

3 The equivalence between the local and global definitions

We prove next that the local and global definitions of causal Bayesian networks are equivalent. To the best of our knowledge, the proof of equivalence has not been published before.

Theorem 1 (Equivalence between local and global compatibility). *Let G be a DAG and \mathbf{P}_* a set of interventional distributions, the following statements are equivalent:*

- i. G is locally compatible with \mathbf{P}_*
- ii. G is globally compatible with \mathbf{P}_*

Proof. (Definition 3 \Rightarrow Definition 4)

Given an intervention $do(\mathbf{X} = \mathbf{x})$, $\mathbf{X} \subseteq \mathbf{V}$, assume that conditions 3:(i-ii) are satisfied. For any arbitrary instantiation \mathbf{v} of variables \mathbf{V} , consistent with $\mathbf{X} = \mathbf{x}$, we can express $P_{\mathbf{x}}(\mathbf{v})$ as

$$\begin{aligned} P_{\mathbf{x}}(\mathbf{v}) &\stackrel{\text{def.3:(i)}}{=} \prod_i P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) \\ &= \prod_{\{i|v_i \in \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) \prod_{\{i|v_i \notin \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) \\ &\stackrel{\text{effectiveness}}{=} \prod_{\{i|v_i \notin \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) \\ &\stackrel{\text{def.3:(ii)}}{=} \prod_{\{i|v_i \notin \mathbf{X}\}} P(v_i \mid \mathbf{pa}_i) \end{aligned} \quad (3)$$

which is the truncated product as desired.

(Definition 4 \Rightarrow Definition 3)

We assume that the truncated factorization holds, i.e., the distribution $P_{\mathbf{x}}(\mathbf{v})$ resulting from any intervention $do(\mathbf{X} = \mathbf{x})$ can be computed as eq. (2).

To prove effectiveness, consider an intervention $do(\mathbf{X} = \mathbf{x})$, and let $v_i \in \mathbf{X}$. Let $Dom(v_i) = \{v_{i1}, v_{i2}, \dots, v_{im}\}$ be the domain of variable V_i , with only one of those values consistent with $\mathbf{X} = \mathbf{x}$. Since $P_{\mathbf{x}}(\mathbf{v})$ is a probability distribution, we must have $\sum_j P_{\mathbf{x}}(V_i = v_{ij}) = 1$. According to eq. (2), all terms not consistent with $\mathbf{X} = \mathbf{x}$ have probability zero, and thus we obtain $P_{\mathbf{x}}(v_i) = 1$, v_i consistent with $\mathbf{X} = \mathbf{x}$.

To show Definition 3:(ii), we consider an ordering $\pi : (v_1, \dots, v_n)$ of the variables, consistent with the graph G induced by the truncated factorization with no intervention $P(\mathbf{v}) = \prod_i P(v_i \mid \mathbf{pa}_i)$. Now, given an intervention $do(\mathbf{X} = \mathbf{x})$

$$P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) = \frac{P_{\mathbf{x}}(v_i, \mathbf{pa}_i)}{P_{\mathbf{x}}(\mathbf{pa}_i)}$$

$$\begin{aligned} &\stackrel{\text{marginal.}}{=} \frac{\sum_{v_j \notin \{V_i, \mathbf{PA}_i\}} P_{\mathbf{x}}(\mathbf{v})}{\sum_{v_j \notin \{\mathbf{PA}_i\}} P_{\mathbf{x}}(\mathbf{v})} \\ &\stackrel{\text{eq.(2)}}{=} \frac{\sum_{v_j \notin \{V_i, \mathbf{PA}_i, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa}_k)}{\sum_{v_j \notin \{\mathbf{PA}_i, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa}_k)} \\ &= P(v_i \mid \mathbf{pa}_i) \times \frac{\sum_{v_j \notin \{V_i, \mathbf{PA}_i, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}, k \neq i} P(v_k \mid \mathbf{pa}_k)}{\sum_{v_j \notin \{\mathbf{PA}_i, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa}_k)} \end{aligned} \quad (4)$$

The last step is due to the fact that variables in $\{V_i, \mathbf{PA}_i\}$ do not appear in the summations in the numerator. Rewriting the numerator, breaking it in relation to variables before and after v_i , we obtain

$$\begin{aligned} &\sum_{v_j \notin \{V_i, \mathbf{PA}_i, \mathbf{X}\}} \prod_{\substack{v_k \notin \mathbf{X} \\ k \neq i}} P(v_k \mid \mathbf{pa}_k) = \\ &\sum_{\substack{v_j \notin \{\mathbf{PA}_i, \mathbf{X}\} \\ j < i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k < i}} P(v_k \mid \mathbf{pa}_k) \sum_{\substack{v_j \notin \mathbf{X} \\ j > i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k > i}} P(v_k \mid \mathbf{pa}_k) \end{aligned} \quad (5)$$

Note that $\sum_{\substack{v_j \notin \mathbf{X} \\ j > i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k > i}} P(v_k \mid \mathbf{pa}_k) = 1$ because all $V_j > V_i$ appear in the summation. Thus, we obtain

$$\begin{aligned} &\sum_{v_j \notin \{V_i, \mathbf{PA}_i, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa}_k) = \\ &\sum_{\substack{v_j \notin \{\mathbf{PA}_i, \mathbf{X}\} \\ j < i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k < i}} P(v_k \mid \mathbf{pa}_k) \end{aligned} \quad (6)$$

Similarly for the denominator,

$$\begin{aligned} &\sum_{v_j \notin \{\mathbf{PA}_i, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa}_k) = \\ &\sum_{\substack{v_j \notin \{\mathbf{PA}_i, \mathbf{X}\} \\ j < i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k < i}} P(v_k \mid \mathbf{pa}_k) \end{aligned} \quad (7)$$

Observe that eqs. (6) and (7) are identical, equation (4) reduces to $P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) = P(v_i \mid \mathbf{pa}_i)$ as desired.

To show Definition 3:(i), we first use the truncated factorization

$$\begin{aligned} P_{\mathbf{x}}(\mathbf{v}) &\stackrel{\text{eq.(2)}}{=} \prod_{\{i, v_i \notin \mathbf{X}\}} P(v_i \mid \mathbf{pa}_i) \\ &\stackrel{\text{def.3:(ii)}}{=} \prod_{\{i, v_i \notin \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) \\ &\stackrel{\text{effectiveness}}{=} \prod_i P_{\mathbf{x}}(v_i \mid \mathbf{pa}_i) \end{aligned} \quad (8)$$

Finally, def. 3:(i) follows from the definition of Markov compatibility (definition 1.2.2 in [Pearl, 2000]). \square

4 Alternative characterization of Causal Bayesian Networks

We state next a local definition of CBNs which focuses on the absence of edges in the causal graph, i.e., the missing-links representing absence of causal influence.

Definition 5 (Missing-link causal Bayesian network). *A DAG G is said to be missing-link compatible with a set of interventional distributions \mathbf{P}_* if and only if the following conditions hold:*

- i. [Markov] $\forall \mathbf{X} \subseteq \mathbf{V}$, $P_{\mathbf{x}}(\mathbf{v})$ is Markov relative to G ;
- ii. [Missing-link] $\forall \mathbf{X} \subset \mathbf{V}, Y \in \mathbf{V}, \mathbf{S} \subset \mathbf{V}$, $P_{\mathbf{x}, \mathbf{s}, \mathbf{pa}_Y}(y) = P_{\mathbf{s}, \mathbf{pa}_Y}(y)$ whenever there is no arrow from \mathbf{X} to Y in G , \mathbf{pa}_Y is consistent with $\{\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}\}$ and $\mathbf{X}, \{Y\}, \mathbf{S}$ are disjoint.
- iii. [Parents do/see] $\forall Y \in \mathbf{V}, \mathbf{X} \subset \mathbf{V}$, $P_{\mathbf{x}, \mathbf{pa}_Y}(y) = P_{\mathbf{x}}(y | \mathbf{pa}_Y)$ whenever \mathbf{pa}_Y is consistent with $\mathbf{X} = \mathbf{x}$ and $\mathbf{X}, \{Y\}$ are disjoint.

Condition (ii) requires that when we set \mathbf{X} to some value while keeping the variables $\mathbf{S} \cup \mathbf{PA}_Y$ constant, the marginal distribution of Y remains unaltered, independent of the value of \mathbf{X} , whenever there is no edge between \mathbf{X} and Y , i.e., an intervention on \mathbf{X} does not change Y 's distribution while holding constant its parents. In addition to the missing-link condition, 5:(iii) describes the relationship inside each family, i.e., the effect on Y should be the same whether observing (seeing) or intervening (doing) on its parents \mathbf{PA}_Y .

Note that the missing-link condition is not sufficient on its own to characterize causal Bayesian networks – condition 5:(iii) is also necessary when there is a link between any two variables. To illustrate, consider a DAG G with only two binary variables A and B , and an edge from A to B . Without condition 5:(iii), the interventional distribution $P_a(b)$ is unconstrained, which allows $P_a(b) \neq P(b | a)$. However, Definition 3 implies $P_a(b) = P(b | a)$ since A is the only parent of B . Condition 5:(iii) ensures this equality.

To facilitate comparison to previous definitions, we next define the notion of *interventional invariance*:

Definition 6 (Interventional invariance). *We say that Y is interventional invariant (II) to \mathbf{X} given \mathbf{Z} , denoted $(Y \perp_{ii} \mathbf{X} | \mathbf{Z})_{\mathbf{P}_*}$, if intervening on \mathbf{X} does not change the interventional distribution of Y given $do(\mathbf{Z} = \mathbf{z})$, i.e., $\forall \mathbf{x}, y, \mathbf{z}$, $P_{\mathbf{x}, \mathbf{z}}(y) = P_{\mathbf{z}}(y)$.*

Note that definitions 2 and 6 represent different types of causal invariance, the former claims invariance given an observation, while the latter claims invariance given an intervention. Interpreting CBNs in these terms, Definition 3 assumes modularity of each family in terms of conditional invariance (i.e., $(Y \perp_{ci} \mathbf{X} | \mathbf{PA}_Y)_{\mathbf{P}_*}, \forall \mathbf{X}$), while Definition 5 expresses the absence of causal effect in terms of interventional invariance (i.e., $(Y \perp_{ii} \mathbf{X} | \mathbf{PA}_Y, \mathbf{S})_{\mathbf{P}_*}, \forall \mathbf{S}, \mathbf{X}$).

We believe that Definition 5 is more intuitive because it relies exclusively on causal relationships in terms of which the bulk of scientific knowledge is encoded. We further discuss this intuition in the next section.

Note that conditional independence claims encoded by the CBNs are of the form $(Y \perp \mathbf{ND}_Y | \mathbf{PA}_Y)_{\mathbf{P}_*}$, and the II

claims are of the form $(Y \perp_{ii} \mathbf{X} | \mathbf{PA}_Y, \mathbf{S})_{\mathbf{P}_*}, \forall \mathbf{X}, \mathbf{S}$. In both cases, \mathbf{PA}_Y is required to separate Y from other variables. In the observational case Y is separated from its non-descendants, while in the experimental one it is separated from all other variables. This is so because in the experimental case, an intervention on a descendant of a variable Z cannot influence Z (as is easily shown by d-separation in the mutilated graph).

A characterization based on Zero Direct Effect

The missing-link definition requires advance knowledge about parent sets, which is not necessarily available in the network construction. In this section, we extend the previous definition and propose a new characterization based on the notion of *Zero direct effect*, which is more aligned with our intuition about causal relationships, especially these emanating from typical experiments.

Definition 7 (Zero direct effect). *Let $\mathbf{X} \subset \mathbf{V}, Y \in \mathbf{V}$ and $\mathbf{S}_{\mathbf{XY}} = \mathbf{V} - \{X, Y\}$.⁴ We say that \mathbf{X} has zero direct effect on Y , denoted $ZDE(\mathbf{X}, Y)$, if*

$$(Y \perp_{ii} \mathbf{X} | \mathbf{S}_{\mathbf{XY}})$$

Now, we introduce the definition of CBNs motivated by this notion:

Definition 8 (Zero direct effect (ZDE) causal Bayesian network). *A DAG G is ZDE compatible with a set of interventional distributions \mathbf{P}_* if the following conditions hold:*

- i. [Markov] $\forall \mathbf{X} \subseteq \mathbf{V}$, $P_{\mathbf{x}}(\mathbf{v})$ is Markov relative to G ;
- ii. [ZDE] $\forall X, Y \in \mathbf{V}$, $ZDE(X, Y)$ whenever there is no arrow from X to Y in G ;
- iii. [Additivity] $\forall \mathbf{X} \subset \mathbf{V}, Z, Y \in \mathbf{V}$, $ZDE(\mathbf{X}, Y)$ and $ZDE(Z, Y) \Rightarrow ZDE(\mathbf{X} \cup \{Z\}, Y)$;
- iv. [Parents do/see] $\forall Y \in \mathbf{V}, \mathbf{X} \subset \mathbf{V}$, $P_{\mathbf{x}, \mathbf{pa}_Y}(y) = P_{\mathbf{x}}(y | \mathbf{pa}_Y)$ whenever \mathbf{pa}_Y is consistent with $\mathbf{X} = \mathbf{x}$ and $\mathbf{X}, \{Y\}$ are disjoint.

The main feature of Definition 8 is condition (ii), which implies that varying X from x to x' while keeping all other variables constant does not change Y 's distribution – this corresponds to an ideal experiment in which all variables are kept constant and the scientist “wiggles” one variable (or set) at a time, and contemplates how the target variable reacts (i.e., *ceteris paribus*).

This condition is supplemented by condition 8:(iii), which has also an intuitive appeal: if experiments show that separate interventions on \mathbf{X} and \mathbf{Z} have no direct effect on Y , then a joint intervention on \mathbf{X} and \mathbf{Z} should also have no direct effect on Y . Conditions (i) and (iv) are the same as in the missing-link definition.

One distinct feature of this new definition emerges when we test whether a given pair $\langle G, \mathbf{P}_* \rangle$ is compatible. First, the modularity condition of Definition 3 requires that each family is invariant to interventions on all subsets of elements “outside” the family, which is combinatorially explosive. In contrast, condition (ii) of Definition 8 involves singleton pairwise experiments which are easier to envision and evaluate.

⁴We use $\{\mathbf{A}, \mathbf{B}\}$ to denote the union of \mathbf{A} and \mathbf{B} .

Put another way, when the ZDE condition does not hold, it implies the existence of an edge between the respective pair of nodes thus providing fewer and easier experiments in testing the structure of the graph. Further, one should test the Markov compatibility of P and the new induced graph G .

We now show that all three local definitions of causal Bayesian networks stated so far are equivalent.

Theorem 2. *Let G be a DAG and \mathbf{P}_* a set of interventional distributions, the following statements are equivalent:*

- i. G is locally compatible with \mathbf{P}_*
- ii. G is missing-link compatible with \mathbf{P}_*
- iii. G is ZDE compatible with \mathbf{P}_*

Note that the notion of “parents set”, though less attached to modularity and invariance, is still invoked by the last two compatibility conditions. We believe therefore that it is an essential conceptual element in the definition of causal Bayesian networks.

5 Equivalence between the local definitions of causal Bayesian network

Definition 9 (Strong Markov Condition). *Each variable is interventionally independent of every other variable after fixing its parents. That is, for all $Y \in \mathbf{V}$ and $\mathbf{X} \subseteq \mathbf{V} - \{Y, \mathbf{PA}_Y\}$ we have*

$$P_{\mathbf{x}, \mathbf{pa}_Y}(y) = P_{\mathbf{pa}_Y}(y), \text{ for all } \mathbf{x}, y, \mathbf{pa}_Y \quad (9)$$

5.1 [Zde-CBN] \Rightarrow [local-CBN]

In this subsection, we assume that the four conditions in the definition of the Zero direct effect causal Bayesian network are valid for a given graph G and set \mathbf{P}_* .

The first result simply extends the Zero direct effect semantics to subset of variables:

Lemma 1. *Zde(\mathbf{W}, Y) holds for every $\mathbf{W} \subseteq \mathbf{V} - \{Y, \mathbf{PA}_Y\}$.*

Proof. Note that \mathbf{W} does not contain parents of Y . Then, [Zde] gives that, for every U in \mathbf{W} , we have $Zde(U, Y)$. But then, it follows directly by [Additivity], that $Zde(\mathbf{W}, Y)$ holds. \square

The next Lemma shows that the strong Markov condition is also valid for G and \mathbf{P}_* .

Lemma 2. *For all $Y \in \mathbf{V}$ and $\mathbf{X} \subset \mathbf{V} - \{Y, \mathbf{PA}_Y\}$, the relation $(Y \perp\!\!\!\perp_{ii} \mathbf{X} \mid \mathbf{PA}_Y)$ holds.*

Proof. Let $\mathbf{T}_1 = \mathbf{V} - \{Y, \mathbf{PA}_Y\}$, and note that $S_{Y\mathbf{T}_1} = \mathbf{PA}_Y$. Since \mathbf{T}_1 does not have parents of Y , by Lemma 1, we have $Zde(\mathbf{T}_1, Y)$, that is

$$P_{\mathbf{t}_1, s_{y\mathbf{t}_1}}(y) = P_{s_{y\mathbf{t}_1}}(y) = P_{\mathbf{pa}_Y}(y)$$

Now, let $\mathbf{T}_2 = \mathbf{V} - \{Y, \mathbf{X}, \mathbf{PA}_Y\}$, and note that $S_{Y\mathbf{T}_2} = \{\mathbf{X}, \mathbf{PA}_Y\}$. Since \mathbf{T}_2 does not have parents of Y , by Lemma 1, we have $Zde(\mathbf{T}_2, Y)$, that is

$$P_{\mathbf{t}_2, s_{y\mathbf{t}_2}}(y) = P_{s_{y\mathbf{t}_2}}(y) = P_{\mathbf{x}, \mathbf{pa}_Y}(y)$$

Since $(\mathbf{T}_1 \cup S_{Y\mathbf{T}_1}) = (\mathbf{T}_2 \cup S_{Y\mathbf{T}_2})$, we obtain

$$P_{\mathbf{x}, \mathbf{pa}_Y}(y) = P_{\mathbf{pa}_Y}(y)$$

\square

Lemma 3. *The condition of [Modularity] is valid for G and \mathbf{P}_* .*

Proof. Fix a variable Y and $\mathbf{X} \subset \mathbf{V} - \{Y\}$. We need to show that

$$P_{\mathbf{x}}(y \mid \mathbf{pa}_Y) = P(y \mid \mathbf{pa}_Y)$$

Applying the condition [Parents do/see] to both sides in the equation above, we obtain

$$P_{\mathbf{x}, \mathbf{pa}_Y}(y) = P_{\mathbf{pa}_Y}(y)$$

and we immediately recognize here a claim of the strong Markov condition. \square

Finally, the observation that the condition [Markov] is present in both definitions, we complete the proof that G is a local causal Bayesian network for \mathbf{P}_* .

5.2 [local-CBN] \Rightarrow [Zde-CBN]

In this subsection, we assume that the two conditions in the definition of the local causal Bayesian network are valid for a given graph G and set \mathbf{P}_* .

Lemma 4. *For all $Y \in \mathbf{V}$ and $\mathbf{X} \subset \mathbf{V} - \{Y, \mathbf{PA}_Y\}$ we have*

$$P_{\mathbf{x}, \mathbf{pa}_Y}(\mathbf{pa}_Y \mid y) = 1$$

whenever $P_{\mathbf{x}, \mathbf{pa}_Y}(y) > 0$, and \mathbf{pa}_Y is compatible with \mathbf{x} .

Proof. This is an immediate consequence of the property of [Effectiveness], in the definition of \mathbf{P}_* . \square

Lemma 5. *The condition [Parents do/see] is valid for G and \mathbf{P}_* .*

Proof. Fix a variable $\mathbf{X} \subset \mathbf{V}$ and consider an arbitrary instantiation \mathbf{v} of variables \mathbf{V} , and \mathbf{pa}_Y consistent with \mathbf{x} .

Consider the intervention $do(\mathbf{X} = \mathbf{x})$, and given the condition [Modularity], $P_{\mathbf{x}}(y \mid \mathbf{pa}_Y) = P(y \mid \mathbf{pa}_Y)$, $Y \notin \mathbf{X}$. Now consider the intervention $do(\mathbf{X} = \mathbf{x}, \mathbf{PA}_Y = \mathbf{pa}_Y)$, and again by the condition [Modularity] $P_{\mathbf{x}, \mathbf{pa}_Y}(y \mid \mathbf{pa}_Y) = P(y \mid \mathbf{pa}_Y)$. The r.h.s. coincide, therefore

$$\begin{aligned} P_{\mathbf{x}}(y \mid \mathbf{pa}_Y) &= P_{\mathbf{x}, \mathbf{pa}_Y}(y \mid \mathbf{pa}_Y) \\ &\stackrel{\text{Bayes thm.}}{=} \frac{P_{\mathbf{x}, \mathbf{pa}_Y}(\mathbf{pa}_Y \mid y) P_{\mathbf{x}, \mathbf{pa}_Y}(y)}{P_{\mathbf{x}, \mathbf{pa}_Y}(\mathbf{pa}_Y)} \\ &\stackrel{\text{effectiveness}}{=} P_{\mathbf{x}, \mathbf{pa}_Y}(\mathbf{pa}_Y \mid y) P_{\mathbf{x}, \mathbf{pa}_Y}(y) \end{aligned} \quad (10)$$

We consider two cases. If $P_{\mathbf{x}, \mathbf{pa}_Y}(y) > 0$, by lemma 4 $P_{\mathbf{x}, \mathbf{pa}_Y}(\mathbf{pa}_Y \mid y) = 1$, and then substituting back in eq. (10) we obtain $P_{\mathbf{x}}(y \mid \mathbf{pa}_Y) = P_{\mathbf{x}, \mathbf{pa}_Y}(y)$. If $P_{\mathbf{x}, \mathbf{pa}_Y}(y) = 0$, substituting back in eq. (10) we obtain $P_{\mathbf{x}}(y \mid \mathbf{pa}_Y) = P_{\mathbf{x}, \mathbf{pa}_Y}(\mathbf{pa}_Y \mid y) * 0 = 0$, and then $P_{\mathbf{x}}(y \mid \mathbf{pa}_Y) = P_{\mathbf{x}, \mathbf{pa}_Y}(y)$. \square

Lemma 6. *The condition [Zde] is valid for G and \mathbf{P}_* .*

Proof. Fix $Y, X \in \mathbf{V}$ such that there is no arrow pointing from X to Y . Let $\mathbf{S}_{\mathbf{XY}} = \mathbf{V} - \{X, Y\}$. We want to show

$$P_{x, \mathbf{s}_{\mathbf{XY}}}(y) = P_{\mathbf{s}_{\mathbf{XY}}}(y), \text{ for all } x, y, \mathbf{s}_{\mathbf{XY}}$$

Note that $\mathbf{PA}_y \subseteq \mathbf{S}_{\mathbf{XY}}$, and then by the [Parent do/see] condition we have to show

$$P_{x, \mathbf{s}'_{\mathbf{XY}}}(y \mid \mathbf{pa}_y) = P_{\mathbf{s}'_{\mathbf{XY}}}(y \mid \mathbf{pa}_y)$$

where $\mathbf{S}'_{\mathbf{XY}} = \mathbf{S}_{\mathbf{XY}} - \{\mathbf{PA}_y\}$.

The condition [Modularity] implies that $P_{x, \mathbf{s}'_{\mathbf{XY}}}(y \mid \mathbf{pa}_y) = P(y \mid \mathbf{pa}_y)$. Again by [Modularity], we obtain $P(y \mid \mathbf{pa}_y) = P_{\mathbf{s}'_{\mathbf{XY}}}(y \mid \mathbf{pa}_y)$. Applying [Parents do/see], [Zde] follows. \square

Lemma 7. *The condition [Additivity] is valid for G and \mathbf{P}_* .*

Proof. Fix $\mathbf{X} \subset \mathbf{V}$ and $Z, Y \in \mathbf{V}$. Let $\mathbf{S}_{\mathbf{xzy}} = \mathbf{V} - \{\mathbf{X}, Y, Z\}$. Assume $Zde(\mathbf{X}, Y)$ and $Zde(Z, Y)$. For the sake of contradiction, suppose that $Zde(\mathbf{X} \cup \{Z\}, Y)$ is false.

We can rewrite it based on the law of total probability,

$$\sum_{\mathbf{pa}_y} P_{\{\mathbf{x}, z\}, \mathbf{s}_{\mathbf{xzy}}}(y \mid \mathbf{pa}_y) P_{\{\mathbf{x}, z\}, \mathbf{s}_{\mathbf{xzy}}}(\mathbf{pa}_y) \neq \sum_{\mathbf{pa}_y} P_{\mathbf{s}_{\mathbf{xzy}}}(y \mid \mathbf{pa}_y) P_{\mathbf{s}_{\mathbf{xzy}}}(\mathbf{pa}_y)$$

Notice that there is only one configuration of \mathbf{pa}_y consistent with $\mathbf{s}_{\mathbf{xzy}}$ in both sides because $\mathbf{PA}_y \subseteq \mathbf{S}_{\mathbf{xzy}}$ and [Effectiveness]. Then, this equation reduces to

$$P_{\{\mathbf{x}, z\}, \mathbf{s}_{\mathbf{xzy}}}(y \mid \mathbf{pa}_y) \neq P_{\mathbf{s}_{\mathbf{xzy}}}(y \mid \mathbf{pa}_y)$$

We reach a contradiction given [Modularity]. \square

The proof for the Missing-link CBN is analogous.

6 Conclusions

We first proved the equivalence between two characterizations of Causal Bayesian Networks, one local, based on modularity, and the other global, based on the truncated product formula. We then introduced two alternative characterizations of CBNs, proved their equivalence with the previous ones, and showed that some of their features make the tasks of empirically testing the network structure, as well as judgmentally assessing its plausibility more manageable.

Another way to look at the results of our analysis is in terms of the information content of CBNs, that is, what constraints a given CBN imposes on both observational and experimental findings. For a probabilistic Bayes network the answer is simple and is given by the set of conditional independencies that are imposed by the d-separation criterion. For a CBN, the truncated product formula (2) imposes conditional independencies on any interventional distribution $P_x(\mathbf{v})$. But this does not sum up the entire information content of a CBN. Equation (2) further tells us that the relationship between any two interventional distributions, say $P_x(\mathbf{v})$ and $P_{x'}(\mathbf{v})$, is not entirely arbitrary; the two distributions constrain each other in various ways. For example, the conditional distributions

$P_x(v_i \mid \mathbf{pa}_i)$ and $P_{x'}(v_i \mid \mathbf{pa}_i)$ must be the same for any unmanipulated family. Or, as another example, for any CBN we have the inequality: $P_x(y) \geq P(x, y)$ [Tian *et al.*, 2006].

A natural question to ask is whether there exists a representation that encodes all constraints of a given type. The modularity property of Definition 2 constitutes such a representation, and so do the missing-link and the ZDE definitions. Each encodes constraints of a given type and our equivalence theorems imply that all constraints encoded by one representation can be reconstructed from the other representation without loss of information.

References

- A. P. Dawid. Influence diagrams for causal modelling and inference. *International Statistical Review*, 70(2):161–189, 2001.
- D. Galles and J. Pearl. An axiomatic characterization of causal counterfactuals. *Foundation of Science*, 3(1):151–182, 1998.
- J.Y. Halpern. Axiomatizing causal reasoning. In G.F. Cooper and S. Moral, editors, *Uncertainty in Artificial Intelligence*, pages 202–210. Morgan Kaufmann, San Francisco, CA, 1998. Also, *Journal of Artificial Intelligence Research* 12:3, 17–37, 2000.
- D. Heckerman and R. Shachter. Decision-theoretic foundations for causal reasoning. *Journal of Artificial Intelligence Research*, 3:405–430, 1995.
- D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.
- S. L. Lauritzen. Causal inference from graphical models. In *Complex Stochastic Systems*, pages 63–107. Chapman and Hall/CRC Press, 1999.
- J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Mateo, CA, 1988.
- J. Pearl. Belief networks revisited. *Artificial Intelligence*, 59:49–56, 1993.
- J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, 2000. Second ed., 2009.
- J.M. Robins. A new approach to causal inference in mortality studies with a sustained exposure period – applications to control of the healthy workers survivor effect. *Mathematical Modeling*, 7:1393–1512, 1986.
- P. Spirtes, C.N. Glymour, and R. Scheines. *Causation, Prediction, and Search*. Springer-Verlag, New York, 1993.
- J. Tian and J. Pearl. A new characterization of the experimental implications of causal Bayesian networks. In *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, pages 574–579. AAAI Press/The MIT Press, Menlo Park, CA, 2002.
- J. Tian, C. Kang, and J. Pearl. A characterization of interventional distributions in semi-Markovian causal models. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*, pages 1239–1244. AAAI Press, Menlo Park, CA, 2006.