

Searching Model Structures Based on Marginal Model Structures

Sung-Ho Kim and Sangjin Lee

*Department of Mathematical Sciences
Korea Advanced Institute of Science and Technology
Daejeon, 305-701, South Korea*

1. Introduction

Graphs are used effectively in representing model structures in a variety of research fields such as statistics, artificial intelligence, data mining, biological science, medicine, decision science, educational science, etc. We use different forms of graphs according to the nature of the random variables involved. For instance, arrows are used when the relationship is asymmetric as when it is causal or temporal, and undirected edges are used when the relationship is associative.

When a random field is Markov with respect to a triangulated graph, i.e., a decomposable graph, which does not have a cycle of length 4 or larger, its corresponding probability model is expressed in a factorized form which facilitates computation over the probability distribution of the random field (Kemeny et al., 1976). This computational feasibility, among others, makes such a Markov random field a most favored random field. Literature is abound in regard to the properties of theMarkov randomfield which isMarkov with respect to a decomposable graph (see Chapter 12 of Whittaker (1990) and Lauritzen (1996)). We call such a random field a decomposable graphical model.

There have been remarkable improvements in learning graphical models in the form of a Bayesian network (Pearl, 1986 & 1988; Heckerman et al., 1995; Friedman & Goldszmidt, 1998; Neil et al., 1999; Neapolitan, 2004) from data. This learning however is mainly instrumented by heuristic searching algorithms and the model searching is usually NP-hard [Chickering (1996)]. A good review is given in Cooper (1999) and Neopolitan (2004) on structural discovery of Bayesian or causal networks from data. Since a Bayesian network can be transformed into a decomposable graph [Lauritzen and Spiegelhalter (1988)], the method of model combination which is proposed in this paper would lead to an improvement in graphical modelling from data. This method would be useful when we don't have data which are large enough for the number of the random variables that are involved in the data. In this situation, it is desirable to develop marginal models of manageable sizes for subsets of variables and then search for a model for the whole set of variables based on the marginal models.

The main idea of the method to be proposed is similar to constraint-based learning as described in Neapolitan (2004) (also see Meek (1995) and Spirtes et al. (2000)) where we

construct a Bayesian network based on a list of constraints which are given in terms of conditional independence among a given set of random variables. But a noteworthy difference between the two is that, while the statements of conditional independencies are an extraction, as for the constraintbased learning, from the probability model of the whole set of the variables involved, the statements of conditional independencies for the method to be proposed are from the marginal probability models of the subsets of variables. This difference in how we extract the statements of conditional independence is the main source of the difference between the two methods.

In deriving the method of the paper, it is imperative that we make use of the relationship between the joint (as against marginal) model structure and its marginal model structure. Kim (2006) introduced a certain type of subgraph, called Markovian subgraph, and investigated its properties as a subgraph of a decomposable graph. Some of the properties play a crucial role in the process of constructing a decomposable graph based on a collection of its Markovian subgraphs. We will elaborate on this in later sections. Kim (2004) called our attention to the relationship between a set of probability models and a set of model structures and proved a theorem to the effect that we may deal with model structures of marginal models in search of the model structure of the joint probability model for the whole set of variables involved in data. In 1 this respect, we will use graphs to represent model structures and compare the joint model with its marginal models using graphs.

This paper consists of 8 sections. Section 2 introduces notations and graphical terminologies along with new concepts such as Markovian subgraph and Markovian subpath. A simple but motivational example is considered in Section 3 with some prelusive remarks of the method to be proposed. Sections 4 and 5 then introduce theorems and a new type of graph that are instrumental for the model-combination. Section 6 describes the model-combining process and it is illustrated in section 7. The paper is concluded in section 8 with summarizing remarks.

2. Notation and preliminaries

We will consider only undirected graphs in the paper. We denote a graph by $\mathcal{G} = (V, E)$, where V is the set of the indexes of the variables involved in \mathcal{G} and E is a collection of ordered pairs, each pair representing that the nodes of the pair are connected by an edge. Since \mathcal{G} is undirected, that (u, v) is in E is the same as that (v, u) is in E . If $(u, v) \in E$, we say that u is a neighbor node of or adjacent to v or vice versa. We say that a set of nodes of \mathcal{G} forms a complete subgraph of \mathcal{G} if every pair of nodes in the set is adjacent to each other. If every node in A is adjacent to all the nodes in B , we will say that A is adjacent to B . A maximal complete subgraph is called a clique of \mathcal{G} , where the maximality is in the sense of set-inclusion. We denote by $C(\mathcal{G})$ the set of cliques of \mathcal{G} .

A path of length n is a sequence of nodes $u = v_0, \dots, v_n = v$ such that $(v_i, v_{i+1}) \in E, i = 0, 1, \dots, n - 1$ and $u \neq v$. If $u = v$, the path is called an n -cycle. If $u \neq v$ and u and v are connected by a path, we write $u \rightleftharpoons v$. We define the connectivity component of u as

$$[u] = \{v \in V; v \rightleftharpoons u\} \cup \{u\}.$$

So, we have

$$v \in [u] \iff u \Rightarrow v \iff u \in [v].$$

We say that a path, $v_1, \dots, v_n, v_1 \neq v_n$, is intersected by A if $A \cap \{v_1, \dots, v_n\} \neq \emptyset$ and neither of the end nodes of the path is in A . We say that nodes u and v are separated by A if all the paths from u and v are intersected by A . In the same context, we say that, for three disjoint sets A, B , and C , A is separated from B by C if all the paths from A to B are intersected by C and write $\langle A|C|B \rangle_{\mathcal{G}}$. A non-empty set B is said to be intersected by A if B is partitioned into three sets B_1, B_2 , and $B \cap A$ and B_1 and B_2 are separated by A in \mathcal{G} . The complement of a set A is denoted by A^c and the cardinality of a set A by $|A|$.

For $A \subset V$, we define an *induced subgraph* of \mathcal{G} confined to A as $\mathcal{G}_A^{ind} = (A, E \cap (A \times A))$. We also define a graph, called a *Markovian subgraph* of \mathcal{G} confined to A , which is formed from \mathcal{G}_A^{ind} by completing the boundaries in \mathcal{G} of the connectivity components of the complement of A and denote it by \mathcal{G}_A . In other words, $\mathcal{G}_A = (A, E_A)$ where

$$E_A = (E \cap A \times A) \cup \{(u, v) \in A \times A; u \text{ and } v \text{ are not separated by } A \setminus \{u, v\} \text{ in } \mathcal{G}\}.$$

Let a path, π say, from u to v is a sequence of edges (u_i, u_{i+1}) with $u_0 = u$ and $u_k = v$. Then we will say that a sequence of edges $(u_{i_1}, u_{i_2}), \dots, (u_{i_r}, u_{i_{r+1}})$, $0 \leq i_1 < i_2 < \dots < i_{r+1} \leq k$, is a Markovian subpath of π .

If $\mathcal{G} = (V, E)$, $\mathcal{G}' = (V, E')$, and $E' \subseteq E$, then we say that \mathcal{G}' is an edge-subgraph of \mathcal{G} and write $\mathcal{G}' \subseteq^e \mathcal{G}$. A subgraph of \mathcal{G} is either a Markovian subgraph, an induced subgraph, or an edge-subgraph of \mathcal{G} . If \mathcal{G}' is a subgraph of \mathcal{G} , we call \mathcal{G} a supergraph of \mathcal{G}' .

Although decomposable graphs are well known in literature, we define them here for completeness.

Definition 2.1. A triple (A, B, C) of disjoint, nonempty subsets of V is said to form a decomposition of \mathcal{G} if $V = A \cup B \cup C$ and the two conditions below both hold:

- (i) A and B are separated by C ;
- (ii) \mathcal{G}_C^{ind} is complete.

By recursively applying the notion of graph decomposition, we can define a decomposable graph.

Definition 2.2. \mathcal{G} is said to be decomposable if it is complete, or if there exists a decomposition (A, B, C) into decomposable subgraphs $\mathcal{G}_{A \cup C}^{ind}$ and $\mathcal{G}_{B \cup C}^{ind}$.

For a decomposable graph, we can find a sequence of cliques C_1, \dots, C_k of \mathcal{G} which satisfies the following condition

[see Proposition 2.17 of Lauritzen (1996)]: with $C_{(j)} = \bigcup_{i=1}^j C_i$ and $S_j = C_j \cap C_{(j-1)} \neq \emptyset$, for all $i > 1$, there is a $j < i$ such that $S_i \subseteq C_j$

By this condition for a sequence of cliques, we can see that S_j is expressed as an intersection of neighboring cliques of \mathcal{G} . If we denote the collection of these S_j 's by $x(\mathcal{G})$, we have, for a decomposable graph \mathcal{G} , that

$$\chi(\mathcal{G}) = \{a \cap b; \quad a, b \in \mathcal{C}(\mathcal{G}), \quad a \neq b\} \quad (1)$$

It is possible for some decomposable graph \mathcal{G} that there are sets, a and b , in $x(\mathcal{G})$ such that $a \subset b$.

The cliques are elementary graphical components and the S_j is obtained as intersection of neighboring cliques. So, we will call the S_j 's prime separators (PSs for short) of the decomposable graph \mathcal{G} . The PSs in a decomposable graph may be extended to separators of prime graphs in any undirected graph, where the prime graphs are defined as the maximal subgraphs without a complete separator in Cox and Wermuth (1999).

3. Simple example with remarks

Graph \mathcal{G} can be represented in the same way as a graphical log-linear model is represented in terms of generators [Fienberg (1980)]. If \mathcal{G} consists of cliques C_1, \dots, C_r , we will write

$$\mathcal{G} = [C_1] \cdots [C_r].$$

For instance, if \mathcal{G} is of five nodes and $C_1 = \{1, 2\}$, $C_2 = \{2, 3\}$, $C_3 = \{3, 4, 5\}$, then $\mathcal{G} = [12][23][345]$. In this context, the terms *graph* and *model structure* are used in the same sense.

Suppose that we are given a pair of simple graphical models where one model is of random variables X_1, X_2, X_3 with their inter-relationship that X_1 is independent of X_3 conditional on X_2 and the other is of X_1, X_2, X_4 with their inter-relationship that X_1 is independent of X_4 conditional on X_2 . From this pair, we can imagine a model structure for the four variables X_1, \dots, X_4 . The two inter-relationships are pictured at the left end of Figure 1. The graph at the top of the two at the left is represented by [12][23] and the one at the bottom by [12][24]. X_1 and X_2 are shared in both models, and assuming that none of the four variables are marginally independent of the others, we can see that the following joint models have the marginals, [12][23] and [12][24]:

$$[12][24][23], \quad [12][24][34], \quad [12][23][34], \quad [12][234], \quad (2)$$

which are displayed in graph in Figure 1. Note that the first three of these four models are submodels or edge-subgraphs of the last one.

It is important to note that some variable(s) are independent of the others, conditional on X_2 in the pair of marginals, and in all the models in (2). That conditional independence takes place conditional on the same variable in the marginal models and also in the joint models underlies the main theme of the method to be proposed in the paper.

In addressing the issue of combining graphical model structures, we can not help using independence graphs and related theories to derive desired results with more clarity and refinement. The conditional independence embedded in a distribution can be expressed to some level of satisfaction by a graph in the form of graph-separateness [see, for example, the separation theorem in p. 67, Whittaker (1990)]. We instrument the notion of conditional independence with some particular sets of random variables in a model, where the sets form a basis of the model structure so that the Markov property among the variables of the model may be preserved between the joint model and its marginals. The sets are

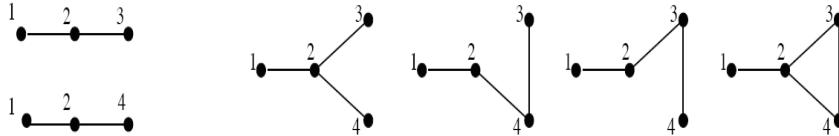


Fig. 1. Two marginal models on the left and the four joint models on the right

prime separators. In the simple example, X_2 forms the basis. Without the variable, X_2 , the conditional independence disappears.

It is shown that if we are given a graphical model with its independence graph, \mathcal{G} , and some of its marginal models, then under the decomposability assumption of the model we can find a graph, say \mathcal{H} , which is not smaller than \mathcal{G} and in which the graph-separateness in the given marginal models is preserved (Theorem 4.3). This graph-separateness is substantiated by the prime separators which are found in the graphs of the marginal models. In combining marginal models into \mathcal{H} , we see to it that these prime separators appear as the only prime separators in \mathcal{H} . This is reflected in the model-combining procedure described in Section 6.

4. Theorems useful for model-combination

Let $\mathcal{G} = (V, E)$ be the graph of a decomposable model and let V_1, V_2, \dots, V_m be subsets of V . The m Markovian subgraphs, $\mathcal{G}_{V_1}, \mathcal{G}_{V_2}, \dots, \mathcal{G}_{V_m}$, may be regarded as the structures of m marginal models of the decomposable model, \mathcal{G} . For simplicity, we write $\mathcal{G}_i = \mathcal{G}_{V_i}$.

Definition 4.1. Suppose there are m Markovian subgraphs, $\mathcal{G}_1, \dots, \mathcal{G}_m$. Then we say that graph \mathcal{H} of a set of variables V is a combined model structure (CMS) corresponding to $\mathcal{G}_1, \dots, \mathcal{G}_m$, if the following conditions hold:

- (ii) $\bigcup_{i=1}^m V_i = V$.
- (ii) $\mathcal{H}_{V_i} = \mathcal{G}_i$ for $i = 1, \dots, m$. That is, \mathcal{G}_i are Markovian subgraphs of \mathcal{H} .

We will call \mathcal{H} a maximal CMS corresponding to $\mathcal{G}_1, \dots, \mathcal{G}_m$ if adding any edge to \mathcal{H} invalidates condition (ii) for at least one $i = 1, \dots, m$. Since \mathcal{H} depends on $\mathcal{G}_1, \dots, \mathcal{G}_m$ we denote the collection of the maximal CMSs by $\Omega(\mathcal{G}_1, \dots, \mathcal{G}_m)$.

According to this definition, a CMS is a Markovian supergraph of each $\mathcal{G}_i, i = 1, \dots, m$. There may be many CMSs that are obtained from a collection of Markovian subgraphs as we saw in (2).

In the theorem below, $\mathcal{C}_{\mathcal{G}}(A)$ is the collection of the cliques which include nodes of A in the graph \mathcal{G} . The proof is intuitive. The symbol, $\langle \cdot | \cdot \cdot | \cdot \rangle$, follows Pearl (1988), and for three disjoint sets, A, B , and C , $\langle A|C|B \rangle_{\mathcal{G}}$ means that A is separated from B by C in \mathcal{G} .

Theorem 4.2. Let $\mathcal{G}' = (V', E')$ be a Markovian subgraph of \mathcal{G} and suppose that, for three disjoint subsets A, B, C of V , $\langle A|B|C \rangle_{\mathcal{G}'}$. Then

- (i) $\langle A|B|C \rangle_{\mathcal{G}}$;
- (ii) For $W \in \mathcal{C}_{\mathcal{G}}(A)$ and $W' \in \mathcal{C}_{\mathcal{G}}(C)$, $\langle W|B|W' \rangle_{\mathcal{G}}$.

Proof. Since

$$\langle A|B|C \rangle_{\mathcal{G}'}, \quad (3)$$

there is no path in \mathcal{G}' between A and C that bypasses B . If (i) does not hold, it is obvious that (3) does not hold either. Now suppose that result (ii) does not hold. Then there must be a path from a node in A to a node in C bypassing B . This implies negation of the condition (3) by the definition of the Markovian subgraph. Therefore, result (ii) must hold.

Recall that if \mathcal{G}_i , $i = 1, 2, \dots, m$, are Markovian subgraphs of \mathcal{G} , then \mathcal{G} is a CMS. For a given set S of Markovian subgraphs, there may be many maximal CMSs, and they are related with S through PSs as in the theorem below.

Theorem 4.3. Let there be Markovian subgraphs \mathcal{G}_i , $i = 1, 2, \dots, m$, of a decomposable graph \mathcal{G} . Then

$$(i) \quad \cup_{i=1}^m \chi(\mathcal{G}_i) \subseteq \chi(\mathcal{G}),$$

$$(ii) \text{ for any maximal CMS } \mathcal{H}, \quad \cup_{i=1}^m \chi(\mathcal{G}_i) = \chi(\mathcal{H}).$$

Proof. See Kim (2006).

For a given set of Markovian subgraphs, we can readily obtain the set of PSs under the decomposability assumption. By (1), we can find $\chi(\mathcal{G})$ for any decomposable graph \mathcal{G} simply by taking all the intersections of the cliques of the graph. An apparent feature of a maximal CMS in contrast to a CMS is stated in Theorem 4.3. Note that, in this theorem, \mathcal{G} is a CMS of \mathcal{G}_i , $i = 1, 2, \dots, m$.

Another important merit of a PS is that if a set of nodes is a PS in a Markovian subgraph, then it is not intersected in any other Markovian subgraphs.

Theorem 4.4. Let \mathcal{G} be a decomposable graph and \mathcal{G}_1 and \mathcal{G}_2 be Markovian subgraphs of \mathcal{G} . Suppose that a set $C \in \chi(\mathcal{G}_1)$ and that $C \subseteq V_2$. Then C is not intersected in \mathcal{G}_2 by any other subset of V_2 .

Proof. Suppose that there are two nodes u and v in C that are separated in \mathcal{G}_2 by a set S . Then, by Theorem 4.2, we have $\langle u|S|v \rangle_{\mathcal{G}}$. Since $C \in \chi(\mathcal{G}_1)$ and \mathcal{G}_1 is decomposable, C is an intersection of some neighboring cliques of \mathcal{G}_1 by equation (1). So, S can not be a subset of V_1 but a proper subset of S can be. This means that there are at least one pair of nodes, v_1 and v_2 , in \mathcal{G}_1 such that all the paths between the two nodes are intersected by C in \mathcal{G}_1 , with v_1 appearing in one of the neighboring cliques and v_2 in another.

Since v_1 and v_2 are in neighboring cliques, each node in C is on a path from v_1 to v_2 in \mathcal{G}_1 . From $\langle u|S|v \rangle_{\mathcal{G}}$, it follows that there is an l -cycle ($l \geq 4$) that passes through the nodes u , v , v_1 , and v_2 in \mathcal{G} . This contradicts the assumption that \mathcal{G} is decomposable. Therefore, there can not be such a separator S in \mathcal{G}_2 .

Among the above three theorems, Theorem 4.3 plays a key role in the method of model-combination and the other two are employed in adding and removing edges during the combining process.

5. Graph of prime separators

In this section, we will introduce a graph of PSs which consists of PSs and edges connecting them. The graph is the same as the undirected graphs that are considered so far in this paper, the nodes being replaced with PSs. Given a decomposable graph \mathcal{G} , the graph of the PSs of \mathcal{G} is defined as follows:

Let $A = \cup_{a \in \chi(\mathcal{G})} a$. Then the graph of the prime separators (GOPs for short) of \mathcal{G} is obtained from \mathcal{G}_A by replacing every PS and all the edges between every pair of neighboring PSs in \mathcal{G}_A with a node and an edge, respectively.

For example, there are three PSs, $\{3, 4\}$, $\{3, 5\}$, and $\{4, 8\}$, in graph \mathcal{G}_1 in Figure 8. Then none of the PSs is conditionally independent of any other among the three PSs. We represent this phenomenon with the graph at the top-left corner in Figure 9, where the GOPs's are the graphs of the line (as against dotted) ovals only. The xGOPs's (short for "expanded GOPs") as appearing in the figure are defined in Section 6 and used in model combining.

We can see conditional independence among the PSs, $\{13, 14\}$, $\{10, 13\}$, $\{10, 19\}$, and $\{10, 21\}$, in graph \mathcal{G}_3 in Figure 8. This conditional independence is depicted in $\text{GOP}_{\mathcal{G}_3}$ in Figure 9. As connoted in $\text{GOP}_{\mathcal{G}_1}$ in Figure 9, a GOP may contain a clique of more than 2 PSs, but it cannot contain a cycle of length 4 or larger if the PSs are from a decomposable graph.

Let \mathcal{G}' be a Markovian subgraph of \mathcal{G} and suppose that, for three PSs, A, B , and C , of \mathcal{G}' , $A \setminus C$ and $B \setminus C$ are separated by C in \mathcal{G}' . Then, by Theorem 4.2, the same is true in \mathcal{G} .

For three sets, A, B , and C , of PSs of a graph \mathcal{G} , if A and B are separated by C , then we have that

$$(\cup_{a \in A} a) \cap (\cup_{b \in B} b) \subseteq (\cup_{c \in C} c) \quad (4)$$

When A, B, and C are all singletons of PSs, the set-inclusion is expressed as

$$A \cap B \subseteq C \quad (5)$$

This is analogous to the set-inclusion relationship among cliques in a junction tree of a decomposable graph (Lauritzen (1996)). A junction tree is a tree-like graph of cliques and intersection of them, where the intersection of neighboring cliques lies on the path which connects the neighboring cliques. As for a junction tree, the sets in (5) are either cliques or intersection of cliques. In the context of a junction tree, the property as expressed in (5) is called the junction property. We will call the property expressed in (4) *PS junction* property, where 'PS' is from 'prime separator.'

The GOPS and the junction tree are different in the following two senses: First, the basic elements are PSs in the GOPS while they are cliques in the junction tree; secondly, the GOPS is an undirected graph of PSs while the junction tree is a tree-like graph of cliques. Some PSs may form a clique in an undirected graph as in graphs \mathcal{G}_1 and \mathcal{G}_4 in Figure 8. This is why GOPS may not necessarily be tree-like graphs. So, two PSs may be separated by a set of PSs. But, since all the PSs in a decomposable graph \mathcal{G} are obtained from the intersections of neighboring cliques in \mathcal{G} , the GOPS of \mathcal{G} is the same as the junction tree of \mathcal{G} with the clique-nodes removed from the junction tree. Whether \mathcal{G} is decomposable or not, expression (4) holds in general.

6. Description of model-combining procedure

We will call a node a *PS node* if it is contained in a PS, and a *non-PS node* otherwise. Theorem 4.4 implies that if, for a given Markovian subgraph \mathcal{G}' , s is the set of the PSs each of which is a neighbor to a PS node v in \mathcal{G}' , then s will also be the set of the neighboring PSs of any PS, say a , such that $v \in a$, in the Markovian subgraph which is obtained by adding the PS, a , to \mathcal{G}' . This is useful in locating PSs for model-combination since PS nodes of a PS always form a complete subgraph.

Other useful nodes in model-combination are the non-PS nodes that are shared by multiple Markovian subgraphs. A simple illustration of the usefulness is given in expression (2). The Markovian subgraphs in Figure 1 share node 1, which determines the meeting points of the subgraphs when they are combined into the maximal CMS, [12][234]. Whether they are PS nodes or not, a set of nodes which are shared by a pair of Markovian subgraphs become meeting points of the subgraphs in the combining process. The shared nodes restrict the possible locations of the PS nodes that are not shared by both of the subgraphs. We will call by xGOPS a GOPS which is expanded with the nodes that are shared with other subgraphs. However we will not distinguish the two and use the terminology "GOPS" when confusion is not likely.

A rule of thumb of model-combination is that we connect two nodes each from different Markovian subgraphs in a given set, say \mathcal{M} , of Markovian subgraphs if the two nodes are not separated by any other nodes in \mathcal{M} . We will formally describe this condition below:

[Separateness condition] Let \mathcal{M} be a set of Markovian subgraphs of \mathcal{G} and \mathcal{H} a maximal CMS of \mathcal{M} . If two nodes are in a graph in \mathcal{M} and they are not adjacent in the graph, then neither are they in \mathcal{H} . Otherwise, adjacency of the nodes in \mathcal{H} is determined by checking separateness of the nodes in \mathcal{M} .

Suppose that \mathcal{M} consists of m Markovian subgraphs, $\mathcal{G}_1, \dots, \mathcal{G}_m$, of \mathcal{G} and we denote by a^i a PS of \mathcal{G}_i . We can then combine the models of \mathcal{M} as follows.

Step 1. We arrange the subgraphs into $\mathcal{G}_{i_1}, \dots, \mathcal{G}_{i_m}$ such that $|V_{i_j} \cap V_{i_{j+1}}| \geq |V_{i_{j+1}} \cap V_{i_{j+2}}|$ for $j = 1, 2, \dots, m-2$. For convenience, let $i_j = j$, $j = 1, 2, \dots, m$. We set $\eta_1 = \{\mathcal{G}_1\}$.

Step 2a. We first put an edge between every pair of PSs, a^1 and a^2 , if

$$a^1 \cap a^2 \neq \emptyset,$$

in such a way that the separateness condition is satisfied with regard to \mathcal{M} . We denote the resulting GOPS by H .

Step 2b. Once the node-sharing PSs are all considered in Step 2a, we need to consider all the PSs a^1 and a^2 such that

$$a^1 \cap (\cup_{a \in \chi(\mathcal{G}_2)} a) = \emptyset \text{ and } a^2 \cap (\cup_{a \in \chi(\mathcal{G}_1)} a) = \emptyset \quad (6)$$

and put edges between a^i , $i = 1, 2$, and every PS in \mathcal{G}_{3-i} that is acceptable under the separateness condition, in addition to the GOPS which is obtained in Step 2a. For example, for each a^1 satisfying (6), we add edges to \mathcal{H} between the a^1 and every possible PS in \mathcal{G}_2 under the separateness condition, and similarly for each of a^2 that satisfy (6). We denote the result of the combination by η_2 .

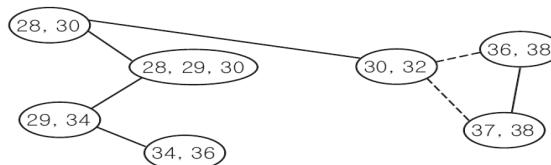


Fig. 2. A graphic display of part of Step 2a corresponding to that the PS of GOPS_5 , $\{28, 30\}$, and the PS of GOPS_6 , $\{30, 32\}$, share node 30 and that $\{28, 30\}$ is adjacent to $\{29, 31, 32, 34\}$ and separated from $\{35, 36, 37, 38\}$ by $\{29, 31, 32, 34\}$. The non-adjacent connectedness is expressed by dashed lines.

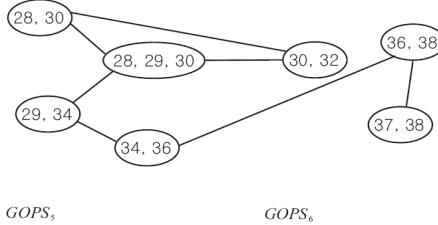


Fig. 3. Step 2a in progress from Figure 2 as for the PS pairs, {28, 29, 30} and {30, 32} and {34, 36} and {36, 38}.

Step 3. Let η_i be the GOPS obtained from the preceding step. Note that η_i can be a set of GOPS's. For each GOPS \mathcal{H} in η_i , we combine \mathcal{H} with \mathcal{G}_{i+1} as in Step 2, where we replace \mathcal{G}_1 and \mathcal{G}_2 with \mathcal{H} and \mathcal{G}_{i+1} , respectively. We repeat this combination with \mathcal{G}_{i+1} for all the graphs \mathcal{H} in η_i , which results in the set, η_{i+1} , of newly combined graphs.

Step 4. If $i + 1 = m$, then stop the process. Otherwise, repeat Step 3.

We will call this process Markovian combination of model structures or MCMoSt for short. The process is summarized in flowcharts in Figures 5 and 6; the former is of the main body of the process and the latter is of checking for the separateness condition. For a brief illustration of the MCMoSt, we will consider the two marginal graphs, \mathcal{G}_5 and \mathcal{G}_6 in Figure 8. This example has only two graphs, so we may skip Step 1.

Figure 9 shows the GOPS_s of two marginal graphs \mathcal{G}_5 and \mathcal{G}_6 . As for \mathcal{G}_5 , the set of PSs in GOPS_s is $\{\{28, 30\}, \{28, 29, 30\}, \{29, 34\}, \{34, 36\}\}$ and it is $\{\{30, 32\}, \{36, 38\}, \{37, 38\}\}$ for \mathcal{G}_6 . The PS of GOPS_s, $\{28, 30\}$, and the PS of GOPS_e, $\{30, 32\}$, share node 30. So we put an edge between the two PS's. In \mathcal{G}_5 , $\{28, 30\}$ is adjacent to $\{29, 31, 32, 34\}$ and is separated from $\{35, 36, 37, 38\}$ by $\{29, 31, 32, 34\}$. This separateness must be preserved, by Theorem 4.2, in the combined model of \mathcal{G}_5 and \mathcal{G}_6 . We represent this non-adjacent connectedness by dashed lines in Figure 2.

The other PSs that share nodes between \mathcal{G}_5 and \mathcal{G}_6 are the pair of $\{28, 29, 30\}$ and $\{30, 32\}$ and the pair of $\{34, 36\}$ and $\{36, 38\}$. We put edges between the PSs in each of these pairs and then check the separateness condition. In \mathcal{G}_5 , $\{37, 38\}$ is separated from $\{28, 29, 30\}$ by $\{31, 32, 34, 35, 36\}$, which is satisfied in the graph in Figure 3. This is the result of Step 2a.

In Step 2b, we can see that the PS, $\{37, 38\}$, of \mathcal{G}_6 is disjoint with all the PS's of \mathcal{G}_5 . In \mathcal{G}_5 , we see that $\{34, 36\}$ separates $\{37, 38\}$ from the remaining six nodes in \mathcal{G}_5 . Thus we put an edge between $\{34, 36\}$ and $\{37, 38\}$ only. This ends up with the combined GOPS in Figure 4.

In combining a pair of graphs, \mathcal{G}_1 and \mathcal{G}_2 say, suppose that an edge is added between a PS, a^1 , in \mathcal{G}_1 and another PS, a^2 , in \mathcal{G}_2 and let N_i , $i = 1, 2$, be the set of the PSs which are adjacent to a^i in \mathcal{G}_i . Then, under the decomposability assumption and the separateness condition, further edge-additions are possible between the PSs in the $(\{a^i\} \cup N_i)$'s only.

An example of this is given in Section 7.

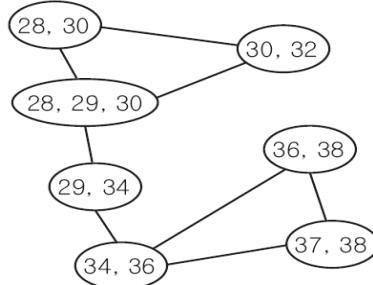


Fig. 4. Step 2b as continued from Figure 3.

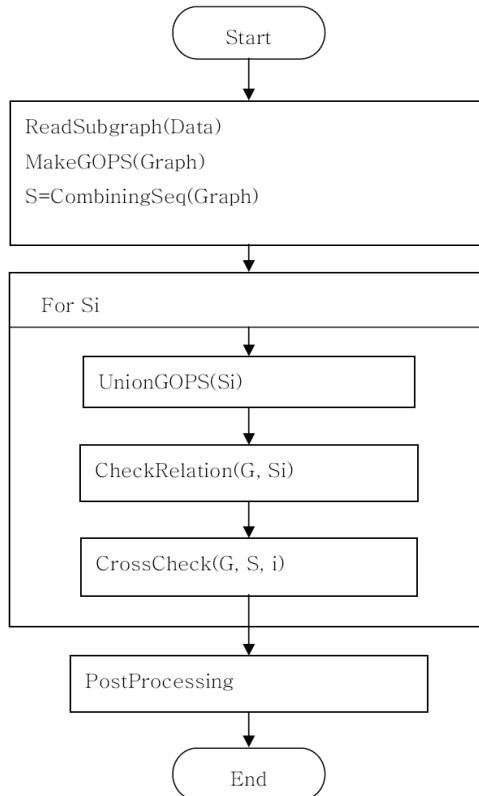


Fig. 5. A flowchart of the model-combining process, MCMoSt. In this chart, S is a sequence of marginal models to be combined; UnionGOPS just puts the two graphs to be combined together; CheckRelation checks if the separateness condition is satisfied between nodes and/or PSs; CrossCheck checks if the combined graph preserves the PSs of the two graphs.

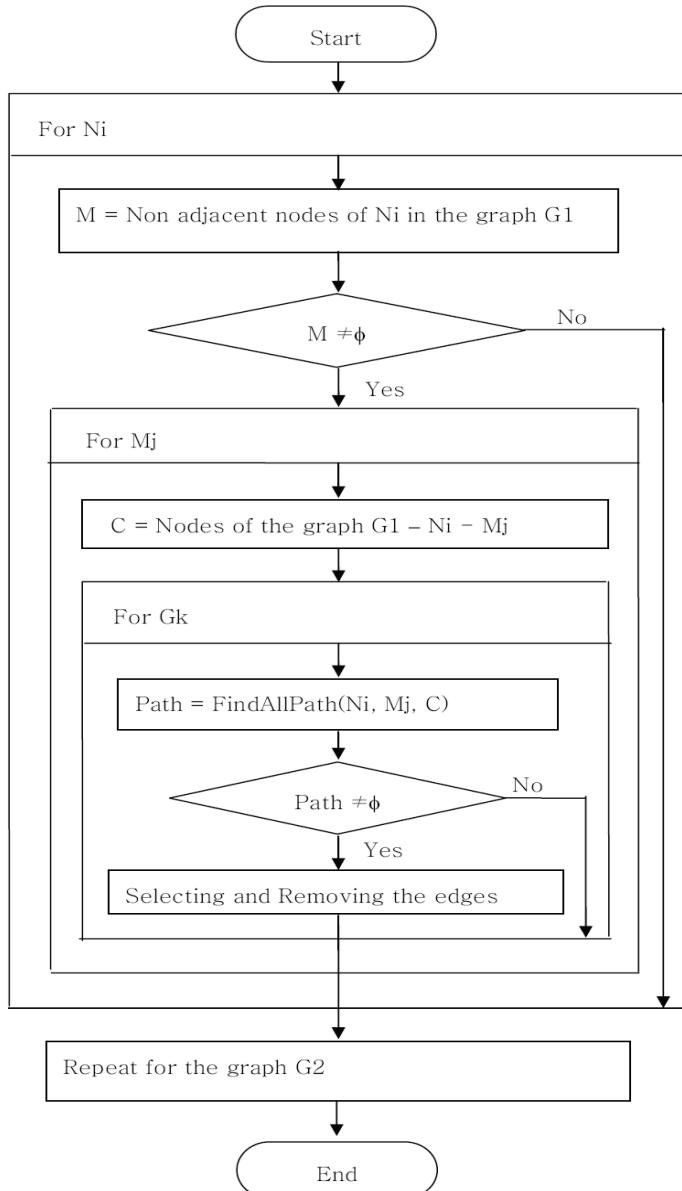
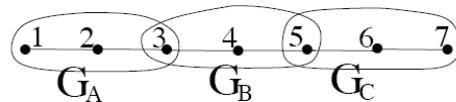


Fig. 6. A flowchart of the process *CheckRelation* which is a main part of MCMoSt. In this chart, we assume combining two graphs, \mathcal{G}_1 and \mathcal{G}_2 say. $FindAllPath(A, B, C)$ finds paths between A and B that are blocked by C ; *Selecting and Removing the edges* means that, for each of the paths which are found in $FindAllPath$, the edges to be removed are selected and removed.

6.1 Time complexity of the procedure

Let $\mathcal{V} = \{V_1, V_2, \dots, V_m\}$. For a given set of \mathcal{G}_A 's, $A \in \mathcal{V}$, we denote by $E^s(\mathcal{V})$ the set of the pairs, u and v , for which there is at least one \mathcal{G}_A such that $\{u, v\} \subseteq A$ and they are not adjacent in \mathcal{G}_A , denote by $E^a(\mathcal{V})$ the set of the pairs, u and v , for which there is at least one \mathcal{G}_A such that $\{u, v\} \subseteq A$ and they are adjacent in \mathcal{G}_A , and let $E^{rem}(\mathcal{V}) = \{\{u, v\} \subseteq V; u \neq v\} \setminus (E^s(\mathcal{V}) \cup E^a(\mathcal{V}))$. For example, in the graph below, $V = \{1, 2, \dots, 7\}$, $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{5, 6, 7\}$, $\mathcal{V} = \{A, B, C\}$, $E^s(\mathcal{V}) = \{\{1, 3\}, \{3, 5\}, \{5, 7\}\}$, $E^a(\mathcal{V}) = \{\{i, i+1\}, i = 1, 2, \dots, 6\}$ and $E^{rem}(\mathcal{V}) = \{\{i, j\}, 1 \leq i < j \leq 7\} \setminus (E^s(\mathcal{V}) \cup E^a(\mathcal{V}))$.



The computing time of MCMoSt depends upon the sizes of the sets such as E^a and E^{rem} of the graphs in \mathcal{M} . A main part of the algorithm is designed for searching for all the possible edges between nodes under the condition that the pairs of nodes in E^s are separated. We use the depth-first search method (Tarjan, 1972) in Step 2 of the combination process to check the separateness between nodes. Suppose we combine \mathcal{G}_1 and \mathcal{G}_2 into a graph \mathcal{H} and obtain E^s , E^a and E^{rem} from \mathcal{G}_1 and \mathcal{G}_2 . Then we search for all the possible edges between nodes in such a way that, if there is a path, π' , in \mathcal{G}_1 or \mathcal{G}_2 which contains u and v on itself and there is a path, π , in \mathcal{H} which also contains u and v on itself, then π' is a Markovian subpath of π .

For two graphs, \mathcal{G}_1 and \mathcal{G}_2 , let $|V_i| = n_i$ with $i = 1, 2$, $|V_1 \cap V_2| = n_{12}$ and $\tilde{n}_i = n_i - n_{12}$. It is well known that the time complexity of the depth-first search method for a graph $\mathcal{G} = (V, E)$ is of order $O(|V| + |E|)$. So the time complexity for the combination is of order $\tilde{n}_{21} \tilde{n}_1^2 O(\tilde{n}_2 + \tilde{e}_2) + \tilde{n}_2^2 O(\tilde{n}_1 + \tilde{e}_1)$ where \tilde{e}_i is the number of edges in the induced subgraph of \mathcal{G}_i on $V_i \setminus V_{3-i}$. As a matter of fact, when we use GOPS's instead of graphs of nodes, the time complexity reduces by a considerable amount. For instance, we can see in Figure 9 that the six GOPS's are composed of 3, 3, 5, 5, 6, 3 PS's, respectively, while the marginal graphs are of ten nodes each. MCMoSt uses PS's and the nodes that are shared between graphs rather than nodes only.

7. Illustration

In this section, we suppose that we are given six marginal models as in Figure 8 each of which is Markovian subgraphs of the graph in Figure 7. As a matter of fact the six marginal models were obtained through a statistical analysis. We first generated data from the model in Figure 7 assuming that all the 40 variables are binary. We then chose six subsets of variables in such a way that the variables are more highly associated within subsets than between them. The six marginal models in Figure 8 were obtained through a statistical analysis of contingency table data. A detailed description of this is given in Kim (2005).

Since our interest is in the model-combining method, we will refrain from any further discussion on this statistical issue.

For notational convenience, we will denote a PS by $c(\cdot)$. For instance, $c(1, 2)$ denotes a PS consisting of nodes 1 and 2. From \mathcal{G}_1 , we have $\chi(\mathcal{G}_1) = \{c(3, 4), c(4, 8), c(3, 5)\}$. If we regard the three PSs as random variables, these PSs are associated. In the same context, we can represent the conditional independence relationship among the PSs via an independence graph based on the corresponding marginal models \mathcal{G}_i . The GOPS's are displayed for each marginal model in Figure 9 along with the nodes which are shared among the marginal models. We will call a variable whose corresponding node is a PS-node a PS variable and similarly for a non-PS variable. Since every non-PS variable is

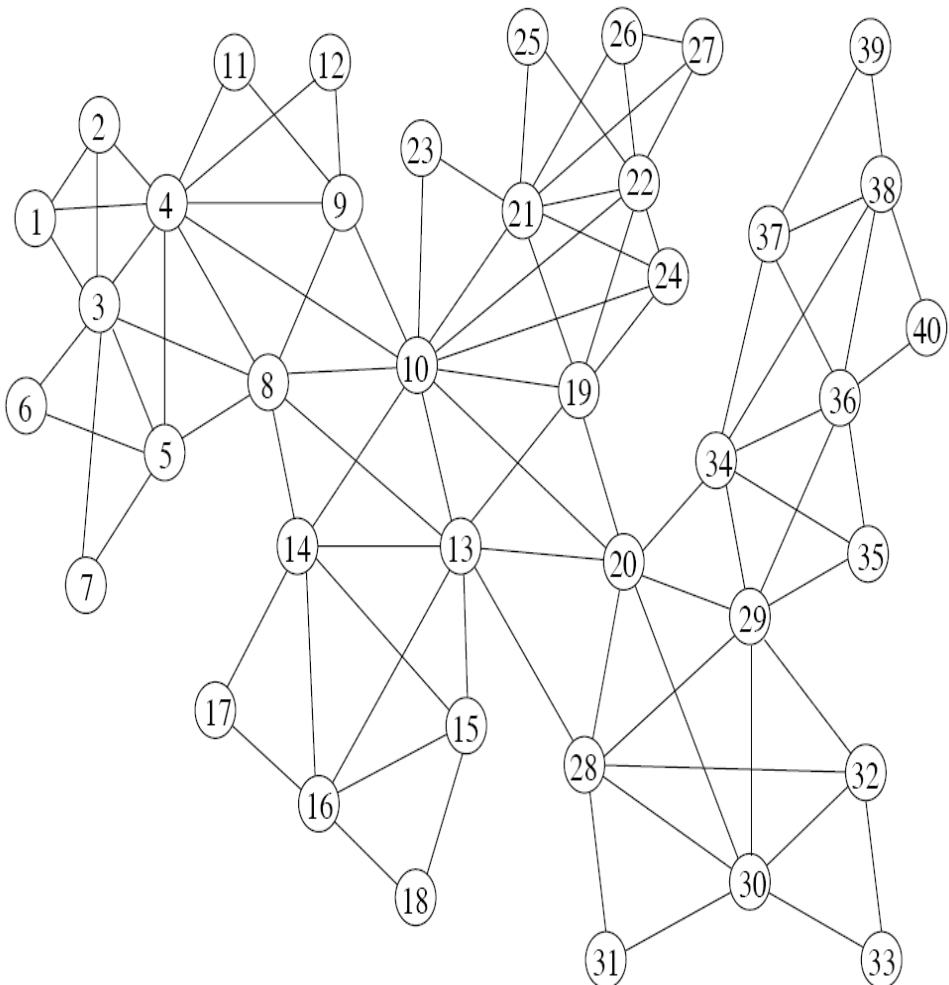


Fig. 7. A model of the 40 variables that are used for illustration

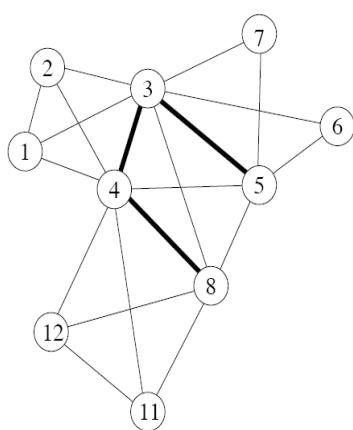
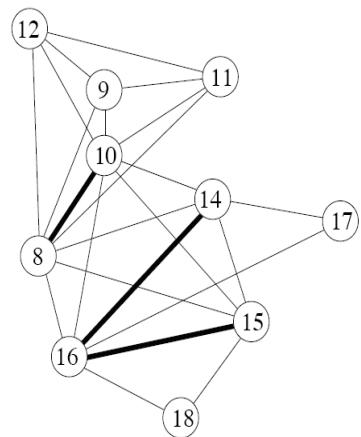
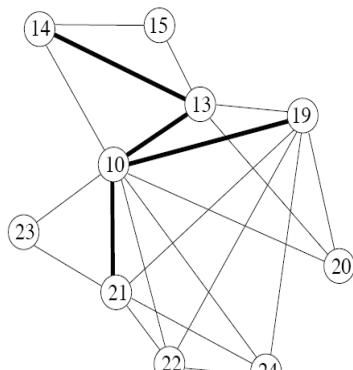
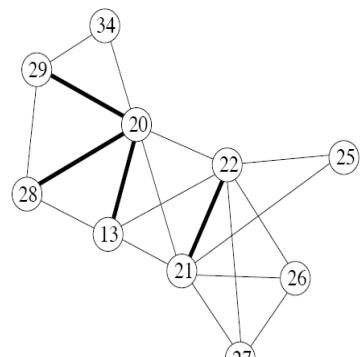
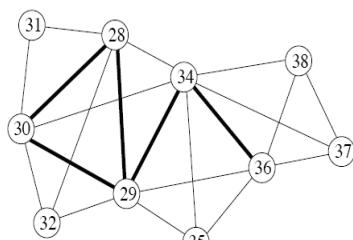
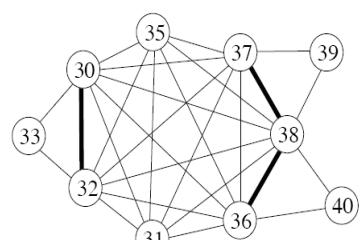
 \mathcal{G}_1  \mathcal{G}_2  \mathcal{G}_3  \mathcal{G}_4  \mathcal{G}_5  \mathcal{G}_6

Fig. 8. Six marginal models of the model in Figure 7. PSs are represented by thick lines. See Figure 9 for the PSs of the six marginal models.

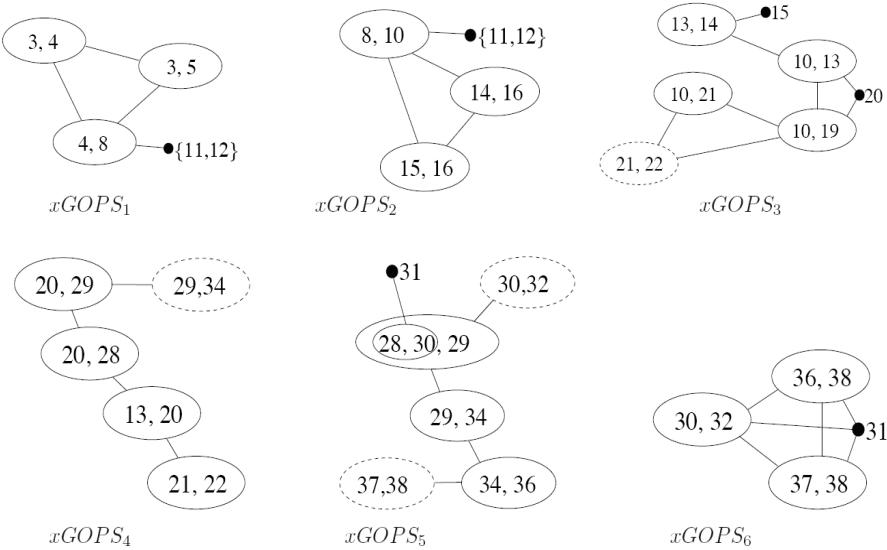


Fig. 9. The GOPS's and x GOPS's of the six marginal models in Figure 8. GOPSi's are the graphs of the line ovals only. x GOPS_i is the independence graph of the PSs of \mathcal{G}_i and the nodes which are shared by \mathcal{G}_i with other marginal models. The oval nested in another oval in x GOPS₅ means that the PS, $c(28, 30)$, is a subset of the PS, $c(28, 29, 30)$. $c(28, 30)$ is a neighbor of node 31 in the graph. Dotted ovals mean that the corresponding set of nodes is a PS in some other marginal models.

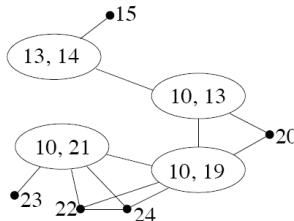


Fig. 10. The graph obtained by linking non-PS variables (bullets) to the PS's of \mathcal{G}_3 in Figure 8.

Separated from other variables by its neighbor PSs, we can represent \mathcal{G}_i by linking each non-PS variable to its neighbor PSs. For example, the graph in Figure 10, which is obtained by adding non-PS variables of \mathcal{G}_3 to the graph, $GOPS_3$, of the PSs of \mathcal{G}_3 . Since every non-PS node has a unique set of neighboring PSs, a graph such as that in Figure 10 is determined uniquely.

According to Theorem 4.3 (i), all the PSs that appear in a marginal model of \mathcal{G} are found in $\chi(\mathcal{G})$. This means that we must make sure that all the PSs of the marginal models appear as PSs in \mathcal{G} . This fact is instrumental to constructing the independence graph of the PSs of the marginal models.

In section 3, we considered a simple problem of model combination. Although the example is very simple, we can see therein that the shared variables, X_1 and X_2 , are like road signs in constructing a model structure of the variables that are involved in the variable-sharing marginal models. As more variables are shared between a pair of marginal models, the possible locations of each variable are more limited and thus model construction for the variables that are involved in either of the two marginal models becomes easier. The variable-sharing between V_i and V_{i+1} , $i = 1, 2, \dots, 5$, is as follows:

$$|V_1 \cap V_2| = 3, |V_2 \cap V_3| = 3, |V_3 \cap V_4| = 4, |V_4 \cap V_5| = 3, |V_5 \cap V_6| = 7.$$

$|V_i \cap V_j| = 0$ when $|i - j| > 1$. So, it is desirable that we begin combining marginal models from the pair of \mathcal{G}_5 and \mathcal{G}_6 .

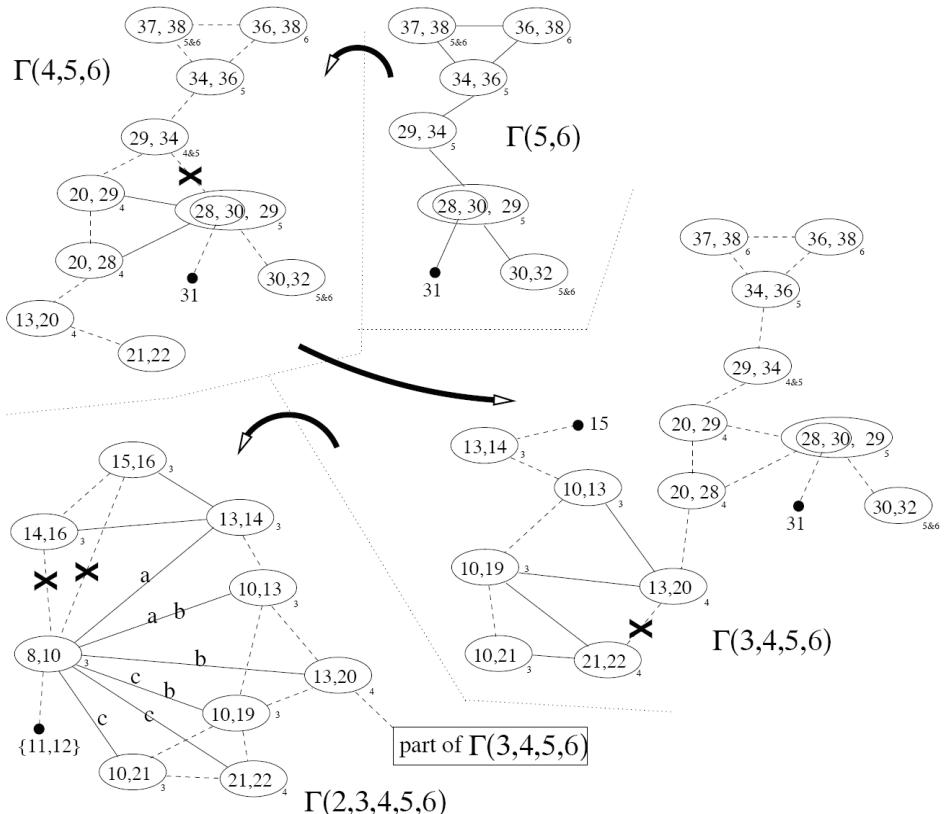


Fig. 11. A model-combining process of marginal models \mathcal{G}_i . $\Gamma(A)$ denotes a maximal independence graph of the PSs of \mathcal{G}_i , $i \in A$. The small numbers at the bottom-right of the ovals are the marginal model labels to which the corresponding PSs belong. $\Gamma(2, 3, 4, 5, 6)$ is of three xGOPS's which are determined by three different groups of edges that are labelled by a, b, and c at the bottom-left corner. Line edges are used when they are newly added; dotted edges for existing edges; and X-marked dotted edges for the existing edges to be removed.

and then keep combining marginal models in the order of $\mathcal{G}_4, \mathcal{G}_3, \mathcal{G}_2, \mathcal{G}_1$.

Figure 11 shows part of the model-combining process of the xGOPS's. The graph in the top-right is the result of combining \mathcal{G}_5 and \mathcal{G}_6 , which is the same as the graph in Figure 3.

For a given collection of the GOPSs of $\mathcal{G}_i, i \in A$, there can be more than one maximal independence graph of the PSs, and we will denote the set of maximal independence graphs by $\Gamma(A)$. When $|\Gamma(A)| = 1$, the independence graph itself will be represented by $\Gamma(A)$.

For notational convenience, we will use the symbol \oplus to denote an outcome of model-combination. For example, we write $\mathcal{G} \oplus \mathcal{G}'$ to express the outcome of combining the two graphs, \mathcal{G} and \mathcal{G}' . Since $\Gamma(5, 6)$ is a single graph, we may express the next combination by $\Gamma(5, 6) \oplus x\text{GOPS}_4$. By applying the same method for the combination, we also obtain a unique graph as in the top-left corner of Figure 11, and denote it by $\Gamma(4, 5, 6)$. Note that the PS, $c(29, 34)$, is shared by both of $\Gamma(5, 6)$ and $x\text{GOPS}_4$. Since $c(20, 28)$ and $c(20, 29)$ share nodes with $c(28, 29, 30)$, we put edges between each of the former two PSs and the last PS. This edge addition conflicts with the separateness of $c(29, 24)$ from $c(20, 28)$. Considering that node 28 is shared by $c(20, 28)$ and $c(28, 29, 30)$ only (since $c(28, 30)$ is contained in $c(28, 29, 30)$, it is ignored here), we can see that the edge between $c(29, 34)$ and $c(28, 29, 30)$ must be deleted. Note that the three PSs, $c(29, 34)$, $c(20, 29)$, and $c(28, 29, 30)$ are on a path and satisfy the PS junction property (4).

$\Gamma(4, 5, 6) \oplus x\text{GOPS}_3 = \Gamma(3, 4, 5, 6)$ is also a single graph as in Figure 11. Since neither of $c(10, 21) \cap c(21, 22)$ and $c(10, 13) \cap c(13, 20)$ is empty, line edges are put between the node-sharing PSs. Then the separateness condition is violated since $c(10, 21)$ and $c(10, 13)$ are no longer separated by $c(10, 19)$, and the three PSs, $c(10, 13)$, $c(10, 19)$, and $c(10, 21)$, share node 10. Thus the edge between $c(21, 22)$ and $c(13, 20)$ is deleted. After this, we check if there are any PSs pertaining to Stage 2 of section 6, which end up with the addition of edges between $c(10, 19)$ and $c(21, 22)$ and between $c(10, 19)$ and $c(13, 20)$.

So far the combination process has produced single graphs. But $\Gamma(3, 4, 5, 6) \oplus x\text{GOPS}_2$ is a set of three graphs. $c(14, 16)$ and $c(13, 14)$ share node 14, and so we put a line edge between them. $c(8, 10)$ shares node 10 with

$$c(10, 13), c(10, 19), \text{ and } c(10, 21), \quad (7)$$

so we can put two line edges, between $c(8, 10)$ and each of the first two of the PSs in (7), and another set of two line edges, between $c(8, 10)$ and each of the last two of the PSs in (7). Note that any of these edge additions violates the separateness of $c(13, 14)$ from $c(10, 19)$ and $c(10, 21)$. So the edges between $c(8, 10)$ and $c(14, 16)$ and between $c(8, 10)$ and $c(15, 16)$ are deleted. Implementing Stage 2 then ends up with the addition of edges between $c(15, 16)$ and $c(13, 14)$ and possibly between $c(8, 10)$ and $c(13, 14)$, where the latter edge can be added along with the edge between $c(8, 10)$ and $c(10, 13)$ out of the three edges between $c(8, 10)$ and each of the PSs in (7). In other words, $c(8, 10)$ can form a clique with $c(10, 13)$ and $c(13, 14)$, with $c(10, 13)$, $c(13, 20)$, and $c(10, 19)$, or with $c(10, 19)$, $c(10, 21)$, and $c(21, 22)$ in an xGOPS in $\Gamma(3, 4, 5, 6) \oplus x\text{GOPS}_2$. This is depicted at the bottom-left corner of Figure 11.

However, the set $c(8, 10) \cup c(13, 14) \cup c(10, 13) \cup c(10, 19) \cup c(10, 21) = \{8, 10, 13, 14, 19, 21\}$ shares node 8 with V_1 , nodes 8, 10, 14 with V_2 , nodes 10, 13, 14 with V_3 , and nodes

13, 21 with V_4 . So it is more likely that $c(8, 10)$ and $c(13, 14)$ belong to the same clique. This is because the grouping of the variables was made so that the variables are more highly associated with each other within the subsets than between them. Based on this observation, we chose the xGOPS in which $c(8, 10)$ forms a clique with $c(10, 13)$ and $c(13, 14)$. We denote this graph by $\Gamma^a(2, 3, 4, 5, 6)$.

We can apply the same argument in combining $\Gamma^a(2, 3, 4, 5, 6)$ with \mathcal{G}_1 , which ends up with the GOPS in Figure 12. The bullets in the figure represent the non-PS variables, which are connected to their neighboring PSs. Note that those neighboring PSs are classified as such mostly due to the non-PS variables. For example, $\{37, 38\}$ is a PS separating node 39 from $V_6 \setminus \{37, 38, 39\}$ in \mathcal{G}_6 .

A PS is itself a complete subgraph and so is a clique of PSs. So we can easily transform the graph in Figure 12 into the undirected graph in Figure 13. This is the maximal CMS of the six marginal models as listed in Figure 8. The model in Figure 7 is fully recovered in the maximal CMS except the 5 thick edges appearing in Figure 13. These additional edges were created because X_4 were missing in \mathcal{G}_2 . If X_4 had been added to \mathcal{G}_2 , then $X_{[4,9]}$ would have separated X_{11}, X_{12} , and $X_{[8,10]}$ from each other, making those additional edges unnecessary. This phenomenon of additional edges leads

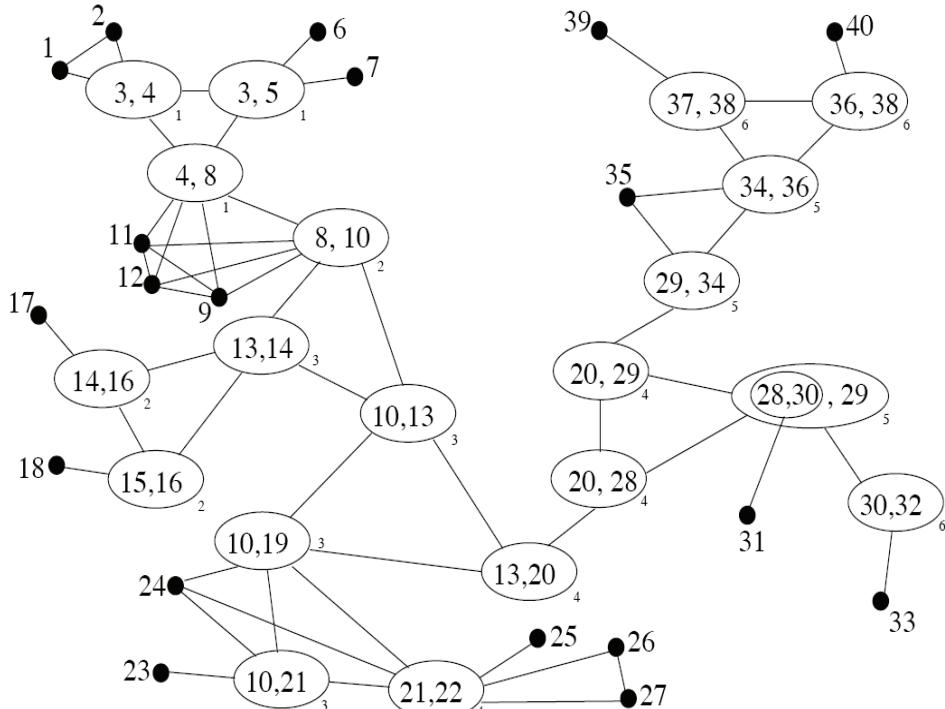


Fig. 12. An independence graph of PSs and non-PS variables. The PSs are in ovals and the dots are for the non-PS variables, and the small numbers at the bottom-right of the ovals are the marginal model labels of which the ovals are PSs.

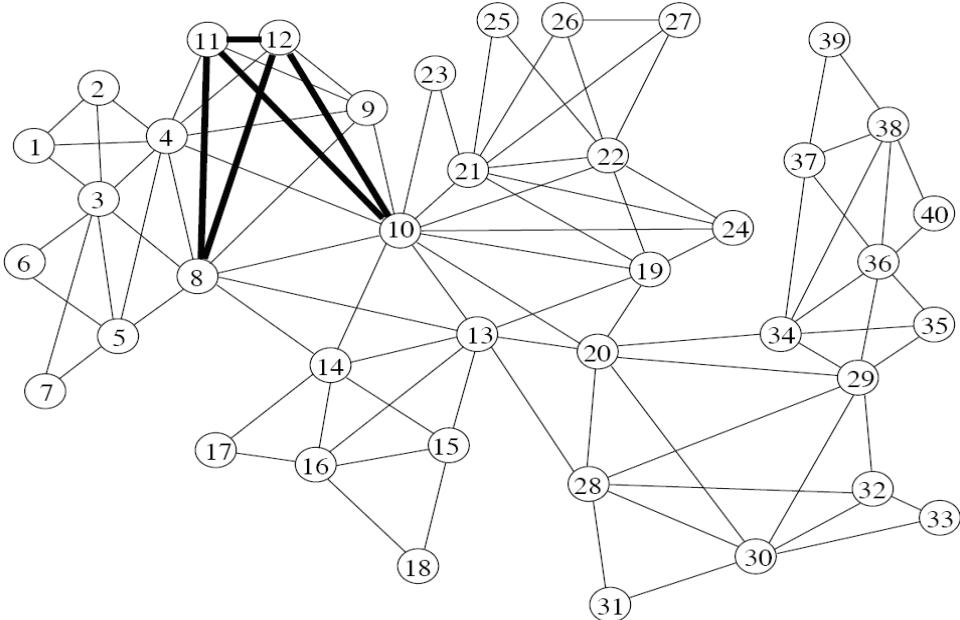


Fig. 13. The combined model structure which is obtained from the independence graph in Figure 12. The thick edges are additional to the model in Figure 7. us to recommend that the variables be grouped into marginal models so that the association between variables is higher within a marginal model than between marginal models.

8. Conclusion

In Section 7, we considered a model, \mathcal{G} , of 40 variables and six marginal models of it. The marginal models have their model structures given in decomposable graphs which are actually Markovian subgraphs of the graph \mathcal{G} . In this context, marginal model and Markovian subgraph may be regarded as synonyms as long as the joint model has a model structure which can be represented via an undirected graph.

In combining marginal models, it is important to make use of the locations of the variables that are shared by the marginal models to be combined. While we use GOPS's of marginal models to construct another GOPS, the locations of the non-PS nodes that are shared by the marginal models to be combined are as important as the PSs in the marginal models. The PS junction property (4) and the separateness condition are instrumental for locating PSs in model-combination. When $|\mathcal{G}_i \oplus \mathcal{G}_j| > 1$, that is, a multiple number of maximal CMS's are obtained, it is desirable that we look for or develop more marginal models from data in order to minimize the resultant maximal CMS's. For instance, the PS {8, 10} of \mathcal{G}_2 in Figure 8 can be connected to the combined graph $\Gamma(3, 4, 5, 6)$ by three different sets of edges as shown at the bottom-left corner of Figure 11. Denote the three different sets of corresponding nodes by A, B, and C. Then if we could develop a model for the set of

variables, $\{8, 10\} \cup A \cup B \cup C$, then the model would help us in choosing one of the three different sets of edges.

In selecting the subsets of variables, it is important that the variables that are highly associated belong to the same set. In other words, when the joint model is graphical with an undirected graph as its model structure, variables that appear as neighbors in the graph are desired to belong to a subset of variables. Otherwise, the model-combination may end up with an unnecessarily large graph. An example is demonstrated by thick edges in Figure 13. The thick edges would not have appeared, if X_9 had been included in \mathcal{G}_1 or X_4 in \mathcal{G}_2 . In this regard, subset selection for marginal modeling is crucial for a successful model-combination.

As mentioned in Section 1, several heuristic searching methods are developed for learning Bayesian networks from data. Since they deal with the whole set of variables involved in data, we can easily run into a sparse data problem for large scale modeling, not to mention the time complexity burden. The marginal-model based approach as proposed in this paper may not suffer from the sparse data problem. Furthermore, in applying our method, the marginal models don't have to be based on observed data only. They may be based on expert opinions, since the method deals with model structures only.

Although the model combination is carried out under the decomposability assumption, we can deal with the marginal models of a graphical model, which are not decomposable, by transforming their model structures into decomposable (i.e., triangulated) graphs. The combined model will then be larger than expected as a trade-off of the graph triangulation made on the marginal models.

9. References

- Chickering, D. (1996). Learning Bayesian networks is NP-complete, In: *Learning from Data*, D. Fisher and H. Lenz (Ed.), 121-130, Springer-Verlag.
- Cooper, G.F.(1999). An overview of the representation and discovery of causal relationships using Bayesian networks, In: *Computation, Causation, and Discovery*, C. Glymour and G.F. Cooper (Ed.), 3-62, The MIT Press, Cambridge, MA.
- Cox, D.R. and Wermuth, N. (1999). Likelihood factorizations for mixed discrete and continuous variables, *Scand. J. Statist.*, 26, 209-220.
- Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data*, The MIT Press, Cambridge, MA.
- Friedman, N. and Goldszmidt, M. (1998). Learning Bayesian networks with local structure, In: *Learning in Graphical Models*, M.I. Jordan, (Ed.), 421-460, Kluwer, Dordrecht, Netherlands.
- Heckerman, D., Geiger, D. and Chickering, D.M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, 20, 197-243.
- Kemeny, J.G., Speed, T.P., Knapp, A.W.(1976). Denumerable Markov Chains, 2nd edition, Springer, New York, NY.
- Kim, S.-H. (2004). Combining decomposable model-structures, Research Report 04-15, Division of Applied Mathematics, KAIST, Daejeon, S. Korea.
- Kim, S.-H. (2005). Log-linear modelling for contingency tables by using marginal model structures, *Research Report* 05-12, Division of Applied Mathematics, KAIST, Daejeon, S. Korea.

- Kim, S.-H. (2006). Properties of Markovian subgraphs of a decomposable graph, *Lecture Notes in AI*, 4293, 15-26.
- Lauritzen, S.L. (1996). *Graphical Models*. Oxford: Oxford University Press.
- Lauritzen, S.L. and Spiegelhalter, D.J. (1988). Local computations with probabilities on graphical structures and their application to expert systems, *J. R. Statist. Soc. B*, 50, 2, 157-224.
- Meek, C. (1995). Causal influence and causal explanation with background knowledge, In: *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference*, P. Besnard and S. Hanks (Ed.), 403-410, Morgan Kaufmann, San Mateo, CA.
- Neapolitan, R.E. (2004). *Learning Bayesian Networks*, Pearson Prentice Hall, Upper Saddle River, NJ.
- Neil, J.R., Wallace, C.S. and Korb, K.B. (1999). Learning Bayesian networks with restricted causal interactions, In: *Uncertainty in Artificial Intelligence*, K.B. Laskey and H. Prade, (Ed.), 486-493, Morgan Kaufmann Publishers, San Francisco, CA.
- Pearl, J. (1986). Fusion, propagation and structuring in belief networks, *Artificial Intelligence*, 29, 241-288.
- Pearl, J. (1988). *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, CA.
- Spirites, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA: MIT Press.
- Tarjan, R. E. (1972). Depth-first search and linear graph algorithms. *SIAM J. Comput.* 1(2), 146-160.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, Wiley, New York, NY.