Faster algorithms for Markov equivalence

Zhongyi Hu   Robin J. Evans

Department of Statistics
University of Oxford

ISI Virtual Conference, July 2021

Algorithms

This talk is largely based on Hu and Evans (2020). In the paper we give new algorithms for:

- determining Markov equivalence in **maximal ancestral graphs**;
- projecting an ADMG to a MAG.

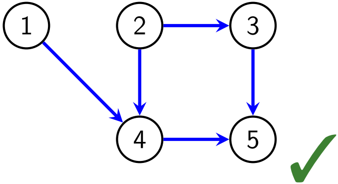
Combining these algorithms is **faster** than any previously provided method for checking Markov equivalence in ADMGs (or MAGs!).

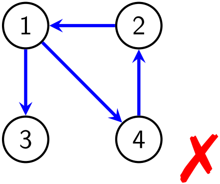
Ali et al. (2009) give an algorithm with complexity $O(ne^4)$, whereas our worst case is $O(ne^2)$.

Typically the algorithm is much faster than this.

Directed Acyclic Graph Models

A directed acyclic graph (DAG) is a directed graph that contains no (directed) cycles.





A **DAG model** imposes the factorization

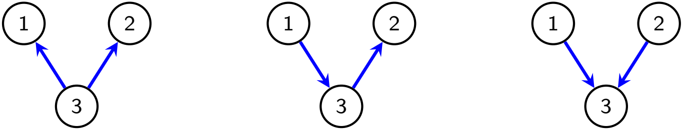
$$p(x_V) = \prod_{i \in V} p(x_i \mid x_{\mathrm{pa}(i)}).$$

where $\mathrm{pa}(i) = \{j : j \to i\}$.

Can also read off independences using **d-separation** (Verma and Pearl, 1990).

Markov Equivalence

With DAGs, sometimes two or more graphs will represent the same model:



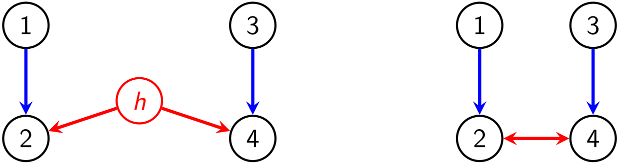
$p(x_3) \cdot p(x_1 \mid x_3) \cdot p(x_2 \mid x_3)$ $\quad$ $X_1 \perp\!\!\!\perp X_2 \mid X_3$ $\quad$ $p(x_1) \cdot p(x_2) \cdot p(x_3 \mid x_1, x_2)$

$X_1 \perp\!\!\!\perp X_2 \mid X_3$ $\quad$ $p(x_1) \cdot p(x_3 \mid x_1) \cdot p(x_2 \mid x_3)$ $\quad$ $X_1 \perp\!\!\!\perp X_2$

We can characterize this using the **skeleton** and the **unshielded colliders**.

Margins of DAG Models

DAG models are not closed under marginalization and conditioning.

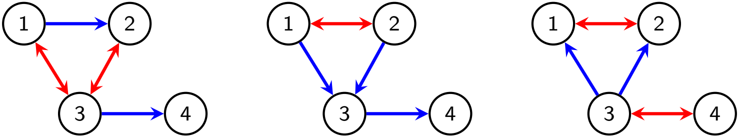


This can be solved using a **latent projection** which introduces bidirected edges ($\leftrightarrow$).

The resulting objects are called **acyclic directed mixed graphs** (ADMGs).

More Markov equivalence

All DAGs are also ADMGs, so unsurprisingly Markov equivalence is still an issue.



Maximal Ancestral Graphs (MAGs)

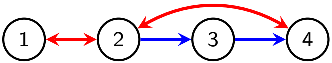
There is a subclass of ADMGs called **maximal ancestral graphs** (MAGs), introduced by Richardson and Spirtes (2002).

They are

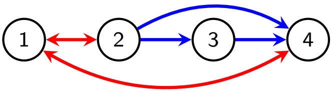
- **ancestral**: meaning no vertex is a bidirected neighbour of an ancestor; and
- **maximal**: meaning any two non-adjacent vertices have an m-separation.

Any ADMG can be 'projected' onto a **Markov equivalent** MAG.

ADMG



MAG



Main Results

We will assume that we have a MAGs with

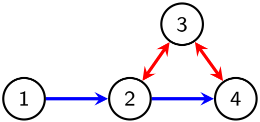
- n vertices, and
- e edges (either $\rightarrow$ or $\leftrightarrow$).

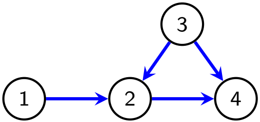
We also assume the graph is connected, so $n = O(e)$.

In this paper we present three things:

1. new non-parametric characterization of Markov equivalence class of a MAG;
2. polynomial time algorithm ($O(ne^2)$) for verifying Markov equivalence between two MAGs;
3. polynomial time algorithm ($O(n^2e)$) for transforming an ADMG into a Markov equivalent MAG.

Discriminating Paths





Both graphs have $1 \perp_m 3$.

Paths between 1 and 4 are $1 \rightarrow 2 \rightarrow 4$ and either:

$$1 \rightarrow 2 \leftrightarrow 3 \leftrightarrow 4 \qquad \text{or} \qquad 1 \rightarrow 2 \leftarrow 3 \rightarrow 4.$$

Note that for the left graph, 3 is a collider, but the right graph it is a non-collider.

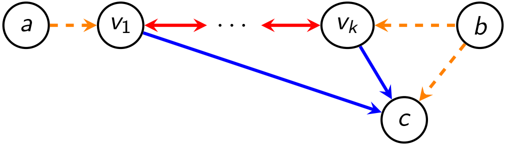
This gives us **different** m-separation sets.

$$1 \perp_m 4 \mid 2 \qquad\qquad 1 \perp_m 4 \mid 2, 3.$$

Discriminating Paths

As well as the same **skeleton** and **unshielded colliders** we need the **discriminating paths** to be the same.

A generic **discriminating path** for b is:



Note that a and c cannot be adjacent.

Since $\mathcal{G}$ is **maximal** there is (at least) one set S that m-separates a and c.

Then notice that any m-separating set S is such that

$$a \perp_m c \mid S \iff \begin{cases} b \in S \text{ and } b \text{ is a non-collider} \\ b \notin S \text{ and } b \text{ is a collider.} \end{cases}$$

Previous work

There are three graphical criteria which can be used to characterize a MAG model:

- *discriminating paths* by Spirtes and Richardson (1997);
- *minimal collider paths* (MCPs) by Zhao et al. (2005);
- *colliders with order* by Ali et al. (2009).

Using the first two approaches naïvely has exponential complexity, and the third has an $O(ne^4)$ algorithm associated with it.

Heads and Tails

Definition

A **head**, H, is a set of vertices such that:

- no vertex in H is an ancestor of any other;
- the set H is bidirected-connected within $A := \text{an}(H)$.

The corresponding **tail** is then:

$$\text{tail}_{\mathcal{G}}(H) := (\text{dis}_A(H) \setminus H) \cup \text{pa}(\text{dis}_A(H)).$$

Note that the tail is the **Markov blanket** for H within $\text{an}(H)$.

In a MAG:

- heads of size 1 are just the individual vertices;
- tails for heads of size 1 are just their parents;
- heads of size 2 just correspond to bidirected edges.

Parameterizing Sets

Define the **parametrizing sets** for a graph $\mathcal{G}$ as

$$\mathcal{S}(\mathcal{G}) := \{H \cup A : H \in \mathcal{H}(\mathcal{G}),\, A \subseteq \mathrm{tail}_{\mathcal{G}}(H)\},$$

where $\mathcal{H}(\mathcal{G})$ is the set of heads in $\mathcal{G}$.

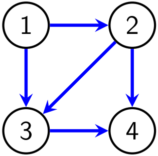
We also define:

$$\mathcal{S}_k(\mathcal{G}) := \{S \in \mathcal{S}(\mathcal{G}) : |S| \leq k\},$$

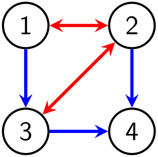
and

$$\widetilde{\mathcal{S}}_3(\mathcal{G}) = \{S \in \mathcal{S}_3(\mathcal{G}) \mid \text{1 or 2 adjacencies among vertices in } \mathcal{G}\}$$

Examples



H	T	$\mathcal{S}$
1	$\emptyset$	$\{1\}$
2	1	$\{2\}, \{1,2\}$
3	1, 2	$\{3\}, \{1,3\}$ $\{2,3\}, \{1,2,3\}$
4	2, 3	$\{4\}, \{2,4\}$ $\{3,4\}, \{2,3,4\}$



H	T	$\mathcal{S}$
1	$\emptyset$	$\{1\}$
2	$\emptyset$	$\{2\}$
1, 2	$\emptyset$	$\{1,2\}$
3	1	$\{3\}, \{1,3\}$
2, 3	1	$\{2,3\}, \{1,2,3\}$
4	2,3	$\{4\}, \{2,4\}$ $\{3,4\}, \{2,3,4\}$

Note that both these graphs have the same collection of **parametrizing sets** (i.e. all sets not containing both 1 and 4).

Preliminary Results

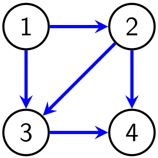
Proposition

Let $\mathcal{G}$ be a MAG and W a set of vertices. Then $W \notin \mathcal{S}(\mathcal{G})$ if and only if $\exists a, b \in W$ such that

$$a \perp_m b \mid C$$

for a set C with $W \subseteq C \cup \{a, b\}$.

Example. For the previous graphs, the sets $\{1, 4\}, \ldots, \{1, 2, 3, 4\}$ were missing, which precisely corresponds to the m-separation $1 \perp_m 4 \mid \{2, 3\}$.



Preliminary Results

Lemma

In a MAG $\mathcal{G}$, we have that there is an edge between a and b if and only if $\{a, b\} \in \mathcal{S}(\mathcal{G})$.

So we can obtain the skeleton just from $\mathcal{S}_2(\mathcal{G})$.

Lemma

If (a, b, c) is an unshielded triple, then b is a collider on this triple if and only if $\{a, b, c\} \in \mathcal{S}(\mathcal{G})$.

So we can also obtain the unshielded colliders from $\mathcal{S}_3(\mathcal{G})$.

Lemma

If π is a discriminating path from a to c for b, then b is a collider if and only if $\{a, b, c\} \in \mathcal{S}(\mathcal{G})$.

$\mathcal{S}_3(\mathcal{G})$ also gives orientations from discriminating paths!

Main Results

Theorem (Hu and Evans, 2020)

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two MAGs. These graphs are Markov equivalent if and only if $\mathcal{S}(\mathcal{G}_1) = \mathcal{S}(\mathcal{G}_2)$.

This is good, but the length of $\mathcal{S}(\mathcal{G})$ may be exponential in n.

Corollary

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two MAGs. These graphs are Markov equivalent if and only if $\mathcal{S}_3(\mathcal{G}_1) = \mathcal{S}_3(\mathcal{G}_2)$.

This is better, because now the list has length $O(n^3)$.
However, the best result we get is:

Corollary

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two MAGs. These graphs are Markov equivalent if and only if $\widetilde{\mathcal{S}}_3(\mathcal{G}_1) = \widetilde{\mathcal{S}}_3(\mathcal{G}_2)$.

Computing $\widetilde{\mathcal{S}}_3$

For a MAG $\mathcal{G}$, we can compute $\widetilde{\mathcal{S}}_3(\mathcal{G})$ by obtaining:

1. heads of size 2, and tails of each vertex;
2. tails of every head of size 2;
3. every head of size 3 in $\widetilde{\mathcal{S}}_3(\mathcal{G})$.

} unshielded colliders/ discriminating paths

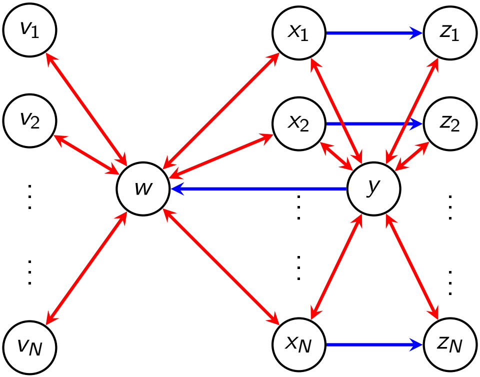
The most computationally difficult task is 3. We must:

- iterate through every bidirected edge $v \leftrightarrow w$ ($O(e)$);
- for each other vertex z ($O(n)$)...
- compute the district of v within the set of ancestors of v, w, z ($O(n + e)$) and check if z is in it.

This gives us $O(ne^2)$ complexity.

Comes from the fact that we (seem to) have to compute the ancestral district for every triple of form $v \leftrightarrow w$ and some z.

Graph achieving max complexity (Figure 5)



Consider $v_i \leftrightarrow w$ for each $i = 1, \ldots, N$, and then note that we need to consider it with each z_j for $j = 1, \ldots, N$.

These sets are all heads of size three.

There are $N^2 = O(e^2)$ such heads, which gives the required complexity.

Markov equivalence for ADMGs

We also give a method for quickly ($O(n^2 e)$) turning an ADMG into a Markov equivalent MAG.

(Note: here e is the number of edges in the MAG!).

Note that length of $\widetilde{\mathcal{S}}_3(\mathcal{G})$ is at most $O(e^2)$, so we can sort the output in $O(e^2 \log e^2) = O(e^2 \log e)$ operations.

Hence we can combine our two algorithms to check equivalence of any ADMGs in at most $O(ne^2)$.

Note it is easy to add in **undirected edges** to a MAG, and we just add those pairs to the graph to obtain $\widetilde{\mathcal{S}}_3(\mathcal{G})$.

Summary

- We can identify the Markov equivalence class of any ADMG using a purely graphical approach in at most $O(ne^2)$ complexity.
- This complexity is attained by family of graphs in Figure 5.
- There is much more to do! We can deduce some conditional independences directly from the parametrizing sets.
- There is also a connection between the parameterizing sets and **imsets** (Studený, 2006).

References

R. A. Ali, T. S. Richardson, and P. Spirtes. Markov equivalence for ancestral graphs. *Annals of Statistics*, 37(5B):2808–2837, 10 2009.

Z. Hu and R. Evans. Faster algorithms for markov equivalence. In *UAI-20*, 2020. arXiv:2007.02310.

T. S. Richardson and P. Spirtes. Ancestral graph Markov models. *Annals of Statistics*, 30(4):962–1030, 08 2002.

P. Spirtes and T. S. Richardson. A polynomial time algorithm for determining DAG equivalence in the presence of latent variables and selection bias, 1997.

M. Studený. *Probabilistic conditional independence structures*. Springer Science & Business Media, 2006.

T. Verma and J. Pearl. Causal networks: Semantics and expressiveness. In *Machine intelligence and pattern recognition*, volume 9, pages 69–76. Elsevier, 1990.

H. Zhao, Z. Zheng, and B. Liu. On the Markov equivalence of maximal ancestral graphs. *Science in China Series A: Mathematics*, 48(4):548–562, Apr 2005.

Thank you!

Latent Projection

Given an ADMG $\mathcal{G}$ with vertices $H \dot{\cup} O$, we define the **latent projection** of $\mathcal{G}$ onto O as the ADMG with vertices O and edges:

- $i \rightarrow j$ if there is a directed path from i to j in $\mathcal{G}$ with any intermediate vertices in H.
- $i \leftrightarrow j$ if there is a trek with arrows into both i and j and all intermediate vertices in H.

Definitions

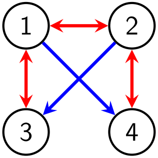
A graph $\mathcal{G}$ is called an **acyclic directed mixed graph** (ADMG) if has only directed ($\rightarrow$) and bidirected ($\leftrightarrow$) edges, and its directed edges form a DAG.

A **path** π is a sequence of distinct, adjacent vertices.

An internal vertex v on a path is called a **collider** if both adjacent edges have arrowheads at v; otherwise it is a **non-collider**.

A path from a to b is said to be **m-connecting** (or **open**) given C if:

- every non-collider is outside C;
- every collider is in $\text{an}_{\mathcal{G}}(C)$.

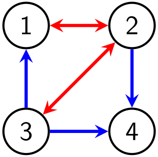


Definitions

For two vertices a, b and a disjoint set of vertices C in $\mathcal{G}$ we say that a and b are **m-separated** by C if there is no m-connecting paths given C.

Example. Consider:

- 1 and 4 given 2, 3, (m-separated);
- 1 and 4 given 2, (not m-separated).



Extension to ADMGs

Further, we show that heads and tails are preserved through the projection. Let $\mathcal{H}(\mathcal{G})$ denote the set of all heads in an ADMG $\mathcal{G}$.

Proposition

If $\mathcal{G}$ is an ADMG, $\mathcal{H}(\mathcal{G}) = \mathcal{H}(\mathcal{G}^m)$ and for every $H \in \mathcal{H}(\mathcal{G})$, $\text{tail}_{\mathcal{G}}(H) = \text{tail}_{\mathcal{G}^m}(H)$.

Independence from Parameterizing Sets

Proposition

Let $\mathcal{G}$ be a MAG and consider $W \notin \mathcal{S}(\mathcal{G})$ where $W = \{a, b\} \cup C$ such that $a, b \notin C$, if the following conditions hold:

$$(i) : \text{for any } S \supset W \text{ we have } S \in \mathcal{S}(\mathcal{G})$$
$$(ii) : \text{for any } C' \subseteq C \text{ we have } \{a, b\} \cup C' = W' \notin \mathcal{S}(\mathcal{G});$$

then $a \perp_m b \mid C$ in $\mathcal{G}$

Characteristic Imset

Note that for DAGs, the parameterizing set is the same as the **characteristic imset** (Studený, 2006).

This suggests that perhaps the parameterizing set is also the characteristic set for MAGs.

We have shown that the associated standard imset defines the model for **some** MAGs... but not all!

Indeed, the bidirected 5-cycle is a counterexample to our conjecture.

Algorithm 2: Projection from ADMG to MAGs

Algorithm 2: Obtain a MAG $\mathcal{G}^m$ from an ADMG $\mathcal{G}$

Input: ADMG $\mathcal{G}$ (list of parents and siblings of each vertex)
Output: Markov equivalent MAG $\mathcal{G}^m$

for $v \in V$ **do**
  compute $\mathrm{an}_{\mathcal{G}}(v) = \{v\} \cup \mathrm{an}_{\mathcal{G}}(\mathrm{pa}_{\mathcal{G}}(v))$;
  compute $\mathrm{tail}_{\mathcal{G}}(v) = (\mathrm{dis}_{\mathrm{an}(v)}(v) \setminus \{v\}) \cup \mathrm{pa}_{\mathcal{G}}(\mathrm{dis}_{\mathrm{an}(v)}(v))$;
  add in $p \to v$ for each $p \in \mathrm{tail}_{\mathcal{G}}(v)$;
end
for v, w *same district* **do**
  if $w \in \mathrm{dis}_{\mathrm{an}(\{v,w\})}(v)$ then add $v \leftrightarrow w$;
end

Notice first loop is $O(n(n+e))$ and second is $O(n^2(n+e))$.

Hence overall complexity is $O(n^2 e)$.

Algorithm 2: convert an ADMG to a Markov equivalent MAG

A projection from Richardson and Spirtes (2002) convert an ADMG $\mathcal{G}$ to a Markov equivalent MAG $\mathcal{G}^m$.

Lemma

Suppose $\mathcal{G}$ is an ADMG, let v, w be two vertices then (i) $v \rightarrow w$ in $\mathcal{G}^m$ if and only if $v \in \text{tail}_{\mathcal{G}}(w)$ and (ii) $v \leftrightarrow w$ in $\mathcal{G}^m$ if and only if $\{v, w\} \in \mathcal{H}(\mathcal{G})$.

For the complexity :

- upper bounded by $O(n^2 e)$;
- the average complexity is at $O(n^2)$ if each edge has an *i.i.d* Bernoulli distribution.

Definitions

Two ADMGs $\mathcal{G}_1$ and $\mathcal{G}_2$ with the same vertex sets, are said to be *Markov equivalent* if any m-separation holds in $\mathcal{G}_1$ if and only if it holds in $\mathcal{G}_2$.

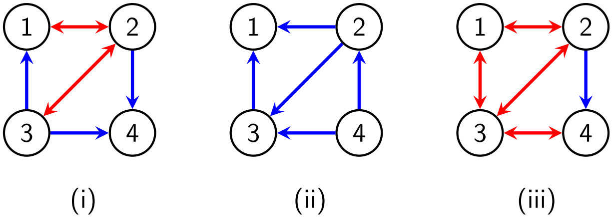


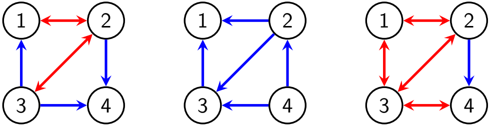
Figure: Three ADMGs where (i) and (ii) are Markov equivalent but (iii) is not.

Projection from ADMGs to Markov equivalent MAGs

Richardson and Spirtes (2002) :

- every pair of vertices $a, b \in \mathcal{V}$ in $\mathcal{G}$ that are connected by an *inducing path* becomes adjacent in $\mathcal{G}^m$;
- an edge connecting a, b in $\mathcal{G}^m$ is oriented as follows: if $a \in \text{an}_{\mathcal{G}}(b)$ then $a \to b$; if $b \in \text{an}_{\mathcal{G}}(a)$ then $b \to a$; if neither is the case, then $a \leftrightarrow b$.
- An *inducing path* between a, b is a collider path such that every collider is in $\text{an}(\{a, b\})$.
- Ancestral relations are preserved.

Examples



(i) (ii) (iii)

Figure	heads	tails	Figure	heads	tails
(i)	1	3	(iii)	1	∅
	2	∅		2	∅
	3	∅		3	∅
	4	2,3		4	2
	1,2	3		1,2	∅
	2,3	∅		1,3	∅
(ii)	1	2,3		2,3	∅
	2	4		3,4	2
	3	2,4		1,2,3	∅
	4	∅		1,3,4	2