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

Robert E. Tillman\textsuperscript{1} \quad Peter Spirtes\textsuperscript{2}

\textsuperscript{1}Carnegie Mellon University
return@cmu.edu

\textsuperscript{2}Carnegie Mellon University
ps7z@andrew.cmu.edu

April 11, 2011
Learning from single i.i.d. dataset

\[ X_1 \quad Y_1 \quad Z_1 \quad U_1 \quad V_1 \quad W_1 \quad t_1 \]
\[ X_2 \quad Y_2 \quad Z_2 \quad U_2 \quad V_2 \quad W_2 \quad t_2 \]
\[ X_3 \quad Y_3 \quad Z_3 \quad U_3 \quad V_3 \quad W_3 \quad t_3 \]
\[ X_4 \quad Y_4 \quad Z_4 \quad U_4 \quad V_4 \quad W_4 \quad t_4 \]
\[ X_5 \quad Y_5 \quad Z_5 \quad U_5 \quad V_5 \quad W_5 \quad t_5 \]
\[ X_6 \quad Y_6 \quad Z_6 \quad U_6 \quad V_6 \quad W_6 \quad t_6 \]
\[ X_7 \quad Y_7 \quad Z_7 \quad U_7 \quad V_7 \quad W_7 \quad t_7 \]
\[ X_8 \quad Y_8 \quad Z_8 \quad U_8 \quad V_8 \quad W_8 \quad t_8 \]
\[ X_9 \quad Y_9 \quad Z_9 \quad U_9 \quad V_9 \quad W_9 \quad t_9 \]
\[ X_{10} \quad Y_{10} \quad Z_{10} \quad U_{10} \quad V_{10} \quad W_{10} \quad t_{10} \]
\[ \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \]
Learning from multiple datasets with overlapping variables
Examples: Learning neural cascades during cognitive tasks

Dataset 1
Dataset 2
Dataset 3
Dataset 4

Learning from multiple datasets with overlapping variables

Tillman and Spirtes

AISTATS 2011
Formal Problem Statement:

Assumptions:

- Underlying structure for variables of interest $\mathcal{V}$
- Observational data for $\mathcal{V}_1, \mathcal{V}_2, \cdots \subset \mathcal{V}$
- If we add latent variables (confounding and selection) to the underlying structure, it can be represented as a DAG

Goal:

- Learn correct underlying structure from the datasets using conditional independence information (highest scoring set of factorizations of joint probability distribution)
- I.e. learn the complete Markov equivalence class with latent variables
Formal Problem Statement:

Assumptions:
- Underlying structure for variables of interest $\mathcal{V}$
- Observational data for $\mathcal{V}_1, \mathcal{V}_2, \cdots \subset \mathcal{V}$
- If we add latent variables (confounding and selection) to underlying structure, it can be represented as a DAG

Goal:
- Learn correct underlying structure from the datasets
  - using conditional independence information (highest scoring set of factorizations of joint probability distribution)
  - i.e. learn the complete Markov equivalence class with latent variables
Formal Problem Statement:

Assumptions:

- Underlying structure for variables of interest $\mathcal{V}$
- Observational data for $\mathcal{V}_1, \mathcal{V}_2, \cdots \subset \mathcal{V}$
- If we add latent variables (confounding and selection) to underlying structure, it can be represented as a DAG

Goal:

- Learn correct underlying structure from the datasets
  - using conditional independence information (highest scoring set of factorizations of joint probability distribution)
  - i.e. learn the complete Markov equivalence class with latent variables
Formal Problem Statement:

Assumptions:

- Underlying structure for variables of interest $\mathcal{V}$
- Observational data for $\mathcal{V}_1, \mathcal{V}_2, \cdots \subseteq \mathcal{V}$
- If we add latent variables (confounding and selection) to underlying structure, it can be represented as a DAG

Goal:

- Learn correct underlying structure from the datasets
  - using conditional independence information (highest scoring set of factorizations of joint probability distribution)
  - i.e. learn the complete Markov equivalence class with latent variables
Formal Problem Statement:

Assumptions:
- Underlying structure for variables of interest $\mathcal{V}$
- Observational data for $\mathcal{V}_1, \mathcal{V}_2, \cdots \subset \mathcal{V}$
- If we add latent variables (confounding and selection) to underlying structure, it can be represented as a DAG

Goal:
- Learn correct underlying structure from the datasets
  - using conditional independence information (highest scoring set of factorizations of joint probability distribution)
  - i.e. learn the complete Markov equivalence class with latent variables
Errors due to latent variables

Confounding:

\[ Z \rightarrow X \rightarrow Y \]
leads to spurious arcs

Conditioning:

\[ Z \rightarrow X \]
leads to spurious dependences

Learning from multiple datasets with overlapping variables

Tillman and Spirtes

AISTATS 2011
Errors due to latent variables

Confounding:

Conditioning:

leads to spurious arcs

leads to spurious dependences
Errors due to latent variables

Confounding:

\[ Z \rightarrow X \rightarrow Y \]

leads to spurious arcs

Conditioning:

\[ X \rightarrow Z \rightarrow Y \]

leads to spurious dependences
Errors due to latent variables

Confounding:

\[ Z \rightarrow X \rightarrow Y \]

leads to spurious arcs

Conditioning:

\[ X \rightarrow Z \rightarrow Y \]

leads to spurious dependences
Maximal Ancestral Graphs (MAGs) represent sets of DAGs with the same conditional independence relationships amongst observed variables.

- Bidirected edges represent confounding
- Undirected edges represent association due to selection bias (conditioning)

Important MAG Properties:
- Closure under conditioning and marginalization
- Separation criteria (m-separation) relating conditional independence to graph topology
Maximal Ancestral Graphs (MAGs) represent sets of DAGs with the same conditional independence relationships amongst observed variables

- bidirected edges represent confounding
- undirected edges represent association due to selection bias (conditioning)

Important MAG Properties

- closure under conditioning and marginalization
- separation criteria ($m$-separation) relating conditional independence to graph topology
Maximal Ancestral Graphs (MAGs) represent sets of DAGs with the same conditional independence relationships amongst observed variables:

- bidirected edges represent confounding
- undirected edges represent association due to selection bias (conditioning)

Important MAG Properties:

- closure under conditioning and marginalization
- separation criteria (m-separation) relating conditional independence to graph topology
Maximal Ancestral Graphs (MAGs) represent sets of DAGs with the same conditional independence relationships amongst observed variables:

- Bidirected edges represent confounding
- Undirected edges represent association due to selection bias (conditioning)

Important MAG Properties:

- Closure under conditioning and marginalization
- Separation criteria (m-separation) relating conditional independence to graph topology
Maximal Ancestral Graphs (MAGs) represent sets of DAGs with the same conditional independence relationships amongst observed variables

- bidirected edges represent confounding
- undirected edges represent association due to selection bias (conditioning)

Important MAG Properties

- closure under conditioning and marginalization
- separation criteria ($m$-separation) relating conditional independence to graph topology
Markov Equivalence and PAGs

Two MAGs are Markov equivalent if they contain the same:

- adjacencies
- immoralities
- discriminating path
- v-structures
Markov Equivalence and PAGs

Two MAGs are Markov equivalent if they contain the same:
- adjacencies
- immoralities
- discriminating path
- v-structures

Partial Ancestral Graphs (PAGs)
Restated Goal:

\[ X_1 \quad Z_1 \]
\[ X_2 \quad Z_2 \]
\[ X_3 \quad Z_3 \]
\[ \ldots \ldots \ldots \]

\[ Y_1^' \quad Z_1^' \quad W_1^' \]
\[ Y_2^' \quad Z_2^' \quad W_2^' \]
\[ Y_3^' \quad Z_3^' \quad W_3^' \]
\[ \ldots \ldots \ldots \]

Learning from multiple datasets with overlapping variables  
Tillman and Spirtes  
AISTATS 2011
The Integration of Overlapping Networks (ION) Algorithm:

**Input:** MAGs/PAGs over different variable sets with overlapping variables

**Output:** set of PAGs
Related Approach: ION Algorithm [Tillman et al., NIPS 2008]

Learning from multiple datasets using ION

- Use FCI [Spirtes et al., UAI 1995; Zhang, UAI 2007] to learn PAG for each dataset
- Input PAGs to ION

Problems/Limitations:

- Statistical errors in FCI can lead to conflicting graphical constraints
- Each FCI instance ignores useful data
- Significant computation time and memory usage (in some cases)

Our approach:

- Learn directly from datasets, always using as many datasets as possible
Related Approach: ION Algorithm [Tillman et al., NIPS 2008]

Learning from multiple datasets using ION

- Use FCI [Spirtes et al., UAI 1995; Zhang, UAI 2007] to learn PAG for each dataset
- Input PAGs to ION

Problems/Limitations:

- Statistical errors in FCI can lead to conflicting graphical constraints
- Each FCI instance ignores useful data
- Significant computation time and memory usage (in some cases)

Our approach:

- Learn directly from datasets, always using as many datasets as possible
### Related Approach: ION Algorithm [Tillman et al., NIPS 2008]

**Learning from multiple datasets using ION**
- Use FCI [Spirtes et al., UAI 1995; Zhang, UAI 2007] to learn PAG for each dataset
- Input PAGs to ION

**Problems/Limitations:**
- **Statistical errors in FCI can lead to conflicting graphical constraints**
- Each FCI instance ignores useful data
- Significant computation time and memory usage (in some cases)

**Our approach:**
- Learn directly from datasets, always using as many datasets as possible
Related Approach: ION Algorithm [Tillman et al., NIPS 2008]

Learning from multiple datasets using ION

- Use FCI [Spirtes et al., UAI 1995; Zhang, UAI 2007] to learn PAG for each dataset
- Input PAGs to ION

Problems/Limitations:

- Statistical errors in FCI can lead to conflicting graphical constraints
- Each FCI instance ignores useful data
- Significant computation time and memory usage (in some cases)

Our approach:

- Learn directly from datasets, always using as many datasets as possible
Related Approach: ION Algorithm [Tillman et al., NIPS 2008]

Learning from multiple datasets using ION

- Use FCI [Spirtes et al., UAI 1995; Zhang, UAI 2007] to learn PAG for each dataset
- Input PAGs to ION

Problems/Limitations:

- Statistical errors in FCI can lead to conflicting graphical constraints
- Each FCI instance ignores useful data
- Significant computation time and memory usage (in some cases)

Our approach:

- Learn directly from datasets, always using as many datasets as possible
Related Approach: ION Algorithm [Tillman et al., NIPS 2008]

Learning from multiple datasets using ION

- Use FCI [Spirtes et al., UAI 1995; Zhang, UAI 2007] to learn PAG for each dataset
- Input PAGs to ION

Problems/Limitations:

- Statistical errors in FCI can lead to conflicting graphical constraints
- Each FCI instance ignores useful data
- Significant computation time and memory usage (in some cases)

Our approach:

- Learn directly from datasets, always using as many datasets as possible
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing

- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]

- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null
- In practice, Fisher test statistic leads to best results

Our Approach:

- Use Fisher test statistic with each dataset measuring the relevant variables
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing
- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]
- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null
- In practice, Fisher test statistic leads to best results

Our Approach:
- Use Fisher test statistic with each dataset measuring the relevant variables
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing

- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]

- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null
- In practice, Fisher test statistic leads to best results

Our Approach:

- Use Fisher test statistic with each dataset measuring the relevant variables
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing
- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]
- Choose appropriate independence test for each dataset
  - Find $p$-value associated with each test statistic
  - Form new test statistic for distribution of $p$-values under null
  - In practice, Fisher test statistic leads to best results

Our Approach:
- Use Fisher test statistic with each dataset measuring the relevant variables
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing
- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]
- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null
- In practice, Fisher test statistic leads to best results

Our Approach:
- Use Fisher test statistic with each dataset measuring the relevant variables
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing

- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]

- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null

In practice, Fisher test statistic leads to best results

Our Approach:

- Use Fisher test statistic with each dataset measuring the relevant variables
Conditional Independence Testing with Multiple Datasets

Concatenating datasets for independence testing

- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]

- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null
- In practice, Fisher test statistic leads to best results

Our Approach:

- Use Fisher test statistic with each dataset measuring the relevant variables
Concatenating datasets for independence testing
- Differences between datasets even after standardizing
- Lead to spurious associations/independences

Pooling $p$-values [Tillman, ICML 2009]
- Choose appropriate independence test for each dataset
- Find $p$-value associated with each test statistic
- Form new test statistic for distribution of $p$-values under null
- In practice, Fisher test statistic leads to best results

Our Approach:
- Use Fisher test statistic with each dataset measuring the relevant variables
The Integration of Overlapping Datasets (IOD) Algorithm

Basic Structure:

1. Begin with complete graph over $\mathcal{V}$
2. First stage of FCI with respect to each variable set $\mathcal{V}_i$

Results in:
- Superset of edges in correct PAG
- Subset of immoralities and discriminating path v-structures

3. Find edges to remove and immoralities/v-structures to orient, s.t. resulting structure is consistent with independence/dependencies
4. Apply final PAG orientation rules to each such structure
The Integration of Overlapping *Datasets* (IOD) Algorithm

**Basic Structure:**

1. Begin with complete graph over $\mathcal{V}$
2. First stage of FCI with respect to each variable set $\mathcal{V}_i$

Results in:
- Superset of edges in correct PAG
- Subset of immoralities and discriminating path v-structures

3. Find edges to remove and immoralities/v-structures to orient, s.t. resulting structure is consistent with independencies/dependencies
4. Apply final PAG orientation rules to each such structure
The Integration of Overlapping Datasets (IOD) Algorithm

Basic Structure:

1. Begin with complete graph over $\mathcal{V}$
2. First stage of FCI with respect to each variable set $\mathcal{V}_i$

Results in

- Superset of edges in correct PAG
- Subset of immoralities and discriminating path v-structures

3. Find edges to remove and immoralities/v-structures to orient, s.t. resulting structure is consistent with independencies/dependencies
4. Apply final PAG orientation rules to each such structure
The Integration of Overlapping *Datasets* (IOD) Algorithm

**Basic Structure:**

1. Begin with complete graph over $\mathcal{V}$
2. First stage of FCI with respect to each variable set $\mathcal{V}_i$

Results in

- Superset of edges in correct PAG
- Subset of immoralities and discriminating path v-structures

3. Find edges to remove and immoralities/v-structures to orient, s.t. resulting structure is consistent with independencies/dependencies

4. Apply final PAG orientation rules to each such structure
**The Integration of Overlapping Datasets (IOD) Algorithm**

**Basic Structure:**

1. Begin with complete graph over $\mathcal{V}$
2. First stage of FCI with respect to each variable set $\mathcal{V}_i$

**Results in**
- Superset of edges in correct PAG
- Subset of immoralities and discriminating path v-structures

3. Find edges to remove and immoralities/v-structures to orient, s.t. resulting structure is consistent with independencies/dependencies
4. Apply final PAG orientation rules to each such structure
Removing edges and adding orientations

Determine whether resulting structure $G$ is consistent with independences/dependences

**Theorem**

- Let $G_{\mathcal{V}_1}^*, G_{\mathcal{V}_2}^*, \ldots$ be marginalizations of the correct structure
- Then $G$ is consistent with $G_{\mathcal{V}_1}^*, G_{\mathcal{V}_2}^*, \ldots$ if
  - $\exists Z \subseteq \mathcal{V}_i$ s.t. $\text{msep}_{G_{\mathcal{V}_i}}(X, Y|Z) = \text{msep}_G(X, Y|Z)$
  - $X, Y$ adjacent in $G_{\mathcal{V}_i} = G$ contains an inducing path w/ resp. to $\mathcal{V}\setminus\mathcal{V}_i$

Furthermore

- Sufficient $Z$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $\mathcal{G}$ is consistent with independences/dependences

**Theorem**

- Let $\mathcal{G}^*_1, \mathcal{G}^*_2, \ldots$ be marginalizations of the correct structure
- Then $\mathcal{G}$ is consistent with $\mathcal{G}^*_1, \mathcal{G}^*_2, \ldots$ if
  - $\exists Z \subseteq V_i$ s.t. $\text{msep}_{\mathcal{G}^*_i}(X, Y|Z) \Rightarrow \text{msep}_\mathcal{G}(X, Y|Z)$
  - $X, Y$ adjacent in $\mathcal{G}^*_i = \mathcal{G}$ contains an inducing path w/ resp. to $V \setminus V_i$

Furthermore

- Sufficient Z sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $\mathcal{G}$ is consistent with independences/dependences

Theorem

- Let $\mathcal{G}_{\mathcal{V}_1}^*, \mathcal{G}_{\mathcal{V}_2}^*, \ldots$ be marginalizations of the correct structure
- Then $\mathcal{G}$ is consistent with $\mathcal{G}_{\mathcal{V}_1}^*, \mathcal{G}_{\mathcal{V}_2}^*, \ldots$ if
  1. $\exists Z \subseteq \mathcal{V}_i$ s.t. $\text{msep}_{\mathcal{G}_{\mathcal{V}_i}^*}(X, Y|Z) \Rightarrow \text{msep}_{\mathcal{G}}(X, Y|Z)$
  2. $X, Y$ adjacent in $\mathcal{G}_{\mathcal{V}_i}^* \Rightarrow \mathcal{G}$ contains an inducing path w/ resp. to $\mathcal{V}\setminus \mathcal{V}_i$

Furthermore

- Sufficient $Z$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $G$ is consistent with independences/dependences

**Theorem**

- Let $G_{V_1}^*, G_{V_2}^*, \ldots$ be marginalizations of the correct structure
- Then $G$ is consistent with $G_{V_1}^*, G_{V_2}^*, \ldots$ if
  1. $\exists Z \subseteq V_i$ s.t. $msep_{G_{V_i}^*}(X, Y|Z) \Rightarrow msep_G(X, Y|Z)$
  2. $X, Y$ adjacent in $G_{V_i}^* \Rightarrow G$ contains an inducing path w/ resp. to $\mathcal{V} \setminus V_i$

Furthermore

- Sufficient $Z$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path $v$-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $\mathcal{G}$ is consistent with independences/dependences

**Theorem**

- Let $\mathcal{G}_1^*, \mathcal{G}_2^*, \ldots$ be marginalizations of the correct structure
- Then $\mathcal{G}$ is consistent with $\mathcal{G}_1^*, \mathcal{G}_2^*, \ldots$ if
  1. $\exists Z \subseteq V_i \text{ s.t. } msep_{\mathcal{G}_i^*}(X, Y | Z) \Rightarrow msep_{\mathcal{G}}(X, Y | Z)$
  2. $X, Y$ adjacent in $\mathcal{G}_i^* \Rightarrow \mathcal{G}$ contains an inducing path w/ resp. to $\mathcal{V} \setminus V_i$

Furthermore

- Sufficient $Z$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $\mathcal{G}$ is consistent with independences/dependences

Theorem

- Let $\mathcal{G}_{\mathcal{V}_1}^*, \mathcal{G}_{\mathcal{V}_2}^*, \ldots$ be marginalizations of the correct structure

- Then $\mathcal{G}$ is consistent with $\mathcal{G}_{\mathcal{V}_1}^*, \mathcal{G}_{\mathcal{V}_2}^*, \ldots$ if
  1. $\exists Z \subseteq \mathcal{V}_i$ s.t. $\text{msep}_{\mathcal{G}_{\mathcal{V}_i}^*}(X, Y|Z) \Rightarrow \text{msep}_{\mathcal{G}}(X, Y|Z)$
  2. $X, Y$ adjacent in $\mathcal{G}_{\mathcal{V}_i}^* \Rightarrow \mathcal{G}$ contains an inducing path w/ resp. to $\mathcal{V}\setminus\mathcal{V}_i$

Furthermore

- Sufficient $Z$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $\mathcal{G}$ is consistent with independences/dependences

**Theorem**

- Let $\mathcal{G}_{\mathcal{V}_1}^*, \mathcal{G}_{\mathcal{V}_2}^*, \ldots$ be marginalizations of the correct structure
- Then $\mathcal{G}$ is consistent with $\mathcal{G}_{\mathcal{V}_1}^*, \mathcal{G}_{\mathcal{V}_2}^*, \ldots$ if
  1. $\exists \mathcal{Z} \subseteq \mathcal{V}_i$ s.t. $\text{msep}_{\mathcal{G}_{\mathcal{V}_i}^*}(X, Y|\mathcal{Z}) \Rightarrow \text{msep}_{\mathcal{G}}(X, Y|\mathcal{Z})$
  2. $X, Y$ adjacent in $\mathcal{G}_{\mathcal{V}_i}^* \Rightarrow \mathcal{G}$ contains an inducing path w/ resp. to $\mathcal{V}\setminus \mathcal{V}_i$

Furthermore

- Sufficient $\mathcal{Z}$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined

Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Removing edges and adding orientations

Determine whether resulting structure $\mathcal{G}$ is consistent with independences/dependences

**Theorem**

- Let $\mathcal{G}^*_{\mathcal{V}_1}, \mathcal{G}^*_{\mathcal{V}_2}, \ldots$ be marginalizations of the correct structure
- Then $\mathcal{G}$ is consistent with $\mathcal{G}^*_{\mathcal{V}_1}, \mathcal{G}^*_{\mathcal{V}_2}, \ldots$ if
  1. $\exists \mathcal{Z} \subseteq \mathcal{V}_i \text{ s.t. } \text{msep}_{\mathcal{G}^*_{\mathcal{V}_i}}(X, Y|\mathcal{Z}) \Rightarrow \text{msep}_{\mathcal{G}}(X, Y|\mathcal{Z})$
  2. $X, Y$ adjacent in $\mathcal{G}^*_{\mathcal{V}_i} \Rightarrow \mathcal{G}$ contains an inducing path w/ resp. to $\mathcal{V}\setminus\mathcal{V}_i$

Furthermore

- Sufficient $\mathcal{Z}$ sets are checked during the first stage of FCI
- Necessary inducing paths can also be determined
- Provides information to determine edges that can be removed and immoralities / discriminating path v-structures that can be oriented
Example

**Truth:**

```
X
Y
Z
V
W
```

**Output:** $V \notin \mathcal{V}_1$, $W \notin \mathcal{V}_2$
Example

Truth:

Output: $V \notin \mathcal{V}_1$, $W \notin \mathcal{V}_2$
Truth:

Output: $V \notin \mathcal{V}_1$, $W \notin \mathcal{V}_2$
Truth:

Output: $V \notin \mathcal{V}_1$, $W \notin \mathcal{V}_2$
Example

Truth:

Output: $V \not\in \mathcal{V}_1$, $W \not\in \mathcal{V}_2$
Correctness and Completeness

If a MAG $G^* = \langle \mathcal{V}, \mathcal{E} \rangle$ describes data generating mechanism

Then, in the large sample limit...

**Theorem (Correctness)**
For any PAG $H$ in the output of IOD, $H$ represents a set of MAGs that are Markov equivalent to $G^*_{|\mathcal{V}_1|}, G^*_{|\mathcal{V}_2|}, \ldots$

**Theorem (Completeness)**
If there exists a PAG $H$ which represents a set of MAGs that are Markov equivalent to $G^*_{|\mathcal{V}_1|}, G^*_{|\mathcal{V}_2|}, \ldots$, then $H$ is in the output of IOD
Correctness and Completeness

If a MAG $G^* = \langle \mathcal{V}, \mathcal{E} \rangle$ describes data generating mechanism

Then, in the large sample limit...

Theorem (Correctness)

For any PAG $H$ in the output of IOD, $H$ represents a set of MAGs that are Markov equivalent to $G^*_1, G^*_2, \ldots$.

Theorem (Completeness)

If there exists a PAG $H$ which represents a set of MAGs that are Markov equivalent to $G^*_1, G^*_2, \ldots$, then $H$ is in the output of IOD.
Correctness and Completeness

If a MAG $G^* = \langle \mathcal{V}, \mathcal{E} \rangle$ describes data generating mechanism

Then, in the large sample limit...

Theorem (Correctness)

For any PAG $H$ in the output of IOD, $H$ represents a set of MAGs that are Markov equivalent to $G^*_{|\mathcal{V}_1|}, G^*_{|\mathcal{V}_2|}, \ldots$.

Theorem (Completeness)

If there exists a PAG $H$ which represents a set of MAGs that are Markov equivalent to $G^*_{|\mathcal{V}_1|}, G^*_{|\mathcal{V}_2|}, \ldots$, then $H$ is in the output of IOD.
Simulations

2 Datasets:
- $|\mathcal{V}| = 8, |\mathcal{V}| = 14$
- $|\mathcal{V}_i| = 7, |\mathcal{V}_i| = 13$

3 Datasets:
- $|\mathcal{V}| = 8, |\mathcal{V}| = 14$
- $|\mathcal{V}_i| = 6, |\mathcal{V}_i| = 12$
Simulations - 2 Datasets, $|\mathcal{V}| = 14$
Simulations - 3 Datasets, $|\mathcal{V}| = 14$

![Graphs showing Precision, Recall, Runtime (seconds), and Memory usage (MB) vs. Sample size for different methods (IOD, ION) across various sample sizes.]
Application: Learning neural cascades during cognitive tasks

- 13 datasets, 160 samples
- Each missing 1-4 variable(s)
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Conclusion

Contributions / Results:

- Correct and complete algorithm for learning from multiple datasets with overlapping variables
- Results relating MAGs to sets of Markov equivalent marginalized MAGs
- Incorporated conditional independence method preventing the pathological cases for ION
- Accuracy and tractability much better than ION

Future Work / Generalizations

- Use datasets where some variables subject to intervention
- Combine with methods based on non-Gaussianity / nonlinearity
- Relax acyclicity assumption
Discussion of “Learning Equivalence Classes of Acyclic Models with Latent and Selection Variables from Multiple Datasets with Overlapping Variables”

Jiji Zhang and Ricardo Silva
Lingnan University/University College London

AISTATS 2011 – Fort Lauderdale, FL
On overlapping variables and partial information

- Three main issues in statistical learning: estimation, computation and identification

  - The problem, then: estimating Markov equivalence classes when independence assessments stop at a particular order
  - T&S can be seen as a generalization in some directions: from incomplete independence assessments to more general equivalence classes
Built-in robustness

- Incorrect decisions on qualitative information might be less likely by pooling data from different sources

“[It] has been suggested that causal discovery methods based solely on associations will find their greatest potential in longitudinal studies conducted under slightly varying conditions, where accidental independencies are destroyed and only structural independencies are preserved.” (Pearl, 2009, p.63)
On selection bias

- The assumption of common structure on separate studies also implies common selection bias

- T&S make the role of structural assumptions very clear, but the one concerning common selection bias structure might not be as believable depending on the domain
Combining different interventional studies

- Structures might change given different interventional studies
Beyond independence constraints

- Cudeck (2000): overlapping factor analysis, finding $\sigma_{15}$
Beyond independence constraints

- Less obvious: what should be done with, e.g., additive error functional constraints

Dataset I: $X_2 = f(X_1, X_5) + \text{error}$

Dataset II: $X_2 = g(X_1) + \text{error}$?
The Bayesian approach

- Latent variable models might not be necessary, neither is the iid assumption.

First likelihood function: \( L(\theta_1 \mid \text{Dataset 1}) \)

Second likelihood function: \( L(\theta_2 \mid \text{Dataset 2}) \)
Related problems: finding substructure by generalizing penalized composite likelihood?

- With/without same parameters, a single score function

\[ S(\theta \mid \text{Data 1, Data 2}) = \alpha_1 L(\theta_1 \mid \text{Data 1}) + \alpha_2 L(\theta_2 \mid \text{Data 2}) + \Omega \]
Other approaches: generalizing penalized composite likelihood?

- Enforcing constraints: link to constrained optimization

Maximize $S(\theta \mid \text{Data 1, Data 2}) = \alpha_1 L(\theta_1 \mid \text{Data 1}) + \alpha_2 L(\theta_2 \mid \text{Data 2}) + \Omega$

Subject to:

$$\text{IM}_1(X_1, X_2, X_3, X_4) = \text{IM}_2(X_1, X_2, X_3, X_4)$$

for the corresponding independent models IM on each subset