

# On the Representation of Pairwise Causal Background Knowledge and Its Applications in Causal Inference

Zhuangyan Fang<sup>1,2</sup>

Ruiqi Zhao<sup>1,3</sup>

Yue Liu<sup>4</sup>

Yangbo He<sup>1\*</sup>

FANGZHUANGYAN@XIAOMI.COM

ZHAORUIQI02@INSUR.COM

LIUYUE\_STATS@RUC.EDU.CN

HEYB@MATH.PKU.EDU.CN

<sup>1</sup> School of Mathematics, Peking University, Beijing, China

<sup>2</sup> Group Information Technology Department, Xiaomi Corporation, Beijing, China

<sup>3</sup> Inspur Industrial Innovation (Shandong) Intelligent Manufacturing Co., Ltd., Shandong, China

<sup>4</sup> Center for Applied Statistics and School of Statistics, Renmin University of China, Beijing, China

**Editor:** Ilya Shpitser

## Abstract

Pairwise causal background knowledge about the existence or absence of causal edges and paths is frequently encountered in observational studies. Such constraints allow the shared directed and undirected edges in the constrained subclass of Markov equivalent DAGs to be represented as a causal maximally partially directed acyclic graph (MPDAG). In this paper, we first provide a sound and complete graphical characterization of causal MPDAGs and introduce a minimal representation of a causal MPDAG. Then, we give a unified representation for three types of pairwise causal background knowledge, including direct, ancestral and non-ancestral causal knowledge, by introducing a novel concept called direct causal clause (DCC). Using DCCs, we study the consistency and equivalence of pairwise causal background knowledge and show that any pairwise causal background knowledge set can be uniquely and equivalently decomposed into the causal MPDAG representing the refined Markov equivalence class and a minimal residual set of DCCs. Polynomial-time algorithms are also provided for checking consistency and equivalence, as well as for finding the decomposed MPDAG and the residual DCCs. Finally, with pairwise causal background knowledge, we prove a sufficient and necessary condition to identify causal effects and surprisingly find that the identifiability of causal effects only depends on the decomposed MPDAG. We also develop a local IDA-type algorithm to estimate the possible values of an unidentifiable effect. Simulations suggest that pairwise causal background knowledge can significantly improve the identifiability of causal effects.

**Keywords:** causal graphical model, pairwise causal background knowledge, Markov equivalence class, causal inference, identifiability, IDA

## 1. Introduction

Causal background knowledge refers to the understanding or consensus of causal and non-causal relations in a system. Such information, as a supplement to data, may be obtained from domain knowledge or experts' judgments (such as smoking causes lung cancer and eating betel nuts causes oral cancer), from common sense (such as a subsequent event is not

---

\*. Corresponding author.

a cause of a prior event), or even from previous experimental studies (such as double-blind experiments or A/B tests). Under the framework of causal graphical models, representing and exploiting causal background knowledge may improve the identifiability of causal structures or causal effects in a study of causal discovery or causal inference (Meek, 1995; Perković, 2020). For example, as shown in Figure 1, consider a simple causal chain with three binary variables: smoking, bronchitis and dyspnea. With observational data only, it is possible to consistently estimate a completed partially directed acyclic graph (CPDAG) shown in Figure 1a, representing a set of statistically equivalent DAGs called Markov equivalent. In this case, neither the causal structure among the three vertices nor the causal effect of smoking on bronchitis is identifiable, since there are three Markov equivalent DAGs and the causal effects estimated based on each of them are not identical. However, if we have already known that smoking can cause dyspnea, then there is only one DAG in the Markov equivalence class satisfying this causal constraint and thus the causal effect of smoking on bronchitis is identifiable.

This paper focuses on representing *pairwise* causal background knowledge and incorporating this knowledge into causal inference assuming no hidden variables or selection biases. We consider three types of pairwise causal background knowledge, including direct, ancestral and non-ancestral causal knowledge. A piece of direct causal knowledge is defined as the presence of a directed edge in a DAG. Direct causal knowledge is natural and has been studied extensively (Dor and Tarsi, 1992; Meek, 1995; Perković et al., 2017; Henckel et al., 2022; Perković, 2020; Witte et al., 2020; Guo and Perković, 2021). A piece of ancestral (non-ancestral) causal knowledge is defined as the presence (absence) of a directed path in a causal DAG. One can learn ancestral (non-ancestral) causal relations from observational data (Fang et al., 2022), or from interventional data since when a variable is intervened, its descendant variables could be changed while other variables usually keep unchanged (He and Geng, 2008). Each non-ancestral relation essentially implies a causal topological order between two variables. Thus, a causal topological order among variables, which is a common type of causal background knowledge in literature (Park and Klabjan, 2017; Wang and Michailidis, 2019), can be translated to pairwise non-ancestral relations equivalently: each variable is not an ancestral variable of its preceding variables in the order.

Existing works on pairwise causal background knowledge mainly focus on direct causal knowledge (Dor and Tarsi, 1992; Meek, 1995; Perković et al., 2017; Henckel et al., 2022; Perković, 2020; Witte et al., 2020; Guo and Perković, 2021). Meek (1995) proved that the set of DAGs in a Markov equivalence class satisfying given direct causal knowledge is non-empty if and only if it can be represented by a causal maximally partially directed acyclic graph (MPDAG), which contains both directed and undirected edges. Benefiting from the compact graphical representation, many researchers discussed the identifiability and efficient estimation of a causal effect, or the estimation of all possible causal effects of a treatment on a response with direct causal knowledge (Perković et al., 2017; Henckel et al., 2022; Perković, 2020; Witte et al., 2020; Guo and Perković, 2021). Recently, Fang and He (2020) further studied non-ancestral causal knowledge and proved that non-ancestral causal knowledge can also be represented exactly by causal MPDAGs. However, causal MPDAGs may fail to represent ancestral causal knowledge. The DAGs in a Markov equivalence class satisfying given ancestral causal knowledge may satisfy some structural constraints that cannot be posed by any causal MPDAG. An example is provided in Example 1.



Figure 1: A CPDAG over three variables including smoking, bronchitis and dyspnea is given in Figure 1a, which represents the Markov equivalent class shown in Figures 1b.

Instead, the representation of ancestral causal knowledge remains under-explored. In the existing studies, ancestral causal knowledge is generally regarded as a constraint on the existence of directed paths (see, for example, Borboudakis and Tsamardinos, 2012). Such a constraint is *global* in the sense that it imposes complex restrictions on the direction of all edges along the paths connecting two nodes. As a result, it becomes challenging to answer some basic queries, such as whether two pieces of ancestral causal knowledge contradict each other, or whether one piece of ancestral causal knowledge can be inferred from others. Existing literature often addresses these queries, as well as other issues related to ancestral causal knowledge, by explicitly or implicitly enumerating all equivalent DAGs within a given Markov equivalence class and examining the paths in each DAG (Borboudakis and Tsamardinos, 2012). However, this enumeration-based approach is infeasible in high dimensional settings, where the number of Markov equivalent DAGs grows exponentially with the number of variables. This limitation also restricts the practical application of pairwise causal background knowledge in causal inference.

In this paper, we first provide a sound and complete graphical characterization of causal MPDAGs, together with their minimal representation. We establish sufficient and necessary conditions under which a partially directed graph qualifies as a causal MPDAG. Next, we introduce a novel representation of pairwise causal background knowledge, called direct causal clauses, which provide a unified way to represent direct, ancestral, and non-ancestral knowledge within a single framework. We further analyze the consistency and equivalence of pairwise causal background knowledge represented by direct causal clauses, and show that any set of pairwise causal background knowledge can be uniquely and equivalently decomposed into (i) the causal MPDAG representing the subset of Markov equivalent DAGs consistent with the given knowledge, and (ii) a minimal residual set of direct causal clauses. We also provide sufficient and necessary conditions under which an MPDAG can exactly represent pairwise causal background knowledge.

Leveraging direct causal clauses, we then propose polynomial-time algorithms for checking the consistency and equivalence of pairwise causal background knowledge, as well as for constructing the corresponding decomposed causal MPDAG and residual direct causal clauses. To the best of our knowledge, these are the first polynomial-time algorithms that address these tasks in the presence of all three types of pairwise causal knowledge.

Finally, we investigate the identifiability of causal effects when pairwise causal background knowledge is available. We find, perhaps surprisingly, that the identifiability depends only on the decomposed MPDAG derived from the background knowledge. When a causal effect is not identifiable, its possible values are determined jointly by the MPDAG

and the residual set of direct causal clauses. For such cases, we develop IDA-type algorithms to estimate the possible effects locally or semi-locally, building on new local orientation rules for CPDAGs with direct causal clauses.

The following three subsections devote to some preliminaries. Unless otherwise stated, we use capital letters such as  $X$  to denote variables or vertices or nodes, and use boldface letters like  $\mathbf{X}$  to denote variable sets or vectors. An instantiation of a variable or vector is denoted by a lowercase letter, such as  $x$  and  $\mathbf{x}$ . We use  $\mathbf{X} \subseteq \mathbf{Y}$ ,  $\mathbf{X} \subsetneq \mathbf{Y}$  and  $\mathbf{X} \not\subseteq \mathbf{Y}$  to denote that  $\mathbf{X}$  is a subset, proper subset and not a subset of  $\mathbf{Y}$ , respectively.

### 1.1 Causal Graphical Models

In this paper, we use  $\mathbf{V}(\mathcal{G})$ ,  $\mathbf{E}(\mathcal{G})$ ,  $\mathbf{E}_d(\mathcal{G})$ , and  $\mathbf{E}_u(\mathcal{G})$  to denote the vertex set (or node set), edge set, set of directed edges, and set of undirected edges of a given graph  $\mathcal{G}$ , respectively. Here,  $\mathcal{G}$  can be a directed, undirected or partially directed graph. The skeleton of  $\mathcal{G}$  is the undirected graph obtained by removing all arrowheads from  $\mathcal{G}$ . For any  $\mathbf{V}' \subseteq \mathbf{V}$ , the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}'$ , denoted by  $\mathcal{G}(\mathbf{V}')$ , is the graph with vertex set  $\mathbf{V}'$  and edge set  $\mathbf{E}' \subseteq \mathbf{E}$  containing all and only edges between vertices in  $\mathbf{V}'$ . The undirected subgraph and directed subgraph of  $\mathcal{G}$  are denoted by  $\mathcal{G}_u$  and  $\mathcal{G}_d$ , respectively. The former is defined as the undirected graph obtained by removing all directed edges, while the latter is the directed graph obtained by removing all undirected edges from  $\mathcal{G}$ . An undirected (or directed) induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}' \subseteq \mathbf{V}$  is the induced subgraph of  $\mathcal{G}_u$  (or  $\mathcal{G}_d$ ) over  $\mathbf{V}'$ .

In a graph  $\mathcal{G}$ ,  $X_i$  is a parent of  $X_j$  and  $X_j$  is a child of  $X_i$  if  $X_i \rightarrow X_j$ , and  $X_i$  is a sibling of  $X_j$  if  $X_i - X_j$ . Two vertices  $X_i$  and  $X_j$  are adjacent and called neighbors of each other if they are connected by an edge. We use  $pa(X_i, \mathcal{G})$ ,  $ch(X_i, \mathcal{G})$ ,  $sib(X_i, \mathcal{G})$ , and  $adj(X_i, \mathcal{G})$  to denote the sets of parents, children, siblings, and adjacent vertices of  $X_i$  in  $\mathcal{G}$ , respectively. A graph is called complete if every two distinct vertices are adjacent. A vertex is simplicial if its neighbors induce a complete subgraph.

A path is a sequence of distinct vertices  $(X_1, \dots, X_n)$  such that any two consecutive vertices are adjacent.  $X_1$  and  $X_n$  are endpoints and the others are intermediate nodes. A path connecting  $X \in \mathbf{X}$  and  $Y \in \mathbf{Y}$  is proper if the intermediate nodes on the path are not in  $\mathbf{X} \cup \mathbf{Y}$ . If every two distinct vertices in a graph are connected by a path, then the graph is connected. A path from  $X_1$  to  $X_n$  is partially directed if  $X_i \leftarrow X_{i+1}$  does not occur in  $\mathcal{G}$  for any  $i = 1, \dots, n-1$  and  $X_i \rightarrow X_{i+1}$  for some  $i = 1, \dots, n-1$ . Moreover, a path from  $X_1$  to  $X_n$  is possibly causal if  $X_i \leftarrow X_j$  does not occur in  $\mathcal{G}$  for any  $i, j = 1, \dots, n$  and  $i < j$ , and is non-causal otherwise (Perković et al., 2017). A path from  $X_1$  to  $X_n$  is directed if  $X_i \rightarrow X_{i+1}$  for every  $i = 1, \dots, n-1$ , and is undirected if  $X_i - X_{i+1}$  for every  $i = 1, \dots, n-1$ . A partially directed (directed, or undirected) cycle is a partially directed (directed, or undirected) path from  $X_1$  to  $X_n$  together with a directed or an undirected edge (a directed edge, or an undirected edge) from  $X_n$  to  $X_1$ . A directed graph is acyclic (DAG) if there are no directed cycles. A partially directed acyclic graph (PDAG) is a partially directed graph without directed cycles. A chain graph is a partially directed graph with no partially directed cycles (Lauritzen and Richardson, 2002). The length of a path (cycle) is the number of edges on the path (cycle). A vertex  $X_i$  is an ancestor of  $X_j$  and  $X_j$  is a descendant of  $X_i$  if there is a directed path from  $X_i$  to  $X_j$  or  $X_i = X_j$ ; the sets of ancestors

and descendants of  $X_i$  in  $\mathcal{G}$  are denoted by  $an(X_i, \mathcal{G})$  and  $de(X_i, \mathcal{G})$ , respectively. A vertex  $X_j$  is a possible descendant of  $X_i$  if there is a possibly causal path from  $X_i$  to  $X_j$ . A chord of a path (cycle) is an edge joining two nonconsecutive vertices on the path (cycle). An undirected graph is chordal if it has no chordless cycle with length greater than three.

Let  $\pi = (X_1, \dots, X_n)$  be a path in  $\mathcal{G}$ .  $X_i$  ( $i \neq 1, n$ ) is a collider on  $\pi$  if  $X_{i-1} \rightarrow X_i \leftarrow X_{i+1}$ , and is a definite non-collider on  $\pi$  if  $X_{i-1} \leftarrow X_i$ , or  $X_i \rightarrow X_{i+1}$ , or  $X_{i-1} - X_i - X_{i+1}$  but  $X_{i-1}$  is not adjacent to  $X_{i+1}$ . Moreover,  $X_i$  is of definite status on  $\pi$  if it is a collider, or a definite non-collider, or an endpoint on  $\pi$  (Guo and Perković, 2021). A path  $\pi$  is of definite status if its nodes are of definite status. For distinct vertices  $X_i, X_j$  and  $X_k$ , if  $X_i \rightarrow X_j \leftarrow X_k$  and  $X_i$  is not adjacent to  $X_k$  in  $\mathcal{G}$ , the triple  $(X_i, X_j, X_k)$  is called a v-structure collided on  $X_j$ . A definite status path  $\pi$  from  $X$  to  $Y$  is d-separated (blocked) by  $\mathbf{Z}$  ( $X, Y \notin \mathbf{Z}$ ) if  $\pi$  has a definite non-collider in  $\mathbf{Z}$  or  $\pi$  has no collider who has a descendant in  $\mathbf{Z}$ , and is d-connected given  $\mathbf{Z}$  otherwise.

Two DAGs are Markov equivalent if they induce the same d-separation relations. Pearl et al. (1989) proved that two DAGs are equivalent if and only if they have the same skeleton and the same v-structures. A Markov equivalence class contains all DAGs equivalent to each other. A Markov equivalence class can be uniquely represented by a completed PDAG, or essential graph, defined as follows:

**Definition 1 (Completed PDAG, Andersson et al. 1997)** *Given a DAG  $\mathcal{G}$ , the completed PDAG (CPDAG) of  $\mathcal{G}$ , denoted by  $\mathcal{G}^*$ , is a PDAG that has the same skeleton as  $\mathcal{G}$ , and a directed edge occurs in  $\mathcal{G}^*$  if and only if it appears in all equivalent DAGs of  $\mathcal{G}$ .*

We assume that the CPDAG  $\mathcal{G}^*$  of the Markov equivalence class containing the underlying DAG  $\mathcal{G}$  is provided, and use  $[\mathcal{G}]$  or  $[\mathcal{G}^*]$  to represent the Markov equivalence class. Andersson et al. (1997) proved that a CPDAG is a chain graph, and its undirected subgraph is the union of disjoint connected chordal graphs, which are called chain components. A causal DAG model consists of a DAG  $\mathcal{G}$  and a distribution  $f$  over the same set  $\mathbf{V}$  such that  $f(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i | pa(x_i, \mathcal{G}))$ .

## 1.2 Pairwise Causal Background Knowledge

In this paper, we mainly consider *pairwise* causal background knowledge, which can be formally defined in terms of constraints as follows.

**Definition 2 (Pairwise Causal Constraints)** *A direct causal constraint denoted by  $X \rightarrow Y$  is a proposition saying that  $X$  is a parent of  $Y$ , that is,  $X$  is a direct cause of  $Y$ . An ancestral causal constraint denoted by  $X \dashrightarrow Y$  is a proposition saying that  $X$  is an ancestor of  $Y$ , that is,  $X$  is a cause of  $Y$ . A non-ancestral causal constraint denoted by  $X \dashv\rightarrow Y$  is a proposition saying that  $X$  is not an ancestor of  $Y$ , that is,  $X$  is not a cause of  $Y$ . Moreover,  $X$  is called the tail and  $Y$  is called the head in the above notions.*

Non-pairwise causal background knowledge will be briefly discussed in Section 3.3. A pairwise causal constraint set is also called a (causal) background knowledge set for short. A pairwise causal constraint set over  $\mathbf{V}$  consists of some of the constraints with heads and tails in  $\mathbf{V}$ . Given a DAG  $\mathcal{G}$ , a pairwise causal constraint set  $\mathcal{B}$  over  $\mathbf{V}(\mathcal{G})$  is said to hold for  $\mathcal{G}$ , or equivalently,  $\mathcal{G}$  is said to satisfy  $\mathcal{B}$ , if every proposition in  $\mathcal{B}$  is true for  $\mathcal{G}$ . We define the restricted Markov equivalence class induced by  $\mathcal{G}^*$  and  $\mathcal{B}$  as follows.

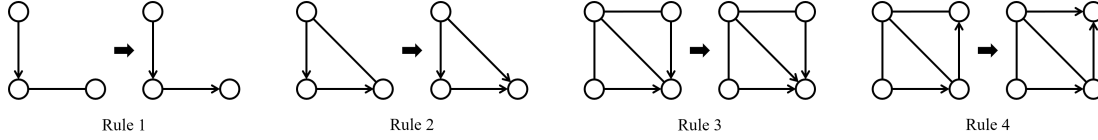


Figure 2: A visualization of Meek's rules. If the graph on the left-hand side of a rule is an induced subgraph of a PDAG, then orient the undirected edge such that the resulting subgraph is the one on the right-hand side of the rule.

**Definition 3 (Restricted Markov Equivalence Class)** *The restricted Markov equivalence class induced by a CPDAG  $\mathcal{G}^*$  and a pairwise causal constraint set  $\mathcal{B}$  over  $\mathbf{V}(\mathcal{G}^*)$ , denoted by  $[\mathcal{G}^*, \mathcal{B}]$ , consists of all equivalent DAGs in  $[\mathcal{G}^*]$  that satisfy  $\mathcal{B}$ .*

If  $\mathcal{B}$  is empty, then  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{G}^*]$ . A restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{B}]$  is empty if none of the DAGs in  $[\mathcal{G}^*]$  satisfies  $\mathcal{B}$ . For example, if two exclusive constraints, say both  $X \dashrightarrow Y$  and  $X \dashleftarrow Y$ , appear in  $\mathcal{B}$ , we have that  $[\mathcal{G}^*, \mathcal{B}] = \emptyset$ . Conversely, we say that  $\mathcal{B}$  is consistent with  $\mathcal{G}^*$  if  $[\mathcal{G}^*, \mathcal{B}] \neq \emptyset$ .

A PDAG is maximal (MPDAG) if it is closed under the four Meek's rules shown in Figure 2 (Meek, 1995). Given a CPDAG  $\mathcal{G}^*$  and a pairwise causal constraint set  $\mathcal{B}$  consistent with  $\mathcal{G}^*$ , the causal MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$  is defined as follows.

**Definition 4 (Causal MPDAG)** *The MPDAG  $\mathcal{H}$  of a non-empty restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{B}]$  induced by a CPDAG  $\mathcal{G}^*$  and a pairwise causal constraint set  $\mathcal{B}$  is a PDAG such that (1)  $\mathcal{H}$  has the same skeleton and v-structures as  $\mathcal{G}^*$ , and (2) an edge is directed in  $\mathcal{H}$  if and only if it appears in all DAGs in  $[\mathcal{G}^*, \mathcal{B}]$ . An MPDAG  $\mathcal{H}$  is a causal MPDAG if there exists a CPDAG  $\mathcal{G}^*$  and a pairwise causal constraint set  $\mathcal{B}$  (possibly empty) consistent with  $\mathcal{G}^*$  such that  $\mathcal{H}$  is the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ .*

It is easy to verify that Definition 4 indeed defines an MPDAG. Clearly, the MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{B}]$  contains the common direct causal relations of all restricted Markov equivalent DAGs in  $[\mathcal{G}^*, \mathcal{B}]$ . Let  $[\mathcal{H}]$  be the set of DAGs which contain all directed edges of  $\mathcal{H}$  and have the same skeleton and v-structures as  $\mathcal{H}$ . Following Definition 4, if  $\mathcal{H}$  is the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ , then every DAG in  $[\mathcal{G}^*, \mathcal{B}]$  belongs to  $[\mathcal{H}]$ , that is,  $[\mathcal{G}^*, \mathcal{B}] \subseteq [\mathcal{H}]$ . An example illustrating that  $[\mathcal{G}^*, \mathcal{B}]$  may be a proper subset of  $[\mathcal{H}]$  is shown by Example 1.

**Example 1** *Figure 3a shows a CPDAG  $\mathcal{G}^*$ , and  $\mathcal{G}_1$  to  $\mathcal{G}_4$  shown in Figures 3b to 3e are DAGs in  $[\mathcal{G}^*]$  satisfying the ancestral causal constraint  $\mathcal{B} = \{X \dashrightarrow Y\}$ . That is, the restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{B}] = \{\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4\}$ . The causal MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{B}]$  is shown in Figure 3f, which has two directed edges  $A \rightarrow Y$  and  $B \rightarrow Y$  as they are both in  $\mathcal{G}_1$  to  $\mathcal{G}_4$ . On the other hand,  $[\mathcal{H}]$  consists of  $\mathcal{G}_1$  to  $\mathcal{G}_6$ , meaning that  $[\mathcal{G}^*, \mathcal{B}] \subsetneq [\mathcal{H}]$ . In summary,  $\mathcal{B}$  implies two direct causal relations,  $A \rightarrow Y$  and  $B \rightarrow Y$ , as well as a constraint that  $X$  is a direct cause of either  $A$  or  $B$ . The MPDAG  $\mathcal{H}$  does not imply the latter constraint.*

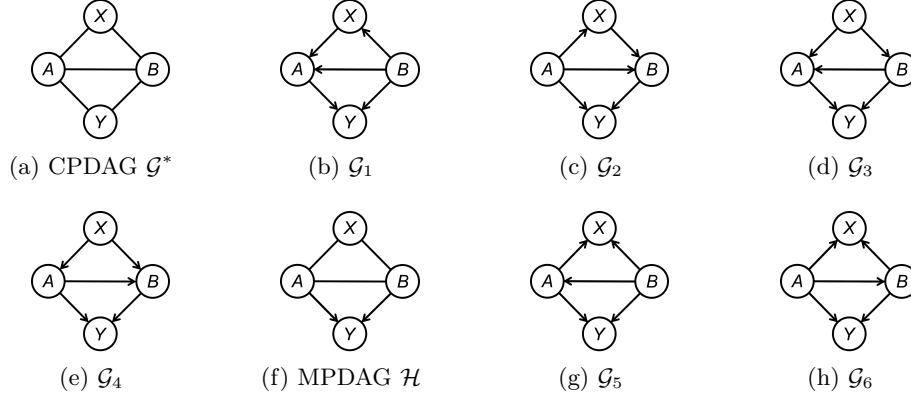


Figure 3: An example of a causal MPDAG. Let  $\mathcal{B} = \{X \dashrightarrow Y\}$  and  $\mathcal{H}$  be the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ .  $[\mathcal{G}^*, \mathcal{B}]$  consists of the DAGs from  $\mathcal{G}_1$  to  $\mathcal{G}_4$ , while  $[\mathcal{H}]$  consists of the DAGs from  $\mathcal{G}_1$  to  $\mathcal{G}_6$ , which indicates that  $[\mathcal{G}^*, \mathcal{B}] \subsetneq [\mathcal{H}]$ .

**Definition 5 (Fully Informative MPDAG)** *A causal MPDAG  $\mathcal{H}$  is fully informative with respect to a restricted Markov equivalence class if the set  $[\mathcal{H}]$  is identical to the restricted Markov equivalence class.*

When  $\mathcal{H}$  is fully informative with respect to  $[\mathcal{G}^*, \mathcal{B}]$ ,  $\mathcal{H}$  can represent  $[\mathcal{G}^*, \mathcal{B}]$  exactly. When  $\mathcal{B}$  only contains direct causal constraints, Meek (1995) proved that the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$  is fully informative, and the MPDAG can be constructed in polynomial time using Meek’s rules. Fang and He (2020) further showed that any non-ancestral causal constraint can be represented equivalently by some direct causal constraints and the corresponding fully informative  $\mathcal{H}$  can also be constructed from  $\mathcal{G}^*$  and  $\mathcal{B}$  in polynomial time.

### 1.3 Intervention Calculus

In order to obtain the effect of an intervention on a response variable, Pearl (2009) employed the notion of *do-operator* to formulate the post-intervention distribution as follows: given a DAG  $\mathcal{G}$  over the vertex set  $\mathbf{V} = \{X_1, \dots, X_n\}$  and  $\mathbf{X} \subseteq \mathbf{V}$ ,

$$f(\mathbf{v} \mid do(\mathbf{X} = \mathbf{x})) = \begin{cases} \prod_{X_i \in \mathbf{V} \setminus \mathbf{X}} f(x_i \mid pa(x_i, \mathcal{G}))|_{\mathbf{x}=\mathbf{x}}, & \text{if } \mathbf{v}|_{\mathbf{X}} = \mathbf{x}, \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Here,  $f(\mathbf{v} \mid do(\mathbf{X} = \mathbf{x}))$  (or  $f(\mathbf{v} \mid do(\mathbf{x}))$  for short) is the post-intervention distribution over  $\mathbf{V}$  after intervening on  $\mathbf{X}$ , by forcing  $\mathbf{X}$  to equal  $\mathbf{x}$ ;  $\mathbf{v}$  is an instantiation of  $\mathbf{V}$ ;  $\mathbf{v}|_{\mathbf{X}} = \mathbf{x}$  means the value of  $\mathbf{X}$  in the instantiation  $\mathbf{v}$  equals  $\mathbf{x}$ . The post-intervention distribution  $f(\mathbf{y} \mid do(\mathbf{x}))$  is defined by integrating out all variables other than  $\mathbf{Y}$  in  $f(\mathbf{v} \mid do(\mathbf{x}))$ . Given a treatment set  $\mathbf{X}$  and a response set  $\mathbf{Y}$ , if there exists an  $\mathbf{x} \neq \mathbf{x}'$  such that  $f(\mathbf{y} \mid do(\mathbf{x})) \neq f(\mathbf{y} \mid do(\mathbf{x}'))$ , then  $\mathbf{X}$  has a causal effect on  $\mathbf{Y}$  (Pearl, 2009). Following the notion of Pearl (2009), we simply use  $f(\mathbf{y} \mid do(\mathbf{x}))$  to represent the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$ .

Given the underlying causal DAG, the post-intervention distribution can be calculated from observational distribution by using a number of criteria. For example, the post-

intervention distribution of a single response  $Y \notin pa(X, \mathcal{G})$  after intervening on a single treatment  $X$  can be calculated by

$$f(y | do(x)) = \int f(y | X = x, pa(X, \mathcal{G}) = u) f(u) du. \quad (2)$$

If  $Y \in pa(X, \mathcal{G})$ , then  $f(y | do(x)) = f(y | do(x'))$  for any two instantiations  $x, x'$  of  $X$ . Equation (2) is a special case of the backdoor adjustment (Pearl, 1995, 2009), and  $pa(X, \mathcal{G})$  is a backdoor adjustment set. However, if the underlying DAG is not fully known,  $f(\mathbf{v} | do(\mathbf{x}))$  may not be *identifiable* (Pearl, 2009). Recently, the identifiability of a causal effect given an MPDAG has been studied (Perković et al., 2017; Perković, 2020). Perković (2020) proved that  $f(\mathbf{v} | do(\mathbf{x}))$  is identifiable if and only if every proper possibly causal path from  $\mathbf{X}$  to  $\mathbf{Y}$  starts with a directed edge in the given MPDAG.

If a causal effect is not identifiable, we can use the IDA framework to estimate all possible causal effects. The original IDA enumerates all possible causal effects of a single treatment  $X$  on a single response  $Y$  given a CPDAG by listing all possible parental sets of  $X$  and adjusting for each of them. To decide whether a set of variables is possible to be the parents of  $X$ , Maathuis et al. (2009, Lemma 3.1) provided a locally valid orientation rule. Recently, Fang and He (2020, Theorem 1) extended the locally valid orientation rule to MPDAGs and proposed a fully local extension of IDA to deal with direct causal and non-ancestral causal constraints. For multiple interventions, Nandy et al. (2017) proposed the joint-IDA. Compared with IDA, this extension is semi-local, which uses Meek’s rules to check the validity of each candidate parental set. However, Meek’s rules are global in the sense that they require an entire PDAG as input. Perković et al. (2017) further extended the joint-IDA to MPDAGs, and the algorithm is called the semi-local IDA. The recent work on efficient adjustment (see, for example, Henckel et al., 2022) also motivates other extensions of IDA, such as the works of Witte et al. (2020); Liu et al. (2020a); Guo and Perković (2021).

In Section 2, we study the graphical characterization of causal MPDAGs and their minimal representation. Section 3 then introduces direct causal clauses and demonstrates how to use them to represent pairwise causal constraints. Algorithms for checking the consistency and equivalence of pairwise causal constraints, as well as for constructing the decomposed causal MPDAG and residual direct causal clauses, are presented in Section 4. In Section 5, we focus on the identifiability of causal effects and the methods for locally or semi-locally estimating all possible causal effects, with simulations also provided. Additional algorithms and detailed proofs are given in the appendices.

## 2. A Graphical Characterization of Causal MPDAGs

In this section, we study the necessary and sufficient conditions for a partially directed graph to be a causal MPDAG, as well as the minimal representation of a causal MPDAG. MPDAGs serve as graphical tools for representing direct and non-ancestral causal constraints, and they play a crucial role in identifying causal effects under pairwise causal constraints, as demonstrated in Section 5.1. Before presenting the main results in Theorems 8 and 13, we note that the results in this section are not directly used in the subsequent sections. Readers primarily interested in the general representation may choose to skip this section and proceed directly to Section 3.



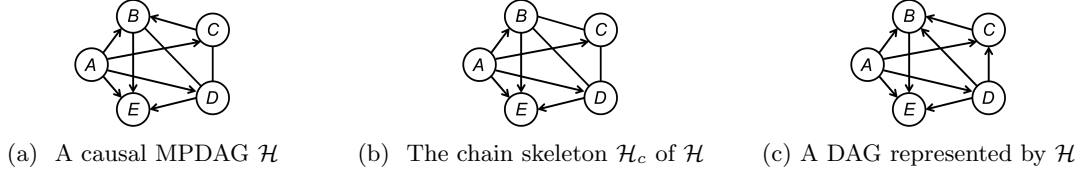


Figure 4: Examples to illustrate the graphical characterization of causal MPDAGs.

We first introduce two concepts related to partially directed graphs.

**Definition 6 (B-component)** *Given a partially directed graph  $\mathcal{G}$ , a B-component  $\mathcal{C}^b$  of  $\mathcal{G}$  is an induced subgraph of  $\mathcal{G}$  over the vertices which are connected by at least one undirected path in  $\mathcal{G}$ .*

The letter “B” in “B-component” comes from “**B**ucket”. We note that, the definition of a B-component is related to *Bucket* defined in Perković (2020). A bucket is a maximal undirected connected subset of the node set of an MPDAG (Perković, 2020). In an MPDAG, the vertex set of a B-component corresponds to a bucket and the induced subgraph over a bucket is a B-component. However, in contrast to buckets, which are defined solely on MPDAGs, B-components are defined for general partially directed graphs. We also remark that the B-component generalizes the concept of a chain component in a chain graph. In fact, for a chain graph, the definition of a B-component degenerates to that of a chain component. However, unlike chain components, a B-component may contain both directed and undirected edges. For example, the MPDAG  $\mathcal{H}$  shown in Figure 4a has three B-components: the induced subgraphs of  $\mathcal{H}$  over  $\{A\}$ ,  $\{E\}$ , and  $\{B, C, D\}$ . The directed edges in a partially directed graph can be divided into two parts: the edges between two B-components (such as  $A \rightarrow B, A \rightarrow E$  in Figure 4a), and the edges within a B-component (such as  $C \rightarrow B$  in Figure 4a).

**Definition 7 (Chain Skeleton)** *Given a partially directed graph  $\mathcal{G}$ , the chain skeleton of  $\mathcal{G}$ , denoted by  $\mathcal{G}_c$ , is the graph obtained from  $\mathcal{G}$  by removing arrowheads of all directed edges in every B-component of  $\mathcal{G}$ .*

According to the definition of a B-component, all undirected edges of  $\mathcal{G}$  appear in B-components of  $\mathcal{G}$ , so the undirected subgraph of  $\mathcal{G}_c$  is a union of the skeletons of the B-components of  $\mathcal{G}$ . Figure 4b displays the chain skeleton of  $\mathcal{H}$  illustrated in Figure 4a. In Figure 4b, the induced subgraph of the chain skeleton  $\mathcal{H}_c$  over  $\{B, C, D\}$  is undirected.

Theorem 8 provides sufficient and necessary conditions for a partially directed graph  $\mathcal{H} = (\mathbf{V}, \mathbf{E})$  to be a causal MPDAG.

**Theorem 8** *A partially directed graph  $\mathcal{H} = (\mathbf{V}, \mathbf{E})$  is a causal MPDAG if and only if  $\mathcal{H}$  satisfies the following conditions.*

- (i) *The chain skeleton  $\mathcal{H}_c$  of  $\mathcal{H}$  is a chain graph.*
- (ii) *The skeleton of each B-component of  $\mathcal{H}$  is chordal.*

- (iii) The vertices in the same B-component have the same parents in  $\mathcal{H}_c$ .
- (iv) For any directed edge  $X \rightarrow Y$  in any B-component  $\mathcal{C}^b$  of  $\mathcal{H}$ ,  $pa(X, \mathcal{H}) \subseteq pa(Y, \mathcal{H}) \setminus \{X\}$  and  $adj(Y, \mathcal{C}^b) \setminus \{X\} \subseteq adj(X, \mathcal{C}^b)$ .

Comparing to the graphical characterizations of essential graphs (Andersson et al., 1997) and intervention essential graphs (Hauser and Bühlmann, 2012), the conditions in Theorem 8 are weaker since these two types of graphs are also causal MPDAGs. That is, these conditions are necessary but not sufficient for a graph to be an essential graph or an intervention essential graph.

In Theorem 8, condition (i) states a global characteristic of the partially directed graph  $\mathcal{H}$ , that is, there are no partially directed circles in the chain skeleton  $\mathcal{H}_c$  of  $\mathcal{H}$ . The last three conditions characterize the graphical structure related to B-components of  $\mathcal{H}$ . Condition (ii) states that the undirected induced subgraphs of  $\mathcal{H}_c$  are chordal, and condition (iii) shows that a vertex out of a B-component is either a parent of all vertices in the B-component, or not a parent of any vertex in the B-component. Condition (iv) indicates that the neighbor and parental sets of the two endpoints of a directed edge in a B-component must satisfy some inclusion relations.

In Section 1.2, a causal MPDAG is defined as an MPDAG that can represent a restricted Markov equivalence class induced by a CPDAG  $\mathcal{G}^*$  and a pairwise causal constraint set  $\mathcal{B}$ . Below, we first show that such a CPDAG  $\mathcal{G}^*$  is unique.

**Proposition 9** *Given a causal MPDAG  $\mathcal{H}$ , any restricted Markov equivalence class that can be represented by  $\mathcal{H}$  is induced by the same CPDAG  $\mathcal{G}^*$  and some pairwise causal constraint set. Moreover, there exists a direct causal constraint set  $\mathcal{B}_d$  such that  $[\mathcal{H}] = [\mathcal{G}^*, \mathcal{B}_d]$ .*

Following Proposition 9, we introduce the notions of a generator and a minimal generator.

**Definition 10 (Generator and Minimal Generator)** *Let  $\mathcal{H}$  be a causal MPDAG and  $\mathcal{G}^*$  be the unique CPDAG of all restricted Markov equivalence classes that can be represented by  $\mathcal{H}$ . A direct causal constraint set (or equivalently, a set of directed edges)  $\mathcal{A}$  is called a generator of  $\mathcal{H}$  if  $[\mathcal{H}] = [\mathcal{G}^*, \mathcal{A}]$ . A generator  $\mathcal{A}$  is called minimal if the number of direct causal constraints in  $\mathcal{A}$  is less than or equal to that in any other generator of  $\mathcal{H}$ .*

Proposition 9 shows that every causal MPDAG has a generator. Below we will show that the minimal generator is unique. A new concept called M-strongly protected is required.

**Definition 11 (M-Strongly Protected)** *Let  $\mathcal{H}$  be a causal MPDAG. A directed edge  $X \rightarrow Y$  in  $\mathcal{H}$  is M-strongly protected if  $X \rightarrow Y$  occurs in at least one of the five configurations in Figure 5 as an induced graph of  $\mathcal{H}$ .*

The following proposition establishes the uniqueness of the minimal generator.

**Proposition 12** *Given a causal MPDAG  $\mathcal{H}$ , a set of directed edges  $\mathcal{A}$  is a minimal generator of  $\mathcal{H}$  if and only if  $\mathcal{A}$  is the set of directed edges which are not M-strongly protected in  $\mathcal{H}$ . Moreover, the minimal generator of  $\mathcal{H}$  is unique.*

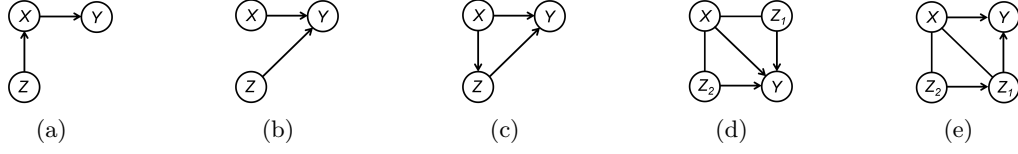


Figure 5: The five configurations of M-strongly protected edges.

Proposition 12 also provides a method to find the minimal generator of a given causal MPDAG. For example, consider the causal MPDAG  $\mathcal{H}$  shown in Figure 4a, the four directed edges  $A \rightarrow B$ ,  $A \rightarrow E$ ,  $B \rightarrow E$  and  $D \rightarrow E$  are M-strongly protected and the other three directed edges  $A \rightarrow C$ ,  $A \rightarrow D$  and  $C \rightarrow B$  are not, so  $\mathcal{A} = \{A \rightarrow C, A \rightarrow D, C \rightarrow B\}$  is the unique minimal generator of  $\mathcal{H}$ . In summary, we have the following theorem.

**Theorem 13** *Let  $\mathcal{H}$  be a causal MPDAG. Then, there exists a unique CPDAG  $\mathcal{G}^*$  and a unique minimal generator  $\mathcal{B}_m$  such that  $[\mathcal{H}] = [\mathcal{G}^*, \mathcal{B}_m]$ .*

The first four configurations of M-strongly protected edges in Definition 11 are exactly the configurations of strongly protected defined for essential graphs. Andersson et al. (1997) proved that every directed edge in an essential graph is strongly protected. Therefore, strongly protected can be used to figure out the directed edges that can be learned from observational data. In contrast, from Proposition 12 and Theorem 13, in a causal MPDAG, a directed edge can be learned from observational data, or can be inferred from the minimal generator if the directed edge is M-strongly protected.

### 3. A Unified Representation of Pairwise Causal Constraints

Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{B}$  be a set of pairwise causal constraints consistent with  $\mathcal{G}^*$ . As discussed in Section 1.2, a causal MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{B}]$  may not be fully informative when  $\mathcal{B}$  contains ancestral causal constraints. That is,  $[\mathcal{G}^*, \mathcal{B}]$  cannot be represented exactly by any causal MPDAG. In this section, we first introduce a new representation, called direct causal clauses, which can be used to represent all types of pairwise causal constraints, and discuss the consistency and equivalence of direct causal clauses. Then, in Section 3.2, we show that any pairwise causal background knowledge set can be equivalently decomposed into a causal MPDAG plus a minimal residual set of direct causal clauses, and prove sufficient and necessary conditions for a causal MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{B}]$  to be fully informative. Finally, Section 3.3 briefly discusses the representation of non-pairwise causal background knowledge.

#### 3.1 Direct Causal Clauses

In this section, we develop a non-graphical tool called direct causal clauses, which can uniformly represent direct, ancestral, and non-ancestral causal constraints.

**Definition 14 (Direct Causal Clause)** *A direct causal clause (DCC for short)  $\kappa$ , also denoted by  $\kappa_t \xrightarrow{\text{or}} \kappa_h$ , over a variable set  $\mathbf{V}$  is a proposition saying that  $\kappa_t$  is a direct cause*

of at least one variable in  $\kappa_h$ , where  $\kappa_t \in \mathbf{V}$  is called the tail of  $\kappa$ , and  $\kappa_h \subseteq \mathbf{V}$  satisfying  $\kappa_t \notin \kappa_h$  is the head set whose elements are called the heads of  $\kappa$ .

When the head set  $\kappa_h$  of a DCC  $\kappa$  is a singleton set, say  $\kappa_h = \{D\}$ ,  $\kappa_t \xrightarrow{\text{or}} \kappa_h$  is equivalent to the proposition that  $\kappa_t$  is a direct cause of  $D$ , denoted by  $\kappa_t \rightarrow D$ . For ease of presentation, we will use  $\kappa_t \xrightarrow{\text{or}} D$  or  $\kappa_t \rightarrow D$  as a shorthand for  $\kappa_t \xrightarrow{\text{or}} \{D\}$ , and we will use them interchangeably throughout this paper. Given a DAG  $\mathcal{G}$  and a DCC  $\kappa$  over the same variable set  $\mathbf{V}$ , we say that  $\kappa$  holds for  $\mathcal{G}$  and  $\mathcal{G}$  satisfies  $\kappa$  if  $\kappa_h \cap \text{ch}(\kappa_t, \mathcal{G}) \neq \emptyset$ .

**Proposition 15** *For any DAG  $\mathcal{G}$  over  $\mathbf{V}$ , we have that (i) a DCC  $\kappa$  over  $\mathbf{V}$  with  $\kappa_h = \emptyset$  never holds for  $\mathcal{G}$ , and (ii) for any DCC  $\kappa$  over  $\mathbf{V}$ ,  $\kappa \iff \bigvee_{D \in \kappa_h} (\kappa_t \rightarrow D)$  for  $\mathcal{G}$ .*

The first statement of Proposition 15 naturally holds since for any  $\kappa$  with  $\kappa_h = \emptyset$ ,  $\kappa_h \cap \text{ch}(\kappa_t, \mathcal{G}) = \emptyset$ , no matter whether  $\text{ch}(\kappa_t, \mathcal{G}) = \emptyset$  or not. Even if for a singleton graph  $\mathcal{G} = (\kappa_t, \emptyset)$  containing  $\kappa_t$  only,  $\kappa_t \xrightarrow{\text{or}} \emptyset$  does not hold for  $\mathcal{G}$ . Proposition 15 also shows that a DCC  $\kappa$  holds for  $\mathcal{G}$  if and only if there exists at least one variable  $D \in \kappa_h$  such that  $\kappa_t \rightarrow D$  holds for  $\mathcal{G}$ . It implies that a DCC is a disjunction of direct causal constraints.

To demonstrate that DCCs can represent pairwise causal constraints, we introduce the concept of a critical set as follows.

**Definition 16 (Critical Set)** *Let  $\mathcal{G}^*$  be a causal MPDAG, and let  $X$  and  $Y$  be two distinct vertices in  $\mathcal{G}^*$ . The critical set of  $X$  with respect to  $Y$  in  $\mathcal{G}^*$ , denoted by  $\mathbf{C}_{XY}(\mathcal{G}^*)$ , consists of all neighbors of  $X$  lying on at least one chordless partially directed path or chordless undirected path from  $X$  to  $Y$ .*

This concept was first introduced for CPDAGs by Fang and He (2020). As an example, consider the CPDAG shown in Figure 7a. The critical set of  $A$  with respect to  $Y$  consists of  $B$  and  $C$ , as  $A - B \rightarrow Y$  and  $A - C \rightarrow Y$  are two chordless partially directed path from  $A$  to  $Y$ . On the other hand, variable  $X$  is not in  $\mathbf{C}_{AY}(\mathcal{G}^*)$ , since  $A - X - C \rightarrow Y$  has a chord  $A - C$  and  $A - X - B \rightarrow Y$  has a chord  $A - B$ . Similarly, the critical set of  $D$  with respect to  $Y$  is  $\{A, X\}$ , for  $D - A - B \rightarrow Y$  and  $D - X - C \rightarrow Y$  are chordless.

**Theorem 17** *Let  $\mathcal{G}^*$  be a CPDAG,  $X, Y \in \mathbf{V}(\mathcal{G}^*)$ , and  $\mathcal{G} \in [\mathcal{G}^*]$ . Denote by  $\mathbf{C}_{XY}(\mathcal{G}^*)$  the critical set of  $X$  with respect to  $Y$  in  $\mathcal{G}^*$ . Then, we have:*

- (i)  $X$  is a direct cause of  $Y$  in  $\mathcal{G}$  if and only if  $X \xrightarrow{\text{or}} Y$  holds for  $\mathcal{G}$ .
- (ii)  $X$  is a cause of  $Y$  in  $\mathcal{G}$  if and only if  $X \xrightarrow{\text{or}} \mathbf{C}_{XY}(\mathcal{G}^*)$  holds for  $\mathcal{G}$ .
- (iii)  $X$  is not a cause of  $Y$  in  $\mathcal{G}$  if and only if  $C \xrightarrow{\text{or}} X$  holds for  $\mathcal{G}$  for every  $C \in \mathbf{C}_{XY}(\mathcal{G}^*)$ .

Theorem 17 extends Lemma 2 in Fang and He (2020) to all pairwise causal constraints. Analogue to Definition 3, we can define the restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{K}]$  induced by a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ , as the subset of  $[\mathcal{G}^*]$  in which every DAG satisfies all DCCs in  $\mathcal{K}$ . Likewise, the MPDAG of a non-empty  $[\mathcal{G}^*, \mathcal{K}]$  can also be defined analogously to Definition 4. Given a CPDAG and a set of pairwise causal constraints, Theorem 17 proves that there is a set of DCCs which induces the same restricted Markov equivalence class as the original constraints.

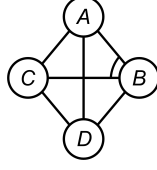


Figure 6: A CPDAG over  $\{A, B, C, D\}$  and a DCC  $B \xrightarrow{\text{or}} \{A, C\}$ , which is visualized by an arc. This example shows that not every DCC can be translated back into pairwise causal constraints.

From Theorem 17, the global path constraints are transformed to the local ones that only put constraints on the edges between  $X$  and its neighbors. A polynomial-time algorithm proposed by Fang et al. (2022, Algorithm 2) can be used to find critical sets, so we can efficiently obtain the equivalent DCCs from a given CPDAG and pairwise causal constraints.

We remark that any pairwise causal constraint can be equivalently transformed into a set of DCCs, but not vice versa. That is, not every set of DCCs can be translated back into pairwise causal constraints. An illustrative example is provided in Figure 6, where the DCC  $B \xrightarrow{\text{or}} \{A, C\}$  is not equivalent to any combination of pairwise causal constraints. This example demonstrates that DCCs can encode more information than pairwise constraints. Assuming that a set of DCCs can be translated back into pairwise causal constraints, Appendix A.1 presents a method for obtaining an equivalent set of pairwise causal constraints.

**Example 2** Consider the CPDAG  $\mathcal{G}^*$  shown in Figure 7a. Recall that  $\mathbf{C}_{AY}(\mathcal{G}^*) = \{B, C\}$  and  $\mathbf{C}_{DY}(\mathcal{G}^*) = \{A, X\}$ . Suppose that  $A \dashrightarrow Y$  holds for the underlying DAG, then by Theorem 17,  $B \xrightarrow{\text{or}} A$  and  $C \xrightarrow{\text{or}} A$  holds. Since  $B \xrightarrow{\text{or}} A \iff B \rightarrow A$  and  $C \xrightarrow{\text{or}} A \iff C \rightarrow A$ , we can represent  $A \dashrightarrow Y$  graphically by orienting  $B \rightarrow A$  and  $C \rightarrow A$  in  $\mathcal{G}^*$ . Furthermore, if  $D \dashrightarrow Y$  holds, then by Theorem 17 we equivalently have  $D \xrightarrow{\text{or}} \{A, X\}$ , meaning that  $D$  is a direct cause of  $A$  or  $X$  in the underlying DAG. Figures 7d to 7f enumerate three possible orientations of the edges  $D-A$  and  $D-X$  in  $\mathcal{G}^*$  with the constraint  $D \xrightarrow{\text{or}} \{A, X\}$ . For any DAG in  $[\mathcal{G}^*, D \dashrightarrow Y]$ , the edge orientations of  $D-A$  and  $D-X$  must be one of the three possibilities shown in Figures 7d to 7f. Conversely, every DAG in  $[\mathcal{G}^*]$  whose edges between  $D$  and  $\{A, X\}$  are identical to one of the configurations shown in Figures 7d to 7f must satisfy the constraint  $D \dashrightarrow Y$ .

Below, we discuss the consistency of a DCC set and the equivalence of two DCC sets with respect to a given CPDAG. We first introduce the definition of consistency as follows.

**Definition 18 (Consistency)** Given a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ ,  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$  if  $[\mathcal{G}^*, \mathcal{K}] \neq \emptyset$ .

Following Definition 18, a set  $\mathcal{K}$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$  is consistent with  $\mathcal{G}^*$  if and only if there exists at least one equivalent DAG  $\mathcal{G} \in [\mathcal{G}^*]$  which satisfies all clauses in  $\mathcal{K}$ . Below, we give a definition of equivalence for two sets of DCCs, which allows us to discuss the consistency of a DCC set via its equivalent reduced form.

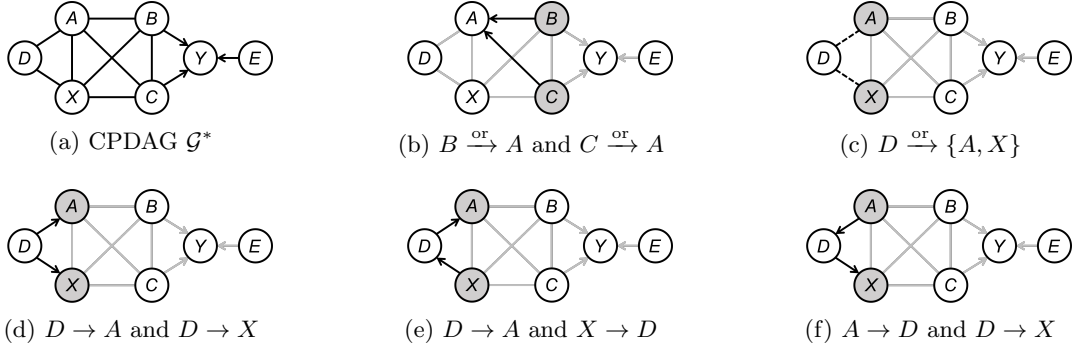


Figure 7: An illustration of a critical set and Theorem 17. Figure 7a shows a CPDAG. Figures 7b and 7c show the equivalent DCCs to the pairwise causal background knowledge  $A \leftrightarrow Y$  and  $D \leftrightarrow Y$ , respectively. Figures 7d to 7f enumerate three possible orientations of the edges  $D - A$  and  $D - X$  in  $\mathcal{G}^*$  with the constraint  $D \xrightarrow{\text{or}} \{A, X\}$ .

**Definition 19 (Equivalence)** *Given a CPDAG  $\mathcal{G}^*$ , two sets  $\mathcal{K}_1$  and  $\mathcal{K}_2$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$  are equivalent with respect to  $\mathcal{G}^*$  if  $[\mathcal{G}^*, \mathcal{K}_1] = [\mathcal{G}^*, \mathcal{K}_2]$ .*

Let  $\kappa := \kappa_t \xrightarrow{\text{or}} \kappa_h$  be a DCC. Denoting  $[\mathcal{G}^*, \{\kappa\}]$  by  $[\mathcal{G}^*, \kappa]$  for convenience and assuming that  $\kappa$  is over  $\mathbf{V}(\mathcal{G}^*)$ , the following proposition gives several equivalent reduced forms of  $\kappa$ .

**Proposition 20** *For any CPDAG  $\mathcal{G}^*$  and any direct causal clause  $\kappa := \kappa_t \xrightarrow{\text{or}} \kappa_h$ , we have that (i)  $[\mathcal{G}^*, \kappa] = [\mathcal{G}^*]$  if  $ch(\kappa_t, \mathcal{G}^*) \cap \kappa_h \neq \emptyset$ , (ii)  $[\mathcal{G}^*, \kappa] = [\mathcal{G}^*, \kappa_t \xrightarrow{\text{or}} (\kappa_h \cap adj(\kappa_t, \mathcal{G}^*))]$ , and (iii)  $[\mathcal{G}^*, \kappa] = [\mathcal{G}^*, \kappa_t \xrightarrow{\text{or}} (\kappa_h \setminus pa(\kappa_t, \mathcal{G}^*))]$ .*

In Proposition 20, the first result holds since every  $\mathcal{G} \in [\mathcal{G}^*]$  satisfies  $\kappa$  when  $ch(\kappa_t, \mathcal{G}^*) \cap \kappa_h \neq \emptyset$ . Therefore, a DCC with  $ch(\kappa_t, \mathcal{G}^*) \cap \kappa_h \neq \emptyset$  is redundant for  $\mathcal{G}^*$ . For any  $\mathcal{G} \in [\mathcal{G}^*]$ ,  $\mathcal{G}$  satisfies  $\kappa$  if and only if there exists at least a variable  $D \in \kappa_h$  such that  $\kappa_t \rightarrow D$  appears in  $\mathcal{G}$ . If such a variable  $D$  exists, it must be adjacent to  $\kappa_t$ , so the second equation holds. Similarly, the third equation holds since  $D$  must not be a direct cause of  $\kappa_t$ . Consequently, we give a reduced form of a set of DCCs  $\mathcal{K}$  in Definition 21.

**Definition 21 (Reduced Form)** *Given a CPDAG  $\mathcal{G}^*$  and a set of DCCs  $\mathcal{K}$  over  $\mathbf{V}(\mathcal{G}^*)$ , a reduced form of  $\mathcal{K}$  with respect to  $\mathcal{G}^*$ , denoted by  $\mathcal{K}(\mathcal{G}^*)$ , is defined as follows.*

$$\mathcal{K}(\mathcal{G}^*) := \{\kappa_t \xrightarrow{\text{or}} (\kappa_h \cap sib(\kappa_t, \mathcal{G}^*)) \mid \kappa \in \mathcal{K} \text{ and } \kappa_h \cap ch(\kappa_t, \mathcal{G}^*) = \emptyset\}. \quad (3)$$

Specifically,  $\mathcal{K}(\mathcal{G}^*) := \emptyset$  if  $\mathcal{K} = \emptyset$ .

**Proposition 22 (Equivalent Reduced Form)** *Given a CPDAG  $\mathcal{G}^*$  and a set of DCCs  $\mathcal{K}$  over  $\mathbf{V}(\mathcal{G}^*)$ , we have that  $\mathcal{K}$  is equivalent to  $\mathcal{K}(\mathcal{G}^*)$  defined in Equation (3) with respect to  $\mathcal{G}^*$ .*

Proposition 22 shows that the reduced form of  $\mathcal{K}$  is equivalent to  $\mathcal{K}$  with respect to  $\mathcal{G}^*$ . Below, we define a subset of  $\mathcal{K}$  restricted on an undirected induced subgraph of  $\mathcal{G}^*$ .

**Definition 23 (Restriction Subset)** Suppose that  $\mathcal{U}$  is an undirected induced subgraph of a CPDAG  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}) \subseteq \mathbf{V}(\mathcal{G}^*)$ , and  $\mathcal{K}$  is a set of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ . The restriction subset of  $\mathcal{K}$  on  $\mathcal{U}$  is defined by

$$\mathcal{K}(\mathcal{U}) := \{\kappa \in \mathcal{K}(\mathcal{G}^*) \mid \{\kappa_t\} \cup \kappa_h \subseteq \mathbf{V}(\mathcal{U})\}. \quad (4)$$

It can be seen that  $\mathcal{K}(\mathcal{G}^*) = \mathcal{K}(\mathcal{G}_u^*)$ . Basically,  $\mathcal{K}(\mathcal{U})$  consists of all clauses in  $\mathcal{K}(\mathcal{G}^*)$  whose tail and heads are all in  $\mathcal{U}$ . The following concept of a potential leaf node is of key importance in checking consistency of a set of DCCs.

**Definition 24 (Potential Leaf Node)** Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ . Given an undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$  and a vertex  $X$  in  $\mathcal{U}$ ,  $X$  is called a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , if  $X$  is a simplicial vertex in  $\mathcal{U}$  and  $X$  is not the tail of any clause in  $\mathcal{K}(\mathcal{U})$ .

We note that if  $\mathcal{U} = (\{X\}, \emptyset)$  only contains a singleton  $X$ , then  $X$  is trivially a potential leaf node in  $\mathcal{U}$  with respect to any  $\mathcal{K}$  that does not contain any DCC of the form  $X \xrightarrow{\text{or}} \emptyset$ . A leaf node of a directed graph is a vertex that has no child. Analogously, a potential leaf node defined above is a vertex that may have no child in at least one  $\mathcal{G}$  in  $[\mathcal{G}^*, \mathcal{K}]$  (see Lemma 55 in Appendix for more details). Now, we present the sufficient and necessary condition for a set of DCCs  $\mathcal{K}$  to be consistent with  $\mathcal{G}^*$ .

**Theorem 25** Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ . Then, the following two statements are equivalent.

- (i)  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ .
- (ii) (Potential-leaf-node condition) Any connected undirected induced subgraph of  $\mathcal{G}^*$  has a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .

The proof of Theorem 25 also motivates a polynomial-time algorithm for checking consistency of a set of DCCs. The details are given in Section 4.1. The potential-leaf-node condition given in Theorem 25 is similar to the fact that any induced subgraph of a DAG has a leaf node. Below we give an example to demonstrate this result.

**Example 3** Recall that in Example 2 we show that, with respect to the CPDAG  $\mathcal{G}^*$  (Figure 7a),  $A \dashrightarrow Y$  is equivalent to  $C \xrightarrow{\text{or}} A$  and  $B \xrightarrow{\text{or}} A$ , and  $D \dashrightarrow Y$  is equivalent to  $D \xrightarrow{\text{or}} \{A, X\}$ . Suppose that we have both  $A \dashrightarrow Y$  and  $D \dashrightarrow Y$ , then the equivalent DCCs  $\mathcal{K}$  consists of  $C \xrightarrow{\text{or}} A$ ,  $B \xrightarrow{\text{or}} A$  and  $D \xrightarrow{\text{or}} \{A, X\}$ . However, since  $\{D, A, X, B\}$  induces an undirected subgraph where none of the vertices is a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ ,  $\mathcal{K}$  is inconsistent by Theorem 25.

Next, we give sufficient and necessary conditions under which two DCC sets,  $\mathcal{K}_1$  and  $\mathcal{K}_2$ , are equivalent with respect to a CPDAG  $\mathcal{G}^*$ . We will show that these conditions can be expressed both in terms of consistency and redundancy, where the latter is defined as follows.

**Definition 26 (Redundancy)** *Given a restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{K}]$  induced by a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ , a DCC  $\kappa$  over  $\mathbf{V}(\mathcal{G}^*)$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  if  $[\mathcal{G}^*, \mathcal{K}] = [\mathcal{G}^*, \mathcal{K} \cup \{\kappa\}]$ . A set  $\mathcal{K}$  of DCCs is redundant with respect to a CPDAG  $\mathcal{G}^*$  if there exists at least one  $\kappa \in \mathcal{K}$  that is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa\}]$ . Otherwise, the set  $\mathcal{K}$  is non-redundant.*

In the remainder of the paper, when the context is clear, we may occasionally omit  $\mathcal{G}^*$  and simply say that “a DCC  $\kappa$  is redundant with respect to  $\mathcal{K}$ ” for brevity. If  $\kappa_t \rightarrow s$  appears in  $\mathcal{G}^*$  for some  $s \in \kappa_h$ , then  $\kappa$  is redundant with respect to any DCC set. If  $\mathcal{K}$  is inconsistent with  $\mathcal{G}^*$ , then any DCC  $\kappa$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ , as  $[\mathcal{G}^*, \mathcal{K}] = [\mathcal{G}^*, \mathcal{K} \cup \{\kappa\}] = \emptyset$ . However, an inconsistent DCC set  $\mathcal{K}$  itself may not be redundant.

According to Definition 26, a DCC  $\kappa$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  if and only if  $\kappa$  holds for all DAGs in  $[\mathcal{G}^*, \mathcal{K}]$ . With this concept, the following Theorem 27 discusses the equivalence of two sets of DCCs.

**Theorem 27** *Given a CPDAG  $\mathcal{G}^*$  and two sets of DCCs  $\mathcal{K}_1$  and  $\mathcal{K}_2$  over  $\mathbf{V}(\mathcal{G}^*)$ , the following statements are equivalent.*

- (i)  $\mathcal{K}_1$  and  $\mathcal{K}_2$  are equivalent given  $\mathcal{G}^*$ .
- (ii) Every DCC in  $\mathcal{K}_1$ , if exists, is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}_2]$ , and every DCC in  $\mathcal{K}_2$ , if exists, is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}_1]$ .
- (iii) For every  $\kappa \in \mathcal{K}_1$ ,  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathcal{K}_2$  is not consistent with  $\mathcal{G}^*$ , and for every  $\kappa \in \mathcal{K}_2$ ,  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathcal{K}_1$  is not consistent with  $\mathcal{G}^*$ .

Theorem 27 builds the relations among equivalence, redundancy and consistency. Moreover, using the third statement of Theorem 27, we can determine whether two sets of DCCs are equivalent by checking the consistency of a series of DCC sets.

We end this subsection by showing that two equivalent non-redundant consistent DCC sets must have the same number of DCCs. The following definition serves as a prerequisite.

**Definition 28 (Minimal Redundancy)** *Given a restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{K}]$  induced by a CPDAG  $\mathcal{G}^*$  and a consistent set  $\mathcal{K}$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ , a DCC  $\kappa$  over  $\mathbf{V}(\mathcal{G}^*)$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  if  $\kappa$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ , but either  $|\kappa_h| = 1$ , or for any proper subset  $\mathbf{s} \subsetneq \kappa_h$ , the DCC  $\kappa_t \xrightarrow{\text{or}} \mathbf{s}$  is not redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ .*

With this definition, we give the following theorem that summarizes key properties of equivalent non-redundant consistent DCC sets over a CPDAG.

**Theorem 29** *Suppose that  $\mathcal{G}^*$  is a CPDAG and  $\mathcal{K}$  is a non-redundant consistent DCC set over  $\mathbf{V}(\mathcal{G}^*)$ . Then:*

- (i) *There exists a unique non-redundant DCC set  $\mathcal{K}'$  over  $\mathbf{V}(\mathcal{G}^*)$  such that  $\mathcal{K}$  is equivalent to  $\mathcal{K}'$  and  $\kappa'$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}']$  for any  $\kappa' \in \mathcal{K}'$ . Moreover, for every  $\kappa \in \mathcal{K}$ , there exists a unique  $\kappa' \in \mathcal{K}'$  such that  $\kappa'_t = \kappa_t$  and  $\kappa'_h \subseteq \kappa_h$ ; likewise, for every  $\kappa' \in \mathcal{K}'$ , there exists a unique  $\kappa \in \mathcal{K}$  such that  $\kappa_t = \kappa'_t$  and  $\kappa_h \subseteq \kappa'_h$ .*



- (ii) For every non-redundant DCC set  $\mathcal{K}'$  over  $\mathbf{V}(\mathcal{G}^*)$  which is equivalent to  $\mathcal{K}$ , it holds that  $|\mathcal{K}| = |\mathcal{K}'|$  and  $\{\kappa_t \mid \kappa \in \mathcal{K}\} = \{\kappa'_t \mid \kappa' \in \mathcal{K}'\}$ .

By definition, a non-redundant DCC set is subset-minimal, meaning that none of its subsets can be inferred from the others. Furthermore, Theorem 29 establishes that a non-redundant DCC set is also cardinality-minimal, implying that the cardinality of the DCC set is the smallest among all equivalent DCC sets. Finally, the first statement of Theorem 29 guarantees the existence and uniqueness of a non-redundant and equivalent DCC set that is element-wise head-minimal, meaning that every DCC in the set has the smallest possible head.

### 3.2 Equivalent Decomposition of Pairwise Causal Constraints

Let  $\mathcal{G}^*$  be a CPDAG,  $\mathcal{B}$  be a set of consistent pairwise causal constraints, Theorem 17 proves that  $\mathcal{B}$  can be equivalently represented by a set of DCCs  $\mathcal{K}$ . Let  $\mathcal{H}$  be the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ . By Definition 4 of MPDAGs,  $\mathcal{H}$  represents all common direct causal relations shared by all DAGs in  $[\mathcal{G}^*, \mathcal{B}]$ . Due to the equivalence,  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{G}^*, \mathcal{K}] = [\mathcal{H}, \mathcal{K}]$ , where  $[\mathcal{H}, \mathcal{K}]$  denotes the subset of  $[\mathcal{H}]$  consisting of DAGs satisfying all clauses in  $\mathcal{K}$ . However, given  $\mathcal{H}$ , some of the DCCs in  $\mathcal{K}$  may become redundant. The following result proves that any set of consistent pairwise causal constraints can be equivalently decomposed into an MPDAG and a non-redundant residual set of DCCs.

**Theorem 30 (Equivalent Decomposition)** *Let  $\mathcal{G}^*$  be a CPDAG,  $\mathcal{B}$  be a set of consistent pairwise causal constraints, and  $\mathcal{H}$  be the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ . Then:*

- (i) *There exists a DCC set  $\mathcal{R}$  such that  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{H}, \mathcal{R}]$ , where none of the DCC  $\kappa \in \mathcal{R}$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R} \setminus \{\kappa\})$ .*
- (ii) *Every DCC set  $\mathcal{R}$  such that  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{H}, \mathcal{R}]$  and none of the DCC  $\kappa \in \mathcal{R}$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R} \setminus \{\kappa\})$  contains the same number of DCCs.*
- (iii) *There exists a unique DCC set  $\mathcal{R}^*$  such that (1)  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{H}, \mathcal{R}^*]$ , (2) none of the DCC  $\kappa \in \mathcal{R}^*$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R}^* \setminus \{\kappa\})$ , and (3) every  $\kappa \in \mathcal{R}^*$ , if  $\mathcal{R}^* \neq \emptyset$ , is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{B}]$ .*
- (iv) *For any DCC set  $\mathcal{R}$  such that  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{H}, \mathcal{R}]$  and none of the DCC  $\kappa \in \mathcal{R}$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R} \setminus \{\kappa\})$ ,  $\mathcal{R} \neq \emptyset$  if and only if  $\mathcal{R}^* \neq \emptyset$ , and when  $\mathcal{R} \neq \emptyset$ , it holds that, (1) for every  $\kappa \in \mathcal{R}$ , there exists a unique  $\kappa^* \in \mathcal{R}^*$  such that  $\kappa_t^* = \kappa_t$  and  $\kappa_h^* \subseteq \kappa_h$ , and (2) for every  $\kappa^* \in \mathcal{R}^*$ , there exists a unique  $\kappa \in \mathcal{R}$  such that  $\kappa_t^* = \kappa_t$  and  $\kappa_h^* \subseteq \kappa_h$ . Here,  $\mathcal{R}^*$  is the unique DCC set defined in (iii).*

Given a CPDAG, Theorem 30 indicates that any set of pairwise causal constraints can be equivalently decomposed into the MPDAG of the induced restricted Markov equivalence class (which is unique), and a residual set of DCCs in which every DCC is non-redundant with respect to the other DCCs and the directed edges in the MPDAG. Although the decomposed residual sets of DCCs are not unique, they all contain the same number of DCCs, which means they are all minimal in terms of cardinality. Moreover, the third

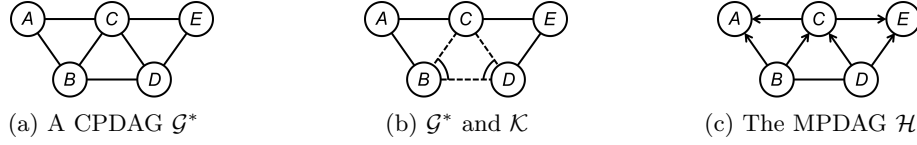


Figure 8: An example of a fully informative MPDAG.

statement of Theorem 30 establishes the existence and uniqueness of an element-wise head-minimal residual set of DCCs where all DCCs are minimally redundant, meaning that each DCC in the set contains the minimum possible number of heads, thereby generalizing Theorem 13 in Section 2. Finally, the last statement characterizes the relationship between the element-wise head-minimal residual set and the other residual sets.

Note that, since the MPDAG  $\mathcal{H}$  contains all direct causal edges that appear in all DAGs in  $[\mathcal{G}^*, \mathcal{B}]$ , every directed edge in the MPDAG of  $[\mathcal{G}^*, \mathcal{R}]$  is also in  $\mathcal{H}$ . That is,  $\mathcal{R}$  cannot bring more directed causal edges other than those in  $\mathcal{H}$ .

When the residual set of DCCs is empty, the MPDAG of the induced restricted Markov equivalence class is fully informative. As mentioned in Section 1.2, a sufficient condition that guarantees the emptiness of a residual set is when  $\mathcal{B}$  only contains direct and non-ancestral causal constraints. Yet, this condition is not necessary, as shown in the following example.

**Example 4** Figure 8a shows a CPDAG  $\mathcal{G}^*$ . Consider two ancestral causal constraints  $B \dashrightarrow E$  and  $D \dashrightarrow A$ . By Theorems 17 and 30,  $B \dashrightarrow E$  and  $D \dashrightarrow A$  are equivalent to  $B \xrightarrow{\text{or}} \{C, D\}$  and  $D \xrightarrow{\text{or}} \{B, C\}$ , respectively, which are visualized by arcs in Figure 8b. Let  $\mathcal{K} = \{B \xrightarrow{\text{or}} \{C, D\}, D \xrightarrow{\text{or}} \{B, C\}\}$ . We first find the MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{K}]$ . If there is a DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$  where  $C \rightarrow B$  is in  $\mathcal{G}$ , then by the constraint  $B \xrightarrow{\text{or}} \{C, D\}$ ,  $B \rightarrow D$  is in  $\mathcal{G}$ . Likewise, by the constraint  $D \xrightarrow{\text{or}} \{B, C\}$ ,  $D \rightarrow C$  is in  $\mathcal{G}$ . However,  $C \rightarrow B \rightarrow D \rightarrow C$  is a directed cycle. Therefore, every DAG in  $[\mathcal{G}^*, \mathcal{K}]$  should have  $B \rightarrow C$ , implying that  $B \rightarrow C$  is in  $\mathcal{H}$ . Similarly,  $D \rightarrow C$  is in  $\mathcal{H}$ . Finally, by Meek's rules we obtain the MPDAG shown in Figure 8c. Notice that  $[\mathcal{H}]$  only contains two DAGs: one has  $B \rightarrow D$  and the other has  $D \rightarrow B$ . Since both DAGs satisfy the constraints  $B \dashrightarrow E$  and  $D \dashrightarrow A$ ,  $[\mathcal{H}] \subseteq [\mathcal{G}^*, \mathcal{B}]$ . On the other hand, as mentioned in Section 1.2 that  $[\mathcal{G}^*, \mathcal{B}] \subseteq [\mathcal{H}]$ , it holds that  $[\mathcal{H}] = [\mathcal{G}^*, \mathcal{B}]$ . Therefore,  $\mathcal{H}$  is fully informative.

Below, we present the necessary and sufficient conditions for an MPDAG to be fully informative.

**Theorem 31** Suppose that  $\mathcal{G}^*$  is a CPDAG,  $\mathcal{K}$  is a set of consistent DCCs, and  $\mathcal{H}$  is the MPDAG of  $[\mathcal{G}^*, \mathcal{K}]$ . Then, the following statements are equivalent.

- (i)  $\mathcal{H}$  is fully informative with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .
- (ii)  $[\mathcal{H}, \emptyset] = [\mathcal{G}^*, \mathcal{B}]$ .
- (iii) For any  $\kappa \in \mathcal{K}$ , either  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{H})$  induces an incomplete subgraph of  $\mathcal{H}$ , or  $\kappa_h \cap \text{ch}(\kappa_t, \mathcal{H}) \neq \emptyset$ .

We note that, in Section 4.2, we give a polynomial-time algorithm to find the MPDAG given a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs. With the found MPDAG, the condition in statement (iii) can also be verified in polynomial-time.

**Example 5** Figure 9a shows a CPDAG  $\mathcal{G}^*$  and a set of DCCs  $\mathcal{K}$  consisting of  $D \xrightarrow{\text{or}} \{E, B\}$ ,  $E \xrightarrow{\text{or}} \{A, C\}$ ,  $E \xrightarrow{\text{or}} \{B, F\}$  and  $G \xrightarrow{\text{or}} \{B, H\}$ , and Figure 9b shows the MPDAG of  $[\mathcal{G}^*, \mathcal{K}]$ . (How to find this MPDAG is left to Example 8, Section 4.2). Since  $E \rightarrow A$ ,  $E \rightarrow F$  and  $G \rightarrow H$  are in the MPDAG,  $E \xrightarrow{\text{or}} \{A, C\}$ ,  $E \xrightarrow{\text{or}} \{B, F\}$  and  $G \xrightarrow{\text{or}} \{B, H\}$  are redundant given  $\mathcal{H}$ . However,  $D \xrightarrow{\text{or}} \{B, E\}$  is not redundant, as  $\{B, E\}$  induces a complete subgraph of  $\mathcal{H}$ . Therefore, in this example, the MPDAG  $\mathcal{H}$  is not fully informative.

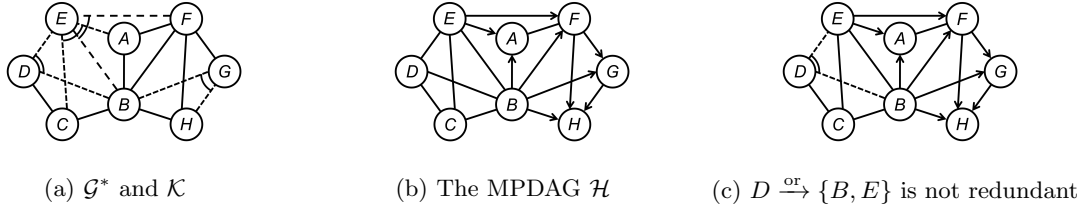


Figure 9: An MPDAG which is not fully informative.

Notice that, given a DCC  $\kappa$ , if  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{H})$  induces an incomplete subgraph of  $\mathcal{H}$ , then  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*)$  induces an incomplete subgraph of  $\mathcal{G}^*$ . Conversely, if  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*)$  induces an incomplete subgraph of  $\mathcal{G}^*$ , then by Rule 1 of Meek’s rules,  $\kappa$  holds for all DAGs in  $[\mathcal{G}^*]$ , and thus  $\kappa$  is redundant. For a given DCC set  $\mathcal{K}$ , removing the above redundant DCCs from the equivalent reduced form  $\mathcal{K}(\mathcal{G}^*)$ , we define a subset of  $\mathcal{K}(\mathcal{G}^*)$  as follows.

$$\mathcal{K}^c(\mathcal{G}_u^*) := \{\kappa \mid \kappa \in \mathcal{K}(\mathcal{G}_u^*) \text{ and the induced subgraph } \mathcal{G}^*(\kappa_h) \text{ is complete}\}. \quad (5)$$

Based on the above argument, we have the following corollary.

**Corollary 32** *Let  $\mathcal{H}$  be an MPDAG representing the restricted Markov equivalence class induced by a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of consistent DCCs, then  $[\mathcal{H}] = [\mathcal{G}^*, \mathcal{K}]$  if and only if for any  $\kappa \in \mathcal{K}^c(\mathcal{G}_u^*)$ ,  $\kappa_h \cap \text{ch}(\kappa_t, \mathcal{H}) \neq \emptyset$  holds.*

### 3.3 Non-Pairwise Causal Background Knowledge

To end this section, we present some discussions on non-pairwise causal background knowledge. Non-pairwise causal background knowledge is also common in practice. For example, *tiered background knowledge* is non-pairwise. Suppose that  $\mathbf{T} := \{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n\}$  defines a disjoint partition of the variable set  $\mathbf{V}$ . The tiered background knowledge is a proposition saying that for all  $X_i \in \mathbf{V}_i$  and  $X_j \in \mathbf{V}_j$  such that  $1 \leq i \leq j \leq n$ , either  $X_i \dashrightarrow X_j$  or  $X_i$  is not adjacent to  $X_j$  (Andrews et al., 2020). Nonetheless, most non-pairwise causal background knowledge can be viewed as Boolean combinations of pairwise causal background

knowledge. For instance, tiered background knowledge can be interpreted and denoted by

$$\text{tbk}^{\mathbf{T}} := \bigwedge_{\substack{X_i \in \mathbf{V}_i, X_j \in \mathbf{V}_j, \\ 1 \leq i \leq j \leq n}} X_i \dashrightarrow X_j \vee (X_i \nrightarrow X_j \wedge X_j \nrightarrow X_i).$$

Moreover, given a CPDAG  $\mathcal{G}^*$ , as  $X_i \nrightarrow X_j \wedge X_j \nrightarrow X_i$  holds for all DAGs in  $[\mathcal{G}^*]$  if and only if  $X_i \notin \text{adj}(X_j, \mathcal{G}^*)$ , we have

$$\text{tbk}^{\mathbf{T}} \iff \bigwedge_{\substack{X_i \in \mathbf{V}_i, X_j \in \mathbf{V}_j, \\ X_i \in \text{adj}(X_j, \mathcal{G}^*), 1 \leq i \leq j \leq n}} X_i \rightarrow X_j.$$

Thus, the results in the main text can be extended to tiered background knowledge.

Unfortunately, directly extending our results to an arbitrary Boolean combination of pairwise causal background knowledge is difficult. Nevertheless, a possible solution exists. Let  $\mathcal{B}$  be a set of Boolean combinations of pairwise causal background knowledge. It is clear that a DAG satisfies all constraints in  $\mathcal{B}$  if and only if  $\bigwedge_{b \in \mathcal{B}} b$  holds for the DAG. Since for any  $b \in \mathcal{B}$ ,  $b$  is a Boolean combination of pairwise causal background knowledge,  $\bigwedge_{b \in \mathcal{B}} b$  can be reformulated into its disjunctive normal form,

$$\bigwedge_{b \in \mathcal{B}} b = \bigvee_{i \in I} \left( \bigwedge_{j \in J} c_{ij} \right), \quad (6)$$

where  $I, J$  are finite indicator sets, and each  $c_{ij}$  is a direct, non-ancestral or ancestral causal constraint. For each  $i \in I$ , let  $\mathcal{B}_i = \{c_{ij} \mid j \in J\}$ , we can then extend our results to each  $\mathcal{B}_i$  and combine them together. For example, to check the consistency of  $\bigwedge_{b \in \mathcal{B}} b$ , it suffices to show that there exists a consistent  $\mathcal{B}_i$ ; to construct the MPDAG representing  $[\mathcal{G}^*, \mathcal{B}]$ , one needs first construct the MPDAG for each  $\mathcal{B}_i$ , then subtract all common direct causal relations from those MPDAGs.

We remark that the bottleneck of the above solution is the computation of Equation (6). Generally, the disjunctive normal form is not unique, and the reformulation may be exponential in complexity. How to efficiently represent Boolean combinations of pairwise causal background knowledge is regarded as future work.

## 4. Polynomial-Time Algorithms

In this section, we present algorithms for checking the consistency and equivalence of DCCs, and for finding the MPDAG and the minimal residual sets of DCCs given a CPDAG and a set of consistent pairwise causal constraints. All proposed algorithms run in polynomial time.

### 4.1 Algorithms for Checking Consistency and Equivalence

Recall that Theorem 25 provides a sufficient and necessary condition for consistency. The proof of Theorem 25 also motivates an algorithm for checking consistency. Algorithm 1 shows the schema. The inputs of Algorithm 1 are a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs over

---

**Algorithm 1** Checking the consistency of a DCC set.
 

---

**Require:** A CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ .

**Ensure:** A Boolean value indicating whether  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ .

```

1: Set  $\mathcal{U} = \mathcal{G}_u^*$  and compute  $\mathcal{K}(\mathcal{U})$  according to Equation (3) and Equation (4).
2: while  $\mathcal{U}$  has a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , do
3:   Find a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  and denote it by  $Y$ .
4:   Update  $\mathcal{U}$  by removing  $Y$  and the edges connected to  $Y$ .
5:   Update  $\mathcal{K}(\mathcal{U})$  by removing all clauses whose heads contain  $Y$ .
6: end while
7: if  $\mathcal{U}$  is an empty graph, then
8:   return True.
9: end if
10: return False.
```

---

$\mathbf{V}(\mathcal{G}^*)$ . It first initializes  $\mathcal{U}$  by  $\mathcal{G}_u^*$ , and then sequentially removes potential leaf nodes from  $\mathcal{U}$  in lines 2-6, until no more potential leaf node can be found.

The complexity of computing  $\mathcal{K}(\mathcal{U})$  is bounded by  $O(|\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^2)$ . Moreover, algorithm 1 runs the while-loop at most  $|\mathbf{V}(\mathcal{G}^*)|$  times. Every time it runs the while-loop, it checks for each vertex in  $\mathcal{U}$  whether the vertex is simplicial and not the tail of any clause in  $\mathcal{K}(\mathcal{U})$ . The complexity of the former is bounded by  $O(|\mathbf{V}(\mathcal{G}^*)|^3)$ , and the complexity of the later is bounded by  $O(|\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|)$ . The complexity of removing  $Y$ , as well as the clauses whose heads contain  $Y$ , is bounded by  $O(|\mathbf{V}(\mathcal{G}^*)|)$  and  $O(|\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|)$ , respectively. Therefore, the complexity of Algorithm 1 is upper bounded by  $O(|\mathbf{V}(\mathcal{G}^*)|^4 + |\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^2)$ , which is polynomial to both  $|\mathbf{V}(\mathcal{G}^*)|$  and  $|\mathcal{K}|$ .

Algorithm 1 is related to the perfect elimination ordering (PEO) of a chordal graph (Blair and Peyton, 1993; Maathuis et al., 2009). In fact, when  $\mathcal{K}$  is consistent, the ordering of the removal potential leaf nodes is a PEO of the chordal graph  $\mathcal{G}_u^*$ . PEOs are important to construct Markov equivalent DAGs. We refer the interested readers to Appendix B.1 for more details. We also remark that, when a pairwise causal background knowledge set contains only direct causal constraints, Algorithm 1 degenerates to the algorithm proposed by Dor and Tarsi (1992).

**Example 6 (continued)** We next use Algorithm 1 to check the consistency of  $\mathcal{K} = \{C \xrightarrow{\text{or}} A, B \xrightarrow{\text{or}} A, D \xrightarrow{\text{or}} \{A, X\}\}$  with respect to  $\mathcal{G}^*$  illustrated in Figure 7a. We first set  $\mathcal{U} = \mathcal{G}_u^*$ , which is the induced subgraph of  $\mathcal{G}^*$  over  $\{A, B, C, D, X\}$ . With respect to  $\mathcal{K}$ , however,  $\mathcal{U}$  has no potential leaf node, meaning that the while-loop (lines 2-6) is not triggered. As  $\mathcal{U}$  is not empty, Algorithm 1 returns False.

In some circumstances, pairwise causal background knowledge is not obtained all at once. Therefore, we also need an approach to sequentially check consistency. That is, given a CPDAG  $\mathcal{G}^*$  and a consistent pairwise causal constraint set  $\mathcal{B}$ , we want to determine whether a newly obtained pairwise causal constraint set is consistent with  $\mathcal{G}^*$  together with  $\mathcal{B}$ . This issue will be investigated in Appendix A.2.

Finally, with Algorithm 1 and Theorem 27, we can check the equivalence of two DCC sets, as shown in Algorithm 2. Note that, to accelerate the procedure, we first check the

---

**Algorithm 2** Checking the equivalence of two DCC sets.

---

**Require:** A CPDAG  $\mathcal{G}^*$  and two sets  $\mathcal{K}_1$  and  $\mathcal{K}_2$  of DCCs over  $\mathbf{V}(\mathcal{G}^*)$ .

**Ensure:** A Boolean value indicating whether  $\mathcal{K}_1$  is equivalent to  $\mathcal{K}_2$  given  $\mathcal{G}^*$ .

```

1: if both  $\mathcal{K}_1$  and  $\mathcal{K}_2$  are consistent with  $\mathcal{G}^*$ , then
2:   for  $\kappa \in \mathcal{K}_1$ , do
3:     if  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathcal{K}_2$  is consistent with  $\mathcal{G}^*$ , then
4:       return False.
5:     end if
6:   end for
7:   for  $\kappa \in \mathcal{K}_2$ , do
8:     if  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathcal{K}_1$  is consistent with  $\mathcal{G}^*$ , then
9:       return False.
10:    end if
11:   end for
12:   return True.
13: else if neither  $\mathcal{K}_1$  nor  $\mathcal{K}_2$  is consistent with  $\mathcal{G}^*$ , then
14:   return True.
15: else
16:   return False.
17: end if

```

---

consistency of  $\mathcal{K}_1$  and  $\mathcal{K}_2$  separately. If neither  $\mathcal{K}_1$  nor  $\mathcal{K}_2$  is consistent with  $\mathcal{G}^*$  then they are equivalent; if one of them is consistent with  $\mathcal{G}^*$  but the other is not, then they are not equivalent.

## 4.2 Algorithms for Finding MPDAGs and Minimal DCC Sets

We first discuss how to find the MPDAG of  $[\mathcal{G}^*, \mathcal{K}]$  induced by a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs consistent with  $\mathcal{G}^*$ . By the definition of an MPDAG, it suffices to find all common direct causal relations shared by all DAGs in  $[\mathcal{G}^*, \mathcal{K}]$ . A new concept is needed before proceeding.

**Definition 33 (Orientation Component)** *Given a CPDAG  $\mathcal{G}^*$  and a DCC set  $\mathcal{K}$  consistent with  $\mathcal{G}^*$ . With respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , a connected undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$  is called an orientation component for a vertex  $X$  if  $X$  is the only potential leaf node in  $\mathcal{U}$ .*

**Example 7** *An orientation component is illustrated by this example. Figure 10a shows a CPDAG  $\mathcal{G}^*$ . Consider  $\mathcal{K} = \{A \xrightarrow{\text{or}} \{X, B, D\}, B \xrightarrow{\text{or}} \{X, A\}, B \xrightarrow{\text{or}} \{X, C, Y\}, X \xrightarrow{\text{or}} \{B, C\}\}$ . Since  $\{X, B, D\} \cap \text{sib}(A, \mathcal{G}^*) = \{X, B\}$  and  $\{X, C, Y\} \cap \text{ch}(B, \mathcal{G}^*) \neq \emptyset$ , by Equation (3),  $\mathcal{K}(\mathcal{G}_u^*)$  consists of  $A \xrightarrow{\text{or}} \{X, B\}$ ,  $B \xrightarrow{\text{or}} \{X, A\}$  and  $X \xrightarrow{\text{or}} \{B, C\}$ , which are visualized by arcs in Figure 10b. First, consider the undirected induced subgraph  $\mathcal{U}$  over  $\{A, B, X\}$ . By definition,  $\mathcal{K}(\mathcal{U})$  consists of  $A \xrightarrow{\text{or}} \{X, B\}$  and  $B \xrightarrow{\text{or}} \{X, A\}$ . Therefore, there is only one potential leaf node in  $\mathcal{U}$ , namely  $X$ , and thus  $\mathcal{U}$  is an orientation component for  $X$ . Next, consider the undirected induced subgraph  $\mathcal{U}$  over  $\{X, B, C\}$ .  $\mathcal{K}(\mathcal{U})$  has only one constraint  $X \xrightarrow{\text{or}} \{B, C\}$ , hence  $\mathcal{U}$  has two potential leaf nodes  $B$  and  $C$ . Finally,*

consider the entire undirected subgraph  $\mathcal{G}_u^*$ .  $\mathcal{G}_u^*$  has only one potential leaf node  $C$ , thus  $\mathcal{G}_u^*$  is an orientation component for  $C$ .

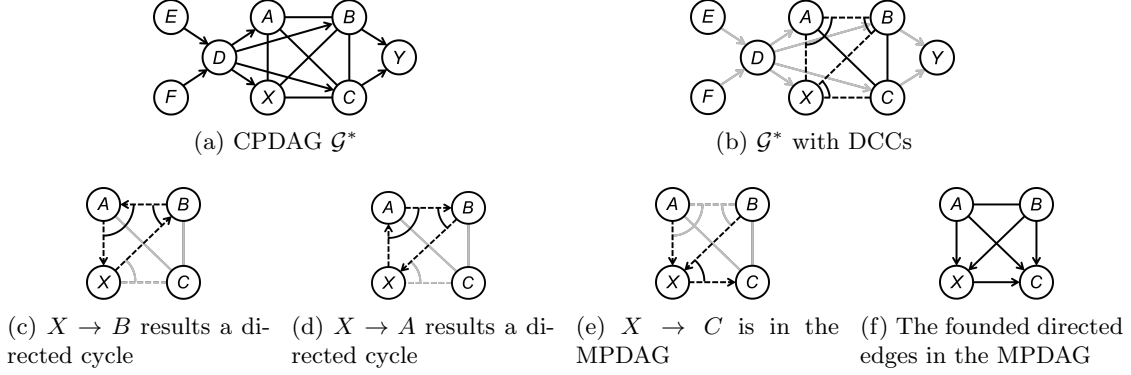


Figure 10: An illustration of an orientation component is given by this example.

The following proposition shows that an orientation component can be used to identify some direct causal relations.

**Proposition 34** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs consistent with  $\mathcal{G}^*$ . For any orientation component  $\mathcal{U}$  for  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  with  $\text{adj}(X, \mathcal{U}) \neq \emptyset$ , all variables in  $\text{adj}(X, \mathcal{U})$  are direct causes of  $X$  in every DAG in  $[\mathcal{G}^*, \mathcal{K}]$ .*

Together with the directed edges in the CPDAG, the orientation components can completely characterize the common direct causal relations given a CPDAG and a set of DCCs.

**Theorem 35** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs consistent with  $\mathcal{G}^*$ , then  $X \rightarrow Y$  is in every DAG in  $[\mathcal{G}^*, \mathcal{K}]$  if and only if  $X \rightarrow Y$  appears in  $\mathcal{G}^*$  or there exists an orientation component for  $Y$  containing  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .*

Graphically, Theorem 35 shows all possible sources of the directed edges in an MPDAG. Consider Figure 10b as an example. As discussed in Example 7, the undirected induced subgraph over  $\{A, B, X\}$  and  $\{A, B, X, C\}$  are orientation components for  $X$  and  $C$  respectively. If  $X \rightarrow B$  is in a DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$  (Figure 10c), then due to the constraint  $B \xrightarrow{\text{or}} \{X, A\}$  we must have that  $B \rightarrow A$  is in  $\mathcal{G}$ . Again, due to the constraint  $A \xrightarrow{\text{or}} \{X, B\}$ ,  $A \rightarrow X$  is in  $\mathcal{G}$ . However,  $X \rightarrow B \rightarrow A \rightarrow X$  is a directed cycle, meaning that  $X \rightarrow B$  is not in any DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$ , and thus  $B \rightarrow X$  should be in the MPDAG. Similarly,  $A \rightarrow X$  is in the MPDAG (Figure 10d). Since  $B \rightarrow X$  and  $X \xrightarrow{\text{or}} \{B, C\}$ ,  $X \rightarrow C$  is also in the MPDAG (Figure 10e). Finally, applying Meek's rules results  $A \rightarrow C$  and  $B \rightarrow C$  (Figure 10f).

Theorem 35 provides an intuitive method to find the MPDAG by finding all orientation components for all variables, but this method is time-consuming as the number of orientation components for a variable may be very large. The following proposition gives a clue to improve this method by finding the exact orientation component of interest.

---

**Algorithm 3** Finding the maximal orientation component for a vertex.

---

**Require:** A CPDAG  $\mathcal{G}^*$ , a set  $\mathcal{K}$  of DCCs consistent with  $\mathcal{G}^*$ , and a vertex  $X$  in  $\mathcal{G}^*$ .

**Ensure:** The maximal orientation component for  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .

- 1: Set  $\mathcal{U}_m$  to be the chain component containing  $X$ , and compute  $\mathcal{K}_m = \mathcal{K}(\mathcal{U}_m)$ .
  - 2: **while**  $\mathcal{U}_m$  is not an orientation component for  $X$ , **do**
  - 3:   Find a potential leaf node in  $\mathcal{U}_m$  and denote it by  $Y$ .
  - 4:   Update  $\mathcal{U}_m$  by removing  $Y$  and the edges connected to  $Y$ .
  - 5:   Update  $\mathcal{K}_m$  by removing all clauses whose heads contain  $Y$ .
  - 6: **end while**
  - 7: **return**  $\mathcal{U}_m$ .
- 

**Proposition 36** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs consistent with  $\mathcal{G}^*$ . Suppose that with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ ,  $\mathcal{U}_1$  and  $\mathcal{U}_2$  are two orientation components for the same vertex  $X$ , then the undirected induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}_1) \cup \mathbf{V}(\mathcal{U}_2)$  is also an orientation component for  $X$ .*

In some literature, the undirected induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}_1) \cup \mathbf{V}(\mathcal{U}_2)$  is called the union of  $\mathcal{U}_1$  and  $\mathcal{U}_2$ . Thus, Proposition 36 indicates that the union of two orientation components for  $X$  is still an orientation component for  $X$ . This result motivates the definition of a maximal orientation component.

**Definition 37 (Maximal Orientation Component, MOC)** *Given a CPDAG  $\mathcal{G}^*$  and a set  $\mathcal{K}$  of DCCs consistent with  $\mathcal{G}^*$ , with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , an orientation component for a variable  $X$  is called maximal if every orientation component for  $X$  is its induced subgraph.*

Based on Definition 37 and Theorem 35, we have the following corollary.

**Corollary 38** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs consistent with  $\mathcal{G}^*$ , then  $X \rightarrow Y$  is in every DAG in  $[\mathcal{G}^*, \mathcal{K}]$  if and only if  $X \rightarrow Y$  is in  $\mathcal{G}^*$  or the maximal orientation component for  $Y$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  contains  $X$ .*

Corollary 38 yields a method to find the MPDAG of a restricted Markov equivalence class. The key step is to find the maximal orientation component for a variable, whose procedure is given in Algorithm 3. Algorithm 3 is a generalization of Algorithm 1. The first step of Algorithm 3 is to set  $\mathcal{U}_m$  to be the chain component containing  $X$ , and compute  $\mathcal{K}_m = \mathcal{K}(\mathcal{U}_m)$ . This is because the maximal orientation component for  $X$  is an induced subgraph of the chain component to  $X$  belongs. If  $\mathcal{U}_m$  is an orientation component for  $X$ , then it is definitely maximal. Otherwise, the while-loop begins. Since  $X$  is not the only potential leaf node in  $\mathcal{U}_m$ , based on Theorem 25, there must be another potential leaf node  $Y$  in  $\mathcal{U}_m$ . We then remove  $Y$  and the edges connected to  $Y$ . The resulting graph is the induced subgraph of  $\mathcal{U}_m$  over  $\mathbf{V}(\mathcal{U}_m) \setminus \{Y\}$ , and we still denote it by  $\mathcal{U}_m$ . Finally, we remove from  $\mathcal{K}_m$  those clauses whose heads include  $Y$ . This is equivalent to setting  $\mathcal{K}_m = \mathcal{K}_m(\mathcal{U}_m)$ . The while-loop ends when  $\mathcal{U}_m$  is an orientation component for  $X$ . Note that, since in each loop we remove one vertex,  $\mathcal{U}_m$  will eventually become an orientation component for  $X$  in the finite number of loops. The correctness of Algorithm 3 is guaranteed by the following theorem.



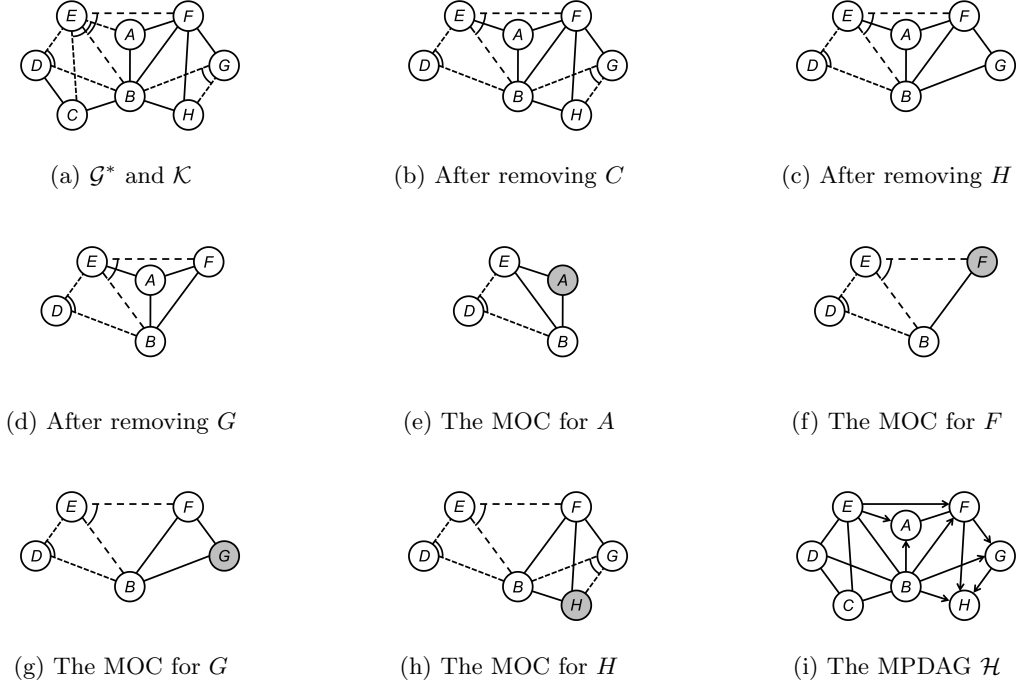


Figure 11: An illustrative example to show how to use Algorithm 3 to find the MPDAG.

**Theorem 39** *The outputted undirected graph of Algorithm 3 is identical to the maximal orientation component for  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .*

Similar to Algorithm 1, the complexity of Algorithm 3 is upper bounded by  $O(|\mathbf{V}(\mathcal{G}^*)|^4 + |\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^2)$ . Applying Algorithm 3 to each vertex separately, we can find the MPDAG of a restricted Markov equivalence class based on Corollary 38, as shown by the following example.

**Example 8** Figure 11a shows a CPDAG  $\mathcal{G}^*$  and a set of DCCs  $\mathcal{K}$  consisting of  $D \xrightarrow{\text{or}} \{E, B\}$ ,  $E \xrightarrow{\text{or}} \{A, C\}$ ,  $E \xrightarrow{\text{or}} \{B, F\}$  and  $G \xrightarrow{\text{or}} \{B, H\}$ . We first show how to use Algorithm 3 to find the maximal orientation component for  $A$ . Note that, since  $\mathcal{G}^*$  is undirected,  $\mathcal{G}_u^* = \mathcal{G}^*$  and  $\mathcal{K}(\mathcal{G}_u^*) = \mathcal{K}$ . As  $C, A, H$  are potential leaf nodes in  $\mathcal{G}_u^*$ ,  $\mathcal{G}_u^*$  is not an orientation component for  $A$ , which triggers the while-loop. In the first loop, we remove  $C$  and the edges connected to  $C$  first, resulting the undirected graph  $\mathcal{U}_m$  shown in Figure 11b, and then remove  $E \xrightarrow{\text{or}} \{A, C\}$  from  $\mathcal{K}$  as it has a head not in  $\mathcal{U}_m$ . The remaining clauses are  $D \xrightarrow{\text{or}} \{E, B\}$ ,  $E \xrightarrow{\text{or}} \{B, F\}$  and  $G \xrightarrow{\text{or}} \{B, H\}$ , which are visualized by arcs in Figure 11b. Since  $H$  is still a potential leaf node in  $\mathcal{U}_m$ , we remove  $H$  and the edges connected to  $H$  from the current graph, and remove  $G \xrightarrow{\text{or}} \{B, H\}$  from the current set of clauses. The result is shown in Figure 11c. In the next two loops, we sequentially remove  $G$  (Figure 11d) and  $F$  (Figure 11e). Finally, we have the undirected graph shown in Figure 11e, which is the maximal orientation component for  $A$ . Similarly, using Algorithm 3 we can find the maximal orientation components for  $F, G$  and  $H$  separately, which are shown in Figures 11f to 11h, respectively. Note that, The maximal orientation components for the remaining

variables are all singleton graphs. Therefore, by Corollary 38,  $E \rightarrow A$ ,  $B \rightarrow A$ ,  $E \rightarrow F$ ,  $B \rightarrow F$ ,  $F \rightarrow G$ ,  $B \rightarrow G$ ,  $B \rightarrow H$ ,  $F \rightarrow H$ , and  $G \rightarrow H$  are all and only directed edges in the MPDAG. The resulting MPDAG is given in Figure 11i.

---

**Algorithm 4** Finding the MPDAG and the minimal residual set (Decomposing pairwise causal constraints).

---

**Require:** A CPDAG  $\mathcal{G}^*$ , a consistent causal constraint set  $\mathcal{B}$ .

**Ensure:** The MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ , a cardinality-minimal residual set of DCCs  $\mathcal{R}$ , and the unique element-wise head-minimal residual set of DCCs  $\mathcal{R}^*$ .

```

1: Construct the equivalent DCCs  $\mathcal{K}$  based on Theorem 17.
2: Let  $\mathcal{H} = \mathcal{G}^*$ .
3: for  $X \in \mathbf{V}(\mathcal{G}^*)$ , do
4:   Find the maximal orientation component for  $X$  according to Algorithm 3 and denote it by  $\mathcal{U}$ .
5:   for  $Y \in \text{adj}(X, \mathcal{U})$ , do
6:     Replace  $Y - X$  by  $Y \rightarrow X$  in  $\mathcal{H}$ .
7:   end for
8: end for
9: Set  $\mathcal{R} = \mathcal{K}$ .
10: while  $\mathcal{R} \neq \emptyset$ , do
11:   if there exists a  $\kappa \in \mathcal{R}$  such that  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R} \setminus \{\kappa\})$  is inconsistent with  $\mathcal{G}^*$ , then
12:     Set  $\mathcal{R} = \mathcal{R} \setminus \{\kappa\}$ .
13:   else
14:     break
15:   end if
16: end while
17: Set  $\mathcal{R}^* = \mathcal{R}$ .
18: for  $\kappa \in \mathcal{R}^*$  do
19:   while there exists a  $D \in \kappa_h$  such that  $\cup_{D' \in \kappa_h \setminus \{D\}} \{D' \rightarrow \kappa_t\} \cup \mathcal{R}^*$  is inconsistent with  $\mathcal{G}^*$ , do
20:     Set  $\kappa_h = \kappa_h \setminus \{D\}$ .
21:   end while
22: end for
23: return  $\mathcal{H}$ ,  $\mathcal{R}$  and  $\mathcal{R}^*$ .

```

---

Algorithm 4 summarizes the procedure for decomposing pairwise causal constraints. Lines 2 to 8 describe how to use Algorithm 3 to construct the MPDAG. To obtain a cardinality-minimal residual set of DCCs  $\mathcal{R}$ , we iteratively remove redundant DCCs  $\kappa$  from  $\mathcal{K}$ , as illustrated in lines 9 to 16. According to Theorem 27,  $\kappa$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R} \setminus \{\kappa\})$  if and only if  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R} \setminus \{\kappa\})$  is inconsistent with  $\mathcal{G}^*$ . Algorithm 1 can be applied to check such consistency. Lines 17 to 22 detail the procedure for obtaining the unique element-wise head-minimal residual DCC set. Starting from the cardinality-minimal set  $\mathcal{R}$ , we prune the heads of each DCC. Specifically, for each

$\kappa \in \mathcal{R}$ , we check whether there exists a  $D \in \kappa_h$  such that  $\cup_{D' \in \kappa_h \setminus \{D\}} \{D' \rightarrow \kappa_t\} \cup \mathcal{R}$  is inconsistent with  $\mathcal{G}^*$ . If so,  $\kappa_t \xrightarrow{\text{or}} \kappa_h \setminus \{D\}$  can be inferred from  $\mathcal{R}$ . Since  $\kappa_t \xrightarrow{\text{or}} \kappa_h \setminus \{D\}$  implies  $\kappa_t \xrightarrow{\text{or}} \kappa_h$ , replacing  $\kappa$  in  $\mathcal{R}$  with  $\kappa_t \xrightarrow{\text{or}} \kappa_h \setminus \{D\}$  yields a DCC set equivalent to  $\mathcal{R}$ . We repeat this pruning procedure until no further reduction is possible. After pruning all DCCs in  $\mathcal{R}$ , we obtain an element-wise head-minimal DCC set. The correctness of this procedure is guaranteed by Definition 28 and statements (iii) and (iv) of Theorem 30.

According to Fang et al. (2022), the complexity of line 1 is at most  $O(|\mathcal{B}| \cdot |\mathbf{V}(\mathcal{G}^*)|^3)$ . Since Algorithm 3 is applied to each vertex separately to find the MPDAG, the complexity of lines 3 to 8 is upper bounded by  $O(|\mathbf{V}(\mathcal{G}^*)|^5 + |\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^3)$ . Next, the while-loop in lines 10 to 16 runs at most  $|\mathcal{K}|$  times, with each iteration taking at most  $O(|\mathbf{V}(\mathcal{G}^*)|^4 + (|\mathcal{K}| + |\mathbf{V}(\mathcal{G}^*)|) \cdot |\mathbf{V}(\mathcal{G}^*)|^2)$  time. Thus, the total complexity of this while-loop is bounded by  $O(|\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^4 + (|\mathcal{K}|^2 + |\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|) \cdot |\mathbf{V}(\mathcal{G}^*)|^2)$ . Similarly, the for-loop in lines 18 to 22 has complexity bounded by  $O(|\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^5 + (|\mathcal{K}|^2 \cdot |\mathbf{V}(\mathcal{G}^*)| + |\mathcal{K}| \cdot |\mathbf{V}(\mathcal{G}^*)|^2) \cdot |\mathbf{V}(\mathcal{G}^*)|^2)$ . In summary, the overall time complexity of Algorithm 4 is polynomial in  $|\mathbf{V}(\mathcal{G}^*)|$ ,  $|\mathcal{K}|$  and  $|\mathcal{B}|$ . Since the sizes of  $\mathcal{B}$  and  $\mathcal{K}$  are both at most quadratic in the number of vertices, the worst-case time complexity of Algorithm 4 is bounded by  $O(|\mathbf{V}(\mathcal{G}^*)|^7)$ .

## 5. Causal Inference with Background Knowledge

In this section, we study the causal inference problem when pairwise causal background knowledge is available. It is well-known that the causal effect of a treatment (or multiple treatments) on a response (or multiple responses) may not be identifiable given a CPDAG, while additional information in background knowledge could make the causal effect identifiable or less uncertain. In Section 5.1, we study the identification condition of a causal effect under restricted Markov equivalence represented by a CPDAG and a set of DCCs, and in Section 5.2, when a causal effect is unidentifiable, we further extend the IDA framework to estimate all possible causal effects.

### 5.1 Identifiability

We first present the definition of causal effect identifiability. This definition is a generalization of Perković (2020, Definition 3.1) to the restricted Markov equivalence class induced by a CPDAG and background knowledge.

**Definition 40 (Causal Effect Identifiability)** *Let  $\mathcal{G}^*$  be a CPDAG over the vertex set  $\mathbf{V}$  and  $\mathcal{B}$  be a consistent pairwise causal constraint set (or a consistent DCC set). Suppose that  $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$  are two disjoint vertex sets. The causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable from  $\mathcal{G}^*$  and  $\mathcal{B}$  (or in  $[\mathcal{G}^*, \mathcal{B}]$ ), if and only if  $f(\mathbf{y} | \text{do}(\mathbf{x}))$  is uniquely computable from any observational distribution Markovian to any  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{B}]$ .*

Our main result of identifiability is given below.

**Theorem 41** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{B}$  be a consistent pairwise causal constraint set (or a consistent DCC set). Denote by  $\mathcal{H}$  the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ . For any two disjoint vertex sets  $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}(\mathcal{G}^*)$ , the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{G}^*, \mathcal{B}]$  if and only if it is identifiable in  $[\mathcal{H}]$ .*

Recall that  $[\mathcal{G}^*, \mathcal{B}] \subseteq [\mathcal{H}]$ , Theorem 41 shows that although  $\mathcal{H}$  carries less knowledge than  $\mathcal{B}$ , it is enough to identify all identifiable causal effects. In other words, the information that cannot be represented by the MPDAG (that is, the residual set of DCCs defined in Section 3.2) contributes nothing to the identifiability.

By Perković (2020), the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{H}]$  if and only if every proper possibly causal path from  $\mathbf{X}$  to  $\mathbf{Y}$  in  $\mathcal{H}$  starts with a directed edge. Therefore, one can graphically identify causal effects by constructing the MPDAG from  $\mathcal{G}^*$  and  $\mathcal{B}$  using Algorithm 4 first.

**Example 9** Consider the MPDAG  $\mathcal{H}$  with two DCCs  $A \xrightarrow{\text{or}} \{B, C\}$  and  $C \xrightarrow{\text{or}} \{A, X\}$  shown in Figure 12a. By Theorem 41, the causal effect of  $X$  on  $Y$  is not identifiable, as  $X - A \rightarrow Y$  and  $X - C \rightarrow Y$  are possibly causal paths on which the first edges are undirected. To verify this result, we enumerate all possible parental sets of  $X$  in Figures 12b to 12e. It can be seen that the causal effect of  $X$  on  $Y$  is definitely zero in Figure 12b, while the causal effects of  $X$  on  $Y$  are possibly non-zero in Figures 12c to 12e.

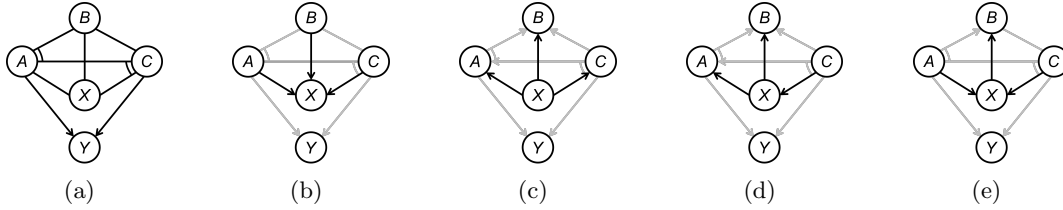


Figure 12: An MPDAG  $\mathcal{H}$  with two DCCs (depicted by arcs) is shown in Figure 12a, and the four possible parental sets of  $X$  are shown in Figures 12b to 12e.

Perković (2020) proved the following formula to calculate identifiable causal effects in an MPDAG. This formula can be directly applied in our setting as Theorem 41 shows that the causal effect identification condition for  $\mathcal{G}^*$  and  $\mathcal{B}$  is the same as that for the MPDAG  $\mathcal{H}$  represents  $[\mathcal{G}^*, \mathcal{B}]$ . Suppose that the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in the MPDAG  $\mathcal{H}$ , then for any distribution  $f$  Markovian to  $\mathcal{G}$ , it holds that

$$f(\mathbf{y} \mid do(\mathbf{x})) = \int \prod_{i=1}^k f(\mathbf{b}_i \mid pa(\mathbf{b}_i, \mathcal{H})) d\mathbf{b},$$

for values  $pa(\mathbf{b}_i, \mathcal{H})$  that are in agreement with  $\mathbf{x}$ , where  $(\mathbf{B}_1, \dots, \mathbf{B}_k) = PCO(an(\mathbf{Y}, \mathcal{H}(\mathbf{V} \setminus \mathbf{X})))$  and  $\mathbf{B} = an(\mathbf{Y}, \mathcal{H}(\mathbf{V} \setminus \mathbf{X})) \setminus \mathbf{Y}$ . The definition of the operator  $PCO$  can be found in Perković (2020, Algorithm 1). This formula generalizes the g-formula of Robins (1986), the truncated factorization formula of Pearl (2009), or the manipulated density formula of Spirtes et al. (2000) to MPDAGs.

The following corollary of Theorem 41 provides conditions under which a pairwise causal background knowledge set can definitely increase the number of identifiable effects.

**Corollary 42** Suppose that  $\mathcal{G}^*$  is a CPDAG over the vertex set  $\mathbf{V}$  and  $\mathcal{K}$  is a set of DCCs consistent with  $\mathcal{G}^*$ , then the following two statements are equivalent.

- (i) At least one unidentifiable effect in  $[\mathcal{G}^*]$  becomes identifiable in  $[\mathcal{G}^*, \mathcal{K}]$ .
- (ii) The DAGs in  $[\mathcal{G}^*, \mathcal{K}]$  have at least one common direct causal relation that is not encoded by a directed edge in  $\mathcal{G}^*$ .

In particular, if  $\mathcal{K}$  is derived from a consistent pairwise causal background knowledge set  $\mathcal{B}$  and there is a direct or non-ancestral causal constraint in  $\mathcal{B}$  which does not hold for all DAGs in  $[\mathcal{G}^*]$ , or there is an ancestral causal constraint  $X \dashrightarrow Y$  in  $\mathcal{B}$  such that  $Y \dashrightarrow X$  does not hold for all DAGs in  $\mathcal{G}^*$ , then at least one unidentifiable effect in  $[\mathcal{G}^*]$  becomes identifiable in  $[\mathcal{G}^*, \mathcal{K}]$ .

Another approach to calculate causal effect is through adjustment. The following theorem gives a sound and complete adjustment criterion.

**Theorem 43** *Let  $\mathcal{G}^*$  be a CPDAG over the vertex set  $\mathbf{V}$  and  $\mathcal{B}$  be a consistent pairwise causal constraint set (or a consistent DCC set). Denote by  $\mathcal{H}$  the MPDAG of  $[\mathcal{G}^*, \mathcal{B}]$ . For any pairwise disjoint vertex sets  $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ ,  $\mathbf{Z}$  is an adjustment set for  $(\mathbf{X}, \mathbf{Y})$  with respect to  $\mathcal{G}^*$  and  $\mathcal{B}$  if and only if  $\mathbf{Z}$  satisfies the b-adjustment criterion relative to  $(\mathbf{X}, \mathbf{Y})$  in  $\mathcal{H}$ .*

The definitions of the adjustment set and the b-adjustment criterion in Theorem 43 are given below.

**Definition 44 (Adjustment Set)** *Let  $\mathcal{G}^*$  be a CPDAG over the vertex set  $\mathbf{V}$  and  $\mathcal{B}$  be a consistent pairwise causal constraint set (or a consistent DCC set). Suppose that  $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$  are pairwise disjoint vertex sets. Then,  $\mathbf{Z}$  is called an adjustment set for  $(\mathbf{X}, \mathbf{Y})$  with respect to  $\mathcal{G}^*$  and  $\mathcal{B}$  if for any DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{B}]$  and observational distribution  $f$  Markovian to  $\mathcal{G}$ , the interventional distribution  $f(\mathbf{y} | do(\mathbf{x}))$  can be calculated by*

$$f(\mathbf{y} | do(\mathbf{x})) = \begin{cases} \int f(\mathbf{y} | \mathbf{x}, \mathbf{z}) f(\mathbf{z}) d\mathbf{z}, & \text{if } \mathbf{Z} \neq \emptyset, \\ f(\mathbf{y} | \mathbf{x}), & \text{otherwise.} \end{cases} \quad (7)$$

**Definition 45 (b-Adjustment Criterion, Perković et al. 2017)** *Let  $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$  be pairwise disjoint vertex sets in an MPDAG  $\mathcal{H}$ , and  $\text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H})$  be the set of variables which are possible descendants of some  $W \notin \mathbf{X}$  lying on a proper possibly causal path from  $\mathbf{X}$  to  $\mathbf{Y}$  in  $\mathcal{H}$ . Then,  $\mathbf{Z}$  satisfies the b-adjustment criterion relative to  $(\mathbf{X}, \mathbf{Y})$  in  $\mathcal{H}$  if: (i) all proper possibly causal paths from  $\mathbf{X}$  to  $\mathbf{Y}$  start with a directed edge in  $\mathcal{H}$ , (ii)  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) = \emptyset$ , and (iii) all proper definite status non-causal paths from  $\mathbf{X}$  to  $\mathbf{Y}$  are blocked by  $\mathbf{Z}$  in  $\mathcal{H}$ .*

Notice that the adjustment criterion in Theorem 43 is the same as that in Perković et al. (2017, Theorem 4.4) for  $[\mathcal{H}]$  induced by an MPDAG  $\mathcal{H}$ . As discussed by Perković (2020), when  $\mathbf{X}$  or  $\mathbf{Y}$  are non-singleton sets, there exist examples showing that some causal effects of  $\mathbf{X}$  on  $\mathbf{Y}$  cannot be identified using adjustment, meaning that not all causal effects can be identified through adjustment. However, under the condition that both  $\mathbf{X}$  and  $\mathbf{Y}$  are singleton sets, the adjustment is sufficient and necessary for identifying causal effects.

Using Theorem 43, the results on optimal adjustment sets for MPDAGs can be naturally extended to general pairwise causal background knowledge. The details can be found in Henckel et al. (2022); Rotnitzky and Smucler (2020).

## 5.2 Estimating Possible Causal Effects

When a causal effect is not identifiable, we can estimate its bounds by enumerating all possible causal effects. Based on the proposed Algorithm 1 for checking consistency, it is straightforward to extend the semi-local IDA (Perković et al., 2017, Algorithm 2) to estimate all possible causal effects of multiple treatments on multiple responses. Moreover, by Theorem 43, the optimal IDA (Witte et al., 2020) and the minimal IDA (Guo and Perković, 2021) can be similarly extended. Thus, in the following, we mainly focus on extending the IDA framework to fully-locally estimate all possible causal effects of a single treatment on a single target. The key result is the following local orientation rules for CPDAGs with DCCs.

**Theorem 46** *Let  $\mathcal{K}$  be a DCC set consistent with a CPDAG  $\mathcal{G}^*$ , and  $\mathcal{H}$  be the MPDAG of  $[\mathcal{G}^*, \mathcal{K}]$ . For any vertex  $X$  and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{H})$ , the following statements are equivalent.*

- (i) *There is a DAG  $\mathcal{G}$  in  $[\mathcal{G}^*, \mathcal{K}]$  such that  $\text{pa}(X, \mathcal{G}) = \mathbf{S} \cup \text{pa}(X, \mathcal{H})$  and  $\text{ch}(X, \mathcal{G}) = \text{sib}(X, \mathcal{H}) \cup \text{ch}(X, \mathcal{H}) \setminus \mathbf{S}$ .*
- (ii) *The restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$  is consistent with  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$ , where  $D_X := \{u \rightarrow X \mid u \in \text{pa}(X, \mathcal{H}) \cup \mathbf{S}\} \cup \{X \rightarrow v \mid v \in \text{sib}(X, \mathcal{H}) \cup \text{ch}(X, \mathcal{H}) \setminus \mathbf{S}\}$ .*

We remark that  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$  is a chordal graph, which can be viewed as a CPDAG (Andersson et al., 1997), and thus the consistency in statement (ii) is well-defined. Theorem 46 provides a method to locally enumerate all possible parental sets of a given  $X$ , which are then used to estimate all possible causal effects. Algorithm 5 shows the procedure.

---

**Algorithm 5** The bgk-IDA algorithm.

---

**Require:** A CPDAG  $\mathcal{G}^*$ , a consistent pairwise causal constraint set  $\mathcal{B}$ , a treatment  $X$ , and a response  $Y$ .

**Ensure:**  $\Theta_X$  which stores all possible causal effects of  $X$  on  $Y$ .

- 1: Derive the DCC set  $\mathcal{K}$  from  $\mathcal{G}^*$  and  $\mathcal{B}$  based on Theorem 17.
  - 2: Construct the MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{K}]$ .
  - 3: **for** each  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{H})$  such that the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$  is consistent with  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$ , where  $D_X$  is defined in Theorem 46,  
**do**
  - 4:   Estimate the causal effect of  $X$  on  $Y$  by adjusting for  $\mathbf{S} \cup \text{pa}(X, \mathcal{H})$ , and add the causal effect to  $\Theta_X$ .
  - 5: **end for**
  - 6: **return**  $\Theta_X$ .
- 

To illustrate Theorem 46, Figure 13 shows the four impossible parental sets of  $X$  in the MPDAG shown in Figure 12a. Recall that  $X$  has 3 siblings, meaning that there are totally 8 candidate parental sets of  $X$ . For example, if we let  $B \rightarrow X$  and  $X \rightarrow \{A, C\}$  (Figure 13a) then  $C \xrightarrow{\text{or}} \{X, A\}$  indicates that  $C \rightarrow A$ , and further causes  $A \rightarrow B$  by the constraint  $A \xrightarrow{\text{or}} \{B, C\}$ . Hence,  $\mathcal{K} \cup \{B \rightarrow X, X \rightarrow A, X \rightarrow C\}$  on  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$  is inconsistent with  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$ , as  $X \rightarrow A \rightarrow B \rightarrow X$  is a directed cycle.

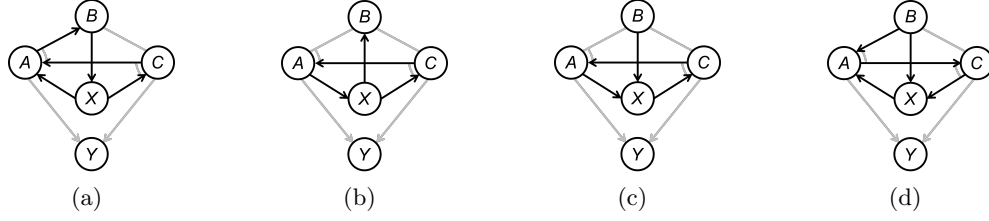


Figure 13: The four impossible parental sets of  $X$  in the MPDAG shown in Figure 12a.

### 5.3 Simulations

We empirically study how pairwise causal background knowledge improves the identifiability of a causal effect in this subsection. We use randomly sampled chordal graphs instead of CPDAGs in our simulations since the impact of the clauses can be separately considered for each chain component of the CPDAG after transforming pairwise causal background knowledge into DCCs (please refer to Lemma 56 in Appendix for details). The chordal graphs were sampled in a reject-sampling manner and the detailed algorithm is in Appendix C.1.

For each combination of  $n \in \{10, 30\}$  and  $e$ , where  $e \in \{10, 15, 20, 25\}$  if  $n = 10$  and  $e \in \{30, 45, 60, 75\}$  if  $n = 30$ , we first sampled 500 chordal graphs with  $n$  vertices and  $e$  edges. Next, for each chordal graph  $\mathcal{G}^*$ , we treated it as a CPDAG and sampled a DAG  $\mathcal{G}$  from  $[\mathcal{G}^*]$ , and assigned each edge in  $\mathcal{G}$  with a weight sampled from  $\text{Uniform}(0.5, 2)$ .  $\mathcal{G}$  was regarded as the underlying true DAG. Then, we sampled a treatment  $X$  as well as a response  $Y$ . Finally, we enumerated all pairs of direct, ancestral, or non-ancestral causal relations in  $\mathcal{G}$  and randomly sampled  $b$  direct, ancestral, or non-ancestral causal constraints from them as background knowledge, and used the bgk-IDA to compute the possible causal effects of  $X$  on  $Y$ , where  $b \in \{0, 1, 2, 3, 4, 5\}$  for  $n = 10$  and  $b \in \{0, 3, 6, 9, 12, 15\}$  for  $n = 30$ . We assumed the data generating mechanism is linear-Gaussian with equal variances, and the causal effects were estimated from the true covariance matrix computed from  $\mathcal{G}$ .

To measure the discrepancy between an estimated multi-set of possible effects and the true effect computed with  $\mathcal{G}$ , we examined four metrics including the causal mean squared error (CMSE) introduced by (Tsirlis et al., 2018; Liu et al., 2020b), the number of possible causal effects, the length of the interval determined by the minimum and maximum values of a set of possible effects, and the Int-MSE introduced by Malinsky and Spirtes (2017). Denote by  $\hat{\Theta}_{XY}$  and  $\theta_{XY}$  the estimated multi-set of possible causal effects and the true causal effect, respectively. The CMSE is defined as

$$\text{CMSE}(\hat{\Theta}_{XY}, \theta_{XY}) = \frac{1}{m} \sum_{i=1}^m [(\hat{\theta}_i - \theta_{XY})^2],$$

where  $m = |\hat{\Theta}_{XY}|$  and  $\hat{\theta}_i \in \hat{\Theta}_{XY}$  is an estimated possible causal effect. The CMSE is calculated as a weighted average of the squared distances between the true effect and each distinct estimated effect in the multi-set, with the weights corresponding to the frequency of each distinct estimate's occurrence in the multi-set. In practice, the CMSE may be inflated if the number of possible effects in the multi-set is large, hence we examined the number of possible causal effects as well as the length of the interval determined by the

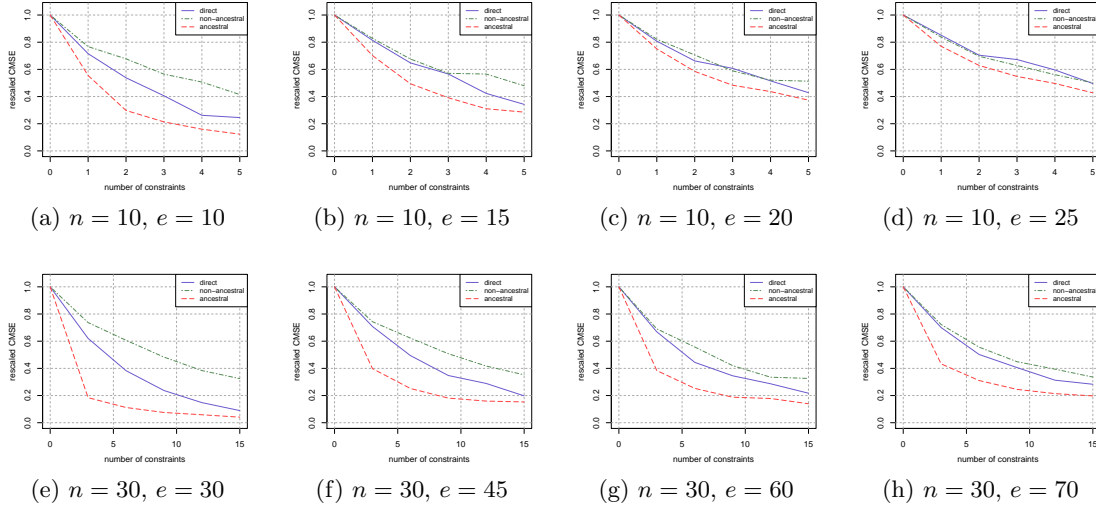


Figure 14: The rescaled CMSEs.

minimum and maximum possible effects. The Int-MSE, introduced by Malinsky and Spirtes (2017), represents the mean absolute distance between the true effect,  $\theta$ , and the interval  $[\hat{\theta}_{\min}, \hat{\theta}_{\max}]$ . The distance  $d$  is defined as  $d = 0$  if the interval covers the true effect and  $d = \min(|\theta - \hat{\theta}_{\min}|, |\theta - \hat{\theta}_{\max}|)$  otherwise. The Int-MSE thus evaluates whether an estimated interval covers the true effect. For all four metrics discussed above, lower values indicate higher identifiability of the causal effect.

Since the causal effects were estimated from the true covariance matrix derived from  $\mathcal{G}$ , and given the correctness of IDA-type algorithms, the range between the minimum and maximum possible causal effects should contain the true effect. Therefore, the Int-MSE should be (nearly) zero. In fact, in our simulations, all Int-MSE values are below  $10^{-8}$ , and non-zero Int-MSE values are due to numerical errors. This result also suggests the correctness of Algorithm 5.

Except for Int-MSE, the other metrics show similar results. We focus on CMSE in what follows, and the results for the other metrics are provided in Appendix C.2. For each sampled DAG and variable pair  $(X, Y)$ , we computed three sequences of CMSEs corresponding to three types of background knowledge. Each sequence contains six elements corresponding to the six possible values of  $b$ . For each sequence, all CMSEs were normalized by the CMSE at  $b = 0$ , yielding the so-called rescaled CMSE.

Figure 14 shows the sequences of rescaled CMSEs under different types of causal constraints. The rescaled CMSE decreases rapidly as the number of constraints increases. Moreover, given the same number of constraints, the rescaled CMSE under ancestral causal constraints is much lower than that under direct causal constraints, which in turn is lower than that under non-ancestral constraints. This phenomenon occurs because ancestral causal constraints are more informative than non-ancestral ones: if  $X$  is a cause of  $Y$ , then  $Y$  cannot be a cause of  $X$ , but not vice versa.



## 6. Discussion

Pairwise causal background knowledge is frequently encountered in real-world problems. Assuming both pairwise causal background knowledge and a sufficiently large observational data set are available, this paper systematically studies the representation of pairwise causal background knowledge, and demonstrates the potential of exploiting pairwise causal background knowledge in causal inference. The main contribution of the paper is three-fold.

Firstly, we investigate the graphical characterization of causal MPDAGs. We present sufficient and necessary graphical conditions for a partially directed graph to be a causal MPDAG. MPDAGs are important in representing common direct causal relations in a restricted Markov equivalent class. Our graphical characterization generalizes the existing results for essential graphs (Andersson et al., 1997) and intervention essential graphs (Hauser and Bühlmann, 2012).

Despite the wide use of causal MPDAGs, they may fail to represent ancestral causal knowledge exactly. Therefore, we develop direct causal clauses to represent all types of pairwise causal background knowledge in a unified form. Because of the local nature of the direct causal clauses, our new representation brings a lot of convenience. As a result, we can now check the consistency of pairwise causal background knowledge, or the equivalence of two pairwise causal background knowledge sets, in polynomial time. Moreover, we prove that any pairwise causal background knowledge set can be decomposed into a causal MPDAG and a minimal residual set of direct causal clauses, and the decomposition can be achieved in polynomial time, too.

The decomposition of pairwise causal background knowledge plays an important role in causal inference. The third contribution of our work is that, we show that the decomposed causal MPDAG entirely determines the identifiability of a causal effect, and the residual direct causal clauses alone contribute nothing to the identifiability, but may reduce the possible values of an unidentifiable effect. We also develop IDA-type algorithms to locally or semi-locally estimate possible causal effects. Using the proven sufficient and necessary identification condition, the adjustment criterion, and the IDA-type algorithms, we can identify causal effects as well as estimate their values or bounds.

There are also many topics worthy of further investigation. First, although the proposed algorithms run in polynomial time, they may not be the most time-efficient. Determining the optimal time complexity for consistency checking and MPDAG construction, as well as developing corresponding optimal algorithms, remains an important direction for future research. Moreover, efficiently enumerating Markov equivalent DAGs that satisfy given pairwise causal background knowledge remains an interesting open problem. The local properties of the DCC representation may help prune the search space and facilitate the design of more efficient enumeration algorithms. Furthermore, our approach assumes that the pairwise causal background knowledge is consistent with the true CPDAG (or the learned CPDAG). Extending our framework to accommodate inconsistent background knowledge is another important direction. Finally, studying the representation of pairwise causal background knowledge in the presence of hidden variables and selection bias is also an interesting direction for future work.

## Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2022ZD0160300). Liu's work was partially supported by the National Natural Science Foundation of China (Grant No. 12201629). We are grateful to the editor and the reviewers for their valuable comments and suggestions, which have significantly improved the quality of the paper.

## Appendix A. Additional Remarks

Some additional topics are discussed in this section.

### A.1 From DCCs to Pairwise Causal Constraints

Assuming that  $[\mathcal{H}, \mathcal{R}]$  is equivalent to some set of pairwise causal constraints, we can obtain the maximal equivalent set of such constraints by using Theorems 17 and 27 and Algorithm 1.

We begin by enumerating all pairs of nodes  $X$  and  $Y$ , and checking whether  $X$  is an ancestor of  $Y$  in all DAGs within  $[\mathcal{H}, \mathcal{R}]$ . This is done by transforming the proposition into the corresponding DCC,  $X \xrightarrow{\text{or}} \mathbf{C}_{XY}(\mathcal{G}^*)$ , using Theorem 17, and then determining whether  $X \xrightarrow{\text{or}} \mathbf{C}_{XY}(\mathcal{G}^*)$  is redundant given  $\mathcal{H}$  and  $\mathcal{R}$ , based on Theorem 27 and the consistency-checking algorithm (Algorithm 1). If  $X \xrightarrow{\text{or}} \mathbf{C}_{XY}(\mathcal{G}^*)$  is found to be redundant, then  $X$  is an ancestor of  $Y$  in all DAGs in  $[\mathcal{H}, \mathcal{R}]$ , indicating that the proposition “ $X$  is an ancestor of  $Y$ ” can be recovered from  $[\mathcal{H}, \mathcal{R}]$ . Similarly, we can check whether the proposition “ $X$  is not an ancestor of  $Y$ ” holds in all DAGs in  $[\mathcal{H}, \mathcal{R}]$ . Denote by  $\mathcal{B}$  the set of all ancestral and non-ancestral causal constraints that hold for all DAGs in  $[\mathcal{H}, \mathcal{R}]$ . This set  $\mathcal{B}$  constitutes the maximal set of pairwise causal constraints that can be translated back from  $[\mathcal{H}, \mathcal{R}]$ .

Moreover, one can iteratively prune the redundant constraints in  $\mathcal{B}$  to further obtain a minimal equivalent set of constraints. More formally, for each  $b \in \mathcal{B}$ , we transform it into its corresponding DCCs, denoted by  $\kappa(b)$ . We then apply Theorem 27 and the consistency-checking algorithm (Algorithm 1) to determine whether  $\kappa(b)$  is redundant given the DCCs derived from  $\mathcal{B} \setminus b$ . The resulting subset of  $\mathcal{B}$  constitutes a minimal set of pairwise causal constraints equivalent to  $[\mathcal{H}, \mathcal{R}]$ .

### A.2 A Sequential Method for Checking Consistency

In this section, we present an approach to sequentially check consistency. Suppose that  $\mathcal{G}^*$  is a CPDAG and  $\mathcal{K}$  is a set of consistent DCCs, and  $\mathcal{H}$  is the MPDAG representing  $[\mathcal{G}^*, \mathcal{K}]$ . A set of DCCs  $\mathcal{K}'$  is called consistent with  $\mathcal{G}^*$  given  $\mathcal{K}$ , if  $\mathcal{K}' \cup \mathcal{K}$  is consistent with  $\mathcal{G}^*$ . For any DCC  $(X \xrightarrow{\text{or}} \mathbf{D}) \in \mathcal{K}'$ , it holds that,

$$\begin{aligned}
 \mathcal{K} \cup \{X \xrightarrow{\text{or}} \mathbf{D}\} \text{ is inconsistent} &\iff X \xrightarrow{\text{or}} \mathbf{D} \text{ does not hold for any DAG in } [\mathcal{G}^*, \mathcal{K}] \\
 &\iff \forall \mathcal{G} \in [\mathcal{G}^*, \mathcal{K}], \text{ch}(X, \mathcal{G}) \cap \mathbf{D} = \emptyset \\
 &\iff \forall \mathcal{G} \in [\mathcal{G}^*, \mathcal{K}], \mathbf{D} \subseteq \text{pa}(X, \mathcal{G}) \\
 &\iff \forall \mathcal{G} \in [\mathcal{G}^*, \mathcal{K}], \mathbf{D} \rightarrow X \\
 &\iff \mathbf{D} \subseteq \text{pa}(X, \mathcal{H}).
 \end{aligned}$$

Therefore, the consistency of  $\mathcal{K}'$  with  $\mathcal{G}^*$  given  $\mathcal{K}$  can be checked sequentially: picking one DCC from the current  $\mathcal{K}'$  at a time; if the clause is consistent with  $\mathcal{G}^*$  given  $\mathcal{K}$ , then adding it to the current  $\mathcal{K}$  and updating the current MPDAG based on  $\mathcal{K}$ .

**Example 10** *Following Example 6, we now use the sequential method to check the consistency of  $\mathcal{K} = \{C \xrightarrow{\text{or}} A, B \xrightarrow{\text{or}} A, D \xrightarrow{\text{or}} \{A, X\}\}$  with respect to  $\mathcal{G}^*$  shown in Figure 7a. If the first DCC chosen from  $\mathcal{K}$  is  $D \xrightarrow{\text{or}} \{A, X\}$ , then the updated MPDAG, denoted by  $\mathcal{H}$ , is*

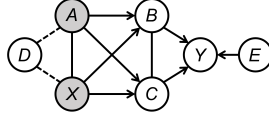


Figure 15: An MPDAG given in Example 10.

the one shown in Figure 15. Since  $A \rightarrow B$  and  $A \rightarrow C$  are both in  $\mathcal{H}$ , no matter what clause will be chosen next, the procedure returns False. Similarly, if we choose  $B \rightarrow A$  and  $C \rightarrow A$  in the first two loops, then based on Rule 4 of Meek's rules, the resulting MPDAG must have edges  $A \rightarrow D$  and  $X \rightarrow D$ . Hence,  $\mathcal{K}$  is inconsistent once we consider the remaining clause  $D \xrightarrow{\text{or}} \{A, X\}$ .

## Appendix B. Proofs

Before presenting all detailed proofs of the theorems, propositions and corollaries of the main text, we introduce some helpful concepts and results in Appendix B.1. In the following paper, for a variable  $X$  and a non-empty variable set  $\mathbf{Y}$ , we will use the notion  $\mathbf{Y} \rightarrow X$  to represent  $Y \rightarrow X$  for all  $Y \in \mathbf{Y}$ , and use the notion  $X \rightarrow \mathbf{Y}$  to represent  $X \rightarrow Y$  for all  $Y \in \mathbf{Y}$ .

### B.1 Preliminaries

We briefly review some graphical properties of chordal graphs and CPDAGs. In a graph, a cycle with length three is called a triangle. A path is called unshielded if none of its three consecutive vertices form a triangle. For a graph  $\mathcal{G}$  over a vertex set  $\mathbf{V}$ ,  $\mathbf{M} \subseteq \mathbf{V}$  is called a clique if  $\mathbf{M}$  induces a complete subgraph. A clique is called maximal if it is not a proper subset of any other cliques. Let  $\mathcal{C}$  be a chordal graph over the vertex set  $\mathbf{V}(\mathcal{C})$ . Any induced subgraph of  $\mathcal{C}$  is chordal. It can be proved that any chordal graph has a simplicial vertex, and moreover, any non-complete chordal graph has two non-adjacent simplicial vertices (Blair and Peyton, 1993). A perfect elimination ordering (PEO) of  $\mathcal{C}$  is a total ordering of the vertices in  $\mathbf{V}(\mathcal{C})$ , denoted by  $\beta = (V_1, V_2, \dots, V_n)$ , such that for any  $V_i$ ,  $i = 1, 2, \dots, n$ ,  $\text{adj}(V_i, \mathcal{C}) \cap \{V_i, V_{i+1}, \dots, V_n\}$  induces a complete subgraph of  $\mathcal{C}$ . An undirected graph is chordal if and only if it has a PEO.

Given a PEO  $\beta = (V_1, V_2, \dots, V_n)$  of  $\mathcal{C}$ , if we orient the edges in  $\mathcal{C}$  such that  $\text{adj}(V_i, \mathcal{C}) \cap \{V_i, V_{i+1}, \dots, V_n\}$  are parents of  $V_i$ , then the resulting directed graph is acyclic and v-structure-free. Conversely, any v-structure-free DAG who has the same skeleton as  $\mathcal{C}$  can be oriented from  $\mathcal{C}$  according to some PEO of  $\mathcal{C}$ .

Let  $\mathcal{G}^*$  be a CPDAG. It was pointed by Maathuis et al. (2009) that (i) no orientation of the edges not oriented in  $\mathcal{G}^*$  will create a directed cycle which includes an edge or edges that were oriented in  $\mathcal{G}^*$ , and (ii) no orientation of an edge not directed in  $\mathcal{G}^*$  can create a new v-structure with an edge that was oriented in  $\mathcal{G}^*$ . As any orientation of the edges in  $\mathcal{G}^*$  which does not create directed cycles or v-structures corresponds to a DAG in  $[\mathcal{G}^*]$ , we can separately orient the undirected edges in each chain component such that every resulting directed graph is a DAG without v-structure.

## B.2 Proof of Proposition 9

We first prove Proposition 9 as it is required in proving Theorem 8.

**Proof** The first claim holds because the definition of a causal MPDAG implies that for any restricted Markov equivalence class  $[\mathcal{G}^*, \mathcal{B}]$  that can be represented by  $\mathcal{H}$ ,  $\mathcal{G}^*$  and  $\mathcal{H}$  have the same skeleton and v-structures.

To prove the second claim, let  $[\mathcal{G}^*, \mathcal{B}]$  be a restricted Markov equivalence class that can be represented by  $\mathcal{H}$ , and let  $\mathcal{B}_d = \mathbf{E}_d(\mathcal{H}) \setminus \mathbf{E}_d(\mathcal{G}^*)$ . It is easy to verify that orienting undirected edges in  $\mathcal{G}^*$  according to  $\mathcal{B}_d$  does not introduce any v-structure or directed cycle and the resulting PDAG is closed under Meek's rules, as the resulting PDAG is exactly  $\mathcal{H}$ . Therefore,  $[\mathcal{H}] = [\mathcal{G}^*, \mathcal{B}_d]$ .  $\blacksquare$

## B.3 Proof of Theorem 8

We first introduce two properties of chordal graphs. For the completeness of the paper, the proofs of these results are also provided.

**Lemma 47** *Let  $\mathcal{C}$  be a connected chordal graph, then the following claims hold for  $\mathcal{C}$ .*

- (i) *For any simplicial vertex  $X$  in  $\mathcal{C}$ , there is a unique maximal clique containing  $X$ .*
- (ii) *If  $\mathcal{C}$  is not complete, then for any non-simplicial vertex  $Y$  in  $\mathcal{C}$  that is adjacent to some simplicial vertex  $X$ ,  $Y$  has a neighbor  $Z \neq X$  which is not adjacent to  $X$ .*

**Proof** [Proof of lemma 47] For any simplicial vertex  $X$  in  $\mathcal{C}$ , if there are two distinct maximal cliques  $\mathbf{M}_i, \mathbf{M}_j$  containing  $X$ , then there exist  $X_i \in \mathbf{M}_i$  and  $X_j \in \mathbf{M}_j$  such that  $X_i$  and  $X_j$  are not adjacent, since otherwise  $\mathbf{M}_i \cup \mathbf{M}_j$  is also a maximal clique. Note that both  $X_i$  and  $X_j$  are adjacent to  $X$ , hence  $X$  is not simplicial, which is contrary to the assumption. This completes the proof of the first claim.

We next prove the second claim. Let  $\mathbf{M}_i$  be the maximal clique containing  $X$ . Since  $X$  is simplicial,  $Y$  is adjacent to  $X$  implies that  $Y \in \mathbf{M}_i$ . Assume, for the sake of contradiction, that every neighbor of  $Y$  other than  $X$  is adjacent to  $X$ , then  $\text{adj}(Y, \mathcal{C}) \subseteq \mathbf{M}_i$ , which means  $\text{adj}(Y, \mathcal{C})$  induces a complete subgraph of  $\mathcal{C}$ . This contradicts the assumption that  $Y$  is non-simplicial. Therefore, the second claim holds true.  $\blacksquare$

Next, we present Lemmas 48-54 in the following. These lemmas contribute to the proof of the necessity of Theorem 8.

**Lemma 48 (Necessity of Condition (ii))** *The skeleton of  $\mathcal{C}^b$  is a chordal graph for any B-component  $\mathcal{C}^b$  of a causal MPDAG  $\mathcal{H}$ .*

**Proof** [Proof of lemma 48] According to Proposition 9, the skeleton of each B-component of a causal MPDAG is an induced subgraph of a chain component of a CPDAG. The result comes from the fact that any induced subgraph of a chordal graph is still chordal.  $\blacksquare$

**Lemma 49 (Necessity of Condition (iii))** *Let  $\mathcal{H}$  be a causal MPDAG and  $\mathcal{C}^b$  be a B-component of  $\mathcal{H}$ . For any vertex  $X \notin \mathbf{V}(\mathcal{C}^b)$ , if  $X \rightarrow Y$  for some vertex  $Y \in \mathbf{V}(\mathcal{C}^b)$ , then  $X \rightarrow Y$  for every vertex  $Y \in \mathbf{V}(\mathcal{C}^b)$ .*

**Proof** [Proof of Lemma 49] If  $|\mathbf{V}(\mathcal{C}^b)| = 1$ , then Lemma 49 naturally holds. Assume that  $|\mathbf{V}(\mathcal{C}^b)| > 1$ , then  $\text{sib}(Y, \mathcal{C}^b) \neq \emptyset$  by the definition of a B-component. Since every B-component is a connected graph, to prove Lemma 49 it suffices to show that  $X \rightarrow Z$  for every  $Z \in \text{sib}(Y, \mathcal{C}^b)$ . This is because every vertex in  $\mathcal{C}^b$  other than  $Y$  is connected to  $Y$  by an undirected path. If the conclusion holds for every  $Z \in \text{sib}(Y, \mathcal{C}^b)$ , then the same argument can be successively applied to every vertex along the path.

For any vertex  $Z \in \text{sib}(Y, \mathcal{C}^b)$ , it is clear that  $X$  and  $Z$  are adjacent, since otherwise by Rule 1 of Meek's rules, it holds that  $Y \rightarrow Z$  is in  $\mathcal{H}$ , which is a contradiction to  $Z \in \text{sib}(Y, \mathcal{C}^b)$ . As  $X \notin \mathbf{V}(\mathcal{C}^b)$ ,  $X$  and  $Z$  must be connected by a directed edge. If  $Z \rightarrow X$ , then  $Z - Y$  can be oriented as  $Z \rightarrow Y$  by Rule 2 of Meek's rules, which also contradicts to  $Z \in \text{sib}(Y, \mathcal{C}^b)$ . Thus, we have  $X \rightarrow Z$ .  $\blacksquare$

**Lemma 50 (Necessity of Condition (i))** *Given a causal MPDAG  $\mathcal{H}$ , the chain skeleton  $\mathcal{H}_c$  of  $\mathcal{H}$  is a chain graph. Furthermore,  $\mathcal{H}_c$  is also an MPDAG.*

**Proof** [Proof of Lemma 50] We first prove that the graph  $\mathcal{H}_c$  is a chain graph, which suffices to show that there are no partially directed cycles in  $\mathcal{H}_c$ . Assume that there is a partially directed cycle in  $\mathcal{H}_c$ , and the cycle is of the following form:  $X_{11} \rightarrow X_{21} - \dots - X_{2n_2} \rightarrow \dots \rightarrow X_{k1} - \dots - X_{kn_k} \rightarrow X_{1n_1} - \dots - X_{11}$ . Based on the definitions of  $\mathcal{H}_c$  and B-component,  $X_{21}, \dots, X_{2n_2}$  are in the same B-component while  $X_{11}$  is not in this B-component. By Lemma 49, it holds that  $X_{11} \rightarrow X_{2n_2}$ . Similarly,  $X_{2n_2} \rightarrow X_{3n_3} \rightarrow \dots \rightarrow X_{kn_k} \rightarrow X_{11}$ , which together with  $X_{11} \rightarrow X_{2n_2}$  gives a directed cycle. Since all directed edges in  $\mathcal{H}_c$  are also in  $\mathcal{H}$ , we have constructed a directed cycle in  $\mathcal{H}$ , which contradicts to the definition of a causal MPDAG.

We then show that  $\mathcal{H}_c$  is an MPDAG. Since we have already proved that there is no (partially) directed cycle in  $\mathcal{H}_c$ ,  $\mathcal{H}_c$  is a PDAG. What remains is to show that  $\mathcal{H}_c$  is closed under the four Meek's rules.

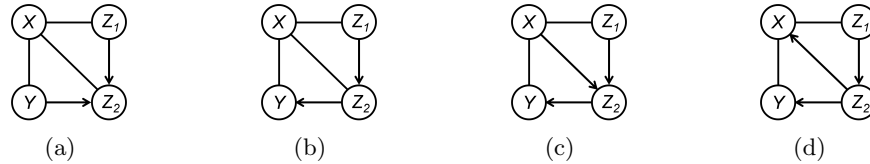


Figure 16: The cases discussed in the proof of Lemma 50

- (i) If  $\mathcal{H}_c$  is not closed under the first Meek's rule, then  $\mathcal{H}_c$  has an induced subgraph  $X \rightarrow Y - Z$ , in which  $X \notin \text{adj}(Z, \mathcal{H}_c)$ . By the construction of  $\mathcal{H}_c$ ,  $Y$  and  $Z$  are in the same B-component of  $\mathcal{H}$  while  $X$  is not in that B-component. According to Lemma 49,  $X \rightarrow Z$  must be in  $\mathcal{H}_c$ , which contradicts the assumption that  $X \notin \text{adj}(Z, \mathcal{H}_c)$ .

- (ii) If  $\mathcal{H}_c$  is not closed under the second Meek's rule, then  $\mathcal{H}_c$  has an induced subgraph consisting of  $X \rightarrow Y \rightarrow Z$  and  $X - Z$ . By the similar argument for (i),  $Y \rightarrow X$  must be in  $\mathcal{H}_c$ , which leads to a contradiction.
- (iii) If  $\mathcal{H}_c$  is not closed under the third Meek's rule, then  $\mathcal{H}_c$  has an induced subgraph with the configuration shown in Figure 16a. In this case,  $X, Y, Z_1, Z_2$  are in the same B-component. Again, by the construction of  $\mathcal{H}_c$ ,  $Y \rightarrow Z_2, Z_1 \rightarrow Z_2$  should not be in  $\mathcal{H}_c$ , which leads to a contradiction.
- (iv) If  $\mathcal{H}_c$  is not closed under the fourth Meek's rule, then one of the configurations shown in Figures 16b-16d must appear in  $\mathcal{H}_c$  as an induced subgraph. In the following, we will prove that none of the configurations is in  $\mathcal{H}_c$ . In fact, if  $\mathcal{H}_c$  has an induced subgraph with the configuration shown in Figure 16b, then by the similar argument for (iii),  $Z_1 \rightarrow Z_2 \rightarrow Y$  should not be in  $\mathcal{H}_c$  as  $X, Y, Z_1, Z_2$  are in the same B-component of  $\mathcal{H}$ . If  $\mathcal{H}_c$  has an induced subgraph with the configuration shown in Figure 16c, then the induced subgraph of  $\mathcal{H}_c$  over  $X, Z_2, Y$  is not closed under the second Meek's rule. Similarly, in Figure 16d, the induced subgraph of  $\mathcal{H}_c$  over  $Z_1, Z_2, X$  is not closed under the second Meek's rule either.

This completes the proof. ■

**Lemma 51** *Given a causal MPDAG  $\mathcal{H}$ , if a directed edge in  $\mathcal{H}$  can be oriented by Rule 1 or Rule 4 of Meek's rules, then the involved vertices of the corresponding Meek's rule are not in the same B-component.*

We remark that, a directed edge  $Y \rightarrow Z$  can be oriented by Rule 1 means that there is an  $X \notin \text{adj}(Z, \mathcal{H})$  such that  $X \rightarrow Y$  is in  $\mathcal{H}$ . In this case,  $X, Y, Z$  are the involved vertices of Rule 1. The meaning of the expression that "a directed edge can be oriented by Rule 4" is similar. We also note that, the condition of Lemma 51 does not rule out the possibility that the edge can also be oriented by Rules 2, 3 or from the background knowledge set.

**Proof** [Proof of Lemma 51] If a directed edge  $Y \rightarrow Z$  can be oriented by Rule 1 of Meek's rules, then there exists a vertex  $X \in \mathbf{V}(\mathcal{H})$  such that  $X \rightarrow Y$  and  $X \notin \text{adj}(Z, \mathcal{H})$ . We need to prove that  $X, Y, Z$  are not in the same B-component of  $\mathcal{H}$ . Assume, for the sake of contradiction, that  $X, Y, Z$  are in the same B-component  $\mathcal{C}^b$ , then there is an undirected path  $Y - W_1 - \dots - W_n - Z$  connecting  $Y$  and  $Z$  where  $n \geq 1$  and  $W_i \in \mathcal{C}^b$  for  $i = 1, 2, \dots, n$ . Since  $Y - W_1$  is undirected, we have  $X \in \text{adj}(W_1, \mathcal{H})$ , otherwise we would have  $Y \rightarrow W_1$  due to Rule 1 of Meek's rules. Similarly,  $Y \in \text{adj}(W_n, \mathcal{H})$  since  $Y \rightarrow Z - W_n$  is in  $\mathcal{H}$ .

If  $n = 1$  or  $X \in \text{adj}(W_n, \mathcal{H})$ , then  $W_n \rightarrow Z$  should be in  $\mathcal{H}$  by Rule 4 of Meek's rules (Figure 17), which contradicts our assumption. Now consider the case where  $n > 1$  and  $X \notin \text{adj}(W_n, \mathcal{H})$ . Since  $X \rightarrow Y$ , by the first Meek's rule we have  $Y \rightarrow W_n$ . Moreover, since  $W_{n-1} - W_n$ , it holds that  $Y \in \text{adj}(W_{n-1}, \mathcal{H})$ . Let  $k := \arg \max_{1 \leq j \leq n-1} X \in \text{adj}(W_j, \mathcal{H})$  denote the largest subscript of the vertex on the path which is adjacent to  $X$ , then by the same argument we can show that  $Y \rightarrow W_i$  for  $i = k + 1, \dots, n$  and  $Y \in \text{adj}(W_k, \mathcal{H})$ . Note that, the induced subgraph of  $\mathcal{H}$  over  $X, Y, W_{k+1}$  and  $W_k$  is the same as the left-hand side of the fourth Meek's rule,  $W_k \rightarrow W_{k+1}$  is in  $\mathcal{H}$ , which is contrary to our assumption.

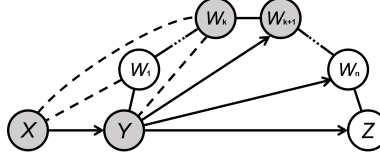


Figure 17: An illustration of the graph structure discussed in the proof of Lemma 51. A dashed undirected edge connecting two vertices indicates they are adjacent, but the direction of the edge is not relevant to the proof.

If a directed edge  $X \rightarrow Y$  can be oriented by Rule 4 of Meek's rules but cannot be oriented by Rule 1 of Meek's rules, then there exist vertices  $Z_1, Z_2 \in \mathcal{H}$  such that  $\mathcal{H}$  has an induced subgraph shown in Figure 16b. Note that,  $Z_2 \rightarrow Y$  can be oriented by Rule 1 of Meek's rules since  $Z_1 \rightarrow Z_2$  and  $Z_1$  is not adjacent to  $Y$ ,  $Z_1, Z_2$  and  $Y$  are not in the same B-component by the first part of the proof. Thus,  $X, Z_1, Z_2$  and  $Y$  are not in the same B-component, which completes the proof. ■

**Lemma 52** *Let  $\mathcal{H}$  be a causal MPDAG and  $\mathcal{C}^b$  be a B-component of  $\mathcal{H}$ . If a directed edge in  $\mathcal{C}^b$  can be oriented by Meek's rules, then it can only be oriented by Rule 2 of Meek's Rules, and the directed edges in the configuration on the left-hand side of Rule 2 are all in  $\mathcal{C}^b$ .*

**Proof** [Proof of Lemma 52] Let  $X \rightarrow Y$  be a directed edge in  $\mathcal{C}^b$  that can be oriented by Meek's rules. If  $X \rightarrow Y$  can be oriented by the first Meek's rule, then there is a  $Z \notin \text{adj}(Y, \mathcal{H})$  such that  $Z \rightarrow X$  is in  $\mathcal{H}$ . By Lemma 49,  $Z \in \mathbf{V}(\mathcal{C}^b)$ . However, this is impossible based on Lemma 51. Therefore,  $X \rightarrow Y$  cannot be oriented by the first Meek's rule.

If  $X \rightarrow Y$  can be oriented by the third Meek's rule, then there are  $Z_1, Z_2 \in \text{adj}(X, \mathcal{H})$  such that  $Z_1 \rightarrow Y \leftarrow Z_2$  is a v-structure in  $\mathcal{H}$  while  $Z_1 \rightarrow X \leftarrow Z_2$  is not in  $\mathcal{H}$ . If  $Z_1 \notin \mathbf{V}(\mathcal{C}^b)$ , then  $Z_1 \rightarrow X \rightarrow Z_2$  by Lemma 49 and the first Meek's rule. On the other hand,  $Z_2 \rightarrow Y$  while  $Z_2 \not\rightarrow X$  implies that  $Z_2 \in \mathbf{V}(\mathcal{C}^b)$  by Lemma 49. However, this is impossible by Lemma 49, since  $Z_1$  and  $Z_2$  are not adjacent. Therefore,  $Z_1 \in \mathbf{V}(\mathcal{C}^b)$ . Similarly,  $Z_2 \in \mathbf{V}(\mathcal{C}^b)$ , meaning that  $X, Y, Z_1$  and  $Z_2$  are in the same B-component of  $\mathcal{H}$ . However, as implied by Proposition 9, any B-component is an induced subgraph of some chain component of a CPDAG, and thus the v-structure  $Z_1 \rightarrow Y \leftarrow Z_2$  is not allowed in  $\mathcal{C}^b$ , leading to a contradiction.

If  $X \rightarrow Y$  can be oriented by the fourth Meek's rule but cannot be oriented by the first Meek's rule, then there are  $Z_1 \in \text{sib}(X, \mathcal{H})$  and  $Z_2 \in \text{sib}(X, \mathcal{H})$  such that  $X - Z_1 \rightarrow Z_2 \rightarrow Y$  and  $Z_1 \notin \text{adj}(Y, \mathcal{H})$ . Since  $Z_1 - X$  and  $Z_2 - X$ , we have that  $Z_1, Z_2, X$  and  $Y$  are in the same B-component  $\mathcal{C}^b$ . However,  $Z_2 \rightarrow Y$  can be oriented by the first Meek's rule, which is contrary to Lemma 51. Therefore,  $X \rightarrow Y$  cannot be oriented by the fourth Meek's rule.

Finally, if  $X \rightarrow Y$  can be oriented by the second Meek's rule, then there is a  $Z$  such that  $X \rightarrow Z \rightarrow Y$  is in  $\mathcal{H}$ . By Lemma 49,  $Z$  is in  $\mathcal{C}^b$ . Thus,  $X \rightarrow Z$  and  $Z \rightarrow Y$  are all in  $\mathcal{C}^b$ . ■



**Lemma 53** *Given a causal MPDAG  $\mathcal{H}$  and a B-component  $\mathcal{C}^b$  of  $\mathcal{H}$ , let  $\mathbf{M}_i$  and  $\mathbf{M}_j$  be two distinct maximal cliques of  $\mathcal{C}^b$  such that  $\mathbf{M}_{ij} := \mathbf{M}_i \cap \mathbf{M}_j \neq \emptyset$ . For any vertex  $X \in \mathbf{M}_i \setminus \mathbf{M}_{ij}$ ,  $Y \in \mathbf{M}_{ij}$ , the directed edge  $X \rightarrow Y$  does not exist in  $\mathcal{C}^b$ .*

**Proof** [Proof of Lemma 53]  $X \in \mathbf{M}_i \setminus \mathbf{M}_{ij}$  implies that there must be a  $Z \in \mathbf{M}_j \setminus \mathbf{M}_{ij}$  such that  $Z$  is not adjacent to  $X$ , since otherwise  $X$  is adjacent to every vertex in  $\mathbf{M}_j$  and consequently  $X \in \mathbf{M}_j$ . Assume that such a directed edge  $X \rightarrow Y$  exists in  $\mathcal{H}$ , then such a  $Z$  must be a child of  $Y$  in  $\mathcal{H}$  based on the first Meek's rule. This contradicts Lemma 52. ■

**Lemma 54 (Necessity of Condition (iv))** *Suppose that  $\mathcal{H}$  is a causal MPDAG and  $\mathcal{C}^b$  is a B-component of  $\mathcal{H}$ . For any directed edge  $X \rightarrow Y$  in  $\mathcal{C}^b$ , it can be proved that (i)  $\text{adj}(Y, \mathcal{C}^b) \setminus \{X\} \subseteq \text{adj}(X, \mathcal{C}^b)$ , and (ii)  $\text{pa}(X, \mathcal{H}) \subseteq \text{pa}(Y, \mathcal{H}) \setminus \{X\}$ .*

**Proof** [Proof of Lemma 54] We first prove the correctness of statement (i). For any directed edge  $X \rightarrow Y$  in a B-component  $\mathcal{C}^b$  of  $\mathcal{H}$ , there is a maximal clique  $\mathbf{M}_i$  of  $\mathcal{C}^b$  containing both  $X$  and  $Y$ . If  $\text{adj}(Y, \mathcal{C}^b) \setminus \{X\} \not\subseteq \text{adj}(X, \mathcal{C}^b)$ , then there exists a vertex  $Z$  satisfying  $Z \in \text{adj}(Y, \mathcal{C}^b) \setminus \{X\}$  but  $Z \notin \text{adj}(X, \mathcal{C}^b)$ , implying that  $Y$  and  $Z$  also belong to another maximal clique  $\mathbf{M}_j$  of  $\mathcal{C}^b$  which does not contain  $X$ . Hence, we have that  $X \in \mathbf{M}_i \setminus \mathbf{M}_{ij}$  and  $Y \in \mathbf{M}_{ij}$ , which contradicts Lemma 53.

We next show that statement (ii) holds. When  $\text{pa}(X, \mathcal{H}) = \emptyset$ , the result is trivial, so we assume that  $\text{pa}(X, \mathcal{H}) \neq \emptyset$ . For any vertex  $Z \in \text{pa}(X, \mathcal{H})$ , if  $Z \notin \mathcal{C}^b$ , then  $Z \rightarrow Y$  by Lemma 49. If  $Z \in \mathcal{C}^b$ , then we have  $Z \in \text{adj}(Y, \mathcal{H})$ , since otherwise  $Z \rightarrow X \rightarrow Y$  implies that  $X \rightarrow Y$  can be oriented by the first Meek's rule, contradicted to Lemma 52. Therefore,  $Z \rightarrow Y$  appears in  $\mathcal{H}$  by applying Rule 2 of Meek's rules. ■

Finally, we present the proof of Theorem 8.

**Proof** [Proof of Theorem 8] The necessity of conditions (i)-(iv) follows from Lemma 50, Lemma 48, Lemma 49, and Lemma 54, respectively. In the following, we will prove the sufficiency of conditions (i) to (iv). Let  $\mathcal{H} = (\mathbf{V}, \mathbf{E})$  be a partially directed graph which satisfies conditions (i)-(iv). Our goal is to show that  $\mathcal{H}$  is an MPDAG and  $\mathcal{H}$  is causal.

We first show that  $\mathcal{H}$  is acyclic. That is, there is no directed cycle in  $\mathcal{H}$ . Assume, for the sake of contradiction, that there exist directed cycles in  $\mathcal{H}$  and let  $\rho = (X_1, X_2, \dots, X_n, X_1)$  be the shortest one. By condition (i), all the vertices on  $\rho$  are in the same B-component, since otherwise the corresponding cycle of  $\rho$  in  $\mathcal{H}_c$  is a partially directed cycle, which contradicts condition (i). If  $n = 3$ , then  $X_3 \in \text{pa}(X_1, \mathcal{H})$  while  $X_3 \notin \text{pa}(X_2, \mathcal{H})$ , meaning that  $\text{pa}(X_1, \mathcal{H}) \not\subseteq \text{pa}(X_2, \mathcal{H}) \setminus \{X_1\}$ , which contradicts condition (iv). If  $n > 3$ , however, condition (iv) implies that  $X_n \rightarrow X_2$ , and thus  $X_n \rightarrow X_2 \rightarrow X_3 \rightarrow \dots \rightarrow X_n$  is a directed cycle of length  $n - 1$ , contrary to the assumption that  $\rho$  is the shortest. Therefore,  $\mathcal{H}$  is acyclic.

To prove that  $\mathcal{H}$  is an MPDAG, it suffices to show that  $\mathcal{H}$  is closed under Meek's rules. We will consider each rule separately in below.

- (i) If  $\mathcal{H}$  is not closed under Rule 1 of Meek's rules, then  $\mathcal{H}$  has an induced subgraph  $X \rightarrow Y - Z$ , in which  $X \notin \text{adj}(Z, \mathcal{H})$ . Since  $Y$  and  $Z$  are connected by an undirected edge, there exists a B-component  $\mathcal{C}^b$  of  $\mathcal{H}$  such that  $Y, Z \in \mathcal{C}^b$ . According to condition (iii),  $X \in \mathcal{C}^b$  since otherwise  $X \rightarrow Z$  should be in  $\mathcal{H}$ . However, if  $X \in \mathcal{C}^b$ , then  $Z \in \text{adj}(Y, \mathcal{C}^b)$  but  $Z \notin \text{adj}(X, \mathcal{C}^b)$ , which contradicts condition (iv). Thus,  $\mathcal{H}$  is closed under the first Meek's rule.
- (ii) If  $\mathcal{H}$  is not closed under Rule 2 of Meek's rules, then  $\mathcal{H}$  has an induced subgraph consists of  $X \rightarrow Y \rightarrow Z$  as well as  $X - Z$ . By condition (iii),  $X, Y, Z$  are in the same B-component. However, for the directed edge  $Y \rightarrow Z$ ,  $X \in \text{pa}(Y, \mathcal{H})$  but  $X \notin \text{pa}(Z, \mathcal{H}) \setminus \{Y\}$ , which contradicts condition (iv). Thus,  $\mathcal{H}$  is closed under the second Meek's rule.
- (iii) If  $\mathcal{H}$  is not closed under Rule 3 of Meek's rules, then  $\mathcal{H}$  has an induced subgraph shown in Figure 16a. The vertices  $X, Y, Z_1, Z_2$  are in the same B-component  $\mathcal{C}^b$ . However, this is impossible as  $Y \in \text{adj}(Z_2, \mathcal{C}^b) \setminus \{Z_1\}$  but  $Y \notin \text{adj}(Z_1, \mathcal{C}^b)$ .
- (iv) If  $\mathcal{H}$  is not closed under Rule 4 of Meek's rules, then  $\mathcal{H}$  has an induced subgraph shown in one of Figures 16b-16d. By the similar argument for (ii), no matter which induced subgraph  $\mathcal{H}$  has, the vertices  $X, Y, Z_1, Z_2$  are in the same B-component  $\mathcal{C}^b$ . However, for the directed edge  $Z_1 \rightarrow Z_2$ , we have  $Y \in \text{adj}(Z_2, \mathcal{C}^b) \setminus \{Z_1\}$  but  $Y \notin \text{adj}(Z_1, \mathcal{C}^b)$ , contrary to condition (iv). Hence,  $\mathcal{H}$  is closed under the fourth Meek's rule.

So far, we have proved that  $\mathcal{H}$  is an MPDAG. It remains to show that  $\mathcal{H}$  is causal. By the definition of a causal MPDAG, it suffices to show that there is a DAG that has the same skeleton and the same v-structures as  $\mathcal{H}$ , or equivalently, there exists an orientation of all undirected edges in  $\mathcal{H}$  that does not create a new v-structure or a directed cycle.

We first claim that (1) no orientation of the undirected edges in  $\mathcal{H}$  will create a directed cycle which includes a directed edge or edges in  $\mathcal{H}_c$  and (2) no orientation of an undirected edge in  $\mathcal{H}$  can create a new v-structure with an edge that was oriented in  $\mathcal{H}_c$ . In fact, the first claim holds because of condition (i), and the second claim holds because of condition (iii).

Based on the above two claims, to prove that  $\mathcal{H}$  is causal, it suffices to show that the skeleton of each B-component of  $\mathcal{H}$  has a perfect elimination ordering whose corresponding DAG contains the existing directed edges in that B-component.

Let  $\mathcal{C}^b$  be a B-component of  $\mathcal{H}$ . If  $\mathcal{C}^b$  is a complete graph with  $n$  vertices, then every vertex is simplicial in  $\mathcal{C}^b$ . We claim that, there exists a vertex  $V_1$  in  $\mathcal{C}^b$  which has no child in  $\mathcal{C}^b$ . In fact, if every vertex in  $\mathcal{C}^b$  has a child in  $\mathcal{C}^b$ , then there will be a directed cycle in  $\mathcal{C}^b$ , which is impossible. Note that, the induced subgraph of  $\mathcal{C}^b$  over  $\mathbf{V}(\mathcal{C}^b) \setminus \{V_1\}$  is still complete. Hence, repeat the above procedure we can find a sequence of vertices  $V_1, V_2, \dots, V_n$ . It can be easily verified that the ordering of the vertices forms a PEO of  $\mathcal{C}^b$ , and the corresponding DAG contains all directed edges in  $\mathcal{C}^b$ .

We then consider the case where  $\mathcal{C}^b$  is not a complete graph. Assume that every simplicial vertex has a child in  $\mathcal{C}^b$ . Let  $X$  be a simplicial vertex and  $\mathbf{M}_i$  be the (unique) maximal clique that contains  $X$  (Lemma 47). Denote by  $\mathbf{S}$  the set of simplicial vertices of  $\mathcal{C}^b$  contained in  $\mathbf{M}_i$ . Since  $\mathbf{M}_i$  induces a complete subgraph of  $\mathcal{C}^b$ ,  $\mathbf{S} \subseteq \mathbf{M}_i$  also induces a complete subgraph of  $\mathcal{C}^b$ . If every simplicial vertex in  $\mathbf{S}$  has a child which is also in  $\mathbf{S}$ , then we can

construct a directed cycle. Thus, there must be a simplicial vertex in  $\mathbf{S}$  whose child is not in  $\mathbf{S}$ . Without loss of generality, we can assume that such a vertex is  $X$ . Notice that,  $\mathcal{C}^b$  is an incomplete connected chordal graph,  $\mathbf{M}_i \setminus \mathbf{S} \neq \emptyset$ , and thus  $X \rightarrow Y$  is in  $\mathcal{C}^b$  for some  $Y \in \mathbf{M}_i \setminus \mathbf{S}$ . As  $\mathbf{S}$  consists of all simplicial vertices contained in  $\mathbf{M}_i$ ,  $Y$  is not simplicial, and thus there is a  $Z \in \text{adj}(Y, \mathcal{C}^b)$  that is not adjacent to  $X$  (Lemma 47). This means that  $\text{adj}(Y, \mathcal{C}^b) \setminus \{X\} \not\subseteq \text{adj}(X, \mathcal{C}^b)$ , which violates condition (iv).

Therefore, we can find a simplicial vertex, denoted by  $V_1$ , that does not have any child in  $\mathcal{C}^b$ . Since the induced subgraph of  $\mathcal{C}^b$  over  $\mathbf{V}(\mathcal{C}^b) \setminus \{V_1\}$  is still chordal, we can repeat the above procedure and find a sequence vertices  $V_1, V_2, \dots, V_n$ . Again, it can be checked that the ordering of the vertices forms a PEO of  $\mathcal{C}^b$ , and the corresponding DAG contains all directed edges in  $\mathcal{C}^b$ . This completes the proof.  $\blacksquare$

#### B.4 Proof of Proposition 12

**Proof** To prove the sufficiency, we first show that  $\mathcal{A}$  is a generator of  $\mathcal{H}$ , and then prove its minimality.

Denote by  $\mathcal{G}^*$  the CPDAG with the same skeleton and v-structures as  $\mathcal{H}$ . To prove that  $\mathcal{A}$  is a generator of  $\mathcal{H}$ , or equivalently, to prove that  $\mathcal{H}$  is the MPDAG of  $[\mathcal{G}^*, \mathcal{A}]$ , by Corollary 38 it suffices to show that for every directed edge  $X \rightarrow Y$  in  $\mathbf{E}_d(\mathcal{H}) \setminus (\mathbf{E}_d(\mathcal{G}^*) \cup \mathcal{A})$ , an orientation component for  $Y$  with respect to  $\mathcal{A}$  and  $\mathcal{G}^*$  contains  $X$ .

By construction of  $\mathcal{A}$ , the edges in  $\mathbf{E}_d(\mathcal{H}) \setminus (\mathbf{E}_d(\mathcal{G}^*) \cup \mathcal{A})$  are all M-strongly protected. On the other hand, as  $\mathcal{H}$  can be viewed as the MPDAG of  $[\mathcal{G}^*, \mathbf{E}_d(\mathcal{H})]$ , the maximal orientation component for  $Y$  with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ , denoted by  $\mathcal{U}$ , contains  $X$ . In the following, we will show that  $\mathcal{U}$  is an orientation component for  $Y$  with respect to  $\mathcal{A}$  and  $\mathcal{G}^*$ .

Suppose that, with respect to  $\mathcal{A}$  and  $\mathcal{G}^*$ ,  $\mathcal{U}$  is not an orientation component for  $Y$ , then by the definition of an orientation component, there exists another potential leaf node  $Z$  in  $\mathcal{U}$  with respect to  $\mathcal{A}$  and  $\mathcal{G}^*$ . Such a variable  $Z$  must have the following properties.

- (P1)  $Z$  is simplicial in  $\mathcal{U}$ .
- (P2) If an undirected edge connected to  $Z$  in  $\mathcal{U}$  is directed in  $\mathcal{A}$ ,  $Z$  is not the tail of that directed edge.
- (P3) For every sibling  $W$  of  $Z$  in  $\mathcal{U}$  such that  $Z \rightarrow W$  is in  $\mathcal{H}$  (such a  $W$  definitely exists),  $Z \rightarrow W$  is not in  $\mathcal{A}$ . (Note that, since every directed edge in  $\mathcal{A}$  is also in  $\mathcal{H}$ ,  $Z \rightarrow W$  is in  $\mathcal{H}$  implies that  $W \rightarrow Z$  is not in  $\mathcal{A}$  either).

The third property comes from that fact that, with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ ,  $\mathcal{U}$  is the maximal orientation component for  $Y$ .

Since  $Z \rightarrow W$  is not in  $\mathcal{A}$ ,  $Z \rightarrow W$  is M-strongly protected in  $\mathcal{H}$ . Note that,  $Z$  and  $W$  are adjacent in  $\mathcal{U}$ , they are in the same chain component of  $\mathcal{G}^*$ . We first prove that  $Z \rightarrow W$  cannot occur in the configurations (b) and (d), and if it occurs in one of the configurations (a), (c) and (e), the involved vertices in those configurations are all in the same chain component of  $\mathcal{G}^*$ .

- (i) If  $Z \rightarrow W$  occurs in the configuration (a), then there exists a vertex  $W_2 \notin \text{adj}(W, \mathcal{H})$  such that  $W_2 \rightarrow Z \rightarrow W$  is in  $\mathcal{H}$ . If  $W_2$  is not in the same chain component as  $Z$  and  $W$ , then  $W_2 \rightarrow Z$  implies that  $W_2 \rightarrow W$  is also in  $\mathcal{H}$ , which contradicts the configuration (a).
- (ii) If  $Z \rightarrow W$  occurs in the configuration (b), then there is a vertex  $W_2 \notin \text{adj}(Z, \mathcal{H})$  such that  $Z \rightarrow W \leftarrow W_2$ , which is a v-structure collided on  $W$ , is in  $\mathcal{H}$ . This means that  $W_2$  is not in the same chain component as  $Z$ , and thus,  $W_2 \rightarrow Z$  should be in  $\mathcal{H}$ , leading to a contradiction.
- (iii) If  $Z \rightarrow W$  occurs in the configuration (c), then there is a vertex  $W_2$  such that  $Z \rightarrow W_2 \rightarrow W$  is in  $\mathcal{H}$ . If  $W_2$  is not in the same chain component as  $Z$  and  $W$ , then  $W_2 \rightarrow W$  implies that  $W_2 \rightarrow Z$  is also in  $\mathcal{H}$ , which contradicts the configuration (c).
- (iv) If  $Z \rightarrow W$  occurs in the configuration (d), then there are vertices  $W_2, W_3 \in \text{sib}(Z, \mathcal{H})$  such that  $W_2$  is not adjacent to  $W_3$  in  $\mathcal{H}$  and  $W_2 \rightarrow W \leftarrow W_3$  is in  $\mathcal{H}$ . It is easy to see that  $W_2, W_3$  and  $\{Z, W\}$  are in three different chain components, and thus,  $W_2 \rightarrow Z \leftarrow W_3$  should be in  $\mathcal{H}$ . This contradicts the configuration (d).
- (v) If  $Z \rightarrow W$  occurs in the configuration (e), then there are vertices  $W_2, W_3 \in \text{sib}(Z, \mathcal{H})$  such that  $W_2 \rightarrow W_3 \rightarrow W$  and  $W_2$  is not adjacent to  $W$  in  $\mathcal{H}$ . If  $W_3$  is not in the same chain component as  $Z$  and  $W$ , then  $W_3 \rightarrow Z$  should be in  $\mathcal{H}$ , contradicted to the configuration. If  $W_2$  not in the same chain component as  $Z, W$  and  $W_3$ , then  $W_2 \rightarrow Z$  should also be in  $\mathcal{H}$ , contradicted to the configuration.

Below, we will consider the configurations (a), (c) and (e) separately and show that although  $Z \rightarrow W$  is M-strongly protected in  $\mathcal{H}$ ,  $Z \rightarrow W$  cannot occur in any of these configurations.

- (i) If  $Z \rightarrow W$  occurs in the configuration (a), then there exists a vertex  $W_2 \notin \text{adj}(W, \mathcal{H})$  such that  $W_2, Z$  and  $W$  are in the same chain component and  $W_2 \rightarrow Z$  is in  $\mathcal{H}$ . If  $W_2 \in \mathbf{V}(\mathcal{U})$ , then  $Z$  is not simplicial in  $\mathcal{U}$ , which contradicts (P1). Thus,  $W_2 \notin \mathbf{V}(\mathcal{U})$ . We claim that the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}) \cup \{W_2\}$ , denoted by  $\mathcal{U}'$ , is also an orientation component for  $Y$  with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ . In fact, with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ , since none of the vertices except for  $Y$  is a potential leaf node in  $\mathcal{U}$ , the vertices in  $\mathbf{V}(\mathcal{U}) \setminus \{Y\}$  are definitely not potential leaf nodes in  $\mathcal{U}'$  either. On the other hand,  $W_2$  is not a potential leaf node in  $\mathcal{U}'$ , as  $W_2 \rightarrow Z$  is in  $\mathbf{E}_d(\mathcal{H})$ , the consistency of  $\mathbf{E}_d(\mathcal{H})$  with  $\mathcal{G}^*$  implies that the only potential leaf node in  $\mathcal{U}'$  must be  $Y$  (Theorem 25). Therefore,  $\mathcal{U}'$  is also an orientation component for  $Y$  with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ . This contradicts the maximality of  $\mathcal{U}$ .
- (ii) If  $Z \rightarrow W$  occurs in the configuration (e), then there are vertices  $W_2, W_3 \in \text{sib}(Z, \mathcal{H})$  such that  $W_2 \rightarrow W_3 \rightarrow W$  and  $W_2$  is not adjacent to  $W$  in  $\mathcal{H}$ . If both  $W_2$  and  $W_3$  are in  $\mathbf{V}(\mathcal{U})$ , then  $Z$  is not a simplicial vertex in  $\mathcal{U}$ , which is again contradicted to (P1). On the other hand, if either  $W_2$  or  $W_3$  is not in  $\mathbf{V}(\mathcal{U})$ , then by the similar argument given in (i), the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}) \cup \{W_2, W_3\}$  is also an orientation component for  $Y$  with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ . This is again contradicted to the maximality of  $\mathcal{U}$ .

- (iii) If  $Z \rightarrow W$  occurs in the configuration (c), then there is a vertex  $W_2$  such that  $Z \rightarrow W_2 \rightarrow W$  is in  $\mathcal{H}$ . If  $W_2 \notin \mathbf{V}(\mathcal{U})$ , then by the similar argument given in (i), the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}) \cup \{W_2\}$  is also an orientation component for  $Y$  with respect to  $\mathbf{E}_d(\mathcal{H})$  and  $\mathcal{G}^*$ , contradicted to the maximality of  $\mathcal{U}$ . Hence,  $W_2 \in \mathbf{V}(\mathcal{U})$ , and by (P3),  $Z \rightarrow W_2$  is also M-strongly protected in  $\mathcal{H}$ . By the same argument given in (i) and (ii),  $Z \rightarrow W_2$  can only occur in the configuration (c). Repeat the above procedure we can find a sequence of vertices  $W_2, W_3, \dots, W_n, \dots$  in  $\text{adj}(Z, \mathcal{U})$ , such that  $Z \rightarrow W_i$  for  $i = 2, 3, \dots$  and  $W \leftarrow W_2 \leftarrow W_3 \leftarrow \dots \leftarrow W_n \dots \leftarrow$ . Note that, since the above procedure never ends, but there are only a finite number of vertices in  $\text{adj}(Z, \mathcal{U})$ , a subpath of  $W \leftarrow W_2 \leftarrow W_3 \leftarrow \dots$  must be a directed cycle, leading to a contradiction.

Therefore, although  $Z \rightarrow W$  is M-strongly protected in  $\mathcal{H}$ ,  $Z \rightarrow W$  cannot occur in any of the configurations (a) to (e). This is impossible. Thus,  $\mathcal{U}$  is an orientation component for  $Y$  with respect to  $\mathcal{A}$  and  $\mathcal{G}^*$ . Consequently,  $\mathcal{A}$  is a generator of  $\mathcal{H}$ .

We then show that  $\mathcal{A}$  is minimal. If there is another generator  $\mathcal{A}^-$  of  $\mathcal{H}$  such that  $|\mathcal{A}^-| < |\mathcal{A}|$ , then there is a directed edge  $X \rightarrow Y$  in  $\mathcal{A}$  which is not in  $\mathcal{A}^-$ . Since  $\mathcal{A}^-$  is a generator of  $\mathcal{H}$  and  $X \rightarrow Y$  is in  $\mathcal{H}$ ,  $X \rightarrow Y$  is either in a v-structure of the form  $X \rightarrow Y \leftarrow Z$ , or can be oriented by Meek's rules. This means that  $X \rightarrow Y$  occurs in at least one configurations labeled by (a) to (e) as an induced subgraph of  $\mathcal{H}$ . Hence,  $X \rightarrow Y$  is M-strongly protected in  $\mathcal{H}$ , which is contrary to the assumption that  $X \rightarrow Y$  is in  $\mathcal{A}$ . This completes the proof of sufficiency.

Next, we prove the necessity. Let  $\mathcal{A}$  be a minimal generator of  $\mathcal{H}$ . We first show that  $\mathcal{A}$  contains all directed edges in  $\mathcal{H}$  that are not M-strongly protected. Suppose that there is a directed edge  $X \rightarrow Y \notin \mathcal{A}$  which is in  $\mathcal{H}$  but not M-strongly protected. Then, since  $\mathcal{A}$  generates  $\mathcal{H}$ ,  $X \rightarrow Y$  is either in a v-structure of the form  $X \rightarrow Y \leftarrow Z$ , or can be oriented by Meek's rules. This means that  $X \rightarrow Y$  occurs in at least one configurations labeled by (a) to (e) as an induced subgraph of  $\mathcal{H}$ . That is,  $X \rightarrow Y$  is M-strongly protected. This contradicts our assumption. We next show that  $\mathcal{A}$  only contains the directed edges in  $\mathcal{H}$  that are not M-strongly protected. In fact, if this is not the case, then the proper subset of  $\mathcal{A}$  which consisting of the directed edges that are not M-strongly protected in  $\mathcal{H}$  is a generator of  $\mathcal{H}$ , meaning that  $\mathcal{A}$  is not minimal, which is a contradiction.

Finally, the uniqueness follows from the fact that the set of directed edges in  $\mathcal{H}$  that are not M-strongly protected in  $\mathcal{H}$  is unique. ■

### B.5 Proof of Theorem 13

**Proof** The conclusion follows directly from Propositions 9 and 12. ■

### B.6 Proof of Proposition 15

**Proof** Both claims hold by the definition of a DCC, and thus we omit the proof. ■

### B.7 Proof of Theorem 17

**Proof** The first statement is clearly true and the third statement can be derived from Fang and He (2020, Lemma 2). Since the second statement is the inverse of the third statement, the proof is completed. ■

### B.8 Proof of Proposition 20

**Proof** The proof follows directly from the definition of a DCC. ■

### B.9 Proof of Proposition 22

**Proof** The proof follows from Proposition 20. ■

### B.10 Proof of Theorem 25

We first prove two lemmas.

**Lemma 55** *Let  $\mathcal{G}^*$  be a CPDAG,  $\mathcal{K}$  be a set of consistent DCCs, and  $\mathcal{U}$  be a connected undirected induced subgraph of  $\mathcal{G}^*$ . For any DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$ , every leaf node in the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{U})$  is a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .*

**Proof** Denote by  $\mathcal{G}_{\text{sub}}$  the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{U})$ . By the definition of an undirected induced subgraph and the fact that  $\mathcal{G}^*$  has no partially directed cycle (Andersson et al., 1997),  $\mathcal{U}$  is the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U})$ . Thus, two vertices are adjacent in  $\mathcal{U}$  if and only if they are adjacent in  $\mathcal{G}_{\text{sub}}$ . Let  $V_{\text{leaf}}$  be a leaf node in  $\mathcal{G}_{\text{sub}}$ . By the definition of a potential leaf node, the conclusion holds if  $\text{adj}(V_{\text{leaf}}, \mathcal{U}) = \emptyset$ . Therefore, to prove the lemma it suffices to show that, (1)  $\text{adj}(V_{\text{leaf}}, \mathcal{U})$  induces a complete subgraph of  $\mathcal{U}$ , and (2)  $V_{\text{leaf}}$  is not the tail of any DCC in  $\mathcal{K}(\mathcal{U})$ .

As  $V_{\text{leaf}}$  is a leaf node in  $\mathcal{G}_{\text{sub}}$ ,  $\text{adj}(V_{\text{leaf}}, \mathcal{G}_{\text{sub}}) \rightarrow V_{\text{leaf}}$  are in  $\mathcal{G}_{\text{sub}}$ . Since  $\text{adj}(V_{\text{leaf}}, \mathcal{G}_{\text{sub}}) \subseteq \text{sib}(V_{\text{leaf}}, \mathcal{G}^*)$ , the configuration  $\text{adj}(V_{\text{leaf}}, \mathcal{G}_{\text{sub}}) \rightarrow V_{\text{leaf}}$  contains no v-structure collided on  $V_{\text{leaf}}$ . Thus,  $\text{adj}(V_{\text{leaf}}, \mathcal{G}_{\text{sub}})$  induces a complete subgraph of  $\mathcal{G}_{\text{sub}}$ , meaning that  $\text{adj}(V_{\text{leaf}}, \mathcal{U})$  induces a complete subgraph of  $\mathcal{U}$ . This completes the proof of statement (1). On the other hand, if there is a  $V_{\text{leaf}} \xrightarrow{\text{or}} \kappa_h$  in  $\mathcal{K}(\mathcal{U})$ , then by Equation (4),  $\kappa_h \subseteq \text{adj}(V_{\text{leaf}}, \mathcal{U}) = \text{adj}(V_{\text{leaf}}, \mathcal{G}_{\text{sub}})$  and  $V_{\text{leaf}} \xrightarrow{\text{or}} \kappa_h$  is in  $\mathcal{K}(\mathcal{G}^*)$ . Notice that  $\mathcal{K}$  is equivalent to  $\mathcal{K}(\mathcal{G}^*)$ ,  $V_{\text{leaf}} \xrightarrow{\text{or}} \kappa_h$  must hold for  $\mathcal{G}$ . Consequently,  $V_{\text{leaf}} \xrightarrow{\text{or}} \kappa_h$  holds for  $\mathcal{G}_{\text{sub}}$ . However,  $V_{\text{leaf}}$  is a leaf node in  $\mathcal{G}_{\text{sub}}$ , which leads to a contradiction. ■

Lemma 55 suggests that if a vertex is not a potential leaf node in some connected undirected induced subgraph  $\mathcal{U}$ , then it cannot be a leaf node in the induced subgraph over  $\mathbf{V}(\mathcal{U})$  of any restricted Markov equivalent DAG.

**Lemma 56** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs. Then,  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$  if and only if  $\mathcal{K}(\mathcal{C})$  is consistent with  $\mathcal{C}$  for any chain component  $\mathcal{C}$ .*

**Proof** Suppose that  $[\mathcal{G}^*, \mathcal{K}(\mathcal{G}_u^*)] \neq \emptyset$  and let  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}(\mathcal{G}_u^*)]$ . For any chain component  $\mathcal{C}$ , let  $\mathcal{G}_{\text{sub}}$  denote the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{C})$ . It is easy to verify that  $\mathcal{G}_{\text{sub}} \in [\mathcal{C}, \mathcal{K}(\mathcal{C})]$ . Conversely, if  $[\mathcal{C}, \mathcal{K}(\mathcal{C})] \neq \emptyset$  for any chain component  $\mathcal{C}$ , then choose  $\mathcal{G}_{\mathcal{C}} \in [\mathcal{C}, \mathcal{K}(\mathcal{C})]$  arbitrarily for each chain component and orient undirected edges in  $\mathcal{G}^*$  according to  $\{\mathcal{G}_{\mathcal{C}}\}$ . That is, orient  $X - Y$  in  $\mathcal{G}^*$  as  $X \rightarrow Y$  if  $X$  is a parent of  $Y$  in the DAG  $\mathcal{G}_{\mathcal{C}}$ , where  $\mathcal{C}$  is the chain component containing  $X$  and  $Y$ . Notice that  $\mathcal{G}_u^*$  is a union of (disjoint) chain components, we have  $\mathcal{K}(\mathcal{G}_u^*) = \bigcup_{\mathcal{C}} \mathcal{K}(\mathcal{C})$ , where  $\mathcal{C}$  is a chain component. It is straightforward to show that the resulting DAG with the orientations defined above is in  $[\mathcal{G}^*, \mathcal{K}(\mathcal{G}_u^*)]$ .  $\blacksquare$

Finally, we present the proof of Theorem 25.

**Proof** [Proof of Theorem 25] We first prove the necessity. If  $\mathcal{K}$  is consistent, then there is a DAG  $\mathcal{G}$  in  $[\mathcal{G}^*, \mathcal{K}]$ . Let  $\mathcal{U}$  be an arbitrary connected undirected induced subgraph of  $\mathcal{G}^*$ , and denote the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{U})$  by  $\mathcal{G}_{\text{sub}}$ . Since any induced subgraph of a DAG is still a DAG,  $\mathcal{G}_{\text{sub}}$  is a DAG, and thus it must have a leaf node  $V_{\text{leaf}}$ . By Lemma 55, we can conclude that  $V_{\text{leaf}}$  is a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ .

We next prove the sufficiency. By Lemma 56,  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$  if and only if  $\mathcal{K}(\mathcal{C})$  is consistent with  $\mathcal{C}$  for any chain component  $\mathcal{C}$ . Therefore, given a chain component  $\mathcal{C}$ , our goal is to prove that  $\mathcal{K}(\mathcal{C})$  is consistent with  $\mathcal{C}$ , providing that the potential-leaf-node condition holds. Based on the analysis in Appendix B.1, we need to construct a PEO of  $\mathcal{C}$  such that the corresponding DAG satisfies all DCCs in  $\mathcal{K}(\mathcal{C})$ . By assumption,  $\mathcal{C}$  has a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , denoted by  $V_1$ . By the definition of a potential leaf node,  $V_1$  is simplicial in  $\mathcal{C}$ . Next, consider the induced subgraph of  $\mathcal{C}$  over  $\mathbf{V}(\mathcal{C}) \setminus \{V_1\}$ , denoted by  $\mathcal{C}_2$ .  $\mathcal{C}_2$  is clearly connected. Hence, by assumption  $\mathcal{C}_2$  has a potential leaf node, denoted by  $V_2$ . Following the above procedure, we have a sequence of undirected graphs  $(\mathcal{C} = \mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m)$  and a sequence of vertices  $(V_1, V_2, \dots, V_m)$ , where  $m = |\mathbf{V}(\mathcal{C})|$ . By the construction, the ordering of the vertices in this sequence forms a PEO of  $\mathcal{C}$ . Denote by  $\mathcal{G}_{\mathcal{C}}$  the corresponding DAG of this PEO. If there is a  $V_i \xrightarrow{\text{or}} \kappa_h \in \mathcal{K}(\mathcal{C})$  which does not hold for  $\mathcal{G}_{\mathcal{C}}$ , then  $\kappa_h \rightarrow V_i$  are in  $\mathcal{G}_{\mathcal{C}}$ . This means  $\kappa_h \subseteq \{V_{i+1}, V_{i+2}, \dots, V_m\}$ , which contradicts the construction of the vertex sequence as  $V_i$  is definitely not a potential leaf node in the induced subgraph  $\mathcal{C}_i$ . Therefore,  $\mathcal{G}_{\mathcal{C}} \in [\mathcal{C}, \mathcal{K}(\mathcal{C})]$ . This completes the proof of Theorem 25.  $\blacksquare$

### B.11 Proof of Theorem 27

**Proof** The equivalence of statements (i) and (ii) follows from the definition of equivalence and redundancy. Observed that when  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ ,

$$\begin{aligned}
 \kappa_t \xrightarrow{\text{or}} \kappa_h \text{ is redundant with respect to } [\mathcal{G}^*, \mathcal{K}] &\iff \forall \mathcal{G} \in [\mathcal{G}^*, \mathcal{K}], \kappa_t \xrightarrow{\text{or}} \kappa_h \text{ holds for } \mathcal{G} \\
 &\iff \forall \mathcal{G} \in [\mathcal{G}^*, \mathcal{K}], \text{ch}(\kappa_t, \mathcal{G}) \cap \kappa_h \neq \emptyset \\
 &\iff \forall \mathcal{G} \in [\mathcal{G}^*, \mathcal{K}], \kappa_h \not\subseteq \text{pa}(\kappa_t, \mathcal{G}) \\
 &\iff \{\kappa_h \rightarrow \kappa_t\} \cup \mathcal{K} \text{ is inconsistent with } \mathcal{G}^*.
 \end{aligned} \tag{8}$$

Therefore, statement (ii) is equivalent to statement (iii) if both  $\mathcal{K}_1$  and  $\mathcal{K}_2$  are consistent with  $\mathcal{G}^*$ . If neither  $\mathcal{K}_1$  nor  $\mathcal{K}_2$  is consistent with  $\mathcal{G}^*$ , then statements (i) and (iii) hold simultaneously. Finally, if  $\mathcal{K}_1$  is consistent with  $\mathcal{G}^*$  but  $\mathcal{K}_2$  is not, then statements (i) does not hold. Thus, we need only to show that statements (iii) does not hold either. In fact, by Equation (8), if  $\cup_{D \in \kappa_h} \{D \rightarrow \kappa_t\} \cup \mathcal{K}_1$  is not consistent with  $\mathcal{G}^*$  for every  $\kappa \in \mathcal{K}_2$ , then every  $\kappa \in \mathcal{K}_2$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}_1]$ . Consequently,  $\mathcal{K}_2 \cup \mathcal{K}_1$  is equivalent to  $\mathcal{K}_1$  given  $\mathcal{G}^*$  and thus is consistent. However, this is impossible, as the union of an inconsistent DCC and a consistent DCC is definitely inconsistent. ■

### B.12 Proof of Theorem 29

Some technical lemmas are required before we present the proof of Theorem 29.

**Lemma 57** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs consistent with  $\mathcal{G}^*$ , then a DCC  $\kappa$  holds for every DAG in  $[\mathcal{G}^*, \mathcal{K}]$  if and only if either  $\kappa_t \rightarrow s$  appears in  $\mathcal{G}^*$  for some  $s \in \kappa_h$  or there exists a connected undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$  containing the vertices in  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*)$  and  $\kappa_t$  such that with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , every potential leaf node in  $\mathcal{U}$  is in  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*)$ .*

**Proof** We first prove the sufficiency. It is clear that  $\kappa_t \rightarrow s$  appears in  $\mathcal{G}^*$  for some  $s \in \kappa_h$  implies that  $\kappa$  holds for every DAG in  $[\mathcal{G}^*, \mathcal{K}]$ . On the other hand, if there exists a connected undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$  containing the vertices in  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*)$  and  $\kappa_t$  such that with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , every potential leaf node in  $\mathcal{U}$  is in  $\kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*)$ , by Lemma 55, for every DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$ , there exists an  $s \in \kappa_h$  such that  $s$  is a leaf node in the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{U})$ . Therefore,  $\kappa_t \rightarrow s$  is in  $\mathcal{G}$ , indicating that  $\kappa$  holds for  $\mathcal{G}$ .

We next prove the necessity. By Theorem 27,  $\cup_{s \in \kappa_h} \{s \rightarrow \kappa_t\} \cup \mathcal{K}$  is not consistent with  $\mathcal{G}^*$ . Therefore, by Theorem 25, there is a connected undirected induced subgraph  $\mathcal{U}_0$  of  $\mathcal{G}^*$  which has no potential leaf node with respect to  $\cup_{s \in \kappa_h} \{s \rightarrow \kappa_t\} \cup \mathcal{K}$  and  $\mathcal{G}^*$ . If  $\kappa_t \rightarrow s$  appears in  $\mathcal{G}^*$  for some  $s \in \kappa_h$ , then the proof is completed. Thus, we assume that  $\kappa_t \nrightarrow s$  in  $\mathcal{G}^*$  for all  $s \in \kappa_h$ . Let  $P = \{s \in \kappa_h \mid s \rightarrow \kappa_t \text{ is in } \mathcal{G}^*\}$  and  $Q = \kappa_h \cap \text{sib}(\kappa_t, \mathcal{G}^*) = \{s \in \kappa_h \mid s - \kappa_t \text{ is in } \mathcal{G}^*\}$ . Since  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ ,  $Q \neq \emptyset$  and  $\mathcal{U}_0$  must contain  $\kappa_t$  and at least one  $s \in Q$ . Now consider the connected undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}_0) \cup Q$ . With respect to  $\cup_{s \in \kappa_h} \{s \rightarrow \kappa_t\} \cup \mathcal{K}$  and  $\mathcal{G}^*$ , none of the vertices in  $Q$  is a potential leaf node as  $s \rightarrow \kappa_t$  for  $s \in Q$ , and none of the vertices in  $\mathbf{V}(\mathcal{U}_0)$  is a potential leaf node as  $\mathcal{U}_0$  has no potential leaf node with respect to  $\cup_{s \in \kappa_h} \{s \rightarrow \kappa_t\} \cup \mathcal{K}$  and  $\mathcal{G}^*$ . As a result,  $\mathcal{U}$  has no potential leaf node with respect to  $\cup_{s \in \kappa_h} \{s \rightarrow \kappa_t\} \cup \mathcal{K}$  and  $\mathcal{G}^*$ .

However,  $\mathcal{U}$  has potential leaf nodes with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ . If a vertex  $p \in \mathbf{V}(\mathcal{U}_0) \setminus Q$  is a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , then  $p$  is also a potential leaf node with respect to  $\cup_{s \in \kappa_h} \{s \rightarrow \kappa_t\} \cup \mathcal{K}$  and  $\mathcal{G}^*$ . Therefore, the potential leaf nodes in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  are all in  $Q$ . ■

The above lemma generalizes Theorem 35 and Corollary 38. When  $\kappa$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ , we have the following corollary.



**Corollary 58** *Let  $\mathcal{G}^*$  be a CPDAG and  $\mathcal{K}$  be a set of DCCs consistent with  $\mathcal{G}^*$ , then a DCC  $\kappa$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  if and only if either  $\kappa_h = \{s\}$  is a singleton set and  $\kappa_t \rightarrow s$  appears in  $\mathcal{G}^*$ , or  $\kappa_h \subseteq \text{sib}(\kappa_t, \mathcal{G}^*)$  and there exists a connected undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$  containing the vertices in  $\kappa_h$  and  $\kappa_t$  such that with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ ,  $\kappa_h$  are all and only potential leaf nodes in  $\mathcal{U}$ .*

The following lemma is the key to prove Theorem 29.

**Lemma 59** *Suppose that  $\mathcal{G}^*$  is a CPDAG and  $\mathcal{K}, \mathcal{K}'$  are two non-redundant consistent DCC sets over  $\mathbf{V}(\mathcal{G}^*)$ , and  $\kappa'$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  for any  $\kappa' \in \mathcal{K}'$ .*

(i) *If  $\kappa \in \mathcal{K}$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa\})]$ , then there exists a unique DCC  $\kappa' \in \mathcal{K}'$  such that  $\kappa'_t = \kappa_t$  and  $\kappa'_h \subseteq \kappa_h$ . Moreover, for any DCC  $\gamma \in \mathcal{K}$  such that  $\gamma \neq \kappa$ , either  $\kappa'_t \neq \gamma_t$  or  $\kappa'_h \not\subseteq \gamma_h$ .*

(ii) *If  $[\mathcal{G}^*, \mathcal{K}] = [\mathcal{G}^*, \mathcal{K}']$ , then  $|\mathcal{K}| = |\mathcal{K}'|$ .*

**Proof** [Proof of Lemma 59] By assumption,  $\mathcal{K}' \cup \mathcal{K}$  and all of its subset are consistent with  $\mathcal{G}^*$ . We first prove statement (i). Without loss of generality, in the following, we assume that  $\mathcal{K} = \{\kappa_1, \dots, \kappa_n\}$ ,  $\mathcal{K}' = \{\kappa'_1, \dots, \kappa'_m\}$ , and  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})]$ . Since  $\kappa'_i$  is not redundant with respect to  $\mathcal{K}' \setminus \{\kappa'_i\}$ ,  $\kappa'_{ih} \cap \text{ch}(\kappa'_{it}, \mathcal{G}^*) = \emptyset$ . Moreover, since  $\kappa'_i$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ , by Corollary 58, for every  $\kappa'_i$ , there exists a connected undirected induced subgraph  $\mathcal{U}'_i$  of  $\mathcal{G}^*$  containing the vertices in  $\kappa'_{ih}$  and  $\kappa'_{it}$  such that with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ ,  $\kappa'_{ih}$  are all and only potential leaf nodes in  $\mathcal{U}'_i$ .

Assuming that the corresponding connected undirected induced subgraphs of  $\kappa'_1, \dots, \kappa'_k$ , which are denoted by  $\mathcal{U}'_1, \dots, \mathcal{U}'_k$ , contain all vertices in  $\kappa_{nt}$  and  $\kappa_{nh} \cap \text{sib}(\kappa_{nt}, \mathcal{G}^*)$ , while the remaining induced subgraphs  $\mathcal{U}'_{k+1}, \dots, \mathcal{U}'_m$  do not contain all vertices in  $\kappa_{nt}$  and  $\kappa_{nh}$ . Since  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})]$ , by Lemma 57 and the fact that  $\kappa_n$  is not redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ , there is a connected undirected induced subgraph  $\mathcal{U}_n$  of  $\mathcal{G}^*$  containing  $\kappa_{nt}$  and a subset  $\mathbf{s}_n$  of  $\kappa_{nh}$  such that all potential leaf nodes in  $\mathcal{U}_n$ , with respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ , are in  $\mathbf{s}_n$ . It is clear that, there is at least one DCC  $\kappa'^* \in \mathcal{K}'$  satisfying that  $\kappa'^*_t$  and all vertices in  $\kappa'^*_h \cap \text{sib}(\kappa'^*_t, \mathcal{G}^*)$  are in  $\mathcal{U}_n$  (this condition is denoted by C1). Since otherwise, the restriction subset of  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$  on  $\mathcal{U}_n$  is identical to that of  $\mathcal{K} \setminus \{\kappa_n\}$ , which violates the assumption that  $\mathcal{K}$  is not redundant.

Denote by  $\{\kappa'^*_1, \dots, \kappa'^*_x\}$  the set of all DCCs in  $\mathcal{K}'$  satisfying the above condition C1. We claim that  $\{\kappa'^*_1, \dots, \kappa'^*_x\} \cap \{\kappa'_1, \dots, \kappa'_k\} \neq \emptyset$ . In fact, if  $\{\kappa'^*_1, \dots, \kappa'^*_x\} \subseteq \{\kappa'_{k+1}, \dots, \kappa'_m\}$ , by the assumption that  $\mathcal{U}'_l$  for every  $l = k+1, \dots, m$  does not contain all vertices in  $\kappa_{nt}$  and  $\kappa_{nh} \cap \text{sib}(\kappa_{nt}, \mathcal{G}^*)$ , the restriction subset of  $\mathcal{K}$  on  $\mathcal{U}'_l$  is identical to that of  $\mathcal{K} \setminus \{\kappa_n\}$ , and hence  $\kappa'^*_1, \dots, \kappa'^*_x$  are redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ . On the other hand, the restriction subset of  $\mathcal{K}' \cup \mathcal{K} \setminus \{\kappa_n\}$  on  $\mathcal{U}_n$  is identical to that of  $\{\kappa'^*_1, \dots, \kappa'^*_x\} \cup (\mathcal{K} \setminus \{\kappa_n\})$ , implying that  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \{\kappa'^*_1, \dots, \kappa'^*_x\} \cup (\mathcal{K} \setminus \{\kappa_n\})]$ . As a result,  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ , which contradicts the non-redundancy of  $\mathcal{K}$ . This completes the proof of the claim.

Without loss of generality, we assume that  $\{\kappa'^*_1, \dots, \kappa'^*_x\} \cap \{\kappa'_1, \dots, \kappa'_k\} = \{\kappa'_1, \dots, \kappa'_d\}$ . We next show that there is at least one  $\kappa'_i \in \{\kappa'_1, \dots, \kappa'_d\}$  such that  $\kappa'_{ih} \cap \mathbf{s}_n \neq \emptyset$ . Consider

the undirected induced subgraph over  $\mathbf{V}(\mathcal{U}'_1), \dots, \mathbf{V}(\mathcal{U}'_d)$  and  $\mathbf{V}(\mathcal{U}_n)$ , which is also the union graph of  $\mathcal{U}'_1, \dots, \mathcal{U}'_d$  and  $\mathcal{U}_n$ . With respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ , if a vertex is not a potential leaf node in one of  $\mathcal{U}'_1, \dots, \mathcal{U}'_d$  and  $\mathcal{U}_n$ , then it is not a potential leaf node in the union graph either. If a vertex is a potential leaf node in  $\mathcal{U}'_i$  ( $i = 1, \dots, d$ ) with respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ , it must be in  $\kappa'_{ih} \cup \{\kappa_{nt}\}$ . However, if  $\kappa'_{ih} \cap \mathbf{s}_n = \emptyset$ , the vertices in  $\kappa'_{ih} \cup \{\kappa_{nt}\}$  are not potential leaf nodes in  $\mathcal{U}_n$  with respect to  $\mathcal{K}' \cup \mathcal{K} \setminus \{\kappa_n\}$ , and thus are not potential leaf nodes in the union graph. Similarly,  $\kappa'_{ih} \cap \mathbf{s}_n = \emptyset$  for any  $i = 1, \dots, d$  implies that the vertices in  $\mathbf{s}_n$  are not potential leaf nodes in any  $\mathcal{U}'_i$ , and thus are not potential leaf nodes in the union graph. As a result, no vertex in the union graph is a potential leaf node under the condition that  $\kappa'_{ih} \cap \mathbf{s}_n = \emptyset$  for any  $i = 1, \dots, d$ , which is impossible as  $\mathcal{K}' \cup \mathcal{K} \setminus \{\kappa_n\}$  is consistent.

Assuming that  $\kappa'_1, \dots, \kappa'_l$  ( $1 \leq l \leq d$ ) are all and only the DCCs such that  $\kappa'_{ih} \cap \mathbf{s}_n \neq \emptyset$  ( $i = 1, \dots, l$ ). We will show that

$$\mathbf{s} := \bigcap_{i=1}^l \kappa'_{ih} \cap \mathbf{s}_n \neq \emptyset.$$

Assume, for contradiction, that  $\bigcap_{i=1}^l \kappa'_{ih} \cap \mathbf{s}_n = \emptyset$ , and consider again the union graph of  $\mathcal{U}'_1, \dots, \mathcal{U}'_d$  and  $\mathcal{U}_n$ . Using the same argument given above, we can prove that with respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ , any potential leaf node in  $\mathcal{U}'_i$  ( $i = l+1, \dots, d$ ) is not a potential leaf node in the union graph. If a vertex  $h_i$  is a potential leaf node in  $\mathcal{U}'_i$  ( $i = 1, \dots, l$ ) but not in  $\mathbf{s}_n$ , it is not a potential leaf node in the union graph either. However, if  $h_i \in \mathbf{s}_n$ , by assumption, there exists a  $\kappa'_j$ ,  $j \neq i$  and  $j \in \{1, \dots, l\}$ , such that  $h_i \notin \kappa'_{jh}$ . Note that,  $h_i \in \mathbf{s}_n \subseteq \kappa_{nh}$ , meaning that  $h_i \neq \kappa_{nt}$ . Therefore, in  $\mathcal{U}'_j$ , with respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ ,  $h_i$  is not a potential leaf node as  $h_i \notin \kappa'_{jh} \cup \{\kappa_{nt}\}$ , and thus  $h_i$  is not a potential leaf node in the union graph. Similarly, any potential leaf node in  $\mathcal{U}_n$  is not a potential leaf node in at least one  $\mathcal{U}'_i$ ,  $i \in \{1, \dots, l\}$ , and thus is not a potential leaf node in the union graph. As a consequence, no vertex in the union graph is a potential leaf node, which leads to a contradiction.

Since  $\mathbf{s} \neq \emptyset$ , by the similar argument we can prove that every potential leaf node in the union graph of  $\mathcal{U}'_1, \dots, \mathcal{U}'_d$  and  $\mathcal{U}_n$  with respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$  is in  $\mathbf{s}$ . According to Lemma 57,  $\kappa'_{it} \xrightarrow{\text{or}} \mathbf{s}$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})]$  for any  $i = 1, \dots, l$ , and thus redundant with respect to  $[\mathcal{G}^*, \mathcal{K}' \cup \mathcal{K}]$ . Note that,  $\mathcal{K}' \cup \mathcal{K}$  is equivalent to  $\mathcal{K}$  due to the minimal redundancy of the DCCs in  $\mathcal{K}'$ ,  $\kappa'_{it} \xrightarrow{\text{or}} \mathbf{s}$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  for any  $i = 1, \dots, l$ . Therefore,  $\kappa'_{ih} = \mathbf{s}$  for any  $i = 1, \dots, l$ , since otherwise  $\mathbf{s} \subsetneq \kappa'_{ih}$  contradicts to the minimal redundancy of  $\kappa'_i$ .

With  $\kappa'_{ih} = \mathbf{s}$  for any  $i = 1, \dots, l$ , we claim that  $l = d$ . Suppose that  $l < d$ , then with respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ , any potential leaf node in  $\mathcal{U}'_i$  ( $i = l+1, \dots, d$ ) is not a potential leaf node in the union graph of  $\mathcal{U}'_1, \dots, \mathcal{U}'_d$  and  $\mathcal{U}_n$ . On the other hand, any potential leaf node in  $\mathcal{U}_n$  is not a potential leaf node in  $\mathcal{U}'_{l+1}, \dots, \mathcal{U}'_d$  as it is in  $\mathbf{s}_n$ . Since  $\kappa'_{ih} = \mathbf{s} \subseteq \mathbf{s}_n$  for  $i = 1, \dots, l$ , any potential leaf node in  $\mathcal{U}'_i$ ,  $i = 1, \dots, l$ , is not a potential leaf node in  $\mathcal{U}'_{l+1}, \dots, \mathcal{U}'_d$ , and  $\mathcal{U}_n$ . Therefore, no vertex in the union graph is a potential leaf node, which leads to a contradiction.

The next step is to prove that there exists a DCC  $\kappa' \in \{\kappa'_1, \dots, \kappa'_d\}$  such that  $\kappa'_t = \kappa_{nt}$  and  $\kappa'_h \subseteq \kappa_{nh}$ . Assume, for contradiction, that none of the DCC in  $\{\kappa'_1, \dots, \kappa'_d\}$  satisfies the condition. Since we proved that  $\kappa'_{ih} = \mathbf{s} \subseteq \kappa_{nh}$  for any  $i = 1, \dots, d$ , by assumption,

$\kappa'_{it} \neq \kappa_{nt}$  for any  $i = 1, \dots, d$ . Firstly, consider  $\mathcal{U}_n$ . With respect to  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$ , since  $\{\kappa'_{it}\}$  are not potential leaf nodes, it holds that  $\kappa'_{it} \notin \mathbf{s}_n$ . Moreover, the restriction subset of  $\mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa_n\})$  on  $\mathcal{U}_n$  is identical to that of  $\{\kappa'_1, \dots, \kappa'_d\} \cup (\mathcal{K} \setminus \{\kappa_n\})$ , which is further identical to that of  $\{\kappa'_1, \dots, \kappa'_d\} \cup (\mathcal{K} \setminus \{\kappa_n\})$ . The latter holds because  $\mathcal{U}'_i$  for  $\kappa'_i \in \{\kappa'_1, \dots, \kappa'_d\} \setminus \{\kappa'_1, \dots, \kappa'_d\}$  does not contain all vertices in  $\kappa_{nt}$  and  $\kappa_{nh} \cap \text{sib}(\kappa_{nt}, \mathcal{G}^*)$ , implying that  $\kappa'_i$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ . Consequently, with respect to  $\{\kappa'_1, \dots, \kappa'_d\} \cup (\mathcal{K} \setminus \{\kappa_n\})$ ,  $\mathbf{s}_n$  are all and only potential leaf nodes in  $\mathcal{U}_n$ . Since  $\kappa'_{it} \neq \kappa_{nt}$  for any  $i = 1, \dots, d$ , with respect to  $\mathcal{K} \setminus \{\kappa_n\}$ ,  $\kappa_{nt}$  is not a potential leaf node in  $\mathcal{U}_n$ . Next, consider  $\mathcal{U}'_i$  for  $i = 1, \dots, d$ . With respect to  $\mathcal{K} \setminus \{\kappa_n\}$ ,  $\{\kappa'_{it}\}$  are not potential leaf nodes. Therefore, in the union graph of  $\mathcal{U}'_1, \dots, \mathcal{U}'_d$  and  $\mathcal{U}_n$ , only the vertices in  $\mathbf{s}_n$  are potential leaf nodes with respect to  $\mathcal{K} \setminus \{\kappa_n\}$ . This implies that  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ , leading to a contradiction.

Finally, to prove the uniqueness in statement (i), it suffices to show that none of the DCC  $\kappa' \in \mathcal{K}' \setminus \{\kappa'_1, \dots, \kappa'_d\}$  satisfies  $\kappa'_t = \kappa_{nt}$  and  $\kappa'_h \subseteq \kappa_{nh} \cap \text{sib}(\kappa_{nt}, \mathcal{G}^*)$ . Assume, for contradiction, that such a DCC  $\kappa'$  exists. Since  $\mathcal{U}_n$  contains all vertices in  $\kappa_{nh} \cap \text{sib}(\kappa_{nt}, \mathcal{G}^*)$  and  $\kappa_{nt}$ ,  $\mathcal{U}_n$  contains  $\kappa'_h$  and  $\kappa'_t$ , meaning that  $\kappa' \notin \{\kappa'_{d+1}, \dots, \kappa'_k\}$ . However, if  $\kappa' \in \{\kappa'_{k+1}, \dots, \kappa'_m\}$ , since  $\mathcal{U}'_i$  does not contain all vertices in  $\kappa_{nt}$  and  $\kappa_{nh} \cap \text{sib}(\kappa_{nt}, \mathcal{G}^*)$ ,  $\kappa'$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ , and hence  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ , which leads to a contradiction.

To complete the proof of statement (i), we assume, for contradiction, that there is another DCC  $\kappa_i \in \mathcal{K}$ ,  $i \neq n$ , such that  $\kappa'_t = \kappa_{it}$  and  $\kappa'_h \subseteq \kappa_{ih}$ . It is clear that  $\mathcal{U}'$  contains all vertices in  $\kappa_{ih}$  and  $\kappa_{it}$ , since otherwise  $\kappa_i$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_i\}]$ . Now, consider  $\mathcal{U}'$  and  $\bigcup_{s \in \kappa_{nh}} \{s \rightarrow \kappa_{nt}\} \cup (\mathcal{K} \setminus \{\kappa_n\})$ . It is easy to verify that none of the vertices in  $\mathcal{U}'$  is a potential leaf node. Therefore,  $\kappa_n$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K} \setminus \{\kappa_n\}]$ , which is contradicted to the non-redundancy  $\kappa_n$ .

We then prove statement (ii). If  $[\mathcal{G}^*, \mathcal{K}] = [\mathcal{G}^*, \mathcal{K}']$ , then any  $\kappa \in \mathcal{K}$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}']$ , and thus redundant with respect to  $[\mathcal{G}^*, \mathcal{K}' \cup (\mathcal{K} \setminus \{\kappa\})]$ . By statement (i),  $n \leq m$ . If  $n < m$ , assume, without generality, that the unique DCC satisfies the condition in statement (i) for  $\kappa_i$  is  $\kappa'_i$ ,  $i = 1, 2, \dots, n$ , then  $\kappa_i$  is redundant with respect to  $[\mathcal{G}^*, \kappa'_i]$ . As a result,  $\kappa'_m$  is redundant with respect to  $[\mathcal{G}^*, \{\kappa'_1, \dots, \kappa'_n\}]$ , leading to a contradiction. Therefore,  $n = m$ . ■

**Proof** [Proof of Theorem 29] We first prove statement (i). For any DCC  $\kappa \in \mathcal{K}$ , let  $\kappa'$  be the DCC such that (1)  $\kappa'_t = \kappa_t$ , (2)  $\kappa'_h \subseteq \kappa_h$ , and (3)  $\kappa'$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ . Such a DCC always exists. In fact, one can first check whether  $\kappa_t \rightarrow s$  is redundant for some  $s \in \kappa_h$ . If exists, then  $\kappa_t \rightarrow s$  is minimally redundant, and if not exists, one can further enumerate and check all 2-elements subsets of  $\kappa_h$ . The enumeration will end because  $\kappa' = \kappa$  is redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$ .

Let  $\mathcal{K}'$  denote the set of all  $\kappa'$  constructed above. It is clear that  $\mathcal{K}'$  is equivalent to  $\mathcal{K}$  and  $|\mathcal{K}'| \leq |\mathcal{K}|$ . If  $\mathcal{K}'$  is redundant, then its non-redundant proper subset  $\mathcal{K}'' \subsetneq \mathcal{K}'$  is equivalent to  $\mathcal{K}$ . By statement (ii) of Lemma 59,  $|\mathcal{K}''| = |\mathcal{K}|$ . This is not possible as  $|\mathcal{K}''| < |\mathcal{K}'|$ . Therefore,  $\mathcal{K}'$  is not redundant.

Assume that there is another non-redundant DCC set  $\mathcal{K}'''$  such that  $\mathcal{K}$  is equivalent to  $\mathcal{K}'''$  and  $\kappa'''$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}]$  for any  $\kappa''' \in \mathcal{K}'''$ . Then,

$\mathcal{K}'''$  is also equivalent to  $\mathcal{K}'$  and  $\kappa'''$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}']$  for any  $\kappa''' \in \mathcal{K}'''$ . By the first statement of Lemma 59, for any  $\kappa' \in \mathcal{K}'$ , there is a unique DCC  $\kappa''' \in \mathcal{K}'''$  such that  $\kappa_t''' = \kappa_t'$  and  $\kappa_h''' \subseteq \kappa_h'$ . Since  $\kappa'$  is also minimally redundant,  $\kappa_h''' = \kappa_h'$ . Therefore, By the first statement of Lemma 59,  $\mathcal{K}''' = \mathcal{K}'$ .

Finally, by the first statement of Lemma 59, for every  $\kappa \in \mathcal{K}$ , there exists a unique  $\kappa' \in \mathcal{K}'$  such that  $\kappa_t' = \kappa_t$  and  $\kappa_h' \subseteq \kappa_h$ ; and for every  $\kappa' \in \mathcal{K}'$ , there exists a unique  $\kappa \in \mathcal{K}$  such that  $\kappa_t' = \kappa_t$  and  $\kappa_h' \subseteq \kappa_h$ .

We next prove statement (ii). By statement (i), there is a unique non-redundant DCC set  $\mathcal{K}''$  over  $\mathbf{V}(\mathcal{G}^*)$  such that  $\mathcal{K}$  is equivalent to  $\mathcal{K}''$  and  $\kappa''$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}']$  for any  $\kappa'' \in \mathcal{K}''$ . Since  $\mathcal{K}'$  is equivalent to  $\mathcal{K}$ ,  $\mathcal{K}''$  is also the unique non-redundant DCC set such that  $\mathcal{K}'$  is equivalent to  $\mathcal{K}''$  and  $\kappa''$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{K}']$  for any  $\kappa'' \in \mathcal{K}''$ . By Lemma 59,  $|\mathcal{K}| = |\mathcal{K}''| = |\mathcal{K}'|$ . Moreover, statement (i) of Lemma 59 indicates that  $\{\kappa_t \mid \kappa \in \mathcal{K}\} = \{\kappa_t'' \mid \kappa'' \in \mathcal{K}''\}$ . therefore,  $\{\kappa_t \mid \kappa \in \mathcal{K}\} = \{\kappa_t'' \mid \kappa'' \in \mathcal{K}''\} = \{\kappa_t' \mid \kappa' \in \mathcal{K}'\}$ .  $\blacksquare$

### B.13 Proof of Theorem 30

**Proof** We first prove statement (i). Theorem 17 proves that  $\mathcal{B}$  can be equivalently represented by a set of DCCs  $\mathcal{K}$ . By Definition 4 of MPDAGs,  $\mathcal{H}$  represents all common direct causal relations shared by all DAGs in  $[\mathcal{G}^*, \mathcal{B}]$ . Due to the equivalence,  $[\mathcal{G}^*, \mathcal{B}] = [\mathcal{G}^*, \mathcal{K}] = [\mathcal{H}, \mathcal{K}]$ . Now consider  $\mathbf{E}_d(\mathcal{H}) \cup \mathcal{K}$ . For any  $\kappa \in \mathcal{K}$ , if  $\kappa$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{K} \setminus \kappa)$ , it can be removed from  $\mathcal{K}$ , yielding a set  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{K} \setminus \kappa)$  that is equivalent to  $\mathbf{E}_d(\mathcal{H}) \cup \mathcal{K}$ . By repeating this procedure, we can iteratively remove all redundant DCCs from  $\mathcal{K}$  until no DCC remains redundant given the others and  $\mathbf{E}_d(\mathcal{H})$ . The remaining DCCs in  $\mathcal{K}$  constitute the desired residual set  $\mathcal{R}$ .

We next prove statement (ii). Let  $\mathcal{R}_1$  and  $\mathcal{R}_2$  be two residual DCC sets. By assumption, every DCC  $\kappa_1 \in \mathcal{R}_1$ , if exists, is not redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup (\mathcal{R}_1 \setminus \{\kappa_1\})$ , and hence  $\mathcal{R}_1 \cap \mathbf{E}_d(\mathcal{H}) = \emptyset$ . Following the similar pruning procedure in the proof of (i), we can further remove redundant DCCs from  $\mathbf{E}_d(\mathcal{H}) \cup \mathcal{R}_1$  and obtain a non-redundant DCC set  $\mathcal{A}_1 \subseteq \mathbf{E}_d(\mathcal{H}) \cup \mathcal{R}_1$  with respect to  $\mathcal{G}^*$ . Clearly,  $\mathcal{A}_1$  is equivalent to  $\mathbf{E}_d(\mathcal{H}) \cup \mathcal{R}_1$ , and  $\mathcal{A}_1$  can be written as a union of two sets, say  $\mathcal{E}_1$  and  $\mathcal{R}'_1$ , where  $\mathcal{E}_1 \subseteq \mathbf{E}_d(\mathcal{H})$  and  $\mathcal{R}'_1 \subseteq \mathcal{R}_1$ . However, if there is a  $\kappa \in \mathcal{R}_1$  but  $\kappa \notin \mathcal{R}'_1$  then  $\kappa$  is redundant with respect to  $\mathcal{E}_1 \cup \mathcal{R}'_1$ , and therefore is redundant with respect to  $\mathbf{E}_d(\mathcal{H}) \cup \mathcal{R}_1 \setminus \{\kappa\}$ , which contradicts the assumption. Therefore,  $\mathcal{R}'_1 = \mathcal{R}_1$  and  $\mathcal{A}_1 = \mathcal{E}_1 \cup \mathcal{R}_1$ . By Theorem 29, there is a unique DCC set  $\mathcal{A}^*$  such that  $\mathcal{A}^*$  is equivalent to  $\mathcal{A}_1$  and every DCC in  $\mathcal{A}^*$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{A}_1]$ , which also means  $\mathcal{A}^*$  is minimally redundant with respect to  $[\mathcal{G}^*, \mathcal{B}]$ . Based on statement (i) of Theorem 29, for any  $\kappa_1 \in \mathcal{R}_1$ , if exists, there is a unique  $\kappa^* \in \mathcal{A}^*$  such that  $\kappa_t^* = \kappa_{1,t}$  and  $\kappa_h^* \subseteq \kappa_{1,h}$ . If  $|\kappa_h^*| = 1$ , then  $\kappa^*$  is a directed edge and  $\kappa^* \in \mathbf{E}_d(\mathcal{H})$ . This means  $\kappa_1 \in \mathcal{R}_1$  is redundant with respect to  $\mathbf{E}_d(\mathcal{H})$ , which is contradicted to our assumption. Therefore, it holds that  $|\kappa_h^*| > 1$ . Similarly, all directed edges in  $\mathcal{E}_1$  are in  $\mathcal{A}^*$ , and hence  $\mathcal{A}^* = \mathcal{E}_1 \cup \mathcal{R}^*$  where  $\mathcal{R}^*$  is the set of DCCs each of which has more than one head and uniquely corresponds to a DCC in  $\mathcal{R}_1$ . By statement (i) of Theorem 29,  $|\mathcal{R}^*| = |\mathcal{R}_1|$ . Due to the uniqueness of the element-wise head-minimal DCC set, applying the same technique given above, we can prove that  $|\mathcal{R}^*| = |\mathcal{R}_2|$ . This completes the proof of statement (ii).

For statement (iii), notice that the DCC set  $\mathcal{R}^*$  in the proof of statement (ii) is a DCC set satisfying conditions (1) to (3). The uniqueness comes from statement (i) of Theorem 29.

Finally, statement (iv) comes from the proof of statement (ii).  $\blacksquare$

#### B.14 Proof of Theorem 31

The proof of Theorem 31 requires the following lemma.

**Lemma 60** *Let  $\mathcal{H}$  be a causal MPDAG. For any vertex  $X$  and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{H})$ , if  $\mathbf{S}$  induces a complete subgraph of  $\mathcal{H}$ , then there is a DAG  $\mathcal{G} \in [\mathcal{H}]$  in which  $\mathbf{S} \subseteq \text{pa}(X, \mathcal{G})$ .*

**Proof** Without loss of generality, we assume that the skeleton of  $\mathcal{H}$ , denoted by  $\mathcal{C}$ , is a connected chordal graph, and  $\mathcal{H}$  is the MPDAG representing  $[\mathcal{C}, \mathcal{B}_d]$  for some direct causal constraints  $\mathcal{B}_d$  (Proposition 9). That is,  $[\mathcal{H}] = [\mathcal{C}, \mathcal{B}_d]$ . We first show that  $\text{pa}(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$  induces a complete subgraph of  $\mathcal{H}$ . In fact, if  $\text{pa}(X, \mathcal{H}) = \emptyset$ , then  $\text{pa}(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$  is complete. If  $\text{pa}(X, \mathcal{H}) \neq \emptyset$ , then for any  $p \in \text{pa}(X, \mathcal{H})$  and  $S \in \mathbf{S}$ ,  $p$  and  $S$  are adjacent in  $\mathcal{H}$ , since otherwise  $X \rightarrow S$  should appear in  $\mathcal{H}$  due to the maximality of  $\mathcal{H}$ . Thus,  $\text{pa}(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$  is also complete.

Fang and He (2020, Theorem 1) proved that for any vertex  $X$  and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{H})$ , the following three statements are equivalent.

- (i) There is a DAG  $\mathcal{G}$  in  $[\mathcal{H}]$  such that  $\text{pa}(X, \mathcal{G}) = \mathbf{S} \cup \text{pa}(X, \mathcal{H})$  and  $\text{ch}(X, \mathcal{G}) = \text{sib}(X, \mathcal{H}) \cup \text{ch}(X, \mathcal{H}) \setminus \mathbf{S}$ .
- (ii) Orienting  $\mathbf{S} \rightarrow X$  and  $X \rightarrow \text{sib}(X, \mathcal{H}) \setminus \mathbf{S}$  in  $\mathcal{H}$  does not introduce any new v-structure collided on  $X$  or any directed triangle containing  $X$ .
- (iii) The induced subgraph of  $\mathcal{H}$  over  $\mathbf{S}$  is complete, and there does not exist an  $S \in \mathbf{S}$  and a  $C \in \text{sib}(X, \mathcal{H}) \setminus \mathbf{S}$  such that  $C \rightarrow S$  is in  $\mathcal{H}$ .

Denote by  $\mathbf{M}$  a maximal clique containing  $\text{pa}(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$  (which definitely exists but may not be unique) and let  $\mathbf{M}_p = \mathbf{M} \cap \text{pa}(X, \mathcal{H}) = \text{pa}(X, \mathcal{H})$ ,  $\mathbf{M}_c = \mathbf{M} \cap \text{ch}(X, \mathcal{H})$  and  $\mathbf{M}_s = \mathbf{M} \cap \text{sib}(X, \mathcal{H})$ . We first show that if  $\mathbf{M}_c = \emptyset$ , then there is a DAG  $\mathcal{G} \in [\mathcal{H}]$  such that  $\text{pa}(X, \mathcal{G}) = \mathbf{M}_s \cup \text{pa}(X, \mathcal{H})$  and  $\text{ch}(X, \mathcal{G}) = \text{sib}(X, \mathcal{H}) \cup \text{ch}(X, \mathcal{H}) \setminus \mathbf{M}_s$ . As  $\mathbf{M}_s \subseteq \mathbf{M}$ ,  $\mathbf{M}_s$  induces a complete subgraph of  $\mathcal{H}$ . Suppose that there is a  $P \in \mathbf{M}_s$  and a  $C \in \text{sib}(X, \mathcal{H}) \setminus \mathbf{M}_s$  such that  $C \rightarrow P$ , then  $C$  is adjacent to every vertex in  $\text{pa}(X, \mathcal{H}) \cup \mathbf{M}_s$ , since otherwise  $X \rightarrow P'$  is in  $\mathcal{H}$  for any  $P' \in \text{pa}(X, \mathcal{H}) \cup \mathbf{M}_s$  which is not adjacent to  $C$ , and contradicts the assumption that  $\text{pa}(X, \mathcal{H}) \cup \mathbf{M}_s \subseteq \text{pa}(X, \mathcal{H}) \cup \text{sib}(X, \mathcal{H})$ . However,  $C$  is adjacent to every vertex in  $\text{pa}(X, \mathcal{H}) \cup \mathbf{M}_s$  implies that  $C$  is adjacent to every vertex in  $\mathbf{M} = \mathbf{M}_s \cup \mathbf{M}_p$ , meaning that  $\mathbf{M}$  is not a maximal clique. This leads to a contradiction. The desired result then follows from Fang and He (2020, Theorem 1).

On the other hand, suppose that  $\mathbf{M}_c \neq \emptyset$  for any maximal clique  $\mathbf{M}$  containing  $\text{pa}(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$ . In the following, we will construct a maximal clique  $\mathbf{M}$  containing  $\text{pa}(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$ , such that orienting  $\mathbf{M}_s \rightarrow X$  and  $X \rightarrow \text{sib}(X, \mathcal{H}) \setminus \mathbf{M}_s$  does not violate statement (iii) of Fang and He (2020, Theorem 1).

Let  $\mathbf{M}^0$  be an arbitrary maximal clique containing  $pa(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$ . If orienting  $\mathbf{M}_s^0 \rightarrow X$  and  $X \rightarrow sib(X, \mathcal{H}) \setminus \mathbf{M}_s^0$  does not violate statement (iii) of Fang and He (2020, Theorem 1), then the proof is completed. If otherwise, let  $\mathbf{C}^0 \subseteq sib(X, \mathcal{H}) \setminus \mathbf{M}_s^0$  be the set of vertices such that for any  $C \in \mathbf{C}^0$ ,  $C \rightarrow P$  for some  $P \in \mathbf{M}_s^0$ . Using the same argument given in the last paragraph, we can prove that  $C$  is adjacent to every vertex in  $pa(X, \mathcal{H}) \cup \mathbf{M}_s^0$  for any  $C \in \mathbf{C}^0$ . Likewise, it can be shown that any two distinct vertices in  $\mathbf{C}^0$ , if exist, are adjacent. Hence,  $\mathbf{C}^0 \cup \mathbf{M}_s^0 \cup pa(X, \mathcal{H})$  is a clique.

Let  $\mathbf{M}^1$  be a maximal clique containing  $\{X\} \cup \mathbf{C}^0 \cup \mathbf{M}_s^0 \cup pa(X, \mathcal{H})$ . By assumption,  $\mathbf{M}_c^1 \neq \emptyset$  and  $\mathbf{S} \subseteq \mathbf{M}_s^0 \subsetneq \mathbf{M}_s^1$ . If orienting  $\mathbf{M}_s^1 \rightarrow X$  and  $X \rightarrow sib(X, \mathcal{H}) \setminus \mathbf{M}_s^1$  does not violate statement (iii) of Fang and He (2020, Theorem 1), then the proof is completed. Otherwise, following the above procedure we can find a new maximal clique  $\mathbf{M}^2$  containing  $\{X\} \cup \mathbf{C}^1 \cup \mathbf{M}_s^1 \cup pa(X, \mathcal{H})$ , where  $\mathbf{C}^1 \subseteq sib(X, \mathcal{H}) \setminus \mathbf{M}_s^1$  be the set of vertices such that for any  $C \in \mathbf{C}^1$ ,  $C \rightarrow P$  for some  $P \in \mathbf{M}_s^1$ . Note that,  $\mathbf{S} \subseteq \mathbf{M}_s^0 \subsetneq \mathbf{M}_s^1 \subsetneq \cdots \subsetneq \mathbf{M}_s^i \subsetneq \cdots \subseteq sib(X, \mathcal{H})$  implies that the above construction will eventually stop as  $sib(X, \mathcal{H})$  is a finite set. That is, we will finally find an  $\mathbf{M} = \mathbf{M}_s^i$  containing  $pa(X, \mathcal{H}) \cup \mathbf{S} \cup \{X\}$  such that orienting  $\mathbf{M}_s \rightarrow X$  and  $X \rightarrow sib(X, \mathcal{H}) \setminus \mathbf{M}_s$  does not violate statement (iii) of Fang and He (2020, Theorem 1), which completes the proof.  $\blacksquare$

**Proof** [Proof of Theorem 31] By definition, (i) is equivalent to (ii), and thus we only prove (i) is equivalent to (iii) in the following.

By definition,  $\mathcal{H}$  is fully informative with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  if and only if  $\kappa_t \xrightarrow{\text{or}} \kappa_h$  holds for all DAGs in  $[\mathcal{H}]$  for any  $\kappa \in \mathcal{K}$ , where  $\kappa := \kappa_t \xrightarrow{\text{or}} \kappa_h$ .

If  $\kappa_h \cap ch(\kappa_t, \mathcal{H}) \neq \emptyset$ , then  $\kappa_t \xrightarrow{\text{or}} \kappa_h$  holds for all DAGs in  $[\mathcal{H}]$ . If  $\kappa_h \cap sib(\kappa_t, \mathcal{H})$  induces an incomplete subgraph of  $\mathcal{H}$ , then there exist  $V_1, V_2 \in \kappa_h \cap sib(\kappa_t, \mathcal{H})$  such that  $V_1, V_2$  are not adjacent. By the first rule of Meek's rules, for any DAG in  $[\mathcal{H}]$ , either  $V_1$  or  $V_2$  is a child of  $\kappa_t$ . Thus,  $\kappa_t \xrightarrow{\text{or}} \kappa_h \cap sib(\kappa_t, \mathcal{H})$  holds for all DAGs in  $[\mathcal{H}]$ .

Conversely, if  $\kappa_h \cap ch(\kappa_t, \mathcal{H}) = \emptyset$  and the induced subgraph of  $\mathcal{H}$  over  $\kappa_h \cap sib(\kappa_t, \mathcal{H})$  is complete, then by Lemma 60, there is a DAG  $\mathcal{G} \in [\mathcal{H}]$  in which  $\kappa_h \cap sib(\kappa_t, \mathcal{H}) \subseteq pa(\kappa_t, \mathcal{G})$ . Therefore,  $\kappa_t \xrightarrow{\text{or}} \kappa_h$  does not hold for  $\mathcal{G}$ .  $\blacksquare$

### B.15 Proof of Corollary 32

**Proof** Corollary 32 follows from Theorem 31 and Equation (5).  $\blacksquare$

### B.16 Proof of Proposition 34

**Proof** By Lemma 55, for any DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$ , the leaf node in the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{U})$ , denoted by  $\mathcal{G}_{\text{sub}}$ , must be  $X$ , as  $X$  is the unique potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ . The result comes from the fact that  $adj(X, \mathcal{G}_{\text{sub}}) = adj(X, \mathcal{U})$ .  $\blacksquare$

### B.17 Proof of Theorem 35

**Proof** The sufficiency follows from Proposition 34, and below we prove the necessity. Since  $\mathcal{H}$  has the same skeleton as  $\mathcal{G}^*$ , it suffices to prove that, for two adjacent variables  $X$  and  $Y$ , if  $X \rightarrow Y$  is not in  $\mathcal{G}^*$  and there is no orientation component for  $Y$  containing  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , then  $X \rightarrow Y$  is not in  $\mathcal{H}$ .

As  $X \rightarrow Y$  is not in  $\mathcal{G}^*$  but  $X$  and  $Y$  are adjacent,  $Y$  is either a parent or a sibling of  $X$  in  $\mathcal{G}^*$ . If  $Y \rightarrow X$  is in  $\mathcal{G}^*$ , then  $Y \rightarrow X$  is in  $\mathcal{H}$ , which completes the proof. Now consider the case where  $X - Y$  is in  $\mathcal{G}^*$ . According to our assumption, with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , every connected undirected induced subgraph containing  $X$  and  $Y$  is not an orientation component for  $Y$ . If there is a such subgraph which is an orientation component for  $X$ , then by Proposition 34 and the definition of an MPDAG,  $Y \rightarrow X$  is in  $\mathcal{H}$ .

On the other hand, with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , if every connected undirected induced subgraph containing  $X$  and  $Y$  is neither an orientation component for  $X$  nor an orientation component for  $Y$ , then by Theorem 25, it either has a potential leaf node which is neither  $X$  nor  $Y$ , or has exactly two potential leaf nodes which are  $X$  and  $Y$ . In the following, we will prove that there is a DAG  $\mathcal{G}_1 \in [\mathcal{G}^*, \mathcal{K}]$  in which  $X \rightarrow Y$  and there is also a DAG  $\mathcal{G}_2 \in [\mathcal{G}^*, \mathcal{K}]$  in which  $Y \rightarrow X$ . According to symmetry, we need only to prove that there is a DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$  in which  $X \rightarrow Y$ .

To prove this claim, it suffices to show that  $\mathcal{K} \cup \{X \rightarrow Y\}$  is consistent with  $\mathcal{G}^*$ . By Theorem 25, any connected undirected induced subgraph  $\mathcal{U}$  has a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ . If  $\mathcal{U}$  does not contain  $X$  or  $Y$ , then it can be easily verify that the potential leaf node of  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  is still a potential lead node with respect to  $\mathcal{K} \cup \{X \rightarrow Y\}$  and  $\mathcal{G}^*$ . If  $\mathcal{U}$  contains both  $X$  and  $Y$ , then we have that,

- (i) if  $\mathcal{U}$  has a potential leaf node with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$  which is neither  $X$  nor  $Y$ , then by the definition of a potential leaf node, such a vertex is still a potential leaf node with respect to  $\mathcal{K} \cup \{X \rightarrow Y\}$  and  $\mathcal{G}^*$ ;
- (ii) if  $\mathcal{U}$  had exactly two potential leaf nodes which are  $X$  and  $Y$ , then with respect to  $\mathcal{K} \cup \{X \rightarrow Y\}$  and  $\mathcal{G}^*$ ,  $Y$  is still a potential leaf node in  $\mathcal{U}$ .

Therefore, any connected undirected induced subgraph of  $\mathcal{G}^*$  has a potential leaf node with respect to  $\mathcal{K} \cup \{X \rightarrow Y\}$  and  $\mathcal{G}^*$ . The desired result follows from Theorem 25.  $\blacksquare$

### B.18 Proof of Proposition 36

**Proof** Since  $\mathcal{U}_1$  is an orientation components for  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , for any  $Y_1 \in \mathbf{V}(\mathcal{U}_1)$  such that  $Y_1 \neq X$ , either  $Y_1$  is a non-simplicial node in  $\mathcal{U}_1$ , or there is a  $\kappa \in \mathcal{K}(\mathcal{U}_1)$  such that  $\kappa_t = Y_1$ . Denote by  $\mathcal{U}_{12}$  the undirected induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{V}(\mathcal{U}_1) \cup \mathbf{V}(\mathcal{U}_2)$ . If  $Y_1$  is a non-simplicial node in  $\mathcal{U}_1$ , then it is a non-simplicial node in  $\mathcal{U}_{12}$ , as two non-adjacent vertices in  $\mathcal{U}_1$  remains non-adjacent in  $\mathcal{U}_{12}$ . Moreover, if there exists a  $\kappa \in \mathcal{K}(\mathcal{U}_1)$  such that  $\kappa_t = Y_1$ , then  $\kappa \in \mathcal{K}(\mathcal{U}_{12})$ . Hence,  $Y_1$  is not a potential leaf node in  $\mathcal{U}_{12}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ . Following the same argument, none of the vertices in  $\mathbf{V}(\mathcal{U}_2) \setminus \{X\}$  is a potential leaf node of  $\mathcal{U}_{12}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ . However, due to the consistency,  $\mathcal{U}_{12}$  should have at least one potential leaf node. Therefore,  $X$

is the only potential leaf node of  $\mathcal{U}_{12}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , which completes the proof. ■

### B.19 Proof of Corollary 38

**Proof** The conclusion follows directly from Proposition 36 and Theorem 35. ■

### B.20 Proof of Theorem 39

**Proof** It is easy to see that the outputted graph of Algorithm 3, denoted by  $\mathcal{U}_{\text{out}}$ , is an orientation component for  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ . Assuming that  $\mathcal{U}_{\text{out}}$  is not maximal, then the true maximal orientation component for  $X$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , denoted by  $\mathcal{U}_{\text{true}}$ , is a proper super graph of  $\mathcal{U}_{\text{out}}$ . Recall that Algorithm 3 removes one vertex from the current graph in each while loop until the current graph is  $\mathcal{U}_{\text{out}}$ . Since  $\mathbf{V}(\mathcal{U}_{\text{out}}) \subsetneq \mathbf{V}(\mathcal{U}_{\text{true}})$ , let  $\{Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}\} = \mathbf{V}(\mathcal{U}_{\text{true}}) \setminus \mathbf{V}(\mathcal{U}_{\text{out}})$ , where the subscript of each  $Y_i$  indicates the number of loops when it is removed. Without loss of generality we can assume that  $i_1 < i_2 < \dots < i_k$ . Based on Algorithm 3, the vertices that are removed before  $Y_{i_1}$ , if exist, are not in  $\mathbf{V}(\mathcal{U}_{\text{out}})$ , since otherwise the while loop ends before considering  $Y_{i_1}$ . Let  $\mathcal{U}_1$  be the undirected graph right before removing  $Y_{i_1}$ . By Algorithm 3, with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ ,  $\mathcal{U}_1$  is not an orientation component for  $X$  and  $Y_{i_1}$  is a potential lead node in  $\mathcal{U}_1$ . This means  $Y_{i_1}$  is a simplicial node in  $\mathcal{U}_1$  and there is no DCC in  $\mathcal{K}(\mathcal{U}_1)$  in which  $Y_{i_1}$  is the tail. Note that  $\mathcal{U}_{\text{true}}$  is the induced subgraph of  $\mathcal{U}_1$  over  $\mathbf{V}(\mathcal{U}_{\text{true}})$  and  $Y_{i_1} \in \mathbf{V}(\mathcal{U}_{\text{true}})$ ,  $Y_{i_1}$  is also a simplicial node in  $\mathbf{V}(\mathcal{U}_{\text{true}})$ . Moreover, there is no DCC in  $\mathcal{K}(\mathcal{U}_{\text{true}})$  in which  $Y_{i_1}$  is the tail, as  $\mathcal{K}(\mathcal{U}_{\text{true}}) \subseteq \mathcal{K}(\mathcal{U}_1)$ . Thus,  $Y_{i_1}$  is a potential leaf node in  $\mathcal{U}_{\text{true}}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , meaning that  $\mathcal{U}_{\text{true}}$  is not an orientation component for  $X$ . This leads to a contradiction, and hence  $\mathcal{U}_{\text{out}}$  is maximal. ■

### B.21 Proof of Theorem 41

The key ingredient for proving the identifiability is the following lemma.

**Lemma 61** *Let  $\mathcal{H}$  be an MPDAG representing  $[\mathcal{G}^*, \mathcal{K}]$  induced by a CPDAG  $\mathcal{G}^*$  and a set of consistent DCCs  $\mathcal{K}$ . Then, for any vertex  $X$ , there is a DAG  $\mathcal{G}$  in  $[\mathcal{G}^*, \mathcal{K}]$  such that  $ch(X, \mathcal{G}) = sib(X, \mathcal{H}) \cup ch(X, \mathcal{H})$ .*

**Proof** It suffices to show that  $\mathcal{K} \cup \{X \rightarrow W \mid W \in sib(X, \mathcal{H})\}$  is consistent with  $\mathcal{G}^*$ . For any connected undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{G}^*$ , a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , if it is not  $X$ , is still a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K} \cup \{X \rightarrow W \mid W \in sib(X, \mathcal{H})\}$  and  $\mathcal{G}^*$ . On the other hand, if  $X$  is the only potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{G}^*$ , then  $adj(X, \mathcal{U}) \rightarrow X$  are in  $\mathcal{H}$ . Consequently,  $W$  is not in  $\mathcal{U}$  since  $W \in sib(X, \mathcal{H})$ , and hence  $X$  is still the only potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K} \cup \{X \rightarrow W \mid W \in sib(X, \mathcal{H})\}$  and  $\mathcal{G}^*$ . The desired result follows from Theorem 25. ■



Perković (2020, Proposition 3.2) showed that,

**Lemma 62 (Perković 2020, Proposition 3.2)** *Let  $\mathcal{H}$  be a causal MPDAG. If there is a proper possibly causal path  $\pi = (X, W, U, \dots, Y)$  from  $X \in \mathbf{X}$  to  $Y \in \mathbf{Y}$  such that  $X \rightarrow W \rightarrow U \rightarrow \dots \rightarrow Y$  is in one DAG  $\mathcal{G}_1 \in [\mathcal{H}]$  and  $X \leftarrow W \rightarrow U \rightarrow \dots \rightarrow Y$  is in another DAG  $\mathcal{G}_2 \in [\mathcal{H}]$ , then there exists a multivariate Gaussian density  $f$  over  $\mathbf{V}(\mathcal{G}^*)$  such that  $f_1(y \mid do(\mathbf{x})) \neq f_2(y \mid do(\mathbf{x}))$ , where  $f_1(y \mid do(\mathbf{x}))$  and  $f_2(y \mid do(\mathbf{x}))$  are interventional distributions computed from two causal models  $(\mathcal{G}_1, f(\mathbf{V}))$  and  $(\mathcal{G}_2, f(\mathbf{V}))$ , respectively.*

Lemma 62 results the following corollary.

**Corollary 63** *Let  $\mathcal{H}$  be an MPDAG representing  $[\mathcal{G}^*, \mathcal{K}]$  for a CPDAG  $\mathcal{G}^*$  and a set of consistent DCCs  $\mathcal{K}$ . If there is a proper possibly causal path  $\pi = (X, W, U, \dots, Y)$  from  $X \in \mathbf{X}$  to  $Y \in \mathbf{Y}$  such that  $X \rightarrow W \rightarrow U \rightarrow \dots \rightarrow Y$  is in one DAG  $\mathcal{G}_1 \in [\mathcal{G}^*, \mathcal{K}]$  and  $X \leftarrow W \rightarrow U \rightarrow \dots \rightarrow Y$  is in another DAG  $\mathcal{G}_2 \in [\mathcal{G}^*, \mathcal{K}]$ , then there exists a multivariate Gaussian density  $f$  over  $\mathbf{V}(\mathcal{G}^*)$  such that  $f_1(y \mid do(\mathbf{x})) \neq f_2(y \mid do(\mathbf{x}))$ , where  $f_1(y \mid do(\mathbf{x}))$  and  $f_2(y \mid do(\mathbf{x}))$  are interventional distributions computed from two causal models  $(\mathcal{G}_1, f(\mathbf{V}))$  and  $(\mathcal{G}_2, f(\mathbf{V}))$ , respectively.*

Corollary 63 holds because  $[\mathcal{G}^*, \mathcal{K}] \subseteq [\mathcal{H}]$ . Thus, we omit the proof. Analogue to Corollary 63, we have,

**Lemma 64** *Let  $\mathcal{H}$  be an MPDAG representing  $[\mathcal{G}^*, \mathcal{K}]$  for a CPDAG  $\mathcal{G}^*$  and a set of consistent DCCs  $\mathcal{K}$ . If there is a proper possibly causal path  $\pi = (X, W, U, \dots, Y)$  from  $X \in \mathbf{X}$  to  $Y \in \mathbf{Y}$  such that  $X \rightarrow W \leftarrow U \rightarrow \dots \rightarrow Y$  and  $X \rightarrow U$  are in one DAG  $\mathcal{G}_1 \in [\mathcal{G}^*, \mathcal{K}]$ , and  $X \leftarrow W \rightarrow U \rightarrow \dots \rightarrow Y$  and  $X \rightarrow U$  are in another DAG  $\mathcal{G}_2 \in [\mathcal{G}^*, \mathcal{K}]$ , then there exists a multivariate Gaussian density  $f$  over  $\mathbf{V}(\mathcal{G}^*)$  such that  $f_1(y \mid do(\mathbf{x})) \neq f_2(y \mid do(\mathbf{x}))$ , where  $f_1(y \mid do(\mathbf{x}))$  and  $f_2(y \mid do(\mathbf{x}))$  are interventional distributions computed from two causal models  $(\mathcal{G}_1, f(\mathbf{V}))$  and  $(\mathcal{G}_2, f(\mathbf{V}))$ , respectively.*

**Proof** Let  $f$  be a multivariate Gaussian density determined by the following linear Gaussian structural equation model,

$$X_i = \sum_{X_j \in pa(X_i, \mathcal{G}_1)} \beta_{ji} X_j + \epsilon_i,$$

where  $\beta_{ji} = 0$  if  $X_j \rightarrow X_i$  is neither  $X \rightarrow U$  nor the edges on the corresponding path of  $\pi$  in  $\mathcal{G}_2$  and  $0 < \beta_{ji} < 1$  otherwise,  $\epsilon_i$  are Gaussian noises with zero means and variances that makes every variable has variance one. (This is possible if we set  $\beta_{ji}$ 's small enough.) It is clear that  $f$  is Markovian to  $\mathcal{G}_2$ , and thus Markovian to  $\mathcal{G}_1$  as  $\mathcal{G}_1$  and  $\mathcal{G}_2$  are Markov equivalent. Moreover, denote by  $\mathcal{G}'_i$  the DAGs obtained by removing from  $\mathcal{G}_i$  the edges that are neither on the corresponding path of  $\pi$  nor  $X \rightarrow U$  ( $i = 1, 2$ ), it can be checked that  $f$  is Markovian to  $\mathcal{G}'_i$  and  $f_i(y \mid do(\mathbf{x})) = f'_i(y \mid do(\mathbf{x}))$ , where  $f'_i(y \mid do(\mathbf{x}))$  is the interventional distributions computed from the causal model  $(\mathcal{G}'_i, f(\mathbf{V}))$ ,  $i = 1, 2$ .

Using the backdoor adjustment, it can be verified that  $E_1(y \mid do(X = 1)) = \sigma_{xy}$  and  $E_2(y \mid do(X = 1)) = (\sigma_{xy} - \sigma_{xw}\sigma_{wy})/(1 - \sigma_{xw}^2)$ . By Wright's rule (Wright, 1921), it can be checked that  $\sigma_{xy} - (\sigma_{xy} - \sigma_{xw}\sigma_{wy})/(1 - \sigma_{xw}^2)$  equals to the product of the edge weights along the path  $X \leftarrow W \rightarrow U \rightarrow \dots \rightarrow Y$  in  $\mathcal{G}_2$ . By our assumption, the edge weights are non-zero,

$E_1(y \mid do(X = 1)) \neq E_2(y \mid do(X = 1))$ , and consequently,  $f_1(y \mid do(\mathbf{x})) \neq f_2(y \mid do(\mathbf{x}))$ . ■

Finally, the proof of Theorem 41 follows a similar argument to that for Perković et al. (2017, Lemma C.1).

**Proof** [Proof of Theorem 41] Based on Perković (2020, Theorem 3.6), the sufficiency of the identification condition holds, since  $[\mathcal{G}^*, \mathcal{K}] \subseteq [\mathcal{H}]$  and the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable from  $\mathcal{H}$  when the condition holds.

To prove the necessity, let  $\pi$  be a possibly causal path from  $X \in \mathbf{X}$  to  $Y \in \mathbf{Y}$  in  $\mathcal{H}$  where the first edge from the side of  $X$  is undirected. Denote by  $\pi^* = (X, W, U, \dots, Y)$ , a shortest subsequence of  $\pi$  with length at least 1, such that  $\pi^*$  is also a possibly causal path from  $X$  to  $Y$ , where the first edge from the side of  $X$  is undirected. It is clear that  $\pi^*(W, Y)$  is unshielded. By Lemma 61, there is a DAG  $\mathcal{G}_2 \in [\mathcal{G}^*, \mathcal{K}]$  such that  $ch(W, \mathcal{G}_2) = sib(W, \mathcal{H}) \cup ch(W, \mathcal{H})$ . Hence, the corresponding path of  $\pi^*$  in  $\mathcal{G}_2$  is  $X \leftarrow W \rightarrow U \rightarrow \dots \rightarrow Y$ , according to the first Meek's rule.

If  $X$  is not adjacent to  $U$  in  $\mathcal{H}$ , then we consider a DAG  $\mathcal{G}_1 \in [\mathcal{G}^*, \mathcal{K}]$  where  $ch(X, \mathcal{G}_1) = sib(X, \mathcal{H}) \cup ch(X, \mathcal{H})$ . By Lemma 61, such a DAG exists. Since  $X$  is not adjacent to  $U$ ,  $\pi^*$  is unshielded, and thus the corresponding path of  $\pi^*$  in  $\mathcal{G}_1$  is  $X \rightarrow W \rightarrow U \rightarrow \dots \rightarrow Y$ . By Corollary 63, the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is not identifiable.

If  $X$  is adjacent to  $U$ , then  $X \rightarrow U$  is in  $\mathcal{H}$ , since otherwise  $X - U$  and  $\pi^*(U, Y)$  form a possibly causal path shorter than  $\pi^*$ , which contradicts our assumption. If  $W \rightarrow U$  is in  $\mathcal{H}$ , then we again consider a DAG  $\mathcal{G}_1 \in [\mathcal{G}^*, \mathcal{K}]$  where  $ch(X, \mathcal{G}_1) = sib(X, \mathcal{H}) \cup ch(X, \mathcal{H})$ . In  $\mathcal{G}_1$ , the corresponding path of  $\pi^*$  is  $X \rightarrow W \rightarrow U \rightarrow \dots \rightarrow Y$ . If  $W - U$  is in  $\mathcal{H}$ , then we consider the DAG  $\mathcal{G}_1 \in [\mathcal{G}^*, \mathcal{K}]$  where  $ch(U, \mathcal{G}_1) = sib(U, \mathcal{H}) \cup ch(U, \mathcal{H})$ . Since  $X \rightarrow U$  and  $U \rightarrow W$  are in  $\mathcal{G}_1$ ,  $X \rightarrow W$  is in  $\mathcal{G}_1$ . Thus, according to Lemma 64, the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is not identifiable. ■

## B.22 Proof of Corollary 42

**Proof** We first prove that (ii)  $\Rightarrow$  (i). Assume that  $X \rightarrow Y$  exists in every DAG in  $[\mathcal{G}^*, \mathcal{K}]$ , but  $X - Y$  is in  $\mathcal{G}^*$ , then the causal effect of  $Y$  on  $X$  is not identifiable in  $[\mathcal{G}^*]$  but becomes identifiable in  $[\mathcal{G}^*, \mathcal{K}]$ . In fact, the causal effect of  $Y$  on  $X$  is 0 in every DAG in  $[\mathcal{G}^*, \mathcal{K}]$ .

Conversely, if the common directed causal relations of the DAGs in  $[\mathcal{G}^*, \mathcal{K}]$  are all encoded by directed edges in  $\mathcal{G}^*$ , then by the definition of a causal MPDAG,  $\mathcal{H} = \mathcal{G}^*$ . By Theorem 41, an effect is identifiable in  $[\mathcal{G}^*]$  if and only if it is identifiable in  $[\mathcal{G}^*, \mathcal{K}]$ . This completes the proof of (i)  $\Rightarrow$  (ii).

Next, suppose that  $\mathcal{K}$  is derived from a consistent pairwise causal background knowledge set  $\mathcal{B}$ . If there is a direct causal constraint in  $\mathcal{B}$  which does not hold for all DAGs in  $[\mathcal{G}^*]$ , then it is clear that an unidentifiable effect becomes identifiable in  $[\mathcal{G}^*, \mathcal{K}]$ , as statement (ii) holds. If there is a non-ancestral causal constraint in  $\mathcal{B}$  which does not hold for all DAGs in  $[\mathcal{G}^*]$ , then by Theorem 17, statement (ii) also holds. Finally, since  $X \leftrightarrow Y$  implies  $Y \leftrightarrow X$ ,  $\mathcal{B}$  is equivalent to  $(Y \leftrightarrow X) \cup \mathcal{B}$  with respect to  $\mathcal{G}^*$ . Thus, if  $Y \leftrightarrow X$  does not hold for all DAG in  $[\mathcal{G}^*]$ , then statement (ii) holds. ■

### B.23 Proof of Theorem 43

Before proving Theorem 43, we first introduce some technical lemmas.

**Lemma 65** (Perković et al. 2017, Lemma C.2) *Let  $\mathcal{H}$  be a causal MPDAG, and  $\mathbf{X}, \mathbf{Y}$  are disjoint subsets of vertices of  $\mathcal{H}$ . Suppose that the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{H}]$ , then*

- (i)  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) = \emptyset$  implies that  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$  for any DAG  $\mathcal{G} \in [\mathcal{H}]$ , and
- (ii)  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) \neq \emptyset$  implies that in  $\mathcal{H}$  there exist  $X \in \mathbf{X}$ ,  $Y \in \mathbf{Y}$ ,  $U \in \mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H})$  and  $W$  such that there are a directed path from  $X$  to  $W$  and unshielded possibly causal paths from  $W$  to  $Y$  and  $U$ , respectively.

**Lemma 66** (Perković et al. 2017, Lemma C.3) *Let  $\mathcal{H}$  be a causal MPDAG, and  $\mathbf{X}, \mathbf{Y}$  are disjoint subsets of vertices of  $\mathcal{H}$ . Suppose that the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{H}]$  and  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) = \emptyset$ , then there is a proper definite status non-causal path from  $X \in \mathbf{X}$  to  $Y \in \mathbf{Y}$  which is not blocked by  $\mathbf{Z}$  in  $\mathcal{H}$  implies that  $\mathbf{Z}$  is not an adjustment set for  $(\mathbf{X}, \mathbf{Y})$  in any DAG  $\mathcal{G} \in [\mathcal{H}]$ .*

Lemmas 65 and 66 result the following Corollaries 67 and 68, respectively.

**Corollary 67** *Let  $\mathcal{H}$  be an MPDAG representing  $[\mathcal{G}^*, \mathcal{K}]$  for a CPDAG  $\mathcal{G}^*$  and a set of consistent DCCs  $\mathcal{K}$ , and  $\mathbf{X}, \mathbf{Y}$  are disjoint subsets of vertices of  $\mathcal{H}$ . Suppose that the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{G}^*, \mathcal{K}]$ , then*

- (i)  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) = \emptyset$  implies that  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) = \emptyset$  for any DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$ , and
- (ii)  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) \neq \emptyset$  implies that in  $\mathcal{H}$  there exist  $X \in \mathbf{X}$ ,  $Y \in \mathbf{Y}$ ,  $U \in \mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H})$  and  $W$  such that there are a directed path from  $X$  to  $W$  and unshielded possibly causal paths from  $W$  to  $Y$  and  $U$ , respectively.

**Corollary 68** *Let  $\mathcal{H}$  be an MPDAG representing  $[\mathcal{G}^*, \mathcal{K}]$  for a CPDAG  $\mathcal{G}^*$  and a set of consistent DCCs  $\mathcal{K}$ , and  $\mathbf{X}, \mathbf{Y}$  are disjoint subsets of vertices of  $\mathcal{H}$ . Suppose that the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{G}^*, \mathcal{K}]$  and  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H}) = \emptyset$ , then there is a proper definite status non-causal path from  $X \in \mathbf{X}$  to  $Y \in \mathbf{Y}$  which is not blocked by  $\mathbf{Z}$  in  $\mathcal{H}$  implies that  $\mathbf{Z}$  is not an adjustment set for  $(\mathbf{X}, \mathbf{Y})$  in any DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{K}]$ .*

The above corollaries hold since  $[\mathcal{G}^*, \mathcal{K}] \subseteq [\mathcal{H}]$  and Theorem 41 proves that the causal effect of  $\mathbf{X}$  on  $\mathbf{Y}$  is identifiable in  $[\mathcal{G}^*, \mathcal{K}]$  if and only if it is identifiable in  $[\mathcal{H}]$ . We omit the proofs here.

The proof of Theorem 43 is analogue to that of Perković et al. (2017, Theorem 4.4).

**Proof** [Proof of Theorem 43] Based on Theorem 4.4 in Perković et al. (2017) and the fact that  $[\mathcal{G}^*, \mathcal{B}] \subseteq [\mathcal{H}]$ , it is clear that  $\mathbf{Z}$  satisfies the b-adjustment criterion relative to  $(\mathbf{X}, \mathbf{Y})$  in  $\mathcal{H}$  implies that  $\mathbf{Z}$  is an adjustment set for  $(\mathbf{X}, \mathbf{Y})$  with respect to  $\mathcal{G}^*$  and  $\mathcal{B}$ . To prove the other direction, we use the proof by contradiction. First, Theorem 41 indicates that the

first condition holds. Assuming the first condition holds but the second condition fails to hold, then by Corollary 67, there exist  $X \in \mathbf{X}$ ,  $Y \in \mathbf{Y}$ ,  $U \in \mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{H})$  and  $W$  such that there are a directed path from  $X$  to  $W$  and unshielded possibly causal paths from  $W$  to  $Y$  and  $U$ , respectively. According to Lemma 61, there is a DAG  $\mathcal{G} \in [\mathcal{G}^*, \mathcal{B}]$  such that every sibling of  $W$  is a child of  $W$  in  $\mathcal{G}$ . Thus, the unshielded possibly causal paths from  $W$  to  $Y$  and  $U$  are directed in  $\mathcal{G}$ , which means  $\mathbf{Z} \cap \text{Forb}(\mathbf{X}, \mathbf{Y}, \mathcal{G}) \neq \emptyset$  for  $\mathcal{G}$ . Consequently, by Perković (2020, Theorem 4.4),  $\mathbf{Z}$  is not an adjustment set for  $(\mathbf{X}, \mathbf{Y})$  in  $\mathcal{G}$ . Finally, the necessity of the third condition is guaranteed by Corollary 68.  $\blacksquare$

## B.24 Proof of Theorem 46

Maathuis et al. (2009) proved that,

**Lemma 69 (Maathuis et al. 2009, Lemma 3.1)** *Given a CPDAG  $\mathcal{G}^*$ , a treatment  $X$ , and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{G}^*)$ , there is a DAG  $\mathcal{G} \in [\mathcal{G}^*]$  such that  $\text{pa}(X, \mathcal{G}) = \mathbf{S} \cup \text{pa}(X, \mathcal{G}^*)$  and  $\text{ch}(X, \mathcal{G}) = \text{sib}(X, \mathcal{G}^*) \cup \text{ch}(X, \mathcal{G}^*) \setminus \mathbf{S}$  if and only if the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{S}$  is complete.*

Given a CPDAG  $\mathcal{G}^*$ , a treatment  $X$  and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{G}^*)$ , and suppose that there is a DAG  $\mathcal{G} \in [\mathcal{G}^*]$  such that  $\text{pa}(X, \mathcal{G}) = \mathbf{S} \cup \text{pa}(X, \mathcal{G}^*)$  and  $\text{ch}(X, \mathcal{G}) = \text{sib}(X, \mathcal{G}^*) \cup \text{ch}(X, \mathcal{G}^*) \setminus \mathbf{S}$ . Regarding  $\mathbf{S} \rightarrow X$  and  $X \rightarrow \text{sib}(X, \mathcal{G}^*) \setminus \mathbf{S}$  as direct causal constraints and denote them by  $\mathcal{K}$ , the MPDAG  $\mathcal{H}$  of  $[\mathcal{G}^*, \mathcal{K}]$  is a chain graph (He and Geng, 2008, Theorem 6). Suppose that, apart from  $\mathcal{K}$ , we have another DCC set  $\mathcal{K}'$ . The following lemma extends Theorem 25 and gives a sufficient and necessary condition to check whether  $\mathcal{K}'$  is consistent with  $\mathcal{G}^*$  given  $\mathcal{K}$  (that is,  $\mathcal{K}' \cup \mathcal{K}$  is consistent with  $\mathcal{G}^*$ ).

**Lemma 70** *Let  $\mathcal{G}^*$  be a CPDAG,  $X$  be a variable in  $\mathcal{G}^*$  and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{G}^*)$ . Let  $\mathcal{K} = \{S \rightarrow X \mid S \in \mathbf{S}\} \cup \{X \rightarrow C \mid C \in \text{sib}(X, \mathcal{G}^*) \setminus \mathbf{S}\}$  and assume that  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ . Denote by  $\mathcal{H}$  the MPDAG of  $[\mathcal{G}^*, \mathcal{K}]$ . For any DCC set  $\mathcal{K}'$ , the following two statements are equivalent.*

- (i)  $\mathcal{K}'$  is consistent with  $\mathcal{G}^*$  given  $\mathcal{K}$ .
- (ii) Any connected undirected induced subgraph of  $\mathcal{H}$  has a potential leaf node with respect to  $\mathcal{K}'$  and  $\mathcal{H}$ .

Note that, in the main text the potential leaf node is defined over a CPDAG instead of an MPDAG. Therefore, for the sake of rigor, we extend the related definitions to chain graph causal MPDAG in the following, where a chain graph causal MPDAG is a causal MPDAG which itself is a chain graph.

**Definition 71** *Given a chain graph causal MPDAG  $\mathcal{H}$  and a set of DCCs  $\mathcal{K}$  over  $\mathbf{V}(\mathcal{H})$ , a reduced form of  $\mathcal{K}$  with respect to  $\mathcal{H}$ , denoted by  $\mathcal{K}(\mathcal{H})$ , is defined as follows.*

$$\mathcal{K}(\mathcal{H}) := \{\kappa_t \xrightarrow{\text{or}} (\kappa_h \cap \text{sib}(\kappa_t, \mathcal{H})) \mid \kappa \in \mathcal{K} \text{ and } \kappa_h \cap \text{ch}(\kappa_t, \mathcal{H}) = \emptyset\}. \quad (9)$$

Similar to Proposition 22, it is easy to verify that a DAG in  $[\mathcal{H}]$  satisfies all constraints in  $\mathcal{K}(\mathcal{H})$  if and only if it satisfies all constraints in  $\mathcal{K}$ .

**Definition 72** *Given an undirected induced subgraph  $\mathcal{U}$  of a chain graph causal MPDAG  $\mathcal{H}$  over  $\mathbf{V}(\mathcal{U}) \subseteq \mathbf{V}(\mathcal{H})$ , and a set of DCCs  $\mathcal{K}$  over  $\mathbf{V}(\mathcal{H})$ , the restriction subset of  $\mathcal{K}$  on  $\mathcal{U}$  given  $\mathcal{H}$  is defined by*

$$\mathcal{K}(\mathcal{U} \mid \mathcal{H}) := \{\kappa \in \mathcal{K}(\mathcal{H}) \mid \{\kappa_t\} \cup \kappa_h \subseteq \mathbf{V}(\mathcal{U})\}. \quad (10)$$

**Definition 73** *Let  $\mathcal{H}$  be a chain graph causal MPDAG and  $\mathcal{K}$  be a set of DCCs over  $\mathbf{V}(\mathcal{H})$ . Given an undirected induced subgraph  $\mathcal{U}$  of  $\mathcal{H}$  and a vertex  $X$  in  $\mathcal{U}$ ,  $X$  is called a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}$  and  $\mathcal{H}$ , if  $X$  is a simplicial vertex in  $\mathcal{U}$  and  $X$  is not the tail of any clause in  $\mathcal{K}(\mathcal{U} \mid \mathcal{H})$ .*

**Proof** [proof of Lemma 70] The proof is similar to that of Theorem 25. We first prove the necessity. If  $\mathcal{K}'$  is consistent with  $\mathcal{G}^*$  given  $\mathcal{K}$ , then there exists a DAG  $\mathcal{G} \in [\mathcal{H}]$  satisfying all constraints in  $\mathcal{K}'$ . Let  $\mathcal{U}$  be an arbitrary connected undirected induced subgraph of  $\mathcal{H}$ , and denote the induced subgraph of  $\mathcal{G}$  over  $\mathbf{V}(\mathcal{U})$  by  $\mathcal{G}_{\text{sub}}$ . Since any induced subgraph of a DAG is still a DAG,  $\mathcal{G}_{\text{sub}}$  is a DAG, and thus it must have a leaf node  $V_{\text{leaf}}$ . As we assume that  $\mathcal{H}$  is a chain graph, following exactly the same argument for proving Lemma 55, we can show that  $V_{\text{leaf}}$  is a potential leaf node in  $\mathcal{U}$  with respect to  $\mathcal{K}'$  and  $\mathcal{H}$ .

We next prove the sufficiency. Since  $\mathcal{H}$  is a chain graph, no orientation of the edges not oriented in  $\mathcal{H}$  will create a directed cycle which includes an edge or edges that were oriented in  $\mathcal{H}$ . Moreover, based on statement (iii) of Theorem 8, no orientation of an edge not directed in  $\mathcal{H}$  can create a new v-structure with an edge that was oriented in  $\mathcal{H}$ . Then, following the same argument for proving Lemma 56, we can show that  $\mathcal{K}'$  is consistent with  $\mathcal{G}^*$  given  $\mathcal{K}$  if and only if  $\mathcal{K}'(\mathcal{C} \mid \mathcal{H})$  is consistent with  $\mathcal{C}$  for any chain component  $\mathcal{C}$  of  $\mathcal{H}$ . The desired result comes from the same construction of PEO given in the proof of Theorem 25. ■

In order to prove Theorem 46, we prove the following Theorem 46', which includes the result provided in Theorem 46.

**Theorem 46'** *Let  $\mathcal{K}$  be a set of DCCs consistent with a CPDAG  $\mathcal{G}^*$ , and  $\mathcal{H}$  be the MPDAG of  $[\mathcal{G}^*, \mathcal{K}]$ . For any vertex  $X$  and  $\mathbf{S} \subseteq \text{sib}(X, \mathcal{H})$ , let*

$$\mathbf{T} = \{X\} \cup ((\text{pa}(X, \mathcal{H}) \cup \mathbf{S}) \cap \text{sib}(X, \mathcal{G}^*)),$$

and

$$D_X = \{u \rightarrow X \mid u \in \text{pa}(X, \mathcal{H}) \cup \mathbf{S}\} \cup \{X \rightarrow v \mid v \in \text{sib}(X, \mathcal{H}) \cup \text{ch}(X, \mathcal{H}) \setminus \mathbf{S}\}.$$

Then, the following statements are equivalent.

- (i) *There is a DAG  $\mathcal{G}$  in  $[\mathcal{G}^*, \mathcal{K}]$  such that  $\text{pa}(X, \mathcal{G}) = \mathbf{S} \cup \text{pa}(X, \mathcal{H})$  and  $\text{ch}(X, \mathcal{G}) = \text{sib}(X, \mathcal{H}) \cup \text{ch}(X, \mathcal{H}) \setminus \mathbf{S}$ .*
- (ii) *The induced subgraph of  $\mathcal{H}$  over  $\mathbf{S}$  is complete and the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\mathbf{M}_{\mathbf{T}})$  given  $\mathcal{G}^*$  is consistent with  $\mathcal{G}^*(\mathbf{M}_{\mathbf{T}})$  for all maximal clique  $\mathbf{M}_{\mathbf{T}}$  of  $\mathcal{G}^*$  containing  $\mathbf{T}$ .*

(iii) Given  $\mathcal{G}^*$ , the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$  is consistent with  $\mathcal{G}^*(\{X\} \cup \text{sib}(X, \mathcal{G}^*))$ .

**Proof** [Proof of Theorem 46'] By Theorem 25, statement (i) implies statement (iii) and statement (iii)  $\Rightarrow$  (ii). Thus, we only prove (ii)  $\Rightarrow$  (i) in the following.

To prove (ii)  $\Rightarrow$  (i), it suffices to show that  $\mathcal{K} \cup D_X$  is consistent with  $\mathcal{G}^*$ . According to Lemma 56, we can consider each chain component of  $\mathcal{G}^*$  separately. Therefore, without loss of generality, we can assume that  $\mathcal{G}^*$  is a connected chordal graph, and that  $\mathcal{K}$  is already in its reduced form with respect to  $\mathcal{G}^*$ . This means we have that,

(P1) for any  $\kappa \in \mathcal{K}$ ,  $\kappa_h \subseteq \text{sib}(\kappa_t, \mathcal{G}^*) = \text{adj}(\kappa_t, \mathcal{G}^*)$ , and the consistency of  $\mathcal{K}$  with  $\mathcal{G}^*$  indicates that  $\kappa_h \neq \emptyset$ .

It can be seen from the definition of consistency that checking the consistency of  $\mathcal{K} \cup D_X$  with  $\mathcal{G}^*$  is equivalent to the following two-steps procedure:

**Step 1.** checking whether  $D_X$  is consistent with  $\mathcal{G}^*$ , and if the consistency holds, then

**Step 2.** checking whether  $\mathcal{K} \setminus D_X$  is consistent with  $\mathcal{G}^*$  given  $D_X$ .

Note that, the intersection of  $\mathcal{K}$  and  $D_X$  may not be empty. Clearly,  $\mathcal{K} \cup D_X$  is consistent if and only if neither of the above two steps returns a negative answer.

We first prove that  $D_X$  is consistent with  $\mathcal{G}^*$ , meaning that step 1 returns a positive answer. Based on statement (ii), the induced subgraph of  $\mathcal{H}$  over  $\mathbf{S}$  is complete, indicating that the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{S}$  is complete as  $\mathcal{H}$  and  $\mathcal{G}^*$  has the same skeleton. If  $pa(X, \mathcal{H}) = \emptyset$ , then  $pa(X, \mathcal{H}) \cup \mathbf{S}$  induces a complete subgraph. If  $pa(X, \mathcal{H}) \neq \emptyset$ , then for any  $p \in pa(X, \mathcal{H})$  and  $S \in \mathbf{S}$ ,  $p$  and  $S$  are adjacent in  $\mathcal{H}$ , since otherwise  $X \rightarrow S$  should appear in  $\mathcal{H}$  due to the maximality of  $\mathcal{H}$ . Moreover, since we have assumed that  $\mathcal{G}^*$  is a connected chordal graph,  $\mathcal{H}$  has no v-structure as discussed in Appendix B.1, and thus,  $pa(X, \mathcal{H})$  induces a complete subgraph. Therefore, it holds that,

(P2)  $pa(X, \mathcal{H}) \cup \mathbf{S}$  induces a complete subgraph of  $\mathcal{G}^*$ .

By Lemma 69 and the definition of  $D_X$ ,  $D_X$  is consistent with  $\mathcal{G}^*$ .

Below we will show that  $\mathcal{K} \setminus D_X$  is consistent with  $\mathcal{G}^*$  given  $D_X$ . Note that, since  $D_X$  contains direct causal constraints only, if we denote the MPDAG representing  $[\mathcal{G}^*, D_X]$  by  $\mathcal{C}^*$ , then  $\mathcal{C}^*$  is a chain graph MPDAG and  $[\mathcal{C}^*] = [\mathcal{G}^*, D_X]$  (He and Geng, 2008, Theorem 6). Now, consider the following four subsets of  $\mathcal{K} \setminus D_X$ :

$$\begin{aligned}\mathcal{K}_p &= \{\kappa \in \mathcal{K} \setminus D_X \mid \exists d \in \kappa_h, \kappa_t \rightarrow d \text{ is in } \mathcal{C}^*\}, \\ \mathcal{K}_u &= \{\kappa \in \mathcal{K} \setminus D_X \mid \forall d \in \kappa_h, \kappa_t - d \text{ is in } \mathcal{C}^*\}, \\ \mathcal{K}_c &= \{\kappa \in \mathcal{K} \setminus D_X \mid \forall d \in \kappa_h, d \rightarrow \kappa_t \text{ is in } \mathcal{C}^*\}, \\ \mathcal{K}_r &= \mathcal{K} \setminus D_X \setminus (\mathcal{K}_p \cup \mathcal{K}_u \cup \mathcal{K}_c).\end{aligned}$$

It is easy to verify that  $\mathcal{K}_p, \mathcal{K}_u, \mathcal{K}_c$  and  $\mathcal{K}_r$  are disjoint and

$$\mathcal{K}_r = \{\kappa \in \mathcal{K} \setminus D_X \setminus \mathcal{K}_p \mid \exists d \in \kappa_h, d \rightarrow \kappa_t \text{ is in } \mathcal{C}^* \text{ and } \exists d \in \kappa_h, d - \kappa_t \text{ is in } \mathcal{C}^*\}.$$

Since the DCCs in  $\mathcal{K}_p$  are already satisfied given  $\mathcal{C}^*$ , to prove that  $\mathcal{K} \setminus D_X$  is consistent with  $\mathcal{G}^*$  given  $D_X$ , it suffices to show that  $\mathcal{K}_u \cup \mathcal{K}_c \cup \mathcal{K}_r$  is consistent with  $\mathcal{G}^*$  given  $D_X$ .

We will give a proof by contradiction. Assuming that  $\mathcal{K}_u \cup \mathcal{K}_c \cup \mathcal{K}_r$  is not consistent with  $\mathcal{G}^*$  given  $D_X$ , then either  $\mathcal{K}_c \neq \emptyset$  or there exists a connected undirected induced subgraph  $\mathcal{W}$  of  $\mathcal{C}^*$  such that  $\mathcal{W}$  has no potential lead node with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{C}^*$ , as indicated by Lemma 70.

**Case 1.** Suppose that  $\mathcal{K}_c \neq \emptyset$ . Let  $\kappa \in \mathcal{K}_c$  be an arbitrary DCC. By the definitions of  $\mathcal{K}_c$  and  $\mathcal{C}^*$  as well as the maximality of  $\mathcal{C}^*$ , it holds that,

(P3) the maximal orientation component for  $\kappa_t$  with respect to  $\mathcal{G}^*$  and  $D_X$ , denoted by  $\mathcal{U}_t$ , satisfies that  $\kappa_h \subseteq pa(\kappa_t, \mathcal{C}^*) = adj(\kappa_t, \mathcal{U}_t)$ .

$pa(\kappa_t, \mathcal{C}^*) = adj(\kappa_t, \mathcal{U}_t)$  is because we have assumed that  $\mathcal{G}^*$  does not contain any directed edge. By (P1),  $\kappa_h \neq \emptyset$ , and thus  $adj(\kappa_t, \mathcal{U}_t) \neq \emptyset$  and  $\mathcal{U}_t$  is not a singleton graph (that is,  $\mathcal{U}_t$  has at least two vertices).

We first show that  $X \in \mathbf{V}(\mathcal{U}_t)$ . Assume, for the sake of contradiction, that  $X \notin \mathbf{V}(\mathcal{U}_t)$ , then the restriction subset of  $D_X$  on  $\mathcal{U}_t$  given  $\mathcal{G}^*$  is empty, as every DCC in  $D_X$  has  $X$  as its tail or head. Thus,  $\mathcal{U}_t$  is not a maximal orientation component for  $\kappa_t$  with respect to  $D_X$  and  $\mathcal{G}^*$ , since  $\mathcal{U}_t$  is chordal and any connected chordal graph has at least two simplicial vertices. This leads to a contradiction.

Recall that  $\mathcal{U}_t$  is connected, hence  $X \in \mathbf{V}(\mathcal{U}_t)$  implies that  $adj(X, \mathcal{U}_t) \neq \emptyset$ . The remaining proof is quite lengthy. We will consider two subcases:  $adj(X, \mathcal{U}_t) \subseteq pa(X, \mathcal{H}) \cup \mathbf{S} = pa(X, \mathcal{C}^*)$  and  $adj(X, \mathcal{U}_t) \not\subseteq pa(X, \mathcal{H}) \cup \mathbf{S} = pa(X, \mathcal{C}^*)$ , and show that in both subcases statement (ii) does not hold. That is, the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\mathbf{M}_T)$  given  $\mathcal{G}^*$  is not consistent with  $\mathcal{G}^*(\mathbf{M}_T)$  for some  $\mathbf{M}_T$ . This completes the proof for case 1.

**Case 1-1.** If  $adj(X, \mathcal{U}_t) \subseteq pa(X, \mathcal{H}) \cup \mathbf{S} = pa(X, \mathcal{C}^*)$ , then  $adj(X, \mathcal{U}_t)$  is a clique in  $\mathcal{U}_t$  based on (P2). Thus,  $X$  is a simplicial vertex in  $\mathcal{U}_t$ . On the other hand,  $adj(X, \mathcal{U}_t) \subseteq pa(X, \mathcal{H}) \cup \mathbf{S}$  implies that  $D_X(\mathcal{U}_t) = \{u \rightarrow X \mid u \in adj(X, \mathcal{U}_t)\}$ , hence  $X$  is a potential leaf node in  $\mathcal{U}_t$  with respect to  $D_X$  and  $\mathcal{G}^*$ . By the definition of  $\mathcal{U}_t$ , it holds that  $X = \kappa_t$ . Thus,  $\{X\} \cup \kappa_h \subseteq \{X\} \cup pa(X, \mathcal{H}) \cup \mathbf{S} \subseteq \mathbf{M}_T$  for some maximal clique  $\mathbf{M}_T$  of  $\mathcal{G}^*$  containing  $\mathbf{T}$ . Recall that  $\kappa \in \mathcal{K}_c \subseteq \mathcal{K}$ , the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\mathbf{M}_T)$  given  $\mathcal{G}^*$  contains  $X \xrightarrow{\text{or}} \kappa_h$  and  $\kappa_h \rightarrow X$  both, which is not consistent with  $\mathcal{G}^*(\mathbf{M}_T)$ .

**Case 1-2.** If  $adj(X, \mathcal{U}_t) \not\subseteq pa(X, \mathcal{H}) \cup \mathbf{S} = pa(X, \mathcal{C}^*)$ , then  $adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*) \neq \emptyset$ . We first claim that,

(P4) none of the vertices in  $\mathbf{V}(\mathcal{U}_t) \setminus (\{\kappa_t, X\} \cup pa(X, \mathcal{C}^*))$  (possibly empty) is simplicial in  $\mathcal{U}_t$ .

In fact, with respect to  $D_X$  and  $\mathcal{G}^*$ , only  $X$  and the vertices in  $adj(X, \mathcal{U}_t) \cap pa(X, \mathcal{C}^*)$  (possibly empty) can be the tails of DCCs in  $D_X(\mathcal{U}_t)$ . By the definition of a potential leaf node and  $\mathcal{U}_t$ , the vertices in  $\mathbf{V}(\mathcal{U}_t) \setminus (\{\kappa_t, X\} \cup pa(X, \mathcal{C}^*))$  are non-simplicial as they are not potential leaf nodes in  $\mathcal{U}_t$  with respect to  $D_X$  and  $\mathcal{G}^*$ .

Now, denote by  $\mathbf{L}$  the set of potential leaf nodes of  $\mathcal{U}_t$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ , then  $\mathbf{L} \subseteq \{\kappa_t, X\} \cup pa(X, \mathcal{C}^*)$  based on (P4) as the vertices in  $\mathbf{L}$  are simplicial in  $\mathcal{U}_t$ . However, as  $\kappa_h \subseteq pa(\kappa_t, \mathcal{C}^*) = adj(\kappa_t, \mathcal{U}_t)$ , if  $X = \kappa_t$ , then  $adj(X, \mathcal{U}_t) = pa(X, \mathcal{C}^*)$ , contradicted to the

assumption of case 1-2. Hence,  $X \neq \kappa_t$ . Since  $\kappa \in \mathcal{K}_c(\mathcal{U}_t)$ ,  $\kappa_t$  is not a potential leaf node in  $\mathcal{U}_t$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . Thus, it holds that,

- (P5) the set of potential leaf nodes of  $\mathcal{U}_t$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ , denoted by  $\mathbf{L}$ , satisfies that  $\mathbf{L} \subseteq \{X\} \cup pa(X, \mathcal{C}^*)$ .

**Case 1-2-1.** If  $X$  is simplicial in  $\mathcal{U}_t$ , then  $adj(X, \mathcal{U}_t)$  is a clique in  $\mathcal{U}_t$ . Recall that  $adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*) \neq \emptyset$ , we next consider two possibilities depends on whether a vertex in  $adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*)$  is simplicial.

First, suppose that there is a  $c \in adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*)$  which is simplicial in  $\mathcal{U}_t$ , then by (P4),  $c \notin \mathbf{V}(\mathcal{U}_t) \setminus (\{\kappa_t, X\} \cup pa(X, \mathcal{C}^*))$ , indicating that  $c = \kappa_t$ . Thus,  $\kappa_t \in adj(X, \mathcal{U}_t)$ .

- (i)  $adj(X, \mathcal{U}_t) = \{\kappa_t\}$ . We claim that this is an impossible case. In fact, it holds that  $adj(\kappa_t, \mathcal{U}_t) = \{X\}$ , since otherwise the vertices in  $adj(\kappa_t, \mathcal{U}_t)$  are adjacent to  $X$  due to the simplicity of  $\kappa_t$ , which contradicts  $|adj(X, \mathcal{U}_t)| = 1$ . Thus,  $\kappa_h = \{X\}$  and  $\kappa_t \rightarrow X$  is in  $\mathcal{H}$ , which contradicts the assumption that  $\kappa_t \in ch(X, \mathcal{C}^*) = ch(X, \mathcal{H}) \cup sib(X, \mathcal{H}) \setminus \mathbf{S}$ .
- (ii)  $adj(X, \mathcal{U}_t) \setminus \{\kappa_t\} \neq \emptyset$  and  $adj(X, \mathcal{U}_t) \setminus \{\kappa_t\} = pa(X, \mathcal{C}^*)$ . Since  $\kappa_t, X$  are both simplicial in  $\mathcal{U}_t$ ,  $\kappa_h \subseteq adj(\kappa_t, \mathcal{U}_t) = \{X\} \cup adj(X, \mathcal{U}_t) \setminus \{\kappa_t\} = \{X\} \cup pa(X, \mathcal{C}^*) = \mathbf{T}$ . Hence,  $\{\kappa_t\} \cup \mathbf{T}$  is a clique in  $\mathcal{G}^*$ . It is then straightforward to check that the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\mathbf{M}_{\mathbf{T}})$  given  $\mathcal{G}^*$  is not consistent with  $\mathcal{G}^*(\mathbf{M}_{\mathbf{T}})$  for any  $\mathbf{M}_{\mathbf{T}}$  containing  $\{\kappa_t\} \cup \mathbf{T}$ .
- (iii)  $adj(X, \mathcal{U}_t) \setminus \{\kappa_t\} \neq \emptyset$  and  $adj(X, \mathcal{U}_t) \setminus \{\kappa_t\} \subsetneq pa(X, \mathcal{C}^*)$ . Below we show that this is also an impossible case. For any  $p \in pa(X, \mathcal{C}^*)$  which is not in  $\mathcal{U}_t$ ,  $p$  is not adjacent to  $\kappa_t$ , since otherwise  $p \rightarrow \kappa_t$  is in  $\mathcal{C}^*$  by Rule 1 of Meek's rules and  $p$  should be included in  $\mathcal{U}_t$ . Denote by  $\mathcal{T}$  the induced subgraph of  $\mathcal{G}^*$  over  $\{\kappa_t, X\} \cup adj(\kappa_t, \mathcal{U}_t)$ . As  $adj(\kappa_t, \mathcal{U}_t) \subseteq \{X\} \cup pa(X, \mathcal{C}^*)$  and  $\{X\} \cup pa(X, \mathcal{C}^*)$  induces a complete subgraph of  $\mathcal{G}^*$ ,  $p \in pa(X, \mathcal{C}^*)$  is adjacent to every vertex in  $adj(\kappa_t, \mathcal{U}_t)$ . Therefore,  $p$  and  $\kappa_t$  are all and only simplicial vertices in  $\mathcal{T}$ . Moreover, due to the consistency of  $\mathcal{K}_c$  and the fact that  $\kappa \in \mathcal{K}_c$ ,  $\mathcal{T}$  is an orientation component for  $p$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . Therefore,  $adj(\kappa_t, \mathcal{U}_t) \rightarrow p$  are in  $\mathcal{H}$ . In particular,  $X \rightarrow p$  is in  $\mathcal{H}$ , which is contrary to the assumption that  $p \in pa(X, \mathcal{C}^*) \subseteq pa(X, \mathcal{H}) \cup sib(X, \mathcal{H})$ .
- (iv)  $adj(X, \mathcal{U}_t) \setminus \{\kappa_t\} \neq \emptyset$  and  $adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*) \setminus \{\kappa_t\} \neq \emptyset$ . This case is not possible either. Notice that none of the vertices in  $adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*) \setminus \{\kappa_t\}$  is simplicial in  $\mathcal{U}_t$ , we have that  $\mathcal{U}_t$  is not complete, and thus  $\mathcal{U}_t$  must have two non-adjacent simplicial vertices. Since  $X$  and  $\kappa_t$  are two adjacent simplicial vertices in  $\mathcal{U}_t$ , there must exist another simplicial vertex  $w$  in  $\mathcal{U}_t$  which is not adjacent to  $X$  and not adjacent to  $\kappa_t$ . However, as  $w \neq X$  and  $w$  is not adjacent to  $X$ ,  $w$  is not the tail of any DCC in  $D_X$ . It implies that  $w \neq \kappa_t$  is a potential leaf node of  $\mathcal{U}_t$  with respect to  $D_X$  and  $\mathcal{G}^*$ . This leads to a contradiction since we assume that  $\mathcal{U}_t$  is the maximal orientation component for  $\kappa_t$  with respect to  $D_X$  and  $\mathcal{G}^*$ .

Next, suppose that none of the vertices in  $adj(X, \mathcal{U}_t) \cap ch(X, \mathcal{C}^*)$  is simplicial in  $\mathcal{U}_t$ . Denote by  $\mathbf{R}$  (possibly empty) the set of simplicial vertices in  $\mathcal{U}_t$  which are adjacent to  $X$ . Let  $\mathcal{T}$  be the induced subgraph of  $\mathcal{U}_t$  over  $\mathbf{V}(\mathcal{U}_t) \setminus \mathbf{R}$ . We will show that,



(P6)  $\mathcal{T}$  is an orientation component for  $X$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ .

Since  $X$  is a simplicial node in  $\mathcal{U}_t$ ,  $X$  must be simplicial in  $\mathcal{T}$ . Hence, by Theorem 25 and the consistency of  $\mathcal{K}_c$ , it suffices to show that for any  $Y \in \mathbf{V}(\mathcal{T}) \setminus \{X\}$ ,  $Y$  is not a potential leaf node in  $\mathcal{T}$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . The proof consists of two claims:

- (i) For any  $Y \in \mathbf{V}(\mathcal{T}) \setminus \{X\}$ ,  $Y$  is not simplicial in  $\mathcal{U}_t$  implies that  $Y$  is not simplicial in  $\mathcal{T}$ . Suppose that there exists a  $Y \in \mathbf{V}(\mathcal{T}) \setminus \{X\}$  which is not simplicial in  $\mathcal{U}_t$  but simplicial in  $\mathcal{T}$ , then  $\text{adj}(Y, \mathcal{U}_t) \setminus \text{adj}(Y, \mathcal{T}) \neq \emptyset$ . This implies that  $\mathbf{R} \neq \emptyset$  as  $\text{adj}(Y, \mathcal{U}_t) \setminus \text{adj}(Y, \mathcal{T}) \subseteq \mathbf{V}(\mathcal{U}_t) \setminus \mathbf{V}(\mathcal{T}) = \mathbf{R}$ . On the other hand, since every vertex in  $\mathbf{R}$  is simplicial in  $\mathcal{U}_t$ ,  $\mathbf{R} \subseteq \text{adj}(Y, \mathcal{U}_t)$ , meaning that  $\text{adj}(Y, \mathcal{U}_t) \setminus \text{adj}(Y, \mathcal{T}) = \mathbf{R}$ . Moreover,  $\mathbf{R} \subseteq \text{adj}(X, \mathcal{U}_t)$  implies that  $Y$  is also adjacent to  $X$ . Notice that  $\text{adj}(Y, \mathcal{T})$  is a clique and  $X \in \text{adj}(Y, \mathcal{T})$ ,  $\text{adj}(Y, \mathcal{U}_t)$  is also a clique, which contradicts our assumption.
- (ii) For any  $Y \in \mathbf{V}(\mathcal{T}) \setminus \{X\}$ ,  $Y$  is simplicial in  $\mathcal{U}_t$  implies that  $Y$  is the tail of some DCC in  $\mathcal{K}_c(\mathcal{T})$ . In fact, if  $Y \notin \mathbf{R} \cup \{X\}$  is simplicial in  $\mathcal{U}_t$ , then it is simplicial in  $\mathcal{T}$ , and there is a DCC in  $\mathcal{K}_c(\mathcal{U}_t)$  whose tail is  $Y$ , since otherwise  $Y$  is a potential leaf node in  $\mathcal{U}_t$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ , which means  $Y \in \text{adj}(X, \mathcal{U}_t) \cap \text{pa}(X, \mathcal{C}^*)$  and thus  $Y \in \mathbf{R}$ , according to (P5). If  $Y$  is not the tail of any DCC in  $\mathcal{K}_c(\mathcal{T})$ , then  $\text{adj}(Y, \mathcal{T}) \subsetneq \text{adj}(Y, \mathcal{U}_t)$ , meaning that  $\text{adj}(Y, \mathcal{U}_t) \cap \mathbf{R} \neq \emptyset$ . Since the vertices in  $\mathbf{R}$  are simplicial in  $\mathcal{U}_t$  and  $\mathbf{R} \subseteq \text{adj}(X, \mathcal{U}_t)$ ,  $Y$  is adjacent to  $X$  in  $\mathcal{U}_t$ . Finally, as  $Y$  is simplicial in  $\mathcal{U}_t$ , by the construction, we have  $Y \in \mathbf{R}$  and hence  $Y \notin \mathbf{V}(\mathcal{T})$ , which contradicts our assumption.

In conclusion,  $\mathcal{T}$  is an orientation component for  $X$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . Note that  $\text{adj}(X, \mathcal{U}_t) \cap \text{ch}(X, \mathcal{C}^*) \neq \emptyset$  and  $\text{adj}(X, \mathcal{U}_t) \cap \text{ch}(X, \mathcal{C}^*) \cap \mathbf{R} = \emptyset$ ,  $\text{adj}(X, \mathcal{U}_t) \cap \text{ch}(X, \mathcal{C}^*) \subseteq \text{adj}(X, \mathcal{T})$ . Hence,  $\text{adj}(X, \mathcal{U}_t) \cap \text{ch}(X, \mathcal{C}^*) \rightarrow X$  are in  $\mathcal{H}$ , which contradicts the assumption that  $\text{ch}(X, \mathcal{C}^*) \subseteq \text{ch}(X, \mathcal{H}) \cup \text{sib}(X, \mathcal{H}) \setminus \mathbf{S}$ .

**Case 1-2-2.** Suppose that  $X$  is not simplicial in  $\mathcal{U}_t$ . We first show that there is a  $\phi \in \text{adj}(X, \mathcal{U}_t) \cap \text{pa}(X, \mathcal{C}^*)$  which is a potential leaf node in  $\mathcal{U}_t$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . In fact, with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ ,  $\mathcal{U}_t$ , which is a connected undirected induced subgraph of  $\mathcal{G}^*$ , must have a potential leaf node, since  $\mathcal{K}_c \subseteq \mathcal{K}$  and  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ . By (P5), these potential leaf nodes are in  $\{X\} \cup \text{pa}(X, \mathcal{C}^*)$ . However, as  $X$  is not simplicial in  $\mathcal{U}_t$  by our assumption, the potential leaf nodes are all in  $\text{pa}(X, \mathcal{C}^*)$ .

Now let  $\phi \in \text{adj}(X, \mathcal{U}_t) \cap \text{pa}(X, \mathcal{C}^*)$  be a potential leaf node in  $\mathcal{U}_t$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . Denote by  $\mathbf{R}$  (possibly empty) the set of simplicial vertices in  $\mathcal{U}_t$  which are adjacent to  $\phi$ . Clearly,  $\{\phi\} \cup \mathbf{R}$  is a clique of  $\mathcal{U}_t$ . Let  $\mathcal{T}$  be the induced subgraph of  $\mathcal{U}_t$  over  $\mathbf{V}(\mathcal{U}_t) \setminus \mathbf{R}$ . By the similar argument given to prove (P6), we can prove that  $\mathcal{T}$  is an orientation component for  $\phi$  with respect to  $\mathcal{K}_c$  and  $\mathcal{G}^*$ . Note that  $X \in \text{adj}(\phi, \mathcal{U}_t)$  and  $X \notin \mathbf{R}$ ,  $X \in \text{adj}(\phi, \mathcal{T})$ . Hence,  $X \rightarrow \phi$  is in  $\mathcal{H}$ , which contradicts the assumption that  $\phi \in \text{pa}(X, \mathcal{H}) \cup \mathbf{S}$ .

**Case 2.** Suppose that  $\mathcal{K}_c = \emptyset$ , then there exists a connected undirected induced subgraph  $\mathcal{W}$  of  $\mathcal{C}^*$  such that  $\mathcal{W}$  has no potential lead node with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{C}^*$  (Lemma 70). Note that, since  $\mathcal{C}^*$  is a chain graph, for any DCC in  $\mathcal{K}_u$ , its heads and tail are in the same chain component of  $\mathcal{C}^*$ . As  $\mathcal{K}$  is consistent with  $\mathcal{G}^*$ ,  $\mathcal{K}_u \subseteq \mathcal{K}$  is also consistent with  $\mathcal{G}^*$ . Then, by Lemma 70, we can conclude that,

(P7)  $\mathcal{K}_u$  is consistent with  $\mathcal{G}^*$  given  $D_X$  because  $\mathcal{K}_u(\mathcal{C}^*) = \mathcal{K}_u = \mathcal{K}_u(\mathcal{G}^*)$  and any connected undirected induced subgraph of  $\mathcal{C}^*$  is also a connected undirected induced subgraph of  $\mathcal{G}^*$ .

Since  $\mathcal{W}$  is undirected and connected,  $\mathcal{W}$  must be an induced subgraph of some chain component of  $\mathcal{C}^*$ . Denote by  $\mathbf{A}$  the set of all potential leaf nodes of  $\mathcal{W}$  with respect to  $\mathcal{K}_u$  and  $\mathcal{C}^*$ . As  $\mathcal{K}_u$  is consistent with  $\mathcal{G}^*$  given  $D_X$ ,  $\mathbf{A} \neq \emptyset$ . Notice that,  $(\mathcal{K}_u \cup \mathcal{K}_r)(\mathcal{W} \mid \mathcal{C}^*) = \mathcal{K}_u(\mathcal{W} \mid \mathcal{C}^*) \cup \mathcal{K}_r(\mathcal{W} \mid \mathcal{C}^*)$ ,  $\mathcal{W}$  has no potential leaf node with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{C}^*$  implies that every vertex in  $\mathbf{A}$  is the tail of some DCC in  $\mathcal{K}_r(\mathcal{W} \mid \mathcal{C}^*)$ . That is,

(P8) for any  $a \in \mathbf{A}$  there is a DCC  $\kappa := (a \xrightarrow{\text{or}} \kappa_h) \in \mathcal{K}_r$  such that (i)  $\kappa_h \cap \text{sib}(a, \mathcal{C}^*) \neq \emptyset$  and  $\kappa_h \cap \text{sib}(a, \mathcal{C}^*) \subseteq \mathbf{V}(\mathcal{W})$ , (ii)  $\kappa_h \setminus \mathbf{V}(\mathcal{W}) \neq \emptyset$  and  $\kappa_h \setminus \mathbf{V}(\mathcal{W}) \subseteq \text{pa}(a, \mathcal{C}^*)$ , and (iii)  $\kappa_h \subseteq \mathbf{V}(\mathcal{W}) \cup \text{pa}(a, \mathcal{C}^*)$ .

Now consider  $\mathcal{G}^*$  and  $D_X$ . For any  $a \in \mathbf{A}$ , let  $\mathcal{U}_a$  be the maximal orientation component for  $a$  with respect to  $D_X$  in  $\mathcal{G}^*$ . It is clear that  $\text{pa}(a, \mathcal{C}^*) = \text{adj}(a, \mathcal{U}_a)$ . Since  $\mathcal{C}^*$  is a chain graph causal MPDAG and  $\mathbf{A} \subseteq \mathbf{V}(\mathcal{W})$  and  $\mathcal{W}$  is an induced subgraph of some chain component of  $\mathcal{C}^*$ , by Theorem 8,  $\text{pa}(a, \mathcal{C}^*) = \text{pa}(a', \mathcal{C}^*)$  for any  $a, a' \in \mathbf{A}$ . Thus,  $\text{adj}(a, \mathcal{U}_a) = \text{adj}(a', \mathcal{U}_{a'})$  for any  $a, a' \in \mathbf{A}$ .

On the other hand, following the same argument for case 1, it can be shown that  $X \in \mathbf{V}(\mathcal{U}_a)$  for any  $a \in \mathbf{A}$ . Moreover, if  $\text{adj}(X, \mathcal{U}_a) \cap \text{ch}(X, \mathcal{C}^*) = \emptyset$  for some  $a \in \mathbf{A}$ , then we can prove that  $X$  is a potential leaf node in  $\mathcal{U}_a$  with respect to  $D_X$  and  $\mathcal{G}^*$ , which means  $X = a$ , and consequently,  $X \in \mathbf{V}(\mathcal{W})$ . As  $\mathcal{W}$  is an induced subgraph of some chain component of  $\mathcal{C}^*$  but  $X$  has no siblings in  $\mathcal{C}^*$ ,  $\mathbf{V}(\mathcal{W}) = \{X\}$ . This is impossible since  $\mathbf{V}(\mathcal{W}) = \{X\}$  implies that  $\kappa_h \cap \text{sib}(a, \mathcal{C}^*) = \emptyset$ , contrary to (P8). Therefore, we have that,

(P9) for all  $a \in \mathbf{A}$ ,  $\text{adj}(X, \mathcal{U}_a) \cap \text{ch}(X, \mathcal{C}^*) \neq \emptyset$ .

Moreover, following the same argument for proving (P4), it can be shown that,

(P10) none of the vertices in  $\mathbf{V}(\mathcal{U}_a) \setminus (\{a, X\} \cup \text{pa}(X, \mathcal{C}^*))$  (possibly empty) is simplicial in  $\mathcal{U}_a$ .

The rest of the proof is similar to that for case 1-2. Let

$$\mathbf{F} = \mathbf{V}(\mathcal{W}) \bigcup_{a \in \mathbf{A}} \mathbf{V}(\mathcal{U}_a),$$

and

$$\mathbf{N} = \bigcup_{a \in \mathbf{A}} \mathbf{V}(\mathcal{U}_a) \setminus (\{X\} \cup \mathbf{A} \cup \text{pa}(X, \mathcal{C}^*)).$$

Denote by  $\mathcal{F}$  the induced subgraph of  $\mathcal{G}^*$  over  $\mathbf{F}$ . Firstly, since  $\mathbf{V}(\mathcal{U}_a) \subseteq \mathbf{F}$  and  $\mathbf{N} \subseteq \mathbf{F}$ , by (P10), it holds that

(P11) none of the vertices in  $\mathbf{N}$  is simplicial in  $\mathcal{F}$ .

Moreover, by the definition of  $\mathbf{A}$ , every vertex in  $\mathbf{V}(\mathcal{W}) \setminus \mathbf{A}$  is either non-simplicial in  $\mathcal{W}$ , or the tail of a DCC in  $\mathcal{K}_u(\mathcal{W} \mid \mathcal{C}^*)$ . Thus,

- (P12)  $w \in \mathbf{V}(\mathcal{W}) \setminus \mathbf{A}$  is non-simplicial in  $\mathcal{W}$  implies that  $w$  is non-simplicial in  $\mathcal{F}$ , and  $w \in \mathbf{V}(\mathcal{W}) \setminus \mathbf{A}$  is the tail of a DCC in  $\mathcal{K}_u(\mathcal{W} \mid \mathcal{C}^*)$  implies that  $w$  is also the tail of a DCC in  $\mathcal{K}_u(\mathcal{W} \mid \mathcal{G}^*)$  based on (P7).

Finally, for set  $\mathbf{A}$ , we have that,

- (P13) for any  $a \in \mathbf{A}$ ,  $a$  is simplicial in  $\mathcal{F}$ , and is also the tail of some DCC in  $\mathcal{K}_r(\mathcal{F} \mid \mathcal{G}^*)$ .

The first claim comes from the simplicity of  $a$  in  $\mathcal{W}$  as well as the fact that

$$\text{adj}(a, \mathcal{F}) = \text{adj}(a, \mathcal{W}) \bigcup_{a' \in \mathbf{A}} \text{adj}(a', \mathcal{U}_{a'}) = \text{adj}(a, \mathcal{W}) \cup \text{adj}(a, \mathcal{C}^*) = \text{adj}(a, \mathcal{W}) \cup \text{pa}(a, \mathcal{C}^*),$$

where the second equality holds because  $\text{adj}(a, \mathcal{U}_a) = \text{adj}(a', \mathcal{U}_{a'})$  for any  $a, a' \in \mathbf{A}$ , and the third equality holds because of the definition of  $\mathcal{U}_a$ . The second claim holds because of the above equation and (P8)-(iii).

Below we will consider two subcases depending on whether  $X$  is simplicial in  $\mathcal{F}$ .

**Case 2-1 (analogue to case 1-2-1).** If  $X$  is simplicial in  $\mathcal{F}$ , then  $\text{adj}(X, \mathcal{F})$  is a clique in  $\mathcal{F}$ . Recall that (P9) says that  $\text{adj}(X, \mathcal{U}_a) \cap \text{ch}(X, \mathcal{C}^*) \neq \emptyset$  for all  $a \in \mathbf{A}$ , we have  $\text{adj}(X, \mathcal{F}) \cap \text{ch}(X, \mathcal{C}^*) \neq \emptyset$  based on the definition of  $\mathcal{F}$ .

If there is a  $c \in \text{adj}(X, \mathcal{F}) \cap \text{ch}(X, \mathcal{C}^*)$  which is simplicial in  $\mathcal{F}$ , then  $c \in \mathbf{V}(\mathcal{W})$  based on (P11), (P12) and (P13). Since  $\mathcal{C}^*$  is a chain graph and  $\mathcal{W}$  is an induced subgraph of some chain component in  $\mathcal{C}^*$ ,  $X \rightarrow \mathbf{V}(\mathcal{W})$  in  $\mathcal{C}^*$ . That is,  $\mathbf{V}(\mathcal{W}) \subseteq \text{ch}(X, \mathcal{C}^*) \cap \text{adj}(X, \mathcal{F})$ . As  $X$  is simplicial,  $\mathbf{V}(\mathcal{W})$  is a clique in  $\mathcal{F}$ .

- (i)  $\text{adj}(X, \mathcal{F}) = \mathbf{V}(\mathcal{W})$ . If there is an  $a \in \mathbf{A} \subseteq \mathbf{V}(\mathcal{W})$  such that  $\text{pa}(a, \mathcal{C}^*) \setminus \{X\} \neq \emptyset$ , then the vertices in  $\text{pa}(a, \mathcal{C}^*) \setminus \{X\}$  are adjacent to  $X$ , since  $a$  is simplicial in  $\mathcal{F}$  and  $\text{pa}(a, \mathcal{C}^*) \subseteq \mathbf{V}(\mathcal{U}_a) \subseteq \mathbf{V}(\mathcal{F})$ . This contradicts  $\text{adj}(X, \mathcal{F}) = \mathbf{V}(\mathcal{W})$ , and thus,  $\text{pa}(a, \mathcal{C}^*) = \{X\}$ . Since  $\text{pa}(w, \mathcal{C}^*) = \text{pa}(w', \mathcal{C}^*)$  for any  $w, w' \in \mathbf{V}(\mathcal{W})$ , we have  $\text{pa}(\mathbf{V}(\mathcal{W}), \mathcal{C}^*) = \{X\}$ . It can be shown by (P12) and (P13) that the induced subgraph of  $\mathcal{G}^*$  over  $\{X\} \cup \mathbf{V}(\mathcal{W})$  is an orientation component for  $X$  with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{G}^*$ , thus  $\mathbf{V}(\mathcal{W}) \rightarrow X$  are in  $\mathcal{H}$ , which contradicts the assumption that  $\mathbf{V}(\mathcal{W}) \subseteq \text{ch}(X, \mathcal{C}^*) \subseteq \text{ch}(X, \mathcal{H}) \cup \text{sib}(X, \mathcal{H}) \setminus \mathbf{S}$ .
- (ii)  $\text{adj}(X, \mathcal{F}) \setminus \mathbf{V}(\mathcal{W}) = \text{pa}(X, \mathcal{C}^*)$ . We claim that the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\mathbf{M}_{\mathbf{T}})$  given  $\mathcal{G}^*$  is not consistent with  $\mathcal{G}^*(\mathbf{M}_{\mathbf{T}})$  for any  $\mathbf{M}_{\mathbf{T}}$  containing  $\mathbf{T} \cup \mathbf{V}(\mathcal{W})$ . In fact, every vertex in  $\mathbf{T} \cup \mathbf{V}(\mathcal{W}) = \{X\} \cup \text{pa}(X, \mathcal{C}^*) \cup \mathbf{V}(\mathcal{W})$  is the tail of some DCC in the restriction subset of  $\mathcal{K} \cup D_X$  on  $\mathcal{G}^*(\mathbf{T} \cup \mathbf{V}(\mathcal{W}))$  given  $\mathcal{G}^*$ , because (1)  $\text{pa}(X, \mathcal{C}^*) \rightarrow X$  and  $X \rightarrow \mathbf{V}(\mathcal{W})$  are in the restriction subset of  $D_X$  on  $\mathcal{G}^*(\mathbf{T} \cup \mathbf{V}(\mathcal{W}))$  given  $\mathcal{G}^*$ , (2) every vertex in  $\mathbf{V}(\mathcal{W}) \setminus \mathbf{A}$  is the tail of some DCC in the restriction subset of  $\mathcal{K}_u$  on  $\mathcal{G}^*(\mathbf{T} \cup \mathbf{V}(\mathcal{W}))$  given  $\mathcal{G}^*$  according to (P12), and (3) every vertex in  $\mathbf{A}$  is the tail of some DCC in the restriction subset of  $\mathcal{K}_r$  on  $\mathcal{G}^*(\mathbf{T} \cup \mathbf{V}(\mathcal{W}))$  given  $\mathcal{G}^*$  according to (P13) and (P8).
- (iii)  $\text{adj}(X, \mathcal{F}) \setminus \mathbf{V}(\mathcal{W}) \subsetneq \text{pa}(X, \mathcal{C}^*)$ . Let  $p \in \text{pa}(X, \mathcal{C}^*)$  such that  $p \notin \text{adj}(X, \mathcal{F})$ . It is clear that none of the vertices in  $\mathcal{W}$  is adjacent to  $p$ , since otherwise  $p \rightarrow \mathbf{V}(\mathcal{W})$  are in  $\mathcal{C}^*$ , and in particular,  $p \rightarrow \mathbf{A}$  are in  $\mathcal{C}^*$  and  $p$  should be included in  $\mathcal{F}$ . By the similar

argument for proving case 1-2-1 we can show that the induced subgraph of  $\mathcal{G}^*$  over  $\{p, X\} \cup \text{adj}(X, \mathcal{F})$  is an orientation component for  $p$  with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$ . In fact, (1)  $p$  is simplicial in the induced subgraph of  $\mathcal{G}^*$  over  $\{p, X\} \cup \text{adj}(X, \mathcal{F})$  since  $p$  is adjacent to the vertices in  $\{X\} \cup \text{adj}(X, \mathcal{F}) \setminus \mathbf{V}(\mathcal{W})$  and  $\{X\} \cup \text{adj}(X, \mathcal{F})$  is a clique, (2) every vertex in  $\{X\} \cup \text{adj}(X, \mathcal{F}) \setminus \mathbf{V}(\mathcal{W})$  is non-simplicial in the induced subgraph of  $\mathcal{G}^*$  over  $\{p, X\} \cup \text{adj}(X, \mathcal{F})$  since  $p$  is not adjacent to any  $w \in \mathbf{V}(\mathcal{W})$  but both  $p$  and  $w$  are neighbors of the vertices in  $\{X\} \cup \text{adj}(X, \mathcal{F}) \setminus \mathbf{V}(\mathcal{W})$ , (3) every vertex in  $\mathbf{V}(\mathcal{W}) \setminus \mathbf{A}$  is the tail of some DCC in  $\mathcal{K}_u(\mathcal{W} \mid \mathcal{G}^*)$ , and (4)  $\mathbf{V}(\mathcal{W}) \cup \text{pa}(a, \mathcal{C}^*) \subseteq \{p, X\} \cup \text{adj}(X, \mathcal{F})$  for every  $a \in \mathbf{A}$  (since the simplicity of  $a$  in  $\mathcal{F}$  implies that  $\text{pa}(a, \mathcal{F}) \subseteq \text{adj}(X, \mathcal{F})$ ) and (P8) implies that every  $a \in \mathbf{A}$  is the tail of some DCC in the restriction subset of  $\mathcal{K}_r$  on  $\mathcal{G}^*(\{p, X\} \cup \text{adj}(X, \mathcal{F}))$  given  $\mathcal{G}^*$ . Therefore,  $X \rightarrow p$  is in  $\mathcal{H}$ , which leads to a contradiction.

- (iv)  $\text{adj}(X, \mathcal{F}) \cap \text{ch}(X, \mathcal{C}^*) \setminus \mathbf{V}(\mathcal{W}) \neq \emptyset$ . Let  $c \in \text{adj}(X, \mathcal{F}) \setminus \mathbf{V}(\mathcal{W})$  such that  $c \in \text{ch}(X, \mathcal{C}^*)$ . It is clear that  $c \in \mathbf{N}$ , and consequently,  $c$  is not simplicial in  $\mathcal{F}$  by (P11). This indicates that  $\mathcal{F}$  is not complete, and thus, there is a simplicial vertex  $w$  in  $\mathcal{F}$  which is not adjacent to  $X$ . However, this is impossible since every simplicial vertex in  $\mathcal{F}$  should be in  $\mathbf{F} \setminus \mathbf{N} \subseteq \mathbf{V}(\mathcal{W}) \cup \text{pa}(X, \mathcal{C}^*) \cup \{X\}$ , which is either  $X$  or adjacent to  $X$ .

Now assume that none of the vertices in  $\text{adj}(X, \mathcal{F}) \cap \text{ch}(X, \mathcal{C}^*)$  is simplicial in  $\mathcal{F}$ . Denote by  $\mathbf{R}$  (possibly empty) the set of simplicial vertices in  $\mathcal{F}$  which are adjacent to  $X$ . Following the same argument for proving (P6), we can show that the induced subgraph of  $\mathcal{F}$  over  $\mathbf{F} \setminus \mathbf{R}$  is an orientation component for  $X$  with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{G}^*$ . Hence,  $\text{adj}(X, \mathcal{F}) \cap \text{ch}(X, \mathcal{C}^*) \rightarrow X$  are in  $\mathcal{H}$ , which leads to a contradiction.

**Case 2-2 (analogue to case 1-2-2).** Suppose that  $X$  is not simplicial in  $\mathcal{F}$ . Denote by  $\mathbf{L}$  the set of potential leaf nodes in  $\mathcal{F}$  with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{G}^*$ . Since  $\mathcal{K}_u \cup \mathcal{K}_r$  is consistent with  $\mathcal{G}^*$ ,  $\mathbf{L} \neq \emptyset$ . Based on (P11), (P12), (P13) and the definition of a potential leaf node,  $\mathbf{L} \subseteq \{X\} \cup \text{pa}(X, \mathcal{C}^*)$ . Moreover, since  $X$  is not simplicial in  $\mathcal{F}$ ,  $\mathbf{L} \subseteq \text{pa}(X, \mathcal{C}^*)$ . Therefore, there is a  $\phi \in \text{adj}(X, \mathcal{F}) \cap \text{pa}(X, \mathcal{C}^*)$  which is a potential leaf node in  $\mathcal{F}$  with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$ . Denote by  $\mathbf{R}$  (possibly empty) the set of simplicial vertices in  $\mathcal{F}$  which are adjacent to  $\phi$ . Following the same proof for case 1-2-2, it can be checked that the induced subgraph of  $\mathcal{F}$  over  $\mathbf{F} \setminus \mathbf{R}$  is an orientation component for  $\phi$  with respect to  $\mathcal{K}_u \cup \mathcal{K}_r$  and  $\mathcal{G}^*$ . Hence,  $X \rightarrow \phi$  is in  $\mathcal{H}$ , which leads to a contradiction.  $\blacksquare$

## Appendix C. Simulations

Some details of the simulations are presented in this section.

### C.1 Generating Chordal Graphs

To generate a chordal graph with  $n$  nodes and  $e$  edges, where  $e \geq n$ , we first randomly generate a connected undirected tree with  $n$  nodes using the following Algorithm 6. Algorithm 6 successively adds  $n - 1$  edges to a graph with  $n$  nodes but without any edge, and every edge Algorithm 6 adds except the first one makes a singleton node connect to the existing non-singleton connected component. Next, we sequentially and randomly add

---

**Algorithm 6** Generating a connected undirected tree.
 

---

**Require:**  $X_1, \dots, X_n$ , which are  $n$  nodes.

**Ensure:** A connected undirected tree over  $X_1, \dots, X_n$ .

- 1: Set  $Q = (X_1)$  and  $P = (X_2, \dots, X_n)$ .
  - 2: Set  $\mathcal{T}$  be a graph with nodes  $X_1, \dots, X_n$  and empty edge set.
  - 3: **for**  $i$  in  $2, \dots, n$ , **do**
  - 4:     Randomly sample a node  $q$  from  $Q$  and a node  $p$  from  $P$ .
  - 5:     Add the undirected edge  $q - p$  to the graph  $\mathcal{T}$ .
  - 6:     Adding  $p$  to  $Q$  and removing  $p$  from  $P$ .
  - 7: **end for**
  - 8: **return**  $\mathcal{T}$ .
- 

---

**Algorithm 7** Generating a chordal graph.
 

---

**Require:**  $X_1, \dots, X_n$ , which are  $n$  nodes, and  $e$  representing the number of edges.

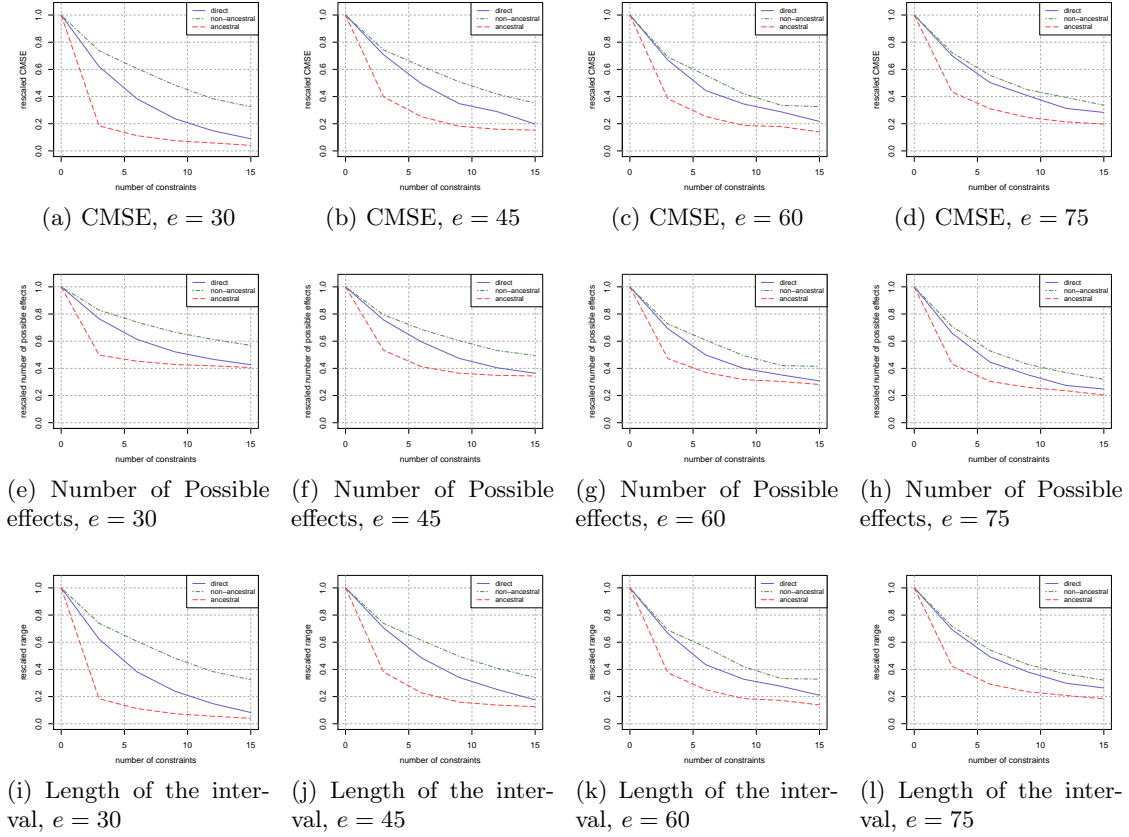
**Ensure:** A connected chordal graph over  $X_1, \dots, X_n$  with  $e$  edges.

- 1: Sample a connected undirected tree  $\mathcal{T}$  using Algorithm 6.
  - 2: Set  $r = e - n + 1$ .
  - 3: **while**  $r \neq 0$ , **do**
  - 4:     Randomly sample two non-adjacent nodes  $p$  and  $q$ .
  - 5:     Add the undirected edge  $q - p$  to  $\mathcal{T}$  and denote the resulting graph by  $\mathcal{T}_{\text{tmp}}$ .
  - 6:     **if**  $\mathcal{T}_{\text{tmp}}$  is chordal, **then**
  - 7:         Set  $r = r - 1$ .
  - 8:         Set  $\mathcal{T} = \mathcal{T}_{\text{tmp}}$ .
  - 9:     **end if**
  - 10: **end while**
  - 11: **return**  $\mathcal{T}$ .
- 

$e - n + 1$  undirected edges to  $\mathcal{T}$ . Every time an undirected edge is added to  $\mathcal{T}$ , we check whether the resulting graph is chordal, by calling `networkx.is_chordal` from the Python package `networkx`. If the resulting graph is chordal, we accept the added edge, otherwise we reject the added edge, re-sample an edge, and check the chordality again. The complete procedure is summarized in Algorithm 7. In our implementation, we keep track of the number of iterations of the while loop. If it exceeds a predefined maximum threshold, the loop is terminated and FAIL is returned. In such case, we restart the process to sample a new chordal graph.

## C.2 Additional Results

Figure 18 shows the results of three metrics in the settings where  $n = 30$ , where the metrics are the causal mean squared error (CMSE) introduced by (Tsirlis et al., 2018; Liu et al., 2020b), the number of possible causal effects, and the length of the interval determined by the minimum and maximum values of a set of possible effects. All scores are rescaled, as described in Section 5.3.

Figure 18: The results of the three metrics when  $n = 30$ .

The results of the above three metrics are similar, as shown in Figure 18. All scores decrease rapidly as the number of constraints increases. Moreover, for the same number of constraints, the scores of providing ancestral causal constraints are much lower than those of providing direct causal constraints, which in turn are significantly lower than those for non-ancestral constraints. This phenomenon arises because ancestral causal constraints are more informative than non-ancestral causal constraints. Specifically, knowing that  $X$  is a cause of  $Y$  implies that  $Y$  is not a cause of  $X$ , but the reverse implication does not hold. In conclusion, pairwise causal background knowledge—particularly ancestral causal background knowledge—can greatly enhance the identifiability of a causal effect.

## References

- Steen A. Andersson, David Madigan, and Michael D. Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *The Annals of Statistics*, 25(2):505–541, 1997.
- 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 *Proceedings of the Twenty-Third International Conference on Artificial Intelligence and Statistics*, volume 108, pages 4002–4011. PMLR, 2020.
- Jean R. S. Blair and Barry Peyton. An introduction to chordal graphs and clique trees. In *Graph Theory and Sparse Matrix Computation*, pages 1–29. Springer New York, 1993.
- Giorgos Borboudakis and Ioannis Tsamardinos. Incorporating causal prior knowledge as path-constraints in Bayesian networks and maximal ancestral graphs. In *Proceedings of the 29th International Conference on Machine Learning*, pages 1799–1806. Omnipress, 2012.
- Dorit Dor and Michael Tarsi. A simple algorithm to construct a consistent extension of a partially oriented graph. Technical Report R-185, Cognitive Systems Laboratory, UCLA, 1992.
- Zhuangyan Fang and Yangbo He. IDA with background knowledge. In *Proceedings of the Thirty-sixth Conference on Uncertainty in Artificial Intelligence*. PMLR, 2020.
- Zhuangyan Fang, Yue Liu, Zhi Geng, and Yangbo He. A local method for identifying causal relations under Markov equivalence. *Artificial Intelligence*, 305:103669, 2022.
- F. Richard Guo and Emilija Perković. Minimal enumeration of all possible total effects in a Markov equivalence class. In *Proceedings of the Twenty-Fourth International Conference on Artificial Intelligence and Statistics*, pages 2395–2403. PMLR, 2021.
- Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs. *Journal of Machine Learning Research*, 13(Aug):2409–2464, 2012.
- YangBo He and Zhi Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547, 2008.
- Leonard Henckel, Emilija Perković, and Marloes H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84:579–599, 2022.
- Steffen L Lauritzen and Thomas S Richardson. Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):321–348, 2002.
- Yue Liu, Zhuangyan Fang, Yangbo He, and Zhi Geng. Collapsible IDA: Collapsing parental sets for locally estimating possible causal effects. In *Proceedings of the Thirty-sixth Conference on Uncertainty in Artificial Intelligence*. PMLR, 2020a.

- Yue Liu, Zhuangyan Fang, Yangbo He, Zhi Geng, and Chunchen Liu. Local causal network learning for finding pairs of total and direct effects. *Journal of Machine Learning Research*, 21(148):1–37, 2020b.
- Marloes H. Maathuis, Markus Kalisch, and Peter Bühlmann. Estimating high-dimensional intervention effects from observational data. *The Annals of Statistics*, 37(6A):3133–3164, 2009.
- Daniel Malinsky and Peter Spirtes. Estimating bounds on causal effects in high-dimensional and possibly confounded systems. *International Journal of Approximate Reasoning*, 88: 371–384, 2017.
- Christopher Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 403–410. Morgan Kaufmann Publishers Inc., 1995.
- Preetam Nandy, Marloes H. Maathuis, and Thomas S. Richardson. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *The Annals of Statistics*, 45(2):647–674, 2017.
- Young Woong Park and Diego Klabjan. Bayesian network learning via topological order. *Journal of Machine Learning Research*, 18(1):3451–3482, 2017.
- Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.
- Judea Pearl. *Causality*. Cambridge University Press, 2009.
- Judea Pearl, Dan Geiger, and Thomas Verma. Conditional independence and its representations. *Kybernetika*, 25(7):33–44, 1989.
- Emilija Perković. Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Proceedings of the Thirty-sixth Conference on Uncertainty in Artificial Intelligence*. PMLR, 2020.
- Emilija Perković, Markus Kalisch, and Marloes H Maathuis. Interpreting and using CPDAGs with background knowledge. In *Proceedings of the Thirty-Third Conference on Uncertainty in Artificial Intelligence*. AUAI press, 2017.
- James Robins. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512, 1986.
- Andrea Rotnitzky and Ezequiel Smucler. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *Journal of Machine Learning Research*, 21(188):1–86, 2020.
- Peter Spirtes, Clark N Glymour, and Richard Scheines. *Causation, Prediction, and Search*. MIT Press, second edition, 2000.



- Konstantinos Tsirlis, Vincenzo Lagani, Sofia Triantafillou, and Ioannis Tsamardinos. On scoring maximal ancestral graphs with the max–min hill climbing algorithm. *International Journal of Approximate Reasoning*, 102:74–85, 2018.
- Pei-Li Wang and George Michailidis. Directed acyclic graph reconstruction leveraging prior partial ordering information. In *Machine Learning, Optimization, and Data Science: 5th International Conference, LOD 2019, Siena, Italy, September 10–13, 2019, Proceedings 5*, pages 458–471. Springer, 2019.
- Janine Witte, Leonard Henckel, Marloes H. Maathuis, and Vanessa Didelez. On efficient adjustment in causal graphs. *Journal of Machine Learning Research*, 21(246):1–45, 2020.
- Sewall Wright. Correlation and causation. *Journal of Agricultural Research*, 20(7):557–585, 1921.