Interventional Causal Discovery in a Mixture of DAGs

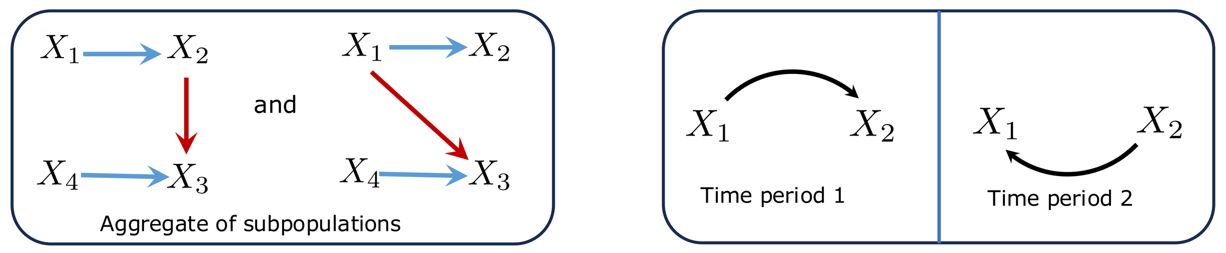
Burak Varıcı[1] Dmitriy A. Katz[2] Dennis Wei[2] Prasanna Sattigeri[2] Ali Tajer[3]

[1]Carnegie Mellon University [2]IBM Research [3]Rensselaer Polytechnic Institute

Motivation

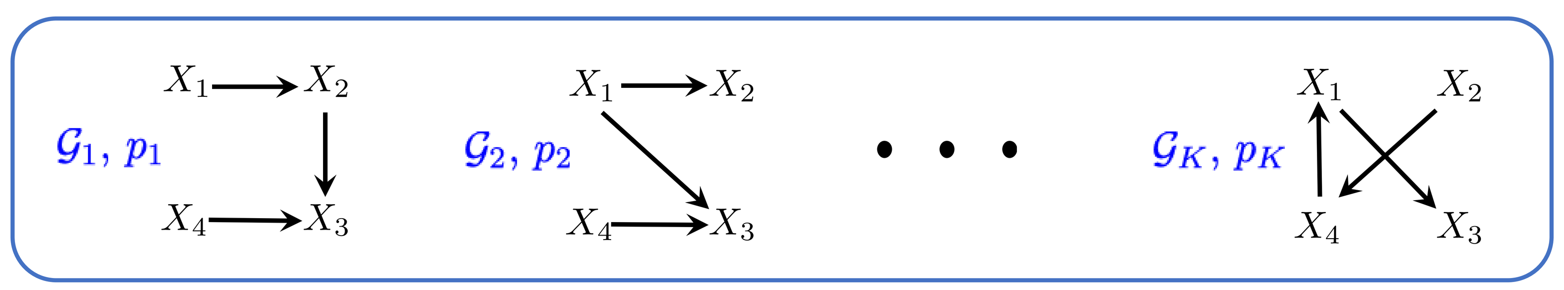
Most real causal systems are complex

- **Time-varying systems**: causal relationships can change over time
- **Cyclic relationships**: e.g., feedback loops
- **Modeling subpopulations**: e.g., subtypes of cancers do not share the same exact biological pathways



These complex models are better modeled by a mixture of DAGs!

Mixture Model



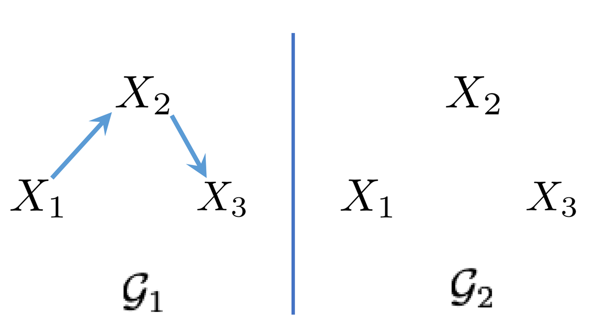
- A mixture of (unknown) K DAGs over n nodes: $\{\mathcal{G}_1, \dots, \mathcal{G}_K\}$
- Mixture distribution:

$$p_{\text{mix}} = \sum_{\ell \in [K]} w_\ell \cdot p_\ell \,, \qquad \text{where} \quad \sum_{\ell \in [K]} w_\ell = 1$$

- True edges: exist in at least one component DAG
- Define mixture parents: $\text{pa}_{\text{mix}}(i) = \bigcup_{\ell \in [K]} \text{pa}_\ell(i)$

Challenges of causal discovery in mixtures

- Single DAG: CI tests on observational data give the skeleton
- Mixture of DAGs: spurious unbreakable dependencies



X_1 and X_2 are inseparable $(1 - 2)$ is an emergent edge

Prior work on **observational** data

- FCI algorithm, graphical restrictions e.g. no cycles (Saeed et al. 2020)
- Longitudinal data, no solution for spurious (Strobl, 2023)
- Necessary-sufficient conditions for **emergent** edges (Varıcı et al. TMLR 2024) (*Separability Analysis for Causal Discovery in Mixture of DAGs*)

Observational data is not sufficient

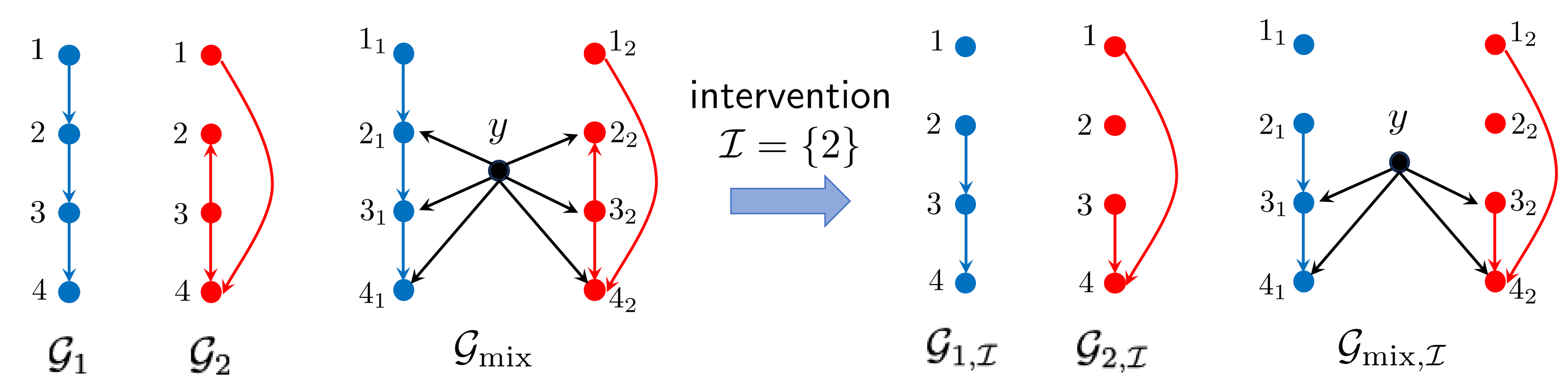
Interventional Causal Discovery

Intervention model: For a subset of nodes $I \subset [n]$, cut off parents

$$p_\ell(x_i \mid x_{\text{pa}(i)}) \to q_i(x_i) \,, \quad \forall i \in I, \ \forall \ell \in [K]$$

$$p_{\ell,\mathcal{I}}(x) \triangleq \prod_{i \in \mathcal{I}} q_i(x_i) \prod_{i \in \mathbf{V} \setminus I} p_\ell(x_i \mid x_{\text{pa}_\ell(i)}) \,, \quad \forall \ell \in [K]$$

interventional mixture dist. : $p_{\text{mix},\mathcal{I}} = \sum_{\ell \in K} w_\ell \cdot p_{\ell,\mathcal{I}}$



- True edges: $\{(1 \to 2), (2 \to 3), (3 \to 2), (3 \to 4), (1 \to 4)\}$
- Mixture parents, e.g., $\text{pa}_{\text{mix}}(2) = \{1, 3\}$ and $\text{pa}_{\text{mix}}(3) = \{2\}$
- Spurious dependencies: $\{(1, 3), (2, 4)\}$

goal: identify true edges (or mixture parents) via interventions

1. Necessary and sufficient sizes of interventions
2. A learning algorithm with near-optimal interventions

Assume interventional mixture faithfulness (standard extension)

Results

Theorem (intervention size): To find $\text{pa}_{\text{mix}}(i)$ of via CI tests

1. interventions with size $|I| \le |\text{pa}_{\text{mix}}(i)| + 1$ are **sufficient**
2. at the worst-case, $|I| = |\text{pa}_{\text{mix}}(i)| + 1$ is **necessary**

Why? To determine whether $j \in \text{pa}_{\text{mix}}(i)$, $I = \text{pa}_{\text{mix}}(i) \cup \{j\}$ suffices

Theorem (int. size - trees): For a mixture of trees

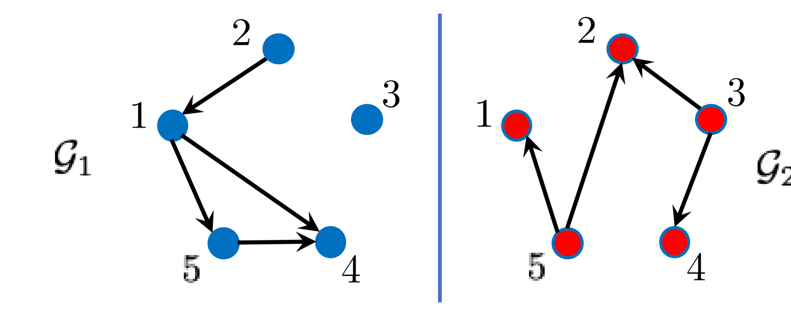
1. interventions with size $|I| \le K + 1$ are **sufficient**
2. at the worst-case, $|I| = K + 1$ is **necessary**

Why? There are at most K paths from j to i, one for each DAG
At the worst case (all disjoint paths), we need to block them all

Learning Algorithm

Step 1. **mixture ancestors/descendants** via atomic interventions

$$\text{des}_{\text{mix}}(i) = \{j : X_j \not\perp\!\!\!\perp X_i\} \quad \text{in} \ p_{\text{mix},\{i\}}\}$$



$\text{anc}_{\text{mix}}(1) = \{2, 5\}$
$\text{anc}_{\text{mix}}(2) = \{3, 5\}$
$\text{anc}_{\text{mix}}(3) = \emptyset$
$\text{anc}_{\text{mix}}(4) = \{1, 2, 3, 5\}$
$\text{anc}_{\text{mix}}(5) = \{1, 2\}$

Repeat steps 2, 3, 4 for all nodes $i \in \mathbf{V}$

Step 2. **Break cycles among ancestors**

- $\mathcal{C}(i)$: all cycles among $\text{anc}_{\text{mix}}(i)$
- $\mathcal{B}(i)$: minimal set that intersects with every cycle in $\mathcal{C}(i)$

e.g. $\mathcal{C}(4) = \{(2,5,2), (2,1,5,2), (1,5,1)\} \ \to \ \mathcal{B}(4) = \{5\}$

- **Cyclic complexity**: $\tau_i \triangleq |\mathcal{B}(i)|$
- $I = \mathcal{B}(i) \cup \{j\}$ for all $j \in \text{anc}_{\text{mix}}(i)$ to refine descendants

Step 3. **Topological layering**: refined descendant sets do not conflict!

- Bottom-up layering: $S_1(4) = \{1\}$, $S_2(4) = \{2\}$, $S_3(4) = \{3, 5\}$

Step 4. **Identify mixture parents**: process each layer sequentially

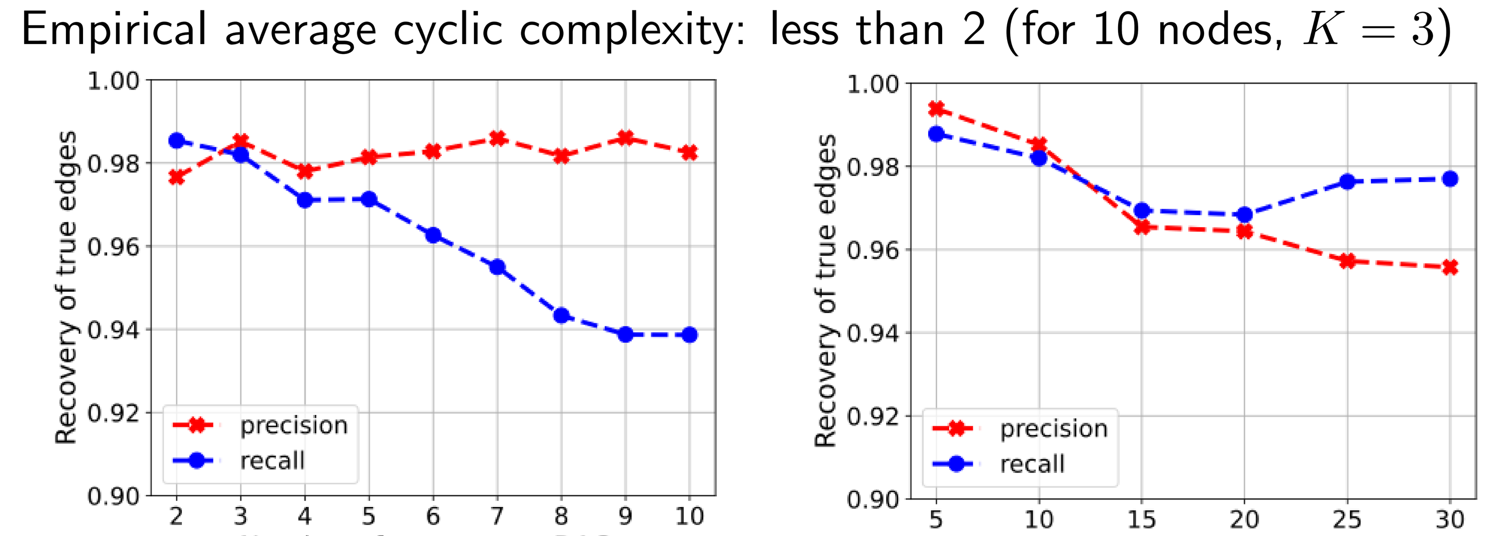
- For every possible parent j, intervene on $\mathcal{B}(i) \cup \hat{\text{pa}}(i) \cup \{j\}$
- e.g. $I = \{1, 5\}$ determines whether $1 \in \text{pa}_{\text{mix}}(4)$

Theorem: Learns all true edges using $\mathcal{O}(n^2)$ interventions, with interv. size at most $|\text{pa}_{\text{mix}}(i)| + \tau_i + 1$ for each node i.

Optimality gap = cyclic complexity of node i

Experiments

- Mixture of Gaussians, vary number of DAGs and graph size
- Strong performance for all settings.
- Empirical average cyclic complexity: less than 2 (for 10 nodes, $K = 3$)



contact: bvarici@andrew.cmu.edu