Justifying Multiply Sectioned Bayesian Networks

Y. Xiang and V. Lesser
Department of Computer Science
University of Massachusetts
Amherst, MA 01003, USA
{yxiang, lesser}@cs.umass.edu

(To appear in ICMAS-2000)

Abstract

We consider multiple agents who's task is to determine the true state of a uncertain domain so they can act properly. If each agent only has partial knowledge about the domain and local observation, how can agents accomplish the task with the least amount of communication? Multiply sectioned Bayesian networks (MSBNs) provide an effective and exact framework for such a task but also impose a set of constraints. The most notable is the hypertree agent organization which prevents an agent from communicating with arbitrarily another agent. Are there simpler frameworks with the same performance but with less restrictions?

We identify a small set of high level choices which logically imply the key representational choices made in MSBNs. The result addresses concerns regarding the necessity of restrictions of the framework. It facilitates comparison with related frameworks and provides guidance to extension of the framework as what can or cannot be traded off.

(**Keywords:** Decentralized interpretation, communication, organization structure, uncertain reasoning, belief network)

1 Introduction

Consider a large uncertain domain populated by a set of agents. The agents' task is to determine what is the true state of the domain so they can act upon it. We can describe the domain with a set of variables. Some variables are not directly observable hence their values can only be inferred based on observation of other variables and background knowledge on their dependence. Furthermore, each agent may only have knowledge about a subset of variables, and can only observe and reason within the subset. How can agents cooperate to accomplish the task with the least amount of communication? We shall term this type of agent systems as cooperative multi-agent distributed interpretation systems (CMADISs).

In the case of a single agent, the problem can be solved by representing the domain knowledge in a Bayesian network (BN) [13] and by performing inference in the BN given observations. As the domain becomes larger and more complex, however, a multiagent solution will be desirable. How should the domain be partitioned among agents? How should each agent represent its subdomain? How should the agents be organized in their activity? What information should they exchange and how, in order to minimize the amount of communication? Can they achieve the same level of accuracy in interpreting the state of the domain as a single agent?

Multiply sectioned Bayesian networks (MSBNs) [16] provide one solution to these issues. A MSBN consists of a set of interrelated Bayesian subnets each of which encodes an agent's knowledge on a subdomain. Agents are organized into a hypertree structure such that inference can be performed in a distributed fashion while answers to queries are exact with respect to probability theory. Each agent only exchanges information with adjacent agents on the hypertree, and each pair of adjacent agents only exchange information on a set of shared variables. The complexity of communication among all agents is linear on the number of agents and the complexity of local inference is the same as if the subnet is a single agent based BN.

Are there simpler alternatives that can achieve the same performance? In other words, are the technical restrictions of MSBN necessary? For example, the hypertree organization of agents prevents an agent from communicating with arbitrarily another agent. Is this necessary? If the answers to these questions are negative, then such concerns are counter-productive and hinders the adoption of MSBN to suitable CMADIS applications.

In this work, we try to address these concerns. We show that given some reasonable fundamental choice/assumptions, the key restrictions of a MSBN, such as a hypertree structure and a d-sepset (defined below) agent interface, are unavoidable. In particular, we identify the choice points in the formation of MSBN. We term fundamental choices as *basic commitments* (BCs). Given the BCs, other choices are entailed. Hence a MSBN or some equivalent follows once we admit the BCs.

The contributions are the following: First, the analysis provides a high-level (vs. technical level) description about the applicability of MSBN and addresses concerns regarding necessity of major restrictions. Second, the results facilitate comparison with alternative frameworks. Third, when needs for extension of MSBN or relaxation of its restrictions arise, the analysis provides a guideline as what can or cannot be traded off.

In Section 2, we briefly overview the MSBN framework with representational choices summarized. Each remaining section identifies some BCs and derives implied choices.

2 Overview of MSBNs

A BN [13] 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 [18, 16] M is a collection of Bayesian subnets that together defines a BN. These subnets are required to satisfy certain conditions. One condition 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 $G = G_0 \sqcup G_1$.

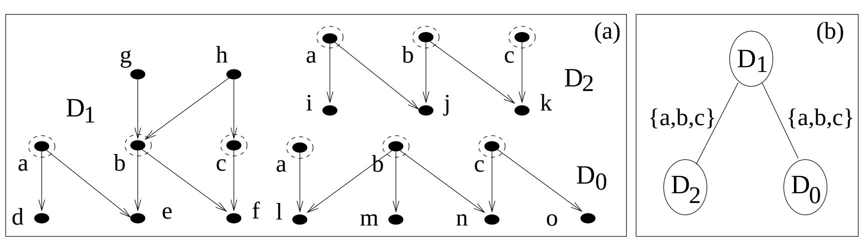

Figure 1: (a) DAGs of an MSBN. D-sepnodes are shown with dotted circles. (b) Hypertree organization of (a).

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`.

Fig. 1 (a) shows three DAGs D_i $(i = 0, 1, 2)$ of a MSBN with the d-sepset between each pair being $\{a, b, c\}$. 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 DAG, is a connected DAG constructible 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 and each link is a hyperlink. The DAGs in Fig. 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. Although DAGs are 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. For instance, multiple paths are formed between k and n after D_2 and D_0 are unioned. A MSBN is then defined as follows:

Definition 3 *An 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. Let x be a variable and $\pi(x)$ be all parents of x in $\mathcal{D}$. For each x, exactly one of its occurrences (in a D_i containing $\{x\} \cup \pi(x)$) is assigned $P(x|\pi(x))$, and each occurrence in other DAGs is assigned a constant table. $\mathcal{P} = \prod_i P_{D_i}$ is the* `jpd`*, where each P_{D_i} is the product of the probability tables associated with nodes in D_i. A triplet $S_i = (N_i, D_i, P_{D_i})$ 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 provide a framework for uncertain reasoning in CMADISs. Each agent holds its partial perspective (a subnet) of a total universe, reasons with local evidence and through communication with other agents, and answers queries or takes actions. Agents may be built by independent vendors with privacy protected with regard to the internal reasoning of each agent. Agents can acquire evidence in parallel while answers to queries are consistent with evidence in the entire system. Applications mostly studied include monitoring and diagnosis of large, complex and multi-component equipment [17] and object oriented BNs [7].

To aid the analysis, we list representational choices of MSBNs below, where the most important ones are 3 and 7.

1. Each agent's belief is represented by probability.
2. The total universe is decomposed into subdomains. For each pair, there exists a sequence of subdomains such that every pair of subdomains adjacent in the sequence shares some variables.
3. Subdomains are organized into a (hyper)tree structure where each hypernode is a subdomain, and each hyperlink represents a non-empty set of shared variables between the two hypernodes.
4. The hypertree satisfies local covering.
5. The dependency structure of each subdomain is represented as a DAG.
6. The union of DAGs for all subdomains is a connected DAG.
7. Each hyperlink is a d-sepset.
8. The joint probability distribution can be expressed as Def. 3.

Below we identify a set of BCs leading to these choices.

3 On connectivity of communication graph

We use *uncertain knowledge*, *belief* and *uncertainty* interchangeably, and make the following basic commitment:

BC 1 *Each agent's belief is represented by probability.*

It directly corresponds to the first choice of Section 2. We shall use *coherence* to describe any assignment of belief consistent with the probability theory.

We consider a total universe $\mathcal{N}$ of variables over which a CMADIS of n agents $A_0, ..., A_{n-1}$ is defined. Each A_i has knowledge over a $N_i \subset \mathcal{N}$, called the *subdomain* of A_i. It is assumed whenever $N_i \cap N_j \neq \emptyset$, the intersection is *small* relative to N_i and N_j. For example, in equipment diagnosis, each N_i is a component including all devices and their input/output. From BC 1, the knowledge of A_i is a probability distribution over N_i, denoted by $P_i(N_i)$.

To minimize communication, we allow agents to exchange only their belief on shared variables (BC 2 below). We take it for granted that for agents to communicate directly, $N_i \cap N_j$ must be nonempty. Note that BC 2 does not restrict the *order* nor the *number* of communications.

BC 2 *A_i and A_j can communicate* directly *only with $P(N_i \cap N_j)$.*

We refer to $P(N_i \cap N_j)$ as a *message* and to direct communication as *message passing*. Paths for message passing can be represented by a *communication graph* (CG): In a graph with n nodes, associate each node with an agent A_i and label it by N_i. Connect each pair of nodes N_i and N_j by a link labeled by $I = N_i \cap N_j$ (called a *separator*) if $I \neq \emptyset$. CG is a *junction graph* [4] over $\mathcal{N}$ whose links represent all potential paths of message passing. As belief of one agent can influence another through a third, CG also represents all potential paths of *indirect* communications. Each agent's belief should potentially be influential in any other, directly or indirectly. Otherwise the system can be split into two. Hence CG is *connected*. We summarize this in Proposition 4,. It is equivalent to the second choice in Section 2.

Proposition 4 *Let H be the communication graph of a CMADIS over $\mathcal{N}$ that observes BC 1 and BC 2. Each agent's belief can in general influence that of each other agent through communication. Then H is connected.*

4 On hypertree organization

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. Those based on message passing, e.g., [13, 9, 5, 15], all convert a multiply connected network into a tree. However, *no* formal arguments can be found, e.g., in [13, 4, 11, 1], which demonstrate convincingly that message passing *cannot* be made coherent in multiply connected networks. This leaves the question whether it is *impossible* to construct such a method or the method remains to be discovered.

The answer to this question ties closely to the necessity of hypertree organization of agents as specified in Def. 2 and restated as the third choice in Section 2. This tie can be seen by noting that the hypertree in Def. 2 is isomorphic to a subgraph of the communication graph H of the same CMADIS: An one-to-one mapping exists between hypernodes in Def. 2 and nodes in H. Each hyperlink in Def. 2 is a link in H but the converse is not true. In what follows, we show that in general, coherent message passing is impossible in multiply connected CGs. The result formally establishes not only the necessity of hypertree structure in CMADIS, but also the necessity of tree topology for message passing based inference in single agent systems. Since a CG is a junction graph, we use a junction graph in our analysis. We first classify loops as follows:

Definition 5 *Let G be a junction graph over $\mathcal{N}$. A loop in G is* `degenerate` *if all separators on the loop are identical. Otherwise, it is* `nondegenerate`*.*

In fig. 2, all loops in (a) are degenerate, and those in (b) and (c) are nondegenerate. In general, a junction graph can have both types of loops.

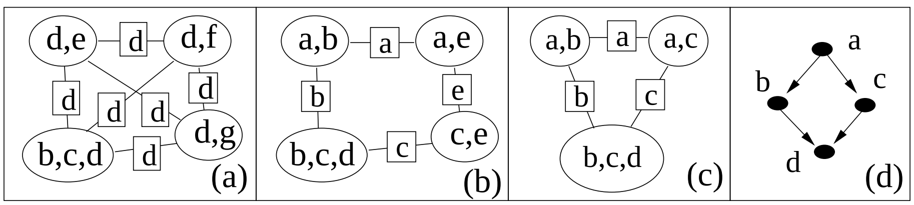

Figure 2: (a-c) Junction graphs with nodes shown in ovals and separators in boxes. (d) A DAG to which (c) is a junction graph.

4.1 Nondegenerate loops

We show that when nondegenerate loops exist, messages are *uninformative*. No matter how messages are manipulated or routed, they cannot become *informative* and it becomes impossible to make message passing coherent.

Consider a domain with the dependence structure in fig. 2 (d) where a, b, c, d are binary, over which a CMADIS of three agents A_i $(i = 0, 1, 2)$ with $U_0 = \{a, b\}$, $U_1 = \{a, c\}$ and $U_2 = \{b, c, d\}$ is defined. Fig. 2 (d) is the junction graph. The local knowledge of 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 $P_1(c) = P_2(c)$. Hence, message passing cannot change any agent's belief. We 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 is called an *initial state* of Cmas3.

Suppose that A_2 observes $d = d_0$. If the agents can update their belief coherently, 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 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. In the following, we show that A_0 and A_1 cannot update their belief coherently based on these messages. Before the general result, we illustrate with a particular initial state. From fig. 2(d), we can independently specify $P(a)$, $P(b|a)$, $P(c|a)$, and $P(d|b,c)$ as follows:

$$
\begin{array}{llll}
P(a_0) = 0.26 & & & \\
P(b_0|a_0) = 0.98 & P(b_0|a_1) = 0.33 & P(c_0|a_0) = 0.02 & P(c_0|a_1) = 0.67 \\
P(d_0|b_0,c_0) = 0.03 & P(d_0|b_0,c_1) = 0.66 & P(d_0|b_1,c_0) = 0.7 & P(d_0|b_1,c_1) = 0.25
\end{array}
$$

From these, we define an initial state s which is consistent:

$$P_0(a,b) = P(a)P(b|a), \quad P_1(a,c) = P(a)P(c|a), \quad P_2(b,c,d) = P(b,c)P(d|b,c),$$

where $P(b,c) = \sum_a P(a)P(b|a)P(c|a)$. After $d = d_0$ is observed by A_2, its messages are $P_2(b|d_0) = (0.448, 0.552)$ and $P_2(c|d_0) = (0.477, 0.532)$.

Consider now a different initial state s' that differs from s by replacing $P'(d|b,c)$ with the following:

$$P'_2(d_0|b_0,c_0) = 0.5336 \quad P'_2(d_0|b_0,c_1) = 0.1154 \quad P'_2(d_0|b_1,c_0) = 0.14 \quad P'_2(d_0|b_1,c_1) = 0.66$$

Note that $P'_2(b,c,d) \neq P_2(b,c,d)$, but $P'_0(a,b) = P_0(a,b)$ and $P'_1(a,c) = P_1(a,c)$. After $d = d_0$ is observed, if we compute the messages $P'_2(b|d_0)$ and $P'_2(c|d_0)$, we will find them to be identical to those obtained from state s. That is, the messages are insensitive to the difference between the two initial states. As a consequence, the new belief in A_0 and A_1 will be identical in both cases. Should the new belief in both cases be different? Using coherent probabilistic inference, we obtain $P(a_1|d_0) = 0.666$ from s, and $P'(a_1|d_0) = 0.878$ from s'. The difference is significant.

We now show that the above phenomenon is not accidental. Without losing generality, we assume that all distributions are strictly positive. Lemma 6 says that for infinitely many different initial states of agent A_2, its messages to A_0 and A_1, however, are identical.

Lemma 6 *Let s be a strictly positive initial state of Cmas3. There exists infinitely many distinct state s', identical to s in $P(a)$, $P(b|a)$ and $P(c|a)$ but is distinct in $P(d|b,c)$ such that the message $P_2(b|d=d_0)$ produced from s' is identical to that produced from s, and so is the message $P_2(c|d=d_0)$.*

Proof: We denote the message component $P_2(b = b_0|d = d_0)$ from state s by $P_2(b_0|d_0)$. We denote the message component from s' by $P'_2(b_0|d_0)$. $P_2(b_0|d_0)$

can be expanded as

$$P_2(b_0|d_0) = P_2(b_0,d_0)/(P_2(b_0,d_0)+P_2(b_1,d_0)) = [1+\frac{P_2(b_1,d_0)}{P_2(b_0,d_0)}]^{-1}$$
$$= [1+\frac{P_2(b_1,c_0,d_0)+P_2(b_1,c_1,d_0)}{P_2(b_0,c_0,d_0)+P_2(b_0,c_1,d_0)}]^{-1} = [1+\frac{P_2(d_0|b_1,c_0)P_2(b_1,c_0)+P_2(d_0|b_1,c_1)P_2(b_1,c_1)}{P_2(d_0|b_0,c_0)P_2(b_0,c_0)+P_2(d_0|b_0,c_1)P_2(b_0,c_1)}]^{-1}.$$

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

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

By assumption, $P_0(a,b) = P'_0(a,b)$, $P_1(a,c) = P'_1(a,c)$ and $P_2(b,c) = P'_2(b,c)$ but $P_2(d|b,c) \neq P'_2(d|b,c)$. If agent A_2 at s' can generate the identical messages $P'_2(b|d_0) = P_2(b|d_0)$ and $P'_2(c|d_0) = P_2(c|d_0)$ (conclusion of the lemma), then $P'_2(d|b,c)$ must be the solutions of the following equations:

$$\frac{P'_2(d_0|b_1,c_0)P_2(b_1,c_0)+P'_2(d_0|b_1,c_1)P_2(b_1,c_1)}{P'_2(d_0|b_0,c_0)P_2(b_0,c_0)+P'_2(d_0|b_0,c_1)P_2(b_0,c_1)} = \frac{P_2(b_1,d_0)}{P_2(b_0,d_0)}$$
$$\frac{P'_2(d_0|b_0,c_1)P_2(b_0,c_1)+P'_2(d_0|b_1,c_1)P_2(b_1,c_1)}{P'_2(d_0|b_0,c_0)P_2(b_0,c_0)+P'_2(d_0|b_1,c_0)P_2(b_1,c_0)} = \frac{P_2(c_1,d_0)}{P_2(c_0,d_0)}$$

Since $P'_2(d|b,c)$ has four independent parameters but is constrained by only two equations, it has *infinitely* many solutions. Each solution defines an initial state s' of Cmas3 that satisfies all conditions in the lemma. □

Lemma 7 says that with the same difference in initial states, a coherent inference will produce distinct results from Cmas3.

Lemma 7 *Let P and P' be strictly positive probability distributions over the DAG of fig. 2 (d) such that they are identical in $P(a)$, $P(b|a)$ and $P(c|a)$ but distinct in $P(d|b,c)$. Then $P(a|d=d_0)$ is distinct to $P'(a|d=d_0)$ in general.*

Proof: We have the following from P and P', respectively:

$$P(a|d_0) = \sum_{b,c} P(a|b,c)P(b,c|d_0) \quad (1)$$
$$P'(a|d_0) = \sum_{b,c} P(a|b,c)P'(b,c|d_0) \quad (2)$$

where we have used $P(a|b,c)$ since P' is identical with P in $P(a)$, $P(b|a)$ and $P(c|a)$. If $P(b,c|d_0) \neq P'(b,c|d_0)$ (which we show below), then in general

$P(a|d_0) \neq P'(a|d_0)$. We also have

$$P(b,c|d_0) = \frac{P(d_0|b,c)P(b,c)}{P(d_0)} = \frac{P(d_0|b,c)P(b,c)}{\sum_{b,c} P(d_0|b,c)P(b,c)},$$
$$P'(b,c|d_0) = \frac{P'(d_0|b,c)P(b,c)}{P'(d_0)} = \frac{P'(d_0|b,c)P(b,c)}{\sum_{b,c} P'(d_0|b,c)P(b,c)}.$$

Since $P(d|b,c) \neq P'(d|b,c)$, in general $P(b,c|d_0) \neq P'(b,c|d_0)$. □

We conclude with the following theorem:

Theorem 8 *Message passing in Cmas3 cannot be coherent in general, no matter how it is performed.*

Proof: By Lemma 6, $P_2(b|d = d_0)$ and $P_2(c|d = d_0)$ are insensitive to the initial states and hence the posteriors (e.g., $P_0(a|d = d_0)$) computed from the messages cannot be sensitive either. However, by Lemma 7, the posteriors should be different in general given different initial states. Hence, correct belief updating cannot be achieved in Cmas3. □

Note that the non-coherence of Cmas3 is due to its non-degenerate loop. From Eqs.(2) and (2), correct inference requires $P(b,c|d_0)$. To pass such a message, a separator must contain $\{b,c\}$, the intersection between U_2 and $U_0 \cup U_1$. The nondegenerate loop signifies the splitting of such a separator (into separators $\{b\}$ and $\{c\}$). The result is the passing of marginals of $P(b,c|d_0)$ (the insensitive messages) and ultimately the incorrect inference.

We can generalize this analysis to an arbitrary nondegenerate loop of length 3 (the loop length of Cmas3), where each of a, b, c, d is a set of variables. The result in Lemmas 6, 7 and Theorem 8 can be similarly derived.

We can further generalize this analysis to an arbitrary nondegenerate loop of length $K > 3$. By clumping $K-2$ adjacent subdomains into one big subdomain Q, the loop is reduced to length 3. Any message passing among the $k-2$ subdomains can be considered as occurring in the same way as before the clumping but "inside" Q. Now the above analysis for an arbitrary nondegenerate loop of length 3 applies. Corollary 9 summarizes the analysis.

Corollary 9 *Message passing in a nondegenerate loop cannot be coherent in general, no matter how it is performed.*

4.2 Degenerate loops

In a degenerate loop, all subdomains share the same separator and it is straightforward to pass the message coherently (we omit details for space limit). However, in practice a CG made of only degenerate loops are rare, and such loops can always be cut open with coherent message passing performed in the resultant tree. Under the assumption that nondegenerate loops are commonplace, we prefer a uniform organization for agents which support coherent message passing no matter what types of loops exist in the CG:

BC 3 *A uniform agent organization regarding loops is preferred.*

By Corollary 9, a tree must be used when non-degenerate loops exist. By BC 3, a tree will be preferred. We summarize in the following proposition which implies the third choice in Section 2, with the understanding that a loopy organization may be used as long as all loops involved are degenerate.

Proposition 10 *Let a CMADIS over $\mathcal{N}$ be one that observes BC 1 through BC 3. Then a tree organization of agents must be used.*

Proposition 10 admits many tree organizations. Jensen [4] showed that coherent message passing may not be achieved with just any tree. In particular, if two subdomains N_i and N_j share a subset I of variables but I is not contained in every subdomain on the path between them in the tree, then coherent message passing is not achievable. To ensure coherent message passing, the tree must be a *junction tree*, where for each pair of N_i and N_j, $N_i \cap N_j$ is contained in every subdomain on the path between N_i and N_j. Hence we have the following proposition:

Proposition 11 *Let a CMADIS over $\mathcal{N}$ be one that observes BC 1 through BC 3. Then a junction tree organization of agents must be used.*

5 On local covering condition

In this section, we show that the local covering condition in Def. 2 is necessary and sufficient to guarantee that the resultant hypertree is a junction tree. The proof is in Appendix.

Theorem 12 *Let N_0, ..., N_{n-1} be a set of subdomains. Start with an empty hypergraph, add each N_i recursively as a hypernode and connect it with an existing hypernode with a hyperlink. The resultant hypergraph is a junction tree iff each hypernode is added according to the local covering condition.*

From Theorem 12, the fourth choice of Section 2 follows.

6 On subdomain separators

Given our commitment to a (hyper) junction tree organization (Theorem 12), it follows that each separator must be chosen such that the message over it is *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 easily 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. Hence we have the following proposition:

Proposition 13 *Let a CMADIS over $\mathcal{N}$ be one that observes BC 1 through BC 3. Then each separator in a tree organization must render the two subtrees conditionally independent.*

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.

7 Choice on subdomain representation

Given a subdomain N_i, the number of parameters to represent the belief of A_i is exponential on $|N_i|$. Graphical models allow more compact representation. We focus on DAG models as they are the most concise with the understanding that other models such as decomposable Markov networks or chain graphs can also be used.

BC 4 *A DAG is used to structure individual agent's knowledge.*

A DAG model admits a causal interpretation of dependence. Once we adopt it for each agent, we must adopt it for the joint belief of all agents:

Proposition 14 *Let a CMADIS over $\mathcal{N}$ be constructed following BC 1, through BC 4. Then each subdomain N_i is structured as a DAG over N_i and the union of these DAGs is a connected DAG over $\mathcal{N}$.*

Proof: If the union of subdomain DAGs is not a DAG, then it has a directed cycle. This contradicts the causal interpretation of individual DAG models. The connectedness is implied by Proposition 4. $\square$

The fifth and sixth choices of Section 2 now follows.

8 On interface between subdomains

We show that the interface between subdomains must be structured as a d-sepset. This is established below through the concept of d-separation [13].

Proposition 15 *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: Sufficiency has been shown in [18].

[Necessity] 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. $\square$

Since d-separation captures all graphically identifiable conditional independencies [13], Proposition 15 implies that d-sepset is the necessary and sufficient *syntactic* condition for conditionally independent separators (Proposition 13) under *all possible subdomain structures and observation patterns.* We emphasize that d-sepset is necessary for the most general case, since by restricting subdomain structure (e.g., some agent contains only "cause" relative to other agents but no "effect") or observation pattern (e.g., some agent has no local observation and only relies on others' observation), the d-sepset requirement may be relaxed. The seventh choice of Section 2 now follows. From Propositions 14, 15 and Theorem 12, the following proposition is implied. The proof is in Appendix.

Proposition 16 *Let a CMADIS over* $\mathcal{N}$ *be constructed following BC 1 through BC 4. Then it must be structured as a hypertree MSDAG.*

9 On belief assignment

By Propositions 14, the structure of a CMADIS is a DAG (we emphasize that it is a consequence of BC 1 through BC 4, not an assumption). Hence a joint probability distribution (jpd) over $\mathcal{N}$ can be defined by specifying local distribution for each node and applying chain rule. In a CMADIS, a node can be internal to an agent or shared. Distribution for an internal node can be specified by the corresponding agent vender.

When a node is shared, it may have different parents in different agents (e.g., b in Fig. 1). Since each shared node is a d-sepnode, Def. 1 implies that for each shared variable x, there exists a subdomain containing all the parents of x in the universe as stated in the following lemma:

Lemma 17 *Let x be a d-sepnode in a hypertree MSDAG. Let the parents of x in D_i be $\pi_i(x)$. Then there exists D_k such that $\pi_k(x) = \bigcup_i \pi_i(x)$.*

If agents are built by the same vendor, then once $P(x|\pi_k(x))$ is specified for x, $P(x|\pi_i(x))$ for each i is implied. If agents are built by different vendors, then it is possible that distributions on a d-sepnode may be incompatible with each other. For instance, in fig. 1, A_0 and A_1 may differ on $P(a)$. We make the following basic commitment for integrating independently built agents into a CMADIS:

BC 5 *Within each agent's subdomain, jpd is consistent with the agent's belief. For shared nodes, jpd supplements each agent's knowledge with others'.*

The key issue is to combine agents' belief on a shared variable to arrive at a common belief. One idea [14] is to interpret the distribution from each agent as obtained from a sample data. The combined $P(x|\pi(x))$ can then be obtained from the combined data sample. In summary, let agents combine their belief for each shared x. Then, for each shared x, let jpd be consistent with $P(x|\pi_k(x))$, and for each internal x, let jpd be consistent with $P(x|\pi(x))$ held by the corresponding agent. It's easy to see that the resultant jpd is precisely the one defined in Def. 3, stated in the following proposition:

Proposition 18 *Let a CMADIS over $\mathcal{N}$ be constructed following BC 1 through BC 5. Then the jpd over $\mathcal{N}$ is identical to that of Def. 3.*

The last choice of Section 2 now follows. Pooling Propositions 16 and 18 together, the MSBN representation is entailed by the BCs:

Theorem 19 *Let a CMADIS over $\mathcal{N}$ be constructed following BC 1 through BC 5. Then it must be represented as a MSBN or some equivalent.*

10 Conclusion

From the following *basic* commitments: [BC 1] exact probabilistic measure of belief, [BC 2] communication by belief over small sets of shared variables, [BC 3] uniform organization of agents regarding loops, [BC 4] DAG for domain structuring, [BC 5] joint belief admitting agents' belief on internal variables and combining their belief on shared ones, we have shown that the resultant representation of a CMADIS is a MSBN or some equivalent.

This result aids comparison with related frameworks. Multiagent inference frameworks based on default reasoning (e.g., DATMS [10] and DTMS [3]) do not admit BC 1, nor does the *blackboard* [12]. Several frameworks for decomposition of probabilistic knowledge has been proposed. *Abstract network* [8] replaces fragments of a centralized BN by *abstract arcs* to improve inference efficiency. *Similarity network* and *Bayesian multinet* [2] represent asymmetric independence where each subnet shares almost all variables with each other subnet. A *nested junction trees* [6] can exploit independence induced by incoming messages to a cluster and it shares all its variables with the nesting cluster. They were not intended for multiagent systems and do not admit BC 2. MSBNs are unique in satisfying both BC 1 and BC 2 in one.

This analysis addresses concerns on restrictions imposed by MSBN. In particular, the two key technical restrictions, hypertree and d-sepset interface, are the consequence of BC 1 and BC 2.

One useful consequence of BC 2 and MSBN is that the internal knowledge of each agent is never transmitted and can remain private. This aids construction of CMADISs by agents from independent vendors. Multiagent systems commonly stand in two extreme: self-interested versus cooperative. MSBN stands in the middle: agents are cooperative and truthful to each other while the internal know-how is protected.

Our analysis provides guidance to extension/relaxations of MSBNs. Less fundamental restrictions can be relaxed, e.g., BC 4 such that other graphical models can be used. BC 3 requires degenerate loops be handled in the same way as nondegenerate loops. If loopy organization of agents are indeed needed, the analysis shows that it is okay as long as loops are degenerate. If subdomain structures and observation patterns are less than general, then the d-sepset restriction can be relaxed.

Acknowledgements

This work is supported by the Research Grant OGP0155425 from the Natural Sciences and Engineering Research Council (NSERC) of Canada, by the National Science Foundation (NSF) under Grant No. IIS-9812755, and by the Defense Advanced Research Projects Agency (DARPA) and Air Force Research Laboratory, Air Force Materiel Command, USAF, under agreement number F30602-99-2-0525. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon.

The views and conclusions contained herein are those of the author(s) and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the National Science Foundation. Defense Advanced Research Projects Agency (DARPA), the Air Force Research Laboratory, or the U.S. Government.

References

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

[2] D. Geiger and D. Heckerman. Knowledge representation and inference in similarity networks and Bayesian multinets. *Artificial Intelligence*, 82:45–72, 1996.

[3] M.N. Huhns and D.M. Bridgeland. Multiagent truth maintenance. *IEEE Trans. Sys., Man, and Cybernetics*, 21(6):1437–1445, 1991.

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

[5] 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.

[6] U. Kjaerulff. Nested junction trees. In *Proc. 13th Conf. on Uncertainty in Artificial Intelligence*, pages 294–301, Providence, Rhode Island, 1997.

[7] D. Koller and A. Pfeffer. Object-oriented Bayesian networks. In D. Geiger and P.P. Shenoy, editors, *Proc. 13th Conf. on Uncertainty in Artificial Intelligence*, pages 302–313, Providence, Rhode Island, 1997.

[8] W. Lam. Abstraction in Bayesian belief networks and automatic discovery from past inference sessions. In *Proc. of AAAI*, pages 257–262, 1994.

[9] 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.

[10] C.L. Mason and R.R. Johnson. DATMS: a framework for distributed assumption based reasoning. In L. Gasser and M.N. Huhns, editors, *Distributed Artificial Intelligence II*, pages 293–317. Pitman, 1989.

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

[12] H.P. Nii. Blackboard systems: the blackboard model of problem solving and the evolution of blackboard architectures. *AI Magazine*, (2):38–53, 1986.

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

[14] D. Poole, A. Mackworth, and R. Goebel. *Computational Intelligence: A Logical Approach*. Oxford University Press, 1998.

[15] 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.

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

[17] Y. Xiang and H. Geng. Distributed monitoring and diagnosis with multiply sectioned Bayesian networks. In *Proc. AAAI Spring symposium on AI in Equipment Service, Maintenance and Support*, 1999.

[18] 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.

Appendix: Proofs

Proof for Theorem 12

[Sufficiency] Clearly, the resultant hypergraph is a tree since each new node is connected to the existing graph by a single link. We show that the tree is also a junction tree.

Suppose the tree resultant from following the local covering condition is not a junction tree. Then there exist a pair of subdomains N_i and N_k, and a third subdomain N_j on the path between them such that $N_i \cap N_k \not\subset N_j$. Suppose N_i is added before N_k. From the procedure for adding nodes, N_j must be added after N_i but before N_k.

Suppose $m \geq 0$ subdomains are on the path between N_j and N_k with the order $N_j, N_{x_1}, N_{x_2}, ..., N_{x_m}, N_k$. If $N_i \cap N_k \not\subset N_{x_m}$, then N_k was not added according to the local covering condition: a contradiction. If $N_i \cap N_k \subset N_{x_m}$, we consider the addition of N_{x_m}. If $N_i \cap N_k \not\subset N_{x_{m-1}}$, then N_{x_m} was not added according to the local covering condition: a contradiction. Otherwise, we consider the addition of $N_{x_{m-1}}$.

Repeating this process for at most m times, we either find a contradiction, or finally end up considering the addition of N_{x_1} which satisfies $N_i \cap N_k \subset N_{x_1}$. Since $N_i \cap N_k \not\subset N_j$, N_{x_1} was not added according to the local covering condition: a contradiction.

[Necessity] Using the above notation, if we do not following the local covering condition in adding nodes, we could add N_i first, followed by N_j connected to N_i, followed by N_k connected to N_j. The resultant tree will not be a junction tree. $\square$

Proof of Proposition 16

From BC 1 through BC 4, it follows that the universe should be structured as a connected DAG (Proposition 14) such that each subdomain is structured as a subDAG. The DAGs should be organized into a hypertree according to the local covering condition (Theorem 12). The interface between individual DAGs should be a d-sepset (Proposition 15). Hence the CMADIS should be structured as a hypertree MSDAG (Def. 2). $\square$