A CHARACTERIZATION OF MARKOV EQUIVALENCE CLASSES FOR ACYCLIC DIGRAPHS[1]

By Steen A. Andersson, David Madigan
and Michael D. Perlman

Indiana University and University of Washington

Undirected graphs and acyclic digraphs (ADG's), as well as their mutual extension to chain graphs, are widely used to describe dependencies among variables in multivariate distributions. In particular, the likelihood functions of ADG models admit convenient recursive factorizations that often allow explicit maximum likelihood estimates and that are well suited to building Bayesian networks for expert systems. Whereas the undirected graph associated with a dependence model is uniquely determined, there may be many ADG's that determine the same dependence (i.e., Markov) model. Thus, the family of all ADG's with a given set of vertices is naturally partitioned into Markov-equivalence classes, each class being associated with a unique statistical model. Statistical procedures, such as model selection or model averaging, that fail to take into account these equivalence classes may incur substantial computational or other inefficiencies. Here it is shown that each Markov-equivalence class is uniquely determined by a single chain graph, the *essential graph*, that is itself simultaneously Markov equivalent to all ADG's in the equivalence class. Essential graphs are characterized, a polynomial-time algorithm for their construction is given, and their applications to model selection and other statistical questions are described.

1. Introduction. The use of directed graphs to represent possible dependencies among statistical variables dates back to Wright (1921) and has generated considerable research activity in the social and natural sciences. Since 1980, particular attention has been directed to graphical Markov models specified by conditional independence relations among the variables, that is, by the Markov properties determined by the graph. Both directed and undirected graphs have found extensive applications, the latter in such areas as spatial statistics and image analysis. Recent books by Whittaker (1990) and Lauritzen (1996) conveniently summarize the statistical perspective on these developments.

Graphical Markov models determined by acyclic directed graphs (ADG's) admit especially simple statistical analyses. In particular, ADG models admit convenient recursive factorizations of their joint probability density functions [Lauritzen, Dawid, Larson and Leimer (1990)], provide an elegant framework

for Bayesian analysis [Spiegelhalter and Lauritzen (1990)] and, in expert system applications, allow simple causal interpretations [Lauritzen and Spiegelhalter (1988)]. In the multinomial and multivariate normal cases, the likelihood function (i.e., both the joint probability density function and the parameter space) factorizes and admits explicit maximum likelihood estimates, which exist with probability 1 [Lauritzen (1996), Andersson and Perlman (1996)]. Furthermore, the only undirected graphical (UDG) models with these properties are the *decomposable* models, that is, the UDG models that have the same Markov properties as ADG models [Dawid and Lauritzen (1993), Andersson, Madigan and Perlman (1996)].

For these reasons, ADG models have become popular across an extraordinary range of applications; see, for example, Lauritzen and Spiegelhalter (1988), Pearl (1988), Neapolitan (1990), Spiegelhalter and Lauritzen (1990), Spiegelhalter, Dawid, Lauritzen and Cowell (1993), Madigan and Raftery (1994) and York, Madigan, Heuch and Lie (1995). Indeed, the vibrant "uncertainty in artificial intelligence" (UAI) community focuses much of its effort on ADG models.

Much of this applied work has adopted a Bayesian perspective: experts specify a prior distribution on competing ADG models. These prior distributions are combined with likelihoods (typically integrated over parameters) to give posterior model probabilities. Model selection algorithms then seek out the ADG models with highest posterior probability, and subsequent inference proceeds conditionally on these selected models [Cooper and Herskovits (1990), Buntine (1994), Spiegelhalter, Dawid, Lauritzen, and Cowell (1993), Heckerman, Geiger and Chickering (1994), Madigan and Raftery (1994)]. Non-Bayesian model selection methods proceed in a similar manner, replacing posterior model probabilities by, for example, penalized maximum likelihoods [Chickering (1995)].

Heckerman, Geiger and Chickering (1994) highlighted a fundamental problem with this general approach. Because several different ADG's may determine the *same* statistical model, that is, may determine the same set of conditional independence restrictions among a given set of random variates, the collection of all possible ADG's for these variates naturally coalesces into one or more classes of *Markov-equivalent* ADG's, where all ADG's within a Markov-equivalence class determine the same statistical model. Model selection algorithms that ignore these equivalence classes face three main difficulties.

1. Repeating analyses for equivalent ADG's leads to significant computational inefficiencies.
2. Ensuring that equivalent ADG's have equal posterior probabilities imposes severe constraints on prior distributions.
3. Weighting individual ADG's in Bayesian model averaging procedures to achieve specified weights for all Markov-equivalence classes is impractical without an explicit representation of these classes.

Treating each Markov-equivalence class as a single model would overcome these difficulties. As Heckerman, Geiger and Chickering (1994) have pointed out, however, a tractable characterization of these equivalence classes has not been available. In the present paper we show that for every ADG D, the equivalence class $[D]$ can be uniquely represented by a certain Markov-equivalent *chain graph* D^{*1}, the *essential graph* associated with the equivalence class.[2] Furthermore, we present an explicit characterization of those graphs G such that $G = D^*$ for some ADG D, then we apply this characterization to obtain a polynomial-time algorithm for constructing D^* from D. This characterization and construction lead to more efficient model selection and model averaging procedures for ADG models, based on essential graphs. Such procedures are discussed briefly in Section 7 and at greater length in Madigan, Andersson, Perlman and Volinsky (1996).

We suggest, therefore, that graphical modelers, both Bayesian and non-Bayesian, may wish to focus their attention on the class of essential graphs rather than ADG's.

Some basic definitions, terminology and results concerning graphs, graphical Markov models and their Markov equivalence are summarized in Appendices A and B, which the reader might review first. In Section 2 the essential graph D^* associated with an ADG D is formally defined and illustrated. Section 3 introduces the notions of irreversible, protected, and strongly protected arrows and relates these to the essential arrows of D, that is, the arrows of D^*.

In Section 4 we show first that D^* is a chain graph, each of whose chain components induces a chordal UDG (Proposition 4.1). Every $D' \in [D]$ can be recovered from D^* by orienting the edges of each (chordal) chain component of D^* in all possible "perfect" ways (Proposition 4.2). The chain graph D^* is itself Markov equivalent to D (Proposition 4.3).

Theorem 4.1, the main result of Section 4, applies Proposition 4.1 to obtain an explicit characterization of those graphs G that can occur as the essential graph D^* for some ADG D. Corollaries 4.1 and 4.2 characterize those UDGs and digraphs that can occur as essential graphs D^* for some ADG D. These results in turn lead to Proposition 4.5, which can be applied to establish the irreducibility of certain Markov chains used for Monte Carlo search procedures over the space of essential graphs (see Section 7).

A polynomial-time algorithm for constructing D^* from D is presented in Section 5.[3] The validity of our algorithm is established in Theorem 5.1 by means of our characterization of essential graphs. In Section 6 we exhibit all essential graphs on four or fewer vertices and note that the number of essential graphs is substantially smaller than the number of ADG's.

In Section 7 we indicate how the Markov-equivalence classes and their associated essential graphs can be used to overcome the three difficulties listed above that complicate model selection and model averaging for ADG models. We also briefly discuss model-search procedures based on equivalence classes and essential graphs.

Markov dependence models determined by chain graphs recently were introduced and developed by Frydenberg and Lauritzen (1989), Lauritzen and Wermuth (1989) and Frydenberg (1990); also see Andersson, Madigan and Perlman (1996). The introduction of chain graphs followed earlier work in this direction by Goodman (1973), Asmussen and Edwards (1983) and Kiiveri, Speed and Carlin (1984). Chain graphs provide much of the focus for current research on modeling statistical dependence; see, for example, Wermuth and Lauritzen (1990) and Cox and Wermuth (1993, 1996). The fact that the essential graph D^* associated with an ADG D is a chain graph that is Markov equivalent to D allows us to conduct statistical inference in the space of essential graphs, rather than in the larger space of individual ADG's; see Section 7, especially (7.2).

2. Markov equivalence of acyclic digraphs: the essential graph D^*. Our development begins with a well-known graph-theoretic criterion for the Markov equivalence of ADG's, given in Theorem 2.1. This was discovered by Verma and Pearl [(1990, Theorem 1; (1992), Corollary 3.2] and, independently, by Frydenberg [(1990), Theorem 5.6] for the more general class of chain graphs; also see Andersson, Madigan and Perlman [(1996), Theorem 3.1]. Frydenberg's result is stated as Theorem B.1 of our Appendix B. For completeness, in Appendix B we also present a direct proof of Theorem 2.1, different from that of Verma and Pearl.

THEOREM 2.1. *Two ADG's are Markov equivalent if and only if they have the same skeleton and the same immoralities* (*see Figure* 1).

We say that two ADG's D_1 and D_2 are *graphically equivalent*, and write $D_1 \sim D_2$, if they have the same skeleton and the same immoralities. By Theorem 2.1, D_1 and D_2 are Markov equivalent if and only if they are graphically equivalent; thus we shall use the term *equivalent* for both notions. The equivalence class containing D is denoted by $[D]$.

While Theorem 2.1 provides a practical criterion for deciding whether two given ADG's are Markov equivalent, it does not directly yield a characterization of the entire equivalence class $[D]$ for a given ADG D. Consider, for example, the following question regarding the nontransitive ADG D_1 in Figure 2: does $[D_1]$ contain a *transitive* ADG? [For the statistical relevance of
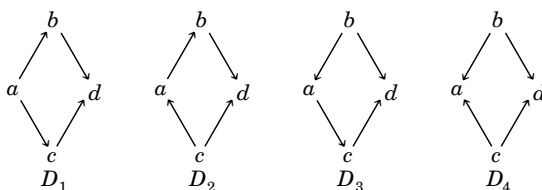

FIG. 1. *The four ADG's with the same skeleton as D_1 and the immorality (b, d, c). The ADG's D_1, D_2, and D_3 have no other immoralities, hence are Markov equivalent by Theorem 2.1. The ADG D_4 has the additional immorality (b, a, c), hence is not Markov equivalent to the others. Thus, $[D_1] = \{D_1, D_2, D_3\}$.*
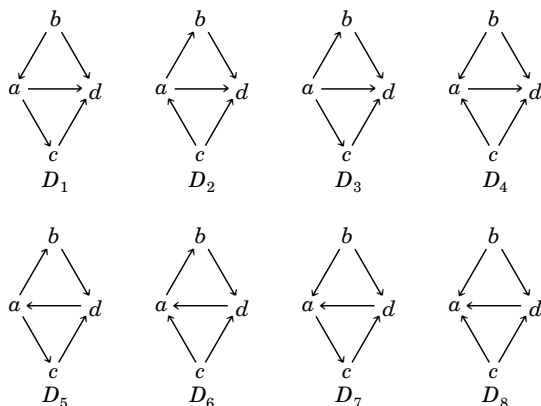
FIG. 2. *The $2^3 = 8$ possible digraphs with the same skeleton as D_1 and the immorality (b, d, c). Of these 8, D_5, D_6, and D_7 are not acyclic, while D_4 and D_8 are acyclic but possess the additional immorality (b, a, c), so $[D_1] = \{D_1, D_2, D_3\}$.*

this question, see Andersson, Madigan, Perlman and Triggs (1995).] Theorem 2.1 does not allow us to answer this question by direct inspection of D_1; instead, we must first determine all members of $[D_1]$, then check each member for transitivity, as follows. Since (b, d, c) is an immorality in D_1, the arrows $b \rightarrow d$ and $c \rightarrow d$ are *essential* in D_1, that is, these arrows must occur in each member of $[D_1]$. The other three edges of D_1 can be oriented in $2^3 = 8$ possible ways, as shown in Figure 2; of these eight digraphs, only five are acyclic, and of these five, only three (D_1, D_2, D_3) possess the same immorality as D_1 and no other. Thus, $[D_1] = \{D_1, D_2, D_3\}$; hence $[D_1]$ does contain a transitive ADG, namely D_3.

Since the number of possible orientations of all arrows that do not participate in any immorality of an ADG D grows exponentially with the number of such arrows, hence superexponentially with the number of vertices, determination of the equivalence class $[D]$ by exhaustive enumeration of possibilities, as in the preceding example, rapidly becomes computationally infeasible as the size of D increases. A closer examination of this example reveals, however, that the arrow $a \rightarrow d$ occurs in every member of $[D_1]$, hence is an essential arrow of D_1 even though it is not involved in any immorality of D_1. Had we been able to identify all three essential arrows of D_1 directly from D_1 itself, it would not have been necessary to consider D_5–D_8 in order to determine $[D_1]$. On the other hand, it appears necessary to determine $[D_1]$ before we can identify the essential arrows of D_1.

Fortunately, this is not the case. A main purpose of the present paper is to develop a polynomial-time algorithm (Section 5) for determining all essential arrows of an ADG D. This is done by introducing and characterizing the essential graph D^* associated with D. Furthermore, questions such as the existence of a transitive member of $[D]$ can be answered by a polynomial-time inspection of D^* itself, without the need for an exhaustive search of $[D]$ [Andersson, Madigan, Perlman and Triggs (1996)].

DEFINITION 2.1. The essential graph D^* associated with D is the graph

$$D^* := \bigcup(D'|D' \sim D),$$

that is, D^* is the smallest graph larger than every $D' \in [D]$.

Thus, D^* is the graph with the same skeleton as D, but where an edge is directed in D^* iff it occurs as a directed edge (that is, arrow) with the same orientation in every $D' \in [D]$; all other edges of D^* are undirected. (See Figure 3 for examples.) The directed edges (that is, arrows) in D^* are called the *essential arrows* of D. Clearly, every arrow that participates in an immorality in D is essential, but D may contain other essential arrows as well, for example, the arrow $a \to d$ in the second graph in Figure 3 and the arrows $a \to d$ and $b \to d$ (verify) in the third graph in Figure 3.[4] We will show that D^* is a chain graph (Proposition 4.1) that is itself Markov equivalent to D (Proposition 4.3), so that D^* contains the same statistical information as D.[5] (Note that D and D^* have the same skeleton and immoralities, so that $D_1 \sim D_2$ iff $D_1^* = D_2^*$.) The complete characterization of essential graphs in Theorem 4.1 involves further restrictions on the configurations of arrows and lines (equivalent to undirected edges) that can occur in D^*.

3. First characterization of the essential arrows of D. By Definition 2.1, an arrow $a \to b$ in an ADG D is essential iff $a \to b \in D'$ for each $D' \in [D]$. Proposition 3.1 shows that, in addition, $a \to b$ must be *protected* in each $D' \in [D]$, that is, must occur in each D' in at least one of the three configurations (a), (b) and (c) shown after Definition 3.2.

To begin, note that an essential arrow $a \to b$ must be *irreversible* in D.

DEFINITION 3.1. Let G be a chain graph. An arrow $a \to b \in G$ is *irreversible* in G if changing $a \to b$ to $a \leftarrow b$ either creates or destroys an immorality or creates a directed cycle.

To determine whether an arrow $a \to b$ is irreversible in G according to Definition 3.1, global knowledge of G is required, since directed cycles of arbitrary length must be considered. For a characterization of irreversibility to be computationally feasible, however, it must be local, that is, must only require consideration of directed cycles of bounded length. For an ADG D,
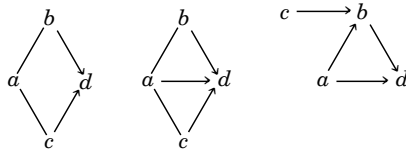

FIG. 3.   *Three examples of essential graphs D^*. In the first example, D is the ADG D_1 of Figure* 1. *In the second example, D is the ADG D_1 of Figure* 2. *In the third example, $D = D^*$ (see Corollary* 4.2).

Lemma 3.1(i) shows that in fact only directed cycles of length 3 need be considered. The following definition is required.

DEFINITION 3.2. Let G be a graph. An arrow $a \to b \in G$ is *protected* in G if $\text{pa}_G(a) \neq \text{pa}_G(b) \setminus \{a\}$.

It is easy to see that $a \to b$ is protected in G if and only if $a \to b$ occurs in at least one of the following six configurations as an induced subgraph of G.
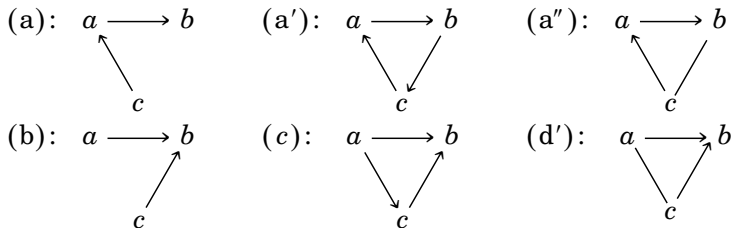

If G is a chain graph, then only (a), (b), (c) or (d') can occur; if $G \equiv D$ is an ADG, then only (a), (b) or (c) can occur. For a general graph G, $a \to b$ is protected in G iff $a \to b$ is protected in the directed graph $D(G)$ obtained by deleting all undirected edges (lines) in G [since $\text{pa}_G(a) = \text{pa}_{D(G)}(a)$].

The arrow $a \to b$ is irreversible in a chain graph G if and only if *either* $a \to b$ occurs in configuration (a) or (b) as an induced subgraph of G *or else* $a \to b$ blocks some directed cycle in G. If $a \to b$ is protected in a chain graph G, then clearly it is irreversible in G. If $G \equiv D$ is an ADG, then the converse is also true.

LEMMA 3.1. *Let D be an ADG.*

(i) *An arrow $a \to b$ is irreversible in D if and only if it is protected in D.*
(ii) *An arrow $a \to b$ is reversible in D if and only if the digraph D' obtained from D by replacing $a \to b$ by $a \leftarrow b$ is acyclic and $D' \sim D$.*

PROOF. (i) Suppose that $a \to b$ is irreversible in D by virtue of blocking some directed cycle in D:
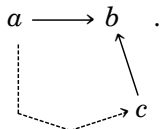
If no edge $a \cdots c$ is present in D then $a \to b$ already occurs in configuration (b) as an induced subgraph of D. If an edge $a \cdots c$ is present in D then either
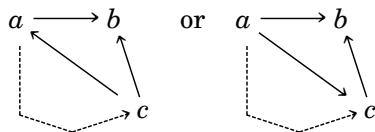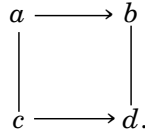

occurs in D. The first case is impossible since it contains a directed cycle.

Thus the second must hold, so $a \to b$ occurs in configuration (c) in D. Thus $a \to b$ is protected in D.

(ii) This assertion is immediate. □

Lemma 3.1(i) is not true for a general chain graph G; the following chain graph provides a counterexample:

$$
\begin{array}{ccc}
a & \longrightarrow & b \\
\big| & & \big| \\
c & \longrightarrow & d.
\end{array}
$$

PROPOSITION 3.1. *Let D be an ADG. An essential arrow $a \to b$ of D is protected in every $D' \in [D]$.*

PROOF. If $a \to b$ is an essential arrow of D then clearly $a \to b$ is irreversible in every $D' \in [D]$; hence, by Lemma 3.1(i), $a \to b$ is protected in every $D' \in [D]$. □

In Proposition 3.1, it is possible a priori that the third vertex c in the protecting configuration (a), (b) or (c) for the essential arrow $a \to b \in D$ may vary with D', that is, $c = c(D')$. In fact this is not the case, but the notion of "protected" must be extended.

DEFINITION 3.3. Let G be a graph. An arrow $a \to b \in G$ is *strongly protected* in G if $a \to b$ occurs in at least one of the following four configurations as an induced subgraph of G:

(a): $a \longrightarrow b$ (b): $a \longrightarrow b$ (c): $a \longrightarrow b$ (d): $a \longrightarrow b$ $(c_1 \neq c_2)$.
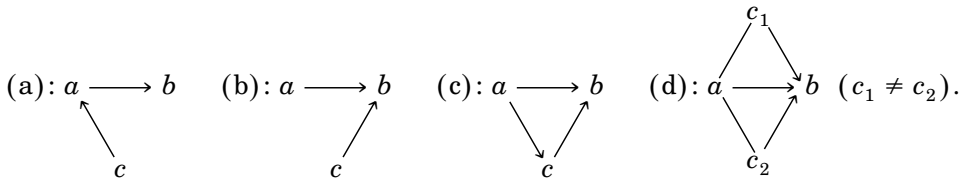
Since (d) ⇒ (d'), "strongly protected" implies "protected", while if $G \equiv D$ is an ADG, then "strongly protected" is equivalent to "protected." For a chain graph G, the definition of "strongly protected" differs from "protected" only in that (d) replaces (d'), but this difference is significant: by Theorem 4.1, every essential graph D^* must be a chain graph and every arrow in D^* (i.e., every essential arrow of D) must be strongly protected in D^* (see the examples in Figure 3). This characterization provides the basis for the polynomial-time algorithm in Section 5 for constructing D^* from D. (Also see Remark 5.1.)

In Corollary 4.2, it is shown that every arrow of an ADG D is essential (i.e., $D = D^*$) if and only if every arrow of D is protected in D. The third graph in Figure 3 provides an example.

The final lemma will be needed for the proof of Theorem 2.1 in Appendix B.

LEMMA 3.2.   *Let D, D' be two ADG's such that $D \sim D'$ but $D \neq D'$. Then there exists a finite sequence $D \equiv D_1, \ldots, D_k \equiv D'$ such that each $D_i \in [D]$ and each consecutive pair D_i, D_{i+1} differ in exactly one edge.*

PROOF.   By the definition of equivalence, D and D' have the same vertex set V and the same skeleton. Let $F \coloneqq \{a_1 \to b_1, \ldots, a_n \to b_n\} \neq \varnothing$ denote the set of edges in D that occur with the opposite orientation in D'. By Lemma 3.1(ii) and induction, it suffices to show that at least one $a_i \to b_i$ is reversible in D.

Suppose, to the contrary, that each $a \to b \in F$ is irreversible in D, hence by Lemma 3.1(i) is protected in D. Let b^* be a minimal element of $\{b_1, \ldots, b_n\}$ with respect to the partial ordering $(V, \leq)$ determined by the ADG D: $a \leq b$ if and only if $a = b$ or there exists a path from a to b in D. Let a^* be a maximal element of $\{a \in V | a \to b^* \in F\}$. Since $a^* \to b^* \in F$, $a^* \to b^*$ cannot occur in an immorality in D. Thus, because $a^* \to b^*$ is protected in D, $a^* \to b^* \in F$ must occur in D *either* in configuration (a) as an induced subgraph of D with $c \to a^* \in F$, *or else* in configuration (c) with either $a^* \to c \in F$ or $c \to b^* \in F$. But the first two possibilities violate the minimality of b^*, while the third violates the maximality of a^*. This completes the proof. $\square$

4. Characterization of the essential graph D^*.

Theorem 4.1, the main result of this section, gives necessary and sufficient conditions for a graph $G \equiv (V, E)$ to be the essential graph D^* for some ADG D. We begin by showing that such a G must be a chain graph. (Most proofs are deferred to the end of this section.)

Let D^{**} denote the smallest chain graph larger than every $D' \in [D]$. That is, D^{**} is the graph obtained from D^* by converting to undirected edges (that is, lines) all those directed edges in D^* that participate in a directed cycle in D^*. Note that this can be done in a single step: suppose that the arrow $a \to b$ occurs in a directed cycle in D^* and that, after converting $a \to b$ into a line, a second arrow $c \to d \in D^*$ now becomes part of a directed cycle:
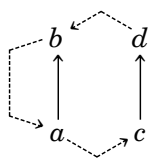
(possibly $a = c$ or $b = d$). Then $c \to d$ was already part of a directed cycle in D^* before $a \to b$ was converted to a line.

Clearly $D \subseteq D^* \subseteq D^{**}$. In fact, the second inclusion is an equality.

PROPOSITION 4.1.   (i) $D^* = D^{**}$, hence D^* is a chain graph.
(ii) For each chain component $\tau \in \mathbf{T}(D^*)$, the induced UDG $(D^*)_\tau$ is chordal.

Next, every ADG $D' \in [D]$ can be recovered from the essential graph D^*.

PROPOSITION 4.2.   A digraph D' is acyclic and equivalent to the ADG D if and only if D' is obtained from D^* by orienting the edges of each (chordal) chain component $(D^*)_\tau$ of D^* in any perfect way.

PROPOSITION 4.3.   Let D be an ADG and D^* its essential graph. Then D and D^* are Markov equivalent.

THEOREM 4.1 (Characterization of D^*).   A graph $G \equiv (V, E)$ is equal to D^* for some ADG D if and only if G satisfies the following four conditions.

(i) G is a chain graph.
(ii) For every chain component τ of G, G_τ is chordal.
(iii) The configuration $a \rightarrow b \text{ --- } c$ does not occur as an induced subgraph of G.
(iv) Every arrow $a \rightarrow b \in G$ is strongly protected in G.

Since both UDG's and ADG's are chain graphs, Theorem 4.1 immediately yields the following two corollaries.

COROLLARY 4.1.   Let G be a UDG. Then $G = D^*$ for some ADG D if and only if G is chordal.

COROLLARY 4.2.   Let G be a digraph. Then $G = D^*$ for some ADG D if and only if G is an ADG and every arrow of G is protected in G; in this case $G = D = D^*$.

PROOF.   Apply Theorem 4.1 and the fact that an arrow is protected in an ADG if and only if it is strongly protected in the ADG. (Note that the chain components of an ADG are just its vertices, hence trivially are chordal.)

The following is an example of an ADG D such that $D = D^*$:
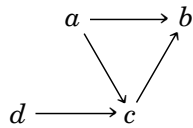
Clearly, each arrow of D is protected in D.
Let $G \equiv (V, E)$ be a chain graph. An arrow $a \rightarrow b$ is an *initial arrow* of G if a is minimal in $\{a' \in V | \exists\ b \in V \ni a' \rightarrow b \in E\}$ with respect to the pre-ordering $(V, \leq)$ determined by G. Note that G has no initial arrows iff G is a

UDG. Clearly an initial arrow $a \to b$ cannot occur in configuration (a) in G, so, if $G = D^*$ for some ADG D, then Theorem 4.1 implies that $a \to b$ must occur in configuration (b), (c) or (d) as an induced subgraph of G. Because D^* is determined by the immoralities of D, one might speculate that in this case, every initial arrow of G must in fact occur in configuration (b) or (d) as an induced subgraph of G, but this is not true in general: consider the initial arrow $a \to b$ of the chain graph (in fact, ADG) $G \equiv D \equiv D^*$ in the figure in the preceding paragraph. It is almost true, however, as seen by the following result, which provides a useful necessary condition for determining whether a given graph G is an essential graph.

PROPOSITION 4.4. *Suppose that $G = D^*$ for some ADG D. For every initial arrow $a \to b$ of G, there exists a vertex $c \in V$ such that $a \to c$ is also an initial arrow of G and $a \to c$ occurs in configuration* (b) *or* (d) *as an induced subgraph of G.*

COROLLARY 4.3. *An ADG D has no essential arrows (i.e., D^* is a UDG) if and only if D has no immoralities.*

PROOF. If D is moral then so is D^*, hence configurations (b) and (d) cannot occur in D^*. Proposition 4.4 implies that D^* has no initial arrows, hence D^* is a UDG. The converse is trivial. □

REMARK 4.1. An initial arrow in D^* need not be initial in D, or vice versa. Consider the following ADG D:

$$d \longrightarrow a \longrightarrow b$$
$$\nearrow$$
$$c \quad .$$
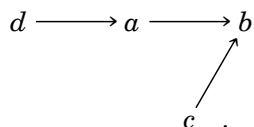
Then $a \to b$ is initial in the associated essential graph D^*:

$$d \longrightarrow a \longrightarrow b$$
$$\nearrow$$
$$c$$
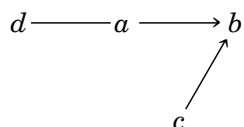
but not in D, whereas $d \to a$ is initial in D but does not occur in D^*.

The final result of this section can be applied to establish the irreducibility of certain Markov chains used for Monte Carlo search algorithms over the space of essential graphs; see Section 7.

PROPOSITION 4.5. *Let G and H be two essential graphs with the same vertex set V. Then there exists a finite sequence $G \equiv G_1, \ldots, G_k \equiv H$ of essential graphs with vertex set V such that each consecutive pair G_i, G_{i+1} dif-*

fer by either (i) *exactly one line* $a - b$, *or* (ii) *exactly one arrow* $a \rightarrow b$, *or* (iii) *exactly two arrows that form an immorality*: $a \rightarrow b \leftarrow c$.
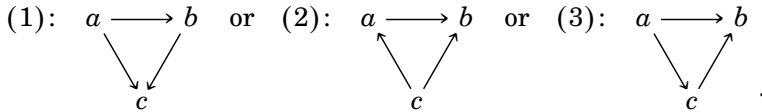
We turn to the proofs. The proof of Proposition 4.1 requires the following five facts.

FACT 1. The configuration $a \rightarrow b - c$ cannot occur as an induced subgraph of D^*.

PROOF. If $a \rightarrow b - c$ occurs as an induced subgraph in D^* (requiring that a and c are not linked), then $a \rightarrow b \leftarrow c$ must occur as an immorality in some $D' \sim D$, hence $b \leftarrow c$ must be an essential arrow, contradicting $b - c \in D^*$. □

FACT 2. If $a \underset{c}{\overset{\nearrow}{\rightarrow}} b$ occurs in D^*, then there exist $D_1, D_2 \in [D]$ such that

$a \underset{c}{\overset{\searrow\!\swarrow}{\rightarrow}} b$ occurs in D_1 and $a \underset{c}{\overset{\nwarrow\!\nearrow}{\rightarrow}} b$ occurs in D_2.

PROOF. Any $D' \in [D]$ must contain either

$$(1): \quad a \overset{\searrow\!\swarrow}{\underset{c}{\longrightarrow}} b \quad \text{or} \quad (2): \quad a \overset{\nwarrow\!\nearrow}{\underset{c}{\longrightarrow}} b \quad \text{or} \quad (3): \quad a \overset{\searrow\!\nearrow}{\underset{c}{\longrightarrow}} b.$$

If (1) were to occur in *no* $D' \in [D]$, then necessarily $c \rightarrow b \in D^*$, contradicting the hypothesis. Thus (1) must occur in some $D_1 \in [D]$. Similarly, (2) must occur in some $D_2 \in [D]$. □

FACT 3. D^{**} has the same immoralities as D (hence, as D^*).

PROOF. Recall that D^{**} is obtained by converting all arrows that occur in directed cycles in D^* into lines. It is evident that D^* has the same immoralities as D. Since $D^* \subseteq D^{**}$, D^{**} can have the same or fewer immoralities than D^*. We shall show it impossible that an immorality $a \rightarrow b \leftarrow c$ occurs in D^* while $a - b \in D^{**}$.

If this were to happen, then $a \rightarrow b$ would be part of a directed cycle $(a, b \equiv b_0, b_1, \ldots, b_k \equiv a)$ in D^*, where $k \geq 2$ and where each edge $b_{i-1} \cdots b_i$ in the cycle occurs as either $b_{i-1} - b_i$ or $b_{i-1} \rightarrow b_i$, $1 \leq i \leq k$. (In particular, $b_1 \neq a, c$.) See the figure:

$$b_k \equiv a \longrightarrow b \equiv b_0 \longleftarrow c.$$
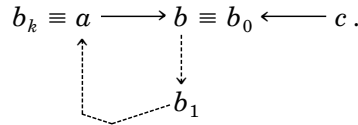
Case 1. Suppose that $b - b_1 \in D^*$. Then there exist ADG's $D_1, D_2 \in [D]$ such that $b_1 \to b \in D_1$ and $b_1 \leftarrow b \in D_2$. Since neither $a \to b \leftarrow b_1$ nor $b_1 \to b \leftarrow c$ can occur as an immorality in D_1, there must be edges $a \cdots b_1$ and $c \cdots b_1$ in D_1. To avoid a cycle, necessarily $a \to b_1 \in D_2$ and $c \to b_1 \in D_2$, so $a \to b_1 \leftarrow c$ forms an immorality in D_2, hence also in D^*. Thus we have a *shorter* directed cycle $(a, b_1, \ldots, b_k \equiv a)$ in D^* such that the immorality $a \to b_1 \leftarrow c$ occurs in D^* but $a - b_1 \in D^{**}$.

Case 2. Suppose that $b \to b_1 \in D^*$. Since D contains no directed cycles, at least one edge in the cycle $(a, b \equiv b_0, b_1, \ldots, b_k \equiv a)$ must be undirected in D^*. Consider the smallest i such that $b_{i-1} - b_i \in D^*$. This i satisfies $2 \le i \le k$ and $b_{i-2} \to b_{i-1} - b_i$ occurs in D^*. By Fact 1, there must be an edge $b_{i-2} \cdots b_i$ in D^*. But $b_{i-2} \leftarrow b_i \notin D^*$, since there is some ADG $D' \in [D]$ containing $b_{i-1} \to b_i$ that consequently would contain a directed triangle. Therefore, either $b_{i-2} - b_i \in D^*$ or $b_{i-2} \to b_i \in D^*$, again producing a *shorter* directed cycle $(a, b_0, \ldots, b_{i-2}, b_i, \ldots, b_k \equiv a)$ in D^* such that the immorality $a \to b_0 \leftarrow c$ occurs in D^* but $a - b_0 \in D^{**}$.

Thus, Cases 1 and 2 together allow us to proceed by induction to reduce to the case where the immorality $a \to b \leftarrow c$ occurs in D^* but $a \to b$ occurs in a directed triangle (a, b, d) in D^* (necessarily, $d \ne c$). The only type of directed triangle (a, b, d) in D^* that does not imply the contradictory existence of an ADG $D' \in [D]$ such that (a, b, d) comprises a directed triangle in D' is pictured here:

$$a \longrightarrow b \longleftarrow c$$
$$\diagdown \quad \diagup$$
$$d$$
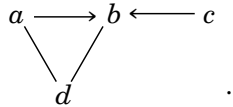
By Fact 2, there exist ADG's $D_1, D_2 \in [D]$ with $a \to d$, $b \to d$ in D_1 and $a \leftarrow d$, $b \leftarrow d$ in D_2. Thus there must be an edge $c \cdots d$ in D_2. (Otherwise $d \to b \leftarrow c$ would form an immorality in D_2, forcing $d \to b \in D^*$, contradicting the occurrence of the undirected edge $d - b$ in D^*). Since the edge $c \cdots d$ must be present in D_1 also, it must be oriented there as $c \to d$ [otherwise (c, b, d) would form a directed triangle]. Thus the configuration

$$a \longrightarrow b \longleftarrow c$$
$$\diagdown \quad \downarrow \quad \diagup$$
$$d$$
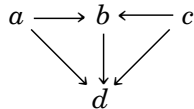
must occur in D_1. This produces the immorality $a \to d \leftarrow c$ in D_1, forcing $a \to d \in D^*$, contradicting the occurrence of $a - d$ in D^*. This establishes Fact 3. □

FACT 4. D^* and D^{**} have no undirected chordless k-cycles, $k \ge 4$.

PROOF. If an undirected chordless k-cycle, $k \ge 4$, occurs in D^* or in D^{**}, then D must have at least one immorality in this cycle. This immorality must also occur in D^*, hence, by Fact 3, also in D^{**}, contradicting the assumption that the cycle is undirected. □

FACT 5.   The configuration $a \to b - c$ cannot occur as an induced sub-graph of D^{**} (i.e., a and c are not linked).

PROOF.   Suppose that $a \to b - c$ occurs as an induced subgraph in D^{**}. Then $a \to b \in D^*$ and hence $a \to b \in D'$ for all $D' \in [D]$. Thus $b \leftarrow c \notin D'$ for all $D' \in [D]$ (otherwise $a \to b \leftarrow c$ forms an immorality in D', hence in D^{**} by Fact 3), so $a \to b \to c$ occurs as an induced subgraph in all $D' \in [D]$, hence also in D^*. Therefore $b \to c$ must be part of a directed cycle $(b, c \equiv c_0, c_1, \ldots, c_k \equiv b)$ in D^* (see the following figure), $k \geq 2$, where, for $1 \leq i \leq k$, the edge $c_{i-1} \cdots c_i$ is either $c_{i-1} - c_i$ or $c_{i-1} \to c_i$. (Note that $c_1 \neq a, b$.)

$$a \longrightarrow b \equiv c_k \longrightarrow c \equiv c_0$$
$$c_1$$
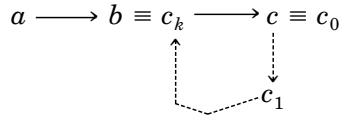
Case 1. Suppose that $c - c_1 \in D^*$. Then there exist ADG's $D_1, D_2 \in [D]$ such that $c \to c_1 \in D_1$ and $c \leftarrow c_1 \in D_2$. Therefore there must be an edge $b \cdots c_1$ in D_2 (else $c \leftarrow c_1$ participates in an immorality), hence also in D_1 and D^*. To avoid a directed cycle, this edge must appear as $b \to c_1$ in D_1. If there were an edge $a \cdots c_1$ in D_1, it must be $a \to c_1$ [otherwise (a, b, c_1) would comprise a directed triangle in D_1], which would imply the immorality $a \to c_1 \leftarrow c$ in D_1, contradicting $c - c_1 \in D^*$. Thus, there is no edge connecting a and c_1 in D_1, hence none in D^*. Therefore the edge $b \cdots c_1$ cannot occur in D^* as $b - c_1$ (by Fact 1) or as $b \leftarrow c_1$ (since $b \to c_1 \in D_1$), hence $b \to c_1 \in D^*$. Thus $a \to b \to c_1$ also occurs as an induced subgraph in D^*, so $b \to c_1$ occurs in a *shorter* directed cycle $(b, c_1, \ldots, c_k \equiv b)$ in D^*.

Case 2. Suppose that $c \to c_1 \in D^*$. Consider the smallest $i \geq 2$ such that $c_{i-1} - c_i \in D^*$. Thus $c_{i-2} \to c_{i-1} - c_i$ occurs in D^*, so by Fact 1, there must be an edge $c_{i-2} \cdots c_i$ in D^*. As in Case 2 of Fact 3, either $c_{i-2} - c_i \in D^*$ or $c_{i-2} \to c_i \in D^*$. Thus $a \to b \to c_0$ occurs as an induced subgraph in D^*, hence $b \to c_0$ occurs in a *shorter* directed cycle $(b, c_0, \ldots, c_{i-2}, c_i, \ldots, c_k \equiv b)$ in D^*.

Cases 1 and 2 together allow us to proceed by induction to reduce to the situation where $a \to b \to c$ occurs as an induced subgraph in D^* but $b \to c$ participates in a directed triangle (b, c, d) in D^*:

$$a \longrightarrow b \longrightarrow c$$
$$d$$
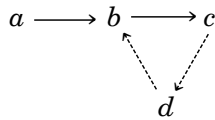
(necessarily, $d \neq a$). The only such directed triangle in D^* that does *not* imply the existence of an ADG $D' \in [D]$, such that (b, c, d) comprises a directed triangle in D', is pictured here:

$$a \longrightarrow b \longrightarrow c$$
$$d \qquad .$$
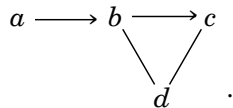
By Fact 2, there exist ADG's $D_1, D_2 \in [D]$ with $b \to d \leftarrow c$ in D_1 and $b \leftarrow d \to c$ in D_2. Thus there must be an edge $a \cdots d$ in D_2 (otherwise $a \to b \leftarrow d$ would form an immorality in D_2, forcing $b \leftarrow d \in D^*$, contradicting the occurrence of the undirected edge $b - d \in D^*$). The edge $a \cdots d$ also must be present in D_1, where it must be oriented as $a \to d$ so that (a, b, d) does not form a directed triangle. Thus the configuration

$$a \longrightarrow b \longrightarrow c$$
$$\searrow \quad \downarrow \quad \swarrow$$
$$d$$
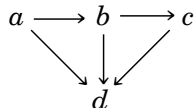
must occur in D_1. This produces the immorality $a \to d \leftarrow c$ in D_1, forcing $d \leftarrow c \in D^*$, contradicting the occurrence of the undirected edge $d - c$ in D^*. Fact 5 is proved. □

PROOF OF PROPOSITION 4.1. (i) We know that $D^* \subseteq D^{**}$. To show that $D^* = D^{**}$, it suffices to show that if an undirected edge $a - b \in D^{**}$, then also $a - b \in D^*$.

Let τ be the unique chain component of D^{**} such that $a - b \in (D^{**})_\tau$. By Fact 4, $(D^{**})_\tau$ is a chordal UDG. Therefore (see Appendix A) $(D^{**})_\tau$ admits two perfect directed versions, D_1 and D_2, such that $a \to b \in D_1$ and $a \leftarrow b \in D_2$.

Now assign perfect orientations to the edges within all other chain components of D^{**}, obtaining two directed graphs, D' and D''. These have the same skeleton as D, D^* and D^{**} and satisfy the following conditions: (1) all arrows in D^{**} also occur as arrows in D' and D''; (2) $(D')_\tau = D_1$ and $(D'')_\tau = D_2$, so $a - b \in D' \cup D''$.

Both D' and D'' are acyclic. For, if D' or D'' has a directed cycle, at least one of the arrows in this cycle must be an arrow in D^{**} (otherwise the cycle must lie entirely within one chain component of D^{**}, hence cannot be directed). Thus if we convert back into lines all arrows in this cycle that came from lines in D^{**}, at least one arrow remains, giving a directed cycle in D^{**}, contradicting its chain graph property.

Next, D' and D'' have the same immoralities as D, D^* and D^{**}, so D' and $D'' \in [D]$. To see this, begin by noting that, since D' and $D'' \subseteq D^{**}$, every immorality in D^{**} must also occur in D' and D''. Suppose that $a \to b \leftarrow c$ is an immorality in D' or D''. This immorality could not have arisen from the configuration $a - b - c$ in D^{**}, since the edges within each chain component of D^{**} are perfectly oriented in D' and D'', nor, by Fact 5, could it have arisen from the configurations $a \to b - c$ or $a - b \leftarrow c$ in D^{**}. Thus the immorality $a \to b \leftarrow c$ must also occur in D^{**}.

Finally, since D' and $D'' \in [D]$, necessarily $D' \cup D'' \subseteq D^*$. But $a - b \in D' \cup D''$, hence $a - b \in D^*$. This completes the proof of (i). Part (ii) follows from Fact 4. □

PROOF OF PROPOSITION 4.2. Since $D^* = D^{**}$, the "if" assertion is established in the proof of Proposition 4.1. To verify "only if," suppose that

$D' \in [D]$. Then any arrow in D^* also occurs in D', while D' can have no immoralities within any chain component of D^* (since D' and D^* have the same immoralities), hence the restriction of D' to each chain component of D^* is perfect. $\square$

PROOF OF PROPOSITION 4.3.   By Proposition 4.1, D^* is a chain graph. Since D and D^* have the same skeleton, by Theorem B.1 of Appendix B it suffices to show that D and D^* have the same minimal complexes. By Fact 3, they have the same immoralities. By Fact 1, D^* can have no minimal complexes other than immoralities; trivially, neither can D, since it is an ADG. $\square$

PROOF OF THEOREM 4.1 ("only if").   Proposition 4.1 implies properties (i) and (ii), while (iii) follows from Fact 1. Property (iv) will be established by means of the following two facts regarding the essential arrows of D. [See Section 3 for the definitions of configurations (a)–(d) and (d').]

FACT 6.   Every essential arrow $a \to b$ of D occurs in at least one of the configurations (a), (b), (c) or (d') as an induced subgraph of D^*. Thus, $a \to b$ is irreversible in D^*.

PROOF.   Suppose that $a \to b \in D^*$ but satisfies neither (a), (b), (c) nor (d') in D^*. Consider the two (distinct) chain components τ_a and τ_b of D^* that contain a and b, respectively. By property (ii), we can construct a directed graph D' from D^* by assigning arbitrary perfect orientations to the edges of $(D^*)_\tau$ for every chain component τ other than τ_a and τ_b, and by assigning perfect orientations starting at a (resp., b) to the edges within τ_a (resp., τ_b), so that all edges within τ_a (τ_b) that involve a (b) are oriented outward from a (b) (see the following figure). By Proposition 4.2, D' is an ADG and $D' \in [D]$.
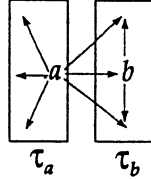
Now construct another directed graph D'', which is identical to D' except that $a \to b \in D'$ is changed to $a \gets b$ in D''. Then D'' is also acyclic, for if it were to contain a directed cycle, then this cycle must include $a \gets b$, hence must include a subgraph $a \gets b \gets c$ of D'' with $c \neq a$. Necessarily $c \notin \tau_b$, since all arrows of D'' within τ_b are oriented outward from b, so $b \gets c \in D^*$. Thus $a \to b \gets c$ occurs in D^*, so, since $a \to b$ cannot satisfy (b) in D^*, there must be an edge $a \cdots c$ in D^*. This edge cannot be $a \to c$ or $a - c$,

otherwise $a \to b$ would satisfy (c) or (d′) in D^*, hence it must appear as $a \leftarrow c$ in D^*. Thus $a \leftarrow c$ must also occur in D'', so the assumed directed cycle in D'' must have contained at least four vertices. Therefore, removing the vertex b from this cycle leaves another directed cycle in D'', which must also occur in D' since D' and D'' coincide except for the edge $a \cdots b$. This is a contradiction, so we conclude that D'' is acyclic.

We shall show that D'' has the same immoralities as D'. If an immorality $c \to a \leftarrow b$ is created in D'' when $a \to b$ is changed to $a \leftarrow b$, necessarily $c \notin \tau_a$, since all arrows of D'' within τ_a are oriented outward from a. Therefore $c \to a \in D^*$, hence $c \to a \to b$ occurs as an induced subgraph in D^*, contradicting the assumed nonoccurrence of (a) in D^*. Next, no immorality $a \to b \leftarrow c$ can occur in D', since D' and $(D')^* = D^*$ have the same immoralities and (b) is assumed not to occur in D^*. Thus D' and D'' have the same immoralities.

It follows that $D'' \in [D]$, whereby $D' \cup D'' \subseteq D^*$. But $a - b \in D' \cup D''$, hence $a - b \in D^*$, contradicting the assumption that $a \to b \in D^*$ and thereby establishing Fact 6. $\square$

FACT 7.   Every essential arrow of D is strongly protected in D^*.

PROOF.   Suppose that $a \to b \in D^*$ but satisfies neither (a), (b), (c) nor (d) in D^*. By Fact 6, $a \to b$ occurs in configuration (d′) for some $c \neq a, b$. Define the chain components τ_a and τ_b as above, and define $\sigma_a = \{c' \in \tau_a | c' \to b \in D^*\}$:
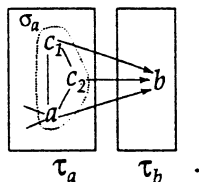

By (d′), a and $c \in \sigma_a$. We assert that G_{σ_a} is a complete subgraph of G_{τ_a} in $G \equiv D^*$.

Let c_1, c_2 be two distinct vertices in σ_a; it must be shown that $c_1 - c_2 \in D^*$. Suppose first that $c_2 = a$. Then an edge $a \cdots c_1$ must occur in D^*, or else $a \to b$ would satisfy (b) in D^*. Since $c_1 \in \sigma_a$, this edge must be $a - c_1$. Next, suppose that $a \neq c_1, c_2$. By the first case, $a - c_1 \in D^*$ and $a - c_2 \in D^*$. Therefore an edge $c_1 \cdots c_2$ must occur in D^*, else $a \to b$ would satisfy (d) in D^*. Since $c_1, c_2 \in \tau_a$, this edge must be $c_1 - c_2$.

Construct a directed graph D' from D^* as follows (see the following figure).

(i) For each chain component τ of $G \equiv D^*$ other than τ_a or τ_b, orient the edges of G_τ perfectly.

(ii) Assign a perfect orientation to the edges of G_{τ_b} starting at b.

(iii) Assign a perfect orientation to the edges of G_{τ_a} so that (α) any edge $a - c$ with $c \in \sigma_a$ becomes $a \leftarrow c$, and (β) any edge $a - d$ with $d \in \tau_a \setminus \sigma_a$ becomes $a \rightarrow d$.
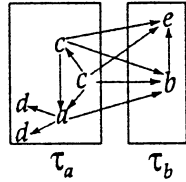
It must be shown that such an orientation exists for G_{τ_a}. Let $c_1, \ldots, c_q \equiv a$ be any ordering of the vertices in σ_a such that a occurs last. Starting at c_1, order the edges of G_{τ_a} by applying maximum cardinality search. The completeness of G_{σ_a} ensures that MCS can reproduce the initial sequence $c_1, \ldots, c_q$. The resulting perfect orientation of the edges within G_{τ_a} determined by this perfect ordering clearly satisfies (α) and (β).

By Proposition 4.2, D' is an ADG and $D' \in [D]$. Now construct a directed graph D'' which is identical to D' except that $a \rightarrow b \in D'$ is changed to $a \leftarrow b$ in D''. If D'' were to contain a directed cycle, then this cycle must include $a \leftarrow b$, hence must include a subgraph $a \leftarrow b \leftarrow c$ of D'' with $c \neq a$. By (ii), $c \notin \tau_b$, so $b \leftarrow c \in D^*$. Thus $a \rightarrow b \leftarrow c$ occurs in D^*, so, since $a \rightarrow b$ cannot satisfy (b) in D^*, there must be an edge $a \cdots c$ in D^*. This edge cannot be $a \rightarrow c$, otherwise $a \rightarrow b$ would satisfy (c) in D^*, hence must appear as either $a - c$ or $a \leftarrow c$ in D^*. If $a - c \in D^*$ then $c \in \tau_a$, hence $c \in \sigma_a$; by (α), this implies that $a \leftarrow c \in D'$ and therefore $a \leftarrow c \in D''$. If $a \leftarrow c \in D^*$, then again $a \leftarrow c$ must occur in both D' and D''. In either case, the assumed directed cycle in D'' cannot consist of the three vertices a, b and c alone, hence must have at least four distinct vertices. Furthermore, since $a \leftarrow c \in D''$, removing b from this directed cycle leaves a shorter directed cycle in D'' which must also occur in D' since D' and D'' coincide except for the edge $a \cdots b$, contradicting the acyclicity of D'. Thus D'' is acyclic.

Now we show that D'' has the same immoralities as D'. If a new immorality $c \rightarrow a \leftarrow b$ is created in D'' when $a \rightarrow b$ is changed to $a \leftarrow b$, then $c \rightarrow a \rightarrow b$ occurs in D'. Necessarily $c \notin \sigma_a$, for otherwise an edge $c \cdots b$ would occur in D^*. Also $c \notin \tau_a \setminus \sigma_a$, otherwise, by ($\beta$), $c \leftarrow a \in D'$. Thus

$c \notin \tau_a$, so $c \rightarrow a \in D^*$. Therefore $c \rightarrow a \rightarrow b$ occurs in D^* as an induced subgraph of D^*, contradicting the assumed nonoccurrence of (a). Next, no immorality $a \rightarrow b \leftarrow c$ can occur in D', since D' and $(D')^* = D^*$ have the same immoralities and (b) is assumed not to occur in D^*. Thus D' and D'' have the same immoralities, so $D' \sim D''$.

It follows that $D'' \in [D]$, hence $D' \cup D'' \subseteq D^*$. But $a - b \in D' \cup D''$, hence $a - b \in D^*$, contradicting the assumed occurrence of $a \rightarrow b$ in D^*. This establishes Fact 7 and thereby completes the proof of the "only if" assertion of Theorem 4.1. $\square$

("if".) Let $G \equiv (V, E)$ be a graph that satisfies properties (i)–(iv) of Theorem 4.1. It must be shown that $G = D^*$ for some ADG D. Let D be a digraph obtained from G by assigning arbitrary perfect orientations to the edges within each (chordal) chain component of G. Note that $D \subseteq G$. We shall show that D is an ADG and that $G = D^*$.

Suppose first that D contains a directed cycle. It cannot lie entirely within one chain component of G, hence at least one of its arrows is also an arrow in G. Therefore it determines a directed cycle in G, contradicting property (i). Thus D is an ADG.

To show that $G \subseteq D^*$, let $\mathbf{D}(G)$ be the collection of all ADG's D' constructed from G by assigning perfect orientations to the edges within each chain component of G (that is, all ADG's D' constructed in the same manner as D). Clearly $G \supseteq D'$, so $G \supseteq \cup(D'|D' \in \mathbf{D}(G))$. Furthermore, any line $a - b \in G$ lies in G_τ for some chain component τ of G. By property (ii), there exist two perfect orientations of the edges in G_τ, one with $a \rightarrow b$ and one with $a \leftarrow b$, so $G = \cup(D'|D' \in \mathbf{D}(G))$. By properties (ii) and (iii), no immorality in D' or D can involve an arrow that had been a line in G, that is, an arrow that lies within a chain component of G. Thus any immorality in D' or D is an immorality in G and conversely, so $D' \sim D$. Therefore $\cup(D'|D' \in \mathbf{D}(G)) \subseteq \cup(D'|D' \sim D) \equiv D^*$, so $G \subseteq D^*$. It remains to show that $G = D^*$.

For this, it suffices to show that

$$A := \{a \in V | \exists \, b \in V \ni a \rightarrow b \in G \text{ and } a - b \in D^*\} = \varnothing.$$

If not, let a be a minimal element of A with respect to the preordering $(V, \leq)$ determined by the chain graph G. Since $a \in A$,

$$B := \{b \in V | a \rightarrow b \in G \text{ and } a - b \in D^*\} \neq \varnothing.$$

Let b be a minimal element of B; then $a \rightarrow b \in G$ and $a - b \in D^*$. By property (iv), $a \rightarrow b$ occurs in at least one of the configurations (a), (b), (c) or (d) as an induced subgraph of G. If (a) were to occur in G, then, since a is minimal in A, $c - a \notin D^*$, hence $c \rightarrow a \in D^*$. But then $c \rightarrow a - b$ occurs as an induced subgraph of D^*, which is impossible by Fact 1. If (b) were to occur in G, then it must also occur in D, so $a \rightarrow b \in D^*$, which is also impossible. If (c) were to occur in G, then the minimality of b implies that $a - c \notin D^*$. Since $G \subseteq D^*$, one of the following two directed triangles must

occur in D^*, again impossible:
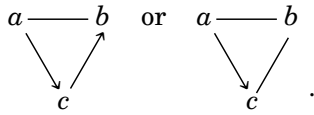
If (d) were to occur in G, then D^* would contain two directed triangles, also impossible. Thus $A = \varnothing$, hence $G = D^*$. The proof of Theorem 4.1 is complete. □

PROOF OF PROPOSITION 4.4. By hypothesis, the set

$$B := \{b' \in V | a \to b' \in G\}$$

is nonempty. Let c be any minimal element of B with respect to the preordering $(V, \leq)$ determined by the chain graph G. Since $a \to c$ is an initial arrow of G, it cannot occur in configuration (a) in G, nor can it occur in configuration (c), by the minimality of c. By Theorem 4.1(iv), therefore, $a \to c$ must occur in configuration (b) or (d) as an induced subgraph of G. □

PROOF OF PROPOSITION 4.5. It suffices to establish the result when H has *no* edges, that is, $H = (V, \varnothing)$. First assume that G contains at least one line (that is, undirected edge), so $G \equiv G_1$ has at least one chain component τ with at least two vertices. Since G_τ is chordal, it has at least one simplicial vertex a [cf. Blair and Peyton (1993), Lemma 2.2]; since G_τ is connected, $\mathrm{bd}_G(a) \neq \varnothing$. Choose any $b \in \mathrm{bd}_G(a)$, so that $a - b \in G$, then remove the line connecting a and b to produce a graph G_2. Since a was simplicial in G_τ, $(G_2)_\tau$ is also chordal. Because G_1 is an essential graph, it is now straightforward to verify that G_2 satisfies the conditions of Theorem 4.1, hence G_2 is also an essential graph. Continue this process (i) of single line removal until reaching an essential graph G_j with no lines. [A related argument appears in Lemma 5 of Frydenberg and Lauritzen (1989).]

If G_j has no arrows (that is, no directed edges), then $G_j = H$ and we are done. Otherwise, we can reach $H \equiv (V, \varnothing)$ by removing arrows from the ADG G_j according to (ii) or (iii) as follows. Let $B \; (\neq \varnothing)$ be the set of all *terminal* vertices of G_j, that is, the set of all $b \in V$ such that b is maximal in V with respect to the preordering $(V, \leq)$ determined by the chain graph G_j. Since G_j has at least one arrow, there must exist at least one $b \in B$ such that $A := \{a \in V | a \to b \in G_j\} \neq \varnothing$. Define $A_0 := \{a \in V | a$ is minimal in $A\} \; (\neq \varnothing)$. By Corollary 4.2, every arrow in G_j is protected in G_j. If A_0 contains only one vertex a, the minimality of a and the maximality of b imply that removal of the arrow $a \to b$ cannot leave any other arrow unprotected in the resulting ADG. If A_0 contains two or more vertices, their minimality implies that no two are adjacent in G_j. As in the first case, it follows that the arrows that these vertices form with b can be removed singly, until only two remain, and then either singly or as a pair,[6] in such a way that after each removal all remaining arrows are protected in the resulting ADG. Again by Corollary 4.2, each such ADG is an essential graph. This process can be continued until A_0

is exhausted, so that b becomes an isolated vertex in the resulting essential graph. Now consider the set of terminal vertices in this new essential graph and repeat the arrow removal process. Eventually all arrows can be removed and H will be reached. The proof is complete. $\square$

5. Construction of the essential graph D^*.   We now present a polynomial-time algorithm to construct the essential graph D^* from an ADG $D \equiv (V, E)$. This algorithm does not require an exhaustive search over the entire equivalence class $[D]$.

THE CONSTRUCTION ALGORITHM.   Define $G_0 := D$. For $i \geq 1$, convert every arrow $a \rightarrow b \in G_{i-1}$ that is not strongly protected in G_{i-1} into a line $a - b$, obtaining a graph G_i. Stop after k steps, where $k \geq 0$ is the smallest nonnegative integer such that $G_k = G_{k+1}$. Necessarily, $k \leq |E|$.

This algorithm produces a sequence $G_0, \ldots, G_k$ of graphs such that

(5.1) $$D \equiv G_0 \subset \cdots \subset G_k = G_{k+1}.$$

Since both arrows of an immorality are strongly protected, each G_i has the same immoralities as D and D^*. Let $n = |V|$. Because the determination of the set of arrows that are not strongly protected in G_{i-1} requires at most $O(n^4)$ operations and because $|E| = O(n^2)$, this algorithm requires at most $O(n^6)$ operations, although it can be implemented in a more efficient fashion.

THEOREM 5.1. The Construction Algorithm is valid: $G_k = D^*$.

PROOF.   If $k = 0$ (i.e., if every arrow of D is protected in D) then the result follows from Corollary 4.2. Thus we may assume that $k \geq 1$.

We begin by showing that $G_k \subseteq D^*$. First, by (5.1), $a \rightarrow b \in G_k \Rightarrow a \rightarrow b \in D \Rightarrow a \leftarrow b \notin D^* \Rightarrow$ either $a \rightarrow b \in D^*$ or $a - b \in D^*$. It remains to show that $a - b \in G_k \Rightarrow a - b \in D^*$. We shall accomplish this by proving that

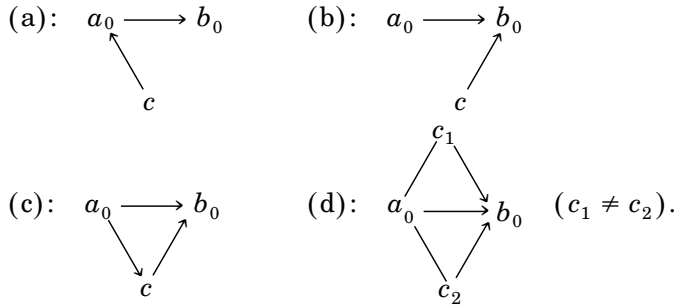$$B := \{b \in V | \exists\, a \in V \ni a - b \in G_k \text{ and } a \rightarrow b \in D^*\} = \varnothing.$$

Suppose that $B \neq \varnothing$. Let b_0 be a minimal element of B with respect to the preordering $(V, \leq)$ determined by the chain graph D^*. Therefore

$$A := \{a \in V | a - b_0 \in G_k \text{ and } a \rightarrow b_0 \in D^*\} \neq \varnothing.$$

For $a \in A$, let $i(a) \in \{1, \ldots, k\}$ be the unique integer such that $a \rightarrow b_0 \in G_{i(a)-1}$ but $a - b_0 \in G_{i(a)}$. Choose $a_0 \in A$ to minimize $i(a)$ over A. Thus, for no $a \in A$ is $a \rightarrow b_0$ converted to $a - b_0$ before $a_0 \rightarrow b_0$ is converted to $a_0 - b_0$ in the sequence $G_0, G_1, \ldots, G_k$. Therefore, a_0 and b_0 satisfy the following four properties.

(P1) $a_0 - b_0 \in G_k$ and $a_0 \rightarrow b_0 \in D^*$.

(P2) $a_0 \rightarrow b_0 \in G_{i(a_0)-1}$ but $a_0 - b_0 \in G_{i(a_0)}$, that is, $a_0 \rightarrow b_0$ is not strongly protected in $G_{i(a_0)-1}$.

(P3) If $a \rightarrow b_0 \in D^*$ but $a - b_0 \in G_k$, then $a \rightarrow b_0 \in G_{i(a_0)-1}$.

(P4) If $b < b_0$ in D^*, then for each $a \in V$ either $a - b \notin G_k$ or $a \rightarrow b \notin D^*$.

By Theorem 4.1(iv), $a_0 \to b_0 \in D^*$ must occur in at least one of the following four configurations as an induced subgraph of D^*:

(a): $a_0 \longrightarrow b_0$ with c

(b): $a_0 \longrightarrow b_0$ with c

(c): $a_0 \longrightarrow b_0$ with c

(d): $a_0 \longrightarrow b_0$ with c_1, c_2 $(c_1 \neq c_2)$.

However, each of these four possibilities leads to a contradiction.

(a) If $c \to a_0 \to b_0$ occurs as an induced subgraph of D^*, apply (P4) with $b = a_0$ and $a = c$ to conclude that $c - a_0 \notin G_k$. But $c \to a_0 \in D^* \Rightarrow c \to a_0 \in D \subset G_k$, hence $c \to a_0 \in G_k$. By (5.1), $c \to a_0 \in G_{i(a_0)-1}$, so by (P2), $c \to a_0 \to b_0$ occurs as an induced subgraph of $G_{i(a_0)-1}$. This implies that $a_0 \to b_0$ is strongly protected in $G_{i(a_0)-1}$, which contradicts (P2).

(b) The occurrence of the immorality $a_0 \to b_0 \leftarrow c$ in D^* implies its occurrence in $D \equiv G_0$. Thus both $a_0 \to b_0$ and $b_0 \leftarrow c$ are strongly protected in G_0, hence in $G_1, \ldots, G_{k-1}$. Therefore $a_0 \to b_0 \in G_k$, which contradicts (P1).

(c) Here, necessarily $c \to b_0 \in D$. By (5.1), either $c \to b_0 \in G_k$ or $c - b_0 \in G_k$. In the first case, $c \to b_0 \in G_{i(a_0)-1}$; in the second case, apply (P3) with $a = c$ to reach the same conclusion. Together with (P2), this implies that one of the following three configurations must occur in $G_{i(a_0)-1}$:

$a_0 \longrightarrow b_0$ with c or $a_0 \longrightarrow b_0$ with c or $a_0 \longrightarrow b_0$ with c.
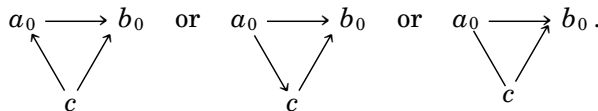
The first configuration is impossible, since $a_0 \to c \in D^* \Rightarrow a_0 \to c \in D \subseteq G_{i(a_0)-1}$. The second configuration is impossible, for otherwise $a_0 \to b_0$ is strongly protected in $G_{i(a_0)-1}$, contradicting (P2). If the third configuration holds, apply (P4) with $b = c$ and $a = a_0$ to deduce that $a_0 - c \notin G_k$, which contradicts the fact that $a_0 - c \in G_{i(a_0)-1}$ in this configuration.

(d) If this configuration occurs as an induced subgraph of D^*, then the immorality $c_1 \to b_0 \leftarrow c_2$ must occur in D and hence in $G_1, \ldots, G_k$. Together with (P2), this implies that

$a \longrightarrow b_0$ with c_1 (dashed), c_2 (dashed)
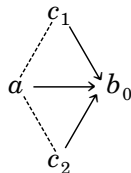
occurs in $G_{i(a_0)-1}$ but that $a_0 \to b_0$ is not strongly protected in $G_{i(a_0)-1}$. Therefore, one of the following three configurations must occur as an induced
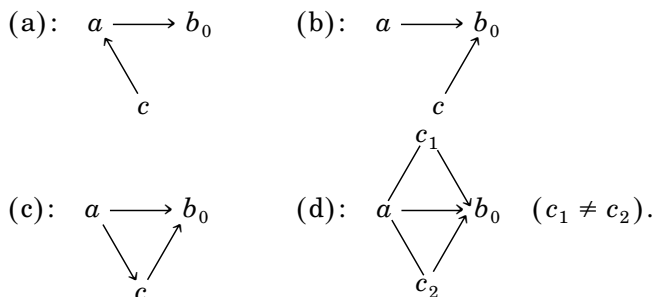
subgraph of $G_{i(a_0)-1}$:
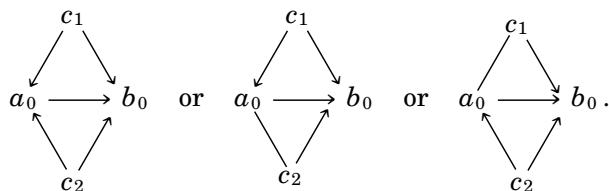
In the first case the immorality $c_1 \to a_0 \leftarrow c_2$ occurs in D and therefore in D^*, contradicting the assumed occurrence of $c_1 - a_0 - c_2$ in D^*. In the second case, either $c_1 \to a_0 \leftarrow c_2$ or $c_1 \to a_0 \to c_2$ must occur as an induced subgraph of D. As before, the immorality leads to a contradiction, so $c_1 \to a_0 \to c_2$ must occur as an induced subgraph of D and hence of $G_1, \ldots, G_{i(a_0)-2}$. Therefore $a_0 \to c_2$ is strongly protected in $G_{i(a_0)-2}$, contradicting the occurrence of $a_0 - c_2$ in $G_{i(a_0)-1}$ in this case. The third case is similar to the second.

Thus, each of the four possible configurations (a), (b), (c) or (d) leads to a contradiction, so $B = \varnothing$, hence $G_k \subseteq D^*$. It remains to show that $G_k = D^*$. For this purpose it suffices to show that

$$B' := \{b \in V | \exists\, a \in V \ni a \to b \in G_k \text{ and } a{-}b \in D^*\} = \varnothing.$$

Suppose that $B' \neq \varnothing$. Let b_0 be a minimal element of B' with respect to the partial ordering $(V, \leq)$ determined by the ADG D (*not* D^*). (Since $D \subseteq G_k$, this partial ordering is compatible with arrows in G_k, i.e., $a \to b \in G_k \Rightarrow a < b \in D$.) Thus there exists $a \in V$ such that $a \to b_0 \in G_k$ and $a - b_0 \in D^*$.

Since $a \to b_0 \in G_k$, $a \to b_0$ must be strongly protected in G_k, hence must occur in one of the following four configurations as an induced subgraph of G_k:

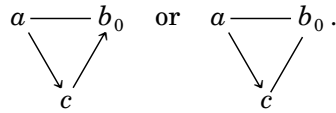(a) If $c \to a \to b_0$ occurs as an induced subgraph of G_k, then it also occurs as such in D. The minimality of b_0 then implies that $c \to a \in D^*$, hence $c \to a - b_0$ occurs as an induced subgraph of D^*, contradicting Fact 1.
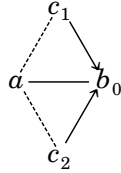
(b) The occurrence of the immorality $a \to b_0 \leftarrow c$ in G_k implies its occurrence in D and hence in D^*, contradicting the fact that $a - b_0 \in D^*$.

(c) If configuration (c) occurs in G_k, the minimality of b_0 implies that $a \to c \in D^*$. Since $G_k \subseteq D^*$ and $a - b_0 \in D^*$, one of the following two
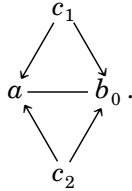
directed triangles must occur in D^*, contradicting Proposition 4.1.

(d) If configuration (d) occurs as an induced subgraph of G_k, then the configuration

must occur as an induced subgraph of D^*. This forces the occurrence of the configuration

in D^* (otherwise D^* would contain a directed triangle). Since $D \subseteq D^*$, the immorality $c_1 \to a \leftarrow c_2$ must occur in D and therefore in G_k, contradicting the assumed occurrence of $c_1 - a - c_2$ in G_k.

Each of the four possible configurations (a), (b), (c) or (d) has led to a contradiction, so $B' = \varnothing$. Therefore $G_k = D^*$ and the proof of Theorem 5.1 is complete. $\square$

REMARK 5.1. The construction algorithm becomes invalid if "strongly protected" is replaced by "protected." The following ADG D provides a counterexample:
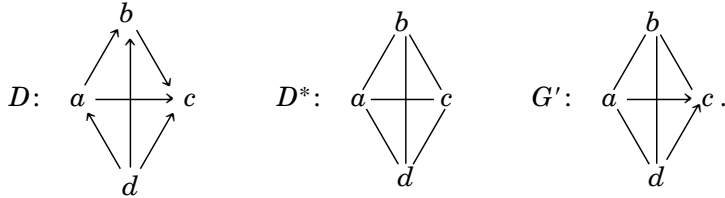
The valid algorithm produces D^* from D after $k = 2$ steps, while the invalid version stops at G' after $k = 2$ steps.

6. A brief catalog of essential graphs. By Theorem 4.1, essential graphs may be viewed as generalizations of chordal graphs. Darroch, Lauritzen and Speed (1980) give a brief catalog of chordal (decomposable) graphs; here we do the same for essential graphs with $n \leq 4$ vertices. In

Table 1 we list all such unlabelled essential graphs,[7] together with their corresponding global Markov properties,[8] then we simply enumerate the corresponding labelled essential graphs D^* and the corresponding labelled ADG's D' in the equivalence class $[D]$.

In applications, of course, different labelled essential graphs represent different statistical models, whereas different labelled ADG's D' corresponding to the same labelled essential graph represent the same statistical model.

Thus, for example, the second essential graph listed in Table 1 corresponds to one labelled D^*: $1—2$ and to two labelled D': $1 \to 2$ and $1 \leftarrow 2$. The fifth essential graph in Table 1 corresponds to three labelled D^*: $1 — 2 — 3$, $1 — 3 — 2$, $2 — 3 — 1$, each representing a different statistical model, and to nine labelled D': $1 \to 2 \to 3$, $1 \leftarrow 2 \leftarrow 3$, $1 \leftarrow 2 \to 3$, $1 \to 3 \to 2$, $1 \leftarrow 3 \leftarrow 2$, $1 \leftarrow 3 \to 2$, $2 \to 1 \to 3$, $2 \leftarrow 1 \leftarrow 3$, $2 \leftarrow 1 \to 3$, representing the same three models.

For $n = 5$ vertices, we have utilized a computer search to find that the total numbers of labelled essential graphs and labelled ADG's are 8,782 and 29,281, respectively. Robinson (1977) gives a recursive formula for the number of labelled ADG's, from which it follows that there are 3,781,503 labelled ADG's for $n = 6$ vertices, but at present no formula is available for the number of labelled essential graphs. It would be of interest to determine the asymptotic behavior of the ratio of these numbers as n approaches infinity.

7. Model selection and model averaging for acyclic digraphs. By focusing on Markov-equivalence classes of ADG's rather than on the individual ADG's themselves, data analysts and expert system builders can overcome several difficulties associated with ADG models. Three such difficulties were listed in Section 1; here we examine these in more detail and indicate how the introduction of essential graphs can help to overcome them.

1. *Computational inefficiencies*. Heckerman, Geiger and Chickering (1994) and Chickering (1995) argue that statistical inference for ADG models should be "score equivalent": in the absence of a priori causal knowledge, Markov-equivalent ADG's should have identical posterior model probabilities (Bayesian) or identical penalized likelihoods (non-Bayesian). Under this criterion, therefore, model selection and model averaging algorithms need visit each Markov-equivalence class only once. However, standard algorithms [e.g., Madigan and Raftery (1994), Madigan and York (1995) and Heckerman, Geiger and Chickering (1994)] fail to treat each Markov-equivalence class of ADG's as a single statistical model and search in the space of ADG's, introducing considerable computational inefficiency. For example, an exhaustive search among all ADG's on four variables would require the calculation of posterior probabilities for all 543 such ADG's, whereas a search over the space of essential graphs (in $1-1$ correspondence with the equivalence classes) would require only 185 such calculations. For five variables, the numbers become 8,782 and 29,281, respectively.
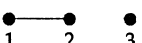
TABLE 1
Essential graphs with n = 2, 3 and 4 vertices
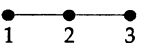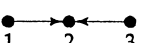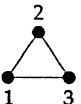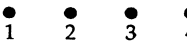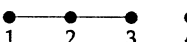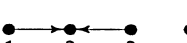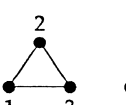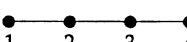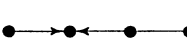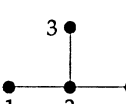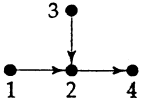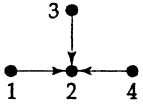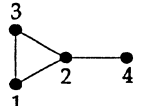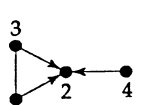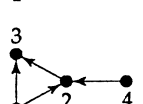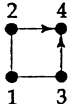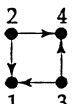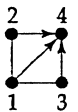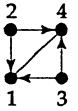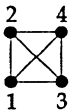
	Essential graph	Markov property	Number of labelled essential graphs	Number of labelled ADGs
$n = 2$		$1 \perp 2$	1	1
		(None)	1	2
Totals:			2	3
$n = 3$		$1 \perp 2 \perp 3$	1	1
		$(1,2) \perp 3$	3	6
		$1 \perp 3|2$	3	9
		$1 \perp 3$	3	3
		(None)	1	6
Totals:			11	25
$n = 4$		$1 \perp 2 \perp 3 \perp 4$	1	1
		$(1,2) \perp 3 \perp 4$	6	12
		$(1,2) \perp (3,4)$	3	12
		$1 \perp 3|2$ $(1,2,3) \perp 4$	12	36
		$1 \perp 3$ $(1,2,3) \perp 4$	12	12
		$(1,2,3) \perp 4$	4	24
		$1 \perp 3|2$ $(1,2) \perp 4|3$	12	48
		$1 \perp (3,4)$ $2 \perp 4|1,3$	24	48
		$1 \perp (3,4)|2$ $(1,3) \perp 4|2$	4	16

TABLE 1 (*Continued*)

Essential graph	Markov property	Number of labelled essential graphs	Number of labelled ADGs
	$1 \perp 3$ $(1,3) \perp 4 \mid 2$	12	12
	$1 \perp 3 \perp 4$	4	4
	$(1,3) \perp 4 \mid 2$	12	96
	$(1,3) \perp 4$	12	24
	$1 \perp 4$ $3 \perp 4 \mid 1,2$	24	24
	$2 \perp 3 \mid 1$ $1 \perp 4 \mid 2,3$	12	36
	$2 \perp 3$ $1 \perp 4 \mid 2,3$	6	6
	$2 \perp 3 \mid 1,4$	6	60

2. *Constraints on prior distributions.* For a Bayesian analysis over the space of all individual ADG models with a fixed vertex set V, score equivalence imposes severe restrictions on the prior distributions that may be used to represent prior knowledge about the parameters of these models. For any individual ADG D, the joint pdf (if it exists) of a global D-Markovian distribution admits the factorization [cf. Lauritzen, Dawid, Larsen and Leimer (1990),

TABLE 1 (*Continued*)

Essential graph	Markov property	Number of labelled essential graphs	Number of labelled ADGs
2 4 / 1 3	$2 \perp 3 \mid 1$	12	36
2 4 / 1 3	$2 \perp 3$	6	12
2 4 / 1 3	(None)	1	24
Totals:		185	543

Theorem 1]

$$(7.1) \qquad f(V) = \Pi\big(f(a|\mathrm{pa}_D(a))|a \in V\big).$$

For categorical data,[9] where each conditional pdf $f(a|\mathrm{pa}_D(a))$ is multinomial, Spiegelhalter and Lauritzen (1990) proposed the now widely accepted conjugate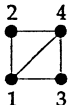 family of Dirichlet prior distributions for the parameters occurring in these conditional multinomial distributions. However, Heckerman, Geiger and Chickering (1994) show that score equivalence *requires that the sum of the parameters of all the Dirichlet distributions associated with each $a \in V$* [*i.e., the Dirichlet distributions for each of the levels of* $\mathrm{pa}_D(a)$] *be identical for all $a \in V$*. Since these sums behave as "equivalent sample sizes" in subsequent Bayesian updating, this constraint severely restricts an expert with more prior knowledge about some variables than others—he must use a single equivalent sample size for *each* of the Dirichlet distributions occurring in the conjugate prior and is therefore unable fully to utilize his prior knowledge.

This difficulty can be overcome by constructing prior distributions over Markov-equivalence classes of ADG models, rather than over the individual ADG models themselves. To accomplish this, represent each equivalence class [D] by its essential graph D^*, then select appropriate prior distributions for the parameters of the chain graph model determined by D^*. More precisely, by Theorem 4.1(ii) of Frydenberg (1990), the joint pdf (if it exists and is positive) of a global D^*-Markovian distribution P admits the factorization

$$(7.2) \qquad f(V) = \Pi\big(f(\tau|\mathrm{bd}_{D^*}(\tau))|\tau \in \mathbf{T}(D^*)\big),$$

where, further, each *marginal* pdf $f(\mathrm{cl}_{D^*}(\tau))$ is global $[(D^*)_{\mathrm{cl}_{D^*}(\tau)}]^m$-Markovian. This in turn implies that each conditional pdf $f(\tau|\mathrm{bd}_{D^*}(\tau))$ is global $(D^*)_\tau$-Markovian.[10] However, by our Theorem 4.1(ii) each $(D^*)_\tau$ is chordal (i.e., decomposable), so therefore we can utilize *hyper-Dirichlet* distributions as prior distributions for the parameters occurring in these conditional pdfs.[11] Since score equivalence is no longer an issue, no constraints are required on the parameters of these hyper-Dirichlet priors.

Furthermore, although the Dirichlet and hyper-Dirichlet families provide considerable flexibility for modeling prior knowledge in the Bayesian analysis of categorical data, more general priors, such as mixtures of Dirichlet distributions, sometimes may be needed to adequately reflect prior knowledge [Bernardo and Smith (1994), page 279]. When working in the space of individual ADG models, however, Geiger and Heckerman (1995) show that the Dirichlet family is the only family of prior distributions that can be used to achieve score equivalence. Working in the space of Markov-equivalence classes, conveniently represented by essential graphs, eliminates the issue of score equivalence and therefore allows the adoption of arbitrary prior distributions on the associated parameters, at least in principle.

3. *Improper model weights.* Madigan and Raftery (1994) and others have argued that basing inference on a single model ignores model uncertainty and leads to poorly calibrated predictions. Bayesian model averaging (BMA) provides a remedy: current BMA procedures average inferences or predictions over all models in the class under consideration, or at least over a subset of the models that receive substantial posterior weight [see Madigan and York (1995) for a review.] When applied naively to ADG models, however, BMA assigns a weight to each Markov-equivalence class that is proportional to its size. Instead, averaging directly over equivalence classes overcomes this problem.

A stochastic search scheme over the space of ADG's based on the Metropolis–Hastings algorithm has been proposed for Bayesian model averaging by Madigan and York (1995).[12] As suggested by the final paragraph of Section 6, the number of essential graphs on n vertices, although substantially smaller than the number of ADG's, will still be too large in most applications to allow an exhaustive analysis,[13] hence search procedures over the space of essential graphs also will be required.

Madigan, Andersson, Perlman and Volinsky (1996) describe several stochastic search procedures for model selection and model averaging, again based on the Metropolis–Hastings algorithm, that act directly on essential graphs rather than ADG's. Such procedures move through the space of essential graphs according to a Markov chain whose transition probabilities are chosen to achieve a desired stationary distribution. Convergence to the stationary distribution requires that the Markov chain be irreducible and aperiodic. By Proposition 4.5, irreducibility will hold whenever the chain has positive probability of moving to any essential graph that differs by *at most two edges* from the current essential graph. However, it follows from the proof

of Proposition 4.5 that in fact irreducibility will hold whenever the chain has positive probability of moving from the current essential graph G to each essential graph G' that has one of the following three properties:

(a) G' differs from G by exactly one edge;

(b) G' is obtained from G by deleting both arrows in an immorality $a \rightarrow b \leftarrow c$, where b is a terminal vertex of G and where a and c are the only parents of b in G;

(c) G' is obtained from G by adding two arrows to form an immorality $a \rightarrow b \leftarrow c$, where b is an isolated vertex of G and where a and c are not adjacent in G.

Aperiodicity can be guaranteed, for example, by ensuring that the chain has positive probability of remaining in its current state.

Nonstochastic model selection and model averaging schemes based on essential graphs also can be developed, analogous to those proposed by Heckerman, Geiger and Chickering (1994) and Madigan and Raftery (1994) for ADG's.

APPENDIX A

Graphs. Our terminology and notation closely follow those of Lauritzen, Dawid, Larsen and Leimer (1990) and Frydenberg (1990), with one exception noted below. A *graph* G is a pair (V, E), where V is a finite set of *vertices* and E, the set of *edges*, is a subset of $E^*(V) \equiv (V \times V) \setminus \{(a, a) | a \in V\}$, that is, a set of ordered pairs of distinct vertices; thus our graphs include no loops or multiple edges. An edge $(a, b) \in E$ whose opposite $(b, a) \in E$ is called an *undirected* edge and appears as a line $a — b$ in our figures, whereas an edge $(a, b) \in E$ whose opposite $(b, a) \notin E$ is called a *directed* edge and appears as an arrow: $a \rightarrow b$.[14] If G contains only undirected edges, it is an *undirected graph* (UDG); if G contains only directed edges it is a *directed graph* (digraph).

It will be convenient to write "$a \rightarrow b \in G$" to indicate that $(a, b) \in E$ but $(b, a) \notin E$; in this case we say that *the arrow* $a \rightarrow b$ *occurs in* G. Similarly, we write "$a — b \in G$" to indicate that $(a, b) \in E$ and $(b, a) \in E$; in this case we say that *the line* $a — b$ *occurs in* G. We write "$a \cdots b \in G$" to indicate that there is an edge of some type between a and b in G.

For each vertex $a \in V$, define $\mathrm{pa}_G(a) := \{b \in V | b \rightarrow a \in G\}$, the set of *parents* of a in G. For any subset $A \subseteq V$, the *boundary* of A in G is the set $\mathrm{bd}_G(A) := \{b \in V \setminus A | (b, a) \in E \text{ for some } a \in A\}$; the *closure* of A in G is the set $\mathrm{cl}_G(A) := \mathrm{bd}_G(A) \cup A$.

A subset $A \subseteq V$ *induces* the subgraph $G_A := (A, E_A)$, where $E_A := E \cap (A \times A)$.

The *skeleton* G^u of a graph $G \equiv (V, E)$ is its underlying undirected graph, that is, $G^u := (V, E^u)$, where $E^u := \{(a, b) | (a, b) \in E \text{ or } (b, a) \in E\}$. Two vertices a, b are called *adjacent* in G if $(a, b) \in E^u$, or, equivalently, if $a \cdots b \in G$. A vertex a is *isolated* if it is not adjacent to any b.

Let a, b and c be three distinct vertices of $G \equiv (V, E)$. The triple (a, b, c) is called an *immorality* of G if the induced subgraph $G_{\{a, b, c\}}$ is $a \to b \leftarrow c$; that is, if the "parents" a and c of b are "unmarried" (i.e., nonadjacent).

A graph $G_2 \equiv (V_2, E_2)$ is said to be *larger* than a graph $G_1 \equiv (V_1, E_1)$, denoted by $G_1 \subseteq G_2$, if $V_1 \subseteq V_2$ and $E_1 \subseteq E_2$. Thus, if $(G_1)^u = (G_2)^u$, then $G_1 \subseteq G_2$ iff G_1 and G_2 differ only in that some directed edges (arrows) in G_1 may be converted into undirected edges (lines) in G_2. We write $G_1 \subset G_2$ if $G_1 \subseteq G_2$ but $G_1 \neq G_2$.

The *union* of a finite collection of subgraphs $\{G_i \equiv (V_i, E_i) | i = 1, \ldots, n\}$ of $G \equiv (V, E)$ is the subgraph $\cup G_i := (\cup V_i, \cup E_i)$. Clearly, $\cup G_i$ is the smallest subgraph larger than each G_i, $i = 1, \ldots, n$.

Let a, b be distinct vertices in $G \equiv (V, E)$. A *path* π of length $n \geq 1$ from a to b in G is a sequence $\pi \equiv \{a_0, a_1, \ldots, a_n\} \subseteq V$ of distinct vertices such that such that $a_0 = a$, $a_n = b$, and either $a_{i-1} \to a_i \in G$ or $a_{i-1} - a_i \in G$ for every $i = 1, \ldots, n$. If $a_{i-1} \to a_i \in G$ for at least one i, the path is *directed*; if this is not the case, the path is *undirected*. A (*directed*) *cycle* is a (directed) path with the modification that $a_0 = a_n$. An arrow $a \to b \in G$ is said to *block a directed cycle in* G if there is a directed path from a to b in G other than $a \to b$ itself.

A UDG $G \equiv (V, E)$ is *complete* if all pairs of vertices are adjacent. Trivially, the empty graph is complete. A subset $A \subseteq V$ is *complete* if its induced subgraph G_A is complete. A complete subset that is maximal with respect to inclusion is called a *clique*. A vertex a is *simplicial* if its boundary $\mathrm{bd}_G(a)$ is complete. A subset $A \subseteq V$ is *connected* in G if for every distinct pair $a, b \in A$, there is a path from a to b in G_A. For pairwise disjoint subsets A ($\neq \varnothing$), B ($\neq \varnothing$) and S of V, A and B are *separated* by S in G if all paths from vertices in A to vertices in B intersect S.

The UDG $G \equiv (V, E)$ is *chordal* if every cycle of length $n \geq 4$ possesses a *chord*, that is, two nonconsecutive adjacent vertices. A total ordering of V is a *perfect ordering* of G if, when each edge of G is oriented in accordance with this ordering, the resulting ADG D is *perfect*, that is, is *acyclic* and *moral* (without immoralities); D is called a *perfect directed version of G*. It is well known that a UDG admits a perfect directed version if and only if it is chordal [cf. Blair and Peyton (1993)]. Furthermore, such a perfect orientation of a chordal UDG G is not unique: in fact, by using maximum cardinality search (MCS) [cf. Blair and Peyton (1993)], the perfect ordering can be started at any vertex in G. Thus, for any distinct vertices $a, b \in V$, a chordal UDG G admits two perfect directed versions, say D_1 and D_2, such that $a \to b \in D_1$ and $a \leftarrow b \in D_2$.

A graph $G \equiv (V, E)$ is called a *chain graph* (equivalently, an *adicyclic graph*) if it contains no directed cycles. Every induced subgraph G_A of G is also a chain graph. Any UDG is trivially a chain graph. A chain graph that is also a digraph is called an *acyclic digraph* (ADG).

An ADG D is *transitive* if $a \to c \in D$ whenever $a \to b \in D$ and $b \to c \in D$.

For the remainder of Appendix A, let $G \equiv (V, E)$ be a chain graph. Then G determines a preordering $(V, \leq)$ as follows: $a \leq b$ iff $a = b$ or there exists a
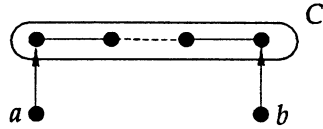
FIG. 4.   *A simple chain graph. Here (a, C, b) is a minimal complex.*

path from a to b in G. A subset $A \subseteq V$ is an *anterior* set if $b \leq a \in A \Rightarrow b \in A$. For a subset $A \subseteq V$, $\mathrm{An}(A)$ denotes the smallest anterior set containing A: $\mathrm{An}(A) = \{b \in V | b \leq a \text{ for some } a \in A\}$.

If both $a \leq b$ and $b \leq a$, then we write $a \approx b$, which occurs iff $a = b$ or there is an undirected path from a to b in G. Frydenberg (1990) notes that $\approx$ is an equivalence relation on V; we denote the set of equivalence classes in V by $\mathbf{T}(G)$. Equivalently, $\mathbf{T}(G)$ is the set of connected components of the undirected graph obtained from G by removing all directed edges. Each $\tau \in \mathbf{T}(G)$ is called a *chain component* of G. A connected UDG has only one chain component, while for an ADG, every chain component consists of a single vertex.

We write $a < b$ if there exists a directed path from a to b. The *future* of a vertex $a \in V$ is the set $\phi(a) := \{b \in V | a < b\}$.

A triple (a, C, b) is called a *complex* in G if C is a connected subset of a chain component $\tau \in \mathbf{T}(G)$ and a and b are two nonadjacent vertices in $\mathrm{bd}_G(\tau) \cap \mathrm{bd}_G(C)$. A complex (a, C, b) is called a *minimal complex* in G if no proper subset $C' \subset C$ forms a complex (a, C', b) in G. Frydenberg (1990) notes that (a, C, b) is a minimal complex in G iff $G_{C \cup \{a, b\}}$ looks like the chain graph of Figure 4. An immorality is the special case of a minimal complex where $|C| = 1$.

The *moral graph* determined by G is the undirected graph $G^m \equiv (V, E^m)$, where $E^m := E^u \cup [\bigcup (E^*(\mathrm{bd}_G(\tau)) | \tau \in \mathbf{T}(G))]$. That is, G^m is G^u augmented by all undirected edges needed to make $\mathrm{bd}_G(\tau)$ complete in G^m for every chain component $\tau \in \mathbf{T}(G)$. Equivalently, G^m is obtained from G^u by adding a line $a \text{ --- } b$ whenever (a, C, b) is a minimal complex in G.

APPENDIX B

Graphical Markov models and Markov equivalence.   We consider multivariate probability distributions P on a product probability space $\mathbf{X} \equiv \times (\mathbf{X}_a | a \in V)$, where V is a finite index set and each $\mathbf{X}_a$ is sufficiently regular to ensure the existence of regular conditional probabilities. Such distributions are conveniently represented by a random variate $X := (X_a | a \in V) \in \mathbf{X}$. For any subset $A \subseteq V$, we define $X_A := (X_a | a \in A)$. Often we abbreviate X_a and X_A by a and A, respectively, and define $X_\emptyset \equiv$ constant.

For three pairwise disjoint subsets A, B and C of V, we write $A \perp B | C [P]$ to indicate that X_A and X_B are conditionally independent given X_C under P.

A graphical Markov model is defined by a collection of conditional independencies among the component random variates $(X_a | a \in V)$, which collection is represented by a chain graph $G \equiv (V, E)$ with vertex set V.

DEFINITION B.1. A probability measure P on $\mathbf{X}$ is said to be *local G-Markovian* if $a \perp [V \setminus \phi(a)] \setminus \mathrm{cl}_G(a) | \mathrm{bd}_G(a)[P]$ $\forall$ $a \in A$.

DEFINITION B.2. A probability measure P on $\mathbf{X}$ is said to be *global G-Markovian* if $A \perp B | S[P]$ whenever S separates A and B in $(G_{\mathrm{An}(A \cup B \cup S)})^m$.

Frydenberg [(1990), page 339] notes that global G-Markovian implies local G-Markovian. The converse is not true in general; see, for example, Andersson, Madigan and Perlman [(1996), Remark A.1], but Lauritzen, Dawid, Larsen and Leimer [(1990), Proposition 4] show that the converse is valid if G is an ADG.

We define the *graphical Markov model* on $\mathbf{X}$ determined by a chain graph G to be the set of all global G-Markovian probability measures on $\mathbf{X}$. (In applications, an additional parametric assumption, such as multivariate normality, is often imposed.)

DEFINITION B.3. Two chain graphs G_1 and G_2 are *Markov equivalent* on a product space $\mathbf{X}$ indexed by V if the classes of global G_1-Markovian and global G_2-Markovian probability measures on $\mathbf{X}$ coincide. If G_1 and G_2 are Markov equivalent on every such product space $\mathbf{X}$, G_1 and G_2 are called *Markov equivalent*.

The following basic result concerning Markov equivalence of chain graphs was first proved by Frydenberg [(1990), Theorem 5.6] for a restricted class of probability measures and by Andersson, Madigan and Perlman [(1996), Theorem 3.1] for the general case. We shall say that two chain graphs are *graphically equivalent* if they have the same skeleton and the same minimal complexes.

THEOREM B.1. *Suppose that for each $a \in V$, the component space X_a of $\mathbf{X}$ contains at least two points. Then two chain graphs G_1 and G_2 are Markov equivalent on $\mathbf{X}$ if and only if they have the same skeleton and the same minimal complexes. Thus, G_1 and G_2 are Markov equivalent if and only if they are graphically equivalent.*

Since the only possible minimal complexes in an ADG are immoralities, Theorem 2.1, the key equivalence theorem for ADG's, follows from Theorem B.1 as a special case. Because the proof of Theorem B.1 is quite complex, however, we present here a direct proof of Theorem 2.1, different from that of Verma and Pearl (1992) in that their notion of "d-separation" is not used.

We require the notion of a *well-numbering* (i.e., *topological sort*) of an ADG $D \equiv (V, E)$, namely, a 1–1 mapping $\nu: V \to \{1, \ldots, n\}$, $n \equiv |V|$, such that $c < d \Rightarrow \nu(c) < \nu(d)$. A straightforward inductive argument shows that every ADG admits at least one well-numbering. Propositions 4 and 5 of Lauritzen, Dawid, Larsen and Leimer (1990) together imply that a probability measure P on $\mathbf{X}$ is global D-Markovian if and only if, for some (and, therefore, for every) well-numbering ν of D,

(B.1) $c \perp \{d \in V | \nu(d) < \nu(c)\} \setminus \text{pa}_D(c) | \text{pa}_D(c) [P]$ $\forall\, c \in V$.

PROOF OF THEOREM 2.1. ("if"). Suppose that D and D' are two ADG's with the same skeleton and same immoralities. In order to show that D and D' are Markov equivalent, by Lemma 3.2 we may assume in addition that D and D' differ in exactly one edge, say $a \to b \in D$ but $b \to a \in D'$. By Lemma 3.1(ii), $a \to b$ is reversible, and therefore unprotected, in D, that is, $\text{pa}_D(a) = \text{pa}_D(b) \setminus \{a\}$. It follows that for any well-numbering ν of D, $\nu(b) = \nu(a) + 1$. This can then be applied to show that $\nu': V \to \{1, \ldots, n\}$ is a well-numbering of D', where ν' is defined as follows: $\nu'(a) = \nu(b)$, $\nu'(b) = \nu(a)$, $\nu'(c) = \nu(c)$ if $c \neq a, b$. Therefore, a probability measure P on $\mathbf{X}$ is global D'-Markovian if and only if

(B.2) $c \perp \{d \in V | \nu'(d) < \nu'(c)\} \setminus \text{pa}_{D'}(c) | \text{pa}_{D'}(c) [P]$ $\forall\, c \in V$.

Since D and D' differ only in the edge $a \cdots b$, $\text{pa}_D(c) = \text{pa}_{D'}(c)$ if $c \neq a, b$, so the conditions in (B.1) and (B.2) coincide when $c \neq a, b$. The remaining conditions in (B.1) and (B.2) are

(B.3) $a \perp \{d \in V | \nu(d) < \nu(a)\} \setminus \text{pa}_D(a) | \text{pa}_D(a) [P]$,

(B.4) $b \perp \{d \in V | \nu(d) < \nu(b)\} \setminus \text{pa}_D(b) | \text{pa}_D(b) [P]$,

and

(B.5) $a \perp \{d \in V | \nu'(d) < \nu'(a)\} \setminus \text{pa}_{D'}(a) | \text{pa}_{D'}(a) [P]$,

(B.6) $b \perp \{d \in V | \nu'(d) < \nu'(b)\} \setminus \text{pa}_{D'}(b) | \text{pa}_{D'}(b) [P]$,

respectively. Since $\text{pa}_{D'}(a) = \text{pa}_D(a) \cup \{b\}$ and $\text{pa}_{D'}(b) = \text{pa}_D(b) \setminus \{a\}$, (B.5) and (B.6) can be rewritten as

(B.7) $a \perp \{d \in V | \nu(d) < \nu(b)\} \setminus \text{pa}_D(a) | [\text{pa}_D(a) \cup \{b\}] [P]$,

(B.8) $b \perp \{d \in V | \nu(d) < \nu(a)\} \setminus \text{pa}_D(b) | [\text{pa}_D(b) \setminus \{a\}] [P]$.

Finally, use the relation $\text{pa}_D(b) = \text{pa}_D(a) \cup \{a\}$ and the following well-known property of conditional distributions to conclude that (B.3) and (B.4) are jointly equivalent to (B.7) and (B.8): for any four random variates X, Y, Z and W,

$X \perp (Y, Z) | W \Leftrightarrow X \perp Y | W$ and $X \perp Z | W, Y \Leftrightarrow X \perp Z | W, Y \Leftrightarrow X \perp Z | W$ and $X \perp Y | W, Z$.

Therefore (B.1) and (B.2) are equivalent, hence D and D' are Markov equivalent.

("only if".) The proof given by Frydenberg [(1990), pages 347–348] for chain graphs applies without change to the special case of ADG's.

ENDNOTES

1. Chain graphs may have both directed and undirected edges but may contain no (partially) directed cycles; they include both ADG's and UDG's ADG's as special cases.
2. The essential graph associated with an (equivalence class of) ADG's was first introduced by Verma and Pearl (1990) as the *completed pattern* associated with the ADG.
3. Chickering (1995) and Meek (1995) also have obtained polynomial-time algorithms for constructing D^* from D.
4. Chickering [(1995), Section 4] notes that, under certain additional assumptions, the essential arrows (= *compelled edges*) of an ADG may indicate causal influences.
5. This statement is valid because we have defined Markov equivalence of chain graphs in terms of the *global* Markov property—see Definition B.3 in Appendix B. If we were to replace the global Markov property by the *local* Markov property, then this statement is not valid in general. The local and global Markov properties of the chain graph D^* need not be equivalent, whereas those of the ADG D must be equivalent—see Appendix B and also the example in Remark 3.4 of Andersson, Madigan and Perlman (1996).
6. This is not an arbitrary choice: removal of only one arrow may leave the other unprotected.
7. The vertices of the graphs in Table 1 are labelled only to allow us to describe the Markov properties.
8. In fact, we only present a parsimonious list of independencies that are *equivalent* to the global Markov properties of the essential graph. Recall that the local and global Markov properties of the essential graph itself may not be equivalent, whereas the local and global Markov properties of any ADG in its equivalence class are equivalent to each other and to the global Markov properties of the essential graph. We use the local Markov properties of such ADG's, together with standard properties of conditional independence, to obtain our parsimonious lists.
9. In this case, the joint and conditional pdf's in (7.1) and (7.2) denote the pdf's for the classification of a *single individual*.
10. Since $f(\mathrm{cl}_{D*}(\tau))$ is global $[(D^*)_{\mathrm{cl}_{D*}(\tau)}]^m$-Markovian, it admits a Gibbs factorization over the cliques $\chi_1, \ldots, \chi_k$ of $[(D^*)_{\mathrm{cl}_{D*}(\tau)}]^m$ [cf. Frydenberg (1990), page 344]. Because each intersection $\chi_i \cap \tau$ is complete in $(D^*)_\tau$, it is contained in at least one clique of $(D^*)_\tau$. Thus, each conditional pdf $f(\tau | \mathrm{bd}_{D*}(\tau))$ admits a Gibbs factorization over the cliques of $(D^*)_\tau$, hence each conditional pdf is global $(D^*)_\tau$-Markovian.
11. Dawid and Lauritzen (1993) introduced hyper-Dirichlet distributions as natural conjugate priors in decomposable models for categorical data. As in the case of Dirichlet priors for multinomial data, hyper-Dirichlet priors allow explicit expressions for posterior model probabilities.
12. George and McCulloch (1994) discuss similar stochastic search procedures for Bayesian model selection in regression analysis.
13. For example, Spiegelhalter, Dawid, Lauritzen and Cowell [(1993), Section 3] and Heckerman, Horvitz and Nathwani (1992) model dependencies in biomedical data by means of ADG's with $n = 20$ and $n = 108$ vertices, respectively.
14. Our notation differs from Frydenberg's in that he uses the notation $a \Rightarrow b$ rather than $a \to b$ in his text, although not in his figures.

REFERENCES

ANDERSSON, S. A., MADIGAN, D. and PERLMAN, M. D. (1996). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scand. J. Statist.* To appear.

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1995). On the relation between conditional independence models determined by finite distributive lattices and by directed acyclic graphs. *J. Statist. Plann. Inference* **48** 25–46.

ANDERSSON, S. A., MADIGAN, D., PERLMAN, M. D. and TRIGGS, C. M. (1996). A graphical characterization of lattice conditional independence models. *Ann. Math. and Artificial Intelligence*. To appear.

ANDERSSON, S. A. and PERLMAN, M. D. (1996). Normal linear regression models with recursive graphical Markov structure. Unpublished manuscript.

ASMUSSEN, S. and EDWARDS, D. (1983). Collapsibility and response variables in contingency tables. *Biometrika* **70** 567–578.

BERNARDO, J. M. and SMITH, A. F. M. (1994). *Bayesian Theory*. Wiley, Chichester.

BLAIR, J. R. S. and PEYTON, B. (1993). An introduction to chordal graphs and clique trees. In *Graph Theory and Sparse Matrix Computation* (A. George, J. R. Gilbert and J. W. H. Liu, eds.) 1–29. Springer, New York.

BUNTINE, W. (1994). Operations for learning with graphical models. *Journal of Artificial Intelligence Research* **2** 159–225.

CHICKERING, D. M. (1995). A transformational characterization of equivalent Bayesian network structures. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence* (P. Besnard and S. Hanks, eds.) 87–98. Morgan Kaufmann, San Mateo, CA.

COOPER, G. F. and HERSKOVITS, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* **9** 309–347.

COX, D. R. and WERMUTH, N. (1993). Linear dependencies represented by chain graphs (with discussion). *Statist. Sci.* **8** 204–218, 247–277.

COX, D. R. and WERMUTH, N. (1996). *Multivariate Dependencies: Models, Analysis, and Interpretation*. Chapman and Hall, London.

DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317.

DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539.

FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Statist.* **17** 333–353.

FRYDENBERG, M. and LAURITZEN, S. L. (1989). Decomposition of maximum likelihood in mixed graphical interaction models. *Biometrika* **76** 539–555.

GEIGER, D. and HECKERMAN, D. (1995). A characterization of the Dirichlet distribution through local and global independence. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence* (P. Besnard and S. Hanks, eds.) 196–207. Morgan Kaufmann, San Mateo, CA.

GEORGE, E. and MCCULLOGH, R. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.

GOODMAN, L. (1973). The analysis of multidimensional contingency tables when some variables are posterior to others: a modified path analysis approach. *Biometrika* **60** 179–192.

HECKERMAN, D., GEIGER, D. and CHICKERING, D. M. (1994). Learning Bayesian networks: the combination of knowledge and statistical data. In *Uncertainty in Artificial Intelligence, Proceedings of the Tenth Conference* (B. Lopez de Mantaras and D. Poole, eds.) 293–301. Morgan Kaufmann, San Mateo, CA.

HECKERMAN, D., HORVITZ, E. and NATHWANI, B. (1992). Toward normative expert systems. I: the PATHFINDER project. *Methods of Information in Medicine* **31** 90–105.

KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–52.

LAURITZEN, S. L. (1996). *Graphical Models*. Oxford Univ. Press.

LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505.

LAURITZEN, S. L. and SPIEGELHALTER, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **50** 157–224.

LAURITZEN, S. L. and WERMUTH, N. (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* **17** 31–57.

MADIGAN, D., ANDERSSON, S. A., PERLMAN, M. D. and VOLINSKY, C. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Comm. Statist. Theory Methods*. To appear.

MADIGAN, D. and RAFTERY, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *J. Amer. Statist. Assoc.* **89** 1535–1546.

MADIGAN, D. and YORK, J. (1995). Bayesian graphical models for discrete data. *Internat. Statist. Rev.* **63** 215–232.

MEEK, C. (1995). Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence* (P. Besnard and S. Hanks, eds.) 403–410. Morgan Kaufmann, San Mateo, CA.

NEAPOLITAN, R. E. (1990). *Probabilistic Reasoning in Expert Systems*. Wiley, New York.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA.

ROBINSON, R. W. (1977). Counting unlabeled acyclic digraphs. In *Proceedings of the Fifth Australian Conference on Combinatorial Mathematics* (C. H. C. Little, ed.) 28–43. Springer-Verlag, Berlin.

SPIEGELHALTER, D. J., DAWID, A. P., LAURITZEN, S. L. and COWELL, R. G. (1993). Bayesian analysis in expert systems (with discussion). *Statist. Sci.* **8** 219–283.

SPIEGELHALTER, D. J. and LAURITZEN, S. L. (1990). Sequential updating of conditional probabilities on directed graphical structures. *Networks* **20** 579–605.

VERMA, T. and PEARL, J. (1990). Equivalence and synthesis of causal models. In *Uncertainty in Artificial Intelligence: Proceedings of the Sixth Conference* (M. Henrion, R. Shachter, L. Kanal and J. Lemmer, eds.) 220–227. Morgan Kaufman, San Francisco.

VERMA, T. and PEARL, J. (1992). An algorithm for deciding if a set of observed independencies has a causal explanation. In *Uncertainty in Artificial Intelligence: Proceedings of the Eighth Conference* (D. Dubois, M. P. Wellman, B. D'Ambrosio and P. Smets, eds.) 323–330. Morgan Kaufman, San Francisco.

WERMUTH, N. and LAURITZEN, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models (with discussion). *J. Roy. Statist. Soc. Ser. B* **52** 21–72.

WHITTAKER, J. L. (1990). *Graphical Models in Applied Multivariate Statistics*. Wiley, New York.

WRIGHT, S. (1921). Correlation and causation. *J. Agricultural Research* **20** 557–585.

YORK, J., MADIGAN, D., HEUCH, I. and LIE, R. T. (1995). Estimating a proportion of birth defects by double sampling: a Bayesian approach incorporating covariates and model uncertainty. *J. Roy. Statist. Soc. Ser. C* **44** 227–242.

STEEN A. ANDERSSON
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, INDIANA 47405

DAVID MADIGAN
MICHAEL D. PERLMAN
DEPARTMENT OF STATISTICS
BOX 354322
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195-4322