QUERY-SPECIFIC CAUSAL GRAPH PRUNING UNDER TIERED KNOWLEDGE

Anonymous authors
Paper under double-blind review

ABSTRACT

We present a systematic method for pruning edges from causal graphs by leveraging tiered knowledge. We characterize conditions under which edges can be removed from a causal graph while preserving the identifiability of (conditional) causal effects. This result enables causal identification on simplified graphs that are substantially smaller than the original graphs. This approach is particularly valuable when researchers are interested in causal relationships within specific tiers while controlling for broader influences from other tiers without fully specifying them. Building on this, we introduce a *query-specific* causal discovery procedure that takes a causal query as an additional input and recovers a reduced graph tailored to the query from observational data. Through theoretical analysis and empirical studies, we demonstrate that our procedure can achieve exponential speedups compared to the existing method when tiered knowledge is available.

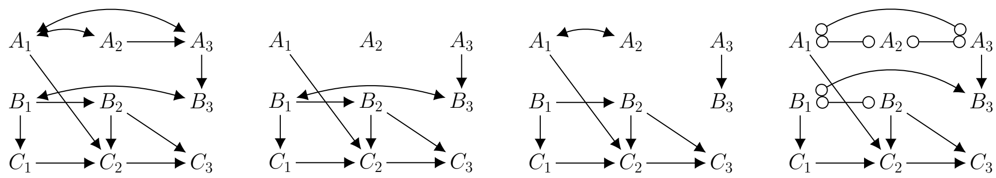
1 INTRODUCTION

Pearl's Causal Hierarchy categorizes queries into three separate layers: associational, interventional, and counterfactual (Pearl & Mackenzie, 2018). The first layer examines statistical associations estimable from observational data, while the latter two layers involve causal interventions and require understanding causal relationships between variables typically established through experimental studies (e.g., randomized controlled trials). We focus on the second layer and consider two widely studied types of interventional queries: causal effects and conditional causal effects. A causal effect quantifies the impact of an intervention on an outcome, exemplified by the question "What is the probability that a patient will recover after the doctor instructs them to undergo surgery?", whereas a conditional causal effect further restricts the query to a subpopulation and asks "What is the probability that a patient *older than 65* will recover after the doctor instructs them to undergo surgery?"

A common approach to answering these causal queries is through causal identification. Given a *causal graph* G and an *observational distribution* $\Pr(\mathbf{V})$, causal identification aims to estimate these causal queries based on G and $\Pr(\mathbf{V})$. This gives rise to the *identifiability* problem, which checks whether the values of causal queries are *uniquely* determined by G and $\Pr(\mathbf{V})$. Sound and complete methods for testing identifiability have been developed, including the identification method in (Tian & Pearl, 2003; Huang & Valtorta, 2006), the ID algorithm (Shpitser & Pearl, 2006), do-calculus (Pearl, 2009) for identifying causal effects, and the IDC algorithm (Shpitser & Pearl, 2008) for identifying conditional causal effects. Among these methods, the choice of causal graph plays a crucial role in determining both identifiability and the values of causal queries. For instance, consider the causal effect of surgery on patient recovery. The causal effect is identifiable if the graph contains a directed edge $Surgery \rightarrow Recovery$ but becomes unidentifiable if a hidden confounder introduces an additional bidirected edge $Surgery \leftrightarrow Recovery$.[1] The causal effect remains identifiable if the graph contains a reversed edge $Surgery \leftarrow Recovery$, but its value becomes zero.

Recent advancements in causal inference have expanded beyond traditional frameworks by integrating background knowledge in addition to the causal graph, leading to improvements in both identifiability and computational efficiency. This includes leveraging the knowledge of functional dependencies (Chen & Darwiche, 2022; 2024; Darwiche, 2021), context-specific independencies (Tikka

[1]The bidirected edge $A \leftrightarrow B$ means $A \leftarrow U \rightarrow B$ where U is a *hidden confounder* causing both A and B. For example, patients' symptoms are potential hidden confounders between "Surgery" and "Recovery".



(a) original graph G (b) pruned graph G' (c) mutilated graph (d) PAG

Figure 1: Pruned, mutilated, and partial ancestral graphs for G. Tiers are indicated using distinct alphabetical letters. For example, tier A may represent contextual variables (e.g., patient characteristics), B the treatment variables (e.g., surgery), and C the outcome variables (e.g., recovery).

et al., 2019; Mokhtarian et al., 2022), and a fully-known observational distribution (Chen & Darwiche, 2025). In this work, we extend this research line by showing yet another benefit of incorporating background knowledge: *it reduces the size of causal graphs required for causal identification.* In particular, we consider *tiered knowledge*, which partitions variables into "tiers" with a fixed causal ordering (Andrews et al., 2020). Specifically, only directed edges can be added from higher to lower tiers. This type of knowledge arises frequently, for example, in instrumental-variable analysis and in temporal reasoning where variables are ordered according to their time steps; see (Andrews et al., 2020) for more details. A key contribution of this work is to introduce an edge pruning technique that leverages tiered knowledge while preserving the identifiability of (conditional) causal effects. This technique allows us to test identifiability using a reduced graph rather than the original graph, while maintaining the soundness and completeness. As a preliminary example, consider the causal graph in Figure 1a, which contains three tiers (distinguished by alphabetical letters). Our edge pruning results enable us to conclude that causal effects $\Pr_{B_1}(B_3, C_3)$ and $\Pr_{B_1}(B_3, C_3|A_1)$ are identifiable in G by showing that they are identifiable in the pruned graph G'. One key advantage of considering these smaller causal graphs is that it facilitates graph specification when edges are either determined through domain knowledge or learned from data. We demonstrate this by presenting a *query-specific* causal discovery algorithm that improves upon the existing method by learning a reduced graph tailored to a specific causal query, resulting in exponential computational speedups.

The paper is structured as follows. We start by reviewing key definitions and methodologies in causal inference and discovery in Section 2. In Section 3, we introduce our main results on graph pruning that preserves the identifiability of (conditional) causal effects. In Section 4, we propose a refinement on the existing causal discovery algorithm, which yields improved computational efficiency. In Section 5, we present experimental results to further demonstrate the effectiveness of our method. We close with some concluding remarks in Section 6. All proofs are included in the Appendix.

2 TECHNICAL PRELIMINARIES

We focus on discrete variables, though all the results presented in this work can be extended to continuous domains. Single variables are denoted by uppercase letters (e.g., X) and their states are denoted by lowercase letters (e.g., x). Sets of variables are denoted by bold uppercase letters (e.g., $\mathbf{X}$) and their instantiations are denoted by bold lowercase letters (e.g., $\mathbf{x}$).

2.1 CASUAL GRAPHS WITH TIERED KNOWLEDGE

Causal graphs are widely used to model the causal relationship among variables. In this work, we do not assume causal sufficiency and consider causal graphs in the form of Acyclic Directed Mixed Graphs (ADMGs), which contain both directed and bidirected edges.

Definition 2.1. (Richardson, 2003a) An acyclic directed mixed graph (ADMG) is a graph that contains directed edges ($\to$) and bidirected edges ($\leftrightarrow$) and in which directed edges do not form cycles.

Figures 1a and 1b are both valid ADMGs. We can interpret the edges in an ADMG as follows. Each directed edge $A \to B$ means that A is a *direct cause* of B. Moreover, we call A a *parent* of B, and B a *child* of A. Each bidirected edge $A \leftrightarrow B$ indicates the existence of a *hidden confounder* causing both A and B. In Figure 1a, for example, A_1 is a direct cause of C_2, and there exists a hidden confounder between B_1 and B_3. Variable A is a *neighbor* of B if there is an edge (directed or bidirected) between A and B. Two variables A, B are said to be in a same *c-component* in a graph if they are connected by a bidirected path, i.e., a path containing only bidirected edges. The variables in a graph can always be partitioned into multiple c-components. For example, variables in Figure 1a can be partitioned into c-components $\{A_1, A_2, A_3\}$, $\{B_1, B_3\}$, $\{B_2\}$, $\{C_1\}$, $\{C_2\}$, $\{C_3\}$.

Tiered knowledge introduces additional constraints on causal graphs by specifying a causal ordering over its observed variables $\mathbf{V}$ (Andrews et al., 2020). By definition, tiered knowledge partitions $\mathbf{V}$ into t tiers (disjoint subsets), denoted $\{\mathbf{V}^1, \ldots, \mathbf{V}^t\}$, where all edges from higher tiers (tiers with lower indexes) to lower tiers (tiers with higher indexes) are directed. That is, for each pair $X \in \mathbf{V}^i$ and $Y \in \mathbf{V}^j$ where $i < j$, only the edge $X \to Y$ is permitted. Consider Figure 1a with tiers $\{\{A_1, A_2, A_3\}, \{B_1, B_2, B_3\}, \{C_1, C_2, C_3\}\}$. The directed edge $A_1 \to C_2$ is allowed since A_1 belongs to a higher tier than C_2, whereas neither $B_2 \to A_2$ nor $B_2 \leftrightarrow A_2$ are allowed since B_2 belongs to a lower tier than A_2. All edges in Figures 1a and 1b satisfy the tiered constraints. Such tiered causal orderings are quite prevalent in practice. For example, in a medical context, variables are naturally partitioned into tiers $\mathbf{V}^1, \mathbf{V}^2, \mathbf{V}^3$, where $\mathbf{V}^1$ represents patient characteristics (e.g., age), $\mathbf{V}^2$ captures medical treatments and adherence, and $\mathbf{V}^3$ reflects the quality of recovery.

For simplicity, we use Γ to denote the *tier mapping* where $\Gamma(X) = i$ for each $X \in \mathbf{V}^{(i)}$. For a set of variables $\mathbf{W}$, let $\Gamma^-(\mathbf{W})$ denote the *minimum* tier index of $\mathbf{W}$, i.e., $\Gamma^-(\mathbf{W}) = min_{W \in \mathbf{W}}\, \Gamma(W)$. Similarly, let $\Gamma^+(\mathbf{W})$ denote the *maximum* tier index of $\mathbf{W}$, i.e., $\Gamma^+(\mathbf{W}) = max_{W \in \mathbf{W}}\, \Gamma(W)$.

2.2 Identifying Conditional Causal Effects

The identification of causal effects and conditional causal effects from observational data has been studied extensively in the past; see, e.g., (Pearl, 2009; Hernán & Robins, 2020; Peters et al., 2017; Imbens & Rubin, 2015; Tian, 2004; Shpitser & Pearl, 2008). Let G be a causal graph and M be a model (parameterization) for G, an intervention $do(\mathbf{x})$ induces a mutilated graph G', obtained from G by removing all the incoming edges of $\mathbf{X}$, and a modified model $M_{\mathbf{x}}$, in which the values of $\mathbf{X}$ are fixed to $\mathbf{x}$ in M. Figure 1c depicts the mutilated graph under interventions on A_3, B_1. The modified model $M_{\mathbf{x}}$ induces an *interventional distribution*, denoted $\Pr_{\mathbf{x}}(\mathbf{V})$, which is used for computing the causal queries. Given a set of treatments $\mathbf{X}$ and a set of outcomes $\mathbf{Y}$, the *causal effect* of $\mathbf{X}$ on $\mathbf{Y}$, denoted $\Pr_{\mathbf{X}}(\mathbf{Y})$, computes the marginal distribution $\Pr_{\mathbf{x}}(\mathbf{Y})$ for each $\mathbf{x}$. The *conditional causal effect*, on the other hand, takes an additional conditioned set $\mathbf{Z}$ and computes $\Pr_{\mathbf{x}}(\mathbf{Y}|\mathbf{Z})$ for each $\mathbf{x}$.

The identifiability problem asks whether a causal query is *uniquely computable* from a causal graph G and an observational distribution $\Pr(\mathbf{V})$. We provide a formal definition for conditional causal-effect identifiability here – unconditional causal-effect identifiability is a special case when $\mathbf{Z} = \emptyset$.

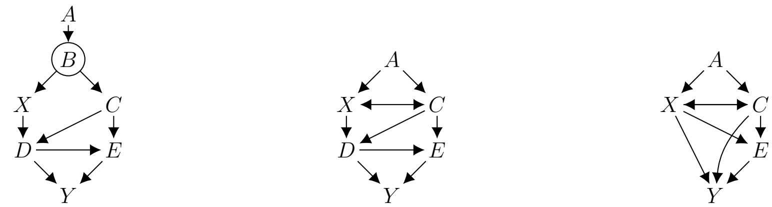
Definition 2.2. (Tian, 2004) The conditional causal effect of $\mathbf{X}$ on $\mathbf{Y}$ given $\mathbf{Z}$ is said to be identifiable with respect to $\langle G, \mathbf{V}\rangle$ if $\Pr^1_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = \Pr^2_{\mathbf{x}}(\mathbf{y}|\mathbf{z})$ for any pair of distribution $\Pr^1, \Pr^2$ induced by G and that satisfy $\Pr^1(\mathbf{V}) = \Pr^2(\mathbf{V}) > 0$.[2]

A causal query is called *unidentifiable* if it is not identifiable. In Figure 1a, $\Pr_{B_1}(B_3, C_3)$ and $\Pr_{A_1,B_1}(C_3|C_2)$ are both identifiable. On the other hand, $\Pr_{A_2}(B_3, C_2)$ and $\Pr_{A_2}(C_2|A_3)$ are unidentifiable. Sound and complete algorithms based on c-components have been developed for identifying both unconditional and conditional causal effects (Tian & Pearl, 2003; Huang & Valtorta, 2006; Shpitser & Pearl, 2008). These methods also provide an identifying formula for estimating the (conditional) causal effect whenever it is identifiable. For example, applying IDENTIFY (Tian & Pearl, 2003) not only confirms the identifiability of $\Pr_{b_1}(b_3, c_3)$ but also yields the following formula: $\Pr_{b_1}(b_3, c_3) = \sum_{a_1,a_3,b_2,c_1,c_2} \Pr(a_1, a_3)\Pr(b_2|b_1)\Pr(c_1|b_1)\Pr(c_2|a_1, b_2, c_1)\Pr(c_3|b_2, c_2)$ $\sum_{b'_1} \Pr(b_3|a_3, b'_1)\Pr(b'_1)$. Further details on these methods are provided in Appendix A.

2.3 Causal Discovery With Tiered Knowledge

Causal discovery learns a Markov equivalence class (MEC) of causal graphs from observational distribution $\Pr(\mathbf{V})$. The MEC contains all the Maximal Ancestral Graphs (MAGs) that exhibit the

[2] We assume strict positivity here since it is commonly required by existing identification methods.



(a) original (b) projected (c) functionally-projected

Figure 2: Original, projected, and functionally projected graphs that preserve identifiability.

same conditional independencies and is represented by Partial Ancestral Graphs (PAGs).[3] A PAG contains edges $\rightarrow$, $\leftrightarrow$, $\circ\!\!\rightarrow$ and $\circ\!\!-\!\!\circ$, where each $\circ$ can be replaced by either an arrowhead or a tail. To illustrate, Figure 1d depicts a PAG representing the MEC that contains the MAG in Figure 1a.

We consider the Fast Causal Inference (FCI) algorithm, a constraint-based causal discovery algorithm that is proven to be sound and complete for finding a *maximally informative* PAG from $\Pr(\mathbf{V})$ (Spirtes et al., 2000; Zhang, 2008). That is, the algorithm is guaranteed to return a PAG that represents exactly the MEC containing the true causal graph G, assuming $\Pr(\mathbf{V})$ is a P-MAP of G.[4] To apply existing causal identification methods (such as IDENTIFY), it is common to include a postprocessing step that recovers the true graph G by replacing every $\circ$ mark in the learned PAG with either an arrowheads or a tail, using approaches that leverage prior knowledge (Meek, 1995; Heckerman et al., 1995; Borboudakis & Tsamardinos, 2012) or incorporate experimental data (He & Geng, 2008; Hyttinen et al., 2013; Hauser & Bühlmann, 2014; Kocaoglu et al., 2017). While this postprocessing step is important, we omit details since it falls outside our focus on causal discovery.[5]

Under tiered knowledge, the FCI algorithm remains sound and complete in discovering PAGs as proved in (Andrews et al., 2020). The approach is based on splitting the original causal graph into multiple subgraphs based on tiers and applying the FCI algorithm iteratively to recover each subgraph; see more details in Section 4 and Appendix B. As we will show later, when a particular (conditional) causal effect is of interest, we can accelerate this process by pruning edges from PAGs.

3 Graph Pruning Under Tiered Knowledge

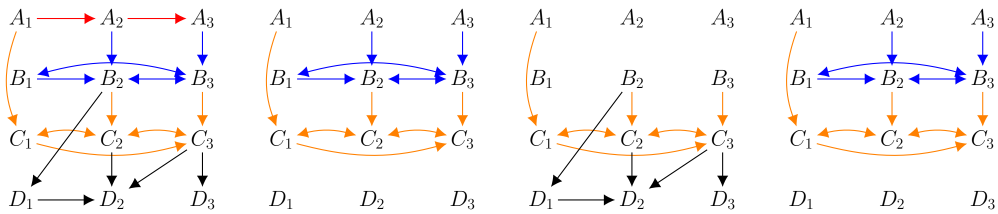
We present methods that prune edges from a causal graph while preserving the identifiability of (conditional) causal effects in the presence of tiered knowledge. This allows us to reduce the identifiability problem in the original graph G to the identifiability problem in a pruned graph G', which is advantageous when G' is easier to attain than G. We call G and G' *ID-equivalent* since they yield equivalent identifiability results – that is, a causal effect is identifiable in G iff it is identifiable in G'.

Such graph-reduction approaches for testing identifiability have been explored in prior works. A common example is the use of the *projection* operation (Verma, 1993; Tian & Pearl, 2002) to remove hidden variables from a DAG G before applying causal identification methods such as those in (Tian & Pearl, 2003; Shpitser & Pearl, 2006). Figure 2b depicts the projected graph for Figure 2a when variable B is hidden. More recently, a refined projection operation called *functional projection* was introduced to eliminate (remove) variables that exhibit *functional dependencies* on their parents (e.g., "legal driving age" is functionally determined by "country") (Chen & Darwiche, 2024). Suppose variable D in Figure 2b exhibits functional dependency on its parents $\{X, C\}$, then eliminating D

[3]MAGs are a subclass of ADMGs that satisfy *ancestrality* and *maximality*. Ancestrality prohibits bidirected edges between X, Y if there is a directed path from X to Y. Maximality ensures that non-adjacent nodes X, Y can always be m-separated by some variable set; see (Richardson & Spirtes, 2002) for more details.

[4]Formally, a PAG is *maximally informative* if each edge mark $\circ$ in the PAG corresponds to an arrowhead in some MAG and an arrow tail in another MAG within the MEC. $\Pr(\mathbf{V})$ is a perfect map (P-MAP) of G if for any disjoint variable sets $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, $(\mathbf{X} \perp\!\!\!\perp \mathbf{Y} | \mathbf{Z})$ in $\Pr$ iff $\mathbf{X}$ and $\mathbf{Y}$ are m-separated by $\mathbf{Z}$ in G; see details of P-MAP in (Pearl, 1988) and m-separation in (Richardson, 2003b).

[5]Another research direction avoids the conversion to MAGS by directly identifying causal queries from PAGs; see, e.g., (Jaber et al., 2022; Perkovic et al., 2017).



(a) original (b) $\Pr_{B_1}(C_1, C_2)$ (c) $\Pr_{C_1}(D_2, D_3|A_1, B_2)$ (d) $\Pr_{B_1}(C_1|B_2, B_3)$

Figure 3: Original graph and its pruned graphs for identifying $\Pr_{B_1}(C_1, C_2)$, $\Pr_{C_1}(D_2, D_3|A_1, B_2)$ and $\Pr_{B_1}(C_1|B_2, B_3)$. Tiers are indicated using distinct alphabetical letters.

using functional projection yields the graph in Figure 2c. Both projected and functionally projected graphs are ID-equivalent to the original graph and hence can be used for testing identifiability.

When a particular causal query is of interest, we show that we can construct an ID-equivalent graph that is guaranteed to be smaller than the original graph by leveraging tiered knowledge. We analyze several theoretical properties of these constructed graphs, including their validity for identifying the given query and their optimality in terms of graph size. We begin with the case of causal effects.

3.1 Edge Pruning For Causal Effects

Unless stated otherwise, let G denote a causal graph in which observed variables $\mathbf{V}$ are partitioned into t tiers $\{\mathbf{V}^1, \ldots, \mathbf{V}^t\}$ with tier mapping Γ. Moreover, let $\mathbf{X}, \mathbf{Y}$ be disjoint sets of treatment variables and outcome variables. We begin with a key graphical notion introduced in (Andrews et al., 2020), which we refer to as a "T-component" for convenience.

Definition 3.1. (Andrews et al., 2020) The i^{th} $(1 \leq i \leq t)$ tiered-component (T-component) is the subgraph of G containing all nodes in G and all edges (l, r) where $max\{\Gamma(l), \Gamma(r)\} = i$.

The edges in each T-component are highlighted with distinct colors in Figure 3a. For simplicity, we denote i^{th} T-component as G^i. The i^{th} T-component contains all edges between variables in tier i, which we refer to as *in-tier edges*, as well as directed edges from variables in higher tiers to variables in tier i, which we refer to as *cross-tier edges*. To illustrate, the T-component G^2 contains in-tier edges $B_1 \to B_2$, $B_1 \leftrightarrow B_3$, $B_2 \leftrightarrow B_3$ and cross-tier edges $A_2 \to B_2$, $A_3 \to B_3$. One observation is that the original graph is equivalent to the union of all T-components, i.e., $G = \bigcup_{i=1}^{t} G^i$.[6]

Our method is based on pruning edges from graph G without impacting the identifiability of the causal effect $\Pr_{\mathbf{X}}(\mathbf{Y})$. The following proposition characterizes the subgraph resulting from pruning.

Proposition 3.2. *Let G' be the union of all T-components G^i with $\Gamma^-(\mathbf{X}) \leq i \leq \Gamma^+(\mathbf{Y})$. Then $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable in G iff it is identifiable in G' and can be computed as $\Pr_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{w} \setminus \mathbf{y}} \Pr(\mathbf{w}) \text{IDENTIFY}_{G'}(\mathbf{x} \cup \mathbf{w}, \mathbf{y} \setminus \mathbf{w})$, where $\mathbf{W}$ is the set of variables whose tier indexes are smaller than $\Gamma^-(\mathbf{X})$ and $\text{IDENTIFY}_{G'}(\mathbf{x} \cup \mathbf{w}, \mathbf{y} \setminus \mathbf{w})$ is the identifying formula returned by IDENTIFY (Tian & Pearl, 2003) for computing $\Pr_{\mathbf{x} \cup \mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ in G'.*[7]

Consider the causal graph G in Figure 3a and the causal effect $\Pr_{B_1}(C_1, C_2)$. Without tiered knowledge, we would typically construct the full graph G containing all edges and use it to identify the causal effect. However, suppose we know in advance that G can be partitioned into four tiers (indicated by different alphabetical letters), we can apply Proposition 3.2 and use the ID-equivalent graph G' depicted in Figure 3b instead. In this case, $\Gamma^+(\mathbf{X}) = \Gamma^+(\{B_1\}) = 2$ and $\Gamma^-(\mathbf{Y}) = \Gamma^-(\{C_1, C_2\}) = 3$, so G' is constructed as the union of T-components G^2 and G^3. In particular, the pruned graph G' contains in-tier edges such as $B_1 \to B_2$ and $C_1 \leftrightarrow C_2$, as well as cross-tier edges such as $A_1 \to C_1$ and $A_2 \to B_2$. Including cross-tier edges in pruned graphs provides several advantages. First, when the pruned graph is learned from observational data using causal discovery, omitting cross-tier edges may introduce incorrect edges in the graph. For example, ignoring

[6] The union of graphs $G_1(\mathbf{V}_1, \mathbf{E}_1), G_2(\mathbf{V}_2, \mathbf{E}_2)$, denoted $G_1 \cup G_2$, has nodes $\mathbf{V}_1 \cup \mathbf{V}_2$ and edges $\mathbf{E}_1 \cup \mathbf{E}_2$.

[7] The completeness of IDENTIFY was proved in (Huang & Valtorta, 2006). The IDENTIFY here can also be replaced by the ID algorithm (Shpitser & Pearl, 2006), which is also based on c-component decompositions. We cannot directly replace IDENTIFY with $\Pr_{\mathbf{x} \cup \mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ since the distribution $\Pr$ is not induced by G'.

$C_3 \to D_2$ and $C_3 \to D_3$ in Figure 1a would produce a false-positive edge between D_2 and D_3. Second, the cross-tier edges capture parent-child relationships that can be leveraged to simplify the identifying formula in Proposition 3.2 by exploiting the Markov assumption (see examples below).

The proposition establishes two key properties of the pruned graph G'. First, the identifiability of the causal effect is preserved, i.e., G and G' are ID-equivalent. For example, we can still apply existing sound and complete methods (such as IDENTIFY) to test the identifiability of $\Pr_{B_1}(C_1, C_2)$ in Figure 3b, which is identifiable in this case. Second, the proposition provides a *specific* identifying formula for estimating the causal effect from G' when it is identifiable. In this example, the identifying formula is $\Pr_{b_1}(c_1, c_2) = \sum_{a_1,a_2,b_2} \Pr(a_1, a_2) \Pr(b_2|a_2, b_1)\Pr(c_2|a_1, b_2, c_1)\Pr(c_1|a_1)$.[8] The form of identifying formula is fixed here since a formula for G' is not necessarily valid for the same causal effect in G. To illustrate, the following formula correctly computes $\Pr_{B_1}(C_1, C_2)$ in G' depicted in Figure 3b: $\Pr_{b_1}(c_1, c_2) = \sum_{b_2}[\sum_{a_1} \Pr(c_2|a_1, b_2, c_1) \Pr(a_1, c_1)][\sum_{a_2} \Pr(a_2) \Pr(b_2|a_2, b_1)]$, yet it does not correctly compute the causal effect in the original graph depicted in Figure 3a. Such errors occur since G' ignores all the dependencies between variables in tiers above $\mathbf{X}$, which can lead to inaccurate estimations if causal effects are identified solely based on the simplified graph G'.

A remaining question is whether the ID-equivalent causal graphs constructed by Proposition 3.2 are optimal in size. Finding *minimal* ID-equivalent causal graphs offers significant benefits for causal graph specification. For graphs specified through human knowledge, only edges in the minimal graphs need to labeled, which can save significant manual effort when the original graphs are large or when we lack knowledge to orient certain edges. For graphs learned from data, using smaller graphs can substantially reduce computational time of causal discovery as we demonstrate later.

We show that the bounds ($\Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y})$) provided in Proposition 3.2 are tight. Specifically, the identifiability of causal effects is no longer preserved if we prune additional edges from G' by increasing $\Gamma^-(\mathbf{X})$ or decreasing $\Gamma^+(\mathbf{Y})$. This holds trivially when $\Gamma^-(\mathbf{X}) > \Gamma^+(\mathbf{Y})$, where the pruned graph G' contains no edges. The next proposition proves the tightness for remaining cases.

Proposition 3.3. *Let $\mathcal{L}, \mathcal{U}, \mathcal{L}', \mathcal{U}'$ be positive integers with $\mathcal{L} \leq \mathcal{U}$, $\mathcal{L}' \leq \mathcal{U}'$, and $\mathcal{L}' > \mathcal{L}$ or $\mathcal{U}' < \mathcal{U}$. There exists a causal graph G and tiered mapping Γ where $\Gamma^-(\mathbf{X}) = \mathcal{L}$, $\Gamma^+(\mathbf{Y}) = \mathcal{U}$, and Proposition 3.2 no longer holds if the bounds $\Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y})$ are replaced with $\mathcal{L}'$ and $\mathcal{U}'$.*

Before discussing edge pruning for conditional causal effects, we highlight an important result on the size of pruned graphs: there exists a class of graphs where the sizes of original graphs are unbounded, while the sizes of their pruned graphs (from Proposition 3.2) are bounded. Such graphs can be constructed, for example, by adding more tiers above B_i's in Figure 3a. The graph size grows infinitely while the size of its pruned graph for identifying $\Pr_{B_1}(C_1, C_2)$ remains constant. This result has important implications for graph specification: when causal graphs are constructed from prior knowledge or learned via causal discovery, it suffices to specify or learn only *a constant number* of edges to identify the causal effect, even if the full causal graph is arbitrarily large. While we will elaborate on this further in Section 4, we present this key result as the proposition below.

Proposition 3.4. *There exists a class of causal graphs G^t with t tiers whose edge count grows unbounded with t, while the pruned graphs (Proposition 3.2) have a constant number of edges.*

3.2 EDGE PRUNING FOR CONDITIONAL CAUSAL EFFECTS

We now consider a more general type of causal query called conditional causal effects. Compared to causal effects, these queries take another set of conditioned variables $\mathbf{Z}$ as input and compute $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$. Analogous to causal effects, we introduce a method to prune edges from causal graphs while preserving identifiability. The pruning criterion, however, is more subtle for conditional causal effects as it depends on the choice of conditioned variables.

We start with the case where the conditioned variables are in higher tiers than treatment variables.

Proposition 3.5. *Let G' be the union of all T-components G^i with $\Gamma^-(\mathbf{X}) \leq i \leq \Gamma^+(\mathbf{Y})$. If $\Gamma^+(\mathbf{Z}) < \Gamma^-(\mathbf{X})$, then $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is identifiable in G iff it is identifiable in G' and can be computed as $\Pr_{\mathbf{x}}(\mathbf{y}|\mathbf{z}) = \frac{\Pr_{\mathbf{x}}(\mathbf{y},\mathbf{z})}{\Pr(\mathbf{z})}$.*

[8]The formula is simplified from the return of IDENTIFY: $\Pr_{b_1}(c_1, c_2) = \sum_{a_1,a_2,a_3,b_2} \Pr(a_1, a_2, a_3) \Pr(b_2|a_2, b_1) \Pr(c_2|a_1, b_2, c_1) \Pr(c_1|a_1)$. This illustrates the benefit of including cross-tier edges.

Algorithm 1 Query-Specific Causal Discovery with Tiered Knowledge

Inputs: Variables $\mathbf{V} = (\mathbf{V}^{(1)}, \dots, \mathbf{V}^{(n)})$, Query $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$
Output: PAG $\mathcal{P}$

/ Preprocessing to determine the min and max tier indexes */*
if $\mathbf{Z} = \emptyset$ or $\Gamma^+(\mathbf{Z}) < \Gamma^-(\mathbf{X})$ **then**
  $maxTier \leftarrow \Gamma^+(\mathbf{Y}),\ minTier \leftarrow \Gamma^-(\mathbf{X})$
else
  $maxTier \leftarrow \Gamma^+(\mathbf{Y} \cup \mathbf{Z}),\ minTier \leftarrow 1$
end if
/ Below is adapted from (Andrews et al., 2020, Algorithm 1) */*
$G \leftarrow$ unconnected graph over $\mathbf{V}$
for $i = maxTier$ to $minTier$ **do**
  $\mathbf{A}_i \leftarrow \bigcup_{j=1}^{i-1} \mathbf{V}^{(j)}$
  $\mathbf{B}_i \leftarrow \mathbf{V}^{(i)}$
  $\mathcal{P}^i \leftarrow$ FCIEXOGENOUS$(\mathbf{A}_i, \mathbf{B}_i)$ ▷ (Andrews et al., 2020, Algorithm 2)
  Add edges in $\mathcal{P}^i$ to $\mathcal{P}$
end for

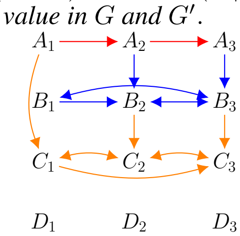
That is, a conditional causal effect is identifiable in the original graph G iff it is identifiable in the pruned graph G'. We stress again that this reduction only holds when the conditioned variables $\mathbf{Z}$ are in tiers above the treatment variables $\mathbf{X}$. The identifying formula contains two parts: an interventional probability $\Pr_{\mathbf{x}}(\mathbf{y}, \mathbf{z})$ and an observational probability $\Pr(\mathbf{z})$. While $\Pr(\mathbf{z})$ can be directly computed from the observational distribution, we apply Proposition 3.2 to estimate $\Pr_{\mathbf{x}}(\mathbf{y}, \mathbf{z})$ by plugging in treatments $\mathbf{X}' = \mathbf{X}$ and outcomes $\mathbf{Y}' = \mathbf{Y} \cup \mathbf{Z}$. Interestingly, Proposition 3.2 always computes $\Pr_{\mathbf{X}'}(\mathbf{Y}')$ using the same graph G' because $\Gamma^-(\mathbf{X}') = \Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y}') = \Gamma^+(\mathbf{Y})$.

Consider again the causal graph in Figure 3a but with a conditional causal effect $\Pr_{B_1}(C_1, C_2|A_2)$. Since $\Gamma^+(\mathbf{Z}) = \Gamma^+(A_2) = 1$ is less than $\Gamma^-(\mathbf{X}) = \Gamma^-(B_1) = 2$, we can apply Proposition 3.5 and test identifiability in the pruned graph G' in Figure 3b. In this case, the conditional causal effect is identifiable in G' according to the IDC algorithm (Shpitser & Pearl, 2006) and can be computed as $\Pr_{b_1}(c_1, c_2|a_2) = \frac{\Pr_{b_1}(c_1,c_2,a_2)}{\Pr(a_2)}$, where $\Pr_{b_1}(c_1, c_2, a_2)$ can be further computed from G' as $\Pr_{b_1}(c_1, c_2, a_2) = \sum_{a_1,b_2} \Pr(a_1, a_2) \Pr(b_2|a_2, b_1) \Pr(c_2|a_1, b_2, c_1) \Pr(c_1|a_1)$ by Proposition 3.2. Our second example examines the same causal graph but a different conditional causal effect $\Pr_{C_1}(D_2, D_3|A_1, B_2)$. Since $\Gamma^+(\{A_1, B_2\}) < \Gamma^-(\{C_1\})$, we can again apply Proposition 3.5 and test identifiability using the pruned graph in Figure 3c, which concludes that $\Pr_{C_1}(D_2, D_3|A_1, B_2)$ is unidentifiable by the IDC algorithm. Similar to the case of unconditional causal effects, the bounds in Proposition 3.6 are tight, which we formulate and prove as Proposition C.2 in the Appendix.

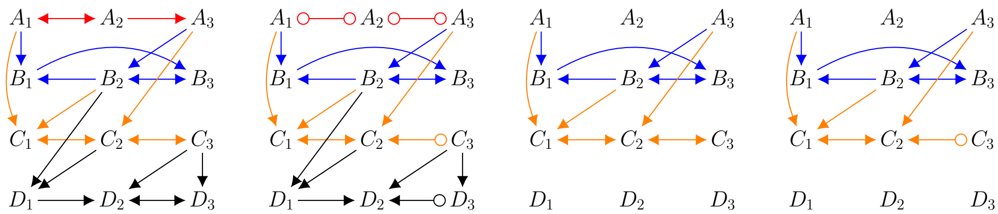
One may naturally ask whether the foregoing edge pruning method remains valid when the condition $\Gamma^+(\mathbf{Z}) < \Gamma^-(\mathbf{X})$ is violated. The answer is negative. To illustrate this, we present a conditional causal effect that is unidentifiable in the original graph but becomes identifiable in the graph constructed via Proposition 3.5. Consider the causal graph in Figure 3a and the conditional causal effect $\Pr_{B_1}(C_1|B_2, B_3)$. The condition required by the proposition no longer holds here since $\Gamma^+(\{B_2, B_3\}) = \Gamma^-(\{B_1\})$. The causal effect is unidentifiable in the original graph according to the IDC algorithm but becomes identifiable in the pruned graph depicted in Figure 3d. This example suggests that the edge pruning method does not, in general, preserve the identifiability of conditional causal effects when a conditioned variable lies in the same or a lower tier than a treatment variable.

However, we can always prune edges below $(\mathbf{Y} \cup \mathbf{Z})$ while preserving the identifiability of $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$. We state this observation as Proposition 3.6 and illustrate with an example below.

Proposition 3.6. *Let G' be the union of all T-components G^i with $i \leq \Gamma^+(\mathbf{Y} \cup \mathbf{Z})$. Then $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is identifiable in G iff it is identifiable in G'. Moreover, it attains the same value in G and G'.*



Consider again the graph in Figure 3a and a conditional causal effect $\Pr_{B_1}(C_1|B_2, B_3)$. The graph on the right depicts the pruned graph based on Proposition 3.6. In contrast to Figure 3d, all edges in tiers at or higher than $\Gamma^+(\{B_2, B_3, C_1\})$ are preserved. We can then apply the IDC algorithm on the pruned graph to conclude that $\Pr_{B_1}(C_1|B_2, B_3)$ is unidentifiable. Although the edge pruning in Proposition 3.6 appears more con-



(a) MAG (b) PAG (c) MAG for $\Pr_{B_1}(C_2)$ (d) PAG for $\Pr_{B_1}(C_2)$

Figure 4: MAG G, PAG for G, MAG required for $\Pr_{B_1}(C_2)$, and PAG learned from Algorithm 1.

servative than that in Propositions 3.2 and 3.5, it is applicable in all cases, regardless of the choice of conditioned variables, and remains effective in improving the efficiency of existing causal discovery algorithms, as we demonstrate in the following sections.

4 Query-Specific Causal Discovery

We now show one key application of edge pruning methods introduced in the previous sections: accelerating causal discovery when causal graphs are learned from data. Given a causal graph (MAG) G and a causal effect $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$, Propositions 3.2-3.6 shows that we can identify $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ in a pruned graph G' while preserving identifiability. For example, testing identifiability of $\Pr_{B_1}(C_2)$ in the pruned graph in Figure 4a yields the same result as in the original graph in Figure 4c.

The same edge pruning techniques extend to PAGs. In particular, we can obtain *partial* PAGs from *full* PAGs by pruning edges using Propositions 3.2-3.6 and use the partial PAGs for the purpose of identifying (conditional) causal effects. To illustrate, we only need to learn the partial PAG in Figure 4d instead of the full PAG in Figure 4b if we are interested in $\Pr_{B_1}(C_2)$. By incorporating additional knowledge or experimental data, the uncertain edge $C_3 \circ\!\!\rightarrow C_2$ may be replaced by $C_3 \leftrightarrow C_2$, yielding a MAG in Figure 4c which can be used to identify the causal effect $\Pr_{B_1}(C_2)$.

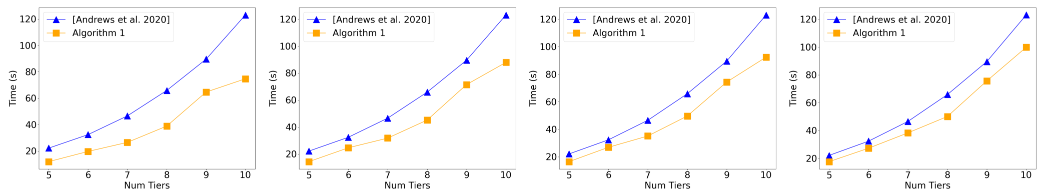
Our approach of query-specific causal discovery (Algorithm 1) for learning partial PAGs is adapted from the method FciTiers introduced in (Andrews et al., 2020). Given a particular $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$, we first apply Proposition 3.2-3.6 to identify the T-components in the pruned graph. For example, the pruned graph for $\Pr_{B_1}(C_2)$ in Figure 4b contains T-components G^2, G^3 by Proposition 3.2. We then iterate through each T-component in the pruned graph and discover the PAG for that T-component. In this example, discovering the PAGs for G^2, G^3 yields Figure 4d. This procedure is detailed in Algorithm 1, where lines 1–6 implement a preprocessing step for finding the T-components in pruned graphs, and lines 7–14 perform causal discovery. The PAG discovery for each T-component follows the same procedure as FciExogenous in (Andrews et al., 2020) and is described in Appendix B.

The following result shows that Algorithm 1 is sound and complete for discovering partial PAGs. That is, Algorithm 1 is guaranteed to find the maximally informative PAG when the observational distribution is a P-MAP of the true causal graph.

Proposition 4.1. *Let G be a MAG, $\mathcal{P}$ be the maximally informative PAG for the MEC of G (with tiered knowledge), and $\mathcal{P}'$ be the result of pruning edges from $\mathcal{P}$ according to Propositions 3.2-3.6. Then Algorithm 1 returns $\mathcal{P}'$ when using G as the conditional independence oracle.*

A key benefit of focusing on a partial (smaller) PAG is the reduction in computational cost during causal discovery. In Section 3, we showed a class of causal graphs with unbounded size that can be reduced to bounded size through edge pruning. This example remains relevant for causal discovery, where the complexity of learning the full PAG can be unbounded, while learning the corresponding partial PAG is computationally bounded. We next provide an additional result demonstrating that the edge pruning technique can yield exponential speedups even when the number of tiers is fixed.

Proposition 4.2. *There exists a class of distributions induced by causal graphs with n nodes and 2 tiers, along with a causal effect, where FciTiers in (Andrews et al., 2020) takes $O(n^2 \cdot 2^n)$ time, while Algorithm 1 takes $O(n^4)$ time, assuming access to a conditional independence oracle.*



(a) $|\mathbf{Z}| = 0$ (b) $|\mathbf{Z}| = 1$ (c) $|\mathbf{Z}| = 2$ (d) $|\mathbf{Z}| = 3$

Figure 5: Comparison of execution times between tiered FCI (Andrews et al., 2020) and Algorithm 1 for varying sizes of the conditioned set $\mathbf{Z}$. The queries are reduced to causal effects when $|\mathbf{Z}| = 0$.

5 Experiments

We present empirical results to illustrate the improved computational efficiency enabled by our graph pruning approach. Specifically, we compare the execution times of FCITIERS (Andrews et al., 2020) and our Algorithm 1 on randomly generated causal graphs and (conditional) causal effects.

We start by generating random causal graphs that contain $T \in \{5, 6, 7, 8, 9, 10\}$ tiers, where each tier contains 10 variables that have either two or three states. This is done by first randomizing at most 5 parents for each variable and then randomly converting 30% of the in-tier (directed) edges into bidirected edges. To ensure sparsity, we bound the maximal number of neighbors of nodes by 4. We then assign a random parameterization for the causal graph and sample $500,000$ data instances from the model. Finally, we construct a conditional causal effect by randomly selecting 3 treatment variables, 3 outcome variables, and a conditioned set of size $\{0, 1, 2, 3\}$ from the causal graph.

Figure 4 compares the execution times of FCITIERS (Andrews et al., 2020) and our Algorithm 1 for learning PAGs from generated samples, averaged over 30 runs. We observed the following patterns. First, Algorithm 1 achieved shorter execution times in all cases, which matches our expectation since edge pruning enables our method to learn a partial PAG tailored to the specific causal effect instead of the full PAG, as illustrated in the previous section. Moreover, for a fixed number of conditioned variables, the performance gap between the two algorithms increases as the number of tiers in the causal graph grows, since additional tiers provide more opportunities to prune edges. Second, for a fixed causal graph, increasing the size of the conditioning set $\mathbf{Z}$ reduces the performance gap. This is because a larger $\mathbf{Z}$ typically results in fewer pruned edges, as the constraint $\Gamma^+(\mathbf{Z}) < \Gamma^-(\mathbf{X})$ in Proposition 3.5 is less likely to hold, making the learned partial PAG closer in size to the full PAG.

In addition to computation time, we evaluate the precision and recall of the adjacencies and arrowheads learned by both methods, as reported in Table 1 in the Appendix.[9] In particular, the PAGs returned by causal discovery methods are compared against the ground truth obtained by running FCITIERS with the true causal graph as the conditional independence oracle. The results show that our method achieves comparable precisions and recalls to FCITIERS, supporting the soundness and completeness of Algorithm 1 in identifying maximally informative partial PAGs.

6 Conclusion

We proposed a graph pruning method that removes edges from causal graphs while preserving the identifiability of (conditional) causal effects under tiered background knowledge. This approach enables more efficient specification of causal graphs for identifying causal effects, whether the graphs are constructed from human knowledge or learned from observational data. We focused on the latter case in this work and developed a causal discovery algorithm that incorporates edge pruning as a preprocessing step, resulting in significant speedups over the existing method. Potential directions for future research include extending the pruning techniques to additional types of queries (such as counterfactual queries) and generalizing the framework to support bidirected edges across tiers.

[9]The adjacency (arrowhead) precision and recall are used in (Scheines & Ramsey, 2016) to measure the soundness and completeness of causal discovery algorithms. Adjacency (arrowhead) precision measures the proportion of predicted edges (arrowheads) that are present in the true PAG, while adjacency (arrowhead) recall measures the proportion of true edges (arrowheads) correctly predicted by the algorithm.

References

Bryan Andrews, Peter Spirtes, and Gregory F. Cooper. On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. In *AISTATS*, volume 108 of *Proceedings of Machine Learning Research*, pp. 4002–4011. PMLR, 2020.

Giorgos Borboudakis and Ioannis Tsamardinos. Incorporating causal prior knowledge as path-constraints in bayesian networks and maximal ancestral graphs. In *ICML*. icml.cc / Omnipress, 2012.

Yizuo Chen and Adnan Darwiche. On the definition and computation of causal treewidth. In *UAI*, volume 180 of *Proceedings of Machine Learning Research*, pp. 368–377. PMLR, 2022.

Yizuo Chen and Adnan Darwiche. Identifying causal effects under functional dependencies. *Entropy*, 26(12), 2024.

Yizuo Chen and Adnan Darwiche. Constrained identifiability of causal effects. *AAAI AICT Workshop*, 2025.

Adnan Darwiche. Causal inference with tractable circuits. In *WHY Workshop, NeurIPS*, 2021. https://arxiv.org/abs/2202.02891.

Alain Hauser and Peter Bühlmann. Two optimal strategies for active learning of causal models from interventional data. *Int. J. Approx. Reason.*, 55(4):926–939, 2014.

Yang-Bo He and Zhi Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(84):2523–2547, 2008.

David Heckerman, Dan Geiger, and David Maxwell Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Mach. Learn.*, 20(3):197–243, 1995.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

Yimin Huang and Marco Valtorta. Identifiability in causal bayesian networks: A sound and complete algorithm. In *AAAI*, pp. 1149–1154. AAAI Press, 2006.

Antti Hyttinen, Frederick Eberhardt, and Patrik O. Hoyer. Experiment selection for causal discovery. *J. Mach. Learn. Res.*, 14(1):3041–3071, 2013.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

Amin Jaber, Adèle H. Ribeiro, Jiji Zhang, and Elias Bareinboim. Causal identification under markov equivalence: Calculus, algorithm, and completeness. In *NeurIPS*, 2022.

Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. In *NIPS*, pp. 7018–7028, 2017.

Christopher Meek. Causal inference and causal explanation with background knowledge. In *UAI*, pp. 403–410. Morgan Kaufmann, 1995.

Ehsan Mokhtarian, Fateme Jamshidi, Jalal Etesami, and Negar Kiyavash. Causal effect identification with context-specific independence relations of control variables. In *AISTATS*, volume 151 of *Proceedings of Machine Learning Research*, pp. 11237–11246. PMLR, 2022.

Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988. ISBN 1558604790.

Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, second edition, 2009.

Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

Emilija Perkovic, Johannes Textor, Markus Kalisch, and Marloes H. Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *J. Mach. Learn. Res.*, 18:220:1–220:62, 2017.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.

Thomas Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, 2003a. ISSN 03036898, 14679469.

Thomas Richardson. Markov properties for acyclic directed mixed graphs. *Scandinavian Journal of Statistics*, 30(1):145–157, 2003b.

Thomas Richardson and Peter Spirtes. Ancestral graph Markov models. *The Annals of Statistics*, 30 (4):962 – 1030, 2002. doi: 10.1214/aos/1031689015. URL https://doi.org/10.1214/aos/1031689015.

Richard Scheines and Joseph D. Ramsey. Measurement error and causal discovery. In *CFA@UAI*, volume 1792 of *CEUR Workshop Proceedings*, pp. 1–7. CEUR-WS.org, 2016.

Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *AAAI*, pp. 1219–1226. AAAI Press, 2006.

Ilya Shpitser and Judea Pearl. Complete identification methods for the causal hierarchy. *J. Mach. Learn. Res.*, 9:1941–1979, 2008.

Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search, Second Edition*. Adaptive computation and machine learning. MIT Press, 2000.

Jin Tian. Identifying conditional causal effects. In *UAI*, pp. 561–568. AUAI Press, 2004.

Jin Tian and Judea Pearl. On the testable implications of causal models with hidden variables. In *UAI*, pp. 519–527. Morgan Kaufmann, 2002.

Jin Tian and Judea Pearl. On the identification of causal effects. *Technical Report*, R-290-L, 2003.

Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Identifying causal effects via context-specific independence relations. In *NeurIPS*, pp. 2800–2810, 2019.

Thomas S. Verma. Graphical aspects of causal models. *Technical Report*, R-191, 1993.

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

A METHODS FOR IDENTIFYING (CONDITIONAL) CAUSAL EFFECTS

We review the sound and complete methods for identifying causal effects and conditional causal effects. We start with the IDENTIFY algorithm (Tian & Pearl, 2003; Huang & Valtorta, 2006) for identifying causal effects $\Pr_{\mathbf{X}}(\mathbf{Y})$.

The IDENTIFY algorithm starts by partition the variables in the subgraph $An(\mathbf{Y})_{G_{\mathbf{X}}}$ into c-components $\mathcal{S} = \{\mathbf{S}_1, \ldots, \mathbf{S}_k\}$, where $An(\mathbf{Y})_{G_{\mathbf{X}}}$ contains nodes that are ancestors of $\mathbf{Y}$ in the mutilated graph under $do(\mathbf{X})$, along with edges between these nodes. For each $\mathbf{S}_i \in \mathcal{S}$, there exists a unique c-component $\mathbf{C}_i$ in the original graph G that contains the variables in $\mathbf{S}_i$.

Let D be the subgraph formed by $\mathbf{S}_i$ in $An(\mathbf{Y})_{G_{\mathbf{X}}}$, and T be the subgraph formed by $\mathbf{C}_i$ in G. We say that $\mathbf{S}_i$ can be identified from $\mathbf{C}_i$ iff D can be obtained from T by recursively applying the following two operations: (1) replace T by the subgraph formed by its c-component that contains $\mathbf{C}_i$; and (2) replace T by $An(\mathbf{C}_i)_T$; see (Tian & Pearl, 2003) for details of the full algorithm. This yields a sound and complete algorithm for testing identifiability: $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable iff each $\mathbf{S}_i$ can be identified from $\mathbf{C}_i$ (Huang & Valtorta, 2006).

The identification methods for conditional causal effects $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$, IDC (Shpitser & Pearl, 2008), is based on a complete reduction of the problem to that of identifying causal effects. The reduction involves finding a *maximal* set of variables $\mathbf{W}$ such that $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z}) = \Pr_{\mathbf{X},\mathbf{W}}(\mathbf{Y}|\mathbf{Z} \setminus \mathbf{W})$, which can be determined using rule 2 of do-calculus. Existing methods for identifying causal effects (such as IDENTIFY or ID (Shpitser & Pearl, 2006)) can then be applied after the reduction to test the identifiability of $\Pr_{\mathbf{X},\mathbf{W}}(\mathbf{Y}|\mathbf{Z} \setminus \mathbf{W})$; see (Shpitser & Pearl, 2008) for more details.

B CAUSAL DISCOVERY WITH TIERED KNOWLEDGE

We briefly review the FCIEXOGENOUS procedure introduced in (Andrews et al., 2020), which is a key component in our query-specific causal discovery in Algorithm 1.

The FCIEXOGENOUS procedure discovers the PAG for the i^{th} T-component from observational data. It takes two sets of variables $\mathbf{A}$ and $\mathbf{B}$ as inputs, where $\mathbf{B}$ are the variables in tier i and $\mathbf{A}$ are the variables in tiers above i. The procedure then discovers the PAG from data by following the standard FCI algorithm (Spirtes et al., 2000) with two modifications. First, the adjacencies between variables in $\mathbf{A}$ are forbidden. Second, all edges between $A \in \mathbf{A}$ and $B \in \mathbf{B}$ are oriented as $A \to B$ before orientation rules $\mathcal{R}_0 - \mathcal{R}_{10}$ in (Zhang, 2008) are applied; see Algorithm 2 for the detailed procedure.

C PROOFS

C.1 PROOF OF PROPOSITION 3.2

We start with the following lemma that shows an identifying formula for computing $\Pr_{\mathbf{X}}(\mathbf{Y})$ in G.

Lemma C.1. *The causal effect* $\Pr_{\mathbf{X}}(\mathbf{Y})$ *is identifiable in* G *and can be computed as* $\Pr_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{w}\setminus\mathbf{y}} \Pr(\mathbf{w}) \Pr_{\mathbf{x}\cup\mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ *iff* $\Pr_{\mathbf{x}\cup\mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ *is identifiable in* G*, where* $\mathbf{W}$ *denotes variables whose tier indexes are smaller than* $\Gamma^-(\mathbf{X})$.

Proof. We first prove that $\Pr_{\mathbf{x}}(\mathbf{y}) = \sum_{\mathbf{w}\setminus\mathbf{y}} \Pr(\mathbf{w}) \Pr_{\mathbf{x}\cup\mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ in all graphs that satisfy tiered knowledge using do-calculus (Pearl, 2009).

$$
\begin{aligned}
\Pr_{\mathbf{x}}(\mathbf{y}) &= \sum_{\mathbf{w}\setminus\mathbf{y}} \Pr_{\mathbf{x}}(\mathbf{y}, \mathbf{w}) = \sum_{\mathbf{w}\setminus\mathbf{y}} \Pr_{\mathbf{x}}(\mathbf{w}) \Pr_{\mathbf{x}}((\mathbf{y} \setminus \mathbf{w})|\mathbf{w}) \\
&= \sum_{\mathbf{w}\setminus\mathbf{y}} \Pr(\mathbf{w}) \Pr_{\mathbf{x}}((\mathbf{y} \setminus \mathbf{w})|\mathbf{w}) \quad \text{(do-calculus, Rule 3 \& } \Gamma^+(\mathbf{W}) < \Gamma^-(\mathbf{X})\text{)} \\
&= \sum_{\mathbf{w}\setminus\mathbf{y}} \Pr(\mathbf{w}) \Pr_{\mathbf{x}\cup\mathbf{w}}(\mathbf{y} \setminus \mathbf{w}) \quad \text{(do-calculus, Rule 2 \& } \Gamma^+(\mathbf{W}) < \Gamma^-(\mathbf{Y} \setminus \mathbf{W})\text{)}
\end{aligned} \tag{1}
$$

This immediately implies that $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable and can be computed using the formula if $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is identifiable. We are left to show that $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is identifiable if

Algorithm 2 FCIEXOGENOUS (Andrews et al., 2020, Algorithm 2)

Inputs: Variable sets $\mathbf{A}, \mathbf{B}$
Output: PAG $\mathcal{P}$

Add $X_i \rightarrow X_j$ to $\mathcal{P}$ for all $X_i \in \mathbf{A}$ and $X_j \in \mathbf{B}$
Add X_i ∘—∘ X_j to $\mathcal{P}$ for all distinct $X_i, X_j \in \mathbf{B}$
$n \leftarrow 0$
repeat
  for all adjacent $(X_i, X_j) \in \mathbf{A} \cup \mathbf{B}$ and subset $\mathbf{S} \subseteq adj(X_i) \setminus \{X_j\}$ where $|\mathbf{S}| = n$ **do**
    if $\mathbf{X}_i \perp\!\!\!\perp X_j | \mathbf{S}$ **then**
      Delete edge (X_i, X_j) from $\mathcal{P}$
      $sepset(X_i, X_j) \leftarrow \mathbf{S}$, $sepset(X_j, X_i) \leftarrow \mathbf{S}$
    end if
  end for
  $n \leftarrow n + 1$
until $n > |adj(X_i) \setminus \{X_j\}|$ for all adjacent $(X_i, X_j) \in \mathbf{A} \cup \mathbf{B}$
Apply $\mathcal{R}_0$ to $\mathcal{P}$
for all adjacent $(X_i, X_j) \in \mathbf{A} \cup \mathbf{B}$ **do**
  / Checking possible d-separating sets pds */*
  if there exists $\mathbf{S} \in pds(X_i, X_j)$ where $X_i \perp\!\!\!\perp X_j | \mathbf{S}$ **then**
    Delete edge (X_i, X_j) from $\mathcal{P}$
    $sepset(X_i, X_j) \leftarrow \mathbf{S}$, $sepset(X_j, X_i) \leftarrow \mathbf{S}$
  end if
end for
for all adjacent $(X_i, X_j) \in \mathbf{B}$ **do**
  Replace edge (X_i, X_j) with X_i ∘—∘ X_j in $\mathcal{P}$
end for
Apply $\mathcal{R}_0 - \mathcal{R}_4, \mathcal{R}_8 - \mathcal{R}_{10}$ to $\mathcal{P}$

$\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable. Suppose $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is unidentifiable, there exists a hedge $\langle H, H' \rangle$ for $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ where $H' \subset H$ are c-forests in G that share a same set of roots (Shpitser & Pearl, 2006).[10] Moreover, H' must only contain variables that are ancestors of $(\mathbf{Y} \setminus \mathbf{W})$ in the mutilated graph $G_{\overline{\mathbf{X}\cup\mathbf{W}}}$ in which all incoming edges of $\mathbf{X} \cup \mathbf{W}$ are removed. This implies that all variables in H' have a tier index between $\Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y} \setminus \mathbf{W})$ (no hedge exists if $\Gamma^+(\mathbf{Y} \setminus \mathbf{W}) < \Gamma^-(\mathbf{X})$). Since H is a c-component containing H', it also contains variables whose tier indexes are between $\Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y} \setminus \mathbf{W})$ since no bidirected edge is allowed across tiers. In this case, $\langle H, H' \rangle$ is also a hedge for $\Pr_{\mathbf{X}}(\mathbf{Y})$, which concludes the unidentifiability of $\Pr_{\mathbf{X}}(\mathbf{Y})$. □

Proof of Proposition 3.2. By Lemma C.1, $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable in G iff $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is identifiable in G. Similarly, $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable in G' iff $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is identifiable in G'. Hence, it suffices to show (1) $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is identifiable in G iff it is identifiable in G', and (2) the formula returned by IDENTIFY$(\mathbf{x} \cup \mathbf{w}, \mathbf{y} \setminus \mathbf{w})$ in G' is valid for identifying $\Pr_{\mathbf{x}\cup\mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ in G.

First observe that $An(\mathbf{Y} \setminus \mathbf{W})_{G_{\overline{\mathbf{X}\cup\mathbf{W}}}}$ and $An(\mathbf{Y} \setminus \mathbf{W})_{G'_{\overline{\mathbf{X}\cup\mathbf{W}}}}$ are identical, where $An(\mathbf{V})_G$ denotes the ancestors of variables $\mathbf{V}$ in graph G. Hence, we obtain the same c-component decomposition $\mathcal{C}' = \{C'_1, \dots C'_m\}$ for $An(\mathbf{Y} \setminus \mathbf{W})_{G_{\overline{\mathbf{X}\cup\mathbf{W}}}}$ and $An(\mathbf{Y} \setminus \mathbf{W})_{G'_{\overline{\mathbf{X}\cup\mathbf{W}}}}$. Moreover, all variables in $\mathcal{C}'$ have tier indexes between $\Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y} \setminus \mathbf{W})$. Hence, they have the same edges between tiers $\Gamma^-(\mathbf{X})$ and $\Gamma^+(\mathbf{Y} \setminus \mathbf{W})$ and the same c-component decomposition $\mathcal{C} = \{C_1, \dots C_m\}$, where $C_1 \supseteq C'_1, \dots, C_m \supseteq C'_m$. Therefore, $\Pr_{\mathbf{X}\cup\mathbf{W}}(\mathbf{Y} \setminus \mathbf{W})$ is identifiable in G iff it is identifiable in G'. Since all variables in $\mathcal{C}$ have the same parents in G and G', the formula returned by IDENTIFY applied on G' can be used to identify $\Pr_{\mathbf{x}\cup\mathbf{w}}(\mathbf{y} \setminus \mathbf{w})$ in G by (Tian & Pearl, 2003, Lemma 4). □

C.2 PROOF OF PROPOSITION 3.3

Proof. The proof is based on constructing a causal graph G such that Proposition 3.2 is violated under any such $\mathcal{L}, \mathcal{U}, \mathcal{L}', \mathcal{U}'$. Let X be a treatment variable where $\Gamma(X) = \mathcal{L}$ and Y be an outcome variable where $\Gamma(Y) = \mathcal{U}$. Suppose $\mathcal{L} = \mathcal{U}$, we simply construct G with two edges: $X \rightarrow Y$ and $X \leftrightarrow Y$. The causal effect $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable in this case. However, any subgraph obtained

[10] A *c-forest* is a c-component in which all nodes have at most one child. The *roots* of a c-component are the nodes that do not have any child.

under ($\mathcal{L}'$,$\mathcal{U}'$) does not contain X, Y; hence, $\Pr_{\mathbf{X}}(\mathbf{Y})$ is identifiable in these subgraphs, which means that the identifiability is not preserved. Suppose now $\mathcal{L} < \mathcal{U}$, let W be a non-treatment variable where $\Gamma(W) = \mathcal{L}$. We construct G with the following edges: $X \to W$, $X \leftrightarrow W$, $W \to Y$, where $\Pr_{\mathbf{X}}(\mathbf{Y})$ is unidentifiable. However, the causal effect becomes identifiable either when $\mathcal{L}' > \mathcal{L}$, which removes X, W from G, or when $\mathcal{U}' < \mathcal{U}$, which removes Y from G. □

C.3 Proof of Proposition 3.4

Proof. Consider the following chain graph $V_1 \to V_2 \to \cdots \to V_t$ where V_i belongs to i^{th} tier and the causal effect $\Pr_{V_{t-1}}(V_t)$. As t grows, the number of edges in the chain graph becomes unbounded. However, the size of pruned graph remains $V_{t-1} \to V_t$, which contains a single edge. □

C.4 Proof of Proposition 3.5

We start by showing that $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is identifiable in G (or G') iff $\Pr_{\mathbf{X}}(\mathbf{Y}, \mathbf{Z})$ is identifiable in G (or G'). Consider the following equality:

$$\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z}) = \frac{\Pr_{\mathbf{X}}(\mathbf{Y}, \mathbf{Z})}{\Pr_{\mathbf{X}}(\mathbf{Z})} = \frac{\Pr_{\mathbf{X}}(\mathbf{Y}, \mathbf{Z})}{\Pr(\mathbf{Z})} \tag{2}$$

according to do-calculus Rule 3. Since $\Pr(\mathbf{Z})$ is fixed by the observational distribution, $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is identifiable iff $\Pr_{\mathbf{X}}(\mathbf{Y}, \mathbf{Z})$ is identifiable. Since $\Gamma^{+}(\mathbf{Z}) < \Gamma^{-}(\mathbf{X})$, the graph G' is the same as the subgraph induced by Proposition 3.2 for $\Pr_{\mathbf{X}}(\mathbf{Y}, \mathbf{Z})$. Hence, $\Pr_{\mathbf{X}}(\mathbf{Y}, \mathbf{Z})$ is identifiable in G iff it is identifiable in G', which concludes that $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is identifiable in G iff it is identifiable in G'. The identifying formula directly follows from Eq. 2.

C.5 Proof of Proposition C.2

Proposition C.2. *Let $\mathcal{L}$,$\mathcal{U}$, $\mathcal{L}'$,$\mathcal{U}'$ be positive integers with $\mathcal{L} \leq \mathcal{U}$, $\mathcal{L}' \leq \mathcal{U}'$, and either $\mathcal{L}' > \mathcal{L}$ or $\mathcal{U}' < \mathcal{U}$. There exists a causal graph G and tiered mapping Γ such that $\Gamma^{-}(\mathbf{X}) = \mathcal{L}$, $\Gamma^{+}(\mathbf{Y}) = \mathcal{U}$, and Proposition 3.5 no longer holds if the bounds $\Gamma^{-}(\mathbf{X})$ and $\Gamma^{+}(\mathbf{Y})$ are replaced with $\mathcal{L}'$ and $\mathcal{U}'$.*

Proof. This follows directly from the proof of Proposition 3.3 by considering the same graphs. □

C.6 Proof of Proposition 3.6

Proof. According to (Shpitser & Pearl, 2008), $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z})$ is identifiable in G iff $\Pr_{\mathbf{X}\cup\mathbf{Z}_1}(\mathbf{Y} \cup \mathbf{Z}_2)$ where $\mathbf{Z}_1, \mathbf{Z}_2$ form a partition of $\mathbf{Z}$. Moreover, $\mathbf{Z}_1$ is the unique maximal subset of $\mathbf{Z}$ such that $\Pr_{\mathbf{X}}(\mathbf{Y}|\mathbf{Z}) = \Pr_{\mathbf{X}\cup\mathbf{Z}_1}(\mathbf{Y}|\mathbf{Z}_2)$ by rule 2 of do-calculus. This set $\mathbf{Z}_1$ is the same in G and G' since $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}_1|\mathbf{X}, \mathbf{Z}_2)_{G_{\overline{\mathbf{X}}\underline{\mathbf{Z}_1}}}$ iff $(\mathbf{Y} \perp\!\!\!\perp \mathbf{Z}_1|\mathbf{X}, \mathbf{Z}_2)_{G'_{\overline{\mathbf{X}}\underline{\mathbf{Z}_1}}}$ when G' only removes edges below $\mathbf{Y} \cup \mathbf{Z}$. □

C.7 Proof of Proposition 4.1

Proof. By construction, the partial PAG $\mathcal{P}' = \bigcup_{i=minTier}^{maxTier} \mathcal{P}^i$ if we check each case in Propositions 3.2-3.6. Moreover, the soundness and completeness of learning each partial graph $\mathcal{P}^i$ follows directly from (Andrews et al., 2020, Lemma 12). □

C.8 Proof of Proposition 4.2

Proof. WLG, let $\mathbf{A} = \{A_1, \ldots, A_n\}$ be the variables in the first tier, and $\mathbf{B} = \{B_1, \ldots, B_n\}$ be the variables in the second tier. We add bidirected edges between every pair of variables in $\mathbf{A}$, directed edges from each A_i to B_i, and directed edges from each B_i to B_{i+1}. Suppose we are interested in the causal effect $\Pr_{B_1}(B_n)$. It takes at least $O(n^2 \cdot 2^n)$ CI tests for tiered FCI algorithm to learn all the edges between variables in $\mathbf{A}$. On the other hand, Proposition 3.2 allows us to ignore all edges among $\mathbf{A}$ and learn all edges by conditioning on at most 2 variables. Since there are $O(n^2)$ possible edge and $O(n^2)$ possible conditioned sets, Algorithm 1 takes at most $O(n^4)$ CI tests. □

T	$\|\mathbf{Z}\|$	Adjacency Precision	Adjacency Recall	Arrowhead Precision	Arrowhead Recall	Time (s)
5	Full	0.96	0.97	0.86	0.94	22.13
	0	0.97	0.97	0.88	0.95	11.89
	1	0.97	0.97	0.87	0.95	14.31
	2	0.97	0.97	0.86	0.95	16.53
	3	0.96	0.97	0.86	0.95	17.64
6	Full	0.95	0.97	0.85	0.95	32.37
	0	0.96	0.98	0.88	0.96	19.62
	1	0.96	0.97	0.86	0.95	24.57
	2	0.95	0.97	0.86	0.95	26.94
	3	0.96	0.97	0.86	0.95	27.30
7	Full	0.95	0.97	0.85	0.95	46.57
	0	0.95	0.98	0.86	0.95	26.50
	1	0.95	0.97	0.87	0.94	31.77
	2	0.95	0.98	0.86	0.95	35.16
	3	0.95	0.97	0.86	0.95	38.33
8	Full	0.94	0.97	0.84	0.94	65.83
	0	0.95	0.97	0.87	0.95	38.80
	1	0.95	0.97	0.86	0.95	45.13
	2	0.94	0.97	0.86	0.95	49.63
	3	0.94	0.97	0.86	0.95	50.09
9	Full	0.93	0.96	0.84	0.94	89.52
	0	0.94	0.97	0.86	0.94	64.60
	1	0.93	0.97	0.85	0.94	71.43
	2	0.93	0.97	0.84	0.94	74.32
	3	0.93	0.97	0.84	0.94	75.64
10	Full	0.94	0.97	0.85	0.94	123.04
	0	0.94	0.97	0.86	0.94	74.69
	1	0.94	0.97	0.85	0.94	87.89
	2	0.94	0.97	0.85	0.94	92.42
	3	0.94	0.97	0.85	0.94	99.82

Table 1: Adjacency Precision/Recall, Arrowhead Precision/Recall, and time for FCITIERS (rows with $|\mathbf{Z}| = Full$) and Algorithm 1 (rows with $|\mathbf{Z}| \in \{0, 1, 2, 3\}$) under various number of tiers (T).