CISPA HELMHOLTZ CENTER FOR INFORMATION SECURITY

Information-Theoretic Causal Discovery in Topological Order

Sascha Xu, Sarah Mameche, Jilles Vreeken
CISPA Helmholtz Center for Information Security

Framework

Assumptions

(1) Information Gain $g : \mathscr{E} \times \mathscr{G} \to \mathbb{R}$ of adding an edge $i \to j$ to graph G:

$$K(P_{X_j} | \mathrm{pa}_j{}^G) - K(P_{X_j} | \mathrm{pa}_j{}^G \cup i)$$

(2) Faithfulness: for a true edge and any graph G we have $g(i \to j, G) > 0$

(3) Path Identifiability: if $\mathrm{pa}_i{}^G \supseteq \mathrm{pa}_i{}^*$ and $i \in \mathrm{an}_j{}^*$ we have $g(i \to j, G) > g(i \to j, G')$

(4) Independence of Conditionals:

$$\mathrm{pa}_j{}^G \supseteq \mathrm{pa}_j{}^* : K(P_{X_j} | \mathrm{pa}_j{}^G) = K(P_{X_j} | \mathrm{pa}_j{}^*)$$

We propose an **oracle** Ω of the most predictive node i in a given graph G as

$$\operatorname*{argmax}_i \left(\min_{j \notin \mathrm{an}_i G} \left(g(i \to j; G) - g(j \to i; G) \right) \right)$$

Under A(3), Ω returns nodes in true topological order.

TOPIC Algorithm

Initialize empty graph G. Iterate d times:

1 **query oracle** for the next node i.

2 **add all outgoing edges** that improve the score,

$$G' = G \cup (i, j) \quad \text{if} \quad g(i \to j; G) > 0$$

3 **prune redundant edges** by score,

$$G' = G \setminus (h, i) \quad \text{if} \quad L(G') < L(G)$$

TOPIC has complexity of $O(d^3)$ for d nodes.

Theorem 1 *When Assumptions (2-4) are met, TOPIC recovers the true causal graph.*

Instantiation

In practice, we use the Minimum Description Length (MDL) to upper-bound Kolmogorov Complexity

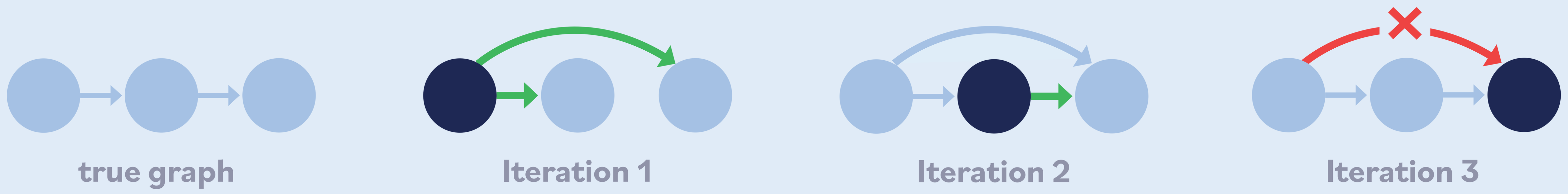
$$K(P_{X_i} | \mathrm{pa}_i{}^G) \approx L(D|M) + L(M)$$

Two-part MDL trades off model fit $L(M)$ and complexity $L(D|M)$. We use domain-specific instantiations of L

Information-Theoretic Causality Compression in the causal direction is greater than in the anti-causal direction

Given a domain-specific score, our algorithm TOPIC

1 finds a **source node** **2** **adds** out-edges **3** **prunes** in-edges



We give identifiability guarantees for

i. Continuous Additive ANM

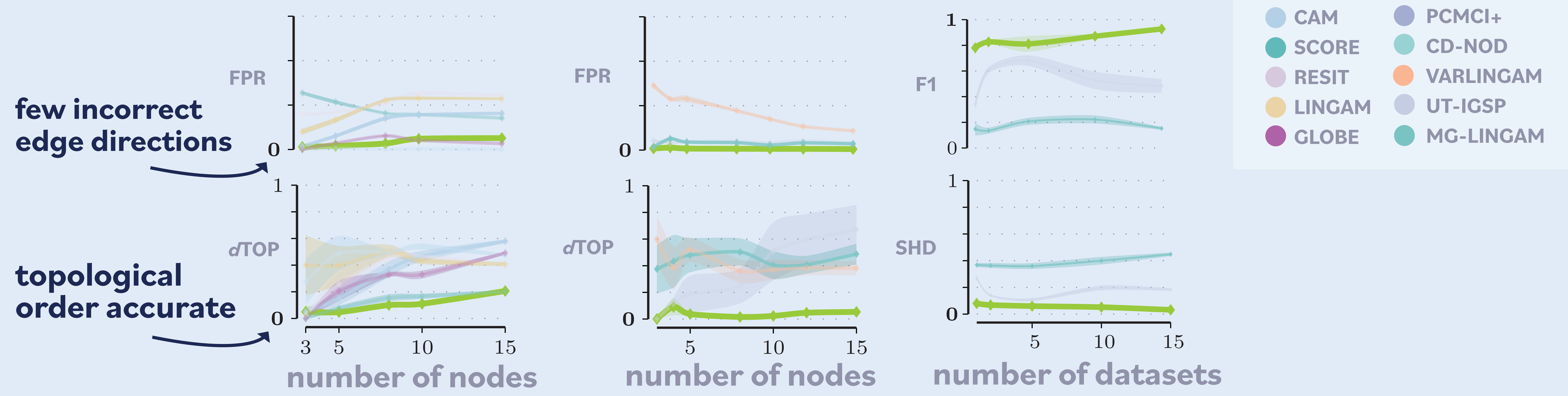
$$X_j = \textstyle\sum_{i \in \mathrm{pa}_j} f_{ij}(X_i) + N_j$$

ii. ANM with Mechanism Shifts

$$X_j = \begin{cases} f_j^c(\mathrm{pa}_j) + N_j & \text{if} \quad j \in I_c \\ f_j(\mathrm{pa}_j) + N_j & \text{otherwise} \end{cases}$$

We perform evaluations against the SOTA in

✓continuous data ✓time series ✓non-iid data



The TOPIC framework discovers the true topological order and causal graph accurately – across different domains

Guarantees

We show under which causal models the compression gain is greater in the causal direction than in the anti-causal direction.

i. **Continuous IID case:** additive causal model with additive noise (ANM),

$$X_j = \textstyle\sum_{i \in \mathrm{pa}_j} f_{ij}(X_i) + N_j$$

Consider at iteration k of TOPIC a resolved node i in G where $\mathrm{pa}_i{}^G = \mathrm{pa}_i{}^*$, then the following holds.

Theorem 2 *Let any causal path $i \to j$ from a resolved node i be a post-nonlinear noise model. If all are identifiable, then TOPIC iterates in topological order of the true graph.*

ii. **Continuous non-IID case:** causal ANM with Independent Causal Mechanism Shifts (IMS),

$$X_j = \begin{cases} f_j^c(\mathrm{pa}_j) + N_j & \text{if} \quad j \in I_c \\ f_j(\mathrm{pa}_j) + N_j & \text{otherwise} \end{cases}$$

over multiple datasets c, where I_c is a list of (unknown) nodes that undergo a causal mechanism shift in c. We assume shifts to be **independent** and **sparse.**

Theorem 3 *The guarantees of Thm 2 hold with high probability in the non-iid case as the number of contexts N_c tends to infinity.*

REGED Lung Cancer Data