Transparency in Sum-Product Network Decompilation Supplementary Material

A Rules of do-Calculus

What follows is an essential graphical tool for deciding under what conditions we can reduce interventional queries to conditional ones. Here, $\mathcal{B}_{\overline{x}}$ denotes the graph obtained after deleting all the edges pointing to X, $\mathcal{B}_{\underline{x}}$ the one resulting after deleting all the edges emerging from X, and $\mathcal{B}_{\overline{x}\underline{z}}$ refers to the graph where both edges incoming to X and stemming from Z are deleted.

Definition A.1 ***(Rules of do-Calculus)*** *Let $\mathcal{B}$ be a BN corresponding to a SEM and $\Pr(\cdot)$ the probability measure induced by it. If **X, Y, Z, W** are disjoint sets, then the following hold:*

Rule 1: $\Pr(\boldsymbol{y}|do(\boldsymbol{x}),\boldsymbol{z},\boldsymbol{w}) = \Pr(\boldsymbol{y}|do(\boldsymbol{x}),\boldsymbol{w})$ *if* $(\boldsymbol{Y}\perp\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{W})_{\mathcal{B}_{\overline{X}}}$.

Rule 2: $\Pr(\boldsymbol{y}|do(\boldsymbol{x}),do(\boldsymbol{z}),\boldsymbol{w}) = \Pr(\boldsymbol{y}|do(\boldsymbol{x}),\boldsymbol{z},\boldsymbol{w})$ *if* $(\boldsymbol{Y}\perp\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{W})_{\mathcal{B}_{\overline{X}\underline{Z}}}$.

Rule 3: $\Pr(\boldsymbol{y}|do(\boldsymbol{x}),do(\boldsymbol{z}),\boldsymbol{w}) = \Pr(\boldsymbol{y}|do(\boldsymbol{x}),\boldsymbol{w})$ *if* $(\boldsymbol{Y}\perp\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{W})_{\mathcal{B}_{\overline{X}\overline{Z(W)}}}$, *where **Z(W)** is the set of **Z**-nodes that are not ancestors of any **W**-node in $\mathcal{B}_{\overline{X}}$.*

B Proofs

B.1 Proof of Proposition 4.1

Using the 3rd rule of Pearl's do-calculus, it suffices to show that $(\mathbf{X}\perp\mathbf{U}\cup(\mathbf{V}\setminus\mathbf{X}))_{\mathcal{B}_{\overline{X}}}$. By assumption, no edges emanate from nodes in $\mathbf{X}$, which implies that each of them will be isolated in $\mathcal{B}_{\overline{X}}$, so the desired independence holds, meaning that $\Pr(\mathbf{U},\mathbf{V}_{-\mathbf{X}}|do(\mathbf{X})) = \Pr(\mathbf{U},\mathbf{V}_{-\mathbf{X}})$. In addition, we have that $\Pr(\mathbf{V}_{-\mathbf{X}}|do(\mathbf{X})) = \sum_{\mathbf{U}}\Pr(\mathbf{U},\mathbf{V}_{-\mathbf{X}}|do(\mathbf{X})) = \sum_{\mathbf{U}}\Pr(\mathbf{U},\mathbf{V}_{-\mathbf{X}}) = \Pr(\mathbf{V}_{-\mathbf{X}})$.

B.2 Proof of Theorem 4.1

In [23], the authors explicitly point out that the decompiled BN has a bipartite structure, with edges only pointing from latent to observable variables. In [15], the decompiled BN only allows for edges from latent variables to other latent or observable variables, so no edges point out from observable variables. In [4], the decompilation process (Algorithm 2) introduces edges between nodes in line 23. All edges stem from some **Z**[S'] to some other node (where **Z** is a mapping from SPN sum nodes to latent variables, and S' is a SPN sum node). Moreover, **Z** is assigned (non-trivial) values in lines 13 and 16. In both lines the output of **Z** is defined to be a latent variable, so putting everything together, the decompilation algorithm only allows for edges stemming from latent variables.

In all cases, the decompiled BN has no edges stemming from observable variables, so the result follows by applying Proposition 4.1.

B.3 Proof of Proposition 4.2

Without loss of generality, we can assume that $k = n$, since otherwise we can just swap indices between X_k and X_n. Applying the law of total probability, we get that

$$\Pr(Y|X_1,\cdots,X_{n-1}) = \sum_{X_n}\Pr(Y,X_n|X_1,\cdots,X_{n-1}) =$$
$$\Pr(Y|X_1,\cdots,X_{n-1},X_n=0)\cdot\Pr(X_n=0|X_1,\cdots,X_{n-1})+$$
$$\Pr(Y|X_1,\cdots,X_{n-1},X_n=1)\cdot\Pr(X_n=1|X_1,\cdots,X_{n-1}) =$$
$$\Pr(Y|X_1,\cdots,X_{n-1},X_n=0)\cdot\Pr(X_n=0|X_1,\cdots,X_{n-1})+$$
$$\Pr(Y|X_1,\cdots,X_{n-1},X_n=0)\cdot\Pr(X_n=1|X_1,\cdots,X_{n-1}) =$$
$$\Pr(Y|X_1,\cdots,X_{n-1},X_n=0)\cdot(\Pr(X_n=0|X_1,\cdots,X_{n-1})+$$
$$\Pr(X_n=1|X_1,\cdots,X_{n-1})) = \Pr(Y|X_1,\cdots,X_{n-1},X_n=0)$$

By symmetry, we also have that

$$\Pr(Y|X_1,\cdots,X_{n-1},X_n=1) = \Pr(Y|X_1,\cdots,X_{n-1}),$$

concluding the proof.

B.4 Proof of Theorem 4.2

Let V be a variable represented by some sum node in $\mathcal{SPN}_{\mathcal{C}}$, and $\mathbf{S}$ be the set of all sum nodes representing V, $\mathbf{S} = \{S|S \text{ is a sum node, Represent}(S) = V\}$. Properties 1 and 2 imply that all $S \in \mathbf{S}$ model the conditional distribution of V, given a configuration of its ancestors in $\mathcal{SPN}_{\mathcal{C}}$, $X_1,\cdots,X_k$. Furthermore, both $\mathcal{B}$ and $\mathcal{SPN}_{\mathcal{C}}$ represent the same distribution, which factorizes according to $\mathcal{B}$, so it satisfies the local Markov property. This means that $\Pr(V|X_1,\cdots,X_k) = \Pr(V|P_1,\cdots,P_m)$, where $P_1,\cdots,P_m$ are V's parents in $\mathcal{B}$, since all of these variables are within the set $\{X_1,\cdots,X_k\}$. Finally, Proposition 4.2 is used in order to identify the variables that should be removed from the conditional. This is done in am implicit way, by just comparing certain induced trees, as follows:

- Pick a variable, X_m, that is represented by a sum node that is closer to the root than all nodes in $\mathbf{S}$. Any variable satisfying this property appears before V in the underlying topological ordering.

- Pick a node, S_m, representing X_m, and let $Trees^{S_m} = \{T | T \in Subtrees^S(\mathcal{SPN}_\mathcal{C})$ for some $S \in \mathbf{S}, S_m \in T\}$ be the set of all induced sub-trees that pass through S_m and end in one of the nodes in $\mathbf{S}$
- Let $t_0, t_1 \in Trees^{S_m}$, such that they both contain exactly the same indicators for all variables, except from X_m. Instead, when it comes to X_m, $\mathbb{1}_{X_m=0} \in t_0, \mathbb{1}_{X_m=1} \in t_1$.
- Furthermore, let $S_0, S_1 \in \mathbf{S}$ be the end nodes of t_0, t_1, respectively. Then S_0 models the distribution of $V|\mathbf{x}_{1:k,-m}, X_m = 0$, while S_1 models $V|\mathbf{x}_{1:k,-m}, X_m = 1$, where $\mathbf{x}_{1:k,-m}$ are the states of the variables $X_1, \cdots, X_{m-1}, X_{m+1}, \cdots, X_k$, which are the same in both conditionals, since t_0, t_1 include the same indicators for each of these variables. By construction, the two conditioning sets only differ in the state of X_m.
- If $S_0 \neq S_1$, then by Proposition 4.2 we conclude X_m and V are not conditionally independent (since $\Pr(V|\mathbf{x}_{1:k,-m}, X_m = 0) \neq \Pr(V|\mathbf{x}_{1:k,-m}, X_m = 1)$), so X_m must be a parent of V in $\mathcal{B}$. On the other hand, if $S_0 = S_1$, we cannot reach a definite conclusion, so we consider a new pair of trees $t_0, t_1 \in Trees^{S_m}$ and repeat the same process. If after considering all such trees no two distinct sum nodes can be found, we then consider another node representing X_m and repeat the steps above. Finally, if this loop terminates without identifying two distinct sum nodes, then $\Pr(V|\mathbf{x}_{1:k,-m}, X_m = 0) = \Pr(V|\mathbf{x}_{1:k,-m}, X_m = 1)$ for all values of $\mathbf{x}_{1:k,-m}$, so again by Proposition 4.2 we can conclude that X_m must not be a parent of V in $\mathcal{B}$.