Characterization of Inclusion Neighbourhood in Terms of the Essential Graph: Lower Neighbours

Milan Studený*
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
studeny@utia.cas.cz

Abstract

The topic of the contribution is to characterize the inclusion neighbourhood of a given equivalence class of Bayesian networks in terms of the respective essential graph in such a way that it can be used efficiently in a method of local search for maximization of a quality criterion. One can distinguish two kinds of inclusion neighbours: upper and lower ones. A previous paper [15] gave a characterization of upper inclusion neighbourhood, and this contribution gives a characterization of lower inclusion neighbourhood.

It is shown here that each inclusion neighbour is uniquely described by a pair $([a,b],C)$ where $[a,b]$ is an unordered pair of distinct nodes and $C \subseteq N \setminus \{a,b\}$ a disjoint set of nodes in the essential graph. The second basic result is that, for given $[a,b]$, the collection of those sets C which correspond to lower inclusion neighbours has a special form. More specifically, given an unordered pair $[a,b]$ of nodes which is not an edge in the essential graph, the respective collection of sets C is the union of (at most) two *tufts* of sets. The least and maximal sets of these two tufts, which determine them, can also be read from the essential graph.

1 Motivation

1.1 Learning Bayesian Networks

Some of the approaches to learning Bayesian networks use the method of maximization of a *quality criterion*, named also 'quality measure' [3] and 'score metric' [4]. Quality criterion is a function, designed by a statistician, which ascribes a real number to data and a network. This number evaluates how the statistical model determined by the network is suitable to explain the occurence

*This research has been supported by the grant GA ČR n. 201/01/1482.

of data. Since the actual aim of a learning procedure is to get a statistical model (defined by a network) reasonable quality criteria do not distinguish between equivalent Bayesian networks, that is, between networks which define the same statistical model. Therefore, from operational point of view, the goal is to learn an equivalence class of Bayesian networks, that is, a class of acyclic directed graphs.

As direct maximization of a quality criterion is typically infeasible the *method of local search* is often used. The main idea of this approach is that suitable concept of neighbourhood is introduced for acyclic directed graphs over a fixed set of nodes N. The point is that the change in the value of a (reasonable) quality criterion is easy to compute for neighbouring graphs. Thus, instead of global maximization of a quality criterion one searches for a local maximum of the criterion with respect to the considered neighbourhood structure and this task is usually computationally feasible. Typical neighbourhood structures used in practice are defined by means of simple graphical operations with considered graphs - for details see [5, 7].

The algorithms of this kind can also be classified according to the method of representation of equivalence classes of networks. In some algorithms, an equivalence class is represented by any of its members which may, however, result in computational complications. In other algorithms, a special representative of each equivalence class is used. The most popular representative of an equivalence class of Bayesian networks is the *essential graph* which is a certain chain graph describing some common features of acyclic directed graphs from the class. The term 'essential graph' was proposed by Andersson, Madigan and Perlman [1]; altenative names 'completed pattern' [18], 'maximally oriented graph for a pattern' [9] and 'completed pdag' [5] also appeared in the literature.

1.2 Inclusion Neighbourhood

There exists a neighbourhood structure (for equivalence classes of Bayesian networks) which has a good theoretical basis. The inverse inclusion of statistical models defined by the networks, which corresponds to the inclusion of conditional independence structures induced by the networks, defines a natural *inclusion ordering* on the collection of equivalence classes. This ordering induces a neighbourhood concept then. More specifically, two different types of neighbouring equivalence classes are assigned to every equivalence class of networks: the *upper* neighbours and *lower* neighbours. Thus, the *inclusion neighbourhood*, sometimes also named 'inclusion boundary neighbourhood' [7, 2], consists of these two parts. There are also some practical reasons for using the inclusion neighbourhood - for details see [4]. Note that Chickering [5] has recently confirmed Meek's conjecture [10] about transformational graphical characterization of the inclusion ordering. A consequence of this result is a graphical description of the inclusion neighbourhood in terms of the collection of graphs in the considered equivalence class (see Section 2.4).

The topic of this contribution and a previous one [15][1] is to characterize the inclusion neighbourhood of a given equivalence class of Bayesian networks in terms of the respective essential graph in such a way that it can be used efficiently in a method of local search for maximization of a quality criterion. Two recent papers were devoted to this problem, but, in author's view, none of them brought a satisfactory solution to the problem.

Chickering, in Section 5 of [5], gave a method which is able to generate tentatively all neighbouring equivalence classes (of a given equivalence class described by the respective essential graph). More specifically, two composite graphical operations applicable to an essential graph and respective legality tests which are able to decide whether the respective graphical operation leads to a real neighbouring equivalence class are designed in that paper. One of the operations and the respective legality test are aimed to generate upper neighbours, the other operation and test correspond to lower neighbours. Although the graphical description of the inclusion neighbourhood in terms of individual networks from Section 2.4 implies that every inclusion neighbour can be reached in this way the method has two drawbacks.

- The first drawback of the method is that it is tentative: different graphical operations may lead to the same equivalence class. Therefore, additional checking must be done to cure this imperfection.

- The second drawback of this mechanistic approach is that it does not allow one to discern possible internal structure of the inclusion neighbourhood.

Auvray and Wehenkel [2] made an attempt at direct characterization of the inclusion neighbourhood. Their characterization of the upper inclusion neighbourhood removes the first drawback. They uniquely characterized and classified neighbouring equivalence classes of a given equivalence class (described in terms of the respective essential graph) by means of certain mathematical objects. However, these object are still unnecessarily complicated which means that their characterization of upper inclusion neighbourhood is too awkward. In particular, the second drawback is not removed by their approach since their approach does not allow one to make out the internal structure of the inclusion neighbourhood. Moreover, their characterization is incomplete: only partial direct characterization of lower neighbours is given in their paper.

1.3 Consistent Characterization of Inclusion Neighbours

In this paper an elegant characterization of the inclusion neighbourhood of a given equivalence class in terms of the respective essential graph is presented. Each inclusion neighbour is uniquely described by a pair $([a,b],C)$ where $[a,b]$ is an unordered pair of distinct nodes and $C \subseteq N \setminus \{a,b\}$ a disjoint set of nodes. More specifically, $[a,b]$ is an edge of the essential graph in case of the

[1] These two papers have the same motivation and basic tools for which reason their first two sections are nearly identical.

upper neighbourhood while $[a,b]$ is a pair of nodes which is not an edge in the essential graph in case of the lower neighbourhood.

The first new observation made in [15] and here is that every inclusion neighbour is uniquely characterized by a pair $([a,b],C)$ of this kind. The second observation is that, for given $[a,b]$, the collection of those sets C which correspond to inclusion neighbours has a special form.

In this contribution, a complete analysis of the lower inclusion neighbourhood is given. In this case, the collection of sets C for a given pair $[a,b]$ which is not an edge in the essential graph has the form of the union of at most two *tufts* of sets. By a tuft is meant a collection of sets with the least set (= the unique minimal set) and with possibly several maximal sets such that every set which contains the least set and which is contained in one of the maximal sets belongs to the collection. In particular, every tuft is completely described by its least set and by the list of its maximal sets. Given an essential graph G^* and a pair of its nodes $[a,b]$ which is not an edge in G^* the least and maximal sets of the respective tufts are characterized directly in terms of G^*.

Recall that an analogous characterization of upper inclusion neighbourhood was presented in [15]. In that case, given an edge $[a,b]$ in the essential graph, the respective collection of sets is a tuft of sets.

Note that the characterization of inclusion neighbours by means of pairs $([a,b],C)$ where $C \subseteq N \setminus \{a,b\}$ and the way how it is done in [15] and in this paper is not incidental. An interesting fact is that, from a certain perspective which is explained in details in Chapter 8 of [17], the pair $([a,b],C)$ has close relation to conditional independence interpretation of the 'move' from the considered equivalence class towards its respective inclusion neighbour.

The proofs given in the Appendix combine the ideas motivated by an arithmetic approach to the description of Bayesian network models from [17] with certain graphical procedures which were already used in [2].

2 Basic Concepts

2.1 Graphical Notions

Graphs considered in this paper have a finite non-empty set N as the set of *nodes* 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 is $a - b$. 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. 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 unordered pair $[a,b]$ of distinct elements of N will be called an *edge in* H (between a and b) if one of the following cases occurs: $a - b$ in H, $a \to b$ in H and $b \to a$ in H. If $\emptyset \neq A \subseteq N$ then 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 all subsets of A).

A set $K \subseteq N$ is *complete* in a hybrid graph H over N if $\forall\, a, b \in N$ $a \neq b$ one has $a - b$ in H. By a *clique* of H will be understood a maximal complete set in H (with respect to set inclusion). The collection of cliques of H will be denoted by *cliques*(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.

An *undirected graph* is a hybrid graph without arrows, that is, $\mathcal{A}(H) = \emptyset$. An undirected graph H is *triangulated* if, for every undirected cycle in H which 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 *directed graph* is a hybrid graph having arrows only, that is, $\mathcal{L}(H) = \emptyset$. An *acyclic directed graph* 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 is that a directed graph is acyclic iff there exists a total ordering of all nodes of N $a_1, \ldots, a_m$, $m \geq 1$ 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 sets, called *blocks*, $B_1, \ldots, B_m$, $m \geq 1$ such that

- if $a - b$ in H then $a, b \in B_i$ for some $1 \leq i \leq m$,
- if $a \to b$ in H then $a \in B_i, b \in B_j$ with $1 \leq i < j \leq m$.

Clearly, every undirected graph and every acyclic directed graph is a chain graph. 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 [13]. In particular, there is no arrow in a chain graph between nodes of a connected set $C \subseteq N$; in other words, the induced subgraph H_C is undirected. Thus, the set of *parents* of C, that is,

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

is disjoint with C if C is connected. The set

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

will be named the set of *neighbours* of C.

Two concepts of ancestor will be distinguished in this paper. If there exists a *descending path* from a to b in H, that is, a sequence of distinct nodes $a =$

$c_1, \ldots, c_n = b$, $n \geq 1$ such that either c_i — c_{i+1} or $c_i \to c_{i+1}$ in H for $i = 1, \ldots, n-1$, then a is called an *ancestor* of b in H. The set of ancestors of b in H will be denoted by $an_H(b)$. If there exists a (strictly) *directed path* in H from a to b, that is, a sequence of distinct nodes $a = d_1, \ldots, d_n = b$, $n \geq 2$ such that $d_i \to d_{i+1}$ in H for $i = 1, \ldots, n-1$, then a is called a *strict ancestor* of b in H. The set of strict ancestors of b in H will be denoted by $An_H(b)$. Observe that $An_H(b) \subset an_H(b)$ and if H is a chain graph then $a \in an_H(b) \setminus An_H(b)$ for any a which belongs to the same connectivity component like b.

2.2 Bayesian Networks and Their Equivalence

A *Bayesian network* is a certain statistical model, that is, a class of (multidimensional probability) distributions, appended to an acyclic directed graph. It could be introduced as the class of distributions (on a fixed sample space) which factorize according to the graph in a certain way. An alternative definition of that class can be given terms of conditional independence restrictions, using the d-separation criterion from [11] or using the moralization criterion from [8] which are known to be equivalent. Because exact definitions of these concepts are not needed in this paper they are omitted. Nevertheless, given an acyclic directed graph G over N, the symbol $\mathcal{I}(G)$ will be used to denote the collection of conditional independence restrictions determined by G. Moreover, the phrase "Bayesian network" will be used as a synonym for an acyclic directed graph throughout the rest of the paper.

An important concept is the concept of (Markov) *equivalence of Bayesian networks*. Two Bayesian networks G_1 and G_2 are considered to be equivalent if they represent the same statistical model, which requirement is typically equivalent to the condition $\mathcal{I}(G_1) = \mathcal{I}(G_2)$. Given a equivalence class $\mathcal{G}$ of Bayesian networks over N the symbol $\mathcal{I}(\mathcal{G})$ will denote the shared collection of conditional independence restrictions $\mathcal{I}(G)$ for $G \in \mathcal{G}$. Verma and Pearl [18] gave a direct graphical characterization of equivalent Bayesian networks which can be used as its formal definition here. 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 (graphically) *equivalent* iff they have the same underlying graph and the same collection of immoralities. The equivalence characterization makes the following definition consistent: given an equivalence class $\mathcal{G}$ of Bayesian networks, a pair $[a, b]$ of distinct nodes is called an *edge in* $\mathcal{G}$ if $[a, b]$ is an edge in some $G \in \mathcal{G}$, which means, it is an edge in every $G \in \mathcal{G}$.

2.3 Essential Graphs

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

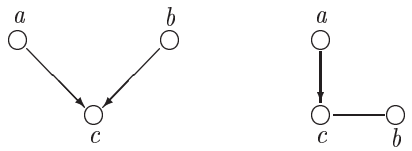

Figure 1: An immorality and a flag.

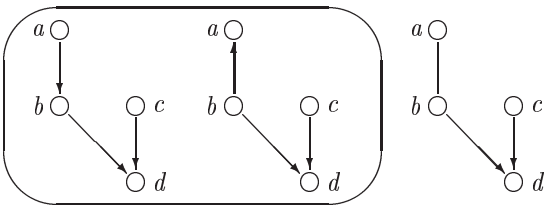

Figure 2: An equivalence class of Bayesian networks and the respective essential graph.

- $a \to b$ 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 2.1 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 given by Andersson, Madigan and Perlman as Theorem 4.1 in [1]. Recently, a simpler alternative characterization has been found in [16] and, independently, in [12]. Because a complete characterization of essential graphs is not needed in this paper it is omitted. However, what is needed is the following observation. It follows from Theorem 4.1 of [1] that every essential graph H (of an equivalence class of Bayesian networks) is a chain graph without flags such that, for every component C of H, the induced subgraph H_C is a triangulated graph. Recall that by a *flag* in a hybrid graph H is meant 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. Note that every chain graph without flags has the following pleasant property: for every component C of H and $a, b \in C$ one has $pa_H(a) = pa_H(b)$; in particular, $pa_H(a) = pa_H(C)$ for any $a \in C$.

To formulate another important fact about essential graphs recall that every chain graph can also be interpreted as a statistical model. Indeed, the moralization criterion from [8], alternatively the c-separation criterion from [14], allows

one to ascribe the collection of conditional independence restrictions $\mathcal{I}(H)$ to every chain graph H over N. This defines the respective statistical model and induces the concept of equivalence for chain graphs over N. A direct graphical characterization of equivalent chain graphs was given by Frydenberg [6]. It follows from that result that two chain graphs without flags are equivalent iff they have the same underlying graph and immoralities - see Observation 2 in [16]. Thus, Frydenberg's result implies the result by Verma and Pearl [18]. Another basic fact about the essential graph G^* of an equivalence class $\mathcal{G}$ is that it is equivalent to every $G \in \mathcal{G}$, that is, $\mathcal{I}(G^*) = \mathcal{I}(G)$ for $G \in \mathcal{G}$ - see Consequence 2 in [16].

2.4 Inclusion Ordering and Neighbourhood

The *inclusion ordering* on the set of equivalence classes of Bayesian networks over a fixed set of nodes N is defined by the binary relation $\mathcal{I}(\mathcal{K}) \subseteq \mathcal{I}(\mathcal{L})$ for equivalence classes $\mathcal{K}$ and $\mathcal{L}$. The symbol $\mathcal{I}(\mathcal{K}) \subset \mathcal{I}(\mathcal{L})$ will denote the *strict inclusion*, that is, the situation when $\mathcal{I}(\mathcal{K}) \subseteq \mathcal{I}(\mathcal{L})$ and $\mathcal{I}(\mathcal{K}) \neq \mathcal{I}(\mathcal{L})$. Finally, the symbol $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$ will mean that $\mathcal{I}(\mathcal{K}) \subset \mathcal{I}(\mathcal{L})$ but there is no equivalence class $\mathcal{G}$ of Bayesian networks over N such that $\mathcal{I}(\mathcal{K}) \subset \mathcal{I}(\mathcal{G}) \subset \mathcal{I}(\mathcal{L})$. If this is the case then $\mathcal{L}$ will be called the *upper neighbour* of $\mathcal{K}$ and $\mathcal{K}$ will be called the *lower neighbour* of $\mathcal{L}$. By the *inclusion neighbourhood* of an equivalence class is understood the collection its upper and lower neighbours.

Transformational characterization of the inclusion ordering from [5] allows one to derive a simple graphical description of the relation $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$ as its consequence - see Lemma 8.5 in [17].

Lemma 2.1 If $\mathcal{K}$ and $\mathcal{L}$ are equivalence classes of Bayesian networks over N then one has $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$ iff there exists $K \in \mathcal{K}$ and $L \in \mathcal{L}$ such that K is made of L by adding of (exactly) one arrow.

2.5 Tuft of Sets

Let $\mathcal{T}$ be a non-empty collection of subsets of N, that is, $\emptyset \neq \mathcal{T} \subseteq \mathcal{P}(N)$, and $\mathcal{T}_{\max}$ denotes the collection of maximal sets in $\mathcal{T}$ (with respect to set inclusion). The collection $\mathcal{T}$ will be called a *tuft* of sets if

- $\mathcal{T}$ has the least set $T_{\min}$, that is, $T_{\min} \in \mathcal{T}$ with $T_{\min} \subseteq T$ for each $T \in \mathcal{T}$,
- every set $T \subseteq N$ with $T_{\min} \subseteq T \subseteq T'$ for some $T' \in \mathcal{T}_{\max}$ belongs to $\mathcal{T}$.

Thus, a tuft of sets $\mathcal{T}$ is determined by its unique least set $T_{\min}$ and by the class of its maximal sets $\mathcal{T}_{\max}$. Alternatively, it can be described by $T_{\min}$ and the class $\{T' \setminus T_{\min}\,;\, T' \in \mathcal{T}_{\max}\}$. More specifically, assume that $A \subseteq N$ and $\mathcal{B}$ is a non-empty class of incomparable subsets of $N \setminus A$, that is, there are no sets $B, B' \in \mathcal{B}$ with $B \subset B'$. Introduce a special notation:

$$\mathsf{TUFT}(A|\mathcal{B}) \,=\, \{\, T = A \cup C\,;\, \exists\, B \in \mathcal{B}\;\; C \subseteq B\,\}\,.$$

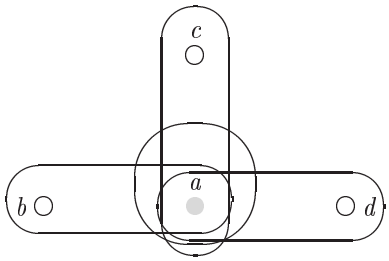

Figure 3: A tuft of sets $\mathsf{TUFT}(\{a\}|\{b\},\{c\},\{d\})$.

Evidently, $\mathcal{T} = \mathsf{TUFT}(A|\mathcal{B})$ is a tuft such that $\mathcal{T}_{\max} = \{A \cup B; B \in \mathcal{B}\}$ and $T_{\min} = A$. Of course, every tuft of subsets of N can be described in this way.

Example 2.2 Suppose $N = \{a,b,c,d\}$ and put $A = \{a\}$. Consider the class $\mathcal{B} = \{\,\{b\},\{c\},\{d\}\,\}$ which is a class of incomparable subsets of $N \setminus A$. Actually, the sets in $\mathcal{B}$ are disjoint. Then $\mathsf{TUFT}(A|\mathcal{B})$ consists of four sets: $\{a\}$, $\{a,b\}$, $\{a,c\}$ and $\{a,d\}$. The tuft is shown in Figure 3.

3 Lower Inclusion Neighbourhood

3.1 Description of Lower Neighbours

By the *lower neighbourhood* of an equivalence class $\mathcal{L}$ of Bayesian networks is understood the collection $o^{\downarrow}(\mathcal{L})$ of equivalence classes $\mathcal{K}$ such that $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$. It follows from Lemma 2.1 that each $L \in \mathcal{L}$ and each ordered pair of distinct nodes which is not an edge in $\mathcal{L}$ may define together an element of $o^{\downarrow}(\mathcal{L})$ and every element of $o^{\downarrow}(\mathcal{L})$ is obtained in this way. Thus, the lower neighbourhood $o^{\downarrow}(\mathcal{L})$ is, in fact, described in terms of elements of $\mathcal{L}$. Nevertheless, the above described correspondence is not a one-to-one mapping because different elements of $\mathcal{L}$ and different choices of the direction of the arrow which is added may yield the same neighbouring class $\mathcal{K}$.

One the other hand, every neighbouring class is uniquely described by a certain pair $([a,b],C)$ where $a,b \in N$, $a \neq b$ and $C \subseteq N \setminus \{a,b\}$. The pair $([a,b],C)$ can be introduced in graphical terms as follows.

> Let $\mathcal{L}$ be an equivalence class of Bayesian networks over N and $\mathcal{K} \in o^{\downarrow}(\mathcal{L})$. Choose $L \in \mathcal{L}$ and $K \in \mathcal{K}$ such that K is obtained from L by adding of an arrow $a \to b$. Then $\mathcal{K}$ will be described by the pair $([a,b],C)$ where $C = pa_L(b)$ and $[a,b]$ is viewed as an unordered pair.

To show that the definition above is consistent one has to show that the pair $([a,b],C)$ does not depend on the choice of L and K and that distinct pairs are ascribed to distinct lower neighbours.

Proposition 3.1 *Let $\mathcal{L}$ be an equivalence class of Bayesian networks, $\mathcal{K}_1, \mathcal{K}_2 \in o^{\downarrow}(\mathcal{L})$. Suppose, for $i = 1, 2$, that graphs $L_i \in \mathcal{L}$ and $K_i \in \mathcal{K}_i$ are given such that K_i is made of L_i by adding of an arrow $a_i \to b_i$ and $C_i = pa_{L_i}(b_i)$. Then $\mathcal{K}_1 = \mathcal{K}_2$ iff $[\{a_1, b_1\} = \{a_2, b_2\}$ and $C_1 = C_2]$.*

The proof of Proposition 3.1, which is given in Section 5.1 of the Appendix, is based on a special arithmetic characterization of equivalence of Bayesian networks. Perhaps one can also prove this result using purely graphical tools, but the given proof is more elegant.

Remark 3.1 Note that above mentioned way of description of lower inclusion neighbourhood is consistent with the way which was used in [15] for description of upper neighbours. Indeed, if $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$ then the pair $([a, b], C)$ which describes $\mathcal{K}$ as one of the lower neighbours of $\mathcal{L}$ coincides with the pair which describes $\mathcal{L}$ as one of the upper neighbours of $\mathcal{K}$.

Example 3.1 To illustrate the concepts introduced above let us consider the essential graph L^* shown at the top of Figure 4 (in a single oval). The respective equivalence class $\mathcal{L}$ of Bayesian networks over $N = \{a, b, c, d, e\}$ is show below it (in a double oval). The figure describes a part of its lower neighbourhood, namely those neighbours which correspond to adding (an edge) $[a, b]$. The third layer of Figure 4 contains acyclic directed graphs obtained from graphs in $\mathcal{L}$ in that way: K_1 is obtained from L_1 by adding $a \to b$, K_2 is obtained from L_1 by adding $b \to a$, K_3 is obtained from L_2 by adding $a \to b$ and K_4 is obtained from L_2 by adding $b \to a$. Two corresponding essential graphs are on the next layer: K_1^* corresponds to K_1 while K_2^* corresponds to K_2, K_3 and K_4. The respective uniquely characterizing pairs are below them.

3.2 Characterization of Lower Neighbourhood

Given an equivalence class $\mathcal{L}$ the aim is to characterize those pairs $([a, b], C)$ which define elements $\mathcal{K} \in o^{\downarrow}(\mathcal{L})$. In this section, this task is answered for a fixed unordered pair of distinct nodes $[a, b]$. The first step to do this is to characterize those lower neighbours which correspond to adding of an arrow $a \to b$. For this purpose put

$$\mathcal{C}^+_{\mathcal{L}}(a \to b) \;=\; \{\, C\,;\; \exists\, L \in \mathcal{L} \text{ such that } b \notin an_L(a) \text{ and } C = pa_L(b)\,\}$$

for every ordered pair of distinct nodes (a, b) such that $[a, b]$ is not an edge in $\mathcal{L}$. Note that the condition $b \notin an_L(a)$ is a necessary and sufficient condition for the graph K obtained from L by adding $a \to b$ to be acyclic.

It follows from what it says in Section 3.1 that $\mathcal{C}^+_{\mathcal{L}}(a \to b) \cup \mathcal{C}^+_{\mathcal{L}}(b \to a)$ is the class of sets which needs to be characterized. Therefore, given an ordered pair (a, b), one first needs to find out when $\mathcal{C}^+_{\mathcal{L}}(a \to b)$ is non-empty and describe that collection in that case. The following concept is useful for that purpose.

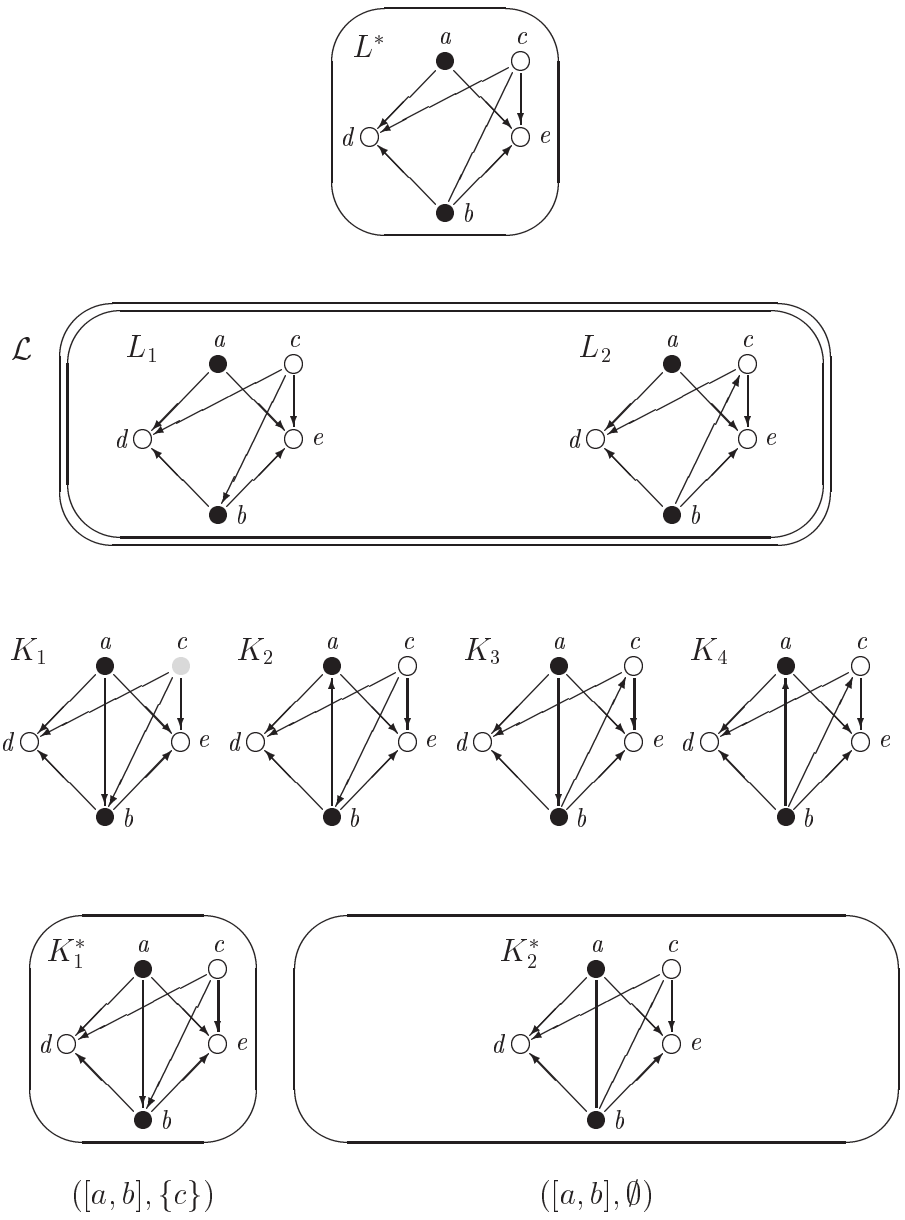

Figure 4: Lower inclusion neighbourhood - an example.

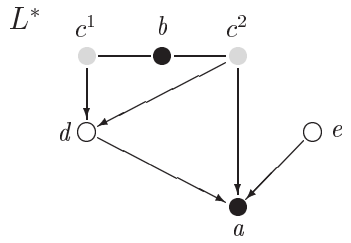

Figure 5: An essential graph with an incomplete set of relative neighbours.

Definition 3.1 Let H be a chain graph without flags and (a, b) an ordered pair of nodes in H such that $[a, b]$ is not an edge in H. We say that $c \in ne_H(b)$ is a *relative neighbour* of b with respect to a and write $c \in re_H(b|a)$ if there exist a descending path in H from c to a which is outside $ne_H(b) \setminus \{c\}$.

Remark 3.2 Note that the set of relative neighbours $re_H(b|a)$ can be equivalently introduced as the set $c \in ne_H(b)$ such that there exists a path in H of the form $c = d_1 - \ldots - d_k \to \ldots \to d_n = a$, $n \geq 2$ such that $1 \leq k \leq n$ and $d_i \notin ne_H(b)$ for $i = 2, \ldots, k$. Indeed, if there exist a path mentioned in Definition 3.1 then consider a path of this kind which cannot be shorthened. Because H has no flags it necessarily has the form mentioned in this remark. This condition is maybe more suitable for verification.

Proposition 3.2 *Let $\mathcal{L}$ be an equivalence class of Bayesian networks, L^* the essential graph of $\mathcal{L}$ and (a, b) an ordered pair of distinct nodes of L^* such that $[a, b]$ is not an edge in L^*. Put $P = pa_{L^*}(b)$, $R = re_{L^*}(b|a)$ and introduce $M = \{d \in ne_{L^*}(b) \setminus R\,;\ d - c$ in L^* for every $c \in R\}$. Then*

(i) *$\mathcal{C}^+_{\mathcal{L}}(a \to b) \neq \emptyset$ iff there exists no (strictly) directed path in L^* from b to a, that is, $b \notin An_{L^*}(a)$ and, moreover, R is a complete set in L^*;*

(ii) *if this is the case then $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(P \cup R|cliques(L^*_M))$ where $cliques(L^*_\emptyset) = \{\emptyset\}$ by convention.*

The proof is given in Section 5.2 of the Appendix.

Remark 3.3 The condition $b \notin An_{L^*}(a)$ is equivalent to the requirement that there is no descending path in L^* from b to a which starts by an arrow. Indeed, if there exists a path like that then consider one which cannot be shorthened. Because L^* has no flags it is necessarily a directed path.

The condition $b \notin An_{L^*}(a)$ need not imply that the set $R = re_{L^*}(b|a)$ is complete in L^*. An example is given in Figure 5 where the set of relative neighbours $R = re_{L^*}(b|a) = \{c^1, c^2\}$ is not complete in L^* despite the fact that there is no directed path in L^* from b to a.

The previous result makes it possible to provide a complete analysis of that part of lower inclusion neighbourhood which corresponds to a given unordered pair of nodes $[a, b]$.

Corollary 3.1 Let $\mathcal{L}$ be an equivalence class of Bayesian networks over N, L^* the essential graph of $\mathcal{L}$ and $[a, b]$ an unordered pair of nodes which is not an edge in L^*. Let us put $P_a = pa_{L^*}(a)$, $R_a = re_{L^*}(a|b)$,

$$M_a = \{d \in ne_{L^*}(a) \setminus R_a \,;\, \forall\, c \in R_a \;\; d - c \text{ in } L^* \}\,,$$

$P_b = pa_{L^*}(b)$, $R_b = re_{L^*}(b|a)$ and

$$M_b = \{d \in ne_{L^*}(b) \setminus R_b \,;\, \forall\, c \in R_b \;\; d - c \text{ in } L^* \}\,.$$

Then the collection $\mathcal{C}$ of those sets $C \subseteq N \setminus \{a, b\}$ such that $([a, b], C)$ describes a lower neighbour $\mathcal{K} \in o^{\downarrow}(\mathcal{L})$ is as follows.

(a) If a and b belong to the same component of L^* then $P_a = P_b$, $a \notin An_{L^*}(b)$, $b \notin An_{L^*}(a)$ and sets R_a and R_b are both complete in L^*. In particular, both $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(P_b \cup R_b \,|\, \mathit{cliques}(L^*_{M_b}))$ and $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(P_a \cup R_a \,|\, \mathit{cliques}(L^*_{M_a}))$ are non-empty and $\mathcal{C}$ is their union.

(b) If $b \in an_{L^*}(a)$ but a and b belong to different components of L^* then $a \notin An_{L^*}(b)$, $R_a = \emptyset$ and $M_a = ne_{L^*}(a)$. These facts imply that the set $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(P_a \,|\, \mathit{cliques}(L^*_{M_a}))$ is non-empty. Moreover, the set $\mathcal{C}^+_{\mathcal{L}}(a \to b)$ is non-empty iff $b \notin An_{L^*}(a)$ and R_b is complete in L^*. If this is the case then $\mathcal{C}^+_{\mathcal{L}}(a \to b)$ has the form $\mathsf{TUFT}(P_b \cup R_b \,|\, \mathit{cliques}(L^*_{M_b}))$. In particular, $\mathcal{C}$ is the union of these two collections.

(c) If $a \in an_{L^*}(b)$ but a and b belong to different components of L^* then the case (b) can be obtained by interchange of a and b.

(d) If a and b belong to incomparable components of L^* then P_a and P_b may differ, $a \notin An_{L^*}(b)$, $b \notin An_{L^*}(a)$, $R_a = R_b = \emptyset$, $M_a = ne_{L^*}(a)$ and $M_b = ne_{L^*}(b)$. Thus, both $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(P_b \,|\, \mathit{cliques}(L^*_{M_b}))$ and $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(P_a \,|\, \mathit{cliques}(L^*_{M_a}))$ are non-empty and $\mathcal{C}$ is their union.

In particular, the collection $\mathcal{C}$ of respective sets is always non-empty.

The proof is given in Section 5.2 of the Appendix.

Example 3.2 To illustrate the previous analysis consider the essential graphs shown in Figure 6. The case (a) from Corollary 3.1 occurs for the upper left graph. More specifically, one has $P_a = P_b = M_a = \emptyset$, $R_a = \{d, e\}$, $R_b = \{c\}$ and $M_b = \{f\}$. Consequently, $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(\{c\} \,|\, \{f\})$ and $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(\{d, e\} \,|\, \emptyset)$.

The case (d) is illustrated by the upper right graph in Figure 6 in which case $P_a = \{c, d\}$, $P_b = \{d, e\}$ and $M_a = M_b = \emptyset$. Thus, $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(\{d, e\} \,|\, \emptyset)$ and $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(\{c, d\} \,|\, \emptyset)$. The case (b) and the situation $b \in An_{L^*}(a)$

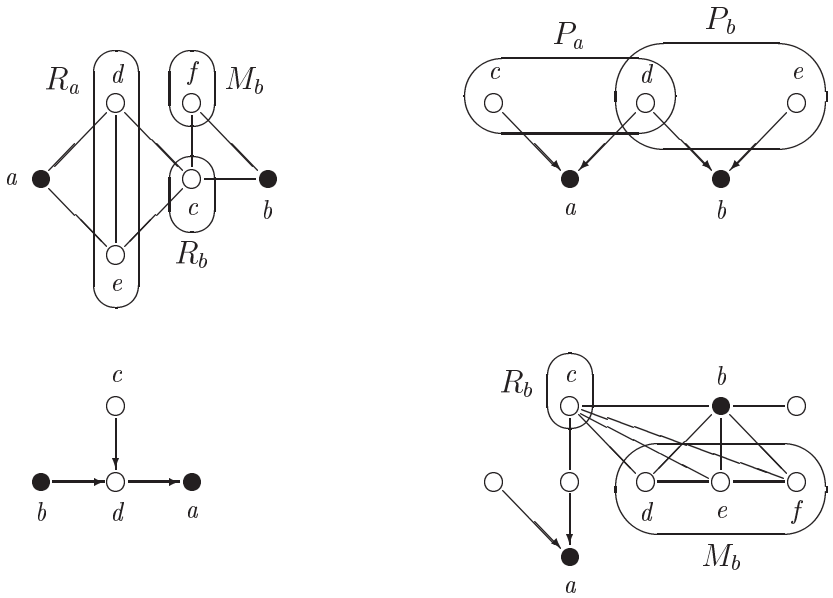

Figure 6: Four essential graphs.

in which case $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \emptyset$ is illustrated by the lower left graph in the figure. One has $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(\{c\} \,|\, \emptyset)$ then. Another subcase of the case (b) was already mentioned in Remark 3.3 - see Figure 5 in which case $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(\{c^2, d, e\} \,|\, \emptyset)$. The case (b) and the situation $\mathcal{C}^+_{\mathcal{L}}(a \to b) \neq \emptyset$ is illustrated by the lower right graph in Figure 6 in which case $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(\{c\} \,|\, \{d, e\}, \{e, f\})$ because $L^*_{M_b}$ has two cliques.

4 Conclusions

In this contribution a characterization of the lower inclusion neighbourhood was presented. Analogous results for the upper inclusion neighbourhood have also been achieved and they were presented in a previous paper [15]. There is internal consistency of both characterizations (see Remark 3.1). This means that the pair $([a, b], C)$, where $[a, b]$ is an unordered pair of nodes and $C \subseteq N \setminus \{a, b\}$ a disjoint set of nodes, which is used in these two papers to characterize uniquely an inclusion neighbour $\mathcal{G}$ of a given equivalence class of Bayesian network $\mathcal{H}$ can be viewed as a natural characteristic of the 'move' between $\mathcal{H}$ and $\mathcal{G}$.

As explained in Section 1 the presented characterization is more elegant than the previous ones. Indeed, Chickering [5] only gave a tentative algorithmic method and Auvrey and Wehenkel [2] characterized every inclusion neighbour by an unordered pair of nodes and by an opaque collection of immoralities,

namely those which are either created or cancelled if an equivalence class is replaced by its inclusion neighbour.

Finally, the characterization presented in this paper has a close connection to an arithmetic method of description of equivalence clases of Bayesian networks (from Chapter 8 of [17]) and leads to conditional independence interpretation of 'moves' in the method of local search.

5 Appendix

5.1 Arithmetic Approach

Special concepts from [17] are recalled in this section. By an *imset* over N will be understood any integer-valued function on $\mathcal{P}(N)$. Arithmetic operations with imsets are defined coordinatewisely. Given $A \subseteq N$ the symbol δ_A will denote a special imset which identifies the set A:

$$\delta_A(B) = \begin{cases} 1 & \text{if } B = A, \\ 0 & \text{if } B \neq A, \end{cases} \qquad \text{for any } B \subseteq N\,.$$

If G is a Bayesian network over N then the *standard imset* for G is given by the formula

$$u_G \;=\; \delta_N - \delta_\emptyset + \sum_{i \in N} \{\, \delta_{pa_G(i)} - \delta_{\{i\} \cup pa_G(i)} \,\}\,. \tag{1}$$

The following result is proved as Consequence 7.1 in [17].

Theorem 5.1 *Bayesian networks K and L over N are equivalent iff $u_K = u_L$.*

In particular, a standard imset $u_\mathcal{G}$ for any equivalence class $\mathcal{G}$ of Bayesian networks can be introduced as the shared standard imset u_G for $G \in \mathcal{G}$. Note that there exists a formula for $u_\mathcal{G}$ in terms of the essential graph which is, however, omitted here. A basic observation is as follows.

Lemma 5.1 Suppose that K, L are Bayesian networks over N and K is made of L by adding of an arrow $a \to b$. Then $u_L - u_K = \delta_C - \delta_{\{a\}\cup C} - \delta_{\{b\}\cup C} + \delta_{\{a,b\}\cup C}$ where $C = pa_L(b)$.

Proof: It follows from the assumption that $pa_K(c) = pa_L(c)$ for every $c \in N \setminus \{b\}$. Thus, after the substitution of (1) into $u_L - u_K$ most of the terms are cancelled. As $pa_L(b) = C$ and $pa_K(b) = \{a\} \cup C$ the result is what Lemma 5.1 says. □

Remark 5.1 Note that the inclusion neighbourhood relation can also be characterized in terms of standard imsets. It is shown in [17] (Consequence 8.4) that, for equivalence classes $\mathcal{K}$ and $\mathcal{L}$, one has $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$ iff $u_\mathcal{L} - u_\mathcal{K}$ is an imset of the form $\delta_C - \delta_{\{a\}\cup C} - \delta_{\{b\}\cup C} + \delta_{\{a,b\}\cup C}$ for some $a, b \in N$, $a \neq b$ and $C \subseteq N \setminus \{a, b\}$. The inclusion ordering can be characterized analogously (see Lemma 8.6 in [17]).

The proof of **Proposition 3.1** can be done on basis of Lemma 5.1 easily.

Proof: If the assumption of Proposition 3.1 are fulfilled then $u_{L_1} = u_{L_2}$ by Theorem 5.1. Moreover, $\mathcal{K}_1 = \mathcal{K}_2$ iff $u_{K_1} = u_{K_2}$ which is equivalent to $u_{L_1} - u_{K_1} = u_{L_2} - u_{K_2}$. However, the latter is equivalent to

$$\delta_{C_1} - \delta_{\{a_1\}\cup C_1} - \delta_{\{b_1\}\cup C_1} + \delta_{\{a_1,b_1\}\cup C_1} = \delta_{C_2} - \delta_{\{a_2\}\cup C_2} - \delta_{\{b_2\}\cup C_2} + \delta_{\{a_2,b_2\}\cup C_2}$$

by Lemma 5.1. This is nothing but the condition that $C_1 = C_2$ and $\{a_1, b_1\} = \{a_2, b_2\}$. □

5.2 Bayesian Network Construction

Given an undirected graph H over N with $|N| = n$, by a *perfect numbering* is meant a total ordering $a_1, \ldots, a_n$ of its nodes such that $ne_H(a_i) \cap \{a_j; j < i\}$ is a complete set in H for every $i = 1, \ldots, n$. It is a well-known fact that every triangulated graph admits a perfect numbering. More specifically, the following statement is true.

Lemma 5.2 Let H be a triangulated graph, $A \subseteq N$ a complete set in H and $a_1, \ldots, a_r$, $r \geq 0$ any ordering of nodes in A. Then there exists a perfect numbering of nodes of H which starts by the sequence $a_1, \ldots, a_r$.

Proof: The required perfect numbering is constructed in the reverse order by means of repeated application of well-known Dirac's lemma - see Lemma 2.9 in [8]. Recall that a node is *simplicial* in an undirected graph H' if $ne_{H'}(a)$ is complete in H'. Dirac's lemma says that if H' is a triangulated graph over N' and N' is not complete in H' then H' has two simplicial nodes $a, b \in N'$ such that $[a, b]$ is not an edge in H'.

Put $H^n = H'$. If N is complete in H^n then the claim of Lemma 5.2 is evident. If this is not the case then the assumption that A is complete in H^n implies by Dirac's lemma that a simplicial node $a_n \in N \setminus A$ exists. Put $H^{n-1} = H^n_{N\setminus\{a_n\}}$ and observe that H^{n-1} is a triangulated graph and A is complete in H^{n-1}. The procedure can be repeated untill the set of nodes of some H^j for $1 \leq j \leq n$ is complete in H^j. □

The preceding lemma makes it possible to construct a Bayesian network on basis of the respective essential graph.

Lemma 5.3 Let H be a chain graph without flags such that, for every component C of H, the graph H_C is triangulated. Moreover, suppose that a perfect numbering of nodes of H_C is prescribed for every component C. Let G be a directed graph made of H in such a way that every line a — b in H is replaced by an arrow $a \to b$ in G provided that a precedes b in the prescribed perfect numbering of the component containing $\{a, b\}$. Then G is a Bayesian network which is equivalent to H.

Proof: To show that G is acyclic a total ordering of nodes consistent with the direction of arrows is constructed. First, since H is a chain graph one can find a chain $C_1, \ldots, C_m$, $m \geq 1$ for H whose blocks are components of H. Second, within each component, nodes are ordered according to the prescribed numbering. Thus, G is acyclic and both G and H are chain graphs without flags with the same underlying graph.

To evidence that they are equivalent one needs to verify that they have the same immoralities (see Section 2.3). By construction, every immorality in H remains in G. Thus, consider an immorality $a \to c \leftarrow b$ in G and show that it is an immorality in H. First, the cases $a \leftarrow c$ in H and $c \to b$ in H are excluded due to the definition of G. Second, the alternative $a - c - b$ in H also cannot occur because the prescribed numberings are perfect: otherwise a, b, c belong to the same component of H and $a, b \in pa_G(c)$ means that $a, b \in ne_H(c)$ and a, b precede c in the prescribed numbering which implies a contradictory conclusion that $[a, b]$ is an edge in H. Third, the alternatives $a \to c - b$ in H and $a - c \leftarrow b$ in H are impossible because H has no flags. This necessitates $a \to c \leftarrow b$ in H which was desired. □

Lemma 5.4 Let H be an essential graph and (a, b) is an ordered pair of nodes in H such that $[a, b]$ is not an edge in H, $b \notin An_H(a)$ and $R \equiv re_H(b|a)$ is complete in H. Put $P = pa_H(b)$ and $M = \{d \in ne_H(b) \setminus R; \forall c \in R \;\; d - c \text{ in } H\}$. Suppose that $X \subseteq M$ is complete in H. Then there exists a Bayesian network G which is equivalent to H, $b \notin an_G(a)$ and $pa_G(b) = P \cup R \cup X$.

Proof: It follows from the assumption that R is complete, the definition of M and the assumption that $X \subseteq M$ is complete that $R \cup X \cup \{b\}$ is a complete set in H. Let C be a component of H containing b. The graph H_C is triangulated (see Section 2.3) which allows one to apply Lemma 5.2 to it. Therefore, there exists a perfect ordering of nodes of H_C in which the nodes from $R \cup X$ precede the node b which precedes the remaining nodes in C. Choose arbitrary perfect orderings for other components of H and apply Lemma 5.3 to get the respective Bayesian network G which is equivalent to H. The fact $pa_G(b) = P \cup R \cup X$ follows from the construction of G. Let us show by contradiction that $b \notin an_G(a)$.
Indeed, otherwise there exists a descending path $\rho : b = d_1, \ldots, d_n = a$, $n \geq 2$ in G. It follows from the fact that H is the essential graph of the equivalence class containing G that ρ is also a descending path in H. As mentioned in Remark 3.3 the assumption $b \notin An_H(a)$ implies that ρ cannot start by an arrow in H for which it has start by a line. Thus, $d_2 \in ne_H(b)$ and one can consider the node d_l which is the last node of ρ in $ne_H(b)$. By the definition of $re_H(b|a)$ one has $d_l \in R = re_H(b|a)$. Thus, $d_l \in R \subseteq pa_G(b)$ implies that $d_l \to b = d_1$ in G which contradicts the fact that G is an acyclic directed graph. □

Now, the proof of **Proposition 3.2** follows.

Proof: I. The first step is to show the necessity of conditions in (i).
Indeed, if $b \in An_{L^*}(a)$ then the respective directed path in L^* is also a directed path in every $L \in \mathcal{L}$ for which $b \in an_L(a)$ for every $L \in \mathcal{L}$ and $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \emptyset$. If $R = re_{L^*}(b|a)$ is not complete in L^* then choose $c^1, c^2 \in R$, $c^1 \neq c^2$ such that $[c^1, c^2]$ is not an edge in

L^*. Like in Remark 3.2 choose respective descending paths $\rho_i : c^i = d^i_1, \ldots, d^i_{n(i)} = a$, $n(i) \geq 2$, $i = 1, 2$ in L^* which cannot be shorthened. Now, consider any $L \in \mathcal{L}$. If $b \to c^i$ in L then observe that ρ_i is a directed path in L (because otherwise an immorality composed of consecutive nodes of the path $b = d^i_0, d^i_1, \ldots, d^i_{n(i)}$ in L exists and this contradicts the fact that L and L^* are equivalent - see Section 2.3). This implies $b \in an_L(a)$. However, if $c^1 \to b \leftarrow c^2$ in L then this is an immorality in L which is not in L^* which again contradicts the fact that they are equivalent. Therefore, there is no $L \in \mathcal{L}$ with $b \notin an_L(a)$ for which $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \emptyset$.

II. The sufficiency of the condition in (i) follows from Lemma 5.4 applied to $H = L^*$ and $X = \emptyset$ which implies that $P \cup R \in \mathcal{C}^+_{\mathcal{L}}(a \to b)$.

III. For (ii) one first needs to show $\mathcal{C}^+_{\mathcal{L}}(a \to b) \subseteq \mathsf{TUFT}(P \cup R \,|\, \mathit{cliques}(L^*_M))$. For this purpose one has to verify the following claims for $C \in \mathcal{C}^+_{\mathcal{L}}(a \to b)$.

(a) $P \cup R \subseteq C$,

(b) $X \equiv C \setminus (P \cup R) \subseteq M$,

(c) X is complete in L^*.

Thus, suppose that $L \in \mathcal{L}$ exists such that $b \notin an_L(a)$ and $C = pa_L(a)$. It follows from the definition of the essential graph that $P = pa_{L^*}(b) \subseteq pa_L(b) = C$. Supposing $c \in R = re_{L^*}(b|a)$ consider a path $\rho : c = d_1 - \ldots - d_k \to \ldots \to d_n = a$, $n \geq 2$, $1 \leq k \leq n$ mentioned in Remark 3.2 which cannot be shorthened. If $b \to c$ in L then ρ is a directed path in L (for the same reason as mentioned in Step I.) which contradicts the fact $b \notin an_L(a)$. Thus, necessarily $b \leftarrow c$ in L which means $c \in pa_L(b) = C$. This concludes the proof of (a).

Now, put $X = C \setminus (P \cup R)$. Clearly, every element of X is a parent of b in L for which it is adjacent to b in L^*. If $d \in X$ then the definition of the essential graph excludes $b \to d$ in L^* for which $X \subseteq ne_{L^*}(b)$ and, therefore, $R \cup X \subseteq ne_{L^*}(b)$. Observe by contradiction that $R \cup X$ is complete in L^*. Indeed, otherwise $c^1, c^2 \in C \setminus P$, $c^1 \neq c^2$ exists such that $[c^1, c^2]$ is not an edge in L^*, and, therefore, in L. This implies that $c^1 \to b \leftarrow c^2$ is an immorality in L which is not in L^* which contradicts the assumption that they are equivalent. Thus, the fact $R \cup X \subseteq ne_{L^*}(b)$ is complete in L^* implies both (c) and (b).

IV. The other inclusion $\mathsf{TUFT}(P \cup R \,|\, \mathit{cliques}(L^*_M)) \subseteq \mathcal{C}^+_{\mathcal{L}}(a \to b)$ in (ii) follows easily from Lemma 5.4 applied to $H = L^*$ and any complete set X in L^*_M. □

Now, the proof of **Corollary 3.1** follows.

Proof: Recall that one needs to characterize $\mathcal{C}^+_{\mathcal{L}}(a \to b) \cup \mathcal{C}^+_{\mathcal{L}}(b \to a)$ using Proposition 3.2. In case (a) the fact that L^* has no flags implies $P_a = P_b$ (see Section 2.3). The main step is to show that both R_a and R_b are complete.

Indeed, suppose for contradiction that R_b is not complete, more specifically that there exist $c^1, c^2 \in re_{L^*}(b|a)$ such that $[c^1, c^2]$ is not an edge in L^*. Like in Remark 3.2 choose paths $\rho_i : c^i = d^i_1 - \ldots - d^i_{n(i)} = a$, $n(i) \geq 2$, $i = 1, 2$ which are outside $ne_{L^*}(b) \setminus \{c^i\}$ and cannot be shorthened. Because a and b belong to the same component of L^* and $b - c^i$ in L^* both paths ρ_i are undirected. The definition of ρ_i implies that neither b belongs to ρ_i nor b is adjacent to a node of ρ_i different from c^i. Let d^1_l be the first node on ρ_1 which is adjacent to a node of ρ_2 and d^2_k be the first node of ρ_2 which is adjacent to d^1_l. It follows from the assumption that c^1 and c^2 are not adjacent that, if $d^1_l = c^1$ then $d^2_k \neq c^2$. This implies that the cycle $b - c^1 = d^1_1 - \ldots - d^1_l - d^2_k -$

$\ldots - d_1^2 = c^2 - b$ has the length at least four. Moreover, by the construction this cycle has no chord which contradicts the fact that the induced subgraph L^*_C for every component C is a triangulated graph - see Section 2.3.
In case (b) there is no descending path from a to b in L^* (otherwise they belong to the same component of L^*) which implies both $a \notin An_{L^*}(b)$ and $R_a = \emptyset$. Similar argument gives $R_b = \emptyset$ in case (c) and can be repeated in case (d). □

References

[1] Andersson S. A., Madigan D. and Perlman, M. D. (1997), A characterization of Markov equivalence classes for acyclic digraphs, *Annals of Statistics* 25, 505–541.

[2] Auvray V. and Wehenkel L. (2002), On the construction of the inclusion boundary neighbourhood for Markov equivalence clases of Bayesian network structures, In Uncertainty in Artificial Intelligence 18 (Darwiche A. and Friedman N. eds.), Morgan Kaufmann, 26–35.

[3] Bouckaert R. R. (1995), Bayesian belief networks: from construction to inference, PhD thesis, University of Utrecht.

[4] Castelo R. (2002), The discrete acyclic digraph Markov model in data mining, PhD thesis, University of Utrecht.

[5] Chickering D. M. (2002), Optimal structure identification with greedy search, *Journal of Machine Learning Research* 3, 507-554.

[6] Frydenberg M. (1990), The chain graph Markov property, *Scandinavian Journal of Statistics* 17, 333–353.

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

[8] Lauritzen S. L. (1996), *Graphical Models*, Clarendon Press, Oxford.

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

[10] Meek C. (1997), Graphical models, selecting causal and statistical models, PhD thesis, Carnegie Mellon University.

[11] Pearl J. (1988), *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo.

[12] Roverato A. (2004), A unified approach to the characterisation of equivalence classes of DAGs, chain graphs with no flags and chain graphs, technical report, Dipartimento di Scienze Sociali Cognitive e Quantitative, University of Modena and Reggio Emilia, submitted to *Annals of Statistics*.

[13] Studený M. (1997), A recovery algorithm for chain graphs, *International Journal of Approximate Reasoning* 17, 265–293.

[14] Studený M. and Bouckaert R. R. (1998), On chain graph models for description of conditional independence structures, *Annals of Statistics* 26, 1434–1495.

[15] Studený M. (2003), Characterization of inclusion neighbourhood in terms of the essential graph: upper neighbours, in Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Nielsen D. H. and Zhang N. L. eds.), Lecture Notes in AI 2711, Springer Verlag, 161-172.

[16] Studený M. (2004), Characterization of essential graphs by means of an operation of legal merging of components, to appear in *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems.*

[17] Studený M. (2004), *On probabilistic conditional independence structures*, a research monograph, to be published by *Springer Verlag.*

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