ALTERNATIVE MARKOV PROPERTIES FOR CHAIN GRAPHS*

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

Indiana University and University of Washington

15 July 1999

Abstract

Graphical Markov models use graphs, either undirected, directed, or mixed, to represent possible dependences among statistical variables. Applications of undirected graphs (UGs) include models for spatial dependence and image analysis, while acyclic directed graphs (ADGs), which are especially convenient for statistical analysis, arise in such fields as genetics and psychometrics and as models for expert systems and Bayesian belief networks. Lauritzen, Wermuth, and Frydenberg (LWF) introduced a Markov property for chain graphs (CG): mixed graphs that can be used to represent simultaneously both structural and associative dependencies and that include both UGs and ADGs as special cases. Cox and Wermuth (CW) introduced block regression, multivariate regression, and concentration regression models for CGs.

In this paper an alternative Markov property (AMP) for CGs is introduced and shown to be the Markov property satisfied by a CW concentration regression model with multivariate normal errors. This model can be decomposed into a collection of conditional normal models, each of which combines the features of multivariate linear regression models and covariance selection models, facilitating the estimation of its parameters. In the general case, necessary and sufficient conditions are given for the equivalence of the LWF and AMP Markov properties of a CG, for the AMP Markov equivalence of two CGs, for the AMP Markov equivalence of a CG to some ADG or decomposable UG, and for other equivalences. For CGs, in some ways the AMP property is a more direct extension of the ADG Markov property than is the LWF property.

1. Introduction.

Graphical Markov models (GMM) use graphs, either undirected, directed, or mixed, to represent possible dependences among the variables of a multivariate probability distribution. The vertices of the graph represent the variables, while the presence (absence) of an edge between two vertices indicates possible dependence (independence) between the two corresponding variables. A GMM is constructed by specifying local dependencies for each vertex in terms of its immediate neighbors, parents or both, yet can represent a complex system of dependencies by means of the global structure of the graph.

Applications of undirected graphs (UGs) include models for spatial dependence and image analysis, while acyclic directed graphs (ADGs)[1] occur in genetics, psychometrics, expert systems, Bayesian belief networks, and many other fields.

---

* This research was supported in part by the U. S. National Science Foundation.

† *Address for correspondence*: Department of Statistics, Box 35-4322, University of Washington, Seattle, WA 98195-4322, USA. *e-mail*: michael@ms.washington.edu.

1 The phrase "directed acyclic graph (DAG)" is more commonly used.

GMMs determined by ADGs admit especially elegant statistical analysis. The likelihood function associated with an ADG Markov model has a convenient recursive factorization which, for categorical or multivariate normal data, yields explicit maximum likelihood estimates and likelihood ratio tests - cf. Lauritzen *et al* (1990), Whittaker (1990), Edwards (1995), Lauritzen (1996), Andersson and Perlman (1998). ADG models allow efficient computational algorithms for exact probability calculations, as well as efficient updating algorithms for Bayesian analysis - cf. Pearl (1988), Lauritzen and Spiegelhalter (1988), Shachter and Kenley (1989), Spiegelhalter *et al* (1993). The only UG models with these properties are the decomposable models, exactly those UG models that are Markov equivalent to some ADG model (Wermuth and Lauritzen (1983), Dawid and Lauritzen (1993), Andersson, Madigan, and Perlman [AMP] (1997a)).

Lauritzen and Wermuth (1989) and Frydenberg (1990) (LWF) introduced a Markov property for adicyclic graphs ≡ "chain graphs"(CG) that generalizes the classical Markov properties for both UGs and ADGs. CGs have both directed and undirected edges and contain no (partially) directed cycles - CG models attempt to simultaneously represent dependencies some of which are structural and some associative. Wermuth and Lauritzen (1990), Cox and Wermuth [CW] (1993, 1996), Højsgaard and Thiesson (1995), and Lauritzen (1996) describe statistical applications of CGs, while Buntine (1995) discusses their usefulness for modeling belief networks.

It has been noted recently that a CG may admit alternative Markov interpretations, hence may simultaneously represent different statistical models ≡ belief networks ([CW] (1993, 1996), [AMP] (1996)). These competing Markov interpretations and the assessment of their relative applicability have stimulated substantial interest. In this paper we describe an alternative Markov property (AMP) for CGs that in some ways is a more direct generalization of the ADG Markov property than is the Lauritzen-Wermuth-Frydenberg (LWF) Markov property for CGs.

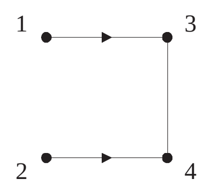


Figure 1.1. A simple chain graph G for which the LWF and AMP properties differ.

To motivate the AMP Markov property, consider the CG in Figure 1.1, which represents a set of conditional independences (CI) satisfied by random variables X_1, X_2, X_3, X_4 $\equiv 1, 2, 3, 4$. The LWF Markov property for G reduces to the CIs

$$1 \perp\!\!\!\perp 4 \mid 2, 3, \quad 2 \perp\!\!\!\perp 3 \mid 1, 4, \quad 1 \perp\!\!\!\perp 2, \tag{1.1}$$

whereas the AMP can be expressed in terms of the CIs

$$1 \perp\!\!\!\perp 4 \mid 2, \quad 2 \perp\!\!\!\perp 3 \mid 1, \quad 1 \perp\!\!\!\perp 2. \tag{1.2}$$

Although both interpretations of the CG G may be useful for modeling ([CW] (1993, p.206)), Cox (1993, p.369) states that "While from one perspective this [condition (1.1)] is easily interpreted, it clearly does not satisfy the requirement of specifying a direct mode of data generation."

By contrast, a direct mode of data generation for the AMP is easily specified. Consider the recursive linear system

$$\begin{aligned} X_1 &= \epsilon_1 \\ X_2 &= \epsilon_2 \\ X_3 &= b_{31}X_1 + \epsilon_3 \\ X_4 &= b_{42}X_2 + \epsilon_4, \end{aligned} \tag{1.3}$$

where b_{31}, b_{42} are non-random scalars and where ϵ_1, ϵ_2, and (ϵ_3, ϵ_4) are mutually independent random errors with zero means, ϵ_1 and ϵ_2 have univariate normal distributions with arbitrary variances, and (ϵ_3, ϵ_4) has a bivariate normal distribution with arbitrary covariance matrix. Then (X_1, X_2, X_3, X_4) satisfies the AMP condition (1.2) for G, but not the LWF condition (1.1) when ϵ_3 and ϵ_4 have nonzero correlation.

Such a linear representation remains valid for a general CG G: the AMP for G is equivalent to the set of conditional independences (CIs) satisfied by the system of block-recursive linear equations naturally associated with G - see Remark 5.1.[2] Each variate is given as a linear function of its parents in the chain graph, together with a normal error term - see (5.6). The errors are independent across blocks, while within each block the errors are (possibly) correlated according to the undirected edges connecting them, as in a Gaussian UG model $\equiv$ 'covariance selection model' (Dempster 1972). The linear model constructed in this way differs from a standard linear structural equation model (SEM): a SEM model usually specifies zeroes in the covariance matrix for the error terms, while the covariance selection model sets to zero elements of the *inverse* error covariance matrix.

For nonsingular multivariate normal distributions, the LWF and AMP Markov properties generally are specified by different sets of constraints among regression coefficients and conditional covariance matrices. For example, consider a normal random vector (X_1, X_2, X_3, X_4) with mean vector (0,0,0,0) and positive definite covariance matrix $\Sigma \equiv (\sigma_{ij} \mid i, j = 1, 2, 3, 4)$. The conditional distribution of (X_3, X_4) given (X_1, X_2) is bivariate normal:

$$\begin{pmatrix} X_3 \\ X_4 \end{pmatrix} \Bigg| \, X_1, X_2 \; \sim \; \mathcal{N}_2\left(\beta \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \Lambda\right) \equiv \; \mathcal{N}_2\left(\begin{pmatrix} \beta_{31}X_1 + \beta_{32}X_2 \\ \beta_{41}X_1 + \beta_{42}X_2 \end{pmatrix}, \Lambda\right), \tag{1.4}$$

where

$$\Lambda := \begin{pmatrix} \sigma_{33} & \sigma_{34} \\ \sigma_{43} & \sigma_{44} \end{pmatrix} - \begin{pmatrix} \sigma_{31} & \sigma_{32} \\ \sigma_{41} & \sigma_{42} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{13} & \sigma_{14} \\ \sigma_{23} & \sigma_{24} \end{pmatrix} \tag{1.5}$$

[2] As noted in Lauritzen (1996, p.154), in general this does not hold for the LWF property; Theorem 4.3 below describes those G for which LWF $\equiv$ AMP.

is the conditional covariance matrix and

$$\beta \equiv \begin{pmatrix} \beta_{31} & \beta_{32} \\ \beta_{41} & \beta_{42} \end{pmatrix} := \begin{pmatrix} \sigma_{31} & \sigma_{32} \\ \sigma_{41} & \sigma_{42} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1} \tag{1.6}$$

is the matrix of regression coefficients of (X_3, X_4) on (X_1, X_2). By (1.4), the AMP condition (1.2) is equivalent to the directly interpretable conditions

$$\beta_{32} = \beta_{41} = 0, \quad \sigma_{12} = 0, \tag{1.7}$$

while the LWF condition (1.1) is equivalent to the less easily interpretable conditions

$$\gamma_{32} = \gamma_{41} = 0, \quad \sigma_{12} = 0, \tag{1.8}$$

where

$$\begin{pmatrix} \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \end{pmatrix} \equiv \gamma := \Lambda^{-1}\beta \tag{1.9}$$

is the natural exponential parameter occurring in the conditional normal distribution (1.4).

Yet another striking contrast between the LWF and AMP Markov properties for the CG G in Figure 1.1 appears when the marginal distribution of (X_1, X_3, X_4) is considered. Under the AMP property, the marginal distribution inherits the independence $1 \perp\!\!\!\perp 4$ implied by (1.2). Under the LWF property, however, no CI holds between X_1 and X_4 in the marginal distribution, despite the absence of an edge between 1 and 4 in G.

Similarly, for a general CG G, under the assumption of multivariate normality, both the LWF and AMP Markov properties imply that the joint distribution factors into a product of conditional normal distributions (see (4.3) and (5.2)), where each conditional distribution involves a regression matrix β and a conditional covariance matrix Λ. Under the AMP (Definition 4.1), the Markov conditions take the form of separate zero restrictions on certain elements of each β and each Λ^{-1} (see (5.3) and (5.4)). By contrast, under the LWF property (Definition 3.3), separate zero restrictions are imposed on certain elements of each $\gamma := \Lambda^{-1}\beta$ and each Λ^{-1} (the natural parameters when the normal distribution is expressed as an exponential family). Since regression coefficients are directly interpretable as indicators of dependence, at least under the assumption of normality, the AMP formulation seems preferable in this regard.

For these reasons, it is surprising that with the notable exception of [CW] (1993, 1996), the study of CG Markov models has been limited to the LWF interpretation.

[CW] (1996, p.34, 205) introduce *joint-response models* for CGs. Their CGs G may have both solid and dashed lines and arrows, with the restriction that the lines in any given chain component (see §2) of G must be either all solid or all dashed, and all arrows entering any given chain component must be either all solid or all dashed. The CI assumptions associated with solid vs. dashed lines and arrows are different: the Markov property for a CG with solid lines and dashed arrows (a CW concentration regression model) agrees with our AMP block-recursive Markov property (see Definition 4.1), while the Markov property for a CG with solid lines and solid arrows (a CW block regression model) agrees with the LWF block-recursive Markov property (see Definition 3.3).

[CW] (1996, p.40) note that whereas the global Markov properties of CW block regression CG models are well known, "no similar results are currently available" for any other cases of their joint-response CG Markov models, including their concentration regression models. Our new AMP global Markov property gives the first such result (Theorem 4.1).

This paper presents a study, begun in [AMP] (1996), of the Markov properties of CG models under the AMP formulation. For a CG that is either a UG or an ADG, the LWF and AMP Markov properties coincide. For a general CG, however, the AMP property in some ways seems a more direct extension of the ADG Markov property than the LWF property.

Our graph-theoretic terminology is introduced in Section 2. An LWF block-recursive Markov property for CGs is formulated in Definition 3.3, then shown in Theorem 3.1 to be equivalent to the LWF global Markov property.

Our AMP block-recursive Markov property for CGs is introduced in Definition 4.1, then shown to be equivalent to a new AMP global Markov property in Theorem 4.1. Theorem 4.3 gives the necessary and sufficient condition on a CG for its LWF and AMP Markov properties to coincide. An AMP local Markov property, two AMP pairwise Markov properties, and their relation to the AMP global Markov property are given in Section 4.

In Section 5 we show that under multivariate normality, the AMP Markov property for a CG G coincides with the Markov property satisfied by a block-recursive normal linear system (a CW concentration regression model) naturally associated with G. This model can be decomposed into a family of multivariate normal models, each of which combines the features of multivariate linear regression models and covariance selection models, facilitating the estimation of its statistical parameters.

In Section 6, the necessary and sufficient condition for the AMP Markov equivalence of two CGs is given, complementing the results of Frydenberg (1990) and [AMP] (1997a) for the LWF Markov equivalence of CGs. Our condition is then applied to determine when a given CG is AMP Markov equivalent to some UG, to some ADG, or to some decomposable UG. We also give necessary and sufficient conditions for the AMP Markov properties of a given CG G to coincide with the LWF Markov properties of at least one (possibly different) CG $\tilde{G}$, and necessary conditions for the converse, thereby demonstrating that each of the classes of AMP and LWF chain graph models contains models not belonging to the other class. Some additional aspects of AMP CG models, such as the essential graph and a pathwise separation criterion for conditional independence (unlike LWF CG models), are discussed briefly in Section 7. Appendices A and B contain the proofs of most main results.

The relations among the various classes of graphical Markov models considered in this paper are illustrated in Figure 6.2.

Except for the factorization (4.3) and the normal case treated in Section 5, the results in this paper are *not* limited to multivariate distributions that admit joint probability density functions.

2. Graph-theoretic Terminology.

Our development of graphs and graphical Markov models follows those of Lauritzen *et al* (1990), Frydenberg (1990), and [AMP] (1997a), but with several significant modifications. A *graph* G is a pair (V, E), where V is a finite set of *vertices* and $E \subseteq \{(v, w) \in V \times V \mid v \neq w\}$ is a set of *edges*, i.e., a set of ordered pairs of distinct vertices. An edge $(v, w) \in E$ whose opposite $(w, v) \in E$ is called an *undirected* edge and appears as a line v—w in our figures; in the text we write v—$w \in G$. An edge $(v, w) \in E$ whose opposite $(w, v) \notin E$, is called a *directed* edge and appears as an arrow $v \to w$ in our figures; in the text we write $v \to w \in G$. Two vertices $v, w \in V$ are *adjacent* in G, written as $v \cdots w \in G$, if $(v, w) \in E$ or $(w, v) \in E$ or both. A graph with only undirected edges is called an *undirected graph* (UG). A graph with only directed edges is a *directed graph* ($\equiv$ *digraph*).

A graph $G' \equiv (V', E')$ is a *weak subgraph* of $G \equiv (V, E)$, denoted by $G' \subseteq G$, if $V' \subseteq V$ and $E' \subseteq E$; we also say that G' is *smaller* than G and G is *larger* than G'. We write $G' \subset G$ if $G' \subseteq G$ but $G' \neq G$. A graph $G' \equiv (V', E')$ is a *subgraph* of $G \equiv (V, E)$, denoted by $G' \sqsubseteq G$, if $G' \subseteq G$ and also $v \to w \in G' \Rightarrow v \to w \in G$.

A subset $A \subseteq V$ *induces* the subgraph $G_A := (A, E_A)$, where $E_A := E \cap (A \times A)$; that is, E_A is obtained from E by retaining all edges with both endpoints in A. If $B \subseteq A \subseteq V$, clearly $(G_A)_B = G_B$ and $G_B \sqsubseteq G_A$.

Each graph $G = (V, E)$ determines two UGs $G^\vee \equiv (V, E^\vee)$, $G^\wedge \equiv (V, E^\wedge)$ defined by

$$E^\vee := \{(v, w) \,|\, (v, w) \in E \,\vee\, (w, v) \in E\},$$
$$E^\wedge := \{(v, w) \,|\, (v, w) \in E \,\wedge\, (w, v) \in E\},$$

respectively. Thus, $G^\vee$ is the *skeleton* of G, i.e., the underlying UG obtained by converting all arrows of G into lines, while $G^\wedge$ is obtained by deleting all arrows of G, so $G^\wedge \sqsubseteq G \subseteq G^\vee$. For any subset $A \subseteq V$, $(G_A)^\vee = (G^\vee)_A$ and $(G_A)^\wedge = (G^\wedge)_A$.

If $G_1^\vee = G_2^\vee$, then $G_1 \subseteq G_2$ iff G_1 and G_2 differ only in that some directed edges of G_1 may occur as undirected edges in G_2, while $G_1 \sqsubseteq G_2$ iff $G_1 = G_2$.

The *union* of a finite collection $(\,G_i \equiv (V_i, E_i) | i \in I)$ of (weak) subgraphs of a graph $G \equiv (V, E)$ is the (weak) subgraph $\cup G_i := (\cup V_i, \cup E_i)$. Clearly, $\cup G_i$ is the smallest (weak) subgraph larger than each G_i, $i \in I$. If $G_i' \sqsubseteq G_i \sqsubseteq G$ for each $i \in I$, then $\cup G_i' \sqsubseteq \cup G_i \sqsubseteq G$.

Let $G \equiv (V, E)$ be a graph and $A \subseteq V$ a subset of vertices. Denote the *boundary* of A in G by

$$\mathrm{bd}(A) \equiv \mathrm{bd}_G(A) := \{v \in V \backslash A \mid (v, a) \in E \text{ for some } a \in A\}$$

and the *closure* of A in G by

$$\mathrm{cl}(A) \equiv \mathrm{cl}_G(A) := A \dot{\cup} \mathrm{bd}(A).$$

The *parents* and *neighbors* of A in G, denoted by

$$\mathrm{pa}(A) \equiv \mathrm{pa}_G(A) := \{v \in V \backslash A \mid v \to a \in G \text{ for some } a \in A\},$$
$$\mathrm{nb}(A) \equiv \mathrm{nb}_G(A) := \{v \in V \backslash A \mid v \text{—} a \in G \text{ for some } a \in A\},$$

respectively, are those vertices $b \in V \setminus A$ that are linked to some $a \in A$ in G by directed edges or by undirected edges, respectively. Thus, $\mathrm{bd}(A) = \mathrm{pa}(A) \cup \mathrm{nb}(A)$.[3] The *children* of A in G are defined as

$$\mathrm{ch}(A) \equiv \mathrm{ch}_G(A) := \{v \in V \setminus A \mid a \to v \in G \text{ for some } a \in A\}.$$

A *path* of length $n \geq 1$ from v to w in G is a sequence $(v_0, v_1, \ldots, v_n)$ of $n+1$ distinct vertices such that $v_0 = v$, $v_n = w$, and $(v_{i-1}, v_i) \in E$ for all $i = 1, \ldots, n$. An *n-cycle* is a path of length $n \geq 3$ with the modification that $v_n = v_0$. If $v_{i-1} \to v_i \in G$ for (all) (at least one) (no) i, the path/cycle is called (*directed*), (*semi-directed*), (*undirected*). If G is a UG (digraph), all paths are undirected (directed). A path (n-cycle) is *chordless* if no two nonconsecutive (nonconsecutive (mod n)) vertices are adjacent.

A UG $G \equiv (V, E)$ is *chordal* if every n-cycle with $n \geq 4$ possesses a *chord*, that is, an edge between two nonconsecutive (mod n) vertices.[4] Clearly, any induced subgraph of a chordal graph is chordal.

A UG $G \equiv (V, E)$ is *complete* if all pairs of vertices are adjacent. A subset $A \subseteq V$ is *complete* if the induced subgraph G_A is complete. A maximal complete subset $C \subseteq V$ is called a *clique.* The UG G is *connected* if, for every distinct $v, w \in V$, there is a path between v and w in G. A subset $A \subseteq V$ is *connected* in G if G_A is connected. The maximal connected subsets are called the *connected components* of G, and V can be uniquely partitioned into the disjoint union of the connected components of G. For pairwise disjoint subsets $A(\neq \emptyset), B(\neq \emptyset)$, and S of V, A and B are *separated* by S in the UG G if all paths in G between A and B intersect S. Note that if $S = \emptyset$, then A and B and separated by S in G if and only if there are *no* paths connecting A and B in G. In this case, A and B are separated by *any* subset S disjoint from A and B.

A graph is called *adicyclic* if it contains no semi-directed cycles. An adicyclic graph is commonly called a *chain graph* (CG). Subgraphs (unions) of CGs are (need not be) chain graphs. An *acyclic digraph* (ADG) is a digraph that contains no directed cycles. Thus, UGs and ADGs are special cases of CGs.

For the remainder of this paper, let $G \equiv (V, E)$ be a chain graph. Define the following binary relations on V:

$$\begin{aligned}
v \leq_G w &\iff \exists \text{ a path in } G \text{ from } v \text{ to } w, \text{ or } v = w,\\
v \ll_G w &\iff \exists \text{ a directed path in } G \text{ from } v \text{ to } w,\\
v <_G w &\iff \exists \text{ a semi}-\text{directed path in } G \text{ from } v \text{ to } w,\\
v \sim_G w &\iff \exists \text{ an undirected path in } G \text{ from } v \text{ to } w \in G, \text{ or } v = w.
\end{aligned}$$

When G is understood, we simply write $v \leq w$, $v \ll w$, $v < w$, and $v \sim w$.

A subset $A \subseteq V$ is called *G-anterior* if $v \in A$ whenever $v \leq a$ for some $a \in A$. Equivalently, A is anterior iff $\mathrm{bd}(A) = \emptyset$, so A is anterior iff it is both coherent and

[3] In general, $pa(A)$ and $nb(A)$ need not be disjoint, but $pa(\sigma)$ and $nb(\sigma)$ are disjoint if G is a chain graph and $\sigma \subseteq \tau$ for some chain component τ of G.

[4] A UG is chordal iff it is *decomposable* - cf. Lauritzen, Speed, and Vijayan (1984, Theorem 2), Whittaker (1990, Proposition 12.4.2), or Lauritzen (1996, Proposition 2.5).

ancestral (see definitions below). If A and B are anterior, then $A \cap B$ is anterior. For any subset $A \subseteq V$, define $\mathrm{At}(A) \equiv \mathrm{At}_G(A) :=$ the smallest G-anterior set containing A. Clearly,

$$\mathrm{At}(A) = \{\, v \in V \mid v \leq a \text{ for some } a \in A \,\}.$$

For any subsets $A, B \subseteq V$, $\mathrm{At}(A \cup B) = \mathrm{At}(A) \cup \mathrm{At}(B)$ (but $\cup$ cannot be replaced by $\cap$).

For $A \subseteq V$, the *expanded subgraph* $G(A)$ is the induced subgraph (see Figures 2.7, 2.8)

$$G(A) := G_{\mathrm{At}(A)}.$$

For $A \subseteq V$, the set of *ancestors* of A in G is defined as

$$\mathrm{an}(A) \equiv \mathrm{an}_G(A) := \big\{ v \in V \backslash A \;\big|\; v \ll a \text{ for some } a \in A \big\}.$$

A subset $A \subseteq V$ is called *G-ancestral* if $\mathrm{an}(A) = \emptyset$. Equivalently, A is ancestral iff $\mathrm{pa}(A) = \emptyset$. If A and B are ancestral, so is $A \cap B$. Therefore, for any subset $A \subseteq V$, $\mathrm{An}(A) \equiv \mathrm{An}_G(A) :=$ the smallest G-ancestral set containing A, is well-defined and is given by

$$\mathrm{An}(A) = A \dot{\cup} \mathrm{an}(A).$$

For any subsets $A, B \subseteq V$, $\mathrm{An}(A \cup B) = \mathrm{An}(A) \cup \mathrm{An}(B)$ (but $\cup$ cannot be replaced by $\cap$).

For $A \subseteq V$, the set of *descendants* of A in G is defined as

$$\mathrm{de}(A) \equiv \mathrm{de}_G(A) := \big\{ v \in V \backslash A \;\big|\; a \ll v \text{ for some } a \in A \big\}.$$

The sets of *nondescendants (proper nondescendants)* of A in G are defined as

$$\begin{aligned} \mathrm{Nd}(A) \equiv \mathrm{Nd}_G(A) &: = V \backslash \mathrm{de}(A), \\ \mathrm{nd}(A) \equiv \mathrm{nd}_G(A) &: = \mathrm{Nd}(A) \backslash A. \end{aligned}$$

As in Frydenberg (1990), let $\mathcal{T} \equiv \mathcal{T}(G)$ denote the set of equivalence classes in V induced by the equivalence relation $\sim_G$. Equivalently, $\mathcal{T}$ is the set of connected components of $G^\wedge \equiv \cup(\, G_\tau \mid \tau \in \mathcal{T}\,)$ (see Figures 2.1a,b). Each vertex $v \in V$ lies in a unique chain component $\tau(v) \in \mathcal{T}$. For each $\tau \in \mathcal{T}$, $\mathrm{nb}(\tau) = \emptyset$, so $\mathrm{bd}(\tau) = \mathrm{pa}(\tau) = \{\mathrm{pa}(v) \mid v \in \tau\}$.

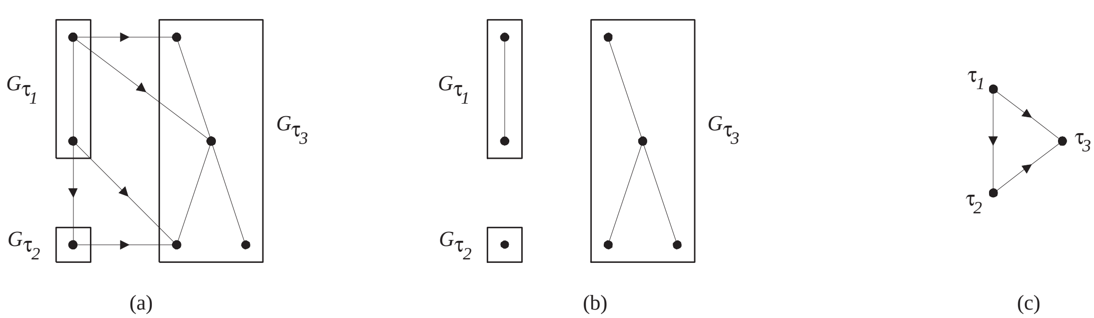


Figure 2.1. (a) A chain graph G with $\mathcal{T}(G) = \{\tau_1, \tau_2, \tau_3\}$. (b) The UG $G^\wedge$. (c) The ADG $\mathcal{D}(G)$.

A subset $A \subseteq V$ is called *G-coherent* if $v \in A$ whenever $v \sim a$ for some $a \in A$, that is, if A is a union of chain components of G. Equivalently, A is coherent iff $\mathrm{nb}(A) = \emptyset$. If A and B are coherent, then $A \cap B$ is coherent. For any $A \subseteq V$, define $\mathrm{Co}(A) \equiv \mathrm{Co}_G(A) :=$ the smallest G-coherent set containing A, so

$$\mathrm{Co}(A) = \{\, v \in V \mid v \sim a \text{ for some } a \in A \,\} = \cup(\tau \in \mathcal{T} \mid \tau \cap A \neq \emptyset).$$

For $A, B \subseteq V$, $\mathrm{Co}(A \cup B) = \mathrm{Co}(A) \cup \mathrm{Co}(B)$. Note that $\mathrm{An}(A) \subseteq \mathrm{Co}(\mathrm{An}(A)) \subseteq \mathrm{At}(A)$.

For $A \subseteq V$, the *extended subgraph* $G[A]$ is defined by (see Figures 2.7, 2.8)

$$G[A] := G_{\mathrm{An(A)}} \cup G^{\wedge}_{\mathrm{Co}(\mathrm{An}(A))}.$$

Note that $G[A] \sqsubseteq G$, i.e., $G[A]$ is in fact a subgraph of G, and that a *directed* edge occurs in $G[A]$ iff it occurs in $G_{\mathrm{An}(A)}$.

A chain component $\tau \in \mathcal{T}$ is *terminal* in G if $\mathrm{ch}_G(\tau) = \emptyset$. A subset $A \subseteq V$ is anterior iff it can be generated from V by stepwise removal of terminal chain components. (Note that the removal of a terminal chain component of G might render other chain components terminal in the remaining graph.) 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.

The chain components of G themselves comprise the vertices of the graph $\mathcal{D}(G) \equiv (\mathcal{T}(G), \mathcal{E}(G))$, where

$$\mathcal{E}(G) := \{\, (\tau, \tau') \in \mathcal{T} \times \mathcal{T} \mid \tau \neq \tau',\ \exists\, v \in \tau,\ v' \in \tau :\ v \to v' \in G \,\}. \tag{2.1}$$

Then $\mathcal{D} \equiv \mathcal{D}(G)$ is in fact an acyclic digraph (ADG) and $\tau \to \tau' \in \mathcal{D}$ iff $v \to v' \in G$ for some $v \in \tau, v' \in \tau'$ (see Figure 2.1c). The chain component τ is terminal in G iff τ is a terminal vertex in $\mathcal{D}$, i.e., $\mathrm{ch}_{\mathcal{D}}(\tau) = \emptyset$. A subset $A \subseteq V$ is G-anterior iff $A = \cup(\tau | \tau \in \mathcal{A})$ for some $\mathcal{D}$-ancestral set $\mathcal{A} \subseteq \mathcal{T}$.

A *k-complex*[5] $(a, b; c_1, \dots, c_k)$ in G is an induced subgraph of the form $a \to c_1 - \cdots - c_k \leftarrow b$ $(k \geq 1)$ as in Figure 2.2. Here $a, b, c_1, \dots, c_k$ are distinct vertices of G and, if $k \geq 2$, $(c_1, \dots, c_k)$ is a chordless undirected path in G. (Note that $(a, b; c_1, \dots, c_k) = (b, a; c_k, \dots, c_1)$.) A k-complex for $k \geq 2$ is called a *multicomplex*. A 1-complex $(a, b; c)$ is also called an *immorality* (Figure 2.4a).

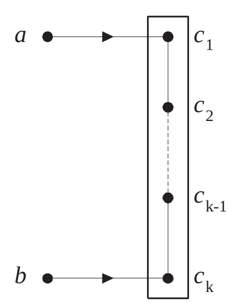


Figure 2.2. The k-complex $(a, b; c_1, \dots, c_k)$.

[5] Frydenberg (1990) uses the term *minimal complex*; we follow Studený (1996, 1997).

The *moralized* k-complex $(a, b; c_1, \ldots, c_k)^{\mathrm{m}}$ is the UG obtained from the complex by adding the undirected edge a—b and converting the arrows $a \to c_1$ and $b \to c_k$ to lines, yielding a chordless undirected $(k+2)$-cycle when $k \geq 2$ and an undirected 3-cycle when $k = 1$. The *moralized graph* G^{m} derived from a CG G is defined to be the UG obtained by moralizing all complexes in G, then converting all remaining arrows of G to lines - see Figures 2.3, 2.7e, 2.8e. Equivalently, G^{m} is obtained by completing $\mathrm{pa}_G(\tau)$ for each $\tau \in \mathcal{T}$, then converting all arrows of G to lines.

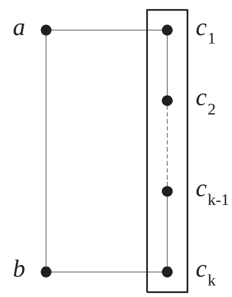


Figure 2.3. The moralized graph G^{m} for the chain graph G in Figure 2.2.

If $\tilde{G} \sqsubseteq G$ then $\tilde{G}^{\mathrm{m}} \subseteq G^{\mathrm{m}}$. (This need not hold if $\tilde{G} \subseteq G$.) In particular, for any subset $A \subseteq V$, $(G_A)^{\mathrm{m}} \subseteq G^{\mathrm{m}}$.[6] More generally, let A, B, C, D be subsets of V such that $D \subseteq C \subseteq A$ and $D \subseteq B \subseteq A$. Then $\mathrm{At}_{G_C}(D) \subseteq \mathrm{At}_{G_A}(B)$, so, since $G_A(B) = G_{\mathrm{At}_{G_A}(B)}$ and $G_C(D) = G_{\mathrm{At}_{G_C}(D)}$,

$$G_C(D)^{\mathrm{m}} \subseteq G_A(B)^{\mathrm{m}}. \tag{2.2}$$

Let a, b, c be distinct vertices of G. A *flag* $[a, b; c]$ in G is an induced subgraph of the form $a \to c$—b (Figure 2.4b). A *triplex* in G is an ordered pair $(\{a, b\}, c)$ such that *either* $(a, b; c)$ is an immorality in G *or else* $[a, b; c]$ or $[b, a; c]$ is a flag in G. Thus, the triplex $(\{a, b\}, c)$ occurs in G iff one of the three graphs shown in Figures 2.4a,b,c occurs as an induced subgraph of G.[7]

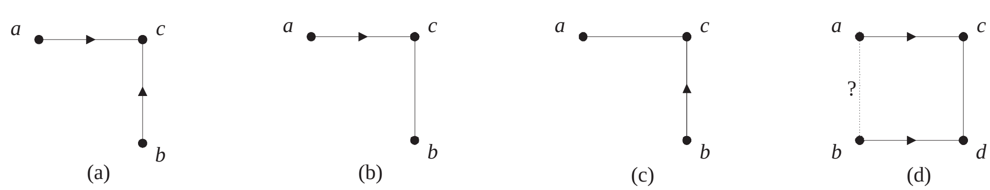


Figure 2.4. (a) The immorality $(a, b; c)$. (b) The flag $[a, b; c]$. (c) The flag $[b, a; c]$. The triplex $(\{a, b\}, c)$ occurs in G iff either (a), (b), or (c) occurs as an induced subgraph. (d) The four configurations that define the 2-biflag $[a, b; c, d]$.

---

[6] In fact, for any distinct $a, b, c_1, \ldots, c_k \in A$, $(a, b; c_1, \ldots, c_k)$ is a complex in G_A iff it is a complex in G.

[7] Our terminology differs from that in [AMP] (1996), where a triplex was called a flag and a 2-biflag was called a double flag.

Let $a, b, c_1, \ldots, c_k$ $(k \geq 2)$ denote distinct vertices of G such that $(c_1, \ldots, c_k)$ is a chordless undirected path of length $k-1$ in G. For $k \geq 3$, a *k-biflag* $[a; c_1, \ldots, c_k]$ in the CG G is an induced subgraph $G_{\{a,c_1,\ldots,c_k\}}$ of the form in Figure 2.5a. Here $[a, c_1; c_2]$ and $[a, c_k; c_{k-1}]$ are flags and $a \to c_i \in G$ for $i = 2, \ldots, k-1$. (Note that $[a; c_1, \ldots, c_k] = [a; c_k, \ldots, c_1]$.) For $k \geq 2$, a *k-biflag* $[a, b; c_1, \ldots, c_k]$ in the CG G is an induced subgraph $G_{\{a,b,c_1,\ldots,c_k\}}$ having one of the four forms indicated in Figure 2.5b. Here $[a, c_k; c_{k-1}]$ and $[b, c_1; c_2]$ are flags, $a \to c_i \in G$ for $i = 1, \ldots, k-1$, $b \to c_i \in G$ for $i = 2, \ldots, k$, and the "?" in Figure 2.5b indicates that either $a — b \in G$, $a \to b \in G$, $a \leftarrow b \in G$, or a and b are not adjacent in G. (Note that $[a, b; c_1, \ldots, c_k] = [b, a; c_k, \ldots, c_1]$.) The four possible forms of the 2-biflag $[a, b; c, d]$ are depicted in Figure 2.4d.

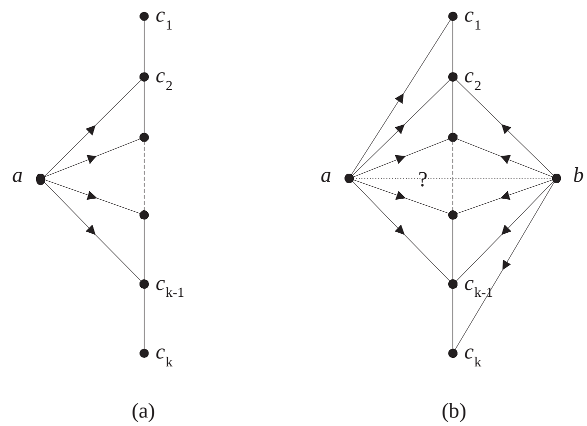


Figure 2.5. (a) The k-biflag $[a; c_1, \ldots, c_k]$, $k \geq 3$.
(b) The four configurations that define the k-biflag $[a, b; c_1, \ldots, c_k]$, $k \geq 2$.

The *augmented* triplex $(\{a, b\}, c)^{\mathrm{a}}$ is the complete UG with vertices a, b, c - cf. Figure 2.6a. The *augmented* 2-biflag $[a, b; c, d]^{\mathrm{a}}$ is the complete UG with vertices a, b, c, d - cf. Figure 2.6b. The *augmented graph* G^{a} derived from a CG G is defined to be the UG obtained by augmenting all triplexes and 2-biflags in G, then converting all remaining directed edges ($\equiv$ arrows) of G into undirected edges ($\equiv$ lines) - see Figures 2.6, 2.7c, and 2.8c. Equivalently, G^{a} is obtained by completing $C \dot{\cup} \mathrm{pa}_G(C)$ for every clique C in $G^{\wedge}$, then converting all arrows in G to lines. Note that $G[A] \sqsubseteq G(A)$, but neither $G[A]^{\mathrm{a}} \subseteq G(A)^{\mathrm{m}}$ nor $G(A)^{\mathrm{m}} \subseteq G[A]^{\mathrm{a}}$ in general - see Figures 2.8c,e and 2.7c,e.

Figure 2.6. (a) The augmented graph G^{a} for the CGs G in Figures 2.4a,b,c.
(b) The augmented graph G^{a} for the CG G in Figure 2.4d.

If $\tilde{G} \sqsubseteq G$ then $\tilde{G}^{\mathrm{a}} \subseteq G^{\mathrm{a}}$. (This need not hold if $\tilde{G} \subseteq G$.) In particular, for any subset $A \subseteq V$, $(G_A)^{\mathrm{a}} \subseteq G^{\mathrm{a}}$.[8] More generally, if A, B, C, D are subsets of V such that $D \subseteq C \subseteq A$ and $D \subseteq B \subseteq A$, then as in (2.2),

$$G_C[D]^{\mathrm{a}} \subseteq G_A[B]^{\mathrm{a}}. \tag{2.3}$$

If $G \equiv (V, E)$ is a UG, then $\mathcal{T}(G) = \mathcal{C}(G) :=$ the set of connected components of G, $\mathrm{Co}(A) = \mathrm{At}(A)$, $\mathrm{An}(A) = A$, $G[A] = G(A) = G_{\mathrm{Co}(A)}$, $G^{\mathrm{a}} = G^{\mathrm{m}} = G$, and $\mathcal{D}(G) = (\mathcal{C}(G), \emptyset)$. If $G \equiv (V, E)$ is an ADG, then $\mathcal{T}(G) = V$, $\mathrm{Co}(A) = A$, $\mathrm{An}(A) = \mathrm{At}(A)$, $G[A] = G(A) = G_{\mathrm{An}(A)}$, $G^{\mathrm{a}} = G^{\mathrm{m}}$, and $\mathcal{D}(G) = G$.

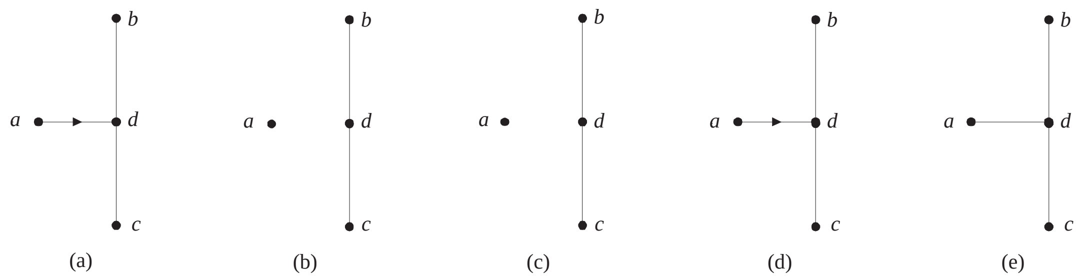


Figure 2.7. (a) A chain graph G. (b) The extended subgraph $G[\{a, b, c\}]$.
(c) The augmented graph $G[\{a, b, c\}]^{\mathrm{a}}$.
(d) The expanded subgraph $G(\{a, b, c\})$. (e) The moralized graph $G(\{a, b, c\})^{\mathrm{m}}$.

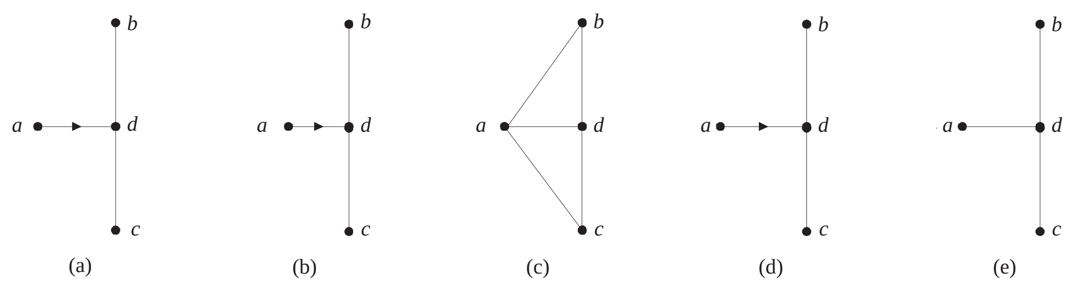


Figure 2.8. (a) A chain graph G. (b) The extended subgraph $G[\{b, c, d\}]$.
(c) The augmented graph $G[\{b, c, d\}]^{\mathrm{a}}$.
(d) The expanded subgraph $G(\{b, c, d\})$. (e) The moralized graph $G(\{b, c, d\})^{\mathrm{m}}$.

[8] In fact, for any distinct $a, b, c \in A$, $(\{a, b\}, c)$ is a triplex in G_A iff it is a triplex in G.

3. The LWF Markov Property for Chain Graphs.

We consider multivariate probability distributions P on a product probability space $\mathbf{X} \equiv \times(\mathbf{X}_v | v \in V)$, where V is a finite index set and each $\mathbf{X}_v$ is sufficiently regular to ensure the existence of regular conditional probabilities. Such a distribution is conveniently represented by a random variate $X \equiv (X_v | v \in V) \in \mathbf{X}$. For any subset $A \subseteq V$, we define $\mathbf{X}_A := \times(\mathbf{X}_v | v \in A)$, $X_A := (X_v | v \in A)$, and $X_\emptyset :=$ constant. We often abbreviate X_v and X_A by v and A, respectively. The marginal distribution of X_A is denoted by P_A.

For three pairwise disjoint subsets $A, B, C \subseteq V$ and a probability measure P on $\mathbf{X}$, write $A \perp\!\!\!\perp B \mid C\,[P]$ to indicate that X_A and X_B are conditionally independent given X_C under P. Trivially, $A \perp\!\!\!\perp B \mid C\,[P]$ if $A = \emptyset$ or $B = \emptyset$, while $A \perp\!\!\!\perp B \mid \emptyset\,[P]$ means that A and B are independent. Dawid (1980) notes that for any four pairwise disjoint subsets $A, B, C, D \subseteq V$,

$$A \perp\!\!\!\perp (B \dot{\cup} C) \mid D\,[P] \iff A \perp\!\!\!\perp B \mid C \dot{\cup} D\,[P] \text{ and } A \perp\!\!\!\perp C \mid D\,[P]. \tag{3.1}$$

A graphical Markov model uses a graph $G \equiv (V, E)$ with vertex set V to specify a Markov property, i.e., a collection of conditional independences, among the component random variates X_v, $v \in V$. Here we shall introduce our block-recursive version of the LWF Markov property for a CG G, then establish its equivalence to Frydenberg's global LWF Markov property. It is this block-recursive version, rather than the global property, that suggests our alternative Markov property introduced in Section 4.

Our block-recursive version of the LWF Markov property is naturally expressed in terms of the local ($\equiv$ global) Markov property determined by the ADG $\mathcal{D} \equiv \mathcal{D}(G) := (\mathcal{T}(G), \mathcal{E}(G)) \equiv (\mathcal{T}, \mathcal{E})$ (cf. (2.1)), the global Markov properties determined by the family of UGs $(G_\tau | \tau \in \mathcal{T})$, and a critical third property which, when suitably altered, yields the AMP block-recursive Markov property. First we review the local and global Markov properties for ADGs and UGs, respectively (Lauritzen *et al* (1990)).

Definition 3.1. (The local Markov property for ADGs.) Let $D \equiv (V, E)$ be an ADG. A probability measure P on $\mathbf{X}$ is said to be *local D-Markovian* if

$$v \perp\!\!\!\perp \big(\mathrm{nd}_D(v) \setminus \mathrm{pa}_D(v)\big) \;\big|\; \mathrm{pa}_D(v)\,[P] \quad \forall v \in V.$$

Definition 3.2. (The global Markov property for UGs.) Let $G \equiv (V, E)$ be a UG. A probability measure P on $\mathbf{X}$ is said to be *global G-Markovian* if $A \perp\!\!\!\perp B \mid S\,[P]$ whenever S separates A and B in G.

Lemma 3.1. Let $G \equiv (V, E)$ be a UG. A probability measure P on $\mathbf{X}$ is global G-Markovian if and only if

$$\forall A \subset V, \quad A \perp\!\!\!\perp \big(V \setminus \mathrm{cl}_G(A)\big) \;\big|\; \mathrm{nb}_G(A)\,[P]. \tag{3.2}$$

Proof. See Appendix A.

Definition 3.3. (The LWF block-recursive Markov property for chain graphs.) Let $G \equiv (V, E)$ be a CG and let $\mathcal{D} = \mathcal{D}(G)$. A probability measure P on $\mathbf{X}$ is said to be *LWF block-recursive G-Markovian* if P satisfies the following three conditions:

(C1) $\forall \tau \in \mathcal{T}: \ \tau \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_{\mathcal{D}}(\tau)\big) \ \big|\ \mathrm{pa}_{\mathcal{D}}(\tau)\,[P]$, i.e., P is local $\mathcal{D}$-Markovian on $\mathbf{X}$;[9]

(C2) $\forall \tau \in \mathcal{T}$: the conditional distribution $P_{\tau|\mathrm{pa}_{\mathcal{D}}(\tau)}$ is global G_τ-Markovian on $\mathbf{X}_\tau$;

(C3) $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau: \quad \sigma \perp\!\!\!\perp \big(\mathrm{pa}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \ \big|\ \mathrm{bd}_G(\sigma)\,[P]$.

The set of all LWF block-recursive G-Markovian P on $\mathbf{X}$ is denoted by $\mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G; \mathbf{X})$.

Proposition 3.1. Let $G \equiv (V, E)$ be a CG. Then $P \in \mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G; \mathbf{X})$ if and only if P satisfies the following *single* condition:

(C4) $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau: \quad \sigma \perp\!\!\!\perp \big(\mathrm{Nd}_{\mathcal{D}}(\tau) \setminus \mathrm{cl}_G(\sigma)\big) \ \Big|\ \mathrm{bd}_G(\sigma)\,[P]$.

Proof. Since $\mathrm{cl}_G(\sigma) = \mathrm{pa}_G(\sigma) \dot\cup \mathrm{cl}_{G_\tau}(\sigma)$ and $\mathrm{bd}_G(\sigma) = \mathrm{pa}_G(\sigma) \dot\cup \mathrm{nb}_G(\sigma)$ when $\sigma \subseteq \tau$, C4 can be stated equivalently as follows:

(C4′) $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau: \quad \sigma \perp\!\!\!\perp \Big(\big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \dot\cup \big(\tau \setminus \mathrm{cl}_{G_\tau}(\sigma)\big)\Big) \ \Big|\ \mathrm{pa}_G(\sigma) \dot\cup \mathrm{nb}_G(\sigma)\,[P]$,

where $\mathrm{nd}_{\mathcal{D}}(\tau) = \mathrm{Nd}_{\mathcal{D}}(\tau) \setminus \tau \supseteq \mathrm{pa}_{\mathcal{D}}(\tau) \supseteq \mathrm{pa}_G(\tau)$. By Lemma 3.1, C2 can be stated in the following equivalent form:

(C2′) $\forall \tau \in \mathcal{T}, \forall \sigma \subset \tau: \quad \sigma \perp\!\!\!\perp \big(\tau \setminus \mathrm{cl}_{G_\tau}(\sigma)\big) \ \big|\ \mathrm{pa}_{\mathcal{D}}(\tau) \dot\cup \mathrm{nb}_G(\sigma)\,[P]$.

Because $\tau = \sigma \dot\cup \mathrm{nb}_G(\sigma) \dot\cup (\tau \setminus \mathrm{cl}_{G_\tau}(\sigma))$, C1 implies that $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau$,

$$\sigma \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_{\mathcal{D}}(\tau)\big) \ \big|\ \big(\tau \setminus \mathrm{cl}_{G_\tau}(\sigma)\big) \dot\cup \mathrm{pa}_{\mathcal{D}}(\tau) \dot\cup \mathrm{nb}_G(\sigma)\,[P],$$

which combines with C2′ via (3.1) to yield the following condition:

(C5) $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau: \quad \sigma \perp\!\!\!\perp \Big(\big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_{\mathcal{D}}(\tau)\big) \dot\cup \big(\tau \setminus \mathrm{cl}_{G_\tau}(\sigma)\big)\Big) \ \Big|\ \mathrm{pa}_{\mathcal{D}}(\tau) \dot\cup \mathrm{nb}_G(\sigma)\,[P]$,

Because

$$\mathrm{pa}_{\mathcal{D}}(\tau) \dot\cup \mathrm{nb}_G(\sigma) = \big(\mathrm{pa}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \dot\cup \mathrm{bd}_G(\sigma),$$

C3 and C5 combine via (3.1) to yield C4′, hence C4.

Conversely, set $\sigma = \tau$ in C4′ and note that $\mathrm{pa}_{\mathcal{D}}(\tau) \supseteq \mathrm{pa}_G(\tau)$ to obtain C1. Similarly, C4 $\equiv$ C4′ yields C2′ and C3. This completes the proof.

Remark 3.1. As shown in the proof of Proposition 3.1, conditions C1 and C2 $\equiv$ C2′ together imply C5, while trivially C5 implies C2′, and C5 with $\sigma = \tau$ yields C1. Thus the *two* conditions C3 and C5 also characterize $\mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G; \mathbf{X})$.

[9] Note that $\mathcal{T}$ determines a coarser factorization of the product space X, i.e., $X = \times(X_\tau | \tau \in \mathcal{T})$.

Definition 3.4. (The LWF global Markov property for chain graphs.) Let $G \equiv (V, E)$ be a CG. A probability measure P on $\mathbf{X}$ is said to be *LWF global G-Markovian* if $A \perp\!\!\!\perp B \mid S\,[P]$ whenever S separates A and B in $G(A\dot{\cup}B\dot{\cup}S)^{\mathrm{m}}$. The set of all LWF global G-Markovian P on $\mathbf{X}$ is denoted by $\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X})$.

If G is an ADG, Lauritzen *et al* (1990) showed that its local and LWF global Markov properties are equivalent; trivially its LWF block-recursive and LWF global Markov properties coincide in this case. If G is a UG, Frydenberg (1990, p.339) noted that its global and LWF global Markov properties are equivalent; its LWF block-recursive and LWF global Markov properties also coincide, both being equivalent to the mutual independence of its connected components and the global G-Markov property within each connected component. Theorem 3.1 states that the LWF block-recursive and LWF global Markov properties are equivalent for general CGs; also see Theorem 2 of Buntine (1995).

Theorem 3.1. Let $G \equiv (V, E)$ be a chain graph. Then

(i) $\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G;\mathbf{X}) \;\; \forall \mathbf{X}$;

(ii) $\mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X}) \;\; \forall \mathbf{X}$.

Therefore $\mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G;\mathbf{X}) = \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X}) =: \mathcal{P}_{\mathrm{LWF}}(G;\mathbf{X}) \;\; \forall \mathbf{X}$.

Proof. See Appendix A.

For a general CG $G \equiv (V, E)$, let $\mathcal{P}^{\mathrm{l}}_{\mathrm{LWF}}(G;\mathbf{X})$ and $\mathcal{P}^{\mathrm{p}}_{\mathrm{LWF}}(G;\mathbf{X})$ denote the sets of all LWF local and pairwise G-Markovian distributions P on $\mathbf{X}$, respectively, defined by Frydenberg (1990, §3) as follows: P is *LWF local G-Markovian* if

$$v \perp\!\!\!\perp \big(\mathrm{Nd}_{\mathcal{D}}(\tau(v)) \setminus \mathrm{cl}_G(v)\big) \,\Big|\, \mathrm{bd}_G(v)\,[P] \quad \forall v \in V \tag{3.3}$$

(i.e., if P satisfies C4 for all singletons $\sigma = \{v\}$), while P is *LWF pairwise G-Markovian* if

$$v \perp\!\!\!\perp w \,\Big|\, \mathrm{Nd}_{\mathcal{D}}(\tau(v)) \setminus \{v, w\}\,[P] \tag{3.4}$$

for all pairs $v, w \in V$ such that $v \not\sim w$ in G and $w \in \mathrm{Nd}_{\mathcal{D}}(\tau(v))$. Frydenberg showed that

$$\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{l}}_{\mathrm{LWF}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{p}}_{\mathrm{LWF}}(G;\mathbf{X}), \tag{3.5}$$

$$\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X}) = \mathcal{P}^{\mathrm{l}}_{\mathrm{LWF}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X}) = \mathcal{P}^{\mathrm{p}}_{\mathrm{LWF}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X}), \tag{3.6}$$

where $\mathcal{C}(\mathbf{X})$ is the class of all distributions P on $\mathbf{X}$ that satisfy the following condition:

$$A \perp\!\!\!\perp B \mid C\dot{\cup}D\,[P] \text{ and } A \perp\!\!\!\perp C \mid B\dot{\cup}D\,[P] \;\Rightarrow\; A \perp\!\!\!\perp (B\dot{\cup}C) \mid D\,[P] \tag{CI5}$$

whenever A, B, C, D are pairwise disjoint subsets of V. Condition CI5 is satisfied whenever P has a *positive* joint probability density function on $\mathbf{X}$ with respect to some product of

σ-finite measures, but CI5 also holds for more general distributions - cf. [AMP] (1997a, Remark 3.3).

4. An Alternative Markov Property for Chain Graphs.

When applied to the CG in Figure 1.1, condition C1 reduces to $1 \perp\!\!\!\perp 2$, C2 is vacuous, while C3 yields the first two LWF Markov conditions in (1.1). In order to obtain instead the first two AMP Markov conditions in (1.2), we need only modify C3 by deleting the subset $\mathrm{nb}_G(\sigma)$ from the conditioning set $\mathrm{bd}_G(\sigma) \equiv \mathrm{pa}_G(\sigma)\dot{\cup}\mathrm{nb}_G(\sigma)$, as follows:

$$\text{(C3*)}\quad \forall\tau \in \mathcal{T}, \forall\sigma \subseteq \tau : \quad \sigma \perp\!\!\!\perp \big(\mathrm{pa}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \mid \mathrm{pa}_G(\sigma)\,[P].$$

Conditions C1, C2, and C3* constitute an alternative block-recursive Markov property for CGs which, unlike the LWF property, can be represented by a block-recursive normal linear system - see (1.3) and (5.6).

Definition 4.1. (The AMP block-recursive Markov property for chain graphs.) Let $G \equiv (V, E)$ be a CG. A probability measure P on $\mathbf{X}$ is said to be *AMP block-recursive G-Markovian* if P satisfies conditions C1, C2, and C3*. The set of all AMP block-recursive G-Markovian P on $\mathbf{X}$ is denoted by $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$.

Unlike condition C4 which characterizes $\mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G;\mathbf{X})$, there does not appear to be a single condition that characterizes $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$. As noted in Remark 3.1, C1 and C2 $\equiv$ C2′ together are equivalent to C5, so the *two* conditions C3* and C5 characterize $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$. Next, C3* (set $\sigma = \tau$) and C5 combine to yield

$$\text{(C6)}\quad \forall\tau \in \mathcal{T}, \forall\sigma \subseteq \tau : \quad \sigma \perp\!\!\!\perp \Big(\big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)\big)\dot{\cup}\big(\tau\setminus\mathrm{cl}_{G_\tau}(\sigma)\big)\Big) \;\Big|\; \mathrm{pa}_G(\tau)\dot{\cup}\mathrm{nb}_G(\sigma)\,[P].$$

Conversely, set $\sigma = \tau$ in C6 to obtain C1, while C6 directly implies C2′. Thus the *two* conditions C3* and C6 also characterize $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$. (Note that C4 $\Rightarrow$ C6 $\Rightarrow$ C5.)

The next lemma is convenient for the proof of Theorem 4.2 below, for the proof of Theorem 4.1 given in Appendix A, and for the proof of Proposition B.1 in Appendix B.

Lemma 4.1. Let G be a chain graph and P a probability measure on $\mathbf{X}$. Then $P \in \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$ if and only if P satisfies the following two conditions with respect to G:

$$\text{(C7)}\quad \forall\tau \in \mathcal{T}, \forall\sigma \subseteq \tau : \quad \sigma \perp\!\!\!\perp \big(\tau\setminus\mathrm{cl}_{G_\tau}(\sigma)\big) \;\big|\; \mathrm{pa}_G(\tau)\dot{\cup}\mathrm{nb}_G(\sigma)\,[P];$$

$$\text{(C8)}\quad \forall\tau \in \mathcal{T}, \forall\sigma \subseteq \tau : \quad \sigma \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \;\big|\; \mathrm{pa}_G(\sigma)\,[P].$$

Proof. We have seen that $P \in \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$ iff P satisfies C3* and C6. Trivially, C6 implies C7. Now set $\sigma = \tau$ in C6 to see that $\forall\tau \in \mathcal{T}, \forall\sigma \subseteq \tau$,

$$\sigma \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)\big) \;\big|\; \big(\mathrm{pa}_G(\tau) \setminus \mathrm{pa}_G(\sigma)\big)\dot{\cup}\mathrm{pa}_G(\sigma)\,[P],$$

while C3* implies that $\forall\tau \in \mathcal{T}, \forall\sigma \subseteq \tau$,

$$\sigma \perp\!\!\!\perp \big(\mathrm{pa}_G(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \;\big|\; \mathrm{pa}_G(\sigma)\,[P].$$

These two relations combine via (3.1) to yield C8.

Conversely, C8 implies C3*. Set $\sigma = \tau$ in C8 to see that $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau$,

$$\sigma \dot\cup \mathrm{nb}_G(\sigma) \dot\cup \big(\tau \backslash \mathrm{cl}_{G_\tau}(\sigma)\big) \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)\big) \mid \mathrm{pa}_G(\tau)\, [P],$$

hence

$$\sigma \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)\big) \mid \mathrm{pa}_G(\tau) \dot\cup \mathrm{nb}_G(\sigma) \dot\cup \big(\tau \backslash \mathrm{cl}_{G_\tau}(\sigma)\big)\, [P].$$

This combines with C7 to yield C6.

Definition 4.2. (The AMP global Markov property for chain graphs.) Let $G \equiv (V, E)$ be a CG. A probability measure P on $\mathbf{X}$ is said to be *AMP global G-Markovian* if $A \perp\!\!\!\perp B \mid S\, [P]$ whenever S separates A and B in $G[A \dot\cup B \dot\cup S]^{\mathrm{a}}$. The set of all AMP global G-Markovian P on $\mathbf{X}$ is denoted by $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G; \mathbf{X})$.

The AMP block-recursive and AMP global Markov properties are also equivalent:

Theorem 4.1. Let $G \equiv (V, E)$ be a chain graph. Then

(i) $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G; \mathbf{X}) \subseteq \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G; \mathbf{X}) \;\; \forall \mathbf{X}$;

(ii) $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G; \mathbf{X}) \subseteq \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G; \mathbf{X}) \;\; \forall \mathbf{X}$.

Therefore $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G; \mathbf{X}) = \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G; \mathbf{X}) =: \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) \;\; \forall \mathbf{X}$.

Proof. See Appendix A.

As noted in Section 3, Frydenberg (1990) introduced LWF local and pairwise Markov properties for CGs that are successively weaker than the LWF global property, and showed the equivalence of all three LWF Markov properties under assumption CI5. We now introduce analogous AMP local and (two) pairwise Markov properties for CGs, then examine their relation to the AMP global property. Although our first AMP pairwise property is equivalent to our AMP local property under CI5, an additional assumption (CI5*) is needed for the AMP local property to be equivalent to the AMP global property. For this reason our AMP local and pairwise Markov properties should not be considered definitive, but we do utilize them in our proofs of Theorems 4.3, 6.1 and 6.2 on Markov equivalence.

Definition 4.3. (An AMP local Markov property for chain graphs.) Let $G \equiv (V, E)$ be a CG and let $\mathcal{D} = \mathcal{D}(G)$. A probability measure P on $\mathbf{X}$ is said to be *AMP local G-Markovian* if P satisfies the following *two* conditions:

(L1) $\forall v \in V:\; v \perp\!\!\!\perp \big(\tau(v) \backslash \mathrm{cl}_{G_{\tau(v)}}(v)\big) \mid \mathrm{nd}_{\mathcal{D}}(\tau(v)) \dot\cup \mathrm{nb}_{G_{\tau(v)}}(v)\, [P]$;

(L2) $\forall v \in V:\; v \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau(v)) \backslash \mathrm{pa}_G(v)\big) \mid \mathrm{pa}_G(v)\, [P]$.

The set of all AMP local G-Markovian P on $\mathbf{X}$ is denoted by $\mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G; \mathbf{X})$.

Definition 4.4. (Two AMP pairwise Markov properties for chain graphs.) Let $G \equiv (V, E)$ be a CG and let $\mathcal{D} = \mathcal{D}(G)$. A probability measure P on $\mathbf{X}$ is said to be *AMP pairwise (pairwise′) G-Markovian* if P satisfies the following conditions P1 and P2 (P1′ and P2′):

(P1) $\forall v \in V,\ \forall w \in \tau(v)\backslash \mathrm{cl}_{G_{\tau(v)}}(v): \quad v \perp\!\!\!\perp w \,\big|\, \mathrm{nd}_{\mathcal{D}}(\tau(v)) \dot\cup \big(\tau(v)\backslash\{v, w\}\big)\, [P];$

(P2) $\forall v \in V,\ \forall w \in \mathrm{nd}_{\mathcal{D}}(\tau(v))\backslash \mathrm{pa}_G(v): \ v \perp\!\!\!\perp w \,\big|\, \mathrm{nd}_{\mathcal{D}}(\tau(v))\backslash\{w\}\, [P].$

(P1′) $\forall v \in V,\ \forall w \in \tau(v)\backslash \mathrm{cl}_{G_{\tau(v)}}(v): \quad v \perp\!\!\!\perp w \,\big|\, \mathrm{nd}_{\mathcal{D}}(\tau(v)) \dot\cup \mathrm{nb}_{G_{\tau(v)}}(v)\, [P];$

(P2′) $\forall v \in V,\ \forall w \in \mathrm{nd}_{\mathcal{D}}(\tau(v))\backslash \mathrm{pa}_G(v): \ v \perp\!\!\!\perp w \,\big|\, \mathrm{pa}_G(v)\, [P].$

The set of all AMP pairwise (pairwise′) G-Markovian P on $\mathbf{X}$ is denoted by $\mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G; \mathbf{X})$ $(\mathcal{P}^{\mathrm{p}'}_{\mathrm{AMP}}(G; \mathbf{X}))$.

If G is a connected UG (arbitrary ADG) then L2, P2, and P2′ (L1, P1, and P1′) are vacuous while L1 and P1 (L2 and P2) reduce to the local and pairwise Markov properties for UGs (ADGs) (cf. Lauritzen (1996)). If G is the CG in Figure 4.1 then L1, P1, and P1′ are vacuous while L2, P2, and P2′ each reduce to the two CIs $a \perp\!\!\!\perp c|b$ and $a \perp\!\!\!\perp d|b$. Note that these do not imply the AMP global Markov property $a \perp\!\!\!\perp cd|b$ for G, even if CI5 is assumed; instead, the less generally applicable assumption CI5* (below) is needed. By contrast, the LWF local (3.3) and pairwise (3.4) Markov properties for G both reduce to the two CIs $a \perp\!\!\!\perp c|bd$ and $a \perp\!\!\!\perp d|bc$, which do imply the LWF ($\equiv$ AMP, by Theorem 4.3) global Markov property $a \perp\!\!\!\perp cd|b$ if CI5 is assumed.

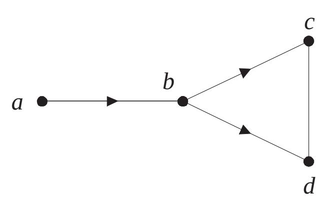


Figure 4.1.

Let $\mathcal{C}^*(\mathbf{X})$ be the set of probabilities P on $\mathbf{X}$ that satisfy the following condition:

(CI5*) $\qquad A \perp\!\!\!\perp B \mid D\, [P]$ and $A \perp\!\!\!\perp C \mid D\, [P] \ \Rightarrow\ A \perp\!\!\!\perp (B \dot\cup C) \mid D\, [P]$

whenever A, B, C, D are pairwise disjoint subsets of V. As noted by Kauermann (1996, pp. 107-8), unlike CI5, CI5* need not hold even if P has a positive joint probability density function on $\mathbf{X}$ with respect to some product measure, but CI5* does hold for normal distributions. Let $\mathcal{C}^{**}(\mathbf{X}) := \mathcal{C}(\mathbf{X}) \cap \mathcal{C}^*(\mathbf{X})$.

Theorem 4.2. Let $G \equiv (V, E)$ be a chain graph. Then

(i) $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{P}^{\mathrm{p}'}_{\mathrm{AMP}}(G;\mathbf{X}) \;\; \forall \mathbf{X}$;

(ii) $\mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X}) = \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X}) \;\; \forall \mathbf{X}$.

(iii) $\mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^*(\mathbf{X}) = \mathcal{P}^{\mathrm{p}'}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^*(\mathbf{X}) \;\; \forall \mathbf{X}$.

(iv) $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X}) = \mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X}) \;\; \forall \mathbf{X}$.

Therefore

$$\begin{aligned} &\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X}) = \mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X}) \\ &= \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X}) = \mathcal{P}^{\mathrm{p}'}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X}) \;\; \forall \mathbf{X}. \end{aligned}$$

Proof. (i) If $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X})$ then P satisfies conditions C6 and C8 (by Theorem 4.1 and Lemma 4.1). Set $\sigma = \{v\}$ and $\tau = \tau(v)$ in C6 and C8 to obtain L1 (by (3.1)) and L2, respectively, so the first inclusion in (i) holds. Since L1 implies P1 and P1′ and L2 implies P2 and P2′ by (3.1), the second inclusion is immediate.

(ii) If $P \in \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X})$ then P satisfies P1, P2, and CI5. Because CI5 is inherited under conditioning, for any $v \in V$ the conditional distribution $P_{\tau(v)|\mathrm{nd}_{\mathcal{D}}(\tau(v))}$ on $\mathbf{X}_{\tau(v)}$ satisfies CI5 and is pairwise $G_{\tau(v)}$-Markovian by P1, hence is local $G_{\tau(v)}$-Markovian (equivalently, P satisfies L1) by the equivalence of the pairwise and local Markov properties under CI5 for UGs (cf. Lauritzen (1996)).

Next, let $\{w_1, \ldots, w_n\}$ be an enumeration of the vertices in $\mathrm{nd}_{\mathcal{D}}(\tau(v)) \backslash \mathrm{pa}_G(v)$. Then P2 implies

$$\begin{aligned} v \perp\!\!\!\perp w_1 \bigm| \{w_2, w_3, \ldots, w_n\} \dot\cup \mathrm{pa}_G(v)\, [P], \\ v \perp\!\!\!\perp w_2 \bigm| \{w_1, w_3, \ldots, w_n\} \dot\cup \mathrm{pa}_G(v)\, [P], \end{aligned}$$

hence by CI5,

$$v \perp\!\!\!\perp (w_1, w_2) \bigm| \{w_3, \ldots, w_n\} \dot\cup \mathrm{pa}_G(v)\, [P].$$

Continue this procedure to obtain L2. Thus $P \in \mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}(\mathbf{X})$.

(iii) The implications P1′ $\Rightarrow L1$ and P2′ $\Rightarrow L2$ are obtained by repeated application of CI5*.

(iv) If $P \in \mathcal{P}^{\mathrm{l}}_{\mathrm{AMP}}(G;\mathbf{X}) \cap \mathcal{C}^{**}(\mathbf{X})$ then P satisfies L1, L2, CI5, and CI5*. To show that $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X})$, by Lemma 4.1 it suffices to show that P satisfies C7 and C8. For $\tau \in \mathcal{T}(G)$, the conditional distribution $P_{\tau|\mathrm{nd}_{\mathcal{D}}(\tau)}$ on $\mathbf{X}_\tau$ satisfies CI5 and is local G_τ-Markovian by L1, hence is global G_τ-Markovian by the equivalence of the local and global Markov properties under CI5 for UGs (cf. Lauritzen (1996)). Therefore P satisfies

(C7′) $\;\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau : \;\; \sigma \perp\!\!\!\perp \big(\tau \backslash \mathrm{cl}_{G_\tau}(\sigma)\big) \bigm| \mathrm{nd}_{\mathcal{D}}(\tau) \dot\cup \mathrm{nb}_{G_\tau}(\sigma)\, [P].$

For $\tau \in \mathcal{T}(G)$ and $\sigma \subseteq \tau$, enumerate the vertices of σ as $\{v_1, \ldots, v_m\}$. Because $\mathrm{pa}_G(v_1) \cup \mathrm{pa}_G(v_2) = \mathrm{pa}_G(\{v_1, v_2\})$, it follows from L2 (by (3.1)) that

$$\begin{aligned} v_1 \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \backslash \mathrm{pa}_G(\{v_1, v_2\})\big) \bigm| \mathrm{pa}_G(\{v_1, v_2\})\, [P], \\ v_2 \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \backslash \mathrm{pa}_G(\{v_1, v_2\})\big) \bigm| \mathrm{pa}_G(\{v_1, v_2\})\, [P], \end{aligned}$$

hence by CI5*,

$$\{v_1, v_2\} \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \backslash \mathrm{pa}_G(\{v_1, v_2\})\big) \bigm| \mathrm{pa}_G(\{v_1, v_2\})\,[P].$$

Continue this procedure to obtain C8. Now set $\sigma = \tau$ in C8 to obtain

$$\tau \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)\big) \bigm| \mathrm{pa}_G(\tau)\,[P],$$

hence by (3.1),

$$\sigma \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)\big) \bigm| \mathrm{pa}_G(\tau) \dot\cup \mathrm{nb}_{G_\tau}(\sigma)\,[P].$$

Since $\mathrm{nd}_{\mathcal{D}}(\tau) = (\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\tau)) \dot\cup \mathrm{pa}_G(\tau)$, this combines with C7$'$ by (3.1) to yield C6, which in turn yields C7. This completes the proof.

The following lemma will be used in the proofs of Theorems 4.3, 6.1, and 6.2.

Lemma 4.2. Let $G \equiv (V, E)$ be a chain graph. For any subset $A \subseteq V$ let P_A be a probability distribution on $\mathbf{X}_A$, and for each $v \in V \backslash A$ let P_v be a probability distribution on $\mathbf{X}_v$. Let P be the product probability measure on $\mathbf{X}$ defined by $P := P_A \times \big(\times (P_v | v \in V \backslash A)\big)$, so that under P, X_A and $(X_v | v \in V \setminus A)$ are mutually independent. Then

(i) $P_A \in \mathcal{P}_{\mathrm{LWF}}(G_A; \mathbf{X_A}) \Rightarrow P \in \mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$;

(ii) $P_A \in \mathcal{P}_{\mathrm{AMP}}(G_A; \mathbf{X_A}) \Rightarrow P \in \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$.

Proof. By the definition of the LWF global Markov property, to prove (i) it suffices to show that $B \perp\!\!\!\perp C \mid S\,[P]$ whenever B, C, S are disjoint subsets of V such that S separates B and C in $G(B \dot\cup C \dot\cup S)^{\mathrm{m}}$. Let $B_A := B \cap A$, $C_A := C \cap A$, $S_A := S \cap A$. Since X_A and $(X_v | v \in V \setminus A)$ are mutually independent under P, it suffices to show that $B_A \perp\!\!\!\perp C_A \mid S_A\,[P_A]$, hence it suffices to show that S_A separates B_A and C_A in $G_A(B_A \dot\cup C_A \dot\cup S_A)^{\mathrm{m}}$. But this follows from the assumed separation of B and C by S in $G(B \dot\cup C \dot\cup S)^{\mathrm{m}} \supseteq G_A(B_A \dot\cup C_A \dot\cup S_A)^{\mathrm{m}}$ (cf. (2.2)). The proof of (ii) is analogous, using (2.3).

For UGs and ADGs the LWF and AMP global Markov properties coincide, since

$$G(A \dot\cup B \dot\cup S)^{\mathrm{m}} = G[A \dot\cup B \dot\cup S]^{\mathrm{a}} = G_{\mathrm{Co}(A \dot\cup B \dot\cup S)}$$

if G is a UG and

$$G(A \dot\cup B \dot\cup S)^{\mathrm{m}} = G[A \dot\cup B \dot\cup S]^{\mathrm{a}} = (G_{\mathrm{An}(A \dot\cup B \dot\cup S)})^{\mathrm{m}}$$

if G is an ADG. The simplest CG for which the LWF and AMP global Markov properties differ is the graph $a \to c — b$ consisting of the flag $[a, b; c]$ alone. The LWF global Markov property for this graph is $a \perp\!\!\!\perp b \mid c$, while the AMP global Markov property is $a \perp\!\!\!\perp b$. The non-occurrence of a flag is necessary and sufficient for these two Markov properties to coincide.

Theorem 4.3. Let $G \equiv (V, E)$ be a chain graph.

(i) If G has no flags, then $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X}) = \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) \ \ \forall \mathbf{X}$.

(ii) If G has at least one flag, then for every $\mathbf{X}$ such that $\mathbf{X}_v$ admits a non-degenerate probability measure for each $v \in V$,

$$\big(\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X}) \cap \mathcal{C}(\mathbf{X})\big) \setminus \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) \neq \emptyset, \tag{4.1}$$

$$\big(\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) \cap \mathcal{C}(\mathbf{X})\big) \setminus \mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X}) \neq \emptyset. \tag{4.2}$$

Proof. Recall that $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$ is characterized by C4$'$ and $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$ by C3* and C5 or, equivalently, by C3* and C6.

(i) If G has no flags, then, since each $\tau \in \mathcal{T}(G)$ is connected in G, $\mathrm{pa}_G(\sigma) = \mathrm{pa}_G(\tau)$ whenever $\emptyset \neq \sigma \subseteq \tau$. Thus C4$'$ and C6 are identical in this case, so $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) \subseteq \mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$. Furthermore, by applying C4$'$ with $\sigma = \tau$ we obtain C3* in this case, while C4$'$ implies C5 for any G, hence $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X}) \subseteq \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$.

(ii) Assume that the flag $[a, b; c]$ occurs in G. If $\mathbf{X}_a, \mathbf{X}_b, \mathbf{X}_c$ each admit non-degenerate probability distributions, we can construct two probability measures $P, P^* \in \mathcal{C}(\mathbf{X})$ such that X_a, X_b, and X_c are nondegenerate Bernoulli random variables under P and P^*, $X_a \perp\!\!\!\perp X_b \mid X_c\,[P]$ but $X_a \not\!\perp\!\!\!\perp X_b\,[P]$, $X_a \perp\!\!\!\perp X_b\,[P^*]$ but $X_a \not\!\perp\!\!\!\perp X_b \mid X_c\,[P^*]$, and $(X_v | v \in V \setminus A) = \text{constant}\,[P], [P^*]$. By Lemma 4.2 with $A = \{a, b, c\}$, $P \in \mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$ and $P^* \in \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$. Since $a \not\!\perp\!\!\!\perp b\,[P]$, P fails to satisfy condition P2 for $(v, w) = (b, a)$, so $P \notin \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G; \mathbf{X})$, hence $P \notin \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$ by Theorem 4.2(i). Similarly, $a \not\!\perp\!\!\!\perp b | c\,[P^*]$ so P^* fails to satisfy (3.4) for $(v, w) = (b, a)$, so $P^* \notin \mathcal{P}^{\mathrm{p}}_{\mathrm{LWF}}(G; \mathbf{X})$, hence $P^* \notin \mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$ by (3.5).

Example 4.1. The CG $G := a \rightarrow b — c — d$ has one flag. By applying (3.1), it can be shown that $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$ is determined by the conditions $(a, b) \perp\!\!\!\perp d \mid c$ and $a \perp\!\!\!\perp (c, d) \mid b$ and $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$ by the conditions $(a, b) \perp\!\!\!\perp d \mid c$ and $a \perp\!\!\!\perp (c, d)$.

Example 4.2. The CG G in Figure 2.7a consisting of the 3-biflag $[a; b, d, c]$ has two flags. By applying (3.1), it can be shown that $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$ is determined by the conditions $b \perp\!\!\!\perp c \mid a, d$ and $a \perp\!\!\!\perp (b, c) \mid d$, while $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$ is determined by the conditions $b \perp\!\!\!\perp c \mid a, d$ and $a \perp\!\!\!\perp (b, c)$. In fact, by Theorem 6.2(i), there exists *no* CG $\tilde{G}$ such that $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(\tilde{G}; \mathbf{X})$, but there does exist a CG $\tilde{G}$ such that $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X}) = \mathcal{P}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X})$. (Take $\tilde{G} = G^{\vee}$.)

Remark 4.1. If D is an ADG and D^* the associated *essential graph* that represents the ADG Markov equivalence class $[D]$ (cf. Section 7), then D^* is a CG that has no flags ([AMP] (1997b), Theorem 4.1). Thus by Theorem 4.3, $\mathcal{P}_{\mathrm{AMP}}(D^*; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(D^*; \mathbf{X})\ \forall \mathbf{X}$.

If P belongs to either $\mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})$ or $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$ and also admits a probability density function f with respect to some σ-finite product measure on $\mathbf{X}$, then by C1, f admits the recursive factorization (cf. Lauritzen *et al* (1990))

$$f(x) = \prod \big(f(x_\tau | x_{\mathrm{pa}_{\mathcal{D}}(\tau)}) \mid \tau \in \mathcal{T}\big), \ x \in \mathbf{X}. \tag{4.3}$$

If f is positive and $P \in \mathcal{P}_{\mathrm{LWF}}(G;\mathbf{X})$, Frydenberg (1990, Theorem 4.1(iii)) notes that each conditional density in (4.3) factors further into a product of "potentials" (not necessarily densities) indexed by the cliques of $(G_{\mathrm{cl}(\tau)})^{\mathrm{m}}$. If $P \in \mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X})$, however, no such further factorization appears to hold in general. Nonetheless, at least under the assumption of multivariate normality, statistical inference for an AMP model is straightforward, as will be shown in Section 5.

5. The AMP Markov Property for Multivariate Normal Distributions.

In this section we set $\mathbf{X}_v = \mathbf{R}$ for $v \in V$, so $\mathbf{X} = \mathbf{R}^V$. For any CG $G \equiv (V,E)$, define

$$\mathbf{N}_V(0,G) := \mathbf{N}_V(0) \cap \mathcal{P}_{\mathrm{AMP}}(G;\mathbf{R}^V), \tag{5.1}$$

where $\mathbf{N}_V(0) := (\mathcal{N}_V(0,\Sigma) \mid \Sigma \in \mathbf{P}(V))$, $\mathcal{N}_V(0,\Sigma)$ is the normal distribution on $\mathbf{R}^V$ with mean 0 and covariance matrix Σ, and $\mathbf{P}(V)$ is the set of all $V \times V$ real positive definite symmetric matrices. If $X \sim P \in \mathbf{N}_V(0,G)$, (4.3) implies that the joint distribution of $X \equiv (X_v | v \in V)$ is determined by the family of conditional distributions $\big(X_\tau | X_{\mathrm{pa}_{\mathcal{D}}(\tau)} \bigm| \tau \in \mathcal{T}\big)$, where $\mathcal{D} = \mathcal{D}(G)$ and $\mathcal{T} = \mathcal{T}(G)$. By normality, each of these conditional distributions has the form of a multivariate linear regression model:

$$X_\tau | X_{\mathrm{pa}_{\mathcal{D}}(\tau)} \sim \mathcal{N}_\tau(\beta_\tau X_{\mathrm{pa}_{\mathcal{D}}(\tau)}, \Lambda_\tau), \tag{5.2}$$

where Λ_τ is the (nonsingular) $\tau \times \tau$ conditional covariance matrix of X_τ given $X_{\mathrm{pa}_{\mathcal{D}}(\tau)}$ and β_τ is the $\tau \times \mathrm{pa}_{\mathcal{D}}(\tau)$ matrix of regression coefficients.

Conditions C2 and C3* now impose additional restrictions on Λ_τ and β_τ, as follows. By (5.2), C2 is equivalent to the restriction $\Lambda_\tau \in \mathbf{P}(G_\tau)$, the set of all $\tau \times \tau$ positive definite real matrices such that $\mathcal{N}_\tau(0,\Lambda_\tau)$ is global G_τ-Markovian. By normality, this is equivalent to the following condition:

$$u,v \in \tau,\ u,v \text{ not adjacent in } G_\tau \ \Rightarrow\ (\Lambda_\tau^{-1})_{uv} = 0, \tag{5.3}$$

that is, Λ_τ satisfies the *covariance selection* model (Dempster (1972), Lauritzen (1996, Section 5.2)) determined by the UG G_τ. Also by (5.2), C3* is equivalent to the restriction $\beta_\tau \in \mathbf{B}_\tau(G)$, where $\mathbf{B}_\tau(G)$ is the set of all β_τ that satisfy the following condition:

$$u,v \in \tau,\ u,v \text{ not adjacent in } G_\tau \ \Rightarrow\ (\Lambda_\tau^{-1})_{uv} = 0, \tag{5.3}$$

that is, Λ_τ satisfies the *covariance selection* model (Dempster (1972)) determined by the UG G_τ. Also by normality, C3* is equivalent to the restriction $\beta_\tau \in \mathbf{B}_\tau(G)$, where $\mathbf{B}_\tau(G)$ is the set of all β_τ that satisfy the following condition:

$$u \in \tau,\ v \in \mathrm{pa}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(u) \ \Rightarrow\ (\beta_\tau)_{uv} = 0. \tag{5.4}$$

Since this is a linear restriction, the conditional distribution (5.2) retains the form of a multivariate linear regression model,[10] in fact, a generalized *seemingly unrelated regressions (SUR) model*, cf. Zellner (1962).

[10] It is important to note that (5.3) and (5.4) impose constraints on Λ_τ and β_τ *separately.* If, in (5.1), "$\mathcal{P}_{AMP}$" is replaced by "$\mathcal{P}_{LWF}$", then (5.2) and (5.3) remain unchanged, but in (5.4), β_τ must be replaced by $\Lambda_\tau^{-1}\beta_\tau$, which, together with Λ_τ^{-1}, comprise the *natural parameters* for the multivariate normal exponential family.

Thus, the likelihood function (LF) of the normal AMP CG model $\mathbf{N}_V(0,G)$ factors according to (5.2) into the product of the LFs of generalized SUR models with covariance selection restrictions. The parameter space

$$\mathbf{P}(G) := \{\Sigma \in \mathbf{P}(V) \,|\, \mathcal{N}_V(0,\Sigma) \in \mathbf{N}_V(0,G)\}$$

of the model $\mathbf{N}_V(0,G)$ factors into the product of the corresponding parameter spaces according to the bijective mapping

$$\begin{aligned} \mathbf{P}(G) &\to \times\big(\,\mathbf{P}(G_\tau)\times\mathbf{B}_\tau(G)\;\big|\;\tau\in\mathcal{T}\big) \\ \Sigma &\mapsto \quad \big((\Lambda_\tau,\beta_\tau)\;\big|\;\tau\in\mathcal{T}\big). \end{aligned} \tag{5.5}$$

The family of matrices $\big((\Lambda_\tau,\beta_\tau)\;\big|\;\tau\in\mathcal{T}\big)$ is called the family of *G-parameters* of Σ.

The maximum likelihood estimate (MLE) of Σ is obtained by first calculating the MLEs of its G-parameters (requiring iterative methods for generalized SUR and covariance selection models), then using these to reconstruct the MLE of Σ. The estimation of Λ_τ may be somewhat simplified by noting that, by C3*, Λ_τ is also the conditional covariance matrix of X_τ given $X_{\mathrm{pa}_G(\tau)}$.

Remark 5.1. The preceding discussion shows that in general it is the AMP, rather than the LWF, block-recursive Markov property for a CG G that is satisfied by the following block-recursive normal linear system naturally associated with G:

$$X_\tau = \beta_\tau X_{\mathrm{pa}_{\mathcal{D}}(\tau)} + \epsilon_\tau, \quad \tau\in\mathcal{T}. \tag{5.6}$$

Here $\epsilon_\tau \sim \mathcal{N}_\tau(0,\Lambda_\tau)$, $\Lambda_\tau\in\mathbf{P}(G_\tau)$, $\beta_\tau\in\mathbf{B}_\tau(G)$, and the ϵ_τ are mutually independent. The model (5.6) is a CW *concentration regression CG model* ([CW] (1993, 1996)). See Spirtes (1995) and Koster (1996) for related results.

6. Markov Equivalence of Chain Graphs.

An interesting, although complicating, feature of ADG models and CG models is the possible non-uniqueness of the graph associated with the model. Unlike UGs, two or more ADGs or CGs may determine the same Markov model. For example, the three ADGs $a\to c\to b$, $a\leftarrow c\leftarrow b$, and $a\leftarrow c\to b$ each specify one Markov condition: $a \perp\!\!\!\perp b \mid c$. This non-uniqueness can lead to computational inefficiency in model selection and to inappropriate specification of prior distributions in Bayesian model averaging (Madigan *et al* (1996)).

Two CGs $G_1=(V,E_1)$ and $G_2=(V,E_2)$ with the same vertex set V are called *LWF (resp., AMP) Markov equivalent* if $\mathcal{P}_{\mathrm{LWF}}(G_1;\mathbf{X})=\mathcal{P}_{\mathrm{LWF}}(G_2;\mathbf{X})$ (resp., $\mathcal{P}_{\mathrm{AMP}}(G_1;\mathbf{X}) = \mathcal{P}_{\mathrm{AMP}}(G_2;\mathbf{X})$) for every product space $\mathbf{X}$ indexed by V. Theorem 4.3 implies that the two notions of Markov equivalence coincide for UGs and for ADGs. Verma and Pearl (1992) proved that two ADGs are Markov equivalent iff they have the same skeleton and same immoralities. Frydenberg (1990) and [AMP] (1997a) showed that two CGs are LWF Markov equivalent iff they have the same skeleton and same complexes. Theorem 6.1 below shows that two CGs are AMP Markov equivalent iff they have the same skeleton and same triplexes. (The proof is given in Appendix B.) Thus, the condition for AMP Markov equivalence of CGs more closely resembles that for ADG Markov equivalence than does the condition for LWF Markov equivalence of CGs, in the sense that both immoralities and triplexes involve only three vertices, while complexes can involve arbitrarily many vertices. This in turn suggests that determining AMP Markov equivalence may require lesser computational complexity than does LWF Markov equivalence.

Theorem 6.1. Two chain graphs $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ with the same vertex set V are AMP Markov equivalent iff they have the same skeleton and the same triplexes.

[AMP] (1997a) give necessary and sufficient conditions for a CG G to be LWF Markov equivalent to some UG, to some ADG, or to some decomposable ($\equiv$ chordal) UG. Theorem 6.1 can be applied to obtain parallel conditions for the AMP Markov equivalences of G. In particular, if G is determined to be Markov equivalent to an ADG, the analysis of the statistical model determined by G is substantially simplified.

We write $G_1 \overset{\mathrm{A}}{\sim} G_2$ ($G_1 \overset{\mathrm{L}}{\sim} G_2$) to indicate that the CGs G_1 and G_2 are AMP (LWF) Markov equivalent. Proposition 6.1 follows directly from Theorem 6.1 and in turn implies Proposition 6.3. The proofs of Propositions 6.2 and 6.4 are presented in Appendix B. [The parallel results regarding LWF Markov equivalence appear in [AMP] (1997a).]

Proposition 6.1. Let G be a chain graph. Conditions (1), (1′), and (1″) are equivalent:

(1) $G \overset{\mathrm{A}}{\sim}$ some (unique) UG.	$G \overset{\mathrm{L}}{\sim}$ some (unique) UG.
(1′) G has no triplexes.	G has no complexes.
(1″) $G \overset{\mathrm{A}}{\sim} G^{\vee}$.	$G \overset{\mathrm{L}}{\sim} G^{\vee}$.

Proposition 6.2. Let G be a chain graph. Conditions (2) and (2′) are equivalent:

(2) $G \overset{\mathrm{A}}{\sim}$ some ADG.	$G \overset{\mathrm{L}}{\sim}$ some ADG.
(2′) G_τ is chordal $\forall \tau \in \mathcal{T}(G)$ and G has no biflags.	$\left(G_{\mathrm{cl}(\tau)}\right)^{\mathrm{m}}$ is chordal $\forall \tau \in \mathcal{T}(G)$.

Remark 6.1. (An AMP - LWF Duality.) For $a \in V$ and $\tau \in \mathcal{T}(G)$, define $\mathrm{ch}_\tau(a) := \mathrm{ch}(a) \cap \tau$ and $\overline{\mathrm{ch}}_\tau(a) := \tau \setminus \mathrm{ch}_\tau(a)$. The following is an equivalent but more explicit version of condition (2′).

(2″) For every chain component $\tau \in \mathcal{T}(G)$:

(a) G_τ is chordal;

(b) for every $a \in \mathrm{bd}(\tau)$ and every non-adjacent pair $c, d \in \overline{\mathrm{ch}}_\tau(a)$, $\overline{\mathrm{ch}}_\tau(a) \setminus \{c, d\}$ separates c and d in G_τ (in particular, $\overline{\mathrm{ch}}_\tau(a) \setminus \{c, d\}$ must be nonempty); and

(c) for every distinct pair $a, b \in \mathrm{bd}(\tau)$ and every pair $c \in \overline{\mathrm{ch}}_\tau(b) \setminus \overline{\mathrm{ch}}_\tau(a)$, $d \in \overline{\mathrm{ch}}_\tau(a) \setminus \overline{\mathrm{ch}}_\tau(b)$, $[\overline{\mathrm{ch}}_\tau(a) \cup \overline{\mathrm{ch}}_\tau(b)] \setminus \{c, d\}$ separates c and d in G_τ (in particular, $[\overline{\mathrm{ch}}_\tau(a) \cup \overline{\mathrm{ch}}_\tau(b)] \setminus \{c, d\}$ must be nonempty and c, d must be non-adjacent).

Now note that by simply replacing "$\overline{\mathrm{ch}}_\tau$" by "$\mathrm{ch}_\tau$" throughout (b) and (c), ($2''$) becomes the necessary and sufficient condition that $G \overset{\mathrm{L}}{\sim}$ some ADG - see Remark 4.2 of [AMP] (1997a). This implies that

$$G \overset{\mathrm{A}}{\sim} \text{ some ADG} \iff \overline{G} \overset{\mathrm{L}}{\sim} \text{ some ADG},$$

where $\overline{G}$ is the CG derived from G as follows: for each $\tau \in \mathcal{T}(G)$ and each a such that $\mathrm{ch}_\tau(a)$ and $\overline{\mathrm{ch}}_\tau(a)$ are nonempty, reverse the roles of $\mathrm{ch}_\tau(a)$ and $\overline{\mathrm{ch}}_\tau(a)$, that is, delete all arrows from a to the vertices in $\mathrm{ch}_\tau(a)$ and insert arrows from a to all vertices in $\overline{\mathrm{ch}}_\tau(a)$. (Clearly $\mathcal{T}(\overline{G}) = \mathcal{T}(G)$ and $\mathcal{D}(\overline{G}) = \mathcal{D}(G)$, so $\overline{G}$ is adicyclic.) Since $\overline{\overline{G}} = G$, also $G \overset{\mathrm{L}}{\sim}$ some ADG iff $\overline{G} \overset{\mathrm{A}}{\sim}$ some ADG. Finally, note that $G = \overline{G}$ iff G has no flags, which is exactly the condition that the AMP and LWF Markov properties for G coincide (Theorem 4.3).

Proposition 6.3. Let G be a chain graph. Conditions (3), ($3'$), and ($3''$) are equivalent:

(3) $G \overset{\mathrm{A}}{\sim}$ some (unique) chordal UG.	$G \overset{\mathrm{L}}{\sim}$ some (unique) chordal UG.
($3'$) $G^\vee$ is chordal and G has no triplexes.	$G^\vee$ is chordal and G has no complexes.
($3''$) $G^\vee$ is chordal and $G \overset{\mathrm{A}}{\sim} G^\vee$.	$G^\vee$ is chordal and $G \overset{\mathrm{L}}{\sim} G^\vee$.

Proposition 6.4. Let G be a chain graph. Conditions ($3'$) and (4) are equivalent:

($3'$) $G^\vee$ is chordal and G has no triplexes.	$G^\vee$ is chordal and G has no complexes.
(4) G_τ is chordal $\forall \tau \in \mathcal{T}(G)$ and G has no triplexes.	$\left(G_{\mathrm{cl}(\tau)}\right)^{\mathrm{m}}$ is chordal $\forall \tau \in \mathcal{T}(G)$ and G has no immoralities.

Propositions 6.1 - 6.4 immediately yield the following corollary:

Corollary 6.1. Let G be a chain graph. Then $G \overset{\mathrm{A}}{\sim}$ ($\overset{\mathrm{L}}{\sim}$) some UG and $G \overset{\mathrm{A}}{\sim}$ ($\overset{\mathrm{L}}{\sim}$) some ADG if and only if $G \overset{\mathrm{A}}{\sim}$ ($\overset{\mathrm{L}}{\sim}$) some decomposable ($\equiv$ chordal) UG, namely, $G^\vee$.

Remark 6.2. By Proposition 6.1 and Theorem 4.3, if $G \overset{\mathrm{A}}{\sim}$ some UG then G has no flags and $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(G; \mathbf{X})\ \forall \mathbf{X}$, hence $G \overset{\mathrm{L}}{\sim}$ some UG. By Proposition 4.1 of [AMP] (1997a), however, this is not necessarily true if $\overset{\mathrm{A}}{\sim}$ and $\overset{\mathrm{L}}{\sim}$ are interchanged. For example, if G is the flag $a \to b$—c then $G \overset{\mathrm{L}}{\sim} a$—$b$—$c$ but $G \overset{\mathrm{A}}{\not\sim}$ any UG with the same vertex set. Similarly, by Proposition 6.2 above and Proposition 4.2 of [AMP] (1997a), $G \overset{\mathrm{L}}{\sim}$ some ADG need not imply that $G \overset{\mathrm{A}}{\sim}$ some ADG, nor vice versa. For example, if G is the 3-biflag in Figure 2.7a then $G \overset{\mathrm{L}}{\sim}$ the ADG obtained from G by converting d—b to $d \to b$ and d—c to $d \to c$ but $G \overset{\mathrm{A}}{\not\sim}$ any ADG, while $\overline{G} \overset{\mathrm{A}}{\sim}$ the ADG obtained from $\overline{G}$ by these same conversions but $\overline{G} \overset{\mathrm{L}}{\not\sim}$ any ADG. The first three sentences of this Remark remain valid with "UG" replaced by "decomposable UG".

We conclude this section by addressing the following crucial question: does the class of AMP chain graph models actually include statistical models that cannot be represented as LWF chain graph models (and vice versa)? If the answer were "no" then the statistical interest of the AMP Markov property would be lessened, but this is not the case.

By Theorem 4.3, if the CG G has at least one flag then $\mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X}) \neq \mathcal{P}_{\mathrm{LWF}}(G;\mathbf{X})$ for at least one $\mathbf{X}$, but this does not exclude the possibility that there may exist some other CG $\tilde{G}$ such that $\mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(\tilde{G};\mathbf{X})\ \forall\mathbf{X}$. However, it follows from Theorem 6.2(i) below that *the necessary and sufficient condition for the non-existence of such $\tilde{G}$ is that G has at least one biflag.* Thus, for the CG G in Figure 1.1, the AMP conditions in (1.2) do not coincide with the LWF conditions determined by *any* chain graph $\tilde{G}$. Similarly, by Theorem 6.2(ii) the LWF conditions in (1.1) do not coincide with the AMP conditions determined by *any* chain graph $\tilde{G}$. Theorem 6.2 is proved in Appendix B.

Theorem 6.2. Let $G \equiv (V, E)$ be a chain graph.

(i) There exists a CG $\tilde{G} \equiv (V, \tilde{E})$ such that $\mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(\tilde{G};\mathbf{X})\ \forall\mathbf{X}$ if and only if G has no biflags.

(ii) There exists a CG $\tilde{G} \equiv (V, \tilde{E})$ such that $\mathcal{P}_{\mathrm{LWF}}(G;\mathbf{X}) = \mathcal{P}_{\mathrm{AMP}}(\tilde{G};\mathbf{X})\ \forall\mathbf{X}$ only if G has no multicomplexes.

Remark 6.3. The necessary condition in (ii) is *not* sufficient. Figure 6.1 gives examples of chain graphs G that do not have multicomplexes but for which there exists no CG $\tilde{G} \equiv (V, \tilde{E})$ such that $\mathcal{P}_{\mathrm{LWF}}(G;\mathbf{X}) = \mathcal{P}_{\mathrm{AMP}}(\tilde{G};\mathbf{X})\ \forall\mathbf{X}$. (This follows from Theorem 6.2(i).)

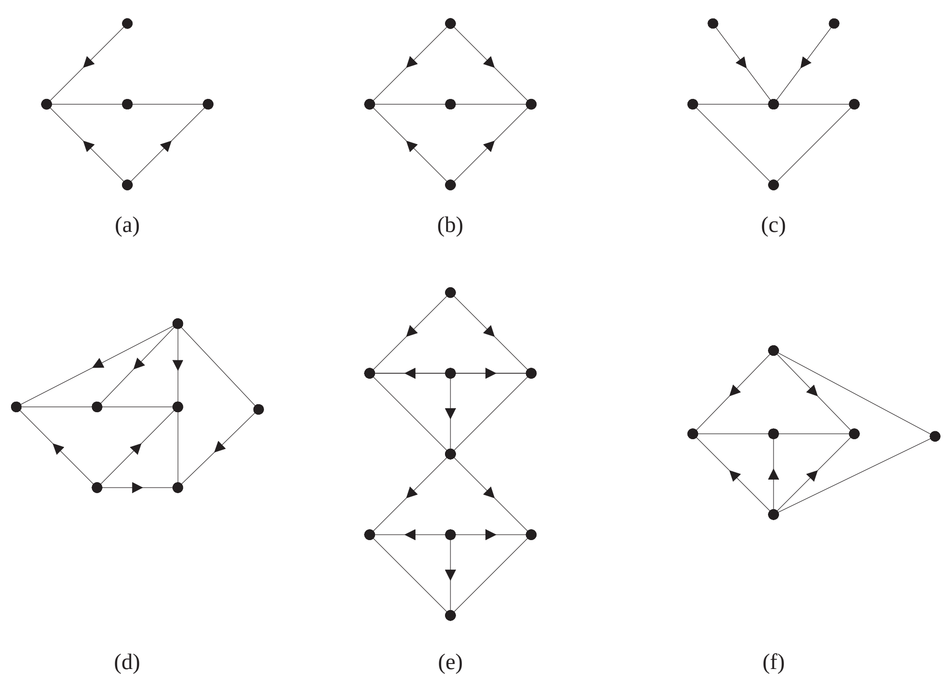


Figure 6.1.

Corollary 6.2. Let $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ be chain graphs with the same vertex set V. Then $\mathcal{P}_{\mathrm{AMP}}(G_1; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(G_2; \mathbf{X})\ \forall \mathbf{X}$ if and only if the following three conditions hold:

(i) $G_1^{\vee} = G_2^{\vee}$;

(ii) G_1 has no biflags;

(iii) $\tilde{G}_1$ and G_2 have the same complexes, where $\tilde{G}_1$ is the CG obtained from G_1 by converting each flag $[a, b; c]$ into an immorality $(a, b; c)$.

Proof. "if": As in the proof of Theorem 6.2(i) in Appendix B, the construction of $\tilde{G}_1$ from G_1 is unambiguous by (ii) and $\mathcal{P}_{\mathrm{AMP}}(G_1; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(\tilde{G}_1; \mathbf{X})\ \forall \mathbf{X}$. By (i) and (iii), $\tilde{G}_1$ and G_2 are LWF Markov equivalent, hence $\mathcal{P}_{\mathrm{AMP}}(G_1; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(G_2; \mathbf{X})\ \forall \mathbf{X}$.

"only if": Conditions (i) and (ii) follow from Proposition B.2 in Appendix B and $\mathcal{P}_{\mathrm{AMP}}(G_1; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(\tilde{G}_1; \mathbf{X})\ \forall \mathbf{X}$ as shown in the proof of Theorem 6.2(i). Thus $\tilde{G}_1$ and G_2 are LWF Markov equivalent, hence have the same complexes.

The relations among the classes of graphical Markov models considered in this section are illustrated in Figure 6.2. The ?? indicate that the sufficient condition for Theorem 6.2(ii) is currently unknown.

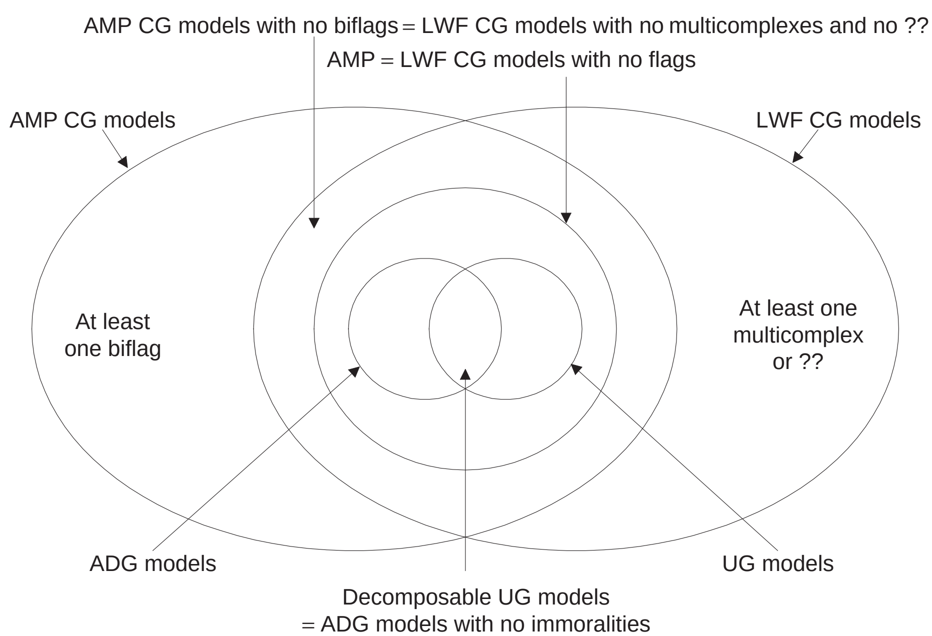


Figure 6.2.

7. Remarks and Extensions.

For an ADG $D \equiv (V, E)$, let $[D]$ denote the ADG Markov equivalence class determined by D, that is, the set of all ADGs $D' \equiv (V, E')$ such that D' is Markov equivalent to D. [AMP] (1997b) have shown that $[D]$ can be uniquely represented by the *essential graph*

$$D^* := \cup\{D' | D' \in [D]\}, \tag{7.1}$$

the smallest graph larger than every $D' \in [D]$. They show that D^* is a chain graph such that $D^* \overset{\text{L}}{\sim} D$ and such that D^* has no flags, hence $D^* \overset{\text{A}}{\sim} D$ by Theorem 4.3. It is readily seen that an arrow $a \to b$ is *essential*, i.e., occurs in the essential graph D^*, iff *it occurs with the same orientation in every* $D' \in [D]$. Clearly both arrows in an immorality of D are essential, but certain other arrows also may be essential – cf. [AMP] (1997b). (Pearl and Verma (1991), Spirtes *et al.* (1993), and Chickering (1995) noted that, under certain additional assumptions, the essential arrows may indicate causal influences. Madigan *et al.* (1996) show that substantial computational efficiencies can be achieved in ADG model selection and Bayesian model averaging by replacing the class of all ADGs by the smaller but equivalent class of essential graphs.

The preceding ideas can be extended from ADG models to AMP chain graph models. The AMP Markov equivalence class $[G]_{AMP}$ for a CG G is uniquely represented by the *essential graph* G^* defined as follows: G^* has the same vertex set and skeleton as G, and an arrow $a \to b$ is *essential*, i.e., occurs in G^*, iff *it occurs with the same orientation in at least one* $G' \in [G]_{AMP}$ *but with the opposite orientation in no* $G'' \in [G]_{AMP}$. It can be shown that if $G \overset{\text{A}}{\sim} D$ for some ADG D, then $G^* = D^*$ as defined in (7.1).

Frydenberg (1990) demonstrated the existence of a unique *largest* ($\equiv$ maximal) CG G_∞ in the LWF Markov equivalence class $[G]_{LWF}$. Trivially,

$$G_\infty = \cup\{G' | G' \in [G]_{LWF}\}. \tag{7.2}$$

Studený (1996, 1997) proposes that G_∞ be used as a unique representative for $[G]_{LWF}$. Since G_∞ is defined with respect to $[G]_{LWF}$ rather than $[G]_{AMP}$, it need not reduce to D^* when $G = D$. For example, if G is the ADG shown in Figure 7.1, then $G^* = G$ but G_∞ replaces the essential arrow $c \to d$ by the line c—d. Similarly, for the second CG G in Figure 7.1, two essential arrows are replaced by lines in G_∞. Thus, if an essential arrow does represent a causal relation, then the AMP essential graph G^* may represent, more completely than G_∞, the set of causal relations determined by G.

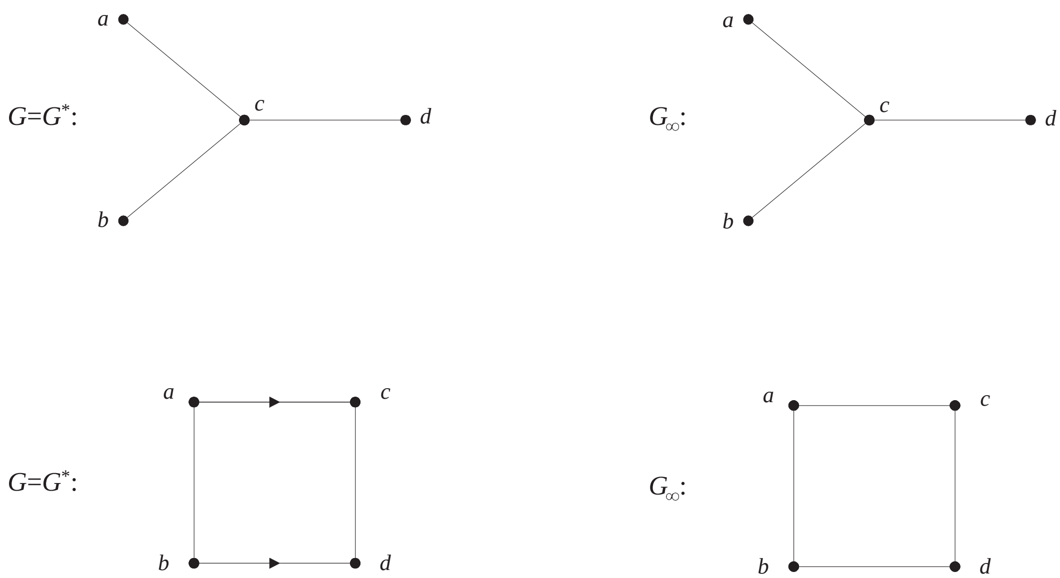


Figure 7.1. Two chain graphs G for which $G = G^* \subset G_\infty$; the first is an ADG.

We remark that, unlike $[G]_{LWF}$, the AMP Markov equivalence class $[G]_{AMP}$ need not contain a unique maximal element, in particular, $\cup\{G'|G' \in [G]_{AMP}\}$ need not be a member of $[G]_{AMP}$. For example, if $G = a \to b — c$ then $[G]_{AMP}$ consists of the three CGs G, $G' := a — b \leftarrow c$ and $G'' := a \to b \leftarrow c \equiv G^*$. Both G and G' are maximal members of $[G]_{AMP}$, while $G \cup G' \cup G'' \equiv G_\infty = a — b — c \notin [G]_{AMP}$. Similarly, $[G]_{AMP}$ need not contain a unique minimal element: if G is an acyclic directed triangle, then G and all other acyclic directed triangles with the same vertices are minimal members of $[G]_{AMP}$.

As exemplified by the CG $G \equiv a \to b — c$ and corresponding essential graph $G^* \equiv a \to b \leftarrow c$ in the preceding paragraph, we see that the essential graph G^* may be *strictly smaller* than G, that is, may contain essential (= causal?) arrows that do not appear in G itself (nor in G_∞).[11] More dramatically, *even if G (and therefore G_∞) is a completely undirected graph, G^* may possess essential (= causal?) arrows.* (See the example in Figure 7.2.) It is of interest to characterize those CGs in which this phenomenon occurs and to investigate its implications (if any) for causal inference.

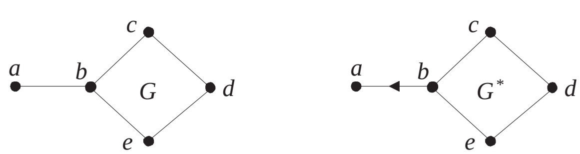


Figure 7.2. A completely undirected chain graph G whose essential graph G^* contains a directed edge $\equiv$ essential arrow.

Like the essential graph D^* for an ADG D, the AMP essential graph G^* for a CG G will play a fundamental role for inference, model selection, and model averaging for

[11] By contrast, if $G = a \to b \to c$ or $a \leftarrow b \to c$, then $G^* = a — b — c = G_\infty$, so G^* can also be *strictly larger* than G.

AMP CGs. For these purposes, in a subsequent paper we shall extend the results of [AMP] (1997b) and Madigan *et al.* (1996) to AMP CGs, in particular extending their characterization of essential graphs for ADGs to essential graphs for AMP CGs. This will lead to polynomial-time algorithms for constructing G^* from G and to irreducible Markov chain Monte Carlo algorithms for model searches over the space of AMP essential graphs.

The standard computational method used to identify valid CIs in ADG models is based on a pathwise separation criterion, called *d-separation*, introduced by Pearl (1988), then elegantly applied by Geiger and Pearl (1988) to establish the completeness of the global Markov property for ADGs (completeness guarantees that *d-separation* uncovers *all* conditional independences implied by the ADG model). Bouckaert and Studený (1995) and Studený and Bouckaert (1998) have generalized this to *c-separation*, a more complicated criterion for identifying valid CIs in LWF CG models, then applied this to establish the completeness of the LWF global Markov property for CGs. Madigan *et al* (1998) have obtained a new pathwise separation criterion for AMP CG models that is simpler than c-separation and more closely resembles the d-separation criterion, due again to the fact that triplexes involve only triples whereas complexes can be of arbitrary length. By Remark 4.1, therefore, the simpler p-separation criterion can be used to determine the Markov properties of the essential graph D^* for any ADG D. Furthermore, p-separation can be used to establish the completeness of the AMP global Markov property for CGs by an argument similar to that of Geiger and Pearl (1988).

For learning and statistical analysis, chain graphs offer considerable expressive power. Under either the LWF or AMP interpretation, CGs can represent many sets of conditional independences that neither ADGs nor UGs alone can represent. As a consequence, CGs encompass many standard statistical model classes (Wermuth and Lauritzen (1990)) and certain neural networks (Buntine (1995)). We speculate that the AMP interpretation will admit simpler Bayesian analysis of CG models than will the LWF interpretation, although for both interpretations the formulation of appropriate hyper-Markov laws (Dawid and Lauritzen (1993)) for non-decomposable models remains problematic.

Appendix A: Proofs of Lemma 3.1, Theorem 3.1, and Theorem 4.1.

Throughout Appendices A and B, the union of pairwise disjoint subsets A, B, C, ... of V often will be denoted by $ABC\cdots$ rather than by $A\dot{\cup}B\dot{\cup}C\cdots$.

Proof of Lemma 3.1. Because $\mathrm{nb}_G(A)$ separates A and $V\setminus\mathrm{cl}_G(A)$ in G for any $A\subset V$, any global G-Markovian P satisfies (3.2). Conversely, suppose that P satisfies (3.2) and that A,B,S are disjoint subsets of V, $A,B\neq\emptyset$, such that S separates A and B in G. Define (see Figure A.1)

$$\begin{aligned} q_A &:= \{v\in V\backslash(ABS)\mid \exists \text{ a path in } G \text{ between } v \text{ and } A \text{ that bypasses } S\},\\ q_B &:= \{v\in V\backslash(ABS)\mid \exists \text{ a path in } G \text{ between } v \text{ and } B \text{ that bypasses } S\},\\ \bar{A} &:= Aq_A, \qquad\qquad \bar{B} := Bq_B, \qquad\qquad q_S := V\backslash(\bar{A}\bar{B}S). \end{aligned}$$

By the assumed separation of A and B by S, $q_A\cap q_B=\emptyset$. Now apply (3.2) to $\bar{A}$ to obtain

$$\bar{A}\perp\!\!\!\perp \big(V\backslash\mathrm{cl}_G(\bar{A})\big)\;\big|\;\mathrm{nb}_G(\bar{A})\,[P]. \tag{A.1}$$

But $\mathrm{nb}_G(\bar{A})\subseteq S$: otherwise, $\mathrm{nb}_G(\bar{A})\cap(\bar{B}q_S)\neq\emptyset$, contradicting the separation of A and B by S or the definition of q_S. Therefore

$$V\backslash\mathrm{cl}_G(\bar{A})=\bar{B}q_S\big(S\backslash\mathrm{nb}_G(\bar{A})\big),$$

hence, by (3.1), (A.1) implies that

$$\bar{A}\perp\!\!\!\perp \bar{B}q_S\mid S\,[P].$$

Thus $A\perp\!\!\!\perp B\mid S\,[P]$, so P is global G-Markovian, as required.

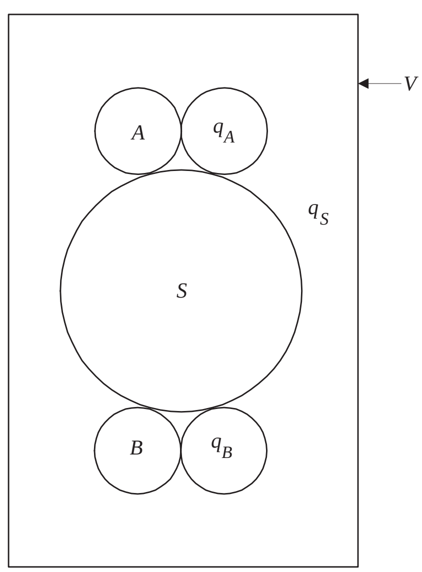


Figure A.1.

Proof of Theorem 3.1. (i) $\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G;\mathbf{X})$: By Proposition 3.1, it suffices to show that if P is LWF global G-Markovian, then P satisfies C4. Thus, we must show that for each $\tau \in \mathcal{T}$ and $\sigma \subseteq \tau$, $\mathrm{bd}_G(\sigma)$ separates σ and $\mathrm{Nd}_{\mathcal{D}}(\tau) \setminus \mathrm{cl}_G(\sigma)$ in H^{m}, where, since $\mathrm{cl}_G(\sigma) = \sigma \dot{\cup} \mathrm{bd}_G(\sigma)$ and $\mathrm{Nd}_{\mathcal{D}}(\tau)$ is G-anterior,

$$H := G\big(\mathrm{Nd}_{\mathcal{D}}(\tau)\big) = G_{\mathrm{Nd}_{\mathcal{D}}(\tau)}.$$

Suppose that π is a path from σ to $\mathrm{Nd}_{\mathcal{D}}(\tau)\setminus\mathrm{cl}_G(\sigma)$ in the UG H^{m}, so that $\pi \subseteq \mathrm{Nd}_{\mathcal{D}}(\tau)$. Let $s \in \sigma \cap \pi$ be the *last* vertex of π that is also in σ, and let $v \in \pi\setminus\sigma$ be the *next* vertex in π after s, so that $s — v \in H^{\mathrm{m}}$. Thus, either (1) $s \leftarrow v \in H$, (2) $s — v \in H$, (3) $s \rightarrow v \in H$, or (4) $s \cdot\!\not\cdot v$ in H but $\exists\, c_1, \ldots, c_k \in \mathrm{Nd}_{\mathcal{D}}(\tau)$ such that $(s, v; c_1, \ldots, c_k)$ is a k-complex in H. In case (1), $v \in \mathrm{pa}_G(\sigma)$, while in case (2), $v \in \mathrm{nb}_G(\sigma)$; in both cases π intersects $\mathrm{bd}_G(\sigma)$, as required. In case (3), $v \in \mathrm{de}_G(\tau) \subseteq \mathrm{de}_{\mathcal{D}}(\tau)$, while in case (4), $c_1 \in \mathrm{de}_G(\tau) \subseteq \mathrm{de}_{\mathcal{D}}(\tau)$, both impossible since $\mathrm{de}_{\mathcal{D}}(\tau) \cap \mathrm{Nd}_{\mathcal{D}}(\tau) = \emptyset$.

(ii) $\mathcal{P}^{\mathrm{b}}_{\mathrm{LWF}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G;\mathbf{X})$: We must show that if P satisfies C4 $\equiv$ C4$'$ w.r.t. G, then P is LWF global G-Markovian. As in Frydenberg (1990, Lemma 3.2), the proof proceeds by induction on $|\mathcal{T}| \equiv |\mathcal{T}(G)|$. If $|\mathcal{T}| = 1$, then G is a (connected) UG and, by Lemma 3.1, C4$'$ implies that P is global G-Markovian, hence is LWF global G-Markovian. Now assume that the result is true whenever $|\mathcal{T}| < n$, consider a CG G with $|\mathcal{T}| = n$, and assume that P satisfies C4 $\equiv$ C4$'$ w.r.t. G. We assert that $A \perp\!\!\!\perp B \mid S\,[P]$ if S separates $A \neq \emptyset$ and $B \neq \emptyset$ in H^{m}, where

$$H := G(ABS) = G_{\mathrm{At}(ABS)}.$$

Let τ be a terminal chain component of G, so that $\mathrm{nd}_{\mathcal{D}}(\tau) = V \setminus \tau$. If $(ABS) \cap \tau = \emptyset$, the assertion follows from the induction hypothesis applied to $G' := G_{V\setminus\tau}$: verify that $G'(ABS) = G(ABS) \equiv H$, so S separates A and B in $G'(ABS)^{\mathrm{m}}$, and that the marginal distribution $P_{V\setminus\tau}$ satisfies C4$'$ w.r.t. G'. The latter follows from C4$'$ for G, from the identities $\mathcal{T}(G') = \mathcal{T}(G) \setminus \tau$, $\mathcal{D}(G') = (\mathcal{D}(G))_{\mathcal{T}(G)\setminus\tau}$, and

$$\begin{aligned}
\mathrm{nd}_{\mathcal{D}(G')}(\tau') &= \mathrm{nd}_{\mathcal{D}(G)}(\tau') \setminus \tau, && \forall \tau' \in \mathcal{T}(G'),\\
\mathrm{pa}_{G'}(\sigma) &= \mathrm{pa}_G(\sigma), && \forall \tau' \in \mathcal{T}(G'), \forall \sigma \subseteq \tau',\\
\mathrm{nb}_{G'}(\sigma) &= \mathrm{nb}_G(\sigma), && \forall \tau' \in \mathcal{T}(G'), \forall \sigma \subseteq \tau',\\
G'_{\tau'} &= G_{\tau'}, && \forall \tau' \in \mathcal{T}(G'),
\end{aligned} \tag{A.2}$$

the first two of which follow from the assumption that τ is terminal in G.

If $(ABS) \cap \tau \neq \emptyset$, define (see Figure A.2):

$$
\begin{array}{lll}
A_1 := A \setminus \tau, & A_1' := A_1 \cap \mathrm{pa}_G(\tau), & A_1'' := A_1 \backslash A_1', \\
B_1 := B \setminus \tau, & B_1' := B_1 \cap \mathrm{pa}_G(\tau), & B_1'' := B_1 \backslash B_1', \\
S_1 := S \setminus \tau, & S_1' := S_1 \cap \mathrm{pa}_G(\tau), & S_1'' := S_1 \backslash S_1', \\
\multicolumn{3}{l}{\;\; p := \mathrm{pa}_G(\tau) \backslash (A_1 B_1 S_1) \subseteq V \backslash (\tau A_1 B_1 S_1),} \\
\multicolumn{3}{l}{p_A := \{v \in p \mid \exists \text{ a path in } H^{\mathrm{m}} \text{ between } v \text{ and } A \text{ that bypasses } S\},} \\
\multicolumn{3}{l}{p_B := \{v \in p \mid \exists \text{ a path in } H^{\mathrm{m}} \text{ between } v \text{ and } B \text{ that bypasses } S\},} \\
\bar{A}_1 := A_1 p_A, & \bar{B}_1 := B_1 p_B, & p_S := p \backslash (p_A p_B), \\
A_2 := A \cap \tau, & B_2 := B \cap \tau, & S_2 := S \cap \tau, \\
\multicolumn{3}{l}{q_A := \{v \in \tau \backslash (A_2 B_2 S_2) \mid \exists \text{ a path in } H^{\mathrm{m}} \text{ between } v \text{ and } A \text{ that bypasses } S\},} \\
\multicolumn{3}{l}{q_B := \{v \in \tau \backslash (A_2 B_2 S_2) \mid \exists \text{ a path in } H^{\mathrm{m}} \text{ between } v \text{ and } B \text{ that bypasses } S\}} \\
\bar{A}_2 := A_2 q_A, & \bar{B}_2 := B_2 q_B, & q_S := \tau \backslash (\bar{A}_2 \bar{B}_2 S_2).
\end{array}
$$

Since S separates A and B in H^{m}, $p_A \cap p_B = q_A \cap q_B = \emptyset$. Furthermore, among the 15 possible pairs of sets formed from the 6 sets $\bar{A}_1, \bar{B}_1, p_S, \bar{A}_2, \bar{B}_2, q_S$, all except the 3 pairs $(\bar{A}_1, \bar{A}_2)$, $(\bar{B}_1, \bar{B}_2)$, and (p_S, q_S) must be separated by S in H^{m}. Since $\mathrm{pa}_G(\tau)$ is complete in H^{m}, it follows that *at most one of* $A_1' p_A, B_1' p_B$, *and* p_S *is nonempty.*

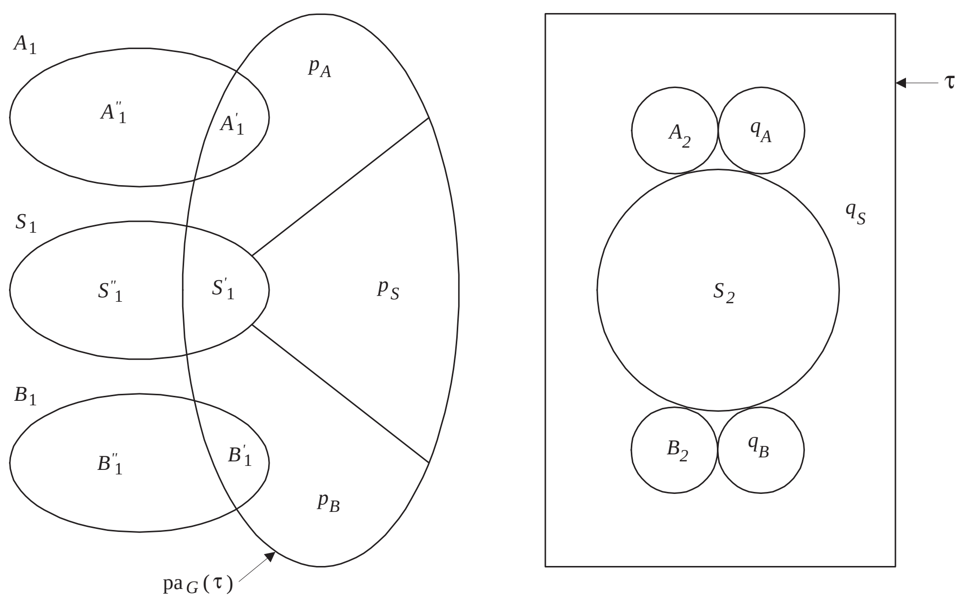


Figure A.2.

Consider first the case $B_1' p_B = p_S = \emptyset$. We shall establish the following three conditional independences (omitting "$[P]$" throughout):

(A.3) $$\bar{A}_1 \perp\!\!\!\perp B_1'' \mid S_1,$$

(A.4) $$\bar{A}_2 \bar{B}_2 q_S S_2 \perp\!\!\!\perp B_1'' \mid \bar{A}_1 S_1,$$

(A.5) $$\bar{B}_2 q_S \perp\!\!\!\perp \bar{A}_1 B_1'' \bar{A}_2 \mid S_1 S_2.$$

Then (A.3) and (A.4) yield

$$\bar{A}_1\bar{A}_2\bar{B}_2 q_S S_2 \perp\!\!\!\perp B_1'' \mid S_1$$

via (3.1), which in turn implies

$$\bar{A}_1\bar{A}_2 \perp\!\!\!\perp B_1'' \mid S_1 S_2. \tag{A.6}$$

But (A.5) implies

$$\bar{A}_1\bar{A}_2 \perp\!\!\!\perp \bar{B}_2 q_S \mid B_1'' S_1 S_2,$$

which combines with (A.6) via (3.1) to yield

$$\bar{A}_1\bar{A}_2 \perp\!\!\!\perp B_1''\bar{B}_2 q_S \mid S_1 S_2. \tag{A.7}$$

Since $B_1' = \emptyset$, $A \perp\!\!\!\perp B \mid S$ as required.

Because $P_{V\setminus\tau}$ satisfies C4 w.r.t. $G' := G_{V\setminus\tau}$, the relation (A.3) will follow from the induction hypothesis applied to G' if it can be shown that (*) S_1 separates $\bar{A}_1$ and B_1'' in $G'(\bar{A}_1 B_1'' S_1)^{\mathrm{m}}$. Because τ is terminal in G and $p_A \subseteq p \subseteq \mathrm{At}_G(ABS)$,

$$G'(\bar{A}_1 B_1'' S_1)^{\mathrm{m}} = G(\bar{A}_1 B_1'' S_1)^{\mathrm{m}} \subseteq H^{\mathrm{m}}. \tag{A.8}$$

Since S separates $\bar{A}_1$ and B_1'' in H^{m} and $\mathrm{At}_G(\bar{A}_1 B_1'' S_1) \cap S_2 = \emptyset$, (*) follows from (A.8).

Next, since $\tau = \bar{A}_2\bar{B}_2 q_S S_2$, $\mathrm{nd}_{\mathcal{D}}(\tau)\setminus\mathrm{pa}_G(\tau) \supseteq A_1'' B_1'' S_1''$, and $\mathrm{pa}_G(\tau) = A_1' p_A S_1'$, apply C4$'$ with $\sigma = \tau$ to obtain

$$\bar{A}_2\bar{B}_2 q_S S_2 \perp\!\!\!\perp A_1'' B_1'' S_1'' \mid A_1' p_A S_1',$$

which yields (A.4). Similarly, since $\mathrm{pa}_G(\bar{B}_2 q_S) \subseteq S_1'$ and $\mathrm{nb}_G(\bar{B}_2 q_S) \subseteq S_2$, apply C4$'$ with $\sigma = \bar{B}_2 q_S$ to obtain

$$\bar{B}_2 q_S \perp\!\!\!\perp \bar{A}_1 B_1'' S_1'' \big(S_1'\setminus\mathrm{pa}_G(\bar{B}_2 q_S)\big)\bar{A}_2\big(S_2\setminus\mathrm{nb}_G(\bar{B}_2 q_S)\big) \mid \mathrm{pa}_G(\bar{B}_2 q_S)\dot{\cup}\mathrm{nb}_G(\bar{B}_2 q_S),$$

which yields (A.5).

The proof for the second case $A_1' p_A = p_S = \emptyset$ is obtained from that for the first case by interchanging "A" and "B". The proof for the third case $A_1' p_A = B_1' p_B = \emptyset$ is obtained from the first case by everywhere omitting q_S and replacing $\bar{A}_1$ by $A_1'' p_S$ and $A_1' p_A$ by p_S, noting that $\mathrm{pa}_G(\tau) = p_S S_1'$ and $p_S = p \subseteq \mathrm{At}_G(ABS)$, eventually yielding the relation

$$A_1'' p_S \bar{A}_2 \perp\!\!\!\perp B_1''\bar{B}_2 \mid S_1 S_2 \tag{A.9}$$

in place of (A.7). Since $A_1' = B_1' = \emptyset$, $A \perp\!\!\!\perp B \mid S$ as required.

Proof of Theorem 4.1. (i) $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X})$: By Lemma 4.1, it suffices to show that if P is AMP global G-Markovian, then P satisfies C7 and C8.

Clearly C7 is trivial when $\sigma = \tau$. To establish C7 when $\sigma \subset \tau$, we must show that $\mathrm{pa}_G(\tau)\dot\cup\mathrm{nb}_G(\sigma)$ separates σ and $\tau\backslash\mathrm{cl}_{G_\tau}(\sigma)$ in H^{a}, where, since $\mathrm{An}_G(\mathrm{cl}_G(\tau)) = \mathrm{An}_G(\tau)$,

$$\begin{aligned} H :=& G[\sigma\dot\cup\mathrm{pa}_G(\tau)\dot\cup\mathrm{nb}_G(\sigma)\dot\cup(\tau\backslash\mathrm{cl}_{G_\tau}(\sigma))] \\ \equiv& G[\mathrm{cl}_G(\tau)] \\ =& G[\tau] \\ =& G_{\mathrm{An}_G(\tau)} \cup G^{\wedge}_{\mathrm{Co}_G(\mathrm{An}_G(\tau))}. \end{aligned}$$

Suppose that π is a path between σ and $\tau\backslash\mathrm{cl}_{G_\tau}(\sigma)$ in H^{a}. Let $s \in \tau\cap\pi$ be the *last* vertex of π that is also in σ and let $v \in \pi\backslash\sigma$ be the *next* vertex in π after s, so that s—$v \in H^{\mathrm{a}}$. This implies that either (1) $s \leftarrow v \in H$, (2) s—$v \in H$, (3) $s \rightarrow v \in H$, or (4) $s \cdot\!\!\!/\!\!\cdot v$ in H but either $(\{s,v\},w)$ is a triplex in H for some $w \in \mathrm{An}_G(\tau)$ or $[s,v;w,y]$ is a 2-biflag in H for some $w,y \in \mathrm{An}_G(\tau)$. In case (1), $v \in \mathrm{pa}_G(\sigma) \subseteq \mathrm{pa}_G(\tau)$, while in case (2), $v \in \mathrm{nb}_G(\sigma)$; in both cases the required separation holds. In case (3), $v \in \mathrm{An}_G(\tau)\cap\mathrm{de}_G(\tau)$, which contradicts the adicyclicity of G since G_τ is connected. In case (4), one of the four configurations in Figure A.3 must occur as an induced subgraph of H. In the first, second, and fourth configurations, $w \in \mathrm{An}_G(\tau)\cap\mathrm{de}_G(\tau)$, again contradicting adicyclicity. In the third, s—$w \in H \Rightarrow w \in \tau$, hence $v \in \mathrm{pa}_G(\tau)$ and the required separation again holds.

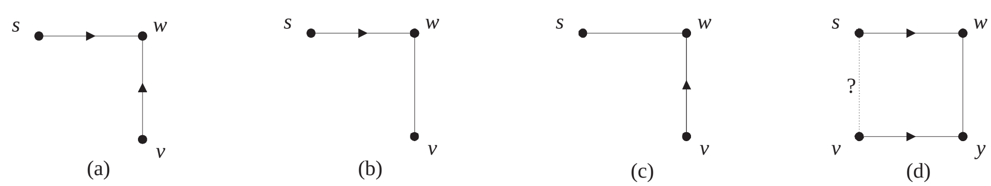


Figure A.3.

For C8, it suffices to show that for each $\tau \in \mathcal{T}$ and $\sigma \subseteq \tau$, $\mathrm{pa}_G(\sigma)$ separates σ and $\mathrm{nd}_{\mathcal{D}}(\tau)\backslash\mathrm{pa}_G(\sigma)$ in H^{a}, where $H := G[\sigma\dot\cup\mathrm{nd}_{\mathcal{D}}(\tau)]$. Since $\mathrm{nd}_{\mathcal{D}}(\tau)$ is G-ancestral, G-coherent, and contains $\mathrm{an}_G(\sigma)$,

$$H = G_{\sigma\dot\cup\mathrm{nd}_{\mathcal{D}}(\tau)} \cup G^{\wedge}_{\mathrm{Nd}_{\mathcal{D}}(\tau)}.$$

Suppose that π is a path from σ to $\mathrm{nd}_{\mathcal{D}}(\tau)\setminus\mathrm{pa}_G(\sigma)$ in H^{a}, so that $\pi \subseteq \mathrm{Nd}_{\mathcal{D}}(\tau)$. Let $s \in \tau\cap\pi$ be the *last* vertex of π that is also in τ, and let $v \in \pi\backslash\tau$ be the *next* vertex in π after s, so that s—$v \in H^{\mathrm{a}}$. However, s—$v \notin H$ since s and v lie in different chain components of G and therefore of H, while $s \rightarrow v \notin H$ since $v \in \pi \subseteq \mathrm{Nd}_{\mathcal{D}}(\tau)$. Thus, either (1) $s \leftarrow v \in H$, or (2) $s \cdot\!\!\!/\!\!\cdot v$ in H but either $(\{s,v\},w)$ is a triplex in H for some $w \in \mathrm{Nd}_{\mathcal{D}}(\tau)$ or $[s,v;w,y]$ is a 2-biflag in H for some $w,y \in \mathrm{Nd}_{\mathcal{D}}(\tau)$. In case (1), $s \in (\sigma\dot\cup\mathrm{nd}_{\mathcal{D}}(\tau))\cap\tau \equiv \sigma$ by the definition of H, so $v \in \mathrm{pa}_G(\sigma)\cap\pi$ and the required separation holds. In case (2), one of the four configurations in Figure A.3 must occur as an induced subgraph of H. The first, second, and fourth configurations are impossible since $w \in \mathrm{Nd}_{\mathcal{D}}(\tau)$. In the third configuration, $v \rightarrow w \in H \Rightarrow w \in \sigma\dot\cup\mathrm{nd}_{\mathcal{D}}(\tau)$ and s—$w \in H \Rightarrow w \in \tau$, hence $w \in \sigma$, so $v \in \mathrm{pa}_G(\sigma)$ and the required separation again holds.

(ii) $\mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G;\mathbf{X}) \subseteq \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X})$: We must show that if P satisfies C7 and C8, then P is AMP global G-Markovian. As in the proof of Theorem 3.1(ii), the proof proceeds by induction on $|\mathcal{T}| \equiv |\mathcal{T}(G)|$. If $|\mathcal{T}| = 1$, then G is a (connected) UG and C7 states that P is global G-Markovian, hence is AMP global G-Markovian. Now assume that the result is true whenever $|\mathcal{T}| < n$, consider a CG G with $|\mathcal{T}| = n$, and assume that P satisfies C7 and (8) w.r.t. G on $\mathbf{X}$. We assert that $A \perp\!\!\!\perp B \mid S\,[P]$ if S separates $A \neq \emptyset$ and $B \neq \emptyset$ in H^{a}, where

$$H := G[ABS] = G_{\mathrm{An}(ABS)} \cup G^{\wedge}_{\mathrm{Co}(\mathrm{An}(ABS))}.$$

Let τ be a terminal chain component of G, so that $\mathrm{nd}_{\mathcal{D}}(\tau) = V \setminus \tau$. If $(ABS) \cap \tau = \emptyset$, the assertion follows from the induction hypothesis applied to $G' := G_{V\setminus\tau}$: verify that $G'[ABS] = G[ABS] \equiv H$, so S separates A and B in $G'[ABS]^{\mathrm{a}}$, and that the marginal distribution $P_{V\setminus\tau}$ satisfies C7 and C8 w.r.t. G'. The latter follows from C7 and C8 for G, from the identities $\mathcal{T}(G') = \mathcal{T}(G) \setminus \tau$ and $\mathcal{D}(G') = (\mathcal{D}(G))_{\mathcal{T}(G)\setminus\tau}$, and from (A.2).

If $(ABS) \cap \tau \neq \emptyset$, define (see Figure A.4):

$$\begin{aligned}
&A_1 := A \setminus \tau, && B_1 := B \setminus \tau, && S_1 := S \setminus \tau,\\
&A_2 := A \cap \tau, && B_2 := B \cap \tau, && S_2 := S \cap \tau,\\
&p := \mathrm{pa}_G(A_2B_2S_2) \setminus (A_1B_1S_1) && && \Omega := \mathrm{pa}_G(\tau) \setminus (pA_1B_1S_1),\\
&p_A := \{v \in p \mid \exists \text{ a path in } H^{\mathrm{a}} \text{ between } v \text{ and } A \text{ that bypasses } S\},\\
&p_B := \{v \in p \mid \exists \text{ a path in } H^{\mathrm{a}} \text{ between } v \text{ and } B \text{ that bypasses } S\},\\
&\bar{A}_1 := A_1p_A, && \bar{B}_1 := B_1p_B, && p_S := p\setminus(p_Ap_B),\\
&q_A := \{v \in \tau\setminus(A_2B_2S_2) \mid \exists \text{ a path in } H^{\mathrm{a}} \text{ between } v \text{ and } A \text{ that bypasses } S\},\\
&q_B := \{v \in \tau\setminus(A_2B_2S_2) \mid \exists \text{ a path in } H^{\mathrm{a}} \text{ between } v \text{ and } B \text{ that bypasses } S\}\\
&\bar{A}_2 := A_2q_A, && \bar{B}_2 := B_2q_B, && q_S := \tau\setminus(\bar{A}_2\bar{B}_2S_2),\\
&c_A := \mathrm{ch}_G(\bar{A}_1) \cap S_2, && c_B := \mathrm{ch}_G(\bar{B}_1) \cap S_2, && c_S := \mathrm{ch}_G(p_S) \cap S_2,\\
&\Delta := S_2\setminus(c_Ac_Bc_S).
\end{aligned}$$

Since S separates A and B in H^{a}, $p \subseteq \mathrm{An}_G(ABS)$, and $q_Aq_Bq_S \subseteq \tau \subseteq \mathrm{Co}_G(\mathrm{An}_G(ABS))$, necessarily $p_A \cap p_B = q_A \cap q_B = \emptyset$ and, among the 15 possible pairs of sets formed from the 6 sets $\bar{A}_1, \bar{B}_1, p_S, \bar{A}_2, \bar{B}_2, q_S$, all except the 3 pairs $(\bar{A}_1, \bar{A}_2)$, $(\bar{B}_1, \bar{B}_2)$, and (p_S, q_S) must be separated by S in H^{a}. It follows from this and from the definitions of H and H^{a} that c_A, c_B, and c_S are also pairwise disjoint and that

$$\mathrm{pa}_G(A_2c_A) \subseteq \bar{A}_1S_1, \tag{A.10}$$

$$\mathrm{pa}_G(B_2c_Bc_S) \subseteq \bar{B}_1S_1p_S, \tag{A.11}$$

$$\mathrm{pa}_G(\Delta) \subseteq S_1, \tag{A.12}$$

$$\mathrm{nb}_G(\bar{A}_2c_A) \subseteq \Delta\ . \tag{A.13}$$

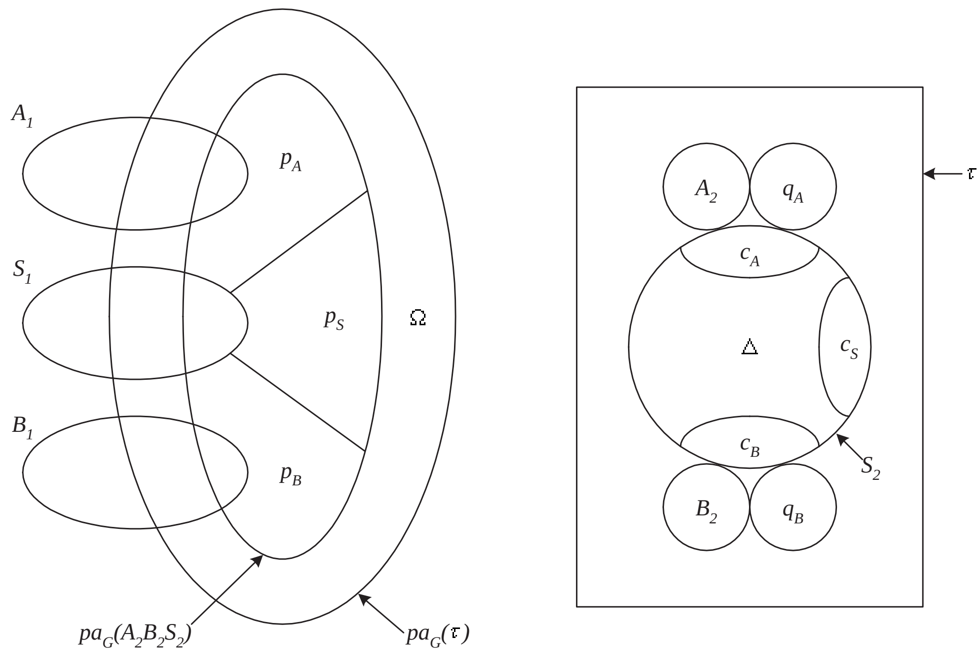


Figure A.4.

We proceed to establish the following four conditional independences (omitting "[P]" throughout):

$$\bar{A}_1 \perp\!\!\!\perp \bar{B}_1 p_S \mid S_1, \tag{A.14}$$

$$B_2 c_B c_S \Delta \perp\!\!\!\perp \bar{A}_1 \Omega \mid \bar{B}_1 S_1 p_S, \tag{A.15}$$

$$A_2 c_A \Delta \perp\!\!\!\perp \bar{B}_1 p_S \Omega \mid \bar{A}_1 S_1, \tag{A.16}$$

$$A_2 c_A \perp\!\!\!\perp \bar{B}_2 c_B c_S q_S \Omega \mid \bar{A}_1 \bar{B}_1 S_1 p_S \Delta\ . \tag{A.17}$$

As with (A.3), the relation (A.14) follows from the induction hypothesis applied to the CG $G' := G_{V \setminus \tau}$ since the marginal distribution $P_{V\setminus\tau}$ satisfies C7 and C8 with respect to G' on $\mathbf{X}_{V\setminus\tau}$, and S_1 separates $\bar{A}_1$ and $\bar{B}_1 p_S$ in $G'[\bar{A}_1 \bar{B}_1 S_1 p_S]^{\mathrm{a}}$. This separation is seen as follows: Because τ is terminal and $p_A p_B p_S \equiv p \subseteq \mathrm{An}_G(ABS)$,

$$G'[\bar{A}_1 \bar{B}_1 S_1 p_S]^{\mathrm{a}} = G[\bar{A}_1 \bar{B}_1 S_1 p_S]^{\mathrm{a}} \subseteq H^{\mathrm{a}}. \tag{A.18}$$

Because S separates $\bar{A}_1$ and $\bar{B}_1 p_S$ in H^{a} and $\mathrm{Co}_G(\mathrm{An}_G(\bar{A}_1 \bar{B}_1 S_1 p_S)) \cap S_2 = \emptyset$, the asserted separation follows from (A.18).

Next, apply C8 with $\sigma = B_2 c_B c_S \Delta$ and then apply (A.11) and (A.12) to obtain

$$B_2 c_B c_S \Delta \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \bar{B}_1 S_1 p_S\big) \mid \bar{B}_1 S_1 p_S.$$

Since $\mathrm{nd}_{\mathcal{D}}(\tau) \setminus (\bar{B}_1 S_1 p_S) \supseteq \bar{A}_1 \Omega$, (A.15) follows by (3.1).

Similarly, apply C8 with $\sigma = A_2 c_A \Delta$ and then apply (A.10) and (A.12) to obtain

$$A_2 c_A \Delta \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus (\bar{A}_1 S_1)\big) \mid \bar{A}_1 S_1.$$

Since $\mathrm{nd}_{\mathcal{D}}(\tau)\backslash(\bar{A}_1 S_1) \supseteq \bar{B}_1 p_S \Omega$, (A.16) follows by (3.1).

To derive (A.17), apply C7 with $\sigma = \bar{A}_2 c_A$, then apply (A.13) and (3.1) to obtain

$$\bar{A}_2 c_A \perp\!\!\!\perp \bar{B}_2 c_B c_S q_S \mid \mathrm{pa}_G(\tau)\Delta. \tag{A.19}$$

Apply C8 with $\sigma = \tau \equiv \bar{A}_2 c_A \bar{B}_2 c_B c_S q_S \Delta$, then use $\mathrm{nd}_{\mathcal{D}}(\tau) \supseteq A_1 B_1 S_1$ and (3.1) to obtain

$$\bar{A}_2 c_A \perp\!\!\!\perp \big((A_1 B_1 S_1)\backslash \mathrm{pa}_G(\tau)\big) \mid \bar{B}_2 c_B c_S q_S \mathrm{pa}_G(\tau)\Delta.$$

This combines with (A.19) via (3.1) to yield

$$\bar{A}_2 c_A \perp\!\!\!\perp \big((A_1 B_1 S_1)\backslash \mathrm{pa}_G(\tau)\big)\bar{B}_2 c_B c_S q_S \mid \mathrm{pa}_G(\tau)\Delta,$$

which in turn yields

$$A_2 c_A \perp\!\!\!\perp \bar{B}_2 c_B c_S q_S \mid \bar{A}_1 \bar{B}_1 S_1 p_S \Omega\Delta. \tag{A.20}$$

Next, (A.16) implies that

$$A_2 c_A \perp\!\!\!\perp \Omega \bar{|} \bar{A}_1 \bar{B}_1 S_1 p_S \Delta,$$

which combines with (A.20) to yield (A.17).

To complete the proof of (ii), delete Ω from (A.15), then combine with (A.14) via (3.1) to obtain

$$\bar{A}_1 \perp\!\!\!\perp \bar{B}_1 B_2 c_B c_S p_S \Delta \mid S_1,$$

which in turn yields

$$\bar{A}_1 \perp\!\!\!\perp \bar{B}_1 B_2 p_S \mid S_1 c_B c_S \Delta\ . \tag{A.21}$$

Next, (A.16) implies

$$A_2 c_A \perp\!\!\!\perp \bar{B}_1 p_S \Omega \mid \bar{A}_1 S_1 \Delta,$$

which combines with (A.17) to yield

$$A_2 c_A \perp\!\!\!\perp \bar{B}_1 \bar{B}_2 c_B c_S p_S q_S \Omega \mid \bar{A}_1 S_1 \Delta$$

and therefore

$$A_2 c_A \perp\!\!\!\perp \bar{B}_1 B_2 p_S \mid \bar{A}_1 S_1 c_B c_S \Delta\ . \tag{A.22}$$

Finally, (A.21) and (A.22) combine to yield

$$\bar{A}_1 A_2 c_A \perp\!\!\!\perp \bar{B}_1 B_2 p_S \mid S_1 c_B c_S \Delta,$$

hence

$$\bar{A}_1 A_2 \perp\!\!\!\perp \bar{B}_1 B_2 p_S \mid S_1 c_A c_B c_S \Delta\ . \tag{A.23}$$

Since $\bar{A}_1 A_2 \supseteq A$, $\bar{B}_1 B_2 \supseteq B$, and $S_1 c_A c_B c_S \Delta = S_1 S_2 = S$, (A.23) implies that $A \perp\!\!\!\perp B \mid S$, as required.

Appendix B: Proofs of Theorem 6.1, Propositions 6.2 and 6.4, Theorem 6.2.

By Lemmas B.1 and B.2 below, it will suffice to prove Theorem 6.1 in the special case where $G_1 \supseteq G_2$. This case is then established by means of Proposition B.1 and Theorem 4.1.

Let $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ be chain graphs with the same vertex set and same skeleton, that is, $G_1^\vee = G_2^\vee$. Construct G_{12} from G_1 by converting a line $a—b \in G_1$ into an arrow $a \to b \in G_{12}$ whenever $a \to b \in G_2$, and similarly construct G_{21} from G_2. Clearly $G_{12}^\vee = G_{21}^\vee = G_1^\vee = G_2^\vee$, $G_1 \supseteq G_{12}$, and $G_2 \supseteq G_{21}$.

We write $G_1 \approx G_2$ if G_1 and G_2 have the same vertex set, the same skeleton, and the same lines, that is, if both $G_1^\vee = G_2^\vee$ and $G_1^\wedge = G_2^\wedge$.

Lemma B.1. Let $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ be chain graphs such that $G_1^\vee = G_2^\vee$.

(i) G_{12} and G_{21} are chain graphs such that $G_{12} \approx G_{21}$.

(ii) If G_1 and G_2 have the same triplexes, then G_{12} and G_{21} each have the same triplexes as G_1 and G_2.

Proof. (i). To see that G_{12} is a CG, suppose to the contrary that G_{12} contains a semi-directed cycle $(v_0, v_1, \ldots, v_n \equiv v_0)$. Since $G_1 \supseteq G_{12}$ and G_1 is adicyclic, each edge $v_{i-1}\cdots v_i \in G_1$ must be a line $v_{i-1}—v_i$. Thus, by the definition of G_{12}, each arrow $v_{i-1} \to v_i$ (resp., line $v_{i-1}—v_i$) that occurs in G_{12} must also occur in G_2, contradicting the adicyclicity of G_2. Similarly, G_{21} is also a CG.

To see that $G_{12} \approx G_{21}$, note that if $a—b \in G_{12} \subseteq G_1$ then $a—b \in G_1$, so $a—b \in G_2$ by the definition of G_{12}, hence $a—b \in G_{21}$ by the definition of G_{21}. Similarly, $a—b \in G_{21} \;\Rightarrow\; a—b \in G_{12}$.

(ii) If the immorality $a \to c \leftarrow b$ occurs in G_1, then, since $G_1 \supseteq G_{12}$, it must also occur in G_{12}. If the flag $a \to c—b$ occurs in G_1, then, since G_1 and G_2 have the same triplexes, either (i) the flag $a \to c—b$ occurs in G_2, in which case it also occurs in G_{12}, or else (ii) either the immorality $a \to c \leftarrow b$ or the flag $a—c \leftarrow b$ occurs in G_2, in which case the immorality $a \to c \leftarrow b$ occurs in G_{12}. Thus each triplex in G_1 also occurs in G_{12}.

Conversely, if the flag $a \to c—b$ occurs in G_{12}, then, since $G_1 \supseteq G_{12}$, either (i) that flag also occurs in G_1, or (ii) $a—c—b$ occurs in G_1. But (ii) is impossible, since it would imply that the flag $a \to c—b$ occurs in G_2, contradicting the assumption that G_1 and G_2 have the same triplexes. If the immorality $a \to c \leftarrow b$ occurs in G_{12}, then either (i) that immorality or one of the flags $a \to c—b$ or $a—c \leftarrow b$ occurs in G_1, or (ii) $a—c—b$ occurs in G_1, which is impossible as before. Thus each triplex in G_{12} also occurs in G_1.

Thus G_1 and G_{12} have the same triplexes, and similarly so do G_2 and G_{21}. This completes the proof.

If, as above, G_1 and G_2 are CGs such that $G_1^\vee = G_2^\vee$, then, following Frydenberg (1990, §5), we define $G_1 \vee G_2$ to be the graph obtained from $G_1 \cup G_2$ by converting to undirected edges ($\equiv$ lines) each directed edge ($\equiv$ arrow) that participates in any semi-directed cycle in $G_1 \cup G_2$. Note that this can be done in a single step: if $a \to b$ occurs in a semi-directed cycle in $G_1 \cup G_2$ and if, after converting $a \to b$ to a line $a—b$, a second arrow $c \to d$ now occurs in a semi-directed cycle, then $c \to d$ must have already occurred in a semi-directed cycle in $G_1 \cup G_2$. Thus $G_1 \vee G_2$ is a chain graph, in fact, *the unique smallest chain graph larger than* G_1 *and* G_2. Clearly $(G_1 \vee G_2)^\vee = (G_1 \cup G_2)^\vee = G_1^\vee = G_2^\vee$ and $G_1 \vee G_2 \supseteq G_1 \cup G_2$.

Lemma B.2. Let G_1 and G_2 be chain graphs such that $G_1 \approx G_2$ and such that G_1 and G_2 have the same triplexes. Then $G_1 \vee G_2$ has the same triplexes as G_1 and G_2.

Proof. First we note two facts:

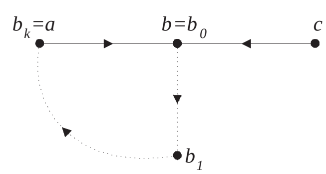
(i) *A line* a—$b \in G_1 \cup G_2$ *iff either:* a—$b \in G_1, G_2$, *or:* $a \to b \in G_1$ *and* $a \leftarrow b \in G_2$ (or vice versa) This follows from the definition of $G_1 \cup G_2$ and the fact that G_1 and G_2 have the same lines.

(ii) G_1, G_2 *and* $G_1 \cup G_2$ *have the same flags and the same immoralities.* If the flag $a \to b$—c (respectively, immorality $a \to b \leftarrow c$) occurs in G_1, then, since G_1 and G_2 have the same lines and the same triplexes, the flag $a \to b$—c (respectively, the immorality $a \to b \leftarrow c$) also occurs in G_2 and therefore in $G_1 \cup G_2$. Conversely, if the flag $a \to b$—c occurs in $G_1 \cup G_2$, then $a \to b \in G_1, G_2$, so by (i) and the fact that G_1 and G_2 have the same triplexes, the flag $a \to b$—c must occur in both G_1 and G_2. Finally, if the immorality $a \to b \leftarrow c$ occurs in $G_1 \cup G_2$, then it must occur in both G_1 and G_2, since $G_1, G_2 \subseteq G_1 \cup G_2$.

To complete the proof of Lemma B.2, by (ii) it suffices to show that $G_1 \vee G_2$ has the same triplexes as $G_1 \cup G_2$. Since $G_1 \cup G_2 \subseteq G_1 \vee G_2$, the following three situations (a), (b), and (c) are the only possible ways in which a triplex might occur in one of these two graphs but not in the other. We shall show that in fact (a), (b), and (c) cannot occur.

(a) *The immorality* $a \to b \leftarrow c$ *occurs in* $G_1 \cup G_2$ *but* a—b—c *occurs in* $G_1 \vee G_2$.

It suffices to show it impossible that an immorality $a \to b \leftarrow c$ occurs in $G_1 \cup G_2$ but a—$b \in G_1 \vee G_2$. For this to occur, $a \to b$ must occur in a semi-directed cycle $(a, b \equiv b_0, b_1, \ldots, b_k \equiv a)$ in $G_1 \cup G_2$, where $k \geq 2$ and where each edge $b_{j-1} \cdots b_j$ $(1 \leq j \leq k)$ in the cycle occurs as either b_{j-1}—b_j or $b_{j-1} \to b_j$ in $G_1 \cup G_2$ (see the following figure). Necessarily $b_1 \neq a, c$; also, we may assume that $b_j \neq c$ for $j = 2, \ldots, k-1$ – otherwise, reverse the roles of a and c. Since G_1 and G_2 are adicyclic, we may consider the *smallest* i, $1 \leq i \leq k$, such that $b_{i-1} \leftarrow b_i \in G_1$ or G_2; say G_1, so necessarily $b_{i-1} \to b_i \in G_2$ by (i).

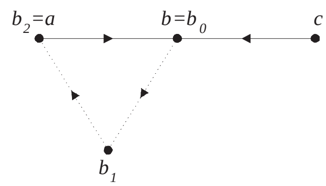


Case 1. Suppose that $i = 1$, so $b \leftarrow b_1 \in G_1$ and $b \to b_1 \in G_2$. Since neither $a \to b \leftarrow b_1$ nor $b_1 \to b \leftarrow c$ can occur as an immorality in G_2, neither can they occur in G_1 (by (ii)), hence there must be edges $a \cdots b_1$ and $c \cdots b_1$ in G_1 and therefore in G_2 and in $G_1 \cup G_2$. To avoid semi-directed 3-cycles in G_2, necessarily $a \to b_1 \in G_2$ and $c \to b_1 \in G_2$, so $a \to b_1 \leftarrow c$ forms an immorality in G_2, hence also in $G_1 \cup G_2$ (by (ii)). (Note that this requires $k \geq 3$, for if $k = 2$ then $a \to b_1 = b_2 \to b_1 \notin G_1 \cup G_2$.) For the immorality $a \to b_1 \leftarrow c$, $a \to b_1$ is part of a *shorter* semi-directed cycle $(a, b_1, \ldots, b_k \equiv a)$ in $G_1 \cup G_2$ that forces a—$b_1 \in G_1 \vee G_2$.

Case 2. Suppose that $i \geq 2$. Thus either $b_{i-2} \to b_{i-1} \leftarrow b_i$ or b_{i-2}—$b_{i-1} \leftarrow b_i$ occurs as a subgraph of G_1, while $b_{i-1} \to b_i \in G_2$. By (ii), necessarily $b_{i-2} \cdots b_i \in G_1$, hence also

$b_{i-2} \cdots b_i \in G_2$. Thus $b_{i-2} \to b_i \in G_2$, since otherwise $(b_{i-2}, b_{i-1}, b_i, b_{i-2})$ would comprise a semi-directed 3-cycle in G_2. Therefore, either $b_{i-2} \to b_i \in G_1 \cup G_2$ or $b_{i-2} — b_i \in G_1 \cup G_2$, producing a *shorter* semi-directed cycle $(a, b \equiv b_0, \ldots, b_{i-2}, b_i, \ldots, b_k \equiv a)$ in $G_1 \cup G_2$ that forces $a — b \in G_1 \vee G_2$.

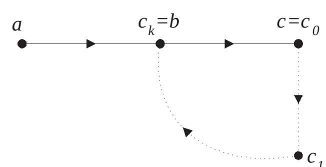
Thus, Cases 1 and 2 together allow us to proceed recursively to reduce (a) to the case where an immorality $a \to b \leftarrow c$ occurs in $G_1 \cup G_2$ but $a \to b$ occurs in a semi-directed *3-cycle* $(a, b \equiv b_0, b_1, b_2 \equiv a)$ in $G_1 \cup G_2$:



Here $k = 2$, so necessarily $i = 2$. But in this case, (a, b, b_1, a) must form a semi-directed 3-cycle in G_2, contradicting its adicyclicity. Thus (a) cannot occur.

(b) *The flag $a \to b — c$ occurs in $G_1 \vee G_2$ but $a \to b \to c$ occurs as an induced subgraph of $G_1 \cup G_2$.*

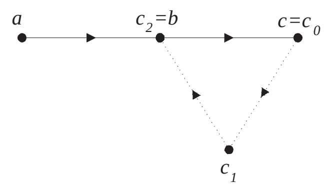
For this to occur, $b \to c$ must be part of a semi-directed cycle $(b, c \equiv c_0, c_1, \ldots, c_k \equiv b)$ in $G_1 \cup G_2$, where $k \geq 2$ and where each edge $c_{j-1} \cdots c_j$ $(1 \leq j \leq k)$ in the cycle occurs as either $c_{j-1} — c_j$ or $c_{j-1} \to c_j$ in $G_1 \cup G_2$ (see the following figure). Necessarily $c_1 \neq a, b$; also, $c_j \neq a$ for any $j = 2, \ldots, k-1$, for otherwise, $a — b — c$ would occur in $G_1 \vee G_2$ rather than $a \to b — c$. Since G_1 and G_2 are adicyclic, we may consider the smallest i, $1 \leq i \leq k$, such that $c_{i-1} \leftarrow c_i \in G_1$ or G_2; say G_1, so necessarily $c_{i-1} \to c_i \in G_2$.



Case 1. Suppose that $i = 1$, so $c \leftarrow c_1 \in G_1$ and $c \to c_1 \in G_2$. Thus $b \to c \leftarrow c_1$ does not occur as an immorality in G_2, hence cannot occur in G_1 (by (ii)), so there must be an edge $b \cdots c_1$ in G_1 and therefore also in G_2 and $G_1 \cup G_2$. Necessarily $b \to c_1 \in G_2$; otherwise, (b, c, c_1, b) would comprise a semi-directed 3-cycle in G_2. If there were an edge $a \cdots c_1$ in G_2, it must be $a \to c_1$ (otherwise, (a, b, c_1, a) would comprise a semi-directed 3-cycle in G_2), which would imply the immorality $a \to c_1 \leftarrow c$ in G_2, hence in G_1 (by (ii)), contradicting $c \leftarrow c_1 \in G_1$. Thus there is no edge $a \cdots c_1$ in G_2, hence none in G_1 or $G_1 \cup G_2$. Therefore, since $b \to c_1 \in G_2$, by (ii) the edge $b \cdots c_1$ cannot occur in G_1 as $b \leftarrow c_1$ nor as $b — c_1$, hence $b \to c_1 \in G_1 \cup G_2$. (Note that this again requires $k \geq 3$, for if $k = 2$ then $b \to c_1 = c_2 \to c_1 \notin G_1 \cup G_2$). Thus, like $a \to b \to c$, $a \to b \to c_1$ occurs as an induced subgraph in $G_1 \cup G_2$, while $b \to c_1$ is part of a *shorter* semi-directed cycle $(b, c_1, \ldots, c_k \equiv b)$ in $G_1 \cup G_2$ that forces the occurrence of $a \to b — c_1$ as a flag in $G_1 \vee G_2$.

Case 2. Suppose that $i \geq 2$. By an argument similar to that in Case 2 of (a), $b \to c$ is part of a *shorter* semi-directed cycle $(b, c_0, \ldots, c_{i-2}, c_i, \ldots, c_k \equiv b)$ in $G_1 \cup G_2$ that forces $b — c \in G_1 \vee G_2$.

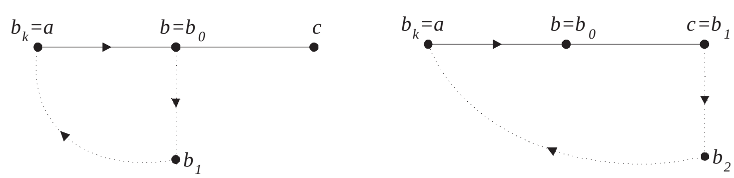
Thus, Cases 1 and 2 together allow us to proceed recursively to reduce (b) to the case where $a \to b \to c$ occurs as an induced subgraph of $G_1 \cup G_2$ but $b \to c$ occurs in a semi-directed *3-cycle* $(b, c \equiv c_0, c_1, c_2 \equiv b)$ in $G_1 \cup G_2$:



Here $k = 2$, so necessarily $i = 2$. But in this case, (b, c, c_1, b) must form a semi-directed 3-cycle in G_2, again contradicting its adicyclicity. Thus (b) cannot occur.

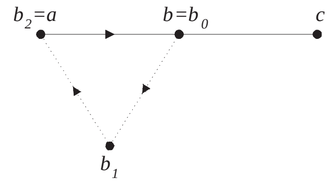
(c) *The flag $a \to b — c$ occurs in $G_1 \cup G_2$ but $a — b — c$ occurs in $G_1 \vee G_2$.*

It suffices to show it impossible that a flag $a \to b — c$ occurs in $G_1 \cup G_2$ but $a — b \in G_1 \vee G_2$. For this to occur, $a \to b$ must occur in a semi-directed cycle $(a, b \equiv b_0, b_1, \ldots, b_k \equiv a)$ in $G_1 \cup G_2$, where $k \geq 2$ and where each edge $b_{i-1} \cdots b_i$ $(1 \leq i \leq k)$ in the cycle occurs as either $b_{i-1} — b_i$ or $b_{i-1} \to b_i$ in $G_1 \cup G_2$. Necessarily $b_1 \neq a$ but both $b_1 \neq c$ and $b_1 = c$ are possible. (These two possibilities are shown in the following figure.) Note that if $b_1 \neq c$ but $b_j = c$ for some $j \geq 2$, then by shortening the cycle this can be reduced to the case $b_1 = c$; hence we may assume that $b_j \neq c$ for all $j \geq 2$. Since G_1 and G_2 are adicyclic, we may consider the smallest i, $1 \leq i \leq k$, such that $b_{i-1} \leftarrow b_i \in G_1$ or G_2; say G_1, so necessarily $b_{i-1} \to b_i \in G_2$. Note that if $b_1 = c$ then $i \geq 2$ (by (ii) and the non-adjacency of a and c in $G_1 \cup G_2$).



Case 1. Suppose that $b_1 \neq c$ and $i = 1$, so $b \leftarrow b_1 \in G_1$ and $b \to b_1 \in G_2$. Thus $a \to b \leftarrow b_1$ (respectively, $b_1 \to b — c$) cannot occur as an immorality (resp., flag) in G_2, so by (ii), neither can they occur as such in G_1. Therefore there must be edges $a \cdots b_1$ and $b_1 \cdots c$ in G_1 and therefore in G_2 and in $G_1 \cup G_2$. To avoid semi-directed 3-cycles in G_2, necessarily $a \to b_1 \in G_2$ and $b_1 \leftarrow c \in G_2$, so $a \to b_1 \leftarrow c$ forms an immorality in G_2, hence also in G_1 (by (ii)). But $b_1 \leftarrow c \in G_1$ implies that (b, c, b_1, b) forms a semi-directed 3-cycle in G_1, contradicting its adicyclicity. Thus Case 1 cannot occur.

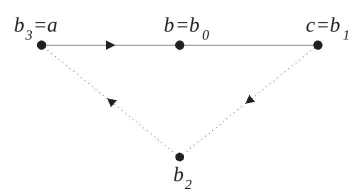
Case 2. Suppose that $b_1 \neq c$ and $i \geq 2$. By an argument similar to that in Case 1 of (a), $a \to b$ is part of a *shorter* semi-directed cycle $(a, b \equiv b_0, \ldots, b_{i-2}, b_i, \ldots, b_k \equiv a)$ (again excluding c) in $G_1 \cup G_2$ that forces $a — b \in G_1 \vee G_2$. Proceed recursively (invoking the impossibility of Case 1 when necessary) to reduce to the case where $a \to b$ is part of a semi-directed 3-cycle $(a, b \equiv b_0, b_1, b_2 \equiv a)$ (again excluding c) in $G_1 \cup G_2$:



Now, either $i = 1$ or $i = 2$. The former is impossible by Case 1, while the latter is also impossible, for it implies that (a, b, b_1, a) forms a semi-directed 3-cycle in G_2. Thus Case 2 cannot occur.

Case 3. Suppose that $b_1 = c$ and $i = 2$, so $c \leftarrow b_2 \in G_1$ and $c \to b_2 \in G_2$. Thus b—$c \leftarrow b_2$ cannot occur as a flag in G_2, hence by (ii) cannot occur as a flag in G_1. Therefore there must be an edge $b \cdots b_2$ in G_1 and G_2. In order that (b, c, b_2, b) not occur as a semi-directed 3-cycle in G_1 (respectively, G_2), necessarily $b \leftarrow b_2 \in G_1$, (resp., $b \to b_2 \in G_2$). Thus $a \to b \leftarrow b_2$ cannot occur as an immorality in G_2, hence not in G_1, so necessarily $a \cdots b_2 \in G_1$, hence $a \cdots b_2 \in G_2$. In order that (a, b, b_2, a) not occur as a semi-directed 3-cycle in G_2, necessarily $a \to b_2 \in G_2$. Therefore $a \to b_2 \leftarrow c$ occurs as an immorality in G_2 but not in G_1, contradicting (ii). Thus Case 3 cannot occur.

Case 4. Suppose that $b_1 = c$ and $i \geq 3$. By an argument similar to that in Case 1 of (a), $a \to b$ is part of a *shorter* semi-directed cycle $(a, b \equiv b_0, c \equiv b_1, \ldots, b_{i-2}, b_i, \ldots, b_k \equiv a)$ in $G_1 \cup G_2$ that forces a—$b \in G_1 \vee G_2$. Proceed recursively (invoking the impossibility of Case 3 when necessary) to reduce to the case where $a \to b$ is part of a semi-directed cycle $(a, b \equiv b_0, c \equiv b_1, b_2, b_3 \equiv a)$ in $G_1 \cup G_2$:



Now, either $i = 2$ or $i = 3$. The former is impossible by Case 3, while the latter is also impossible, since it entails a semi-directed 4-cycle in G_2. Thus Case 4 cannot occur, so (c) cannot occur. This completes the proof of Lemma B.2.

Remark B.1. Lemmas B.1 and B.2 together serve the same function in our characterization of AMP Markov equivalence of CGs as does Proposition 5.4 of Frydenberg (1990) for his characterization of LWF Markov equivalence of CGs.

Proof of Theorem 6.1. ("if"): Here, G_1 and G_2 are CGs such that $G_1^\vee = G_2^\vee$ and such that G_1 and G_2 have the same triplexes. By Lemma B.1, G_{12} and G_{21} are CGs such that $G_{12} \approx G_{21}$ and have the same triplexes as G_1 and G_2. By Lemma B.2 applied to G_{12} and G_{21}, $G_{12} \vee G_{21}$ is a CG that also has the same triplexes as G_1 and G_2. Summarizing, the five CGs G_1, G_{12}, $G_{12} \vee G_{21}$, G_{21}, and G_2 satisfy the relations

$$G_1^\vee = G_{12}^\vee = (G_{12} \vee G_{21})^\vee = G_{21}^\vee = G_2^\vee, \tag{B.1}$$

$$G_1 \supseteq G_{12} \subseteq (G_{12} \vee G_{21}) \supseteq G_{21} \subseteq G_2, \tag{B.2}$$

and all five have the same triplexes. The AMP Markov equivalence of G_1 and G_2 now follows from Proposition B.1.

Proposition B.1. Let $G \equiv (V, E)$ and $\tilde{G} \equiv (V, \tilde{E})$ be chain graphs such that $G^\vee = \tilde{G}^\vee$ and G and $\tilde{G}$ have the same triplexes. If in addition $G \supseteq \tilde{G}$, then

(i) $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X}) \subseteq \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G; \mathbf{X}) \ \ \forall \mathbf{X}$;

(ii) $\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G; \mathbf{X}) \subseteq \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X}) \ \ \forall \mathbf{X}$.

Thus $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) = \mathcal{P}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X}) \ \ \forall \mathbf{X}$ (by Theorem 4.1).

Proof. Note first that for any subset $A \subseteq V$,

$$\mathrm{pa}_G(A) \subseteq \mathrm{pa}_{\tilde{G}}(A), \tag{B.3}$$

$$\mathrm{An}_G(A) \subseteq \mathrm{An}_{\tilde{G}}(A), \tag{B.4}$$

$$\mathrm{pa}_{\tilde{G}}(A) \subseteq \mathrm{pa}_G(A) \cup \mathrm{nb}_G(A), \tag{B.5}$$

$$\mathrm{nb}_{\tilde{G}}(A) \subseteq \mathrm{nb}_G(A), \tag{B.6}$$

$$\mathrm{cl}_{\tilde{G}}(A) \subseteq \mathrm{cl}_G(A), \tag{B.7}$$

$$\mathrm{Co}_{\tilde{G}}(\mathrm{An}_{\tilde{G}}(A)) \subseteq \mathrm{Co}_G(\mathrm{An}_G(A)), \tag{B.8}$$

The inclusions (B.3) – (B.6) follow immediately from the relation $G \supseteq \tilde{G}$, and (B.7) follows from (B.5) and (B.6). To verify (B.8), note that $b \in \mathrm{Co}_{\tilde{G}}(\mathrm{An}_{\tilde{G}}(A))$ iff $b \in A$ or there is a path of the form $b \!-\!\!-\! \cdots \!-\!\!-\! c \to \cdots \to a$ in $\tilde{G}$ from b to some $a \in A$. (This path may be directed, undirected, or semi-directed.) This path also occurs as a path in G, but one or more arrows $v \to w$ in $\tilde{G}$ may change to lines $v \!-\!\!-\! w$ in G. If this results in the occurrence of a subpath of the form $u \to v \!-\!\!-\! w$ in G, then, since G and $\tilde{G}$ have the same triplexes, $u \cdots w \in G$. Since G is adicyclic, necessarily $u \to w \in G$, resulting in a shorter path from b to a in G that bypasses v. This reduction can be continued until a path of the form $b \!-\!\!-\! \cdots \!-\!\!-\! c' \to \cdots \to a$ in G is obtained, hence $b \in \mathrm{Co}_G(\mathrm{An}_G(A))$.

For the remainder of the proof, set $\mathcal{T} = \mathcal{T}(G)$, $\tilde{\mathcal{T}} = \mathcal{T}(\tilde{G})$, $\mathcal{D} = \mathcal{D}(G)$, and $\tilde{\mathcal{D}} = \mathcal{D}(\tilde{G})$.

(i) Assume that $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X})$. To establish that $P \in \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(G; \mathbf{X})$, by Lemma 4.1 it suffices to show that P satisfies conditions C7 and C8 for G:

(C7) $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau : \quad \sigma \perp\!\!\!\perp \big(\tau \backslash \mathrm{cl}_{G_\tau}(\sigma)\big) \,\big|\, \mathrm{pa}_G(\tau) \dot{\cup} \mathrm{nb}_G(\sigma) \, [P]$;

(C8) $\forall \tau \in \mathcal{T}, \forall \sigma \subseteq \tau : \quad \sigma \perp\!\!\!\perp \big(\mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)\big) \,\big|\, \mathrm{pa}_G(\sigma) \, [P]$.

Clearly C7 is trivial when $\sigma = \tau$. To establish C7 when $\sigma \subset \tau$, since $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X})$ it suffices to show that $\mathrm{pa}_G(\tau)\dot{\cup}\mathrm{nb}_G(\sigma)$ separates σ and $\tau\backslash\mathrm{cl}_{G_\tau}(\sigma)$ in $\tilde{H}^{\mathrm{a}}$, where, since $\mathrm{An}_{\tilde{G}}(\mathrm{cl}_G(\tau)) = \mathrm{An}_{\tilde{G}}(\tau)$,

$$
\begin{aligned}
\tilde{H} :=& \tilde{G}[\sigma\dot{\cup}\mathrm{pa}_G(\tau)\dot{\cup}\mathrm{nb}_G(\sigma)\dot{\cup}(\tau\backslash\mathrm{cl}_{G_\tau}(\sigma))] \\
\equiv& \tilde{G}[\mathrm{cl}_G(\tau)] \\
=& \tilde{G}[\tau] \\
=& \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)} \cup \tilde{G}^{\wedge}_{\mathrm{Co}_{\tilde{G}}(\mathrm{An}_{\tilde{G}}(\tau))}.
\end{aligned}
$$

(Note that $\sigma \subset \tau \Rightarrow \mathrm{nb}_G(\sigma) \neq \emptyset$.) By (B.8),

$$\mathrm{Co}_{\tilde{G}}(\mathrm{An}_{\tilde{G}}(\tau)) \subseteq \mathrm{Co}_G(\mathrm{An}_G(\tau)). \tag{B.9}$$

We must show that for any path $\tilde{\pi}$ between σ and $\tau\backslash\mathrm{cl}_{G_\tau}(\sigma)$ in $\tilde{H}^{\mathrm{a}}$,

$$\tilde{\pi} \cap \big(\mathrm{pa}_G(\tau)\dot{\cup}\mathrm{nb}_G(\sigma)\big) \neq \emptyset. \tag{B.10}$$

Let $s \in \tilde{\pi}\cap\sigma$ be the *last* vertex of $\tilde{\pi}$ that is also in σ and let $v \in \tilde{\pi}\backslash\sigma$ be the *next* vertex in $\tilde{\pi}$ after s, so that s—$v \in \tilde{H}^{\mathrm{a}}$ (hence $v \in \mathrm{Co}_G\mathrm{An}_G(\tau)$ by (B.9)). By the definition of $\tilde{H}^{\mathrm{a}}$, either (1) s—$v \in \tilde{H}$, (2) $s \rightarrow v \in \tilde{H}$, (3) $s \leftarrow v \in \tilde{H}$, or (4) $s \not\cdots v$ in $\tilde{H}$ but either $(\{s, v\}, w)$ is a triplex in $\tilde{H}$ for some $w \in \mathrm{Co}_G(\mathrm{An}_G(\tau))$ (by (B.9)) or $[s, v; w, y]$ is a 2-biflag in $\tilde{H}$ for some $w, y \in \mathrm{Co}_G(\mathrm{An}_G(\tau))$ (by (B.9)) (see Figure A.3). In case (1) s—$v \in \tilde{G}$ by the definition of $\tilde{H}$, so s—$v \in G$ by (B.6). Thus $v \in \tilde{\pi}\cap\mathrm{nb}_G(\sigma)$, so (B.10) holds. In case (2) $s \rightarrow v \in \tilde{G}$, so either $s \rightarrow v \in G$ or s—$v \in G$ by (B.5). The former is impossible since $s \in \tau$, $v \in \mathrm{Co}_G\mathrm{An}_G(\tau)$, and G is adicyclic, while as in (1) the latter implies that (B.10) holds. In case (3), $s \leftarrow v \in \tilde{G}$, so either $s \leftarrow v \in G$ or s—$v \in G$ by (B.5). The former implies that $v \in \tilde{\pi}\cap\mathrm{pa}_G(\tau)$ so (B.10) holds, while as in (1) the latter also implies that (B.10) holds.

In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 must occur as an induced subgraph of $\tilde{H}$. If (a) occurs in $\tilde{H}$, then $s \rightarrow w \in \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)}$ and $v \rightarrow w \in \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)}$, so $s \not\cdots v$ in $\tilde{G}$ (since $s \cdots v \in \tilde{G} \Rightarrow s \cdots v \in \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)} \Rightarrow s \cdots v \in \tilde{H}$). Thus $s \rightarrow w \leftarrow v$ occurs as an immorality in $\tilde{G}$, so, since $\tilde{G}$ and G have the same skeleton and triplexes, by (B.5) either $s \rightarrow w \leftarrow v$, $s \rightarrow w$—v, or s—$w \leftarrow v$ must occur as an (induced) subgraph of G. Since G is adicyclic, the first two cases are impossible because $s \in \sigma \subseteq \tau$ and $w \in \mathrm{Co}_G(\mathrm{An}_G(\tau))$. In the third case, $s \in \tau \Rightarrow w \in \tau$, so $v \in \tilde{\pi}\cap\mathrm{pa}_G(\tau)$, hence (B.10) holds.

If (b) occurs in $\tilde{H}$, then $s \rightarrow w$—v occurs as a subgraph of $\tilde{G}$, hence by (B.5) and (B.6) either $s \rightarrow w \in G$ or s—w—v occurs as a subgraph of G. As in (a), the former is impossible since $s \in \tau$ and $w \in \mathrm{Co}_G(\mathrm{An}_G(\tau))$. In the latter case, $(\{s, v\}, w)$ is not a triplex in G, hence not in $\tilde{G}$, so $s \cdots v \in \tilde{G}$, hence $s \cdots v \in G$. Since G is adicyclic, necessarily s—$v \in G$, hence $v \in \tilde{\pi}\cap\mathrm{nb}_G(\sigma)$, so (B.10) holds.

If (c) occurs in $\tilde{H}$ then, since $s \in \sigma \subseteq \mathrm{An}_{\tilde{G}}(\tau)$, s—$w \leftarrow v$ occurs as an induced subgraph (a flag) of $\tilde{G}$, hence by (B.5) and (B.6) either s—$w \leftarrow v$ or s—w—v occurs as an induced subgraph of G. In the former case, $w \in \tau$ so $v \in \tilde{\pi}\cap\mathrm{pa}_G(\tau)$, hence (B.10) holds. The latter case is impossible since G and $\tilde{G}$ have the same triplexes.

If (d) occurs in $\tilde{H}$, then $s \to w \in \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)}$ and $v \to y \in \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)}$, so $s \not\cdots y$ in $\tilde{G}$ (since $s \cdots y \in \tilde{G} \Rightarrow s \cdots y \in \tilde{G}_{\mathrm{An}_{\tilde{G}}(\tau)} \Rightarrow s \cdots y \in \tilde{H}$). Thus $s \to w — y$ occurs as a flag in $\tilde{G}$, so, since $\tilde{G}$ and G have the same skeleton and triplexes, by (B.5) and (B.6) must also occur as a flag in G. In particular, $s \to w \in G$, which is impossible as in (a) and (b).

Thus (B.10) holds in all permissible cases, so C7 is established.

To establish C8, since $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(\tilde{G}; \mathbf{X})$ it suffices to show that $\mathrm{pa}_G(\sigma)$ separates σ and $\sigma' := \mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)$ in $\tilde{H}^{\mathrm{a}}$, where

$$\tilde{H} := \tilde{G}[\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau)].$$

Since $\mathrm{nd}_{\mathcal{D}}(\tau)$ is G-ancestral, G-coherent, and contains $\mathrm{an}_G(\sigma)$, it follows from (B.8) that the vertex set of $\tilde{H}$ satisfies

$$\text{(B.11)} \qquad \mathrm{Co}_{\tilde{G}}(\mathrm{An}_{\tilde{G}}(\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau))) \subseteq \tau \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau) \equiv \mathrm{Nd}_{\mathcal{D}}(\tau).$$

We must show that for any path $\tilde{\pi}$ between σ' and σ in $\tilde{H}^{\mathrm{a}}$,

$$\text{(B.12)} \qquad \tilde{\pi} \cap \mathrm{pa}_G(\sigma) \neq \emptyset.$$

Let $s \in \tilde{\pi} \cap \sigma'$ be the *last* vertex of $\tilde{\pi}$ that is also in σ' and let $v \in \tilde{\pi} \setminus \sigma'$ be the *next* vertex in $\tilde{\pi}$ after s, so that $s — v \in \tilde{H}^{\mathrm{a}}$. By (B.11), $v \in \mathrm{Nd}_{\mathcal{D}}(\tau) \setminus \sigma' = \tau \dot{\cup} \mathrm{pa}_G(\sigma)$, and (B.12) holds if $v \in \mathrm{pa}_G(\sigma)$, so we may assume that $v \in \tau$. By the definition of $\tilde{H}^{\mathrm{a}}$, either (1) $s — v \in \tilde{H}$, (2) $s \leftarrow v \in \tilde{H}$, (3) $s \to v \in \tilde{H}$, or (4) $s \not\cdots v$ in $\tilde{H}$ but either $(\{s, v\}, w)$ is a triplex in $\tilde{H}$ for some $w \in \mathrm{Nd}_{\mathcal{D}}(\tau)$ (by (B.11)) or $[s, v; w, y]$ is a 2-biflag in $\tilde{H}$ for some $w, y \in \mathrm{Nd}_{\mathcal{D}}(\tau)$ (by (B.11)) (see Figure A.3). In case (1) $s — v \in \tilde{G}$ by the definition of $\tilde{H}$, so $s — v \in G$, which is impossible since $s \notin \tau$. In case (2) $s \leftarrow v \in \tilde{G}$, so either $s \leftarrow v \in G$ or $s — v \in G$. The former is impossible since $s \in \sigma' \subseteq \mathrm{nd}_{\mathcal{D}}(\tau)$, while the latter is impossible since $s \notin \tau$.

In case (3) $s \to v \in \tilde{G}$, so either $s — v \in G$ which is impossible since $s \notin \tau$, or else $s \to v \in G$. In the latter case $v \notin \sigma$ (since $s \notin \mathrm{pa}_G(\sigma)$), so $v \in \tau \setminus \sigma$. (*) By the definition of $\tilde{H}$, $s \to v \in \tilde{H} \Rightarrow v \in \mathrm{An}_{\tilde{G}}(\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau))$, so, since $v \notin \sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau)$, a subgraph of the form $s \to v \to v_1 \to \cdots \to v_n \in \sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau)$ must occur in $\tilde{G}$ ($n \geq 1$). Because $v \in \tau$, by (B.5) the subgraph $s \to v — v_1 — \cdots — v_n$ must occur in G, so $v_n \in \tau$, hence $v_n \in \sigma$. Since G and $\tilde{G}$ have the same triplexes, necessarily $s \to v_1 \in G$, hence $s \to v_1 \in \tilde{G}$. Continue this reduction to conclude that $s \to v_n \in G$, implying $s \in \mathrm{pa}_G(\sigma)$, which is impossible since $s \in \sigma'$.

In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 must occur as an induced subgraph of $\tilde{H}$. If (a) occurs in $\tilde{H}$, either $s \to w \leftarrow v$, $s — w \leftarrow v$, $s — w — v$, or $s \to w — v$ must occur as a subgraph of G. The first two cases are impossible since $v \in \tau$ and $w \in \mathrm{Nd}_{\mathcal{D}}(\tau)$. The third case is impossible since $v \in \tau$ but $s \notin \tau$. In the fourth case, $w \in \tau \setminus \sigma$ (since $v \in \tau$ and $s \notin \mathrm{pa}_G(\sigma)$). Now apply the argument beginning at (*) in the preceding paragraph with v replaced by w to obtain a similar contradiction.

If (b) occurs in $\tilde{H}$, either $s — w — v$ or $s \to w — v$ occurs as a subgraph of G. The former case is impossible as in (a). In the latter case, $w \in \tau \setminus \sigma$ (since $v \in \tau$ and $s \notin \mathrm{pa}_G(\sigma)$), so again apply the argument beginning at (*) with v replaced by w to obtain the same contradiction. If (c) occurs in $\tilde{H}$, either s—w—v or s—$w \leftarrow v$ occurs as a subgraph of G. The former is impossible as in (a), while the latter is impossible since $v \in \tau$ and $w \in \mathrm{Nd}_{\mathcal{D}}(\tau)$.

If (d) occurs in $\tilde{H}$, either $s \to w$—$y \leftarrow v$, s—w—$y \leftarrow v$, s—w—y—v, or $s \to w$—y—v occurs as a subgraph of G. The first and second cases are impossible since $v \in \tau$ and $y \in \mathrm{Nd}_{\mathcal{D}}(\tau)$, while the third case is impossible since $v \in \tau$ but $s \notin \tau$. In the fourth case $w \in \tau\backslash\sigma$ (since $v, y \in \tau$ and $s \notin \mathrm{pa}_G(\sigma)$), so again apply the argument beginning at (*) with v replaced by w to obtain the same contradiction.

Thus (B.12) holds in all permissible cases, so C8 is established. This completes the proof of (i).

(ii) Assume that $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X})$. To establish that $P \in \mathcal{P}^{\mathrm{b}}_{\mathrm{AMP}}(\tilde{G};\mathbf{X})$, by Lemma 4.1 it suffices to show that P satisfies C7 and C8 for $\tilde{G}$; we denote these conditions by $\tilde{\mathrm{C7}}$ and $\tilde{\mathrm{C8}}$, respectively:

($\tilde{\mathrm{C7}}$) $\forall\tilde{\tau} \in \tilde{\mathcal{T}}, \forall\tilde{\sigma} \subseteq \tilde{\tau}:\quad \tilde{\sigma} \perp\!\!\!\perp \big(\tilde{\tau}\backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})\big) \;\big|\; \mathrm{pa}_{\tilde{G}}(\tilde{\tau})\dot{\cup}\mathrm{nb}_{\tilde{G}}(\tilde{\sigma})\,[P]$;

($\tilde{\mathrm{C8}}$) $\forall\tilde{\tau} \in \tilde{\mathcal{T}}, \forall\tilde{\sigma} \subseteq \tilde{\tau}:\quad \tilde{\sigma} \perp\!\!\!\perp \big(\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \setminus \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})\big) \;\big|\; \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})\,[P]$.

Clearly $\tilde{\mathrm{C7}}$ is trivial when $\tilde{\sigma} = \tilde{\tau}$. To establish $\tilde{\mathrm{C7}}$ when $\tilde{\sigma} \subset \tilde{\tau}$, since $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G;\mathbf{X})$ it suffices to show that $\mathrm{pa}_{\tilde{G}}(\tilde{\tau})\dot{\cup}\mathrm{nb}_{\tilde{G}}(\tilde{\sigma})$ separates $\tilde{\sigma}$ and $\tilde{\tau}\backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})$ in H^{a}, where

$$H := G[\tilde{\sigma}\dot{\cup}\mathrm{pa}_{\tilde{G}}(\tilde{\tau})\dot{\cup}\mathrm{nb}_{\tilde{G}}(\tilde{\sigma})\dot{\cup}(\tilde{\tau}\backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma}))] = G[\mathrm{cl}_{\tilde{G}}(\tilde{\tau})].$$

(Note that $\tilde{\sigma} \subset \tilde{\tau} \Rightarrow \mathrm{nb}_{\tilde{G}}(\tilde{\sigma}) \neq \emptyset$.) We must show that for any path π between $\tilde{\sigma}$ and $\tilde{\tau}\backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})$ in H^{a},

$$\pi \cap \big(\mathrm{pa}_{\tilde{G}}(\tilde{\tau})\dot{\cup}\mathrm{nb}_{\tilde{G}}(\tilde{\sigma})\big) \neq \emptyset. \tag{B.13}$$

Since $\tilde{G} \subseteq G$, $\tilde{\tau} \subseteq \tau$ for a unique $\tau \in \mathcal{T}$. It follows from (B.5), (B.7), the inclusion $\tilde{\tau} \subseteq \tau$, the connectedness of G_τ, and the adicyclicity of G that

$$\begin{aligned}
\mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau})) &\equiv \mathrm{cl}_{\tilde{G}}(\tilde{\tau})\dot{\cup}\mathrm{an}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau})) \\
&\subseteq \mathrm{cl}_{\tilde{G}}(\tilde{\tau}) \cup \mathrm{an}_G(\tau) \qquad \text{(B.14)}\\
&\subseteq \mathrm{cl}_G(\tau) \cup \mathrm{an}_G(\tau) \\
&= \mathrm{An}_G(\tau) \qquad \text{(B.15)}\\
&\subseteq \mathrm{Nd}_G(\tau). \qquad \text{(B.16)}
\end{aligned}$$

We shall establish (B.13) in two steps.

Step 1. If $\pi \not\subseteq \tau$, *then* $\pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau}) \neq \emptyset$, *so* (B.13) *holds:*

Let s, v denote consecutive vertices in π such that $s \in \tau$, $v \notin \tau$. Since s—$v \in H^{\mathrm{a}}$, either (1) s—$v \in H$, (2) $s \to v \in H$, (3) $s \leftarrow v \in H$, or (4) $s \not\sim v$ in H but either $(\{s, v\}, w)$ is a triplex in H for some $w \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ or $[s, v; w, y]$ is a 2-biflag in H for some $w, y \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ (see Figure A.3). In case (1) s—$v \in G$, contradicting $s \in \tau$, $v \notin \tau$. In case (2) $s \to v \in G$ and $v \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ by the definition of H, hence $v \in \mathrm{Nd}_G(\tau)$

by (B.16), again a contradiction. In case (3), $s \leftarrow v \in G$ and $s \in \tau \cap \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ by the definition of H, so $s \in \mathrm{cl}_{\tilde{G}}(\tilde{\tau}) \equiv \tilde{\tau}\dot{\cup}\mathrm{pa}_{\tilde{G}}(\tilde{\tau})$ by (B.14). Since $s \leftarrow v \in \tilde{G}$ by (B.3), either $v \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau})$ or $s \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau})$, hence $\pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau}) \neq \emptyset$.

In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 must occur as an induced subgraph of H. In cases (a), (b), and (d), $s \rightarrow w \in G$ but $w \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau})) \subseteq \mathrm{Nd}_G(\tau)$, again a contradiction as in (2). If (c) occurs in H then s—$w \leftarrow v$ occurs in G and $w \in \tau \cap \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ by the definition of H, so $w \in \mathrm{cl}_{\tilde{G}}(\tilde{\tau}) \equiv \tilde{\tau}\dot{\cup}\mathrm{pa}_{\tilde{G}}(\tilde{\tau})$ by (B.14). Since $w \leftarrow v \in \tilde{G}$ by (B.3), either $v \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau})$ as required, or $w \in \mathrm{pa}_{\tilde{G}}(\tilde{\tau})$. In the latter case, $\tilde{t} \leftarrow w \in \tilde{G}$ for some $\tilde{t} \in \tilde{\tau}$, so either $\tilde{t} \leftarrow w \in G$ or $\tilde{t}$—$w \in G$ by (B.5). The former is impossible since $w \in \tau$ and $\tilde{\tau} \subseteq \tau$. In the latter case we have $\tilde{t} \leftarrow w \leftarrow v \in \tilde{G}$ and $\tilde{t}$—$w \leftarrow v \in G$. Since $\tilde{G}$ and G have the same triplexes and G is adicyclic, necessarily $\tilde{t} \leftarrow v \in G$, hence $\tilde{t} \cdots v \in \tilde{G}$. Since $\tilde{G}$ is adicyclic, necessarily $\tilde{t} \leftarrow v \in \tilde{G}$, so again $v \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau})$ as required. This completes Step 1.

Step 2. If $\pi \subseteq \tau$, *then* (B.13) *holds:*

For any two consecutive vertices $s, v \in \pi$, s—$v \in H^{\mathrm{a}}$, so either (1) s—$v \in H$, (2) $s \rightarrow v \in H$, (3) $s \leftarrow v \in H$, or (4) $s \not\sim v$ in H but either $(\{s,v\}, w)$ is a triplex in H for some $w \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ or $[s, v; w, y]$ is a 2-biflag in H for some $w, y \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau}))$ (see Figure A.3). In case (1), s—$v \in G$. In case (2) $s \rightarrow v \in G$, which is impossible since $s, v \in \tau$; similarly, (3) is impossible. In case (4), $w \in \mathrm{An}_G(\mathrm{cl}_{\tilde{G}}(\tilde{\tau})) \subseteq \mathrm{An}_G(\tau)$ by (B.15) while either $s \rightarrow w \in G$ or $v \rightarrow w \in G$ (or both), again impossible since $s, v \in \tau$. Thus s—$v \in G$ for each consecutive pair $s, v \in \pi$.

Therefore, the length of π must be ≥ 2 since $\tilde{\sigma}$ and $\tilde{\tau} \backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})$ cannot contain adjacent vertices in $\tilde{G}$. For any *three* consecutive vertices $s, w, v \in \pi$, s—w—v must occur in G. If either (a) $s \rightarrow w \leftarrow v$, (b) s—$w \leftarrow v$, or (c) $s \rightarrow w$—v occurs in $\tilde{G}$, then $s \cdots v \in \tilde{G}$ since G and $\tilde{G}$ have the same triplexes. Therefore $s \cdots v \in G$ since G and $\tilde{G}$ have the same skeletons, hence s—$v \in G$ since G is adicyclic. Thus we can eliminate the vertex w to produce a *shorter* path $\pi' \subseteq \pi \subseteq \tau$ that also connects $\tilde{\sigma}$ and $\tilde{\tau} \backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})$ in G, with s—$v \in G$ for each consecutive pair $s, v \in \pi'$. Continue this reduction procedure until we obtain a path $\pi'' \subseteq \pi \subseteq \tau$ that also connects $\tilde{\sigma}$ and $\tilde{\tau} \backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})$ in G, such that none of the configurations (a), (b), or (c) occur in $\tilde{G}$ for any three consecutive vertices $s, w, v \in \pi''$. This implies that π'' must have the form $s_0 \leftarrow \cdots \leftarrow s_l$— $\cdots$ — $s_{l+m} \rightarrow \cdots \rightarrow s_{l+m+r}$ in $\tilde{G}$, where $s_0 \in \tilde{\sigma} \subset \tilde{\tau}$, $s_{l+m+r} \in \tilde{\tau} \backslash \mathrm{cl}_{\tilde{G}_{\tilde{\tau}}}(\tilde{\sigma})$, and $l + m + r \geq 2$. If $l \geq 1$ or $r \geq 1$ then $\pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\tau}) \neq \emptyset$, while if $l = r = 0$ but $m \geq 2$ then $\pi \cap \mathrm{nb}_{\tilde{G}}(\tilde{\sigma}) \neq \emptyset$. In either case (B.13) holds, so Step 2 is complete and $\tilde{\mathrm{C7}}$ is established.

To establish $\tilde{\mathrm{C8}}$, since $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G; \mathbf{X})$ it suffices to show that $\mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$ separates $\tilde{\sigma}$ and $\tilde{\sigma}' := \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \setminus \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$ in H^{a}, where

$$H := G[\tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})].$$

Since $\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ is $\tilde{G}$-ancestral and contains $\mathrm{an}_{\tilde{G}}(\tilde{\sigma})$, it follows from (B.4) that

$$\mathrm{An}_G(\tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})) = \tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \equiv \tilde{\sigma}\dot{\cup}\tilde{\sigma}'\dot{\cup}\mathrm{pa}_{\tilde{G}}(\tilde{\sigma}). \tag{B.17}$$

We must show that for any path π between $\tilde{\sigma}'$ and $\tilde{\sigma}$ in H^{a},

$$\pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\sigma}) \neq \emptyset. \tag{B.18}$$

By considering a subpath of π if necessary, we may assume that only one vertex of π (its *initial* vertex) lies in $\tilde{\sigma}'$ and only one vertex of π (its *terminal* vertex) lies in $\tilde{\sigma}$. As above, $\tilde{\tau} \subseteq \tau$ for a unique $\tau \in \mathcal{T}$.

Step 1. If $\pi \not\subseteq \tau$, *then* (B.18) *holds:*

Assume that (B.18) fails. Let s, v denote consecutive vertices in π such that $s \in \tau$ and $v \notin \tau$. By assumption, $s, v \notin \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$. Since $s—v \in H^{\mathrm{a}}$, either (1) $s—v \in H$, (2) $s \to v \in H$, (3) $s \leftarrow v \in H$, or (4) $s \not\cdot\!\cdot v$ in H but either $(\{s, v\}, w)$ is a triplex in H for some $w \in \tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ (by (B.17)) or $[s, v; w, y]$ is a 2-biflag in H for some $w, y \in \tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ (by (B.17)) (see Figure A.3). In case (1) $s—v \in G$, contradicting $s \in \tau$, $v \notin \tau$. In case (2) $s \to v \in G \supseteq \tilde{G}$ and $s, v \in \tilde{\sigma}\dot{\cup}\tilde{\sigma}'$ by (B.17) and the definition of H, so $v \in \tilde{\sigma}' \subseteq \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ (since $\tilde{\sigma} \subseteq \tilde{\tau} \subseteq \tau$), hence $s \in \tilde{\sigma} \subseteq \tilde{\tau}$, which is impossible since $s \to v \in \tilde{G}$. In case (3) $s \leftarrow v \in G \supseteq \tilde{G}$ while $v \in \tilde{\sigma}'$ and $s \in \tilde{\sigma}$ as in (2), so $v \in \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$, again a contradiction.

In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 must occur as an induced subgraph of H. If (a) or (d) occurs in H then $s \to w \in G \supseteq \tilde{G}$ while $v \in \tilde{\sigma}'$ and $s \in \tilde{\sigma} \subseteq \tilde{\tau}$ as in (2), contradicting $w \in \tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \subseteq \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$.

If (c) occurs in H, then $s—w \leftarrow v$ occurs as a subgraph (not necessarily an induced subgraph) of $G \supseteq \tilde{G}$ while $w \in \tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ and $v \in \tilde{\sigma}\dot{\cup}\tilde{\sigma}'$ by (B.17). Now $v \notin \mathrm{pa}_{\tilde{G}}(\tilde{\sigma}) \Rightarrow w \notin \tilde{\sigma} \Rightarrow w \in \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$, and $v \notin \tau \Rightarrow v \notin \tilde{\sigma} \Rightarrow v \in \tilde{\sigma}' \Rightarrow s \notin \tilde{\sigma}' \Rightarrow s \notin \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \Rightarrow s \in \tilde{\tau}\dot{\cup}\mathrm{de}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. Thus neither $s \to w$ nor $s—w$ can occur in $\tilde{G}$, hence $s \leftarrow w \in \tilde{G}$. Therefore $s \leftarrow v \in G$ and $s \leftarrow v \in \tilde{G}$, since G and $\tilde{G}$ have the same triplexes and skeletons and are adicyclic. Thus $s \notin \tilde{\sigma}$ because $v \notin \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$. Since $s \in \tilde{\tau}\dot{\cup}\mathrm{de}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ and $\tilde{\sigma} \subseteq \tilde{\tau}$ (a connected subset in $\tilde{G}$), there exists a path $(s_0, s_1, \ldots, s_n \equiv s)$ of length $n \geq 1$ from $\tilde{\sigma}$ to s in $\tilde{G}$. (Note that $s_i \neq v, w$ for $i = 0, 1, \ldots, n$ since $v, w \in \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$.) Because $\tilde{G} \subseteq G$, $(s_0, s_1, \ldots, s_n \equiv s)$ must also be a path from $\tilde{\sigma}$ to s in G; furthermore this path must be *undirected* in G since $s_0, s \in \tau$. Thus the subgraph $s_{n-1}—s_n—w$ occurs in G while $s_{n-1}—s_n \leftarrow w$ occurs in $\tilde{G}$, so necessarily $s_{n-1} \leftarrow w \in \tilde{G}$ and $s_{n-1}—w \in G$, again because G and $\tilde{G}$ have the same triplexes and skeletons and are adicyclic. Therefore the subgraph $s_{n-2}—s_{n-1}—w$ occurs in G while $s_{n-2}—s_{n-1} \leftarrow w$ occurs in $\tilde{G}$, so $s_{n-2} \leftarrow w \in \tilde{G}$ and $s_{n-2}—w \in G$. Continue this process to conclude that $s_0 \leftarrow w \in \tilde{G}$ and $s_0—w \in G$, hence the subgraphs $s_0 \leftarrow w \leftarrow v$ and $s_0—w \leftarrow v$ occur in $\tilde{G}$ and G, respectively. Therefore $s_0 \leftarrow v \in G$, hence $s_0 \leftarrow v \in \tilde{G}$, so $v \in \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$, again a contradiction.

If (b) occurs in H then $s \to w—v$ occurs as a subgraph of $G \supseteq \tilde{G}$ while $s \in \tilde{\sigma}\dot{\cup}\tilde{\sigma}'$ and $w \in \tilde{\sigma}\dot{\cup}\mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ by (B.17). Now $s \notin \mathrm{pa}_{\tilde{G}}(\tilde{\sigma}) \Rightarrow w \notin \tilde{\sigma} \Rightarrow w \in \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \Rightarrow s \notin \tilde{\tau} \Rightarrow s \in \tilde{\sigma}'$ (*so s must be the unique initial vertex of* π) $\Rightarrow v \notin \tilde{\sigma}'$, while $v \notin \tau \Rightarrow v \notin \tilde{\tau} \Rightarrow v \in \tilde{\sigma}'\dot{\cup}\mathrm{de}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \Rightarrow v \in \mathrm{de}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \Rightarrow v \notin \tilde{\sigma}$, so v is *not* the terminal vertex of π. Because we have assumed that after leaving its initial vertex $s \in \tau$ the path π exits τ at v, π must eventually re-enter τ in order to reach its terminal vertex in $\tilde{\sigma} \subset \tau$. Thus there must exist a *second* pair s', v' of consecutive vertices in π such that $s' \in \tau, v' \notin \tau$ (possibly $v' = v$, but $s' \neq s$.) Now repeat the argument of Step 1 applied to the pair s', v' to conclude that s' is the initial vertex of π, hence $s' = s$, again a contradiction. This completes Step 1.

Step 2. If $\pi \subseteq \tau$, *then* (B.18) *holds:*

Assume that (B.18) fails. For any two consecutive vertices $s, v \in \pi$, $s - v \in H^{\mathrm{a}}$, so either (1) $s - v \in H$, (2) $s \to v \in H$, (3) $s \leftarrow v \in H$, or (4) $s \cdot\!\!/\!\!\cdot v$ in H but either $(\{s,v\},w)$ is a triplex in H for some $w \in \tilde\sigma \dot\cup \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde\tau)$ (by (B.17)) or $[s,v;w,y]$ is a 2-biflag in H for some $w, y \in \tilde\sigma \dot\cup \mathrm{nd}_{\tilde{\mathcal{D}}}(\tilde\tau)$ (by (B.17)) (see Figure A.3). By assumption, $s, v \notin \mathrm{pa}_{\tilde G}(\tilde\sigma)$. In case (1), $s - v \in G$. In case (2) $s \to v \in G$, which is impossible since $s, v \in \tau$; similarly, (3) is impossible. In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 must occur as an induced subgraph of H. If (a) or (d) occurs in H then $s, v \in \tilde\sigma \dot\cup \tilde\sigma'$, $s \to w \in G \supseteq \tilde G$, and $v \to w \in G \supseteq \tilde G$ (in (a)) or $v \to y \in G \supseteq \tilde G$ (in (d)) by (B.17) and the definition of H. Therefore $s, v \notin \tilde\sigma$ since $w, y \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde\tau)$, hence both $s, v \in \tilde\sigma'$, a contradiction. If (b) or (c) occurs in H then $s \to w - v$ or $s - w \leftarrow v$ occurs as a subgraph of G, contradicting $s, v \in \tau$. Thus $s - v \in G$ for each consecutive pair $s, v \in \pi$.

Therefore, the length of π must be ≥ 2 since $\tilde\sigma'$ and $\tilde\sigma$ cannot contain adjacent vertices in $\tilde G$. By the argument in the second paragraph of Step 2 for C$\tilde{7}$, π must contain a subpath $\pi'' \subseteq \pi \subseteq \tau$ of the form $s_0 \leftarrow \cdots \leftarrow s_l - \cdots - s_{l+m} \to \cdots \to s_{l+m+r}$ in $\tilde G$, where $s_0 \in \tilde\sigma'$, $s_{l+m+r} \in \tilde\sigma$, and $l + m + r \geq 2$. If $r = 0$ then $s_0 \in \tilde\sigma' \cap \mathrm{de}_{\tilde{\mathcal{D}}}(\tilde\tau) = \emptyset$, which is impossible. If $r \geq 1$ then $s_{l+m+r-1} \in \mathrm{pa}_{\tilde G}(\tilde\sigma)$, contradicting the assumption that (B.18) fails.

Thus this assumption is disproved, so Step 2 is complete and C$\tilde{8}$ is established. This completes the proof of (ii), so Proposition B.1 is established.

Remark B.2. Note that both parts (i) and (ii) of Proposition B.1 reduce to Theorem 4.1(i) when $G = \tilde G$. Thus, both Theorems 4.1(i) and 6.1("if") follow from Proposition B.1 and Theorem 4.1(ii).

Proof of Theorem 6.1. ("only if"): First suppose that $G_1^\vee \neq G_2^\vee$. We shall show that if $a, b \in V$ are such that $a \cdot\!\!/\!\!\cdot b$ in G_1 but $a \cdots b$ in G_2, then

$$\text{(B.19)} \qquad \big(\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_2; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)\big) \setminus \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \neq \emptyset.$$

Let $P = \mathcal{N}_V(0, \Sigma)$, where $\Sigma = (\sigma_{vw} | v, w \in V)$ with $\sigma_{vv} = 1$ for every $v \in V$, $\sigma_{vw} = 0$ for $\{v, w\} \neq \{a, b\}$, $\sigma_{ab} = \rho \neq 0$, and $|\rho|$ is small enough that Σ is positive definite, so $P \in \mathcal{C}^{**}(\mathbf{R}^V)$. Note that $(a,b) \perp\!\!\!\perp v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp v_m\,[P]$ where $\{v_1, \dots, v_m\} := V \setminus \{a, b\}$. It follows trivially from Lemma 4.2(ii) with $A = \{a, b\}$ that $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_2; \mathbf{R}^V)$. For $G = G_1$, however, $a \cdot\!\!/\!\!\cdot b$ implies that either $b \in \tau(a) \setminus \mathrm{cl}_{G_{\tau(a)}}(a)$, $b \in \mathrm{nd}_{\mathcal{D}}(\tau(a)) \setminus \mathrm{pa}_G(a)$, or $a \in \mathrm{nd}_{\mathcal{D}}(\tau(b)) \setminus \mathrm{pa}_G(b)$. In the first case P1 (and P1$'$) fails for P when $\{v, w\} = \{a, b\}$ since $a \not\perp\!\!\!\perp b\,[P]$, while in the second and third cases P2 (and P2$'$) fails for P when $(v, w) = (a, b)$ or (b, a), respectively, for the same reason. Thus $P \notin \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$, so $P \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$ by Theorem 4.2(i), hence (B.19) holds.

Next suppose that $G_1^\vee = G_2^\vee$ but that the triplex $(\{a, b\}, c)$ occurs in G_1 but not in G_2 (recall Figure 2.4). We shall show that

$$\text{(B.20)} \qquad \big(\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)\big) \setminus \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_2; \mathbf{R}^V) \neq \emptyset,$$

$$\text{(B.21)} \qquad \big(\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_2; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)\big) \setminus \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \neq \emptyset.$$

Let $P_1 = \mathcal{N}_V(0, \Sigma_1)$ and $P_2 = \mathcal{N}_V(0, \Sigma_1^{-1})$, where $\Sigma_1 = (\sigma_{vw} | v, w \in V)$ with $\sigma_{vv} = 1$ for every $v \in V$, $\sigma_{vw} = 0$ for $\{v, w\} \not\subseteq \{a, b, c\}$ $(v \neq w)$, $\sigma_{ab} = 0$, $\sigma_{ac} = \sigma_{bc} = \rho \neq 0$, and $|\rho|$ is small enough that Σ_1 is positive definite, so $P_1, P_2 \in \mathcal{C}^{**}(\mathbf{R}^V)$. Note that $(a,b,c) \perp\!\!\!\perp v_1 \perp\!\!\!\perp \cdots \perp\!\!\!\perp v_{m-1}\,[P_i]$ for $i = 1, 2$, where $V\backslash\{a,b,c\} =: \{v_1, \ldots, v_{m-1}\}$. Because the triplex $(\{a,b\}, c)$ occurs in G_1, one of the three configurations (1) $a \to c \leftarrow b$, (2) $a \to c$—b, or (3) a—$c \leftarrow b$ must occur as an induced subgraph of G_1. Since $a \perp\!\!\!\perp b\,[P_1]$, it follows from Lemma 4.2(ii) with $A = \{a,b,c\}$ that $P_1 \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$. Next, because G_1 and G_2 have the same skeletons but $(\{a,b\}, c)$ does not occur as a triplex in G_2, one of the six configurations (4) $a \to c \to b$, (5) $a \leftarrow c \to b$, (6) a—$c \to b$, (7) a—c—b, (8) $a \leftarrow c \leftarrow b$, or (9) $a \leftarrow c$—b must occur as an induced subgraph of G_2. Since $a \perp\!\!\!\perp b|c\,[P_2]$, it follows from Lemma 4.2(ii) with $A = \{a,b,c\}$ that $P_2 \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_2; \mathbf{R}^V)$.

If (1) occurs in $G \equiv G_1$ then either $b \in \tau(a)\backslash \mathrm{cl}_{G_{\tau(a)}}(a)$ or $b \notin \tau(a)$. In the first case, the CI in P1 fails for $(v,w) = (a,b)$ under P_2 because $c \notin \mathrm{nd}_{\mathcal{D}}(\tau(a))\dot\cup(\tau(a)\backslash\{a,b\})$ by (1) and $a \not\perp\!\!\!\perp b\,[P_2]$. In the second case, either $b \in \mathrm{nd}_{\mathcal{D}}(\tau(a))\backslash \mathrm{pa}_G(a)$ or $a \in \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash \mathrm{pa}_G(b)$, again by (1). Since $c \notin \mathrm{nd}_{\mathcal{D}}(\tau(a))\backslash\{b\}$ and $c \notin \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash\{a\}$ by (1), the CI in P2 fails either for $(v,w) = (a,b)$ or for $(v,w) = (b,a)$ under P_2 because $a \not\perp\!\!\!\perp b\,[P_2]$. If (2) occurs in $G \equiv G_1$ then $a \in \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash \mathrm{pa}_G(b)$ and $c \notin \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash\{a\}$, hence the CI in P2 fails for $(v,w) = (b,a)$ under P_2, again because $a \not\perp\!\!\!\perp b\,[P_2]$. Case (3) is similar to (2). We conclude that $P_2 \notin \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$, so $P_2 \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$ by Theorem 4.2(i), hence (B.20) holds.

If (4) or (6) occurs in $G \equiv G_2$ then $a \in \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash \mathrm{pa}_G(b)$ and $c \in \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash\{a\}$, hence the CI in P2 fails for $(v,w) = (b,a)$ under P_1 because $a \not\perp\!\!\!\perp b|c\,[P_1]$. Cases (8) and (9) are similar to (4) and (6), respectively. If (7) occurs in $G \equiv G_2$ then $b \in \tau(a)\backslash \mathrm{cl}_{G_{\tau(a)}}(a)$ and $c \in \mathrm{nd}_{\mathcal{D}}(\tau(a))\dot\cup(\tau(a)\backslash\{a,b\})$, hence the CI in P1 fails for $(v,w) = (a,b)$ under P_1 because $a \not\perp\!\!\!\perp b|c\,[P_1]$. If (5) occurs in $G \equiv G_2$ then either $b \in \tau(a)\backslash \mathrm{cl}_{G_{\tau(a)}}(a)$ or $b \notin \tau(a)$. In the first case, the CI in P1 fails for $(v,w) = (a,b)$ under P_1 because $c \in \mathrm{nd}_{\mathcal{D}}(\tau(a))\dot\cup(\tau(a)\backslash\{a,b\})$ by (5) and $a \not\perp\!\!\!\perp b|c\,[P_1]$. In the second case, either $b \in \mathrm{nd}_{\mathcal{D}}(\tau(a))\backslash \mathrm{pa}_G(a)$ or $a \in \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash \mathrm{pa}_G(b)$, again by (5). Since $c \in \mathrm{nd}_{\mathcal{D}}(\tau(a))\backslash\{b\}$ and $c \in \mathrm{nd}_{\mathcal{D}}(\tau(b))\backslash\{a\}$ by (5), the CI in P2 fails either for $(v,w) = (a,b)$ or for $(v,w) = (b,a)$ under P_1, again because $a \not\perp\!\!\!\perp b|c\,[P_1]$. As above we conclude that $P_1 \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_2; \mathbf{R}^V)$, hence (B.21) holds and the proof is complete.

Proof of Proposition 6.2. (2) $\Rightarrow$ (2$'$): By Theorem 6.1, G has the same skeleton and triplexes as some ADG D; note that every triplex in D must occur as an immorality in D. If G has a 2-biflag $[a,b;c,d]$, either orientation of the undirected edge c—d will replace one of the flags $[a,d;c]$ or $[b,c;d]$ in G by a non-immorality in D, a contradiction. If G has a k-biflag $[a; c_1, \ldots, c_k]$ with the flags $[a, c_1; c_2]$ and $[a, c_k; c_{k-1}]$, then the immoralities $(a, c_1; c_2)$ and $(a, c_k; c_{k-1})$ must occur in D, hence $c_1 \to c_2 \in D$ and $c_k \to c_{k-1} \in D$. Now any orientation of the undirected edges c_i—c_{i+1}, $i = 2, \ldots, k-2$ will create at least one immorality in D that does not correspond to a triplex in G, again a contradiction. Similarly, G cannot have a k-biflag $[a,b; c_1, \ldots, c_k]$, $k \geq 3$. Finally, G cannot possess a chordless undirected n-cycle C for any $n \geq 4$, since any acyclic orientation of such C must create at least one immorality.

(2$'$) $\Rightarrow$ (2): We shall construct an ADG D that has the same skeleton as G and whose immoralities exactly correspond to the triplexes (if any) of G. First, construct a graph $\tilde{G} \subseteq G$ with the same skeleton as G by converting each flag $[a,b;c]$ in G into an immorality $(a,b;c)$; this process is unambiguous since G has no 2-biflags. Then:

(i) $\tilde{G}$ is a CG: If $\tilde{G}$ were not adicyclic, it would contain a semi-directed cycle $C \equiv$

$(c_0, c_1, \ldots, c_n \equiv c_0)$, $n \geq 3$, with $c_{i-1} \to c_i \in \tilde{G}$ for at least one $i \in \{1, \ldots, n\}$, say $i = 1$. Since G is adicyclic, C must determine a fully undirected cycle in G. Thus, by the construction of $\tilde{G}$, there must exist a vertex $a \neq c_0, c_1$ such that $[a, c_0; c_1]$ is a flag in G; in particular, $a \not\cdots c_0$ in G. Necessarily $a \neq c_{n-1}$ since $c_{n-1} \cdots c_0 \in G$, and, if $n \geq 4$, $a \neq c_j$ for any $j = 2, \ldots, n-2$, else $(c_1, \ldots, c_j, c_1)$ would constitute a semi-directed cycle in G. But $[a, c_2; c_1]$ cannot occur as a flag in G, else $c_1 \leftarrow c_2 \in \tilde{G}$, hence $a \cdots c_2 \in G$. By the adicyclicity of G, this can occur only if $a \to c_2 \in G$. Similarly, since $[a, c_{i+1}; c_i]$ cannot occur as a flag in G for $i = 2, \ldots, n-1$, necessarily $a \to c_{i+1} \in G$ for $i = 2, \ldots, n-1$. In particular, $a \to c_n \in G$, contradicting the non-adjacency of a and $c_n \equiv c_0$ in G.

(ii) By the construction of $\tilde{G}$, each triplex in G corresponds to an immorality in $\tilde{G}$.

(iii) Since G has no 3-biflags $[a; c_1, c_2, c_3]$ or $[a, b; c_1, c_2, c_3]$, each immorality in $\tilde{G}$ corresponds to a triplex in G. (But not every flag in $\tilde{G}$ need correspond to a triplex in G.)

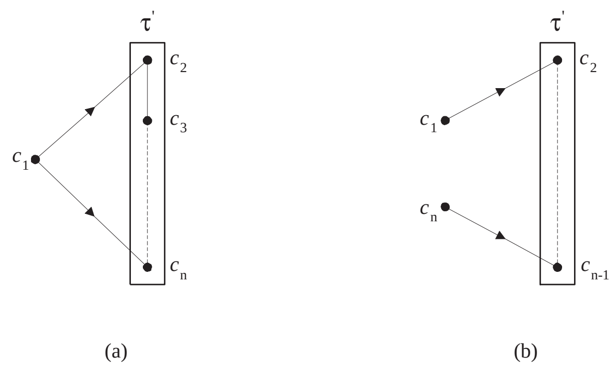


Figure B.1. The induced subgraph $\tilde{G}_C$ for the chordless cycle $C \equiv \{c_0, c_1, \ldots, c_n \equiv c_0\}$.

(iv) $(\tilde{G}_{\mathrm{cl}_{\tilde{G}}(\tilde{\tau})})^{\mathrm{m}}$ is chordal $\forall\ \tilde{\tau} \in \mathcal{T}(\tilde{G})$: If not, then $(\tilde{G}_{\mathrm{cl}_{\tilde{G}}(\tilde{\tau})})^{\mathrm{m}}$ has a chordless n-cycle $C \equiv (c_0, c_1, \ldots, c_n \equiv c_0) \subseteq \mathrm{cl}_{\tilde{G}}(\tilde{\tau}) \equiv \tilde{\tau} \dot{\cup} \mathrm{pa}_{\tilde{G}}(\tilde{\tau})$, $n \geq 4$. Since $\tilde{G} \subseteq G$, $\tilde{\tau} \subseteq \tau$ for some $\tau \in \mathcal{T}(G)$. Thus, since $((\tilde{G}_{\mathrm{cl}_{\tilde{G}}(\tilde{\tau})})^{\mathrm{m}})_{\tilde{\tau}} = \tilde{G}_{\tilde{\tau}} = G_{\tilde{\tau}}$ and G_τ is chordal, C cannot lie entirely within $\tilde{\tau}$. Because $\mathrm{pa}_{\tilde{G}}(\tilde{\tau})$ is complete in $(\tilde{G}_{\mathrm{cl}_{\tilde{G}}(\tilde{\tau})})^{\mathrm{m}}$, either (a) exactly one vertex of C lies in $\mathrm{pa}_{\tilde{G}}(\tilde{\tau})$, say c_1, or (b) exactly two vertices of C, necessarily consecutive, lie in $\mathrm{pa}_{\tilde{G}}(\tilde{\tau})$, say c_1 and c_n. In case (a), $c_2, c_n \in \tilde{\tau}$, hence $c_1 \cdots c_2 \in \tilde{G}$ and $c_1 \cdots c_n \in \tilde{G}$, so $c_1 \to c_2 \in \tilde{G}$ and $c_1 \to c_n \in \tilde{G}$ by (i) and the connectedness of $\tilde{G}_{\tilde{\tau}}$ (see Figure B.1a). But then $c_1 — c_2 \in G$ and $c_1 — c_n \in G$ by the definition of $\tilde{G}$, so C is a chordless cycle in G_τ, contradicting the chordality of G_τ. In case (b), similarly $c_1 \to c_2 \in \tilde{G}$ and $c_n \to c_{n-1} \in \tilde{G}$ (see Figure B.1b), while $c_1 — c_2 \in G$ and $c_n — c_{n-1} \in G$. Therefore $C \subseteq \tau$, hence c_1 and c_n cannot be adjacent in G, as G_τ is chordal. Furthermore, by the definition of $\tilde{G}$, there must exist vertices a, b (possibly $a = b$) such that $[a, c_n; c_{n-1}]$ and $[b, c_1; c_2]$ are flags in G. Therefore $b \neq c_1, c_2, c_3$ and, since C is a chordless cycle in $(\tilde{G}_{\mathrm{cl}_{\tilde{G}}(\tilde{\tau})})^{\mathrm{m}}$, $b \neq c_4, \ldots, c_n$. Similarly, $a \neq c_1, \ldots, c_n$. Since $c_i — c_{i+1} \in \tilde{G} \subseteq G$ for $i = 2, \ldots, n-2$, it follows from the definition of $\tilde{G}$ and the adicyclicity of G that $b \to c_i \in G$, $i = 3, \ldots, n-1$, and $a \to c_i \in G$, $i = n-2, \ldots, 2$. Thus, at least one of $[a; c_1, \ldots, c_n]$, $[b; c_1, \ldots, c_n]$, or $[a, b; c_1, \ldots, c_n]$ must occur as an n-biflag in G, contradicting $(2')$.

It follows from (iv) and Proposition 4.2 of [AMP] (1997a) that the undirected edges of $\tilde{G}$ can be oriented to yield an ADG D that is LWF Markov equivalent to $\tilde{G}$, that is, $\tilde{G}$ and D have the same skeleton and complexes. Thus every immorality in D is an immorality in $\tilde{G}$ hence, by (iii), corresponds to a triplex in G, while by (ii), every triplex of G corresponds to an immorality in D. Therefore, $G \overset{\mathrm{A}}{\sim} D$ by Theorem 6.1, so (2) holds.

Remark B.3. By (iv), $\tilde{G}$ has no 2-biflags and no complexes other than immoralities. (But $\tilde{G}$ might have k-biflags for $k \geq 3$.)

Remark B.4. Once we have obtained one ADG $D \overset{\mathrm{A}}{\sim} G$ by the construction in the proof of Proposition 4.2, we can generate *all* ADGs $D' \overset{\mathrm{A}}{\sim} G$ by the process described in Remark 4.1 of [AMP] (1997a).

Proof of Proposition 6.4. $(3') \Rightarrow (4)$: this is immediate since every induced subgraph of a chordal UG is chordal.

$(4) \Rightarrow (3')$: It suffices to show that $G^{\vee}$ is chordal. If not, then $G^{\vee}$ contains a chordless n-cycle C, $n \geq 4$. By (4), C cannot lie entirely within any chain component of G, hence G_C must contain at least one directed edge. Since G is adicyclic, in fact G_C must contain at least two opposing directed edges. A straightforward argument now shows that G_C, and hence G, must contain at least one triplex, contradicting (4).

The following proposition will be used in the proof of Theorem 6.2.

Proposition B.2. Let $G_1 \equiv (V, E_1)$ and $G_2 \equiv (V, E_2)$ be chain graphs such that $\mathcal{P}_{\mathrm{AMP}}(G_1; \mathbf{X}) = \mathcal{P}_{\mathrm{LWF}}(G_2; \mathbf{X})\ \forall \mathbf{X}$. Then:

(i) $G_1^{\vee} = G_2^{\vee}$;

(ii) G_1 has no biflags;

(iii) G_2 has no multicomplexes.

Proof. (i) Suppose that $G_1^{\vee} \neq G_2^{\vee}$. We shall show that if $a, b \in V$ are such that $a \cdots b$ in G_1 but $a \cdot\!/\!\cdot b$ in G_2, then

$$\left(\mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)\right) \setminus \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V) \neq \emptyset, \tag{B.22}$$

and that if $a \cdots b$ in G_2 but $a \cdot\!/\!\cdot b$ in G_1, then

$$\left(\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)\right) \setminus \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \neq \emptyset. \tag{B.23}$$

Let $P \equiv \mathcal{N}_V(0, \Sigma) \in \mathcal{C}^{**}(\mathbf{R}^V)$ be as defined in the proof of Theorem 6.1 ("only if"). It follows trivially from Lemma 4.2(ii) (resp., (i)) with $A = \{a, b\}$ that if $a \cdots b$ in G_1 (resp., $a \cdots b$ in G_2), then $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$ (resp., $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$). Furthermore, if $a \cdot\!/\!\cdot b$ in G_2 (resp., $a \cdot\!/\!\cdot b$ in G_1), then, since $a \not\perp\!\!\!\perp b\,[P]$, it is readily verified that $P \notin \mathcal{P}^{\mathrm{p}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$ (resp., $P \notin \mathcal{P}^{\mathrm{p}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$), hence $P \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$ by (3.5) (resp., $P \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$ by Theorem 4.2(i)). Thus (B.22) and (B.23) hold.

(ii) Suppose that $G_1^\vee = G_2^\vee$ but G_1 has a k-biflag of the type $[a; c_1, \ldots, c_k]$, $k \geq 3$ (Figure 2.5a). Either $c_1 \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$ or $c_k \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_1))$, so by symmetry we may assume the former. We shall show that (B.22) again holds.

Let P denote the distribution of a normal random vector $X \in \mathbf{R}^V$ such that $\mathrm{Cov}(X_A)$ is nonsingular and $X_v = 0$ for $v \in V \setminus A$, where $A := \{a, c_1, \ldots, c_k\}$, so that $P \in \mathcal{C}^{**}(\mathbf{R}^V)$. First we assert that if $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$ then

$$c_1 \perp\!\!\!\perp c_k \mid S\,[P], \tag{B.24}$$

where $S := \{c_2, \ldots, c_{k-1}\} \cap \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$. ($S$ depends on G_2 but not on P.) Begin by noting that $c_1 \cdot\!\!\not\!-\!\!\cdot c_k$ in G_2 since $G_1^\vee = G_2^\vee$. Thus, by (3.5) and (3.4) with $(v, w) = (c_k, c_1)$ and by the structure of X,

$$c_1 \perp\!\!\!\perp c_k \mid \big(\mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k)) \setminus \{c_1, c_k\}\big) \cap A\,[P]. \tag{B.25}$$

Either $a \notin \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$ or $a \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$. In the first case, (B.24) follows from (B.25). In the second case it follows from (B.25) that

$$c_1 \perp\!\!\!\perp c_k \mid aS\,[P], \tag{B.26}$$

while, since $a \cdot\!\!\not\!-\!\!\cdot c_k$ in G_2, it also follows from (3.5) and (3.4) with $(v, w) = (c_k, a)$ that

$$a \perp\!\!\!\perp c_k \mid \big(\mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k)) \setminus \{a, c_k\}\big) \cap A\,[P], \tag{B.27}$$

whence

$$a \perp\!\!\!\perp c_k \mid c_1 S\,[P]. \tag{B.28}$$

Because P satisfies CI5, (B.26) and (B.28) together imply that (B.24) holds here also.

Now specify $X_A \equiv (X_v | v \in A)$ as follows: $X_a = Y$, $X_{c_1} = U_1$, $X_{c_i} = U_1 + \cdots U_i + Y$ for $i = 2, \ldots, k-1$, and $X_{c_k} = U_1 + \cdots U_k$, where $U_1, \ldots, U_k, Y$ are mutually independent $\mathcal{N}(0,1)$ random variables. Then $\mathrm{Cov}(X_A)$ is nonsingular, and it is readily verified by means of (5.2)-(5.4) that $P_A \in \mathbf{N}_A(0, (G_1)_A)$ (cf. (5.1)), hence $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)$ by Lemma 4.2(ii). However, P does not satisfy (B.24): if $S = \emptyset$ then

$$\mathrm{Cov}(X_{c_1}, X_{c_k} \mid X_S) = \mathrm{Cov}(U_1, U_1) > 0, \tag{B.29}$$

contradicting (B.24), while if $S \neq \emptyset$ then it is straightforward to show that

$$\mathrm{Cov}(X_{c_1}, X_{c_k} \mid X_S) = -\mathrm{Cov}(U_1, Y \mid U_1 + \cdots + U_l + Y) > 0, \tag{B.30}$$

where $l = \min\{i | c_i \in S\}$ $(2 \leq l \leq k-1)$, again contradicting (B.24). Thus $P \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$, hence (B.22) holds.

Suppose now that $G_1^\vee = G_2^\vee$ but G_1 has a k-biflag of the type $[a, b; c_1, \ldots, c_k]$, $k \geq 2$ (Figures 2.4d and 2.5b). By symmetry we may again assume that $c_1 \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$.

To see that (B.22) holds here, define P and X as above (B.24) but with $A = \{a, b, c_1, \ldots, c_k\}$. We now assert that there exists a subset $S \subseteq \{b, c_2, \ldots, c_{k-1}\}$ (possibly empty) depending only on G_2 such that if $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$, then

$$a \perp\!\!\!\perp c_k \mid c_1 S \, [P]. \tag{B.31}$$

Because $G_1^{\vee} = G_2^{\vee}$, $a \not\sim c_k$ in G_2. Either (a) $a \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$ or (b) $a \notin \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$. In case (a), it follows from (3.4) with $(v, w) = (c_k, a)$ and from the structure of X that

$$a \perp\!\!\!\perp c_k \mid \big(\mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k)) \setminus \{a, c_k\}\big) \cap A \, [P]. \tag{B.32}$$

which yields (B.31) with $S := \{b, c_2, \ldots, c_{k-1}\} \cap \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$. (Recall that $c_1 \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$). In case (b) $c_k \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(a))$, so we can apply (3.4) with $(v, w) = (a, c_k)$ to obtain

$$a \perp\!\!\!\perp c_k \mid \big(\mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(a)) \setminus \{a, c_k\}\big) \cap A \, [P]. \tag{B.33}$$

Either (c) $c_1 \notin \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(a))$, which is impossible under (b) since $c_1 \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(c_k))$ and G_2 is adicyclic, or (d) $c_1 \in \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(a))$. If (d) holds then (B.33) yields (B.31) with $S := \{b, c_2, \ldots, c_{k-1}\} \cap \mathrm{Nd}_{\mathcal{D}(G_2)}(\tau(a))$, so the assertion is established.

Now specify $X_A \equiv (X_v | v \in A)$ as follows: $X_a = Y$, $X_b = Z$, $X_{c_1} = U_1 + Y$, $X_{c_i} = U_1 + \cdots U_i + Y + Z$ for $i = 2, \ldots, k-1$, and $X_{c_k} = U_1 + \cdots + U_k + Z$, where $U_1, \ldots, U_k, Y, Z$ are mutually independent $\mathcal{N}(0,1)$ random variables. Then $\mathrm{Cov}(X_A)$ is nonsingular, and it is readily verified by means of (5.2)-(5.4) that $P_A \in \mathbf{N}_A(0, (G_1)_A)$ (cf. (5.1)), hence $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)$ by Lemma 4.2(ii). However, P does not satisfy (B.31), since it is straightforward to verify that

$$\mathrm{Cov}(X_a, X_{c_k} \mid X_{c_1}, X_S) = \mathrm{Cov}(U_1, Y \mid U_1 + Y) < 0. \tag{B.34}$$

Thus $P \notin \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2; \mathbf{R}^V)$, hence (B.22) also holds in this case.

(iii) Suppose that $G_1^{\vee} = G_2^{\vee}$ but G_2 has a k-complex $(a, b; c_1, \ldots, c_k)$, $k \geq 2$ (Figure 2.2). We shall show that (B.23) holds.

Define P and X as above (B.31), again with $A = \{a, b, c_1, \ldots, c_k\}$. We assert that at least one of the following CI relations holds simultaneously for every such $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{AMP}}(G_1; \mathbf{R}^V)$. In each relation S (and i) is fixed and depends only on G_1, not on P:

$$a \perp\!\!\!\perp c_k \mid S \, [P] \quad (S = \emptyset, \{c_1\}, \{c_{k-1}\}, \text{ or } \{b\}); \tag{B.35}$$

$$b \perp\!\!\!\perp c_{k-1} \mid S \, [P] \quad (S = \emptyset \text{ or } \{c_{k-2}\} \text{ (or } \{a\} \text{ (if } k = 2)); \tag{B.36}$$

$$b \perp\!\!\!\perp c_i \quad [P] \quad (i \in \{1, \ldots, k-2\}); \tag{B.37}$$

$$a \perp\!\!\!\perp c_i \mid S \, [P] \quad (S = \emptyset \text{ or } \{c_1\},\ i \in \{2, \ldots, k-1\}); \tag{B.38}$$

$$b \perp\!\!\!\perp c_1 \mid S \, [P] \quad (S = \emptyset, \{c_2\}, \{c_k\}, \text{ or } \{a\}); \tag{B.39}$$

$$a \perp\!\!\!\perp c_2 \mid S \, [P] \quad (S = \emptyset \text{ or } \{c_3\} \text{ (or } \{b\} \text{ (if } k = 2)); \tag{B.40}$$

$$a \perp\!\!\!\perp c_i \quad [P] \quad (i \in \{3, \ldots, k\}); \tag{B.41}$$

$$b \perp\!\!\!\perp c_i \mid S \, [P] \quad (S = \emptyset \text{ or } \{c_k\},\ i \in \{2, \ldots, k-1\}). \tag{B.42}$$

Recall that for any $v \in V$,

$$V = \mathrm{nd}_{\mathcal{D}}(\tau(v)) \dot{\cup} \mathrm{de}_{\mathcal{D}}(\tau(v)) \dot{\cup} \tau(v), \tag{B.43}$$

where $\mathcal{D} \equiv \mathcal{D}(G_1)$ and $\tau(v) \equiv \tau_{G_1}(v)$. Because $a \not\sim c_k$ in G_1, we may set $v = a$ to see that either (1) $c_k \in \mathrm{nd}_{\mathcal{D}}(\tau(a)) \setminus \mathrm{pa}_{G_1}(a)$, (2) $a \in \mathrm{nd}_{\mathcal{D}}(\tau(c_k)) \setminus \mathrm{pa}_{G_1}(c_k)$, or (3) $c_k \in \tau(a) \setminus \mathrm{nb}_{G_1}(a)$. In case (1), it follows from Theorem 4.2(i), property P2$'$ with $(v, w) = (a, c_k)$, and the structure of X, that

$$a \perp\!\!\!\perp c_k \mid \mathrm{pa}_{G_1}(a) \cap A \, [P],$$

which implies (B.35) since $\mathrm{pa}_{G_1}(a) \cap A = \emptyset$ or $\{c_1\}$.

In case (2), it follows similarly from P2$'$ with $(v, w) = (c_k, a)$ that

$$a \perp\!\!\!\perp c_k \mid \mathrm{pa}_{G_1}(c_k) \cap A \, [P],$$

and $\mathrm{pa}_{G_1}(c_k) \cap A \subseteq \{b, c_{k-1}\}$. If $b \notin \mathrm{pa}_{G_1}(c_k)$ or $c_{k-1} \notin \mathrm{pa}_{G_1}(c_k)$ then (B.35) holds with $S = \emptyset$, $\{c_{k-1}\}$, or $\{b\}$. Next, assume that $b, c_{k-1} \in \mathrm{pa}_{G_1}(c_k)$, so $\mathrm{pa}_{G_1}(b) \cap A = \emptyset$ and $\mathrm{pa}_{G_1}(c_{k-1}) \cap A = \emptyset$ or $\{c_{k-2}\}$ (or $\{a\}$ if $k = 2$). By applying (B.43) with $v = b$, we obtain that either (a) $c_{k-1} \in \mathrm{nd}_{\mathcal{D}}(\tau(b)) \setminus \mathrm{pa}_{G_1}(b)$, (b) $b \in \mathrm{nd}_{\mathcal{D}}(\tau(c_{k-1})) \setminus \mathrm{pa}_{G_1}(c_{k-1})$, or (c) $c_{k-1} \in \tau(b) \setminus \mathrm{nb}_{G_1}(b)$. In case (a), apply P2$'$ with $(v, w) = (b, c_{k-1})$ to obtain (B.36) with $S = \emptyset$. In case (b), apply P2$'$ with $(v, w) = (c_{k-1}, b)$ to obtain (B.36). In case (c), apply P1$'$ with $(v, w) = (b, c_{k-1})$ to obtain

$$b \perp\!\!\!\perp c_{k-1} \mid \big(\mathrm{nd}_{\mathcal{D}}(\tau(b)) \dot{\cup} \mathrm{nb}_{G_1}(b)\big) \cap A \, [P].$$

Here $a, c_1, \ldots, c_{k-2} \notin \mathrm{nb}_{G_1}(b)$, while $c_{k-1}, c_k \notin \mathrm{nd}_{\mathcal{D}}(\tau(b)) \dot{\cup} \mathrm{nb}_{G_1}(b)$ by (c) and the assumption that $b \in \mathrm{pa}_{G_1}(c_k)$. Thus (B.36) holds with $S = \emptyset$ or $\{a\}$ if $c_1, \ldots, c_{k-2} \notin \mathrm{nd}_{\mathcal{D}}(\tau(b))$, so assume that $c_i \in \mathrm{nd}_{\mathcal{D}}(\tau(b))$ for at least one $i \in \{1, \ldots, k-2\}$. Since $\mathrm{pa}_{G_1}(b) \cap A = \emptyset$, it follows from P2$'$ with $(v, w) = (b, c_i)$ that (B.37) holds.

In case (3), it follows from P1$'$ with $(v, w) = (a, c_k)$ that

$$a \perp\!\!\!\perp c_k \mid \big(\mathrm{nd}_{\mathcal{D}}(\tau(a)) \dot{\cup} \mathrm{nb}_{G_1}(a)\big) \cap A \, [P].$$

Here $b, c_2, \ldots, c_{k-1} \notin \mathrm{nb}_{G_1}(a)$, so if $b, c_2, \ldots, c_{k-1} \notin \mathrm{nd}_{\mathcal{D}}(\tau(a))$ then (B.35) holds with $S = \emptyset$ or $\{c_1\}$. If $c_i \in \mathrm{nd}_{\mathcal{D}}(\tau(a))$ for some $i \in \{2, \ldots, k-1\}$, then it follows from P2$'$ with $(v, w) = (a, c_i)$ that (B.38) holds, since $\mathrm{pa}_{G_1}(a) \cap A \subseteq \{c_1\}$. Finally, if $b \in \mathrm{nd}_{\mathcal{D}}(\tau(a))$, then by (3) this implies that (α) $b \to c_k \in G_1$. Since the k-complexes $(a, b; c_1, \ldots, c_k)$ and $(b, a; c_k, \ldots, c_1)$ are identical we can now repeat this entire argument, beginning at (B.43), with $a, b, c_1, \ldots, c_k$ replaced by $b, a, c_k, \ldots, c_1$, to conclude that *either* at least one of (B.39), (B.40), (B.41), (B.42) holds, *or else* (3$'$) $c_1 \in \tau(b) \setminus \mathrm{nb}_{G_1}(b)$ and (α') $a \to c_1 \in G_1$. But (3), (α), (3$'$), and (α') together contradict the adicyclicity of G_1. Thus the assertion is established.

Now specify $X_A \equiv (X_v | v \in A)$ as follows: $X_a = Y$, $X_b = Z$, $X_{c_i} = U_1 + \cdots U_i + Y + iZ$ for $i = 1, \ldots, k$, where $U_1, \ldots, U_k, Y, Z$ are mutually independent $\mathcal{N}(0,1)$ random variables. Then $\mathrm{Cov}(X_A)$ is nonsingular, and it can be verified, either by applying Theorem 4.1 of Frydenberg (1990) or by verifying (3.3) for $(P_A, (G_2)_A)$ and applying (3.6), that $P_A \in$

$\mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}((G_2)_A;\mathbf{R}^A)$, hence $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(G_2;\mathbf{R}^V) \cap \mathcal{C}^{**}(\mathbf{R}^V)$ by Lemma 4.2(i). However, P does not satisfy *any* of the relations (B.35) - (B.42). This can be verified by showing that each of the covariances and conditional covariances corresponding to the independences and CIs in (B.35) - (B.42) is nonzero. For example, regarding (B.35),

$$\begin{aligned}
\mathrm{Cov}(X_a, X_{c_k}) &= \mathrm{Cov}(Y,Y) > 0,\\
\mathrm{Cov}(X_a, X_{c_k} \mid X_{c_1}) &= (k-1)\mathrm{cov}(Y, Z \mid U_1 + Y + Z) < 0,\\
\mathrm{Cov}(X_a, X_{c_k} \mid X_{c_{k-1}}) &= \mathrm{cov}(Y, Z \mid U_1 + \cdots + U_{k-1} + Y + (k-1)Z) < 0,\\
\mathrm{Cov}(X_a, X_{c_k} \mid X_b) &= \mathrm{Cov}(Y,Y) > 0.
\end{aligned}$$

The remaining cases are treated similarly. This completes the proof.

Remark B.5. In the final paragraph of the preceding proof, it is somewhat easier to specify $X_A \equiv (X_v | v \in A)$ as follows: $X_a = Y$, $X_b = Z$, $X_{c_i} = W_i + \rho^{i-1}Y + \rho^{k-i}Z$ for $i = 1, \ldots, k$, where $0 < |\rho| < 1$, $W \equiv (W_1, \ldots, W_k)$, Y, and Z are mutually independent, Y and $Z \sim \mathcal{N}(0,1)$, and $W \sim \mathcal{N}_k(0, \Lambda)$ with $\Lambda \equiv (\lambda_{ij})$ a serial correlation matrix given by $\lambda_{ij} = \rho^{|i-j|}$, whose inverse is tridiagonal. Again $\mathrm{Cov}(X_A)$ is nonsingular and $P_A \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}((G_2)_A;\mathbf{R}^A)$, but here the joint distributions of $(X_a, X_b, X_{c_1}, \ldots, X_{c_k})$ and $(X_b, X_a, X_{c_k}, \ldots, X_{c_1})$ are identical. Thus the relations (B.35) - (B.38) are equivalent to (B.39) - (B.42), hence only the former need be considered.

Remark B.6. We conjecture that (B.23) holds in (ii) and (B.22) holds in (iii).

Proof of Theorem 6.2. The "only if" statements in (i) and (ii) follow directly from Proposition B.2(i) and (ii) respectively. As in the proof of Proposition B.1, again use the abbreviations $\mathcal{T} = \mathcal{T}(G)$, $\tilde{\mathcal{T}} = \mathcal{T}(\tilde{G})$, $\mathcal{D} = \mathcal{D}(G)$, and $\tilde{\mathcal{D}} = \mathcal{D}(\tilde{G})$.

(i) "if": As in the second half of the proof of Proposition 6.2, construct the graph $\tilde{G} \subseteq G$ by converting each flag in G into an immorality in $\tilde{G}$. As in that proof, because G has no biflags this construction is unambiguous, $\tilde{G}$ is adicyclic, each triplex in G corresponds to an immorality in $\tilde{G}$, and each immorality in $\tilde{G}$ corresponds to a triplex in G. In order to show that $\mathcal{P}_{\mathrm{LWF}}(\tilde{G};\mathbf{X}) \subseteq \mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X})$, as in the proof of Proposition B.1(i) it suffices to show that if $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(\tilde{G};\mathbf{X})$ then P satisfies conditions C7 and C8 for G.

To establish C7 when $\sigma \subset \tau$, because $P \in \mathcal{P}^{\mathrm{g}}_{\mathrm{LWF}}(\tilde{G};\mathbf{X})$ it suffices to show that $\mathrm{pa}_G(\tau)\dot\cup\mathrm{nb}_G(\sigma)$ separates σ and $\tau\backslash\mathrm{cl}_{G_\tau}(\sigma)$ in $\tilde{H}^{\mathrm{m}}$, where, since $\mathrm{At}_{\tilde{G}}(\mathrm{cl}_G(\tau)) = \mathrm{At}_{\tilde{G}}(\tau)$,

$$\tilde{H} := \tilde{G}\big(\sigma\dot\cup\mathrm{pa}_G(\tau)\dot\cup\mathrm{nb}_G(\sigma)\dot\cup(\tau\backslash\mathrm{cl}_{G_\tau}(\sigma))\big) \equiv \tilde{G}(\mathrm{cl}_G(\tau)) = \tilde{G}(\tau).$$

Since $\mathrm{At}_{\tilde{G}}(A) \subseteq \mathrm{At}_G(A)$ for any $A \subseteq V$, the vertex set of $\tilde{H}$ satisfies

$$\mathrm{At}_{\tilde{G}}(\tau) \subseteq \mathrm{At}_G(\tau). \tag{B.44}$$

We must show that (B.10) holds for any path $\tilde{\pi}$ between σ and $\tau\backslash\mathrm{cl}_{G_\tau}(\sigma)$ in $\tilde{H}^{\mathrm{m}}$.

Let $s \in \tilde{\pi}\cap\sigma$ be the *last* vertex of $\tilde{\pi}$ that is also in σ and let $v \in \tilde{\pi}\backslash\sigma$ be the *next* vertex in $\tilde{\pi}$ after s, so that $s — v \in \tilde{H}^{\mathrm{m}}$ (hence $v \in \mathrm{At}_G(\tau)$ by (B.44)). By the definition of $\tilde{H}^{\mathrm{m}}$, either (1) s—$v \in \tilde{H}$, (2) $s \to v \in \tilde{H}$, (3) $s \leftarrow v \in \tilde{H}$, or (4) $s \not\sim v$ in $\tilde{H}$ but $(s, v; w_1, \ldots, w_k)$ is a complex in $\tilde{H}$ (hence in $\tilde{G}$) for some $w_1, \ldots, w_k \in \mathrm{At}_G(\tau)$, $k \geq 1$. In case (1) s—$v \in \tilde{G}$, so s—$v \in G$ and thus $v \in \tilde{\pi} \cap \mathrm{nb}_G(\sigma)$, implying (B.10). In case (2) $s \to v \in \tilde{G}$, so either $s \to v \in G$ or s—$v \in G$. The former contradicts the adicyclicity of G since $s \in \tau$ and $v \in \mathrm{At}_G(\tau)$, while the latter implies (B.10) as in (1). In case (3) $s \leftarrow v \in \tilde{G}$, so either $s \leftarrow v \in G$ or s—$v \in G$. The former implies that $v \in \tilde{\pi} \cap \mathrm{pa}_G(\tau)$ so (B.10) holds, while as in (1) the latter also implies that (B.10) holds.

In case (4), either $s \to w_1 \in G$, which contradicts the adicyclicity of G, or s—$w_1 \in G$. In the latter case $w_k \in \tau$, so if $w_k \leftarrow v \in G$ then $v \in \mathrm{pa}_G(\tau)$, again implying (B.10). Alternatively, if w_k—$v \in G$ then s—w_1—$\cdots$—w_k—v occurs as a chordless undirected path in G while $s \to w_1$—$\cdots$—$w_k \leftarrow v$ occurs as an induced subgraph in $\tilde{G}$. By the construction of $\tilde{G}$ from G and the adicyclicity of G, this implies that there exist vertices $a, b \in V \setminus \{s, w_1, \ldots, w_k, v\}$ (possibly $a = b$) such that s—$w_1 \leftarrow a$ and $b \to w_k$—v occur as flags in G and such that $a \to w_i \in G$ for $i = 2, \ldots, k$ and $b \to w_i \in G$ for $i = 1, \ldots, k-1$. But this implies that at least one of the biflags $[a; s, w_1, \ldots, w_k, v]$, $[b; s, w_1, \ldots, w_k, v]$, or $[b, a; s, w_1, \ldots, w_k, v]$ occurs in G, a contradiction. Thus C7 holds.

To establish C8, it suffices to show that $\mathrm{pa}_G(\sigma)$ separates σ and $\sigma' := \mathrm{nd}_{\mathcal{D}}(\tau) \setminus \mathrm{pa}_G(\sigma)$ in $\tilde{H}^{\mathrm{m}}$, where

$$\tilde{H} := \tilde{G}(\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau)).$$

Since $\mathrm{At}_G(\sigma) = \mathrm{At}_G(\tau)$ and $\mathrm{Nd}_{\mathcal{D}}(\tau) \equiv \tau \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau)$ is G-anterior, the vertex set of $\tilde{H}$ satisfies

$$\mathrm{At}_{\tilde{G}}(\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau)) \subseteq \mathrm{Nd}_{\mathcal{D}}(\tau). \tag{B.45}$$

We must show that (B.12) holds for any path $\tilde{\pi}$ between σ' and σ in $\tilde{H}^{\mathrm{m}}$

Let $s \in \tilde{\pi} \cap \sigma'$ be the *last* vertex of $\tilde{\pi}$ also in σ' and let $v \in \tilde{\pi} \setminus \sigma' \subseteq \mathrm{At}_{\tilde{G}}(\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau))$ be the *next* vertex in $\tilde{\pi}$ after s, so that s—$v \in \tilde{H}^{\mathrm{m}}$. By (B.45), $v \in \mathrm{Nd}_{\mathcal{D}}(\tau) \setminus \sigma' = \tau \dot{\cup} \mathrm{pa}_G(\sigma)$; since (B.12) holds if $v \in \mathrm{pa}_G(\sigma)$, we may assume that $v \in \tau$. Either (1) s—$v \in \tilde{H}$, (2) $s \leftarrow v \in \tilde{H}$, (3) $s \to v \in \tilde{H}$, or (4) $s \not\sim v$ in $\tilde{H}$ but $(s, v; w_1, \ldots, w_k)$ is a complex in $\tilde{H}$ (hence in $\tilde{G}$) for some $w_1, \ldots, w_k \in \mathrm{At}_{\tilde{G}}(\sigma \dot{\cup} \mathrm{nd}_{\mathcal{D}}(\tau))$, $k \geq 1$. In case (1) s—$v \in \tilde{G}$ so s—$v \in G$, which is impossible since $s \notin \tau$. In case (2) $s \leftarrow v \in \tilde{G}$, so either s—$v \in G$, which is impossible as in (1), or $s \leftarrow v \in G$, which is impossible since $s \in \sigma' \subseteq \mathrm{nd}_{\mathcal{D}}(\tau)$.

In case (3) $s \to v \in \tilde{G}$, hence either s—$v \in G$ which is impossible as in case (1), or else (*) $s \to v \in G$. In the latter case $v \notin \sigma$ since $s \in \sigma'$, so $v \in \tau \setminus \sigma$. Therefore $v \in \mathrm{At}_{\tilde{G}}(\sigma) \setminus \sigma$, so there exists a path $\phi \equiv (v, v_1, \ldots, v_n)$ of length $n \geq 1$ in $\tilde{G}$ from v to some $v_n \in \sigma$. Since $v \in \tau$ and $\tilde{G} \subseteq G$, the subgraph $s \to v$—v_1—$\cdots$—v_n must occur in G. If $s \not\sim v_1$ in G then by the construction of $\tilde{G}$, $v \leftarrow v_1 \in \tilde{G}$, contradicting the definition of ϕ, hence $s \to v_1 \in G$ by the adicyclicity of G. Similarly, $s \to v_i \in G$ for $i = 2, \ldots, n$, but $s \to v_n \in G$ contradicts the fact that $s \notin \mathrm{pa}_G(\sigma)$.

In case (4), either $w_k \leftarrow v \in G$, which contradicts the adicyclicity of G by (B.45), or w_k—$v \in G$. In the latter case $w_1 \in \tau$, so if s—w_1 then $s \in \tau$, contradicting $s \in \sigma'$. Alternatively, if $s \to w_1 \in G$, apply the argument in (3) beginning at (*) with v replaced by w_1 to obtain a similar contradiction. Thus C8 holds, so $\mathcal{P}_{\mathrm{LWF}}(\tilde{G}; \mathbf{X}) \subseteq \mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X})$.

To show that $\mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X}) \subseteq \mathcal{P}_{\mathrm{LWF}}(\tilde{G};\mathbf{X})$, it suffices to show that if $P \in \mathcal{P}_{\mathrm{AMP}}(G;\mathbf{X})$ then P satisfies condition C4 for $\tilde{G}$, which we denote by $\tilde{\mathrm{C4}}$:

$$(\tilde{\mathrm{C4}})\quad \forall\tilde{\tau} \in \tilde{\mathcal{T}}, \forall\tilde{\sigma} \subseteq \tilde{\tau}:\quad \tilde{\sigma} \perp\!\!\!\perp \left(\mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \setminus \mathrm{cl}_{\tilde{G}}(\tilde{\sigma})\right) \,\Big|\, \mathrm{bd}_{\tilde{G}}(\tilde{\sigma})\,[P].$$

Thus it suffices to show that $\mathrm{bd}_{\tilde{G}}(\tilde{\sigma}) \equiv \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})\dot{\cup}\mathrm{nb}_{\tilde{G}}(\tilde{\sigma})$ separates $\tilde{\sigma}$ and $\mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \setminus \mathrm{cl}_{\tilde{G}}(\tilde{\sigma})$ in H^{a}, where, since $\mathrm{An}_G(\mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})) = \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$,

$$H := G[\mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})] = G_{\mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})} \cup G^{\wedge}_{\mathrm{Co}_G(\mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}))}.$$

We must show that for any path $\pi =: (v_0, v_1, \ldots, v_n)$ between some $v_0 \in \tilde{\sigma}$ and some $v_n \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau}) \setminus \mathrm{cl}_{\tilde{G}}(\tilde{\sigma})$ in H^{a},

$$\pi \cap \left(\mathrm{pa}_{\tilde{G}}(\tilde{\sigma})\dot{\cup}\mathrm{nb}_{\tilde{G}}(\tilde{\sigma})\right) \neq \emptyset. \tag{B.46}$$

First assume that $n = 1$, so $v_1 \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. Because $v_0 \in \tilde{\sigma} \subseteq \tilde{\tau}$, $v_0 \to v_1$ cannot occur in $\tilde{G}$. Since $v_0 - v_1 \in H^{\mathrm{a}}$, either (1) $v_0 \to v_1 \in H$, (2) $v_0 \leftarrow v_1 \in H$, (3) $v_0 - v_1 \in H$, or (4) $v_0 \not\sim v_1$ in H but either $(\{v_0, v_1\}, w)$ is a triplex in H for some $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ or $[v_0, v_1; w, y]$ is a 2-biflag in H for some $w, y \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ (Figure A.3 with s, v replaced by v_0, v_1). In case (1) $v_0 \to v_1 \in G$ so $v_0 \to v_1 \in \tilde{G}$, a contradiction. In case (2) $v_0 \leftarrow v_1 \in G$ so $v_0 \leftarrow v_1 \in \tilde{G}$, hence $v_1 \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$, so (B.46) holds. In case (3) $v_0 - v_1 \in G$, so either $v_0 \leftarrow v_1 \in \tilde{G}$ or $v_0 - v_1 \in \tilde{G}$, hence $v_1 \in \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$ or $v_1 \in \mathrm{nb}_{\tilde{G}}(\tilde{\sigma})$, respectively, so (B.46) holds.

In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 (with s, v replaced by v_0, v_1) must occur as an induced subgraph of H. In cases (a), (b), and (d), $v_0 \to w \in G$ hence $v_0 \to w \in \tilde{G}$, but $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$, a contradiction as in (1). In case (c), $v_1, w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ and $v_0 \in \tilde{\sigma} \subseteq \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$, so $v_0 - w \leftarrow v_1$ must occur as an induced subgraph (a flag) of G. By the construction of $\tilde{G}$, $v_0 \to w \in \tilde{G}$, again a contradiction as in (1).

Now assume that $n \geq 2$ but that (B.46) does not hold. We assert that there exists a subset $\{v'_1, \ldots, v'_m\} \subseteq \{v_1, \ldots, v_n\}$ such that $\pi' := (v'_0, v'_1, \ldots, v'_m)$ is a chordless undirected path from $v'_0 \equiv v_0$ to $v'_m \equiv v_n$ in G that is semi-directed in $\tilde{G}$ with $v'_0 \to v'_1 \in \tilde{G}$. But this is impossible since $v_n \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$, so we will conclude that (B.46) must hold.

By an argument similar to that for $n = 1$, $v'_0 - v_1 \in G$ and $v'_0 \to v_1 \in \tilde{G}$. Set $v'_1 = v_1$ (possibly temporarily), so $v'_1 \notin \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. Now $v'_1 - v_2 \in H^{\mathrm{a}}$, so either (1) $v'_1 \to v_2 \in H$, (2) $v'_1 \leftarrow v_2 \in H$, (3) $v'_1 - v_2 \in H$, or (4) $v'_1 \not\sim v_2$ in H but either $(\{v'_1, v_2\}, w)$ is a triplex in H for some $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ or $[v'_1, v_2; w, y]$ is a 2-biflag in H for some $w, y \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 (with s, v replaced by v'_1, v_2) must occur as an induced subgraph of H. Cases (1), (2), (4a), (4b), and (4d) are impossible since $v'_1 \notin \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$.

In case (4c), $v'_0 - v'_1 - w \leftarrow v_2$ must occur as a subgraph (not necessarily induced) of G, while either $v'_0 \to v'_1 \to w \leftarrow v_2$, $v'_0 \to v'_1 - w \leftarrow v_2$, or $v'_0 \to v'_1 \leftarrow w \leftarrow v_2$ must occur as a subgraph of $\tilde{G}$. The first two cases are impossible since $v'_0 \in \tilde{\sigma} \subseteq \tilde{\tau}$ and $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. In the third case, if $v'_0 \cdots w$ in G then $v'_0 - w \in G$ by the adicyclicity of G, while either $v'_0 - w \in \tilde{G}$ or $v'_0 \leftarrow w \in \tilde{G}$ since $v'_0 \in \tilde{\tau}$ and $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. In both instances, by the construction of $\tilde{G}$ the flag $[v_2, v'_0; w]$ cannot occur in G, hence $v'_0 \leftarrow v_2 \in G$ and so $v_2 \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$, contradicting the assumption that (B.46) fails. If, however, $v'_0 \not\cdots w$ in G then G must possess a 3-biflag of the form $[a; v'_0, v'_1, w]$ or $[a, b; v'_0, v'_1, w]$, again a contradiction.

In case (3), $v'_1 — v_2 \in G$. If $v'_0 \cdots v_2$ in G then $v'_0 — v_2 \in G$ by the adicyclicity of G, while $v'_0 \to v_2 \in \tilde{G}$ since (B.46) is assumed to be false, so in this case we may redefine $v'_1 = v_2$. Now assume that $v'_0 \not\cdots v_2$ in G, so either $v'_0 \to v'_1 \leftarrow v_2$, $v'_0 \to v'_1 — v_2$, or $v'_0 \to v'_1 \to v_2$ must occur as an induced subgraph of $\tilde{G}$. In the first case a 3-biflag of the form $[a; v'_0, v'_1, v_2]$ or $[a, b; v'_0, v'_1, v_2]$ must occur in G, contradicting our hypothesis. In the second and third cases, we may define $v'_2 = v_2$, so that (v'_0, v'_1, v'_2) forms a chordlesss undirected path in G that is semi-directed in $\tilde{G}$ with $v'_0 \to v'_1$. Thus if $n = 2$, the required assertion is thereby established.

If $n \geq 3$, we may assume inductively that for some $k \in \{2, \ldots, n-1\}$ we have constructed a subset $\{v'_1, \ldots, v'_i\} \subseteq \{v_1, \ldots, v_k\}$ with $v'_i = v_k$ such that $(v'_0, v'_1, \ldots, v'_i)$ is a chordless undirected path in G that is semi-directed in $\tilde{G}$ with $v'_0 \to v'_1 \in \tilde{G}$, so $v'_i \notin \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. Now $v'_i — v_{k+1} \in H^{\mathrm{a}}$, so either (1) $v'_i \to v_{k+1} \in H$, (2) $v'_i \leftarrow v_{k+1} \in H$, (3) $v'_i — v_{k+1} \in H$, or (4) $v'_i \not\cdots v_{k+1}$ in H but either $(\{v'_i, v_{k+1}\}, w)$ is a triplex in H for some $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$ or $[v'_i, v_{k+1}; w, y]$ is a 2-biflag in H for some $w, y \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. In case (4), one of the four configurations (a), (b), (c), or (d) in Figure A.3 (with s, v replaced by v'_i, v_{k+1}) must occur as an induced subgraph of H. Cases (1), (2), (4a), (4b), and (4d) are impossible since $v'_i \notin \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$.

In case (4c), $v'_i — w \leftarrow v_{k+1}$ must occur as a subgraph (not necessarily induced) of G, while either $v'_i \to w \leftarrow v_{k+1}$, $v'_i — w \leftarrow v_{k+1}$, or $v'_i \leftarrow w \leftarrow v_{k+1}$ must occur as a subgraph of $\tilde{G}$. The first two cases are impossible since $v'_0 \in \tilde{\tau}$ and $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$. In the third case, define

$$r := \min\{j | 0 \leq j \leq i, v'_j \cdots w \in G\}.$$

If $r = 0$ then $v'_0 — w \in G$ by the adicyclicity of G, while, since $v'_0 \in \tilde{\tau}$ and $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$, either $v'_0 — w \in \tilde{G}$ or $v'_0 \leftarrow w \in \tilde{G}$. In both instances the flag $[v_{k+1}, v'_0; w]$ cannot occur in G by the construction of $\tilde{G}$, hence $v'_0 \leftarrow v_{k+1} \in G$ and so $v_{k+1} \in \pi \cap \mathrm{pa}_{\tilde{G}}(\tilde{\sigma})$, contradicting the assumption that (B.46) fails. If $1 \leq r \leq i-1$, then $v'_r — w \in G$ by the adicyclicity of G, while $v'_r \leftarrow w \in \tilde{G}$ because $v'_0 \in \tilde{\tau}$ and $w \in \mathrm{Nd}_{\tilde{\mathcal{D}}}(\tilde{\tau})$; clearly these relations also hold if $r = i$. If we now define

$$l := \max\{j | 1 \leq j \leq r, v'_{j-1} \to v'_j \in \tilde{G}\},$$

then an $(r-l+3)$-biflag of the form $[a; v'_{l-1}, v'_l, \ldots, v'_r, w]$ or $[a, b; v'_{l-1}, v'_l, \ldots, v'_r, w]$ must occur in G, again impossible.

In case (3), $v'_i — v_{k+1} \in G$. Define r as above but with w replaced by v_{k+1}, so $v'_r — v_{k+1} \in G$ by the adicyclicity of G if $r \leq i-1$ or by (3) if $r = i$. If $r = 0$ then $v'_0 \to v_{k+1} \in \tilde{G}$ since (B.46) is assumed false, so we may redefine $v'_1 = v_{k+1}$. If $1 \leq r \leq i$, then either $v'_r \leftarrow v_{k+1} \in \tilde{G}$, $v'_r — v_{k+1} \in \tilde{G}$, or $v'_r \to v_{k+1} \in \tilde{G}$. In the first case, an $(r-l+3)$-biflag of the form $[a; v'_{l-1}, v'_l, \ldots, v'_r, w]$ or $[a, b; v'_{l-1}, v'_l, \ldots, v'_r, w]$ must occur in G, where l is as defined above. In the second and third cases, we may define $v'_{i+1} = v_{k+1}$ if $r = i$ or redefine $v'_{r+1} = v_{k+1}$ if $r < i$. Thus in all possible cases we have constructed a subset $\{v'_1, \ldots, v'_h\} \subseteq \{v_1, \ldots, v_{k+1}\}$ with $v'_h = v_{k+1}$ such that $(v'_0, v'_1, \ldots, v'_h)$ is a chordless undirected path in G that is semi-directed in $\tilde{G}$ with $v'_0 \to v'_1 \in \tilde{G}$.

Now proceed by induction on k to establish the required assertion and thereby establish (B.46). Thus $\tilde{C4}$ holds, so $\mathcal{P}_{\mathrm{AMP}}(G; \mathbf{X}) \subseteq \mathcal{P}_{\mathrm{LWF}}(\tilde{G}; \mathbf{X})$. This completes the proof.

Acknowledgements. We wish to thank Wray Buntine, David Cox, Steven Gillispie, Clark Glymour, Soren Højsgaard, Steffen Lauritzen, Michael Levitz, Chris Meek, Thomas Richardson, Peter Spirtes, Milan Studený, Chris Triggs, Nanny Wermuth, and Joe Whittaker for helpful comments and suggestions.

References

Andersson, S. A., Madigan, D., and Perlman, M. D. (1996). An alternative Markov property for chain graphs. In *Uncertainty in Artificial Intelligence: Proceedings of the Twelfth Conference* (F. Jensen and E. Horvitz, eds.), 40-48. Morgan Kaufmann, San Francisco.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997a). On the Markov equivalence of chain graphs, undirected graphs, and acyclic digraphs. *Scand. J. Statist.* **24** 81-102.

Andersson, S. A., Madigan, D., and Perlman, M. D. (1997b). A characterization of Markov equivalence classes for acyclic digraphs. *Ann. Statist.* **25** 505-541.

Andersson, S. A. and Perlman, M. D. (1998). Normal linear regression models with recursive graphical Markov structure. *J. Multivariate Analysis*, **66** 133-187.

Bouckaert, R. and Studený, M. (1995). Chain graphs: semantics and expressiveness. In *Symbolic and Qualitative Approaches to Reasoning and Uncertainty* (Ch. Froidevaux and J. Kohlas, eds.) 69-76. Springer-Verlag, Berlin.

Buntine, W. L. (1995). Chain graphs for learning. In *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference* (P. Besnard and S. Hanks eds.), 46-54. Morgan Kaufmann, San Francisco.

Chickering, D. M. (1995). A transformational characterization of equivalent Bayesian network structures. In *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference* (P. Besnard and S. Hanks, eds.), 87-98. Morgan Kaufman, San Francisco.

Cox, D. R. (1993). Causality and graphical models. *Bull. Internat. Statist. Inst.*, Proceedings 49th Session, **1** 363-372.

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. (1980). Conditional independence for statistical operations. *Ann. Statist.* **8** 1272-1317.

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

Dempster, A. P. (1972). Covariance selection. *Biometrics* **28** 157-175.

Edwards, D. (1995). *Introduction to Graphical Modelling.* Springer, New York.

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

Geiger, D. and Pearl, J. (1988). On the logic of influence diagrams. *Proc. 4th Workshop on Uncertainty in Artificial Intelligence* (R. Shachter, T. Levitt, L. Kanal, J. Lemmer, eds.), 136-147. North-Holland, Amsterdam.

Højsgaard, S. and Thiesson, B. (1995). BIFROST - Block recursive models induced from relevant knowledge, observations, and statistical techniques. *Comp. Statist. and Data Analysis* **19** 155-175.

Kauermann, G. (1996). On a dualization of graphical Gaussian models. *Scand. J. Statist.* **23** 105-116.

Koster, J. T. A. (1996). Markov properties of non-recursive causal models. *Ann. Statist.* **24**, 2148-2177.

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

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., Speed, T. P., and Vijayan, K. (1984). Decomposable graphs and hypergraphs. *J. Austral. Math. Soc. Series A* **36** 12-29.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion.) *J. Royal Statist. Soc. Series 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. M. (1996). Bayesian model averaging and model selection for Markov equivalence classes of acyclic digraphs. *Commun. Statist. - Theory Meth.* **25** 2493-2519.

Madigan, D., Perlman, M. D., and Levitz, M. (1998). A new pathwise separation criterion for chain graphs. In preparation.

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

Pearl, J. and Verma, T. (1991). A theory of inferred causation. In *Proceedings of the Second International Conference on Principles of Knowledge Representation* (J. Allen, R. Fikes, E. Sandewall, eds.), 441-452. Morgan Kaufmann, San Francisco.

Shachter, R. D. and Kenley, C. R. (1989). Gaussian influence diagrams. *Management Sci.* **35** 527-550.

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.

Spirtes, P. (1995). Directed cyclic graphical representation of feedback models. In *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference* (P. Besnard and S. Hanks, eds.), 491-498. Morgan Kaufmann, San Francisco.

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction, and Search.* Springer, New York.

Studený, M. (1996). On separation criterion and recovery algorithm for chain graphs. In *Uncertainty in Artificial Intelligence: Proceedings of the Twelfth Conference* (F. Jensen and E. Horvitz, eds.), 509-516. Morgan Kaufmann, San Francisco.

Studený, M. (1997). A recovery algorithm for chain graphs. *Internat. J. Approximate Reasoning* **17** 265-293.

Studený, M. and Bouckaert, R. R. (1998). On chain graph models for description of conditional independence structure. *Ann. Statist.* **26** 1434-1495.

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. Wellman, B. D'Ambrosio, 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.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias. *J. Amer. Statist. Assoc.* **57** 348-368.