Supplementary material for "Causal Learning for Partially Observed Stochastic Dynamical Systems"

Søren Wengel Mogensen
Department of Mathematical Sciences
University of Copenhagen
Copenhagen, Denmark

Daniel Malinsky
Department of Computer Science
Johns Hopkins University
Baltimore, MD, USA

Niels Richard Hansen
Department of Mathematical Sciences
University of Copenhagen
Copenhagen, Denmark

This supplementary material contains proofs that were omitted in the paper. It also contains the potential parent and potential sibling criteria and reports the results of a small simulation study illustrating the cost and the impact of the potential step in the learning algorithm.

A PROOFS OF LEMMAS 5 AND 6

Lemma 5. The independence model $\mathcal{I}(\mathcal{G})$ satisfies left and right {decomposition, weak union, composition} and left {redundancy, intersection, weak composition}. Furthermore, $\langle A, B \mid C\rangle \in \mathcal{I}(\mathcal{G})$ whenever $B = \emptyset$.

Proof. Left redundancy, left and right decomposition and left and right composition follow directly from the definition of μ-separation. Left and right weak union are also immediate. Left weak composition follows from left redundancy, left decomposition and left composition. It is also clear that $\langle A, B \mid C\rangle \in \mathcal{I}(\mathcal{G})$ if $B = \emptyset$.

For left intersection, consider a μ-connecting walk, $\omega = \langle \nu_1, e_1, \ldots, e_n, \nu_{n+1}\rangle$ from $\delta = \nu_1 \in A \cup C$ to $\beta = \nu_{n+1} \in B$ given $A \cap C$. This walk is by definition non-trivial. Consider now the shortest possible non-trivial subwalk of ω of the form $\tilde{\omega} = \langle \nu_i, e_i, \ldots, e_n, \nu_{n+1}\rangle$ such that $\nu_i \in (A \cup C) \setminus (A \cap C)$. Such a subwalk always exists and it is μ-connecting either from A to B given C or from C to B given A. □

Lemma 6. $\mathcal{I}(\mathcal{G})$ satisfies cancellation.

Proof. The contrapositive of $A \perp_\mu B \mid C \cup \{\delta\} \Rightarrow A \perp_\mu B \mid C$ is $A \not\perp_\mu B \mid C \Rightarrow A \not\perp_\mu B \mid C \cup \{\delta\}$. So we have that $A \perp_\mu C_1 \cup \{\delta\} \mid C \cup \{\delta\}$, $\delta \perp_\mu C_2 \cup A \mid C \cup A$, and $A \not\perp_\mu B \mid C$ and want to show that $A \not\perp_\mu B \mid C \cup \{\delta\}$. Note that $A \perp_\mu \delta \mid C \cup \{\delta\}$ by right decomposition.

There exists a μ-connecting walk ω from $\alpha \in A$ to some $\beta \in B$ given C, and we argue that this walk is also μ-connecting given $C \cup \{\delta\}$. Suppose not, for contradiction. Note that $\alpha \notin C$ so $\alpha \notin C \cup \{\delta\}$ since by factorization $A, C, \{\delta\}$ are disjoint. Also every collider on ω is in $\mathrm{an}(C)$ so it is in $\mathrm{an}(C \cup \{\delta\})$. Thus if ω is not μ-connecting given $C \cup \{\delta\}$ it must be because there is some non-collider on ω which is not in C but is in $C \cup \{\delta\}$, i.e., the non-collider is δ. Choose now a subwalk of ω between some (possibly different) $\alpha \in A$ and δ such that no non-endpoint node of this subwalk is in $A \cup \{\delta\}$. Again, $\alpha \notin C \cup \{\delta\}$. Such a subwalk always exists.

There are two possibilities: either there is an arrowhead into δ on this subwalk of ω or there is not. In the first case, the subwalk of ω from α into δ is μ-connecting given $C \cup \{\delta\}$, i.e., $A \not\perp_\mu \delta \mid C \cup \{\delta\}$. Contradiction. In the second case, we consider a collider ε on the subwalk between α and δ (if there is no collider on the walk, then the directed walk from δ to α is μ-connecting given $C \cup A$). Either $\varepsilon \in C_1$, $\varepsilon \in C_2$, or there is a (non-trivial) directed walk from ε to some ε' that is either in C_1 or C_2. If $\varepsilon \in C_1$, there is a μ-connecting subwalk of ω from α to $\varepsilon \in C_1$ given C. Since there are no non-colliders on this walk in $\{\delta\}$, it is also μ-connecting given $C \cup \{\delta\}$. If $\varepsilon \in C_2$, likewise there is a μ-connecting walk from δ to C_2 given $C \cup A$ (note that there are no non-colliders in A on this walk by choice of α). Either way, contradiction.

If $\varepsilon \notin C$, we consider concatenating one of the aforementioned walks to ε with the directed path ω' from ε to $\varepsilon' \in C$. Either δ appears on ω' or it does not. In the first case, then there is an arrowhead at δ on ω' and so $A \not\perp_\mu \delta \mid C \cup \{\delta\}$ as before. In the latter case, there are two subcases to consider: either there is some vertex in A on ω' or there is not. If there is, choose $\alpha' \in A$ on ω' such that there are no vertices in A nearer to ε on ω'. Then the the walk from δ to α' is μ-connecting given $C \cup A$. If there is no vertex in A on ω', then by concatenating a subwalk of ω to ω' we get a μ-connecting walk from α or δ to ε' in C_1 or C_2 given $C \cup \{\delta\}$ or $C \cup A$, respectively. In any case, contradiction. □

B PROOF OF THEOREM 8

In this section, we first prove some lemmas and then use these to prove Theorem 8.

Lemma 19. If $A \perp_\mu B \mid C$ and $A \perp_\mu D \mid C$, then $A \perp_\mu B \mid C \cup D$.

Proof. This follows from right composition, right weak union, and right decomposition of μ-separation. □

Lemma 20. Assume $\gamma \in \mathrm{an}(A \cup B \cup C)$ and $\alpha, \gamma \notin C$. If there is a walk between $\alpha \in A$ and γ such that no non-collider is in C and every collider is in $\mathrm{an}(C)$, and there is a μ-connecting walk from γ to $\beta \in B$ given C, then there is a μ-connecting walk from A to B given C.

If $\omega = \langle \nu_1, e_1, \nu_2, \dots, e_n, \nu_{n+1} \rangle$ is a walk, then the inverse, ω^{-1}, is the walk $\langle \nu_{n+1}, e_n, \nu_n, \dots, e_1, \nu_1 \rangle$.

Proof. If $\gamma \in \mathrm{an}(C)$, then simply compose the walks. Assume $\gamma \notin \mathrm{an}(C)$. If $\gamma \in \mathrm{an}(A)$ let π denote the directed path from γ to $\bar{\alpha} \in A$. We have that there is no node in C on π and composing π^{-1} with the μ-connecting walk from γ to B gives a μ-connecting walk from $\bar{\alpha} \in A$ to $\beta \in B$ given C. If $\gamma \in \mathrm{an}(B)$ compose the walk from α to γ with the directed path from γ to B (which is μ-connecting given C as $\gamma \notin \mathrm{an}(C)$). □

Lemma 21. Assume that $\mathcal{I}$ satisfies left weak composition, left intersection, and left decomposition. If $A \cap D = \emptyset$ then

$$\langle A, B \mid C \cup D \rangle \in \mathcal{I}, \langle D, B \mid C \cup A \rangle \in \mathcal{I} \Rightarrow \langle A \cup D, B \mid C \rangle \in \mathcal{I}.$$

Proof. By left weak composition $\langle A \cup C, B \mid C \cup D \rangle \in \mathcal{I}, \langle D \cup C, B \mid C \cup A \rangle \in \mathcal{I}$. It follows by left intersection that $\langle A \cup C \cup D, B \mid C \rangle \in \mathcal{I}$ and by left decomposition the result follows. □

Lemma 22. Let $\mathcal{D} = (V, E)$ be a DG, and let $\alpha, \beta \in V$. Then $\alpha \notin \mathrm{pa}_{\mathcal{D}}(\beta)$ if and only if $\alpha \perp_\mu \beta \mid V \setminus \{\alpha\}$.

In the following proofs, we will use $\sim$ to denote an arbitrary edge.

Proof. Assume first that $\alpha \notin \mathrm{pa}_{\mathcal{D}}(\beta)$, and consider a walk between α and β that has a head at β, $\alpha \sim \dots \sim \gamma \to \beta$. We must have that $\alpha \neq \gamma$ and therefore the walk is not μ-connecting given $V \setminus \{\alpha\}$.

Assume instead that $\alpha \perp_\mu \beta \mid V \setminus \{\alpha\}$. The edge $\alpha \to \beta$ would constitute a μ-connecting walk given $V \setminus \{\alpha\}$ and therefore we must have that $\alpha \notin \mathrm{pa}_{\mathcal{D}}(\beta)$. □

Theorem 8. Assume that $\mathcal{I}$ is an independence model that satisfies left {redundancy, intersection, decomposition, weak union, weak composition}, right {decomposition, composition}, is cancellative, and furthermore $\langle A, B \mid C \rangle \in \mathcal{I}$ whenever $B = \emptyset$. Let $\mathcal{D}$ be a DG. Then $\mathcal{I}$ satisfies the pairwise Markov property with respect to $\mathcal{D}$ if and only if it satisfies the global Markov property with respect to $\mathcal{D}$.

Proof. It follows directly from the definitions and Lemma 22 that the global Markov property implies the pairwise Markov property. Assume that $\mathcal{I}$ satisfies the pairwise Markov property w.r.t. $\mathcal{D}$ and let $A, B, C \subseteq V$. Assume $A \perp_\mu B \mid C$. We wish to show that $\langle A, B \mid C \rangle \in \mathcal{I}$.

Assume $|V| = n > 0$. We will proceed using reverse induction on $|C|$. As the induction base, $C = V$. The result follows by noting that $\langle V, B \mid V \rangle \in \mathcal{I}$ by left redundancy of $\mathcal{I}$. By left decomposition of $\mathcal{I}$, we get $\langle A, B \mid V \rangle \in \mathcal{I}$.

For the induction step, consider a node $\gamma \notin C$. Note first that if $A \subseteq C$, then the result once again follows using left redundancy and then left decomposition, and therefore assume that $A \setminus C \neq \emptyset$, and take $\alpha \in A \setminus C$ (note that $\alpha = \gamma$ is allowed). Assume first that we cannot choose α and γ such that $\alpha \neq \gamma$. This means that $C = V \setminus \{\alpha\}$. By right decomposition of $\mathcal{I}(\mathcal{G})$ we have that $A \perp_\mu \beta \mid C$ for all $\beta \in B$, and by left decomposition of $\mathcal{I}(\mathcal{G})$ we have $\alpha \perp_\mu \beta \mid C$. If $B = \emptyset$, then the result follows by assumption, and else by the pairwise Markov property and Lemma 22 we have $\langle \alpha, \beta \mid C \rangle \in \mathcal{I}$ for all $\beta \in B$ and by right composition of $\mathcal{I}$ we have $\langle \alpha, B \mid C \rangle \in \mathcal{I}$. By left weak composition, we have $\langle A, B \mid C \rangle \in \mathcal{I}$.

Now assume $\gamma \neq \alpha$. We split the proof into two cases, (i) and (ii), depending on whether or not we can choose γ as an ancestor to $A \cup B \cup C$.

Case (i): $\gamma \in \mathrm{an}(A \cup B \cup C)$
We have that $\gamma \perp_\mu B \mid C$ or $A \perp_\mu \gamma \mid C$ by Lemma 20. We split into two subcases, (i-1) and (i-2).

Case (i-1): $\gamma \perp_\mu B \mid C$
By left composition of $\mathcal{I}(\mathcal{G})$, $A \cup \{\gamma\} \perp_\mu B \mid C$ and by left weak union $A \cup \{\gamma\} \perp_\mu B \mid C \cup \{\gamma\}$ as well as $A \cup \{\gamma\} \perp_\mu B \mid C \cup (A \setminus \{\gamma\})$. By the induction hypothesis and noting that $C \cup \{\gamma\} \neq C \neq C \cup (A \setminus \{\gamma\})$, $\langle A \cup \{\gamma\}, B \mid C \cup \{\gamma\} \rangle \in \mathcal{I}$, and $\langle A \cup \{\gamma\}, B \mid C \cup (A \setminus \{\gamma\}) \rangle \in \mathcal{I}$. By left decomposition of $\mathcal{I}$ and Lemma 21, the result follows.

Case (i-2): $A \perp_\mu \gamma \mid C$
In this case, we can assume that $\gamma \notin A$, as otherwise by left decomposition of $\mathcal{I}(\mathcal{G})$ we would also have $\gamma \perp_\mu$

$B \mid C$ which is case (i-1). Moreover, either $\gamma \perp_\mu B \mid C$ or $\gamma \perp_\mu A \setminus C \mid C$, as otherwise $A \perp_\mu B \mid C$ would not hold (Lemma 20). $\gamma \perp_\mu B \mid C$ is the above case, so assume that $\gamma \not\perp_\mu B \mid C$ and $\gamma \perp_\mu A \setminus C \mid C$. Using right weak union of $\mathcal{I}(\mathcal{G})$, we have $A \perp_\mu \gamma \mid C \cup \{\gamma\}$ and $\gamma \perp_\mu A \setminus C \mid C \cup A$. Using the induction assumption, we have that $\langle A, \gamma \mid C \cup \{\gamma\}\rangle \in \mathcal{I}$ and $\langle \gamma, A \setminus C \mid C \cup A\rangle \in \mathcal{I}$. We have $A \perp_\mu B \mid C$ and $A \perp_\mu \gamma \mid C$ and using right composition and right weak union of $\mathcal{I}(\mathcal{G})$, we obtain $A \perp_\mu B \cup \{\gamma\} \mid C \cup \{\gamma\}$. Using the induction assumption we have that $\langle A, B \mid C \cup \{\gamma\}\rangle \in \mathcal{I}$. Assume to obtain a contradiction that $A \not\perp_\mu \delta \mid C \cup \gamma$ and $\gamma \not\perp_\mu \delta \mid C \cup A$ for some $\delta \in C$. We know that $A \perp_\mu \gamma \mid C$ and by using the contrapositive of Lemma 19 this means that $A \not\perp_\mu \delta \mid C$. Similarly, we obtain that $\gamma \not\perp_\mu \delta \mid C$. We note that $\gamma \not\perp_\mu B \mid C$ and by Lemma 20 this means that $A \not\perp_\mu B \mid C$ which is a contradiction. Therefore, we have that for each $\delta \in C$, either $A \perp_\mu \delta \mid C \cup \gamma$ (and therefore also $A \setminus C \perp_\mu \delta \mid C \cup \gamma$) or $\gamma \perp_\mu \delta \mid C \cup A$. Using the induction assumption, right composition of $\mathcal{I}$, the cancellation property and left weak composition of $\mathcal{I}$ we arrive at the conclusion.

Case (ii): If one cannot choose a $\gamma \in \text{an}(A \cup B \cup C)$ such that $\gamma \notin C$ and $\gamma \neq \alpha$, then $\text{an}(A \cup B \cup C) = C \cup \{\alpha\}$. Assume this and furthermore assume that $\gamma \notin \text{an}(A \cup B \cup C)$. We will first argue that $A \perp_\mu B \mid C \cup \{\gamma\}$. If this was not the case there would be a μ-connecting walk, ω, from A to $\beta \in B$ given $C \cup \{\gamma\}$ on which γ was a collider and furthermore every collider was in $C \cup \{\gamma\}$. Consider now the last occurrence of γ on this walk, and the subwalk of ω, $\gamma \sim \ldots \sim \theta \sim \ldots \to \beta$. Let θ be the node in $\text{an}(A \cup B \cup C)$ which is the closest to γ on the walk. Then there must be a tail at θ, and this means that $\theta = \alpha$ as otherwise the walk would be closed. In this case, the subwalk from α to β would also be μ-connecting given C which is a contradiction.

It also holds that $\gamma \perp_\mu B \mid C \cup A$ as every parent of a node in B is in $C \cup A$. Using the induction assumption we have that $\langle A, B \mid C \cup \{\gamma\}\rangle \in \mathcal{I}$ and $\langle \gamma, B \mid C \cup A\rangle \in \mathcal{I}$ and using Lemma 21 and left decomposition of $\mathcal{I}$ we obtain $\langle A, B \mid C\rangle \in \mathcal{I}$.

□

C PROOF OF LEMMA 11

Lemma 11. Let $\mathcal{I}$ be a local independence model. Then it satisfies left {redundancy, decomposition, weak union, weak composition} and right {decomposition, composition} and furthermore $\langle A, B \mid C\rangle \in \mathcal{I}$ whenever $B = \emptyset$. If $\mathcal{F}_t^A \cap \mathcal{F}_t^C = \mathcal{F}_t^{A\cap C}$ holds for all $A, C \subseteq V$ and $t \in [0, T]$, then left intersection holds.

Proof. Left redundancy: We note that $\mathcal{F}_t^{A\cup C} = \mathcal{F}_t^C$ from which the result follows.

Left decomposition: Assume that $A_1 \cup A_2 \not\to_\lambda B \mid C$. We wish to show that $A_1 \not\to_\lambda B \mid C$.

$$
\begin{aligned}
E(\lambda_t^\beta \mid \mathcal{F}_t^{A_1\cup C}) &= E\big(\underbrace{E(\lambda_t^\beta \mid \mathcal{F}_t^{A_1\cup A_2\cup C})}_{=E(\lambda_t^\beta \mid \mathcal{F}_t^C)} \mid \mathcal{F}_t^{A_1\cup C}\big) \\
&= E(\lambda_t^\beta \mid \mathcal{F}_t^C)
\end{aligned}
$$

Left weak union: Simply note that the conditioning σ-algebra stays the same in the conditional expectation which is assumed to be $\mathcal{F}_t^C$-adapted and therefore also $\mathcal{F}_t^{C\cup D}$-adapted.

Left weak composition: The conditioning σ-algebra again stays the same in the conditional expectation.

Right decomposition and right composition follow directly from the coordinate-wise definition of local independence.

Left intersection: We note that $E(\lambda_t^\beta \mid \mathcal{F}_t^{A\cup C})$ by assumption has an $\mathcal{F}_t^A$-adapted and an $\mathcal{F}_t^C$-adapted version, thus it has a version, which is adapted w.r.t. the filtration $\mathcal{F}_t^A \cap \mathcal{F}_t^C = \mathcal{F}_t^{A\cap C}$.

Finally, it is clear that $\langle A, B \mid C\rangle \in \mathcal{I}$ if $B = \emptyset$ as this makes the condition void. □

D PROOFS, SECTION 5

Lemma 16. Subalgorithm 1 outputs the separability graph of $\mathcal{I}$, $\mathcal{S}$, and furthermore $\mathcal{N} \subseteq \mathcal{S}$.

Proof. In Subalgorithm 1, we only remove edges $\alpha \ast\!\!\to \beta$ when we have found a set $C \subseteq V \setminus \{\alpha\}$ that separates β from α. The DMGs $\mathcal{G}_0$ and $\mathcal{N}$ are Markov equivalent and therefore the same separation holds in $\mathcal{I}(\mathcal{N})$. Such an edge would always be μ-connecting from α to β given C as $\alpha \notin C$ and therefore we know it to be absent in $\mathcal{N}$. This means that the output of the algorithm is a supergraph of $\mathcal{N}$.

The graph $\mathcal{G}$ in Subalgorithm 1 is always a supergraph of $\mathcal{G}_0$ and therefore $D_{\mathcal{G}_0}(\alpha, \beta) \subseteq D_{\mathcal{G}}(\alpha, \beta)$. If there exists a set that separates β from α then $D_{\mathcal{G}_0}(\alpha, \beta)$ does and by the above inclusion we are always sure to test this set. This means that the output is the separability graph. □

Lemma 17. Subalgorithm 2 outputs a supergraph of $\mathcal{N}$.

Proof. By Lemma 16, $\mathcal{N} \subseteq \mathcal{S}$. We also know that if there is an edge $\alpha \to \beta$ in $\mathcal{S}$ then $\alpha \in u(\beta, \mathcal{I}(\mathcal{G}_0)) =$

$u(\beta, \mathcal{I}(\mathcal{N})) = u(\beta, \mathcal{I})$. Assume there is an unshielded W-structure ${}_w(\alpha, \beta, \gamma)$ in $\mathcal{S}$. The edge between α and β in $\mathcal{S}$ means that β cannot be separated from α in $\mathcal{I}(\mathcal{N})$ and therefore there exists for every $C \subseteq V \setminus \{\alpha\}$ a μ-connecting walk from α to β given C. By definition of μ-connecting walks this has a head at (the final) β. The W-structure is unshielded, that is, $\alpha \to \gamma$ is not in $\mathcal{S}$. This means that we have previously found a separating set $S_{\alpha,\gamma}$, such that $\langle \alpha, \gamma \mid S_{\alpha,\beta} \rangle \in \mathcal{I}(\mathcal{N})$ and $\alpha \notin S_{\alpha,\gamma}$. We know that there exists a μ-connecting walk ω, from α to β given $S_{\alpha,\gamma}$ in $\mathcal{N}$ as $\alpha \in u(\beta, \mathcal{I}(\mathcal{N}))$. If $\beta \notin S_{\alpha,\gamma}$ then we can compose ω with the edge $\beta \to \gamma$ which gives a μ-connecting walk from α to γ given $S_{\alpha,\gamma}$ which is a contradiction, and therefore the edge $\beta \to \gamma$ cannot be in $\mathcal{N}$. If $\beta \in S_{\alpha,\gamma}$ then we can argue analogously and obtain that $\beta \leftrightarrow \gamma$ cannot be in $\mathcal{N}$. □

Theorem 18. The algorithm defined by first doing the separation step, then the pruning, and finally the potential step outputs $\mathcal{N}$, the maximal element of $[\mathcal{G}_0]$.

Proof. By Lemma 17, the output after the first two steps is a supergraph of $\mathcal{N}$. In the potential step, an edge $\alpha \to \beta$ is only removed if α is not a potential parent of β in $\mathcal{I}$. We know that if the edge is in $\mathcal{N}$ then α is a potential parent of β in $\mathcal{I}(\mathcal{N}) = \mathcal{I}(\mathcal{G}_0) = \mathcal{I}$ (Mogensen and Hansen, 2018) and by contraposition of this result it follows that every directed edge removed is not in $\mathcal{N}$. The same argument applies in the case of a bidirected edge and therefore the output is a supergraph of $\mathcal{N}$.

If we consider some edge $\alpha \overset{e}{\to} \beta$ in the output graph, then either α is a potential parent of β, in which case e is also in $\mathcal{N}$, or $\mathcal{I}(\mathcal{G} - e) \cap \mathcal{L}_n \neq \emptyset$. Assume the latter. We have that $\mathcal{G}_0 \subseteq \mathcal{G}$, and therefore $\mathcal{I}(\mathcal{G} - e) \subseteq \mathcal{I}(\mathcal{G}_0)$ if e is not in $\mathcal{G}_0$. The above intersection is non-empty and therefore there is some triple which is in both $\mathcal{I}(\mathcal{G} - e)$ and $\mathcal{L}_n$, and by $\mathcal{I}(\mathcal{G} - e) \subseteq \mathcal{I}(\mathcal{G}_0)$ it is also in $\mathcal{I}(\mathcal{G}_0)$. But by definition $\mathcal{L}_n$ contains only triples not in $\mathcal{I}(\mathcal{G}_0)$, so this is a contradiction. Therefore, e must be in $\mathcal{G}_0$ and also in $\mathcal{N}$ as $\mathcal{G}_0 \subseteq \mathcal{N}$. One can argue analogously for the bidirected edges. We conclude that the output graph is equal to $\mathcal{N}$, the maximal element of $[\mathcal{G}_0]$. □

E POTENTIAL PARENT/SIBLINGS

Consider an independence model, $\mathcal{I}$, over V and let $\alpha, \beta \in V$. The set $u(\beta, \mathcal{I})$ is defined in Subsection 5.1.1. As described in Subsection 5.1 the below definitions define a list of independence tests which one can conduct to directly construct $\mathcal{N}$. This was proven by Mogensen and Hansen (2018). However, the list is very large and one can construct $\mathcal{N}$ in a more efficient manner. If e.g. $|V| = 10$, then for each choice of γ in (s2) we can choose C in 2^8 different ways (omitting sets C containing γ as such an independence would hold trivially for any independence model satisfying left redundancy and left decomposition).

Definition 23. We say that α and β are *potential siblings* in the independence model $\mathcal{I}$ if (s1)-(s3) hold:

(s1) $\beta \in u(\alpha, \mathcal{I})$ and $\alpha \in u(\beta, \mathcal{I})$,

(s2) for all $\gamma \in V$, $C \subseteq V$ such that $\beta \in C$,

$$\langle \gamma, \alpha \mid C \rangle \in \mathcal{I} \Rightarrow \langle \gamma, \beta \mid C \rangle \in \mathcal{I},$$

(s3) for all $\gamma \in V$, $C \subseteq V$ such that $\alpha \in C$,

$$\langle \gamma, \beta \mid C \rangle \in \mathcal{I} \Rightarrow \langle \gamma, \alpha \mid C \rangle \in \mathcal{I}.$$

Definition 24. We say that α is a *potential parent* of β in the independence model $\mathcal{I}$ if (p1)-(p4) hold:

(p1) $\alpha \in u(\beta, \mathcal{I})$,

(p2) for all $\gamma \in V$, $C \subseteq V$ such that $\alpha \notin C$,

$$\langle \gamma, \beta \mid C \rangle \Rightarrow \langle \gamma, \alpha \mid C \rangle,$$

(p3) for all $\gamma, \delta \in V$, $C \subseteq V$ such that $\alpha \notin C, \beta \in C$,

$$\langle \gamma, \delta \mid C \rangle \Rightarrow \langle \gamma, \beta \mid C \rangle \vee \langle \alpha, \delta \mid C \rangle,$$

(p4) for all $\gamma \in V, C \subseteq V$, such that $\alpha \notin C$,

$$\langle \beta, \gamma \mid C \rangle \Rightarrow \langle \beta, \gamma \mid C \cup \{\alpha\} \rangle.$$

F SIMULATION STUDY

We conducted a small simulation study to empirically evaluate the cost and impact of the third step in the learning algorithm, the potential step. This step is computationally expensive as it involves testing the potential parent/siblings conditions, see above.

We simulated a random DMG on 5 nodes by first drawing p_d from a uniform distribution on $[0, 1/2]$ and p_b from a uniform distribution on $[0, 1/4]$. We then generated independent Bernoulli random variates, $\{b_{\langle \alpha,\beta \rangle}\}$, each with success parameter p_d, and one for each ordered pair of nodes, $\langle \alpha, \beta \rangle$. The edge $\alpha \to \beta$ was included if $b_{\langle \alpha,\beta \rangle} = 1$. For each unordered pair of nodes, $\{\alpha, \beta\}$, we did analogously, using p_b as success parameter. We discarded graphs for which the maximal Markov equivalent graph had more then 15 edges.

Simulating 800 random DMGs, we saw that on average the first step required 90 independence tests and removed

26 edges. The second step removed 1.1 edge on average (it does not use any additional independence tests), while the third required an additional 77 independence tests. On average the third step removed 0.8 edge. This simulation is very limited and simple, however, it does indicate that the potential step of the learning algorithm constitutes a substantial part of the computational cost while not removing a lot of edges.