© Board of the Foundation of the Scandinavian Journal of Statistics 1997. Published by Blackwell Publishers Ltd, 108 Cowley Road, Oxford OX4 1JF, UK and 350 Main Street, Malden, MA 02148, USA. Vol. 24: 81–102, 1997

On the Markov Equivalence of Chain Graphs, Undirected Graphs, and Acyclic Digraphs

STEEN A. ANDERSSON

Indiana University

DAVID MADIGAN and MICHAEL D. PERLMAN

University of Washington

ABSTRACT. Graphical Markov models use undirected graphs (UDGs), acyclic directed graphs (ADGs), or (mixed) chain graphs to represent possible dependencies among random variables in a multivariate distribution. Whereas a UDG is uniquely determined by its associated Markov model, this is not true for ADGs or for general chain graphs (which include both UDGs and ADGs as special cases). This paper addresses three questions regarding the equivalence of graphical Markov models: when is a given chain graph Markov equivalent (1) to some UDG? (2) to some (at least one) ADG? (3) to some decomposable UDG? The answers are obtained by means of an extension of Frydenberg's (1990) elegant graph-theoretic characterization of the Markov equivalence of chain graphs.

Key words: acyclic directed graph, chain graph, chordal graph, conditional independence, decomposable graph, graphical models, Markov equivalence, undirected graph

1. Introduction

The use of graphs to represent dependence relations among random variables, first introduced by Wright (1921), has generated considerable research activity, especially since the early 1980s. Particular attention has been devoted to graphical Markov models, where a graph is used to specify certain conditional independence relations among the variables. These models have found diverse applications, including image analysis, spatial statistics, categorical data analysis, pedigree analysis, and expert systems. The recent books by Pearl (1988), Whittaker (1990), Edwards (1995), and Lauritzen (1996) conveniently summarize the theoretical and applied aspects of graphical Markov models.

In these graphical models, the vertices of the graph represent random variables and the links between the vertices (which can be either directed or undirected) represent the absence of conditional independence. Much of the research on graphical models in the 1980s focused on "pure" graphs, i.e. either undirected graphs, where all the links are undirected, or acyclic digraphs, where all the links are directed and no directed cycles occur. For example, the acyclic digraphs (ADG) of Fig. 1(a), (b), and (c) embody the assumption that β and γ are conditionally independent given α. (We now use the phrase "acyclic digraph" rather than the more common, but inaccurate, "directed acyclic graph" after Brian Alspach kindly pointed out to us that these adjectives do not commute.) The undirected graph (UDG) of Fig. 1(d) represents the same assumption, although in general, directed and undirected graphs with the same vertices and links may represent different sets of conditional independencies.

In the mid-1980s, Lauritzen, Wermuth, and Frydenberg introduced graphical models defined by "chain graphs" (Lauritzen & Wermuth, 1989; Frydenberg, 1990). This followed earlier work in this direction by Goodman (1973), Asmussen & Edwards (1983), and Kiiveri *et al.* (1984). Chain graphs may have both directed and undirected edges, and include both the ADGs and UDGs as special cases. The defining property of a chain graph is that it

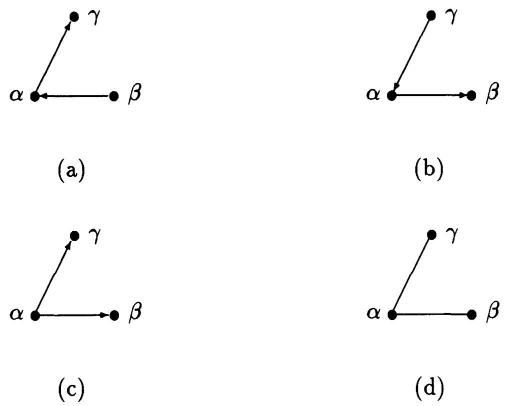

Fig. 1. Three simple acyclic digraphs (ADGs) and a simple undirected graph (UDG). Here α, β, and γ represent (possibly dependent) random variables. All four graphical models embody the Markov assumption that β and γ are conditionally independent given α.

contains no cycles involving one or more directed edges. Chain graphs provide much of the focus for current research on modelling statistical dependence, see for example, Lauritzen (1989), Wermuth & Lauritzen (1990), and Cox & Wermuth (1993).

While chain graphs do provide a more general modelling class, ADG models admit especially simple statistical analysis, and have become popular across a diverse range of applications; see, for example, Lauritzen & Spiegelhalter (1988), Pearl (1988), Neapolitan (1990), Spiegelhalter & Lauritzen (1990), Spiegelhalter *et al.* (1993), Madigan & Raftery (1994) and York *et al.* (1995). In particular, ADG models provide a convenient recursive factorization of the joint probability density function (see (3.1)), an elegant framework for Bayesian analysis, and, in expert system applications, a simple causal interpretation. 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. Furthermore, the only UDGs that provide these conveniences are the *decomposable* UDG models; these decomposable UDGs are exactly those that are Markov equivalent to (i.e. have the same Markov properties as) some ADG (see Fig. 2 and corollary 4.1 below).

Whereas a UDG is uniquely determined by its associated Markov model (see remark 3.1), this need not be true for ADGs (see Fig. 1) or for chain graphs. In particular, a given chain graph G may be Markov equivalent to some UDG, to some ADG(s), or to both. In particular, if G is determined to be equivalent to an ADG, then its statistical analysis can be simplified substantially.

These considerations suggest the following three questions.

1. When is a chain graph, G, Markov equivalent to *some* (necessarily unique) UDG?
2. When is a chain graph, G, Markov equivalent to *some* (not necessarily unique) ADG? In this case, is there an efficient algorithm to generate *at least one* (or *all*) Markov equivalent ADGs from G?
3. When is a chain graph, G, Markov equivalent to *some* (necessarily unique) decomposable UDG?

Question 1 was essentially answered by Frydenberg (1990, p. 348). Prop. 8.2 of Lauritzen & Wermuth (1989) presents a sufficient condition for a given chain graph, G, to be Markov equivalent to a *given* ADG. Further, they conjecture that their condition is also necessary.

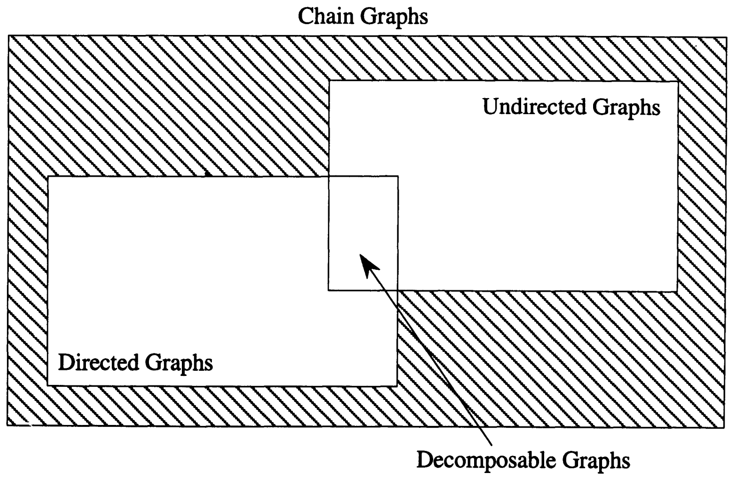

Fig. 2. Four classes of graphical Markov models. Chain graphs include UDGs, ADGs, and decomposable UDGs as special cases.

Their question is related to, but not identical to, our question 2. Similarly, their prop. 8.3 addresses a question related to our question 3.

In this paper we resolve questions 1, 2, and 3 in a unified fashion, in propositions 4.1, 4.2, and 4.3, respectively. Our main result is proposition 4.2, where we provide necessary and sufficient conditions for a given chain graph, G, to be Markov equivalent to some ADG. Also, we present the required algorithm for generating *at least one* Markov equivalent ADG from G. (See remark 4.1 regarding how to generate *all* Markov equivalent ADGs from G.) This work is based on our extension to general probability measures (see theorem 3.1) of Frydenberg's (1990) elegant graph-theoretic characterization of the Markov equivalence of two chain graphs. Thus our arguments are graph-theoretic rather than probabilistic.

Various authors have shown (with varying degrees of generality) that the intersection of the classes of UDG models and ADG models is the class of decomposable UDG models (Wermuth, 1980, prop. 5; Wermuth & Lauritzen, 1983, prop. 5; Asmussen & Edwards, 1983, coroll. 3.4; Kiiveri *et al.*, 1984, coroll., p. 39). By combining our answers to questions 1, 2, and 3, we obtain a purely graph-theoretic demonstration of this result in complete generality (corollary 4.1).

2. Definitions and notation

Our development closely follows that of Frydenberg (1990), with one exception noted below. We consider multivariate probability distributions on a product probability space $\mathscr{X} \equiv \times_{\alpha \in V} \mathscr{X}_\alpha$, where V is a finite index set and each $\mathscr{X}_\alpha$ is sufficiently regular to ensure the existence of regular conditional probabilities. Such distributions are conveniently represented by a random variate $X := (X_\alpha : \alpha \in V) \in \mathscr{X}$. For any subset $A \subseteq V$, we define $X_A := (X_\alpha : \alpha \in A)$. We often abbreviate X_α and X_A by α and A, respectively, and define $X_\emptyset \equiv$ constant.

A graphical model is defined by a collection of conditional independencies among the component random variates $(X_\alpha : \alpha \in V)$, which collection is represented by a *graph* $G \equiv (V, E)$ with vertex set V. The set of edges, E, is a subset of $E^*(V) \equiv (V \times V) \setminus \{(\alpha, \alpha) \mid \alpha \in V\}$, i.e. a set of ordered pairs of distinct vertices. An edge $(\alpha, \beta) \in E$ whose opposite $(\beta, \alpha) \in E$ also, is called an *undirected* edge and appears as a line

$\alpha — \beta$ in our figures, whereas an edge $(\alpha, \beta) \in E$ whose opposite $(\beta, \alpha) \notin E$, is called a *directed* edge and appears as an arrow: $\alpha \to \beta$. (Our notation differs from Frydenberg's in this regard: he uses the notation $\alpha \Rightarrow \beta$ instead of $\alpha \to \beta$ in his text, although not in his figures.)

If $A \subseteq V$ is a subset of the vertex set, it induces a subgraph $G_A = (A, E_A)$, where the edge set $E_A \equiv E \cap (A \times A)$ is obtained from G by retaining all edges with both endpoints in A. If $A \subseteq B \subseteq V$, clearly $G_A = (G_B)_A$.

For a graph $G = (V, E)$, we will denote the *skeleton* of G by $G^u = (V, E^u)$, where $E^u = \{(\alpha, \beta) \mid (\alpha, \beta) \in E \text{ or } (\beta, \alpha) \in E\}$; G^u is just the underlying undirected graph associated with G. For any subset $A \subseteq V$, $(G_A)^u = (G^u)_A$. Two vertices $\alpha, \beta \in V$ are called *adjacent* in G if $(\alpha, \beta) \in E^u$.

For a subset $A \subseteq V$, $\mathrm{bd}(A) \equiv \{\beta \in V \backslash A \mid (\beta, \alpha) \in E \text{ for some } \alpha \text{ in } A\}$ and $\mathrm{cl}(A) \equiv \mathrm{bd}(A) \cup A$ denote the *boundary* and *closure* of A in G, respectively. For any vertex $\alpha \in V$, the *neighbours* of α in G are those vertices linked to α by undirected edges. For any vertex $\alpha \in V$ and any subset $A \subseteq V$, $\mathrm{ch}_A(\alpha) \equiv \{\beta \in A \mid (\alpha, \beta) \in E \text{ and } (\beta, \alpha) \notin E\}$ is the set of all *children* of α that occur in A. The *parents* of α in G, denoted by $\mathrm{pa}(\alpha) \equiv \{\beta \in V \mid (\beta, \alpha) \in E \text{ and } (\alpha, \beta) \notin E\}$, are those vertices linked to α by directed edges. For $A \subseteq V$, define $\mathrm{pa}(A) = \bigcup_{\alpha \in A} \mathrm{pa}(\alpha) \backslash A$. For subsets $A \subseteq B \subseteq V$, we write $\mathrm{bd}_B(A)$ and $\mathrm{cl}_B(A)$ to denote the boundary and closure of A in the induced subgraph G_B.

A *path* π of length $n \geqslant 1$ from α to β in G is a sequence $\pi \equiv \{\alpha_0, \alpha_1, \dots, \alpha_n\} \subseteq V$ of distinct vertices such that $\alpha_0 = \alpha$, $\alpha_n = \beta$, and $(\alpha_{i-1}, \alpha_i) \in E$ for all $i = 1, \dots, n$. If (α_{i-1}, α_i) is directed for at least one i, we call the path *directed*, and if this is not the case, we call it *undirected*. A *cycle* is a path with the modification that $\alpha_n = \alpha_0$. The *non-descendants* of α are those $\beta \in V$ such that there is no directed path from α to β.

A graph is called a *chain graph* if it does not contain any directed cycles. If the graph has only undirected edges it is an *undirected graph* (UDG). If all the edges are directed, and the graph contains no directed cycles, the graph is said to be an *acyclic digraph* (ADG).

Let $G \equiv (V, E)$ be a UDG. A set of vertices $A \subseteq V$ is *connected* in G if, for every distinct $\alpha, \beta \in A$, there is a path from α to β in G. For pairwise disjoint subsets $A(\neq \emptyset)$, $B(\neq \emptyset)$, and S of V, A and B are *separated* by S in G if all paths within G from vertices in A to vertices in B include vertices in S. Note that if $S = \emptyset$, then A and B are separated by S in G if and only if there are *no* paths from A to B in G. In this case, A and B are separated by *any* subset S disjoint from A and B.

An undirected graph is *complete* if all pairs of vertices are joined by an edge. Trivially, the empty graph is complete. A subset $A \subseteq V$ is *complete* in G if it induces a complete subgraph.

A subset $A \subseteq V$ is *simplicial* in the UDG, G, if its boundary is complete. A pair (A, B) of non-empty subsets of V is said to form a *decomposition* of G if $V = A \cup B$, $A \cap B$ is complete, and $A \cap B$ separates $A \backslash B$ from $B \backslash A$ in G. When this is the case we say that (A, B) decomposes G into the *components* G_A and G_B. If the sets A and B in (A, B) are both proper subsets of V, the decomposition is *proper*. An undirected graph is said to be *decomposable* if it is complete, or if there exists a proper decomposition (A, B) into decomposable subgraphs G_A and G_B.

A UDG is *chordal* if every cycle of length $n \geqslant 4$ possesses a *chord*, that is, two non-consecutive vertices that are neighbours. A well-known result states that a UDG is decomposable if and only if it is chordal, cf. Lauritzen *et al.* (1984, th. 2) or Whittaker (1990, prop. 12.4.2).

For the remainder of this section, let $G \equiv (V, E)$ be a chain graph. If $A \subseteq V$ and $\alpha, \beta \in A$, we write $\alpha \leqslant_A \beta$ if $\alpha = \beta$ or there is a path from α to β in G_A. If both $\alpha \leqslant_A \beta$ and $\beta \leqslant_A \alpha$, then we write $\alpha \approx_A \beta$. When $A = V$, we write $\leqslant (\equiv \leqslant_G)$ for $\leqslant_V$ and $\approx (\equiv \approx_G)$ for $\approx_V$. Frydenberg (1990) notes that $\approx$ is an equivalence relation on V and denotes the induced set of equivalence classes by $\mathscr{T}(G)$, the set of *chain components* of G. Equivalently, $\mathscr{T}(G)$ is the set of connected components of the undirected graph obtained from G by removing all directed edges. Note that for each $\tau \in \mathscr{T}(G)$, $\text{bd}(\tau) = \text{pa}(\tau) \equiv \{\text{pa}(\alpha) \mid \alpha \in \tau\}$. Both UDGs and ADGs are special cases of chain graphs: the chain components of a UDG are its connected components, while for an ADG, every chain component consists of a single vertex.

For any subset $A \subseteq V$, the induced subgraph G_A is also a chain graph, and $\mathscr{T}(G_A) = \{\tau \cap A \mid \tau \in \mathscr{T}(G)\}$. Thus, if $\tau \in \mathscr{T}(G)$ and $\tau \subseteq A$, then $\tau \in \mathscr{T}(G_A)$.

Let $\alpha < \beta$ stand for the statement "there exists a directed path from α to β". We define the *future* of α in G by $\phi(\alpha) = \{\beta \mid \alpha < \beta\}$, and the *past* of α in G by $\pi(\alpha) = \{\beta \mid \beta < \alpha\}$. If $\tau \in \mathscr{T}(G)$, the future or the past is the same for all vertices in τ and thus we may use $\phi(\tau)$ and $\pi(\tau)$ to denote their common future and past in G, respectively. Furthermore, we call τ *terminal* in G if $\phi(\tau)$ is empty and *initial* in G if $\pi(\tau)$ is empty.

A subset $A \subseteq V$ is called a *G-anterior* set (or simply an *anterior* set if G is understood) if it can be generated by stepwise removal of terminal chain components. (Note that the removal of a terminal chain component might render other chain components terminal in the remaining graph.) Equivalently, it can be shown that A is anterior if and only if $\beta \in A$ whenever $\alpha \in A$, $\beta \in V$, and $\beta \leqslant \alpha$. Also equivalently, A is anterior if and only if $\text{bd}(A) = \emptyset$. If A is anterior and τ is a terminal chain component of G such that $\tau \subseteq A$, then τ is a terminal chain component of G_A.

If A and B are anterior sets, then $A \cap B$ and $A \cup B$ are anterior. If $B \subseteq A \subseteq V$ and B is G-anterior, then B is G_A-anterior. If $B \subseteq A \subseteq V$ and A is G-anterior, then B is G-anterior if and only if B is G_A-anterior.

For any subsets $B \subseteq A \subseteq V$, let $\text{an}_A(B)$ denote the smallest G_A-anterior subset of A that contains B. Clearly $\text{an}_A(B) = \{\alpha \in A \mid \alpha \leqslant_A \beta \text{ for some } \beta \in B\}$, so if A if G-anterior and $B \subseteq A$, then $\text{an}_A(B) = \text{an}(B)$. Note that for any subsets $B \subseteq A$ and $C \subseteq A$, $\text{an}_A(B \cup C) = \text{an}_A(B) \cup \text{an}_A(C)$ (but $\cup$ cannot be replaced by $\cap$).

We call a triple (α, B, β) a *complex* in G if B is a connected subset of a chain component $\tau \in \mathscr{T}(G)$ and α and β are two non-adjacent vertices in $\text{bd}(\tau) \cap \text{bd}(B)$. Further, we call a complex (α, B, β) a *minimal* complex in G if no proper subset $B' \subset B$ forms a complex (α, B', β). Every complex (α, B', β) contains at least one minimal complex, (α, B, β), $B \subseteq B'$. A minimal complex (α, B, β) is called an *immorality* if B contains only one vertex. Frydenberg (1990) notes that (α, B, β) is a minimal complex in G if and only if $G_{B \cup \{\alpha, \beta\}}$ looks like the chain graph of Fig. 3. For any subset $A \subseteq V$ such that $\{\alpha, B, \beta\} \subseteq A$, (α, B, β) is a minimal complex in G_A if and only if it is a minimal complex in G.

A chain graph $\tilde{G} \equiv (\tilde{V}, \tilde{E})$ is *larger* than the chain graph $G \equiv (V, E)$, indicated by $G \subseteq \tilde{G}$, if $V \subseteq \tilde{V}$ and $E \subseteq \tilde{E}$. Thus, if $V = \tilde{V}$ and $G^u = \tilde{G}^u$, then $G \subseteq \tilde{G}$ if and only if they differ only in that some directed edges (arrows) in G may be converted to undirected edges (lines) in $\tilde{G}$. In this case, each $\tau \in \mathscr{T}(G)$ is contained in a *unique* $\tilde{\tau} \equiv \psi(\tau) \in \mathscr{T}(\tilde{G})$, and $\psi: \mathscr{T}(G) \to \mathscr{T}(\tilde{G})$ is an order-preserving mapping, i.e. a poset homeomorphism, under the partial orders induced by $\leqslant_G$ and $\leqslant_{\tilde{G}}$, respectively. Furthermore, $\text{bd}_G(\tau) = \text{bd}_{\tilde{G}}(\tau)$ and $\text{cl}_G(\tau) = \text{cl}_{\tilde{G}}(\tau)$.

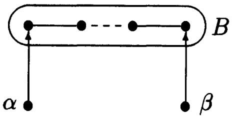

Fig. 3. A simple chain graph. Here (α, B, β) is a minimal complex.

The *moral graph*, G^m, determined by G, is defined to be the undirected graph $G^m \equiv (V, E^m)$, where $E^m = E^u \cup \bigcup_{\tau \in \mathcal{T}(G)} E^*(\mathrm{bd}(\tau))$. That is, G^m is the skeleton G^u augmented by all (undirected) edges needed to make $\mathrm{bd}(\tau)$ complete in G^m for every chain component $\tau \in \mathcal{T}(G)$. Equivalently, G^m is obtained from G^u by adding an undirected edge $\alpha - \beta$ whenever (α, B, β) is a minimal complex in G; thus $G \subseteq G^m$.

For any subset $A \subseteq V$, $(G_A)^m = (A, (E_A)^m)$, where $(E_A)^m = (E_A)^u \cup \bigcup_{\tau' \in \mathcal{T}(G_A)} E^*(\mathrm{bd}_A(\tau'))$; thus $(G_A)^m \subseteq (G^m)_A$. For every chain component $\tau \in \mathcal{T}(G)$, $(G_{\mathrm{cl}(\tau)})^m$ is the skeleton $(G_{\mathrm{cl}(\tau)})^u$ augmented by those undirected edges $\alpha - \beta$ such that (α, B, β) is a minimal complex for some $B \subseteq \tau$. Furthermore, $(G_{\mathrm{cl}(\tau)})^m = (G^m)_{\mathrm{cl}(\tau)}$ for every terminal chain component $\tau \in \mathcal{T}(G)$.

3. The Markov properties for chain graphs

First, we describe informally the Markov properties specified by ADGs and UDGs, then do so formally for chain graphs, which include both ADGs and UDGs as special cases. The Markov property specified by an ADG embodies the hypothesis that for each vertex $\alpha \in V$, the parents of α are the only direct influences on α. That is, a probability measure P on $\mathcal{X}$ satisfies the (local) Markov properties specified by G if, under P, each $X_\alpha, \alpha \in V$, is conditionally independent of its non-descendants given its parents. When P admits a joint density $p \equiv p(V)$ with respect to some product measure on $\mathcal{X}$, this property implies a recursive factorization of p given by:

$$p(V) = \prod_{\alpha \in V} p(\alpha \mid \mathrm{pa}(\alpha)). \tag{3.1}$$

The class of probability models that can be defined in this way was introduced by Wermuth & Lauritzen (1983) and Kiiveri *et al.* (1984) and are a subclass of their class of recursive causal models.

A probability measure P on $\mathcal{X}$ satisfies the (local) Markov property specified by a UDG (see for example Fig. 1(d)) if, under P, each $X_\alpha, \alpha \in V$ is conditionally independent of all others given its neighbours. If P admits a positive joint density p on $\mathcal{X}$, this (local) Markov property does *not* imply a recursive factorization of p, unless the UDG is also decomposable. The simplicity of decomposable models has been exploited in a number of contexts—see for example Lauritzen & Spiegelhalter (1988), Madigan & Mosurski (1990), Dawid & Lauritzen (1993), and Madigan & Raftery (1994). For a more detailed exposition of Markov properties with respect to ADGs and UDGs, we refer the reader to Lauritzen *et al.* (1990) and Dawid & Lauritzen (1993).

Frydenberg (1990) defined three versions of the Markov property for a general chain graph $G \equiv (V, E)$. First, for three pairwise disjoint subsets A, B, and C of V, and P a probability measure on $\mathcal{X}$, we write $A \perp B \mid C[P]$ if X_A and X_B are conditionally independent given X_C under P. Trivially, $A \perp B \mid C[P]$ if $A = \emptyset$ or $B = \emptyset$. If A, B, and C are not disjoint, then $A \perp B \mid C[P]$ is defined to mean $[A \setminus (B \cup C)] \perp [B \setminus (A \cup C)] \mid C[P]$. Dawid (1980) showed that the following properties hold for any probability measure P:

(CI1) $A \perp B \mid C[P]$ implies $B \perp A \mid C[P]$
(CI2) $A \perp B \cup C \mid D[P]$ implies $A \perp B \mid D[P]$
(CI3) $A \perp B \cup C \mid D[P]$ implies $A \perp B \mid C \cup D[P]$
(CI4) ($A \perp B \mid D[P]$ and $A \perp C \mid D \cup B[P]$) implies $A \perp B \cup C \mid D[P]$

whenever A, B, C, and D are disjoint subsets of V. Note that the converse implication in (CI4) is also valid by (CI2) and (CI3).

Definition 3.1

A probability measure P on $\mathscr{X}$ is said to be:

(P) *Pairwise G-Markovian* if $\alpha \perp \beta \mid [V \backslash \phi(\alpha)] \backslash \{\alpha, \beta\}[P]$ whenever $\beta \notin \phi(\alpha)$ and β and α are not adjacent;

(L) *Local G-Markovian* if $\alpha \perp [V \backslash \phi(\alpha)] \backslash \mathrm{cl}(\alpha) \mid \mathrm{bd}(\alpha)[P]$ for all α;

(G) *Global G-Markovian* if $A \perp B \mid S[P]$ whenever S separates A and B in $(G_{\mathrm{an}(A \cup B \cup S)})^m$.

Frydenberg (1990, p. 339) notes that if G is a UDG, then P is global G-Markovian if and only if $A \amalg B \mid S[P]$ whenever S separates A and B in G. For ADGs, Lauritzen *et al.* (1990) show that (G) $\Leftrightarrow$ (L) $\Leftrightarrow$ (P) for any P on $\mathscr{X}$. For general chain graphs, Frydenberg (1990) shows that (G) $\Rightarrow$ (L) $\Rightarrow$(P) for any P on $\mathscr{X}$, while if P satisfies the following property CI5, then (P), (L), and (G) are equivalent for P (Frydenberg, 1990, th. 3.3):

(CI5) $\quad A \perp B \mid D \cup C[P]$ and $A \perp C \mid D \cup B[P]$ implies $A \perp B \cup C \mid D[P]$

whenever A, B, C, and D are disjoint subsets of V. Note that CI5 is satisfied whenever P has a *positive* joint probability density with respect to some product measure on $\mathscr{X}$.

Definition 3.2

Two chain graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are called *Markov equivalent* if, for every product space $\mathscr{X}$ indexed by V, the classes of *global* G_1-Markovian and *global* G_2-Markovian probability measures P on $\mathscr{X}$ coincide. In this case we write $G_1 \overset{\mathrm{M}}{\sim} G_2$.

Th. 5.6 of Frydenberg (1990) states that two chain graphs are Markov equivalent if and only if they have the same skeleton and the same minimal complexes. However, his treatment of Markov equivalence is entirely restricted to probability measures P satisfying CI5. In the Appendix we remove this restriction and establish the following stronger result, which is of independent interest.

Theorem 3.1

Two chain graphs $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ are Markov equivalent if and only if G_1 and G_2 have the same skeleton and the same minimal complexes.

Proof. See Appendix.

Remark 3.1. It follows immediately from theorem 3.1 that two UDGs are Markov equivalent if and only if they are identical. Thus a UDG is uniquely determined by its associated Markov model. However, the same Markov model may also be specified by other chain graphs, including ADGs (see corollary 4.1).

Remark 3.2. Theorem 3.1 also implies that if a chain graph G is Markov equivalent to some ADG, then G can have no minimal complexes other than immoralities. The converse is false—see example 4.1 and subsequent discussion.

Remark 3.3. If a probability measure has a *strictly positive* density, f, with respect to some product measure $\mathscr{X}$, then P satisfies CI5. That this strict positivity condition on f is not necessary for P to satisfy CI5 can be shown by the example of the uniform (with respect to Lebesgue measure on $\mathscr{X} \equiv \mathfrak{R}^3$) probability distribution on the skewed cube given by the convex hull of the points $(0, 0, 0)$, $(1, 0, 0)$, $(0, 1, 0)$, $(1, 1, 0)$, $(0, 1, 1)$, $(1, 1, 1)$, $(0, 2, 1)$, $(1, 2, 1)$ in $\mathfrak{R}^3$. Under this distribution, the three relations $X \perp Y \mid Z$, $X \perp Z \mid Y$, and

$X \perp (Y, Z)$ are valid, but $Y \perp Z \mid X$ is not, so CI5 is satisfied. Finally, any random vector in $\mathfrak{R}^3$ of the form $(X, Y \equiv g(X), Z \equiv h(X))$, where g and h are strictly increasing functions and X is a non-degenerate random variable, does *not* satisfy CI5.

Remark 3.4. Because the global, local, and pairwise Markov properties coincide for probability measures that satisfy CI5, Frydenberg's Markov equivalence theorem (1990, th. 5.6) is valid regardless of whether Markov equivalence is defined in terms of the global, local, or pairwise property. This implies that the "only if" assertion in our theorem 3.1 also remains valid if we define Markov equivalence in terms of the local or pairwise, rather than global, Markov property. However, the "if" assertion in theorem 3.1 does not remain valid in this case. Consider, for example, the pair of graphs G_1 and G_2 in Fig. 4. Because G_2 is an ADG, its global, local, and pairwise Markov properties coincide; they are given by the single condition $12 \perp 34 \mid 5$. For the UDG G_1, the global Markov property is also given by $12 \perp 34 \mid 5$, whereas its local and pairwise Markov properties are the following:

local: $1 \perp 34 \mid 25$, $2 \perp 34 \mid 15$, $3 \perp 12 \mid 45$, $4 \perp 12 \mid 35$;

pairwise: $1 \perp 3 \mid 245$, $1 \perp 4 \mid 235$, $2 \perp 3 \mid 145$, $2 \perp 4 \mid 135$.

If X and Y are independent and non-degenerate random variables, then $(X_1, X_2, X_3, X_4, X_5) \equiv (X, X, X, X, Y)$ satisfies the local and pairwise Markov properties for G_1 but not for G_2. However, $G_1^u = G_2^u$ and G_1 and G_2 have the same ($\equiv$no) minimal complexes. [The construction of such examples is facilitated by props 1 and 2 of Matúš (1992).]

Fig. 4. Two chain graphs with identical global Markov properties but non-identical local and pairwise Markov properties.

4. The Markov equivalences of a chain graph

While our primary result concerns the Markov equivalence of chain graphs and ADGs (proposition 4.2 below), we also present conditions for Markov equivalence of chain graphs with UDGs (proposition 4.1) and decomposable UDGs (proposition 4.3).

Proposition 4.1

Let G be a chain graph. The following are equivalent:

(1) *G is Markov equivalent to some (necessarily unique) UDG.*

(1′) *G has no minimal complexes.*

(1″) $G \overset{\mathrm{M}}{\sim} G^u$.

Proof. This follows directly from theorem 3.1.

Example 4.1. Consider the chain graphs G_3, G_4, G_5, and G_6 in Fig. 5. Since G_3 has no minimal complexes, G_3 is Markov equivalent to the UDG $G_3^u \equiv G_4$, by proposition 4.1. Since

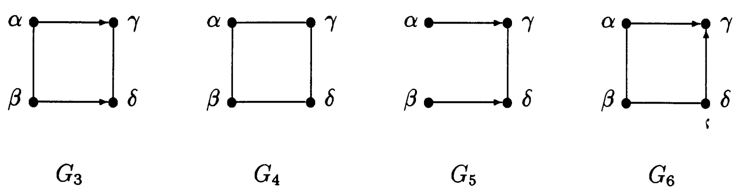

Fig. 5. Four chain graphs. The first two are Markov equivalent.

G_5 has one minimal complex, it is not Markov equivalent to any UDG. Lastly, G_6 has one immorality, hence is not Markov equivalent to any UDG.

In view of proposition 4.1 and remark 3.2, one might surmise that a chain graph G is Markov equivalent to some ADG if and only if G has no minimal complexes other than immoralities. Were this so, then *every* UDG would be Markov equivalent to some ADG, but this is known to be false: only *decomposable* UDGs are Markov equivalent to some ADG (see corollary 4.1). Thus, the non-decomposable UDG, G_4, in example 4.1 is not Markov equivalent to any ADG: any orientation of the edges of G_4 must produce one immorality not present in G_4. Note that G_4 has exactly one chain component τ, namely $\tau = G_4$ itself, and this chain component is not decomposable.

Similarly, the chain graph G_3 in example 4.1 is not Markov equivalent to any ADG. (Indeed, G_3 is Markov equivalent to G_4, and hence to no ADG.) As with G_4, G_3 has no minimal complexes, but any orientation of the edge between γ and δ must produce an immorality not present in G_3. Note that G_3 has exactly two chain components, $\tau_1 \equiv \{\alpha, \beta\}$ and $\tau_2 \equiv \{\gamma, \delta\}$. Looking ahead to proposition 4.2, we see that the difficulty arises because, for $G = G_3$, the moral graph $(G_{\mathrm{cl}(\tau_2)})^m \equiv G_4$ is not decomposable. (Recall that the global Markov property is defined in terms of certain moral graphs.)

Next, the chain graph G_5 in example 4.1 has one minimal complex, which is not an immorality, hence by remark 3.2, G_5 cannot be Markov equivalent to any ADG: either orientation of the single undirected edge in G_5 will produce an immorality not present in G_5. Here, G_5 has three chain components, $\tau_1 = \{\alpha\}$, $\tau_2 = \{\beta\}$, and $\tau_3 = \{\gamma, \delta\}$, and for $G = G_5$, $(G_{\mathrm{cl}(\tau_3)})^m \equiv G_4$ is not decomposable.

Finally, the chain graph G_6 in example 4.1 has one minimal complex, an immorality. There are four possible orientations of the two undirected edges of G_6 (see Fig. 6), of which three produce no additional immoralities, hence produce ADGs to which G_6 is Markov equivalent. Here, G_6 has two chain components, $\tau_1 = \{\alpha, \beta, \delta\}$ and $\tau_2 = \{\gamma\}$. For $G = G_6$, both $(G_{\mathrm{cl}(\tau_1)})^m$ ($= G_{\tau_1}$) and $(G_{\mathrm{cl}(\tau_2)})^m$ (= the complete triangle with vertices α, γ, and δ) are decomposable, hence G_6 satisfies condition (2′) of proposition 4.2 below, our main result concerning the Markov equivalence of chain graphs and ADGs.

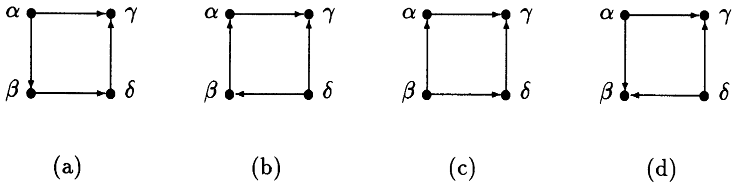

Fig. 6. The four possible orientations of G_6. By theorem 3.1, the first three ADGs are Markov equivalent to G_6 while the fourth is not.

First we require a lemma:

Lemma 4.1
Let G be a chain graph such that $(G_{\mathrm{cl}(\tau)})^m$ is decomposable for every chain component $\tau \in \mathscr{T}(G)$. Then G has no minimal complexes other than immoralities.

Proof. If (α, B, β) is a minimal complex in G, then (α, B, β) determines a chordless n-cycle in $(G_{\mathrm{cl}(\tau)})^m$, where $\tau \in \mathscr{T}(G)$ is the unique chain component containing B and where $n = |B| + 2$ (see Fig. 3). Thus, if $|B| \geqslant 2$ then $(G_{\mathrm{cl}(\tau)})^m$ is not decomposable.

The converse of lemma 4.1 is *not* true: any non-decomposable UDG provides a counter-example, for example, G_4 in Fig. 5. The chain graph G_3 also provides a counter-example.

Proposition 4.2
Let $G \equiv (V, E)$ be a chain graph. The following are equivalent:
(2) *G is Markov equivalent to some (not necessarily unique) ADG, D.*
(2′) *$(G_{\mathrm{cl}(\tau)})^m$ is decomposable for all $\tau \in \mathscr{T}(G)$.*

Proof. (2) $\Rightarrow$ (2′): Suppose that $G \overset{\mathrm{M}}{\sim} D$ but that $(G_{\mathrm{cl}(\tau)})^m$ is not decomposable for some $\tau \in \mathscr{T}(G)$. Then $(G_{\mathrm{cl}(\tau)})^m$ must contain at least one chordless cycle, $C \equiv \{\alpha_0, \alpha_1, \ldots, \alpha_n \equiv \alpha_0\}$, $n \geqslant 4$, with $C \subseteq \mathrm{cl}(\tau) \equiv \mathrm{bd}(\tau) \cup \tau$. First, $C \nsubseteq \tau$; otherwise, any acyclic orientation of the edges of this chordless cycle would create at least one immorality in D that was not present in G_τ (since G_τ is undirected), hence which was not present in G. By theorem 3.1, this would violate the assumption that $G \overset{\mathrm{M}}{\sim} D$. Thus $|C \cap \mathrm{bd}(\tau)| \geqslant 1$. Also, $|C \cap \mathrm{bd}(\tau)| \leqslant 2$, since $\mathrm{bd}(\tau)$ is complete in $(G_{\mathrm{cl}(\tau)})^m$ and C is chordless in $(G_{\mathrm{cl}(\tau)})^m$.

If $|C \cap \mathrm{bd}(\tau)| = 1$, assume without loss of generality that $C \cap \mathrm{bd}(\tau) = \alpha_1$. Because $\alpha_1 \in \mathrm{bd}(\tau) = \mathrm{pa}(\tau)$ and α_1 is adjacent in G to both $\alpha_2 \in \tau$ and to $\alpha_n \in \tau$, necessarily both $\alpha_1 \to \alpha_2$ and $\alpha_1 \to \alpha_n$ in G (see Fig. 7a). If $|C \cap \mathrm{bd}(\tau)| = 2$, assume without loss of generality that $C \cap \mathrm{bd}(\tau) = \{\alpha_1, \alpha_2\}$, in which case necessarily both $\alpha_1 \to \alpha_n$ and $\alpha_2 \to \alpha_3$ in G (see Fig. 7b). Since C is chordless and $n \geqslant 4$, it follows that any possible orientation of the edges $\alpha_2 — \cdots — \alpha_n$ in the first case, or $\alpha_3 — \cdots — \alpha_n$ in the second case, will create at least one immorality in D that was not present in G, again contradicting the assumption that $G \overset{\mathrm{M}}{\sim} D$.

(2′) $\Rightarrow$ (2): Let G be a chain graph satisfying (2′). For each chain component $\tau \in \mathscr{T}(G)$, the edges of the decomposable UDG $(G_{\mathrm{cl}(\tau)})^m$ can be oriented by means of the maximum cardinality search (MCS) algorithm to convert $(G_{\mathrm{cl}(\tau)})^m$ into a *perfect* digraph $D^{(\tau)}$, i.e. $D^{(\tau)}$ is an acyclic digraph with no immoralities (see Blair & Peyton (1993, th. 2.5)). The MCS

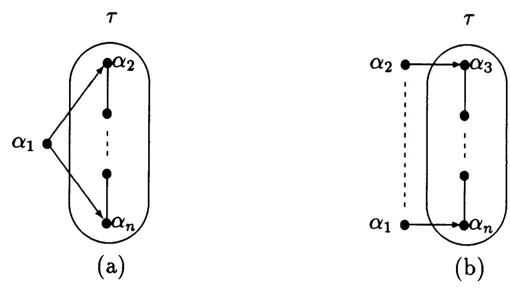

Fig. 7. The induced subgraph G_C. In case (b), α_1 and α_2 need not be adjacent in G, even though they are adjacent in $(G_{\mathrm{cl}(\tau)})^m$.

algorithm begins by assigning the number 1 to any vertex in $\mathrm{cl}(\tau)$, then assigning the numbers $2, \ldots, |\mathrm{cl}(\tau)|$ consecutively to the remaining vertices in $\mathrm{cl}(\tau)$, selecting each time the vertex with the most previously numbered neighbours in $(G_{\mathrm{cl}(\tau)})^m$, breaking ties arbitrarily. Thus we may begin at any vertex in $\mathrm{bd}(\tau)$, then, since $\mathrm{bd}(\tau)$ is complete in $(G_{\mathrm{cl}(\tau)})^m$, number the remaining vertices in $\mathrm{bd}(\tau)$ before numbering any of the vertices in τ. With such a numbering of $\mathrm{cl}(\tau)$, the orientation in $D^{(\tau)}$ of any edge between a vertex in $\mathrm{bd}(\tau)$ and a vertex in τ agrees with the orientation of that edge in G.

Let D be the digraph obtained from G by orienting all undirected edges in G as follows: for every $\tau \in \mathscr{T}(G)$, orient all edges in G_τ according to their orientation in $D^{(\tau)}$. Then D is acyclic: if D possesses a directed cycle, C, then C cannot lie entirely within any chain component $\tau \in \mathscr{T}(G)$ (since $D^{(\tau)}$ is acyclic), hence G_C contains at least one directed edge. But then C is also a directed cycle in the chain graph G, which is impossible.

Thus, D is an ADG with the same skeleton as G. To show that $G \overset{\mathrm{M}}{\sim} D$, by theorem 3.1 we must show that D has the same minimal complexes as G. By lemma 4.1, G has no minimal complexes other than immoralities. Since any immorality of G is also an immorality of D, it suffices to show that D has no immoralities not already in G. Such a new immorality, say (α, γ, β), must involve at least one undirected edge in G, say $\alpha - \gamma$. (Otherwise, this immorality would also occur in G.) Thus $\alpha, \gamma \in \tau$ for some $\tau \in \mathscr{T}(G)$, and $\beta \in \mathrm{cl}(\tau) = \mathrm{bd}(\tau) \cup \tau$. Then (α, γ, β) is also an immorality in $D^{(\tau)}$, since α and β cannot be adjacent in $(G_{\mathrm{cl}(\tau)})^m$. But $D^{(\tau)}$ is perfect, so this yields a contradiction, thus completing the proof.

The ADG, D, constructed in the proof of the implication $(2') \Rightarrow (2)$, is obtained from G by orienting all undirected edges in each induced subgraph G_τ, $\tau \in \mathscr{T}(G)$, according to the orientation of these edges in the corresponding ADG, $D^{(\tau)}$. Thus the orientations of the remaining edges in $D^{(\tau)}$ are not needed in this construction. The following algorithm makes use of this observation to construct efficiently a Markov equivalent ADG, D, from a chain graph G satisfying $(2')$:

The orientation algorithm

For every chain component $\tau \in \mathscr{T}(G)$, perform a modified maximum cardinality search (MCS) on G_τ as follows:

1. Choose $\alpha_1 = \arg\max_{\alpha \in \tau} \{|\mathrm{pa}_G(\alpha)|\}$, breaking ties arbitrarily.
2. For each $i = 2, \ldots, |\tau|$, having chosen $\alpha_1, \ldots, \alpha_{i-1} \in \tau$, now choose:

   $\alpha_i = \arg\max_{\alpha \in \tau \setminus \{\alpha_1, \ldots, \alpha_{i-1}\}} \{|\mathrm{pa}_G(\alpha)| + |\{j \in \{1, \ldots, i-1\}: (\alpha_j, \alpha) \in E\}|\}$,

   breaking ties arbitrarily. This produces an ordering $\{\alpha_1, \ldots, \alpha_{|\tau|}\}$ of all the vertices in τ.
3. Orient the edges of G_τ in accordance with this ordering, i.e. orient $\alpha_j \to \alpha_i$ if and only if $(\alpha_j, \alpha_i) \in E$ and $j < i$.
4. All edges in G are now oriented, producing the required ADG, D.

In order to verify that a chain graph G does indeed satisfy $(2')$, first apply the orientation algorithm to G, obtaining for each $\tau \in \mathscr{T}(G)$ an ordering $\alpha_1, \ldots, \alpha_{|\tau|}$, of the vertices in τ. Next, arbitrarily order the vertices in $\mathrm{bd}(\tau)$: $\beta_1, \ldots, \beta_{|\mathrm{bd}(\tau)|}$. Finally apply the "test for zero fill-in" of Tarjan & Yannakakis (1984, p. 571) to the reverse ordering $\alpha_{|\tau|}, \ldots, \alpha_1, \beta_{|\mathrm{bd}(\tau)|}, \ldots, \beta_1$ of the vertices of $(G_{\mathrm{cl}(\tau)})^m$. If this reversed ordering is "zero fill-in" for $(G_{\mathrm{cl}(\tau)})^m$, then $(G_{\mathrm{cl}(\tau)})^m$ is chordal, hence decomposable.

Remark 4.1. Once we have constructed *one* ADG, D, that is Markov equivalent to a chain graph G satisfying $(2')$, we can generate all such Markov equivalent ADGs as follows. From

D, construct the *essential graph* D^* determined by D by means of the algorithm given by Andersson *et al.* (1996). Here D^* is the graph obtained from D by converting to undirected edges those directed edges in D that occur with the *opposite* orientation in at least one ADG that is Markov equivalent to D (hence to G). Andersson *et al.* (1996) show that D^* is a chain graph, each of whose chain components is decomposable. Furthermore, they show that all ADGs that are Markov equivalent to D (hence to G) can be obtained by assigning all possible perfect ($\equiv$acyclic and moral) orientations to the edges in these decomposable chain components of D^*.

Remark 4.2. An equivalent but more explicit version of condition (2′) can be obtained from the following observation. For any chain component $\tau \in \mathscr{T}(G)$, $(G_{\mathrm{cl}(\tau)})^m$ is decomposable if and only if:

(i) G_τ is decomposable, and

(ii) For every $\alpha \in \mathrm{bd}(\tau)$, and every non-adjacent pair $\gamma, \delta \in \mathrm{ch}_\tau(\alpha)$, $\mathrm{ch}_\tau(\alpha)\setminus\{\gamma, \delta\}$ separates γ and δ in G_τ (in particular $\mathrm{ch}_\tau(\alpha)\setminus\{\gamma, \delta\}$ must be non-empty), and

(iii) For every distinct pair $\alpha, \beta \in \mathrm{bd}(\tau)$, and every pair $\gamma \in \mathrm{ch}_\tau(\alpha)\setminus\mathrm{ch}_\tau(\beta)$, $\delta \in \mathrm{ch}_\tau(\beta)\setminus\mathrm{ch}_\tau(\alpha)$, $[\mathrm{ch}_\tau(\alpha) \cup \mathrm{ch}_\tau(\beta)]\setminus\{\gamma, \delta\}$ separates γ and δ in G_τ (in particular, $[\mathrm{ch}_\tau(\alpha) \cup \mathrm{ch}_\tau(\beta)]\setminus\{\gamma, \delta\}$ must be non-empty and γ and δ must be non-adjacent).

To verify this assertion, first note that the failure of either (i), (ii), or (iii) would imply the existence of a chordless n-cycle in $(G_{\mathrm{cl}(\tau)})^m$ for some $n \geqslant 4$, hence $(G_{\mathrm{cl}(\tau)})^m$ would be non-decomposable. Conversely, if $(G_{\mathrm{cl}(\tau)})^m$ is non-decomposable then $(G_{\mathrm{cl}(\tau)})^m$ contains a chordless n-cycle, $C, n \geqslant 4$. If $C \subseteq \tau$, then (i) fails. If $C \nsubseteq \tau$, proceed as in the proof of the implication (2) $\Rightarrow$ (2′) in proposition 4.2 to conclude that either (ii) or (iii) must fail (cf. Fig. 7).

Remark 4.3. A chain graph model with strictly positive joint densities (or limits of such distributions) is called a *decomposable block recursive model* by Højsgaard & Thiesson (1995), if it satisfies condition (2′). They note that such a model admits a recursive factorization of the form (3.1), and hence is Markov equivalent to some ADG.

Proposition 4.3

Let G be a chain graph. The following are equivalent:

(3) *G is Markov equivalent to some (necessarily unique) decomposable UDG.*

(3′) *G has no minimal complexes and G^u is decomposable.*

(3″) *$G \overset{\mathrm{M}}{\sim} G^u$ and G^u is decomposable.*

Proof. This follows directly from proposition 4.1.

By combining propositions 4.1, 4.2, and 4.3, we can demonstrate in complete generality that the intersection of the classes of graphical Markov models determined by all UDGs and by all ADGs respectively, is the class of graphical Markov models determined by all decomposable UDGs (corollary 4.1). This will follow directly from proposition 4.4, whose proof is purely graph-theoretic.

Proposition 4.4

Let G be a chain graph. Then

(1′) *G has no minimal complexes, and*

(2′) *$(G_{\mathrm{cl}(\tau)})^m$ is decomposable for all chain components $\tau \in \mathscr{T}(G)$,*

if and only if

(3′) *G has no minimal complexes and G^u is decomposable.*

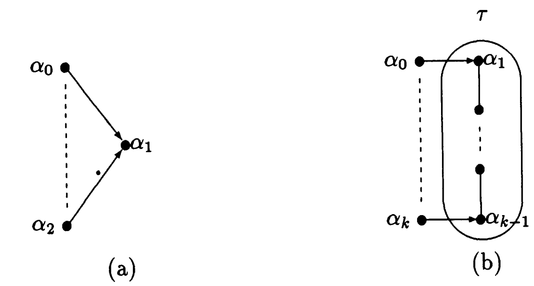

Fig. 8. The induced subgraph G_C, where C is a chordless n-cycle in G^u.

Proof. $(3') \Rightarrow (1')$ and $(2')$: It suffices to show that G satisfies $(2')$. Because G has no minimal complexes, $(G_{\text{cl}(\tau)})^m = (G_{\text{cl}(\tau)})^u = (G^u)_{\text{cl}(\tau)}$ for every $\tau \in \mathscr{T}(G)$. However, $(G^u)_{\text{cl}(\tau)}$ is decomposable, since every induced subgraph of a decomposable UDG is decomposable.

$(1')$ and $(2') \Rightarrow (3')$: It suffices to show that G^u is decomposable. If not, then G^u contains a chordless n-cycle, $C \equiv \{\alpha_0, \alpha_1, \ldots, \alpha_n \equiv \alpha_0\}$, $n \geqslant 4$. By $(2')$, C cannot lie entirely within any chain component of G, hence G_C must have at least one directed edge. Since G is a chain graph, G_C therefore must have at least *two* directed edges, say (α_0, α_1) and (α_k, α_{k-1}), where $2 \leqslant k \leqslant n$, with opposing orientations in G_C. Furthermore, we may select these two edges such that either $\alpha_1 = \alpha_{k-1}$ (hence $k = 2$—see Fig. 8(a)) or else $\{\alpha_1, \ldots, \alpha_{k-1}\}$ forms an *undirected* path in G_C (hence $3 \leqslant k \leqslant n$—see Fig. 8(b)). In the first case, $(\alpha_0, \alpha_1, \alpha_2)$ forms an immorality in G (since C is chordless), violating $(1')$. In the second case, let $\tau \in \mathscr{T}(G)$ be the unique chain component of G such that $\{\alpha_1, \ldots, \alpha_{k-1}\} \in \tau$. If $k = n - 1$ or n, then $\alpha_0 \in \text{bd}(\tau)$ and $\alpha_k \in \text{bd}(\tau)$, so C is a chordless cycle in $(G^u)_{\text{cl}(\tau)} = (G_{\text{cl}(\tau)})^u$, violating $(2')$. If $3 \leqslant k \leqslant n-2$ (in which case $n \geqslant 5$), then $\alpha_0 \in \text{bd}(\tau)$ and $\alpha_k \in \text{bd}(\tau)$, but α_0 and α_k are not adjacent in G. But then α_0 and α_k *are* adjacent in $(G_{\text{cl}(\tau)})^m$, hence $\{\alpha_0, \alpha_1, \ldots, \alpha_{k+1} \equiv \alpha_0\}$ forms a chordless $(k+1)$-cycle in $(G_{\text{cl}(\tau)})^m$, again violating $(2')$. This completes the proof.

Corollary 4.1
Let G be a chain graph. Then
(1) *G is Markov equivalent to some UDG, and*
(2) *G is Markov equivalent to some ADG,*
if and only if
(3) *G is Markov equivalent to some decomposable UDG (namely, G^u).*

Proof. This follows immediately from propositions 4.1–4.4.

It is easily seen from the proof of the implication $(1')$ and $(2') \Rightarrow (3')$ that proposition 4.4 remains true if $(1')$ is replaced by the weaker condition $(1''')$, as follows:

Proposition 4.5
Let G be a chain graph. Then
$(1''')$ *G has no immoralities, and*
$(2')$ $(G_{\text{cl}(\tau)})^m$ *is decomposable for all chain components* $\tau \in \mathscr{T}(G)$,
if and only if
$(3')$ *G has no minimal complexes and G^u is decomposable.*

G_7 G_8

Fig. 9. Two chain graphs with no minimal complexes other than immoralities, yet which are not Markov equivalent to any UDG or ADG.

Note, however, that propositions 4.4 and 4.5 are not true if (3′) is replaced by the following weaker condition:

(3‴) *G has no immoralities and G^u is decomposable.*

The chain graph G_5 in example 4.1 provides a simple counter-example.

Propositions 4.4 and 4.5 give equivalent graphical characterizations of those chain graphs that are Markov equivalent simultaneously to some UDG and to some ADG. It is natural, then, to seek graphical characterizations of those chain graphs that are *not* Markov equivalent to any UDG or to any ADG (the shaded region in Fig. 2).

It follows immediately from theorem 3.1 that if G has at least one minimal complex that is not an immorality, then G cannot be Markov equivalent to any UDG or to any ADG: for example, G_3 in Fig. 5. However, the condition is not necessary: the chain graphs G_7 and G_8 in Fig. 9 have no minimal complexes other than immoralities, yet by propositions 4.1 and 4.2, are not Markov equivalent to any UDG or ADG. The final result of this section, proposition 4.6, presents three graphical conditions, (4′)–(4‴), each of which is both necessary and sufficient for a chain graph not to be Markov equivalent to any UDG or ADG.

Proposition 4.6

Let G be a chain graph. The following are equivalent:

(4) *G is not Markov equivalent to any UDG or to any ADG.*

(4′) *G has at least one minimal complex, and $(G_{\mathrm{cl}(\tau)})^m$ is non-decomposable for at least one chain component $\tau \in \mathscr{T}(G)$.*

(4″) (*a*) *G has at least one minimal complex that is not an immorality; or*

(*b*) *G has at least one immorality, and $(G_{\mathrm{cl}(\tau)})^m$ is non-decomposable for at least one chain component $\tau \in \mathscr{T}(G)$.*

(4‴) (*a*) *G has at least one minimal complex that is not an immorality; or*

(*b*) *G has at least one immorality, and, for at least one chain component $\tau \in \mathscr{T}(G)$, either*

(*i*) *G_τ is non-decomposalbe; or*

(*ii*) *There exists an $\alpha \in \mathrm{bd}(\tau)$ and there exists a non-adjacent pair $\gamma, \delta \in \mathrm{ch}_\tau(\alpha)$, such that there is a path π in τ from γ to δ with $[\pi\backslash\{\gamma, \delta\}] \cap \mathrm{ch}_\tau(\alpha) = \emptyset$; or*

(*iii*) *There exists a distinct pair $\alpha, \beta \in \mathrm{bd}(\tau)$ and there exists a pair $\gamma \in \mathrm{ch}_\tau(\alpha)\backslash\mathrm{ch}_\tau(\beta)$, $\delta \in \mathrm{ch}_\tau(\beta)\backslash\mathrm{ch}_\tau(\alpha)$, such that there exists a path π in τ from γ to δ with $[\pi\backslash\{\gamma, \delta\}] \cap [\mathrm{ch}_\tau(\alpha) \cup \mathrm{ch}_\tau(\beta)] = \emptyset$.*

Proof. This follows immediately from propositions 4.1 and 4.2 and remark 4.2.

Acknowledgements

Research supported in part by the US National Science Foundation and the US National Security Agency.

References

Andersson, S. A., Madigan, D. & Perlman, M. D. (1996). A characterization of Markov equivalence classes of acyclic digraphs. To appear in *Ann. Statist.*

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

Blair, J. R. S. & Peyton, B. (1993). An introduction to chordal graphs and clique trees. In *Graph theory and sparse matrix computations* (eds A. George, J. R. Gilbert & J. W. H. Liu), 1–29. Springer, New York.

Cox, D. R. & Wermuth, N. (1993). Linear dependencies represented by chain graphs. *Statist. Sci.* **8,** 204–218; 247–277.

Dawid, A. P. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8,** 598–617.

Dawid, A. P. & Lauritzen, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21,** 1272–1317.

Edwards, D. (1995). *Introduction to graphical modelling.* Springer, New York.

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

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

Højsgaard, S. & Thiesson, B. (1995). BIFROST—Block recursive models induced from relevant knowledge, observations, and statistical techniques. *Computat. Statist. Data Anal.* **19,** 155–175.

Kiiveri, H., Speed, T. P. & Carlin, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36,** 30–52.

Lauritzen, S. L. (1989). Mixed graphical association models (with discussion). *Scand. J. Statist.* **16,** 273–306.

Lauritzen, S. L. (1996). *Graphical association models.* Oxford University Press, Oxford.

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

Lauritzen, S. L., Speed, T. P. & Vijayan, K. (1984). Decomposable graphs and hypergraphs. *J. Austral. Math. Soc. Ser. A* **36,** 12–29.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N. & Leimer, H.-G. (1990). Independence properties of directed Markov fields. *Networks* **20,** 491–505.

Madigan, D. & Mosurski, K. R. (1990). An extension of the results of Asmussen and Edwards on collapsibility in contingency tables. *Biometrika* **77,** 315–319.

Madigan, D. & 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.

Matúš, F. (1992). On equivalence of Markov properties over undirected graphs. *J. Appl. Probab.* **29,** 745–749.

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.

Spiegelhalter, D. J. & Lauritzen, S. L. (1990). Sequential updating of conditional propabilities on directed graphical structures. *Networks* **20,** 579–605.

Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S. L. & Cowell, R. G. (1993). Bayesian analysis in expert systems. *Statist. Sci.* **8,** 219–283.

Tarjan, R. E. & Yannakakis, M. (1984). Simple linear time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce hypergraphs. *SIAM J. Comput.* **13,** 566–579.

Wermuth, N. (1980). Linear recursive equations, covariance selection and path analysis. *J. Amer. Statist. Assoc.* **75,** 963–972.

Wermuth, N. & Lauritzen, S. L. (1983). Graphical and recursive models for contingency tables. *Biometrika* **70,** 537–552.

Wermuth, N. & Lauritzen, S. L. (1990). On substantive research hypotheses, conditional independence graphs and graphical chain models. *J. Roy. Statist. Soc. Ser. B* **52,** 21–50.
Whittaker, J. (1990). *Graphical models in applied multivariate statistics.* Wiley, Chichester.
Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20,** 557–585.
York, J., Madigan, D., Heuch, I. & Lie, R. T. (1995). Estimating a proportion of birth defects by double sampling: a Bayesian approach incorporating covariates and model uncertainty. *Appl. Statist.* **94,** 227–242.

Received January 1995, in final form January 1996

M. D. Perlman, Department of Statistics, Box 354322, University of Washington, Seattle WA 98195, USA.

Appendix. Proof of theorem 3.1

Theorem 3.1 extends the Markov equivalence theorem of Frydenberg (1990, th. 5.6) to general probability measures. Frydenberg's restriction to measures that satisfy CI5 can be removed, provided that Markov equivalence is defined in terms of the *global* Markov property. The outline of our proof will closely follow Frydenberg (1990). First, however, we must extend to general UDGs a result obtained by Dawid & Lauritzen (1993, prop. B.7) for the special case of decomposable graphs.

Lemma A.1
Let (C, D) be a decomposition of an undirected graph $G = (V, E)$ and let P be a probability measure on $\mathscr{X} \equiv \times_{\alpha \in V} \mathscr{X}_\alpha$. Then P is global G-Markovian if and only if
(*i*) *P_C is global G_C-Markovian,*
(*ii*) *P_D is global G_D-Markovian, and*
(*iii*) *$C \perp D \mid C \cap D[P]$.*

Proof. The result is easily verified if $C = V$ or $D = V$, so we may assume that the decomposition is proper.

First assume that P is global G-Markovian. In order to establish (i), it suffices to show that if S separates A and B in G_C, then S separates A and B in G. If the latter fails, there exists a path $\pi = \{\alpha_0, \ldots, \alpha_n\} \subseteq V \equiv C \cup D$ such that $\alpha_0 \in A$, $\alpha_n \in B$, and $\pi \cap S = \emptyset$. By assumption, $\pi \nsubseteq C$, hence $\alpha_i \in D \backslash C$ for some i, $1 \leqslant i \leqslant n-1$ (see Fig. 10). Since $C \cap D$ separates $C \backslash D$ and $D \backslash C$, there exist integers j and k such that $0 \leqslant j \leqslant i-1$, $i+1 \leqslant k \leqslant n$, and $\alpha_j, \alpha_k \in C \cap D$. Because $C \cap D$ is complete, α_j and α_k must be adjacent in G, hence $\pi' \equiv \{\alpha_0, \ldots, \alpha_j, \alpha_k, \ldots, \alpha_n\}$ is a path in G_C between A and B that does not intersect S, contradicting the assumption that S separates A and B in G_C. Thus (i) must hold, and similarly (ii). Lastly, (iii) holds by the definition of a decomposition.

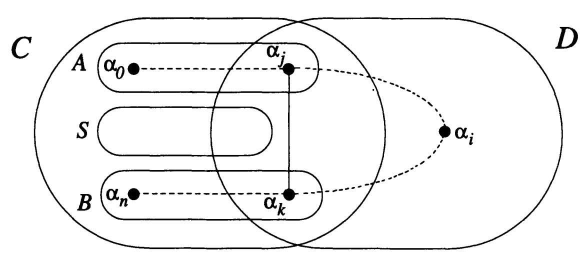

Fig. 10. The path π (dotted line) between A and B in $V \equiv C \cup D$.

Conversely, suppose that P satisfies (i), (ii), and (iii). We must show that

$$A \perp B \mid S[P] \quad \text{whenever } S \text{ separates } A \text{ and } B \text{ in } G. \tag{A.1}$$

Partition $A = A_1 \cup A_2 \cup A_3$, $B = B_1 \cup B_2 \cup B_3$, $S = S_1 \cup S_2 \cup S_3$, where $A_1 = A \cap (C \backslash D)$, $A_2 = A \cap (C \cap D)$, $A_3 = A \cap (D \backslash C)$, etc. Note that either $A_2 = \emptyset$ or $B_2 = \emptyset$ (or both): otherwise, since $C \cap D$ is complete, A and B could not be separated in G. Without loss of generality, assume that $A_2 = \emptyset$. Define $E = (C \cap D) \backslash (B \cup S) \equiv (C \cap D) \backslash (B_2 \cup S_2)$, so $B_2 \cup S_2 \cup E = C \cap D$ (see Fig. 11). By (iii) and CI2,

$$(A_1 \cup B_1 \cup S_1) \perp (B_3 \cup S_3) \mid B_2 \cup S_2 \cup E[P]. \tag{A.2}$$

We shall establish (A.1) by considering a series of six cases.

Case 1. $A_3 = B_2 = B_3 = \emptyset$, so $A = A_1$ and $B = B_1$. Here (A.2) takes the form:

$$(A_1 \cup B_1 \cup S_1) \perp S_3 \mid S_2 \cup E[P]. \tag{A.3}$$

By assumption, $S \equiv S_1 \cup S_2 \cup S_3$ separates A_1 and B_1 in G, hence $S_1 \cup S_2$ separates A_1 and B_1 in G_C. In addition, either (a) $S_1 \cup S_2$ separates A_1 and E in G_C, or (b) $S_1 \cup S_2$ separates B_1 and E in G_C (or both). If (a) holds, it follows from (i) that

$$A_1 \perp (B_1 \cup E) \mid S_1 \cup S_2[P]. \tag{A.4}$$

Also, by (A.3) and CI3,

$$A_1 \perp S_3 \mid B_1 \cup S_1 \cup S_2 \cup E[P]. \tag{A.5}$$

By CI4, (A.4) and (A.5) combine to yield

$$A_1 \perp (B_1 \cup S_3 \cup E) \mid S_1 \cup S_2[P], \tag{A.6}$$

which, by CI2 and CI3, implies that

$$A_1 \perp B_1 \mid S_1 \cup S_1 \cup S_3[P]; \tag{A.7}$$

this is (A.1) in case 1. If (b) holds, we again obtain (A.7) by reversing the roles of A_1 and B_1.

Case 2. $A_3 = B_1 = B_3 = \emptyset$, so $A = A_1$ and $B = B_2$. Here (A.2) takes the form:

$$(A_1 \cup S_1) \perp S_3 \mid B_2 \cup S_2 \cup E[P]. \tag{A.8}$$

By assumption, $S \equiv S_1 \cup S_2 \cup S_3$ separates A_1 and B_2 in G, hence $S_1 \cup S_2$ separates A_1 and B_2 in G_C. Furthermore, since $C \cap D$ is complete, $S_1 \cup S_2$ separates A_1 and E in G_C. It follows from (i) that

$$A_1 \perp (B_2 \cup E) \mid S_1 \cup S_2[P]. \tag{A.9}$$

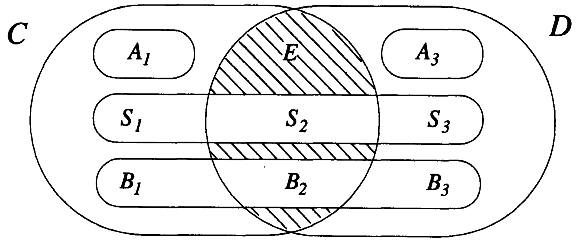

Fig. 11. The partitions of A, B, and S when $A_2 = \emptyset$; E (shaded) $= (C \cap D) \backslash (B \cup S)$.

Also, by (A.8) and CI3,

$$A_1 \perp S_3 \mid B_2 \cup S_1 \cup S_2 \cup E[P], \tag{A.10}$$

which combines with (A.9) to yield

$$A_1 \perp (B_2 \cup S_3 \cup E) \mid S_1 \cup S_2[P]. \tag{A.11}$$

Therefore, by CI2 and CI3, we obtain

$$A_1 \perp B_2 \mid S_1 \cup S_2 \cup S_3[P], \tag{A.12}$$

which is (A.1) in case 2.

Case 3. $A_3 = B_1 = B_2 = \emptyset$, so $S = A_1$ and $B = B_3$. Now (A.2) becomes

$$(A_1 \cup S_1) \perp (B_3 \cup S_3) \mid S_2 \cup E[P]. \tag{A.13}$$

By assumption, $S \equiv S_1 \cup S_2 \cup S_3$ separates A_1 and B_3 in G. Also, either (a) $S_1 \cup S_2$ separates A_1 and E in G_C, or (b) $S_2 \cup S_3$ separates B_3 and E in G_D (or both). If (a) holds, it follows from (i) that

$$A_1 \perp E \mid S_1 \cup S_2[P]. \tag{A.14}$$

By (A.13) and CI3,

$$A_1 \perp (B_3 \cup S_3) \mid S_1 \cup S_2 \cup E[P]. \tag{A.15}$$

But (A.14) and (A.15) combine by CI4 to yield

$$A_1 \perp (B_3 \cup S_3 \cup E) \mid S_1 \cup S_2[P]. \tag{A.16}$$

Thus, by CI2 and CI3, we obtain

$$A_1 \perp B_3 \mid S_1 \cup S_2 \cup S_3[P], \tag{A.17}$$

which is (A.1) in case 3. If (b) holds, (A.17) is obtained by reversing the roles of (A_1, C) and (B_3, D) and replacing (i) by (ii).

Case 4. $A_3 = \emptyset$ (so $A = A_1$) and B arbitrary (so $B = B_1 \cup B_2 \cup B_3$). We will establish (A.1) by combining cases 1, 2, and 3. By assumption, S separates A_1 and $B_1 \cup B_2 \cup B_3$ in G, hence $S' \equiv B_2 \cup S$ separates A_1 and B_1 in G. Thus, by case 1,

$$A_1 \perp B_1 \mid B_2 \cup S[P]. \tag{A.18}$$

Also, S separates A_1 and B_2 in G, so case 2 applies to yield

$$A_1 \perp B_2 \mid S[P]. \tag{A.19}$$

Then (A.18) and (A.19) combine by CI4 to give

$$A_1 \perp (B_1 \cup B_2) \mid S[P]. \tag{A.20}$$

Next, since $B_3 \cup S$ also separates A_1 and $B_1 \cup B_2$ in G, (A.20) yields

$$A_1 \perp (B_1 \cup B_2) \mid B_3 \cup S[P]. \tag{A.21}$$

However, by case 3,

$$A_1 \perp B_3 \mid S[P], \tag{A.22}$$

which combines with (A.21) via CI4 to yield

$$A_1 \perp (B_1 \cup B_2 \cup B_3) \mid S[P]; \tag{A.23}$$

this is (A.1) in case 4.

Case 5. $A_1 = \emptyset$ (so $A = A_3$) and B arbitrary. This follows from case 4 by reversing the roles of C and D.

Case 6 (the general case). $A = A_1 \cup A_3$ and B arbitrary. By assumption, S separates $A_1 \cup A_3$ and B in G, hence $S' \equiv A_3 \cup S$ separates A_1 and B in G. Thus by case 4,

$$A_1 \perp B \mid A_3 \cup S[P]. \tag{A.24}$$

Next, by case 5,

$$A_3 \perp B \mid S[P]. \tag{A.25}$$

Then, (A.24) and (A.25) combine via CI4 to yield

$$(A_1 \cup A_3) \perp B \mid S[P], \tag{A.26}$$

which is (A.1) in case 6. This completes the proof of lemma A.1.

Lemma A.2 (cf. Frydenberg, 1990, prop. 2.2)
Let $G \equiv (V, E)$ be an undirected graph, P a probability measure on $\mathcal{X}$, and C a proper subset of V. If C is simplicial in G then the following are equivalent:

(*i*) *P is global G-Markovian*;
(*ii*) *$P_{V\setminus C}$ is global $G_{V\setminus C}$-Markovian, $P_{\mathrm{cl}(C)}$ is global $G_{\mathrm{cl}(C)}$-Markovian, and* $(V\setminus \mathrm{cl}(C)) \perp C \mid \mathrm{bd}(C)[P]$.

Proof. This follows from lemma A.1, since $(V\setminus C, \mathrm{cl}(C))$ forms a decomposition of G with $(V\setminus C) \cap \mathrm{cl}(C) = \mathrm{bd}(C)$.

Lemma A.3 (cf. Frydenberg, 1990, lem. 3.1)
If τ is a terminal chain component in a chain graph $G \equiv (V, E)$ and P is global G-Markovian, then $P_{V\setminus\tau}$ is global $G_{V\setminus\tau}$-Markovian.

Proof. We must show that $A \perp B \mid S[P]$, if A, B, and S are subsets of $V\setminus\tau$ such that S separates A from B in $((G_{V\setminus\tau})_{\mathrm{an}_{V\setminus\tau}(A \cup B \cup S)})^m \equiv (G_{\mathrm{an}_{V\setminus\tau}(A \cup B \cup S)})^m$. Because $V\setminus\tau$ is a G-anterior set, however, $\mathrm{an}_{V\setminus\tau}(A \cup B \cup S) = \mathrm{an}(A \cup B \cup S)$, so S separates A from B in $(G_{\mathrm{an}(A \cup B \cup S)})^m$. Since P is global G-Markovian, $A \perp B \mid S$.

Lemma A.4 (cf. Frydenberg, 1990, lem. 3.2)
If $G \equiv (V, E)$ is a chain graph and P is global G-Markovian, then P is global G^m-Markovian.

Proof. Since G^m is an undirected graph, it suffices to show that $A \perp B \mid S[P]$ whenever S separates A from B in G^m. However, $(G_{\mathrm{an}(A \cup B \cup S)})^m \subseteq G^m$, so S also separates A from B in $(G_{\mathrm{an}(A \cup B \cup S)})^m$ and the result follows.

Lemma A.5 (cf. Frydenberg, 1990, th. 3.3)
Let $G \equiv (V, E)$ be a chain graph and P a probability measure on $\mathcal{X}$. Then the following are equivalent:

(*i*) *P is global G-Markovian*;
(*ii*) P_A *is global* $(G_A)^m$*-Markovian for all G-anterior subsets* $A \subseteq V$.

Proof. To prove that (i) $\Rightarrow$ (ii), suppose that P is global G-Markovian and A is a G-anterior set. Because A is obtained from V by stepwise removal of terminal chain components, it follows from lemma A.3 that P_A is global G_A-Markovian. By lemma A.4, P_A is global $(G_A)^m$-Markovian. The implication (ii) $\Rightarrow$ (i) is immediate from the definition of the global Markov property and its restatement for undirected graphs immediately following definition 3.1.

Proposition A.1 (cf. Frydenberg, 1990, coroll. 3.4)
Let $G \equiv (V, E)$ *be a chain graph,* τ *a terminal chain component, and P a probability measure on* $\mathscr{X}$. *The following are equivalent*:
(*i*) *P is global G-Markovian*;
(*ii*) $P_{V\setminus\tau}$ *is global* $G_{V\setminus\tau}$*-Markovian and P is* G^m*-Markovian*;
(*iii*) $P_{V\setminus\tau}$ *is global* $G_{V\setminus\tau}$*-Markovian,* $P_{\mathrm{cl}(\tau)}$ *is global* $(G_{\mathrm{cl}(\tau)})^m$*-Markovian, and* $\tau \perp (V\setminus\mathrm{cl}(\tau)) \mid \mathrm{bd}(\tau)[P]$.

Proof. That (i) implies (ii) follows from lemmas A.3 and A.4. That (ii) implies (iii) follows from lemma A.2 (applied to G^m) and the facts that τ is simplicial in G^m and that $(G^m)_{\mathrm{cl}(\tau)} = (G_{\mathrm{cl}(\tau)})^m$.

Assume now that (iii) holds. To verify (i), by lemma A.5 it suffices to show that P_A is global $(G_A)^m$-Markovian for all G-anterior sets $A \subseteq V$. If $A \subseteq V\setminus\tau$, then A is $G_{V\setminus\tau}$-anterior, so by lemma A.5 applied to $G_{V\setminus\tau}$, P_A is global $(G_A)^m$-Markovian. If $A \nsubseteq V\setminus\tau$, then $\tau \subseteq A$. We shall apply lemma A.2 with (G, C, P) replaced by $((G_A)^m, \tau, P_A)$ to conclude that P_A is global $(G_A)^m$-Markovian. In order that lemma A.2 be applicable, we must verify the following four facts:

(a) τ is simplicial in $(G_A)^m$: since τ is terminal in G, it is terminal in G_A. Therefore the boundary of τ with respect to $(G_A)^m$ is $\mathrm{bd}_A(\tau)$, which is complete in $(G_A)^m$.
(b) $P_{A\setminus\tau}$ is global $((G_A)^m)_{A\setminus\tau}$-Markovian: it is straightforward to verify that $A\setminus\tau$ is a $G_{V\setminus\tau}$-anterior subset of $V\setminus\tau$, so by lemma A.5 applied to $G_{V\setminus\tau}$ and $P_{V\setminus\tau}$, $P_{A\setminus\tau}$ is global $(G_{A\setminus\tau})^m$-Markovian. Since $((G_A)^m)_{A\setminus\tau} \supseteq ((G_A)_{A\setminus\tau})^m = (G_{A\setminus\tau})^m$, (b) holds.
(c) $P_{\mathrm{cl}_A(\tau)}$ is global $((G_A)^m)_{\mathrm{cl}_A(\tau)}$-Markovian: since A is G-anterior, $\mathrm{cl}_A(\tau) = \mathrm{cl}(\tau)$. Also, $((G_A)^m)_{\mathrm{cl}(\tau)} = ((G_A)_{\mathrm{cl}(\tau)})^m = (G_{\mathrm{cl}(\tau)})^m$, so (c) follows from the second relation in (iii).
(d) $\tau \perp (A\setminus\mathrm{cl}_A(\tau)) \mid \mathrm{bd}_A(\tau)[P]$: This follows from the third relation in (iii) and CI2, since $\mathrm{bd}_A(\tau) = \mathrm{bd}(\tau)$ and $\mathrm{cl}_A(\tau) = \mathrm{cl}(\tau)$.

Lemma A.6 (cf. Frydenberg, 1990, lem. 5.1)
Let $\tilde{\tau}$ *be a terminal chain component of a chain graph* $\tilde{G} \equiv (V, \tilde{E})$ *and let* τ *be a connected subset of* $\tilde{\tau}$. *Let* $G \equiv (V, E)$ *be the chain graph which differs from* $\tilde{G}$ *only in that all edges in* $\tilde{G}$ *between* $\tilde{\tau}\setminus\tau$ *and* τ *are changed into arrows towards* τ *in G. If* τ *is simplicial in* $\tilde{G}^m$, *or equivalently, if* $(\tilde{G}_{\mathrm{cl}(\tau)})^m = (G_{\mathrm{cl}(\tau)})^m$, *then G and* $\tilde{G}$ *are Markov equivalent.*

Proof. First recall that $\mathrm{bd}(\tau)$ and $\mathrm{cl}(\tau)$ are the same relative to G and to $\tilde{G}$. Then the proof of the result is obtained from that of lem. 5.1 of Frydenberg (1990) with the following minor changes: replace "Markovian" by "global Markovian", "coroll. 3.4" by "proposition A.1 (the equivalence of (i) and (ii))", and "prop. 2.2" by "lemma A.2". [In the fourth line of Frydenberg's proof, "$\tilde{G}$" should be "$\tilde{G}^m$".]

Proposition A.2 (cf. Frydenberg, 1990, prop. 5.2)
Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$ and $G \subseteq \tilde{G}$, i.e. $\tilde{G}$ might have lines where G has arrows. If each chain component $\tau \in \mathscr{T}(G)$ is simplicial in $(\tilde{G}_{\mathrm{cl}(\tau)})^m$, or equivalently, if $(\tilde{G}_{\mathrm{cl}(\tau)})^m = (G_{\mathrm{cl}(\tau)})^m$, then G and $\tilde{G}$ are Markov equivalent.

Proof. The proof of this result follows that of prop. 5.2 of Frydenberg (1990) if "Markovian" is replaced by "global Markovian", "lem. 5.1" by "lemma A.6", and "coroll. 3.4" by "proposition A.1 (the equivalence of (i) and (iii))". [In the eighth line of Frydenberg's proof, "P" should be "$P_{W(t)}$"; in the ninth line, "$P_{\mathrm{cl}(\tau_t)}$" should be "$P_{\overline{\mathrm{cl}}(\tau_t)}$".]

Lemma A.7 (cf. Frydenberg, 1990, p. 346)
Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$ and $G \subseteq \tilde{G}$. Then G and $\tilde{G}$ have the same minimal complexes if and only if any minimal complex in G is a minimal complex in $\tilde{G}$.

Proof. "only if" is trivial. "if": let (α, B, β) be a minimal complex in $\tilde{G}$. Note that $G_{\{\alpha, \beta\} \cup B}$ and $\tilde{G}_{\{\alpha, \beta\} \cup B}$ may differ only in that some lines in $\tilde{G}_B$ might occur as arrows in G_B. But any occurrence of arrows in G_B would create at least one minimal complex in $G_{\{\alpha, \beta\} \cup B}$ which, by hypothesis, must also exist in $\tilde{G}_{\{\alpha, \beta\} \cup B}$, violating the fact that $\tilde{G}_B$ has only lines, not arrows. Thus, $G_{\{\alpha, \beta\} \cup B}$ and $\tilde{G}_{\{\alpha, \beta\} \cup B}$ are identical, hence $\{\alpha, B, \beta\}$ is also a minimal complex in $G_{\{\alpha, \beta\} \cup B}$.

Lemma A.8 ($\equiv$prop. 5.3 of Frydenberg, 1990)
Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$ and $G \subseteq \tilde{G}$. The following are equivalent:

(*i*) *Each chain component $\tau \in \mathscr{T}(G)$ is simplicial in $(\tilde{G}_{\mathrm{cl}(\tau)})^m$;*
(*ii*) *G and $\tilde{G}$ have the same minimal complexes.*

Proof. Apply lemma A.7—see Frydenberg (1990, pp. 346–347).

Proposition A.3 (cf. Frydenberg, 1990, p. 347)
Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$ and $G \subseteq \tilde{G}$. If G and $\tilde{G}$ have the same minimal complexes, then G and $\tilde{G}$ are Markov equivalent.

Proof. This follows immediately from proposition A.2 and lemma A.8.

If $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ are two chain graphs such that G and $\tilde{G}$ have the same vertex set, define $G \cup \tilde{G} = (V, E \cup \tilde{E})$. Let $G \vee \tilde{G}$ be the graph obtained from $G \cup \tilde{G}$ by changing into a line any arrow that is part of a directed cycle in $G \cup \tilde{G}$. Frydenberg (1990, p. 347) notes that $G \vee \tilde{G}$ is a chain graph; clearly $G \vee \tilde{G}$ is the (unique) smallest chain graph larger than both G and $\tilde{G}$.

Proposition A.4 ($\equiv$prop. 5.4 of Frydenberg, 1990)
Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be two chain graphs such that $G^u = \tilde{G}^u$. Then G and $\tilde{G}$ have the same minimal complexes if and only if they both have the same minimal complexes as $G \vee \tilde{G}$.

Our main result is now at hand:

Theorem 3.1 (cf. Frydenberg, 1990, th. 5.6)
Two chain graphs $G \equiv (V, E)$ and $G \equiv (V, \tilde{E})$ are Markov equivalent if and only if $G^u = \tilde{G}^u$ and G and $\tilde{G}$ have the same minimal complexes.

Proof. Since $G \vee \tilde{G}$ is larger than both G and $\tilde{G}$, the "if" assertion is a direct consequence of propositions A.3 and A.4. The "only if" assertion is proved by Frydenberg (1990, pp. 347–348).

Corollary A.1 (cf. Frydenberg, 1990, prop. 5.7)
For any chain graph G, there exists a unique largest chain graph that is Markov equivalent to G.

Proof. This follows from proposition A.4 and theorem 3.1—see Frydenberg.