London Math Society Durham Symposium on
Mathematical Aspects of Graphical Models.
June 30 - July 10, 2008

Markovian Combination of Decomposable Model Structures: MCMoSt

Sung-Ho Kim & Sangjin Lee

Department of Mathematical Sciences
Korea Advanced Institute of Science and Technology

Contents

1 Introduction

- **Problem and motivation**

1. **Consider a problem of developing a graphical model of 30 item score variables (X's) and 20 or more cognitive/knowledge state variables (U's).**
2. **We use the model for diagnosing knowledge states where knowledge states are predicted via conditional probabilities.**
3. **X's are observable and U's are not, and assume they are all binary.**
4. **Cognitive diagnosis and model structure**

- **Two common sense issues in large-scale modeling:**

1. **Sparseness of data (Koehler, 1986; Maydeu-Olivares and Joe, 2005; Kim(in revision))**
2. **Model/time complexity (Chickering, 1996)**

- Koehler, K.J., 1986. Goodness-of-fit tests for log-linear models in sparse contingency tables. JASA, 81(394), 483-493.
- Maydeu-Olivares, A., Joe, H., 2005. Limited- and full-information estimation and goodness-of -fit testing in 2^n contingency tables: a unified framework. JASA, 100(471), 1009-1020.
- Chickering, D. (1996). Learning Bayesian networks is NP-complete, In: *Learning from Data*, D. Fisher and H. Lenz (Ed.), 121-130, Springer-Verlag.
- Kim (in revision). Estimate-based goodness-of-fit test for large sparse multinomial distributions.

• Fienberg and Kim (1999) and Kim (2006) 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 joint model structure [Theorems 3 and 4, Fienberg and Kim].

- Fienberg and Kim (1999). Combining conditional log-linear structures, *JASA*, 445(94), 229-239.
- Kim (2006). Conditional log-linear structures for log-linear modelling, CSDA, 50(8), 2044-2064.

But not between marginal and joint models!

• **Consider two marginal models,** [12][23] **and** [12][24]**, which are possible from each of**

$$[12][24][23],\ [12][24][34],\ [12][23][34],\ [12][234].$$

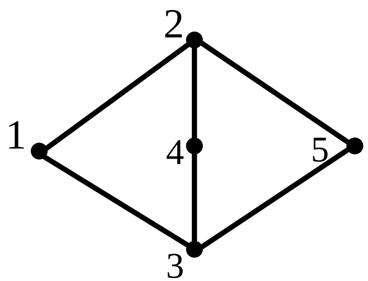
• **How can we find a joint model structure from a given set of marginal model structures?**

2 Preliminaries

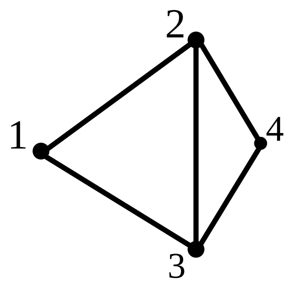
- **We will consider only undirected graphs.**
- **For a subset $A \subseteq V$, we denote by $\mathcal{G}_A = (A, E_A)$ the subgraph of $\mathcal{G} = (V, E)$ confined to A where**

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

We will call $\mathcal{G}_A$ the Markovian subgraph of $\mathcal{G}$ confined to A.



(a)



(b)

- **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 $\mathcal{G}' \subseteq^e \mathcal{G}$.**

- **According to the definition of a decomposable graph, 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 $S_j = C_j \cap C_{(j-1)}$,**

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

- **We denote the collection of these S_j's by $\chi(\mathcal{G})$.**

- Lauritzen (1996). *Graphical Models.* Oxford: Oxford University Press.

• The cliques are elementary graphical components and the S_j is obtained as intersection of neighboring cliques. We will call the S_j's prime separators of the decomposable graph $\mathcal{G}$.

• Prime graphs are defined as the maximal subgraphs without a complete separator in Cox and Wermuth(1999).

• The prime separators in a decomposable graph may be extended to separators of prime graphs in any undirected graph.

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

- **We will denote by $M(\mathcal{G})$ the collection of the distributions that are globally Markov with respect to $\mathcal{G}$, i.e., if, for three disjoint subsets A, B, C of V, $\mathbf{X}_A$ and $\mathbf{X}_B$ are conditionally independent given $\mathbf{X}_C$, then A and B are separated by C in $\mathcal{G}$.**

3 Distribution and interaction graphs

- **For a distribution P, let $\mathcal{G}(P)$ be the interaction graph of P.**

Theorem 1. *(Corollary 3.4 in Kim (2004))* **For a distribution P of $\mathbf{X}_V$ and $A \subseteq V$,**

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

- **For a collection $\mathcal{V}$ of subsets of V, let**

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

Theorem 2. *(Theorem 3.6 in Kim(2004))* **For a collection $\mathcal{V}$ of subsets of V with an undirected graph $\mathcal{G}$,**

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

- Kim (2004). Combining decomposable model structures, RR 04-15, Division of Applied Mathematics, KAIST.

4 Combined model structures

- **Let $\mathcal{G} = (V, E)$ be decomposable and let $V_1, V_2, \cdots, V_m$ be subsets of V. For simplicity, we write $\mathcal{G}_i = \mathcal{G}_{V_i}$.**

Definition 3. 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 $i = 1, \cdots, m$.

- **Let $\mathcal{C}_{\mathcal{G}}(A)$ denote the collection of the cliques which include nodes of A in the graph $\mathcal{G}$.**

Lemma 4. 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 $W' \in \mathcal{C}_{\mathcal{G}}(C)$, $\langle W|B|W'\rangle_{\mathcal{G}}$.

Theorem 5. *(Kim, 2006)* **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 $\mathcal{H}$,**

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

Theorem 6 (Unique existence). *(Kim, 2006)* **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 $\mathcal{G} \subseteq^e \mathcal{H}^*$.**

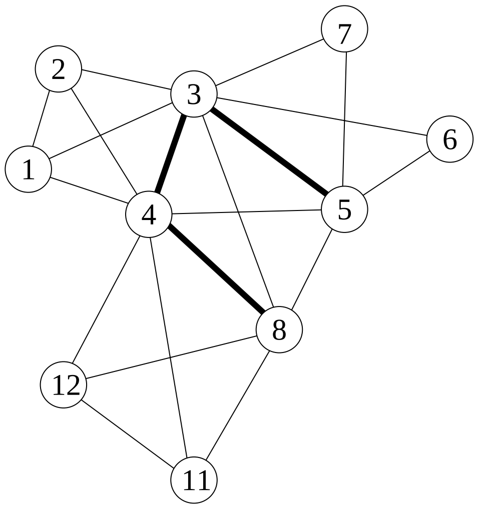
- Kim. (2006). Properties of Markovian subgraphs of a decomposable graph, LNAI 4293, 15-26.

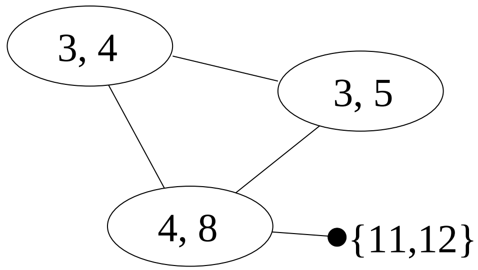
Theorem 7 (Invariance of PS)**. Let $\mathcal{G}$ be a decomposable graph and $\mathcal{G}_1$ and $\mathcal{G}_2$ be Markovian subgraphs of $\mathcal{G}$. Suppose that a set $C \in \chi(\mathcal{G}_1)$ and that $C \subseteq V_2$. Then C is not intersected in $\mathcal{G}_2$ by any other subset of V_2.**

- **We will call a node PS-node if it is a component of a PS; otherwise, a non-PS-node.**

5 Graph of prime separators (GOPS)

Definition 8. Let $A = \cup_{a \in \chi(\mathcal{G})} a$. Then the graph of the prime separators (GOPS for short) of $\mathcal{G}$ is obtained from $\mathcal{G}_A$ by replacing every PS and all the edges between every pair of neighboring PSs in $\mathcal{G}_A$ with a node and an edge, respectively.





6 Algorithm

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

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

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

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

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

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

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

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

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

the separateness condition, and similarly for each of a^2 that satisfy (2). We denote the result of the combination by η_2.

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

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

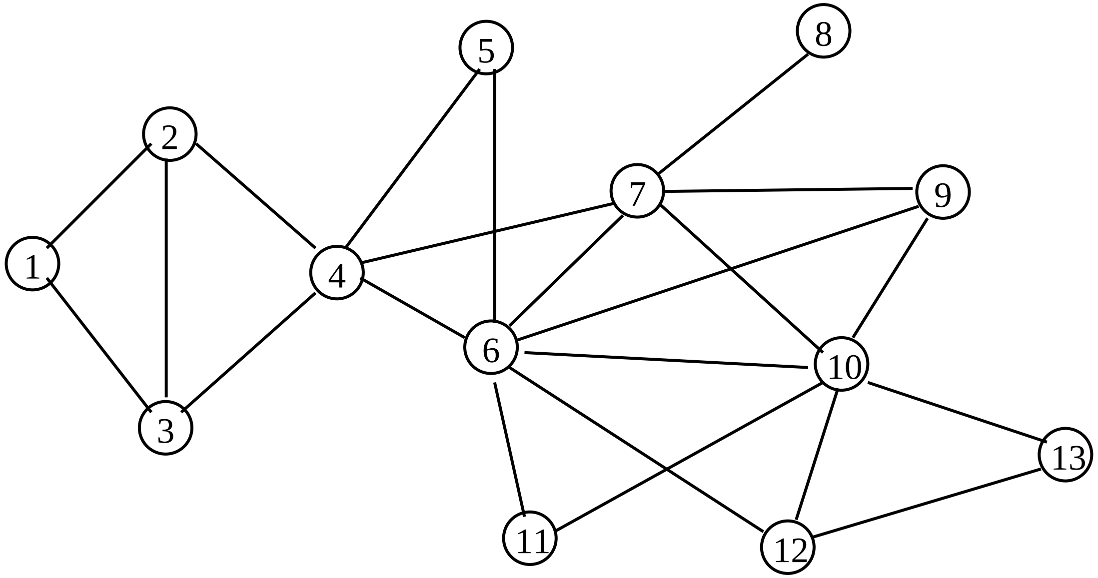


Figure 1: A model of 13 variables.

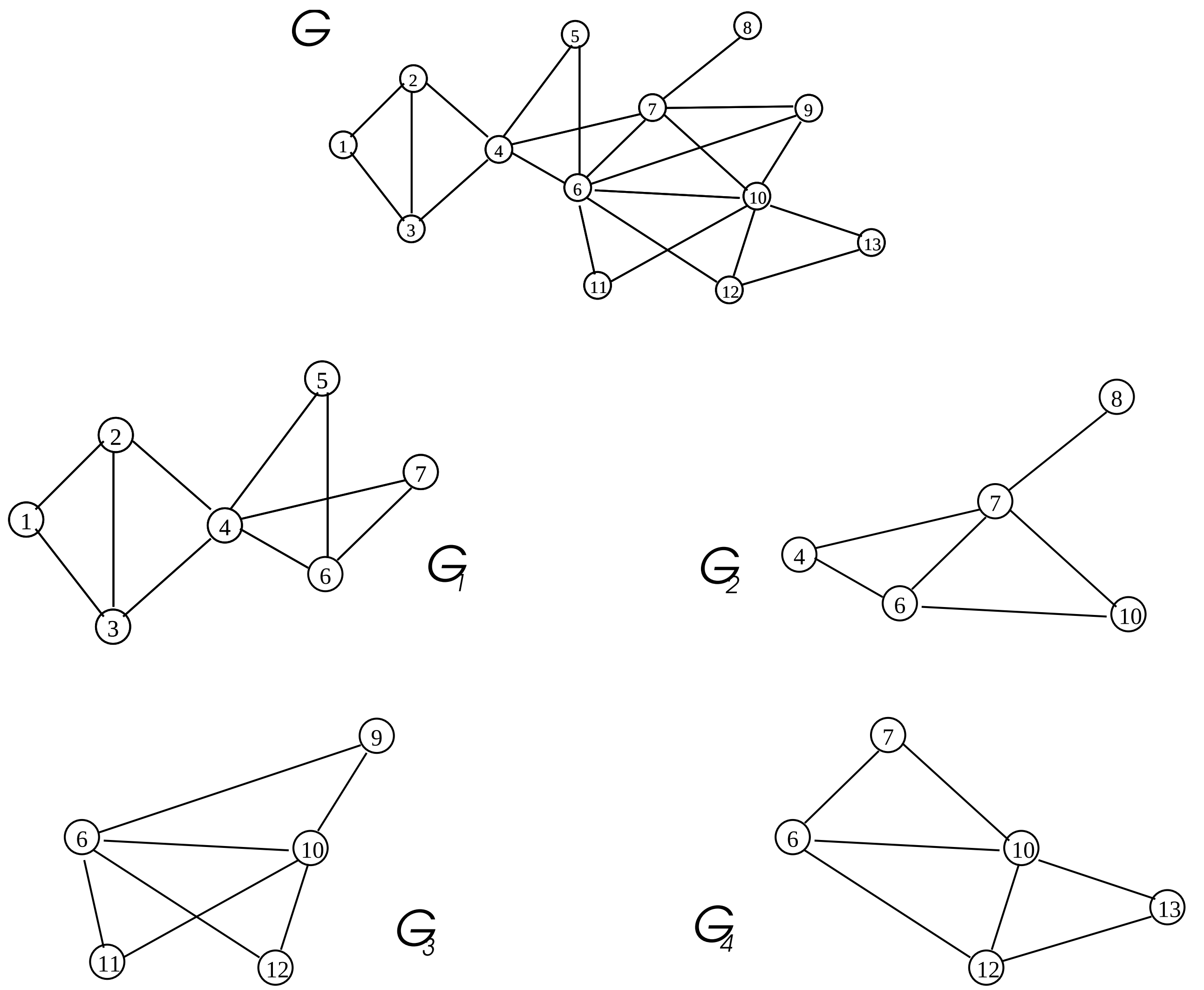


Figure 2: Marginals of the model in Figure 1.

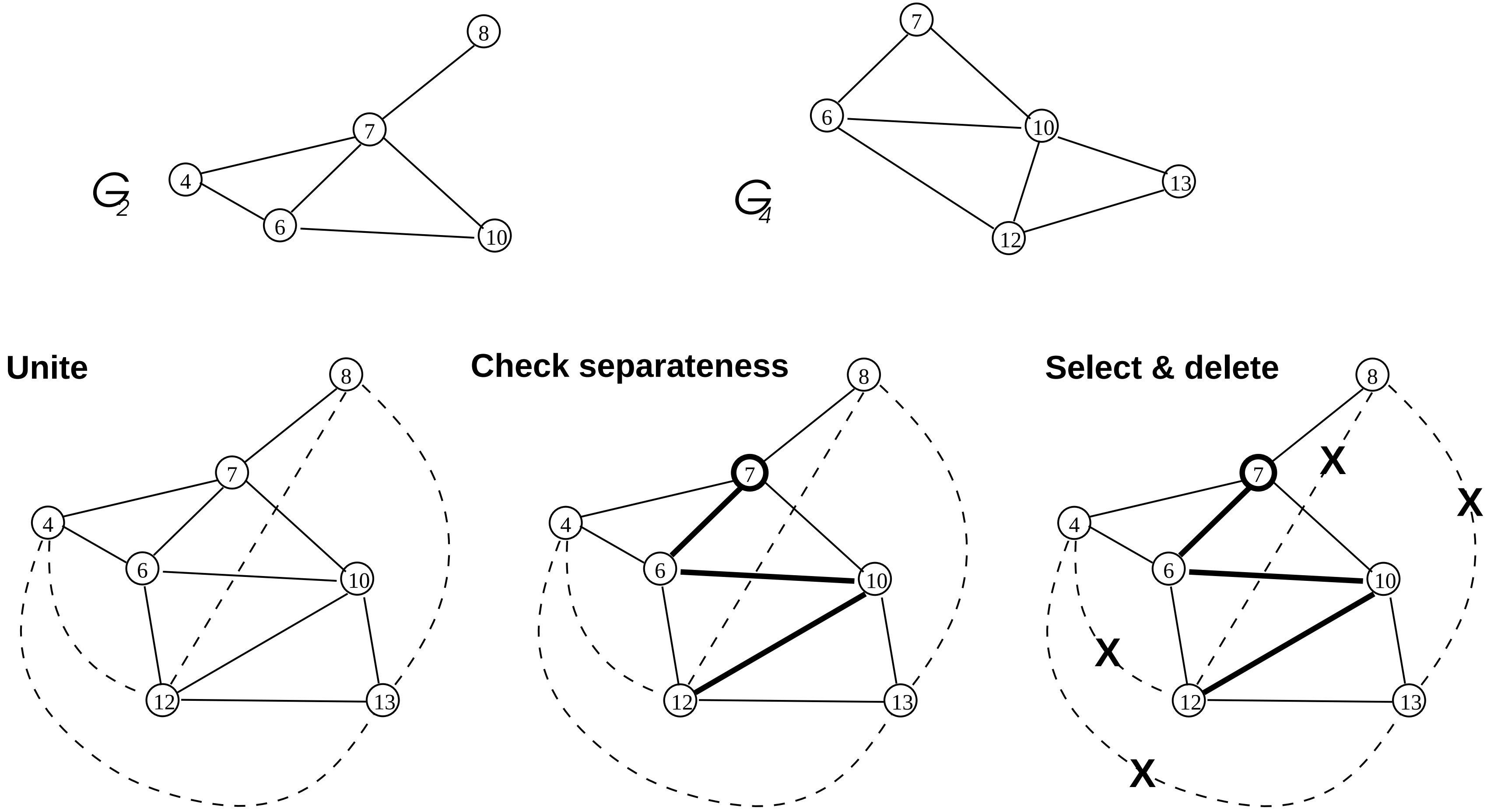


Figure 3: Combining $\mathcal{G}_2$ and $\mathcal{G}_4$ of the model in Figure 1.

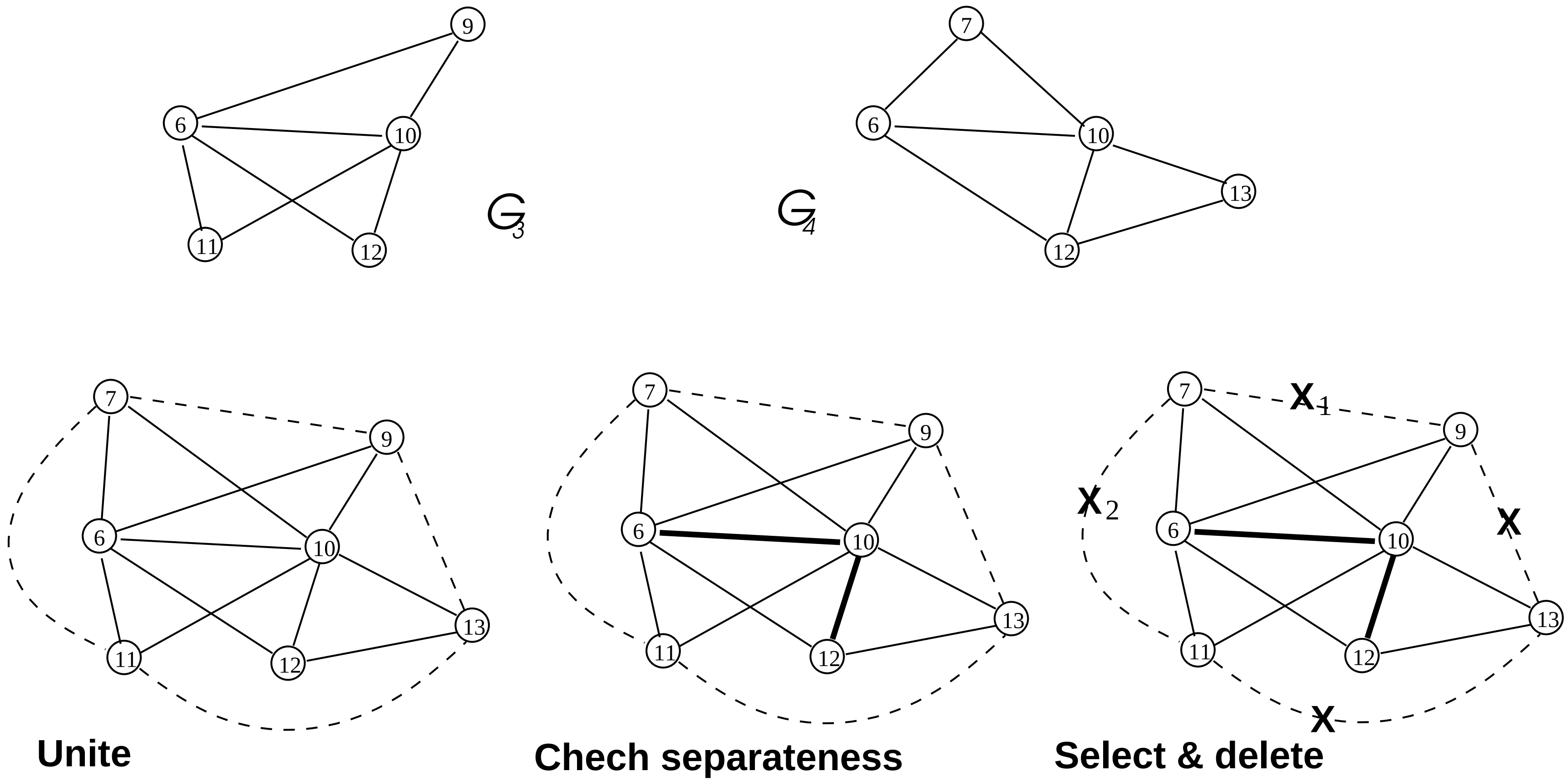
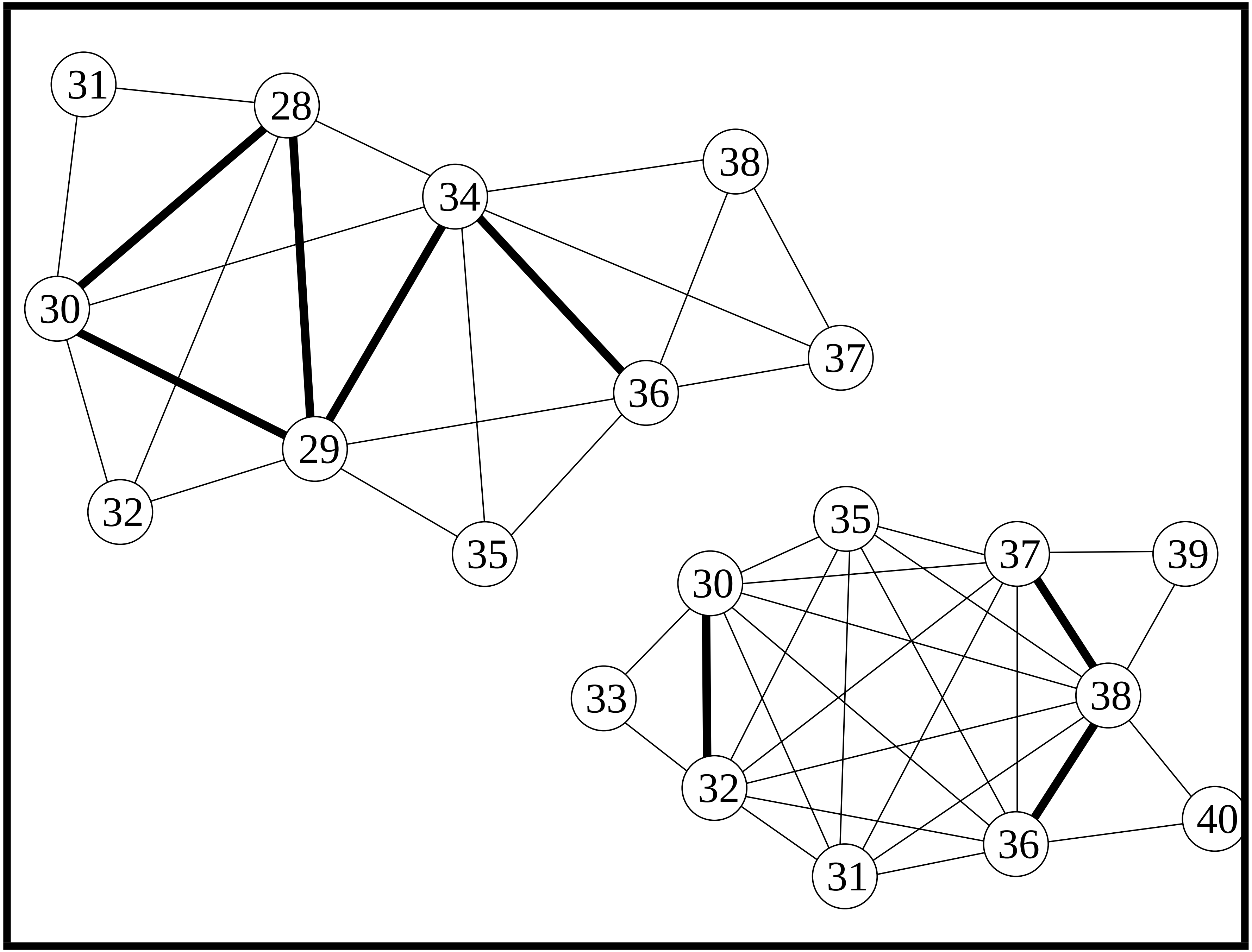


Figure 4: Combining $\mathcal{G}_3$ and $\mathcal{G}_4$ of the model of 13 variables.


31
28
38
34
30
37
36
29
32
35
35
37
39
30
33
38
32
40
36
31

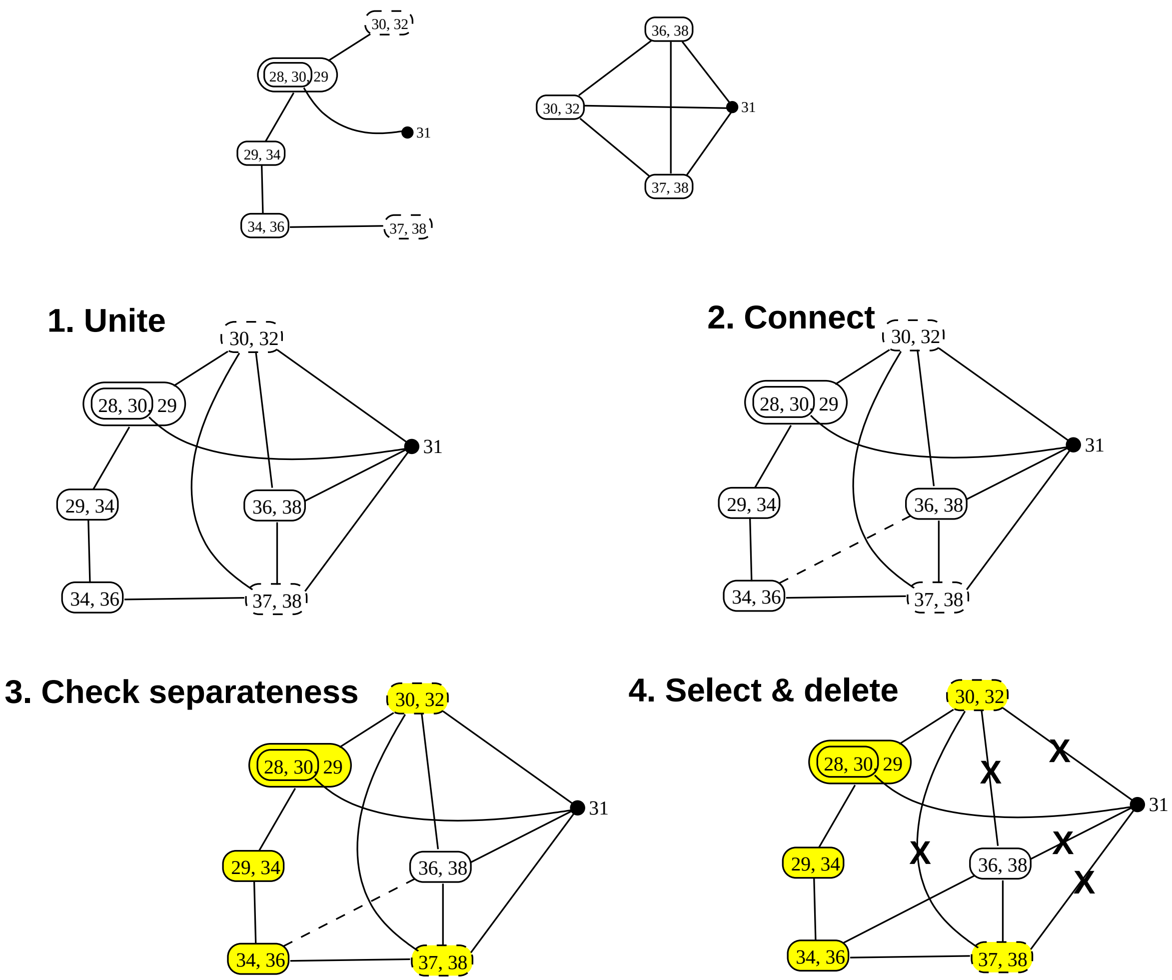


Figure 5: Combining $xGOPS_5$ and $xGOPS_6$ of the model of 40 variables.

7 Time complexity of the procedure

- **For two graphs, $\mathcal{G}_1$ and $\mathcal{G}_2$, let $|V_i| = n_i$ with $i = 1, 2$, $|V_1 \cap V_2| = n_{12}$ and $\tilde{n}_i = n_i - n_{12}$.**
- **It is well known that the time complexity of the depth-first search method (Tarjan, 1972) for a graph $\mathcal{G} = (V, E)$ is of order $O(|V| + |E|)$.**
- **So the time complexity for the combination is of order $\tilde{n}_1^2 O(\tilde{n}_2 + \tilde{e}_2) + \tilde{n}_2^2 O(\tilde{n}_1 + \tilde{e}_1)$, where $\tilde{e}_i$ is the number of edges in the induced subgraph of $\mathcal{G}_i$ on $V_i \setminus V_{3-i}$.**

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

8 Illustration-1

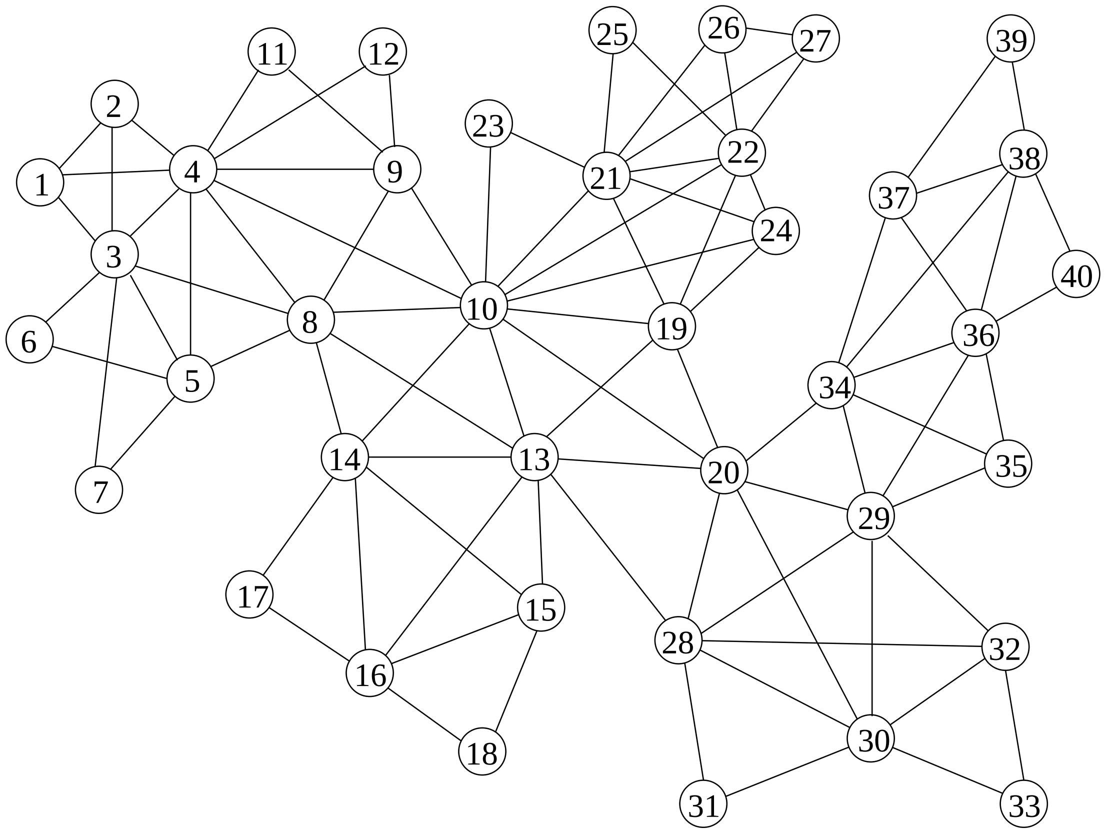


Figure 6: A model of 40 binary variables

	1	2	3	4	5	6	7	8	11	12	9	16	17	18	10	14	15	23	19	24	21	13	22	20	25	26	27	29	28	34	30	31	32	35	38	37	36	33	39	40
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	1	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	0	1	1	0	0	1	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	1	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1	1
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	1	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0

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

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

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

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

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

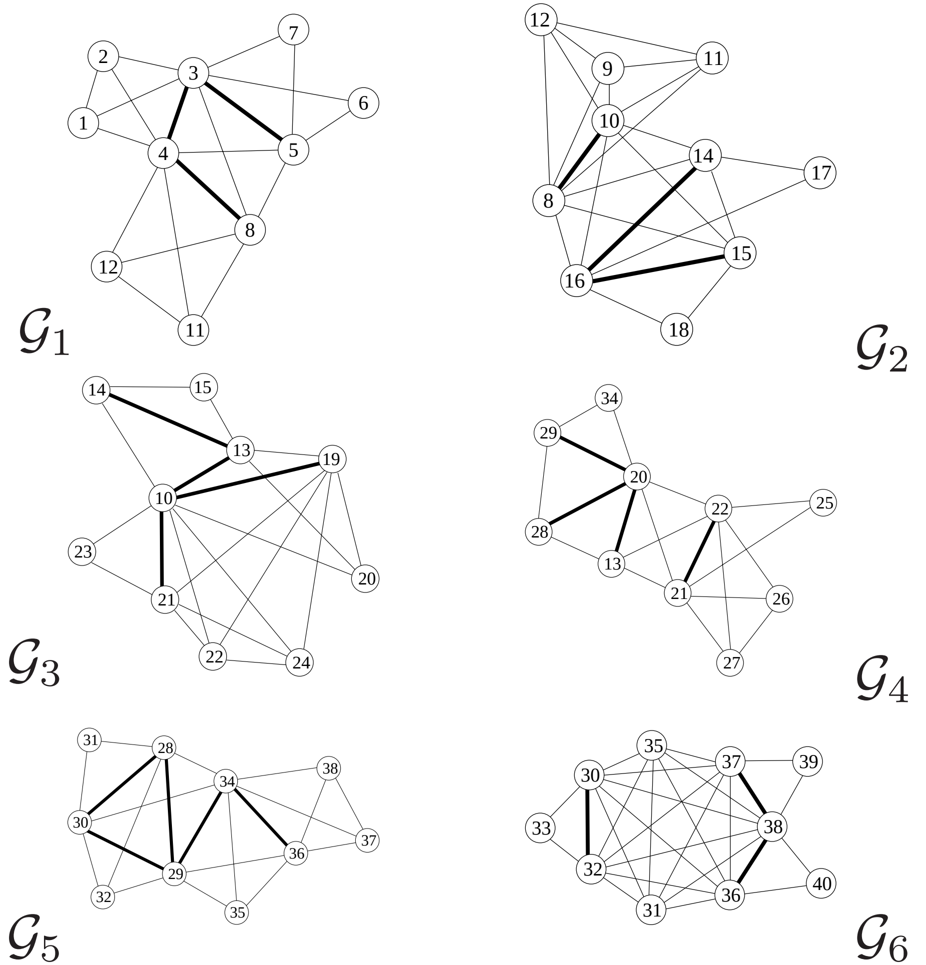


Figure 8: Marginal models of the model in Figure 6 for the 6 subsets of variables which are listed in Table 1. $\mathcal{G}_i$ is the decomposable log-linear model for subset V_i. PSs are represented by thick lines. See Figure 9 for the PSs of the 6 marginal models

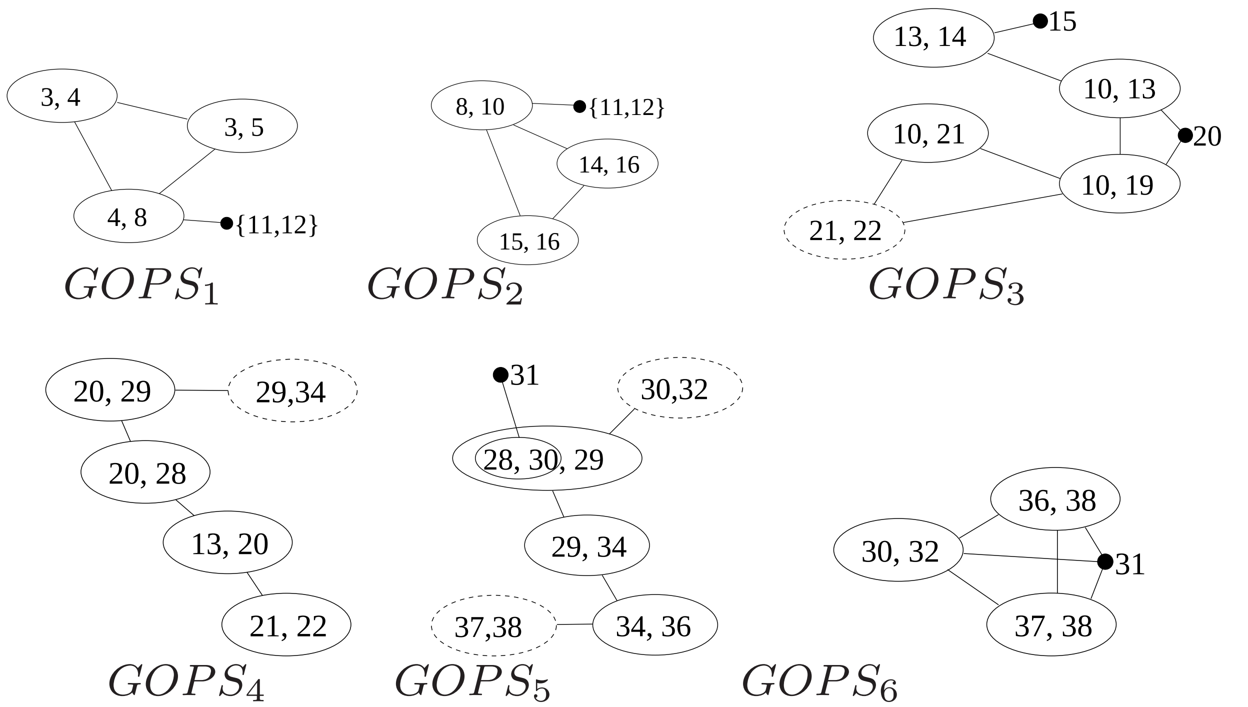


Figure 9: The GOPS's of the six marginal models in Figure 8.

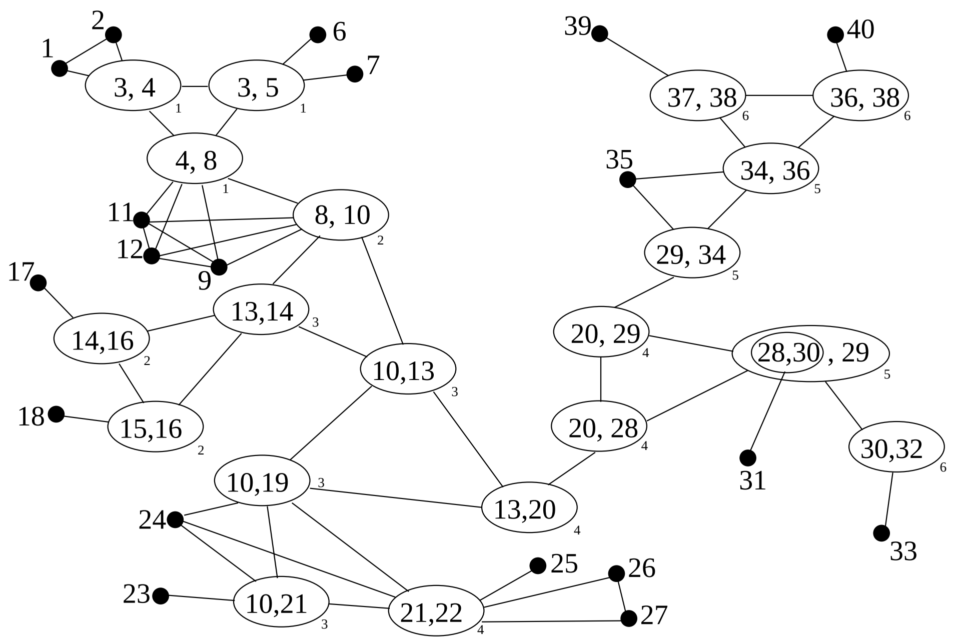


Figure 10: An independence graph of PSs and non-PS variables. The PSs are in ovals and the dots are for the non-PS variables, and the small numbers at the bottom-right of the ovals are the submodel labels of which the ovals are PSs.

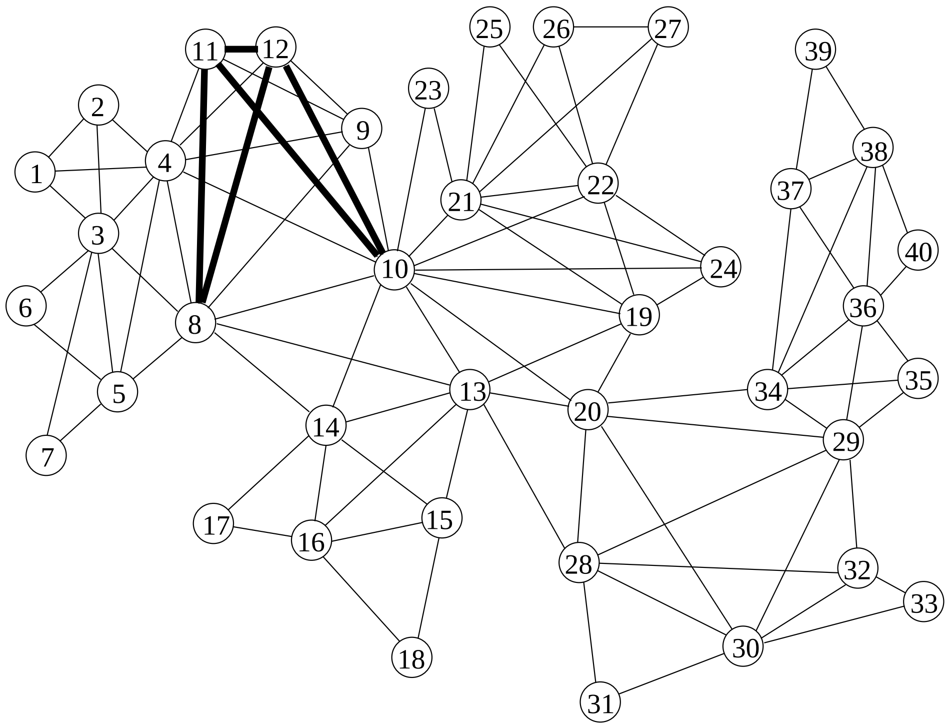
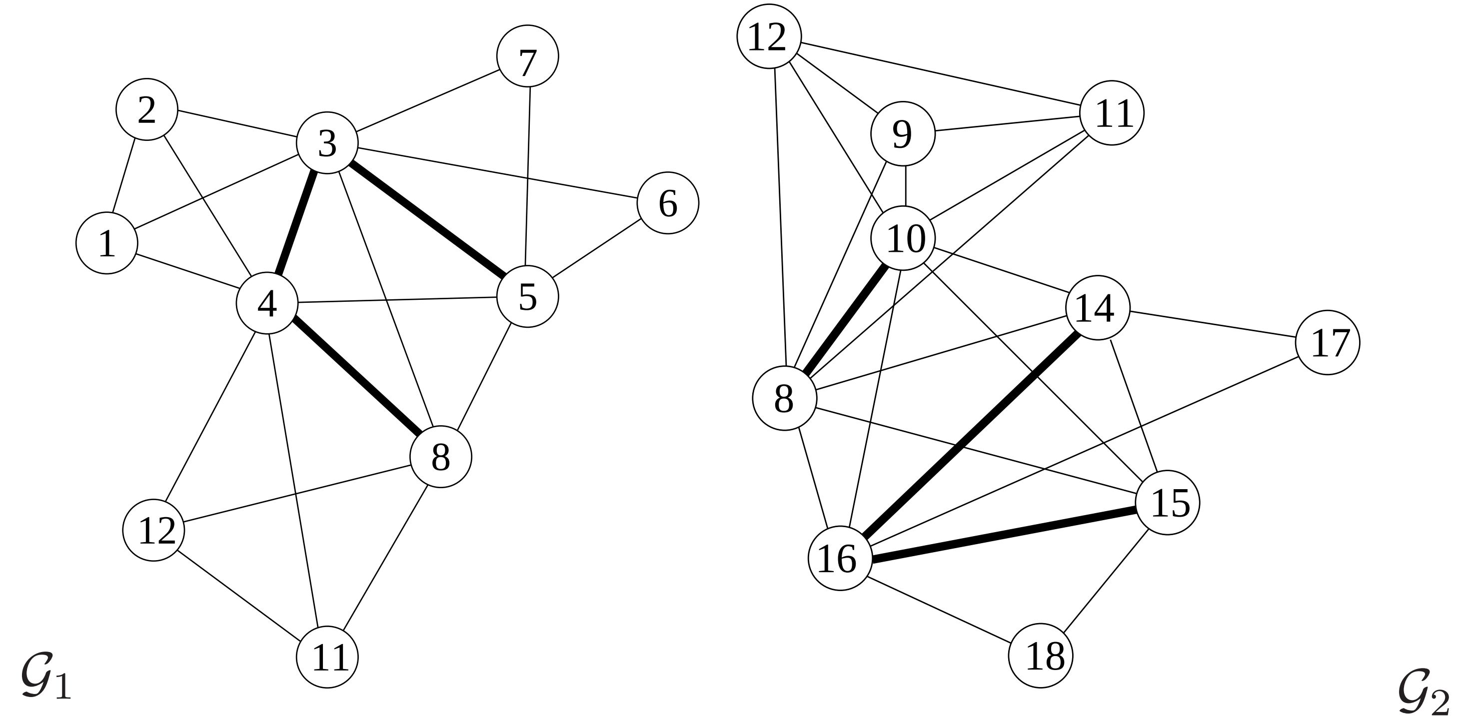


Figure 11: The combined model structure which is obtained from the independence graph in Figure 10. The thick edges are additional to the true model in Figure 6.


2
7
3
6
1
4
5
8
12
11
$\mathcal{G}_1$
12
9
11
10
14
17
8
15
16
18
$\mathcal{G}_2$

9 Illustration-2

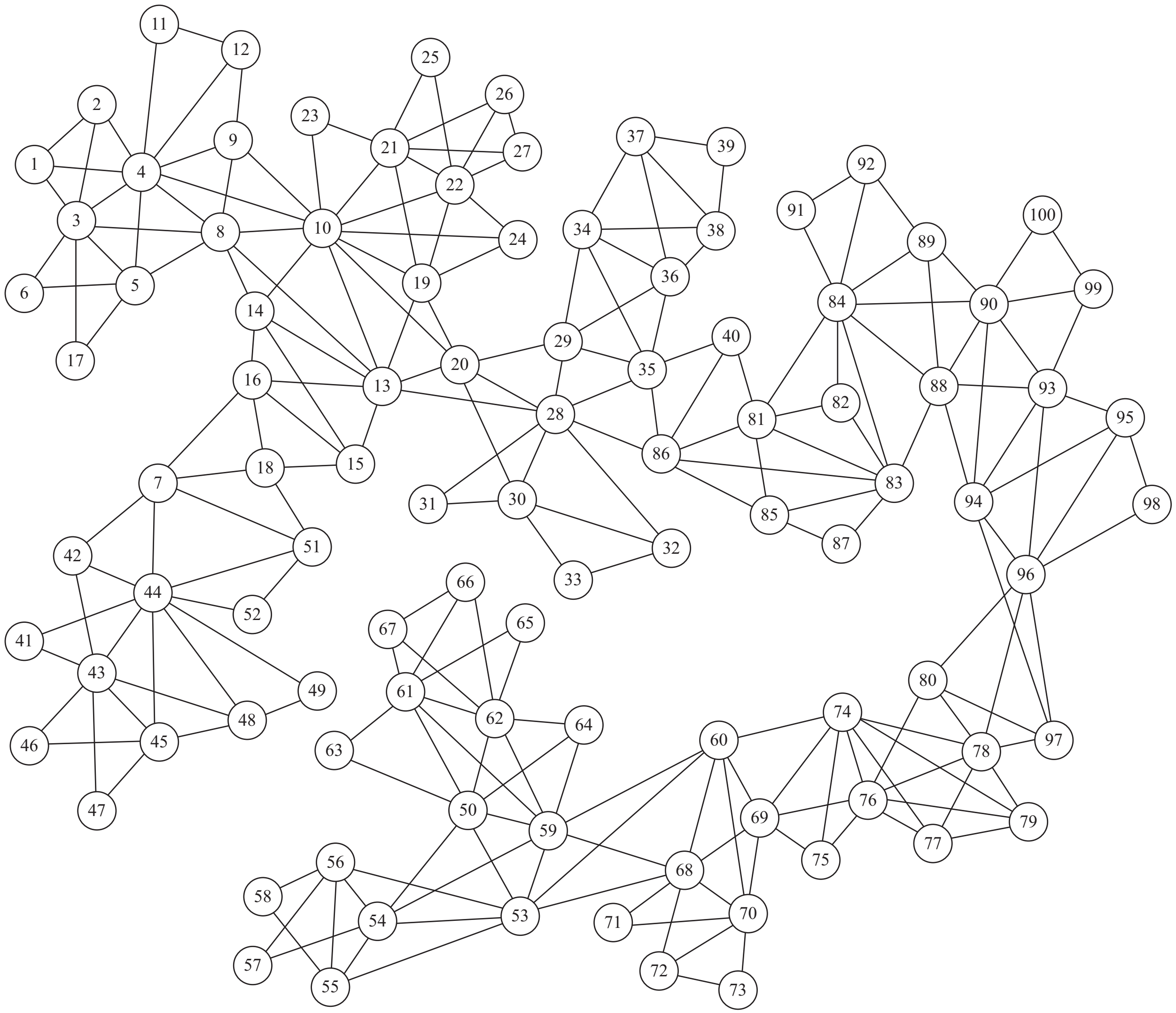


Figure 12: A model of 100 variables

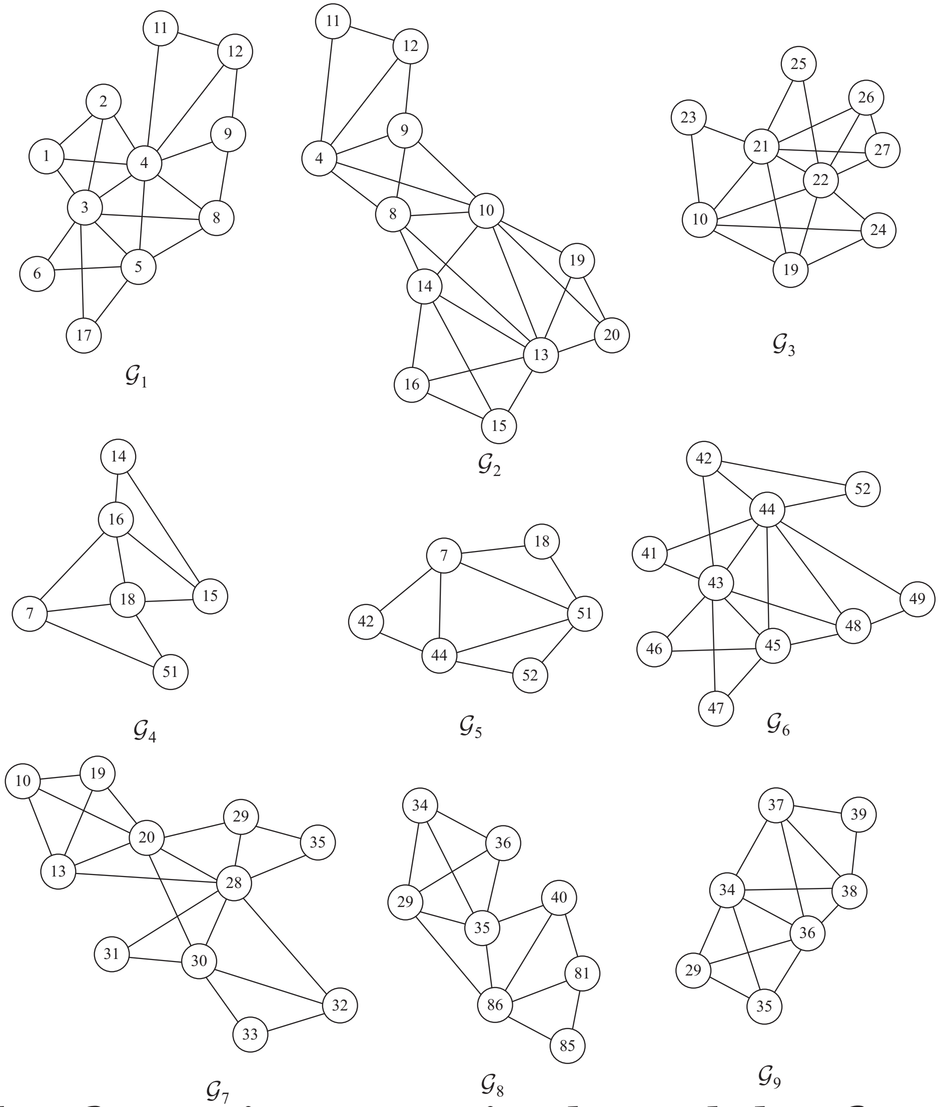


Figure 13: The first nine marginal models, $\mathcal{G}_1, \cdots, \mathcal{G}_9$, of the model in Figure 12.

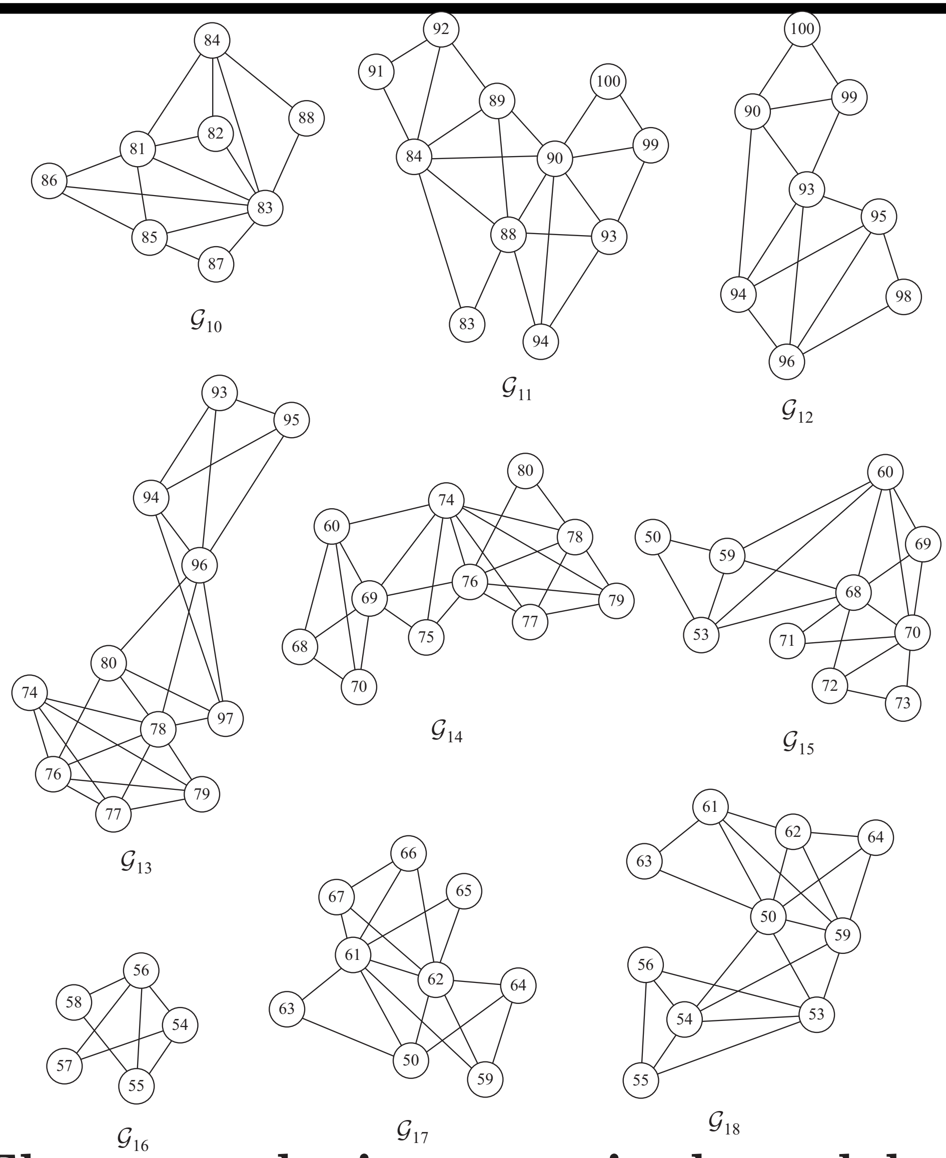


Figure 14: The second nine marginal models, $\mathcal{G}_{10}, \cdots, \mathcal{G}_{18}$, of the model in Figure 12.

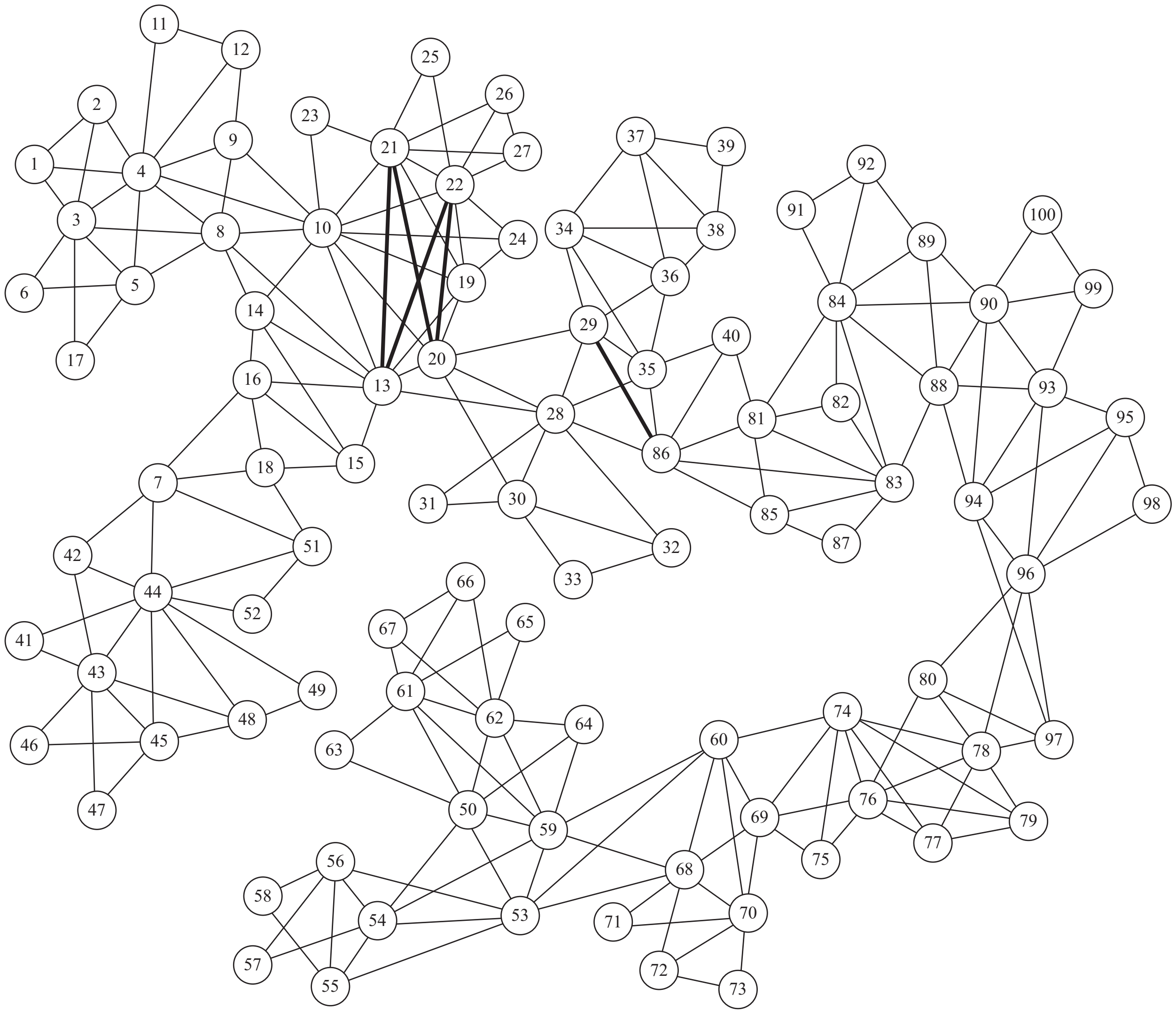


Figure 15: The combined result of the 18 marginal models in Figures 13 and 14. The thick edges are additional to the true model in Figure 12.

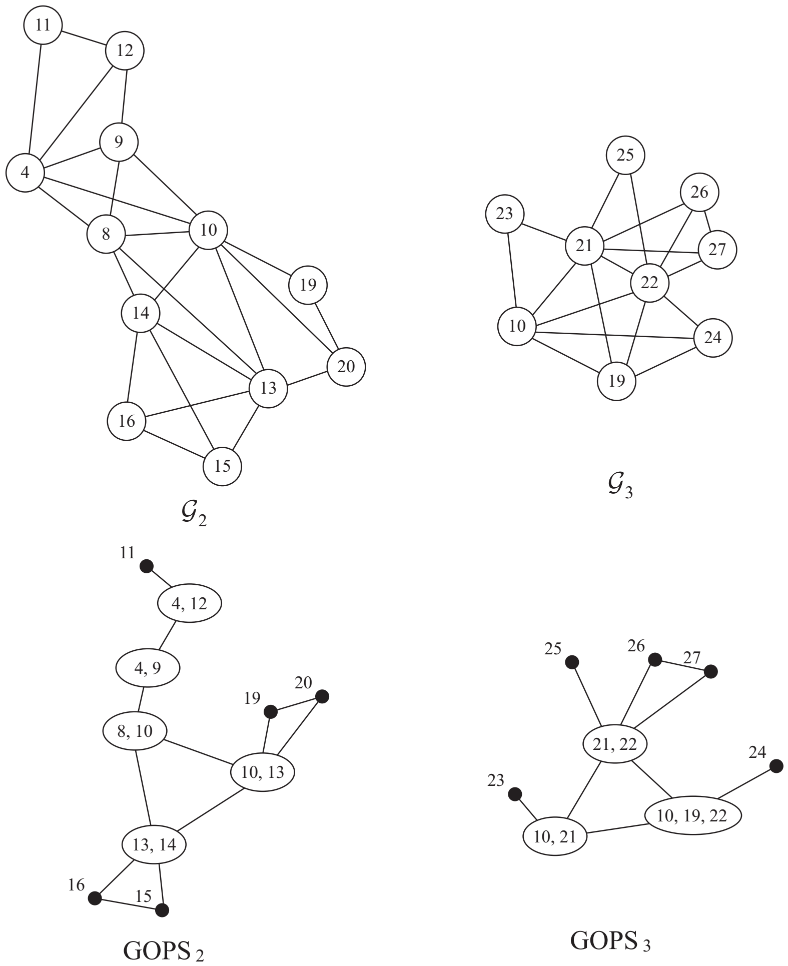


Figure 16: Two marginal models which include nodes 10 and 19.

10 Concluding Remarks

1. **The main idea of MCMoSt is similar to constraint-based learning[CBL] (Meek 1995; Spirtes, Glymour & Scheines, 2000; Neapolitan, 2004) where we construct a Bayesian network based on a list of constraints which are given in terms of conditional independence among a given set of random variables.**

- Meek, C. (1995). Causal influence and causal explanation with background knowledge, UAI 11, 403-410.

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

- Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*, 2nd ed.

But a noteworthy difference between the two is that, while the statements of conditional independencies, as for the CBL, are an extraction from the probability model of the whole set of the variables involved, the statements for MCMoSt are from the marginal probability models of the subsets of variables.

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

3. **Although the model combination is carried out under the decomposability assumption, we can deal with the marginal models of a graphical model, which are not decomposable, by transforming their model structures into decomposable (i.e., triangulated) graphs.**

11 Where to go?

- **Combining model structures other than UG's.**
- **Models with no observed data.**
- **Consistency of models (Dawid and Lauritzen, 1993)**
- **Robustness of prediction/classification (Kim, 2005; Fienberg and Kim, 2007)**

- Dawid and Lauritzen (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *The Annals of Statistics*, 21, 3, 1272-1317.
- Kim (2005). Stochastic ordering and robustness in classification from a Bayesian network, Decision Support Systems, 39 (3), 253-266.
- Fienberg and Kim (2007). Positive association among three binary variables and cross-product ratios, Biometrika 94(4), 999-1005.

Thank You!