Causal Discovery

Introduction to PAGs and the (T)FCI algorithm

Christine Winther Bang,
Department of Mathematical Sciences, October 21 2025

UNIVERSITY OF COPENHAGEN

Preparation for exercises

You can already now have a look at the page "(T)FCI exercises" and run the following code from exercise 4.1:

```
fcires <- fci(nlsdata, sparsity = 0.05)
```

And exercise 5.1:

```
tfcires <- tfci(nlsdata, sparsity = 0.05,
                order = c("r1", "r6", "r12"))
```

These functions are **very** slow!

No unmeasured confounding?

Until now: DAGs, CPDAGs and MPDAGs $\Rightarrow$ Assume no unmeasured (aka unobserved or latent) confounders.

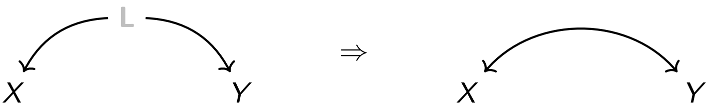
This is a strong assumption!

No unmeasured confounding?

Until now: DAGs, CPDAGs and MPDAGs $\Rightarrow$ Assume no unmeasured (aka unobserved or latent) confounders.

This is a strong assumption!

Now: Use *bidirected* edges to represent unmeasured confounding:



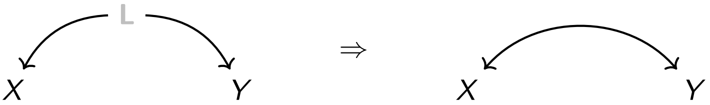
Unmeasured confounding

Now: Use *bidirected* edges to represent unmeasured confounding:

$$X \leftarrow L \rightarrow Y \quad \Rightarrow \quad X \leftrightarrow Y$$

Unmeasured confounding

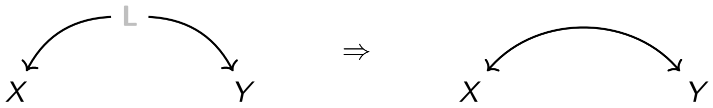
Now: Use *bidirected* edges to represent unmeasured confounding:



Why do we need bidirected edges?

Unmeasured confounding

Now: Use *bidirected* edges to represent unmeasured confounding:

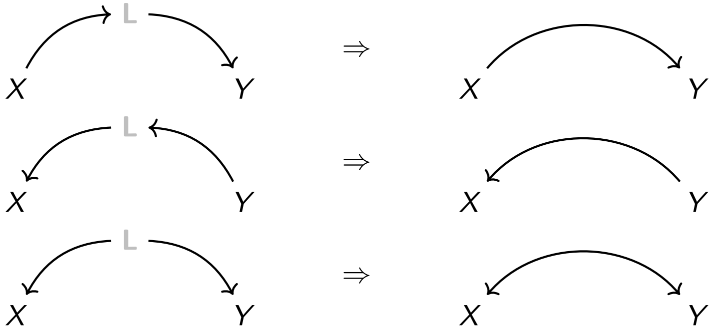


Why do we need bidirected edges?

- When constructing expert graphs we can include "unobserved confounders".
- When *estimating* a graph we do not know beforehand where to add unobserved confounders!

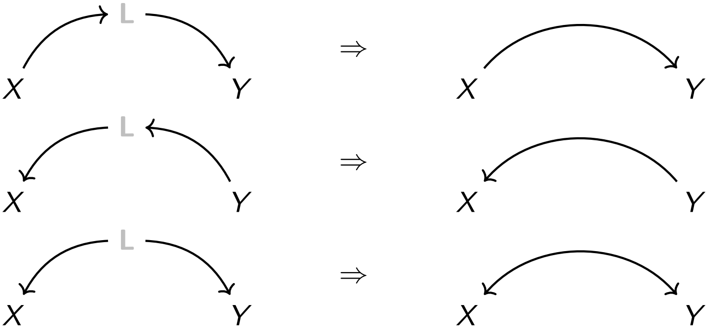
Unmeasured confounding

We can represent both latent confounders and mediators:



Unmeasured confounding

We can represent both latent confounders and mediators:

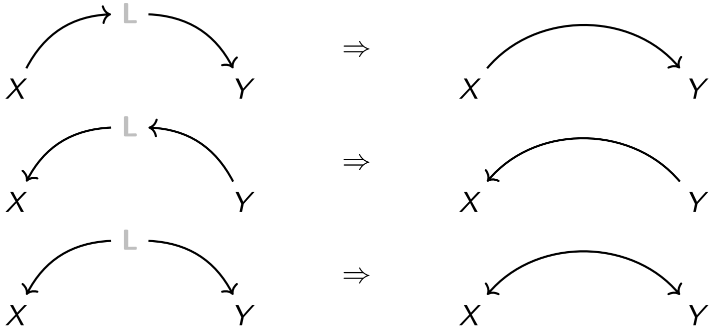


$X \not\perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y \mid L$

$X \not\perp\!\!\!\perp Y$

Unmeasured confounding

We can represent both latent confounders and mediators:

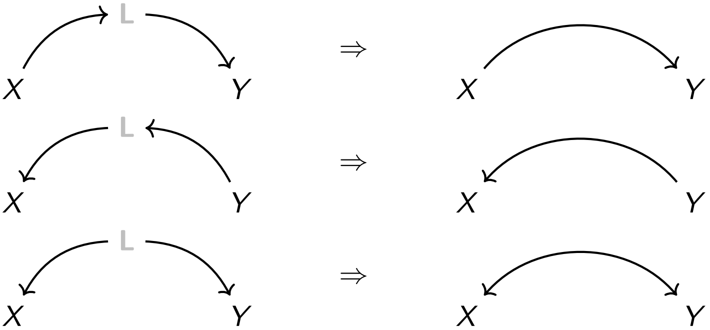


$X \not\perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y \mid L$

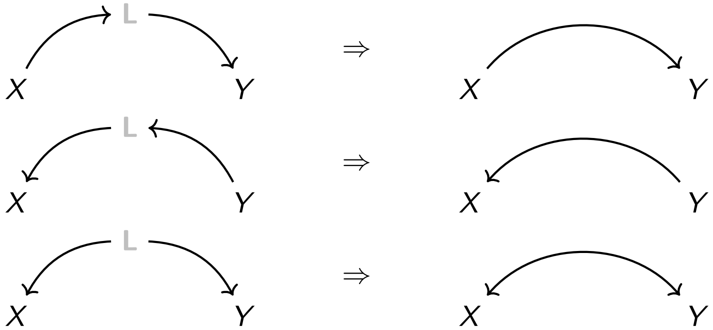
$X \not\perp\!\!\!\perp Y$

If we only observe X and Y we cannot distinguish the six DAGs!

No unmeasured confounding?



No unmeasured confounding?



$X \not\perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y \mid L$

$X \not\perp\!\!\!\perp Y$

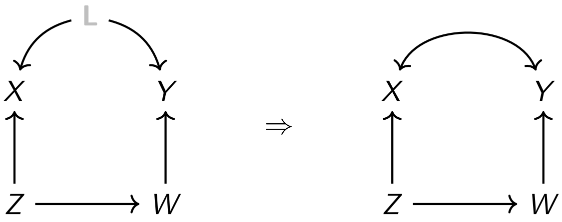
Here we are *marginalising out* the node L. DAGs are not closed under marginalisation $\Rightarrow$ we need another type of graph.

Maximal Ancestral Graphs (MAGs)

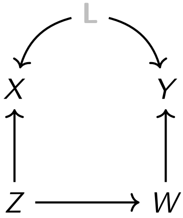
- May contain *directed* ($\rightarrow$) and *bidirected* ($\leftrightarrow$) edges. These imply the following in the underlying DAG:
  - $X \rightarrow Y$: X is a parent or ancestor of Y.
  - $X \leftrightarrow Y$: X is not a parent or ancestor of Y and Y is not an parent or ancestor of X.
- Has no directed cycles: We cannot have both $X \rightarrow \ldots \rightarrow Y$ and $X \leftarrow \ldots \leftarrow Y$.
- Is *ancestral*: If $X \leftrightarrow Y$ then we cannot have $X \rightarrow \ldots \rightarrow Y$ or $X \leftarrow \ldots \leftarrow Y$.
- Is *maximal*: No edges can be added without changing the independence model

m-separation in MAGs generalises *d-separation* in DAGs.

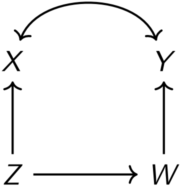
Examples

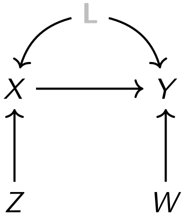


Examples

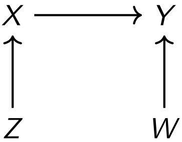


$\Rightarrow$





$\Rightarrow$



Partial Ancestral Graphs (PAGs)

Multiple MAGs can encode same independence model $\Rightarrow$ we can at best estimate an equivalence class. This is represented by a PAG.

MAGs include two types of edges (directed and bidirected). We might know one *edgemark* but not the other. PAGs can have three types of edgemarks with the following interpretations:

Tail ($-$): this tail is present in all MAGs in the equivalence class.

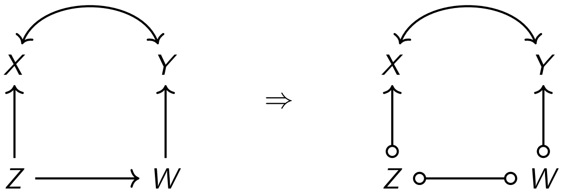
Arrowhead ($>$): this arrowhead is present in all MAGs in the equivalence class.

Circle mark ($\circ$): at least one MAG in the equivalence class has an edge where this is a tail, and at least one has an edge where this is an arrowhead.

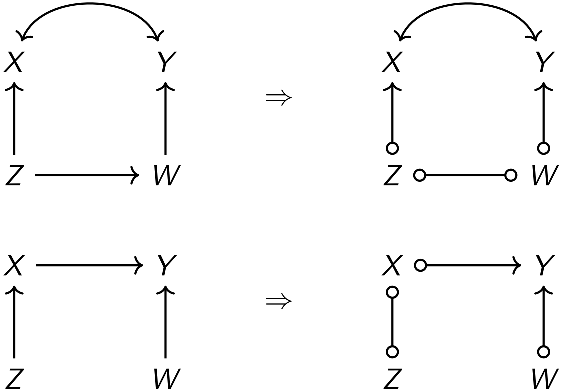
Edges in PAGs

Edge type	**Interpretation**
Directed $X \rightarrow Y$	X is an ancestor (or parent) of Y, and there may be unobserved confounding between.
Bidirected $X \leftrightarrow Y$	There is unobserved confounding between X and Y and no ancestral ancestral relation.
Partially dir. $X \circ\!\!\rightarrow Y$	Either $X \rightarrow Y$ or $X \leftrightarrow Y$.
Non-directed $X \circ\!\!-\!\!\circ Y$	Either $X \rightarrow Y$, $X \leftarrow Y$ or $X \leftrightarrow Y$.

Examples



Examples



The Fast Causal Inference (FCI) algorithm

Proceeds as the PC but has two additional steps (to find correct skeleton and v-structures) and need more orientation rules (to output the maximally informative PAG).

(I) Start with fully connected graph.

(II) Remove adjacencies for pairs of nodes conditionally independent given a subset of the adjacent nodes.

(III) Orient v-structures.

(IV) Remove adjacencies for pairs of nodes conditionally independent given a subset of the possible separating sets.

(V) Orient v-structures.

(VI) Find more arrowheads and tails using orientation rules.

(Spirtes et al. 1999, Ali et al. 2005, Zhang 2008)

The temporal FCI (TFCI) algorithm

Modification of the FCI algorithm. Uses temporal/tiered background knowledge in two ways:

1. Skips independence tests conditioning on the future.
2. Orients cross-tier edges.
   (a) As directed edges if there is no unobserved confounding across tiers (Andrews et al. 2020).
   (b) As partially directed edges otherwise (Bang & Didelez 2025).

(b) needs more orientation rules than the original FCI algorithm (Wang et al. 2024, Venkateswaran & Perković 2024)

Does not necessarily output a graph that represents a (restricted) equivalence class!

References

Ali, R. A., Richardson, T., Spirtes, P., & Zhang, J. (2005). Orientation rules for constructing markov equivalence classes of maximal ancestral graphs. Technical Report 476, Dept. of Statistics, University of Washington.

Andrews, B., Spirtes, P., & Cooper, G. F. (2020). On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. In International Conference on Artificial Intelligence and Statistics. PMLR, 4002-4011.

Bang, C. W. & Didelez, V. (2025). Constraint-based causal discovery with tiered background knowledge and latent variables in single or overlapping datasets Proceedings of the Fourth Conference on Causal Learning and Reasoning. PMLR, 1116-1146.

Richardson, T., & Spirtes, P. (2002). Ancestral graph Markov models. The Annals of Statistics, 30(4), 962-1030.

Spirtes P., Meek C., & Richardson T. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In: Glymour C, Cooper G, editors. Computation, Causation, and Discovery. MIT press, 211-52.

Venkateswaran, A., & Perković, E. (2024). Towards complete causal explanation with expert knowledge. arXiv preprint arXiv:2407.07338.

Wang, T. Z., Tao, L., & Zhou, Z. H. (2024). New rules for causal identification with background knowledge. arXiv preprint arXiv:2407.15259.

Zhang, J. (2008). On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. Artificial Intelligence, 172(16-17), 1873-1896.