An Algebraic Characterization of Equivalent Bayesian Networks

by

S.K.M. Wong and Dan Wu

Technical Report CS-02-02

February 2002

Department of Computer Science

University of Regina

Regina, Saskatchewan

Canada S4S 0A2

ISSN 0828-3494

ISBN 0-7731-0435-6

Abstract

In this paper, we propose an *algebraic* characterization for equivalent classes of Bayesian networks. Unlike the other characterizations, which are based on the graphical structure of Bayesian networks, our algebraic characterization is derived from the intrinsic algebraic structure of Bayesian networks, i.e., joint probability distribution factorization. The new proposed algebraic characterization not only provides us with a new perspective to look into equivalent Bayesian networks, but also suggests simple and efficient methods for determining equivalence of Bayesian networks and identifying compelled edges in Bayesian networks.

1 Introduction

Bayesian networks [6], [3] have been well established as a knowledge based system for uncertainty information processing. Based on incomplete or uncertain information from a problem domain, Bayesian networks can help people make decision under uncertain situation using probability.

Being considered a knowledge based system, Bayesian networks have a set of conditional probability distributions, whose product yields a *joint probability distribution*, as its knowledge base; and an efficient inference engine [5], [4] for computing posterior probability based on observed evidence.

The success of Bayesian networks relies on its capability of acquiring and representing conditional independency [2] information in a problem domain. Informally, a Bayesian network consists of a graphical structure, i.e., a directed acyclic graph (DAG) representing conditional indenpendencies and a set of conditional probability distributions, whose product yields a joint probability distribution. The set of conditional probability distributions, being considered the knowledge base, together with efficient inference engines [5], [4] for computing posterior probability based on observed evidence composes a knowledge based system for dealing with uncertainty.

It has been noted that different Bayesian networks may be equivalent in the sense that they actually represent the same joint probability distribution (and thus conditional independency information as well) [7], even though they have different graphical structures. Research has been done to determine whether two given Bayesian networks are equivalent or not and to find out invariant portions in the DAGs [7, 1] of equivalent Bayesian networks. Graphical criterion has been proposed to determine if two different Bayesian networks are equivalent or not [7]. More recently, Chickering [1] reported a transformational characterization of equivalent Bayesian networks based on the notion of covered edges in a DAG. An algorithm was also designed to find out all edges in a DAG that maintain their directionality in equivalent Bayesian networks.

In this paper, instead of providing another graphical criterion, we propose an *algebraic* characterization for describing equivalent Bayesian networks. This algebraic characterization was inspired by our understanding that the joint probability distribution defined by the set of conditional probability distributions is the essence of a Bayesian network, while the DAG is only an *auxiliary* graphical structure which helps represent conditional independency information inferred by the joint probability distribution. Therefore, we believe that joint probability distribution of a Bayesian network itself alone should possess the capability to describe and characterize equivalent Bayesian networks. By studying different forms of the factorization of a joint probability distribution defined by a Bayesian network, it is discovered that one can use a particular form of the factorization to describe and characterize equivalent Bayesian networks. This discovery further confirms our understanding that the joint probability distribution alone is sufficient to capture the characteristic of equivalent Bayesian networks. Moreover, this algebraic characterization suggests that the problem of identifying compelled edges in a Bayesian network can be viewed from a different perspective such that a simple and efficient algorithm can be easily designed.

The paper is organized as follows. We review pertinent background knowledge in section 2. The proposed algebraic characterization for equivalent Bayesian network will be presented in section 3. Section 4 discusses several applications of the new algebraic characterization, in particular, we propose a method for identifying all compelled edges in a DAG. We conclude our paper in section 5.

2 Background

We assume readers have familiarities with Bayesian networks to the extent in [6]. We thus will quickly go through the notions that will be used in the paper.

Let U be a finite set of variables, we use $p(U)$ to denote a *joint probability distribution* (jpd) over U. We call $p(V)$, $V \subset U$, *marginal* (distribution) of $p(U)$ and

$p(X|Y)$ *conditional probability distribution* (CPD). By XY, where $X \subseteq U$, $Y \subseteq U$, we mean $X \cup Y$. According to definition of conditional probability,

$$p(X|Y) = \frac{p(XY)}{p(Y)}, \text{whenever } p(Y) > 0,$$

we thus say in the above expression, the denominator $p(Y)$ is *absorbed* by the numerator $p(XY)$ to yield conditional $p(X|Y)$. It is noted that $p(Y)$ can be absorbed by $P(X)$ to yield $p(X - Y|Y)$ if and only if $Y \subset X$. The concept of "absorbed" is very important in our disscussion in this paper.

A *Bayesian network* (BN) is a tuple $(\mathcal{D},\ P)$, where (a) $\mathcal{D} =< U,\ E >$ is a *directed acyclic graph* (DAG), as a qualitative part, with $U = \{X_1,\ \ldots,\ X_n\}$ as the nodes (variables) of DAG and $E = \{(X_i,\ X_j) \mid X_i,\ X_j \in \mathcal{D}\}$ as the set of directed edges (from X_i to X_j) of $\mathcal{D}$; and (b) $P = \{p(X_i|pa(X_i)) \mid X_i \in U\}$ is a set of *conditionals* as quantitative part, where $pa(X_i)$ denotes the parents of node X_i in DAG D, such that,

$$p(U) = \prod_{i=1}^{n} p(X_i|pa(X_i)). \tag{1}$$

We will use the term BN and DAG interchangeably if no confusion arises.

Definition 1 [1] Two Bayesian networks are *equivalent* if they define the same probability distribution.

We use $BN1 \approx BN2$ (or $\mathcal{D}_1 \approx \mathcal{D}_2$) to denote that $BN1$ and $BN2$ are equivalent. The relation $\approx$ induces a set of equivalence classes over Bayesian networks.

An edge $(X,\ Y)$ in a DAG $\mathcal{D}$ is *compelled* if for any DAG $\mathcal{D}' \approx \mathcal{D}$, $(X,\ Y)$ is also in $\mathcal{D}'$. Otherwise, it is *reversible*. The notion of *covered* edge is important for the study in [1]. An edge $(X,\ Y)$ is *covered* in a DAG if $pa(Y) = pa(X) \cup X$. A $v - structure$ in a DAG $\mathcal{D}$ is an ordered triple of nodes $(X,\ Y,\ Z)$ such that (1) $\mathcal{D}$ contains edges $(X,\ Y)$ and $(Z,\ Y)$, and (2) X and Z are not directly connected in $\mathcal{D}$. The *skeleton* of a DAG is the undirected graph obtained by dropping the directionality of every edge in a DAG.

Verma and Pearl characterizes equivalent DAGs using the notions of skeleton and $v - structure$ in DAGs.

Theorem 1 [7] Two DAGs are equivalent if and only if they have the same skeleton and the same $v-structures$.

Chickering characterizes equivalent DAGs using the notion of covered edges in a DAG.

Theorem 2 [1] Let $\mathcal{D}$ be a DAG containing an edge $(X,\ Y)$ and $\mathcal{D}'$ be a DAG identical to $\mathcal{D}$ except that the edge $(X,\ Y)$ in $\mathcal{D}$ is oriented as $(Y,\ X)$ in $\mathcal{D}'$, $\mathcal{D} \approx \mathcal{D}'$ if and only if $(X,\ Y)$ is a covered edge in $\mathcal{D}$.

Because of the symmetry between $\mathcal{D}$ and $\mathcal{D}'$, we have:

Corollary 1 $(Y,\ X)$ is also a covered edge in $\mathcal{D}'$.

Chickering further shows that a property holds for all pairs of equivalent Bayesian networks that differ by a single covered edge orientation if and only if that property holds for Bayesian networks in the equivalent class. In other words, a given property is invariant over all equivalent Bayesian networks if and only if we can show that the property is invariant to a pair of DAGs that differ in a single reversal of covered edge.

3 The Algebraic Characterization

In this section, we will present an algebraic characterization that describes the equivalent class of Bayesian networks. We begin with an illustrative example.

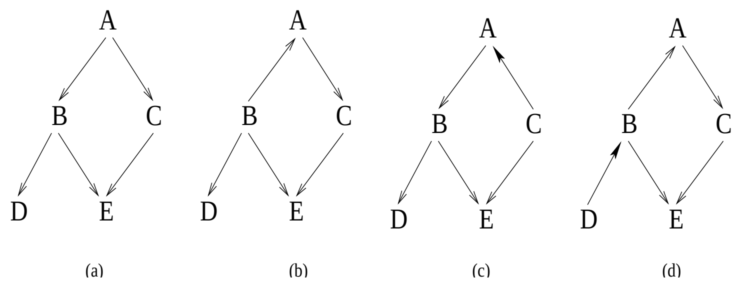

Figure 1: Equivalent Bayesian networks.

Example 1 Consider the BN shown in Figure 1 (a), whose corresponding jpd is written as:

$$\begin{aligned} p(ABCDE) &= p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|BC), && (2)\\ &= p(A) \cdot \frac{p(BA)}{p(A)} \cdot \frac{p(CA)}{p(A)} \cdot \frac{p(DB)}{p(B)} \cdot \frac{p(EBC)}{p(BC)}, && (3)\\ &= \frac{p(BA) \cdot p(CA) \cdot p(DB) \cdot p(EBC)}{p(A) \cdot p(B) \cdot p(BC)}. && (4) \end{aligned}$$

It is noticed that equation (3) was obtained by rewriting equation (2) according to the definition of conditional probability; equation (4) was obtained by canceling numerator $p(A)$ and denominator $p(A)$ in equation (3). It perhaps should be emphasized here that equation (4) and equation (2) are identities.

For Bayesian networks, we are interested in the jpd being factorized as a product of CPDs according to the topological structure of DAG. The fact that equation (4) and equation (2) are identities imposes the following question: *if we are given equation (4), can we derive any other CPD product factorization other than the one in equation (2)?*

In order to turn equation (4) into CPD factorization, it is obvious that we need to absorb each denominator in equation (4), i.e., $p(A)$, $p(B)$, $p(BC)$. The question imposed above can thus be easily answered by the following demonstrations:

$$\begin{aligned} p(ABCDEF) &= \frac{p(BA) \cdot p(CA) \cdot p(DB) \cdot p(EBC)}{p(A) \cdot p(B) \cdot p(BC)}, && (5)\\ &= p(BA) \cdot \frac{p(CA)}{p(A)} \cdot \frac{p(DB)}{p(B)} \cdot \frac{p(EBC)}{p(BC)}, && (6)\\ &= p(A) \cdot p(B|A) \cdot p(C|A) \cdot p(D|B) \cdot p(E|BC), && (7)\\ &= p(B) \cdot p(A|B) \cdot p(C|A) \cdot p(D|B) \cdot p(E|BC). && (8) \end{aligned}$$

It is noted that in equation (5), denominator $p(A)$ was absorbed by numerator $p(CA)$; denominator $p(B)$ was absorbed by numerator $p(DB)$; denominator $p(BC)$ was absorbed by numerator $p(EBC)$; this scheme for denominator absorption yields equation (6), which further yields equation (7) and (8). Both equation (7) and (8) represents BN as shown in Figure 1 (a), (b), respectively.

It perhaps should be mentioned that the reason we obtained the equations (7) and (8) was due to the scheme for denominator absorption. It is also noted that for equation (4) we have another 2 schemes for denominator absorption. One scheme for denominator absorption is shown below:

$$
\begin{aligned}
p(ABCDEF) &= \frac{p(BA)\cdot p(CA)\cdot p(DB)\cdot p(EBC)}{p(A)\cdot p(B)\cdot p(BC)}, && (9)\\
&= \frac{p(BA)}{p(A)}\cdot p(CA)\cdot \frac{p(DB)}{p(B)}\cdot \frac{p(EBC)}{p(BC)}, && (10)\\
&= p(B|A)\cdot p(C)\cdot p(A|C)\cdot p(D|B)\cdot p(E|BC), && (11)\\
&= p(B|A)\cdot p(A)\cdot p(C|A)\cdot p(D|B)\cdot p(E|BC). && (12)
\end{aligned}
$$

The BNs corresponding to equation (11) and (12) are shown in Figure 1 (c), (a).

The other scheme for denominator absorption is:

$$
\begin{aligned}
p(ABCDEF) &= \frac{p(BA)\cdot p(CA)\cdot p(DB)\cdot p(EBC)}{p(A)\cdot p(B)\cdot p(BC)}, && (13)\\
&= \frac{p(BA)}{p(B)}\cdot \frac{p(CA)}{p(A)}\cdot p(DB)\cdot \frac{p(EBC)}{p(BC)} && (14)\\
&= p(A|B)\cdot p(C|A)\cdot p(B)\cdot p(D|B)\cdot p(E|BC), && (15)\\
&= p(A|B)\cdot p(C|A)\cdot p(D)\cdot p(B|D)\cdot p(E|BC). && (16)
\end{aligned}
$$

The BNs corresponding to equation (15) and (16) are shown in Figure 1 (b), (d).

So far We have tried all different schemes for denominator absorption and it results in four different BNs shown in Figure 1. Obviously, all these 4 BNs are the equivalent and form an equivalent class of BNs.

The above examples actually demonstrate that the fraction factorization of a jpd defined by a BN characterizes the equivalent class of BNs and all equivalent BNs in the same class can be obtained from the fraction factorization. We are now to show that this observation is generally true.

Definition 2 Consider a BN with its jpd as follows:

$$
\begin{aligned}
p(X_1X_2\ldots X_n) &= p(X_1)\cdot p(X_2|pa(X_2))\cdot \ldots \cdot p(X_n|pa(X_n)), && (17)\\
&= \frac{p(X_1)}{1}\cdot \frac{p(X_2,\ pa(X_2))}{p(pa(X_2))}\cdot\ldots\cdot \frac{p(X_n,\ pa(X_n))}{p(pa(X_n))}, && (18)\\
&= \prod_{i,\ j} \frac{p(X_i,\ pa(X_i))}{p(pa(X_j))}, && (19)
\end{aligned}
$$

where $\{X_i,\ pa(X_i)\} \neq \{pa(X_j)\}$ for any $1 \leq i,\ j \leq n$. Equation (17) is called *Bayesian factorization*. Each $p(X_i|pa(X_i))$ in equation (17) is called a *factor*. We call the equation in (18) *fraction factorization* of the BN. The jpd obtained in equation (19) (by canceling any applicable numerator and denominator in equation (18)) is called the *intrinsic factorization* of the BN.

For instance, the factorization in equation (4) is the *intrinsic* factorization of the BN shown in Figure 1.

It is worth mentioning that the number of denominators in intrinsic factorization is always less than the number of numerators.

Before we present the main result of this paper, consider the following lemma.

Lemma 1 If $\mathcal{D}$ is a DAG containing an edge $(X,\ Y)$ and $\mathcal{D}^{'}$ is a DAG identical to $\mathcal{D}$ except that the edge $(X,\ Y)$ in $\mathcal{D}$ is now oriented as $(Y,\ X)$ in $\mathcal{D}^{'}$, then

(a). $pa^{'}(Y) = pa(X)$,

(b). $\{X\} \cup pa^{'}(X) = \{Y\} \cup pa(Y)$,

where $pa^{'}(Y)$ is the parent set for node Y in $\mathcal{D}^{'}$, $pa(X)$ is the parent set for node X in $\mathcal{D}$.

Proof: Since $(X,\ Y)$ is a covered edge in $\mathcal{D}$, it follows $pa(Y) = \{X\} \cup pa(X)$. Similarly, since $(Y,\ X)$ is a covered edge in $\mathcal{D}^{'}$, it follows $pa^{'}(X) = \{Y\} \cup pa^{'}(Y)$. Because $(X,\ Y)$ is the only edge that differs in $\mathcal{D}$ and $\mathcal{D}^{'}$, it then follows $pa^{'}(X) = pa(X) \cup \{Y\}$, $pa^{'}(Y) = pa(Y) - \{X\}$. Moreover, $pa^{'}(Y) = pa(Y) - \{X\} = (\{X\} \cup pa(X)) - \{X\} = pa(X)$; $\{X\} \cup pa^{'}(X) = \{X\} \cup \{Y\} \cup pa^{'}(Y) = \{X\} \cup \{Y\} \cup pa(X) = \{Y\} \cup pa(Y)$.

We now present the characterization that describes an equivalent class of Bayesian networks.

Theorem 3 Two BNs are equivalent if and only if they have the same intrinsic factorization.

Proof: The proof for the "if" part is trivial. If two BNs have the same intrinsic factorization, they must define the same jpd. We thus only need to prove the "only if" part.

According to [1], we only need to show our claim holds for two equivalent BNs, say $BN1$ and $BN2$, that differ by a single covered edge orientation. We further assume that in $BN1$, we have a covered edge $(X,\ Y)$, while in $BN2$, we have the edge $(Y,\ X)$. Consider the fraction factorizations for both $BN1$ and $BN2$. For any node Z of $BN1$ and $BN2$ such that $Z \neq X \neq Y$, we have the factor $p(Z|pa(Z))$, i.e., $\frac{p(Z,\ pa(Z))}{p(pa(Z))}$, in the fraction factorization of both $BN1$ and $BN2$. We thus focus on the factors for nodes X and Y in $BN1$ and $BN2$. First consider the edge $(X,\ Y)$ in $BN1$, since $(X,\ Y)$ is a covered edge in $BN1$, it follows that $pa(Y) = pa(X) \cup \{X\}$. In the fraction factorization of $BN1$, we must have $\frac{p(Y,\ pa(Y))}{p(pa(Y))} \cdot \frac{p(X,\ pa(X))}{p(pa(X))}$ (i.e., $p(Y|pa(Y)) \cdot p(X|pa(X))$), which can be simplified as $\frac{p(Y,\ pa(Y))}{p(pa(X))}$. Similarly, consider the edge $(Y,\ X)$ in $BN2$, since $(Y,\ X)$ is also a covered edge in $BN2$ by Corollary 1, it follows that $pa^{'}(X) = pa^{'}(Y) \cup \{Y\}$. In the fraction factorization of $BN2$, we must have $\frac{p(X,\ pa^{'}(X))}{p(pa^{'}(X))} \cdot \frac{p(Y,\ pa^{'}(Y))}{p(pa^{'}(Y))}$ (i.e., $p(X|pa^{'}(X)) \cdot p(Y|pa^{'}(Y))$), which can be simplified as $\frac{p(X,\ pa^{'}(X))}{p(pa^{'}(Y))}$. By lemma (1), it follows immediately that $\frac{p(Y,\ pa(Y))}{p(pa(X))}$ is identical to $\frac{p(X,\ pa^{'}(X))}{p(pa^{'}(Y))}$. This indicates that the fraction factorization for $BN1$ and $BN2$ are identical which further results in the fact that $BN1$ and $BN2$ have the same intrinsic factorization.

Corollary 2 All the BNs in the same equivalent class have the same intrinsic factorization.

Theorem (3) and corollary (2) indicate that the intrinsic factorization for an equivalent class of Bayesian networks is *unique* and it characterizes and describes the whole equivalent class of Bayesian networks algebraically.

4 Applications of the Algebraic Characterization

In this section, we describe some applications of the algebraic characterization.

4.1 Checking Equivalence of Two Bayesian Networks

Both [7] and [1] proposed graphical criterions for determining whether two given Bayesian networks are equivalent or not. In [7], in order to determine whether two

Bayesian networks are equivalent or not, a graphical structure, called *rudimentary pattern* of the DAG, is constructed for each of the DAG to be determined, two Bayesian networks are equivalent if and only if they have the same rudimentary pattern.

Under our algebraic characterization, given two Bayesian networks $BN1$ and $BN2$, according to theorem 3 and Corollary 2, we can first compute their intrinsic factorizations respectively. If their intrinsic factorizations are identical, then $BN1$ and $BN2$ are equivalent. This checking procedure is a pure algebraic one and does not involve the construction of a secondary graphical structure, i.e., rudimentary pattern.

4.2 Identifying Compelled Edge in a DAG

Compelled edge identification is very important for learning Bayesian networks as discussed in [1]. An algorithm for identifying all compelled edges in a given DAG was given in [1]. This algorithm consists of two sub-algorithms, one for ordering all the edges in a DAG, the other then labels all the edges in the DAG as either *compelled* or *reversible*. In this section, we propose a much simple method for the same purpose but using the algebraic characterization. The idea is very simple. Recall in example (1), we have demonstrated that from the intrinsic factorization in equation (4), we could derive all Bayesian networks that are equivalent to the one shown in Figure 1 (a) by absorbing all the denominators in equation (4) differently. We now formalize this method of denominator absorption and the notion of denominator absorption scheme.

Considering a Bayesian network $BN = (\mathcal{D},\ P)$, there is no difficulty at all to obtain its intrinsic factorization as shown below:

$$p(X_1X_2\dots X_n) \;=\; \prod_{i,\ j}\frac{p(X_i,\ pa(X_i))}{p(pa(X_j))}, \tag{20}$$

where $\{X_i,\ pa(X_i)\} \neq \{pa(X_j)\}$, for any $1 \leq i,\ j \leq n$. The equation above describes the class of equivalent Bayesian networks according to Corollary 2. In order to obtain each Bayesian network in the equivalent class, we only need to turn the intrinsic factorization in equation (20) into different Bayesian factorizations each of which corresponds to a Bayesian network in the equivalent class. It is important to note that turning the intrinsic factorization in equation (20) into a Bayesian factorization actually means absorbing all denominators in equation (20). More precisely, let

$$N = \{p(X_{i_1},\ pa(X_{i_1})),\ p(X_{i_2},\ pa(X_{i_2})),\ \ldots,\ p(X_{i_m},\ pa(X_{i_m}))\},$$

where each element of the set N is a numerator in equation (20). Similarly, let

$$D = \{p(pa(X_{j_1})),\ p(pa(X_{j_2})),\ \ldots,\ p(pa(X_{j_n}))\},$$

where each element of the set D is a denominator in equation (20). Absorbing denominators in D can be considered as a mapping such that each element $p(pa(X_{j_k}))$, $1 \leq k \leq n$, of D is being mapped to a *distinct* element $p(X_{i_{k'}},\ pa(X_{i_{k'}}))$, $1 \leq k' \leq m$, of N, where $pa(X_{j_k}) \subset \{X_{i_{k'}},\ pa(X_{i_{k'}})\}$. We call such a mapping a *denominator absorption scheme*. Since the number of numerators and the number of denominators in the intrinsic factorization (20) is finite, we can enumerate all possible denominator absorption schemes. If a denominator $d = p(pa(X_{j_k})) \in D$ is always mapped (absorbed) to a numerator $c = p(X_{i_{k'}},\ pa(X_{i_{k'}}))$ in all possbile denominator absorption schemes, it then follows immediately that regardless of which denominator absorption scheme one choose to yield Bayesian factorization, this d will always be absorbed by c to yield the conditional $p(\{X_{i_{k'}},\ pa(X_{i_{k'}})\} - pa(X_{j_k}) \mid pa(X_{j_k}))$ in every resulting Bayesian factorization obtained by following any denominator absorption scheme. This indicates that the edges $(X_i,\ X_j)$, where $X_i \in pa(X_{j_k})$, $X_j \in \{X_{i_{k'}},\ pa(X_{i_{k'}})\} - pa(X_{j_k})$ appears in every Bayesian network in the equivalent class, i.e., they are compelled edges.

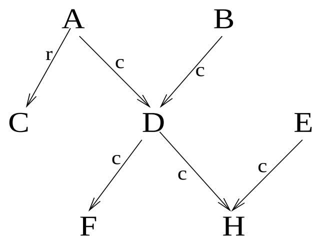

Figure 2: Equivalent Bayesian networks. The edge labelled with ‘r‘ is reversible. The edge labelled with ‘c‘ is compelled.

Example 2 Consider the Bayesian network shown in Figure 2. Its intrinsic factorization is as follows in equation (21):

$$p(ABCDEFH) = p(A) \cdot p(B) \cdot p(C|A) \cdot p(D|AB) \cdot p(E)$$

$$\begin{aligned} & \cdot p(F|D) \cdot p(H|DE), \\ = \; & p(B) \cdot p(CA) \cdot \frac{p(DAB)}{p(AB)} \cdot p(E) \cdot \frac{p(FD)}{p(D)} \cdot \\ & \frac{p(HDE)}{p(DE)}. \end{aligned} \tag{21}$$

Let $N = \{p(B),\ p(CA),\ p(DAB),\ p(E),\ p(FD),\ p(FDE)\}$ be the set of numerators and $D = \{p(AB),\ p(D),\ p(DE)\}$ be the set of denominators in equation (21), respectively. In order to turn equation (21) into a Bayesian factorization, it is easy to verify that in all possible denominator absorption schemes, denominator $p(AB)$ must be absorbed by $p(DAB)$, denominator $p(D)$ must be absorbed by $p(FD)$, and denominator $p(DE)$ must be absorbed by $p(HDE)$. This means that the edges $(A,\ D)$, $(B,\ D)$, $(D,\ F)$, $(D,\ H)$, and $(E,\ H)$ are compelled edges.

An algorithm can easily be designed to find out all compelled edges in a DAG, due to the space limit, the complete algorithm will appear in an forthcoming separate paper.

5 Conclusion

In this paper, we have proposed an algebraic characterization for describing and characterizing an equivalent class of Bayesian networks. It has been demonstrated that this new characterization really capture the essence of equivalent Bayesian networks algebraically without resorting to any other secondary graphical strucutres and criterition. Based on this characterization, we also proposed a method for identifying all compelled edges in a DAG. This method can be easily understood and implemented, compared with previously developed algorithm.

References

[1] D.M. Chickering. A transformational characterization of equivalent bayesian network structures. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 87–98. Morgan Kaufmann Publishers, 1995.

[2] A.P. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society*, 41B:1–31, 1979.

[3] F.V. Jensen. *An Introduction to Bayesian Networks.* UCL Press, 1996.

[4] F.V. Jensen, S.L. Lauritzen, and K.G. Olesen. Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly*, 4:269–282, 1990.

[5] S.L. Lauritzen and D.J. Spiegelhalter. Local computation with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, 50:157–244, 1988.

[6] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Francisco, California, 1988.

[7] T. Verma and J. Pearl. Equivalence and synthesis of causal models. In *Sixth Conference on Uncertainty in Artificial Intelligence*, pages 220–227. GE Corporate Research and Development, 1990.