A Transformational Characterization of Markov Equivalence between DAGs with Latent Variables

JiJi Zhang & Peter Spirtes

March 24, 2005

Technical Report No. CMU-PHIL-169

Philosophy

Methodology

Logic

Carnegie Mellon

Pittsburgh, Pennsylvania 15213

A Transformational Characterization of Markov Equivalence between DAGs with Latent Variables

Jiji Zhang & Peter Spirtes

March 24, 2005

Abstract

Different directed acyclic graphs (DAGs) may be Markov equivalent in the sense that they entail the same conditional independence relations among the observed variables. Markov equivalence between DAGs (with no latent variables) has been characterized in various ways, each of which has been found useful for certain purposes. In particular, Chickering's transformational characterization is useful in deriving properties shared by Markov equivalent DAGs, and, with certain generalization, is needed to prove the asymptotic correctness of a search procedure over Markov equivalence classes, known as the GES algorithm.

Maximal ancestral graphs (MAGs) are a generalization of DAGs that can represent the observable conditional independence relations (as well as some causal features) of DAG models with latent variables. Thus Markov equivalence between DAGs with latent variable is reduced to Markov equivalence between the corresponding MAGs. However, no characterization of Markov equivalent MAGs is yet available that is analogous to Chickering's transformational characterization. The main contribution of the current paper is to establish such a characterization for directed MAGs, which we expect will have similar uses as it does for DAGs.

1 INTRODUCTION

Markov equivalence between directed acyclic graphs (DAGs) has been characterized in several ways (e.g., Verma and Pearl 1990, Chickering 1995, Andersson et al. 1997). All of them have been found useful for various purposes. In particular, the transformational characterization provided by Chickering (1995) — that two DAGs are Markov equivalent if and only if one can be transformed to the other by a sequence of single edge reversals that preserve Markov equivalence — is useful in deriving properties shared by Markov equivalent DAGs. Moreover, when generalized, the transformational characterization implies the asymptotic correctness of the GES algorithm, an efficient search procedure over Markov equivalence classes of DAGs (Meek 1996, Chickering 2002).

In many situations, however, we need also to consider DAGs with latent variables. Indeed there are cases where no DAGs can perfectly explain the observed conditional independence relations unless latent variables are introduced. Such latent variable models, fortunately, can be represented by ancestral graphical models (Richardson and Spirtes 2002), in that for any DAG with latent variables, there is a (maximal) ancestral graph that captures the exact observable conditional independence relations as well as some of the causal relations entailed by that DAG. Since ancestral graphs do not explicitly include latent variables, they provide, among other virtues, a finite search space of latent variable models (Spirtes et al. 1997).

Markov equivalence for ancestral graphs has been characterized in ways analogous to the one given by Verma and Pearl (1990) for DAGs (Spirtes and Richardson 1996, Ali et al. 2004). However, no characterization is yet available that is analogous to Chickering's transformational characterization. In this paper we establish one for directed ancestral graphs. Specifically we show that two directed maximal ancestral graphs are Markov equivalent if and only if one can be transformed to the other by a sequence of single mark changes — adding or dropping an arrowhead — that preserve Markov equivalence. This characterization we expect will have similar uses as Chickering's does for DAGs. In particular, it is a step towards justifying the application of the GES algorithm to MAGs, and hence to latent variable DAG models.

The paper is organized as follows. The remainder of this section introduces the relevant definitions and notations. We then present the main result in section 2, drawing on some facts proved in Zhang and Spirtes (2005) and Ali et al. (2005). We conclude the paper in section 3 with a discussion of the potential application, limitation and generalization of our result.

1.1 DIRECTED ANCESTRAL GRAPHS

In full generality, an ancestral graph can contain three kinds of edges: directed edge ($\rightarrow$), bi-directed edge ($\leftrightarrow$) and undirected edge ($-$). In this paper, however, we will confine ourselves to directed ancestral graphs — which do not contain undirected edges — until section 3, where we explain why our result does not hold for general ancestral graphs. The class of directed ancestral graphs, due to its inclusion of bi-directed edges, is suitable for representing observed conditional independence structures in the presence of latent confounders (see Figure 1).

By a **directed mixed graph** we denote an arbitrary graph that can have two kinds of edges: directed and bi-directed. The two ends of an edge we call **marks** or **orientations**. So the two marks of a bi-directed edge are both **arrowheads** ($>$), while a directed edge has one arrowhead and one **tail** ($-$) as its marks. Sometimes we say an edge is **into** (or **out of**) a vertex if the mark of the edge at the vertex is an arrowhead (or a tail). The meaning of the standard graph theoretical concepts, such as **parent/child**, **(directed) path**, **ancestor/descendant**, etc., remains the same in mixed graphs. Furthermore, if there is a bi-directed edge between two vertices A and B ($A \leftrightarrow B$), then A is called a **spouse** of B and B a spouse of A.

Definition 1 (ancestral). *A directed mixed graph is* **ancestral** *if*

(a1) there is no directed cycle; and

(a2) for any two vertices A and B, if A is a spouse of B (i.e., $A \leftrightarrow B$), then A is not an ancestor of B.

Clearly DAGs are a special case of directed ancestral graphs (with no bi-directed edges). Condition (a1) is just the familiar one for DAGs. Condition (a2), together with (a1), defines a nice feature of arrowheads — that is, an arrowhead implies non-ancestorship. This motivates the term "ancestral" and induces a natural causal interpretation of ancestral graphs (see, e.g., Richardson and Spirtes 2003).

Mixed graphs encode conditional independence relations by essentially the same graphical criterion as the well-known *d-separation* for DAGs, except that in mixed graphs colliders can arise in more edge configurations than they do in DAGs. Given a path u in a mixed graph, a non-endpoint vertex V on u is called a **collider** if the two edges incident to V on u are both into V, otherwise V is called a **non-collider**.

Definition 2 (m-separation). *In a mixed graph, a path u between vertices A and B is* **active** *(***m-connecting***) relative to a set of vertices* $\mathbf{Z}$ $(A, B \notin \mathbf{Z})$ *if*

i. every non-collider on u is not a member of $\mathbf{Z}$;

ii. every collider on u is an ancestor of some member of $\mathbf{Z}$.

A and B are said to be **m-separated** *by $\mathbf{Z}$ if there is no active path between A and B relative to $\mathbf{Z}$.*

The following property is true of DAGs: if two vertices are not adjacent, then there is a set of some other vertices that m-separates (d-separates) the two. This, however, is not true of directed ancestral graphs in general, which motivates the following definition.

Definition 3 (maximality). *A directed ancestral graph is said to be* **maximal** *if for any two non-adjacent vertices, there is a set of vertices that m-separates them.*

It is shown in Richardson and Spirtes (2002) that every non-maximal ancestral graph has a unique supergraph that is ancestral and maximal, and it is easy to construct the maximal supergraph given a non-maximal ancestral graph. This justifies considering only those ancestral graphs that are maximal (MAGs). From now on, we focus on directed maximal ancestral graphs, which we will refer to as DMAGs. A notion closely related to maximality is that of inducing path:

Definition 4 (inducing path). *In an ancestral graph, a path u between A and B is called an* **inducing path** *if every non-endpoint vertex on u is a collider and is an ancestor of either A or B.*

Richardson and Spirtes (2002) proved that the presence of an inducing path is necessary and sufficient for two vertices not to be m-separated by any set. So, to show that a graph is maximal, it suffices to demonstrate that there is no inducing path between any two non-adjacent vertices in the graph.

Given any DAG with (or without) latent variables, the conditional independence relations as well as the causal relations among the observed variables can be represented by a DMAG that includes only the observed variables. The DMAG is constructed as follows: for every pair of observed variables, O_i and O_j, put an edge between them if and only if they are not d-separated by any set of other observed variables in the given DAG, and mark an arrowhead at O_i (O_j) on the edge if it is *not* an ancestor of O_j (O_i) in the given DAG.

For example, Figure 1(a) is a DAG with latent variables $\{L1, L2, L3\}$. Figure 1(b) depicts the DMAG (G1) resulting from the above construction. The m-separation relations in G1 correspond exactly to the d-separation relations over $\{X1, X2, X3, X4, X5\}$ in Figure 1(a). By contrast, no DAG without extra latent variables has the exact same d-separation relations. Furthermore, the orientations in G1 accurately represent the ancestor relationships — which, upon natural interpretations, are causal relationships — among the observed variables in 1(a). (This, however, is not the case with G2.)

1.2 MARKOV EQUIVALENCE

A DMAG represents the set of joint distributions that satisfy its *global Markov property*, i.e., the set of distributions of which the conditional independence relations entailed by m-separations in the DMAG hold. Hence, if two DMAGs share the same m-separation structures, then they represent the same set of distributions.

Definition 5 (Markov equivalence). *Two DMAGs $\mathcal{G}_1, \mathcal{G}_2$ (with the same set of vertices) are* **Markov equivalent** *if for any three disjoint sets of vertices* $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, $\mathbf{X}$ *and* $\mathbf{Y}$ *are m-separated by* $\mathbf{Z}$ *in* $\mathcal{G}_1$ *if and only if* $\mathbf{X}$ *and* $\mathbf{Y}$ *are m-separated by* $\mathbf{Z}$ *in* $\mathcal{G}_2$.

Figure 1(c), for example, is a DMAG Markov equivalent to 1(b). It is well known that two DAGs are Markov equivalent if and only if they have the same adjacencies

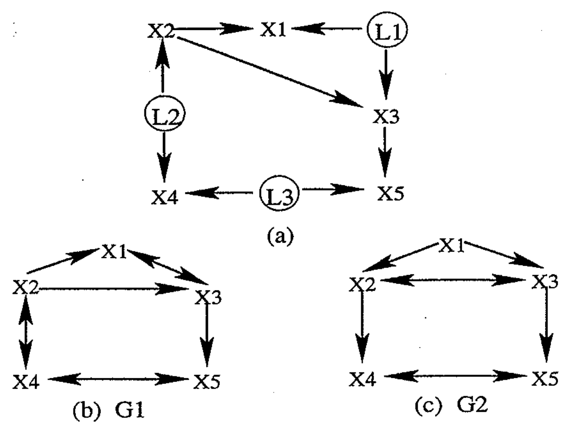

Figure 1: (a): A DAG with latent variables; (b): A DMAG that captures both the conditional independence and causal relations among the observed variables represented by (a); (c): A DMAG that entails the right conditional independence relations but not the right causal relations in (a).

and the same unshielded colliders (Verma and Pearl 1990). (A triple $\langle A, B, C\rangle$ is said to be **unshielded** if A, B are adjacent, B, C are adjacent but A, C are not adjacent.) The conditions are still necessary for Markov equivalence between DMAGs, but are not sufficient. For two DMAGs to be equivalent, some shielded colliders have to be present in both or neither of the graphs. The next definition is related to this.

Definition 6 (discriminating path). *In a DMAG, a path between X and Y, $u = \langle X, \cdots, W, V, Y\rangle$, is a* **discriminating path** *for V if*

i. u includes at least three edges (i.e., at least four vertices as specified);

ii. V is adjacent to an endpoint Y on u; and

iii. X is not adjacent to Y, and every vertex between X and V is a collider on u and is a parent of Y.

Discriminating paths behave similarly to unshielded triples in that if $u = \langle X, \cdots, W, V, Y\rangle$ is discriminating for V, then $\langle W, V, Y\rangle$ is a (shielded) collider if and only if every set that m-separates X and Y excludes V; it is a non-collider if and only if every set that m-separates X and Y contains V. The following proposition is proved in Spirtes and Richardson (1996)[1].

Proposition 1. *Two DMAGs over the same set of vertices are Markov equivalent if and only if*

(e1) They have the same adjacencies;

(e2) They have the same unshielded colliders;

(e3) If a path u is a discriminating path for a vertex B in both graphs, then B is a collider on the path in one graph if and only if it is a collider on the path in the other.

2 A Transformational Property of DMAGs

We present the main result of the paper in this section, namely Markov equivalent DMAGs can be transformed to each other by a sequence of single mark changes that preserve Markov equivalence. We first describe in section 2.1 two corollaries from Zhang and Spirtes (2005) and Ali et al. (2005) which our arguments will rely upon. Section 2.2 establishes sufficient and necessary conditions for a single mark change to preserve equivalence. The theorems are then presented in section 2.3.

2.1 Loyal Equivalent Graph

Given a MAG $\mathcal{G}$, a mark (or edge) in $\mathcal{G}$ is **invariant** if it is present in all MAGs Markov equivalent to $\mathcal{G}$. Invariant marks are particularly important for causal inference because data alone usually can't distinguish between members of a Markov equivalence class. An algorithm for detecting all invariant arrowheads in a MAG is given by Ali et al. (2005), and one for further detecting all invariant tails is presented in Zhang and Spirtes (2005). In what follows we list some facts proved in Zhang and Spirtes (2005) which will lead to two propositions needed for our current purpose.

[1]The conditions are also valid for maximal ancestral graphs that contain undirected edges.

Zhang and Spirtes (2005) give a set of orientation rules by which one can construct, given an arbitrary MAG, a complete representation of the invariant marks of the MAG. The representation they use is a graph that can contain three kinds of marks: arrowhead ($>$), tail ($-$) and circle ($\circ$), and hence in general six kinds of edges: $\circ\!\!-\!\!\circ$, $\circ\!\!\rightarrow$, $\rightarrow$, $\leftrightarrow$, $-$ and $\circ\!\!-$; although in the case of DMAGs, only the first four are needed. Intuitively circles purport to indicate that the corresponding mark could be either an arrowhead or a tail in the Markov equivalence class of $\mathcal{G}$, whereas a non-circle mark represents a mark shared by all members of the equivalence class.

Given a MAG $\mathcal{G}$, the construction starts with $\mathcal{P}_0$, a graph having the same adjacencies as $\mathcal{G}$ does, but every edge therein is of the form $\circ\!\!-\!\!\circ$. Then a set of orientation rules is applied to change some of the circles into arrowheads or tails. In particular, the following five orientation rules are sufficient to identify all invariant arrowheads in $\mathcal{G}$ ($*$ is used as a meta-symbol that represents whatever mark that may be present at an end of an edge, i.e., a circle, an arrowhead, or a tail):

$\mathcal{R}0$ For every triple $\alpha *\!\!-\!\!* \beta \circ\!\!-\!\!* \gamma$ s.t. α, γ are not adjacent, if it is an unshielded collider in $\mathcal{G}$, then orient the triple as $\alpha *\!\!\rightarrow \beta \leftarrow\!\!* \gamma$.

$\mathcal{R}1$ If $\alpha *\!\!\rightarrow \beta \circ\!\!-\!\!* \gamma$, and α and γ are not adjacent, then orient the triple as $\alpha *\!\!\rightarrow \beta \rightarrow \gamma$.

$\mathcal{R}2$ If $\alpha \rightarrow \beta *\!\!\rightarrow \gamma$ or $\alpha *\!\!\rightarrow \beta \rightarrow \gamma$, and $\alpha *\!\!-\!\!\circ \gamma$, then orient $\alpha *\!\!-\!\!\circ \gamma$ as $\alpha *\!\!\rightarrow \gamma$.

$\mathcal{R}3$ If $\alpha *\!\!\rightarrow \beta \leftarrow\!\!* \gamma$, $\alpha *\!\!-\!\!\circ \theta \circ\!\!-\!\!* \gamma$, α and γ are not adjacent, and $\theta *\!\!-\!\!\circ \beta$, then orient $\theta *\!\!-\!\!\circ \beta$ as $\theta *\!\!\rightarrow \beta$.

$\mathcal{R}4$ If $u = \langle \theta, ..., \alpha, \beta, \gamma \rangle$ is a discriminating path between θ and γ for β, and $\beta \circ\!\!-\!\!* \gamma$; then if $\beta \rightarrow \gamma$ appears in $\mathcal{G}$, orient $\beta \circ\!\!-\!\!* \gamma$ as $\beta \rightarrow \gamma$; otherwise orient the triple $\langle \alpha, \beta, \gamma \rangle$ as $\alpha \leftrightarrow \beta \leftrightarrow \gamma$.

It is to be understood that $\mathcal{R}0$ is applied first to $\mathcal{P}_0$, and then $\mathcal{R}1 - \mathcal{R}4$ are applied repeatedly and exhaustively, i.e., until no more circles can be oriented. Let $\mathcal{P}$ be the resulting graph, which will be called the **partial ancestral graph (PAG)** of $\mathcal{G}$. We state three facts from Zhang and Spirtes (2005). First of all, $\mathcal{R}0 - \mathcal{R}4$ are sound and complete with respect to invariant arrowheads in the following sense (Lemma 2 and Theorem 2 in Zhang and Spirtes (2005)):

Fact 1. *Every non-circle mark in $\mathcal{P}$ is invariant. That is, all arrowheads and tails in $\mathcal{P}$ are shared by all DMAGs that are Markov equivalent to $\mathcal{G}$. Furthermore, every invariant arrowhead in $\mathcal{G}$ appears in $\mathcal{P}$.*

We also need the following fact, which is proved in Lemma 6 in Zhang and Spirtes (2005) (see also Corollary 4.1 in Ali et al. 2005).

Fact 2. *For any A o—oB and vertex C in $\mathcal{P}$, $C \leftrightarrow A$ appears in $\mathcal{P}$ if and only if $C \leftrightarrow B$ appears in $\mathcal{P}$.*

Lastly, let the **circle component** of $\mathcal{P}$ — denoted by $\mathcal{P}^c$ — be the induced subgraph that consists of all o—o edges in $\mathcal{P}$. The following fact, as a special case of Lemma 16 in Zhang and Spirtes (2005), is key to the first proposition we want to establish in the current section.

Fact 3. *Let $\mathcal{H}$ be the graph that results from changing all $\circ\!\!\rightarrow$ edges in $\mathcal{P}$ into $\rightarrow$, and orienting $\mathcal{P}^c$ into a directed acyclic graph with no unshielded colliders. Then $\mathcal{H}$ is a DMAG and is Markov equivalent to $\mathcal{G}$.*

The main aim of the current section is to show that for any DMAG, there is a Markov equivalent DMAG such that every bi-directed edge therein is invariant, and every directed edge in the given DMAG is retained. This obviously will rely on Fact 3 just stated. The only extra piece of fact needed is the following:

Fact 4. *Let $\mathcal{G}^*$ be a directed ancestral graph over a set of vertices $\mathbf{V}$. If $\mathcal{G}^*$ has no unshielded colliders, then there is a directed acyclic graph $\mathcal{D}$ over $\mathbf{V}$ such that*

(1) $\mathcal{D}$ has the same adjacencies as $\mathcal{G}^*$ does;

(2) $\mathcal{D}$ has no unshielded colliders;

(3) every directed edge in $\mathcal{G}^*$ is also in $\mathcal{D}$.

Proof. Here is a formal construction of the DAG. The ancestor relationship in $\mathcal{G}^*$ naturally induces a partial order over $\mathbf{V}$. Extend this partial order to a total order. Put an edge $A \to B$ in $\mathcal{D}$ if and only if A and B are adjacent in $\mathcal{G}^*$, and A precedes B in the order. The resulting $\mathcal{D}$ is obviously a DAG, and has the same adjacencies as $\mathcal{G}^*$ does. Furthermore, all the directed edges in $\mathcal{G}^*$ are retained in $\mathcal{D}$, because the ordering that yields $\mathcal{D}$ respects the ordering in $\mathcal{G}^*$. So the only difference between $\mathcal{G}^*$ and $\mathcal{D}$ is that bi-directed edges in the former are changed into directed edges in the latter. It follows that all arrowheads, and hence all colliders, in $\mathcal{D}$ are also in $\mathcal{G}^*$. Since there is no unshielded collider in $\mathcal{G}^*$, there is no unshielded collider in $\mathcal{D}$. $\square$

Proposition 2. *Given any DMAG $\mathcal{G}$, there exists a DMAG $\mathcal{H}$ such that*

(1) $\mathcal{H}$ is Markov equivalent to $\mathcal{G}$;

(2) every bi-directed edge in $\mathcal{H}$ is invariant;

(3) every directed edge in $\mathcal{G}$ is also in $\mathcal{H}$.

Proof. Let $\mathcal{P}$ be the PAG of $\mathcal{G}$, and $\mathcal{G}^*$ be the corresponding subgraph of $\mathcal{G}$ over the vertices of $\mathcal{P}^c$, the circle component of $\mathcal{P}$. It is clear that $\mathcal{G}^*$ does not contain any unshielded collider, otherwise it would have been introduced into $\mathcal{P}$ by $\mathcal{R}0$. Hence, by Fact 4, there is a DAG $\mathcal{D}$ over the vertices of $\mathcal{P}^c$ with the same adjacencies and no unshielded colliders such that all directed edges in $\mathcal{G}^*$ are retained in $\mathcal{D}$. Let $\mathcal{H}$ be the graph resulting from changing all $\circ\!\!\rightarrow$ edges in $\mathcal{P}$ into $\rightarrow$, and orienting $\mathcal{P}^c$ into $\mathcal{D}$. By Fact 3, $\mathcal{H}$ is a DMAG and is Markov equivalent to $\mathcal{G}$. Also, since every non-circle mark in $\mathcal{P}$ is invariant (Fact 1), and the extra orientations do not introduce new bi-directed edges, every bi-directed edge in $\mathcal{H}$ is invariant. Finally, any directed edge in $\mathcal{G}$ is either in $\mathcal{G}^*$, or corresponds to a directed edge or a $\circ\!\!\rightarrow$ edge in $\mathcal{P}$. In either case, the directed edge is also in $\mathcal{H}$ by our construction. $\square$

We will call $\mathcal{H}$ in Proposition 2 a **Loyal Equivalent Graph (LEG)** of $\mathcal{G}$. In general a DMAG could have multiple LEGs. A distinctive feature of the LEGs is that they have the fewest bi-directed edges among the Markov equivalent DMAGs[2]. Drton

[2]Proposition 2 is actually a special case of Corollary 18 in Zhang and Spirtes (2005), which, for general MAGs, also asserts that the LEGs have the fewest undirected edges as well.

and Richardson (2004) explored the statistical significance of this fact for bi-directed graphs (graphs that contain only bi-directed edges). Roughly speaking, if the LEGs of a bi-directed graph are DAGs, then fitting is easy; otherwise fitting is not easy (in a specific technical sense).

Another feature which is particularly relevant to our argument is that between a DMAG and any of its LEGs, only one kind of differences is possible, namely, some bi-directed edges in the DMAG are oriented as directed edges in its LEG. For a simple illustration, compare the graphs in Figure 2, where H1 is a LEG of G1, and H2 is a LEG of G2.

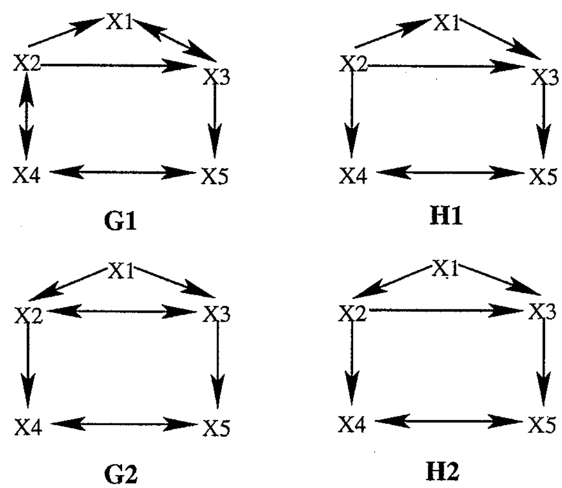

Figure 2: A LEG of G1 (H1) and a LEG of G2 (H2)

A directed edge in a DMAG is called **reversible** if there is another Markov equivalent DMAG in which the direction of the edge is reversed. It would be convenient to have the following proposition in order to prove Theorem 2 below.

Proposition 3. *Let $A \rightarrow B$ be any reversible edge in a DMAG $\mathcal{G}$. For any vertex C (distinct from A and B), there is an invariant bi-directed edge between C and A if and only if there is an invariant bi-directed edge between C and B.*

Proof. Since $A \rightarrow B$ is reversible (which means neither of the two marks of the edge is invariant), in $\mathcal{P}$ — the PAG of $\mathcal{G}$ — the edge between A and B would be A o—oB

by Fact 1. For any C, if there is an invariant bi-directed edge between C and A, by Fact 1, $C \leftrightarrow A$ would appear in $\mathcal{P}$. By Fact 2, $C \leftrightarrow B$ is also in $\mathcal{P}$, and hence is an invariant bi-directed edge in $\mathcal{G}$. Conversely, if there is an invariant bi-directed edge between C and B in $\mathcal{G}$, the same argument shows that there would also be an invariant bi-directed edge between C and A. □

In particular, if $\mathcal{H}$ is a LEG of a DMAG, then $A \rightarrow B$ being reversible implies that A and B have the same set of spouses, as every bi-directed edge in $\mathcal{H}$ is invariant.

2.2 Equivalence-Preserving Mark Change

Eventually we will show that two Markov equivalent DMAGs can be connected by a sequence of equivalence-preserving mark changes. It is thus desirable to give some simple graphical conditions under which a single mark change would preserve equivalence. The next lemma presents necessary and sufficient conditions under which adding an arrowhead to a directed edge (i.e., changing the directed edge to a bi-directed one) preserves Markov equivalence. By symmetry, they are also the conditions for dropping an arrowhead from a bi-directed edge while preserving Markov equivalence.

Lemma 1. *Let $\mathcal{G}$ be an arbitrary DMAG, and $A \rightarrow B$ an arbitrary directed edge in $\mathcal{G}$. Let $\mathcal{G}'$ be the graph identical to $\mathcal{G}$ except that the edge between A and B is $A \leftrightarrow B$. (In other words, $\mathcal{G}'$ is the result of simply changing the mark at A on $A \rightarrow B$ from an tail into an arrowhead.) $\mathcal{G}'$ is a DMAG and Markov equivalent to $\mathcal{G}$ if and only if*

(t1) there is no directed path from A to B other than $A \rightarrow B$;

(t2) For any $C \rightarrow A$ in $\mathcal{G}$, $C \rightarrow B$ is also in $\mathcal{G}$; and for any $D \leftrightarrow A$ in $\mathcal{G}$, either $D \rightarrow B$ or $D \leftrightarrow B$ is in $\mathcal{G}$;

(t3) There is no discriminating path for A on which B is the endpoint adjacent to A.

Proof. We first show that each of the conditions is necessary (**only if**). Obviously if (t1) fails, $\mathcal{G}'$ will not be ancestral (see Definition 1). The failure of (t2) could be due to one of the following two cases:

Case 1: there is a vertex C which is a parent of A but not a parent of B. If B and C are not adjacent, then there is an unshielded collider in $\mathcal{G}'$ but not in $\mathcal{G}$, and hence the two graphs are not Markov equivalent, according to Proposition 1. If B and C are adjacent, then $\mathcal{G}$ can't be ancestral (unless we have $C \to B$).

Case 2: there is a vertex C which is a spouse of A but not a parent or spouse of B. Again, if B and C are not adjacent, the two graphs can't be Markov equivalent because there is an unshielded collider in $\mathcal{G}$ but not in $\mathcal{G}'$. If B and C are adjacent, the edge between them must be $B \to C$ in $\mathcal{G}$ by the supposition. But then there is an almost directed cycle in $\mathcal{G}$, and hence $\mathcal{G}$ is not ancestral.

If (t3) fails, that is, there is a discriminating path $u = \langle U, \cdots, V, A, B \rangle$ for A. If the edge between V and A is into A, then $\mathcal{G}$ and $\mathcal{G}'$ are not Markov equivalent, because (e3) in Proposition 1 is violated. If, on the other hand, the edge between V and A is not into A, then it must be $A \to V$. By the definition of discriminating path (Definition 6), V is a parent of B. So we have $A \to V \to B \leftrightarrow A$ in $\mathcal{G}'$, an almost directed cycle, which means $\mathcal{G}'$ is not ancestral.

Next, we demonstrate the sufficiency of the conditions (**if**). Suppose (t1)-(t3) are met. We first verify that $\mathcal{G}'$ is a DMAG, i.e., it is both ancestral and maximal. Suppose for contradiction that $\mathcal{G}'$ is not ancestral. Since $\mathcal{G}$ is ancestral, and $\mathcal{G}'$ differs from $\mathcal{G}$ only regarding the edge between A and B, in $\mathcal{G}'$ the violation of the definition of ancestral graphs (Definition 1) must involve the edge between A and B. So it can't be a violation of (a1), because a directed cycle would not involve $A \leftrightarrow B$. If it is a violation of (a2), i.e., there is an almost directed cycle in $\mathcal{G}'$. That cycle includes $A \leftrightarrow B$, which means either A is an ancestor of B or B is an ancestor of A in $\mathcal{G}'$. The former case contradicts (t1), and the latter case yields a directed cycle in $\mathcal{G}$. So there can't be any violation of (a2) in $\mathcal{G}'$. Hence $\mathcal{G}'$ is ancestral.

To show that $\mathcal{G}'$ is maximal, Suppose for the sake of contradiction that there is an inducing path u in $\mathcal{G}'$ between two non-adjacent vertices, D and E. Then u must include $A \leftrightarrow B$, otherwise u would also be an inducing path in $\mathcal{G}$. Furthermore, A is not an endpoint of u, otherwise u is still an inducing path in $\mathcal{G}$ (in fact, there would be an almost directed path in $\mathcal{G}$ in that case). Suppose, without loss of generality, that D is the endpoint closer to A on u than it is to B. We show that some vertex on $u(D, A)$ other than A is B's spouse. Suppose not — that is, suppose no vertex between D and A on u is B's spouse, we argue by induction that every vertex on $u(A, D)$, and in particular D, is a parent of B. By (t2), the vertex adjacent to A on $u(D, A)$ is either a parent or a spouse of B, but it is not a spouse by supposition, so it is a parent. In the inductive step, suppose the first n vertices next to A on $u(D, A)$ are B's parents, then the $n + 1$'st vertex V must be adjacent to B, otherwise the sub-path of u between V and B forms a discriminating path for A, which violates (t3). By supposition, V is not a spouse of B, i.e., it is not the case that $V \leftrightarrow B$. It can't be $V \leftarrow B$ either, because in that case there would be an almost directed cycle in $\mathcal{G}'$ (as the vertex before V, by the inductive hypothesis, is a parent of B), which we have shown to be impossible. So V must be a parent of B. Thus we have shown that every vertex on $u(A, D)$, and in particular D, is a parent of B. Then B must be an ancestor of E, because by the definition of inducing path (Definition 4), B is an ancestor of either D or E. So D is an ancestor of E, which means the vertex adjacent to E on u must be an ancestor of E. But the edge between that vertex and E must be into that vertex (as it is a collider on u by definition), so there is a directed or almost directed cycle in $\mathcal{G}'$, which we have shown to be absent. Hence a contradiction. So some vertex on $u(D, A)$ other than A is a spouse of B. Let C be such a vertex on $u(D, A)$. Replacing $u(C, B)$ on u with $C \leftrightarrow B$ yields an inducing path between D and E in $\mathcal{G}$, which contradicts the fact that $\mathcal{G}$ is maximal.

Having shown that $\mathcal{G}'$ is a DMAG, we now verify that $\mathcal{G}$ and $\mathcal{G}'$ satisfy the conditions for Markov equivalence in Proposition 1. Obviously they have the same adjacencies, and share the same colliders except possibly A. But A will not be a collider in an unshielded triple, for condition (t2) requires that any vertex that is incident to an edge into A is also adjacent to B. So the only worry is that a triple $\langle C, A, B \rangle$ might be discriminated by a path, but (t3) guarantees that there is no such path. Therefore, $\mathcal{G}'$ is Markov equivalent to $\mathcal{G}$. $\square$

We say a mark change is *legitimate* when the conditions in Lemma 1 are satisfied. Recall that for DAGs the basic unit of equivalence-preserving transformation is (covered) edge reversal (Chickering 1995). In the current paper we treat an edge reversal as simply a special case of two consecutive mark changes. That is, a reversal of $A \rightarrow B$ is simply to first add an arrowhead at A (to form $A \leftrightarrow B$), and then to drop the arrowhead at B (to form $A \leftarrow B$). An edge reversal is said to be legitimate if both of the two consecutive mark changes are legitimate. Given Lemma 1, it is straightforward to check the validity of the following condition for legitimate edge reversal. (We use $\mathbf{Pa}_{\mathcal{G}}/\mathbf{Sp}_{\mathcal{G}}$ to denote the set of parents/spouses of a vertex in $\mathcal{G}$.)

Lemma 2. *Let $\mathcal{G}$ be an arbitrary DMAG, and $A \rightarrow B$ an arbitrary directed edge in $\mathcal{G}$. The reversal of $A \rightarrow B$ is legitimate if and only if $\mathbf{Pa}_{\mathcal{G}}(B) = \mathbf{Pa}_{\mathcal{G}}(A) \cup \{A\}$ and $\mathbf{Sp}_{\mathcal{G}}(B) = \mathbf{Sp}_{\mathcal{G}}(A)$.*

When there is no bi-directed edge in $\mathcal{G}$, that is, when $\mathcal{G}$ is a DAG, the condition in Lemma 2 is reduced to the familiar definition for *covered edge*, i.e., $\mathbf{Pa}_{\mathcal{G}}(B) = \mathbf{Pa}_{\mathcal{G}}(A) \cup \{A\}$ (Chickering 1995). The condition given by Drton and Richardson (2004) for a bi-directed edge in a bi-directed graph to be orientable as a directed edge in either direction ($\mathbf{Sp}_{\mathcal{G}}(B) = \mathbf{Sp}_{\mathcal{G}}(A)$) can be viewed as another special case of the above lemma.

2.3 Transformation between Equivalent DMAGs

We first state two intermediate theorems crucial for the main result we are heading for. The first one says if the differences between two Markov equivalent DMAGs $\mathcal{G}$ and $\mathcal{G}'$ are all of the following sort: a directed edge is in $\mathcal{G}$ while the corresponding edge is bi-directed in $\mathcal{G}'$, then there is a sequence of legitimate mark changes that transforms one to the other. The second one says that if every bi-directed edge in $\mathcal{G}$ and every bi-directed edge in $\mathcal{G}'$ are invariant, then there is a sequence of legitimate mark changes (edge reversals) that transforms one to the other. The proofs follow the strategy of Chickering's proof for DAGs.

Theorem 1. *Let $\mathcal{G}$ and $\mathcal{G}'$ be two Markov equivalent DMAGs. If every bi-directed edge in $\mathcal{G}$ is also in $\mathcal{G}'$, and every directed edge in $\mathcal{G}'$ is also in $\mathcal{G}$, then there is a sequence of legitimate mark changes that transforms one to the other.*

Proof. We prove that there is a sequence of transformation from $\mathcal{G}$ to $\mathcal{G}'$, the reverse of which will be a transformation from $\mathcal{G}'$ to $\mathcal{G}$. Specifically we show that as long as

$\mathcal{G}$ and $\mathcal{G}'$ are different, there is always a legitimate mark change that can eliminate a difference between them. The theorem then follows from a simple induction on the number of differences.

The antecedent of the theorem implies that the differences between $\mathcal{G}$ and $\mathcal{G}'$ are all of the same sort: a directed edge ($\rightarrow$) is in $\mathcal{G}$ while the corresponding edge in $\mathcal{G}'$ is bi-directed ($\leftrightarrow$). Let

$$\mathbf{Diff} = \{y|\text{ there is a } x \text{ such that } x \rightarrow y \text{ is in } \mathcal{G} \text{ and } x \leftrightarrow y \text{ is in } \mathcal{G}'\}$$

It is clear that $\mathcal{G}$ and $\mathcal{G}'$ are identical if and only if $\mathbf{Diff} = \emptyset$. We claim that if $\mathbf{Diff}$ is not empty, there is a legitimate mark change that eliminates a difference. Choose $B \in \mathbf{Diff}$ such that no proper ancestor of B in $\mathcal{G}$ is in $\mathbf{Diff}$. Let

$$\mathbf{Diff}_B = \{x|x \rightarrow B \text{ is in } \mathcal{G} \text{ and } x \leftrightarrow B \text{ is in } \mathcal{G}'\}$$

Since $B \in \mathbf{Diff}$, $\mathbf{Diff}_B$ is not empty. Choose $A \in \mathbf{Diff}_B$ such that no proper descendant of A in $\mathcal{G}$ is in $\mathbf{Diff}_B$. The claim is that changing $A \rightarrow B$ to $A \leftrightarrow B$ in $\mathcal{G}$ is a legitimate mark change.

To see this is so, let us verify the conditions stated in Lemma 1. First, suppose condition (t1) is violated, that is, suppose there is another directed path $d = \langle A, \cdots, C, B\rangle$ from A to B besides $A \rightarrow B$. d is not present in $\mathcal{G}'$, otherwise $\mathcal{G}'$ is not a MAG due to the presence of $A \leftrightarrow B$. So some edge on d in $\mathcal{G}'$ must be bi-directed. If the edge is $C \leftrightarrow B$, then C belongs to $\mathbf{Diff}_B$, but is a proper descendant of A in $\mathcal{G}$, which contradicts our choice of A. If the edge is between another pair of vertices, say $D \leftrightarrow E$ (s.t. $D \rightarrow E$ is in $\mathcal{G}$), then E is in $\mathbf{Diff}$, but is a proper ancestor of B, which contradicts our choice of B. So can't be any directed path from A to B in $\mathcal{G}$ other than $A \rightarrow B$. Condition (t1) stands.

Next we check condition (t2). For the first part, let C be any parent of A in $\mathcal{G}$. C must also be a parent of A in $\mathcal{G}'$, otherwise A is in $\mathbf{Diff}$, but is a proper ancestor of B in $\mathcal{G}$, which contradicts our choice of B. It follows that C and B are adjacent, for otherwise $\langle C, A, B\rangle$ is an unshielded collider in $\mathcal{G}'$ but not in $\mathcal{G}$, contrary to the assumption that they are Markov equivalent. Then C must be a parent of B in $\mathcal{G}$, otherwise $\mathcal{G}$ is not ancestral.

For the second part, let D be any spouse of A (i.e., $D \leftrightarrow A$) in $\mathcal{G}$. D is also a spouse of A in $\mathcal{G}'$ by our assumption. It follows that D and B are adjacent, for otherwise $\langle D, A, B \rangle$ is an unshielded collider in $\mathcal{G}'$ but not in $\mathcal{G}$. But D cannot be a child of B in $\mathcal{G}$, for otherwise $\mathcal{G}$ is not ancestral. Hence D is either a parent or a spouse of B.

Finally, suppose condition (t3) is violated, that is, suppose there is a discriminating path $u = \langle U, \cdots, V, A, B \rangle$ for A. By the definition of discriminating path, V is a parent of B. It follows that the edge between A and V is not $A \rightarrow V$, for otherwise $A \rightarrow V \rightarrow B$ would be a directed path from A to B, which has been shown to be absent. Hence the edge between V and A must be bi-directed, $V \leftrightarrow A$. Further note that by our assumption about the difference between $\mathcal{G}$ and $\mathcal{G}'$, every arrowhead in $\mathcal{G}$ is also in $\mathcal{G}'$, which implies that every collider in $\mathcal{G}$ is also in $\mathcal{G}'$. In particular, every vertex between U and A on u is also a collider on u in $\mathcal{G}'$.

Now we prove by induction that every vertex between U and A on u, including A, is a parent of B in $\mathcal{G}'$, contradicting the fact that $A \leftrightarrow B$ is in $\mathcal{G}'$. Let W be the vertex next to U on u. Since U and B are not adjacent by the definition of discriminating path, $\langle U, W, B \rangle$ is an unshielded non-collider in $\mathcal{G}$ (because W is a parent of B in $\mathcal{G}$ by the definition of discriminating path). Because $\mathcal{G}$ and $\mathcal{G}'$ are Markov equivalent, $\langle U, W, B \rangle$ should also be a non-collider in $\mathcal{G}'$. But W is a collider on u in $\mathcal{G}$, and hence also a collider in $\mathcal{G}'$, which means the edge between U and W is into W. Thus $W \rightarrow B$ is in $\mathcal{G}'$, otherwise $\langle U, W, B \rangle$ would be an unshielded collider in $\mathcal{G}'$. This establishes the base case. In the inductive step, suppose the first n vertices after U on u are all parents of B in $\mathcal{G}'$, then we have an discriminating path for the $n+1$'st vertex between D and B in both graphs. Since the two graphs are Markov equivalent, the $n+1$'st vertex must be a parent of B as well, otherwise (e3) in Proposition 1 would be violated. This finishes our induction. So, in particular, A should be a parent of B in $\mathcal{G}'$, a contradiction. Thus condition (t3) also obtains.

Therefore, we can always identify a legitimate mark change to eliminate a difference as long as $\mathcal{G}$ and $\mathcal{G}'$ are still different. An induction on the number of differences between $\mathcal{G}$ and $\mathcal{G}'$ would do to complete the argument. $\square$

Obviously a DMAG and any of its LEGs satisfy the antecedent of Theorem 1, so they can be transformed to each other by a sequence of legitimate mark changes. Steps 0-2, in Figure 3, for example, portraits a stepwise transformation from G1 to H1.

Theorem 2. *Let $\mathcal{G}$ and $\mathcal{G}'$ be two Markov equivalent MAGs. If every bi-directed edge in $\mathcal{G}$ and every bi-directed edge in $\mathcal{G}'$ are invariant, then there is a sequence of legitimate mark changes that transforms one to the other.*

Proof. Without loss of generality, we prove that there is a transformation from $\mathcal{G}$ to $\mathcal{G}'$. It follows from the assumption that $\mathcal{G}$ and $\mathcal{G}'$ have the same set of bi-directed edges, and hence all differences between $\mathcal{G}$ and $\mathcal{G}'$ are of the same sort: $\rightarrow$ is in $\mathcal{G}$, while $\leftarrow$ is in $\mathcal{G}'$. Let

$$\mathbf{Diff} = \{y| \text{ there is a } x \text{ such that } x \rightarrow y \text{ is in } \mathcal{G} \text{ and } x \leftarrow y \text{ is in } \mathcal{G}'\}$$

Clearly $\mathcal{G}$ and $\mathcal{G}'$ are identical if and only if $\mathbf{Diff} = \emptyset$. We claim that if $\mathbf{Diff}$ is not empty, we can always identify a legitimate edge reversal (that is, two legitimate mark changes in a row) that eliminates a difference in direction.

Suppose $\mathbf{Diff}$ is not empty. We can choose a vertex $B \in \mathbf{Diff}$ such that no proper ancestor of B in $\mathcal{G}$ is in $\mathbf{Diff}$. Let

$$\mathbf{Diff}_B = \{x|x \rightarrow B \text{ is in } \mathcal{G} \text{ and } x \leftarrow B \text{ is in } \mathcal{G}'\}$$

Since $B \in \mathbf{Diff}$, $\mathbf{Diff}_B$ is not empty. Choose $A \in \mathbf{Diff}_B$ such that no proper descendant of A in $\mathcal{G}$ is in $\mathbf{Diff}_B$. Then changing $A \rightarrow B$ to $A \leftarrow B$ in $\mathcal{G}$ is a legitimate edge reversal.

To justify this claim, we verify the conditions in Lemma 2. Note that $A \rightarrow B$, by our choice, is a reversible edge in $\mathcal{G}$ (for $A \leftarrow B$ is in $\mathcal{G}'$, which is Markov equivalent to $\mathcal{G}$). It thus follows directly from Proposition 3 (and the assumption about bi-directed edges in $\mathcal{G}$) that $\mathbf{Sp}_{\mathcal{G}}(B) = \mathbf{Sp}_{\mathcal{G}}(A)$.

The argument for $\mathbf{Pa}_{\mathcal{G}}(B) = \mathbf{Pa}_{\mathcal{G}}(A) \cup \{A\}$ is virtually the same as Chickering's proof for DAGs. For any parent C of A in $\mathcal{G}$, C is also a parent of A in $\mathcal{G}'$, otherwise A is in $\mathbf{Diff}$ and is a proper ancestor of B in $\mathcal{G}$, which contradicts our choice of B. It follows that C is adjacent to B, otherwise $\langle C, A, B\rangle$ is an unshielded collider in $\mathcal{G}'$ but not one in $\mathcal{G}$, which would contradict the Markov equivalence between $\mathcal{G}$ and $\mathcal{G}'$.

Then C must be a parent of B in $\mathcal{G}$, otherwise $\mathcal{G}$ is not ancestral. Conversely, let $D \neq A$ be any other parent of B in $\mathcal{G}$. D must be adjacent to A, otherwise $\langle D, B, A \rangle$ is an unshielded collider in $\mathcal{G}$ but not one in $\mathcal{G}'$. D is not a spouse of A in $\mathcal{G}$, otherwise D is a spouse of B, according to what we just showed. So D is either a parent or a child of A in $\mathcal{G}$. Suppose it is a child of A, that is, $A \rightarrow D \rightarrow B$ is in $\mathcal{G}$. We derive a contradiction from this. Since $A \leftarrow B$ is in $\mathcal{G}'$, $A \rightarrow D \rightarrow B$ does not appear in $\mathcal{G}'$. That means either $A \leftarrow D$ or $D \leftarrow B$ (or both) is in $\mathcal{G}'$. In the former case, D is in $\mathbf{Diff}$ and is a proper ancestor of B in $\mathcal{G}$, which contradicts our choice of B. In the latter case, D is in $\mathbf{Diff}_B$ and is a proper descendant of A in $\mathcal{G}$, which contradicts our choice of A. Hence D can't be a child of A in $\mathcal{G}$, which means it is a parent of A in $\mathcal{G}$.

Note that after an edge reversal, no new bi-directed edge is introduced, so the assumption that every bi-directed edge is invariant still holds for the new graph. Hence we can always identify a legitimate edge reversal to eliminate a difference in direction as long as $\mathcal{G}$ and $\mathcal{G}'$ are still different. An easy induction on the number of differences between $\mathcal{G}$ and $\mathcal{G}'$ would do to complete the argument. $\square$

Since a LEG (of any MAG) only contains invariant bi-directed edges, two LEGs can always be transformed to each other via a sequence of legitimate mark changes according to the above theorem. For example, steps 2-4 in Figure 3 constitute a transformation from H1 (a LEG of G1) to H2 (a LEG of G2).

We are ready to prove the main result of this paper.

Theorem 3. *Two DMAGs $\mathcal{G}$ and $\mathcal{G}'$ are Markov equivalent if and only if there exists a sequence of single mark changes in $\mathcal{G}$ such that*

1. *after each mark change, the resulting graph is also a DMAG and is Markov equivalent to $\mathcal{G}$;*

2. *after all the mark changes, the resulting graph is $\mathcal{G}'$.*

Proof: The "if" part is trivial – since every mark change preserves the equivalence, the end is of course Markov equivalent to the beginning. Now suppose $\mathcal{G}$ and $\mathcal{G}'$ are equivalent. We show that there exists such a sequence of transformation. By

Proposition 2, there is a LEG $\mathcal{H}$ for $\mathcal{G}$ and a LEG $\mathcal{H}'$ for $\mathcal{G}'$. By Theorem 1, there is a sequence of legitimate mark changes s_1 that transforms $\mathcal{G}$ to $\mathcal{H}$, and there is a sequence of legitimate mark changes s_3 that transforms $\mathcal{H}'$ to $\mathcal{G}'$. By Theorem 2, there is a sequence of legitimate mark changes s_2 that transforms $\mathcal{H}$ to $\mathcal{H}'$. Concatenating s_1, s_2 and s_3 yields a sequence of legitimate mark changes that transforms $\mathcal{G}$ to $\mathcal{G}'$. $\square$

As a simple illustration, Figure 3 gives the steps in transforming G1 to G2 according to Theorem 3. That is, G1 is first transformed to one of its LEGs, H1; H1 is then transformed to H2, a LEG of G2. Lastly, H2 is transformed to G2.

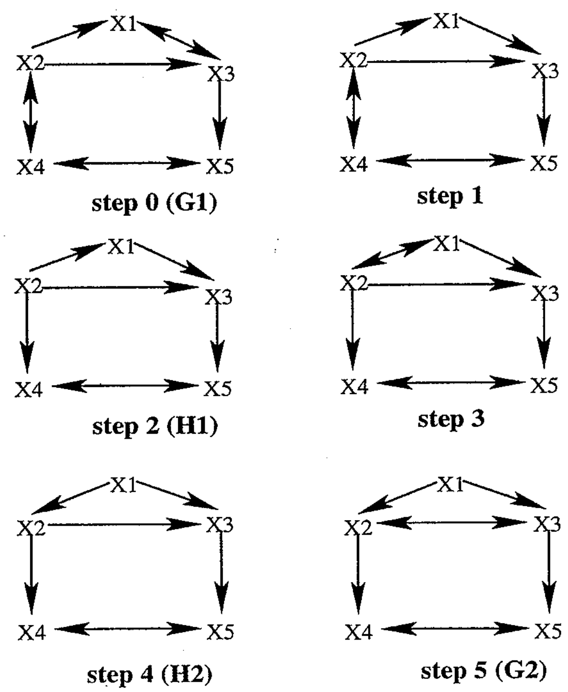

Figure 3: A transformation from G1 to G2

Theorems 1 and 2, as they are currently stated, are special cases of Theorem 3, but the proofs of them actually achieve a little more than they claim. The transformations constructed in the proofs of Theorems 1 and 2 are efficient in the sense that every mark change in the transformation eliminates a difference between the current DMAG and the target. So the transformations consist of as many mark changes as the number of differences at the beginning. By contrast, the transformation constructed in Theorem 3 may take some "detours", in that some mark changes in the way actually increase rather than decrease the difference between $\mathcal{G}$ and $\mathcal{G}'$. (This is not the case in Figure 3, but if, for example, we chose different LEGs for G1 or G2, there would be detours.) We believe that no such detour is really necessary, that is, there is always a transformation from $\mathcal{G}$ to $\mathcal{G}'$ consisting of as many mark changes as the number of differences between them. But we are yet unable to prove this conjecture.

3 Conclusion

In this paper we established a transformational property for Markov equivalent directed MAGs, which is a generalization of the transformational characterization of Markov equivalent DAGs given by Chickering (1995). It implies that no matter how different two Markov equivalent graphs are, there is a sequence of Markov equivalent graphs in between such that the adjacent graphs differ in only one edge. It could thus simplify derivations of invariance properties across a Markov equivalence class — in order to show two arbitrary Markov equivalent DMAGs share something in common, we only need to consider two Markov equivalent DMAGs with the minimal difference. Indeed, Chickering (1995) used his characterization to derive that Markov equivalent DAGs have the same number of parameters under the standard CPT parameterization (and hence would receive the same score under the typical penalized-likelihood type metrics). The discrete parameterization of DMAGs is currently under development[3]. We think our result will prove useful to show similar facts once the discrete parameterization is available.

The property, however, does not hold exactly for general MAGs, which may also contain undirected edges[4]. A simple counterexample is given in Figure 4. When we

[3]Drton and Richardson (2005) provide a parameterization for bi-directed graphs with binary variables, for which the problem of parameter equivalence does not arise because no two different bi-directed graphs are Markov equivalent.

[4]Undirected edges are motivated by the need to represent the presence of selection variables, features that influence which units are sampled (that are conditioned upon in sampling).

include undirected edges, the requirement of ancestral graphs is that the endpoints of undirected edges are of zero in-degree — that is, if a vertex is an endpoint of an undirected edge, then no edge is into that vertex (see Richardson and Spirtes (2002) for details). So, although the two graphs in Figure 4 are Markov equivalent MAGs, M1 cannot be transformed to M2 by a sequence of single legitimate mark changes, as adding any single arrowhead to M1 would make it non-ancestral. Therefore, for general MAGs, the transformation may have to include a stage of changing the undirected subgraph to a directed one in a wholesale manner.

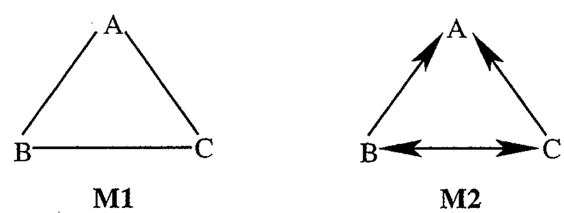

Figure 4: A simple counterexample with general MAGs: M1 can't be transformed into M2 by a sequence of legitimate single mark changes.

The transformational characterization for Markov equivalent DAGs was generalized, as a conjecture, to a transformational characterization for DAG I-maps by Meek (1996), which was later shown to be true by Chickering (2002). A graph is an I-map of another if the set of conditional independence relations entailed by the former is a subset of the conditional independence relations entailed by the latter. This generalized transformational property is used to prove the asymptotic correctness of the GES algorithm, an efficient search algorithm over the Markov equivalence classes of DAGs. The extension of both the property and the GES algorithm to MAGs is now under our investigation.

References

Ali, R.A., T. Richardson, and P. Spirtes (2004) Markov Equivalence for Ancestral Graphs. Department of Statistics, University of Washington, Tech Report 466.

Ali, R.A., T. Richardson, P. Spirtes, and J. Zhang (2005) Orientation Rules for Constructing Markov Equivalence Classes for Maximal Ancestral Graphs. Department of Statistics, University of Washington, Tech Report 476.

Andersson, S., D. Madigan, and M. Pearlman (1997) A Characterization of Markov Equivalence Classes of Acyclic Digraphs, in *Ann. Statist.* 25(2): 505-541.

Chickering, D.M. (1995) A transformational characterization of equivalent Bayesian network structures, in *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence*, 87-98. Morgan Kaufmann.

Chickering, D.M. (2002) Optimal Structure Identification with Greedy Search, in *Journal of Machine Learning Research*, 3:507-554.

Drton, M., and T. Richardson (2004) Graphical Answers to Questions About Likelihood Inference in Gaussian Covariance Models. Department of Statistics, University of Washington, Tech Report 467.

Drton, M., and T. Richardson, T. (2005) Binary Models for Marginal Independence. Department of Statistics, University of Washington, Tech Report 474.

Meek, C. (1996) Graphical Models: Selecting Causal and Statistical Models. Carnegie Mellon University, Philosophy Department, PhD Thesis.

Richardson, T., and P. Spirtes (2002) Ancestral Markov Graphical Models, in *Ann. Statist.* 30(4): 962-1030.

Richardson, T., and P. Spirtes (2003) Causal Inference via Ancestral Graph Models, in *Highly Structured Stochastic Systems.* Eds. P. Green, N. Hjort, and S. Richardson. Oxford University Press.

Spirtes, P., and T. Richardson (1996) A Polynomial Time Algorithm For Determining DAG Equivalence in the Presence of Latent Variables and Selection Bias, in *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics.*

Spirtes, P., T. Richardson, and C. Meek (1997) Heuristic Greedy Search Algorithms for Latent Variable Models, in *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics.*

Verma, T., and J. Pearl (1990) Equivalence and Synthesis of Causal Models, in *Proceedings of 6th Conference on Uncertainty in Artificial Intelligence*, 220-227.

Zhang, J., and P. Spirtes (2005) A Characterization of Markov Equivalence Classes for Ancestral Graphical Models. Department of Philosophy, Carnegie Mellon University, Tech Report 168.

A Transformational Characterization of Markov Equivalence between DAGs with Latent Variables

Jiji Zhang & Peter Spirtes

March 24, 2005

Abstract

Different directed acyclic graphs (DAGs) may be Markov equivalent in the sense that they entail the same conditional independence relations among the observed variables. Markov equivalence between DAGs (with no latent variables) has been characterized in various ways, each of which has been found useful for certain purposes. In particular, Chickering's transformational characterization is useful in deriving properties shared by Markov equivalent DAGs, and, with certain generalization, is needed to prove the asymptotic correctness of a search procedure over Markov equivalence classes, known as the GES algorithm.

Maximal ancestral graphs (MAGs) are a generalization of DAGs that can represent the observable conditional independence relations (as well as some causal features) of DAG models with latent variables. Thus Markov equivalence between DAGs with latent variable is reduced to Markov equivalence between the corresponding MAGs. However, no characterization of Markov equivalent MAGs is yet available that is analogous to Chickering's transformational characterization. The main contribution of the current paper is to establish such a characterization for directed MAGs, which we expect will have similar uses as it does for DAGs.

1 INTRODUCTION

Markov equivalence between directed acyclic graphs (DAGs) has been characterized in several ways (e.g., Verma and Pearl 1990, Chickering 1995, Andersson et al. 1997). All of them have been found useful for various purposes. In particular, the transformational characterization provided by Chickering (1995) — that two DAGs are Markov equivalent if and only if one can be transformed to the other by a sequence of single edge reversals that preserve Markov equivalence — is useful in deriving properties shared by Markov equivalent DAGs. Moreover, when generalized, the transformational characterization implies the asymptotic correctness of the GES algorithm, an efficient search procedure over Markov equivalence classes of DAGs (Meek 1996, Chickering 2002).

In many situations, however, we need also to consider DAGs with latent variables. Indeed there are cases where no DAGs can perfectly explain the observed conditional independence relations unless latent variables are introduced. Such latent variable models, fortunately, can be represented by ancestral graphical models (Richardson and Spirtes 2002), in that for any DAG with latent variables, there is a (maximal) ancestral graph that captures the exact observable conditional independence relations as well as some of the causal relations entailed by that DAG. Since ancestral graphs do not explicitly include latent variables, they provide, among other virtues, a finite search space of latent variable models (Spirtes et al. 1997).

Markov equivalence for ancestral graphs has been characterized in ways analogous to the one given by Verma and Pearl (1990) for DAGs (Spirtes and Richardson 1996, Ali et al. 2004). However, no characterization is yet available that is analogous to Chickering's transformational characterization. In this paper we establish one for directed ancestral graphs. Specifically we show that two directed maximal ancestral graphs are Markov equivalent if and only if one can be transformed to the other by a sequence of single mark changes — adding or dropping an arrowhead — that preserve Markov equivalence. This characterization we expect will have similar uses as Chickering's does for DAGs. In particular, it is a step towards justifying the application of the GES algorithm to MAGs, and hence to latent variable DAG models.

The paper is organized as follows. The remainder of this section introduces the relevant definitions and notations. We then present the main result in section 2, drawing on some facts proved in Zhang and Spirtes (2005) and Ali et al. (2005). We conclude the paper in section 3 with a discussion of the potential application, limitation and generalization of our result.

1.1 DIRECTED ANCESTRAL GRAPHS

In full generality, an ancestral graph can contain three kinds of edges: directed edge ($\rightarrow$), bi-directed edge ($\leftrightarrow$) and undirected edge (—). In this paper, however, we will confine ourselves to directed ancestral graphs — which do not contain undirected edges — until section 3, where we explain why our result does not hold for general ancestral graphs. The class of directed ancestral graphs, due to its inclusion of bi-directed edges, is suitable for representing observed conditional independence structures in the presence of latent confounders (see Figure 1).

By a **directed mixed graph** we denote an arbitrary graph that can have two kinds of edges: directed and bi-directed. The two ends of an edge we call **marks** or **orientations**. So the two marks of a bi-directed edge are both **arrowheads** (>), while a directed edge has one arrowhead and one **tail** (–) as its marks. Sometimes we say an edge is **into** (or **out of**) a vertex if the mark of the edge at the vertex is an arrowhead (or a tail). The meaning of the standard graph theoretical concepts, such as **parent/child**, **(directed) path**, **ancestor/descendant**, etc., remains the same in mixed graphs. Furthermore, if there is a bi-directed edge between two vertices A and B ($A \leftrightarrow B$), then A is called a **spouse** of B and B a spouse of A.

Definition 1 (ancestral). *A directed mixed graph is* **ancestral** *if*

(a1) there is no directed cycle; and

(a2) for any two vertices A and B, if A is a spouse of B (i.e., $A \leftrightarrow B$), then A is not an ancestor of B.

Clearly DAGs are a special case of directed ancestral graphs (with no bi-directed edges). Condition (a1) is just the familiar one for DAGs. Condition (a2), together with (a1), defines a nice feature of arrowheads — that is, an arrowhead implies non-ancestorship. This motivates the term "ancestral" and induces a natural causal interpretation of ancestral graphs (see, e.g., Richardson and Spirtes 2003).

Mixed graphs encode conditional independence relations by essentially the same graphical criterion as the well-known *d-separation* for DAGs, except that in mixed graphs colliders can arise in more edge configurations than they do in DAGs. Given a path u in a mixed graph, a non-endpoint vertex V on u is called a **collider** if the two edges incident to V on u are both into V, otherwise V is called a **non-collider**.

Definition 2 (m-separation). *In a mixed graph, a path u between vertices A and B is* **active** *(***m-connecting***) relative to a set of vertices $\mathbf{Z}$ ($A, B \notin \mathbf{Z}$) if*

i. every non-collider on u is not a member of $\mathbf{Z}$;

ii. every collider on u is an ancestor of some member of $\mathbf{Z}$.

A and B are said to be **m-separated** *by $\mathbf{Z}$ if there is no active path between A and B relative to $\mathbf{Z}$.*

The following property is true of DAGs: if two vertices are not adjacent, then there is a set of some other vertices that m-separates (d-separates) the two. This, however, is not true of directed ancestral graphs in general, which motivates the following definition.

Definition 3 (maximality). *A directed ancestral graph is said to be* **maximal** *if for any two non-adjacent vertices, there is a set of vertices that m-separates them.*

It is shown in Richardson and Spirtes (2002) that every non-maximal ancestral graph has a unique supergraph that is ancestral and maximal, and it is easy to construct the maximal supergraph given a non-maximal ancestral graph. This justifies considering only those ancestral graphs that are maximal (MAGs). From now on, we focus on directed maximal ancestral graphs, which we will refer to as DMAGs. A notion closely related to maximality is that of inducing path:

Definition 4 (inducing path). *In an ancestral graph, a path u between A and B is called an* **inducing path** *if every non-endpoint vertex on u is a collider and is an ancestor of either A or B.*

Richardson and Spirtes (2002) proved that the presence of an inducing path is necessary and sufficient for two vertices not to be m-separated by any set. So, to show that a graph is maximal, it suffices to demonstrate that there is no inducing path between any two non-adjacent vertices in the graph.

Given any DAG with (or without) latent variables, the conditional independence relations as well as the causal relations among the observed variables can be represented by a DMAG that includes only the observed variables. The DMAG is constructed as follows: for every pair of observed variables, O_i and O_j, put an edge between them if and only if they are not d-separated by any set of other observed variables in the given DAG, and mark an arrowhead at O_i (O_j) on the edge if it is *not* an ancestor of O_j (O_i) in the given DAG.

For example, Figure 1(a) is a DAG with latent variables $\{L1, L2, L3\}$. Figure 1(b) depicts the DMAG (G1) resulting from the above construction. The m-separation relations in G1 correspond exactly to the d-separation relations over $\{X1, X2, X3, X4, X5\}$ in Figure 1(a). By contrast, no DAG without extra latent variables has the exact same d-separation relations. Furthermore, the orientations in G1 accurately represent the ancestor relationships — which, upon natural interpretations, are causal relationships — among the observed variables in 1(a). (This, however, is not the case with G2.)

1.2 MARKOV EQUIVALENCE

A DMAG represents the set of joint distributions that satisfy its *global Markov property*, i.e., the set of distributions of which the conditional independence relations entailed by m-separations in the DMAG hold. Hence, if two DMAGs share the same m-separation structures, then they represent the same set of distributions.

Definition 5 (Markov equivalence). *Two DMAGs $\mathcal{G}_1, \mathcal{G}_2$ (with the same set of vertices) are* **Markov equivalent** *if for any three disjoint sets of vertices $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, $\mathbf{X}$ and $\mathbf{Y}$ are m-separated by $\mathbf{Z}$ in $\mathcal{G}_1$ if and only if $\mathbf{X}$ and $\mathbf{Y}$ are m-separated by $\mathbf{Z}$ in $\mathcal{G}_2$.*

Figure 1(c), for example, is a DMAG Markov equivalent to 1(b). It is well known that two DAGs are Markov equivalent if and only if they have the same adjacencies

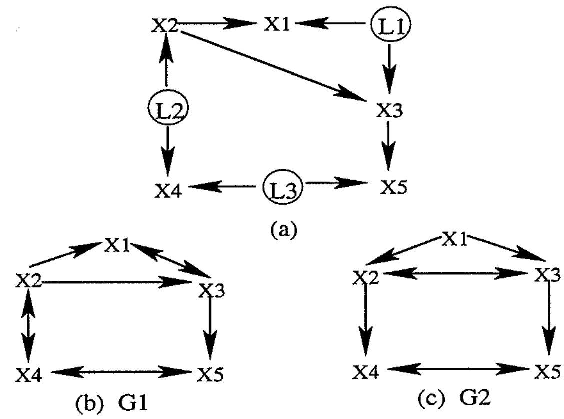

Figure 1: (a): A DAG with latent variables; (b): A DMAG that captures both the conditional independence and causal relations among the observed variables represented by (a); (c): A DMAG that entails the right conditional independence relations but not the right causal relations in (a).

and the same unshielded colliders (Verma and Pearl 1990). (A triple $\langle A, B, C\rangle$ is said to be **unshielded** if A, B are adjacent, B, C are adjacent but A, C are not adjacent.) The conditions are still necessary for Markov equivalence between DMAGs, but are not sufficient. For two DMAGs to be equivalent, some shielded colliders have to be present in both or neither of the graphs. The next definition is related to this.

Definition 6 (discriminating path). *In a DMAG, a path between X and Y, $u = \langle X, \cdots, W, V, Y\rangle$, is a* **discriminating path** *for V if*

i. u includes at least three edges (i.e., at least four vertices as specified);

ii. V is adjacent to an endpoint Y on u; and

iii. X is not adjacent to Y, and every vertex between X and V is a collider on u and is a parent of Y.

Discriminating paths behave similarly to unshielded triples in that if $u = \langle X, \cdots, W, V, Y\rangle$ is discriminating for V, then $\langle W, V, Y\rangle$ is a (shielded) collider if and only if every set that m-separates X and Y excludes V; it is a non-collider if and only if every set that m-separates X and Y contains V. The following proposition is proved in Spirtes and Richardson (1996)[1].

Proposition 1. *Two DMAGs over the same set of vertices are Markov equivalent if and only if*

(e1) They have the same adjacencies;

(e2) They have the same unshielded colliders;

(e3) If a path u is a discriminating path for a vertex B in both graphs, then B is a collider on the path in one graph if and only if it is a collider on the path in the other.

2 A Transformational Property of DMAGs

We present the main result of the paper in this section, namely Markov equivalent DMAGs can be transformed to each other by a sequence of single mark changes that preserve Markov equivalence. We first describe in section 2.1 two corollaries from Zhang and Spirtes (2005) and Ali et al. (2005) which our arguments will rely upon. Section 2.2 establishes sufficient and necessary conditions for a single mark change to preserve equivalence. The theorems are then presented in section 2.3.

2.1 Loyal Equivalent Graph

Given a MAG $\mathcal{G}$, a mark (or edge) in $\mathcal{G}$ is **invariant** if it is present in all MAGs Markov equivalent to $\mathcal{G}$. Invariant marks are particularly important for causal inference because data alone usually can't distinguish between members of a Markov equivalence class. An algorithm for detecting all invariant arrowheads in a MAG is given by Ali et al. (2005), and one for further detecting all invariant tails is presented in Zhang and Spirtes (2005). In what follows we list some facts proved in Zhang and Spirtes (2005) which will lead to two propositions needed for our current purpose.

[1]The conditions are also valid for maximal ancestral graphs that contain undirected edges.

Zhang and Spirtes (2005) give a set of orientation rules by which one can construct, given an arbitrary MAG, a complete representation of the invariant marks of the MAG. The representation they use is a graph that can contain three kinds of marks: arrowhead ($>$), tail ($-$) and circle ($\circ$), and hence in general six kinds of edges: $\circ\!\!-\!\!\circ$, $\circ\!\!\rightarrow$, $\rightarrow$, $\leftrightarrow$, $-$ and $\circ\!\!-$; although in the case of DMAGs, only the first four are needed. Intuitively circles purport to indicate that the corresponding mark could be either an arrowhead or a tail in the Markov equivalence class of $\mathcal{G}$, whereas a non-circle mark represents a mark shared by all members of the equivalence class.

Given a MAG $\mathcal{G}$, the construction starts with $\mathcal{P}_0$, a graph having the same adjacencies as $\mathcal{G}$ does, but every edge therein is of the form $\circ\!\!-\!\!\circ$. Then a set of orientation rules is applied to change some of the circles into arrowheads or tails. In particular, the following five orientation rules are sufficient to identify all invariant arrowheads in $\mathcal{G}$ ($*$ is used as a meta-symbol that represents whatever mark that may be present at an end of an edge, i.e., a circle, an arrowhead, or a tail):

$\mathcal{R}0$ For every triple $\alpha *\!\!-\!\!* \beta \circ\!\!-\!\!* \gamma$ s.t. α, γ are not adjacent, if it is an unshielded collider in $\mathcal{G}$, then orient the triple as $\alpha *\!\!\rightarrow \beta \leftarrow\!\!* \gamma$.

$\mathcal{R}1$ If $\alpha * \rightarrow \beta \circ - * \gamma$, and α and γ are not adjacent, then orient the triple as $\alpha *\!\!\rightarrow \beta \rightarrow \gamma$.

$\mathcal{R}2$ If $\alpha \rightarrow \beta *\!\!\rightarrow \gamma$ or $\alpha *\!\!\rightarrow \beta \rightarrow \gamma$, and $\alpha *\!\!-\!\!\circ \gamma$, then orient $\alpha *\!\!-\!\!\circ \gamma$ as $\alpha *\!\!\rightarrow \gamma$.

$\mathcal{R}3$ If $\alpha *\!\!\rightarrow \beta \leftarrow\!\!* \gamma$, $\alpha *\!\!-\!\!\circ \theta \circ\!\!-\!\!* \gamma$, α and γ are not adjacent, and $\theta *\!\!-\!\!\circ \beta$, then orient $\theta *\!\!-\!\!\circ \beta$ as $\theta *\!\!\rightarrow \beta$.

$\mathcal{R}4$ If $u = \langle \theta, ..., \alpha, \beta, \gamma \rangle$ is a discriminating path between θ and γ for β, and $\beta \circ\!\!-\!\!* \gamma$; then if $\beta \rightarrow \gamma$ appears in $\mathcal{G}$, orient $\beta \circ\!\!-\!\!* \gamma$ as $\beta \rightarrow \gamma$; otherwise orient the triple $\langle \alpha, \beta, \gamma \rangle$ as $\alpha \leftrightarrow \beta \leftrightarrow \gamma$.

It is to be understood that $\mathcal{R}0$ is applied first to $\mathcal{P}_0$, and then $\mathcal{R}1 - \mathcal{R}4$ are applied repeatedly and exhaustively, i.e., until no more circles can be oriented. Let $\mathcal{P}$ be the resulting graph, which will be called the **partial ancestral graph (PAG)** of $\mathcal{G}$. We state three facts from Zhang and Spirtes (2005). First of all, $\mathcal{R}0 - \mathcal{R}4$ are sound and complete with respect to invariant arrowheads in the following sense (Lemma 2 and Theorem 2 in Zhang and Spirtes (2005)):

Fact 1. *Every non-circle mark in $\mathcal{P}$ is invariant. That is, all arrowheads and tails in $\mathcal{P}$ are shared by all DMAGs that are Markov equivalent to $\mathcal{G}$. Furthermore, every invariant arrowhead in $\mathcal{G}$ appears in $\mathcal{P}$.*

We also need the following fact, which is proved in Lemma 6 in Zhang and Spirtes (2005) (see also Corollary 4.1 in Ali et al. 2005).

Fact 2. *For any $A \circ\!\!-\!\!\circ B$ and vertex C in $\mathcal{P}$, $C \leftrightarrow A$ appears in $\mathcal{P}$ if and only if $C \leftrightarrow B$ appears in $\mathcal{P}$.*

Lastly, let the **circle component** of $\mathcal{P}$ — denoted by $\mathcal{P}^c$ — be the induced subgraph that consists of all $\circ\!\!-\!\!\circ$ edges in $\mathcal{P}$. The following fact, as a special case of Lemma 16 in Zhang and Spirtes (2005), is key to the first proposition we want to establish in the current section.

Fact 3. *Let $\mathcal{H}$ be the graph that results from changing all $\circ\!\!\rightarrow$ edges in $\mathcal{P}$ into $\rightarrow$, and orienting $\mathcal{P}^c$ into a directed acyclic graph with no unshielded colliders. Then $\mathcal{H}$ is a DMAG and is Markov equivalent to $\mathcal{G}$.*

The main aim of the current section is to show that for any DMAG, there is a Markov equivalent DMAG such that every bi-directed edge therein is invariant, and every directed edge in the given DMAG is retained. This obviously will rely on Fact 3 just stated. The only extra piece of fact needed is the following:

Fact 4. *Let $\mathcal{G}^*$ be a directed ancestral graph over a set of vertices $\mathbf{V}$. If $\mathcal{G}^*$ has no unshielded colliders, then there is a directed acyclic graph $\mathcal{D}$ over $\mathbf{V}$ such that*

(1) $\mathcal{D}$ has the same adjacencies as $\mathcal{G}^*$ does;

(2) $\mathcal{D}$ has no unshielded colliders;

(3) every directed edge in $\mathcal{G}^*$ is also in $\mathcal{D}$.

Proof. Here is a formal construction of the DAG. The ancestor relationship in $\mathcal{G}^*$ naturally induces a partial order over $\mathbf{V}$. Extend this partial order to a total order. Put an edge $A \rightarrow B$ in $\mathcal{D}$ if and only if A and B are adjacent in $\mathcal{G}^*$, and A precedes B in the order. The resulting $\mathcal{D}$ is obviously a DAG, and has the same adjacencies as $\mathcal{G}^*$ does. Furthermore, all the directed edges in $\mathcal{G}^*$ are retained in $\mathcal{D}$, because the ordering that yields $\mathcal{D}$ respects the ordering in $\mathcal{G}^*$. So the only difference between $\mathcal{G}^*$ and $\mathcal{D}$ is that bi-directed edges in the former are changed into directed edges in the latter. It follows that all arrowheads, and hence all colliders, in $\mathcal{D}$ are also in $\mathcal{G}^*$. Since there is no unshielded collider in $\mathcal{G}^*$, there is no unshielded collider in $\mathcal{D}$. $\square$

Proposition 2. *Given any DMAG $\mathcal{G}$, there exists a DMAG $\mathcal{H}$ such that*

(1) $\mathcal{H}$ is Markov equivalent to $\mathcal{G}$;

(2) every bi-directed edge in $\mathcal{H}$ is invariant;

(3) every directed edge in $\mathcal{G}$ is also in $\mathcal{H}$.

Proof. Let $\mathcal{P}$ be the PAG of $\mathcal{G}$, and $\mathcal{G}^*$ be the corresponding subgraph of $\mathcal{G}$ over the vertices of $\mathcal{P}^c$, the circle component of $\mathcal{P}$. It is clear that $\mathcal{G}^*$ does not contain any unshielded collider, otherwise it would have been introduced into $\mathcal{P}$ by $\mathcal{R}0$. Hence, by Fact 4, there is a DAG $\mathcal{D}$ over the vertices of $\mathcal{P}^c$ with the same adjacencies and no unshielded colliders such that all directed edges in $\mathcal{G}^*$ are retained in $\mathcal{D}$. Let $\mathcal{H}$ be the graph resulting from changing all $\circ\!\!\rightarrow$ edges in $\mathcal{P}$ into $\rightarrow$, and orienting $\mathcal{P}^c$ into $\mathcal{D}$. By Fact 3, $\mathcal{H}$ is a DMAG and is Markov equivalent to $\mathcal{G}$. Also, since every non-circle mark in $\mathcal{P}$ is invariant (Fact 1), and the extra orientations do not introduce new bi-directed edges, every bi-directed edge in $\mathcal{H}$ is invariant. Finally, any directed edge in $\mathcal{G}$ is either in $\mathcal{G}^*$, or corresponds to a directed edge or a $\circ\!\!\rightarrow$ edge in $\mathcal{P}$. In either case, the directed edge is also in $\mathcal{H}$ by our construction. $\square$

We will call $\mathcal{H}$ in Proposition 2 a **Loyal Equivalent Graph (LEG)** of $\mathcal{G}$. In general a DMAG could have multiple LEGs. A distinctive feature of the LEGs is that they have the fewest bi-directed edges among the Markov equivalent DMAGs[2]. Drton

[2]Proposition 2 is actually a special case of Corollary 18 in Zhang and Spirtes (2005), which, for general MAGs, also asserts that the LEGs have the fewest undirected edges as well.

and Richardson (2004) explored the statistical significance of this fact for bi-directed graphs (graphs that contain only bi-directed edges). Roughly speaking, if the LEGs of a bi-directed graph are DAGs, then fitting is easy; otherwise fitting is not easy (in a specific technical sense).

Another feature which is particularly relevant to our argument is that between a DMAG and any of its LEGs, only one kind of differences is possible, namely, some bi-directed edges in the DMAG are oriented as directed edges in its LEG. For a simple illustration, compare the graphs in Figure 2, where H1 is a LEG of G1, and H2 is a LEG of G2.

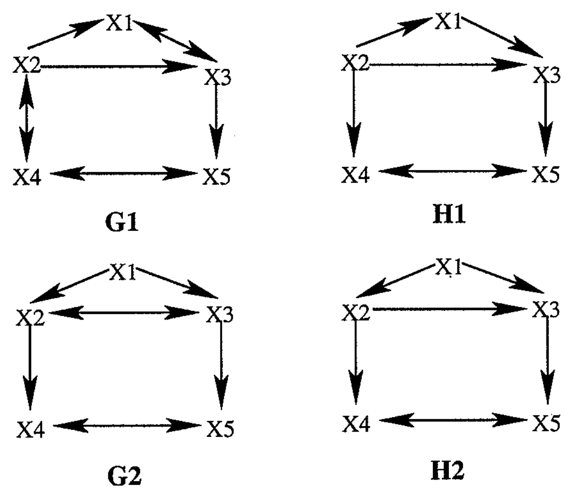

Figure 2: A LEG of G1 (H1) and a LEG of G2 (H2)

A directed edge in a DMAG is called **reversible** if there is another Markov equivalent DMAG in which the direction of the edge is reversed. It would be convenient to have the following proposition in order to prove Theorem 2 below.

Proposition 3. *Let $A \rightarrow B$ be any reversible edge in a DMAG $\mathcal{G}$. For any vertex C (distinct from A and B), there is an invariant bi-directed edge between C and A if and only if there is an invariant bi-directed edge between C and B.*

Proof. Since $A \rightarrow B$ is reversible (which means neither of the two marks of the edge is invariant), in $\mathcal{P}$ — the PAG of $\mathcal{G}$ — the edge between A and B would be $A \circ\!\!-\!\!\circ B$

by Fact 1. For any C, if there is an invariant bi-directed edge between C and A, by Fact 1, $C \leftrightarrow A$ would appear in $\mathcal{P}$. By Fact 2, $C \leftrightarrow B$ is also in $\mathcal{P}$, and hence is an invariant bi-directed edge in $\mathcal{G}$. Conversely, if there is an invariant bi-directed edge between C and B in $\mathcal{G}$, the same argument shows that there would also be an invariant bi-directed edge between C and A. $\square$

In particular, if $\mathcal{H}$ is a LEG of a DMAG, then $A \rightarrow B$ being reversible implies that A and B have the same set of spouses, as every bi-directed edge in $\mathcal{H}$ is invariant.

2.2 Equivalence-Preserving Mark Change

Eventually we will show that two Markov equivalent DMAGs can be connected by a sequence of equivalence-preserving mark changes. It is thus desirable to give some simple graphical conditions under which a single mark change would preserve equivalence. The next lemma presents necessary and sufficient conditions under which adding an arrowhead to a directed edge (i.e., changing the directed edge to a bi-directed one) preserves Markov equivalence. By symmetry, they are also the conditions for dropping an arrowhead from a bi-directed edge while preserving Markov equivalence.

Lemma 1. *Let $\mathcal{G}$ be an arbitrary DMAG, and $A \rightarrow B$ an arbitrary directed edge in $\mathcal{G}$. Let $\mathcal{G}'$ be the graph identical to $\mathcal{G}$ except that the edge between A and B is $A \leftrightarrow B$. (In other words, $\mathcal{G}'$ is the result of simply changing the mark at A on $A \rightarrow B$ from an tail into an arrowhead.) $\mathcal{G}'$ is a DMAG and Markov equivalent to $\mathcal{G}$ if and only if*

(t1) there is no directed path from A to B other than $A \rightarrow B$;

(t2) For any $C \rightarrow A$ in $\mathcal{G}$, $C \rightarrow B$ is also in $\mathcal{G}$; and for any $D \leftrightarrow A$ in $\mathcal{G}$, either $D \rightarrow B$ or $D \leftrightarrow B$ is in $\mathcal{G}$;

(t3) There is no discriminating path for A on which B is the endpoint adjacent to A.

Proof. We first show that each of the conditions is necessary (**only if**). Obviously if (t1) fails, $\mathcal{G}'$ will not be ancestral (see Definition 1). The failure of (t2) could be due to one of the following two cases:

Case 1: there is a vertex C which is a parent of A but not a parent of B. If B and C are not adjacent, then there is an unshielded collider in $\mathcal{G}'$ but not in $\mathcal{G}$, and hence the two graphs are not Markov equivalent, according to Proposition 1. If B and C are adjacent, then $\mathcal{G}$ can't be ancestral (unless we have $C \to B$).

Case 2: there is a vertex C which is a spouse of A but not a parent or spouse of B. Again, if B and C are not adjacent, the two graphs can't be Markov equivalent because there is an unshielded collider in $\mathcal{G}$ but not in $\mathcal{G}'$. If B and C are adjacent, the edge between them must be $B \to C$ in $\mathcal{G}$ by the supposition. But then there is an almost directed cycle in $\mathcal{G}$, and hence $\mathcal{G}$ is not ancestral.

If (t3) fails, that is, there is a discriminating path $u = \langle U, \cdots, V, A, B \rangle$ for A. If the edge between V and A is into A, then $\mathcal{G}$ and $\mathcal{G}'$ are not Markov equivalent, because (e3) in Proposition 1 is violated. If, on the other hand, the edge between V and A is not into A, then it must be $A \to V$. By the definition of discriminating path (Definition 6), V is a parent of B. So we have $A \to V \to B \leftrightarrow A$ in $\mathcal{G}'$, an almost directed cycle, which means $\mathcal{G}'$ is not ancestral.

Next, we demonstrate the sufficiency of the conditions (**if**). Suppose (t1)-(t3) are met. We first verify that $\mathcal{G}'$ is a DMAG, i.e., it is both ancestral and maximal. Suppose for contradiction that $\mathcal{G}'$ is not ancestral. Since $\mathcal{G}$ is ancestral, and $\mathcal{G}'$ differs from $\mathcal{G}$ only regarding the edge between A and B, in $\mathcal{G}'$ the violation of the definition of ancestral graphs (Definition 1) must involve the edge between A and B. So it can't be a violation of (a1), because a directed cycle would not involve $A \leftrightarrow B$. If it is a violation of (a2), i.e., there is an almost directed cycle in $\mathcal{G}'$. That cycle includes $A \leftrightarrow B$, which means either A is an ancestor of B or B is an ancestor of A in $\mathcal{G}'$. The former case contradicts (t1), and the latter case yields a directed cycle in $\mathcal{G}$. So there can't be any violation of (a2) in $\mathcal{G}'$. Hence $\mathcal{G}'$ is ancestral.

To show that $\mathcal{G}'$ is maximal, Suppose for the sake of contradiction that there is an inducing path u in $\mathcal{G}'$ between two non-adjacent vertices, D and E. Then u must include $A \leftrightarrow B$, otherwise u would also be an inducing path in $\mathcal{G}$. Furthermore, A is not an endpoint of u, otherwise u is still an inducing path in $\mathcal{G}$ (in fact, there would be an almost directed path in $\mathcal{G}$ in that case). Suppose, without loss of generality, that D is the endpoint closer to A on u than it is to B. We show that some vertex on $u(D, A)$ other than A is B's spouse. Suppose not — that is, suppose no vertex between D and A on u is B's spouse, we argue by induction that every vertex on $u(A, D)$, and in particular D, is a parent of B. By (t2), the vertex adjacent to A on $u(D, A)$ is either a parent or a spouse of B, but it is not a spouse by supposition, so it is a parent. In the inductive step, suppose the first n vertices next to A on $u(D, A)$ are B's parents, then the $n + 1$'st vertex V must be adjacent to B, otherwise the sub-path of u between V and B forms a discriminating path for A, which violates (t3). By supposition, V is not a spouse of B, i.e., it is not the case that $V \leftrightarrow B$. It can't be $V \leftarrow B$ either, because in that case there would be an almost directed cycle in $\mathcal{G}'$ (as the vertex before V, by the inductive hypothesis, is a parent of B), which we have shown to be impossible. So V must be a parent of B. Thus we have shown that every vertex on $u(A, D)$, and in particular D, is a parent of B. Then B must be an ancestor of E, because by the definition of inducing path (Definition 4), B is an ancestor of either D or E. So D is an ancestor of E, which means the vertex adjacent to E on u must be an ancestor of E. But the edge between that vertex and E must be into that vertex (as it is a collider on u by definition), so there is a directed or almost directed cycle in $\mathcal{G}'$, which we have shown to be absent. Hence a contradiction. So some vertex on $u(D, A)$ other than A is a spouse of B. Let C be such a vertex on $u(D, A)$. Replacing $u(C, B)$ on u with $C \leftrightarrow B$ yields an inducing path between D and E in $\mathcal{G}$, which contradicts the fact that $\mathcal{G}$ is maximal.

Having shown that $\mathcal{G}'$ is a DMAG, we now verify that $\mathcal{G}$ and $\mathcal{G}'$ satisfy the conditions for Markov equivalence in Proposition 1. Obviously they have the same adjacencies, and share the same colliders except possibly A. But A will not be a collider in an unshielded triple, for condition (t2) requires that any vertex that is incident to an edge into A is also adjacent to B. So the only worry is that a triple $\langle C, A, B \rangle$ might be discriminated by a path, but (t3) guarantees that there is no such path. Therefore, $\mathcal{G}'$ is Markov equivalent to $\mathcal{G}$. □

We say a mark change is *legitimate* when the conditions in Lemma 1 are satisfied. Recall that for DAGs the basic unit of equivalence-preserving transformation is (covered) edge reversal (Chickering 1995). In the current paper we treat an edge reversal as simply a special case of two consecutive mark changes. That is, a reversal of $A \rightarrow B$ is simply to first add an arrowhead at A (to form $A \leftrightarrow B$), and then to drop the arrowhead at B (to form $A \leftarrow B$). An edge reversal is said to be legitimate if both of the two consecutive mark changes are legitimate. Given Lemma 1, it is straightforward to check the validity of the following condition for legitimate edge reversal. (We use $\mathbf{Pa}_{\mathcal{G}}/\mathbf{Sp}_{\mathcal{G}}$ to denote the set of parents/spouses of a vertex in $\mathcal{G}$.)

Lemma 2. *Let $\mathcal{G}$ be an arbitrary DMAG, and $A \rightarrow B$ an arbitrary directed edge in $\mathcal{G}$. The reversal of $A \rightarrow B$ is legitimate if and only if $\mathbf{Pa}_{\mathcal{G}}(B) = \mathbf{Pa}_{\mathcal{G}}(A) \cup \{A\}$ and $\mathbf{Sp}_{\mathcal{G}}(B) = \mathbf{Sp}_{\mathcal{G}}(A)$.*

When there is no bi-directed edge in $\mathcal{G}$, that is, when $\mathcal{G}$ is a DAG, the condition in Lemma 2 is reduced to the familiar definition for *covered edge*, i.e., $\mathbf{Pa}_{\mathcal{G}}(B) = \mathbf{Pa}_{\mathcal{G}}(A) \cup \{A\}$ (Chickering 1995). The condition given by Drton and Richardson (2004) for a bi-directed edge in a bi-directed graph to be orientable as a directed edge in either direction ($\mathbf{Sp}_{\mathcal{G}}(B) = \mathbf{Sp}_{\mathcal{G}}(A)$) can be viewed as another special case of the above lemma.

2.3 Transformation between Equivalent DMAGs

We first state two intermediate theorems crucial for the main result we are heading for. The first one says if the differences between two Markov equivalent DMAGs $\mathcal{G}$ and $\mathcal{G}'$ are all of the following sort: a directed edge is in $\mathcal{G}$ while the corresponding edge is bi-directed in $\mathcal{G}'$, then there is a sequence of legitimate mark changes that transforms one to the other. The second one says that if every bi-directed edge in $\mathcal{G}$ and every bi-directed edge in $\mathcal{G}'$ are invariant, then there is a sequence of legitimate mark changes (edge reversals) that transforms one to the other. The proofs follow the strategy of Chickering's proof for DAGs.

Theorem 1. *Let $\mathcal{G}$ and $\mathcal{G}'$ be two Markov equivalent DMAGs. If every bi-directed edge in $\mathcal{G}$ is also in $\mathcal{G}'$, and every directed edge in $\mathcal{G}'$ is also in $\mathcal{G}$, then there is a sequence of legitimate mark changes that transforms one to the other.*

Proof. We prove that there is a sequence of transformation from $\mathcal{G}$ to $\mathcal{G}'$, the reverse of which will be a transformation from $\mathcal{G}'$ to $\mathcal{G}$. Specifically we show that as long as

$\mathcal{G}$ and $\mathcal{G}'$ are different, there is always a legitimate mark change that can eliminate a difference between them. The theorem then follows from a simple induction on the number of differences.

The antecedent of the theorem implies that the differences between $\mathcal{G}$ and $\mathcal{G}'$ are all of the same sort: a directed edge ($\rightarrow$) is in $\mathcal{G}$ while the corresponding edge in $\mathcal{G}'$ is bi-directed ($\leftrightarrow$). Let

$$\mathbf{Diff} = \{y| \text{ there is a } x \text{ such that } x \rightarrow y \text{ is in } \mathcal{G} \text{ and } x \leftrightarrow y \text{ is in } \mathcal{G}'\}$$

It is clear that $\mathcal{G}$ and $\mathcal{G}'$ are identical if and only if $\mathbf{Diff} = \emptyset$. We claim that if $\mathbf{Diff}$ is not empty, there is a legitimate mark change that eliminates a difference. Choose $B \in \mathbf{Diff}$ such that no proper ancestor of B in $\mathcal{G}$ is in $\mathbf{Diff}$. Let

$$\mathbf{Diff}_B = \{x|x \rightarrow B \text{ is in } \mathcal{G} \text{ and } x \leftrightarrow B \text{ is in } \mathcal{G}'\}$$

Since $B \in \mathbf{Diff}$, $\mathbf{Diff}_B$ is not empty. Choose $A \in \mathbf{Diff}_B$ such that no proper descendant of A in $\mathcal{G}$ is in $\mathbf{Diff}_B$. The claim is that changing $A \rightarrow B$ to $A \leftrightarrow B$ in $\mathcal{G}$ is a legitimate mark change.

To see this is so, let us verify the conditions stated in Lemma 1. First, suppose condition (t1) is violated, that is, suppose there is another directed path $d = \langle A, \cdots, C, B\rangle$ from A to B besides $A \rightarrow B$. d is not present in $\mathcal{G}'$, otherwise $\mathcal{G}'$ is not a MAG due to the presence of $A \leftrightarrow B$. So some edge on d in $\mathcal{G}'$ must be bi-directed. If the edge is $C \leftrightarrow B$, then C belongs to $\mathbf{Diff}_B$, but is a proper descendant of A in $\mathcal{G}$, which contradicts our choice of A. If the edge is between another pair of vertices, say $D \leftrightarrow E$ (s.t. $D \rightarrow E$ is in $\mathcal{G}$), then E is in $\mathbf{Diff}$, but is a proper ancestor of B, which contradicts our choice of B. So can't be any directed path from A to B in $\mathcal{G}$ other than $A \rightarrow B$. Condition (t1) stands.

Next we check condition (t2). For the first part, let C be any parent of A in $\mathcal{G}$. C must also be a parent of A in $\mathcal{G}'$, otherwise A is in $\mathbf{Diff}$, but is a proper ancestor of B in $\mathcal{G}$, which contradicts our choice of B. It follows that C and B are adjacent, for otherwise $\langle C, A, B\rangle$ is an unshielded collider in $\mathcal{G}'$ but not in $\mathcal{G}$, contrary to the assumption that they are Markov equivalent. Then C must be a parent of B in $\mathcal{G}$, otherwise $\mathcal{G}$ is not ancestral.

For the second part, let D be any spouse of A (i.e., $D \leftrightarrow A$) in $\mathcal{G}$. D is also a spouse of A in $\mathcal{G}'$ by our assumption. It follows that D and B are adjacent, for otherwise $\langle D, A, B\rangle$ is an unshielded collider in $\mathcal{G}'$ but not in $\mathcal{G}$. But D cannot be a child of B in $\mathcal{G}$, for otherwise $\mathcal{G}$ is not ancestral. Hence D is either a parent or a spouse of B.

Finally, suppose condition (t3) is violated, that is, suppose there is a discriminating path $u = \langle U, \cdots, V, A, B\rangle$ for A. By the definition of discriminating path, V is a parent of B. It follows that the edge between A and V is not $A \rightarrow V$, for otherwise $A \rightarrow V \rightarrow B$ would be a directed path from A to B, which has been shown to be absent. Hence the edge between V and A must be bi-directed, $V \leftrightarrow A$. Further note that by our assumption about the difference between $\mathcal{G}$ and $\mathcal{G}'$, every arrowhead in $\mathcal{G}$ is also in $\mathcal{G}'$, which implies that every collider in $\mathcal{G}$ is also in $\mathcal{G}'$. In particular, every vertex between U and A on u is also a collider on u in $\mathcal{G}'$.

Now we prove by induction that every vertex between U and A on u, including A, is a parent of B in $\mathcal{G}'$, contradicting the fact that $A \leftrightarrow B$ is in $\mathcal{G}'$. Let W be the vertex next to U on u. Since U and B are not adjacent by the definition of discriminating path, $\langle U, W, B\rangle$ is an unshielded non-collider in $\mathcal{G}$ (because W is a parent of B in $\mathcal{G}$ by the definition of discriminating path). Because $\mathcal{G}$ and $\mathcal{G}'$ are Markov equivalent, $\langle U, W, B\rangle$ should also be a non-collider in $\mathcal{G}'$. But W is a collider on u in $\mathcal{G}$, and hence also a collider in $\mathcal{G}'$, which means the edge between U and W is into W. Thus $W \rightarrow B$ is in $\mathcal{G}'$, otherwise $\langle U, W, B\rangle$ would be an unshielded collider in $\mathcal{G}'$. This establishes the base case. In the inductive step, suppose the first n vertices after U on u are all parents of B in $\mathcal{G}'$, then we have an discriminating path for the $n + 1$'st vertex between D and B in both graphs. Since the two graphs are Markov equivalent, the $n + 1$'st vertex must be a parent of B as well, otherwise (e3) in Proposition 1 would be violated. This finishes our induction. So, in particular, A should be a parent of B in $\mathcal{G}'$, a contradiction. Thus condition (t3) also obtains.

Therefore, we can always identify a legitimate mark change to eliminate a difference as long as $\mathcal{G}$ and $\mathcal{G}'$ are still different. An induction on the number of differences between $\mathcal{G}$ and $\mathcal{G}'$ would do to complete the argument. $\square$

Obviously a DMAG and any of its LEGs satisfy the antecedent of Theorem 1, so they can be transformed to each other by a sequence of legitimate mark changes. Steps 0-2, in Figure 3, for example, portraits a stepwise transformation from G1 to H1.

Theorem 2. *Let $\mathcal{G}$ and $\mathcal{G}'$ be two Markov equivalent MAGs. If every bi-directed edge in $\mathcal{G}$ and every bi-directed edge in $\mathcal{G}'$ are invariant, then there is a sequence of legitimate mark changes that transforms one to the other.*

Proof. Without loss of generality, we prove that there is a transformation from $\mathcal{G}$ to $\mathcal{G}'$. It follows from the assumption that $\mathcal{G}$ and $\mathcal{G}'$ have the same set of bi-directed edges, and hence all differences between $\mathcal{G}$ and $\mathcal{G}'$ are of the same sort: $\rightarrow$ is in $\mathcal{G}$, while $\leftarrow$ is in $\mathcal{G}'$. Let

$$\mathbf{Diff} = \{y|\text{ there is a } x \text{ such that } x \rightarrow y \text{ is in } \mathcal{G} \text{ and } x \leftarrow y \text{ is in } \mathcal{G}'\}$$

Clearly $\mathcal{G}$ and $\mathcal{G}'$ are identical if and only if $\mathbf{Diff} = \emptyset$. We claim that if $\mathbf{Diff}$ is not empty, we can always identify a legitimate edge reversal (that is, two legitimate mark changes in a row) that eliminates a difference in direction.

Suppose $\mathbf{Diff}$ is not empty. We can choose a vertex $B \in \mathbf{Diff}$ such that no proper ancestor of B in $\mathcal{G}$ is in $\mathbf{Diff}$. Let

$$\mathbf{Diff}_B = \{x|x \rightarrow B \text{ is in } \mathcal{G} \text{ and } x \leftarrow B \text{ is in } \mathcal{G}'\}$$

Since $B \in \mathbf{Diff}$, $\mathbf{Diff}_B$ is not empty. Choose $A \in \mathbf{Diff}_B$ such that no proper descendant of A in $\mathcal{G}$ is in $\mathbf{Diff}_B$. Then changing $A \rightarrow B$ to $A \leftarrow B$ in $\mathcal{G}$ is a legitimate edge reversal.

To justify this claim, we verify the conditions in Lemma 2. Note that $A \rightarrow B$, by our choice, is a reversible edge in $\mathcal{G}$ (for $A \leftarrow B$ is in $\mathcal{G}'$, which is Markov equivalent to $\mathcal{G}$). It thus follows directly from Proposition 3 (and the assumption about bi-directed edges in $\mathcal{G}$) that $\mathbf{Sp}_{\mathcal{G}}(B) = \mathbf{Sp}_{\mathcal{G}}(A)$.

The argument for $\mathbf{Pa}_{\mathcal{G}}(B) = \mathbf{Pa}_{\mathcal{G}}(A) \cup \{A\}$ is virtually the same as Chickering's proof for DAGs. For any parent C of A in $\mathcal{G}$, C is also a parent of A in $\mathcal{G}'$, otherwise A is in $\mathbf{Diff}$ and is a proper ancestor of B in $\mathcal{G}$, which contradicts our choice of B. It follows that C is adjacent to B, otherwise $\langle C, A, B\rangle$ is an unshielded collider in $\mathcal{G}'$ but not one in $\mathcal{G}$, which would contradict the Markov equivalence between $\mathcal{G}$ and $\mathcal{G}'$.

Then C must be a parent of B in $\mathcal{G}$, otherwise $\mathcal{G}$ is not ancestral. Conversely, let $D \neq A$ be any other parent of B in $\mathcal{G}$. D must be adjacent to A, otherwise $\langle D, B, A\rangle$ is an unshielded collider in $\mathcal{G}$ but not one in $\mathcal{G}'$. D is not a spouse of A in $\mathcal{G}$, otherwise D is a spouse of B, according to what we just showed. So D is either a parent or a child of A in $\mathcal{G}$. Suppose it is a child of A, that is, $A \rightarrow D \rightarrow B$ is in $\mathcal{G}$. We derive a contradiction from this. Since $A \leftarrow B$ is in $\mathcal{G}'$, $A \rightarrow D \rightarrow B$ does not appear in $\mathcal{G}'$. That means either $A \leftarrow D$ or $D \leftarrow B$ (or both) is in $\mathcal{G}'$. In the former case, D is in **Diff** and is a proper ancestor of B in $\mathcal{G}$, which contradicts our choice of B. In the latter case, D is in $\mathbf{Diff}_B$ and is a proper descendant of A in $\mathcal{G}$, which contradicts our choice of A. Hence D can't be a child of A in $\mathcal{G}$, which means it is a parent of A in $\mathcal{G}$.

Note that after an edge reversal, no new bi-directed edge is introduced, so the assumption that every bi-directed edge is invariant still holds for the new graph. Hence we can always identify a legitimate edge reversal to eliminate a difference in direction as long as $\mathcal{G}$ and $\mathcal{G}'$ are still different. An easy induction on the number of differences between $\mathcal{G}$ and $\mathcal{G}'$ would do to complete the argument. $\square$

Since a LEG (of any MAG) only contains invariant bi-directed edges, two LEGs can always be transformed to each other via a sequence of legitimate mark changes according to the above theorem. For example, steps 2-4 in Figure 3 constitute a transformation from H1 (a LEG of G1) to H2 (a LEG of G2).

We are ready to prove the main result of this paper.

Theorem 3. *Two DMAGs $\mathcal{G}$ and $\mathcal{G}'$ are Markov equivalent if and only if there exists a sequence of single mark changes in $\mathcal{G}$ such that*

1. *after each mark change, the resulting graph is also a DMAG and is Markov equivalent to $\mathcal{G}$;*

2. *after all the mark changes, the resulting graph is $\mathcal{G}'$.*

Proof: The "if" part is trivial – since every mark change preserves the equivalence, the end is of course Markov equivalent to the beginning. Now suppose $\mathcal{G}$ and $\mathcal{G}'$ are equivalent. We show that there exists such a sequence of transformation. By

Proposition 2, there is a LEG $\mathcal{H}$ for $\mathcal{G}$ and a LEG $\mathcal{H}'$ for $\mathcal{G}'$. By Theorem 1, there is a sequence of legitimate mark changes s_1 that transforms $\mathcal{G}$ to $\mathcal{H}$, and there is a sequence of legitimate mark changes s_3 that transforms $\mathcal{H}'$ to $\mathcal{G}'$. By Theorem 2, there is a sequence of legitimate mark changes s_2 that transforms $\mathcal{H}$ to $\mathcal{H}'$. Concatenating s_1, s_2 and s_3 yields a sequence of legitimate mark changes that transforms $\mathcal{G}$ to $\mathcal{G}'$. $\square$

As a simple illustration, Figure 3 gives the steps in transforming G1 to G2 according to Theorem 3. That is, G1 is first transformed to one of its LEGs, H1; H1 is then transformed to H2, a LEG of G2. Lastly, H2 is transformed to G2.

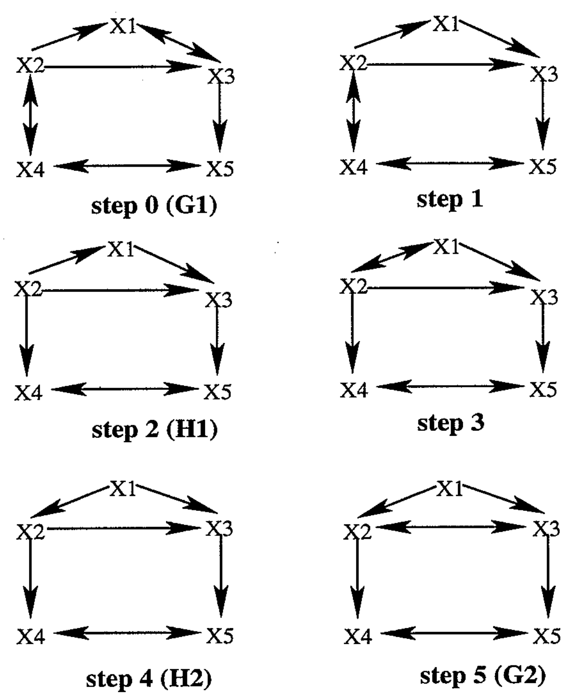

Figure 3: A transformation from G1 to G2

Theorems 1 and 2, as they are currently stated, are special cases of Theorem 3, but the proofs of them actually achieve a little more than they claim. The transformations constructed in the proofs of Theorems 1 and 2 are efficient in the sense that every mark change in the transformation eliminates a difference between the current DMAG and the target. So the transformations consist of as many mark changes as the number of differences at the beginning. By contrast, the transformation constructed in Theorem 3 may take some "detours", in that some mark changes in the way actually increase rather than decrease the difference between $\mathcal{G}$ and $\mathcal{G}'$. (This is not the case in Figure 3, but if, for example, we chose different LEGs for G1 or G2, there would be detours.) We believe that no such detour is really necessary, that is, there is always a transformation from $\mathcal{G}$ to $\mathcal{G}'$ consisting of as many mark changes as the number of differences between them. But we are yet unable to prove this conjecture.

3 Conclusion

In this paper we established a transformational property for Markov equivalent directed MAGs, which is a generalization of the transformational characterization of Markov equivalent DAGs given by Chickering (1995). It implies that no matter how different two Markov equivalent graphs are, there is a sequence of Markov equivalent graphs in between such that the adjacent graphs differ in only one edge. It could thus simplify derivations of invariance properties across a Markov equivalence class — in order to show two arbitrary Markov equivalent DMAGs share something in common, we only need to consider two Markov equivalent DMAGs with the minimal difference. Indeed, Chickering (1995) used his characterization to derive that Markov equivalent DAGs have the same number of parameters under the standard CPT parameterization (and hence would receive the same score under the typical penalized-likelihood type metrics). The discrete parameterization of DMAGs is currently under development[3]. We think our result will prove useful to show similar facts once the discrete parameterization is available.

The property, however, does not hold exactly for general MAGs, which may also contain undirected edges[4]. A simple counterexample is given in Figure 4. When we

[3]Drton and Richardson (2005) provide a parameterization for bi-directed graphs with binary variables, for which the problem of parameter equivalence does not arise because no two different bi-directed graphs are Markov equivalent.

[4]Undirected edges are motivated by the need to represent the presence of selection variables, features that influence which units are sampled (that are conditioned upon in sampling).

include undirected edges, the requirement of ancestral graphs is that the endpoints of undirected edges are of zero in-degree — that is, if a vertex is an endpoint of an undirected edge, then no edge is into that vertex (see Richardson and Spirtes (2002) for details). So, although the two graphs in Figure 4 are Markov equivalent MAGs, M1 cannot be transformed to M2 by a sequence of single legitimate mark changes, as adding any single arrowhead to M1 would make it non-ancestral. Therefore, for general MAGs, the transformation may have to include a stage of changing the undirected subgraph to a directed one in a wholesale manner.

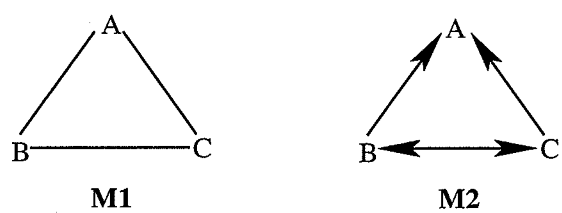

Figure 4: A simple counterexample with general MAGs: M1 can't be transformed into M2 by a sequence of legitimate single mark changes.

The transformational characterization for Markov equivalent DAGs was generalized, as a conjecture, to a transformational characterization for DAG I-maps by Meek (1996), which was later shown to be true by Chickering (2002). A graph is an I-map of another if the set of conditional independence relations entailed by the former is a subset of the conditional independence relations entailed by the latter. This generalized transformational property is used to prove the asymptotic correctness of the GES algorithm, an efficient search algorithm over the Markov equivalence classes of DAGs. The extension of both the property and the GES algorithm to MAGs is now under our investigation.

References

Ali, R.A., T. Richardson, and P. Spirtes (2004) Markov Equivalence for Ancestral Graphs. Department of Statistics, University of Washington, Tech Report 466.

Ali, R.A., T. Richardson, P. Spirtes, and J. Zhang (2005) Orientation Rules for Constructing Markov Equivalence Classes for Maximal Ancestral Graphs. Department of Statistics, University of Washington, Tech Report 476.

Andersson, S., D. Madigan, and M. Pearlman (1997) A Characterization of Markov Equivalence Classes of Acyclic Digraphs, in *Ann. Statist.* 25(2): 505-541.

Chickering, D.M. (1995) A transformational characterization of equivalent Bayesian network structures, in *Proceedings of Eleventh Conference on Uncertainty in Artificial Intelligence*, 87-98. Morgan Kaufmann.

Chickering, D.M. (2002) Optimal Structure Identification with Greedy Search, in *Journal of Machine Learning Research*, 3:507-554.

Drton, M., and T. Richardson (2004) Graphical Answers to Questions About Likelihood Inference in Gaussian Covariance Models. Department of Statistics, University of Washington, Tech Report 467.

Drton, M., and T. Richardson, T. (2005) Binary Models for Marginal Independence. Department of Statistics, University of Washington, Tech Report 474.

Meek, C. (1996) Graphical Models: Selecting Causal and Statistical Models. Carnegie Mellon University, Philosophy Department, PhD Thesis.

Richardson, T., and P. Spirtes (2002) Ancestral Markov Graphical Models, in *Ann. Statist.* 30(4): 962-1030.

Richardson, T., and P. Spirtes (2003) Causal Inference via Ancestral Graph Models, in *Highly Structured Stochastic Systems.* Eds. P. Green, N. Hjort, and S. Richardson. Oxford University Press.

Spirtes, P., and T. Richardson (1996) A Polynomial Time Algorithm For Determining DAG Equivalence in the Presence of Latent Variables and Selection Bias, in *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics.*

Spirtes, P., T. Richardson, and C. Meek (1997) Heuristic Greedy Search Algorithms for Latent Variable Models, in *Proceedings of the 6th International Workshop on Artificial Intelligence and Statistics.*

Verma, T., and J. Pearl (1990) Equivalence and Synthesis of Causal Models, in *Proceedings of 6th Conference on Uncertainty in Artificial Intelligence*, 220-227.

Zhang, J., and P. Spirtes (2005) A Characterization of Markov Equivalence Classes for Ancestral Graphical Models. Department of Philosophy, Carnegie Mellon University, Tech Report 168.