Graphical Reasoning with Bayesian Networks

Ildikó Flesch and Peter Lucas
Department of Information and Knowledge Systems
Institute for Computing and Information Sciences
Radboud University Nijmegen, The Netherlands
Email: {ildiko,peterl}@cs.ru.nl

Abstract

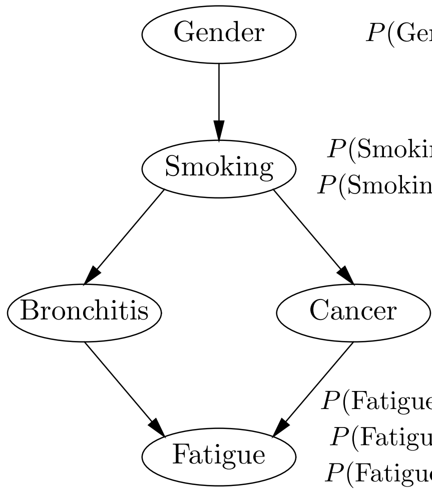
Nowadays, Bayesian networks are seen by many researchers as standard tools for reasoning with uncertainty. Despite the fact that Bayesian networks are graphical representations, representing dependence and independence information, normally the emphasis of the visualisation of the reasoning process is on showing changes in the associated marginal probability distributions due to entering observations, rather than on changes in the associated graph structure. In this paper, we argue that it is possible and relevant to look at Bayesian network reasoning as reasoning with a graph structure, depicting changes in the dependence and independence information. We propose a new method that is able to modify the graphical part of a Bayesian network to bring it in accordance with available observations. In this way, Bayesian network reasoning is seen as reasoning about changing dependences and independences as reflected by changes in the graph structure.

1 Introduction

Bayesian networks are examples of probabilistic graphical models that are powerful tools for data analysis and problem solving in areas involving uncertainty, such as medicine [2]. A Bayesian network consists of two parts: (1) an acyclic directed graph that represents the dependences and independences in a domain of concern, and (2) a joint probability distribution of a set of random variables that is associated with the vertices. It is a very convenient formalism for specifying probabilistic information; by taking into account the independences represented by the graph, usually much less probabilistic information needs to be specified than would be required otherwise.

Bayesian networks are used in problem solving. This is normally accomplished by instantiating the random variables that have been observed; subsequently, the probability distributions of the individual random variables that have not been observed are computed, taking into account the influence of the observations. The results of the computation are visualised by plots. As we only consider discrete probability distributions, the plots are then bar-graphs.

While the bar-graphs are informative to the user, the entered observation may, under certain conditions, change the dependences and independences in a Bayesian network. The surprising fact, however, is that normally the graph structure is kept unchanged in the reasoning process. One may wonder why the changes are not clearly indicated to the user, thus supporting the user's understanding of how dependences and independences change in the face of observed evidence. The authors of this paper believe that without the possibility of displaying the changes in the graphical part of a Bayesian network, reasoning with Bayesian



$P(\text{Gender} = \text{male}) = 0.6$

$P(\text{Smoking} = \text{yes} \mid \text{Gender} = \text{male}) = 0.6$
$P(\text{Smoking} = \text{yes} \mid \text{Gender} = \text{female}) = 0.2$

$P(\text{Bronchitis} = \text{yes} \mid \text{Smoking} = \text{yes}) = 0.7$
$P(\text{Bronchitis} = \text{yes} \mid \text{Smoking} = \text{no}) = 0.001$
$P(\text{Cancer} = \text{yes} \mid \text{Smoking} = \text{yes}) = 0.3$
$P(\text{Cancer} = \text{yes} \mid \text{Smoking} = \text{no}) = 0.1$
$P(\text{Fatigue} = \text{yes} \mid \text{Bronchitis} = \text{yes}, \text{Cancer} = \text{yes}) = 0.9$
$P(\text{Fatigue} = \text{yes} \mid \text{Bronchitis} = \text{yes}, \text{Cancer} = \text{no} = 0.1$
$P(\text{Fatigue} = \text{yes} \mid \text{Bronchitis} = \text{no}, \text{Cancer} = \text{yes}) = 0.8$
$P(\text{Fatigue} = \text{yes} \mid \text{Bronchitis} = \text{no}, \text{Cancer} = \text{no}) = 0$

Figure 1: Acyclic directed graph (left) and associated joint probability distribution (right) of a fictitious Bayesian network concerning causes and consequences of bronchitis and cancer; only half of the probabilistic information is given: the complementary probabilities can be obtained by use of the axiom $P(X = \text{no}) = 1 - P(X = \text{yes})$.

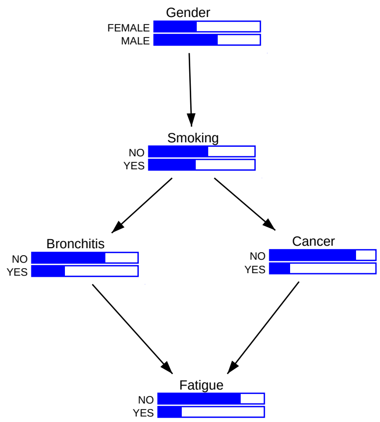
networks is incomplete. In this paper, we propose a new method for graphical reasoning with Bayesian networks. It is expected that a user's understanding of the graphical part of the reasoning process endorses the exploitation of Bayesian networks as problem solvers.

The results obtained by the research discussed in this paper are two-fold. Firstly, we develop a method of graphical reasoning with Bayesian networks that is mathematically sound; secondly, we show that this form of reasoning can be looked upon as reasoning with a class of acyclic directed graphs, rather than reasoning with a single graph. As far as we know, it is the first time that reasoning in Bayesian networks is looked upon in this particular fashion.

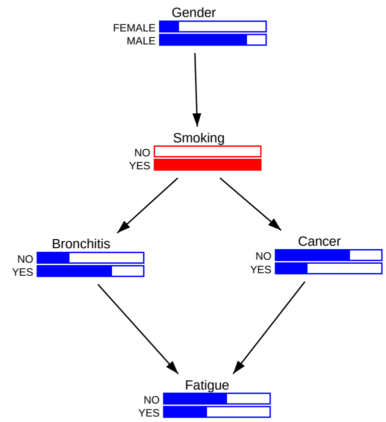
The paper is organised as follows. In the next section, the ideas underlying this research are motivated further. The basic concepts used in this paper are next reviewed in Section 3. Subsequently, in Section 4, a new equivalence relation on Bayesian networks is developed, taking into account random variables that have been observed. In Section 5, work that is related to the research presented in this paper is reviewed and compared to our work. Finally, in Section 6, we summarise what has been achieved and consider further research.

2 Motivating Example

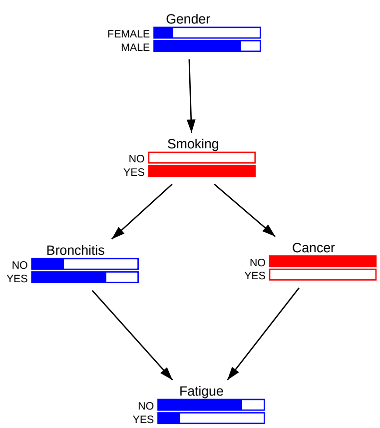
Before going into the details of Bayesian networks and their representation of (in)dependence information, we demonstrate that it is to some extent possible to draw conclusion about dependence and independence of random variables, or variables for short, in Bayesian networks using conventional Bayesian network reasoning, just by looking at probability distributions. The method used is straightforward: by looking at changes in the marginal probability distributions visualised for individual variables, and instantiating some of the other variables, it is possible to conclude that two or more variables are dependent of one another. However,



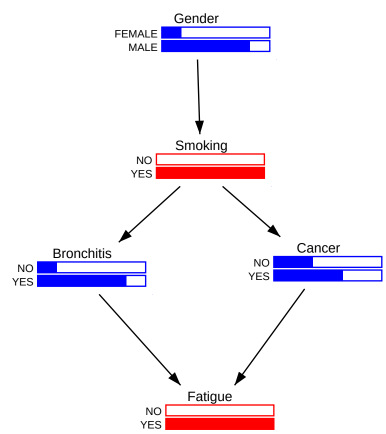
(a) Bronchitis and Cancer are dependent



(b) Bronchitis and Cancer are independent



(c) Proof that Bronchitis and Cancer are independent



(d) Bronchitis and Cancer are dependent

Figure 2: Bayesian networks with associated marginal probability distributions obtained from the specification given in Figure 1; marginal probability distributions (a) and distributions obtained after entering observations concerning smoking (b), smoking and cancer (c) and smoking and fatigue (d).

dependences and independences can also be read off from a graph by using particular subgraph structure rules, to be reviewed in the next section. Readers not familiar with these rules, should for the moment simply try to develop some intuition of reasoning in Bayesian networks.

Consider the specification of a Bayesian network shown in Figure 1; it consists of an acyclic directed graph and a joint probability distribution, factorised in terms of local probability distributions of the form

$$P(X_v \mid X_{\pi(v)})$$

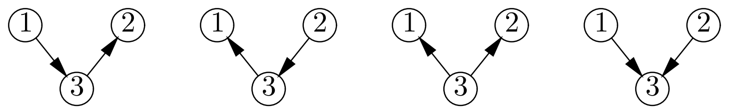
where v is a vertex, $\pi(v)$ the set of parents of vertex v, and X_v is the random variable associated with the vertex v. For the sake of simplicity, we assume here that vertices and variables have the same name.

The marginal probability distributions $P(X_v)$, for each vertex v in the graph, are shown in Figure 2(a); these have been computed by probabilistic inference. Using the rules for extracting (in)dependence information from acyclic directed graphs, the variables 'Bronchitis' and 'Cancer' in Figure 2(a) are dependent, as they are connected via the variables 'Smoking'. However, in Figure 2(b) the variables 'Bronchitis' and 'Cancer' have become *in*dependent, as their common cause (Smoking) has been observed. This should be interpreted as saying that once we know that somebody smokes, also knowing that somebody has cancer does not change our beliefs about whether or not the person has bronchitis; 'Smoking' is the variable that completely explains the dependence that exists between 'Cancer' and 'Bronchitis'. Figure 2(c) proves that this is indeed a correct interpretation of this reasoning process, as the probability distributions of 'Bronchitis' in Figure 2(b) and Figure 2(c) are exactly the same, even though in Figure 2(c) we know, in addition to the fact that the person smokes, that the person has cancer. In contrast, in Figure 2(d) both probability distributions of the random variables 'Bronchitis' and 'Cancer' have changed in comparison to Figure 2(b), despite the fact that 'Smoking' was observed as well. The reason for this is that observing that the person has fatigue, which is a common consequence of bronchitis and cancer, has again made these two random variables dependent of each other.

Clearly, it is possible to reason about changes in dependences and independences in Bayesian networks by looking at changes in the underlying probability distributions. However, it would have been much more convenient if these changes had been visualised by changing the graphical part of the Bayesian network, e.g., by the addition or deletion of arcs. Current Bayesian network packages, however, do not offer this capability. In fact, as we will see, things are not as easy as simply adding and deleting arcs, because by reasoning with an acyclic directed graph we may move beyond this class of probabilistic graphical models, i.e., the result may no longer be an acyclic directed graph. In the remainder of the paper, we will develop the necessary theory to be able to offer this kind of support.

3 Mathematical Preliminaries

Even though the graphical part of a Bayesian network is an acyclic directed graph, we need other graphical representations as well in order to develop our theory of graphical reasoning. In particular, we sometimes need to replace arcs by lines. We first summarise some bits and pieces of graph theory required in the remainder of the paper and some of the theory of statistical independence, which this paper has taken as a starting point.



(a) serial connection (b) serial connection (c) divergent connection (d) convergent connection

Figure 3: Connections as representations of conditional (in)dependences.

3.1 Some Elements of Graph Theory

We assume the reader has some familiarity with notions from graph theory (cf. [2, 3]), such as graph, undirected graph, acyclic directed graph, also called ADG[1] for short in the following, vertex, arc (a directed edge $u \to v$ between two vertices u and v), line (an undirected edge $u - v$ between two vertices u and v). Here, if $u \to v$, then u is called the parent of v. The set of parents of a vertex $v \in V$ is denoted by $\pi(v)$. A graph G is denoted by $G = (V, E)$, where V is the set of vertices and E is the set of edges, i.e., arcs, lines, or both arcs and lines.

A *path* in a graph $G = (V, E)$ is a sequence of unique vertices $v_1, v_2, \ldots, v_n$, with possible exception of v_1, v_n, where either $v_i \to v_{i+1}$ or $v_i - v_{i+1} \in E$ for each i, $1 \leq i \leq n-1$. If all arcs on a path have the same direction, and the path consists of at least *one* arc it is called a *directed path*. A *directed cycle* is a directed path with $v_1 = v_n$. A *trail* τ is a sequence of vertices, where either $v_i \to v_{i+1}$ or $v_{i+1} \to v_i$, or $v_i - v_{i+1}$ are unique edges in graph G for each i. The set of descendants of a vertex $v \in V$, denoted by $\delta(v)$, is the largest set of vertices $U \subseteq V \setminus \{v\}$, where v is connected to each $u \in U$ by a directed path. Let $G = (V, E)$ be an ADG, then if for $W \subseteq V$ it holds that $\pi(v) \subseteq W$ for all $v \in W$, then W is called an *ancestral set*. By $an(W)$ is denoted the *smallest* ancestral set containing W.

In addition to undirected graphs and ADGs we need the concepts of *mixed graph*, which is a graph that contains arcs and lines, and *chain graph*, which is a mixed graph without directed cycles. The notion of chain graph is essential to the remainder of the paper.

3.2 Independence Information

Let $U, W, S \subseteq V$ be disjoint sets of vertices. Let $G = (V, E)$ be an undirected graph. Then, if *each* path between each vertex in U and each vertex in W contains a vertex in S, then U and W are said to be *u-separated* by S; otherwise, they are said to be *u-connected*.

Let $u, w \in (V \setminus S)$ be distinct vertices in the ADG $G = (V, E)$, connected to each other by the trail τ. Then τ is said to be *blocked* by $S \subseteq V$ in G if one of the following conditions is satisfied: (*i*) $s \in S$ appears on the trail τ, and the arcs of τ meeting at vertex s constitute a serial or divergent connection (See Figure 3); (*ii*) $s \notin S$ and $\delta(s) \cap S = \varnothing$, i.e., if s appears on the trail τ then neither s nor any of its descendants occurs in S, and the arcs meeting at s on τ constitute a convergent connection as shown in Figure 3, where vertex 3 is called a *collider* or common child. Then, if *each* trail τ in G between each $u \in U$ and each $w \in W$ is blocked by S, sets U and W are said to be *d-separated* by S; otherwise, U and W are *d-connected* by S.

From an ADG G we can derive its associated undirected *moral graph* G^m that allows reading off all conditional dependences in G, and is constructed by the *moralisation* procedure:

[1]The abbreviation DAG is also frequently used.

(i) add lines to all non-connected vertices, which have a common child, and (ii) replace each arc with a line in the resulting graph. The correspondence between d-separation and moralisation is established by the following proposition:

Proposition 1 *Let $G = (V, E)$ be an ADG and let $U, W, S \subseteq V$ be disjoint sets of vertices. Then, U and W are d-separated by S iff U and W are u-separated in the moral graph of the set of vertices $an(U \cup W \cup S)$, i.e., $G^m_{an(U \cup W \cup S)}$.*
Proof: See Ref. [2], page 72. $\square$

This means that we can choose between two different methods for determining whether two sets of vertices are dependent or independent given a third, possibly empty, set of vertices. By the d-separation method we consider trails between two vertices, look subsequently at the form of the connections at vertices based on Figure 3, and then decide whether the vertices are dependent or not. In the moralisation procedure, the relevant part of the ADG is transformed into an undirected graph. We can then use u-separation, which is simpler than d-separation, to find out whether two sets of vertices are (in)dependent given a third set of vertices.

Finally, for chain graphs we have the notion of *c-separation*, which is closely related to the notion of d-separation (cf. [3, 6] for details). For example, the vertices v_1 and v_4 in the trail $v_1 \rightarrow v_2 - v_3 \leftarrow v_4$ become c-connected by $S = \{v_2\}$, whereas, if it is the only trail, v_1 and v_4 are c-separated by $S = \varnothing$.

All the graph separation notions (u-, d- and c-separation), can be used to define independence relations $\perp\!\!\!\perp_G$, where the presence of lines or arcs represent dependences and the absence of lines or arcs independences. For example, for an ADG G we say that

$$U \perp\!\!\!\perp_G W \mid Z,$$

holds if U and W are d-separated by Z. Complementary to the independence relation $\perp\!\!\!\perp_G$, there is the dependence relation $\not\perp\!\!\!\perp_G$ (as sets of vertices are either dependent or independent given a third set of vertices).

Let P be a *joint probability distribution*, or JPD for short, of the set of discrete random variables X. Instead of random variables we will write variables, for short. Let $U, W, Z \subseteq V$ be disjoint sets of vertices corresponding to sets of variables in the set X, then X_U is said to be *conditionally independent* of X_W *given* X_Z relative to P, denoted by $U \perp\!\!\!\perp_P W \mid Z$, if

$$P(X_U \mid X_W, X_Z) = P(X_U \mid X_Z), \text{ with } P(X_W, X_Z) > 0 \ . \tag{1}$$

An ADG G that is a graphical representation associated with a joint probability distribution P, is called a directed *independence map*, *I-map* for short, if it respects all the dependences in P, i.e.,

$$U \perp\!\!\!\perp_G W \mid Z \ \Rightarrow \ U \perp\!\!\!\perp_P W \mid Z,$$

for all disjoint sets $U, W, Z \subseteq V$. It is said that P obeys the *global directed Markov property* relative to G [2].

A *Bayesian network*, BN for short, is a pair $\mathcal{B} = (G, P)$, where G is an I-map of P. The I-map relation means that the joint probability distribution P obeys all independence statements relative to G. Hence, the graphical part of a Bayesian network can never contain an independence that does not hold for the associated JPD, but it may contain a dependence that does not hold for the JPD.

As an example, consider the Bayesian network from Figure 1 again; some of the dependence and independence statements that hold for the graphical part G of the Bayesian network are:

$\{\text{Bronchitis}\} \not\perp\!\!\!\perp_G \{\text{Cancer}\} \mid \varnothing$
$\{\text{Bronchitis}\} \perp\!\!\!\perp_G \{\text{Cancer}\} \mid \{\text{Smoking}\}$
$\{\text{Bronchitis}\} \not\perp\!\!\!\perp_G \{\text{Cancer}\} \mid \{\text{Smoking, Fatigue}\}$
$\{\text{Smoking}\} \not\perp\!\!\!\perp_G \{\text{Fatigue}\} \mid \{\text{Bronchitis}\}$
$\{\text{Smoking}\} \not\perp\!\!\!\perp_G \{\text{Fatigue}\} \mid \{\text{Cancer}\}$
$\{\text{Smoking}\} \perp\!\!\!\perp_G \{\text{Fatigue}\} \mid \{\text{Bronchitis,Cancer}\}$

The first three (in)dependence statements are interesting. The first statement says that 'Bronchitis' and 'Cancer' are dependent; the reason for this is that there is a common course, smoking, which renders these vertices dependent. As soon as we know the common course 'Smoking', all dependence between 'Bronchitis' and 'Cancer' is explained, and thus, 'Bronchitis' and 'Cancer' become independent, as signified by the second statement. The third statement says that if in addition to 'Smoking' 'Fatigue' is observed, 'Bronchitis' and 'Cancer' have again become dependent through the common consequence 'Fatigue'. Note that the independence relation $\perp\!\!\!\perp_P$ shows a similar behaviour, as can be verified by looking at Figure 2; G is an I-map of P.

We have seen that independence information can be represented in different form, e.g., in graphical form or hidden within a probability distribution. However, usually these representations are not unique. We, therefore, need to take into account that different ADGs may encode the same independence information. We only consider equivalence of ADGs:

Definition 1 ***(independence equivalence** [6]) Let G, G' be two ADGs and let $\perp\!\!\!\perp_G$ and $\perp\!\!\!\perp_{G'}$ be their independence relations, defined by the d-separation criterion. If $\perp\!\!\!\perp_G = \perp\!\!\!\perp_{G'}$, then the two graphs are said to be* independent equivalent *to one another.*

Independence equivalence of ADGs can also be defined in an other way. The undirected version of G is called the *skeleton* of G. A subgraph $G' = (V', E')$ of an ADG $G = (V, E)$, with $V' = \{u, w, z\} \subseteq V$, $E' = E \cap (V' \times V')$, is called an *immorality* if $u \to z, w \to z \in E'$ and $u \to w, u \leftarrow w \notin E'$. We now have the following theorem by Verma and Pearl [7]:

Theorem 1 *Two ADGs are* independence equivalent *with each other iff they have the same skeleton and the same set of immoralities.*

It now appears that classes of independence equivalent ADGs can be uniquely described by means of chain graphs, called essential graphs, which thus act as class representatives [1]; they are defined as follows:

Definition 2 ***(essential graph)** Let E denote an equivalence class of acyclic directed graphs that are independence equivalent. The* essential graph *G^* is then the smallest graph larger than any of the acyclic directed graphs G in the equivalence class E; formally*

$$G^* := \bigcup \{G \mid G \in E\}, \tag{2}$$

where the union of graphs, denoted by $\bigcup$, is obtained by taking the union of their vertex sets and their sets of arcs, where arcs connecting the same vertices, but pointing in different directions, are replaced by a line.

Note that G^* is the least upper bound of all the graphs that it represents.

4 Taking into Account Observations

Problem solving using a Bayesian network involves instantiating variables, and computing the new probability distribution. We also consider the consequences for the (in)dependence information in this section.

4.1 Probability Updating

In using a Bayesian network when solving a problem at a certain instance, variables are instantiated to values, that have been observed. Thus, the following mutually disjoint sets are distinguished:

- O, *observed vertices* (shown as shaded circles or ellipses in diagrams), and
- U, *unobserved vertices* (shown as non-shaded circles or ellipses in diagrams).

Likewise, we also distinguish between observed variables X_O (e.g., whether or not the person smokes) and *unobserved variables*, X_U. A variable cannot be both observed and unobserved.

The JPD obtained by incorporating observed variables, called the *observed joint probability distribution* P_O, is defined as follows:

$$P_O(X_V) := P(X_V \mid X_O = x_O). \tag{3}$$

The definition of observed joint probability distribution gives rise to the introduction of independences; in the following proposition the consequences with respect to the independence relationships between observed and unobserved variables are explored.

Proposition 2 *Let V be a set of vertices with associated set of variables X_V. Furthermore, let $O' \subseteq O$, $V' \subseteq V \setminus O'$, then, it holds that $X_{O'}$ and $X_{V'}$ are independent with respect to P_O, i.e., for all possible values $x_{O'}$ and $x_{V'}$ it holds that $P_O(X_{O'} = x_{O'}, X_{V'} = x_{V'})$ $= P_O(X_{O'} = x_{O'})P_O(X_{V'} = x_{V'})$.*

The following specific properties with respect to P_O are a consequence of the proposition above: (1) sets of observed variables are independent of the unobserved variables, and (2) sets of observed variables are mutually independent.

4.2 Transformation due to Observations

If the observation of a variable gives rise to the creation of a new dependence, then it appears that one proper way in which the new dependence can be represented is by inserting lines into the graph. Let $G = (V, E)$ be an ADG, and let $O \subseteq V$ be the set of observed vertices. Furthermore, let $u, w \in U$, $u \neq w$, be two unobserved vertices and let $o \in O$ be either the common child or a descendant of the common child of an immorality with u and w as parents. Then, the line (u, w), which is inserted into graph G, is called a *moral line*. The set of moral lines of graph G is denoted by $M_G(O)$; e.g., $3-4$ in Figure 4(b) is a moral line with $O = \{2\}$. Clearly, it holds that $M_G(\varnothing) = \varnothing$.

Moral lines play a crucial role in the context of the observation of variables, since they depict the new dependences created by the observations. However, by inserting a line into an ADG, the result is no longer an ADG. Sometimes, the result is not even a chain graph, but rather a mixed graph that is cyclic. However, even though moral lines can be inserted into a graph, we still want to keep the chain graph property. This would allow us to apply c-separation to the graphical model, which enables us to uncover the independence relation from the graph. The following transformation essentially repairs these 'side effects' of moral line insertion.

Let $G = (V, E)$ be an ADG with observed vertices $O \subseteq V$. Then, the *arc–line transformation*, denoted by T_A is defined as

$$T_A : G = (V, E) \mapsto G_A = (V, E_T),$$

where E_T denotes the smallest set of edges defined as follows. For each $(u, v) \in E$:

- if $u \in an(\{v\} \cup O)$, $v \in an(\{u\} \cup O)$ and $u, v \notin O$, then $u - v \in E_T$;
- otherwise, $u \to v \in E_T$.

Proposition 3 *Let G be an ADG with set of observed vertices O. Then, the resulting graph G_A obtained by the arc-line transformation is a chain graph.*

Proof: Suppose that G_A consists of the directed cycle $v_1 \to v_2 \to \cdots \to v_m - v_1$, and, thus, is not a chain graph. This can happen because of one of the following two reasons: (i) arc (v_1, v_m) in G and there is a path from vertex v_m to an observed vertex $o \in O$, or (ii) arc (v_m, v_1) in G and there is a path from vertex v_1 to vertex $o \in O$. But then, all arcs in the directed cycle also satisfy the first condition of the arc–line transformation and should have been replaced by a line as well; contradiction. Thus, graph G_A does not include directed cycles, and, hence, is a chain graph. $\square$

To determine the set of independence statements following specific properties of P_O, we need to take into account that (1) observed variables are independent of the remaining, unobserved variables, and (2) that sets of observed variables are independent. Based on these properties, these created independences are depicted by labelling arcs as *semi-observed arcs*, which connect an unobserved and an observed vertex, and as *observed arcs*, which connect two observed vertices. The sets of semi-observed and observed arcs are denoted by E_S and E_O, respectively. The conditional independence set mirrored by the graph G_O, denoted by $\perp\!\!\!\perp_{G_O}$, is then defined by *observed c-separation*, explained below.

An ADG $G = (V, E)$ can be transformed into a chain graph $G_O = T_O(G)$ by an *observation transformation*, denoted by T_O, which includes the additional (in)dependences obtained by observational knowledge, as follows:

$$T_O : G = (V, E) \mapsto G_O = (V, (E_T \cup M_G(O)) \setminus (E_S \cup E_O)),$$

where E_T, E_S and E_O are defined as above. Thus, (semi)observed arcs are removed; observed c-separation is then nothing else then c-separation applied to G_O.

Proposition 4 *Let $B = (G, P)$ be a BN with observed variables X_O. Then, $G_O = T_O(G)$ is a chain graph, which is an I-map of P_O.*

Proof: The I-map property implies that each dependence in P_O should also be represented in the graph G_O. As (semi)observed arcs mirror independences created in P_O, the I-map property could only be lost because of two reasons: (i) the insertion of moral lines into the graph, (ii) the replacement of arcs by lines by the arc–line transformation. (i) Note that

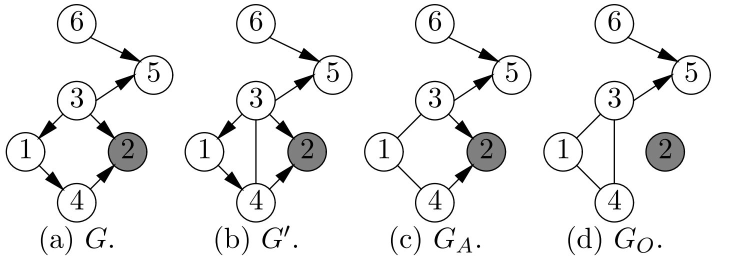


Figure 4: Example of arc–line and observation transformation.

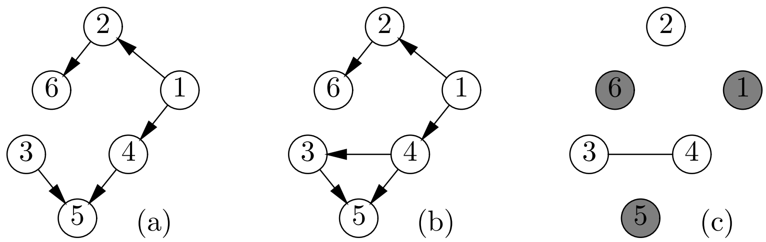


Figure 5: Example of observed equivalence graphs.

moral lines depict additional dependence between parents, reflecting dependence because of an *observed* collider. (*ii*) Replacing arcs by lines, we do not change arcs involved in an immorality into a line by the arc–line transformation. We conclude that any dependence added to the graph is valid; moreover, no valid dependences are removed. □

Consider the ADG shown in Figure 4(a), with $O = \{2\}$. Here, the moral line $3 - 4$ needs to be inserted into the graph; however, the resulting graph would contain a directed cycle, see (b). Therefore, we first apply the arc–line transformation to the graph G to remove this potential directed cycle, by transforming the arcs $3 \to 1$ and $1 \to 4$ into lines, as is shown in (c). The resulting chain graph $G_O = T_O(G)$ obtained by applying the observation transformation is shown in (d).

4.3 Observed Equivalence Classes

An interesting question is whether it is possible that two different ADGs that are independence equivalent are again independence equivalent after entering observations. Related to this question is the issue whether two ADGs that are not independence equivalent can become independence equivalent after taking into account observations. These questions will be briefly explored in this section.

Let G and G' be two ADGs, then they are said to be *observed independence equivalent* with each other with respect to the set of observed vertices O, if it holds that $\perp\!\!\!\perp_{G_O} = \perp\!\!\!\perp_{G'_O}$. Chain graphs obtained by the observation transformation are used as a basis to establish equivalence, using observed c-separation.

As an example, consider Figure 5 with the graphs G^a and G^b as the two left-hand side graphs, which are *not* independence equivalent. However, when taking the observed variables $X_O = \{X_1, X_5, X_6\}$, the graphs G^a_O and G^b_O do become observed independence equivalent; both graphs are now represented by the graph in Figure 5(c).

Independence equivalence between graphs implies observed independence equivalence:

Proposition 5 *Let G and G' be two independence equivalent ADGs, then $G_O = T_O(G)$ and $G'_O = T_O(G')$ are again independence equivalent.*

An observed independence equivalent class can be uniquely represented by an *observed essential graph*, which simply is an essential graph taking into account the special nature of observed vertices.

Thus, independence equivalent ADGs remain equivalent, possibly as a chain graph that is not an ADG, after taking into account observations; in addition, ADGs that originally were not equivalent, can become independence equivalent after the processing of observations.

4.4 Graphical Reasoning Illustrated

We illustrate the theory developed above by considering the Bayesian network shown in Figure 6(a), which is an extension of the network of Figure 1.

As a result of the observation of indolence, the variables 'Bronchitis' and 'Cancer' become directly dependent, indicated by a line connecting the corresponding vertices shown in Figure 6(b). This dependence cannot be changed, except when observing either or both of these variables. In addition, most of the arcs have been changed into lines, with the exception of the arc pointing towards 'Weightloss', which has not changed as observing indolence has not changed the dependence information concerning this variable. Note that it still holds that $\{\text{Smoking}\} \perp\!\!\!\perp_{G_O} \{\text{Fatigue}\} \mid \{\text{Bronchitis, Cancer}\}$, as the observation of indolence is unable to create a direct dependence between smoking and fatigue.

After observing fatigue, the variables 'Bronchitis' and 'Cancer' become directly dependent, as indicated in graph (c). Furthermore, similar to graph (b), some arcs will be replaced by lines. Three arcs that are related to this observed variable are removed from the graph. The resulting graph, (c), expresses the information that smoking still depends on gender and affects the occurrence of bronchitis and cancer. If smoking *is* observed, bronchitis and cancer are still dependent of each other.

Graph (d) represents the effects of observation of the variable 'Smoking'. In this case, the lower part of the graph remains unchanged, since the joint probability distribution of bronchitis and cancer already incorporates the observation of smoking. Therefore, we are also able to remove some arcs from the original graph related to smoking. We conclude that the graph transformations are consistent with our intuition.

5 Related Work

Richardson and Spirtes have investigates the properties of maximal ancestral graphs (MAGs), which are hybrid graphs containing directed and undirected edges as well as double directed edges, under variable selection and marginalisation [5]. They have shown that MAGs are closed under these two operations. Variable selection is related to the concept of observed variables introduced in this paper. However, the work by Richardson and Spirtes does not focus on reasoning with a Bayesian network, but, instead, considers the representation of selection effects in probabilistic graphical models. In addition, the work does not consider the changes in the relationships between equivalence classes when taking into account observed variables.

Van der Gaag has studied the effects on the efficiency of probabilistic inference when taking into account observed variables [4]. The transformation proposed in the paper by Van

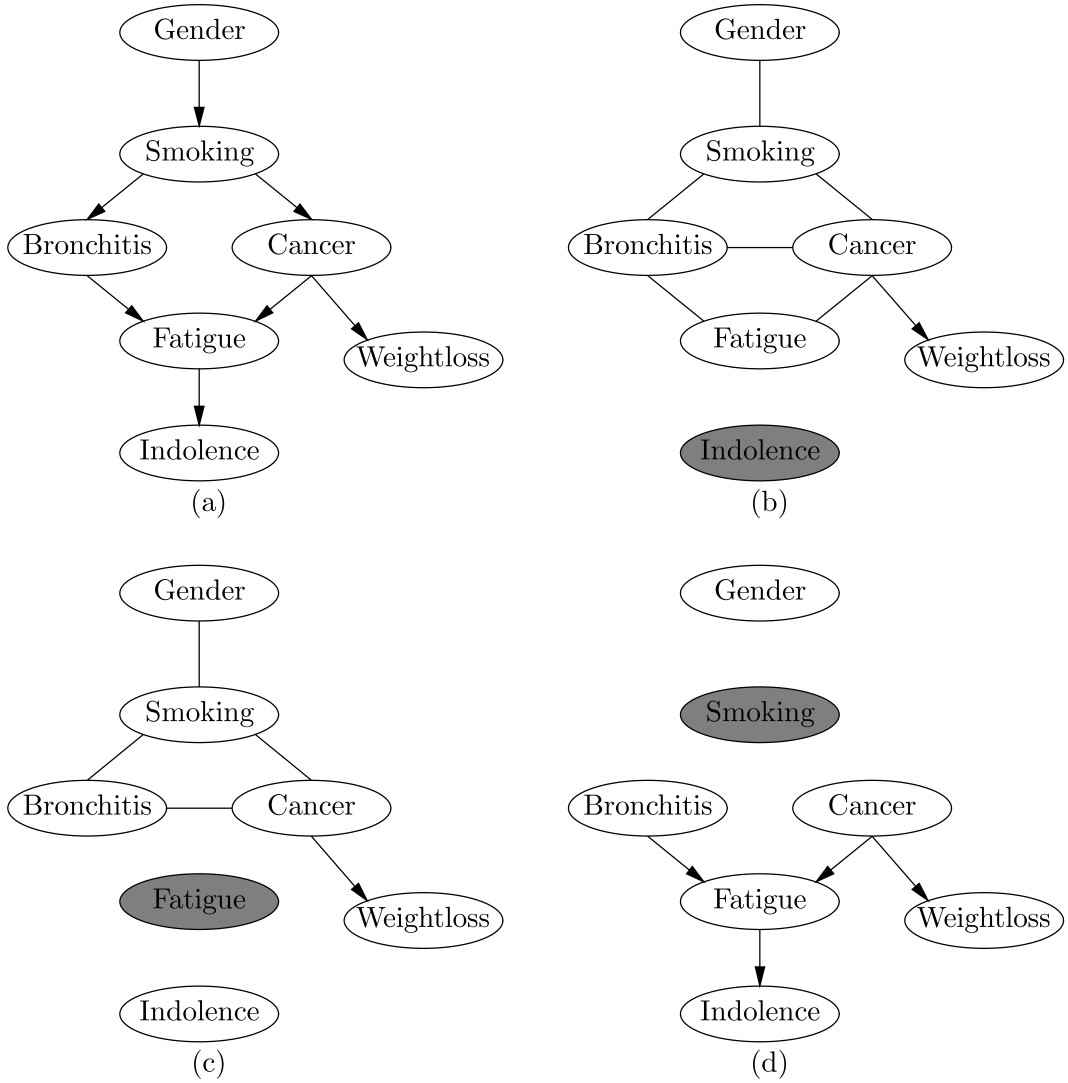


Figure 6: Example transformations due to various observations.

der Gaag is different from the one proposed in our paper, as it was necessary to keep some of the semi-observed arcs, as dependences created by the observation of variables were not represented by the insertion of extra lines, as in our paper. In addition, Van der Gaag's work neither considers chain graphs, nor independence equivalence.

6 Discussion

In the paper, we have developed a method for graphical reasoning with Bayesian networks. We have looked upon Bayesian network reasoning as logical reasoning with (in)dependence information, rather than reasoning with the non-graphical aspects of a probability distribution only, which is a more common view on Bayesian-network reasoning. Although reasoning with independence information can also be viewed as reasoning in terms of a logical language, where set theory is augmented by the independence predicate $\perp\!\!\!\perp$, we have chosen to use graphs for the representation of the reasoning process. While there can be no doubt that Bayesian networks and related probabilistic graphical models, seen as compact representations of JPDs, thank their existence to a major extent to the associated graphical notation, normally the graphical representation is ignored once the Bayesian network is used to solve problems. In this paper, we show that it is possible and natural to use graphs, in particular chain graphs, as a formalism for the representation of changes in the dependence and independence information in Bayesian networks. It is, of course, still possible to inspect the updated associated probability distribution.

Facilities that assist with the graphical reasoning of a Bayesian network may enhance the insight of the user when applying a Bayesian network to problem solving; how this may influence the problem-solving process is something that needs to be explored further.

Acknowledgements. This work has been partially funded by the Netherlands Science Foundation (NWO) (project 612.066.201).

References

[1] Andersson, S. A., Madigan, D., and Perlman, M. D. A Characterization of Markov Equivalence Classes for Acyclic Digraphs. *Annals of Statistic 25* (1997), 505–541.

[2] Cowell, R. G., Dawid, A. P., Lauritzen, S. L., and Spiegelhalter, D. J. *Probabilistic Networks and Expert Systems.* Springer-Verlag New York, 1999.

[3] Flesch, I., and Lucas, P.J.F. Markov Equivalence in Bayesian Networks. Technical Report, Radboud University Nijmegen, NIII-R0436, 2004.

[4] Gaag, L. C. van der. *Evidence Absorption for Belief Networks.* Technical Report RUU-CS-93-35, Utrecht University, 1993.

[5] Richardson, T., and Spirtes, P. Ancestral graph Markov models. *Annals of Statistics 30* (2002), 962–1030.

[6] Studený, M. *Probabilistic Conditional Independence Structures.* Springer-Verlag London, 2005.

[7] Verma, T., and Pearl, J. Equivalence and synthesis of causal models. In UAI'90 (San Francisco, CA, 1990), Morgan Kaufmann, pp. 220–227.