Score-based vs constraint-based causal learning in the presence of confounders

Sofia Triantafillou and Ioannis Tsamardinos

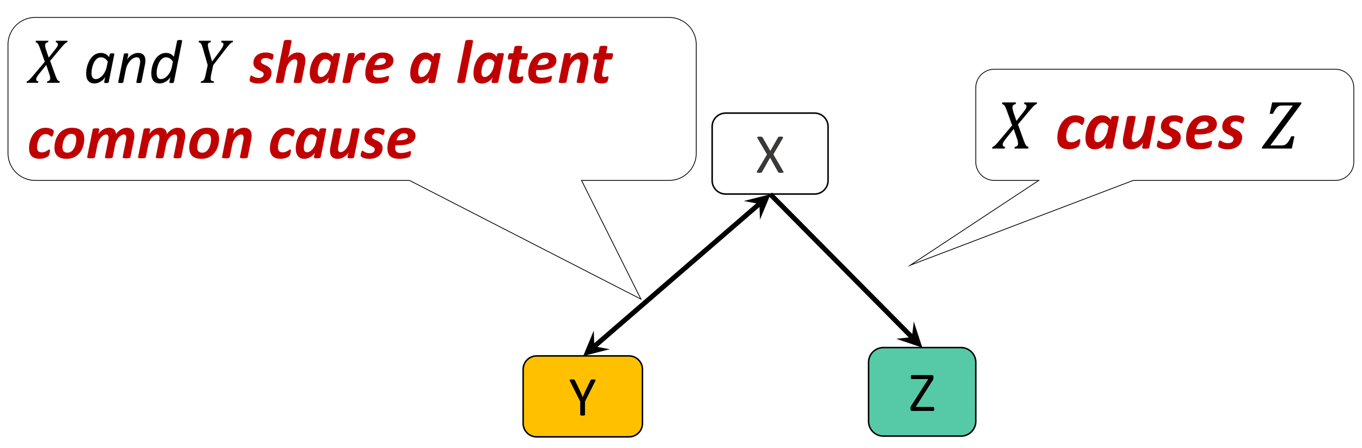
Department of Computer Science, University of Crete

Abstract

- We compare score-based and constraint-based learning in the presence of latent confounders for continuous data sets.
- We develop a greedy search strategy to identify the best fitting Maximal Ancestral Graph (MAG)[1] under the assumption of multivariate normality.
- We use Residual Iterative Conditional Fitting (RICF) [2] is used to obtain MLE parameters for a MAG G and score the graph.
- We use decomposition results [3, 4] designed for Semi-Markov Causal Models to implement an efficient greedy search.
- We compare to FCI [5] and CFCI [6]. Simple greedy search performs worse than CFCI and often better than FCI.
- Results indicate potential benefits from combining approaches.

Maximal Ancestral Graphs

- Extensons of Bayesian Networks that allow modeling causally insufficient systems (i.e. the presence of latent variables).
- No directed cycles, no almost cycles: $X \to \cdots \to Y \leftrightarrow X$.



- Under multivariate normality, a MAG G defines a system of linear equations

$$V_i = \sum_{j \in Pa_G(i)} \beta_{ij} V_j + \epsilon_i, \qquad i \in \{1, \dots V\},$$

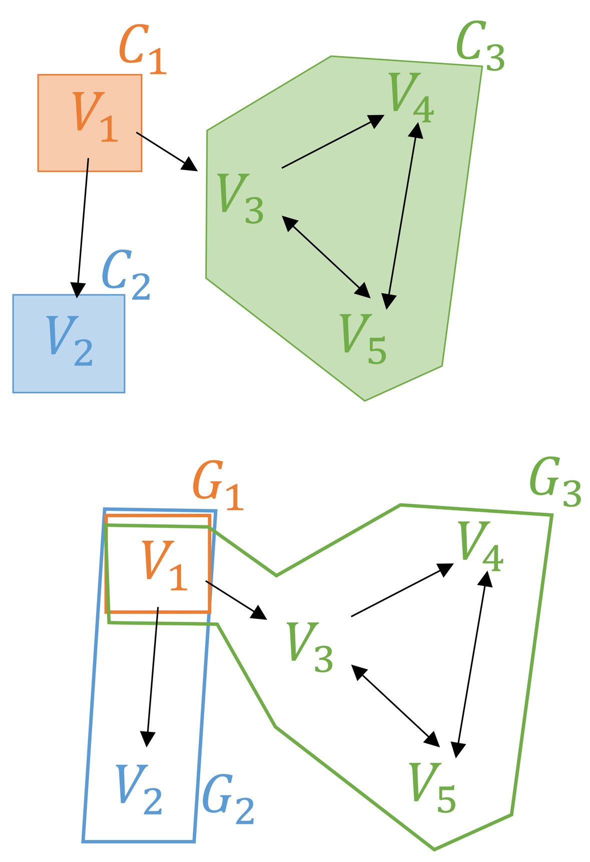
$$\beta_{ij} = 0 \; if \; V_j \to V_i \notin G$$

$$\omega_{ij} = cov(\epsilon_i, \epsilon_j), \omega_{ij} = 0 \; if \; V_j \leftrightarrow V_i \notin G$$

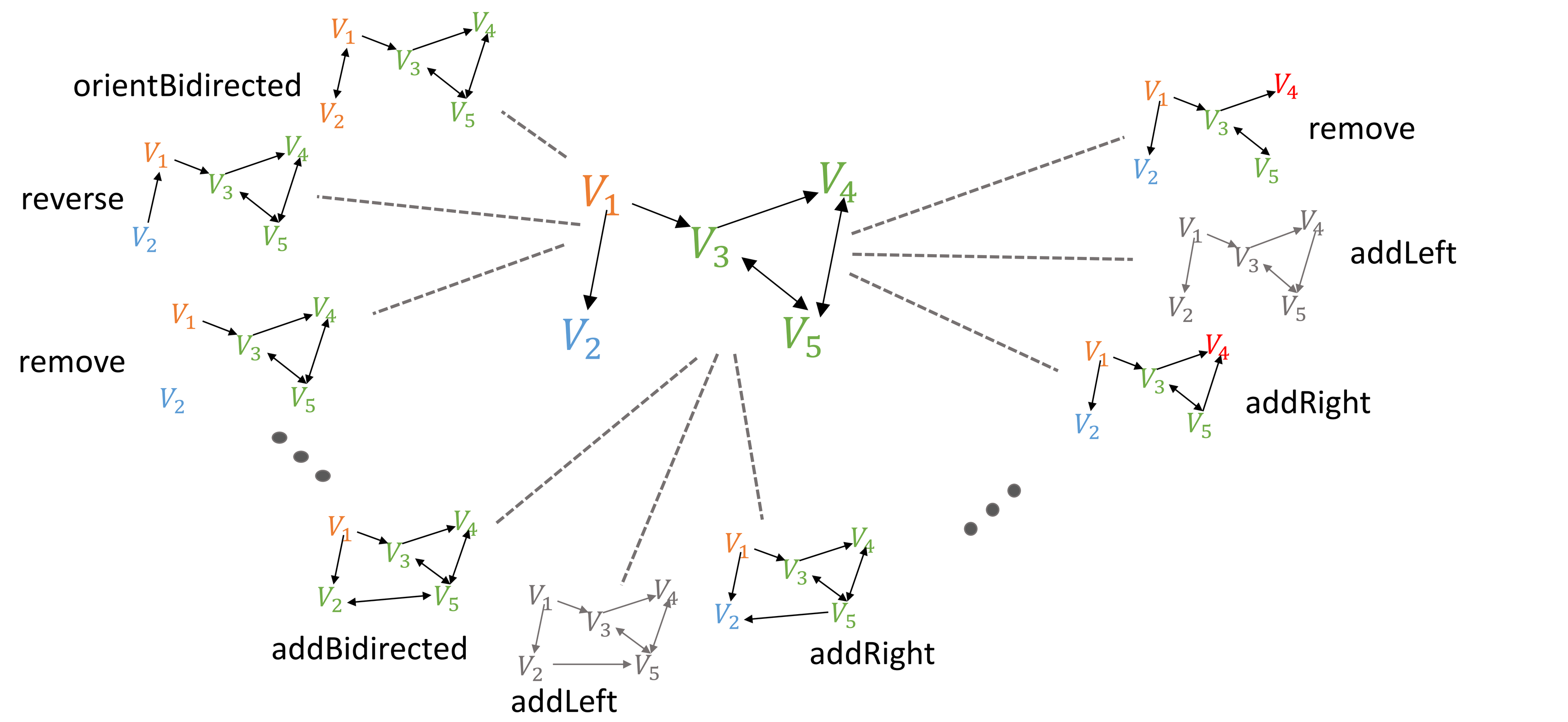
- Use RICF to identify MLE $\hat{\beta}, \hat{\omega}$ and likelihood $L_G(\hat{\beta}, \hat{\omega}|D)$.
- Use BIC to score $(G, \hat{\beta}, \hat{\omega})$.

Score Decomposition

- Edges that are connected with a bi-directed path form a confounded – component (c-component).
- A MAG can be split into maximal c-components $\{C_1, \dots, C_m\}$.
- Likelihood can be decomposed c-components [3, 4].
- $L_G = \sum_k L_G(G_k)$
- G_K: The restriction of G on $C_k \cup Pa_G(C_k)$.

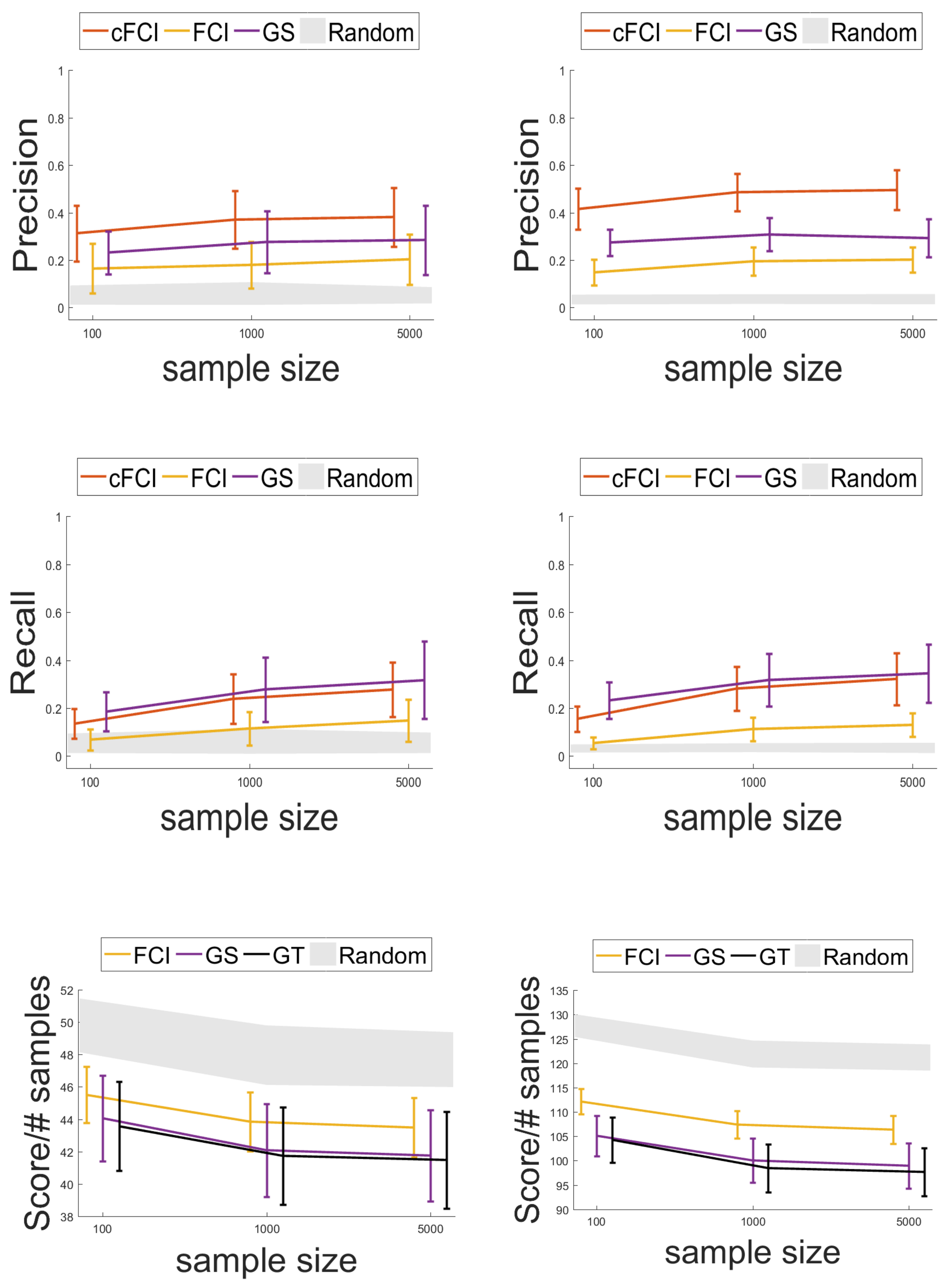


Greedy search



- Start from the empty graph.
- In each step, try all possible changes that do not create (almost) cycles.
- Recompute the score only for connected components that have changed.
  - Split connected components when a $\leftrightarrow$ edge is removed / becomes directed.
  - Merge connected components when a $\leftrightarrow$ edge is added.
- Incrementally recompute ancestral relationships when a $\to, \leftarrow$ edge is added.

Experiments



- 100 graphs.
- 10*/20/50 variables.
- 3*/5 maximum parents per variable .

- Precision: $\frac{TP \; edges}{\# \; edges \; in \; \hat{G}}$.
- Recall: $\frac{TP \; edges}{\# \; edges \; in \; G}$.
- Score: Penalized negative log likelihood (lowest for best model).

- Best precision: cFCI.
- Best recall: GS.
- GS score is close to ground truth score.
- GS is much slower than both FCI, CFCI*.

*not shown.

References

1. TS Richardson and P Spirtes. Ancestral graph Markov models. The Annals of Statistics.
2. J Tian and J Pearl. On the identification of causal effects. Technical Report.
3. M Drton, M Eichler, and TS Richardson. Computing maximum likelihood estimates in recursive linear models with correlated errors. JMLR.
4. C Nowzohour, M Maathuis, and P Bühlmann. Structure learning with bow-free acyclic path diagrams, arXiv preprint.

contact: striant@csd.uoc.gr