From Propositional to Lifted Causal Inference

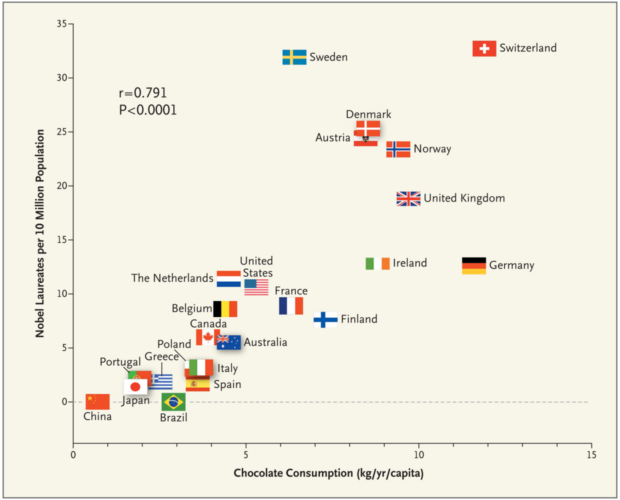
Research Seminar – Institute for Data Science Foundations, Hamburg University of Technology (TUHH)

Malte Luttermann

German Research Center for Artificial Intelligence (DFKI)

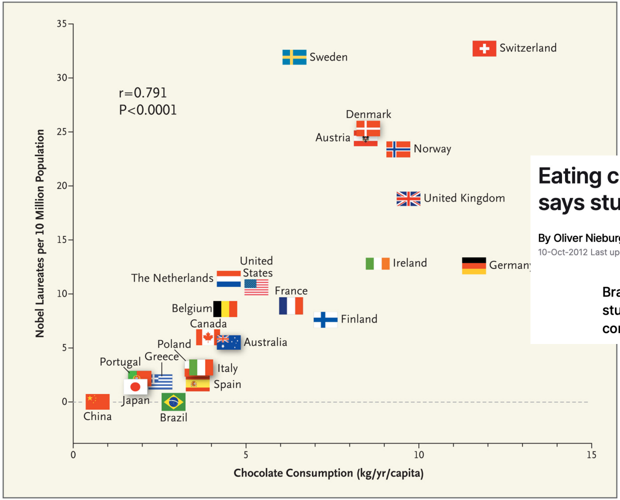
March 5, 2025

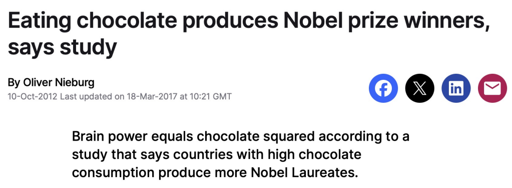
Why We Should Care About Causality



Franz H. Messerli (2012). »Chocolate Consumption, Cognitive Function, and Nobel Laureates«. *New England Journal of Medicine* 367, pp. 1562–1564.

Why We Should Care About Causality





Eating chocolate produces Nobel prize winners, says study

By Oliver Nieburg
10-Oct-2012 Last updated on 18-Mar-2017 at 10:21 GMT

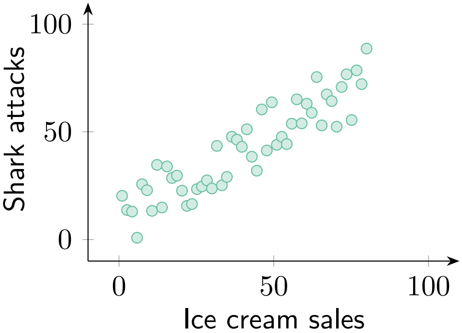
Brain power equals chocolate squared according to a study that says countries with high chocolate consumption produce more Nobel Laureates.

Franz H. Messerli (2012). »Chocolate Consumption, Cognitive Function, and Nobel Laureates«. *New England Journal of Medicine* 367, pp. 1562–1564.

Why We Should Care About Causality

An Ice Cream Example

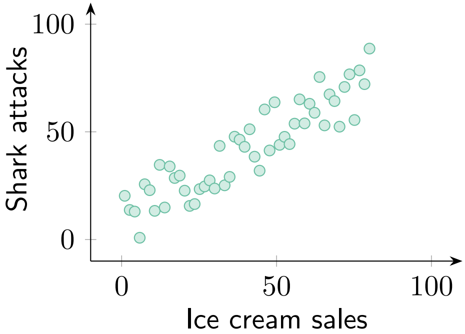
- Correlation $\neq$ causation



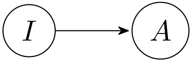
Why We Should Care About Causality

An Ice Cream Example

- Correlation $\neq$ causation



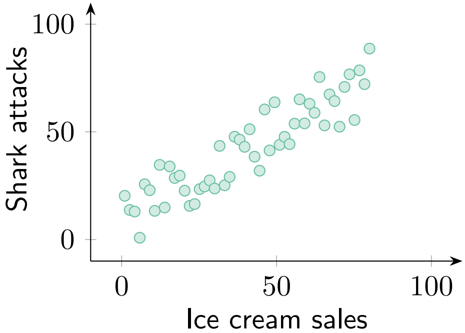
- Possible causal explanations:

1. 

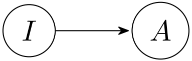
Why We Should Care About Causality

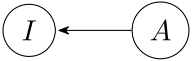
An Ice Cream Example

- Correlation $\neq$ causation



- Possible causal explanations:

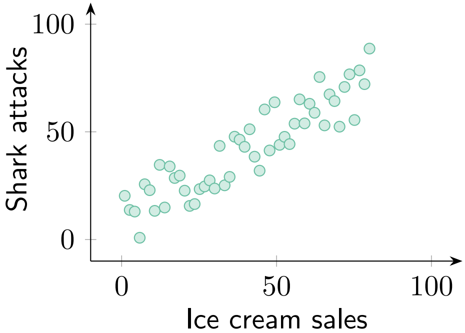
1. 

2. 

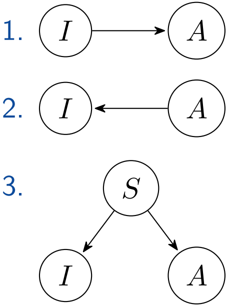
Why We Should Care About Causality

An Ice Cream Example

- Correlation $\neq$ causation

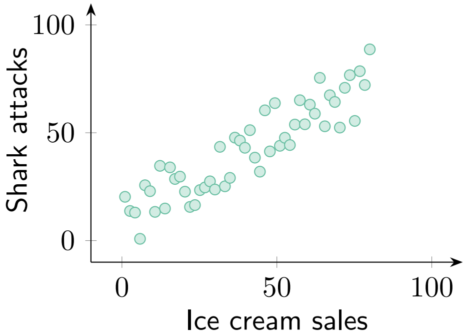


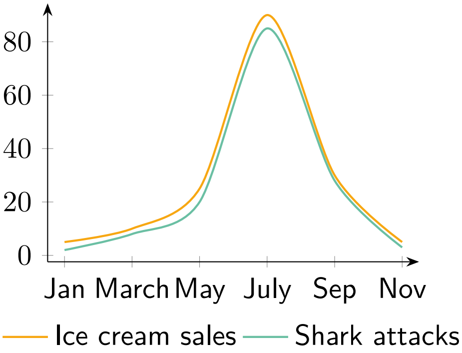
- Possible causal explanations:



Why We Should Care About Causality

Explanation of the Ice Cream Example Data

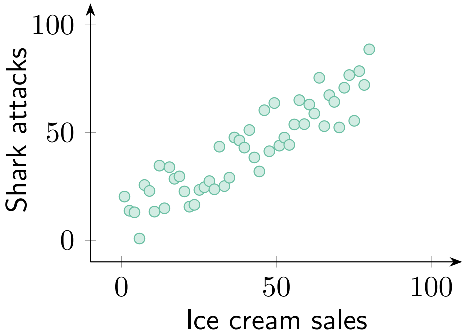




Why We Should Care About Causality

Learnings from the Ice Cream Example

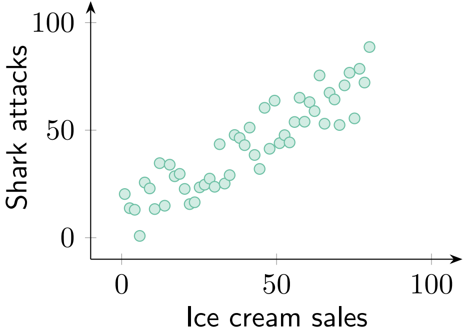
- For *prediction*, correlation is sufficient



Why We Should Care About Causality

Learnings from the Ice Cream Example

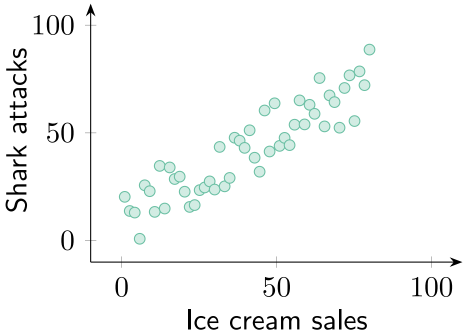
- For *prediction*, correlation is sufficient
  - E.g., knowing ice cream sales suffices to predict shark attacks



Why We Should Care About Causality

Learnings from the Ice Cream Example

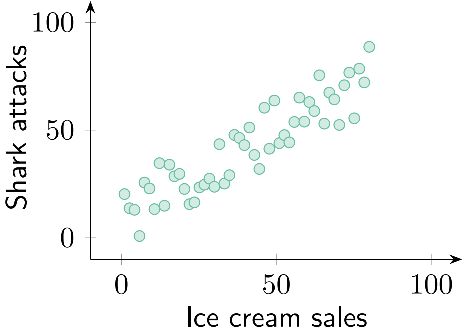
- For *prediction*, correlation is sufficient
  - E.g., knowing ice cream sales suffices to predict shark attacks
- For *decision making (acting)*, causal information is required



Why We Should Care About Causality

Learnings from the Ice Cream Example

- For *prediction*, correlation is sufficient
  - E.g., knowing ice cream sales suffices to predict shark attacks
- For *decision making (acting)*, causal information is required
  - E.g., Reducing ice cream sales will *not* reduce shark attacks



Causal Models

A *causal model* consists of

1. a causal graph G, and
2. a probability distribution P.

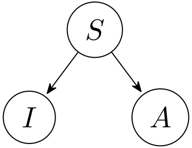
Remarks:

- G and P must be compatible (i.e., P must factorise according to G)
- More in-depth definitions possible, e.g., via a set of differential equations

Causal Models

A *causal model* consists of

1. a causal graph G, and
2. a probability distribution P.



$+$

A	I	S
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮

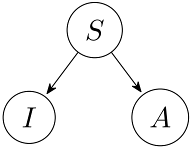
Causal Models

A *causal model* consists of

1. a causal graph G, and
2. a probability distribution P.

$P(S)$
$\vdots$

S	$P(I \mid S)$
$\vdots$	$\vdots$
$\vdots$	$\vdots$



S	$P(A \mid S)$
$\vdots$	$\vdots$
$\vdots$	$\vdots$

Causal Models

Where Does the Causal Model Come From?

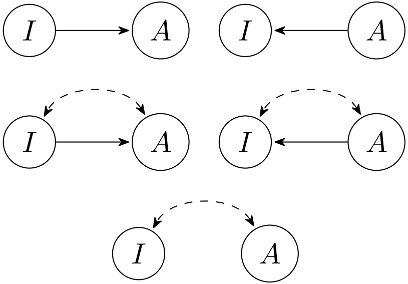
- In general, there is no unique causal graph that explains the data

Observational Data

A	I
⋮	⋮
⋮	⋮

$\longrightarrow$

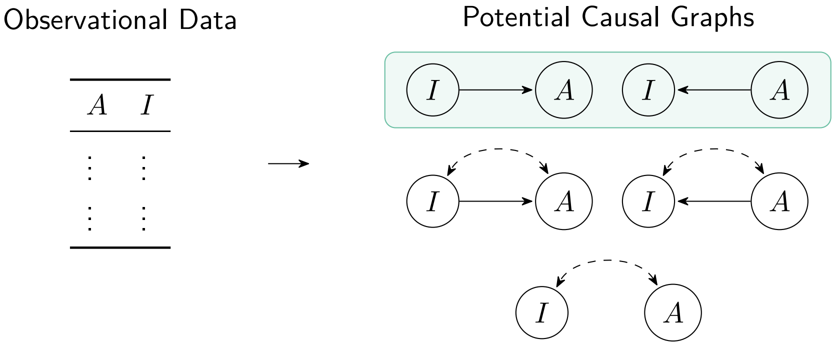
Potential Causal Graphs



Causal Models

Where Does the Causal Model Come From?

- In general, there is no unique causal graph that explains the data



- Common assumptions:
  - Causal sufficiency: No unobserved confounders
  - Acyclicity: No directed cycles

Application of Causal Models

Simpson's Paradox

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell (2016). *Causal Inference in Statistics: A Primer*. 1st. Wiley.

Application of Causal Models

Simpson's Paradox

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

The paradox:

- For men, taking the drug has a benefit
- For women, taking the drug has a benefit as well
- For all people combined, taking the drug has *no* benefit

Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell (2016). *Causal Inference in Statistics: A Primer*. 1st. Wiley.

Application of Causal Models

How to Resolve the Paradox?

- Should a person take the drug?
  - Considering the data alone is not sufficient
  - We need to understand the causal mechanisms that lead to the data

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

Application of Causal Models

How to Resolve the Paradox?

- Should a person take the drug?
  - Considering the data alone is not sufficient
  - We need to understand the causal mechanisms that lead to the data

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

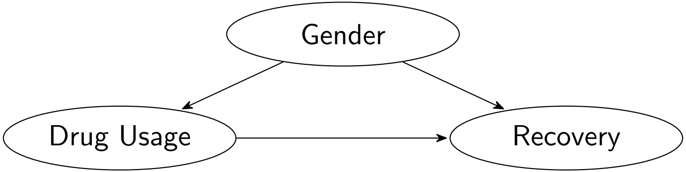
- Taking the drug has less benefit for women
- Women are more likely to take the drug than men

Application of Causal Models

How to Resolve the Paradox?

- Should a person take the drug?
  - Considering the data alone is not sufficient
  - We need to understand the causal mechanisms that lead to the data

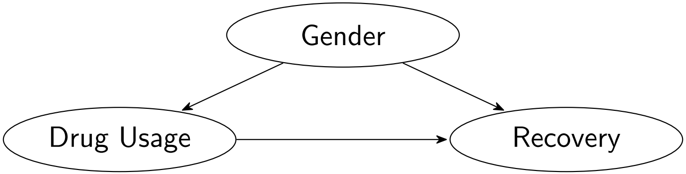
	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$



Application of Causal Models

Computing the Effect of Actions

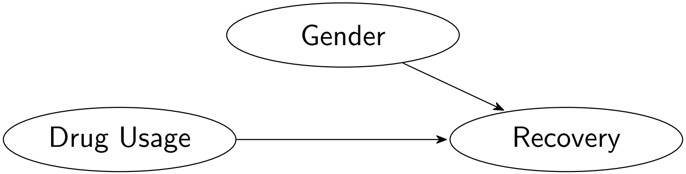
- Should a person take the drug?
  - Need to compute the causal effect of taking the drug on recovery
  - Apply the notion of an intervention $do(D = d)$



Application of Causal Models

Computing the Effect of Actions

- Should a person take the drug?
  - Need to compute the causal effect of taking the drug on recovery
  - Apply the notion of an intervention $do(D = d)$



- Average causal effect:

$$ACE = \mathbb{E}[R \mid do(D = 1)] - \mathbb{E}[R \mid do(D = 0)] = ?$$

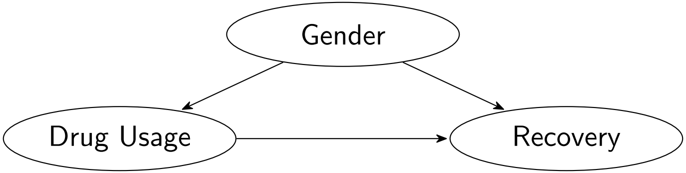
- If $ACE > 0$, taking the drug has a benefit

Interventions: Computing the Effects of Actions

- Average causal effect:

$$\begin{aligned} ACE &= \mathbb{E}[R \mid do(D=1)] - \mathbb{E}[R \mid do(D=0)] \\ &= P(R=1 \mid do(D=1)) - P(R=1 \mid do(D=0)) = ? \end{aligned}$$

- To compute the ACE, we have to compute the interventional distributions:
  - $P(R=1 \mid do(D=1)) = ?$
  - $P(R=1 \mid do(D=0)) = ?$

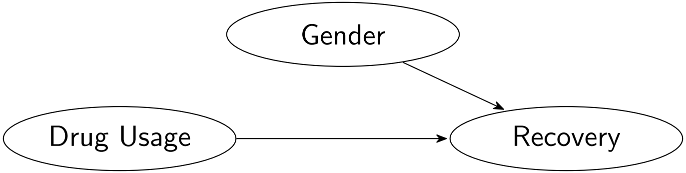


Interventions: Computing the Effects of Actions

- Average causal effect:

$$\begin{aligned} ACE &= \mathbb{E}[R \mid do(D=1)] - \mathbb{E}[R \mid do(D=0)] \\ &= P(R=1 \mid do(D=1)) - P(R=1 \mid do(D=0)) = \,? \end{aligned}$$

- To compute the ACE, we have to compute the interventional distributions:
  - $P(R=1 \mid do(D=1)) = \,?$
  - $P(R=1 \mid do(D=0)) = \,?$



- Need to remove incoming influences on D
- Need to segregate the data w.r.t. G (»adjust for G«)

Interventions: Computing the Effects of Actions

- $P(R = 1 \mid do(D = 1))$

$$= \sum_{g \in \{0,1\}} P(R = 1 \mid D = 1, G = g) P(G = g)$$

$$= P(R = 1 \mid D = 1, G = 1) P(G = 1) + P(R = 1 \mid D = 1, G = 0) P(G = 0)$$

$$= 0.93 \cdot (87 + 270) \,/\, 700 + 0.73 \cdot (263 + 80) \,/\, 700$$

$$= 0.832$$

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

Interventions: Computing the Effects of Actions

- $P(R = 1 \mid do(D = 1))$

$$= \sum_{g \in \{0,1\}} P(R = 1 \mid D = 1, G = g) P(G = g)$$
$$= P(R = 1 \mid D = 1, G = 1) P(G = 1) + P(R = 1 \mid D = 1, G = 0) P(G = 0)$$
$$= 0.93 \cdot (87 + 270) \,/\, 700 + 0.73 \cdot (263 + 80) \,/\, 700$$
$$= 0.832$$

- $P(R = 1 \mid do(D = 0)) = 0.7818$

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

Interventions: Computing the Effects of Actions

- $P(R = 1 \mid do(D = 1))$

$$= \sum_{g \in \{0,1\}} P(R = 1 \mid D = 1, G = g) P(G = g)$$
$$= P(R = 1 \mid D = 1, G = 1) P(G = 1) + P(R = 1 \mid D = 1, G = 0) P(G = 0)$$
$$= 0.93 \cdot (87 + 270) \,/\, 700 + 0.73 \cdot (263 + 80) \,/\, 700$$
$$= 0.832$$

- $P(R = 1 \mid do(D = 0)) = 0.7818$
- $ACE = 0.832 - 0.7818 = 0.0502 > 0$, i.e., taking the drug has a benefit

	Recovery rate *with* drug	Recovery rate *without* drug
Men ($357 / 700 = 0.51$)	$81 / 87 = 0.93$	$234 / 270 = 0.87$
Women ($343 / 700 = 0.49$)	$192 / 263 = 0.73$	$55 / 80 = 0.69$
Combined	$273 / 350 = 0.78$	$289 / 350 = 0.83$

Interventions: Computing the Effects of Actions

Adjustment Formula

In general:

- Given an intervention $do(X = x)$, we need to block all backdoor paths
  - A backdoor path from X to Y is a non-causal path, i.e., a path that remains after removing all outgoing edges of X

Interventions: Computing the Effects of Actions

Adjustment Formula

In general:

- Given an intervention $do(X = x)$, we need to block all backdoor paths
  - A backdoor path from X to Y is a non-causal path, i.e., a path that remains after removing all outgoing edges of X
- E.g., backdoor paths can be blocked by adjusting for the parents $Pa(X)$ of X
- Adjustment formula for parent adjustment:

$$
\begin{aligned}
&P(Y = y \mid do(X = x)) \\
&\quad = \sum_{pa(x)} P(Y = y \mid X = x, Pa(X) = pa(x)) \cdot P(Pa(X) = pa(x))
\end{aligned}
$$

Interventions: Computing the Effects of Actions

Adjustment Formula

In general:

- Given an intervention $do(X = x)$, we need to block all backdoor paths
  - A backdoor path from X to Y is a non-causal path, i.e., a path that remains after removing all outgoing edges of X
- E.g., backdoor paths can be blocked by adjusting for the parents $Pa(X)$ of X
- Adjustment formula for parent adjustment:

$$\begin{aligned} &P(Y = y \mid do(X = x)) \\ &\quad = \sum_{pa(x)} P(Y = y \mid X = x, Pa(X) = pa(x)) \cdot P(Pa(X) = pa(x)) \end{aligned}$$

Note:

- Not always all parents for adjustment needed
- Other adjustment sets possible (that block all backdoor paths)

Interventions: Computing the Effects of Actions

Truncated Product Formula

- Adjustment formula is for a single intervention $do(X = x)$
- Can be generalised to multiple interventions $do(X_1 = x_1, \ldots, X_\ell = x_\ell)$:

$$P(Y_1 = y_1, \ldots, Y_k = y_k \mid do(X_1 = x_1, \ldots, X_\ell = x_\ell))$$
$$= \prod_{i=1}^{k} P(Y_i = y_i \mid Pa(Y_i) = pa(Y_i))$$

Interventions: Computing the Effects of Actions

Truncated Product Formula

- Adjustment formula is for a single intervention $do(X = x)$
- Can be generalised to multiple interventions $do(X_1 = x_1, \ldots, X_\ell = x_\ell)$:

$$P(Y_1 = y_1, \ldots, Y_k = y_k \mid do(X_1 = x_1, \ldots, X_\ell = x_\ell))$$
$$= \prod_{i=1}^{k} P(Y_i = y_i \mid Pa(Y_i) = pa(Y_i))$$

- Without intervening, the distribution is given by

$$P(Y_1 = y_1, \ldots, Y_k = y_k, X_1 = x_1, \ldots, X_\ell = x_\ell)$$
$$= \prod_{i=1}^{k} P(Y_i = y_i \mid Pa(Y_i) = pa(Y_i)) \prod_{i=1}^{\ell} P(X_i = x_i \mid Pa(X_i) = pa(X_i))$$

- Here, $\{Y_1, \ldots, Y_k\} \cup \{X_1, \ldots, X_\ell\}$ is a partition of all random variables

From Propositional to Lifted Causal Inference

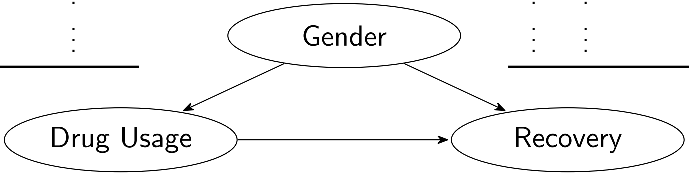
Representation of Causal Models

- Remember: *Causal model* = Causal graph + probability distribution
- E.g., causal Bayesian network

$P(G)$
⋮

G	$P(D \mid G)$
⋮	⋮
⋮	⋮

G	D	$P(R \mid G, D)$
⋮	⋮	⋮
⋮	⋮	⋮

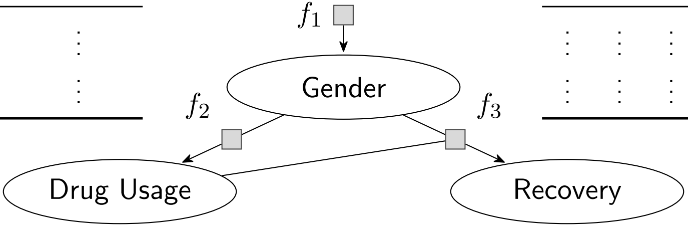


- We will use causal factor graphs instead of causal Bayesian networks
- Every Bayesian network can be transformed into an equivalent factor graph

G	$\phi_1(G)$
$\vdots$	$\vdots$

G	D	$\phi_2(G, D)$
$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$

G	D	R	$\phi_3(G, D, R)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$



From Propositional to Lifted Causal Inference

Factor Graphs as Causal Models – Semantics

- A factor graph compactly encodes a full joint probability distribution
- Semantics is given by a product over all factors:

$$P(\boldsymbol{R} = \boldsymbol{r}) = \frac{1}{Z} \prod_{j=1}^{m} \phi_j(\mathcal{R}_j = \boldsymbol{r}_j)$$

- Originally an undirected model, but can be extended to encode causal knowledge

---

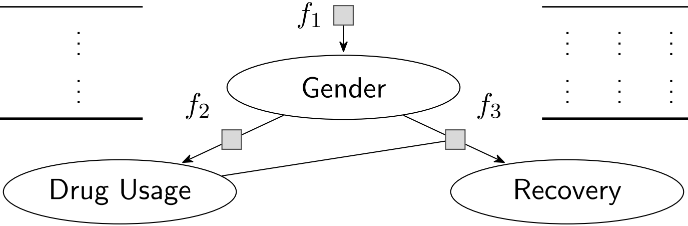
Brendan J. Frey (2003). »Extending Factor Graphs so as to Unify Directed and Undirected Graphical Models«. *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI-2003)*. Morgan Kaufmann Publishers Inc., pp. 257–264.

- Example:

$$P(g, d, r) = \frac{1}{Z} \cdot \phi_1(g) \cdot \phi_2(g, d) \cdot \phi_3(g, d, r)$$

G	$\phi_1(G)$
⋮	⋮

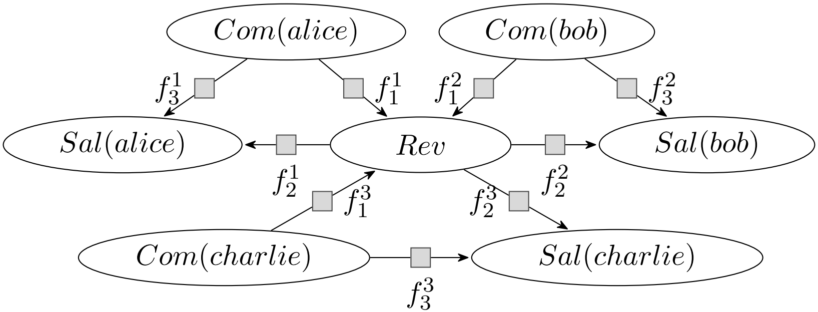
G	D	$\phi_2(G, D)$
⋮	⋮	⋮
⋮	⋮	⋮



G	D	R	$\phi_3(G, D, R)$
⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮

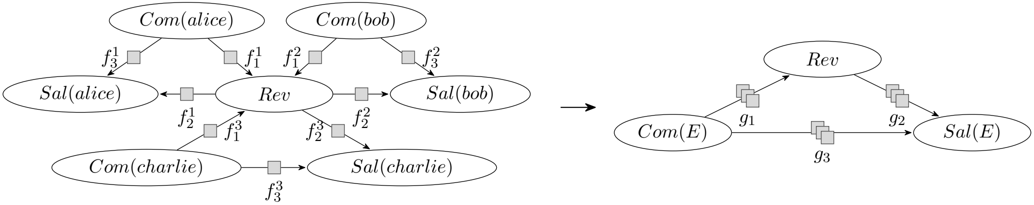
From Propositional to Lifted Causal Inference

- Common assumption: Data is independent and identically distributed (i.i.d.)
- Often not true in practice (especially in relational data)
- Our goal: Represent individual objects and their relationships



The Idea Behind Lifting

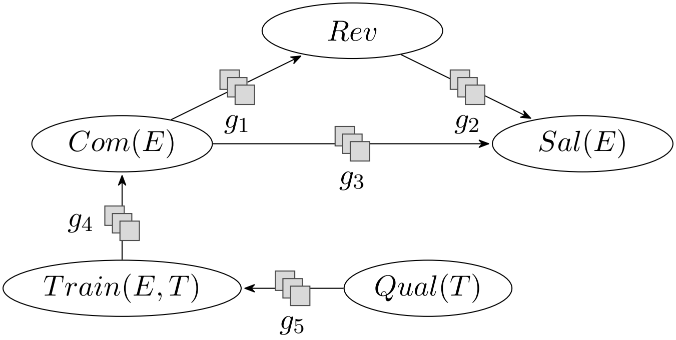
- The model becomes very large with many objects (e.g., employees)
- Assumption: There are symmetries, i.e., indistinguishable objects
- Idea: Group indistinguishable objects and reason over sets of objects



David Poole (2003). »First-Order Probabilistic Inference«. *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-2003)*. Morgan Kaufmann Publishers Inc., pp. 985–991.

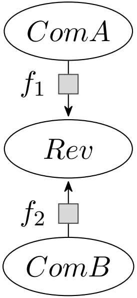
The Idea Behind Lifting

- Lifting uses a representative of indistinguishable individuals for computations
  - Logical variables to represent groups (sets) of random variables
  - Parfactors to represent sets of factors
- Lifting exploits symmetries to speed up inference



Exploitation of Symmetries

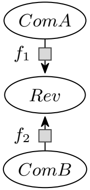
- Consider a subgraph to illustrate the idea
- $ComA$ for $Com(alice)$, $ComB$ for $Com(bob)$



$ComA$	Rev	$\phi_1(ComA, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

$ComB$	Rev	$\phi_2(ComB, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

Exploitation of Symmetries



$ComA$	Rev	$\phi_1(ComA, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

$ComB$	Rev	$\phi_2(ComB, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

- Assume we want to compute $P(Rev)$:

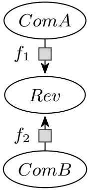
$$P(Rev) = \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} P(a, Rev, b)$$

Exploitation of Symmetries

- Assume we want to compute $P(Rev)$:

$$P(Rev) = \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} P(a, Rev, b)$$

$$= \frac{1}{Z} \cdot \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} \phi_1(a, Rev) \cdot \phi_2(b, Rev)$$



$ComA$	Rev	$\phi_1(ComA, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

$ComB$	Rev	$\phi_2(ComB, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

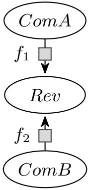
Exploitation of Symmetries

- Assume we want to compute $P(Rev)$:

$$P(Rev) = \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} P(a, Rev, b)$$

$$= \frac{1}{Z} \cdot \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} \phi_1(a, Rev) \cdot \phi_2(b, Rev)$$

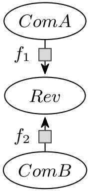
$$= \frac{1}{Z} \cdot \sum_{a \in \text{range}(ComA)} \phi_1(a, Rev) \cdot \sum_{b \in \text{range}(ComB)} \phi_2(b, Rev)$$



$ComA$	Rev	$\phi_1(ComA, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

$ComB$	Rev	$\phi_2(ComB, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

Exploitation of Symmetries



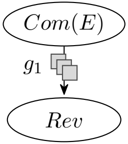
$ComA$	Rev	$\phi_1(ComA, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

$ComB$	Rev	$\phi_2(ComB, Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

- Assume we want to compute $P(Rev)$:

$$
\begin{aligned}
P(Rev) &= \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} P(a, Rev, b) \\
&= \frac{1}{Z} \cdot \sum_{a \in \text{range}(ComA)} \sum_{b \in \text{range}(ComB)} \phi_1(a, Rev) \cdot \phi_2(b, Rev) \\
&= \frac{1}{Z} \cdot \sum_{a \in \text{range}(ComA)} \phi_1(a, Rev) \cdot \sum_{b \in \text{range}(ComB)} \phi_2(b, Rev) \\
&= \frac{1}{Z} \cdot \Big(\sum_{a \in \text{range}(ComA)} \phi_1(a, Rev) \Big)^2 = \frac{1}{Z} \cdot \Big(\sum_{b \in \text{range}(ComB)} \phi_2(b, Rev) \Big)^2
\end{aligned}
$$

Exploitation of Symmetries



$Com(E)$	Rev	$\phi'_1(Com(E), Rev)$
high	high	φ_1
high	low	φ_2
low	high	φ_3
low	low	φ_4

- With $\mathrm{dom}(E) = \{alice, bob\}$:

$$P(Rev) = \frac{1}{Z} \cdot \left(\sum_{a \in \mathrm{range}(ComA)} \phi_1(a, Rev) \right)^2$$

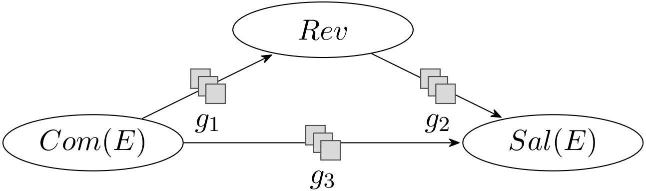
$$= \frac{1}{Z} \cdot \left(\sum_{b \in \mathrm{range}(ComB)} \phi_2(b, Rev) \right)^2$$

$$= \frac{1}{Z} \cdot \left(\sum_{c \in \mathrm{range}(Com(E))} \phi'_1(c, Rev) \right)^{|\mathrm{dom}(E)|}$$

Parametric Causal Factor Graphs (PCFGs)

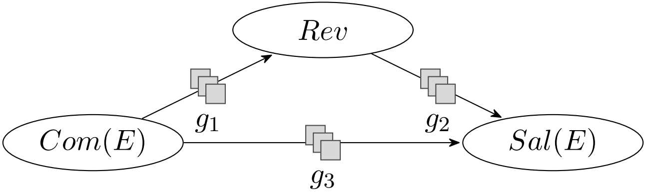
- Logical variables to represent groups of random variables
- Full joint probability distribution encoded by a product over all ground factors:

$$P(\boldsymbol{R} = \boldsymbol{r}) = \frac{1}{Z} \prod_{\phi_j \in \boldsymbol{\Phi}} \prod_{\phi_k \in \mathrm{gr}(\phi_j)} \phi_k(\mathcal{R}_k = \boldsymbol{r}_k)$$



Malte Luttermann, Mattis Hartwig, et al. (2024). »Lifted Causal Inference in Relational Domains«. *Proceedings of the Third Conference on Causal Learning and Reasoning (CLeaR-2024)*. PMLR, pp. 827–842.

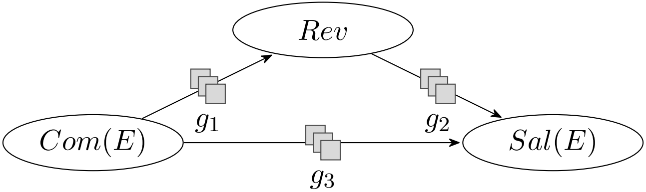
Lifted Causal Inference in PCFGs



- Is it worth the costs to send an employee to a training course?

$$P(Rev \mid do(Com(alice) = \text{high})) - P(Rev \mid do(Com(alice) = \text{low})) = ?$$

Lifted Causal Inference in PCFGs



- Is it worth the costs to send an employee to a training course?

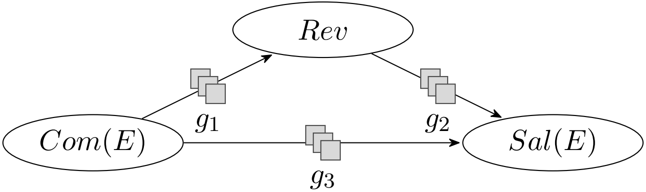
$$P(Rev \mid do(Com(alice) = \text{high})) - P(Rev \mid do(Com(alice) = \text{low})) = ?$$

- What effect has sending all employees to a training course on the revenue?

$$P(Rev \mid do(Com(E) = \text{high})) - P(Rev \mid do(Com(E) = \text{low})) = ?$$

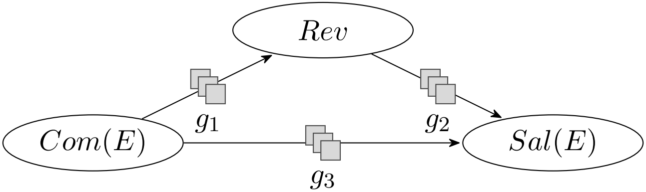
Lifted Causal Inference in PCFGs

- E.g., $P(Rev \mid do(Com(E) = \text{high}))$
  - Sets fixed value $Com(E) = \text{high}$
  - Removes incoming influences from $Com(E)$ (truncated product formula)



Lifted Causal Inference in PCFGs

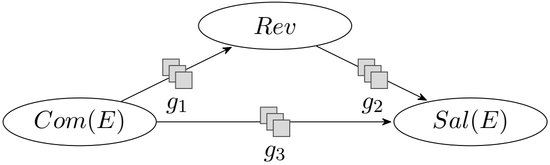
- E.g., $P(Rev \mid do(Com(E) = \mathrm{high}))$
  - Sets fixed value $Com(E) = \mathrm{high}$
  - Removes incoming influences from $Com(E)$ (truncated product formula)



- $do(Com(E) = \mathrm{high})$ is shorthand for $do(Com(e_1) = \mathrm{high}, \dots, Com(e_k) = \mathrm{high})$, where $\mathrm{dom}(E) = \{e_1, \dots, e_k\}$
- In non-lifted model, every $e_i \in \mathrm{dom}(E)$ has to be considered separately

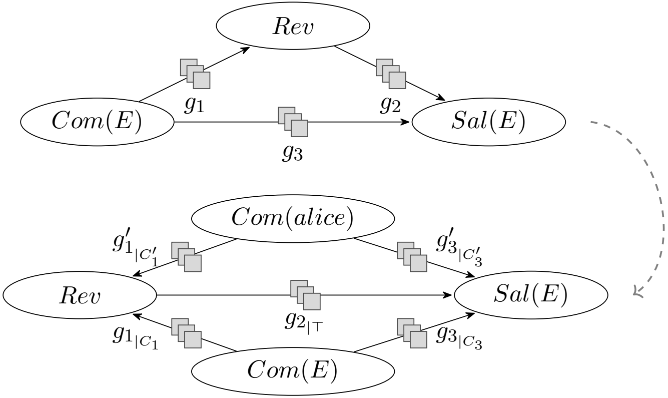
Lifted Causal Inference in PCFGs

- An intervention on a propositional random variable requires splitting of nodes
- E.g., $P(Rev \mid do(Com(alice) = high))$
  - Removes $alice$ from $Com(E)$
  - Adds an additional node $Com(alice)$



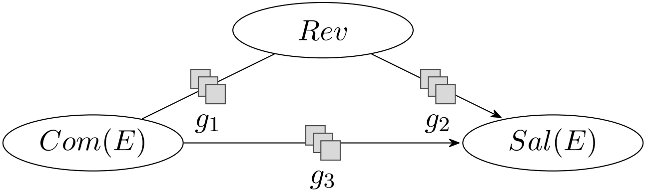
Lifted Causal Inference in PCFGs

- An intervention on a propositional random variable requires splitting of nodes
- E.g., $P(Rev \mid do(Com(alice) = high))$
  - Removes $alice$ from $Com(E)$
  - Adds an additional node $Com(alice)$



Partially Directed Parametric Causal Factor Graphs

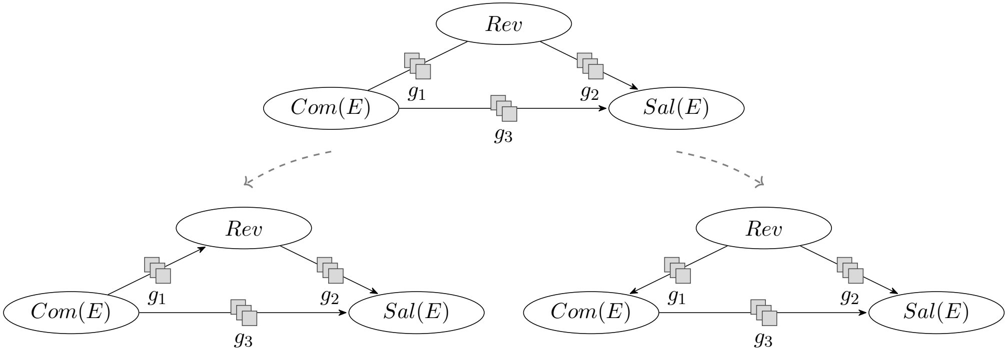
- Often not all causal relationships are known
- Directed edges to represent known causal relationships
- Undirected edges for relationships with unknown causal directions



Malte Luttermann, Tanya Braun, et al. (2024b). »Estimating Causal Effects in Partially Directed Parametric Causal Factor Graphs«. *Proceedings of the Sixteenth International Conference on Scalable Uncertainty Management (SUM-2024)*. Springer, pp. 265–280.

Lifted Causal Inference in Partially Directed PCFGs

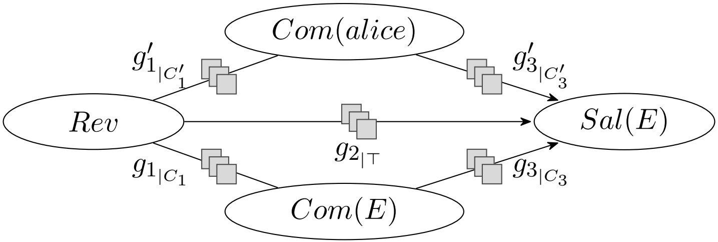
- An intervention is defined on a fully directed graph
- E.g., $P(Rev \mid do(Com(E) = \text{high}))$
  - Sets fixed value $Com(E) = \text{high}$
  - Removes incoming influences from $Com(E)$ (truncated product formula)



Lifted Causal Inference in Partially Directed PCFGs

General algorithm:

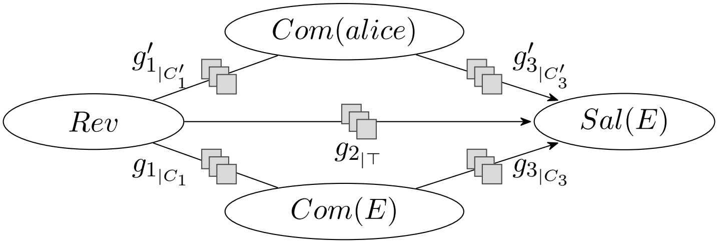
1. Split nodes of interventional variables (avoid grounding as much as possible)
2. Enumerate relevant edge directions to compute the effect of an action



Lifted Causal Inference in Partially Directed PCFGs

General algorithm:

1. Split nodes of interventional variables (avoid grounding as much as possible)
2. Enumerate relevant edge directions to compute the effect of an action

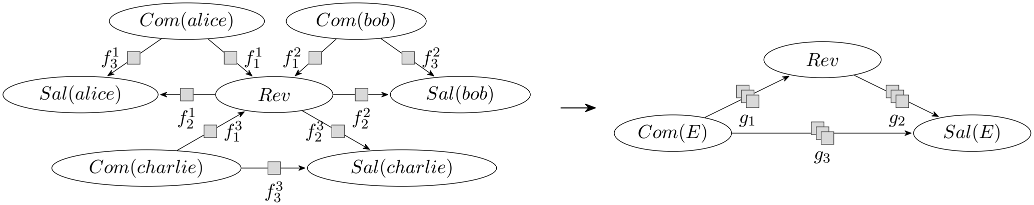


Theorem

To compute the effect of an intervention, it is sufficient to consider the directions of the undirected edges that are connected to the random variables on which we intervene.

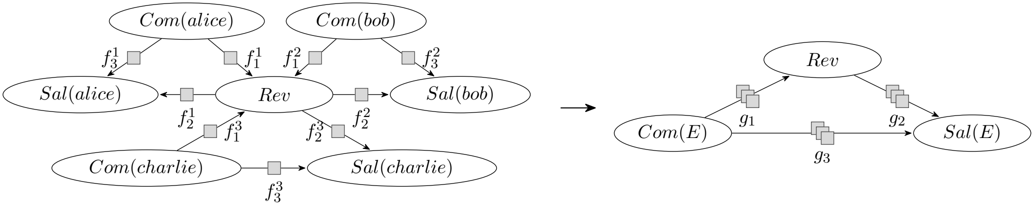
How to Obtain a Lifted Causal Model?

- The Advanced Colour Passing (ACP) algorithm compresses a factor graph
  - Start with a causal factor graph and find symmetric subgraphs
  - Symmetric subgraphs can be grouped to obtain a lifted model



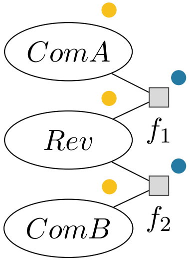
How to Obtain a Lifted Causal Model?

- The Advanced Colour Passing (ACP) algorithm compresses a factor graph
  - Start with a causal factor graph and find symmetric subgraphs
  - Symmetric subgraphs can be grouped to obtain a lifted model



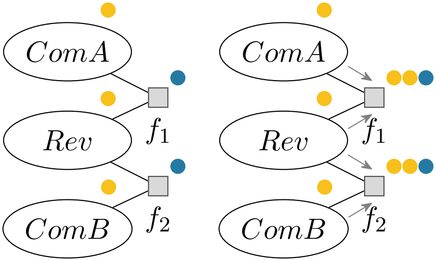
- ACP originally operates on undirected factor graphs
- Can be extended to causal (i.e., fully directed) factor graphs
- Extending it to partially directed factor graphs might be more difficult

The Advanced Colour Passing Algorithm



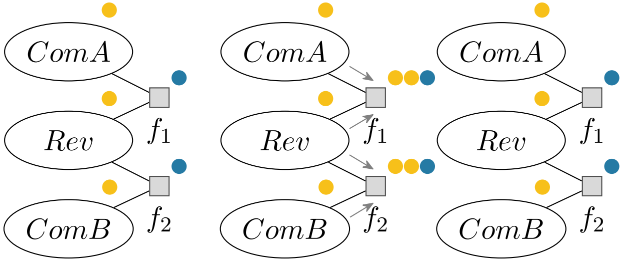
Malte Luttermann, Tanya Braun, et al. (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.

The Advanced Colour Passing Algorithm



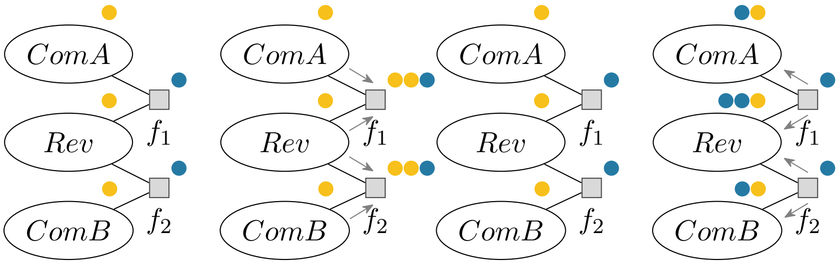
Malte Luttermann, Tanya Braun, et al. (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.

The Advanced Colour Passing Algorithm



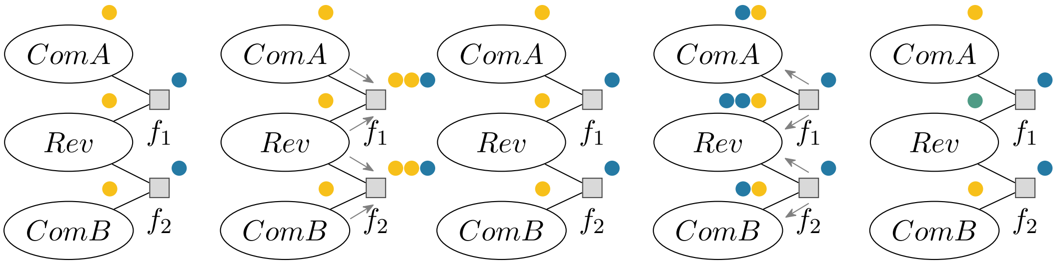
Malte Luttermann, Tanya Braun, et al. (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.

The Advanced Colour Passing Algorithm



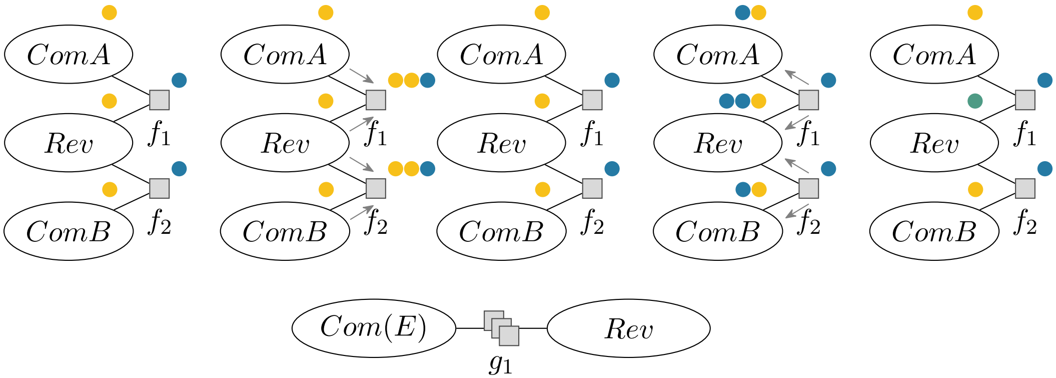
Malte Luttermann, Tanya Braun, et al. (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.

The Advanced Colour Passing Algorithm



Malte Luttermann, Tanya Braun, et al. (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.

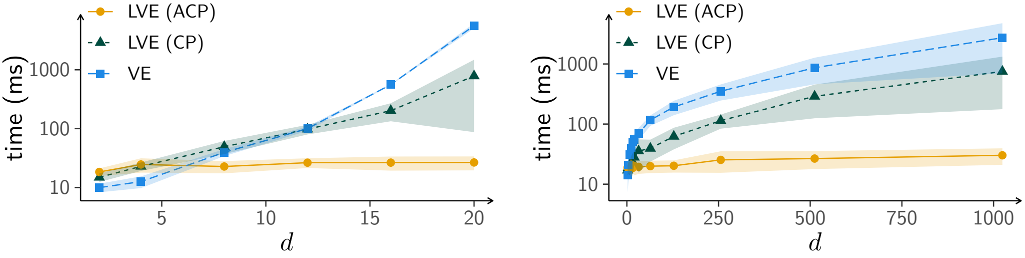
The Advanced Colour Passing Algorithm



Malte Luttermann, Tanya Braun, et al. (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.

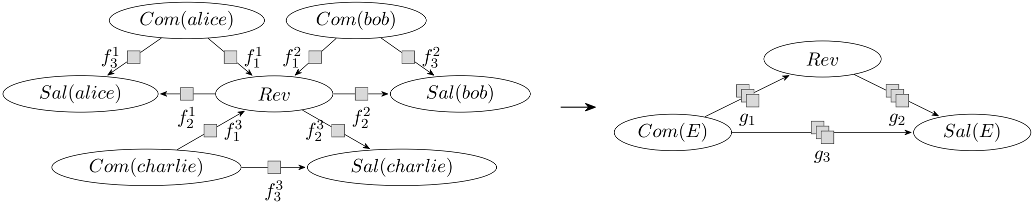
Experimental Results

- Comparison of run times for lifted inference
- d is the domain size and controls the size of the input factor graph



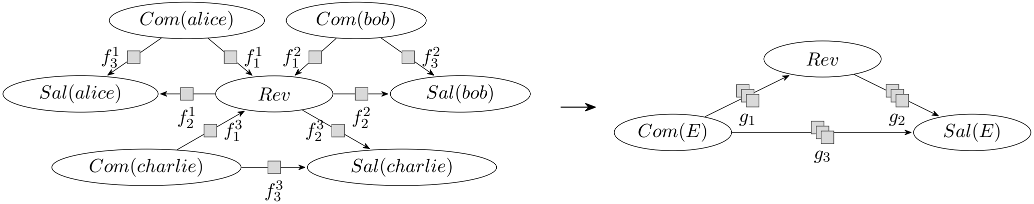
Conclusions

- For *prediction tasks*, correlation is sufficient
- For *decision making*, causal information is required
- Data is often relational (not i.i.d.)
- Lifting exploits symmetries to speed up probabilistic and causal inference



Conclusions

- For *prediction tasks*, correlation is sufficient
- For *decision making*, causal information is required
- Data is often relational (not i.i.d.)
- Lifting exploits symmetries to speed up probabilistic and causal inference



- Future research: Relax assumptions (e.g., hidden confounders, cycles, ...)

References I

- Brendan J. Frey (2003). »Extending Factor Graphs so as to Unify Directed and Undirected Graphical Models«. *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI-2003)*. Morgan Kaufmann Publishers Inc., pp. 257–264.
- Malte Luttermann, Tanya Braun, Ralf Möller, and Marcel Gehrke (2024a). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press, pp. 20500–20507.
- — (2024b). »Estimating Causal Effects in Partially Directed Parametric Causal Factor Graphs«. *Proceedings of the Sixteenth International Conference on Scalable Uncertainty Management (SUM-2024)*. Springer, pp. 265–280.
- Malte Luttermann, Mattis Hartwig, Tanya Braun, Ralf Möller, and Marcel Gehrke (2024). »Lifted Causal Inference in Relational Domains«. *Proceedings of the Third Conference on Causal Learning and Reasoning (CLeaR-2024)*. PMLR, pp. 827–842.

References II

Franz H. Messerli (2012). »Chocolate Consumption, Cognitive Function, and Nobel Laureates«. *New England Journal of Medicine* 367, pp. 1562–1564.

Judea Pearl, Madelyn Glymour, and Nicholas P. Jewell (2016). *Causal Inference in Statistics: A Primer*. 1st. Wiley.

David Poole (2003). »First-Order Probabilistic Inference«. *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI-2003)*. Morgan Kaufmann Publishers Inc., pp. 985–991.