What Necessitate Multiply Sectioned Bayesian Networks?

Yang Xiang

Technical Report CS-97-05
Dec. 1997

Department of Computer Science
University of Regina
Regina, Saskatchewan, CANADA
S4S 0A2

ISSN 0828-3494
ISBN 0-7731-0352-X

What Necessitate Multiply Sectioned Bayesian Networks?

Y. Xiang
Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2, yxiang@cs.uregina.ca

Abstract

Multiply sectioned Bayesian networks (MSBNs) provide a coherent framework for probabilistic reasoning in cooperative multi-agent distributed interpretation systems (CMADISs). Previous work on MSBNs focuses on the *sufficiency* of MSBNs for representation and inference with uncertain knowledge in CMADISs. Since several representation choices were made in the formation of a MSBN, it appears unclear whether certain choices were necessary. For example, it is unclear why a hypertree organization of agents was imposed.

This study focuses on the *necessity* of MSBNs for representation of uncertain knowledge in CMADISs. We identify a small set of fundamental choices which logically implies a MSBN or some equivalent representations. We consider privacy of agents to be essential if we are to allow agents developed by *independent* vendors so that vendors' know-how can be protected. We found that the privacy of agents plays an important role in this necessity analysis. The study provides insights into the MSBN framework and valuable guidances to multiagent system researchers whether they are satisfied with the framework or unsatisfied with the restrictions imposed.

1 Introduction

Multiply sectioned Bayesian networks (MSBNs) provide a coherent framework for probabilistic reasoning in cooperative multi-agent distributed interpretation systems (CMADISs) (15). MSBNs are an extension of Bayesian networks (BNs) (11). A MSBN consists of a set of interrelated Bayesian subnets that collectively define a BN (16). Each subnet encodes an agent's uncertain knowledge about a subdomain. Agents are organized into a hypertree structure such that probabilistic inference can be performed coherently in a modular and distributed fashion. Previous work by Xiang (15) establishes the *sufficiency* of MSBNs for representation and inference with uncertain knowledge in CMADISs. Since several technical choices were made in the formation of MSBN framework, it appears unclear whether certain choices were necessary. For example, it is unclear why a hypertree organization of agents was imposed.

In this study, we focus on the *necessity* of MSBNs for representation of uncertain knowledge in CMADISs. We identify the choice points in the formation of MSBN framework. We shall term some fundamental choices made in the process as the *basic commitments*. Given the basic commitments, other choices follow logically. We shall term these choices as *secondary commitments*, or simply *commitments*. When we refer to several commitments as a group, some basic and some secondary, we shall call them just *commitments*.

The identification of these basic commitments answers the question "what are the conditions under which a MSBN or some equivalent is the necessary representation?" It provides a high-level (vs. the technical level) description about the applicability of MSBNs and a valuable guidance to practitioners in CMADISs.

In Section 1, we briefly overview the theory of MSBNs taken from (16; 15). Each of the subsequent sections identifies some basic or secondary commitments and derives certain aspects of the MSBN framework. The framework logically follows when all the commitments and their consequences unfold, as summarized in Section 8.

2 Overview of MSBNs

A BN S is a triplet (N, D, P) where N is a set of domain variables, D is a DAG whose nodes are labeled by elements of N, and P is a joint probability distribution (jpd) over N. A MSBN M is a collection of Bayesian subnets that together defines a BN. These subnets are required to satisfy certain conditions that permit the construction of distributed inference algorithms. One of these conditions requires that nodes shared by different subnets form a *d-sepset*, as defined below.

Let $G_i = (N_i, E_i)$ $(i = 0, 1)$ be two graphs. The graph $G = (N_0 \cup N_1, E_0 \cup E_1)$ is referred to as the *union* of G_0 and G_1, denoted by $G = G_0 \sqcup G_1$.

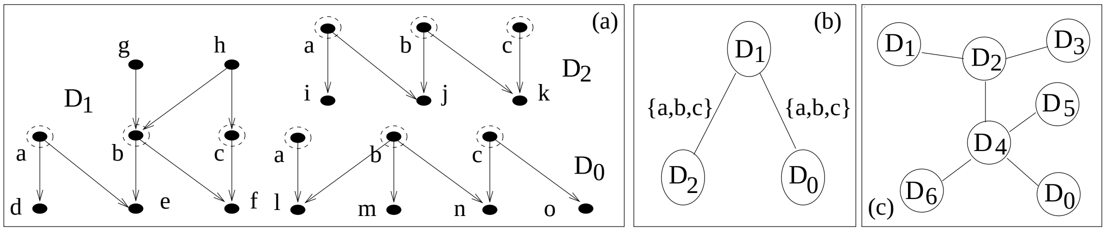

Figure 1: (a) DAGs of an MSBN with each d-sepnode shown with a dotted circle. (b) Hypertree organization of DAGs in (a). (c) A general hypertree MSDAG (unrelated to (a)).

Definition 1 *Let $D_i = (N_i, E_i)$ $(i = 0, 1)$ be two DAGs such that $D = D_0 \sqcup D_1$ is a DAG. The intersection $I = N_0 \cap N_1$ is a* `d-sepset` *between D_0 and D_1 if for every $x \in I$ with its parents π in D, either $\pi \subseteq N_0$ or $\pi \subseteq N_1$. Each $x \in I$ is called a* `d-sepnode`.

Using the concept of d-separation (11), it has been shown that when a pair of subnets are isolated from M, their d-sepset renders them conditionally independent. Figure 1 (a) shows three DAGs D_i $(i = 0, 1, 2)$ of a MSBN with the d-sepset between each pair as $\{a, b, c\}$ although in general, d-sepsets between different pairs of DAGs may differ.

Just as the structure of a BN is a DAG, the structure of a MSBN is a multiply sectioned DAG (MSDAG) with a hypertree organization:

Definition 2 *A* `hypertree MSDAG` *$\mathcal{D} = \bigsqcup_i D_i$, where each D_i is a connected DAG, is a DAG built by the following procedure:*

Start with an empty graph (no node). Recursively add a DAG D_k, called a `hypernode`, *to the existing MSDAG $\bigsqcup_{i=0}^{k-1} D_i$ subject to the constraints:*

[d-sepset] For each D_j $(j < k)$, $I_{jk} = N_j \cap N_k$ is a d-sepset when the two DAGs are isolated.

[Local covering] There exists D_i $(i < k)$ such that, for each D_j $(j < k; j \neq i)$, we have $I_{jk} \subseteq N_i$. For an arbitrarily chosen such D_i, I_{ik} is the `hyperlink` *between D_i and D_k which are said to be* `adjacent`.

Note that a hypertree MSDAG is a *tree* where each node is a hypernode as defined above and each link is a hyperlink. The DAGs in Figure 1 (a) can be organized into the trivial hypertree MSDAG in (b), where each hypernode is labeled by a DAG and each hyperlink is labeled by a d-sepset. Figure 1 (c) depicts a general hypertree MSDAG. Although DAGs of a MSBN should be organized into a hypertree, each DAG may be multiply connected, e.g., D_1. Moreover, there can be multiple paths between a pair of nodes in different DAGs in a hypertree MSDAG. For instance, multiple paths are formed between k and n after D_2 and D_0 are unioned. The local covering condition ensures that for any undirected cycle across two adjacent DAGs, both of its two paths are through the corresponding d-sepset. Together with the d-sepset condition, they ensure that in a hypertree structured M, each hyperlink renders the two parts of M that it connects conditionally independent. A MSBN is then defined as follows:

Definition 3 *A MSBN M is a triplet $(\mathcal{N}, \mathcal{D}, \mathcal{P})$. $\mathcal{N} = \bigcup_i N_i$ is the* `total universe` *where each N_i is a set of variables. $\mathcal{D} = \bigsqcup_i D_i$ (a hypertree MSDAG) is the* `structure` *where nodes of each DAG D_i are labeled by elements of N_i. $\mathcal{P} = \prod_i P_i(N_i) / \prod_k P_k(I_k)$ is the* `jpd`. *Each $P_i(N_i)$ is a distribution over N_i such that whenever D_i and D_j are adjacent in $\mathcal{D}$, the marginalizations of $P_i(N_i)$ and $P_j(N_j)$ onto the d-sepset I_{ij} are identical. Each $P_k(I_k)$ is such a marginal distribution over a hyperlink of $\mathcal{D}$. Each triplet $S_i = (N_i, D_i, P_i)$ is called a* `subnet` *of M.*

Two subnets S_i and S_j are said to be adjacent if D_i and D_j are adjacent.

MSBNs forms a coherent framework for probabilistic reasoning in CMADISs. Each agent holds its partial perspective of a large problem domain, accesses a local evidence source, communicates with other agents *infrequently*, reasons with the local evidence and limited global evidence, and answers queries or takes actions. It has been shown that if all agents are cooperative (vs self-interested), and each pair of adjacent agents are conditionally independent given their shared variables and have common initial belief on the shared variables, then a joint system belief is well defined which is identical to each agent's belief within its subdomain and supplemental to the agent's belief outside the subdomain. Even though multiple agents may acquire evidence asynchronously in parallel, the communication operations of MSBNs ensure that the answers to queries from each agent are consistent with evidence acquired in the entire system after each communication. Since communication is infrequent, the operations also ensure that between two successive communications, the answers to queries for each agent are consistent with all local evidence gathered so far and are consistent with all evidence gathered in the entire system up to the last communication. Therefore, a MSBN can be characterized as one of functionally accurate, cooperative distributed systems (9). Potential applications include decision support to cooperative human users in uncertain domains and troubleshooting a complex system by multiple knowledge based subsystems developed by independent vendors (15).

In the following sections, we identify major choices in uncertain knowledge representation in a CMADIS that lead to a MSBN.

3 Choice on measures of belief

In a CMADIS, agents are given the task to determine cooperatively what is true in a large uncertain problem domain. We shall use the terms *uncertain knowledge*, *belief* and *uncertainty* interchangeably.

Cox (2) demonstrated that the axioms of probability theory are a necessary consequence of intuitive properties of measures of belief. As summerized by Horvitz et al. (4), these properties can be termed as *clarity, scalar continuity, completeness, context dependency, hypothetical conditioning, complementarity*, and *consistency.*

We assume the *seven fundamental properties* and make them as our basic commitments. According to Cox, we must then accept probabilities or their monotonic transformations as agents' measures of belief. We shall term this as our commitment to *probability.* We shall use the term *coherence* to describe any assignment of measures of belief that is consistent with the probability theory, as we have been up to this point in the paper.

We consider a problem domain consisting of a total universe $\mathcal{N}$ of variables over which a CMADIS of n agents $A_0, A_1, ..., A_{n-1}$ is defined. Each agent A_i has its local knowledge over a subset $N_i \subset \mathcal{N}$, called the *subdomain* of the agent. From our commitment to probability, it follows that this knowledge takes the form of a probability distribution over N_i, denoted by $P_i(N_i)$.

4 Choice on privacy of agents

We assume that for each agent A_i, its knowledge over N_i is *private.* The privacy has two levels: At the first level, variables of N_i that are not shared by other agents are known only to A_i. Formally, if agent A_j shares variables $N_j \cap N_i \neq \phi$ with agent A_i, then elements of $N_i \setminus N_j$ are unknown to A_j.

At the second level, except the marginal distributions (marginals) of $P_i(N_i)$ onto variables shared with other agents, the exact form of $P_i(N_i)$ is unknown to other agents. Formally, if agent A_j shares variables $N_j \cap N_i \neq \phi$ with agent A_i, then A_i only makes $\sum_{N_i \setminus N_j} P_i(N_i)$ known to A_j but not $P_i(N_i)$. We shall term this as our basic commitment to *privacy of agents.* The privacy is essential if we are to allow agents developed by *independent* vendors so that vendors' know-how can be protected.

We take it granted that in order for two agents to communicate probabilistic knowledge without jeopardizing their privacy, they must share a subset of common variables. That is, if agents A_i and A_j are to communicate *directly*, it must be the case that $N_i \cap N_j \neq \phi$. We shall term this as our basic commitment to *communication by shared variables.*

The paths for direct communication can be represented graphically as follows: Construct a graph with n nodes. Each node corresponds to an agent A_i and is labeled by N_i. Hence the graph has an one-to-one mapping between nodes of the graph and the agents of the CMADIS. For each pair of nodes N_i and N_j, connect them by a link labeled by $I = N_i \cap N_j$ (called a *separator*) if $I \neq \phi$. The resultant is a *junction graph* (5) whose links represent all potential paths of *direct* communication between agents. That is, an agent can directly communicate with another agent if and only if there is a link between them in the junction graph.

Furthermore, since the knowledge of one agent can influence the knowledge of another agent through a third agent, the junction graph also represents all potential paths of *indirect* communications between agents. In a CMADIS, each agent's knowledge should potentially be influential in any other agent, directly or indirectly. Otherwise the system can be split into two separate systems without affecting the performance of each. This is equivalent to the condition that there exists a path between any pair of nodes, which implies that the junction graph is *connected.*

From our commitments to probability, to privacy of agents, and to communication by shared variables, it also follows that the *content* of direct communication between a pair of agents must be a probability distribution over the common variables. We shall refer to this distribution as a *message* and direct communication as *message passing.*

5 On the necessity of hypertree

The difficulty of coherent inference in multiply connected (with loops) graphical models of probabilistic knowledge is well known and many inference algorithms have been proposed to tackle the issue, e.g., (11; 8; 12; 3; 6; 13). Those algorithms that are based on message passing, e.g., (11; 8; 6; 13), all explore the tree topology by converting a multiply connected network into a tree. However, when one searches through the literature, e.g., (11; 5; 10; 1), *no* formal arguments can be found which demonstrate convincingly that message passing *cannot* be made coherent in multiply connected networks. This leaves the question whether it is impossible to construct a method of coherent message passing in multiply connected networks or it is possible bu the method remains to be found. In what follows, we show that the former is the case.

A junction graph can be multiply connected in general. A loop is *degenerate* if all separators on the loop are identical (Figure 2 (a)). Otherwise, the loop is *nondegenerate* ((b) and (c)). In general, a junction graph can have both types of loops.

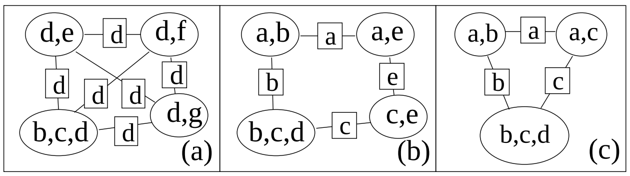

Figure 2: Junction graphs with nodes shown in ovals and separators shown in boxes.

5.1 Degenerate loops

We shall refer to a node in a junction graph as a *cluster*. Clusters in a degenerate loop must be *complete* (pairwise connected) since each pair has the identical intersection. These clusters can be organized into a star subgraph by arbitrarily selecting one cluster as the center and removing links between other clusters. Message passing through a degenerate loop can then be performed using the star structure coherently[1] and efficiently as follows:

If observations are received at a peripheral cluster, propagation of new belief first from the cluster to the center and then from the center to all other clusters will update all clusters in the star. Observations received at multiple clusters can be propagated to the center, combined and distributed back to update all clusters (11). Each agent, either at the center or not, can respond to messages efficiently. If the agent is not at the center, it sends a message to the center when observations are received and updates belief when the reply from the center is received. If the agent is at the center, it receives messages from other clusters, updates its own belief and then distributes updated belief in the reply messages.

Although the same coherence could be achieved using the original complete structure, the communication cannot be performed as efficiently as mentioned above. Now every agent has the same number of message paths. No one knows who should assume the role of center. Either a logical center is selected through some voting mechanism (as there exists no physical center), which after incurring the voting overhead returns to the above star structure, or agents will communicate randomly, which in the best case may (or may never) reach coherent belief after more messages are sent than using the star structure.

When everything else is the same, we shall prefer a formalism that is more efficient. We shall term this as our basic commitment to *efficiency*. Based on the above analysis, it follows that a star structure is *always* more efficient than the complete structure for a degenerate loop. Hence, we shall insist in using a star structure to organize clusters in every degenerate loop.

5.2 Nondegenerate loops

We show that in general belief updating in a nondegenerate loop *cannot* be performed coherently through message passing. In what follows, we construct a CMADIS with a nondegenerate loop where no message passing can update agents' belief coherently.

Consider a simple CMADIS of three agents A_0, A_1 and A_2 with $U_0 = \{a, b\}$, $U_1 = \{a, c\}$ and $U_2 = \{b, c, d\}$, where a, b, c, d are binary variables[2]. Figure 2 (c) shows the corresponding junction graph.

The local knowledge of three agents are $P_0(a, b)$, $P_1(a, c)$ and $P_2(b, c, d)$, respectively. We assume that their belief are initially consistent, namely, the marginal distributions satisfy $P_0(a) = P_1(a)$, $P_0(b) = P_2(b)$, and $P_1(c) = P_2(c)$. Since their belief are consistent, message passing cannot change any agent's belief. We shall refer to this CMADIS as *Cmas3*. Any given $P_0(a, b)$, $P_1(a, c)$ and $P_2(b, c, d)$ subject to the above consistency constraint is called an *initial knowledge state* of Cmas3.

Suppose that agent A_2 subsequently observes $d = d_0$. If the agents can update their belief coherently, then their new belief should be $P_0(a, b|d = d_0)$, $P_1(a, c|d = d_0)$ and $P_2(b, c, d|d = d_0)$. For A_2, $P_2(b, c, d|d = d_0)$ can be obtained locally. However, for A_0 and A_1 to update their belief, they must rely on communication, namely, the message $P_2(b|d = d_0)$ sent by A_2 to A_0 and the message $P_2(c|d = d_0)$ sent by A_2 to A_1.

Clearly, the new belief of A_0 and A_1, $P_0(a, b|d = d_0)$ and $P_1(a, c|d = d_0)$, should be sensitive to A_2's initial knowledge $P_2(b, c, d)$. In other words, everything else being the same, given different initial knowledge $P_2(b, c, d)$ of A_2, the new belief of A_0 and A_1 should be different as well. In the following theorem, we show that this is not always the case, which disproves that

[1] In fact, the coherence depends on an additional requirement of conditional independence to be discussed in Section 6.

[2] For those who might think this system to be trivial, we can make each of a, b, c, d a set of variables and our conclusion can still be drawn by the same argument.

it is always possible for agents to update their belief coherently in a nondegenerate loop.

Theorem 4 *There exists an infinite set of initial knowledge states of Cmas3 such that the following conditions hold:*

1. *At each state in the set,* $P_0(a,b)$ *is identical and so are* $P_1(a,c)$ *and* $P_2(b,c)$.
2. $P_2(d|b,c)$ *at each state in the set is distinctive.*
3. *At each state in the set, the resultant message* $P_2(b|d=d_0)$ *is identical and so is the message* $P_2(c|d=d_0)$.

Before proving the theorem, we give an intuitive interpretation. Condition 1 says that the initial knowledge of A_0 and A_1, and part of the initial knowledge of A_2 ($P_2(b,c)$) remain the same across the states. Condition 2 says that the initial knowledge of A_2 is different across the states since $P_2(b,c,d) = P_2(d|b,c) * P_2(b,c)$. Note that since $P_2(b,c)$ remains the same, the consistency among agents is maintained even though $P_2(b,c,d)$ differs across the states. Condition 3 says that the difference in the initial knowledge of A_2 cannot cause different new belief in A_0 and A_1, which is the conclusion we want to establish.

Proof:

Without losing generality, we assume that all distributions involved are strictly positive. To simplify notations, we shall denote the message component $P_2(b=b_0|d=d_0)$ by $P_2(b_0|d_0)$. It can be expanded as

$$P_2(b_0|d_0) = P_2(b_0,d_0)/(P_2(b_0,d_0)+P_2(b_1,d_0))$$

$$= \frac{1}{1+\frac{P_2(b_1,d_0)}{P_2(b_0,d_0)}} = \frac{1}{1+\frac{P_2(b_1,c_0,d_0)+P_2(b_1,c_1,d_0)}{P_2(b_0,c_0,d_0)+P_2(b_0,c_1,d_0)}}$$

$$= \frac{1}{1+\frac{P_2(d_0|b_1,c_0)P_2(b_1,c_0)+P_2(d_0|b_1,c_1)P_2(b_1,c_1)}{P_2(d_0|b_0,c_0)P_2(b_0,c_0)+P_2(d_0|b_0,c_1)P_2(b_0,c_1)}}.$$

Similarly, the message component $P_2(c_0|d_0)$ can be expanded as

$$P_2(c_0|d_0) = \frac{1}{1+\frac{P_2(d_0|b_0,c_1)P_2(b_0,c_1)+P_2(d_0|b_1,c_1)P_2(b_1,c_1)}{P_2(d_0|b_0,c_0)P_2(b_0,c_0)+P_2(d_0|b_1,c_0)P_2(b_1,c_0)}}.$$

We shall use s^0 to denote a particular initial knowledge state and label an agent's knowledge at s^0 by a superscript (e.g., $P_2^0(d|b,c)$). We now consider a different state s^1 that satisfies $P_0^1(a,b) = P_0^0(a,b)$, $P_1^1(a,c) = P_1^0(a,c)$ and $P_2^1(b,c) = P_2^0(b,c)$ (condition 1). If agent A_2 at s^1 can generate the identical messages $P_2^1(b|d_0) = P_2^0(b|d_0)$ and $P_2^1(c|d_0) = P_2^0(c|d_0)$ (condition 3), then $P_2^1(d|b,c)$ must be the solutions of the following equations:

$$\frac{P_2^1(d_0|b_1,c_0)P_2^0(b_1,c_0)+P_2^1(d_0|b_1,c_1)P_2^0(b_1,c_1)}{P_2^1(d_0|b_0,c_0)P_2^0(b_0,c_0)+P_2^1(d_0|b_0,c_1)P_2^0(b_0,c_1)}$$

$$= \frac{P_2^0(d_0|b_1,c_0)P_2^0(b_1,c_0)+P_2^0(d_0|b_1,c_1)P_2^0(b_1,c_1)}{P_2^0(d_0|b_0,c_0)P_2^0(b_0,c_0)+P_2^0(d_0|b_0,c_1)P_2^0(b_0,c_1)}$$

and

$$\frac{P_2^1(d_0|b_0,c_1)P_2^0(b_0,c_1)+P_2^1(d_0|b_1,c_1)P_2^0(b_1,c_1)}{P_2^1(d_0|b_0,c_0)P_2^0(b_0,c_0)+P_2^1(d_0|b_1,c_0)P_2^0(b_1,c_0)}$$

$$= \frac{P_2^0(d_0|b_0,c_1)P_2^0(b_0,c_1)+P_2^0(d_0|b_1,c_1)P_2^0(b_1,c_1)}{P_2^0(d_0|b_0,c_0)P_2^0(b_0,c_0)+P_2^0(d_0|b_1,c_0)P_2^0(b_1,c_0)}.$$

Since $P_2^1(d|b,c)$ has four independent parameters but is constrained by only two equations, it has *infinite* number of solutions (condition 3). Each solution defines an initial knowledge state of the Cmas3 that satisfies all conditions in the theorem. □

In the following, we give an example that further illustrates the implication of Theorem 4.

Example 5 Let the initial knowledge state s^0 of Cmas3 be defined as follows:

$p_0^0(a_0,b_0) = 0.2548$	$p_0^0(a_0,b_1) = 0.0052$
$p_0^0(a_1,b_0) = 0.2442$	$p_0^0(a_1,b_1) = 0.4958$
$p_1^0(a_0,c_0) = 0.0052$	$p_1^0(a_0,c_1) = 0.2548$
$p_1^0(a_1,c_0) = 0.4958$	$p_1^0(a_1,c_1) = 0.2442$
$p_2^0(b_0,c_0) = 0.16871$	$p_2^0(b_0,c_1) = 0.33029$
$p_2^0(b_1,c_0) = 0.33229$	$p_2^0(b_1,c_1) = 0.16871$
$p_2^0(d_0|b_0,c_0) = 0.03$	$p_2^0(d_0|b_0,c_1) = 0.66$
$p_2^0(d_0|b_1,c_0) = 0.7$	$p_2^0(d_0|b_1,c_1) = 0.25$

Let the initial knowledge state s^1 be defined identically except the following:

$p_2^1(d_0|b_0,c_0) = 0.533604$	$p_2^1(d_0|b_0,c_1) = 0.115431$
$p_2^1(d_0|b_1,c_0) = 0.14$	$p_2^1(d_0|b_1,c_1) = 0.66$

Since $p_0^0(a_0) = p_1^0(a_0) = 0.26$, A_0 and A_1 are consistent. Since $p_0^0(b_0) = p_2^0(b_0) = 0.499$, A_0 and A_2 are consistent. Since $p_1^0(c_0) = p_2^0(c_0) = 0.501$, A_1 and A_2 are consistent.

In both states, after $d = d_0$ is observed by A_2, its messages are $p_2^0(b_0|d_0) = p_2^1(b_0|d_0) = 0.448$ and $p_2^0(c_0|d_0) = p_2^1(c_0|d_0) = 0.477$. Hence, A_0 and A_1 will be unable to update their belief differently.

Next, we illustrate the difference in the new belief produced by a coherent probabilistic inference. To this end, we assume the following independence relations among the variables monitored by the three agents: Variables b and c are conditionally independent given a, and a and d are conditionally independent given b and c. Note that these assumptions are fully *consistent* with agents' knowledge specified above. Assuming these relations and assuming that each agent's knowledge is correct within its subdomain, we can derive the jpd of the domain as

$$p(a, b, c, d) = p_0(a, b)p_1(c|a)p_2(d|b, c).$$

Using this jpd and coherent probabilistic calculation, from s^0, A_0 and A_1 should update their belief on a to $p^0(a_1|d_0) = 0.666$. From s^1, on the other hand, A_0 and A_1 should update their belief on a to $p^1(a_1|d_0) = 0.878$. The difference is significant.

From Theorem 4, it follows that in general belief updating cannot be performed coherently through message passing in a nondegenerate loop. Since replacing a degenerate loop by a star structure renders inference more efficient, and coherent inference precludes nondegenerate loops, it follows from our commitments to probability and to efficiency that agents should be organized into a structure that has *no* loops. The only such structure is a *tree* structure. In other words, some paths in the junction graph should be disallowed such that the resultant subgraph is a tree. We shall term this as our commitment to a *hypertree organization*.

6 On conditional independent separators

Given our commitment to hypertree organization, it follows that separators in the hypertree play an important role in agents' communication as each separator is the only information channel between the two subtrees that it separates. Therefore, each separator must be chosen such that its distribution (which is the message passed over the corresponding link) is *always sufficient* to convey all the relevant information from one subtree to the other. Formally, this means that all variables in one subtree are conditionally independent of all variables in the other subtree given the separator.

It can be shown formally that when the separator renders the two subtrees conditionally independent, if new observations are obtained in one subtree, coherent belief update in the other subtree can be achieved by simply passing the updated distribution on the separator. On the other hand, if the separator does not render the two subtrees conditionally independent, belief updating by passing only the separator distribution will not be coherent in general. We shall term this as our commitment to *conditional independent separators*.

This commitment requires the CMADIS designer to partition the domain among agents such that intersections of subdomains form conditional independent separators in a hypertree organization. It has been shown (15) that if agents are cooperative, the subdomains are organized into a hypertree, the links on the hypertree are conditionally independent separators, and agents' belief on their separators are consistent, then a unique joint belief of all agents in the CMADIS exists and can be expressed as

$$p(\mathcal{N}) = (\prod_i p_i(N_i))/(\prod_j p_j(I_j)),$$

where each I_j is a separator in the hypertree such that A_j is one of the two agents with $N_j \supset I_j$.

7 Choice on subdomain representation

Although probability theory follows from the seven fundamental properties for measures of belief, the traditional representation using the jpd is impractical (14) and against our commitment to efficiency. Given a problem domain of k variables, the jpd is specified by $O(2^k)$ parameters, which imposes heavy burden to knowledge acquisition.

To make the representation practical, the jpd must be specified *compactly* in terms of distributions over small subsets of the domain variables. This requirement gives rise to the representation of jpd using Bayesian networks (BNs) (11), decomposable Markov networks (DMNs) (17), and chain graphs (CGs) (7). All these representations factorize the jpd into distributions over small subsets of variables by exploring conditional independence among subsets portrayed by a graphical model. The jpd is then defined coherently by these distributions, for example, according to the *chain rule* in BNs (11). We shall refer to these compact representations based on graphical dependence models as *belief networks*.

In the previous section, we have already factorized the jpd according to a hypertree topology. Since each subdomain may still be large enough to preclude the representation of $p_i(N_i)$ as an unstructured distribution, it then follows that for each subdomain N_i, $p_i(N_i)$ should itself be structured into a belief network. We shall focus on the structuring using BNs and term this as our basic commitment to *DAG models*. When subdomains adjacent in the hypertree are structured into BNs, they impose constraints to each other through their separators. Each subdomain is structured into a DAG. However, when DAGs for different subdomains are joined together, they may form a directed loop. This implies that if individual subdomains are structured into DAGs, the entire domain should be structured into something like a MSDAG. Now the only gap to arrive at a MSBN is to show that the separator between subdomains must be structured as a d-sepset. This is established in Theorem 6 through the concept of d-separation (11). Since d-separation captures all graphically identifiable conditional independencies, the theorem shows that d-sepset is the necessary and sufficient *syntactic* condition for conditionally independent separators.

Theorem 6 *Let $D_i = (N_i, E_i)$ $(i = 0, 1)$ be two DAGs such that $D = D_0 \sqcup D_1$ is a DAG. $N_0 \setminus N_1$ and $N_1 \setminus N_0$ are d-separated by $I = N_0 \cap N_1$ iff I is a d-sepset.*

Proof:

We show only the necessity as the sufficiency has been shown in (16).

Suppose there exists $x \in I$ with distinct parents y and z in D such that $y \in N_0$ but $y \notin N_1$, and $z \in N_1$ but $z \notin N_0$. Note that the condition disqualifies I from being a d-sepset, and this is the only way that I may become disqualified. Now y and z are not d-separated given x and hence $N_0 \setminus N_1$ and $N_1 \setminus N_0$ are not d-separated by I. □

8 Conclusion

Throughout our analysis, we have made the following *basic* commitments:

1. Seven fundamental properties of measures of belief.
2. Privacy of agents.
3. Communication by shared variables.
4. Efficiency in inference.
5. DAG models for subdomain structuring.

From these basic commitments, we have shown that subdomains controlled by agents should each be represented as a Bayesian subnet, should be organized into a hypertree, and should be separated by d-sepsets. From some simple graphical arguments, it then follows that the structure of the resultant CMADIS is a MS-DAG (Definition 2). By further comparison between the equation in Section 6 and Definition 3, it follows that the resultant representation is a MSBN. In summary, we have shown that from these basic commitments, MSBNs or some equivalent formalisms follow logically as the representation of uncertain knowledge in CMADISs.

This result provides insight to the MSBN framework by revealing the fundamental properties of cooperative multiagent uncertain reasoning that are behind the technical details of the framework. Practitioners who find these basic commitments satisfactory are assured a representation formalism that delivers what is possible given these commitments. Researchers who find the restrictions imposed by MSBNs unsatisfactory are directed to evaluation and relaxation of these basic commitments rather than the technical details of MSBNs.

References

[1] E. Castillo, J. Gutierrez, and A. Hadi. *Expert Systems and Probabilistic Network Models.* Springer, 1997.

[2] R.T. Cox. Probability, frequency and reasonable expectation. *American J. of Physics*, 14(1):1–13, 1946.

[3] M. Henrion. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In J.F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 149–163. Elsevier Science Publishers, 1988.

[4] E. Horvitz, D.E. Heckerman, and C.P. Langlotz. A framework for comparing alternative formalisms for plausible reasoning. In *Proc. 5th Natl. Conf. on Artificial Intelligence*, pages 210–214, Philadelphia, 1986.

[5] F.V. Jensen. *An introduction to Bayesian networks.* UCL Press, 1996.

[6] F.V. Jensen, S.L. Lauritzen, and K.G. Olesen. Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly*, (4):269–282, 1990.

[7] S.L. Lauritzen. *Graphical models.* Oxford, 1996.

[8] S.L. Lauritzen and D.J. Spiegelhalter. Local computation with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society, Series B*, (50):157–244, 1988.

[9] V.R. Lesser and D.D. Corkill. Functionally accurate, cooperative distributed systems. *IEEE Trans. on Systems, Man and Cybernetics*, SMC-11(1):81–96, 1981.

[10] R.E. Neapolitan. *Probabilistic Reasoning in Expert Systems.* John Wiley and Sons, 1990.

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

[12] R.D. Shachter. Probabilistic inference and influence diagrams. *Operations Research*, 36(4):589–604, 1988.

[13] R.D. Shachter, B. D'Ambrosio, and B.A. Del Favero. Symbolic probabilistic inference in belief networks. In *Proc. 8th Natl. Conf. on Artificial Intelligence*, pages 126–131, 1990.

[14] P. Szolovits and S.G. Pauker. Categorical and probabilistic reasoning in medical diagnosis. *Artificial Intelligence*, (11):115–144, 1978.

[15] Y. Xiang. A probabilistic framework for cooperative multi-agent distributed interpretation and optimization of communication. *Artificial Intelligence*, 87(1-2):295–342, 1996.

[16] Y. Xiang, D. Poole, and M. P. Beddoes. Multiply sectioned Bayesian networks and junction forests for large knowledge based systems. *Computational Intelligence*, 9(2):171–220, 1993.

[17] Y. Xiang, S.K.M. Wong, and N. Cercone. A 'microscopic' study of minimum entropy search in learning decomposable Markov networks. *Machine Learning*, 26(1):65–92, 1997.