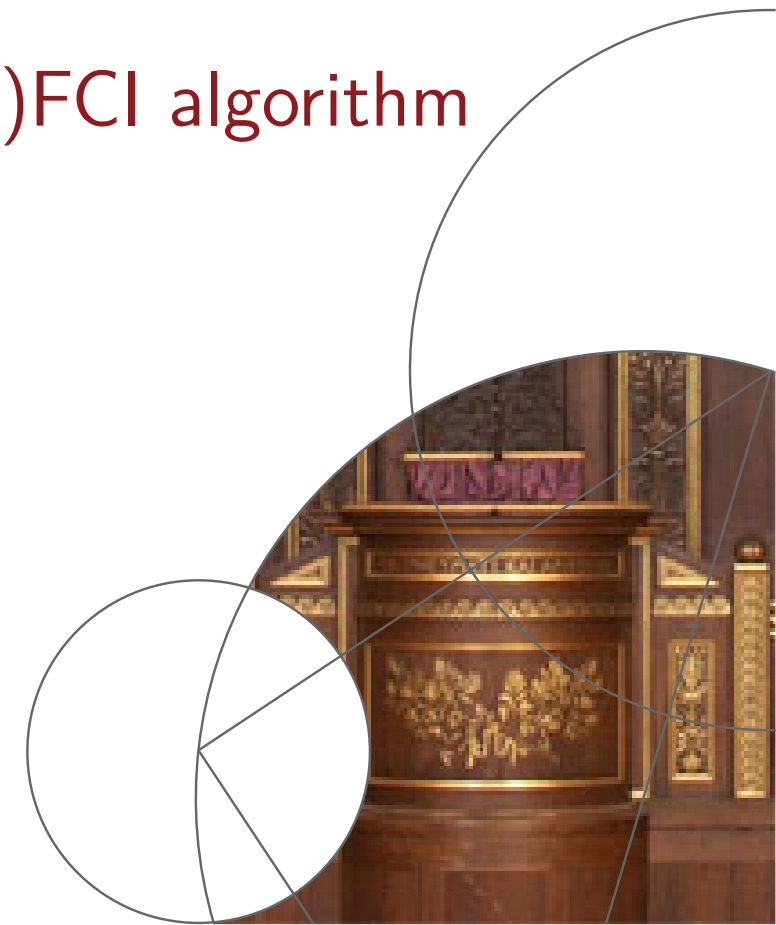


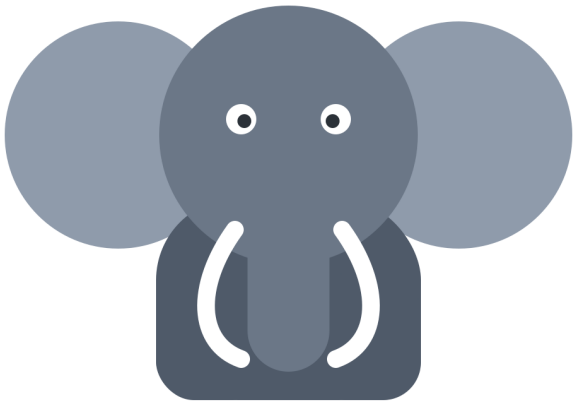


Introduction to PAGs and the (T)FCI algorithm

Anne Helby Petersen



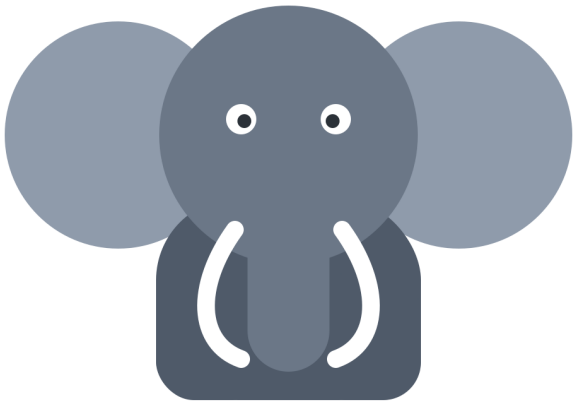
The elephant in the room



We have assumed **no unobserved confounding** so far.



The elephant in the room

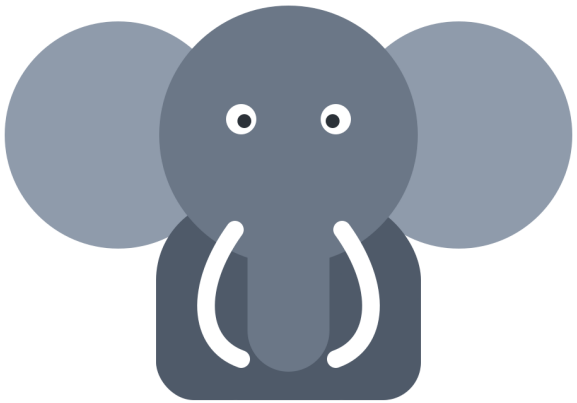


We have assumed **no unobserved confounding** so far.

This is not reasonable for the example considered today – nor for most observational health science studies.



The elephant in the room



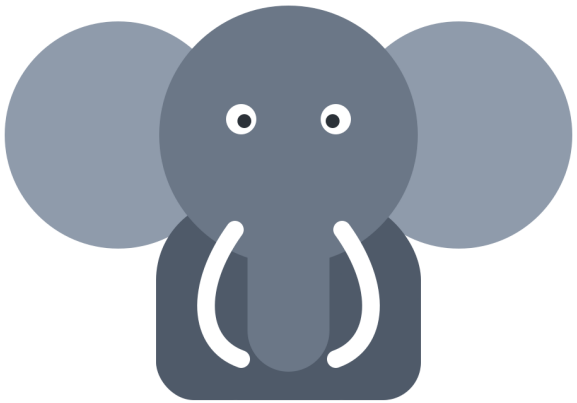
We have assumed **no unobserved confounding** so far.

This is not reasonable for the example considered today – nor for most observational health science studies.

Question for the remainder of the day: *What can we learn from data without this assumption?*



The elephant in the room



We have assumed **no unobserved confounding** so far.

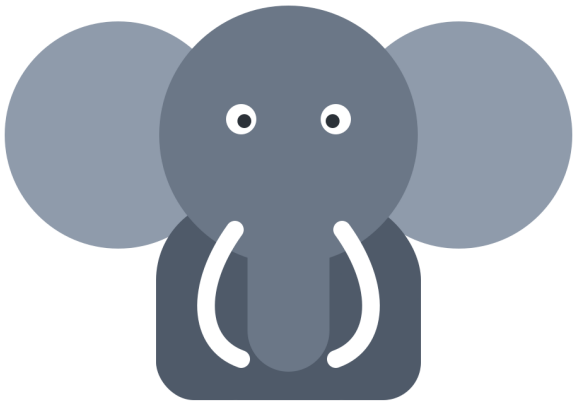
This is not reasonable for the example considered today – nor for most observational health science studies.

Question for the remainder of the day: *What can we learn from data without this assumption?*

Answer: Something!



The elephant in the room



We have assumed **no unobserved confounding** so far.

This is not reasonable for the example considered today – nor for most observational health science studies.

Question for the remainder of the day: *What can we learn from data without this assumption?*

Answer: Something! But (of course!) a lot less.



Out of scope: unobserved selection

There is a second kind of unobserved variable: unobserved **selection** (conditioning on an unobserved collider).

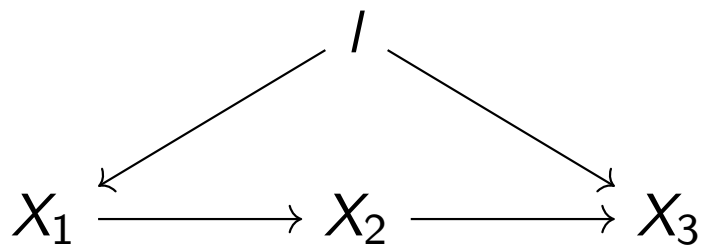
Example: We study smoking and prostate cancer, but only include people who are alive at age 80, to get "enough" cancer cases. Both smoking and prostate cancer affect death, so we have conditioned on a collider.

In this session we assume **no unobserved selection**, only unobserved confounding. Possible to extend results presented here to case w/ unobs. selection, but complicated enough already...



Representing unobserved confounding

Idea: Add (specific) unobserved confounders as latent variables:



The usual DAG calculus still works (d-separation, back-door and front-door criteria), but we can only condition on observed variables ($=$ not I).



DAGs with latent variables

Pros:

- Good for describing our **causal assumptions**.
- Can be used to deduce which effects are **identifiable** under those assumptions.
- Can guide the **estimation** procedure.



DAGs with latent variables

Pros:

- Good for describing our **causal assumptions**.
- Can be used to deduce which effects are **identifiable** under those assumptions.
- Can guide the **estimation** procedure.

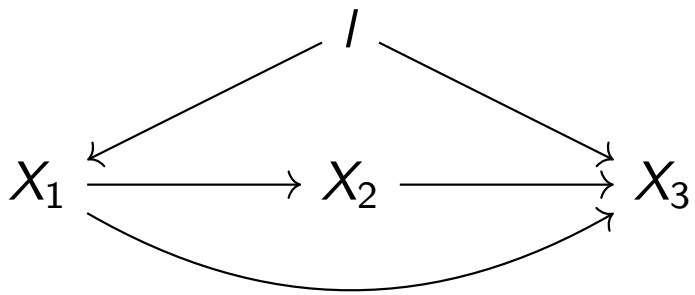
Cons:

- Requires one to have a **complete understanding of latent (unobserved) structure** – we have to "know what we don't know (observe)".
- **Cannot be learned from data**. Very different DAGLs give rise to the same sets of conditional independences.
    - This is *not* a power or sample size issue. It is a fundamental mathematical limitation.



What can we then learn from data?

Overall issue: When there is unobserved confounding, we cannot generally distinguish between direct and indirect effects.

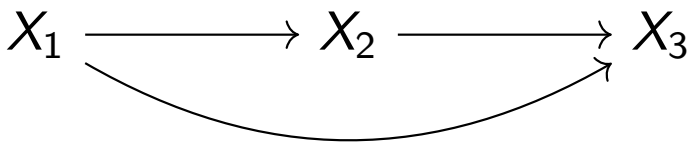


What we can (sometimes) learn: Ancestral relationships, i.e. whether X is causally "prior" (ancestor) to Y.

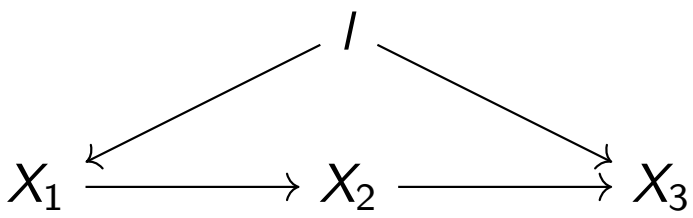


Five DAGLs, same observed ancestral relationships

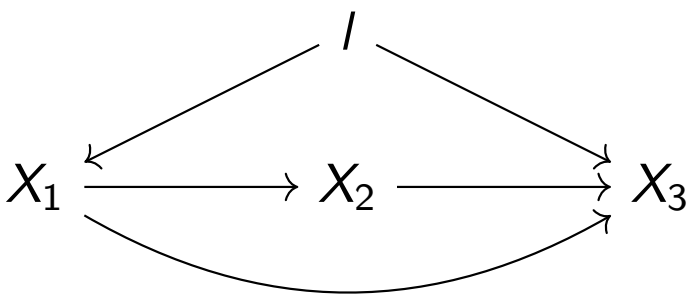
A



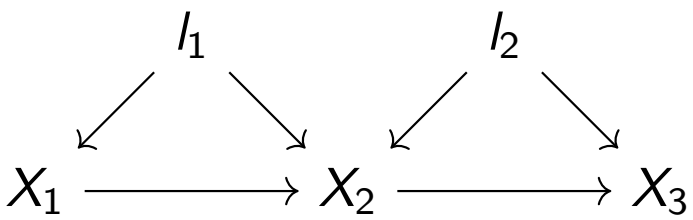
B



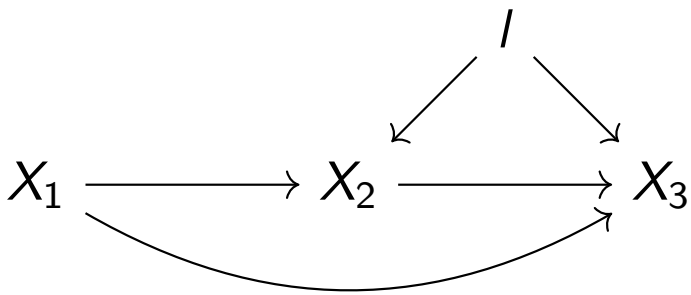
C



D



E



5 DAGs in 1: The maximal ancestral graph

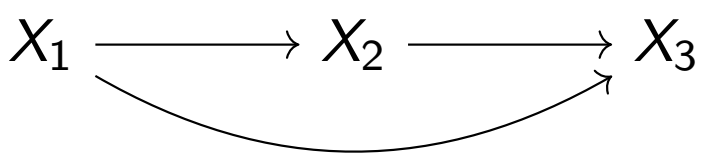
- For a given DAG(L) there exists a unique graphs that summarizes all ancestral relations in the observed variables with as few edges as possible. This is the **maximal ancestral graph** (MAG).
- **Edge interpretation** is different from DAGs: In a MAG, $X \to Y$ means that X is an ancestor of Y, but not necessarily that there is a direct effect.



5 DAGs in 1: The maximal ancestral graph

- For a given DAG(L) there exists a unique graphs that summarizes all ancestral relations in the observed variables with as few edges as possible. This is the **maximal ancestral graph** (MAG).
- **Edge interpretation** is different from DAGs: In a MAG, $X \to Y$ means that X is an ancestor of Y, but not necessarily that there is a direct effect.

MAG (for all of A–E)

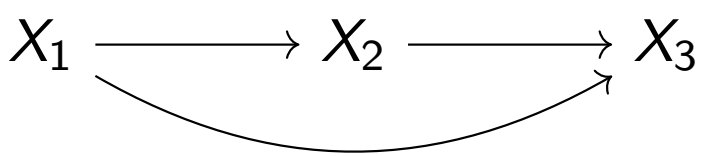




5 DAGs in 1: The maximal ancestral graph

- For a given DAG(L) there exists a unique graphs that summarizes all ancestral relations in the observed variables with as few edges as possible. This is the **maximal ancestral graph** (MAG).
- **Edge interpretation** is different from DAGs: In a MAG, $X \to Y$ means that X is an ancestor of Y, but not necessarily that there is a direct effect.

MAG (for all of A–E)

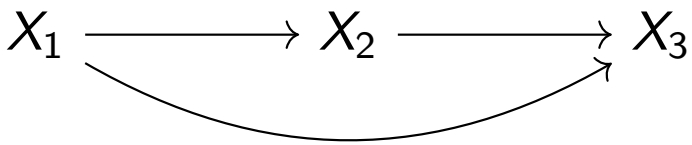


Note: The five DAGs clearly do **not** describe the same causal mechanism, but they share the same ancestral relationships.

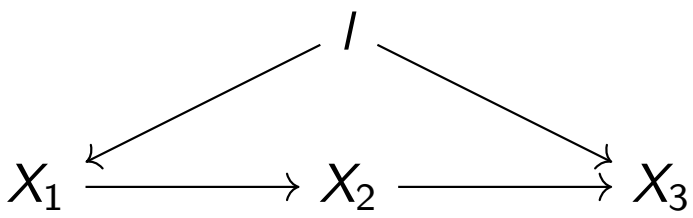


Five DAGLs, same observed ancestral relationships

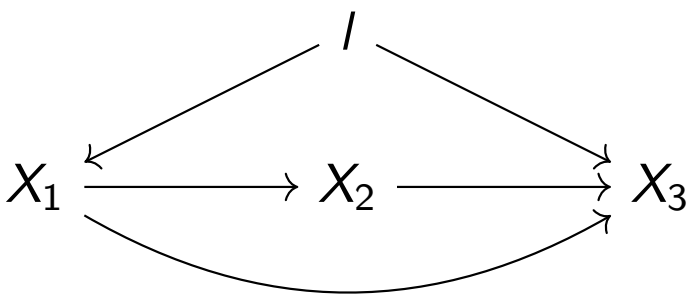
A



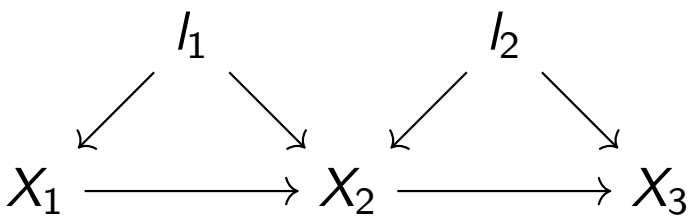
B



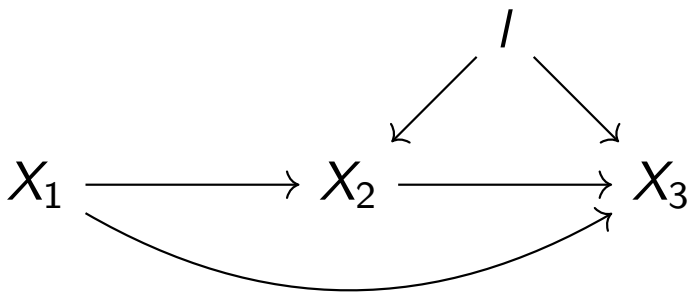
C



D



E



Same MAG, different causal implications

Effect identifiability when only observing X_1, X_2 and X_3:

DAG	**Direct effect**		**Total effect**
	$X_1 \to X_2$	$X_2 \to X_3$	$X_1 \to X_3$	$X_1 \to X_3$
A	Identified w/o adjustment	Back-door adjustment for X_1	Identified w/o adjustment	Identified w/o adjustment
B	Identified w/o adjustment	Back-door adjustment for X_1	Identified w/o adjustment	Front-door adjustment for X_2
C	Identified w/o adjustment	Back-door adjustment for X_1	Not identifiable	Not identifiable
D	Not identifiable	Not identifiable	Not identifiable	Not identifiable
E	Identified w/o adjustment	Not identifiable	Not identifiable	Identified w/o adjustment



Same MAG, different causal implications

Effect identifiability when only observing X_1, X_2 and X_3:

DAG	**Direct effect**			**Total effect**
	$X_1 \to X_2$	$X_2 \to X_3$	$X_1 \to X_3$	$X_1 \to X_3$
A	Identified w/o adjustment	Back-door adjustment for X_1	Identified w/o adjustment	Identified w/o adjustment
B	Identified w/o adjustment	Back-door adjustment for X_1	Identified w/o adjustment	Front-door adjustment for X_2
C	Identified w/o adjustment	Back-door adjustment for X_1	Not identifiable	Not identifiable
D	Not identifiable	Not identifiable	Not identifiable	Not identifiable
E	Identified w/o adjustment	Not identifiable	Not identifiable	Identified w/o adjustment

$\Rightarrow$ The same ancestral information is compatible with very different causal structures, e.g. for mediation analysis.



Let's then estimate a MAG, I guess?



Let's then estimate a MAG, I guess?

... Yeah, no, we can't. They're "too informative". Several MAGs give rise to the same conditional independences.



A realistic goal: Partial ancestral graphs (PAGs)

- For DAGs, we had to change our objective to CPDAGs. For MAGs, we have to change our objective to **PAGs**.
- A partial ancestral graph (PAG) represents a **family (equivalence class) of MAGs**, i.e., all MAGs that share the same conditional independences.
- A PAG shows which variables are **causally connected** and hence which are definitely *not*.
- A PAG cannot always determine neither 1) the **direction** of ancestral relationships, nor 2) whether an association is **confounding** or an ancestral relation.
- PAGs can be **estimated from data** under reasonable assumptions. But they are difficult to interpret and often not very informative. . .



The four edge types of a PAG

$X_1 \to X_2$ X_1 is an ancestor of X_2, and there may also be unobserved confounding.

$X_1 \leftrightarrow X_2$ Unobserved confounding, but no ancestral relationship in either direction.

$X_1 \circ\!\!\to X_2$ X_1 is an ancestor of X_2, or there is unobserved confounding, or both.

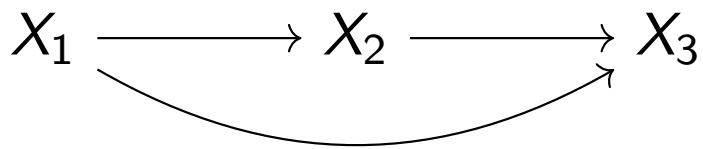
$X_1 \circ\!\!-\!\!\circ X_2$ There is some relationship between X_1 and X_2; either may be an ancestor of the other, with or without confounding.

If no edge: Neither is an ancestor of the other and there is no unobserved confounder.



From MAG to PAG

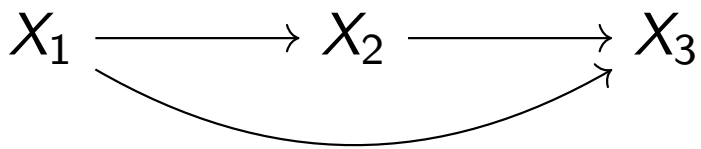
MAG for all of A–E



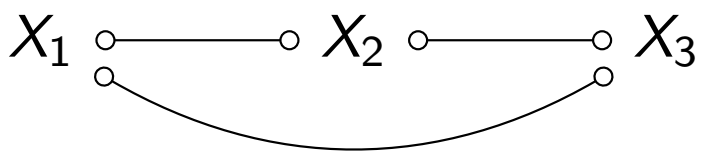


From MAG to PAG

MAG for all of A–E



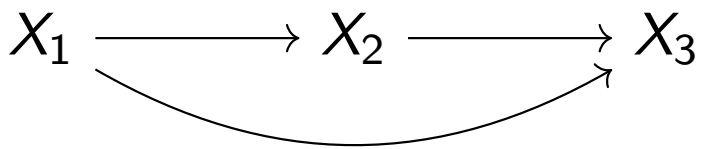
PAG for all of A–E



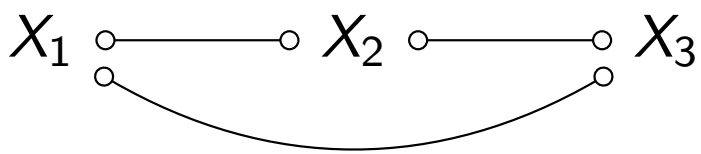


From MAG to PAG

MAG for all of A–E



PAG for all of A–E

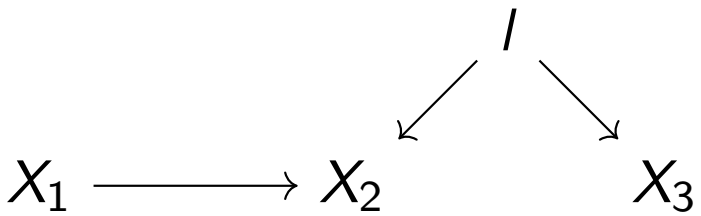


This is a **completely uninformative** graph: fully connected, and every edge undetermined. Learning this PAG tells us nothing.

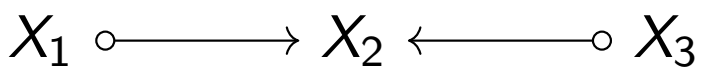


A more informative PAG

DAG with latent I



Corresponding PAG



What the PAG says:

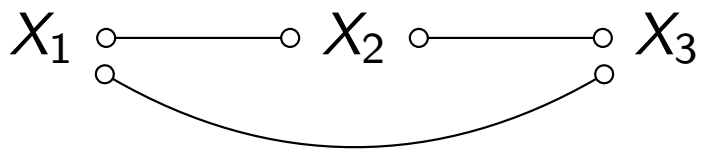
- No edge between X_1 and X_3: Neither is an ancestor of the other.
- $X_1 \circ\!\!\rightarrow X_2$: X_2 is *not* an ancestor of X_1. If related causally, X_1 is the cause. It may also be confounding only, or both.
- $X_2 \leftarrow\!\!\circ X_3$: X_2 is *not* an ancestor of X_3. If related causally, X_3 is the cause. It may also be confounding only, or both.

Note: This is the **totality** of what conditional independences alone can tell us about this causal mechanism.



Or time may help

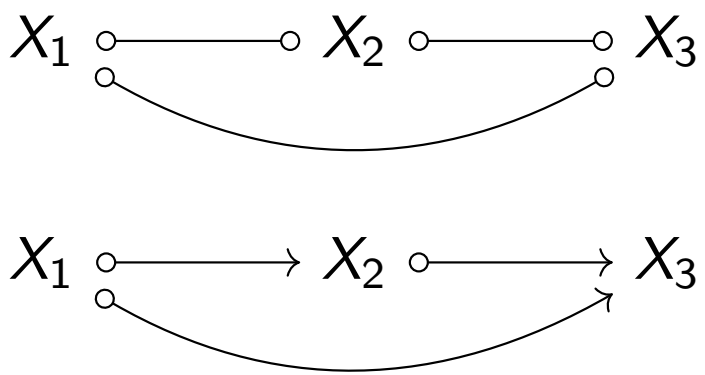
Assume X_1 is prior to X_2, which is prior to X_3. Then we can restrict the PAG:





Or time may help

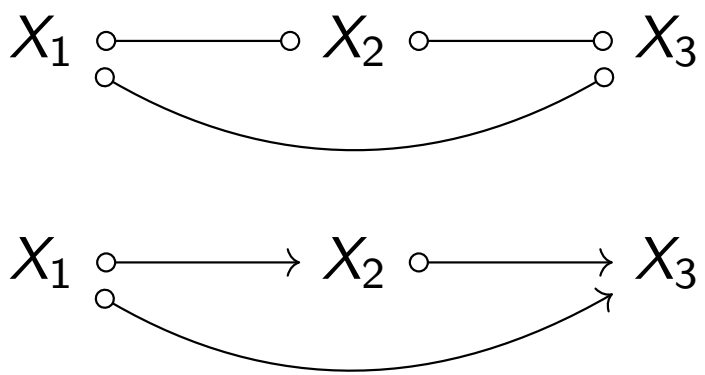
Assume X_1 is prior to X_2, which is prior to X_3. Then we can restrict the PAG:





Or time may help

Assume X_1 is prior to X_2, which is prior to X_3. Then we can restrict the PAG:

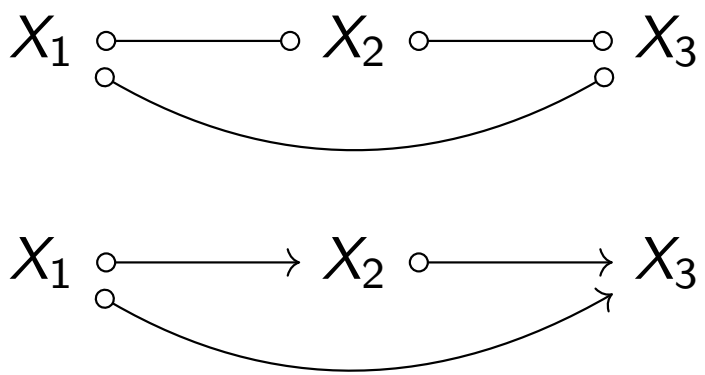


Open (mathematical) research question whether or not this is the totality of information that can be derived from the temporal background knowledge + conditional independences.



Or time may help

Assume X_1 is prior to X_2, which is prior to X_3. Then we can restrict the PAG:



Open (mathematical) research question whether or not this is the totality of information that can be derived from the temporal background knowledge + conditional independences.

In either case, as you will see in exercises: Appears useful in practice.



Summary: DAGs vs. MAGs and PAGs

- **DAGs with latent variables** are intuitive, but cannot be recovered from observational data, no matter the sample size + requires a thorough specification of latent (unobserved) structure.
- **MAGs and PAGs** (MAGs) encode ancestral relations (total effects), but cannot generally distinguish between direct vs. indirect effects
- **PAGs** summarize what is learnable in the presence of **arbitrary unobserved confounding**.
  - The possibility of confounding is built into the graph structure $\Rightarrow$ limits how detailed the description can be.
  - (Temporal) background knowledge appears to help. But mathematical theory still being developed.
  - Much less intuitive to interpret. Advice: Focus on what the graph rules out, rather than what it states.



The Fast Causal Inference (FCI) algorithm

Extension of PC for learning PAGs:

1. Start with fully connected graph.
2. Remove adjacencies for pairs of nodes conditionally independent given a subset of the adjacent nodes.
3. Orient v-structures.
4. Remove adjacencies for pairs of nodes conditionally independent given a subset of the possible separating sets.
5. Orient v-structures.
6. Apply additional orientation rules (more than for PC).

Mathematically correct and complete (for infinite data + perfect test). (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.
    1. As directed edges if assuming no unobserved confounding across tiers (Andrews et al. 2020).
    2. As partially directed edges otherwise (Bang & Didelez 2025).
3. Needs more orientation rules than the original FCI algorithm (Wang et al. 2024, Venkateswaran & Perković 2024)[1]

[1] Open question what and how many rules are needed to ensure the output is maximally informative. But we know it's "sound" (makes no false statements if we had infinite data + perfect statistical test) using the original FCI orientation rules only (Bang & Didelez 2025).



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.
    1. As directed edges if assuming no unobserved confounding across tiers (Andrews et al. 2020).
    2. As partially directed edges otherwise (Bang & Didelez 2025).
3. Needs more orientation rules than the original FCI algorithm (Wang et al. 2024, Venkateswaran & Perković 2024)[1]

[1] Open question what and how many rules are needed to ensure the output is maximally informative. But we know it's "sound" (makes no false statements if we had infinite data + perfect statistical test) using the original FCI orientation rules only (Bang & Didelez 2025).



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.

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.

