Markovian combination of decomposable model structures

by

Sung-Ho Kim and Sangjin Lee

BK21 Research Report

09 - 12

June 22, 2009

DEPARTMENT OF MATHEMATICAL SCIENCES

The second stage of BK21 Fostering A World Class Talent

KAIST 한국과학기술원 Korea Advanced Institute of Science and Technology

Markovian combination of decomposable model structures

Sung-Ho Kim and Sangjin Lee
Korea Advanced Institute of Science and Technology, Daejeon, 305-701, South Korea [1]

Abstract

Suppose that we are interested in modeling for a random vector $\mathbf{X}$ and that we are given a set of graphical decomposable models, $\mathcal{G}_1, \cdots, \mathcal{G}_m$, for subvectors of $\mathbf{X}$ each of which share some variables with at least one of the other models. Under the assumption that the model of $\mathbf{X}$ is graphical and decomposable, we propose an approach of searching for model structures of $\mathbf{X}$ based on the given decomposable graphical models. A main idea in this approach is that we combine $\mathcal{G}_1, \cdots, \mathcal{G}_m$ using graphs of prime separators (section 2). When the true graphical model for the whole data is decomposable, prime separators in a marginal model are also prime separators in a maximal combined model of the marginal models. This property plays a key role in model-combination. The proposed approach is applied to a simulated data set of 40 binary variables and to a model of 100 variables for illustration.

Keywords: combined model structure; decomposable graph; edge-subgraph; graph-separateness; interaction graph; Markovian subgraph; prime separator

1 Introduction

Since the work of Darroch, Lauritzen, and Speed (1980), statistical graphical models have accumulated popularity among statisticians and others in relevant research fields due to their powerful interpretive and pedagogical value in statistical modeling. The model structure embodies the formal representation of independence and conditional independence relationships. This representativeness can be employed, from a graph-theoretic point of view, not only for building a graphical model from data but also for combining graphical models.

Fienberg and Kim (1999) and Kim (2006a) considered a problem of combining conditional graphical log-linear structures and derived a combining rule for them based on the relation between the log-linear model and its conditional version. A main feature of the relation is that conditional log-linear structures appear as parts of their original model structure [see Theorems 3 and 4 in Fienberg and Kim]. The relationship becomes more explicit when the distribution is multivariate normal. Let $\mathbf{X}$ be a normal random vector. The precision matrix of the conditional distribution of a subvector $\mathbf{X}_1$ given the remaining part of $\mathbf{X}$ is the same as the $\mathbf{X}_1$ part of the precision matrix of $\mathbf{X}$ [Section 5.7, Whittaker (1990)]. Marginals of a joint probability distribution are not in general represented as parts of the joint distribution. However, there is a way that we can express explicitly the relationship between joint and

[1] Address for correspondence: Sung-Ho Kim, Department of Mathematical Sciences, KAIST, Yuseong-gu, Daejeon, 305-701, S. Korea.
E-mail: sung-ho.kim@kaist.edu

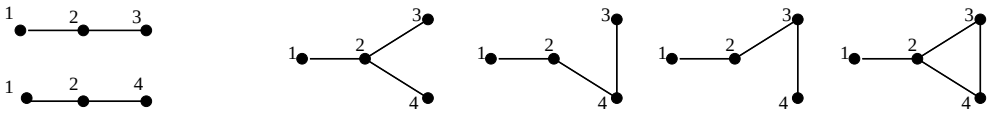


Figure 1: Two marginal models (Pair-1) on the left and the four graphs of the models in (1)

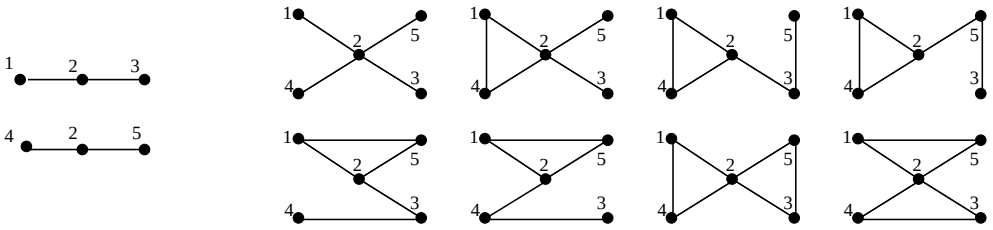


Figure 2: Two marginal models (Pair-2) on the left and the eight graphs of the models in (2)

marginal distributions under the assumption that the joint (as against marginal) probability model is graphical and decomposable (Kim, 2006b).

Suppose that we are given a pair (call it Pair-1) 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 $X_1, \cdots, X_4$. The two inter-relationships are pictured at the left end of Figure 1. We will use the notation $[\cdot]\cdots[\cdot]$ as used in Fienberg(1980) to represent a model. The graph at the top of the two at the left is of the model [12][23] and the one at the bottom is of the model [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 models have marginals in Pair-1:

$$[12][24][23],\ [12][24][34],\ [12][23][34],\ [12][234], \tag{1}$$

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

We consider another pair (call it Pair-2) of simple marginals, [12][23] and [24][25], where only one variable is shared. In this case, we have a longer list of combined models as follows:

$$\begin{aligned}&[12][24][23][25], [124][23][25], [124][23][35], [124][25][35], [125][23][34], [125][24][34],\\ &[124][235], [125][234],\end{aligned} \tag{2}$$

which are displayed in Figure 2. Model structures [124][235] and [125][234] are maximal in the sense of set inclusion among these eight models.

It is important to note that some variable(s) are independent of the others, conditional on X_2 in each of the two pairs of marginals, Pair-1 and Pair-2, and in all the models in (1) and (2). That conditional independence takes place conditional on the same variable in the marginal models and also in the combined (or joint) models underlies the main theme of 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 model and its marginals. The sets are called prime separators and defined in section 2. In the above two simple examples, X_2 forms the basis. Without the variable, X_2, the conditional independence disappears in the examples.

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 3.3). 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 $\mathcal{H}$. Based on this observation, we will propose an algorithm for combining marginal graphical models to the effect that one can build a decomposable graphical model of a large number of random variables.

While we will consider a problem of building decomposable graphical models from a collection of marginal models, there have been remarkable improvements in learning graphical models in the form of a Bayesian network [Pearl (1986, 1988)] from data. This learning however is mainly instrumented by heuristic searching algorithms since the model searching is usually NP-hard [Chickering (1996)]. A good review is given in Cooper (1999) 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 and applied in this paper would lead to an improvement in graphical modelling from data.

In section 2 we introduce notation and graphical terminologies to use; some of the terminologies are prime separator and Markovian subgraph. In section 3 we describe stochastic properties concerning the relation between a graph and a type of its Markovian subgraph and introduce basic notions and a tool for model combination and presents some important results that are instrumental for model combination. In section 4 we then define a special type of graph which is called a graph of prime separators or GOPS for short, and describe the combining procedure that is proposed in this paper. In section 5, we describe basic rules and conditions that underlie the proposed algorithm and explain how the algorithm works for model combination. The algorithm is applied to a simulated data set of 40 binary variables and to a model of 100 variables in Section 6. The paper closes at section 7 with some concluding remarks.

2 Notation and Terminologies

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 $\mathcal{C}(\mathcal{G})$ the set of cliques of $\mathcal{G}$.

A path of length n is a sequence of nodes $u = v_0, \cdots, v_n = v$ such that $(v_i, v_{i+1}) \in E$, $i = 0, 1, \cdots, n-1$ 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

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

We say that a path, $v_1, \cdots, v_n$, $v_1 \neq v_n$, is intersected by A if $A \cap \{v_1, \cdots, v_n\} \neq \emptyset$ 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

$$E_A = (E \cap A \times A) \cup \{(u, v) \in A \times A;\ \ u \text{ and } v \text{ are not separated by } A \setminus \{u, v\} \text{ in } \mathcal{G}\}. \qquad (3)$$

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 $(u_{i_1}, u_{i_2}), \cdots, (u_{i_r}, u_{i_{r+1}})$, $0 \leq i_1 < i_2 < \cdots < i_{r+1} \leq k$, 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 $C_1, \cdots, C_k$ of $\mathcal{G}$ which satisfies the following condition [see Proposition 2.17 of Lauritzen (1996)]: with $C_{(j)} = \cup_{i=1}^{j} C_i$ and $S_j = C_j \cap C_{(j-1)} \neq \emptyset$,

$$\text{for all } i > 1, \text{ there is a } j < i \text{ such that } S_i \subseteq C_j. \qquad (4)$$

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 $\chi(\mathcal{G})$, we have, for a decomposable graph $\mathcal{G}$, that

$$\chi(\mathcal{G}) = \{a \cap b; \;\; a, b \in \mathcal{C}(\mathcal{G}), \; a \neq b\}. \tag{5}$$

It is possible for some decomposable graph $\mathcal{G}$ that there are sets, a and b, in $\chi(\mathcal{G})$ 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 Markovian Subgraphs and Combined Models

For a probability distribution P of $\mathbf{X}_V$, let the logarithm of the density of P be expanded into interaction terms and let the set of the maximal domain sets of these interaction terms be denoted by $\Gamma(P)$, where maximality is in the sense of set-inclusion. We will call the set, $\Gamma(P)$, the generating class of P and denote by $\mathcal{G}(\Gamma(P)) = (V, E)$ the interaction graph of P which satisfies, under the hierarchy assumption for probability models,

$$(u, v) \in E \iff \{u, v\} \subseteq a \quad \text{for some } a \in \Gamma(P). \tag{6}$$

When confusion is not likely, we will use Γ instead of $\Gamma(P)$.

It is well known in literature [Pearl and Paz (1987)] that if a probability distribution on $\mathbf{X}_V$ is positive, then the three types of Markov property, pairwise Markov (PM), locally Markov (LM), and globally Markov (GM) properties relative to an undirected graph, are equivalent. Furthermore, for any probability distribution, it holds that

$$(GM) \Longrightarrow (LM) \Longrightarrow (PM)$$

[see Proposition 3.8 in Lauritzen (1996)]. So, we will write $M(\mathcal{G})$ instead of $M_G(\mathcal{G})$ and we will simply say that a distribution P is Markov with respect to $\mathcal{G}$ when $P \in M_G(\mathcal{G})$.

Let $\mathcal{V}$ be a set of subsets of V. We will define a collection of distributions,

$$\tilde{L}(\mathcal{G}_A, \; A \in \mathcal{V}) = \{P; \;\; P_A \in M(\mathcal{G}_A), \; A \in \mathcal{V}\}.$$

$\tilde{L}(\mathcal{G}_A, \; A \in \mathcal{V})$ is the collection of the distributions each of whose marginals is Markov with respect to its corresponding Markovian subgraph of $\mathcal{G}$.

Theorem 3.1. *For a collection $\mathcal{V}$ of subsets of V with an undirected graph $\mathcal{G}$,*

$$M(\mathcal{G}) \subseteq \tilde{L}(\mathcal{G}_A, \; A \in \mathcal{V}).$$

Proof. See the proof of Theorem 4.6 in Kim(2004). □

Theorem 3.1 lays the groundwork for model-combination since it shows the relationship between a graphical model with its graph $\mathcal{G}$ and a collection of Markovian subgraphs of $\mathcal{G}$. The set $M(\mathcal{G})$ of the probability distributions each of which is Markov with respect to $\mathcal{G}$ is contained in the set $\tilde{L}(\mathcal{G}_A,\ A \in \mathcal{V})$ of the distributions each of which has its marginals Markov with respect to their corresponding Markovian subgraphs $\mathcal{G}_A$, $A \in \mathcal{V}$. This result sheds light on our efforts in searching for $M(\mathcal{G})$ since it can be found as a subset of $\tilde{L}(\mathcal{G}_A,\ A \in \mathcal{V})$.

Let $\mathcal{G} = (V, E)$ be the graph of a decomposable model and let $V_1, V_2, \cdots, V_m$ be subsets of V. The m Markovian subgraphs, $\mathcal{G}_{V_1}, \mathcal{G}_{V_2}, \cdots, \mathcal{G}_{V_m}$, may be regarded as the structures of m submodels of the decomposable model. In this context, we may refer to a Markovian subgraph as a *marginal model structure*. These terms reflect that our goal is to find the model structure $\mathcal{G}$ based on a collection of marginal models. For simplicity, we write $\mathcal{G}_i = \mathcal{G}_{V_i}$.

Definition 3.2. *Suppose there are m Markovian subgraphs, $\mathcal{G}_1, \cdots, \mathcal{G}_m$. Then we say that graph $\mathcal{H}$ of a set of variables V is a* combined model structure *(CMS) corresponding to $\mathcal{G}_1, \cdots, \mathcal{G}_m$, if the following conditions hold:*
(i) $\cup_{i=1}^m V_i = V$.
(ii) $\mathcal{H}_{V_i} = \mathcal{G}_i$, for $i = 1, \cdots, m$. That is, $\mathcal{G}_i$ are Markovian subgraphs of $\mathcal{H}$.

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

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

Theorem 3.3. Let there be Markovian subgraphs $\mathcal{G}_i$, $i = 1, 2, \cdots, m$, of a decomposable graph $\mathcal{G}$. Then

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

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

Proof. See the proof of Theorem 4 in Kim (2006b).

4 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. If $\mathcal{G}_1$ is an interaction graph, 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.

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 $GOPS_3$ in Figure 9. As connoted in $GOPS_1$ in Figure 9, a GOPS 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.

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, $\langle\cdot|\cdot|\cdot\rangle$, 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 4.1. *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

$$\langle A|B|C\rangle_{\mathcal{G}'}, \tag{7}$$

there is no path in $\mathcal{G}'$ between A and C that bypasses B. If (i) does not hold, it is obvious that (7) 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 (7) by the definition of the Markovian subgraph. Therefore, result (ii) must hold. □

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 4.1, the same is true in $\mathcal{G}$.

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

$$(\cup_{a\in A} a) \cap (\cup_{b\in B} b) \subseteq (\cup_{c\in C} c)\,. \tag{8}$$

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

$$A \cap B \subseteq C. \tag{9}$$

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 (9) are either cliques or intersection of cliques. In the context of a junction tree, the property as expressed in (9) is called the junction property. We will call the property expressed in (8) *PS junction* property, where 'PS' is from 'prime separator.'

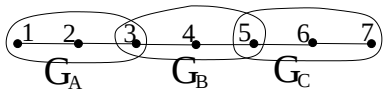
The GOPS and the junction tree are different in the following two senses. First, the basic elements are PSs in the GOPS while they are cliques in the junction tree; secondly, the GOPS is an undirected

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 (8) holds in general.

5 Markovian model-combining procedure

We can obtain a maximal CMS, $\mathcal{H}^*$ say, by adding edges, if any, to the graph $\mathcal{G}$ of Theorem 3.3 in such a way that result (ii) of Theorem 3.3 holds. In reality, however, we do not know the true model $\mathcal{G}$ and the Markovian subgraphs $\mathcal{G}_i$'s are obtained based on the corresponding marginal sets of data. If we let $\mathcal{V} = \{V_1, V_2, \cdots, V_m\}$, then $\tilde{L}(\mathcal{G}_A,\ A \in \mathcal{V})$ is the family of the probability models P of $\mathbf{X}_V$ for which $P_A \in M(\mathcal{G}_A)$ for all $A \in \mathcal{V}$. Since $\mathcal{H}^*_A = \mathcal{G}_A$, every pair of nodes, u and v, that are separated by a set of nodes in any of the graphs $\mathcal{G}_A,\ A \in \mathcal{V}$, are also separated in $\mathcal{H}^*$ by the same set of nodes. However, if a pair of nodes, u and v, are not separated in any of the graphs $\mathcal{G}_A,\ A \in \mathcal{V}$, the two nodes may or may not be separated in $\mathcal{H}^*$.

For a given set of $\mathcal{G}_A$'s, 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, $V = \{1, 2, \cdots, 7\}$, $A = \{1, 2, 3\}$, $B = \{3, 4, 5\}$, $C = \{5, 6, 7\}$, $\mathcal{V} = \{A, B, C\}$, $E^s(\mathcal{V}) = \{\{1, 3\}, \{3, 5\}, \{5, 7\}\}$, $E^a(\mathcal{V}) = \{\{i, i+1\}, i = 1, 2 \cdots, 6\}$ and $E^{rem}(\mathcal{V}) = \{\{i, j\},\ 1 \leq i < j \leq 7\} \setminus (E^s(\mathcal{V}) \cup E^a(\mathcal{V}))$.



Before describing the model-combining process in a formal manner, we will show that if a set of nodes is a PS in a Markovian subgraph, then it is not intersected in any other Markovian subgraphs.

Theorem 5.1. *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 4.1, 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 (5). 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 $\mathcal{G}_2$. □

We will call a node a *PS node* if it is contained in a PS, and a *non-PS node* otherwise. Theorem 5.1 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 expressions (1) and (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, $[12][234]$. On the other hand, when combining the models in Pair-2, i.e., $[12][23]$ and $[24][25]$, node 4 can form a clique with either $\{1, 2\}$ or $\{2, 3\}$ in the maximal CMS of models in Pair-2.

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.

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, $\mathcal{G}_1, \cdots, \mathcal{G}_m$, 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 $\mathcal{G}_{i_1}, \cdots, \mathcal{G}_{i_m}$ such that $|V_{i_j} \cap V_{i_{j+1}}| \geq |V_{i_{j+1}} \cap V_{i_{j+2}}|$ for $j = 1, 2, \cdots, m-2$. For convenience, let $i_j = j$, $j = 1, 2, \cdots, m$. 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

$$a^1 \cap a^2 \neq \emptyset, \tag{10}$$

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

$$a^1 \cap \left(\cup_{a \in \chi(\mathcal{G}_2)} a\right) = \emptyset \text{ and } a^2 \cap \left(\cup_{a \in \chi(\mathcal{G}_1)} a\right) = \emptyset \tag{11}$$

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 (11), we add edges to 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 (11). We denote the result of the combination by η_2.

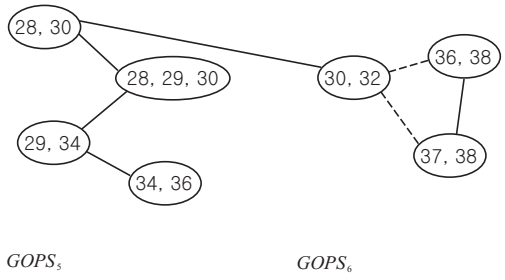


Figure 3: A graphic display of part of Step 2a corresponding to that the PS of GOPS$_5$, $\{28, 30\}$, and the PS of GOPS$_6$, $\{30, 32\}$, share node 30 and that $\{28, 30\}$ is adjacent to $\{29, 31, 32, 34\}$ and separated from $\{35, 36, 37, 38\}$ by $\{29, 31, 32, 34\}$. The non-adjacent connectedness is expressed by dashed lines.

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. 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 8 shows the GOPSs of two marginal graphs $\mathcal{G}_5$ and $\mathcal{G}_6$. As for $\mathcal{G}_5$, the set of PSs in GOPS$_1$ 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$_5$, $\{28, 30\}$, and the PS of GOPS$_6$, $\{30, 32\}$, 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 separated from $\{35, 36, 37, 38\}$ by $\{29, 31, 32, 34\}$. This separateness must be preserved, by Theorem 4.1, in the combined model of $\mathcal{G}_5$ and $\mathcal{G}_6$. We represent this non-adjacent connectedness by dashed lines in Figure 3.

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 4. 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 5.

In combining a pair of subgraphs, $\mathcal{G}_1$ 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 $\mathcal{N}_i$, $i = 1, 2$, 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 $(\{a^i\} \cup \mathcal{N}_i)$'s only.

The computing time of MCMoSt depends upon the sizes of the sets such as E^a and E^{rem} of the

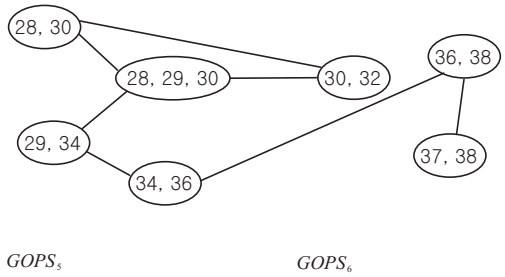


Figure 4: Step 2a in progress from Figure 3 as for the PS pairs, $\{28, 29, 30\}$ and $\{30, 32\}$ and $\{34, 36\}$ and $\{36, 38\}$.

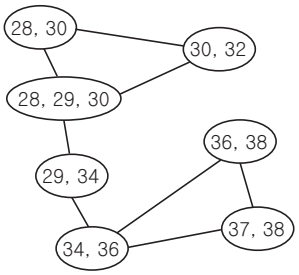


Figure 5: Step 2b as continued from Figure 4.

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.

The overall time complexity of this algorithm depends upon the number of the nodes that are not shared between graphs. 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 $\tilde{n}_1^2 O(\tilde{n}_2 + \tilde{e}_2) + \tilde{n}_2^2 O(\tilde{n}_1 + \tilde{e}_1)$, where $\tilde{e}_i$ is the number of edges in the induced subgraph of $\mathcal{G}_i$ on $V_i \backslash V_{3-i}$. 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.

6 Applications

We applied the algorithm, MCMoSt, to two sets of marginal models, one of 40 binary variables and the other of 100 variables. Using a decomposable model, say $\mathcal{G}$, of the 40 variables, we generated a data

set, constructed marginal models of 10 variables, combined the marginal models into a decomposable model of the 40 variables, and then compared the combined result with $\mathcal{G}$. As for the case of 100 variables, we did the same thing as for the 40 variables except that we use a simulated data set. We instead split the model of 100 variables into several marginal models in such a way that neighboring nodes in the original model are also neighbors in the marginal models. These applications are described in the following two subsections.

6.1 A model of 40 binary variables

In this subsection, we use a simulated data set of 40 binary variables which is obtained from the graphical log-linear model as in Figure 6, and demonstrate a modelling procedure with the data. The model in Figure 6 is decomposable, and, by Theorem 7 of Pearl (1988, p. 112), the graph can be transformed into a directed acyclic graph. So the simulated data can be generated by following the direction of the arrows in the directed graph.

The number of categorical variables that can be handled at once for log-linear modelling and the complexity of a model are limited up to the computational capacity of a computer. Our computer (IBM PC) could handle up to 10 binary variables at once at a relatively good speed of a few seconds or minutes. For any larger model with more than 10 variables, it would take hours or days with the computer. So, we applied the approach as described in the preceding section.

6.1.1 Selection of marginal contingency tables

The log-linear model is a model of association among the variables. The association between a pair of categorical variables is reflected in the coefficient of the logistic regression model of one variable of the pair upon the other (Chapter 3 of Hosmer and Lemeshow (1989)). It is thus reasonable to apply a regression method to select subsets of variables that are associated higher with each other within subsets than they are between subsets.

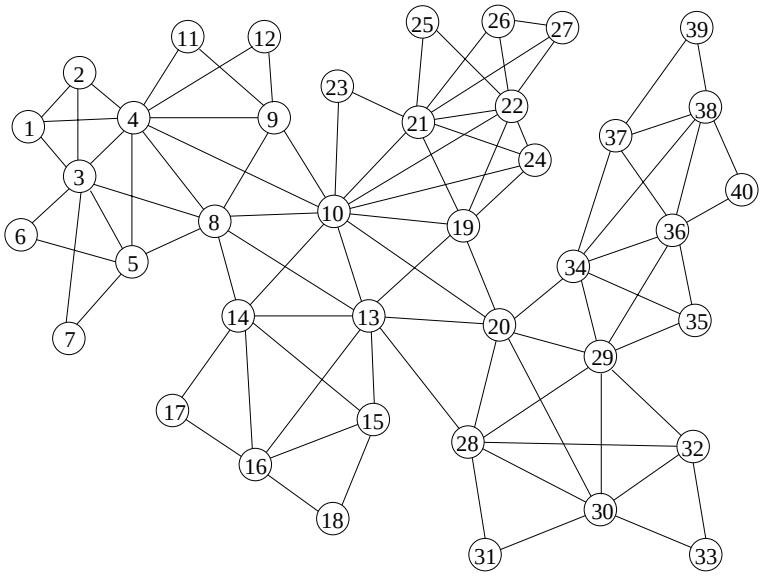


Figure 6: A model of the 40 variables that are used for the simulation study

Since the logistic regression analysis requires dealing with the whole set of variables, we run into the same, large modelling problem. So, we apply a nonparametric, tree regression method as an effective alternative by using the **tree** and its related commands in S-plus (Chambers and Hastie, 1992). In the tree regression analysis for variable X^*, we select regressor variables one after another in such a way that, if variables $X_1, X_2, \cdots, X_i$ are selected already with their outcomes, $x_1, x_2, \cdots, x_i$, respectively, then we select the variable X' for which $var(X^*|X_1 = x_1, X_2 = x_2, \cdots, X_i = x_i) - E(var(X^*|X_1 = x_1, X_2 = x_2, \cdots, X_i = x_i, X')) > 0$ is maximized.

Using the simulated data set of size 250,000, we applied the tree regression to every of the 40 variables and selected a set of regressor variables which explains about 80% of the total variation of the response variable.

The selected regressor variables are listed in the form of a matrix in Figure 7. The entries in the matrix are either 0 or 1. The regressor variables for the variable X_j are listed as 1's in the jth column. For instance, variables X_4, X_6, X_7, X_8 are selected as the regressor variables for X_5. The variables are

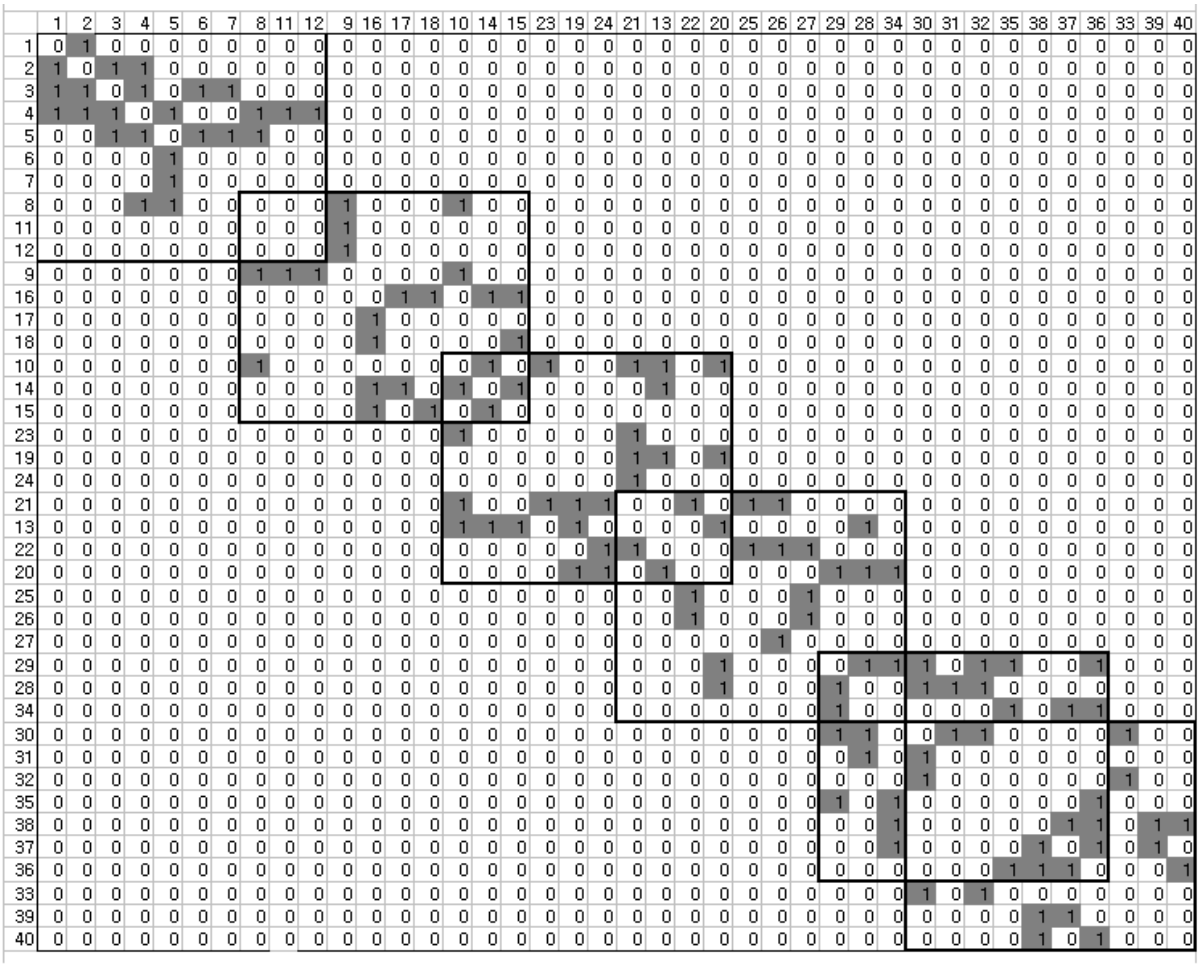


Figure 7: A matrix of regressor-response relationships for the 40 variables as obtained from a tree regression analysis. The 1's in column j are the indicators of the regressor variables for the response variable X_j. The six blocks correspond to the six subsets of variables listed in Table 1.

not arranged in their index order but are arranged so that more highly associated variables appear next to each other.

We denote by $\psi(j)$ the selected set of the indexes of the regressor variables for X_j. That $i \in \psi(j)$ does not necessarily imply that $j \in \psi(i)$. For instance, $\psi(7) = \{3, 5\}$, while $\psi(3) = \{2, 4, 5\}$. This asymmetry in the regressor-response relationship is not unusual. The regression tree we consider is constructed by selecting regressor variables one after another in such a way that the most informative variable for a particular response variable may be selected. So, the fact that $\psi(3) = \{2, 4, 5\}$ means that X_7 is less informative for X_3 than those in $\psi(3)$. This asymmetry can also be due to sampling error. Although we have seen some instances of asymmetry in the regressor-response relationship, symmetry is a prevalent feature in the relationship, as we can see in the matrix.

We grouped the 40 variables into 6 subsets of 10 variables in such a way that the variables share more variables as regressor variables within subsets of variables than the variables share between subsets. The grouping can be carried out easily from the arrangement in Figure 7, and the six subsets are listed in Table 1. As noted in the figure and the table, subsets i and $i + 1$, $i = 1, 2, \cdots, 5$, share a nonempty set of variables. In particular, subsets 5 and 6 share as many as 7 variables. Once an arrangement of variables such as in Figure 7 is obtained, the grouping is subject to the optimal number of variables that our computer can handle at once and the variable-sharing between neighboring subsets of variables. For example, we need a subset such as V_5 as a bridge between V_4 and V_6 since $V_4 \cap V_6 = \emptyset$.

Table 1: The indexes of the variables in the 6 subsets, $V_1, \cdots, V_6$.

$$
\begin{aligned}
V_1 &= \{1, 2, 3, 4, 5, 6, 7, 8, 11, 12\} \\
V_2 &= \{8, 9, 10, 11, 12, 14, 15, 16, 17, 18\} \\
V_3 &= \{10, 13, 14, 15, 19, 20, 21, 22, 23, 24\} \\
V_4 &= \{13, 20, 21, 22, 25, 26, 27, 28, 29, 34\} \\
V_5 &= \{28, 29, 30, 31, 32, 34, 35, 36, 37, 38\} \\
V_6 &= \{30, 31, 32, 33, 35, 36, 37, 38, 39, 40\}
\end{aligned}
$$

Table 2: Goodness-of-fit levels of the six marginal models

Marginal model	d.f.	Pearson χ^2	p-value
1	567	547.50	0.714
2	645	667.41	0.263
3	601	589.07	0.628
4	649	679.25	0.199
5	617	591.89	0.760
6	604	621.53	0.302

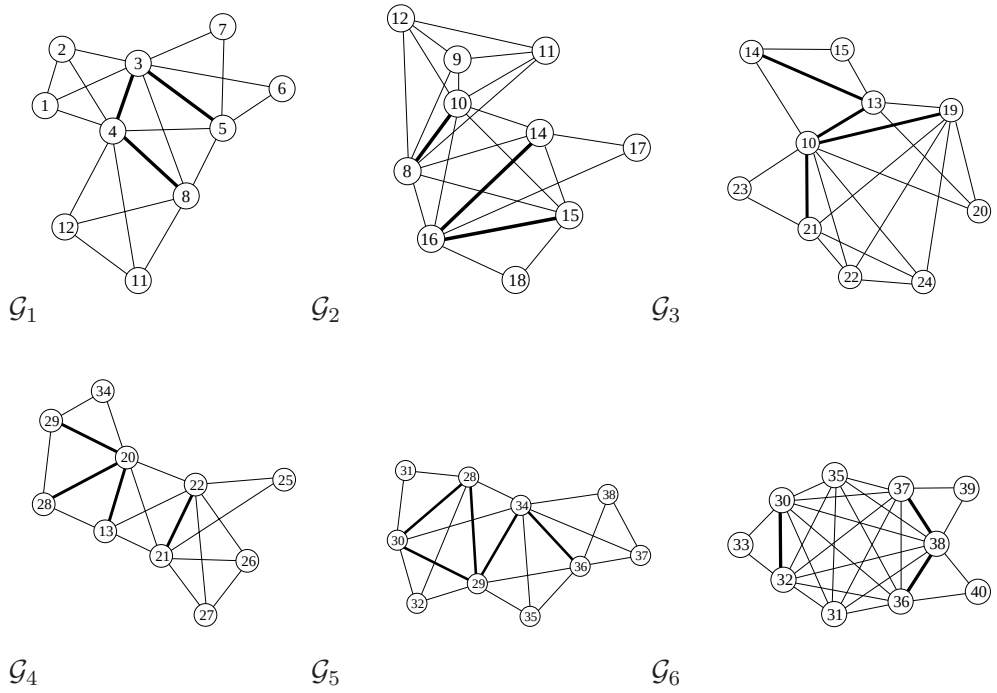


Figure 8: Marginal models of the model in Figure 6 for the 6 subsets of variables which are listed in Table 1. $\mathcal{G}_i$ is the decomposable log-linear model for subset V_i. PSs are represented by thick lines. See Figure 9 for the PSs of the 6 marginal models.

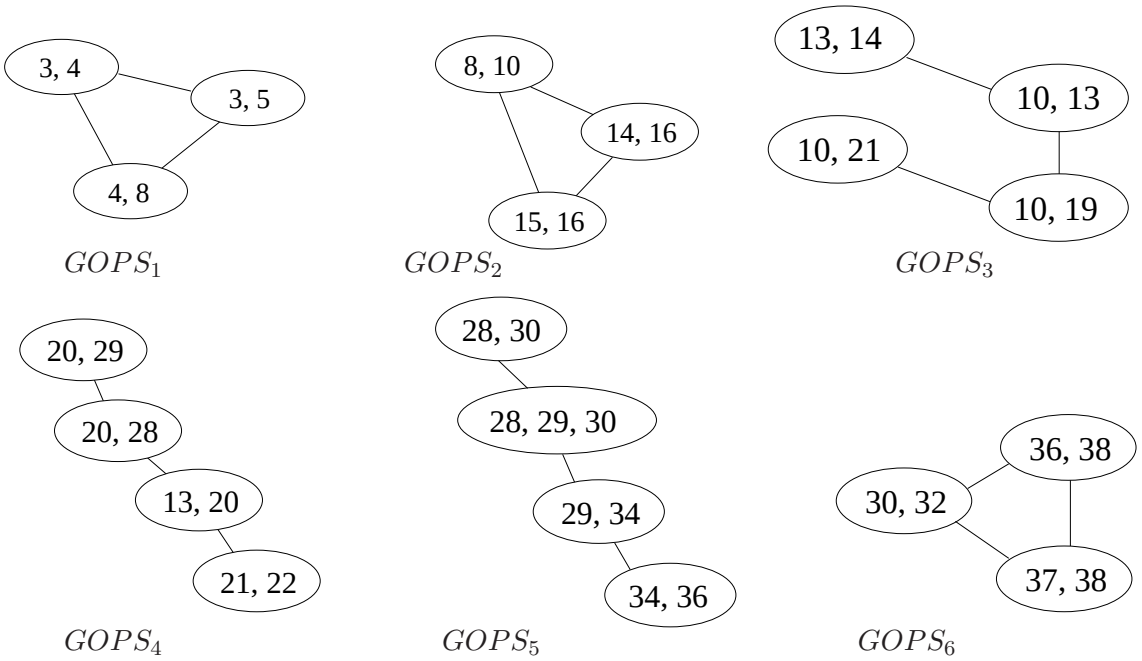


Figure 9: The GOPS's of the six marginal models in Figure 8.

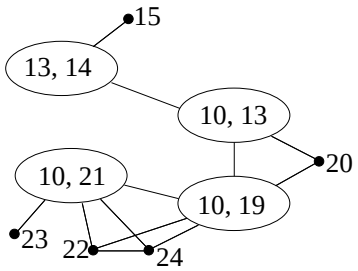


Figure 10: The graph obtained by linking non-PS variables (bullets) to the PS's of $\mathcal{G}_3$.

6.1.2 Marginal log-linear models for the 6 subsets of variables

For every subset of 10 variables, we found, by applying a backward deletion method, a decomposable log-linear model that fits well to the corresponding marginal set of data. In applying the backward deletion method, we began with an all k-way interaction model. Once the k value was found for a subset of variables with an acceptable goodness-of-fit level, we began removing interaction terms until an appropriate model was reached. Figure 8 displays the model structures for the six subsets of variables and their goodness-of-fit results are listed in Table 2. The p-values of the goodness-of-fit tests are all larger than or equal to 0.199. We denote by $\mathcal{G}_i$ the model structure of the variables indexed in V_i and by $c(k_1, k_2, \cdots, k_r)$ the PS which consists of variables, $X_{k_1}, X_{k_2}, \cdots, X_{k_r}$.

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 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.

6.1.3 Combination of marginal models

The variable-sharing between V_i and V_{i+1}, $i = 1, 2, \cdots, 5$, 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 $\mathcal{G}_6$, and then keep combining marginal models in the order of $\mathcal{G}_4,\ \mathcal{G}_3,\ \mathcal{G}_2,\ \mathcal{G}_1$.

Models $\mathcal{G}_5$ and $\mathcal{G}_6$ are combined in Section 5 and then we combine the GOPS in Figure 5 with GOPS $_4$ in Figure 9 by applying the MCMoSt. We repeat this combining process until all the GOPS's in Figure 9 are combined to obtain the combined GOPS in Figure 11. This combining process took 2 seconds by MCMoSt with an IBM compatible PC and MATLAB 7.0.

A PS is itself a complete subgraph and so is a clique of PSs. So we can easily transform the graph in Figure 11 into the undirected graph in Figure 12. This is a maximal CMS of the six marginal models as listed in Figure 8. The true model in Figure 6 is fully recovered in the maximal CMS except the 5 thick

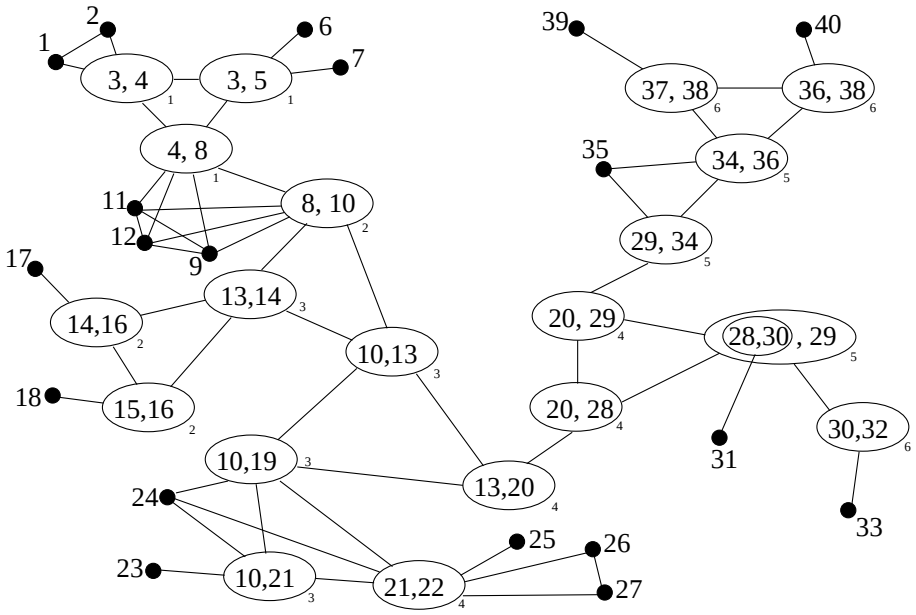


Figure 11: An independence graph of PSs and non-PS variables. The PSs are in ovals and the dots are for the non-PS variables, and the small numbers at the bottom-right of the ovals are the submodel labels of which the ovals are PSs.

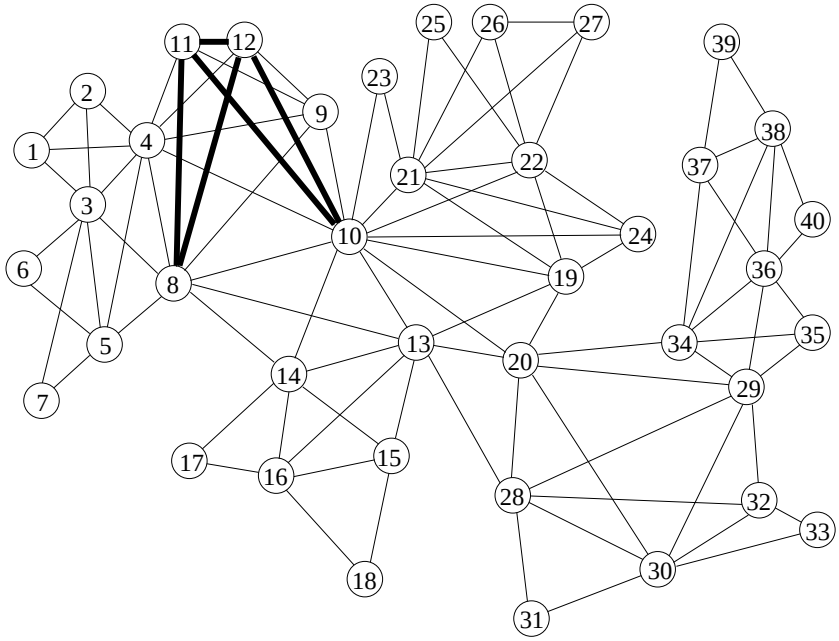


Figure 12: The combined model structure which is obtained from the independence graph in Figure 11. The thick edges are additional to the true model in Figure 6.

edges appearing in Figure 12. These additional edges were created because both X_4 and X_9 were not contained in any of the marginal models. If X_4 had been added to $\mathcal{G}_2$, then $\mathbf{X}_{\{4,9\}}$ would have separated X_{11}, X_{12}, and $\mathbf{X}_{\{8,10\}}$ from each other, making those additional edges unnecessary. This phenomenon

of additional edges leads 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.

6.2 A model of 100 variables

In this subsection, we intend to check availability of MCMoSt when there are a large number of marginal models to combine. We consider a decomposable model, say $\mathcal{G}^*$, of 100 variables as given in Figure 13. For this model, we did not use data to select subsets of variables but use the model structure to split it into 18 marginal models in such a way that non-adjacent nodes in $\mathcal{G}^*$ are not neighbors in any of the marginal models. The 18 marginal models or Markovian subgraphs are in Figures 16 and 17. As displayed in these figures, the marginal models are of sizes from 5 to 12 and none of them is isolated from the rest. Table 3 lists the variables of the marginal models and the neighboring models of each marginal model.

The combining process took 37 seconds until we got the combined result in Figure 15, where we see four thick edges which are additional to the edges in the true model, $\mathcal{G}^*$. This addition is due to the same reason as is for the additional edges in Figure 12.

The four thick edges, $(13, 20), (13, 22), (20, 21), (20, 22)$, took place when combining marginal models $\mathcal{G}_2$ and $\mathcal{G}_3$. Note that the set $\{10, 19\}$ forms a PS in the true model but not in any of the marginal models (see Figure 14.) Thus those nodes which are separated by $\{10, 19\}$ in the true model become neighbors to each other in the combined result. A similar argument applies to the additional edge $(29, 86)$. The set $\{28, 35\}$ is a PS in the true model but not in any of the marginal models. For

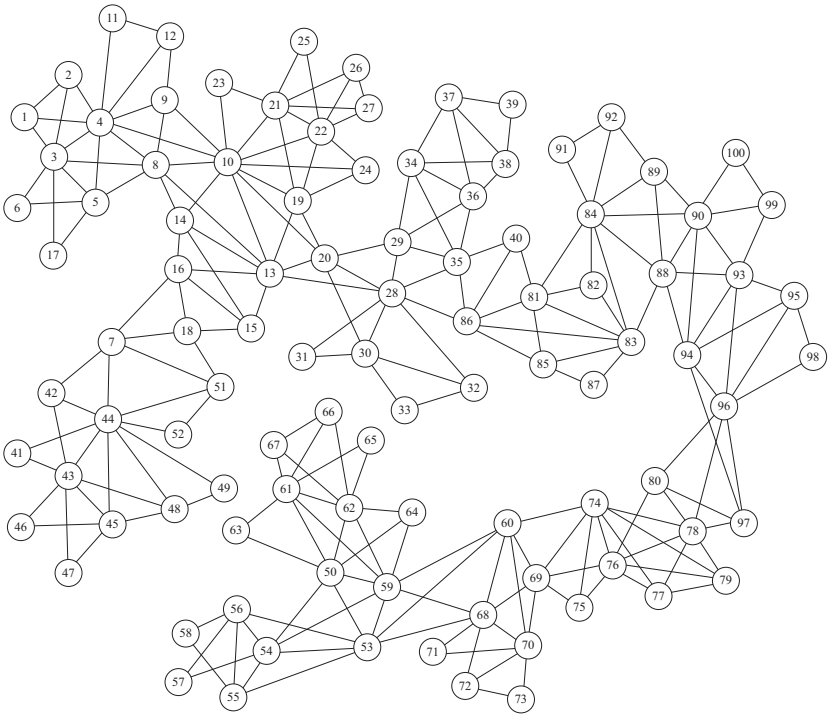


Figure 13: A model of 100 variables

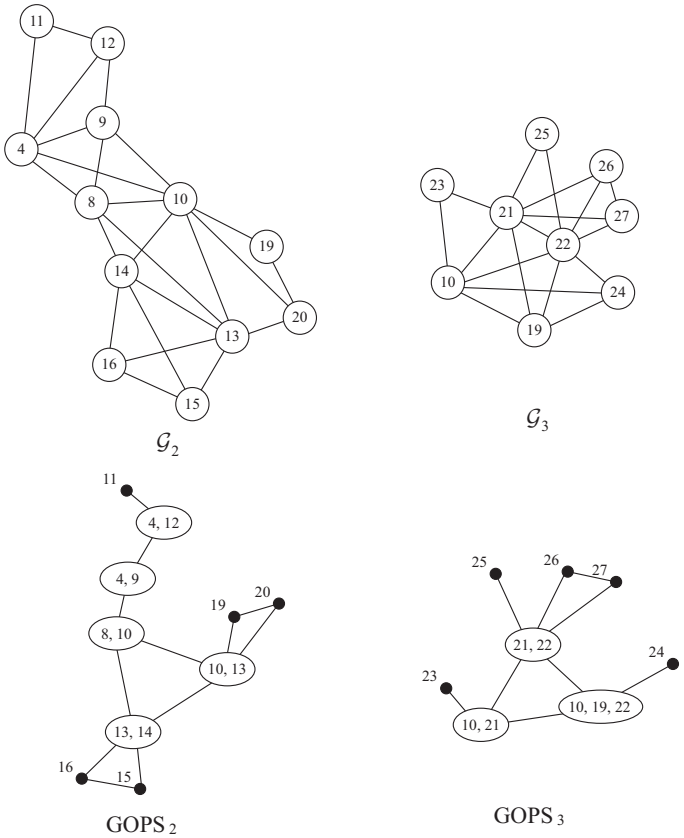


Figure 14: Two marginal models of the model in Figure 13 which include nodes 10 and 19.

instance, the marginal model $\mathcal{G}_7$ contains the nodes, 28 and 35, but they do not form a PS therein.

7 Concluding remarks

In this paper, we propose an approach of combining marginal model structures that are decomposable. A main idea behind the approach is that we group the random variables into several subsets such that the association among the variables are higher within subsets than between subsets. The sizes of the subsets of random variables are bounded by the computing capacity of the computer to use. It is desirable that the subset sizes are as large as possible since smaller subsets would end up with more edges which are not contained in the actual model.

In combining marginal models, we need to take into consideration the number of variables that are shared by neighboring marginal models. As more variables are shared, the model-combination becomes easier since the shared variables are used as road signs in constructing a maximal CMS of a given set of marginal models. 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 (8) and the separateness condition are instrumental for locating PSs in model-combination.

In the two applications, we assumed true models. But in reality, we do not know true models but only have marginal models that are developed from subsets of data involving manageable sizes of

variables or different sources of data. When we have a data set of a full set of random variables, we can build a model as in subsection 6.1. Even though only marginal models are provided, we can apply MCMoSt to obtain combined results as long as none of the models are isolated from the rest.

In the applications, the sizes of the marginal models are between 5 and 12. In combining models, it is desirable that there are smaller number of larger marginal models. For instance, let $V_1 \cup V_2 \cup V_3 = V$. Then combining $\mathcal{G}_{V_1}$, $\mathcal{G}_{V_2}$, and $\mathcal{G}_3$ is at most as efficient as combining $\mathcal{G}_{V_1 \cup V_2}$ and $\mathcal{G}_3$. This is because the combined result of $\mathcal{G}_{V_1}$ and $\mathcal{G}_{V_2}$ may at best be the same as $\mathcal{G}_{V_1 \cup V_2}$.

Consider two marginal models, $\mathcal{G}_i = (V_i, E_i)$, $i = 1, 2$. As mentioned at the end of section 5, the time complexity is at most of order $\max\{|V_1|^2(|V_2| + |E_2|),\ |V_2|^2(|V_1| + |E_1|)\}$. As the models share more variables, the complexity decreases accordingly.

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.

Acknowledgements

This work was supported by Korea Research Foundation Grant (KRF-2003-015-C00096).

References

Hosmer, D.W. and Lemeshow, S. 1989. Applied Logistic Regression. New York: Wiley.

Asmussen, S. and Edwards, D. (1983). Collapsibility and response variables in contingency tables, *Biometrika*, **70**(3), 567-578.

Bergsma, W.P. and Rudas, T. (2002). Marginal models for categorical data, *Ann. Statist.*, **30**(1), 140-159.

Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J. (1984). *Classification and Regression Trees*. Belmont, CA.: Wardsworth International Group.

Chambers, J.M. and Hastie, T.J. (1992). *Statistical Models in S*. Pacific Grove, CA.: Wadsworth & Brooks/ Cole Advanced Books & Software.

Chickering,D. (1996). Learning Bayesian networks is NP-complete, in D. Fisher and H. Lenz (Eds.) *Learning from Data* Springer-Verlag, 121-130.

Chickering, D.M. and Heckerman, D.(1999). Fast learning from sparse data, *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 109-115.

Cooper, G.F.(1999). An overview of the representation and discovery of causal relationships using Bayesian networks, in C. Glymour and G.F. Cooper (Eds.) *Computation, Causation, and Discovery*, The MIT Press, 3-62.

Cox, D.R. and Wermuth, N. (1999). Likelihood factorizations for mixed discrete and continuous variables, *Scand. J. Statist.*, **26**, 209-220.

Dawid, A.P. (1979). Conditional independence in statistical theory, *J. R. Statist. Soc.* B, **41**(1), 1-31.

Didelez, V. and Edwards, D. (2004). Collapsibility of graphical CG-regression models. *Scand. J. Statist.*, **31**(4), 535-552.

Edwards, D. (1995). *Introduction to Graphical Modelling*. Springer-Verlag.

Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data* Cambridge, MA: The MIT Press.

Fienberg, S.E. and Kim, S.-H. (1999). Combining conditional log-linear structures, *J. Am. Statist. Ass.*, **445**(94), 229-239.

Friedman, N. and Goldszmidt, M. (1998). Learning Bayesian networks with local structure, in M.I. Jordan, (Ed.), *Learning in Graphical Models* Dordrecht, Netherlands: Kluwer, 421-460.

Heckerman, D., Geiger, D. and Chickering, D.M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning*, **20**, 197-243.

Kiiveri, H., Speed, T.P. and Carlin, J.B. (1984). Recursive causal models, *J. Aust. Math. Soc.* A, **36**, 30-52.

Kim, S.-H. (2004). Combining decomposable model-structures, Research Report 04-15, Division of Applied Mathematics, KAIST, Daejeon, 305-701, S. Korea.

Kim, S.-H. (2006a). Conditional log-linear structures for log-linear modelling, *Computational Statistics and Data Analysis*, **50**(8), 2044-2064.

Kim, S.-H. (2006b). Properties of Markovian subgraphs of a decomposable graph, *Lecture Notes in Artificial Intelligence*, LNAI **4293**, 15-26.

Lauritzen, S.L. (1996). *Graphical Models.* Oxford: Oxford University Press.

Lauritzen, S.L., Speed, T.P. and Vijayan, K. (1984). Decomposable graphs and hypergraphs, *J. Aust. Math. Soc.* A, **36**, 12-29.

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.

Neil, J.R., Wallace, C.S. and Korb, K.B. (1999). Learning Bayesian networks with restricted causal interactions, in K.B. Laskey and H. Prade, (Eds.), *Uncertainty in Artificial Intelligence* San Francisco, California: Morgan Kaufmann Publishers, 486-493.

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*, San Mateo, CA.: Morgan Kaufmann.

Pearl, J. and Paz, A. (1987). Graphoids: a graph based logic for reasoning about relevancy relations. In *Advances in Artificial Intelligence II*, (ed. B.D. Boulay, D. Hogg, and L. Steel), Amsterdam: North-Holland, pp. 357-63.

Robins, J.M., Scheines, R., Spirtes, P., and Wasserman, L. (2003). Uniform consistency in causal inference, *Biometrika*, **90**, 3, 491-515.

Tarjan, R. E. (1972). Depth-first search and liear graph algorithms. *SIAM J. Comput.* **1**(2), 146-160.

Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *J. Am. Statist. Ass.*, **75**(373), 963-972.

Wermuth, N. and Lauritzen, S.L. (1983). Graphical and recursive models for contingency tables, *Biometrika*, **70**(3), 537-552.

Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*, New York: Wiley

Appendix: Tables and graphs for Subsection 6.2

Table 3: The 18 marginal models of the model in Figure 13.

Marginal models	Variables	Neighbor models
1	1, 2, 3, 4, 5, 6, 8, 9, 11, 12, 17	2
2	4, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20	1, 3, 4, 7
3	10, 19, 21, 22, 23, 24, 25, 26, 27	2, 7
4	7, 14, 15, 16, 18, 51	2, 5
5	7, 18, 42, 44, 51, 52	4, 6
6	41, 42, 43, 44, 45, 46, 47, 48, 49, 52	5
7	10, 13, 19, 20, 28, 29, 30, 31, 32, 33, 35	2, 3, 8, 9
8	29, 34, 35, 36, 40, 81, 85, 86	7, 9, 10
9	29, 34, 35, 36, 37, 38, 39	7, 8
10	81, 82, 83, 84, 85, 86, 87, 88	8, 11
11	83, 84, 88, 89, 90, 91, 92, 93, 94, 99, 100	10, 12, 13
12	90, 93, 94, 95, 96, 98, 99, 100	11, 13
13	74, 76, 77, 78, 79, 80, 93, 94, 95, 96, 97	11, 12, 14
14	60, 68, 69, 70, 74, 75, 76, 77, 78, 79, 80	13, 15
15	50, 53, 59, 60, 68, 69, 70, 71, 72, 73	14, 17, 18
16	54, 55, 56, 57, 58	18
17	50, 59, 61, 62, 63, 64, 65, 66, 67	15, 18
18	50, 53, 54, 55, 56, 59, 61, 62, 63, 64	15, 16, 17

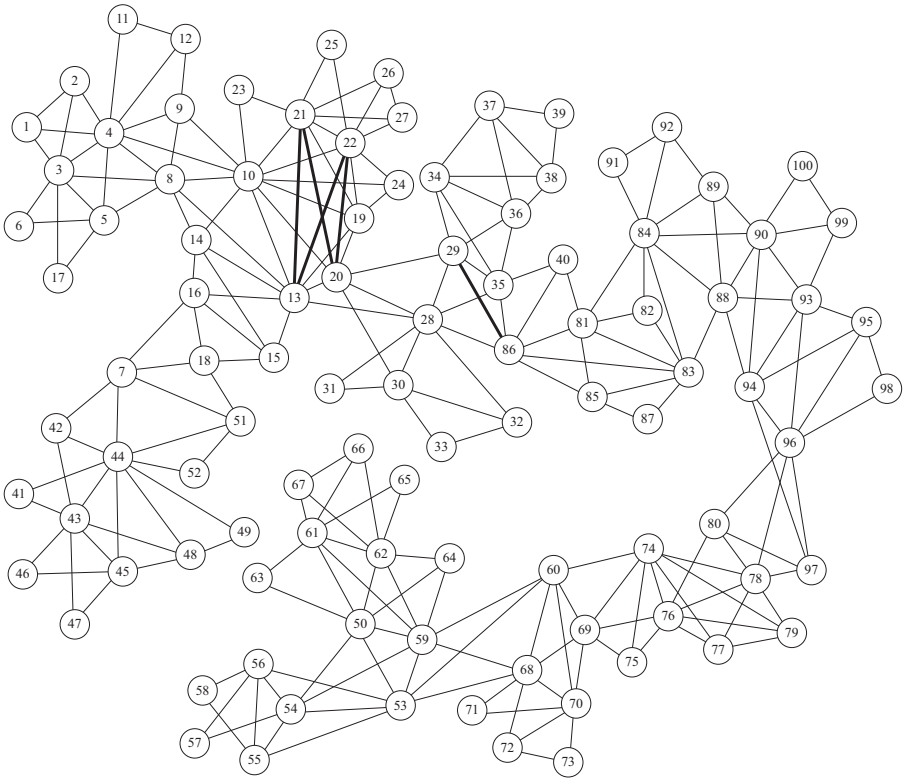


Figure 15: The combined result of the 18 marginal models in Figures 16 and 17. The thick edges are additional to the true model in Figure 13.

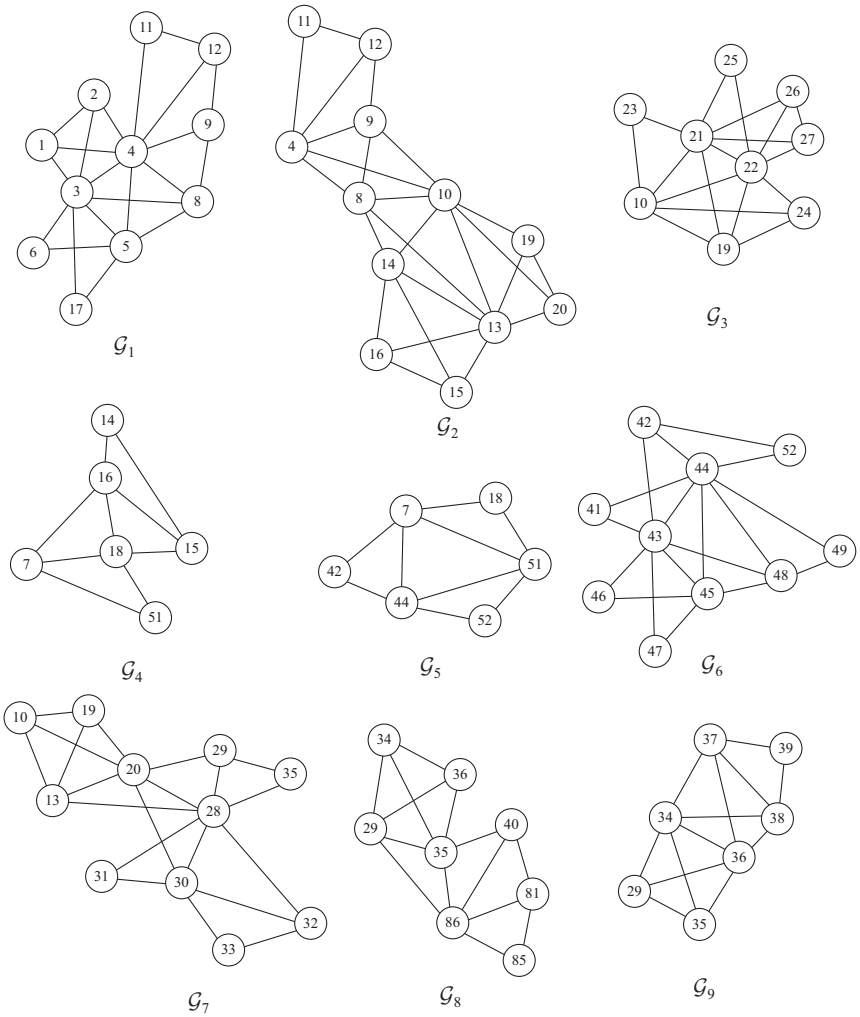


Figure 16: The first nine marginal models, $\mathcal{G}_1, \cdots, \mathcal{G}_9$, of the model in Figure 13.

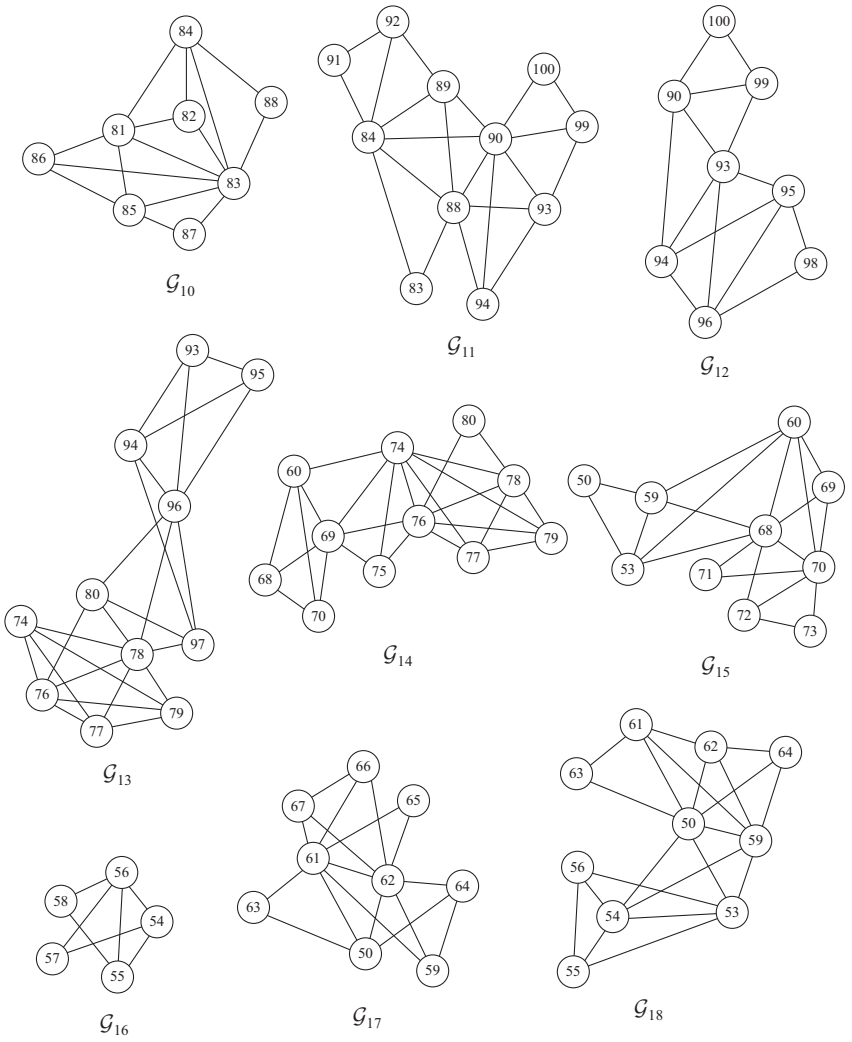


Figure 17: The second nine marginal models, $\mathcal{G}_{10}, \cdots, \mathcal{G}_{18}$, of the model in Figure 13.