Extensions of Undirected and Acyclic, Directed Graphical Models

Thomas Richardson[1]
Statistics Department, University of Washington
(tsr@stat.washington.edu)

Abstract

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, 1988, Lauritzen *et al.* 1988), and in inferring causal structure (Pearl and Verma, 1991; Cooper and Herskovits 1992; Spirtes, Glymour and Scheines, 1993).

In section 1 I will survey a number of extensions of the DAG framework based on directed graphs and chain graphs (Lauritzen and Wermuth 1989; Frydenberg 1990; Koster 1996; Andersson, Madigan and Perlman 1996). Those based on directed graphs include models based on directed cyclic and acyclic graphs, possibly including latent variables and/or selection bias (Pearl, 1988; Spirtes, Glymour and Scheines 1993; Spirtes 1995; Spirtes, Meek, and Richardson 1995; Richardson 1996a, 1996b; Koster 1996; Pearl and Dechter 1996; Cox and Wermuth, 1996).

In section 2 I state two properties, motivated by causal and spatial intuitions, that the set of conditional independencies entailed by a graphical model might satisfy. I proceed to show that the sets of independencies 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. The examination of these properties and others like them may provide insight into the current vigorous debate concerning the applicability of chain graphs under different global Markov properties.

1. Graphs and Probability Distributions

An *undirected* graph *UG* is an ordered pair (**V**,**U**), where **V** is a set of vertices and **U** is a set of undirected edges X—Y between vertices.[2] Similarly, a *directed* graph *DG* is an ordered pair (**V**,**D**) where **D** is a set of directed edges X→Y between vertices in **V**. A *directed cycle* consists of a sequence of edges $X_1 \to X_2 \ldots \to X_n \to X_1$ ($n \geq 2$). If a directed graph *DG* contains no directed cycles it is said to be *acyclic*, otherwise it is *cyclic*. An edge X→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. $X \rightleftarrows Y$.

---

[1] [1]I thank P. Spirtes, C. Glymour, D. Madigan, M. Perlman and J. Besag for helpful conversations. Research for this paper was supported by the Office of Naval Research through contract number N00014-93-1-0568.

[2]Bold face (**X**) indicate sets; plain face (X) indicates individual elements; italics (*U*) indicates a graph or a path.

I will consider directed graphs (cyclic or acyclic) in which **V** is partitioned into three disjoint sets **O** (Observed), **S** (Selection) and **L** (Latent), written *DG*(**O**,**S**,**L**). The interpretation of this definition is that *DG* represents a causal mechanism, **O** represents the subset of the variables that are observed, **S** represents a set of 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 **L** are not observed and for this reason are called 'latent'.[3]

A *mixed graph* contains both directed and undirected edges. A *partially directed cycle* in a mixed graph *G* is a sequence of n distinct vertices $X_1,\ldots X_n$, ($n\geq 3$), and $X_{n+1}\equiv X_1$, such that (a) $\forall i$ ($1\leq i\leq n$) either X_i—X_{i+1} or $X_i\rightarrow X_{i+1}$, and (b) $\exists j$ ($1\leq j\leq n$) such that $X_j\rightarrow X_{j+1}$.

A *chain graph CG* is a mixed graph in which there are no partially directed cycles. Koster (1996) considers a class of *reciprocal* graphs containing directed and undirected edges in which partially directed cycles are allowed. I do not consider such graphs separately here, though many of the comments which apply to LWF chain graphs apply also to reciprocal graphs since the former are a subclass of the latter.

To make clear which kind of graph is being referred to I will use *UG* for undirected graphs, *DG* for directed graphs, *AG* for acyclic directed graphs, *CG* for chain graphs, and *G* to denote 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 $X_i\leftarrow X_{i+1}$ ($1\leq i\leq n$). A *directed path from* X *to* Y is a path of the form X→…→Y.[4]

1.2 Global Markov Properties Associated with Graphs

A *Global Markov Property* associates a set of conditional independence relations with a graph *G*.[5] In an undirected graph *UG*, for disjoint sets of vertices **X**, **Y** and **Z**, (**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 **Z**, then **X** and **Y** are said to be *separated* by **Z**.

Undirected Global Markov Property ($\models_S$):

$UG\models_S \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if **X** and **Y** are separated by **Z** in *UG*.[6]

In a graph *G*, X is a *parent* of Y, (and Y is a *child* of X) if there is an edge X→Y in *G*. X is an *ancestor* of Y (and Y is a *descendant* of X) if X=Y, or there is a directed path X→…→Y from X

[3]Note that we use the terms 'variable' and 'vertex' interchangeably.

[4]Path is defined here as a sequence of edges, rather than vertices; in a 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.

[5]Often global Markov conditions are introduced as a means for deriving the consequences of a set of local Markov conditions. Here I merely define the Global property in terms of the relevant graphical criterion.

[6]'$\mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$' means that '**X** is independent of **Y** given **Z**'; if $\mathbf{Z}=\emptyset$, the abbreviation $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$ is used; if **X**, **Y** and/or **Z** are singleton sets {V}, then brackets are omitted e.g. $V \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$, instead of $\{V\} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$.

to Y. A pair of consecutive edges on a path *P* in *G* 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 *P*, otherwise A is a *non-collider* on *P*.

For distinct vertices X and Y, and set $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X,Y\}$, a path *P* between X and Y given **Z** is said to *d-connect* X and Y given **Z** if every collider on *P* is an ancestor of a vertex in **Z**, and no non-collider on *P* is in **Z**. Disjoint sets **X** and **Y** are said to be *d-connected* given **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 **Z**. If there is no such path then **X** and **Y** are said to be *d-separated* given **Z** (see Pearl, 1988).

Global Markov Property for Directed Graphs; d-separation ($|=_{DS}$):

$DG \models_{DS} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if **X** and **Y** are d-separated by **Z** in *DG*.

For *DG*(**O**,**S**,**L**), and disjoint subsets $\mathbf{X} \dot{\cup} \mathbf{Y} \dot{\cup} \mathbf{Z} \subseteq \mathbf{O}$ we define:

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

Since, under the interpretation of *DG* (**O**,**S**,**L**), the only observed variables are in **O**, we do not observe conditional independence relations involving variables in **L**. Similarly, since samples are drawn from a subpopulation in which all variables in **S** were conditioned on, **S** is conditioned upon in every conditional independence relation we observe to hold in the sample. Thus this definition gives the set of conditional independencies in the observed distribution P(**O**|**S**). (See Spirtes and Richardson, this volume; Spirtes, Meek and Richardson, 1996; Cox and Wermuth, 1996.)

Two different Global Markov properties have been proposed for Chain Graphs. In both definitions a conditional independence relation is entailed if sets **X** and **Y** are separated by **Z** in an undirected graph whose edges are a superset of those in the original chain graph.

A vertex V in a chain graph is said to be *anterior* to a set **W** if there is a path *P* from V to some $W \in \mathbf{W}$ in which all directed edges ($X \rightarrow Y$) on the path (if any) are such that Y is between X and W on *P*, *Ant*(**W**)={V| V is anterior to **W**}. Let *CG*(**W**) denote the induced subgraph of *CG* obtained by removing all vertices in **V****W** and edges with an endpoint in **V****W**. A *complex* in *CG* is an induced subgraph with the following form: $X \rightarrow V_1 — \ldots — V_n \leftarrow Y$ ($n \geq 1$). 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.

Lauritzen-Wermuth-Frydenberg Global Markov Property ($|=_{LWF}$):

$CG \models_{LWF} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if **X** is separated from **Y** by **Z** in Moral(*CG*(Ant(**X**∪**Y**∪**Z**)))

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*(**W**) ={V | V is connected to some $W \in \mathbf{W}$}. The *extended subgraph*, *Ext*(*CG*(**W**)), has vertex set Con(**W**) and contains all directed edges in *CG*(**W**), and all undirected edges in *CG* (Con(**W**)). A triple of vertices <X,Y,Z> is said to form a *triplex* in *CG* if the induced subgraph *CG*({X,Y,Z}) is either X→Y—Z, X→Y←Z, or X—Y←Z. A triplex is *augmented* by adding the X—Z edge. *Aug*(*CG*) is the undirected graph formed by augmenting all triplexes in *CG* and replacing all directed edges with undirected edges.

Andersson-Madigan-Perlman Global Markov Property ($\models_{AMP}$)

$CG \models_{AMP} \mathbf{X} \perp\!\!\!\perp \mathbf{Y} \mid \mathbf{Z}$ if **X** is separated from **Y** by **Z** in Aug(Ext(*CG*(Anc($\mathbf{X} \cup \mathbf{Y} \cup \mathbf{Z}$))))

where Anc(**W**)={V| V is an ancestor of some W∈**W**}.

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. In this sense chain graphs with either property are a generalization of both acyclic, directed graphs and undirected graphs.

Examples:

Figure 1: Two Chain Graphs

The conditional independence relations associated with these chain graphs are:

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

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

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

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

1.3 Completeness

For a given global Markov property R, and graph *G* with vertex set **V**, a distribution P is said to be *G-Markovian*$_R$ if for disjoint subsets **X**,**Y** and **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 given global Markov property is said to be *weakly complete* if for all disjoint sets **X**,**Y** and **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.

Except for the AMP property, all of the global Markov properties here are known to be weakly complete (Geiger, 1990; Frydenberg, 1990). For general directed graphs, d-separation, and for chain graphs, the LWF Markov property, have been shown to be strongly complete. (Spirtes 1995; Meek 1995; Spirtes *et al.* 1993; Studený and Bouckaert, 1996.)

2 Inseparability and Related Markov Properties

In this section I will introduce two properties, motivated by spatial and causal intuitions.

Distinct vertices X and Y are *inseparable*$_R$ in *G* under Markov Property R if there is no set **W** such that $G \models_R \mathrm{X} \perp\!\!\!\perp \mathrm{Y} \mid \mathbf{W}$. If X and Y are not inseparable$_R$, they are *separable*$_R$. Let $[G]_R^{Ins}$ 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* (**O**,**S**,**L**), $[DG(\mathbf{O},\mathbf{S},\mathbf{L})]_{DS}^{Ins}$ is defined to have vertex set **O**.

For an undirected graph model $[UG]^{\text{Ins}}_{S}$ 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 LWF or AMP $[G]^{\text{Ins}}_{R}$ is simply the undirected graph formed by replacing all directed edges with undirected edges. 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, 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 (recall that in *DG*(**O**,**S**,**L**), we restrict ourselves to the observed conditional independencies), inseparability$_{DS}$ is *not* a sufficient condition for there to be an edge between a pair of variables. An *inducing path* between X and Y is a path *P* between X and Y on which (i) every vertex in $\mathbf{O} \cup \mathbf{S}$ is a collider on *P*, and (ii) every collider is an ancestor of X, Y or **S**.[7] In a directed graph *DG*(**O**,**S**,**L**), variables X,Y$\in$**O**, are inseparable$_{DS}$ if and only if there is an inducing path between X and Y in *DG*(**O**,**S**,**L**).[8]

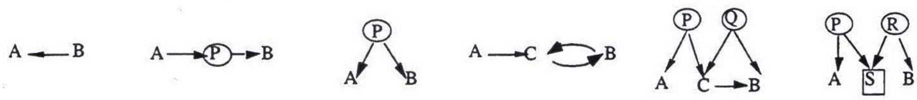

Figure 2: Examples of directed graph models in which A and B are inseparable$_{DS}$, (variables in **L** are circled; variables in **S** are boxed; variables in **O** are marked with nothing).

2.1 '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 $\langle X \equiv X_0, X_1, \ldots X_n \equiv B, X_{n+1}, \ldots X_{n+m} \equiv Y \rangle$ 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 $[G]^{\text{Ins}}_{R}$. The set of vertices between X and Y under property R is denoted *Between*$_R$(X,Y).

Between$_R$ Separated: A model *G* is *between$_R$ separated*, if for all pairs of vertices X, Y and sets **W** (X,Y$\notin$**W**): $G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \Rightarrow G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{Between}_R(X,Y)$

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.

$[G]^{\text{Ins}}_{R}$

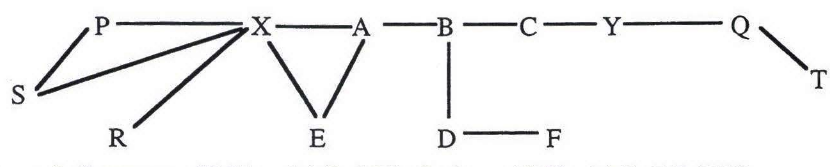

Figure 2: Between$_R$(X,Y) = {A,B,C,E}, CoCon$_R$(X,Y)={A,B,C,D,E,F}
P,Q,R,S,T are vertices not in CoCon$_R$(X,Y)

---

[7]The notion of an inducing path was first introduced, for acyclic directed graphs with latent variables in Verma and Pearl (1990), it was subsequently extended in Spirtes, Meek and Richardson (1995).

[8]Inseparability is a necessary and sufficient condition for there to be an edge between a pair of variables in a Partial Ancestral Graph (PAG), (Richardson 1996a), which represents structural features common to a given Markov equivalence class of directed graphs, possibly with latent and/or selection variables.

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 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.[9]

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: We give here the proof for undirected graph models. 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 P 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 P, and $V \in \mathbf{W} \backslash \text{Between}_S(X,Y)$. But if V is on P, then P 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 1: $CG_1 \models_{LWF} A \perp\!\!\!\perp D \mid \{B,C\}$, so A and D are separable$_{LWF}$, but $\text{Between}_{LWF}(A,D)=\{C\}$ and $CG_1 \not\models_{LWF} A \perp\!\!\!\perp D \mid \{C\}$. For the AMP property note that $CG_2 \models_{AMP} B \perp\!\!\!\perp D \mid \{A,C\}$ but $\text{Between}_{AMP}(B,D)=\{C\}$, and yet $CG_2 \not\models_{AMP} B \perp\!\!\!\perp D \mid \{C\}$.

2.2 'Co-Connection Determined' Models

A vertex W will be said to be *co-connected* $_R$ to X and Y in G if:

(i) there is a sequence of vertices $\langle X,A_1,A_2\ldots A_n,W\rangle$ in G which does not contain Y such that 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 which does not contain X such that consecutive pairs of variables in the sequence are inseparable$_R$ in G under R.

Let $\text{CoCon}_R(X,Y) = \{V \mid V \text{ is co-connected}_R \text{ to X and Y}\}$.

It is easy to see that B will be co-connected$_R$ to X and Y in G, if and only if (i) B is not separated from Y by X in $[G]_R^{\text{Ins}}$, and (ii) B is not separated from X by Y in $[G]_R^{\text{Ins}}$.

[9] Where for A and B to be 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).

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

A model G will be said to be *co-connection$_R$ determined*, if for all pairs of vertices X, Y and sets **W** (X,Y∉**W**): $G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \Leftrightarrow G \models_R X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{CoCon}_R(X,Y)$

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

Theorem 2

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

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

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

Proof: We present here the proof for undirected graphs. The proof for directed graph models is given in the appendix; For reasons of space the proof for AMP chain graphs is not included though it is quite similar to the proof for (i) and (ii).

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$_S$') suffices to show 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 **W** in UG.∴

In fact, for undirected graphs $UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \Leftrightarrow UG \models_S X \perp\!\!\!\perp Y \mid \mathbf{W} \cap \text{Between}_S(X,Y)$, 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 $\text{CoCon}_{LWF}(B,C) = \{D\}$ and $CG_1 \not\models_{LWF} B \perp\!\!\!\perp C \mid \{D\}$. In contrast, chain graphs are co-connection$_{AMP}$ determined.

2.3 Discussion.

The two Markov properties presented here are based on the intuition that only vertices which, in some sense, come "between" X and Y should be relevant to whether or not two vertices in a graph are entailed to be independent. Both of these properties are satisfied by undirected graphs, and by all forms of directed graph model. Since neither of these properties are satisfied by chain graphs under the LWF interpretation these properties capture a qualitative difference between undirected and directed graphs, and LWF chain graphs. In this respect, at least, AMP chain graphs are less dissimilar to directed and undirected graphs since chain graphs are co-connection$_{AMP}$ determined.

Since the pioneering work of Sewall Wright (1921), models based on directed graphs have been used to model causal relations, and data generating processes. Strotz and Wold (1958), Spirtes *et al.* (1993) and Pearl(1995) develop a theory of intervention for directed graph models which makes it possible to calculate the effect of intervening in a system in certain ways. Models allowing directed graphs with cycles, have been used for over 50 years in econometrics, and allow the possible of representing certain kinds of feedback, or two-way interaction. Besag (1974) gives several spatial-temporal data generating processes whose limiting spatial distributions satisfy the Markov property with respect to a naturally associated undirected graph..

In contrast Cox (1993) states that chain graphs under the LWF Markov property do "not satisfy the requirement of specifying a direct mode of data generation." This statement is given additional support by the failure of LWF chain graphs to satisfy either of the properties given above. AMP chain graphs seem more compatible with a data generating process since chain graphs are co-connection$_{AMP}$ determined (See also Andersson *et al.* 1996).

Models which are co-connection determined have a very different character from those which are not: in a co-connection determined model, the inclusion or exclusion of vertices that are not co-connected to X and Y from some set **W** is irrelevant to the question of whether X and Y are entailed to be independent given **W**. Given that a large class of well-understood models which can be interpreted directly as data generating processes possess this property, it would seem that a researcher would have to have a quite particular justification for using an LWF chain graph to model a given system.

3. Appendix - Proofs

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

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

where |**T**(*U*)| is the cardinality of **T**(*U*).*U* is a *minimal d-connecting path* between X and Y given $\mathbf{Z}\cup\mathbf{S}$, if *U* d-connects X and Y given $\mathbf{Z}\cup\mathbf{S}$ and there is no other path *U′* that d-connects X and Y given **Z** such that *size*(*U′*) < *size*(*U*). If there is a path that d-connects X and Y given **Z** then there is at least one minimal d-connecting path between X and Y given **Z**. In the following proofs *U*(A,B) denotes the subpath of *U*between vertices A and B.

Lemma 1: If U is a minimal d-connecting path between X and Y given **Z**∪**S** in DG(**O**,**S**,**L**) then for each collider C_i on U that is not an ancestor of **S**, there is a directed path D_i from C_i to some vertex in **Z**, such that D_i intersects U only at C_i, D_i and D_j do not intersect (i≠j) and no vertex on any path D_i is in **S**.

Proof: Let D_i be a shortest acyclic directed path from a collider C_i on U to a member of **Z**, where C_i is not an ancestor of **S** . We will prove that D_i does not intersect U except at C_i by showing that if such a point of intersection existed then U would not be minimal, contrary to our assumption. See the figure 4 below:

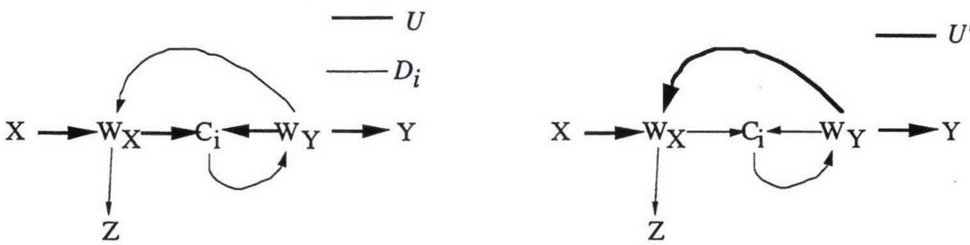

Figure 4. Finding a d-connecting path U' of smaller size than U.

Form the path U' in the following way: If D_i intersects U at a vertex other than C_i then let W_X be the vertex closest to X on U that is on both D_i and U, and let W_Y be the vertex closest to Y on U that is on both D_i and U. Suppose without loss of generality that W_X is after W_Y on D_i. Let U' be the concatenation of $U(X,W_X)$, $D_i(,W_X,W_Y)$, and $U(W_Y,Y)$. It is now easy to show that U' d-connects X and Y given **Z**∪**S**, and $size(U') < size(U)$ because, U' contains no more colliders than U and a shortest directed path from W_X to a member of **Z** is shorter than D_i. Hence U is not minimal, contrary to the assumption.

Next, we will show that if U is minimal, then D_i and D_j (i≠j) do not intersect. Suppose this is false. See figure 5 below:

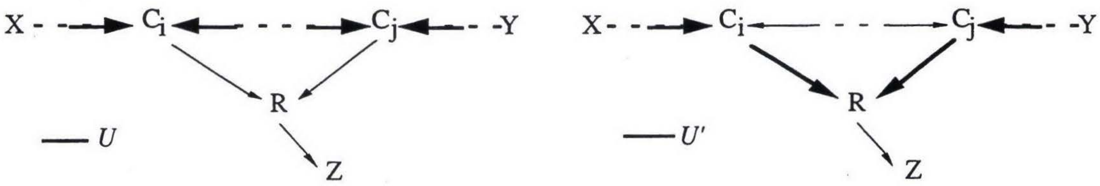

Figure 5. Finding a d-connecting path of smaller size.

Let the vertex on D_i closest to C_i that is also on D_j be R. Let U' be the concatenation of $U(X,C_i)$, $D_i(C_i,R)$, $D_j(R,C_j)$, and $U(C_j,Y)$. It is now easy to show that U' d-connects X and Y given **Z**∪**S** and $size(U') < size(U)$ because C_i and C_j are not colliders on U', the only collider on U' that may not be on U is R, and the length of a shortest path from R to a member of **Z** is less than the length of a shortest path from C_i to a member of **Z**. Hence U is not minimal, contrary to the assumption. Since each C_i is not an ancestor of **S**, it follows directly that no vertex on any path D_i is in **S**.

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

Proof: Since U is a d-connecting path given $\mathbf{S}\cup\mathbf{R}$ every collider on U that is not an ancestor of $\mathbf{S}$ is an ancestor of a vertex in $\mathbf{Z}$. Denote the colliders on U, that are not ancestors of $\mathbf{S}$ as $C_1,\ldots C_k$. Let D_j be a shortest directed path from C_j to some vertex $R_j\in\mathbf{R}$. It follows by the previous Lemma that D_j and U intersect only at C_j, and that D_j and $D_{j'}$ $(j\neq j')$ do not intersect. We now construct a sequence of vertices X_i in $\mathbf{O}$, s.t. each X_i is either on U or is on a directed path D_j from C_j to R_j.

Base Step: Let $X_0\equiv X$.

Inductive Step: If X_i is on some path D_j then let W be C_j otherwise, if X_i is on U, then let W be X_i. Let V be the next vertex on U, after W, such that $V\in\mathbf{O}$. If there is no vertex $C_{j'}$ between W and V on U, then let $X_{i+1}\equiv V$. Otherwise let C_{j*} be the first collider on U that is not an ancestor of $\mathbf{S}$, and let X_{i+1} be the first vertex in $\mathbf{O}$ on the directed path D_{j*} (such a vertex is guaranteed to exist since R_{j*}, the endpoint of D_{j*} is in $\mathbf{O}$).

It follows from the construction that if B is on U, 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 U, then $U(X_i,X_{i+1})$ is a path on which no vertex, except the endpoints, is in $\mathbf{O}$, and every collider is an ancestor of $\mathbf{S}$. Thus $U(X_i,X_{i+1})$ d-connects X_i and X_{i+1} given $\mathbf{R}\cup\mathbf{S}$ for any $\mathbf{R}\subseteq\mathbf{O}\backslash\{X_i,X_{i+1}\}$. So X_i and X_{i+1} are inseparable$_{DS}$.

If X_i lies on some path D_j, but X_{i+1} is on U, then the path P formed by concatenating the directed path $X_i\leftarrow\ldots C_j$ and $U(C_j,X_{i+1})$ again is such that, excepting the endpoints, no vertex on P is in $\mathbf{O}$, and every collider on P 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 U, can be handled similarly.

This completes the proof.∴

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

Proof: This follows directly from Lemma 2

Corollary 2: If U 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 U that is an ancestor of $\mathbf{Z}$, but not $\mathbf{S}$, D is a shortest directed path from C to some $Z\in\mathbf{Z}$, then $Z\in\text{CoCon}_{DS}(X,Y)$.

Proof: By Lemma 1, D does not intersect U except at C. It follows from Corollary 1 that $C\in\text{Between}_{DS}(X,Y)$. Hence there is a sequence of vertices $<X\equiv X_0,X_1,\ldots X_n\equiv C, X_{n+1},\ldots X_{n+m}\equiv Y>$ in $\mathbf{O}$ such that consecutive pairs of vertices are inseparable$_{DS}$. Let the sequence of vertices on D that are in $\mathbf{O}$ be $<C\equiv V_1,\ldots V_r\equiv Z>$. Since, by hypothesis C is not an ancestor of $\mathbf{S}$, it follows that no vertex on D is in $\mathbf{S}$.

Hence $D(V_i,V_{i+1})$ is a directed path from V_i to V_{i+1} on which, with the exception of the endpoints, every vertex is in **L** and is a non-collider on D, it follows that V_i and V_{i+1} are inseparable$_{DS}$ in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$. Thus the sequences $<X\equiv X_0,X_1,\ldots X_n\equiv C \equiv V_1,\ldots V_r\equiv Z>$, and $<Y\equiv X_{n+m},\ldots X_n\equiv C\equiv V_1,\ldots V_r\equiv Z>$ establish that $Z\in$ CoCon$_{DS}$(X,Y) in $DG(\mathbf{O},\mathbf{S},\mathbf{L})$.∴

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

Proof: Suppose, for a contradiction, that $DG(\mathbf{O},\mathbf{S},\mathbf{L}) \models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W}$, 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 P 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 $\mathbf{S}\cup\mathbf{W}$. Clearly it is not possible for a collider on P 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 P, s.t. $B\in\mathbf{S}\cup\mathbf{W}$, but $B\notin \mathbf{S}\cup(\mathbf{W}\cap\text{Between}_{DS}(X,Y))$. Clearly 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: A directed graph $DG(\mathbf{O},\mathbf{S},\mathbf{L})$ is co-connection$_{DS}$ determined.

Since $\text{Between}_{DS}(X,Y)\subseteq\text{CoCon}_{DS}(X,Y)$, the proof of Theorem 1 (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 that $DG \models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W}\cap\text{CoCon}_{DS}(X,Y)$, but that $DG(\mathbf{O},\mathbf{S},\mathbf{L}) \not\models_{DS} X \perp\!\!\!\perp Y \mid \mathbf{W}$. It then follows that there is some path which d-connects X and Y given $\mathbf{W}\cup\mathbf{S}$. Let P be a minimal d-connecting path 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 P 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 P 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 $C\in\mathbf{W}\backslash\text{CoCon}_{DS}(X,Y)$.

Consider a shortest directed path D from C to some vertex W in **W**. It follows from Lemma 1, and the minimality of P that D does not intersect P 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 P d-connects given $\mathbf{S}\cup\mathbf{W}\cap\text{CoCon}_{DS}(X,Y)$, which is a contradiction.∴

I do not include here the proof that chain graphs are co-connection$_{AMP}$ determined. The property follows from the fact that when the extended subgraph is augmented, the only edges that are added are between vertices that are both inseparable$_{AMP}$ from some common third vertex. This is an important difference between augmentation, used in the AMP Markov property, and moralization, used in the LWF Markov property.

References

ANDERSSON, S. A., D. MADIGAN, M. D. PERLMAN (1996). An alternative Markov property for chain graphs. *Uncertainty in Artificial Intelligence, Proceedings, 12th Conference.*

BESAG, J. (1974) On spatial-temporal models and Markov fields. Transactions of the 7th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. Academia, Prague, pp.47-55.

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

COX, D.R. (1993). Causality and Graphical Models. *Bulletin of the International Statistical Institute.* Proceedings, 49th Session, **1** 363-372.

COX, D.R., N. WERMUTH (1996). Multivariate Dependencies. Chapman and Hall, London.

FRYDENBERG, M. (1990) The Chain Graph Markov Property. *Scandinavian Journal of Statistics* **17**.

GEIGER, D. (1990). Graphoids: a qualitative framework for probabilistic inference. Ph.D. dissertation, Computer Science Department, UCLA.

HECKERMAN, D., D. GEIGER, D.M. CHICKERING (1994). Learning Bayesian networks: the combination of knowledge and statistical data. In *Uncertainty in Artificial Intelligence, Proceedings, 10th Conference*, Morgan Kaufman, CA.

KOSTER, J.T.A. (1996). Markov Properties of Non-Recursive Causal Models. *Ann. Statist.* **24**

LAURITZEN, S.L., A.P. DAWID, B. LARSEN, H. LEIMER (1990). Independence properties of directed Markov fields, *Networks* **20**

LAURITZEN S.L., D.J. SPIEGELHALTER (1988). Local computation with probabilities in graphical structures and their applications to expert systems (with discussion). *JRSS B*, **50**, No.2.

LAURITZEN, S.L., N. WERMUTH (1989). Graphical models for association between variables, some of which are qualitative and some quantitative. *Annals of Statistics* **17** pp. 31-57.

MEEK, C. (1995) Strong completeness and faithfulness in Bayesian networks. *Uncertainty in Artificial Intelligence, Proceedings, 11th Conference*, Morgan Kaufman, CA.

PEARL, J. (1988). Probabilistic Reasoning in Intelligent Systems. Morgan Kaufman, CA.

PEARL, J. (1995). Causal diagrams for empirical research (with discussion) *Biometrika* **82**

PEARL, J., R. DECHTER (1996). Identifying independencies in causal graphs with feedback. *Uncertainty in Artificial Intelligence, Proceedings, 12th Conference,* Morgan Kaufman, CA.

RICHARDSON, T. (1996a). A discovery algorithm for directed cyclic graphs. Uncertainty *in Artificial Intelligence, Proceedings, 12th Conference,* Morgan Kaufman, CA.

RICHARDSON, T. (1996b). A polynomial-time algorithm for deciding Markov equivalence of directed cyclic graphical models. *Uncertainty in Artificial Intelligence, Proceedings, 12th Conference*, Morgan Kaufman, CA.

SPIRTES, P. (1995). Directed cyclic graphical representations of feedback models. In *Uncertainty in Artificial Intelligence, Proceedings, 11th Conference,* Morgan Kaufman, CA.

SPIRTES, P., C. GLYMOUR, R. SCHEINES (1993). Causation, Prediction and Search. Lecture Notes in Statistics 81, Springer-Verlag.

SPIRTES, P., C. MEEK, T. RICHARDSON (1995). Causal inference in the presence of latent variables and selection bias. *Uncertainty in Artifcial Intelligence, Proceedings, 11th Conference.* Morgan Kaufman, San Mateo, CA.

STROTZ, R.H. and WOLD, H.O.A. (1960) Recursive versus Non-recursive Systems: An Attempt at Synthesis. *Econometrica* **28** pp.417-427. (Also in *Causal Models in the Social Sciences,* H.M. Blalock Jr. ed., Chicago: Aldine Atherton, 1971)

STUDENÝ , M., R. BOUCKAERT (1996). On chain graph models for description of conditional independence structures. Manuscript, submitted for publication.

VERMA, T., J. PEARL (1990). Equivalence and synthesis of causal models. Proceedings of the Sixth Conference on Uncertainty in AI. Association for Uncertainty in AI, Inc., Mountain View, CA.

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