Constructing the Maximal Prime Decomposition of Bayesian networks

C.J. Butz
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2, E-mail: butz@cs.uregina.ca

Abstract *The maximal prime decomposition (MPD) of a Bayesian network is a hierarchical structure, which represents conditional independency information. The MPD representation has shown to facilitate probabilistic inference in uncertainty management. One method for building the MPD involves applying the moralization and triangulation procedures to the given Bayesian network. An alternative method constructs the MPD using certain independencies encoded in a Bayesian network.*

In this paper, we analyze these two methods with respect to the construction and representation of the root level in the MPD. Our comparison reveals that the latter method can be seen as only requiring the moralization procedure. A second difference is that the former method represents the root level of the MPD as a jointree, while the later represents it as an acyclic hypergraph. Finally, our investigation of these two different approaches to the construction of the MPD yields the introduction of a new hybrid construction algorithm.

1 Introduction

Probability theory is attractive for the management of uncertain knowledge due to its sound mathematical foundation. A *Bayesian network* [2] consists of a *directed acyclic graph* (DAG) and a corresponding set of conditional probability distributions. The *probabilistic conditional independencies* [3] encoded in the DAG indicate that the product of the conditional distributions is a unique joint probability distribution. In practice, probabilistic inference is carried out on a secondary representation of a Bayesian network. Traditionally, Algorithm 1 is applied to transform a Bayesian network into a *jointree* [2]. More specifically, the *moralization* and *triangulation* procedures [2] are applied to the DAG creating an *acyclic hypergraph* (a *chordal* undirected graph) [4].

More recently, however, two works [1, 5] have suggested methods (see Algorithms 2 and 3) for the *maximal prime decomposition* (MPD) of Bayesian networks. The MPD has been shown to facilitate probabilistic inference.

In this paper, we analyze these two methods with respect to the construction and representation of the root level in the MPD. Our comparison reveals that the latter method can be seen as only requiring the moralization procedure; the triangulation procedure is ignored. A second difference is that the former method represents the root level of the MPD as a jointree, while the later represents it as an acyclic hypergraph. Experimental results have shown that the acyclic hypergraph representation is more desirable for probabilistic inference than the jointree representation. Finally, our analysis of the two construction methods (Algorithms 2 and 3) leads to the introduction of a new construction method (Algorithm 4). This hybrid approach involves the first part of Algorithm 2 and the latter part of Algorithm 3.

This paper is organized as follows. Section 2 contains a review of Bayesian networks. In Section 3, we review the MPD construction method suggested in [1], while in Section 4 we do the same for the method given in [5]. In Section 5, the comparison of these two methods leads to the introduction of a third method for constructing the MPD representation. The conclusion is presented in Section 6.

2 Bayesian Networks

Let X, Y, Z be pairwise disjoint subsets of U. The *conditional independence* [3] of Y and Z given X is denoted $I(Y, X, Z)$. The conditional independencies encoded in the *Bayesian network* [4] in Fig. 1 on $U = \{a, b, c, d, e, f, g, h, i, j, k\}$ indicate that the joint probability distribution $p(U)$ can be written as

$$\begin{aligned} p(U) \quad = \quad & p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|b) \cdot p(e|b) \cdot p(f|d, e) \cdot \\ & p(g|b) \cdot p(h|c, f) \cdot p(i|g) \cdot p(j|g, h, i) \cdot p(k|h). \end{aligned}$$

Henceforth, the terms Bayesian network and DAG will be used interchangeably. Algorithm 1 will transform a DAG into a jointree.

Algorithm 1 .
1. Moralize D to obtain the undirected graph D^m.
2. Triangulate D^m to obtain D^t.
3. Identify the maximal cliques $h_1, h_2, \ldots, h_n$ of D^t to obtain the acyclic hypergraph $H = \{h_1, h_2, \ldots, h_n\}$.
4. Organize H as a jointree J.

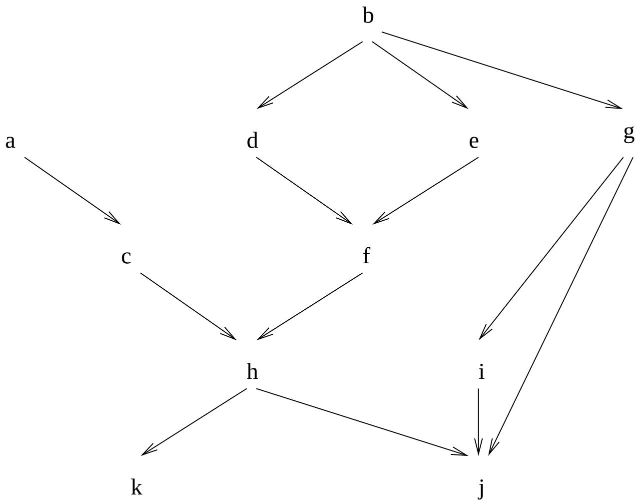

Figure 1. A *Bayesian network* D**.**

Example 1 *Consider the Bayesian network in Figure 1. By step (1), the moralization* D^m *of* D *is shown in Figure 2. Since* D^m *is not triangulated, i.e.,* D^m *is not a chordal undirected graph, edges need to be added to make it so. By step (2), a minimum triangulation* D^t *of* D^m *can be obtained by adding the two edges* (b, f) *and* (f, g)*, as shown by the dashed lines in Figure 3. The maximal cliques of the triangulated graph* D^t *are* $bdef$, bfg, fgh, cfh, ac, hk, *and* $ghij$. *These cliques are organized as a jointree* J*, as illustrated in Figure 4.*

3 Maximal Prime Decomposition of Bayesian Networks

Olesen and Madsen [1] proposed that a given Bayesian network be represented by its unique *maximal prime decomposition* (MPD). Although the MPD is an hierarchical structure, our focus here is only on the root level. The root network is a jointree.

Algorithm 2 will construct a jointree representing the root level of the MPD representation of a given Bayesian network D.

Algorithm 2 *[1]*
1. Construct a conventional jointree J *using a minimal triangulation.*
2. Aggregate any two cliques C_1 *and* C_2 *where the separator* S *is not complete in the moralization of* D*.*

Example 2 *Consider the Bayesian network* D *in Figure 1. A conventional jointree* J *constructed by step (1) is shown in Figure 4. The separators* bf *and* fh *in* J *are each not complete in the moralization of* D*. Hence, the jointree constructed by step (2) is illustrated in Figure 5.*

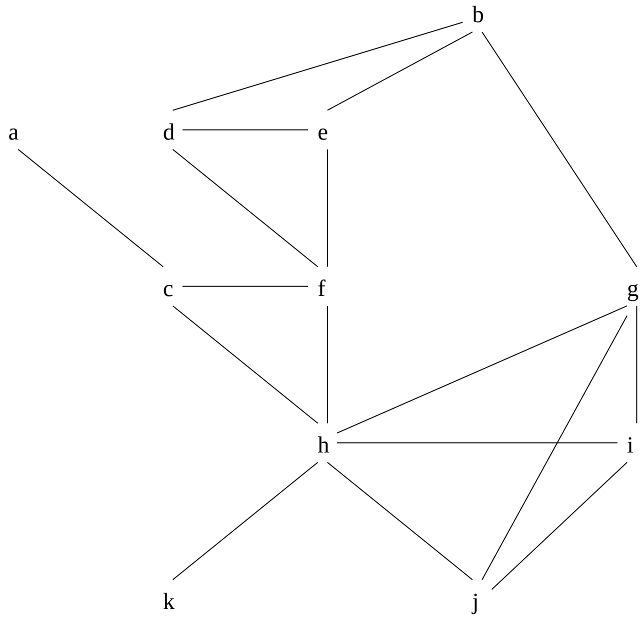

Figure 2. The *moralization* D^m **of the Bayesian network** D **in Figure 1.**

4 An Alternative MPD Method

In this section, we review a second method [5] for constructing the root level of the MPD of Bayesian networks.

Pearl states that in the strictest sense Bayesian networks are hypergraphs (see page 125 in [2]). The *Bayesian hypergraph* D^h defined by a given Bayesian network D is:

$$D^h = \{a_i P_i \mid a_i \text{ is a variable in } D\},$$

where P_i is the *parent set* [2] of variable a_i in D. By definition, a given Bayesian network D defines a *unique* Bayesian hypergraph D^h.

The *Bayesian hypergraph* D^h, defined by the DAG D in Figure 1, is illustrated in Figure 6.

The *separation* method [5] can infer CIs encoded in an undirected graph. The set of CIs encoded in a hypergraph $\mathcal{H}$ is denoted $CI(\mathcal{H})$. For example, the following CIs

$$I(a, c, bdefghijk),\ I(k, h, abcdefgij),$$
$$I(j, ghi, abcdefk),\ I(ij, gh, abcdefk),$$
$$I(bde, fg, achijk),\ I(ac, fh, bdegijk)$$

can be inferred from the hypergraph D^h in Figure 6.

We are primarily interested in a special subset of $CI(\mathcal{H})$. Given $I(Y, X, Z)$ in $CI(H)$, we call X a *Lien sepset*, if the following two conditions are both satisfied:

(i) X is contained by a hyperedge in H, and

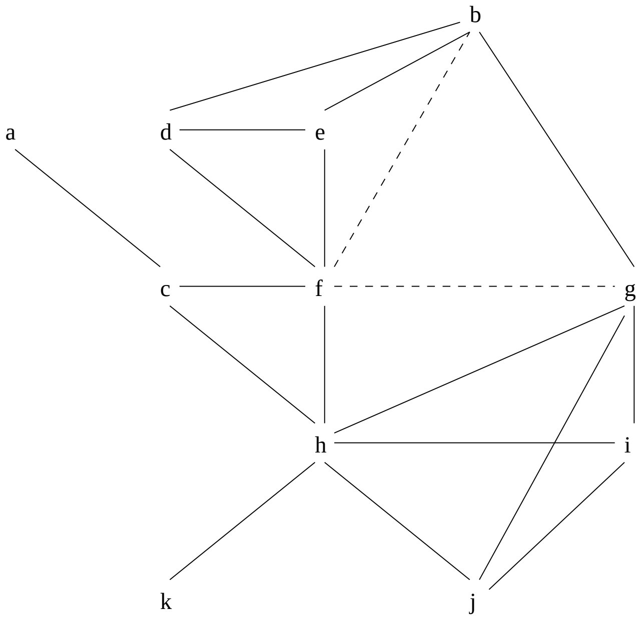

Figure 3. A *triangulated* graph D^t for the Bayesian network D in Figure 1; D^t is obtained from the undirected graph of D^m in Figure 2 by adding the two undirected edges (b, f) and (f, g),

(ii) $I(YX_1, X_2, Z)$ is not in $CI(H)$,

where $X_1X_2 = X$, $X_1 \neq \emptyset$, and $X_2 \neq \emptyset$. The *Lien independencies* of a hypergraph H, denoted $LI(H)$, are defined as:

$$LI(H) = \{I(Y, X, Z) \mid I(Y, X, Z) \text{ is in } CI(H) \text{ and } X \text{ is a Lien sepset}\}.$$

For instance, $I(bde, fg, chij)$ is in $CI(H)$ but not $LI(H)$, since fg is *not* contained by any hyperedge in H.

Algorithm 3 will construct an acyclic hypergraph representing the root level of the MPD representation of a given Bayesian network D.

Algorithm 3 *[5]*
1. Compute the Bayesian hypergraph D^h.
2. Build the acyclic hypergraph using $LI(\mathcal{H}^h)$.

Example 3 *Consider the Bayesian network in Figure 1. The Bayesian hypergraph D^h is shown in Figure 6. The Lien independencies of D^h are*

$$LI(D^h) = \{I(a, c, bdefghijk),\ I(k, h, abcdefgij),\ I(ack, fh, bdegij),\ I(ij, gh, abcdefk)\}.$$

The set $LI(D^h)$ of independencies define the Lien hypergraph in Figure 7.

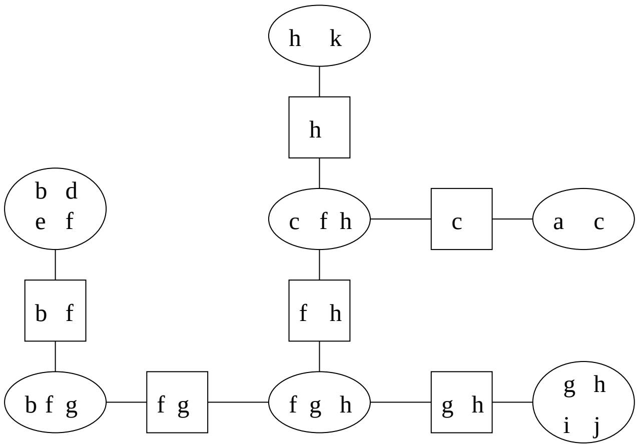

Figure 4. A traditional jointree J for the Bayesian network D in Figure 1.

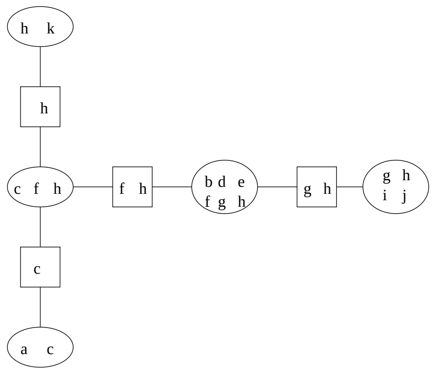

Figure 5. The root level of the MPD of the Bayesian network in Figure 1.

5 Comparing the Construction Methods

We begin by contrasting these two methods. The similarities between the methods suggest a hybrid approach for constructing the MPD representation of Bayesian networks.

Olesen and Madsen [1] suggest that the root level of the MPD be a fixed jointree. On the contrary, Wong et al. [5] propose that the root level of the MPD be an acyclic hypergraph. Experimental results, including [6], have shown that fixing a jointree requires extra computation for processing some probabilistic queries. On the other hand, an acyclic hypergraph can always be pruned to remove the irrelevant variables with respect to a given query.

The similarities we now present between the methods in [1] and [5] lead to the introduction of a hybrid approach to constructing the maximal prime decomposition of Bayesian networks. Our method is based on graphical procedures and also on inferred independency information.

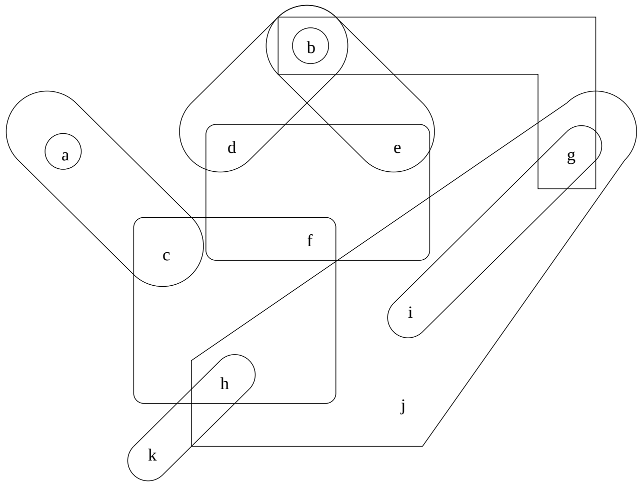

Figure 6. The *Bayesian hypergraph D^h* **defined by the DAG in Figure 1.**

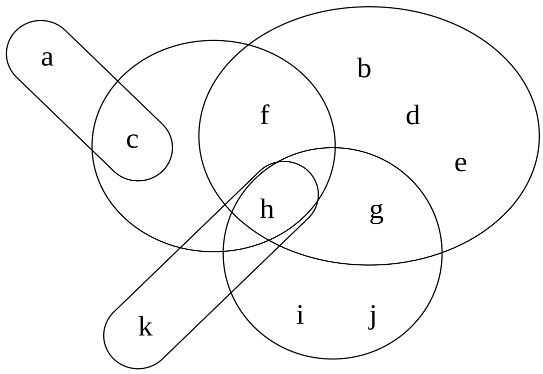

Figure 7. The root level of the MPD of the Bayesian network in Figure 1.

Lemma 1 *Let D be a Bayesian network and D^h the Bayesian hypergraph. Then the graph $G(D^h)$ of D^h is the moralization D^m of D.*

Lemma 2 *Let $\mathcal{H}$ be a hypergraph and $G(\mathcal{H})$ be its undirected graph. Then $CI(\mathcal{H}) = CI(G(\mathcal{H}))$.*

Given the moralization D^m of D, $LI(D^m)$ is defined as the set $I(Y, X, Z)$, where $I(Y, X, Z)$ is in $CI(D^m)$, X is a subset of some family set F_i for some variable a_i in D, $I(YX_1, X_2, Z)$ is not in $CI(D^m)$, $X_1X_2 = X$, $X_1 \neq \emptyset$, and $X_2 \neq \emptyset$. For instance, $I(bde, fg, chij)$ is in $CI(D^m)$ but not $LI(D^m)$ as fg is not a subset of a family set of D.

Algorithm 4 will construct an acyclic hypergraph representing the root level of the MPD representation of a given Bayesian network D.

Algorithm 4 .
1. Compute the moralization D^m of D.
2. Build the acyclic hypergraph using $LI(D^m)$.

Example 4 *Consider the Bayesian network D in Figure 1 and its moralization D^m in Figure 2. The Lien independencies of D^m are*

$$\begin{aligned} LI(D^m) \;\; = \;\; & \{I(a, c, bdefghijk),\ I(k, h, abcdefgij), \\ & I(ack, fh, bdegij),\ I(ij, gh, abcdefk)\}. \end{aligned}$$

The main algorithm in [4] will construct the acyclic hypergraph in Figure 7 from the set $LI(D^m)$ of conditional independencies.

6 Conclusion

Two works [1, 5] have recently suggested that Bayesian networks be represented in a hierarchical fashion. The root level of [1] is a fixed jointree, while that of [5] is an acyclic hypergraph. Experimental results [6] have demonstrated that the acyclic hypergraph representation is more desirable than a jointree for processing probabilistic queries. Our analysis of the two construction methods (Algorithms 2 and 3) lead to the introduction of a new construction method (Algorithm 4). This hybrid approach involves the first part of Algorithm 2 and the latter part of Algorithm 3.

References

[1] Olesen, K.G. and Madsen, A.L. Maximal prime subgraph decomposition of Bayesian networks, *IEEE Transactions on Systems, Man, and Cybernetics*, B, 32(1):21-31, 2002.

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

[3] Wong, S.K.M., Butz, C.J. and Wu, D. On the implication problem for probabilistic conditional independency, *IEEE Transactions on Systems, Man, and Cybernetics*, A, 30(6): 785–805, 2000.

[4] Wong, S.K.M. and Butz, C.J. Constructing the dependency structure of a multi-agent probabilistic network, *IEEE Transactions on Knowledge and Data Engineering*, 13(3):395–415, 2001.

[5] Wong, S.K.M., Butz, C.J. and Wu, D. On undirected representations of Bayesian networks, *ACM SIGIR Workshop on Mathematical/Formal Models in Information Retrieval*, 52-59, 2001.

[6] Wong, S.K.M., Butz, C.J. and Wu, D. Probabilistic reasoning in Bayesian networks: a relational database approach, to appear in *Sixteenth Canadian Conference on Artificial Intelligence*, 2003.