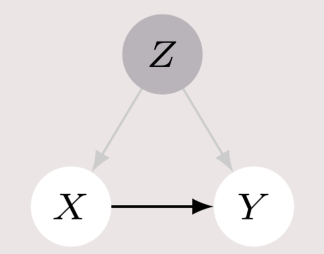
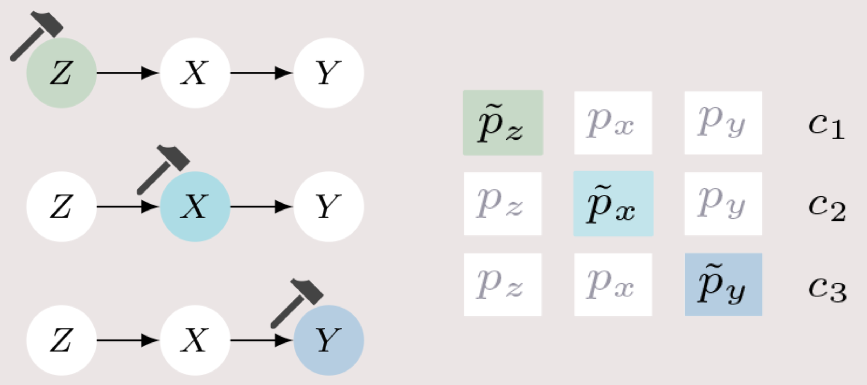
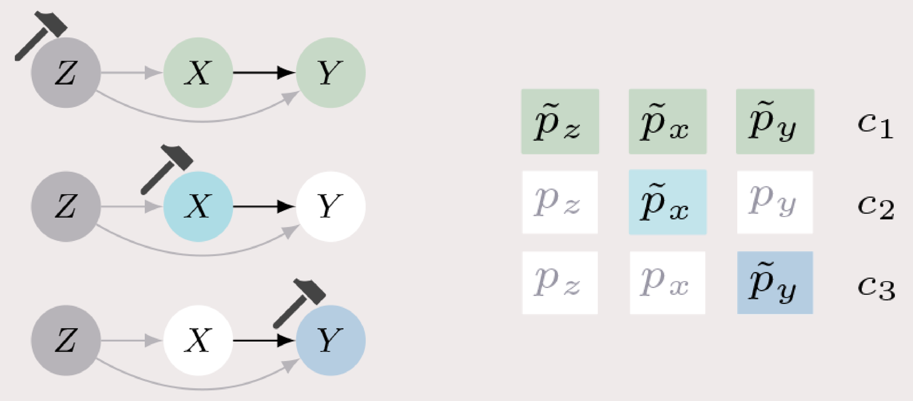
Identifying Confounding from Causal Mechanism Shifts

Sarah Mameche, Jilles Vreeken, David Kaltenpoth
CISPA Helmholtz Center for Information Security

MOTIVATION







1 **Confounders** latent common causes
Problem source of spurious correlations

2 **Different Contexts** (e.g. hospitals, experiments)
Problem distribution shifts (e.g. by intervention)

Goal Discovering confounders and causal directions
Insight Changes of confounders are reflected in the observed distribution of confounded nodes

PROBLEM SETTING

- **Given** Observed variables X and latent confounders Z in a set of contexts C
- **Causal Mechanism Shifts** modeled as set partitions $\Pi_i^* = \{\pi_i^1, \dots, \pi_i^{k_i}\}$ of C
- **Independent changes** $P(\Pi^*) = \prod_{V_i} P(\Pi_i^*)$, by modularity of causal mechanisms
- **Idea** Confounders create measurable dependencies in observed set partitions

MEASURING DEPENDENCY OF MECHANISM SHIFTS

- **Mutual Information (MI)** of partitions $\mathrm{I}(\Pi_1, \Pi_2) = \sum_{ij} \frac{n_{ij}}{N} \log \frac{n_{ij} N}{u_i v_j}$
- **Expected MI** under independence $\mathbb{E}[\mathrm{I}(\Pi_1', \Pi_2')] = \sum_{ij} \sum_{n_{ij}} \mathrm{I}(n_{ij}) \mathcal{P}(n_{ij} \mid u, v, N)$ — hypergeometric distribution (Vinh et al. 2010)
- **Confounding Test** for a pair Π_1, Π_2 $t = \frac{I(\Pi_1, \Pi_2) - \mathbb{E}[\mathrm{I}(\Pi_1', \Pi_2')]}{\sqrt{\mathrm{Var}(\mathrm{I}(\Pi_1', \Pi_2'))}}$

IDENTIFYING CONFOUNDING USING MI

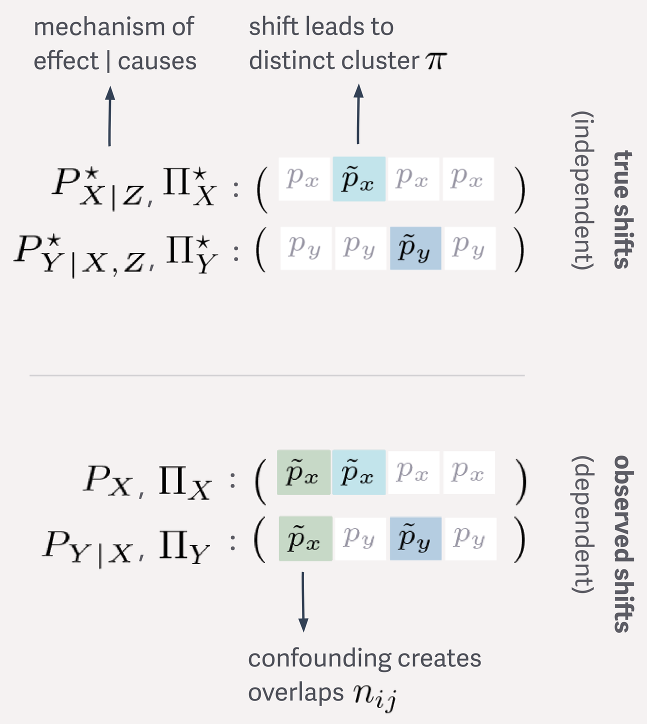
- **Sparse changes** $p = P(\Pi^*(C) \neq \Pi^*(C')) < 0.5$ as key assumption, based on invariance of causal mechanisms
- **Pairwise Confounding** We confirm that we can use MI over partitions to test whether a variable pair is confounded

Lemma 1 We can identify pairwise confounding with a power of 1 in the limit, $\lim_{n_c \to \infty} \mathcal{P}(t > q_{1-\alpha}) \to 1$ and conversely have $\lim_{n_c \to \infty} \mathcal{P}(t > q_{1-\alpha}) \to \alpha$ for unconfounded variables for quantile $q_{1-\alpha}$ of the normal distribution for any $\alpha > 0$.

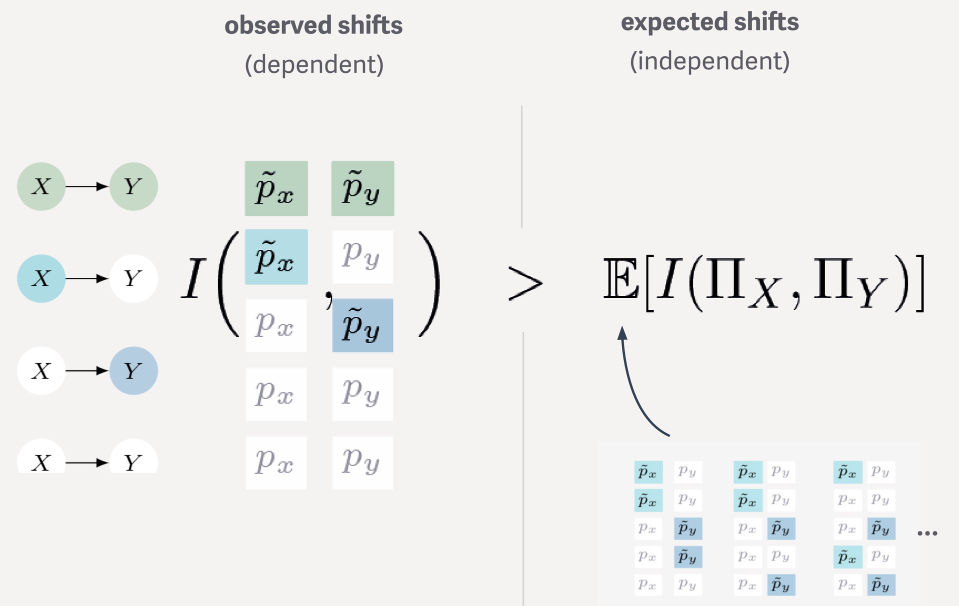
- **Unknown causal directions** We obtain consistency for the multivariate case with unknown causal directions when combining our test with the Minimal Shift Score *(Perry et al. 2022)* to discover the causal directions under confounding

Theorem 1 (informal) The graph and partitions minimizing the number of causal mechanism shifts are the unique minumum of the total correlation given by $\sum_i I(\Pi_i, \Pi_{>i} \mid \Pi_{<i})$ with high probability.

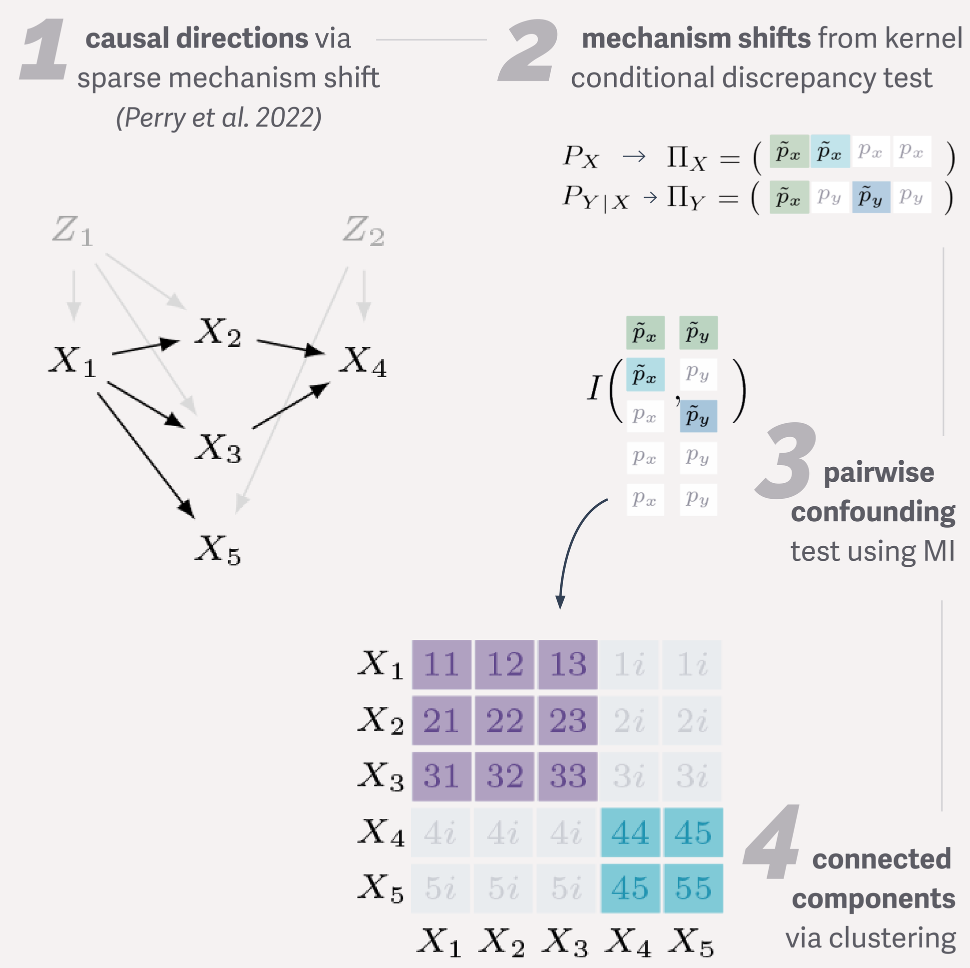
EXAMPLE SETTING



MEASURING CONFOUNDING

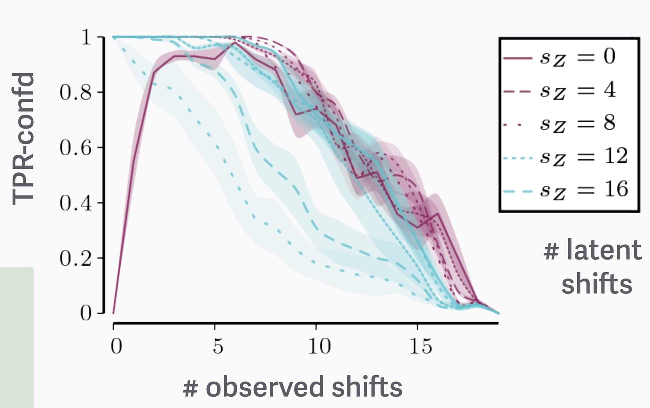


DISCOVERING CONFOUNDERS



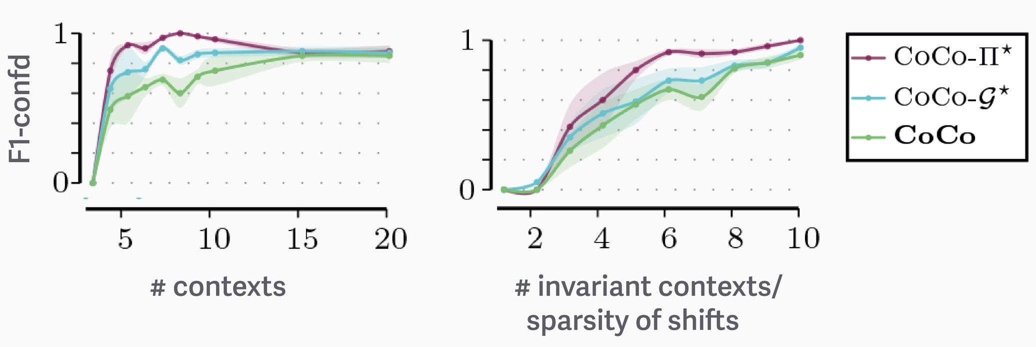
SIMULATIONS

- effect of the number of mechanism changes



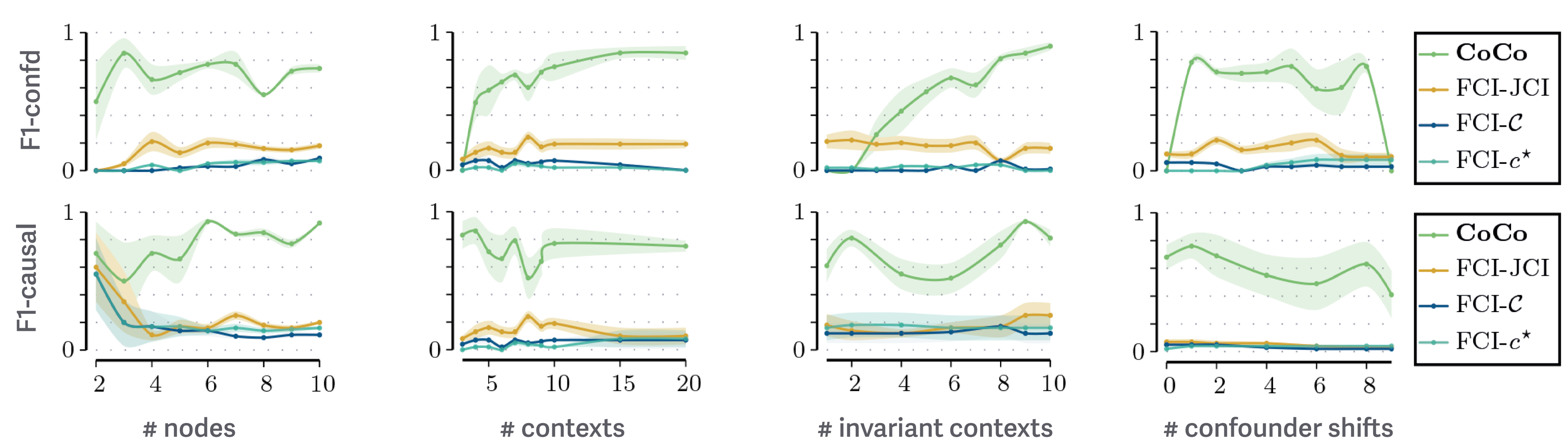
matches identifiability under sparse shift assumption

- oracles for 1 causal directions *(blue)* and 2 mechanism shifts *(purple)* compared to full method *(green)*



Synthetic Data

- confounded variable pairs
- causal directions under confounding



Flow Cytometry Data

- protein cells in different interventional conditions *(Sachs et al. 2006)*
- discover confounding effects of PKC *(solid green)*
- discover potential confounding or feedback between Raf and Mek

consistent with observations: joint shift of Raf and Mek under reagent U0126