A Transformational Characterization of Markov Equivalence for Directed Maximal Ancestral Graphs

Jiji Zhang
Philosophy Department
Carnegie Mellon University
Pittsburgh, PA 15213
jiji@andrew.cmu.edu

Peter Spirtes
Philosophy Department
Carnegie Mellon University
Institute for Machine and Human Cognition
University of West Florida
ps7z@andrew.cmu.edu

Abstract

The conditional independence relations present in a data set usually admit multiple causal explanations | typically represented by directed graphs | which are Markov equivalent in that they entail the same conditional independence relations among the observed variables. Markov equivalence between directed acyclic graphs (DAGs) 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 justify a search procedure over Markov equivalence classes, known as the GES algorithm. Markov equivalence between DAGs with latent variables has also been characterized, in the spirit of Verma and Pearl (1990), via maximal ancestral graphs (MAGs). The latter can represent the observable conditional independence relations as well as some causal features of DAG models with latent variables. However, no characterization of Markov equivalent MAGs is yet available that is analogous to the transformational characterization for Markov equivalent DAGs. The main contribution of the current paper is to establish such a characterization for directed MAGs, which we expect will have similar uses as Chickering's characterization does for DAGs.

1 INTRODUCTION

Markov equivalence between directed acyclic graphs (DAGs) has been characterized in several ways (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) is useful in deriving properties shared by Markov equivalent DAGs. Moreover, when generalized to the I-map relationship, the transformational characterization warrants an e±cient search procedure over Markov equivalence classes of DAGs, known as the GES algorithm (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 the sense 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 causal relations entailed by that DAG. Since ancestral graphs do not explicitly include latent variables, they provide, among other virtues, a ¯nite 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, which we expect will have similar uses as it does for DAGs. 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.

The paper is organized as follows. The remaining of this section introduces the relevant de¯nitions 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 (*!*), bi-directed edge (*$*) and undirected edge (*¡¡*). In this paper, however, we will con¯ne 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.

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 $ B*), then *A* is called a **spouse** of *B* and *B* a spouse of *A*.

De¯nition 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 $ 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), de¯nes 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 con¯gurations than they do in DAGs. Given a path *u* in a

(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

2.1 LOYAL EQUIVALENT GRAPH

Given a MAG, a mark (or edge) therein is **invariant** if it is present in all MAGs equivalent to the given one. Invariant marks are particularly important for causal inference because in general data can only determine up to a Markov equivalence class of graphs. An algorithm of detecting all invariant arrowheads in a MAG is given by Ali et al. (2005), and that of further detecting all invariant tails is presented in Zhang and Spirtes (2005). The following is a special case of Corollary 18 in Zhang and Spirtes (2005).

Proposition 2. *Given any DMAG G, there exists a DMAG H Markov equivalent to G such that all bi-directed edges in H are invariant, and every directed edge in G is also in H.*

We will call *H* in Proposition 2 a **Loyal Equivalent Graph (LEG)** of *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 Markov equivalent DMAGs[2]. Drton and Richardson (2004) explored the statistical signi¯cance 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 ¯tting is easy; otherwise ¯tting is not easy (in a speci¯c technical sense).

Another feature which will be particularly relevant to our argument is that between a DMAG and any of its LEGs, only one kind of di®erences 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. (Note that *X*4 $ *X*5 is invariant, which is why no DAG without latent variables can represent the observable conditional independence structure entailed by Figure 1(a)).

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. To prove Theorem 2 below, we also need a fact that immediately follows from Corollary 4.1 in Ali et al. (2005).

Proposition 3. *Let A ! B be any reversible edge in a DMAG 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.*

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

[2] For general MAGs, Corollary 18 in Zhang and Spirtes (2005) also asserts that the LEGs have the fewest undirected edges as well.

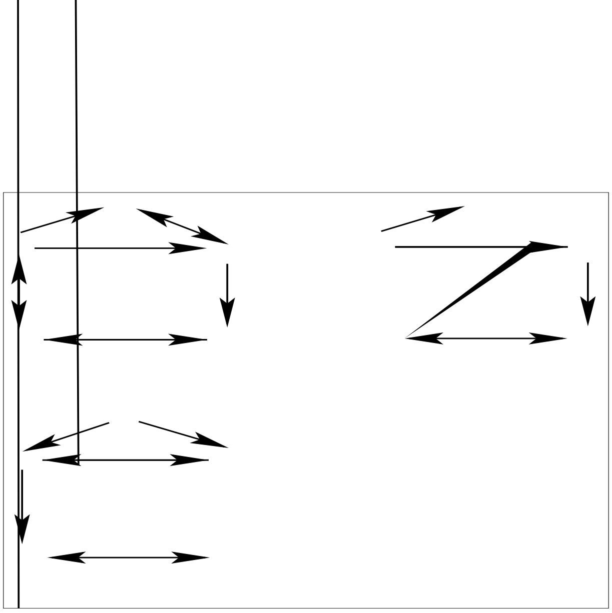

Figure 2: A LEG of G1 (H1) and a LEG of G2 (H2)

2.2 LEGITIMATE MARK CHANGE

Eventually we will show that any 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. Lemma 1 below gives necessary and su±cient 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 symis Markov equivalent to G. To see that G' is ancestral, note that it only di®ers from G, an ancestral graph, regarding the edge between A and 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 G and G' are all of the same sort: a directed edge ($\rightarrow$) is in G while the corresponding edge in G' is bi-directed ($\leftrightarrow$). Let

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

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

$$\mathrm{Diff}_B = \{x \mid x \rightarrow B \text{ is in } G \text{ and } x \leftrightarrow B \text{ is in } G'\}$$

Since $B \in \mathrm{Diff}$, Diff_B is not empty. Choose $A \in \mathrm{Diff}_B$ such that no proper descendant of A in G is in Diff_B. The claim is that changing $A \rightarrow B$ to $A \leftrightarrow B$ in G is a legitimate mark change — that is, it satisfies the conditions stated in Lemma 1.

The verifications of conditions (t1) and (t2) in Lemma 1 take advantage of the specific way by which we choose A and B. For example, if condition (t1) were violated, i.e., there were a directed path d from A to B other than $A \rightarrow B$, then in order for G' to be ancestral, d would not be directed in G', which implies that some edge on d would be bi-directed in G'. It is then easy to derive a contradiction to our choice of A or B in the first place. The verification of (t2) is similarly easy (which uses the fact that G and G' have the same unshielded colliders).

To show that (t3) also holds, suppose for contradiction that there is a discriminating path $u = \langle D, \cdots, C, A, B \rangle$ for A in G. By Definition 7, C is a parent of B. It follows that the edge between A and C is not $A \rightarrow C$, for otherwise $A \rightarrow C \rightarrow B$ would be a directed path from A to B, which has been shown to be absent. Hence the edge between C and A is bi-directed, $C \leftrightarrow A$ (because C, Definition 7, is a collider on u). Then the antecedent of the theorem implies that $C \leftrightarrow A$ is also in G'. Moreover, the antecedent implies that every arrowhead in G is also in G', which entails that in G' every vertex between D and A is still a collider on u. It is then easy to prove by induction that every vertex between D and A on u is also a parent of B in G' (using the fact that G' is Markov equivalent to G), and hence u is also discriminating for A in G' (see, e.g., Lemma 3.5 in Ali et al. 2004). But A is a collider on u in G' but not in G, which contradicts (e3) in Proposition 1. □

Obviously a DMAG and any of its LEGs satisfy the antecedent of Theorem 1, so they87(a)-63vs-9-G

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

1. *after each mark change, the resulting graph is also a DMAG and is Markov equivalent to G;*
2. *after all the mark changes, the resulting graph is 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 G and G' are equivalent. We show that there exists such a sequence of transformation. By Proposition 2, there is a LEG H for G and a LEG H' for G'. By Theorem 1, there is a sequence of legitimate mark changes s_1 that transforms G to H, and there is a sequence of legitimate mark changes s_3 that transforms H' to G'. By Theorem 2, there is a sequence of legitimate mark changes s_2 that transforms H to H'. Concatenating s_1, s_2 and s_3 yields a sequence of legitimate mark changes that transforms G to G'. □

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

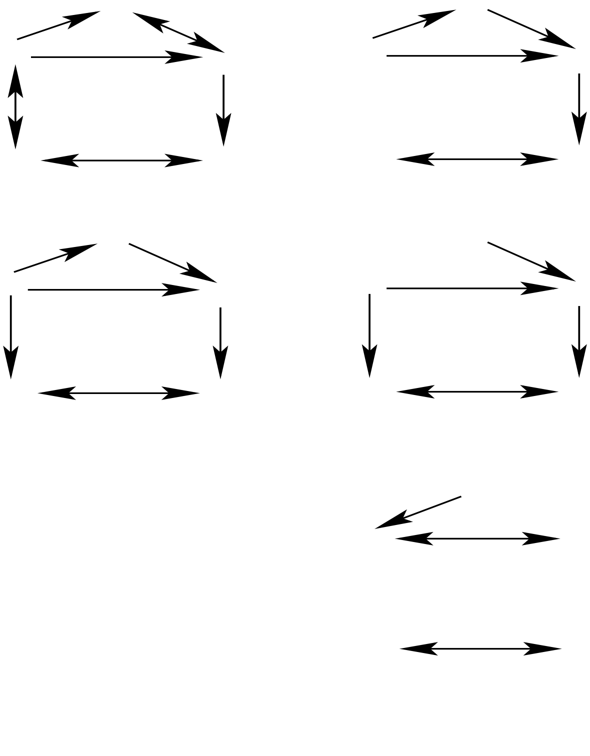

Figure 3: A transformation from G1 to G2

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 e±cient in the sense that every mark change in the transformation eliminates a di®erence between the current DMAG and the target. So the transformations consist of as many mark changes as the number of di®erences 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 di®erence between G and G'. (This is not the case in Figure 3, but if, for example, we chose di®erent 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 G to G' consisting of as many mark changes as the number of di®erences 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 di®erent two Markov equivalent graphs are, there is a sequence of Markov equivalent graphs in between such that the adjacent graphs di®er 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 di®erence. 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[4]. We expect that our result will come in handy 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[5].

[4] 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 di®erent bi-directed graphs are Markov equivalent.

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

A simple counterexample is given in Figure 4. When we 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.

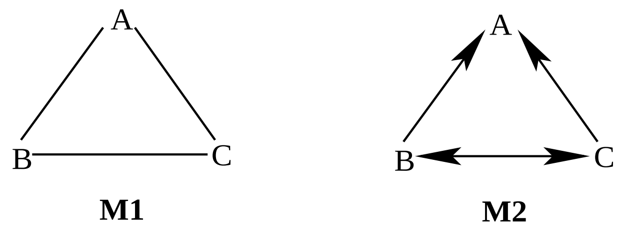

Figure 4: A simple counterexample with general MAGs: M1 can't be transformed into M2 by a sequence of legitimate single mark changes.

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 general-