Supplement to - A Bayesian Approach to Constraint Based Causal Inference

Abstract

This article contains additional results and proofs related to §3.3 'Unfaithful inference: DAGs vs. MAGs' in the UAI-2012 submission 'A Bayesian Approach to Constraint Based Causal Inference'.

This final version after feedback includes detailed proofs for Lemma 3/4, and added inference rules for indirect dependence in §2. The discussion of causal inference from optimal uDAGs in section three has been extended, at the expense of the result for theoretical completeness in the large sample limit which will now appear in a separate paper on inference from uDAGs.

This supplement has three main parts: section one defines causal- and graphical model terminology used throughout the supplement. Section two describes readingin/dependencies from uDAGs; it extends results in (Bouckaert, 1995) through the added assumption of an underlying *faithful MAG*. The third section focuses on the mapping from optimal uDAGs to logical causal statements as used in the BCCD algorithm. The supplement follows the numbering in the main submission.

1 Notation and terminology

A causal model $\mathcal{G}_C$ is a directed acyclic graph (DAG) over a set of variables $\mathbf{V}$ where the arcs represent causal interactions. A directed path from A to B in such a graph indicates a *causal relation* $A \Rightarrow B$ in the system, where cause A *influences* the value of its effect B, but not the other way around. An edge $A \rightarrow B$ in $\mathcal{G}_C$ indicates a *direct* causal link such that A influences B, but not the other way around. A causal relation $A \Rightarrow B$ implies a probabilistic dependence $A \not\perp\!\!\!\perp B$.

The joint probability distribution induced by a causal DAG $\mathcal{G}_C$ factors according to a *Bayesian network* (BN): a pair $\mathcal{B} = (\mathcal{G}, \boldsymbol{\Theta})$, where $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ is DAG over random variables $\mathbf{V}$, and the parameters $\boldsymbol{\theta}_V \subset \boldsymbol{\Theta}$ represent the conditional probability of variable $V \in \mathbf{V}$ given its parents $\mathbf{Pa}(V)$ in the graph $\mathcal{G}$. Probabilistic independencies can be read from the graph $\mathcal{G}$ via the *d-separation* criterion: X is conditionally independent of Y given $\mathbf{Z}$, denoted $X \perp\!\!\!\perp Y \,|\, \mathbf{Z}$, iff there is no unblocked path between X and Y in $\mathcal{G}$ conditional on the nodes in $\mathbf{Z}$, see (Pearl, 1988; Neapolitan, 2004). A *minimal* independence $X \perp\!\!\!\perp Y \,|\, [\mathbf{Z}]$ implies that removing any node from the set surrounded by square brackets turns it into a dependence, and vice versa.

Independence relations between arbitrary subsets of variables from a causal DAG can be represented in the form of a *(maximal) ancestral graph* (MAG) $\mathcal{M}$, an extension of the class of DAGs that is closed under marginalization and selection. In addition to directed arcs, MAGs can contain *bi-directed arcs* $X \leftrightarrow Y$ (indicative of marginalization) and undirected edges $X \relbar Y$ (indicative of selection), see (Richardson and Spirtes, 2002).

The *equivalence class* $[\mathcal{G}]$ of a graph $\mathcal{G}$ is the set of all graphs that are indistinguishable in terms of (Markov) implied independencies. For a DAG or MAG $\mathcal{G}$, the corresponding equivalence class $[\mathcal{G}]$ can be represented as a *partial ancestral graph* (PAG) $\mathcal{P}$, which keeps the skeleton (adjacencies) and all invariant edge marks, i.e. tails ($-$) and arrowheads ($>$) that appear in all members of the equivalence class, and turns the remaining non-invariant edge marks into circles ($\circ$) (Zhang, 2008). A *potentially directed path* (p.d.p.) is a path in a PAG that could be oriented into a directed path by changing circle marks into appropriate tails/arrowheads. For an edge $A \ast\!\!\rightarrow B$ in $\mathcal{P}$, the invariant arrowhead at B signifies that B is *not* a cause of A. An edge $A \rightarrow B$ implies a direct, causal link $A \Rightarrow B$.

A *logical causal statement* L is statement about pres-

ence or absence of causal relations between two or three variables of the form $(X \Rightarrow Y)$, $(X \Rightarrow Y) \lor (X \Rightarrow Z)$, or $(X \not\Rightarrow Y) \equiv \neg(X \Rightarrow Y)$.

A DAG $\mathcal{G}$ is an (unfaithful) **uDAG** approximation to a MAG $\mathcal{M}$ over a set of nodes $\mathbf{X}$, iff for any probability distribution $p(\mathbf{X})$, generated by an underlying causal graph faithful to $\mathcal{M}$, there is a set of parameters $\mathbf{\Theta}$ such that the Bayesian network $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$ encodes the same distribution $p(\mathbf{X})$. The uDAG is **optimal** if there exists no uDAG to $\mathcal{M}$ with fewer free parameters.

We use $\mathcal{D}$ to indicate a data set over variables $\mathbf{V}$ from a distribution that is faithful to some (larger) underlying causal DAG $\mathcal{G}_C$. $\mathbb{L}$ denotes the set of possible causal statements L over variables in $\mathbf{V}$. We use $\mathbf{M_X}$ for the set of MAGs over $\mathbf{X}$, and $\mathbf{M_X}(L)$ to denote the subset that entails logical statement L. We also use $\mathcal{G}$ to explicitly indicate a DAG, $\mathcal{M}$ for a MAG, and $\mathcal{P}$ for a PAG.

2 Inference from uDAGs

2.1 Reading in/dependencies from uDAGs

A uDAG is a DAG for which we do not know if it is faithful or not. Reading in/dependence relations from a uDAG goes as follows:

Lemma 3. Let $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$ be a Bayesian network over a set of nodes $\mathbf{X}$, with $\mathcal{G}$ a uDAG for a MAG $\mathcal{M}$ that is faithful to a distribution $p(\mathbf{X})$. Let $\mathcal{G}_{X \| Y}$ be the graph obtained by eliminating the edge $X - Y$ from $\mathcal{G}$ (if present), then:

$$(X \perp\!\!\!\perp_{\mathcal{G}} Y \mid \mathbf{Z}) \Rightarrow (X \perp\!\!\!\perp_p Y \mid \mathbf{Z}),$$

$$(X \perp\!\!\!\perp_{\mathcal{G}_{X \| Y}} Y \mid \mathbf{Z}) \land (X \not\perp\!\!\!\perp_{\mathcal{G}} Y \mid \mathbf{Z}) \Rightarrow (X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}).$$

(alternatively, in a more compact formulation:)

If $\mathbf{Z}$ is a set that d-separates X and Y in $\mathcal{G}_{X \| Y}$, then

$$(X \perp\!\!\!\perp_{\mathcal{G}} Y \mid \mathbf{Z}) \Leftrightarrow (X \perp\!\!\!\perp_p Y \mid \mathbf{Z}),$$

Proof sketch. The independence rule ($\Rightarrow$) follows (Pearl, 1988). The dependence rule (through negation) is similar to the 'coupling' theorem (3.11) in (Bouckaert, 1995), but stronger. As we assume a faithful MAG, a dependence $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$ cannot be destroyed by in/excluding a node U that has no unblocked path in the underlying MAG to X and/or Y given $\mathbf{Z}$. This eliminates one of the preconditions in the coupling theorem (see below). □

In words: "in a uDAG $\mathcal{G}$ we can apply standard d-separation to infer in/dependence '$X \perp\!\!\!\perp_p Y \mid \mathbf{Z}$?', provided the set $\mathbf{Z}$ d-separates nodes X and Y when edge

$X - Y$ (if present) is removed". So, all independencies from d-separation remain valid, but identifiable dependencies put restrictions on the set $\mathbf{Z}$.

For a more formal proof, we first introduce the notion of coupling : in a DAG $\mathcal{G}$, two variables X and Y are *coupled* given $\mathbf{Z}$, denoted $\rangle X, \mathbf{Z}, Y \langle_{\mathcal{G}}$, if: $(X \to Y) \in \mathcal{G}$, $Pa(Y)_{\mathcal{G}} \subseteq (X \cup \mathbf{Z})$, and $X \perp\!\!\!\perp_{\mathcal{G}} Y \mid \mathbf{Z}$ in $\mathcal{G}_{X \| Y}$ (or vice versa for X).

The relevance of this notion comes courtesy of the following result, based on the graphoid axioms for independence:

Theorem 3.11 (Bouckaert, 1995) In a DAG $\mathcal{G}$, if two variables X and Y are *coupled* given $\mathbf{Z}$ in $\mathcal{G}$ then $X \not\perp\!\!\!\perp Y \mid \mathbf{Z}$.

The difference is that Lemma 3 does not require the explicit inclusion of $Pa(Y)_{\mathcal{G}} \subseteq (X \cup \mathbf{Z})$ in the separating set. This is a direct consequence of the (stronger) assumption of an underlying faithful MAG $\mathcal{M}$, as we now show in the detailed proof of Lemma 3:

Proof. Let $X \to Y$ in $\mathcal{G}$, and let X and Y be coupled given disjoint sets $\mathbf{Z} \cup \mathbf{W}$, with $\mathbf{W} \subset Pa(Y)$, so that $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z} \cup \mathbf{W}$ by Theorem 3.11. Now assume that also $X \perp\!\!\!\perp_{\mathcal{G}_{\|}} Y \mid \mathbf{Z}$, then there is no unblocked path given $\mathbf{Z}$ from X to any $W \in \mathbf{W}$ in $\mathcal{G}$, otherwise the $W \longrightarrow Y$ would imply an unblocked path $\langle X, .., W, Y \rangle$ given $\mathbf{Z}$, in contradiction with $X \perp\!\!\!\perp_{\mathcal{G}_{\|}} Y \mid \mathbf{Z}$. As a result, $\forall W \in \mathbf{W} : X \perp\!\!\!\perp_{\mathcal{G}} W \mid \mathbf{Z}$ in $\mathcal{G}$, so also (Pearl, 1988) $X \perp\!\!\!\perp_p W \mid \mathbf{Z}$, and so by definition also $X \perp\!\!\!\perp_{\mathcal{M}} W \mid \mathbf{Z}$ in the underlying faithful MAG $\mathcal{M}$.

We now show by contradiction that this implies that X and Y are dependent given $\mathbf{Z}$, i.e. $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$:

Suppose that $X \perp\!\!\!\perp_p Y \mid \mathbf{Z}$ while given $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z} \cup \mathbf{W}$ and $X \perp\!\!\!\perp_p \mathbf{W} \mid \mathbf{Z}$. For a faithful MAG $\mathcal{M}$ this implies that X and Y are m-separated given only $\mathbf{Z}$, but *not* given $(\mathbf{Z} \cup \mathbf{W})$. An unblocked path π given a set $(\mathbf{Z} \cup \mathbf{W})$ means that all noncolliders on π are not in $\mathbf{Z} \cup \mathbf{W}$ and all colliders on π are in $An(\mathbf{Z} \cup \mathbf{W} \cup \mathbf{S})$. A path π in $\mathcal{M}$ blocked by the removal of $\mathbf{W}$ therefore contains one or more colliders in $An(\mathbf{W})$, as any collider in $\mathbf{Z}$ that blocks the path π would also block the path given $(\mathbf{Z} \cup \mathbf{W})$ before. Let W be the first collider in $An(\mathbf{W})$ encountered along the unblocked path π given $(\mathbf{Z} \cup \mathbf{W})$, then faithfulness implies $X \not\perp\!\!\!\perp_p W \mid \mathbf{Z}$, contrary the given. Therefore the assumption $X \perp\!\!\!\perp Y \mid \mathbf{Z}$ cannot hold, and so we can infer $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$. □

Note that this does *not* follow from the (in)dependence axioms - otherwise it would hold always.

As a consequence unblocked paths are always preserved.

Corollary 3a. Let $\mathcal{G}$ be a uDAG for a faithful MAG $\mathcal{M}$, then all unblocked paths in $\mathcal{M}$ are preserved in $\mathcal{G}$, i.e. then $X \not\perp\!\!\!\perp_{\mathcal{M}} Y \,|\, \mathbf{Z}$ implies $X \not\perp\!\!\!\perp_{\mathcal{G}} Y \,|\, \mathbf{Z}$.

Proof. Follows immediately from Lemma 3. □

We can extend Lemma 3 to a two-step, *indirect* dependence, similar to Example 1 in the main article:

Lemma 3b. Let $\mathcal{G}$ be a uDAG for a faithful MAG $\mathcal{M}$. Let X, Y, $\mathbf{Z}$, W be disjoint (sets of) nodes with $\langle X, W, Y \rangle \in \mathcal{G}$, for which $X \perp\!\!\!\perp_{\mathcal{G}\|} W \,|\, \mathbf{Z}$ and $W \perp\!\!\!\perp_{\mathcal{G}\|} Y \,|\, \mathbf{Z}$. If $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, \mathbf{Z} \cup W$, then $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

Proof. By Lemma 3, $X \not\perp\!\!\!\perp_p W \,|\, \mathbf{Z}$ and $W \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$, and so there are unblocked paths from X and Y to W given $\mathbf{Z}$ in $\mathcal{M}$. These paths in $\mathcal{M}$ cannot be both *into* W, because that would imply that X and Y are m-connected given $\mathbf{Z} \cup W$, so $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z} \cup W$, contrary the given $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, \mathbf{Z} \cup W$. But that means that W is a noncollider in $\mathcal{M}$ on an unblocked path given $\mathbf{Z}$ from X to Y, and so $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$ without W. □

It is easy to see that, similar to the direct edge in Lemma 3, the path $\langle X, W, Y \rangle$ is also the *only* unblocked path between X and Y given $\mathbf{Z}$. This turns out to hold generally for identifiable dependence in a uDAG $\mathcal{G}$ to a faithful MAG $\mathcal{M}$: if there is *one and only one* unblocked path in $\mathcal{G}$ from X to Y given $\mathbf{Z}$, then $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$. Though perhaps intuitively obvious ('one unblocked path implies there is no other path/mechanism that can cancel out the dependence due to this one'), to prove it we need to relate observations in $\mathcal{G}$ to configurations in $\mathcal{M}$, and deduce the dependency holds from there.

We first show this for an extended version of Lemma 3b, in which there is now a single, unblocked directed path from X to Y given $\mathbf{Z} \subset An(Y)$ (so no unblocked paths 'created' by $\mathbf{Z}$). After that we do the general version.

Lemma 3c. Let $\mathcal{G}$ be a uDAG for a faithful MAG $\mathcal{M}$. Let X, Y, $\mathbf{Z}$, $\mathbf{W} = \{W_1, .., W_k\}$ be disjoint (sets of) nodes. If $\pi = \langle X \to W_1 \to ..W_k \to Y \rangle$ is the *only* unblocked path from X to Y given $\mathbf{Z}$ in $\mathcal{G}$, then $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

Proof. By induction. We show that at each step, if there was an unblocked path in $\mathcal{M}$ between X and W_i given $\mathbf{Z}$, then there is also one between X and W_{i+1}. We do this by deriving a contradiction from the assumption that a necessary collider in $\mathcal{M}$ is not in the set $\mathbf{Z}$.

Let $\mathbf{V} = Pa(\mathbf{W})_{\mathcal{G}} \setminus (\mathbf{W} \cup \mathbf{Z})$ represent the set all parents in $\mathcal{G}$ from nodes in $\mathbf{W}$ that are not on π or in $\mathbf{Z}$. Let $\mathbf{V}_i \subseteq \mathbf{V}$ denote a minimal subset of nodes from $\mathbf{V}$ needed to block all alternative paths between $W_i \to W_{i+1}$ in $\mathcal{G}$, such that $W_i \perp\!\!\!\perp_{\mathcal{G}\|} W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i$. By Lemma 3 this implies:

$$W_i \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i \tag{1}$$

The invariant after induction step i is:

$$X \not\perp\!\!\!\perp_p W_i \,|\, \mathbf{Z} \tag{2}$$

By definition W_i separates all its predecessors on the unblocked path π in $\mathcal{G}$ from all its successors given $\mathbf{Z} \cup \mathbf{V}$, so that in particular:

$$X \perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i \cup W_i \tag{3}$$

Below (in B) we show by contradiction that this implies that:

$$W_i \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \tag{4}$$

For the proof in (B) we first show in (A) that:

$$\forall \mathbf{V}'_i \subseteq (\mathbf{V}_i \setminus V) : X \perp\!\!\!\perp_p V \,|\, \mathbf{Z} \cup \mathbf{V}'_i \tag{5}$$

From this it follows that the unblocked path in $\mathcal{M}$ given $\mathbf{Z}$ corresponding to eq.(2) cannot be blocked by any node from $\mathbf{V}_i$, and so also

$$X \not\perp\!\!\!\perp_p W_i \,|\, \mathbf{Z} \cup \mathbf{V}_i \tag{6}$$

Eqs. (6) and (1) correspond to unblocked paths in $\mathcal{M}$ from to W_i to X and W_{i+1} given $\mathbf{Z} \cup \mathbf{V}_i$. Node W_i cannot be a collider between these two paths in $\mathcal{M}$, otherwise $X \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i \cup W_i$, contrary to (3), and so it follows that:

$$X \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i \tag{7}$$

But then if $X \perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup (\mathbf{V}_i \setminus V)$, then V must be a collider with unblocked paths in $\mathcal{M}$ to X and W_{i+1} given $\mathbf{Z} \cup (\mathbf{V}_i \setminus V)$. But that implies that $X \not\perp\!\!\!\perp_p V \,|\, \mathbf{Z} \cup (\mathbf{V}'_i \setminus V)$, in contradiction with eq.(5), and so $X \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup (\mathbf{V}_i \setminus V)$. This argument can be repeated to eliminate all nodes $\mathbf{V}_i$, so that we find:

$$X \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \tag{8}$$

which corresponds to the invariant for the next step. By induction over the entire path this ultimately proves that $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

The remaining elements in this process are detailed below:

The invariant in eq.(2) holds at the induction base for edge $X \to W_1$. In $\mathcal{G}$ we find that $\mathbf{Z}$ blocks all other

paths between X and W_1, otherwise an alternative unblocked path from X to Y given $\mathbf{Z}$ would exist, contrary the given. Therefore $X \perp\!\!\!\perp_{\mathcal{G}\parallel} W_1 \,|\, \mathbf{Z}$, from which we conclude $X \not\perp\!\!\!\perp_p W_1 \,|\, \mathbf{Z}$ by Lemma 3, and so eq.(2) is satisfied.

Part (A): X is independent of any node in $\mathbf{V}_i$ given $\mathbf{Z}$, eq.(5).
Let $V_j \in \mathbf{V}_i \cap Pa(W_j)$ be a node that is a parent from W_j (with $j \in \{0,..,i\}$) needed to block some secondary path (apart from π) in $\mathcal{G}$ between W_i and W_{i+1}. This path consists of one half as a directed path $\pi_1 = V_j \to W_j.. \to W_i$, and the second half as an unblocked path $\pi_2 = V_j.. \to W_{i+1}$ given $\mathbf{Z}\cup(\mathbf{V}_i\backslash V_j)$ in $\mathcal{G}$, connected by noncollider V_j. The path π_2 may contain zero, one or more collider nodes (that are ancestors of/) from $(\mathbf{V}_i \backslash V_j)$. Denote these as $V_{k(1)},..,V_{k(m)}$ respectively, as encountered along π_2 when going from V_j to W_{i+1}. By contradiction. Suppose that $X \not\perp\!\!\!\perp_p V_j \,|\, \mathbf{Z}$, then there is an unblocked path $X \not\perp\!\!\!\perp_{\mathcal{M}} V_j \,|\, \mathbf{Z}$ in $\mathcal{M}$, and so (Corollary 3a) there is an unblocked path π' in $\mathcal{G}$. If this unblocked path π' in $\mathcal{G}$ does *not* go via W_j (the child of V_j, on π), then the path $\pi'+\pi_1+W_i \to ..Y$ is an alternative unblocked path between X and Y via V_j given $\mathbf{Z}$, and so is not allowed. If it *does* go via W_j and there are no collider nodes along π_2 (i.e. $m = 0$ in the sequence $V_{k(1)},..,V_{k(m)}$), then the path $\pi' + \pi_2 + W_{i+1} \to ..Y$ is an alternative unblocked path between X and Y via V_j given $\mathbf{Z}$, and so is also not allowed. Alternatively, if there are one or more collider nodes along π_2 (i.e. with $m \geq 1$ in the sequence $V_{k(1)},..,V_{k(m)}$), then the fact that π' can only go via W_j if this node is (ancestor of) a collider in $\mathcal{G}$ that is (ancestor of) a node in $\mathbf{Z}$, implies that there is at least one leg $\{W_{k(l)}, W_{k(l+1)}\}$ in the corresponding sequence $W_{k(0)}(= W_j), W_{k(1)},..,W_{k(m)}, W_{k(m+1)}(= W_{i+1})$ for which $k(l) \leq j$ and $W_{k(l)} \prec W_{k(l+1)}$ along π. As node $W_{k(l)}$ is an ancestor of (or equal to) node W_j, which in turn was (ancestor of) a collider in $\mathcal{G}$ that is (ancestor of) a node in $\mathbf{Z}$, it means that the path $X.. \to W_{k(l)} \leftarrow .. \to W_{k(l+1)}.. \to Y$ is an alternative unblocked path between X and Y given $\mathbf{Z}$, and so again not allowed.

In short: assuming $X \not\perp\!\!\!\perp_p V_j \,|\, \mathbf{Z}$ results in a contradiction with the original assumption of a single unblocked path π connecting X and Y in $\mathcal{G}$, and therefore $X \perp\!\!\!\perp_p V_j \,|\, \mathbf{Z}$ must apply for any node from $\mathbf{V}_i$ needed to create the $W_i \perp\!\!\!\perp_{\mathcal{G}\parallel} W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i$. But if $X \perp\!\!\!\perp_p V \,|\, \mathbf{Z}$ holds for all $V \in \mathbf{V}_i$, then (by faithfulness) for an arbitrary node $V' \in (\mathbf{V}_i \backslash V)$ it must hold that $X \perp\!\!\!\perp_p V \,|\, \mathbf{Z} \cup V'$, otherwise, to create the dependence through conditioning on V', there have to be unblocked paths from X (and V) to V' given $\mathbf{Z}$ in $\mathcal{M}$, contradicting $X \perp\!\!\!\perp_p V' \,|\, \mathbf{Z}$. This argument can be extended to arbitrary subsets of $\mathbf{V}_i$, ergo:

$\forall \mathbf{V}' \subseteq (\mathbf{V}_i \backslash V) : X \perp\!\!\!\perp_p V \,|\, \mathbf{Z} \cup \mathbf{V}'$. (end-of-proof part A)

Part (B): two successive nodes along π are dependent given $\mathbf{Z}$, eq.(4).
By contradiction. Assume the invariant $X \not\perp\!\!\!\perp_p W_i \,|\, \mathbf{Z}$ holds up to node W_i along the path, and suppose that $W_i \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i$, but $W_i \perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}$. Then there is at least one $V \in \mathbf{V}_i$ needed to block all paths between W_i and W_{i+1} in $\mathcal{G}$ that is a collider between unblocked paths from these two nodes given $\mathbf{Z} \cup (\mathbf{V}_i \backslash V)$ in $\mathcal{M}$, necessary to create the dependence. By (A) we know that no subset of nodes in $\mathbf{V}_i$ can block the unblocked path in $\mathcal{M}$ between X and W_i given $\mathbf{Z}$, and so we also have (in particular): $X \not\perp\!\!\!\perp_p W_i \,|\, \mathbf{Z} \cup (\mathbf{V}_i \backslash V)$, corresponding to an unblocked path in $\mathcal{M}$ between X and W_i given $\mathbf{Z} \cup (\mathbf{V}_i \backslash V)$. Node W_i cannot be a collider between these paths in $\mathcal{M}$, because that would imply $X \not\perp\!\!\!\perp_{\mathcal{M}} W_{i+1} \,|\, \mathbf{Z} \cup \mathbf{V}_i \cup W_i$, contrary (3). But if W_i is a noncollider between these paths in $\mathcal{M}$, then without W_i the path from X to V is unblocked given $\mathbf{Z} \cup (\mathbf{V}_i \backslash V)$, contrary to (A).

As W_i has to be either a collider or a noncollider between these unblocked paths in $\mathcal{M}$, it follows that the assumption of a node $V \in \mathbf{V}_i$ needed in the conditioning set to create the dependence between W_i and W_{i+1} is false. Hence they must also be dependent without conditioning on $\mathbf{V}_i$, or in other words: $W_i \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}$. (end-of-proof part B) $\square$

We can now formulate the general version to infer dependence from arbitrary, single unblocked paths.

Lemma 4. Let $\mathcal{G}$ be a uDAG for a faithful MAG $\mathcal{M}$. Let X, Y, and $\mathbf{Z}$ be disjoint (sets of) nodes. If $\pi = \langle X,..,Y \rangle$ is the *only* unblocked path from X to Y given $\mathbf{Z}$ in $\mathcal{G}$, then $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

Proof. The path π can be split into three parts: $\pi = \pi_1 + \pi_2 + \pi_3$, with $\pi_1 = X \leftarrow .. \leftarrow U$, the part of π that is a directed path into X, $\pi_2 = U \to .. \to C_1 \leftarrow .. \to C_k \leftarrow .. \leftarrow V$, the part with directed paths into colliders C_i along π, and $\pi_3 = V \to .. \to Y$, a directed path into Y. Note that any π can be written as a combination of one, two or all three subpaths from $\{\pi_1, \pi_2, \pi_3\}$, possibly with X and/or Y taking the role of U and/or V. For example, the case in Lemma 3c corresponds to $\pi = \pi_3$ with $X = V$.

For the proof, we first show in part (A) that each of the three subpaths π_1, π_2, π_3 represents a dependency $U \not\perp\!\!\!\perp_p X \,|\, \mathbf{Z}$, $U \not\perp\!\!\!\perp_p V \,|\, \mathbf{Z}$, and $V \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$, corresponding to unblocked paths in $\mathcal{M}$ given $\mathbf{Z}$. Then we show in part (B) that these can be stitched together in any combination to obtain $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

Part (A):

Subpaths π_1 and π_3 satisfy the antecedent of Lemma 3c, and so represent identifiable dependencies $U \not\perp\!\!\!\perp_p X \,|\, \mathbf{Z}$, and $V \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

For each pair of nodes W_i, W_j on the path π_2 it holds that $\mathbf{Z}$ blocks all alternative paths π'_{ij} between them in $\mathcal{G}$ (except along π_2), otherwise the path $\pi' = \langle X, .., W_i \rangle + \pi'_{ij} + \langle W_j, .., Y \rangle$ is an alternative unblocked path in $\mathcal{G}$ between X and Y given $\mathbf{Z}$: whether $W_{i/j} \in An(\mathbf{Z})$ is a collider or noncollider along π', it does not block the path. As we assumed that π was the *only* unblocked path between X and Y, it follows there is no unblocked path π'_{ij} in $\mathcal{G}$ given $\mathbf{Z}$.
This implies in particular that for each successive pair of nodes W_i, W_{i+1} along π_2 it holds that $W_i \perp\!\!\!\perp_{\mathcal{G}\|} W_{i+1} \,|\, \mathbf{Z}$, and so (by Lemma 3) that $W_i \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}$.

Furthermore, each node W_i that is not a collider along π in $\mathcal{G}$ is also not a collider between its neighbouring legs $W_{i-1} - W_i - W_{i+1}$ along the corresponding unblocked path in $\mathcal{M}$, otherwise conditioning on $\mathbf{Z} \cup W_i$ would unblock a path in $\mathcal{M}$, whereas in $\mathcal{G}$ it implies $W_{i-1} \perp\!\!\!\perp_{\mathcal{G}} W_{i+1} \,|\, \mathbf{Z} \cup W_i$, and so there is an unblocked path in $\mathcal{M}$ without W_i, corresponding to $W_{i-1} \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}$.

But if W_i *is* a collider along π in $\mathcal{G}$ then it is also a collider between its neighbouring legs $W_{i-1} - W_i - W_{i+1}$ along the corresponding unblocked path in $\mathcal{M}$. The single unblocked path implies that there is a subset $\mathbf{Z}' \subset (\mathbf{Z} \setminus W_i)$ (in case collider W_i in $\mathcal{G}$ is itself part of $\mathbf{Z}$) such that both $W_{i-1} \perp\!\!\!\perp_{\mathcal{G}\|} W_i \,|\, \mathbf{Z}'$ and $W_i \perp\!\!\!\perp_{\mathcal{G}\|} W_{i+1} \,|\, \mathbf{Z}'$. This subset also separates W_{i-1} and W_{i+1} in $\mathcal{G}$ (otherwise it would not block all alternative paths to W_i) so that $W_{i-1} \perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}'$. That implies that W_i is a collider in $\mathcal{M}$ between unblocked paths from W_{i-1} and W_{i+1} given $\mathbf{Z}'$, i.e. $W_{i-1} \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}' \cup W_i$. We can expand $\mathbf{Z}'$ to include all nodes in $\mathbf{Z}$ that are not descendant of W_i in $\mathcal{G}$. The remaining subset $\mathbf{Z}^* = \mathbf{Z} \setminus \mathbf{Z}'$ contains only descendants of W_i in $\mathcal{G}$ and can only destroy this dependence if it blocks at least one leg, say $W_{i-1} - W_i$, of this unblocked path in $\mathcal{M}$ given $\mathbf{Z}'$, so that $W_{i-1} \not\perp\!\!\!\perp_p W_i \,|\, \mathbf{Z}$. This also implies an unblocked path in $\mathcal{G}$ from W_{i-1} to a node $Z^* \in \mathbf{Z}^*$ given $\mathbf{Z} \setminus Z^*$ that does *not* go via W_i. Node W_{i+1} cannot have a similar alternative path to Z^* in $\mathcal{G}$ that does not go via W_i, because that would imply an alternative unblocked path between X and Y, bypassing W_i. Therefore, similar to the situation in Lemma 3c, W_{i+1} and Z^* can be separated (in $\mathcal{G}$) by some set including W_i, whereas in $\mathcal{M}$ they are dependent given W_i (blocking the path $W_{i-1} - \mathbf{Z}^* .. \to W_i \leftarrow ..W_{i+1}$). The contradiction implies that the assumption the nodes in $\mathbf{Z}^*$ can block the dependence via W_i given $\mathbf{Z}'$ is false, and hence that again $W_{i-1} \not\perp\!\!\!\perp_p W_{i+1} \,|\, \mathbf{Z}$.

As this applies to each overlapping triple we can extend the dependence (similar to Lemma 3c) along the entire path π_2 to obtain $U \not\perp\!\!\!\perp_p V \,|\, \mathbf{Z}$.

Part (B):

If π consists of just a single subpath π_i, then the dependence is already shown above. For combinations we can connect the subpaths on root nodes U and V along π in $\mathcal{G}$ in the same fashion: Node U cannot be a collider between π_1 and π_2 in $\mathcal{M}$, because in $\mathcal{G}$ conditioning on U cannot unblock any new paths (as U was already in $An(\mathbf{Z})$, and so $X \perp\!\!\!\perp_{\mathcal{G}} V \,|\, \mathbf{Z} \cup U$, and so without U there is an unblocked path in $\mathcal{M}$ corresponding to $X \not\perp\!\!\!\perp_p V \,|\, \mathbf{Z}$. Similarly V cannot be a collider between π_2 and π_3, and therefore also for a single unblocked path $\pi = \pi_1 + \pi_2 + \pi_3$ in $\mathcal{G}$ it holds that $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$.

For empty π_2, we also find that $U(=V)$ cannot be a collider between π_1 and π_3 in $\mathcal{M}$, as conditioning on U blocks the last unblocked path in $\mathcal{G}$ between X and Y. Otherwise, any path in $\mathcal{G}$ unblocked by adding U to the conditioning set $\mathbf{Z}$ goes via a collider $C \in An(U)$. But C already has an unblocked path to $X(/Y)$ given $\mathbf{Z}$, which means that the path $\pi' = X \leftarrow .. \to C \to .. \to U \to .. \to Y$ (or vice versa for Y) is an alternative unblocked path in $\mathcal{G}$ given $\mathbf{Z}$. This is contrary the orignal assumption, and therefore conditioning on U blocks the path π but cannot open up any new path in $\mathcal{G}$, and therefore $X \perp\!\!\!\perp_p Y \,|\, \mathbf{Z} \cup U$. This implies U must be a noncollider connecting π_1 and π_3 in $\mathcal{M}$, and therefore again $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$. $\qquad\square$

A powerful way to obtain more dependence statements is to eliminate nodes from the conditioning set $\mathbf{Z}$ that can be shown not to be needed to ensure the dependence.

Lemma 4a Let $\mathcal{G}$ be a uDAG for a faithful MAG $\mathcal{M}$. Let X, Y, and $\mathbf{Z}$ be disjoint (sets of) nodes such that $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$. Let $\mathbf{Z}' \subseteq \mathbf{Z}$ be a subset such that for each $Z \in \mathbf{Z}'$ there are no (disjoint) unblocked paths $\pi_X = \langle X, .., Z \rangle$ and $\pi_Y = \langle Z, .., Y \rangle$ between X and Y in $\mathcal{G}$ given $\mathbf{Z} \setminus Z$, then $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z} \setminus \mathbf{Z}'$.

Proof. The given $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$ establishes the existence of an unblocked path π in $\mathcal{M}$ given $\mathbf{Z}$. All nodes $Z \in \mathbf{Z}^* \subseteq \mathbf{Z}$ that are (descendants of) colliders along this unblocked path π have unblocked paths to both X and Y given $\mathbf{Z} \setminus Z$ (or given the union of $(\mathbf{Z}^* \setminus Z)$ and any subset $(\mathbf{Z} \setminus Z)$), and are therefore (by Corollary 3a) not in $\mathbf{Z}'$. So, removing any (subset of) node(s) $\mathbf{Z}'$ from $\mathbf{Z}$ cannot introduce a noncollider on π, nor remove a necessary collider from π. Hence the unblocked path π remains unblocked in $\mathcal{M}$ given $\mathbf{Z} \setminus \mathbf{Z}'$, and so (by faithfulness) the dependence $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z} \setminus \mathbf{Z}'$ also holds. $\qquad\square$

This approach can be extended to read even more dependencies. For example, the single unblocked path requirement in Lemma 4 can be relaxed, ultimately leading to a graphical criterion to read dependencies from uDAGs. However, a full analysis of inference from uDAGs would go far beyond the scope of the current article. Instead we focus on the mapping to the logical causal statements in the BCCD algorithm.

3 Causal statements from uDAGs

3.1 Minimal in/dependencies

From (Claassen and Heskes, 2011b) we know that causal information can be found by identifying variables Z that either *make* or *break* an independence relation between $\{X, Y\}$:

1. $X \perp\!\!\!\perp_p Y \,|\, [\mathbf{W} \cup Z] \;\;\vdash\;\; (Z \Rightarrow X) \vee (Z \Rightarrow Y)$,

2. $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{W} \cup [Z] \;\;\vdash\;\; Z \not\Rightarrow (\{X,Y\} \cup \mathbf{W})$.

In words: a *minimal* conditional independence identifies the presence of at least one from two causal relations, whereas a dependence identifies the absence of causal relations.

To infer a minimal independence $X \perp\!\!\!\perp_p Y \,|\, [\mathbf{Z}]$ from a uDAG we need to establish that in a given independence $X \perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$ all nodes $Z \in \mathbf{Z}$ are noncollider on some unblocked path between X and Y given the other nodes $\mathbf{Z}_{\backslash Z}$.

Lemma 4b. Let $\mathcal{G}$ be a uDAG to a faithful MAG $\mathcal{M}$. Then $X \perp\!\!\!\perp_p Y \,|\, [\mathbf{Z}]$ can be read from $\mathcal{G}$, iff we can infer $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, [\mathbf{Z}]$, and $\forall Z \in \mathbf{Z} : X, Y \not\perp\!\!\!\perp_p Z \,|\, \mathbf{Z}_{\backslash Z}$.

Proof. In words: it suffices to establish that given a separating set $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, \mathbf{Z}$ in $\mathcal{G}$, each node Z in $\mathbf{Z}$ is dependent on both X and Y given the others.
Clearly, if the independence is not minimal in $\mathcal{G}$ then we cannot infer it is minimal in $\mathcal{M}$ (otherwise $\mathcal{M} = \mathcal{G}$ is a trivial counter), so we can start from $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, [\mathbf{Z}]$. If there is a node Z for which it is *not* possible to establish a dependence to X and Y given the rest, then there exists a corresponding MAG in which Z is independent from X/Y given $\mathbf{Z}_{\backslash Z}$, and so by Lemma 4a not always needed in the minimal independence.
If it *does* hold, then each node $Z \in \mathbf{Z}$ is noncollider on some unblocked path between X and Y given all the others: for each node there are unblocked paths π_{XZ} and π_{ZY} from X and Y to Z given $\mathbf{Z}_{\backslash Z}$, connected by noncollider Z (otherwise not $X \perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$), which makes $\pi_{XY} = \pi_{XZ} + \pi_{ZY}$ the required unblocked path between X and Y. Therefore $\mathbf{Z}$ is needed to block all paths in $\mathcal{M}$, and so it also represents a minimal independence in $\mathcal{M}$, i.e. $X \perp\!\!\!\perp_p Y \,|\, [\mathbf{Z}]$. $\qquad\square$

Note that to establish in Lemma 4b that a node Z has unblocked paths to both X and Y given the others we can either show that $X \not\perp\!\!\!\perp_p Z \,|\, \mathbf{Z}_{\backslash Z}$ and $Z \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}_{\backslash Z}$ hold, or directly show that $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}_{\backslash Z}$ can be inferred from uDAG $\mathcal{G}$.

Identifying a node that breaks an independence from a uDAG follows straightforward from the definition:

Corollary 4c. Let $\mathcal{G}$ be a uDAG to a faithful MAG $\mathcal{M}$. Then $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{W} \cup [Z]$ can be read from $\mathcal{G}$, iff $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, \mathbf{W}$, and both $X \not\perp\!\!\!\perp_p Z \,|\, \mathbf{W}$ and $Z \not\perp\!\!\!\perp_p Y \,|\, \mathbf{W}$ can be inferred.

Proof. If Z has unblocked paths to X and Y given $\mathbf{W}$, then Z is a collider between these paths in $\mathcal{M}$ (otherwise not $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, \mathbf{W}$), and so including Z makes them dependent, i.e. $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{W} \cup [Z]$. $\qquad\square$

It means that for inferring (both types of) causal information from uDAGs, reading dependencies remains the crucial bottleneck. From Lemma 4 we know that the existence of a single unblocked path is sufficient to infer a dependence, but that would miss out on many others. One way to increase the number of readable dependencies is to identify patterns of nodes that can invalidate a given unblocked path π in uDAG $\mathcal{G}$: if we find these patterns are not present for said path, then we can also infer the dependence. For that we introduce the following notion:

Definition. In a uDAG $\mathcal{G}$, a node Z lies on an *(indirect) triangle detour* for an edge $X \rightarrow Y$, iff Z is a non-collider on a triangle with X and Y in $\mathcal{G}$, or $X \rightarrow (Z' \rightarrow Y) \leftarrow Z$ in $\mathcal{G}$. A node Z lies on an *(indirect) collider detour* for $X \rightarrow Y$, iff $X \rightarrow Z \leftarrow Y$ in $\mathcal{G}$, or if it has disjoint incoming directed paths from X and Y via (only) other (indirect) collider detour nodes for X and Y.

The relevance lies in the following property:

Lemma 4d. In a uDAG $\mathcal{G}$ to a faithful MAG $\mathcal{M}$, an edge $X \rightarrow Y$ is guaranteed to imply $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$ if $\mathbf{Z}$ contains all (indirect) triangle detour nodes for $X \rightarrow Y$, but no (indirect) collider detour nodes.

Proof sketch. If $X \rightarrow Y$ in $\mathcal{G}$, then this tells us that $X \not\perp\!\!\!\perp_p Y \,|\, An(Y)_{\mathcal{G}}$. Suppose X and Y are not adjacent in $\mathcal{M}$, then this implies that either:
(1) a set of nodes $\mathbf{U}$, necessary for separating X and Y in $\mathcal{M}$, is not in $An(Y)_{\mathcal{G}}$, and/or
(2) a set of nodes $\mathbf{W}$ that unblock a path between m-separated X and Y in $\mathcal{M}$ are in $An(Y)_{\mathcal{G}}$.

In case of (1): let U be the first node from $\mathbf{U}$ in (some global ordering that satisfies) the partial order induced by uDAG $\mathcal{G}$, then $X \rightarrow U \leftarrow Y$ in $\mathcal{G}$, and so U is part

of a collider detour for $X \to Y$. This follows from the fact that all nodes in $\mathbf{U}$ are part of some minimal separating subset (though not necessarily all together), and so U has a directed path to at least X or Y in $\mathcal{M}$, and is a noncollider on some unblocked path given the other nodes in a minimal separating set $\mathbf{Z}_i$ containing U. Therefore, conditional on any subset $An(Y)_{\backslash X}$ that still includes Y, U has an unblocked path to X in $\mathcal{M}$, so $X \to U$ in $\mathcal{G}$; similar for any subset $An(Y)_{\backslash Y}$ that still includes X, U has an unblocked path to Y in $\mathcal{M}$, so $Y \to U$ in $\mathcal{G}$. Similar for subsequent alternative blocking nodes from $\mathbf{U}$, except that now these may (or may not) be separated from X and/or Y in $\mathcal{M}$ by preceding nodes from $\mathbf{U}$ in $\mathcal{G}$, in which case they have an unblocked path to one or more of those nodes from $\mathbf{U}$, and so are part of an in direct collider detour.

In case of (2): let $\mathbf{Z}$ be the subset of predecessors of Y that are in $An(X, Y)_{\mathcal{M}}$, and let $\mathbf{W}$ be its complement $\mathbf{W} = An(Y)_{\mathcal{G}} \setminus \mathbf{Z}$. Then $X \perp\!\!\!\perp_p Y \mid \mathbf{Z}$, but there is also at least one unblocked path $\pi = \langle X, .., (Y) \rangle$ in $\mathcal{M}$ (partly) via a subset of collider nodes $\{W_1, .., W_k\} \subseteq \mathbf{W}$. All nodes along π (including X) have an unblocked path to Y in $\mathcal{M}$ given $\mathbf{Z} \cup \mathbf{W}$, and so arcs $\pi \to Y$ in $\mathcal{G}$. Using $\pi_{\backslash X}$ as shorthand for all nodes along π except X (and Y), then if $\pi_{\backslash X} \prec X$ in $\mathcal{G}$, then the same holds for arcs $\pi_{\backslash X} \to X$, and so all non-endpoint nodes along π form triangle detours for the edge $X \to Y$. If $\pi_{\backslash W} \prec W$, then W has unblocked paths to all other nodes along π in $\mathcal{M}$, and so $\pi_{\backslash W} \to W \to Y$ in $\mathcal{G}$, which again means they all form an (indirect) triangle detour for $X \to Y$.

For an arc $X \to Y$ in $\mathcal{G}$, if X and Y are also adjacent in $\mathcal{M}$ then they are dependent given any set. If not, then in case of (1) including nodes from $\mathbf{U}$ may separate them, but these nodes are all part of an (indirect) collider detour for $X \to Y$ in $\mathcal{G}$, and so excluding these from $\mathbf{Z}$ avoids destroying the dependence. Similarly, in case of (2) the dependence can be the result of an unblocked path due to conditioning on non-ancestors of X and Y in $\mathcal{M}$, but these are then all part of (indirect) triangle detours in $\mathcal{G}$, and so as long as all of these are included in $\mathbf{Z}$ the dependence $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$ is ensured. □

We can string these dependencies together to form longer paths.

Corollary 4e. In a uDAG $\mathcal{G}$ to a faithful MAG $\mathcal{M}$, if W is a noncollider between non-adjacent X and Y, and we can infer that $X \not\perp\!\!\!\perp_p W \mid \mathbf{Z}$ and $W \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$, then it also follows that $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$.

Proof. By contradiction: assume $X \perp\!\!\!\perp_p Y \mid \mathbf{Z}$. From the given there are unblocked paths π_{XW} and π_{WY} in $\mathcal{M}$ given $\mathbf{Z}$, and the assumption implies W would

need to be a collider between these paths in $\mathcal{M}$. There cannot exist alternative directed paths out of W to X and/or Y in $\mathcal{M}$: these would need to be blocked by $\mathbf{Z}$ to ensure the independence, but that would unblock the collider path, resulting in $X \perp\!\!\!\perp_p Y \mid \mathbf{Z}$, contrary the assumed.

As X and Y are not adjacent in $\mathcal{G}$ we can choose $\mathbf{U}$ from $An(X, Y, W)_{\mathcal{G}}$ such that there is only one unblocked path (edge) between X and W in $\mathcal{G}$, corresponding to an unblocked path π'_{XW} in $\mathcal{M}$, and likewise π'_{WY} in $\mathcal{M}$ for edge $W - Y$ in $\mathcal{G}$. By Lemma 4 this also implies identifiable dependency $X \not\perp\!\!\!\perp_p Y \mid \mathbf{U}$. But by construction, adding W to the conditioning set will separate X and Y in $\mathcal{G}$, giving $X \perp\!\!\!\perp_p Y \mid \mathbf{U} \cup W$ by Lemma 3. This implies that W cannot be a collider between π'_{XW} and π'_{WY} in $\mathcal{M}$, whereas the original assumption implies it must be. This means that the assumption must be false, and so indeed $X \not\perp\!\!\!\perp_p Y \mid \mathbf{Z}$. □

We can use the uDAG rules above to test observed (minimal) independencies in $\mathcal{G}$ for the required dependencies in Lemma 4b/c: if there are multiple unblocked paths for a given dependence, then validating any one of them via 4d/e corresponds to identifying an unblocked path in faithful MAG $\mathcal{M}$, which is sufficient to infer the dependence.

We can try to find additional uDAG rules to read even more dependencies, but that would neglect another important piece of information, namely that the uDAG is also *optimal*.

3.2 Causal inference from optimal uDAGs

In general, Lemmas 3-4 assert different dependencies for different uDAG members of the same equivalence class. For *optimal* uDAGs (oDAGs for short) additional information can be inferred.

Lemma 5a. If $\mathcal{G}$ is an *optimal* uDAG to a faithful MAG $\mathcal{M}$, then all in/dependence statements that can be inferred for any uDAG instance of the corresponding equivalence class $[\mathcal{G}]$ are valid.

Proof. All (DAG) instances in an equivalence class $[\mathcal{G}]$ can describe the same distribution with the same in/dependencies, and have the same number of free parameters. Therefore, if one is a valid (optimal) uDAG to the faithful MAG $\mathcal{M}$, then they all are. That means that all in/dependence statements derived for any of these via proper uDAG inference rules, e.g. Lemma 3, are valid in $\mathcal{M}$. □

Even though there can be different oDAGs (optimal uDAGs) for a MAG $\mathcal{M}$, it does mean that no edge in a given oDAG $\mathcal{G}$ can be removed without either

requiring an invariant bi-directed edge in the corresponding equivalence class, or implying in/dependence statements not present in $\mathcal{M}$.

For example, knowing that Fig.1(c) is an *optimal* uDAG implies $X \not\perp\!\!\!\perp Y$, whereas this does not follow for 'ordinary' uDAGs (Lemmas 3-4h only give $X \perp\!\!\!\perp_{\mathcal{G}} Y \mid ZW$).
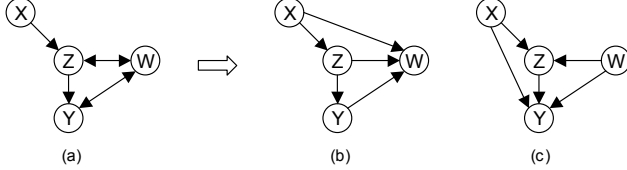

Figure 1: (a) MAG with invariant bi-directed edges ($\mathcal{R}4b$), (b) optimal uDAG if $r_W \leq (r_Y - 1)r_Z + 1$, (c) idem, if $r_W \geq (r_Y - 1)r_Z + 1$; with r_X the multiplicity of random variable X, etc.

It also follows that inference is most naturally done on the PAG representation $\mathcal{P}$ of the graph. In this paper we focus on deriving causal statements. For that we need to establish a connection between the underlying faithful MAG $\mathcal{M}$ and an optimal uDAG representation $\mathcal{G}$ (where we ignore selection bias).

Lemma 5b. For a faithful MAG $\mathcal{M}$, an optimal uDAG $\mathcal{G}$ is a member of an equivalence class $[\mathcal{M}']$ obtained by (only) adding arcs to $\mathcal{M}$, necessary to eliminate an arrowhead from a bi-directed edge in the PAG $\mathcal{P}(\mathcal{M}')$, until no more invariant bi-directed edges are left.

Proof sketch. If the PAG $\mathcal{P}(\mathcal{M})$ does not contain a bi-directed edge, then there exists a DAG representative of the corresponding equivalence class $[\mathcal{M}]$, see Theorem 2 in (Zhang, 2008). As fewer edges and fewer invariant edge marks require fewer free parameters, any such DAG is also optimal.

If the PAG $\mathcal{P}(\mathcal{M})$ *does* contain edges with two invariant arrowheads then a uDAG approximation is needed. From (Claassen and Heskes, 2011a,b) we know that all invariant arrowheads at a node Z on a bi-directed edge $Z \leftrightarrow Y$ in a PAG are inferred from (minimal) conditional independencies $U \perp\!\!\!\perp_p V \mid [\mathbf{W}]$ with $Y \in (\{U, V\} \cup \mathbf{W})$ that are destroyed by conditioning on the arrowhead node Z. In this minimal dependence $U \not\perp\!\!\!\perp_p V \mid \mathbf{W} \cup [Z]$ node Z has distinct unblocked incoming paths in $\mathcal{P}$ from U and V given $\mathbf{W}$ (so Z also has another invariant arrowhead to some other node, apart from Y). As a uDAG leaves every unblocked path in $\mathcal{M}$ intact, the only way to eliminate the invariant arrowhead is to 'hide' the conditional independence, by either adding an edge $U-V$, or adding edges to extend the required separating set $\mathbf{W}$. But

additional nodes in the separating set can only hide the independence if at least one of these, say W', helps to block all paths between Z and, say, U. But then W' and V would also need to be separated in $\mathcal{M}$ (otherwise $W' \in \mathbf{W}$), and so the invariant arrowhead at Z still follows from a conditional independence destroyed by Z, i.e. $W' \not\perp\!\!\!\perp_p V \mid .. \cup [Z]$, *unless* an edge is added between the two separated nodes.

In short: to eliminate an invariant arrowhead $Z \leftarrow\!* Y$ edges need to be added in $\mathcal{M}$ between the two separated nodes in a non-empty subset of (minimal) conditional independencies destroyed by Z to obtain an unfaithful MAG $\mathcal{M}'$. Each added edge is in the form of an arc with the arrowhead at Y (or arbitrary orientation if $Y \in \mathbf{W}$), *unless* this necessarily results in an almost directed cycle (not permitted in a MAG)), in which case the added edge itself becomes a bi-directed edge, which then has to be eliminated in a subsequent step. How to find which (minimal set of) edges need to be added in each step is not important to us here.

Once all required edges have been added, the collider(s) at Z is/are no longer invariant, and the newly implied possible dependence in the MAG $\mathcal{M}'$ via Z is compensated for by the implied dependencies via the added edges, in combination with parameter constraints that ensure these separate paths cancel out each other exactly. After this step the PAG $\mathcal{P}(\mathcal{M}')$ is recomputed. This process is repeated until all invariant bi-directed edges have been eliminated. At that point a there is a DAG instance $\mathcal{G}$ in the equivalence class $[\mathcal{M}']$, which is a uDAG to the faithful MAG $\mathcal{M}$, for which the number of free parameters can be calculated.

Choosing different arrowheads to eliminate in each step can lead to different uDAGs with different numbers of free parameters: the smallest one(s) correspond to the optimal uDAG(s) $\mathcal{G}$ to the faithful MAG $\mathcal{M}$. $\square$

Note that a given MAG can have different optimal uDAG representations, possibly depending on the multiplicity of the variables as well.

Having established a connection between an optimal uDAG $\mathcal{G}$ and the underlying faithful MAG $\mathcal{M}$, we can translate this information into causal inference from the PAG representation $\mathcal{P}(\mathcal{G})$ of the observed uDAG. Fortunately the inference rule for absent causal relations takes a particularly simple form, identical to that for regular, faithful PAGs in (Claassen and Heskes, 2010). It uses the notion of a potentially directed path (p.d.p.), introduced in §1.

Lemma 5. Let $\mathcal{G}$ be an optimal uDAG to a faithful MAG $\mathcal{M}$, then the absence of a causal relation $X \not\Rightarrow Y$ can be identified, iff there is no potentially directed path from X to Y in the PAG $\mathcal{P}$ of $[\mathcal{G}]$.

Proof. From Lemma 5a we know that the optimal uDAG $\mathcal{G}$ is obtained by (only) adding arcs between variables in the MAG $\mathcal{M}$ to eliminate invariant bi-directed edges, until no more are left. By construction, all arrowheads on arcs added in each step to obtain the next $\mathcal{M}'$ satisfy the non-ancestor relations in $\mathcal{M}$. Therefore, any remaining invariant arrowhead in the corresponding PAG $\mathcal{P}(\mathcal{M}')$ matches a non-ancestor relation in the original MAG $\mathcal{M}$. For a MAG all nodes not connected by a potentially directed path (p.d.p.) in the corresponding PAG have a definite non-ancestor relation in the underlying causal graph, see Theorem 2 in (Claassen and Heskes, 2010). As all unblocked paths in $\mathcal{M}$ are left intact in $\mathcal{M}'$ and $\mathcal{G}$, and a p.d.p. is by definition an unblocked path given the empty set, adding edges at each step can only *hide* non-ancestor relations still identifiable in the previous step, but never introduce new ones.

For an optimal uDAG $\mathcal{G}$ it holds that $\mathcal{P}(\mathcal{G}) = \mathcal{P}(\mathcal{M}')$ (at least for one of the possible MAG solutions), and so if there is no p.d.p from X to Y in the PAG $\mathcal{P}(\mathcal{G})$, then there is no p.d.p. from X to Y in $\mathcal{M}'$, and so also none in $\mathcal{M}$, which implies the absence of a causal relation $X \not\Rrightarrow Y$. But no more then these can be inferred, as the uDAG also matches itself as faithful MAG, and for that MAG the nodes not connected by a p.d.p. in $\mathcal{P}$ are all that can be identified. $\square$

For causal alternatives a result can also be found. Like for 'regular' uDAGs, it is based on identifying minimal conditional independencies via Lemma 4b: start from a minimal separating set $X \perp\!\!\!\perp_{\mathcal{G}} Y \,|\, [\mathbf{Z}]$ in the graph $\mathcal{P}(\mathcal{G})$, and establish dependencies from each node $Z \in \mathbf{Z}$ to X and Y given the other $\mathbf{Z}_{\setminus Z}$.

Naturally, to identify independencies in *optimal* uDAGs we can use all uDAG Lemmas from §3.1, including in particular the 'only one unblocked path' result (Lemma 4). But we now also utilize a variant of the triangle/collider detours in Lemma 4d, based on the fact that if an edge in $\mathcal{P}(\mathcal{G})$ cannot 'hide' an invariant bi-directed edge, then it cannot invalidate that edge as a dependence.

Lemma 6a. Let $\mathcal{G}$ be an optimal uDAG to a faithful MAG $\mathcal{M}$, and $\mathcal{P}$ the corresponding PAG of $\mathcal{G}$. Then, an edge $X - Y$ in $\mathcal{P}$ corresponds to an identifiable dependence if all nodes Z in a triangle with X and Y either satisfy:
(1) $Z \ast\!\!\to X$ and/or $Z \ast\!\!\to Y$ are *not* in $\mathcal{P}$, or
(2) $Z \,\text{---}\!\ast X$ and/or $Z \,\text{---}\!\ast Y$ are in $\mathcal{P}$.

Proof. The proof of Lemma 4d for regular uDAGs showed that if an edge $X - Y$ was present in $\mathcal{G}$ but not in the underlying MAG $\mathcal{M}$, then it showed in the presence of either a triangle or collider detour. For optimal uDAGs collider detours do not apply, as they only introduce additional arrowheads in $\mathcal{G}$ using more parameters, and so do not appear in the construction of a optimal uDAG in Lemma 5b. Remains to show that a node in a triangle in $\mathcal{P}(\mathcal{G})$ cannot correspond to a triangle detour:

If (1) applies, then if Z was oriented as a collider between X and Y (removing edge $X - Y$) then it would represent the same equivalence class but with fewer parameters, and so the fact that this did not occur implies this is not the case for Z.

If (2) applies, then Z is definitely a noncollider between X and Y (though not necessarily ancestor of), and so Z cannot have an invariant bi-directed edge to either that is 'hidden' in $\mathcal{G}$ by edge $X - Y$.

If this holds for all nodes in a triangle with edge $X - Y$, then there is no node that can hide an implicit collider that invalidates the edge, and so the edge represents a direct dependence. $\square$

We can extend this to identifiable dependencies by finding a path that can be validated through Lemma 6a. In this we use the term *base path* to indicate a path/edge that is not itself a triangle detour of another path/edge in $\mathcal{P}$.

Corollary 6b. Let $\mathcal{G}$ be an optimal uDAG to a faithful MAG $\mathcal{M}$, and $\mathcal{P}$ the corresponding PAG of $\mathcal{G}$. Then a dependence $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$ can be inferred if there is an unblocked base path in $\mathcal{P}$ between X and Y given $\mathbf{Z}$ along which all edges can be verified to represent a direct dependence.

Proof. If we can validate all edges along the path, e.g. by Lemma 4 or 6a, then we can string these together similar to Corollary 4e, to establish the existence of an unblocked path in the underlying faithful MAG $\mathcal{M}$ between X and Y given $\mathbf{Z}$, which ensures the dependence $X \not\perp\!\!\!\perp_p Y \,|\, \mathbf{Z}$. $\square$

And one special alternative to validate edges as dependencies from already identified in/dependencies:

Lemma 6c. Let $\mathcal{G}$ be an optimal uDAG to a faithful MAG $\mathcal{M}$, and $\mathcal{P}$ the corresponding PAG of $\mathcal{G}$. Then for adjacent $X - Y$ in $\mathcal{P}$ a dependence $X \not\perp\!\!\!\perp_p Y \,|\, Z$ can be inferred, if there exist identifiable $X \not\perp\!\!\!\perp_p Y \,|\, W$ and $X \perp\!\!\!\perp_p Z \,|\, [W]$.

Proof. If $X - Y$ in $\mathcal{P}(\mathcal{G})$, then if $X \not\perp\!\!\!\perp_p Y \,|\, W$ and $X \perp\!\!\!\perp_p Z \,|\, [W]$, then $X \not\perp\!\!\!\perp_p Y \,|\, Z$ Proof: if also $X - Y$ in $\mathcal{M}$ then X and Y are dependent given any set, so also given Z. If $X \not\perp\!\!\!\perp_p Y$ but not adjacent in $\mathcal{M}$ then they remain dependent given Z, as if Z blocks all paths between X and Y, and W blocks all paths between X and Z, then W would also block all paths between X

and Y, contrary $X \not\perp\!\!\!\perp_p Y \mid W$. Finally, if $X \perp\!\!\!\perp_p Y$ in $\mathcal{M}$, but conditioning on W unblocks a path between them, then W cannot have a directed path to X (or Y). But that means that $X \perp\!\!\!\perp_p Z \mid [W]$ implies that W does have a directed path to Z in $\mathcal{M}$, and so if W unblocks the path, then so does descendant Z, and so $X \not\perp\!\!\!\perp_p Y \mid Z$. $\qquad\square$

Note that Lemmas 6a/c does not identify the minimal independencies themselves, but only verifies the dependencies (in Lemma 4b) required to find them. Also note that Lemma 6c builds on minimal independencies already found, which implies a recursive approach is needed to find the full mapping.

For every oDAG we may infer additional causal information by applying the standard causal inference rules on the statements obtained via Lemmas 5-8. Together this results in the mapping from each oDAG $\mathcal{G}$ to the set of logical causal statements L as used in the BCCD algorithm. Interestingly enough, for optimal uDAGs up to four nodes the mapping is *identical* to that for regular, faithful DAGs. Only at five or more nodes the distinction becomes relevant.

3.3 Further improvements

We do not claim that the current oDAG mapping is complete in the sense that it is guaranteed to extract the maximum amount of causal information from all possible oDAGs. A brute-force check showed that the combined inference rules above cover all information for optimal uDAGs up to five nodes.

Additional improvements can be made: the (optimal) uDAG inference rules can be extended to infer even more dependencies and/or causal information. If possible this should take the form of an easy-to-use sound complete graphical criterion in the vein of *d*-separation, but based on our impressions so far this seems rather ambitious. Instead of only extracting logically valid statements we can weigh different implied causal statements by the proportion of underlying MAGs in which it holds. This would produce a more informative, weighted mapping to a list of causal statements L, where only the entries $p(L|\mathcal{G}) \geq 0.5$ need to be kept. It seems not possible to infer this type of mapping from graphical rules alone, and so it would have to rely on brute-force computation.

For larger graphs some form of Monte Carlo sampling can be applied to obtain the (weighted) mapping to causal statements. It would be very expensive to compute, but could provide valuable information to decide on borderline cases, or simply as an independent confirmation for parts of the inferred structure. Scoring MAGs or even PAGs directly would

eliminate the need for 'unfaithful inference' altogether. This would simplify the entire process considerably, but that still leaves the problem of the huge number of possible graphs to consider (score) for larger sets of variables. Sampling MAGs could also be employed to obtain or confirm probability estimates for causal information derived from *combinations* of separately obtained logical statements, as described in (Claassen and Heskes, 2011b). We expect such combinations to be highly dependent, but especially for less certain ones, e.g. two statements with $p(L_{1,2}|\mathcal{D}) = 0.6$, the difference between a fully independent estimate: $p(L_1|\mathcal{D}) \cdot p(L_2|\mathcal{D}) = 0.36$, and a fully dependent estimate: $min\,(p(L_1|\mathcal{D}), p(L_2|\mathcal{D})) = 0.6$, is considerable. Having a means to obtain a more principled reliability estimate for the combination can improve the overall accuracy of the BCCD algorithm.

References

R. Bouckaert. *Bayesian Belief Networks: From Construction to Inference*. PhD thesis, University of Utrecht, 1995.

T. Claassen and T. Heskes. A structure independent algorithm for causal discovery. In *ESANN-19*, pages 309–314, 2011a.

T. Claassen and T. Heskes. Causal discovery in multiple models from different experiments. In *NIPS-23*, pages 415–423. 2010.

T. Claassen and T. Heskes. A logical characterization of constraint-based causal discovery. In *Proc. of the 27th Conference on Uncertainty in Artificial Intelligence*, 2011b.

R. Neapolitan. *Learning Bayesian Networks*. Prentice Hall, 1st edition, 2004.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman Publishers, San Mateo, CA, 1988.

T. Richardson and P. Spirtes. Ancestral graph Markov models. *Ann. Stat.*, 30(4):962–1030, 2002.

J. 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.