International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
© World Scientific Publishing Company

CHARACTERIZATION OF ESSENTIAL GRAPHS BY MEANS OF THE OPERATION OF LEGAL MERGING OF COMPONENTS

MILAN STUDENÝ

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, Prague, 18208, Czech Republic
studeny@utia.cas.cz

Received (received date)
Revised (revised date)

One of the most common ways of representing classes of equivalent Bayesian networks is the use of *essential graphs* which are also known in the literature as completed patterns or completed pdags. The name essential graph was proposed by Andersson, Madigan and Perlman who also gave a graphical characterization of essential graphs. In this paper an alternative characterization of essential graphs is presented. The main observation is that every essential graph is the largest chain graph within a special class of chain graphs. More precisely, every equivalence class of Bayesian networks is contained in an equivalence class of chain graphs without flags (= certain induced subgraphs). A special operation of *legal merging* of (connectivity) components for a chain graph without flags is introduced. This operation leads to an algorithm for finding the essential graph on the basis of any graph in that equivalence class of chain graphs without flags which contains the equivalence class of a Bayesian network. In particular, the algorithm may start with any Bayesian network.

Keywords: chain graph, acyclic directed graph, essential graph, flag, legal merging.

1. Introduction

This is to explain the motivation in brief. Several approaches to learning Bayesian networks are based on the idea of maximization of a score metric. Some of these approaches [6] use the method of travelling in a search space which is often the collection of equivalence classes of Bayesian networks over a fixed set of variables.

For this purpose one needs a suitable representative of an equivalence class of this type to be kept in memory of a computer. A quite popular representative of an equivalence class of Bayesian networks is its *essential graph*. This term was introduced by Andersson, Madigan and Perlman [1] but some alternative names for this graph have appeared in the literature: completed pattern [15], maximally oriented graph for a pattern [9] and completed pdag [6]. A graphical characterization of essential graphs was presented in [1] as Theorem 4.1.

The author's plan has been to develop an algebraic method of description of the equivalence classes of Bayesian networks by means of certain integral vectors, that is, vectors whose components are integers. The hope is that this approach could be utilized to make existing local search methods of learning Bayesian networks more effective - for a description of the main ideas of this algebraic approach see Chapter 8 in [13]. In connection with the effort to establish the relationship between essential graph representation and integral vector representation, an alternative characterization of essential graphs was found as a byproduct. The characterization is based on a special operation of *legal merging* of components and leads to an algorithm for converting a Bayesian network into the respective essential graph presented in this paper. Unlike former algorithms presented in [9] and [5], this alternative algorithm avoids indicating essential (= compelled) arrows. Note that similar characterization of essential graphs has been found independently in a recent paper [11] by Rovenato who has had another motivation source.

2. Basic concepts

2.1. *Graphical concepts*

Graphs considered in this paper will have a finite non-empty set N as the set of *nodes*, which will be represented by small circles in pictures, and two possible types of edges. An undirected edge or a *line* over N is a subset of N of cardinality two, that is, an unordered pair $\{a, b\}$ where $a, b \in N$, $a \neq b$. The respective notation $a - b$ corresponds to its pictorial representation by means of an undirected link between circles which represent the nodes a and b. A directed edge or an *arrow* over N is an ordered pair (a, b) where $a, b \in N$, $a \neq b$. The notation $a \to b$ reflects its pictorial representation by a directed link from a circle representing the node a to a circle representing the node b.

A *hybrid graph* (over N) is a graph without multiple edges, that is, a triplet $H = (N, \mathcal{L}(H), \mathcal{A}(H))$ where N is a set of nodes, $\mathcal{L}(H)$ a set of lines over N and $\mathcal{A}(H)$ a set of arrows over N such that whenever $(a, b) \in \mathcal{A}(H)$ then $(b, a) \notin \mathcal{A}(H)$ and $\{a, b\} = \{b, a\} \notin \mathcal{L}(H)$. A pair $[a, b]$ of distinct elements of N will be called an *edge in* H (between a and b) if one of the following options occurs: $a - b$ in H, $a \to b$ in H and $b \to a$ in H.

An *undirected graph* is a hybrid graph having lines only, that is, $\mathcal{A}(H) = \emptyset$. A *directed graph* is a hybrid graph having arrows only, that is, $\mathcal{L}(H) = \emptyset$. Given $\emptyset \neq A \subseteq N$ the *induced subgraph* H_A of H is the triplet $(A, \mathcal{L}(H) \cap \mathcal{P}(A), \mathcal{A}(H) \cap (A \times A))$ where $\mathcal{P}(A)$ denotes the power set of A (= the collection of subsets of A).

A set $K \subseteq N$ is *complete* in a chain graph H over N if $\forall\, x, y \in N$ $x \neq y$ one has $x - y$ in H. A set $C \subseteq N$ is *connected* in H if for every $a, b \in N$ there exists an undirected path connecting them, that is, a sequence of distinct nodes $a = c_1, \ldots, c_n = b$, $n \geq 1$ such that $c_i - c_{i+1}$ in H for $i = 1, \ldots, n-1$. Connectivity *components* of H are maximal connected sets in H (with respect to set inclusion). Of course, components of a directed graph are singletons (= sets of cardinality one).

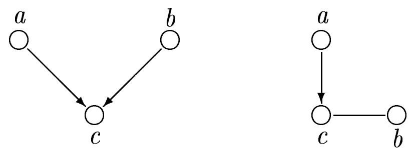

Fig. 1. An immorality and a flag.

An *acyclic directed graph* (ADG) is a directed graph without directed cycles, that is, without any sequence $d_1, \ldots, d_n, d_{n+1} = d_1$, $n \geq 3$ such that $d_1, \ldots, d_n$ are distinct and $d_i \to d_{i+1}$ in H for $i = 1, \ldots, n$. The well-known fact (whose proof is omitted) is that a directed graph is acyclic iff there exists a total ordering of all nodes of N (= ordering into a sequence) $a_1, \ldots, a_m$, $m \geq 1$ which is *consistent* with the direction of arrows, that is, whenever $a_i \to a_j$ in H then $i < j$.

A *chain graph* is a hybrid graph H for which there exists a chain, that is, an ordered partitioning of N into non-empty blocks $B_1, \ldots, B_m$, $m \geq 1$ such that

- if $a — b$ in H then $a, b \in B_i$ for some i,
- if $a \to b$ in H then $a \in B_i, b \in B_j$ with $i < j$.

Of course, both undirected graph and acyclic directed graph are special cases of a chain graph. Note that a chain for a given chain graph is not determined uniquely. An equivalent definition of a chain graph is that it is a hybrid graph H without semi-directed cycles, that is, without any sequence $d_1, \ldots, d_n, d_{n+1} = d_1$, $n \geq 3$ such that $d_1, \ldots, d_n$ are distinct, $d_1 \to d_2$ in H and $\forall\, i = 2, \ldots, n$ either $d_i \to d_{i+1}$ or $d_i — d_{i+1}$ in H - see Lemma 2.1 in [12]. In particular, in a chain graph H over N, there is no arrow between nodes of a connected set $C \subseteq N$; in other words, the induced subgraph H_C is undirected. In particular, the set of *parents* of C, that is,

$$pa_H(C) = \{a \in N\,;\, \exists\, b \in C \;\; a \to b \text{ in } H\}$$

is disjoint with C. The set

$$ne_H(C) = \{a \in N \setminus C\,;\, \exists\, b \in C \;\; a — b \text{ in } H\}$$

will be named the set of *neighbors* of C. The notation $pa_H(x)$, respectively $ne_H(x)$, for $x \in N$ will sometimes substitute $pa_H(\{x\})$, respectively $ne_H(\{x\})$. A *descending path* in a chain graph H is a sequence $d_1, \ldots, d_n$, $n \geq 1$ of distinct nodes such that $\forall\, i = 1, \ldots, n-1$ either $d_i \to d_{i+1}$ in H or $d_i — d_{i+1}$ in H.

2.2. *Bayesian networks*

A *Bayesian network* is a certain (usually discrete) statistical model appended to an ADG. It can be introduced as the class of multidimensional probability distributions (on a fixed sample space) which factorize according to the ADG in a certain way. Note that some authors prefer to define Bayesian network as one distribution

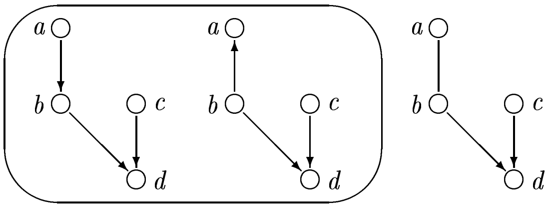

Fig. 2. An equivalence class of Bayesian networks and the respective essential graph.

belonging to this class. An alternative definition of that class can be formulated terms of conditional independence, using the d-separation criterion from [10]. Since exact definitions of related concepts are not needed in this paper, they are omitted. Thus, throughout the rest of this paper, the phrase "Bayesian network" will be understood as a synonym for an ADG.

However, the concept that is needed is the concept of *equivalence of Bayesian networks*. A statistician would expect the following definition: two Bayesian networks are equivalent if they represent the same statistical model. In standard situations this requirement is equivalent to the condition that the d-separation criterion induces the same collections of conditional independence statements for both ADGs. Fortunately, Verma and Pearl [15] gave a direct graphical characterization of equivalent Bayesian networks which will serve as its definition in this paper.

The *underlying graph* of a hybrid graph H over N is an undirected graph H^u over N such that a — b in H^u iff $[a, b]$ is an edge in H. An *immorality* in H is an induced subgraph of H shown in the left-hand picture of Figure 1, that is, the configuration $a \to c \leftarrow b$ where a, b, c are distinct nodes and the pair $[a, b]$ is not an edge in H. Two Bayesian networks G_1, G_2 over N are called *equivalent* iff they have the same underlying graph and the same collection of immoralities.

2.3. *Equivalence class representation*

An equivalence class $\mathcal{G}$ of Bayesian networks (over N) can be described by its *essential graph* which is a hybrid graph G^* (over N) such that

- $a \to b$ in G^* if and only if $a \to b$ in G for every $G \in \mathcal{G}$,
- a — b in G^* if and only if there exist $G_1, G_2 \in \mathcal{G}$ such that $a \to b$ in G_1 and $b \to a$ in G_2.

Example 1. Let us consider the equivalence class $\mathcal{G}$ of Bayesian networks over $N = \{a, b, c, d\}$ shown on the left-hand side of Figure 2. The respective essential graph G^* is in the right-hand picture of the same figure.

A graphical characterization of essential graphs was presented in [1]. Let us recall the definitions of relevant concepts. An undirected graph H is *chordal* (alternative names are *triangulated* and *decomposable*) if, for every undirected cycle in H which

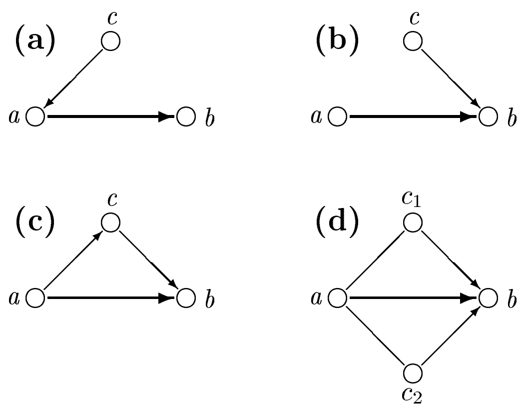

Fig. 3. A strongly protected arrow $a \to b$.

has the length at least four, that is, for any sequence $c_1, \ldots, c_n, c_{n+1} = c_1$, $n \geq 4$ where $c_1, \ldots, c_n$ are distinct and $c_i — c_{i+1}$ in H for $i = 1, \ldots, n$, there exists a *chord* in H, that is, an edge $c_i — c_j$ where $1 \leq i, j \leq n$ and $1 < j - i < n - 1$. A *flag* in a hybrid graph H is an induced subgraph of H shown in the right-hand picture of Figure 1, that is, the configuration $a \to c — b$ where a, b, c are distinct nodes and the pair $[a, b]$ is not an edge in H. An arrow $a \to b$ in a hybrid graph H is called *strongly protected* if it belongs to at least one induced subgraph of H of the types **(a)**-**(d)** shown in Figure 3.

Remark 1. Note that there are other possible ways of representing equivalence classes of Bayesian networks. One of them uses the concept of the largest chain graph which is based on results of Frydenberg [7]. Let us call by a (minimal) *complex* in a hybrid graph H a special induced subgraph of H, namely, the configuration $a \to c_1 — \ldots — c_k \leftarrow b$, $k \geq 1$ where $c_i — c_{i+1}$ for $i = 1, \ldots, k-1$ and there is no other edge in H between distinct nodes $a, b, c_1, \ldots, c_k$. Of course, an immorality is a special case of a complex. Frydenberg [7] gave a characterization of equivalent chain graphs which can be viewed as a generalization of the result by Verma and Pearl [15]: two chain graphs are equivalent iff they have the same underlying graph and the same (minimal) complexes. Moreover, Frydenberg [7] also showed that every equivalence class $\mathcal{H}$ of chain graphs over N has a distinguished representative H_∞ which can be obtained as follows:

- $a \to b$ in H_∞ iff $a \to b$ in H for every $H \in \mathcal{H}$,
- $a — b$ in H_∞ iff $a — b$ in H for at least one $H \in \mathcal{H}$.

He named this graph, which is the graph with the maximum number of lines in $\mathcal{H}$, the *largest chain graph* (LCG) of $\mathcal{H}$. This naturally leads to an idea mentioned in [12]: every equivalence class $\mathcal{G}$ of Bayesian networks over N is a subclass of the

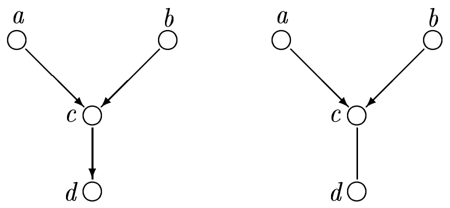

Fig. 4. An essential graph and the respective largest chain graph.

uniquely determined (wider) equivalence class $\mathcal{G}^{\times}$ of chain graphs over N, and one can represent $\mathcal{G}$ by means of the LCG of $\mathcal{G}^{\times}$. In general, the LCG of $\mathcal{G}^{\times}$ and the essential graph of $\mathcal{G}$ may differ - the simplest example is in Figure 4.

Another alternative is an algebraic approach described in Chapter 8 of [13]. The main idea of that approach is to represent every equivalence class of Bayesian networks by a special integral (= integer-valued) vector, called *standard imset*, whose components correspond to subsets of N. This approach can be viewed as a special case of a more general method used to describe models of probabilistic conditional independence structure which is described in details in previous chapters of [13].

Lemma 1. *A hybrid graph H over N is an essential graph of an equivalence class of Bayesian networks over N iff the following four conditions are fulfilled:*

(i) *H is a chain graph;*
(ii) *for every connectivity component C of H the induced subgraph H_C is chordal;*
(iii) *H has no flags;*
(iv) *every arrow in H is strongly protected.*

Proof. See Theorem 4.1 in [1]. □

Consequence 1. *Let H be the essential graph of an equivalence class of Bayesian networks over N and $\emptyset \neq A \subseteq N$ is a set closed under parents in H, that is, $x \in pa_H(y)$ and $y \in A$ implies $x \in A$. Then H_A is the essential graph of an equivalence class of Bayesian networks over A.*

Proof. It suffices to show that the conditions (i)-(iv) from Lemma 1 for H imply the same conditions for H_A. Because the induced subgraph of a chain graph is a chain graph and the induced subgraph of a chordal (undirected) graph is a chordal graph this is easy for (i)-(iii). To verify (iv) the assumed condition on A is needed. Supposing $a \to b$ is an edge in H_A one has $a \to b$ in H and by (iv) it belongs to an induced subgraph of H of one of the types **(a)**-**(d)** in Figure 3. The fact $a, b \in A$ and the form of these induced subgraphs implies $c \in A$, respectively $c_1, c_2 \in A$, which means that H_A has the respective induced subgraph as well. □

3. Chain graphs without flags

The goal of this paper is to characterize every essential graph as a distinguished member of a certain class of chain graphs. Actually, the concept of essential graph can be viewed as an analogy of the concept of LCG. The only difference is that the collection of chain graphs is replaced by the collection of chain graphs without flags. These chain graphs are slightly special in certain respects. For example, the parent set of a node is an invariant of a connectivity component.

Observation 1. Let H be a chain graph without flags and C a connectivity component of H. Then $pa_H(x) = pa_H(y)$ for every $x, y \in C$; in particular, $pa_H(x) = pa_H(C)$.

Proof. If $z \to x - y$ in H then $z \to y$ in H because of the absence of flags and semi-directed cycles in H. Thus, $x - y$ in H implies $pa_H(x) \subseteq pa_H(y)$ and hence $pa_H(x) = pa_H(y)$ by symmetry. If $x, y \in C$ the use an inductive argument. □

Equivalence of chain graphs without flags is characterized in a simple way.

Observation 2. Two chain graphs without flags are equivalent iff they have the same underlying graph and immoralities.

Proof. Use Frydenberg's characterization [7] of equivalent chain graphs mentioned in Remark 1. Observe that the only (minimal) complex in a chain graph without flags is an immorality. □

Consequence 2. *Let G be a Bayesian network over N and $\mathcal{G}$ denotes the equivalence class of Bayesian networks containing G. If G^* the essential graph of $\mathcal{G}$ then G and G^* are equivalent (chain graphs).*

Proof. Due to the construction of G^*, the hybrid graphs G and G^* have the same underlying graph and immoralities. As a result of Lemma 1, G^* is a chain graph without flags and the same claim is evident for G. Then use Observation 2. □

Lemma 2. *A chain graph H without flags is equivalent to a Bayesian network iff, for every connectivity component C of H, the induced subgraph H_C is chordal.*

Proof. Use Proposition 4.2 and Remark 4.2 in [2] where chain graphs equivalent to ADGs are characterized. A more explicit characterization in Remark 4.2 there contains three conditions: (i), (ii) and (iii). The reader can consult [2] and verify by means of Observation 1 that the conditions (ii) and (iii) are always fulfilled for chain graphs without flags. The last condition (i) requires exactly what is mentioned in Lemma 2. □

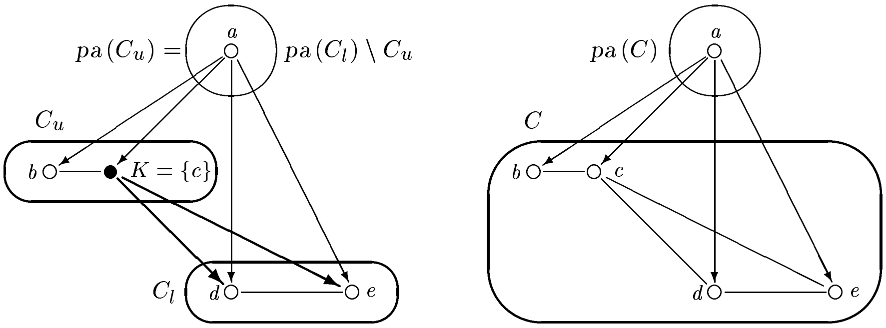

Fig. 5. The situation before and after legal merging.

4. Legal merging of components

To distinguish the components which are to be merged, one of them will be named the *upper* component and denoted by C_u while the other one will be named the *lower* component and denoted by C_l.

Definition 1. Let H be a chain graph without flags and C_u, C_l two components of H such that

(i) $K \equiv pa_H(C_l) \cap C_u$ is a non-empty complete set in H,
(ii) $pa_H(C_u) = pa_H(C_l) \setminus C_u$.

If all arrows from (nodes in) C_u to (nodes in) C_l are replaced by lines then the resulting hybrid graph H' is said to arise from H by *legal merging* of the (upper) component C_u and the (lower) component C_l.

Remark 2. The assumptions that the set K from (i) is non-empty and H is a chain graph imply that C_u and C_l are distinct. Moreover, one can derive by Observation 1 that $a \to b$ in H for $a \in C_u$, $b \in C_l$ iff $a \in K$ (and $b \in C_l$). By the condition (ii) from Definition 1 $pa_H(C_u) \subseteq pa_H(C_l)$ and this observation allows one to show that the only type of an edge between C_u and C_l is an arrow from K to C_l. Indeed, by contradiction: if $a - b$ for $a \in C_u$, $b \in C_l$ then $C_u = C_l$ and if $a \leftarrow b$ for $a \in C_u$, $b \in C_l$ then $b \in pa_H(C_u) \setminus pa_H(C_l)$ because of $C_l \cap pa_H(C_l) = \emptyset$. The situation described in Definition 1 is illustrated by Figure 5.

The next result justifies the concept of legal (component) merging.

Statement 1. *Assuming the situation from Definition 1, the constructed graph H' is a chain graph without flags which is equivalent to the original graph H and has $C \equiv C_u \cup C_l$ as a component.*

Proof. I. The graph H' is a chain graph.

Suppose for the sake of contradiction that H' contains a semi-directed cycle $d_1, \ldots, d_n, d_{n+1} = d_1$, $n \geq 3$ with $d_1 \to d_2$. Then it necessarily contains at least one line between C_u and C_l. Put $j = \min\{i\,;\, 2 \leq i \leq n+1,\ d_i \in C_u \cup C_l\}$ and $k = \max\{i\,;\, 2 \leq i \leq n+1,\ d_i \in C_u \cup C_l\}$. Clearly, $j \neq k$ and the section $d_k, \ldots, d_{n+1} = d_1, d_2, \ldots, d_j$ is a descending path which does not contain an edge between C_u and C_l (recall that $d_1 \to d_2$ in H'). Thus, it is a path in a chain graph H, which implies $d_k \in C_u$ and $d_j \in C_l$ (other options are excluded). Observe that $d_{j-1} \to d_j$ in H (because if $j > 2$ and $d_{j-1} - d_j$ then $d_{j-1} \in C_l$, which contradicts the definition of j). Similarly observe $d_{j-1} \notin C_u$ (if $j = 2$ then $d_{j-1} - d_j$ in H' contradicts the fact $d_1 \to d_2$). Thus, $d_{j-1} \in pa_H(C_l) \setminus C_u$ and, by the assumption (ii) from Definition 1 and Observation 1, $d_{j-1} \in pa_H(C_u) = pa_H(d_k)$. Therefore, $d_{j-1}, d_k, \ldots, d_{n+1} = d_1, \ldots, d_{j-1}$ is a semi-directed cycle in H which contradicts the assumption.

II. The graph H' has no flags.

Suppose for the sake of contradiction that H' has an induced subgraph $a \to c - b$. Clearly, $a \to c$ in H and the option $c - b$ in H is excluded. If $c \to b$ in H then $c \in C_u$, $b \in C_l$ and, by the assumption (ii) and Observation 1, $a \in pa_H(C_u) \subseteq pa_H(C_l) = pa_H(b)$ which contradicts the assumption that $[a, b]$ is not an edge. If $b \to c$ in H then $b \in C_u$, $c \in C_l$ and $b \in pa_H(C_l) \cap C_u$, $a \in pa_H(C_l)$. This allows one to derive a contradictory conclusion that $[a, b]$ is an edge: in case $a \in C_u$ one has $a, b \in pa_H(C_l) \cap C_u$ and can use the assumption (i), while in case $a \notin C_u$ the assumption (ii) and Observation 1 can be used to write $a \in pa_H(C_l) \setminus C_u \subseteq pa_H(C_u) = pa_H(b)$.

III. The graphs H and H' are equivalent.

Use Observation 2. To this end, suppose for the sake of contradiction that there is an immorality $a \to c \leftarrow b$ in H which is cancelled in H'. The situation where a flag is created in H' has already been excluded in Step II. Therefore, necessarily $a - c - b$ in H' and hence $c \in C_l$, $a, b \in C_u$. By the assumption (i), a contradictory conclusion that $[a, b]$ is an edge in H can be derived.

IV. The set $C \equiv C_u \cup C_l$ is a component of H'.

The fact that C is connected is evident. Its maximality can be verified by contradiction: if $a \in C$, $a - b$ in H' but $b \notin C$ then by the construction of H' we get $a - b$ in H and the fact that both C_u and C_l are components of H implies $b \in C_u \cup C_l = C$. □

Consequence 3. *The operation of legal merging of components of a chain graph without flags preserves the validity of the condition that induced subgraphs for components are chordal.*

Proof. Use Lemma 2 and Statement 1. □

5. Characterization of essential graphs

Lemma 3. *Let H be a graph without flags. If no pair of components C_u and C_l fulfills the conditions from Definition 1, then each arrow in H is strongly protected.*

Proof. Let $a \to b$ be an arrow in H, denote by C_u the component of H containing a and by C_l the component of H containing b. The aim is to show that $a \to b$ belongs to one of the configurations **(a)**-**(d)** shown in Figure 3. Since the conditions (i)-(ii) from Definition 1 are not met, one of the following three cases occurs:

(I) $pa_H(C_u) \setminus pa_H(C_l) \neq \emptyset$,
(II) $pa_H(C_l) \setminus [C_u \cup pa_H(C_u)] \neq \emptyset$,
(III) $pa_H(C_l) \cap C_u$ is not complete.

If (I) occurs then by Observation 1 find $c \in pa_H(C_u) \setminus pa_H(C_l) = pa_H(a) \setminus pa_H(b)$ and observe that $a \to b$ belongs to the configuration **(a)** from Figure 3.

If (II) occurs then take $c \in pa_H(C_l) \setminus [C_u \cup pa_H(C_u)] = pa_H(b) \setminus [C_u \cup pa_H(C_u)]$. Thus, the cases $a \leftarrow c$ in H and $a — c$ in H are excluded. If $a \to c$ in H then $a \to b$ belongs to the configuration **(c)** otherwise it belongs to the configuration **(b)**.

If (III) occurs then there exist distinct $c_1, c_2 \in pa_H(C_l) \cap C_u$ such that the pair $[c_1, c_2]$ is not an edge. By Observation 1 conclude that $c_1 \to b \leftarrow c_2$ is an immorality in H. If either $a \in \{c_1, c_2\}$ or $[a, c_i]$ is not an edge for at least one i then $a \to b$ belongs to the configuration **(b)**. If $a \notin \{c_1, c_2\}$ and $[a, c_i]$ is an edge for $i = 1, 2$ then the fact $c_i \in C_u$ implies $a — c_i$ and $a \to b$ belongs to the configuration **(d)**. □

The preceding lemma leads to the following characterization of essential graphs.

Statement 2. *Let H be a chain graph without flags which is equivalent to a Bayesian network. Then H is an essential graph iff no pair of its components can be legally merged.*

Proof. I. The sufficiency of the condition follows from preceding lemmas.

Use Lemma 1. The conditions (i) and (iii) are assumed, (ii) follows from Lemma 2 and (iv) from Lemma 3.

II. The necessity of the condition can be verified by contradiction with the help of its sufficiency.

Indeed, otherwise put $H_1 = H$ and start a process of creating a sequence $H_1, H_2, \ldots$ of equivalent graphs: if the operation of legal component merging can be applied to H_i, then perform this operation and construct a graph H_{i+1}. By Statement 1, H_{i+1} is again a chain graph without flags which is equivalent to H_i (and therefore to H) but has more lines. Since the number of edges is finite, the process has to terminate after a finite number of steps and the resulting graph H_n, $n \geq 1$ must be an essential graph (by Step I). The fact $H_1 \neq H_n$ is in contradiction with the uniqueness of the essential graph. □

Definition 2. Let H_1 and H_2 be two equivalent chain graphs. One says that H_1 is *larger* than H_2 if $a — b$ in H_2 implies $a — b$ in H_1 and $a \to b$ in H_2 implies that $a \to b$ in H_1 or $a — b$ in H_1.

Of course, the operation "being larger" is a partial ordering. For example, the graph on the right-hand side of Figure 4 is larger than the graph on the left-hand side.

Consequence 4. *Let G be a Bayesian network over N, $\mathcal{G}$ the equivalence class of Bayesian networks containing G and $\mathcal{H}(G)$ the collection of chain graphs without flags which are equivalent to G (evidently $\mathcal{G} \subseteq \mathcal{H}(G)$). Then a hybrid graph over N is the essential graph of $\mathcal{G}$ iff it is the largest graph in $\mathcal{H}(G)$.*

Proof. Let G^* denote the essential graph of $\mathcal{G}$. By Lemma 1 and Consequence 2, $G^* \in \mathcal{H}(G)$. Let us show that G^* is larger than any other graph $H \in \mathcal{H}(G)$. Put $H_1 = H$ and start the process of creating a sequence of graphs $H_1, H_2, \ldots$ in $\mathcal{H}(G)$ described in Step II of the proof of Statement 2. The process has to terminate after a finite number of steps, resulting in a graph $H_n \in \mathcal{H}(G)$. The only difference from the situation in the proof of Statement 2 is that now the case $n = 1$ can occur. By Statement 2, $H_n = G^*$, which implies that G^* is larger than H. Thus, G^* is the unique "largest" graph in $\mathcal{H}(G)$.

To show the converse, consider the largest chain graph H in $\mathcal{H}(G)$ and observe (by contradiction with the aid of Statement 1) that its components cannot be legally merged. Then use Statement 2. □

6. Algorithms

The above results lead to the following procedure which is 'blind' in the sense that one has to look for components which can be merged. For this reason the procedure is not fully deterministic.

BLIND ALGORITHM

Let H be any chain graph without flags which is equivalent to a Bayesian network G. If the operation of legal merging of some components C_u and C_l of H can be done, then perform it and replace H by the resulting graph. The algorithm stops when legal merging of components is not possible.

Of course, one can take a Bayesian network G in place of H. It follows from Statements 1 and 2 (and Consequence 4) that the algorithm stops with the essential graph of the equivalence class of G.

However, in the case of a Bayesian network even a more specific version of this algorithm can be designed. In this procedure, the tentative step of looking for components to be merged is removed and the procedure is 'guided' by an ordering of nodes which is consistent with the direction of arrows. Once the ordering is fixed, the algorithm is fully deterministic. The procedure is based on the following concept, which is formally introduced for a general hybrid graph although it has reasonable sense for a chain graph without flags.

Definition 3. Let H be a hybrid graph over N and $a \in N$. By a *parent component* for $a \in N$ will be meant any connectivity component P of H such that

(i) $pa_H(a) \cap P$ is non-empty and complete set in H,
(ii) $pa_H(P) = pa_H(a) \setminus P$.

Observation 3. Let H be a chain graph without flags over N and $a \in N$. Then a parent component for a is uniquely determined if it exists.

Proof. Suppose for the sake of contradiction that two distinct components P_1 and P_2 of H satisfy the conditions (i) and (ii) from Definition 3. Owing to (i) for P_i there exists $d_i \in pa_H(a) \cap P_i$ for $i = 1, 2$. Since $d_1 \notin P_2$, respectively $d_2 \notin P_1$, by (ii) derive $d_1 \in pa_H(a) \setminus P_2 = pa_H(P_2)$, respectively $d_2 \in pa_H(a) \setminus P_1 = pa_H(P_1)$. By Observation 1 $d_1 \to d_2$, respectively $d_2 \to d_1$, which is impossible in a hybrid graph. $\square$

GUIDED ALGORITHM

Let G be a Bayesian network over N and $a_1, \ldots, a_m$, $m \geq 1$ a total ordering of its nodes which is consistent with the direction of arrows in G. Let us construct a sequence of hybrid graphs $G^1, \ldots, G^m$ over N as follows:

- *put $G^1 = G$,*
- *for $k = 2, \ldots, m$ check whether there exists a parent component P for a_k in G^{k-1}. If it exists then define G^k as the hybrid graph obtained from G^{k-1} by replacing arrows from $pa_{G^{k-1}}(a_k) \cap P$ to a_k by lines; otherwise put $G^k = G^{k-1}$.*

The algorithm stops with G^m.

The following lemma established the consistency of the steps of the algorithm.

Lemma 4. *Let G be a Bayesian network over N. Then the guided algorithm defines a sequence of chain graphs without flags which are equivalent to G and a potential change of a graph is realized by means of the operation of legal component merging. More specifically, if $2 \leq k \leq m$ then $\{a_k\} \equiv C_l$ is a component of G^{k-1} and if a parent component P for a_k exists in G^{k-1} then it is uniquely determined and $P \equiv C_u \subseteq \{a_1, \ldots, a_{k-1}\}$. In particular, if the ordering $a_1, \ldots a_m$ is fixed then the algorithm if fully deterministic, that is, $G^1, \ldots, G^m$ is uniquely determined.*

Proof. The idea is to show by induction on $k = 2, \ldots, m+1$ that the following four statements are true:

(a) G^{k-1} is a chain graph without flags which is equivalent to G,
(b) if $a_i \to a_j$ in G^{k-1} then $i < j$,
(c) $\forall\, l \geq k \quad ne_{G^{k-1}}(a_l) = \emptyset$,
(d) if a parent component P for a_k in G^{k-1} exists then it is uniquely determined and $P \subseteq \{a_1, \ldots, a_{k-1}\}$.

The condition (c) implies that $\{a_k\}$ is a component of G^{k-1}. Therefore, if a parent component P for a_k in G^{k-1} exists then the assumptions for legal merging of $C_u = P$ and $C_l = \{a_k\}$ in G^{k-1} from Definition 1 are fulfilled and G^k arises from G^{k-1} by this operation (compare Definitions 3 and 1 and recall what is mentioned in Remark 2). The above observation implies the claims in Lemma 4 then.

Since $G^1 = G$ the conditions (a)-(c) for $k = 2$ follow easily from the assumptions. As concerns (d), for any fixed $2 \leq k \leq m$, the uniqueness of P follows from (a) by Observation 3 while the fact $P \subseteq \{a_1, \ldots, a_{k-1}\}$ is implied by (b) and (c).

Indeed, the condition (i) from Definition 3 implies that there exists $a_i \in P$ such that $a_i \to a_k$ in G^{k-1}. Suppose $a_l \in P$. If $a_l = a_i$ then $l < k$ by (b). If $a_l \neq a_i$ then the fact that P is a connectivity component of G^{k-1} implies that there exists an undirected path in G^{k-1} between a_l and a_i. In particular, $ne_{G^{k-1}}(a_l) \neq \emptyset$ which implies $l < k$ by (c).

If the conditions (a)-(d) hold for some $2 \leq k \leq m$ then the induction step, that is, the proof that (a)-(d) hold for $k+1$, can be done as follows. The condition (a), that is, the claim that G^k is a chain graph without flags which is equivalent to G, follows from the induction hypothesis and Statement 1. As concerns (b), the assumption $a_i \to a_j$ in G^k implies $a_i \to a_j$ in G^{k-1} by the definition G^k and this fact implies $i < j$ by the induction hypothesis. The condition (c), that is, $\forall\, l > k$ $ne_{G^k}(a_l) = \emptyset$ also follows from the induction hypothesis.

Indeed, by the construction of G^k the only difference between G^k and G^{k-1} can occur for edges between P and a_k. As $P \subseteq \{a_1, \ldots, a_{k-1}\}$ by (d) for k one has $ne_{G^k}(a_l) = ne_{G^{k-1}}(a_l)$ for $l > k$. Thus, $ne_{G^k}(a_l) = \emptyset$ follows from (c) for k.

The observation that if $k < m$ then the condition (d) for $k+1$ is implied by the conditions (b) and (c) for $k+1$ has already been made above. The condition (d) for $m+1$ is an empty claim. □

Statement 3. *Let G be a Bayesian network over N and $a_1, \ldots, a_m$, $m \geq 1$ a total ordering of its nodes which is consistent with the direction of arrows in G. The graph G^m obtained by the guided algorithm is the essential graph of the equivalence class of G.*

Proof. Let H denote the essential graph of the equivalence class of G. The first observation is as follows.

I. For every $1 \leq k \leq m$, $H_{\{a_1,\ldots,a_k\}}$ is an essential graph over $\{a_1, \ldots, a_k\}$.

Indeed, observe first that $\{a_1, \ldots, a_k\}$ is closed under parents in H: $a_i \to a_j$ in H, $j \leq k$ implies $a_i \to a_j$ in G by the definition of essential graph. Since $a_1, \ldots, a_m$ is consistent with the direction of arrows in G one has $i < j \leq k$. By Consequence 1, $H_{\{a_1,\ldots,a_k\}}$ is the essential graph of an equivalence class of ADGs over $\{a_1, \ldots, a_k\}$.

The basic idea of the proof of Statement 3 is to verify the following two facts by induction on $k = 1, \ldots, m$

[**A**] the edges in G^k between nodes of $\{a_1, \ldots, a_k\}$ share type and direction (of potential arrows) with corresponding edges in H, that is, $G^k_{\{a_1,\ldots,a_k\}} = H_{\{a_1,\ldots,a_k\}}$,

[**B**] the edges in G^k which hit $\{a_{k+1}, \ldots, a_m\}$, that is, those edges which are <u>not</u> edges between nodes of $\{a_1, \ldots, a_k\}$, share type and direction (of potential arrows) with corresponding edges in G.

As $G^1 = G$ the conditions [A] and [B] are evident for $k = 1$. To verify the induction step suppose $k \geq 2$ and observe the following series of facts II - VII.

II. All edges in G^{k-1} between $\{a_1, \ldots, a_{k-1}\}$ and a_k are arrows towards a_k.

Indeed, this follows from the induction hypothesis [B] for $k-1$ and the assumption that G is a directed graph and $a_1, \ldots, a_m$ is consistent with the direction of arrows in G.

III. Every edge in H between $\{a_1, \ldots, a_{k-1}\}$ and a_k is either a line or an arrow towards a_k.

Indeed, suppose for contradiction that $a_l \leftarrow a_k$ in H for some $l < k$. Then $a_l \leftarrow a_k$ in G by the definition of essential graph which contradicts the assumption that $a_1, \ldots, a_m$ is consistent with the direction of arrows in G.

IV. If $a_l - a_k$ in H for some $l < k$ then the component P in G^{k-1} containing a_l is a parent component for a_k.

It suffices to verify the conditions (i) and (ii) from Definition 3. Note that by Consequence 2, Observation 2 and the construction of $G^1, \ldots, G^m$ the graph G, H and G^{k-1} have the same underlying graph.

For (i) observe, by Step II, that $a_l \rightarrow a_k$ in G^{k-1} which implies that the set $pa_{G^{k-1}}(a_k) \cap P$ contains a_l and is, therefore, non-empty. The fact that the set is complete can be verified by contradiction. Indeed, otherwise distinct $x, y \in P$ exist such that $x \rightarrow a_k \leftarrow y$ is an immorality in G^{k-1}. As G^{k-1} and H are equivalent (by Lemma 4) it is an immorality in H by Observation 2. Because $a_l - a_k$ in H, the fact that H has no flags and semi-directed cycles allows one to derive that $x \rightarrow a_l \leftarrow y$ is an immorality in H. Hence, it is an immorality in G^{k-1} which implies a contradictory conclusion that a_l and $\{x, y\}$ do not belong to the same component P of G^{k-1}.

The first step to verify (ii) is to show that $pa_{G^{k-1}}(P) \subseteq pa_{G^{k-1}}(a_k) \setminus P$. Suppose $a_i \in pa_{G^{k-1}}(P)$. Because G^{k-1} is a chain graph without flags (Lemma 4) by Observation 1 $a_i \rightarrow a_l$ in G^{k-1}. By contradiction observe that $i < k$: otherwise $i \geq k$ implies $a_i \rightarrow a_l$ in G by the induction hypothesis [B] for $k-1$, which gives a contradictory conclusion $i < l < k$. Thus, $a_i, a_l \in \{a_1, \ldots a_{k-1}\}$ and by the induction hypothesis [A] for $k-1$ observe $a_i \rightarrow a_l$ in H. As H has no flags and semi-directed cycles $a_i \rightarrow a_l - a_k$ in H implies $a_i \rightarrow a_k$ in H and $a_i \rightarrow a_k$ in G by the definition of essential graph. Hence, by the induction hypothesis [B] for $k-1$ get $a_i \rightarrow a_k$ in G^{k-1}. Thus, $a_i \in pa_{G^{k-1}}(a_k)$; the facts $a_l \in P$ and $a_i \rightarrow a_l$ in G^{k-1} imply $a_i \notin P$.

The second step is to verify (ii) is to show $pa_{G^{k-1}}(a_k) \setminus P \subseteq pa_{G^{k-1}}(P)$. Suppose $a_j \in pa_{G^{k-1}}(a_k) \setminus P$. By the induction hypothesis [B] for $k-1$ derive $a_j \rightarrow a_k$ in G (and $a_l \rightarrow a_k$ in G). By contradiction observe that $[a_l, a_j]$ is an edge in G: otherwise $a_l \rightarrow a_k \leftarrow a_j$ is an immorality in G which is an immorality in H (by Consequence 2 and Observation 2). This contradicts the assumption $a_l - a_k$ in H. Thus, $[a_l, a_j]$ must be an edge both in G and H. The fact $a_j \rightarrow a_k$ in G also implies $j < k$ and by Step III observe that either $a_j - a_k$ in H or $a_j \rightarrow a_k$ in H. One can show that the alternative $a_j - a_k$ leads to contradiction as follows: the facts $a_j - a_k - a_l$ in H and $[a_j, a_l]$ is an edge in H imply $a_j - a_l$ in H because of the absence of semi-directed cycles in H. Hence, the induction hypothesis [A] for $k-1$ gives $a_j - a_l$ in G^{k-1} which implies a contradictory conclusion $a_j \in P$. Thus, $a_j \rightarrow a_k - a_l$ in H which implies $a_j \rightarrow a_l$ in H. The induction hypothesis [A] for $k-1$ yields $a_j \rightarrow a_l$ in G^{k-1}, that is, $a_j \in pa_{G^{k-1}}(P)$.

V. If there exists a parent component P for a_k in G^{k-1} then there exists $l < k$ such that $a_l - a_k$ in H.

First observe that $G' \equiv G^{k-1}_{\{a_1,\ldots,a_k\}}$ is a chain graph without flags for which the operation of legal component merging is applicable. Indeed, by the induction hypothesis [B] for $k-1$

and the assumption that $a_1,\ldots,a_m$ is consistent with the direction of arrows in G observe that every edge in G^{k-1} between $\{a_1,\ldots,a_k\}$ and $\{a_{k+1},\ldots,a_m\}$ is an arrow towards the node in $\{a_{k+1},\ldots,a_m\}$. Thus, $pa_{G'}(A) = pa_{G^{k-1}}(A)$ for any set $A \subseteq \{a_1,\ldots,a_k\}$. This fact together with Lemma 4 allows one to show that both $C_l \equiv \{a_k\}$ and $C_u \equiv P \subseteq \{a_1,\ldots,a_{k-1}\}$ are components of G' and, moreover, the conditions (i) and (ii) from Definition 1 are fulfilled for G'.

The claim of Step V can be verified by contradiction. If no $l<k$ with $a_l — a_k$ in H exists then by Step III observe that every edge in H between $\{a_1,\ldots,a_{k-1}\}$ and a_k is an arrow towards a_k. This together with Step II and the induction hypothesis [A] for $k-1$ implies $G^{k-1}_{\{a_1,\ldots,a_k\}} = H_{\{a_1,\ldots,a_k\}}$ which is an essential graph over $\{a_1,\ldots,a_k\}$ by Step I. This implies by Statement 2 that the operation of legal component merging is not applicable to $G' = G^{k-1}_{\{a_1,\ldots,a_k\}} = H_{\{a_1,\ldots,a_k\}}$ and this contradicts an earlier observation.

VI. If there exists a parent component P for a_k in G^{k-1} then for each $i<k$ such that $a_i \to a_k$ in G^{k-1} one has $[a_i — a_k$ in $H]$ iff $a_i \in P$.

The necessity of the condition follows from Step IV which says that the component of G^{k-1} containing a_i is a parent component for a_k, that is, P (it is uniquely determined by Lemma 4).

For sufficiency of the condition use the claim in Step V which implies the existence of $l<k$ such that $a_l — a_k$ in H. By Step IV (and Lemma 4) $a_l \in P$. Therefore, an undirected path in G^{k-1} connecting a_i and a_l exists. The nodes of the path are in P and, therefore, in $\{a_1,\ldots,a_{k-1}\}$ (by Lemma 4). The induction hypothesis [A] for $k-1$ says that the path is in H as well. As $a_l — a_k$ in H, $[a_i, a_k]$ is an edge in H and H has no semi-directed cycles one has $a_i — a_k$ in H.

VII. Induction step: the conditions [A] and [B] are valid for k.

It follows from Lemma 4 that G^k may only differ from G^{k-1} in edges between $\{a_1,\ldots,a_{k-1}\}$ and a_k. The induction hypothesis [A] for $k-1$, therefore, implies $G^k_{\{a_1,\ldots,a_{k-1}\}} = G^{k-1}_{\{a_1,\ldots,a_{k-1}\}} = H_{\{a_1,\ldots,a_{k-1}\}}$. To verify [A] for k it suffices to show that the edges between $\{a_1,\ldots,a_{k-1}\}$ and a_k have the same type and direction (of potential arrows) in G^k and H. By Steps IV and V conclude that a parent component for a_k in G^{k-1} exists iff there exists a line in H between $\{a_1,\ldots,a_{k-1}\}$ and a_k. Therefore, if the parent component does not exist the claim follows from Steps II and III. If a parent component P for a_k in G^{k-1} exists then by Step VI observe $pa_{G^{k-1}}(a_k) \cap P = \{a_l\,;\, l<k,\ a_l — a_k \text{ in } H\}$. The graph G^k is made from G^{k-1} by replacing the arrows from this set to a_k by lines which implies, with the help of Steps II and III, what is desired.

The induction hypothesis [B] for $k-1$ and the way of constructing of G^k from G^{k-1} also easily implies [B] for k, which concludes the proof of the induction step. □

Remark 3. Note that the question of verification of the existence of a parent component for a node a in a chain graph H without flags is a local question in the sense it only concerns the nodes in $pa_H(a)$ (and their parents). Indeed, the reader can easily show that, provided $pa_H(a) \neq \emptyset$, the conditions (i) and (ii) from Definition 3 are valid for a component P of H iff the set

$$K = \{x \in pa_H(a)\,;\, \text{no } y \in pa_H(a) \text{ with } x \to y \text{ in } H \text{ exists}\}$$

is complete and has just the remaining nodes of $pa_H(a)$ as parents, that is,

(j) $\forall\, x, y \in K \quad$ if $x \neq y$ then $x - y$ in H,
(jj) $\forall\, x \in K \quad pa_H(x) = pa_H(a) \setminus K$.

Moreover, the set K is nothing but $pa_H(a) \cap P$ then. Note that as H has no flags, if (j) is verified, then by Observation 1 it suffices to verify (jj) for one element $x \in K$ only. The advantage of the guided algorithm is that it guarantees that every its iteration is a chain graph without flags.

7. Conclusions

The presented results hopefully clarify the role of the essential graph in the class of chain graphs which are equivalent to a fixed Bayesian network and explain its relationship to the respective largest chain graph (LCG). Note again that the results presented in Statements 1 and 2 here have also been (independently) achieved in [11] in which paper, moreover, an analogous characterization of LCGs is given. On the other hand, this paper brings a new effective algorithm for the construction of the essential graph on basis of any Bayesian network in Statement 3.

However, to judge the contribution of these results to local search methods for learning Bayesian networks thoroughly, other theoretical questions need to be answered. For example, quite an important task is the task to characterize natural neighbors of an equivalence class of Bayesian networks, that is, neighboring equivalence classes in sense of the "inclusion boundary" mentioned in [8]. Note that Chickering [6] gave recently an algorithm which, on basis of the respective essential graph, tentatively generates neighbors of an equivalence class of Bayesian networks and is able to generate all neighbors; similar (but only partial) results were presented in [4]. A direct characterization of the neighborhood in terms of the essential graph will be presented in [14].

In the author's view, a direct characterization of the (whole) neighborhood in terms of a suitable representative of the equivalence class is desirable for making learning methods based on local search even more effective. The suitable representative may appear to be the essential graph or perhaps the integral (= integer-valued) vector from [13] named *standard imset*. The main advantage of that algebraic approach [13] is that reasonable classic score criteria are linear functions of standard imsets. A relationship between essential graph representation and standard imset representation has already been established but an elegant characterization of neighbors in terms of standard imsets remains to be an open task.

Acknowledgements

This research was supported by the grants GAČR n. 201/01/1482 and GAAVČR n. A1075104.

1. S. A. Andersson, D. Madigan and M. D. Perlman, "A characterization of Markov equivalence classes for acyclic digraphs", *Annals of Statistics* **25** (1997) 505–541.
2. S. A. Andersson, D. Madigan and M. D. Perlman, "On the Markov equivalence of chain graphs, undirected graphs and acyclic digraphs", *Scandinavian Journal of Statistics* **24** (1997) 81–102.
3. S. A. Andersson, D. Madigan and M. D. Perlman, "Alternative Markov properties for chain graphs", *Scandinavian Journal of Statistics* **28** (2001) 33–85.
4. V. Auvray and L. Wehenkel, "On the construction of the inclusion boundary neighbourhood for Markov equivalence clases of Bayesian network structures", in *Uncertainty in Artificial Intelligence 18*, eds. A. Darwiche and N. Friedman (Morgan Kaufmann, San Francisco, 2002) pp. 26–35.
5. D. M. Chickering, "A tranformational characterization of equivalent Bayesian network structures", in *Uncertainty in Artificial Intelligence 11*, eds. P. Besnard and S. Hanks (Morgan Kaufmann, San Francisco, 1995) pp. 87–98.
6. D.M. Chickering, "Optimal structure identification with greedy search", to appear in *Journal of Machine Learning Research* (2003).
7. M. Frydenberg, "The chain graph Markov property", *Scandinavian Journal of Statistics* **17** (1990) 333–353.
8. T. Kočka and R. Castelo, "Improved learning of Bayesian networks", in *Uncertainty in Artificial Intelligence 17*, eds. J. Breese and D. Koller (Morgan Kaufmann, San Francisco, 2001) pp. 269–276.
9. C. Meek, "Causal inference and causal explanation with background knowledge", in *Uncertainty in Artificial Intelligence 11*, eds. P. Besnard and S. Hanks (Morgan Kaufmann, San Francisco, 1995) pp. 403–410.
10. J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, San Mateo, 1988).
11. A. Rovenato, "A unified approach to the characterisation of equivalence classes of DAGs, chain graphs with no flags and chain graphs", a technical report, Dipartimento di Scienze Sociali Cognitive e Quantitative, University of Modena and Reggio Emilia (2003), submitted to *Annals of Statistics.*
12. M. Studený, "A recovery algorithm for chain graphs", *International Journal of Approximate Reasoning* **17** (1997) 265–293.
13. M. Studený, "On probabilistic conditional independence structures", a research monograph, Institute of Information Theory and Automation, Prague, Czech Republic, January 2003, offerred to *Springer Verlag.*
14. M. Studený, "Characterization of inclusion neighbourhood in terms of the essential graph: upper neighbours", be submitted to Proceedings of ECSQARU 2003.
15. T. Verma and J. Pearl "Equivalence and synthesis of causal models", in *Uncertainty in Artificial Intelligence 6*, eds. P. P. Bonissone, M. Henrion, L. N. Kanal and J. F. Lemmer (Elsevier, New York, 1991) pp. 220–227.
16. M. Volf and M. Studený, "A graphical characterization of the largest chain graphs", *International Journal of Approximate Reasoning* **20** (1999) 209–236.