Arrowhead completeness from minimal conditional independencies*

Tom Claassen, Tom Heskes
Radboud University Nijmegen
The Netherlands
{tomc,tomh}@cs.ru.nl

Abstract

We present two inference rules, based on so called *minimal conditional independencies*, that are sufficient to find *all* invariant arrowheads in a single causal DAG, even when selection bias may be present. It turns out that the set of seven graphical orientation rules that are usually employed to identify these arrowheads are, in fact, just different instances/manifestations of these two rules. Contains detailed proofs to main results.

1 Introduction

Learning causal relations from data is still considered a challenging task, especially when latent variables and selection bias may be present. Even in these circumstances, some methods, like the FCI-algorithm [2], are able to infer valid causal information from conditional (in)dependencies in an observed probability distribution. For example, they may identify that variable X causally affects variable Y, or, conversely, that variable Y is *not* a cause of X, and can often distinguish between direct and indirect causal relations.

In this article, we focus on the second aspect: detectable absence of causal influence. In the graphical output produced by the FCI-algorithm, an arrowhead at a node Y on an edge from X in the graph represents causal information of the form 'Y does *not* cause X'. We want to show that two rules, based on the ones in [12] for combining multiple models, form a sound and complete set to uncover *all* identifiable arrowheads in this output.

Section 2 introduces some basic terminology and the assumptions made. Section 3 describes the current state-of-the-art approach to learning all invariant arrowheads from a given distribution. Section 4 generalizes the link between observed *minimal* conditional (in)dependencies and causal relations, to allow for possible selection bias. Section 5 contains the main result: that two rules are sufficient to derive all arrowheads, *without* relying on the detailed structure of the graph. The last two sections discuss some extensions and potential applications for this approach.

2 Graphical model preliminaries

2.1 Terminology

First some standard graphical model concepts used throughout the article. A *directed graph* $\mathcal{G}$ is a pair $\langle \mathbf{V}, \mathbf{E} \rangle$, where $\mathbf{V}$ is a set of vertices or nodes and $\mathbf{E}$ is a set of edges between pairs of nodes, represented by arrows $X \rightarrow Y$. A *path* $\pi = \langle V_0, \ldots, V_n \rangle$ between V_0 and V_n in $\mathcal{G}$ is a sequence of distinct vertices such that for $0 \leq i \leq n-1$, V_i and V_{i+1} are connected by an edge in $\mathcal{G}$. A

*Technical Report, Radboud University Nijmegen, iCIS/IS/ML, April 2011. (added details and corrections w.r.t. Nov. 2010 version)

directed path is a path that is traversed entirely in the direction of the arrows. A *directed acyclic graph* (DAG) is a directed graph that does not contain a directed path from any node to itself. A vertex X is an *ancestor* of Y (and Y is a *descendant* of X) if there is a directed path from X to Y in $\mathcal{G}$ or if $X = Y$. A vertex Z is a *collider* on a path $\pi = \langle \ldots, X, Z, Y, \ldots \rangle$ if it contains the subpath $X \rightarrow Z \leftarrow Y$, otherwise it is a *noncollider*. A *trek* is a path that does not contain any collider.

For disjoint (sets of) vertices X, Y and $\mathbf{Z}$ in a DAG $\mathcal{G}$, X is *d-connected* to Y conditional on $\mathbf{Z}$ (possibly empty), iff there exists an unblocked path $\pi = \langle X, \ldots, Y \rangle$ between X and Y given $\mathbf{Z}$, i.e. such that every collider on π is an ancestor of some $Z \in \mathbf{Z}$ and every noncollider on π is not in $\mathbf{Z}$. If not, then all such paths are blocked, and X is said to be *d-separated* from Y; see [1, 2] for details.

Definition 1. Two nodes X and Y are *minimally conditionally independent* given a set of nodes $\mathbf{Z}$, denoted $[X \perp\!\!\!\perp Y \,|\, \mathbf{Z}]$, iff X is conditionally independent of Y given a *minimal* set of nodes $\mathbf{Z}$. Here minimal, indicated by the square brackets, implies that the relation does not hold for any proper subset $\mathbf{Z}' \subsetneq \mathbf{Z}$ of the (possibly empty) set $\mathbf{Z}$.

A *causal DAG* $\mathcal{G}_C$ is a graphical model in the form of a DAG where the arrows represent direct causal interactions between variables in a system [3]. There is a causal relation $X \Rightarrow Y$ iff there is a directed path from X to Y in $\mathcal{G}_C$. Absence of such a path is denoted $X \not\Rightarrow Y$. The *causal Markov condition* links the structure of a causal graph to its probabilistic concomitant, [1]: two variables X and Y in a causal DAG $\mathcal{G}_C$ are dependent given a set of nodes $\mathbf{Z}$, iff they are connected by a path π in $\mathcal{G}_C$ that is unblocked given $\mathbf{Z}$.

When there are possible latent (hidden) and selection variables in a DAG, the distribution over the subset of observed variables can be represented by a *(maximal) ancestral graph* (MAG) [4]. Different MAGs can represent the same distribution, but only the (invariant) features shared by all MAGs that can represent that distribution carry identifiable causal information. The *(complete) partial ancestral graph* (cPAG) $\mathcal{P}$ that represents the equivalence class $[\mathcal{G}]$ of a MAG $\mathcal{G}$ is a graph with either a tail '$-$', arrowhead '$>$' or circle mark '$\circ$' at each end of an edge, such that $\mathcal{P}$ has the same adjacencies as $\mathcal{G}$, and there is a tail or arrowhead on an edge in $\mathcal{P}$, iff it is invariant in $[\mathcal{G}]$, otherwise it has a circle mark [5]. Absence of an edge between two variables X and Y in a graph $\mathcal{G}$ is sometimes explicitly indicated as $X \not\!\!-\!\!\!\!\!\times Y$. The cPAG of a given MAG is unique and maximally informative for $[\mathcal{G}]$. We use PAGs as a concise and intuitive graphical representation of (all) conditional (in)dependence relations between nodes in an observed distribution; see [4, 5] for more information on how to read independencies directly from a MAG/PAG using the m-separation criterion, which is essentially just the d-separation criterion, only applied to MAGs.

2.2 Assumptions

Below we briefly discuss the main assumptions in this article:

- causal DAG
- causal faithfulness
- causal Markov
- large sample limit
- possible latent variables and selection bias
- no deterministic relations / (blocking) interventions

We assume that the systems we consider correspond to some underlying causal DAG $\mathcal{G}_C$ over a great many observed and unobserved nodes.

Throughout this article we adopt the *causal faithfulness assumption*, which implies that all and only the conditional independence relations entailed by the causal Markov condition applied to the true causal DAG will hold in the joint probability distribution over the variables in $\mathcal{G}_C$. For an in-depth discussion of the justification and connection between these assumptions, see [6].

We assume that the large sample limit distributions $P(\mathbf{V})$ are known and can be used to obtain categorical statements about probabilistic (in)dependencies between sets of nodes. Finite data precludes such definite decisions, which has to be accounted for in the conclusions. In this article, however, we are mainly concerned with the validity of our results in the large sample limit (for if they do not apply even there, then they are heuristic at best). Note, however, the remarks at the end of section 6.

In real experiments, we do not observe (in)dependencies between variables in the underlying causal structure directly, but through the data obtained from the experiment. This introduces the possibility of *selection bias*, see [7]. As a result, dependencies may be observed that do not originate from the *true* causal structure, but from the data gathering process. In this way, it is possible to generate distributions that cannot be reproduced faithfully by *any* causal DAG structure, but can be represented by ancestral graphs [4]. Loosely speaking, selection on a node in a causal DAG makes all ancestors dependent, while leaving conditional independencies intact, by destroying (connecting) all unshielded colliders ($X \rightarrow Z \leftarrow Y$ with no edge between X and Y) that have a directed path to it. Consequently, the detection of an unshielded collider Z in observed data also signifies the absence of selection bias on Z or any of its descendants. In this article we explicitly do *not* assume the absence of selection bias, i.e. all conclusions remain valid, whether selection effects are present or not.

Finally, we assume there are no deterministic relations in $\mathcal{G}_C$, nor interventions that effectively block or overrule the causal influence of one node on another. Both these assumptions are implied by faithfulness.

3 Learning causal models

Usually, causal discovery starts from an observed probability distribution over a subset of variables, and then tries to determine which variables do or do not have a directed path to each other in the underlying causal DAG. Here we only consider causal relations that, in the large sample limit, can be discovered from the *conditional (in)dependencies* that are present in the data, and ignore other properties of the distribution that can yield causal information, such as non-Gaussianity and non-linear features, see [8, 9]. From the previous section we know that this independence structure can be represented in the form of a MAG $\mathcal{G}$. However, many different MAGs can represent the same observed probabilty distribution. Therefore, it is only the *invariant features*, such as missing edges and arrowhead- and tail marks at certain edges, that are common to *all* MAGs in the equivalence class $[\mathcal{G}]$, that carry definite, identifiable causal information. The challenge of causal discovery is how to identify these invariant features from a given data set.

The *Fast Causal Inference* (FCI) algorithm [2] was one of the first algorithms that was able to validly infer causal relations from conditional independence statements in the large sample limit, even in the presence of latent and selection variables. It consists of an efficient search for a conditional independence between each pair of variables, to identify the skeleton of the underlying causal MAG, followed by an orientation stage to identify invariant tail and arrowhead marks. It was shown to be sound [7], although not yet complete. Ali et al. [10] showed that a set of seven graphical orientation rules (see below) is sufficient to identify all invariant arrowheads in the equivalence class $[\mathcal{G}]$, given a single MAG $\mathcal{G}$. Later, Zhang [5] extended this result to another set of seven rules that could orient the remaining invariant tails, and proved the resulting procedure to obtain the PAG is not only sound, but also complete.

In the rules, the following notion is employed: in a MAG $\mathcal{G}$, a path $\pi = \langle X, Z_1, (\ldots, Z_k,) Z, Y\rangle$ is a *discriminating path* for Z, iff X is not adjacent to Y, and every node $\{Z_1, \ldots, Z_k\}$ is a collider on π and is a parent of Y.

The rules to obtain the PAG $\mathcal{P}$ start from a given MAG $\mathcal{G}$. However, when starting from data instead of a MAG, each rule is formulated in terms of an equivalent set of (in)dependence statements, see e.g. [7] for more details. In that case, a preliminary rule is used to obtain the skeleton of $\mathcal{P}$:

$\mathcal{R}0$a For every pair of variables $\{X, Y\}$, if there is a set **Z** such that $X \perp\!\!\!\perp Y \,|\, \mathbf{Z}$, then remove the edge $X \ast\!\!\!\not\!\!-\!\!\ast Y$ from $\mathcal{P}$, and record $bf Z \in Sepset(X, Y)$

All edges in the skeleton of $\mathcal{P}$ initially get circle marks $\circ\!\!-\!\!\circ$. For the next step, the seven graphical orientation rules to find all invariant arrowheads are given below, where we follow the enumeration in [5] in referring to the rules. Here tails $-$ and arrowheads $>$ on edges represent invariant marks that have already been identified; circle marks $\circ$ are either not yet identified or non-invariant marks, and the asterisk $*$ is a meta-symbol that stands for any of the other three ('does not matter'). Figure 1 provides a graphical description.

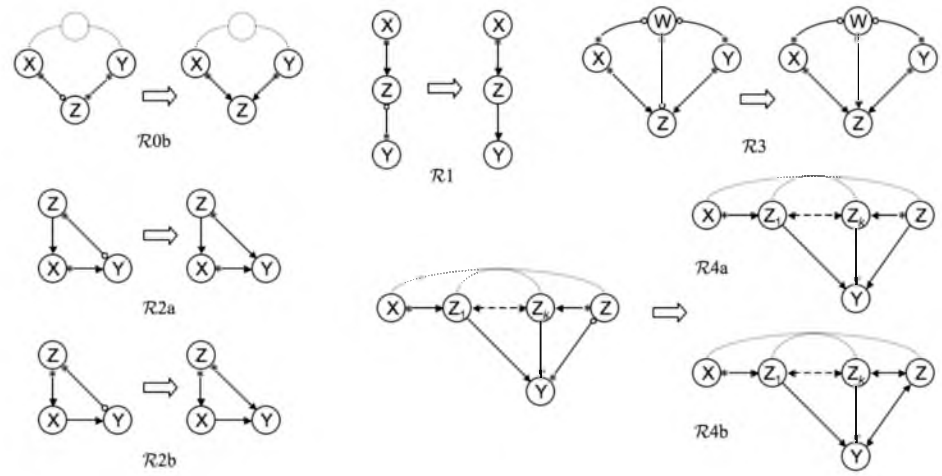

Figure 1: Rules $\mathcal{R}0\text{b}-\mathcal{R}4\text{b}$, arrowhead orientation rules

$\mathcal{R}$0b For every triple $X \ast\!\!-\!\!\ast Z \circ\!\!-\!\!\ast Y$, with X and Y not adjacent, if $Z \notin Sepset(X,Y)$, then orient triple as $X \ast\!\!\rightarrow Z \leftarrow\!\!\ast Y$.

$\mathcal{R}$1 If $X \ast\!\!\rightarrow Z \circ\!\!-\!\!\ast Y$, and X and Y are not adjacent, then orient as $Z \longrightarrow Y$.

$\mathcal{R}$2a If $Z \longrightarrow X \ast\!\!\rightarrow Y$ and $Z \ast\!\!-\!\!\circ Y$, then orient as $Z \ast\!\!\rightarrow Y$.

$\mathcal{R}$2b If $Z \ast\!\!\rightarrow X \longrightarrow Y$ and $Z \ast\!\!-\!\!\circ Y$, then orient as $Z \ast\!\!\rightarrow Y$.

$\mathcal{R}$3 If $X \ast\!\!\rightarrow Z \leftarrow\!\!\ast Y$, $X \ast\!\!-\!\!\circ W \circ\!\!-\!\!\ast Y$, X and Y are not adjacent and $W \ast\!\!-\!\!\circ Z$, then orient as $W \ast\!\!\rightarrow Z$.

$\mathcal{R}$4a If $u = \langle X, \ldots, Z_k, Z, Y\rangle$ is a discriminating path between X and Y for Z, and $Z \circ\!\!-\!\!\ast Y$, then if $Z \in Sepset(X,Y)$, then orient $Z \longrightarrow Y$.

$\mathcal{R}$4b If $u = \langle X, \ldots, Z_k, Z, Y\rangle$ is a discriminating path between X and Y for Z, and $Z \circ\!\!-\!\!\ast Y$, then if $Z \notin Sepset(X,Y)$, then orient $Z_k \longleftrightarrow Z \longleftrightarrow Y$.

These seven rules are sound and complete: rule $\mathcal{R}$0b needs to be completed first, after which rules $\mathcal{R}1-\mathcal{R}4$b can be repeatedly fired in arbitrary order until no more new instances apply. Each invariant arrowhead $X \ast\!\!\rightarrow Y$ in the PAG $\mathcal{G}$ identified in this way, corresponds to the absence of causal relation $Y \not\Rightarrow X$ in the underlying causal DAG $\mathcal{G}_C$. In addition, also some invariant tails $Z \longrightarrow Y$ are found in rules $\mathcal{R}$1 and $\mathcal{R}$4a. Out of these, only the tails added by $\mathcal{R}$1 *definitely* imply the causal relation $Z \Rightarrow Y$; the tails added by $\mathcal{R}$4a, and indeed by any of the other tail orientation rules in [5], do not necessarily imply a causal relation, as they cannot exclude selection bias on (a descendant of) Z as the origin of the tail.

We refer to the combination of the original FCI-algorithm, together with the complete set of orientation rules, as the *extended FCI-algorithm.* In this article, however, we ignore the final tail orientation rules and the details behind the initial adjacency search, and only focus on the arrowhead orientation stage, as depicted in Algorithm 1.

The first part of the algorithm, lines $1-5$, aims to find the skeleton of the graph by eliminating edges between nodes X and Y that are separated by some set $\mathbf{Z}$. The 'in some clever way' not only has to do with avoiding as many unneccessary conditional independence tests as possible, but also with the fact that without causal sufficiency (i.e. with latent common causes a.k.a confounders) it is no longer sufficient to restrict the search for separating sets to the so called *Markov blanket* [11] of the nodes, see also [7]. The second part, lines $6-15$, is the orientation stage that identifies all invariant arrowheads in the output PAG, as well as some invariant tails. The repeated loop reflects the fact that orienting one edge mark may lead to subsequent orientations. In the next section we look at alternatives to this orientation procedure.

Input : independence oracle, fully $\circ\!\!-\!\!\circ$ connected graph over $\mathbf{V}$
Output : PAG $\mathcal{P}$
for all $\{X, Y\} \in \mathbf{V}$ **do**
  search *in some clever way* for a $X \perp\!\!\!\perp Y \mid \mathbf{Z}$
  if found, eliminate $X \ast\!\!\!\not\!-\!\!\ast Y$ from $\mathcal{P}$ (=$\mathcal{R}$0a)
    record $\mathbf{Z}$ in $Sepset(X, Y)$
end for
$\mathcal{R}$0b: orient $X \ast\!\!\rightarrow Z \leftarrow\!\!\ast Y$, iff $X \ast\!\!\!\not\!-\!\!\ast Y$ and $Z \notin Sepset(X, Y)$
repeat
  $\mathcal{R}1$: orient $Z \longrightarrow Y$, if $X \ast\!\!\!\not\!-\!\!\ast Y$ and $X \ast\!\!\rightarrow Z$
  $\mathcal{R}$2a: orient $Z \ast\!\!\rightarrow Y$, if $Z \longrightarrow X \ast\!\!\rightarrow Y$
  $\mathcal{R}$2b: orient $Z \ast\!\!\rightarrow Y$, if $Z \ast\!\!\rightarrow X \longrightarrow Y$
  $\mathcal{R}3$: orient $W \ast\!\!\rightarrow Z$, if $X \ast\!\!\rightarrow Z \leftarrow\!\!\ast Y$, $X \ast\!\!-\!\!\circ W \circ\!\!-\!\!\ast Y$ and $X \ast\!\!\!\not\!-\!\!\ast Y$
  if $\langle X, Z_1, \ldots, Z_k, Z, Y \rangle$ is a discriminating path for Z, then
    $\mathcal{R}$4a: orient $Z \longrightarrow Y$, iff $Z \in Sepset(X, Y)$
    $\mathcal{R}$4b: orient $Z_k \longleftrightarrow Z \longleftrightarrow Y$, iff $Z \notin Sepset(X, Y)$
until no new orientations found

Algorithm 1: FCI for invariant arrowheads

4 Causal relations from independence patterns

This section generalizes the result in theorem 1 of [12] to allow for the possibility of selection bias.

For that, let $\mathcal{G}_C$ be a causal DAG over variables $\mathbf{V}$, with hidden variables $\mathbf{L}$ and selection variables $\mathbf{S}$, so that $\mathbf{O} = \mathbf{V} \setminus (\mathbf{L} \cup \mathbf{S})$ is the set of observed variables. Let $\mathcal{G}'$ be the corresponding (marginalized) maximal ancestral graph of $\mathcal{G}_C$ over $\mathbf{O}$, so $\mathcal{G}' \equiv \mathcal{G}_C[^{\mathbf{S}}_{\mathbf{L}}$ in [4], and let X, Y, $\mathbf{Z}$ and W represent disjoint (subsets of) nodes (possibly empty) from $\mathbf{O}$. Then:

Lemma 1. An observed minimal conditional independence $[X \perp\!\!\!\perp Y \mid \mathbf{Z}]$ implies that every $Z \in \mathbf{Z}$ has a directed path to X, Y and/or $S \in \mathbf{S}$ in $\mathcal{G}_C$.[1]

Proof. We construct a directed path from an arbitrary $Z_1 \in \mathbf{Z}$ to either X, Y or $\mathbf{S}$ in $\mathcal{G}_C$. Similar to Theorem 1 in [13], Z_1 must be the (only) noncollider on some path π_1 in $\mathcal{G}'$ connecting X and Y given all the other nodes $(\mathbf{Z} \setminus Z_1) \cup \mathbf{S}$ (otherwise Z_1 would not be needed in $\mathbf{Z}$). If the edge along π_1 on either side of Z_1 is undirected, then there is a directed path from Z_1 to some $S \in \mathbf{S}$ in G_C (lemma 3.9 in [4] + acyclicity), and we are done. If there is no such undirected edge, then follow π_1 in the direction of the outgoing arrow (choose either branch if Z_1 has two outgoing arrows along π_1) until either X, Y or a collider on π_1 that is an ancestor of one of the remaining nodes in $\mathbf{Z} \setminus Z_1$, say Z_2, is encountered. (A collider that is ancestor of $\mathbf{S}$ would also be unblocked, but in this case the collider cannot have a directed path to $\mathbf{S}$, since that would imply that the initial arrow from Z_1 would have been an undirected edge. As we are constructing a strictly directed path this also holds for all other nodes encountered along that path in the remainder of this proof). If X or Y is found then a directed path has been found and we are done. If not then we can continue the directed path from the collider encountered on π_1 to its descendant node $Z_2 \in \mathbf{Z} \setminus Z$. This node in turn must be the only noncollider on some other path π_2 that m-connects X and Y given all nodes $\mathbf{Z} \setminus Z_2$. Again this path can be followed in the direction of the arrows until either X or Y or a collider that is ancestor of one of the nodes in $\mathbf{Z} \setminus \{Z_1, Z_2\}$ is encountered. (This cannot be one of the previous nodes since that would imply the existence of a directed path.) We can continue, and as long as neither X nor Y is reached we will find new nodes from $\mathbf{Z}$ until all have been encountered. At that point the final node will lie on a trek connecting X and Y that cannot be blocked by any other node in $\mathbf{Z}$, and therefore will have a directed path to X or Y. By construction, that means that if $Z_1 \notin An(\mathbf{S})$, then there is a directed path from Z_1 to either X or Y in $\mathcal{G}_C$. As Z_1 was chosen arbitrarily it holds for all $Z \in \mathbf{Z}$. □

[1] Many thanks to Peter Spirtes for pointing out that a similar observation was already made in [7] (corollary to lemma 14), although it was only used to prove correctness of the FCI-algorithm and never used as an orientation rule.

Lemma 2. A node W that creates the conditional dependence $X \not\perp\!\!\!\perp Y \,|\, \mathbf{Z} \cup W$ — so, without W, $X \perp\!\!\!\perp Y \,|\, \mathbf{Z}$ — implies that there are no directed paths from W to X, Y, $\mathbf{Z}$ or $\mathbf{S}$ in $\mathcal{G}_C$.

Proof. Node W must be a (descendant of a) collider on some path π m-connecting X and Y given $\mathbf{Z} \cup W$, otherwise it would not be needed to unblock the path. Any directed path in $\mathcal{G}_C$ from W to a node in $\mathbf{Z} \cup \mathbf{S}$ implies that W is not needed, for then conditioning on W cannot unblock a collider that is not already unblocked by conditioning on $\mathbf{Z} \cup \mathbf{S}$. No directed paths from W to $\mathbf{Z}$ implies that if there existed a directed path from W to X or Y, then it could not be blocked by any $Z \in \mathbf{Z}$. But then such a path would make W a noncollider on an unblocked path between X and Y given $\mathbf{Z}$: starting from X, let θ_X be the first collider encountered along π that is unblocked by conditioning on W, and similarly θ_Y the first collider along π starting from Y, (possibly $\theta_X = \theta_Y$, but $\{\theta_X, \theta_Y\} \notin \mathbf{Z}$ (otherwise W not needed)); then the paths $\langle X, \theta_X, W \rangle$ and $\langle W, \theta_Y, Y \rangle$ are into W and unblocked given $\mathbf{Z} \cup W$, so a directed path $W \Rightarrow X$ would make W a noncollider on unblocked path $\langle X, \theta_X, W, Y \rangle$ given $\mathbf{Z}$, contradicting $X \perp\!\!\!\perp Y \,|\, \mathbf{Z}$; idem for $W \Rightarrow Y$. □

Together, the two rules allow to infer causal relations, even in the presence of selection bias:

Corollary 3. In a causal DAG $\mathcal{G}_C$, if there is a minimal conditional independence $[X \perp\!\!\!\perp Y \,|\, \mathbf{Z}]$, and a conditional dependence $X \not\perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z$ created by some $Z \in \mathbf{Z}$, then there is a directed path $Z \Rightarrow Y$ in $\mathcal{G}_C$.
Proof. Analogous to Lemma 1 in [12]: the minimal conditional independence $[X \perp\!\!\!\perp Y \,|\, \mathbf{Z}]$ implies directed paths from Z to X, Y and/or $\mathbf{S}$ (lemma 1), of which both options $Z \not\Rightarrow X$ and $Z \not\Rightarrow \mathbf{S}$ are eliminated by $X \not\perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z$ (lemma 2), leaving only $Z \Rightarrow Y$.

Finally, the standard result for the exclusion of direct interactions between nodes, i.e. not mediated by any of the other observed nodes in $\mathcal{G}_C$, still applies when selection bias may be present:

Lemma 4. A (conditional) independence $X \perp\!\!\!\perp Y \,|\, \mathbf{Z}$ implies the absence of direct causal paths $X \Rightarrow Y$ or $X \Leftarrow Y$ in $\mathcal{G}_C$ between X and Y that are not mediated by nodes in $\mathbf{Z}$.

Proof. From the faithfulness assumption, with or without selection bias, the existence of a directed path between X and Y not via any node in $\mathbf{Z}$ would result in $X \not\perp\!\!\!\perp Y \,|\, \mathbf{Z}$, see [2].

It is used to eliminate direct links between variables. For empty $\mathbf{Z}$ it implies absence of any causal path between X and Y. This lemma relies on the assumption that there are no external interventions on nodes in the underlying causal DAG $\mathcal{G}_C$ that effectively block any causal influence from other nodes, see section 2.2.

5 Arrowheads from minimal independencies

This section derives the main result of this report: that the seven graphical orientation rules, $\mathcal{R}$0b$-\mathcal{R}$4b in Figure 1, for inferring all invariant arrowheads in (the PAG of) a causal DAG are, in fact, different manifestations of just two rules, that both start from a minimal conditional independence

Theorem 1. In a PAG $\mathcal{G}$, all invariant arrowheads $X \ast\!\!\rightarrow Y$ are instances of rules

(1) $U \not\perp\!\!\!\perp V \,|\, \mathbf{W} \cup Y$, created by Y from a *minimal* $[U \perp\!\!\!\perp V \,|\, \mathbf{W}]$, with $X \in (U \cup V \cup \mathbf{W})$, or

(2) a minimal $[Y \perp\!\!\!\perp Z \,|\, \mathbf{W} \cup X]$, with an arrowhead at $Z \ast\!\!\rightarrow X$ from either rule (1) or rule (2).

Proof.sketch. Both rules are sound, as they are direct applications of Lemmas 1 and 2. The proof that they are also complete follows from the lemmas below, by induction on the graphical orientation rules $\mathcal{R}$0b$-\mathcal{R}$4b, showing that none of them introduces a violation of Theorem 2. As these rules are sufficient for arrowhead completeness [10, 5], it follows that the theorem holds for all invariant arrowheads. □

For the remainder of the proof, we show that in each case, if the antecedent of one of the graphical orientation rules $\mathcal{R}0b-\mathcal{R}4b$, see Figure 1, is satisfied, then the antecedent of rules (1) and/or (2) is also satisfied, and the consequent of rule (1)/(2) then matches the consequent of the graphical rule.

Lemma 5. The arrowheads at Z from rules $\mathcal{R}0b$, $\mathcal{R}3$, and $\mathcal{R}4b$ are covered by rule (1) and the arrowhead at Y from rule $\mathcal{R}1$ is covered by rule (2).

Proof. Implied directly by the corresponding patterns in Figure 1:

- $\mathcal{R}0b$: If this rule fires, then it implies $[X \perp\!\!\!\perp Y \,|\, \mathbf{W}]$ for some set $\mathbf{W}$ (possibly empty), with $X \not\perp\!\!\!\perp Y \,|\, \mathbf{W} \cup Z$. Therefore rule (1) applies and Z gets arrowheads on the edges from X and Y in $\mathcal{G}$, just as in the consequent of $\mathcal{R}0b$ in Figure 1.
- $\mathcal{R}1$: Implies $[X \perp\!\!\!\perp Y \,|\, \mathbf{Z}]$ with $Z \in \mathbf{Z}$ and an arrowhead at Z from a rule that fired before. If no violations before $\mathcal{R}1$ fires, then rule (2) applies, and there is an arrowhead at $Z \longrightarrow Y$ in $\mathcal{G}$, just as in the consequent of $\mathcal{R}1$.
- $\mathcal{R}3$: Implies $[X \perp\!\!\!\perp Y \,|\, \mathbf{W}]$ with $W \in \mathbf{W}$, and $X \not\perp\!\!\!\perp Y \,|\, \mathbf{W} \cup Z$. Therefore rule (1) applies, to give $W *\!\!\rightarrow Z$ in $\mathcal{G}$, just as in $\mathcal{R}3$.
- $\mathcal{R}4b$: By construction of the discriminating path, $\mathcal{R}4b$ implies $[X \perp\!\!\!\perp Y \,|\, \mathbf{Z}]$, with $\{Z_1, \ldots, Z_k\} \in \mathbf{Z}$, but $Z \notin \mathbf{Z}$ as $X \not\perp\!\!\!\perp Y \,|\, \mathbf{Z} \cup Z$. Therefore rule (1) applies, resulting in the addition of $Z_k *\!\!\rightarrow Z \leftarrow\!\!* Y$ to $\mathcal{G}$, just as in $\mathcal{R}4b$.

□

Lemma 6. The arrowheads at Y from rules $\mathcal{R}2b$, $\mathcal{R}4a$, and $\mathcal{R}4b$ are covered by rules (1) and (2).

Proof. If no violations before $\mathcal{R}2b$ fires, then the arrowhead at $Z *\!\!\rightarrow X$ either appeared by rule (1) as a node X that creates the dependency $U \not\perp\!\!\!\perp V \,|\, \mathbf{W} \cup X$ from $[U \perp\!\!\!\perp V \,|\, \mathbf{W}]$, with $Z \in (U \cup V \cup \mathbf{W})$ (case 1), or by rule (2), as a minimal conditional independence $[X \perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z]$, with an already established arrowhead at $U *\!\!\rightarrow Z$, for which either U and Y are also independent given $\mathbf{W} \cup Z$ (case 2a), or not (case 2b).
Note: $Y \notin \mathbf{W}$, otherwise path $\langle X, Y, Z, U \rangle$ would be unblocked; also: Y is a descendant of X.

For the three cases:

1) If conditioning on X creates $U \not\perp\!\!\!\perp V \,|\, \mathbf{W} \cup X$, then conditioning on Y as a descendant of X implies $U \not\perp\!\!\!\perp V \,|\, \mathbf{W} \cup Y$, and so rule (1) also applies to $Z *\!\!\rightarrow Y$.

2a) If $Y \perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z$, then also $[Y \perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z]$, as no subset can block the path between Y and U via X, and so rule (2) applies to $Z \longrightarrow Y$.

2b) If $Y \not\perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z$, then there is an unblocked path π between U and Y given $\mathbf{W} \cup Z$. The path π is *into* Y, since otherwise the path $\langle X, Y \rangle + \pi$ would be an unblocked path between X and U given $\mathbf{W} \cup Z$, contrary to $[X \perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z]$. Therefore, conditioning on collider Y on the path creates the dependency $X \not\perp\!\!\!\perp U \,|\, \mathbf{W} \cup Z \cup Y$, and so rule (1) applies.

In $\mathcal{R}4a$ and $\mathcal{R}4b$, the arrowhead at Y is simply an instance of $\mathcal{R}2b$ with $Z_k = X$. □

This leaves rule $\mathcal{R}2a$ as the only remaining case to prove. For that we use three intermediate results. First we have the observation:

Lemma 7. If two nodes X and Y are conditionally independent given a set of nodes $\mathbf{Z}$, $X \perp\!\!\!\perp Y \,|\, \mathbf{Z}$, then an arbitrary node V can be:

(a) part of the conditional independence, i.e. $V \in (X \cup Y \cup \mathbf{Z})$,

(b) conditionally independent of X and/or Y given $\mathbf{Z}$, i.e. $(V \perp\!\!\!\perp X \,|\, \mathbf{Z}) \vee (V \perp\!\!\!\perp Y \,|\, \mathbf{Z})$, or

(c) (descendant of) a collider between U and V, such that $X \not\perp\!\!\!\perp Y \,|\, \mathbf{Z} \cup V$.

Proof. If neither (a) nor (b), i.e. $V \notin (X \cup Y \cup \mathbf{Z})$ and $V \not\perp\!\!\!\perp \{X, Y\} \mid \mathbf{Z}$, then there are paths $\pi_X = \langle X, \ldots, V\rangle$ and $\pi_Y = \langle Y, \ldots, V\rangle$ in the corresponding graph that are unblocked given $\mathbf{Z}$. Node V has to be a collider on the path $\pi = \pi_X + \pi_Y$, otherwise π would be unblocked given $\mathbf{Z}$ (as $V \notin \mathbf{Z}$), which would make X and Y dependent, contrary to $X \perp\!\!\!\perp Y \mid \mathbf{Z}$. But then conditioning on $\mathbf{Z} \cup V$ will make them dependent, i.e. then (c). □

Note that if $\mathbf{Z}$ is a *minimal* set that makes X and Y independent, then case (b) does not imply that it is also minimal for $V \perp\!\!\!\perp (X/Y) \mid \mathbf{Z}$, as shown by the example in Figure 2: from $[X \perp\!\!\!\perp Y \mid \{Z_1, Z_2\}]$, for node V we find $V \perp\!\!\!\perp X \mid \{Z_1, Z_2\}$, as none of the other options in lemma 7 applies, ... but this is only *minimal* for subset $[V \perp\!\!\!\perp X \mid Z_2]$.

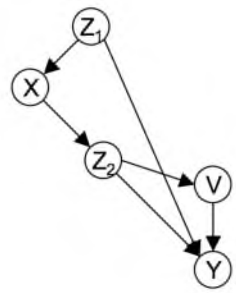

Figure 2: Example of case (b) in lemma 7 with 'minimal' only for subset

In the proof of $\mathcal{R}2$a we also use:

Lemma 8. In an ancestral graph $\mathcal{G}$, if a node Z unblocks a blocked path $\pi = \langle U, \ldots, V\rangle$ between two nodes U and V given some set $\mathbf{W}$, then there are unblocked paths from both U and V into Z relative to $\mathbf{W}$, and so $Z \not\perp\!\!\!\perp \{U, V\} \mid \mathbf{W}$.

Proof. A path π is unblocked given $\mathbf{W}$ if all non-colliders on π are not in $\mathbf{W}$ and all colliders are in $An(\mathbf{W})$. As $An(\mathbf{W}) \subseteq An(\mathbf{W} \cup Z)$, adding a node Z can only unblock on a collider, so a blocked path π can only be unblocked by conditioning on a node Z that is (a descendant of) a collider X on the path (possibly $X = Z$). No node $W \in \mathbf{W}_{\setminus X}$ blocks the path $X \Rightarrow Z$ (otherwise conditioning on Z would not be needed), therefore if π is unblocked relative to $\mathbf{W}$, then so are the two paths $\pi_U = \langle U, ..(, X, ..), Z\rangle$ and $\pi_V = \langle Z, (.., X,).., V\rangle$, which implies $Z \not\perp\!\!\!\perp \{U, V\} \mid \mathbf{W}$. □

Finally, we need the following result:

Lemma 9. In an ancestral graph $\mathcal{G}$, if there is an edge $Y * - * Z$, and there are nodes such that $Z \perp\!\!\!\perp U \mid \mathbf{W}$ and $Z \not\perp\!\!\!\perp U \mid \mathbf{W} \cup Y$, then there is a node $V \in (U \cup \mathbf{W})$ such that $[Z \perp\!\!\!\perp V \mid \mathbf{Q}]$ and $Z \not\perp\!\!\!\perp V \mid \mathbf{Q}' \cup Y$. In words, if conditioning on a node Y destroys (unblocks) some conditional independence for a neighbouring node Z, then the same holds for at least some *minimal* conditional independence between Z and one of the other nodes involved.

Proof. By definition, there is a $\mathbf{W}' \subseteq \mathbf{W}$ such that $[U \perp\!\!\!\perp Z \mid \mathbf{W}']$. If then also $U \not\perp\!\!\!\perp Z \mid \mathbf{W}' \cup Y$, then the lemma applies with $W = U$ and $\mathbf{Q} = \mathbf{W}'$. If not, i.e. if $U \perp\!\!\!\perp Z \mid \mathbf{W}' \cup Y$, then we can show that there is a node $W \in \mathbf{W}$ for which the lemma holds. From the original $U \not\perp\!\!\!\perp Z \mid \mathbf{W} \cup Y$, by lemma 8, there is an unblocked path $\pi = \langle U, \ldots, Y\rangle$ into Y given $\mathbf{W}$. The path π contains one or more (say k) colliders, some of which are (ancestors of) nodes from $\mathbf{W}$, but not from $\mathbf{W}'$ (otherwise the path to Y would also be unblocked given $\mathbf{W}'$, which, together with edge $Z * \!\!\rightarrow Y$, would imply $U \not\perp\!\!\!\perp Z \mid \mathbf{W}' \cup Y$, contrary to the assumed). Number the colliders as $W_1, \ldots, W_k$, as they are encountered along π when starting from Y, such that $\pi = U * \!\!\rightarrow W_k \longleftrightarrow \ldots \longleftrightarrow W_2 \leftarrow \rightarrow W_1 \longleftrightarrow Y$ (ignoring/marginalizing other, intermediate nodes on the path, that are not in $\mathbf{W}$). By induction: if there is no unblocked trek between W_1 and Z given $\mathbf{W}$, then they are (minimally) conditionally independent given some set $\mathbf{Q}_1 \subset \mathbf{W}$ (possibly empty), but dependent given Y, as the paths from both W_1 and Z into Y are not blocked by any node from $\mathbf{W}$ (as a 'colliderless' subpath of unblocked path π, resp. direct edge to Y), and so the lemma is satisfied. If not, i.e. if there is an unblocked trek between W_1 and Z given $\mathbf{W}$, then this trek is *out of* W_1, otherwise the path $\langle U, W_k, .., W_1, Z\rangle$ would be unblocked relative to $\mathbf{W}$, making U and Z dependent given $\mathbf{W}$, contrary the given. But then for W_2, if there is no unblocked trek to Z, then $[W_2 \perp\!\!\!\perp Z \,|\, \mathbf{Q}_2]$, with $W_1 \in \mathbf{Q}_2$, because it is the only node from $\mathbf{W}$ that blocks the trek $W_2 \longleftrightarrow W_1 \longrightarrow Z$. But that also means that the path from W_2 to Y is unblocked given $\mathbf{Q}_2$, and so $W_2 \not\!\perp\!\!\!\perp Z_2 \,|\, \mathbf{Q}_2 \cup Y$. If not, then the unblocked trek to Z is (again) *out of* W_2, otherwise $U \not\!\perp\!\!\!\perp Z \,|\, \mathbf{W}$, contrary the given. This applies to all successive colliders W_i on the path π. But if all, up to and including W_k, have an unblocked trek into Z, then no unblocked path between U and Z implies that W_k is needed to block $U \ast\!\!\rightarrow W_k \longrightarrow Z$, and so *all* W_i on π are in $\mathbf{W}'$, implying an unblocked path to Y, and so $U \not\!\perp\!\!\!\perp Z \,|\, \mathbf{W}' \cup Y$.

We can now show:

Lemma 10. The arrowhead at Y from $\mathcal{R}$2a is covered by rules (1) and (2).

Proof. Assuming no violations before $\mathcal{R}$2a fires, then if the arrowhead at $X \ast\!\!\rightarrow Y$ originates from rule (2), then the edge appears as $X \longrightarrow Y$, and is therefore also an instance of $\mathcal{R}$2b, which we already found to be valid. If $X \ast\!\!\rightarrow Y$ originates from rule (1), then there is a minimal $[U \perp\!\!\!\perp V \,|\, \mathbf{W}]$, with $X \in (U \cup V \cup \mathbf{W})$, and the node Y creates $U \not\!\perp\!\!\!\perp V \,|\, \mathbf{W} \cup Y$. By lemma 7 there are now three cases for node Z:

(a) $Z \in (U \cup V \cup \mathbf{W})$,

(b) $Z \perp\!\!\!\perp U \,|\, \mathbf{W}$, (and/or $Z \perp\!\!\!\perp V \,|\, \mathbf{W}$)

(c) $U \not\!\perp\!\!\!\perp V \,|\, \mathbf{W} \cup Z$.

For case (a), both X and Z are in $(U \cup V \cup \mathbf{W})$, and so if rule (1) applies to $X \ast\!\!\rightarrow Y$, it also applies to $Z \ast\!\!\rightarrow Y$. Case (c) cannot occur, as that would imply $Z \not\Rightarrow (U \cup V \cup \mathbf{W})$ by lemma 2, while $\mathcal{R}$2a has $Z \longrightarrow X$, with $X \in (U \cup V \cup \mathbf{W})$.
For the remaining case (b), w.l.o.g. we assume $Z \perp\!\!\!\perp U \,|\, \mathbf{W}$. Then from lemma 9 it follows that there is at least one minimal conditional independence between Z and some node from $(U \cup \mathbf{W})$ that is destroyed by conditioning on Y. Therefore, the arrowhead $Z \ast\!\!\rightarrow Y$ is then covered by rule (1). □

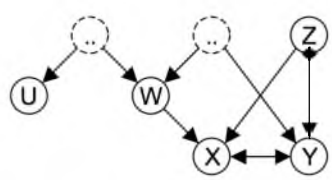

Figure 3: Example of non-minimal case (b) in lemma 10.

Example. Figure 3 shows an instance of case (b) for $\mathcal{R}$2a where the initial separating set is not minimal. Here, $\mathcal{R}$2a applies to $Z \ast\!\!\rightarrow Y$, after $X \ast\!\!\rightarrow Y$ is derived via rule (1) from $[U \perp\!\!\!\perp X \,|\, W]$ with $U \not\!\perp\!\!\!\perp X \,|\, W \cup Y$ (the origin of the edge $Z \longrightarrow X$ is not depicted). By lemma 7, for node Z indeed $U \perp\!\!\!\perp Z \,|\, W$ holds (case b), but not as a minimal independence, as $[U \perp\!\!\!\perp Z \,|\, \varnothing]$. As a result, edge $Z \ast\!\!\rightarrow Y$ does not follow from rule (1) applied to this combination of nodes as conditioning on Y does not make U and Z dependent, i.e. $U \perp\!\!\!\perp Z \,|\, Y$. However, as in the proof of lemma 10, Z is minimally conditionally independent of 'eliminated' node W, but dependent when conditioning on Y. Therefore, rule (1) applies to $[W \perp\!\!\!\perp Z \,|\, \varnothing]$ and $W \not\!\perp\!\!\!\perp Z \,|\, Y$, from which follows that $Z \ast\!\!\rightarrow Y$.

We can now complete the proof of the main result:

Proof of theorem 1. Follows from the arrowhead completeness of rules $\mathcal{R}$0b-$\mathcal{R}$4b, the fact that after $\mathcal{R}$0a the theorem holds (no arrowheads), in combination with the proof in lemmas 5-10 that none of the rules $\mathcal{R}$0b-$\mathcal{R}$4b introduces a violation of the theorem, if there was no violation prior to firing of the rule. □

6 Extensions

We can extend the previous results to improve scope and efficiency of the causal relations identified from minimal conditional (in)dependencies. First we note that, for each pair of nodes $\{X, Y\}$, the standard implementation of the FCI-algorithm already finds only one minimal conditional independence, see lines 1-5 of Algorithm 1. The next results shows that this is also sufficient to find all invariant arrowheads in our new approach.

Lemma 11. For each pair of nodes $\{X, Y\}$ in the graph, finding a *single* minimal separating set $\mathbf{Z}$, i.e. such that $[X \perp\!\!\!\perp Y \,|\, \mathbf{Z}]$ (if it exists), is sufficient to orient *all* invariant arrowheads, using the two rules in Theorem 1.

Proof sketch. This stems from the fact that the graphical orientation rules are defined on sets of adjacent nodes, which ensures that most nodes are almost always needed to separate two nonadjacent nodes in the same rule, and so will be found as part of the separating set, no matter how large/variable the set of nodes to block all paths between the two can be. Note: once a minimal set is found for a pair of nodes, then all remaining nodes are checked to see if including them destroys the independence (so rule (1) applies). □

The essential steps in the detailed proof per rule are very similar to that for theorem 1. Again, $\mathcal{R}2$a turns out to be the only 'difficult' one.

For that, we use the following result:

Lemma 12. In an ancestral graph $\mathcal{G}$, if there are (sets of) nodes U, Y, Z and $\mathbf{W}$, such that $[U \perp\!\!\!\perp Z \,|\, \mathbf{W}]$ and $U \not\perp\!\!\!\perp Z \,|\, \mathbf{W} \cup Y$, with $Z *\!\!\rightarrow Y$ in $\mathcal{G}$, then there is a node V (possibly $V = U$), such that for all sets $\mathbf{Q}$ for which $[V \perp\!\!\!\perp Z \,|\, \mathbf{Q}]$ it holds that $V \not\perp\!\!\!\perp Z \,|\, \mathbf{Q} \cup Y$.
In words: if conditioning on a node Y destroys (unblocks) some minimal conditional independence for a neighbouring node Z, then it does so in all minimal independencies between Z and at least one node in $\mathcal{G}$.

Proof. Follows along the lines of lemma 9. We show that if lemma 12 does not hold for U, then we can (repeatedly) find another node V to which the lemma can be applied in turn, until we eventually find one that satisfies the lemma. Note: in this proof $U *\!\!\rightarrow Y$ denotes any (unblocked) trek in $\mathcal{G}$ between U and Y that is into Y, i.e. not necesarily an edge (except for the neighbors $Z *\!\!\rightarrow Y$). Similar for $U \longleftrightarrow Y$ etc.
From the proof of lemma 9, there is an unblocked path $\pi = U *\!\!\rightarrow W_k \longleftrightarrow \ldots \longleftrightarrow W_2 \longleftrightarrow W_1 \longleftrightarrow Y$ that is into Y given $\mathbf{W}$, for some $k \geq 0$. If any of the nodes W_i does not have an unblocked (trek) path $W_i \longrightarrow Z$ relative to $\mathbf{W}$, then let V be the first such W_i encountered along π when starting from Y, so that $[V \perp\!\!\!\perp Z \,|\, \mathbf{W}']$ and $V \not\perp\!\!\!\perp Z \,|\, \mathbf{W}' \cup Y$, for some $\mathbf{W}' \subsetneq \mathbf{W}$ (in $\mathbf{W}$, all nodes up to W_i are needed to make $V = W_i$ independent of Z, which also ensures an unblocked path from V to Y; also note that unblocked treks relative to $\mathbf{W}$ imply unblocked treks relative to $\mathbf{W}'$), and so the lemma can be applied with V instead of U and $\mathbf{W}'$ instead of $\mathbf{W}$.
If not, (so all W_i have an unblocked trek $W_i \longrightarrow Z$ relative to $\mathbf{W}$), then let V be the first node in $\mathcal{G}$, encountered along π when starting from U, that is *not* in $\mathbf{W}$. If there is no such V, then no node in $\mathcal{G}$ can block the path π, and all nodes W_i are needed in all sets that make U and Z independent, ensuring the path to Y is unblocked, so conditioning on Y will destroy the independence, ergo satisfying the lemma with $V = U$. If V is encountered somewhere between $W_{i+1} \leftarrow\!\!* V *\!\!\rightarrow W_i$, (with $i \geq 1$, $W_{k+1} = U$), then V cannot have an unblocked path to Z given $\mathbf{W}$, because then, i.e. with $V \notin \mathbf{W}$, if the path to Z is *out of.* V, that would imply an unblocked path $U *\!\!\rightarrow \ldots \longleftrightarrow W_{i+1} \leftarrow\!\!* V \longrightarrow Z$, or, if the path to Z is *into* V, it would imply an unblocked $U *\!\!\rightarrow \ldots \longleftrightarrow W_{i+1} \longleftarrow V \leftarrow\!\!* Z$ or an unblocked $U *\!\!\rightarrow \ldots \longleftrightarrow W_{i+1} \longleftrightarrow V \leftarrow\!\!* Z$ (as then collider V is an ancestor of W_i), contrary to $[U \perp\!\!\!\perp Z \,|\, \mathbf{W}]$. Therefore, V is minimally conditionally independent of Z given some subset $\mathbf{W}' \subseteq \mathbf{W}$, that includes all $\{W_i, .., W_1\}$ (because of the unblocked treks $W_i \longrightarrow Z$), and so also has an unblocked path to Y. Therefore lemma 12 can be applied to this V with $[V \perp\!\!\!\perp Z \,|\, \mathbf{W}']$ and $V \not\perp\!\!\!\perp Z \,|\, \mathbf{W}' \cup Y$.
That leaves the case where V is encountered on the path $W_1 \leftarrow\!\!* V *\!\!\rightarrow Y$ (possibly $W_1 = U$). If V does not have an unblocked trek to Z relative to $\mathbf{W}$, then V is (minimally) conditionally independent of Z given some subset $\mathbf{W}' \subseteq \mathbf{W}$, and we can again apply lemma 12 with V in place of U. If V does have an unblocked trek to Z relative to $\mathbf{W}$, then this trek has to appear in the form $W_1 \longleftrightarrow V \leftarrow\!\!* Z$, otherwise, if either of the unblocked treks $W_1 \leftarrow\!\!* V$ and $V *\!\!-\!\!* Z$ is out of V, then

the path $U *\!\!\rightarrow \ldots \longleftrightarrow W_1 \leftarrow\!\!* V \longrightarrow Z$ would be unblocked, contrary to $[U \perp\!\!\!\perp Z \mid \mathbf{W}]$. So both treks $W_1 \longleftrightarrow V \leftarrow\!\!* Z$ are into V. Now, if V is part of *some* minimal independence between U and Z, then by lemma 1, V has a directed path to U and/or Z. But V cannot have a directed path to U, for then the trek $U \longleftarrow V \leftarrow\!\!* Z$ must be blocked by some node from $\mathbf{W}$, otherwise U and Z would be dependent without V, which in turn would unblock $W_1 \longleftrightarrow V \leftarrow\!\!* Z$, as V would be a collider that is an ancestor of $\mathbf{W}$, contrary to $[U \perp\!\!\!\perp Z \mid \mathbf{W}]$. But then V also cannot have a directed path to Z, as that would have to be blocked by some node from $\mathbf{W}$ as well, otherwise $U \ldots W_1 \longleftrightarrow V \longrightarrow\!\!* Z$ would be unblocked, which in turn would unblock $W_1 \longleftrightarrow V \leftarrow\!\!* Z$, as collider V is again ancestor of some node in $\mathbf{W}$, making U and Z dependent given $\mathbf{W}$, contrary the initial assumption. Therefore, if both treks $W_1 \longleftrightarrow V \leftarrow\!\!* Z$ are into V, then V is *not* part of any minimal conditional independence between U and Z, and so cannot prevent the original $[U \perp\!\!\!\perp Z \mid \mathbf{Q}]$ with $U \not\perp\!\!\!\perp Z \mid \mathbf{Q} \cup Y$ being found, with $\forall \mathbf{Q} : \{W_1, ..W_k\} \subseteq \mathbf{Q}$, guaranteeing the unblocked paths to Y, and so the lemma applies with $V = U$.
This recursive application of lemma 12 to nodes, conditionally independent of Z, that are closer and closer to Y along the initial path π can continue at most until there is a direct edge $U *\!\!\rightarrow Y$ in $\mathcal{G}$: at that point, for any set $\mathbf{Q}$ that (minimally) separates U and Z, conditioning on $\mathbf{Q} \cup Y$ will make them dependent, and so the lemma is satisfied. □

It turns out it is actually easier to use a more restricted variant of theorem 1 to prove lemma 11:

Lemma 13. In a PAG $\mathcal{P}$, all invariant arrowheads $Z *\!\!\rightarrow Y$ are instances of

(1′) $U \not\perp\!\!\!\perp V \mid \mathbf{W} \cup Y$, created from $[U \perp\!\!\!\perp V \mid \mathbf{W}]$, with $Z \in (U \cup V \cup \mathbf{W})$, and where for all sets $\mathbf{W}' : [U \perp\!\!\!\perp V \mid \mathbf{W}']$ the paths from U and V to Y are unblocked relative to $\mathbf{W}'$, and either $Z \in \{U, V\}$ or (necessarily) $Z \in \mathbf{W}'$.

(2′) $[X \perp\!\!\!\perp Y \mid \mathbf{W}]$ with $Z \in \mathbf{W}$, and $Z \not\Rightarrow (X \cup \mathbf{S})$ from either case (1′) or case (2′), and Z in all sets $\mathbf{W}' : [X \perp\!\!\!\perp Y \mid \mathbf{W}']$.

In words: case (1) only needs to be applied to instances where Z is always part of the minimal conditional independence, and that will always be unblocked when conditioning on Y. Case (2) also only needs to be applied to instances where Z is always part of the minimal conditional independence.

Proof. For each arrowhead rule:

- $\mathcal{R}$0b fires on any (minimal) conditional independence $X \perp\!\!\!\perp Y \mid \mathbf{W}$ between X and Y, and for any such $\mathbf{W}$, including Z will unblock the path $\langle X, Z, Y \rangle$, so case (1′) applies,
- $\mathcal{R}$1 node Z is part of *any* set (minimal or not) that separates X and Y, and so case (2′) applies,
- $\mathcal{R}$3 similar to $\mathcal{R}$0b, fires on a node W that is part of all sets separating X and Y, and including Z will unblock the path $\langle X, Z, Y \rangle$, and so case (1′) applies,
- $\mathcal{R}$4b (arrowheads at Z) all nodes $Z_1, \ldots, Z_k$ are part of all sets separating X and Y, and including Z then makes them dependent, so case (1′) applies,
- $\mathcal{R}$2b for instance (1) in lemma 6, if case (1′) applies to $Z *\!\!\rightarrow X$, then it also applies to $Z *\!\!\rightarrow Y$, as X is never part of the minimal conditional independence involving Z, and so unblocked paths to X imply unblocked paths to Y; for instance (2a), Z is present in all sets that make X and U independent, and so also in all sets that make Y and U independent (as it implies $Z \longrightarrow Y$), and so case (2′) applies; for instance (2b), if $Y \not\perp\!\!\!\perp U \mid \mathbf{W} \cup Z$ holds for all sets for which $[X \perp\!\!\!\perp U \mid \mathbf{W} \cup Z]$, then it is an instance of case (1′) with $V = X$. If not, then there is *some* $\mathbf{W}'$ for which $[X \perp\!\!\!\perp U \mid \mathbf{W}' \cup Z]$ and *not* $Y \not\perp\!\!\!\perp U \mid \mathbf{W}' \cup Z$. But as Z is needed in all sets that block a path $\pi = U.. *\!\!\rightarrow Z \longrightarrow X$ between U and X, it means that Z is also needed in all sets that separate U and Y, because if there is any remaining unblocked path π from U to either X or Z, then $\pi +$ either $X \longrightarrow Y$ or $Z \longrightarrow Y$ is an unblocked path from U into Y. Therefore Z is also needed in all sets that separate U and Y, which, together with $Z \not\Rightarrow (U \cup \mathbf{S})$, implies that it is an instance of case (2′).
- $\mathcal{R}$4a/b (arrowhead at Y) instances of $\mathcal{R}$2b with $Z_k = X$,
- $\mathcal{R}$2a if the arrowhead between X and Y originates from case (2′) then $X \longrightarrow Y$, and so is instance of $\mathcal{R}$2b; if from case (1′) and it holds before, then instance (c) in lemma 7 still cannot occur. The proof of instance (b) (lemma 9) shows that Z appears as one of the separated nodes in *some* minimal conditional independence $[W \perp\!\!\!\perp Z \mid \mathbf{Q}]$ that is destroyed by conditioning on Y, and so is an instance of lemma 12, which implies that it satisfies case (1′). Finally, if instance (a) (in lemma 7), and arrowhead $X *\!\!\rightarrow Y$ results from case (1′), i.e. X appears in all $\mathbf{W} : [U \perp\!\!\!\perp V \mid \mathbf{W}]$, and for all of these there are unblocked paths

from U and V into Y relative to $\mathbf{W}$, then if $Z \in \{U, V\}$ or Z appears in all sets $\mathbf{W}$, then both X and Z satisfy the definition of case $(1')$ for Z. If not, i.e. if Z does not appear in *all* minimal sets $\mathbf{W}$ that can separate U and V, then Z cannot have edges to both U and V (otherwise it would be needed in all sets, because it cannot be a collider between them, as $Z \longrightarrow X$ and X has a directed path to at least U or V, by lemma 1). Therefore, in that case (by lemma 7), Z is conditionally independent of U or V given some $\mathbf{W}$, and so (lemma 9) also minimally conditionally independent, and so (lemma 12) also in some combination that satisfies case $(1')$. □

(Note that $\mathcal{R}$2b is not simply an instance of lemma 12, as that only says that X has a conditional independence with some node $V \in \mathbf{W}$ that will always be destroyed by conditioning on Y. It does not guarantee that Z is also always a part of this set.)

So all arrowhead rules are covered by lemma 13. As the two cases in lemma 13 are just a restricted form of the cases in theorem 1, it follows that all rules are also covered by 1 if just a (any) single minimal independence is found between each pair (if it exists).

The rules in theorem 1 already allow identification of many causal relations $X \Rightarrow Y$ and absence thereof. However, sometimes, we can (only) identify a causal relation indirectly from two or more steps. As before, we assume that $\mathcal{G}_C$ is the causal DAG over a set of nodes $\mathbf{V}$, possibly including latent variables and selection variables, and that $\mathcal{P}$ is the corresponding completed PAG over the subset of observed variables.

Lemma 14. The transitive closure of the invariant arcs found by rule (2), all correspond to identifiable, definite causal relations in the underlying causal DAG $\mathcal{G}_C$.

Proof. By corollary 3, all arrowheads $X *\!\!\rightarrow Y$ found by rule (2) in Theorem 1 correspond to definite causal relations $X \Rightarrow Y$. The binary relation $X \Rightarrow Y$ is *transitive*, as $X \Rightarrow Y$ and $Y \Rightarrow Z$ also implies that $X \Rightarrow Z$ (there is at least a directed path from X to Z via Y in the underlying DAG $\mathcal{G}_C$). Therefore all elements in the transitive closure also correspond to definite causal relations. □

Note: this lemma corresponds to orientation rule $\mathcal{R}$8a in [5].

All results we have obtained so far do not depend on the structure of the graph, that is, they apply to arbitrary sets of nodes, irrespective of the graphical structure between the nodes in the PAG. In particular, they do not need to be adjacent. Nevertheless, if we record all instances identified by rules (1) and (2) in Theorem 1, and then apply corollary 3 and lemma 14 to infer (absence of) causal relations, then it is straightforward to transfer this structure independent information on relations between pairs of nodes directly to *edges* in the graph $\mathcal{P}$.

Lemma 15. Let $(X, Y; \mathbf{Z}) \in \mathbf{S}_{CI}$, record the fact that all $Z \in \mathbf{Z}$ occur in *some* minimal (conditional) independence for two nodes X and Y, and let $\mathbf{M}_C$ contain all recorded explicit (absence of) causal relations $X \Rightarrow Y$ in $\mathcal{G}_C$, then

- $(X, Y; *) \in \mathbf{S}_{CI}$ implies that $X \not\ast\!\!-\!\!\ast Y$ in $\mathcal{P}$,
- $(X, Y; *) \notin \mathbf{S}_{CI}$ and $(X \Rightarrow Y) \in \mathbf{M}_C$ implies that $X \,{-}\!\!\ast\, Y$ in $\mathcal{P}$,
- $(X, Y; *) \notin \mathbf{S}_{CI}$ and $(Y \not\Rightarrow X) \in \mathbf{M}_C$ implies that $X *\!\!\rightarrow Y$ in $\mathcal{P}$.

Proof. Follows directly from the definition of a (completed) PAG in combination with the validity of lemmas 1, 2, and 4, together with corollary 3 and lemma 14. □

It implies that *all* invariant arcs $X \longrightarrow Y$ in the PAG $\mathcal{P}$ established in this way are *definite* causal relations, i.e. identifiable, whether there is selection bias present or not. In general, however, invariant arcs $X \longrightarrow Y$ in a PAG may also be produced by selection on (a descendant of) X, without the actual existence of a directed path $X \Rightarrow Y$ in the underlying causal DAG $\mathcal{G}_C$.

Note that it does not imply that all definite causal relations can be identified this way. Indeed, there are examples of such relations that do not rely on rule (2). Then again, there can also exist identifiable definite causal relations that do take this form, but that do not start from an invariant tail in the PAG at all. Similarly, for some instances, the identifiable absence of a causal relation $X \not\Rightarrow Y$ does not follow directly from Theorem 1 for arbitrary single minimal independencies, as in lemma

11. Still, this never pertains to adjacent nodes, and can always be identified using Theorem 2 in [12]. More details on this, and how to read (definite) causal and other relations from a PAG, will be the subject of another article.

7 Conclusion

We have shown that the results underpinning causal discovery from multiple models in [12] can be extended to allow for the presence of selection bias. As such, the corresponding two rules are sufficient to directly identify *all* invariant arrowheads in a PAG. This in contrast to the seven graphical orientation rules [10, 5] that are normally used to accomplish this feat. The rules can be incorporated efficiently in a standard structure learning algorithm, like FCI [2]. As the rules do not depend on the detailed structure of the underlying graph, they offer the interesting possibility to work towards a more robust form of causal inference for finite data, by considering multiple combinations of nodes to decide on (the absence of) a specific causal relation.

The MCI-algorithm introduced in [12] was sound, but by no means complete. Arrowhead completeness for single models is a first step towards a completeness result for multiple models. Next step is to look at invariant tails in a PAG, see [5]. Current results suggest that two additional rules, similar to the ones in Theorem 1, are sufficient to cover all invariant tails in a single model as well. Furthermore, a few more rules can be applied to eliminate edges. Whether or not this approach will be able to cover all possible causal information that can be inferred from multiple models remains an open question.

Acknowledgement

This research was funded by NWO Vici grant 639.023.604.

References

[1] J. Pearl, *Causality: models, reasoning and inference*. Cambridge University Press, 2000.

[2] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*. Cambridge, Massachusetts: The MIT Press, 2nd ed., 2000.

[3] J. Zhang, “Causal reasoning with ancestral graphs,” *Journal of Machine Learning Research*, vol. 9, pp. 1437 – 1474, 2008.

[4] T. Richardson and P. Spirtes, “Ancestral graph Markov models,” *Ann. Stat.*, vol. 30, no. 4, pp. 962–1030, 2002.

[5] J. Zhang, “On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias,” *Artificial Intelligence*, vol. 172, no. 16-17, pp. 1873 – 1896, 2008.

[6] J. Zhang and P. Spirtes, “Detection of unfaithfulness and robust causal inference,” *Minds and Machines*, vol. 2, no. 18, pp. 239–271, 2008.

[7] P. Spirtes, C. Meek, and T. Richardson, “An algorithm for causal inference in the presence of latent variables and selection bias,” in *Computation, Causation, and Discovery*, pp. 211–252, 1999.

[8] S. Shimizu, P. Hoyer, A. Hyvärinen, and A. Kerminen, “A linear non-Gaussian acyclic model for causal discovery,” *Journal of Machine Learning Research*, vol. 7, pp. 2003–2030, 2006.

[9] P. Hoyer, D. Janzing, J. Mooij, J. Peters, and B. Schölkopf, “Nonlinear causal discovery with additive noise models,” in *Advances in Neural Information Processing Systems 21 (NIPS*2008)*, pp. 689–696, 2009.

[10] R. Ali, T. Richardson, P. Spirtes, and J. Zhang, “Towards characterizing markov equivalence classes for directed acyclic graphs with latent variables,” in *Proceedings of the 21th Conference on Uncertainty in Artificial Intelligence*, pp. 10–17, AUAI Press, 2005.

[11] J. Pellet and A. Elisseef, “Using Markov blankets for causal structure learning,” *Journal of Machine Learning Research*, vol. 9, pp. 1295–1342, 2008.

[12] T. Claassen and T. Heskes, “Causal discovery in multiple models from different experiments,” in *Advances in Neural Information Processing Systems 23* (J. Lafferty, C. K. I. Williams, R. Zemel, J. Shawe-Taylor, and A. Culotta, eds.), pp. 415–423, 2010.

[13] T. Claassen and T. Heskes, “Learning causal network structure from multiple (in)dependence models,” in *Proceedings of the Fifth European Workshop on Probabilistic Graphical Models*, 2010.