An Algorithm for Combining Decomposable Graphical Models

by

Sung-Ho Kim and Sangjin Lee

Applied Mathematics
Research Report
07-09
November 12, 2007

DEPARTMENT OF MATHEMATICAL SCIENCES

KAIST 한국과학기술원
Korea Advanced Institute of Science and Technology

An Algorithm for Combining Decomposable Graphical Models

Sung-Ho Kim [a,*], Sangjin Lee [a]

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

Abstract

We propose an algorithm for combining decomposable graphical models and apply it for building decomposable graphical log-linear models which involve a large number of variables. A main idea in this algorithm is that we group the random variables that are involved in the data into several subsets of variables, build graphical log-linear models for the marginal data, and then combine the marginal models using graphs of prime separators (section 2). The application of the algorithm to a data set of 40 binary variables is very successful, yielding a model which is mostly the same as the true one.

Key words: Combined model structure, Graph-separateness, Interaction graph; Markovian subgraph, Prime separator.

1 Introduction

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

* Corresponding author.
Email addresses: shkim@amath.kaist.ac.kr (Sung-Ho Kim), Lee.Sang.jin@kaist.ac.kr (Sangjin Lee).

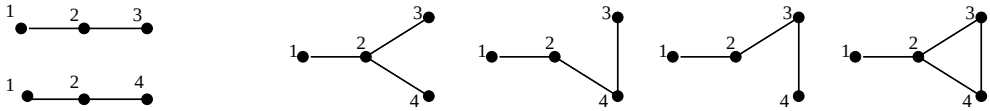

Fig. 1. Two marginal models (Pair-1) on the left and the four graphs of the models in (1)

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:

$$[12][24][23][25], [124][23][25], [124][23][35], [124][25][35], [125][23][34], [125][24][34], [124][235], [125][234]. \tag{2}$$

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.

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)]. 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 8). 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.

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. The procedure is applied successfully to a simulated data set in section 5. Finally, concluding remarks are given in section 6.

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. 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 connected by an edge. A maximal complete subgraph is called a clique of $\mathcal{G}$, where the maximality is in the sense of set-inclusion. We denote by $\mathcal{C}(G)$ 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

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

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

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

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

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 $S_j = C_j \cap C_{(j-1)} \neq \emptyset$,

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

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

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

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 write $\mathcal{G}(P)$ instead of $\mathcal{G}(\Gamma(P))$.

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 $P \in M_G(\mathcal{G})$.

For $A \subset V$, we denote by $\mathcal{J}_A$ the collection of the connectivity components in $\mathcal{G}^{ind}_{A^c}$ and let

$$\beta(\mathcal{J}_A) = \{bd(B);\ B \in \mathcal{J}_A\}.$$

We also define $\bar{\Gamma}(P_A)$ as

$$\bar{\Gamma}(P_A) = (\Gamma(P) \cap A) \cup \beta(\mathcal{J}_A). \tag{7}$$

From this, it follows that

$$\beta(\mathcal{J}_A) \preceq \bar{\Gamma}(P_A) \preceq \mathcal{C}(\mathcal{G}(\bar{\Gamma}(P_A))).$$

The second $\preceq$ holds since it is possible that, for some $B \in \mathcal{J}_A$, $bd(B)$ is a strict subset of a clique in $\mathcal{G}(\bar{\Gamma}(P_A))$.

The following result is immediate from (7).

Theorem 3 For a distribution P of $\mathbf{X}_V$ and $A \subseteq V$,

$$\mathcal{G}(\bar{\Gamma}(P_A)) = \mathcal{G}(P)_A.$$

Proof: By definition, the interaction graph corresponding to the right hand side of (7) is $\mathcal{G}(P)_A$. Thus the result follows.

From this theorem and the fact that $\Gamma(P_A) \preceq \bar{\Gamma}(P_A)$, we have

Corollary 4 For a distribution P of $\mathbf{X}_V$ and $A \subseteq V$,

$$P_A \in M(\mathcal{G}(P)_A).$$

From Theorem 3, we can also derive a result concerning both the relationship between a distribution P and a graph $\mathcal{G}$ and the relationship between P_A and $\mathcal{G}_A$.

Corollary 5 For a distribution P of $\mathbf{X}_V$ and $A \subseteq V$, suppose that $P \in M(\mathcal{G})$ for an undirected graph $\mathcal{G}$. Then

$$P_A \in M(\mathcal{G}_A).$$

Proof: Since $P \in M(\mathcal{G})$, we have $\mathcal{G}(P) \subseteq^e \mathcal{G}$. This implies that $\mathcal{G}(P)_A \subseteq^e \mathcal{G}_A$. So, by Corollary 4, we have the desired result.

If we regard $\mathcal{G}$ as an interaction graph of a distribution P, then Corollary 4 says that $P_A \in M(\mathcal{G}_A)$, which means that P_A is Markov with respect to $\mathcal{G}_A$. We call $\mathcal{G}_A$ a Markovian subgraph of $\mathcal{G}$ in this context.

For $A \subseteq V$, we define $M(\mathcal{G})_A$ and $L(\mathcal{G}_A)$ as

$$M(\mathcal{G})_A = \{P_A;\ \ P \in M(\mathcal{G})\}$$

and

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

$M(\mathcal{G})_A$ is the set of the marginal distributions on $\mathbf{X}_A$ of a distribution P which is Markov with respect to $\mathcal{G}$; $L(\mathcal{G}_A)$ is the set of the distributions of $\mathbf{X}_V$ whose marginal P_A on $\mathbf{X}_A$ is Markov with respect to $\mathcal{G}_A$.

By definition and Corollary 5, we have the following:

$$L(\mathcal{G}) = M(\mathcal{G}),$$

$$M(\mathcal{G}) \subseteq L(\mathcal{G}_A), \qquad \text{(by Corollary 5)} \qquad (8)$$
$$P \in L(\mathcal{G}_A) \Longleftrightarrow P_A \in M(\mathcal{G}_A)$$

and

$$M(\mathcal{G})_A \subseteq M(\mathcal{G}_A).$$

The last expression holds since, if a distribution Q is in $M(\mathcal{G})_A$, it means that $Q = P_A$ for some distribution P in $M(\mathcal{G})$, and so, by Corollary 5, it follows that $Q \in M(\mathcal{G}_A)$.

It follows from (8) that, for $A, B \subseteq V$,

$$M(\mathcal{G}) \subseteq L(\mathcal{G}_A) \cap L(\mathcal{G}_B).$$

We will derive a generalized version of this result below.

Let $\mathcal{V}$ be a set of subsets of V. We will define another 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 6 *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: Let $P \in M(\mathcal{G})$. Then, by (8), $P \in L(\mathcal{G}_A)$ for $A \in \mathcal{V}$. By definition, $P_A \in M(\mathcal{G}_A)$. Since this holds for all $A \in \mathcal{V}$, it follows that $P \in \tilde{L}(\mathcal{G}_A,\ A \in \mathcal{V})$. This completes the proof.

Theorem 6 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 marginal models 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 7 *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 8 Let there be Markovian subgraphs $\mathcal{G}_i$, $i = 1, 2, \cdots, m$, of a decomposable graph $\mathcal{G}$. Then

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

(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 (2006).

If $\mathcal{G}$ is not a maximal CMS of $\mathcal{G}_i$, $i = 1, 2, \cdots, m$, then we can construct a maximal CMS, $\mathcal{H}^*$ say, by adding edges to $\mathcal{G}$. For notational convenience, let $\mathcal{V} = \{V_1, V_2, \cdots, V_m\}$. Then for $A \in \mathcal{V}$, we have $\mathcal{H}^*_A = \mathcal{G}(P)_A$. If P is the distribution of $\mathbf{X}_V$ and we put $\mathcal{G} = \mathcal{G}(P)$ in Theorem 6, then we end up with the summarizing expression

$$M(\mathcal{G}(P)) \subseteq M(\mathcal{H}^*) \subseteq \tilde{L}\left(\mathcal{G}(P)_A,\ A \in \mathcal{V}\right), \tag{9}$$

where the first inequality follows since $\mathcal{G}(P) \subseteq^e \mathcal{H}^*$. Since $P \in M(\mathcal{G}(P))$, expression (9) implies that P is also Markov relative to the maximal CMS, $\mathcal{H}^*$, and P_A's are Markov relative to their corresponding $\mathcal{G}(P)_A$'s.

While Theorem 8 shows the relationship among $\mathcal{G}$, $\mathcal{H}^*$, and $\{\mathcal{G}_A,\ A \in \mathcal{V}\}$ from a graphical point of view, the inequalities in expression (9) say that the true probability model lies in the set of the probability models that are obtained based on $\{\mathcal{G}_A,\ A \in \mathcal{V}\}$.

4 Combining Procedure

We will begin this section with introducing 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 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

$$(\cup_{a \in A} a) \cap (\cup_{b \in B} b) \subseteq (\cup_{c \in C} c)\,. \tag{10}$$

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

$$A \cap B \subseteq C. \tag{11}$$

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 (11) are either cliques or intersection of cliques. We will call the property in (10) the PS junction property.

Suppose that we have a contingency table of a large number of random variables whose model structure is graphical and that we cannot handle the whole table at once for modelling. In this situation, we propose to develop several marginal log-linear models and use them for building a model for the whole data set.

In selecting subsets of random variables, it is important to have the random variables associated more highly within subsets than between subsets. This way of subset-selection would end up with subsets of random variables where random variables that are neighbors in the graph of the model structure of the whole data set are more likely to appear in the same marginal model. Once marginal models are

obtained, we construct a maximal CMS based on the collection of the GOPS's of the marginal models.

Suppose there are m marginal models. For operational convenience, it is recommended that we start combining a pair of marginal models which shares more random variables than any other pairs of marginal models. As more random variables are shared between a pair of marginal models, say $\mathcal{G}_1$ and $\mathcal{G}_2$, it is more likely that the model-combination gets easier, since the shared variables restrict the locations of the other random variables in $V_1 \cup V_2$ in the model-combination. Once the first pair of marginal models are combined, it is desirable that a marginal model is selected which shares most random variables with $V_1 \cup V_2$, continue model-combination until all the marginal models are combined into a model which is a maximal CMS of the m marginal models.

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 define $\eta_1 = \{\mathcal{G}_1\}$.

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

$$a^1 \cap a^2 \neq \emptyset, \tag{12}$$

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

$$a^1 \cap \left(\cup_{a \in \chi(\mathcal{G}_2)} a\right) = \emptyset \text{ and } a^2 \cap \left(\cup_{a \in \chi(\mathcal{G}_1)} a\right) = \emptyset \tag{13}$$

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 (13), 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 (13). We denote the result of the combination by η_2.

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.

Figure 1 is obtained by applying this procedure to Pair-1 in section 1. Since $\{2\}$ is the only PS in both of the marginal models, the resulting GOPS consists of $\{2\}$ only. Node 1 is shared between the marginal models. So the maximal CMS, which is at the right end of Figure 1, is obtained by applying the Separateness condition to the nodes 3 and 4.

In combining a pair of Markovian subgraphs, say $\mathcal{G}_1$ and $\mathcal{G}_2$, 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. If the two subgraphs share no nodes, it is desirable to find a marginal model from given data which shares variables (nodes) with both of the subgraphs rather than simply make a list of all the possible patterns of edge-additions between the subgraphs, $\mathcal{G}_1$ and $\mathcal{G}_2$, under the Separateness condition.

5 Application

We use a simulated data set of 40 binary variables which is obtained from the graphical log-linear model as in Figure 2, and apply the combining procedure to the data. The model in Figure 2, without the thick edges, 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 proposed combining procedure.

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 by applying the regression tree method (Chambers and Hastie, 1992), and we ended up with the six subsets as listed in Table 1. We find, by the regression tree method, a subset, b_i say, of variables which are mostly informative for X_i, $i = 1, 2 \cdots, 40$.

We then have 40 subsets of variables, $b_i \cup \{i\},\ i = 1, 2, \cdots, 40$. Arranging these 40 subsets so that those subsets which share more variables among themselves are put together yields Table 1. We may also use the mutual information or the Kullback-Leibler information divergence method (Whittaker, 1990) for this grouping.

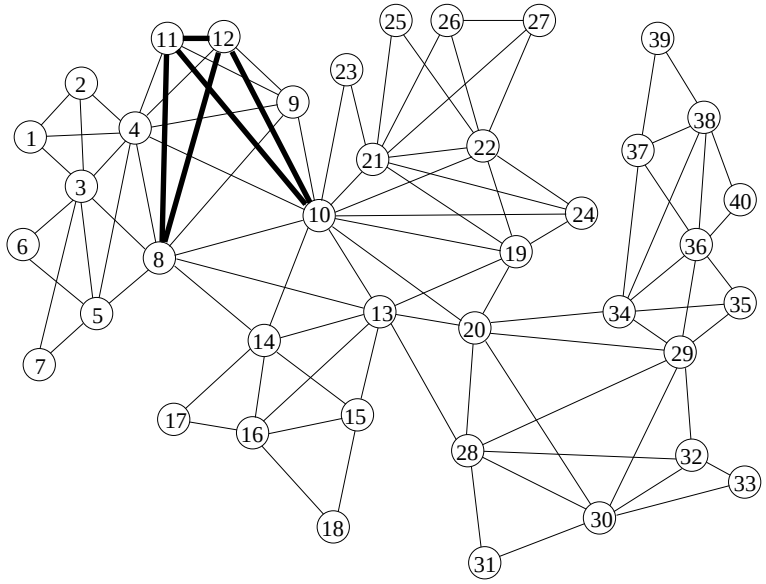

Fig. 2. The true model, without the thick edges, used for application. The thick edges which are additional to the true model are resulted from the combining process.

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

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

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

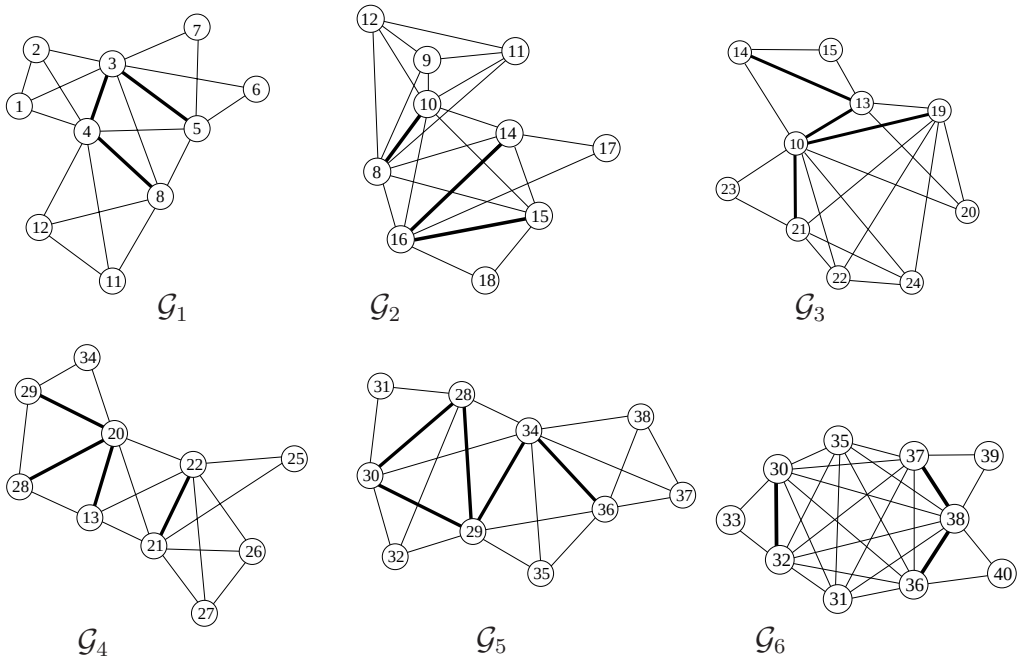

Fig. 3. Marginal models of the model in Figure 2 for the 6 subsets of variables which are listed in Table 1. $\mathcal{G}_i$ is the decomposable log-linear model for subset V_i. PSs are represented by thick lines.

As noted in 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. The marginal models corresponding to the 6 sets of variables are given in Figure 3. We obtained these models by applying the backward deletion method starting from an all k-way interaction model whose goodness-of-fit level was acceptable. The p-values of the goodness-of-fit tests of the models in the figure are all larger than or equal to 0.199 as shown in Table 2.

The true model structure in Figure 2 is fully recovered in the maximal CMS except the 5 thick edges. These additional edges were created because X_4 were missing in V_2. Also see $\mathcal{G}_2$ in Figure 3. If X_4 had been added to V_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 Concluding Remarks

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. When two marginal models, $\mathcal{G}_1 = (V_1,\ E_1)$ and $\mathcal{G}_2 =$

$(V_2,\ E_2)$, do not share any variables, the overall time complexity of this method is of order $|V_1|^2(O(|V_2| + |E_2|) + |V_2|^2(O(|V_1| + |E_1|)$ (Tarjan, 1972). When the two models share nodes, the time complexity can be expressed by $|\widetilde{V}_1|^2(O(|\widetilde{V}_2| + |\widetilde{E}_2|) + |\widetilde{V}_2|^2(O(|\widetilde{V}_1| + |\widetilde{E}_1|)$ where $\widetilde{\mathcal{G}}_i = (\widetilde{V}_i,\ \widetilde{E}_i) = (\mathcal{G}_i)_{V_1 \cap V_2}$, $i = 1, 2$.

The PS junction property (10) and the Separateness condition are instrumental for locating PSs in model-combination. When more than one combined model are created, we can reduce the ambiguity or the number of combined models by modelling for a marginal set of the variables which are involved in the ambiguity or by examining which of the variables that are involved in the ambiguity belong to which marginal models.

The proposed method is developed for combining decomposable marginal models but it can be extended to combining general forms of undirected graphs provided that prime subgraphs (Olesen and Madsen, 2002) of the marginal models may be transformed into cliques. Since the theorems in section 3 hold for a variety of distributions, the proposed method is not limited to contingency table data only. When combining large marginal models, we may use parts of them which are useful for the combining including those that are shared by the models, which will reduce the time complexity of the proposed method.

References

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

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* Cambridge, MA: The MIT Press.

Kim, S.-H. (2006). 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.

Olesen, K.G. and Madsen, A.L. (2002). Maximal prime subgraph decomposition of Bayesian networks, *IEEE Trans. Systems, Man and Cybernetics, Part B: Cybernetics*, **32**, 21-31.

Pearl, J. (1988). *Probabilistic Reasoning In Intelligent Systems: Networks of Plausible Inference,* San Mateo, CA.: Morgan Kaufmann.

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

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