Comparing Hierarchical Markov networks and Multiply Sectioned Bayesian networks

C.J. Butz and H. Geng

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: {butz,gengho}@cs.uregina.ca

Abstract. *Multiply sectioned Bayesian networks* (MSBNs) were originally proposed as a modular representation of uncertain knowledge by sectioning a large *Bayesian network* (BN) into smaller units. More recently, *hierarchical Markov networks* (HMNs) were developed in part as an hierarchical representation of the flat BN.
In this paper, we compare the MSBN and HMN representations. The MSBN representation does not specify how to section a BN, nor is it a faithful representation of BNs. On the contrary, a given BN has a *unique* HMN representation, which encodes *precisely* those independencies encoded in the BN. More importantly, we show that failure to encode known independencies can lead to unnecessary computation in the MSBN representation. These results, in particular, suggest that HMNs may be a more natural representation of BNs than MSBNs.

1 Introduction

Probabilistic reasoning with *Bayesian networks* (BNs) [5] has been an active field of research over the past two decades. To facilitate the inference process, a BN is represented as a secondary network, usually a (decomposable) *Markov network* (MN) [5]. Several researchers, however, have suggested alternative representations of BNs, including *hierarchical Markov networks* (HMNs) [10], *multiply sectioned Bayesian networks* (MSBNs) [12], *multiple undirected graphs* [6], *nested jointrees* [2], and *maximal prime decompositions* [4]. Our discussion here focuses on the HMN and MSBN representations.

As the name suggests, the *hierarchical Markov network* (HMN) framework represents a BN as a hierarchy of MNs. It was also shown in [10] that HMNs have several advantages over the MN, multiple undirected graphs, and nested jointree representations. Very recently, it was shown in [1] that the HMN representation has the same advantages over the *maximal prime decomposition* [4] representation. Hence, the HMN representation seems to be a favorable framework for representing uncertain knowledge.

On the other hand, MSBNs were originally proposed as a modular representation of a large and sparse BN. By sectioning one BN into several smaller units, inference computation can be performed on one local network in a more efficient manner than on one conventional MN. Xiang [11] showed that Srinivas's work in [7] was actually an application of a special case of MSBN to hierarchical model-based diagnosis. The MSBN representation supports object-oriented inference, as emphasized by Koller and Pfeffer [3]. As MSBNs seem to be another desirable representation of uncertainty, it is natural to compare the HMN and MSBN representations.

Despite its name, we first show in this paper that a MSBN is in fact a two-level hierarchy of MNs. Although the HMN representation is guaranteed to encode precisely those independencies in a BN [10], we next show that a MSBN does not. This is a crucial difference as efficient probabilistic inference is based on utilizing independencies. We explicitly demonstrate in Ex. 7 that failure to represent known independencies leads to unnecessary computation. Moreover, Xiang et al.[13] point out some limitations of sectioning a BN as a MSBN. The MSBN technique makes the natural localization assumption. Hence, localization does not dictate exactly what should be the boundary conditions between different subnets [13]. In order to provide a coherent framework for probabilistic inference, technical constraints are imposed. Xiang, Olesen and Jensen [15] recently acknowledged that how to satisfy these technical constraints may not be obvious to a practitioner. This means that the MSBN representation itself does not indicate how the BN is to be sectioned, while the knowledge engineer may not know how to satisfy the technical constraints required to make a workable MSBN. On the other hand, our constructed HMN representation is *unique* for a given BN [10]. This sectioning is defined solely by the structure of the BN. It does not involve any technical constraints, nor does it require any type of practitioner input. Our analysis then suggests that it is perhaps more useful to represent a given BN as a HMN rather than as a MSBN.

This paper is organized as follows. In Section 2, we review BNs and MNs. We outline the MSBN and HMN representations in Section 3. In Section 4, we compare these two representations. The conclusion is given in Section 5.

2 Background Knowledge

Let U be a finite set of discrete random variables, each with a finite set of mutually exclusive states. Obviously, it may be impractical to define a joint distribution on U directly: for example, one would have to specify 2^n entries for a distribution over n binary variables. BNs utilize *conditional independencies* [9] to facilitate the acquisition of probabilistic knowledge.

Let X, Y and Z be disjoint subsets of variables in R. Let x, y, and z denote arbitrary values of X, Y and Z, respectively. We say Y and Z are *conditionally independent* given X under the joint probability distribution p, denoted $I(Y, X, Z)$, if

$$p(y \mid x, z) = p(y \mid x), \tag{1}$$

whenever $p(x, z) > 0$. $I(Y, X, Z)$ can be equivalently written as

$$p(y, x, z) = \frac{p(y, x) \cdot p(x, z)}{p(x)}. \tag{2}$$

A *Bayesian network* (BN) [5] is a pair $\mathcal{B} = (D, C)$. In this pair, D is a *directed acyclic graph* (DAG) on a set U of variables, and $C = \{p(a_i|P_i) \mid a_i \in D\}$ is the corresponding set of *conditional probability tables* (CPTs), where P_i denotes the *parent set* of variable a_i in the DAG D. The *family set* of a variable $a_i \in D$, denoted F_i, is defined as $F_i = \{a_i\} \cup P_i$. The *d-separation* method [5] can be used to read independencies from a DAG. For instance, $I(d, b, e)$, $I(c, \emptyset, f)$, $I(h, g, i)$ and $I(defh, b, g)$ all hold by d-separation in the DAG D in Fig. 1.

Example 1. Consider the BN $\mathcal{B} = (D, C)$, where D is the DAG in Fig. 1 on $U = \{a, b, c, d, e, f, g, h, i, j, k\} = abcdefghijk$, and C is the corresponding set of CPTs. The conditional independencies encoded in the DAG D indicate that the product of the CPTs in C defines a *unique* joint probability distribution $p(U)$:

$$p(U) = p(a) \cdot p(b) \cdot p(c|a) \cdot p(d|b) \cdot p(e|b) \cdot p(f|d, e) \cdot p(g|b) \cdot p(h|c, f) \cdot p(i|g) \\ \cdot p(j|g, h, i) \cdot p(k|h). \quad (3)$$

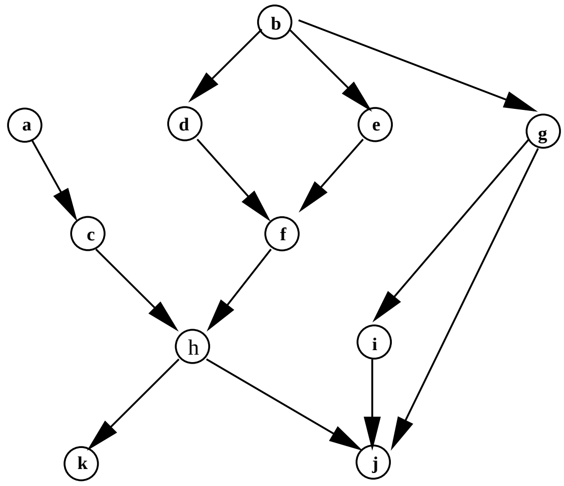

Fig. 1. A Bayesian network on variables $U = \{a, b, c, d, e, f, g, h, i, j, k\}$.

To facilitate probabilistic inference, a BN is usually transformed into a *Markov network* (MN), which Pearl [5] calls a *decomposable* MN. A MN consists of a *triangulated* (*chordal*) graph together with a potential defined over each maximal clique of D^t (defined below). Given a DAG D, the *moralization* D^m of D is the undirected graph defined as

$$D^m = \{(a, b) \mid a, b \in F_i \text{ for the family set } F_i \text{ of each variable } a_i \in D\}. \quad (4)$$

If necessary, edges are added to D^m to obtain a *triangulated* graph D^t. The *maximal cliques* (maximal complete subgraphs) of D^t are organized as *jointree* J. Finally, the CPTs of the BN are assigned to nodes of J.

Example 2. Consider the BN $\mathcal{D} = (D, C)$ above. The *moralization* D^m of D is shown in Fig. 2. A minimum *triangulation* D^t can be obtained by adding the two edges (b, f) and (f, g) to D^m. 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 shown in Fig. 3.

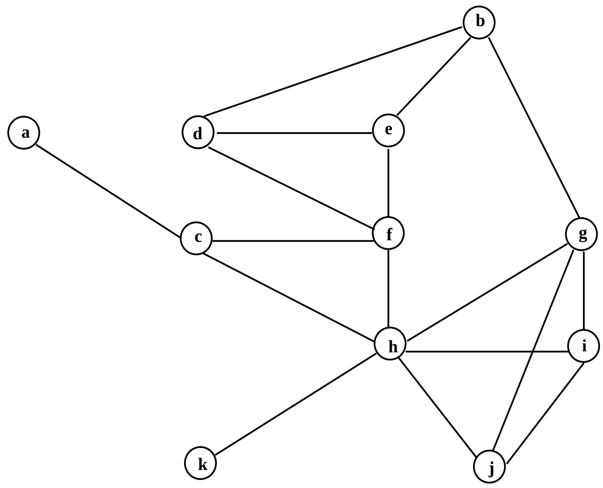

Fig. 2. The *moralization* D^m of the DAG D in Fig. 1.

Example 3. Given the BN in Fig. 1, one possible MN is illustrated in Fig. 3. This MN expresses the joint distribution in Ex. 1 as

$$p(U) = \frac{p(bdef) \cdot p(bfg) \cdot p(fgh) \cdot p(ghij) \cdot p(cfh) \cdot p(ac) \cdot p(hk)}{p(bf) \cdot p(fg) \cdot p(gh) \cdot p(fh) \cdot p(c) \cdot p(h)}. \quad (5)$$

Although the MN representation facilitates the probabilistic inference process, it may not represent all of the independencies in a BN. For instance, while the BN in Ex. 1 encodes $I(h, g, i)$, this conditional independence of h and i given g is not encoded in the MN in Ex. 3. This undesirable characteristic has lead to the proposal of other representations of BNs [2,4,10], including the hierarchical Markov network representation discussed in the next section.

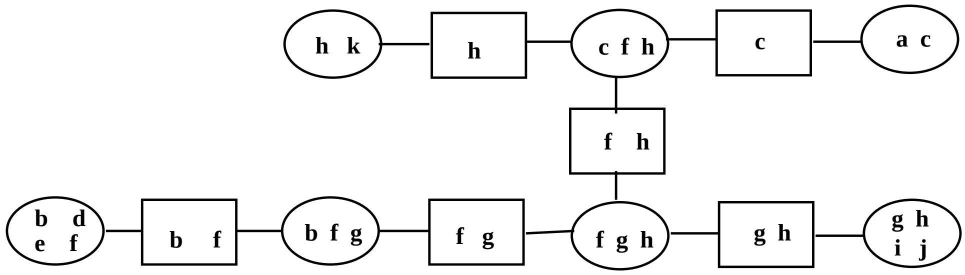

Fig. 3. One possible *Markov network* (MN) of the BN in Fig. 1.

3 The MSBN and HMN representations

Here we review two favorable representations of probabilistic knowledge, namely, multiply sectioned Bayesian networks and hierarchical Markov networks.

3.1 Multiply Sectioned Bayesian networks

Multiply sectioned Bayesian networks (MSBNs) [12,13,15] were originally proposed as a modular representation of a large and sparse BN.

Despite its name, a MSBN is a two-level hierarchy of MNs. There are only two differences between a MSBN and a traditional MN. First, the root level MN in a MSBN is not necessarily obtained via the moralization and triangulation procedures. Second, each node in the root level MN has a local MN nested in it. One technical constraint imposed on the root level MN is that, for any variable appearing in more than one node, there exists a node containing its parent set.

Example 4. The BN in Fig. 1 can be represented by the MSBN in Fig. 4. (The root level MN satisfies the MSBN restriction, since the parent set $\{a\}$ of c is contained in the node $\{a, c\}$, the parent set $\{b\}$ of g is contained in the node $\{b, c, d, e, f, g, h\}$, and the parent set $\{c, f\}$ of h is contained in the node $\{b, c, d, e, f, g, h\}$.) This MSBN encodes the following independency information:

$$p(U) = \frac{p(ac) \cdot p(bcdefgh) \cdot p(ghij) \cdot p(hk)}{p(c) \cdot p(gh) \cdot p(h)}, \tag{6}$$

$$p(bcdefgh) = \frac{p(cfh) \cdot p(def) \cdot p(bde) \cdot p(bg)}{p(f) \cdot p(de) \cdot p(b)}. \tag{7}$$

It is perhaps worthwhile here to elaborate on the MSBN construction process. Given the root MN J for a MSBN representation, an embedded MN J_X is constructed for each node X of J by the following four steps:

(i) compute the *subDAG* D_X of DAG D onto the subset X of variables,
(ii) apply the *MSBN moralization* to D_X giving the undirected graph $D_X^{m'}$,

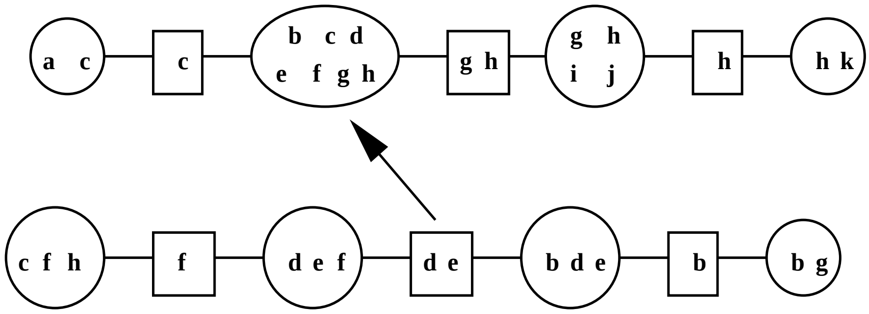

Fig. 4. One possible *multiply sectioned Bayesian network* (MSBN) of the BN in Fig. 1.

(iii) triangulate $D_X^{m'}$ if necessary,
(iv) construct a jointree J_X for the triangulated graph in (iii).

The new notions of subDAG and MSBN moralization are now defined. Given a BN D on U, the subDAG D_X of D onto the subset X of U is defined as:

$$D_X = \{\ (a, b) \mid (a, b) \in D \text{ and } a, b \in X\ \}.$$

Since the parent set of a node might not be contained in the same subDAG as the node itself, the moralization procedure is modified as follows. The *MSBN moralization* $D_X^{m'}$ of a subDAG D_X means the moralization in Eq. (4) except that the family set F_i of variable a_i is defined with respect to D_X.

Example 5. Consider the node $X = bcdefgh$ in the MSBN root MN in Fig. 4. By definition, the subDAG D_X is:

$$D_X = \{\ (b, d), (b, e), (b, g), (c, h), (d, f), (e, f), (f, h)\ \}.$$

The MSBN moralization $D_X^{m'}$ is then:

$$D_X^{m'} = \{\ (b, d),\ (b, e),\ (b, g),\ (c, f),\ (d, e),\ (d, f),\ (e, f)\ \}.$$

No additional edges have to be added to $D_X^{m'}$, as it is already a triangulated graph. The maximal cliques of $D_X^{m'}$ are cfh, def, bde, bg. There is only one jointree J_X for these four cliques, namely, the one shown in Fig. 4.

On the other hand, consider the root jointree node $ghij$ in Fig. 4. The subDAG D_{ghij} is

$$D_{ghij} = \{\ (g, i), (g, j), (h, j), (i, j)\ \},$$

and the MSBN moralization $D_{ghij}^{m'}$ is

$$D_{ghij}^{m'} = \{\ (g, h), (g, i), (g, j), (h, i), (h, j), (i, j)\ \}.$$

Not only is $D^{m'}_{ghij}$ a triangulated graph, but it is in fact a *complete* graph, i.e., $D^{m'}_{ghij}$ has only one maximal clique. By definition, the jointree J_X for $D^{m'}_{ghij}$ has only one node. Since any jointree defined by a single node does *not* encode any independencies, the embedded jointree J_X for the root jointree node $ghij$ is not illustrated in Fig. 4. The important point is that the MSBN representation does not encode any independencies for the root jointree node $ghij$. Similar remarks hold for the root jointree nodes ac and hk in Fig. 4.

By sectioning one BN into several smaller subnets, inference computation can be performed on one subnet at a time in a more efficient manner. Instead of updating the entire MN as in a traditional approach to probabilistic inference, the MSBN approach only updates the embedded MN for the node currently under consideration in the root level MN.

The MSBN representation is quite robust as it can be applied to large diagnostic systems [15], in either a single agent or a multi-agent paradigm [12], and supports object-oriented inference as emphasized in [3].

3.2 Hierarchical Markov networks

In [10], Wong et al. suggested that a BN be transformed into a *hierarchical Markov network* (HMN). An HMN is a hierarchy of MNs (jointrees).

Due to space limitations, we use an example to illustrate HMNs, and refer readers to [10] for a thorough discussion on the automated procedure for constructing the unique HMN representation of a given BN.

Example 6. The BN in Fig. 1 can be represented by the unique HMN in Fig. 5. This HMN encodes the following independency information:

$$p(U) = \frac{p(ac) \cdot p(cfh) \cdot p(hk) \cdot p(bdefgh) \cdot p(ghij)}{p(c) \cdot p(h) \cdot p(fh) \cdot p(gh)}, \tag{8}$$

$$p(cf) = p(c) \cdot p(f), \tag{9}$$

$$p(bdefgh) = \frac{p(fh) \cdot p(def) \cdot p(bde) \cdot p(bg)}{p(f) \cdot p(de) \cdot p(b)}, \tag{10}$$

$$p(bde) = \frac{p(bd) \cdot p(be)}{p(b)}, \tag{11}$$

$$p(ghi) = \frac{p(gh) \cdot p(gi)}{p(g)}. \tag{12}$$

In [10], it was shown that HMNs have several advantages over the MN, multiple undirected graphs, and nested jointree representations. In particular, the HMN can optimize queries using independencies that would go unnoticed in other representations [10]. More recently, it was explicitly demonstrated in [1] that HMNs possess several important characteristics, which the maximal prime decomposition representation does not. In the next section, we bring the elegance of the HMN representation down to bear on the MSBN representation.

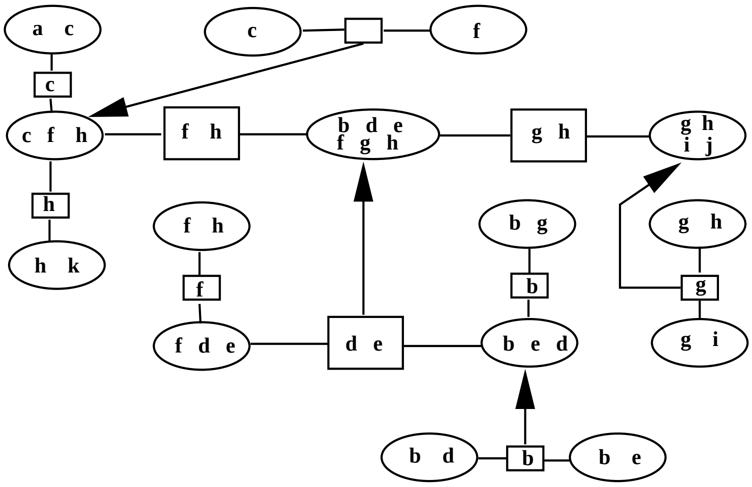

Fig. 5. The *hierarchical Markov network* (HMN) for the DAG D in Fig. 1.

4 Comparing the MSBN and HMN representations

In this section, we present a comprehensive comparison of the HMN and MSBN representations based on seven relevant factors.

(i) Assumptions: The MSBN representation makes the natural localization assumption. Hence, *localization does not dictate exactly what should be the boundary conditions between different subnets* [13]. The HMN representation does not require any assumptions when sectioning a BN.

(ii) Technical constraints: In order to make a workable MSBN, technical constraints such as the *d-sepset* condition need to be imposed [13]. The HMN representation does not impose any technical constraints.

(iii) Practitioner input: The sectioning of a BN into a MSBN is performed by a knowledge engineer. Xiang et al. [15] recently acknowledged that how to satisfy these technical constraints may not be obvious to a practitioner. Constructing a HMN from a BN is an *automated* procedure; it does not require any human input.

(iv) Restriction on the number of levels: By definition, the MSBN representation always has precisely two levels. The number of levels in a HMN is determined solely by the structure of a BN, and is not confined to two levels. For instance, the HMN in Fig. 5 has three levels.

(v) Uniqueness: While there may be multiple MSBN representations for a given BN, the constructed HMN representation is always *unique* [10].

(vi) Faithfulness: Given a BN, the HMN representation is guaranteed to be *equivalent* [10], whereas the MSBN is not. In other words, the HMN encodes those and only those independencies in the BN. In our running example, an independence can be obtained from Eq. (3) if and only if it can be using Eqs. (8)-(12). For instance, the conditional independence $I(h, g, i)$ of h and i given g is encoded in the BN and it is encoded in the HMN (see Eq. (12)). However, $I(h, g, i)$ is *not* encoded in the MSBN of Fig. 4.

(vii) Probabilistic Inference: Both the HMN and MSBN representations can perform local query processing. However, the MSBN approach to optimization needs to be somewhat qualified, as the next example demonstrates.

Example 7. In the HMN and MSBN representations, let us process the query $p(c|f = 0)$, assuming for simplicity that all variables are binary. By definition,

$$p(c \mid f = 0) = \frac{p(c, f = 0)}{p(f = 0)}. \tag{13}$$

The MSBN can use its only embedded jointree as follows. Two additions are required to compute $p(c, f = 0)$ from the stored distribution $p(c, f, h)$. One more addition is required to derive $p(f = 0)$ from $p(c, f = 0)$. Two divisions are needed to compute the desired result $p(c|f = 0)$ using $p(c, f = 0)$ and $p(f = 0)$. Thus, the MSBN approach requires three additions and two divisions to compute $p(c|f = 0)$. On the contrary, the HMN approach requires zero additions and zero divisions to compute $p(c|f = 0)$. The reason is that the HMN encodes $I(c, \emptyset, f)$ meaning that Eq. (13) can be rewritten as:

$$p(c \mid f = 0) = \frac{p(c, f = 0)}{p(f = 0)} = \frac{p(c) \cdot p(f = 0)}{p(f = 0)} = p(c). \tag{14}$$

The marginal $p(c)$ is already stored in the HMN representation.

Query optimization means taking advantage of independencies during processing. Our HMN encodes $I(c, \emptyset, f)$, which is given in the BN. Since c and f are unconditionally independent, $p(c|f = 0) = p(c)$ as shown in Eq. (14). Since $p(c)$ is a marginal already stored in the HMN, the query $p(c|f = 0)$ can be answered without any additional computation. On the contrary, the MSBN sacrifices $I(c, \emptyset, f)$. Failure to represent $I(c, \emptyset, f)$ leads to unnecessary work in the MSBN computation of $p(c|f = 0)$, as Ex. 7 demonstrates.

5 Conclusion

In this paper, we exposed the MSBN representation as a very limited hierarchical representation; one that always consists of precisely two levels. More importantly, we explicitly demonstrated that the MSBN may not represent all the independencies encoded in a BN. This has important practical ramifications. As Ex. 7

explicitly shows, failure to represent known independencies results in unnecessary computation in the MSBN representation. Moreover, the MSBN approach is not unique, makes assumptions, and imposes technical constraints. Xiang et al. [15] recently acknowledged that how to satisfy these technical constraints may not be obvious to a practitioner. On the contrary, the HMN is a unique and equivalent representation of BNs, which does not require assumptions, technical constraints, and practitioner input. Our analysis then suggests that HMNs may be a better representation of BNs than MSBNs.

References

1. C.J. Butz, Q. Hu and X.D. Yang: Critical remarks on the maximal prime decomposition of Bayesian networks. To appear in *Proc. 9th International Conf. on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing* 2003.
2. U. Kjaerulff: Nested junction trees. In *Proc. 13th Conf. on Uncertainty in Artificial Intelligence*, 302–313, 1997.
3. D. Koller and A. Pfeffer: Object-oriented Bayesian networks. In *Thirteenth Conference on Uncertainty in Artificial Intelligence*, 302–313, 1997.
4. K.G. Olesen and A.L. Madsen: Maximal prime subgraph decomposition of Bayesian networks. *IEEE Transactions on Systems, Man, and Cybernetics*, B, 32(1):21-31, 2002.
5. J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Francisco, California, 1988.
6. R.D. Shachter: A graph-based inference method for conditional independence. *Proc. 7th Conf. on Uncertainty in Artificial Intelligence*, 353–360, 1991.
7. S. Srinivas: A probabilistic approach to hierarchical model-based diagnosis. *Proc. 10th Conf. on Uncertainty in Artificial Intelligence*, 538–545, 1994.
8. S.K.M. Wong and C.J. Butz: Constructing the dependency structure of a multi-agent probabilistic network. *IEEE Trans. Knowl. Data Eng.*, 13(3):395–415, 2001.
9. S.K.M. Wong, C.J. Butz, and D. Wu: On the implication problem for probabilistic conditional independency. *IEEE Transactions on Systems, Man, and Cybernetics*, A, 30(6): 785–805, 2000.
10. S.K.M. Wong, C.J. Butz, and D. Wu: On undirected representations of Bayesian networks. *ACM SIGIR Workshop on Mathematical/Formal Models in Information Retrieval*, 52-59, 2001.
11. Y. Xiang: Optimization of inter-subnet belief updating in multiply sectioned Bayesian networks. *Proc. 11th Conf. on Uncertainty in Artificial Intelligence*, 565–573, 1995.
12. Y. Xiang: *Probabilistic Reasoning in Multi-Agent Systems: A Graphical Models Approach*. Cambridge Publishers, 2002.
13. Y. Xiang, D. Poole, and M. Beddoes: Exploring localization in Bayesian networks for large expert systems. *Proc. 8th Conf. on Uncertainty in Artificial Intelligence*, 344–351, 1992.
14. Y. Xiang, D. Poole, and M. Beddoes: Multiply sectioned Bayesian networks and junction forests for large knowledge based systems. *Computational Intelligence*, 9(2):171–220, 1993.
15. Y. Xiang, K.G. Olesen, and F.V. Jensen. Practical issues in modeling large diagnostic systems with multiply sectioned Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 14(1):59–71, 2000.