Characterization of essential graphs by means of an operation of legal component merging

Milan Studený
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, Prague, 18208, Czech Republic

Abstract

One of the most common ways of representing classes of equivalent Bayesian networks is the use of *essential graphs*. These chain graphs are also known in the literature as completed patterns or completed pdags. The name essential graph was proposed by Andersson, Madigan and Perlman (1997a) who also gave a graphical characterization of essential graphs. In this contribution 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.

1 Introduction

The author apologizes for his very sketchy explanation of his motivation which is caused in part by page limitation. Several approaches to learning Bayesian networks are based on the idea of maximization of a score metric. Some of these approaches (Chickering, 2002) 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 computer memory. A quite popular representative of an equivalence class of Bayesian networks is its *essential graph*. This term was introduced by Andersson, Madigan and Perlman in (1997a) but some alternative names for this graph have appeared in the literature: completed pattern (Verma and Pearl, 1991), maximally oriented graph for a pattern (Meek, 1995) and completed pdag (Chickering, 2002). A graphical characterization of essential graphs was presented in (Andersson et. al., 1997a) as Theorem 4.1.

The author's plan is to develop an algebraic method of description of the equivalence classes by means of certain integral vectors. The hope is that this approach could be utilized to make existing local search methods of learning Bayesian networks more effective - for a brief description of the main idea of this algebraic approach see (Studený, 2002). 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. Unlike former algorithms (Meek, 1995) and (Chickering, 1995), this alternative algorithm avoids indicating essential (= compelled) arrows.

2 Basic concepts

2.1 Graphical concepts

Graphs considered in this contribution 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 $\{a, b\} = \{b, a\} \notin \mathcal{L}(H)$.

An *undirected graph* is a hybrid graph having only lines, i.e., $\mathcal{A}(H) = \emptyset$. A *directed graph* is a hybrid graph having only arrows, i.e., $\mathcal{L}(H) = \emptyset$. 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 $C \subseteq N$ is *connected* in H if for every $a, b \in N$ there exists an undirected path connecting them, that is, there 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 inclusion). Of course, components of a directed graph are singletons (= sets of cardinality one).

An *acyclic directed graph* (ADG) is a directed graph without directed cycles, that is, without any sequence $d_1, \ldots, d_n, d_{n+1}$, $n \geq 3$ such that $d_1, \ldots, d_n$ are distinct, $d_{n+1} = d_1$ and $d_i \to d_{i+1}$ in H for $i = 1, \ldots, n$. The well-known fact (whose proof is left to the reader) is that a directed graph is acyclic iff there exists a total ordering of all nodes of N (= ordering into a sequence) $b_1, \ldots, b_m$, $m \geq 1$ which is *consistent* with the direction of arrows, i.e., whenever $b_i \to b_j$ 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 (Studený, 1997). 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. This implies that the set of *parents* of C i.e.,

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

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 understood by the set of *neighbours* of C. The notation $pa_H(x)$, respectively $ne_H(x)$, for $x \in N$ will sometimes substitute $pa_H(\{x\})$, respectively $ne_H(\{x\})$. *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. 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.

2.2 Bayesian networks

A *Bayesian network* is a certain (usually discrete) statistical model appended to an ADG.

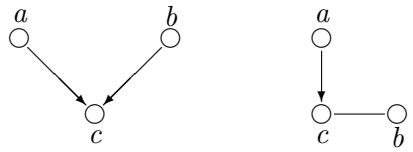

Figure 1: An immorality and a flag.

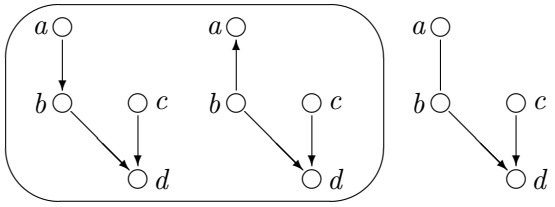

Figure 2: An equivalence class of ADGs and the respective essential graph.

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 a Bayesian network as one distribution belonging to this class. An alternative definition of that class can be formulated terms of conditional independence, using the d-separation criterion from (Pearl, 1988). Since exact definitions of related concepts are not needed in this contribution, they are omitted. Thus, the phrase "Bayesian network" can be understood as a synonym for an ADG throughout the rest of this contribution.

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 (1991) gave a direct graphical characterization of equivalent Bayesian networks which will serve as its definition in this contribution.

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 $a \to b$ in H, $b \to a$ in H or $a - b$ 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 and $[a, b]$ 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_* such that

- $a \to b$ in G_* iff $a \to b$ in G for every $G \in \mathcal{G}$,
- $a - b$ in G_* iff there are $G_1, G_2 \in \mathcal{G}$ such that $a \to b$ in G_1 and $b \to a$ in G_2.

Example. Let us consider an equivalence class $\mathcal{G}$ of ADGs 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.

Graphical characterization of essential graphs was presented in (Andersson et. al. 1997a). 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 of 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 $[a, b]$ 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

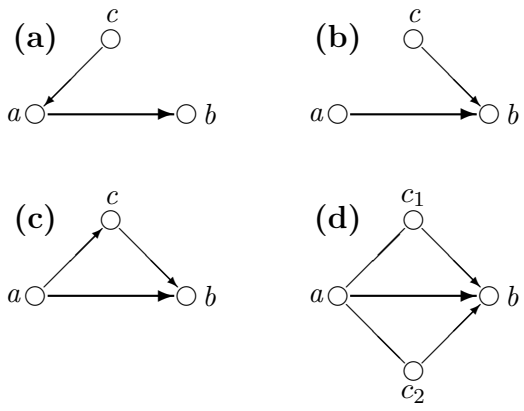

Figure 3: A strongly protected arrow $a \to b$.

on results of Frydenberg (1990). Let us call a *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 (1990) gave characterization of equivalent chain graphs which can be viewed as generalisation of the result by Verma and Pearl (1991): two chain graphs are equivalent iff they have the same underlying graph and the same complexes. Moreover, Frydenberg (1990) 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_∞ if $a \to b$ in H for every $H \in \mathcal{H}$,
- $a - b$ in H_∞ if $a - b$ in H for at least one $H \in \mathcal{H}$.

He named this 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 (Studený, 1997): every equivalence class $\mathcal{G}$ of Bayesian networks over N is a subclass of the 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 option is an algebraic approach outlined in (Studený, 2002). The idea is to represent an 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 that are described in (Studený, 2001).

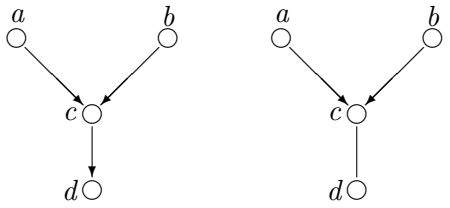

Figure 4: An essential graph and the respective largest chain graph.

LEMMA 1 A hybrid graph H over N is an essential graph of an equivalence class of ADGs 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 (Andersson et. al., 1997a). □

3 Chain graphs without flags

The goal of this contribution is to characterize every essential graph as a distinguished member of a certain class of chain graphs. Actually, the concept of an essential graph can be viewed as an analogy of the concept of LCG. The only diference 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. Use the induction then. □

Equivalence of chain graphs without flags is simple.

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

Proof: Use Frydenberg's (1990) characterization of equivalent chain graphs mentioned in Remark 1. Observe that the only complex in a chain graph without flags is an immorality. □

CONSEQUENCE 1 Let G be a Bayesian network over N, $\mathcal{G}$ the equivalence class of ADGs containing G, and 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 observation 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 (Andersson et. al., 1997b) where chain graphs equivalent to ADGs are characterized. More explicit characterization in Remark 4.2 there contains 3 conditions: (i), (ii) and (iii). The reader can consult (Andersson et. al., 1997b) and verify by means of Observation 1 that the conditions (ii) and (iii) are always fulfilled for chain graphs without flags. The condition (i) requires exactly what is mentioned in Lemma 2. □

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.

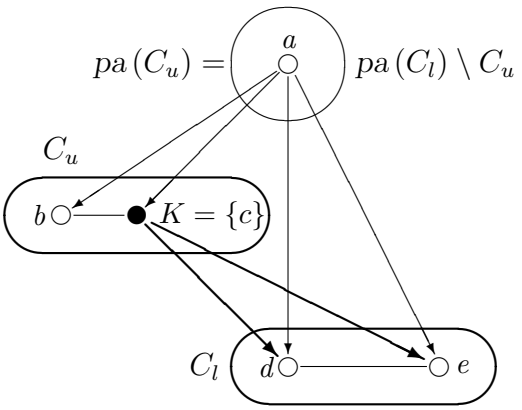

Figure 5: The situation before legal merging.

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 assumption that the set K is non-empty implies that C_u and C_l are distinct. Moreover, by Observation 1 we derive that $a \to b$ in H for $a \in C_u$, $b \in C_l$ iff $a \in K$. By (ii) $pa_H(C_u) \subseteq pa_H(C_l)$ and this observation allows one to derive that the only type of an edge between C_u and C_l is an arrow from C_u 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 $a \leftarrow b$ then $b \in pa_H(C_u) \setminus pa_H(C_l)$ since $C_l \cap pa_H(C_l) = \emptyset$. 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 of Definition 1, the constructed graph H' is a chain graph without flags, which is equivalent to the original graph H, and has $C_u \cup C_l$ 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 $d_1 \rightarrow d_2$. 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. Thus, it is a path in a chain graph H, which implies $d_k \in C_u$ and $d_j \in C_l$ (recall that $d_1 \rightarrow d_2$). Observe that $d_{j-1} \rightarrow d_j$ 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 $d_1 \rightarrow d_2$). Thus, $d_{j-1} \in pa_H(C_l) \setminus C_u$ and, by the assumption (ii) and Observation 1, $d_{j-1} \in pa_H(C_u) = pa_H(d_k)$. Therefore, $d_j, d_k, \ldots, d_{n+1} = d_1, \ldots, d_{j-1}$ 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 $a \rightarrow c - b$. Clearly, $a \rightarrow c$ in H and the option $c - b$ in H is excluded. If $c \rightarrow b$ in H then $c \in C_u$, $b \in C_l$ and, by (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 $b \rightarrow c$ 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 could use (i), while in case $a \notin C_u$ (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 $a \rightarrow c \leftarrow b$ 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 (i) we derive a contradictory conclusion that $[a, b]$ is an edge in H.
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 2 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 assumptions of Definition 1, then every arrow in H is strongly protected.

Proof: Let $a \rightarrow b$ 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 $a \rightarrow b$ 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 $a \rightarrow b$ 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 $a \leftarrow c$ and $a - c$ in H are excluded. If $a \rightarrow c$ then $a \rightarrow b$ 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 $[c_1, c_2]$ is not an edge. By Observation 1 conclude that $c_1 \rightarrow b \leftarrow c_2$ 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 $a \rightarrow b$ 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 $a \rightarrow b$ belong 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.
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 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 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 $a \to b$ 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 3 Let G be a Bayesian network over N, $\mathcal{G}$ the equivalence class of ADGs 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 1, $G_* \in \mathcal{H}(G)$. 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 "maximal" 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

Consequence 3 leads to the following procedure.

BLIND ALGORITHM
Let H be any chain graph without flags which is equivalent to an ADG 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 is not possible.

Of course, one can take a Bayesian network G in place of H. It follows from Consequence 3 and Statement 2 that the algorithm stops with the essential graph of the equivalence class of G. However, in the case of a Bayesian network an even more specific version of this algorithm can be utilized. It is based on the concept of a *parent component* for a node (which corresponds to the situation when the lower component in legal merging is a singleton).

OBSERVATION 3 Let G be a chain graph over N without flags and $c \in N$ such that $ne_G(c) = \emptyset$. Then a component P of G satisfying

(i) $pa_G(c) \cap P$ is non-empty and complete,

(ii) $pa_G(c) \setminus P = pa_G(P)$,

is uniquely determined if it exists.

Proof: Suppose for the sake of contradiction that two distinct components P_1 and P_2 satisfy (i) and (ii). Find $d_i \in pa_G(c) \cap P_i$ for $i = 1, 2$. Since $d_1 \notin P_2$ resp. $d_2 \notin P_1$, by (ii) derive $d_1 \in pa_G(c) \setminus P_2 = pa_G(P_2)$, resp. $d_2 \in pa_G(c) \setminus P_1 = pa_G(P_1)$. By Observation 1 $d_1 \to d_2$, resp. $d_2 \to d_1$, which is impossible in a hybrid graph. □

GUIDED ALGORITHM
Let G be an ADG over N and $a_1, \ldots, a_n$, $n \geq 1$ a total ordering of its nodes consistent with the direction of arrows in G. Let us construct a sequence of hybrid graphs $G^1, \ldots, G^n$ over N as follows:

- *put $G^1 = G$,*
- *for $k = 2, \ldots, n$ check whether there exists a component P of G^{k-1} satisfying the*

conditions (i)-(ii) from Observation 3 for $c = a_k$. *If yes then* G^k *arises from* G^{k-1} *by legal merging of* $C_u = P$ *and* $C_l = \{a_k\}$, *otherwise* $G^k = G^{k-1}$.

The algorithm stops with G^n.

STATEMENT 3 *The graph* G^n *obtained by the guided algorithm above is the essential graph of the equivalence class of* G.

The proof of Statement 3 is quite long and therefore it is omitted (page limitation).

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). However, to assess the contribution to local search methods for learning Bayesian networks thoroughly, other theoretical questions need to be answered. For example, quite an important task is to characterize natural neighbours of an equivalence class of Bayesian networks, that is, neighbouring equivalence classes in sense of the "inclusion boundary" mentioned in (Kočka and Castello, 2001). Note that Chickering (2002) has recently characterized these neighbours in terms of a whole equivalence class.

In the author's view, a characterization 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 vector from (Studený, 2002) named *standard imset*. The main advantage of that algebraic approach is that reasonable score criteria are linear functions of standard imsets. The author has already established a relationship between essential graph representation and standard imset representation (still in notes written by hand) but an elegant characterization of neighbours remains an open task.

Acknowledgements

This research was supported by the grants GAČR n. 201/01/1482 and GAAVČR n. A107104.

References

Andersson, S.A., Madigan, D., and Perlman, M.D. (1997a). A characterization of Markov equivalence classes for acyclic digraphs. Annals of Statistics 25, 505-541.

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

Andersson, S.A., Madigan, D., and Perlman, M.D. (2001). Alternative Markov properties for chain graphs. Scandinavian Journal of Statistics 28, 33-85.

Chickering, D.M. (1995). A tranformational characterization of equivalent Bayesian network structures. In Uncertainty in Artificial Intelligence 11 (P. Besnard, S. Hanks eds.), Morgan Kaufmann, 87-98.

Chickering, D.M. (2002). Optimal structure identification with greedy search. Submitted to Journal of Machine Learning Research.

Frydenberg, M. (1990). The chain graph Markov property. Scandinavian Journal of Statistics 17, 333-353.

Kočka, T. and Castello, R. (2001). Improved learning of Bayesian networks. In Uncertainty in Artificial Intelligence 17 (J. Breese, D. Koller eds.), Morgan Kaufmann, 269-276.

Meek, C. (1995). Causal inference and causal explanation with background knowledge. In Uncertainty in Artificial Intelligence 11 (P. Besnard, S. Hanks eds.) Morgan Kaufmann, 403-410.

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann.

Studený, M. (1997). A recovery algorithm for chain graphs. International Journal of Approximate Reasoning 17, 265-293.

Studený, M. (2001). On mathematical description of probabilistic conditional independence structures. DrSc thesis, Institute of Information Theory and Automation, Prague, Czech Republic.

Studený, M. (2002). An algebraic approach to learning Bayesian networks. Poster from the 7th Valencia International Meeting on Bayesian Statistics, Tenerife, Spain. Available at http://www.utia.cas.cz/user_data/studeny/p1.html

Verma, T. and Pearl, J. (1991). Equivalence and synthesis of causal models. In Uncertainty in Artificial Intelligence 6 (P.P. Bonissone, M. Henrion, L.N. Kanal, J.F. Lemmer eds.), Elsevier, 220-227.

Volf, M. and Studený, M. (1999). A graphical characterization of the largest chain graphs. International Journal of Approximate Reasoning 20, 209-236.