Appendix: Causal Structure Discovery from Distributions Arising from Mixtures of DAGs

Basil Saeed [1] **Snigdha Panigrahi** [2] **Caroline Uhler** [1 3]

A. Proof of Proposition 2.1

We begin by recalling the definition of an inducing path from Richardson and Spirtes (2002), specialized to ancestral graphs.

Definition A.1. *A path $v_1, \ldots, v_n$ in an ancestral graph $\mathcal{G}$ is* inducing *if v_1 and v_n are not adjacent in $\mathcal{G}$ and for all $i \in \{2, \ldots, n-1\}$, we have*

$$v_{i-1} \leftrightarrow v_i \leftrightarrow v_{i+1} \qquad \text{and} \qquad v_i \in \mathrm{an}_{\mathcal{G}}(\{v_1, v_n\}).$$

Richardson and Spirtes (2002) showed the following condition for an ancestral graph to be maximal.

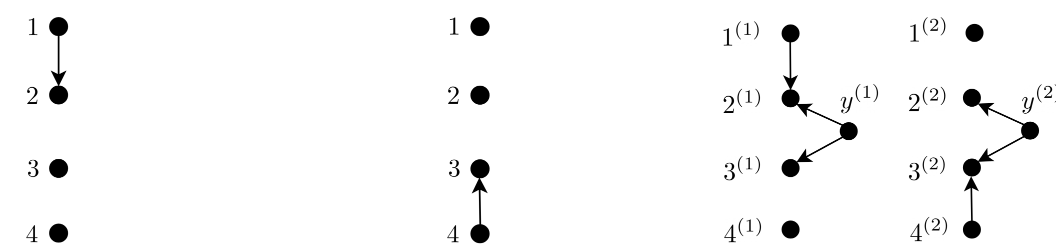
Lemma A.2 ((Richardson and Spirtes, 2002))**.** *An ancestral graph $\mathcal{M}$ is maximal if and only if $\mathcal{G}$ does not contain any inducing paths.*

This allows us to prove Proposition 2.1.

Proof of Proposition 2.1. We show that the graph resulting from Algorithm 1 does not contain inducing paths. Let $\mathcal{M}$ be the output of the algorithm. Suppose we have vertices $v_1, \ldots, v_n$ where $v_{i-1} \leftrightarrow v_i \leftrightarrow v_{i+1}$ for all $i \in \{2, \ldots, n-1\}$ in $\mathcal{M}$. Then by step 1 of the algorithm, we must have $v_1, \ldots, v_n \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$, implying that $v_1 \leftrightarrow_{\mathcal{M}} v_n$, and hence the path is not inducing. □

B. Counter-example for the Markov property of the mother graph

In the following, we provide a counter-example for the Markov property of the mother-graph representation introduced by Strobl (2019b;a). We first remark that the Markov property in Strobl (2019a) generalizes that of Strobl (2019b) in the following sense: if the Markov property of the latter is satisfied, then the former is satisfied. Hence, we here provide a counter-example for the former, which can serve as a counter-example for both.



(a) $\mathcal{D}^{(1)}$ (b) $\mathcal{D}^{(2)}$ (c) $\mathcal{D}_m$

Figure 1: (c) shows the mother graph $\mathcal{D}_m$ associated with the DAGs $\mathcal{D}^{(1)}$ and $\mathcal{D}^{(2)}$ in (a) and (b).

[1]Laboratory for Information and Decision Systems and Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA, USA [2]Department of Statistics, University of Michigan, Ann Arbor, MI, USA [3]Department of Biosystems Science and Engineering, ETH Zurich, Switzerland. Correspondence to: Caroline Uhler <cuhler@mit.edu>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

We start by recalling a few definitions from Strobl (2019b) using notation native to our development. Given a mixture of DAGs with distribution p_μ where $p^{(j)}$ factorizes according to $\mathcal{D}^{(j)}$, the mother graph $\mathcal{D}_m = (V_m, D_m)$ has nodes $V_m := [V] \cup \{y^{(1)}, \ldots, y^{(K)}\}$ and directed edges

$$D_m := \bigcup_{1 \leq j \leq K} \{y^{(j)} \to v^{(j)} : v \in V \setminus V_{\text{INV}}\} \cup \{u^{(j)} \to v^{(j)} : u \to_{\mathcal{D}^{(j)}} v\}.$$

An example of the mother graph is shown in Figure 1. A variable $c^{(j)} \in [V]$ in the mother graph is called an *m-collider* if and only if at least one of the following conditions hold:

- $a^{(j)} \to c^{(j)} \leftarrow b^{(j)}$, where $a, b \in V \cup \{y\}$
- $a^{(j)} \to c^{(j)} \leftarrow y^{(j)}$ and $y^{(k)} \to c^{(k)} \leftarrow b^{(k)}$ where $a, b \in V$.

An *m-path* exists between $[A]$ and $[B]$ in the mother graph if and only if there exists a sequence of triples between $[A]$ and $[B]$ such that at least one of the following two conditions is true for each triple in the sequence:

- $a^{(j)} {*\!\!-\!\!*} c^{(j)} {*\!\!-\!\!*} b^{(j)}$ with $a, b, c \in V \cup \{y\}$
- $a^{(j)} \to c^{(j)} \leftarrow y^{(j)}$ and $y^{(k)} \to c^{(k)} \leftarrow b^{(k)}$ where $a, b, c \in V$.

Finally, $[A]$ and $[B]$ are said to be *m-d-connected given* $[C]$ if and only if there exists an m-path between $[A]$ and $[B]$ such that the following two conditions hold:

- $c^{(j)} \in [C]$ for every m-collider on the path, where $c \in V$
- $a^{(j)} \notin [C]$ for every non-m-collider on the path, where $c \in V \cup \{y\}$.

Now, the Markov property for the mother graph states that if $[A]$ and $[B]$ are not m-d connected given $[C]$ in the mother graph, then $X_A \perp\!\!\!\perp X_B \mid X_C$ in p_μ (Strobl, 2019b;a).

We now provide a counter example for this Markov property. For this, consider the mother graph in Figure 1c over $V = \{1, 2, 3, 4\}$. Note that according to the definition of m-d-connection, $[\{1\}]$ and $[\{4\}]$ are not m-d-connected given $[\{2, 3\}]$. Hence, the Markov property should imply that $X_1 \perp\!\!\!\perp X_4 | X_2, X_3$ in any mixture distribution whose mother graph is as shown. In the following, construct a mixture distribution where this is not satisfied.

For simplicity, let $p_J(1) = p_J(2) = \frac{1}{2}$. Define $p^{(1)}(x_V)$ as

$$\begin{aligned} p^{(1)}(x_1) &= \mathcal{N}(x_1; 0, 1), \\ p^{(1)}(x_2|x_1) &= \mathcal{N}(x_2; x_1, 1), \\ p^{(1)}(x_3) &= \mathcal{N}(x_3; 0, 1), \\ p^{(1)}(x_4) &= \mathcal{N}(x_4; 0, 1), \end{aligned}$$

and $p^{(2)}(x_V)$ as

$$\begin{aligned} p^{(2)}(x_1) &= \mathcal{N}(x_1; 0, 1), \\ p^{(2)}(x_2) &= \mathcal{N}(x_2; 0, 1), \\ p^{(2)}(x_3|x_4) &= \mathcal{N}(x_3; x_4, 1), \\ p^{(2)}(x_4) &= \mathcal{N}(x_4; 0, 1). \end{aligned}$$

Clearly, $p^{(1)}(x_V)$ and $p^{(2)}(x_V)$ factorize according to $\mathcal{D}^{(1)}$ of Figure 1a and $\mathcal{D}^{(2)}$ of Figure 1b, respectively. Now,

$$\begin{aligned}
p_\mu(x_1,x_2,x_3,x_4) &= \sum_{j\in\{1,2\}} p_J(j)p^{(j)}(x_1,x_2,x_3,x_4)\\
&= \frac{1}{2}\frac{1}{(2\pi)^2}\Big(e^{-\frac{x_1^2}{2}}e^{-\frac{x_3^2}{2}}e^{-\frac{x_4^2}{2}}e^{-\frac{(x_2-x_1)^2}{2}} + e^{-\frac{x_1^2}{2}}e^{-\frac{x_2^2}{2}}e^{-\frac{x_4^2}{2}}e^{-\frac{(x_3-x_4)^2}{2}}\Big)\\
&= \frac{1}{2}\frac{1}{(2\pi)^2}e^{-\frac{x_1^2}{2}}e^{-\frac{x_2^2}{2}}e^{-\frac{x_3^2}{2}}e^{-\frac{x_4^2}{2}}\Big(e^{x_2x_1}e^{-\frac{x_1^2}{2}} + e^{x_3x_4}e^{-\frac{x_4^2}{2}}\Big),
\end{aligned}$$

which cannot be written as

$$f(x_1,x_2,x_3)g(x_2,x_3,x_4)$$

for any f, g, implying that $X_1 \not\perp\!\!\!\perp X_4 \mid X_2, X_3$ in p_μ. □

C. Proof of Lemma 3.3

Proof of Lemma 3.3. By the assumption, $p^{(j_1)}(x_V)$ factors according to $\mathcal{D}^{(j_1)}$. Hence, it is sufficient to define a distribution $\widetilde{p}_{X_V,J}(x_v,j)$ over $X_V \cup \{J\}$ that factors according to $\widetilde{\mathcal{D}}^{(j)}$, with $J \in \{j_1, j_2\}$ for an arbitrarily chosen $j_2 \in \{1,\dots,K\} \setminus \{j_1\}$, such that

$$\widetilde{p}_{X_V|J}(x_V|j_1) = p^{(j_1)}(x_V).$$

Then, the factorization with respect to $\widetilde{\mathcal{D}}^{(j_1)}$ along with the two d-separation statements in the hypothesis of the lemma would imply

$$\begin{aligned}
p^{(j_1)}(x_A,x_B|x_C) &= \frac{\sum_{x_{V\setminus(A\cup B\cup C)}} p^{(j_1)}(x_V)}{\sum_{x_{V\setminus C}} p^{(j_1)}(x_V)}\\
&= \frac{\sum_{x_{V\setminus(A\cup B\cup C)}} \widetilde{p}(x_V|j_1)}{\sum_{x_{V\setminus C}} \widetilde{p}(x_V|j_1)}\\
&= \widetilde{p}(x_A,x_B|x_C,j_1)\\
&= \widetilde{p}(x_A|x_C)\widetilde{p}(x_B|x_C,j_1).
\end{aligned}$$

To complete the proof, we define such a distribution $\widetilde{p}$. First let $V_y := \mathrm{ch}_{\widetilde{\mathcal{D}}^{(j_1)}}(y)$ and note that

$$\begin{aligned}
\widetilde{p}(x_V,j) &= \widetilde{p}_J(j)\prod_{v\in V}\widetilde{p}(x_v|x_{\mathrm{pa}_{\widetilde{\mathcal{D}}^{(j_1)}}(v)},j)\\
&= \widetilde{p}_J(j)\prod_{v\in V_y}\widetilde{p}(x_v|x_{\mathrm{pa}_{\mathcal{D}^{(j_1)}}(v)},j)\prod_{v\in V\setminus V_y}\widetilde{p}(x_v|x_{\mathrm{pa}_{\mathcal{D}^{(j_1)}}(v)}).
\end{aligned}$$

Define

$$\widetilde{p}_J(j) := \begin{cases} p_J(j_1) & j=j_1\\ 1-p_J(j_1) & j=j_2\end{cases},$$

$$\widetilde{p}(x_v|x_{\mathrm{pa}_{\mathcal{D}^{(j)}}(v)}) := p(x_v|x_{\mathrm{pa}_{\mathcal{D}^{(j)}}(v)}) \quad \forall v\in V\setminus V_y.$$

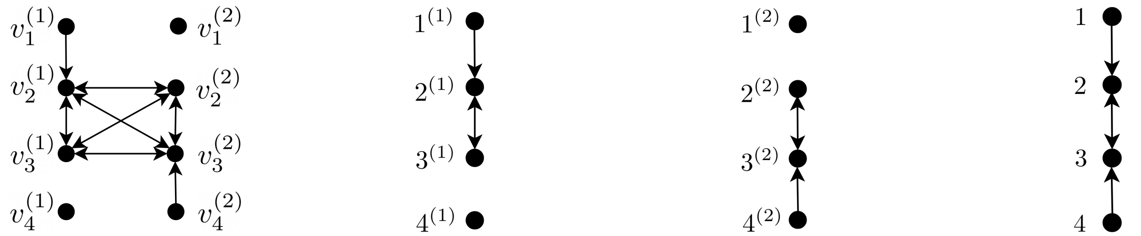
Now, for each $v \in V_y$, define

$$U(v) := \mathrm{pa}_{\mathcal{D}^{(j_1)}}(v) \cap \mathrm{pa}_{\mathcal{D}^{(j_2)}}(v)$$

and

$$D(v) := \mathrm{pa}_{\mathcal{D}^{(j_2)}(v)} \setminus \mathrm{pa}_{\mathcal{D}^{(j_1)}(v)},$$

and choose an arbitrary fixed value for $x_{\mathrm{pa}_{\mathcal{D}^{(i)}(v)}\setminus \mathrm{pa}_{\mathcal{D}^{(j)}(v)}}$ and denote it by $x'_d(v)$.



(a) $\mathcal{M}_\mu$ (b) $\mathcal{M}^{(1)}$ (c) $\mathcal{M}^{(2)}$ (d) $\mathcal{M}_\cup$

Figure 2

Then define for all $v \in V_y$,

$$\widetilde{p}(x_v|x_{\mathrm{pa}_{\mathcal{D}^{(j)}}(v)}, j) := \begin{cases} p_{X_v|X_{\mathrm{pa}_{\mathcal{D}^{(j_1)}}(v)},J}(x_v|x_{\mathrm{pa}_{\mathcal{D}^{(j_1)}}(v)}, j_1) & j = j_1 \\ p_{X_v|X_{U(v)},X_{D(v)},J}(x_v|x_{U(v)}, x'_d(v), j_2) & j = j_2 \end{cases}.$$

Now, one easily checks that this distribution indeed satisfies the factorization property, which completes the proof. $\square$

D. Proof of Lemma 4.3

The ancestral property follows directly since we impose the order compatibility assumption of Definition 4.1. In the following, we show maximality using the definition of inducing path and the associated maximality condition in Section A.

Proof of Lemma 4.3. Suppose we have a path $v_1 \leftrightarrow v_2 \leftrightarrow \dots v_{n-1} \leftrightarrow v_n$ in $\mathcal{M}_\cup$. Then, for all $m \in \{1, \dots, n-1\}$, we must have some $j \in \{1, \dots, K\}$ such that $v_m^{(j)} \leftrightarrow v_{m+1}^{(j)}$ in $\mathcal{M}^{(j)}$, implying that for all m, we must have a j such that $v_m^{(j)}, v_{m+1}^{(j)} \in \mathrm{ch}_{\mathcal{D}^{(j)}}(y)$ and hence a j such that $v_m^{(j)}, v_{m+1}^{(j)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$. But by construction of $\mathcal{D}_\mu$, this implies that $v_m^{(j)} v_{m+1}^{(j)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$ for all $j \in \{1, \dots, K\}$. Therefore, for any j, we have $v_1^{(j)} \cdots, v_n^{(j)} \in \mathrm{ch}_{\mathcal{D}^{(j)}}(y)$, and hence Algorithm 1 adds an edge between $v_1^{(j)}$ and $v_n^{(j)}$ in $\mathcal{M}^{(j)}$, resulting in an edge between v_1 and v_n in $\mathcal{M}_\cup$. Therefore, the path $v_1, \dots, v_n$ is not inducing in $\mathcal{M}_\cup$. $\square$

E. Proof of Theorem 4.4

Since we assume that $A, B, C \subseteq V$, i.e., these sets do not contain y, then $[A]$ and $[B]$ are d-separated in $\mathcal{D}_\mu$ given $[C]$ if and only if they are d-separated in the marginal MAG of $\mathcal{D}_\mu$ w.r.t. $\{y\}$ obtained from Algorithm 1. We refer to this MAG as the *mixture MAG* and denote it by $\mathcal{M}_\mu$. We will make use of this MAG in parts of the following proof since it simplifies the arguments.

One thing to note about $\mathcal{M}_\mu$ is that if we remove the edges of the form $u^{(j)} \circ\!\!-\!\!\circ v^{(i)}$ for $u, v \in V$ and $i \neq j$, then we obtain a bijection between the edges of $\mathcal{M}_\mu$ and the union of all the edges of $\mathcal{M}^{(j)}$ for all j. Figure 2 illustrates this for an example. Hence, we can alternatively think of the union graph as having directed edges

$$D_\cup := \{u \to v : u, v \in V,\ \exists_i\ u^{(i)} \to_{\mathcal{M}_\mu} v^{(i)}\},$$

and bidirected edges

$$B_\cup := \{u \leftrightarrow v : u, v \in V,\ \exists_i\ u^{(i)} \leftrightarrow_{\mathcal{M}_\mu} v^{(i)}\}.$$

We prove Theorem 4.4 in 3 main steps. *First*, in Lemma E.5 we show that for any d-connecting path between a and b given C in $\mathcal{M}_\cup$, we can find a d-connecting path between $a^{(i)}$ and $b^{(k)}$ given $[C]$ in $\mathcal{M}_\mu$. *Second*, in Lemma E.7 we show the converse: that for any d-connecting path $a^{(i)}$ and $b^{(k)}$ given $[C]$ in $\mathcal{M}_\mu$, we can find a d-connecting path between a and b given C in $\mathcal{M}_\cup$. *Finally*, in Lemma E.8 we show that this equivalence implies that for any disjoint sets $A, B, C \subseteq V$, A and B are d-separated in $\mathcal{M}_\cup$ if and only if $[A]$ and $[B]$ are d-separated in $\mathcal{M}_\mu$ given $[C]$.

The proof strategy in Lemmas E.5 and E.7 relies on concatenating d-connecting paths given C of the form $P_1 = \langle v_1, \dots, v_n \rangle$ and $P_2 = \langle v_n \dots, v_m \rangle$ together to create longer d-connecting paths given C of the form $P = \langle v_1, \dots, v_m \rangle$. When doing so, we must take care to ensure that v_n is active on the longer path, i.e., we must ensure that v_n is a collider on the path P if and only if $v_n \in C$.

E.1. A connecting path in $\mathcal{M}_\cup$ implies an analogus one in $\mathcal{M}_\mu$

We begin by proving some auxiliary results for step 1.

Lemma E.1 (Bidirected Connections). *If $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} b^{(k)}$ for any $i, k \in \{1, \ldots, K\}$, then $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} b^{(j)}$ for all $j \in \{1, \ldots, K\} \setminus \{i\}$.*

Proof. $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} b^{(k)}$ implies that $a^{(i)}, b^{(k)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$. By construction of $\mathcal{D}_\mu$, this implies $a^{(j)}, b^{(j)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$ for all $j \in \{1, \ldots, K\}$, and hence step 1 of Algorithm 1 will add the bidirected edges $a^{(i)} \leftrightarrow b^{(j)}$ for all $j \in \{1, \ldots, K\}$. Step 3 will only remove it if $a^{(i)}$ and $a^{(j)}$ are ancestors of one another in $\mathcal{D}_\mu$, which could happen only if $j = i$. Hence, $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} b^{(j)}$ for all $j \in \{1, \ldots, K\} \setminus \{i\}$. □

Lemma E.2 (Bidirected district). *Assume $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} b^{(j)}$ and $c^{(k)} \leftrightarrow_{\mathcal{M}_\mu} d^{(l)}$.*

- *If $i \neq l$, then $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} d^{(l)}$.*
- *If $i = l$, then*
  - *$a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} d^{(l)}$ if neither $a^{(i)}, d^{(l)}$ is an ancestor of another in $\mathcal{M}_\mu$,*
  - *$a^{(i)} \rightarrow_{\mathcal{M}_\mu} d^{(l)}$ if $a^{(i)} \in \mathrm{an}_{\mathcal{M}_\mu}(d^{(l)})$; or*
  - *$a^{(i)} \leftarrow_{\mathcal{M}_\mu} d^{(l)}$ if $d^{(l)} \in \mathrm{an}_{\mathcal{M}_\mu}(a^{(i)})$.*

Proof. $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} b^{(j)}$ and $c^{(k)} \leftrightarrow_{\mathcal{M}_\mu} d^{(l)}$ implies that $a^{(\iota)}, b^{(\iota)}, c^{(\iota)}, d^{(\iota)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$ for all $\iota \in \{1, \ldots, K\}$. Hence, step 1 of Algorithm 1 will add $a^{(i)} \leftrightarrow_{\mathcal{M}_\mu} d^{(l)}$. If $i \neq l$, then $a^{(i)}$ and $d^{(l)}$ cannot be ancestors of one another, implying that step 3 will not remove this bidirected edge. If $i = l$, then the edge will be removed and replaced with the appropriate directed edge if one of $a^{(i)}$ or $d^{(l)}$ is an ancestor of the other. Otherwise, the bidirected edge will remain. □

Lemma E.3 (Arrow tip lemma). *Under the ordering assumption in Definition 4.1, if a directed edge $a \rightarrow_{\mathcal{M}_\cup} b$ exists in $\mathcal{M}_\cup$, then we must have $a^j \rightarrow_{\mathcal{M}_\mu} b^j$ for some j in $\mathcal{M}_\mu$. If a bidirected edge $a \leftrightarrow_{\mathcal{M}_\cup} b$ exists in $\mathcal{M}_\cup$, then we must have $a^j \leftrightarrow_{\mathcal{M}_\mu} b^j$ for some j in $\mathcal{M}_\mu$.*

Proof. The proof follows directly from the definition of the union graph. □

Lemma E.4 (Changing Arrowtips Lemma). *Under the ordering assumption in Definition 4.1, if $a^{(j)} {*\!\!\rightarrow}_{\mathcal{M}_\mu} b^{(j)}$ but not $a^{(k)} {*\!\!\rightarrow}_{\mathcal{M}_\mu} b^{(k)}$ (same type of edge) for some $j \neq k$, then we must have $b^{(j)} \leftrightarrow b^{(k)}$.*

Proof. The ordering assumption does not allow $a^{(j)} \rightarrow_{\mathcal{M}_\mu} b^{(j)}$ and $a^{(k)} \leftrightarrow_{\mathcal{M}_\mu} b^{(k)}$ (and vice versa). Hence, we must only look at the existence of $a^{(j)} {*\!\!\rightarrow} b^{(j)}$ and the in-existence of an edge between $a^{(k)}$ and $b^{(k)}$.

First, we note that if step 1 of Algorithm 1 defining $\mathcal{M}_\mu$ adds $b^{(j)} \leftrightarrow b^{(k)}$, then it will remain since step 2 does not modify edges but only adds them, while step 3 will never remove an edge $b^{(j)} \leftrightarrow b^{(k)}$ since neither can be an ancestor or a descendant of the other in $\mathcal{D}_\mu$.

Now, if $a^{(j)} \rightarrow_{\mathcal{D}^{(j)}} b^{(j)}$ but not $a^{(k)} \rightarrow_{\mathcal{D}^{(k)}} b^{(k)}$ for some k, then we must have $b \in V \setminus V^I$ and hence $b^{(\iota)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$ for all $\iota \in \{1, \ldots, K\}$ by construction of $\mathcal{D}_\mu$. Therefore, step 1 of Algorithm 1 will add $b^{(j)} \leftrightarrow b^{(k)}$.

For the other case we must check that $a^{(j)} {\circ\!\!\rightarrow} b^{(j)}$ was added by the algorithm that created $\mathcal{M}_\mu$. In all steps, the algorithm will only add such an edge if $b \in V \setminus V_{\mathrm{INV}}$ and hence $b^{(j)} \leftrightarrow b^{(k)}$ must have been added in step 1. □

Lemma E.5 (Step 1). *Under the ordering compatibility assumption in Definition 4.1, if there is a connecting path between a and b given some $C \subseteq V \setminus \{a, b\}$ in $\mathcal{M}_\cup$ ending in an arrow head (or tail respectively) incident to b, then there is a connecting path between $a^{(i)}$ and $b^{(k)}$ given $[C]$ in $\mathcal{M}_\mu$ for some $i, k \in \{1, \ldots, K\}$ that also ends in an arrow head (or tail respectively) towards $b^{(k)}$.*

Proof. We use induction on the number of edges in the connecting path in $\mathcal{M}_\cup$. The base case for 1 edge follows directly from Lemma E.3.

Now assume we have a d-connecting path given C consisting of $m+1$ edges in $\mathcal{M}_\cup$: $P_\cup = \langle a, \ldots, d, b\rangle$ ending in an arrow head (or tail respectively). Consider the sub-path $\langle a, \ldots, d\rangle$ with m edges. By the inductive hypothesis, there is a path $P_\mu = \langle a^{(i)}, \ldots, d^{(j)}\rangle$ in $\mathcal{M}_\mu$ that is d-connecting given $[C]$, for some i, j, ending in the same tip. In the following, we show that we can always find a path of the form $\langle d^{(j)}, \ldots, b^{(k)}\rangle$ for some k that can be joined together with P_μ to create a path $\langle a^{(i)}, b^{(k)}\rangle$ that is d-connecting given $[C]$. We do this by considering all the different cases for the tips of the edges $c{*\!\!-\!\!*}_{\mathcal{M}_\cup} d$ and $d{*\!\!-\!\!*}_{\mathcal{M}_\cup} b$.

Before discussing the different cases, note that if the edge $d^{(j)}{*\!\!-\!\!*}_{\mathcal{M}_\mu} b^{(k)}$ exists and is of the same type as the edge $d{*\!\!-\!\!*}_{\mathcal{M}_\cup} b$, then we can create the desired d-connecting path $\widetilde{P}_\mu$ from $a^{(j)}$ to $b^{(i)}$ given $[C]$ by concatenating this edge with P_μ, since:

$$\begin{aligned} d \text{ is active on } Q_\cup &\Rightarrow \big(d \text{ is a collider on } Q_\cup \Leftrightarrow d \in C\big)\\ &\Rightarrow \big(d^{(j)} \text{ is a collider on } \widetilde{P}_\mu \Leftrightarrow d^{(j)} \in [C]\big)\\ &\Rightarrow d^{(j)} \text{ is active on } \widetilde{P}_\mu, \end{aligned}$$

where the second implication follows because the path P_μ ends in the correct type of arrow tip by the inductive hypothesis (I.H.). Hence, in what follows, it is sufficient to

$$\text{assume either } d^{(j)}{*\!\!-\!\!*} b^{(k)} \text{ is not in } \mathcal{M}_\mu \text{ or is not the same edge type as } d{*\!\!-\!\!*} b \text{ in } \mathcal{M}_\cup. \tag{1}$$

(i) case $c{*\!\!\rightarrow} d {\leftarrow\!\!*} b$ in $\mathcal{M}_\cup$:

Since d is a collider on the path $Q_\cup$, we must have $d \in C$, and hence $d^{(j)} \in [C]$ for all j. Hence the path P_μ is of the form $\langle a^{(i)}, \ldots, \gamma^{(\iota)}, d^{(j)}\rangle$, where $\gamma^{(\iota)} {*\!\!\rightarrow} d^{(j)}$ for some $\gamma \in V$ and $\iota \in \{1, \ldots, K\}$ by the I.H. Furthermore, by Lemma E.3, we must have $d^{(\iota)} {\leftarrow\!\!*}_{\mathcal{M}_\mu} b^{(k)}$ for some $\iota, k \in \{1, \ldots, K\}$. Since we assumed in (1) that this isn't true for $\iota = j$, then by Lemma E.4 we must have $d^{(j)} \leftrightarrow d^{(\iota)}$, creating the path ${*\!\!\rightarrow} d^{(j)} \leftrightarrow d^{(\iota)} {\leftarrow\!\!*} b^{(k)}$ that is d-connected given $[C]$ (recall $d^{(j)}, d^{(\iota)} \in [C]$). Concatenating $d^{(j)} \leftrightarrow d^{(\iota)} {\leftarrow\!\!\circ} b^{(k)}$ to P_μ gives the desired d-connecting path $\langle a^{(i)}, \ldots, d^{(j)}, d^{(\iota)}, b^{(k)}\rangle$.

For the remaining cases, we begin by recalling that the edge $d{*\!\!-\!\!*}_{\mathcal{M}_\cup} b$ must exist since $d^{(k)}{*\!\!-\!\!*}_{\mathcal{M}_\mu} b^{(k)}$ for some k by Lemma E.3. Now, let $\alpha^{(j)}$ be the node on the path P_μ closest to $a^{(i)}$ such that all nodes between $\alpha^{(j)}$ and $d^{(j)}$ have the same index j, i.e., all of these are contained in the same MAG $\mathcal{M}^{(j)}$. This means that the node preceding $\alpha^{(j)}$ on this path, call it $\gamma^{(\kappa)}$, either has a different index (i.e., a part of a different $\mathcal{M}^{(\kappa)}$), or $\alpha^{(j)} = a^{(i)}$.

Call $P_\mu^{(j)} = \langle \alpha^{(j)}, \ldots, d^{(j)}\rangle$ the subpath of P_μ from $\alpha^{(j)}$ to $d^{(j)}$. This path is completely contained in $\mathcal{M}^{(j)}$. If it is possible to find a path $P_\mu^{(k)} = \langle \alpha^{(k)}, \ldots, d^{(k)}\rangle$ in $\mathcal{M}^{(k)}$ that is analogous to $P_\mu^{(j)}$ (same types of edges), then we can replace the segment $P_\mu^{(j)}$ of P_μ with $P_\mu^{(k)}$ to obtain a connecting path between $a^{(i)}$ and $d^{(k)}$ given $[C]$. Then, concatenating $d^{(k)}{*\!\!-\!\!*} b^{(k)}$ gives us the desired connecting path from $a^{(i)}$ to $b^{(k)}$ given $[C]$ in $\mathcal{M}_\mu$.

Hence, in checking the remaining cases, we further

$$\text{assume that it is not possible to find a path } P_\mu^{(k)} \text{ in } \mathcal{M}^{(k)}. \tag{2}$$

Therefore, walking along the path $P_\mu^{(j)}$ backwards starting at $d^{(j)}$ until $\alpha^{(j)}$, we will eventually find an edge $\beta^{(j)}{*\!\!-\!\!*}\delta^{(j)}$ such that $\beta^{(k)}{*\!\!-\!\!*}\delta^{(k)}$ is not an edge. Take the first such edge. Now, if this edge was $\beta^{(j)} \leftrightarrow \delta^{(j)}$, then by Lemma E.1, we must have $\beta^{(j)} \leftrightarrow \delta^{(k)}$, implying that we can concatenate the subpath of P_μ of the form $\langle a^{(i)}, \ldots, \beta^{(j)}\rangle$ with $\beta^{(j)} \leftrightarrow \delta^{(k)}$ and the subpath of $P_\mu^{(k)}$ of the form $\langle \delta^{(k)}, \ldots, b^{(k)}\rangle$ to create the desired d-connecting path given $[C]$. Next we look at the situations where we do not have $\beta^{(j)} \leftrightarrow \delta^{(j)}$, considering each remaining case on the arrowheads of $c{*\!\!-\!\!*} d {*\!\!-\!\!*} b$ in $\mathcal{M}_\cup$ separately.

(ii) case $c \leftarrow d \rightarrow b$ in $\mathcal{M}_\cup$: This case is depicted in Figure 3a. If the first edge found is of the form $\beta^{(j)} \leftarrow \delta^{(j)}$ where $\beta^{(k)} \leftarrow \delta^{(k)}$ is not present (see Figure 3b), then by Lemmas E.4 and E.2, we must have $\beta^{(j)} \leftrightarrow b^{(k)}$ (Figure 3d). Replacing the segment $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ of P_μ with $\beta^{(j)} \leftrightarrow b^{(k)}$ gives the desired path.

Otherwise, if we have $\beta^{(j)} \rightarrow \delta^{(j)}$ instead (Figure 3c), then Lemmas E.4 and E.2 again say that we must have $\delta^{(j)} \leftrightarrow b^{(k)}$ (Figure 3e). The subpath of P_μ of the form $\langle \delta^{(j)}, c^{(j)} \rangle$ shown in Figure 3e is connecting given $[C]$ by the $I.H.$ Starting at $\delta^{(j)}$ and walking towards $c^{(j)}$, we can find a collider that is in $[C]$ (shown in Figure 3f). This collider must be a descendant of $\delta^{(j)}$ Hence, $\delta^{(j)}$ is active given $[C]$ on the path $\beta^{(j)} \rightarrow \delta^{(j)} \leftrightarrow b^{(k)}$ since it is a collider whose descendant is in $[C]$. Replacing the segment $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ in P_μ with this path gives the desired connecting path given $[C]$.

(iii) case $c \rightarrow d \rightarrow b$ in $\mathcal{M}_\cup$: Proceeding similarly, if the edge found is of the form $\beta^{(j)} \leftarrow \delta^{(j)}$, then we must have $\beta^{(j)} \leftrightarrow b^{(k)}$ similar to before and for the same reasons. Furthermore, we can find a d-connecting path by performing a concatenation similar to the one we did before: replace the segment $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ of P_μ with $\beta^{(j)} \leftrightarrow_{\mathcal{M}_\mu} b^{(k)}$. This is illustrated in Figure 4a,

If, otherwise, the edge found is of the form $\beta^{(j)} \rightarrow \delta^{(j)}$. We can conclude that we have the bidirected edge $\delta^{(j)} \leftrightarrow d^{(k)}$ by applying the Lemmas E.2 and E.4 again.

If there is a collider on the subpath between $\langle \delta^{(j)}, \ldots, c^{(j)} \rangle$, then any such collider must be in $[C]$ since P_μ is d-connecting given $[C]$ (see Figure 4b). Furthermore, one of these colliders will be a descendant of $\delta^{(j)}$, and we can apply similar logic to that in Case (ii) to show that the path obtained by replacing the segment $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ of P_μ with $\delta^{(j)} \leftrightarrow d^{(k)}$ is d-connecting given $[C]$.

Otherwise, no such collider exists between $\delta^{(j)}$ and $c^{(j)}$ and hence $c^{(j)}$ is a descendant of $\delta^{(j)}$ (see Figure 4c). Therefore, $b^{(k)}$ is a descendant of $\delta^{(k)}$ by the ordering compatibility assumption, and Algorithm 1 adds the directed edge $\delta^{(k)} \rightarrow b^{(k)}$ since $\delta^{(k)}$ and $b^{(k)}$ will both be in $\mathrm{ch}_{\mathcal{D}_\mu}(y)$. This further implies that $\delta^{(j)}, b^{(j)} \in \mathrm{ch}_{\mathcal{D}_\mu}(y)$, so Algorithm 1 will add an edge between these two nodes. The ordering assumption once again ensures that this edge is of the form $\delta^{(j)} \rightarrow b^{(j)}$.

(iv) case $c \leftarrow d \leftarrow b$ in $\mathcal{M}_\cup$. Proceeding similarly, if we have the edge $\beta^{(j)} \rightarrow \delta^{(j)}$, then we can follow the same logic to

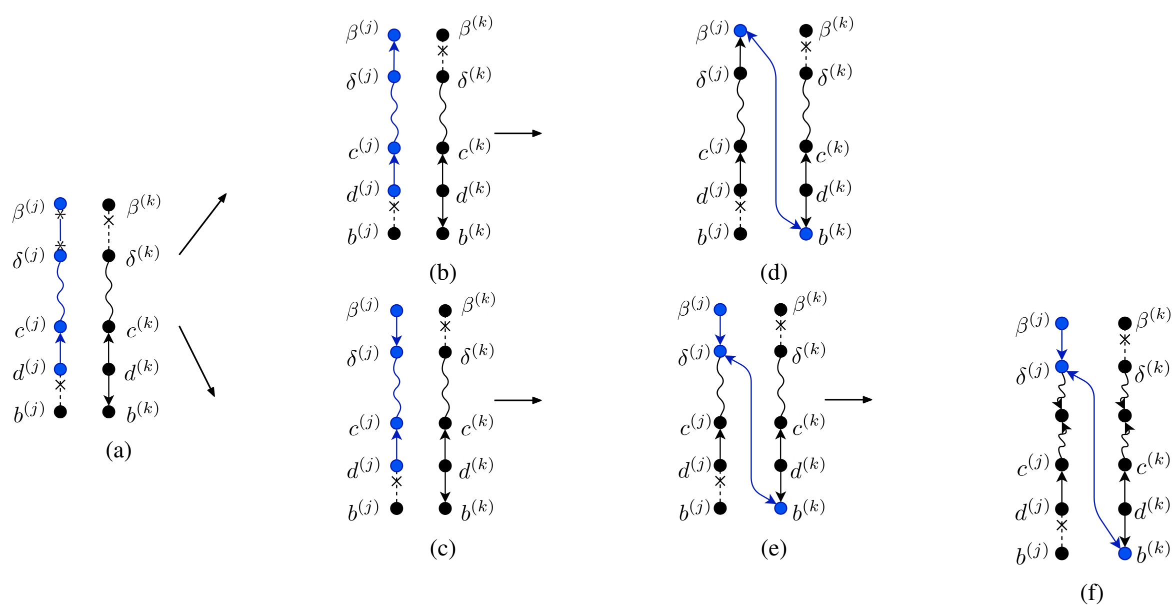


Figure 3: An illustration of the logic in the proof of Lemma E.5, case (ii). We do not plot all possible edges in order to reduce clutter. Instead, we plot non-edges using an × superimposed on a dashed line. Furthermore, we indicate paths between two nodes with a squiggly line. (a), (b) and (c) show the relevant segment of the path P_μ in blue; (d), (e) and (f) show the segment that replaces $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ on P_μ to create the desired d-connecting path in blue.

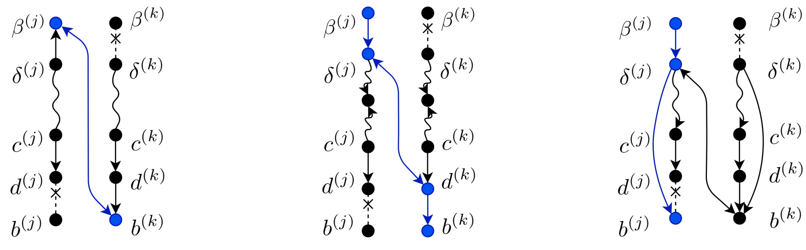


Figure 4: An illustration of the d-connecting paths constructed by following the logic of case (iii) in the proof of Lemma E.5. In each of (a), (b) and (c), the segment that replaces $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ on P_μ to create the desired d-connecting path is colored in blue.

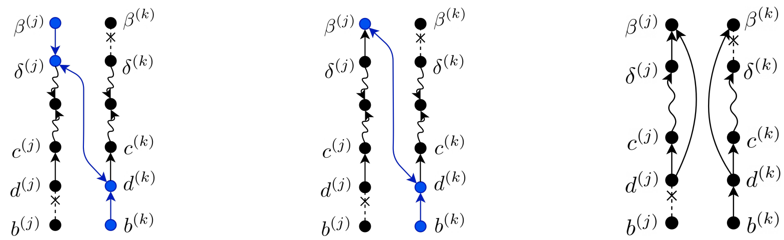


Figure 5: An illustration of the d-connecting paths constructed by following the logic of case (iv) in the proof of lemma E.5. In each of (a) and (b), the segment that replaces $\langle \beta^{(j)}, \ldots, d^{(j)} \rangle$ on P_μ to create the desired d-connecting path is colored in blue.

create the d-connecting path (see Figure 5a).

Otherwise, $\beta^{(j)} \leftarrow \delta^{(j)}$, and we have the bidirected edge $\beta^{(j)} \leftrightarrow d^{(k)}$, and we again check for colliders between $\beta^{(j)}$ and $d^{(j)}$.

If there is a collider, it will be both in $[C]$ and a descendant of $d^{(k)}$ in$\mathcal{M}_\mu$, and we can find the desired d-connecting path with the same logic followed previously (see Figure 5b).

If there is no such collider, then $\beta^{(j)}$ will be a descendant of $d^{(j)}$, and using a similar argument to that used for Figure 4c, we can conclude that we have directed edges $\beta^{(j)} \leftarrow_{\mathcal{M}_\mu} d^{(j)}$ and $\beta^{(k)} \leftarrow_{\mathcal{M}_\mu} d^{(k)}$ (see Figure 5c). In such a scenario, we can repeat the logic for the node β in place of the node c: we continue walking along the path $P_\mu^{(j)}$ starting from $\beta^{(j)}$ until $\alpha^{(j)}$ is reached or until we find another edge along this path that does not exist on $P_\mu^{(k)}$. If the former happens first, we deal with the case like we would have if $P_\mu^{(k)}$ and $P_\mu^{(j)}$ had identical edges. If the latter happens first, then we recursively repeat the logic of case (iv).

This completes the proof. □

E.2. A d-connecting path in $\mathcal{M}_\mu$ implies an analogous d-connecting path in $\mathcal{M}_\cup$

Again, we begin with some auxiliary results.

Lemma E.6 (At most 1 bidirected edge)**.** *If there exists a connecting path between $a^{(i)}$ and $b^{(k)}$ given some $[C]$ where $a, b \in V$ and $C \subseteq V \setminus \{a, b\}$ in $\mathcal{M}_\mu$, then there must exist a path $\widetilde{P}_\mu$ between $a^{(i)}$ and $b^{(k)}$ that is also connecting given $[C]$ that contains at most one bidirected edge.*

Proof. Since $a^{(i)}$ and $b^{(k)}$ are connected given $[C]$ in $\mathcal{M}_\mu$, then they must also be connected given $[C]$ in $\mathcal{D}_\mu$. Let P_μ denote the path connecting $a^{(i)}$ to $b^{(k)}$ given $[C]$ in $\mathcal{D}_\mu$. Let $P_\mu = \langle a^{(i)}, u_1, \ldots, u^{(l)} \rangle$ and let u_x, u_z be the first and last occurrences of the vertex y on P_μ, respectively, if any. Since y has an in-degree of 0, neither u_x nor u_z can be a collider. Hence, we can concatentate the paths $P_1 = \langle a^{(i)}, \ldots, u_x \rangle$ and $P_2 = \langle u_z, \ldots, b^{(k)} \rangle$ to get a connecting path given $[C]$ in $\mathcal{D}_\mu$.

Now, if u_{x-1} is neither an ancestor nor a descendant of d_{z+1}, then in $\mathcal{M}_\mu$, we will have the path $a^{(i)}, \dots, u_{x-1}, u_{z+1}, \dots, b^{(k)}$ by virtue of Algorithm 1, since it adds a bidirected edge between any pair of children of y. This is a path from $a^{(i)}$ to $b^{(k)}$ that is also connected given $[C]$ that contains only 1 bidirected edge.

Otherwise, (W.L.O.G) $u_{x-1} \in \mathrm{an}_{\mathcal{D}_\mu}(u_{z+1})$, i.e., there is a directed path from u_{x-1} to u_{z+1} in $\mathcal{D}_\mu$. Step 3 of Algorithm 1 adds the edge $u_{x-1} \to u_{z+1}$ to $\mathcal{M}_\mu$ to create the path $\widetilde{P}_\mu := \langle a^{(i)}, \dots, u_{x-1}, u_{z+1}, \dots, b^{(k)} \rangle$. This path is from $a^{(i)}$ to $b^{(k)}$ and passes through no bidirected edges. If this path is active, then we are done. If this path is not active, then, since $\langle a^{(i)} \dots, u_{x-1} \rangle$ and $\langle u_{z+1}, \dots, b^{(k)}$ are active, $\widetilde{P}_\mu$ must be inactive by virtue of $u_{x-1} \in [C]$. But since P_μ in $\mathcal{D}_\mu$ is connecting, this implies that u_{x-1} must have been a collider on that path, hence we have the edge $u_{x-2} \to u_{x-1}$ in $\mathcal{D}_\mu$ and $\mathcal{M}_\mu$. Step 2 of Algorithm 1 adds $u_{x-2} \to u_{z+1}$ in such a case. Then, the path $\langle a^{(i)}, \dots, u_{x-1}, u_{z+1}, \dots, b^{(k)} \rangle$ must be connecting from $a^{(i)}$ to $b^{(k)}$ given $[C]$, which completes the proof. □

Lemma E.7 (A Connecting Path in $\mathcal{M}_\mu$ implies a connecting path in $\mathcal{M}_\cup$)**.** *Under the assumptionin Defintion 4.1, if there is a connecting path between $a^{(i)}$ and $b^{(k)}$ given some $[C]$ in $\mathcal{M}_\mu$ for some $i, k \in \{1, \dots, K\}$, where $C \subseteq V \setminus \{a, b\}$, then there is a connecting path between a and b given C in $\mathcal{M}_\cup$.*

Proof. By Lemma E.6, we must have a connecting path in $\mathcal{M}_\mu$ between $a^{(i)}$ and $b^{(k)}$ given $[C]$ that passes through at most 1 bidirected edge. If there exist paths that pass through no bidirected edges, take any such path. Otherwise, take any path that passes through 1 bidirected edge. Call this path $P_\mu = \langle a^{(i)} =: u_0^{(i)}, u_1^{(i)}, \dots, u_m^{(k)} := b^{(k)} \rangle$.

By the structure of $\mathcal{M}_\mu$ discussed in the beginning of this section, only a bidirected edge can connect a node $u_x^{(i)}$ to a node $u_{x+1}^{(k)}$ in $\mathcal{M}_\mu$ for $i \neq k$. Hence, if there is no bidirected edge on this path, then all the nodes $u_0^{(i)}, \dots, u_m^{(i)}$ will be contained in the same MAG $\mathcal{M}^{(i)}$. Each edge along this d-connecting path given $[C]$ will show up in $\mathcal{M}_\cup$, and hence we can create a path $\langle u_0, \dots, u_m \rangle$ that is d-connecting given C in $\mathcal{M}_\cup$.

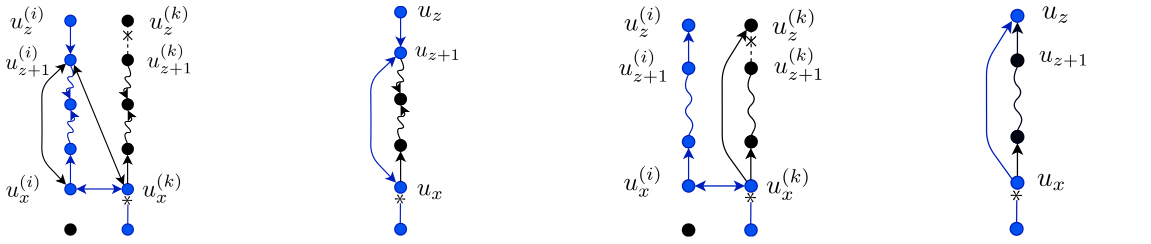
In the case where P_μ contains a bidirected edge, let us label the nodes incident as $u_x^{(i)} \leftrightarrow u_{x+1}^{(k)}$. The segments $\langle u_0^{(i)}, \dots, u_x^{(i)} \rangle$ and $\langle u_{x+1}^{(k)}, \dots, u_m^{(k)} \rangle$ will each be contained in $\mathcal{M}^{(i)}$ and $\mathcal{M}^{(k)}$ respectively, and hence we can find d-connecting paths $\langle u_0, \dots, u_x \rangle$ and $\langle u_{x+1}, \dots, u_m \rangle$ in $\mathcal{M}_\cup$ that are each d-connecting given C. We must now show that we can connect these paths to create a d-connecting path given C from $u_0 = a$ to $u_m = b$ in $\mathcal{M}_\cup$.

Of course, there is no difficulty if the bidirected edge $u_x \leftrightarrow u_{x+1}$ appears in $\mathcal{M}_\cup$, since we can connect these two subpaths with this bidirected edge and have the desired connecting path. The difficulty is when this edge does not appear. From the definition of $\mathcal{M}_\cup$, we can see that this only happens when the bidirected edge connects $u_x^{(i)}$ and $u_{x+1}^{(k)}$ for $i \neq k$, i.e., the bidirected edge is not contained in any MAG $\mathcal{M}^{(j)}$ for any j. We split the remainder into two cases.

(i) case $u_x = u_{x+1}$. If $u_x^{(i)}$ and $u_{x+1}^{(k)}$ are both colliders on P_μ, then we must have $u_x, u_{x+1} \in C$. Then $c = d$ will be an active collider given C in $\mathcal{M}_\cup$ on the path obtained by concatenating $\langle u_0, \dots, u_x \rangle$ and $\langle u_{x+1}, \dots, u_m \rangle$ in $\mathcal{M}_\cup$, and hence we have our d-connecting path given C. We therefore assume, W.L.O.G., that $u_x^{(i)}$ is not a collider on P_μ.

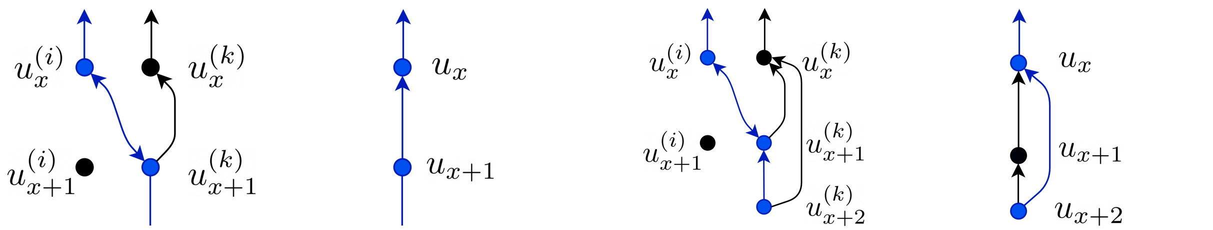
If there is a path $\langle u_0^{(k)}, \dots, u_x^{(k)} \rangle$ in $\mathcal{M}_\mu$ where every pair of adjacent vertices $u_n^{(k)}, u_{n+1}^{(k)}$ on this path are connected by the same edge type as the pair $u_n^{(i)}, u_{n+1}^{(i)}$ in P_μ, then we can replace the segment of $\langle u_0^{(i)}, \dots, u_x^{(i)} \rangle$ of P_μ with $\langle u_0^{(k)}$ to $u_x^{(k)} \rangle$ to obtain a path that is d-connecting given $[C]$ and contained completely in $\mathcal{M}^{(k)}$, meaning that we can find the desired d-connecting path given C in $\mathcal{M}_\cup$. If no such path exists in $\mathcal{M}_\mu$, then starting at $u_x^{(i)}$ and walking backwards along P_μ towards $u_0^{(i)}$, we will find an edge $u_z^{(i)} \ast\!\!-\!\!\ast_{\mathcal{M}_\mu} u_{z+1}^{(i)}$ where $u_z^{(k)} \ast\!\!-\!\!\ast_{\mathcal{M}_\mu} u_{z+1}^{(k)}$ is not an edge. Take the first such edge found (i.e., the edge closest to $u_x^{(i)}$ that satisfies this; see Figure 6a).

If $u_z^{(i)} \to_{\mathcal{M}_\mu} u_{z+1}^{(i)}$, then by Lemmas E.4 and E.2, there is a bidirected edge $u_{z+1}^{(i)} \leftrightarrow_{\mathcal{M}_\mu} u_x^{(k)}$, implying that step 1 of Algorithm 1 adds another bidirected edge $u_{z+1}^{(k)} \leftrightarrow u_x^{(k)}$. If $u_{z+1}^{(i)}$ is not a descendant of $u_x^{(i)}$, then the bidirected edge $u_{z+1}^{(k)} \leftrightarrow_{\mathcal{M}_\mu} u_x^{(k)}$ would not be removed by step 3 of Algorithm 1 and hence will appear in $\mathcal{M}_\mu$. Furthermore, we will have collliders $\alpha^{(i)}$ and $\gamma^{(i)}$ between $u_{z+1}^{(i)}$ and $u_x^{(i)}$ that are in $[C]$ that will be descendants of $u_{z+1}^{(i)}$ and $u_x^{(i)}$ respectively. The ordering assumption ensures that α and γ are descendants of u_{z+1} and u_x in $\mathcal{M}_\cup$, respectively. Hence, the path $\langle u_0, \dots, u_{z+1}, u_x, \dots, u_m \rangle$ in $\mathcal{M}_\cup$ is d-connecting in $\mathcal{M}_\cup$ given C. Figures 6a and 6b illustrate this.



(a) (b) (c) (d)

Figure 6: An illustration of the logic for case (i) for the proof of Lemma E.7. In (a) and (c), we color in blue the relevant segments of the d-connecting path in $\mathcal{M}_\mu$, while in (b) and (d), we color in blue the relevant segments of the constructed d-connecting path in $\mathcal{M}_\cup$.



(a) (b) (c) (d)

Figure 7: An illustration of the logic for case (ii) for the proof of Lemma E.7. In (a) and (c), we color in blue the relevant segments of the d-connecting path in $\mathcal{M}_\mu$, while in (b) and (d), we color in blue the relevant segments of the constructed d-connecting path in $\mathcal{M}_\cup$.

Now we check the case where $u_z^{(i)} \leftarrow u_{z+1}^{(i)}$. If $u_z^{(i)}$ is not a descendant of $u_x^{(i)}$, then we can construct a path in $\mathcal{M}_\cup$ by a similar argument to the above. If $u_z^{(i)}$ is a descendant of $u_x^{(i)}$, then by Lemma E.2, there is a directed edge $u_z^{(i)} \leftarrow_{\mathcal{M}_\mu} u_x^{(i)}$, which appears as $u_z \leftarrow_{\mathcal{M}_\cup} u_x$. We can use this to construct a path in $\mathcal{M}_\cup$ as shown in Figures 6c and 6d. This path is active since $u_x^{(i)}$ is not a collider, and hence $u_x^{(i)} \notin [C]$, implying that $u_x \notin C$.

(ii) case $u_x \neq u_{x+1}$: Step 1 of Algorithm 1 adds the bidirected edge $u_x^{(k)} \leftrightarrow u_{x+1}^{(k)}$, which will show up in $\mathcal{M}_\cup$ as an edge $u_x \leftrightarrow u_{x+1}$ unless it is removed by step 3; so this is the only case we must check. Assume W.L.O.G. that this edge is removed by step 3 because $u_x^{(k)}$ is a descendant of $u_{x+1}^{(k)}$ in $\mathcal{D}_\mu$ and therefore in $\mathcal{M}_\mu$. Then a directed edge $u_x^{(i)} \leftarrow u_{x+1}^{(i)}$ will be added instead, which appears in $\mathcal{M}_\cup$ as $u_x \leftarrow u_{x+1}$. The only case where we cannot join $\langle u_0, \ldots, u_x \rangle$ and $\langle u_{x+1}, \ldots, u_m$ in $\mathcal{M}_\cup$ together using this directed edge $u_x \leftarrow u_{x+1}$ to create a d-connected path given C is when $u_{x+1}^{(k)}$ is in $[C]$, and hence is a collider on P_μ. This implies that we have $u_{x+1}^{(k)} \leftarrow_{\mathcal{M}_\mu} u_{x+2}^{(k)}$. in which case step 2 of Algorithm 1 would have added the edge $u_{x+1}^{(k)} \leftarrow u_{x+2}^{(k)}$, which appears as $u_{x+1} \leftarrow_{\mathcal{M}_\cup} u_{x+2}$. This edge can be used to create the d-connecting path given C given by $\langle u_0, \ldots, u_x, u_{x+2}, \ldots, u_m \rangle$ in $\mathcal{M}_\cup$. This is illustrated in Figure 7 and completes the proof. □

E.3. The main result

Finally, we use the results of the first two steps to prove the following.

Theorem E.8. *Under the assumption in Definition 4.1, for any disjoint $A, B, C \subseteq V$, $[A]$ and $[B]$ are d-separated given $[C]$ in $\mathcal{D}_\mu$ if and only if A and B are d-separated given C in $\mathcal{M}_\cup$.*

Proof. Since $\mathcal{M}_\mu$ is the marginal MAG in $\mathcal{D}_\mu$ with respect to the vertex y, the d-separation statements involving subsets not including y are the same in both. By proposition 2.1, $\mathcal{M}_\mu$ is a MAG, hence d-separation in $\mathcal{M}_\mu$ is compositional (Sadeghi and Lauritzen, 2014); therefore for $A, B, C \subseteq V$ disjoint it holds that

$$\Big\{[A] \text{ sep from } [B] \text{ in } \mathcal{M}_\mu \text{ given } [C]\Big\} \\ \Leftrightarrow \Big\{a^i \text{ sep from } b^k \text{ in } \mathcal{M}_\mu \text{ given } [C] \text{ for all } a^i \in [A], b^k \in [B]\Big\}.$$

Now Lemmas E.5 and E.7 imply

$$\Big\{a^i \text{ sep from } b^k \text{ in } \mathcal{M}_\mu \text{ given } [C] \text{ for all } a^i \in [A], b^k \in [B]\Big\} \\ \Leftrightarrow \Big\{a \text{ sep from } b \text{ given } C \text{ for all } a \in A, b \in B\Big\}.$$

Finally, since $\mathcal{M}_\cup$ is a MAG, applying compositionality gives

$$\Big\{a \text{ sep from } b \text{in } \mathcal{M}_\cup \text{ given } C \text{ for all } a \in A, b \in B\Big\} \Leftrightarrow \Big\{A \text{ sep from } B \text{ given } C \text{ in } \mathcal{M}_\cup\Big\},$$

which completes the proof. □

F. Proof of Proposition 4.6

$u \leftrightarrow_{\mathcal{M}_\cup} v$ implies $u^{(j)} \leftrightarrow_{\mathcal{M}^{(j)}} v^{(j)}$ for some j, which implies $u \leftarrow y \rightarrow v$ in $\mathcal{D}_\mu$. Hence, $u, v \in V \setminus V_{\text{INV}}$. By definition of V_{INV}, this implies the claim.

G. Additional Experimental Results

G.1. Synthetic Data

In the following, we present figures for the experiments described in Section 5 for additional values of K and n, and when $p(j)$ is not uniform over the mixture components. Figures **??** shows the normalized SHD plot in evaluating the union graph as described in the main paper, while Figures **??** shows the true and false positives in predicted $V \setminus V_{\text{INV}}$. Finally, Figure 12 shows the result of K-means clustering.

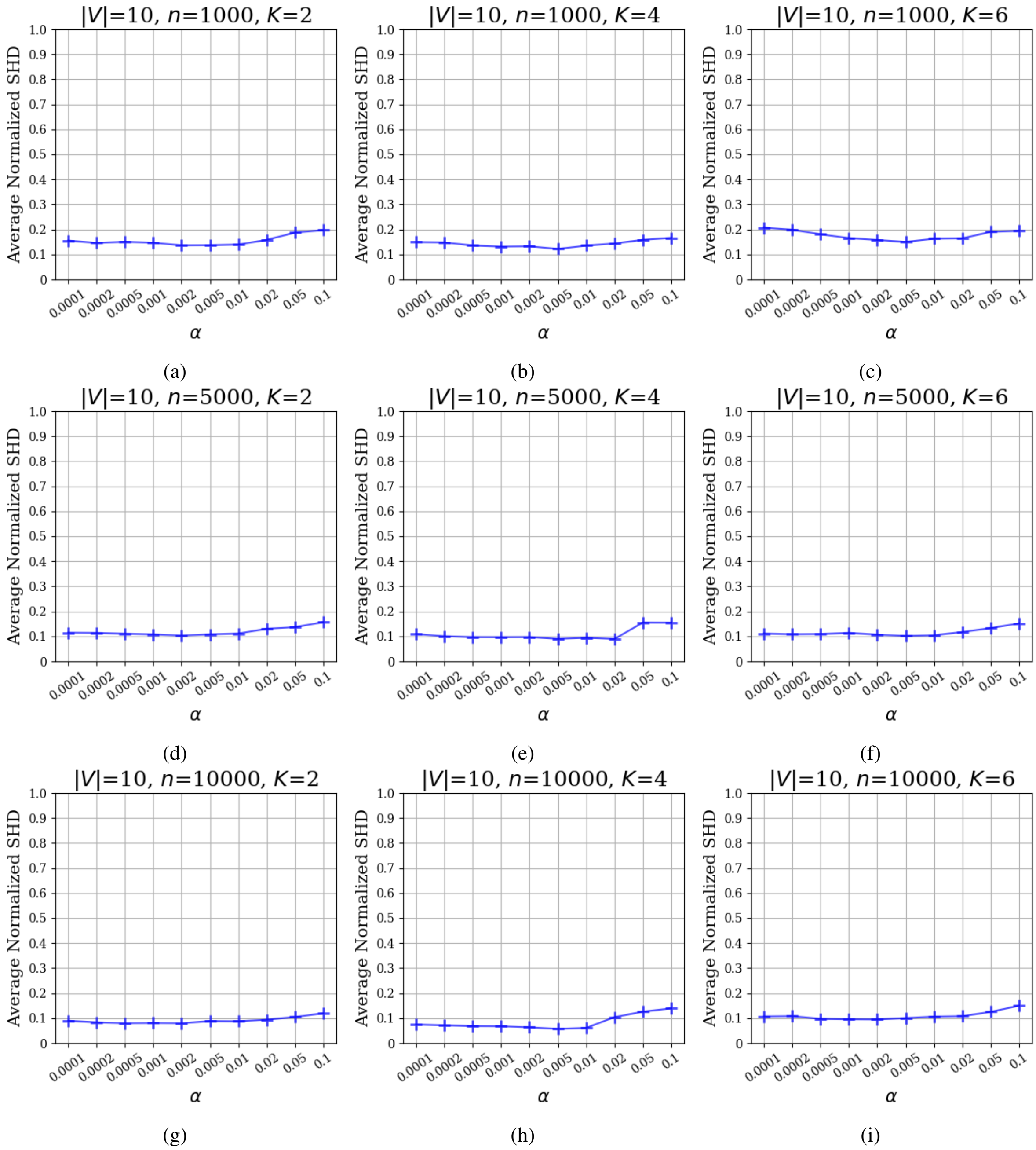


Figure 8: Normalized SHD evaluating the estimation of the union graph from mixture data using FCI for $K \in \{2, 4, 6\}$ and $n \in \{1000, 5000, 10000\}$. We take $p(j)$ uniform over the mixture components.

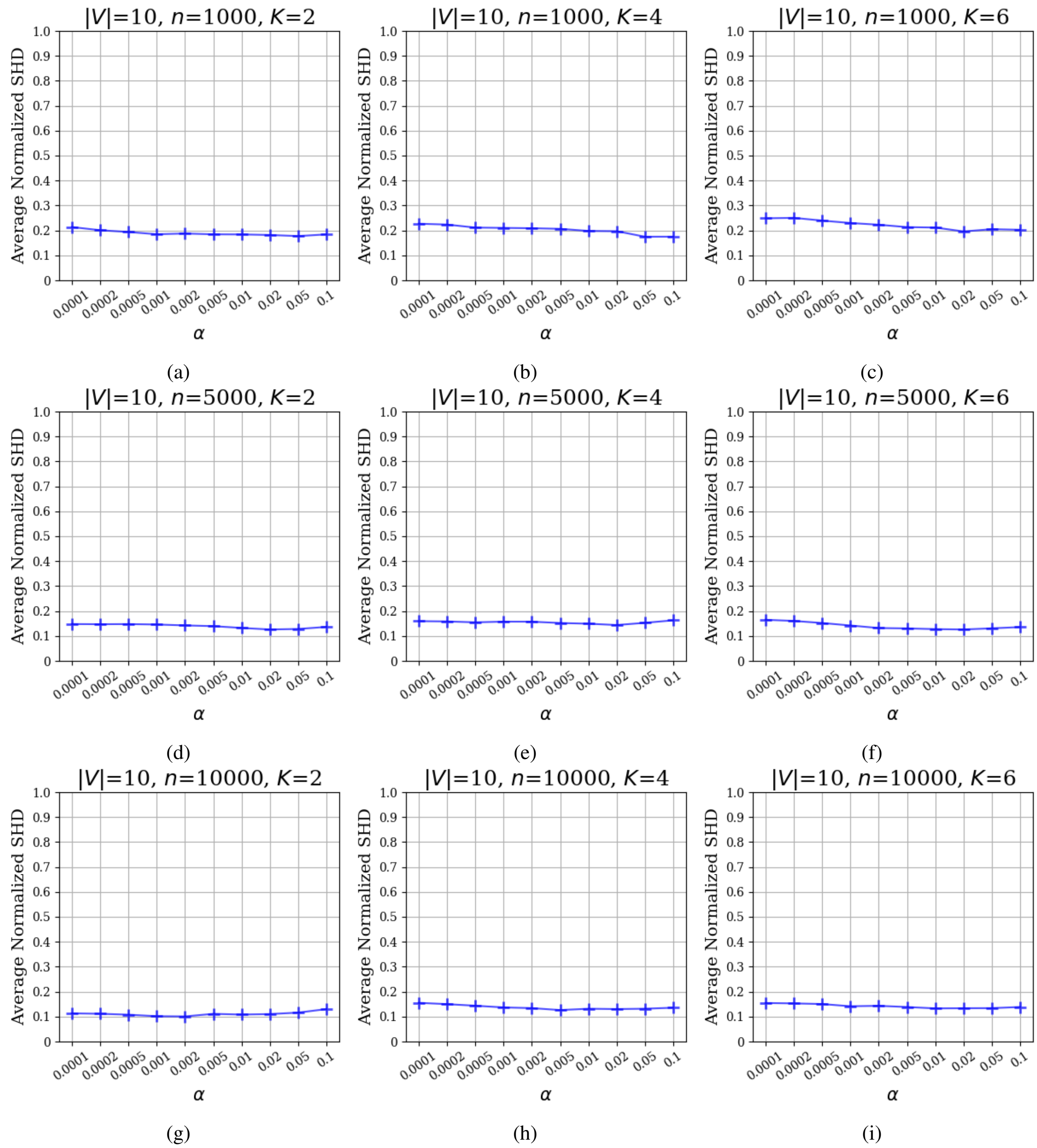


Figure 9: Normalized SHD evaluating the estimation of the union graph from mixture data using FCI for $K \in \{2, 4, 6\}$ and $n \in \{1000, 5000, 10000\}$. We take $p(j)$ to be Dirichlet with parameter 2.

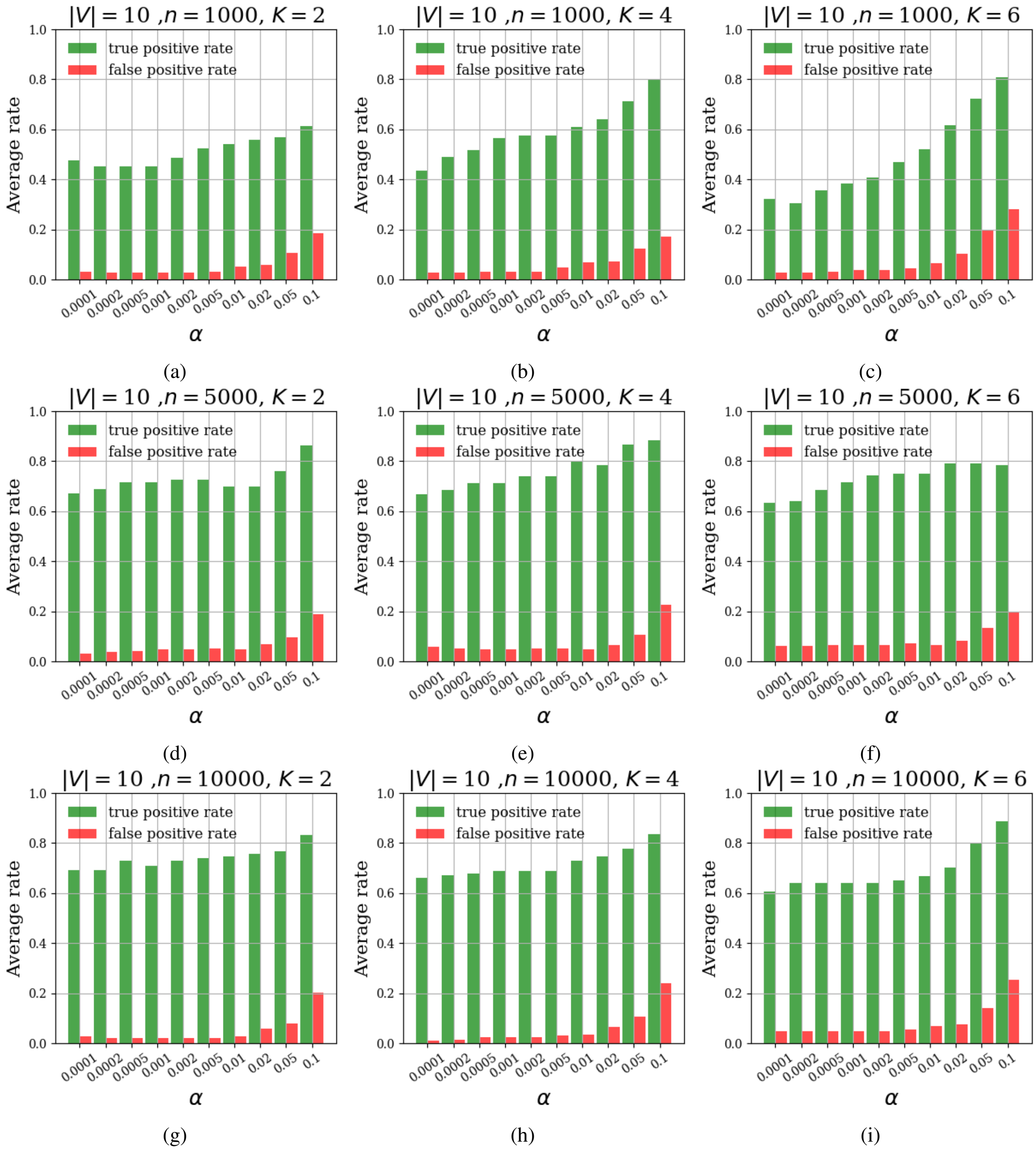


Figure 10: True and false positive rates in estimating $V \setminus V_{\text{INV}}$ using Proposition 4.6 applied to the PAG $\widehat{\mathcal{P}}_{\cup}$ estimated by running FCI on the mxiture data. The figures show the results for $K \in \{2, 4, 6\}$ and $n \in \{1000, 5000, 10000\}$. We take $p(j)$ to be uniform.

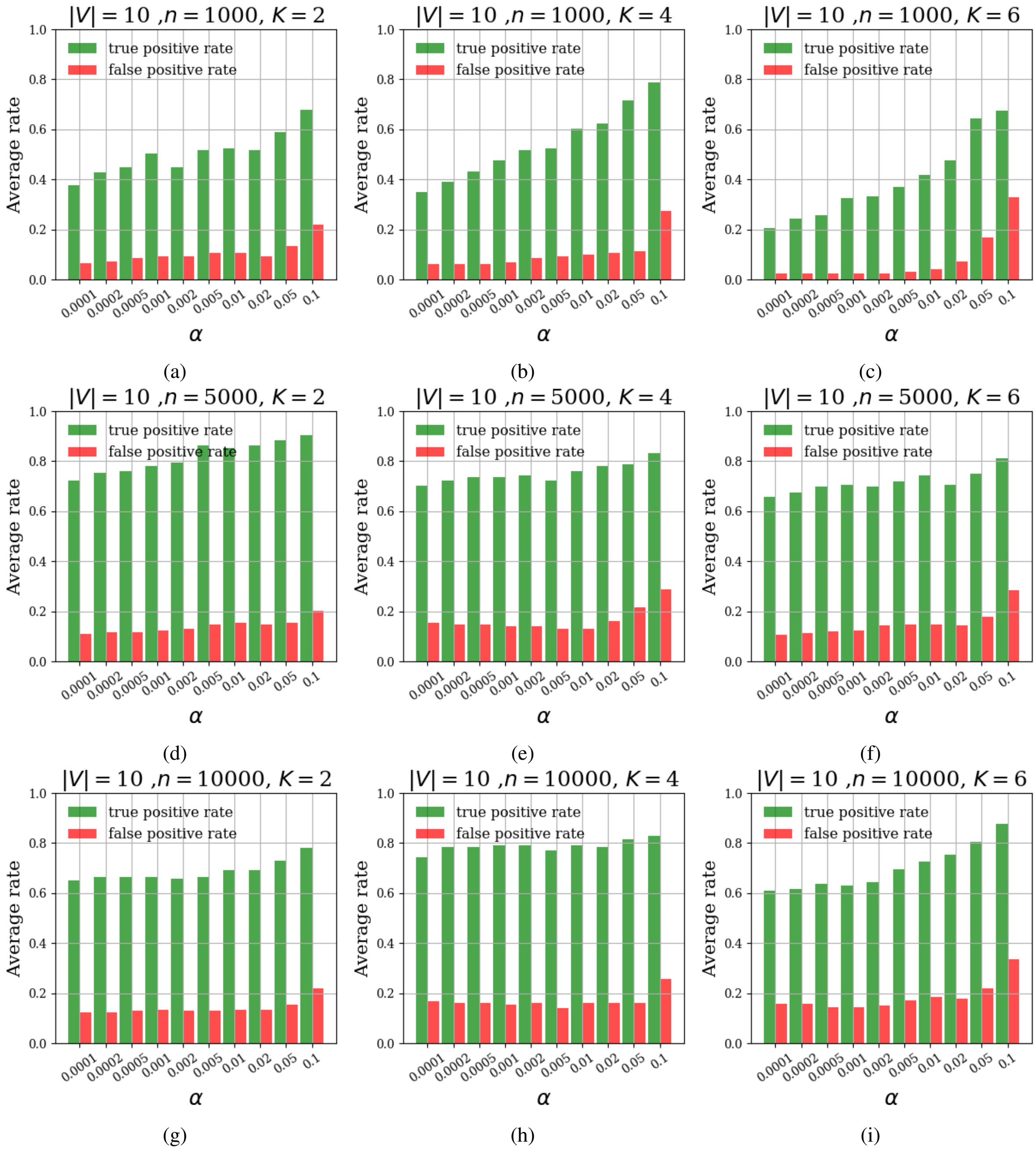
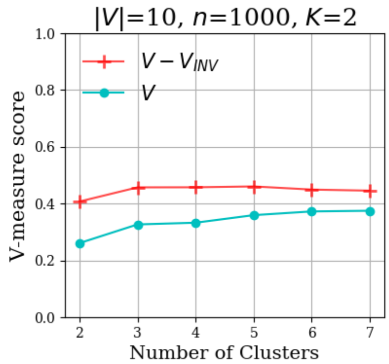
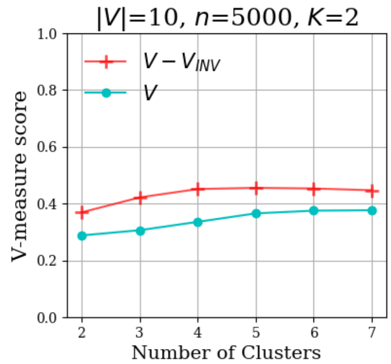


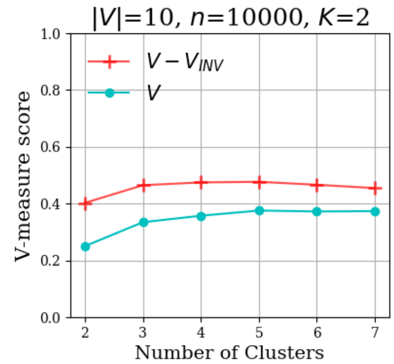
Figure 11: True and false positive rates in estimating $V \setminus V_{\text{INV}}$ using Proposition 4.6 applied to the PAG $\widehat{\mathcal{P}}_{\cup}$ estimated by running FCI on the mxiture data. The figures show the results for $K \in \{2, 4, 6\}$ and $n \in \{1000, 5000, 10000\}$. We take $p(j)$ to be Dirichlet with parameter 2



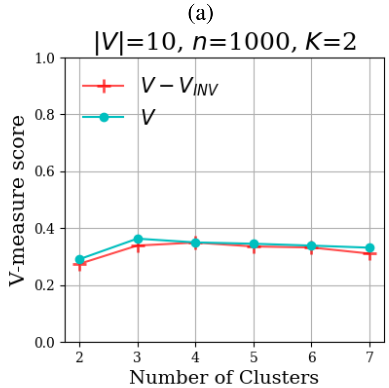
(a)



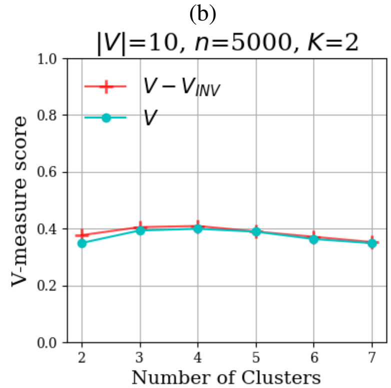
(b)



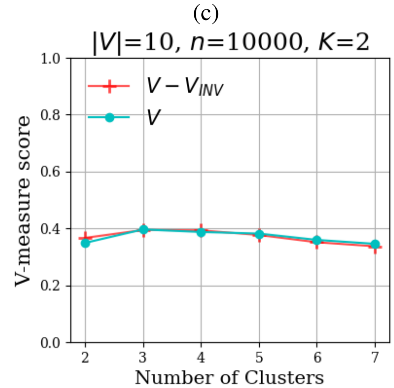
(c)



(d)



(e)



(f)

Figure 12: A comparison of clustering when all the variables are used as features vs. when only the variables in the estimated set $V \setminus V_{\text{INV}}$ are used as features. In generating figures (a), (b) and (c), $V \setminus V_{\text{INV}}$ has descendants in the generating model, while in figures (d), (e) and (f), $V \setminus V_{\text{INV}}$ has no descendants.

G.2. Real Data

Here, we present the output of FCI on the T cell mixture data referenced in section 5.2.

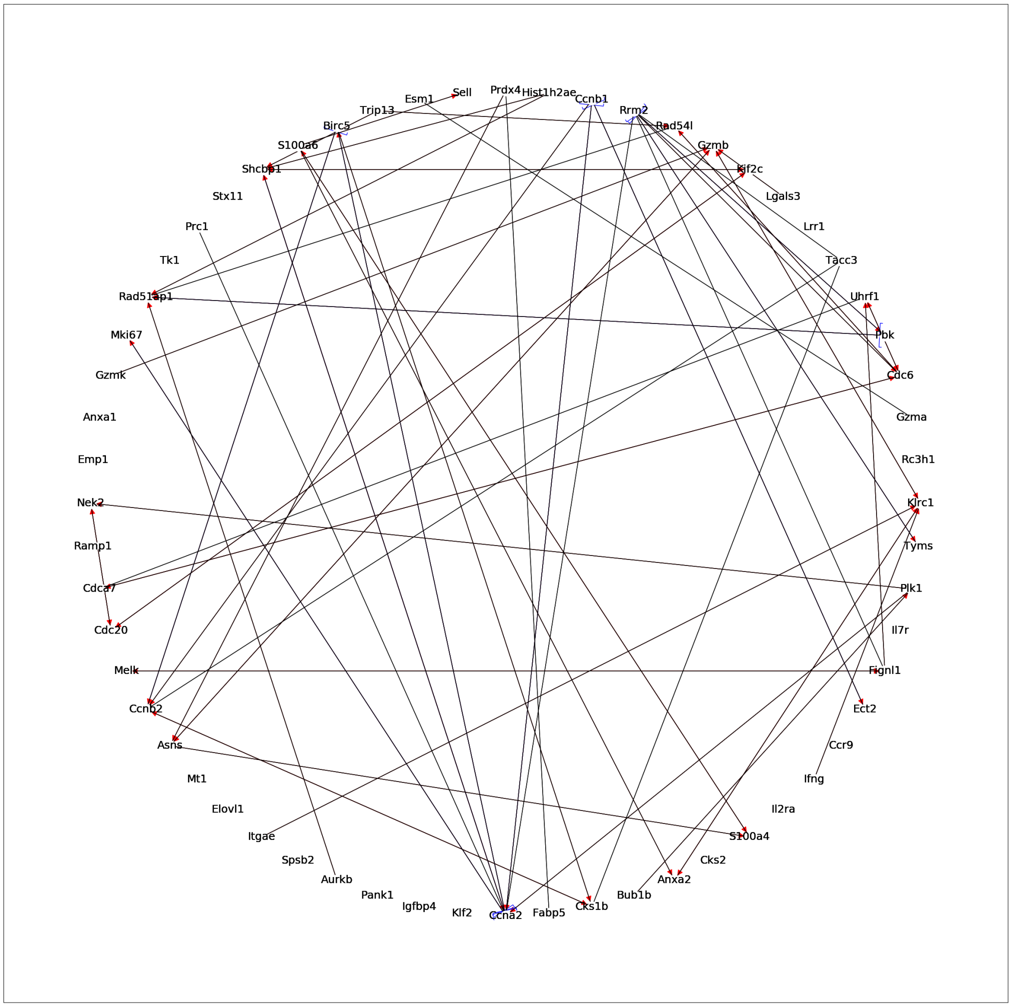


Figure 13: The PAG learned using FCI on the T cell mixture data. The inferred arrowheads are shown in red, while the inferred arrowtails are shown as blue brackets.

REFERENCES

Richardson, T. and Spirtes, P. (2002). Ancestral graph Markov models. *The Annals of Statistics*, 30(4):962–1030.

Sadeghi, K. and Lauritzen, S. (2014). Markov properties for mixed graphs. *Bernoulli*, 20(2):676–696.

Strobl, E. V. (2019a). The global Markov property for a mixture of DAGs. *arXiv preprint arXiv:1909.05418*.

Strobl, E. V. (2019b). Improved causal discovery from longitudinal data using a mixture of DAGs. In *Proceedings of Machine Learning Research*, volume 104, pages 100–133.