Causal Discovery for Max-Linear Bayesian Networks

Francesco Nowell

TU Berlin

20.08.2025

About

A PC Algorithm for Max-Linear Bayesian Networks (FN+ 2025)

https://arxiv.org/abs/2508.13967

Carlos Améndola (TU Berlin)

Ben Hollering (MPI-MiS Leipzig)

Preliminaries

What is Causal Discovery?

Structural Equation Models

Let $\mathcal{G} = (V, E)$ be a directed acyclic graph with $V = \{1, \dots n\}$.
A random vector $X = (X_1, \dots X_n)$ is distributed according to a structural equation model on $\mathcal{G}$ if

$$X_i = f_i(X_{\mathrm{pa}(i)}, \varepsilon_i) \ ,$$

where $\mathrm{pa}(i)$ is the set of parents of i and ε_i is the (Gaussian) error at i.

Structural Equation Models

Let $\mathcal{G} = (V, E)$ be a directed acyclic graph with $V = \{1, \dots n\}$.
A random vector $X = (X_1, \dots X_n)$ is distributed according to a structural equation model on $\mathcal{G}$ if

$$X_i = f_i(X_{\mathrm{pa}(i)}, \varepsilon_i) \ ,$$

where $\mathrm{pa}(i)$ is the set of parents of i and ε_i is the (Gaussian) error at i.

Example (Linear structural equation model)

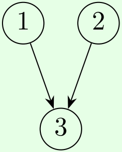


Figure: The collider DAG

$$X_1 = \varepsilon_1$$
$$X_2 = \varepsilon_2$$
$$X_3 = c_{13}X_1 + c_{23}X_2 + \varepsilon_3$$

Intuitively: Arrows represent causal relationships.

Causal Discovery

TASK: Given data which comes from a SEM X on $\mathcal{G}$, recover $\mathcal{G}$.

IDEA: relate conditional independence in X to combinatorial *separation criteria* in $\mathcal{G}$.
Example: $d-$separation for linear SEMs (Verma and Pearl, 1990 [6])

Causal Discovery

TASK: Given data which comes from a SEM X on $\mathcal{G}$, recover $\mathcal{G}$.

IDEA: relate conditional independence in X to combinatorial *separation criteria* in $\mathcal{G}$.
Example: $d-$separation for linear SEMs (Verma and Pearl, 1990 [6])

Definition

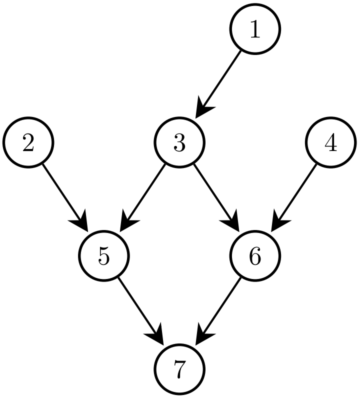
Two nodes $i, j \in V$ in a DAG are *$d-$connected* given $K \subset V \setminus ij$ if there exists an undirected path π from i to j such that:

- Any center node of any *collider* along π lies in $K \cup an(K)$
- No non-collider along π lies in K.

If no $d-$connecting path exists, we write $[i \perp_d j | K]$ and say that i and j are *$d-$separated* given K.

d-separation:examples

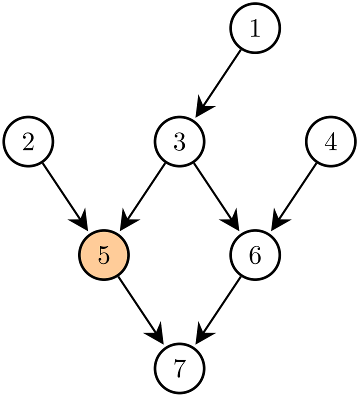
$$K = \emptyset$$



Connected: 2 and 5, 2 and 7, 5 and 6
Separated: 2 and 3, 3 and 4, 1 and 4

d-separation:examples

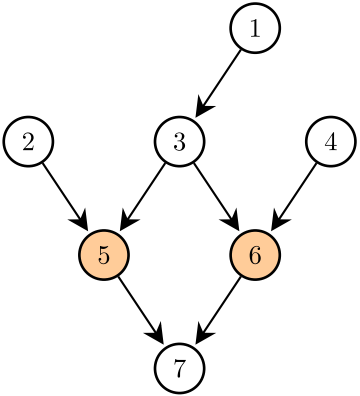
$$K = \{5\}$$



Connected: 2 and 3, 2 and 1, 1 and 7
Separated: 2 and 7, 2 and 4, 1 and 4

d-separation:examples

$$K = \{5, 6\}$$



Connected: 2 and 4, 1 and 4, 1 and 2
Separated: 3 and 7, 1 and 7

Linear SEMs and d-separation

Theorem

Linear SEMs are faithful to d−separation, i.e:

$$[X_i \perp\!\!\!\perp X_j \mid X_K] \;\; \textit{holds in } X \quad \Longleftrightarrow \quad [i \perp_d j | K] \;\; \textit{holds in } \mathcal{G} \tag{1}$$

for any X distributed according to a linear SEM on $\mathcal{G}$.

Equivalently, the entire CI structure of X is encoded in its *d-separation Global Markov property*.

$$\text{global}_d(\mathcal{G}) := \Big\{ [i \perp\!\!\!\perp j \mid K] \quad \text{s.t} \;\; [i \perp_d j | K] \text{ holds in } \mathcal{G} \Big\}. \tag{2}$$

The PC algorithm (Spirtes and Glymour,[5])

Constraint based causal discovery algorithm.
Input: A method for testing CI in a distribution X on $\mathcal{G}$ faithful to $\perp_d$.
(equivalently: $\text{global}_d(\mathcal{G})$)
Output: A partially oriented graph approximating $\mathcal{G}$

Step 1: Reconstruct the undirected skeleton of $\mathcal{G}$ by querying $\text{global}_d(\mathcal{G})$. (Skeleton Retrieval)

Step 2: Orient the unshielded colliders in the skeleton (Edge Orientation)

The PC algorithm (Spirtes and Glymour,[5])

Constraint based causal discovery algorithm.
Input: A method for testing CI in a distribution X on $\mathcal{G}$ faithful to $\perp_d$.
(equivalently: $\mathrm{global}_d(\mathcal{G})$)
Output: A partially oriented graph approximating $\mathcal{G}$

Step 1: Reconstruct the undirected skeleton of $\mathcal{G}$ by querying $\mathrm{global}_d(\mathcal{G})$. (Skeleton Retrieval)

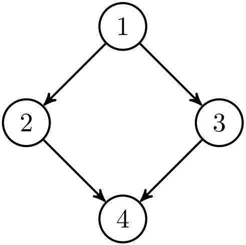
Step 2: Orient the unshielded colliders in the skeleton (Edge Orientation)

Theorem

PC outputs a representative of the Markov equivalence class of $\mathcal{G}$. Its worst-case complexity is in $\mathcal{O}(n^{d+2})$, where $n = |V|$ and $d := \max_{v \in V} \mathrm{indeg}(v)$.

PC algorithm example: Skeleton retrieval

Consider the *diamond* DAG $\mathcal{G}$



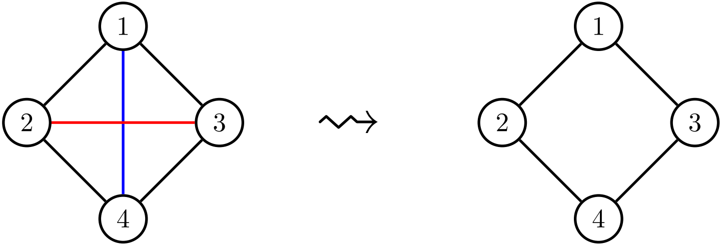
and its $d-$separation global markov property

$$\text{global}_d(\mathcal{G}) := \left\{ [2 \perp\!\!\!\perp 3 \mid 1] \, , [1 \perp\!\!\!\perp 4 \mid 23] \right\}$$

PC algorithm example: Skeleton retrieval

$$\text{global}_d(\mathcal{G}) := \Big\{ \; [2 \perp\!\!\!\perp 3 \mid 1] \;\; , [1 \perp\!\!\!\perp 4 \mid 23] \; \Big\}$$

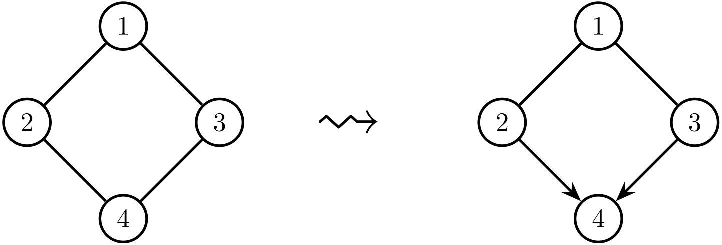
"Start with the complete graph and delete the edge $\{i, j\}$ whenever $[i \perp\!\!\!\perp j \mid K] \in \text{global}_d(\mathcal{G})$ for some K"



PC algorithm example: Edge orientation

$$\text{global}_d(\mathcal{G}) := \left\{ \; [2 \perp\!\!\!\perp 3 \mid 1] \;\; , [1 \perp\!\!\!\perp 4 \mid 23] \; \right\}$$

"For any unshielded triple $\{i, j, k\}$:
orient as $i \to j \leftarrow k$ if $[i \perp\!\!\!\perp j \mid k] \notin \text{global}_d(\mathcal{G})$."



What is a Max-Linear Bayesian Network?

Max-Linear Bayesian Networks (MLBNs)

Let $\mathcal{G}$ be DAG on n nodes with edge weights $c_{ij} \geq 0$ for $i \to j \in \mathcal{G}$.
A random vector $X = (X_1, \ldots X_n)$ is distributed according to a *max-linear model* on $\mathcal{G}$ if

$$X_i = \bigvee_{j \in \mathrm{pa}(i)} c_{ij} X_j \vee Z_i, \qquad c_{ij}, Z_i \geq 0 \tag{3}$$

where $\vee = \max$, $\mathrm{pa}(i)$ is the set of parents of i in $\mathcal{G}$, and the Z_i are independent, atom-free, continuous random variables.

Max-Linear Bayesian Networks (MLBNs)

Let $\mathcal{G}$ be DAG on n nodes with edge weights $c_{ij} \geq 0$ for $i \to j \in \mathcal{G}$.
A random vector $X = (X_1, \dots X_n)$ is distributed according to a *max-linear model* on $\mathcal{G}$ if

$$X_i = \bigvee_{j \in \mathrm{pa}(i)} c_{ij} X_j \vee Z_i, \qquad c_{ij}, Z_i \geq 0 \tag{3}$$

where $\vee = \max$, $\mathrm{pa}(i)$ is the set of parents of i in $\mathcal{G}$, and the Z_i are independent, atom-free, continuous random variables.

Example

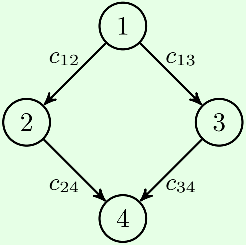
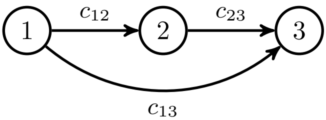


Figure: The diamond

$$X_1 = Z_1$$
$$X_2 = c_{12} X_1 \vee Z_2 = \max(c_{12} X_1, Z_2)$$
$$X_3 = c_{13} X_1 \vee Z_3 = \max(c_{13} X_1, Z_3)$$
$$X_4 = c_{24} X_2 \vee c_{34} X_3 \vee Z_4 = \max(c_{24} X_2, c_{34} X_3, Z_4)$$

Challenges of the Max-Linear setting

The conditional independence structure of a MLBN depends on the choice of edge weights:

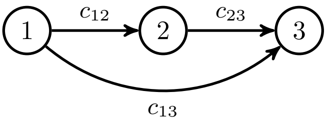


$$\text{The CI statements which hold are } \begin{cases} \{1 \perp\!\!\!\perp 3|2\} & \text{if } c_{13} \leq c_{12}c_{23} \\ \emptyset & \text{if } c_{13} > c_{12}c_{23}. \end{cases}$$

In particular: MLBNs are **not** faithful to d-separation.

Challenges of the Max-Linear setting

The conditional independence structure of a MLBN depends on the choice of edge weights:



The CI statements which hold are $\begin{cases} \{1 \perp\!\!\!\perp 3 | 2\} & \text{if } c_{13} \leq c_{12}c_{23} \\ \emptyset & \text{if } c_{13} > c_{12}c_{23}. \end{cases}$

In particular: MLBNs are **not** faithful to d-separation.
This motivated the C^**-separation* criterion of Améndola et. al [2].

C^*-separation

Let $(\mathcal{G}, C)$ be a weighted DAG with vertex set V and edge set E.
For $i, j \in V$, let $P(i, j)$ denote the set of all directed paths from i to j.

C^*-separation

Let $(\mathcal{G}, C)$ be a weighted DAG with vertex set V and edge set E.
For $i, j \in V$, let $P(i, j)$ denote the set of all directed paths from i to j.

- A path $\pi \in P(i, j)$ is *critical* in $(\mathcal{G}, C)$ if its weight $\omega_C(\pi) = \Pi_{e \in \pi} c_e$ is maximal over all paths in $P(i, j)$.

C^*-separation

Let $(\mathcal{G}, C)$ be a weighted DAG with vertex set V and edge set E.
For $i, j \in V$, let $P(i,j)$ denote the set of all directed paths from i to j.

- A path $\pi \in P(i,j)$ is *critical* in $(\mathcal{G}, C)$ if its weight $\omega_C(\pi) = \Pi_{e\in\pi} c_e$ is maximal over all paths in $P(i,j)$.
- For $K \subset V$, the *critical DAG* $\mathcal{G}^*_K(C)$ is the graph with vertex set V and edges determined by the condition

$$i \to j \in \mathcal{G}^*_K(C) \iff |P(i,j)| \geq 1 \text{ and no critical directed path from } i \text{ to } j \text{ intersects } K.$$

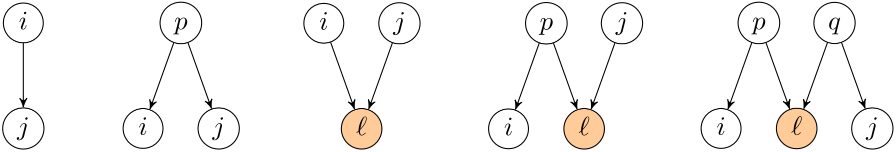
C^*-separation

Let $(\mathcal{G}, C)$ be a weighted DAG with vertex set V and edge set E.
For $i, j \in V$, let $P(i, j)$ denote the set of all directed paths from i to j.

- A path $\pi \in P(i, j)$ is *critical* in $(\mathcal{G}, C)$ if its weight $\omega_C(\pi) = \Pi_{e \in \pi} c_e$ is maximal over all paths in $P(i, j)$.
- For $K \subset V$, the *critical DAG* $\mathcal{G}^*_K(C)$ is the graph with vertex set V and edges determined by the condition

$$i \to j \in \mathcal{G}^*_K(C) \iff |P(i,j)| \geq 1 \text{ and no critical directed path from } i \text{ to } j \text{ intersects } K.$$

- Two nodes $i, j \in V$ are *C^*−connected* given $K \subset V \setminus ij$ if there exists an $i - j$ path in $\mathcal{G}^*_K(C)$ of one of the five types below.



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

C^*-separation:part II

Let $i, j, K \subset V$. If no C^*−connecting path exists in $\mathcal{G}^*_K(C)$, we say that i and j are *C^*−separated* given K, and write $[i \perp_{C^*} j \mid K]$.

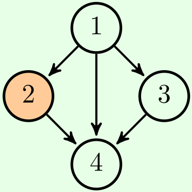
C^*-separation:part II

Let $i, j, K \subset V$. If no C^*-connecting path exists in $\mathcal{G}^*_K(C)$, we say that i and j are *C^*-separated* given K, and write $[i \perp_{C^*} j \mid K]$.

Example

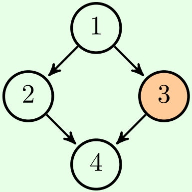
Let $(\mathcal{G}, C)$ be the the diamond  from Example 1 with C chosen such that $1 \to 3 \to 4$ is the unique critical $1 - 4$ path.

$\mathcal{G}^*_{\{2\}}(C)$



"1 and 4 are C^*-connected given 2"

$\mathcal{G}^*_{\{3\}}(C)$



$[1 \perp_{C^*} 4 \mid 3]$

C^*-separation: part III

Theorem ([2], Theorem 6.2)

MLBNs are faithful to C^*-separation. If X is distributed according to a MLBN on $\mathcal{G}$, then

$$[i \perp_{C^*} j \mid K] \text{ holds in } (\mathcal{G}, C) \iff [X_i \perp\!\!\!\perp X_j \mid X_K] \text{ holds in } X$$

In other words,

$$\text{global}_{C^*}(\mathcal{G}, C) = \left\{[i \perp\!\!\!\perp j | K] \text{ s.t } [i \perp_{C^*} j \mid K] \text{ holds in } (\mathcal{G}, C)\right\}$$

encodes the entire CI information of X.

C^*-separation: part III

Theorem ([2], Theorem 6.2)

MLBNs are faithful to C^*-separation. If X is distributed according to a MLBN on $\mathcal{G}$, then

$$[i \perp_{C^*} j \mid K] \text{ holds in } (\mathcal{G}, C) \iff [X_i \perp\!\!\!\perp X_j \mid X_K] \text{ holds in } X$$

In other words,

$$\text{global}_{C^*}(\mathcal{G}, C) = \left\{[i \perp\!\!\!\perp j | K] \text{ s.t } [i \perp_{C^*} j \mid K] \text{ holds in } (\mathcal{G}, C)\right\}$$

encodes the entire CI information of X.

Theorem ([2], Corollary 5.9)

For any choice of weights C:

$$\text{global}_d(\mathcal{G}) \subset \text{global}_*(\mathcal{G}, C)$$

Our contribution

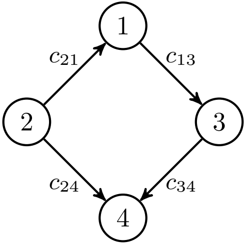
In our work “A PC algorithm for Max-Linear Bayesian Networks”, we...

- Investigate the output of the PC algorithm upon replacing $\text{global}_d(\mathcal{G})$ with $\text{global}_*(\mathcal{G}, C)$.
- Introduce a modified PC algorithm for Causal Discovery in MLBNs.
- Develop a new edge orientation rule which allows for additional identifiability.
- Implement the algorithm in `julia` and perform tests.

Skeleton retrieval and the weighted transitive reduction

The PC algorithm deletes additional edges!

Consider the *21-diamond* with edge weights chosen such that $c_{24} < c_{21}c_{13}c_{34}$ and corresponding structural equations:



$$
\begin{aligned}
X_1 &= c_{21}X_2 \vee Z_1 \\
X_2 &= Z_2 \\
X_3 &= c_{13}X_1 \vee Z_3 = c_{13}(c_{21}Z_2 \vee Z_1) \vee Z_3 \\
X_4 &= c_{24}X_2 \vee c_{34}X_3 \vee Z_4
\end{aligned}
$$

The set of conditional independence statements which hold in $X = (X_1, \ldots X_4)$ is

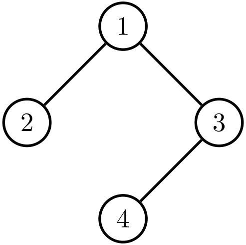
$$
\begin{aligned}
&1 \perp\!\!\!\perp 4|\{3\}\ ,\ 1 \perp\!\!\!\perp 4|\{2,3\} \\
&2 \perp\!\!\!\perp 3|\{1\}\ ,\ 2 \perp\!\!\!\perp 3|\{1,4\} \\
&2 \perp\!\!\!\perp 4|\{1\}\ ,\ 2 \perp\!\!\!\perp 4|\{3\}\ ,\ 2 \perp\!\!\!\perp 4|\{1,3\}.
\end{aligned}
$$

Thus...

The PC algorithm deletes additional edges

$1 \perp\!\!\!\perp 4|\{3\}$, $1 \perp\!\!\!\perp 4|\{2,3\}$
$2 \perp\!\!\!\perp 3|\{1\}$, $2 \perp\!\!\!\perp 3|\{1,4\}$
$2 \perp\!\!\!\perp 4|\{1\}$, $2 \perp\!\!\!\perp 4|\{3\}$, $2 \perp\!\!\!\perp 4|\{1,3\}$.

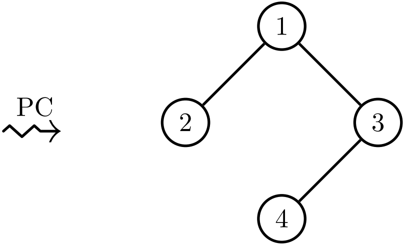
$\overset{\text{PC}}{\rightsquigarrow}$



Reason: $c_{24} < c_{21}c_{13}c_{34}$ implies that $2 \to 4$ is not *critical* in $(\mathcal{G}, C)$.

The PC algorithm deletes additional edges

$1 \perp\!\!\!\perp 4|\{3\}$, $1 \perp\!\!\!\perp 4|\{2,3\}$
$2 \perp\!\!\!\perp 3|\{1\}$, $2 \perp\!\!\!\perp 3|\{1,4\}$
$2 \perp\!\!\!\perp 4|\{1\}$, $2 \perp\!\!\!\perp 4|\{3\}$, $2 \perp\!\!\!\perp 4|\{1,3\}$.

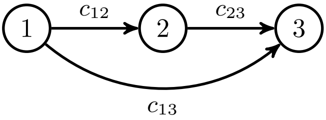


Reason: $c_{24} < c_{21}c_{13}c_{34}$ implies that $2 \to 4$ is not *critical* in $(\mathcal{G}, C)$.
Skeleton retrieval no longer retrieves the undirected skeleton of $\mathcal{G}$.
This is a feature, not a bug!

Weighted Transitive Reduction

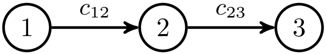
The *weighted transitive reduction* of $(\mathcal{G}, C)$ is the weighted DAG $(\mathcal{G}_C^{\mathrm{tr}}, C^{tr})$ on n nodes with weighted edges determined as follows:

$i \to j \in \mathcal{G}_C^{\mathrm{tr}}$ with weight c_{ij} $\quad :\Longleftrightarrow$ The edge $i \to j$ is the unique critical path from i to j in $\mathcal{G}$.



$\rightsquigarrow$

$c_{13} < c_{12}c_{23}$



Theorem (FN+ 2025)

$\mathcal{G}_C^{\mathrm{tr}}$ is the sparsest subgraph of $\mathcal{G}$ capable of encoding the same CI statements as $(\mathcal{G}, C)$.

Skeleton Retrieval in MLBNs

Theorem (FN+ 2025)

Applying the Skeleton Retrieval Step of the PC algorithm to the set

$$\text{global}_*(\mathcal{G}, C) = \left\{[i \perp\!\!\!\perp j|K] \ s.t \ [i \perp_{C^*} j \mid K] \ \textit{holds in} \ (\mathcal{G}, C)\right\}$$

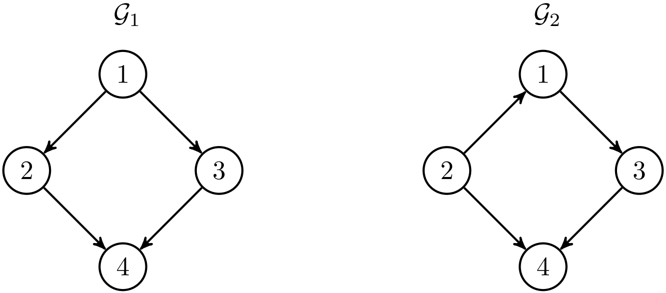
retrieves the undirected skeleton of $\mathcal{G}_C^{\text{tr}}$.

With a modified edge orientation step, the unshielded colliders of $\mathcal{G}_C^{\text{tr}}$ can also be determined in polynomial time.

The PCstar algorithm

Induced cycles* can be oriented!

$\mathcal{G}_1$ and $\mathcal{G}_2$ are Markov equivalent w.r.t $\perp_d$: $\text{global}_d(\mathcal{G}_1) = \text{global}_d(\mathcal{G}_2)$



However their C^*-Markov properties differ for any choice of weights.

More generally: *Induced cycles* of $\mathcal{G}$ containing a unique collider triple may be oriented under certain additional assumptions!

PCSTAR

Algorithm 1: PCSTAR

Input : A complete set of CI statements $\text{global}_*(\mathcal{G}, C)$ coming from a graphical model faithful to C^*-separation

Output: A CPDAG approximating the sparsest graph with the same Markov property as $(\mathcal{G}, C)$ for an appropriate choice of weights

Recover $\text{skel}(\mathcal{G}_C^{\text{tr}})$ by applying skeleton retrieval to $\text{global}_*(\mathcal{G}, C)$.

Detect and orient the unshielded colliders

Optional: Orient all identifiable induced cycles

Theorem (FN+ 2025)

The output of PCSTAR is a CPDAG with the same undirected skeleton, unshielded colliders, and orientable induced cycles as $\mathcal{G}_C^{\text{tr}}$. *Without the optional cycle orientation step, its complexity is* $\mathcal{O}(n^{d+2})$, *where* $n = |V|$ *and* $d = \max_{v \in V} \text{indeg}(v)$.

Implementation and experiments

We implemented PCSTAR in julia and ran it on data produced by randomly generated DAGs:

$\lvert V \rvert$	d	# edges of $\mathcal{G}$	# edges of $\mathcal{G}_C^{\text{tr}}$	# recovered edges (w/o cycle orientation)	# recovered edges (with cycle orientation)
10	2	9.17	8.66	3.88	4.43
	3	13.13	10.54	6.38	7.54
	4	16.13	11.57	6.99	8.33
	5	19.63	12.15	6.94	8.60
15	2	15.63	14.89	8.53	9.58
	3	17.46	15.95	10.02	11.31
	4	22.74	18.21	12.83	14.59
	5	28.30	19.61	14.17	15.92
20	2	20.36	19.74	11.51	12.47
	3	22.57	21.31	13.74	15.32
	4	28.13	24.38	17.85	20.14

`https://github.com/fpnowell/starskeleton`

Caveats and future directions

- No specialized (non-parametric) CI testing for MLBNs
- Comparison with score-based approaches to causal discovery [1]
- Further investigation of the combinatorial structure of $\text{global}_*(\mathcal{G}, C)$ [3]
- Interventions and "do"-calculus [4] for MLBNs

Thank you! Questions?

A PC Algorithm for Max-Linear Bayesian Networks

https://arxiv.org/abs/2508.13967

[1] Mark Adams, Kamillo Ferry, and Ruriko Yoshida. *Inference for max-linear Bayesian networks with noise.* 2025. arXiv: 2505.00229 [stat.ML]. URL: https://arxiv.org/abs/2505.00229 (cit. on p. 42).

[2] Carlos Améndola et al. "Conditional independence in max-linear Bayesian networks". In: *The Annals of Applied Probability* 32.1 (Feb. 2022). Publisher: Institute of Mathematical Statistics, pp. 1–45. ISSN: 1050-5164, 2168-8737. DOI: 10.1214/21-AAP1670. URL: https://projecteuclid.org/journals/annals-of-applied-probability/volume-32/issue-1/Conditional-independence-in-max-linear-Bayesian-networks/10.1214/21-AAP1670.full (visited on 10/08/2024) (cit. on pp. 21, 22, 29, 30).

[3] Tobias Boege et al. *Polyhedral Aspects of Maxoids.* arXiv:2504.21068 [math]. Apr. 2025. DOI: 10.48550/arXiv.2504.21068. URL: http://arxiv.org/abs/2504.21068 (visited on 07/14/2025) (cit. on p. 42).

[4] Judea Pearl. *Causality.* 2nd ed. Cambridge University Press, 2009 (cit. on p. 42).

[5] Peter Spirtes and Clark Glymour. "An Algorithm for Fast Recovery of Sparse Causal Graphs". en. In: *Social Science Computer Review* 9.1 (Apr. 1991). Publisher: SAGE Publications Inc, pp. 62–72. ISSN: 0894-4393.

DOI: 10.1177/089443939100900106. URL: https://doi.org/10.1177/089443939100900106 (visited on 10/07/2024) (cit. on pp. 13, 14).

[6] Thomas Verma and Judea Pearl. "Equivalence and synthesis of causal models". In: *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*. 1990, pp. 255–270 (cit. on pp. 7, 8).