Towards Learning Differentially Private Probabilistic Relational Models I

AnoMed Seminar

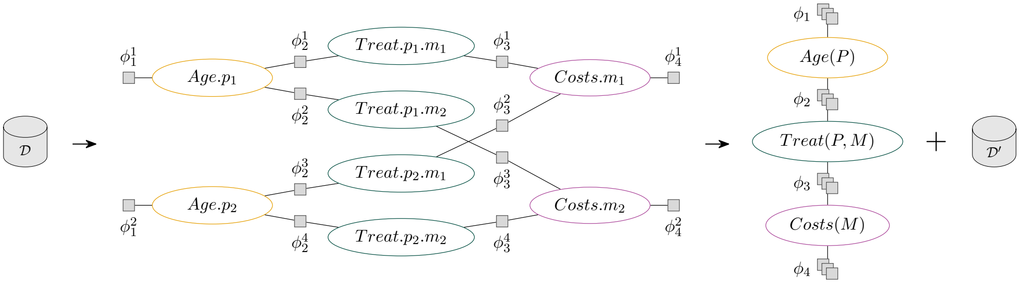
Malte Luttermann

German Research Center for Artificial Intelligence (DFKI)

April 24, 2025

Motivation

- Goal: Learn a differentially private probabilistic relational model (DP PRM)
- Reason over cohorts of patients using a lifted representation
- Sample from the DP PRM to create new publishable datasets

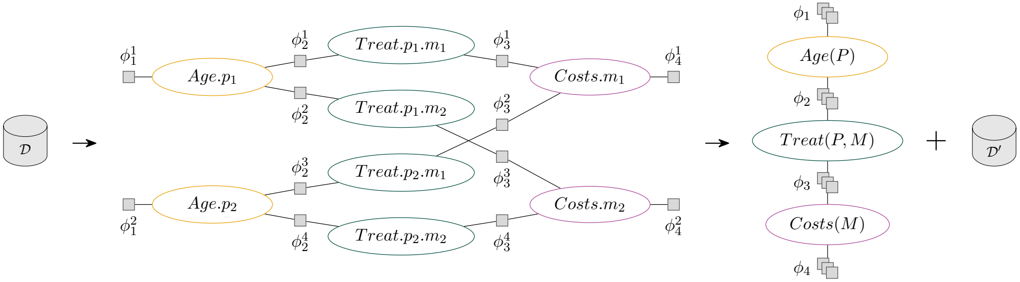


Overview

(Luttermann, Möller, and Hartwig, 2024)

Pipeline to generate synthetic relational data via probabilistic relational models:

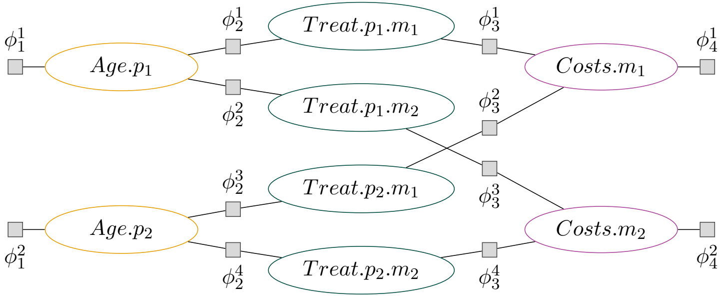
1. Construction of a factor graph
2. Transformation of the factor graph into a PRM (e.g., a parametric factor graph)
3. Sampling the PRM to generate synthetic data samples



Factor Graphs

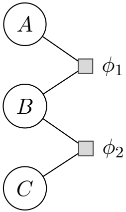
- A factor graph G compactly encodes a full joint probability distribution
- Semantics of G over a set of factors $\boldsymbol{\Phi} = \{\phi_1, \ldots, \phi_m\}$ is given by

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



Indistinguishability in Factor Graphs

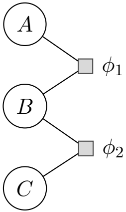
$$P(B) = \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} P(A = a, B, C = c)$$



A	B	$\phi_1(A,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

Indistinguishability in Factor Graphs

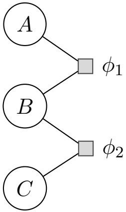


A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

$$
\begin{aligned}
P(B) &= \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} P(A = a, B, C = c) \\
&= \frac{1}{Z} \cdot \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} \phi_1(A = a, B) \cdot \phi_2(C = c, B)
\end{aligned}
$$

Indistinguishability in Factor Graphs

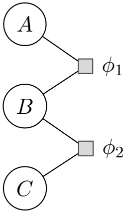


A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

$$
\begin{aligned}
P(B) &= \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} P(A = a, B, C = c) \\
&= \frac{1}{Z} \cdot \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} \phi_1(A = a, B) \cdot \phi_2(C = c, B) \\
&= \frac{1}{Z} \cdot \sum_{a \in \mathrm{range}(A)} \phi_1(A = a, B) \cdot \sum_{c \in \mathrm{range}(C)} \phi_2(C = c, B)
\end{aligned}
$$

Indistinguishability in Factor Graphs



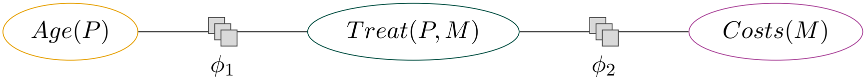
A	B	$\phi_1(A,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

$$
\begin{aligned}
P(B) &= \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} P(A=a, B, C=c) \\
&= \frac{1}{Z} \cdot \sum_{a \in \mathrm{range}(A)} \sum_{c \in \mathrm{range}(C)} \phi_1(A=a, B) \cdot \phi_2(C=c, B) \\
&= \frac{1}{Z} \cdot \sum_{a \in \mathrm{range}(A)} \phi_1(A=a, B) \cdot \sum_{c \in \mathrm{range}(C)} \phi_2(C=c, B) \\
&= \frac{1}{Z} \cdot \Big(\sum_{a \in \mathrm{range}(A)} \phi_1(A=a, B) \Big)^2 = \frac{1}{Z} \cdot \Big(\sum_{c \in \mathrm{range}(C)} \phi_2(C=c, B) \Big)^2
\end{aligned}
$$

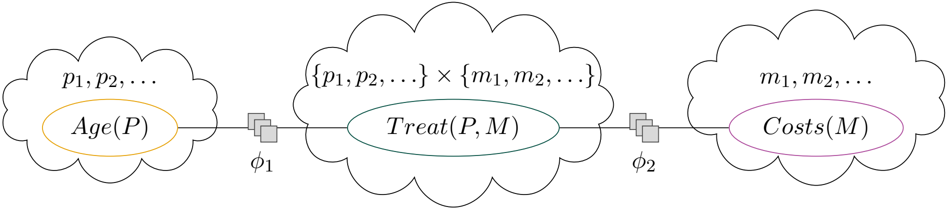
Probabilistic Relational Models: Parametric Factor Graphs

- Compact encoding of a joint probability distribution on a lifted (first-order) level
- Use a representative of indistinguishable objects for computations
- Logical variables to represent groups of random variables
  - $\mathrm{dom}(P) = \{p_1, p_2, \ldots\}$
  - $\mathrm{dom}(M) = \{m_1, m_2, \ldots\}$



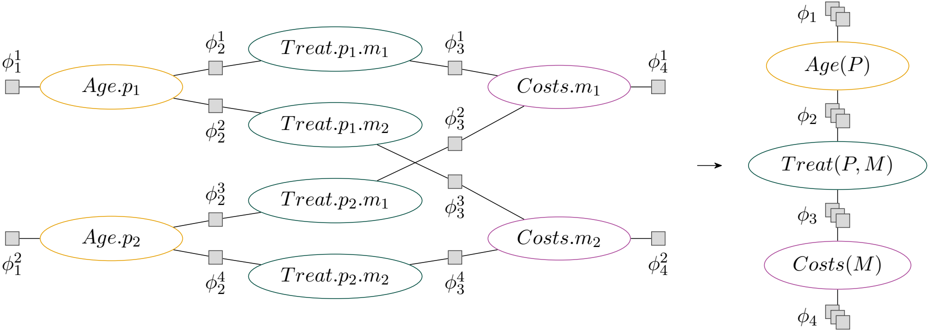
Probabilistic Relational Models: Parametric Factor Graphs

- Compact encoding of a joint probability distribution on a lifted (first-order) level
- Use a representative of indistinguishable objects for computations
- Logical variables to represent groups of random variables
  - $\mathrm{dom}(P) = \{p_1, p_2, \ldots\}$
  - $\mathrm{dom}(M) = \{m_1, m_2, \ldots\}$



Constructing a Lifted Representation

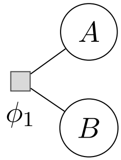
- Idea: Detect symmetric subgraphs by passing colours around
- Assign colours to random variables depending on their ranges and evidence
- Assign colours to factors depending on their potentials



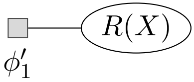
Commutative Factors

- Use histograms to detect commutativity of factors
- Each histogram must be mapped to a unique value
- If a factor is commutative, its neighbours might be grouped

A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_2
false	false	φ_3

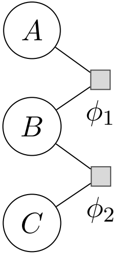


$\#_X[R(X)]$	$\phi'_1(\#_X[R(X)])$
$[2, 0]$	φ_1
$[1, 1]$	φ_2
$[0, 2]$	φ_3



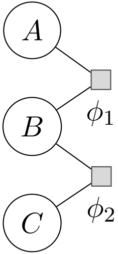
Exchangeable Factors

- Both factor graphs entail equivalent semantics
- Histograms can be used as a filter condition to find permutations



A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

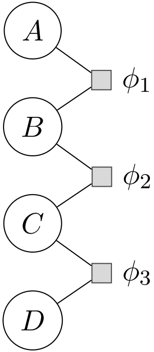


A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

B	C	$\phi_2(B, C)$
true	true	φ_1
true	false	φ_3
false	true	φ_2
false	false	φ_4

The Advanced Colour Passing Algorithm

(Luttermann, Braun, et al., 2024)



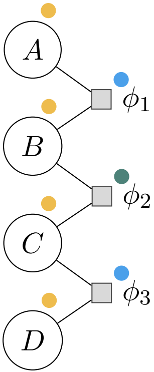
A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	D	$\phi_3(C, D)$
true	true	φ_1
true	false	φ_3
false	true	φ_2
false	false	φ_4

B	C	$\phi_2(B, C)$
true	true	φ_5
true	false	φ_6
false	true	φ_6
false	false	φ_7

The Advanced Colour Passing Algorithm

(Luttermann, Braun, et al., 2024)



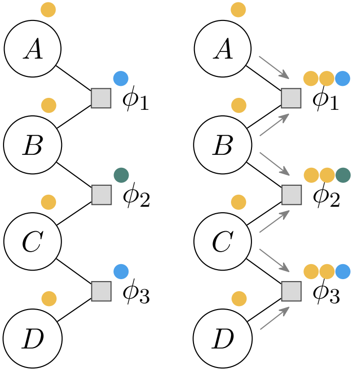
A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

D	C	$\phi_3(D, C)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

B	C	$\phi_2(B, C)$
true	true	φ_5
true	false	φ_6
false	true	φ_6
false	false	φ_7

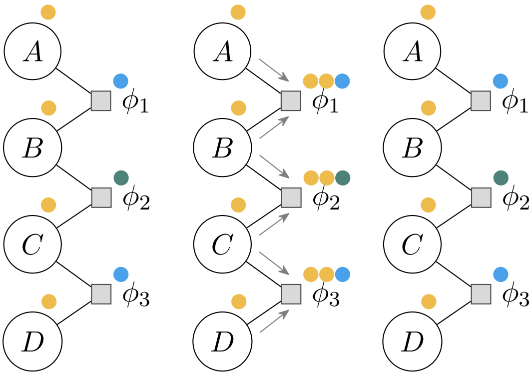
The Advanced Colour Passing Algorithm

(Luttermann, Braun, et al., 2024)



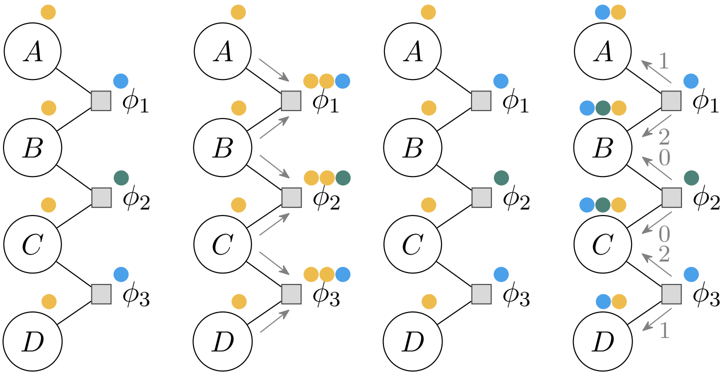
The Advanced Colour Passing Algorithm

(Luttermann, Braun, et al., 2024)



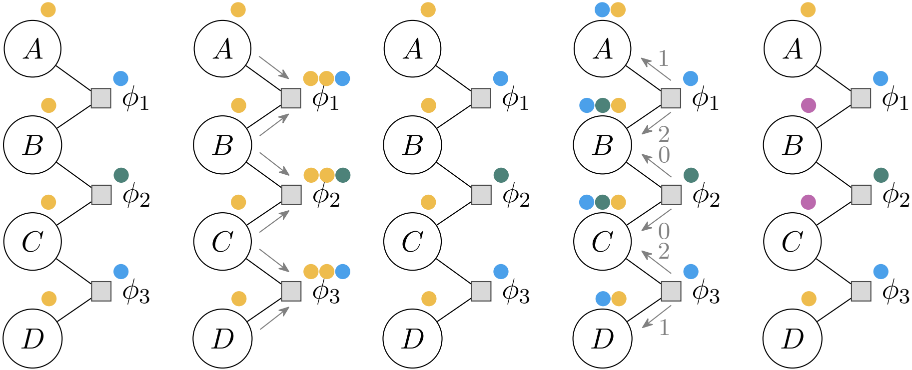
The Advanced Colour Passing Algorithm

(Luttermann, Braun, et al., 2024)



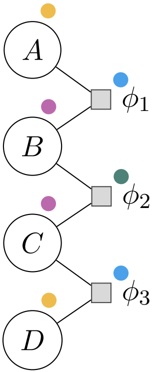
The Advanced Colour Passing Algorithm

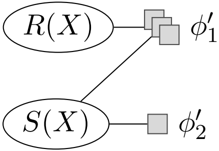
(Luttermann, Braun, et al., 2024)



The Advanced Colour Passing Algorithm

(Luttermann, Braun, et al., 2024)





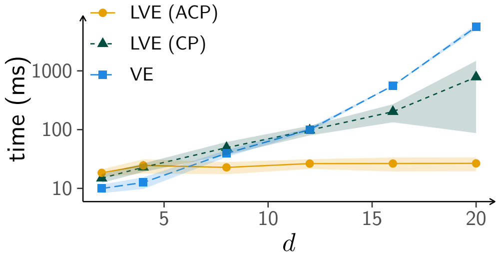
$R(X)$	$S(X)$	$\phi'_1(R(X), S(X))$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

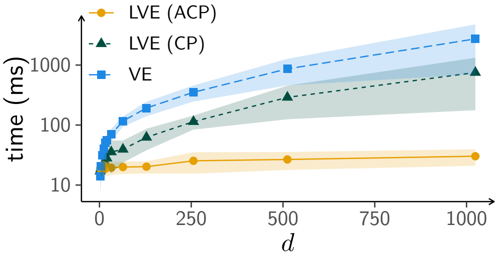
$\#_X[S(X)]$	$\phi'_2(\#_X[S(X)])$
$[2, 0]$	φ_5
$[1, 1]$	φ_6
$[0, 2]$	φ_7

The Advanced Colour Passing Algorithm – Experiments

(Luttermann, Braun, et al., 2024)

- Comparison of run times for lifted inference
- d is the domain size and controls the size of the input factor graph
- Left: Factor graphs with 1 commutative factor (no permuted argument orders)
- Right: Factor graphs where the arguments of 3 percent of the factors are permuted (no commutative factors)

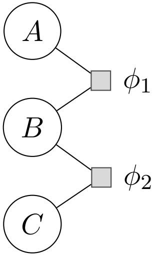




Approximate Lifted Model Construction

(Luttermann, Speller, et al., 2025)

- So far: Strict equality between potentials
  - $\varphi_1 = \varphi'_1$, $\varphi_2 = \varphi'_2$, $\varphi_3 = \varphi'_3$, $\varphi_4 = \varphi'_4$
- Next: Allow for a small deviation between potentials for practical applicability
  - $\varphi_1 \approx \varphi'_1$, $\varphi_2 \approx \varphi'_2$, $\varphi_3 \approx \varphi'_3$, $\varphi_4 \approx \varphi'_4$



A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C, B)$
true	true	φ'_1
true	false	φ'_2
false	true	φ'_3
false	false	φ'_4

ε-Equivalence

(Luttermann, Speller, et al., 2025)

- Potentials $\varphi_1 \in \mathbb{R}^+$ and $\varphi_2 \in \mathbb{R}^+$ are ε-*equivalent* if

$$\varphi_1 \in [\varphi_2 \cdot (1-\varepsilon), \varphi_2 \cdot (1+\varepsilon)], \text{ and}$$
$$\varphi_2 \in [\varphi_1 \cdot (1-\varepsilon), \varphi_1 \cdot (1+\varepsilon)]$$

- Factors $\phi_1(R_1, \ldots, R_n)$ and $\phi_2(R'_1, \ldots, R'_n)$ are ε-*equivalent* if all potentials in their potential tables are ε-equivalent
- E.g., with $\varepsilon = 0.1$, $\phi_1(A, B)$ and $\phi_2(C, B)$ are ε-equivalent:

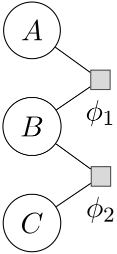
A	B	$\phi_1(A, B)$
true	true	0.8
true	false	0.3
false	true	0.5
false	false	0.2

C	B	$\phi_2(C, B)$
true	true	0.84
true	false	0.31
false	true	0.51
false	false	0.22

Grouping ε-Equivalent Factors

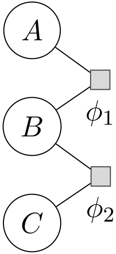
(Luttermann, Speller, et al., 2025)

- Potential tables must be identical to exploit exponentiation
- How to choose φ_i^*?



A	B	$\phi_1(A, B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C, B)$
true	true	φ'_1
true	false	φ'_2
false	true	φ'_3
false	false	φ'_4



A	B	$\phi_1(A, B)$
true	true	φ_1^*
true	false	φ_2^*
false	true	φ_3^*
false	false	φ_4^*

C	B	$\phi_2(C, B)$
true	true	φ_1^*
true	false	φ_2^*
false	true	φ_3^*
false	false	φ_4^*

Grouping ε-Equivalent Factors

(Luttermann, Speller, et al., 2025)

For a group of pairwise ε-equivalent factors $\boldsymbol{G} = \{\phi_1, \ldots, \phi_k\}$, we want to set

$$\phi_1 = \phi^*, \ldots, \phi_k = \phi^*$$

such that

$$\phi^* = \arg\min_{\phi_j} \sum_{\phi_i \in \boldsymbol{G}} Err(\phi_i, \phi_j),$$

where $Err(\phi_i, \phi_j)$ is the sum of squared deviations between ϕ_i's and ϕ_j's potentials:

$$Err(\phi_i, \phi_j) = \sum_{r_1, \ldots, r_n} \Big(\phi_i(r_1, \ldots, r_n) - \phi_j(r_1, \ldots, r_n)\Big)^2$$

Grouping ε-Equivalent Factors

(Luttermann, Speller, et al., 2025)

For a group of pairwise ε-equivalent factors $\boldsymbol{G} = \{\phi_1, \ldots, \phi_k\}$, we want to set

$$\phi_1 = \phi^*, \ldots, \phi_k = \phi^*$$

such that

$$\phi^* = \arg\min_{\phi_j} \sum_{\phi_i \in \boldsymbol{G}} Err(\phi_i, \phi_j),$$

where $Err(\phi_i, \phi_j)$ is the sum of squared deviations between ϕ_i's and ϕ_j's potentials:

$$Err(\phi_i, \phi_j) = \sum_{r_1, \ldots, r_n} \Big(\phi_i(r_1, \ldots, r_n) - \phi_j(r_1, \ldots, r_n)\Big)^2$$

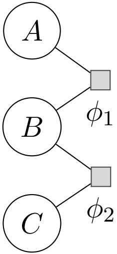
Thus, ϕ^* is chosen as the arithmetic mean of the potentials in $\boldsymbol{G}$:

$$\phi^*(\boldsymbol{r}) = \frac{1}{k} \sum_{i=1}^{k} \phi_i(\boldsymbol{r})$$

Grouping ε-Equivalent Factors

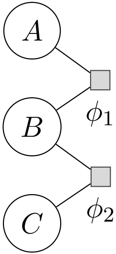
(Luttermann, Speller, et al., 2025)

- Choose φ_i^* as the row-wise arithmetic mean of the potentials



A	B	$\phi_1(A,B)$
true	true	φ_1
true	false	φ_2
false	true	φ_3
false	false	φ_4

C	B	$\phi_2(C,B)$
true	true	φ_1'
true	false	φ_2'
false	true	φ_3'
false	false	φ_4'



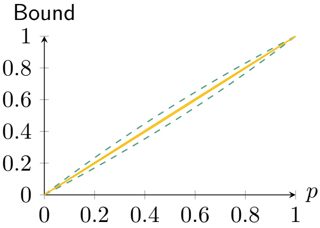
A	B	$\phi_1(A,B)$
true	true	$(\varphi_1 + \varphi_1') \ / \ 2$
true	false	$(\varphi_2 + \varphi_2') \ / \ 2$
false	true	$(\varphi_3 + \varphi_3') \ / \ 2$
false	false	$(\varphi_4 + \varphi_4') \ / \ 2$

C	B	$\phi_2(C,B)$
true	true	$(\varphi_1 + \varphi_1') \ / \ 2$
true	false	$(\varphi_2 + \varphi_2') \ / \ 2$
false	true	$(\varphi_3 + \varphi_3') \ / \ 2$
false	false	$(\varphi_4 + \varphi_4') \ / \ 2$

Bounding the Change in Query Results

(Luttermann, Speller, et al., 2025)

- Bounds for $m = 10$ (left), $m = 100$ (middle), and $m = 1000$ (right) factors
- Dashed line: $\varepsilon = 0.01$, solid line: $\varepsilon = 0.001$
- x-axes depict the original probability p, y-axes reflect the bound on the change in p

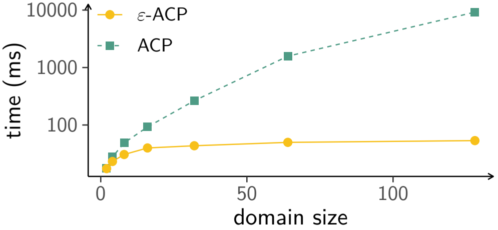


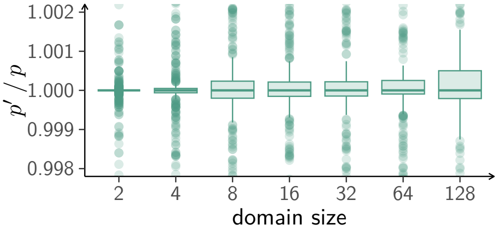
- Bounds apply to arbitrary queries and factor graphs

The ε-Advanced Colour Passing Algorithm – Experiments

(Luttermann, Speller, et al., 2025)

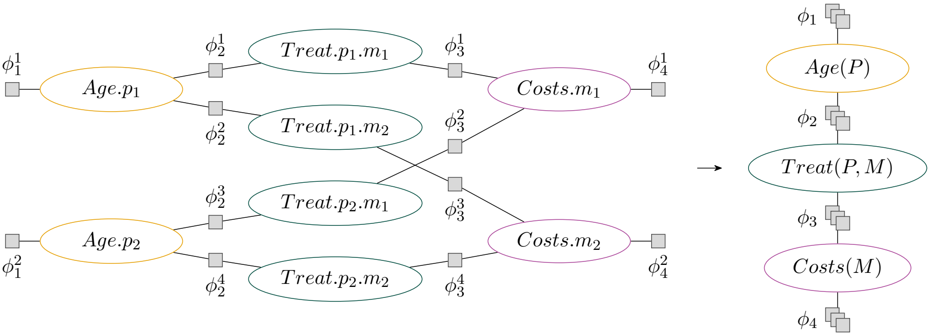
- ε-ACP is a generalisation of the Advanced Colour Passing Algorithm that assigns identical colours to groups of ε-equivalent factors (instead of equivalent factors)
- Left: Comparison of run times for lifted inference
- Right: Quotients of query results p' in the modified factor graph and p in the original factor graph





Conclusion

- First steps towards learning differentially private probabilistic relational models
- Problem of learning a (non-DP) PRM from a propositional factor graph solved



References

- Malte Luttermann, Tanya Braun, Ralf Möller, and Marcel Gehrke (2024). »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.
- Malte Luttermann, Ralf Möller, and Mattis Hartwig (2024). »Towards Privacy-Preserving Relational Data Synthesis via Probabilistic Relational Models«. *Proceedings of the Forty-Seventh German Conference on Artificial Intelligence (KI-2024)*. Springer, pp. 175–189.
- Malte Luttermann, Jan Speller, Marcel Gehrke, Tanya Braun, Ralf Möller, and Mattis Hartwig (2025). »Approximate Lifted Model Construction«. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2025)*. IJCAI Organization.