On the Completeness of Causal Discovery in the Presence of Latent Confounding with Tiered Background Knowledge

Bryan Andrews
University of Pittsburgh

Peter Spirtes
Carnegie Mellon University

Gregory F. Cooper
University of Pittsburgh

Abstract

The discovery of causal relationships is a core part of scientific research. Accordingly, over the past several decades, algorithms have been developed to discover the causal structure for a system of variables from observational data. Learning ancestral graphs is of particular interest due to their ability to represent latent confounding implicitly with bi-directed edges. The well-known FCI algorithm provably recovers an ancestral graph for a system of variables encoding the sound and complete set of causal relationships identifiable from observational data.[1] Additional causal relationships become identifiable with the incorporation of background knowledge; however, it is not known for what types of knowledge FCI remains complete. In this paper, we define tiered background knowledge and show that FCI is sound and complete with the incorporation of this knowledge.

1 INTRODUCTION

Directed acyclic graphs (DAGs) have become widely studied and applied in causal modeling and discovery (Glymour and Cooper, 1999; Spirtes et al., 2000; Pearl, 2009); their simple directed structure provides an interpretable representation for causality and facilitates systematic learning procedures. Indeed, given an infinite amount of observational data from a system of variables, there exist algorithms capable of learning the set of DAGs equivalent to the true causal DAG (Spirtes et al., 2000; Chickering, 2002; Colombo and Maathuis, 2014).[2] Accordingly, any edge in common among the learned DAGs may be interpreted causally in the large sample limit. This guarantee comes at the cost of several assumptions, including the standard causal Markov and faithfulness assumptions, and the assumption that there are no unmeasured common causes—no latent confounding. The latter assumption is generally referred to as causal sufficiency.

Causal DAG learning algorithms map the implications of a graphical separation criterion (d-separation) on DAGs to conditional independence in the data. The aforementioned causal Markov and faithfulness assumptions ensure that separation in the true causal graph implies conditional independence in the data and vice versa. Unfortunately, this process is underdetermined in the sense that multiple DAGs may be consistent with the patterns of conditional independence in the data. Therefore, these algorithms learn a set of equivalent DAGs and label any edge that varies between two or more of the learned DAGs as causally ambiguous. Fortunately, background knowledge, such as specifying one variable as the cause of another, can further refine the set of DAGs, thereby increasing the number of identifiable causal relationships. In addition, causal DAG learning algorithms retain asymptotic correctness for general background knowledge (Meek, 1995).

However, the causal sufficiency assumption, required by causal DAG learning algorithms, is violated with some regularity. The fast causal inference (FCI) algorithm was developed to discover causal relationships from observational data without this assumption. When relaxing causal sufficiency, an alternative representation is needed; FCI uses maximal ancestral graphs (MAGs). MAGs are similar to DAGs, but additionally include bi-directed edges that can represent statistical associations due strictly to latent confounding. FCI maps the implications of a graphical separa-

[1] The proof of FCI's correctness assumes access to a conditional independence oracle. Thus, FCI is correct in the large sample limit using an asymptotically correct conditional independence test.

[2] DAGs are conditional independence structures; by equivalent we mean in terms of conditional independence.

Proceedings of the 23$^{\text{rd}}$International Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

tion criterion (m-separation) on MAGs to conditional independence in the data. Unfortunately, this process is underdetermined in the same sense as causal DAG learning. Therefore, FCI learns the set of equivalent MAGs consistent with the data. FCI is sound and complete in the sense that, given a conditional independence oracle, it recovers the set of MAGs that encode only and all identifiable causal relationships from observational data (Spirtes et al., 1999; Zhang, 2008). This set is represented by a summary graph called a partial ancestral graph (PAG). As with causal DAG learning, background knowledge can increase the number of identifiable causal relationships for FCI, but it is unknown for what types of knowledge FCI remains complete. In this paper, we define tiered background knowledge and show that FCI is sound and complete with the incorporation of this knowledge.

1.1 Incorporating Background Knowledge

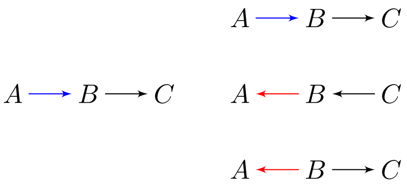

Figure 1: An example of resolving unidentifiability in causal DAG learning with background knowledge.

Figure 1 illustrates how background knowledge can resolve ambiguity in causal DAG learning. Data generated from the causal DAG on the left is consistent with the DAGs on the right. However, if we know that A causes B, shown by the blue $A \rightarrow B$ edges, then we can rule out any DAG on the right that violates this information, shown by the red $A \leftarrow B$ edges. Chris Meek extended the set of orientation rules for causal DAG learning to cases with background knowledge and proved sound and completeness (Meek, 1995). Using these rules, an algorithm can asymptotically recover the DAGs consistent with the patterns of conditional independence in the data and any background knowledge.

Figure 2 illustrates how background knowledge can resolve ambiguity in causal MAG learning. Data generated from a system represented by the causal MAG on the left is consistent with the MAGs on the right. If we know that A causes B, shown by the blue $A \rightarrow B$ edges, then we can rule out any MAG on the right that violates this information, shown by the red $A \leftrightarrow B$ edges. Unfortunately, it is unknown under what types of background knowledge FCI remain sound and complete. In fact, it is straightforward to show that FCI is not complete with background knowledge that specifies one variable as the cause of another.

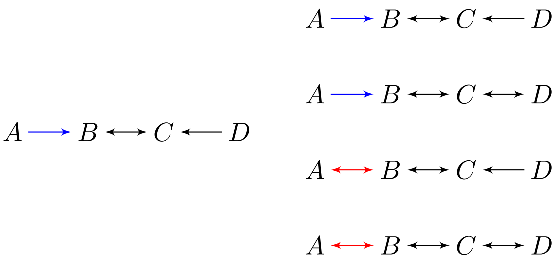

Figure 2: An example of resolving unidentifiability in causal MAG learning with background knowledge.

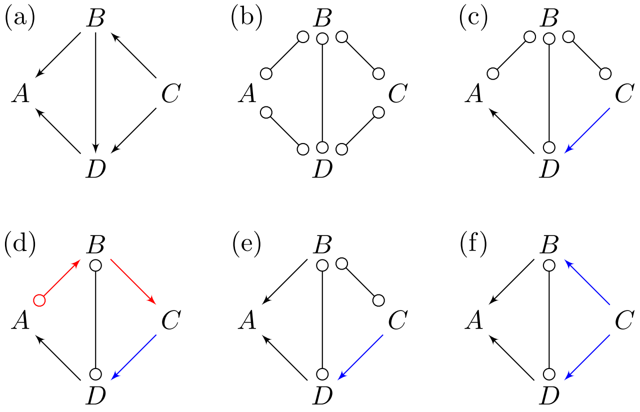

Figure 3: An example of the incompleteness of FCI with background knowledge.

Figure 3 provides an example of the incompleteness of FCI with background knowledge. In the example, FCI is not complete with background knowledge that C causes D, shown by the blue $C \rightarrow D$ edges. Data generated from a system represented by the causal MAG in (a) is consistent with all of the MAGs represented by the graph in (b), where circle edge marks represent uncertainty.[3] The graph in (b) is a PAG that summarizes the set of MAGs learned by FCI with a conditional independence oracle and no background knowledge. If we know that C causes D, then we can rule out any MAG that violates this knowledge. Accordingly, FCI with the background knowledge that C causes D returns the PAG in (c).

However, the PAG in (c) admits the possibility that the edge between A and B has an arrowhead directed into B, shown by the red $A {\circ\!\!\rightarrow} B$ edge in the graph in (d). If this is the case, then after applying FCI's orientation rules, B will be identified as a cause of C, shown by the red $B \rightarrow C$ edge in the graph in

[3] In a PAG, $X {\circ\!\!\rightarrow} Y$ means that the set of graphs represented by the PAG contains a graph with $X \rightarrow Y$ and a graph with $X \leftrightarrow Y$.

(d). However, any graph represented by the graph in (d) will violate the later discussed ancestral property and therefore cannot represent a MAG. It follows that B causes A; the PAG in (e) illustrates the sound and complete set of MAGs with the background knowledge that C causes D. Thus, FCI is not complete with background knowledge that specifies one variable as the cause of another.

1.2 Tiered Background Knowledge

In this paper, we show that FCI is sound and complete with tiered background knowledge. By tiered background knowledge, we mean any knowledge where the variables may be partitioned into two or more mutually exclusive and exhaustive subsets among which there is a known causal order.

For example, consider a data set measuring gene expression in yeast. Furthermore, suppose the data set was curated in multiple labs and that each lab was responsible for collecting both wild-type (observational) data and gene-knockout (experimental) data. If we add variables that measure the lab where the data were collected and the set of active experiments for each recorded instance, then it is reasonable to assume that these variables causally precede the gene expression variables. Thus, this scenario admits two causal tiers.

Definition *tiered background knowledge*: We say that a MAG satisfies tiered background knowledge if the variables may be partitioned into $n > 1$ disjoint subsets (tiers) $\boldsymbol{T} = \{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n\}$ and for all $A \in \boldsymbol{T}_i$ and $B \in \boldsymbol{T}_j$ such that $1 \leq i < j \leq n$

(i) A is an ancestor of B or

(ii) A and B are not adjacent.

In Figure 3, the PAG in (f) is the output of FCI with a conditional independence oracle and the tiered background knowledge that tier $\boldsymbol{T}_1 = \{C\}$ causally precedes tier $\boldsymbol{T}_2 = \{A, B, D\}$, the implications of which are shown with blue edges. Tiered background knowledge arises in many different situations including but not limited to (i) instrumental variables, (ii) data from multiple contexts and interventions, and (iii) temporal data with contemporaneous confounding.

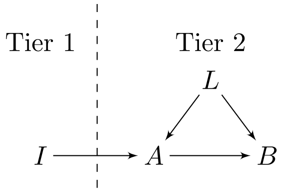

Figure 4: An example of an instrumental variable.

Figure 4 illustrates an example of an instrumental variable. Under assumptions, instrumental variables can be applied to estimate cause-effect relationships of interest (Angrist and Pischke, 2008). In Figure 4, suppose we are interested in the relationship between A and B and the following assumptions hold: (1) I is a known cause of A, (2) I affects B only through A, and (3) I and B are not confounded. The instrument I can be placed within tiered background knowledge, where tier $\boldsymbol{T}_1 = \{I\}$ causally precedes tier $\boldsymbol{T}_2 = \{A, B\}$. In this example, L is used to represent that A and B have a latent common cause. FCI will derive a PAG that is consistent with an instrumental-variable analysis, although in general it will not be as informative as the output of such an analysis.

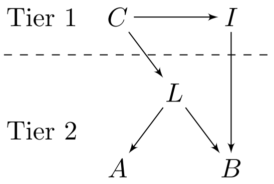

Figure 5: An example of context and interventions.

Figure 5 illustrates an example with multiple contexts and interventions. Consider again the data set measuring gene expression in yeast. In that example, each lab where a subset of the data were collected represents a different context (C) and each gene-knockout represents a different intervention (I).

Recent work has shown how to perform causal discovery using data obtained from multiple contexts and under experimental manipulation by explicitly representing the context and interventions as additional variables in the data set (Zhang et al., 2015; Magliacane et al., 2016). In this paradigm, it is often assumed that the context and interventions are exogenous with respect to the other variables. The knowledge that context C and intervention I are exogenous with respect to A and B can be encoded with tiered background knowledge, where tier $\boldsymbol{T}_1 = \{C, I\}$ causally precedes $\boldsymbol{T}_2 = \{A, B\}$. Again, L is used to represent that A and B have a latent common cause.

Figure 6 illustrates an example of temporal data with contemporaneous confounding. Applying FCI to time series data has recently become an application of interest (Entner and Hoyer, 2010; Malinsky and Spirtes, 2018). The knowledge provided by the temporal ordering can be encoded with tiered background knowledge where tier $\boldsymbol{T}_1 = \{A_1, B_1\}$ causally precedes tier $\boldsymbol{T}_2 = \{A_2, B_2\}$ which causally precedes tier $\boldsymbol{T}_3 = \{A_3, B_3\}$. Here, each L_i is used to represent that A_i and B_i have

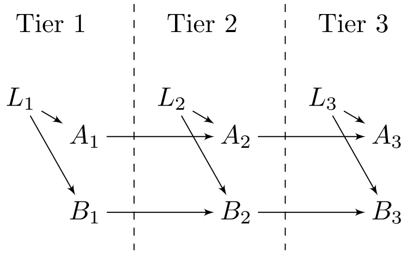

Figure 6: An example of a time series.

a contemporaneous confounder. However, if there was an edge $L_1 \to L_2$, then L_1 would be a cross-lag confounder for A_1 and B_2; in this case, tiered background knowledge would no longer be applicable (A_1 and B_2 would be adjacent, but A_1 would not be an ancestor of B_2). Nevertheless, tiered background knowledge is applicable in a wide range of scenarios.

Up to this point, we have refrained from mentioning selection bias. As it turns out, FCI is sound and complete in the presence of latent confounding and selection (Zhang, 2008). To keep the current paper tractable, we assume no selection bias and leave handling selection to future research.

1.3 Outline

In Section 2, we introduce concepts specific to latent variable modeling in order to facilitate the statement and discussion of the proofs presented in this paper. In Section 3, we state and illustrate the proof strategy for showing that FCI is sound and complete with tiered background knowledge; the details of the formal proofs appear in the supplement. In Section 4, we state our conclusions.

2 BACKGROUND

We assume the reader is familiar with fundamental graphical modeling concepts, such as DAGs and d-separation; a discussion of these concepts may be found in (Koller and Friedman, 2009). In this section, we introduce concepts specific to latent variable modeling in order to facilitate the statement and discussion of the proofs presented in this paper. Because we assume no selection bias, the definitions below are stated with no latent selection variables. More details on the concepts outlined in this section and on selection may be found in (Spirtes et al., 2000).

Definition *mixed graph*: A mixed graph is a vertex edge graph that can contain two kinds of edges: directed ($\to$) and bi-directed ($\leftrightarrow$) with at most one edge between any two variables.

Let $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$ be a mixed graph with variables $\boldsymbol{V} = \{X_1, \ldots, X_p\}$. We say that X_1 is a *parent* of X_2 and that X_2 is a *child* of X_1 if the directed edge $X_1 \to X_2$ is in $\mathcal{G}$. Similarly, we say that X_1 is a *spouse* of X_2 if the bi-directed edge $X_1 \leftrightarrow X_2$ is in $\mathcal{G}$. More generally, we say that X_1 and X_2 are *adjacent* if there is any kind of edge between X_1 and X_2 in $\mathcal{G}$.

If there is a sequence of distinct adjacent variables $\langle X_i, X_{i+1}, \ldots, X_{i+k}, \rangle$ for $k \geq 1$ in $\mathcal{G}$, then we say that there is a *path* between X_i and X_{i+k}. A *collider* occurs on a path when two edges are directed into the same variable $X_1 {*\!\!\to} X_2 \leftarrow\!\!* X_3$ and that collider is an *unshielded collider* if X_1 and X_3 are not adjacent.[4] If every edge along a path π is directed $X_{i+j} \to X_{i+j+1}$ for $0 \leq j < k$ then we say π is a *directed path* from X_i to X_{i+k}. Furthermore, we say X_i is an *ancestor* of X_{i+k} and X_{i+k} is a *descendant* of X_i. We additionally define every variable as an ancestor and descendant of itself.

Definition *ancestral graph*: A mixed graph $\mathcal{G}$ is ancestral if

(i) there are no *directed cycles*;
(ii) there are no *almost directed cycles*.

That is, if there is a directed path from X_1 to X_2, then there is neither (i) a directed edge $X_1 \leftarrow X_2$ nor (ii) a bi-directed edge $X_1 \leftrightarrow X_2$ in $\mathcal{G}$.

Definition *m-separation*: In a mixed graph, a path π between variables X and Y is *active*, or *m-connecting*, relative to a set of variables $\boldsymbol{Z}$ $(X, Y \notin \boldsymbol{Z})$ if

(i) every non-collider on π is not a member of $\boldsymbol{Z}$;
(ii) every collider on π has a descendant in $\boldsymbol{Z}$.

$\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be *m-separated* by $\boldsymbol{Z}$ if there is no active path between any $X \in \boldsymbol{X}$ and any $Y \in \boldsymbol{Y}$ relative to $\boldsymbol{Z}$.

Definition *maximal*: An ancestral graph is said to be maximal if for any two non-adjacent variables, there is a set of variables that m-separates them.

Accordingly, a MAG is a mixed graph that is both maximal and ancestral. The edges of a MAG may be interpreted as follows

> $X \to Y$ implies X is a cause of Y;[5]
>
> $X \leftrightarrow Y$ implies neither X nor Y is a cause of the other; X and Y are associated strictly due to latent confounding.

[4] An asterisk edge mark denotes that we are agnostic to that mark (it may be either a tail, arrowhead, or circle).

[5] This interpretation is not valid when allowing for selection bias.

By saying X is a cause of Y, we mean there is a directed path from X to Y in the underlying causal DAG. Furthermore, $X \rightarrow Y$ does not necessarily rule out the possibility of a latent common cause between X and Y. Algorithm 1 in (Triantafillou and Tsamardinos, 2015) provides a nice intuition for the edges of a MAG.

Let $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$ be a DAG where the variables may be partitioned into two distinct subsets $\boldsymbol{V} = \boldsymbol{O} \cup \boldsymbol{L}$ which are the observed and latent variables, respectively. A MAG $\mathcal{H}$ represents $\mathcal{G}$ over the observed variables $\boldsymbol{O}$ if and only if

(i) The variables of $\mathcal{H}$ are $\boldsymbol{O}$;

(ii) X and Y are adjacent in $\mathcal{H}$ if and only if X and Y are d-connected in $\mathcal{G}$ conditional on every subset of $\boldsymbol{O}$;

(iii) $X \rightarrow Y$ in $\mathcal{H}$ if X and Y are adjacent in $\mathcal{H}$ and X is an ancestor of Y in $\mathcal{G}$;

(iv) $X \leftrightarrow Y$ in $\mathcal{H}$ if X and Y are adjacent in $\mathcal{H}$ and X is not an ancestor of Y and vice versa in $\mathcal{G}$.

$\mathcal{H}$ represents the marginalization of $\mathcal{G}$ in the sense that if $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are distinct subsets of $\boldsymbol{O}$, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated in $\mathcal{G}$ conditioned on $\boldsymbol{Z}$ if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are m-separated in $\mathcal{H}$ conditioned on $\boldsymbol{Z}$.

Definition *Markov equivalence*: Two MAGs $\mathcal{G}$ and $\mathcal{H}$ over the same set of variables are Markov equivalent if for any three disjoint sets of variables $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}$, it holds that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are m-separated by $\boldsymbol{Z}$ in $\mathcal{H}$ if and only if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are m-separated by $\boldsymbol{Z}$ in $\mathcal{G}$. We use $[\mathcal{G}]$ denote the Markov equivalence class of MAG $\mathcal{G}$.

Definition *partial ancestral graph*: Let $\mathcal{G}$ be a MAG and $\boldsymbol{G} \subseteq [\mathcal{G}]$ be a set of MAGs Markov equivalent to $\mathcal{G}$.[6] A PAG $\mathcal{P}$ for $\boldsymbol{G}$ is a vertex edge graph with three kinds of possible edge marks and four kinds of edges: $\{\rightarrow, \leftrightarrow, \circ\!\!-\!\!\circ, \circ\!\!\rightarrow\}$, such that $\mathcal{P}$ has the same adjacencies as $\mathcal{G}$ and every non-circle edge mark in $\mathcal{P}$ occurs in every member of $\boldsymbol{G}$. If every circle edge mark in $\mathcal{P}$ corresponds to an edge mark that varies among the members of $\boldsymbol{G}$, $\mathcal{P}$ is called the maximally informative (abbreviated as m.i.) PAG for $\boldsymbol{G}$.

3 METHOD

Let $\mathcal{G} = (\boldsymbol{V}, \boldsymbol{E})$ be the DAG that represents the true underlying causal model and that the variables may be partitioned into two distinct subsets $\boldsymbol{V} = \boldsymbol{O} \cup \boldsymbol{L}$ which are the observed and latent variables, respectively. Furthermore, let $\mathcal{H}$ be the MAG that represents $\mathcal{G}$ over the observed variables $\boldsymbol{O}$. Suppose that we have been provided with tiered background knowledge and that $\boldsymbol{G} \subseteq [\mathcal{H}]$ is the set of MAGs Markov equivalent to $\mathcal{H}$ which satisfy the provided knowledge. In this section, we show that FCI using $\mathcal{H}$ as a conditional independence oracle and incorporating the provided knowledge returns the m.i. PAG for $\boldsymbol{G}$. That is, FCI is sound and complete with tiered background knowledge.

In general, our strategy is to add dummy variables to the underlying causal MAG such that the edge orientations required by tiered background knowledge are entailed by one or more applications of FCI's orientation rules. The dummy variables must be added such that running FCI on the modified graph only changes the orientations changed by incorporating our knowledge. Since FCI is sound and complete and incorporating tiered background knowledge into FCI is equivalent to running FCI on a modified graph, FCI with tiered background knowledge is sound and complete. For an example, see Figure 7.

Assumptions: The true underlying causal model for a system of variables is a DAG $\mathcal{G}$ with latent variables, but no selection. The MAG $\mathcal{H}$ that represents $\mathcal{G}$ over the observed variables satisfies the causal Markov and faithfulness conditions.

3.1 Exogenous Background Knowledge

In order to facilitate the proofs presented in this paper, we define two additional types of background knowledge.

Definition *exogenous background knowledge*: Let $\mathcal{G} = (\boldsymbol{O}, \boldsymbol{E})$ be a MAG where the variables may be partitioned into two distinct subsets $\boldsymbol{O} = \boldsymbol{A} \cup \boldsymbol{B}$. We say that $\mathcal{G}$ satisfies the exogenous background knowledge that $\boldsymbol{A}$ is exogenous with respect to $\boldsymbol{B}$, denoted $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$, if for all $A \in \boldsymbol{A}$ and $B \in \boldsymbol{B}$

(i) A is an ancestor of B or

(ii) A and B are not adjacent.

Definition *modified background knowledge*: Let $\mathcal{G} = (\boldsymbol{O}, \boldsymbol{E})$ be a MAG where the variables may be partitioned into two distinct subsets $\boldsymbol{O} = \boldsymbol{A} \cup \boldsymbol{B}$. We say that $\mathcal{G}$ satisfies modified background knowledge, denoted $mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$, if $\mathcal{G}$ satisfies $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$ and for all $A, A' \in \boldsymbol{A}$

(i) A and A' are not adjacent.

The Markov equivalence class of a MAG $\mathcal{G}$ may be constrained with $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$, denoted $[\mathcal{G}] + ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$. This repre-

[6] A PAG is usually defined for a proper Markov equivalence class rather than a subset. We define PAGs for subsets to facilitate the use of background knowledge.

sents the set of MAGs Markov equivalent to $\mathcal{G}$ that satisfy $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$. Analogously, $[\mathcal{G}] + mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$ represents the set of MAGs Markov equivalent to $\mathcal{G}$ that satisfy $mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$.

3.2 Proof Concepts

We define several graphical operations in order to facilitate the proofs. Note that if one of the operations defined below is applied to $[\mathcal{G}]$, it may be interpreted as applying that operation to every member of $[\mathcal{G}]$. Furthermore, $[\mathcal{G}]$ may be represented graphically as a m.i. PAG; the two may be thought of synonymous. For MAGs $\mathcal{G}$ and $\mathcal{H}$, we define the following concepts.[7]

Definition $\textsc{Edges}(\mathcal{G}, \boldsymbol{S})$: Let $\boldsymbol{S}$ be a subset of the variables in $\mathcal{G}$. $\textsc{Edges}(\mathcal{G}, \boldsymbol{S})$ is the subset of edges in $\mathcal{G}$ connecting two members of $\boldsymbol{S}$.

Definition $\textsc{Ins}(\mathcal{H}, \textsc{Edges}(\mathcal{G}, \boldsymbol{S}))$: Let $\boldsymbol{S}$ be a subset of the variables in common between $\mathcal{G}$ and $\mathcal{H}$. $\textsc{Ins}(\mathcal{H}, \textsc{Edges}(\mathcal{G}, \boldsymbol{S}))$ is the graph resulting from inserting $\textsc{Edges}(\mathcal{G}, \boldsymbol{S})$ into $\mathcal{H}$. That is, $\textsc{Ins}(\mathcal{H}, \textsc{Edges}(\mathcal{G}, \boldsymbol{S}))$ contains the union of the edges in $\mathcal{H}$ and $\textsc{Edges}(\mathcal{G}, \boldsymbol{S})$.

Definition $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$: Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{W}$ be three disjoint sets of variables that partition the variables in $\mathcal{G}$. $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$ is the graph constructed by removing from $\mathcal{G}$ the variables in $\boldsymbol{W}$ and the edges connecting two members of $\boldsymbol{A} \cup \boldsymbol{W}$. That is, $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$ contains the variables in $\boldsymbol{A} \cup \boldsymbol{B}$ and the edges in $\textsc{Edges}(\mathcal{G}, \boldsymbol{A} \cup \boldsymbol{B}) \setminus \textsc{Edges}(\mathcal{G}, \boldsymbol{A})$. For example, we may derive the PAG in Figure 7(c) by applying $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\cdot)$ to the PAG in Figure 7(b).

Definition $\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$: Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{W}'$ be three disjoint sets of variables that partition the variables in $\mathcal{G}$. $\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$ is the graph constructed by adding to $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$ the variables in $\boldsymbol{W} = \{W_1, W_2\}$ and the directed edges $W_1 \rightarrow A \leftarrow W_2$ for all $A \in \boldsymbol{A}$. That is, $\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})$ contains the variables in $\boldsymbol{A} \cup \boldsymbol{B} \cup \boldsymbol{W}$ and the union of the edges in $\textsc{Edges}(\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G}), \boldsymbol{A} \cup \boldsymbol{B})$ and $\{W \rightarrow A \mid A \in \boldsymbol{A}, W \in \boldsymbol{W}\}$. For example, we may derive the PAG in Figure 7(d) by applying $\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\cdot)$ to the PAG in Figure 7(b).

Definition $\textsc{Fci}(\mathcal{G})$: $\textsc{Fci}(\mathcal{G})$ is the output of running FCI using $\mathcal{G}$ as a conditional independence oracle.

Definition $\textsc{Fci}(\mathcal{G} + mbk^{\boldsymbol{A}}_{\boldsymbol{B}})$: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two disjoint sets of variables that partition the variables in $\mathcal{G}$. $\textsc{Fci}(\mathcal{G} + mbk^{\boldsymbol{A}}_{\boldsymbol{B}})$ is the output of running FCI using $\mathcal{G}$ as a conditional independence oracle and incorporating modified background knowledge $mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$. For example, the PAG in Figure 7(c) is the output of FCI with $mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$ when the MAG in Figure 7(a) is used as a conditional independence oracle. The details of running FCI with $mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$ are provided in Algorithm 2.

Definition $\mathcal{G}(\boldsymbol{S})$ (or $[\mathcal{G}](\boldsymbol{S})$): Let $\boldsymbol{S}$ be a subset of the variables in $\mathcal{G}$. $\mathcal{G}(\boldsymbol{S})$ is the induced subgraph of $\mathcal{G}$ over the subset of variables $\boldsymbol{S}$. That is, $\mathcal{G}$ contains the variables in $\boldsymbol{S}$ and the edges in $\textsc{Edges}(\mathcal{G}, \boldsymbol{S})$. For example, the dashed rectangle in Figure 7(d) denotes the induced subgraph over the variables $\{A_1, A_2, B_1, B_2, B_3, B_4\}$.

We may now formalize tiered background knowledge in terms of exogenous background knowledge. Informally, tiered background knowledge is defined by a partition of the variables $\boldsymbol{T} = \{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n\}$ where $\boldsymbol{T}_1$ is the subset of variables in the first tier, $\boldsymbol{T}_2$ is the subset of variables in the second tier, and so forth. The tiers define a causal ordering over the variables where a variable in $\boldsymbol{T}_i$ can cause a variable in $\boldsymbol{T}_j$ for for all $1 \leq i < j \leq n$, but not the other way around.

Definition *tiered background knowledge*: Let $\mathcal{G}$ be a MAG where the variables may be partitioned into $n > 1$ disjoint subsets $\boldsymbol{T} = \{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n\}$. Let $\boldsymbol{A}_i = \bigcup_{j=1}^{i-1} \boldsymbol{T}_j$, $\boldsymbol{B}_i = \boldsymbol{T}_i$, and $\boldsymbol{O}_i = \boldsymbol{A}_i \cup \boldsymbol{B}_i$. We say that $\mathcal{G}$ satisfies the tiered background knowledge given by $\boldsymbol{T}$, denoted $tbk^{\boldsymbol{T}}$, if $\mathcal{G}(\boldsymbol{O}_i)$ satisfies $ebk^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}$ for all $1 \leq i \leq n$.

The Markov equivalence class of a MAG $\mathcal{G}$ may be constrained with $tbk^{\boldsymbol{T}}$, denoted $[\mathcal{G}] + tbk^{\boldsymbol{T}}$. This represents the set of MAGs Markov equivalent to $\mathcal{G}$ that satisfy $tbk^{\boldsymbol{T}}$.

3.3 Theoretical Results

This section provides a summary of the main theoretical results, the details of which may be found in the supplement. The numbering of the lemmas and theorems follows the numbering in the supplement.

Lemma 8. Let $\mathcal{G} = (\boldsymbol{O}, \boldsymbol{E})$ be a MAG and $\boldsymbol{A}$ and $\boldsymbol{B}$ be two disjoint sets of variables that partition $\boldsymbol{O}$. If $\mathcal{G}$ satisfies $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$, then $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}([\mathcal{G}] + ebk^{\boldsymbol{A}}_{\boldsymbol{B}}) \equiv \textsc{Fci}(\mathcal{G} + mbk^{\boldsymbol{A}}_{\boldsymbol{B}})$.

This lemma shows that if $\mathcal{G}$ satisfies $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$, then running FCI using $\mathcal{G}$ as a conditional independence oracle and incorporating modified background knowledge $mbk^{\boldsymbol{A}}_{\boldsymbol{B}}$ recovers the sound and complete set of edges that connect two members of $\boldsymbol{B}$. If we do not care about the edges between the members of $\boldsymbol{A}$—this may be the case if $\boldsymbol{A}$ contains instrumental variables or $\boldsymbol{A}$ contains context and intervention variables—then the application of this lemma is interesting in its own right.

Figure 7 illustrates an example of the proof strategy for Lemma 8. In the example, $\mathcal{G}$ has variables $\boldsymbol{A} = \{A_1, A_2\}$ and $\boldsymbol{B} = \{B_1, B_2, B_3, B_4\}$ and is shown in (a). $[\mathcal{G}] + ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$ is the Markov equivalence class of $\mathcal{G}$

[7] MAGs are not necessarily closed under the defined operations; lemmas in the supplement prove when they are.

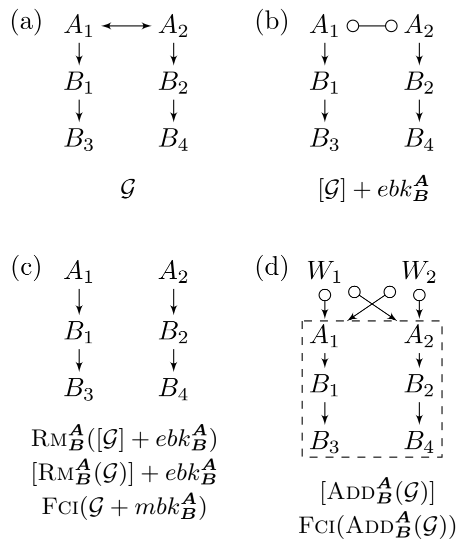

Figure 7: An illustration for Lemma 8.

constrained with $ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$ and is shown in (b). The proof is completed by showing the following equivalences:

(i) $\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}([\mathcal{G}] + ebk^{\boldsymbol{A}}_{\boldsymbol{B}}) \equiv [\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})] + ebk^{\boldsymbol{A}}_{\boldsymbol{B}}$ illustrated in Figure 7(c).

(ii) $[\textsc{Rm}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})] + ebk^{\boldsymbol{A}}_{\boldsymbol{B}} \equiv [\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})](\boldsymbol{A} \cup \boldsymbol{B})$ illustrated in Figure 7(c,d).

(iii) $[\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G})](\boldsymbol{A} \cup \boldsymbol{B}) \equiv \textsc{Fci}(\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G}))(\boldsymbol{A} \cup \boldsymbol{B})$ illustrated in Figure 7(d).

(iv) $\textsc{Fci}(\textsc{Add}^{\boldsymbol{A}}_{\boldsymbol{B}}(\mathcal{G}))(\boldsymbol{A} \cup \boldsymbol{B}) \equiv \textsc{Fci}(\mathcal{G} + mbk^{\boldsymbol{A}}_{\boldsymbol{B}})$ illustrated in Figure 7(c,d).

Lemma 12. Let $\mathcal{G} = (\boldsymbol{O}, \boldsymbol{E})$ be a MAG and $\boldsymbol{T} = \{\boldsymbol{T}_1, \dots, \boldsymbol{T}_n\}$ be a partitioning of $\boldsymbol{O}$. Let $\boldsymbol{A}_i = \bigcup_{j=1}^{i-1} \boldsymbol{T}_j$, $\boldsymbol{B}_i = \boldsymbol{T}_i$, and $\boldsymbol{O}_i = \boldsymbol{A}_i \cup \boldsymbol{B}_i$. If $\mathcal{G}$ satisfies $tbk^{\boldsymbol{T}}$, then $\textsc{Rm}^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}(([\mathcal{G}] + tbk^{\boldsymbol{T}})(\boldsymbol{O}_i)) \equiv \textsc{Fci}(\mathcal{G}(\boldsymbol{O}_i) + mbk^{\boldsymbol{A}_i}_{\boldsymbol{B}_i})$ for all $1 \leq i \leq n$.

This lemma extends the results of Lemma 8 by showing that, if $\mathcal{G}$ satisfies $tbk^{\boldsymbol{T}}$, then for all $1 \leq i \leq n$, running FCI on $\boldsymbol{O}_i$ using $\mathcal{G}(\boldsymbol{O}_i)$ as a conditional independence oracle and incorporating modified background knowledge $mbk^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}$ recovers the sound and complete set of edges that connect two members of $\boldsymbol{B}_i$. Accordingly, the proof strategy for Theorem 1 is to repeatedly apply Lemma 12 until we recover the m.i. PAG for $[\mathcal{G}]+tbk^{\boldsymbol{T}}$.

Figure 8 illustrates an example of the proof strategy for Theorem 1. The meta graph in (a) shows the general layout of a graph that satisfies tiered background knowledge. The variables are partitioned into three mutually exclusive and exhaustive subsets: $\boldsymbol{T}_1, \boldsymbol{T}_2$, and $\boldsymbol{T}_3$, with a known causal order. Figure 8, (b), (c), and (d) depicts the first, second, and third applications of Lemma 12, respectively. Each application recovers the sound and complete set of edges involving a member of $\boldsymbol{B}_i$. Therefore, since every variable is a member of $\boldsymbol{B}_i$ within some application of Lemma 12, we recover the m.i. PAG for $[\mathcal{G}] + tbk$.

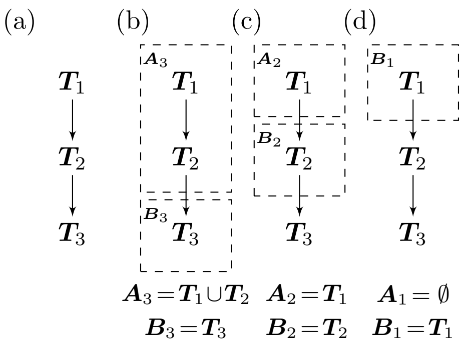

Figure 8: An illustration for Theorem 1.

Theorem 1. Let $\mathcal{G} = (\boldsymbol{O}, \boldsymbol{E})$ be a MAG and $\boldsymbol{T} = \{\boldsymbol{T}_1, \dots, \boldsymbol{T}_n\}$ be a partitioning of $\boldsymbol{O}$. If $\mathcal{G}$ satisfies $tbk^{\boldsymbol{T}}$, then FCI with $tbk^{\boldsymbol{T}}$ is sound and complete in the sense that, given a conditional independence oracle, FCI with $tbk^{\boldsymbol{T}}$ returns the m.i. PAG for $[\mathcal{G}]+tbk^{\boldsymbol{T}}$.

Let $\boldsymbol{A}_i = \bigcup_{j=1}^{i-1} \boldsymbol{T}_j$, $\boldsymbol{B}_i = \boldsymbol{T}_i$, and $\boldsymbol{O}_i = \boldsymbol{A}_i \cup \boldsymbol{B}_i$. Furthermore, let $\mathcal{P}_i = \textsc{Fci}(\mathcal{G}(\boldsymbol{O}_i) + mbk^{\boldsymbol{A}_i}_{\boldsymbol{B}_i})$ be the result of running FCI on $\boldsymbol{O}_i$ using $\mathcal{G}(\boldsymbol{O}_i)$ as a conditional independence oracle and incorporating modified background knowledge $mbk^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}$. By Lemma 12, $\mathcal{P}_i = \textsc{Rm}^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}(([\mathcal{G}] + tbk^{\boldsymbol{T}})(\boldsymbol{O}_i))$ and, accordingly, $\textsc{Edges}(\mathcal{P}_i, \boldsymbol{O}_i) \equiv \textsc{Edges}(\textsc{Rm}^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}(([\mathcal{G}]+tbk^{\boldsymbol{T}})(\boldsymbol{O}_i)), \boldsymbol{O}_i)$.

The details of running FCI with $tbk^{\boldsymbol{T}}$ are provided in Algorithm 1. By construction, FCI with $tbk^{\boldsymbol{T}}$ returns a PAG $\mathcal{P}$ with the variables in $\boldsymbol{O}$ and the edges in $\bigcup_{i=1}^{n} \textsc{Edges}(\mathcal{P}_i, \boldsymbol{O}_i)$.

Note that $\textsc{Edges}(\textsc{Rm}^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}(([\mathcal{G}] + tbk^{\boldsymbol{T}})(\boldsymbol{O}_i)), \boldsymbol{O}_i)$ is the sound and complete set of edges involving a member of $\boldsymbol{B}_i$. Therefore, since every variable in $\boldsymbol{O}$ is a member of $\boldsymbol{B}_i$ for some $1 \leq i \leq n$, $\bigcup_{i=1}^{n} \textsc{Edges}(\textsc{Rm}^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}(([\mathcal{G}] + tbk^{\boldsymbol{T}})(\boldsymbol{O}_i)), \boldsymbol{O}_i) \equiv \textsc{Edges}([\mathcal{G}]+tbk, \boldsymbol{O})$. It follows that

$$\begin{aligned}
\textsc{Edges}(\mathcal{P}, \boldsymbol{O}) &\equiv \bigcup_{i=1}^{n} \textsc{Edges}(\mathcal{P}_i, \boldsymbol{O}_i) \\
&\equiv \bigcup_{i=1}^{n} \textsc{Edges}(\textsc{Rm}^{\boldsymbol{A}_i}_{\boldsymbol{B}_i}(([\mathcal{G}] + tbk^{\boldsymbol{T}})(\boldsymbol{O}_i)), \boldsymbol{O}_i) \\
&\equiv \textsc{Edges}(\mathcal{G} + tbk^{\boldsymbol{T}}, \boldsymbol{O}).
\end{aligned}$$

Therefore, FCI with $tbk^{\boldsymbol{T}}$ is sound and complete in the sense that, given a conditional independence oracle, FCI with $tbk^{\boldsymbol{T}}$ returns the m.i. PAG for $[\mathcal{G}]+tbk^{\boldsymbol{T}}$. □

3.4 FCI with Tiered Background Knowledge

In order to formulate FCI, we first need to define the following concept.

Definition *possible d-separating set*: $X \in pds(X_i, X_j)$ if and only if $X \notin \{X_i, X_j\}$, and there is a path π between X_i and X in $\mathcal{G}$ such that for every subpath $\langle X_k, X_l, X_m \rangle$ of π either X_l is a collider on π or X_k and X_m are adjacent.

Algorithm 1 runs FCI with tiered background information. As noted above, tiered background information may be formalized in terms of exogenous background knowledge. Accordingly, Algorithm 2 is called as a subroutine where Algorithm 2 runs FCI with exogenous background knowledge. The details of both algorithms are provided below.

```
Algorithm 1: FciTiers($\boldsymbol{T}$)
Input: Disjoint sets (tiers) $\boldsymbol{T} = \{\boldsymbol{T}_1, \ldots, \boldsymbol{T}_n\}$
Output: PAG $\mathcal{P}$
Form an unconnected graph $\mathcal{P}$ over $\bigcup_{i=1}^{n} \boldsymbol{T}_i$ ;
for $i = n$ to $1$ do
    $\boldsymbol{A}_i \leftarrow \bigcup_{j=1}^{i-1} \boldsymbol{T}_j$ ;
    $\boldsymbol{B}_i \leftarrow \boldsymbol{T}_i$ ;
    $\boldsymbol{O}_i \leftarrow \boldsymbol{A}_i \cup \boldsymbol{B}_i$ ;
    $\mathcal{P}_i$ = FciExogenous($\boldsymbol{A}_i, \boldsymbol{B}_i$) ;
    Add Edges($\mathcal{P}_i, \boldsymbol{O}_i$) to $\mathcal{P}$ ;
end
return $\mathcal{P}$
```

```
Algorithm 2: FciExogenous($\boldsymbol{A}, \boldsymbol{B}$)
Input: Disjoint sets $\boldsymbol{A}$ and $\boldsymbol{B}$
Output: PAG $\mathcal{P}$
Form an unconnected graph $\mathcal{P}$ over $\boldsymbol{A} \cup \boldsymbol{B}$ ;
Add $X_i \rightarrow X_j$ to $\mathcal{P}$ for all $X_i \in \boldsymbol{A}$ and $X_j \in \boldsymbol{B}$ ;
Add $X_i \circ\!\!-\!\!\circ X_j$ to $\mathcal{P}$ for all $X_i, X_j \in \boldsymbol{B}$ ;
$n \leftarrow 0$ ;
repeat
    forall pairs of adjacent vertices $(X_i, X_j) \in \boldsymbol{A} \cup \boldsymbol{B}$ and subset $\boldsymbol{S} \subseteq adj(X_i) \setminus \{X_j\}$ s.t. $|\boldsymbol{S}| = n$ do
        if $X_i \perp\!\!\!\perp X_j | \boldsymbol{S}$ then
            Delete edge $X_i \ast\!\!-\!\!\ast X_j$ from $\mathcal{P}$ ;
            $sepset(X_i, X_j) \leftarrow \boldsymbol{S}$ ;
            $sepset(X_j, X_i) \leftarrow \boldsymbol{S}$ ;
        end
    end
    $n \leftarrow n + 1$ ;
until $n > |adj(X_i) \setminus \{X_j\}|$ for all pairs of adjacent vertices $(X_i, X_j) \in \boldsymbol{A} \cup \boldsymbol{B}$;
Apply $\mathcal{R}_0$ to $\mathcal{P}$;
forall pairs of adjacent vertices $(X_i, X_j) \in \boldsymbol{A} \cup \boldsymbol{B}$ do
    if there exists $\boldsymbol{S} \in pds(X_i, X_j)$ s.t. $X_i \perp\!\!\!\perp X_j | \boldsymbol{S}$ then
        Delete edge $X_i \ast\!\!-\!\!\ast X_j$ from $\mathcal{P}$ ;
        $sepset(X_i, X_j) \leftarrow \boldsymbol{S}$ ;
        $sepset(X_j, X_i) \leftarrow \boldsymbol{S}$ ;
    end
end
forall pairs of adjacent vertices $(X_i, X_j) \in \boldsymbol{B}$ do
    Replace edge $X_i \ast\!\!-\!\!\ast X_j$ with $X_i \circ\!\!-\!\!\circ X_j$ in $\mathcal{P}$ ;
end
Exhaustively apply $\mathcal{R}_0$ - $\mathcal{R}_4$, $\mathcal{R}_8$ - $\mathcal{R}_{10}$ to $\mathcal{P}$ ;
return $\mathcal{P}$
```

In each iteration of the main loop in Algorithm 1, there are two sets of variables: $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ where $\boldsymbol{A}_i$ is exogenous with respect to $\boldsymbol{B}_i$. Algorithm 2 runs with $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ as input to obtain PAG $\mathcal{P}_i$. The edges Edges($\mathcal{P}_i, \boldsymbol{O}_i$) are added to the final graph $\mathcal{P}$. After looping through every combination of $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$, $\mathcal{P}$ contains the sound and complete set of edges.

Algorithm 2 is identical to the standard FCI algorithm with the exception that adjacencies between the members of $\boldsymbol{A}$ are forbidden and edges between any $A \in \boldsymbol{A}$ and $B \in \boldsymbol{B}$ are oriented $A \rightarrow B$. Orientation rules ($\mathcal{R}_0$ - $\mathcal{R}_4$, $\mathcal{R}_8$ - $\mathcal{R}_{10}$) may be found in the supplement.

4 CONCLUSIONS

In causal structure discovery, multiple graphs are often consistent with the patterns of conditional independence in the data. Generally, these patterns do not precisely resolve all of the causal relationships among the observed variables. Fortunately, the addition of background knowledge can often help resolve unresolved causal relationships. However, it was not known when FCI, a causal discovery algorithms that can model latent confounding, is sound and complete with background knowledge.

In this paper, we show that FCI is sound and complete when given tiered background knowledge. Tiered background knowledge arises in many different situations, including but not limited to instrumental variables, data from multiple contexts and interventions, and temporal data with contemporaneous-confounding.

The proof that FCI is complete provides algorithm developers with an assurance that this aspect of the algorithm's development is finished. It also provides users with knowledge that the algorithm is able to find all of the causal relationships that are identifiable from tiered background knowledge and observational data, under the typical assumptions.

Acknowledgements

The research reported in this paper was supported by grant U54HG008540 awarded by the National Human Genome Research Institute through funds provided by the trans-NIH Big Data to Knowledge (BD2K) initiative, grant #4100070287 from the Pennsylvania Department of Health (DOH), and grant IIS-1636786 from the National Science Foundation. Author BA also received support from training grant T15LM007059 from the National Library of Medicine. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of these funding agencies.

References

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Chickering, D. M. (2002). Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554.

Colombo, D. and Maathuis, M. H. (2014). Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research*, 15:3741–3782.

Entner, D. and Hoyer, P. O. (2010). On causal discovery from time series data using FCI. In *Proceedings of the European Workshop on Probabilistic Graphical Models*, pages 121–128.

Glymour, C. and Cooper, G. F. (1999). *Computation, Causation, and Discovery*. MIT Press.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT press.

Magliacane, S., Claassen, T., and Mooij, J. M. (2016). Joint causal inference on observational and experimental datasets. *arXiv preprint arXiv:1611.10351*.

Malinsky, D. and Spirtes, P. (2018). Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of the ACM SIGKDD Workshop on Causal Discovery*, pages 23–47.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 403–410.

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

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press.

Spirtes, P., Meek, C., and Richardson, T. S. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In Glymour, C. and Cooper, G. F., editors, *Computation, Causation, and Discovery*. MIT Press.

Triantafillou, S. and Tsamardinos, I. (2015). Constraint-based causal discovery from multiple interventions over overlapping variable sets. *Journal of Machine Learning Research*, 16:2147–2205.

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

Zhang, K., Huang, B., Zhang, J., Schölkopf, B., and Glymour, C. (2015). Discovery and visualization of nonstationary causal models. *arXiv preprint arXiv:1509.08056*.