View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CiteSeerX

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'.

It has three main parts: section one defines causal- and graphical model terminology used throughout the supplement. Section two relates to the mapping from uDAGs to logical causal statements; much of this builds on (Bouckaert, 1995). The third section details the result that in the large sample limit all required minimal independencies and dependencies can be found. The supplement follows the numbering in the original 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 AB.

The joint probability distribution induced by a causal DAG $\mathcal{G}_C$ factors according to a *Bayesian network* (BN): a pair $\mathcal{B} = (\mathcal{G}, \mathbf{\Theta})$, where $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ is DAG over random variables $\mathbf{V}$, and the parameters $\boldsymbol{\theta}_V \subset \mathbf{\Theta}$ 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 $X — Y$ (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 $A {*\!\!\rightarrow} B$ 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 presence or absence of causal relations between two or three variables of the form $(X \Rightarrow Y)$, $(X \Rightarrow Y) \vee (X \Rightarrow Z)$, 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 $\mathcal{G}_C$. Ł 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

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

$$(X \perp\!\!\!\perp_G Y \,|\, \mathbf{Z}) \Leftrightarrow (X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}),$$

Proof. See §3.3.3 (Bouckaert, 1995) starting from the graphoid axioms. Our assumption of an underlying faithful MAG is stronger. Among other things it implies that a node U cannot have a dependence with a node X (or Y) given $\mathbf{Z}$ if there is no unblocked path between U and X in $\mathcal{G}$ conditional on $\mathbf{Z}$. As a result, a dependence relation between X and Y given $\mathbf{Z}$ cannot be destroyed by including or excluding such a node U from the set $\mathbf{Z}$. This applies whether U is in the parent set of X or Y in $\mathcal{G}$ or not, relaxing one of the coupling criteria. □

So, all independencies from d-separation remain valid, but identifiable dependencies put restricitions on the set $\mathbf{Z}$ For *optimal* uDAGs additional information can be inferred.

Lemma 4. 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 instances of the corresponding equivalence class $[\mathcal{G}]$ via Lemma 3 are valid.

Proof. All instances in an equivalence class can describe the same distribution, with the same in/dependencies, and have the same number of free parameters. Therefore, if one is a optimal uDAG to the faithful MAG $\mathcal{M}$, then they all are, and a validly inferred statement, e.g. via Lemma 3, for any of these is therefore valid. □

For absent causal relation this ultimately reduces to:

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 of $[\mathcal{G}]$.

Proof. The optimal uDAG $\mathcal{G}$ is obtained by (only) adding edges between variables in the MAG $\mathcal{M}$ to eliminate invariant bi-directed edges, until no more are left. At that point the uDAG is a representative of the corresponding equivalence class $\mathcal{P}$, see Theorem 2 in (Zhang, 2008). For any faithful MAG all and only the nodes not connected by a p.d.p. in the corresponding PAG have a definite non-ancestor relation in the underlying causal graph. At least one uDAG instance in the equivalence class of an optimal uDAG over a given skeleton leaves the partial order of the original MAG intact, because extra parameters are needed to force additional arrowheads on edges into v-structures that run counter to the ancestor relation in the MAG, which means the uDAG was not optimal. Therefore, any remaining invariant arrowhead in the PAG $\mathcal{P}$ matches a non-ancestor relation in the original MAG. So all nodes not connected by a p.d.p. in $\mathcal{P}$ are definite non-ancestors, as they certainly would not be connected by such a path in a graph containing less edges and a superset of invariant arrowheads on those edges. But no more then these can be inferred, as the uDAG also matches itself as MAG, and for that MAG the nodes not connected by a p.d.p. in $\mathcal{P}$ are all that can be identified. □

For causal alternatives a similar, but more complicated criterion can be found:

Lemma 6. Let $\mathcal{G}$ be an optimal uDAG to a faithful MAG $\mathcal{M}$, then $(Z \Rightarrow X) \vee (Z \Rightarrow Y)$ may be inferred if $X \perp\!\!\!\perp_G Y \,|\, [\mathbf{Z}]$ in $\mathcal{G}$, and for each $Z \in \mathbf{Z}$, either a path to X or Y with an invariant tail at Z can be validated, or both paths to X and Y can be validated, where validation implies that no edge along the unblocked path from X/Y to Z is part of the collider part of a shielded collider triangle.

We do not claim that the current uDAG mapping, in combination with deduction on standard causal properties, is complete in the sense that they are guaranteed to extract the maximum amount of information from all possible uDAGs. However, a brute-force check showed that they cover all valid mappings for uDAGs up to five nodes.

3 Finding invariant arrowheads

The submission mentions the fact that, even though individual optimal uDAGs can hide in/dependencies present in the underlying MAG, in the large sample limit a strategic search using uDAGs is still guaranteed to find the entire skeleton and all invariant arrowheads. The practical relevance of this result is limited, as it can involve scoring DAGs over large sets of nodes, not to mention infinite sample sizes, which is why it is largely left out of the current submission.

The proof is still fairly convoluted, so as a guide to the steps involved (in reverse order):

- At Theorem 1, the entire skeleton and all invariant arrowheads are found.
- Before that, Lemma 13 and Corollary 14 show that we can find all nodes that destroy an independence, from which to find the arrowheads from uDAGs;
- Similarly, Lemmas 10 and 11, together with Corollary 12, show the skeleton is found. First assuming causally sufficiency, then for the general case.
- These lemmas rely on Lemma 8 and Corollary 9, on (in)dependence relations that hold between nodes in a minimal conditional independence.

3.1 Minimal independencies

We start by showing that all minimal independencies $X \perp\!\!\!\perp_M Y \mid [\mathbf{Z}]$ are guaranteed to show up in a optimal uDAG over just those nodes. The proof is split in two parts: first for the case where all separating nodes $\mathbf{Z}$ are adjacent to either X and/or Y. Then the general proof that also allows non-adjacent separating nodes: assuming causal sufficiency this case cannot occur, but without causal sufficiency it can, invalidating many causal discovery methods that rely on finding separating sets from neighbouring nodes. It was this possibility that inspired/necessitated the development of the FCI-algorithm (Spirtes et al., 1999, 2000). The proof for this general case, dubbed 'FCI-DSep', follows naturally from the causally sufficient case. Both proofs use Lemma 7 to derive a contradiction between the presence of an edge $X - Y$ in the uDAG over $(X, Y, \mathbf{Z})$ and the assumption that the uDAG is optimal.

First three results used in subsequent proofs.

Lemma 7. Let $\mathcal{G}$ be a DAG over nodes $\mathbf{X}$. Let $\mathbf{X}_C \subseteq \mathbf{X}$ be a fully connected subset in $\mathcal{G}$ (clique), that all share the same external parents (from $\mathbf{X} \backslash \mathbf{X}_C$). Then all arcs between nodes in $\mathbf{X}_C$ are reversible, and $\mathbf{X}_C$ forms a fully connected circle component in the equivalence class of $\mathcal{G}$.

Proof. The DAG $\mathcal{G}$ induces a partial order in which the nodes $\mathbf{X}_C$ (can) form a successive block. Each relative ordering of $\mathbf{X}_C$ leaves all in/dependencies intact, and is therefore member of the same equivalence class as $\mathcal{G}$. As all different orderings imply that all arcs (parent-child relation) between nodes in $\mathbf{X}_C$ can be interchanged, and so none of the tail-arrowheads are invariant, and therefore obtain a circle mark in the corresponding PAG; see also (Meek, 1995). □

In the proof to show that we can find all required (minimal) independencies, we use the next result about dependencies that apply to all nodes in the conditioning set:

Lemma 8. In a faithful MAG $\mathcal{M}$, let $X \perp\!\!\!\perp_M Y \mid [\mathbf{Z}]$, $Z \in \mathbf{Z}$ and $\mathbf{Z}' \subseteq \mathbf{Z}_{\backslash Z}$, then:

1. Z is dependent on both X and Y given *all* $\mathbf{Z}_{\backslash Z}$,
2. Z is dependent on (at least one of) X and/or Y given any subset $\mathbf{Z}'$,
3. if $X \notin An(Y)$, then Z is dependent on Y given any $\mathbf{Z}' \cup X$, and
4. all $\{Z_i, Z_j\} \in \mathbf{Z}$ are dependent conditional on any set $\mathbf{Z}' \subseteq \mathbf{Z}_{\backslash \{Z_i, Z_j\}} \cup \{X, Y\}$.

Proof. (1.) Follows from directly from 'minimal' (so it is the only noncollider on some unblocked path given all others);
(2.) By theorem 3.x, each $Z \in \mathbf{Z}$ has a directed path in $\mathcal{M}$ to X and/or Y, and is therefore dependent given the empty set; each node in $\mathbf{Z}$ is either adjacent to X in $\mathcal{M}$, or, if not adjacent to X, connected to it via a sequence of bi-directed edges (Spirtes et al., 1999). Same holds for Y (although non-adjacent nodes may differ). Every path π between X and Y that is only blocked by noncollider Z given $\mathbf{Z}$ is either a direct edge or a sequence of colliders. Any subset $\mathbf{Z}' \subset \mathbf{Z}$ that lacks one of the (descendants of) these colliders on π may break the dependency along the path π, but opens up a new dependence via the directed path from the first eliminated node along π (in both directions, starting from Z) to either X and/or Y. As a result, conditional on any set at least one dependence (unblocked path) will remain intact.
(3.) If Z was already dependent on Y given $\mathbf{Z}'$, then conditioning on collider X can never block that path. If not, then $\mathbf{Z}'$ is a strict subset of $\mathbf{Z}_{\backslash Z}$, by (1). By (2) it then has an unblocked path π to X given $\mathbf{Z}'$. As $\mathbf{Z}$ is minimal, there is now also an unblocked path π' given $\mathbf{Z}'$ between X and Y, as not all remaining $\mathbf{Z}$ are included. If X is not an ancestor of Y then both these paths are into X, and so conditioning on collider X will unblock the path $\pi + \pi'$ from Z to Y, and so $Z \not\perp\!\!\!\perp_P Y \mid \mathbf{Z}' \cup X$.
(4.) If $X \in An(Y)$, then all $\mathbf{Z}$ are adjacent to descendant Y (Colombo et al., 2011), so then conditioning

on any set that includes Y will make them dependent; vice versa for Y. If neither is ancestor of the other, then all paths are into X and Y. By (2), if both Z_i and Z_j are dependent on X, or both on Y, then including these as collider will unblock a path, making them dependent. Same holds if $\mathbf{Z}'$ contained all other nodes, for then Z_i and Z_j would have unblocked paths to both X and Y (or each other). If $\mathbf{Z}'$ did not contain all other nodes, then (as $\mathbf{Z}$ was minimal) there is an unblocked path between X and Y. So in that case, if Z_i and Z_j have unblocked paths to different nodes from (X, Y) given $\mathbf{Z}'$, then also conditioning on both $X+Y$ will unblock the path $Z_i - X - Y - Z_j$, leaving them dependent again, ergo $Z_i \not\perp\!\!\!\perp_P Z_j \,|\, \mathbf{Z}' \cup \{X, Y\}$. □

This lemma also implies the following independence result, used in the proof for causal sufficiency:

Corollary 9. Let $X \perp\!\!\!\perp_M Y \,|\, [\mathbf{Z}]$ in a faithful MAG $\mathcal{M}$, and let $\mathcal{G}$ be a optimal uDAG over $\{X, Y\} \cup \mathbf{Z}$. If all $\mathbf{Z}$ are both non-descendants of Y and adjacent to Y, then no two nodes from $(X \cup \mathbf{Z})$ are independent conditioned on any subset of other nodes in $\mathcal{G}$ that includes Y.

Proof. There are two cases, depending on whether X is part of the pair or not. Case (1): for the pair $\{Z_i, Z_j\} \subseteq \mathbf{Z}$, the path $Z_i *\!\!\rightarrow Y \leftarrow\!\!* Z_j$ is unblocked in $\mathcal{M}$ given any subset that includes Y. Case (2): for the pair $\{X, Z\}$, with $Z \in \mathbf{Z}$, for any strict subset $\mathbf{Z}' \subsetneq (\mathbf{Z} \setminus Z)$ there is an unblocked path from Y to X given $\mathbf{Z}'$, and so the path $Z *\!\!\rightarrow Y \leftarrow\!\!* ..X$ is unblocked in $\mathcal{M}$ given any set $\mathbf{Z}' \cup Y$; for the subset $\mathbf{Z}' = (\mathbf{Z} \setminus Z)$, by Lemma 8, there is an unblocked path between X and Z given $\mathbf{Z}'$ which cannot be blocked by conditioning on Y. □

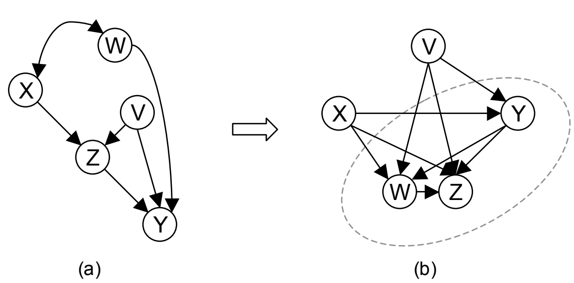

Figure 1: Illustration for proof of Lemma 10: (a) MAG with $X \perp\!\!\!\perp Y \,|\, [V, W, Z]$, (b) (non-optimal) uDAG with arc $X \longrightarrow Y$

When causal sufficiency applies, all minimal independencies can be found from the right uDAGs:

Lemma 10. Let $X \perp\!\!\!\perp_M Y \,|\, [\mathbf{Z}]$ be a minimal conditional independence in a faithful MAG $\mathcal{M}$ over $\{X, Y\} \cup \mathbf{Z}$, and let $\mathcal{G}$ be an optimal uDAG approximation to $\mathcal{M}$. If all separating nodes are adjacent to either X or Y, i.e. if either $\mathbf{Z} \subseteq Adj(X)$ or $\mathbf{Z} \subseteq Adj(Y)$ in $\mathcal{M}$, then (in the large sample limit) $\mathcal{G}$ implies $X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$.

Proof. By contradiction. Let $X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$, and let $\mathcal{G}$ be the (alleged) optimal uDAG over $(X, Y, \mathbf{Z})$ with edge $X - Y$. As it is a minimal independence, in the underlying MAG all nodes $\mathbf{Z}$ are ancestor of either X or Y, see (Spirtes et al., 1999), so at least one of these, say Y is not ancestor of any. As stated, we assume that the independence is not of the FCI-DSep form that requires nodes not adjacent to either X or Y in the MAG to separate them. Therefore, in the marginal MAG over $(X, Y, \mathbf{Z})$ all nodes $\mathbf{Z}$ are adjacent to (non-ancestor) Y. In $\mathcal{G}$, the set $\mathbf{Z}$ can be divided into a subset $\mathbf{W} \subseteq \mathbf{Z}$ containing descendants of both X and Y in $\mathcal{G}$, and a set $\mathbf{V} = \mathbf{Z} \setminus \mathbf{W}$ that are ancestor of at least one of $\{X, Y\}$. Note that the set $\mathbf{V}$ d-separates X and Y in $\mathcal{G}_{X \| Y}$. Also note that $\mathbf{W}$ cannot be empty, as otherwise $\mathcal{G}$ would erroneously imply $X \not\perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$.

We now show that if this $\mathbf{W}$ exists, then it must be of a very specific, heavily connected form. First, all nodes in $\mathbf{W}$ have an incoming arc from Y, as they are assumed to be descendant of Y in $\mathcal{G}$, and were adjacent to Y in $\mathcal{M}(X, Y, \mathbf{Z})$, and the skeleton of $\mathcal{G}$ must be a superset of $\mathcal{M}$.

Next, all nodes in $\mathbf{W}$ must be fully connected to each other in $\mathcal{G}$. Otherwise, $\mathcal{G}$ would imply a conditional independence of the form $W_i \perp\!\!\!\perp_P W_j \,|\, .. \cup Y$, which is impossible by Corollary 9. Also, all nodes in $\mathbf{W}$ must be connected to all nodes in $\mathbf{V}$, as otherwise again a conditional independence $W_i \perp\!\!\!\perp_P V_j \,|\, .. \cup Y$ is implied which is not present according to Corollary 9. By the same token, all $\mathbf{W}$ are directly connected to X, because any implied independence $W_i \perp\!\!\!\perp_P X \,|\, .. \cup Y$ would again be prohibited by Corollary 9. So all $\mathbf{W}$ are descendants of and adjacent to X implying all incoming arcs $X \longrightarrow \mathbf{W}$.
Now assume (case 1): $X \longrightarrow Y$ in $\mathcal{G}$. Then all edges $\mathbf{V} - Y$ are arcs into Y:

- any node $V_i \in \mathbf{V}$ that is not adjacent to X must be connected as $X \longrightarrow Y \longleftarrow V_i$, otherwise it implies an independence with Y in the conditioning set, which is forbidden by Corollary 9,
- any node $V_i \in \mathbf{V}$ that is adjacent to X and has an arc $Y \longrightarrow V_i$ must also have an arc $X \longrightarrow V_i$ (by acyclicity, given $X \longrightarrow Y \longrightarrow V_i$), but that would make V_i part of $\mathbf{W}$, so it too must have an arc $V_i \longrightarrow Y$

By acyclicity, that also implies that all edges between $\mathbf{V} - \mathbf{W}$ are arcs into $\mathbf{W}$, which makes $\mathbf{W} \cup Y$ a fully connected set with the same external parents $\mathbf{V} \cup X$. Next assume (case 2): $X \longleftarrow Y$ in $\mathcal{G}$. Then every node $V_i \in \mathbf{V}$ must be connected to X, otherwise it would imply an impossible independence $X \perp\!\!\!\perp_P V_i \,|\, .. \cup Y$ (by Corollary 9). All edges $V_i - X$ must be arcs into X, otherwise acyclicity would also require an arc $V_i \longleftarrow Y$, which would make $V_i \in \mathbf{W}$. But that makes $\mathbf{W} \cup X$ a fully connected set with the same external parents $\mathbf{V} \cup Y$. By Lemma 7, in both cases this means the graph $\mathcal{G}$ is equivalent with (in the same equivalence class as) a DAG $\mathcal{G}'$, in which either X or Y is the last node in the induced order, with only incoming arcs. But for that instance the edge $X - Y$ would not be needed, and so at least one of the equivalent graphs would need fewer parameters than $\mathcal{G}$, and be returned as the minimal instead. Ergo, if $X \perp\!\!\!\perp_M Y \,|\, [\mathbf{Z}]$, then the optimal DAG approximation $\mathcal{G}(X, Y, \mathbf{Z})$ has no edge $X - Y$, and the independence can therefore be identified from the graph. □

The case for minimal separating sets in general, i.e. that may involve 'FCI-DSEP' nodes in the separating set $\mathbf{Z}$ that are not adjacent to either X or Y in the marginal $\mathcal{M}$ goes similar, but with a somewhat modified 'adjacent to Y' part. First note that this remaining case implies that neither X nor Y is an ancestor of the other (Colombo et al., 2011), and that in the marginal MAG the non-adjacent nodes $\mathbf{Q} \subset \mathbf{Z}$ are connected to X and/or Y via a sequence of bi-directed edges, i.e. $\pi = X \longleftrightarrow Z_i \longleftrightarrow .. \longleftrightarrow Z_j \longleftarrow\!\!* \, Q_k$ in $\mathcal{M}$. Note that nodes 'nonadjacent to X' can still be adjacent to Y: the FCI-DSEP case only implies that there is no set $\mathbf{Z} \subseteq Adj(X)$ or $\mathbf{Z} \subseteq Adj(Y)$ that separates X and Y.

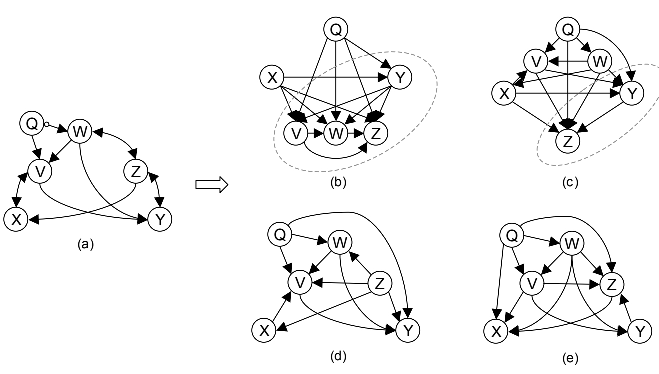

Figure 2: Illustration for proof of Lemma 11: (a) MAG with $X \perp\!\!\!\perp Y \,|\, [V, W, Z, Q]$ with non-adjacent node Q, (b)/(c) (non-optimal) uDAGs with arc $X \longrightarrow Y$, (d)/(e) different optimal uDAGs

We can now do the general proof for minimal separating sets:

Lemma 11. Let $X \perp\!\!\!\perp_M Y \,|\, [\mathbf{Z}]$ be a minimal conditional independence in a faithful MAG $\mathcal{M}$ over $\{X, Y\} \cup \mathbf{Z}$. If $\mathcal{G}$ is a optimal DAG approximation to $\mathcal{M}$, then $\mathcal{G}$ implies $X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$.

Proof. By contradiction. The case for all $\mathbf{Z}$ adjacent is covered in Lemma 10. That leaves the case with a nonempty subset $\mathbf{Q} \subset \mathbf{Z}$ that is not adjacent to either X or Y in $\mathcal{M}$. As before, we use $\mathbf{W} \subseteq \mathbf{Z}$ to denote the nodes in the (optimal) DAG $\mathcal{G}$, that are descendants of both X and Y, and $\mathbf{V} = \mathbf{Z} \setminus \mathbf{W}$ to denote the ones that have not. Again, $\mathbf{W}$ is not empty, otherwise an edge $X - Y$ would erroneously imply $X \not\perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$, and so we derive a contradiction from the assumption of non-empty $\mathbf{W}$ through equivalent structures.

First, all $W \in \mathbf{W}$ are connected to X and Y by a direct incoming arc. Suppose there is no edge $W - Y$, such that $W \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}_{WY}$, for some set $\mathbf{Z}_{WY}$. Then by Lemma 8(2) W is dependent on ancestor X (in $\mathcal{G}$), i.e. there is an unblocked path π between X and W given $\mathbf{Z}_{WY}$. If either π or the edge $X - Y$ is out of X, then X is the only noncollider on an unblocked path between Y and W, and must therefore be included in the set $\mathbf{Z}_{WY}$. But that would imply an independence that does not hold in $\mathcal{M}$ by Lemma 8(3). If both π and the edge $X - Y$ are into X, then (any descendant of) X cannot be included in $\mathbf{Z}_{WY}$, for then there would still be an unblocked path. But as W is a descendant of X, that would leave the path $Y \longrightarrow X \longrightarrow .. \longrightarrow W$ unblocked, and so there is no set in $\mathcal{G}$ that can separate W and Y without contradiciting $\mathcal{M}$. Ergo, all nodes $\mathbf{W}$ have a direct, incoming arc from Y; and similarly an incoming arc from X.

Like before, $\mathbf{W}$ itself is also fully connected in $\mathcal{G}$, otherwise it would imply some conditional independence $W_i \perp\!\!\!\perp_P W_j \,|\, \mathbf{Z}_{W_i \| W_j} \cup \{X, Y\}$, which does not hold in the MAG: by Lemma 8(4).

By the same rationale, all $\mathbf{V}$ are connected to all nodes $\mathbf{W}$ in $\mathcal{G}$. If not, then $\mathcal{G}$ would again imply some conditional independence $V_i \perp\!\!\!\perp_P W_j \,|\, \mathbf{Z}_{V_i \| W_j} \cup \{X, Y\}$, which does cannot hold in $\mathcal{M}$. (Note that X and Y may not necessarily be *needed* in the separating set between V_i and W_j, but they never unblock a path to $\mathbf{W}$ (as noncolliders with outgoing arcs to $\mathbf{W}$).

Finally, using Lemma 8(2) above, we show that *at least* either all $\mathbf{V}$ are connected to X in $\mathcal{G}$, or all $\mathbf{V}$ are connected to Y (and probably many to both). Suppose the edge $X - Y$ has the form $X \longrightarrow Y$ in $\mathcal{G}$, and assume there is a set $\mathbf{V}_{Q \| Y}$ that separates Q and Y, so there is no edge $Q - Y$ in $\mathcal{G}$ or $\mathcal{M}$. (Note that there can be no node from the colliders in $\mathbf{V} \longrightarrow \mathbf{W} \longleftarrow Y$ in the separating set.) By Lemma 8(2), any set $\mathbf{V}_{Q \| Y}$ that separates Q and Y in $\mathcal{M}$, also makes Q and X

dependent in $\mathcal{M}$, and so also in $\mathcal{G}$. But then X is needed to block the path $Q..X \longrightarrow Y$ in $\mathcal{G}$ given the other separating nodes, and so $X \in \mathbf{V}_{Q\|Y}$. But if X is necesarily included in $\mathbf{V}_{Q\|Y}$, then by Lemma 8(3) Q and Y are conditionally dependent in $\mathcal{M}$ given $\mathbf{V}_{Q\|Y}$, implying an edge $Q - Y$ in $\mathcal{G}$. So, if $X \longrightarrow Y$ in $\mathcal{G}$, then all $\mathbf{V}$ are adjacent to Y in $\mathcal{G}$ as well. Similarly, if $X \longleftarrow Y$ in $\mathcal{G}$, then all $\mathbf{V}$ are adjacent to X. Therefore, if there is an edge $X - Y$ in $\mathcal{G}$, then either all nodes $\mathbf{V}$ are adjacent to X, or all nodes $\mathbf{V}$ are adjacent to Y.

We now show that for each node $V_i \in \mathbf{V}$ this link must be an arc out of V_i. Assume $X \longrightarrow Y$ in $\mathcal{G}$, so V_i is adjacent to Y, then:

- if $V_i \longrightarrow X$ in $\mathcal{G}$, then acyclicity requires $V_i \longrightarrow Y$,
- if $V_i \longleftarrow X$ in $\mathcal{G}$, then this also requires $V_i \longrightarrow Y$, otherwise $V_i \in \mathbf{W}$,
- if there is no link $V_i - X$ in $\mathcal{G}$, then, like before, any set that separates them in $\mathcal{M}$ unblocks a path to Y, and so including Y makes them dependent. Therefore Y is not in a separating set between V_i and X in $\mathcal{G}$, and so it has to be a collider between them, which implies again $V_i \longrightarrow Y$.

Similar for $V_i \longrightarrow X$ in case of $X \longleftarrow Y$.

So if there is an arc $X \longrightarrow Y$ in $\mathcal{G}$, then all nodes $\mathbf{V}$ have arcs into Y, making $\mathbf{W} \cup Y$ a fully connected component with the same parents $\mathbf{V} \cup X$. As a result, the edge $X \longrightarrow Y$ can be eliminated by putting Y last in the queue. The resulting DAG $\mathcal{G}'$ is equivalent with $\mathcal{G}$ but requires fewer parameters and would therefore be returned instead. Therefore, if $\mathcal{G}$ is a optimal DAG approximation to $\mathcal{M}$, then $\mathcal{G}$ implies $X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$. $\square$

So all (minimal) conditional independencies that are present in $\mathcal{M}$ can be inferred from a optimal uDAG $\mathcal{G}$ over an appropriate subset of nodes. To illustrate this principle, consider the minimal independence $X \perp\!\!\!\perp_M Y \,|\, [\{V, W, Z, Q\}]$, with node Q not adjacent to either X or Y, in the MAG in Figure 2(a). Depending on the multiplicity of the nodes, either (d) or (e) will be found as the optimal uDAG in the large limit. Both correctly exhibit the independence $X \perp\!\!\!\perp_P Y \,|\, \{V, W, Z, Q\}$, even though node Q is now adjacent to X or Y. However, only (d) allows to infer directly from the model that this independence is also minimal; for (e) this follows only *indirectly* from the absence of the independence in smaller models, as node W blocks both $X \longleftarrow W \longrightarrow Y$ and $X \longleftarrow W \longrightarrow Z \longleftarrow Y$.

Corollary 12. For a MAG $\mathcal{M}$ over variables $\mathbf{V}$ that is faithful to a distribution $p(\mathbf{V})$, all (minimal) conditional independencies needed to obtain the skeleton of $\mathcal{M}$ can be found using uDAGs over subsets of variables up to the minimum size of the largest required minimal conditional independence.

Proof. From Lemma 11 we know that any $X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$ will show up in a uDAG over $\{X, Y\} \cup \mathbf{Z}$. So if there is a constant K such that for all separable pairs of nodes $\{X, Y\}$ there is a set $\mathbf{Z}$ such that $X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$, with size bounded by $|\{X, Y\} \cup \mathbf{Z}| \leq K$, then we are guaranteed to find all of these in one or more uDAGs up to size K. From this we can reconstruct the entire skeleton of $\mathcal{M}$. $\square$

3.2 Minimal dependencies

Once we find a (minimal) independence, we can also find the nodes that destroy it:

Lemma 13. Let $\mathcal{M}$ be a faithful MAG over nodes $\{X, Y, W\} \cup \mathbf{Z}$, for which $X \perp\!\!\!\perp_M Y \,|\, [\mathbf{Z}]$ and $X \not\perp\!\!\!\perp_M Y \,|\, \mathbf{Z} \cup [W]$. Then, if $\mathcal{G}$ is an optimal uDAG approximation to $\mathcal{M}$, then $X \not\perp\!\!\!\perp_P Y \,|\, \mathbf{Z} \cup [W]$ can be inferred from $\mathcal{G}$, provided it is (already) known that $X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$.

Proof. Invariant bi-directed edges in the equivalence class of a MAG can only be eliminated by adding one or more edges between unshielded colliders at the bi-directed edge. An optimal uDAG only adds such edges until all invariant bi-directed edges are eliminated. Therefore, if conditioning on W does not make X and Y dependent given $\mathbf{Z}$, then X and Y are not an inducing pair for any invariant arrowhead in $\mathcal{M}$, and so adding an edge $X - Y$ is always superfluous and always results in a uDAG requiring more parameters than the smallest one without that edge. Therefore, if we know that $X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$, and X and Y are connected in $\mathcal{G}$, then W must be the node that makes them dependent, and so $W \not\Rightarrow (X, Y, \mathbf{Z})$. If X and Y are not connected in $\mathcal{G}$, then all nodes $\mathbf{Z}$ are again needed to separate them, and then including W must make X and Y dependent, otherwise it would imply an invalid independence, which in a uDAG means a path $X \longrightarrow W \longleftarrow Y$. If X and Y are not connected in $\mathcal{G}$ and conditioning on W leaves them independent, then it automatically implies that W also does not make them dependent in $\mathcal{M}$. Remains to be shown that if X and Y are not connected in $\mathcal{G}$, then W cannot be a collider between them in a optimal uDAG if W does not make them dependent in $\mathcal{M}$. If there are edges or unblocked paths $X - W$ and $W - Y$ in the (underlying) MAG $\mathcal{M}$ over nodes $(X, Y, W, \mathbf{Z})$ conditional on all $\mathbf{Z}$, then without W this would imply an unblocked path between X and Y given $\mathbf{Z}$ (via W) in $\mathcal{M}$ *unless* W is a collider on this path. But then conditioning on W would make X and Y dependent, contrary the assumption that *not* $X \not\perp\!\!\!\perp_P Y \,|\, \mathbf{Z} \cup [W]$. Therefore, if W is the 'last' node (no descendants) in the optimal uDAG $\mathcal{G}$, then if W does not destroy the minimal independence $X \perp\!\!\!\perp_P Y \,|\, \mathbf{Z}$ then at least one of the arcs $X \longrightarrow W$ or $W \longleftarrow Y$ is redundant and will not present in $\mathcal{G}$. It follows that if the subpath $X \longrightarrow W \longleftarrow Y$ is present in $\mathcal{G}$, then W is necessarily a node that makes them dependent in $\mathcal{M}$ given $\mathbf{Z}$, and so correctly implies $X \not\perp\!\!\!\perp_P Y \,|\, \mathbf{Z} \cup [W]$, and therefore also $W \not\Rightarrow (X, Y, \mathbf{Z})$. □

So, even though we can find the nodes that create a dependency, it is not guaranteed to be from a single model (as for minimal independence). But if the corresponding independence is *known* to exist, then it can be identified from that model, either directly or indirectly.

An example is shown in Figure 3, where the MAG in (a) contains the (minimal) conditional independence $X \perp\!\!\!\perp_P Y \,|\, [Z]$, together with $X \not\perp\!\!\!\perp_P Y \,|\, Z \cup [W]$. Both (b) and (c) represent (potentially) optimal uDAGs to (a): which one is chosen depends on the multiplicity of the nodes. But where (b) shows both $X \perp\!\!\!\perp_P Y \,|\, Z$ and $X \not\perp\!\!\!\perp_P Y \,|\, ZW$, (c) only allows the positive identification of $X \not\perp\!\!\!\perp_P Y \,|\, ZW$, as the conditional independence is no longer present. Still, if we learned from the model over (X, Y, Z) that $X \perp\!\!\!\perp_P Y \,|\, Z$, then it is possible to infer that W is the node that destroys that independence, and so cannot be an ancestor of any.

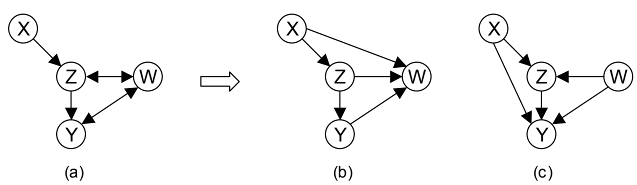

Figure 3: (a) MAG with two bi-directed arcs, (b) optimal uDAG with $X \longrightarrow W$, (c) optimal uDAG with $X \longrightarrow Y$

If we can find all required (minimal) independencies for the skeleton, and all nodes that create a subsequent dependency, then we have enough to identify all invariant arrowheads:

Corollary 14. For a MAG $\mathcal{M}$ over variables $\mathbf{V}$ that is faithful to a distribution $p(\mathbf{V})$, all conditional dependencies needed to obtain all invariant arrowheads in the skeleton of $\mathcal{M}$ can be found using uDAGs over subsets of variables up to the minimum size of the largest required conditional dependence.

Proof. From Lemma 13 we know that all nodes W that break some independence $X \not\perp\!\!\!\perp_P Y \,|\, \mathbf{Z} \cup [W]$ can be inferred from a uDAG over $\{X, Y, W\} \cup \mathbf{Z}$, provided we already know that $X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$. From Lemma 11 we know that all such $X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$ show up in a uDAG over $\{X, Y\} \cup \mathbf{Z}$. So if there is a constant K such that for all separable pairs of nodes $, Y\}$ there is a set $\mathbf{Z}$ such that $X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$, with size bounded by $|\{X, Y\} \cup \mathbf{Z}| \leq K$, then we are guaranteed to find all subsequent dependencies in one or more uDAGs up to size $K + 1$. Combining these is enough to cover all instances of cases(1) and (2) in Claassen and Heskes (2011)), which are sufficient to find all invariant arrowheads. □

The main result of this section can now be stated as:

Theorem 1. Let $\mathcal{M}$ be a faithful MAG over variables $\mathbf{V}$. Let K be a bound on the size of the separating sets, such that $\forall \{X, Y\} \in \mathbf{V}, X \notin Adj_M(Y), \exists \mathbf{Z}, |\mathbf{Z}| \leq K : X \perp\!\!\!\perp_P Y \,|\, [\mathbf{Z}]$. Then in the large sample limit the entire skeleton and all invariant arrowheads of $\mathcal{M}$ can be obtained from optimal uDAGs over subsets size $|\mathbf{X}| \leq K + 3$.

Proof. With causal statements obtained from the corresponding MAGs over these subsets of nodes, the LoCI algorithm is already known to be sound and complete in the large sample limit. This theorem states that, provided the possible loss of faithfulness is properly accounted for, uDAGs can also provide enough information to find at least the entire skeleton and invariant arrowheads. This follows immediately from corollaries 12 and 14. □

References

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

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

D. Colombo, M. Maathuis, M. Kalisch, and T. Richardson. Learning high-dimensional dags with latent and selection variables (uai2011). Technical report, ArXiv, Zurich, 2011.

C. Meek. Causal inference and causal explanation with background knowledge. In *UAI*, pages 403–410. Morgan Kaufmann, 1995.

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.

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*, pages 211–252. 1999.

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

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.