Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence

Canadian AI 2026
25-29 May 2026,Vancouver,BC

Visualizing the Elimination of Arbitrary Variables in Bayesian Networks as Compound Bayesian Networks

Cory J. Butz[†,*], Brandon C. Sasyniuk[†]
[†] Department of Computer Science, University of Regina, Regina, Canada

Abstract

Research on *Bayesian network* (BN) inference continues to this day along two main fronts: scalable inference and deepening our understanding of the semantics of intermediate inference steps. In this theoretical paper, falling in the latter direction, we give a novel graphical representation of eliminating arbitrary variables from discrete BNs. This includes methods that represent both multiplication and marginalization operations and involves extending classical BNs to *compound* BNs. Our main result formally establishes a one-to-one correspondence between intermediate numeric factorizations and graphical representations.

Keywords: Bayesian networks, variable elimination, inference, probabilistic reasoning

1. Introduction

Ongoing research into *Bayesian network* (BN) [1] inference continues to advance along two major directions: developing faster and more scalable inference methods [2–4], and deepening our understanding of the BN semantics [5–8]. In the former direction, BNs can be compiled into various types of probabilistic circuits, whose nodes typically represent the sum and product numeric operations, and which can be exploited by clustering, pruning, and parallelization [2] or implemented on GPU-based architectures [3, 4].

In the latter direction, investigations focus on semantics such as fusing BNs [5] or safely augmenting a given BN with new variables, a process known as *inverse marginalization* [6]. A notable contribution of [6] is the augmentation of an arbitrary variable into a BN, that is, the variable is not a special case (such as either a leaf node or a root node with one child). This feature directly corresponds with current work on marginalizing an arbitrary variable from a BN using the arc-reversal technique [7, 8]. However, one remaining open problem is understanding the intermediate steps when eliminating an arbitrary variable from a BN.

This paper gives a novel graphical representation of variable elimination in discrete BNs. The key to our approach is changing the representation of a classical BN, namely, a closed curve no longer represents a variable in the problem domain as historically done but instead represents the left side of a BN conditional probability table. The directed edges represent parent-child relationships as usual. One salient feature of our approach is that the numeric operations of multiplication and summation can be graphically represented. Our main result is that our graphical representation maintains a precise one-to-one correspondence with the numeric factorizations during inference. This work then fills a long-standing gap in the probabilistic reasoning literature.

The remainder of this paper is organized as follows. Section 2 reviews BNs, exact inference, and the fundamental task of variable elimination. Our novel BN representation and theoretical foundation is introduced in Section 3. Section 4 contains conclusions.

2. Background Knowledge

We review variable elimination in BNs, assuming some familiarity with the topic [1, 9].

[*] cory.butz@uregina.ca

A discrete *Bayesian network* (BN) [1] on finite set of variables (nodes) $U = \{v_1, v_2, \dots, v_n\}$ is a pair $(\mathcal{B}, \mathcal{C})$. $\mathcal{B}$ is a *directed acyclic graph* (DAG) with vertex set U. $\mathcal{C}$ is the set of *conditional probability tables* (CPTs) $\{p(v_i|P_i) \mid i = 1, 2, \dots, n\}$, where P_i are the parents (immediate predecessors) of v_i in $\mathcal{B}$. One salient feature of BNs is that the product of the BN CPTs is a *joint probability distribution* $p(U)$.

Example 1. *The BN on $U = \{a, b, \dots, l\}$ in Fig. 1 (a), defines a joint distribution as:*

$$p(U) \; = \; p(a) \cdot p(b|a) \cdots p(l|i, j), \tag{2.1}$$

where CPTs are not shown due to space considerations.

Exact probabilistic inference is the task of computing a query $p(W|Z = z)$ posed to a BN on U, where W and Z are disjoint subsets of U and z is *evidence*, the observed domain value of Z.

Our paper focuses on eliminating all variables outside of W and Z, the central task of BN inference [10]. A variable v is eliminated from a BN $\mathcal{B}$ as a two-part process: (i) multiply together all CPTs involving v, and (ii) sum out v from the product in part (i). The other variables not in the query are removed recursively. Note that a CPT $p(X|Y)$ is produced after every numeric step, with X and Y called the *left side* and *right side*, respectively.

Example 2. *Arbitrary variable f can be eliminated from the BN in Fig. 1 (a) following these sum and product numeric operations:*

$$\begin{aligned} \sum_f p(f|b,c) \cdot p(h|e,f) \cdot p(i|f) \; &= \; \sum_f p(f,h|b,c,e) \cdot p(i|f) \\ &= \; \sum_f p(f,h,i|b,c,e) \\ &= \; p(h,i|b,c,e). \end{aligned}$$

3. A Novel Representation of BNs and Arbitrary Variable Elimination

We motivate the need for a new depiction of BNs and then show how such a fresh depiction can represent arbitrary variable elimination in BNs.

In BNs, random variables in a problem domain and vertices (nodes) in a DAG are often used interchangeably. However, for other models (e.g. influence diagrams and decision graphs [11]) that contain decision variables and utility functions, a distinction is kept between variables and vertices, as a vertex does not necessarily represent a variable. In this paper, we advocate that this distinction is, in fact, required to depict arbitrary variable elimination. Probabilistic network CPTs are of the form $p(X|Y)$, where X is a single variable and Y is a (possibly empty) set of variables [11]. This perspective conflicts with variable elimination since eliminating a variable can yield CPTs where X is a non-singleton set, as clearly seen in Example 2. Consequently, if one is to depict variable elimination, a fresh perspective on variables and vertices is needed.

We propose overlaying closed curves on a DAG to depict the left sides of the CPTs.

Definition 1. *Let $\mathcal{F}$ be the CPT factorization obtained from an initial BN. The graph of $\mathcal{F}$, denoted $\mathcal{G}$ and called a compound BN, has a closed curve around the left side X, for each CPT $p(X|Y) \in \mathcal{F}$.*

Example 3. *Recall the BN in Figure 1(a) defining the CPT factorization $\mathcal{F}$ of the joint distribution in Eq. (2.1). The graph $\mathcal{G}$ of $\mathcal{F}$ is obtained by overlaying a closed curve around each of the variables, as illustrated in Fig. 1 (b).*

We now describe how to read the CPT information represented by a compound BN $\mathcal{G}$.

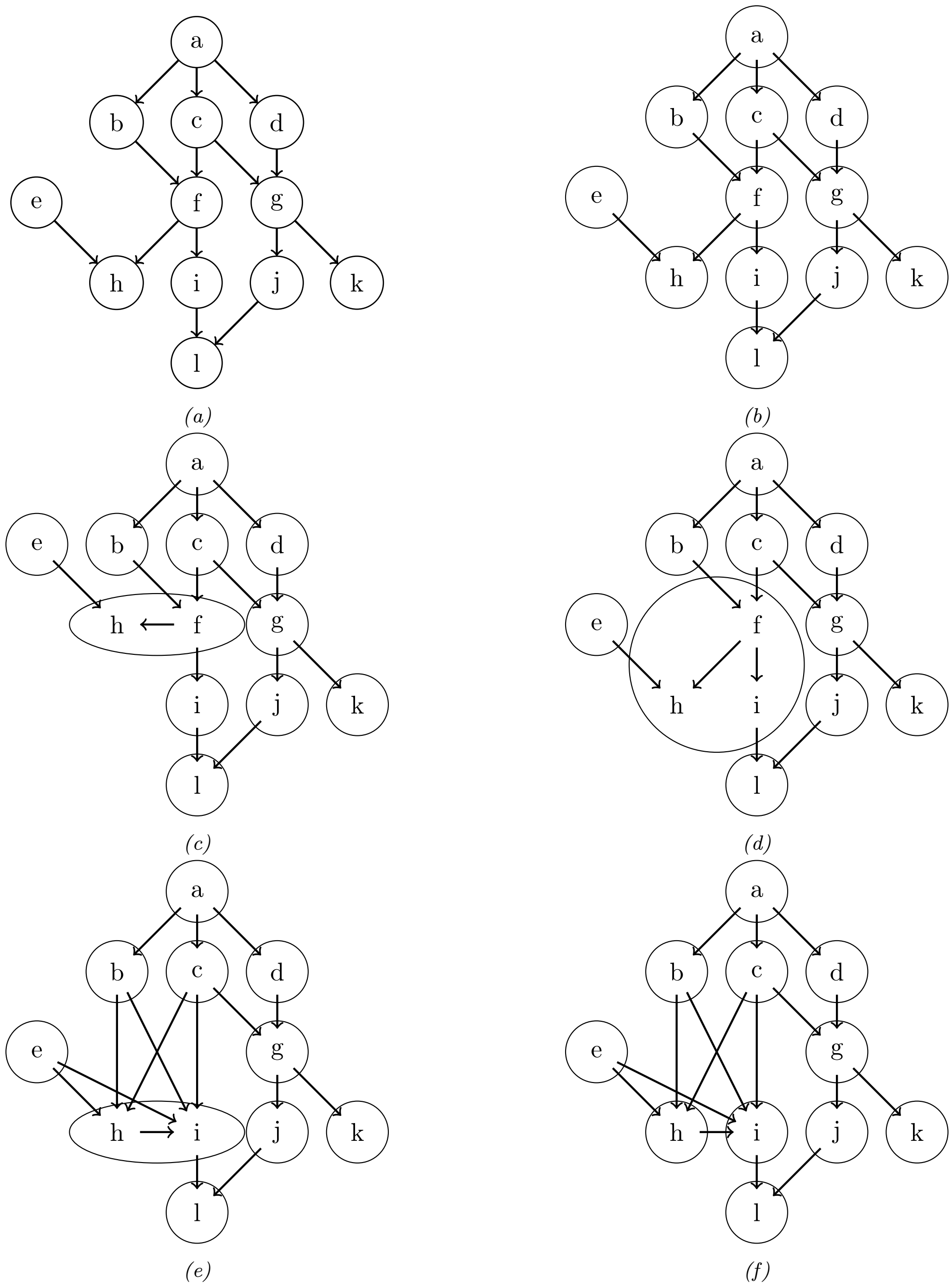

Figure 1. A classical depiction of a BN in (a) can be alternatively represented as the compound BN in (b), allowing for the precise depiction of all intermediate steps during the elimination of arbitrary variable f, as subsequently described in Example 5.

Definition 2. *Let $\mathcal{G}$ be a compound BN. The CPT factorization $\mathcal{F}$ of $\mathcal{G}$ is defined as follows. The number of closed curves in $\mathcal{G}$ represents the number of CPTs in the factorization. The set of vertices in a closed curve represents a CPT's left side X. This CPT's right side Y is the set of vertices not in X and having a directed edge to any vertex in X.*

Example 4. *The compound BN in Figure 1 (b) depicts the CPT factorization in Eq. (2.1).*

We focus on depicting the multiplication of two CPTs during variable elimination.

Definition 3. *Consider the multiplication of two CPTs $p(X_1|Y_1)$ and $p(X_2|Y_2)$, yielding the new CPT $p(X_1X_2|Y_1Y_2 - X_1X_2)$:*

$$p(X_1X_2|Y_1Y_2 - X_1X_2) \ = \ p(X_1|Y_1) \cdot p(X_2|Y_2).$$

This multiplication is depicted by deleting the closed curves around X_1 and X_2, adding a closed curve around X_1X_2, and leaving the directed edges undisturbed.

Now consider the summation operation in variable elimination. The summation step may involve adding new directed edges. To maintain acyclicity and uniqueness [6, 7], one *topological order*, denoted $\prec$, of the variables in the initial BN will be fixed. Any new directed edge (v_i, v_j) has $v_i \prec v_j$. In our examples, $\prec$ will be alphabetical.

When a variable v is to be summed out of the network, v will only appear in one CPT and, more specifically, will be a member of the variables in the left side of the CPT [12].

Definition 4. *Depict the summation of v from the CPT $p(v, X|Y)$,*

$$p(X|Y) = \sum_v p(v, X|Y)$$

by: (i) replacing the closed curve around vX with one around X; (ii) adding directed edges from every parent of v to every child of v, from every spouse (parent of a common child) of v to every child of v, and from every child of v_i to every other child v_j of v such that $v_i \prec v_j$; and (iii) delete vertex v and all edges involving v.

We can now depict every multiplication operation and every summation operation applied when eliminating arbitrary variables from an initial BN.

Example 5. *Consider the compound BN in Figure 1 (b). The first multiplication to eliminate arbitrary variable f, $p(f, h|b, c, e) = p(f|b, c) \cdot p(h|e, f)$, is depicted in Figure 1 (c). Similarly, Figure 1 (d) shows the next multiplication, $p(f, h, i|b, c, e) = p(f, h|b, c, e) \cdot p(i|f)$. Finally, Figure 1 (e) illustrates the summation, $p(h, i|b, c, e) = \sum_f p(f, h, i|b, c, e)$.*

The vital point in Example 5 is that all the numeric computations needed for the elimination of the arbitrary variable f are graphically represented as compound BNs in Figure 1.

We conclude this section by formally establishing our main claim: a one-to-one correspondence between intermediate numeric factorizations and graphical representations.

Theorem 1. *Let $\mathcal{F}_0$ be the CPT factorization of an initial BN and let $\mathcal{G}_0$ be the graph of $\mathcal{F}_0$. Let $\mathcal{F}_1$, $\mathcal{F}_0$, ..., $\mathcal{F}_k$ be the respective CPT factorizations when applying variable elimination. Let $\mathcal{G}_i$ be the graph obtained from $\mathcal{G}_{i-1}$ by making the modifications in Definitions 3 and 4 accordingly, $i = 1, \ldots, k$. Then $\mathcal{G}_k$ is the graph of $\mathcal{F}_k$ and vice versa.*

Proof. ($\Longrightarrow$) We prove the claim by mathematical induction. (Basic step) Consider an initial BN comprised of a DAG over n vertices and CPTs based on the structure of the DAG:

$$p(U) \ = \ p(v_1|P_1) \cdot p(v_2|P_2) \cdots p(v_n|P_n).$$

Definition 1 can be immediately applied to depict this factorization. (Inductive step) The CPT factorization obtained after $k-1$ multiplication and summation operations,

$$p(X_1|Y_1) \cdot p(X_2|Y_2) \cdots p(X_i|Y_i) \cdot p(X_j|Y_j) \cdots p(X_l|Y_l), \tag{3.1}$$

is precisely depicted by applying Definitions 3 and 4 as required, where $l < n$. There are two choices for the next numeric operation applied by variable elimination. Suppose the next operation is $p(X_i|Y_i) \cdot p(X_j|Y_j)$. Definition 3 would replace the closed curves around X_i and X_j with one around X_iX_j, thereby representing the factorization:

$$p(X_1|Y_1) \cdot p(X_2|Y_2) \cdots p(X_iX_j|Y_iY_j - X_iX_j) \cdots p(X_l|Y_l). \tag{3.2}$$

However, Equation (3.2) is exactly the factorization that variable elimination builds. Suppose the next operation is $\sum_v p(X_i|Y_i)$. Definition 4 would replace the closed curve around X_i with one around $X_i - v$. However, when adding new directed edges (v', v'') according to the fixed topological order $\prec$, variable v'' is a member of X_i. Thus, when reading CPTs from the graph, the new edges do not affect any other CPT in Equation (3.1), giving

$$p(X_1|Y_1) \cdot p(X_2|Y_2) \cdots p(X_i - v|Y_i) \cdot p(X_j|Y_j) \cdots p(X_l|Y_l). \tag{3.3}$$

Equation (3.3) is the CPT factorization numerically constructed by variable elimination.

($\Longleftarrow$) A similar argument holds. □

It is known that any compound CPTs constructed during BN inference can be equivalently rewritten into singleton CPTs defining a classical BN [7, 8]. Thus, our depiction of a compound BN can always be equivalently expressed as a classical BN. For example, recall the compound BN in Figure 1 (e). With $h \prec i$, the non-singleton CPT $p(h, i|b, c, e)$ can be rewritten using the chain rule as $p(h, i|b, c, e) = p(h|b, c, e) \cdot p(i|b, c, e, h)$. Thus, the new singleton CPT for h is $p(h|b, c, e)$, and that for i is $p(i|b, c, e, h)$. Figure 1 (f) depicts the equivalent classical BN.

While numerous attempts depict variable elimination during exact inference in discrete BNs, they either illustrate probability information before inference, hold only for special cases of elimination, or go beyond the realm of BNs. Zhang and Poole [13] and Shafer [10] maintain a sub-BN provided that the variable being eliminated is a leaf variable (one without children). Conversely, Kjaerulff and Madsen [11] depict a valid sub-BN when the variable being eliminated is a root variable having at most one child. If evidence is observed prior to variable elimination, Shafer [10] can depict the evidence variables, but this requires that every evidence variable be a recursively defined root variable. Similarly, Darwiche [9] deletes the edges emanating from evidence variables, but this graphical depiction ends once variable elimination begins. Madsen [14] can depict arbitrary variable elimination, but steps outside the realm of BNs by using *domain graphs*, involving undirected and directed edges. Butz et al. [15] do the same with their graphical representation involving black and white nodes, called *Darwinian networks*.

While acknowledging that advancements in semantic knowledge are hard to quantify, graphs, with their visual appeal, have long been a common metaphor for conceptual dependencies, including *semantic networks* [16], *constraint networks* [17], *inference networks* [18], *conceptual dependencies* [19], *conceptual structures* [20], and *Bayesian networks*. Graph concepts are entrenched in our language, i.e., "lines of reasoning," "threads of thought," "connected ideas," and "far-fetched arguments" [1], not to mention graph terminology such as "the path of least resistance," "to cross paths," and "to take the path less chosen." Visual representation of concepts and ideas is clearly beneficial to human understanding.

4. **Conclusion**

Research on BN inference is still of current interest, typically focusing on faster scalable inference, such as that conducted in probabilistic circuits [2–4] or a better understanding of semantics when fusing [5] or augmenting [6] BNs. Whereas [6] considers sufficient topological conditions to allow for the safe augmentation of a BN with another arbitrary variable, here we consider the opposite, namely, depicting the topology of the intermediate structures when eliminating an arbitrary variable from a BN. The key realization is to pioneer a new depiction of classical BNs as seen in Fig. 1 (a)-(b). This allows for the multiplication and summation operations to be graphically depicted during elimination. Our main result in Theorem 1 is that during arbitrary variable elimination in BNs a one-to-one correspondence is maintained between the factorization and the graph of the factorization. We have pointed out that although many attempts have been made to visualize variable elimination, none of these handle arbitrary variable elimination or require moving into frameworks such as domain graphs [14] and Darwinian networks [15]. Future work can explore incorporating additional semantics, such as using dashed closed curves to indicate that the probabilities of an intermediate CPT do not agree with those of the joint distribution [12].

References

[1] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.

[2] C. J. Butz, A. Santoscoy-Rivera, and C. E. Lewis. "Paths Not Taken: Structure-Based Pruning in PSDD Learning and Inference". In: *Proceedings of the Fortieth AAAI Conference on Artificial Intelligence*. AAAI Press, 2026, to appear.

[3] G. Gala, C. de Campos, A. Vergari, and E. Quaeghebeur. "Scaling Continuous Latent Variable Models as Probabilistic Integral Circuits". In: *Advances in Neural Information Processing Systems 37*. 2024.

[4] J. Maene, V. Derkinderen, and P. Zuidberg Dos Martires. "KLay: Accelerating Arithmetic Circuits for Neurosymbolic AI". In: *Proceedings of the Thirteenth International Conference on Learning Representations*. Singapore: ICLR, 2025.

[5] P. Torrijos, J. M. Puerta, J. A. Aledo, and J. A. Gámez. "Bayesian Network Structural Consensus via Greedy Min-Cut Analysis". In: *Proceedings of the Fortieth AAAI Conference on Artificial Intelligence*. 2026, to appear.

[6] J. Kwisthout and S. Renooij. "Inverse Marginalisation for Safely Expanding Bayesian Networks". In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Springer, Cham, 2025.

[7] C. J. Butz, A. L. Madsen, and J. S. Oliveira. "Fast Arc Reversal". In: *Proceedings of the Twelfth International Conference on Probabilistic Graphical Models*. 2024.

[8] C. J. Butz, A. L. Madsen, and J. S. Oliveira. "On Fast arc-reversal". In: *International Journal of Approximate Reasoning* 186 (2025), p. 109488.

[9] A. Darwiche. *Modeling and Reasoning with Bayesian Networks*. Los Angeles, CA: Cambridge University Press, 2009.

[10] G. Shafer. *Probabilistic Expert Systems*. Philadelphia, PA: Philadelphia: Society for Industrial and Applied Mathematics, 1996.

[11] U. B. Kjærulff and A. L. Madsen. *Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis*. 2nd. New York, NY: Springer, 2013.

[12] C. J. Butz and W. Yan. "The Semantics of Intermediate CPTs in Variable Elimination". In: *Proc. of the Fifth European Workshop on Probabilistic Graphical Models*. 2010, pp. 41–49.

[13] N. L. Zhang and D. Poole. "A simple approach to Bayesian network computations". In: *Proceedings of the Tenth Canadian Artificial Intelligence Conference*. 1994, pp. 171–178.

[14] A. L. Madsen. "An empirical evaluation of possible variations of lazy propagation". In: *Proceedings of the Twentieth Uncertainty in Artificial Intelligence*. 2004, pp. 366–373.

[15] C. J. Butz, J. S. Oliveira, and A. E. dos Santos. "Darwinian Networks". In: *Proceedings of the Twenty-Eighth Canadian Artificial Intelligence Conference*. 2015, pp. 16–29.

[16] W. A. Woods. "What's in a Link: Foundations for Semantic Networks". In: ed. by D. Bobrow and A. Collins. New York, NY, USA, 1975, pp. 35–82.

[17] U. Montanari. "Networks of constraints: Fundamental properties and applications to picture processing". In: *Information Sciences* 7 (1974), pp. 95–132.

[18] R. O. Duda, P. E. Hart, and N. J. Nilsson. "Subjective Bayesian Methods for Rule-Based Inference Systems". In: *Proceedings of the National Computer Conference and Exposition*. 1976, 1075–1082.

[19] R. C. Schank. "Conceptual dependency: A theory of natural language understanding". In: *Cognitive Psychology* 3.4 (1972), pp. 552–631.

[20] J. F. Sowa. "Conceptual Structures: Information Processing in Mind and Machine". In: Reading, Mass.: Addison-Wesley, 1984.