Article

Undirected Structural Markov Property for Bayesian Model Determination

Xiong Kang, Yingying Hu and Yi Sun *

College of Mathematics and System Science, Xinjiang University, Urumqi 830046, China
* Correspondence: brian@xju.edu.cn

Abstract: This paper generalizes the structural Markov properties for undirected decomposable graphs to arbitrary ones. This helps us to exploit the conditional independence properties of joint prior laws to analyze and compare multiple graphical structures, while being able to take advantage of the common conditional independence constraints. This work provides a theoretical support for full Bayesian posterior updating about the structure of a graph using data from a certain distribution. We further investigate the ratio of graph law so as to simplify the acceptance probability of the Metropolis–Hastings sampling algorithms.

1. Introduction

A probabilistic graphical model (PGM) or a structured probabilistic model (SPM) is a statistical model that consists of a graph and a distribution family for which the graph encodes the conditional independence information between random variables. Such models always associate with independence models, arise naturally in multivariate analysis and can provide certain versatility and convenience in analyzing complex data with large scales, while independence models are the sets of conditional independence constraints encoded by graphs via the global Markov property.

It is known that different classes of graphs with different interpretations of independence have been developed in the past decades, and the reader can refer to [1–4] for details. One of the most important classes of graphs in graphical models is undirected graphs (UGs). Their corresponding Markov models are often known as undirected graphical models or Markov networks [1,2]. These models have been found to have many applications in a wide range of areas such as econometrics, medical science, artificial intelligence [5–7] and so on. Our research in this paper is related to the work in the area of the structure determination of these models with the Bayesian method.

The main objective of Bayesian structure learning is to learn the structure of a graph from data. Meanwhile, Bayesian structure learning requires a clear illustration of a prior distribution about graphical structures, which is termed as a graph law. Statisticians have proposed some approaches to calculate the prior law of a graph. The simplest graph law is the uniform distribution in [8]. Additionally, the Erdős-Rényi random graph model is also used to indicate the graph law in [9]. Furthermore, a characterization of graph law with the form of exponential family is proposed by [10]. However, how to simplify this prior law is a significant task for us, especially in the posterior inference of graphical structures. In view of this, the structural Markov property is first proposed for the purpose of characterizing the conditional independence of the structure of a graph. The structural Markov properties require that the structures of distinct components of graphs are conditionally independent given the existence of a separating component; see [10]. These properties reflect the conditional independence at the structural level. It has been proved that a graph law is

structural Markov if and only if it is a member of the clique exponential family given the support condition as the set of decomposable undirected graphs; see [10]. Further, a weaker support condition of equivalent characterization for graph laws is given via closure operation of graphical structures in [11].

Indeed, the structural Markov property is an extension of the hyper Markov property, which was proposed in [12] and reflects the global Markov property at the parameter level. These hyper Markov properties are used to describe the conditional independence properties of a distribution of random variables or statistical quantities in graphical models. The hyper Markov laws arise naturally as sampling distributions of maximum likelihood estimators and as prior or posterior distributions in Bayesian inference.

Recently, a weaker version of the structural Markov properties for decomposable graphs was introduced in [13], where the authors provided an analogous clique-separator factorization for the graph law. These weakly structural Markov properties require that the separator is complete. It has been shown that this provides a more flexible family of graph prior laws to use in full Bayesian posterior updating.

It should be pointed out that all the work in [8,10,13] only focuses on decomposable graphical models. However, based on conditional independence and graphical separation, the structural Markov properties might be extended to non-decomposable undirected graphical models. The aim of this paper is filling this gap in the field of graphical models. Further, we focus on a full Bayesian method for the posterior updating of graph laws via the observed data from a certain distribution, and we also prove that this full Bayesian posterior of graph law is feasible and reasonable. Finally, as examples, we illustrate our theory with detailed investigations of two significant cases based on the graphical Gaussian models and the multinomial models, respectively.

The outline of this paper is organized as follows. In Section 2, we introduce the terminologies and conceptions used in this paper. Section 3 first investigates the structural Markov properties for non-decomposable graphs, and then exploits the joint prior laws of a random sample distribution for full Bayesian inference. Section 4 gives two examples such as the inverse Wishart distribution and the Dirichlet distribution to study the posterior updating of graph laws in details. Further, we discuss some details about the computation for the structural Markov graph laws in Section 5. Finally, in Section 6, we give the conclusion of this paper.

2. Preliminaries

For terms and symbols, we follow the references [10,12] as many theoretical frameworks of this paper are constructed and developed based on them. Several concrete notions and terminologies used in this paper will be given in the following for clarity and consistency.

2.1. Graphical Terminologies and Notation

A graph $G = (V, E)$ consists of a finite set of vertices $V(G) = \{v_1, v_2, \ldots, v_p\}$ and the set of edges $E(G) \subseteq V(G) \times V(G) = \{(u, v) : u, v \in V(G)\}$. An edge (u, v) of G is said to be undirected if (u, v) is an unordered pair. A graph G is said to be an undirected graph if all its edges are undirected. Unless otherwise specified, here G is always assumed to be undirected, simple and connected throughout the paper.

For $A \subseteq V(G)$, an induced subgraph of G on A will be denoted by $G_A = (A, E(G_A))$, where $E(G_A) = \{(u, v) \in E(G) : u, v \in A\}$. All subgraphs in this paper are induced subgraphs. A is complete (or a clique) if any two different vertices $u, v \in A$ are adjacent, i.e., $(u, v) \in E(G)$. A graph is a clique if its vertex set is a clique. For $A, B \subseteq V(G)$, a clique G_A is a maximal clique if G_B is incomplete for any superset $B \supset A$. Two vertices u and v are considered to be neighbors if $(u, v) \in E(G)$. For $A \subseteq V(G)$, the boundary $bd(A)$ is the set of vertices in $V(G) \setminus A$ that are neighbors of vertices in A. G can be collapsible onto A if every connected component of $V(G) \setminus A$ has a complete boundary in G.

For any subsets A, B and C of $V(G)$, we say that C separates A from B, and write $A \perp\!\!\!\perp B | C[G]$, if any path in G between some $u \in A$ and $v \in B$ contains a vertex in C. Usually, we call C a separator of A and B. Separators that are cliques are called clique separators.

For any disjoint subsets A, B and S of $V(G)$, we say (A, B, S) forms a decomposition if (i) $A \cup B \cup S = V(G)$; (ii) $A \perp\!\!\!\perp B | S[G]$, and (iii) S is a clique separator in G. A decomposition (A, B, S) is said to be proper if the sets $A \cup S$ and $B \cup S$ are both proper subsets of $V(G)$.

Definition 1 ([14]). *Let $G = (V, E)$ be an undirected graph. A graph G is reducible if its vertex set contains a clique separator, otherwise G is said to be prime. E.g., G is prime if G is a clique, while G is reducible if G is a disconnected graph. An induced subgraph G_U is a maximal prime subgraph of G if it satisfies*

(i) G_U is prime, and

(ii) $\forall W \subseteq V(G)$ s.t. $U \subset W$, G_W is reducible.

In Figure 1, it is easy to find that G_1 is prime since there is no clique separator in G_1. However, G_2 is reducible because of a clique separator $S = \{a, c\}$ in G_2.
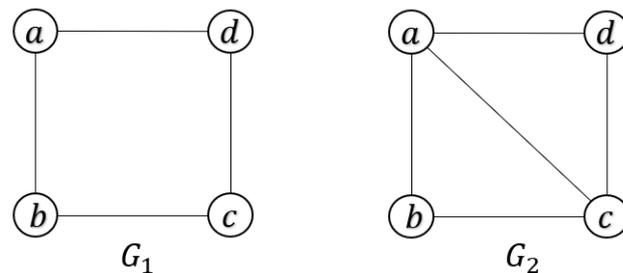

Figure 1. G_1 is a prime graph and G_2 is a reducible graph.

Definition 2 ([14]). *A proper decomposition (A, B, S) of an undirected graph G is stated to form a prime decomposition if $G_{A \cup S}$ and $G_{B \cup S}$ are prime, or $G_{A \cup S}$ and $G_{B \cup S}$ can be recursively decomposed into pairwise different maximal prime subgraphs of G.*

In particular, G is decomposable if $G_{A \cup S}$ and $G_{B \cup S}$ are complete, or they are both decomposable subgraphs of G. Note that the prime decomposition of arbitrary undirected graphs is a generalization of that of chordal graphs. For instance, in Figure 2, G is a non-decomposable undirected graph with $V(G) = \{a, b, c, d, e\}$, which involves two maximal prime subgraphs G_{U_1} and G_{U_2}, with $U_1 = \{a, b, c\}$ and $U_2 = \{b, c, d, e\}$, respectively, and a clique separator $S = U_1 \cap U_2 = \{b, c\}$. It is obvious that $(\{a\}, \{d, e\}, \{b, c\})$ forms a prime decomposition of G. Additionally, we find that G_{U_1} is complete since all its pairs of vertices are joined, while G_{U_2} is incomplete because the vertices between b and d, or c and e, are not joined.
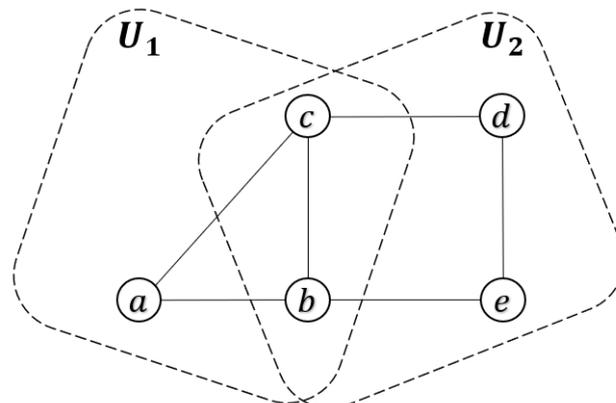

Figure 2. A prime decomposition for an undirected graph G.

It is worthwhile to point out that all the maximal prime subgraphs of an undirected graph can form a perfect sequence in a certain way. If there exists a proper decomposition of an undirected graph G, then G admits a perfect sequence $(U_1, U_2, \ldots, U_k)$ of maximal prime subgraphs, so that for each $j = 2, \ldots, k$, there exists some $h \in \{1, 2, \ldots, j-1\}$, and we have

$$S_j = U_j \cap (\bigcup_{i=1}^{j-1} U_i) \subseteq U_h,$$

where S_j are clique separators actually. Specifically, G is decomposable if its all maximal prime subgraphs are complete (cliques).

In a PGM, a vertex v denotes a random variable X_v, which takes values in a space $\mathfrak{X}_v$. Let $X = X_{V(G)} = (X_v)_{v \in V(G)}$ be a p-dimensional random vector on some product space $\prod_{v \in V(G)} \mathfrak{X}_v$ with P or θ representing its distribution. All the concerned distributions in the present paper are assumed to be positive and closed under marginalization and conditioning with respect to the type of a joint distribution family. For the sake of simplicity, we use $\mathcal{P}$ to represent the set of all positive distributions over X. For $A, B \subseteq V(G)$, θ_A will denote the marginal distribution of X_A and $\theta_{B|A}$ the conditional distribution of X_B given $X_A = x_A$.

Let $\mathfrak{U}$ be the set of undirected graphs with fixed vertex set $V(G)$. A probability distribution of a random graph G, which takes values in $\mathfrak{U}$, is said to be a law, denoted by $\mathfrak{G}$. Further, define $\mathfrak{U}(A, B, S)$ to be the set of undirected graphs for which (A, B, S) is a prime decomposition.

2.2. Independence Model and Collapsibility

Given a finite set N, for $A, B, C \subseteq N$, an independence model, denoted by $\mathcal{I}$, is the set of triplets of the form $\langle A, B | C \rangle$, which are termed as conditional independence statements. A graphical independence model is an independence model induced by a graph. For a graph $G \in \mathfrak{U}$, the graphical independence model of G can be defined as

$$\mathcal{I}(G) = \{ \langle A, B | C \rangle : A \perp\!\!\!\perp B | C [G] \text{ for } A, B, C \subseteq V(G) \}.$$

Obviously, $\mathcal{I}(G)$ is the set of triples $\langle A, B | C \rangle$, encoding its global Markov property over G.

It should be pointed out that the conditional independence of a statistical model in [15,16] shares the same properties of graph separation in [2], i.e., for a graphical independence model $\mathcal{I}(G)$, it has the following properties:

1. for all $A, B \subseteq V(G)$, $\langle A, B | A \rangle \in \mathcal{I}(G)$, $\langle A, B | B \rangle \in \mathcal{I}(G)$ and $\langle A, B | A \cap B \rangle \in \mathcal{I}(G)$;
2. if $\langle A, B | C \rangle \in \mathcal{I}(G)$, then $\langle B, A | C \rangle \in \mathcal{I}(G)$;
3. if $\langle A, B | C \rangle \in \mathcal{I}(G)$, and $U \subseteq A$, then $\langle U, B | C \rangle \in \mathcal{I}(G)$;
4. if $\langle A, B | C \rangle \in \mathcal{I}(G)$, and $U \subseteq A$, then $\langle A, B | C \cup U \rangle \in \mathcal{I}(G)$;
5. if $\langle A, B | C \rangle \in \mathcal{I}(G)$, and $\langle A, W | B \cup C \rangle \in \mathcal{I}(G)$, then $\langle A, B \cup W | C \rangle \in \mathcal{I}(G)$.

In particular, the following property holds when A, B, C are disjoint.

If $\langle A, B | C \rangle \in \mathcal{I}(G)$ and $\langle A, C | B \rangle \in \mathcal{I}(G)$, then $\langle A, B \cup C | \varnothing \rangle \in \mathcal{I}(G)$.

Further, a graphical independence model $\mathcal{I}(G)$ has a natural projection operation on $D \subseteq V(G)$ that

$$\mathcal{I}(G)_D = \{ \langle A, B | C \rangle \in \mathcal{I}(G) : A, B, C \subseteq D \}.$$

It is worthwhile to point out that $\mathcal{I}(G)_D \subseteq \mathcal{I}(G_D)$, where $\mathcal{I}(G_D)$ is the independence model induced by the induced subgraph G_D.

Definition 3 (CI-collapsibility). *Let G be a fixed undirected graph in $\mathfrak{U}$. For $D \subseteq V(G)$, $\mathcal{I}(G)$ can be conditional independence collapsible (CI-collapsible) onto D if $\mathcal{I}(G)_D = \mathcal{I}(G_D)$.*

CI-collapsibility reflects the consistence of conditional independence relations induced by G_D and those induced by G, but constrained on D.

We say a distribution P is Markov with respect to G if for $A, B, C \subseteq V(G)$, it holds that

$$\langle A, B | C \rangle \in \mathcal{I}(G) \Rightarrow X_A \perp\!\!\!\perp X_B | X_C \, [P],$$

where $X_A \perp\!\!\!\perp X_B | X_C[P]$ represents the assertion that X_A is independent of X_B given X_C under P.

In order to ensure that various distributions and those of statistical quantities are Markov with respect to G, we now are in a position to review the graphical models within the framework of undirected graphs. A graphical model, denoted by $\mathcal{P}(G)$, is a statistical model such that

$$\mathcal{P}(G) = \{P \in \mathcal{P} : \ X_A \perp\!\!\!\perp X_B | X_C[P] \text{ for } \langle A, B | C \rangle \in \mathcal{I}(G)\}.$$

For the Markov distribution family $\mathcal{P}(G)$, we say that it is faithful to G if there exists a distribution $P^* \in \mathcal{P}(G)$ such that $\mathcal{I}(P^*) = \mathcal{I}(G)$, where

$$\mathcal{I}(P) = \{\langle A, B | C \rangle : X_A \perp\!\!\!\perp X_B | X_C[P] \text{ for } A, B, C \subseteq V(G)\}.$$

All the graphical models concerned throughout this paper are assumed to be faithful to G. Such an assumption is called "Faithfulness Assumption" [17]. In fact, this assumption is broad and mild since Gaussian distribution families and multinomial distribution families satisfy the faithfulness assumption.

Moreover, a statistical model $\mathcal{P}(G)$ also admits a natural projection operation on $D \subseteq V(G)$, denoted by $\mathcal{P}(G)_D$, which is defined as follows:

$$\mathcal{P}(G)_D = \left\{ P_D = \int_{V(G) \setminus D} dP : P \in \mathcal{P}(G) \right\}.$$

Generally, $\mathcal{P}(G)_D$ is not equal to $\mathcal{P}(G_D)$, but it is obviously shown that $\mathcal{P}(G)_D \supseteq \mathcal{P}(G_D)$.

Definition 4 (M-collapsibility). *Let G be a fixed undirected graph in $\mathfrak{U}$. For $D \subseteq V(G)$, $\mathcal{P}(G)$ can be model collapsible (M-collapsible) onto D if $\mathcal{P}(G)_D = \mathcal{P}(G_D)$.*

M-collapsibility indicates that the marginal distribution family is identical to the distribution family induced by G_D.

Theorem 1. *Let G be a fixed undirected graph in $\mathfrak{U}$ and $D \subseteq V(G)$. Then, the following statements are equivalent.*

1. *G is graphical collapsible onto D;*
2. *$\mathcal{I}(G)$ is CI-collapsible onto D;*
3. *$\mathcal{P}(G)$ is M-collapsible onto D.*

Proof. See Appendix A. □

Let $H_j = \bigcup_{i=1}^{j} U_i$ denote the histories set for each $j = 1, 2, \ldots, k$. By Theorem 1, we can obtain the following result.

Proposition 1. *Let G be a fixed graph in $\mathfrak{U}$ and G has a perfect sequence $(U_1, U_2, \ldots, U_k)$ of maximal prime subgraphs. Then, the following statements hold for each $j \in \{1, 2, \ldots, k\}$.*

1. *G can be graphical collapsible onto H_j;*
2. *$\mathcal{I}(G_{H_j}) = \mathcal{I}(G)_{H_j}$;*
3. *$\mathcal{P}(G_{H_j}) = \mathcal{P}(G)_{H_j}$.*

Proof. This can be easily obtained from the meaning of collapsibility and Theorem 1. □

3. Structural Markov Graph Laws for Full Bayesian Inference

3.1. Basic Concepts and Properties

We begin with the definition of the structural Markov property of [10].

Definition 5. *A graph law* $\mathfrak{G}(G)$ *over* $\mathfrak{U}$ *is structural Markov if*

$$G_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | \{G \in \mathfrak{U}(A, B, S)\} \ [\mathfrak{G}],$$

where $\mathfrak{G}(\mathfrak{U}(A, B, S)) > 0$ *and* $\mathfrak{U}(A, B, S)$ *is the set of undirected graphs for which* (A, B, S) *is a prime decomposition.*

Specifically, if G is decomposable in $\mathfrak{U}$, Definition 5 degenerates to that defined in [10].

The structural Markov property indicates that the structures of different induced subgraphs are conditionally independent when the event $\{G \in \mathfrak{U}(A, B, S)\}$ happens; see Figure 3 as an illustration.
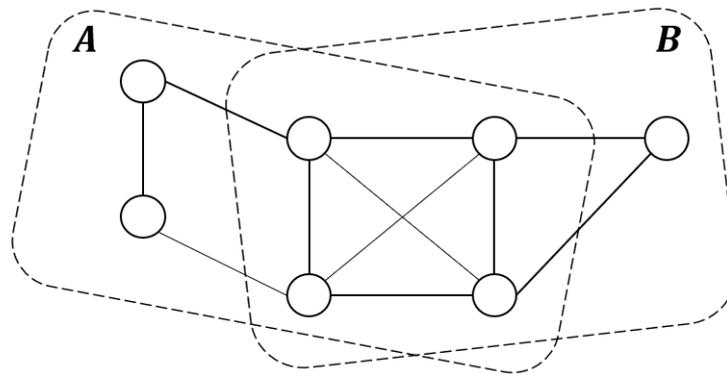

Figure 3. A representation of the structural Markov property for non-decomposable undirected graphs: $A \cap B$ is complete and separates A from B.

Proposition 2. *Let G be a fixed undirected graph in $\mathfrak{U}$. For any subsets A, B and S of $V(G)$ satisfying $A \cup B \cup S = V(G)$, if $\mathfrak{G}(G)$ is structural Markov, then*

$$G_{A \cup S} \perp\!\!\!\perp G_{B \cup S}$$

whenever S is complete and separates A from B in G.

Proof. By Definition 5, the existence of the remaining edges in $G_{A \cup S}$ is independent of those in $G_{B \cup S}$ since S is complete and separates A from B in G. Therefore, we are naturally left with a statement of marginal independence $G_{A \cup S} \perp\!\!\!\perp G_{B \cup S}$ since the term G_S is redundant. Hence, the result follows. □

Proposition 2 indicates that different components of undirected graphs are conditionally independent provided that their corresponding separators are complete. In order to illustrate our results with detailed investigations, we give a non-decomposable graph G in which $A \cap B$ separates A from B while $A \cap B$ is incomplete in Figure 4. We can easily find that the two subgraphs G_B and G_A have possible common edges in $A \cap B$, which make the existence of the remaining edges in G_A dependent of those in G_B. In other words, these dependencies will disappear as long as $A \cap B$ is complete.
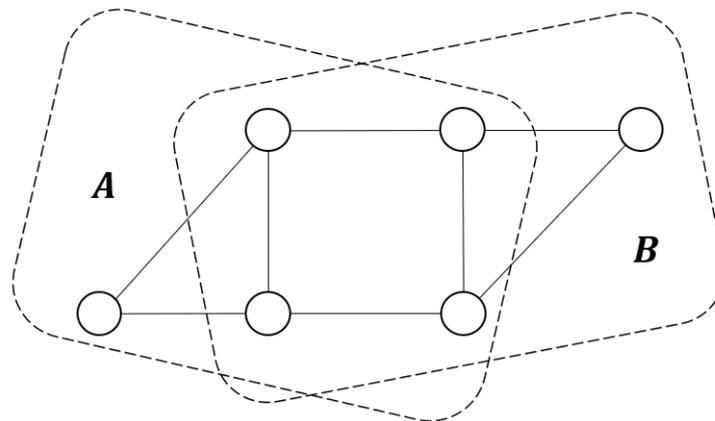
Figure 4. $A \cap B$ separates A from B while $A \cap B$ is incomplete.

It also implies that an arbitrary undirected graph can be denoted by the graph product of its induced subgraphs as

$$G = G_{A \cup S} \otimes G_{B \cup S}, G \in \mathfrak{U}(A, B, S).$$

The structural Markov property can be well-characterized by the above operation.

Proposition 3. *Let π be the density of a graph law $\mathfrak{G}$ with respect to the counting measure on $\mathfrak{U}$. Suppose that $G, G' \in \mathfrak{U}(A, B, S)$. Then,*

1. $G_{A \cup S} \otimes G'_{B \cup S} \in \mathfrak{U}(A, B, S)$ *and* $G'_{A \cup S} \otimes G_{B \cup S} \in \mathfrak{U}(A, B, S)$;
2. *if $\mathfrak{G}$ is structural Markov on $\mathfrak{U}$, then*

$$\pi(G)\pi(G') = \pi(G_{A \cup S} \otimes G'_{B \cup S})\pi(G'_{A \cup S} \otimes G_{B \cup S}).$$

Proof. See Appendix A. □

For any subset $C \subseteq A$, define $G_A^{(C)}$ to be the graph on A such that $G_A^{(C)}$ is complete in C and empty otherwise.

Proposition 4. *Let G be a fixed graph in $\mathfrak{U}$ and G has a perfect sequence $(U_1, U_2, \ldots, U_k)$ of maximal prime subgraphs. If G has a structural Markov graph law $\mathfrak{G}$ with the density π, then the density π can be factorized as*

$$\pi(G) = \frac{\prod\limits_{j=1}^{k} \pi(G_{U_j})}{\prod\limits_{j=2}^{k} \pi(G_{S_j})}. \tag{1}$$

Proof. See Appendix A. □

3.2. Joint Distribution Law

In this section, we will investigate how the structural Markov laws interact with the hyper Markov laws when they are considered as the joint prior laws.

Hyper Markov laws are motivated by the property that graph decomposition allows one to decompose a prior or posterior distribution into the product of marginal distributions on corresponding maximal prime subgraphs. For a fixed graph $G \in \mathfrak{U}(A, B, S)$, any prior or posterior distribution of $\theta \in \mathcal{P}(G)$ is uniquely characterized by its marginals $\theta_{A \cup S}$ and $\theta_{B \cup S}$, taking values in $\mathcal{P}(G_{A \cup S})$ and $\mathcal{P}(G_{B \cup S})$, respectively.

Following [12], to be specific, a probability distribution of a random distribution θ, which takes values in $\mathcal{P}(G)$, is said to be a law, denoted by $\mathcal{L}$. For $A \subseteq V(G)$, the marginal law of θ_A will be denoted by $\mathcal{L}_A$ and $\mathcal{L}_{B|A}$ will denote the conditional law of $\theta_{B|A}$.

Here, we give the definitions of weak and strong hyper Markov properties.

Definition 6 ([12], Weak and strong hyper Markov). *Suppose that G is a fixed graph in $\mathfrak{U}(A, B, S)$ and $\theta \in \mathcal{P}(G)$. Let $\mathcal{L}(\theta)$ be a law of θ. We say that $\mathcal{L}(\theta)$ is weak hyper Markov over G if*

$$\theta_{A\cup S} \perp\!\!\!\perp \theta_{B\cup S} | \theta_S \ [\mathcal{L}].$$

Further, we say that $\mathcal{L}(\theta)$ is strong hyper Markov over G if

$$\theta_{A\cup S} \perp\!\!\!\perp \theta_{B\cup S|S} \ [\mathcal{L}].$$

Let X be a random sample from $\theta \in \mathcal{P}(G)$. The conditional independence property of the joint distribution law $(P, \mathcal{L})$ for the pair (X, θ) on $G \in \mathfrak{U}$ can be characterized as follows.

Proposition 5. *Let G be a fixed undirected graph in $\mathfrak{U}$ with a prime decomposition (A, B, S). X is a random sample from $\theta \in \mathcal{P}(G)$. Then, the joint distribution law of (X, θ) satisfies:*

1. *if $\mathcal{L}(\theta)$ is weak hyper Markov with respect to G, then*

$$(X_{A\cup S}, \theta_{A\cup S}) \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S}) | (X_S, \theta_S) \ [P, \mathcal{L}];$$

2. *if $\mathcal{L}(\theta)$ is strong hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S|S}) \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S}) | X_S \ [P, \mathcal{L}].$$

Proof. See Appendix A. □

It is worth mentioning that the hyper Markov property does not hold for the cases where separators are not complete. For instance, the graph G_1 in Figure 1 is incomplete, and we do not have $\theta_{\{a,b,c\}} \perp\!\!\!\perp \theta_{\{a,d,c\}} | \theta_{\{a,c\}}$ or $\theta_{\{b,a,d\}} \perp\!\!\!\perp \theta_{\{b,c,d\}} | \theta_{\{b,d\}}$. However, it is worthwhile to point out that the corresponding pairwise Markov property $X_b \perp\!\!\!\perp X_d | X_{\{a,c\}}$ or $X_a \perp\!\!\!\perp X_c | X_{\{b,d\}}$ holds under P if P is Markov with respect to G_1.

Let Θ be the family of Markov distributions over $\mathfrak{U}$ and $\mathfrak{L}$ the family of hyper Markov laws over $\mathfrak{U}$. For the sake of discussion, it is necessary for us to reconsider the notion of hyper compatibility, which was first proposed by [10], to characterize families of laws for every graph.

Definition 7 (Hyper compatibility). *Let $\mathcal{L}, \mathcal{L}' \in \mathfrak{L}$ be the laws of $\theta \in \Theta$ with respect to G and G' on $\mathfrak{U}$, respectively. For $A \subseteq V(G)$, we say $\mathfrak{L}$ is hyper compatible on $\mathfrak{U}$ if $\mathcal{L}_A(\theta) = \mathcal{L}'_A(\theta)$ whenever G, G' are collapsible onto A and $G_A = G'_A$.*

Here $\mathfrak{L}$ is always assumed to be hyper compatible over $\mathfrak{U}$. Based on the arguments above, some significant conditional independence properties of such joint law $(\mathfrak{L}, \mathfrak{G})$ can be investigated as the following.

Proposition 6. *Suppose that G has a graph law $\mathfrak{G}$ over $\mathfrak{U}$. If θ has a law $\mathcal{L}$ from a hyper compatible family $\mathfrak{L}$ over $\mathfrak{U}$, then*

$$\theta_{A\cup S} \perp\!\!\!\perp G_{B\cup S} | (G_{A\cup S}, \{G \in \mathfrak{U}(A, B, S)\}) \ [\mathfrak{L}, \mathfrak{G}].$$

Proof. Suppose that $G \in \mathfrak{U}(A, B, S)$. Since G is collapsible onto both $A \cup S$ and $B \cup S$, by hyper compatibility, $\mathcal{L}_{A\cup S}$ can only take values in $\mathfrak{L}(G_{A\cup S})$ for any $\mathcal{L} \in \mathfrak{L}$. □

Theorem 2. *Suppose that $\mathfrak{G}(G)$ is structural Markov over $\mathfrak{U}$. For any $\mathcal{L} \in \mathfrak{L}$,*

1. *if $\mathcal{L}(\theta)$ is weak hyper Markov, then*

$$(\theta_{A\cup S}, G_{A\cup S}) \perp\!\!\!\perp (\theta_{A\cup S}, G_{B\cup S}) | (\theta_S, \{G \in \mathfrak{U}(A, B, S)\}) \ [\mathfrak{L}, \mathfrak{G}];$$

2. *if $\mathcal{L}(\theta)$ is strong hyper Markov, then*

$$(\theta_{A\cup S}, G_{A\cup S}) \perp\!\!\!\perp (\theta_{B\cup S|S}, G_{B\cup S}) | \{G \in \mathfrak{U}(A, B, S)\}\ [\mathfrak{L}, \mathfrak{G}].$$

Proof. See Appendix A. □

Theorem 2 reflects the conditional independence properties at both parameter and structural level.

Further, for any $G \in \mathfrak{U}$, let X be a random sample from a distribution $\theta \in \Theta$ on $\mathfrak{U}$. If G is assigned the prior law $\mathfrak{G}$ and θ is assigned the prior law $\mathcal{L}$, then a joint distribution law is thereby created for (X, θ, G).

Proposition 7. *Suppose that $\mathfrak{G}(G)$ is structural Markov on $\mathfrak{U}$. Let X be a random sample from $\theta \in \Theta$. For $\mathcal{L} \in \mathfrak{L}$,*

1. *if $\mathcal{L}(\theta)$ is weak hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S}) \perp\!\!\!\perp G_{B\cup S} | (X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\})\ [\Theta, \mathfrak{L}, \mathfrak{G}];$$

2. *if $\mathcal{L}(\theta)$ is strong hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S|S}) \perp\!\!\!\perp G_{B\cup S} | (X_S, \{G \in \mathfrak{U}(A, B, S)\})\ [\Theta, \mathfrak{L}, \mathfrak{G}].$$

Proof. See Appendix A. □

The conditional independence property of any such joint distribution law of (X, θ, G) can be characterized as follows.

Theorem 3. *Suppose that $\mathfrak{G}(G)$ is structural Markov on $\mathfrak{U}$. Let X be a random sample from $\theta \in \Theta$ on $\mathfrak{U}$. For $\mathcal{L} \in \mathfrak{L}$,*

1. *if $\mathcal{L}(\theta)$ is weak hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S}, G_{A\cup S}) \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S}, G_{B\cup S}) | (X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\})\ [\Theta, \mathfrak{L}, \mathfrak{G}];$$

2. *if $\mathcal{L}(\theta)$ is strong hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S}, G_{A\cup S}) \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S|S}, G_{B\cup S}) | (X_S, \{G \in \mathfrak{U}(A, B, S)\})\ [\Theta, \mathfrak{L}, \mathfrak{G}].$$

Proof. See Appendix A. □

Theorem 3 reflects that a random sample can be determined by both hyper and structural parameters, which will play a significant role in full Bayesian inference.

Corollary 1. *Suppose that $\mathfrak{G}(G)$ is structural Markov on $\mathfrak{U}$. Let X be a random sample from $\theta \in \Theta$ on $\mathfrak{U}$. For $\mathcal{L} \in \mathfrak{L}$,*

1. *if $\mathcal{L}(\theta)$ is weak hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S}) \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S}) | (X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\})\ [\Theta, \mathfrak{L}, \mathfrak{G}];$$

2. *if $\mathcal{L}(\theta)$ is strong hyper Markov, then*

$$(X_{A\cup S}, \theta_{A\cup S}) \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S|S}) | (X_S, \{G \in \mathfrak{U}(A, B, S)\})\ [\Theta, \mathfrak{L}, \mathfrak{G}].$$

Proof. It can be easily obtained from Theorem 3. □

Corollary 1 can be considered as a generalization of Proposition 5 since G is a random undirected graph on $\mathfrak{U}$ with a prime decomposition (A, B, S). Without loss of generality,

when the event $\{G \in \mathfrak{U}(A, B, S)\}$ happens, i.e., given a graph G with a prime decomposition (A, B, S), we can deduce from Corollary 1 that

$$(X_{A \cup S}, \theta_{A \cup S}) \perp\!\!\!\perp (X_{B \cup S}, \theta_{B \cup S}) | (X_S, \theta_S) \; [P, \mathcal{L}].$$

3.3. Posterior Updating for Graph Law

Our research in this section aims to identify the structure of models via the Bayesian approach. Based on our results in Section 3.2, in the following, we will use data from a certain distribution to learn the structure of a graph.

We assume that G has a structural Markov graph law $\mathfrak{G}$ over $\mathfrak{U}$. For $\theta \in \Theta$, let θ have a law from a hyper compatible family $\mathfrak{L}$. Let $X^{(n)} = (X^1, X^2, \dots, X^n)$ denote a random sample of n observations from θ. If we focus on the density of posterior graph law $\pi(G|x^{(n)}, \theta)$ with its conjugated prior graph law $\pi(G)$, then the full Bayesian posterior graph law follows:

$$\pi(G|x^{(n)}, \vartheta) = \frac{1}{Z} \pi(G) \ell(\theta|G) p(x^{(n)}|\theta; G), \theta \in \Theta, G \in \mathfrak{U},$$

where Z is a normalizing constant and ϑ is a hyperparameter that characterizes the law of θ. In general, it is hardly for us to estimate the structure of a graph G since the hyper parameter ϑ is unknown.

In the following, we investigate the properties of structural Markov laws when used as priors for models.

Proposition 8. *If the prior graph law $\mathfrak{G}(G)$ is structural Markov on $\mathfrak{U}$, then the posterior graph law, obtained by conditioning on data $X^{(n)} = x^{(n)}$, is structural Markov on $\mathfrak{U}$.*

Proof. By the conditional independence and Theorem 3, we can easily find that

$$G_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | (X^{(n)}, \theta, \{G \in \mathfrak{U}(A, B, S)\}).$$

$\square$

Proposition 9. *Assume that the prior graph law $\mathfrak{G}(G)$ is structural Markov and $\mathcal{L}(\theta)$ is strong hyper Markov on $\mathfrak{U}$. Then, the following properties hold:*

1. *The posterior graph law obtained by conditioning on data $X^{(n)} = x^{(n)}$ is structural Markov with respect to $\mathfrak{U}$;*
2. *The marginal data distribution of $X^{(n)}$ is Markov with respect to $\mathfrak{U}$;*
3. *The posterior law of θ conditioning on $X^{(n)} = x^{(n)}$ is Markov with respect to $\mathfrak{U}$.*

Proof. By the conditional independence and Theorem 3, we have

$$G_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | (X^{(n)}, \{G \in \mathfrak{U}(A, B, S)\}).$$

This implies (i).

To prove (ii), by the conditional independence and Theorem 3, we have

$$X_{A \cup S}^{(n)} \perp\!\!\!\perp X_{B \cup S}^{(n)} | (X_S^{(n)}, \{G \in \mathfrak{U}(A, B, S)\}).$$

In particular, if G is given from $\mathfrak{U}(A, B, S)$, then

$$X_{A \cup S}^{(n)} \perp\!\!\!\perp X_{B \cup S}^{(n)} | X_S^{(n)}.$$

From Theorem 3, we have

$$\theta_{S \cup S} \perp\!\!\!\perp \theta_{B \cup S} | (X^{(n)}, \theta_S, \{G \in \mathfrak{U}(A, B, S)\}),$$

which implies (iii). □

Our Bayesian approaches call for a strong hyper Markov prior law on θ with respect to $G \in \mathfrak{U}$. By Proposition 9, the posterior law of θ, given G, has a density ℓ of the following form:

$$\ell(\theta|x^{(n)}, G) = \frac{\prod_{U \in \mathcal{U}} \ell(\theta_U|x_U^{(n)})}{\prod_{S \in \mathcal{S}} \ell(\theta_S|x_S^{(n)})}, \theta \in \Theta, \tag{2}$$

where $\mathcal{U}$ is the set of maximal prime subgraphs of G and $\mathcal{S}$ is the set of corresponding clique separators.

If $\mathfrak{G}(G)$ is structural Markov and $\mathcal{L}(\theta)$ is strong hyper Markov with respect to G, then the posterior graph law of G will be given by

$$\ell(G|x^{(n)}, \vartheta) \propto \frac{\prod_{U \in \mathcal{U}} \pi(G_U)\ell(\theta_U|x_U^{(n)})}{\prod_{S \in \mathcal{S}} \pi(G_S)\ell(\theta_S|x_S^{(n)})}, \theta \in \Theta, G \in \mathfrak{U}. \tag{3}$$

It is worthwhile to point out that (3) indicates that the posterior graph law of G will preserve the structural Markov property under the hyper compatible laws. This result coincides with Proposition 8. Further, this updating may be performed locally by (3), which implies that the posterior graph laws on each maximal prime subgraphs of G are only dependent of the posterior of hyper compatible laws on the maximal prime subgraph.

4. Two Special Cases
4.1. Graphical Gaussian Models and the Inverse Wishart Law

A graphical Gaussian model is defined by a p-dimensional multivariate Gaussian distribution with the expected value μ and covariance matrix Σ, i.e.,

$$P(X) = N_p(\mu, \Sigma).$$

For simplicity, we assume that the model has zero mean in the following. Define $\mathbf{K} = \Sigma^{-1}$ to be the precision matrix of G, where

$$\mathbf{K} \in \mathbf{M}_p^+ : \mathbf{K}_{uv} = 0 \text{ for all } (u, v) \notin E(G),$$

where $\mathbf{M}_p^+$ denotes the set of $p \times p$ positive definite matrices. For any matrix $\mathbf{M} \in \mathbf{M}_p^+$, $\mathbf{M}_A$ will denote the $|A| \times |A|$ matrix obtained by $(\mathbf{M}_{uv})_{(u,v) \in A^2}$. It has been shown that the global, local and pairwise Markov properties are equivalent in graphical Gaussian models; see [2]. We therefore conclude that the graphical Gaussian distribution P is Markov with respect to G if and only if

$$\mathbf{K}_{uv} = 0 \quad \Leftrightarrow \quad X_u \perp\!\!\!\perp X_v | X_{V(G)\setminus\{u,v\}}. \tag{4}$$

Let $x^{(n)}$ be observations of $n \times p$ sample matrix $X^{(n)}$, a random sample of size n from the graphical Gaussian distribution $N_p(0, \Sigma)$, and let $\mathbf{S} = x^{(n)}(x^{(n)})^T$ denote the observed sum-of-products matrix. Then, for any $U \in \mathcal{U}$,

$$p(x_U^{(n)}|\Sigma_U) = \frac{1}{(2\pi)^{|U|}} |\Sigma_U|^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2} tr(\Sigma_U^{-1}\mathbf{S}_U) \right\},$$

where $|U|$ is the cardinality of U, and $|\Sigma_U|$ is the determinant of Σ_U. It is similar for $p(x_S^{(n)}|\Sigma_S)$, $S \in \mathcal{S}$.

The inverse Wishart distribution is also termed inverse Wishart law, denoted by $IW(\delta, \mathbf{\Phi})$. It is as the prior for the graphical Gaussian distribution $N_p(0, \Sigma)$. Conditioning

on (4), $\boldsymbol{\Sigma}$ has a hyper inverse Wishart prior law, denoted by $HIW(\delta, \boldsymbol{\Phi})$. The marginal density $\ell(\boldsymbol{\Sigma}_U | \boldsymbol{\Phi}_U)$ is of the form

$$\ell(\boldsymbol{\Sigma}_U | \boldsymbol{\Phi}_U) = |\boldsymbol{\Sigma}_U|^{-\frac{\delta+|U|-1}{2}} \exp\left\{-\frac{1}{2} tr(\boldsymbol{\Sigma}_U^{-1} \boldsymbol{\Phi}_U)\right\}.$$

It is already shown in [12] that the hyper inverse Wishart law satisfies the strong hyper Markov property, which would allow us to compute the posterior updating of Σ by the margins of maximal prime subgraphs of the graph G. That is, for any $U \in \mathcal{U}$,

$$\ell(\boldsymbol{\Sigma}_U | X^{(n)} = x^{(n)}) = \ell(\boldsymbol{\Sigma}_U | X_U^{(n)} = x_U^{(n)})$$

with the density

$$\ell(\boldsymbol{\Sigma}_U | x_U^{(n)}) \propto \boldsymbol{\Sigma}_U|^{-\frac{\delta+n+|U|-1}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\Sigma}_U^{-1}(\boldsymbol{\Phi}_U + \mathbf{S}_U)\right\}.$$

We conclude that

$$\mathcal{L}(\boldsymbol{\Sigma}_U | X_U^{(n)} = x_U^{(n)}) = IW((\delta + n; \mathbf{S}_U + \boldsymbol{\Phi}_U).$$

Consequently, if we assign a prior law of form (1) for G, then from Proposition 8 we can conclude that the posterior graph law of G, given data $X^{(n)} = x^{(n)}$ from the Gaussian distribution $N_p(0, \boldsymbol{\Sigma})$, can be obtained through (3) with a density of the following form:

$$\pi(G | x^{(n)}, \boldsymbol{\Phi}) \propto \frac{\prod_{U \in \mathcal{U}} \pi(G_U) \ell(\boldsymbol{\Sigma}_U | \boldsymbol{\Phi}_U + \mathbf{S}_U)}{\prod_{S \in \mathcal{S}} \pi(G_S) \ell(\boldsymbol{\Sigma}_S | \boldsymbol{\Phi}_S + \mathbf{S}_S)}, G \in \mathfrak{U}. \tag{5}$$

4.2. Multinomial Models and the Dirichlet Law

Suppose that all the variables $(X_1, X_2, \ldots, X_p)$ are discrete-valued. Let $V(G)$ denote the contingency table by $I = I_1 \times I_2 \times \cdots \times I_p$, where I_h is a finite set for each $h \in \{1, 2, \cdots, p\}$. An element $i \in I$ is referred to as a cell in this table. Based on this, $(X_1, X_2, \ldots, X_p)$ will take value in finite sets $I = (I_1, I_2, \ldots, I_p)$. Indeed, I is a discrete-valued random vector whose distribution θ is assumed to be Markov with respect to G. Then,

$$\theta(i) = \frac{\prod_{U \in \mathcal{U}} \theta(i_U)}{\prod_{S \in \mathcal{S}} \theta(i_S)}, i \in I, \tag{6}$$

where $\theta(i_U) \in (0, 1), \theta(i_S) \in (0, 1)$ and $\sum_{i \in I} \theta(i) = 1$.

Let $x^{(n)}$ be observations of $X^{(n)}$, a random sample from θ. $X^{(n)}$ is an $n \times p$ matrix where each row denotes an observation of I. The distribution of $X^{(n)}$ is the multinomial distribution with index n and probabilities θ, denoted by $M(n, \theta)$. Then, the likelihood function $p(x^{(n)} | \theta, G)$ has the form

$$p(x_U^{(n)} | \theta_U) = \prod_{i_U \in I_C} \theta(i_U)^{n(i_U)}, U \in \mathcal{U},$$

where $I_U = \prod_{u \in U} I_u$, $\theta_U = (\theta(i_U))_{i_U \in I_U}$, and $n(i_U)$ counts the number of elements of $x_U^{(n)}$ from the marginal cell i_U. It is similar for $p(x_S^{(n)} | \theta_S)$, $S \in \mathcal{S}$.

The Dirichlet distribution is also termed Dirichlet law, denoted by $D(\alpha)$, where $\alpha = (\alpha(i))_{i \in I}$ are hyper parameters. It is used as the prior for multinomial distribution $M(n, \theta)$. It is shown that the Dirichlet law satisfies strong hyper Markov property; see [12]. Thus, we have

$$\ell(\theta_U | \alpha_U) = \prod_{i_U \in I_U} \theta(i_U)^{\alpha(i_U)-1}, U \in \mathcal{U},$$

and then the posterior law can be written as

$$\ell(\theta_U|x^{(n)}, \alpha_U) = \prod_{i_U \in I_U} \theta(i_U)^{\alpha(i_U)+n(i_U)-1}, U \in \mathcal{U}.$$

Based on the above arguments, we can conclude that $\mathcal{L}(\theta_U|x_U^{(n)}) = D(\alpha_U + n_U)$. Further, if we assign a prior law of form (1) for G, by Proposition 8, the posterior graph law of G, given data $X^{(n)} = x^{(n)}$ obtained from θ, has density in the following way:

$$\pi(G|x^{(n)}, \alpha) \propto \frac{\prod_{U \in \mathcal{U}} \pi(G_U)\ell(\alpha_U + n)}{\prod_{S \in \mathcal{S}} \pi(G_S)\ell(\alpha_S + n)}, G \in \mathfrak{U}.$$

4.3. An Example on Simulated Data

4.3.1. Dataset Description

In this section, we present the results for one application to a real dataset. We analyze a labor force survey dataset, which is available from [18]. This dataset is used to analyze the multivariate associations among income, education and family background on 1002 males in the American labor force. Here, we briefly describe these variables in this dataset.

- **inc:** The income of the respondents.
- **deg:** Tespondents' highest educational degree.
- **chi:** The number of children of the respondents.
- **pin:** The income of the respondents' parents.
- **pde:** The highest educational degree of respondents' parents.
- **pch:** The number of children of respondents' parents.
- **age:** Respondents' age in years.

4.3.2. Experiments and Results

We consider the posterior graph law of G in Equation (5), a Gibbs sampler can then be formed by using the following conditional posteriors:

1. $X \sim N_p(0, \mathbf{\Sigma})$;
2. $\mathbf{\Sigma}|X, \mathbf{\Phi} \sim IW(n + \delta - 1, \mathbf{S} + \mathbf{\Phi})$.

For the prior graph law of G, following from Example 3.5 in [10], we consider an Erdős-Rényi random graph model prior on each edge (u, v) with

$$\pi(G) \propto \left(\frac{\varphi}{1-\varphi}\right)^{|E(G)|},$$

where the parameter $\varphi \in (0, 1)$ is a prior probability of existing edges. In this case, we set $\varphi = 0.5$. We use the inverse Wishart law $IW(\delta, \mathbf{\Phi})$ as a prior for the covariance matrix over the graph G, with $\delta = 7$ and $\mathbf{\Phi} = \mathbf{I}_7$ as an identity matrix here.

By using the function above, we simulate $n = 1002$ observations. The experimental results are implemented by R package for 5000 iterations with 2500 as burn-in as follows:

The experimental results on this dataset are displayed in Figures 5 and 6. The estimated posterior probabilities of the size of the graphs are shown in the left of Figure 5, which shows that our algorithm mainly visits graphs with sizes between nine and twelve edges. The figure on the right exhibits the estimated posterior probabilities of all visited graphs with various sizes, and also shows that more than 15 different graphs are visited. The graph in Figure 6 is the selected graph with the highest posterior probability from these visited graphs.
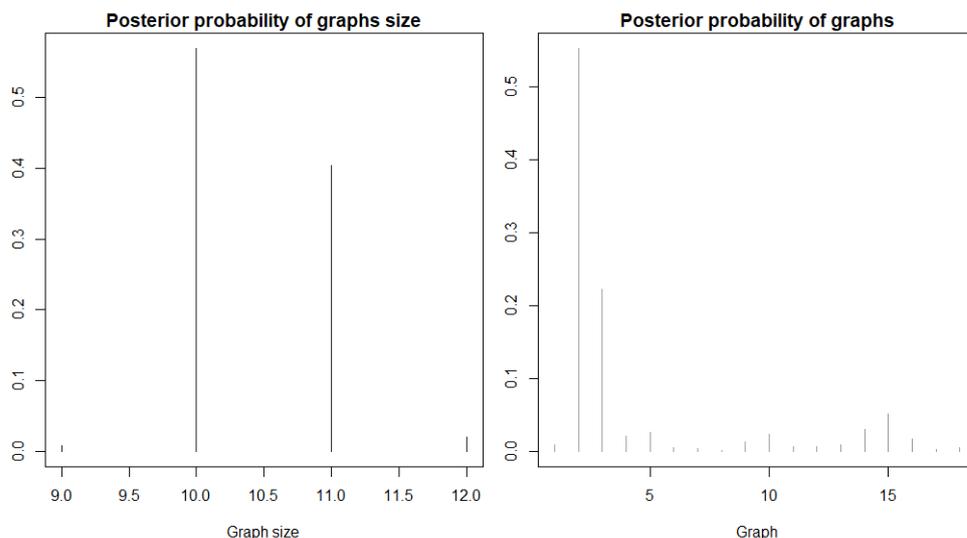
Figure 5. The figure in the left is the estimated posterior probabilities of the size of the graphs. The figure in the right is the estimated posterior probabilities of all visited graphs.
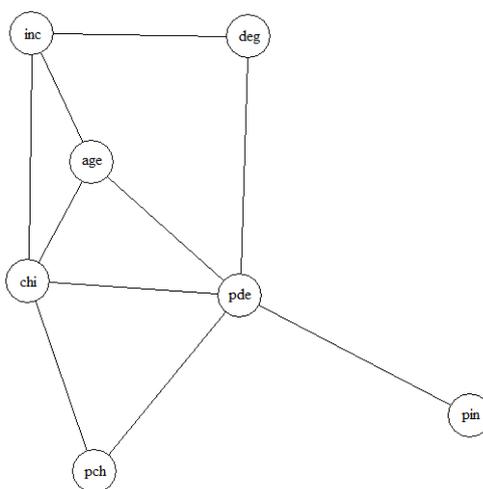

Figure 6. The figure is the inferred graph with the highest posterior probability.

The results also suggest that the respondents' income has relationships with their own education and age. It is also shown that the income of respondents' parents is only related to their education.

5. Computations

In this section, we aim to design an algorithm to take samples that we are interested in, such as decomposable undirected graphs, from the structural Markov graph law $\mho$ on $\mathfrak{U}$.

5.1. Ratio for Graph Law

Model comparison plays an important role in statistical analysis, especially in solving the problem of the ratio of distributions of variables in different states. We consider a graph itself as a random variable into the construction of this ratio between two undirected

graphs G' and G, where G' is obtained from G by removing or adding one edge. This ratio can be written as

$$\Lambda(G' : G) = \frac{\pi(G')}{\pi(G)}.$$

The main objective of this next section is to greatly simplify this complex calculation under the assumption that the graph law $\mathfrak{G}$ is structural Markov on $\mathfrak{U}$. For the sake of convenience, we define $\eta_U = \pi(G_U)$ and $\zeta_S = \pi(G_S)$ for $U \in \mathcal{U}$ and $S \in \mathcal{S}$.

In Figure 7, it is a special case where G' is obtained from G by removing the edge (u, v), which is exactly in one prime component $U \in \mathcal{U}$ of G.
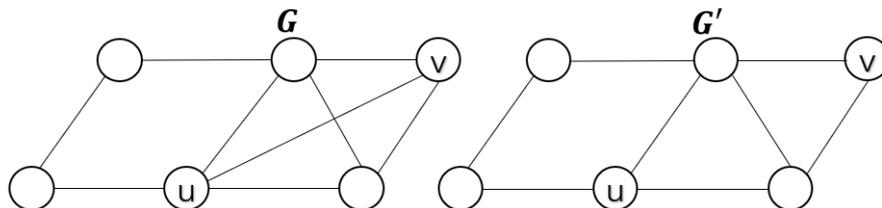

Figure 7. G' is obtained from G by removing the edge (u, v).

Proposition 10. *Let G be a fixed graph in $\mathfrak{U}$ and G has a perfect sequence $(U_1, U_2, \ldots, U_k)$ of maximal prime subgraphs. Suppose that G' is obtained from G by removing the edge (u, v). Then,*

1. *if u and v are contained in exactly one maximal prime subgraph U_j of G, then*

$$\Lambda(G' : G) = \frac{\eta_{U_j'}}{\eta_{U_j}}, j = 1, 2, \ldots, k;$$

2. *if u and v are contained in both two neighboring maximal prime subgraphs U_j, U_{j+1} of G, then*

$$\Lambda(G' : G) = \frac{\eta_{W'}}{\eta_W}, \tag{7}$$

where $W = U_j \cup U_{j+1}$ in G.

Proof. See Appendix A $\quad\square$

In Figure 8, it is a certain case where G' is obtained from G by adding the edge (u, v) within two neighboring prime components U_i and U_j of G such that $u \in U_i$ and $v \in U_j$.
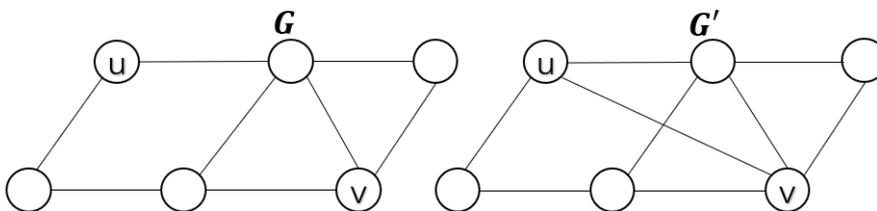

Figure 8. G' is obtained from G by adding the edge (u, v).

Proposition 11. *Let G be a fixed graph in $\mathfrak{U}$ and G has a perfect sequence $(U_1, U_2, \ldots, U_k)$ of maximal prime subgraphs. Suppose that G' is obtained from G by adding the edge (u, v). Then,*

1. *if u and v are contained in exactly one incomplete prime subgraph U_h, then*

$$\Lambda(G' : G) = \frac{\eta_{U_h'}}{\eta_{U_h}}.$$

2.　*if $U_i \ni u$ and $U_j \ni v$ are the two distinct maximal prime subgraphs of G, then there are some prime components $U_i = U_{h_1}, U_{h_2}, \ldots, U_{h_m} = U_j$ such that*

$$\Lambda(G' : G) = \frac{\eta_{T'}}{\eta_T}, \tag{8}$$

where $T = U_{h_1} \cup U_{h_2} \cup \cdots \cup U_{h_m}$.

Proof. See Appendix A.　□

In particular, if *G* is a decomposable graph in $\mathfrak{U}$, then we have the following results.

Lemma 1 ([19])**.** *Let G be a decomposable graph in $\mathfrak{U}$ and G has a perfect cliques sequence $(U_1, U_2, \ldots, U_k)$. Suppose that G' is decomposable, obtained from G by removing or adding one edge (u, v). Then,*

1.　*If G' is obtained from G by removing the edge (u, v), then u and v must belong to a clique U_j of G;*
2.　*If G' is obtained from G by adding the edge (u, v), then there exist two different cliques $U_i \ni u$ and $U_j \ni v$ such that $S = U_i \cap U_j$ is complete and separates U_i and U_j.*

Corollary 2. *Let G be a decomposable graph in $\mathfrak{U}$ and G has a perfect cliques sequence $(U_1, U_2, \ldots, U_k)$. Suppose that G' is decomposable, obtained from G by removing or adding one edge (u, v). Then,*

1.　*If G' is obtained from G by removing the edge (u, v) within U_j, then*

$$\Lambda(G' : G) = \frac{\eta_{U_u} \eta_{U_v}}{\eta_0 \eta_U},$$

where $U_u = U_j \setminus \{v\}$, $U_v = U_j \setminus \{u\}$ and $U_0 = U_j \setminus \{u, v\}$;
2.　*If G' is obtained from G by adding the edge (u, v) such that $u \in U_i$ and $v \in U_j$, then the ratio $\Lambda(G' : G)$ is*

$$\Lambda(G' : G) = \frac{\zeta_S \zeta_{S_0}}{\zeta_{S_u} \zeta_{S_v}},$$

where $S = U_i \cap U_j$, $S_u = S \cup \{u\}$, $S_v = S \cup \{v\}$ and $S_0 = S \cup \{u, v\}$.

Proof. We first give the proof of 1. If $(u, v) \in E(G)$ and $(u, v) \notin E(G')$, by Lemma 1, the deleted edge (u, v) must belong to a single clique U_j. It is worthwhile to point out that all of U_u, U_v and U_0 are cliques in *G* and G'. Then,

$$\eta_{U'_j} = \frac{\eta_{U_u} \eta_{U_v}}{\eta_{U_0}},$$

which combines with (A19) gives the result. The proof of 2 is similar.　□

5.2. Sampling Decomposable Graphs from Structural Markov Graph Laws

We now take a random graph on $\mathfrak{U}$ as the initial state to design the Markov chain Monte Carlo (MCMC) sampler for sampling from a structural Markov graph law. This technique relies on small perturbations to the edge set of a graph, indicating that one edge could be removed or added.

A reversible jump MCMC sampler is introduced for posterior sampling of decomposable graphical models, which relies on making single edge additions and removals; see [8]. We now use this jump MCMC methodology for our sample from structural Markov law in further details.

Let *G* denote a state variable and G' the destination variable where G' is obtained from a random graph *G* by removing or adding one edge, and so *G* would take the chain to the

destination G' with probability $q(G, G')$, which ensures detailed balance with respect to the target distribution $\pi(G)$. Then, the Metropolis–Hastings acceptance ratio can be written as

$$\alpha(G', G) = \min\left\{1, \frac{\pi(G')q(G', G)}{\pi(G)q(G, G')}\right\}. \tag{9}$$

In fact, the Equation (9) is not the only choice yielding detailed balance. In particular, in order to reduce the error caused by excessive proportion, we can make the following adjustment that

$$\alpha(G', G) = \min\left\{1, \frac{\pi(G')}{\pi(G)}\right\} \times \min\left\{1, \frac{q(G', G)}{q(G, G')}\right\}.$$

In general, since the proposal kernel, which we will set as symmetric, that is, $q(G, G') = q(G', G)$. Consequently, it is indicated that the acceptance probability is only dependent on the relative densities, which will only require us to compute

$$\alpha(G', G) = \min\{1, \Lambda(G' : G)\}.$$

We randomly select a pair of vertices $u, v \in V(G)$. If $(u, v) \in E(G)$, then it is removed. If $(u, v) \notin E(G)$, then it is added. Let $G^{+(u,v)}$ denote the graph, which is obtained from G by adding the edge (u, v), and similarly for $G^{-(u,v)}$. Let $G^{(t)}$ denote the state of G at time t and let $\mathfrak{U}^*$ be the set of decomposable undirected graphs with vertex set $V(G)$. We begin with an ER random graph as its initial state, and then a Metropolis–Hastings algorithm for sampling decomposable graphs from a structural Markov graph law $\mathfrak{G}$ can be constructed in the following Algorithm 1:

Algorithm 1 A Metropolis–Hastings algorithm for sampling decomposable graphs from a structural Markov graph law.

Input: An ER random graph $G \in \mathfrak{U}$.
Output: A set of decomposable graph from $\mathfrak{U}$.
  Set $G^{(0)} = G$
  for $t = 0, 1, 2, \ldots.$ **do**
    if $(u, v) \in E(G^{(t)})$ and $G^{-(u,v)} \in \mathfrak{U}^*$ **then**
      set $G^{(t+1)} = G^{-(u,v)}$ with probability $\min\left(\frac{\eta_{u_u}\eta_{u_v}}{\eta_{u_0}\eta_u}, 1\right)$
    else if $(u, v) \notin E(G^{(t)})$ and $G^{+(u,v)} \in \mathfrak{U}^*$ **then**
      set $G^{(t+1)} = G^{+(u,v)}$ with probability $\min\left(\frac{\zeta_S\zeta_{S_0}}{\zeta_{S_u}\zeta_{S_v}}, 1\right)$
    else
      $G^{(t+1)} = G^{(t)}$
    end if
  end for
  return A set of decomposable graphs.

Based on our results in Section 5.1, this algorithm implies that the acceptance probability can be obtained by only evaluating the marginal likelihood of corresponding subsets of $V(G)$ at each step when sampling from a posterior graph law in Proposition 8 or Proposition 9.

6. Conclusions

The main contribution of this paper is to define the structural Markov properties of [10] for non-decomposable undirected graphs. It is shown that an arbitrary undirected graph can be primely decomposed into the sum of several prime subgraphs. Based on the prime decomposition of undirected graphs and conditional independence, the structural Markov properties can be naturally extended to arbitrary undirected graphs.

Then, we propose a full Bayesian method for estimating the structure of a graph. This method requires that our observed data are from a certain distribution. By using our results, we have shown that the computation of posterior updating of graph law can be determined by the prime components margins, which would make the computation of the posterior graph law greatly simplified.

It should be pointed that all our research only focuses on undirected graphs. However, other classes of graphs, such as chain graphs or ancestral graphs, may have more interesting and valuable properties that can reflect the conditional independence of the graph structure in the problem of models determination. In the future, we will detail them at length.

Author Contributions: Methodology, Y.S.; Validation, X.K.; Writing—review & editing, Y.H. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Appendix A. Proofs of Some Main Theorems and Propositions

Proof of Theorem 1. The equivalence of (i) and (ii) can be implied by [Corollary 2.5] [20]. So, it suffices to show that (ii) $\Leftrightarrow$ (iii). We first give the proof of (ii) $\Rightarrow$ (iii). Firstly, we know that
$$\mathcal{P}(G) = \{P : \mathcal{I}(G) \subseteq \mathcal{I}(P)\}.$$

By the meaning of $\mathcal{P}(G)$, we define

$$\mathcal{P}(G_D) = \{Q : \mathcal{I}(G_D) \subseteq \mathcal{I}(Q)\}. \tag{A1}$$

For $P \in \mathcal{P}(G)$, let $R = P_D \in \mathcal{P}(G)_D$. By CI-collapsibility, we have

$$\mathcal{I}(G_D) = \mathcal{I}(G)_D \subseteq \mathcal{I}(R). \tag{A2}$$

So, we implied that $R \in \mathcal{P}(G_D)$ by (A1) and (A2). From which, it follows that $\mathcal{P}(G)_D \subseteq \mathcal{P}(G_D)$. Hence, the result follows by $\mathcal{P}(G)_D \supseteq \mathcal{P}(G_D)$. Conversely, under the "Faithfulness Assumption", there is a $P^* \in \mathcal{P}(G)$ such that $\mathcal{I}(P^*) = \mathcal{I}(G)$, implying $\mathcal{I}(P_D^*) = \mathcal{I}(P^*)_D = \mathcal{I}(G)_D$. By M-collapsibility, we know that $P_D^* \in \mathcal{P}(G_D)$, which gives $\mathcal{I}(G_D) \subseteq \mathcal{I}(P_D^*)$. Hence, we have $\mathcal{I}(G_D) \subseteq \mathcal{I}(G)_D$. The result follows since it is easy to obtain that $\mathcal{I}(G)_D \subseteq \mathcal{I}(G_D)$. $\square$

Proof of Proposition 3. By the graph product operation, since S is complete and separates A from B, then for any $G, G' \in \mathfrak{U}(A, B, S)$, (A, B, S) is a prime decomposition of the graph $G_{A \cup S} \otimes G'_{B \cup S}$ with vertex set $V(G)$, and so is $G'_{A \cup S} \otimes G_{B \cup S}$. They imply that (i) holds. As for (ii), if $\mathfrak{G}$ is structural Markov on $\mathfrak{U}$, then

$$\begin{aligned} \pi(G) &= \pi(G_{A \cup S} \otimes G_{B \cup S}) \\ &= \pi(G_{A \cup S} | \{G \in \mathfrak{U}(A, B, S)\}) \pi(G_{B \cup S} | \{G \in \mathfrak{U}(A, B, S)\}), \end{aligned}$$

and similarly we can have the same result for $\pi(G')$. From (i), we have

$$\begin{aligned} \pi(G_{A \cup S} \otimes G'_{B \cup S}) = &\pi(G_{A \cup S} | \{G \in \mathfrak{U}(A, B, S)\}) \\ &\times \pi(G'_{B \cup S} | \{G \in \mathfrak{U}(A, B, S)\}), \end{aligned}$$

and so is $\pi(G'_{A\cup S} \otimes G_{B\cup S})$. Thus, our results by the above arguments follow. $\square$

Proof of Proposition 4. Let $H_j = \bigcup_{i=1}^{j} U_i$, $S_j = U_j \cap H_{j-1}$, $j \in \{2,3,\ldots,k\}$. Since $(H_{j-1} \setminus S_j, U_j \setminus S_j, S_j)$ forms a prime decomposition of G_{H_j}, for each $j \in \{2,3,\ldots,k\}$, we have

$$G_{H_j} = G_{U_j} \otimes G_{H_{j-1}}.$$

For $j = k$, since $S_k = U_k \cap H_{k-1}$ is complete, then

$$G_{S_k} = G_{H_k}^{(S_k)} = G_{U_k}^{(S_k)} \otimes G_{H_{k-1}}^{(S_k)}.$$

Whence we have

$$\pi(G_{H_k})\pi(G_{S_k}) = \pi(G_{U_k} \otimes G_{H_{k-1}})\pi(G_{U_k}^{(S_k)} \otimes G_{H_{k-1}}^{(S_k)}).$$

By Proposition 3, we can obtain

$$\pi(G_{H_k})\pi(G_{S_k}) = \pi(G_{U_k} \otimes G_{U_k}^{(S_k)})\pi(G_{H_{k-1}} \otimes G_{H_{k-1}}^{(S_k)})$$
$$= \pi(G_{U_k})\pi(G_{H_{k-1}}).$$

The equation (1) can be obtained recursively. $\square$

Proof of Proposition 5. Suppose that (A, B, S) forms a prime decomposition of G. Since G can be graphical collapsible onto $A \cup S$, by Theorem 1, $\theta_{A\cup S}$ only takes values in $\mathcal{P}(G_{A\cup S})$. This implies that $X_{A\cup S}$ can be obtained from $\theta_{A\cup S}$ actually. Then, we obtain

$$X_{B\cup S} \perp\!\!\!\perp \theta_{A\cup S} | \theta_{B\cup S}. \tag{A3}$$

From (A3), we deduce
$$X_{B\cup S} \perp\!\!\!\perp \theta_{A\cup S} | (\theta_{B\cup S}, \theta_S). \tag{A4}$$

By the meaning of the hyper Markov property, it follows that $\theta_{A\cup S} \perp\!\!\!\perp \theta_{B\cup S} | \theta_S$. Combing this with (A4) and the axioms of conditional independence gives

$$\theta_{A\cup S} \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S}) | (X_S, \theta_S), \tag{A5}$$

which implies that
$$X_{A\cup S} \perp\!\!\!\perp (X_{B\cup S}, \theta_{B\cup S}) | (X_S, \theta_{A\cup S}). \tag{A6}$$

Together, (A5) and (A6) yield the result. The proof for the strong case follows similar steps. $\square$

Proof of Theorem 2. The weak hyper Markov property states that

$$\theta_{A\cup S} \perp\!\!\!\perp \theta_{B\cup S} | (\theta_S, G, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A7}$$

Since $G \in \mathfrak{U}(A, B, S)$, then $G = G_{A\cup S} \otimes G_{B\cup S}$. Thus, from (A7) we deduce

$$\theta_{A\cup S} \perp\!\!\!\perp \theta_{B\cup S} | (\theta_S, G_{A\cup S}, G_{B\cup S}, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A8}$$

From Proposition 6, we obtain

$$\theta_{A\cup S} \perp\!\!\!\perp G_{B\cup S} | (\theta_S, G_{A\cup S}, \{G \in \mathfrak{U}(A, B, S)\}),$$

which gives the result that

$$\theta_{A\cup S} \perp\!\!\!\perp (\theta_{B\cup S}, G_{B\cup S}) | (\theta_S, G_{A\cup S}, \{G \in \mathfrak{U}(A, B, S)\}) \tag{A9}$$

by combining with (A8).

Again, by Proposition 6 and the structural Markov property, we have

$$G_{A \cup S} \perp\!\!\!\perp (\theta_{B \cup S}, G_{B \cup S}) | \{G \in \mathfrak{U}(A, B, S)\}.$$

Then, we have

$$G_{A \cup S} \perp\!\!\!\perp (\theta_{B \cup S}, G_{B \cup S}) | (\theta_S, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A10}$$

Thus, our result follows from (A9) and (A10). The proof for the strong case follows similar steps. □

Proof of Proposition 7. By Theorem 2, we obtain

$$(\theta_{A \cup S}, G_{A \cup S}) \perp\!\!\!\perp (\theta_{B \cup S}, G_{B \cup S}) | (\theta_S, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A11}$$

From (A11), we deduce

$$\theta_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | (\theta_S, \{G \in \mathfrak{U}(A, B, S)\}).$$

Whence we have

$$\theta_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | (X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A12}$$

By conditional independence property and Theorem 2,

$$X_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | (X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\}).$$

Thus, we have

$$X_{A \cup S} \perp\!\!\!\perp G_{B \cup S} | (X_S, \theta_S, \theta_{A \cup S}, \{G \in \mathfrak{U}(A, B, S)\}), \tag{A13}$$

which combines with (A12) to give the result. The proof of the strong case is similar, so we omit it for simplicity. □

Proof of Theorem 3. Since X is a random sample from θ and $\mathcal{L}(\theta)$ is hyper Markov with respect to G, then by Proposition 5,

$$(X_{A \cup S}, \theta_{A \cup S}) \perp\!\!\!\perp (X_{B \cup S}, \theta_{B \cup S}) | (X_S, \theta_S, G, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A14}$$

Since $G \in \mathfrak{U}(A, B, S)$, $G = G_{A \cup S} \otimes G_{B \cup S}$. Then, from (A14) we can find that

$$(X_{A \cup S}, \theta_{A \cup S}) \perp\!\!\!\perp (X_{B \cup S}, \theta_{B \cup S}) | (X_S, \theta_S, G_{A \cup S}, G_{B \cup S}, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A15}$$

From Proposition 7, we obtain

$$(X_{A \cup S}, \theta_{A \cup S}) \perp\!\!\!\perp G_{B \cup S} | (X_S, \theta_S, G_{A \cup S}, \{G \in \mathfrak{U}(A, B, S)\}),$$

which combines with (A15) to give the result that

$$(X_{A \cup S}, \theta_{A \cup S}) \perp\!\!\!\perp (X_{B \cup S}, \theta_{B \cup S}, G_{B \cup S}) | (X_S, \theta_S, G_{A \cup S}, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A16}$$

Additionally, from the structurally Markov property and Proposition 7, we have

$$G_{A \cup S} \perp\!\!\!\perp (X_{B \cup S}, \theta_{B \cup S}, G_{B \cup S}) | (X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A17}$$

From (A17), we deduce

$$G_{A \cup S} \perp\!\!\!\perp (X_{B \cup S}, \theta_{B \cup S}, G_{B \cup S}) | (X_{B \cup S}, \theta_{A \cup S}, X_S, \theta_S, \{G \in \mathfrak{U}(A, B, S)\}). \tag{A18}$$

So, the result follows from (A16) and (A18). Similar proof can be given for the case of the strong hyper Markov law. □

Proof of Proposition 10. We first give the proof of (i). Suppose that $G, G' \in \mathfrak{U}$. If $\mathfrak{G}$ is structural Markov on $\mathfrak{U}$, then we have

$$
\Lambda(G' : G) = \frac{\pi(G')}{\pi(G)} = \frac{\prod\limits_{j=1}^{k} \pi(G'_{U_j})}{\prod\limits_{j=2}^{k} \pi(G'_{S_j})} \times \frac{\prod\limits_{j=2}^{k} \pi(G_{S_j})}{\prod\limits_{j=1}^{k} \pi(G_{U_j})} \tag{A19}
$$
$$
= \frac{\eta_{U'_j}}{\eta_{U_j}},\ j \in \{1, 2, \ldots, k\}.
$$

The proof of (ii) is given as follows. It is obvious that W' is prime in G'. Consequently, G' has a perfect maximal prime subgraphs sequence $(U_1, \ldots, U_{j-1}, W', U_{j+2}, \ldots, U_k)$, and then the Equation (7) follows by using (i). □

Proof of Proposition 11. The proof of (i) follows similar steps to that of Proposition 10. To give the proof of (ii), let $\mathcal{T}$ be the junction tree with vertices being all maximal prime subgraphs of G. The construction of $\mathcal{T}$ can be referred to [21]. Since u, v are in two different maximal prime subgraphs of G, we then connect the U_i and U_j in $\mathcal{T}$. Then, we will obtain a unique cycle. Without a loss of generality, the vertices on this cycle are denoted by $(U_i = U_{h_1}, U_{h_2}, \ldots, U_{h_m} = U_j)$, where U_{h_t} and $U_{h_{t+1}}$ are connected by an edge in $\mathcal{T}$. Then, it is easy to see that $\mathcal{U} \setminus \{U_i = U_{h_1}, U_{h_2}, \ldots, U_{h_m} = U_j\} \cup T$ is the set of all the maximal prime subgraphs of G'. So, by applying (i), Equation (8) follows. □

References

1. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; Adaptive Computation and Machine Learning; MIT Press: Cambridge, MA, USA, 2009.
2. Lauritzen, S.L. *Graphical Models*; Oxford University Press: New York, NY, USA, 1996.
3. Richardson, T. A factorization criterion for acyclic directed mixed graphs. In Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 18–21 June 2009.
4. Richardson, T.; Spirtes, P. Ancestral graph Markov models. *Ann. Stat.* **2002**, *30*, 962–1030. [CrossRef]
5. Iqbal, K.; Buijsse, B.; Wirth, J. Gaussian Graphical Models Identify Networks of Dietary Intake in a German Adult Population. *J. Nutr. Off. Organ Am. Inst. Nutr.* **2016**, *146*, 646–652. [CrossRef] [PubMed]
6. Larranaga, P.; Moral, S. Probabilistic graphical models in artificial intelligence. *Appl. Soft Comput.* **2011**, *11*, 1511–1528. [CrossRef]
7. Verzilli, C.J.; Stallard, N.; Whittaker, J.C. Bayesian graphical models for genomewide association studies. *Am. J. Hum. Genet.* **2006**, *79*, 100–112. [CrossRef] [PubMed]
8. Giudici, P.; Green, P.J. Decomposable graphical Gaussian model determination. *Biometrika* **1999**, *86*, 785–801. [CrossRef]
9. Madigan, D.; Raftrey, A.E. Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Amer. Stat. Assoc.* **1994**, *89*, 1535–1546. [CrossRef]
10. Byrne, S.; Dawid, A.P. Structural Markov graph laws for Bayesian model uncertainty. *Ann. Stat.* **2015**, *43*, 1647–1681. [CrossRef]
11. Li, B.C. Support condition for equivalent characterization of graph laws. *Sci. Sin. Math.* **2022**, *52*, 467–474.
12. Dawid, A.P.; Lauritzen, S.L. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Stat.* **1993**, *21*, 1272–1317. [CrossRef]
13. Green, P.J.; Thomas, A. A structural Markov property for decomposable graph laws that allows control of clique intersections. *Biometrika* **2018**, *105*, 19–29. [CrossRef]
14. Leimer, H.G. Optimal decomposition by clique separators. *Discret. Math.* **1993**, *113*, 99–123. [CrossRef]
15. Dawid, A.P. Conditional independence in statistical theory. *J. R. Stat. Soc. B.* **1979**, *41*, 1–15. [CrossRef]
16. Dawid, A.P. Conditional independence for statistical operations. *Ann. Stat.* **1980**, *8*, 598–617. [CrossRef]
17. Meek, C. *Strong Completeness and Faithfulness in Bayesian Networks*; Morgan Kaufmann: San Francisco, CA, USA, 1995.
18. Hoff, P.D. Extending the rank likelihood for semiparametric copula estimation. *Ann. Appl. Stat.* **2007**, *23*, 103-122.
19. Frydennberg, M; Lauritzen, S.L. Decomposition of maximum likelihood in mixed graphical interaction models. *Biometrika* **1989**, *76*, 539–555.

20. Asmussen, S.; Edwards, D. Collapsibility and response variables in contingency tables. *Biometrika* **1983**, *70*, 567–578. [CrossRef]
21. Wang, X.F.; Guo, J.H. Junction trees of general graphs. *Front. Math. China* **2008**, *3*, 399–413. [CrossRef]