To appear in 'Learning in Graphical Models'
M.I.T. Press, 1998, M. Jordan (Ed.)

NNM for prepare

NI97041-NNM

CHAIN GRAPHS AND SYMMETRIC ASSOCIATIONS

THOMAS S. RICHARDSON
Statistics Department
University of Washington
tsr@stat.washington.edu

Abstract. Graphical models based on chain graphs, which admit both directed and undirected edges, were introduced by by Lauritzen, Wermuth and Frydenberg as a generalization of graphical models based on undirected graphs, and acyclic directed graphs. More recently Andersson, Madigan and Perlman have given an alternative Markov property for chain graphs. This raises two questions: How are the two types of chain graphs to be interpreted? In which situations should chain graph models be used and with which Markov property?

The undirected edges in a chain graph are often said to represent 'symmetric' relations. Several different symmetric structures are considered, and it is shown that although each leads to a different set of conditional independences, none of those considered corresponds to either of the chain graph Markov properties.

The Markov properties of undirected graphs, and directed graphs, including latent variables and selection variables, are compared to those that have been proposed for chain graphs. It is shown that there are qualitative differences between these Markov properties. As a corollary, it is proved that there are chain graphs which do not correspond to any cyclic or acyclic directed graph, even with latent or selection variables.

1. Introduction

The use of acyclic directed graphs (often called 'DAG's) to simultaneously represent causal hypotheses and to encode independence and conditional independence constraints associated with those hypotheses has proved fruitful in the construction of expert systems, in the development of efficient updating algorithms (Pearl [22]; Lauritzen and Spiegelhalter [19]), and in inferring causal structure (Pearl and Verma [25]; Cooper and Herskovits [5]; Spirtes, Glymour and Scheines [31]).

Likewise, graphical models based on undirected graphs, also known as Markov random fields, have been used in spatial statistics to analyze data from field trials, image processing, and a host of other applications (Hammersley and Clifford [13]; Besag [4]; Speed [29]; Darroch *et al.* [8]). More recently, chain graphs, which admit both directed and undirected edges have been proposed as a natural generalization of both undirected graphs and acyclic directed graphs (Lauritzen and Wermuth [20]; Frydenberg [11]). Since acyclic directed graphs and undirected graphs can both be regarded as special cases of chain graphs it is undeniable that chain graphs are a generalization in this sense.

The introduction of chain graphs has been justified on the grounds that this admits the modelling of 'simultaneous responses' (Frydenberg [11]), 'symmetric associations' (Lauritzen and Wermuth [20]) or simply 'associative relations', as distinct from causal relations (Andersson, Madigan and Perlman [1]). The existence of two different Markov properties for chain graphs raises the question of what *sort* of symmetric relation is represented by a chain graph under a given Markov property, since the two properties are clearly different. A second related question concerns whether or not there are modelling applications for which chain graphs are particularly well suited, and if there are, which Markov property is most appropriate.

One possible approach to clarifying this issue is to begin by considering causal systems, or data generating processes, which have a symmetric structure. Three simple, though distinct, ways in which two variables, X and Y, could be related symmetrically are: (a) there is an unmeasured, 'confounding', or 'latent' variable that is a common cause of both X and Y; (b) X and Y are both causes of some 'selection' variable (conditioned on in the sample); (c) there is feedback between X and Y, so that X is a cause of Y, and Y is a cause of X. In fact situations (a) and (b) can easily be represented by DAGs through appropriate extensions of the formalism (Spirtes, Glymour and Scheines [31]; Cox and Wermuth [7]; Spirtes, Meek and Richardson [32]). In addition, certain kinds of linear feedback can also be modelled with directed cyclic graphs (Spirtes [30]; Koster [16]; Richardson [26, 27, 28]; Pearl and Dechter [24]). Each of these situations leads to a different set of conditional independences. However, perhaps surprisingly, none of these situations, nor any combination of them, lead in general to either of the Markov properties associated with chain graphs.

The remainder of the paper is organized as follows: Section 2 contains definitions of the various graphs considered and their associated Markov properties. Section 3 considers two simple chain graphs, under both the original Markov property proposed by Lauritzen, Wermuth and Frydenberg, and the alternative given by Andersson, Madigan and Perlman. These are compared to the corresponding directed graphs obtained by replacing the undirected edges with directed edges in accordance with situations (a), (b) and (c) above. Section 4 generalizes the results of the previous section: two properties are presented, motivated by causal and spatial intuitions, that the set of conditional independences entailed by a graphical model might satisfy. It is shown that the sets of independences entailed by (i) an undirected graph via separation, and (ii) a (cyclic or acyclic) directed graph (possibly with latent and/or selection variables) via d-separation, satisfy both properties. By contrast neither of these properties, in general, will hold in a chain graph under the Lauritzen-Wermuth-Frydenberg (LWF) interpretation. One property holds for chain graphs under the Andersson-Madigan-Perlman (AMP) interpretation, the other does not. Section 5 contains a discussion of data-generating processes associated with different graphical models, together with a brief sketch of the causal intervention theory that has been developed for directed graphs. Section 6 is the conclusion, while proofs not contained in the main text are given in Section 7.

2. Graphs and Probability Distributions

This section introduces the various kinds of graph considered in this paper, together with their associated Markov properties.

2.1. UNDIRECTED AND DIRECTED GRAPHS

An *undirected graph*, UG, is an ordered pair $(\mathbf{V}, \mathbf{U})$, where $\mathbf{V}$ is a set of vertices and $\mathbf{U}$ is a set of undirected edges $X - Y$ between vertices.[1]

Similarly, a *directed graph*, DG, is an ordered pair $(\mathbf{V}, \mathbf{D})$ where $\mathbf{D}$ is a set of directed edges $X \rightarrow Y$ between vertices in $\mathbf{V}$. A *directed cycle* consists of a sequence of n distinct edges $X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_n \rightarrow X_1$ $(n \geq 2)$. If a directed graph, DG, contains no directed cycles it is said to be *acyclic*, otherwise it is *cyclic*. An edge $X \rightarrow Y$ is said to be *out of* X and *into* Y; X and Y are the *endpoints* of the edge. Note that if cycles are permitted there may be more than one edge between a given pair of vertices e.g. $X \leftarrow Y \leftarrow X$. Figure 1 gives examples of undirected and directed graphs.

2.2. DIRECTED GRAPHS WITH LATENT VARIABLES AND SELECTION VARIABLES

Cox and Wermuth [7] and Spirtes *et al.* [32] introduce directed graphs in which $\mathbf{V}$ is partitioned into three disjoint sets $\mathbf{O}$ (Observed), $\mathbf{S}$ (Selection)

[1] Bold face $(\mathbf{X})$ denote sets; italics (X) denote individual vertices; greek letters (π) denote paths.

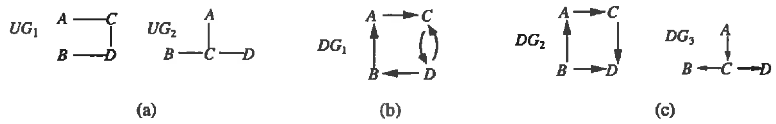

Figure 1. (a) undirected graphs; (b) a cyclic directed graph; (c) acyclic directed graphs

and **L** (Latent), written $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$ (where DG may be cyclic). The interpretation of this definition is that DG represents a causal or data-generating mechanism; **O** represents the subset of the variables that are observed; **S** represents a set of *selection* variables which, due to the nature of the mechanism selecting the sample, are conditioned on in the subpopulation from which the sample is drawn; the variables in **L** are not observed and for this reason are called *latent*.[2]

Example: Randomized Trial of an Ineffective Drug with Unpleasant Side-Effects[3]

A simple causal mechanism containing latent and selection variables is given in Figure 2. The graph represents a randomized trial of an ineffective drug with unpleasant side-effects. Patients are randomly assigned to the treatment or control group (A). Those in the treatment group suffer unpleasant side-effects, the severity of which is influenced by the patient's general level of health (H), with sicker patients suffering worse side-effects. Those patients who suffer sufficiently severe side-effects are likely to drop out of the study. The selection variable (Sel) records whether or not a patient remains in the study, thus for all those remaining in the study $Sel = Stay\ In$. Since unhealthy patients who are taking the drug are more likely to drop out, those patients in the treatment group who remain in the study tend to be healthier than those in the control group. Finally health status (H) influences how rapidly the patient recovers. This example is of interest because, as should be intuitively clear, a simple comparison of the recovery time of the patients still in the treatment and control groups at the end of the study will indicate faster recovery among those in the treatment group. This comparison falsely indicates that the drug has a beneficial effect, whereas in fact, this difference is due entirely to the side-effects causing the sicker patients in the treatment group to drop out of the study.[4] (The only difference between the two graphs in Figure 2 is that in $DG_1(\mathbf{O}_1, \mathbf{S}_1, \mathbf{L}_1)$

[2]Note that the terms *variable* and *vertex* are used interchangeably.

[3]I am indebted to Chris Meek for this example.

[4]For precisely these reasons, in real drug trials investigators often go to great lengths to find out why patients dropped out of the study.

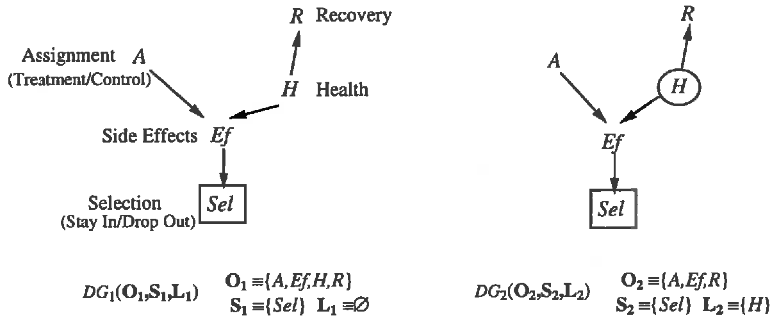

Figure 2. Randomized trial of an ineffective drug with unpleasant side effects leading to drop out. In $DG_1(\mathbf{O_1}, \mathbf{S_1}, \mathbf{L_1})$, $H \in \mathbf{O_1}$, and is observed, while in $DG_2(\mathbf{O_2}, \mathbf{S_2}, \mathbf{L_2})$ $H \in \mathbf{L_2}$ and is unobserved (variables in $\mathbf{L}$ are circled; variables in $\mathbf{S}$ are boxed; variables in $\mathbf{O}$ are not marked).

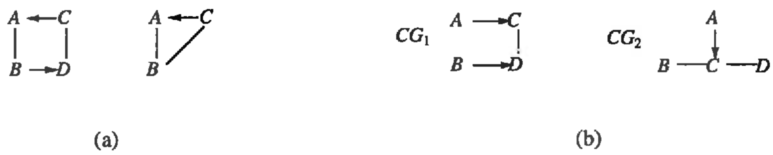

Figure 3. (a) mixed graphs containing partially directed cycles; (b) chain graphs.

health status (H) is observed so $H \in \mathbf{O_1}$, while in $DG_2(\mathbf{O_2}, \mathbf{S_2}, \mathbf{L_2})$ it is not observed so $H \in \mathbf{L_2}$.)

2.3. MIXED GRAPHS AND CHAIN GRAPHS

In a *mixed graph* a pair of vertices may be connected by a directed edge or an undirected edge (but not both). A *partially directed cycle* in a mixed graph G is a sequence of n distinct edges $\langle E_1, \ldots, E_n \rangle$, ($n \geq 3$), with endpoints X_i, X_{i+1} respectively, such that:

(a) $X_1 \equiv X_{n+1}$,
(b) $\forall i$ $(1 \leq i \leq n)$ either $X_i — X_{i+1}$ or $X_i \rightarrow X_{i+1}$, and
(c) $\exists j$ $(1 \leq j \leq n)$ such that $X_j \rightarrow X_{j+1}$.

A *chain graph* CG is a mixed graph in which there are no partially directed cycles (see Figure 3). Koster [16] considers classes of reciprocal graphs containing directed and undirected edges in which partially directed cycles are allowed. Such graphs are not considered separately here, though many of the comments which apply to LWF chain graphs also apply to reciprocal graphs since the former are a subclass of the latter.

To make clear which kind of graph is being referred to UG will denote undirected graphs, DG directed graphs, CG chain graphs, and G a graph which may be any one of these. A *path* between X and Y in a graph G (of whatever type) consists of a sequence of edges $\langle E_1, \ldots, E_n \rangle$ such that there exists a sequence of distinct vertices $\langle X \equiv X_1, \ldots, X_{n+1} \equiv Y \rangle$ where E_i has *endpoints*, X_i and X_{i+1} $(1 \leq i \leq n)$, i.e. E_i is $X_i — X_{i+1}, X_i \rightarrow X_{i+1}$, or $X_i \leftarrow X_{i+1}$ $(1 \leq i \leq n)$.[5] If no vertex occurs more than once on the path then the path is *acyclic*, otherwise it is *cyclic*. A directed path from X to Y is a path of the form $X \rightarrow \cdots \rightarrow Y$.

2.4. THE GLOBAL MARKOV PROPERTY ASSOCIATED WITH UNDIRECTED GRAPHS

A *global Markov property* associates a set of conditional independence relations with a graph G.[6] In an undirected graph UG, for disjoint sets of vertices $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, ($\mathbf{Z}$ may be empty), if there is no path from a variable $X \in \mathbf{X}$, to a variable $Y \in \mathbf{Y}$, that does not include some variable in $\mathbf{Z}$, then $\mathbf{X}$ and $\mathbf{Y}$ are said to be *separated* by $\mathbf{Z}$.

Undirected Global Markov Property; separation ($\models_U$)
$UG \models_U \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if $\mathbf{X}$ and $\mathbf{Y}$ are separated by $\mathbf{Z}$ in UG.[7]

Thus the undirected graphs in Figure 1(a) entail the following conditional independences via separation:

$$UG_1 \models_U \quad A \perp\!\!\!\perp D \mid C;\ A \perp\!\!\!\perp D \mid \{B,C\};\ B \perp\!\!\!\perp C \mid D;\ B \perp\!\!\!\perp C \mid \{A,D\};$$
$$A \perp\!\!\!\perp B \mid C\ A \perp\!\!\!\perp B \mid D;\ A \perp\!\!\!\perp B \mid \{C,D\}$$

$$UG_2 \models_U \quad A \perp\!\!\!\perp B \mid C;\ A \perp\!\!\!\perp B \mid \{C,D\};\ A \perp\!\!\!\perp D \mid C;\ A \perp\!\!\!\perp D \mid \{B,C\};$$
$$B \perp\!\!\!\perp D \mid C;\ B \perp\!\!\!\perp D \mid \{A,C\}.$$

Here, and throughout, all and only 'elementary' independence relations of the form $\{X\} \perp\!\!\!\perp \{Y\} \mid \mathbf{Z}$ ($\mathbf{Z}$ may be empty) are listed. For instance, note that UG_1 also entails $\{A\} \perp\!\!\!\perp \{B,D\} \mid \{C\}$.

[5] 'Path' is defined here as a sequence of edges, rather than vertices; in a directed cyclic graph a sequence of vertices does not in general define a unique path, since there may be more than one edge between a given pair of vertices. (Note that in a chain graph there is at most one edge between each pair of vertices.)

[6] Often global Markov conditions are introduced as a means for deriving the consequences of a set of local Markov conditions. Here the global property is defined directly in terms of the relevant graphical criterion.

[7] '$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$' means that '$\mathbf{X}$ is independent of $\mathbf{Y}$ given $\mathbf{Z}$'; if $\mathbf{Z} = \emptyset$, the abbreviation $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ is used. When convenient braces are omitted from singleton sets $\{V\}$, e.g. $V \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ instead of $\{V\} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$

2.5. THE GLOBAL MARKOV PROPERTY ASSOCIATED WITH DIRECTED GRAPHS

In a directed graph DG, X is a *parent* of Y, (and Y is a *child* of X) if there is a directed edge $X \rightarrow Y$ in G. X is an *ancestor* of Y (and Y is a *descendant* of X) if there is a directed path $X \rightarrow \cdots \rightarrow Y$ from X to Y, or $X \equiv Y$. Thus 'ancestor' ('descendant') is the transitive, reflexive closure of the 'parent' ('child') relation. A pair of consecutive edges on a path π in DG are said to *collide at vertex* A if both edges are into A, i.e. $\rightarrow A \leftarrow$, in this case A is called a *collider on* π, otherwise A is a *non-collider on* π. Thus every vertex on a path in a directed graph is either a collider, a non-collider, or an endpoint. For distinct vertices X and Y, and set $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$, a path π between X and Y is said to *d-connect* X and Y given $\mathbf{Z}$ if every collider on π is an ancestor of a vertex in $\mathbf{Z}$, and no non-collider on π is in $\mathbf{Z}$. For disjoint sets $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$, if there is an $X \in \mathbf{X}$, and $Y \in \mathbf{Y}$, such that there is a path which d-connects X and Y given $\mathbf{Z}$ then $\mathbf{X}$ and $\mathbf{Y}$ are said to be *d-connected* given $\mathbf{Z}$. If no such path exists then $\mathbf{X}$ and $\mathbf{Y}$ are said to be *d-separated* given $\mathbf{Z}$ (see Pearl [22]).

Directed Global Markov Property; d-separation ($\models_{DS}$)
$DG \models_{DS} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by $\mathbf{Z}$ in DG.

Thus the directed graphs in Figure 1(b,c) entail the following conditional independences via d-separation:

$$DG_1 \models_{DS} B \perp\!\!\!\perp C \mid \{A, D\};$$

$$DG_2 \models_{DS} B \perp\!\!\!\perp C \mid A;\ A \perp\!\!\!\perp D \mid \{B, C\};$$

$$DG_3 \models_{DS} A \perp\!\!\!\perp B \mid C;\ A \perp\!\!\!\perp B \mid \{C, D\};\ A \perp\!\!\!\perp D \mid C;\ A \perp\!\!\!\perp D \mid \{B, C\};\ B \perp\!\!\!\perp D \mid C;\ B \perp\!\!\!\perp D \mid \{A, C\}.$$

Note that the conditional independences entailed by DG_3 under d-separation are precisely those entailed by UG_2 under separation.

2.6. THE GLOBAL MARKOV PROPERTY ASSOCIATED WITH DIRECTED GRAPHS WITH LATENT AND SELECTION VARIABLES

The global Markov property for a directed graph with latent and/or selection variables is a natural extension of the global Markov property for directed graphs. For $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$, and $\mathbf{X} \dot{\cup} \mathbf{Y} \dot{\cup} \mathbf{Z} \subseteq \mathbf{O}$ define:

$$DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \models_{DS} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \text{ if and only if } DG \models_{DS} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z} \cup \mathbf{S}.$$

In other words, the set of conditional independence relations entailed by $DG(\mathbf{O},\mathbf{S},\mathbf{L})$ is exactly the subset of those independence relations entailed by the directed graph DG, in which no latent variables occur, and the conditioning set always includes (implicitly) all the selection variables in $\mathbf{S}$. Since, under the interpretation of $DG(\mathbf{O},\mathbf{S},\mathbf{L})$, the only observed variables are in $\mathbf{O}$, conditional independence relations involving variables in $\mathbf{L}$ are not observed. Similarly, samples are drawn from a subpopulation in which all variables in $\mathbf{S}$ are conditioned on, e.g. in the example in Section 2.2, the only patients observed were those for which $Sel = Stay\ In$. Thus the variables in $\mathbf{S}$ will be conditioned upon in every conditional independence relation observed to hold in the sample. Hence $DG(\mathbf{O},\mathbf{S},\mathbf{L})$ entails a set of conditional independences which hold in the observed distribution $\mathrm{P}(\mathbf{O} \mid \mathbf{S} = \mathbf{In})$. (See Spirtes and Richardson [33]; Spirtes, Meek and Richardson [32]; Cox and Wermuth [7].) Thus the graph $DG_1(\mathbf{O_1},\mathbf{S_1},\mathbf{L_1})$, shown in Figure 2, entails the following conditional independences:

$$DG_1(\mathbf{O_1},\mathbf{S_1},\mathbf{L_1}) \models_{DS} A \perp\!\!\!\perp R \mid H;\ A \perp\!\!\!\perp R \mid \{H, Ef\},$$

since $DG_1 \models_{DS} A \perp\!\!\!\perp R \mid \{H, Sel\};\ A \perp\!\!\!\perp R \mid \{H, Ef, Sel\}$.

However, the graph $DG_2(\mathbf{O_2},\mathbf{S_2},\mathbf{L_2})$ does not entail any independences, since health status is unobserved, $H \notin \mathbf{O_2}$, so neither of the above mentioned independences entailed by the graph $DG_1(\mathbf{O_1},\mathbf{S_1},\mathbf{L_1})$ is entailed by $DG_2(\mathbf{O_2},\mathbf{S_2},\mathbf{L_2})$.

2.7. GLOBAL MARKOV PROPERTIES ASSOCIATED WITH CHAIN GRAPHS

There are two different global Markov properties which have been proposed for chain graphs. In both definitions a conditional independence relation is entailed if sets $\mathbf{X}$ and $\mathbf{Y}$ are separated by $\mathbf{Z}$ in an undirected graph the vertices of which are a subset of those in the chain graph, while the edges are a superset of those occurring between these vertices in the original chain graph.[8]

2.7.1. *The Lauritzen-Wermuth-Frydenberg chain graph Markov property*

A vertex V in a chain graph is said to be *anterior* to a set $\mathbf{W}$ if there is a path π from V to some $W \in \mathbf{W}$ in which all directed edges $(X \to Y)$ on the path (if any) are such that Y is between X and W on π, $Ant(\mathbf{W}) = \{V \mid V \text{ is anterior to } W\}$. Let $CG(\mathbf{W})$ denote the *induced subgraph* of CG obtained by removing all vertices in $\mathbf{V}\backslash\mathbf{W}$ and all edges with an endpoint

[8]More recently, both of these Markov properties have been re-formulated in terms of a separation criteria that may be applied to the original chain graph, rather than an undirected graph derived from it (see Studený and Bouckaert [36], Andersson *et al.* [2]).

Figure 4. Contrast between the LWF and AMP Markov properties. Undirected graphs used to test $A \perp\!\!\!\perp D \mid \{B, C\}$ in CG_1 under (a) the LWF property, (b) the AMP property. Undirected graphs used to test $B \perp\!\!\!\perp D \mid C$ in CG_2 under (c) the LWF property, (d) the AMP property.

in $\mathbf{V} \setminus \mathbf{W}$. A *complex* in CG is an induced subgraph with the following form: $X \rightarrow V_1 - \cdots - V_n \leftarrow Y$ ($n \geq 1$). A complex is *moralized* by adding the undirected edge $X - Y$. $Moral(CG)$ is the undirected graph formed by moralizing all complexes in CG, and then replacing all directed edges with undirected edges.

LWF Global Markov Property for Chain Graphs ($\models_{LWF}$)
$CG \models_{LWF} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if $\mathbf{X}$ is separated from $\mathbf{Y}$ by $\mathbf{Z}$ in the undirected graph Moral(CG(Ant($\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$))).

Hence the chain graphs in Figure 3(b) entail the following conditional independences under the LWF Markov property:

$$CG_1 \models_{LWF} A \perp\!\!\!\perp B;\ A \perp\!\!\!\perp D \mid \{B, C\};\ B \perp\!\!\!\perp C \mid \{A, D\};$$

$$CG_2 \models_{LWF} A \perp\!\!\!\perp B \mid C;\ A \perp\!\!\!\perp B \mid \{C, D\};\ A \perp\!\!\!\perp D \mid C;\ A \perp\!\!\!\perp D \mid \{B, C\};\ B \perp\!\!\!\perp D \mid C;\ B \perp\!\!\!\perp D \mid \{A, C\}.$$

Notice that the conditional independences entailed by CG_2 under the LWF Markov property are the same as those entailed by DG_3 under d-separation, and UG_2 under separation (see Figure 1).

2.7.2. *The Andersson-Madigan-Perlman chain graph Markov property*
In a chain graph vertices V and W are said to be *connected* if there is a path containing only undirected edges between V and W, $Con(\mathbf{W}) = \{V \mid V$ is connected to some $W \in \mathbf{W}\}$. The extended subgraph, $Ext(CG, \mathbf{W})$, has vertex set Con($\mathbf{W}$) and contains all directed edges in $CG(\mathbf{W})$, and all undirected edges in CG(Con($\mathbf{W}$)). A vertex V in a chain graph is said to be an *ancestor* of a set $\mathbf{W}$ if there is a path π from V to some $W \in \mathbf{W}$ in which all edges on the path are directed ($X \rightarrow Y$) and are such that Y is between X and W on π.[9] (See Figure 5.) Now let

[9] Note that other authors, e.g. Lauritzen [17], have used 'ancestral' to refer to the set named 'anterior' in Section 3.

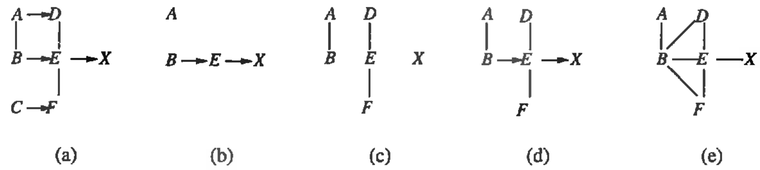

Figure 5. Constructing an augmented and extended chain graph: (a) a chain graph CG; (b) directed edges in $\text{Anc}(\{A, E, X\})$; (c) undirected edges in $\text{Con}(\text{Anc}(\{A, E, X\}))$ (d) $\text{Ext}(CG, \text{Anc}(\{A, E, X\}))$; (e) $\text{Aug}(\text{Ext}(CG, \text{Anc}(\{A, E, X\})))$.

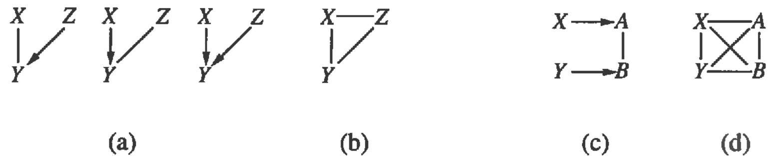

Figure 6. (a) Triplexes $\langle X, Y, Z\rangle$ and (b) the corresponding augmented triplex. (c) A chain graph with a bi-flag $\langle X, A, B, Y\rangle$ and two triplexes $\langle A, B, Y\rangle$, $\langle X, A, B\rangle$; (d) the corresponding augmented chain graph.

$$Anc(\mathbf{W}) = \{V \mid V \text{ is an ancestor of some } W \in \mathbf{W}\}.$$

A triple of vertices $\langle X, Y, Z\rangle$ is said to form a *triplex* in CG if the induced subgraph $CG(\{X, Y, Z\})$ is either $X \rightarrow Y - Z$, $X \rightarrow Y \leftarrow Z$, or $X - Y \leftarrow Z$. A triplex is *augmented* by adding the $X - Z$ edge. A set of four vertices $\langle X, A, B, Y\rangle$ is said to form a *bi-flag* if the edges $X \rightarrow A$, $Y \rightarrow B$, and $A - B$ are present in the induced subgraph over $\{X, A, B, Y\}$. A bi-flag is *augmented* by adding the edge $X - Y$. $Aug(CG)$ is the undirected graph formed by augmenting all triplexes and bi-flags in CG and replacing all directed edges with undirected edges (see Figure 6). Now let

$$\text{Aug}[CG; \mathbf{X}, \mathbf{Y}, \mathbf{Z}] = \text{Aug}(\text{Ext}(CG, \text{Anc}(\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}))).$$

AMP Global Markov Property ($\models_{AMP}$)
$CG \models_{AMP} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if $\mathbf{X}$ is separated from $\mathbf{Y}$ by $\mathbf{Z}$ in the undirected graph $\text{Aug}[CG; \mathbf{X}, \mathbf{Y}, \mathbf{Z}]$.

Hence the conditional independence relations associated with the chain graphs in Figure 3(b) under the AMP global Markov property are:

$CG_1 \models_{AMP}$ $A \perp\!\!\!\perp B$; $A \perp\!\!\!\perp B \mid C$; $A \perp\!\!\!\perp B \mid D$; $A \perp\!\!\!\perp D$; $A \perp\!\!\!\perp D \mid B$; $B \perp\!\!\!\perp C$; $B \perp\!\!\!\perp C \mid A$;

$CG_2 \models_{AMP}$ $A \perp\!\!\!\perp B$; $A \perp\!\!\!\perp B \mid D$; $A \perp\!\!\!\perp D$; $A \perp\!\!\!\perp D \mid B$; $B \perp\!\!\!\perp D \mid \{A, C\}$.

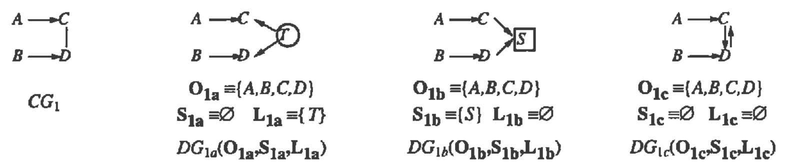

Figure 7. A chain graph and directed graphs in which C and D are symmetrically related.

Both LWF and AMP properties coincide with separation (d-separation) for the special case of a chain graph which is an undirected (acyclic, directed) graph. Thus chain graphs with either property are a generalization of both acyclic, directed graphs and undirected graphs. Cox and Wermuth [7] distinguish between the LWF and AMP Markov properties by using dashed lines, X ---- Y, in chain graphs under the AMP property.

2.8. MARKOV EQUIVALENCE AND COMPLETENESS

Two graphs G_1, G_2 under global Markov properties R_1, R_2 respectively are said to be *Markov equivalent* if $G_1 \models_{R_1} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if and only if $G_2 \models_{R_2} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$. Thus CG_2 under the LWF Markov property, DG_3 under d-separation, and UG_2 under separation are all Markov equivalent. For a given global Markov property R, and graph G with vertex set $\mathbf{V}$, a distribution P is said to be *G-Markovian$_R$* if for disjoint subsets $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, $G \models_R \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ implies $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ in P. A global Markov property is said to be *weakly complete* if for all disjoint sets $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, such that $G \not\models_R \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ there is a G-Markovian$_R$ distribution P in which $\mathbf{X} \not\perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$. The property R is said to be *strongly complete* if there is a *G-Markovian$_R$* distribution P in which $G \models_R \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if and only if $\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ in P. All of the global Markov properties here are known to be strongly (and hence weakly) complete (Geiger [12]; Frydenberg [11]; Spirtes [30]; Meek [21]; Spirtes *et al.* [31]; Studený and Bouckaert [36]; Andersson *et al.* [2]).

3. Directed Graphs with Symmetric Relations

In this section the Markov properties of simple directed graphs with symmetrically related variables are compared to those of the corresponding chain graphs. In particular, the following symmetric relations between variables X and Y are considered: (a) X and Y have a latent common cause; (b) X and Y are both causes of some selection variable; (c) X is a cause of Y, and Y is a cause of X, as occurs in a feedback system.

The conditional independences relations entailed by the directed graphs

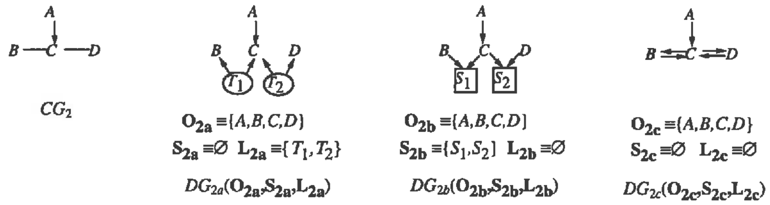

Figure 8. A chain graph and directed graphs in which the pairs of vertices B and C, and C and D, are symmetrically related.

in Figure 7 are:

$$DG_{1a}(\mathbf{O_{1a}},\mathbf{S_{1a}},\mathbf{L_{1a}}) \models_{DS} A \perp\!\!\!\perp B;\ A \perp\!\!\!\perp B \mid C;\ A \perp\!\!\!\perp B \mid D;\ A \perp\!\!\!\perp D;\ A \perp\!\!\!\perp D \mid B;\ B \perp\!\!\!\perp C;\ B \perp\!\!\!\perp C \mid A;$$

$$DG_{1b}(\mathbf{O_{1b}},\mathbf{S_{1b}},\mathbf{L_{1b}}) \models_{DS} A \perp\!\!\!\perp B \mid C;\ A \perp\!\!\!\perp B \mid D;\ B \perp\!\!\!\perp C \mid D;\ A \perp\!\!\!\perp D \mid C;\ A \perp\!\!\!\perp B \mid \{C,D\}; A \perp\!\!\!\perp D \mid \{B,C\}; B \perp\!\!\!\perp C \mid \{A,D\};$$

$$DG_{1c}(\mathbf{O_{1c}},\mathbf{S_{1c}},\mathbf{L_{1c}}) \models_{DS} A \perp\!\!\!\perp B;\ A \perp\!\!\!\perp B \mid \{C,D\}.$$

It follows that none of these directed graphs is Markov equivalent to CG_1 under the LWF Markov property. However, $DG_{1a}(\mathbf{O_{1a}},\mathbf{S_{1a}},\mathbf{L_{1a}})$ is Markov equivalent to CG_1 under the AMP Markov property. Turning now to the directed graphs shown in Figure 8, the following conditional independence relations are entailed:

$$DG_{2a}(\mathbf{O_{2a}},\mathbf{S_{2a}},\mathbf{L_{2a}}) \models_{DS} A \perp\!\!\!\perp B;\ A \perp\!\!\!\perp B \mid D;\ A \perp\!\!\!\perp D;\ A \perp\!\!\!\perp D \mid B;\ B \perp\!\!\!\perp D;\ B \perp\!\!\!\perp D \mid A;$$

$$DG_{2b}(\mathbf{O_{2b}},\mathbf{S_{2b}},\mathbf{L_{2b}}) \models_{DS} A \perp\!\!\!\perp B \mid C;\ A \perp\!\!\!\perp B \mid \{C,D\};\ A \perp\!\!\!\perp D \mid C;\ A \perp\!\!\!\perp D \mid \{B,C\};\ B \perp\!\!\!\perp D \mid C;\ B \perp\!\!\!\perp D \mid \{A,C\};$$

$DG_{2c}(\mathbf{O_{2c}},\mathbf{S_{2c}},\mathbf{L_{2c}})$ does not entail any conditional independences.

It follows that none of these directed graphs is Markov equivalent to CG_2 under the AMP Markov property. However, $DG_{2b}(\mathbf{O_{2b}},\mathbf{S_{2b}},\mathbf{L_{2b}})$ is Markov equivalent to CG_2 under the LWF Markov property. Further, note that $DG_{2b}(\mathbf{O_{2b}},\mathbf{S_{2b}},\mathbf{L_{2b}})$ is also Markov equivalent to UG_2 (under separation) and DG_3 (under d-separation) in Figure 1(a).

There are two other simple symmetric relations that might be considered: (d) X and Y have a common child that is a latent variable; (e) X and Y have a common parent that is a selection variable. However, without additional edges X and Y are entailed to be independent (given $\mathbf{S}$)

in these configurations, whereas this is clearly not the case if there is an edge between X and Y in a chain graph.

Hence none of the simple directed graphs with symmetric relations corresponding to CG_1 are Markov equivalent to CG_1 under the LWF Markov property, and likewise none of those corresponding to CG_2 are Markov equivalent to CG_2 under the AMP Markov property. In the next section a stronger result is proved: in fact there are no directed graphs, however complicated, with or without latent and selection variables, that are Markov equivalent to CG_1 and CG_2 under the LWF and AMP Markov properties, respectively.

4. Inseparability and Related Properties

In this section two Markov properties, motivated by spatial and causal intuitions, are introduced. It is then shown that these Markov properties hold for all undirected graphs, and all directed graphs (under d-separation) possibly with latent and selection variables. Distinct vertices X and Y are *inseparable*$_R$ in G under Markov Property R if there is no set $\mathbf{W}$ such that $G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W}$. If X and Y are not inseparable$_R$, they are *separable*$_R$. Let $[G]^{\mathrm{Ins}}_R$ be the undirected graph in which there is an edge $X - Y$ if and only if X and Y are inseparable$_R$ in G under R. Note that in accord with the definition of $\models_{DS}$ for $DG(\mathbf{O},\mathbf{S},\mathbf{L})$, only vertices $X, Y \in \mathbf{O}$ are separable$_{DS}$ or inseparable$_{DS}$, thus $[DG(\mathbf{O},\mathbf{S},\mathbf{L})]^{\mathrm{Ins}}_{DS}$ is defined to have vertex set $\mathbf{O}$.

For an undirected graph model $[UG]^{\mathrm{Ins}}_S$ is just the undirected graph UG. For an acyclic, directed graph (without latent or selection variables) under d-separation, or a chain graph under either Markov property $[G]^{\mathrm{Ins}}_R$ is simply the undirected graph formed by replacing all directed edges with undirected edges, hence for any chain graph CG, $[CG]^{\mathrm{Ins}}_{AMP} = [CG]^{\mathrm{Ins}}_{LWF}$.

In any graphical model, if there is an edge (directed or undirected) between a pair of variables then those variables are inseparable$_R$. For undirected graphs, acyclic directed graphs, and chain graphs (under either Markov property), inseparability$_R$ is both a necessary and a sufficient condition for the existence of an edge between a pair of variables. However, in a directed graph with cycles, or in a (cyclic or acyclic) directed graph with latent and/or selection variables, inseparability$_{DS}$ is not a sufficient condition for there to be an edge between a pair of variables (recall that in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$, the entailed conditional independences are restricted to those that are observable).

An *inducing path* between X and Y in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$ is a path π between X and Y on which (i) every vertex in $\mathbf{O} \cup \mathbf{S}$ is a collider on π, and (ii) every collider is an ancestor of X, Y or $\mathbf{S}$.[10] In a directed graph, $DG(\mathbf{O},\mathbf{S},\mathbf{L})$,

[10]The notion of an inducing path was first introduced for acyclic directed graphs with

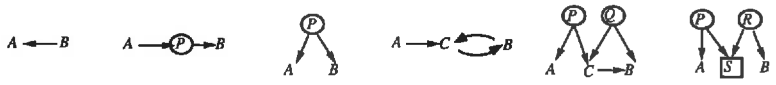

Figure 9. Examples of directed graphs $DG(\mathbf{O},\mathbf{S},\mathbf{L})$ in which A and B are inseparable$_{DS}$.

variables $X, Y \in \mathbf{O}$, are inseparable$_{DS}$ if and only if there is an inducing path between X and Y in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$.[11] For example, C and D were inseparable$_{DS}$ in $DG_{1a}(\mathbf{O_{1a}},\mathbf{S_{1a}},\mathbf{L_{1a}})$, and $DG_{1b}(\mathbf{O_{1b}},\mathbf{S_{1b}},\mathbf{L_{1b}})$, while in $DG_{1c}(\mathbf{O_{1c}},\mathbf{S_{1c}},\mathbf{L_{1c}})$ A and B were the only separable$_{DS}$ variables. Figure 9 contains further examples of graphs in which vertices are inseparable$_{DS}$.

4.1. 'BETWEEN SEPARATED' MODELS

A vertex B will be said to be *between*$_R$ X and Y in G under Markov property R, if and only if there exists a sequence of *distinct* vertices $\langle X \equiv X_0, X_1, \ldots, X_n \equiv B, X_{n+1}, \ldots, X_{n+m} \equiv Y\rangle$ in $[G]_R^{\text{Ins}}$ such that each consecutive pair of vertices X_i, X_{i+1} in the sequence are inseparable$_R$ in G under R. Clearly B will be between$_R$ X and Y in G if and only if B lies on a path between X and Y in $[G]_R^{\text{Ins}}$. The set of vertices between X and Y under property R in graph G is denoted $Between_R(G; X, Y)$, abbreviated to $Between_R(X, Y)$, when G is clear from context. Note that for any chain graph CG, $\text{Between}_{LWF}(CG; X, Y) = \text{Between}_{AMP}(CG; X, Y)$, for all vertices X and Y.

Between$_R$ Separated Models

A model G is *between*$_R$ *separated*, if for all pairs of vertices X, Y and sets $\mathbf{W}$ $(X, Y \notin \mathbf{W})$:
$G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \Longrightarrow G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{Between}_R(G; X, Y)$
(where $\{X, Y\} \cup \mathbf{W}$ is a subset of the vertices in $[G]_R^{\text{Ins}}$).

It follows that if G is between$_R$ separated, then in order to make some (separable) pair of vertices X and Y conditionally independent, it is always sufficient to condition on a subset (possibly empty) of the vertices that lie on paths between X and Y.

The intuition that only vertices on paths between X and Y are relevant to making X and Y independent is related to the idea, fundamental to much

latent variables in Verma and Pearl [37]; it was subsequently extended to include selection variables in Spirtes, Meek and Richardson [32].

[11] Inseparability$_{DS}$ is a necessary and sufficient condition for there to be an edge between a pair of variables in a Partial Ancestral Graph (PAG), (Richardson [26, 27]; Spirtes *et al.* [32, 31]), which represents structural features common to a Markov equivalence class of directed graphs.

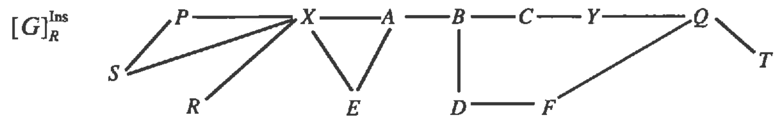

Figure 10. $\text{Between}_R(G; X, Y) = \{A, B, C, D, E, F, Q\}$, $\text{CoCon}_R(G; X, Y) = \{A, B, C, D, E, F, Q, T\}$, while P, R and S, are vertices not in $\text{CoCon}_R(G; X, Y)$

of graphical modelling, that if vertices are *dependent* then they should be *connected* in some way graphically. This is a natural correspondence, present in the spatial intuition that only contiguous regions interact directly, and also in causal principles which state that if two quantities are dependent then they are causally connected (Hausman [14]).[12]

Theorem 1 **(i)** *All undirected graphs H are between$_S$ separated.*
(ii) *All directed graphs $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$ are between$_{DS}$ separated.*

Proof: The proof for undirected graph models is given here. It is easy to see that the proof carries over directly to directed graphs without selection or latent variables (i.e. $\mathbf{V} = \mathbf{O}$, $\mathbf{S} = \mathbf{L} = \emptyset$) replacing 'separated' by 'd-separated', and 'connected' by 'd-connected'. The proof for directed graphs with latent and/or selection variables is in the appendix.

Suppose, for a contradiction, $UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W}$, but $UG \not\models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{Between}_S(X, Y)$. Then there is a path π in UG connecting X and Y given $\mathbf{W} \cap \text{Between}_S(X, Y)$. Since this path does not connect given $\mathbf{W}$, it follows that there is some vertex V on π, and $V \in \mathbf{W} \setminus \text{Between}_S(X, Y)$. But if V is on π, then π constitutes a sequence of vertices $\langle X \equiv X_0, X_1, \ldots, X_n \equiv V, X_{n+1}, \ldots, X_{n+m} \equiv Y \rangle$ such that consecutive pairs of vertices are inseparable$_S$ (because there is an edge between each pair of variables). Hence $V \in \text{Between}_S(X, Y)$, which is a contradiction. □

In general, chain graphs are not between$_{LWF}$ separated or between$_{AMP}$ separated. This is shown by CG_1 and CG_2 in Figure 3:

$$CG_1 \models_{LWF} A \perp\!\!\!\perp D \mid \{B, C\},$$

so A and D are separable$_{LWF}$, but $\text{Between}_{LWF}(CG_1; A, D) = \{C\}$ and

$$CG_1 \not\models_{LWF} A \perp\!\!\!\perp D \mid \{C\}.$$

For the AMP property note that

$$CG_2 \models_{AMP} B \perp\!\!\!\perp D \mid \{A, C\},$$

[12] Where 'A and B are causally connected' means that either A is a cause of B, B is a cause of A, or they share some common cause (or some combination of these).

but $\text{Between}_{AMP}(CG_2; B, D) = \{C\}$, and yet

$$CG_2 \not\models_{AMP} B \perp\!\!\!\perp D \mid \{C\}.$$

4.2. 'CO-CONNECTION DETERMINED' MODELS

A vertex W will be said to be *co-connected*$_R$ to X and Y in G if X, Y and W are vertices in $[G]_R^{\text{Ins}}$ satisfying:

(i) There is a sequence of vertices $\langle X, A_1, A_2, \ldots, A_n, W\rangle$ in $[G]_R^{\text{Ins}}$ which does not contain Y and consecutive pairs of variables in the sequence are inseparable$_R$ in G under R.

(ii) There is a sequence of vertices $\langle W, B_1, B_2, \ldots, B_m, Y\rangle$ in $[G]_R^{\text{Ins}}$ which does not contain X and consecutive pairs of variables in the sequence are inseparable$_R$ in G under R.

Let $CoCon_R(G; X, Y) = \{V \mid V \text{ is co-connected}_R \text{ to } X \text{ and } Y \text{ in } [G]_R^{\text{Ins}}\}$.

It is easy to see that B will be co-connected$_R$ to X and Y in G if and only if (a) B is not separated from Y by X in $[G]_R^{\text{Ins}}$, and (b) B is not separated from X by Y in $[G]_R^{\text{Ins}}$. Note that for any chain graph CG, and vertices X, Y, $\text{CoCon}_{LWF}(CG; X, Y) = \text{CoCon}_{AMP}(CG; X, Y)$.

Clearly $\text{Between}_R(G; X, Y) \subseteq \text{CoCon}_R(G; X, Y)$, so being co-connected$_R$ to X and Y is a weaker requirement than being between$_R$ X and Y. Both $\text{Between}_R(G; X, Y)$ and $\text{CoCon}_R(G; X, Y)$ are sets of vertices which are topologically 'in between' X and Y in $[G]_R^{\text{Ins}}$.

Co-Connection$_R$ Determined Models

A model G will be said to be *co-connection*$_R$ *determined*, if for all pairs of vertices X, Y and sets $\mathbf{W}$ $(X, Y \notin \mathbf{W})$:
$G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \iff G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \cap CoCon_R(G; X, Y)$
(where $\{X, Y\} \cup \mathbf{W}$ is a subset of the vertices in $[G]_R^{\text{Ins}}$).

This principle states that the inclusion or exclusion of vertices that are not in $\text{CoCon}_R(X, Y)$ from some set $\mathbf{W}$ is irrelevant to whether X and Y are entailed to be independent given $\mathbf{W}$.

Theorem 2 (i) *Undirected graph models are co-connection*$_S$ *determined.*

(ii) *Directed graphs, possibly with latent and/or selection variables, are co-connection*$_{DS}$ *determined.*

(iii) *Chain graphs are co-connection*$_{AMP}$ *determined.*

Proof: Again the proof for undirected graphs is given here. The proofs for directed graphs and AMP chain graphs are given in the appendix.

Since $\text{Between}_S(X, Y) \subseteq \text{CoCon}_S(X, Y)$, an argument similar to that used in the proof of Theorem 1 (replacing 'Between$_S$' with 'CoCon$_S$') shows that if $UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W}$ then $UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_S(X, Y)$. Conversely, if $UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_S(X, Y)$ then X and Y are separated by $\mathbf{W} \cap \text{CoCon}_S(X, Y)$ in UG. Since $\mathbf{W} \cap \text{CoCon}_S(X, Y) \subseteq \mathbf{W}$, it follows that X and Y are separated by $\mathbf{W}$ in UG.[13] □

For undirected graphs $UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \Longrightarrow UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{Between}_S(X, Y)$, i.e. undirected graphs could be said to be between$_S$ *determined*. Chain graphs are not co-connection$_{LWF}$ determined. In CG_1 B and C are separable$_{LWF}$, since $CG_1 \models_{LWF} B \perp\!\!\!\perp C \mid \{A, D\}$, but $CG_1 \not\models_{LWF} B \perp\!\!\!\perp C \mid \{D\}$ and $\text{CoCon}_{LWF}(CG_1; B, C) = \{D\}$. In contrast, chain graphs *are* co-connection$_{AMP}$ determined.

5. Discussion

The two Markov properties presented in the previous section are based on the intuition that only vertices which, in some sense, come 'between' X and Y should be relevant as to whether or not X and Y are entailed to be independent. Both of these properties are satisfied by undirected graphs and by all forms of directed graph model. Since chain graphs are not between$_R$ separated under either Markov property, this captures a qualitative difference between undirected and directed graphs, and chain graphs. On the other hand since chain graphs are co-connection$_{AMP}$ determined, in this respect, at least, AMP chain graphs are more similar to directed and undirected graphs.

5.1. DATA GENERATING PROCESSES

Since the pioneering work of Sewall Wright [38] in genetics, statistical models based on directed graphs have been used to model causal relations, and data generating processes. Models allowing directed graphs with cycles have been used for over 50 years in econometrics, and allow the possibility of representing linear feedback systems which reach a deterministic equilibrium subject to stochastic boundary conditions (Fisher [10]; Richardson [27]). Besag [3] gives several spatial-temporal data generating processes whose limiting spatial distributions satisfy the Markov property with respect to a naturally associated undirected graph. These data generating processes are time-reversible and temporally stationary. Thus there are data generating mechanisms known to give rise to the distributions described by undirected and directed graphs.

[13] This is the 'Strong Union Property' of separation in undirected graphs (Pearl [22]).

Cox [6] states that chain graphs under the LWF Markov property "do not satisfy the requirement of specifying a direct mode of data generation." However, Lauritzen[14] has recently sketched out, via an example, a dynamic data generating process for LWF chain graphs in which a pair of vertices joined by an undirected edge, $X - Y$, arrive at a stochastic equilibrium, as $t \to \infty$; the equilibrium distribution being determined by the parents of X and Y in the chain graph.

A data generation process corresponding to a Gaussian AMP chain graph may be constructed via a set of linear equations with correlated errors (Andersson *et al.* [1]). Each variable is given as a linear function of its parents in the chain graph, together with an error term. The distribution over the error terms is given by the undirected edges in the graph, as in a Gaussian undirected graphical model or 'covariance selection model' (Dempster [9]), for which Besag [3] specifies a data generating process. 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.

The existence of a data generating process for a particular chain graph (under either Markov property) is important since it provides a full justification for using this structure. As has been shown in this paper, the mere fact that two variables are 'symmetrically related' does not, on its own, justify the use of a chain graph model.

5.2. A THEORY OF INTERVENTION IN DIRECTED GRAPHS

Strotz and Wold [35], Spirtes *et al.* [31] and Pearl [23] develop a theory of causal intervention for directed graph models which makes it sometimes possible to calculate the effect of an ideal intervention in a causal system. Space does not permit a detailed account of the theory here, however, the central idea is very simple: manipulating a variable, say X, modifies the structure of the graph, removing the edges between X and its parents, and instead making a 'policy' variable the sole parent of X. The relationships between all other variables and their parents are not affected; it is in this sense that the intervention is 'ideal', only one variable is directly affected.[15]

Example: Returning to the example considered in section 2.2, hypothetically a researcher could intervene to directly determine whether or not the patient suffers the side-effects, e.g. by giving all of the patients (in both the

[14] Personal communication.

[15] It should also be noted that for obvious physical reasons it may not make sense to speak of manipulating certain variables, e.g. the age or sex of an individual.

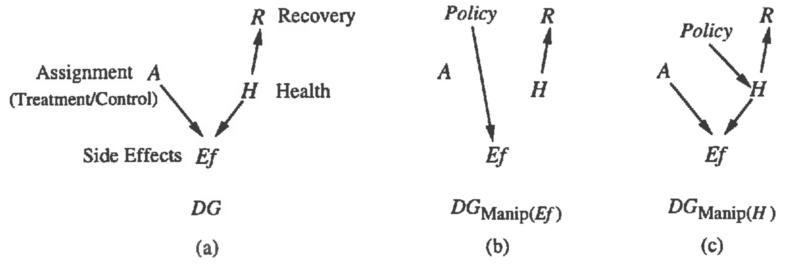

Figure 11. Intervening in a causal system: (a) before intervention; (b) intervening to directly control side-effects; (c) intervening to directly control health status.

treatment and control groups) something which either prevents or precipitates the side-effects. The graph in Figure 11(b) shows the result of such an intervention. After the intervention, which group the patient was assigned to initially (treatment/control) becomes independent of whether the patient suffers side-effects, as would be expected. The graph in Figure 11(c) shows the result of intervening to control the patient's health status directly.

One common objection to this intervention theory is that it makes a distinction between models that are statistically equivalent,[16] and hence which could not be differentiated purely on the basis of observational data. For example the graphs $A \to B$ and $A \leftarrow B$, are statistically equivalent,[17] and yet the effect on B of intervening to manipulate A will clearly be different. This is true, but misses the point: scientists are often able to control certain variables directly, and to perform controlled experiments, thus certain models can often be ruled out on the basis of background knowledge. This objection is also over-simplistic: when there are more than two variables it is often the case that all Markov equivalent models share certain structural features in common, even when latent and/or selection variables may be present (Verma and Pearl [37]; Frydenberg [11]; Spirtes and Verma [34]; Richardson [28]; Spirtes and Richardson [33]). Thus knowing that the data was generated by a model in a particular Markov equivalence class, even if which particular model is unknown, may be enough to predict the results of certain interventions (see Spirtes *et al.* [31]; Pearl [23]). A theory of intervention constitutes an important part of a causal data generating mechansim; specification of the dynamic behaviour of the system is another element, which may be of great importance in settings where feedback is present (Richardson [27, Ch. 2];).

In the absence of a theory of intervention for chain graphs, a resear-

[16]In the sense that they represent the same set of distributions.
[17]This observation is the basis for the slogan "Correlation is not Causation."

cher would be unable to answer questions concerning the consequences of intervening in a system with the structure of a chain graph. However, Lauritzen[18] has recently given, in outline, a theory of intervention for LWF chain graphs, which is compatible with the data generating process he has proposed. Such an intervention theory would appear to be of considerable use in applied settings when the substantive research hypotheses are causal in nature.

6. Conclusion

The examples given in this paper make clear that there are many ways in which a pair of variables may be symmetrically related. Further, different symmetric relationships, in general, will lead to quite different Markov properties. In particular, as has been shown, there are qualitative differences between the Markov properties associated with undirected and directed graphs (possibly with latent and/or selection variables), and either of those associated with chain graphs. Consequently, the Markov structure of a chain graph does not, in general, correspond to any symmetric relationship that can be described by a directed graph model via marginalizing or conditioning. For this reason, the inclusion of an undirected edge, rather than a directed edge, in a hypothesized chain graph model, should not be regarded as being 'weaker' or 'safer', substantively, than the inclusion of a directed edge.

This paper has shown that there are many symmetric relations which do *not* correspond to chain graphs. However, this leaves open the interesting question of which symmetric relations chain graphs *do* correspond to. A full answer to this question would involve the specification, in general, of a data generating process for chain graphs (under a given Markov property), together with an associated theory of intervention.

Acknowledgements

I would like to thank Julian Besag, David Cox, Clark Glymour, Steffen Lauritzen, David Madigan, Chris Meek, Michael Perlman, Richard Scheines, Peter Spirtes, Milan Studený, and Nanny Wermuth for helpful conversations on this topic. I am also grateful to three anonymous reviewers for useful comments and suggestions. Finally, I would like to gratefully acknowledge the Isaac Newton Institute for Mathematical Sciences, Cambridge, England, UK, where the revised version of this paper was prepared.

[18] Personal communication.

7. Proofs

In $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$ suppose that μ is a path that d-connects X and Y given $\mathbf{Z} \cup \mathbf{S}$, C is a collider on μ, and C is not an ancestor of $\mathbf{S}$. Let $length(C, \mathbf{Z})$ be 0 if C is a member of $\mathbf{Z}$; otherwise it is the length of a shortest directed path δ from C to a member of $\mathbf{Z}$. Let

$$Coll(\mu) = \{C \mid C \text{ is a collider on } \mu, \text{ and } C \text{ is not an ancestor of } \mathbf{S}\}.$$

Then let

$$size(\mu, \mathbf{Z}) = |\text{Coll}(\mu)| + \sum_{C \in \text{Coll}(\mu)} \text{length}(C, \mathbf{Z})$$

where $|\text{Coll}(\mu)|$ is the cardinality of $\text{Coll}(\mu)$. A path μ is a *minimal* acyclic d-connecting path between X and Y given $\mathbf{Z} \cup \mathbf{S}$, if μ is acyclic, d-connects X and Y given $\mathbf{Z} \cup \mathbf{S}$, and there is no other acyclic path μ' that d-connects X and Y given $\mathbf{Z} \cup \mathbf{S}$ such that $\text{size}(\mu', \mathbf{Z}) < \text{size}(\mu, \mathbf{Z})$. If there is a path that d-connects X and Y given $\mathbf{Z}$ then there is at least one minimal acyclic d-connecting path between X and Y given $\mathbf{Z}$.[19]

In the following proofs $\mu(A, B)$ denotes the subpath of μ between vertices A and B.

Lemma 1 *If μ is a minimal acyclic d-connecting path between X and Y given $\mathbf{Z} \cup \mathbf{S}$ in $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$, $\mathbf{Z} \cup \{X, Y\} \subseteq \mathbf{O}$, then for each collider C_i on μ that is not an ancestor of $\mathbf{S}$, there is a directed path δ_i from C_i to some vertex in $\mathbf{Z}$, such that δ_i intersects μ only at C_i, δ_i and δ_j do not intersect $(i \neq j)$ and no vertex on any path δ_i is in $\mathbf{S}$.*

Proof: Let δ_i be a shortest acyclic directed path from a collider C_i on μ to a member of $\mathbf{Z}$, where C_i is not an ancestor of $\mathbf{S}$, and hence no vertex on δ_i is in $\mathbf{S}$. We will now be prove that δ_i does not intersect μ except at C_i by showing that if such a point of intersection exists, then μ is not minimal, contrary to the assumption.

Form a path μ' in the following way: if δ_i intersects μ at a vertex other than C_i then let W_X be the vertex closest to X on μ that is on both δ_i and μ, and let W_Y be the vertex closest to Y on μ that is on both δ_i and μ. Suppose without loss of generality that W_X is after W_Y on δ_i. Let μ' be the concatenation of $\mu(X, W_X)$, $\delta_i(W_X, W_Y)$, and $\mu(W_Y, Y)$. It is now easy to show that μ' d-connects X and Y given $\mathbf{Z} \cup \mathbf{S}$. (See Figure 12.) Moreover $\text{size}(\mu', \mathbf{Z}) < \text{size}(\mu, \mathbf{Z})$ because μ' contains no more colliders than μ and a

[19]It is not hard to prove that in a DG, if there is a path (cyclic or acyclic) d-connecting X and Y given $\mathbf{Z}$, then there is an *acyclic* path d-connecting X and Y given $\mathbf{Z}$.

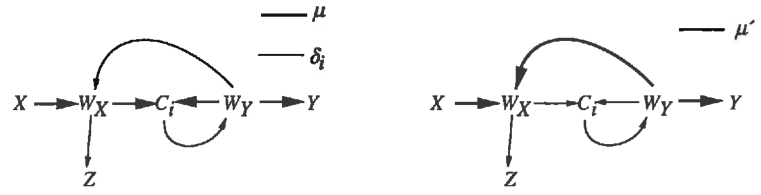

Figure 12. Finding a d-connecting path μ' of smaller size than μ, in the case where μ intersects with a directed path δ_i from a collider C_i to a vertex in $\mathbf{Z}$.

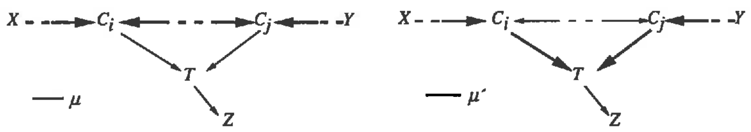

Figure 13. Finding a d-connecting path μ' of smaller size than μ, in the case where two directed paths, δ_i and δ_j, intersect.

shortest directed path from W_X to a member of $\mathbf{Z}$ is shorter than δ_i. Hence μ is not minimal, contrary to the assumption.

It remains to be shown that if μ is minimal, then δ_i and δ_j $(i \neq j)$ do not intersect. Suppose this is false. (See Figure 13.)

Let the vertex on δ_i closest to C_i that is also on δ_j be T. Let μ' be the concatenation of $\mu(X, C_i)$, $\delta_i(C_i, T)$, $\delta_j(T, C_j)$, and $\mu(C_j, Y)$. It is now easy to show that μ' d-connects X and Y given $\mathbf{Z} \cup \mathbf{S}$ and $\text{size}(\mu', \mathbf{Z}) < \text{size}(\mu, \mathbf{Z})$ because C_i and C_j are not colliders on μ', the only collider on μ' that may not be on μ is T, and the length of a shortest path from T to a member of $\mathbf{Z}$ is less than the length of a shortest path from C_i to a member of $\mathbf{Z}$. Hence μ is not minimal, contrary to the assumption. □

Lemma 2 *If μ is a minimal d-connecting path between X and Y, given $\mathbf{Z} \cup \mathbf{S}$ in $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$, B is a vertex on μ, and $\mathbf{Z} \cup \{X, Y, B\} \subseteq \mathbf{O}$, then there is a sequence of vertices $\langle X \equiv X_0, X_1, \ldots, X_n \equiv B, X_{n+1}, \ldots, X_{n+m} \equiv Y \rangle$ in $\mathbf{O}$, such that X_i and X_{i+1} $(0 \leq i < n+m)$ are inseparable$_{DS}$ in $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$.*

Proof: Since μ is a d-connecting path given $\mathbf{S} \cup \mathbf{Z}$ every collider on μ that is not an ancestor of $\mathbf{S}$ is an ancestor of a vertex in $\mathbf{Z}$. Denote the colliders on μ that are not ancestors of $\mathbf{S}$ as $C_1, \ldots, C_k$. Let δ_j be a shortest directed path from C_j to some vertex $Z_j \in \mathbf{Z}$. It follows by Lemma 1 that δ_j and μ intersect only at C_j, and that δ_j and $\delta_{j'}$ $(j \neq j')$ do not intersect, and no vertex on some path δ_j is in $\mathbf{S}$. A sequence of vertices X_i in $\mathbf{O}$, such that each X_i is either on μ or is on a directed path δ_j from C_j to Z_j can now be constructed:

Base Step: Let $X_0 \equiv X$.

Inductive Step: If X_i is on some path δ_j then define W_{i+1} to be C_j; otherwise, if X_i is on μ, then let W_{i+1} be X_i. Let V_{i+1} be the next vertex on μ, after W_{i+1}, such that $V_{i+1} \in \mathbf{O}$. If there is no vertex $C_{j'}$ between W_{i+1} and V_{i+1} on μ, then let $X_{i+1} \equiv V_{i+1}$. Otherwise let C_{j^*} be the first collider on μ, after W_{i+1}, that is not an ancestor of $\mathbf{S}$, and let X_{i+1} be the first vertex in $\mathbf{O}$ on the directed path δ_{j^*} (such a vertex is guaranteed to exist since Z_{j^*}, the endpoint of δ_{j^*}, is in $\mathbf{O}$). It follows from the construction that if B is on μ, and $B \in \mathbf{O}$, then for some i, $X_i \equiv B$.

Claim: X_i and X_{i+1} are inseparable$_{DS}$ in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$ under d-separation. If X_i and X_{i+1} are both on μ, then $\mu(X_i, X_{i+1})$ is a path on which every non-collider is in $\mathbf{L}$, and every collider is an ancestor of $\mathbf{S}$. Thus $\mu(X_i, X_{i+1})$ d-connects X_i and X_{i+1} given $\mathbf{Z} \cup \mathbf{S}$ for any $\mathbf{Z} \subseteq \mathbf{O} \backslash \{X_i, X_{i+1}\}$. So X_i and X_{i+1} are inseparable$_{DS}$. If X_i lies on some path δ_j, but X_{i+1} is on μ, then the path π formed by concatenating the directed path $X_i \leftarrow \cdots \leftarrow C_j$ and $\mu(C_j, X_{i+1})$ again is such that every non-collider on π is in $\mathbf{L}$, and every collider is an ancestor of $\mathbf{S}$, hence again X_i and X_{i+1} are inseparable$_{DS}$. The cases in which either X_{i+1} alone, or both X_i and X_{i+1} are not on μ can be handled similarly. □

Corollary 1 *If B is a vertex on a minimal d-connecting path π between X and Y given $\mathbf{Z} \cup \mathbf{S}$ in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$, $\mathbf{Z} \cup \{X, Y, B\} \subseteq \mathbf{O}$, then $B \in$ Between$_{DS}(X, Y)$.*

Proof: This follows directly from Lemma 2 □

Corollary 2 *If μ is a minimal d-connecting path between X and Y given $\mathbf{Z} \cup \mathbf{S}$ in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$, C is a collider on μ that is an ancestor of $\mathbf{Z}$ but not $\mathbf{S}$, δ is a shortest directed path from C to some $Z \in \mathbf{Z}$, and $\mathbf{Z} \cup \{X, Y, C\} \subseteq \mathbf{O}$, then $Z \in$ CoCon$_{DS}(X, Y)$.*

Proof: By Lemma 1, δ does not intersect μ except at C. Let the sequence of vertices on δ that are in $\mathbf{O}$ be $\langle V_1, \ldots, V_r \equiv Z\rangle$. It follows from the construction in Lemma 1 that there is a sequence of vertices $\langle X \equiv X_0, X_1, \ldots, X_n \equiv V_1, X_{n+1}, \ldots, X_{n+m} \equiv Y\rangle$ in $\mathbf{O}$ such that consecutive pairs of vertices are inseparable$_{DS}$. Since, by hypothesis, C is not an ancestor of $\mathbf{S}$, it follows that no vertex on δ is in $\mathbf{S}$. Hence $\delta(V_i, V_{i+1})$ is a directed path from V_i to V_{i+1} on which, with the exception of the endpoints, every vertex is in $\mathbf{L}$ and is a non-collider on δ, it follows that V_i and V_{i+1} are inseparable$_{DS}$ in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$. Thus the sequences $\langle X \equiv X_0, X_1, \ldots, X_n \equiv V_1, \ldots, V_r \equiv Z\rangle$ and $\langle Y \equiv$

$X_{n+m}, \ldots, X_n \equiv V_1, \ldots, V_r \equiv Z\rangle$ establish that $Z \in \text{CoCon}_{DS}(X, Y)$ in $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$. □

Theorem 1 (i) *A directed graph $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$ is between$_{DS}$ separated under d-separation.*

Proof: Suppose, for a contradiction, that $DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W}$, $(\mathbf{W} \cup \{X, Y\} \subseteq \mathbf{O})$, but $DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \not\models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{Between}_{DS}(X, Y)$. In this case there is some minimal path π d-connecting X and Y given $\mathbf{S} \cup (\mathbf{W} \cap \text{Between}_{DS}(X, Y))$ in $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$, but this path is not d-connecting given $\mathbf{S} \cup \mathbf{W}$. It is not possible for a collider on π to have a descendant in $\mathbf{S} \cup (\mathbf{W} \cap \text{Between}_{DS}(X, Y))$, but not in $\mathbf{S} \cup \mathbf{W}$. Hence there is some non-collider B on π, s.t. $B \in \mathbf{S} \cup \mathbf{W}$, but $B \notin \mathbf{S} \cup (\mathbf{W} \cap \text{Between}_{DS}(X, Y))$. This implies $B \in \mathbf{W} \backslash \text{Between}_{DS}(X, Y)$, and since $\mathbf{W} \subseteq \mathbf{O}$, it follows that $B \in \mathbf{O}$. But in this case by Corollary 1, $B \in \text{Between}_{DS}(X, Y)$, which is a contradiction. □

Theorem 2 (ii) *A directed graph $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$ is co-connection$_{DS}$ determined.*

Proof: Since $\text{Between}_{DS}(X, Y) \subseteq \text{CoCon}_{DS}(X, Y)$, the proof of Theorem 1 above, replacing 'between$_{DS}$' with 'co-connected$_{DS}$', suffices to show that if $DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W}$ then $DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_{DS}(X, Y)$.

To prove the converse, suppose, for a contradiction, $DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_{DS}(X, Y)$, but $DG(\mathbf{O}, \mathbf{S}, \mathbf{L}) \not\models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W}$, where $\mathbf{W} \cup \{X, Y\} \subseteq \mathbf{O}$. It then follows that there is some minimal d-connecting path π between X and Y in $DG(\mathbf{O}, \mathbf{S}, \mathbf{L})$ given $\mathbf{W} \cup \mathbf{S}$. Clearly it is not possible for there to be a non-collider on π which is in $\mathbf{S} \cup (\mathbf{W} \cap \text{CoCon}_{DS}(X, Y))$, but not in $\mathbf{S} \cup \mathbf{W}$. Hence it follows that there is some collider C on π which has a descendant in $\mathbf{S} \cup \mathbf{W}$, but not in $\mathbf{S} \cup (\mathbf{W} \cap \text{CoCon}_{DS}(X, Y))$. Hence C is an ancestor of $\mathbf{W} \backslash \text{CoCon}_{DS}(X, Y)$, but not $\mathbf{S}$. Consider a shortest directed path δ from C to some vertex W in $\mathbf{W}$. It follows from Lemma 1, and the minimality of π that δ does not intersect π except at C. It now follows by Corollary 2, that $W \in \text{CoCon}_{DS}(X, Y)$. Therefore if C is an ancestor of a vertex in $\mathbf{S} \cup \mathbf{W}$, then C is also an ancestor of a vertex in $\mathbf{S} \cup (\mathbf{W} \cap \text{CoCon}_{DS}(X, Y))$. Hence π d-connects X and Y given $\mathbf{S} \cup (\mathbf{W} \cap \text{CoCon}_{DS}(X, Y))$, which is a contradiction. □

Lemma 3 *Let CG be a chain graph with vertex set $\mathbf{V}$; $X, Y \in \mathbf{V}$ and $\mathbf{W} \subseteq \mathbf{V} \backslash \{X, Y\}$. Let H be the undirected graph* $\text{Aug}[CG; X, Y, \mathbf{W}]$. *If there*

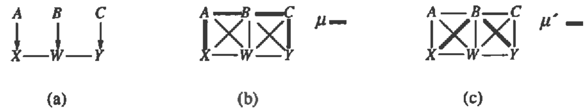

Figure 14. (a) A chain graph CG with $\text{CoCon}_{AMP}(CG; X, Y) = \{B, W\}$; (b) a path μ in $\text{Aug}[CG; X, Y, \{W\}]$; (c) a path μ' in $\text{Aug}[CG; X, Y, \{W\}]$ every vertex of which occurs on μ and is in $\text{CoCon}_{AMP}(CG; X, Y)$.

is a path μ connecting X and Y in H, then there is a path μ' connecting X and Y in H such that if V is a vertex on μ' then V is on μ, and $V \in CoCon_{AMP}(X, Y) \cup \{X, Y\}$.

Proof: If X and Y are adjacent in H then the claim is trivial since $\mu' \equiv \langle X, Y \rangle$ satisfies the lemma. Suppose then that X and Y are not adjacent in H.

Let the vertices on μ be $\langle X \equiv X_1, \ldots, X_n \equiv Y \rangle$. Let α be the greatest j such that X_j is adjacent to X in H. Let β be the smallest $k > \alpha$ such that X_k is adjacent to Y in H. (Since X and Y are not adjacent $\alpha, \beta < n$.)

It is sufficient to prove that $\{X_\alpha, \ldots, X_\beta\} \subseteq \text{CoCon}_{AMP}(X, Y)$, since then the path $\mu' \equiv \langle X, X_\alpha, \ldots, X_\beta, Y \rangle$ satisfies the conditions of the lemma.

This can be proved by showing that there is a path in $[CG]^{Ins}_{AMP}$ from X to each X_i ($\alpha \leq i \leq \beta$) which does not contain Y. A symmetric argument shows that there is also a path from Y to X_i ($\alpha \leq i \leq \beta$) in $[CG]^{Ins}_{AMP}$, which does not contain X. The proof is by induction on i.

Base case: $i = \alpha$. Since X is adjacent to X_α in H, either there is a (directed or undirected) edge between X and X_i in CG, or the edge was added via augmentation of a triplex or bi-flag in $\text{Ext}(CG, \text{Anc}(\{X, Y\} \cup \mathbf{W}))$. In the former case there is nothing to prove since X and X_i are adjacent in $[CG]^{Ins}_{AMP}$. If the edge was added via augmentation of a triplex then there is a vertex T such that $\langle X, T, X_i \rangle$ is a triplex in CG, hence T is adjacent to X and X_i in $[CG]^{Ins}_{AMP}$. Since X and Y are not adjacent in H, $T \not\equiv Y$, so $\langle X, T, X_i \rangle$ is a path which satisfies the claim. If the edge was added via augmentation of a bi-flag then there are two vertices T_0, T_1, forming a bi-flag $\langle X, T_0, T_1, X_i \rangle$. From the definition of augmentation it then follows that T_0 and T_1 are adjacent to X and X_i in H. Since we suppose that X and Y are not adjacent in H it follows that neither T_0 nor T_1 can be Y. Hence $\langle X, T_0, T_1, X_i \rangle$ is a path in $[CG]^{Ins}_{AMP}$ satisfying the claim.

Inductive case: $i > \alpha$; suppose that there is a path from X to X_{i-1} in $[CG]^{Ins}_{AMP}$ which does not contain Y.

Since $i - 1 < \beta$, X_{i-1} is not adjacent to Y in H. By a similar proof to that in the base case it can easily be shown that there is a path from X_{i-1} to X_i in $[CG]^{Ins}_{AMP}$ which does not contain Y. This path may then be

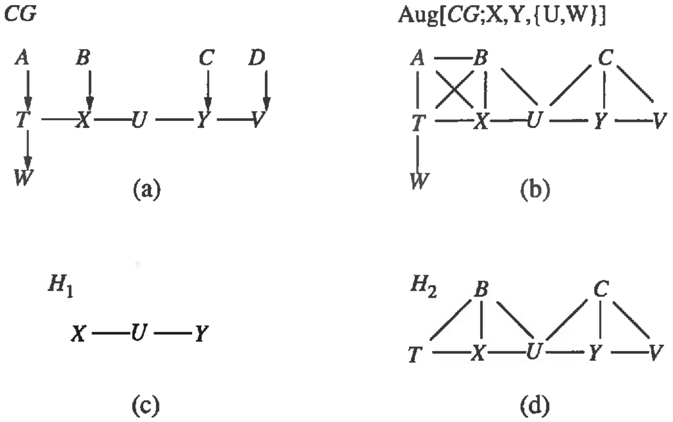

Figure 15. (a) A chain graph, CG, in which $\text{CoCon}_{AMP}(CG; X, Y) = \{U\}$; (b) the undirected graph $\text{Aug}[CG; X, Y, \{U, W\}]$; (c) H_1, the induced subgraph of $\text{Aug}[CG; X, Y, \{U, W\}]$ over $\text{CoCon}_{AMP}(X, Y) \cup \{X, Y\} = \{U, X, Y\}$; (d) the undirected graph H_2, $\text{Aug}[CG; X, Y, \{U, W\} \cap \text{CoCon}_{AMP}(X, Y)]$.

concatenated with the path from X to X_{i-1} (whose existence is guaranteed by the induction hypothesis) to form a path connecting X and X_{i-1} in $[CG]^{Ins}_{AMP}$ which does not contain Y. □

Lemma 4 *Let CG be a chain graph, and let H_1 be the induced subgraph of $Aug[CG; X, Y, \mathbf{W}]$ over $CoCon_{AMP}(X, Y) \cup \{X, Y\}$. Let H_2 be the undirected graph $Aug[CG; X, Y, \mathbf{W} \cap CoCon_{AMP}(X, Y)]$. H_1 is a subgraph of H_2.*

Proof: We first prove that if a vertex V is in H_1 then V is in H_2. If V occurs in H_1 then $V \in \text{CoCon}_{AMP}(X, Y) \cup \{X, Y\}$. Clearly X and Y occur in both H_1 and H_2, so suppose that $V \in \text{CoCon}_{AMP}(X, Y)$.

It follows from the definition of the extended graph that if V is a vertex in $\text{Ext}(CG, \mathbf{T})$ then there is a path consisting of undirected edges from V to some vertex in $\mathbf{T}$. Since V is in $\text{Ext}(CG, \text{Anc}(\{X, Y\} \cup \mathbf{W}))$ there is a path π of the form $\langle V \equiv X_0 - \cdots - X_n \rightarrow \cdots \rightarrow X_{n+m} \equiv W \rangle$ in CG, where $W \in \{X, Y\} \cup \mathbf{W}$, and $n, m \geq 0$. Let X_k be the first vertex on π which is in $\{X, Y\} \cup \mathbf{W}$, i.e. $\forall i \; (0 \leq i < k) \; X_i \notin \{X, Y\} \cup \mathbf{W}$. Now, if $X_k \in \mathbf{W}$ then since $V \in \text{CoCon}_{AMP}(X, Y)$, and $\pi(V, X_k)$ is a path from V to X_k which does not contain X or Y, it follows that $X_k \in \mathbf{W} \cap \text{CoCon}_{AMP}(X, Y)$. Hence V occurs in $\text{Ext}(CG, \text{Anc}(\mathbf{W} \cap \text{CoCon}_{AMP}(X, Y)))$, and so also in H_2. Alternatively, if $X_k \in \{X, Y\}$, then again X_k occurs in $\text{Ext}(CG, \text{Anc}(\{X, Y\}))$, and thus in H_2.

Hence, if there is an edge $A - B$ in H_1, then A and B occur in H_2. There are three reasons why there may be an edge in H_1:

(a) There is an edge (directed or undirected) between A and B in CG. It then follows immediately that there is an edge between A and B in H_2.

(b) The edge between A and B in H_1 is the result of augmenting a triplex in $\text{Ext}(CG, \text{Anc}(\{X,Y\} \cup \mathbf{W}))$. Then there is some vertex T such that $\langle A, T, B\rangle$ forms a triplex in $\text{Ext}(CG, \text{Anc}(\{X,Y\} \cup \mathbf{W}))$. Since, by hypothesis, $A, B \in \text{CoCon}_{AMP}(X,Y)$, it follows that $T \in \text{CoCon}_{AMP}(X,Y) \cup \{X,Y\}$, and hence T occurs in H_1. It then follows by the previous reasoning that T is in H_2, and so the triplex is also present in $\text{Ext}(CG, \text{Anc}(\{X,Y\} \cup (\mathbf{W} \cap \text{CoCon}_{AMP}(X,Y))))$. Hence there is an edge between A and B in H_2.

(c) The edge between A and B in H_1 is the result of augmenting a bi-flag in $\text{Ext}(CG, \text{Anc}(\{X,Y\} \cup \mathbf{W}))$. This case is identical to the previous one, except that there are two vertices T_0, T_1, such that $\langle A, T_0, T_1, B\rangle$ forms a bi-flag in $\text{Ext}(CG, \text{Anc}(\{X,Y\} \cup \mathbf{W}))$. As before, it follows from the hypothesis that $A, B \in \text{CoCon}_{AMP}(X,Y)$, that $T_0, T_1 \in \text{CoCon}_{AMP}(X,Y) \cup \{X,Y\}$, hence T_0 and T_1 occur in H_2 and the bi-flag is in $\text{Ext}(CG, \text{Anc}(\{X,Y\} \cup (\mathbf{W} \cap \text{CoCon}_{AMP}(X,Y))))$. Thus the A—B edge is also present in H_2. □

Theorem 2 (iii) *Chain graphs are co-connection$_{AMP}$ determined.*

Proof: $(CG \models_{AMP} X \perp\!\!\!\perp Y \mid \mathbf{W} \Rightarrow CG \models_{AMP} X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_{AMP}(X,Y))$
Let H be $\text{Aug}[CG; X, Y, \mathbf{W}]$. Since $CG \models_{AMP} X \perp\!\!\!\perp Y \mid \mathbf{W}$, X and Y are separated given $\mathbf{W}$ in H.
Claim: X and Y are separated in H by $\mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)$. Suppose, for a contradiction, that there is some path μ in H, connecting X and Y on which there is no vertex in $\mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)$. It then follows from Lemma 3 that there is a path μ' in H composed only of vertices on μ which are in $\text{CoCon}_{AMP}(X,Y)$. Since no vertex on μ is in $\mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)$, it then follows that no vertex on μ' is in $\mathbf{W}$. So X and Y are not separated by $\mathbf{W}$ in H, contradicting the hypothesis.

However, $\text{Aug}[CG; X, Y, \mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)]$ is a subgraph of H, so X and Y are separated by $\mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)$ in $\text{Aug}[CG; X, Y, \mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)]$. Thus $CG \models_{AMP} X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)$.

$(CG \models_{AMP} X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_{AMP}(X,Y) \Rightarrow CG \models_{AMP} X \perp\!\!\!\perp Y \mid \mathbf{W})$
The proof is by contraposition. Suppose that there is a path μ from X to Y in $\text{Aug}[CG; X, Y, \mathbf{W}]$. Lemma 3 implies that there is a path μ' from X to Y in $\text{Aug}[CG; X, Y, \mathbf{W}]$ every vertex of which is in $\{X,Y\} \cup \text{CoCon}_{AMP}(X,Y)$. It then follows from Lemma 4 that this path exists in $\text{Aug}[CG; X, Y, \mathbf{W} \cap \text{CoCon}_{AMP}(X,Y)]$. □

References

1. S. A. Andersson, D. Madigan, and M. D. Perlman. An alternative Markov property for chain graphs. In F. V. Jensen and E. Horvitz, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 12^{th} *Conference*, pages 40–48, San Francisco, 1996. Morgan Kaufmann.
2. S. A. Andersson, D. Madigan, and M. D. Perlman. A new pathwise separation criterion for chain graphs. In preparation, 1997.
3. J. Besag. On spatial-temporal models and Markov fields. In *Transactions of the* 7^{th} *Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 47–55. Academia, Prague, 1974.
4. J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Royal Statist. Soc. Ser. B*, 36:302–309, 1974.
5. G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.
6. D. R. Cox. Causality and graphical models. In *Proceedings,* 49^{th} *Session*, volume 1 of *Bulletin of the International Statistical Institute*, pages 363–372, 1993.
7. D. R. Cox and N. Wermuth. *Multivariate Dependencies: Models, Analysis and Interpretation.* Chapman and Hall, London, 1996.
8. J. N. Darroch, S. L. Lauritzen, and T. P. Speed. Markov fields and log-linear models for contingency tables. *Ann. Statist.*, 8:522–539, 1980.
9. A. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.
10. F. M. Fisher. A correspondence principle for simultaneous equation models. *Econometrica*, 38(1):73–92, 1970.
11. M. Frydenberg. The chain graph Markov property. *Scandin. J. Statist.*, 17:333–353, 1990.
12. D. Geiger. *Graphoids: a qualitative framework for probabilistic inference.* PhD thesis, UCLA, 1990.
13. J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. Unpublished manuscript, 1971.
14. D. Hausman. Causal priority. *Nous*, 18:261–279, 1984.
15. D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. In B. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 10^{th} *Conference*, pages 293–301, San Francisco, 1994. Morgan Kaufmann.
16. J. T. A. Koster. Markov properties of non-recursive causal models. *Ann. Statist.*, 24:2148–2178, October 1996.
17. S. L. Lauritzen. *Graphical Models.* Number 81 in Oxford Statistical Science Series. Springer-Verlag, 1993.
18. S. L. Lauritzen, A. P. Dawid, B. Larsen, and H.-G. Leimer. Independence properties of directed Markov fields. *Networks*, 20:491–505, 1990.
19. S. L. Lauritzen and D. J. Spiegelhalter. Local computation with probabilities in graphical structures and their application to expert systems (with discussion). *J. Royal Statist. Soc. Ser. B*, 50(2):157–224, 1988.
20. S. L. Lauritzen and N. Wermuth. Graphical models for association between variables, some of which are qualitative and some quantitative. *Ann. Statist.*, 17:31–57, 1989.
21. C. Meek. Strong completeness and faithfulness in Bayesian networks. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 11^{th} *Conference*, pages 403–410, San Francisco, 1995. Morgan Kaufmann.
22. J. Pearl. *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufman, 1988.
23. J. Pearl. Causal diagrams for empirical research (with discussion). *Biometrika*, 82:669–690, 1995.
24. J. Pearl and R. Dechter. Identifying independencies in causal graphs with feedback. In F. V. Jensen and E. Horvitz, editors, *Uncertainty in Artificial Intelligence:*

Proceedings of the 12^{th} *Conference*, pages 454–461, San Francisco, 1996. Morgan Kaufmann.

25. J. Pearl and T. Verma. A theory of inferred causation. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*, pages 441–452, San Mateo, CA, 1991. Morgan Kaufmann.
26. T. S. Richardson. A discovery algorithm for directed cyclic graphs. In F. V. Jensen and E. Horvitz, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 12^{th} *Conference*, pages 454–461, San Francisco, 1996. Morgan Kaufmann.
27. T. S. Richardson. *Models of feedback: interpretation and discovery.* PhD thesis, Carnegie-Mellon University, 1996.
28. T. S. Richardson. A polynomial-time algorithm for deciding Markov equivalence of directed cyclic graphical models. In F. V. Jensen and E. Horvitz, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 12^{th} *Conference*, pages 462–469, San Francisco, 1996. Morgan Kaufmann.
29. T. P. Speed. A note on nearest-neighbour Gibbs and Markov distributions over graphs. *Sankhya Ser. A*, 41:184–197, 1979.
30. P. Spirtes. Directed cyclic graphical representations of feedback models. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 11^{th} *Conference*, pages 491–498, San Francisco, 1995. Morgan Kaufmann.
31. P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction and Search.* Lecture Notes in Statistics. Oxford University Press, 1996.
32. P. Spirtes, C. Meek, and T.S. Richardson. Causal inference in the presence of latent variables and selection bias. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 11^{th} *Conference*, pages 403–410, San Francisco, 1995. Morgan Kaufmann.
33. P. Spirtes and T. S. Richardson. A polynomial-time algorithm for determining dag equivalence in the presence of latent variables and selection bias. In D. Madigan and P. Smyth, editors, *Preliminary papers of the Sixth International Workshop on AI and Statistics, January 4-7, Fort Lauderdale, Florida*, pages 489–501, 1997.
34. P. Spirtes and T. Verma. Equivalence of causal models with latent variables. Technical Report CMU-PHIL-33, Department of Philosophy, Carnegie Mellon University, October 1992.
35. R. H. Strotz and H. O. A. Wold. Recursive versus non-recursive systems: an attempt at synthesis. *Econometrica*, 28:417–427, 1960. (Also in Causal Models in the Social Sciences, H.M. Blalock Jr. ed., Chicago: Aldine Atherton, 1971).
36. M. Studený and R. Bouckaert. On chain graph models for description of conditional independence structures. *Ann. Statist.*, 1996. Accepted for publication.
37. T. Verma and J. Pearl. Equivalence and synthesis of causal models. In M. Henrion, R. Shachter, L. Kanal, and J. Lemmer, editors, *Uncertainty in Artificial Intelligence: Proceedings of the* 12^{th} *Conference*, pages 220–227, San Francisco, 1996. Morgan Kaufmann.
38. S. Wright. Correlation and Causation. *J. Agricultural Research*, 20:557–585, 1921.