Limits in Causal Discovery and the Path Forward

Mateusz Gajewski

Poznań University of Technology,
IDEAS NCBR

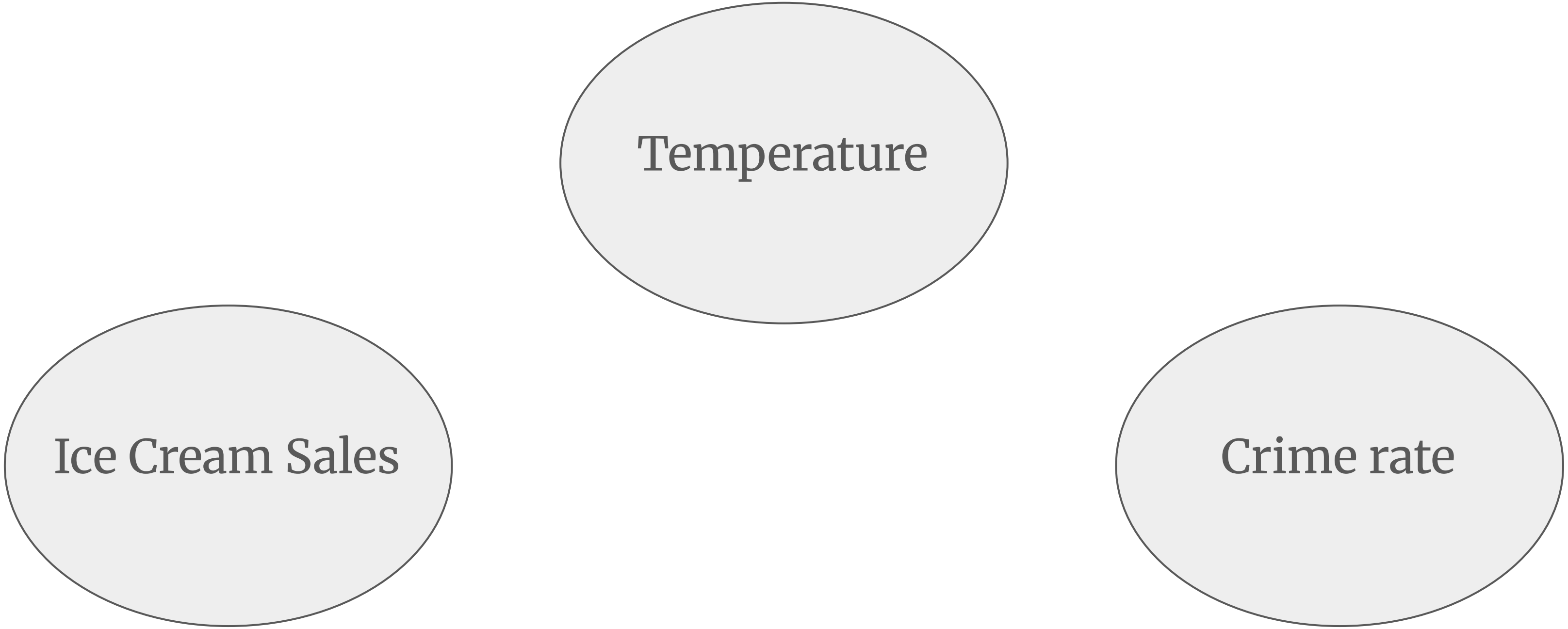
Mateusz Olko

University of Warsaw,
IDEAS NCBR

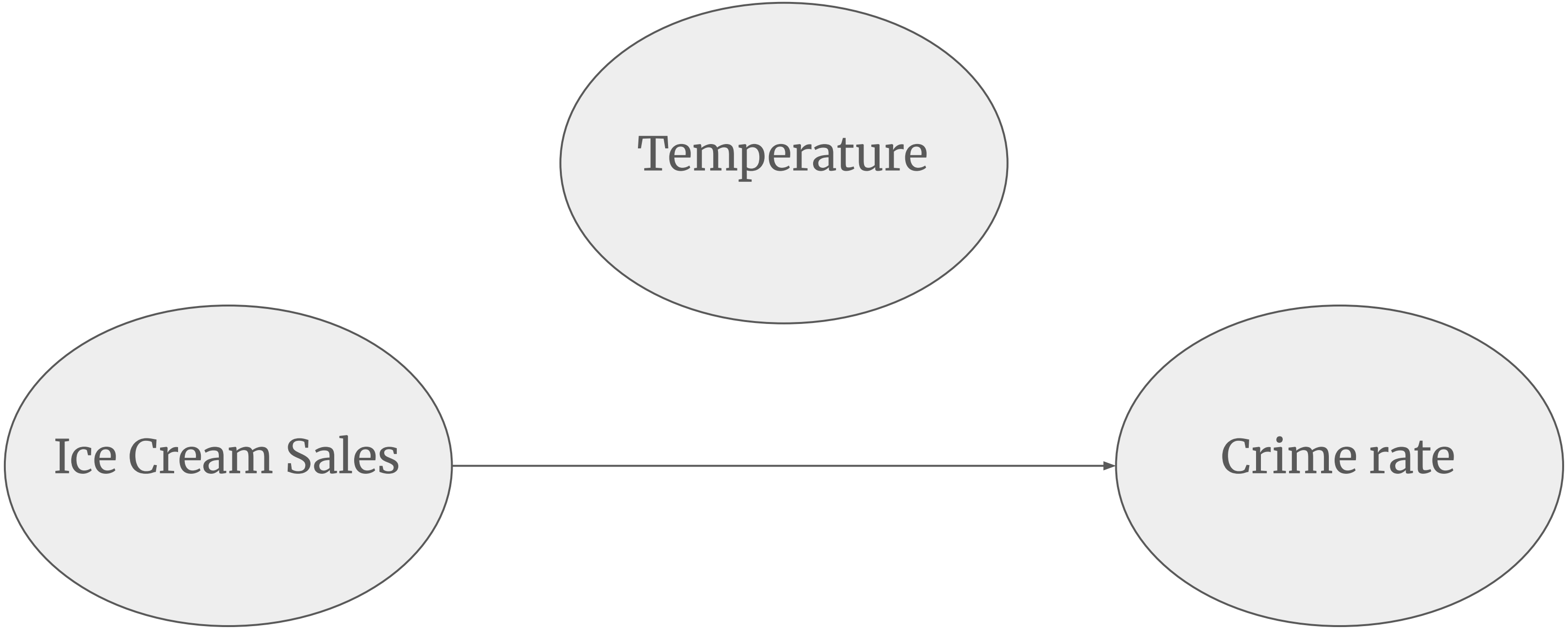
Plan of this presentation

- Introduction to causal discovery and recent advances.
- Faithfulness and limitations of causal discovery.
- The path forward.

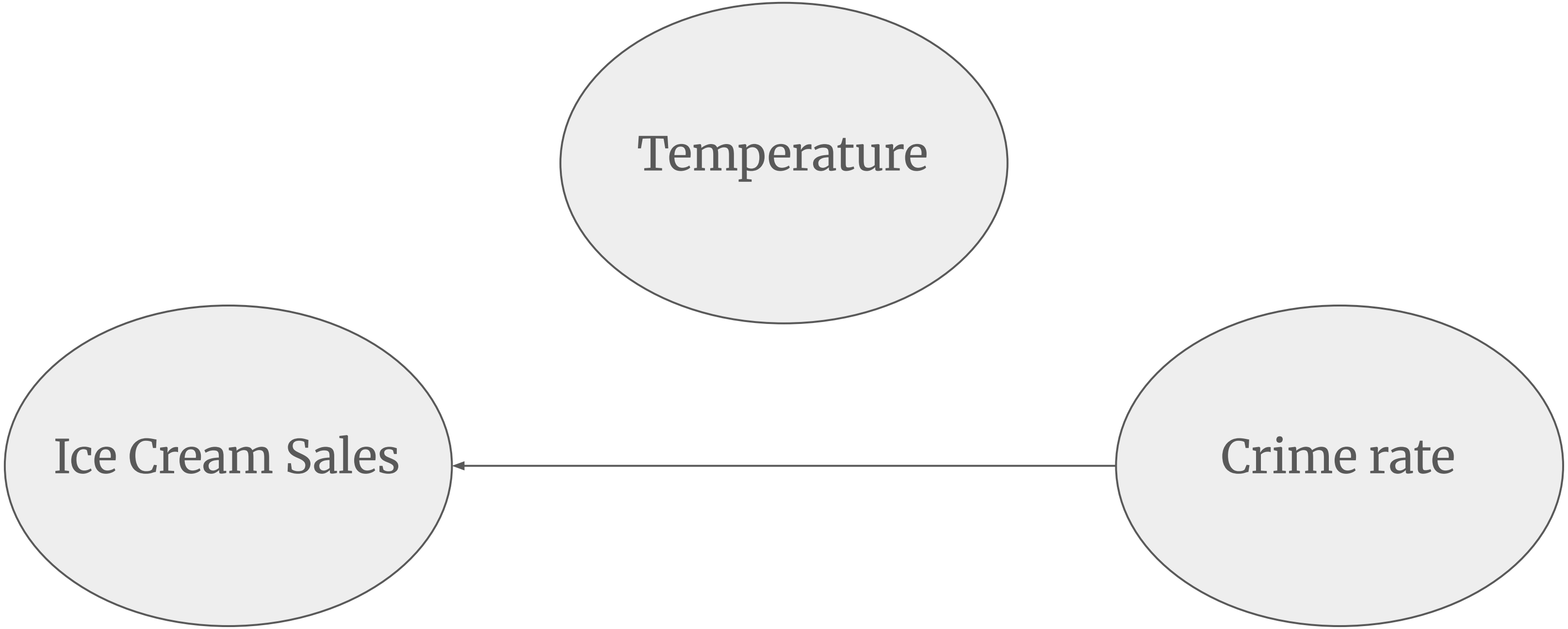
Causation & Correlation



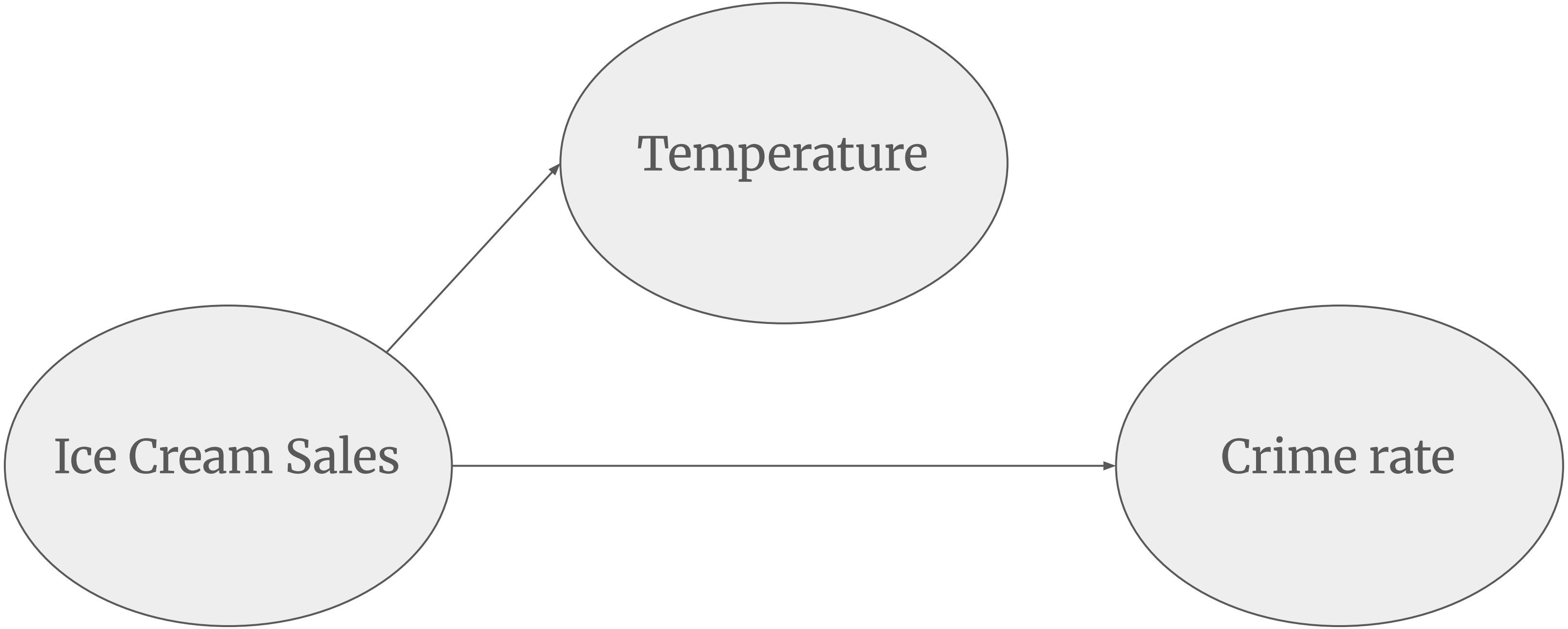
Causation & Correlation



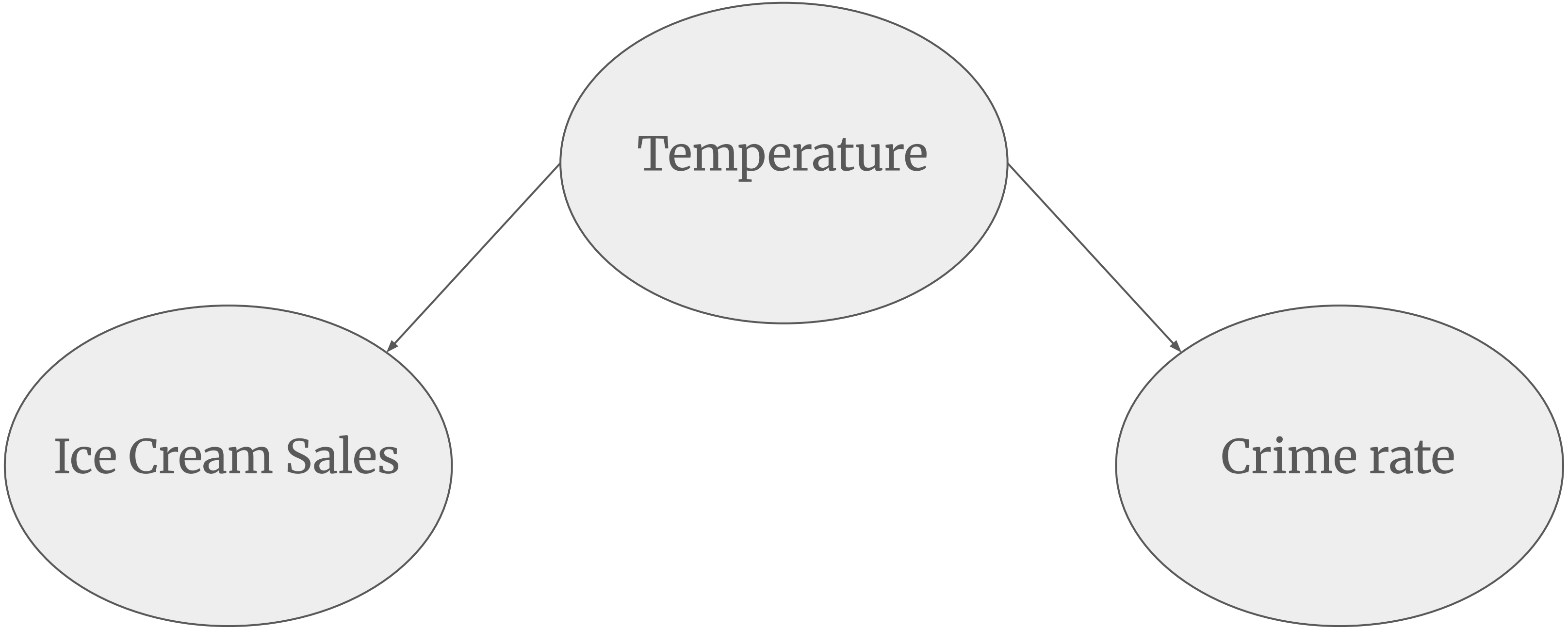
Causation & Correlation



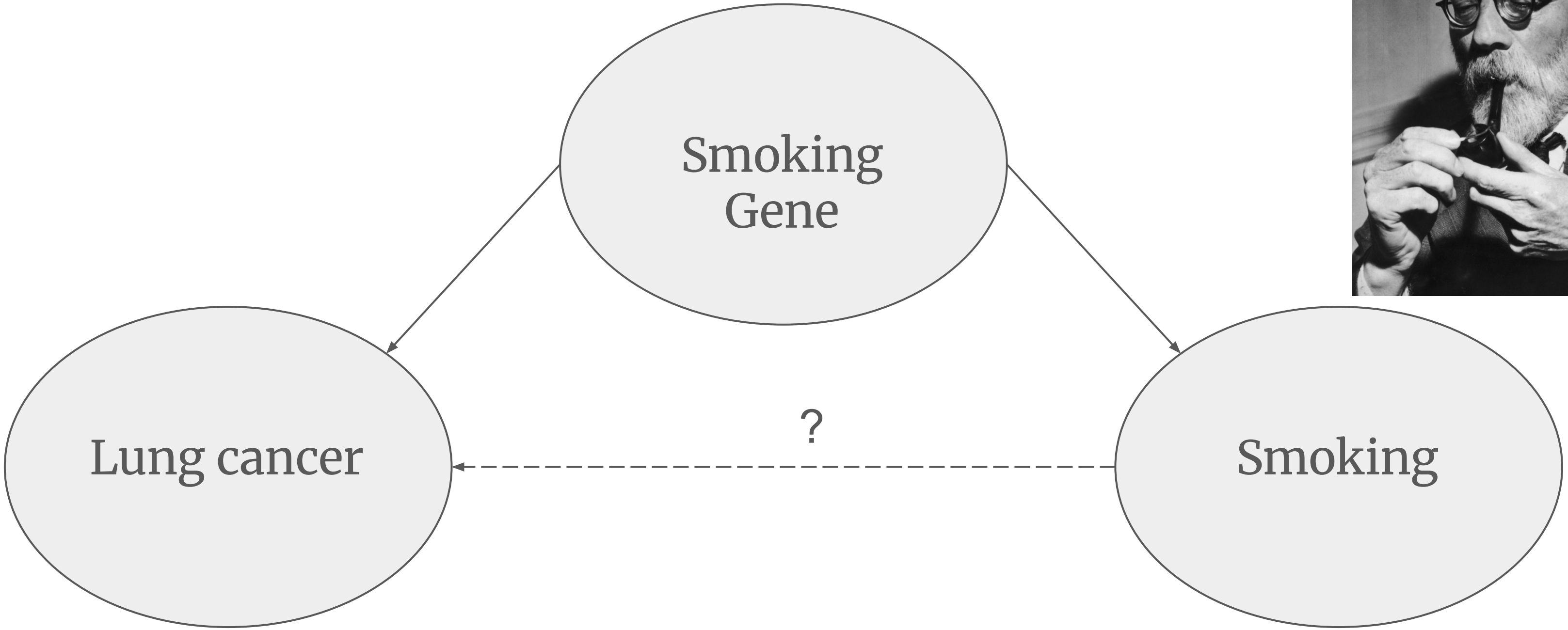
Causation & Correlation

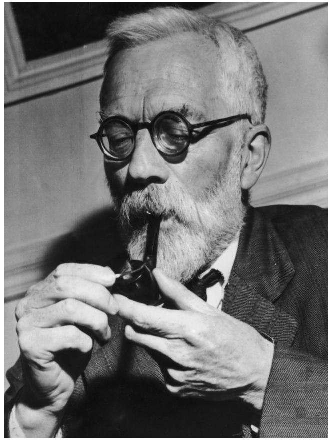


Causation & Correlation

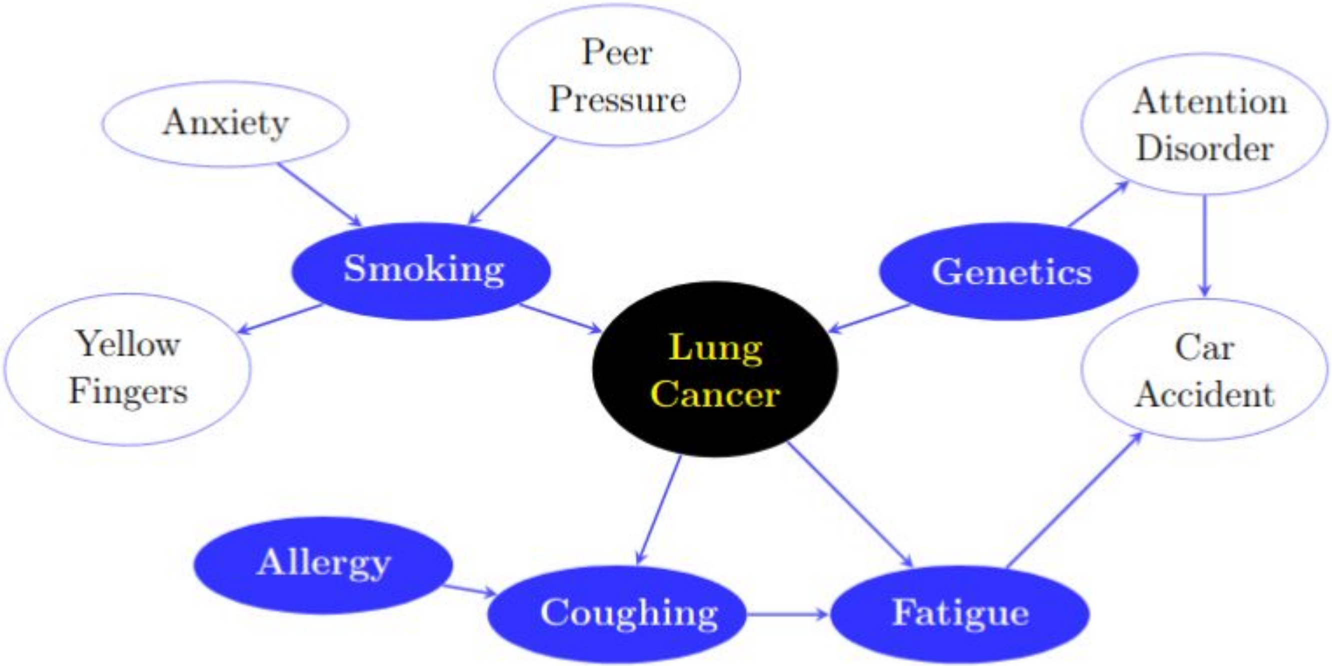


Causation & Correlation





Causality



Yu, Kui, et al. "Causality-based feature selection: Methods and evaluations."

Causality use cases

- Biology
- Neurosciences
- Earth Sciences
- Economy

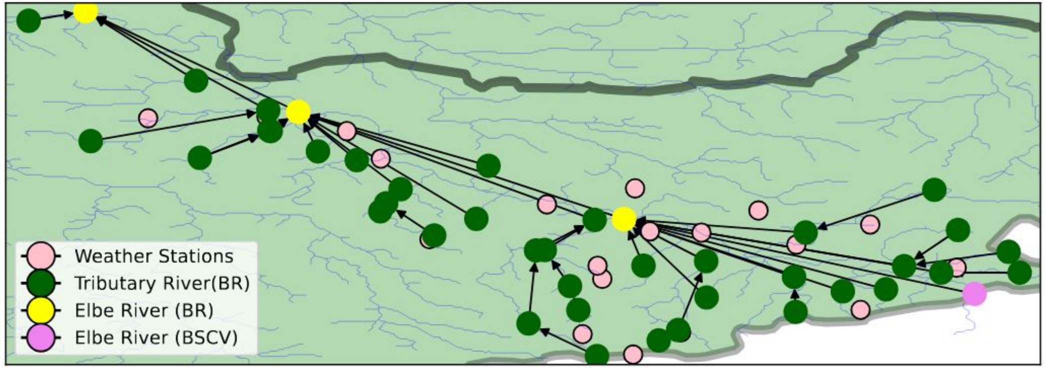
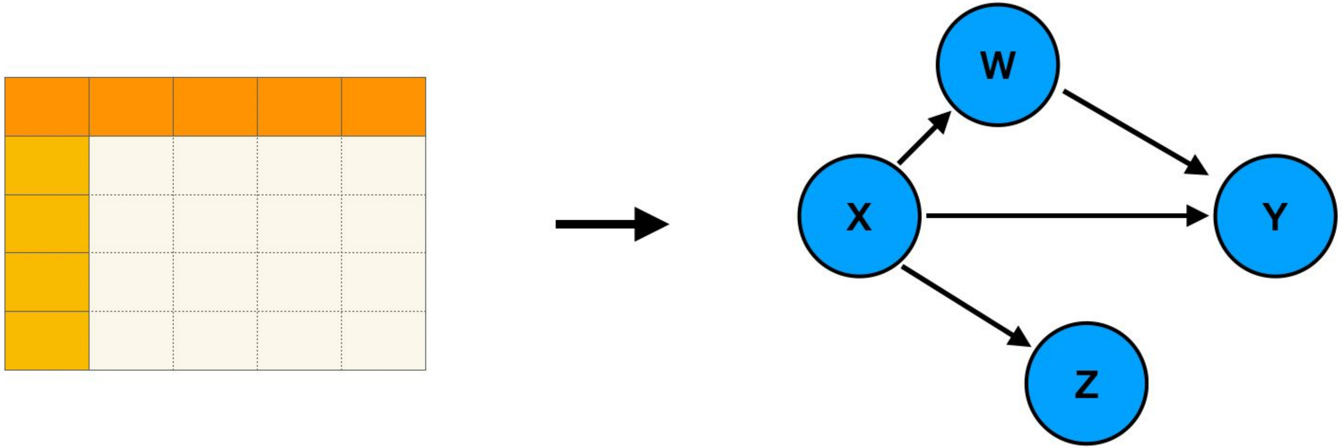


Figure 8: Causal ground truth graph of RiversFlood. This graph is a subset of RiversEastGermany. Weatherstations that were used to investigate the general precipitation levels are depicted in pink.

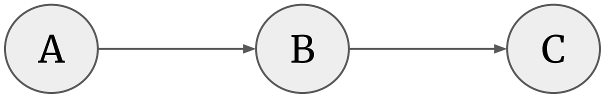
Stein, Gideon, et al. "CausalRivers--Scaling up benchmarking of causal discovery for real-world time-series."

Causal Discovery

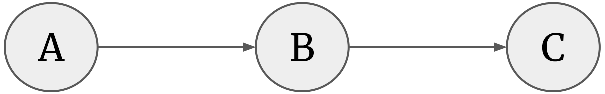


https://towardsdatascience.com/causal-discovery-6858f9af6dcb/

Causal discovery - PC

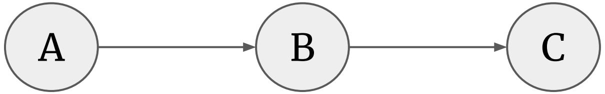


Causal discovery - PC



$$A \perp\!\!\!\perp B$$
$$A \perp\!\!\!\perp C$$
$$B \perp\!\!\!\perp C$$

Causal discovery - PC



$$A \perp\!\!\!\perp B \mid C$$ ❌

$$A \perp\!\!\!\perp C \mid B$$ ✅

$$B \perp\!\!\!\perp C \mid A$$ ❌

Causal discovery - modern developments

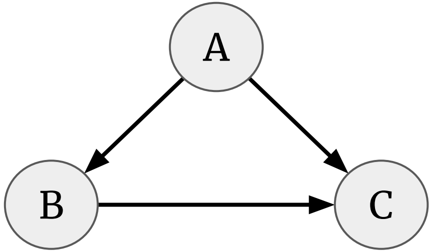
- Recent developments in causal discovery focuses on scalability and efficiency of the methods.
- Modelling complex and nonlinear relations.
- Integrating into causal discovery process, tools known from deep learning.

$$\begin{aligned} \min_{G} \quad & score(G) \\ \text{s.t.} \quad & G \in \text{DAGs} \end{aligned} \quad \iff \quad \begin{aligned} \min_{W \in \mathbb{R}^{d \times d}} \quad & score(W) \\ \text{s.t.} \quad & h(W) = 0 \end{aligned}$$

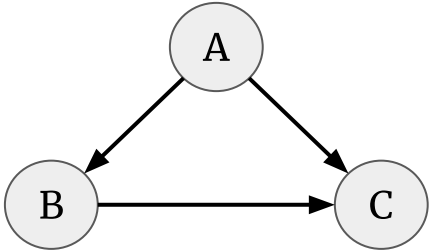
Zheng, Xun, et al. "Dags with no tears: Continuous optimization for structure learning."

The limitations

The faithfulness



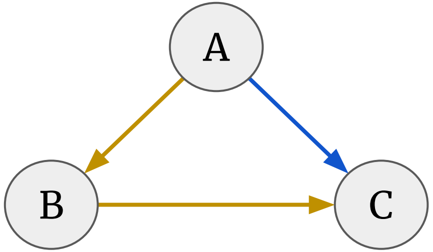
The faithfulness



$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$

$A := \varepsilon_a$
$B := 2A + \varepsilon_b$
$C := 3B - 6A + \varepsilon_c$

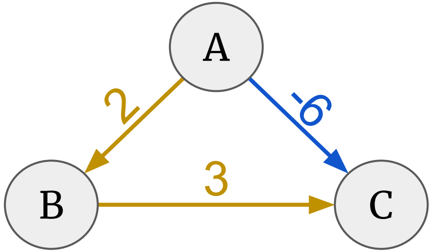
The faithfulness



$$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$$

$$A := \varepsilon_a$$
$$B := 2A + \varepsilon_b$$
$$C := 3B - 6A + \varepsilon_c$$

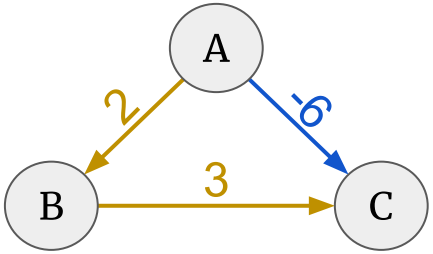
The faithfulness



$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$

$A := \varepsilon_a$
$B := 2A + \varepsilon_b$
$C := 3B - 6A + \varepsilon_c$

The faithfulness



$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$

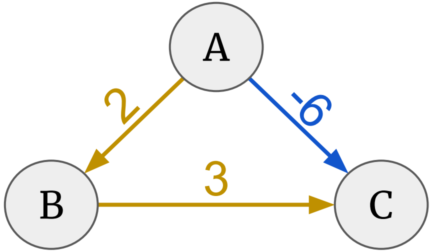
$A := \varepsilon_a$

$B := 2A + \varepsilon_b$

$C := 3(2A + \varepsilon_b) - 6A + \varepsilon_c$

$2 * 3 - 6 = 0$

The faithfulness

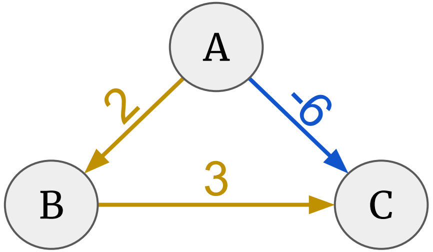


$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$

$$A := \varepsilon_a$$
$$B := 2A + \varepsilon_b$$
$$C := 3\varepsilon_b + \varepsilon_c$$

just indep. noise terms

The faithfulness

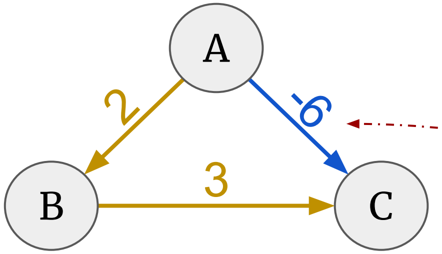


$$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$$

$$A := \varepsilon_a$$
$$B := 2A + \varepsilon_b$$
$$C := 3\varepsilon_b + \varepsilon_c$$

$$A \perp\!\!\!\perp C$$

The faithfulness



$\varepsilon_a, \varepsilon_b, \varepsilon_c \sim N(0,1)$

$A := \varepsilon_a$

$B := 2A + \varepsilon_b$

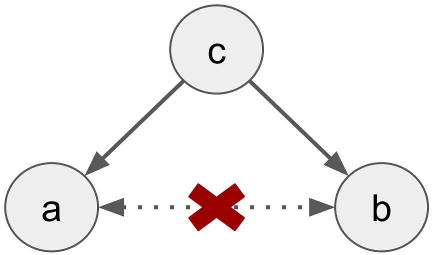
$C := 3\varepsilon_b + \varepsilon_c$

$A \perp\!\!\!\perp C$

The faithfulness

Conditional independencies in data correspond to causal structure in the underlying graph.

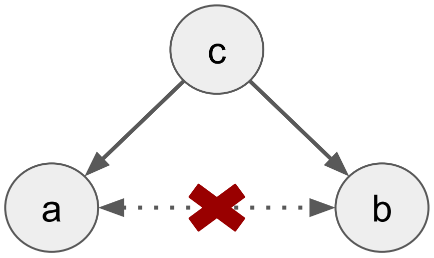
$$a \perp\!\!\!\perp b \mid c \quad \Longleftrightarrow$$



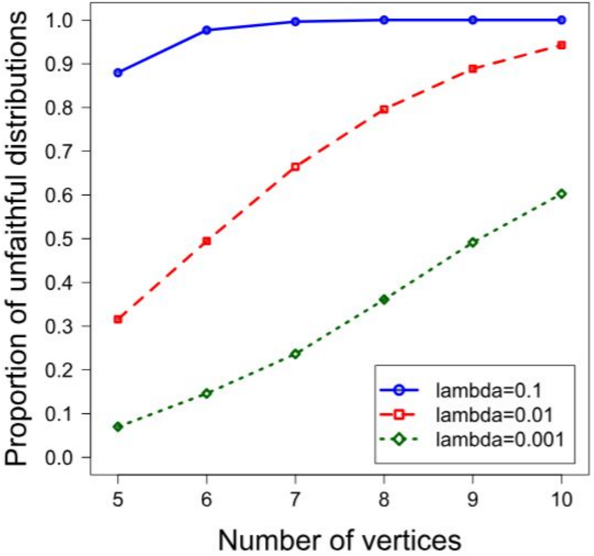
Lambda-strong faithfulness

Thresholded correlation between variables in data correspond to causal structure in the underlying graph.

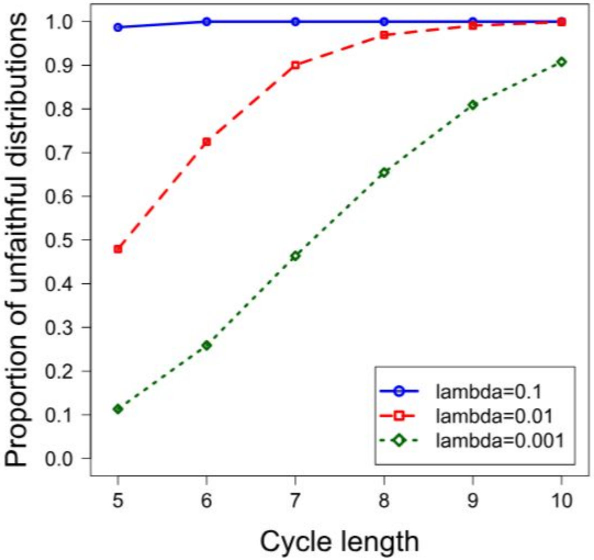
$\rho(a, b \mid c) <= \lambda \quad \Longleftrightarrow$



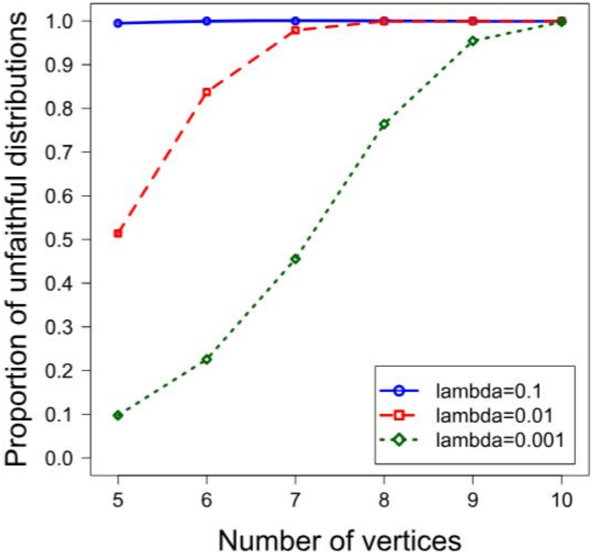
Geometry of the faithfulness assumption in linear systems



(a) trees T_p



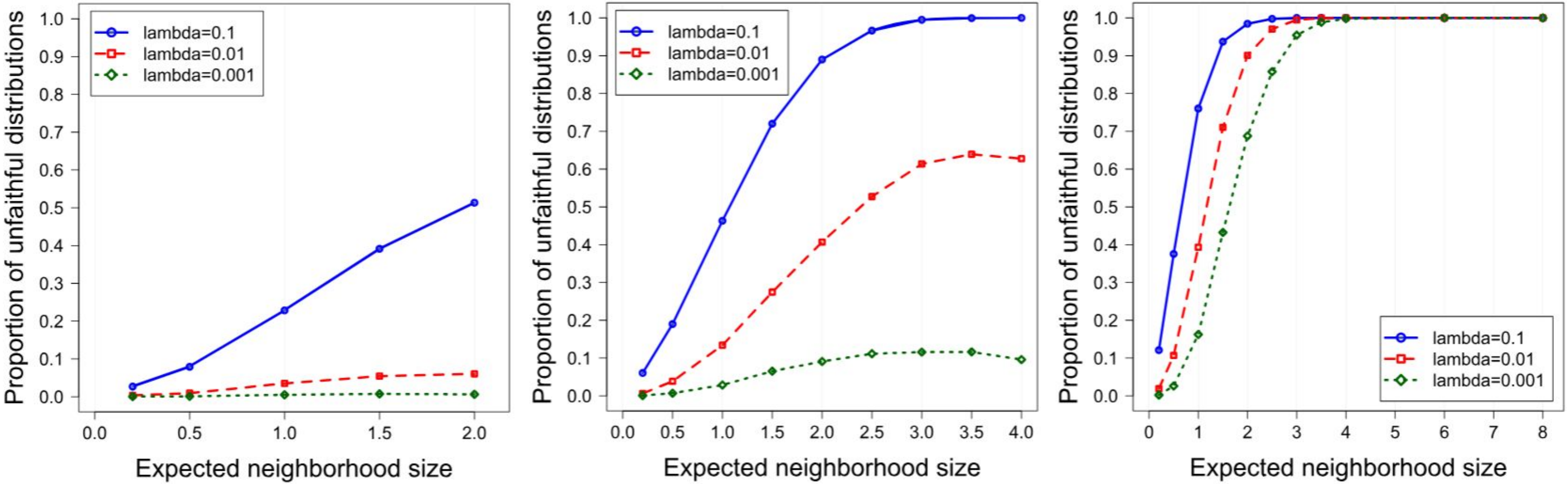
(b) cycles C_p



(c) bipartite graphs $K_{2,p-2}$

Uhler et al., Geometry of the faithfulness assumption in causal inference

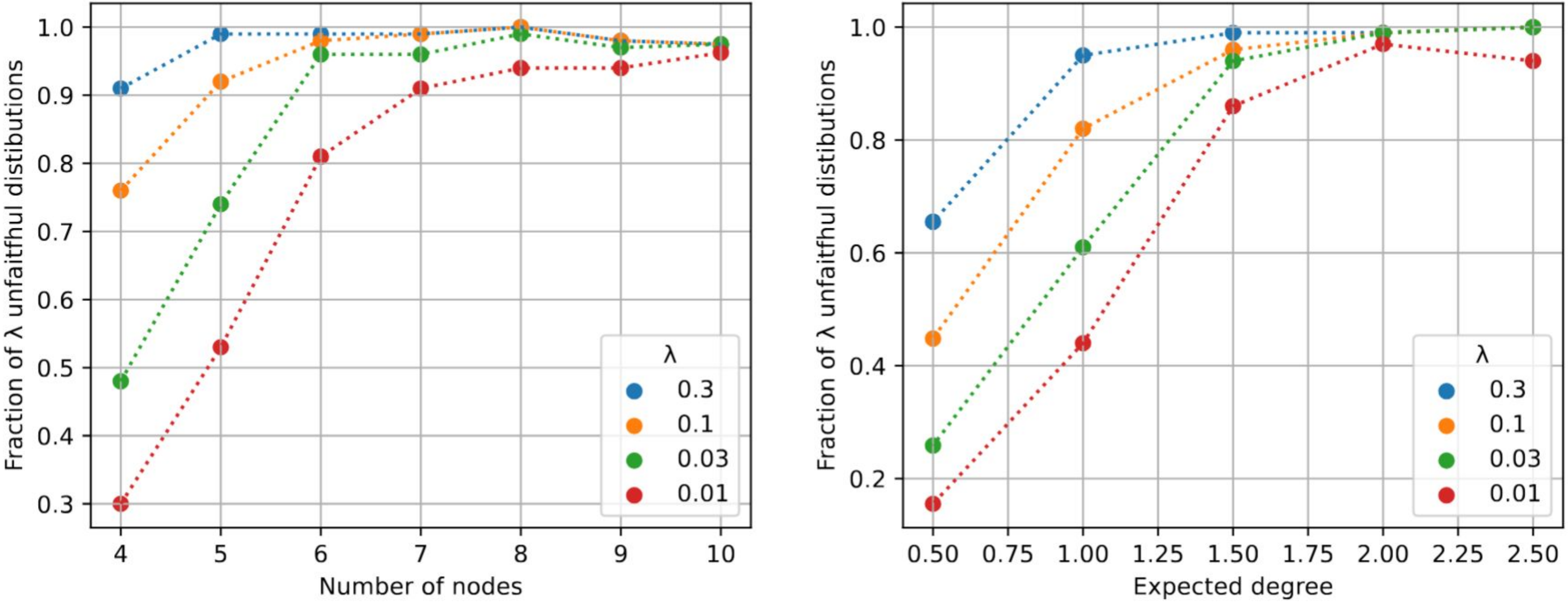
Geometry of the faithfulness assumption in linear systems



(a) 3-node DAGs (b) 5-node DAGs (c) 10-node DAGs

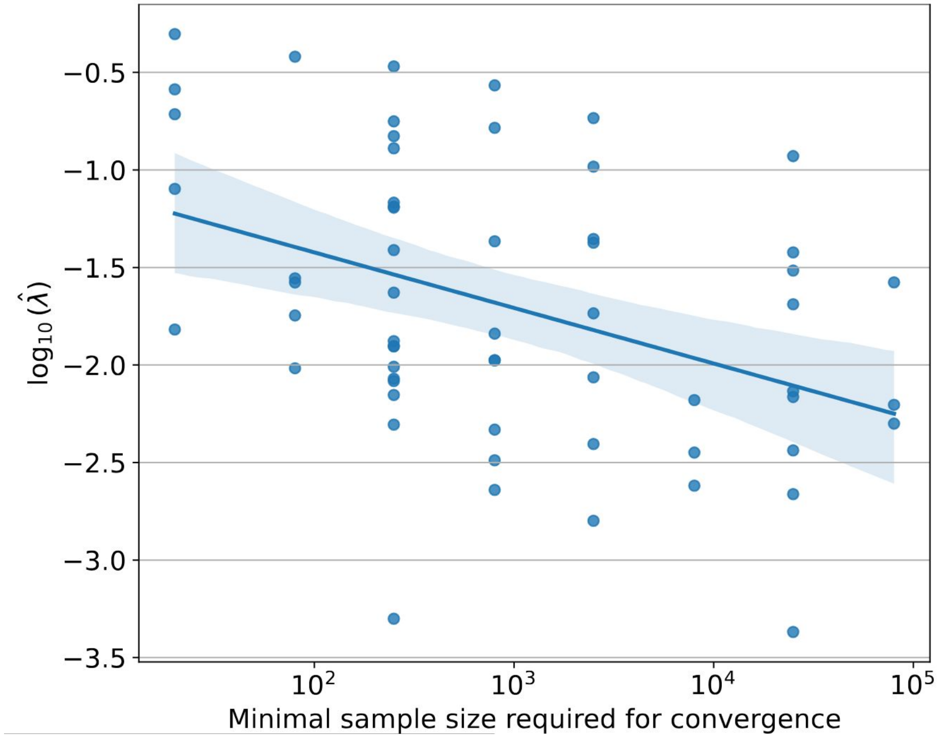
Uhler et al., Geometry of the faithfulness assumption in causal inference

Non-linear geometry of faithfulness



Olko et al., Since Faithfulness Fails: The Performance Limits of Neural Causal Discovery

Lambda vs data requirements

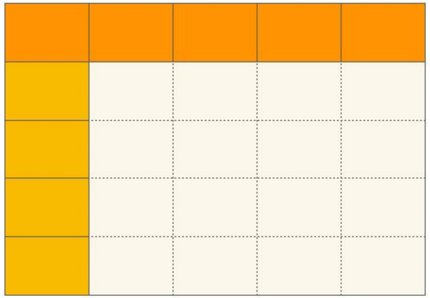


Olko et al., Since Faithfulness Fails: The Performance Limits of Neural Causal Discovery

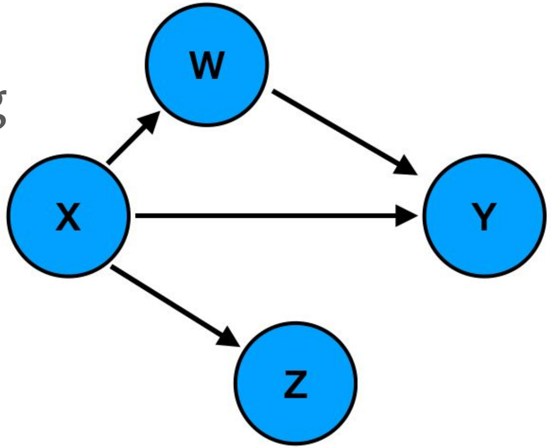
The path forward

- Amortized approach
- Local causal discovery
- Grounding evaluations in real world systems

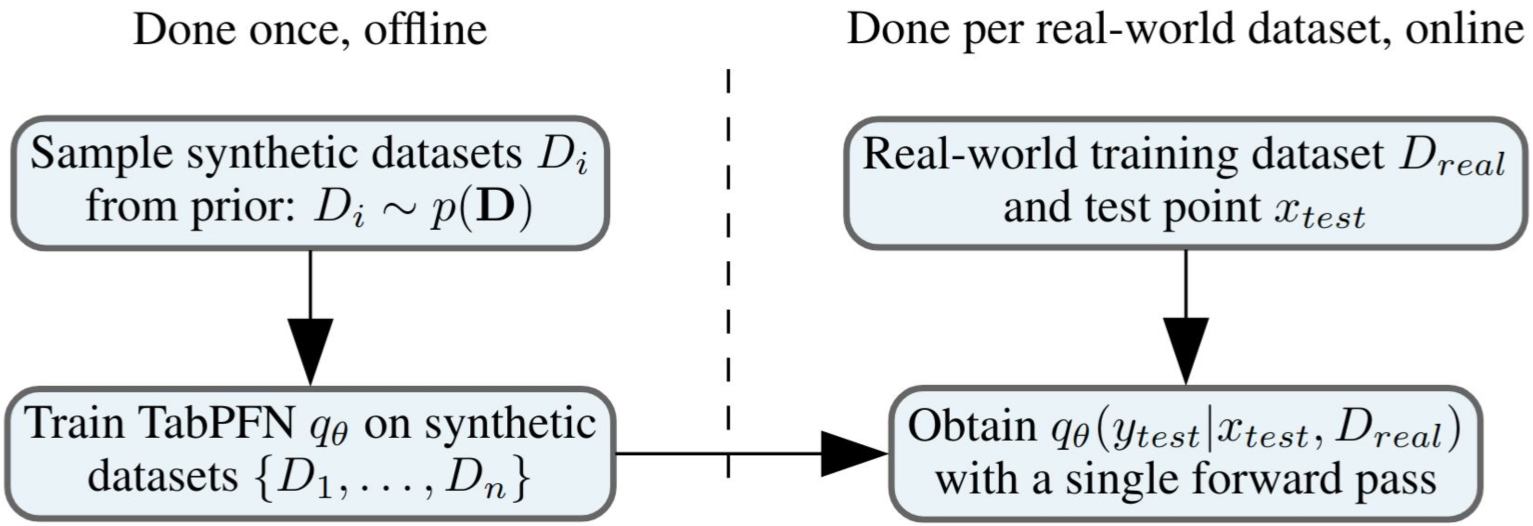
Amortized approach



Deep learning model

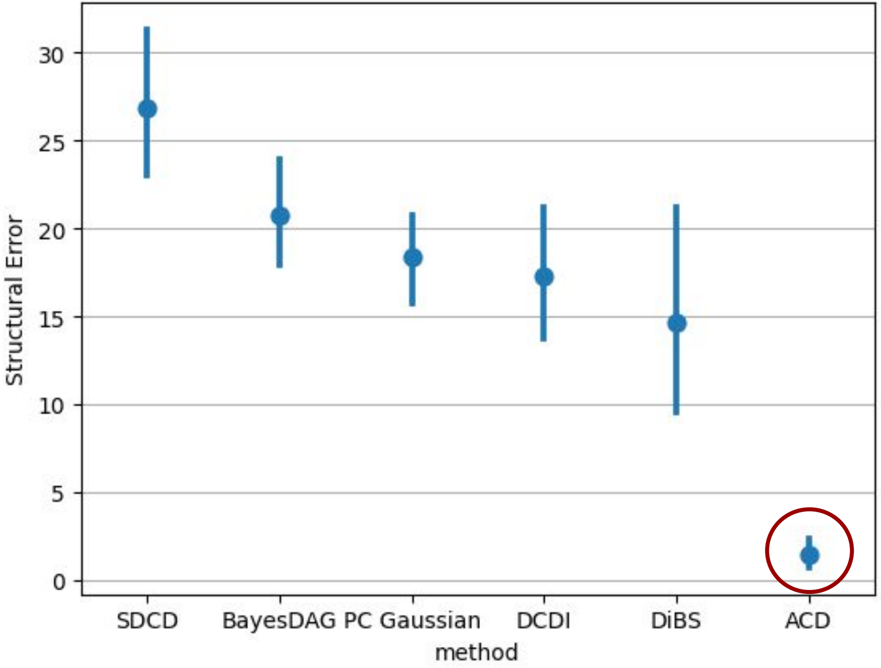


Amortized approach - Bayesian



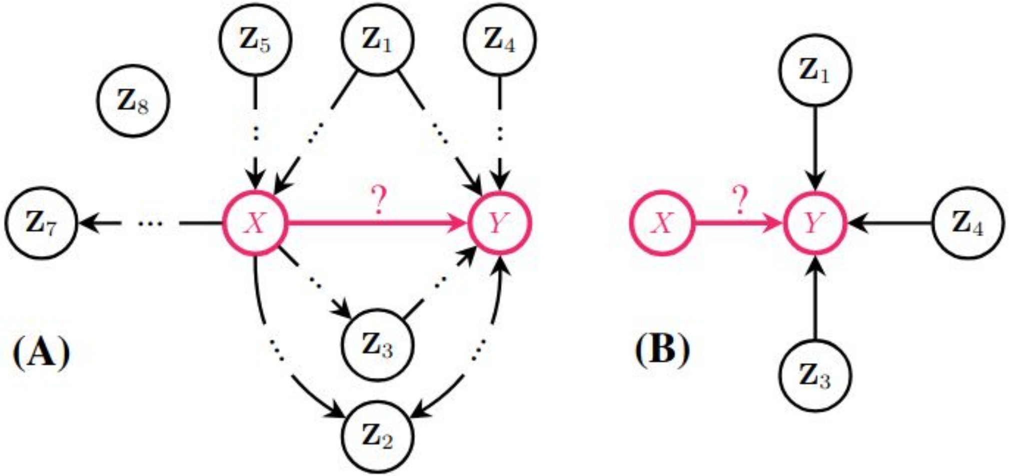
Hollmann et al., TabPFN: A Transformer That Solves Small Tabular Classification Problems in a Second

Amortized approach - Bayesian



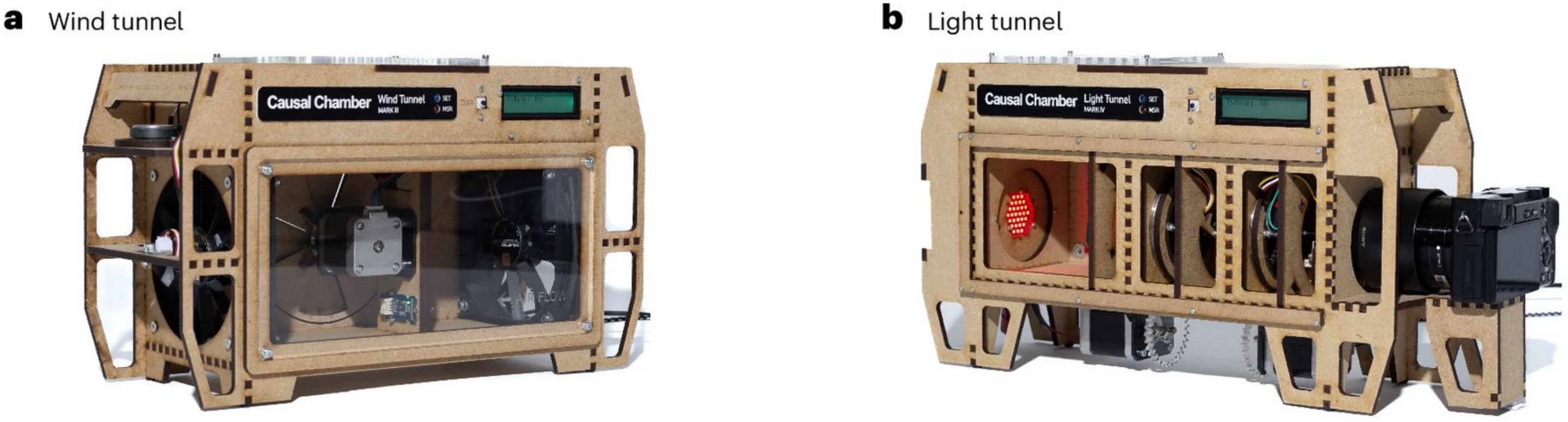
Local causal discovery

- Do not discover a full causal graph.
- Partition specific graph nodes together.
- Easier task, but when defined right can allow to recover useful informations.



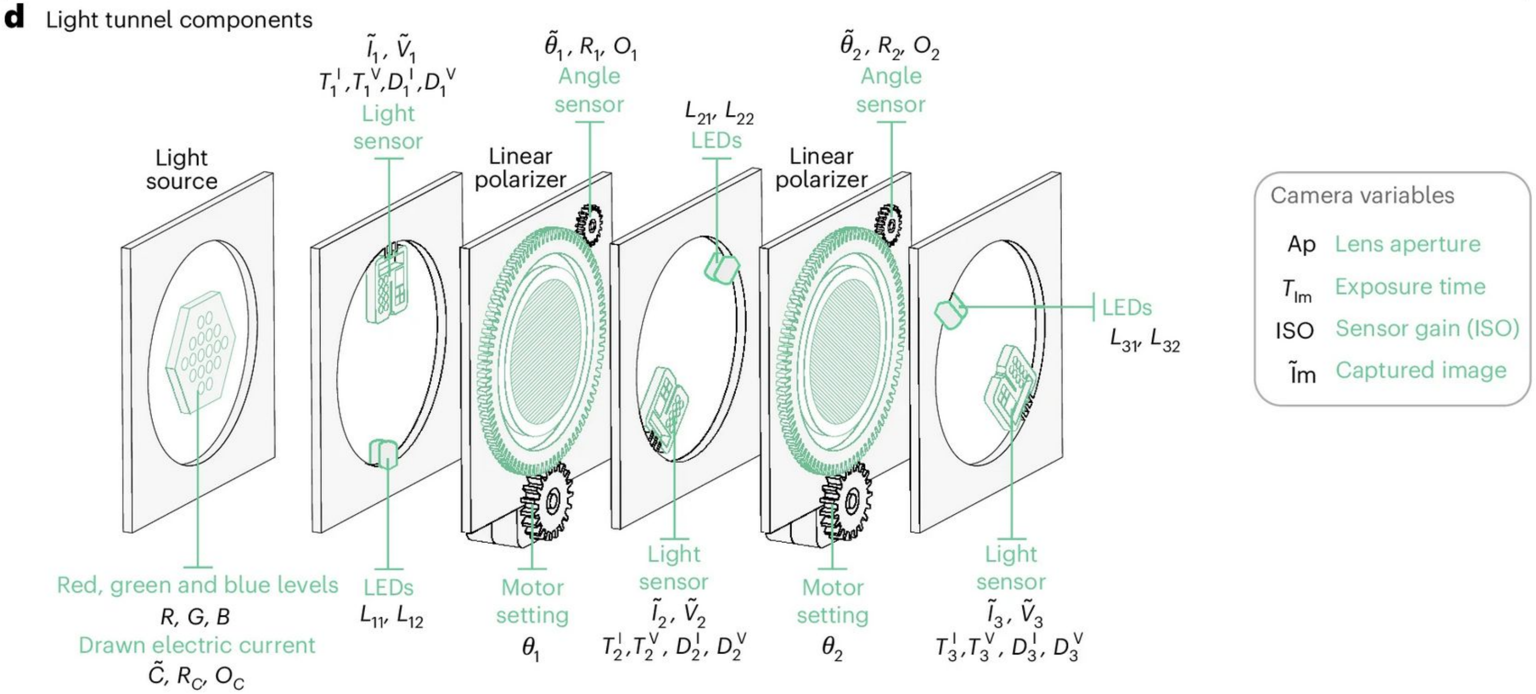
Maasch, Jacqueline, et al. "Local Causal Discovery for Structural Evidence of Direct Discrimination."

Grounding evaluations in a real world systems



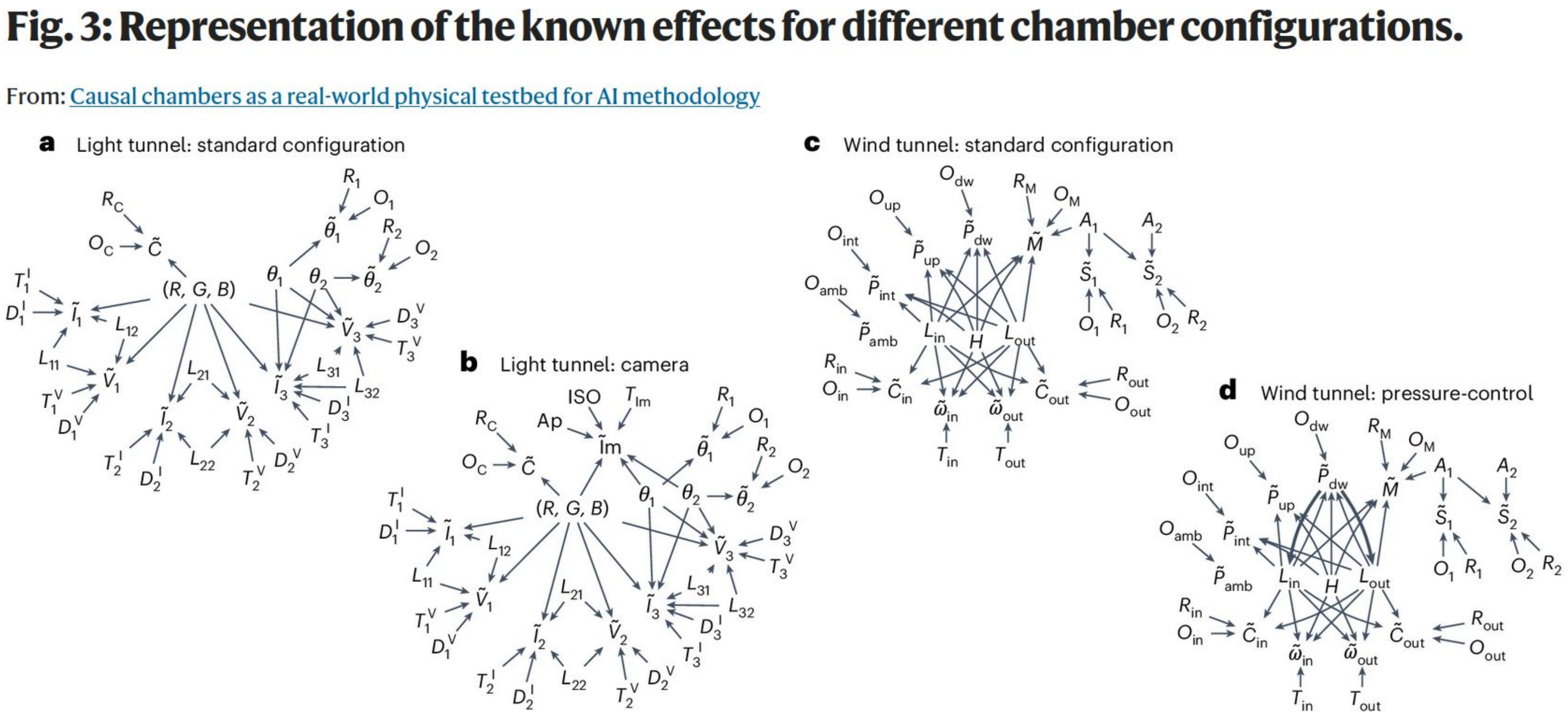
Gamella, Juan L., Jonas Peters, and Peter Bühlmann. "Causal chambers as a real-world physical testbed for AI methodology."

Grounding evaluations in a real world systems



Gamella, Juan L., Jonas Peters, and Peter Bühlmann. "Causal chambers as a real-world physical testbed for AI methodology."

Grounding evaluations in a real world systems



Gamella, Juan L., Jonas Peters, and Peter Bühlmann. "Causal chambers as a real-world physical testbed for AI methodology."

Thank you for your attention!