APPENDIX A: EXAMPLE RUN

We show below an example run of the operation Fbsplit\((K, L, G)\) presented in Figure 1 of the main text. Let \(G\) be the CG below, \(K = \{1, 2, 3, 4, 6, 7\}\), and \(L = \{3, 4, 7\}\).

\[
\begin{array}{c}
1 \\
\downarrow \\
3 \\
\end{array} \quad \begin{array}{c}
2 \\
\downarrow \quad 6 \\
\downarrow \\
4 \\
\downarrow \\
7 \\
\end{array}
\]

According to line 1, \(L_1 = \{3, 4\}\) and \(L_2 = \{7\}\). After having executed lines 2-4 for \(i = 1\), \(G\) looks like the CG below.

\[
\begin{array}{c}
1 \\
\downarrow \quad 2 \\
\downarrow \quad 6 \\
\downarrow \\
3 \\
\downarrow \\
4 \\
\downarrow \\
7 \\
\end{array}
\]

After having executed lines 5-8 for \(i = 1\), \(G\) looks like the CG below. Note that according to line 6, \(K_j = \{6, 7\}\).

\[
\begin{array}{c}
5 \\
\downarrow \\
1 \quad 2 \quad 6 \\
\downarrow \\
3 \quad 4 \quad 7 \\
\end{array}
\]

We show below an example run of the operation Fbmerge\((L, R, G)\) presented in Figure 1 of the main text. Let \(G\) be the CG below, and \(L = \{1, 2, 6\}\), and \(R = \{3, 4, 7\}\).

\[
\begin{array}{c}
1 \\
\downarrow \quad 2 \\
\downarrow \quad 6 \\
\downarrow \\
3 \\
\downarrow \\
4 \\
\downarrow \\
7 \\
\end{array}
\]

According to line 1, \(R_1 = \{3, 4\}\) and \(R_2 = \{7\}\). After having executed lines 2-4 for \(i = 1\), \(G\) looks like the CG below.

\[
\begin{array}{c}
5 \\
\downarrow \\
1 \quad 2 \quad 6 \\
\downarrow \\
3 \quad 4 \quad 7 \\
\end{array}
\]

After having executed lines 5-8 for \(i = 1\), \(G\) looks like the CG below. Note that according to line 6, \(K_j = \{1, 2, 3, 4\}\).

\[
\begin{array}{c}
5 \\
\downarrow \\
1 \quad 2 \quad 6 \\
\downarrow \\
3 \\
\downarrow \\
4 \\
\downarrow \\
7 \\
\end{array}
\]

After having executed lines 2-4 for \(i = 2\), \(G\) looks like the CG below.
After having executed lines 5-8 for $i = 1$, $G$ looks like the CG below. Note that according to line 6, $L_j = \{1, 2\}$.

We show below an example run of the operation Construct $\beta(G, \alpha, \beta)$ presented in Figure 2 of the main text. Let $G$ be the CG below and $\alpha = \{(1), \{5, 6\}, \{3, 4\}, \{2\})$.

Initially, $H = G$. In the first iteration of the algorithm, $C = \{5, 6\}, \beta = \{(5, 6)\}$, and the nodes $\{5, 6\}$ get removed from $H$. The algorithm jumps to line 3. In the second iteration, $C = \{3, 4\}, \beta = \{(3, 4), \{5, 6\}\}$, and the nodes $\{3, 4\}$ get removed from $H$. The algorithm jumps to line 3. In the third iteration, $C = \{1, 2\}, \beta = \{(1, 2), \{3, 4\}, \{5, 6\}\}$, and the nodes $\{1, 2\}$ get removed from $H$. The algorithm halts because $H = \emptyset$.

We finally show below an example run of the algorithm Method B3($G, \alpha, \beta$) presented in Figure 2 of the main text. Let $G$ be the CG above and $\alpha = \{(1), \{5, 6\}, \{3, 4\}, \{2\})$. As shown above, $\beta = \{(1, 2), \{3, 4\}, \{5, 6\}\}$ after having executed line 1. In the first iteration of the algorithm, $C = \{2\}, K = \{1, 2\}$, $G$ gets modified into the CG below by line 6 with $L = \{2\}$

and $\beta = \{(1), \{2\}, \{3, 4\}, \{5, 6\}\}$ by line 7, $G$ does not get modified by line 10 but $\beta = \{(1), \{2, 3, 4\}, \{5, 6\}\}$ by line 11. The algorithm jumps to line 3. In the second iteration, $C = \{2\}, K = \{2, 3, 4\}, G$ does not get modified by line 6 but $\beta = \{(1), \{3, 4\}, \{2\}, \{5, 6\}\}$ by line 7, $G$ does not get modified by line 10 but $\beta = \{(1), \{3, 4\}, \{2, 5, 6\}\}$ by line 11. The algorithm jumps to line 3. In the third iteration, $C = \{2\}, K = \{2, 5, 6\}, G$ does not get modified by line 6 but $\beta = \{(1), \{3, 4\}, \{5, 6\}, \{2\}\}$ by line 7. The algorithm jumps to line 2. In the fourth iteration, $C = \{3, 4\}, K = \{3, 4\}, G$ gets modified into the CG below by line 10 with $L = \{3, 4\}$ and $R = \{5, 6\}$

and $\beta = \{(1), \{3, 4, 5, 6\}, \{2\}\}$ by line 11. The algorithm jumps to line 3. In the fifth iteration, $C = \{3, 4\}, K = \{3, 4, 5, 6\}, G$ gets modified into the CG below by line 6 with $L = \{3, 4\}$

and $\beta = \{(1), \{5, 6\}, \{3, 4\}, \{2\}\}$ by line 7. The algorithm halts because $\beta = \alpha$.

**APPENDIX B: PROOFS**

**Lemma 1.** Let $M$ denote an independence model, and $\alpha$ a chain $C_1, \ldots, C_n$. If $M$ is a graphoid, then there exits a unique CG $G_\alpha$ that is a MI map of $M$ relative to $\alpha$. Specifically, for each node $X$ of each block $C_k$ of $\alpha$, $Bd_{G_\alpha}(X)$ is the smallest subset $B$ of $\cup_{j=1}^{k}C_j \setminus \{X\}$ s.t. $X \perp_{M \cup_{j=1}^{k}C_j \setminus \{X\}} B|B$.

**Proof.** Let $X$ and $Y$ denote any two non-adjacent nodes of $G_\alpha$. Let $k^*$ denote the smallest $k$ s.t. $X, Y \in \cup_{j=1}^{k^*}C_j$. Assume without loss of generality that $X \in C_{k^*}$. Then, $X \perp_{M \cup_{j=1}^{k^*}C_j \setminus \{X\}} Bd_{G_\alpha}(X)|Bd_{G_\alpha}(X)$ by construction of $G_\alpha$ and, thus, $X \perp_{M \cup_{j=1}^{k^*}C_j \setminus \{X, Y\}} B|B$ weak union. Then, $G_\alpha$ satisfies the pairwise block-recursive Markov property w.r.t. $M$ and, thus, $G_\alpha$ is an MI map of $M$. In fact, $G_\alpha$ is a MI map of $M$ by construction of $Bd_{G_\alpha}(X)$.

Assume to the contrary that there exists another CG $H_\alpha$ that is a MI map of $M$ relative to $\alpha$. Let $X$ denote any node s.t. $Bd_{G_\alpha}(X) \neq Bd_{H_\alpha}(X)$. Let $X \in C_k$. Then, $X \perp_{M \cup_{j=1}^{k}C_j \setminus \{X\}} Bd_{G_\alpha}(X)|Bd_{G_\alpha}(X)$ and $X \perp_{M \cup_{j=1}^{k}C_j \setminus \{X\}} Bd_{H_\alpha}(X)|Bd_{H_\alpha}(X)$ because $G_\alpha$ and $H_\alpha$ are MI maps of $M$. Then, $X \perp_{M \cup_{j=1}^{k}C_j \setminus \{X\}}$.

1 By convention, $X \perp_{M \cup_{j=1}^{k}C_j \setminus \{X\}}$.\)
Lemma 2. Let $G$ and $H$ denote two CGs s.t. $I(H) \subseteq I(G)$. For any component $C$ of $G$, there exists a unique component of $H$ that is maximal in $H$ from the set of components of $H$ that contain a descendant of $C$ in $G$.

Proof. By definition of CG, there exists at least one such component of $H$. Assume to the contrary that there exist two such components of $H$, say $K$ and $K'$. Note that $Pa_H(K) \cap K' = \emptyset$ and $Pa_H(K') \cap K = \emptyset$ by definition of $K$ and $K'$. Note also that no node of $K$ or $Pa_H(K)$ is a descendant of $K'$ in $H$ by definition of $K$. This implies that $K' \perp H \cup Pa_H(K) \setminus Pa_H(K')$ and, thus, $K \perp H | Pa_H(K) \cup Pa_H(K')$ by weak union and symmetry.

That $K$ and $K'$ contain some descendants $k$ and $k'$ of $C$ in $G$ implies that there are descending routes from $C$ to $k$ and $k'$ in $G$ s.t. the nodes in the routes are descendant of $C$ in $G$. Thus, there is a route between $k$ and $k'$ in $G$ s.t. the nodes in the route are descendant of $C$ in $G$. Note that no node in this route is in $Pa_H(K) \cup Pa_H(K')$ by definition of $K$ and $K'$. Then, $K \not \perp_G K' | Pa_H(K) \cup Pa_H(K')$. However, this contradicts the fact that $I(H) \subseteq I(G)$ because, as shown, $K \perp H | Pa_H(K) \cup Pa_H(K')$.

Lemma 3. Let $G$ and $H$ denote two CGs s.t. $I(H) \subseteq I(G)$. Let $\alpha$ denote a chain that is consistent with $H$. If no descendant of a node $X$ in $G$ is to the left of $X$ in $\alpha$, then the descendants of $X$ in $G$ are descendant of $X$ in $H$ too.

Proof. Let $D$ denote the descendants of $X$ in $G$. Let $C$ denote the component of $G$ that contains $X$. Note that the descendants of $C$ in $G$ are exactly the set $D$. Then, there exists a unique component of $H$ that is maximal in $H$ from the set of components of $H$ that contain a node from $D$, by Lemma 2.

Let $K$ denote the component of $H$ that contains $X$. Note that $K$ is a component of $H$ that is maximal in $H$ from the set of components of $H$ that contain a node from $D$, since no node of $D$ is to the left of $X$ in $\alpha$. It follows from the paragraph above that $K$ is the only such component of $H$.

Lemma 4. Let $G_\alpha$ denote the MI map of the independence model induced by a CG $G$ relative to a chain $\alpha$. Then, Method B3($G_\alpha)$ returns $G_\alpha$.

Proof. We start by proving that Method B3 halts at some point. When Method B3 is done with the rightmost block of $\alpha$, the rightmost block of $\beta$ contains all and only the nodes of the rightmost block of $\alpha$. When Method B3 is done with the second rightmost block of $\alpha$, the rightmost block of $\beta$ contains all and only the nodes of the rightmost block of $\alpha$, whereas the second rightmost block of $\beta$ contains all and only the nodes of the second rightmost block of $\alpha$. Continuing with this reasoning, one can see that when Method B3 is done with all the blocks of $\alpha$, $\beta$ coincides with $\alpha$ and thus Method B3 halts.

That Method B3 halts at some point implies that it performs a finite sequence of $m$ modifications to $G$ due to the fbsplit and fbmerging in lines 6 and 10. Let $G_t$ denote the CG resulting from the first $t$ modifications to $G$, and let $G_0 = G$. Specifically, Method B3 constructs $G_{t+1}$ from $G_t$ by either

- adding an edge $X \rightarrow Y$ due to line 3 of Fbsplit or Fbmerge,
- adding an edge $X \rightarrow Y$ due to line 4 of Fbsplit or Fbmerge,
- performing all the component splits due to lines 5-8 of Fbsplit, or
- performing all the component mergings due to lines 5-8 of Fbmerge.

Note that none of the modifications above introduces new separation statements. This is trivial to see for the first and second modification. To see it for the third and fourth modification, recall that the splits and the mergings are part of a fbsplit and a fbmerging respectively and, thus, they are feasible. Therefore, $I(G_{t+1}) \subseteq I(G_t)$ for all $0 \leq t < m$ and, thus, $I(G_m) \subseteq I(G_0)$.

We continue by proving that $G_t$ is consistent with $\beta$ for all $0 \leq t < m$. Since this is true for $G_0$ due to line 1, it suffices to prove that if it is true for $G_t$, then it is true for $G_{t+1}$ for all $0 \leq t < m$. We consider the following four cases.

Case 1 Method B3 constructs $G_{t+1}$ from $G_t$ by adding an edge $X \rightarrow Y$ due to line 3 of Fbsplit or Fbmerge. It suffices to note that $X$ and $Y$ are in the same block of $G_t$ and $\beta$. 

Case 2 Method B3 constructs $G_{t+1}$ from $G_t$ by adding an edge $X \rightarrow Y$ due to line 4 of Fbsplit.
It suffices to note that $X$ is to the left of $Y$ in $\beta$, because $G_t$ is consistent with $\beta$.

**Case 3** Method $B3$ constructs $G_{t+1}$ from $G_t$ by adding an edge $X \rightarrow Y$ due to line 4 of Fbmerge. Note that $X$ is to the left of $R$ in $\beta$, because $\beta$ is consistent with $G_t$. Then, $X$ is to the left of $L$ in $\beta$, because $L$ is the left neighbor of $R$ in $\beta$ and $X \notin L$. Then, $X$ is to the left of $Y$ in $\beta$, because $Y \in L$.

**Case 4** Method $B3$ constructs $G_{t+1}$ from $G_t$ by either performing all the component splits due to lines 5-8 of Fbsplit or performing all the component mergings due to lines 5-8 of Fbmerge. Note that the splits and the mergings are feasible, since they are part of a fsplit and a fbmerging respectively. Therefore, $G_{t+1}$ is a CG. Moreover, note that $\beta$ is modified immediately after the fsplit and the fbmerging so that it is consistent with $G_{t+1}$.

Note that $G_m$ is not only consistent with $\beta$ but also with $\alpha$ because, as shown, $\beta$ coincides with $\alpha$ when Method $B3$ halts. In order to prove the lemma, i.e. that $G_m = G_\alpha$, all that remains to prove is that $I(G_\alpha) \subseteq I(G_m)$. To see it, note that $G_m = G_\alpha$ follows from $I(G_\alpha) \subseteq I(G_m)$, $I(G_m) \subseteq I(G_\alpha)$, the fact that $G_m$ is consistent with $\alpha$, and the fact that $G_\alpha$ is the unique MI map of $I(G_\alpha)$ relative to $\alpha$. Recall that $G_\alpha$ is guaranteed to be unique by Lemma 1, because $I(G_\alpha)$ is a graphoid.

The rest of the proof is devoted to prove that $I(G_\alpha) \subseteq I(G_m)$. Specifically, we prove that if $I(G_\alpha) \subseteq I(G_t)$ then $I(G_\alpha) \subseteq I(G_{t+1})$ for all $0 \leq t < m$. Note that this implies that $I(G_\alpha) \subseteq I(G_m)$ because $I(G_\alpha) \subseteq I(G_0)$ by definition of MI map. First, we prove it when Method $B3$ constructs $G_{t+1}$ from $G_t$ by either performing all the component splits due to lines 5-8 of Fbsplit or performing all the component mergings due to lines 5-8 of Fbmerge. Note that the splits and the mergings are feasible, since they are part of a fsplit and a fbmerging respectively. Therefore, $I(G_{t+1}) = I(G_t)$. Thus, $I(G_\alpha) \subseteq I(G_{t+1})$ because $I(G_\alpha) \subseteq I(G_t)$.

Now, we prove that if $I(G_\alpha) \subseteq I(G_t)$ then $I(G_\alpha) \subseteq I(G_{t+1})$ when Method $B3$ constructs $G_{t+1}$ from $G_t$ by adding a directed or undirected edge due to lines 3 and 4 of Fbsplit and Fbmerge. Specifically, we prove that if there is an $S$-active route $\rho^{AB}_{t+1}$ between two nodes $A$ and $B$ in $G_{t+1}$, then there is an $S$-active route between $A$ and $B$ in $G_{t+1}$. We prove this result by induction on the number of occurrences of the added edge in $\rho^{AB}_{t+1}$.

We assume without loss of generality that the added edge occurs in $\rho^{AB}_{t+1}$ as few or fewer times than in any other $S$-active route between $A$ and $B$ in $G_{t+1}$. We call this the minimality property of $\rho^{AB}_{t+1}$. If the number of occurrences of the added edge in $\rho^{AB}_{t+1}$ is zero, then $\rho^{AB}_{t+1}$ is an $S$-active route between $A$ and $B$ in $G_t$ too and, thus, there is an $S$-active route between $A$ and $B$ in $G_a$ since $I(G_a) \subseteq I(G_t)$. Assume as induction hypothesis that the result holds for up to $n$ occurrences of the added edge in $\rho^{AB}_{t+1}$. We now prove it for $n+1$ occurrences. We consider the following four cases.

**Case 1** Method $B3$ constructs $G_{t+1}$ from $G_t$ by adding an edge $X - Y$ due to line 3 of Fbsplit. Note that $X - Y$ occurs in $\rho^{AB}_{t+1}$. Assume that $X - Y$ occurs in a collider section of $\rho^{AB}_{t+1}$. Note that $X$ and $Y$ must be in the same component of $G_t$ for line 3 of Fbsplit to add an edge $X - Y$. This component also contains a node $Z$ that is in $S$ because, otherwise, $\rho^{AB}_{t+1}$ would not be $S$-active in $G_{t+1}$. Note that there is a route $X - ... - Z - ... - Y$ in $G_t$. Then, we can replace any occurrence of $X - Y$ in a collider section of $\rho^{AB}_{t+1}$ with $X - ... - Z - ... - Y$, and thus construct an $S$-active route between $A$ and $B$ in $G_{t+1}$ that violates the minimality property of $\rho^{AB}_{t+1}$. Since this is a contradiction, $X - Y$ only occurs in non-collider sections of $\rho^{AB}_{t+1}$. Let $\rho^{AB}_{t+1} = \rho^{AX}_{t+1} \cup X - Y \cup \rho^{AB}_{t+1}$. Note that $X, Y \notin S$ because, otherwise, $\rho^{AB}_{t+1}$ would not be $S$-active in $G_{t+1}$. For the same reason, $\rho^{AX}_{t+1}$ and $\rho^{AB}_{t+1}$ are $S$-active in $G_{t+1}$. Then, there are $S$-active routes $\rho^{AX}_{t+1}$ and $\rho^{AB}_{t+1}$ between $A$ and $X$ and between $Y$ and $B$ in $G_{t+1}$ by the induction hypothesis.

Let $X - X' - ... - Y' - Y$ be a route in $G_t$ s.t. the nodes in $X' - ... - Y'$ are in $L$. Such a route must exist for line 3 of Fbsplit to add an edge $X - Y$. Note that $X$ and $X'$ are adjacent in $G_\alpha$ since $I(G_\alpha) \subseteq I(G_t)$. In fact, $X \rightarrow X'$ is in $G_\alpha$. To see it, recall that Method $B3$ is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$ in $\alpha$. Then, $K$ only contains nodes from $C$ or from blocks to the left of $C$ in $\alpha$. However, $X \notin C$ because $X \in K \setminus L$ and $L = K \cap C$. Then, $X$ is to the left of $C$ in $\alpha$. Thus, $X \rightarrow X'$ is in $G_\alpha$ because $X' \in L \subseteq C$. Likewise, $Y \rightarrow Y'$ is in $G_\alpha$. Note also that $X' - ... - Y'$ is in $G_\alpha$. To see it, note that the adjacencies in $X' - ... - Y'$ are preserved in $G_a$ since $I(G_a) \subseteq I(G_t)$. Moreover, these adjacencies correspond to undirected edges in $G_a$, because the nodes in $X' - ... - Y'$ are in $L$ and thus in the same block of $G_a$, since $L \subseteq C$. Furthermore, a node in $X' - ... - Y'$ is in $S$ because, otherwise, $\rho^{AX}_{t+1} \cup X - X' - ... - Y' - Y \cup \rho^{YB}_{t+1}$
would be an $S$-active route between $A$ and $B$ in $G_{t+1}$ that would violate the minimality property of $\rho^A_{t+1}$. Then, $\rho^A_{t+1} \cup X \rightarrow X' - \ldots - Y' \leftarrow Y \cup \rho^B_{t+1}$ is an $S$-active route between $A$ and $B$ in $G_\alpha$.

**Case 2** Method B3 constructs $G_{t+1}$ from $G_t$ by adding an edge $X \rightarrow Y$ due to line 4 of Fb-split. Note that $X \rightarrow Y$ occurs in $\rho^A_{t+1}$. Assume that $X \rightarrow Y$ occurs as a collider edge in $\rho^A_{t+1}$, i.e. $X \rightarrow Y$ occurs in a subroute of $\rho^A_{t+1}$ of the form $X \rightarrow Y - \ldots - Z \leftarrow W$. Note that a node in $Y - \ldots - Z$ is in $S$ because, otherwise, $\rho^A_{t+1}$ would not be $S$-active in $G_{t+1}$. Let $X \rightarrow X' - \ldots - Y' \rightarrow Y$ be a route in $G_t$ s.t. the nodes in $X' - \ldots - Y'$ are in $L$. Such a route must exist for line 4 of Fb-split to add an edge $X \rightarrow Y$. Then, we can replace $X \rightarrow Y - \ldots - Z \leftarrow W$ with $X \rightarrow X' - \ldots - Y' - Y - \ldots - Z \leftarrow W$ in $\rho^A_{t+1}$, and thus construct an $S$-active route between $A$ and $B$ in $G_{t+1}$ that violates the minimality property of $\rho^A_{t+1}$. Since this is a contradiction, $X \rightarrow Y$ never occurs as a collider edge in $\rho^A_{t+1}$. Let $\rho^A_{t+1} = \rho^A_{t+1} \cup X \rightarrow Y \cup \rho^B_{t+1}$. Note that $X, Y \notin S$ because, otherwise, $\rho^A_{t+1}$ would not be $S$-active in $G_{t+1}$. For the same reason, $\rho^A_{t+1}$ and $\rho^B_{t+1}$ are $S$-active in $G_{t+1}$.

Then, there are $S$-active routes $\rho^A_{t+1}$ and $\rho^B_{t+1}$ between $A$ and $X$ and between $Y$ and $B$ in $G_\alpha$ by the induction hypothesis.

Let $X \rightarrow X' - \ldots - Y' - Y$ denote a route in $G_t$ s.t. the nodes in $X' - \ldots - Y'$ are in $L$. Such a route must exist for line 4 of Fb-split to add an edge $X \rightarrow Y$. Note that $X' - \ldots - Y'$ is in $G_\alpha$. To see it, note that the adjacencies in $X' - \ldots - Y'$ are preserved in $G_\alpha$ since $I(G_\alpha) \subseteq I(G_t)$. Moreover, these adjacencies correspond to undirected edges in $G_\alpha$, because the nodes in $X' - \ldots - Y'$ are in $L$ and thus in the same block of $G_\alpha$, since $L \subseteq C$. Furthermore, a node in $X' - \ldots - Y'$ is in $S$ because, otherwise, $\rho^A_{t+1} \cup X \rightarrow X' - \ldots - Y' - Y \cup \rho^B_{t+1}$ would be an $S$-active route between $A$ and $B$ in $G_{t+1}$ that would violate the minimality property of $\rho^A_{t+1}$. Moreover, note that $X$ and $X'$ are adjacent in $G_\alpha$ since $I(G_\alpha) \subseteq I(G_t)$. In fact, $X \rightarrow X'$ is in $G_\alpha$. To see it, recall that Method B3 is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$ in $\alpha$. Then, no block to the left of $K$ in $\beta$ has a node from $C$ or from a block to the right of $C$ in $\alpha$. Note that $X$ is to the left of $K$ in $\beta$, because $\beta$ is consistent with $G_t$. Thus, $X \rightarrow X'$ is in $G_\alpha$ since $X' \in L \subseteq C$. Likewise, note that $Y'$ and $Y$ are adjacent in $G_\alpha$ since $I(G_\alpha) \subseteq I(G_t)$. In fact, $Y' \leftarrow Y$ is in $G_\alpha$. To see it, note that $K$ only contains nodes from $C$ or from blocks to the left of $C$ in $\alpha$. However, $Y \notin C$ because $Y \in K \setminus L$ and $L = K \cap C$. Then, $Y$ is to the left of $C$ in $\alpha$. Thus, $Y' \leftarrow Y$ is in $G_\alpha$ because $Y' \in L \subseteq C$. Then, $\rho^A_{t+1} \cup X \rightarrow X' - \ldots - Y' \leftarrow Y \cup \rho^B_{t+1}$ is an $S$-active route between $A$ and $B$ in $G_\alpha$.

**Case 3** Method B3 constructs $G_{t+1}$ from $G_t$ by adding an edge $X \rightarrow Y$ due to line 3 of Fbmerge. Note that $X \rightarrow Y$ occurs in $\rho^A_{t+1}$. We consider two cases.

**Case 3.1** Assume that $X \rightarrow Y$ occurs in a collider section of $\rho^A_{t+1}$. Let $\rho^A_{t+1} = \rho^A_{t+1} \cup Z \rightarrow X' - \ldots - X - Y - \ldots - Y' \leftarrow W \cup \rho^B_{t+1}$. Note that $Z, W \notin S$ because, otherwise, $\rho^A_{t+1}$ would not be $S$-active in $G_{t+1}$. For the same reason, $\rho^A_{t+1}$ and $\rho^B_{t+1}$ are $S$-active in $G_{t+1}$.

Then, there are $S$-active routes $\rho^A_{t+1}$ and $\rho^B_{t+1}$ between $A$ and $Z$ and between $W$ and $B$ in $G_\alpha$ by the induction hypothesis.

Let $R_t$ denote the component of $G_t$ in $R$ that Fbmerge is processing when the edge $X \rightarrow Y$ gets added. Recall that Method B3 is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$ in $\alpha$. Then, $R_t$ only contains nodes from $C$ or from blocks to the left of $C$ in $\alpha$. In other words, $R_t \subseteq \bigcup_{j=1}^{k-1} C_j \setminus \{X, Y\}$ where $C_k = C$ (recall that $X, Y \in L \subseteq C$). Therefore, $X \notin G_\alpha \cup \bigcup_{j=1}^{k-1} C_j \setminus \{X, Y\}$ because $X$ and $Y$ must be in $P_{G_\alpha}(R_t)$ for line 3 of Fbmerge to add an edge $X \rightarrow Y$. Then, $X$ and $Y$ are adjacent in $G_\alpha$ because, otherwise, $X \notin G_\alpha \cup \bigcup_{j=1}^{k-1} C_j \setminus \{X, Y\}$ which would contradict that $I(G_\alpha) \subseteq I(G_t)$. In fact, $X \rightarrow Y$ is in $G_\alpha$ because $X$ and $Y$ are in the same block of $\alpha$, since $X, Y \in L \subseteq C$. Note that $X' \rightarrow \ldots - X$ and $Y \rightarrow \ldots - Y'$ are in $G_\alpha$. To see it, note that the adjacencies in $X' - \ldots - X$ and $Y - \ldots - Y'$ are preserved in $G_\alpha$ since $I(G_\alpha) \subseteq I(G_t)$. Moreover, these adjacencies correspond to undirected edges in $G_\alpha$, because the nodes in $X' - \ldots - X$ and $Y - \ldots - Y'$ are in $L \subseteq C$. Then, $X' - \ldots - X$ and $Y - \ldots - Y'$ is in $G_\alpha$. Furthermore, a node in $X' - \ldots - X - Y - \ldots - Y'$ is in $S$ because, otherwise, $\rho^A_{t+1}$ would not be $S$-active in $G_{t+1}$. Note also that $Z$ and $X'$ are adjacent in $G_\alpha$ since

---

5Note that maybe $A = X$ and/or $Y = B$.  
6Note that maybe $Y = Z$ and/or $W = X$.  
7Note that maybe $X' = Y'$.  
8Note that maybe $X' = Y'$.  
9Note that maybe $A = Z$, $X' = X$, $Y' = Y$, $W = Z$ and/or $W = B$. 
Assume that $I(G_\alpha) \subseteq I(G_1)$. In fact, $Z \rightarrow X'$ is in $G_\alpha$. To see it, recall that Method B3 is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$. Then, no block to the left of $L$ in $\beta$ has a node from $C$ or from a block to the right of $C$ in $\alpha$. Note that $Z$ is to the left of $L$ in $\beta$, because $\beta$ is consistent with $G_t$. Thus, $Z \rightarrow X'$ is in $G_\alpha$ since $X' \in L \subseteq C$. Likewise, $Y' \leftarrow W$ is in $G_\alpha$. Then, $\rho^Z \cup Z \rightarrow X' - \ldots - X - Y - \ldots - Y' \leftarrow W \cup \rho^W$ is an $S$-active route between $A$ and $B$ in $G_\alpha$.

**Case 3.2** Assume that $X \rightarrow Y$ occurs in a non-collider section of $\rho_{\alpha+1}^A$. Note that this implies that $G_t$ has a descending route from $X$ to $A$ or to a node in $S$, or from $Y$ to $B$ or to a node in $S$. Assume without loss of generality that $G_t$ has a descending route from $Y$ to $B$ or to a node in $S$.

Let $R_i$ denote the component of $G_t$ in $R$ that Fmerge is processing when the edge $X - Y$ gets added. Let $L_Y$ denote the component of $G_t$ that contains the node $Y$. Let $D$ denote the component of $G_\alpha$ that is maximal in $G_\alpha$ from the set of components of $G_t$ that contain a descendant of $L_Y$ in $G_t$. Recall that $D$ is guaranteed to be unique by Lemma 2, because $I(G_\alpha) \subseteq I(G_1)$. We now show that some $d \in D$ is a descendant of $R_i$ in $G_t$. We consider four cases.

**Case 3.2.1** Assume that $D \cap L_Y \neq \emptyset$. It suffices to consider any $d \in R_i$. To see it, recall that Method B3 is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$ in $\alpha$. Then, $R_i$ only contains nodes from $C$ or from blocks to the left of $C$ in $\alpha$. Thus, $d$ is not to the right of the nodes of $D \cap L_Y$ in $\alpha$, since $L_Y \subseteq L \subseteq C$. Moreover, $d$ is not to the left of the nodes of $D \cap L_Y$ in $\alpha$ because, otherwise, there would be a contradiction with the definition of $D$. Then, $d \in D$.

**Case 3.2.2** Assume that $D \cap L_Y = \emptyset$ and $D \cap R_i \neq \emptyset$. It suffices to consider any $d \in D \cap R_i$.

**Case 3.2.3** Assume that $D \cap L_Y = \emptyset$, $D \cap R_i = \emptyset$, and some $d \in D$ was a descendant of some $r \in R_i$ in $G_0$. Recall that Method B3 is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$ in $\alpha$. Then, $R_i$ only contains nodes from $C$ or from blocks to the left of $C$ in $\alpha$. Then, $r$ was not in the blocks of $\alpha$ previously considered, since $r \in R_i$. Therefore, no descendant of $r$ in $G_0$ is currently to the left of $r$ in $\beta$ and, thus, the descendants of $r$ in $G_0$ are descendant of $r$ in $G_t$ by Lemma 3, because $I(G_1) \subseteq I(G_0)$ and $\beta$ is consistent with $G_t$. Then, $d$ is a descendant of $r$ and thus of $R_i$ in $G_t$.

**Case 3.2.4** Assume that $D \cap L_Y = \emptyset$, $D \cap R_i = \emptyset$, and no node of $D$ was a descendant of a node of $R_i$ in $G_0$. As shown in Case 3.2.3, the descendants of any node $r \in R_i$ in $G_0$ are descendant of $r$ in $G_t$ too. Therefore, no descendant of $r$ in $G_0$ was to the left of the nodes of $D$ in $\alpha$ because, otherwise, a descendant of $r$ and thus of $L_Y$ in $G_t$ would be to the left of the nodes of $D$ in $\alpha$, which would contradict the definition of $D$. Recall that no descendant of $r$ in $G_0$ was in $D$ either. Note also that the nodes of $D$ are to the left of the nodes of $R_i$ in $\alpha$, by definition of $D$ and the fact that $D \cap R_i = \emptyset$. These observations have two consequences. First, the components of $G$ containing a node from $D$ were still in $H$ when any component of $G$ containing a node from $R_i$ became a terminal component of $H$ in Construct $\beta$. Thus, Construct $\beta$ added the components of $G$ containing a node from $D$ to $\beta$ after having added the components of $G$ containing a node from $R_i$. Second, Construct $\beta$ did not interchange in $\beta$ any component of $G$ containing a node from $D$ with any component of $G$ containing a node from $R_i$. Recall that Method B3 is currently considering the block $C$ of $\alpha$, and that it has previously considered all the blocks of $\alpha$ to the right of $C$ in $\alpha$. Note that the nodes of $D$ were not in the blocks of $\alpha$ previously considered because, otherwise, $C$ and thus the nodes of $L_Y$ (recall that $L_Y \subseteq L \subseteq C$) would be to the left of $D$ in $\alpha$, which would contradict the definition of $D$. Therefore, the nodes of $D$ are currently still to the left of $R_i$ in $\beta$. Note that the only component to the left of $R_i$ in $\beta$ that contains a descendant of $L_Y$ in $G_t$ is precisely $L_Y$, because $L$ is the left neighbor of $R$ in $\beta$, $L_Y \subseteq L$, and $\beta$ is consistent with $G_t$. However, $D \cap L_Y = \emptyset$. Thus, $D$ contains no descendant of $L_Y$ in $G_t$, which contradicts the definition of $D$. Thus, this case never occurs.

We continue with the proof of Case 3.2. Let
\[ \rho_{t+1}^{AB} = \rho_{t+1}^{AX} \cup X - Y \cup \rho_{t+1}^{YB} \]  

Note that if \( X, Y \notin S \) then, otherwise, \( \rho_{t+1}^{AB} \) would not be S-active in \( G_{t+1} \). For the same reason, \( \rho_{t+1}^{AX} \) and \( \rho_{t+1}^{YB} \) are S-active in \( G_{t+1} \). Note that X and Y must be in \( PA_{G_{t}}(R_{t}) \) for line 3 of Fmerge to add an edge \( X \rightarrow Y \). Then, no descendant of \( R_{t} \) in \( G_{t} \) is in \( S \) because, otherwise, there would be an S-active route \( \rho_{t}^{XY} \) between \( X \) and \( Y \) in \( G_{t} \) and, thus, \( \rho_{t}^{AX} \cup \rho_{t}^{YB} \cup \rho_{t}^{YB} \) would be an S-active route between \( A \) and \( B \) in \( G_{t+1} \) that would violate the minimality property of \( \rho_{t}^{AB} \). Then, there is an S-active descending route \( \rho_{t}^{Ed} \) from some \( r \in R_{t} \) to some \( d \in D \) in \( G_{t} \) because, as shown, \( D \) contains a descendant of \( R_{t} \) in \( G_{t} \). Then, \( \rho_{t}^{AX} \cup X \rightarrow X' \ldots \rightarrow r \cup \rho_{t}^{Ed} \) is an S-active route between \( A \) and \( d \) in \( G_{t+1} \). \(^{11}\) Likewise, \( \rho_{t}^{YB} \cup Y \rightarrow Y' \ldots \rightarrow r \cup \rho_{t}^{Ed} \) is an S-active route between \( B \) and \( d \) in \( G_{t+1} \), where \( \rho_{t}^{YB} \) denotes the route resulting from reversing \( \rho_{t}^{YB} \). \(^{12}\) Therefore, there are S-active routes \( \rho_{t}^{Ad} \) and \( \rho_{t}^{Bd} \) between \( A \) and \( d \) and between \( B \) and \( d \) in \( G_{a} \) by the induction hypothesis.

Recall that we assumed without loss of generality that \( G_{t} \) has a descending route from \( Y \) to a node \( E \) s.t. \( E = B \) or \( E \in S \). Note that \( E \) is a descendant of \( L_{Y} \) in \( G_{1} \) and, thus, \( E \) is a descendant of \( d \) in \( G_{a} \) by definition of \( D \) and the fact that \( d \in D \). Let \( \rho_{t}^{Ed} \) denote the descending route from \( d \) to \( E \) in \( G_{a} \). Assume without loss of generality that \( G_{a} \) has no descending route from \( d \) to \( B \) or to a node of \( S \) that is shorter than \( \rho_{t}^{Ed} \). We now consider two cases.

**Case 3.2.5** Assume that \( E = B \). Note that \( \rho_{t}^{Ed} \) is S-active in \( G_{a} \) by definition and the fact that \( d \notin S \). To see the latter, recall that no descendant of \( R_{t} \) in \( G_{t} \) (among which is \( d \)) is in \( S \). Thus, \( \rho_{t}^{Ad} \cup \rho_{t}^{Ed} \) is an S-active route between \( A \) and \( B \) in \( G_{a} \).

**Case 3.2.6** Assume that \( E \in S \). Let \( \rho_{t}^{Ed} \) and \( \rho_{t}^{Ed} \) denote the routes resulting from reversing \( \rho_{t}^{Ed} \) and \( \rho_{t}^{Ed} \). Consider the route \( \rho_{t}^{Ad} \cup \rho_{t}^{Ed} \) between \( A \) and \( B \) in \( G_{a} \). If this route is S-active, then we are done. If it is not S-active in \( G_{a} \), then \( d \) occurs in a collider section of \( \rho_{t}^{Ad} \cup \rho_{t}^{Ed} \) that has no node in \( S \). Then, we can replace each such occurrence of \( d \) with \( \rho_{t}^{Ed} \cup \rho_{t}^{Ed} \) and, thus construct an S-active route between \( A \) and \( B \) in \( G_{a} \).

---

\(^{10}\)Note that maybe \( A = X \) and/or \( Y = B \).

\(^{11}\)Note that maybe \( X' = r \).

\(^{12}\)Note that maybe \( Y' = r \).

**Case 4** Method B3 constructs \( G_{t+1} \) from \( G_{t} \) by adding an edge \( X \rightarrow Y \) due to line 4 of Fmerge. Note that \( X \rightarrow Y \) occurs in \( \rho_{t+1}^{AB} \). We consider two cases.

**Case 4.1** Assume that \( X \rightarrow Y \) occurs as a collider edge in \( \rho_{t+1}^{AB} \). Let \( \rho_{t+1}^{AB} = \rho_{t+1}^{AX} \cup X \rightarrow Y \cup \rho_{t+1}^{YB} \). \(^{13}\) Note that \( X \notin S \) because, otherwise, \( \rho_{t+1}^{AB} \) would not be S-active in \( G_{t+1} \). For the same reason, \( \rho_{t+1}^{AX} \) is S-active in \( G_{t+1} \). Then, there is an S-active route \( \rho_{t+1}^{AX} \) between \( A \) and \( X \) in \( G_{a} \) by the induction hypothesis. Let \( R_{i} \) denote the component of \( G_{i} \) in \( R \) that Fmerge is processing when the edge \( X \rightarrow Y \) gets added. Recall that Method B3 is currently considering the block \( C \) of \( \alpha \), and that it has previously considered all the blocks of \( \alpha \) to the right of \( C \) in \( \alpha \). Then, \( R_{i} \) only contains nodes from \( C \) or from blocks to the left of \( C \) in \( \alpha \). In other words, \( R_{i} \subseteq \cup_{j=1}^{k} C_{j} \setminus \{X, Y\} \) where \( k^{*} \) is the smallest \( k \) s.t. \( X, Y \in \cup_{j=1}^{k} C_{j} \) (recall that \( Y \in L \subseteq C \)). Therefore, \( X \not\subseteq \cup_{j=1}^{k} C_{j} \cup \{X, Y\} \) because \( X \) and \( Y \) must be in \( PA_{G_{t}}(R_{t}) \) for line 4 of Fmerge to add an edge \( X \rightarrow Y \). Then, \( X \) and \( Y \) are adjacent in \( G_{a} \). Note that \( X \) and \( Y \) are adjacent in \( G_{a} \) because, otherwise, \( X \not\subseteq \cup_{j=1}^{k} C_{j} \cup \{X, Y\} \) which would contradict that \( I(G_{a}) \subseteq I(G_{1}) \). In fact, \( X \rightarrow Y \) is in \( G_{a} \). To see it, recall that Method B3 is currently considering the block \( C \) of \( \alpha \), and that it has previously considered all the blocks of \( \alpha \) to the right of \( C \) in \( \alpha \). Then, no block to the left of \( L \) in \( \beta \) has a node from \( C \) or from a block to the right of \( C \) in \( \alpha \). Note that \( X \) is to the left of \( R \) in \( \beta \), because \( \beta \) is consistent with \( G_{1} \). Then, \( X \) is to the left of \( L \) in \( \beta \), because \( L \) is the left neighbor of \( R \) in \( \beta \) and \( X \notin L \). Thus, \( X \rightarrow Y \) is in \( G_{a} \) because \( Y \in L \subseteq C \). We now consider two cases.

**Case 4.1.1** Assume that \( \rho_{t+1}^{YB} = Y \ldots \rightarrow Y \not\subseteq X \cup \rho_{t+1}^{X} \). Note that a node in \( Y \ldots \rightarrow Y \not\subseteq \cup_{j=1}^{k} C_{j} \cup \{X, Y\} \) which would contradict that \( I(G_{a}) \subseteq I(G_{1}) \). Moreover, these adjacencies correspond to undirected edges in \( G_{a} \) because the nodes in \( Y \ldots \rightarrow Y \) are in \( L \) since \( Y \in L \) and, thus, they are in the same block of \( G_{a} \) since \( L \subseteq C \). Then,

---

\(^{13}\)Note that maybe \( A = X \) and/or \( Y = B \).
we can obtain a MI map of $H$ of $I$.

Note that $I$ denotes a chain that is consistent with $H$. Let $G_\alpha$ denote the MI map of $I(G)$ relative to $\alpha$. Recall that $G_\alpha$ is guaranteed to be unique by Lemma 1, because $I(G)$ is a graphoid. Note that $I(H) \subseteq I(G)$ implies that $G_\alpha$ is a subgraph of $H$. To see it, note that $I(H) \subseteq I(G)$ implies that we can obtain a MI map of $I(G)$ relative to $\alpha$ by just removing edges from $H$. However, $G_\alpha$ is the only MI map of $I(G)$ relative to $\alpha$.

Then, it follows from the proof of Lemma 4 that line 2 transforms $G$ into $G_\alpha$ by a sequence of directed and undirected edge additions and feasible splits and mergings s.t. after each operation in the sequence $G$ is a CG and $I(G_\alpha) \subseteq I(G)$.

\begin{proof}

Note from line 1 that $\alpha$ denotes a chain that is consistent with $H$. Let $G_\alpha$ denote the MI map of $I(G)$ relative to $\alpha$. Recall that $G_\alpha$ is guaranteed to be unique by Lemma 1, because $I(G)$ is a graphoid. Note that $I(H) \subseteq I(G)$ implies that $G_\alpha$ is a subgraph of $H$. To see it, note that $I(H) \subseteq I(G)$ implies that we can obtain a MI map of $I(G)$ relative to $\alpha$ by just removing edges from $H$. However, $G_\alpha$ is the only MI map of $I(G)$ relative to $\alpha$.

Then, it follows from the proof of Lemma 4 that line 2 transforms $G$ into $G_\alpha$ by a sequence of directed and undirected edge additions and feasible splits and mergings, and that after each operation in the sequence $G$ is a CG and $I(G_\alpha) \subseteq I(G)$. Thus, after each operation in the sequence $I(H) \subseteq I(G)$ because $I(H) \subseteq I(G_\alpha)$ since, as shown, $G_\alpha$ is a subgraph of $H$. Finally, line 3 transforms $G$ from $G_\alpha$ to $H$ by a sequence of edge additions. Of course, after each edge addition $G$ is a CG and $I(H) \subseteq I(G)$ because $G_\alpha$ is a subgraph of $H$.
\end{proof}

\begin{theo}

For any probability distribution $p$ for which the composition property holds, the CKES algorithm finds a CG that is inclusion optimal w.r.t. $p$.
\end{theo}

\begin{proof}

Let $R, S$ and $T$ be three random variables. Let $H(R|S)$ denote the conditional entropy of $R$ given $S$. Then, $H(R|S, T) \leq H(R|S)$ and, moreover, $H(R|S, T) = H(R|S)$ if and only if $R \perp T|S$ (Cover and Thomas, 1991, Chapter 2). Therefore, $H(Y|\text{Bd}_G(Y))$ stays the same when removing the edge $X \to Y$ from $G$ in line 4. Likewise, $H(X|\text{Bd}_G(X))$ and $H(Y|\text{Bd}_G(Y))$ stay the same when removing $X \to Y$ from $G$ in line 5. On the other hand, $H(Y|\text{Bd}_G(Y))$ decreases when adding the edge $X \to Y$ to $G$ in line 6. Likewise, $H(X|\text{Bd}_G(X))$ or $H(Y|\text{Bd}_G(Y))$ decreases when adding the edge $X \to Y$ to $G$ in line 7. Let us define the score of a CG $G$ as $\sum_{X \subseteq Y} H(X|\text{Bd}_G(X))$. Now, note that the algorithm cannot enter an endless loop, i.e. it cannot perform a sequence of edge additions and removals so that the CGs before and after the sequence coincide. To see it, assume the contrary and note that such a sequence must contain both edge additions and removals. However, such a sequence would imply that the score of the CG after the sequence is smaller than the score of the CG before the sequence, which is a contradiction. If the algorithm cannot enter an endless loop, then the algorithm must reach a CG $G$ such that no edge can be added or removed from it in lines 3-7, because the number of CGs is finite. Moreover, $I(G) \subseteq I(p)$. To see it, assume the contrary. Then, according to the local Markov property and the composition property, there must exist two nodes $X$ and $Y$ such that $X \notin \text{Sd}_G(Y) \cup \text{Bd}_G(Y)$, $X \notin \text{Bd}_G(Y)$ but $X \in Y|\text{Bd}_G(Y)$ and $Y \notin \text{Sd}_G(Y) \cup \text{Bd}_G(Y)$. Then, a new edge can be added to $G$ in line 6 or 7, which is a contradiction. Obviously, $I(G) \subseteq I(p)$ still holds after $G$ has been updated in line 8. Therefore, the next execution of lines 3-7 may remove edges from $G$ but it never adds edges to $G$. Consequently, the algorithm must reach a CG $G$ such that no edge can be removed from it, because the number of edges that can be removed is finite. At this point, the algorithm executes line 8 repeatedly. This implies that the algorithm terminates at some point, because any CG can be transformed in any other equivalent CG via a sequence of feasible
merges and splits (Studený et al., 2009, Corollary 7).

Finally, we show that there is no model $G'$ such that $I(G) \subset I(G') \subset I(p)$ when the algorithm has terminated in $G$. Assume the contrary. Theorem 1 says that there must exist a model $H$ such that $I(G) = I(F) \subset I(H) \subset \ldots \subset I(G') \subset I(p)$ and $H$ can be reached by a single edge removal from a CG $F$ which is in the equivalence class of $G$.

However this edge must be between two nodes $X$ and $Y$ in $F$ such that $X \in Bd_F(Y)$ and $X \perp_P Y | Bd_F(Y) \setminus X$ because $I(H) \subset I(p)$. This now contradicts the assumption since we know that the algorithm only can terminate if there exists no such edge for any CG in the equivalence class of $G$. Hence, we have a contradiction.

\[\square\]

APPENDIX C: ADDITIONAL EXPERIMENT

In this experiment, we want to show how the algorithms handle an unfaithful probability distribution that satisfies the composition property. Chickering and Meek (2002) performed a similar experiment and we used the same setup. Specifically, we used the $w$-structure seen in Figure 1a as CG structure. From this structure, 10 Gaussian probability distributions were generated and sampled into sample sets in the same way as we did for the experiment in our paper. We then removed the $H$ variable from the sample sets. This meant that the algorithms tried to model the (in)dependencies in Figure 1a with only $A$, $B$, $C$ and $D$ as nodes. It can be shown that this is impossible (i.e. that there exists no faithful CG for this probability distribution) and that the inclusion optimal CGs are those seen in Figure 1b (Chickering and Meek, 2002). As performance measures, we computed the recall of the separations over \{A, B, C, D\} in the 5-node CG sampled, and the precision of the separations in the 4-node CGs learnt. This meant that it was impossible to find a CG with perfect precision and recall, but that the best inclusion optimal CG had precision 1 and recall 0.57. The results obtained are shown in Figure 1. The table show average results over the 10 sample sets in the experiment. The results for categorical data are similar and, thus, they are not included here.

Running the example above by hand with perfect information, it can be seen that the CKES algorithm will end up in one of the CGs seen in Figure 1b. The LCD algorithm on the other hand ends up in one of the CGs shown in Figure 1c, which do not include the original independence model. The experimental results obtained with large sample sets (30000 samples) are shown in Figure 1 and coincide with the theoretical results just described. For the smaller sample sets (100 samples), the same trend observed in the experiment in our paper can be seen here (i.e. the CKES algorithm learning a CG with higher precision and the LCD algorithm learning a CG with higher recall). Therefore, we reach here the same conclusion as we did previously: Since the goal of structural learning is to find a model representing an I map of the probability distribution at hand, the CKES algorithm achieves better results than the LCD algorithm.

Table 1: Average Results For Gaussian Data.

<table>
<thead>
<tr>
<th>CG structure</th>
<th>Recall CKES</th>
<th>Recall LCD</th>
<th>Precision CKES</th>
<th>Precision LCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=100</td>
<td>0.54</td>
<td>0.61</td>
<td>0.75</td>
<td>0.44</td>
</tr>
<tr>
<td>n=30000</td>
<td>0.57</td>
<td>0.57</td>
<td>1.00</td>
<td>0.57</td>
</tr>
</tbody>
</table>

References

Chickering, D. M. and Meek, C. Finding Optimal Bayesian Networks. In Proceedings of the 18th Con-
Appendix of “An inclusion optimal algorithm for chain graph structure learning”

ference on Uncertainty in Artificial Intelligence, 94-102, 2002.
