Causality for Machine Learning II

Matej Zečević

7th Int'l Summer School on Data Science (SSDS 2022),
Day 4, 13th October 2022

Overview of Today's Programme:

- ❑ ~~**Done**: *Causality for Machine Learning 1*, where we explore the basic concepts from causality that are being used in current ML research~~
  ~~9 am – 10:30 am (including Q&A)~~

- ❑ **Now**: *Causality for Machine Learning 2*, where we discuss some selected publications and on-going work on the interface of causality and ML
  10:45 am – 12:15 am (including Q&A)

- ❑ In the afternoon: *Hands-on session: Causality for Machine Learning,* where you get to experience some code for using causal ML first hand
  1:30 pm – 3 pm

Many works in Causality x AI/ML check out the sessions at CDG for example:

Causality Discussion Group

Join the **Session** (Zoom)

Join the **Community** (Slack)

Join the **Mailing List** (G-Groups)

Towards Causal Representation Learning

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, Yoshua Bengio

Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style

Julius von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, Francesco Locatello

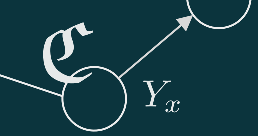
Weakly supervised causal representation learning

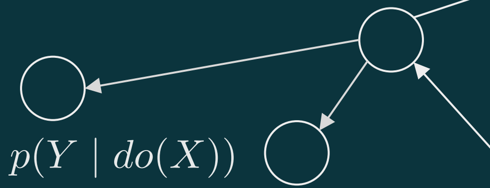
Johann Brehmer, Pim de Haan, Phillip Lippe, Taco Cohen

...

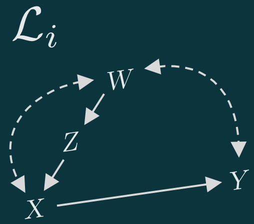
The session recordings are freely available,
and the authors discuss their paper best naturally.

Therefore, this session will focus on several works from **our** lab





I | XAI meets Causality



TECHNISCHE UNIVERSITÄT DARMSTADT

XAI Establishes a Common Ground Between Machine Learning and Causality

old name: Structural Causal Interpretation Theorem

Matej Zečević, Devendra Singh Dhami, Constantin Rothkopf, Kristian Kersting

arxiv: 2110.02395

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Hypothesis 1 (HMM Conversion, short HMMC). *The parts of the HMM that are being used for encoding the causal relationships of the variables of interest can be formally captured by a corresponding SCM (def. in Sec.2), in short this "equivalence" can be denoted as HMM ≡ SCM.*

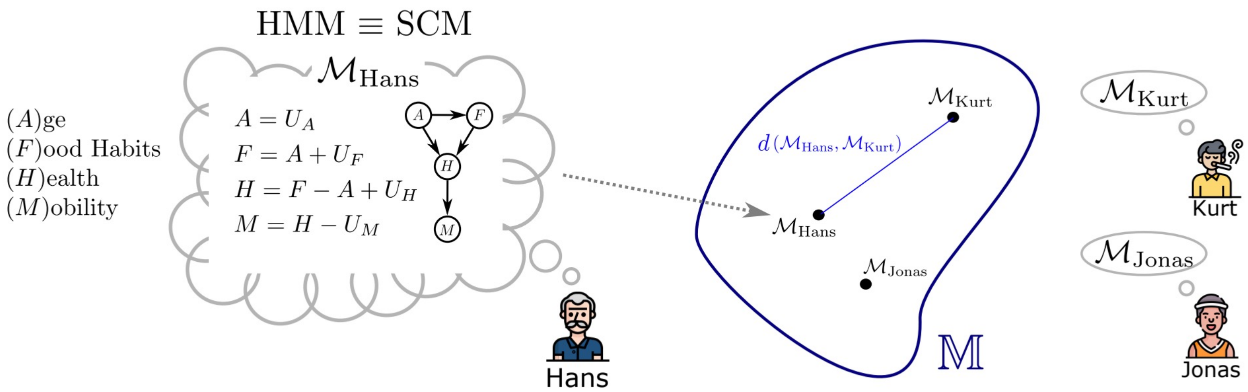


Figure 1: **HMMC Hypothesis and Linear SCM Metric Space.** Left: Accepting Hyp.1 means that the HMM of Hans is an SCM. Right: Different linear SCM (from different individuals) can be compared, an example metric space for $(\mathbb{M}, d)$ is given by Prop.1. (Best viewed in color.)

TECHNISCHE UNIVERSITÄT DARMSTADT

SCM $\mathcal{M}$

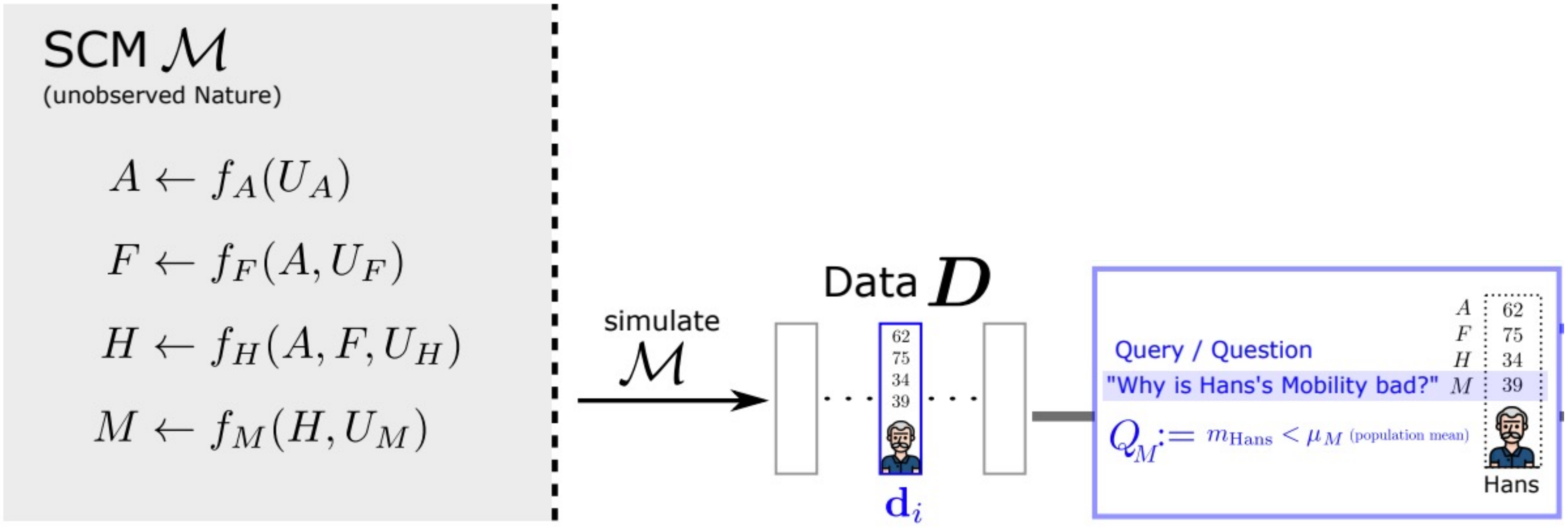
(unobserved Nature)

$$A \leftarrow f_A(U_A)$$

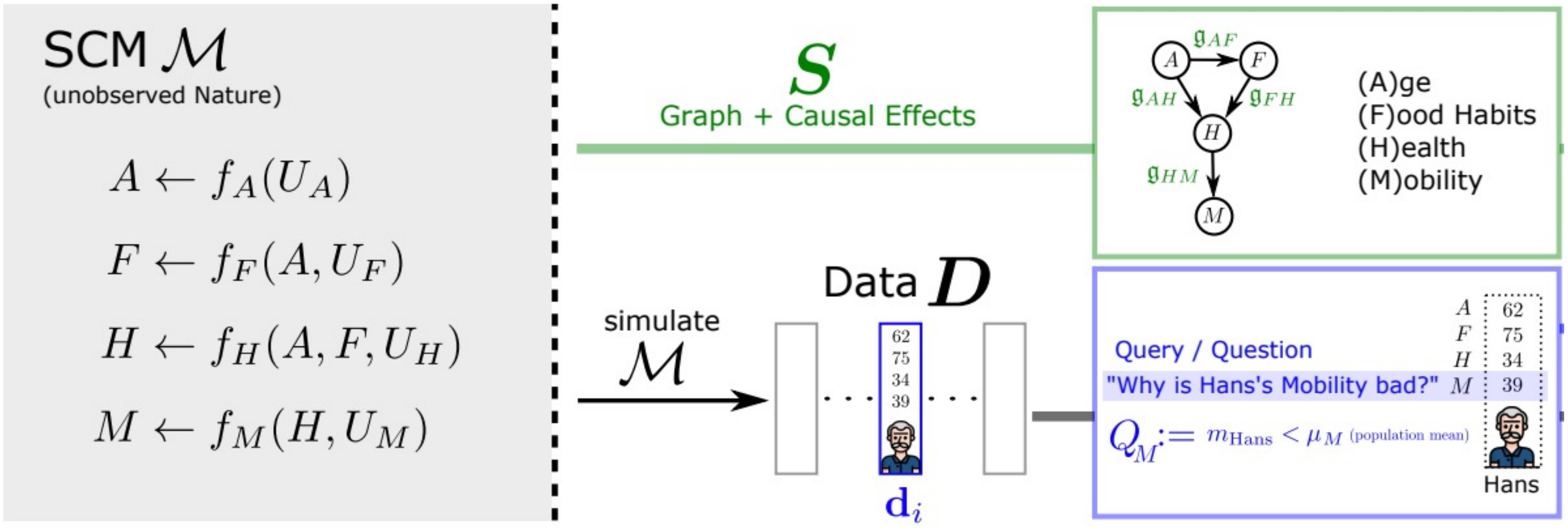
$$F \leftarrow f_F(A, U_F)$$

$$H \leftarrow f_H(A, F, U_H)$$

$$M \leftarrow f_M(H, U_M)$$



TECHNISCHE UNIVERSITÄT DARMSTADT


SCM $\mathcal{M}$
(unobserved Nature)
$A \leftarrow f_A(U_A)$
$F \leftarrow f_F(A, U_F)$
$H \leftarrow f_H(A, F, U_H)$
$M \leftarrow f_M(H, U_M)$
S
Graph + Causal Effects
$\mathfrak{g}_{AF}$
$\mathfrak{g}_{AH}$
$\mathfrak{g}_{FH}$
$\mathfrak{g}_{HM}$
A
F
H
M
(A)ge
(F)ood Habits
(H)ealth
(M)obility
simulate $\mathcal{M}$
Data D
62
75
34
39
$\mathbf{d}_i$
Query / Question
"Why is Hans's Mobility bad?"
$Q_M := m_{\text{Hans}} < \mu_M$ (population mean)
A 62
F 75
H 34
M 39
Hans

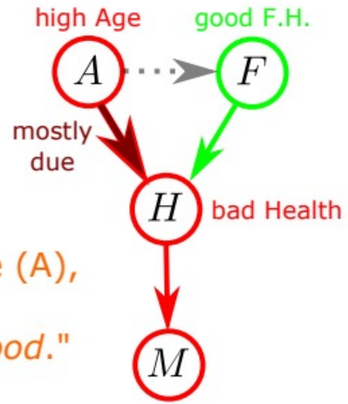
$$\mathfrak{I}(\boldsymbol{Q}_M, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D})$$

"Hans' Mobility (M) is bad

because of his *bad* Health (H)

which is *mostly* due to his *high* Age (A),

although his Food Habits (F) are *good*."



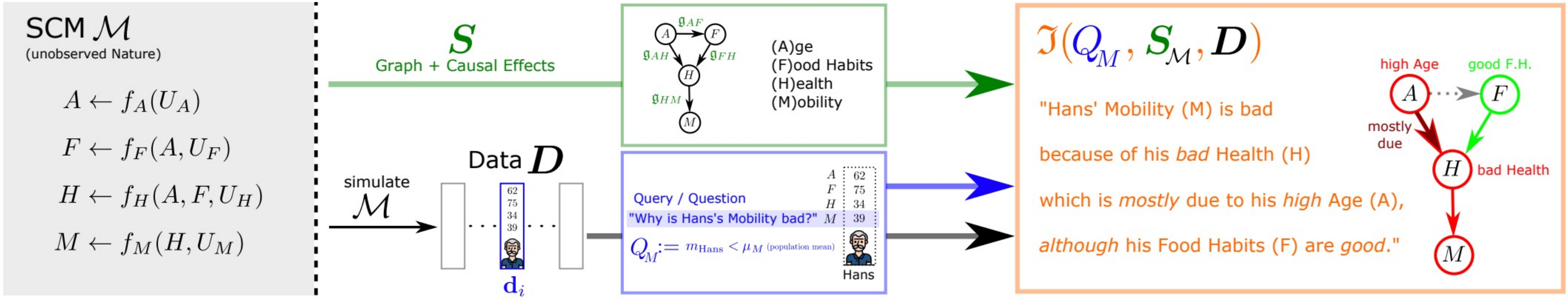


Figure 2: **SCI for the "Causal Hans" Example.** Left: the (unobserved) SCM $\mathcal{M}$. Middle: The why-question Q_M is an individual-level query derived from some population $\boldsymbol{D}$ originating in $\mathcal{M}$. The graph weighted by causal effect of each causal pair of $\mathcal{M}$ is called $\boldsymbol{S}_{\mathcal{M}}$. Right: the Structural Causal Interpretation $\mathfrak{I}(Q_M, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D})$. (Best viewed in color.)

TECHNISCHE UNIVERSITÄT DARMSTADT

Definition 5 (FOL Interpretation Rules). *Let C_{YX} denote a causal scenario (see Def.4), let $s(x) \in \{-1, 1\}$ be the sign of a scalar, let $R_i \in \{<, >\}$ be a binary ordering relation and let $\mathcal{Z}_X = \{|\mathfrak{g}(Z \to X)| : Z \in \mathrm{Pa}_X\}$ be the set of absolute parental causal effects onto X. We define FOL-based rule functions $\mathcal{R}_i(C_{YX}, R_1, R_2, s, \mathcal{Z}_X) \in \{-1, 0, 1\}$,*

($\mathcal{R}_1$) *Excitation. If $R_1 \neq R_2$, then:* $\quad R_1(s(\mathfrak{g}(Y \to X)), 0) \land (R_2(y, \mu_Y) \lor R_1(x, \mu_X))$,

($\mathcal{R}_2$) *Inhibition. If $R_1 \neq R_2$, then:* $\quad R_1(s(\mathfrak{g}(Y \to X)), 0) \land R_1(y, \mu_Y) \land R_2(x, \mu_X)$,

($\mathcal{R}_3$) *Preference. If $|\mathcal{Z}_X| > 1$, then* $\quad Y \iff \arg\max_{Z \in \mathcal{Z}_X} Z$,

indicating for each rule $\mathcal{R}_i$ how the causal relation $Y \to X$ satisfies that rule (e.g. for $\mathcal{R}_1$ either Over- or Under-Excitation when $\mathcal{R}_1 \neq 0$ and No Excitation if $\mathcal{R}_1 = 0$).

Definition 6 (SCI). *Let $Q_X, \boldsymbol{S}_\mathcal{M}$ like before be a why-question and the causal effect weighted graph for SCM $\mathcal{M}$, and let $\boldsymbol{D} \in \mathbb{R}^{n \times |\mathbf{V}|}$ denote our data set. We define a recursion*

$$\mathfrak{I}(Q_X, \boldsymbol{S}_\mathcal{M}, \boldsymbol{D}) = \left(\bigoplus_{Z \in \mathrm{Pa}(X)} \mathcal{R}(Z \to X), \bigoplus_{Z \in \mathrm{Pa}(X)} \mathfrak{I}(Q_Z, \boldsymbol{S}_\mathcal{M}, \boldsymbol{D}) \right) \tag{1}$$

where $\bigoplus_{i=1}^n v_i = (v_1, \ldots, v_n)$ denotes concatenation and in slight abuse of notation $\mathcal{R}$ checks each rule $\mathcal{R}_i$ from Def.5, and the recursion's base case is being evaluated at the roots of the causal path to X, that is, for some $Z \in \mathbf{V}$ with a path $Z \to \cdots \to X$ we have

$$\mathfrak{I}(Q_Z, \boldsymbol{S}_\mathcal{M}, \boldsymbol{D}) = \emptyset. \tag{2}$$

We call $\mathfrak{I}(Q_X, \boldsymbol{S}_\mathcal{M}, \boldsymbol{D})$ Structural Causal Interpretation.

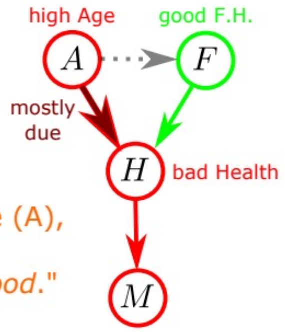
$\mathfrak{I}(Q_M, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D})$

"Hans' Mobility (M) is bad

because of his *bad* Health (H)

which is *mostly* due to his *high* Age (A),

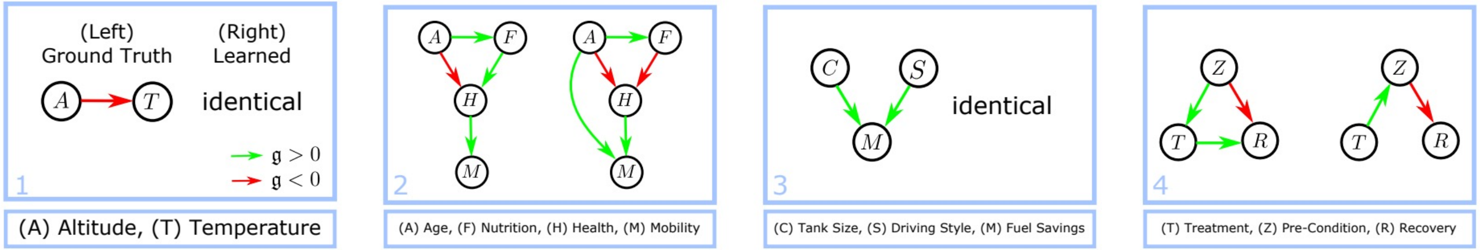
although his Food Habits (F) are *good*."



$$
\begin{aligned}
&\mathfrak{I}(Q_M, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D}) \\
&= ((\mathcal{R}_1 = -1), \textstyle\bigoplus_{Z \in \{A,F\}} \mathfrak{I}(Q_H, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D})) \\
&= ((\mathcal{R}_1 = -1), (((\mathcal{R}_1 = 1, \mathcal{R}_3 = 1), \mathfrak{I}(Q_A, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D})), ((\mathcal{R}_2 = 1), \mathfrak{I}(Q_F, \boldsymbol{S}_{\mathcal{M}}, \boldsymbol{D}))), \\
&= ((\mathcal{R}_1 = -1), (((\mathcal{R}_1 = 1, \mathcal{R}_3 = 1), \emptyset), ((\mathcal{R}_2 = 1), \emptyset).
\end{aligned}
$$

Theorem 1. *Any GIM is interpretable in the SCI-sense.*

TECHNISCHE UNIVERSITÄT DARMSTADT



1	*"Why is the Temperature at the Matterhorn low?"*	(*Question*)
	"The Temperature at the Matterhorn is low because of the high Altitude."	(Ground Truth)
	"The Temperature at the Matterhorn is low because of the high Altitude."	**(Learned)**
2	*"Why is Hans's Mobility bad?"*	
	"Hans's Mobility is bad because of his bad Health which is mostly due to his high Age, although his Food Habits are good."	
	"Hans's Mobility, in spite his high Age, is bad mostly because of his bad Health which is bad mostly due to his good Food Habits."	
3	*"Why is your personal car's left Mileage low?"*	
	"Your left Mileage is low because of your small Car and your bad Driving Style."	
	"Your left Mileage is low because of your small Car and your bad Driving Style."	
4	*"Why did Kurt not Recover?"*	
	"Kurt did not Recover because of his bad Pre-condition, although he got Treatment."	
	"Kurt did not Recover because of his bad Pre-Condition, which were bad although he got Treatment."	

Table 1: **Quality of Learned Interpretations.** We chose the simple, popular NOTEARS from Zheng et al. (2018) as our GIM for generating the SCI. Subtle differences between interpretations exist e.g., the interpretation 4 is right on the top-level but for the wrong reasons, that is $T \rightarrow Z$ instead of $T \rightarrow R$. Variable letters are capitalized. (Best viewed in color.)

TECHNISCHE UNIVERSITÄT DARMSTADT

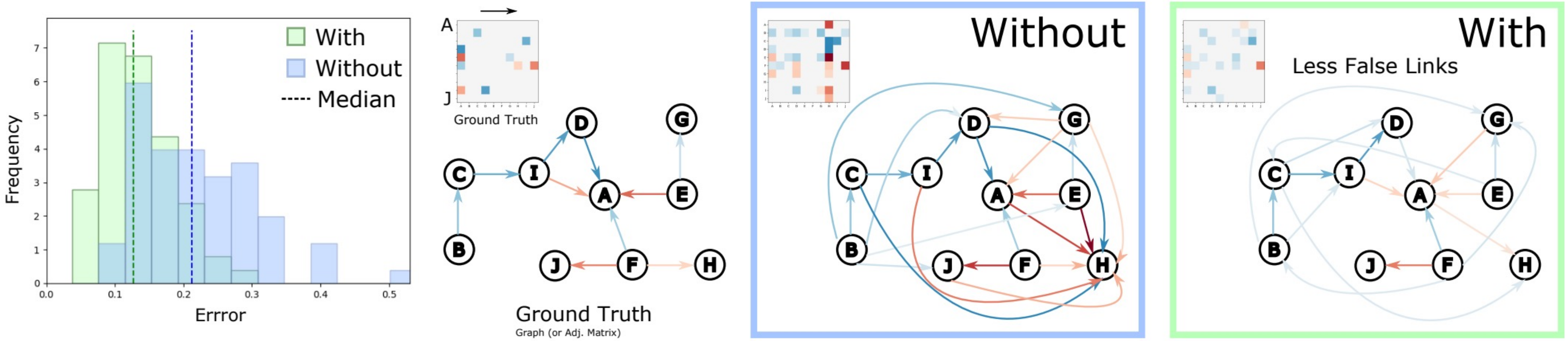


Figure 3: **Quality of Learned Graphs with SCI Regularization.** Left: error distributions when performing graph learning with/-out SCI regularization (which is simply an added penalty term for inconsistent SCIs), next to is the ground truth graph. Right (boxes): the predicted graphs, showing a decreased number of false positives. (Best viewed in color.)

TECHNISCHE
UNIVERSITÄT
DARMSTADT

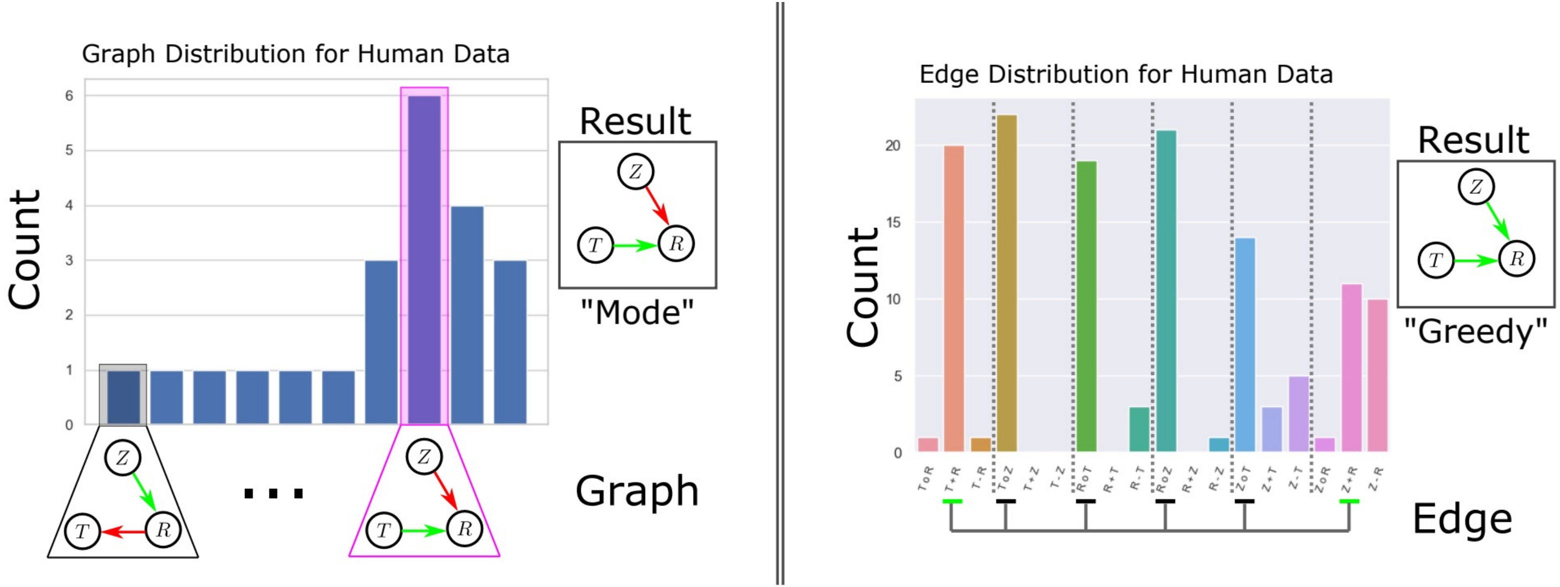
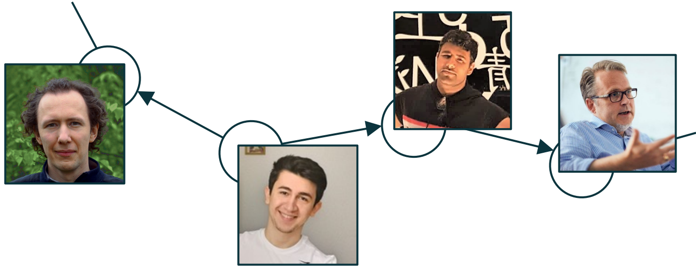


Figure 4: **Human Study.** $N = 22$ subjects data visualized as distributions over graphs (left) and edges (right). We present two ways of *aggregating* the data. Top, the final graph is the mode of the distribution. Bottom, look at each possible pair and choose each the edge type greedily from the distribution. (Best viewed in color.)

Humans: *"Hans's Mobility is bad because of his bad Health which is mostly due to his high Age, although his Food Habits are good."*

Machines: *"Hans's Mobility, in spite his high Age, is bad mostly because of his bad Health which is bad mostly due to his good Food Habits."*

2 | Large Language Models and Causality



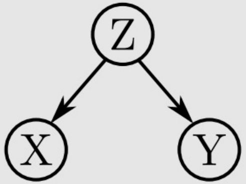
Can Foundation Models Talk Causality?

Moritz Willig*, Matej Zečević*, Devendra Singh Dhami, Kristian Kersting

UAI CRL Workshop 2022

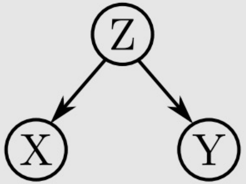
TECHNISCHE UNIVERSITÄT DARMSTADT

Causal Assumptions



“Z is common cause of X and Y”
“X and Y are causally unrelated”

Causal Assumptions

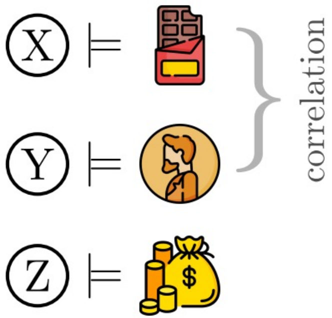


"Z is common cause of X and Y"
"X and Y are causally unrelated"

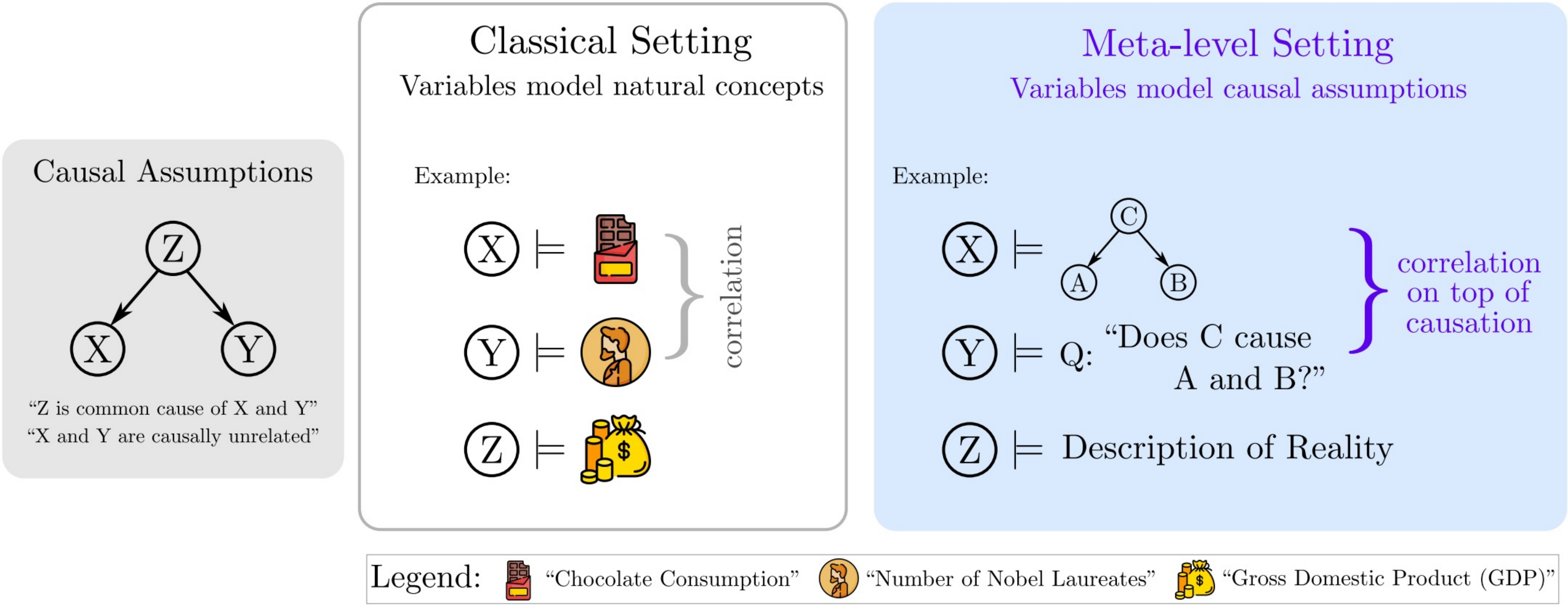
Classical Setting

Variables model natural concepts

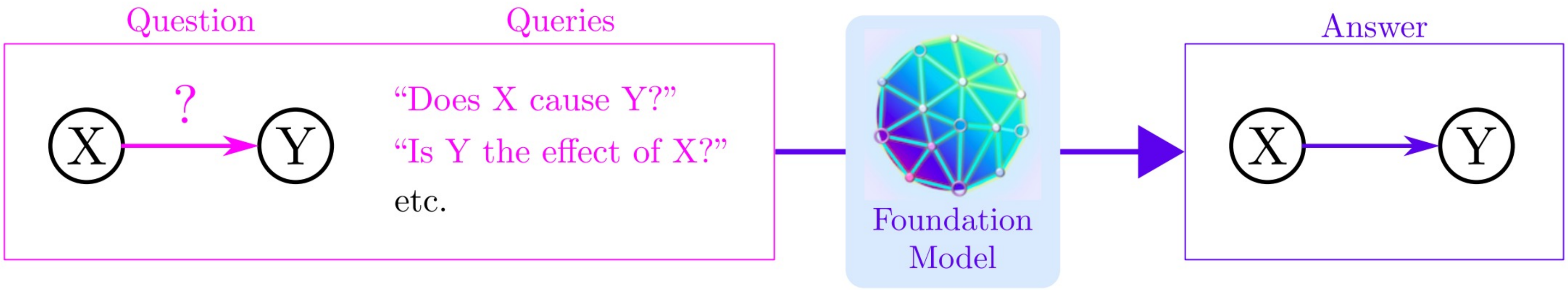
Example:



Legend: "Chocolate Consumption" "Number of Nobel Laureates" "Gross Domestic Product (GDP)"



TECHNISCHE UNIVERSITÄT DARMSTADT



TECHNISCHE
UNIVERSITÄT
DARMSTADT

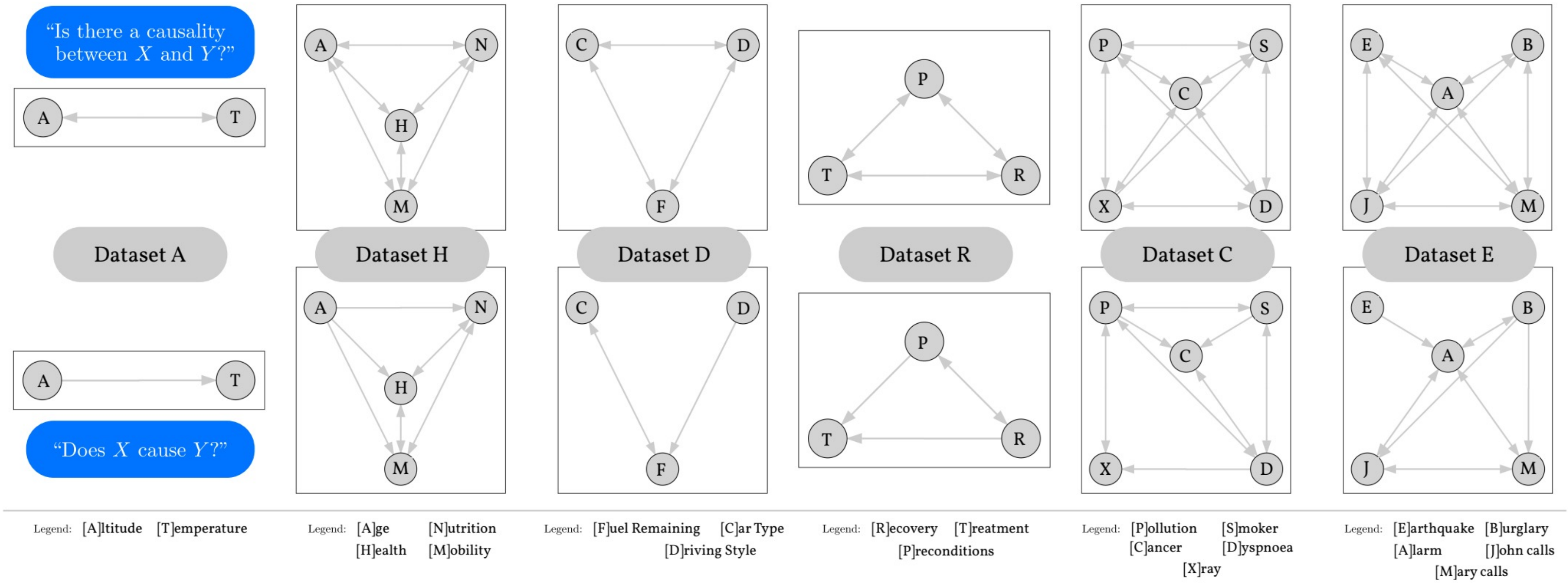
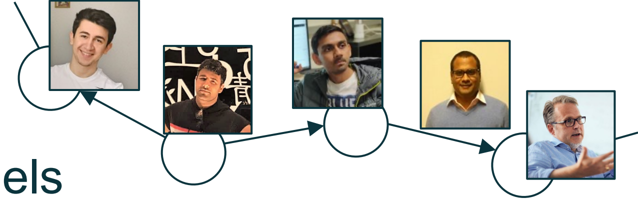


Figure 3: **Asymmetric Query Wording Implies Unidirectedness.** Language FM naive graph predictions on data sets that provide a causal graph (FM-O is shown). Top row, predictions with a symmetric query wording, bottom row, predictions with an asymmetric query wording. Surprisingly, the FM is capable of deciding multiple edges uniquely (and correctly) when switching to the asymmetric formulation without explicit guarantees to such behavior.

TECHNISCHE UNIVERSITÄT DARMSTADT

3 | SCMs and Other Model Families

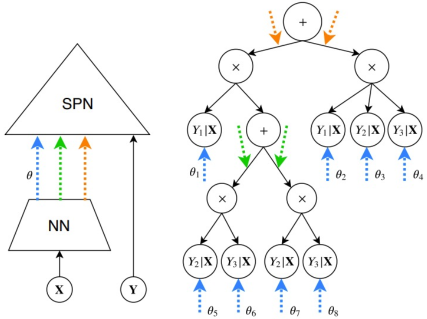


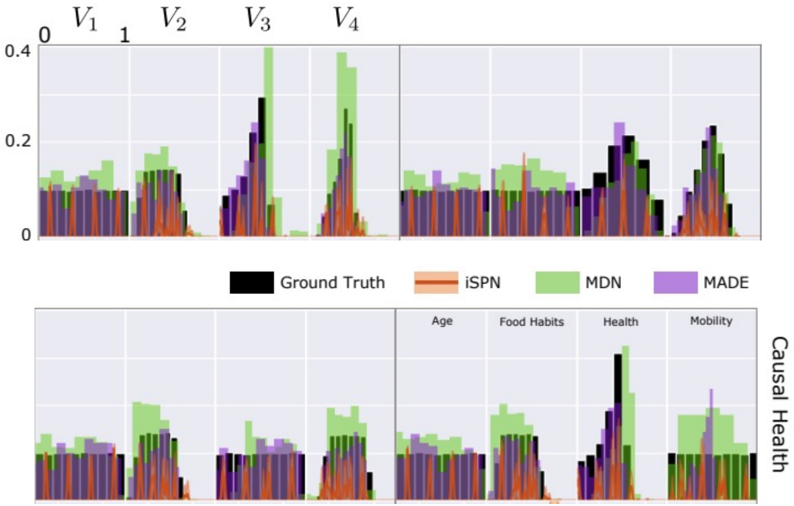
Interventional Sum-Product Networks:
Causal Inference with Tractable Probabilistic Models

Matej Zečević, Devendra Singh Dhami, Athresh Karanam, Sriraam Natarajan, Kristian Kersting

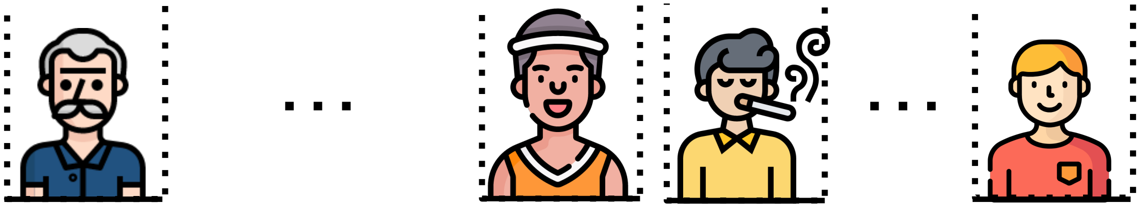
NeurIPS 2021

Definition 1 (Interventional Sum-Product Network). *An interventional sum-product network (iSPN) is the joint model $m(\mathbf{G}, \mathbf{D}) = g(\mathbf{D}; \boldsymbol{\psi} = f(\mathbf{G}; \boldsymbol{\theta}))$, where $g(\cdot)$ is a SPN, $f(\cdot)$ a non-parametric function approximator and $\boldsymbol{\psi} = f(\mathbf{G})$ are shared parameters.*



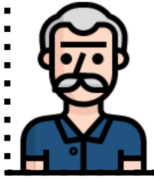


Individuals



Individuals and their Descriptions

$$
\begin{array}{c}
A \\ F \\ H \\ M
\end{array}
\begin{pmatrix}
62 & \dots & \dots & 18 & 24 & \dots & 21 & \dots \\
48 & \dots & \dots & 60 & 20 & \dots & 32 & \dots \\
34 & \dots & \dots & 90 & 40 & \dots & 64 & \dots \\
39 & \dots & \dots & 75 & 37 & \dots & 66 & \dots
\end{pmatrix}
$$



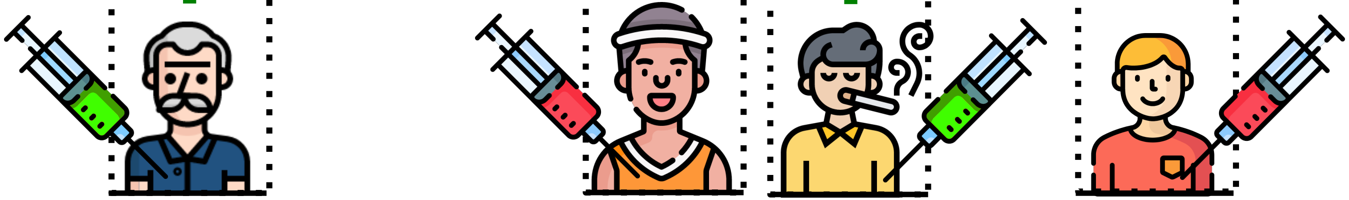
$A =$Age $\quad F =$Food Habits (or Nutrition) $\quad H =$Health $\quad M =$Mobility

TECHNISCHE UNIVERSITÄT DARMSTADT

THE UNIVERSITY OF TEXAS AT DALLAS

Interventions (e.g. a Vaccine)

$$
\begin{array}{c}
A \\ F \\ \color{red}{H} \\ M
\end{array}
\begin{pmatrix}
62 & \cdots & \cdots & 18 & 24 & \cdots & 21 & \cdots \\
48 & \cdots & \cdots & 60 & 20 & \cdots & 32 & \cdots \\
+ & \cdots & \cdots & - & + & \cdots & - & \cdots \\
+ & \cdots & \cdots & - & + & \cdots & - & \cdots
\end{pmatrix}
$$



A =Age F =Food Habits (or Nutrition) H =Health M =Mobility

Hidden Reality: SCM

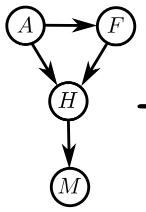
Structural Causal Model

$$A = U(0,100)$$
$$F = \frac{1}{2}A + \mathcal{N}(10,10)$$
$$H = \frac{1}{100}(100 - A^2) + \frac{1}{2}F + \mathcal{N}(40,30)$$
$$M = \frac{1}{2}H + \mathcal{N}(20,10)$$

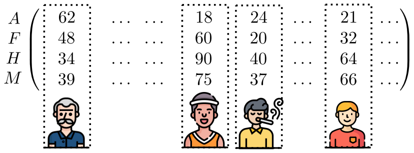
(A)ge
(F)ood Habits
(H)ealth
(M)obility

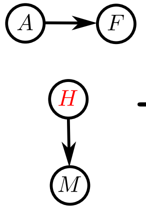
$$A = U(0,100)$$
$$F = \frac{1}{2}A + \mathcal{N}(10,10)$$
$$H = U(0,100)$$
$$M = \frac{1}{2}H + \mathcal{N}(20,10)$$

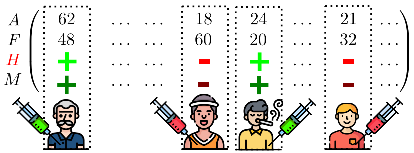
G



D







A =Age $\quad$ F =Food Habits (or Nutrition) $\quad$ H =Health $\quad$ M =Mobility

TECHNISCHE UNIVERSITÄT DARMSTADT

THE UNIVERSITY OF TEXAS AT DALLAS

We ask…

…can we perform *tractable* causal inference?
…can we find a purely computational view on causality?

TECHNISCHE UNIVERSITÄT DARMSTADT
THE UNIVERSITY OF TEXAS AT DALLAS

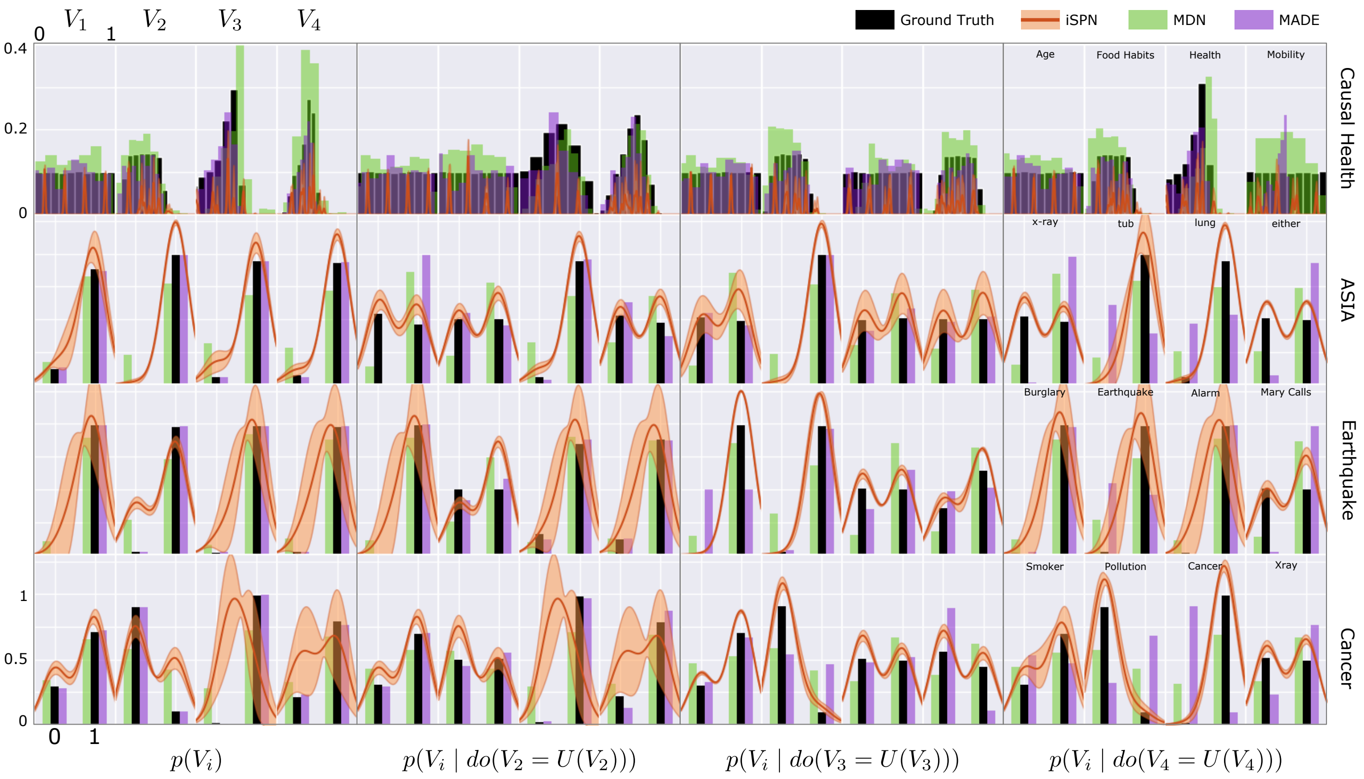
The BN-SPN Relation

- Zhao et al., 2015 presented a compilation method
  between BN and SPN
  that resulted in **degenerate** BNs incapable of causal inference
  as pointed out by Papantonis & Belle 2020.

- We consider *extension* of the SPN model class.
  We reason that the **over-parameterization** framework can be used
  to achieve tractable causal inference with SPN.

  The natural candidate for extension are **Gated** (or **Conditional**) **SPN**
  as presented by Shao et al., 2019.

QI and Q3



Q1: How is the estimation quality for interventional distributions?

Q3: How is the performance relative to generative models?

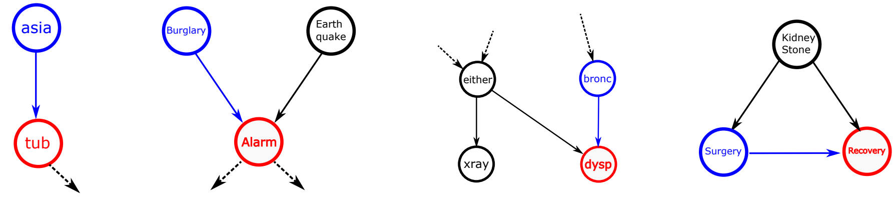
TECHNISCHE UNIVERSITÄT DARMSTADT

THE UNIVERSITY OF TEXAS AT DALLAS

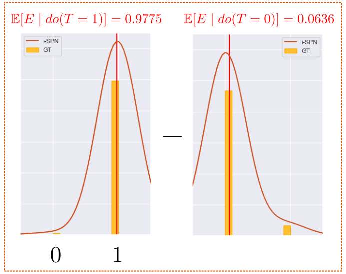
Q4

Confounding	Conditioning	Ground Truth[1]	CausalML	DoWhy	iSPN	
No	0.0374	0.0397 (0.04)	0.0397	0.0397	**0.0347** ✓	ATE(asia, tub)
No	0.9271	0.9337 (0.93342)*	0.9337	0.9337	‡ **0.9139** ✓	ATE(Burglary, Alarm)
Yes	0.6766	0.6703 (0.667586)	0.6703	0.6697	**0.6551** ✓	ATE(bronc, dysp)
Yes	−0.0457	0.0537 (0.05)	−0.0454	0.0538	**0.0545** ✓	ATE(Surgery, Recovery)

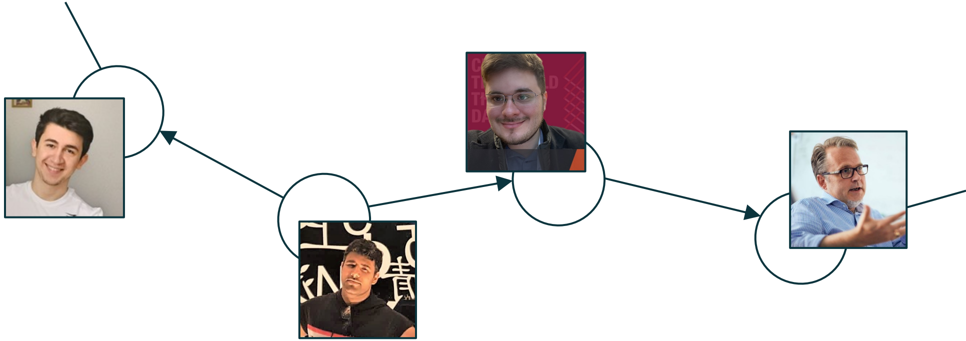
$$\text{ATE}(T, E) := \mathbb{E}[E \mid do(T = 1)] - \mathbb{E}[E \mid do(T = 0)]$$



‡ Visual



Q4: How is the performance relative to causal models?



Relating Graph Neural Networks to Structural Causal Models

Matej Zečević, Devendra Singh Dhami, Petar Veličković, Kristian Kersting

arxiv: 2109.04173

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Variational Inference 101

- Assuming the existence of unobserved but relevant variables $\mathbf{Z}$
  to jointly model the phenomenon with observations $p(\mathbf{X}, \mathbf{Z})$
  we resort to **optimization** to get the *posterior* $p(\mathbf{Z}|\mathbf{X})$

$$q^*(Z) = \arg\min_{q \in \mathcal{Q}} \mathrm{KL}(q(\mathbf{Z}) \,||\, p(\mathbf{Z} \mid \mathbf{X}))$$

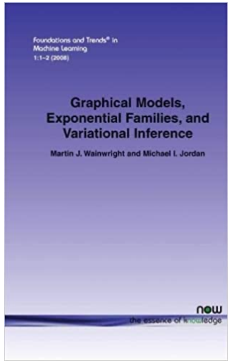
  where $\mathcal{Q}$ denotes a selected variational family of distributions (e.g. MoG).

- Intractability of the evidence $p(\mathbf{X}) = \int p(\mathbf{X}, \mathbf{Z})\, d\mathbf{Z}$
  leads to a **lower bound** (ELBO)

$$\log p(\mathbf{X}) - \mathrm{KL}(q(\mathbf{Z}) \,||\, p(\mathbf{Z} \mid \mathbf{X})) =$$
$$\mathbb{E}_q[\log p(\mathbf{X} \mid \mathbf{Z})] - \mathrm{KL}(q(\mathbf{Z}) \,||\, p(\mathbf{Z}))$$

  with Variational Auto-Encoder (VAE) being a neural variant.

Pointers to Variational Inference References



- Wainwright & Jordan, “**Graphical Models, Exponential Families, and Variational Inference**”, 2008.
- David Blei, “**Variational Inference: A Review for Statisticians”,** JASA 2017.
- Philipp Hennig Lecture, “**Variational Inference**”, 2021.

  `https://www.youtube.com/watch?v=TcgeofQJYyM`
- Jaan Altosaar Blog, “**Tutorial - What is a variational autoencoder?**”.

  `https://jaan.io/what-is-variational-autoencoder-vae-tutorial/`

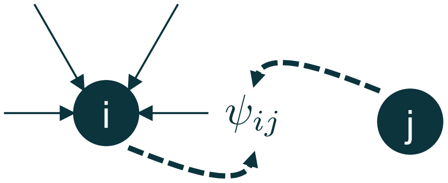
Graph Neural Networks 101

- Graph Neural Networks (GNN) place an **inductive bias on graphs**.
  Permutation-equivariant application $f(\mathbf{D}, \mathbf{A}_G)$
  of permutation-invariant functions $g(\mathbf{d}_i, \mathbf{D}_{\mathcal{N}_i^G})$.

- Three different flavors: convolution $\subseteq$ attention $\subseteq$ message-passing.
  The most general formulation in terms of messages is given by:

$$\mathbf{h}_i = \phi\left(\mathbf{d}_i, \bigoplus_{j \in \mathcal{N}_i^G} \psi(\mathbf{d}_i, \mathbf{d}_j)\right)$$

$G \supseteq \mathcal{N}^G$



Pointers to GNN References

- Petar Veličković Lecture, “**Theoretical Foundations of GNN**”, Cambridge, 2021.

  https://www.youtube.com/watch?v=uF53xsT7mjc

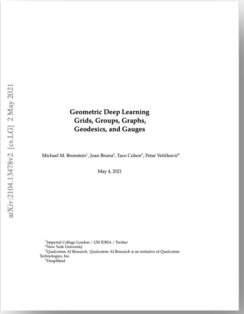
- Bronstein et al., “**Geometric Deep Learning**”, arXiv, 2021.

  Bronstein Lecture, “**GDL: The Erlangen Programme of ML**”, ICLR, 2021.

  https://www.youtube.com/watch?v=w6Pw4MOzMuo

- Nikolas Adaloglou Tutorial, “**How GNN work: […]**”, 2021.

  https://theaisummer.com/graph-convolutional-networks/


arXiv:2104.13478v2 [cs.LG] 2 May 2021

**Geometric Deep Learning
Grids, Groups, Graphs,
Geodesics, and Gauges**

Michael M. Bronstein[1], Joan Bruna[2], Taco Cohen[3], Petar Veličković[4]

May 4, 2021

[1]Imperial College London / USI IDSIA / Twitter
[2]New York University
[3]Qualcomm AI Research. Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.
[4]DeepMind

From First Principles

- Peter Holland & Don Rubin (1986):

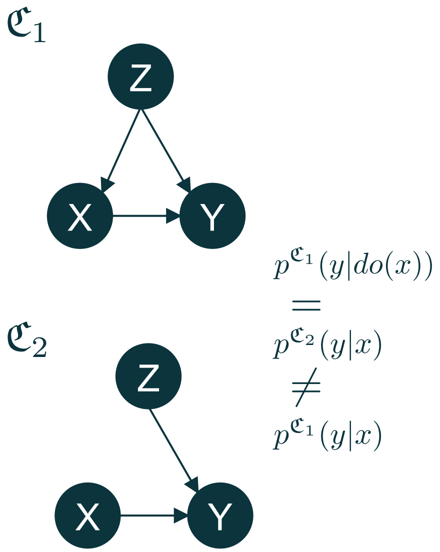
  "No causation without manipulation"

  Not true, as **identifiability** suggests (*do*-calculus etc.):

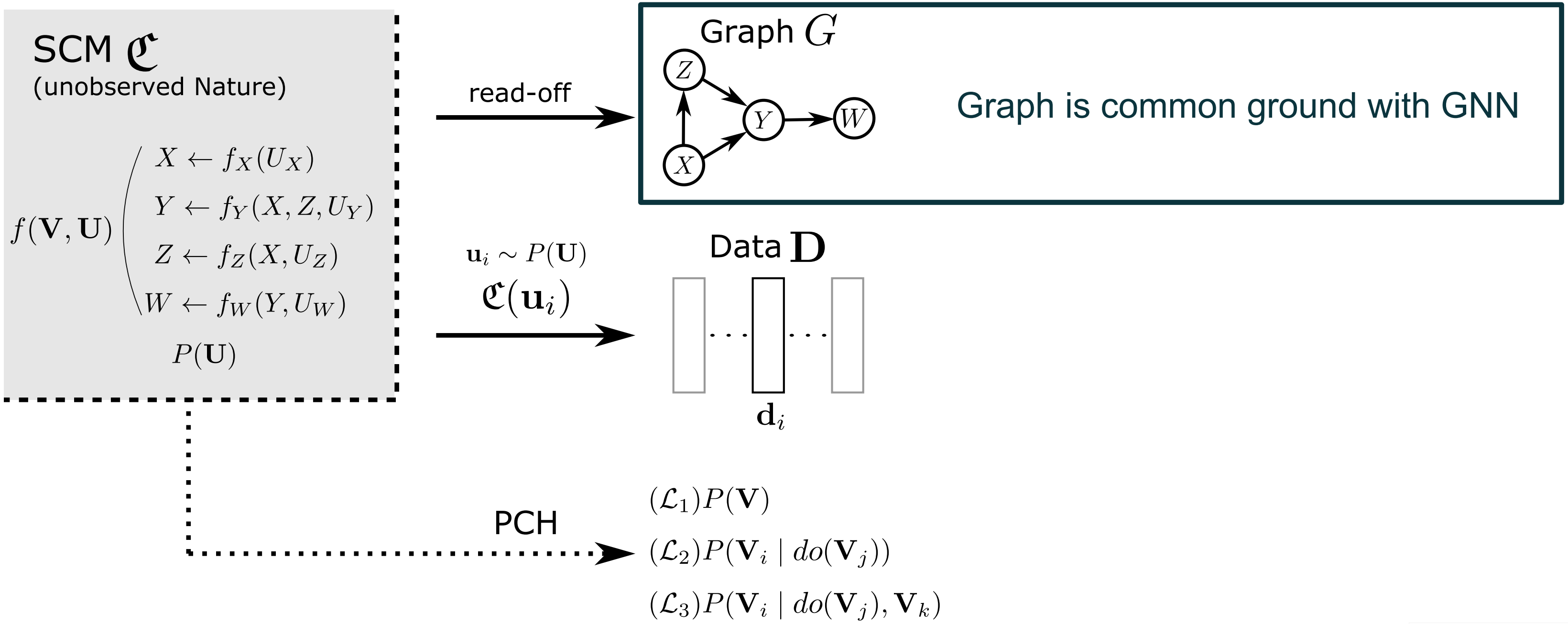
  $$p(y|do(x)) = \sum_z p(y|x,z)p(z)$$

  But yes, interventions are very important and at the core of causality.

- What does it mean to intervene on a GNN?

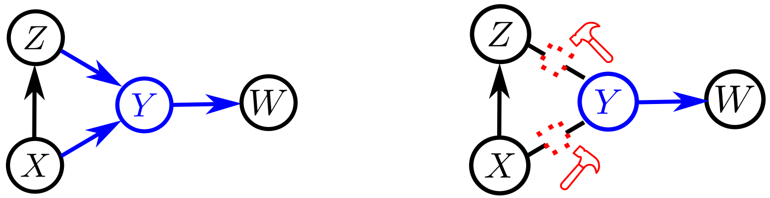


Looking at the SCM



TECHNISCHE UNIVERSITÄT DARMSTADT

Looking at the GNN



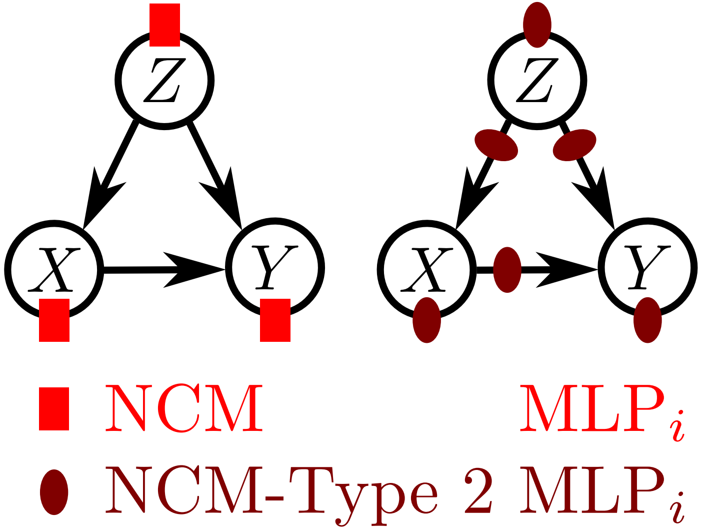
$$f(\mathbf{D}, G \mid \color{red}{do(\mathbf{V}_j)}) = \begin{bmatrix} — & g(\mathbf{d}_X, \mathbf{D}_{\mathcal{M}_X}) & — \\ — & \color{blue}{g(\mathbf{d}_Y, \mathbf{D}_{\color{red}{\mathcal{M}_Y}})} & — \\ — & g(\mathbf{d}_Z, \mathbf{D}_{\mathcal{M}_Z}) & — \\ — & g(\mathbf{d}_W, \mathbf{D}_{\mathcal{M}_W}) & — \end{bmatrix}$$

$$\color{red}{\mathcal{M}_i} = \{j \mid j \in \color{blue}{\mathcal{N}_i}, \\ j \notin pa_i \iff i \in \mathbf{V_j}\}$$

NCM & NCM-Type 2

- NCM: use separate neural networks to model structural equations $|\mathbf{V}|$

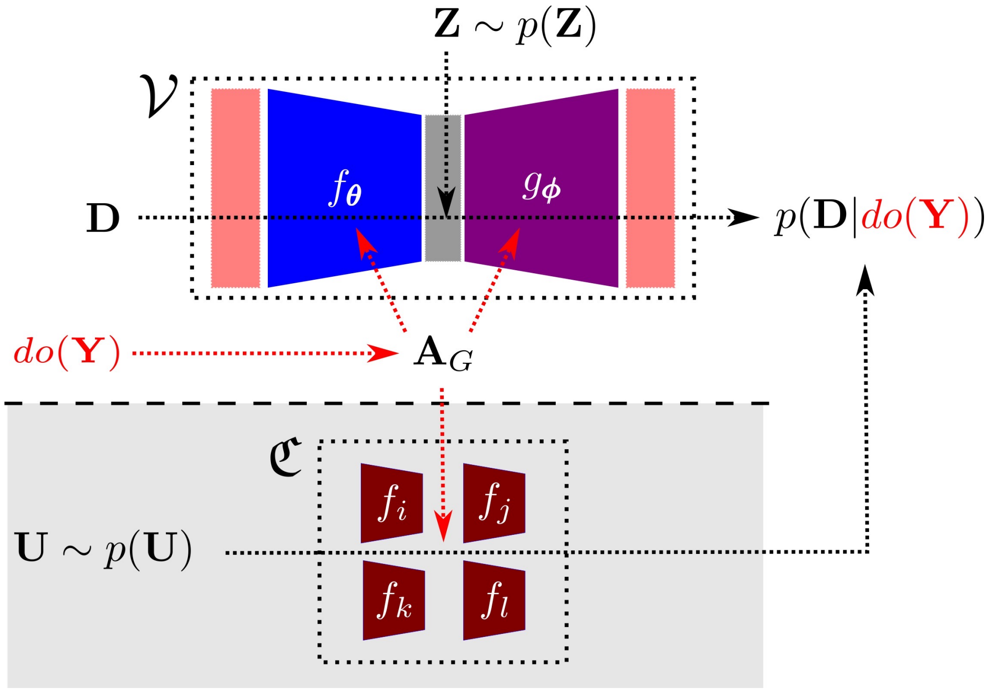
  NCM-T2: … to model a causal relation tuple (plus exogenous variables) $|\mathcal{E}| + |\mathbf{V}|$



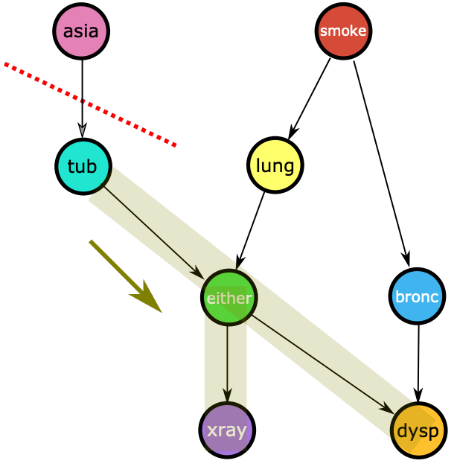
Three Flavours of GNN-based Neuro-Causality

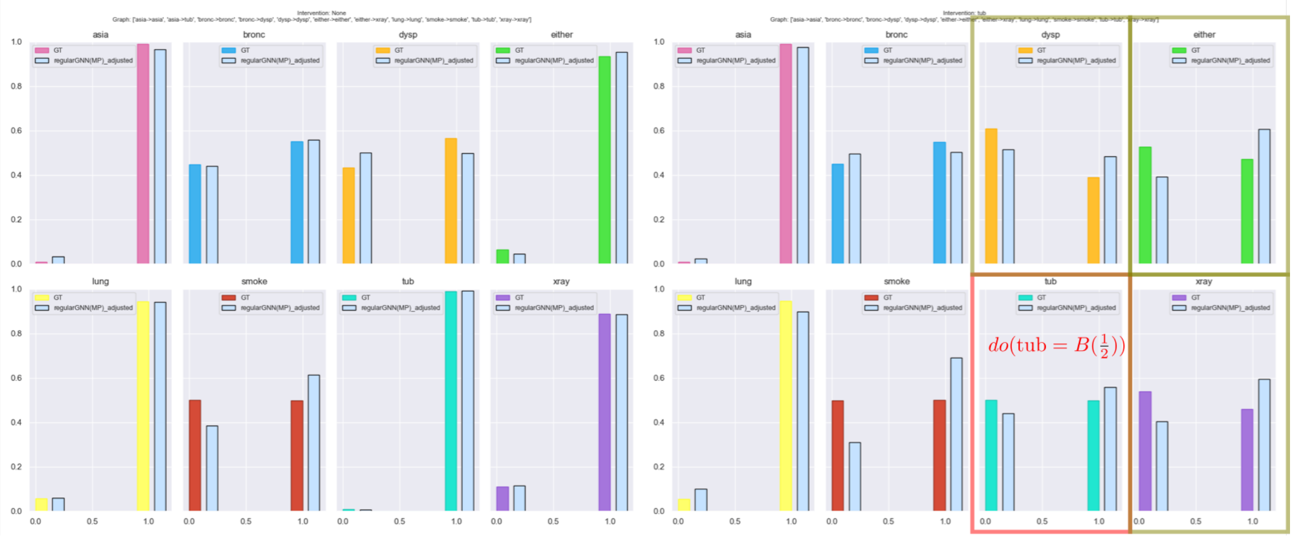
Model	Expressivity (PCH)	Training Difficulty & Cost
iGNN (iVGAE)	*Interventional*	*Easy & Low*
SCM-GNN	*Complete*	*Difficult & High*
NCM-Type 2	*Complete*	*Easy & High*

Looking at the iVGAE



Estimation





S. Lauritzen, D. Spiegelhalter, JRSS, 1988.

4 Linear Programs and Causal Semantics

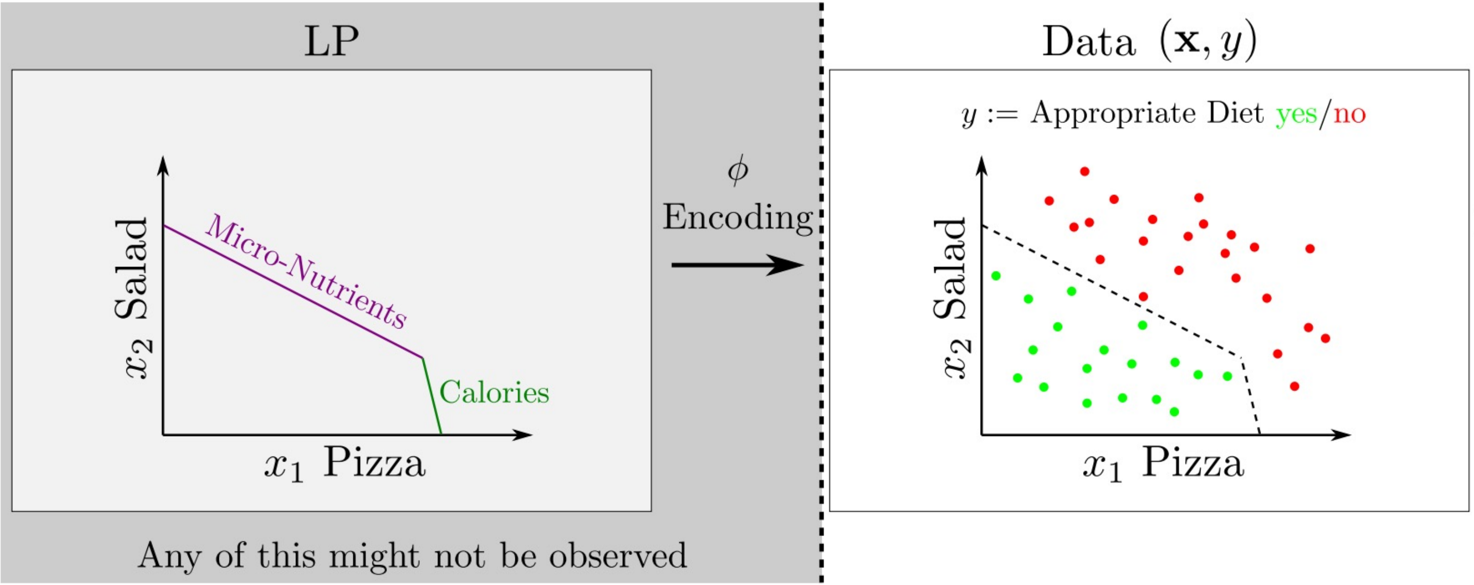
TECHNISCHE UNIVERSITÄT DARMSTADT

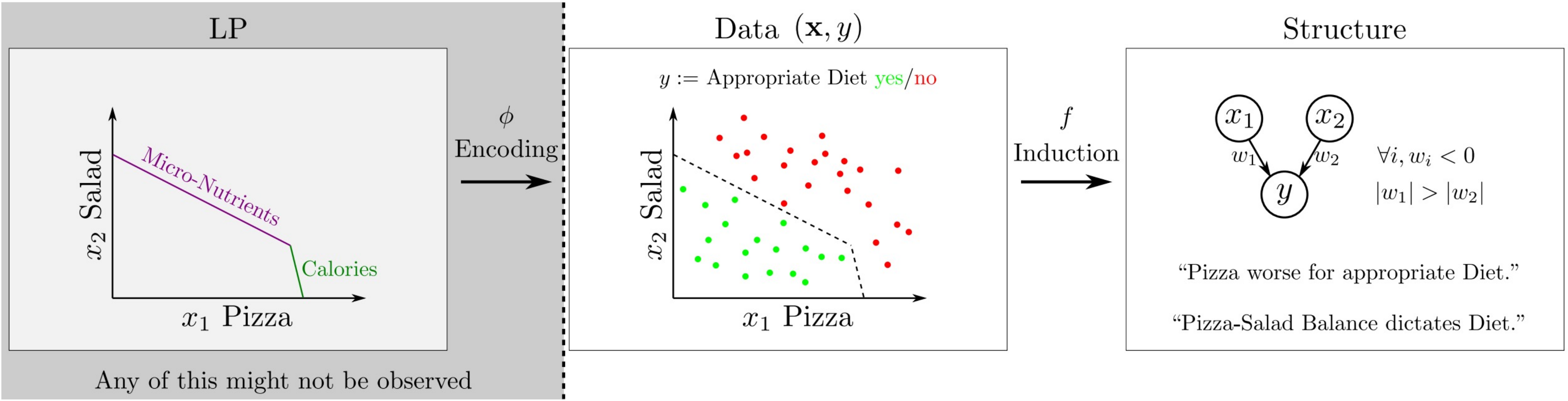
Finding Structure and Causality in Linear Programs

Matej Zečević, Florian Busch, Devendra Singh Dhami, Kristian Kersting

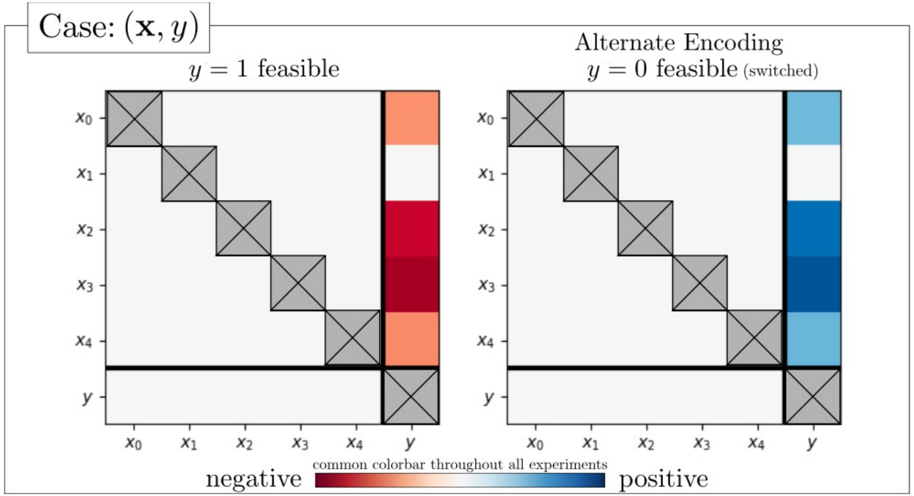
ICLR OSC Workshop 2022

TECHNISCHE UNIVERSITÄT DARMSTADT

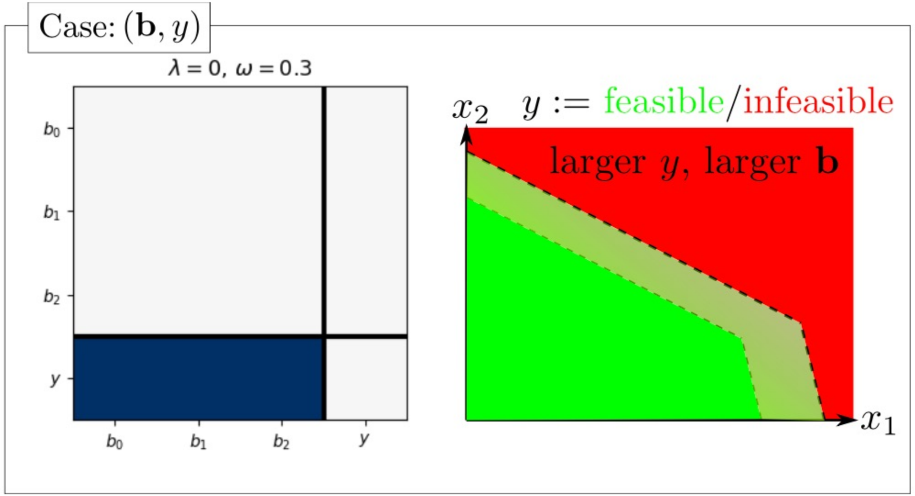

LP
Micro-Nutrients
Calories
x_2 Salad
x_1 Pizza
Any of this might not be observed
ϕ
Encoding
Data $(\mathbf{x}, y)$
$y :=$ Appropriate Diet yes/no
x_2 Salad
x_1 Pizza



TECHNISCHE UNIVERSITÄT DARMSTADT

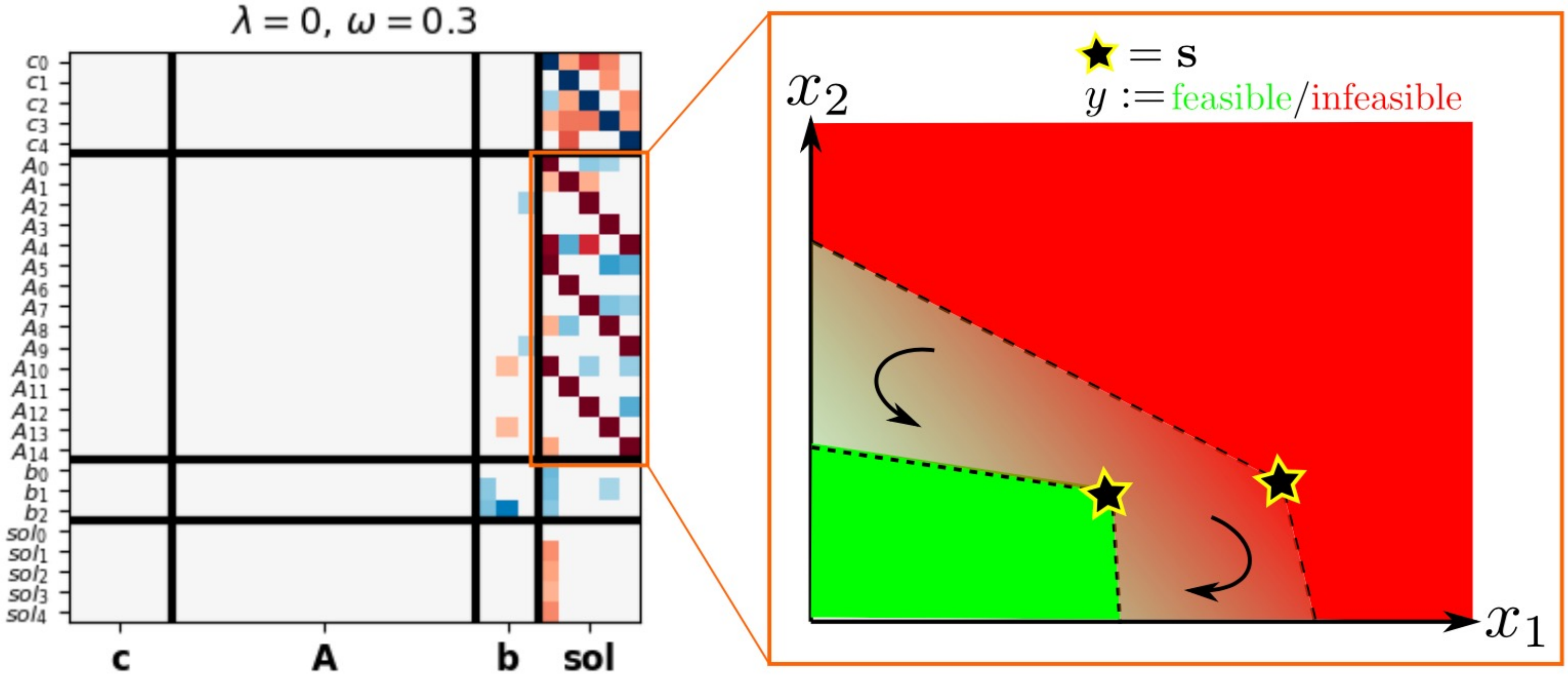


Case: $(\mathbf{x}, y)$
$y = 1$ feasible
Alternate Encoding
$y = 0$ feasible (switched)
x_0
x_1
x_2
x_3
x_4
y
common colorbar throughout all experiments
negative
positive

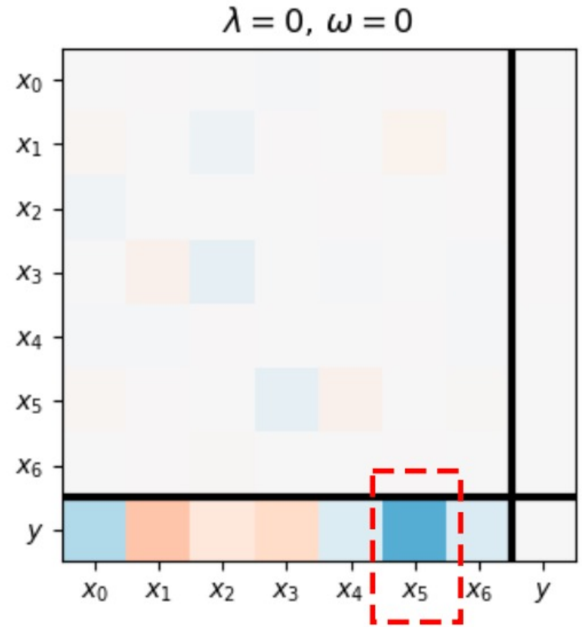


Case: $(\mathbf{b}, y)$
$\lambda = 0,\ \omega = 0.3$
b_0
b_1
b_2
y
$y :=$ feasible/infeasible
x_2
larger y, larger $\mathbf{b}$
x_1

$(\mathbf{c}, \mathbf{A}, \mathbf{b}, \mathbf{s})$

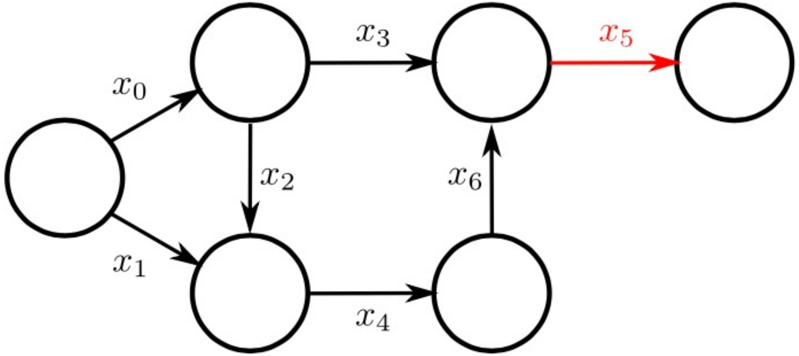


Case: $(\mathbf{x}, y)$

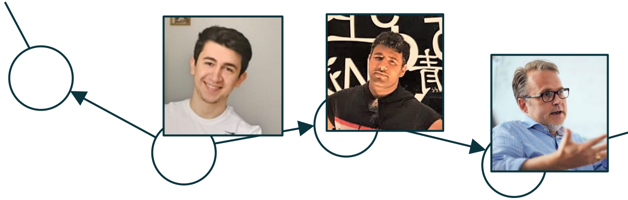


Fixed SP

y denotes valid path



"x_5 must be parth of shortest path."



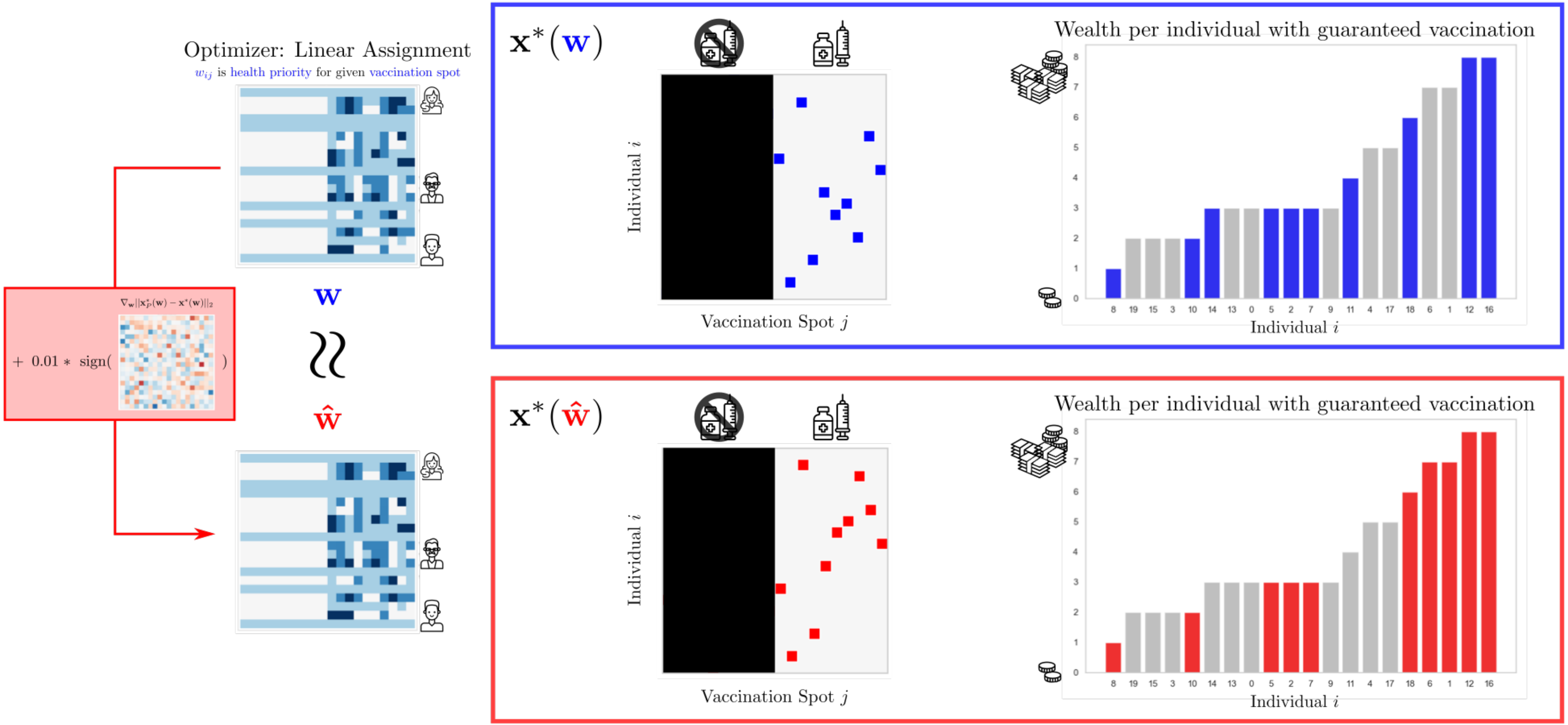
Can Linear Programs Have Adversarial Examples? A Causal Perspective

old name: Intriguing Parameters of
Structural Causal Models

Matej Zečević, Devendra Singh Dhami, Kristian Kersting

arxiv:2105.12697

TECHNISCHE
UNIVERSITÄT
DARMSTADT



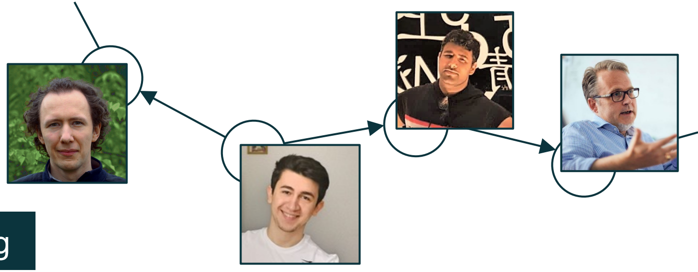
TECHNISCHE UNIVERSITÄT DARMSTADT

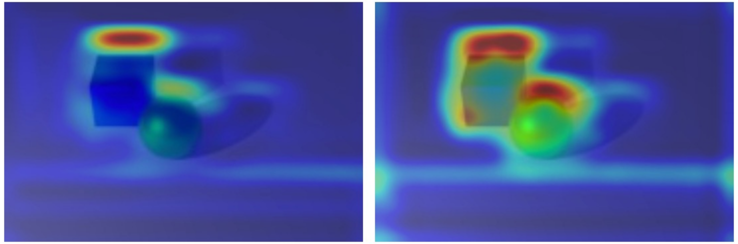
Z | Other Directions for Causal AI/ML

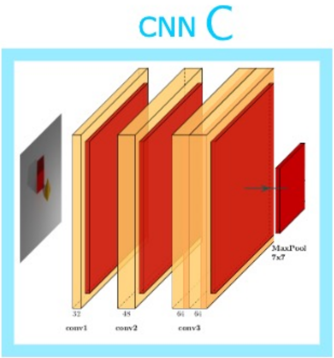
The Causal Loss: Driving Correlation to Imply Causation

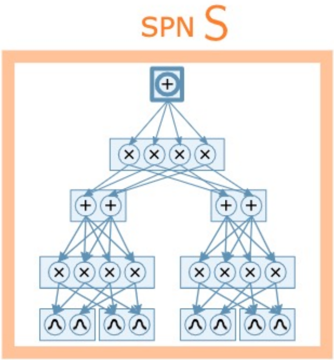
Moritz Willig, Matej Zečević, Devendra Singh Dhami, Kristian Kersting

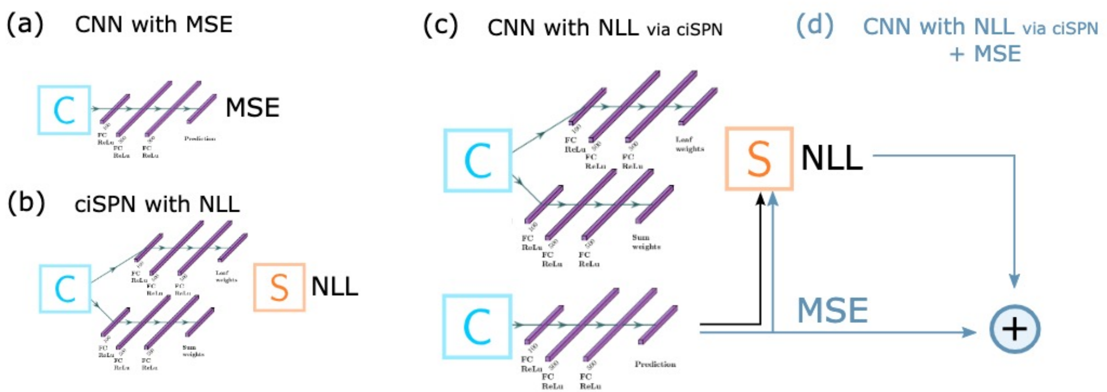
arxiv: 2110.12052







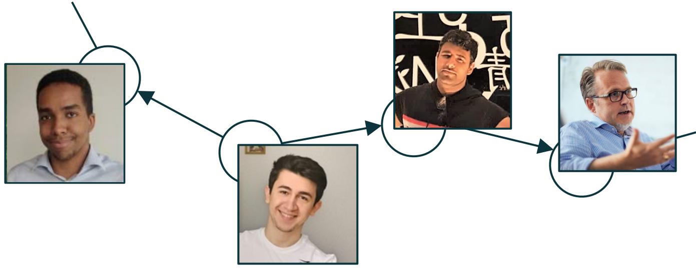


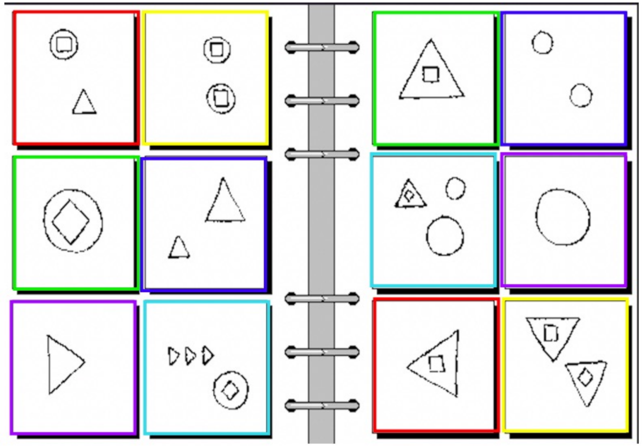


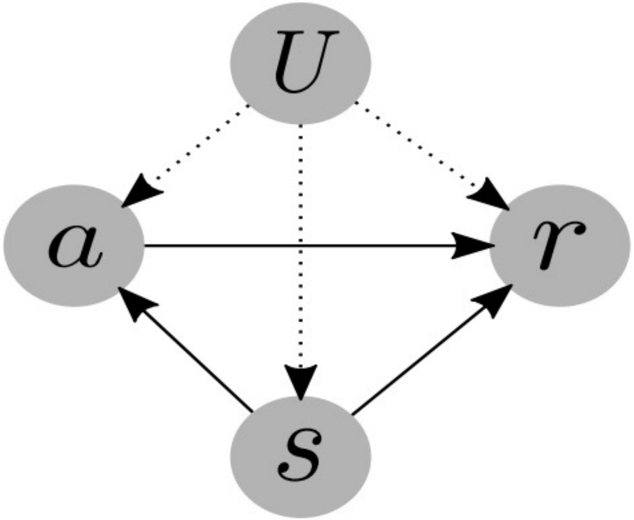
Towards a Solution to Bongard Problems: A Causal Approach

Salah Youssef, Matej Zečević, Devendra Singh Dhami, Kristian Kersting

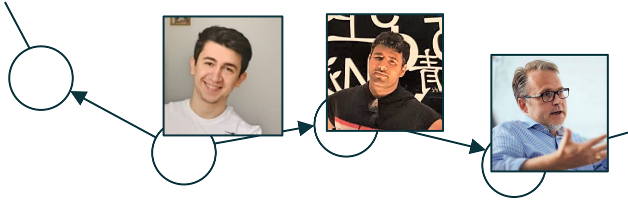
arXiv: 2206.07196







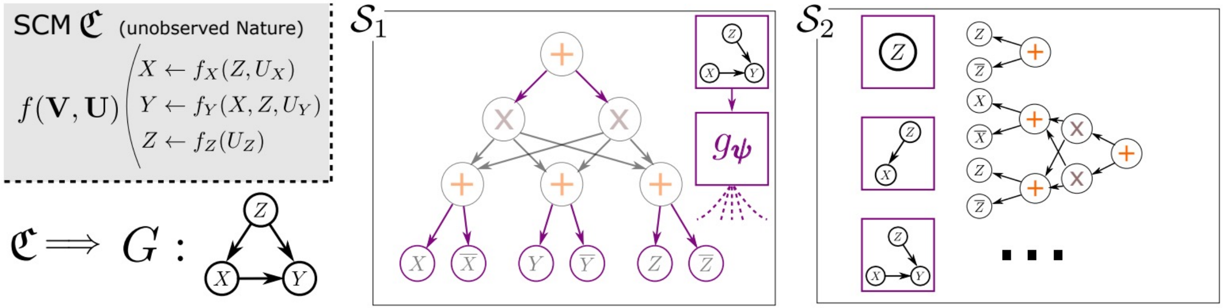
TECHNISCHE UNIVERSITÄT DARMSTADT

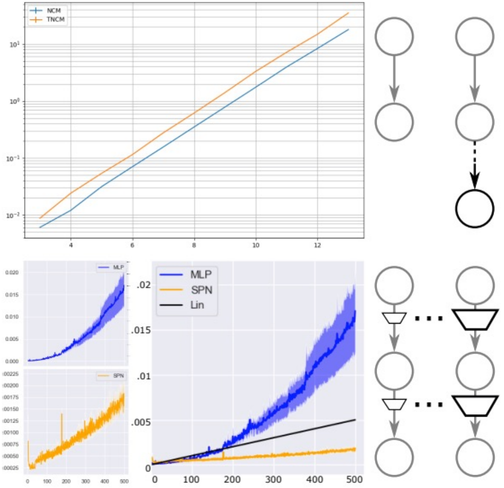


On the Tractability of Neural-Causal Inference

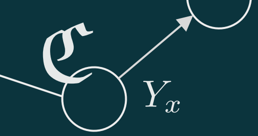
Matej Zečević, Devendra Singh Dhami, Kristian Kersting

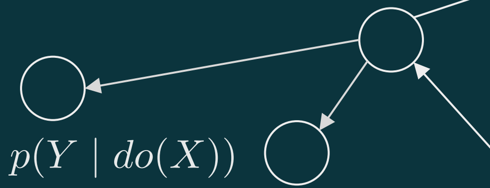
arxiv: 2110.12052





TECHNISCHE UNIVERSITÄT DARMSTADT

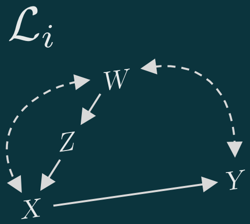




A

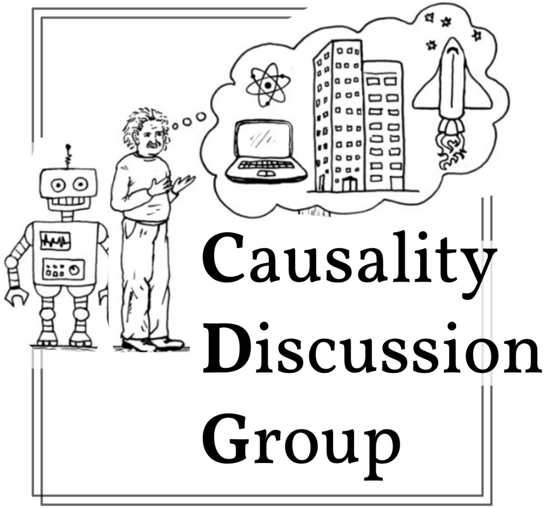
Reminder: Announcements

Opportunities following SSDS 2022



Join the weekly discussion

(or simply listen..)



Join the community
via
discuss.causality.link

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Join the virtual event
ncsi.cause-lab.net

neuro (Causal ∧ Symbolic) AI

Workshop at the 36th Conference on Neural Information Processing Systems (NeurIPS)

December 2022

Visit **ncsi.cause-lab.net**

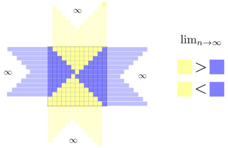
TECHNISCHE UNIVERSITÄT DARMSTADT

Stay in Touch via Social Media

Twitter
@matej_zecevic

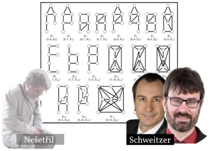
The Infinite is Useful

Talking about the foundation of mathematics usually involves at some initial point the discussion on set theory, which one arguably considers to be a "theory...



Proving Nešetřil's Conjecture: Minimal Asymmetric Graphs

In 1988 at a seminar in Oberwolfach (located in Germany and by many mathematicians considered as a sort of "mecca for mathematicians" [1]) the czech...



Sports Excellence: The Greatest Shohei Ono

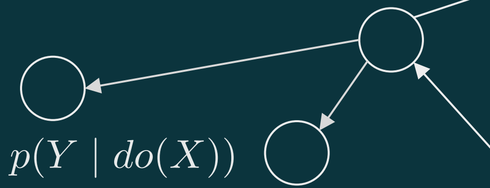
Judo (柔道, japanese for "gentle way") is a martial art centered around throwing techniques for close-quarters combat (opposed to for instance Karate, another japanese martial...

Open-access articles on my Blog

(pretty much on anything but mostly on nerdy topics)

matej-zecevic.de

TECHNISCHE
UNIVERSITÄT
DARMSTADT



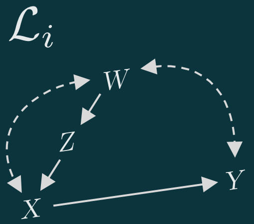
We finished CML I + II !

Questions?

See you at the Hands-on Code Session

matej.zecevic@tu-darmstadt.de | https://www.matej-zecevic.de

Further, my gratitude and thanks go out to Kristian Kersting, Devendra Dhami, all our collaborators,
the AIML lab @ TU Darmstadt and finally the organizers of SSDS 2022



TECHNISCHE
UNIVERSITÄT
DARMSTADT