Learning model structures based on marginal model structures of undirected graphs

by

Sung-Ho Kim

BK21 Research Report

09 - 04

March 25, 2009

DEPARTMENT OF MATHEMATICAL SCIENCES

Learning model structures based on marginal model structures of undirected graphs

BY SUNG-HO KIM
Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology
Daejeon, South Korea
sung-ho.kim@kaist.edu

SUMMARY

Suppose that we are given a set $\mathcal{M}$ of model structures for different sets of random variables and that we want to use these model strictures in search of a model $\mathcal{H}$ for the set of random variables that are involved in at least one of the models in $\mathcal{M}$. Assuming that the true model structure is of an undirected graph, we investigate the relationship between a model and its marginal model and then derive a rule of using the information that are found in the models in $\mathcal{M}$ for building $\mathcal{H}$. Under the assumption that the models in $\mathcal{M}$ are from the same data, we propose a method of combining models in $\mathcal{M}$ and illustrate it through examples.

Some key words: Combined model structure; Graphical compatibility; Graph-separateness; Interaction graph; Markovian subgraph; Prime separator; Self-connected separator.

1 INTRODUCTION AND PROBLEM

Suppose that we are given a set of models whose model structures are in the form of undirected graphs and that we want to build a model of the variables that are involved in at least one of the models in the given set. We may not have enough observed data for all the variables that we are interested in or we may have observed data for only a part of the variables in addition to pieces of information about the models for some subsets of variables. For example, suppose that we are given two model structures of graphical models (Whittaker, 1990; Lauritzen, 1996), $\mathcal{G}_1$ and $\mathcal{G}_2$, which are depicted in Figure 1. $\mathcal{G}_1$ is of variables X_1, X_2, X_3 and $\mathcal{G}_2$ of variables X_1, X_3, X_4. X_1 and X_3 are shared by the two graphs. If they are marginal models from a graphical model $\mathcal{H}$, it is necessary that the inter-relationships among each of the sets $\{X_1, X_2, X_3\}$ and $\{X_1, X_3, X_4\}$, as indicated in the two graphs in the figure, are satisfied in $\mathcal{H}$. In this respect, it is imperative that, given a set $\mathcal{M}$ of model structures of graphical models, we check whether the model structures are compatible with each other among themselves before exploring existence of a model structure of which all of the model structures in $\mathcal{M}$ are marginal models. We aim to address this issue under the assumption that the model structures of all the graphical models that are considered in this paper are of undirected graphs.

Fienberg & 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

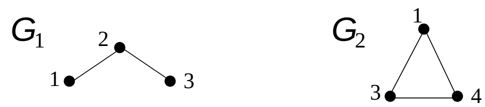

Figure 1: Two model structures of graphical models.

& Kim (1999)]. 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 marginal distributions under the assumption that the joint (as against marginal) probability model is graphical and decomposable (Kim, 2006b).

In addressing the issue of information reuse in the form 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 for decomposable graphs and self-connected (SC) separators for non-decomposable graphs and defined in sections 2 and 4 respectively.

It is shown that if we are given a graphical model with its independence graph, $\mathcal{G}$, and some of its marginal models, then we can find, under the assumption that the graphical model is decomposable, 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. This graph-separateness is substantiated by the prime separators and SC-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 $\mathcal{H}$ when the graphs are decomposable (Kim & Lee, 2008). In this paper, we will extend the graphical combination of marginal models to the case where the marginal models are not necessarily decomposable.

There have been a number of papers applying marginal models for data analysis during the last 15 years or so. Some of the applications are for parameter estimation of a model based on medical data from crossover experiments (Balagtas, Becker & Lang, 1995), for estimating joint probabilities by applying the iterative proportional fitting technique (Molenberghs & Lesaffre (1999), for analyzing sociological data (Becker, 1994; Becker, Minick & Yang, 1998), and for analyzing contingency table data with ordinal response variables (Colombi & Forcina, 2001). In most of these applications of marginal models to multivariate statistical problems, we impose structural restrictions on certain subsets of variables that are involved in a given data set (Liang, Zeger & Qaqish, 1992; Glonek & McCullagh, 1995; Bergsma, 1997; Bartolucci & Forcina, 2002; Bergsma & Rudas, 2002; Rudas & Bergsma, 2004). There also have been remarkable improvements in learning graphical models in the form of a Bayesian network (Pearl 1986, 1988; Meek, 1995; Spirtes, Glymour & Scheines, 2000; Neapolitan, 2004) from data. This learning however is mainly instrumented by heuristic searching algorithms since the model searching is usually NP-hard (Chickering, 1996). In our proposed method of structural learning, we will assume that only the information which is embedded in a given set of marginal model structures is available.

This paper is organized in 7 sections. After introducing graphical terminologies and notation in section 2, we derive a result which shows how two sets of graphs, where the graphs in one set are some type of subgraphs of the graphs in the other set, are related in stochastic context and introduce a type of graph, called a combined model structure (CMS), with regard to the relationship of the two sets of graphs. In section 4, we consider some types of separators of undirected graphs, called a self-connected separator and a prime separator, and use them to further investigate the relationship between the two sets of graphs. We then consider the notion of graphical compatibility (Dawid & Studeny, 1999) in section 5 as a necessary relationship between graphs and show existence of a CMS of set of graphs when the compatibility condition is satisfied among the graphs. In section 6, we propose a combination method of graphs as a way of information reuse from a given set of marginal models. Finally, in section 7, we close the paper with some discussion and concluding remarks.

2 NOTATION AND PRELIMINARIES

We will consider only undirected graphs in the paper. We denote a graph by $\mathcal{G} = (V, E)$, where V is the set of the indexes of the variables involved in $\mathcal{G}$ and E is a collection of ordered pairs, each pair representing that the nodes of the pair are connected by an edge. Since $\mathcal{G}$ is undirected, that (u, v) is in E is the same as that (v, u) is in E. If $(u, v) \in E$, we say that u is a neighbor node of or adjacent to v or vice versa. We say that a set of nodes of $\mathcal{G}$ forms a complete subgraph of $\mathcal{G}$ if every pair of nodes in the set is adjacent to each other. If every node in A is adjacent to all the nodes in B, we will say that A is adjacent to B. The set of all the neighbor nodes of a node v in $\mathcal{G}$ is denoted by $bd_{\mathcal{G}}(v)$; if v becomes a set, A say, we define $bd_{\mathcal{G}}(A) = \cup_{v\in A} bd_{\mathcal{G}}(v) \setminus A$. We define the closure of a set A as $cl_{\mathcal{G}}(A) = bd_{\mathcal{G}}(A) \cup A$. 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

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

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 $\langle A|C|B\rangle_{\mathcal{G}}$. The complement of a set A is denoted by A^c and the cardinality of a set A by $|A|$. For two collection of sets, A and B, we write $A \preceq B$ if, for every set a in A, there exists a set b in B such that $a \subseteq b$.

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

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 $\mathcal{G}' \subseteq_e \mathcal{G}$. If $\mathcal{G}'$ is a subgraph of $\mathcal{G}$, we call $\mathcal{G}$ a supergraph of $\mathcal{G}'$. For a graph $\mathcal{G}$, we will denote the set of nodes of $\mathcal{G}$ by $V(\mathcal{G})$.

The cliques are elementary graphical components and so we will call the intersection of neighboring cliques a prime separator of the decomposable graph $\mathcal{G}$. The prime separators in a decomposable graph may be extended to separators of prime graphs in some graphs, where the prime graphs are defined as the maximal subgraphs without a complete separator in Cox & Wermuth (1999). If $\mathcal{G}$ is not decomposable, separators are not obtained as intersections of neighboring cliques.

3 Marginal models and Markovian subgraphs

Suppose that we are given a probability model, P, with its interaction graph, $\mathcal{G}$, and some of its marginal models are also given whose interaction graphs are Markovian subgraphs of $\mathcal{G}$. In this section, we will look into the relationship between P and the marginal models through the relationship between $\mathcal{G}$ and the Markovian subgraphs.

A distribution P is said to be *globally Markov* with respect to a graph $\mathcal{G}$ if, for a triple (A, B, C) of disjoint subsets A, B, C of V, random vectors X_A and X_B are conditionally independent given an outcome of random vector X_C whenever A is separated from B by C in $\mathcal{G}$.

In addition to the global Markov property, we will consider another property for a probability distribution. A distribution P with probability function f is said to be *factorized* (FA) according to $\mathcal{G}$ [Section 3.2, Lauritzen (1996)] if for all $c \in \mathcal{C}(\mathcal{G})$ there exist non-negative functions ψ_c that depend on x through x_c only such that

$$f(x) = \prod_{c \in \mathcal{C}(\mathcal{G})} \psi_c(x).$$

We will denote the collection of the distributions that are globally Markov with respect to $\mathcal{G}$ by $M_G(\mathcal{G})$.

For a probability distribution P of 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,

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

It is well known in literature (Pearl & Paz, 1987) that if a probability distribution on 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

$$(FA) \Longrightarrow (GM) \Longrightarrow (LM) \Longrightarrow (PM) \tag{1}$$

[see Proposition 3.8 in Lauritzen (1996)]. Under the positivity condition of the probability distribution, we have $(FA) \Longleftrightarrow (PM)$ by Hammersley and Clifford (1971). From this and expression (1), it follows, under the positivity condition, that

$$(FA) \Longleftrightarrow (GM).$$

For notation convenience, 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

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

Let P_A be the marginal of P on X_A. We then define $\bar{\Gamma}(P_A)$ as

$$\bar{\Gamma}(P_A) = (\Gamma(P) \cap A) \cup \beta(\mathcal{J}_A). \tag{2}$$

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 (2).

THEOREM 1. For a distribution P of 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 (2) 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 1. For a distribution P of X_V and $A \subseteq V$,

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

From Theorem 1, 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 2. For a distribution P of X_V and $A \subseteq V$, suppose that $P \in M(\mathcal{G})$ for a 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 1, we have the desired result. □

We call $\mathcal{G}_A$ a Markovian subgraph of $\mathcal{G}$ in the context of Corollary 2.

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 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 X_V whose marginal P_A on X_A is Markov with respect to $\mathcal{G}_A$.

By definition and Corollary 2, we have the following:

$$\begin{aligned} L(\mathcal{G}) &= M(\mathcal{G}), \\ M(\mathcal{G}) &\subseteq L(\mathcal{G}_A), \qquad \text{(by Corollary 2)} \\ P \in L(\mathcal{G}_A) &\Longleftrightarrow P_A \in M(\mathcal{G}_A) \end{aligned} \tag{3}$$

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 2, it follows that $Q \in M(\mathcal{G}_A)$.

It follows from (3) 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 set 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 set of the distributions each of whose marginals is Markov with respect to its corresponding Markovian subgraph of $\mathcal{G}$.

THEOREM 2. *For a collection $\mathcal{V}$ of subsets of V with a 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 (3), $P \in L(\mathcal{G}_A)$ 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 2 shows the relationship between a graphical model with its graph $\mathcal{G}$ and a set 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 graphical 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 graphical model. In this context, we may refer to a Markovian subgraph as a marginal model structure. For simplicity, we write $\mathcal{G}_i = \mathcal{G}_{V_i}$.

DEFINITION 1. *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) of $\mathcal{G}_1, \cdots, \mathcal{G}_m$, 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 *of $\mathcal{G}_1, \cdots, \mathcal{G}_m$ if adding any edge to $\mathcal{H}$ invalidates condition (ii) for at least one $i = 1, \cdots, m$.*

Let $\mathcal{M}$ be the collection of $\mathcal{G}(P)_A,\ \ A \in \mathcal{V}$. We can construct a maximal CMS, $\mathcal{H}^*$ say, by adding edges, if any, to $\mathcal{G}$ in such a way that condition (ii) of Definition 1 is satisfied. Since

$$\mathcal{H}^*_A = \mathcal{G}(P)_A,$$

if we put $\mathcal{G} = \mathcal{G}(P)$ in Theorem 2, we end up with a summarizing expression,

$$M(\mathcal{G}(P)) \subseteq M(\mathcal{H}^*) \subseteq \tilde{L}\left(\mathcal{G}(P)_A,\ A \in \mathcal{V}\right), \tag{4}$$

where the first inequality follows since $\mathcal{G}(P) \subseteq_e \mathcal{H}^*$. Since $P \in M(\mathcal{G}(P))$, expression (4) implies that P is also Markov relative to $\mathcal{H}^*$.

If two nodes u and v are separated in a Markovian subgraph of a graph $\mathcal{G}$, then so are they in $\mathcal{G}$ by the property of a graph. We can extend this result to disjoint sets. Let $\mathcal{C}_{\mathcal{G}}(A)$ denote the collection of the cliques which include nodes of A in $\mathcal{G}$.

THEOREM 3. (Theorem 4.2, Kim & Lee (2008)) *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 $W' \in \mathcal{C}_{\mathcal{G}}(C)$, $\langle W|B|W'\rangle_{\mathcal{G}}$.*

4 MARKOVIAN SUBGRAPHS OF UNDIRECTED GRAPHS

Consider two Markovian subgraphs of $\mathcal{G}$ on A and B, $\mathcal{G}_A$ and $\mathcal{G}_B$. Then by the transitivity property of the Markovian marginalization (Kim, 2006b), $(\mathcal{G}_A)_B$ is also a Markovian subgraph of $\mathcal{G}$, and similarly for $(\mathcal{G}_B)_A$. Furthermore, we can see, by definition, that

$$(\mathcal{G}_A)_B = (\mathcal{G}_B)_A = \mathcal{G}_{A\cap B}. \tag{5}$$

DEFINITION 2. *For three disjoint and exhaustive subsets, A, B, and C, of $V = V(\mathcal{G})$, we will call C a self-connected (SC) separator in $\mathcal{G}$, if the following conditions hold:*

(i) $\langle A|C|B\rangle_{\mathcal{G}}$.

(ii) $\mathcal{G}_C^{ind}$ *is connected.*

(iii) $\mathcal{G}_C^{ind}$ *does not contain any n-cycle ($n > 3$) nor a clique of $\mathcal{G}$ which consists of more than two nodes.*

(iv) $\mathcal{G}_{A\cup C}^{ind}$ *and* $\mathcal{G}_{B\cup C}^{ind}$ *each consists of n-cycles ($n > 3$) or cliques of $\mathcal{G}$ only.*

According to the definition, we can see that, a SC-separator is given as a union of some intersections of n-cycles ($n > 3$) or cliques. For example, in Figure 2, the SC-separator, $\{1, 2, 4, 5, 6\}$, is the union of the intersections of the following two pairs of cycles:

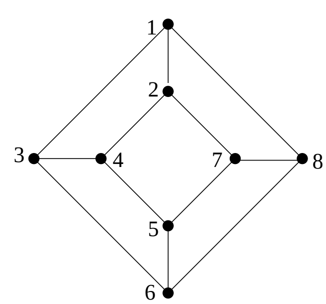

Figure 2: An undirected graph of 8 nodes.

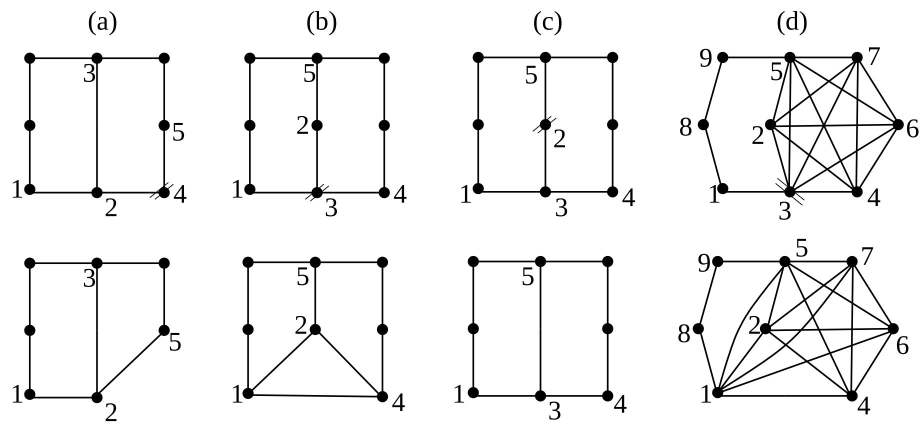

Figure 3: Undirected graphs and their Markovian subgraphs. The slant "=" on node v means that v is removed from the graph.

pair 1 : $\{1,2,4,5,6,8\}$ and $\{1,2,3,4\}$
pair 2 : $\{1,2,4,5,6,8\}$ and $\{3,4,5,6\}$

In this respect, it follows that, if $\mathcal{G}$ is a decomposable graph, then all of its SC-separators are prime separators.

Note that condition (iv) in the definition does not imply condition (ii). For example, if, in Figure 2, we let $A = \{3,4\}$, $B = \{7,8\}$, and $C = \{1,2,5,6\}$, then conditions (i), (iii), (iv) are satisfied for the three sets, but (ii) is not.

If $\mathcal{G}$ is Markovian-marginalized over a node v which is included in a SC-separator of the graph $\mathcal{G}$, then new SC-separators are created in $\mathcal{G}_{V\setminus\{v\}}$. For example, in column (b) of Figure 3, node 3 is removed from the graph at the top and the resultant Markovian subgraph is given at the bottom, where $\{1,2,5\}$ and $\{2,4,5\}$ are new SC-separators. In column (c), node 2 is removed and the removal yields a new SC-separator, $\{3,5\}$, in the Markovian subgraph. In column (d), the set $\{3,5\}$ is a SC-separator and removal of node 3 creates a new SC-separator $\{1,5\}$. Note that $\{1,3,5,8,9\}$ forms a 5-cycle and $\{2,3,4,5,6,7\}$ forms a clique and that the removal creates a new clique, $\{1,2,4,5,6,7\}$, and a new SC-separator $\{1,5\}$. On the other hand, removal of a node which is not a member of a SC-separator does not create any new SC-separator as we see in column (a) of Figure 3.

Let M and S be, respectively, a set of nodes to be removed and a set of nodes which form SC-separators. Since Markovian subgraphs are not dependent upon the order of node-removal, we can begin node-removal with the nodes in $M \cap S$ or with the nodes in $M \setminus S$. The only difference is that the removal of a node in $M \setminus S$ simply reduces the size of a cycle or a clique while the removal of a node in $M \cap S$ creates new SC-separators.

THEOREM 4. *Let $\mathcal{G}'$ be a Markovian subgraph of an undirected graph $\mathcal{G}$. If A is a SC-separator in $\mathcal{G}'$, then there exists a SC-separator, S, in $\mathcal{G}$ such that $A \cap S \neq \emptyset$.*

Proof. Since A is a SC-separator in $\mathcal{G}'$, we can find disjoint sets, B and C, in $V(\mathcal{G}') \setminus A$ such that $A \cup B \cup C = V(\mathcal{G}')$ and $\langle B|A|C\rangle_{\mathcal{G}'}$. Then, by Theorem 3, it follows that $\langle B|A|C\rangle_{\mathcal{G}}$. Let $D = V(\mathcal{G}) \setminus V(\mathcal{G}')$. Then, by the property of an undirected graph, we have $\langle B|A \cup D|C\rangle_{\mathcal{G}}$. Now, we have only to show (i) that the set $A \cup D$ is itself a SC-separator or (ii) that $A \cup D$ contains a

SC-separator as a subset in $\mathcal{G}$.

In case (i), we have the desired result. In case (ii), there are two possibilities. One possibility is that there is a SC-separator A' in $A \cup D$ such that $A \subseteq A'$, and the other that A is itself a SC-separator in $\mathcal{G}$. In the former situation, at least one node is removed from A' in the marginalization of $\mathcal{G}$, and in the latter situation the removal takes place outside the neighborhood of A in $\mathcal{G}$. In the latter situation, A itself is a SC-separator in $\mathcal{G}$; and in the former situation, if node $v \in A'$ is removed from $\mathcal{G}$, all of its neighbor nodes become adjacent to each other, which means that new SC-separators are created in $\mathcal{G}_{v^c}$ as in panels (b), (c), and (d) in Figure 3, where $v^c = V(\mathcal{G}) \setminus \{v\}$. If multiple nodes, $v_1, \cdots, v_r$, are removed from $A' \setminus A$, we can see by the same argument that we have new SC-separators in $\mathcal{G}_R$ where $R = V(\mathcal{G}) \setminus \{v_1, \cdots, v_r\}$. This completes the proof. □

From this theorem, we can see that a SC-separator, S say, in a Markovian subgraph of $\mathcal{G}$ means that there is a SC-separator in $\mathcal{G}$ which shares at least one node with S. An analogous but more tangible result holds when $\mathcal{G}$ is decomposable. In the theorem below, $\chi(\mathcal{G})$ is the set of all the prime separators of a decomposable graph $\mathcal{G}$.

THEOREM 5. (Theorem 4 of Kim (2006b)) 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 $\mathcal{H}$,
$$\cup_{i=1}^{m} \chi(\mathcal{G}_i) = \chi(\mathcal{H}).$$

The above two theorems say that,
(a) when $\mathcal{G}$ is decomposable, every prime separator that is found in a Markovian subgraph of $\mathcal{G}$ is also found in $\mathcal{G}$; but
(b) when $\mathcal{G}$ is not decomposable, every SC-separator, A say, that is found in a Markovian subgraph of $\mathcal{G}$ has at least one SC-separator in $\mathcal{G}$ which shares at least one node with A.

There is another noteworthy difference between the two types of graphs. In a decomposable graph, if a node which is included in a prime separator is removed, then a new clique is formed by the nodes of the cliques that share the prime separator. This means that the prime separator disappears with no trace left. On the other hand, if a node which is included in a SC-separator is removed from a non-decomposable graph, then new SC-separators are created as shown in Figure 3 unless the SC-separator is shared by neighboring cliques only. From this, we can see that prime separators in a decomposable graph may easily be lost in its Markovian subgraphs when at least one of the nodes in a prime separator is removed. On the other hand, node-removal from a SC-separator, S say, in a non-decomposable graph create new SC-separators in the Markovian subgraph which share nodes with S if the number of the removed nodes from S is less than $|S|$.

5 GRAPHICAL COMPATIBILITY AND EXISTENCE OF CMS'S

For $C \subseteq V(\mathcal{G}) \cap V(\mathcal{H})$, $\mathcal{G}$ and $\mathcal{H}$ are said to be C-compatible (Dawid & Studeny, 1999) if

$$\mathcal{G}_C = \mathcal{H}_C.$$

For graphs, $\mathcal{G}_1, \cdots, \mathcal{G}_k$, and sets of nodes, $A_1, \cdots, A_{k-1}$, if $\mathcal{G}_i$ and $\mathcal{G}_{i+1}$ are A_i-compatible for $i = 1, 2, \cdots, k-1$, then we say that $\mathcal{G}_1$ and $\mathcal{G}_k$ are compatible with regard to $\mathcal{G}_i$, $i = 2, \cdots, k-1$.

Denote $V(\mathcal{G}_i)$ by V_i and suppose that $(\mathcal{G}_1)_{A_1} = (\mathcal{G}_2)_{A_1}$ and $(\mathcal{G}_2)_{A_2} = (\mathcal{G}_3)_{A_2}$. If there exist graphs, $\mathcal{G}'$ and $\mathcal{G}''$, such that

$$\mathcal{G}'_{V_1} = \mathcal{G}_1 \text{ and } \mathcal{G}'_{V_2} = \mathcal{G}_2$$

and

$$\mathcal{G}''_{V_2} = \mathcal{G}_2 \text{ and } \mathcal{G}''_{V_3} = \mathcal{G}_3,$$

then it follows that $\mathcal{G}'_{A_2} = \mathcal{G}''_{A_2}$ since

$$\mathcal{G}'_{A_2} = (\mathcal{G}'_{V_2})_{A_2} = \mathcal{G}^2_{A_2} = \mathcal{G}^3_{A_2} = (\mathcal{G}''_{V_3})_{A_2} = \mathcal{G}''_{A_2},$$

where the first and the last equality hold by (5) and the inequality, $A_2 \subseteq V_2 \cap V_3$.

If we assume that $\mathcal{G}_1, \cdots, \mathcal{G}_k$ are Markovian subgraphs of an undirected graph $\mathcal{G}$, then there must exist such graphs as $\mathcal{G}'$ and $\mathcal{G}''$ for every pair of Markovian subgraphs that share at least one node.

Let $\langle i \rangle = \{1, 2, \cdots, i\}$. Suppose that $V_i \cap V_{i+1} \neq \emptyset$ for $i = 1, 2, \cdots, k-1$ and that we have a graph $\mathcal{G}^{\langle j \rangle}$ for $1 < j < k$ whose Markovian subgraphs are $\mathcal{G}_i$, $i = 1, 2, \cdots, j$. Then there must exist $\mathcal{G}^{\langle j+1 \rangle}$ of which $\mathcal{G}^{\langle j \rangle}$ and $\mathcal{G}_{j+1}$ are Markovian subgraphs. Otherwise, the assumption for $\mathcal{G}$ becomes invalid. We state this in a formal manner below.

THEOREM 6. *If two graphs, $\mathcal{G}$ and $\mathcal{H}$, are C-compatible for $C = V(\mathcal{G}) \cap V(\mathcal{H})$, then there exists a CMS of $\mathcal{G}$ and $\mathcal{H}$.*

Proof. When $\mathcal{G} = \mathcal{H}$, the result is trivial since a graph is a CMS of itself. Suppose that $|V(\mathcal{G}) \setminus V(\mathcal{H})| = 1$. Then we can construct a graph $\mathcal{H}^1$ of which $\mathcal{G}$ and $\mathcal{H}$ are Markovian subgraphs as described below.

Let $\{\alpha\} = V(\mathcal{G}) \setminus V(\mathcal{H})$. Then we can think of the following three cases:

(*i*) $bd_{\mathcal{G}}(\alpha) = C$ and there exists a connectivity component g in $\mathcal{H}$ for which $C \subseteq cl_{\mathcal{H}}(g)$.

(*ii*) $bd_{\mathcal{G}}(\alpha) = C$ and there does not exist any connectivity component as in (i) but a connectivity component g' for which $\emptyset \subset cl_{\mathcal{H}}(g') \cap C \subset C$.

(*iii*) $bd_{\mathcal{G}}(\alpha) \subset C$, i.e., $\langle \alpha | bd_{\mathcal{G}}(\alpha) | C \setminus cl_{\mathcal{G}}(\alpha) \rangle_{\mathcal{G}}$.

In case (*i*): In this case, $C \subseteq cl_{\mathcal{H}}(g)$. So node α may be attached to any clique in $\mathcal{H}$ that is connected to C in $\mathcal{H}$, in such a way that $h \cup \{\alpha\}$ may form a new clique in $\mathcal{H}^1$.

In case (*ii*): In this case, α is attached to $\mathcal{H}$ to form $\mathcal{H}^1$ such that $bd_{\mathcal{H}^1}(\alpha) = C$.

In case (*iii*): If there exists a connectivity component g in $\mathcal{H}$ such that $bd_{\mathcal{G}}(\alpha) \subseteq cl_{\mathcal{H}}(g)$ and $C \setminus bd_{\mathcal{G}}(\alpha) = C \setminus cl_{\mathcal{H}}(g)$, then α can be attached to any clique in $\mathcal{H}$ that is connected to $bd_{\mathcal{G}}(\alpha)$ to form a new clique in $\mathcal{H}^1$. If there is no such a connectivity component in $\mathcal{H}$, then we attach α to $\mathcal{H}$ such that $bd_{\mathcal{H}^1}(\alpha) = C$.

Now suppose that $|V(\mathcal{G}) \setminus V(\mathcal{H})| > 1$. Let $\mathcal{G}_0 = \mathcal{G}_C$ and $V(\mathcal{G}) \setminus V(\mathcal{H}) = \{\alpha_1, \cdots, \alpha_k\}$. Let $C_i = C \cup \{\alpha_1, \cdots, \alpha_i\}$ and $\mathcal{G}_i = \mathcal{G}_{C_i}$ for $i = 1, \cdots, k$, where $\mathcal{G}_i \subseteq_M \mathcal{G}_{i+1}$, $i = 0, 1, \cdots, k-1$. Then, by the transitivity property of Markovian subgraphs (Theorem 6 in Kim (2006b)), we have $\mathcal{G}_i \subseteq_M \mathcal{G}$, $i = 0, 1, \cdots, k$.

By applying the above argument, we can obtain a graph $\mathcal{H}^i$ of which $\mathcal{H}^{i-1}$ and $\mathcal{G}_i$ are Markovian subgraphs for $i = 1, \cdots, k$, where $\mathcal{H}^0 = \mathcal{H}$. By the transitivity property of Markovian subgraphs, we have $\mathcal{H} \subseteq_M \mathcal{H}^k$. Therefore, $\mathcal{H}$ and $\mathcal{G} = \mathcal{G}_k$ are Markovian subgraphs of $\mathcal{H}^k$. This completes the proof. $\square$

This theorem can be extended to a set of graphs where each graph is compatible with at least one of the other graphs of the set as shown in the following corollary.

COROLLARY 3. *For graphs $\mathcal{G}_i$, $i = 1, 2, \cdots, m$, let $\mathcal{G}^{\langle i \rangle}$ be a graph of which $\mathcal{G}_j$ is a Markovian subgraph, $j \leq i$. If $\mathcal{G}^{\langle i \rangle}$ and $\mathcal{G}_{i+1}$ are C_i-compatible with $C_i = V(\mathcal{G}^{\langle i \rangle}) \cap V(\mathcal{G}_{i+1})$, $i = 1, 2, \cdots, m-1$, then there exists a CMS of $\mathcal{G}_i$, $i = 1, 2, \cdots, m$.*

Proof. Since $\mathcal{G}^{\langle m-1 \rangle}$ and $\mathcal{G}_m$ are C_{m-1}-compatible by the condition of the corollary, there exists, by Theorem 6, a CMS, I, of $\mathcal{G}^{\langle m-1 \rangle}$ and $\mathcal{G}_m$. By the transitivity property of Markovian subgraphs, I is a CMS of $\mathcal{G}_i$, $i = 1, 2, \cdots, m$. $\square$

6 MARKOVIAN COMBINATION OF MARGINAL MODELS

In the proof of Theorem 6, we considered, to show existence of a CMS, how we can add an edge between a node in $V(\mathcal{G}) \setminus V(\mathcal{H})$ and another node in $\mathcal{H}$ with no confliction with the node-separateness that is found in at least one of the graphs. The two graphs in Figure 1 are $\{1, 3\}$-compatible and their CMS's are as in Figure 4. As for the two graphs in Figure 1, consider adding edges between node 4 in $V(\mathcal{G}_2) \setminus V(\mathcal{G}_1)$ and some nodes in $\mathcal{G}_1$. Because of the node-separateness in $\mathcal{G}_1$, node 4 can only be adjacent to nodes 1 and 2 or to nodes 2 and 3 as in Figure 4.

Since a CMS, $\mathcal{H}$ say, of a pair of compatible graphs, $\mathcal{G}'$ and $\mathcal{G}''$ say, is obtained in the form of attaching the nodes in $V(\mathcal{G}') \setminus V(\mathcal{G}'')$ (or $V(\mathcal{G}'') \setminus V(\mathcal{G}')$) to $\mathcal{G}''$ (or $\mathcal{G}'$), it may be regarded as combining the two graphs together. We will call this combination a *Markovian combination* in the sense that

$$M(\mathcal{H}) \subseteq \tilde{L}(\mathcal{G}', \mathcal{G}'');$$

in other words, a probability model P which is globally Markov with respect to $\mathcal{H}$ has its marginals, $P_{V(\mathcal{G}')}$ and $P_{V(\mathcal{G}'')}$, globally Markov with respect to $\mathcal{G}'$ and $\mathcal{G}''$ respectively.

Since a maximal CMS has a better property than CMS's in the context of Theorem 5, we will propose a combination method for maximal CMS's based on a set of marginal model structures. In the combination, it is imperative that node-separateness is preserved between a graph and its Markovian subgraph. This is reflected in the combination process in such a way that the following condition is satisfied:

[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 $\mathcal{M}$.

Two main rules of Markovian combination are 'union' and 'check of separateness.' We will describe each of them below.

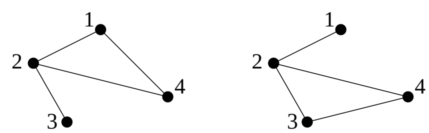

Figure 4: Two CMS's of the graphs in Figure 1.

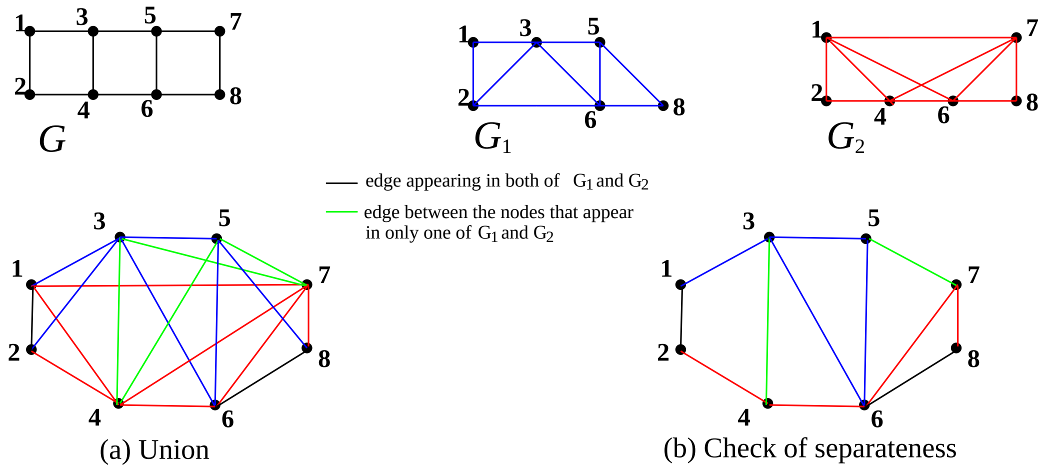

Figure 5: Markovian combination of graphs. The Markovian subgaphs $\mathcal{G}_1$ and $\mathcal{G}_2$ of $\mathcal{G}$ are combined in two steps, union and check of separateness. Different colors are used for $\mathcal{G}_1$ (in blue) and $\mathcal{G}_2$ (in red). When an edge appears in both of the graphs, it is in black; an edge is colored green when its two nodes are not in the same graph of $\mathcal{G}_1$ or $\mathcal{G}_2$.

Union. Suppose we have two Markovian subgraphs, $\mathcal{G}_1$ and $\mathcal{G}_2$, of a graph $\mathcal{G}$. If nodes u and v are not separated in any of $\mathcal{G}_1$ and $\mathcal{G}_2$, we put an edge between the two nodes. We do the same for all the pairs of nodes that are not separated in any of the subgraphs.

If two nodes are not in the same subgraph, then we put an edge between them. If two nodes are shared by $\mathcal{G}_1$ and $\mathcal{G}_2$ and they are connected by an edge in one subgraph but not in the other, we leave them separated. We denote the graph resulting from this operation by $\mathcal{G}^*$.

Check of separateness. We check if the separateness that is found in $\mathcal{G}_1$ and $\mathcal{G}_2$ holds in $\mathcal{G}^*$ also. If an edge in $\mathcal{G}^*$ is in conflict with the separateness of some pair of nodes, we remove the edge from $\mathcal{G}^*$. We denote the graph resulting from this operation by $\mathcal{G}^{**}$.

This combining process is illustrated in Figure 5. Note that in panel (a), edges, (3,4), (3,7), (4,5), (5,7), are created since the nodes in each of the pairs are not in the same graph of $\mathcal{G}_1$ or $\mathcal{G}_2$. Two of the edges are removed in panel (b) since their existence are in conflict with the node-separateness that is embedded in $\mathcal{G}_1$ and $\mathcal{G}_2$. The combined result contains two edges more than the true graph $\mathcal{G}$ in Figure 5. It is interesting to note in this figure that $\mathcal{G}_1$ and $\mathcal{G}_2$ are decomposable while none of $\mathcal{G}$ and the combined graph is. This is an example that the Markovian combination of decomposable graphs does not necessarily produce a decomposable combined graph.

Another illustration is given in Figure 6 where the graph $\mathcal{G}$ is not a chain of cycles as in Figure 5 but a more general form of undirected graphs. A 4-cycle $\{3, 4, 7, 8\}$ is surrounded by a 7-cycle $\{1, 2, 5, 6, 9, 10, 11\}$ in the graph. The combined graph which appears in panel (b) contains all the edges in the graph $\mathcal{G}$ in addition to the edges (1,4), (1,7), (5,7), and (7,10). These four edges appeared in $\mathcal{G}_1$ or $\mathcal{G}_2$ and are not in conflict with any node-separateness that is found in $\mathcal{G}_1$ and $\mathcal{G}_2$. Note, in Figures 5 and 6, that the black edges in panel (a) which appear in both of the graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$ in each of the figures, are preserved in the combined graph in panel (b). This is consequential on the fact that the adjacency of a pair of nodes in both of $\mathcal{G}_1$ and $\mathcal{G}_2$ is in no conflict with the

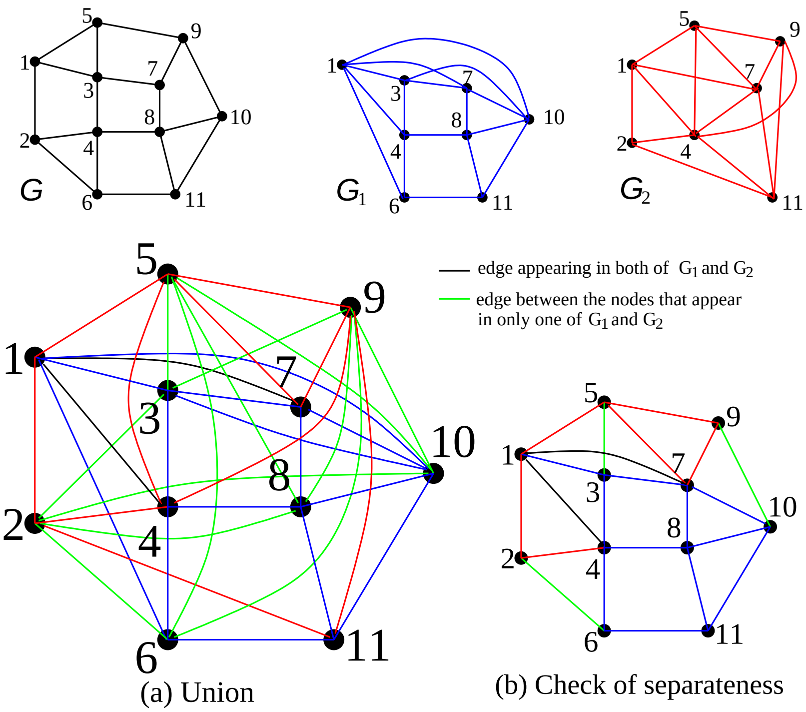

Figure 6: Markovian combination of graphs. In panel (a), there are 12 green edges for the pairs of nodes that do not appear in the same graph of $\mathcal{G}_1$ or $\mathcal{G}_2$. Three of the green edges remain in panel (b).

node-separateness in both of the graphs.

The combined graphs which are obtained through the two operations are maximal CMS's of a given set of Markovian subgraphs as shown in the theorem below.

THEOREM 7. *The combination process by the two operations of Union and Check of separateness produces a maximal CMS.*

Proof. Let $\mathcal{M}$ be a set of Markovian subgraphs of a graph. The "Union" operation puts an edge between a pair of nodes, u and v say, unless u and v are both in a graph in $\mathcal{M}$ and separated therein. Denote the graph from this operation by $\mathcal{G}^*$. It is obvious that $\mathcal{G}' \subseteq_e \mathcal{G}^*_{V(\mathcal{G}')}$ for every $\mathcal{G}' \in \mathcal{M}$.

The "Check of separateness" operation removes edges from $\mathcal{G}^*$ in such a way that the following condition is satisfied for every $\mathcal{G}'$ in $\mathcal{M}$:

For any pair of non-adjacent nodes u and v in $\mathcal{G}'$ and a set C in $\mathcal{G}'$ which is disjoint with $\{u, v\}$,

$$\langle u|C|v\rangle_{\mathcal{G}'} \text{ if and only if } \langle u|C|v\rangle_{\mathcal{G}^*}. \tag{6}$$

Denote a graph obtained from this check-of-separateness by $\mathcal{G}^{**}$. Then any pair of non-adjacent nodes, u and v say, in $\mathcal{G}^{**}$ mean either (*i*) that they are non-adjacent in at least one of the graphs in $\mathcal{M}$ or (*ii*) that they belong to different graphs each and putting an edge between the nodes incurs conflicts with the node-separateness in some of the graphs in $\mathcal{M}$.

Therefore, adding any edge to $\mathcal{G}^{**}$ into another graph $\mathcal{G}''$ disqualifies $\mathcal{G}''$ as a CMS of the graphs in $\mathcal{M}$. This means that $\mathcal{G}''$ is a maximal CMS of the graphs in $\mathcal{M}$. □

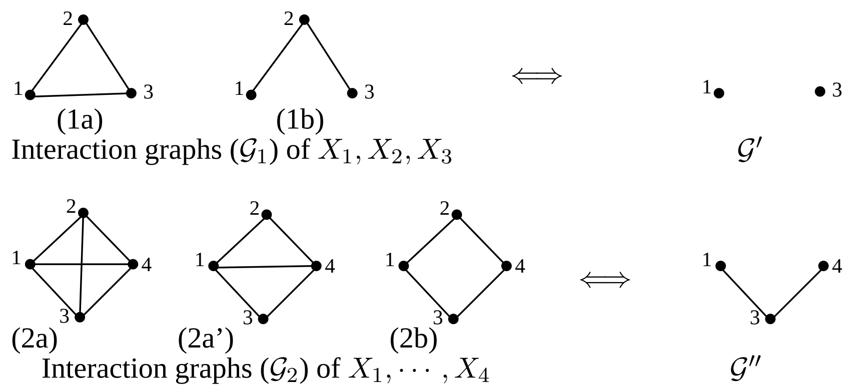

Figure 7: Some simple examples where each of the graphs in the right column are not a Markovian subgraph of any of the graphs on the left-hand side of $\Longleftrightarrow$.

7 Further discussion

In Theorem 2, we are given a set of Markovian subgraphs of $\mathcal{G}$. But in reality, we are often given a set of marginal model structures that are assumed to be interaction graphs of the marginal models. The interaction graphs may not be Markovian subgraphs of the unknown $\mathcal{G}$. In this case, maximal CMS's may not contain $\mathcal{G}$ as an edge-subgraph. Simple examples of this situation are displayed in Figure 7. In the first row of the figure are two interaction graphs ($\mathcal{G}_1$) for X_1, X_2, X_3 and a subgraph $\mathcal{G}'$ which is not Markovian with respect to $\mathcal{G}_1$, and similarly in the second row for $X_1, \cdots, X_4$. Under the hierarchy assumption for contingency tables, none of the graphical log-linear models (1a), (2a), and (2a') is compatible with the graphical submodels at the right ends of the corresponding rows by Theorem 2.3 of Asmussen and Edwards (1983). The model $\mathcal{G}'$ in Figure 7 is possible with the graphical log-linear model (1b) in the figure when

$$E[(P(X_{\{1,3\}} = x_{\{1,3\}}|X_2)] = P(X_1 = x_1)P(X_3 = x_3) \quad \text{for all} \quad x_{\{1,3\}} \in \mathcal{X}_{\{1,3\}}, \tag{7}$$

where $\mathcal{X}_i$ is the support of X_i and $\mathcal{X}_a = \prod_{i \in a} \mathcal{X}_i$. The graphical log-linear model $\mathcal{G}''$ in Figure 7 is also possible from the graphical model (2b) in the figure. Instances of this phenomenon follow.

Example 1. *Probability distributions corresponding to some of the graphs in Figure 7.* We will present contingency tables for which the pair of models, (1b) and $\mathcal{G}'$ in Figure 7, are possible and so are the pair of models, (2b) and $\mathcal{G}''$. When X_i and X_j are conditionally independent given X_k, we will simply write $i \perp\!\!\!\perp j | k$.

(a) Concerning models (1b) and $\mathcal{G}'$:

x_2	x_1	x_3	$P(X = x)$	x_2	x_1	x_3	$P(X = x)$
0	0	0	1/24	1	0	0	2/24
		1	3/24			1	6/24
	1	0	2/24		1	0	1/24
		1	6/24			1	3/24

This distribution satisfies that $1 \perp\!\!\!\perp 3 | 2$ and $1 \perp\!\!\!\perp 3$.

(b) Concerning models (2b) and $\mathcal{G}''$:

x_2	x_3	x_1	x_4	$P(X=x)$	x_2	x_3	x_1	x_4	$P(X=x)$
0	0	0	0	1/42	1	0	0	0	3/42
			1	2/42				1	1/42
		1	0	2/42			1	0	6/42
			1	4/42				1	2/42
	1	0	0	2/42		1	0	0	6/42
			1	4/42				1	2/42
		1	0	1/42			1	0	3/42
			1	2/42				1	1/42

This distribution satisfies the conditional independencies displayed in graph (2b) in Figure 7. The marginal for $X_{\{1,3,4\}}$ satisfies the conditional independence $1 \perp\!\!\!\perp 4|3$. □

Although we have seen examples where subgraphs of graphical log-linear models are not Markovian, Markovian subgraphs are usual situations under the hierarchy assumption for models. As indicated in (7), in order for a subgraph to be non-Markovian, a certain set of equations must be satisfied between the set of parameters of a joint model and that of its interested non-Markovian subgraph. This implies that non-Markovian subgraphs are a rare situation under the hierarchy assumption as long as interaction graphs are concerned. Furthermore, when the distribution is Normal, we can see by its density function that the subgraphs are Markovian. Based on this point of view on Markovian subgraphs, we have assumed in this paper that all the interaction graphs of subsets V_i of random variables are Markovian.

The combination of model structures is in two steps, Union and Check of separateness. Suppose we combine the graphs in $\mathcal{M}$. At the 'Union' step, we put an edge between every pair of nodes unless there exists at least one graph in $\mathcal{M}$ where both of the nodes appear and are not adjacent; at the 'Check of separateness' step, we then remove an edge when its existence is in conflict with the node-separateness in the graphs in $\mathcal{M}$. In this process, we don't need data but the model structures. In this sense, the proposed method reuses the information that is embedded in the marginal model structures for learning structures of a larger set of random variables which are involved in at least one of the graphs in $\mathcal{M}$.

REFERENCES

Balagtas, C.C., Becker, M.P. & Lang, J.B. (1995). Marginal modelling of categorical data from crossover experiments, *Appl. Statist.* **44**, 63-77.

Bartolucci, F. & Forcina, A. (2002). Extended RC association models allowing for order restrictions and marginal modeling, *J. Am. Statist. Assoc.* **97**, 1192-9.

Becker, M.P. (1994). Analysis of repeated categorical measurements using models for marginal distributions: an application to trends in attitudes on legalized abortion. In *Sociological Methodology*, Ed. P.V. Marsden, pp. 229-65. Oxford: Blackwell.

Becker, M.P., Minick, S. & Yang, I. (1998). Specifications of models for cross-classified counts: comparisons of the log-linear model and marginal model perspectives, *Sociological Methods and Research* **26**, 511-29.

Bergsma, W.P. (1997). *Marginal Models for Categorical Data.* Tilburg: Tilburg University Press.

Bergsma, W.P. & Rudas, T. (2002). Marginal models for categorical data, *Ann. Statist.* **30**, 140-59.

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

Colombi, R. & Forcina, A. (2001). Marginal regression models for the analysis of positive association of ordinal response variables, *Biometrika* **88**, 1007-19.

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

Dawid, A. P. & Studeny, M. (1999). Conditional products: An alternative approach to conditional independence. In *Artificial Intelligence and Statistics 99*, Eds. D. Heckerman & J. Whittaker, pp. 32-40. Morgan Kaufmann.

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

Glonek, G.J.N. & McCullagh, P. (1995). Multivariate logistic models, *J. R. Statist. Soc.* B **57**, 533-46.

Hammersley, J.M. & Clifford, P.E. (1971). Markov fields on finite graphs and lattices. *Unpublished manuscript.*

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.

Kim, S.-H. & Lee, S. (2008). Searching model structures based on marginal model structures. In *New Developments in Robotics, Automation and Control*, Eds. A. Lazinica, pp. 355-376. Vienna, Austria: In-Tech Education and Publishing.

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

Lauritzen, S.L. & 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.

Liang, K.-Y., Zeger, S. L. & Qaqish, B. (1992). Multivariate regression analyses for categorical data. *J. R. Statist. Soc.* B **54**, 3-40.

Meek, C. (1995). Causal influence and causal explanation with background knowledge, *Uncertainty in Artificial Intelligence* 11, 403-410.

Molenberghs, G. & Lesaffre, E. (1999). Marginal modelling of multivariate categorical data, *Statistics in Medicine* **18**, 2237-2255.

Neapolitan, R.E. (2004). *Learning Bayesian Networks*, Upper Saddle River, NJ: Pearson Prentice Hall.

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. & Paz, A. (1987). Graphoids: a graph based logic for reasoning about relevancy relations. In *Advances in Artificial Intelligence II*, Eds. B.D. Boulay, D. Hogg & L. Steel, pp. 357-363. Amsterdam: North-Holland, .

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd ed. New York: Springer-Verlag.

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