confounders with a given confidence level. For the causal bandit setup, we propose a two-stage approach where the first step learns a subgraph of the underlying causal graph to construct a set of POMISs, and the second phase learns the optimal arm among the POMISs. We show that the requirement of learning only a subgraph leads to significant savings in terms of interventional samples and consequently, regret. The main contributions of our work are as follows:

- We characterize the necessary and sufficient set of latent confounders in the induced subgraph on ancestors of the reward node that we need to learn/detect in order to identify all the POMISs for a causal bandit setup when the underlying causal graph is unknown.

- We propose a two-phase algorithm for causal bandits with unknown causal graphs containing confounders. The first phase involves learning the induced subgraph on reward's ancestors along with a subset of latent confounders to identify all the POMISs. The next phase involves a standard bandit algorithm, e.g., upper confidence bound (UCB) algorithm. Our theoretical analysis establishes an upper bound on the cumulative regret of the overall algorithm.

2 Preliminaries and Problem Setup

We start with an overview of the causal bandit problem and other relevant background needed on causal models. Structural causal model (SCM) is a tuple $\mathcal{M} = \langle \mathbf{V}, \mathbf{U}, \mathbf{F}, P(\mathbf{U}) \rangle$ where $\mathbf{V} = \{V_i\}_{i=1}^n \cup \{Y\}$ is the set of observed variables, $\mathbf{U}$ is the set of independent exogenous variables, $\mathbf{F}$ is the set of deterministic structural equations and $P(\mathbf{U})$ is the distribution for exogenous variables [18]. The equations f_i map the parents ($\mathsf{Pa}(V_i)$) and a subset of exogenous variables $\mathbf{U}_i \subseteq \mathbf{U}$, to the value of variable V_i, i.e., $V_i = f_i(\mathsf{Pa}(V_i), \mathbf{U}_i)$. We consider the causal bandit setup where all the observed variables $V_i \in \mathbf{V}$ are discrete with the domain $\Omega(V_i) = [K] := \{1, 2, 3, \ldots, K\}$, and the reward Y is binary, i.e., $\Omega(Y) = \{0, 1\}$. We can associate a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ with every SCM, where the vertices $\mathbf{V}$ correspond to the observed variables and edges $\mathbf{E}$ consist of directed edges $V_i \rightarrow V_j$ when $V_i \in \mathsf{Pa}(V_j)$ and bi-directed edges between V_i and V_j ($V_i \leftrightarrow V_j$) when they share some common

unobserved variable, also called latent confounder. We restrict ourselves to semi-Markovian causal models in which every unobserved variable has no parents and has exactly two children, both of which are observed [9]. An intervention on a set of variables $\mathbf{W} \subseteq \mathbf{V}$, denoted by $do(\mathbf{W})$, induces a post-interventional DAG ($\mathcal{G}_{\overline{\mathbf{W}}}$) with incoming edges to vertices $\mathbf{W}$ removed . We can broadly classify interventions into hard interventions, where variables are set to a fixed realization denoted by $do(\mathbf{W} = \mathbf{w})$, and stochastic interventions, where instead of a fixed realization we have $\mathbf{W} \sim \mathbb{P}(.)$, where $\mathbb{P}$ is a probability measure over the domain $\Omega(\mathbf{W})$. We denote the sub-model induced under hard intervention by $\mathcal{M}_{\mathbf{W}=\mathbf{w}}$ and the one induced under stochastic intervention by $\mathcal{M}_{\mathbf{W}}$. In the context of causal bandits, an arm or action corresponds to hard intervention on a subset of variables other than the reward. The goal of the agent is to identify the intervention that maximizes the expected reward. The performance of an agent is measured in terms of cumulative regret R_T.

$$R_T := T \max_{\mathbf{W} \subseteq \mathbf{V}} \max_{\mathbf{w} \in [K]^{|\mathbf{W}|}} \mathbb{E}[Y|do(\mathbf{W} = \mathbf{w})] - \sum_{t=1}^{T} \mathbb{E}[Y|do(\mathbf{W}_t = \mathbf{w}_t)], \tag{1}$$

where $do(\mathbf{W}_t = \mathbf{w}_t)$ represents the intervention selected by the agent in round t. We use the notation $\Delta_{do(\mathbf{w})}$ to define the sub-optimality gap of the corresponding arm $do(\mathbf{W} = \mathbf{w})$. We denote the descendants, ancestors and children of a vertex V_i by $\mathsf{De}(V_i)$, $\mathsf{An}(V_i)$ and $\mathsf{Ch}(V_i)$ respectively. We use the notation $\mathsf{Bi}(V_i, \mathcal{G})$ to denote the set of vertices having bidirected edges to V_i except the reward node Y. We refer to the induced graph between observable variables as the observable graph. The **transitive closure** of a graph, denoted by $\mathcal{G}^{tc}$, encodes the ancestral relationship in $\mathcal{G}$. That is, the directed edge $V_i \rightarrow V_j$ is included in $\mathcal{G}^{tc}$ only when $V_i \in \mathsf{An}(V_j)$. The **transitive reduction**, denoted by $Tr(\mathcal{G}) = (\mathbf{V}, \mathbf{E}^r)$, is a graph with the minimum number of edges such that the transitive closure is the same as $\mathcal{G}$. The connected component (c-component) of the DAG $\mathcal{G}$, containing vertex V_i, is denoted by $\mathsf{CC}(V_i)$, which is the maximal set of all vertices in $\mathcal{G}$ that have a path to V_i, consisting only of bi-directed edges [19]. For a subset of vertices $\mathbf{W} \subseteq \mathbf{V}$, we define $\mathsf{CC}(\mathbf{W}) := \bigcup_{W_i \in \mathbf{W}} \mathsf{CC}(W_i)$. In a DAG, a subset of nodes $\mathbf{W}$ d-separates two nodes V_i and V_j when it effectively blocks all paths between them, denoted as $V_i \perp\!\!\!\perp_d V_j$. Blocking is a graphical criterion associated with d-separation [18]. A probability distribution is said to be faithful to a graph if and only if every conditional independence (CI) statement can be inferred from d-separation statements in the graph. Faithfulness is a commonly used assumption in the existing work on causal discovery [13, 20]. We assume that the following form of the interventional faithfulness assumption holds in our setup.

Assumption 2.1. *Consider two disjoint subsets of nodes $\mathbf{W} \subseteq \mathbf{V}$ and $V_{\mathbf{W}} = \mathbf{V} \setminus \mathbf{W}$. Consider a hard intervention $do(\mathbf{W} = \mathbf{w})$ and the corresponding induced sub-model $\mathcal{M}_{\mathbf{W}=\mathbf{w}}$. We assume that for any stochastic intervention $do(\mathbf{U})$ over any $\mathbf{U} \subseteq V_{\mathbf{W}}$, the CI statement $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y}|\mathbf{Z})_{\mathcal{M}_{\mathbf{W}=\mathbf{w},\mathbf{U}}}$ holds if and only if there is a corresponding d-separation statement in the post-interventional causal graph, i.e., $(\mathbf{X} \perp\!\!\!\perp_d \mathbf{Y}|\mathbf{Z})_{\mathcal{G}_{\overline{\mathbf{W}},\mathbf{U}}}$ where $\mathbf{X}, \mathbf{Y}$, and $\mathbf{Z}$ are disjoint subsets of $V_{\mathbf{W}}$. The CI statements are with respect to the post-interventional joint probability distribution.*

The main consequence of faithfulness assumptions is that any independence relations observed in the data are due to the underlying structure of the causal graph that generated the data, rather than arising from random coincidence. This significantly narrows down the scope of possible causal graphs that could have produced the observational or interventional data.

3 Possibly Optimal Arms in Causal Bandits with Unknown Causal Graph

The optimal intervention in a causal bandit setup is not restricted to the parent set of the reward node when the reward node Y is confounded with any node in its ancestors $\mathsf{An}(Y)$ [5]. For instance, consider SCM $X_1 = U_1$ and $X_2 = X_1 \oplus U_2$ and reward $Y = X_2 \oplus U_2$, where $U_1 \sim Ber(0.5)$ and $U_2 \sim Ber(0.5)$. Note that X_2 and reward Y are confounded in this SCM. The optimal intervention in this case is $do(X_1 = 1)$ since $\mathbb{E}[Y|do(X_1 = 1)] = 1$. The intervention on the parent of the reward $(\mathsf{Pa}(Y) = X_2)$ is suboptimal because $\mathbb{E}[Y|do(X_2 = 0)] = \mathbb{E}[Y|do(X_2 = 1)] = 0.5$. The example shows that it is possible to construct SCMs where optimal intervention is on ancestors of the reward node instead of parents when reward node is confounded with one of its ancestors. We revisit some definitions and results from their work.

Definition 3.1. *(Unobserved Confounder (UC)-Territory[5]) Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ with a reward node Y and let $\mathcal{H}$ be $\mathcal{G}[\mathsf{An}(Y)]$. A set of variables $\mathbf{T} \subseteq V(\mathcal{H})$ containing Y is called an **UC-territory** on $\mathcal{G}$ with respect to Y if $De_{\mathcal{H}}(\mathbf{T}) = \mathbf{T}$ and $\mathsf{CC}_{\mathcal{H}}(\mathbf{T}) = \mathbf{T}$.*

(a) $\mathcal{G}$ (b) $\mathcal{G}_1$ (c) $\mathcal{G}_2$ (d) $\mathcal{G}_3$
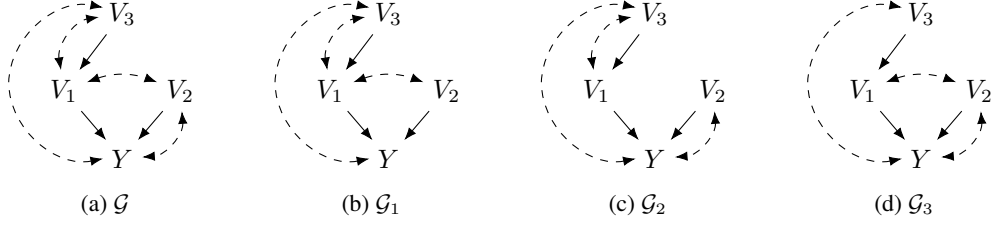
Figure 1: True Causal Graph $\mathcal{G}$ with four other graphs each with one missing bi-directed edge.

A UC-territory is minimal if none of its subsets are UC-territories. A minimal UC-territory denoted by MUCT$(\mathcal{G}, Y)$, can be constructed by extending a set of variables, starting from the reward $\{Y\}$, alternatively updating the set with the c-component and descendants of the set until there is no change.

Definition 3.2. *(Interventional Border)[5] Let T be a minimal UC-territory on $\mathcal{G}$ with respect to Y. Then, $X = \mathsf{Pa}(T) \setminus T$ is called an interventional border for $\mathcal{G}$ w.r.t. Y denoted by $IB(\mathcal{G}, Y)$.*

Lemma 3.1. *[5] For causal graph $\mathcal{G}$ with reward Y, $IB(\mathcal{G}_{\overline{\boldsymbol{W}}}, Y)$ is a POMIS, for any $\boldsymbol{W} \subseteq \boldsymbol{V} \setminus \{Y\}$.*

Although the graphical characterization in Lemma 3.1 provides a means to enumerate the complete set of POMISs, it comes with exponential time complexity. The authors also propose an efficient algorithm for enumerating all POMISs in [5]. However, this requires knowing the true causal graph, and without it, one has to consider interventions on all possible subsets of nodes, which are exponentially many. One naive approach to tackle the problem is to learn the full causal graph with all confounders to list all POMISs. However, a question arises: *Do we need to learn/detect all possible confounders since the goal is to find POMISs and not the full graph?*

Before answering the above question, we start with an example considering the causal graphs in Figure 1. Using Lemma 3.1, the set of POMISs for the true graph $\mathcal{G}$ is $\mathcal{I}_{\mathcal{G}} = \{\phi, \{V_1\}, \{V_2\}, \{V_3\}, \{V_1, V_2\}\}$. However, for $\mathcal{G}_1$ which has the bidirected edge $V_2 \leftrightarrow Y$ missing, the set of POMISs is $\mathcal{I}_{\mathcal{G}_1} = \{\phi, \{V_2\}, \{V_1, V_2\}\}$. Also for $\mathcal{G}_2$ which has the bidirected edge $V_1 \leftrightarrow V_2$ missing, the set of POMISs is $\mathcal{I}_{\mathcal{G}_2} = \{\phi, \{V_1\}, \{V_2\}, \{V_1, V_2\}\}$. In both cases, we miss at least one POMIS, and since it is possible to construct an SCM compatible with the true causal graph $\mathcal{G}$ where any arm in POMIS is optimal, if this arm is not learned, we can suffer linear regret [5]. Although the graph $\mathcal{G}_3$ has the bidirected edge $V_1 \leftrightarrow V_3$ missing, it still has the same set of POMISs as the true graph, i.e., $\mathcal{I}_{\mathcal{G}_3} = \{\phi, \{V_1\}, \{V_2\}, \{V_3\}, \{V_1, V_2\}\}$. This example shows that only a subset of latent confounders affect the POMISs learned from the graph. Theorem 3.1 characterizes all the latent variables that need to be learned, ensuring that the POMISs learned from a sparser causal graph match all those in the true causal graph.

Theorem 3.1. *Consider a causal graph $\mathcal{G}(\boldsymbol{V}, \boldsymbol{E})$ and another causal graph $\mathcal{G}'$ such that they have the same vertex set and directed edges but differ in bidirected edges, with the bidirected edges in $\mathcal{G}'$ being a subset of the bidirected edges in $\mathcal{G}$. The graphs will yield different collections of POMISs if and only if there exists some $Z \in \mathsf{An}(Y)$ such that either (a) or (b) is true:*

(a) There is a bi-directed edge between Z and Y in $\mathcal{G}$ but not in $\mathcal{G}'$.

(b) Neither of the graphs $\mathcal{G}'$ and $\mathcal{G}$ have a bidirected edge between Z and Y, and there exists a bidirected edge in $\mathcal{G}$ between some $X \in MUCT(\mathcal{G}'_{\overline{\mathsf{Pa}(Z), \mathsf{Bi}(Z, \mathcal{G}')}}, Y)$ and Z but not in $\mathcal{G}'$.

Suppose we have access to the induced observable subgraph $\mathcal{G}'$ on ancestors of the reward node. We can start by testing for latent confounders between Y and any node in $\mathsf{An}(Y)$. Then, we need to test for latent confounders between any pair $Z \in \mathsf{An}(Y)$ such that Z and Y don't have a bi-directed edge between them, and $X \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathsf{Pa}(Z), \mathsf{Bi}(Z, \mathcal{G}')}}, Y)$ until there are no new pairs to test. Theorem 3.1 can be useful because depending on the underlying causal graph, it saves us the number of latent confounders we need to test. For instance, consider a causal graph that has the reward Y with n different parent nodes, i.e., $\mathsf{Pa}(Y) = \{V_1, V_2, \ldots, V_n\}$, with no edges between the parents. In cases where every parent of Y is confounded with Y, or when none of them is confounded with Y, we only need to test for $|\mathsf{An}(Y)|$ latent variables, as implied by Theorem 3.1. However, in the worst-case scenario, we would need to test $\binom{|\mathsf{An}(Y)|+1}{2}$ latent variables when the true graph only has the confounders $V_1 \leftrightarrow Y$ and $V_i \leftrightarrow V_{i+1}$ for all $i = 1, .., n-1$. The exact number of latents

4

we need to test can range from $|\mathsf{An}(Y)|$ to $\binom{|\mathsf{An}(Y)|+1}{2}$ depending on the true graph. One issue still remains: we need a sample-efficient algorithm to learn the induced observable graph over $\mathsf{An}(Y)$ and to test the presence of confounders, which is addressed in upcoming sections.

4 Finite Sample Causal Discovery Algorithm

In this section, we propose a sample-efficient algorithm to learn causal graphs with latent confounders. We propose a two-phase approach. In the first phase, the algorithm learns the observable graph structure, i.e., the induced graph between observed variables. In the second phase, it detects the latent confounders. In the next section, we use the proposed discovery algorithm to construct the algorithm for causal bandits with an unknown graph. We begin by proposing two Lemmas to learn the ancestrality relations and latent confounders using interventions.

Lemma 4.1. *Consider a causal graph $\mathcal{G}(\boldsymbol{V}, \boldsymbol{E})$ and $\boldsymbol{W} \subseteq \boldsymbol{V}$. Furthermore, let $X, T \in \boldsymbol{V} \setminus \mathbf{W}$ be any two variables. Under the faithfulness Assumption 2.1 $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\boldsymbol{w}}}}$ if and only if for any $\boldsymbol{w} \in [K]^{|\boldsymbol{W}|}$, we have $P(t|do(\boldsymbol{w})) \neq P(t|do(\boldsymbol{w}), do(x))$ for some $x, t \in [K]$.*

Lemma 4.2. *Consider two variables X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$ and a set of variables $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \boldsymbol{W}$ and $X_i, X_j \notin \mathbf{W}$. Under the faithfulness Assumption 2.1 there is latent confounder between X_i and X_j if and only if for any $\boldsymbol{w} \in [K]^{|\boldsymbol{W}|}$, we have $P(x_j \mid do(x_i), do(\boldsymbol{W} = \boldsymbol{w})) \neq P(x_j \mid x_i, do(\boldsymbol{W} = \boldsymbol{w}))$ for some realization $x_i, x_j \in [K]$.*

These Lemmas are modified versions of Lemma 1 in [13] and Interventional Do-see test in [13], respectively. The difference between Lemma 4.1 and Lemma 1 in [13] is that we have an inequality test that can be used in the sample-efficient discovery instead of a statistical independence test. The Interventional Do-see test in [13] is valid for adjacent nodes only; however, our Lemma 4.2 can be used to test presence of latent confounder between any pair of nodes. This is because the condition in Lemma 4.2, $X_j \notin \mathsf{An}(X_i)$, can always be satisfied for any pair by flipping the order when one node is an ancestor of the other. In order to provide theoretical guarantees on sampling complexity, the inequality conditions are not enough; we need to assume certain gaps similar to [7, 8, 16].

Assumption 4.1. *Consider a causal graph $\mathcal{G}(\boldsymbol{V}, \boldsymbol{E})$ and $\boldsymbol{W} \subseteq \boldsymbol{V}$. Furthermore, let $X, T \in \boldsymbol{V} \setminus \mathbf{W}$ be any two variables. Then, we have $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\boldsymbol{w}}}}$ if and only if for any $\boldsymbol{w} \in [K]^{|\boldsymbol{W}|}$, we have $|P(t|do(\boldsymbol{w})) - P(t|do(\boldsymbol{w}), do(x))| > \epsilon$ for some $x, t \in [K]$, where $\epsilon > 0$ is some constant.*

Assumption 4.2. *Consider two variables X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$ and a set of variables $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and $X_i, X_j \notin \mathbf{W}$. There is a latent confounder or a bidirected edge between X_i and X_j if and only if for any $\boldsymbol{w} \in [K]^{|\boldsymbol{W}|}$, we have $|P(x_j \mid do(x_i), do(\boldsymbol{W} = \boldsymbol{w})) - P(x_j \mid x_i, do(\boldsymbol{W} = \boldsymbol{w}))| > \gamma$ for some realization $x_i, x_j \in [K]$ and some constant $\gamma > 0$.*

4.1 Learning the Observable Graph

We propose Algorithm 1 to learn the transitive closure under any arbitrary intervention $do(\mathbf{W})$, denoted by $\mathcal{G}^{tc}_{\overline{\mathbf{W}}}$. We use the Assumption 4.1 to bound the number of samples for ancestrality tests. We start with an empty graph and add edges by running ancestrality tests for all pairs of nodes in $\mathbf{V} \setminus \mathbf{W}$, resulting in the transitive closure $\mathcal{G}^{tc}_{\overline{\mathbf{W}}}$. We recall that the transitive reduction $Tr(\mathcal{G}) = (\mathbf{V}, \mathbf{E}^r)$ of a DAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ is unique, with $\mathbf{E}^r \subseteq \mathbf{E}$, and it can be computed in polynomial time [21]. Also, note that $Tr(\mathcal{G}) = Tr(\mathcal{G}^{tc})$. We propose a randomized Algorithm 2 similar to the one proposed in [13] that repeatedly uses Algorithm 1 to learn the observable graph structure. The motivation behind the randomized Algorithm 2 is Lemma 5 from [13], which states that for any edge (X_i, X_j), consider a set of variables $\mathbf{W}$ such that $\{W_i : \pi(W_i) > \pi(X_i) \ \& \ W_i \in \mathsf{Pa}(X_j)\} \subseteq \mathbf{W}$ where π is any total order that is consistent with the partial order implied by the DAG, i.e., $\pi(X) < \pi(Y)$ iff $X \in \mathsf{An}(Y)$. In this case, the edge (X_i, X_j) will be present in the graph $Tr(\mathcal{G}_{\overline{\mathbf{W}}})$.

Algorithm 2 randomly selects $\mathbf{W}$, computes the transitive reduction of the post-interventional graph, and finally accumulates all edges found in the transitive reduction across iterations. Algorithm 2 takes a parameter $d_{\max}$, which must be greater than or equal to the highest degree for our theoretical guarantees to hold. Algorithm 1 selects a realization $w \in [K]^{|\mathbf{W}|}$, takes B samples from intervention $do(\mathbf{W} = \mathbf{w})$, and A samples from every $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ for all $X_i \in \mathbf{V} \setminus \mathbf{W}$ and $x_i \in [K]$ interventions. We provide the sampling complexity guarantee for Algorithm 2 in Theorem 4.1.

Algorithm 1: Learn the Transitive Closure of the Causal Graph under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$

---

1 **Function** `LearnTransitiveClosure(`$\mathbf{V}, \mathbf{W}, \delta_1, \delta_2$`):`

2 $\quad$ $\mathbf{E} = \emptyset$, Fix some $\mathbf{w} \in [K]^{|\mathbf{W}|}$ and A $= max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2nK^2}{\delta_1}$ and $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$

3 $\quad$ Get B samples from $do(\mathbf{W} = \mathbf{w})$

4 $\quad$ Get A samples from every $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ $\forall X_i \in \mathbf{V} \setminus \mathbf{W}$ and $\forall x_i \in [K]$

5 $\quad$ **for** *every pair* $X_i, X_j \in \mathbf{V} \setminus \mathbf{W}$ **do**

6 $\quad\quad$ Use the Interventional Data to Test if $(X_i \in \mathsf{An}(X_j))_{\mathcal{G}_{\overline{\mathbf{W}}}}$

7 $\quad\quad$ **if** $\exists\, x_i, x_j \in [K]$ *s.t.* $|\widehat{P}(x_j|do(\mathbf{w})) - \widehat{P}(x_j|do(\mathbf{w}), do(x_i))| > \frac{\epsilon}{2}$ **then**

8 $\quad\quad\quad$ $\mathbf{E} \longleftarrow \mathbf{E} \cup (X_i, X_j)$

9 $\quad$ **return** The graph's transitive closure $(\mathbf{V}, \mathbf{E})$ and All Interventional data

---

Algorithm 2: Learn the Observable Graph

---

1 **Function** `LearnObservableGraph(`$\mathbf{V}, \alpha, d_{max}, \delta_1, \delta_2$`):`

2 $\quad$ $\mathbf{E} = \emptyset$ & $\mathcal{I}Data = \emptyset$

3 $\quad$ **for** $i = 1 \; : \; 8\alpha\, d_{max}\, \log(n)$ **do**

4 $\quad\quad$ $\mathbf{S} = \emptyset$

5 $\quad\quad$ **for** $V_i \in \mathbf{V}$ **do**

6 $\quad\quad\quad$ $\mathbf{W} \leftarrow \mathbf{W} \cup V_i$ with probability $1 - \frac{1}{2d_{max}}$

7 $\quad\quad$ $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}, \text{Data}_{\overline{\mathbf{W}}} = \mathsf{LearnTransitiveClosure}(\mathbf{V}, \mathbf{W}, \delta_1, \delta_2)$

8 $\quad\quad$ Compute the transitive reduction $Tr(\mathcal{G}_{\overline{\mathbf{W}}}^{tc})$ & add any missing edges from $Tr(\mathcal{G}_{\overline{\mathbf{W}}}^{tc})$ to $\mathbf{E}$

9 $\quad\quad$ $\mathcal{I}Data = \mathcal{I}Data \cup \text{Data}_{\overline{\mathbf{W}}}$ (Keep Saving Interventional Data)

10 $\quad$ **return** The observable graph structure $(\mathbf{V}, \mathbf{E})$ and interventional data samples in $\mathcal{I}Data$

Theorem 4.1. *Algorithm 2 learns the true observable graph with probability at least* $1 - \frac{1}{n^{\frac{\alpha}{2d_{max}} - 2}} - 8\alpha d_{max} \log(n)(n\delta_1 + \delta_2)$ *with* $8\alpha d_{max} \log n(KAn + B)$ *interventional samples. If we set* $\alpha = \frac{2d_{max} \log(\frac{2}{\delta} + 2)}{\log n}$, $\delta_1 = \frac{\delta}{32\alpha d_{max} n \log n}$ *and* $\delta_2 = \frac{\delta}{32\alpha d_{max} \log n}$, *then Algorithm 2 learns the observable graph with probability at least* $1 - \delta$. *(We have* $A = max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2nK^2}{\delta_1}$ *and* $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$*)*

4.2 Learning the Latent Confounders

Assumption 4.2 can be used to test for latents between any pair of observed variables. For any variables X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$, we need access to interventional samples $do(\mathbf{W} = \mathbf{w})$ such that $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$. In addition to simple causal effects we need to estimate the conditional causal effect of the form $P(x_j|x_i, do(\mathbf{W} = \mathbf{w}))$. To bound the number of samples required to ensure accurate estimation of the conditional causal effects, we rely on Assumption 4.3. Note that Assumption 4.3 does not restrict the applicability of our algorithm; it simply assumes that under an intervention $do(\mathbf{W} = \mathbf{w})$, either the probability of observing a realization $X_i = x_i$ is zero or is lower-bounded by some constant $\eta > 0$.

Assumption 4.3. *For any variable $X_i \in \mathbf{V}$ and any intervention $do(\mathbf{W} = \mathbf{w})$ where $\mathbf{W} \subseteq \mathbf{V}$ and $\mathbf{w} \in [K]^{|\mathbf{W}|}$, we assume that either $P(x_i|do(\mathbf{W} = \mathbf{w})) = 0$ or $P(x_i|do(\mathbf{W} = \mathbf{w})) \geq \eta > 0$.*

Lemma 4.3. *Consider two nodes X_i and X_j s.t. $X_j \notin \mathsf{An}(X_i)$ and suppose that Assumptions 2.1 4.2 hold and we have access to $max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2K^2}{\delta_1}$ samples from $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ $\forall x_i \in [K]$ and $\frac{16}{\eta\gamma^2} \log(\frac{2K^2}{\delta_3}) + \frac{1}{2\eta^2} \log(\frac{2K^2}{\delta_4})$ samples from $do(\mathbf{W} = \mathbf{w})$ for a fixed $w \in [K]^{|\mathbf{W}|}$ and $\mathbf{W} \subseteq \mathbf{V}$ such that $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and $X_i, X_j \notin \mathbf{W}$. Then, with probability at least $1 - \delta_1 - \delta_3 - \delta_4$, we have a latent confounder between X_i and X_j iff $\exists\, x_i, x_j \in [K]$ s.t. $\left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w})) \right| > \frac{\gamma}{2}$.*

Our full discovery algorithm, which learns the observable graph structure and all latent variables along with the sampling complexity guarantee, is provided in supplementary material A.9. Suppose the constant gaps ϵ and γ in Assumptions 4.1 and 4.2 are close; then we have $C > \frac{1}{\eta}A \geq \frac{1}{\eta}B$. The

constant $0 < \eta < 1$ is usually small in practical scenarios, so the quantity C is much greater than both B and A. This implies that the number of samples required to test the presence of latent variables is greater than that required to learn ancestral relations.

Algorithm 3: Sketch of Algorithm for causal bandits with unknown graph structure

---

1 Calculate $\alpha, \delta_1, \delta_2, \delta_3, \delta_4$ as in Theorem 5.1
2 $\mathcal{G}^{tc}$ = LearnTransitiveClosure($\mathbf{W} = \phi, \frac{\delta}{2n}, \frac{\delta}{n}$)
3 $\mathcal{G}, \mathcal{I}Data$ = LearnObservableGraph($\mathsf{An}(Y)_{\mathcal{G}^{tc}}, \alpha, d_{max}, \delta_1, \delta_2$)
4 # Learn the bi-directed edges between reward Y and all nodes $X_i \in \mathsf{An}(Y)$ and update $\mathcal{G}$.
5 **for** *every* $X_i \in \mathsf{An}(Y)_{\mathcal{G}^{tc}}$ **do**
6 $\quad$ $\mathcal{G}$ = DetectLatentConfounder($\mathcal{G}, X_i, X_j, \delta_2, \delta_3, \delta_4, \mathcal{I}Data$) (Algorithm 4)

7 **while** *There is a new pair that is tested* **do**
8 $\quad$ Find a new pair (Z, X) s.t. $Z \in \mathsf{An}(Y)$ such that Z and Y don't have a bi-directed edge between them
$\quad\quad$ in $\mathcal{G}$ and $X \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathsf{Pa}(Z), \mathsf{Bi}(Z, \mathcal{G})}}, Y)$ and test for the latent and update $\mathcal{G}$.
9 $\quad$ $\mathcal{G}$ = DetectLatentConfounder($\mathcal{G}, Z, X, \delta_2, \delta_3, \delta_4, \mathcal{I}Data$)

10 Learn the set of POMISs $\mathcal{I}_{\mathcal{G}}$ from the graph $\mathcal{G}$ (Using Algorithm 1 from [5]).
11 Run UCB algorithm over the arm set $A = \{\Omega(I) \mid \forall I \in \mathcal{I}_{\mathcal{G}}\}$.

---

5 Algorithm for Causal Bandits with Unknown Graph Structure

Algorithm 3 first learns the transitive closure of the graph $\mathcal{G}^{tc}$ to find ancestors of the reward node Y, since POMISs are subsets of $\mathsf{An}(Y)$. The next step is to learn the observed graph structure among the reward Y and nodes in $\mathsf{An}(Y)$. Instead of detecting the presence of confounders between all pairs of nodes in $\mathsf{An}(Y)$, we focus on identifying the necessary and sufficient ones, as characterized by Theorem 3.1. This approach is sample-efficient since it tests for fewer latent confounders. The last step of Algorithm 3 is to run a simple bandit algorithm, e.g., UCB algorithm [22], to identify the optimal arm from the POMISs. Given that Assumptions 4.1, 4.2, and 4.3 hold, and the reward is binary ($Y \in \{0, 1\}$), we provide a worst-case regret bound for Algorithm 3 in Theorem 5.1.

Theorem 5.1. *Algorithm 3 learns the true set of POMISs with probability at least $1 - 2\delta$. Under the event that it learns POMISs correctly, the cumulative regret is bounded as follows:*

$$R_T \leq Kn \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{4n^2 K^2}{\delta} \;+\; \frac{8}{\epsilon^2} \log \frac{4nK^2}{\delta} \;+$$

$$8\alpha d_{max}\left(KA|\mathsf{An}(Y)| \;+\; \max(B, C)\right) \log\left(|\mathsf{An}(Y)|\right) \;+\; \sum_{\mathbf{s} \in \{\Omega(I) | \forall I \in \mathcal{I}_{\mathcal{G}}\}} \Delta_{do(\mathbf{s})}\left(1 + \frac{\log T}{\Delta_{do(\mathbf{s})}^2}\right),$$

where A and B are given by line 2 of Algorithm 1, and $C = \frac{16}{\eta\gamma^2} \log\left(\frac{2n^2 K^2}{\delta_3}\right) + \frac{1}{2\eta^2} \log\left(\frac{2n^2 K^2}{\delta_4}\right)$ with $\alpha = \frac{2d_{max} \log\left(\frac{4}{\delta} + 2\right)}{\log\left|\mathsf{An}(Y)\right|}$, $\delta_1 = \frac{\delta}{64\alpha d_{max}\left|\mathsf{An}(Y)\right| \log\left|\mathsf{An}(Y)\right|}$ and $\delta_2 = \delta_3 = \delta_4 = \frac{\delta}{64\alpha d_{max} \log\left|\mathsf{An}(Y)\right|}$.

The first four terms in the regret bound correspond to the interventional samples required to learn the ancestors of the reward node and the set of POMISs ($\mathcal{I}_{\mathcal{G}}$). The fifth term corresponds to the regret incurred during the UCB phase.

6 Experiments

Theorem 5.1 establishes the worst-case upper bound for cumulative regret when we need to test latent confounders between all pairs of nodes within $\mathsf{An}(Y)$. However, Algorithm 3 selectively examines only a subset of latent confounders sufficient to infer the true POMIS set. Although the advantage is hard to quantify in general, we demonstrate it using simulations on randomly generated graphs. We sample a random ordering σ among the vertices. Then, for each nth node, we determine its in-degree as $X_n = \max(1, \mathrm{Bin}(n-1, \rho))$, followed by selecting its parents through uniform sampling from the preceding nodes in the ordering. Finally, we chordalize the graph using the elimination algorithm [23], employing an elimination ordering that is the reverse of σ. Additionally, we introduce a confounder between every pair of nodes with a probability of ρ_L. For all the simulations, we randomly sample 50 causal graphs with different values of densities ρ and ρ_L and assume that all variables are binary for

simplicity, i.e., $K = 2$. We set the value of δ to 0.99, and the gaps $\gamma = \epsilon = 0.01$ and $\eta = 0.05$. We plot interventional samples used to learn the induced observable graph on $\mathsf{An}(Y)$ with and without latent confounders, as well as the samples required to learn the POMIS set by Algorithm 3. The width of confidence interval is set to 2 standard deviations.
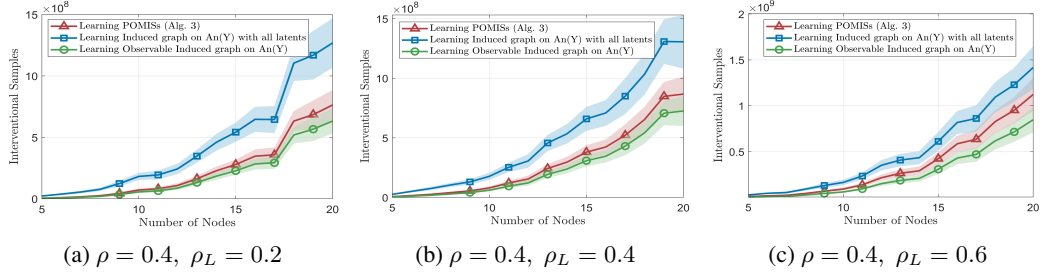

(a) $\rho = 0.4, \ \rho_L = 0.2$

(b) $\rho = 0.4, \ \rho_L = 0.4$

(c) $\rho = 0.4, \ \rho_L = 0.6$

Figure 2: Simulations to demonstrate advantage of Algorithm 3 over learning all possible latents.

(a) $\rho = 0.4, \ \rho_L = 0.2$

(b) $\rho = 0.4, \ \rho_L = 0.4$

(c) $\rho = 0.4, \ \rho_L = 0.6$
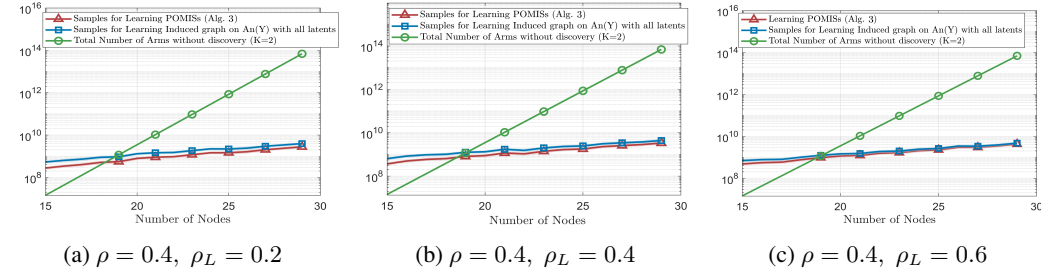
Figure 3: Simulations to demonstrate advantage of discovery for causal bandits.

(a) Nodes $n = 10$
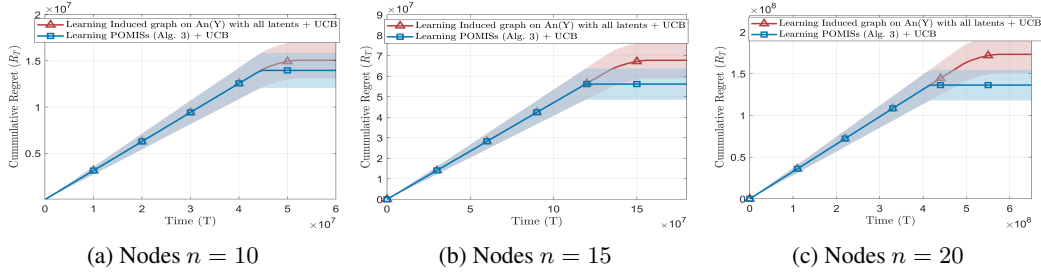
(b) Nodes $n = 15$

(c) Nodes $n = 20$

Figure 4: Cumulative regret for Algorithm 3 versus learning all possible latents ($\rho = \rho_L = 0.3$).

Algorithm 3 requires fewer samples than learning the induced graph on $\mathsf{An}(Y)$ with all confounders. However, the advantage is reduced as ρ_L increases. The plots in Figure 3 compare the exponentially growing arms in causal bandits with interventions samples used by our algorithm to learn the reduced action set in the form of POMISs. We also run the UCB algorithm on the learned POMIS set and plot the cumulative regret in Figure 4. Since the number of time steps T is on the order of 10^8, it is not feasible to store and plot cumulative regret for every time step over multiple randomly sampled graphs; therefore, we downsample the cumulative regret to show the overall trend. The downsampling, along with the large scale of the y-axis, makes the regret in the discovery phase appear linear with a fixed slope, although it is piece-wise linear if we zoom in. Also, the UCB phase converges very fast compared to the discovery phase because the number of POMISs for randomly sampled graphs is small. We plot the results for graphs with 10, 15, and 20 nodes, and in all cases, we can see the advantage of partial discovery compared to full discovery, resulting in lower regret.

7   Conclusion

We show that partial discovery is sufficient to achieve sublinear regret for causal bandits with an unknown causal graph containing latent confounders. Without relying on causal discovery, one has to consider interventions on all possible subsets of nodes, which is infeasible. Thus, we propose a two-phase approach where the first phase learns the induced subgraph on the ancestors of the reward node, along with a subset of confounders, to construct the set of possibly optimal arms, and the second phase involves the application of the UCB algorithm.

8

Acknowledgements

References

[1] Finnian Lattimore, Tor Lattimore, and Mark D Reid. Causal bandits: Learning good interventions via causal inference. *Advances in Neural Information Processing Systems*, 29, 2016.

[2] Rajat Sen, Karthikeyan Shanmugam, Alexandros G Dimakis, and Sanjay Shakkottai. Identifying best interventions through online importance sampling. In *International Conference on Machine Learning*, pages 3057–3066. PMLR, 2017.

[3] Yangyi Lu, Amirhossein Meisami, Ambuj Tewari, and William Yan. Regret analysis of bandit problems with causal background knowledge. In *Conference on Uncertainty in Artificial Intelligence*, pages 141–150. PMLR, 2020.

[4] Vineet Nair, Vishakha Patil, and Gaurav Sinha. Budgeted and non-budgeted causal bandits. In *International Conference on Artificial Intelligence and Statistics*, pages 2017–2025. PMLR, 2021.

[5] Sanghack Lee and Elias Bareinboim. Structural causal bandits: Where to intervene? *Advances in neural information processing systems*, 31, 2018.

[6] Lai Wei, Muhammad Qasim Elahi, Mahsa Ghasemi, and Murat Kocaoglu. Approximate allocation matching for structural causal bandits with unobserved confounders. *Advances in Neural Information Processing Systems*, 36, 2024.

[7] Yangyi Lu, Amirhossein Meisami, and Ambuj Tewari. Causal bandits with unknown graph structure. *Advances in Neural Information Processing Systems*, 34:24817–24828, 2021.

[8] Mikhail Konobeev, Jalal Etesami, and Negar Kiyavash. Causal bandits without graph learning. *arXiv preprint arXiv:2301.11401*, 2023.

[9] Alan Malek, Virginia Aglietti, and Silvia Chiappa. Additive causal bandits with unknown graph. *arXiv preprint arXiv:2306.07858*, 2023.

[10] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

[11] Xinpeng Shen, Sisi Ma, Prashanthi Vemuri, and Gyorgy Simon. Challenges and opportunities with causal discovery algorithms: application to alzheimer's pathophysiology. *Scientific reports*, 10(1):2975, 2020.

[12] Alessio Zanga, Elif Ozkirimli, and Fabio Stella. A survey on causal discovery: Theory and practice. *International Journal of Approximate Reasoning*, 151:101–129, 2022.

[13] Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. *Advances in Neural Information Processing Systems*, 30, 2017.

[14] Kun Zhang, Biwei Huang, Jiji Zhang, Clark Glymour, and Bernhard Schölkopf. Causal discovery from nonstationary/heterogeneous data: Skeleton estimation and orientation determination. In *IJCAI: Proceedings of the Conference*, volume 2017, page 1347. NIH Public Access, 2017.

[15] Karthikeyan Shanmugam, Murat Kocaoglu, Alexandros G Dimakis, and Sriram Vishwanath. Learning causal graphs with small interventions. *Advances in Neural Information Processing Systems*, 28, 2015.

[16] Kristjan Greenewald, Dmitriy Katz, Karthikeyan Shanmugam, Sara Magliacane, Murat Kocaoglu, Enric Boix Adsera, and Guy Bresler. Sample efficient active learning of causal trees. *Advances in Neural Information Processing Systems*, 32, 2019.

[17] Muhammad Qasim Elahi, Lai Wei, Murat Kocaoglu, and Mahsa Ghasemi. Adaptive online experimental design for causal discovery, 2024.

[18] Judea Pearl. *Causality*. Cambridge university press, 2009.

[19] Jin Tian and Judea Pearl. A general identification condition for causal effects. In *Aaai/iaai*, pages 567–573, 2002.

[20] Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventional data. *International Journal of Approximate Reasoning*, 55(4):926–939, 2014.

[21] Alfred V. Aho, Michael R Garey, and Jeffrey D. Ullman. The transitive reduction of a directed graph. *SIAM Journal on Computing*, 1(2):131–137, 1972.

[22] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[23] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

A Supplemental Material

A.1 Review of d-separation:

Consider three disjoint sets of nodes $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ in the causal graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$. The sets of nodes $\mathbf{X}$ and $\mathbf{Y}$ are d-separated given $\mathbf{Z}$, denoted by $(\mathbf{X} \perp\!\!\!\perp_d \mathbf{Y}|\mathbf{Z})_{\mathcal{G}}$, if and only if there exists no path, directed or undirected, between any node in set $\mathbf{X}$ and any node in set $\mathbf{Y}$ such that for every collider on the path, either the collider itself or one of its descendants is included in the set $\mathbf{Z}$, and no other non-collider nodes on the path are included in the set $\mathbf{Z}$. (A collider on a path is a node with both arrows converging, e.g., B is a collider on the path ABC in $A \rightarrow B \leftarrow C$).

A.2 Pearl's Rules of do-Calculus ([18]):

Let $\mathcal{G}$ represent the causal DAG, and let P denote the probability distribution induced by the corresponding causal model. For any disjoint subsets of variables $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, and $\mathbf{W}$, the following rules apply:

Rule 1: (Insertion/deletion of observations):

$$P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) \quad \text{if} \quad (\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z}|\mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}}}. \tag{2}$$

Rule 2: (Action/observation exchange):

$$P(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{z}, \mathbf{w}) \quad \text{if} \quad (\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z}|\mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}\underline{\mathbf{Z}}}}. \tag{3}$$

Rule 3: (Insertion/deletion of actions):

$$P(\mathbf{y}|do(\mathbf{x}), do(\mathbf{z}), \mathbf{w}) = P(\mathbf{y}|do(\mathbf{x}), \mathbf{w}) \quad \text{if} \quad (\mathbf{Y} \perp\!\!\!\perp_d \mathbf{Z}|\mathbf{X}, \mathbf{W})_{\mathcal{G}_{\overline{\mathbf{X}}, \overline{\mathbf{Z}(\mathbf{W})}}}, \tag{4}$$

where $\mathbf{Z}(\mathbf{W})$ is the set of nodes in $\mathbf{Z}$ that are not ancestors of any of the nodes in $\mathbf{W}$ in the graph $\mathcal{G}_{\overline{\mathbf{X}}}$.

A.3 Function to detect presence of Latent Confounder

Algorithm 4: Function to detect presence of Latent Confounder

---
1 **Function** DetectLatentConfounder($\mathcal{G}, X_i, X_j, \delta_2, \delta_3, \delta_4, \mathcal{I}Data$)**:**
2      $C = \frac{16}{\eta\gamma^2} \log(\frac{2n^2K^2}{\delta_3}) + \frac{1}{2\eta^2} \log(\frac{2n^2K^2}{\delta_4})$, $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$
3   if $X_j \in \mathsf{An}(X_i)$ swap them.
4   Find interventional data sets $do(\mathbf{W} = \mathbf{w})$ and $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ from $\mathcal{I}Data$ s.t.
         $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$
5   Get $\max(0, B - C)$ new samples for $do(\mathbf{W} = \mathbf{w})$
6   **if** $\exists\, x_i, x_j \in [K]$ s.t. $|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w}))| > \frac{\gamma}{2}$ **then**
7   Add bi-dirceted edge $X_i \leftrightarrow X_j$ to graph $\mathcal{G}$
8   **return** Updated Causal Graph $\mathcal{G}$

A.4 Proof of Theorem 3.1:

Before proving Theorem 3.1, we state and prove two Lemmas.

Lemma A.1. *It is necessary to learn/detect the latent confounders between reward node Y and any node $X \in \mathsf{An}(Y)$ in causal graph $\mathcal{G}$ to learn all the POMISs correctly and hence avoid linear regret.*

Before proceeding to the proof, we recall an important result from [5]: For a causal graph $\mathcal{G}$ with reward variable Y, $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ is a POMIS for any $\mathbf{W} \subseteq \mathbf{V} \setminus Y$.

Proof: Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ with a node $X \in \mathsf{An}(Y)$ such that there exists a latent confounder between X and the reward Y. Suppose we do not detect the presence of the confounder and have access to another causal graph $\mathcal{G}'$ with everything the same as $\mathcal{G}$ except that there is no confounder between X and Y. We show that there exists one such POMIS that we cannot learn from $\mathcal{G}'$, which actually exists in the true causal graph $\mathcal{G}$. To prove this, consider a set of nodes

11

$\mathbf{W} = \mathsf{Pa}(X) \cup \mathsf{Ch}(\mathsf{Pa}(X)) \cup \mathsf{CC}(X) \setminus \{X, Y\}$. For the graph $\mathcal{G}'$, note that $X \notin \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, and also there $\nexists Z \in \mathsf{Ch}(\mathsf{Pa}(X)) \setminus \{X\}$ s.t. $Z \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$. This implies that $\nexists Z \in \mathsf{Pa}(X)$ s.t. $Z \in \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$. However, for the true graph $\mathcal{G}$, we have a different $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ for the same definition of $\mathbf{W}$ because it contains the bi-directed edge between X and Y, which implies that $X \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$, and as a result, $\mathsf{Pa}(X) \subseteq \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Also, in the case $\mathsf{Pa}(X) = \emptyset$, we have a different POMIS. On this side, note that $X \notin \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, which implies that along the causal path from X to Y, there must be one node Z such that $Z \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, which implies either X or one of its descendants on the path from X to Y is in $\mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, which is not the case for $\mathcal{G}$ since $X \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Thus, we have different interventional boundary or POMIS for the two causal graphs $\mathcal{G}$ and $\mathcal{G}'$ given the above choice of $\mathbf{W}$, even if X has no parents.

The next step is to show that the particular POMIS $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ cannot be learned from the DAG $\mathcal{G}'$, i.e., $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y) \neq \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$ for any $\mathbf{W}' \subseteq \mathbf{V}$. We need to show this because of the graphical characterization of POMISs in Lemma 3.1. Using the definition of $\mathbf{W}$, note that $\mathsf{Pa}(X) \subseteq \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ and for all $Z \in \mathsf{Ch}(\mathsf{Pa}(X)) \setminus \{X\}$, there exists either $Z \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ or $\mathsf{De}(Z) \setminus \{Y\} \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Also, if there are such nodes in $\mathsf{CC}(X) \setminus \{X, Y\}$ which do not have a path to X comprised of directed edges only, call such set of nodes T. If $T \neq \phi$, then for all $t \in T$, we have either $t \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ or $\mathsf{De}(t) \setminus \{Y\} \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Also, note that $\nexists Z \in \mathsf{De}(X) \cup \{X\}$ such that $Z \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Now consider DAG $\mathcal{G}'$ with the bi-directed edge between X and Y missing. Assume by contradiction $\exists \mathbf{W}' \subseteq \mathbf{V}$ such that $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y) = \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$. This, however, using the aforementioned characterization of $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ implies that $\nexists Z \in \mathsf{Ch}(\mathsf{Pa}(X)) \setminus \{X\}$ such that $Z \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$ and also $\nexists t \in T$ such that $t \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$ using the aforementioned definition of T. However, note that we need $\mathsf{Pa}(X) \subseteq \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$, which under the given choice of $\mathbf{W}$ is only possible when $X \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$, which would require is a bi-directed edge between X and Y in the DAG $\mathcal{G}'$, which is a contradiction. Also, for the case when $\mathsf{Pa}(X) = \phi$, we have a contradiction because we require the following to be true: $\nexists Z \in \mathsf{De}(X) \cup \{X\}$ such that $Z \in \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$. For the given choice of $\mathbf{W}$, it implies that there is a bi-directed edge between X and Y in the DAG $\mathcal{G}'$, which is a contradiction. Thus, by contradiction, we show that $\nexists \mathbf{W}' \subseteq \mathbf{V}$ such that $\mathcal{G}'$, i.e., $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y) \neq \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$. Thus, we will miss at least one POMIS given that we do not consider latent confounders between the reward node Y and any node $X \in \mathsf{An}(Y)$, and may incur linear regret. This completes the proof of Lemma A.1.

Lemma A.2. *Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ and another graph $\mathcal{G}'$ such that they have the same vertex set and directed edges but differ in bidirected edges, with the bidirected edges in $\mathcal{G}'$ being a subset of the bidirected edges in $\mathcal{G}$. The graphs will yield different collections of POMISs if there exists some $Z \in \mathsf{An}(Y)$ such that either (a) or (b) is true:*

(a) *There is a bi-directed edge between Z and Y in $\mathcal{G}$ but not in $\mathcal{G}'$.*

(b) *Neither of the graphs $\mathcal{G}'$ and $\mathcal{G}$ have a bidirected edge between Z and Y, and there exists a bidirected edge in $\mathcal{G}$ between some $X \in MUCT(\mathcal{G}'_{\overline{\mathsf{Pa}(Z), \mathsf{Bi}(Z, \mathcal{G}')}}, Y)$ and Z but not in $\mathcal{G}'$.*

Proof: The first half of Lemma A.2, i.e., "The graphs will yield different collections of POMISs if there exists some $Z \in \mathsf{An}(Y)$ such that there is a bi-directed edge between Z and Y in $\mathcal{G}$ but not in $\mathcal{G}'$," is the same as Lemma A.1, and the same proof applies here. The reason is that in graph $\mathcal{G}'$, we miss a latent variable between reward and one of its ancestors, which was actually present in the true graph $\mathcal{G}$. We only need to proof the second half of Lemma A.2 i.e. graphs will yield different collections of POMISs if there exists some $Z \in \mathsf{An}(Y)$ such that (b) is true. Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ and another DAG $\mathcal{G}'$ such that they have the same vertex set and directed edges, but differ in bi-directed edges. Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ and another DAG $\mathcal{G}'$ such that they have the same vertex set and directed edges, but differ in bi-directed edges. We show that if neither of the graphs $\mathcal{G}'$ and $\mathcal{G}$ have a bidirected edge between Z and Y, and there exists a bidirected edge in $\mathcal{G}$ between some $X \in \mathrm{MUCT}(\mathcal{G}', Y) \setminus \mathsf{Pa}(Z)$ and Z, then there exists one such POMIS that we cannot learn from $\mathcal{G}'$, which actually exists in the true causal graph $\mathcal{G}$. To prove this, consider a set of nodes $\mathbf{W} = \mathsf{Pa}(Z) \cup \mathsf{Ch}(\mathsf{Pa}(Z) \setminus \mathsf{An}(X)) \cup \mathsf{Bi}(Z, \mathcal{G}') \setminus \{X, Z, Y\}$. For the graph $\mathcal{G}'$, note that $Z \notin \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, and also there $\nexists N \in \mathsf{Ch}(\mathsf{Pa}(Z) \setminus \mathsf{An}(X)) \setminus \{Z\}$ s.t. $N \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$. This implies that $\nexists N \in \mathsf{Pa}(Z) \setminus \mathsf{An}(X)$ s.t. $N \in \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$. However, for the true graph $\mathcal{G}$, we have a different $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ for the same definition of $\mathbf{W}$ because it contains the bi-directed edge between X and Z, which implies that $Z \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$, and as a result, $\mathsf{Pa}(Z) \setminus \mathsf{An}(X) \subseteq \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Also, in the case $\mathsf{Pa}(Z) \setminus \mathsf{An}(X) = \emptyset$, we have different a POMIS.

On this side, note that $Z \notin \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, which implies that along the causal path from Z to Y, there must be one node N such that $N \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, which implies either Z or one of its descendants on the path from Z to Y is in $\mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}}}, Y)$, which is not the case for $\mathcal{G}$ since $Z \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Thus, we have different interventional boundary or POMIS for the two causal graphs $\mathcal{G}$ and $\mathcal{G}'$ given the above choice of $\mathbf{W}$.

The next step is to show that the particular POMIS $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ cannot be learned from the DAG $\mathcal{G}'$, i.e., $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y) \neq \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$ for any $\mathbf{W}' \subseteq \mathbf{V}$. We need to show this because of the graphical characterization of POMISs in Lemma 3.1. Using the definition of $\mathbf{W}$, note that $\mathsf{Pa}(Z) \setminus \mathsf{An}(X) \subseteq \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ and for all $N \in \mathsf{Ch}(\mathsf{Pa}(Z) \setminus \mathsf{An}(X)) \setminus \{Z\}$, there exists either $N \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ or $\mathsf{De}(N) \setminus \{Y\} \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Also, if there are such nodes in $\mathsf{Bi}(Z, \mathcal{G}') \setminus \{X, Z, Y\}$ which do not have a path to Z comprising of directed edges only, call such set of nodes T. If $T \neq \phi$, then for all $t \in T$, we have either $t \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ or $\mathsf{De}(t) \setminus \{Y\} \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Also, note that $\nexists N \in \mathsf{De}(Z) \cup \{Z\}$ such that $N \in \mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}'}}, Y)$. Now consider the DAG $\mathcal{G}'$ with the bi-directed edge between Z and Y missing. Assume by contradiction $\exists \mathbf{W}' \subseteq \mathbf{V}$ such that $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y) = \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$. This, however, using the aforementioned characterization of $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ implies that $\nexists N \in \mathsf{Ch}(\mathsf{Pa}(Z) \setminus \mathsf{An}(X)) \setminus \{Z\}$ such that $N \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$ and also $\nexists t \in T$ such that $t \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$ for the aforementioned definition of T. However, note that we need $\mathsf{Pa}(Z) \setminus \mathsf{An}(X) \subseteq \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$, which under the given choice of $\mathbf{W}$ is only possible when $Z \in \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$, which would require a bi-directed edge between Z and X in the DAG $\mathcal{G}'$, which is a contradiction. Also, for the case when $\mathsf{Pa}(Z) = \phi$, we have a contradiction because we require the following to be true: $\nexists N \in \mathsf{De}(Z) \cup \{Z\}$ such that $N \in \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$. For the given choice of $\mathbf{W}$, it implies that there is a bi-directed edge between Z and X in the DAG $\mathcal{G}'$, which is a contradiction. Thus, by contradiction, we show that $\nexists \mathbf{W}' \subseteq \mathbf{V}$ such that $\mathcal{G}'$, i.e., $\mathrm{IB}(\mathcal{G}_{\overline{\mathbf{W}}}, Y) \neq \mathrm{IB}(\mathcal{G}'_{\overline{\mathbf{W}'}}, Y)$. This implies that we miss atleast one POMIS when either of statements (a) and (b) hold. This proves the second half of Lemma A.2.

We now proceed to the formal proof for Theorem 3.1:

Theorem. 3.1: *Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ and another DAG $\mathcal{G}'$ such that they have the same vertex set and directed edges but differ in bidirected edges, with the bidirected edges in $\mathcal{G}'$ being a subset of the bidirected edges in $\mathcal{G}$. The graphs will yield different collections of POMISs if and only if there exists some $Z \in \mathsf{An}(Y)$ such that either (a) or (b) is true:*

(a) There is a bi-directed edge between Z and Y in $\mathcal{G}$ but not in $\mathcal{G}'$.

(b) Neither of the graphs $\mathcal{G}'$ and $\mathcal{G}$ have a bidirected edge between Z and Y, and there exists a bidirected edge in $\mathcal{G}$ between some $X \in MUCT(\mathcal{G}'_{\overline{\mathsf{Pa}(Z)}, \mathsf{Bi}(Z, \mathcal{G}')}, Y)$ and Z but not in $\mathcal{G}'$.

Proof: One direction for Theorem 3.1 is proved already in Lemma A.2. We only to need to prove the other direction which is that two causal graphs $\mathcal{G}$ and $\mathcal{G}'$ such that they have the same vertex set and directed edges, but differ in bi-directed edges will yield same collections POMISs when neither of statements (a) and (b) is true. Note when neither of (a) or (b) is true the graphs $\mathcal{G}$ and $\mathcal{G}'$ might still have a different set of bi-directed edges. We will have two possible scenarios here. Suppose $\mathcal{G}$ has a bi-directed edge between some $Z \in \mathsf{An}(Y)$ and some $X \in \mathsf{An}(Y)$, such that there is a bi-directed edge between pair of vertices (Z, Y) and (X, Y) in both the graphs and the bi-directed edge between X and Z is absent in $\mathcal{G}'$. Further, assume neither of statements (a) and (b) hold. In this case, despite the absence of a bi-directed edge between X and Z in $\mathcal{G}'$, the graphs will yield the same set of POMISs. This is because $Z \notin MUCT(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ for some set of nodes $\mathbf{W}$ only when $Z \in \mathbf{W}$, and the same is the case for $\mathcal{G}$ because they share a bi-directed edge between Z and Y. By symmetry, we have the argument hold for X as well. So, the presence or absence of bi-directed edges between X and Z does not change the set of POMISs learned from the graph when both X and Z are confounded with reward Y already. Thus, we can delete all such bi-directed edges one by one from $\mathcal{G}$ while the set of POMISs learned from each of the intermediate causal graphs stays the same. Consider the second scenario, where $\mathcal{G}$ has bi-directed edges between a node $Z \in \mathsf{An}(Y)$, such that there is no bi-directed edge between Z and Y in both graphs ($\mathcal{G}$ and $\mathcal{G}'$) and a node X that has the following characteristics: $X \in \mathrm{MUCT}(\mathcal{G}', Y)$ but $X \notin \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathsf{Pa}(Z)}, \mathsf{Bi}(Z, \mathcal{G}')}, Y)$. However, the bi-directed edge between X and Z is absent in $\mathcal{G}'$. Further, assume neither of statements (a) and (b) hold. The condition $X \in \mathrm{MUCT}(\mathcal{G}', Y)$ but $X \notin \mathrm{MUCT}(\mathcal{G}'_{\overline{\mathsf{Pa}(Z)}, \mathsf{Bi}(Z, \mathcal{G}')}, Y)$ implies that either $\exists N \in \mathsf{Pa}(Z)$ such that $N \in \mathrm{MUCT}(\mathcal{G}', Y)$ or $\exists N \in \mathsf{Bi}(Z, \mathcal{G}')$ such that $N \in \mathrm{MUCT}(\mathcal{G}', Y)$ Since bi-directed edges in $\mathcal{G}'$

are a subset of bi-directed edges in $\mathcal{G}$, we have: Either $\exists N \in \mathsf{Pa}(Z)$ such that $N \in \mathrm{MUCT}(\mathcal{G}, Y)$ or $\exists N \in \mathsf{Bi}(Z, \mathcal{G})$ such that $N \in \mathrm{MUCT}(\mathcal{G}, Y)$. Since any MUCT is closed under the $\mathsf{De}(.)$ and $\mathsf{CC}(.)$ operations, i.e., for any MUCT say $\mathbf{T}$ we have $\mathsf{De}(\mathbf{T}) = \mathbf{T}$ and $\mathsf{CC}(\mathbf{T}) = \mathbf{T}$ if $\exists N \in \mathsf{Pa}(Z)$ such that $N \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ or $\exists N \in \mathsf{Bi}(Z, \mathcal{G})$ such that $N \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$, we already have $Z \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ using the definition of MUCT. The bi-directed edge between X and Z will play a role only when $\nexists N \in \mathsf{Pa}(Z)$ such that $N \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ and $\nexists N \in \mathsf{Bi}(Z, \mathcal{G})$ such that $N \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ for some set of $\mathbf{W}$. Also note that $X \notin \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ when $\nexists N \in \mathsf{Pa}(Z)$ such that $N \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$ and $\nexists N \in \mathsf{Bi}(Z, \mathcal{G})$ such that $N \in \mathrm{MUCT}(\mathcal{G}_{\overline{\mathbf{W}}}, Y)$. Thus absence or presence of bi-directed edge between X and Z will have no effect on POMISs learned from graph $\mathcal{G}$ in this scenario as well. Combining both of the scenarios when neither of the conditions of (a) and (b) hold, all other bi-directed edges from $\mathcal{G}$, which are absent in $\mathcal{G}'$, can be removed one by one from $\mathcal{G}$ while keeping the POMISs learned from both the intermediate graphs the same. Since $\mathcal{G}$ and $\mathcal{G}$ only differ in bi-directed edges, with bi-directed edges in $\mathcal{G}'$ being a subset of those in $\mathcal{G}$, eventually both graphs will become identical, which proves the statement: Two graphs $\mathcal{G}$ and $\mathcal{G}'$ will have the same POMISs if neither of the statements (a) or (b) hold true. This completes the proof of the Theorem 3.1.

A.5 Proof of Lemma 4.1:

Consider a causal graph $\mathcal{G}(\mathbf{V}, \mathbf{E})$ and $\mathbf{W} \subseteq \mathbf{V}$. Furthermore, let $X, T \in \mathbf{V} \setminus \mathbf{W}$ be any two variables. Fix some realization $\mathbf{w} \in [K]^{|\mathbf{W}|}$. Under post interventional faithfulness Assumption 2.1 we want to prove: $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{w}}}} \iff P(t|do(\mathbf{w})) \neq P(t|do(\mathbf{w}), do(x))$ for some $x, t \in [K]$.

Forward Direction ($\Longrightarrow$): $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{w}}}} \implies P(t|do(\mathbf{w})) \neq P(t|do(\mathbf{w}), do(x))$ for some $x, t \in [K]$. By contradiction, assume $P(t|do(\mathbf{w})) = P(t|do(\mathbf{w}), do(x)), \forall x, t \in [K]$. This implies that $P(t|do(\mathbf{w}), do(x)) = P(t|do(\mathbf{w})) =$ some function of only t and $\mathbf{w}$. This implies that for the sub-model $\mathcal{M}_{\mathbf{W}=\mathbf{w}, X}$ the following CI statements holds: $(T \perp\!\!\!\perp X)_{\mathcal{M}_{\mathbf{W}=\mathbf{w}, X}}$. However, note that if $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{w}}}}$, then we still have $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{W},X}}}$. This implies there is a directed path from X to T in the post-interventional graph $\mathcal{G}_{\overline{\mathbf{W},X}}$. Therefore, we have: $(T \not\perp\!\!\!\perp_d X)_{\mathcal{G}_{\overline{\mathbf{W},X}}}$. Note that under the post interventional faithfulness Assumption 2.1, the CI statement $(T \perp\!\!\!\perp X)_{\mathcal{M}_{\mathbf{W}=\mathbf{w}, X}}$ can hold only if the d-separation statement holds $(T \perp\!\!\!\perp_d X)_{\mathcal{G}_{\overline{\mathbf{W},X}}}$, which is clearly a contradiction. This completes the proof for the forward direction.

Reverse Direction ($\Longleftarrow$): $(X \in \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{w}}}} \iff P(t|do(\mathbf{w})) \neq P(t|do(\mathbf{w}), do(x))$ for some $x, t \in [K]$. We prove the contrapositive statement instead, i.e., $(X \notin \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{w}}}} \implies P(t|do(\mathbf{w})) = P(t|do(\mathbf{w}), do(x)), \forall x, t \in [K]$. Note that $(X \notin \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{w}}}}$ clearly implies that $(X \notin \mathsf{An}(T))_{\mathcal{G}_{\overline{\mathbf{W},X}}}$ which implies that $(T \perp\!\!\!\perp_d X)_{\mathcal{G}_{\overline{\mathbf{W},X}}}$. Thus, using Rule 3 of Pearl's do calculus, we have: $P(t|do(\mathbf{w}), do(x)) = P(t|do(\mathbf{w})), \forall x, t \in [K]$. This completes the proof of the reverse direction.

A.6 Proof of Lemma 4.2:

Consider two variables X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$ and a set of variables $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$. Fix some realization $\mathbf{w} \in [K]^{|\mathbf{W}|}$. Under the post-interventional faithfulness Assumption 2.1 we want to show that: There is latent confounder between X_i and X_j $\iff P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) \neq P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w}))$ for some realization $x_i, x_j \in [K]$.

Forward Direction ($\Longrightarrow$): There is latent confounder between X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$ $\implies P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) \neq P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w}))$ for some realization $x_i, x_j \in [K]$. By contradiction assume $P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) = P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w})) \forall x_i, x_j \in [K]$. Recall that: $X_j = f_j(\mathsf{Pa}(X_j), \mathbf{U_j})$. Since there is latent confounder between X_i and X_j call it L_{ij}. Also note that $L_{ij} \in \mathbf{U_j}$. Define $\mathbf{U'_j} := \mathbf{U_j} \setminus \{L_{i,j}\}$

$$P(x_j \mid do(x_i), do(\mathbf{W})) = P(x_j \mid do(x_i), do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}))) \tag{5}$$

where the interventions $do(\mathsf{Pa}(X_i))$ and $do(\mathsf{Pa}(X_j)))$ are consistent with $do(x_i)$ and $do(\mathbf{W} = \mathbf{w})$. Note that there is also an application of Pearl's do-calculus Rule 3 because by definition of set $\mathbf{W}$ we have $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$. All the extra intervention targets can simply be deleted, and we are left with intervention on X_i, $\mathsf{Pa}(X_i)$, and $\mathsf{Pa}(X_j)$.

$$P(x_j \mid do(x_i), do(\mathbf{W})) = \sum_{\mathbf{u'_j},\, l_{i,j}} P(x_j \mid do(x_i), do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij})$$

$$\times\, P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij}) \tag{6}$$

We have another application of Pearl's do-calculus Rule 3 because interventions on observed variables don't affect unobserved variables, as there are no causal/directed paths from observed to unobserved variables. Also we have:

$$P(x_j \mid x_i, do(\mathbf{W})) = P(x_j \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}))) \tag{7}$$

Note that there is also an application of Pearl's do-calculus Rule 3 because by definition of set $\mathbf{W}$ we have $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$. All the extra intervention targets can simply be deleted, and we are left with conditioning on $X_i = x_i$ and interventions on $\mathsf{Pa}(X_i)$ and $\mathsf{Pa}(X_j)$.

$$P(x_j \mid x_i, do(\mathbf{W})) = \sum_{\mathbf{u'_j},\, l_{i,j}} P(x_j \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij})$$

$$\times\, P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij} \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\})) \tag{8}$$

Using Pearl's do-calculus Rule 2, we can replace the conditioning $X_i = x_i$ with the intervention $do(x_i)$ in $P(x_j \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij})$ because $X_j \notin \mathsf{An}(X_i)$ and $\mathsf{Pa}(X_i)$ are already intervened on. Also, the latent confounder L_{ij} is conditioned on, so there is no open backdoor path from X_i to X_j. Thus, we have:

$$P(x_j \mid x_i, do(\mathbf{W})) = \sum_{\mathbf{u'_j},\, l_{i,j}} P(x_j \mid do(x_i), do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij})$$

$$\times\, P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij} \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\})) \tag{9}$$

From the Equations 6 and 9 and assumption $P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) = P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w}))$ $\forall x_i, x_j \in [K]$ we have:

$$\sum_{\mathbf{u'_j},\, l_{i,j}} P(x_j \mid do(x_i), do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij})$$

$$\times\, (P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij} \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\})) - (P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij})) = 0 \tag{10}$$

Since probabilities are positive, whenever $P(x_j \mid do(x_i), do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij}) \geq 0$, we must have:

$$P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij} \mid x_i, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\})) = P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij}). \tag{11}$$

However, since we know that L_{ij} is a confounder between X_i and X_j, we have an edge $L_{ij} \to X_i$ in the causal graph, which implies that under any intervention $do(\mathbf{Z})$ such that $X_i \notin \mathbf{Z}$, we must have $(L_{ij} \not\perp\!\!\!\perp X_i)_{\mathcal{M}_{\mathbf{Z}=\mathbf{z}}}$ by interventional faithfulness Assumption 2.1. This implies that there exists a realization x_i^* and l_{ij}^* such that:

$$P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij}^* \mid x_i^*, do(\mathsf{pa}(X_i)), do(\mathsf{pa}(X_j) \setminus \{x_i\})) \neq P(\mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij}^*) \tag{12}$$

Now, using the combination $do(\mathbf{W} = \mathbf{w})$ and a special choice of realizations x_i^* and l_{ij}^*, we must have at least one special realization x_j^* such that: $P(x_j^* \mid do(x_i^*), do(\mathsf{Pa}(X_i)), do(\mathsf{Pa}(X_j) \setminus \{x_i\}), \mathbf{U'}_j = \mathbf{u'}_j, L_{ij} = l_{ij}^*) > 0$. Combining this with Equations 12 and 10, we conclude for some $x_i^*, x_j^* \in [K]$, we have $P(x_j^* \mid do(x_i^*), do(\mathbf{W} = \mathbf{w})) \neq P(x_j^* \mid x_i^*, do(\mathbf{W} = \mathbf{w}))$. Thus this leads to contradiction. Thus if there is a latent confounder between X_i and $X_j \implies P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) \neq P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w}))$ for some realization $x_i, x_j \in [K]$. This completes the proof of the forward direction.

Reverse Direction ($\impliedby$): For a pair of variables X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$, if $P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) \neq P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w}))$ for some realizations $x_i, x_j \in [K]$, then there is a latent confounder between X_i and X_j. We prove the contrapositive statement instead, i.e., if there is no latent confounder between X_i and X_j, then $P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) = P(x_j \mid x_i, do(\mathbf{W} = \mathbf{w}))$, $\forall x_i, x_j \in [K]$. Note that by construction, we have: $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$. For such choice of set $\mathbf{W}$ and the fact that $X_j \notin \mathsf{An}(X_i)$ and there is no latent confounder between X_i and X_j, we have $(X_j \perp\!\!\!\perp X_i)_{\mathcal{G}_{X_i \overline{\mathbf{W}}}}$. Thus, from Pearl's do-calculus Rule 2, we have $P(x_j \mid do(x_i), do(\mathbf{W} = \mathbf{w})) = P(x_j \mid x_i, do(\overline{\mathbf{W}} = \mathbf{w}))$, $\forall x_i, x_j \in [K]$. This completes the proof of the reverse direction.

A.7 Proof of Theorem 4.1:

We state and prove two lemmas regarding the sampling complexity for running ancestrality tests and sampling complexity for Algorithm 1 to learn the transitive closure of the causal graph under an arbitrary intervention, and then proceed to the proof of Theorem 4.1.

Lemma A.3. *Suppose that the Assumption 4.1 holds and we have access to $max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2K^2}{\delta_1}$ samples from $do(X_i = x_i, \mathbf{W} = \mathbf{w}) \, \forall x_i \in [K]$ and $\frac{8}{\epsilon^2} \log \frac{2K^2}{\delta_2}$ samples from $do(\mathbf{W} = \mathbf{w})$ for a fixed $w \in [K]^{|\mathbf{W}|}$ and $\mathbf{W} \subseteq \mathbf{V}$. Then, with probability at least $1 - \delta_1 - \delta_2$, we have $(X_i \in \mathsf{An}(X_j))_{\mathcal{G}_{\overline{\mathbf{W}}}}$ if and only if $\exists \, x_i, x_j \in [K]$ s.t. $\left| \widehat{P}(x_j \mid do(\boldsymbol{w})) - \widehat{P}(x_j \mid do(\boldsymbol{w}), do(x_i)) \right| > \frac{\epsilon}{2}$.*

Proof: Suppose that Assumption 4.1 holds and we have access to $max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2K^2}{\delta_1}$ samples from $do(X_i = x_i, \mathbf{W} = \mathbf{w}) \, \forall x_i \in [K]$ and $\frac{8}{\epsilon^2} \log \frac{2K^2}{\delta_2}$ samples from $do(\mathbf{W} = \mathbf{w})$ for a fixed $w \in [K]^{|\mathbf{W}|}$ and $\mathbf{W} \subseteq \mathbf{V}$, we want to show that with probability at least $1 - \delta_1 - \delta_2$, we have the following:

$$(X_i \in \mathsf{An}(X_j))_{\mathcal{G}_{\overline{\mathbf{w}}}} \iff \exists \, x_i, x_j \in [K] \text{ s.t. } \left| \widehat{P}(x_j | do(\mathbf{w})) - \widehat{P}(x_j | do(\mathbf{w}), do(x_i)) \right| > \frac{\epsilon}{2}. \quad (13)$$

Using Hoeffding's inequality with A samples from intervention $do(x_i, \mathbf{w})$,

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \geq \sqrt{\frac{1}{2A} \log \frac{2K^2}{\delta_1}} \quad w.p. \text{ at most } \frac{\delta_1}{K^2}. \quad (14)$$

If we choose $A = max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2K^2}{\delta_1}$, we have:

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \geq \frac{\epsilon}{4} \quad w.p. \text{ at most } \frac{\delta_1}{K^2}. \quad (15)$$

Similarly, using Hoeffding's inequality with B samples from intervention $do(\mathbf{w})$,

$$\left| \widehat{P}(x_j | do(\mathbf{w})) - P(x_j | do(\mathbf{w})) \right| \geq \sqrt{\frac{1}{2A} \log \frac{2K^2}{\delta_1}} \quad w.p. \text{ at most } \frac{\delta_2}{K^2}. \quad (16)$$

If we choose $B = \frac{8}{\epsilon^2} \log \frac{2K^2}{\delta_2}$, we have:

$$\left| \widehat{P}(x_j | do(\mathbf{w})) - P(x_j | do(\mathbf{w})) \right| \geq \frac{\epsilon}{4} \quad w.p. \text{ at most } \frac{\delta_2}{K^2}. \quad (17)$$

Since the realization $\mathbf{w} \in [K]^{|\mathbf{W}|}$ is fixed, but $x_i, x_j \in [K]$, we have a total of K^2 possible bad events when estimates are not good. With the given choice of number of samples A and B, we have:

$$\left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w})) \right| \leq \frac{\epsilon}{4} \quad \forall x_i, x_j \in [K] \ \ w.p. \ at \ least \ 1 - \delta_1, \quad (18)$$

$$\left| \widehat{P}(x_j|do(\mathbf{w})) - P(x_j|do(\mathbf{w})) \right| \leq \frac{\epsilon}{4} \quad \forall x_j \in [K] \ \ w.p. \ at \ least \ 1 - \delta_2. \quad (19)$$

Now when both the estimates are accurate which has a probability of at least $1 - \delta_1 - \delta_2$, we consider two possible scenarios. Suppose that $X_i \notin \mathsf{An}(X_j)$ in $\mathcal{G}_{\overline{\mathbf{W}}}$. In this case by Pearl's do-calculus Rule 3 we have $\left| P(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(\mathbf{w})) \right| = 0$, $\forall x_i, x_j \in [K]$. By triangular inequality we have the following:

$$\left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|do(\mathbf{w})) \right| \leq \left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w})) \right| +$$

$$\left| \widehat{P}(x_j|do(\mathbf{w})) - P(x_j|do(\mathbf{w})) \right| \leq \frac{\epsilon}{2} \quad \forall x_i, x_j \in [K]. \quad (20)$$

However, when $X_i \in \mathsf{An}(X_j)$ in $\mathcal{G}_{\overline{\mathbf{W}}}$ under Assumption 4.1 we must have some configuration say $x_i, x_j \in [K]$ for any $\mathbf{w} \in [K]^{|\mathbf{W}|}$ such that $\left| P(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(\mathbf{w})) \right| > \epsilon$. By triangular inequality when $X_i \in \mathsf{An}(X_j)$ in $\mathcal{G}_{\overline{\mathbf{W}}}$, $\exists x_i, x_j \in [K]$ such that

$$\left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|do(\mathbf{w})) \right| \geq \left| P(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(\mathbf{w})) \right|$$

$$- \left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w})) \right| - \left| \widehat{P}(x_j|do(\mathbf{w})) - P(x_j|do(\mathbf{w})) \right| > \frac{\epsilon}{2}. \quad (21)$$

Thus, using Assumption 4.1 with the given choice of number of samples with probability at least $1 - \delta_1 - \delta_2$, we have the following result:

$$(X_i \in \mathsf{An}(X_j))_{\mathcal{G}_{\overline{\mathbf{W}}}} \iff \exists x_i, x_j \in [K] \ s.t. \ \left| \widehat{P}(x_j|do(\mathbf{w})) - \widehat{P}(x_j|do(\mathbf{w}), do(x_i)) \right| > \frac{\epsilon}{2}. \quad (22)$$

This completes the proof for Lemma A.3.

Lemma A.4. *Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, with probability at least $1 - n\delta_1 - \delta_2$ with a maximum $KAn + B$ interventional samples. If we set $\delta_1 = \frac{\delta}{2n}$ and $\delta_2 = \frac{\delta}{2}$, then Algorithm 1 learns true transitive closure with probability at least $1 - \delta$.*

Proof: In order to prove that Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, we recall from the proof of Lemma A.3 that the test for ancestrality works perfectly under the event that the causal effects of the form $P(x_j|do(x_i), do(\mathbf{w}))$ and $P(x_j|do(\mathbf{w}))$ are estimated accurately with an error of at most $\frac{\epsilon}{4}$ for all $x_i, x_j \in [K]$ and any fixed $\mathbf{w} \in [K]^{\mathbf{W}}$. Now, since Algorithm 1 takes $A = \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{2nK^2}{\delta_1}$ samples from $do(\mathbf{W} = \mathbf{w})$ and $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$ samples from every $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ for all $X_i \in \mathbf{V} \setminus \mathbf{W}$ and for all $x_i \in [K]$. The total number of intervention samples collected is clearly at most $KAn + B$. In order to show that Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, with high probability, we have to demonstrate that Algorithm 1 is able to estimate all causal effects with a maximum error of $\frac{\epsilon}{4}$ with high probability so that all the ancestrality tests work perfectly, as implied by the Proof of LemmaA.3.

Using Hoeffding's inequality with $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$ samples from intervention $do(\mathbf{w})$. For a fixed $X_j \in \mathbf{V} \setminus \mathbf{W}$ we have:

$$\left| \widehat{P}(x_j | do(\mathbf{w})) - P(x_j | do(\mathbf{w})) \right| \leq \frac{\epsilon}{4} \;\; \forall x_j \in [K] \;\; w.p. \text{ at least } 1 - \frac{\delta_2}{n}. \tag{23}$$

Using the union bound we have the following:

$$\left| \widehat{P}(X_j = x_j | do(\mathbf{w})) - P(X_j = x_j | do(\mathbf{w})) \right| \leq \frac{\epsilon}{4} \;\; \forall x_j \in [K] \;,\; \forall X_j \in \mathbf{V} \setminus \mathbf{W} \;\; w.p. \text{ at least } 1 - \delta_2. \tag{24}$$

Now, consider a fixed $X_i, X_j \in \mathbf{V} \setminus \mathbf{W}$, and using $A = \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{2nK^2}{\delta_1}$ samples for every $x_i \in [K]$, we have the following using the Hoeffding's inequality:

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \leq \min(\frac{\epsilon}{4}, \frac{\gamma}{4}) \;\; \forall x_i, x_j \in [K] \;\; w.p. \text{ at least } 1 - \frac{\delta_1}{n} \tag{25}$$

Using the union bound we have the following:

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \leq \min(\frac{\epsilon}{4}, \frac{\gamma}{4})$$
$$\forall x_i, x_j \in [K] \;,\; \forall X_j \in \mathbf{V} \setminus (\mathbf{W} \cup \{X_i\}) \;\; w.p. \text{ at least } 1 - \delta_1 \tag{26}$$

Again using the union bound over all intervention targets $X_i \in \mathbf{V}$ we have the following:

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \leq \min(\frac{\epsilon}{4}, \frac{\gamma}{4})$$
$$\forall x_i, x_j \in [K] \;,\; \forall X_i \in \mathbf{V} \setminus \mathbf{W} \;,\; \forall X_j \in \mathbf{V} \setminus (\mathbf{W} \cup \{X_i\}) \;\; w.p. \text{ at least } 1 - n\delta_1 \tag{27}$$

From Equations 24 and 27, using the union bound with probability at least $1 - n\delta_1 - \delta_2$, all the causal effects are estimated within an error of $\frac{\epsilon}{4}$ from the true values, ensuring that all ancestrality tests work perfectly under this good event. Thus, Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, with $KAn + B$ intervention samples with probability of at least $1 - n\delta_1 - \delta_2$. Also, if we set $\delta_1 = \frac{\delta}{2n}$ and $\delta_2 = \frac{\delta}{2}$, then Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$ with a probability of $1 - \delta$, with $KAn + B$ samples, where $A = \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{4n^2 K^2}{\delta}$ and $B = \frac{8}{\epsilon^2} \log \frac{4nK^2}{\delta}$. This completes the proof of Lemma A.4.

In order to prove Theorem 4.1, we revisit the statement of Lemma A.3: Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, with $KAn + B$ intervention samples with a probability of at least $1 - n\delta_1 - \delta_2$. Algorithm 2 randomly samples a target set $\mathbf{W}$ and calls Algorithm 1 to learn the active true transitive closure of the post-interventional graph, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$. For every iteration, Algorithm 2 computes transitive reduction $Tr(\mathcal{G}_{\overline{\mathbf{W}}}^{tc})$ and updates all the edges to construct the observable graph. To prove the results in Theorem 4.1, we rely on Lemma 5 from [13], which is stated below:

Lemma A.5. *[13] Consider a graph $\mathcal{G}$ with observed variables $\mathbf{V}$ and an intervention set $\mathbf{W} \subseteq \mathbf{V}$. Consider post-interventional observable graph $\mathcal{G}_{\overline{\mathbf{W}}}$ and a variable $X_j \in \mathbf{V} \setminus \mathbf{W}$. Let $X_i \in \mathsf{Pa}(X_j)$ be such that all the parents of X_j above X_i in partial order are included in the intervention set $\mathbf{W}$. This implies that $\{W_i : \pi(W_i) > \pi(X_i) \;\&\; W_i \in \mathsf{Pa}(X_j)\} \subseteq \mathbf{W}$. Then, the directed*

edge $(X_i, X_j) \in \mathbf{E}(Tr(\mathcal{G}_{\overline{\mathbf{W}}}))$. *The properties of transitive reduction yields* $Tr(\mathcal{G}_{\overline{\mathbf{W}}}) = Tr(\mathcal{G}_{\overline{\mathbf{W}}}^{tc})$. *Consequently, the transitive reduction of* $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, *i.e.,* $Tr(\mathcal{G}_{\overline{\mathbf{W}}}^{tc}) = Tr(\mathcal{G}_{\overline{\mathbf{W}}})$ *may be used to learn the directed edge* (X_i, X_j).
(Note: $\mathbf{E}(\mathbf{G})$ *denotes the edges of the graph* $\mathcal{G}$ *and* π *is any total order that is consistent with the partial order implied by the DAG, i.e.,* $\pi(X) < \pi(Y)$ *iff X is an ancestor of Y).*

Assume that the number of the direct parents of X_j above X_i is d_{ij} where $d_{ij} \leq d_{max}$. Let $\mathcal{E}_i(X_j)$ be the following event: $X_i, X_j \notin \mathbf{W}$ & $\{W_i : \pi(W_i) > \pi(X_i)$ & $W_i \in \mathsf{Pa}(X_j)\} \subseteq \mathbf{W}$. The probability of this event for one run of the outer loop in Algorithm 2 with the assumption that $2d_{max} >= 2$ is given by:

$$P[\mathcal{E}_i(X_j)] = \frac{1}{4d_{max}^2}(1 - \frac{1}{2d_{max}})^{d_{ij}} \geq \frac{1}{4d_{max}^2}(1 - \frac{1}{2d_{max}})^{2d_{max}} \geq \frac{1}{d_{max}^2}\frac{1}{16}. \quad (28)$$

The last inequality holds for $2d_{max} >= 2$ because $(1 - \frac{1}{x})^x \geq 0.25, \quad \forall x \geq 2$. Based on Lemma A.5, the event $\mathcal{E}_i(X_j)$ implies that the directed edge (X_j, X_j) will be present in $Tr(\mathcal{G}_{\overline{\mathbf{W}}}^{tc})$ and will be learned. The outer loop runs for $8\alpha d_{max}\log(n)$ iterations and elements of the set $\mathbf{W}$ are independently sampled. The probability of failure, i.e., the event under consideration does not happen for all runs of the outer loop in Algorithm 2, is bounded as follows:

$$P[(\mathcal{E}_i(V))^c] \leq (1 - \frac{1}{16\,d_{max}^2})^{8\alpha d_{max}\log(n)} \leq e^{-\frac{\alpha}{2d_{max}}\log(n)} = \frac{1}{n^{\frac{\alpha}{2d_{max}}}}. \quad (29)$$

For a graph with a total number of variables n, the total number of such bad events will be $\binom{n}{2}$ since a graph can have at most $\binom{n}{2}$ edges. Using union bounding, the probability of bad events for every pair of variables is given by:

$$P[Failure] \leq \binom{n}{2} \times \frac{1}{n^{\frac{\alpha}{2d_{max}}}} \leq \frac{1}{n^{\frac{\alpha}{2d_{max}}-2}}. \quad (30)$$

Under the event the Algorithm 1 learns the correct transitive closure $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$ for all the $8\alpha d_{max}\log n$ randomly sampled intervention sets $\mathbf{W} \subseteq \mathbf{V}$, above derivations show that we will be able to learn all the edges in the true observable with a probability of at least $1 - \frac{1}{n^{\frac{\alpha}{2d_{max}}-2}}$.

Now recall the results in Lemma A.3: Algorithm 1 learns the true transitive closure under any intervention, i.e., $\mathcal{G}_{\overline{\mathbf{W}}}^{tc}$, with $KAn + B$ intervention samples with a probability of at least $1 - n\delta_1 - \delta_2$. Combining the two results above using the union bound, we have the following series result:

Algorithm 2 learns the true observable graph with a probability of at least $1 - \frac{1}{n^{\frac{\alpha}{2d_{max}}-2}} - 8\alpha d_{max}\log(n)(n\delta_1 + \delta_2)$ with a maximum $8\alpha d_{max}\log n(KAn + B)$ interventional samples. Also, if we set $\alpha = \frac{2d_{max}\log(\frac{2}{\delta}+2)}{\log n}$, $\delta_1 = \frac{\delta}{32\alpha d_{max}n\log n}$, and $\delta_2 = \frac{\delta}{32\alpha d_{max}\log n}$, then Algorithm 2 learns the true observable graph with a probability of at least $1 - \delta$. Where $A = \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right)\log\frac{2nK^2}{\delta_1}$ and $B = \frac{8}{\epsilon^2}\log\frac{2nK^2}{\delta_2}$. This completes the proof of Theorem 4.1.

A.8 Proof of Lemma 4.3:

Consider two nodes X_i and X_j s.t. $X_j \notin \mathsf{An}(X_i)$ and suppose that Assumptions 4.2 4.3 holds and we have access to $max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2})\log\frac{2K^2}{\delta_1}$ samples from $do(X_i = x_i, \mathbf{W} = \mathbf{w}) \,\forall x_i \in [K]$ and $\frac{16}{\eta\gamma^2}\log(\frac{2K^2}{\delta_3}) + \frac{1}{2\eta^2}\log(\frac{2K^2}{\delta_4})$ from $do(\mathbf{W} = \mathbf{w})$ for a fixed $w \in [K]^{|\mathbf{W}|}$ and $\mathbf{W} \subseteq \mathbf{V}$ such that $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$. Then with probability at least $1 - \delta_1 - \delta_3 - \delta_4$, we want to show that:

There exists a latent confounder between X_i and $X_j \iff$

$$\exists\, x_i, x_j \in [K] \; s.t. \; \big| \, \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w})) \, \big| > \frac{\gamma}{2}. \quad (31)$$

Using Hoeffding's inequality with A samples from intervention $do(x_i, \mathbf{w})$.

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \geq \sqrt{\frac{1}{2A} \log \frac{2K^2}{\delta_1}} \quad w.p. \ at \ most \ \frac{\delta_1}{K^2}. \quad (32)$$

If we choose $A = max(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}) \log \frac{2K^2}{\delta_1}$, we have:

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \geq \frac{\gamma}{4} \quad w.p. \ at \ most \ \frac{\delta_1}{K^2}. \quad (33)$$

Using Hoeffding's inequality with C samples from intervention $do(x_i)$.

$$\left| \widehat{P}(x_j | x_i, do(\mathbf{w})) - P(x_j | x_i, do(\mathbf{w})) \right| \geq \sqrt{\frac{1}{2C_{x_i}} \log \frac{2K^2}{\delta_3}} \quad w.p. \ at \ most \ \frac{\delta_3}{K^2}. \quad (34)$$

Where C_{x_i} is the number of samples where $X_i = x_i$ among the C samples for the intervention $do(\mathbf{w})$. Note the we can't directly control C_{x_i} and it's value depends on the true interventions distribution $P(x_i, do(\mathbf{w}))$ along-with the number of samples C. Suppose if we can set $C_{x_i} \geq \frac{8}{\gamma^2} \log \frac{2K^2}{\delta_3}$, we have:

$$\left| \widehat{P}(x_j | x_i, do(\mathbf{w})) - P(x_j | x_i, do(\mathbf{w})) \right| \geq \frac{\gamma}{4} \quad w.p. \ at \ most \ \frac{\delta_3}{K^2}. \quad (35)$$

We need to find the number of samples C such that $C_{x_i} \geq \frac{8}{\gamma^2} \log \frac{2K^2}{\delta_3}$. Using the Hoeffding's bound we have:

$$P(C_{x_i} \geq CP(x_i | do(\mathbf{w})) - \eta) \geq 1 - 2e^{-2\eta^2/C}. \quad (36)$$

Let $\frac{\delta_4}{K^2} = 2e^{-2\eta^2/C}$, which implies $\eta = \sqrt{\frac{C}{2} \log \frac{2K^2}{\delta_4}}$. Thus we have:

$$P\left(C_{x_i} \geq CP(x_i | do(\mathbf{w})) - \sqrt{\frac{C}{2} \log \frac{2K^2}{\delta_4}} \right) \geq 1 - \frac{\delta_4}{K^2} \quad (37)$$

$$C_{x_i} \geq CP(x_i | do(\mathbf{w})) - \sqrt{\frac{C}{2} \log \frac{2K^2}{\delta_4}} \quad w.p. \ at \ least \ 1 - \frac{\delta_4}{K^2}. \quad (38)$$

Using Assumption 4.3, we have $P(x_i | do(\mathbf{w})) = 0$ or $P(x_i | do(\mathbf{w})) \geq \eta$. Note that if $P(x_i | do(\mathbf{w})) = 0$, the event will never happen, and we don't care about the accuracy of the estimate $\widehat{P}(x_j | x_i, do(\mathbf{w}))$ because it is already initialized to zero. Now the equation above can be rewritten as:

$$C_{x_i} \geq C\eta - \sqrt{\frac{C}{2} \log \frac{2K^2}{\delta_4}} \quad w.p. \ at \ least \ 1 - \frac{\delta_4}{K^2}. \quad (39)$$

Since we want $C_{x_i} \geq \frac{8}{\gamma^2} \log \frac{2K^2}{\delta_3}$ with high probability, we have the following relationship:

$$C\eta - \sqrt{\frac{C}{2} \log \frac{2K^2}{\delta_4}} \geq \frac{8}{\gamma^2} \log \frac{2K^2}{\delta_3} \quad (40)$$

Solving the equation for number of samples C we get:

$$C \geq \frac{4\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2} + \ln\left(\frac{2K^2}{\delta_4}\right) + \sqrt{8\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2}\ln\left(\frac{2K^2}{\delta_4}\right) + \ln^2\left(\frac{2K^2}{\delta_4}\right)}}{4\eta^2} \tag{41}$$

In order to make the expression simpler we choose the number of samples C as follows:

$$C = \frac{4\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2} + \ln\left(\frac{2K^2}{\delta_4}\right) + \sqrt{8\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2}\ln\left(\frac{2K^2}{\delta_4}\right) + \ln^2\left(\frac{2K^2}{\delta_4}\right) + \left(4\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2}\right)^2}}{4\eta^2} \tag{42}$$

$$C = \frac{4\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2} + \ln\left(\frac{2K^2}{\delta_4}\right) + \sqrt{\left(4\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2} + \ln\left(\frac{2K^2}{\delta_4}\right)\right)^2}}{4\eta^2} \tag{43}$$

$$C = \frac{4\eta\frac{8\log\left(\frac{2K^2}{\delta_3}\right)}{\gamma^2} + \ln\left(\frac{2K^2}{\delta_4}\right)}{2\eta^2} \tag{44}$$

$$C = \frac{16}{\eta\gamma^2}\log(\frac{2K^2}{\delta_3}) + \frac{1}{2\eta^2}\log(\frac{2K^2}{\delta_4}) \tag{45}$$

Suppose we take C samples for intervention $do(\mathbf{w})$ as given above. Now, from Equations 35, 39, and 40, using the union bound, we have the following:

$$\left|\widehat{P}(x_j|x_i, do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w}))\right| \geq \frac{\gamma}{4} \ \ w.p. \ at \ most \ \frac{\delta_3 + \delta_4}{K^2}. \tag{46}$$

Since the realization $\mathbf{w} \in [K]^{|\mathbf{W}|}$ is fixed, but $x_i, x_j \in [K]$, we have a total of K^2 possible bad events when estimates are not good. With the given choice of number of samples A and C, we have:

$$\left|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w}))\right| \leq \frac{\gamma}{4} \ \forall x_j \in [K] \ \ w.p. \ at \ least \ 1 - \delta_1. \tag{47}$$

$$\left|\widehat{P}(x_j|x_i, do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w}))\right| \leq \frac{\gamma}{4} \ \forall x_i, x_j \in [K] \ \ w.p. \ at \ least \ 1 - \delta_3 - \delta_4. \tag{48}$$

Now under the good event when both the estimates are accurate which has a probability of at least $1 - \delta_1 - \delta_3 - \delta_4$, we consider two possible scenarios. Suppose that there is no latent confounder between X_i and X_j. In this case by Lemma 4.2 we have $\left|P(x_j|do(x_i), do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w}))\right| = 0$, $\forall x_i x_j \in [K]$. By triangular inequality we have the following:

$$\left|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w}))\right| \leq \left|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w}))\right| +$$
$$\left|\widehat{P}(x_j|x, do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w}))\right| \leq \frac{\gamma}{2} \ \forall x_i, x_j \in [K]. \tag{49}$$

However, when there is a latent confounder between X_i and X_j, in this case, under Assumption 4.2, we must have some configuration, say $x_i, x_j \in [K]$, for any $\mathbf{w} \in [K]^{|\mathbf{W}|}$, such that

$\left| P(x_j|do(x_i), do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w})) \right| > \gamma$. By triangular inequality when there is a latent confounder between X_i and X_j, $\exists\ x_i, x_j \in [K]$ such that:

$$
\left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w})) \right| \geq \left| P(x_j|do(x_i), do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w})) \right|
$$

$$
- \left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w})) \right| - \left| \widehat{P}(x_j|x_i, do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w})) \right| > \frac{\gamma}{2}
$$

$$(50)$$

Thus, using Assumption 4.2 with the given choice of number of samples with probability at least $1 - \delta_1 - \delta_3 - \delta_4$, we have the following result:

There exists a latent confounder between X_i and X_j $\iff$

$$
\exists\ x_i, x_j \in [K]\ s.t.\ \left| \widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w})) \right| > \frac{\gamma}{2}.
$$

$$(51)$$

This completes the proof for Lemma 4.3.

A.9   Algorithm to learn the causal graph will all the latent confounders

---
Algorithm 5: Learn the Causal Graph along-with the Latent Confounders

---
1 **Function** LearnCausalGraph$(\alpha, d_{max}, \delta_1, \delta_2, \delta_3, \delta_4)$**:**
2 $\quad$ $\mathcal{G}, \mathcal{ID}ata =$ LearnObservableGraph$(\alpha, d_{max}, \delta_1, \delta_2)$
3 $\quad$ $C = \frac{16}{\eta\gamma^2} \log(\frac{2n^2K^2}{\delta_3}) + \frac{1}{2\eta^2} \log(\frac{2n^2K^2}{\delta_4})$, $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$
4 $\quad$ **for** *every pair $X_i, X_j \in V \setminus W$* **do**
5 $\quad\quad$ If $X_j \in \mathsf{An}(X_i)$, swap them.
6 $\quad\quad$ Find interventional data sets $do(\mathbf{W} = \mathbf{w})$ and $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ from $\mathcal{ID}ata$ s.t. $(\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$
7 $\quad\quad$ Get $\max(0, C - B)$ new samples for $do(\mathbf{W} = \mathbf{w})$
8 $\quad\quad$ **if** $\exists\ x_i, x_j \in [K]\ s.t.\ |\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w}))| > \frac{\gamma}{2}$ **then**
9 $\quad\quad\quad$ Add bi-directed edge $X_i \leftrightarrow X_j$ to graph $\mathcal{G}$
10 $\quad$ **return** The Causal Graph with Latent Confounders $\mathcal{G}$
11 **End Function**

Theorem A.1. *Algorithm 5 learns the true causal graph with latents with probability at least* $1 - \frac{2}{n^{\frac{\alpha}{2d_{max}}-2}} - 8\alpha d_{max} \log(n)(n\delta_1 + (\delta_2 + \delta_3 + \delta_4))$ *with a maximum of* $8\alpha d_{max} \log n(KAn + \max(B, C))$ *interventional samples. If we set* $\alpha = \frac{2d_{max} \log(\frac{4}{\delta}+2)}{\log n}$, $\delta_1 = \frac{\delta}{64\alpha d_{max} n \log n}$ *and* $\delta_2 = \delta_3 = \delta_4 = \frac{\delta}{64\alpha d_{max} \log n}$, *then Algorithm 5 learns the true causal graph with probability at least* $1 - \delta$. *(A and B are given by line 2 of Algorithm 1 and C is given by line 3 of Algorithm 5.)*

Theorem A.1 gives the sample complexity guarantee for Algorithm 5 to learn the true causal graph, including all latent confounders, with a given confidence level. The Algorithm 5 first calls Algorithm 2 to learn the observable graph structure. We have already proved in Theorem 4.1 that Algorithm 2 learns the true observable graph with a probability of at least $1 - \frac{1}{n^{\frac{\alpha}{2d_{max}}-2}} - 8\alpha d_{max} \log(n)(n\delta_1 + \delta_2)$ with a maximum of $8\alpha d_{max} \log n(KAn + B)$ interventional samples. The next phase in Algorithm 5 is to learn/detect latent confounders between any pair of variables. For all pairs of nodes X_i and X_j such that $X_j \notin \mathsf{An}(X_i)$, we define a set of nodes $\mathbf{W}_{ij} \subseteq \mathbf{V}$ such that $X_i, X_j \notin \mathbf{S}_i$, where $\mathbf{W}_{ij} = (\mathsf{Pa}(X_i) \cup \mathsf{Pa}(X_j) \setminus \{X_i\})$. Also, note that $|\mathbf{W}_{ij}| \leq 2d_{max}$. Let us define the event $\mathcal{E}_{ij} = [\mathbf{W}_{ij} \subseteq \mathbf{W}\ \&\ X_j, X_i \notin \mathbf{W}]$. The probability of this event for one run of the outer loop in Algorithm 2 with the assumption that $2d_{max} \geq 2$ is given by:

$$P[\mathcal{E}_{ij}] = \frac{1}{4d_{max}^2}(1 - \frac{1}{2d_{max}})^{|\mathbf{W}_{ij}|} \geq \frac{1}{4d_{max}^2}(1 - \frac{1}{2d_{max}})^{2d_{max}} \geq \frac{1}{d_{max}^2}\frac{1}{16}. \tag{52}$$

The last inequality holds for $d_{\max} \geq 2$. Note that we reuse all the interventional data samples from Algorithm 2 in Algorithm 5. Under Assumption 4.2, if the event $\mathcal{E}_{ij}$ happens with a large enough number of samples, we can detect the presence or absence of latent confounders between X_i and X_j. The outer loop runs for $8\alpha d_{max} \log(n)$ iterations, and the elements of the set $\mathbf{W}$ are independently sampled. The probability of failure, i.e., the event under consideration does not happen for all runs of the outer loop in Algorithm 2, is bounded as follows:

$$P[\mathcal{E}_{ij}^c] \leq (1 - \frac{1}{16\,d_{max}^2})^{8\alpha d_{max} \log(n)} \leq e^{-\frac{\alpha}{2d_{max}}\log(n)} = \frac{1}{n^{\frac{\alpha}{2d_{max}}}}. \tag{53}$$

For a graph with a total number of variables n, the total number of such bad events will be $\binom{n}{2}$ since a graph can have at most $\binom{n}{2}$ edges. Using union bounding, the probability of bad events for every pair of variables is given by:

$$P[Failure] \leq \binom{n}{2} \times \frac{1}{n^{\frac{\alpha}{2d_{max}}}} \leq \frac{1}{n^{\frac{\alpha}{2d_{max}}-2}}. \tag{54}$$

This implies with a probability of $1 - \frac{1}{n^{\frac{\alpha}{2d_{\max}}-2}}$, we will be able to find an appropriate interventional dataset to test the presence of latent confounders between any pair of variables using Assumption 4.2 after running Algorithm 2. We still need to make sure we have enough interventional samples to be able to test the latents. This is because we need to accurately estimate conditional effects to carry out the test, as in Assumption 4.2. We first consider estimation of the causal effect $\widehat{P}(x_j|do(x_i), do(\mathbf{w}))$ for any randomly sampled set $\mathbf{W}$. Now, consider a fixed $X_i, X_j \in \mathbf{V} \setminus \mathbf{W}$. We have access to $\max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right)\log \frac{2nK^2}{\delta_1}$ samples for every $x_i \in [K]$. We have already shown that under the good event, we have the following:

$$\left|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - P(x_j|do(x_i), do(\mathbf{w}))\right| \leq \min(\frac{\epsilon}{4}, \frac{\gamma}{4})$$
$$\forall x_i, x_j \in [K]\,, \ \forall X_i \in \mathbf{V} \setminus \mathbf{W}\,, \ \forall X_j \in \mathbf{V} \setminus (\mathbf{W} \cup \{X_i\}) \ \ w.p. \ at \ least \ 1 - n\delta_1 \tag{55}$$

Now, we consider estimation of the conditional causal effects, i.e., $\widehat{P}(x_j|x_i, do(\mathbf{w}))$. Note the while running the Algorithm 2 we have access to $B = \frac{8}{\epsilon^2}\log\frac{2nK^2}{\delta_2}$ samples form intervention $do(\mathbf{w})$ and in the step 7 of Algorithm 5 we add more samples to the data set and have access to at least $C = \frac{16}{\eta\gamma^2}\log(\frac{2n^2K^2}{\delta_3}) + \frac{1}{2\eta^2}\log(\frac{2n^2K^2}{\delta_4})$ samples instead. Now, consider a fixed $X_i, X_j \in \mathbf{V} \setminus \mathbf{W}$. With access to C samples as given above, following from Equation 48 in the Proof of Lemma 4.3, we have the following result:

$$\left|\widehat{P}(x_j|x_i, do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w}))\right| \leq \frac{\gamma}{4} \ \ \forall x_i, x_j \in [K] \ \ w.p. \ at \ least \ 1 - \frac{\delta_3}{n^2} - \frac{\delta_4}{n^2}. \tag{56}$$

Note that in the above equation, we have $\frac{\delta_3}{n^2}$ and $\frac{\delta_4}{n^2}$ instead of δ_3 and δ_4 as in Equation 48, because here in the number of samples C, we also have $\frac{\delta_3}{n^2}$ and $\frac{\delta_4}{n^2}$ instead of δ_3 and δ_4 when compared to the number of samples in Equation 45. Now, using the union bound we have the following:

$$\left|\widehat{P}(x_j|x_i, do(\mathbf{w})) - P(x_j|x_i, do(\mathbf{w}))\right| \leq \frac{\gamma}{4}$$
$$\forall x_i, x_j \in [K]\,, \ \forall X_j \in \mathbf{V} \setminus (\mathbf{W} \cup \{X_i\}) \ \ w.p. \ at \ least \ 1 - \frac{\delta_3}{n} - \frac{\delta_4}{n}. \tag{57}$$

Again using the union bound over all $X_i \in \mathbf{V} \setminus \mathbf{W}$ we have the following:

$$\left| \widehat{P}(x_j | do(x_i), do(\mathbf{w})) - P(x_j | do(x_i), do(\mathbf{w})) \right| \leq \frac{\gamma}{4}$$

$$\forall x_i, x_j \in [K], \ \forall X_i \in \mathbf{V} \setminus \mathbf{W}, \ \forall X_j \in \mathbf{V} \setminus (\mathbf{W} \cup \{X_i\}) \ \ w.p. \ at \ least \ 1 - \delta_3 - \delta_4 \quad (58)$$

This implies that under the good event, for every randomly sampled intervention set $\mathbf{W} \subseteq \mathbf{V}$, the estimate of the conditional causal effect is accurate within the desired $\frac{\gamma}{4}$ threshold. This would imply that the test for detection of latent variables is perfect under this good event. We have already shown that to ensure we have access to sufficient datasets to detect latent variables between any pair of nodes, the $8\alpha d_{\max} \log n$ randomly sampled target sets in Algorithm 2 $\mathbf{W}$ are sufficient. Combining these results with the results from Theorem 4.1, we have the following:

The Algorithm 5 learns the true causal graph along with all latents with a probability of at least $1 - \frac{1}{n^{\frac{\alpha}{2d_{max}} - 2}} - \frac{1}{n^{\frac{\alpha}{2d_{max}} - 2}} - 8\alpha d_{max} \log(n)(n\delta_1 + \delta_2) - 8\alpha d_{max} \log(n)(\delta_3 + \delta_4) = 1 - \frac{2}{n^{\frac{\alpha}{2d_{max}} - 2}} - 8\alpha d_{max} \log(n)(n\delta_1 + (\delta_2 + \delta_3 + \delta_4))$ with a maximum $8\alpha d_{max} \log n(KAn + \max(B, C))$ interventional samples. Also If we set $\alpha = \frac{2d_{max} \log(\frac{4}{\delta} + 2)}{\log n}$, $\delta_1 = \frac{\delta}{64\alpha d_{max} n \log n}$ and $\delta_2 = \delta_3 = \delta_4 = \frac{\delta}{64\alpha d_{max} \log n}$, then Algorithm 2 learns the true causal graph with all latents with a probability of at least $1 - \delta$. We have:

$$A = \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{2nK^2}{\delta_1}, \ B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}, \quad (59)$$

$$C = \frac{16}{\eta\gamma^2} \log\left(\frac{2K^2}{\delta_3}\right) + \frac{1}{2\eta^2} \log\left(\frac{2K^2}{\delta_4}\right). \quad (60)$$

This completes the proof for Theorem A.1.

A.10 Full Version of The Algorithm 3 and Proof of Theorem 5.1

Algorithm 3 or its full version (Algorithm 6) starts by learning the transitive closure of the graph, denoted as $\mathcal{G}^{tc}$. This is because $\mathcal{G}^{tc}$ can give us $\mathsf{An}(Y)$, and every possible POMIS is a subset of $\mathsf{An}(Y)$. Thus, we can restrict ourselves to ancestors of the read node. From Lemma $A.4$, we can learn the transitive closure $\mathcal{G}^{tc}$ with a probability of at least $1 - \delta$ with a maximum of $KAn + B$ interventional samples by setting $\delta_1 = \frac{\delta}{2n}$ and $\delta_2 = \frac{\delta}{2}$. Then, Algorithm 1 learns the true transitive closure with a probability of at least $1 - \delta$. (We have $A = \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{2nK^2}{\delta_1}$ and $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$ as in line 2 of Algorithm 1). Thus, the total interventional samples for this step turn out to be: $Kn \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{4n^2 K^2}{\delta} + \frac{8}{\epsilon^2} \log \frac{4nK^2}{\delta}$.

The next step is to learn the complete observable graph induced on the reward node and its ancestors and then learn/detect only a subset of latent confounders which are characterized to be necessary and sufficient to learn the true set of POMISs (Theorem 3.1). Although this step saves us interventional samples compared to the full discovery Algorithm 5, which learns/detects latents between all pairs of variables, the exact saving will depend on the structure of the underlying causal graph.

For the regret upper bound, we can use the results from Theorem A.1 to bound the number of interventional samples for learning the true POMIS set from the ancestors of the reward node. This implies that given the true set of ancestors of the reward $\mathsf{An}(Y)$, we can learn the true POMIS set with a probability of at least $1 - \delta$ using $8\alpha d_{\max}\left(KA|\mathsf{An}(Y)| + B\right) \log\left(|\mathsf{An}(Y)|\right)$ interventions, where A and B are given by line 2 of Algorithm 1, and C is given by line 3 of Algorithm 5 by setting $\alpha = \frac{2d_{\max} \log(\frac{4}{\delta} + 2)}{\log |\mathsf{An}(Y)|}$, $\delta_1 = \frac{\delta}{64\alpha d_{\max} |\mathsf{An}(Y)| \log |\mathsf{An}(Y)|}$, and $\delta_2 = \delta_3 = \delta_4 = \frac{\delta}{64\alpha d_{\max} \log |\mathsf{An}(Y)|}$.

24

Algorithm 6: Full version of Algorithm for causal bandits with unknown graph structure

---

1 Set the Parameter δ, d_{max}
2 Calculate $\alpha, \delta_1, \delta_2, \delta_3, \delta_4$ as in Theorem 5.1
3 $\mathcal{G}^{tc}$ = LearnTransitiveClosure($\mathbf{W} = \phi, \frac{\delta}{2n}, \frac{\delta}{n}$)
4 $\mathcal{G}, \mathcal{ID}ata$ = LearnObservableGraph(An$(Y)_{\mathcal{G}^{tc}}, \alpha, d_{max}, \delta_1, \delta_2$)
5 $C = \frac{16}{\eta\gamma^2} \log(\frac{2n^2K^2}{\delta_3}) + \frac{1}{2\eta^2} \log(\frac{2n^2K^2}{\delta_4})$, $B = \frac{8}{\epsilon^2} \log \frac{2nK^2}{\delta_2}$
6 #Learn the bi-directed edges between reward Y and all nodes $X_i \in$ An(Y) and update $\mathcal{G}$.
7 **for** *every* $X_i \in$ An$(Y)_{\mathcal{G}^{tc}}$ **do**
8 $\quad$ Set $X_j := Y$
9 $\quad$ Find interventional data sets $do(\mathbf{W} = \mathbf{w})$ and $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ from $\mathcal{ID}ata$ s.t.
   $\quad$ (Pa$(X_i) \cup$ Pa$(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$
10 $\quad$ Get $\max(0, B - C)$ new samples for $do(\mathbf{W} = \mathbf{w})$
11 $\quad$ **if** $\exists\, x_i, x_j \in [K]$ *s.t.* $|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w}))| > \frac{\gamma}{2}$ **then**
12 $\quad\quad$ Add bi-dirceted edge $X_i \leftrightarrow X_j$ to graph $\mathcal{G}$

13 **while** *There is a new pair that is tested* **do**
14 $\quad$ Find a new pair (Z, X) s.t. $Z \in$ An(Y) such that Z and Y don't have a bi-directed edge
   $\quad\quad$ between them in $\mathcal{G}$ and $X \in$ MUCT$(\mathcal{G}_{\overline{\mathsf{Pa}(Z), \mathsf{Bi}(Z, \mathcal{G})}}, Y)$
15 $\quad$ # test for the latent between the pair (Z, X) and update $\mathcal{G}$.
16 $\quad$ Set $X_i := Z, X_j := X$
17 $\quad$ if $X_j \in$ An(X_i) swap them.
18 $\quad$ Find interventional data sets $do(\mathbf{W} = \mathbf{w})$ and $do(X_i = x_i, \mathbf{W} = \mathbf{w})$ from $\mathcal{ID}ata$ s.t.
   $\quad\quad$ (Pa$(X_i) \cup$ Pa$(X_j) \setminus \{X_i\}) \subseteq \mathbf{W}$ and X_i & $X_j \notin \mathbf{W}$
19 $\quad$ Get $\max(0, B - C)$ new samples for $do(\mathbf{W} = \mathbf{w})$
20 $\quad$ **if** $\exists\, x_i, x_j \in [K]$ *s.t.* $|\widehat{P}(x_j|do(x_i), do(\mathbf{w})) - \widehat{P}(x_j|x_i, do(\mathbf{w}))| > \frac{\gamma}{2}$ **then**
21 $\quad\quad$ Add bi-directed edge $X_i \leftrightarrow X_j$ to graph $\mathcal{G}$

22 Learn the set of POMISs $\mathcal{I}_{\mathcal{G}}$ from the graph $\mathcal{G}$ Using Algorithm 1 in [5].
23 Run UCB algorithm over the arm set $\mathcal{A} = \{\Omega(I) \mid \forall I \in \mathcal{I}_{\mathcal{G}}\}$.

The last phase is just running the UCB algorithm over the set of all possibly optimal arms, i.e., $\mathcal{A} = \{\Omega(I) \mid \forall I \in \mathcal{I}_{\mathcal{G}}\}$. This phase has a regret bound of $\sum_{\mathbf{s} \in \{\Omega(I)|\forall I \in \mathcal{I}_{\mathcal{G}}\}} \Delta_{do(\mathbf{s})}\left(1 + \frac{\log T}{\Delta^2_{do(\mathbf{s})}}\right)$ [22]. Now combining all the results we have the following:

Algorithm 3 learns the true set of POMISs $\mathcal{I}_{\mathcal{G}}$ with probability at least $1 - \delta - \delta = 1 - 2\delta$, and under the good event E that it learns POMISs correctly, the cumulative regret is bounded as follows:

$$R_t \leq Kn \max\left(\frac{8}{\epsilon^2}, \frac{8}{\gamma^2}\right) \log \frac{4n^2K^2}{\delta} + \frac{8}{\epsilon^2} \log \frac{4nK^2}{\delta}$$

$$\tag{61}$$

$$+ 8\alpha d_{max}\left(KA\big|\mathsf{An}(Y)\big| + \max(B, C)\right) \log\left(\big|\mathsf{An}(Y)\big|\right) + \sum_{\mathbf{s} \in \{\Omega(I)|\forall I \in \mathcal{I}_{\mathcal{G}}\}} \Delta_{do(\mathbf{s})}\left(1 + \frac{\log T}{\Delta^2_{do(\mathbf{s})}}\right),$$

where A and B are given by line 2 of Algorithm 1, and C is given by line 3 of Algorithm 5 by setting $\alpha = \frac{2d_{max} \log\left(\frac{4}{\delta} + 2\right)}{\log\big|\mathsf{An}(Y)\big|}$, $\delta_1 = \frac{\delta}{64\alpha d_{max}\big|\mathsf{An}(Y)\big| \log\big|\mathsf{An}(Y)\big|}$ and $\delta_2 = \delta_3 = \delta_4 = \frac{\delta}{64\alpha d_{max} \log\big|\mathsf{An}(Y)\big|}$. This completes the proof of the Theorem 5.1

A.11 Additional simulations to demonstrate the advantage of Algorithm 3 over full discovery

We present additional results to demonstrate that partial discovery by Algorithm 3 to construct POMISs saves us interventional samples compared to full discovery, i.e., learning all possible latents for randomly sampled causal graphs with a variety of density parameters ρ and ρ_L. The procedure to sample the graphs is the same as described in the main paper. Figure 5 shows that for a wide range of
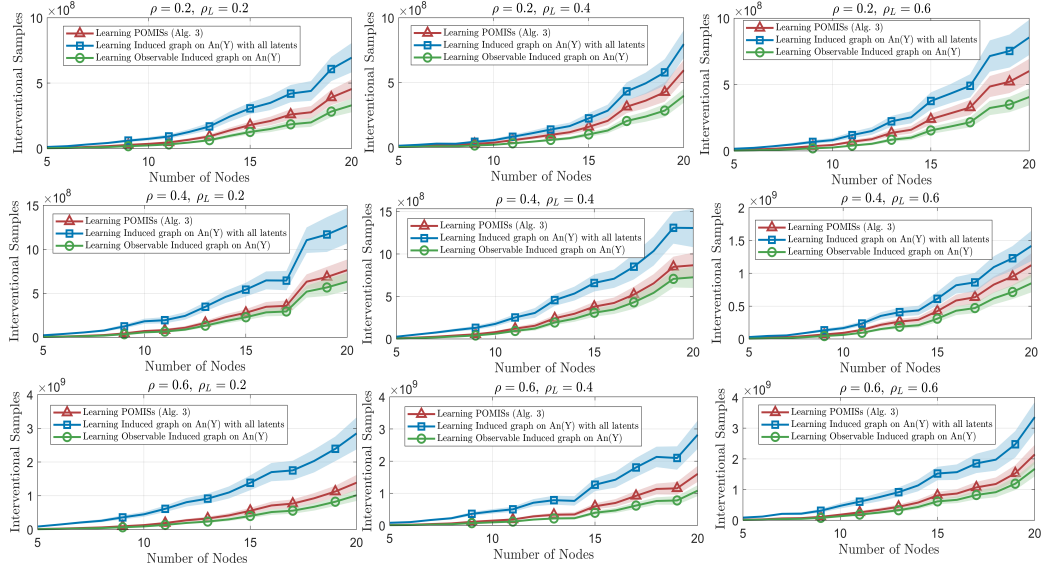
Figure 5: Additional simulations to demonstrate the advantage of Algorithm 3 over full graph discovery (Learning all possible latents)

density parameters ρ and ρ_L, Algorithm 3 can learn the true POMIS set with fewer interventional samples compared to full discovery. The exact advantage depends on the density parameter ρ_L for latent confounders, as seen from the plot in Figure 5.