Exploiting Innocuousness in Bayesian Networks

Alexander Motzek and Ralf Möller

Institute of Information Systems, Universität zu Lübeck, Germany
{motzek,moeller}@ifis.uni-luebeck.de

Abstract. Boolean combination functions in Bayesian networks, such as noisy-or, are often credited a property stating that inactive dependences (e.g., observed to *false*) do not "cause any harm" and an arc becomes vacuous and could have been left out. However, in classic Bayesian networks we are not able to express this property in local CPDs. By using novel ADBNs, we formalize the innocuousness property in CPDs and extend previous work on context-specific independencies. With an explicit representation of innocuousness in local CPDs, we provide a higher causal accuracy for CPD specifications and open new ways for more efficient and less-restricted reasoning in (A)DBNs.

1 Introduction

Boolean combination functions in Bayesian networks (BNs) are often credited a property stating that if a dependence is observed to be inactive (i.e., a precondition observed to be *false*) it shall not "cause any harm" and its arc becomes vacuous, i.e., could have been left out. We call such a property an "*innocuousness*" property of conditional probability distributions (CPDs). We cannot specify such an innocuousness property in CPDs, nor formalize it up to now. Notwithstanding, vacuous dependencies have shown to be of valuable interest for efficient reasoning in Bayesian networks, and an innocuousness property is widely associated with, e.g., noisy-or combination functions. Further, being able to explicitly specify an innocuousness property in CPDs would allow for more precise representations of the world, as demanded and emphasized by Pearl [8]. Formalizing an innocuousness property can almost be achieved with context-specific independencies (CSIs) introduced by Boutilier et al. [2], but consider the following example: Say, random variable X is conditionally dependent on Y and C and we specify a CPD $P(X|Y,C)$. With CSIs we can specify that X becomes independent of Y in a specific context $C = c \in \mathrm{dom}(C)$, but X stays dependent on Y in another context $C = c' \in \mathrm{dom}(X)$. Boutilier et al. [2] formalize: $P(X|Y,c) = P(X|c)$ holds, if $\forall y, y' \in \mathrm{dom}(Y), \forall x \in \mathrm{dom}(X) : P(x|y,c) = P(x|y',c)$, but there exists a $c' \in \mathrm{dom}(C)$ s.t. $\exists y, y' \in \mathrm{dom}(Y), \exists x \in \mathrm{dom}(X) : P(x|y,c') \neq P(x|y',c')$. However, if we would like to formalize an innocuousness property stating that a context $C = c \in \mathrm{dom}(C)$ "removes" *itself* (and not only *another* variable), i.e., we would like to specify "$P(X|Y,c) = P(X|Y)$" in a CPD, we run into the problem that a formal definition is neither available nor easily possible: The allegedly irrelevant random variable C in question is in fact the one that ought to be relevant for specifying the independence.

Motzek and Möller [7] describe a novel form of Bayesian networks, called Activator Dynamic Bayesian Networks (ADBN). While they focus on the exploitation of activators in terms of graphical models, activators, as we will see in this paper, allow a formal definition of an innocuousness property. Further, they only consider activator sets when studying new possibilities in graphical models. We show that by considering innocuousness properties in CPD specifications, previously imposed restrictions of [7] can be significantly relaxed.

Independencies in Bayesian networks and graphical models in general have been extensively studied for efficient inference, notably by Zhang and Poole exploiting causal independencies [13], and have been extended with Boutilier et al.'s contextual independencies [2] in [9]. Still, a contextual independence where a context itself becomes independent was not considered in these works, and this hampers ways of more efficient reasoning and representations of causalities. Boolean combination functions have undergone notable considerations in works by Henrion [5], Srinivas [11], and Antonucci [1] introducing extensions to cope with imprecision. Cozman [3] provides formal definitions and specifies properties of combination functions leading to an axiomization of the noisy-or function, where we find an "accountability" property, which goes into the direction of defining an innocuousness property. However, a formal definition of innocuousness itself as "an inactive node does not cause any harm" is still missing. Although, the counterpart "only an active node causes harm" is mentioned as an "amechanistic" property in Heckerman and Breese [4] as well as Zagorecki and Druzdzel [12], but their work follows a different direction focusing on easier parametrization of CPDs.

The contribution of this paper can be summarized as follows. By formalizing a yet unexpressed innocuousness property in CPDs, we are able to more accurately represent causalities in CPDs, and we relax restrictions previously posed on graphical models. Based on graph enumeration techniques we quantitatively explore new relaxations of syntactic restrictions of graphical models for Bayesian networks.

We discuss preliminaries on ADBNs in Section 2 and introduce novel principles of graphical models using a running example. Afterwards, in Section 3 we introduce the innocuousness property and provide a formal definition using activator random variables. Subsequently we exploit the innocuousness property for relaxing restrictions posed on (A)DBNs in Section 4 and we show in Section 5 that the utility of (A)DBNs is significantly enhanced by exploiting innocuousness properties. We conclude in Section 6.

2 Activator Dynamic Bayesian Networks (ADBN)

After initial notations and definitions used throughout this paper, we demonstrate a running example for ADBNs. We consider an example from [7], which outlines restrictions of classic DBNs and motivates the use of *cyclic* ADBNs.

Notation 1 (State Variables) *Let X_i^t be the random variable for the i^{th} state X_i at time t, where X_i^t is assignable to a value $x_i \in \mathrm{dom}(X_i^t)$. Let $\boldsymbol{X}^t$ be the vector of all n state variables at time t, s.t.,*

$$\boldsymbol{X}^t = \left(X_1^t, \ldots, X_n^t\right)^\intercal .$$

Let $P(X_i^t = x_i)$ (or $P(x_i^t)$ for brevity) denote the probability of state X_i having x_i as a value at time t. If $\mathrm{dom}(X) = \{true, false\}$ we write $+x^t$ for the event $X^t = true$ and $\neg x^t$ for $X^t = false$ as usual. If X_i^t is unspecified and not fixed by evidence, $P(X_i^t)$ denotes the probability distribution of X_i^t w.r.t. all possible values in $\mathrm{dom}(X_i)$.

Definition 1 (Dynamic Bayesian Network). *A DBN is a tuple $(B_0, B_\rightarrow)$ with B_0 defining an initial Bayesian network (BN) representing time $t = 0$, containing all state variables X_i^0 in $\boldsymbol{X}^0$, and a consecutively repeated Bayesian network fragment $B_\rightarrow$ defining state dependencies between X_i^s and X_j^t, with $X_i^s \in \boldsymbol{X}^s, X_j^t \in \boldsymbol{X}^t, s \leq t$. By repeating $B_\rightarrow$ for every time step $t > 0$, a DBN $(B_0, B_\rightarrow)$ is unfolded into a BN defining its semantics as a joint probability over all random variables $P(\boldsymbol{X}^{0:t^\intercal})$. Notwithstanding, for every random variable X_i^t a local CPD, e.g., as a CPT, is defined.*

State dependencies defined in $B_\rightarrow$ are limited, s.t. no cyclic dependencies are created during unfolding.

Example 1 (Running Example & Motivation for ADBNs). Let us assume that in a company one is concerned with regulatory compliance over time. Business documents are exchanged and might contain manipulated information. Receiving such documents might influence an employee becoming corrupt at time t, which, further, might influence other employees. As an employee might *indeliberately* become corrupt, we say he becomes *credulous*. We represent the credulousness state of an employee, say, Claire, Don, and Earl, by respective random variables C^t, D^t, E^t. An influence can only occur if a message is passed from employee X to Y at t. We represent a message exchange by a random variable M_{XY}^t.We assume messages are only passed from Claire to Don to Earl. We can model these influences correctly in a DBN with $B_\rightarrow$ consisting of state random variables C^t, D^t, E^t and message variables M_{CD}^t, M_{DE}^t. Every state X^t depends on its previous state X^{t-1}, and D^t conditionally depends on C^t and M_{CD}^t, and respectively, E^t on D^t, M_{DE}^t. For every random variable, an appropriate CPD is defined. ▲

This example shows that simple influences can be correctly modeled in a DBN. However, we would like to model that messages can potentially be passed between *every* employee, which would render every state variable dependent on every other state variable. Clearly, this would cause cycles in $B_\rightarrow$, which are syntactically forbidden in Bayesian networks. In DBNs, cycles are usually resolved over time in a *diagonal* fashion, where states of t only influence $t + 1$ (Fig. 1, gray). However, we already used "time" for our modeling perspective, and bending dependencies over time causes conflicts with causality. The diagonal structure implies that receiving a message does not ultimately render another person credulous, but only in the consecutive timeslice. This immediately constrains the use of timeslices to infinitesimally small timeslices. If our observations of passed messages are temporally coarser, say, daily, or if high-frequency updates are too costly, then we clash with causality as indirect influences (e.g., C influences E *through* D) are not covered anymore, and we are bound to observations which do not require the anticipation of indirect influences (see later Prop. 1).

Fortunately, Motzek and Möller [7] show that ADBNs can actually be based on *cyclic* graphs completely sound with Bayesian network semantics as long as observations (message transfers) fulfill certain restrictions, based on the following definitions and theorems.

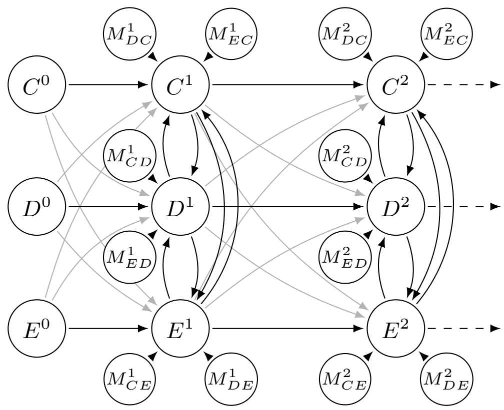

Fig. 1. A correctly represented world using an ADBN (black) for Ex. 1. Syntactic DAG constraints of BNs prevented desired cyclic intra-state dependencies and diagonal inter-state dependencies were enforced (hinted in light gray). In the diagonal case, M^t_{XY} represents $M^{t-1\,t}_{XY}$, i.e., M^t_{XY} affects the dependency of state Y^t on X^{t-1}.

Definition 2 (Activator Random Variables). *We use the notation A_{XY} for a so called activator random variable which activates a dependency of random variable Y on X in a given context. Let* $\mathrm{dom}(A_{XY}) = \{\mathit{true}, \mathit{false}\}$ *(extensions to non-boolean domains are straightforward). We define the* deactivation *criterion* $A_{XY} = \mathit{false}$ *as*

$$\forall x, x' \in \mathrm{dom}(X), \forall y \in \mathrm{dom}(Y), \forall \boldsymbol{z} \in \mathrm{dom}(\boldsymbol{Z}) : \\ P(y|x, \neg a_{XY}, \boldsymbol{z}) = P(y|x', \neg a_{XY}, \boldsymbol{z}) = P(y|*, \neg a_{XY}, \boldsymbol{z}) \ , \tag{1}$$

where $*$ represents a wildcard and $\boldsymbol{Z}$ further dependencies of Y.

The activation *criterion describes a situation where Y becomes dependent on X, i.e., the CPD entry for y is not uniquely identified by just $^{+}a_{XY}$ and $\boldsymbol{z}$, hence*

$$\exists x, x' \in \mathrm{dom}(X), \exists y \in \mathrm{dom}(Y), \exists \boldsymbol{z} \in \mathrm{dom}(\boldsymbol{Z}) : \\ P(y|x, {}^{+}a_{XY}, \boldsymbol{z}) \neq P(y|x', {}^{+}a_{XY}, \boldsymbol{z}) \ . \tag{2}$$

Let $A^{s\,t}$ describe a matrix of activator random variables between time s and t,

$$A^{s\,t} = \begin{pmatrix} A^{s\,t}_{11} & \cdots & A^{s\,t}_{1n} \\ \vdots & \ddots & \vdots \\ A^{s\,t}_{n1} & \cdots & A^{s\,t}_{nn} \end{pmatrix} .$$

Let $\boldsymbol{A}^{s\,t}_i$ denote the i^{th} column of $A^{s\,t}$ and let $\boldsymbol{\mathcal{A}}^{s\,t}$ denote the corresponding column vector of all entries of $A^{s\,t}$. For brevity, we write A^t for A^{tt} (excluding A^{tt}_{kk}), and correspondingly for A^t_{ij}, $\boldsymbol{A}^t_i$ and $\boldsymbol{\mathcal{A}}^t$.

In fact, in Ex. 1 message transfers (M^t_{XY}) take the role of activator random variables and we actually obtain an *Activator* DBN from Ex. 1.

Definition 3 (Activator Dynamic Bayesian Network). *A repeated ADBN fragment $B'_{\rightarrow}$ consists of dependencies between state variables X^s_i and X^t_j, $t-1 \leq s \leq t$ (Markov-1) and matrices $A^{s\,t}$ of activators. Let $A^{s\,t}_{ij}$ be the activator random variable influencing X^t_j regarding a dependency on X^s_i, such that X^t_j's local CPD follows Eq. 2 and Eq. 1. Every activator is assigned a prior probability. An ADBN is then syntactically defined by $(B_0, B'_{\rightarrow})$ defining its semantics as a well-defined joint probability over all random variables $P(\boldsymbol{X}^{0:t^{\intercal}}, \boldsymbol{\mathcal{A}}^{01:tt^{\intercal}})$.*

Note that activators in an ADBN are classic random variables and are part of the modeled domain (message transfers in Ex. 1), i.e., activators are no auxiliary variables. ADBNs are restricted in order to comply with a Bayesian network:

Theorem 1 (Bayesian Network Soundness). *For every combination, i.e., an arbitrary instantiation ${\boldsymbol{\mathcal{A}}_*}^{1:t}$ of $\boldsymbol{\mathcal{A}}^{1:t}$, an ADBN $(B_0, B'_{\rightarrow})$ corresponds to a Bayesian network, if for all t, ${\boldsymbol{\mathcal{A}}_*}^{t}$ satisfies the acyclicity constraint:*

$$\begin{aligned} \forall x, y, z \in \boldsymbol{X}^t : \mathfrak{A}(x,z)^t, \mathfrak{A}(z,y)^t \rightarrow \mathfrak{A}(x,y)^t \\ \neg\exists q : \mathfrak{A}(q,q)^t \ , \end{aligned} \tag{3}$$

with a predicate $\mathfrak{A}(i,j)^t$ defined as

$$\mathfrak{A}(i,j)^t = \begin{cases} \mathit{false} & \mathit{if} \quad \neg a^t_{ij} \\ \mathit{true} & \mathit{otherwise} \end{cases} .$$

A proof for this theorem can be found in [7, Sec. 3].

Theorem 1 means that if a specific structure of an DBN is not known in advance or is changing over time, an ADBN can intrinsically adapt itself to observations. Well-defined semantics is obtained in a (cyclic) ADBN, if only certain combinations of $\boldsymbol{\mathcal{A}}^{1:t}$ are instantiated, enforced by minimal sets of observations.

Example 2 (Restriction Example). Continuing Ex. 1, we observe a message transfer from Claire to Don ($^+m^1_{CD}$) and from Don to Earl ($^+m^1_{DE}$), and we can neglect all other transfers, i.e., $\neg m^1_{DC}, \neg m^1_{ED}, \neg m^1_{CE}, \neg m^1_{EC}$. These observations satisfy Thm. 1 and thus lead to a valid Bayesian network, even though it is based on a cyclic graph. To fully evaluate all implications of the observations, we have to anticipate an indirect influence from Claire to Earl through Don during timeslice 1. A diagonal "classic" DBN (as in Fig. 1, gray) cannot anticipate this indirection and would lead to spurious results. ▲

While cyclic ADBNs are syntactically restricted, Ex. 1 and Ex. 2 demonstrate that diagonal acyclic alternatives are significantly more restricted:

Proposition 1 (Diagonal ADBN Restrictions). *A classic, "diagonal" (as in Fig. 1, gray) (A)DBN is restricted in its usage to special observation sets. Indirect influences are spread over multiple timesteps and possible indirect influences inside one timestep cannot be considered. This restricts a DBN to observations where indirect influences strictly do not occur, i.e., no two activators* A^t_{*i} *and* A^t_{i*} *are allowed to be active, i.e. the set of probably active activators must form a bipartite digraph with uniformly directed edges (cf. [7, Prop. 1]).*

In the following, we introduce a novel property of CPDs, which significantly relaxes restrictions opposed by Thm. 1.

3 Innocuousness

We introduced innocuousness informally as "an inactive node does not cause any harm", but were unable to give a formal definition for such a property in CPDs of classic (D)BNs. Often "accountability", i.e., $P(+x|\neg *) = 0$ [3], is confused with the innocuousness property, but causally $P(+x|\neg *) = 0$ can also represent that exactly one *false*-dependence is responsible for $P(+x|\neg *)$ being 0.

As an extension to context-specific independencies (CSIs) from Boutilier et al. [2], we define a concept of *innocuousness contexts*, with fewer restrictions of CSIs. CSIs provide a formal definition for a variable X becoming independent of a variable Y in a context $\boldsymbol{C} = \boldsymbol{c} \in \text{dom}(\boldsymbol{C})$, where $X, Y \notin \boldsymbol{C}$. This allows us to specify properties such as $P(X|Y, \boldsymbol{c}) = P(X|\boldsymbol{c})$ in local CPDs. But, $X, Y \notin \boldsymbol{C}$ prevents us from specifying that a context $\boldsymbol{C} = \boldsymbol{c}$ removes one of *its own* random variables $C \in \boldsymbol{C}$, e.g., "$P(X|Y, c) = P(X|Y)$". Using activators in ADBNs we extend Boutilier's work to innocuousness contexts. We formally define that in a context $\boldsymbol{C} = \boldsymbol{c}$, a context variable $C \in \boldsymbol{C}$ can itself becomes irrelevant, which we call self-reflexive independence. Let us say a context $C = c \in \text{dom}(C)$, if it shall represent that X becomes independent of C, given $C = c$, i.e. $P(X|c, A_{CX}, \boldsymbol{Z}) = P(X|A_{CX}, \boldsymbol{Z})$. Using an activator-enriched CPD we define this to hold for binary activator random variables if

$$\begin{aligned} &\forall x \in \text{dom}(X), \forall \boldsymbol{z} \in \text{dom}(\boldsymbol{Z}) : \\ &\qquad P(x|c, {+a_{CX}}, \boldsymbol{z}) = P(x|c, \neg a_{CX}, \boldsymbol{z}) = P(x|*, \neg a_{CX}, \boldsymbol{z}) \ , \end{aligned} \tag{4}$$

where $\boldsymbol{Z}$ represents remaining further dependencies of X. Extensions to non-boolean activator random variables are straightforward.

This means, given $C = c$, A_{CX} becomes irrelevant for X, i.e., X becomes independent of A_{CX}. As A_{CX} can be instantiated in any form now, we can also say $\neg a_{CX}$. According to the deactivation criterion of an activator, X then becomes independent of C given $\neg a_{CX}$, or rather X becomes independent of C given $\neg c$, which is exactly what we intended.

Now, assume to specify a CPD $P(X|C, A_{CX}, Q, \boldsymbol{Z})$, where the innocuousness property of a variable is only in place in a further context. For example, there exists a variable Q that activates the innocuousness property of C only given $Q = q \in \mathrm{dom}(X)$. In this case, Eq. 4 only holds for specific $\boldsymbol{z} \in \mathrm{dom}(\boldsymbol{Z})$. This means, one innocuousness context is defined by instantiations of multiple random variables. Moreover, one random variable might stand in multiple different innocuousness contexts.

Notation 2 (Innocuousness Contexts) *Activator random variables are marked with a dot, e.g., $\dot{A}_{YX}$, if they are subject to become irrelevant in specific contexts. We denote a context in which Y is innocuous for X as a so called* innocuousness context *as a left superscript on A_{YX}. If a context is met and Y is innocuous for X, we say that A_{YX} stands in the innocuousness context. For the first example, this would be*

$$P(X|C, {}^{C=c}\dot{A}_{CX}, \boldsymbol{Z}) \ ,$$

with which we can also denote a previously discussed toggle variable Q: Only in the context $Q = q$ and *$C = c$, X becomes independent of C, as A_{CX} becomes freely instantiable. For this situation we write*

$$P(X|C, {}^{Q=q,C=c}\dot{A}_{CX}, Q, \boldsymbol{Z}) \ ,$$

where $\boldsymbol{Z}$ represents further dependencies of X, but without X, Q, C and A_{CX}.

Notation 3 (Innocuousness Context Vectors) *Variables might become innocuous in multiple contexts. Multiple innocuousness contexts $\varphi_{A_{YX}}$ of one activator A_{YX} are encapsulated in a vector $\boldsymbol{\varphi}_{A_{YX}}$ and are delimited by ; . An innocuousness context vector $\boldsymbol{\varphi}_{A_{YX}}$ can also be seen as a Boolean formula, where all contexts are disjunctions and a context is a conjunction of instantiations.*

This notation allows to mark contexts, in which an activator becomes irrelevant and could have been chosen to be deactive, and thus modifies the topology. Definition 4 describes the explicit specification of innocuousness in CPDs.

Definition 4 (Activator Innocuousness). *Let $\Phi_{A_{YX}}$ be the vector of random variables used in a context $\varphi_{A_{YX}}$ associated with A_{YX}. Every innocuousness context $\varphi_{A_{YX}} \in \boldsymbol{\varphi}_{A_{YX}}$ is then defined to hold*

$$\begin{aligned} &\forall x \in \mathrm{dom}(X), \forall \boldsymbol{z} \in \mathrm{dom}(\boldsymbol{Z}) : P(x|\varphi_{A_{YX}}, {}^{+}a_{YX}, \boldsymbol{z}) = P(x|\varphi_{A_{YX}}, \neg a_{YX}, \boldsymbol{z}) \\ &= P(x|\{\varphi_{A_{YX}} \backslash y \in \mathrm{dom}(Y)\}, y, \neg a_{YX}, \boldsymbol{z}) = P(x|\{\varphi_{A_{YX}} \backslash y\}, *, \neg a_{YX}, \boldsymbol{z}) \ , \end{aligned} \tag{5}$$

with remaining arbitrary dependencies of X on other random variables $\boldsymbol{Z}$ and $\boldsymbol{z}$ as an arbitrary instantiation of those, excluding A_{YX} and $\Phi_{A_{YX}}$.

Frankly, with Def. 4 we can formulate the same CSIs as Boutilier et al. [2], but, further, we can specify previously mentioned self-reflexive independences.We are thus able to explicitly express $P(X|\{\varphi_{A_{YX}}\backslash y\}, y, A_{YX}, \boldsymbol{Z}) = P(X|\{\varphi_{A_{YX}}\backslash y\}, A_{YX}, \boldsymbol{Z})$ as demonstrated in the following example.

Example 3 (Activator Innocuousness). Continuing Example 2, let us assume a noisy-or combination function for each CPD of a state X^t. With a noisy-or combination, every activator random variable (a message transfer) $\dot{M}^t_{XY}$ stands in the innocuousness context $\varphi_{M^t_{XY}} = \neg x^t$. We can now explicitly represent that Claire is not influenced by a non-credulous Earl, i.e., $P(C^t|C^{t-1}, D^t, \neg e^t, A^t_{DC}, A^t_{EC}) = P(C^t|C^{t-1}, D^t, A^t_{DC}, A^t_{EC})$, by fixing

$$\begin{aligned}&\forall\ C^t, C^{t-1}, D^t, A^t_{DC}, A^t_{EC} :\\ &\quad P(C^t|C^{t-1}, D^t, \neg e^t, A^t_{DC}, {+a}^t_{EC}) = P(C^t|C^{t-1}, D^t, \neg e^t, A^t_{DC}, \neg a^t_{EC})\\ &\qquad\qquad\qquad\qquad\qquad\qquad \overset{\text{(by Def. 2)}}{=} P(C^t|C^{t-1}, D^t, {+e}^t, A^t_{DC}, \neg a^t_{EC})\end{aligned}$$

in the respective CPD specification of C^t (likewise for non-creduluos Don). ▲

We see that an arc in $B_{\rightarrow}$ representing a dependency of X on Y can become vacuous in a context of the variable Y itself, which was previously impossible to formalize and impossible to define in a CPD without activators. This is beneficial for more efficient reasoning and a higher causal accuracy of independence declarations in all DBNs with activator random variables.

Further, Motzek and Möller [7] did not consider any properties of CPDs for possible acyclicity constraints in ADBNs, and only focus on defined activator sets. In the next section, we consider innocuousness properties of CPDs and relax restrictions posed on graphical models.

4 Exploiting Innocousness

By considering properties of CPDs of state variables $\boldsymbol{X}^t$, we relax restrictions of Thm. 1 by supporting innocuousness contexts as further acyclicity constraints. Note that in an ADBN, these checks and constraints are only sufficient conditions for achieving sound results and are not required for necessary calculations, if, e.g., observations can be trusted to fulfill these restrictions.

Theorem 2 (Bayesian Network Soundness Revised). *For every set of arbitrary instantiation of $\boldsymbol{\mathcal{A}}^{1:t}$ and $\boldsymbol{X}^{0:t}$, written $\boldsymbol{\mathcal{A}_*}^{1:t}$, $\boldsymbol{X}^{0:t}_*$, an ADBN $(B_0, B'_{\rightarrow})$ corresponds to a Bayesian network, if for all t, $\boldsymbol{\mathcal{A}_*}^t$ and $\boldsymbol{X}^t_*$ satisfy the new acyclicity constraint:*

$$\begin{aligned}\forall x, y, z \in \boldsymbol{X}^t : \mathfrak{A}(x,z)^t, \mathfrak{A}(z,y)^t \rightarrow \mathfrak{A}(x,y)^t\\ \neg\exists q : \mathfrak{A}(q,q)^t\ ,\end{aligned} \tag{6}$$

with a predicate $\mathfrak{A}(i,j)^t$ defined as

$$\mathfrak{A}(i,j)^t = \begin{cases} \mathit{false} & \mathit{if} \quad \neg a^t_{ij} \vee \boldsymbol{\varphi}_{A^t_{ij}} \\ \mathit{true} & \mathit{otherwise} \end{cases}\ ,$$

with the innocuousness context vector $\boldsymbol{\varphi}_{A_{ij}^t}$ seen as a disjunction of multiple contexts $\varphi_{A_{ij}^t}$ for activator A_{ij}^t, as defined in Def. 4. Given a correspondence to a Bayesian network an ADBN's semantics is well-defined and the complete joint probability over all random variables is straightforwardly specified by the product of all locally defined CPDs,

$$P(\boldsymbol{X}^{0:t^{\intercal}}, \boldsymbol{\mathcal{A}}^{1:t^{\intercal}}) = P(\boldsymbol{X}^{0:t-1^{\intercal}}, \boldsymbol{\mathcal{A}}^{1:t-1^{\intercal}}) \cdot \prod_i P(X_i^t | \boldsymbol{X}^{t^{\intercal}} \backslash X_i^t, \dot{\boldsymbol{A}}_i^{t^{\intercal}}, X_i^{t-1}) \cdot P(\boldsymbol{\mathcal{A}}^{t^{\intercal}}) \ . \quad (7)$$

Theorem 2 means, every (cyclic) ADBN corresponds to a sound Bayesian network, if only certain combinations of $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})$ are instantiated. Minimal sets of observations, i.e., partial instantiations of $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})$, have to enforce that during inference only valid combinations are used.

In the following, we prove Thm. 2 by showing that any instantiation of $\boldsymbol{\mathcal{A}}^{1:t}$, $\boldsymbol{X}^{0:t}$ holding Eq. 6 is topologically equivalent to some instantiation of $\boldsymbol{\mathcal{A}}^{1:t}$ holding Eq. 3, and thus is a valid Bayesian network with straightforward joint probability.

Proof (Theorem 2). According to Thm. 1 every (cyclic) ADBN is a Bayesian network and its semantic joint probability is well-defined as the product of all locally defined CPDs, if an instantiation of $\boldsymbol{\mathcal{A}}^{1:t}$ holds Eq. 3. Motzek and Möller [7, Proof 1] show, that for every of such combinations, a topological order $\prec$ must exist, s.t. by reversing Bayes' chain rule in Eq. 7 we obtain a joint probability distribution, which belongs to a valid Bayesian network.

Definition 5 (Topology Equivalence). *Given an ADBN $(B_0, B_{\rightarrow})$, an instantiation $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})_1$ is topologically equivalent to an instantiation $(\emptyset, \boldsymbol{\mathcal{A}}^{1:t})_2$, if for both the same topological order $\prec$ exists in $(B_0, B_{\rightarrow})$.*

Generally, in an acyclic ADBN, for every arbitrary instantiation $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})_*$ the same topological order $\prec_*$ exists. In a cyclic ADBN, the topological order is defined (at "runtime") by a minimal set of deactive $\boldsymbol{\mathcal{A}}^{1:t}$ holding to Thm. 1. In that case, some state variables $\boldsymbol{X}^t$ become independent of state variables $\boldsymbol{X}_E^t$, which previously created cycles and prohibited a topological order $\prec$. Note that the set of activators $\boldsymbol{\mathcal{A}}^{1:t}$ only are necessary conditions for creating a topological order and only follow a lexicographic order. However, under Def. 4, an active activator A_+^t might stand in a context $\varphi_{A_+^t}$, which renders A_+^t innocuous or irrelevant. It is straightforward from Def. 4 that A_+^t can then be seen as deactive from a topological perspective, which we call topologically deactive. Two sets of instantiations $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})_1$, $(\emptyset, \boldsymbol{\mathcal{A}}^{1:t})_2$ then share the same topological order $\prec$, if the set of topologically-deactive activators in $(\boldsymbol{\mathcal{A}}^{1:t})_1$ is a superset of deactive activators in $(\boldsymbol{\mathcal{A}}^{1:t})_2$ and a topological order exists for $(\boldsymbol{\mathcal{A}}^{1:t})_2$ (i.e., holds Thm. 1).

With Def. 5, every instantiation $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})_1$ holding Thm. 2 is *topologically* equivalent to an instantiation $(\emptyset, \boldsymbol{\mathcal{A}}^{1:t})_2$ holding Thm. 1 for which a joint probability function is well-defined based on the same topological order $\prec$. Under this topological order $(B_0, B_{\rightarrow})$ is a Bayesian network and Proof 1 in [7] is analogous. □

Note that while the joint probability of two topologically equivalent instantiations follows the same topological order, the results/outcomes of both must not be the same, as we need to consider priors of $\boldsymbol{\mathcal{A}}^{1:t}$.

Theorem 2 shows and it is proven that ADBNs cannot only be based on cyclic graphs and handle *acyclic* activator observations, but also *cyclic* activator observations can be made when considering specific CPD properties. The following example demonstrates the observation of a cyclic activator constellation.

Example 4 (Restriction Relaxation Example). Continuing Ex. 3, we now, observe $\neg e^1$ (in addition to previous observations), but $+m^1_{ED}$. The observations $+m^1_{ED}$, $+m^1_{DE}$ obviously lead to a cycle, which is not allowed according to Thm. 1 and is neither allowed in diagonal networks (Prop. 1). However, we find that the observation $\neg e^1$ meets the innocuousness context $\varphi_{M^1_{ED}}$ (Ex. 3, noisy-or), i.e., E^1 is innocuous for D^1 given $\neg e^1$. Therefore, this observation fixes $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})$ to instantiations that an ADBN can handle (Thm. 2). In fact, observations in this example fix all instantiations of $(\boldsymbol{X}^{0:t}, \boldsymbol{\mathcal{A}}^{1:t})$ to be *topologically equivalent* to the one from the previous example, i.e., both observations are topologically equivalent.

Note that still this observation cannot be handled by a diagonal alternative, as we need to anticipate an indirect influence: $\neg e^1$ tells us *indirectly* something about C^1, e.g., that $\neg c^1$ is now more likely than without the new observations. ▲

This example shows that the cyclic model can handle a larger set of observation constellations in contrast to a diagonal alternative. The next section generalizes these advantages for a general model.

5 Discussion and Comparison

In this section, we investigate how cyclic ADBNs compare to classic diagonal (A)DBNs. As discussed before, only certain combinations of $(\boldsymbol{X}^t, \boldsymbol{\mathcal{A}}^t)$ in a timestep t lead to valid Bayesian networks, which means (partial) observations of $(\boldsymbol{X}^t, \boldsymbol{\mathcal{A}}^t)$ have to fulfill certain restrictions. Further, we explore how the exploitation of innocuousness properties can relax these restrictions. We find that this exploitation significantly allows for more observation sets to be handled, and that cyclic ADBNs heavily outperform their diagonal counterparts w.r.t. expressivity.

In cyclic ADBNs, instantiations of $(\boldsymbol{X}^t, \boldsymbol{\mathcal{A}}^t)$ during a timestep t were restricted according to Thm. 1 and are relaxed due to Thm. 2. For diagonal ADBNs, instantiations are restricted, s.t. no indirect influences can occur (Prop. 1). Notwithstanding, innocuousness properties can also relax this restriction. For a comparison, let us consider Ex. 1 consisting of N employees, i.e., state variables $\boldsymbol{X}^t$, and likewise $N(N-1)$ message exchange variables in every network fragment $B_{\rightarrow}$.

Without considering CPD innocuousness properties, i.e., we do not exploit contexts from $\boldsymbol{X}^t$, we find that the number of possible $\boldsymbol{\mathcal{A}}^t$ combinations in a cyclic ADBN corresponds the number of DAGs [10, Seq. A003024]. In a classic diagonal ADBN no indirect effects are anticipated, and thus, no "interlocking" (possibly active) activator combinations of $\boldsymbol{\mathcal{A}}^t$ are allowed. We find this as the number of uniformly directed bipartite graphs, where isolated nodes belong to a fixed group [10, Seq. $A001831$]. For every of these combinations we have 2^N combinations of all $\boldsymbol{X}^t$.

To emphasize the effect of exploiting innocuousness context, we consider that $q\%$ out of all N state variables $\boldsymbol{X}^t$ in an ADBN fragment $B'_{\rightarrow}$ are innocuous states $\boldsymbol{X}_Q$, meaning that every state X_i^t "is not harmed" by any of these $X_Q^t \in \boldsymbol{X_Q}^t$ if $\neg x_Q^t$. This implies that every activator $\dot{A}_{ij}^t$ has the context $\varphi_{A_{ij}^t} = \neg x_i^t$, if $X_i^t \in \boldsymbol{X_Q}^t$. Thus, $\text{rank}(\boldsymbol{X_Q}^t) = Q = \lfloor N \cdot q \rfloor$, for which flooring operations lead to wavy lines in Fig. 2.

In an cyclic ADBN we obtain the total number $\mathcal{N}^{\mathcal{O}}{}_{N,Q}$ of allowed combinations $(\boldsymbol{X}^t, \boldsymbol{\mathcal{A}}^t)$ in a timestep t with Q innocuous nodes according to Thm. 2 as

$$\mathcal{N}^{\mathcal{O}}{}_{N,Q} = 2^{N-Q} \cdot \sum_{k=0}^{Q} 2^{k(N-1+N-k)} \cdot \text{A003024}_{N-k} \cdot \binom{Q}{k} . \tag{8}$$

$\mathcal{N}^{\mathcal{O}}{}_{N,Q}$ origins from the consideration that we can have between $k = 0$ to $k = Q$ "deactive" innocuous nodes. Thus, activators between $N-k$ nodes are still bound to DAG combinations, for which we have A003024_{N-k} many with 2^{N-Q} instantiations of $\boldsymbol{X}^t$. For every of those DAG combinations we have k deactive nodes, whose $N-1$ activators are free, i.e. $2^{k(N-1)}$ combinations, and $N-k$ active nodes, whose activators with the k deactive nodes are free, i.e. $2^{(N-k)k}$ further combinations. For each combination, we have $\binom{Q}{k}$ options to choose which (labeled) innocuous states are deactive.

Notwithstanding, the restriction that only indirect-free combinations of $\boldsymbol{\mathcal{A}}^t$ are allowed in diagonal ADBNs (Prop. 1) is also relaxed by considering innocuousness properties of $\boldsymbol{X}^t$. To enumerate these, we need the number of uniformly directed bipartite graphs with groups of size n, m, $N = n + m$, which is

$$A001831'_{N,n} = \binom{N}{n} \cdot (2^n - 1)^{N-n} . \tag{9}$$

With Q innocuous-nodes we then find the total number $\mathcal{N}^{/}{}_{N,Q}$ of allowed combinations in diagonal (A)DBNs to be

$$\mathcal{N}^{/}{}_{N,Q} = 2^{N-Q} \cdot \sum_{k=0}^{Q} \sum_{n=0}^{N-k} 2^{k(N+n-1)} \cdot A001831'_{N-k,n} \cdot \binom{Q}{k} \tag{10}$$

$\mathcal{N}^{/}{}_{N,Q}$ origins from the same considerations as $\mathcal{N}^{\mathcal{O}}{}_{N,Q}$, but the activators of the k deactive nodes are not completely free anymore. As (active) activators of the second group of size m interlock with activators of the deactive nodes.

Figure 2 shows a comparison of $\mathcal{N}^{\mathcal{O}}{}_{N,Q}$ and $\mathcal{N}^{/}{}_{N,Q}$ for $0 < N \leq 25$ and different Q. Note that even in a logarithmic plot, a cyclic ADBN has an exponential advantage in favor of a classic acyclic (A)DBN.

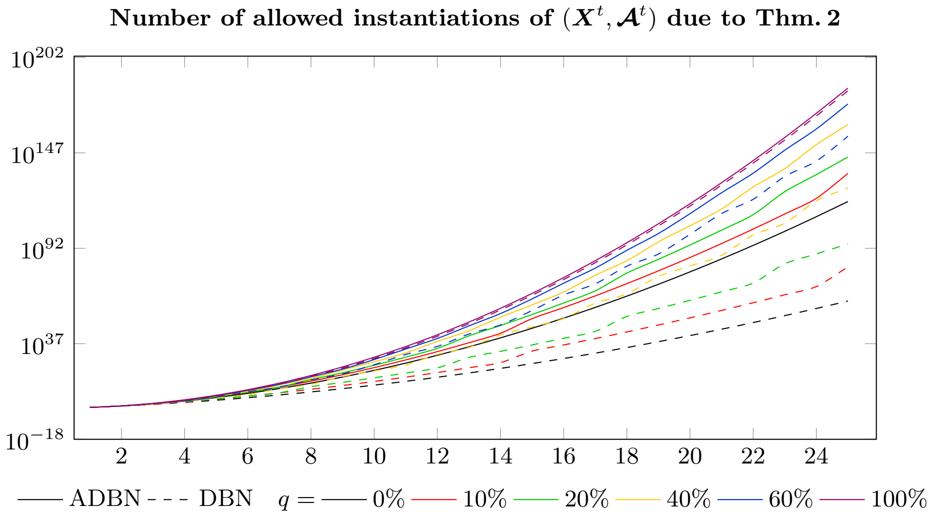

Fig. 2. Cyclic ADBNs ($\mathcal{N}^{\mathcal{O}}{}_{N,Q}$, solid) clearly outperform classic diagonal DBNs ($\mathcal{N}^{/}{}_{N,Q}$, dashed) in the number of allowed instantiations of $(\boldsymbol{X}^t, \boldsymbol{\mathcal{A}}^t)$. Note that for a full noisy-or network (100%) all possible graph structures ($\boldsymbol{\mathcal{A}}^t$ combinations) are allowed in the case of $\forall i \neg x_i$, which draws $\mathcal{N}^{/}{}_{N,N}$ near $\mathcal{N}^{\mathcal{O}}{}_{N,N}$. Still, even in this extreme case a cyclic ADBN outperforms a classic DBN by two orders of magnitude (*semi-logarithmic plot*).

6 Conclusion

In this paper we have formalized an innocuousness property of random variables, which is often associated with Boolean combination functions for general CPDs. Based on a formalization with random variables taking the role of activators, we relax restrictions on graphical models for the use in Bayesian networks and have given a quantitative evaluation of restrictions posed on such networks. This is beneficial for working with graphical models representing a process over time requiring the anticipation of indirect influences under a free choice of temporal granularity. Further, by providing a formal definition for innocuousness in CPDs, we gain the ability to formally represent that in specific contexts a dependency is causally irrelevant, opening new ways for more efficient inference and a higher causal accuracy in specifying CPDs.

Still, like in any other DBN, operations remain computationally intractable with respect to dimension complexity (number of state variables), and this demands approximate inference techniques. Considering our formalization that certain dependencies, i.e. arcs, are irrelevant in specific situations and an resulting BN might turn out to be singly connected, approximate inference techniques can heavily benefit from ADBNs and, here newly defined, innocuousness properties. Future work is dedicated to new inference techniques and extensions to relational Bayesian networks [6].

Acknowledgments

This work was partly supported by the Seventh Framework Programme of the European Commission as part of the PANOPTESEC integrated research project (GA 610416).

References

1. Antonucci, A.: The Imprecise Noisy-OR Gate. In: 14th International Conference on Information Fusion. pp. 1–7. IEEE (2011)
2. Boutilier, C., Friedman, N., Goldszmidt, M., Koller, D.: Context-Specific Independence in Bayesian Networks. In: 12th Conference on Uncertainty in Artificial Intelligence. pp. 115–123 (1996)
3. Cozman, F.G.: Axiomatizing Noisy-OR. In: 16th Eureopean Conference on Artificial Intelligence. p. 979 (2004)
4. Heckerman, D., Breese, J.S.: Causal Independence for Probability Assessment and Inference Using Bayesian Networks. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 26(6), 826–831 (1996)
5. Henrion, M.: Practical Issues in Constructing a Bayes Belief Network. International Journal of Approximate Reasoning 2(3), 337 (1988)
6. Jaeger, M.: Relational Bayesian Networks. In: 13th Conference on Uncertainty in Artificial Intelligence. pp. 266–273 (1997)
7. Motzek, A., Möller, R.: Indirect Causes in Dynamic Bayesian Networks Revisited. In: 24th International Joint Conference on Artificial Intelligence. pp. 703–709. AAAI (2015)
8. Pearl, J.: Reasoning with Cause and Effect. AI Magazine 23(1), 1–83 (2002)
9. Poole, D., Zhang, N.L.: Exploiting Contextual Independence in Probabilistic Inference. Journal Of Artificial Intelligence Research 18, 263–313 (2003)
10. Sloane, N.J.A.: The On-Line Encyclopedia of Integer Sequences. http://oeis.org/. OEIS Foundation Inc. (2015), Sequences A003024 & A001831.
11. Srinivas, S.: A Generalization of the Noisy-OR Model. In: 9th International Conference on Uncertainty in Artificial Intelligence. pp. 208–215 (1993)
12. Zagorecki, A., Druzdzel, M.J.: Probabilistic Independence of Causal Influences. In: 3rd European Workshop on Probabilistic Graphical Models. pp. 325–332 (2006)
13. Zhang, N.L., Poole, D.: Exploiting Causal Independence in Bayesian Network Inference. Journal of Artificial Intelligence Research 5, 301–328 (1996)