Characterization of Inclusion Neighbourhood in Terms of the Essential Graph

Milan Studený [1]

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic

Abstract

The question of efficient characterization of inclusion neighbourhood is crucial in some methods for learning (equivalence classes of) Bayesian networks. In this paper, neighbouring equivalence classes of a given equivalence class of Bayesian networks are characterized efficiently in terms of the respective essential graph. One can distinguish two kinds of inclusion neighbours: upper and lower ones. This paper reveals the hidded internal structure of both parts of the 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 $C \subseteq N \setminus \{a,b\}$ is a disjoint set of nodes in the essential graph. Upper neighbours correspond to edges in the essential graph, while lower neighbours correspond to pairs of nodes that are not edges in the essential graph. Given a pair $[a,b]$ of distinct nodes in the essential graph, the class of those sets C that $([a,b],C)$ encodes an inclusion neighbour is charecterized. The class has a special form; it is uniquely determined by certain distinguished sets. These distinguished sets of the class can be read directly from the essential graph.

Key words: learning Bayesian networks, inclusion neighbourhood, essential graph.
1991 MSC: 68T30, 68R10.

Email address: `studeny@utia.cas.cz` (Milan Studený).
[1] This research has been supported by the grants GAČR n. 201/04/0393 and GAAVČR n. A1075104.

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] or “score metric” [4]. A 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 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 an operational point of view, the goal is to learn an equivalence class of Bayesian networks, that is, a class of acyclic directed graphs. Note for explanation that in this paper the attention is restricted to conditional independence interpretation of acyclic directed graphs. This interpretation differs from causal interpretation of these graphs [14] in which case the equivalence relationship does not make sense.

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 a 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 [5,8].

Algorithms of this kind can also be classified according to the way to represent equivalence classes of networks. In some algorithms, an equivalence class is represented by any of its members which may, however, result in computational inefficiency. This is because some of the equivalence classes can be quite big and algorithms can stick in them. 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” [21], “maximally oriented graph for a pattern” [10] 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" [8,2], 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 [11] about a 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 paper 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 the 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 were proposed there. A legality test is able to decide whether the corresponding graphical operation leads to a real neighbouring equivalence class. One of the operations and the respective legality test are aimed to obtain upper neighbours; the other operation and the other test are designed to generate 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 method is that it has no sensible guidance. It is more likely a blind automatic procedure ignoring any 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 objects are still unnecessarily complicated which means that their characterization of the upper inclusion neighbourhood is too awkward. In particular, the second drawback is not removed by their approach because their approach does not allow one to make out the internal structure of the inclusion neighbourhood. Moreover, their characterization is incomplete: only partial characterization of lower neighbours is given in [2].

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 described. Note that the results of the paper were already presented at conferences [17,18].

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 in the essential graph for upper neighbours and $[a, b]$ is a pair of nodes which is not an edge in the essential graph for lower neighbours. The first new observation 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 the respective inclusion neighbours has a special form.

A complete analysis of the upper inclusion neighbourhood is presented. In this case, the collection of sets C for a given edge $[a, b]$ has the form of a *tuft*. This means that it is 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 of sets is completely described by its least set and by the list of its maximal sets. Given an essential graph G^* and an edge $[a, b]$ in G^* the least and maximal sets of the respective tuft of sets are characterized directly in terms of G^*.

A further result of the paper is an analogous description of the lower inclusion neighbourhood. 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. Given an essential graph G^* and a pair of its distinct nodes $[a, b]$ which is not an edge in G^* the least and maximal sets of the respective tufts are also characterized directly in terms of G^*.

Note that the description of inclusion neighbours by means of pairs $([a, b], C)$ where $C \subseteq N \setminus \{a, b\}$ and the way to introduce these pairs used in this paper is not incidental. An interesting fact is that, from a certain perspective which is explained in detail in Chapter 8 of [20], the pair $([a, b], C)$ has a close relation to conditional independence interpretation of the 'move' from the considered equivalence class towards its respective inclusion neighbour.

Basic concepts and facts, including a new special concept of a tuft, are recalled in Section 2. Section 3 is devoted to the upper inclusion neighbourhood, Section 4 to the lower inclusion neighbourhood. The proofs given in the Appendix combine the ideas motivated by an arithmetic approach to the description of Bayesian network models from [20] 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 non-empty 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 or $b \to a$ in H (note that the role of a and b is interchangeable here). 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 $\forall\, a, b \in K$ $a \neq b$ one has $a - b$ in H. A maximal complete set in H (with respect to set inclusion) will be called a *clique* (of H). The collection of all cliques of H will be denoted by *cliques*(H).

A set $C \subseteq N$ is *connected* in H if, for every $a, b \in C$, 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, $\mathcal{A}(H) = \emptyset$. 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 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 $c_i - c_j$ in H 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 $i = 1, \ldots, n$. A 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 [15]. 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,

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

is disjoint with C if C is connected. The set

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

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

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

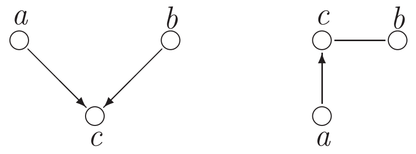

Fig. 1. An immorality and a flag.

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 a node 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 a node b in H will be denoted by $An_H(b)$. Observe that $An_H(b) \subseteq an_H(b)$ and if H is a chain graph then $a \in an_H(b) \setminus An_H(b)$ for any node a which belongs to the same connectivity component as b.

Two types of configurations of three nodes in a graph H will play an important role in the paper. 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. A *flag* in 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.

2.2 Bayesian Networks and Their Equivalence

A *Bayesian network* is a certain statistical model, that is, a class of (multidimensional probability) distributions, associated with 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 in terms of conditional independence restrictions, using the d-separation criterion from [12] or using the moralization criterion from [9], 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.

Remark 2.1 Some authors [7] define a Bayesian network as a pair consisting of an acyclic directed graph and a (discrete) probability distribution which factorizes according to the graph. This paper deals with *learning structure* of a Bayesian network given by the graph. Therefore, a Bayesian network is

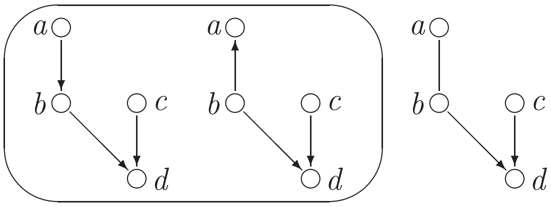

Fig. 2. An equivalence class of Bayesian networks and the respective essential graph.

understood as a class of probability distributions having the same structure of this kind.

An important concept is the concept of *equivalence of Bayesian networks*. Two Bayesian networks G_1 and G_2 are considered to be Markov 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 an 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 [21] gave a direct graphical characterization of equivalent Bayesian networks which can be used as its formal definition here. 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, an unordered 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 implies that 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

- $a \to b$ 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 [19] and, independently, in [13]. 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 in [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. Note that any 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 $a \in C$. Of course, every Bayesian network is a chain graph without flags.

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 [9], alternatively the c-separation criterion from [16], 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 Lemma 2 in [19]. Thus, Frydenberg's result implies the result by Verma and Pearl [21]. Another basic fact about the essential graph G^* of an equivalence class of Bayesian networks $\mathcal{G}$ 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 Corollary 2 in [19].

Remark 2.2 Note that there are other possible ways to represent an equivalence class $\mathcal{G}$ of Bayesian networks. One of them is to use the *largest chain graph* of the collection of chain graphs that are equivalent to (any) $G \in \mathcal{G}$ – see [6] for this concept and [15] for further details. Another alternative is brought by the arithmetic approach presented in Section 8.4 of [20] which offers the concept of a *standard imset* – see Section A.1 in the Appendix.

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 $\mathcal{I}(\mathcal{K}) \subseteq \mathcal{I}(\mathcal{L})$ 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 an *upper neighbour* of $\mathcal{K}$ and $\mathcal{K}$ will be called a *lower neighbour* of $\mathcal{L}$. The *inclusion neighbourhood* of an equivalence class is the collection its upper and lower neighbours.

The 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 [20].

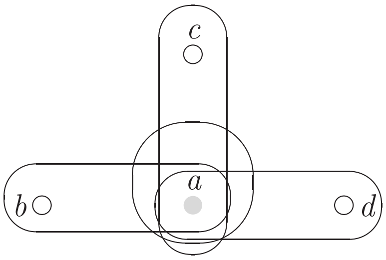

Fig. 3. A tuft of sets $\mathsf{TUFT}(\{a\}|\{b\},\{c\},\{d\})$.

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 the addition of (exactly) one arrow, which is another way of saying that L is made of K by the removal of one arrow.

Remark 2.3 The relation $\mathcal{I}(\mathcal{K}) \subseteq \mathcal{I}(\mathcal{L})$ corresponds to the situation that the statistical model associated with $K \in \mathcal{K}$ contains the statistical model associated with $L \in \mathcal{L}$. The networks in $\mathcal{K}$ have more edges than networks in $\mathcal{L}$ then. The reader may ask why $\mathcal{L}$ is supposed to be 'above' $\mathcal{K}$ in this paper (and not conversely). The terminology used in this paper simply emphasizes the conditional independence interpretation of considered statistical models which is in the center of the author's interests – for more detailed justification and explanation of a wider arithmetic perspective see Remark 8.10 in [20].

2.5 Tufts 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 the following special notation:

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

Evidently, $\mathcal{T} = \mathsf{TUFT}(A|\mathcal{B})$ is a tuft of sets 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.

Remark 2.4 The concept of a tuft appears to be suitable in the context of the problem considered in this paper. Since the author has not been aware of any standard name for a collection of sets satisfying the required conditions a special short word has been proposed to name it. The word "tuft" hopefully indicates what is substantial: the collection has one root and several leaves which originate from the root.

3 Upper Inclusion Neighbourhood

3.1 Description of Upper Neighbours

By the *upper neighbourhood* of an equivalence class $\mathcal{K}$ of Bayesian networks we understand the collection $o^{\uparrow}(\mathcal{K})$ of those equivalence classes $\mathcal{L}$ for which $\mathcal{I}(\mathcal{K}) \sqsubset \mathcal{I}(\mathcal{L})$. It follows from Lemma 2.1 that each $K \in \mathcal{K}$ and each edge in $\mathcal{K}$ define together an element of $o^{\uparrow}(\mathcal{K})$ and every element of $o^{\uparrow}(\mathcal{K})$ is obtained in this way. Thus, the upper neighbourhood $o^{\uparrow}(\mathcal{K})$ is, in fact, described in terms of elements of $\mathcal{K}$. Nevertheless, the above described correspondence is not a one-to-one mapping because different elements of $\mathcal{K}$ may yield the same neighbouring class $\mathcal{L}$.

One the other hand, every neighbouring class is uniquely characterized 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 $\mathcal{K}$ be an equivalence class of Bayesian networks over N, $\mathcal{L} \in o^{\uparrow}(\mathcal{K})$. Choose $K \in \mathcal{K}$ and $L \in \mathcal{L}$ such that L is obtained from K by the removal of an arrow $a \to b$ in K. Then $\mathcal{L}$ will be described by the pair $([a,b],C)$ where $C = pa_K(b) \setminus \{a\}$.

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

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

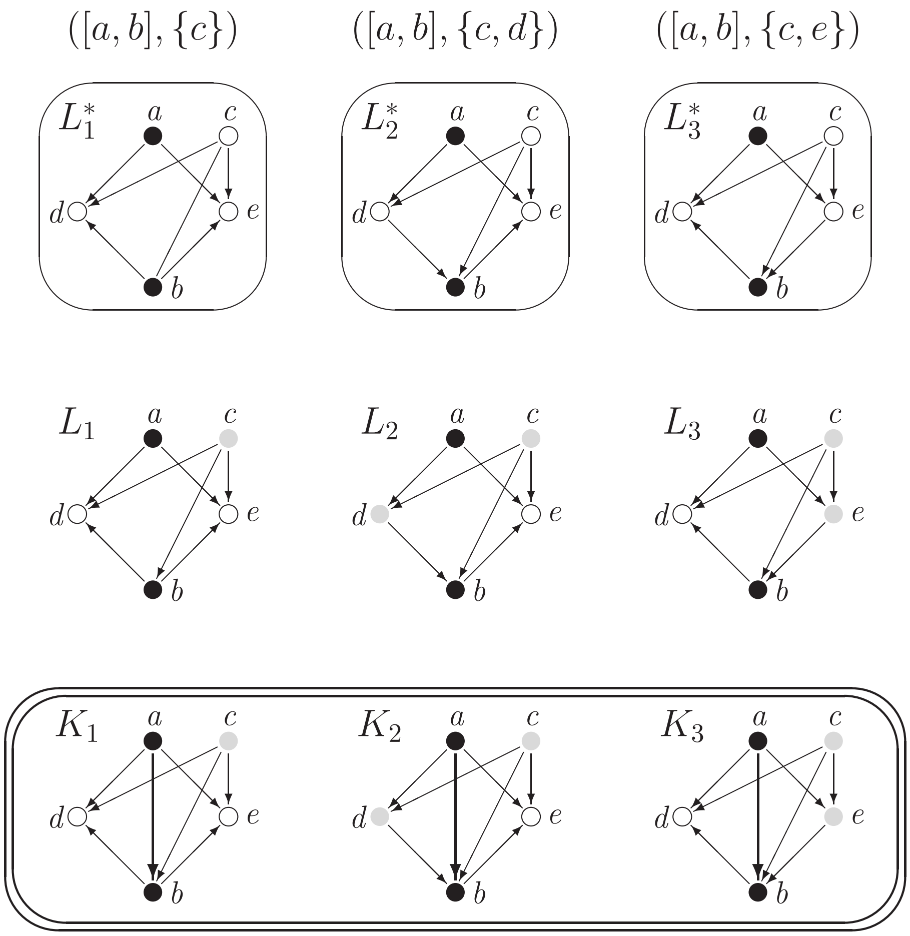

Fig. 4. An equivalence class of Bayesian networks and some of its upper neighbours.

The proof of Proposition 3.1, which is given in Section A.1 of the Appendix, is based on a special arithmetic characterization of equivalence of Bayesian networks. Note that one can perhaps prove this result using purely graphical tools, but the given proof is more elegant.

Example 3.1 To illustrate the concepts introduced above let us consider an equivalence class $\mathcal{K}$ of Bayesian networks over $N = \{a, b, c, d, e\}$ shown in the lower layer of Figure 4. For every $i = 1, 2, 3$, an acyclic directed graph L_i is obtained from $K_i \in \mathcal{K}$ by the removal of the arrow $a \to b$ (see the medium layer of the figure). In this example, each of K_i, $i = 1, 2, 3$, establishes a different neighbouring class $\mathcal{L}_i \in o^{\uparrow}(\mathcal{K})$ – the respective essential graphs are in the upper layer of Figure 4. As $pa_{K_2}(b) \setminus \{a\} = \{c, d\}$ the equivalence class containing L_2 is characterized by the pair $([a, b], \{c, d\})$.

Remark 3.1 The pair $([a, b], C)$ which describes uniquely an upper inclusion neighbour $\mathcal{L} \in o^{\uparrow}(\mathcal{K})$ was introduced in terms of individual networks from $\mathcal{K}$ and $\mathcal{L}$. If $\mathcal{K}$ and $\mathcal{L}$ are represented by the respective essential graphs K^* and L^* then $[a, b]$ is simply the edge of K^* which is not an edge in L^*. The question of how to define C in terms of (the pair) K^* and L^* have not been examined so far by the author. On the other hand, the pair $([a,b],C)$ is obtained immediately if $\mathcal{K}$ and $\mathcal{L}$ are represented by means of their standard imsets – see Section A.1 in the Appendix.

3.2 Characterization of Upper Neighbourhood

Given an equivalence class $\mathcal{K}$ of Bayesian networks the next step is to characterize those pairs $([a,b],C)$ which encode elements of $o^{\uparrow}(\mathcal{K})$. In this section, this task is answered for a fixed unordered pair of distinct nodes $[a,b]$. To this end we first consider an ordered pair of distinct nodes (a,b) and introduce a special collection of subsets of $N \setminus \{a,b\}$:

$$\mathcal{C}^{-}_{\mathcal{K}}(a \to b) \;=\; \{\, C\,;\; \exists\, K \in \mathcal{K} \text{ such that } a \to b \text{ in } K \text{ and } C = pa_K(b) \setminus \{a\}\,\}\,.$$

It follows from what it says in Section 3.1 that $\mathcal{C}^{-}_{\mathcal{K}}(a \to b) \cup \mathcal{C}^{-}_{\mathcal{K}}(b \to a)$ is the class of sets which has to be characterized. Nevertheless, the reader can learn in the sequel (see the proof of Corollary 3.1) that only one of the collections $\mathcal{C}^{-}_{\mathcal{K}}(a \to b)$ and $\mathcal{C}^{-}_{\mathcal{K}}(b \to a)$ actually forms their union. Therefore, given an ordered pair (a,b), one needs to find out whether the class $\mathcal{C}^{-}_{\mathcal{K}}(a \to b)$ is non-empty and to describe that collection of sets then.

Proposition 3.2 *Let $\mathcal{K}$ be an equivalence class of Bayesian networks, K^* the essential graph of $\mathcal{K}$ and (a,b) an ordered pair of distinct nodes of K^*. Put $P = pa_{K^*}(b) \setminus \{a\}$ and $M = \{c \in ne_{K^*}(b) \setminus \{a\}\,;\; [a,c] \text{ is an edge in } K^*\}$. Then the following statements are true.*

(i) *$\mathcal{C}^{-}_{\mathcal{K}}(a \to b) \neq \emptyset$ iff $[a,b]$ is an edge in K^* and $b \notin pa_{K^*}(a)$, that is, either $a \to b$ in K^* or $a - b$ in K^*.*
(ii) *If $a \to b$ in K^* or $a - b$ in K^* then $\mathcal{C}^{-}_{\mathcal{K}}(a \to b) = \mathsf{TUFT}(P|cliques(K^*_M))$ where $cliques(K^*_{\emptyset}) = \{\emptyset\}$ by a convention.*

The proof is given in Section A.3 of the Appendix.

Corollary 3.1 Let $\mathcal{K}$ be an equivalence class of Bayesian networks over N, K^* the essential graph of $\mathcal{K}$ and $[a,b]$ an edge in K^*. Then the collection of those sets $C \subseteq N \setminus \{a,b\}$ such that $([a,b],C)$ describes an upper neighbour $\mathcal{L} \in o^{\uparrow}(\mathcal{K})$ is a tuft of sets $\mathsf{TUFT}(P|cliques(K^*_M))$ where

(a) if $a \to b$ in K^* then $P = pa_{K^*}(b) \setminus \{a\}$ and $M = ne_{K^*}(b)$,
(b) if $a \leftarrow b$ in K^* then $P = pa_{K^*}(a) \setminus \{b\}$ and $M = ne_{K^*}(a)$,
(c) if $a - b$ in K^* then $P = pa_{K^*}(b) = pa_{K^*}(a)$ and $M = ne_{K^*}(a) \cap ne_{K^*}(b)$.

Proof: Recall that one needs to characterize $\mathcal{C}^{-}_{\mathcal{K}}(a \to b) \cup \mathcal{C}^{-}_{\mathcal{K}}(b \to a)$ with the aid of Proposition 3.2. In the case (a) observe that $\mathcal{C}^{-}_{\mathcal{K}}(b \to a) = \emptyset$ by (i).

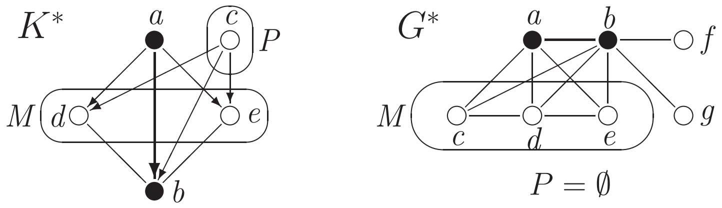

Fig. 5. Two essential graphs.

As concerns $\mathcal{C}^-_{\mathcal{K}}(a \to b)$, the fact that K^* has no flags implies $M = ne_{K^*}(b)$. The case (b) is analogous to the case (a). In the case (c) realize that the fact that K^* has no flags implies $pa_{K^*}(b) = pa_{K^*}(a)$, which is the set P from Proposition 3.2 both for the pair (a, b) and for the pair (b, a). The fact that K^* is a chain graph implies that the set $ne_{K^*}(a) \cap ne_{K^*}(b)$ coincides with the set M from Proposition 3.2 in both cases. Hence, one has $\mathcal{C}^-_{\mathcal{K}}(a \to b) = \mathsf{TUFT}(P, \mathit{cliques}(K^*_M)) = \mathcal{C}^-_{\mathcal{K}}(b \to a) \neq \emptyset$. $\square$

Example 3.2 To illustrate the previous result consider the essential graphs shown in Figure 5. The case (a) from Corollary 3.1 occurs for the graph K^* in the left-hand picture of the figure. More specifically, one has $P = \{c\}$, $M = \{d, e\}$ and $\mathit{cliques}(K^*_M) = \{\,\{d\}, \{e\}\,\}$. Thus, the class of sets $C \subseteq N \setminus \{a, b\}$ such that $([a, b], C)$ describes an upper neighbour of the respective equivalence class is $\mathsf{TUFT}(\{c\}|\{d\}, \{e\})$, that is, the class which involves three sets: $\{c\}$, $\{c, d\}$ and $\{c, e\}$. Indeed, it was shown in Example 3.1 that those upper neighbours of the respective equivalence class which correspond to the 'removal' of $a \to b$ are characterized by pairs $([a, b], \{c\})$, $([a, b], \{c, d\})$ and $([a, b], \{c, e\})$.

If the graph G^* in the right-hand picture of Figure 5 is considered then the case (c) from Corollary 3.1 occurs. One has $P = \emptyset$, $M = \{c, d, e\}$ and $\mathit{cliques}(G^*_M) = \{\,\{c, d\}, \{d, e\}\,\}$. The respective class $\mathsf{TUFT}(\emptyset|\{c, d\}, \{d, e\})$ has six sets, namely $\emptyset$, $\{c\}$, $\{d\}$, $\{e\}$, $\{c, d\}$ and $\{d, e\}$.

4 Lower Inclusion Neighbourhood

4.1 Description of Lower Neighbours

The *lower neighbourhood* of an equivalence class $\mathcal{L}$ of Bayesian networks is 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 $\mathcal{L}$. However, 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 $\mathcal{K}$.

Every neighbouring class can be uniquely described by a 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 the addition 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 correct 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 4.1 *Let $\mathcal{L}$ be an equivalence class of Bayesian networks and $\mathcal{K}_1, \mathcal{K}_2 \in o^{\downarrow}(\mathcal{L})$. 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 the addition 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 is given in Section A.1 of the Appendix.

Example 4.1 To illustrate the above result let us consider the essential graph L^* shown at the top of Figure 6 (in a single oval). The respective equivalence class $\mathcal{L}$ of Bayesian networks over $N = \{a,b,c,d,e\}$ is shown below it (in a double oval). The figure describes a part of its lower neighbourhood, namely those neighbours which correspond to the addition of (an edge) $[a,b]$. The third layer of Figure 6 contains acyclic directed graphs obtained from graphs in $\mathcal{L}$ in that way: K_1 is obtained from L_1 by the addition of $a \to b$, K_2 is obtained from L_1 by the addition of $b \to a$, K_3 is obtained from L_2 by the addition of $a \to b$ and K_4 is obtained from L_2 by the addition of $b \to a$. 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.

Remark 4.1 Note that above mentioned way to describe lower inclusion neighbours is consistent with the way which was used in Section 3.1 to describe 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 $\mathcal{K}$.

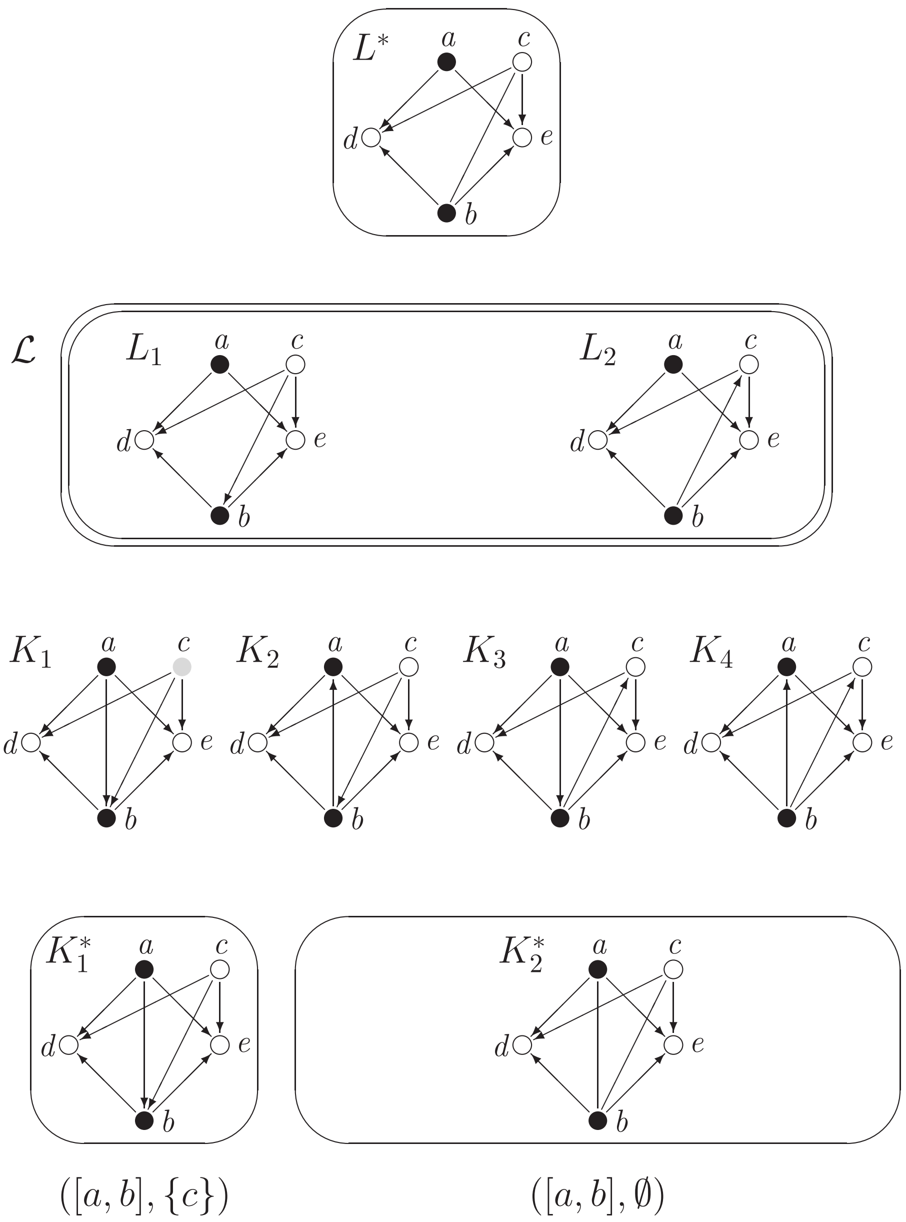

Fig. 6. Lower inclusion neighbours – an example.

4.2 Characterization of Lower Neighbourhood

Given an equivalence class $\mathcal{L}$ of Bayesian networks the aim is to characterize those pairs $([a,b],C)$ which define elements of $o^{\downarrow}(\mathcal{L})$. 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 the addition of an arrow $a \to b$. For this purpose we put

$$\mathcal{C}^+_{\mathcal{L}}(a \to b) \;=\; \{\, C\,;\; \exists\, L \in \mathcal{L} \text{ such that } b \not\in an_L(a) \text{ and } C = pa_L(b)\,\}$$

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

It follows from what it says in Section 4.1 that $\mathcal{C}^+_{\mathcal{L}}(a \to b) \cup \mathcal{C}^+_{\mathcal{L}}(b \to a)$ is the class of sets which should be characterized. Therefore, given an ordered pair (a,b), one first needs to find out in which case $\mathcal{C}^+_{\mathcal{L}}(a \to b)$ is non-empty and describe that collection then. The following concept is useful for that purpose.

Definition 4.1 Let H be a chain graph without flags and (a,b) an ordered pair of distinct 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 exists a descending path in H from c to a which is outside $ne_H(b) \setminus \{c\}$.

Remark 4.2 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 $d_i \not\in ne_H(b)$ for $i = 2, \ldots, k$. Indeed, if there exist a path mentioned in Definition 4.1 then consider a path of this kind which cannot be shortened. Because H has no flags it necessarily has the form mentioned in this remark. This condition is maybe more suitable for verification.

Example 4.2 To illustrate the concept of a relative neighbour consider the essential graph L^* shown in Figure 7. In this case, the set of relative neighbours of b with respect to a coincides with the set of its neighbours. Indeed, for $i = 1, 2$, there exists a descending path from c^i to a in L^* which is outside the rest of $ne_{L^*}(b)$.

Proposition 4.2 *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 $P = \mathrm{pa}_{L^*}(b)$, $R = \mathrm{re}_{L^*}(b|a)$ and introduce $M = \{d \in \mathrm{ne}_{L^*}(b) \setminus R\,;\;\; d - c \text{ in } L^* \text{ for every } c \in R\,\}$. 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, $b \not\in \mathrm{An}_{L^*}(a)$ 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 a convention.*

The proof is given in Section A.4 of the Appendix.

Remark 4.3 The condition $b \not\in An_{L^*}(a)$ from (i) in Proposition 4.2 is equivalent to the requirement that there is no descending path in L^* from b to a

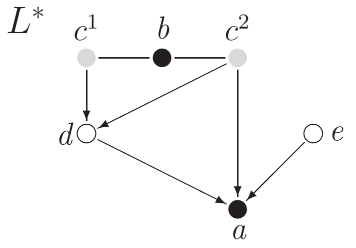

Fig. 7. An essential graph with an incomplete set of relative neighbours.

which starts by an arrow. Indeed, if there exists a path like that then consider one which cannot be shortened. 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 7 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 the lower inclusion neighbourhood which corresponds to a given unordered pair of nodes $[a, b]$.

Corollary 4.1 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 $\mathcal{K} \in o^{\downarrow}(\mathcal{L})$ can be obtained 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 the sets R_a and R_b are both complete in L^*. In particular, both the class $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(P_b \cup R_b \,|\, cliques(L^*_{M_b}))$ and the class $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(P_a \cup R_a \,|\, cliques(L^*_{M_a}))$ are non-empty and the collection $\mathcal{C}$ 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 collection $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(P_a \,|\, cliques(L^*_{M_a}))$ 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^*. In this case $\mathcal{C}^+_{\mathcal{L}}(a \to b)$ has the form $\mathsf{TUFT}(P_b \cup R_b \,|\, cliques(L^*_{M_b}))$. 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 the respective sets is always non-empty.

Proof: Recall that one needs to characterize $\mathcal{C}^+_{\mathcal{L}}(a \to b) \cup \mathcal{C}^+_{\mathcal{L}}(b \to a)$ by means of Proposition 4.2. In the 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 4.2 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 shortened. 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 which is 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 $b — c^1 = d^1_1 — \ldots — d^1_l — d^2_k — \ldots — d^2_1 = c^2 — b$ 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 the 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 $R_a = re_{L^*}(a|b) = \emptyset$. This implies $M_a = ne_{L^*}(a)$. A similar argument gives $R_b = \emptyset$ in the case (c) and can also be repeated in the case (d). $\square$

Example 4.3 To illustrate the previous analysis consider the essential graphs shown in Figure 8. The case (a) from Corollary 4.1 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 8 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 $b \in An_{L^*}(a)$ in which case $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \emptyset$ is illustrated by the lower left graph in the figure. One has $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(\{d\} \,|\, \emptyset)$ then.

Another subcase of the case (b), namely if R_b is not complete, was already mentioned in Remark 4.3 – see Figure 7 in which case $\mathcal{C}^+_{\mathcal{L}}(a \to b) = \emptyset$ and $\mathcal{C}^+_{\mathcal{L}}(b \to a) = \mathsf{TUFT}(\{c^2, d, e\} \,|\, \emptyset)$.

The case (b) and the situation $\mathcal{C}^+_{\mathcal{L}}(a \to b) \neq \emptyset$ is illustrated by the lower right graph in Figure 8 in which case $R_b = \{c\}$ and $M_b = \{d, e, f\}$. Thus,

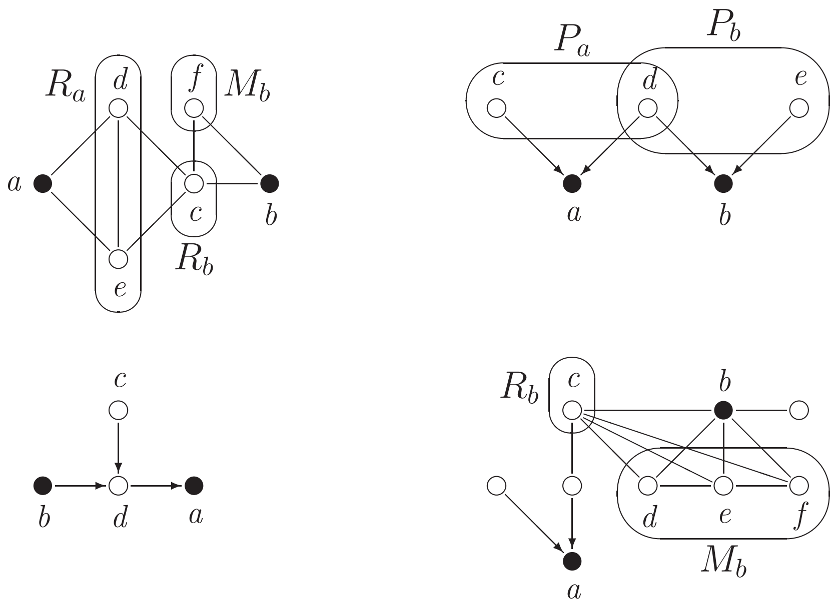

Fig. 8. Four essential graphs.

$\mathcal{C}^+_{\mathcal{L}}(a \to b) = \mathsf{TUFT}(\{c\} \,|\, \{d, e\}, \{e, f\})$ because $L^*_{M_b}$ has two cliques.

5 Conclusions

In this paper both the characterization of the upper inclusion neighbourhood and the characterization of the lower inclusion neighbourhood were presented. There is internal consistency of both characterizations (see Remark 4.1). This implies 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 to characterize uniquely an inclusion neighbour $\mathcal{G}$ of a given equivalence class $\mathcal{H}$ can be viewed as a natural characteristic of the 'move' between $\mathcal{H}$ and $\mathcal{G}$. The presented characterization also has a close connection to an arithmetic method for describing equivalence classes of Bayesian networks developed in Chapter 8 of [20] and leads to conditional independence interpretation of 'moves' in the method of local search. More specifically, a pair $([a, b], C)$ corresponds to an elementary conditional independence statement $a \perp\!\!\!\perp b \,|\, C$ and the respective change in the value of a quality criterion corresponds to this interpretation – for details see Chapter 8 in [20].

As mentioned in Section 1 the presented characterization of the inclusion neighbourhood is qualitatively different from the previous ones, offered by other authors. 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. As concerns the upper inclusion characterization, the result in Corollary 3.1 is essentially equivalent to those in [2,5] but it is more elegant and specific.

Essential improvement is brought by the lower inclusion characterization. Auvrey and Wehenkel only gave an incomplete partial characterization of the lower neighbourhood and Chickering offered a complex graphical procedure to search for 'legal' moves towards lower neighbours. His algorithm can indeed enter 'blind alleys'. To illustate this claim note that, if an arrow $a \to b$ is going to be added, then Chickering's algorithm may try all those moves which, in our description, correspond to sets C which are supersets of $pa(b)$ satisfying $C \setminus pa(b) \subseteq ne(b) \setminus ne(a)$. However, it is shown in Proposition 4.2 that one should only consider certain supersets of $pa(b) \cup re(b|a)$. For example, if one considers the lower right graph in Figure 8 then Chickering's procedure must repeat the respective legality test for each of 16 subsets of the set $\{c, d, e, f\}$ while our characterization gives directly all six 'legal' moves mentioned in Example 4.3. Therefore, the observations made in Section 4.2 can make some computational procedures even more effective.

The unique description of inclusion neighbourhood in terms of pairs $([a, b], C)$ also implies that the maximal number of inclusion neighbours $n \cdot (n-1) \cdot 2^{n-3}$, where $n = |N|$, is achieved for the essential graph which has N as the only clique. On the other hand, the minimal number of inclusion neighbours is $n \cdot (n - 1) \cdot 2^{-1}$ which is achieved for the essential graph over N without edges. These observations lead to a hypothesis that the number of inclusion neighbours increases with the amount of edges. The observations also indicate that, from the point of view of computational efficiency, it does not seem to be a good idea to start a local search learning procedure with a graph which has a high number of edges.

In the author's view, two main contributions of the paper are as follows:

- the hidden internal structure of the inclusion neighbourhood is revealed,
- the presented inclusion neighbourhood description is more detailed than the previous ones, nearly ready to implement.

A Appendix

A.1 Arithmetic Approach

Special concepts from [20] are recalled in this section. An *imset* over N is an integer-valued function on the power set $\mathcal{P}(N)$. 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

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

Remark A.1 To explain a wider perspective note that a special class of *structural imsets* was proposed in [20] to describe all possible conditional independence structures induced by discrete probability distributions. Informally said, a structural imset encodes a certain factorization formula for a probability distribution which is equivalent to a collection of conditional independence restrictions. However, to describe Bayesian network structures it is suitable to consider only a certain subclass of the class of structural imsets, namely the class of standard imsets.

The following result is proved as Corollary 7.1 in [20].

Theorem A.1 *Bayesian networks K and L over N are equivalent if and only if $u_K = u_L$.*

In particular, the 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 A.1 Suppose that K, L are Bayesian networks over N and L is made of K by the removal of an arrow $a \to b$ in K (that is, K is made of L by the addition 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 $C = pa_K(b) \setminus \{a\}$, that is, $C = pa_L(b)$.

Proof: It follows from the assumption that $pa_K(c) = pa_L(c)$ for every node $c \in N \setminus \{b\}$. Thus, after the substitution of (A.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 A.1 says. $\square$

Remark A.2 Note that the inclusion neighbourhood relation can also be characterized in terms of standard imsets. It is shown in [20] (Corollary 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 [20].

The proof of **Proposition 3.1** can be done on basis of Lemma A.1 easily.

Proof: The assumptions of Proposition 3.1 imply $K_1, K_2 \in \mathcal{K}$ which means they are equivalent. Hence, $u_{K_1} = u_{K_2}$ by Theorem A.1. By the same argument, $\mathcal{L}_1 = \mathcal{L}_2$ iff $u_{L_1} = u_{L_2}$ which is equivalent to $u_{L_1} - u_{K_1} = u_{L_2} - u_{K_2}$. However, the latter condition 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 A.1. This is nothing else than the condition that $C_1 = C_2$ and $\{a_1, b_1\} = \{a_2, b_2\}$. $\square$

The proof of **Proposition 4.1** is similar.

Proof: The assumptions of Proposition 4.1 say that L_1 and L_2 are equivalent, that is, $u_{L_1} = u_{L_2}$ by Theorem A.1. Moreover, $\mathcal{K}_1 = \mathcal{K}_2$ iff $u_{K_1} = u_{K_2}$, that is, $u_{L_1} - u_{K_1} = u_{L_2} - u_{K_2}$. By Lemma A.1, this can be written as follows:

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

that is, the condition that $C_1 = C_2$ and $\{a_1, b_1\} = \{a_2, b_2\}$. $\square$

A.2 Bayesian Network Construction

Given an undirected graph H over N with $|N| = n$, a *perfect numbering* is 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 A.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 [9]. 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 A.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 until the set of nodes of some H^j for $1 \leq j \leq n$ is complete in H^j. $\square$

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

Lemma A.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 $a \rightarrow b$ 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 $a \rightarrow c \leftarrow b$ in G and show that it is an immorality in H. First, the cases $a \leftarrow c$ in H and $c \rightarrow b$ 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 $a \rightarrow c - b$ in H and $a - c \leftarrow b$ in H are impossible because H has no flags. This necessitates $a \rightarrow c \leftarrow b$ in H which was desired. $\square$

Corollary A.1 Let H be an essential graph, (a,b) an ordered pair of nodes in H such that either $a \to b$ in H or $a - b$ in H. Put $P = pa_H(b) \setminus \{a\}$ and introduce the set $M = \{c \in ne_H(b) \setminus \{a\}\,;\ [a,c] \text{ is an edge in } H\}$. Suppose that $X \subseteq M$ is complete in H. Then there exists a Bayesian network G which is equivalent to H such that $a \to b$ in G and $pa_G(b) \setminus \{a\} = P \cup X$.

Proof: Every essential graph is a chain graph which satisfies the assumptions on H from Lemma A.3 (c.f. Section 2.3). Let C denote the component of H which contains b. Evidently $X \cup \{b\} \subseteq M \cup \{b\} \subseteq C$ and $X \cup \{b\}$ is a complete set in H_C. Moreover, if $a - b$ in H then $a \in C$ and $X \cup \{a,b\}$ is a complete set in H_C. By Lemma A.2 there exists a perfect numbering of nodes of H_C in which the nodes of X, respectively the nodes of $X \cup \{a\}$, precede the node b and b precedes the remaining nodes of C. After that, arbitrary perfect numberings of nodes of remaining components of H are chosen (by Lemma A.2) and Lemma A.3 is applied to get the respective Bayesian network G. It follows from the construction that $pa_G(b) = \{a\} \cup P \cup X$. $\square$

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

Proof: Recall that an equivalence class $\mathcal{K}$ of Bayesian networks is considered, (a,b) is an ordered pair of its distinct nodes and

$$\mathcal{C}^-_{\mathcal{K}}(a \to b) \ = \ \{\, C\,;\ \exists\, K \in \mathcal{K} \text{ such that } a \to b \text{ in } K \text{ and } C = pa_K(b) \setminus \{a\}\,\}\,.$$

This set is non-empty iff there exists $K \in \mathcal{K}$ such that $a \to b$ in K. It follows from the definition of the essential graph K^* that $a \to b$ in K^* or $a - b$ in K^* then. Thus, the necessity of the condition in (i) of Proposition 3.2 is evident. Its sufficiency follows from Corollary A.1 which implies that $P \equiv pa_{K^*}(b) \setminus \{a\} \in \mathcal{C}^-_{\mathcal{K}}(a \to b)$ (put $X = \emptyset$). Moreover, Corollary A.1 also implies $\mathsf{TUFT}(P|\mathit{cliques}(K^*_M)) \subseteq \mathcal{C}^-_{\mathcal{K}}(a \to b)$ because elements of this tuft of sets are just the sets of the form $P \cup X$ where $X \subseteq M$ is complete in K^* – see Section 2.5. This is one inclusion in the condition (ii) of Proposition 3.2.

The last step is to verify $\mathcal{C}^-_{\mathcal{K}}(a \to b) \subseteq \mathsf{TUFT}(P|\mathit{cliques}(K^*_M))$. If $K \in \mathcal{K}$ such that $a \to b$ in K and $C = pa_K(b) \setminus \{a\}$ then $P \subseteq C$ by the definition of the essential graph. Put $X = C \setminus P$ and observe that $X \subseteq M$. Indeed, for any $c \in X \subseteq C \setminus P$ one has $a \to b \leftarrow c$ in K. Of course, this implies $c \in ne_{K^*}(b)$. If $[a,c]$ is not an edge in K then $a \to b \leftarrow c$ is an immorality in K and, therefore, in K^* which contradicts the assumption $c \notin P$. Thus, $[a,c]$ is an edge in K^* which implies $c \in M$ (see the definition of M in Proposition 3.2). Finally, observe that X is complete as otherwise distinct $c,d \in X$ exist such that $c \to b \leftarrow d$ is an immorality in K and, therefore, in K^* which implies a contradictory conclusion $c,d \in P$. $\square$

Lemma A.4 Let H be an essential graph and (a,b) is an ordered pair of distinct 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 A.2 to it. Therefore, there exists a perfect numbering 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 numberings for other components of H and apply Lemma A.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 4.3 the assumption $b \notin An_H(a)$ implies that ρ cannot start by an arrow in H for which reason it has to 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 $d_l \rightarrow b = d_1$ in G which contradicts the fact that G is an acyclic directed graph. $\square$

Now, the proof of **Proposition 4.2** follows.

Proof: Recall that $\mathcal{L}$ is an equivalence class of Bayesian networks, (a,b) is an ordered pair of its distinct nodes such that $[a,b]$ is not an edge in $\mathcal{L}$ and

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

This class is non-empty iff there exists $L \in \mathcal{L}$ such that $b \notin an_L(a)$. In the condition (i) of Proposition 4.2 the existence of such $L \in \mathcal{L}$ is characterized in terms of the essential graph L^*, namely by two requirements: $b \notin An_{L^*}(a)$ and the set of relative neighbours $R \equiv re_{L^*}(b|a)$ is complete in L^*.
I. The first step is to show the necessity of these two requirements.
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 reason $b \in an_L(a)$ for every $L \in \mathcal{L}$ and the class

$\mathcal{C}^+_{\mathcal{L}}(a \to b)$ is empty. If $b \notin An_{L^*}(a)$ and $R = re_{L^*}(b|a)$ 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 4.2 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 shortened. The fact that L^* is a chain graph, the definition of $re_{L^*}(b|a)$ and the assumption $b \notin An_{L^*}(a)$ imply that there is no edge in L^* between b and $\{d^i_2, \ldots, d^i_{n(i)}\}$ for which the path $b \equiv d^i_0, d^i_1, \ldots, d^i_{n(i)}$ cannot be shortened. Now, consider arbitrary $L \in \mathcal{L}$. If $c^1 \to b \leftarrow c^2$ in L then this is an immorality in L which is not in L^* which contradicts the fact that L and L^* are equivalent – see Section 2.3. Thus, $b \to c^i$ in L for some i and we 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 they are equivalent). This observation implies $b \in an_L(a)$. Therefore, there is no $L \in \mathcal{L}$ with $b \notin an_L(a)$ for which reason the collection $\mathcal{C}^+_{\mathcal{L}}(a \to b)$ is empty.

II. The sufficiency of those two requirements from the condition in (i) of Proposition 4.2 follows from Lemma A.4 applied to $H = L^*$ and $X = \emptyset$ which implies that for $P \equiv pa_{L^*}(b)$ one has $P \cup R \in \mathcal{C}^+_{\mathcal{L}}(a \to b)$.

III. To verify the condition (ii) of Proposition 4.2 one first needs to show $\mathcal{C}^+_{\mathcal{L}}(a \to b) \subseteq \mathsf{TUFT}(P \cup R \,|\, \mathit{cliques}(L^*_M))$ where the set M is defined by $M = \{d \in ne_{L^*}(b) \setminus R;\ d - c \text{ in } L^* \text{ for every } c \in R\}$.

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 $C = pa_L(b)$. 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 4.2 which cannot be shortened. 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 reason 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 this reason $X \subseteq ne_{L^*}(b)$ 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 required conditions (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 the condition (ii) of Proposition 4.2 follows easily from Lemma A.4 applied to $H = L^*$ and any complete set X in L^*_M. $\square$

Acknowledgements

I wish to thank reviewers and volume editors for their comments and positive attitude.

References

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

[2] Auvray V. and Wehenkel L., 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 (2002), 26–35.

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

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

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

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

[7] Jensen F. V., *Bayesian Networks and Decision Graphs*, Springer Verlag (2001), New York.

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

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

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

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

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

[13] Roverato A., A unified approach to the characterisation of equivalence classes of DAGs, chain graphs with no flags and chain graphs, accepted in *Scandinavian Journal of Statistics*.

[14] Spirtes P., Glymour C. and Scheines R., *Causation, Prediction, and Search*, Spinger Verlag (1993), New York.

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

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

[17] Studený M., 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 (2003), 161–172.

[18] Studený, M., Characterization of inclusion neighbourhood in terms of the essential graph: lower neighbours, in Proceedings of WUPES 2003 (Vejnarová J. ed.), School of Economics Prague (2003), 243–264.

[19] Studený M., Characterization of essential graphs by means of an operation of legal merging of components, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 12 (2004), pp. 43–62.

[20] Studený M., *Probabilistic Conditional Independence Structures*, a research monograph, to be published by Springer Verlag in 2004, copyright 2005.

[21] Verma T. and Pearl J., 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 (1991), 220–227.