Sensing is life

Quantitative probing

Validating causal models using quantitative domain knowledge

Daniel Grünbaum, Maike Stern, Elmar W. Lang (R&D @ ams OSRAM / Universität Regensburg)
07.11.22

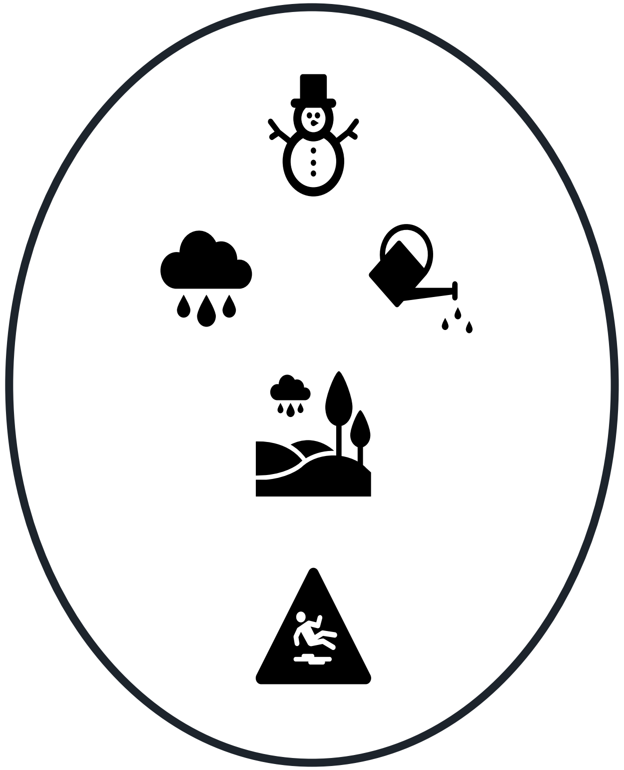
Graph-based Causal Inference

Pearl‘s sprinkler example

Example: „Does my lawn get slippery if I turn on the sprinkler?“

Given: observational data of the form

Season	Sprinkler	Rain	Wet	Slippery
Summer	1	0	1	1
Spring	0	1	1	1
Winter	0	0	0	0



Goal: Determine all causal mechanisms!

Causal end-to-end analysis

Connect causal discovery and causal estimation

Domain knowledge

Nothing causes `Season`

`Slippery` causes nothing

`Sprinkler` does not cause `Rain`

`Sprinkler` directly causes `Wet`

Data

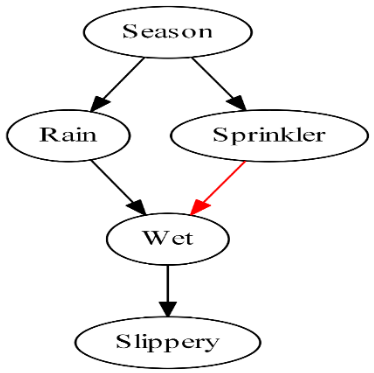
Season	Sprinkler	Rain	Wet	Slippery
Summer	1	0	1	1
Spring	0	1	1	1
Winter	0	0	0	0

Causal estimates

`Wet` does not causally affect `Rain`

Turning on `Sprinkler` strongly increases `Slippery`

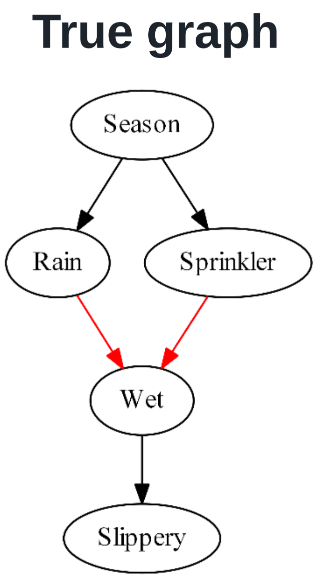
Graphical model

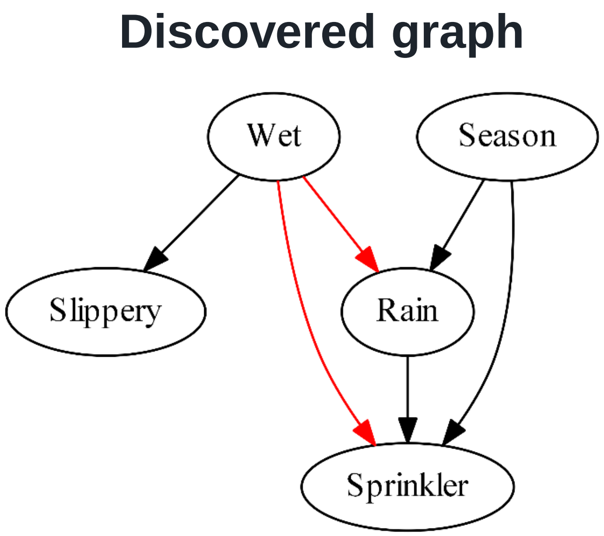


Daniel Grünbaum, Maike Stern, Elmar W. Lang

Motivation

When can we trust the result of a causal analysis?





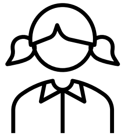
Red: edges required by domain knowledge

What if we messed up and the resulting model is wrong?

- In academia: no one cares after the paper is accepted
- In industry: …

Motivation

A causal horror story in three acts



Boss

My daughter slipped on the lawn and broke her arm.
How can I avoid such accidents?
I suspect it‘s the sprinkler!

Gardener

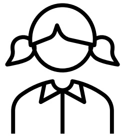
Finally causal gardening school pays off!
I will evaluate observational data and return with safety advice.

Motivation

A causal horror story in three acts

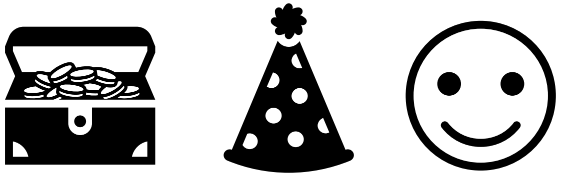
Gardener

My causal analysis says that turning on the sprinkler does not make the lawn more slippery! Turn it on as often as you like!



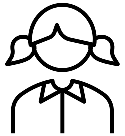
Boss

Thanks, you get a raise for your efforts!



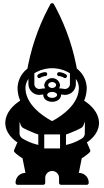
Motivation

A causal horror story in three acts



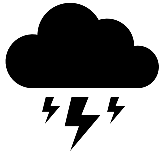
Boss

I turned on the sprinkler and she slipped again!
You are fired!!!
Damn you and your causal models!



Gardener

Noooooo, I have a family to feed!
Curse you, causal models!
I should have gone to correlation-based gardening school instead!

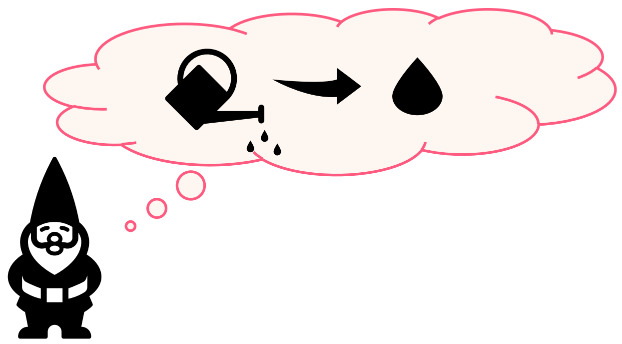
 Causal models are abandoned forever… 

Quantitative probing

Applying the scientific method to causality

This could have been avoided by using quantitative domain knowledge:

- We know that the lawn becomes more wet if we turn on the sprinkler. ➔ **ATE of Sprinkler on Wet > 0**
- We can probe our model by checking if it correctly recovers this and other known effects (quantitative probes).



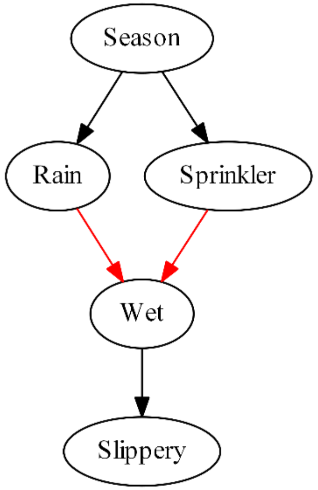
A scientific theory cannot be proven, but only falsified. If our attempts to falsify it fail, this increases our belief in the validity of the theory. *(paraphrasing Popper)*

Daniel Grünbaum, Maike Stern, Elmar W. Lang

Quantitative probing

Applying the scientific method to causality

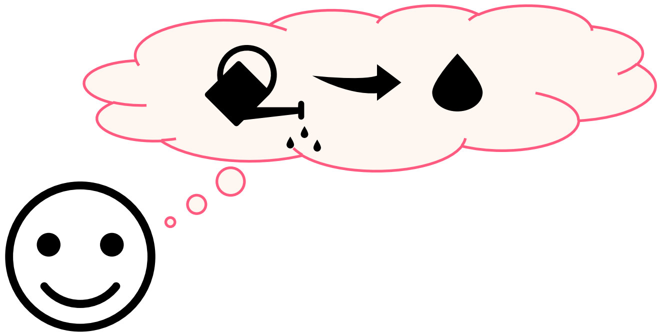
True graph



Passed validations (2/2):

Effect Type	Treatment	Outcome	Estimated	Expected
overall	Sprinkler	Wet	0.62	greater than 0
overall	Wet	Slippery	0.81	greater than 0

No failed validations.



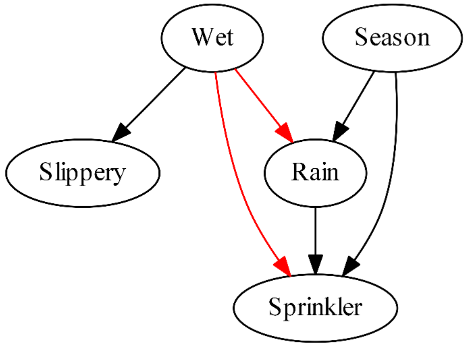
Trust in true model is increased!

Daniel Grünbaum, Maike Stern, Elmar W. Lang

Quantitative probing

Applying the scientific method to causality

Discovered graph



Passed validations (1/2):

Effect Type	Treatment	Outcome	Estimated	Expected
overall	Wet	Slippery	0.81	greater than 0

Failed validations (1/2):

Effect Type	Treatment	Outcome	Estimated	Expected
overall	Sprinkler	Wet	0.00	greater than 0

This model cannot be trusted!

Misspecified model is detected!

Simulation study

Collecting evidence for the effectiveness of quantitative probing

1) Ground truth

- Draw random DAG over 7 nodes
- Draw random binary conditional probability distributions over each node
- Choose random nontrivial target effect (ATE) and 50% of effects as quantitative probes (including tolerance of 0.1)
- Calculate true values of target effect and quantitative probes

2) Learned model

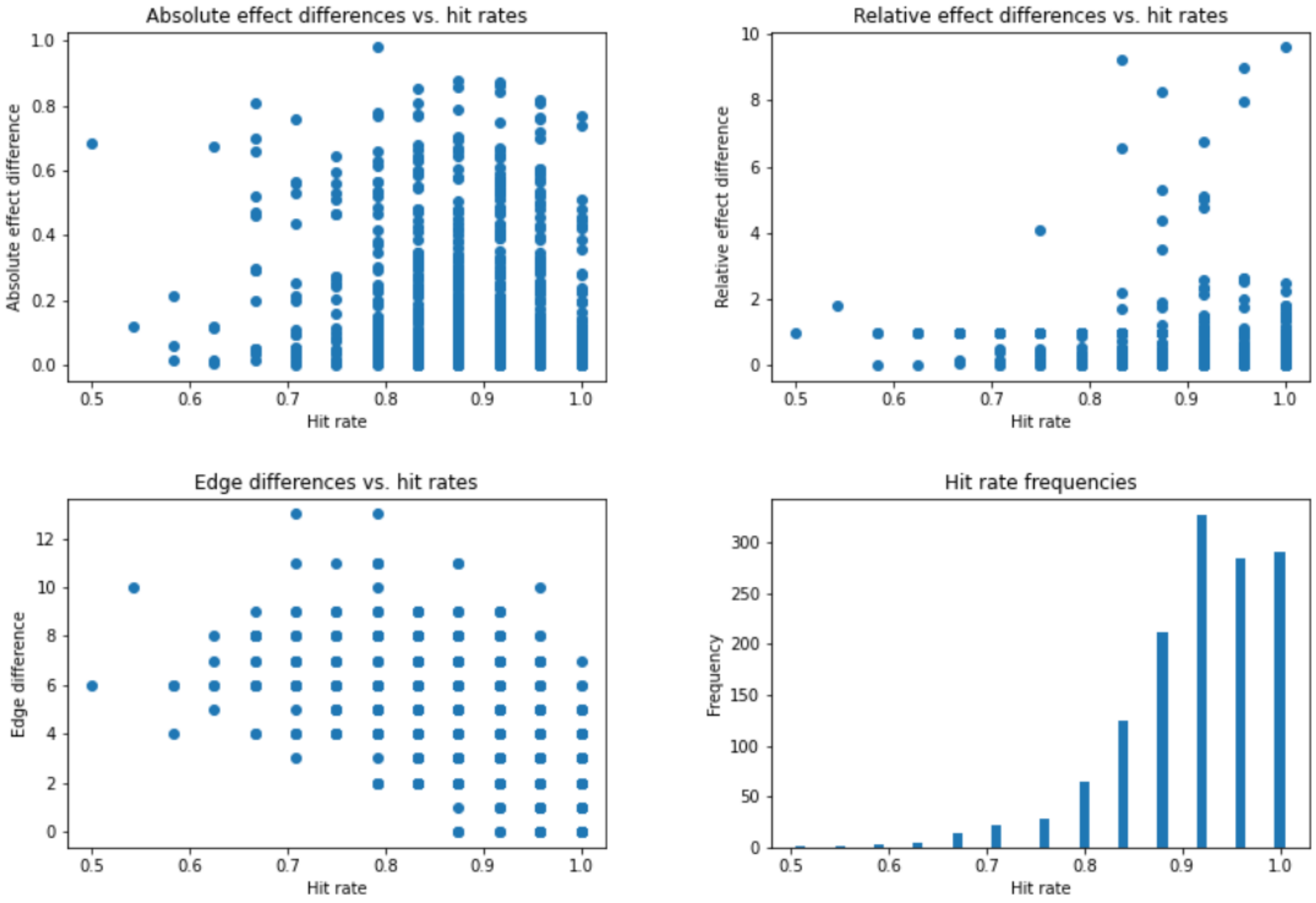
- Simulate 1000 samples from the true model
- Pass some edges of the true graph as domain knowledge
- Run causal discovery (greedy equivalence search) on domain knowledge and data
- Use resulting graph, do-calculus and linear regression to estimate target effect and quantitative probes
- Evaluate estimates of graph, target effect and quantitative probes (➔ hit rate)

Repeat many times.

Are models with a high hit rate better at recovering the causal graph and the target effect?

Simulation study

Evidence for the binary linear case (raw data)



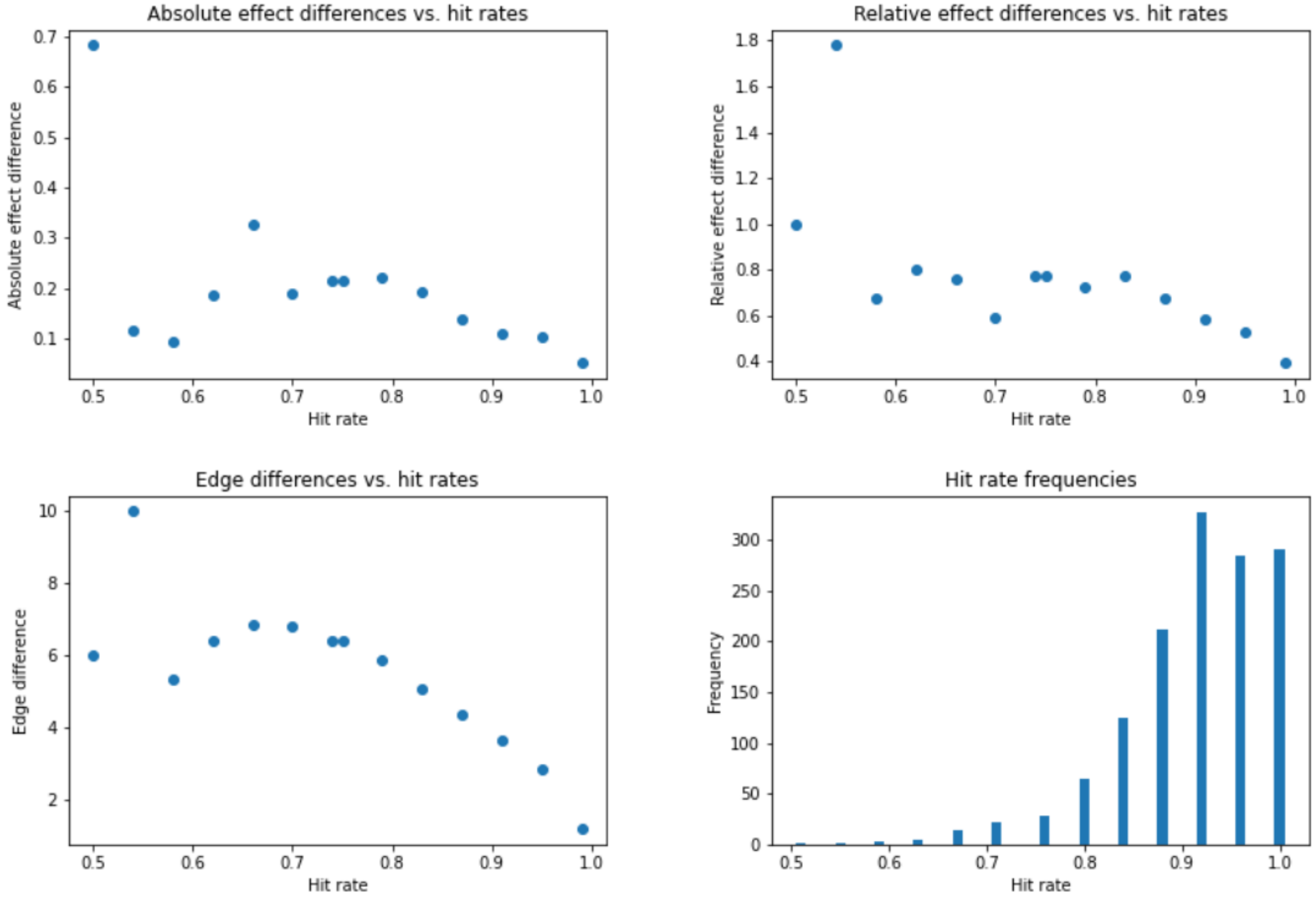
1. **most runs achieve high hit rate**
2. **hard to identify any relation between hit rate and success of the analysis**

Daniel Grünbaum, Maike Stern, Elmar W. Lang

ams OSRAM

Simulation study

Evidence for the binary linear case (aggregated data)



1. most runs achieve high hit rate
2. higher hit rate ➔ better estimation of target effect
3. higher hit rate ➔ better recovery of causal graph
4. linear association?

Outlier analysis

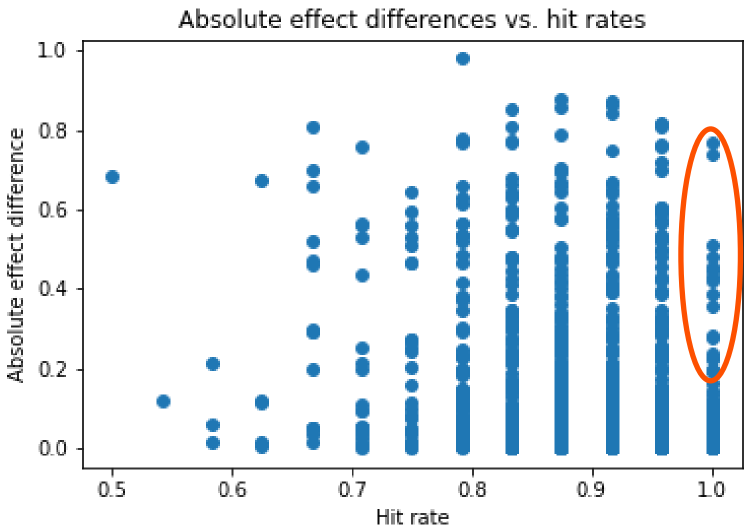
The ones who got away

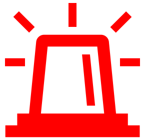
How to **identify weaknesses** of quantitative probing? Inspect outliers!

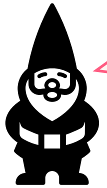
Outlier: run with high hit rate, but incorrect target effect estimation

14 out of 1378 runs where

- perfect hit rate is achieved
- the absolute error for the target effect is at least 0.2

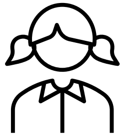






This model
can be trusted!

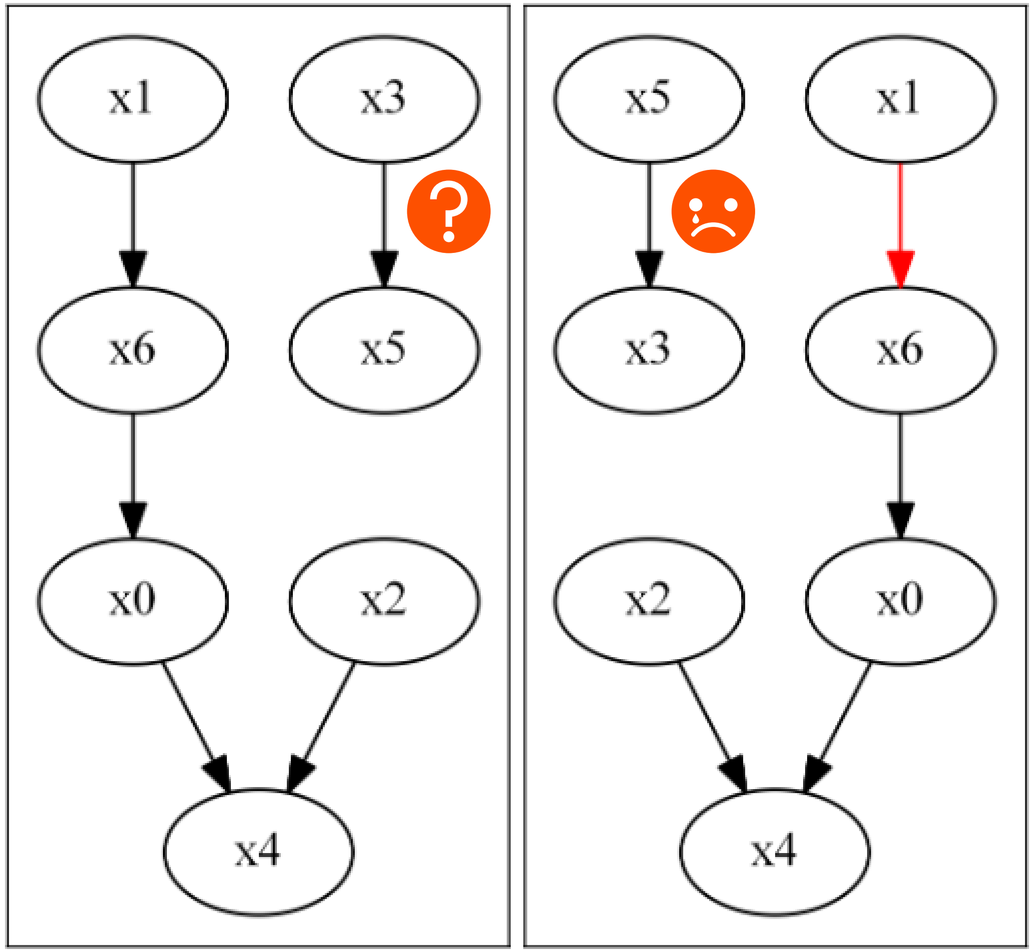
You‘re fired!



Outlier analysis

Connectivity issues

True graph **Discovered graph**



Target effect

Incorrect edge

Missing edge

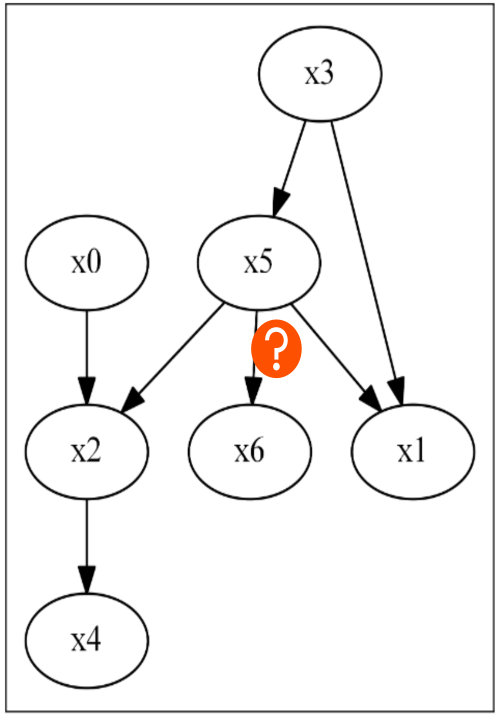
Domain knowledge

The quantitative probes are not in the same component as the target effect!

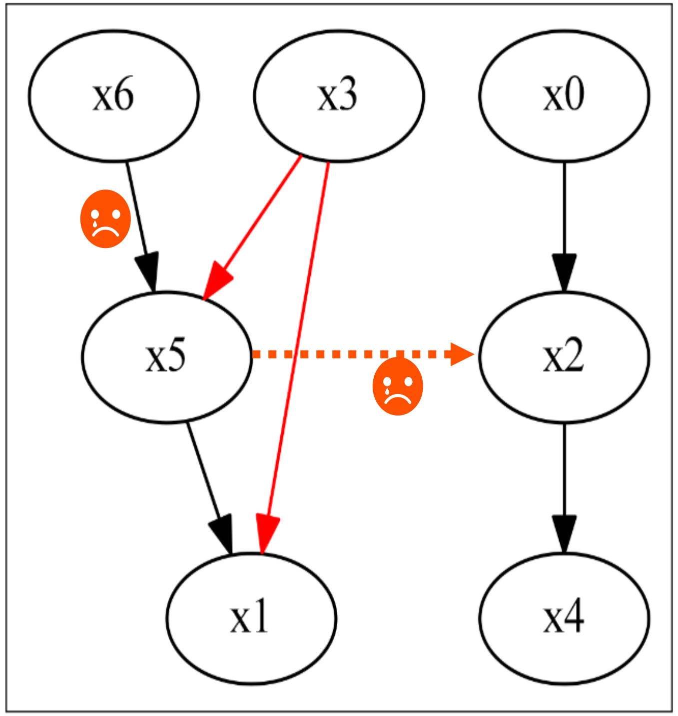
Outlier analysis

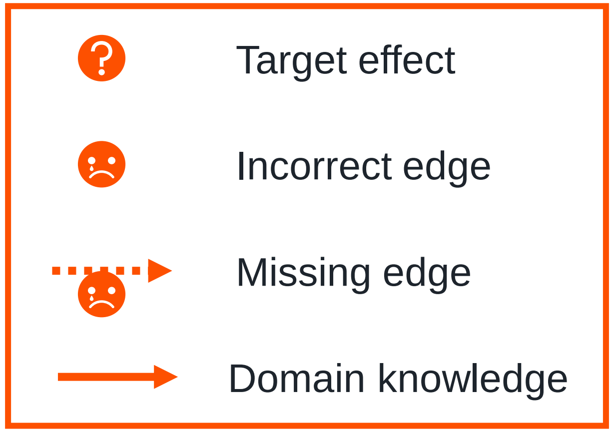
Probe coverage

True graph



Discovered graph





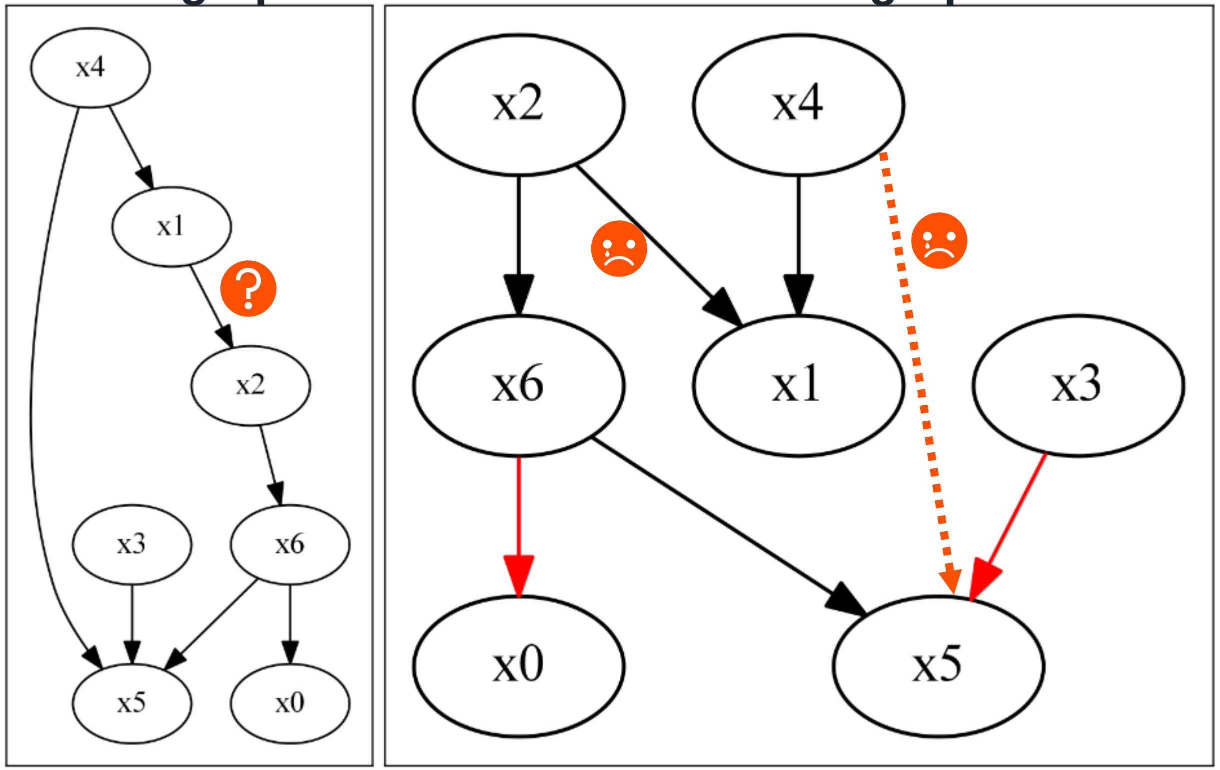
The effect of x3 on x6 has not been used as a quantitative probe!

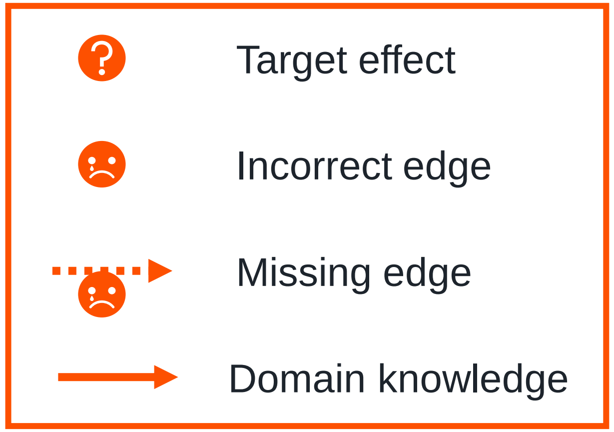
Daniel Grünbaum, Maike Stern, Elmar W. Lang

Outlier analysis

Probe tolerance

True graph **Discovered graph**





The tolerance for the effect of x4 on x2 (0.07 vs. 0) has been set too high (0.1)!

Open questions

Digging deeper

Points that require further investigation:

- **Theory**

  Is the relation between hit rate and performance really linear? Why?

- **Probe selection**

  Can we quantify the usefulness of probes based on certain properties? Are probes more effective if they are closer to the target effect?

- **Probe coverage**

  How many probes do we need to ensure a certain quality of a model with a perfect hit rate?

- **Probe tolerance**

  How does the success of quantitative probing depend on the chosen tolerances?

Your turn!

More quantitative probing

Full paper on arXiv: https://arxiv.org/abs/2209.03013

Quantitative probing: Validating causal models using quantitative domain knowledge

Daniel Grünbaum, Maike L. Stern, Elmar W. Lang

We present quantitative probing as a model-agnostic framework for validating causal models in the presence of quantitative domain knowledge. The method is constructed as an analogue of the train/test split in correlation-based machine learning and as an enhancement of current causal validation strategies that are consistent with the logic of scientific discovery. The effectiveness of the method is illustrated using Pearl's sprinkler example, before a thorough simulation-based investigation is conducted. Limits of the technique are identified by studying exemplary failing scenarios, which are furthermore used to propose a list of topics for future research and improvements of the presented version of quantitative probing. The code for integrating quantitative probing into causal analysis, as well as the code for the presented simulation-based studies of the effectiveness of quantitative probing is provided in two separate open-source Python packages.

Custom quantitative probing experiments in Python with **qprobing**: https://github.com/MLResearchAtOSRAM/qprobing

qprobing Public

The qprobing package provides functionality for evaluating the effectiveness of quantitative probing as a method for validating causal models. Developed by Daniel Grünbaum (@dg46).

Python 1

ams OSRAM

Summary

What to keep in mind about quantitative probing

Valuable quantitative domain knowledge can be harnessed

Successful probes increase our trust in the model, whereas failed probes can trigger necessary reevaluations of the modelling assumptions

Quantitative probing follows the logic of scientific discovery

Versatile and largely model-agnostic framework

Open-source Python package for custom experiments is available on GitHub

Your turn!

More causal inference

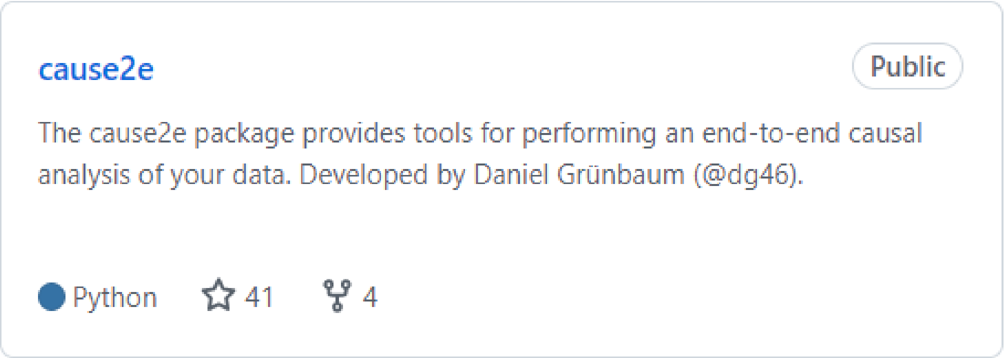
Causal Inference Working Group:

- Regular exchange for causality enthusiasts and curious newcomers
- About 40 active members from academia and industry
- Wiki for schedule and more info: https://gitlab.com/causal-inference/working-group/-/wikis
- Feel free to join or drop by for just one meeting!

Contact me to learn or discuss about causal inference!

daniel.gruenbaum@ams-osram.com

Causal end-to-end analysis in Python with **cause2e**:
https://github.com/MLResearchAtOSRAM/cause2e



Thank you! Questions?