Mutual conditional independence and its applications to model selection in Markov networks

Niharika Gauraha¹ · Swapan K. Parui²

Published online: 21 July 2020 © The Author(s) 2020

Abstract

The fundamental concepts underlying Markov networks are the conditional independence and the set of rules called Markov properties that translate conditional independence constraints into graphs. We introduce the concept of mutual conditional independence in an independent set of a Markov network, and we prove its equivalence to the Markov properties under certain regularity conditions. This extends the notion of similarity between separation in graph and conditional independence in probability to similarity between the mutual separation in graph and the mutual conditional independence in probability. Model selection in graphical models remains a challenging task due to the large search space. We show that mutual conditional independence property can be exploited to reduce the search space. We present a new forward model selection algorithm for graphical log-linear models using mutual conditional independence. We illustrate our algorithm with a real data set example. We show that for sparse models the size of the search space can be reduced from $O(n^3)$ to $O(n^2)$ using our proposed forward selection method rather than the classical forward selection method. We also envision that this property can be leveraged for model selection and inference in different types of graphical models.

Keywords Markov networks · Mutual conditional independence · Graphical models · Graphical log-linear models · Forward model selection

Mathematics Subject Classification (2010) 97K99 · 00A69 · 97R99

Niharika Gauraha
niharika@kth.se

Swapan K. Parui
swapan.parui@gmail.com

¹ Division of Computational Science and Technology, KTH Royal Institute of Technology, Stockholm, Sweden

² Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata, India
1 Introduction

A Markov network is a way of specifying conditional independence constraints between components of a multivariate distribution. Markov properties are the set of rules that determine how conditional independence constraints are translated into a graph (see [11] and [10]). The three Markov properties usually considered for Markov networks are pairwise, local and the global Markov properties. These Markov properties are equivalent to one another for positive probability distributions, see [12]. We introduce the concept of mutual conditional independence for an independent set of a Markov network. We derive an alternative formulation for the three Markov properties (local, pairwise and global) using mutual conditional independence. This alternative formulation is then used to prove the equivalence between mutual conditional independence property (MCIP) and Markov properties, under positive probability distribution assumption. This extends the notion of similarity between separation in graph and conditional independence in probability to similarity between the mutual separation in graph and the mutual conditional independence in probability.

We study the applicability of MCIP for learning and inference in graphical models. Model selection in graphical models still remains a difficult problem due to the large search space, we show that mutual conditional independence property can be used to reduce the search space. We propose a new forward model selection algorithm for graphical log-linear models where we exploit mutual conditional independence check to reduce the search space. We illustrate our algorithm with a real data set example. For some sparse graphs, we show that the size of the search space can be reduced drastically from $O(n^3)$ to $O(n^2)$ using our proposed forward selection algorithm rather than classical forward selection method. We also discuss that in general for any sparse and unknown underlying graph structure, the reduction in the size of the search space depends on the its maximum sized independent sets. Finally, we envision that the MCIP can be leveraged for learning and inference in different types of graphical models. Our contributions are summarized as follows:

- We introduce the concept of mutual conditional independence in an independent set of a Markov network, and we prove its equivalence to the Markov properties.
- We extend the concept of similarity between separation in graph and conditional independence in probability to similarity between mutual separation in graph and mutual conditional independence in probability.
- We propose a new forward model selection algorithm for graphical log-linear models, where mutual conditional independence is employed to reduce the search space.
- We show that, for sparse models the size of the search space can be reduced from $O(n^3)$ to $O(n^2)$ using our proposed forward selection method rather than the classical forward selection method.
- We also envision that this property can be leveraged for model selection and inference in different types of graphical models.

The discussion below is organized as follows. In Section 2, we start with a brief overview and the mathematical foundations of the theory of Markov networks and Graphical Log-Linear Models (GLLM). In Section 3 a proof is provided that MCIP holds within the elements of an independent set. Section 4 involves deriving the global Markov property using the MCIP and proving equivalence between them. In Section 5, we discuss an application of mutual conditional independence relations in terms of model selection for GLLMs. In Section 6, we give computational details that we used for model selection. In Section 7, we conclude and discuss future scope and applicability of MCIP. The proposed model selection algorithm is illustrated step by step with a real data set example in Appendix.
2 An overview and mathematical foundations

A graphical model is a technique for representation of the conditional independences between variables in a multivariate probability distribution. The nodes or vertices in the graph correspond to random variables. The absence of an edge between two random variables denotes a conditional independence relation between them. In the literature, several classes of graphs with various conditional independence interpretations have been described. Undirected graphical models (Markov Networks) and directed acyclic graphs based graphical models (Bayesian Networks) are the most commonly known graphical models. In this article we only consider undirected graphical models, also known as Markov random fields or Markov networks. For details on the foundation of the theory of Markov networks, see [11, 15, 20] and [17].

We now briefly discuss the necessary background theory, present some existing results and introduce notations that will be used throughout the paper.

2.1 Graph theory

A graph $G$ is a pair $G = (V, E)$, where $V$ is the set of vertices and $E$ is the set of edges. A graph is said to be an undirected graph if its vertices are connected by undirected edges. We consider only simple graphs that have neither self loops nor multiple edges. An independent set of a graph $G$ is a subset $S \subset V$ such that no two nodes in $S$ are adjacent. An independent set is said to be maximal if no node can be added to $S$ without violating the independent set property. A clique of a graph $G$ is a subset $C$ of the vertices such that all vertices in $C$ are mutually adjacent. A clique is said to be maximal if no vertex can be added to $C$ without violating the clique property. Note that given a complete list of maximal independent sets or a complete list of maximal cliques of a graph, the graph is uniquely determined.

2.2 Conditional independence

Suppose that $X$, $Y$ and $Z$ are random variables with joint distribution $f$. The random variables $X$ and $Y$ are said to be conditionally independent given the variable $Z$, denoted by $X \perp\perp Y \mid Z$, if the following holds.

$$f(X, Y \mid Z) = f(X \mid Z)f(Y \mid Z)$$ (1)

$$f(X \mid Y, Z) = f(X \mid Z)$$ (2)

Next, we define some properties of conditional independence in terms of graphoid axioms. For an alternative set of the complete axioms, see [8].

Symmetry : $X \perp\perp Y \mid Z \iff Y \perp\perp X \mid Z$ (3)

Decomposition : $X \perp\perp (Y \cup W) \mid Z \iff X \perp\perp Y \mid Z$ (4)

WeakUnion : $X \perp\perp (Y \cup W) \mid Z \iff X \perp\perp Y \mid (Z \cup W)$ (5)

Contraction : $X \perp\perp (Y) \mid Z$ and $X \perp\perp (W) \mid (Z \cup Y) \implies X \perp\perp (Y \cup W) \mid (Z)$ (6)

Intersection : $X \perp\perp (Y) \mid (Z \cup W)$ and $X \perp\perp (W) \mid (Z \cup Y) \implies X \perp\perp (Y \cup W) \mid (Z)$ (7)

A semi-graphoid is a dependency model which satisfies the first four properties, listed above. If, in addition, the intersection property (7) holds, it is called a graphoid. In probability theory, conditional independence as defined in (2), is a semi-graphoid. If $f$ is a positive probability distribution, then conditional independence is a graphoid. Graph separation in...
an undirected graph satisfies graphoid axioms. For further details on graphoid axioms, see [11] and [6].

### 2.3 Markov properties of undirected graphs

We define the following three Markov properties for undirected graphs. Let \( G = (V, E) \) be an undirected graph and \( f \) be a probability distribution over \( G \). The probability distribution \( f \) satisfies the pairwise Markov property \((P)\) for the graph \( G \), if for every pair of non-adjacent vertices \( X \) and \( Y \), \( X \) is independent of \( Y \) given the rest.

\[
X \perp\!\!\!\!\perp Y \mid (V \setminus X, Y) \tag{8}
\]

It satisfies the local Markov property \((L)\), if every variable \( X \) is conditionally independent of its non-neighbours in the graph, given its neighbours.

\[
X \perp\!\!\!\!\perp (V \setminus X \cup bd(X)) \mid bd(X), \tag{9}
\]

where \( bd(X) \) denotes boundary or neighbors of \( X \). The global Markov property \((G)\) is said to be satisfied if for any disjoint subsets of nodes \( A, B, C \) such that \( C \) separates \( A \) and \( B \) on the graph the distribution satisfies the following:

\[
A \perp\!\!\!\!\perp B \mid C. \tag{10}
\]

A probabilistic independence model that satisfies graphoid axioms with respect to \( G \), the following holds. For the proof, see [11, 14] and [6].:

\[
(G) \iff (L) \iff (P). \tag{11}
\]

### 2.4 Markov network graphs and Markov network

A Markov network graph is an undirected graph \( G = (V, E) \) where \( V = \{X_1, X_2, ..., X_n\} \) corresponds to the random variables of a multivariate distribution. A Markov network is a tuple \( M = (G, \psi) \) where \( G \) is a Markov network graph, \( \psi = \{\psi_1, \psi_2, ..., \psi_m\} \) is a set of non-negative functions for each clique \( C_i \in G \forall i = 1 \ldots m \) and the joint probability density function (pdf) can be decomposed into factors as

\[
f(x) = \frac{1}{Z} \prod_{a \in C_m} \psi_a(x),
\]

where \( Z \) is the normalizing constant.

**Theorem 1** (Hammersley-Clifford theorem) \( \) Let \( M = (G, \psi) \) be a Markov network and let the probability density function \( f \) of \( X = \{X_1, X_2, ..., X_n\} \) be positive. Then \( X \) satisfies the global Markov property with respect to the graph \( G \) if and only if it factorizes as follows.

\[
f(x) = \frac{1}{Z} \prod_{a \in C_m} \psi_a(x),
\]

where \( C_m \) are the maximal cliques of \( G \) and \( \psi_a(x) \) depends on \( x \) through \( x_a = (x_v)_{v \in a} \) only.

It follows from the above theorem that if a positive probability distribution factorizes with respect to \( G \), then it also satisfies all Markov properties (pairwise, local and global) with respect to \( G \), see [13].
2.5 Graphical log linear models

In this section, we briefly review graphical log-linear models for contingency tables. A contingency table is a table of counts that summarizes relationship between factors or categorical variables (see [2]). In the context of contingency tables, a Log Linear Model (LLM) is a linear model in the log scale of the expected cell counts. The LLM basically models the association and interaction among factors of a contingency table. For illustration, we consider a three dimensional table with factors X, Y and Z. Suppose, the factors X, Y and Z have I, J and K levels respectively, then we have an $I \times J \times K$ contingency table. The following notations are defined for each elementary cell $(i, j, k)$ for $i = 1 \ldots I$, for $j = 1 \ldots J$ and for $k = 1 \ldots K$

$n_{ijk} =$ the observed counts in the cell $(i, j, k)$
$m_{ijk} =$ the expected counts in the cell $(i, j, k)$
$\hat{m}_{ijk} =$ the Maximum Likelihood Estimate of $m_{ijk}$

The following notations are used for sums of elementary cell counts. where “.” represents summation over that factor.

$n_{i..} = \sum_{jk} n_{ijk}$

$L = n_{..} =$ total number of observations

We use the terms [X], [Y] and [Z] to denote the main effects, and the terms [XY], [YZ], [XZ] and [XYZ] denote two-factor and three-factor interactions. We are particularly interested in a class of LLMs that can be represented by graphs, called graphical log-linear models (GLLMs). In GLLMs, the vertices correspond to the factors and the edges correspond to the two-factor interactions. A LLM is said to be graphical if it contains all the lower order terms which can be derived from the variables contained in a higher-order term, then the model also contains the higher order interactions. In the previous example of a three-factor contingency table, if a model includes all the two factor interactions [XY], [YZ] and [XZ], then it must also contain the three factor interaction [XYZ]. We usually represent a graphical model as a set of maximal cliques, which is [XYZ] in this case. We note that there is a one-to-one correspondence between GLLMs and graphs (see [4]). For more details on GLLM we refer to the books [3, 4] and [1] and a review article [7].

3 Mutual conditional independence in Markov networks

In this section, we prove that the elements of an independent set are mutually conditionally independent given the rest.

Theorem 2 (Mutual Conditional Independence in Markov networks) Let $G$ be a Markov network graph and $f$ be a positive probability distribution that supports the conditional independence relations required to satisfy pairwise (= local=global) Markov property for $G$, then the elements of an independent set $I$ of $G$ are mutually conditionally independent given the rest $(V \setminus I)$.

Proof Without loss of generality we can assume that the first $k$ elements of $X$ form an Independent set $I$. Let $I = \{X_1, X_2, ..., X_k\}$ be an independent set of $G$. Since $\{X_1, X_2, ..., X_k\}$

\[ \]
are mutually non-adjacent, when we condition on \( V \setminus I \) or equivalently when we remove the nodes \( V \setminus I \) from \( G \), the remaining vertices \( \{X_1, X_2, ..., X_k\} \) are disconnected which implies in probability complete “conditional” independence among the vertices of \( I \).

Since \( \{X_1, X_2, ..., X_k\} \) form an independent set, they belong to separate cliques, say \( X_i \in C_i \), for \( i = 1 \) to \( k \), where \( C_i \) is a maximal clique in \( G \). Without loss of generality we can assume that there are exactly \( k \) maximal cliques. From Theorem (Hammersley-Clifford theorem) the probability distribution \( f \) factorizes as follows.

\[
f \propto \psi_1(X_1, Y_1) \psi_2(X_2, Y_2) ... \psi_k(X_k, Y_k)
\]

where \( Y_i \) is the set of nodes that connects two or more \( C_i \)'s and each \( \{X_i, Y_i\} \) forms a maximal clique in \( G \) (see Fig. 1). It can be noted that \( Y_i \) can be empty in case of a disconnected graph and also \( \bigcup Y_i = V \setminus I \).

Then conditional probability \( f(I \mid V \setminus I) \) can be expressed as

\[
f(I \mid Y_1 = y_1, ..., Y_k = y_k) \propto \psi_1(X_1, y_1) \psi_2(X_2, y_2) ... \psi_k(V_k, k_k)
\]

\[
\implies f(I \mid V \setminus I) \propto \phi_1(X_1) \phi_2(X_2) ... \phi_k(X_k),
\]

where each \( \phi_i \) is a function of the corresponding variable \( X_i \) only as the corresponding \( y_i \) is fixed, i.e., \( \phi_1(X_1) = \psi_1(X_1, y_1) \). Hence \( \{X_1, X_2, ..., X_k\} \) are mutually conditionally independent given \( \{V \setminus I\} \).

\[\square\]

4 Mutual conditional independence and the Markov properties

In this section, we represent an alternative way to derive conditional independence relations required for satisfying the Markov properties of the Markov networks. More specifically, we prove equivalence between Mutual Conditional Independence Property (MCIP) and pairwise Markov property, and from (11) it follows that MCIP is equal to the local and global Markov properties as well.

**Theorem 3** (Equivalence of MCIP and Markov properties) Let \( G \) be a Markov network graph and let \( f \) be a positive probability distribution over \( G \). The probability distribution \( f \)

![Fig. 1](image.png) An example of an independent set in a graph, where dotted lines denote that on removal of the sets of nodes \( Y_1, ..., Y_k \), the nodes \( X_1, ..., X_k \) are disconnected
satisfies the mutual conditional independence relations implied by the maximal independent sets of $G$ if and only if the conditional independence relations required to satisfy pairwise (local and global) Markov property for $G$ also holds in $f$.

Proof. We prove equivalence of MCIP and pairwise Markov property. Then under the assumption of positive distribution, MCIP is equal to the local and the global Markov properties as well.

Let $C_1, C_2, \ldots, C_m$ be the complete list of the maximal independent sets of nodes of $G$ then $V = C_1 \cup C_2 \cdots \cup C_m$ and $E = \{(x, y) : (x, y) \notin C_i \forall i = 1 \cdots m\}$.

First, we assume that the conditional independence relations required to satisfy pairwise Markov property holds in $f$ with respect to $G$. It follows that the elements of $C_i$ are mutually conditionally independent given the set $V \setminus C_i$ from Theorem 2. It follows that the pairwise Markov property $\implies$ MCIP.

We recall that the Mutual conditional independence implies pairwise conditional independence. Since the elements of a $C_i$ are mutually conditionally independent given the set $V \setminus C_i$, they are also pairwise independent given $V \setminus C_i$. Let us suppose that $\{x, y\} \in V$ is a pair of nodes which is not an edge. Then $\{x, y\}$ is an independent set and thus is contained in some $C_i$ and hence is pairwise independent given the rest. Therefore,

pairwise Markov Property $\iff$ MCIP

We illustrate the proof by an example as follows. Let us consider the Markov network as given in Fig. 2.

The set of all maximal independent sets for this graph is:

$$S = \{\{A, C, F\}, \{A, C, G\}, \{A, E\}, \{B, D, F\}, \{B, D, G\}, \{B, E\}\}$$

First consider the maximal independent set $C_1 = \{A, C, F\}$, and let us suppose that the MCIP holds which implies that $A$, $C$ and $F$ are mutually independent given the rest $\{B, D, E, G\}$. Or, equivalently independence relation can be expressed as:

$$A \perp\!\!\!\!\!\perp C \perp\!\!\!\!\!\perp F \mid (B, D, E, G)$$

Applying weak union graphoid axiom (5) to the above independence relation we get

$$A \perp\!\!\!\!\!\perp F \mid (B, D, E, G) \cup C$$

$$C \perp\!\!\!\!\!\perp F \mid (B, D, E, G) \cup A$$

$$A \perp\!\!\!\!\!\perp C \mid (B, D, E, G) \cup F$$

Fig. 2 A Markov network graph
Applying similar arguments for the other sets of maximal independent sets it can be shown that for every non-adjacent pair \( x, y \in V \)

\[
x \perp \!
\!\!\perp y \mid V \setminus \{x, y\},
\]

which is also a definition of pairwise Markov property.

Conversely, suppose that the pairwise Markov property holds. We now show that MCIP also holds. From Theorem 1, it is clear that under positive distribution assumption, the probability distribution \( f \) satisfies pairwise Markov property with respect to the graph \( G \) if and only if it factorizes as follows.

\[
f(x) \propto \psi_1(A, B) \psi_2(A, D) \psi_3(B, C) \psi_4(C, D, E) \psi_5(E, F, G)
\]

Consider the pdf of \( (A, C, F) \) conditioned on \( (B, D, E, G) \), we get the factorization of pdf as follows.

\[
f(A, C, F \mid B = b, D = d, E = e, G = g) \propto \psi_1(A, b) \psi_2(A, d) \psi_3(C, b) \psi_4(C, d, e) \psi_5(F, e, g) \propto \phi_1(A) \phi_2(C) \phi_3(F)
\]

From the above factorization of pdf, it follows that \( (A, C, F) \) are mutually independent conditioned on \( (B, D, E, G) \).

Similarly, it can be shown that the mutual conditional independence relations hold for the remaining maximal independent sets. Hence it follows that

\[
MCIP \iff \text{pairwise Markov property}
\]

Applying (11), we get the following equivalence relation that completes the proof.

\[
MCIP \iff P \iff L \iff G
\]

## 5 Applications and illustrations

In this section we develop a forward model selection algorithm for graphical log-linear models exploiting mutual conditional independence property, and we illustrate the algorithm through an example. We also discuss how our proposed algorithm is better than the classical forward model selection algorithm in terms of reduced search space (and hence in terms of reduced computational complexity). In particular, we show that the size of the search space explored by the proposed forward selection algorithm is reduced to \( \mathcal{O}(n^2) \) for the case of star graphs, which is \( \mathcal{O}(n^3) \) for the classical forward selection methods.

### 5.1 Model selection using mutual conditional independence

Estimation of conditional independence structures is an important problem. In the context of GLLMs, the goal of model selection is to choose a smallest graphical model from a class of graphical models under consideration that best fits the data and has the least number of edges (i.e., the number of interaction terms). Since computing the exact independence structure is intractable, greedy algorithms are an appropriate way to tackle this problem. Most existing model selection algorithms are based on forward selection, backward elimination or a combination of both. For detailed discussions on model selection in a graphical log-linear case, we refer to [3–5, 9] and [19]. These greedy algorithms mostly use the conditional independence property as the model selection criterion, and thus may still have a huge search space even for a modest number of variables. We propose a forward model
selection algorithm which is based on the concept of MCIP. We use the mutual conditional independence check to reduce the search space for the local search algorithm.

In this approach, we start with the null model (a complete independence model) and our main focus is to find All the Maximal Independent Sets (AMIS) of the underlying graphical model based on the data sample. We maintain two lists tempAMIS and AMIS. The tempAMIS contains a list of subsets of factors to be tested for MCIP and the AMIS list contains MISs (for which the data supports MCIP). At each step, first we move all the elements from tempAMIS to AMIS for which the data supports MCIP. Then we pick a largest set from tempAMIS. The most significant edge between its elements is found and the required two factor and higher order terms are added into the present model. Then we split all the sets in tempAMIS that contain both end points of the newly added edge. At this point tempAMIS may contain duplicates or proper subsets within tempAMIS or members of AMIS. We eliminate the duplicates and make tempAMIS irreducible such that no elements of tempAMIS are proper subsets of another member of tempAMIS or any member of AMIS. We repeat the process until tempAMIS is empty. If the union of all members of AMIS is not the exhaustive set of all nodes, then create singleton sets for each missing node and add these singleton sets to AMIS. Finally, the algorithm returns AMIS that determines the graph structure uniquely.

For example, let us consider a five-factor contingency table, where the factors are denoted as 1, 2, 3, 4, and 5. Initial model assumption is the null model (all factor form an MIS), tempAMIS = \{\{1, 2, 3, 4, 5\}\} and AMIS = \{\emptyset\}. Let us suppose that the complete independence model does not fit the data and also assume that the edge (1, 2) is the most significant edge. We add this edge into the complete independence model. Since, we assumed that the set \{1, 2, 3, 4, 5\} is a MIS at the beginning, but after adding the edge (1, 2), it is no longer a MIS. In fact, now the assumption for the MISs are tempAMIS = \{\{1, 3, 4, 5\}, \{2, 3, 4, 5\}\}. Now let us suppose that the data supports the MCI condition for the set \{2, 3, 4, 5\}. So we remove it from tempAMIS and add it to AMIS. As the next step we try to find the most significant edge within the set \{1, 3, 4, 5\}. It is important to note that the node pairs for consideration are only (1, 3), (1, 4) and (1, 5), since the remaining combinations formed by the subset \{3, 4, 5\} are contained in the set \{2, 3, 4, 5\} which is assumed to be an independent set. Under the assumption that the set \{2, 3, 4, 5\} is an independent set, the node pairs (3, 4), (3, 5) and (4, 5) are also pairwise independent and they will not be considered as candidate edges. At each step the procedure continues in this way until tempAMIS becomes empty. Finally, if the union of all members of AMIS is not the exhaustive set of all nodes \{1, 2, 3, 4, 5\}, then create singleton sets for each missing node and add those singleton sets to the list AMIS.

### 5.2 An illustration of MCIP based model selection algorithm

Since the proposed algorithm (Algorithm 1) is more notational than conceptual, we begin the development with an example. First, we define the test statistic used to compare models in the following, where a smaller model is tested against a (larger) saturated model to see whether the smaller model is an inadequate explanation of the data. The likelihood ratio chi-square test statistic \(G^2\), to test a model against the saturated model is defined as

\[
G^2 = 2 \sum_i O_i \log \frac{O_i}{E_i},
\]

where \(O\) denotes the observed cell count and \(E\) denotes the expected cell count. Under the null hypothesis \(G^2\) is distributed as \(\chi^2\) with the appropriate degrees of freedom (see [4]). In general, to test a model \(M_1\) against another model \(M_2\), where the model \(M_2\)
Algorithm 1 Forward selection algorithm.

**Input:** A $p$-factor Contingency table, factors are labelled as $\{X_1, X_2, \ldots, X_p\}$

**Output:** AMIS:= All Maximal Independent Set

**Initialize:**

$AMIS = NULL$;

$tempAMIS = \{(X_1, X_2, \ldots, X_p)\}$

**while** $tempAMIS \neq NULL$ **do**

  **for each** $s \in tempAMIS$ **do**
  
  if MCI relation holds amongst factors in $s$ then
  
  $tempAMIS = tempAMIS - s$ ;
  
  $AMIS = AMIS \cup s$ ;

  **end**

**end**

**if** $tempAMIS \neq NULL$ **then**

  Select the first largest element, $s \in tempAMIS$

  Find the most significant edge $(v_1, v_2) \subset s$, such that $(v_1, v_2) \not\subset s' \in AMIS$

  **for each** $t \in tempAMIS$ **do**

  if $(v_1, v_2) \in t$ then

  $t_1 = t - v_1$ ;
  
  $t_2 = t - v_2$ ;

  $tempAMIS = \{tempAMIS - t\} \cup t_1 \cup t_2$ ;

  **end**

  **end**

  Eliminate duplicates from tempAMIS.

  Make tempAMIS irreducible (it does not contain a proper subset).

**end**

If the union of all members of AMIS is not exhaustive set of all nodes $\{X_1, \ldots, X_p\}$, then create singleton sets for each missing node and add those sets to the AMIS.

**return** $AMIS$

is strictly larger than the model $M_1$, we first compute the $G^2$ saturated model test for each model, say $G^2_1$ and $G^2_2$ for $M_1$ and $M_2$ respectively. Then the test of the adequacy of the smaller log-linear model $M_1$ can be obtained by subtraction from the saturated model test statistics as $G^2 = G^2_1 - G^2_2$. The degrees of freedom for the chi-square is the difference in the degrees of freedom for models $M_1$ and $M_2$. The advantage of using $G^2$ is that it simplifies the process of testing models against each other, see [4]. Note that we do not address the multiple testing problem here, and we refer to [18]. The topic itself is a separate research problem that is out of scope of this work.

**Example 1** (Forward Model Selection for the Rienis Dataset) We illustrate the proposed algorithm (Algorithm 1) using the Rienis dataset. For details about the Reinis dataset, see [16]. The Reinis data is shown in Table 1.

The results obtained in each iteration is summarized in Table 2. Each step is illustrated in more detail in Appendix. Note that each iteration involves the following 3 steps:

1. Testing for MCI: We first check if MCIP holds for each member of tempAMIS.
Table 1  The reinis dataset

| Family | Protein | Systol | Phys | Mental | Smoke | No | Yes | No | Yes |
|--------|---------|--------|------|--------|-------|----|-----|----|-----|
| neg    | < 3     | < 140  | no   |        |       | 44 | 40  | 112| 67  |
|        |         |        | yes  |        |       | 129| 145 | 12 | 23  |
|        |         |        | ≥ 140| no     |       | 35 | 12  | 80 | 33  |
|        |         |        |      | yes    |       | 109| 67  | 7  | 9   |
|        | ≥ 3     | < 140  | no   |        |       | 23 | 32  | 70 | 66  |
|        |         |        |      | yes    |       | 50 | 80  | 7  | 13  |
|        |         |        | ≥ 140| no     |       | 24 | 25  | 73 | 57  |
|        |         |        |      | yes    |       | 51 | 63  | 7  | 16  |
| pos    | < 3     | < 140  | no   |        |       | 5  | 7   | 21 | 9   |
|        |         |        | yes  |        |       | 9  | 17  | 4  | 1   |
|        |         |        | ≥ 140| no     |       | 4  | 3   | 11 | 8   |
|        |         |        |      | yes    |       | 14 | 17  | 5  | 2   |
|        | ≥ 3     | < 140  | no   |        |       | 7  | 3   | 14 | 14  |
|        |         |        |      | yes    |       | 9  | 16  | 2  | 3   |
|        |         |        | ≥ 140| no     |       | 4  | 0   | 13 | 11  |
|        |         |        |      | yes    |       | 5  | 14  | 4  | 4   |

2. Addition of a new edge: At each step we add the most significant edge as long as the significance level is below a cut-off value.

3. Redundancy elimination: Before moving to the next step we make sure that the list tempAMIS is irreducible and contains no duplicate.

Table 2  The illustration of iterations of the proposed forward models selection algorithm using Reinis dataset

| Iter | Edge | currModel | tempAMIS | AMIS |
|------|------|-----------|----------|------|
| 0    | —    | [A][B][C][D][E][F] | [A, B, C, D, E, F] | [∅] |
| 1    | BC   | [A][BC][D][E][F]   | [A, B, D, E, F], [A, C, D, E, F] | [∅] |
| 2    | BE   | [A][BC][BE][D][F]  | [A, B, D, F], [A, C, D, E, F] | [∅] |
| 3    | AC   | [AC][BC][BE][D][F] | [A, B, D, F], [A, D, E, F], [C, D, E, F] | [∅] |
| 4    | AD   | [AC][AD][BC][BE][F]| [A, B, F], [B, D, F], [A, E, F], [C, D, E, F] | [∅] |
| 5    | DE   | [AC][AD][BC][BE][DE][F]| [A, E, F], [C, D, F] | [A, D, F] | [∅] |
| 6    | AE   | [AC][ADE][BC][BE][F] | [A, B, F], [B, D, F], [C, D, F] | [C, E, F] | [∅] |

Iter corresponds to iteration number, Edge corresponds to the added edge in each iteration, currModel corresponds to the present model, tempAMIS corresponds to the AMIS list to be tested for mutual conditional independence and AMIS corresponds to the list of MISs.
The algorithm returns the AMIS \( \{ \{ A, D, F \}, \{ C, E, F \}, \{ B, D, F \} \{ B, E, F \} \} \). A graph structure can be determined uniquely from the AMIS as given in Fig. 3.

5.3 A comparison of the search spaces for forward selection algorithms

In this section, we compare the sizes of the search spaces explored by the classical forward selection algorithm (Chapter 6.1, [4]) vs the proposed forward selection algorithm (1). Though the worst case performance of the proposed algorithm is the same as the classical forward selection algorithm, we show that for some sparse models the size of the search space can be reduced from \( O(n^3) \) to \( O(n^2) \).

For illustration, we compare the total search spaces explored by the classical forward selection algorithm and the MCIP based forward selection algorithm for a simple example. Let us consider a star shaped graph, as given in Fig. 4. Let us suppose that the order of significance of the two factor interactions is \((1, 2), (1, 3), (1, 4)\) and \((1, 5)\). Assuming the graph structure is not known, the classical forward selection algorithm will search the space as follows:

1. To search for the most significant edge \((1, 2)\) it would search the entire \( \binom{5}{2} \) combination on top of the NULL or complete independence model \((1)(2)(3)(4)(5)\). We denote the intermediate model as \( M_1 = NULL + (1, 2) \).
2. To decide for the next significant edge it will search for the remaining entire possible edges \( \binom{5}{2} - 1 \). Suppose the outcome of this step is another intermediate model, \( M_2 = M_1 + (1, 3) \).
3. To decide for the next significant edge it will search for the remaining entire possible edges \( \binom{5}{2} - 2 \). We denote the outcome of this step as, \( M_3 = M_2 + (1, 4) \).
4. To decide for the last significant edge it will search for the remaining entire possible edges \( \binom{5}{2} - 3 \). The final model is \( M_4 = M_3 + (1, 5) \).
5. In total, the algorithm fits \( 4 \times \binom{5}{2} - (1 + 2 + 3) = 36 \) models to decide for the final star shaped graph shown in Fig. 4.

The above step can be generalized for any graph with \( n \) number of nodes, and \( k \) number of edges. The effective size of the search space can be given as

\[
(k) \times \binom{n}{2} = \frac{n(n-1)}{2} \quad \text{or} \quad \frac{(k-1)(k-2)}{2}
\]

\[
\approx O(k \times n^2) \approx O(n^3), \text{ when } k \sim n
\]

Now, let us look into the steps and the number of intermediate fits required by the proposed forward selection method using MCI test.
1. It will start with the NULL model, and tempAMIS = \{\{1, 2, 3, 4, 5\}\}. To search for the most significant edge \((1, 2)\) it will also search the entire \(\binom{5}{2}\) combinations on top of the NULL model. An intermediate model is, say, \(M_1 = NULL + (1, 2)\), and tempAMIS = \{\{1, 3, 4, 5\}, \{2, 3, 4, 5\}\}.

2. Before deciding for the next significant edge, the algorithm will detect that the set \(\{2, 3, 4, 5\}\) is an independent set, and it is removed from the list tempAMIS. Now, it will search for the most significant edge in the set \(\{1, 3, 4, 5\}\). Note that there are only three candidate edges \((1, 3)\), \((1, 4)\) and \((1, 5)\), since pairwise independence also holds for the remaining pairs of nodes \((3, 4)\), \((3, 5)\) and \((4, 5)\) as they are subsets of \(\{2, 3, 4, 5\}\) which is already shown to be an independent set. Suppose the outcome of this step is another intermediate model, \(M_2 = M_1 + (1, 3)\), and tempAMIS = \{\{1, 4, 5\}\}.

3. To decide for the next significant edge, it will search for the edges with combinations from the set \(\{1, 4, 5\}\), which requires two model fits considering the additional edges \((1, 4)\) and \((1, 5)\) on top of the previous intermediate model \(M_2\). Suppose the outcome of this step is another intermediate model, \(M_3 = M_2 + (1, 4)\), and tempAMIS = \{\{1, 5\}\}.

4. To decide for the last edge it will fit only one model. The final model is, \(M = M_3 + (1, 5)\) and tempAMIS is empty.

5. In total, the algorithm fits \(\binom{5}{2} + 3 + 2 + 1 = 16\) models to decide for the final star shaped graph.

The above step can be generalized for any star graph with \(n\) number of nodes, the effective size of the search space explored by the proposed algorithm can be given as follows.

\[
\frac{n}{2} + 1 + 2 + ... + (n - 2) \\
= \frac{n(n - 1)}{2} + \frac{(n - 1)(n - 2)}{2} \\
\approx O(n^2)
\]

In general, for any sparse graph, where the size of the maximum independent set(s) is proportional to \(n\), then irrespective of the underlying graph structure the search space can be reduced using our proposed algorithm. For example, for sparse graphs as given in Fig. 5 it is easy to show that the search space is reduced from \(O(n^3)\) to \(O(n^2)\). For sparse graph-2,
the max-sized independent sets are \{2, 4, 5, 6, 7\} and \{3, 4, 5, 6, 7\}, which can be revealed after a couple of iterations and the search space is reduced significantly from subsequent iterations, see Table 3. Also, for Sparse Graph-3 in (b) of Fig. 5, the max-sized independent set is \{4, 5, 6, 7, 8, 9\}, which can be explored after 3-4 iterations and thus the search space is reduced from cubic to square order of the number of nodes.

Table 3 The illustration of iterations and search spaces in terms of the number of candidate edges to explore in each iteration, for Sparse Graph-2

| Iter | Edge | currModel | tempAMIS                  | AMIS               | # of candidate edges |
|------|------|-----------|---------------------------|--------------------|----------------------|
| 0    | 12   | [1][2][3][4][5][6][7]  | \{1, 2, 3, 4, 5, 6, 7\} | \{\}               | 0                    |
| 1    | 13   | [12][13][4][5][6][7]  | \{1, 3, 4, 5, 6, 7\},  | \{\}               | 0                    | (7)                 |
| 2    | 23   | [123][4][5][6][7]     | \{1, 4, 5, 6, 7\},     | \{\}               | 0                    | (7)                 |
| 3    | 23   | [123][4][5][6][7]     | \{2, 3, 4, 5, 6, 7\},  | \{\}               | 0                    | (7)                 |
| 4    | 14   | [123][14][5][6][7]    | \{1, 5, 6, 7\}         | \{2, 4, 5, 6, 7\},  | 4                    |
| 5    | 15   | [123][14][15][6][7]   | \{1, 6, 7\}            | \{2, 4, 5, 6, 7\},  | 3                    |
| 6    | 16   | [123][14][15][16][7]  | \{1, 7\}               | \{2, 4, 5, 6, 7\},  | 2                    |
| 7    | 17   | [123][14][15][16][17] | \{\}                   | \{2, 4, 5, 6, 7\},  | 1                    |

Iter corresponds to iteration number, Edge corresponds to the added edge in each iteration, currModel corresponds to the present model, tempAMIS corresponds to the AMIS list to be tested for mutual conditional independence, AMIS corresponds to the list of MISs, and # of candidate edges corresponds to the search space in terms of candidate edges to explore in each iteration.
6 Computational details

All the experimental results in this paper were carried out using R 3.4.0, with the packages\texttt{gRim} and \texttt{MASS}. All packages used are available at \url{http://CRAN.R-project.org}. We implemented the new forward selection algorithm in R, the code is available at github: \url{https://github.com/niharikag/gMCI}.

7 Conclusion

The notion of conditional independence and Markov properties are fundamental for graphical models. We have discussed different Markov properties for the class of Markov networks. We have introduced the concept of mutual conditional independence in an independent set of a Markov network. Particularly, the similarity between separation in a graph and conditional independence in probability has been extended to similarity between the mutual separation in a graph and the mutual conditional independence in probability. We have developed a new forward model selection algorithm for log-linear models, where the mutual conditional independence is used to reduce the search space. Specifically, for some sparse graphs the reduction in the search space from $O(n^3)$ to $O(n^2)$ has been shown. However, it will be an interesting problem for future research to explore, for any sparse graph, how much the size of the search space can be reduced using the proposed forward selection algorithm as compared to the classical forward selection algorithm.

We conclude with the remark that in the literature most perspectives for the Markov networks are Markov properties and clique factorization. MCIP enables us to take a new perspective on Markov properties and factorization of the corresponding probability distributions. This new perspective could be useful to understand the structures and problems of the Markov random fields. The MCIP can be a promising new direction for model selection and inference in Markov networks.

Acknowledgements The authors are grateful for the constructive inputs given by anonymous reviewers and the editor, which helped us to improve the manuscript.

Funding Information Open access funding provided by Royal Institute of Technology.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix: Illustration of the proposed forward model selection algorithm

We begin by fitting the complete independence model. The vertices $A$, $B$, $C$, $D$, $E$ and $F$ correspond to the factors “smoke”, “mental”, “phys”, “systol”, “protein” and “family”
respectively. Under assumption of the complete independence model, we get the closed form expression for the estimates of the expected cell counts as follows.

\[ \hat{m}_{ijklm} = \frac{n_i \ldots n_j \ldots n_k \ldots n_l \ldots n_m \ldots n}{n^5} \]

For details on the derivation of the above expression, we refer to the book [3]. After having computed the (estimates of) expected cell counts, the \( G^2 \) statistic for this model is computed using (12), and we get the \( G^2 = 843.957 \) (df : 57, p-value : 0). Since observed value of
### Table 6  Model fitting

| Ad. Edge | Model | d.f. | $G^2$  |
|----------|-------|------|--------|
| AB       | [AB][BC][D][E][F] | 55  | 157.9852 |
| AD       | [AD][BC][E][F]     | 55  | 146.9529 |
| AE       | [AE][BC][D][F]     | 55  | 140.5849 |
| AF       | [AF][BC][D][E]     | 55  | 156.8614 |
| BD       | [B][BC][D][E][F]   | 55  | 157.4851 |
| BE       | [B][BC][BE][D][F]  | 55  | 140.0559 |
| BF       | [B][BC][BF][D][E]  | 55  | 153.2537 |
| DE       | [D][BC][BE][F]     | 55  | 145.1760 |
| DF       | [D][BC][DF][E]     | 55  | 154.9815 |
| EF       | [E][BC][DF][EF]    | 55  | 154.9815 |

### Table 7  Model comparison

| Ad. Edge | d.f. | $G^2$ | p-value |
|----------|------|-------|---------|
| AB       | 1    | 9.6637| 0.0018  |
| AD       | 1    | 11.0324| 0.0008  |
| AE       | 1    | 17.4003| 0.0000  |
| AF       | 1    | 1.0686| 0.3012  |
| BD       | 1    | 0.5000| 0.4794  |
| BE       | 1    | 17.9292| 0.0000  |
| BF       | 1    | 4.7315| 0.02961 |
| DE       | 1    | 12.8092| 0.0003  |
| DF       | 1    | 1.1237| 0.2891  |
| EF       | 1    | 3.0037| 0.0830  |

### Table 8  Model fitting

| Ad. Edge | Model | d.f. | $G^2$  |
|----------|-------|------|--------|
| AC       | [AC][BC][BE][D][F] | 54  | 112.5749 |
| AD       | [AD][BC][BE][F]     | 54  | 129.0236 |
| AE       | [AE][BC][BE][D][F]  | 54  | 122.6556 |
| AF       | [AF][BC][BE][D]     | 54  | 138.9873 |
| CD       | [C][BC][BE][CD][F]  | 54  | 139.9605 |
| CE       | [C][BC][BE][CE][F]  | 54  | 136.1561 |
| CF       | [C][BC][BE][CF][D]  | 54  | 139.8857 |
| DE       | [D][BC][BE][DE][F]  | 54  | 127.2467 |
| DF       | [D][BC][BE][DF]     | 54  | 138.9321 |
| EE       | [E][BC][BE][D][EF]  | 54  | 137.0522 |
Table 9 Model comparison

| Ad. Edge | d.f. | $G^2$   | p-value |
|----------|------|---------|---------|
| AC       | 1    | 27.4810 | 0.0000  |
| AD       | 1    | 11.0323 | 0.0009  |
| AE       | 1    | 17.4003 | 0.0000  |
| AF       | 1    | 1.0687  | 0.3012  |
| CD       | 1    | 0.0954  | 0.7574  |
| CE       | 1    | 3.8998  | 0.0483  |
| CF       | 1    | 0.1702  | 0.6799  |
| DE       | 1    | 12.8092 | 0.0003  |
| DF       | 1    | 1.1238  | 0.2891  |
| EF       | 1    | 3.0037  | 0.0831  |

The test statistic is unexpectedly large, hence we conclude that the data fails to support the complete independence model. The data structures are initialized as follows.

\[
\text{currModel} = [A][B][C][D][E][F]
\]

\[
\text{tempAMIS} = \{ A, B, C, D, E, F \}
\]

\[
\text{AMIS} = \{ \emptyset \}
\]

As mentioned before, at each step we add the most significant edge as long as the significance level is below a cut-off value (we use cut-off of $\alpha = 0.05$). As a first step we compare all the models with a single edge added to the model of complete independence. We use the $G^2$ statistics for comparing models (12). Table 4 gives the model fit and Table 5 summarizes the test results.

The model with edge $(B, C)$ has the largest difference in $G^2$ (or the smallest $p$-value), we choose this models as the current model. Also the set containing the factors $(B, C)$ gets separated as follows.

\[
\text{currModel} = [A][BC][D][E][F]
\]

\[
\text{tempAMIS} = \{ A, B, D, E, F \}, \{ A, C, D, E, F \}
\]

\[
\text{AMIS} = \{ \emptyset \}
\]

Before moving to the next step we make sure that the list tempAMIS is irreducible and contains no duplicate.

As a next step, we first check that if MCIP holds for the members of tempAMIS. The MCI test for the elements of tempAMIS $\{A, B, D, E, F\}$ and $\{A, D, C, E, F\}$ gives the $G^2$

Table 10 Model fitting

| Ad. Edge | Model                        | d.f. | $G^2$   |
|----------|------------------------------|------|---------|
|          | $[AC][BC][BE][D][F]$        | 54   | 112.5749|
| AB       | $[AC][BC][BE][D][F]$        | 53   | 112.5736|
| AD       | $[AD][AC][BC][BE][F]$       | 53   | 101.5426|
| AF       | $[AC][AF][BC][BE][D]$       | 53   | 111.5063|
| BD       | $[AC][BC][BD][BE][F]$       | 53   | 112.0748|
| BF       | $[AC][BC][BE][BF][D]$       | 53   | 107.8434|
| DF       | $[AC][BC][BE][DF]$          | 53   | 111.4511|
Mutual conditional independence and its applications...

Table 11  Model comparison

| Ad. Edge | d.f. | \(G^2\) | p-value |
|----------|------|---------|---------|
| AB       | 1    | 0.0013  | 0.9715  |
| AD       | 1    | 11.0323 | 0.0009  |
| AF       | 1    | 1.0687  | 0.3012  |
| BD       | 1    | 0.5001  | 0.4795  |
| BF       | 1    | 4.7316  | 0.0296  |
| DF       | 1    | 1.1238  | 0.2891  |

Statistics as 113.566 (df: 52, p-value: 0) and 125.16 (df: 52, p-value: 0) respectively. Since observed value of the test statistic is unexpectedly large, hence MCI does not hold for \{A, B, D, E, F\} and \{A, C, D, E, F\}.

We consider the set \{A, B, D, E, F\}. We compare the current model with all the models with an additional edge from the set \{A, B, D, E, F\}. It is described in the Tables 6 and 7.

The term (B, E) is added to the current model. The data structures are updated as follows.

\[
\text{currModel} = [A][BC][BE][D][F] \\
\text{tempAMIS} = \{\{A, B, D, F\}, \{A, D, E, F\}, \{A, C, D, E, F\}\} \\
\text{AMIS} = \emptyset
\]

There is no duplicate, but irreducibility check suggests that \{A, D, E, F\} \subset \{A, C, D, E, F\}, the set \{A, D, E, F\} is redundant and hence it is removed from the list tempAMIS. Accordingly the data structures get updated as follows.

\[
\text{currModel} = [A][BC][BE][D][F] \\
\text{tempAMIS} = \{\{A, B, D, F\}, \{A, C, D, E, F\}\} \\
\text{AMIS} = \emptyset
\]

As a next step, we first check if MCIP holds for the set \{A, B, D, F\}. The MCI test for this set gives the \(G^2\) statistic as 67.5 (df: 44, p-value: 0.013). Since observed value of the test statistic is unexpectedly large, hence the data does not support the assumption of \(A \perp B \perp D \perp F\)|(C, E). Note that we have already verified that MCIP does not hold for \{A, C, D, E, F\} in the previous step.

Now, we consider an additional edge from the set \{A, C, D, E, F\}, Table 8 gives the model fit and Table 9 summarizes the test results.

Table 12  Model fitting

| Ad. Edge | Model | d.f. | \(G^2\) |
|----------|-------|------|---------|
| CD       | [AC][AD][BC][BE][F] | 53   | 101.5426 |
| CE       | [AC][AD][BC][BE][CE][F] | 52   | 97.6428  |
| CF       | [AC][AD][BC][BE][CF] | 52   | 101.3724 |
| DE       | [AC][AD][BC][BE][DE][F] | 52   | 88.8999  |
| EF       | [AC][AD][BC][BE][EF] | 52   | 98.5389  |
Table 13  Model comparison

| Ad. Edge | d.f. | $G^2$  | p-value |
|----------|------|--------|---------|
| CD       | 1    | 0.5206 | 0.4706  |
| CE       | 1    | 3.8998 | 0.0483  |
| CF       | 1    | 0.1702 | 0.6799  |
| DE       | 1    | 12.6427| 0.0004  |
| EF       | 1    | 3.0037 | 0.0831  |

The term $(A, C)$ is added to the current model. Accordingly the modified data structure is given below.

\[
\text{currModel} = [AC][BC][BE][D][F]
\]
\[
\text{tempAMIS} = \{\{A, B, D, F\}, \{A, D, E, F\}, \{C, D, E, F\}\}
\]
\[
\text{AMIS} = \{\emptyset\}
\]

There is no duplicate and tempAMIS can not be reduced further. So, we move on to the next step. The $G^2$ statistics 67.500 (df : 44, p-value : 0.0129) , 66.025 (df : 44, p-value : 0.0174) and 93.066 (df : 44, p-value : 0) of the MCI tests for the sets $\{A, B, D, F\}$, $\{A, D, E, F\}$ and $\{C, D, E, F\}$ respectively indicates that the data does not supports the MCI relations for the sets. We consider adding an additional edge from the set $\{A, B, D, F\}$ to the current model, the details are given in the Tables 10 and 11.

The model with edge $(A, D)$ has the largest difference in $G^2$, so we add the edge $(A, D)$ to the current model. The data structures get updated as follows (It must be noted that since $\{D, E, F\} \subset \{C, D, E, F\}$, the set $\{D, E, F\}$ is redundant and it is removed from tempAMIS).

\[
\text{currModel} = [AC][AD][BC][BE][F]
\]
\[
\text{tempAMIS} = \{\{A, B, F\}, \{B, D, F\}, \{A, E, F\}, \{C, D, E, F\}\}
\]
\[
\text{AMIS} = \{\emptyset\}
\]

We perform the MCI test for the sets $\{A, B, F\}$, $\{B, D, F\}$ and $\{A, E, F\}$. The $G^2$ statistics, 38.915 (df : 32, p-value : 0.1865), 39.271 (df : 32, p-value : 0.1762) and 59.043 (df : 32, p-value : 0.0025) indicate that the MCI holds for the sets $\{A, B, F\}$ and $\{B, D, F\}$, and data fails to support the MCI for the set $\{A, E, F\}$. The sets $\{A, B, F\}$ and $\{B, D, F\}$ are moved from tempAMIS to AMIS.

\[
\text{currModel} = [AC][AD][BC][BE][F]
\]
\[
\text{tempAMIS} = \{\{A, E, F\}, \{C, D, E, F\}\}
\]
\[
\text{AMIS} = \{\{A, B, F\}, \{B, D, F\}\}
\]

Now, we look for the most significant edge in the set $\{C, D, E, F\}$. Test details are given in Tables 12 and 13.

Table 14  Model fitting

| Ad. Edge | Model | d.f. | $G^2$  |
|----------|-------|------|--------|
|          | [AC][AD][BC][BE][DE][F] | 52   | 88.89989 |
| AE       | [AC][AD][AE][BC][BE][DE][F] | 51   | 65.8469  |
We select the model with an additional edge (D, E). The set \{C, D, E, F\} gets divided into the subsets \{C, D, F\} and \{C, E, F\}.

\[
\text{currModel} = [AC][AD][BC][BE][DE][F]
\]
\[
\text{tempAMIS} = \{\{A, E, F\}, \{C, D, F\}, \{C, E, F\}\}
\]
\[
\text{AMIS} = \{\{A, B, F\}, \{B, D, F\}\}
\]

We perform the MCI test for \{C, D, F\} and \{C, E, F\}, the $G^2$ statistic is 35.475 ($df$: 32, $p$-value: 0.3077) and 42.534 ($df$: 32, $p$-value: 0.1009), we conclude that the data supports the MCI relation for the sets. The sets are removed from the tempAMIS and added to the AMIS. We get the intermediate data structures as

\[
\text{currModel} = [AC][AD][BC][BE][DE][F]
\]
\[
\text{tempAMIS} = \{\{A, B, F\}\}
\]
\[
\text{AMIS} = \{\{C, E, F\}, \{C, D, F\}, \{A, B, F\}, \{B, D, F\}\}
\]

We consider the set \{A, E, F\}, there is only one candidate edge (A, E) from the set to be considered, as (A, F) $\subset$ \{A, B, F\} and (E, F) $\subset$ \{C, E, F\}. In the following, from the model comparison test as given in Tables 14 and 15, we find that the test statistic is unexpectedly large, hence the data goes in favour of the larger model with an additional edge [AE].

Finally, the data structures get updated as follows.

\[
\text{currModel} = [AC][AD][AE][BC][BE][DE][F]
\]
\[
\text{tempAMIS} = \{\emptyset\}
\]
\[
\text{AMIS} = \{\{C, E, F\}, \{C, D, F\}, \{A, B, F\}, \{B, D, F\}\}
\]

As the list tempAMIS becomes empty, we stop with the model $[AC][ADE][BC][BE][F]$. The algorithm returns the \{\{A, D, F\}, \{C, E, F\}, \{B, D, F\}\} AMIS. A graph structure can be determined uniquely from the AMIS as given in Fig. 3.

References

1. Agresti, A. Categorical Data Analysis, 3rd edn. Wiley-Interscience, New York (2012)
2. Andersen, A.H.: Multidimensional contingency tables. Scand. J. Statist. 3, 115–127 (1974)
3. Bishop, Y.M.M., Fienberg, S.E., Holland, P.W.: Discrete Multivariate Analysis: Theory and Practice. The MIT Press, Cambridge (1989)
4. Christensen, R. Log-Linear Models and Logistic Regression, 2nd edn. Springer, Berlin (1997)
5. Dahinden, C., Kalisch, M., Buhlmann, P.: Decomposition and model selection for large contingency tables. Biometrical Journal 52, 233–252 (2010)
6. Dawid, A.P.: Conditional independence in statistical theory. J. R. Stat. Soc. 41(1), 1–31 (1979)
7. Gauraha, N.: Graphical log-linear models: fundamental concepts and applications. Journal of Modern Applied Statistical Methods 16(1), 545–577 (2017)
8. Geiger, D., Pearl, J.: Logical and algorithmic properties of conditional independence and graphical models. Ann. Stat. 24(4), 2001–2021 (1993)
9. Goodman, L.A.: The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. Technometrics 13, 31–66 (1971)
10. Jordan, M.I.: Graphical models. Stat. Sci. 19(1), 140–155 (2004)
11. Lauritzen, S.L. Graphical Models, 2nd edn. Oxford University Press Inc., New York (1996)
12. Matus, F.: On equivalence of markov properties over undirected graphs. J. Appl. Probability 29, 745–749 (1992)
13. Matus, F.: On conditional independence and log-convexity. Annales de l’I.H.P. Probabilités et statistiques 44(4), 1137–1147 (2012)
14. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco (1988)
15. Preston, C.: Random Fields. Springer, Berlin (1974)
16. Reinis, Z., Pokorny, J., Basika, V., Tiserova, J., Goricin, K., Horakova, D., Stuchlikova, E., Havranek, T., Hrabovsky, F.: Prognostic significance of the risk profile in the prevention of coronary heart disease. Bratis. Lek. Listy 76, 137–150 (1981)
17. Spitzer, F.: Random fields and interacting particle systems. M.A.A.Summer Seminar Notes (1971)
18. Webb, G.I., Petitjean, F.: A multiple test correction for streams and cascades of statistical hypothesis tests. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1255–1264. ACM (2016)
19. Wermuth, N.: Model search among multiplicative models. Biometrics 32, 253–263 (1976)
20. Whittaker, J. Graphical Models in Applied Multivariate Statistics, 2nd edn. Wiley, Chichester (1990)

Publisher’s note  Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.