De-Causalizing NAT-Modeled Bayesian Networks for Inference Efficiency

Yang Xiang and Dylan Loker

University of Guelph, Canada

Abstract. Conditional independence encoded in Bayesian networks (BNs) avoids combinatorial explosion on the number of variables. However, BNs are still subject to exponential growth of space and inference time on the number of causes per effect variable in each conditional probability table (CPT). A number of space-efficient local models exist that allow efficient encoding of dependency between an effect and its causes, and can also be exploited for improved inference efficiency. We focus on the Non-Impeding Noisy-AND Tree (NIN-AND Tree or NAT) models due to its multiple merits. In this work, we develop a novel framework, *de-causalization* of NAT-modeled BNs, by which causal independence in NAT models can be exploited for more efficient inference. We demonstrate its exactness and efficiency impact on inference based on lazy propagation (LP).

1 Introduction

Conditional independence encoded in BNs avoids combinatorial explosion on the number of variables. However, BNs are still subject to exponential growth of space and inference time on the number of causes per effect variable in each CPT. A number of space-efficient local models exist that allow efficient encoding of dependency between an effect and its causes. They include noisy-OR [9], noisy-MAX [4, 2], context-specific independence (CSI) [1], recursive noisy-OR [5], Non-Impeding Noisy-AND Tree (NIN-AND Tree or NAT) [13], DeMorgan [6], tensor-decomposition [10], and cancellation model [11].

We consider expressing BN CPTs as, or compressing them into, NAT models [13]. The merits of NAT models include being based on simple causal interactions (reinforcement and undermining), expressiveness (recursive mixture, multi-valued), and generality (generalizing noisy-OR, noisy-MAX [15], and DeMorgan [13]). Since causal independence encoded in a NAT model is orthogonal to CSI, NAT models provide an alternative to CSI for efficient local modeling in BNs.

Local models not only reduce space and time to acquire CPT parameters, they can also be exploited to improve inference efficiency. Through multiplicative factorization (MF) of NAT-modeled BNs, inference based on LP [7] was made up to two orders of magnitude faster in very sparse BNs [15]. In this work, we develop a framework alternative to MF which is referred to as *de-causalization* of NAT-modeled BNs, where causal independence in NAT models can be exploited for more efficient inference. We also evaluate its impact on LP efficiency.

The remainder of the paper is organized as follows. Section 2 reviews the background on NAT modeling. In Sections 3 and 4, we present how to de-causalize dual and direct NIN-AND gate models. The tree-width of the de-causalized representation is further reduced in Section 5. This de-causalization is extended to

a general NAT model in Section 6 and then to a NAT-modeled BN in Section 7. The impact of de-causalization is empirically evaluated in Section 8.

2 Background on NAT Models

This section briefly reviews background on NAT models. More details can be found in [13, 14]. A NAT model is defined over an effect e and a set of n causes $C = \{c_1, ..., c_n\}$ that are multi-valued, where $e \in D_e = \{e^0, ..., e^\eta\}$ ($\eta \geq 1$) and $c_i \in \{c_i^0, ..., c_i^{m_i}\}$ ($i = 1, ..., n, m_i \geq 1$). C and e form one family (a child variable plus its parents) in a BN, whose dependence is quantified by a CPT by default. Values e^0 and c_i^0 are *inactive*. Other values (may be written as e^+ or c_i^+) are *active*. A higher index often means higher intensity (graded or ordinal variables), but that is not necessary (see [14] for generalization to nominal variables).

A causal event is a *success* or *failure* depending on if e is active up to a given value, is *single-* or *multi-causal* depending on the number of active causes, and is *simple* or *congregate* depending on value range of e. For instance, $P(e^k \leftarrow c_i^j) = P(e^k | c_i^j, c_z^0 : \forall z \neq i)$ ($j > 0$) is probability of a *simple single-causal success*, and

$$P(e \geq e^k \leftarrow c_1^{j_1}, ..., c_q^{j_q}) = P(e \geq e^k | c_1^{j_1}, ..., c_q^{j_q}, c_z^0 : c_z \in C \setminus X)$$

is probability of a *congregate multi-causal success*, where $j_1, ..., j_q > 0$, $X = \{c_1, ..., c_q\}$ ($q > 1$). The latter may be denoted as $P(e \geq e^k \leftarrow \underline{x}^+)$. Interactions among causes may be reinforcing or undermining as defined below.

Definition 1 *Let e^k be an active effect value, $R = \{W_1, ..., W_m\}$ ($m \geq 2$) be a partition of a set $X \subseteq C$ of causes, $S \subset R$, and $Y = \cup_{W_i \in S} W_i$. Sets of causes in R reinforce each other relative to e^k, iff $\forall S \, P(e \geq e^k \leftarrow \underline{y}^+) \leq P(e \geq e^k \leftarrow \underline{x}^+)$. They undermine each other iff $\forall S \, P(e \geq e^k \leftarrow \underline{y}^+) > P(e \geq e^k \leftarrow \underline{x}^+)$.*
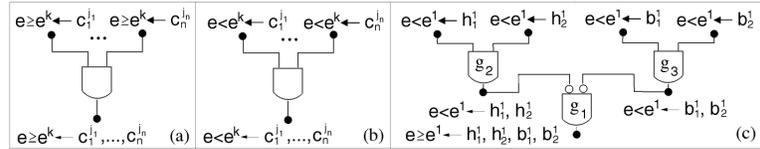
Fig. 1. A direct NIN-AND gate (a), a dual NIN-AND gate (b), and a NAT (c).

A NAT has multiple NIN-AND gates. A *direct* gate involves disjoint sets of causes $W_1, ..., W_m$. Each input event is a success $e \geq e^k \leftarrow \underline{w}_i^+$ ($i = 1, ..., m$), e.g., Fig. 1 (a) where each W_i is a singleton. The output event is $e \geq e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+$. The probability of output event of a direct NIN-AND gate is

$$P(e \geq e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+) = \prod_{i=1}^{m} P(e \geq e^k \leftarrow \underline{w}_i^+). \tag{1}$$

Direct gates encode undermining causal interactions. Each input event of a *dual* gate is a failure $e < e^k \leftarrow \underline{w}_i^+$, e.g., Fig. 1 (b). The output event is $e < e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+$. The probability of output event of a dual NIN-AND gate is

$$P(e < e^k \leftarrow \underline{w}_1^+, ..., \underline{w}_m^+) = \prod_{i=1}^{m} P(e < e^k \leftarrow \underline{w}_i^+). \tag{2}$$

Dual gates encode reinforcement causal interactions.

Fig. 1 (c) shows a NAT, where causes h_1 and h_2 reinforce each other, and so do b_1 and b_2. However, the two groups undermine each other. From the NAT and probabilities of its input events, in the general form $P(e^k \leftarrow c_i^j)$ $(j, k > 0)$, called *single-causals*, $P(e \geq e^1 \leftarrow h_1^1, h_2^1, b_1^1, b_2^1)$ can be obtained. From the single-causals and all derivable NATs [12], CPT $P(e|h_1, h_2, b_1, b_2)$ is uniquely specified [13]. A NAT model is specified by the topology and a set of single-causals with a space linear on n.

A BN where the CPT of every family of size 3 or larger is a NAT model is a *NAT-modeled BN*. A discrete BN where every CPT is tabular has a space complexity of $O(N \kappa^n)$, where N is the number of variables, κ is the size of largest variable domains, and $n+1$ is the largest family size. On the other hand, a NAT-modeled BN has a linear space complexity of $O(N \kappa n)$. The efficiency of NAT-modeled BNs can extend to inference. A MF framework has been developed [15], where each NAT model in the BN is converted into a hybrid network segment by exploiting causal independence. The multiplicatively factorized NAT-model BN allows up to two orders of magnitude speedup in LP for very sparse BNs.

3 De-Causalizing Dual NIN-AND Gate Models

First, we de-causalize a dual NIN-AND gate model, a building block of NAT models. It has been shown that a dual NIN-AND gate over multi-valued variables is equivalent to noisy-MAX [15]. A BN segment can be used [3] to structure noisy-MAX models. Given the equivalence between a dual NIN-AND gate and a noisy-MAX model, any structuring of noisy-MAX is also applicable to dual gates. Nevertheless, we give below an alternative justification of the mapping of dual gates to the BN segment, that is direct and hence more intuitive.
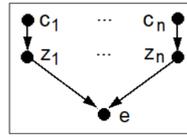
Fig. 2. The DAG structure of BN segment of an NIN-AND gate model.

Fig. 2 shows the BN segment structure, where root variables are the n causes and the leaf is the effect e. For each cause c_i, a probabilistic auxiliary child variable z_i is introduced, whose domain is D_e. It represents the impact of cause c_i to effect e. The CPT at z_i, referred as *single-causal* (SC) CPT, is

$$P(z_i = e^j | c_i = c_i^k) = \begin{cases} 1, & \text{if } e^j = e^0 \text{ and } c_i^k = c_i^0, \\ P(e^j \leftarrow c_i^k), & \text{if } e^j > e^0 \text{ and } c_i^k > c_i^0. \end{cases} \quad (3)$$

The 1st formula says that when c_i is inactive, it cannot render e active. The 2nd formula expresses the impact to e when c_i is active. The CPT at e, referred to as a *MAX* CPT, encodes a *MAX* function as follows, where the domain of every variable is D_e and the number of α_i variables is finite.

$$P(\tau | \alpha_1, \alpha_2, ...) = \begin{cases} 1, \text{ if } \tau = MAX(\alpha_1, \alpha_2, ...), \\ 0, \text{ otherwise}. \end{cases} \quad (4)$$

For the MAX CPT at e, τ is substituted by e and $\alpha_1, \alpha_2, ...$ by $z_1, ..., z_n$.

Definition 2 *Let G be the DAG in Fig. 2 over $C = \{c_1, ..., c_n\}$ and e, and CP be the set of CPTs specified by Eqns. (3) and (4). Then $\Phi = (C, e, G, CP)$ is the* **BN segment for a dual NIN-AND gate model**.

We show below that the BN segment Φ is equivalent to the dual gate model illustrated in Fig. 1 (b).

Proposition 1 *Let $\Phi = (C, e, G, CP)$ be a BN segment for a dual NIN-AND gate model. Then the CPT $P(e|c_1, ..., c_n)$ from Φ defined by the marginalized product*

$$\sum_{z_1, ..., z_n} (P(e|z_1, ..., z_n) \prod_{i=1}^{n} P(z_i|c_i))$$

is the same as that defined by the dual NIN-AND gate model.

Proof: The dual NIN-AND gate model is characterized by Eqn. (2). When each set of causes is a singleton, Eqn. (2) becomes the following where, without losing generality, $c_1, ..., c_m$ are active and $c_{m+1}, ..., c_n$ are inactive:

$$P(e < e^k \leftarrow c_1^+, ..., c_m^+) = \prod_{i=1}^{m} P(e < e^k \leftarrow c_i^+), \quad (k = 1, ..., \eta).$$

It is equivalent to

$$P(e \leq e^k \leftarrow c_1^+, ..., c_m^+) = \prod_{i=1}^{m} P(e \leq e^k \leftarrow c_i^+), \quad (k = 0, ..., \eta - 1),$$

which is a cumulative causal distribution. If Φ has the same cumulative distribution (which we show below), then $P(e|c_1, ..., c_n)$ from Φ is also the same as that of the dual gate model.

Assume that $c_1, ..., c_m$ are active and $c_{m+1}, ..., c_n$ are inactive. In Φ, since the CPT by Eqn. (4) encodes the *MAX* function, we have

$$P(e \leq e^k \leftarrow c_1^+, ..., c_m^+) = \sum_{MAX(z_1, ..., z_n) \leq e^k} P(z_1, ..., z_n | c_1^+, ..., c_m^+, c_{m+1}^0, ..., c_n^0).$$

Since $MAX(z_1, ..., z_n) \leq e^k$ iff $z_i \leq e^k$ for $i = 1, ..., n$, the above is equal to

$$\sum_{z_1 \leq e^k, ..., z_n \leq e^k} P(z_1, ..., z_n | c_1^+, ..., c_m^+, c_{m+1}^0, ..., c_n^0)$$

$$= \sum_{z_1 \leq e^k} \cdots \sum_{z_n \leq e^k} P(z_1, ..., z_n | c_1^+, ..., c_m^+, c_{m+1}^0, ..., c_n^0).$$

By the DAG structure of Φ, z_i is independent of z_j for $i \neq j$ given c_i. Hence, the above equals

$$\sum_{z_1 \leq e^k} \cdots \sum_{z_n \leq e^k} (P(z_1|c_1^+) \cdots P(z_m|c_m^+) P(z_{m+1}|c_{m+1}^0) \cdots P(z_n|c_n^0))$$

$$= \sum_{z_1 \leq e^k} P(z_1|c_1^+) \cdots \sum_{z_m \leq e^k} P(z_m|c_m^+) \cdots \sum_{z_{m+1} \leq e^k} P(z_{m+1}|c_{m+1}^0) \cdots \sum_{z_n \leq e^k} P(z_n|c_n^0).$$

Since $\sum_{z_i \leq e^k} P(z_i|c_i^0) = 1$ for $i = m+1, ..., n$, the above equals

$$\sum_{z_1 \leq e^k} P(z_1|c_1^+) \; ... \sum_{z_n \leq e^k} P(z_m|c_m^+) = \prod_{i=1}^{m} P(z_i \leq e^k \leftarrow c_i^+).$$

From Eqn. (3), the above equals $\prod_{i=1}^{m} P(e \leq e^k \leftarrow c_i^+)$. Hence, we have

$$P(e \leq e^k \leftarrow c_1^+, ..., c_m^+) = \prod_{i=1}^{m} P(e \leq e^k \leftarrow c_i^+). \qquad \square$$

4 De-Causalizing Direct NIN-AND Gate Models

Next, we de-causalize a direct NIN-AND gate model, another building block of NAT models. The structure of BN segment is the same as Fig. 2. However, the domain of each auxiliary variable z_i is $D_a = \{e^0, ..., e^\eta, aaci\}$, where an extra value *aaci* (all above causes inactive) is added to D_e. Its semantics are elaborated below. When values of z_i are compared, the relation $e^0 < ... < e^\eta < aaci$ is assumed. Note that events in Fig. 1 are causal events, while events in Fig. 2 are not. Hence the name *de-causalization*.

The CPT at z_i, referred to as a *single-causal-plus* (SC$^+$) CPT, is the following, where $^+$ signifies the enlarged domain of z_i beyond D_e:

$$\begin{cases} P(z_i = aaci | c_i = c_i^0) = 1, \\ P(z_i = e^j | c_i = c_i^k) = P(e^j \leftarrow c_i^k), \quad (e^j > e^0, c_i^k > c_i^0). \end{cases} \qquad (5)$$

The 1st formula explicitly signifies that c_i is inactive (the above cause is inactive). The 2nd formula covers all cases where c_i is active. The CPT at e, referred to as *PMIN* CPT, encodes a pseudo-MIN (PMIN) function below over a finite set of arguments, where each argument has domain D_a and function range is D_e:

$$PMIN(\alpha_1, \alpha_2, ...) = \begin{cases} e^0, & if \; \forall_i \; \alpha_i = aaci, \\ MIN(\alpha_1', ..., \alpha_m'), & if \; \alpha_1', ..., \alpha_m' \neq aaci \; (m > 0). \end{cases}$$

The PMIN CPT at e is the following:

$$P(\tau | \alpha_1, \alpha_2, ...) = \begin{cases} 1, \; if \; \forall_i \; \alpha_i = aaci \; \wedge \; \tau = e^0, \\ 1, \; if \; \alpha_1', ..., \alpha_m' \neq aaci \; (m > 0) \; \wedge \; \tau = MIN(\alpha_1', ..., \alpha_m'). \end{cases} \qquad (6)$$

We define the BN segment below and establish its soundness.

Definition 3 *Let G be the DAG in Fig. 2 over $C = \{c_1, ..., c_n\}$ and e, and CP be the set of CPTs specified by Eqns. (5) and (6). Then $\Phi = (C, e, G, CP)$ is the* **BN segment for a direct NIN-AND gate model***.*

We show below that the BN segment Φ is equivalent to the dual gate model illustrated in Fig. 1 (a). The proof is omitted due to space limit.

Proposition 2 *Let $\Phi = (C, e, G, CP)$ be a BN segment for a direct NIN-AND gate model. Then the CPT $P(e|c_1, ..., c_n)$ from Φ defined by the marginalized product*

$$\sum_{z_1,...,z_n} (P(e|z_1, ..., z_n) \prod_{i=1}^{n} P(z_i|c_i))$$

is the same as that defined by the direct NIN-AND gate model.

5 Reducing Tree-Width of BN Segment

The complexity of probabilistic reasoning with a BN is critically dependent on its tree-width. By reducing the tree-width of BN segments, the tree-width of a BN may also be reduced. The BN segments presented in previous sections have a tree-width of n. Below, we take advantage of deterministic CPTs in Eqns. (4) and (6), and apply parent divorcing [8] to reduce the tree-width of these BN segments from n to 2.

Fig. 3 shows the enhanced DAG structure. A total of $n - 2$ deterministic auxiliary variables y_i are introduced. Since the DAG is a directed tree where each node has no more than two parents, its tree-width is 2.
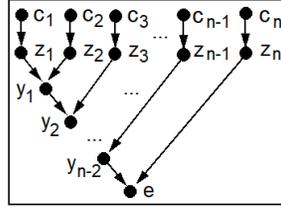
Fig. 3. The DAG structure of BN segment by applying parent divorcing.

For the enhanced BN segment of a dual gate model, the domain of each y_i is D_e. The CPT at each y_i ($i = 1, ..., n-2$) and e is a MAX CPT defined by Eqn. (4). It can be shown that the collection of CPTs at y_i and e is equivalent to the single MAX CPT $P(e|z_1, ..., z_n)$ described in Section 3. We omit the proof due to space considerations.

Assume that all cause variables have the same domain size $\eta + 1$ as e. The total size of the CPT collection is $(n - 1)(\eta + 1)^3$, while the single CPT has a size of $(\eta + 1)^{n+1}$. For $n = \eta = 4$, the two sizes are 375 and 3125.

For the enhanced BN segment of a direct gate model, the domain of each y_i is D_a. The CPT at e is a PMIN CPT defined by Eqn. (6), where condition variables are y_{n-2} and z_n. When one of y_{n-2} and z_n is not $aaci$, the MIN function is trivialized. The CPT at each y_i ($i = 1, ..., n-2$), referred to as a *PMIN$^+$* CPT, encodes the following pseudo-MIN-plus (PMIN$^+$) function:

$$PMIN^+(\alpha_1, \alpha_2) = \begin{cases} aaci, & if \; \alpha_i = aaci \; (i = 1, 2), \\ MIN(\alpha'_1, \alpha'_m), & if \; \alpha'_1, \alpha'_m \neq aaci \; (m > 0). \end{cases}$$

When $m = 1$, the MIN function is trivial. The PMIN$^+$ CPT at each y_i is the following, where τ is substituted by y_i, and α_i are substituted by parents of y_i:

$$P(\tau|\alpha_1, \alpha_2) = \begin{cases} 1, \; if \; \alpha_i = aaci \; (i = 1, 2) \; \wedge \; \tau = aaci, \\ 1, \; if \; \alpha'_1, \alpha'_m \neq aaci \; (m > 0) \; \wedge \; \tau = MIN(\alpha'_1, \alpha'_m). \end{cases} \quad (7)$$

The 1st formula signifies that all causes above y_i are inactive, so that the non-impeding behavior of a direct gate model is enabled. It can be shown that the collection of CPTs at y_i and e are equivalent to the single PMIN CPT described in Section 4. The size of the CPT collection is $(n-2)(\eta+2)^3 + (\eta+1)(\eta+2)^2$, while the single PMIN CPT has a size of $(\eta+1)(\eta+2)^n$. For $n = \eta = 4$, the two sizes are 612 and 6480.

6 De-Causalizing NAT Models

A NAT model generally consists of multiple NIN-AND gates organized into a tree. To de-causalize a general NAT model, we apply the BN segment for each gate and interface the gate segments so that the BN segment of the NAT model encodes the exact CPT of the NAT model. If an NIN-AND gate feeds into another in the NAT model, its effect variable is replaced with a *quasi-effect* variable.
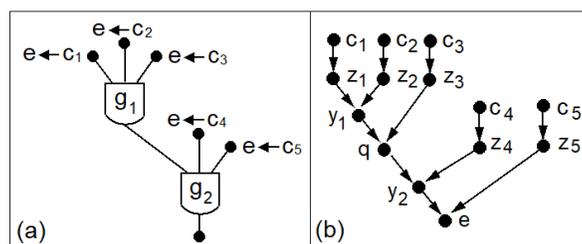
Fig. 4. (a) A NAT. (b) The enhanced BN segment.

Consider the NAT in Fig. 4 (a), where labels of causal events have been simplified (e.g., input events to gates) or omitted (e.g., output events). Suppose that the leaf gate g_2 is dual. Then g_1 is direct. The (enhanced) BN segment of the NAT is shown in (b). The BN segment of g_1 consists of causes variables c_i ($i = 1, 2, 3$), probabilistic auxiliary variables z_i ($i = 1, 2, 3$), deterministic auxiliary variable y_1, and quasi-effect variable q. This segment can be implemented as in Sections 4 and 5, except that the variable e there is renamed as q.

The BN segment of g_2 consists of causes variables c_i ($i = 4, 5$), quasi-effect variable q as an input from g_1, probabilistic auxiliary variables z_j ($j = 4, 5$), deterministic auxiliary variable y_2, and effect variable e. This segment can be implemented as in Sections 3 and 5, except that the quasi-effect variable q should be treated in the same way as probabilistic auxiliary variables z_j ($j = 4, 5$).

Next, suppose that the leaf gate g_2 is direct and g_1 is dual. The BN segment of the NAT is the same as in Fig. 4 (b). However, the BN segment of dual gate g_1 must be modified relative to that of Sections 3 and 5. In Sections 3 and 5, auxiliary variables z_i and y_i, as well as the effect e, have the domain D_e. This is no longer valid. Since g_1 is not the leaf gate, it now feeds into the direct gate g_2. To support non-impeding behavior of the direct gate, domains of z_i, y_i, and quasi-effect q have to be enlarged into D_a.

Due to this enlargement, SC CPTs cannot be applied to z_i ($i = 1, 2, 3$), and MAX CPTs cannot be applied to y_1 and q. Instead, auxiliary variables z_i ($i = 1, 2, 3$) adopt SC^+ CPTs (Eqn. (5)). A new form of CPT is needed for y_1 and q. It

is referred to as *PMAX$^+$* CPTs, and encodes the following pseudo-MAX-plus (PMAX$^+$) function, where domain of each argument and function range are D_a:

$$PMAX^+(\alpha_1, \alpha_2) = \begin{cases} aaci, & if \ \alpha_i = aaci \ (i = 1, 2), \\ MAX(\alpha'_1, \alpha'_m), & if \ \alpha'_1, \alpha'_m \neq aaci \ (m > 0). \end{cases}$$

The PMAX$^+$ CPTs at y_1 and q are the following:

$$P(\tau|\alpha_1, \alpha_2) = \begin{cases} 1, \ if \ \alpha_i = aaci \ (i = 1, 2) \ \wedge \ \tau = aaci, \\ 1, \ if \ \alpha'_1, \alpha'_m \neq aaci \ (m > 0) \ \wedge \ \tau = MAX(\alpha'_1, \alpha'_m). \end{cases} \quad (8)$$

The BN segment of the direct leaf gate g_2 can be encoded as Sections 4 and 5, except that the quasi-effect q should be treated in the same way as auxiliary variables z_4 and z_5.

Table 1. Summary of variable domains

		Auxiliary variable	Quasi-effect
Level 0	Dual gate	D_e	NA
	Direct gate	D_a	NA
Level 1	Dual gate	D_a	D_a
	Direct gate	D_a	D_e
Level 2+	Dual gate	D_a	D_a
	Direct gate	D_a	D_a

In general, domains of auxiliary variables (both probabilistic and deterministic) and quasi-effect variables should be set as summarized in Table 1, where level 0 is the leaf level. The primary criteria are to keep the domain as small as possible, while ensuring non-impeding behavior of direct gates.

CPTs for auxiliary, quasi-effect, and effect variables should be set as summarized in Table 2, where the last column refers to effect (level 0) or quasi-effect (level 1+). The primary criteria are to maintain exact CPT as corresponding NIN-AND gate model, while ensuring non-impeding behavior of direct gates downstream.

Table 2. Summary of variable CPTs

Level	Gate	Probabilistic aux	Deterministic aux	(Quasi)-effect
0	Dual	SC CPT	MAX CPT	MAX CPT
	Direct	SC$^+$ CPT	PMIN$^+$ CPT	PMIN CPT
1	Dual	SC$^+$ CPT	PMAX$^+$ CPT	PMAX$^+$ CPT
	Direct	SC$^+$ CPT	PMIN$^+$ CPT	PMIN CPT
2+	Dual	SC$^+$ CPT	PMAX$^+$ CPT	PMAX$^+$ CPT
	Direct	SC$^+$ CPT	PMIN$^+$ CPT	PMIN$^+$ CPT

It can be shown formally that the collection of CPTs in the BN segment of the NAT model ensures the exact $P(e|c_1, ..., c_n)$ of the NAT model. We omit the formal analysis due to space restriction. Instead, we demonstrate the exactness empirically in Section 8.

7 De-Causalizing NAT-Modeled Bayesian Networks

To de-causalize a NAT-modeled BN, for each NAT model family (child e plus parents $c_1, ..., c_n$), delete the directed link from each parent to the child (as well

as the CPT of the child), reconnect the family by the de-causalizing BN segment, and assign a CPT to each variable (except $c_1, ..., c_n$) as presented above.

Consider the example NAT-modeled BN in Fig. 5, where the NAT model over family of v_8 is shown with simplified labeling, and all variables are ternary. The gate g_3 is direct and remaining gates are dual.
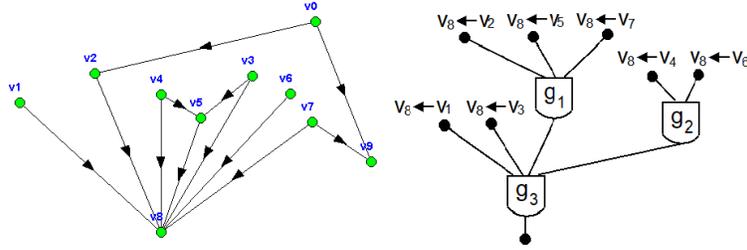
Fig. 5. Left: DAG of a NAT-model BN. Right: NAT-model over family of v_8.

The de-causalized BN is shown in Fig. 6. For causes v_i ($i = 1, ..., 7$) in that order, the probabilistic auxiliary variables are x_{10}, x_{16}, x_{11}, x_{20}, x_{17}, x_{21}, x_{18}, respectively. For gate g_2, the quasi-effect is q_{13}. For gate g_1, the deterministic auxiliary variable is y_{19} and the quasi-effect is q_{12}. For gate g_3, the deterministic auxiliary variables are y_{14} and y_{15}. A BN with the DAG in Fig. 5 (left), where all variables are ternary and all CPTs are tabular, has 6642 numerical parameters (values in all CPTs). The de-causalized BN has 489 parameters.
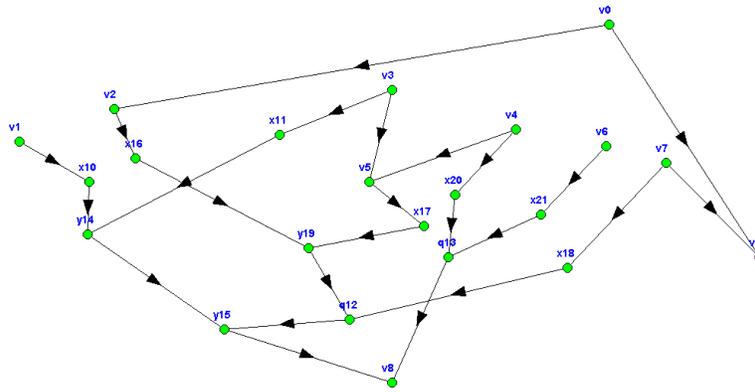
Fig. 6. The NAT-modeled BN in Fig. 5 after de-causalization.

Let a NAT-modeled BN be over the set V of variables and its de-causalized BN be over the set $V \cup W$ of variables, where W is the set of all auxiliary and quasi-effect variables. Let $P(V)$ be the joint probability distribution (JPD) of the NAT-modeled BN, and $P(V, W)$ be the JPD of the de-causalized BN. Since for each replaced BN family, the CPT $P(e|c_1, ..., c_n)$ specified by the de-causalizing seg-

ment is equal to the original NAT CPT of the family, we have

$$\sum_{w \in W} P(V, W) = P(V).$$

The de-causalized BN can be used for probabilistic reasoning using any standard inference algorithm. Only observations over variables in V can be entered, as variables in W are not observable. In Section 8, we demonstrate the posterior marginals thus computed from de-causalized BNs are exact as computed from original NAT-modeled BNs.

8 Experiments

The 1st experiment evaluates the space of a NAT model-CPT as a tabular CPT (TAB), of a de-causalized CPT without parent divorcing (DEC), and of a de-causalized CPT with parent divorcing (DPD). The numbers of causes per CPT are $n = 5, 7, 9, 11$. The uniform domain sizes of variables in each CPT are $d = 3, 5, 7$. For each combination of (n, d), 30 random NAT topologies are generated. Hence, a total of $4 * 3 * 30 = 360$ distinctly structured NAT models are evaluated.

Fig. 7 show spaces of CPTs in $log10$ by TAB, DEC, and DPD when $d = 7$. Due to space consideration, we omit presentation of result for $d = 3, 5$. Space of TAB CPTs are completely determined by n and d, and are constant. Spaces of both DEC and DPD CPTs are sensitive to the NAT topology, but DPD CPTs are only slightly so. DEC CPTs are often more space-efficient than TAB CPTs. But for some NAT topologies, they are less efficient. For instance, the 11th DEC CPT for $n = 11$ and $d = 7$ is less efficient than the TAB CPT, whose NAT has two gates and one of them has 10 inputs. DPD CPTs are always the most efficient, and are 5 orders of magnitude more efficient than TAB CPTs for $n = 11$ and $d = 7$.

Fig. 7. Spaces (s) of CPTs as TAB, DEC, and DPD, where $d = 7$.

The 2nd experiment evaluates the impact of de-causalization on inference efficiency, where the inference method is LP. We simulated NAT-modeled BNs with 100 variables per BN. The maximum number of parents per variable in each BN is bounded at $m = 6, 8, 10, 12$, respectively. The uniform domain size of all variables is controlled at $s = 2, 3$, respectively. The structural density of BNs is controlled by adding $w = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55\%$ of links to a singly connected network, respectively. Hence, there are a total of $4 \times 2 \times 11 = 88$ distinct (m, s, w) combinations. For each combination, we simulated 10 BNs. This amounts to a total of 880 NAT-modeled BNs.

For each NAT-modeled BN, we created a *normalized* BN (NM-BN) where each NAT model is expanded into a tabular CPT, and a de-causalized BN (DC-BN). Both NM-BN and DC-BN are compiled for LP, conditioned on the same observation over 10% of randomly selected variables. For each pair of NM-BN and

DC-BN, LP resulted in the same posterior marginals, which empirically demonstrates exactness of de-causalization. LP runtimes for $m = 10, 12$, using a desktop of 3.4 GHz clock speed, are summarized in Fig. 8. Runtimes for $m = 6, 8$ are omitted due to space restriction.
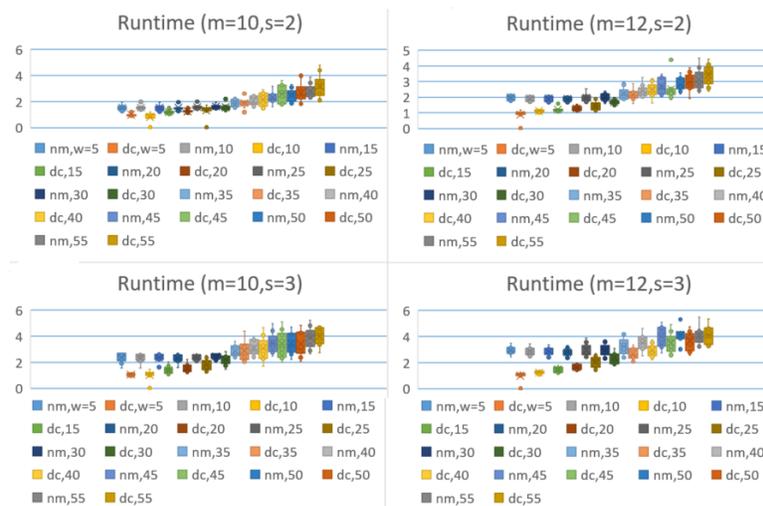
Fig. 8. LP runtimes (msec in log10) for NM-BNs and DC-BNs where $m = 10, 12$.

The inference becomes harder as m, s and w grow. For sparse BN structures, as inference becomes harder, DC-BNs become more advantageous than NM-BNs. For instance, with $w = 5$, as m and s grow, the runtime by DC-BNs become significantly less than NM-BNs. At $(m = 12, s = 3, w = 5)$, LPs with DC-BNs are two orders of magnitude fasters than NM-BNs.

Furthermore, as m and s grow, the range of structural densities where DC-BNs are more efficient than NM-BNs grows as well. For instance, for $(m = 6, s = 3)$, DC-BNs and NM-BNs tie in runtime around $w = 20$. As m grows to 8, 10, 12, the corresponding structural density grows to $w = 30, 50, 55$, respectively.

9   Conclusion

The main contribution of this work is the novel de-causalization framework, by which a NAT-modeled BN is converted into a de-causalized BN for inference computation. An existing alternative is the MF framework. In comparison, the MF frame has a limitation where one potential for each NIN-AND gate is exponential on η (domain size of effect e). In the de-causalization framework, no component is exponential. We demonstrated that the de-causalized BNs support exact inference, and can speed up LP inference by up to two orders of magnitude for a wide range of sparse BN structures. We are currently investigating the efficiency impact of de-causalization, when inference is performed through sum-product networks (SPNs).

Acknowledgement

References

1. C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in Bayesian networks. In E. Horvitz and F. Jensen, editors, *Proc. 12th Conf. on Uncertainty in Artificial Intelligence*, pages 115–123, 1996.

2. F.J. Diez. Parameter adjustment in Bayes networks: The generalized noisy OR-gate. In D. Heckerman and A. Mamdani, editors, *Proc. 9th Conf. on Uncertainty in Artificial Intelligence*, pages 99–105. Morgan Kaufmann, 1993.

3. F.J. Diez and M.J. Druzdzel. Canonical probabilistic models for knowledge engineering. Technical report cisiad-06-01, UNED, 2007.

4. M. Henrion. Some practical issues in constructing belief networks. In L.N. Kanal, T.S. Levitt, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 3*, pages 161–173. Elsevier Science Publishers, 1989.

5. J.F. Lemmer and D.E. Gossink. Recursive noisy OR - a rule for estimating complex probabilistic interactions. *IEEE Trans. on System, Man and Cybernetics, Part B*, 34(6):2252–2261, Dec 2004.

6. P.P. Maaskant and M.J. Druzdzel. An independence of causal interactions model for opposing influences. In M. Jaeger and T.D. Nielsen, editors, *Proc. 4th European Workshop on Probabilistic Graphical Models*, pages 185–192, Hirtshals, Denmark, 2008.

7. A.L. Madsen and F.V. Jensen. Lazy propagation: A junction tree inference algorithm based on lazy evaluation. *Artificial Intelligence*, 113(1-2):203–245, 1999.

8. K.G. Olesen, U. Kjrulff, F. Jensen, F.V. Jensen, B. Falck, S. Andreassen, and S.K. Andersen. A munin network for the median nerve-a case study on loops. *Applied Artificial Intelligence*, 3(2-3):385–403, 1989.

9. J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

10. J. Vomlel and P. Tichavsky. An approximate tensor-based inference method applied to the game of Minesweeper. In *Proc. 7th European Workshop on Probabilistic Graphical Models, Springer LNAI 8745*, pages 535–550, 2012.

11. S. Woudenberg, L.C. van der Gaag, and C. Rademaker. An intercausal cancellation model for Bayesian-network engineering. *Inter. J. Approximate Reasoning*, 63:3247, 2015.

12. Y. Xiang. Acquisition and computation issues with NIN-AND tree models. In P. Myllymaki, T. Roos, and T. Jaakkola, editors, *Proc. 5th European Workshop on Probabilistic Graphical Models*, pages 281–289, Finland, 2010.

13. Y. Xiang. Non-impeding noisy-AND tree causal models over multi-valued variables. *International J. Approximate Reasoning*, 53(7):988–1002, 2012.

14. Y. Xiang and Q. Jiang. NAT model based compression of Bayesian network CPTs over multi-valued variables. *Computational Intelligence (online DOI: 10.1111/coin.12126; paper version in press)*, 2017.

15. Y. Xiang and Y. Jin. Efficient probabilistic inference in Bayesian networks with multi-valued NIN-AND tree local models. *Int. J. Approximate Reasoning*, 87:67–89, 2017.