# Combining decomposable model-structures [1]

Sung-Ho Kim
Division of Applied Mathematics, Korea Advanced Institute of Science and Technology
Daejeon, 305-701, South Korea. e-mail: sung-ho.kim@kaist.ac.kr

**Summary**
An issue of combining marginal models is addressed under the assumption that the true model is graphically decomposable. A main theme which is instrumental throughout the paper is that decomposability is preserved between a model and its submodel and that some type of separators in the graph of a model is found in a decomposable graphical model and also in a collection of its submodels. These separators, which are called prime separators in this paper, are a guideline for model combination. A theory for the guideline is proposed showing how we may use prime separators for drawing a blueprint based on which a combined model is formed.

## 1 Introduction

Fienberg and Kim (1999) 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 therein]. 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 3.5, Edwards (1995)]. 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. This issue will be addressed in this paper.

[1] AMS 1991 subject classifications: Primary 62H05, 62M40; secondary 62A01, 62-09.
*Key words and phrases.* prime separator, conditional independence, edge-subgraph, marginal model, Markov property, Markovian subgraph, maximal combined model structure.
This work was supported by grant No. R02-2001-000-00005-0 from the Korea Science & Engineering Foundation.


Figure 1: Two marginal models

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 in Figure 1. We will use the notation $[\cdot]\cdots[\cdot]$ as used in Fienberg(1980) to represent a model. The left graph is of the model [12][23] and the right one 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, the following models are possible corresponding to Pair-1:

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

Note that we can obtain the pair in Figure 1 from each of these models and that among these four models, the first three 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:

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

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.

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 will show in this paper that the graph-separateness

is invariant between a set of marginal models and some joint model under the decomposability assumption of the model. This invariance is closely related to the statements that a decomposable graph is triangulated [Darroch, Lauritzen, Speed (1980); Leimer (1989)] and that decomposability is preserved in graph-collapsing [Theorem 4.3].

While we will consider a problem of learning decomposable graphical models from a collection of marginal models, there has 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 in this paper would lead to an improvement in learning graphical models from data.

The paper is organized in 7 sections. Section 2 introduces notation and graphical terminologies to use, and defines a type of separator of a decomposable graph which is a useful tool for combining models. Section 3 discusses properties of distributions that are Markov relative to an undirected graph and those that are also Markov relative to its subgraphs. In section 4, we derive theorems on combining decomposable models to a larger decomposable model, and this section ends with a paragraph which summarizes the main theory of the paper putting probability models and graph theory together in one expression [i.e., expression (12)]. The idea of Section 4 is illustrated in section 5 using simple models. Larger models are used in Section 6 for another illustration. Some issues are discussed further, in section 7, concerning the model-combination based on graph theory,

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

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

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

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

For a subset $A \subseteq V$, we denote by $\mathcal{G}_A = (A, E_A)$ the subgraph of $\mathcal{G} = (V, E)$ confined to $A$ where

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

In particular, we will call $\mathcal{G}_A$ the Markovian subgraph of $\mathcal{G}$ confined to $A$.

While a Markovian subgraph $\mathcal{G}_A$ of $\mathcal{G}$ is used to represent a submodel for $(X_i)_{i \in A}$, an *induced* subgraph is simply a part of $\mathcal{G}$. For $A \subset V$, an induced subgraph of $\mathcal{G}$ confined to $A$ is defined as $\mathcal{G}_A^{ind} = (A,\ E \cap (A \times A))$. For $A \subset V$, we denote by $\mathcal{J}_A$ the collection of the connectivity components in $\mathcal{G}_{A^c}^{ind}$ and let

$$\beta(\mathcal{J}_A) = \{bd(B);\ B \in \mathcal{J}_A\}. \quad (4)$$

Then $E_A$ in (3) can be expressed as

$$E_A = [E \cup \{B \times B;\ B \in \beta(\mathcal{J}_A)\}] \cap A \times A.$$

Note that, if we regard $\{B \times B;\ B \in \beta(\mathcal{J}_A)\}$ as a set of edges, the set makes the boundary of $A^c$ complete.

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 $\mathcal{G}'$.

According to the definition of a decomposable graph (see Definition A.2 in Appendix A), we can find a sequence of cliques $C_1, \cdots, C_k$ of a decomposable graph $\mathcal{G}$ which satisfies the following condition [see Proposition 2.17 of Lauritzen (1996)]: with $C_{(j)} = \cup_{i=1}^{j} C_i$ and $S_j = C_j \cap C_{(j-1)}$,

$$\text{for all } i > 1, \text{ there is a } j < i \text{ such that } S_i \subseteq C_j. \quad (5)$$

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

$$\chi(\mathcal{G}) = \{a \cap b;\ \ a, b \in \mathcal{C}(\mathcal{G}),\ a \neq b\}. \tag{6}$$

Since $S_j$ is obtained as intersection of neighboring cliques, we will call the $S_j$'s connectors of cliques or prime separators ("c" for clique) for short.

## 3 Distribution, interaction graph, and Markovian subgraph

We will use a boldface for a random vector, and so $\mathbf{X}_A$ denotes a vector of random variables $X_i$, $i \in A$. We use the symbol $\cdot \perp\!\!\!\perp \cdot | \cdot$, following Dawid (1979), to represent conditional independence. A distribution $P$ is said to be *globally Markov* with respect to an undirected graph $\mathcal{G}$ if, for a triple $(A, B, S)$ of disjoint subsets $A, B, S$ of $V$,

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B | \mathbf{X}_S \tag{7}$$

whenever $A$ is separated from $B$ by $S$ in $\mathcal{G}$. For convenience' sake, we will write (7) simply as $A \perp\!\!\!\perp B | S$.

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* according to $\mathcal{G}$ [Lauritzen (1996), section 3.2] if for all $c \in \mathcal{C}(\mathcal{G})$ there exist non-negative functions $\psi_c$ 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 $M_G(\mathcal{G})$, and by $M_F(\mathcal{G})$ the collection of the distributions that are factorized according to $\mathcal{G}$.

When $\mathcal{G}$ is decomposable, the two sets are closely related as in

**Theorem 3.1.** *Let $\mathcal{G}$ be decomposable. Then*

$$M_F(\mathcal{G}) = M_G(\mathcal{G}).$$

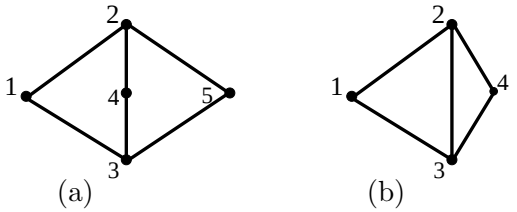



Figure 2: Some undirected graphs

**Proof:** See Proposition 3.19 in Lauritzen (1996). □

For a probability distribution $P$ of $\mathbf{X}_V$ we denote by $\mathcal{G}(\Gamma(P)) = (V, E)$ the interaction graph of $P$ which satisfies, under the hierarchy assumption for probability models, that

$$(u, v) \in E \iff \text{interaction is permitted between } X_u \text{ and } X_v \text{ under } P. \tag{8}$$

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,

$$(u, v) \in E \iff \{u, v\} \subseteq a \quad \text{for some } a \in \Gamma(P). \tag{9}$$

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

**Theorem 3.2.** For a distribution $P$,

$$P \in M_F(\mathcal{G}(\Gamma(P))) \quad \text{and} \quad P \in M_G(\mathcal{G}(\Gamma(P))).$$

**Proof:** By definition, $\Gamma(P) \preceq \mathcal{C}\,(\mathcal{G}(\Gamma(P)))$. Thus it follows that $P \in M_F(\mathcal{G}(\Gamma(P)))$, and so does $P \in M_G(\mathcal{G}(\Gamma(P)))$ by Proposition 3.8 in Lauritzen (1996). □

If graph (a) in Figure 2 is $\mathcal{G} = \mathcal{G}(\Gamma(P))$ for some distribution $P$ and $A = \{1, 2, 3, 4\}$, there are two sets, $\{2, 5\}$ and $\{3, 5\}$, in $\Gamma(P)$, which have 5 as an element in common. If we write the density $f$ of $P$ for the graph in a factorized form,

$$f(x) = g_{12}(x)g_{13}(x)g_{24}(x)g_{34}(x)g_{25}(x)g_{35}(x),$$

where $g_{uv}(x)$ is a non-negative function and $x$ depends upon the set $\{X_u, X_v\}$ through the vector $x_{\{u,v\}}$. The density of the marginal distribution for $\mathbf{X}_A$ is

$$f(x_A) = g_{12}(x)g_{13}(x)g_{24}(x)g_{34}(x) \left[ \int_{\mathcal{X}_5} g_{25}(x)g_{35}(x)d\mu(x_5) \right],$$

where $\mathcal{X}_5$ is the support of $X_5$ and $\mu$ is a measure on $\mathcal{X}_5$. The integral can be interpreted as a sum when $X_5$ is discrete. The bracket is a function of $X_2$ and $X_3$ at the most, which we denote by $g'_{23}(x)$. Inserting this into the above expression yields

$$f(x_A) = g_{12}(x)g_{13}(x)g_{24}(x)g_{34}(x)g'_{23}(x).$$

From this, we can see that the largest possible generating class of $P_A$ is given by

$$\begin{aligned}\bar{\Gamma}(P_A) &= (\Gamma(P) \setminus \{\{2,5\},\{3,5\}\}) \cup \{\{2,3\}\} \\ &= \{\{1,2\},\ \{1,3\},\ \{2,3\},\ \{2,4\},\ \{3,4\}\}.\end{aligned}$$

Graph (b) in Figure 2 is $\mathcal{G}(\bar{\Gamma}(P_A))$.

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

$$\bar{\Gamma}(P_A) = (\Gamma(P) \cap A) \cup \beta(\mathcal{J}_A), \tag{10}$$

where $\beta(\mathcal{J}_A)$ is defined in (4).

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 (10).

**Theorem 3.3.** For a distribution $P$ of $\mathbf{X}_V$ and $A \subseteq V$,

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

where $\mathcal{G}(\Gamma(P))_A$ is the Markovian subgraph of $\mathcal{G}(\Gamma)$ confined to $A$.

**Proof:** By definition, the interaction graph corresponding to the right hand side of (10) is $\mathcal{G}(\Gamma(P))_A$. Thus the result follows. □

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

**Corollary 3.4.** For a distribution $P$ of $\mathbf{X}_V$ and $A \subseteq V$,

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

From Theorem 3.3, 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 3.5.** For a distribution $P$ of $\mathbf{X}_V$ and $A \subseteq V$, suppose that $P \in M_G(\mathcal{G})$ for an undirected graph $\mathcal{G}$. Then

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

**Proof:** Since $P \in M_G(\mathcal{G})$, we have $\mathcal{G}(\Gamma(P)) \subseteq^e \mathcal{G}$. This implies that $\mathcal{G}(\Gamma(P))_A \subseteq^e \mathcal{G}_A$. So, by Corollary 3.4, we have the desired result. □

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 $P \in M(\mathcal{G})$. From the perspective of Corollary 3.4, we have called and will call $\mathcal{G}_A$ a Markovian subgraph of $\mathcal{G}$ and $\mathcal{G}$ a Markovian supergraph of $\mathcal{G}_A$

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 3.5, we have the following:

$$\begin{gathered} L(\mathcal{G}) = M(\mathcal{G}), \\ M(\mathcal{G}) \subseteq L(\mathcal{G}_A), \qquad \text{(by Corollary 3.5)} \\ P \in L(\mathcal{G}_A) \Longleftrightarrow P_A \in M(\mathcal{G}_A) \end{gathered} \tag{11}$$

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

It follows from (11) 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 3.6.** *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 (11), $P \in L(\mathcal{G}_A)$ 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. □

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 $\tilde{L}(\mathcal{G}_A,\ A \in \mathcal{V})$. In subsequent sections, we will explore how $\mathcal{G}_A,\ A \in \mathcal{V}$, are used in search of $\mathcal{G}$.

# 4 Combined model structures

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* or a *marginal model*. 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 4.1.** *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) 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 formally by $MXS(\mathcal{G}_1, \cdots, \mathcal{G}_m)$.*

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

In the lemma 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, $\langle \cdot | \cdot | \cdot \rangle$, is explained in Appendix A.

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

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

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

The following theorem is similar to Corollary 2.8 in Lauritzen (1996), but it is different in that an induced subgraph is considered in that corollary while a Markovian subgraph is considered here.

**Theorem 4.3.** *Every Markovian subgraph of a decomposable graph is decomposable.*

**Proof:** Suppose that a Markovian subgraph $\mathcal{G}_A$ of a decomposable graph $\mathcal{G}$ is not decomposable. Then there must exist a chordless cycle, say $C$, of length $\geq 4$ in $\mathcal{G}_A$. Denote the nodes on the cycle by $\nu_1, \cdots, \nu_l$ and assume that they form a cycle in that order where $\nu_1$ is a neighbor of $\nu_l$.

We need to show that $C$ itself forms a cycle in $\mathcal{G}$ or is contained in a chordless cycle of length $> l$ in $\mathcal{G}$. By Lemma 4.2, there is no edge in $\mathcal{G}$ between any pair of non-neighboring

nodes on the cycle. If $C$ itself forms a cycle in $\mathcal{G}$, our argument is done. Otherwise, we will show that the nodes $\nu_1, \cdots, \nu_l$ are on a cycle which is larger than $C$. Without loss of generality, we may consider the case where there is no edge between $\nu_1$ and $\nu_2$. If there is no path in $\mathcal{G}$ between the two nodes other than the path which passes through $\nu_3, \cdots, \nu_l$, then, since $C$ forms a chordless cycle in $\mathcal{G}_A$, there must exist a path between $v_1$ and $v_2$ other than the path which passes through $v_3, \cdots, v_l$. Thus the nodes $\nu_1, \cdots, \nu_l$ must lie in $\mathcal{G}$ on a chordless cycle of length $> l$. This completes the proof. □

This theorem and expression (6) imply that, as for a decomposable graph $\mathcal{G}$, the prime separators are always given in the form of a complete subgraph in $\mathcal{G}$ and in its Markovian subgraphs. This property may help us in searching for a graphical model based on a collection of the graphs of its submodels.

Lemma 4.2 states that a separator of a Markovian subgraph of $\mathcal{G}$ is also a separator of $\mathcal{G}$. We will next see that every maximal CMS is decomposable provided that all the Markovian subgraphs, $\mathcal{G}_1, \cdots, \mathcal{G}_m$, are decomposable.

**Theorem 4.4.** *Let $\mathcal{G}_1, \cdots, \mathcal{G}_m$ be decomposable. Then every MXS$(\mathcal{G}_1, \cdots, \mathcal{G}_m)$ is also decomposable.*

**Proof:** Suppose that there is a maximal CMS, say $\mathcal{H}$, which contains an $n$-cycle ($n \geq 4$) and let $A$ be the set of the nodes on the cycle. Since $\mathcal{H}$ is maximal, we can not add any edge to it. This implies that no more than three nodes of $A$ are included in any of the $V_i$'s, since any four or more nodes of $A$ that are contained in a $V_i$ form a cycle in $\mathcal{G}_i$, which is impossible due to the decomposability of the $\mathcal{G}_i$'s. Hence, the cycle in $\mathcal{H}$ may become a clique by edge-additions on the cycle, contradicting that $\mathcal{H}$ is maximal. Therefore, $\mathcal{H}$ must be decomposable. □

Theorem 4.4 does not hold for a CMS. For example, in Figure 6, graph $\mathcal{G}$ is not decomposable. However, the Markovian subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are both decomposable. And $\chi(\mathcal{G}) = \{\{4\},\ \{7\}\}$, $\chi(\mathcal{G}_1) = \{\{2,3\},\ \{5,6\},\ \{8,9\}\}$, and $\chi(\mathcal{G}_2) = \{\{4\},\ \{7\}\}$. Note that, for $\mathcal{H}$ in the figure, $\chi(\mathcal{H}) = \chi(\mathcal{G}_1) \cup \chi(\mathcal{G}_2)$, which holds true in general as is shown in Theorem 4.7 below. The theorem characterizes a maximal CMS and the proposed method of model combination is rooted in that theorem. Before stating the theorem, we will show that if a set of nodes is a prime separator in a Markovian subgraph, then it is not intersected in another.

**Theorem 4.5.** *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*

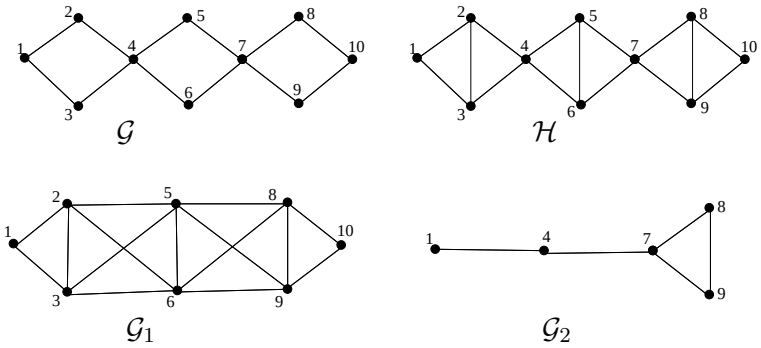



Figure 3: An example of a non-decomposable graph ($\mathcal{G}$) whose Markovian subgraphs ($\mathcal{G}_1$, $\mathcal{G}_2$) are decomposable. Graph $\mathcal{H}$ is a maximal CMS of $\mathcal{G}_1$ and $\mathcal{G}_2$ which is defined in section 3.

*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 Lemma 4.2, we have $\langle u|S|v\rangle_{\mathcal{G}}$. Since $C \in \chi(\mathcal{G}_1)$ and $\mathcal{G}_1$ is decomposable, $C$ is an intersection of some neighboring cliques of $\mathcal{G}_1$ by equation (6). 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 $\mathcal{G}_2$. □

This theorem states that, if $\mathcal{G}$ is decomposable, a prime separator in a Markovian subgraph of $\mathcal{G}$ is either a prime separator or a complete subgraph in any other Markovian subgraph of $\mathcal{G}$. If the set of the prime separator is contained in only one clique of a Markovian subgraph, the set is embedded in the clique. For a subset $V'$ of $V$, if we put $\mathcal{G}_1 = \mathcal{G}$ and $\mathcal{G}_2 = \mathcal{G}_{V'}$ in Theorem 4.5, we have the following corollary:

**Corollary 4.6.** Let $\mathcal{G}$ be a decomposable graph and suppose that a set $C \in \chi(\mathcal{G})$ and that $C \subseteq V' \subset V$. Then $C$ is not intersected in a Markovian subgraph $\mathcal{G}_{V'}$ of $\mathcal{G}$ by any other subset of $V'$.

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 prime separators as in the theorem below.

**Theorem 4.7.** 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 Appendix B.

For a given set of Markovian subgraphs, we can readily obtain the set of prime separators under the decomposability assumption. By (6), we can find $\chi(\mathcal{G})$ for any decomposable graph $\mathcal{G}$ simply by taking all the intersections of the cliques of the graph. An apparent feature of a maximal CMS in contrast to a CMS is stated in Theorem 4.7.

For a set $\mathcal{M}$ of Markovian subgraphs of a graph $\mathcal{G}$, there can be more than one maximal CMS of $\mathcal{M}$. But there is only one such maximal CMS that contains $\mathcal{G}$ as its edge-subgraph.

**Theorem 4.8.** *Suppose there are $m$ Markovian subgraphs $\mathcal{G}_1, \cdots, \mathcal{G}_m$ of a decomposable graph $\mathcal{G}$. Then there exists a unique maximal CMS $\mathcal{H}^*$ of the $m$ Markovian subgraphs such that $\mathcal{G} \subseteq^e \mathcal{H}^*$.*

**Proof:** By Theorem 4.7 (i), we have

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

If $\cup_{i=1}^{m} \chi(\mathcal{G}_i) = \chi(\mathcal{G})$, then since $\mathcal{G}$ is decomposable, $\mathcal{G}$ itself is a maximal CMS. Otherwise, let $\chi' = \chi(\mathcal{G}) - \cup_{i=1}^{m} \chi(\mathcal{G}_i) = \{A_1, \cdots, A_g\}$. Since $A_1 \notin \cup_{i=1}^{m} \chi(\mathcal{G}_i)$, we may add edges so that $\cup_{C \in \mathcal{C}_{\mathcal{G}}(A_1)} C$ becomes a clique, and the resulting graph $\mathcal{G}^{(1)}$ becomes a CMS of $\mathcal{G}_1, \cdots, \mathcal{G}_m$ with $\chi(\mathcal{G}^{(1)}) - \cup_{i=1}^{m} \chi(\mathcal{G}_i) = \{A_2, \cdots, A_g\}$.

We repeat the same clique-merging process for the remaining $A_i$'s in $\chi'$. Since each clique-merging makes the corresponding prime separator disappear into the merged, new clique while maintaining the resulting graph as a CMS of $\mathcal{G}_1, \cdots, \mathcal{G}_m$, the clique-merging creates a CMS of $\mathcal{G}_1, \cdots, \mathcal{G}_m$ as an edge-supergraph of the preceding graph. Therefore, we obtain a maximal CMS, say $\mathcal{H}^*$, of $\mathcal{G}_1, \cdots, \mathcal{G}_m$ at the end of the sequence of the clique-merging processes for all the prime separators in $\chi'$. $\mathcal{H}^*$ is the desired maximal CMS as an edge-supergraph of $\mathcal{G}$.

Since the clique-merging begins with $\mathcal{G}$ and, for each prime separator in $\mathcal{G}$, the set of the cliques which meet at the prime separator only is uniquely defined, the uniqueness of $\mathcal{H}^*$ follows. □

The unique existence of maximal CMS for a given model structure provides us with a useful guideline for searching, based on a set of marginal model structures, the maximal CMS which contains the actual model structure as an edge-subgraph. Since the maximal CMS is at least as large as the actual model $\mathcal{G}$, we have only to work with the maximal CMS to find $\mathcal{G}$ by removing edges as a measure of goodness-of-fit suggests.

Let $P$ be a distribution of $\mathbf{X}_V$ and $\mathcal{V}$ be a collection of subsets of $V$. Result (ii) of Theorem 4.7 is useful in searching for $\mathcal{G}(\Gamma(P))$ from a collection of graphs $\mathcal{G}(\bar{\Gamma}(P_A)),\quad A \in \mathcal{V}$, which, by Theorem 3.3, is the same as the collection $\mathcal{S}$ of $\mathcal{G}(\Gamma(P))_A,\quad A \in \mathcal{V}$. Note that $\mathcal{G}(\Gamma)_A$ are Markovian subgraphs of $\mathcal{G}(\Gamma)$. Thus, under the decomposability assumption, there is a unique maximal CMS, $\mathcal{H}^*$, by Theorem 4.8, based on $\mathcal{S}$ which contains $\mathcal{G}(\Gamma(P))$ as an edge-subgraph. Since $\mathcal{H}^*_A = \mathcal{G}(\Gamma(P))_A$, if we put $\mathcal{G} = \mathcal{G}(\Gamma(P))$ in Theorem 3.6, we end up with the summarizing expression

$$M(\mathcal{G}(\Gamma(P))) \subseteq M(\mathcal{H}^*) \subseteq \tilde{L}\,(\mathcal{G}(\Gamma(P))_A,\ A \in \mathcal{V})\,, \tag{12}$$

where the first inequality follows since $\mathcal{G}(\Gamma(P)) \subseteq^e \mathcal{H}^*$. Since $P \in M(\mathcal{G}(\Gamma(P)))$, expression (12) implies that $P$ is also Markov relative to the maximal CMS, $\mathcal{H}^*$. In reality, $\mathcal{G}(\Gamma(P))_A$'s are determined based on data. Once we obtain $\mathcal{H}^*$, the model for $\mathbf{X}_V$ can be determined by removing edges from $\mathcal{H}^*$ as the data suggest.

Expression (12) holds if the separateness of nodes in $\mathcal{G}(\Gamma)_A$, $A \in \mathcal{V}$, is preserved in $\mathcal{H}^*$. Since $\mathcal{H}^*$ is a CMS of $\mathcal{G}(\Gamma)_A$, $A \in \mathcal{V}$, the separateness in $\mathcal{H}^*$ must also be preserved in $\mathcal{G}(\Gamma)_A$, $A \in \mathcal{V}$. We will call this condition of preservation of separateness between a set of Markovian subgraphs and its corresponding maximal CMS the *separateness condition*. This condition is of practical use in model combination as will be illustrated in next section.

## 5 Illustration of model combination with simple models

For an illustration of model combination, we will consider a couple of examples where relatively simple models are considered. We will make use of result (ii) of Theorem 4.7 for the model combination by forming a graph of the prime separators based on a given set of Markovian subgraphs.

When a graphical model is decomposable, the prime separators of the model structure each form a complete subgraph, and they are related to each other in the context of graph separateness. If the model structure is given in the form of an interaction graph, the graph separateness can be interpreted in terms of conditional independence. Since the prime separators are themselves separators, they can form an undirected graph where each prime separator corresponds to a node in the same way as a clique forms a node in a join tree of cliques [Pearl (1988)]. However, the graph of prime separators cannot be a join tree because, if $l$ ($l > 2$) prime separators share a clique, then the prime separators form a clique in the graph of the prime separators. For simplicity, we will call the "graph of prime separators" GOPS.

We will call a node in a prime separator a connector-node and a node not included in a prime separator a non-separator node. As for the model $\mathcal{G}$ and its Markovian subgraphs, $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$, in Figure 4, we have that $\chi(\mathcal{G}_1) = \{2\}$, $\chi(\mathcal{G}_2) = \{\{2\}, \{4\}\}$, and $\chi(\mathcal{G}_3) = \{\{2\}, \{7\}\}$. The GOPS of the prime separators is given in panel (a) of Figure 5, which is obtained from $\mathcal{G}_1, \mathcal{G}_2$, and $\mathcal{G}_3$ in Figure 4 under the separateness condition since the GOPS is the Markovian subgraph of a maximal CMS of the three Markovian subgraphs in Figure 4. In $\mathcal{G}_2$, nodes 2 and 5 are connected by a path, and $\langle 5|2|7\rangle_{\mathcal{G}_3}$ holds. Thus, we have the GOPS as in panel (a) of Figure 5. According to result (12), we have only to observe the separateness condition in constructing a maximal CMS based on a given set of Markovian sugbgraphs of a model. Maximal CMSs are constructed by adding the non-separator nodes, 1, 3, 5, 6, 8, and 9, to the GOPS in Figure 5. Denote one of the maximal CMSs by $\mathcal{H}$ and the neighborhood of

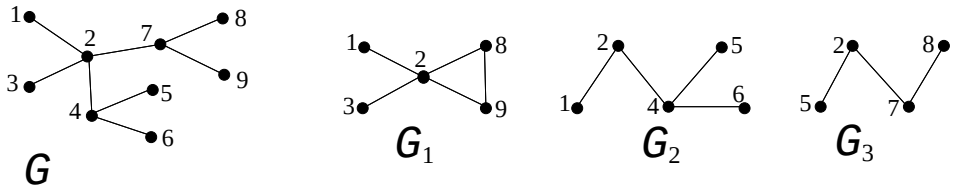



Figure 4: A tree-shape model structure $\mathcal{G}$ and its Markovian subgraphs

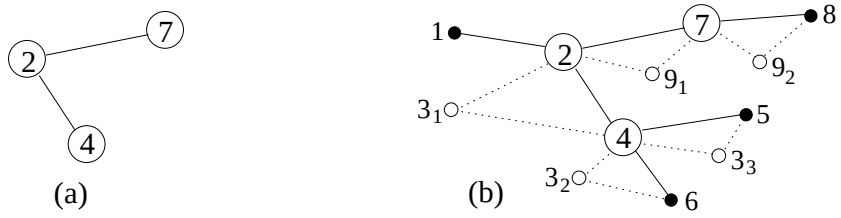



Figure 5: The GOPS (panel (a)) from $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ in Figure 4 and the corresponding maximal CMSs (panel (b)). The two nodes, $9_1$ and $9_2$, mean that the two locations are possible for $X_9$, and similarly for the three nodes of $X_3$.

Table 1: Neighborship between non-separator node and prime separator node for the Markovian subgraphs in Figure 4.

| Non-separator node ($v$) | prime separators as neighbors of $v$ | Relevant Markovian subgraphs |
|---|---|---|
| 1 | $\{2\}$ | $\mathcal{G}_1$, $\mathcal{G}_2$ |
| 3 | $\{2\}, \{4\}$ | $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$ |
| 5 | $\{4\}$ | $\mathcal{G}_2$, $\mathcal{G}_3$ |
| 6 | $\{4\}$ | $\mathcal{G}_2$ |
| 8 | $\{7\}$ | $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$ |
| 9 | $\{2\}, \{7\}$ | $\mathcal{G}_1$, $\mathcal{G}_3$ |

node $v$ in $\mathcal{H}$ by $ne_{\mathcal{H}}(v)$. Then, $\chi(\mathcal{H}) = \{\{2\}, \{4\}, \{7\}\}$. From $\mathcal{G}_1$ and $\mathcal{G}_2$, we can see that $ne_{\mathcal{H}}(1) \cap \chi(\mathcal{H}) = \{\{2\}\}$. As for node 8, $\mathcal{G}_1$ and $\mathcal{G}_3$ imply that $ne_{\mathcal{H}}(8) \cap \chi(\mathcal{H}) = \{\{7\}\}$. The neighborship between prime separator and non-separator node is summarized in Table 1. Note, in this table, that nodes 3 and 9 can have multiple prime separators as neighbors. For instance, as for node 3, we can see from the three Markovian subgraphs that node 3 may have some of nodes 2, 4, 5, and 6 as its neighbors, since nodes 3 is separated from nodes 1 and 8 by node 2 and so are nodes 4, 5, and 6. Table 1 is reflected in panel (b) of Figure 5. Notice that graph $\mathcal{G}$ in Figure 4 is an edge-subgraph of the maximal CMS in panel (b) which consists of nodes $1, 2, 3_1, 4, \cdots, 9_2$.

We combined the Markovian subgraphs in Figure 4, first by constructing a GOPS and then by adding non-separator nodes to the GOPS. By definition, prime separators are separated by prime separators only. Thus, if a GOPS is constructed from a set of Markovian subgraphs, the model combination is done simply by adding the non-separator nodes to the GOPS. However, the GOPS may not be constructed without considering the separateness of non-separator nodes, as is shown in constructing the GOPS in Figure 5. In another example with the Markovian subgraphs in Figure 6, we will see how non-separator nodes are used for the construction of a GOPS.

As for the three Markovian subgraphs, $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ in Figure 6, we have that $\chi(\mathcal{G}_1) = \{\{2, 3\}\}$, $\chi(\mathcal{G}_2) = \{\{3, 4\}\}$, and $\chi(\mathcal{G}_3) = \{\{2, 3\}, \{3, 6\}\}$. Thus, if we denote by $\mathcal{H}'$ a maximal CMS of the three Markovian subgraphs, then $\chi(\mathcal{H}') = \{\{2, 3\}, \{3, 4\}, \{3, 6\}\}$. There seems no separateness among the three prime separators if we do not consider the separateness of the non-separators. We may thus begin with a complete graph of the prime separators and check if the complete graph violates any separateness that lies in the three Markovian subgraphs. From the three

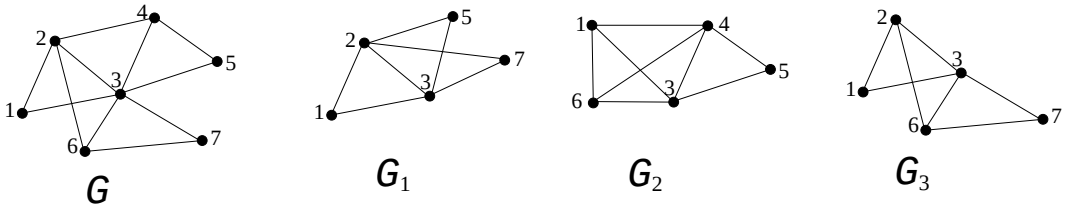



Figure 6: A decomposable model ($\mathcal{G}$) and its Markovian subgraphs

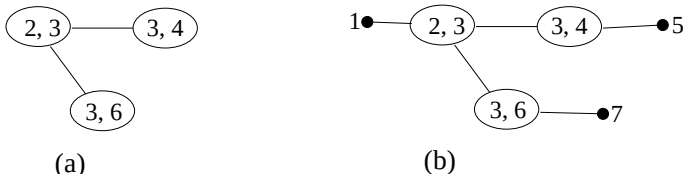



Figure 7: The GOPS (panel (a)) from $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ in Figure 6 and a node-added GOPS (panel (b)).

Table 2: Neighborship between non-separator node and prime separator node for the Markovian subgraphs in Figure 6.

| Non-separator node ($v$) | prime separators as neighbors of $v$ | Relevant Markovian subgraphs |
|---|---|---|
| 1 | $\{2,3\}$ | $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$ |
| 5 | $\{3,4\}$ | $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$ |
| 7 | $\{3,6\}$ | $\mathcal{G}_1$, $\mathcal{G}_3$ |

Markovian subgraphs in Figure 6, we can see that

$$\langle 1|\{2,3\}|\{5,6,7\}\rangle, \quad \langle 5|\{3,4\}|\{1,6\}\rangle, \quad \langle 5|\{2,3\}|\{1,7\}\rangle, \quad \text{and} \quad \langle 7|\{3,6\}|\{1,2\}\rangle. \tag{13}$$

From this separateness, we can see that $ne_{\mathcal{H}'}(1) = \{2,3\}$ and $ne_{\mathcal{H}'}(7) = \{3,6\}$. As for node 5, the two graph-separations in the middle of (13) suggest a clique of nodes 2, 3, 4, and 5. But since $\{3,4\}$ is a prime separator, $1 \sim 6$ in $\mathcal{G}_2$, and $1 \not\sim 6$ in $\mathcal{G}_3$, the GOPS in panel (a) of Figure 7 follows. The three non-separator nodes, 1, 5, and 7, are added to the GOPS in panel (b) of Figure 7. The neighborship of the non-separator nodes with the three prime separators in Figure 6 is summarized in Table 2. Note that the graph in panel (b) of Figure 7 is the same as the $\mathcal{G}$ in Figure 6.

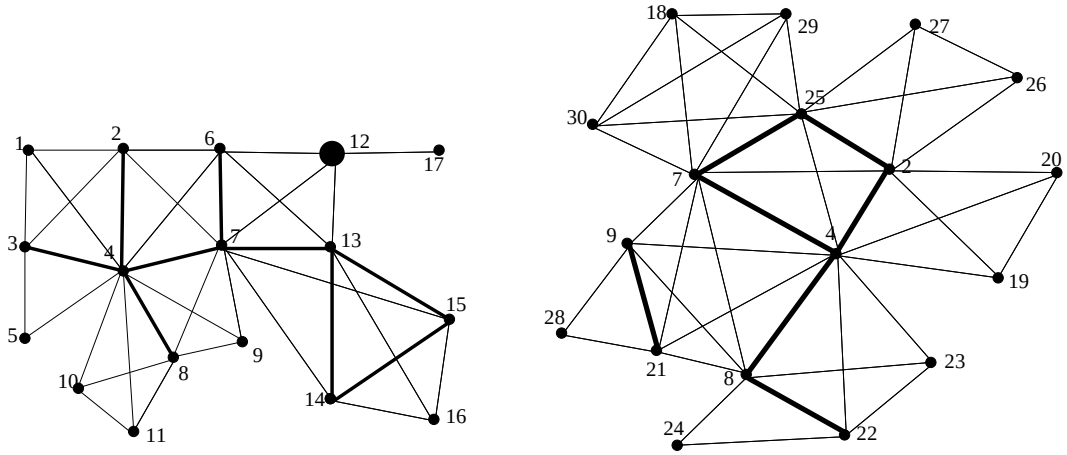
We have seen through the two examples that a GOPS is constructed from a given set of Markovian subgraphs under the separateness condition. Since a non-separator node has at least one prime separator as neighbors in a maximal CMS, adding non-separator nodes to a GOPS is relatively easy.

# 6 Illustration of model combination with larger models

This section is a continuation of the preceding section in the sense that we consider combining larger models than those considered in the preceding section. The Markovian subgraphs to consider are given in Figure 8, where $\mathcal{G}_1$ is of 17 nodes and $\mathcal{G}_2$ is of 18 nodes. As is described below, combining larger Markovian subgraphs does not necessarily mean a higher complexity of the combining process. When Markovian subgraphs share some prime separators or nodes, the shared nodes can be very informative for locating some of the variables involved in the Markovian subgraphs in the construction process of a maximal CMS.

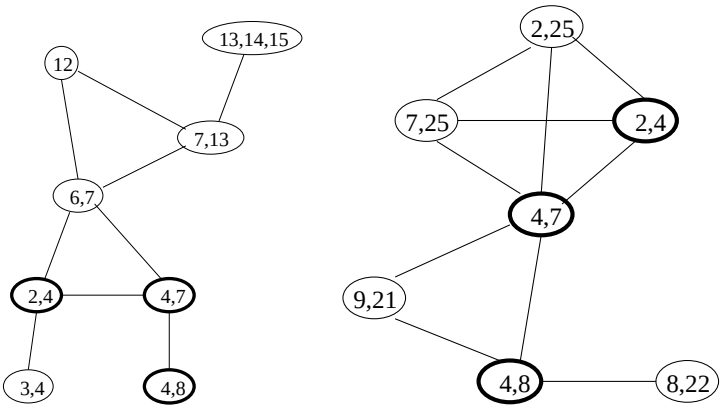
As is displayed in Figure 8(a), the prime separators, $\{12\}$, $\{6,7\}$, and $\{7,13\}$, form a clique, and so do the prime separators, $\{2,4\}$, $\{4,7\}$, and $\{6,7\}$. The former three prime separators form a clique in Figure 8, and so do the latter three prime separators. By Lemma 4.2, we know that the separateness by prime separators in a subgraph, $\mathcal{G}'$ say, is preserved in a graph for which $\mathcal{G}'$ is a Markovian subgraph. We may thus combine the GOPSs of the submodels first and then expand the combined GOPS by adding the non-separator nodes onto the combined GOPS.

Let us consider combining the two model structures $\mathcal{G}_1$ and $\mathcal{G}_2$ which are displayed in panels (a) and (b) of Figure 8, respectively. The GOPS of $\mathcal{G}_2$ is given in Figure 9(b). Note that there are three prime separators common in the two GOPSs. They are $\{2,4\}$, $\{4,7\}$, and $\{4,8\}$. Considering this node-sharing, we can combine the two GOPSs into the GOPS as in panel (iii)




(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$

Figure 8: Markovian subgraphs $\mathcal{G}_1$ and $\mathcal{G}_2$. Prime separators are distinguished by thick lines and a big bullet.




(a) GOPS of $\mathcal{G}_1$ (b) GOPS of $\mathcal{G}_2$

Figure 9: The GOPSs of $\mathcal{G}_1$ and $\mathcal{G}_2$. Thick circled prime separators are shared by both GOPSs.

of Figure 10, which is described below.

We begin with the GOPS of $\mathcal{G}_1$ and the combination proceeds by adding the prime separators in $\chi(\mathcal{G}_2)$ as follows:

**As for prime separators** $\{8, 22\}$ **and** $\{9, 21\}$**:** From the Markov property among the three shared prime separators, $\{2, 4\}$, $\{4, 7\}$, $\{4, 8\}$, and the prime separators, $\{8, 22\}$, $\{9, 12\}$, as appearing in the GOPS of $\mathcal{G}_2$, we can add the prime separators, $\{8, 22\}$ and $\{9, 21\}$, as in panel (ii) of Figure 10.

**As for prime separators** $\{2, 25\}$ **and** $\{7, 25\}$**:** According to the GOPS of $\mathcal{G}_2$, $\{2, 25\} \perp\!\!\!\perp \{4, 8\} | \{4, 7\}$, so $\{2, 25\}$ may form a clique either with $\{2, 4\}, \{4, 7\}, \{6, 7\}$ or with $\{2, 4\}, \{3, 4\}$. But $\{2, 25\}$ and $\{7, 25\}$ share the node 25 and $\{7, 25\}$ can not form a clique with $\{2, 4\}, \{3, 4\}$ since the set of prime separators $\{2, 4\}, \{4, 7\}, \{6, 7\}$ share the node 7 with the prime separator $\{7, 25\}$. Thus, $\{2, 25\}$ must be part of the clique which includes $\{2, 4\}, \{4, 7\}, \{6, 7\}$, and so must be $\{7, 25\}$ since the two prime separator share the node 25 and there is no other prime separator in both of the GOPSs that contains the node. The result is depicted in panel (iii) of Figure 10.

Once a combined GOPS is obtained, adding non-separator nodes is relatively straightforward. For instance, in $\mathcal{G}_1$ in Figure 8, nodes 10, 11 are neighbors of the prime separator $\{4, 8\}$ only and thus independent of all the other nodes conditional on the prime separator. This is represented in graph by connecting the nodes directly to the prime separator $\{4, 8\}$ as in panel (i) of Figure 11. The other non-separator nodes are added in the same manner.

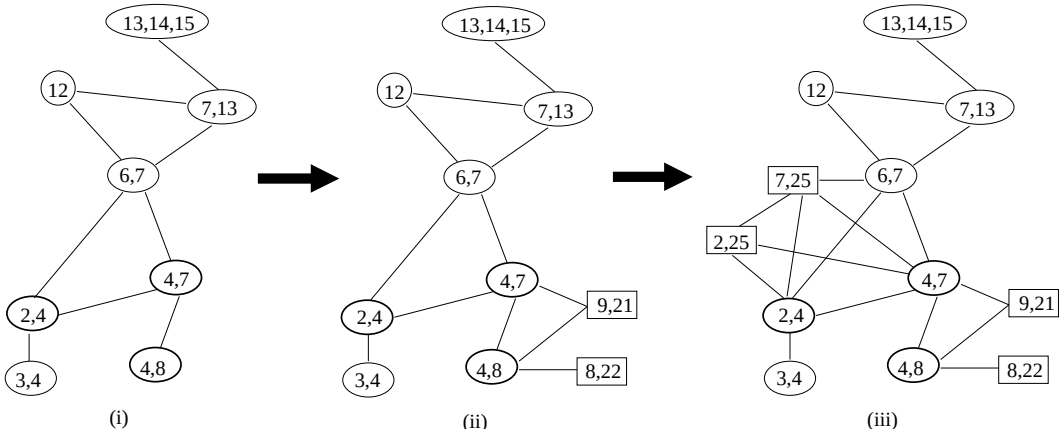



Figure 10: A graphic display of the combining process of the two GOPSs in Figure 9. The boxes are the prime separators of $\mathcal{G}_2$.

Panel (i) of Figure 11 is a graphic display of the addition of the non-separator nodes,

$$X_1, X_5, X_{10}, X_{11}, X_{16}, X_{17},$$

of $\mathcal{G}_1$, onto the GOPS of $\chi(\mathcal{G}_1) \cup \chi(\mathcal{G}_2)$; panel (ii) results from a further addition of the non-separator nodes,

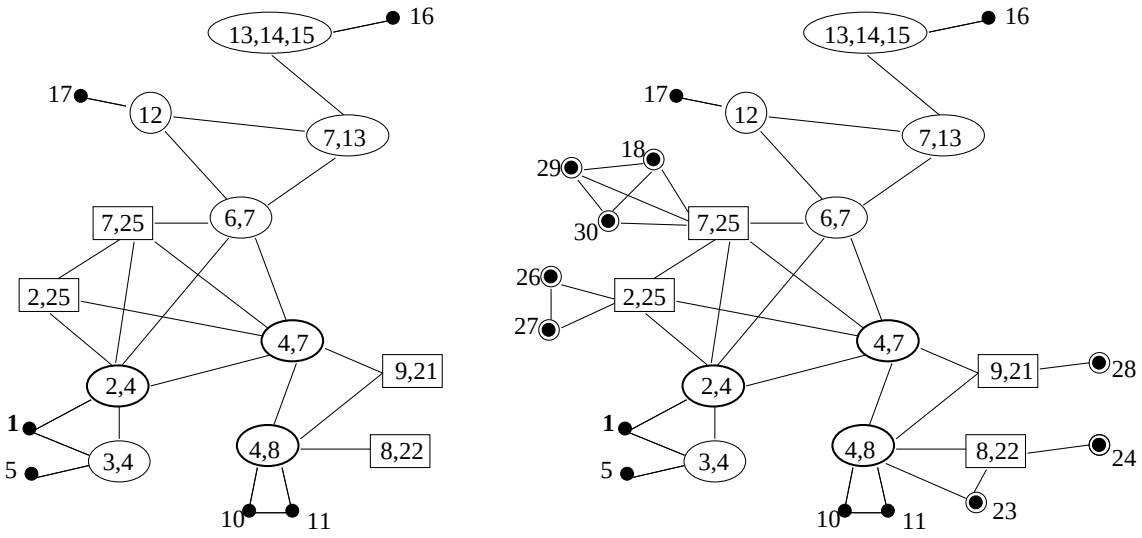
$$X_{18}, X_{19}, X_{20}, X_{23}, X_{24}, X_{26}, X_{27}, X_{28}, X_{29}, X_{30},$$

of $\mathcal{G}_2$.

Transforming the graph in panel (ii) of Figure 11 into an undirected graph is done by unpacking the prime separators in such a way that all the nodes are located properly under the Markov (or conditional independence) condition that is imposed by the graph in panel (ii). For instance, the nodes 13, 14, 15, and 16 form a clique and so do the nodes 6, 7, 12, and 13. The resulting graph is in Figure 13.

Note that nodes 19 and 20 are not included in the graph in Figure 13. There are four different ways of locating the nodes as displayed in Figure 12. Let the clique $\{1, 2, 3, 4\}$ be denoted by $C_1$ and the clique $\{3, 4, 5\}$ by $C_2$. Then the four ways of node-addition are $(i)$ making $C_1 \cup \{19, 20\}$ a clique, $(ii)$ making $C_2 \cup \{19, 20\}$ a clique, $(iii)$ making $C_1 \cup \{19\}$ and $C_2 \cup \{20\}$ two new cliques, and $(iv)$ making $C_1 \cup \{20\}$ and $C_2 \cup \{19\}$ two new cliques. The node-addition is determined by the conditional independence regarding the nodes 19 and 20 as indicated in $\mathcal{G}_2$ of Figure 8:

$$\{19, 20\} \perp\!\!\!\perp (V_2 \setminus \{2, 4, 19, 20\}) \mid \{2, 4\}. \tag{14}$$




(i) Nodes (bullets) in $V_1$ are added

(ii) Nodes (bullets in circles) in $V_2$ are added to the graph in panel (i).

Figure 11: Addition of the non-separator nodes onto the GOPS of $\chi(\mathcal{G}_1) \cup \chi(\mathcal{G}_2)$. The circles represent the prime separators of $\mathcal{G}_1$, the boxes the prime separators of $\mathcal{G}_2$ only, and the bullets represent nodes. For the addition of nodes 19 and 20, see Figure 12.

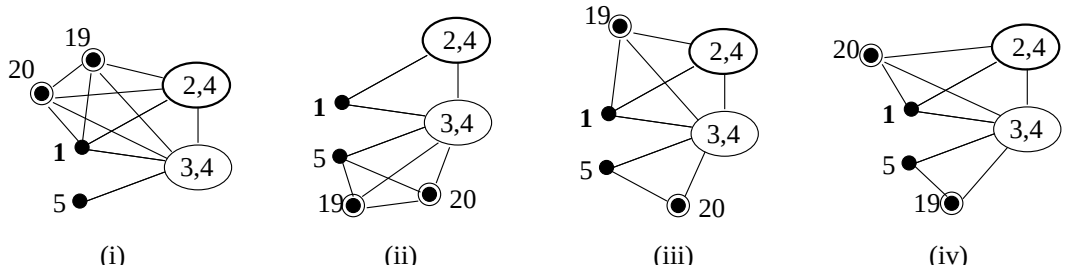



Figure 12: Four situations of node-addition of $X_{19}$ and $X_{20}$ onto the GOPS of $\chi(\mathcal{G}_1) \cup \chi(\mathcal{G}_2)$

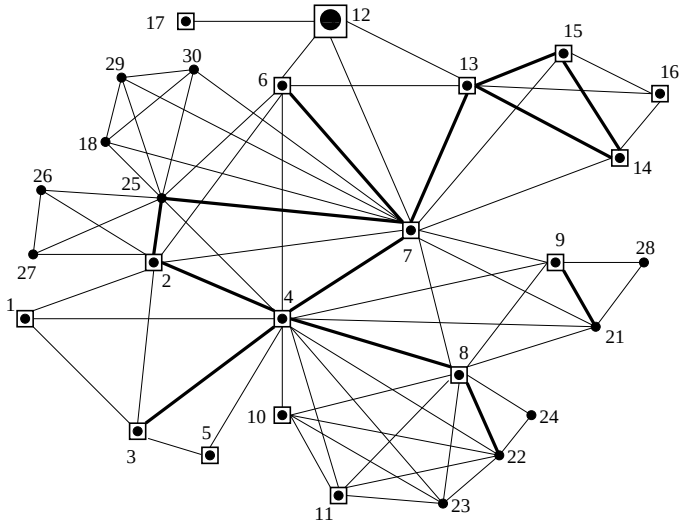



Figure 13: The maximal CMS of $\mathcal{G}_1$ and $\mathcal{G}_2$ with nodes 19 and 20 not included. The boxed nodes are from $\mathcal{G}_1$ and the rest are from $\mathcal{G}_2$; some nodes are shared by both marginal models. Node 12 is an prime separator.

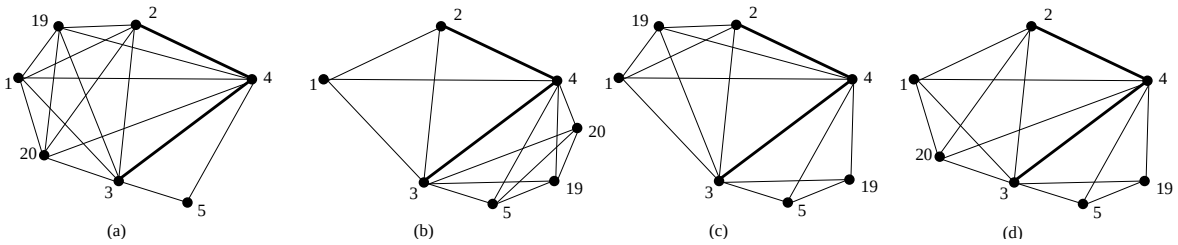



Figure 14: The parts of graph corresponding to the four node-additions of nodes 19 and 20. Panels a, b, c, and d correspond respectively to situations $(i)$, $(ii)$, $(iii)$, and $(iv)$ of node-addition as described in the paragraph which includes expression (14). Replacing each panel with the cliques of $\{1, 2, 3, 4, 5\}$ in Figure 13 yields a maximal CMS for $\mathcal{G}_1$ and $\mathcal{G}_2$.

Independence graphs corresponding to the four situations (i.e., the four panels of Figure 12) are given in Figure 14. One of the four structures can be glued onto the bottom-left corner of Figure 13 as any further information concerning nodes 19 and 20 may suggest.

# 7 Further discussion and conclusion

In Theorem 4.7, 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, derived maximal CMSs may not contain $\mathcal{G}$ as an edge-subgraph. Simple examples of this situation are displayed in Figure 15. In the first row of the figure are two interaction graphs ($\mathcal{G}^1$) for $X_1, X_2, X_3$ and a subgraph $\mathcal{G}^1_{\{1,3\}}$ 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 $\mathcal{G}^1_{\{1,3\}}$ is possible with the graphical log-linear model (1b) in the figure when

$$E[(P(\mathbf{X}_{\{1,3\}} = \mathbf{x}_{\{1,3\}}|X_2)] = P(X_1 = x_1)P(X_3 = x_3) \quad \text{for all} \quad x_{\{1,3\}} \in \mathcal{X}_{\{1,3\}}, \tag{15}$$

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 $\mathcal{G}^2_{\{1,3,4\}}$ is also possible from the graphical model (2b) in the figure. Instances of this phenomenon follow.

**Example 7.1.** We will present contingency tables for which the pair of models, (1b) and $\mathcal{G}^1_{\{1,3\}}$ in Figure 15, are possible and are the pair of models, (2b) and $\mathcal{G}^2_{\{1,3,4\}}$.

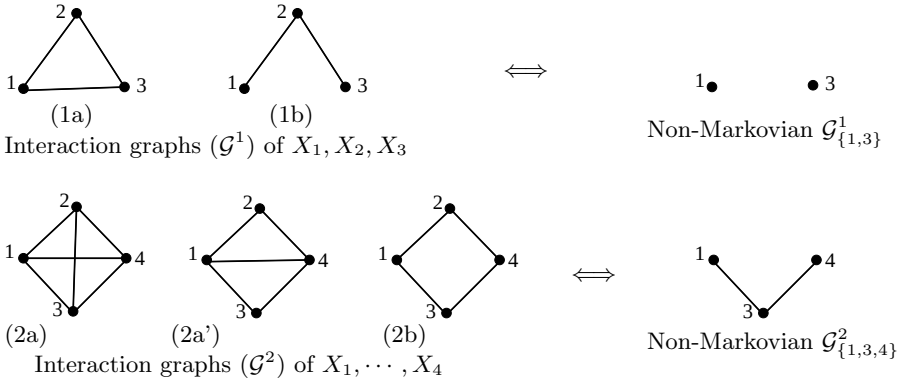



Figure 15: Some simple examples where subgraphs are not Markovian.

**(a)** Concerning models (1b) and $\mathcal{G}^1_{\{1,3\}}$:

| $x_2$ | $x_1$ | $x_3$ | $P(\mathbf{X}=\mathbf{x})$ | $x_2$ | $x_1$ | $x_3$ | $P(\mathbf{X}=\mathbf{x})$ |
|---|---|---|---|---|---|---|---|
| 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 $\mathcal{G}^2_{\{1,3,4\}}$:

| $x_2$ | $x_3$ | $x_1$ | $x_4$ | $P(\mathbf{X}=\mathbf{x})$ | $x_2$ | $x_3$ | $x_1$ | $x_4$ | $P(\mathbf{X}=\mathbf{x})$ |
|---|---|---|---|---|---|---|---|---|---|
| 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 15. The marginal for $\mathbf{X}_{\{1,3,4\}}$ 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 (15), 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. This "Markovian" assumption on subgraphs of interaction graphs is similar to the faithfulness assumption on distributions with respect to a directed acyclic graph (DAG)

or a causal graph, which is described in Spirtes et al. (2000). We say that a distribution $P$ is faithful to a DAG $\mathcal{G}$ if all independence and conditional independence relationships that are embedded in $P$ are displayed in $\mathcal{G}$. The probabilistic independencies that are representable by a DAG are described in Pearl (1988, section 3.3) and Lauritzen (1966, section 3.2). Robins et al. (2003) discuss the relationship between a joint distribution and its marginal distribution under the faithfulness assumption, where data are available for the marginal distribution only.

Finally, Theorem 4.7 is a main result of this paper. The model combination is recommended to be done in two steps, the first being that we construct a GOPS and the second that we add non-separator nodes to the GOPS. Several GOPSs are possible from a given set of Markovian subgraphs, and each of them is obtained in such a way that the separateness lying in a given set of Markovian subgraphs is preserved in the GOPS. The number of the possible GOPSs from a set, $Q$ say, of Markovian subgraphs of a model becomes smaller as the set $Q$ has more information about the model.

In practice, the Markovian subgraphs are often given in the form of interaction graphs. Suppose that we are given a data set for a set, $V$ say, of random variables, and that, for every subset $A$ of $V$ with $|A| \leq r$ for some $0 < r < |V|$, we can obtain an interaction graph which is appropriate to $A$. Then we can collect information about the model structure of $V$ as much as possible under the restriction that $|A| \leq r$, where the information is presented in the form of interaction graphs.

Once a maximal CMS is obtained, we may use it as an initial model structure in search of an appropriate model for data. Bergsma and Rudas (2002) proposed theorems by which we can use the parameter estimates of marginal models in developing a model for the whole data. If the data are for mixed continuous and discrete variables, the methods that are instrumental for modelling the whole data using the maximal CMS as an initial model structure include the methods of likelihood factorization by Cox and Wermuth (1999) and the application of the notion of model collapsibility as proposed by Didelez and Edwards (2004).

## Appendix A: Graphical terminology

Let $\mathcal{G} = (V, E)$ be an undirected graph. If $(u, v) \in E$, we say that $u$ is a neighbor node of $v$ or vice versa and write it as $u \sim v$. 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 $u \in V$, we define $ne(v) = \{u \in V;\ \ v \sim u \text{ in } \mathcal{G}\}$ and, for $A \subseteq V$, $bd(A) = \cup_{v \in A} ne(v) \setminus A$. 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 notation $\langle \cdot|\cdot|\cdot\rangle_{\mathcal{G}}$ follows Pearl (1988). 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 $A^c$. The cardinality of a set $A$ will be denoted by $|A|$. For two collections A, B of sets, if, for every $a \in A$, there exists a set $b$ in $B$ such that $a \subseteq b$, we will write $A \preceq B$.

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

**Definition A.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 A.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 $\mathcal{G}_{B\cup C}^{ind}$.

## Appendix B: Proof of Theorem 4.7

We will first prove result (i). For a subset of nodes $V_j$, the following holds:

(i') If $V_j$ does not contain a subset which is a prime separator of $\mathcal{G}$, then $\chi(\mathcal{G}_j) = \emptyset$.

(ii') Otherwise, i.e., if there are prime separators, $C_1, \cdots, C_r$, of $\mathcal{G}$ as subsets of $V_j$,

(ii'-a) if there are no nodes in $V_j$ that are separated by any of $C_1, \cdots, C_r$ in $\mathcal{G}$, then $\chi(\mathcal{G}_j) = \emptyset$.

(ii'-b) if there is at least one of the prime separators, say $C_s$, such that there are a pair of nodes, say $u$ and $v$, in $V_j$ such that $\langle u|C_s|v\rangle_{\mathcal{G}}$, then $\chi(\mathcal{G}_j) \neq \emptyset$.

We note that, since $\mathcal{G}$ is decomposable, the condition that $V_j$ contains a separator of $\mathcal{G}$ implies that $V_j$ contains a prime separator of $\mathcal{G}$. As for (i'), every pair of nodes, say $u$ and $v$, in $V_j$ have at least one path between them that bypasses $V_j \setminus \{u, v\}$ in the graph $\mathcal{G}$ since $V_j$ does not contain any prime separator of $\mathcal{G}$. Thus, (i') follows.

On the other hand, suppose that there are prime separators, $C_1, \cdots, C_r$, of $\mathcal{G}$ as a subset of $V_j$. The result (ii'-a) is obvious, since for each of the prime separators, $C_1, \cdots, C_r$, the rest of the nodes in $V_j$ are on one side of the prime separator in $\mathcal{G}$.

As for (ii'-b), let there be two nodes, $u$ and $v$, in $V_j$ such that $\langle u|C_s|v\rangle_{\mathcal{G}}$. Since $\mathcal{G}$ is decomposable, $C_s$ is an intersection of neighboring cliques in $\mathcal{G}$, and the nodes $u$ and $v$ must appear in some (not necessarily neighboring) cliques that are separated by $C_s$. Thus, the two nodes are separated by $C_s$ in $\mathcal{G}_j$ with $C_s$ as a prime separator in $\mathcal{G}_j$. Any proper subset of $C_s$ can not separate $u$ from $v$ in $\mathcal{G}$ and in any of its Markovian subgraphs.

From the results (i') and (ii'), it follows that

(iii') if $C \in \chi(\mathcal{G})$ and $C \subseteq V_j$, then either $C \in \chi(\mathcal{G}_j)$ or $C$ is contained in only one clique of $\mathcal{G}_j$.

(iv') the fact that $\chi(\mathcal{G}_j) = \emptyset$ does not necessarily imply that $\chi(\mathcal{G}) = \emptyset$.

To check if $\chi(\mathcal{G}_j) \not\subseteq \chi(\mathcal{G})$ for any $j \in \{1, 2, \cdots, m\}$, suppose that $C \in \chi(\mathcal{G}_j)$ and $C \notin \chi(\mathcal{G})$. This implies, by Lemma 4.2, that $C$ is a separator but not a prime separator in $\mathcal{G}$. Thus, there is a proper subset $C'$ of $C$ in $\chi(\mathcal{G})$. By (iii'), $C' \in \chi(\mathcal{G}_j)$ or is contained in only one clique of $\mathcal{G}_j$. However, neither is possible, since $C' \subset C \in \chi(\mathcal{G}_j)$ and $C$ is an intersection of cliques of $\mathcal{G}_j$. Therefore,

$$\chi(\mathcal{G}_j) \subseteq \chi(\mathcal{G}) \quad \text{for all } j.$$

This proves result (i) of the theorem.

We will now prove result (ii). If $\cup_{i=1}^{m}\chi(\mathcal{G}_i) = \emptyset$, then, since all the $\mathcal{G}_i$'s are decomposable by Theorem 4.3, they are complete graphs themselves. So, by definition, the maximal CMS must be a complete graph of $V$. Thus, the equality of the theorem holds.

Next, suppose that $\cup_{i=1}^{m}\chi(\mathcal{G}_i) \neq \emptyset$. Then there must exist a marginal model structure, say $\mathcal{G}_j$, such that $\chi(\mathcal{G}_j) \neq \emptyset$. Let $A \in \chi(\mathcal{G}_j)$. Then, by Theorem 4.5, $A$ is either a prime separator or embedded in a clique of $\mathcal{G}_i$ if $A \subseteq V_i$ for $i \neq j$. Since a prime separator is an intersection of cliques by equation (6), the prime separator itself is a complete subgraph. Thus, by the definition of maximal CMS and by Lemma 4.2, $A \in \chi(\mathcal{H})$. This implies that $\cup_{i=1}^{m}\chi(\mathcal{G}_i) \subseteq \chi(\mathcal{H})$.

To show that the set inclusion in the last expression comes down to equality, we will suppose that there is a set $B$ in $\chi(\mathcal{H}) \setminus (\cup_{i=1}^{m}\chi(\mathcal{G}_i))$ and show that this leads to a contradiction to the condition that $\mathcal{H}$ is a maximal CMS. $\mathcal{H}$ is decomposable by Theorem 4.4. So, $B$ is the intersection of the cliques in $\mathcal{C}_{\mathcal{H}}(B)$ which is defined right before Lemma 4.2. By supposition, $B \notin \chi(\mathcal{G}_i)$ for all $i = 1, \cdots, m$. This means either (a) that $B \subseteq V_j$ for some $j$ and $B \subseteq C$ for only one clique $C$ of $\mathcal{G}_j$ by Corollary 4.6 or (b) that $B \not\subseteq V_j$ for all $j = 1, \cdots, m$. In both of the situations, $B$ need not be a prime separator in $\mathcal{H}$, since the $\mathcal{G}_i$ are decomposable and so $B \cap V_i$ are complete in $\mathcal{G}_i$ in both of the situations. In other words, edges may be added to $\mathcal{H}$ so that $\mathcal{C}_{\mathcal{H}}(B)$ becomes a clique, which contradicts that $\mathcal{H}$ is a maximal CMS. This completes the proof. □

## References

Andersson, S.A., Madigan, D. and Perlman, M.D.(1997). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs, *Scandinavian Journal of Statistics*, **24**, 81-102.

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, *The Annals of Statistics*, **30**(1), 140-159.

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, *Scandinavian Journal of Statistics*, **26**, 209-220.

Darroch, J.N., Lauritzen, S.L. and Speed, T.P. (1980). Markov fields and log-linear interaction models for contingency tables, *The Annals of Statistics*, **8(3)**, 522-539.

Dawid, A.P. (1979). Conditional independence in statistical theory, *Journal of Royal Statistical Society*, B, **41(1)**, 1-31.

Didelez, V. and Edwards, D. (2004). Collapsibility of graphical CG-regression models. *Scandinavian Journal of Statistics*. To appear.

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, *Journal of The American Statistical Association*, **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, *Journal of The Australian Mathematical Soc. A*, **36**, 30-52.

Lauritzen, S.L., Speed, T.P. and Vijayan, K. (1984). Decomposable graphs and hypergraphs, *Journal of the Australian Mathematical 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, *Journal of Royal Statistical Soc.* B, **50(2)**, 157-224.

Leimer, H.-G. (1989). Triangulated graphs and marked vertices, in *Graph theory in memory of G. A. Dirac* (eds: L. D. Andersen et al.). *Annals of Discrete Mathematics*, **41**, 311-324.

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.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd ed. Cambridge, MA: MIT Press.

Wermuth, N. (1980). Linear recursive equations, covariance selection, and path analysis. *Journal of The American Statistical Assoc.*, **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