On causal equivalence with tiered background knowledge (Supplementary material)

Christine W. Bang[1,2] **Vanessa Didelez**[1,2]

[1] Faculty of Mathematics and Computer Science, University of Bremen, Bremen, Germany
[2] Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany

A TERMINOLOGY

Nodes and edges. We define a *graph* $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ as a collection of *nodes* (or *vertices*) $\mathbf{V}$ and *edges* $\mathbf{E}$. Edges can be either *undirected* $A - B$ or *directed* $A \to B$. By $A \ast\!\!-\!\!\ast B$ we denote an arbitrary edge, i.e. this serves as a placeholder for either a directed or undirected edge. Two nodes $A, B \in \mathbf{V}$ are *adjacent* in $\mathcal{G}$ if $\{A \ast\!\!-\!\!\ast B\} \in \mathbf{E}$. No node can be adjacent to itself, and there can be at most one edge between any pair of nodes. We say that an edge of the form $A \to B$ is directed out of A (into B), and we then say that A is a *parent* of B. If there is an undirected edge between two nodes $A - B$, we say that A and B are *neighbours*. Let $A \in \mathbf{V}$ be a node in a graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, then $\mathrm{ne}_{\mathcal{G}}(A)$/$\mathrm{adj}_{\mathcal{G}}(A)$/$\mathrm{pa}_{\mathcal{G}}(A)$ is the set of neighbours/adjacent nodes/parents of A in $\mathcal{G}$. A graph is *complete* if all its nodes are adjacent. The *skeleton* of a graph is the undirected graph obtained by replacing its directed edges with undirected edges.

Subgraphs. We call $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$ a *subgraph* of $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ if $\mathbf{V}' \subseteq \mathbf{V}$ and $\mathbf{E}' \subseteq \mathbf{E}$. By $\mathcal{G}_u = (\mathbf{V}, \mathbf{E}_u)$ we denote the *undirected subgraph* of $\mathcal{G}$, where $\mathbf{E}_u$ is obtained from $\mathbf{E}$ by removing all directed edges. Correspondingly, $\mathcal{G}_d = (\mathbf{V}, \mathbf{E}_d)$ is the *directed subgraph* of $\mathcal{G}$, where $\mathbf{E}_d$ is obtained from $\mathbf{E}$ by removing all undirected edges. Let $\mathbf{A} \subseteq \mathbf{V}$, then the *induced subgraph* of $\mathcal{G}$ over $\mathbf{A}$ is $\mathcal{G}_{\mathbf{A}} = (\mathbf{A}, \mathbf{E}_{\mathbf{A}})$ where $\mathbf{E}_{\mathbf{A}} \subseteq \mathbf{E}$ contains all the edges between the nodes in $\mathbf{A}$.

Paths and cycles. A *path* $\pi = \langle V_1, V_2, ..., V_{K-1}, V_K\rangle$ from $V_1 \in \mathbf{V}$ to $V_K \in \mathbf{V}$ of length K consists of a sequence of distinct nodes where $V_i \in \mathrm{adj}(V_{i+1})$ for $1 \leq i < K$. A path from a set $\mathbf{A} \subseteq \mathbf{V}$ to another set $\mathbf{B} \subseteq \mathbf{V}$ is a path from some $A \in \mathbf{A}$ to some $B \in \mathbf{B}$. The *subpath* of π from V_i to V_j for $1 \leq i \leq j \leq K$ is $\pi(V_i, V_j) = \langle V_i, V_{i+1}, \ldots, V_{j-1}, V_j\rangle$. Let $\mathcal{G}' = (\mathbf{V}, \mathbf{E}')$ be a graph with same skeleton as $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ but possibly $\mathbf{E}' \neq \mathbf{E}$, then for a path π in $\mathcal{G}$ its *corresponding path* in $\mathcal{G}'$ is the path π' in $\mathcal{G}'$ consisting of the same nodes as π. An *undirected path* consists only of undirected edges. A *directed path* from V_1 to V_K has all edges oriented towards V_K, i.e. $V_j \to V_{j+1}$ for all $1 \leq j < K$; then V_1 is an *ancestor* of V_K (V_K is a *descendant* of V_1). A path from V_1 to V_K that contains both directed and undirected edges with at least one edge $V_j \to V_{j+1}$ for some $1 \leq j < K$ directed towards B and no edge $V_j \leftarrow V_{j+1}$ for any $1 \leq j < K$ is a *partially directed path* from V_1 to V_K. An undirected (directed) path from V_1 to V_K combined with an undirected (directed) path from V_K to V_1 we call an *undirected (directed) cycle*. An undirected or partially directed path from V_1 to V_K combined with a directed or partially directed path from V_K to V_1 we call a *partially directed cycle*.

(Partially) directed acyclic graphs. A graph consisting of only undirected edges is an *undirected graph*. An undirected graph is *chordal* if every cycle of length ≥ 4 has an adjacent pair of non-consecutive nodes. A *directed acyclic graph* (DAG) is a graph containing only directed edges and no directed cycles. A partially directed acyclic graph (PDAG) is a graph containing both directed and undirected edges and no directed cycles; DAGs and undirected graphs are special cases of PDAGs. A *chain graph* is a PDAG that does not have any partially directed cycles. The *chain components* of a chain graph are the undirected subgraphs.

Colliders, (un-) shielded and v-structures. We call a triple $\langle A, B, C\rangle$ *unshielded* if A and B are adjacent, B and C are adjacent, and A and C are not adjacent. We call a path unshielded if all triples on the path are unshielded. If a triple of the form $A \to B \leftarrow C$ occurs, we call B a *collider*, and if the triple is unshielded we call it a *v-structure*.

d-separation

Definition A.1 (d-connecting). *Let π be a path in some PDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, and let $\mathbf{C} \subset \mathbf{V}$. If (i) every collider V on π, or a descendant of V, is in $\mathbf{C}$, and (ii) no non-collider on π is in $\mathbf{C}$, then π is d-connecting given $\mathbf{C}$.*

If there exists a path from a set of nodes $\mathbf{A}$ to another set of nodes $\mathbf{B}$, where $\mathbf{A} \cap \mathbf{B} = \emptyset$, that is d-connecting given $\mathbf{C}$,

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

we say that $\mathbf{A}$ and $\mathbf{B}$ are *d-connected* given $\mathbf{C}$. If no such path exists, we say that $\mathbf{A}$ and $\mathbf{B}$ are *d-separated* given $\mathbf{C}$, and we denote this by

$$\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C}$$

We define an *independence model* $\mathcal{I}(\mathcal{G})$ induced by a graph $\mathcal{G}$ as the collection of all d-separations in $\mathcal{G}$:

$$(\mathbf{A} \perp_d \mathbf{B} \mid \mathbf{C}) \in \mathcal{I}(\mathcal{G}) \Leftrightarrow A \text{ and } B \text{ are d-sep. by } \mathbf{C} \text{ in } \mathcal{G}$$

Markov equivalence and CPDAGs. We say that two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are *Markov equivalent* if they induce the same independence model: $\mathcal{I}(\mathcal{G}_1) = \mathcal{I}(\mathcal{G}_2)$; an *equivalence class* is a class of Markov equivalent graphs. A *completed partially directed acyclic graph* (CPDAG) represents an equivalence class of DAGs, and can consist of undirected as well as directed edges: Undirected edges represent edges for which there exists at least one DAG in the equivalence class where the edge is oriented in one direction, and at least one DAG, where it is oriented in the opposite direction. Directed edges represent edges that must be identical in every DAG contained in the equivalence class. Two DAGs belong to the same equivalence class if and only if they have the same skeleton and the same v-structures [Verma and Pearl, 1990]. A graph is *maximally informative* if no additional edge can be oriented without restricting the equivalence class. A *restricted equivalence class* is a class of Markov equivalent graphs, that encode some additional common information. A *maximally oriented partially directed acyclic graph* (MPDAG) represents a restricted equivalence class.

B PREVIOUS RESULTS

B.1 MEEK'S RULES

An equivalence class of DAGs is uniquely characterised by the skeleton and v-structures [Verma and Pearl, 1990], but more directed edges might be shared among the DAGs in the class. Meek [1995] introduced a set of four orientation rules (Figure B.1), often referred to as *Meek's rules*, for which the graphical output will be maximally informative. Given the correct skeleton and v-structures of some equivalence class, repeated application of rules 1-3 outputs a CPDAG. Given the correct skeleton and v-structures, and additional background knowledge, repeated application of rules 1-4 outputs an MPDAG.

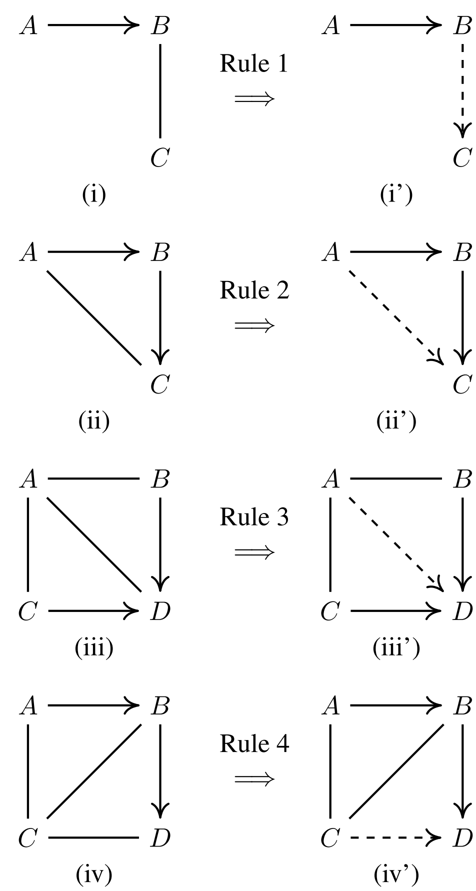

Figure B.1: Meek's rules. If (i), (ii), (iii) or (iv) occur as an induced subgraph of some PDAG, then orient them as (i'), (ii'), (iii') or (iv'), respectively.

B.2 ADJUSTMENT CRITERION

In a CPDAG $\mathcal{C} = (\mathbf{V}, \mathbf{E})$, a path $\pi = \langle V_1, \ldots, V_K \rangle$ is *possibly causal* from V_1 to V_K if it does not contain an edge $V_i \leftarrow V_{i+1}$ with $1 \leq i < K$. Otherwise it is *non-causal* from V_1 to V_K.

Definition B.1 (b-possibly causal [Perković et al., 2017]). *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be an MPDAG and let $\pi = \langle V_1, \dots, V_K \rangle$ be a path in $\mathcal{G}$. Then π is* b-possibly causal *from V_1 to V_K in $\mathcal{G}$ if and only if no edge $V_i \leftarrow V_j$, $1 \leq i < j \leq K$ is in $\mathcal{G}$. Otherwise, π is* b-non-causal path *in $\mathcal{G}$.*

B.3 IDA-ALGORITHM

Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a graph, and let $\mathbf{X} \subseteq \mathbf{V}$. Then we denote the set of parents of $\mathbf{X}$ in $\mathcal{G}$ by $\mathrm{pa}_{\mathcal{G}}(\mathbf{X}) = \bigcup_{X \in \mathbf{X}} \mathrm{pa}_{\mathcal{G}}(X)$.

Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be an MPDAG, let $X \in \mathbf{V}$ and let $\mathbf{S} \subseteq \mathrm{ne}_{\mathcal{G}}(X)$. Then $\mathcal{G}_{\mathbf{S} \to X}$ is the PDAG obtained by orienting all undirected edges $Z - X$ to $Z \to X$ if $Z \in \mathbf{S}$ and $Z \leftarrow X$ if $Z \in \mathrm{ne}_{\mathcal{G}}(X) \backslash \mathbf{S}$. A set of nodes $\mathbf{P} \subseteq \mathbf{V}$ is a *valid* (*jointly valid*) parent set of X ($\mathbf{X}$) if there exists a DAG $\mathcal{D}$ in the class represented by $\mathcal{G}$ for which $\mathrm{pa}_{\mathcal{D}}(X) = \mathbf{P}$ ($\mathrm{pa}_{\mathcal{D}}(\mathbf{X}) = \mathbf{P}$).

Algorithm 1: Locally obtaining valid parent sets from a tiered MPDAG using local IDA [Maathuis et al., 2009]

input : Tiered MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, node $X \in \mathbf{V}$
output : Multiset $\mathbf{PA}_{\mathcal{G}}^{\text{local}}(X)$

$\mathbf{PA}_{\mathcal{G}}^{\text{local}}(X) = \emptyset$
forall $\mathbf{S} \subseteq \mathrm{ne}_{\mathcal{G}}(X)$ **do**
  if $\mathcal{G}_{\mathbf{S} \to X}$ *has no new v-structure with X as collider* **then**
    add $\mathrm{pa}_{\mathcal{G}}(X) \cup \mathbf{S}$ to $\mathbf{PA}_{\mathcal{G}}^{\text{local}}(X)$
end

Algorithm 2: Semi-locally obtaining jointly valid parent sets from a tiered MPDAG using joint IDA [Nandy et al., 2017]

input : Tiered MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, set of nodes $\mathbf{X} \subseteq \mathbf{V}$, $\mathbf{X} = \{X_1, \dots, X_k\}$
output : Multiset $\mathbf{PA}_{\mathcal{G}}^{\text{joint}}(\mathbf{X})$

Obtain $\mathcal{G}_u$ and $\mathcal{G}_d$ from $\mathcal{G}$
Obtain the connected components of $\mathcal{G}_u$ that contain at least one node of $\mathbf{X}$: $\mathcal{G}_{u,1}, \dots, \mathcal{G}_{u,l}$ for $l \leq k$
for $i = 1, \dots, l$ **do**
  Let $\mathbf{PA}_i$ be the multiset of all jointly valid parent sets of the nodes of $\mathbf{X}$ in $\mathcal{G}_{u,i}$ obtained by constructing all DAGs in the (restricted) equivalence class represented by $\mathcal{G}_{u,i}$.
end
Construct $\mathbf{PA}_u$ by taking all possible combinations of $\mathbf{PA}_1, \dots, \mathbf{PA}_l$
$\mathbf{PA}_{\mathcal{G}}^{\text{joint}}(\mathbf{X}) = \{\mathbf{PA}_1' \cup \mathrm{pa}_{\mathcal{G}_d}(X_1), \dots, \mathbf{PA}_k' \cup \mathrm{pa}_{\mathcal{G}_d}(X_k)\}$
where $(\mathbf{PA}_1', \dots, \mathbf{PA}_k') \in \mathbf{PA}_u$.

C SIMULATION STUDY

Simulations were done in R version 4.2.1 using the `pcalg` package version 2.7-8, and random DAGs were simulated using the `randDAG` function. We simulated 8 different types of DAGs: The DAGs had either 10, 25, 50 or 100 nodes, and the structure was either dense or sparse. Sparse graphs had an expected number of adjacent nodes of 2, while dense graphs had an expected number of adjacent nodes of 5. Each DAG type was simulated three times, using either the Erdös-Rényi method, power-law method or geometric method.

We assumed that the full tiered ordering of the nodes assigned them to 5 tiers of equal size; hence, the tier size was either 2, 5, 10 or 20 depending on the number of nodes in the graph. We compared the full knowledge of the five tiers to four combinations of early or late, and more or less detailed knowledge. An overview of the tiered orderings can be found in Figure C.1. For each DAG, we constructed its CPDAG, and for each combination of DAG and tiered ordering τ_{full} (full knowledge), τ_{early1} (early simple), τ_{early2} (early detailed), τ_{late1} (late detailed) or τ_{late2} (late detailed), we constructed the tiered MPDAG. For each MPDAG, the number of additional directed edges compared to its corresponding CPDAG was counted. The above was repeated 1000 times for each combination of DAG type and simulation method; i.e. a total of 24,000 simulations.

The differences between the tiered MPDAGs and the corresponding CPDAGs are visualised in the boxplots in Figure C.2 and in Figure 5 in the main text. In Figure C.2 and Figure 5 we consider the number of new directed edges divided by the total number of edges in the graphs; the raw numbers are depicted in Figure C.3.

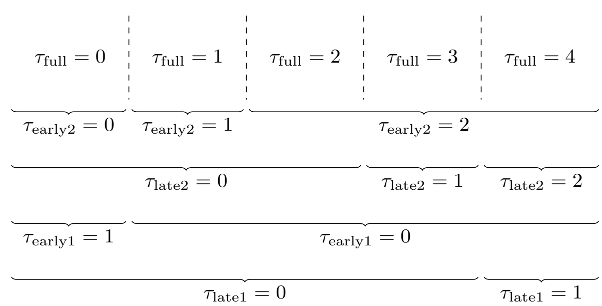

Figure C.1: Overview of the tiered orderings used for the simulation study. The tiered ordering τ_{full} is the full ordering of the nodes. The orderings τ_{early1} and τ_{late1} assign the nodes to two tiers: The main difference between these two is that τ_{early1} is able to distinguish the earliest tier, while τ_{early2} is able to distinguish the latest tier. The tiered orderings τ_{early2} and τ_{late2} assign the nodes to three tiers: While τ_{early2} contains knowledge of early tiers, τ_{late2} contains knowledge of later tiers.

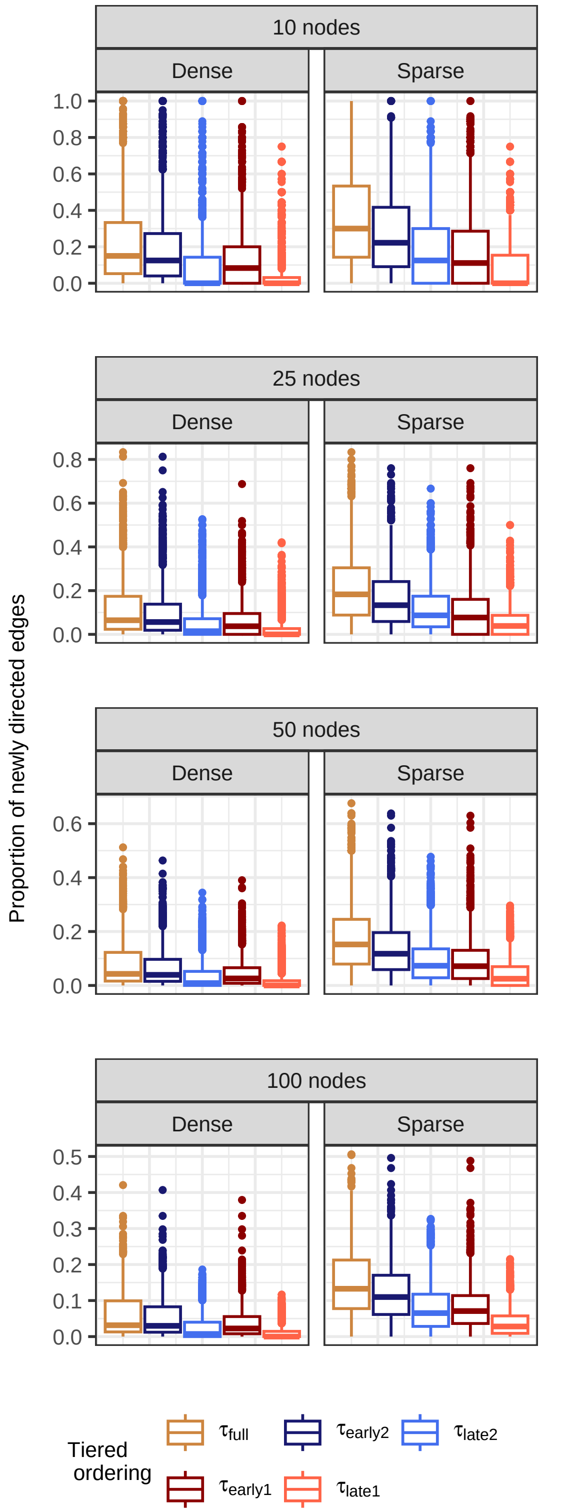

Figure C.2: Results of the simulation study. 24,000 random DAGs with 10, 25, 50 or 100 nodes were generated; half of them sparse, the other half dense. For each random DAG and each tiered ordering, the tiered MPDAG was constructed and the difference in number of directed edges to its corresponding CPDAG was computed and divided by the total number of edges.

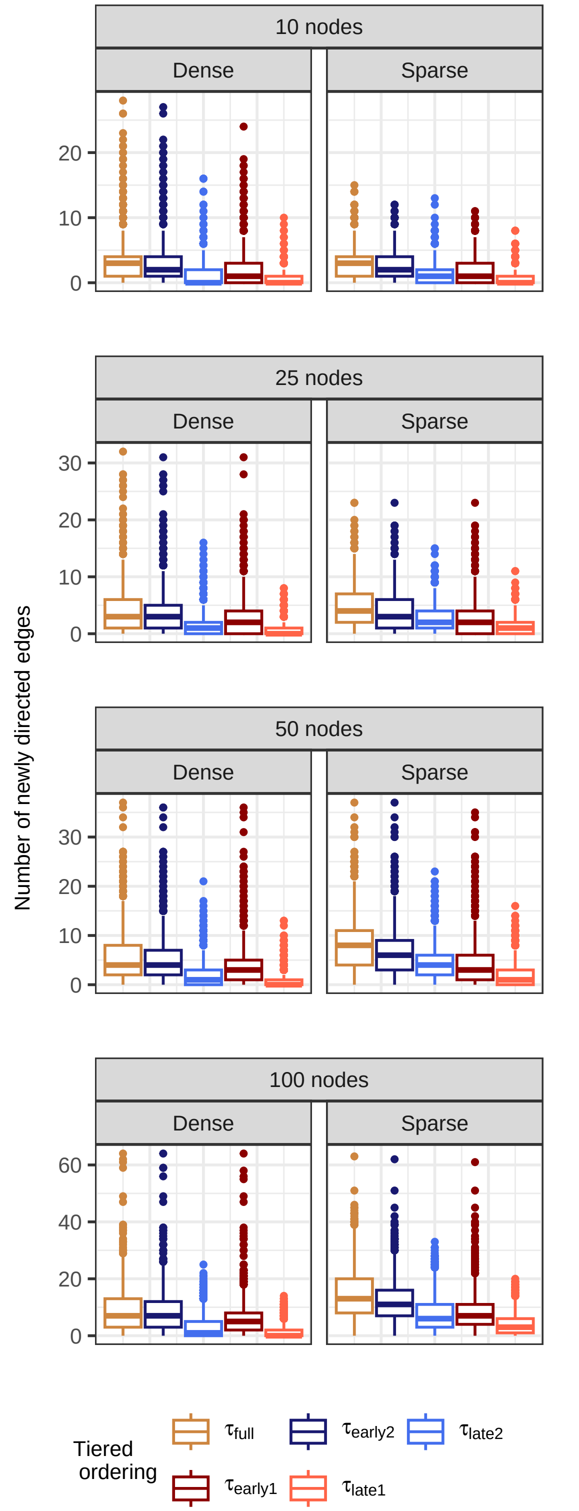

Figure C.3: Results of the simulation study. 24,000 random DAGs with 10, 25, 50 or 100 nodes were generated; half of them sparse, the other half dense. For each random DAG and each tiered ordering, the tiered MPDAG was constructed and the difference in number of directed edges to its corresponding CPDAG was computed.

D PROOFS FOR SECTION 3

D.1 PROOF OF LEMMA 1

Proof. Since the MPDAG is unambiguously defined by the equivalence class and tiered ordering, if $\mathcal{G}$ is an MPDAG, then by construction it is the MPDAG of $\mathcal{C}$ relative to τ. Hence, we need to show that it is in fact an MPDAG.

We proceed in two steps: (1) We show that in $\mathcal{C}^\tau$ an induced subgraph like in Figure B.1 (i) can occur, while no induced subgraphs like in Figures B.1 (ii)-(iv) can occur. (2) Let $\mathcal{C}^{\tau,n}$ be the graph obtained by applying Meek's 1st rule to $\mathcal{C}^\tau$ n times. We will show that Figure B.1 (i) can occur as an induced subgraph of $\mathcal{C}^{\tau,n}$, while Figures B.1 (ii)-(iv) cannot occur as induced subgraphs of $\mathcal{C}^{\tau,n}$. This means that the resulting graph $\mathcal{G}$ is maximally informative, and we conclude that it is an MPDAG.

(1) *Rule 1:* Assume that there is an undirected induced subgraph of $\mathcal{C}$ over $\{A, B, C\} \subseteq \mathbf{V}$ with adjacencies as in Figure B.1 (i). We can obtain a triple with orientations identical to Figure B.1 (i) in $\mathcal{C}^\tau$ if we have $A - B - C$ in $\mathcal{C}$ and $\tau(A) < \tau(B) = \tau(C)$; then τ will force $A \to B$.

Rule 2: Assume that there is an induced subgraph of $\mathcal{C}$ over $\{A, B, C\} \subseteq \mathbf{V}$ with adjacencies as in Figure B.1 (ii). Since $\mathcal{C}$ does not contain any partially directed cycles, this subgraph will have either 3, 2 or 0 directed edges. The case with 3 directed edges is not relevant, as well as any orientation of $\langle A, B, C\rangle$ other than $A \to B \to C$; however, the latter cannot occur in $\mathcal{C}$ since $\mathcal{C}$ is maximally informative. Hence, only an undirected subgraph in $\mathcal{C}$ allows for a subgraph like Figure B.1 (ii) in $\mathcal{C}^\tau$. If there are edges $A \to B$ and $B \to C$ in $\mathcal{C}^\tau$ they must have been forced by τ through $\tau(A) < \tau(B) < \tau(C)$. By transitivity this implies $\tau(A) < \tau(C)$, and $A \to C$ will be oriented by τ as well.

Rule 3: If B.1 (iii) is an induced subgraph of $\mathcal{C}^\tau$, then it is also an induced subgraph of $\mathcal{C}$, since the v-structure cannot be newly forced by τ. However, B.1 (iii) cannot be an induced subgraph of $\mathcal{C}$ since $\mathcal{C}$ is maximally informative.

Rule 4: Assume that there is an induced subgraph of $\mathcal{C}$ over $\{A, B, C, D\} \subseteq \mathbf{V}$ with adjacencies as in Figure B.1 (iv). For the case to be non-trivial, we exclude any subgraphs with other directed edges than $A \to B$ and $B \to D$; since $\mathcal{C}$ does not have any partially directed cycles, the subgraph must be undirected. If $A \to B \to D$ occurs in $\mathcal{C}^\tau$ it must be forced by τ through $\tau(A) < \tau(B) < \tau(D)$. Either $\tau(A) < \tau(C)$, $\tau(A) = \tau(C)$, or $\tau(A) > \tau(C)$. If $\tau(A) < \tau(C)$ or $\tau(A) > \tau(C)$ then it follows that $A \to C$ or $A \leftarrow C$ according to τ. If $\tau(A) = \tau(C)$, then by transitivity $\tau(C) < \tau(B) < \tau(D)$, and we orient $B \leftarrow C \to D$ according to τ.

(2) *Rule 1:* Assume that there is an undirected induced subgraph of $\mathcal{C}$ over $\{A, B, C\} \subseteq \mathbf{V}$ with adjacencies as in Figure B.1 (i). Assume that there is an undirected, unshielded path $\langle V_1, \ldots V_K = A\rangle$ of length $K > 1$ in $\mathcal{C}$ with $V_{K-1} \notin \mathrm{adj}_{\mathcal{C}}(B)$. Assume that $\tau(V_1) < \tau(V_2)$ such that $V_1 \to V_2$ in $\mathcal{C}^\tau$ and assume that $n \geq K - 1$: then $V_1 \to \ldots \to A \to B$ in $\mathcal{C}^{\tau,n}$, and we obtain B.1 (i).

Rule 2: Assume that there is an induced subgraph of $\mathcal{C}$ over $\{A, B, C\} \subseteq \mathbf{V}$ with adjacencies as in Figure B.1 (ii). By the same argument as above, only an undirected induced subgraph of $\mathcal{C}$ can lead to an induced subgraph like B.1 (ii) in $\mathcal{C}^{\tau,n}$. Moreover, by the argument above, we know that Figure B.1 (ii) does not occur as an induced subgraph of $\mathcal{C}^\tau$; hence, we consider the case where $\tau(A) = \tau(B) = \tau(C)$ and this subgraph is undirected. The only way that $A \to B$ can be directed in $\mathcal{C}^{\tau,n}$ and not in $\mathcal{C}^\tau$ is if there is an undirected unshielded path $\langle V_1, \ldots, V_K = A\rangle$ in $\mathcal{C}$ of length $K > 1$ in $\mathcal{C}$ with $V_{K-1} \notin \mathrm{adj}_{\mathcal{C}}(B)$ where $\tau(V_1) < \tau(V_2) = \tau(V_3) = \ldots = \tau(A)$ and $n \geq K - 1$ such that $V_1 \to \ldots \to A \to B$ in $\mathcal{C}^{\tau,n}$. In order for $A - C$ to remain undirected in $\mathcal{C}^{\tau,n}$, it must be the case that $V_{K-1} \in \mathrm{adj}_{\mathcal{C}}(C)$. If $V_{K-2} \notin \mathrm{adj}_{\mathcal{C}}(C)$ then $V_{K-1} \to C - B$ and $C \to B$ will be directed by Meek's 1st rule; hence, assume $V_{K-2} \in \mathrm{adj}_{\mathcal{C}}(C)$. Assume now that $V_j \in \mathrm{adj}_{\mathcal{C}}(C)$ for some $1 \leq j \leq K - 2$. Either (a) $V_{j-1} \notin \mathrm{adj}_{\mathcal{C}}(C)$ or (b) $V_{j-1} \in \mathrm{adj}_{\mathcal{C}}(C)$. (a) If $V_{j-1} \notin \mathrm{adj}_{\mathcal{C}}(C)$ then $V_j \to C - B$ occurs and it must then be the case that $V_j \in \mathrm{adj}_{\mathcal{C}}(B)$ in order for $C - B$ not to be directed as $C \to B$ or create a new v-structure, such that $B \to C$ would have been in $\mathcal{C}$. We then have $A \to B \ast\!\!-\!\!\ast V_j$: this cannot be a v-structure since then $A \to B$ would have been oriented in $\mathcal{C}$ and if $B \to V_j$ we would have had cycle; hence $V_j \in \mathrm{adj}_{\mathcal{C}}(A)$. Then $V_j \in \mathrm{adj}_{\mathcal{C}}(V_{K-1})$ since otherwise $V_{K-1} \to A \ast\!\!-\!\!\ast V_j$ would have been a v-structure or we would have had a cycle; by the same argument, $V_j \in \mathrm{adj}_{\mathcal{C}}(V_{K-2})$, and we can proceed until we obtain $V_j \in \mathrm{adj}_{\mathcal{C}}(V_{j+2})$, which is a contradiction. (b) Assume instead that $V_{j-1} \in \mathrm{adj}_{\mathcal{C}}(C)$ such that $V_j - C$ remains undirected. If $V_{j-2} \notin \mathrm{adj}_{\mathcal{C}}(C)$, we obtain a contradiction as above; hence, assume that $V_{j-2} \in \mathrm{adj}_{\mathcal{C}}(C)$. We can proceed with this until we obtain $V_1 \in \mathrm{adj}_{\mathcal{C}}(C)$. By transitivity, $\tau(V_1) < \tau(C)$ and we obtain $V_1 \to C - A$ in $\mathcal{C}^\tau$. In order to obtain $A - C$ in $\mathcal{C}^{\tau,n}$, we must have $V_1 \in \mathrm{adj}_{\mathcal{C}}(A)$. By the same reasoning as above, the path then cannot be unshielded, and we obtain a contradiction.

Rule 3: If B.1 (iii) is an induced subgraph of $\mathcal{C}^{\tau,n}$, then it is also an induced subgraph of $\mathcal{C}$, since the v-structure cannot be newly forced by Meek's 1st rule. However, B.1 (iii) cannot be an induced subgraph of $\mathcal{C}$ since $\mathcal{C}$ is maximally informative.

Rule 4: Consider the induced subgraph of $\mathcal{C}$ over $\{A, B, C, D\} \subseteq \mathbf{V}$ with adjacencies as in Figure B.1 (iv). By the same argument as above, only an undirected induced subgraph of $\mathcal{C}$ can lead to an induced subgraph like B.1 (iv) in $\mathcal{C}^{\tau,n}$. Moreover, by the argument above, we know that Figure B.1 (iv) does not occur as an induced subgraph of $\mathcal{C}^\tau$; hence, we consider the case where $\tau(A) = \tau(B) =$

$\tau(C) = \tau(D)$ and this subgraph is undirected. The only way that $A \to B$ can be directed in $\mathcal{C}^{\tau,n}$ and not in $\mathcal{C}^{\tau}$ is if there is an undirected unshielded path $\langle V_1, \ldots, V_K = A\rangle$ in $\mathcal{C}$ of length $K > 1$ with $V_{K-1} \notin \mathrm{adj}_{\mathcal{C}}(B)$. Assume that $\tau(V_1) < \tau(V_2) = \tau(V_3) = \ldots = \tau(A)$ such that $V_1 \to V_2$ in $\mathcal{C}^{\tau}$ and $n \geq K - 1$ applications of Meek's 1st rule results in $V_2 \to \ldots \to A \to B$ in $\mathcal{C}^{\tau,n}$. If $V_{K-1} \notin \mathrm{adj}_{\mathcal{C}}(C)$ then $A \to C$ will be forced by Meek's 1st rule. Hence, we assume that $V_{K-1} \in \mathrm{adj}_{\mathcal{C}}(C)$. To obtain Figure B.1 (iv) in $\mathcal{C}^{\tau,n}$ we require $C - B$ to be undirected; hence, we can proceed the in a similar way as for Rule 2 and obtain a contradiction. □

D.2 PROOF OF THEOREM 1

Proof. Assume that $\mathcal{C}$ is the CPDAG of which $\mathcal{G}$ is constructed, and τ the tiered ordering. Let $\mathcal{C}^{\tau}$ denote the graph obtained by orienting edges in $\mathcal{C}$ according to τ, and let $\mathcal{C}^{\tau,n}$ be the graph obtained by applying Meek's 1st rule to $\mathcal{C}^{\tau}$ n times. By Lemma 1, there exists an N such that for $n = N$ we have $\mathcal{G} = \mathcal{C}^{\tau,n}$; hence, we can without loss of generality assume $\mathcal{C}^{\tau,n}$ to be maximally informative. Since $\mathcal{C}$ does not contain any partially directed cycles, any partially directed cycle in $\mathcal{G}$ must be either (i) forced by τ, or (ii) forced by Meek's 1st rule. Hence, any partially directed cycle in $\mathcal{C}^{\tau}$ or $\mathcal{C}^{\tau,n}$ must correspond to an undirected cycle in $\mathcal{C}$: Let $\langle V_1, \ldots V_K\rangle$ combined with $V_1 - V_K$ be an undirected cycle in $\mathcal{C}$. We will show that (i) the corresponding cycle in $\mathcal{C}^{\tau}$ cannot be partially directed, and (ii) the corresponding cycle in $\mathcal{C}^{\tau,n}$ cannot be partially directed.

(i) Without loss of generality, assume that $\tau(V_1) < \tau(V_2)$ such that the edge $V_1 \to V_2$ is oriented in $\mathcal{C}^{\tau}$. If $\tau(V_1) < \tau(V_K)$ we will not obtain a partially directed cycle; therefore, assume that $\tau(V_K) \leq \tau(V_1)$. If for any $2 \leq i \leq K-1$: $\tau(V_i) > \tau(V_{i+1})$, again, it is no longer a partially directed cycle; therefore, assume $\tau(V_i) \leq \tau(V_{i+1})$ for all $2 \leq i \leq K - 1$. This then implies that $\tau(V_2) \leq \tau(V_K) \leq \tau(V_1)$. This is a contradiction to transitivity since we assumed $\tau(V_1) < \tau(V_2)$. We conclude that there cannot exist a partially directed cycle in $\mathcal{C}^{\tau}$.

(ii) By the above, there cannot be any partially directed cycles in $\mathcal{C}^{\tau}$; hence, if $\mathcal{C}^{\tau,n}$ contains a partially directed cycle, it must be forced through Meek's 1st rule; then $\tau(V_1) = \tau(V_2) = \ldots = \tau(V_K)$. Assume that there is an undirected unshielded path $\langle W_1, \ldots, W_m = V_1, V_2\rangle$ in $\mathcal{C}$, $m > 1$, with $\tau(W_1) < \tau(W_2) = \tau(W_3) = \ldots = \tau(V_1)$ such that $W_1 \to W_2$ in $\mathcal{C}^{\tau}$, and assume that $n \geq m-1$ such that $W_1 \to W_2 \to \ldots \to W_{m-1} \to V_1 \to V_2$ is in $\mathcal{C}^{\tau,n}$. If $W_{m-1} \notin \mathrm{adj}_{\mathcal{C}}(V_K)$ the edge $V_1 \to V_K$ follows from Meek's 1st rule and we no longer have a partially directed cycle; therefore, assume that $W_{m-1} \in \mathrm{adj}_{\mathcal{C}}(V_K)$. Either (a) $W_{m-2} \notin \mathrm{adj}_{\mathcal{C}}(V_K)$ or (b) $W_{m-2} \in \mathrm{adj}_{\mathcal{C}}(V_K)$. (a) In this case $W_{m-1} \to V_K$ by Meek's 1st rule. If $W_{m-1} \notin \mathrm{adj}_{\mathcal{C}}(V_{K-1})$, then $V_K \to V_{K-1}$ and we no longer have a partially directed cycle; assume $W_{m-1} \in \mathrm{adj}_{\mathcal{C}}(V_{K-1})$. We can then proceed until we obtain $W_{m-1} \in \mathrm{adj}_{\mathcal{C}}(V_2)$, which is a contradiction. (b) If $W_{m-3} \notin \mathrm{adj}_{\mathcal{C}}(V_K)$, then $W_{m-2} \to V_K$ by Meek's 1st rule, and we obtain a contradiction as above. Hence, assume $W_{m-3} \in \mathrm{adj}_{\mathcal{C}}(V_K)$. We can then proceed until we obtain $W_1 \in \mathrm{adj}_{\mathcal{C}}(V_K)$. By transitivity $\tau(W_1) < \tau(V_K)$ and the orientation $W_1 \to V_K$ is forced by τ. Assume that $W_1 \to V_i$ for some $2 < i \leq K$, then if $W_1 \notin \mathrm{adj}_{\mathcal{C}}(V_{i-1})$, then $V_i \to V_{i-1}$ and we no longer have a partially directed cycle. Hence, assume that $W_1 \in \mathrm{adj}_{\mathcal{C}}(V_{i-1})$ for all $2 < i \leq K$. Then $W_1 \in \mathrm{adj}_{\mathcal{C}}(V_2)$ and for $m = 2$ we have a contradiction. Assume $m > 2$, then $W_1 \in \mathrm{adj}_{\mathcal{C}}(V_1)$ since otherwise we would have either a cycle or a v-structure $W_1 \to V_2 \leftarrow V_1$, such that $V_1 \to V_2$ would have been oriented in $\mathcal{C}$. Then $W_1 \in \mathrm{adj}_{\mathcal{C}}(W_{m-1})$ since otherwise $W_{m-1} \to V_1$ would have been oriented in $\mathcal{C}$. We can proceed with this reasoning until we obtain $W_1 \in \mathrm{adj}_{\mathcal{C}}(W_3)$, which is a contradiction. □

D.3 PROOF OF COROLLARY 1

Proof. In order to show that $\mathcal{G}$ is a chain graph it is sufficient to show that it does not contain any partially directed cycles, which is the case due to Theorem 1. Hence, we only need to show that the chain components are chordal: Assume that $\mathcal{C}$ is the CPDAG from which $\mathcal{G}$ is constructed. Assume π is a chordless undirected cycle of length ≥ 4 in $\mathcal{G}$; then π must have been an undirected cycle in $\mathcal{C}$. Since $\mathcal{C}$ does not have any chordless undirected cycles, and since the procedure of orienting edges according to a tiered ordering or Meek's 1st rule does not delete edges or create partially directed cycles (c.f. Theorem 1), this is a contradiction. □

D.4 PROOF OF COROLLARY 2

The proof of Corollary 2 follows directly from the following result:

Corollary D.1. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a tiered MPDAG, and let $\pi = \langle V_1, \ldots, V_K\rangle$ be a path in $\mathcal{G}$. Then π is b-possibly causal from V_1 to V_K if and only if it is possibly causal from V_1 to V_K.*

Proof. "If" Assume that π is possibly causal from V_1 to V_K. Then there is no V_i, V_j on π with $i < j$ with $V_i \leftarrow V_j$ in $\mathcal{G}$, since otherwise $\langle V_i, \ldots, V_j\rangle$ combined with $\langle V_j, V_i\rangle$ would constitute a partially directed cycle in $\mathcal{G}$, which would be a contradiction to Theorem 1.

"Only if" Assume instead that π is not possibly causal from V_1 to V_K. Then there is an edge $V_i \leftarrow V_{i+1}$ for some $1 \leq i \leq k$ on π. Then $\mathcal{G}$ contains V_i, V_j on π with $i < j$ with $V_i \leftarrow V_j$ and no path in $\mathcal{G}$ is then b-possibly causal from V_1 to V_K; in particular, π is not b-possibly causal from V_1 to V_K. □

D.5 PROOF OF COROLLARY 3

The proofs of the validity of the output of the local IDA-algorithm and the joint IDA-algorithm rely on the fact that in a CPDAG, no orientation of the undirected edges can lead to a new v-structure, or a cycle, that includes an edge that is already directed in the CPDAG [Meek, 1995]. It is straightforward to show that the same is true for tiered MPDAGs:

Lemma D.1. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a tiered MPDAG, and let $\mathcal{G}_u$ and $\mathcal{G}_d$ be the undirected and the directed parts of $\mathcal{G}$ respectively. No orientation of the edges in $\mathcal{G}_u$ can create either (i) a v-structure in $\mathcal{G}$ that includes an edge in $\mathcal{G}_d$, or (ii) a cycle in $\mathcal{G}$ that includes an edge in $\mathcal{G}_d$.*

Proof. (i) By Lemma 1 we know that $\mathcal{G}$ is maximal relative to Meek's 1st rule; this implies that no unshielded triple of the form $X_i \to X_j - X_k$ can occur in $\mathcal{G}$.

(ii) Assume that we could orient the edges in $\mathcal{G}_u$ such that we would create a cycle in $\mathcal{G}$ including an edge from $\mathcal{G}_d$. This would require a cycle in $\mathcal{G}$ consisting of at least one directed part and at least one undirected part; however, this would constitute a partially directed cycle, which is a contradiction to Theorem 1. □

Proof of Corollary 3. We will first consider the joint IDA, and we follow the proof of Theorem 5.1 in Nandy et al. [2017]: Let $\mathcal{G}_{u,1}, \dots, \mathcal{G}_{u,n}$ denote the chain components of $\mathcal{G}_u$. Assume that only $\mathcal{G}_{u,1}, \dots, \mathcal{G}_{u,l}$ contain a node from $\mathbf{X}$. By Lemma D.1 we can orient each component $\mathcal{G}_{u,1}, \dots, \mathcal{G}_{u,l}$ into DAGs independently of the rest of the graph and obtain all valid parent sets from these. The multiplicity statement follows directly from Nandy et al. [2017].

We will now turn to the local IDA and we will follow the proof of Lemma 3.1 in Maathuis et al. [2009], which shows the following result: Let $X \in \mathbf{V}$ and let $\mathbf{S} \subset \mathrm{ne}_{\mathcal{G}}(X)$, then $\mathcal{G}_{\mathbf{S}\to X}$ does not create new v-structures with X as a collider if and only if there exists a DAG $\mathcal{D}$ in the (restricted) equivalence class represented by $\mathcal{G}$ for which $\mathrm{pa}_{\mathcal{D}}(X) = \mathrm{pa}_{\mathcal{G}}(X) \cup \mathbf{S}$. The "if" part is trivial, we show the "only if" part. As argued above, Lemma D.1 allows us to consider each connected component of $\mathcal{G}_u$ separately. Assume that X is in $\mathcal{G}_{u,i}$, we then need to show that we can orient $\mathcal{G}_{u,i}$ into a DAG without any new v-structures, where $\mathbf{S}$ is the parent set of X. In order to show that such an orientation exists, Maathuis et al. [2009] rely on two facts (1) the induced subgraph over $X \cup \mathbf{S}$ is complete, and (2) $\mathcal{G}_{u,i}$ is chordal. By Corollary 1 we know that (2) is satisfied. Since orienting edges from $\mathbf{S}$ into X does not create any new v-structures, all nodes in $\mathbf{S}$ must be adjacent in $\mathcal{G}$; since $\mathbf{S} \subseteq \mathrm{ne}_{\mathcal{G}}(X)$ it follows that the induced subgraph over $X \cup \mathbf{S}$ is complete. The rest follows from the proof of Lemma 3.1 in Maathuis et al. [2009]. □

E PROOFS FOR SECTION 4

E.1 PROOF OF THEOREM 2

Proof. We will make use of the following result: Let $\pi = \langle V_1, V_2, \dots, V_K \rangle$ be an unshielded path in $\mathcal{C}_u$, then π is unshielded in $\mathcal{C}$ as well: If for any subpath $V_{k-1} - V_k - V_{k+1}$ of π there were an edge $V_{k-1} *\!\!-\!\!* V_{k+1}$ in $\mathcal{C}$ that was not in $\mathcal{C}_u$, then this edge would be directed; combined with $V_{k-1} - V_k - V_{k+1}$ this would then create a partially directed cycle, which cannot occur in $\mathcal{C}$ since it is a CPDAG.

"Only if": (i): Assume that (i) is violated. Let $\pi_1 = \langle V_1, \dots, V_K \rangle$ be an unshielded path in $\mathcal{C}_u^{\tau_1}$ with $\pi_2 = \langle V_1, \dots, V_K \rangle$ being the corresponding path in $\mathcal{C}_u^{\tau_2}$, and assume that the first cross-tier edge on π_1 is not the same as the first cross-tier edge on π_2. Additionally, assume that π_1 and π_2 are both earliest.

Since π_1 and π_2 are unshielded and undirected, the corresponding paths in the underlying DAGs cannot contain colliders: They are either directed or they contain a subpath of the form $V_{k-1} \leftarrow V_k \to V_{k+1}$. In the latter case, either all cross-tier edges on π_1 will be on $\pi_1(V_1, V_k)$ or $\pi_1(V_k, V_K)$, or they will both contain cross-tier edges; similarly for π_2. It will then be sufficient to show that either $\pi_1(V_1, V_k) \neq \pi_2(V_1, V_k)$ or $\pi_1(V_k, V_K) \neq \pi_2(V_k, V_K)$. Moreover, since we assume all background knowledge to be correct, the paths must agree on the direction. Hence, we can without loss of generality assume that the corresponding paths in the underlying DAGs are directed from V_1 to V_K.

Assume that the first cross-tier edge on π_1 is $V_i \to V_{i+1}$ for $1 \leq i \leq K$, while the first cross-tier edge on π_2 is $V_j \to V_{j+1}$ with $i < j \leq K$. Let π_1' be the path in $\mathcal{G}_1$ corresponding to π_1, and let π_2' be the corresponding path in $\mathcal{G}_2$. Since only Meek's 1st rule applies (c.f. Lemma 1), the subpath $\pi_1'(V_1, V_i)$ will remain undirected since no new arrowheads are oriented into this subpath. Assume for contradiction that for some V_h with $1 \leq h \leq i-1$ there were a node $W \in \mathrm{adj}_{\mathcal{C}_u}(V_h)$ with $\tau_1(W) < \tau_1(V_h)$ such that $W \to V_h$ in $\mathcal{C}_u^{\tau_1}$. Then the path $\pi' = \langle W, V_h, V_{h+1}, \dots, V_K \rangle$ in $\mathcal{C}_u^{\tau_1}$ would be earlier than π_1, and π_1 would contain the subpath $\langle V_h, V_{h+1}, \dots, V_K \rangle$ of π', which is a contradiction since we assumed π_1 to be earliest. The subpath $\pi_1'(V_i, V_K)$ will be directed: $V_i \to V_{i+1}$ is forced by τ_1, and we will then be able to iteratively orient each node on $\langle V_{i+1}, \dots, V_K \rangle$ in the direction of V_K according to Meek's 1st rule when constructing $\mathcal{G}_1$, c.f. Lemma 1. Analogously, the subpath of $\pi_2'(V_1, V_j)$ is undirected, while the subpath $\pi_2'(V_j, V_K)$ is directed in $\mathcal{G}_2$. Hence, we have that $\pi_1'(V_i, V_j) \neq \pi_2'(V_i, V_j)$: It then follows that $\mathcal{G}_1 \neq \mathcal{G}_2$.

(ii): Assume that (ii) is violated. Let $V_i *\!\!-\!\!* V_j$ be an edge for which $\mathcal{C}_u^{\tau_1}$ and $\mathcal{C}_u^{\tau_2}$ disagree on whether it is directed or not. Since $V_i *\!\!-\!\!* V_j$ is only contained on shielded paths, it can only be oriented by background knowledge c.f. Lemma 1, since

Meek's 1st rule does not apply. It follows that $\mathcal{G}_1 \neq \mathcal{G}_2$.

"If": Since $\mathcal{G}_1$ and $\mathcal{G}_2$ are constructed from the same CPDAG, they will agree on every edge that is directed in $\mathcal{C}$; hence, we will consider $\mathcal{C}_u$. Assume that (i) and (ii) are both satisfied. By (ii) we know that $\mathcal{G}_1$ and $\mathcal{G}_2$ will agree on the orientation of any fully shielded edge, so we need to show that they will also agree on the orientation of any edge that is not fully shielded; we will consider the unshielded paths.

Let $\pi_1 = \langle V_1, \ldots, V_K \rangle$ be an unshielded path in $\mathcal{C}_u^{\tau_1}$ and let $\pi_2 = \langle V_1, \ldots, V_K \rangle$ be the corresponding path in $\mathcal{C}_u^{\tau_2}$. Assume that $\mathcal{D}_1 \in [\mathcal{C}]$ is a DAG giving rise to τ_1 and $\mathcal{D}_2 \in [\mathcal{C}]$ is a DAG giving rise to τ_2. By the same argument as above, we may assume that either (a) the corresponding paths in $\mathcal{D}_1$ and $\mathcal{D}_2$ are directed from V_1 to V_K, or (b) the corresponding path in $\mathcal{D}_1$ contains $V_{k-1} \leftarrow V_k \rightarrow V_{k+1}$ for some $2 \leq k \leq K-1$; i.e. the subpaths will be directed from V_k to V_1 and from V_k to V_K, and the corresponding path in $\mathcal{D}_2$ contains $V_{l-1} \leftarrow V_l \rightarrow V_{l+1}$ for some $2 \leq l \leq K-1$; i.e. the subpaths will be directed from V_l to V_1 and from V_l to V_K, or (c) the corresponding path in one DAG is directed from V_1 to V_K, and the corresponding path in the other DAG contains a subpath $V_{k-1} \leftarrow V_k \rightarrow V_{k+1}$ for some $2 \leq k \leq K-1$. Since (b) is the most general case, we will only consider this; (a) and (c) can be verified in a similar way.

Either $\pi_1(V_1, V_k)$ and $\pi_2(V_1, V_l)$ will have a cross-tier edge, $\pi_1(V_k, V_K)$ and $\pi_2(V_l, V_K)$ will have a cross-tier edge, or they will all have a cross-tier edge. We consider the most general case where they all have a cross-tier edge, and assume that the first cross-tier edge on $\pi_1(V_k, V_K)$ and $\pi_2(V_l, V_K)$ is $V_i \rightarrow V_{i+1}$ and that the first cross-tier edge on $\pi_1(V_1, V_k)$ and $\pi_2(V_1, V_l)$ is $V_j \rightarrow V_{j-1}$. Let π_1' be the path in $\mathcal{G}_1$ corresponding to π_1, and let π_2' be the corresponding path in $\mathcal{G}_2$. By similar arguments as above, it then follows that $\pi_1'(V_j, V_i) = \pi_2'(V_j, V_i)$ will remain undirected, $\pi_1'(V_1, V_j) = \pi_2'(V_1, V_j)$ will be directed from V_j to V_1, and $\pi_1'(V_i, V_K) = \pi_2'(V_i, V_K)$ will be directed from V_i to V_K. The case where π_1 and π_2 only have a single cross-tier edge is special case of this. Hence, $\pi_1' = \pi_2'$. □

E.2 PROOF OF COROLLARY 4

Proof. Let $\mathcal{G}_1$ be the MPDAG obtained from $\mathcal{C}$ relative to τ_1, and let $\mathcal{G}_2$ be the MPDAG obtained from $\mathcal{C}$ relative to τ_2. Assume that (i) and (ii) are satisfied. If $\mathcal{C}_u^{\tau_1}$ does not have any additional oriented edges, then $\mathcal{G}_1 = \mathcal{G}_2$ by Theorem 2.

Assume that (i), (ii), and (iii) are satisfied. Let $\pi_1 = \langle V_1, \ldots V_K \rangle$ be an earliest unshielded path in $\mathcal{C}_u^{\tau_1}$ and let $V_i \rightarrow V_{i+1}$ be the first cross-tier edge on π_1. Let π_1' be the corresponding path in $\mathcal{G}_1$. Then $\pi_1'(V_1, V_i)$ will be undirected and $\pi_1'(V_1, V_i)$ will be directed, by similar arguments as in the proof of Theorem 2. Let $\pi_2 = \langle V_1, \ldots V_K \rangle$ be the path in $\mathcal{C}_u^{\tau_2}$ corresponding to π_1 and assume that $V_i - V_{i+1}$ is not a cross-tier edge in $\mathcal{C}_u^{\tau_2}$. Let π_2' be the corresponding path in $\mathcal{G}_2$. Either π_2 will have at least one cross-tier edge, or it will have no cross-tier edges. If π_2 has no cross-tier edges, then π_2' will be undirected: Since π_1' will be directed from V_i to V_K, $\mathcal{G}_1$ will be contained in $\mathcal{G}_2$. Assume instead that π_2 has at least one cross-tier edge and that the first cross-tier edge is $V_j \rightarrow V_{j+1}$. Then by (i) this is also a cross-tier edge on π_1. Since $V_j \rightarrow V_{j+1}$ is not the first cross-tier edge on π_1 it follows that $i \leq j$; since $V_i \rightarrow V_{i+1}$ is not a cross-tier edge on π_2 we conclude that $i < j$. By similar arguments as in the proof of Theorem 2 we then know that $\pi_1'(V_1, V_i) = \pi_2'(V_1, V_i)$ are undirected, $\pi_1'(V_j, V_K) = \pi_2'(V_j, V_K)$ are directed, and $\pi_1'(V_i, V_j) \neq \pi_2'(V_i, V_j)$ since $\pi_1'(V_i, V_j)$ is directed and $\pi_2'(V_i, V_j)$ is undirected. Then $\mathcal{G}_1$ will be contained in $\mathcal{G}_2$ and τ_1 will be more informative than τ_2.

Assume that (i), (ii) and (iv) are satisfied. Following the proof of Theorem 2, the fully shielded edges can only be oriented by background knowledge and $\mathcal{G}_1$ will be contained in $\mathcal{G}_2$, and τ_1 will be more informative than τ_2.

Assume that (i), (ii), (iii) and (iv) are all satisfied. Then by the same arguments as above, $\mathcal{G}_1$ will be contained in $\mathcal{G}_2$, and τ_1 will be more informative than τ_2. □

References

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.

Christopher Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, pages 403–410, 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.

Emilija Perković, Markus Kalisch, and Maloes H Maathuis. Interpreting and using cpdags with background knowledge. *arXiv preprint arXiv:1707.02171*, 2017.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270, 1990.