An Efficient Search-and-Score Algorithm
for Ancestral Graphs Using Multivariate Information Scores for Complex Non-Linear and Categorical Data

Nikita Lagrange, Hervé Isambert

📍 **Poster Session 2 #45267**
📅 Tuesday, July 15 | 🕐 4:30–7:00 p.m. (PDT)

Motivation

o **Likelihood estimation** is central but hard with **unobserved latent variables**
o **Standard methods** assume **Gaussian data** and fail on **complex real-world cases**
o **Need efficient, non-parametric** approaches for **non-linear** and **categorical data**

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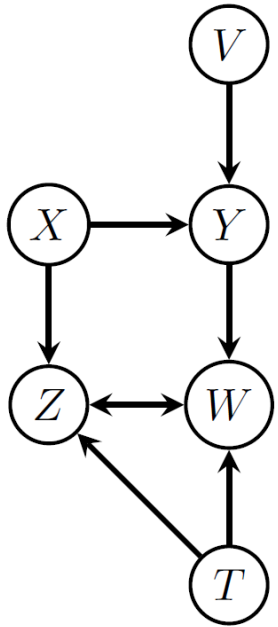
Contribution

o Likelihood decomposition of ancestral graphs **using multivariate information**
o **Efficient two-step search-and-score** algorithm with local **ac-connected subsets**
o **Outperforms** state-of-the-art on **non-linear** and **categorical** data
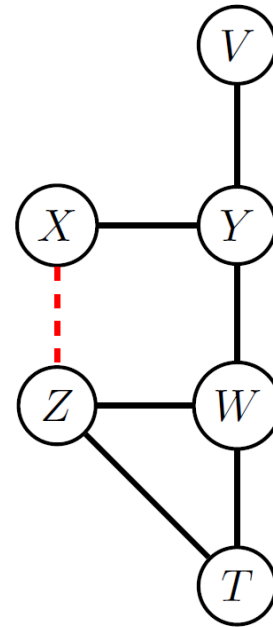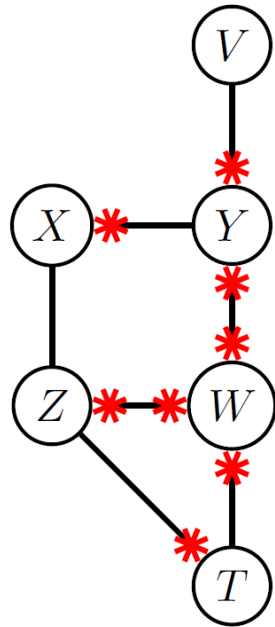
$$\mathcal{L}_{\mathcal{D}|\mathcal{G}} = \frac{1}{Z_{\mathcal{D},\mathcal{G}}} \exp\left(N \sum_{C \subseteq V}^{\text{ac-connected}} (-1)^{|C|} I(C)\right)$$

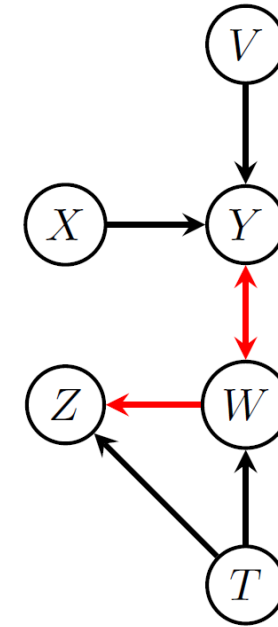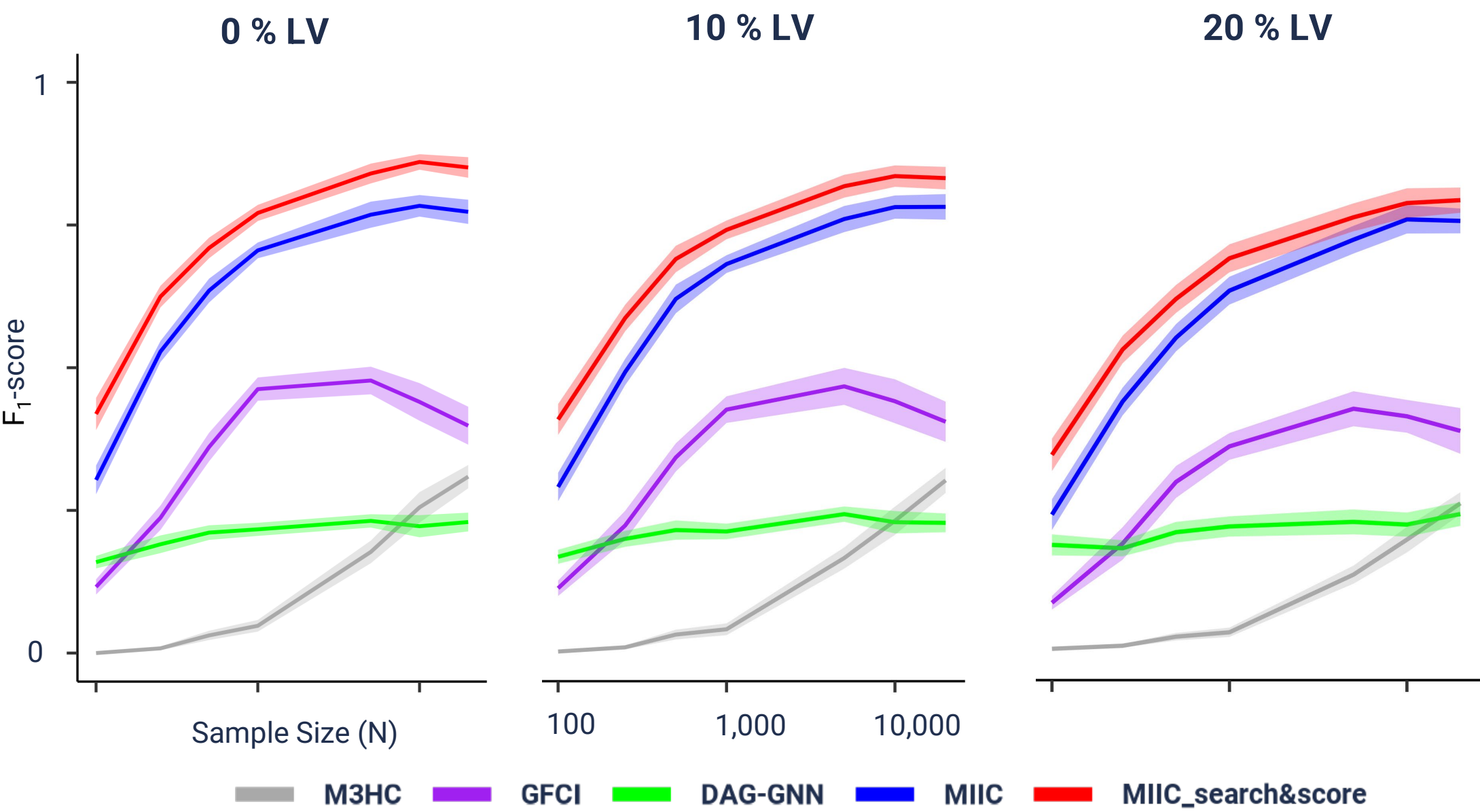# Method: MIIC_search&score Two-Step Algorithm

(i) Output from MIIC

(ii) **Step 1**: Edge orientation priming and Edge removal

(iii) **Step 2**: Edge orientation

Experimental Results: Non-Linear Continuous

0 % LV   10 % LV   20 % LV

F_1-score

Sample Size (N)

100   1,000   10,000

M3HC   GFCI   DAG-GNN   MIIC   MIIC_search&score

Conclusion

o **Introduced** a **decomposition** of ancestral graph likelihood based on **multivariate information** over *ac*-**connected subsets**

o **Reliably learns complex models** from **limited data**, including **non-linear, non-Gaussian**, and **categorical cases with hidden variables**

o **Simplified two-collider approximation** enables **scalability**, at the cost of **missing deeper dependencies**

nikita.lagrange@tutanota.com
nikitalagrange.github.io

github.com/miicTeam/miicsearchscore