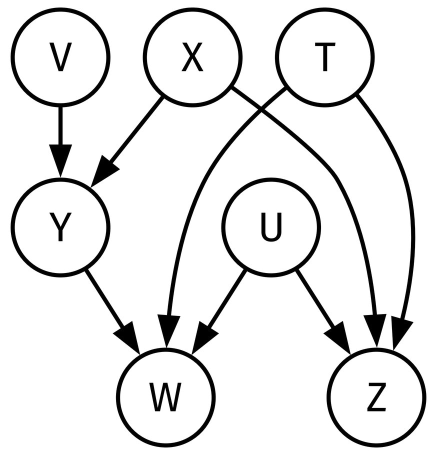
An Efficient Search-and-Score Algorithm for Ancestral Graphs Using Multivariate Information Scores

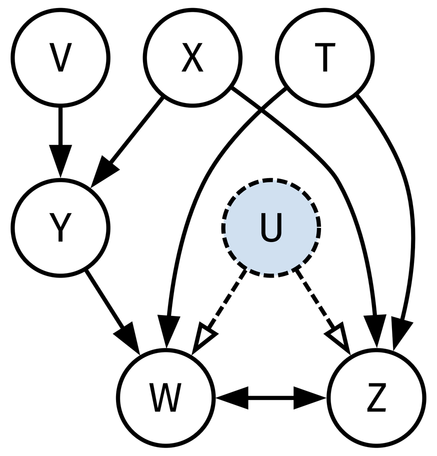
Nikita Lagrange, Hervé Isambert, CNRS, Institut Curie, Paris

Motivation



Likelihood of **Bayesian Networks (BN)**

$$\mathcal{L}_{\mathcal{D}|\mathcal{G}_{\mathrm{BN}}} = \frac{1}{Z_{\mathcal{D},\mathcal{G}_{\mathrm{BN}}}} \exp\left(-N \sum_{X_i \in \boldsymbol{V}}^{\text{vertices}} H(X_i \mid \mathbf{Pa}_{X_i})\right)$$



Likelihood of **Ancestral Graphs (AG)** ?

$$\mathcal{L}_{\mathcal{D}|\mathcal{G}_{\mathrm{AG}}} = \frac{1}{Z_{\mathcal{D},\mathcal{G}_{\mathrm{AG}}}} \exp\big(-N\, H(p,q)\big)$$

$$H(p,q) = \sum_x -p(x)\log q(x) = -\overset{?}{\sum_{\boldsymbol{C}\subseteq\boldsymbol{V}}} (-1)^{|\boldsymbol{C}|} I(\boldsymbol{C})$$

Theory

Code: github.com/miicTeam/miicsearchscore

Definition: *ac*-connected subset

A subset $\boldsymbol{C}$ is said to be *ac*-connected if $\forall X, Y \in \boldsymbol{C}$, either:

- X and Y are connected through any type of edge, or
- there exists an ***ac*-connecting path** between X and Y given $\boldsymbol{C}$.

An ***ac*-connecting path** between X and Y given $\boldsymbol{C}$ is a collider path

$$X{*}\!\rightarrow Z_1 \leftrightarrow \cdots \leftrightarrow Z_k \leftarrow\!{*}Y,$$

such that each $Z_i \in \mathbf{An}_{\mathcal{G}}(\{X,Y\}\cup\boldsymbol{C})$, that is, with $Z_i \in \boldsymbol{C}$ or through an ancestor path, *i.e.*, $Z_i \rightarrow \cdots \rightarrow T$ with $T \in \{X,Y\}\cup\boldsymbol{C}$.

Theorem: likelihood of ancestral graphs[1]

The likelihood $\mathcal{L}_{\mathcal{D}|\mathcal{G}}$ of an ancestral graph $\mathcal{G}$ is decomposable in terms of multivariate cross-information $I(\boldsymbol{C})$, summed over all *ac*-connected subsets of variables $\boldsymbol{C}$:

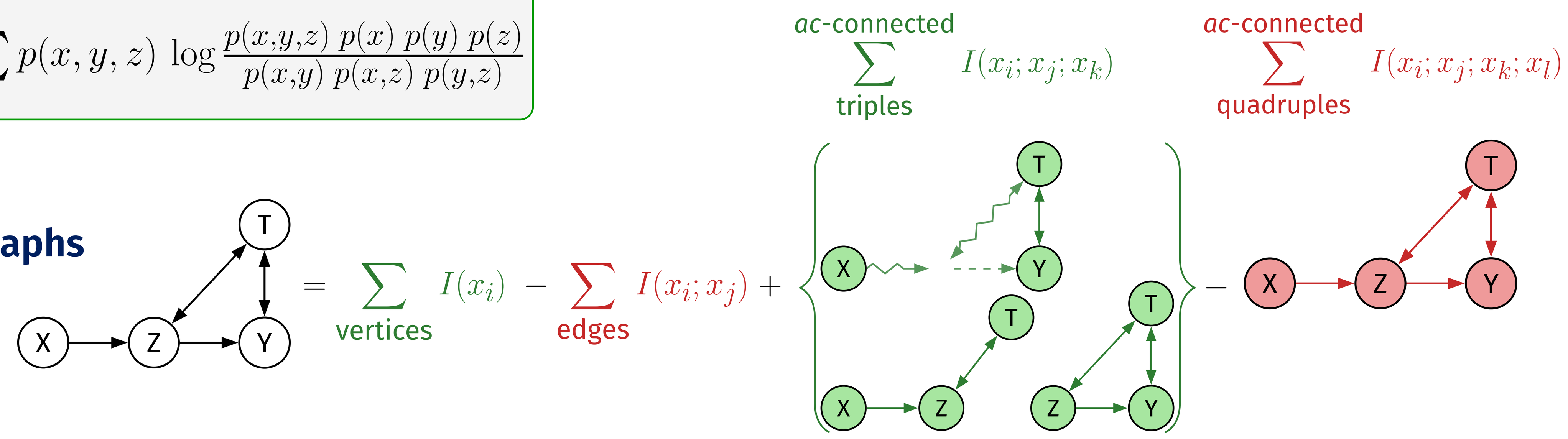
$$\mathcal{L}_{\mathcal{D}|\mathcal{G}} = \frac{1}{Z_{\mathcal{D},\mathcal{G}}} \exp\left(N \sum_{\boldsymbol{C}\subseteq\boldsymbol{V}}^{ac\text{-connected}} (-1)^{|\boldsymbol{C}|} I(\boldsymbol{C})\right)$$

where N is the number of iid samples in dataset $\mathcal{D}$, and $Z_{\mathcal{D},\mathcal{G}}$ is a data- and model-dependent normalization constant.

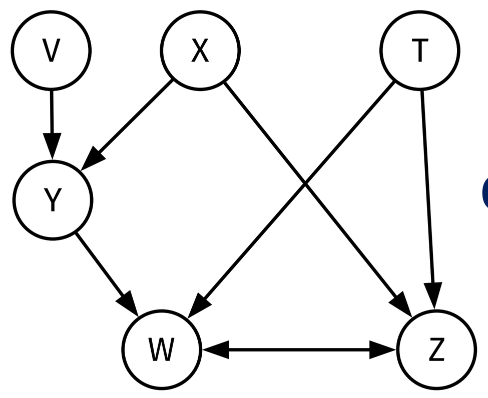
Proposition: estimation of cross-information for *ac*-connected subset C

$$I(\boldsymbol{C}) = \sum p(x,y,z)\,\log \frac{q(x,y,z)\;q(x)\;q(y)\;q(z)}{q(x,y)\;q(x,z)\;q(y,z)} \equiv \sum p(x,y,z)\,\log \frac{p(x,y,z)\;p(x)\;p(y)\;p(z)}{p(x,y)\;p(x,z)\;p(y,z)}$$

Cross-entropy decomposition of **ancestral graphs**



MIIC_search&score



Output from MIIC

MIIC: Multivariate Information-based Inductive Causation[2,3,4]

MIIC: information-theoretic causal discovery (handles latent variables)

Conditional mutual information → conditional independence

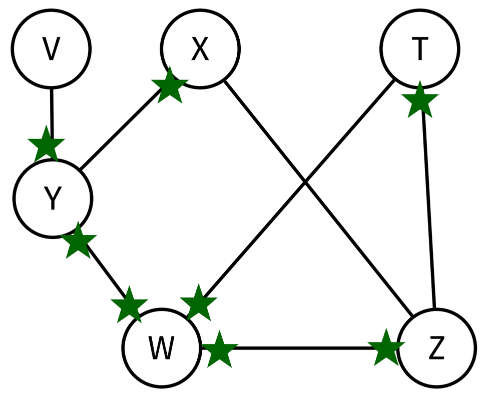
$$I'(X;Y \mid Z) = I(X;Y \mid Z) - \frac{k_{X;Y|Z}}{N}$$

$k_{X;Y|Z}$: fNML (factorized Normalized Maximum Likelihood) complexity term

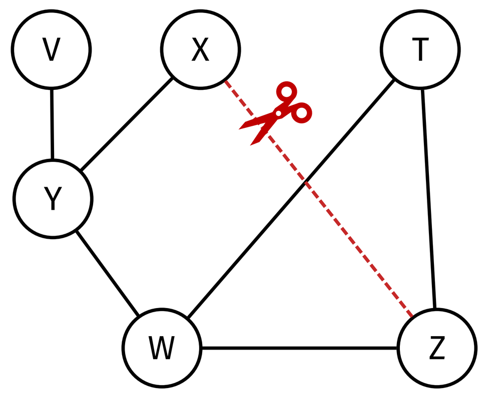
Penalizes model complexity • Prevents overfitting • Favors simpler, robust graphs • Removes spurious edges

Step 1

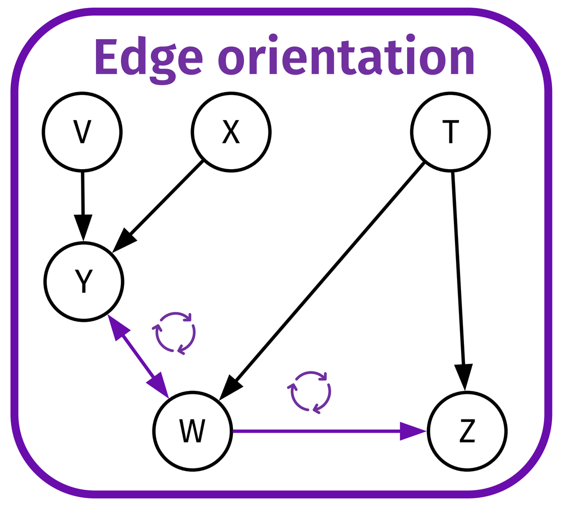
Edge orientation priming



Edge removal



Step 2



Step 1 – Local node score minimization

$$\min_{\mathbf{Pa}'_{X_i}} H(X_i \mid \mathbf{Pa}'_{X_i}) + \text{complexity}$$

$$\mathbf{Pa}'_{X_i} \subseteq \mathbf{Pa}_{X_i} \cup \mathbf{Sp}_{X_i} \cup \mathbf{Ne}_{X_i}$$

With:

- $\mathbf{Pa}_{X_i}$: parents of X_i, e.g. $Y \rightarrow X_i$
- $\mathbf{Sp}_{X_i}$: spouses of X_i, e.g. $Y \leftrightarrow X_i$
- $\mathbf{Ne}_{X_i}$: neighbors of X_i, e.g. $Y - X_i$

Step 2 – Local edge score minimization

$$\min_{\star \in \{\rightarrow,\leftarrow,\leftrightarrow\}} X \star Y = -I(X;Y \mid \mathbf{Pa}'_{\star}) + \text{cplx}$$

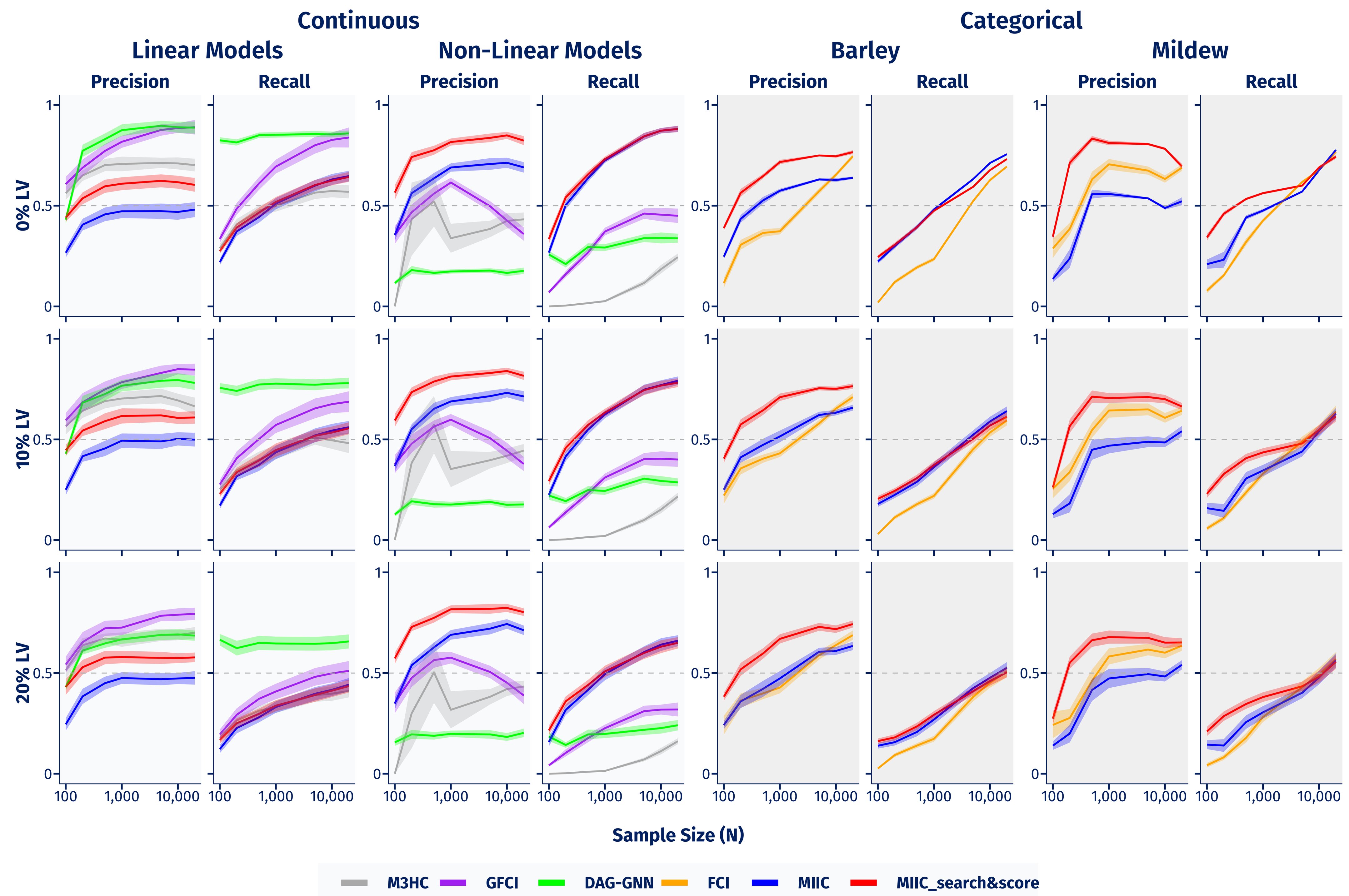
With:

$$\mathbf{Pa}'_{X\rightarrow Y} = \mathbf{Pa}_Y \cup \mathbf{Sp}_Y \backslash X$$

$$\mathbf{Pa}'_{X\leftarrow Y} = \mathbf{Pa}_X \cup \mathbf{Sp}_X \backslash Y$$

$$\mathbf{Pa}'_{X\leftrightarrow Y} = \mathbf{Pa}'_{X\rightarrow Y} \cup \mathbf{Pa}'_{X\leftarrow Y}$$

Benchmarks



Key Takeaways

Results

- **High precision & robust recall**
- **Strong on non-linear, non-Gaussian, and categorical data**
- Works with small samples
- Competitive scalability
- **No strong assumptions**

Limitations

- **Limited local scores**
- No undirected edges
- Outperformed on purely linear Gaussian data (GFCI & DAG-GNN best)

Perspectives

- **Higher-order collider paths**
- MCMC or oracle search
- Mixed and time-series data

References

[1] Lagrange N., Isambert H. (2025). *An Efficient Search-and-Score Algorithm for Ancestral Graphs using Multivariate Information Scores for Complex Non-linear and Categorical Data*. In ICML.

[2] Ribeiro-Dantas M., Li H., Cabeli V., et al. (2024). *Learning interpretable causal networks from large-scale data: application to 400,000 breast cancer records*. iScience.

[3] Cabeli V., Li H., Ribeiro-Dantas M., et al. (2021). *Reliable causal discovery using mutual information supremum principle*. In Why21 @ NeurIPS.

[4] Li H., Cabeli V., Sella N., Isambert H. (2019). *Constraint-based causal structure learning with consistent separating sets*. In NeurIPS.