In M. Croitoru, S. Rudolph, N. Wilson, J. Howse, and O. Corby (Eds.),
GKR 2011, LNAI 7205, Berlin Heidelberg: Springer-Verlag, pp. 1-17, 2012.

TECHNICAL REPORT
R-384
May 2011

Local Characterizations of Causal Bayesian Networks*

Elias Bareinboim
Cognitive Systems Laboratory
Computer Science Department
University of California Los Angeles
CA 90095
eb@cs.ucla.edu

Carlos Brito
Computer Science Department
Federal University of Ceará
carlos@lia.ufc.br

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California Los Angeles
CA 90095
judea@cs.ucla.edu

Abstract

The standard definition of causal Bayesian networks (CBNs) invokes a global condition according to which the distribution resulting from any intervention can be decomposed into a truncated product dictated by its respective mutilated subgraph. We provide alternative formulations which emphasizes local aspects of the causal process and can serve therefore as more meaningful criterion for testing coherence and network construction. We first examine a definition based on "modularity" and prove its equivalence to the global definition. We then introduce two new definitions, the first interprets the missing edges in the graph, and the second interprets "zero direct effect" (i.e., *ceteris paribus*). We show that these two formulations are equivalent but carry different semantic content.

1 Introduction

Nowadays, graphical models are standard tools for encoding distributional and causal information [Pearl, 1988; Spirtes *et al.*, 1993; Heckerman and Shachter, 1995; Lauritzen, 1999; Pearl, 2000; Dawid, 2001; Koller and Friedman, 2009]. One of the most popular representations is a *causal Bayesian network*, namely, a directed acyclic graph (DAG) G which, in addition to the traditional conditional independencies also conveys causal information, and permits one to infer the effects of interventions. Specifically, if an external intervention fixes any set $\mathbf{X}$ of variables to some constant $\mathbf{x}$, the DAG permits us to infer the resulting post-intervention distribution, denoted by $P_{\mathbf{x}}(\mathbf{v})$, [1] from the pre-intervention distribution $P(\mathbf{v})$.

The standard reading of post-interventional probabilities invokes cutting off incoming arrows to the manipulated variables and leads to a "truncated product" formula [Pearl, 1993], also known as "manipulation theorem" [Spirtes *et al.*, 1993] and "G-computation formula" [Robins, 1986]. A local characterization of CBNs invoking the notion of conditional invariance was presented in [Pearl, 2000, p.24] and will be shown here to imply (and be implied by) the truncated product formula. This characterization requires the network builder to judge whether the conditional probability $P(Y \mid \mathbf{PA_y})$ for each parents-child family remains invariant under interventions outside this family. [Tian and Pearl, 2002] provides another characterization with respect to three norms of coherence called Effectiveness, Markov and Recursiveness, and showed their use in learning and identification when the causal graph is not known in advance.

In this paper, we use the concepts of "conditional invariance" and "interventional invariance" to formulate and compare several definitions of CBNs. The first assures invariance of conditional probabilities for each family, while the other two assure the invariance of the distribution of each variable under different interventions. We show that these three definitions are equivalent to the global one, and lead to the same predictions under interventions.

The rest of the paper is organized as follows. In Section 2, we introduce the basic concepts, and present the standard global and local definitions of CBNs together with discussion of their features. In Section 3, we prove the equivalence between these two definitions. In Section 4, we introduce two new definitions which explicitly interprets the missing links in the graph as representing absence of causal influence. In Section 5, we prove the equivalence between these definitions and the previous ones. Finally, we provide concluding remarks in Section 6.

2 Causal Bayesian networks and interventions

A causal Bayesian network (also known as a *Markovian model*) consists of two mathematical objects: (i) a DAG G, called a causal graph, over a set $\mathbf{V} = \{V_1, ..., V_n\}$ of vertices, and (ii) a probability distribution $P(\mathbf{v})$, over the set $\mathbf{V}$ of discrete variables that correspond to the vertices in G. The interpretation of such a graph has two components, probabilistic and causal.[2]

*This work was supported in parts by National Institutes of Health #1R01 LM009961-01, National Science Foundation #IIS-0914211 and #IIS-1018922, and Office of Naval Research #N000-14-09-1-0665 and #N00014-10-1-0933.

[1] [Pearl, 2000] used the notation $P(\mathbf{v} \mid set(\mathbf{t}))$, $P(\mathbf{v} \mid do(\mathbf{t}))$, or $P(\mathbf{v} \mid \hat{\mathbf{t}})$ for the post-intervention distribution, while [Lauritzen, 1999] used $P(\mathbf{v} \parallel \mathbf{t})$.

[2] A more refined interpretation, called functional, is also common [Pearl, 2000], which, in addition to interventions, supports counterfactual readings. The functional interpretation assumes determinis-

The probabilistic interpretation [Pearl, 1988] views G as representing conditional independence restrictions on P: each variable is independent of all its non-descendants given its parents in the graph. This property is known as the *Markov condition* and characterizes the Bayesian network absent of any causal reading. These conditional independence restrictions imply that the joint probability function $P(\mathbf{v}) = P(v_1, ..., v_n)$ factorizes according to the product:

$$P(\mathbf{v}) = \prod_i P(v_i \mid \mathbf{pa_i}) \tag{1}$$

where $\mathbf{pa_i}$ are (assignments of) the parents of variables V_i in G.

The causal interpretation views the arrows in G as representing potential causal influences between the corresponding variables and, alternatively, the absence of arrows represents no direct causal influence between the corresponding variables. In this interpretation, the factorization of eq. (1) still holds, but the factors are further assumed to represent autonomous data-generation processes, that is, each family conditional probability $P(v_i \mid \mathbf{pa_i})$ represents a stochastic process by which the values of V_i are assigned in response to the values $\mathbf{pa_i}$ (previously chosen for V_i's parents), and the stochastic variation of this assignment is assumed independent of the variations in all other assignments in the model.

Moreover, each assignment process remains invariant to possible changes in the assignments processes that govern other variables in the system. This invariance assumption is known as modularity and it enables us to predict the effects of interventions, whenever interventions are described as specific modification of some factors in the product of eq. (1). The most elementary intervention considered is the *atomic* one, where a set $\mathbf{X}$ of variables is fixed to some constant $\mathbf{X} = \mathbf{x}$. The following definitions will facilitate subsequent discussions.

Definition 1 (Interventional distributions). *Let $P(\mathbf{v})$ be a probability distribution on a set $\mathbf{V}$ of variables, and let $P_{\mathbf{x}}(\mathbf{v})$ denote the distribution resulting from the intervention do$(\mathbf{X} = \mathbf{x})$ that sets a subset $\mathbf{X}$ of variables to constant $\mathbf{x}$. Denote by $\mathbf{P}_*$ the set of all interventional distributions $P_{\mathbf{x}}(\mathbf{v}), \mathbf{X} \subseteq V$, including $P(\mathbf{v})$, which represents no intervention (i.e., $\mathbf{X} = \emptyset$). Additionally, $\mathbf{P}_*$ is such that for all $\mathbf{X} \subseteq V$, the following property holds:*

i. *[Effectiveness] $P_{\mathbf{x}}(v_i) = 1$, for all $V_i \in \mathbf{X}$ whenever v_i is consistent with $\mathbf{X} = \mathbf{x}$;*

Definition 2 (Conditional invariance). *We say that Y is conditional invariant to $\mathbf{X}$ given $\mathbf{Z}$, denoted $(Y \perp\!\!\!\perp_{ci} \mathbf{X} \mid \mathbf{Z})_{\mathbf{P}_*}$, if intervening on $\mathbf{X}$ does not change the conditional distribution of Y given $\mathbf{Z} = \mathbf{z}$, i.e., $\forall \mathbf{x}, y, \mathbf{z}, P_{\mathbf{x}}(y \mid \mathbf{z}) = P(y \mid \mathbf{z})$.*

To capture the intuition behind atomic interventions, [Pearl, 2000] proposed the following local definition of causal Bayesian networks:

Definition 3 (Causal Bayesian network [Pearl, 2000, p.24]). *A DAG G is said to be locally compatible with a set of interventional distributions $\mathbf{P}_*$ if and only if the following conditions hold for every $P_{\mathbf{x}} \in P_*$:*

i. *[Markov] $P_{\mathbf{x}}(\mathbf{v})$ is Markov relative to G;*

ii. *[Modularity] $(V_i \perp\!\!\!\perp_{ci} \mathbf{X} \mid \mathbf{PA_i})_{\mathbf{P}_*}$, for all $V_i \notin \mathbf{X}$ whenever $\mathbf{pa_i}$ is consistent with $\mathbf{X} = \mathbf{x}$.* [3]

We shall show (Sec. 3) that modularity permits us to answer queries about the effect of interventions, or causal effects. A *causal effect* of variable X on variable Y written $P_x(y)$, stands for the probability that variable Y attains value y if we enforce uniformly over the population the constraint $X = x$. The standard definition of causal Bayesian networks is based on a global compatibility condition, which makes explicit the joint post-intervention distribution under any arbitrary intervention.

Definition 4 (Global causal Bayesian network [Pearl, 2000]). *A DAG G is said to be globally compatible with a set of interventional distributions $\mathbf{P}_*$ if and only if the distribution $P_{\mathbf{x}}(\mathbf{v})$ resulting from the intervention do$(\mathbf{X} = \mathbf{x})$ is given by the following expression:*

$$P_{\mathbf{x}}(\mathbf{v}) = \begin{cases} \prod_{\{i \mid V_i \notin \mathbf{X}\}} P(v_i \mid \mathbf{pa_i}) & \mathbf{v} \text{ consistent with } \mathbf{x}. \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Equation (2) is also known as the *truncated factorization product* of eq. (1), with factors corresponding to the manipulated variables removed. The truncated factorization follows from Definition 3 because, assuming modularity, the post-intervention probabilities $P(v_i \mid \mathbf{pa_i})$ corresponding to variables in X are either 1 or 0, while those corresponding to unmanipulated variables remain unaltered.

The two definitions emphasize different aspects of the causal model; Definition 3 ensures that each conditional probability $P(v_i \mid \mathbf{pa_i})$ (locally) remains invariant under interventions that do not include directly V_i, while Definition 4 ensures that each manipulated variable is not influenced by its previous parents (before the manipulation), and every other variable is governed by its pre-interventional process. Because the latter invokes theoretical conditions on the data-generating process, it is not directly testable, and the question whether a given implemented intervention conforms to an investigator's intention (e.g., no side effects) is discernible only through the testable properties of the truncated product formula (2). Definition 3 provides in essence a series of local tests for Equation (2), and the equivalence between the two (Theorem 1, below) ensures that *all* empirically testable properties of (2) are covered by the local tests provided by Definition 3.

tic functional relationships between variables in the model, some of which may be unobserved. Complete axiomatizations of deterministic counterfactual relations are given in [Galles and Pearl, 1998; Halpern, 1998].

[3] Explicitly, modularity states: $P(v_i|\mathbf{pa_i}, do(\mathbf{s})) = P(v_i|\mathbf{pa_i})$ for any set $\mathbf{S}$ of variables disjoint of $\{V_i, \mathbf{PA_i}\}$.

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 ⇒ 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

$$\begin{aligned} P_{\mathbf{x}}(\mathbf{v}) & \overset{\text{def.3:}(i)}{=} \prod_i P_{\mathbf{x}}(v_i \mid \mathbf{pa_i}) \\ & = \prod_{\{i \mid v_i \in \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa_i}) \prod_{\{i \mid v_i \notin \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa_i}) \\ & \overset{\text{effectiveness}}{=} \prod_{\{i \mid v_i \notin \mathbf{X}\}} P_{\mathbf{x}}(v_i \mid \mathbf{pa_i}) \\ & \overset{\text{def.3:}(ii)}{=} \prod_{\{i \mid v_i \notin \mathbf{X}\}} P(v_i \mid \mathbf{pa_i}) \end{aligned} \tag{3}$$

which is the truncated product as desired.

(Definition 4 ⇒ 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 $Dom(v_i) = \{v_{i1}, v_{i2}, ..., v_{im}\}$ 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 $\pi : (v_1, ..., v_n)$ 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 $do(\mathbf{X} = \mathbf{x})$

$$\begin{aligned} P_{\mathbf{x}}(v_i \mid \mathbf{pa_i}) & = \frac{P_{\mathbf{x}}(v_i, \mathbf{pa_i})}{P_{\mathbf{x}}(\mathbf{pa_i})} \\ & \overset{\text{marginal.}}{=} \frac{\sum_{v_j \notin \{V_i, \mathbf{PA_i}\}} P_{\mathbf{x}}(\mathbf{v})}{\sum_{v_j \notin \{\mathbf{PA_i}\}} P_{\mathbf{x}}(\mathbf{v})} \\ & \overset{\text{eq.}(2)}{=} \frac{\sum_{v_j \notin \{V_i, \mathbf{PA_i}, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa_k})}{\sum_{v_j \notin \{\mathbf{PA_i}, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa_k})} \\ & = P(v_i \mid \mathbf{pa_i}) \times \\ & \quad \frac{\sum_{v_j \notin \{V_i, \mathbf{PA_i}, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}, k \neq i} P(v_k \mid \mathbf{pa_k})}{\sum_{v_j \notin \{\mathbf{PA_i}, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa_k})} \end{aligned} \tag{4}$$

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

$$\begin{aligned} & \sum_{v_j \notin \{V_i, \mathbf{PA_i}, \mathbf{X}\}} \prod_{\substack{v_k \notin \mathbf{X} \\ k \neq i}} P(v_k \mid \mathbf{pa_k}) = \\ & \sum_{\substack{v_j \notin \{\mathbf{PA_i}, \mathbf{X}\} \\ j < i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k < i}} P(v_k \mid \mathbf{pa_k}) \sum_{\substack{v_j \notin \mathbf{X} \\ j > i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k > i}} P(v_k \mid \mathbf{pa_k}) \end{aligned} \tag{5}$$

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

$$\begin{aligned} & \sum_{v_j \notin \{V_i, \mathbf{PA_i}, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa_k}) = \\ & \qquad \sum_{\substack{v_j \notin \{\mathbf{PA_i}, \mathbf{X}\} \\ j < i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k < i}} P(v_k \mid \mathbf{pa_k}) \end{aligned} \tag{6}$$

Similarly for the denominator,

$$\begin{aligned} & \sum_{v_j \notin \{\mathbf{PA_i}, \mathbf{X}\}} \prod_{v_k \notin \mathbf{X}} P(v_k \mid \mathbf{pa_k}) = \\ & \qquad \sum_{\substack{v_j \notin \{\mathbf{PA_i}, \mathbf{X}\} \\ j < i}} \prod_{\substack{v_k \notin \mathbf{X} \\ k < i}} P(v_k \mid \mathbf{pa_k}) \end{aligned} \tag{7}$$

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

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

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

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] $\forall Y \in \mathbf{V}, \mathbf{X} \subset \mathbf{V}$, $P_{\mathbf{x},\mathbf{pa_y}}(y) = P_{\mathbf{x}}(y \mid \mathbf{pa_y})$ 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 $P_a(b) \neq P(b \mid a)$. However, Definition 3 implies $P_a(b) = P(b \mid a)$ 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 $(Y \perp\!\!\!\perp_{ii} \mathbf{X} \mid \mathbf{Z})_{\mathbf{P}_*}$, 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., $(Y \perp\!\!\!\perp_{ci} \mathbf{X} \mid \mathbf{PA_y})_{\mathbf{P}_*}, \forall \mathbf{X}$), while Definition 5 expresses the absence of causal effect in terms of interventional invariance (i.e., $(Y \perp\!\!\!\perp_{ii} \mathbf{X} \mid \mathbf{PA_y}, \mathbf{S})_{\mathbf{P}_*}, \forall \mathbf{S}, \mathbf{X}$).

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 $(Y \perp\!\!\!\perp \mathbf{ND_Y} \mid \mathbf{PA_y})_{\mathbf{P}_*}$, and the *II* claims are of the form $(Y \perp\!\!\!\perp_{ii} X \mid \mathbf{PA_y}, \mathbf{S})_{\mathbf{P}_*}, \forall X, \mathbf{S}$. In both cases, $\mathbf{PA_y}$ is required to separate Y from other variables. In the observational case Y is separated from its nondescendants, 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 $\mathbf{S_{XY}} = \mathbf{V} - \{X, Y\}$.* [4] *We say that $\mathbf{X}$ has zero direct effect on Y, denoted $ZDE(\mathbf{X}, Y)$, if*

$$(Y \perp\!\!\!\perp_{ii} \mathbf{X} \mid \mathbf{S_{xy}})$$

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] $\forall Y \in \mathbf{V}, \mathbf{X} \subset \mathbf{V}$, $P_{\mathbf{x},\mathbf{pa_y}}(y) = P_{\mathbf{x}}(y \mid \mathbf{pa_y})$ 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 "wriggles" 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 $< G, \mathbf{P}_* >$ 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.

[4] We 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*

$$P_{\mathbf{x},\mathbf{pa_y}}(y) = P_{\mathbf{pa_y}}(y), \textit{ for all } \mathbf{x}, y, \mathbf{pa_y} \tag{9}$$

5.1 [Zde-CBN] ⇒ [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. □

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 $\mathbf{T_2} = V - \{Y, \mathbf{X}, \mathbf{PA_Y}\}$, 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

$$P_{\mathbf{t_2}, s_{y\mathbf{t_2}}}(y) = P_{s_{y\mathbf{t_2}}}(y) = P_{x,\mathbf{pa_y}}(y)$$

Since $(\mathbf{T_1} \cup S_{Y\mathbf{T_1}}) = (\mathbf{T_2} \cup S_{Y\mathbf{T_2}})$, we obtain

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

□

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

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

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

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] ⇒ [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 $\mathbf{P}_*$. □

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

$$\begin{aligned} P_{\mathbf{x}}(y \mid \mathbf{pa_y}) &= P_{\mathbf{x},\mathbf{pa_y}}(y \mid \mathbf{pa_y}) \\ &\overset{\text{Bayes thm.}}{=} \frac{P_{\mathbf{x},\mathbf{pa_y}}(\mathbf{pa_y} \mid y) P_{\mathbf{x},\mathbf{pa_y}}(y)}{P_{\mathbf{x},\mathbf{pa_y}}(\mathbf{pa_y})} \\ &\overset{\text{effectiveness}}{=} P_{\mathbf{x},\mathbf{pa_y}}(\mathbf{pa_y} \mid y) P_{\mathbf{x},\mathbf{pa_y}}(y) \end{aligned} \tag{10}$$

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 $P_{\mathbf{x}}(y \mid \mathbf{pa_y}) = P_{\mathbf{x},\mathbf{pa_y}}(y)$. □

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 $\mathbf{S_{XY}} = \mathbf{V} - \{X, Y\}$. We want to show

$$P_{x,\mathbf{s_{xy}}}(y) = P_{\mathbf{s_{xy}}}(y), \text{for all } x, y, \mathbf{s_{xy}}$$

Note that $\mathbf{PA_y} \subseteq \mathbf{S_{xy}}$, and then by the [Parent do/see] condition we have to show

$$P_{x,\mathbf{s'_{xy}}}(y \mid \mathbf{pa_y}) = P_{\mathbf{s'_{xy}}}(y \mid \mathbf{pa_y})$$

where $\mathbf{S'_{xy}} = \mathbf{S_{xy}} - \{\mathbf{PA_y}\}$.

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

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 $\mathbf{S_{xzy}} = \mathbf{V} - \{\mathbf{X}, Y, Z\}$. 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,

$$\sum_{\mathbf{pa_y}} P_{\{\mathbf{x},z\},\mathbf{s_{xzy}}}(y \mid \mathbf{pa_y}) P_{\{\mathbf{x},z\},\mathbf{s_{xzy}}}(\mathbf{pa_y}) \neq \sum_{\mathbf{pa_y}} P_{\mathbf{s_{xzy}}}(y \mid \mathbf{pa_y}) P_{\mathbf{s_{xzy}}}(\mathbf{pa_y})$$

Notice that there is only one configuration of $\mathbf{pa_y}$ consistent with $\mathbf{s_{xzy}}$ in both sides because $\mathbf{PA_y} \subseteq \mathbf{S_{xzy}}$ and [Effectiveness]. Then, this equation reduces to

$$P_{\{\mathbf{x},z\},\mathbf{s_{xzy}}}(y \mid \mathbf{pa_y}) \neq P_{\mathbf{s_{xzy}}}(y \mid \mathbf{pa_y})$$

We reach a contradiction given [Modularity]. □

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 $P_x(v_i|\mathbf{pa_i})$ and $P_{x'}(v_i|\mathbf{pa_i})$ 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.