Do we become wiser with time?
On causal equivalence with tiered background knowledge

Abstract

Equivalence classes of DAGs (represented by CPDAGs) may be too large to provide useful causal information. Here, we address incorporating *tiered* background knowledge yielding restricted equivalence classes represented by 'tiered MPDAGs'. Tiered knowledge leads to considerable gains in informativeness and computational efficiency: We show that construction of tiered MPDAGs only requires application of Meeks 1st rule, and that tiered MPDAGs (unlike general MPDAGs) are chain graphs with chordal components. This entails simplifications e.g. of determining valid adjustment sets for causal effect estimation. Further, we characterise when one tiered ordering is more informative than another, providing insights into useful aspects of background knowledge.

1 INTRODUCTION

We consider equivalence classes of DAGs represented by completed partially directed acyclic graphs (CPDAGs), occurring as outputs of causal discovery algorithms. A first characterisation of equivalent DAGs was given by Verma and Pearl [1990] and a full characterisation of CPDAGs by Andersson et al. [1997]. Often, domain expertise provides additional information about shared features of the graphs in a class. Restricting the equivalence class by background knowledge yields a type of partially directed acyclic graph (PDAG) that is potentially much more informative due to more induced edge orientations. Meek [1995] provided a set of orientation rules to obtain a graph encoding the maximal implied information, and the resulting graph is then a maximally oriented partially directed acyclic graph (MPDAG) [Perković et al., 2017]. While a CPDAG represents an independence model common to all DAGs in the equivalence class, an MPDAG represents an independence model as well as additional causal or directional information that is common to all DAGs in a restricted equivalence class. DAGs and CPDAGs are special cases of MPDAGs; DAGs are MPDAGs with full (or sufficient) background knowledge, while CPDAGs are MPDAGs with no (or redundant) background knowledge. A general characterisation of MPDAGs was given by Fang et al. [2022]. The interpretation of MPDAGs was described in detail by Perković et al. [2017] and is considerably more involved than that of CPDAGs.

Background knowledge can be induced by, e.g., well-established causal or logical relations. Some kinds of knowledge, e.g. temporal or sequential structures, imply that the nodes can be partitioned into ordered tiers. Tiered background knowledge is typically unambiguous and it seems intuitively obvious that it must be useful. Indeed, implementations of algorithms for constraint-based causal discovery with a given tiered structure exist (see, e.g., Scheines et al. [1998], Petersen et al. [2021], and Foraita et al. [2022]), but the general properties of these restricted model classes have not yet been investigated. Here, we provide the first formal in-depths analysis of equivalence classes restricted by tiered knowledge. We show that there are several desirable properties, distinguishing tiered from other kinds of background knowledge. Thus, we focus on MPDAGs arising from imposing a tiered ordering, which we term 'tiered MPDAGs'. We show that under the key properties of completeness and transitivity tiered background knowledge cannot induce partially directed cycles, and, moreover, that tiered MPDAGs are chain graphs with chordal chain components. This allows us to, e.g., apply common methods for identifying causal effects using CPDAGs to tiered MPDAGs without any additional processing.

When eliciting background knowledge, there may be more effort involved to achieve more detail. For instance, data from cohort studies are organised in a readily available tiered structure. But within the same wave it may be possible to further subdivide the nodes using logical, temporal or similar expertise; e.g. the first wave of a children's cohort may be composed of variables before and after birth, pertain-

Submitted to the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023). **To be used for reviewing only**.

ing to mother or child etc. It is therefore also interesting to characterise when different tiered restrictions are redundant versus when they are most informative.

In Section 2 we formalise the concepts of (tiered) background knowledge and restricted equivalence classes. We then provide a formal characterisation of tiered MPDAGs: Section 3 describes some of their properties, and Section 4 compares different tiered orderings in terms of informativeness. In Section 4.3, we illustrate which types of tiered knowledge are particularly informative via simulation. Section 5 addresses how tiered information structurally differs from other types of background knowledge. Throughout, we rely on standard notation for (causal) graphical models, an overview of relevant definitions can be found in Section A of the Supplement; all proofs can be found in Sections D and E of the Supplement.

2 BACKGROUND KNOWLEDGE

While the DAGs in an equivalence class have exactly the same conditional independencies, they can still have vastly different causal implications. In this section we introduce a smaller, and possibly more informative, subclass of causal graphs using background knowledge.

We define that *background knowledge* $\mathcal{K} = (\mathcal{R}, \mathcal{F})$ consists of a set of *required edges* $\mathcal{R}$ and a set of *forbidden edges* $\mathcal{F}$.

Definition 1. *A graph $\mathcal{G}$ encodes background knowledge $\mathcal{K} = (\mathcal{R}, \mathcal{F})$ if all of the edges in $\mathcal{R}$ and none of the edges in $\mathcal{F}$ are present in $\mathcal{G}$.*

2.1 RESTRICTED EQUIVALENCE CLASSES

In our setup we only consider correct background knowledge in the sense that it agrees with an underlying (unknown) true DAG; more specifically, we assume the given CPDAG and background knowledge to be consistent with each other.

Let $\mathcal{C}$ be a CPDAG, then by $[\mathcal{C}]$ we denote the equivalence class of DAGs represented by $\mathcal{C}$.

Definition 2 (Meek [1995])**.** *A CPDAG $\mathcal{C}$ and background knowledge $\mathcal{K} = (\mathcal{R}, \mathcal{F})$ are* consistent *if and only if there exists a DAG $\mathcal{D} \in [\mathcal{C}]$ such that all of the edges in $\mathcal{R}$ and none of the edges in $\mathcal{F}$ are in $\mathcal{D}$.*

Consistent background knowledge is imposed on a CPDAG by orienting the corresponding undirected edges. In turn, this may allow us to orient further undirected edges, e.g. to avoid directed cycles. Meek [1995] showed that maximal edge orientations implied by given background knowledge are obtained under a set of four orientation rules, also known as Meek's rules (see Figure B.1 in the Supplement). The resulting graph then no longer represents an equivalence class, but rather a *restricted equivalence class*.

Algorithm 1: Constructing $\mathcal{C}^{\mathcal{K}}$

input : CPDAG $\mathcal{C} = (\mathbf{V}, \mathbf{E})$ and consistent background knowledge $\mathcal{K} = (\mathcal{R}, \mathcal{F})$.
output : PDAG $\mathcal{C}^{\mathcal{K}} = (\mathbf{V}, \mathbf{E}')$

```
E' = E
forall {V_i - V_j} ∈ E do
    if {V_i → V_j} ∈ F then
        replace {V_i - V_j} with {V_i ← V_j} in E'
        else if {V_i → V_j} ∈ R then
            replace {V_i - V_j} with {V_i → V_j} in E'
end
```

More formally, the construction is as follows: Let $\mathcal{C} = (\mathbf{V}, \mathbf{E})$ be a CPDAG and let $\mathcal{K} = (\mathcal{R}, \mathcal{F})$ be background knowledge consistent with $\mathcal{C}$. First, orient edges in $\mathcal{C}$ according to $\mathcal{K}$ as in Algorithm 1, and let $\mathcal{C}^{\mathcal{K}}$ denote the PDAG obtained from this procedure. Second, orient additional edges by repeated application of Meek's rules 1-4 until no further change; let $\mathcal{G}$ denote the resulting PDAG. Then $\mathcal{G}$ is the *maximally oriented partially directed acyclic graph* (MPDAG) obtained from $\mathcal{C}$ relative to $\mathcal{K}$ [Meek, 1995].

For a given CPDAG $\mathcal{C}$ and background knowledge $\mathcal{K}$, the MPDAG obtained from $\mathcal{C}$ relative to $\mathcal{K}$ is unique. However, the origin of an MPDAG is not unique: Let $\mathcal{C}$ be a CPDAG and let $\mathcal{K}_1$ and $\mathcal{K}_2$ be two distinct sets of background knowledge. If repeated applications of Meek's rules to $\mathcal{C}^{\mathcal{K}_1}$ and $\mathcal{C}^{\mathcal{K}_2}$ lead to the same MPDAG, then we say that $\mathcal{K}_1$ and $\mathcal{K}_2$ are *equivalent* given $\mathcal{C}$. We say that $\mathcal{K}_2$ is *redundant* relative to $\mathcal{K}_1$ if $\mathcal{K}_1 \subseteq \mathcal{K}_2$ and $\mathcal{K}_1$ and $\mathcal{K}_2$ are equivalent. We say that a PDAG $\mathcal{G}_1$ is *contained in* another PDAG $\mathcal{G}_2$ if they have the same skeleton and every directed edge in $\mathcal{G}_2$ is also in $\mathcal{G}_1$. We say that $\mathcal{K}_1$ is more *informative* than $\mathcal{K}_2$ if the MPDAG relative to $\mathcal{K}_1$ is contained in the MPDAG relative to $\mathcal{K}_2$

2.2 TIERED BACKGROUND KNOWLEDGE

In this work we focus on background knowledge about tiered structures.

Definition 3 (Partial tiered ordering)**.** *Let $\mathcal{G}$ be a PDAG with node set $\mathbf{V}$ of size p, and let $T \in \mathbb{N}$, $T \leq p$. A (partial) tiered ordering of the nodes in $\mathbf{V}$ is a map $\tau : \mathbf{V} \mapsto \{1, \ldots, T\}^p$ that assigns each node $V \in \mathbf{V}$ to a unique tier $t \in \{1, \ldots, T\}$.*

A tiered ordering is partial if multiple nodes are assigned to the same tier. The following properties are implied by the definition and reflect what kind of background knowledge is encoded in a tiered ordering:

(Uniqueness) A node is assigned to no more than one tier.

(Completeness) Every node belongs to a tier.

(Transitivity) If $\tau(A) \leq \tau(B)$ and $\tau(B) \leq \tau(C)$ then $\tau(A) \leq \tau(C)$.

A tiered ordering imposes background knowledge on a graph by demanding that no directed edges point from later tiers into earlier tiers, i.e. specifying the forbidden edges accordingly $\mathcal{F} = \{\{A \leftarrow B\} : \tau(A) < \tau(B), A, B \in \mathbf{V}\}$. Thus, a tiered ordering provides information on the absence of ancestral relations, but not on their presence. Combining tiered knowledge with a PDAG, it might be possible to construct some ancestral relations which allow us to orient undirected edges as illustrated in Example 1.

Example 1. *Assume that we are given $\mathbf{V} = \{A, B\}$ and tiered ordering τ with $\tau(A) < \tau(B)$. This corresponds to the background knowledge $\mathcal{K}$ with forbidden set $\mathcal{F} = \{A \leftarrow B\}$ and no required edges: In this case it could be possible that A is a parent of B or that there is no edge between them. However, if we additionally knew that A and B are adjacent, then our background knowledge would result in the edge orientation $\{A \rightarrow B\}$.*

As illustrated, the cross-tier edges play an important role and are defined as follows.

Definition 4. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a PDAG and τ a tiered ordering of $\mathbf{V}$. An edge $\{A \rightarrow B\} \in \mathbf{E}$ is a cross-tier edge (relative to τ) if $\tau(A) < \tau(B)$.*

With tiered knowledge τ, all cross-tier edges of a PDAG will be directed. Conversely, $A - B$ only occurs if $\tau(A) = \tau(B)$. Since a tiered ordering τ of a node set $\mathbf{V}$ unambiguously implies a forbidden edge set we can refer to the MPDAG obtained from a CPDAG relative to a tiered ordering τ, rather than referring to the forbidden edges implied by τ.

We will refer to MPDAGs relative to exclusively tiered background knowledge as 'tiered MPDAGs'. This is in contrast to general MPDAGs, which can arise from any kind of background knowledge.

Example 2 (Equivalence class restricted by tiered ordering)**.** *Figure 1 shows a DAG $\mathcal{D}$ and a tiered ordering τ of the nodes in $\mathcal{D}$. The differences between the equivalence class of $\mathcal{D}$ and the restricted equivalence class of $\mathcal{D}$ relative to τ are illustrated in Figure 2. The CPDAG $\mathcal{C}$ represents the equivalence class of $\mathcal{D}$; as only two out of seven edges are directed the conditional independencies alone do not contain much causal information. Meanwhile, the restricted equivalence class represented by tiered MPDAG $\mathcal{G}$ relative to τ is much smaller, here six out of seven edges are oriented. $\mathcal{G}$ will naturally contain the same adjacencies and v-structures as $\mathcal{C}$, the dashed cross-tier edges $A \rightarrow C$ and $C \rightarrow F$ are implied by the forbidden directions, and the dotted edges $C \rightarrow D$ and $F \rightarrow G$ are consequences of Meek's 1st rule which prohibits new v-structures. Due to these last additionally implied orientations of previously undirected edges, restricted equivalence classes given tiered background knowledge might be even smaller, and thus more informative, than one might initially expect.*

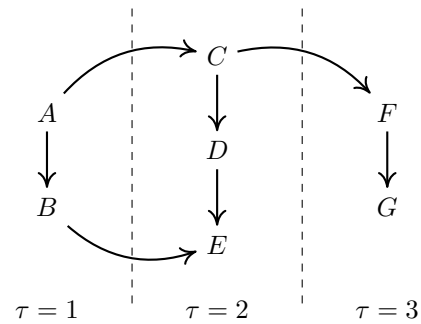

Figure 1: DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ with tiered ordering τ and three tiers: A and B are assigned to tier 1, while C, D and E are assigned to tier 2, and F and G are assigned to tier 3.

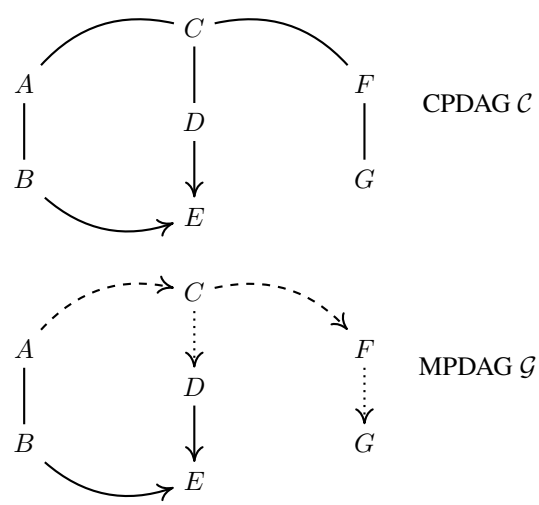

Figure 2: The CPDAG $\mathcal{C}$ representing the equivalence class of $\mathcal{D}$, and the tiered MPDAG $\mathcal{G}$ representing the restricted equivalence class of $\mathcal{D}$ relative to τ.

3 PROPERTIES OF TIERED MPDAGS

When incorporating general background knowledge into a CPDAG, it might be necessary to apply all of Meek's rules 1-4 in order to obtain a maximally informative graph. Meek's 1st rule ensures that no new v-structures are created, while rules 2-4 all concern preventing directed cycles. By construction, tiered knowledge imposes an ordering of the nodes; using that this ordering is transitive and complete, the following lemma shows that Meek's 1st rule is sufficient to construct a maximally informative graph. This is a strong result, which will help us prove further results in this and the following sections.

Lemma 1. *Let $\mathcal{C} = (\mathbf{V}, \mathbf{E})$ be a CPDAG and let τ be a tiered ordering of the nodes $\mathbf{V}$. Let $\mathcal{C}^\tau$ be the PDAG obtained according to Algorithm 1 and let $\mathcal{G}$ be the PDAG obtained by repeatedly orienting edges in $\mathcal{C}^\tau$ according to Meek's rule 1 until no further change occurs. Then $\mathcal{G}$ is the MPDAG obtained from $\mathcal{C}$ relative to τ.*

Note that so far we have taken a given CPDAG as the starting to which tiered background knowledge is added. Alternatively, we can start at an earlier stage with a PDAG $\mathcal{G}$ that contains directed edges only if they belong to v-structures with other edges being undirected. In this case, the tiered ordering can be incorporated into $\mathcal{G}$ by orienting the cross-tier edges and then applying all four of Meeks rules to achieve maximality. Lemma 1 therefore highlights the extra orientations implied by tiered knowledge on top of the usual orientations.

While general MPDAGs might contain partially directed cycles, it turns out that when background knowledge arises from tiered structures the completeness and transitivity of tiered orderings ensure that no partially directed cycles can occur:

Theorem 1. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a tiered MPDAG, then $\mathcal{G}$ does not have any partially directed cycles.*

The implications of Theorem 1 allow us to work with tiered MPDAGs in a similar way as with CPDAGs. The same does not hold for MPDAGs in general [Perković et al., 2017]; we will elaborate on this in the section below.

It was shown by Andersson et al. [1997] that CPDAGs are chain graphs with chordal chain components, which is useful for many purposes. Given Theorem 1 it becomes straightforward to show that the the same holds for tiered MPDAGs:

Corollary 1. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a tiered MPDAG, then $\mathcal{G}$ is a chain graph with chordal chain components.*

3.1 INTERPRETATION OF UNDIRECTED PATHS

In a CPDAG $\mathcal{C} = (\mathbf{V}, \mathbf{E})$, an undirected path π between two nodes $A, B \in \mathbf{V}$ indicates that there exist DAG $\mathcal{D}_1 \in [\mathcal{C}]$ and another DAG $\mathcal{D}_2 \in [\mathcal{C}]$ such that A is an ancestor of B in $\mathcal{D}_1$ and B is an ancestor of A in $\mathcal{D}_2$. This is not necessarily the case in a general MPDAG $\mathcal{G} = (\mathbf{V}, \mathbf{E}')$: If there is an edge $A \to B$ in $\mathcal{G}$ not on π, i.e. $\mathcal{G}$ has a partially directed cycle, then there exists a DAG $\mathcal{D}_1$ represented in the class by $\mathcal{G}$ in which the path corresponding to π is directed from A to B, but there cannot be a DAG $\mathcal{D}_2$ in the class represented by $\mathcal{G}$ in which B is an ancestor of A since this would create a cycle. Hence, an undirected path between two nodes in an MPDAG does not necessarily mean that the path can be directed either way. Hence it is necessary to check multiple paths in the graph in order to determine whether one path can be directed. Since partially directed cycles do not occur in tiered MPDAGs, the above issue does not occur and, hence, the interpretation of undirected paths in tiered MPDAGs is the same as in CPDAGs.

3.2 ADJUSTMENT IN TIERED MPDAGS

The *generalised adjustment criterion* [Perković et al., 2018] determines whether a set of nodes in a CPDAG constitutes a valid adjustment set in every DAG in the equivalence class. This criterion checks for possibly causal paths, i.e. paths for which there are DAGs in the equivalence in which these paths are causal. Hence, any undirected path is possibly causal. However, as described in Section 3.1, this interpretation does not hold for general MPDAGs, and the notion of possibly causal does not transfer directly from CPDAGs to MPDAGs. In order to tackle this issue, Perković et al. [2017] introduced the notion of *b-possibly causal* paths, which is a stronger requirement but ensures that there are in fact paths in the equivalence class that are causal. With this notion an adjustment criterion for general MPDAGs, the *b-adjustment criterion*, can be given [Perković et al., 2017]. To determine whether a path is b-possibly causal, one needs to check multiple paths in the graph, which can be computationally heavy for large and dense graphs. For tiered MPDAGs, the definition of b-possibly causal paths simplifies to the definition of possibly causal paths, and the generalised adjustment criterion for CPDAGs is valid for tiered MPDAGs as well:

Corollary 2. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a tiered MPDAG, and let $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \subseteq \mathbf{V}$ be pairwise disjoint node sets. Then $\mathbf{Z}$ satisfies the generalised adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$ if and only if it satisfies the b-adjustment criterion relative to $(\mathbf{X}, \mathbf{Y})$ in $\mathcal{G}$.*

A related result is shown in van der Zander and Liskiewicz [2016]: They introduce a class of graphs called *restricted chain graphs*, which are chain graphs with (1) chordal chain components, and (2) no unshielded triples of the form $A \to B - C$. They give a sound and complete adjustment criterion for graphs of this type, and they provide an algorithm to find adjustment sets. Clearly, a tiered MPDAG is a type of restricted chain graph, and the results of van der Zander and Liskiewicz [2016] hold for tiered MPDAGs.

3.3 IDA FOR TIERED MPDAGS

Covariate adjustment in a CPDAG (MPDAG) requires an adjustment set to be a valid in every DAG in the (restricted) equivalence class. The IDA-algorithm [Maathuis et al., 2009], instead, finds an adjustment set for each DAG in the class. Enumerating all DAGs in an equivalence class is a computationally heavy task. The 'local' IDA algorithm utilises the fact that if a valid adjustment set exists, then

the parent set is always valid [Pearl, 2009] and considers the possible parents, i.e. all sets that are parent sets in some DAG in the equivalence class. However, as general MPDAGs can contain partially directed cycles, it cannot be verified locally whether a set of nodes is a possible parent set, and a semi-local version was introduced to tackle this [Perković et al., 2017]. The joint IDA-algorithm determines the joint parent sets semi-locally by orienting subgraphs [Nandy et al., 2017]. Similar to the local IDA-algorithm, this approach fails for general MPDAGs due to potential partially directed cycles. To tackle this issue Perković et al. [2017] introduced an additional step to check whether the oriented subgraphs are valid. In contrast, for tiered MPDAGs no additional steps are needed, and the original local and joint IDA both remain valid:

Corollary 3. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a tiered MPDAG. Let $\mathbf{PA}_{\mathcal{G}}(X)$ denote the multiset of parent sets of X in all DAGs represented by $\mathcal{G}$, and let $\mathbf{PA}_{\mathcal{G}}^{\text{local}}(X)$ denote the multiset of parent sets of X obtained from the local IDA algorithm. Then $\mathbf{PA}_{\mathcal{G}}(X)$ and $\mathbf{PA}_{\mathcal{G}}^{\text{local}}(X)$ contain the same distinct elements. Moreover, let $\mathbf{PA}_{\mathcal{G}}^{\text{joint}}(X)$ denote the multiset of parent sets of X obtained from the joint IDA algorithm. Then $\mathbf{PA}_{\mathcal{G}}(X)$ and $\mathbf{PA}_{\mathcal{G}}^{\text{joint}}(X)$ contain the same distinct elements and the ratios of multiplicities of any two elements are the same.*

In order to adapt the local IDA-algorithm to general MPDAGs, Fang and He [2020] introduced a set of local orientation rules to verify whether a set of nodes is a possible parent set of a given node, yielding a fully local version of the IDA that can handle general MPDAGs. While this reduces computation time for general MPDAGs, it is not necessary for tiered MPDAGs due to Corollary 3.

Finally, the optimal IDA algorithm [Witte et al., 2020] is only semi-local and it includes an additional step to check other parts of the graph: This algorithm is essentially not local as the optimal adjustment set is not likely to be the parent set [Witte et al., 2020, Henckel et al., 2022]. In this case, tiered MPDAGs do not have an advantage over general MPDAGs. However, the definition of the optimal adjustment set does simplify in a similar fashion as in Section 3.2.

4 COMPARING TIERED BACKGROUND KNOWLEDGE

In this section, we investigate how different tiered background knowledge can be compared. This is relevant in situations where different experts are consulted or where eliciting more detailed knowledge may require more effort. It also provides insights into what kind of background knowledge is especially valuable.

Throughout, we will only compare tiered orderings that are compatible in the sense that they do not contradict each other on the ordering of the nodes; this is in line with our assumption of all background knowledge being correct. Hence, the orderings can only disagree on the status of an edge being a cross-tier edge, not the direction of it.

Consider two different tiered orderings τ_i and τ_j. If for all $A, B \in \mathbf{V} : \tau_i(A) < \tau_i(B) \Rightarrow \tau_j(A) < \tau_j(B)$, then τ_j is *finer* than τ_i, and τ_i is *coarser* than τ_j. In this case we have for the respective sets of forbidden edges that $\mathcal{F}_i \subseteq \mathcal{F}_j$.

Note that a finer tiered ordering can be redundant compared to the coarser one; otherwise it must be more informative. However, two tiered orderings can be different without one being finer or coarser than the other. They can then either be equivalent, or one can be more informative than the other, or they are incomparable.

To compare tiered orderings on a given CPDAG, we must compare the resulting tiered MPDAGs. We give a graphical criterion for their equivalence in Section 4.1. Evidently, this provides a criterion for redundancy. More interestingly, this also provides insight into when one ordering is more informative than another as addressed in Section 4.2.

4.1 EQUIVALENCE OF TIERED MPDAGS

First, we need some further terminology:

Definition 5. *Let $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ be a PDAG, let τ be a tiered ordering of the nodes $\mathbf{V}$, and let π_1 and π_2 be two arbitrary paths in $\mathcal{G}$. If π_1 contains a node V with $\tau(V) < \tau(W)$ for all nodes W on π_2, we say that π_1 is earlier than π_2; correspondingly, π_2 is later than π_1. A path is earliest if it does not contain subpaths of any earlier paths.*

We say that an edge is *fully shielded* if it does not occur on any unshielded path.

In the following, $\mathcal{C}_u^{\tau}$ refers to the graph obtained by first omitting every directed edge in a graph $\mathcal{C}$, and then orienting edges in $\mathcal{C}_u$ according to the tiered ordering τ.

Theorem 2. *Let $\mathcal{C}$ be a CPDAG and let τ_1 and τ_2 be distinct tiered orderings. Let $\mathcal{G}_1$ be the tiered MPDAG obtained from $\mathcal{C}$ relative to τ_1 and $\mathcal{G}_2$ the tiered MPDAG obtained from $\mathcal{C}$ relative to τ_2. Then $\mathcal{G}_1 = \mathcal{G}_2$ if and only if*

(i) $\mathcal{C}_u^{\tau_1}$ and $\mathcal{C}_u^{\tau_2}$ agree on the first cross-tier edge on any earliest unshielded path and

(ii) $\mathcal{C}_u^{\tau_1}$ and $\mathcal{C}_u^{\tau_2}$ agree on any fully shielded cross-tier edge.

A path can have at most two first cross-tier edges. In the case where a path contains two first cross-tier edges, the above condition requires $\mathcal{C}_u^{\tau_1}$ and $\mathcal{C}_u^{\tau_2}$ to agree on both cross-tier edges.

Example 3 (Equivalent tiered orderings)**.** *We again consider the DAG $\mathcal{D}$ and the tiered ordering τ in Figure 1 from*

Example 2. Now we compare τ to the tiered ordering τ' with

$$\tau'(A) = \tau'(B) = 1$$
$$\tau'(C) = \tau'(D) = \tau'(E) = \tau'(F) = \tau'(G) = 2$$

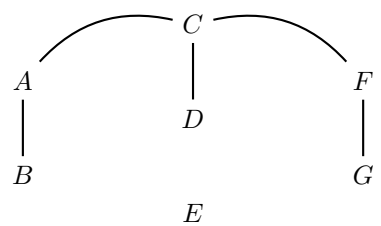

Figure 3: The undirected subgraph $\mathcal{C}_u$ of the CPDAG $\mathcal{C}$ from Figure 2.

In Figure 3 we have the undirected subgraph $\mathcal{C}_u$ of the CPDAG $\mathcal{C}$ of $\mathcal{D}$. There are two earliest unshielded paths in $\mathcal{C}_u$: $B - A - C - F - G$ and $B - A - C - D$. Clearly, $\mathcal{C}_u^{\tau}$ and $\mathcal{C}_u^{\tau'}$ agree on the first cross-tier edges on these paths: On $\langle B, A, C, F, G\rangle$ the first cross-tier edge is $A \to C$, and on $\langle B, A, C, D\rangle$ the first cross-tier edge is $A \to C$. Note that in these graphs, no edge is fully shielded. Orienting edges in $\mathcal{C}_u^{\tau}$ and $\mathcal{C}_u^{\tau'}$ repeatedly according to Meek's 1st rule results in the same MPDAG. Since τ is finer than τ', it follows that τ is redundant relative to τ' given $\mathcal{C}$.

4.2 COMPARING TIERED ORDERINGS

We have now characterised when two tiered orderings produce the same MPDAG; as a consequence of this, Theorem 2 also gives rise to useful insights concerning when tiered knowledge is more informative.

Corollary 4. *Let $\mathcal{C} = (\mathbf{V}, \mathbf{E})$ be a CPDAG and let τ_1 and τ_2 be two distinct tiered orderings of $\mathbf{V}$. Assume that*

(i) every first cross-tier edge on an earliest unshielded path in $\mathcal{C}_u^{\tau_2}$ is also a cross-tier edge in $\mathcal{C}_u^{\tau_1}$, and

(ii) every fully shielded cross-tier edge in $\mathcal{C}_u^{\tau_2}$ is also a cross-tier edge in $\mathcal{C}_u^{\tau_1}$,

then τ_1 is more informative than τ_2 given $\mathcal{C}$ if

(iii) $\mathcal{C}_u^{\tau_1}$ has a first cross-tier edge on an earliest unshielded path, that is not a cross-tier edge in $\mathcal{C}_u^{\tau_2}$, or

(iv) $\mathcal{C}_u^{\tau_1}$ has more fully shielded cross-tier edges than $\mathcal{C}_u^{\tau_2}$.

In Corollary 4, (i) and (ii) ensure that the MPDAG $\mathcal{G}_1$ relative to τ_1 is contained in the MPDAG $\mathcal{G}_2$ relative to τ_2. Additional information can be obtained in two ways: Either directly from the tiered ordering, reflected in condition (iv), or indirectly through Meek's 1st rule, reflected in condition (iii).

Consider first condition (iii). This condition does not mention the number of cross-tier edges, it is only important to know the earliest.

Example 4. *Consider again the DAG $\mathcal{D} = (\mathbf{V}, \mathbf{E})$ from Example 2 and Example 3. Consider now the tiered ordering τ_1 with*

$$\tau_1(A) = \tau_1(B) = 1$$
$$\tau_1(C) = 2$$
$$\tau_1(D) = \tau_1(E) = 3$$
$$\tau_1(F) = 4, \text{ and}$$
$$\tau_1(G) = 5$$

Clearly, τ_1 results in the same MPDAG as τ and τ'; since τ_1 is finer than both τ and τ' it is redundant given $\mathcal{C}$. Consider now the tiered ordering τ_2 of $\mathbf{V}$ with

$$\tau_2(A) = \tau_2(B) = \tau_2(C) = 1$$
$$\tau_2(D) = \tau_2(E) = 2$$
$$\tau_2(F) = 3, \text{ and}$$
$$\tau_2(G) = 4$$

Let $\mathcal{G}'$ be the MPDAG relative to τ_2 given $\mathcal{C}$. Then $\mathcal{G}'$ has the same edge orientations as the MPDAG relative to τ, τ' and τ_1, except for $\{A \to C\}$ which remains undirected in $\mathcal{G}'$. In this case, τ_1 is finer than τ_2, but τ_1 is not redundant. In fact, τ_1 is more informative than τ_2 due to condition (iii) of Corollary 4. In addition, τ and τ' are also both more informative than τ_2, even though τ_2 assigns the nodes to more tiers.

Consider condition (iv) in Corollary 4. This suggests that every fully shielded cross-tier edge provides unique information. Hence, there is an immediate gain in information from each additional fully shielded cross-tier edge in $\mathcal{C}_u$, since edges of this type can not be oriented by Meek's 1st rule. Hence, in the complete subgraphs of $\mathcal{C}_u$, no tiered background knowledge is redundant.

Example 5 (Fully shielded edges)**.** *Consider the simple case of a CPDAG $\mathcal{C} = (\mathbf{V}, \mathbf{E})$ with three nodes $\mathbf{V} = \{A, B, C\}$, where $\mathcal{C}$ is complete: $\mathbf{E} = \{\{A - B\}, \{B - C\}, \{A - C\}\}$. Assume that the true ordering τ_α assigns the nodes to individual tiers with $\tau_\alpha(A) < \tau_\alpha(B) < \tau_\alpha(C)$. There are then three types of partial orderings that are compatible with τ_α: Orderings that assigns the nodes to the same tier, e.g. an ordering τ_β with $\tau_\beta(A) = \tau_\beta(B) = \tau_\beta(C)$, or orderings that assigns two nodes to the same tier and the third to an individual tier, e.g. τ_γ and τ_δ with $\tau_\gamma(A) = \tau_\gamma(B) < \tau_\gamma(C)$ and $\tau_\delta(A) < \tau_\delta(B) = \tau_\delta(C)$.*

Figure 4 shows that the MPDAGs $\mathcal{G}_\alpha$ (relative to τ_α), $\mathcal{G}_\beta$ (relative to τ_β), $\mathcal{G}_\gamma$ (relative to τ_γ), and $\mathcal{G}_\delta$ (relative to τ_δ). Note that they are all distinct: All oriented edges are implied

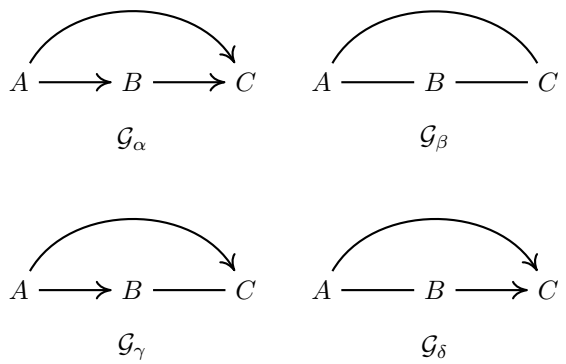

Figure 4: Top left: MPDAG $\mathcal{G}_\alpha$ relative to τ_α; note that this is equal to the underlying DAG. Top right: MPDAG $\mathcal{G}_\beta$ relative to τ_β; note that this is equal to the CPDAG $\mathcal{C}$. Bottom left: MPDAG $\mathcal{G}_\gamma$ relative to τ_γ. Bottom right: MPDAG $\mathcal{G}_\delta$ relative to τ_δ.

by the tiered background knowledge and no edge has been oriented as a consequence of Meek's 1st rule. Here, τ_δ and τ_γ are incomparable, and they are both more informative than τ_β. Moreover, τ_α is more informative than all the other orderings.

4.3 SIMULATION STUDY

Corollary 4 shows that in graphs with many unshielded paths we can potentially obtain a large amount of additional information, and the earlier we are be able to identify the direction of a causal path, the more information we can gain. In summary: (1) early knowledge is in general more beneficial than late knowledge, even if the late knowledge is more detailed (c.f. Example 4), and (2) since we expect unshielded paths to occur more frequently in sparse graphs, the effect of Meek's 1st rule is expected to be more pronounced in sparse than in dense graphs. In order to investigate to which degree these two features occur in practice, we conducted a simulation study.

We generated random DAGs and for each random DAG, we considered five different tiered orderings: full knowledge (τ_{full}), early detailed knowledge (τ_{early2}), late detailed knowledge (τ_{late2}), early simple knowledge (τ_{early1}) and late simple knowledge (τ_{late1}). For each DAG, we constructed its CPDAG, and for each combination of DAG and tiered ordering, we constructed the tiered MPDAG. For each MPDAG, we counted the difference in number of directed edges between the MPDAG and its corresponding CPDAG, and we divided this number by the total number of edges. This measures the fraction of edges that cannot be oriented in the CPDAG, but can be oriented in the tiered MPDAG. A detailed description of the study can be found in Section C in the Supplement.

Figure 5 shows that using the full knowledge of the orderings unsurprisingly provides most new orientations. Moreover, we see that early knowledge is more beneficial than late, even though they are equally detailed, which is in line with (1) above. Interestingly, in dense graphs early knowledge can also induce more oriented edges than late knowledge, even if the later knowledge is more detailed, which is also in line with (1). Additionally, we see that the advantage of adding tiered background knowledge is relatively larger in sparse graphs than dense graphs, which is in line with (2).

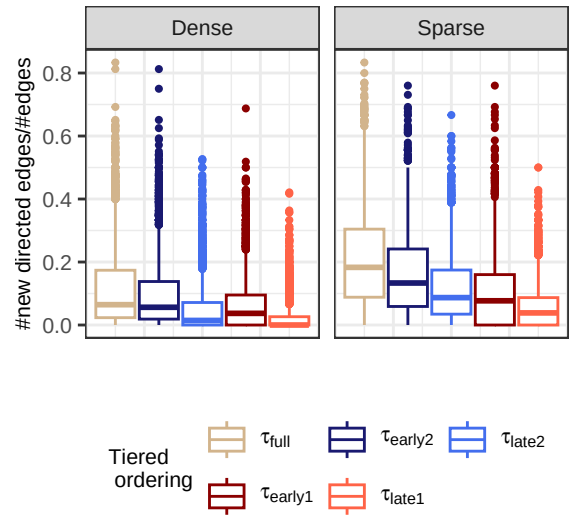

Figure 5: Results of one setting of the simulation study. 6000 random DAGs with 25 nodes were generated; half of them sparse, the other half dense. For each random DAG and each tiered ordering, the tiered MPDAG was constructed and the difference in number of directed edges to its corresponding CPDAG was computed and divided by the total number of edges.

We performed the same procedure for random DAGs with node sets of size 10, 50 and 100, for which we found analogous results. These results are depicted in Figure C.2 in the Supplement.

5 RELATION TO OTHER WORK

Subject matter background knowledge can come from different sources and take different forms, and previous work provides results for other knowledge than tiered one. A distinct type of causal background knowledge is, for instance, obtainable when experimentation is possible, see Hauser and Bühlmann [2012] for a characterisation of interventional equivalence classes of DAGs. While a tiered ordering is given before learning a graph, experiments can be performed iteratively, and the choice of most informative interventions might depend on the given intermediate graphical structure. It has been shown in Eberhardt [2008] and Hauser

and Bühlmann [2014] that the most informative strategy is to intervene on nodes in the largest undirected complete subgraphs: this yields most new edge orientations, including those following from Meek's rules. Since the knowledge obtained from an intervention is local, this type of background knowledge lacks the completeness of tiered knowledge. This means that all four of Meek's rules might apply after orienting edges according to interventional knowledge, resulting in additional orientations within a complete subgraphs, in contrast to tiered background knowledge.

Mooij et al. [2020] considered context variables. These can be seen as a special case of tiered knowledge: Some of the variables, the context variables, form an earlier tier, and others, the system variables, form later tiers. However, there can be additional knowledge about presence/absence of relations between context variables, or their causal relations may not be of interest: In these cases we are no longer in the tiered framework.

Background knowledge about non-ancestral (pairwise) relations is considered by Fang and He [2020], and can be seen as non-complete version of tiered knowledge. They show that knowledge of non-ancestral relations can be translated to a set of direct causal relations,i.e. directed edges, for a given PDAG. In contrast, the completeness and transitivity of tiered knowledge subsumes such relations through the orientations of cross-tier edges. In addition, Fang et al. [2022] discuss ancestral background knowledge on node pairs. This is surprisingly different by nature and cannot necessarily be encoded graphically.

A time structure, and thus tiered ordering, is of course naturally given for time series [Malinsky and Spirtes, 2018, Runge et al., 2019]. The typical setting of time series and their standard data structure being one long observation instead of multiple i.i.d. observations often leads to an additional stationarity assumption of the causal structure. While this requires different inferential techniques it is evident that our results on the resulting equivalence classes, such as absence of partially directed cycles, still apply.

A different line of work on using background knowledge relaxes the assumption of causal sufficiency. Latent variables can be accommodated in maximal ancestral graphs (MAGs) (Richardson and Spirtes [2002]), and the corresponding equivalence classes are represented by partial ancestral graphs (PAGs), see characterization by Ali et al. [2009]. Different orientation rules are needed, and a set of ten rules were introduced by Zhang [2008] ensuring the maximally informative PAG. For added background knowledge it has not yet been shown, in general, that these ten rules yield a maximally informative graph. However, this has been shown under the extra assumptions of no cross-tier confounding and no selection bias [Andrews et al., 2020]. Moreover, in this case it turns out that not all ten orientation rules are needed, similarly to our Lemma 1. We therefore conjecture that analogous results to those of section 3 extend to PAGs with tiered background knowledge under the assumptions of no selection bias and no cross-tier confounding. Further work is still needed to relax the often implausible assumption of no cross-tier confounding.

6 DISCUSSION

By formalising equivalence classes restricted by tiered orderings, we provided some new insights: Tiered MPDAGs do not have partially directed cycles and are chain graphs with chordal chain components; this makes them easier to handle and interpret, e.g. for causal effect estimation. We have given a characterisation of tiered MPDAGs which clarified what can be gained by adding tiered knowledge and what will still remain unknown. Sparse graphs, in particular, will benefit much from edge orientations implied by the tiered ordering; further, eliciting background knowledge to separate out early tiers is especially informative.

In summary, we believe that oftentimes background knowledge comes in the form of a tiered ordering. Moreover, tiered knowledge can be expected to be reliable especially when based on temporal information. Tiered MPDAGs are, therefore, at least as plausible as their corresponding CPDAGs without background knowledge. In addition, tiered MPDAGs will also be at least as informative as their corresponding CPDAGs – in practice they will often be much more informative. While it is self-evident that any prior knowledge should be exploited for causal structure learning, we have illustrated how specific aspects and how much the prior knowledge results in an information gain that goes well beyond the orientation of cross-tier edges.

References

R Ayesha Ali, Thomas S Richardson, and Peter Spirtes. Markov equivalence for ancestral graphs. *The Annals of Statistics*, 37(5B):2808–2837, 2009.

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 *International Conference on Artificial Intelligence and Statistics*, pages 4002–4011. PMLR, 2020.

Frederick Eberhardt. Almost optimal intervention sets for causal discovery. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 161–168, 2008.

Zhuangyan Fang and Yangbo He. Ida with background knowledge. In *Conference on Uncertainty in Artificial Intelligence*, pages 270–279. PMLR, 2020.

Zhuangyan Fang, Ruiqi Zhao, Yue Liu, and Yangbo He. On the representation of causal background knowledge and its applications in causal inference. *arXiv preprint arXiv:2207.05067*, 2022.

Ronja Foraita, Janine Witte, Claudia Bornhorst, Wencke Gwozdz, Valeria Pala, Lauren Lissner, Fabio Lauria, Lucia A Reisch, Denes Molnar, Stefaan De Henauw, et al. A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a european cohort of children and adolescents. *medRxiv*, 2022.

Alain Hauser and Peter Bühlmann. Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs. *The Journal of Machine Learning Research*, 13(1):2409–2464, 2012.

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

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*, 84(2):579–599, 2022.

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. Causal structure learning from multivariate time series in settings with unmeasured confounding. In *Proceedings of 2018 ACM SIGKDD workshop on causal discovery*, pages 23–47. PMLR, 2018.

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.

Joris M Mooij, Sara Magliacane, and Tom Claassen. Joint causal inference from multiple contexts. *Journal of Machine Learning Research*, 21:1–108, 2020.

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.

Judea Pearl. *Causality*. Cambridge university press, 2009.

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

Emilija Perković, Johannes Textor, Markus Kalisch, and Marloes H Maathuis. Complete graphical characterization and construction of adjustment sets in markov equivalence classes of ancestral graphs. *The Journal of Machine Learning Research*, 18, 2018.

Anne H Petersen, Merete Osler, and Claus T Ekstrøm. Data-driven model building for life-course epidemiology. *American Journal of Epidemiology*, 190(9):1898–1907, 2021.

Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D Mahecha, Jordi Muñoz-Marí, et al. Inferring causation from time series in earth system sciences. *Nature communications*, 10(1):2553, 2019.

Richard Scheines, Peter Spirtes, Clark Glymour, Christopher Meek, and Thomas Richardson. The tetrad project: Constraint based aids to causal model specification. *Multivariate Behavioral Research*, 33(1):65–117, 1998.

Benito van der Zander and Maciej Liskiewicz. Separators and adjustment sets in markov equivalent dags. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

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.

Janine Witte, Leonard Henckel, Marloes H Maathuis, and Vanessa Didelez. On efficient adjustment in causal graphs. *Journal of Machine Learning Research*, 21:246, 2020.

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