Towards Neuro-Causality

Relating Graph Neural Networks to Structural Causal Models

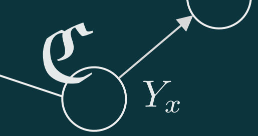
Matej Zečević

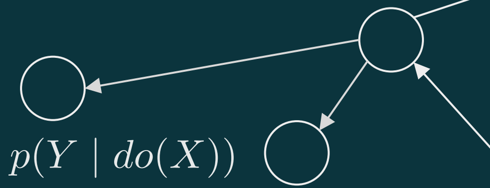
TECHNISCHE UNIVERSITÄT DARMSTADT

Learning on Graphs and Geometry
Reading Group
9th November 2021

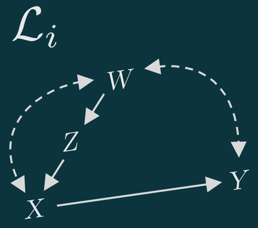
TL;DR

Merging Causality with modern Deep Learning is not futile.
We collectively as researches have already discovered promising directions.





I | Why?



TECHNISCHE UNIVERSITÄT DARMSTADT

Recent Attention

Judea Pearl @yudapearl

1/ This paper deserves attention for deriving new theoretical connections between GNNs and SCMs. Moreover, it allows causal information to be provided the natural way, i.e., via DAGs, then translated to GNN. What is not clear to this reader, though it may be implicit in the text,

> **Mehmet Süzen** @memosisland · Sep 11
> #graph #deeplearning meets #causality via SCM
> arxiv.org/abs/2109.04173
> fyi @yudapearl @mmbronstein @eliasbareinboim

12:20 PM · Sep 12, 2021 · Twitter Web App

50 Retweets **3** Quote Tweets **238** Likes

Elias Bareinboim @eliasbareinboim

1/3 This seems to be a nice extension of the work causal-neural connection (arxiv.org/pdf/2107.00793...) for GNNs, but I haven't fully read it (congrats, @kerstingAIML & team!). Still, we answered precisely this question in Q11 in the FAQ (p. 52), also attached for your convenience.

Elias Bareinboim @eliasbareinboim

3/3 It's not one approach over the other but they are complementary, & understanding one can help with the other. It would be impossible to devise what we did w/o understanding the do-calc, causal bayes nets, PCH, CHT. I think the FAQ, App. C4 & discussion around make this clear.

11:22 PM · Sep 12, 2021 · Twitter Web App

Emanuele @iperboreo_

Replying to @devendratweetin @PetarV_93 and @kerstingAIML

I suspect this is going to have a big impact the field of "NNs&causality".
Great job!

Matej Zečević @matej_zecevic

1/2 Thanks @yudapearl! We agree with @eliasbareinboim. In fact 3 different flavors of NCM (SCM-GNN, NCM-T2, iVGAE) already follow from our paper. Their merit differ in terms of time-/space-complexities regarding representation, optimization, and inference.

> **Judea Pearl** @yudapearl · Sep 12
> Replying to @eliasbareinboim
> Thanks for explaining the gains. In particular, I am intrigued by the facility to easily generate compatible scenarios.

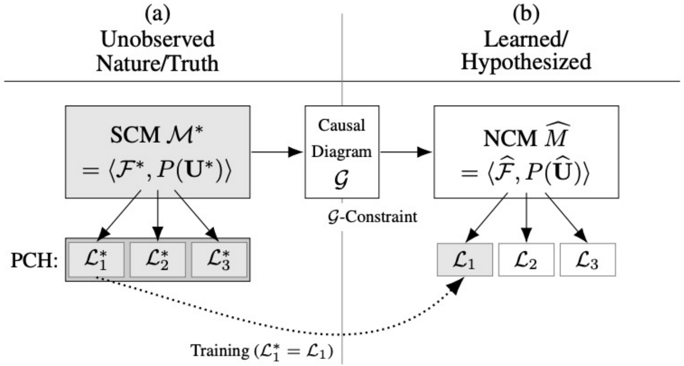
...

Theoretical Milestone

The Causal-Neural Connection:
Expressiveness, Learnability, and Inference

Kevin Xia, Kai-Zhan Lee, Yoshua Bengio, Elias Bareinboim

arxiv:2107.00793

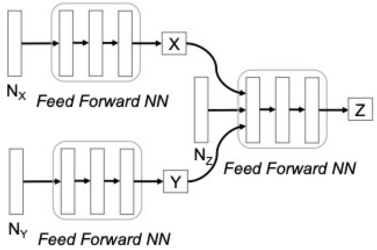


Previous Foundations

CausalGAN:
Learning Causal Implicit Generative Models with Adversarial Training

Murat Kocaoglu, Christopher Snyder, Alexandros Dimakis, Sriram Vishwanath

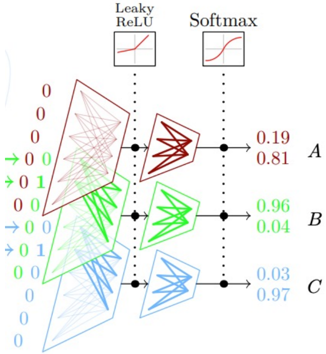
ICLR 2018



Learning Neural Causal Models from Unknown Interventions

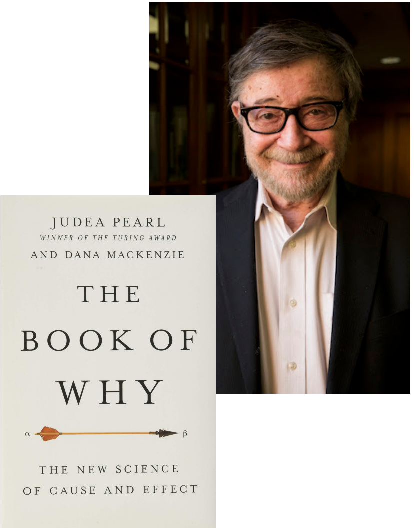
Nan Rosemary Ke, Olexa Bilaniuk, Anirudh Goyal, Stefan Bauer, Hugo Larochelle, Bernhard Schölkopf, Michael C. Mozer, Chris Pal, Yoshua Bengio

arxiv:1910.01075



TECHNISCHE UNIVERSITÄT DARMSTADT

Why?

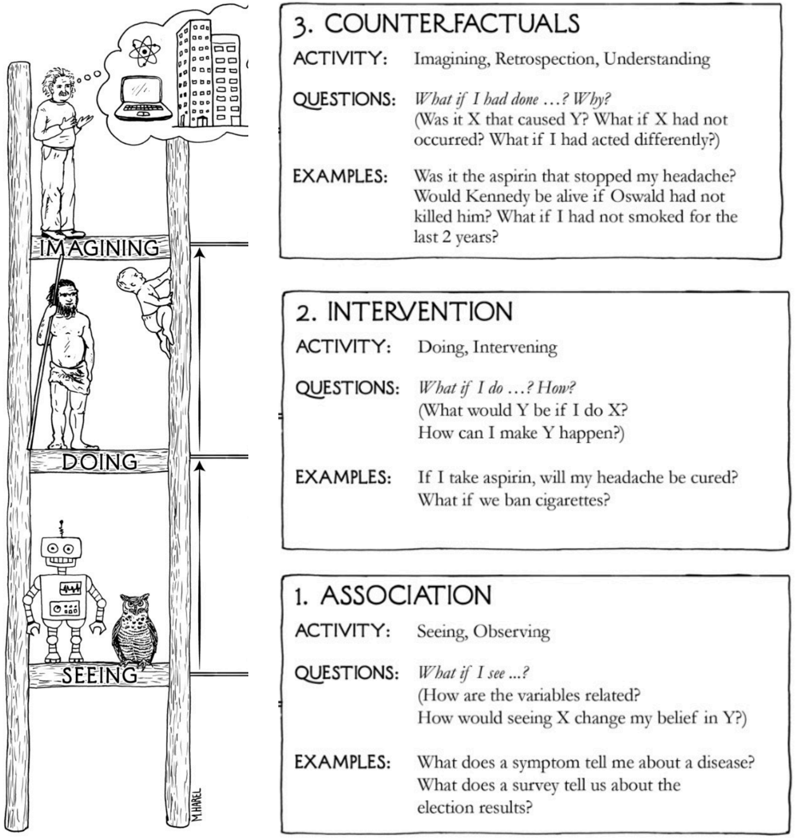


"To Build Truly Intelligent Machines, Teach Them Cause and Effect"

"All the impressive achievements of deep learning amount to just curve fitting"

Judea Pearl in "The Book of Why" and in an interview with quantamagazine in 2018

Why?



Pearl Causal Hierarchy (PCH), “The Book of Why” (2018)

On Pearl’s Hierarchy and
the Foundations of Causal Inference

Elias Bareinboim, Juan Correa, Duligur Ibeling, Thomas Icard

causalai.net/r60

Theorem 1. *[Causal Hierarchy Theorem (CHT), formal version] With respect to the Lebesgue measure over (a suitable encoding of $\mathcal{L}_3$-equivalence classes of) SCMs, the subset in which any* PCH *collapse occurs is measure zero.* ■

TECHNISCHE
UNIVERSITÄT
DARMSTADT

What would be if not?

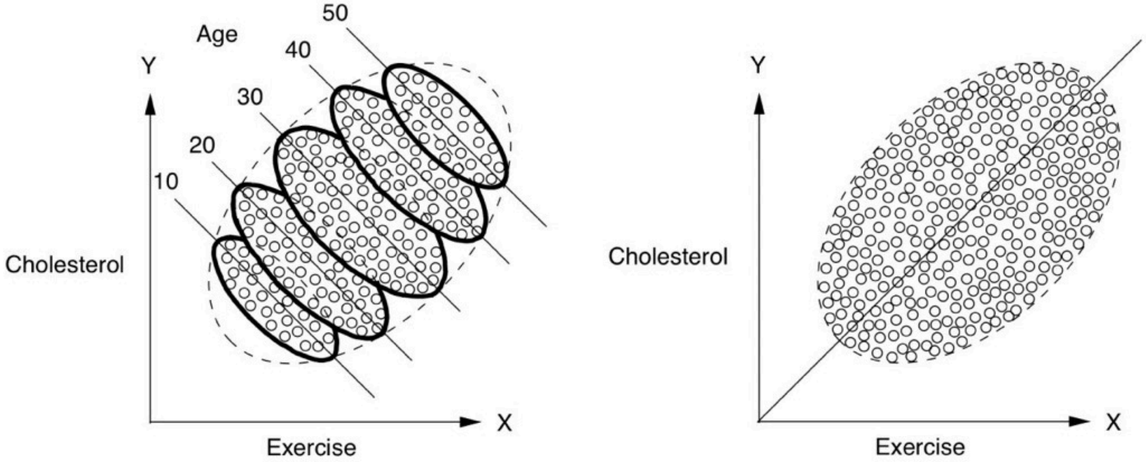


Figure from “The Book of Why” (2018) by
Judea Pearl, Dana Mackenzie

TECHNISCHE UNIVERSITÄT DARMSTADT

What would be if not?

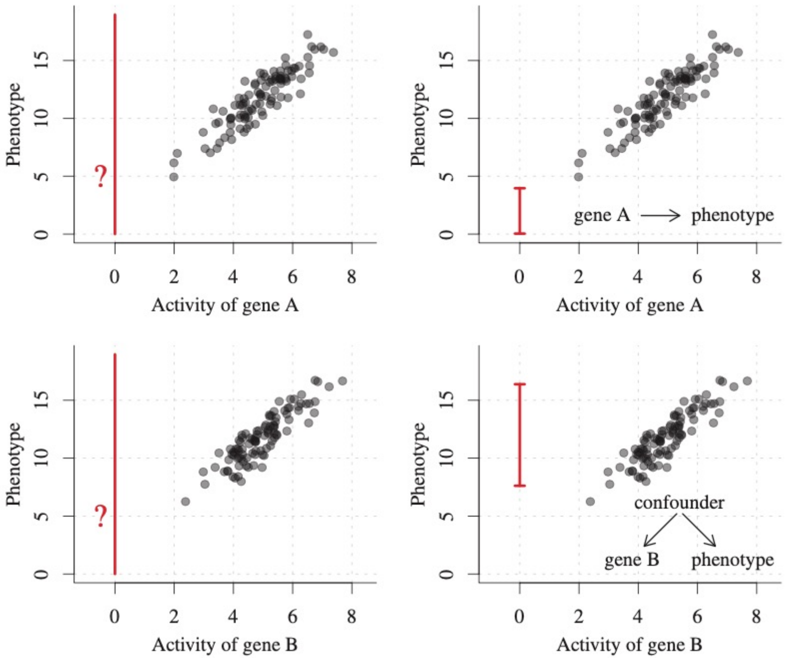


Figure from "Elements of Causal Inference" (2017) by
Jonas Peters, Dominik Janzing and Bernhard Schölkopf

Akin to: What is Artificial Intelligence?

"It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable."
- John McCarthy (2007)

What is Neuro-Causality?

It is the science and engineering of
integrating causality with neural-based methodologies.

TECHNISCHE
UNIVERSITÄT
DARMSTADT

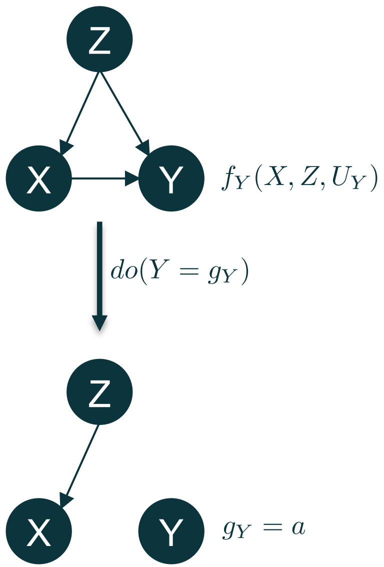
2 Primer

Causal Inference 101

- **Structural Causal Model** (SCM) $\mathfrak{C} = (\mathbf{S}, P(\mathbf{U}))$
  describe the mechanistic relations of variables $\mathbf{V}$
  $$V_i := f_i(pa(V_i), U_i), \quad \text{where } i = 1, \dots, d$$
  where $P(\mathbf{U})$ factorizes the exogenous variables. (Markovian-SCM)

- An **intervention** actively replaces the origin mechanism $f_{\mathbf{W}}$
  with a new mechanism, denoted by $do(\mathbf{W} = g_{\mathbf{w}})$.

- An important view on **valuation** is given by:
  $$p^{\mathfrak{C}}(\mathbf{w} \mid do(\mathbf{z})) = \sum_{\{\mathbf{u} \mid \mathbf{W}_{\mathbf{z}}(\mathbf{u}) = \mathbf{w}\}} p(\mathbf{u})$$
  where $\mathbf{W}_{\mathbf{z}} : \mathbf{U} \mapsto \mathbf{W}$.



Pointers to Causal Inference References

- Judea Pearl, “**Causality**”, Cambridge University Press, 2009.

- Peters et al., “**Elements of Causal Inference**”, MIT Press, 2017.

- Elias Bareinboim Lecture “**Causal Data Science**”, 2019.

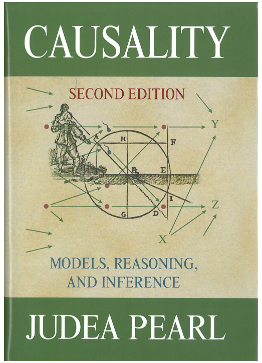
  https://www.youtube.com/watch?v=dUsokjG4DHc

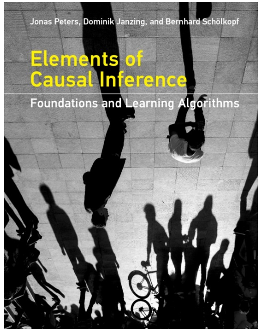
- Brady Neal’s Free Online Course ”**Introduction to Causal Inference**”, 2020.

  https://www.bradyneal.com/causal-inference-course

- Jonas Peters Lecture Series “**Causality**”, 2017.

  https://www.youtube.com/watch?v=zvrcyqcN9Wo





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* $p(\mathbf{Z}|\mathbf{X})$

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

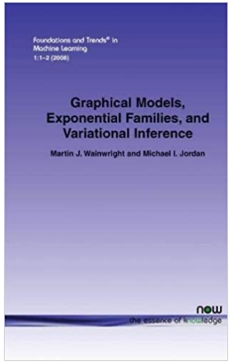
  where $\mathcal{Q}$ denotes a selected variational family of distributions (e.g. MoG).

- 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/`

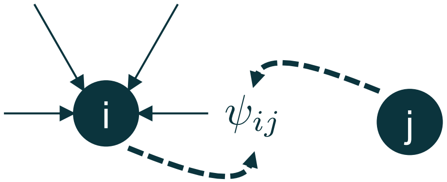
Graph Neural Networks 101

- 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:

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

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



TECHNISCHE UNIVERSITÄT DARMSTADT

Pointers to GNN References

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

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

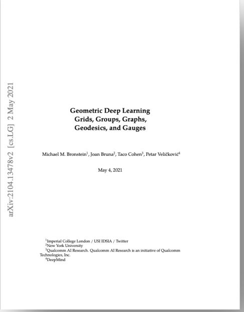
- Bronstein et al., “**Geometric Deep Learning**”, arXiv, 2021.

  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

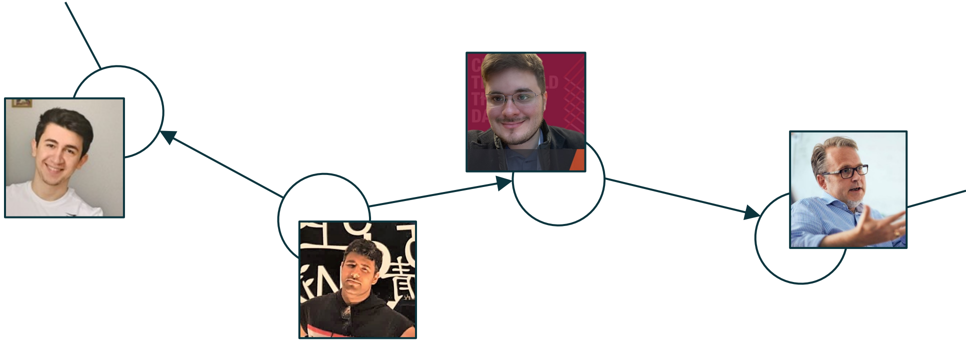
Geometric Deep Learning
Grids, Groups, Graphs,
Geodesics, and Gauges

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

3 | GNNs, NCMs and SCMs



Relating Graph Neural Networks to Structural Causal Models

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

arxiv: 2109.04173

TECHNISCHE UNIVERSITÄT DARMSTADT

From First Principles

- Peter Holland & Don Rubin (1986):

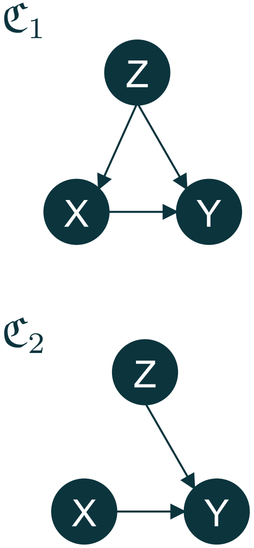
  “No causation without manipulation”

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

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

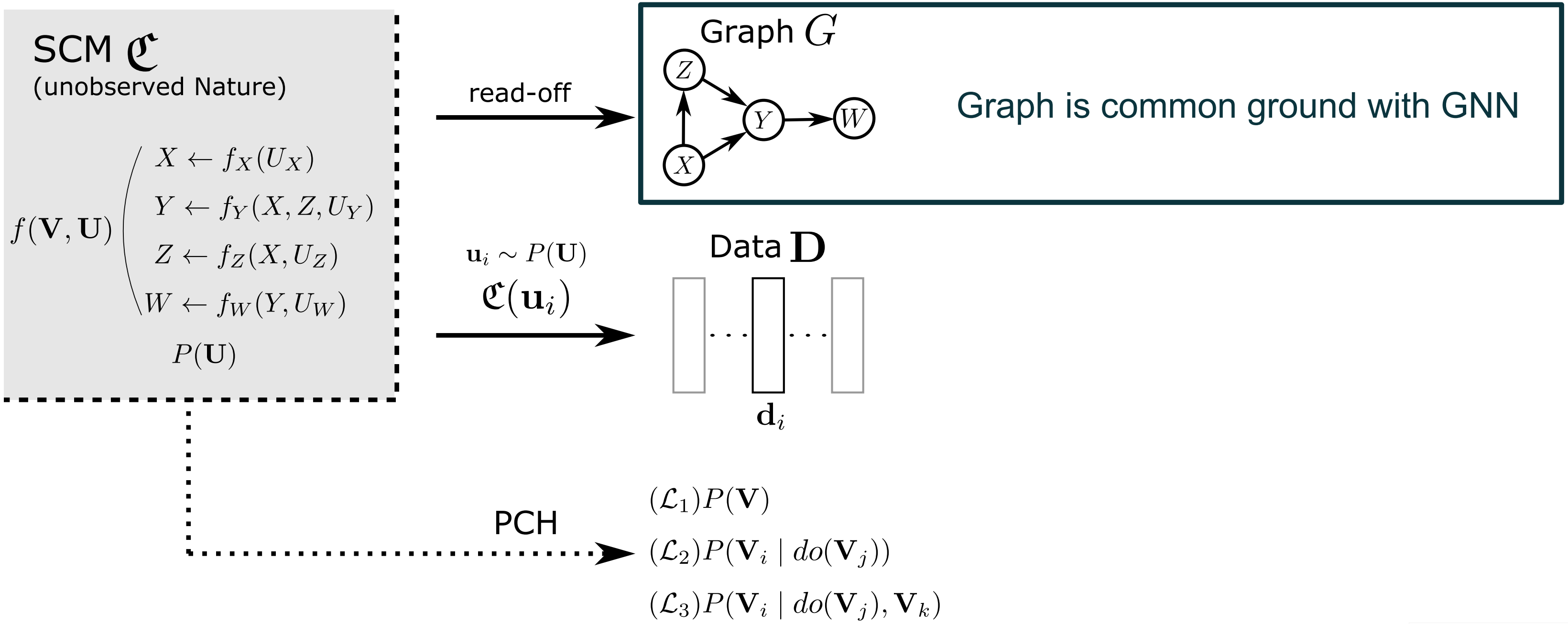
  But yes, interventions are very important and at the core of causality.

- What does it mean to intervene on a GNN?

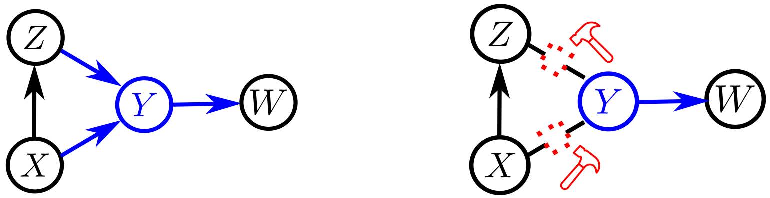


$$p^{\mathfrak{C}_1}(y|do(x)) = p^{\mathfrak{C}_2}(y|x) \neq p^{\mathfrak{C}_1}(y|x)$$

Looking at the SCM



Looking at the GNN



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

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

Interventional GNN

Definition 1 ***(Interventions within GNN.)*** *An intervention* $\mathbf{x}$ *on the corresponding set of variables* $\mathbf{X} \subseteq \mathbf{V}$ *within a GNN layer* $f(\mathbf{D}, \mathbf{A}_G)$*, denoted by* $f(\mathbf{D}, \mathbf{A}_G | \, do(\mathbf{X} = \mathbf{x}))$*, is defined as a modified layer computation,*

$$\mathbf{h}_i = \phi\left(\mathbf{d}_i, \bigoplus_{j \in \mathcal{M}_i^G} \psi(\mathbf{d}_i, \mathbf{d}_j)\right), \tag{1}$$

where the intervened local neighborhood is given by

$$\mathcal{M}_i^G = \{j \mid j \in \mathcal{N}_i^G, j \notin \mathrm{pa}_i \iff i \in \mathbf{X}\} \tag{2}$$

where $\mathcal{N}^G$ *denotes the regular graph neighborhood. A GNN layer that computes Eq.1 is said to be interventional.*

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Concerning $\mathcal{L}_3$

- Def.1 allows the GNN to go beyond $\mathcal{L}_1$ associational to $\mathcal{L}_2$ interventional.
  But what about *counterfactuals* ($\mathcal{L}_3$; the full PCH)?

- Interpreting the "messages" within the general GNN formulation
  in terms of the **structural equation dependencies** (or causal effects)
  allows for a conversion between SCM and GNN:

Theorem 1 ***(GNN-SCM Conversion.)*** *Consider the most general formulation of a message-passing GNN node computation* $\mathbf{h}_i{:}\mathcal{F} \mapsto \mathcal{F}'$. *For any SCM* $\mathfrak{C}{=}(\mathbf{S}, P(\mathbf{U}))$ *there exists always a choice of feature spaces* $\mathcal{F}, \mathcal{F}'$ *and shared functions* ϕ, ψ, *such that for all structural equations* $f{\in}\mathbf{S}$ *it holds that* $\mathbf{h}_i = f_i$.

Structural Equation Dependencies

- Decomposing a structural equation into causal dependencies on the parental-level.

  For linear SCM, this decomposition is *purely* parental.

$$
\begin{aligned}
X \leftarrow f_X(Z, U_X) &= f_{XZ}(Z) + f_{U_X}(U_X, \cdots) \\
Y \leftarrow f_Y(Z, X, U_Y) &= f_{YZ}(Z) + f_{YX}(X) + f_{U_Y}(U_Y, \cdots) \\
Z \leftarrow f_Z(U_Z) &= f_{U_Z}(U_Z, \cdots)
\end{aligned}
$$

- A binary SCM Example:

$$\mathfrak{C} = (\{f_X(Z, U_X) := Z \wedge U_X,\ f_Z(U_Z) := U_Z\}, \quad P(U_X, U_Z))$$

$$
f_X = \begin{array}{cc} U_X = 0 & U_X = 1 \end{array}
$$

$$
f_X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{array}{l} Z = 1 \\ Z = 0 \end{array} \iff \begin{array}{l} f_{XZ}(Z) + f_{U_X}(U_X, Z) \\ \quad [Z] + [U_X - (Z \vee U_X)]. \end{array}
$$

What Thm.1 does not talk about

- Thm.1 presents a direct conversion from SCM to GNN.

  Thus establishing GNN as a Neural Causal Model (NCM) variant.

  However, Thm.1 does *not* talk about **optimization**.

  The GNN message-computing function is *shared.*

$$V_i = \phi\left(v_i, \bigoplus_{j \in \mathcal{N}_i^G} \psi(v_i, v_j) \right)$$

$$\psi(i,j) = \begin{cases} f_{XZ}(Z), & \text{for } i = X, j = Z \\ f_{YZ}(Z), & \text{for } i = Y, j = Z \\ f_{YX}(X), & \text{for } i = Y, j = X \end{cases}$$

  While the structural equations are specific functions *for each* variable.

- What happens if we violate the sharedness-property?

A more fine-grained NCM

- An actual neural parameterization of an SCM.

  Thereby capable of all three levels of the PCH.

  Opposed to NCM, it models on the *edge* level.

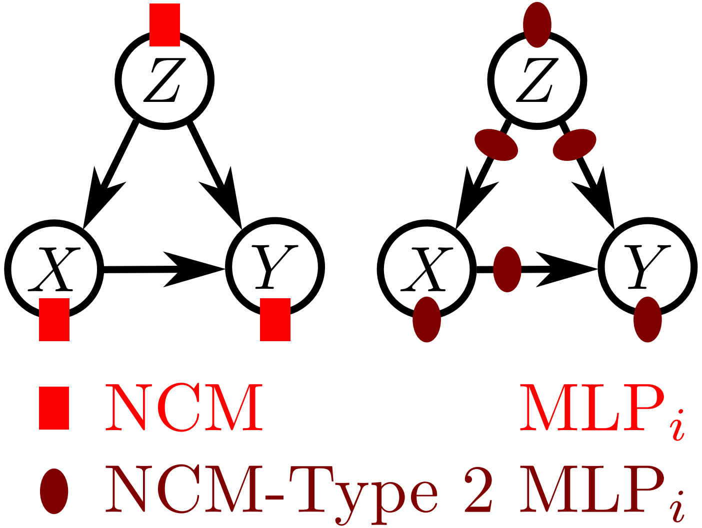
Corollary 1 ***(NCM-Type 2.)*** *Allowing for the violation of sharedness of* ψ *as depicted in Thm.1 and choosing* $\mathcal{F} = \mathcal{F}' = \mathbf{U} \cup \mathbf{V}$ *to be the union over endo- and exogenous variables,* $\phi(i, \dots) = f_{U_i}(U_i, \mathcal{A}_i) + \sum(\dots)$ *to be a sum-aggregator with noise term selection with* $\mathcal{A}_i \in 2^{|\,\mathrm{pa}_i\,|}$*, and* $\psi = \{f^{ij}_{\boldsymbol{\theta}}\}$ *to be the dependency terms of the structural equations* f_i *modelled as feedforward neural networks. Then the computation layer* $\{\mathbf{h}_i\}_i^{|V|}$ *is a special case of the NCM as defined in (Xia et al. 2021).*

TECHNISCHE UNIVERSITÄT DARMSTADT

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) $|\mathcal{E}| + |\mathbf{V}|$

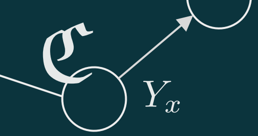


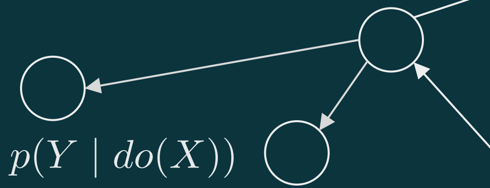
Three Flavours of GNN-based Neuro-Causality

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

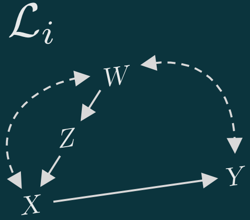
Our Research Selection

We decide on investigating **iGNN** further.
From the neuro-causal viewpoint, this model is especially interesting since it **trades-off expressivity against model description.**





4 | GNN-based Causal Inference



Constructing GNN from Causal Graphs

- Any SCM will imply a causal graph.

  This very graph can be used for constructing a GNN-layer:

  Definition 2 ***($\mathfrak{G}$-GNN construction.)*** *Let $\mathfrak{G}$ be the graph induced by SCM $\mathfrak{C}$. A GNN layer $f(\mathbf{D}, \mathbf{A}_G)$ for which $G = \mathfrak{G}$ is said to be $\mathfrak{G}$-constructed.*

  Following Def.1, an intervention will be "**natural**" to the way it occurs in an SCM:

  Proposition 1 ***(Graph Mutilation Equivalence.)*** *Let $\mathfrak{C}$ be an SCM with graph $\mathfrak{G}$ and let f be a $\mathfrak{G}$-GNN layer. An intervention $do(\mathbf{X}), \mathbf{X} \subseteq V$, on both $\mathfrak{C}$ and f produces the same mutilated graph $\mathfrak{G}'$.*

A Causal Generative Model based on iGNN

- The iGNN (Def.1) is a computation layer able to parameterize a generative model.

  First consider a VGAE which uses GNN-layers:

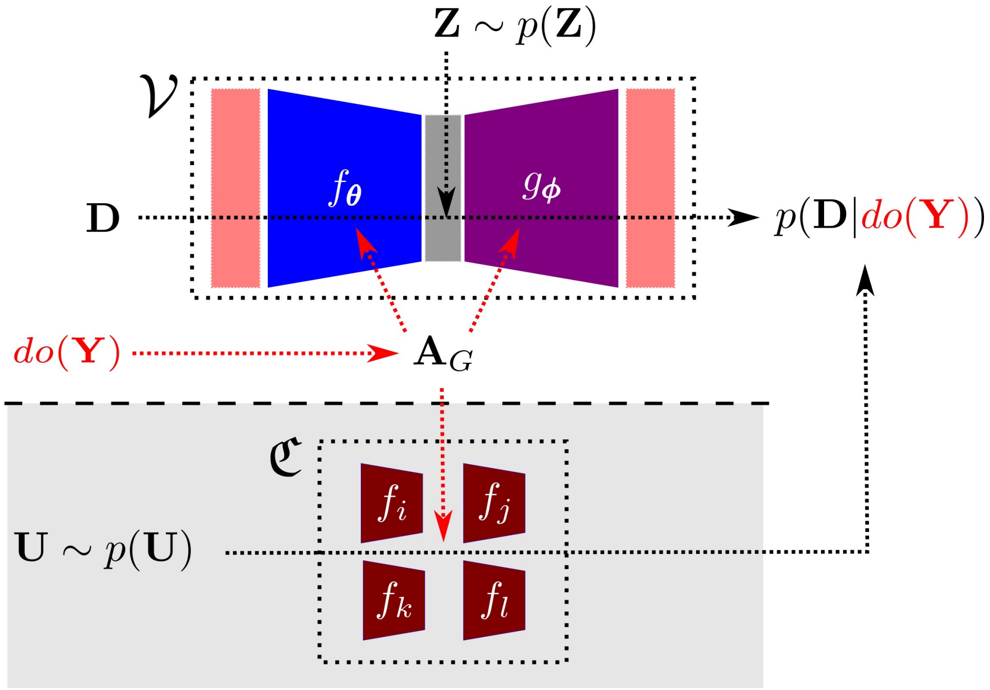
  Definition 4 ***(Variational Graph Auto-Encoder.)*** *We define a VGAE* $\mathcal{V}{=}(q(\mathbf{Z}|\mathbf{D}), p(\mathbf{D}|\mathbf{Z}))$, *as a data-generative model, with inference and generator model respectively s.t.* $q{:=}f_{\boldsymbol{\theta}}(\mathbf{D}, \mathbf{A})$ *is a GNN layer and* $p{:=}g_{\boldsymbol{\phi}}(\mathbf{D}, \mathbf{Z})$ *some parameterized data* $\mathbf{D}$ *dependent model, where* $\boldsymbol{\theta}, \boldsymbol{\phi}$ *are the variational parameters.*

  Following consider an **iGNN-based VGAE**:

  Definition 5 ***(Interventional VGAE.)*** *An interventional VGAE is a data-generative VGAE* $\mathcal{V}{=}(q, p)$ *where both* q, p *are set to be interventional GNN layers* $f_i(\mathcal{D}, \mathbf{A}_G | \, do(\mathbf{X}))$ *where* $\mathcal{D} = \{\mathbf{D}, \mathbf{Z}\}$ *respectively.*

TECHNISCHE UNIVERSITÄT DARMSTADT

Looking at the iVGAE



Measuring Causal Expressivity

- Remember that iVGAE are **not** SCM.

  They trade expressivity against model description.

  Therefore, agreement with a PCH-level $\mathcal{L}_i$ can only be achieved *partially:*

  Definition 3 ***(Partial $\mathcal{L}_i$-Consistency.)*** *Consider a model $\mathcal{M}$ capable of partially emitting PCH, that is $\mathcal{L}_i$ for $i \in \{1,2\}$, and an SCM $\mathfrak{C}$. $\mathcal{M}$ is said to be partially $\mathcal{L}_i$-consistent w.r.t. $\mathfrak{C}$ if $\mathcal{L}_i(\mathcal{M}) \subset \mathcal{L}_i(\mathfrak{C})$ with $|\mathcal{L}_i(\mathcal{M})| > 0$.*

  The *strength* of the consistency we will observe will depend on our data.

  Note that $|\mathcal{L}_1| = 1$ while $|\mathcal{L}_2| \to \infty$.

The Strength of GNN-based Causal Inference

- Pick *any* SCM of your choice.

  There *will* be a corresponding iVGAE to agree on an arbitrarily large subset

  of interventional distributions (including the observational distribution):

Theorem 3 ***(Expressivity.)*** *For any SCM $\mathfrak{C}$ there exists an iVGAE $\mathcal{V}(\boldsymbol{\theta}, \boldsymbol{\phi})$ for which $\mathcal{V}$ is $\mathcal{L}_2$-consistent w.r.t $\mathfrak{C}$.*

TECHNISCHE UNIVERSITÄT DARMSTADT

Causal Hierarchy Theorem for iVGAE

- Bareinboim et al. proved the CHT.

  CHT ensures that causal inference "will still make sense".

  I.e., the PCH remains strict and thereby all inter-layer inferences follow a **single** direction:

Corollary 2 ***(iVGAE Partial Causal Hierarchy Theorem.)*** *Consider the sets of all SCM and iVGAE, Ω, Υ, respectively. If for all $\mathcal{V} \in \Upsilon$ it holds that $\mathcal{V}$ is $\mathcal{L}_1^p(\mathcal{V}) = \mathcal{L}_1^p(\mathfrak{C}) \implies \mathcal{L}_2^q(\mathcal{V}) = \mathcal{L}_2^q(\mathfrak{C})$ with $\mathfrak{C} \in \Omega$, where $\mathcal{L}_i^p \subset \mathcal{L}_i$ is a selection p, q over the set of all level 1,2 distributions respectively, then we say that layer 2 of iVGAE collapses relative to $\mathfrak{C}$. On the Lebesgue measure over SCMs, the subset in which layer 2 of iVGAE collapses to layer 1 has measure zero.*

The Weakness of GNN-based Causal Inference

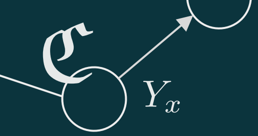
- The iVGAE model class is incapable of counterfactual reasoning ($\mathcal{L}_3$).

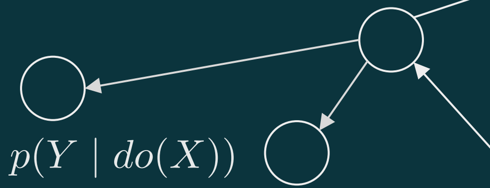
  Remember, we trade expressivity against model description (or compression).

Corollary 3 ***(iVGAE Limitation.)*** *For any SCM* $\mathfrak{C}$ *there exists no iVGAE* $\mathcal{V}$ *such that* $\mathcal{V}$ *is* $\mathcal{L}_3$*-consistent w.r.t.* $\mathfrak{C}$.

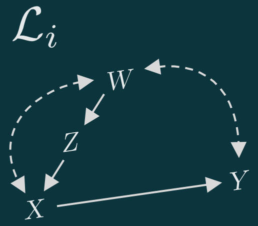
What about practical causal inference?

For the first time we inspected a neural model – which is not an SCM.
Further, we considered their causal prowess in relation to SCM.





5 | Identifiability & Estimation



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Identifiability 101

- **Identification** considers the process of stating causal quantities in terms of statistical ones.
  Identifiability considers the question whether a causal quantity is statistically expressible i.f.p.

- In **Markovian SCM**, given the graph G and observational distribution $p(\mathbf{V})$
  any interventional quantity is computable via the *adjustment formula*.

  Unfortunately, Markovianity is too strict.
  I.e., we practically **never** observe all relevant factors about the phenomenon under investigation.

Neural Identifiability

- For neuro-causality, an equivalence in identifiability need be established first.

  For SCM, there won't exist two $(\mathcal{L}_1; G)$-agreeing descriptions that disagree on $\mathcal{L}_2$.

  For NCM, this should occur by definition.

Definition 6 ***(Neural Identifiability.)*** *Again, let Ω, Υ be the sets of SCMs and corresponding $\mathfrak{G}$-GNN based iVGAE. For any pair $(\mathfrak{C}, \mathcal{V}) \in \Omega \times \Upsilon$, a causal effect $p(\mathbf{V}_i | \, do(\mathbf{V}_j))$ is called neurally identifiable iff the pair agrees on both the causal effect [$p^{\mathfrak{C}}(\mathbf{V}_i | \, do(\mathbf{V}_j)) = p^{\mathcal{V}}(\mathbf{V}_i | \, do(\mathbf{V}_j))$] and the observational distributions [$p^{\mathfrak{C}}(\mathbf{V}) = p^{\mathcal{V}}(\mathbf{V})$].*

TECHNISCHE UNIVERSITÄT DARMSTADT

NI is preserved for iVGAE

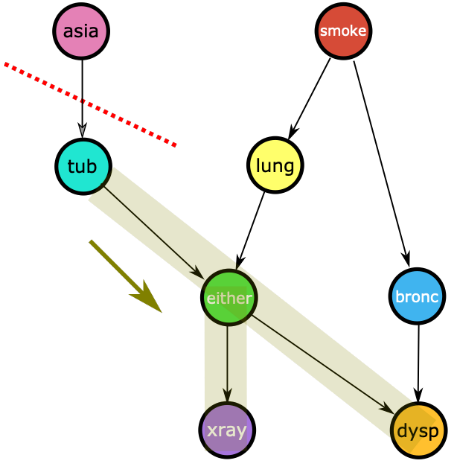
- Pearl's *do*-calculus provides us with *any* identification.

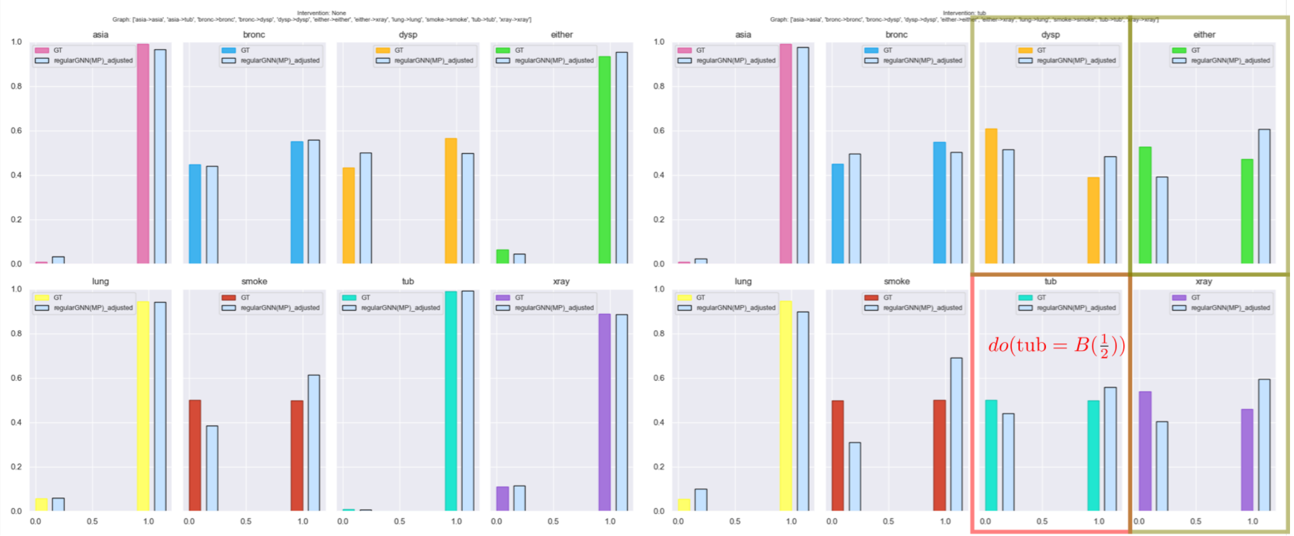
  Xia et al., and we, prove that this is also **achievable with neuro-causality**.

Theorem 4 ***(Dual Identification.)*** *Consider the causal quantity of interest* $Q = p^{\mathfrak{C}}(\mathbf{V}_i |\, do(\mathbf{V}_j))$, $\mathfrak{G}$ *the true causal graph underlying SCM* $\mathfrak{C} \in \Omega$ *and* $p(\mathbf{V})$ *the observational distribution.* Q *is neurally identifiable from iVGAE* $\mathcal{V} \in \Upsilon$ *with* $\mathfrak{G}$*-GNN modules iff* Q *is identifiable from* $\mathfrak{G}$ *and* $p(\mathbf{V})$.

TECHNISCHE UNIVERSITÄT DARMSTADT

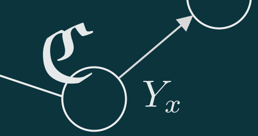
Estimation

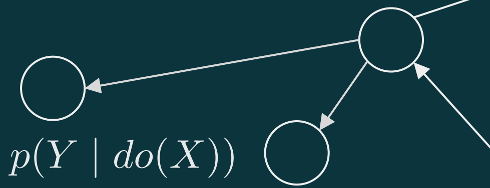




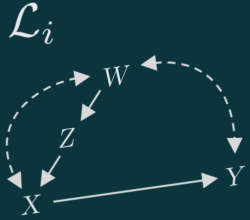
S. Lauritzen, D. Spiegelhalter, JRSS, 1988.

TECHNISCHE UNIVERSITÄT DARMSTADT





6 Neuro-Causality Onwards



TECHNISCHE UNIVERSITÄT DARMSTADT

Recap

- We collectively as researchers are **tightening** the integration of Causality and ML.

- A theoretical milestone by Xia, Lee, Bengio, and Bareinboim on **Neural Causal Models** (NCM).

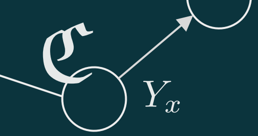
- Our work extended the lore with three new more neuro-causality approaches
  iGNN/iVGAE (Def.1/5), **SCM-GNN** (Thm.1), and **NCM-Type 2** (Cor.1),
  and results on feasibility, expressivity, and identifiability (Thms.3,4; Cors.3,4).

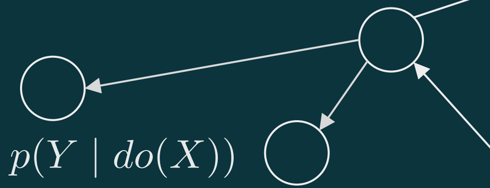
Onwards

- A different formulation of the *intervention within GNN* (Def.1) or
  of *generative* models based on iGNN (Def.5)
  could possibly allow for the complete PCH or more efficacy.

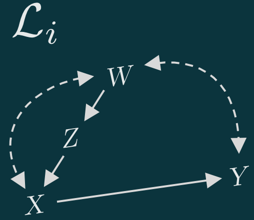
- Practical scaling and deployment in downstream-tasks.

- Exploiting causal properties for unresolved mysteries of Deep Learning, e.g., interpretability.

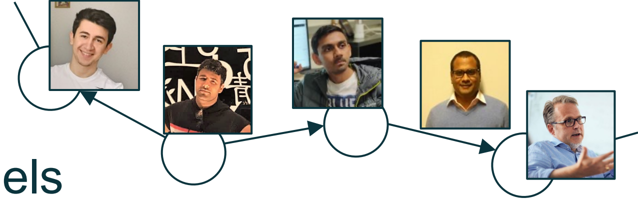




Z | Causality for Machine Learning



TECHNISCHE
UNIVERSITÄT
DARMSTADT

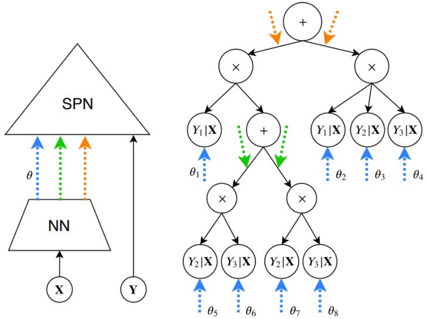


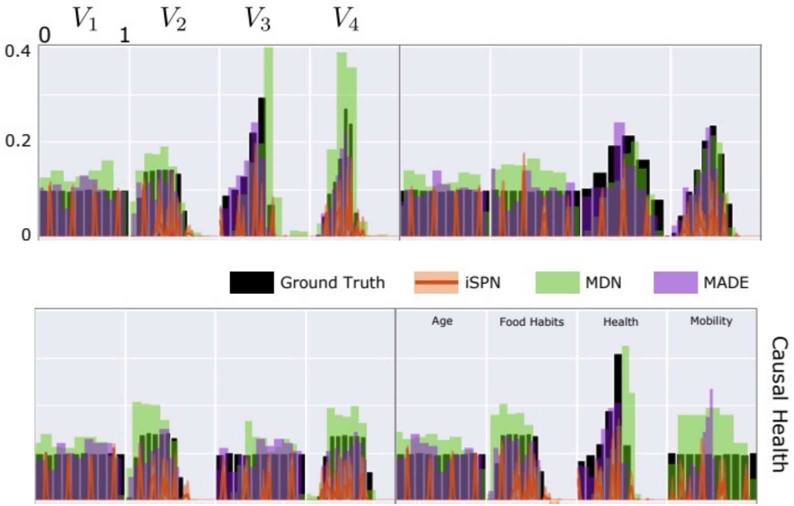
Interventional Sum-Product Networks:
Causal Inference with Tractable Probabilistic Models

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

Accepted to 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.*

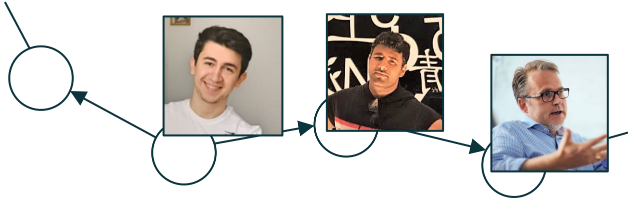


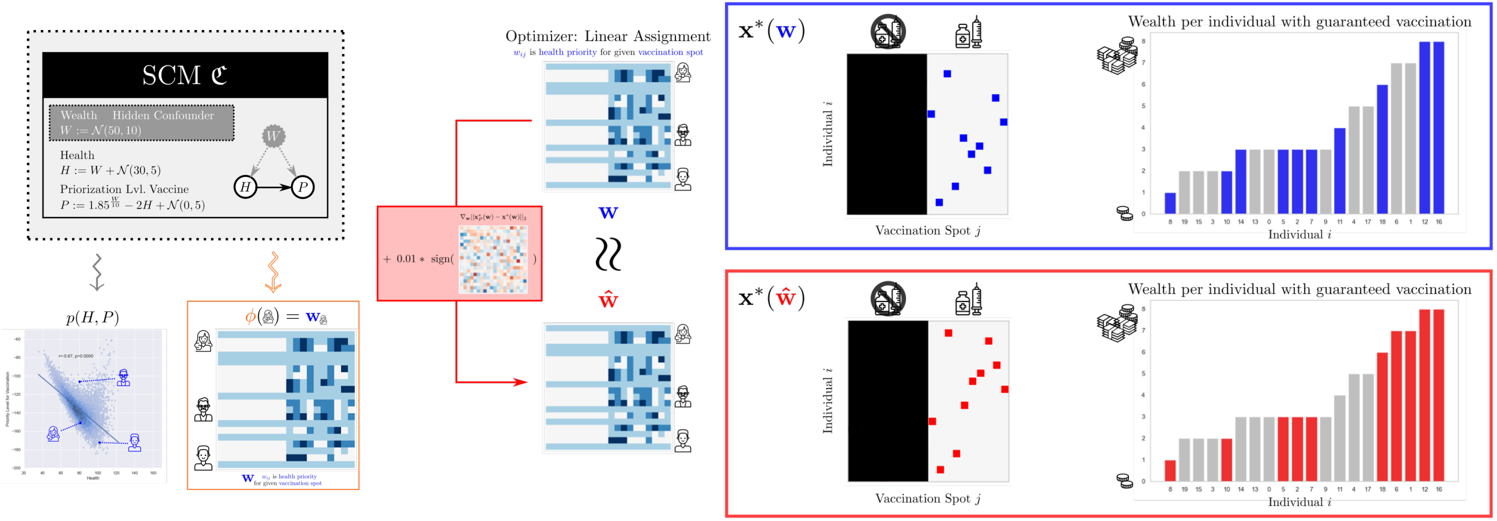


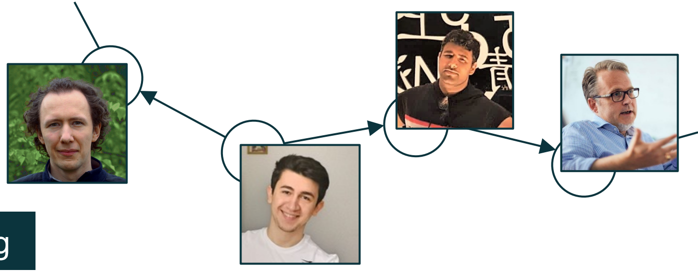
Intriguing Parameters of Structural Causal Models

Matej Zečević, Devendra Singh Dhami, Kristian Kersting

arxiv:2105.12697



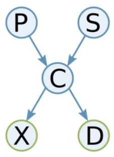


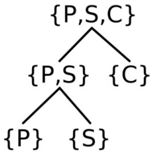


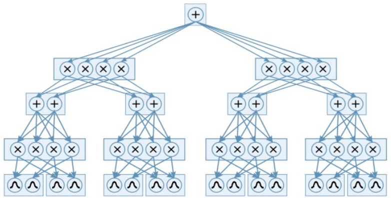
The Causal Loss: Driving Correlation to Imply Causation

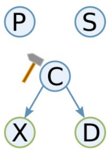
Moritz Willig, Matej Zečević, Devendra Singh Dhami, Kristian Kersting

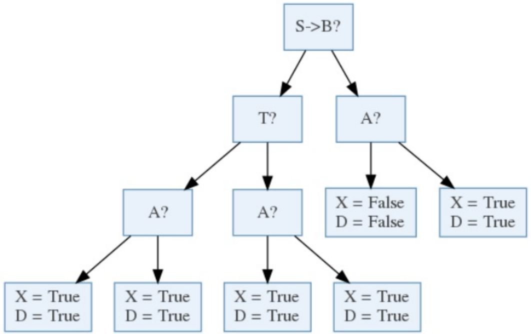
arxiv: 2110.12052

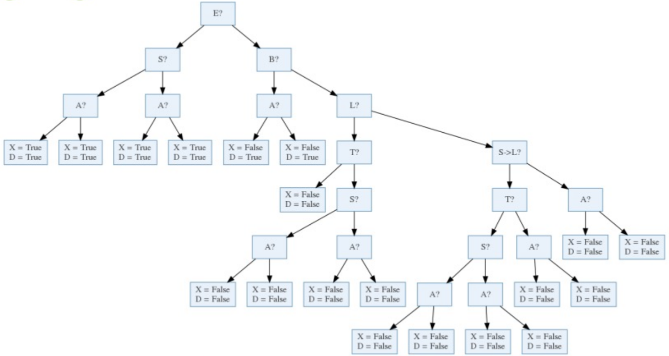


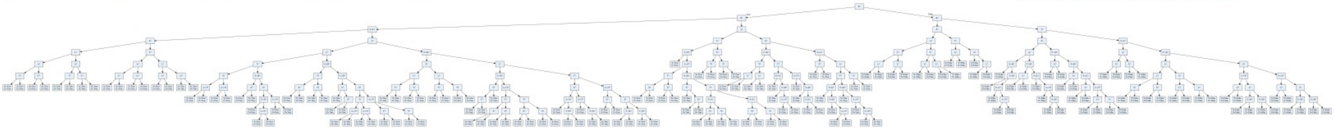




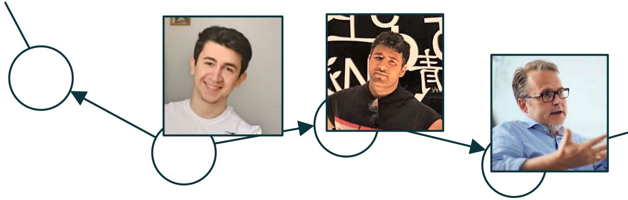








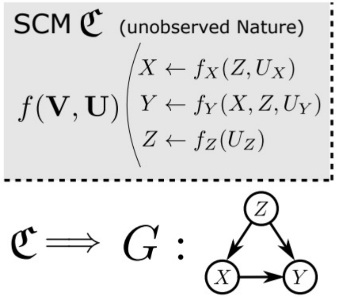
TECHNISCHE UNIVERSITÄT DARMSTADT

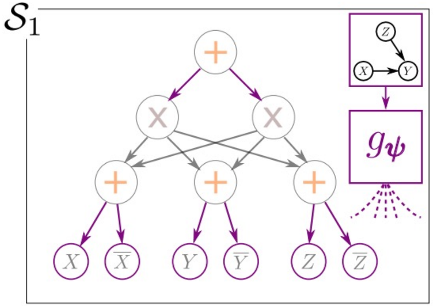


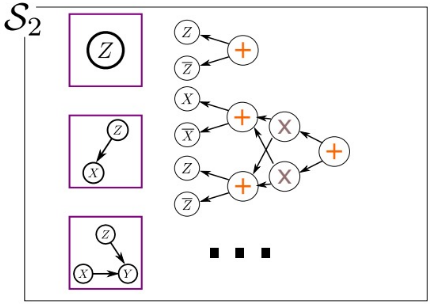
On the *Tractability* of Neural-Causal Inference

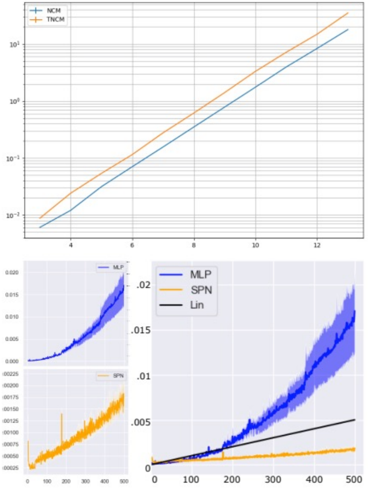
Matej Zečević, Devendra Singh Dhami, Kristian Kersting

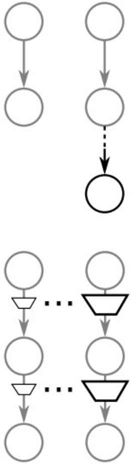
arxiv: 2110.12052









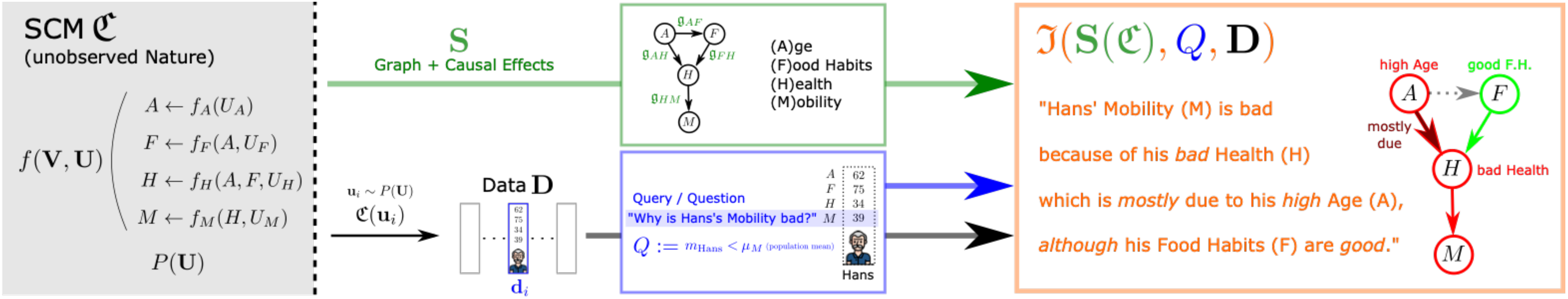


TECHNISCHE UNIVERSITÄT DARMSTADT

Structural Causal Interpretation Theorem

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

arxiv: 2110.02395



TECHNISCHE UNIVERSITÄT DARMSTADT

"As X-rays are to the surgeon, <u>graphs</u> are for causation."

-Judea Pearl in Causality (2009)

"As graphs are to the causality, causal nets are for AI."

"As X-rays are to the surgeon, graphs are for causation."

-Judea Pearl in Causality (2009)

Thank you!

Questions?

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 and AIML@TU Darmstadt