Geometry and Algebra of Complex Causal Networks

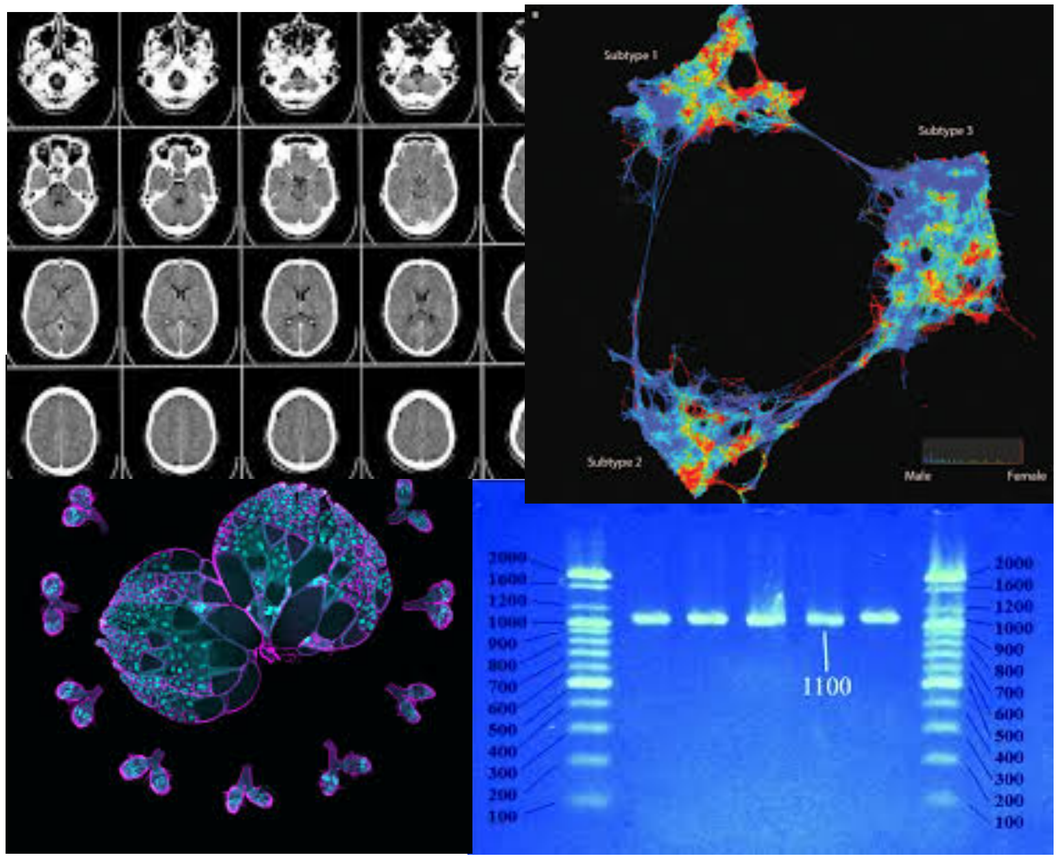
AIToGeLiS Meeting
MPI-CBG, Dresden

Liam Solus, KTH Royal Institute of Technology

Two major breakthroughs of the 21st century resulted in unprecedented amounts of data, sparking an industry shift to *data-driven techniques*.

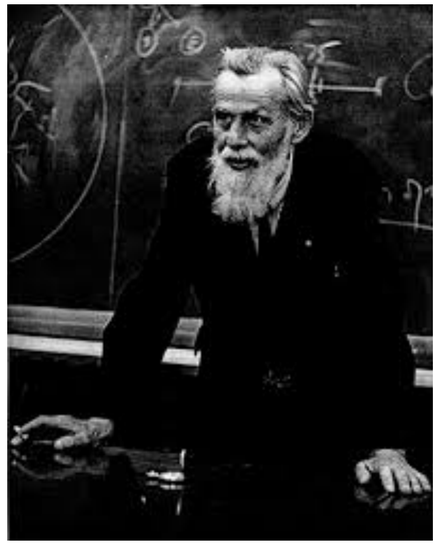
The Web 2.0 Revolution

Biological and Medical Tech



New capabilities to make predictions via old ideas.

- **Neural Nets:** Pitts and McCulloch, A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics* 5 (1943): 115-133.
- **Randomization:** Pierce and Jastrow, (1885). On Small Differences in Sensation. *Memoirs of the National Academy of Sciences*. 3 (1885): 73–83.
- **Randomized Controlled Trials:** Fisher, *The design of experiments,* (1935).
- **Hypothesis Testing:** Pearson, *Statistical Hypothesis Testing,* (1900).



Warren McCulloch

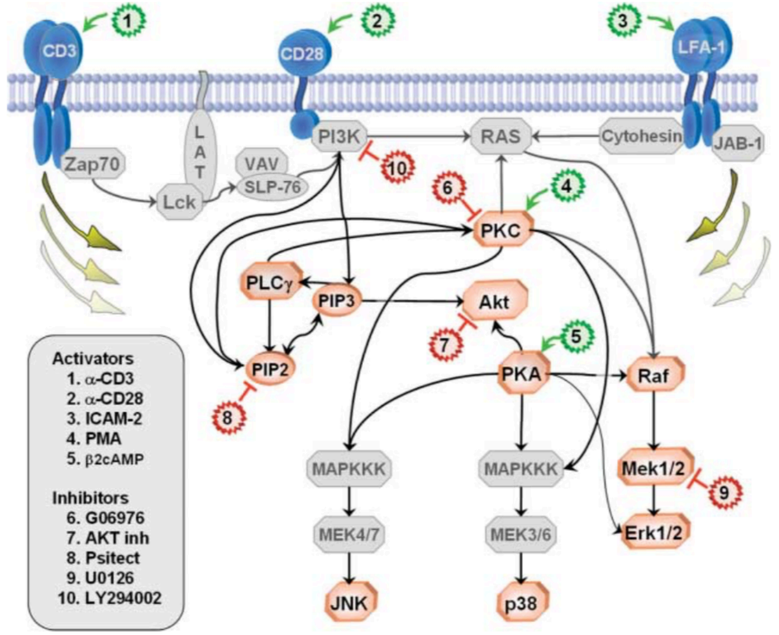
Charles Pierce

Ronald Fisher

Karl Pearson

Different goals for internet tech and bio-tech.

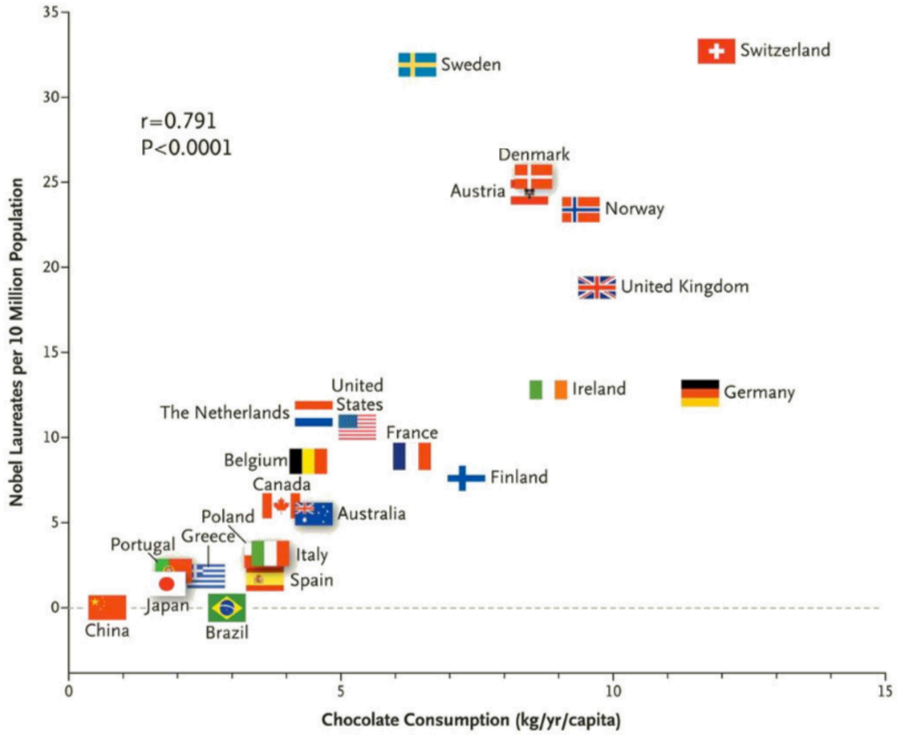
Internet tech focuses mainly on prediction accuracy



Biological Tech focuses on discovering laws and mechanisms that lead to an outcome

A probabilistic prediction:

From a list of possible candidates for a Nobel Prize, I should predict the winner as the person who eats the most chocolate.*



There is a significant positive correlation between chocolate consumption (per capita) and the number of Nobel Laureates (per capita).

-Messerli, 2012

A standard ML model may learn this pattern and use it to make such a prediction.

*A deeper description of this example can be found in *Elements of Causal Inference by Peters, Janzing and Schölkopf (2017).*

Recent perspectives in AI/ML embrace the biological perspective.

Knowledge of the underlying causal laws and mechanisms can help produce better predictions.

Rapid expansion of the ML subfield known as causality.

- Applications in both internet tech and biomedical fields
- A fundamental goal is to discover causal relations in complex systems of jointly distributed random variables

The problem of Causal Discovery. *Given data, can we discover the causal relations that guide the data-generating process?*

Representing causal relations:

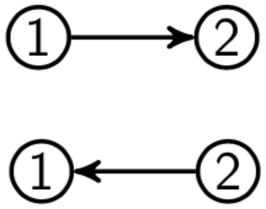
A structural equation model (SEM):

$$X_1 = N_1$$
$$X_2 = 4X_1 + N_2$$

$N_i \sim N(0,1)$ and independent

Chain rule:

$$f_{\mathbf{X}}(x_1, x_2) = f_{X_1}(x_1) f_{X_2|X_1}(x_2 \mid x_1)$$
$$= f_{X_2}(x_2) f_{X_1|X_2}(x_1 \mid x_2)$$



An intervention:

Perturb X_1: $N_1 \sim N(0,3)$

Changing the distribution of X_1 changes the distribution of X_2:

$$X_2 \sim N(0,49)$$

Changing the distribution of $X_2 \mid X_1$ **does not** change the distribution of X_1:

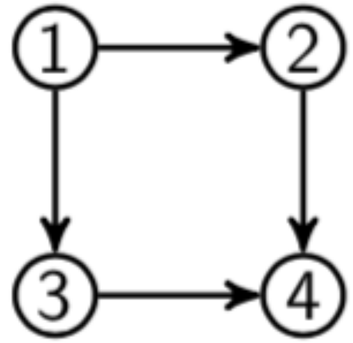
$$N_2 \sim N(0,3)$$

$$X_1 = N_1 \sim N(0,1)$$
$$X_2 = 4X_1 + N_2$$

Representing causal relations:

$[p] := \{1, \ldots, p\}$

$G = ([p], E)$ a **directed acyclic graph (DAG)**



$\mathbf{X} = [X_1, \ldots, X_p]^T$ is **Markov** to G if

$$f_{\mathbf{X}}(x_1, \ldots, x_p) = \prod_{i \in [p]} f_{X_i | \mathbf{X}_{pa_G(i)}}(x_i \,|\, \mathbf{x}_{pa_G(i)})$$

where $\text{pa}_G(i) = \{k \in [p] : k \to i \in E\}$.

The **DAG model** for G:

$$\mathscr{M}(G) = \{\mathbf{X} : \mathbf{X} \text{ Markov to } G\}.$$

Given random sample from $\mathbf{X}$ *can we learn* G*?*
(No, but some causal relations)

Theorem.

$\mathscr{M}(G) = \{\mathbf{X} : \mathbf{X}$ satisfies Global MP w.r.t. $G\}$

.

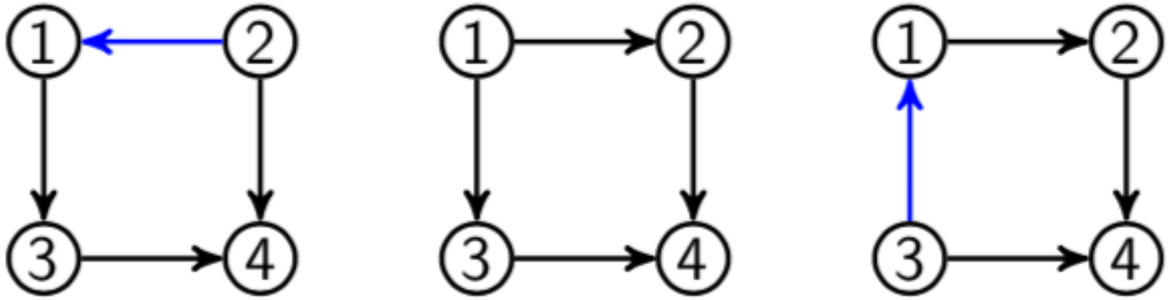
The **Global Markov Property** w.r.t. G**:**

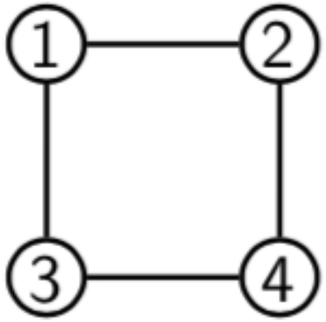
$\mathbf{X}_A \perp \mathbf{X}_B \,|\, \mathbf{X}_C$ whenever A and B are **d-separated** given C in G.

Representing causal relations:

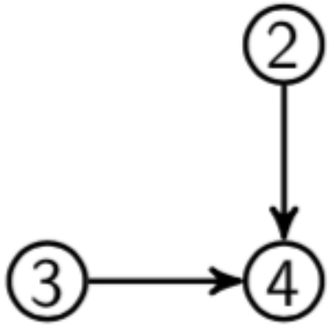
Theorem (Verma, Pearl, 1989). $\mathscr{M}(G) = \mathscr{M}(H)$ if and only if G and H have the same **skeleton** and **v-structures**.

Markov Equivalence Class (MEC):





skeleton

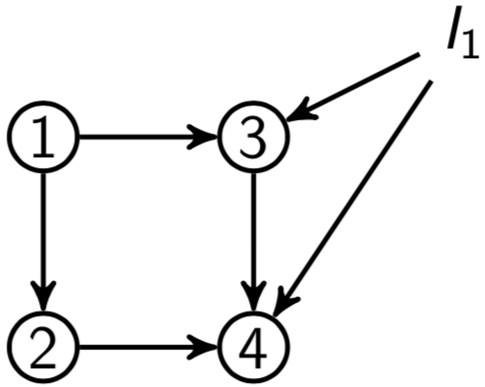


v-structure

With Experimental Data:

$\mathcal{I} = \{I_0 = \emptyset, I_1, \ldots, I_K\}, I_k \subseteq [p]$ **intervention targets**

$\mathcal{I} = \{I_0 = \emptyset, I_1 = \{3,4\}\}$

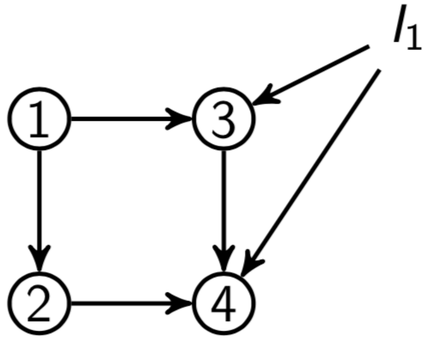


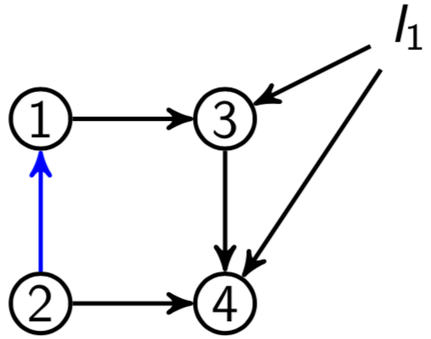
$(f^{(0)}, \ldots, f^{(K)})$ = an **interventional setting**

- $f^{(k)} \in \mathcal{M}(G)$ for all k
- $f^{(k)}(x_i \mid \mathbf{x}_{\text{pa}_G(i)}) = f^{(0)}(x_i \mid \mathbf{x}_{\text{pa}_G(i)})$ for all $i \notin I_k$

$\mathcal{M}(G, \mathcal{I}) = \{(f^{(0)}, \ldots, f^{(K)})$ for G and $\mathcal{I}\}$

Theorem (Yang et al., 2018). $\mathcal{M}(G, \mathcal{I}) = \mathcal{M}(H, \mathcal{I})$ if and only if $G^{\mathcal{I}}$ and $H^{\mathcal{I}}$ have the same **skeleton** and **v-structures**.





Causal discovery algorithms

For learning MECs $\mathscr{M}(G)$:

- **PC algorithm** (Glymour, Spirtes, 1993)
- **GES** (Chickering, 2001)
- **Imset LinOpt** (Studeny, 2006)
- **GreedySP** (LS, Wang, Uhler, 2021)
  - **GrASP** (Lam et al., 2023)
  - **BOSS** (Andrews et al., 2023)

For learning I-MECs $\mathscr{M}(G, \mathscr{I})$:

- **IFCI algorithm** (Kocaoglu et al., 2019)
- **GIES** (Hauser, Bühlmann, 2012)
- **QIGTreeLearn** (Hollering, LS, Johnson, 2024)
- **IGSP** (Wang, LS, Yang, Uhler, 2017)

Powered by polyhedral geometry

Best performers on the benchmarking platform for causal discovery methods:
Benchpress (Rios, Kuipers, Moffa, 2022)

Question. *Can we identify more (or all) of the edges of the causal DAG without collecting (expensive) experimental data?*

Idea (SEMs). Use additional assumptions on the structural equations*:*

- **LiNGAM models** (Shimizu et al., 2006):

  $X_i = \sum_{k \in \text{pa}_G(i)} \lambda_{ki} X_k + N_i$ with N_i non-Gaussian.

- **Equal variances** (Peters and Bühlmann, 2014):

  $X_i = \sum_{k \in \text{pa}_G(i)} \lambda_{ki} X_k + N_i$ with $N_i \sim N(0, \omega)$ for all i.

Idea (discrete data). Use observable ***context-specific CI (CSI) relations****:*

$$X_A \perp X_B \,|\, X_C, X_D = x_D$$

(Tikka et al., 2019)

Carrier $\perp X_1, \ldots, X_m$ *| Exposed = No*

"Whether or not a child is a carrier of chicken pox is independent of all other background factors given that they haven't been exposed."

Discrete Data: CStree models.

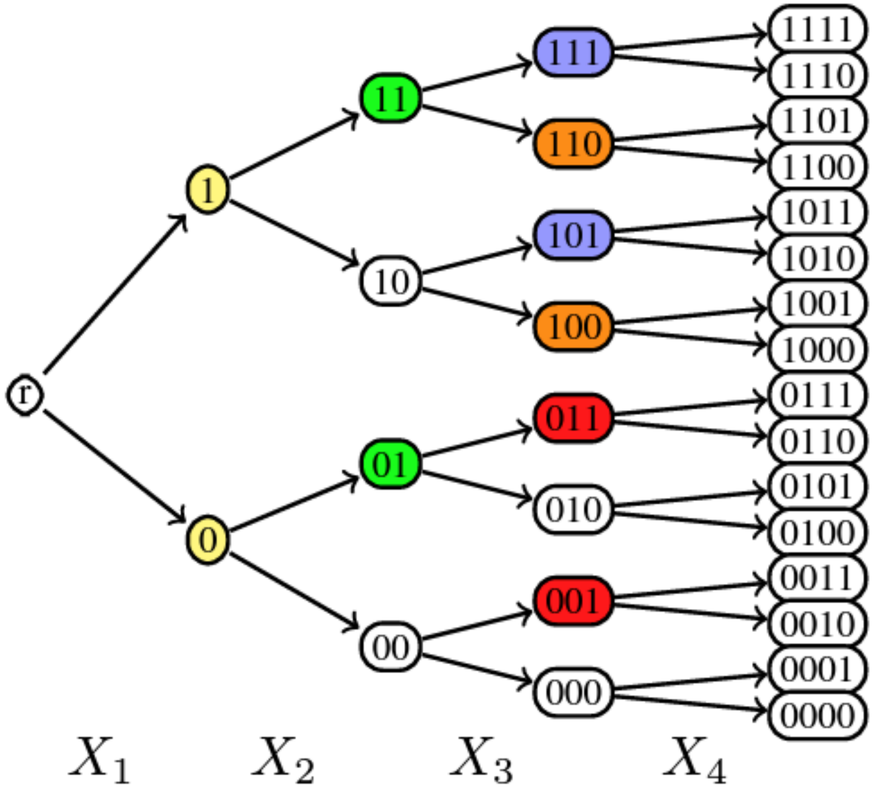
Building a **CStree model** (Duarte, LS, 2021):

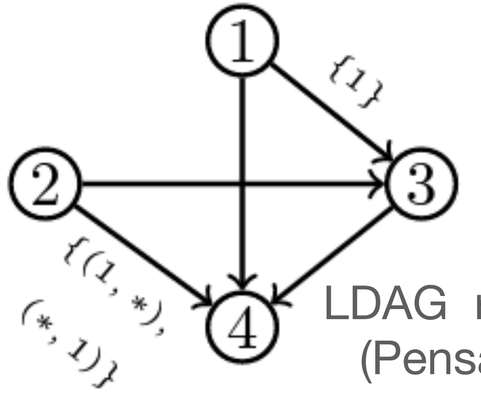
- variable ordering: $\pi = \pi_1 \cdots \pi_p$
- relations $\mathscr{C}_{\pi,i} = \{X_{\pi_i} \perp \mathbf{X}_{[\pi_1:\pi_{i-1}]\setminus S} \,|\, \mathbf{X}_S = \mathbf{x}_S\}$ such that the sets

  $$S_{\pi,i}(\mathbf{x}_S) = \{\mathbf{x}_{[\pi_1:\pi_{i-1}]} \text{ that agrees with } \mathbf{x}_S\}$$

  partition the joint state space of $X_{\pi_1}, \ldots, X_{\pi_{i-1}}$.
- $\mathbf{s}$ = union of all $S_{\pi,i}(\mathbf{x}_S)$
- $\mathscr{T} = (\pi, \mathbf{s})$ a **CStree**
- $\text{pa}_{\mathscr{T}}(\mathbf{x}_{[\pi_1:\pi_{i-1}]}) = S$ for $\mathbf{x}_{[\pi_1:\pi_{i-1}]} \in S_{\pi,i}(\mathbf{x}_S)$
- The **CStree Model** for $\mathscr{T}$ is

$$\mathscr{M}(\mathscr{T}) = \left\{ \mathbf{X} : f_{\mathbf{X}}(\mathbf{x}) = \prod_{i \in [p]} f(x_{\pi_i} \,|\, \mathbf{x}_{\text{pa}_{\mathscr{T}}(\mathbf{x}_{\pi_1:\pi_{i-1}})}) \right\}.$$



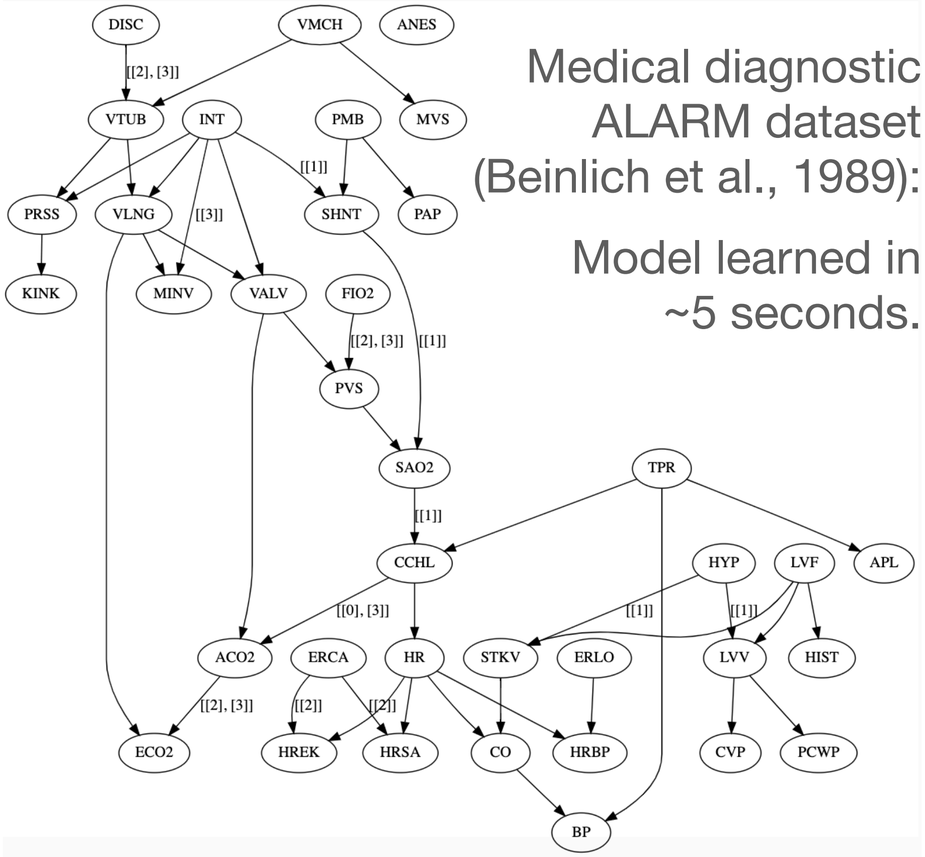


LDAG representation (Pensar et al., 2015)

Scalable Learning of CStree models.

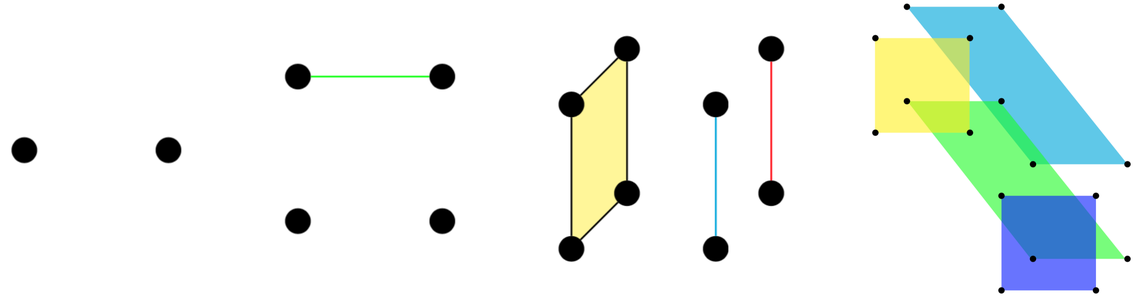
CStree algorithm (Rios, Markham, LS, 2024):

- bound the size of the sets S defining the contexts: $|S| \leq \beta$.
- Requires enumeration of CStrees with $|S| \leq \beta$, which solves a case of a family of problems proposed by (Alon, Balogh, 2023).
- **Theorem.** There are $1 - \binom{i}{2} + \sum_{k=1}^{i} i^{d_k}$ stagings of level i satisfying $|S| \leq 2$.
- **Theorem.** Local computations time complexity $\mathcal{O}(p2^m |\mathcal{S}_{m,\beta}| d^{\beta})$



Medical diagnostic ALARM dataset (Beinlich et al., 1989):

Model learned in ~5 seconds.



Gaussian data: Colored DAG models

- $G = ([p], E)$ a DAG and $\mathbf{X} = [X_1, \ldots X_p]^T$ where

$$X_i = \sum_{k \in \mathrm{pa}_G(i)} \lambda_{ki} X_k + N_k,$$

$N_i \sim N(0, \omega_i)$ independent and $\lambda_{ki} = 0$ if $k \to i \notin E$

- $\Lambda = [\lambda_{ki}] \in \mathbb{R}^{p \times p}$ and $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_p) \in \mathbb{R}^{p \times p}_{>0}$.
- The **Gaussian DAG model** for G is:

$$\mathscr{M}(G) = \{\Sigma \in \mathrm{PD}^{p \times p} : \Sigma = (\mathbf{1} - \Lambda)^{-T} \Omega (\mathbf{1} - \Lambda)^{-1}\}.$$

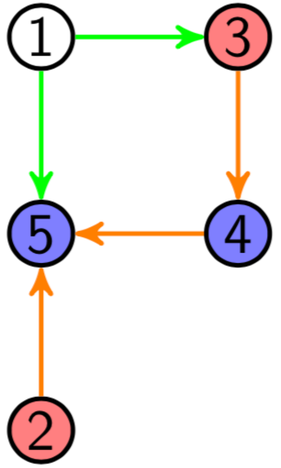
- **Partial homogeneity constraints:**
  - **vertex coloring:**

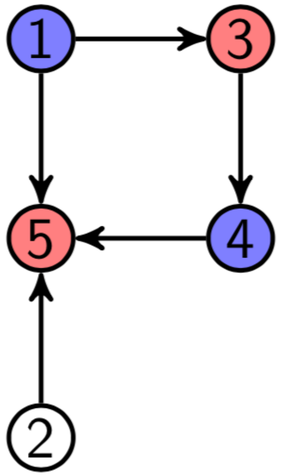
$$c : [p] \longrightarrow [d_V]; \quad \omega_i = \omega_k \Longrightarrow c(i) = c(k)$$

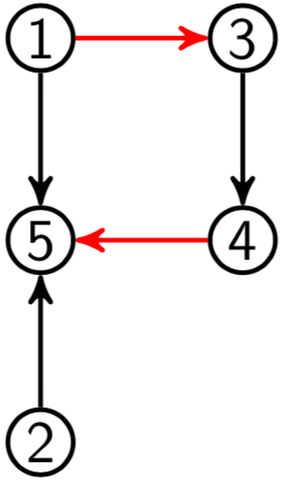
  - **edge coloring:**

$$c : E \longrightarrow [d_E]; \quad \lambda_{ji} = \lambda_{\ell k} \Longrightarrow c(ij) = c(k\ell)$$

Colored DAG models $\mathscr{M}(G, c)$:







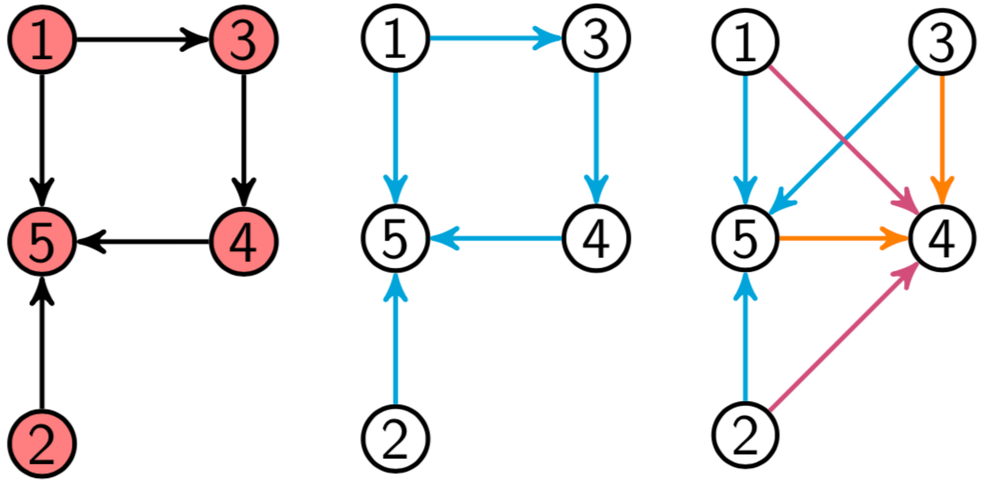
Structural identifiability results.

- **Theorem (Peters, Bühlmann, 2012).** Vertex-colored DAGs with a single color have model equivalence classes of size 1; i.e., they are **structurally identifiable**.

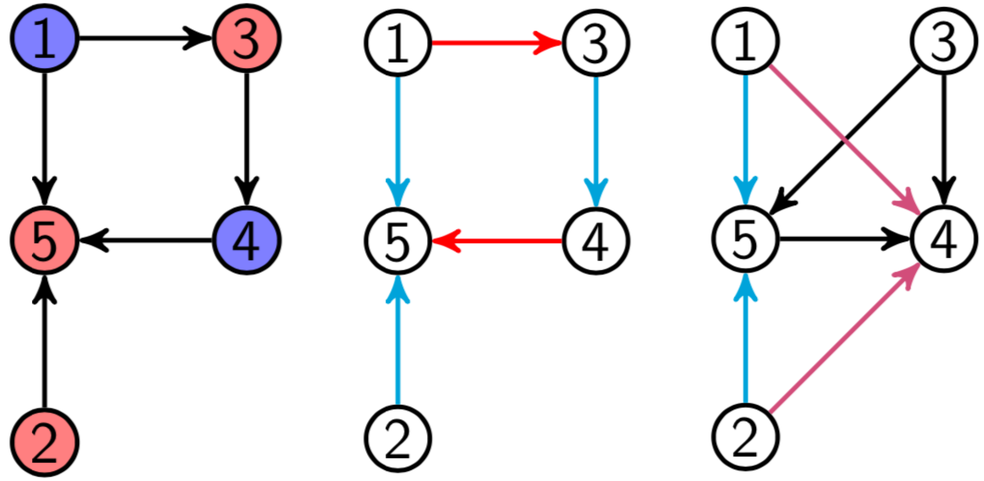
- **Theorem (Wu and Drton, 2023).** Characterization of model equivalence classes of vertex-colored models.

- **Theorem (Boege, Kubjas, Misra, LS, 2024).** Edge-colored DAGs with a single edge color are structurally identifiable.

- **Theorem (Boege, Kubjas, Misra, LS, 2024).** **BPEC-DAGs** are structurally identifiable.



Structurally Identifiable



MEC characterization

MEC characterization ?

Structurally Identifiable ?

Structural identifiability results.

Proof idea:

The kernel $\ker(\phi^*_{G,c})$ of the pullback

$$\phi^*_{G,c} : \mathbb{C}[\Sigma] \longrightarrow \mathbb{C}[\Lambda, \Omega]$$

of the parametrization map

$$\phi_{G,c} : (\Lambda, \Omega) \longmapsto (1-\Lambda)^{-T}\Omega(1-\Lambda)^{-1}$$

is the set of all polynomials vanishing on the model $\mathscr{M}(G, c)$.

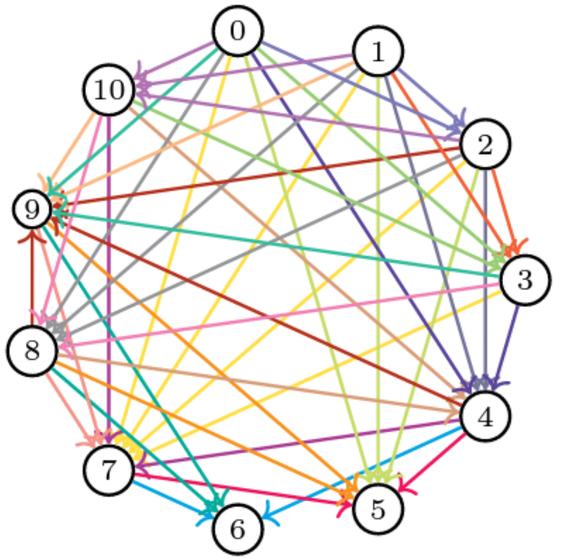
If $c(ij) = c(k\ell)$, $\ker(\phi^*_{G,c})$ contains a polynomial

$$|\Sigma_{j\cup \mathrm{pa}_G(j)\setminus i}||\Sigma_{\mathrm{pa}_G(\ell)}| - |\Sigma_{\ell\cup \mathrm{pa}_G(\ell)\setminus k}||\Sigma_{\mathrm{pa}_G(j)}|$$

Given (H, c'), we show any minimal generating set of $\ker(\phi^*_{G,c})$ cannot generate $\ker(\phi^*_{H,c'})$.

A causal discovery algorithm:

- **Greedy Edge-Colored Search (GECS):**
  - edge-colored extension of GES
  - currently learns BPEC-DAGs
  - github.com/soluslab/coloredDAGs



- 0 = fixed acidity
- 1 = volatile acidity
- 2 = citric acid
- 4 = chlorides
- 7 = density

BPEC-DAG representation of the causal relations between 11 different biochemical properties relevant in white wine quality.

Questions.

Exercises:

1. For the SEM $X_3 = \lambda_{13}X_1 + \lambda_{23}X_2 + N_3, X_1 \sim N(0,\omega_1), X_1 \sim N(0,\omega_2), N_3 \sim N(0,\omega_3)$ for $G = 1 \to 3 \leftarrow 2$:
    1. convince yourself that G is identifiable via Markov equivalence.
    2. intervene at X_3, and convince yourself that the edges of the v-structure are causal.
2. Convince yourself that $H = 1 \to 2 \to 3$ is not identifiable, but it is identifiable if you intervene on X_2.
3. Convince yourself that $G = 1 \to 2$ is identifiable when we assume a Gaussian model with nodes 1 and 2 having the same color.
4. Draw the staged tree and LDAG representations of all CStree models on 3 binary variables. For each tree, associate the variables to some events so that the context-specific relations make sense to you.

Open Questions:

5. Enumerate the ways to partition the d-dimensional cube $[0, 1]^d$ into non-overlapping faces of co-dimension at most 3.
6. Give an algebraic proof of the result of Peters and Bühlmann.

Considerations for applications:

7. Think of some data sets where there may be clustering of direct causal relations.
8. Think of some data sets that may naturally contain context-specific CI relations.

Thank you for listening!

- Boege, Kubjas, Misra, and LS. *Colored Gaussian DAG models. arXiv: 2404.04024* (2024).
- Hollering, Johnson and LS. *Hyperplane representations of interventional characteristic imset polytopes.* arXiv: 2404.18500 (2024).
- Rios, Markham, and LS. *Scalable structure learning for sparse context-specific causal models.* arXiv: 2402.07762 (2024).
- Duarte and LS. *Algebraic geometry of discrete interventional models.* To appear in the EMS special issue on Varieties, Polyhedra and Computation (2024).
- Duarte and LS. *A new characterization of discrete decomposable models.* Proceedings of the American Mathematical Society (2023).
- Linusson, Restadh, and LS. *Greedy causal discovery is geometric.* SIAM Journal on Discrete Mathematics (2023).
- Linusson, Restadh, and LS. *On the edges of characteristic imset polytopes.* Submitted (2023).
- Duarte and LS. *Representation of context-specific causal models with observational and interventional data. Submitted (2022).*
- LS, Wang, and Uhler. *Consistency guarantees for greedy permutation-based causal inference algorithms.* Biometrika (2021).
- Wang, Uhler, and LS. *Permutation-based causal inference algorithms with interventions.* The Proceedings of Neural Information Processing Systems (NeurIPS) (2017).

digital futures

Vetenskapsrådet

WASP | WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM