Learning equivalence classes of acyclic models with latent and selection variables from multiple datasets with overlapping variables

Robert E. Tillman
Carnegie Mellon University
Pittsburgh, PA
rtillman@cmu.edu

Peter Spirtes
Carnegie Mellon University
Pittsburgh, PA
ps7z@andrew.cmu.edu

Abstract

While there has been considerable research in learning probabilistic graphical models from data for predictive and causal inference, almost all existing algorithms assume a single dataset of i.i.d. observations for all variables. For many applications, it may be impossible or impractical to obtain such datasets, but multiple datasets of i.i.d. observations for different subsets of these variables may be available. Tillman *et al.* [2009] showed how directed graphical models learned from such datasets can be integrated to construct an equivalence class of structures over all variables. While their procedure is correct, it assumes that the structures integrated do not entail contradictory conditional independences and dependences for variables in their intersections. While this assumption is reasonable asymptotically, it rarely holds in practice with finite samples due to the frequency of statistical errors. We propose a new correct procedure for learning such equivalence classes directly from the multiple datasets which avoids this problem and is thus more practically useful. Empirical results indicate our method is not only more accurate, but also faster and requires less memory.

1 INTRODUCTION

Probabilistic graphical models are widely used for predictive and causal inference. Learning the structure of these models from data has remained an active research area for several decades. Almost all existing structure learning methods, however, assume that a single dataset of i.i.d. observations for all variables of interest is available. For many applications, such datasets may be impossible or impractical to obtain. For instance, fMRI researchers are often unable to obtain reliable measurements of activation in every brain region of interest each time an fMRI machine is used with a particular individual due to differences across individuals, fMRI machine settings, and other random factors. Instead, such researchers often obtain multiple datasets which each measure different subsets of these brain regions, e.g. a researcher interested in brain regions X, Y, and Z may obtain i.i.d. observations for three different individuals resulting in three datasets with useful recordings only for regions X and Y, Y and Z, and X and Z, respectively. Data are structured similarly in other domains, e.g. econometric models of the U.S. and U.K. economies share some but not all variables due to differences in financial recording conventions and U.S. states report both federal and state-specific educational testing variables [Tillman *et al.*, 2009]. We will say that a set of datasets has *overlapping variables* when data are structured in this way.

While the above problem is superficially similar to learning structure from a dataset where individual cell values are missing at random, e.g. questionnaire data where some individuals randomly skip questions, it is a fundamentally different (and much harder) type of missing data problem: if we concatenate datasets with overlapping variables to form a single dataset with missing cells, the cells which are missing do not occur at random; instead, there is a highly structured pattern to the missing data and certain subsets of variables are never jointly recorded in the concatenated dataset. Tillman *et al.* [2009] showed that the state of the art Structural EM algorithm [Friedman, 1998] for learning directed graphical models from datasets with cell values missing at random is unsuccessful (in terms of both accuracy and computational tractabil-

Appearing in Proceedings of the 14^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2011, Fort Lauderdale, FL, USA. Volume 15 of JMLR: W&CP 15. Copyright 2011 by the authors.

ity) when used with a dataset formed by concatenating multiple datasets with overlapping variables. Thus, a fundamentally different approach is needed.

Tillman *et al.* [2009] proposed the first such approach, the integration of overlapping networks (ION) algorithm, based on the observations in Danks [2005]. ION learns an equivalence class of directed acyclic structures from multiple datasets with overlapping variables under the assumption that the datasets correspond to the same data generating process and thus do not entail contradictory conditional independences and dependences given perfect statistical information. While this is a reasonable theoretical assumption (which we also make in the procedure described later in this paper), the way it is incorporated into how ION combines information from different datasets leads to many practical problems, which result from the fact that statistical information is never perfect. ION assumes that equivalence classes of structures over the variables recorded in each dataset can first be obtained using a structure learning algorithm which detects possible latent confounders (unobserved variables that may be common causes of two or more observed variables) such as FCI [Spirtes *et al.*, 1999]. The ION algorithm then takes these structures as input and, through a series of graph theoretic operations, attempts to learn an equivalence class of possible structures which may correspond to the true data generating process for the union of variables measured in each dataset. However, since each equivalence class in the input set is learned independently using a different dataset, it is often the case that different equivalence classes in the input set entail contrary conditional independences and dependences due to statistical errors (even when the datasets correspond to the same data generating process). When ION encounters contradictory input, it may either produce inaccurate results or (frequently) no results since it cannot find any structures which are consistent with the entire input set. Another serious practical limitation of ION is that it often requires significant memory resources and computation time even for small numbers of variables.

In this paper, we present a new method for learning equivalence classes of directed graphical models from multiple datasets with overlapping variables that is asymptotically correct and complete, effectively deals with the above problem of contrary statistical information from different datasets, and requires significantly less memory and computation time than ION. Rather than learning structures for each dataset independently and attempting to integrate the results, we learn equivalence classes directly from the multiple datasets. This allows us to avoid the problem of contradictory input and also results in a more robust learning procedure since we can often use more than one dataset at once when performing any statistical test, which results in a more accurate test. Section 2 provides necessary background material; section 3 presents new graph theoretic results which our learning algorithm relies on; section 4 describes the method for testing conditional independence using multiple datasets with overlapping variables that our learning algorithm uses; section 5 describes and discusses the learning procedure; section 6 presents empirical results with real and artificial data; finally section 7 discusses conclusions and future research.

2 BACKGROUND

A *mixed graph* $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a graph consisting of edges in $\mathcal{E}$ which connect distinct nodes in $\mathcal{V}$. Edges in mixed graphs can be of three types: (i) directed ($\rightarrow$), (ii) undirected ($-$), and bidirected ($\leftrightarrow$). An edge between two nodes X and Y *points toward* Y if it has an arrowhead at Y. Nodes connected by an edge are *adjacent*. An *edge sequence* is any ordered set of nodes $\langle V_1, \ldots, V_n \rangle$ such for $1 \leq i < n$, V_i and V_{i+1} are adjacent. If each node in an edge sequence occurs exactly once then it is a *path*. An edge sequence or path is *directed* if each edge between consecutive nodes points in the same direction. X is an *ancestor* of Y and Y is a *descendant* of X if there is a directed path which points from X to Y. If such a path consists of only X and Y, then X is a *parent* of Y and Y is a *child* of X. X and Y are *spouses* if they are connected by a bidirected edged. $\mathbf{Adj}_{\mathcal{G}}^{X}$, $\mathbf{Pa}_{\mathcal{G}}^{X}$, $\mathbf{Ch}_{\mathcal{G}}^{X}$, $\mathbf{An}_{\mathcal{G}}^{X}$, $\mathbf{De}_{\mathcal{G}}^{X}$, and $\mathbf{Sp}_{\mathcal{G}}^{X}$ refer to the set of adjacencies, parents, children, ancestors, descendants, and spouses of X in $\mathcal{G}$, respectively. $\mathcal{G}$ contains a *directed cycle* between two nodes X and Y if $X \rightarrow Y$ is in $\mathcal{E}$ and $Y \in \mathbf{An}_{\mathcal{G}}^{X}$. $\mathcal{G}$ contains an *almost directed cycle* between two nodes X and Y if $X \leftrightarrow Y$ is in $\mathcal{E}$ and $Y \in \mathbf{An}_{\mathcal{G}}^{X}$.

A path $\langle X, Z, Y \rangle$ is a *v-structure (collider)* if both edges along the path point towards Z, e.g. $X \rightarrow Z \leftarrow Y$, $X \leftrightarrow Z \leftarrow Y$. Such a path is *unshielded* if X and Y are not adjacent. A path is an *immorality* if it is unshielded and a v-structure. A *trek* between X and Y is a path with no v-structures. A path $\langle V_1, \ldots, V_n \rangle$ is an *inducing path* between V_1 and V_n relative to $\mathbf{Z} \subseteq \mathcal{V}$ if for $1 < i < n$, (i) if $V_i \notin \mathbf{Z}$, then $\langle V_{i-1}, V_i, V_{i+1} \rangle$ is a v-structure, and (ii) if $\langle V_{i-1}, V_i, V_{i+1} \rangle$ is a v-structure, then $V_i \in \mathbf{An}_{\mathcal{G}}^{V_1} \cup \mathbf{An}_{\mathcal{G}}^{V_n}$. A path $\langle V_1, \ldots, V_{n-2}, V_{n-1}, V_n \rangle$ in $\mathcal{G}$ is *discriminating* for V_{n-1} and $\langle V_{n-2}, V_{n-1}, V_n \rangle$ is a *discriminated triple* if (i) for $1 < i < n-1$, $\langle V_{i-1}, V_i, V_{i+1} \rangle$ is a v-structure and $V_i \in \mathbf{Pa}_{\mathcal{G}}^{V_n}$ and (ii) V_1 and V_n are not adjacent.

A *maximal ancestral graph (MAG)* is a mixed graph that is useful for modeling systems with possible latent confounders. They are a natural extension of directed acyclic graphs (DAGs), which are simply special cases of MAGs where all edges are directed. Bidirected edges in MAGs indicate that the corresponding nodes have an unobserved common cause, while undirected edges indicate that the corresponding nodes have an association due to the presence of sample selection bias.[1] Below, we define MAGs formally.

Definition 2.1 (Maximal ancestral graph (MAG))**.** A mixed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is a *maximal ancestral graph (MAG)* if
(i) $\mathcal{G}$ does not contain any directed cycles or almost directed cycles and for any undirected edge $X - Y$ in $\mathcal{G}$, X and Y have no parents or spouses ($\mathcal{G}$ is *ancestral*)
(ii) For any distinct nodes $V_i, V_j \in \mathcal{V}$, if V_i and V_j are not adjacent in $\mathcal{G}$, then $\mathcal{G}$ contains no inducing paths between V_i and V_j with respect to $\varnothing$ ($\mathcal{G}$ is *maximal*).

The first condition used to define MAGs simply extends the acyclicity property of DAGs. The second ensures that MAGs have a separation criteria that connects their topology to individual conditional independence relations in the same way that the well known *d-separation* criterion [Verma and Pearl, 1991] is used to connect DAG topology to such relations.

Definition 2.2 (m-separation)**.** A path $\langle V_1, \ldots, V_m \rangle$ in a MAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ is *active (m-connected)* relative to $\mathbf{Z} \subseteq \mathcal{V} \backslash \{V_1, V_n\}$ if for $1 < i < n$
(i) if $V_i \in \mathbf{Z}$, then $\langle V_{i-1}, V_i, V_{i+1} \rangle$ is a v-structure
(ii) if $\langle V_{i-1}, V_i, V_{i+1} \rangle$ is a v-structure, then $\{V_i\} \cup \mathbf{De}_{\mathcal{G}}^{V_i}$ and $\mathbf{Z}$ have nonempty intersection.
If there are no active paths between nodes X and Y relative to $\mathbf{Z}$ in $\mathcal{G}$, then X and Y are *m-separated* relative to $\mathbf{Z}$, denoted $msep_{\mathcal{G}}(X, Y|\mathbf{Z})$. If X and Y are not m-separated with respect to $\mathbf{Z}$, this is denoted $\neg msep_{\mathcal{G}}(X, Y|\mathbf{Z})$.

If a MAG consists of only directed edges (it is a DAG), then m-separation reduces to d-separation. A MAG essentially represents the collection of DAGs over the observed and unobserved variables represented by the MAG which have the same d-separation and ancestral relations among the observed variables [Zhang, 2008].

Most structure learning algorithms assume that the *Markov* and *faithfulness* conditions hold (see Spirtes *et al.* [2000] for a discussion and justifications). When these conditions hold and the true data generating mechanism can be represented as a DAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$, then d-separation in $\mathcal{G}$ is equivalent to conditional independence in an associated joint probability distribution over variables corresponding to the nodes in $\mathcal{G}$. Thus, if we are only able to observe a subset of these variables $\mathcal{W}$, conditional independence among these variables is equivalent to m-separation in a MAG $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ which respects the d-separation and ancestral relations in $\mathcal{G}$. In general, however, there will be more than one MAG for which conditional independence in an associated joint probability distribution is equivalent to m-separation. Two MAGs which have exactly the same sets m-separation relations are said to be *Markov equivalent*. This occurs if and only if two MAGs share the same adjacencies, immoralities, and discriminated triples which are v-structures [Zhang, 2007]. The complete set of MAGs which have exactly the same sets m-separation relations is a *Markov equivalence class* of MAGs.

This leads to another type of structure which is used to represent Markov equivalence classes of MAGs, referred to as a *partial ancestral graph (PAG)* [Richardson and Spirtes, 2002; Zhang, 2007]. PAGs are mixed graphs with a third type of endpoint, ∘. Whenever an edge has a ∘-marked endpoint this indicates that there is at least one MAG in the Markov equivalence class that has an arrowhead at that endpoint and at least one such MAG that has a tail at that endpoint [Zhang, 2007]. PAGs are analogous to *PDAGs* or *patterns* which are used to represent Markov equivalence classes of DAGs.

A number of structure learning algorithms directly use the results of conditional independence tests (or similarly likelihood-based scoring) to derive the set of (Markov equivalent) graphical structures which could have generated some observed data, subject to assumptions. The FCI algorithm [Spirtes *et al.*, 1999; Zhang, 2007] is a structure learning algorithm for single datasets which detects possible latent confounders and sample selection bias using conditional independence tests. It outputs a PAG representing the complete Markov equivalence class of MAGs over the observed variables. The ION algorithm for multiple datasets with overlapping variables also detects such information (with respect to the complete variable set). However, it outputs a *set* of PAGs since observing only subsets of the complete variable set in a given dataset further underdetermines the true data generating process. When some variables are never jointly observed in a single dataset, some possible conditional independences or dependences are unknown. Thus, the equivalence class may consist of multiple Markov equivalence classes, which requires multiple PAGs to represent.

3 MARGINAL MAG PROPERTIES

An attractive property of MAGs is that they are closed under marginalization and conditioning [Richardson

[1] See [Zhang, 2007] for a example.

and Spirtes, 2002]. If $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ and $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ are both MAGs where $\mathcal{W} \subset \mathcal{V}$ and $msep_{\mathcal{H}}(X, Y|\mathbf{Z}) \Leftrightarrow msep_{\mathcal{G}}(X, Y|\mathbf{Z})$ for all $X, Y \in \mathcal{W}$ and $\mathbf{Z} \subseteq \mathcal{W}\backslash\{X, Y\}$, then $\mathcal{H}$ is a *marginal MAG* with respect to $\mathcal{G}$, i.e. $\mathcal{H}$ is the graph which results after marginalizing the variables $\mathcal{V}\backslash\mathcal{W}$ from $\mathcal{G}$. Below, we provide a correct procedure for marginalizing variables from MAGs, which restates results from Richardson and Spirtes [2002].

Theorem 3.1 (Richardson and Spirtes [2002] Definition 4.2.1, Theorem 4.12, Corollary 4.19, Theorem 4.18)**.** Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a MAG and let $\mathcal{W} \subset \mathcal{V}$. Then the marginal MAG $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ with respect to $\mathcal{G}$ is formed by applying the following steps:
(i) Make $\mathcal{H}$ an undirected graph such that for $X, Y \in \mathcal{W}$, X and Y are adjacent in $\mathcal{H}$ if and only if $\forall \mathbf{Z} \subseteq \mathcal{W}\backslash\{X, Y\}$, there is an active path between X and Y with respect to $\mathbf{Z}$ in $\mathcal{G}$
(ii) For each undirected edge X — Y in $\mathcal{F}$, if there does not exist a trek $\pi = \langle X, \ldots, Y \rangle$ in $\mathcal{G}$ such that no edges along π are bidirected and either (a) all directed edges along π towards X or (b) all directed edges along π point towards Y, make X — Y bidirected in $\mathcal{H}$
(iii) For each undirected edge X — Y in $\mathcal{F}$, if there exists a trek $\pi = \langle X, \ldots, Y \rangle$ in $\mathcal{G}$ such that no edges along π are bidirected and all directed edges along π point toward X, and there does not exist a trek $\pi' = \langle X, \ldots, Y \rangle$ such that no edges along π' are bidirected and all directed edges along π' point toward Y, make X — Y directed towards X in $\mathcal{H}$.

The goal of the learning algorithm for multiple datasets with overlapping variables described in section 5 is essentially to find all MAGs whose marginal MAGs for each set of variables observed in a dataset entail the same conditional independences and dependences observed in the data for those variables. In order to avoid repetitively going through the marginalization procedure above, we give two conditions below which can be used to check, for a MAG $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ and a MAG $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ where $\mathcal{W} \subset \mathcal{V}$, whether $\mathcal{H}$ is a marginal MAG with respect to $\mathcal{G}$.

Theorem 3.2. Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ and $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ be MAGs where $\mathcal{W} \subset \mathcal{V}$. $\mathcal{H}$ is Markov equivalent to the marginal MAG with respect to $\mathcal{G}$ if the following hold for all $X, Y \in \mathcal{W}$.
(i) if X and Y are adjacent in $\mathcal{H}$, then $\mathcal{G}$ has an inducing path between X and Y with respect to $\mathcal{V}\backslash\mathcal{W}$
(ii) $msep_{\mathcal{H}}(X, Y|\mathbf{Z}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{Z})$ for every $\mathbf{Z} \subseteq \mathcal{W}\backslash\{X, Y\}$,

The proof of theorem 3.2 is included in the appendix.

4 ROBUST CONDITIONAL INDEPENDENCE

Let $X \perp\!\!\!\perp_D Y|\mathbf{Z}$ indicate that X and Y were found to be conditionally independent given $\mathbf{Z}$ using some conditional independence test with a dataset D. As noted in section 1, a major shortcoming of ION is that each dataset is dealt with independently which often leads to contradictory statistical information. In order to have a more robust test for conditional independence, we would like to deal with all of the data at once and use every dataset containing the relevant variables for a given conditional independence test whenever we test for a particular conditional independence. A simplistic way of doing this is to concatenate the values for the relevant variables from each such dataset and use a standard conditional independence test for single datasets. This, however, can lead to a number of well known statistical problems and often does not work well in practice, as shown in Tillman [2009]. Another well studied solution, which often works much better in practice, involves using metaanalysis methods. Tillman [2009] used these methods for structure learning with multiple datasets which all contained the same variable sets and found that among well known metaanalysis methods, Fisher's method [Fisher, 1950] performed best (see Tillman [2009] or Fisher [1950]) for details). We will thus use this method for testing conditional independence in the learning algorithm described in section 5. This method assumes we have an appropriate conditional independence test that can be used with single datasets. Let $\mathbf{D} = \langle D_1, \ldots, D_n \rangle$ be multiple datasets with overlapping variables over variable sets $\langle \mathcal{V}_1, \ldots, \mathcal{V}_n \rangle$, respectively, and $\mathcal{V} = \bigcup_{i=1}^{n} \mathcal{V}_i$. Algorithm 1 shows how this method, as described in Tillman [2009], can be adapted for the overlapping variables case to test whether X and Y are conditionally independent given $\mathbf{Z}$ for any $X, Y \in \mathcal{V}$ and $\mathbf{Z} \subseteq \mathcal{V}\backslash\{X, Y\}$ at a significance level α. We will use $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{Z}$ to indicate conditional independence using this method (algorithm 1 returns `true`) and $X \perp\!\!\!\perp_{D_i} Y|\mathbf{Z}$ to indicate conditional independence using the specified test for a single datasets D_i. In algorithm 1, $F^{-1}_{\chi^2_k}$ denotes the χ^2 quantile function with k degrees of freedom. Note that algorithm 1 assumes that X, Y, and $\mathbf{Z}$ are jointly measured in at least one dataset (since the χ^2 quantile function is undefined for 0 degrees of freedom).

5 INTEGRATION OF OVERLAPPING *DATASETS*

In this section, we describe our learning algorithm for multiple datasets with overlapping variables, Integration of Overlapping *Datasets* (IOD). IOD takes

```
Input  : D, $X, Y, \mathbf{Z}$, $\alpha$
Output: true or false
$k \leftarrow 0$
foreach $D_i \in \mathbf{D}$ do
    if $\{X, Y\} \cup \mathbf{Z} \subseteq \mathcal{V}_i$ then
        $p_i \leftarrow p$-value associated with $X \perp\!\!\!\perp_{D_i} Y | \mathbf{Z}$
        $k \leftarrow k + 1$
    else
        $p_i \leftarrow 1$
    end
end
if $-2 \sum_{i=1}^{n} \log(p_i) < F^{-1}_{\chi^2_{2k}}(1 - \alpha)$ then return true
else return false
```

Algorithm 1: Test Independence

multiple datasets as input and outputs a complete equivalence class of PAGs over the union of variables measured in each dataset representing MAGs that may correspond (indistinguishably, based on conditional independence information) to the true data generating mechanism responsible for each dataset. Let $\mathbf{D} = \langle D_1, \ldots, D_n \rangle$ be datasets with overlapping variables over variable sets $\langle \mathcal{V}_1, \ldots, \mathcal{V}_n \rangle$, respectively, and let $\mathcal{V} = \bigcup_{i=1}^{n} \mathcal{V}_i$. We will assume each dataset corresponds to the same data generating process and thus, given perfect statistical information, should not entail contrary conditional independences and dependences for common sets of variables. We will also assume that the Markov and faithfulness conditions hold and the data generating process is acyclic, but may contain latent and selection variables, provided that a variable is a selection variable if and only if it is a selection variable in the true data generating process with respect to $\mathcal{V}$ and all variable sets $\langle \mathcal{V}_1, \ldots, \mathcal{V}_n \rangle$ which contain the variable. Proofs are included in the appendix.

The first stage of the algorithm, shown as algorithm 2, obtains the independence facts and inducing paths, in data structures **Sepset** and **IP**, respectively, necessary to build the equivalence class that will be output by exploiting theorem 3.1 as well as partially oriented mixed graphs $\mathcal{G}$ and $\mathcal{G}_1, \ldots, \mathcal{G}_n$ which are used in the second stage of the algorithm. The independence facts recorded in **Sepset** are obtained by applying the initial steps of FCI to the variable sets $\mathcal{V}_1, \ldots, \mathcal{V}_n$ and using algorithm 1 with **D** for independence testing. In order to obtain the necessary independence facts, algorithm 2 accesses a set **PossSep**$(\{X, Y\}, \mathcal{H})$ for a given graph $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$. This set is constructed by considering each $Z \in \mathcal{W} \backslash \{X, Y\}$ and path $\langle V_1, \ldots, V_m \rangle$ in $\mathcal{H}$ such that $V_1 = X$, $V_m = Y$ and for some $1 < j < m$, $V_j = Z$. If in such a path, either (i) for $1 < k < j$, either V_{k-1} and V_{k+1} are adjacent or $\langle V_{k-1}, V_k, V_{k+1} \rangle$ is a v-structure or (ii) for $j < k < m$, either V_{k-1} and V_{k+1} are adjacent or $\langle V_{k-1}, V_k, V_{k+1} \rangle$ is a v-structure, Z is added to **PossSep**$(\{X, Y\}, \mathcal{H})$. The graphs $\mathcal{G}_1, \ldots, \mathcal{G}_n$ output by algorithm 2 contain the adjacencies of the sets of MAGs which represent the independence and dependence facts of the joint distribution which generated **D** over the variable sets $\mathcal{V}_1, \ldots, \mathcal{V}_n$. These graphs are used to generate **IP** as required by theorem 3.1. $\mathcal{G}$ contains a (not necessarly proper) superset of the edges and (not necessarly proper) subset of the immoralities in any MAG over $\mathcal{V}$ which represents the independence and dependence facts of the joint distribution which generated **D**.

The second stage of the algorithm, shown as algorithm 3, uses $\mathcal{G}$, $\mathcal{G}_1, \ldots, \mathcal{G}_n$, **Sepset**, **IP**, output by algorithm 2, to construct the equivalence class of MAGs over $\mathcal{V}$ which represent the independence and dependence facts of the joint distribution which generated **D**. Algorithm 3 considers the edges that may be removed and possible immoralities and discriminated triples which may oriented in $\mathcal{G}$ to produce a graph in this equivalence class. $\mathcal{P}(\mathbf{X})$ is used to represent the powerset of **X**. Once a graph has been constructed which contains a possible set of adjaencies, immoralities, and discriminated triples which are v-structures, the complete set of orientation rules from Zhang [2007] which are necessary to convert this graph into a PAG ($\mathcal{R}1, \ldots, \mathcal{R}10$) are invoked. In order to determine whether this candidate PAG should be included in the equivalence class, it is converted to a MAG in its represented equivalence class. If this MAG satisfies the conditions of theorem 3.1, then the PAG that was previously created is added to the equivalence class.

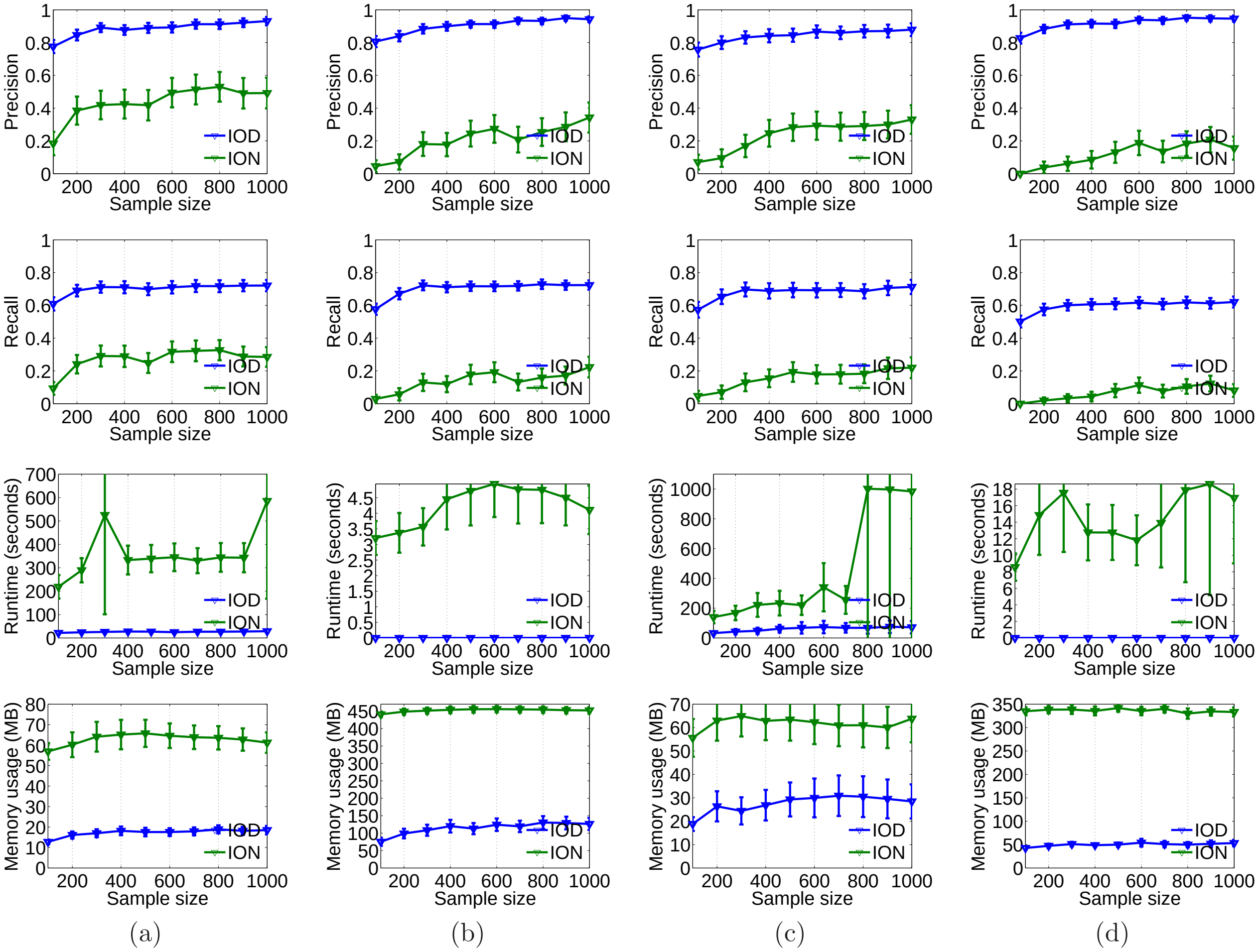
The following two theorems show that IOD is *correct*, or each MAG in the equivalence class that is output is consistent with **D**, and *complete*, or if there exists a MAG $\mathcal{H}$ that is consistent with **D**, then $\mathcal{H}$ is contained in the equivalence class that is output.

Theorem 5.1 (correctness)**.** $\forall \mathcal{J} \in \mathbf{G}$, resulting from algorithm 3, $\mathcal{J}$ is a PAG such that for any MAG $\mathcal{J}'$ in the equivalence class represented by $\mathcal{J}$, for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{Z} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{Z} \Leftrightarrow msep_{\mathcal{J}'}(X, Y | \mathbf{Z})$.

Theorem 5.2 (completeness)**.** Let $\mathcal{H}$ be a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{Z} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{Z} \Leftrightarrow msep_{\mathcal{H}}(X, Y | \mathbf{Z})$. Then $\exists \mathcal{J} \in \mathbf{G}$ such that $\mathcal{H}$ can be formed by orienting $\circ$-marked edges in $\mathcal{J}$.

6 EXPERIMENTAL RESULTS

We compared the performance of IOD to ION in simulation under 4 different scenarios: (a) 2 datasets which each measure 7 out of 8 total variables, e.g. $\{X_1, X_2,$



(a) (b) (c) (d)

Figure 1: Precision, Recall, Runtime, and Memory usage for scenarios (a), (b), (c), and (d)

X_3, X_4, X_5, X_6, $X_7\}$ and $\{X_1$, X_2, X_3, X_4, X_5, X_6, $X_8\}$, (b) 2 datasets which each measure 13 out of 14 total variables, (c) 3 datasets which each measure 6 out of 8 total variables, (d) 3 datasets which each measure 12 out of 14 total variables. For each case, we generated a random directed acyclic structure with 8 or 14 variables to represent some ground truth data generating process using the MCMC procedure in Melançon *et al.* [2000]. We then sampled from this structure using different multivariate Gaussian parameterizations for each dataset. We used the resulting datasets as input to IOD and the structures which resulted from using each of these datasets as input to FCI (separately) as input to ION. The partial correlation-based Fisher Z-transformation conditional independence test, which is correct for multivariate Gaussian distributions, was used for all single dataset conditional independence tests. We always used the significance level $\alpha = .05$. We compared the accuracy of the MAGs in the equivalence classes output by IOD and ION to the true data generating graph using two metrics: *precision*: the number of edges in a structure that are in the ground truth structure (with correct orientations) / the number of edges in that structure; *recall*: the number of edges in a structure that are in the ground truth structure (with correct orientations) / the number of edges in the ground truth structure. For each scenario, we repeated this procedure 100 times for 10 different sample sizes (of each dataset) ranging from 100 to 1000. Figure 1 shows, for each scenario, the best precision and recall scores for MAGs in the equivalences classes output by IOD and ION averaged over the 100 trials. We also measured total runtime and maximum memory usage and report these averages. Error bars indicate 95% confidence intervals. We see that in each scenario IOD significantly outperforms ION in precision and recall. Closer examination of the results reveals the relatively poor performance of ION is largely due to ION returning no results for many of the 100 trails due to statistical errors, as mentioned in section 1. We also see that IOD is noticeably faster and requires significantly less memory than ION.

We also applied IOD to real datasets with overlap-

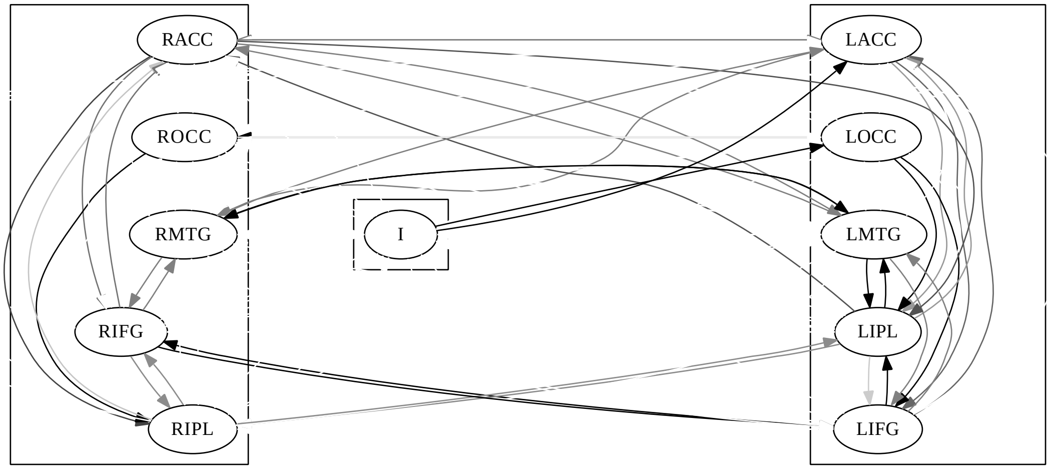


Figure 2: Edges in fMRI datasets equivalence class

ping variables from an fMRI experiment where subjects were presented pairs of words and asked whether they rhymed. Activity in the following brain regions was recorded: left occipital cortex (LOCC), left middle temporal gyrus (LMTG), left anterior cingulate (LACC), left inferior frontal gyrus (LIFG), left inferior parietal (LIPL), right occipital cortex (ROCC), right middle temporal gyrus (RMTG), right anterior cingulate (RACC), right inferior frontal gyrus (RIFG), right inferior parietal (RIPL). For each measurement, an input variable (I) indicates the presentation of rhyming or non-rhyming words. 160 measurements were obtained for 13 subjects, which resulted in 13 datasets where 1-4 variables were not recorded. Figure 2 shows every edge that is present in some MAG in the equivalence class. The darkness of edges indicates the percentage of MAGs in the equivalence class which contain the edge, where darker edges are present in more MAGs than lighter edges. While we do not have ground truth for this data, our results are consistent with domain knowledge which indicates a cascade of interactions in the left side of the brain after the presentation of the stimulus eventually leading to a cascade in the right side. Some edges observed are also consistent with the results in Ramsey *et al.* [2010], which analyzed this data using a different method.

7 CONCLUSION

Learning from multiple datasets with overlapping variables is an important problem with many practical applications that has mostly been ignored in the existing structure learning literature. As data collection increases, such data may become ubiquitous for many problems researchers are interested in. Developing methods which can take advantage of such data may prevent researchers from engaging in the costly and time consuming process of collecting new data for a specific variable set of interest and allow analysis of variable sets which may be impossible to jointly observe in a single dataset, e.g. due to privacy or ethics.

Tillman *et al.* [2009] provided the first correct procedure for addressing this problem, but their procedure often is not useful in practice since each dataset is dealt with separately and this often leads to contradictory statistical information. We introduced the IOD algorithm, which we showed is asymptotically correct and complete in section 5, to provide a new structure learning procedure for such data which can effectively deal with this problem, as is shown in our simulations. By working with as many datasets as possible whenever performing a conditional independence test, IOD not only avoids problems which result from contradictory statistical information, but also is more accurate and robust since statistical decisions are often made using more data than would be used by ION. In section 6 we also showed that IOD significantly outperforms ION not only in terms of accuracy, but also in terms of runtime and memory usage. We also showed that IOD can be successfully applied to real fMRI data.

IOD and ION both learn structure using only the conditional independences and dependences entailed by a dataset. Recently, a number of algorithms have emerged which use non-Gaussianity [Shimizu *et al.*, 2006] and nonlinearity [Hoyer *et al.*, 2009] observed in the data to learn structure. These algorithms typically produce an equivalence class that is much smaller than the Markov equivalence class, often a unique structure. Thus, a significant open problem is how to adapt such methods to use multiple datasets with overlapping variables and (hopefully) produce an equivalence

```
Input   : $\mathbf{D}$
Output: $\mathcal{G}$, $\mathcal{G}_1, \ldots, \mathcal{G}_n$, Sepset, IP
$\mathcal{G} \leftarrow \langle \mathcal{V}, \mathcal{E} \rangle$ where $X \circ\!\!-\!\!\circ Y \in \mathcal{E}$ for all $X, Y \in \mathcal{V}$
for $1 \leq i \leq n$ do
    $\mathcal{G}_i \leftarrow \langle \mathcal{V}_i, \mathcal{E}_i \rangle$ where $X \circ\!\!-\!\!\circ Y \in \mathcal{E}_i$ for all $X, Y \in \mathcal{V}_i$
    $j \leftarrow 0$
    while $j \leq |\mathcal{V}_i| - 2$ and there exists an edge
    $X \circ\!\!-\!\!\circ Y \in \mathcal{E}_i$ such that $\left|\mathbf{Adj}^X_{\mathcal{G}_i} \backslash \{Y\}\right| \geq j$ do
        foreach $X \circ\!\!-\!\!\circ Y \in \mathcal{E}_i$ do
            if $\exists \mathbf{S} \subseteq \mathbf{Adj}^X_{\mathcal{G}_i} \backslash \{Y\}$ such that $|\mathbf{S}| = j$ and
            $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S}$ then
                $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i) \leftarrow \mathbf{S}$
                Remove $X \circ\!\!-\!\!\circ Y$ from $\mathcal{G}_i$ and $\mathcal{G}$
            end
        end
        $j \leftarrow j + 1$
    end
    foreach unshielded path $\langle X, Z, Y \rangle$ in $\mathcal{G}_i$ do
        if $Z \notin \mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ then make
        $\langle X, Z, Y \rangle$ an immorality in $\mathcal{G}_i$
    end
    foreach adjacent $X, Y$ in $\mathcal{G}_i$ do
        if $\exists \mathbf{S} \subseteq \mathbf{PossSep}(\{X, Y\}, \mathcal{G}_i)$ such that
        $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S}$ then
            $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i) \leftarrow \mathbf{S}$
            Remove $X \circ\!\!-\!\!\circ Y$ from $\mathcal{G}_i$ and $\mathcal{G}$
        else
            Add $\langle \{X, Y\}, \mathcal{V}_i \rangle$ to IP
        end
    end
    foreach unshielded path $\langle X, Z, Y \rangle$ in $\mathcal{G}_i$ do
        if $Z \notin \mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ then
            if an edge exists between either $X$ and $Z$
            or $Y$ and $Z$ in $\mathcal{G}$ then orient the edge to
            have an arrowhead endpoint at $Z$
        end
    end
end
```

Algorithm 2: Obtain Independence Facts, Inducing Paths, and Initial Structure

class that is significantly smaller than the equivalence classes produced by IOD and ION. Another open problem is how to use background knowledge about the true data generating process to efficiently guide the IOD search procedure. We hope to address these problems in future research.

Acknowledgements

We thank David Danks and Clark Glymour for helpful comments. RET was funded by a grant from the James S. McDonnell Foundation.

```
Input   : $\mathcal{G}$, $\mathcal{G}_1, \ldots, \mathcal{G}_n$, Sepset, IP
Output: $\mathbf{G}$
RemEdges $\leftarrow \varnothing$
foreach adjacent $X, Y \in \mathcal{G}$ do
    if for $1 \leq i \leq n$,
    $\{X, Y\} \cup \mathbf{Adj}^X_{\mathcal{G}} \cup \mathbf{PossSep}(\{X, Y\}, \mathcal{G}) \nsubseteq \mathcal{V}_i$ and
    $\{X, Y\} \cup \mathbf{Adj}^Y_{\mathcal{G}} \cup \mathbf{PossSep}(\{X, Y\}, \mathcal{G}) \nsubseteq \mathcal{V}_i$
    then add $\{X, Y\}$ to RemEdges
end
$\mathbf{G} \leftarrow \varnothing$
foreach $\mathbf{E} \in \mathcal{P}(\mathbf{RemEdges})$ do
    Let $\mathcal{H}$ be the induced subgraph of $\mathcal{G}$ containing
    only edges between pairs of variables not in $\mathbf{E}$
    PossImm $\leftarrow \varnothing$
    foreach $Z \in \mathcal{V}$ and pair $X, Y \in \mathbf{Adj}^Z_{\mathcal{G}}$ do
        if $\tau = \langle X, Z, Y \rangle$ can be made an immorality
        in $\mathcal{H}$ and for $1 \leq i \leq n$, either $Z \notin \mathcal{V}_i$ or
        $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is undefined then add $\tau$
        to PossImm
    end
    foreach $\mathbf{t} \in \mathcal{P}(\mathbf{PossImm})$ do
        $\mathcal{H}_{\mathbf{t}} \leftarrow \mathcal{H}$
        Make each $\tau \in \mathbf{t}$ an immorality in $\mathcal{H}_{\mathbf{t}}$
        Let $\mathbf{J}$ be the graphs resulting from iteratively
        applying rules $\mathcal{R}1 - \mathcal{R}10$ in Zhang [2007] to
        $\mathcal{H}_{\mathbf{t}}$ and whenever a discriminating path is
        found in $\mathcal{R}4$, creating two graphs where the
        discriminated collider has either orientation
        foreach $\mathcal{J} \in \mathbf{J}$ do
            $\mathcal{J}' \leftarrow \mathcal{J}$
            If an edge in $\mathcal{J}'$ has only one $\circ$ endpoint,
            make that endpoint a tail
            Let $\mathcal{K}$ be the induced subgraph of $\mathcal{J}'$
            containing only $\circ\!\!-\!\!\circ$ edges
            Orient $\mathcal{K}$ such that $\mathcal{K}$ has no immoralities
            For each edge in $\mathcal{K}$, give the corresponding
            edge in $\mathcal{J}'$ the same endpoint orientations
            if (i) $\mathcal{J}'$ is a MAG, (ii) each Sepset set
            corresponds to an m-separation in $\mathcal{J}'$,
            and (iii) for each $\langle \{X, Y\}, \mathcal{V}_i \rangle \in \mathbf{IP}$, there
            is an inducing path between $X$ and $Y$ with
            respect to $\mathcal{V} \backslash \mathcal{V}_i$ in $\mathcal{J}'$ then add $\mathcal{J}$ to $\mathbf{G}$
        end
    end
end
```

Algorithm 3: Construct the Equivalence Class

References

D. Danks. Scientific coherence and the fusion of experimental results. *The British Journal for the Philosophy of Science*, 56:791–807, 2005.

R. A. Fisher. *Statistical Methods for Research Workers*. Oliver and Boyd, London, 11th edition, 1950.

N. Friedman. The Bayesian structural EM algorithm. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998.

P. O. Hoyer, D. Janzing, J. M. Mooij, J. Peters, and B. Schölkopf. Nonlinear causal discovery with additive noise models. In *Advances in Neural Information Processing Systems 21*, 2009.

G. Melançon, I. Dutour, and M. Bousquet-Mélou. Random generation of dags for graph drawing. Technical Report INS-R0005, Centre for Mathematics and Computer Sciences, 2000.

J. D. Ramsey, S. J. Hanson, C. Hanson, Y. O. Halchenko, R. A. Poldrack, and C. Glymour. Six problems for causal inference from fMRI. *NeuroImage*, 49(2):1545–1558, 2010.

T. Richardson and P. Spirtes. Ancestral graph markov models. *Annals of Statistics*, 30(4):962–1030, 2002.

S. Shimizu, P. Hoyer, A. Hyvärinen, and A. Kerminen. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7:1003–2030, 2006.

P. Spirtes, C. Meek, and T. Richardson. Causal inference in the presence of latent variables and selection bias. In C. Glymour and G. Cooper, editors, *Computation, Causation and Discovery*, pages 211–252. MIT Press, 1999.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search.* MIT Press, 2nd edition, 2000.

R. E. Tillman, D. Danks, and C. Glymour. Integerating locally learned causal structures with overlapping variables. In *Advances in Neural Information Processing Systems 21*, 2009.

R. E. Tillman. Structure learning with independent non-identically distributed data. In *Proceedings of the 26th International Conference on Machine Learning*, 2009.

T. S. Verma and J. Pearl. Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, 1991.

J. Zhang. *Causal Inference and Reasoning in Causally Insufficient Systems.* PhD thesis, Carnegie Mello University, 2006.

J. Zhang. A characterization of markov equivalence classes for causal models with latent variables. In *Proceedings of Uncertainty in Artificial Intelligence*, 2007.

J. Zhang. Causal reasoning with ancestral graphs. *Journal of Machine Learning Research*, 9:1437–1474, 2008.

APPENDIX - SUPPLEMENTARY MATERIAL

Proofs

First, we need the following result to prove theorem 3.2.

Theorem 7.1 (Richardson and Spirtes [2002] Theorem 4.2)**.** Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a MAG and $\mathcal{W} \subset \mathcal{V}$. Then for $X, Y \in \mathcal{W}$ there is an inducing path between X and Y in $\mathcal{G}$ with respect to $\mathcal{V} \backslash \mathcal{W}$ if and only if $\forall \mathbf{Z} \subseteq \mathcal{W} \backslash \{X, Y\}$, there is an active path between X and Y with respect to $\mathbf{Z}$ in $\mathcal{G}$.

Theorem 3.2. Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ and $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ be MAGs where $\mathcal{W} \subset \mathcal{V}$. $\mathcal{H}$ is Markov equivalent to the marginal MAG with respect to $\mathcal{G}$ if the following hold for all $X, Y \in \mathcal{W}$
(i) if X and Y are adjacent in $\mathcal{H}$, then $\mathcal{G}$ has an inducing path between X and Y with respect to $\mathcal{V} \backslash \mathcal{W}$
(ii) $msep_{\mathcal{H}}(X, Y|\mathbf{Z}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{Z})$ for every $\mathbf{Z} \subseteq \mathcal{W} \backslash \{X, Y\}$,

Proof. Let $\mathcal{J} = \langle \mathcal{W}, \mathcal{F}' \rangle$ be the MAG which results from marginalizing $\mathcal{V} \backslash \mathcal{W}$ from $\mathcal{G}$ according to theorem 3.1.

If there is an edge between X and Y in $\mathcal{H}$, then there is an inducing path between X and Y with respect to $\mathcal{V} \backslash \mathcal{W}$ in $\mathcal{G}$. Also, $\forall \mathbf{Z} \subseteq \mathcal{W} \backslash \{X, Y\}$, there is an active path between X and Y with respect to $\mathbf{Z}$ by theorem 7.1. This entails an edge between X and Y in $\mathcal{J}$ by theorem 3.1(i). If X and Y are not adjacent in $\mathcal{H}$ then $\exists \mathbf{Z} \subseteq \mathcal{W} \backslash \{X, Y\}$ such that $msep_{\mathcal{H}}(X, Y|\mathbf{Z}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{Z}) \Rightarrow msep_{\mathcal{J}}(X, Y|\mathbf{Z}) \Rightarrow X \notin \mathbf{Adj}^{Y}_{\mathcal{J}}$. Thus, $\mathcal{H}$ and $\mathcal{J}$ have strictly the same adjacencies.

Let $\langle X, Z, Y \rangle$ be an immorality in $\mathcal{H}$. Then, $\exists \mathbf{S} \subseteq \mathcal{W} \backslash \{X, Y, Z\}$ such that $msep_{\mathcal{H}}(X, Y|\mathbf{S}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{S}) \Rightarrow msep_{\mathcal{J}}(X, Y|\mathbf{S}) \Rightarrow \langle X, Z, Y \rangle$ is an immorality in $\mathcal{J}$ since X and Z are adjacent in $\mathcal{J}$, Y and Z are adjacent in $\mathcal{J}$, X and Y are not adjacent in $\mathcal{J}$, and there is a conditioning set which m-separates X and Y in $\mathcal{J}$ which does not include Z. Let $\langle X, Z, Y \rangle$ be unshielded, but not an immorality in $\mathcal{H}$. Then, $\exists \mathbf{S} \subseteq \mathcal{W} \backslash \{X, Y, Z\}$ such that $msep_{\mathcal{H}}(X, Y|\mathbf{S} \cup \{Z\}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{S} \cup \{Z\}) \Rightarrow msep_{\mathcal{J}}(X, Y|\mathbf{S} \cup \{Z\}) \Rightarrow \langle X, Z, Y \rangle$ is not an immorality in $\mathcal{J}$ since X and Z are adjacent in $\mathcal{J}$, Y and Z are adjacent in $\mathcal{J}$, X and Y are not adjacent in $\mathcal{J}$, and Z is in a conditioning set which m-separates X and Y in $\mathcal{J}$. Thus, $\mathcal{H}$ and $\mathcal{J}$ have strictly the same immoralities.

Now let $\langle \phi_1, \ldots, \phi_k \rangle$ be all of the discriminating paths in $\mathcal{H}$ which have discriminated triples corresponding to discriminated triples in $\mathcal{J}$ and for $1 \leq i \leq k$, let $\phi_i = \langle W_{i,1}, \ldots, W_{i,n_i} \rangle$, i.e. $\langle W_{i,n_i-2}, W_{i,n_i-1}, W_{i,n_i} \rangle$ is the discriminated triple for ϕ_i. Then for $1 \leq i \leq k$, $\langle W_{i,n_i-2}, W_{i,n_i-1}, W_{i,n_i} \rangle$ is a v-structure in $\mathcal{H}$ if and only if every set which m-separates $W_{i,1}$ and W_{i,n_i} in $\mathcal{H}$ does not contain W_{i,n_i-1}, and $\langle W_{i,n_i-2}, W_{i,n_i-1}, W_{i,n_i} \rangle$ is not a v-structure in $\mathcal{H}$ if and only if every set which m-separates $W_{i,1}$ and W_{i,n_i} in $\mathcal{H}$ contains W_{i,n_i-1}. Due to this mutual exclusivity, if $\langle W_{i,n_i-2}, W_{i,n_i-1}, W_{i,n_i} \rangle$ is a v-structure in $\mathcal{H}$ and not a v-structure in $\mathcal{J}$, or vice versa, then there exists at least one pair $\{X, Y\} \subseteq \mathcal{W}$ such that $\exists \mathbf{S}, \mathbf{T} \subseteq \mathcal{W} \backslash \{X, Y\}$ such that $msep_{\mathcal{H}}(X, Y|\mathbf{S})$ and $\neg msep_{\mathcal{H}}(X, Y|\mathbf{T})$, but $\neg msep_{\mathcal{J}}(X, Y|\mathbf{S})$ and $msep_{\mathcal{J}}(X, Y|\mathbf{T})$. However, since $msep_{\mathcal{H}}(X, Y|\mathbf{S}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{S}) \Rightarrow msep_{\mathcal{J}}(X, Y|\mathbf{S})$, it must be the case that $\langle W_{i,n_i-2}, W_{i,n_i-1}, W_{i,n_i} \rangle$ is a v-structure in $\mathcal{J}$ if and only if $\langle W_{i,n_i-2}, W_{i,n_i-1}, W_{i,n_i} \rangle$ is a v-structure in $\mathcal{H}$ for $1 \leq i \leq k$.

Thus, since $\mathcal{H}$ and $\mathcal{J}$ have strictly the same adjacencies, immoralities, and discriminated triples that are v-structures, they are Markov equivalent. Thus, $msep_{\mathcal{H}}(X, Y|\mathbf{Z}) \Leftrightarrow msep_{\mathcal{J}}(X, Y|\mathbf{Z}) \Leftrightarrow msep_{\mathcal{G}}(X, Y|\mathbf{Z})$, for $X, Y \in \mathcal{W}$ and $\mathbf{Z} \subseteq \mathcal{W} \backslash \{X, Y\}$. □

For theorem 5.1, we first need several results. The following theorem, which is a restatement of lemmas and a theorem from Spirtes *et al.* [1999], and the corollary which follows says that algorithm 2 performs every conditional independence test necessary to determine the correct equivalence class.

Theorem 7.2 (Spirtes *et al.* [1999], Lemma 12, Lemma 13, Theorem 5)**.** Let **Sepset** be constructed as in algorithm 2 and for $1 \leq i \leq n$, let $\mathcal{K}_i = \langle \mathcal{V}_i, \mathcal{E}_i \rangle$ be a MAG. Then for $1 \leq i \leq n$, if for all $X, Y \in \mathcal{V}_i$ such that $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is defined, $msep_{\mathcal{K}_i}(X, Y|\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i))$, then for all $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Rightarrow msep_{\mathcal{K}_i}(X, Y|\mathbf{S})$.

Corollary 7.1. Let **Sepset** be constructed as in algorithm 2 and $\mathcal{K} = \langle \mathcal{V}, \mathcal{E} \rangle$ be a MAG. Then for $1 \leq i \leq n$, if for all $X, Y \in \mathcal{V}_i$ such that $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is defined, $msep_{\mathcal{K}}(X, Y|\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i))$, then for all $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Rightarrow msep_{\mathcal{K}}(X, Y|\mathbf{S})$.

Proof. Let i be any integer between 1 and n such that for all $X, Y \in \mathcal{V}_i$ such that $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is defined, $msep_{\mathcal{K}}(X, Y|\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i))$. Let X and Y be any such $X, Y \in \mathcal{V}_i$ such that $\exists \mathbf{S} \subseteq \mathcal{V}_i$ such that $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S}$, and let $\mathbf{S}$ be any such set. Assume $\neg msep_{\mathcal{K}}(X, Y|\mathbf{S})$. Now, let $\mathcal{K}_i$ be the marginal MAG for $\mathcal{K}$ with respect to $\mathcal{V}_i$. Since MAGs are closed under marginalization, $\neg msep_{\mathcal{K}}(X, Y|\mathbf{S}) \Rightarrow \neg msep_{\mathcal{K}_i}(X, Y|\mathbf{S})$. However, also since MAGs are closed under marginalization, it must then be the case that for all $X', Y' \in \mathcal{V}_i$ such that $\mathbf{Sepset}(\{X', Y'\}, \mathcal{G}_i)$ is defined, $msep_{\mathcal{K}_i}(X', Y'|\mathbf{Sepset}(\{X', Y'\}, \mathcal{G}_i))$. Thus, by corollary 7.1, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Rightarrow msep_{\mathcal{K}_i}(X, Y|\mathbf{S})$, which is a contradiction. Thus, $msep_{\mathcal{K}}(X, Y|\mathbf{S})$. □

We can now show that for any $\mathcal{J}$ added to $\mathbf{G}$, the output set of algorithm 3, the corresponding mixed graph which results at line 21 before $\mathcal{J}$ is added to this set is a MAG which entails every conditional independence and dependence true in the marginal distributions which generate each $D_i \in \mathbf{D}$.

Lemma 7.1. Let $\mathcal{J}$ be a partially oriented mixed graph that is added to $\mathbf{G}$ in algorithm 3 at line 22. Then the corresponding graph $\mathcal{J}'$ resulting at line 21 is a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{J}'}(X, Y|\mathbf{S})$.

Proof. If $\mathcal{J}$ is added to $\mathbf{G}$, then $\mathcal{J}'$ must be a MAG since condition (i) at line 22 must be satisfied. Since condition (ii) at line 22 must also be satisfied, by corollary 7.1 it must be the case that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$, and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{Z} \Rightarrow msep_{\mathcal{J}'}(X, Y|\mathbf{Z})$. For each $\mathcal{G}_i$ constructed in algorithm 2 and $X, Y \in \mathcal{V}_i$, X and Y are not adjacent if and only if $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is defined. Thus, also by corollary 7.1, $\langle \{X, Y\}, \mathcal{G}_i \rangle \in \mathbf{IP}$ if and only if X and Y are adjacent in a MAG $\mathcal{H}_i$ such that for $X', Y' \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X', Y'\}$, $X' \perp\!\!\!\perp_{\mathbf{D}} Y'|\mathbf{S} \Leftrightarrow msep_{\mathcal{H}_i}(X', Y'|\mathbf{S})$. Since condition (iii) at line 22 must also be satisfied, by theorem 3.2, for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{J}'}(X, Y|\mathbf{S})$. □

Next we need to show that any $\mathcal{J}$ is a PAG. To do this, we first show that if $\mathcal{H}$ is a marginal MAG for $\mathcal{J}'$, then orientations in $\mathcal{J}'$ which correspond to immoralities in $\mathcal{H}$ must be the same as in $\mathcal{H}$.

Lemma 7.2. Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle$ and $\mathcal{H} = \langle \mathcal{W}, \mathcal{F} \rangle$ be MAGs such that $\mathcal{W} \subset \mathcal{V}$ and for any $X, Y \in \mathcal{W}$ and $\mathbf{S} \subseteq \mathcal{W} \backslash \{X, Y\}$, $msep_{\mathcal{H}}(X, Y|\mathbf{S}) \Leftrightarrow msep_{\mathcal{G}}(X, Y|\mathbf{S})$. If $\langle X, Z, Y \rangle$ is an immorality in $\mathcal{H}$, then (i) if X and Z are adjacent in $\mathcal{G}$, then the edge between X and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z and (ii) if Y and Z are adjacent in $\mathcal{G}$, then the edge between Y and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z.

Proof. Since $\mathcal{H}$ is a MAG and $\langle X, Z, Y \rangle$ is an immorality in $\mathcal{H}$, $\exists \mathbf{S} \subseteq \mathcal{W} \backslash \{X, Z, Y\}$ such that $msep_{\mathcal{H}}(X, Y|\mathbf{S}) \Rightarrow msep_{\mathcal{G}}(X, Y|\mathbf{S})$. Since X and Z are adjacent in $\mathcal{H}$ and Y and Z are adjacent in $\mathcal{H}$, there are inducing paths between X and Z and between Y and Z in $\mathcal{G}$ by theorem 7.1. Thus, if either X and Z are adjacent in $\mathcal{G}$ or Y and Z are adjacent in $\mathcal{G}$, then either (i) $\langle X, Z, Y\rangle$ is a triple in $\mathcal{G}$, (ii) $\langle X, Z, Y\rangle$ is not a triple in $\mathcal{G}$, but $\mathcal{G}$ contains a path $\langle X, Z, V_1, \ldots, V_n, Y\rangle$ where $\langle Z, V_1, \ldots, V_n, Y\rangle$ is an active path with respect to $\mathbf{S}$ in $\mathcal{G}$, or (iii) $\langle X, Z, Y\rangle$ is not a triple in $\mathcal{G}$, but $\mathcal{G}$ contains a path $\langle X, V_1, \ldots, V_n, Z, Y\rangle$. where $\langle X, V_1, \ldots, V_n, Z\rangle$ is an active path with respect to $\mathbf{S}$ in $\mathcal{G}$.

Case (i):
Since $msep_{\mathcal{G}}(X, Y|\mathbf{S})$, $\langle X, Z, Y\rangle$ must be a v-structure in $\mathcal{G}$ since for any other orientations of the endpoints at Z along $\langle X, Z, Y\rangle$, $\langle X, Z, Y\rangle$ is an active path with respect to $\mathbf{S}$ since $Z \notin \mathbf{S}$. Thus, the edge between X and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z and the edge between Y and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z.

Case (ii):
Since $\langle Z, V_1, \ldots, V_n, Y\rangle$ is an active path with respect to $\mathbf{S}$ in $\mathcal{G}$ and $Z \notin \mathbf{S}$, $\langle X, Z, V_1, \ldots, V_n, Y\rangle$ must be an active path with respect to $\mathbf{S}$ in $\mathcal{G}$ unless $\langle X, Z, V_1\rangle$ is a v-structure in $\mathcal{G}$. Thus, the edge between X and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z.

Case (iii):
By symmetry with case (ii), $\langle V_n, Z, Y\rangle$ must be v-structure in $\mathcal{G}$. Thus, the edge between Y and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z. □

Now we show that $\mathcal{J}$ must also have these orientations.

Corollary 7.2. Let $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}\rangle$ be a PAG and $\mathcal{H} = \langle \mathcal{W}, \mathcal{F}\rangle$ be a MAG such that $\mathcal{W} \subset \mathcal{V}$ and for any $X, Y \in \mathcal{W}$, $\mathbf{S} \subseteq \mathcal{W}\backslash\{X, Y\}$, and MAG $\mathcal{K}$ represented by $\mathcal{G}$, $msep_{\mathcal{H}}(X, Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{K}}(X, Y|\mathbf{S})$. If $\langle X, Z, Y\rangle$ is an immorality in $\mathcal{H}$, then (i) if X and Z are adjacent in $\mathcal{G}$, then the edge between X and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z and (ii) if Y and Z are adjacent in $\mathcal{G}$, then the edge between Y and Z in $\mathcal{G}$ contains an arrowhead endpoint at Z.

Proof. This follows from the fact that lemma 7.2 can be applied to $\mathcal{H}$ and any such MAG $\mathcal{K}$, so the orientation is invariant in the Markov equivalence class for $\mathcal{K}$ and hence present in $\mathcal{G}$. □

Finally, we can show that $\mathcal{J}$ is a PAG.

Lemma 7.3. Any partially oriented mixed graph $\mathcal{J} \in \mathbf{G}$, the output set of algorithm 3, is a PAG.

Proof. For any $\mathcal{J}$ added to $\mathbf{G}$ at line 22 of algorithm 3, the associated $\mathcal{J}'$ created at line 21 is a MAG such that for $1 \le i \le n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i\backslash\{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{J}'}(X, Y|\mathbf{S})$ by lemma 7.1 and theorem 3.2 since $\mathcal{J}'$ must satisfy conditions (i), (ii), and (iii) at line 22 of algorithm 3. Thus, there exists a PAG $\mathcal{K}$ with the same adjacencies, immoralities, and v-structures at discriminated triples as $\mathcal{J}'$, i.e. $\mathcal{J}'$ is a MAG represented by $\mathcal{K}$. The partially oriented mixed graph $\mathcal{H}_{\mathbf{t}}$ resulting at line 14 must have the same adjacencies and immoralities as $\mathcal{K}$ since $\mathcal{R}1, \ldots, \mathcal{R}10$ do not add or remove edges or creat additional immoralities. Any oriented endpoints along edges of $\mathcal{H}_{\mathbf{t}}$ which do not correspond to immoralities in $\mathcal{K}$ must have been made at line 27 of algorithm 2 to the corresponding edge in $\mathcal{G}$. Any such orientation corresponds to an endpoint of an immorality in a MAG $\mathcal{L} = \langle \mathcal{V}_i, \mathcal{F}\rangle$ for some $1 \le i \le n$ such that for $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i\backslash\{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{L}}(X, Y|\mathbf{S})$. Thus, by corollary 7.2, $\mathcal{K}$ must have these orientations. Since $\mathcal{H}_{\mathbf{t}}$ thus contains no orientations that are not present in $\mathcal{K}$, has the same adjacencies and immoralities as $\mathcal{K}$, and the orientation rules $\mathcal{R}1, \ldots, \mathcal{R}10$ are correct and complete for obtaining a fully oriented PAG [Zhang, 2007], $\mathcal{J}$ is a PAG. □

The last result we need to prove theorem 5.1 is that the steps in lines 18-21 of algorithm 3 produce a MAG from a PAG. This result is proven in Zhang [2006].

Theorem 7.3 (Zhang [2006] Lemma 3.3.4)**.** Let $\mathcal{G}$ be a fully oriented PAG and $\mathcal{H}$ be the mixed graph which results from orienting $\circ$-marked edges in $\mathcal{G}$ as follows:
(i) if an edge has one $\circ$ endpoint, then make this endpoint a tail
(ii) if an edge has two $\circ$ endpoints then let this edge have the same orientations as the corresponding edge in graph which results from orienting edges in the induced subgraph of $\mathcal{G}$ containing only edges with two $\circ$ endpoints such that this graph is a DAG with no immoralities
Then $\mathcal{H}$ is a MAG in the equivalence class represented by $\mathcal{G}$.

Now, we can prove that algorithm 3 produces a correct equivalence class.

Theorem 5.1 (correctness)**.** $\forall \mathcal{J} \in \mathbf{G}$, resulting from algorithm 3, $\mathcal{J}$ is a PAG such that for any MAG $\mathcal{J}'$ in the equivalence class represented by $\mathcal{J}$, for $1 \le i \le n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i\backslash\{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{J}'}(X, Y|\mathbf{S})$.

Proof. By lemma 7.3, if $\mathcal{J}$ is added to $\mathbf{G}$ at line 22 of algorithm 3, then $\mathcal{J}$ is a PAG. By theorem 7.3, the corresponding mixed graph $\mathcal{J}'$ resulting at line 21 of algorithm 3 is a MAG in the equivalence class represented by $\mathcal{J}$. By lemma 7.1, for $1 \le i \le n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i\backslash\{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y|\mathbf{S} \Leftrightarrow msep_{\mathcal{J}'}(X, Y|\mathbf{S})$. Any MAG in the equivalence class represented by $\mathcal{J}$ other than $\mathcal{J}'$ is, by definition, Markov equivalent to $\mathcal{J}'$ and thus entails the same conditional independences and dependences. □

Now, in order to prove theorem 5.2, we need several results. We first show that any MAG $\mathcal{H}$ which entails every conditional independence and dependence true in the marginal distributions which generate each $D_i \in \mathbf{D}$ must contains a subset of the edges and superset of the immoralities of the partially oriented mixed graph $\mathcal{G}$ produced by algorithm 2.

Lemma 7.4. Let $\mathcal{H}$ be a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$. Then the partially oriented mixed graph $\mathcal{G}$ which results from algorithm 2 contains a superset of the edges in $\mathcal{H}$.

Proof. $\mathcal{G}$ is initially constructed at line 1 of algorithm 2 as a complete graph with no orientations. An edge between two nodes X and Y in $\mathcal{G}$ is only removed in algorithm 2 if for some $1 \leq i \leq n$ such that $X, Y \in \mathcal{V}_i$, $\exists \mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$ such that $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Rightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$. For any such $\mathbf{S}$, there is an active path between X and Y with respect to $\mathbf{S}$ in $\mathcal{H}$ unless X and Y are not adjacent in $\mathcal{H}$. □

Lemma 7.5. Let $\mathcal{H}$ be a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$. Then the partially oriented mixed graph $\mathcal{G}$ which results from algorithm 2 contains a subset of the immoralities in $\mathcal{H}$ among common edges.

Proof. If $\langle X, Z, Y \rangle$ is an immorality in $\mathcal{G}$ and $\langle X, Z, Y \rangle$ is a path in $\mathcal{H}$, then X and Y are not adjacent in $\mathcal{H}$ by lemma 7.4. Since $\langle X, Z, Y \rangle$ is an immorality in $\mathcal{G}$, then for some $1 \leq i \leq n$, $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is defined, but $Z \notin \mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$. Thus, for some $1 \leq i \leq n$, there exists a MAG $\mathcal{H}_i = \langle \mathcal{V}_i, \mathcal{F} \rangle$ such that for $X', Y' \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X', Y'\}$, $X' \perp\!\!\!\perp_{\mathbf{D}} Y' | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}_i}(X', Y'|\mathbf{S})$ and $\langle X, Z, Y \rangle$ is an immorality in $\mathcal{H}_i$. Thus, by lemma 7.2, $\langle X, Z, Y \rangle$ is an immorality in $\mathcal{H}$. □

Now we show that some graph with the same adjacencies as $\mathcal{H}$ is considered in algorithm 3.

Lemma 7.6. Let $\mathcal{H}$ be a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$. Then some partially oriented mixed graph considered at line 7 of algorithm 3 must contain the same adjacencies as $\mathcal{H}$.

Proof. By lemma 7.4, $\mathcal{H}$ must contain a subset of the edges in the partially oriented mixed graph $\mathcal{G}$. Let X and Y be two adjacent nodes in $\mathcal{G}$. If X and Y are not adjacent in $\mathcal{H}$, then $\exists \mathbf{S} \subseteq \mathcal{V} \backslash \{X, Y\}$ such that $msep_{\mathcal{H}}(X, Y|\mathbf{S})$ since $\mathcal{H}$ is a MAG. By corollary 7.1, for any such $\mathbf{S}$, it must be the case that for $1 \leq i \leq n$, $\mathbf{S} \nsubseteq \mathcal{V}_i \backslash \{X, Y\}$. Also, by corollary 7.1, for one such $\mathbf{S}$, it must be the case that either $\mathbf{S} \subseteq \mathbf{Adj}^X_{\mathcal{H}}$ or $\mathbf{S} \subseteq \mathbf{PossSep}(\{X, Y\}, \mathcal{H})$ and for one such $\mathbf{S}$, it must be the case that either $\mathbf{S} \subseteq \mathbf{Adj}^Y_{\mathcal{H}}$ or $\mathbf{S} \subseteq \mathbf{PossSep}(\{X, Y\}, \mathcal{H})$. Thus, any such X and Y must be added to **RemEdges** at line 3 of algorithm 3 so some partially oriented mixed graph with the same adjacencies as $\mathcal{H}$ must be considered at line 7 of algorithm 3. □

Now we show that some graph with the same adjacencies and immoralities as $\mathcal{H}$ is considered in algorithm 3.

Lemma 7.7. Let $\mathcal{H}$ be a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$. Then some partially oriented mixed graph considered at line 7 of algorithm 3 must contain the same adjacencies and immoralities as $\mathcal{H}$.

Proof. By lemma 7.6, some $\mathcal{H}'$ must be considered at line 7 of algorithm 3 which has the same adjacencies as $\mathcal{H}$. Since by lemma 7.5 $\mathcal{H}$ must contain a superset of the immoralities among common edges in the partially oriented mixed graph $\mathcal{G}$ and $\mathcal{H}'$ contains a subset of the edges in $\mathcal{G}$, $\mathcal{H}$ must contain a superset of the immoralities in $\mathcal{H}'$. Let $\langle X, Z, Y \rangle$ be an immorality in $\mathcal{H}$ that is not present in $\mathcal{H}'$. If for some $1 \leq i \leq n$, $Z \in \mathcal{V}_i$ and $\mathbf{Sepset}(\{X, Y\}, \mathcal{G}_i)$ is defined, then $\langle X, Z, Y \rangle$ would be made an immorality at line 27 of algorithm 2. Thus, $\langle X, Z, Y \rangle$ must be included in **PossImm** so some partially oriented mixed graph with the same adjacencies and immoralities as $\mathcal{H}$ must be considered at line 14 of algorithm 3. □

Now, we can prove that algorithm 3 produces a complete equivalence class.

Theorem 5.2 (completeness)**.** Let $\mathcal{H}$ be a MAG such that for $1 \leq i \leq n$, $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$. Then $\exists \mathcal{J} \in \mathbf{G}$ such that $\mathcal{H}$ can be formed by orienting $\circ$-marked edges in $\mathcal{J}$.

Proof. By lemma 7.7, some $\mathcal{H}_{\mathbf{t}}$ must be considered at line 14 of algorithm 3 which has the same adjacencies and immoralities as $\mathcal{H}$. Any oriented endpoints along edges of $\mathcal{H}_{\mathbf{t}}$ which do not correspond to immoralities must have been made at line 27 of algorithm 2 to the corresponding edge in $\mathcal{G}$. Any such orientation corresponds to an endpoint of an immorality in a MAG $\mathcal{L} = \langle \mathcal{V}_i, \mathcal{F} \rangle$ for some $1 \leq i \leq n$ such that for $X, Y \in \mathcal{V}_i$ and $\mathbf{S} \subseteq \mathcal{V}_i \backslash \{X, Y\}$, $msep_{\mathcal{L}}(X, Y|\mathbf{S}) \Leftrightarrow X \perp\!\!\!\perp_{\mathbf{D}} Y | \mathbf{S} \Leftrightarrow msep_{\mathcal{H}}(X, Y|\mathbf{S})$ by corollary 7.1. Thus, by corollary 7.2, any PAG $\mathcal{J}$ which represents $\mathcal{H}$ must have these orientations. Thus, since the orientation rules $\mathcal{R}1, \ldots, \mathcal{R}10$ are correct and complete for obtaining a fully oriented PAG [Zhang, 2007], $\mathcal{J} \in \mathbf{J}$

constructed at line 15 of algorithm 3. When $\mathcal{J}$ is considered at line 16 of algorithm 3, the corresponding $\mathcal{J}'$ constructed at line 21 of algorithm 3 is a MAG that is Markov equivalent to $\mathcal{H}$ by theorem 7.3. Thus, $\mathcal{J}'$ satisfies conditions (i) and (ii) considered at line 22. By theorem 7.1, $\mathcal{J}'$ also satisfies condition (iii) considered at line 22. Thus, $\mathcal{J}$ is added to $\mathbf{G}$. □