Learning Probabilistic Relational Models

IFIS Institute Meeting

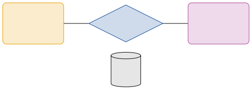
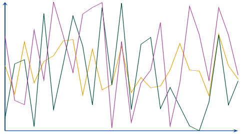
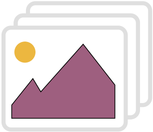
Malte Luttermann

German Research Center for Artificial Intelligence (DFKI)

January 28, 2025

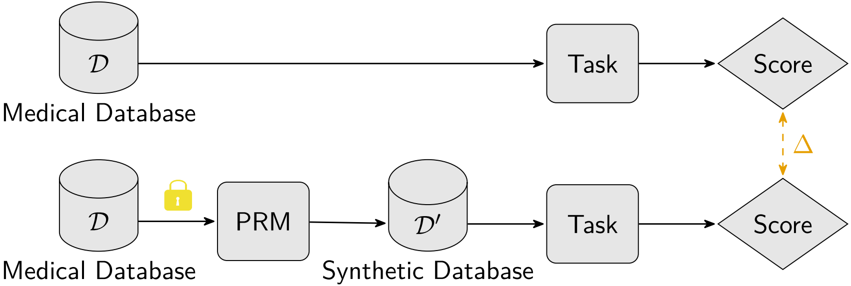
Overview of the AnoMed Project

- AnoMed = Anonymisation for medical applications
- Competence cluster with many partners
- Goal: Make medical data accessible without revealing sensitive information
  - Relational data
  - Time series data
  - Image data and video data

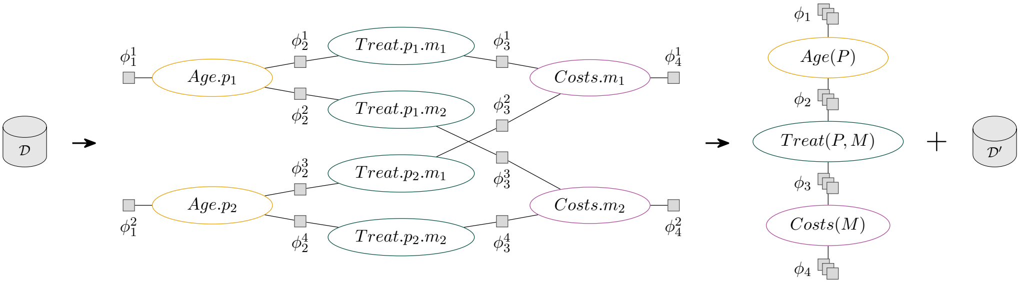
Work Package »Data Synthesis via Probabilistic Relational Models«

- Idea: Synthesise data to publish it without revealing sensitive information
- Trade-off between
  - Protection of sensitive information
  - Utility of the data



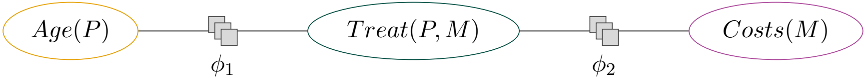
Approach

- 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



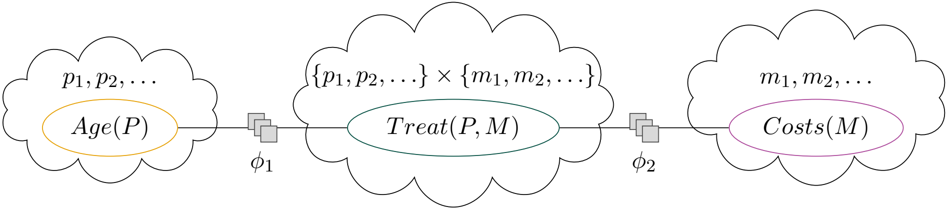
Probabilistic Relational Models: Parametric Factor Graphs

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



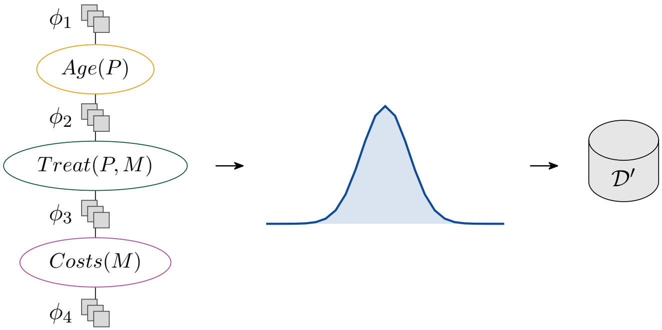
Probabilistic Relational Models: Parametric Factor Graphs

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



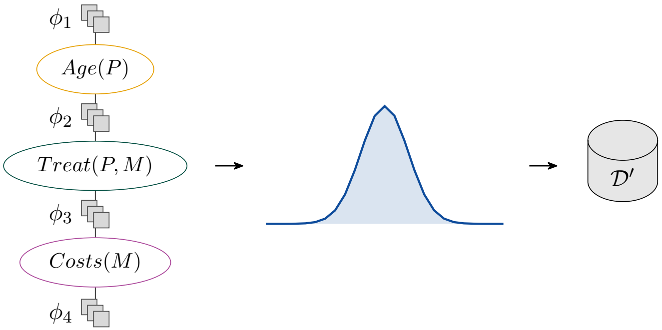
Generating New Synthetic Data Points

- Sample from the encoded probability distribution
- Release data sets for further use without privacy leakage



Generating New Synthetic Data Points

- Sample from the encoded probability distribution
- Release data sets for further use without privacy leakage



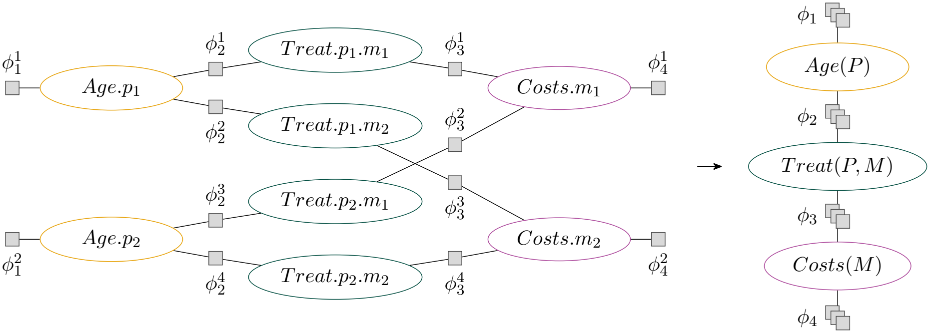
Problem

How to obtain the parametric factor graph?

Constructing a Lifted Model

Overview

- Start with a propositional probabilistic model (can be obtained from data)
- Apply a colour passing procedure to detect symmetric subgraphs
- Lift the model by grouping symmetric subgraphs



Constructing a Lifted Model

Initial Colour Assignments

Colour assignments to random variables according to their ranges:

- $\text{range}(Age.p_i) = \{\geq 18, < 18\}$
- $\text{range}(Treat.p_i.m_j) = \{\text{true}, \text{false}\}$
- $\text{range}(Costs.m_i) = \{\text{low}, \text{medium}, \text{high}\}$

Colour assignments to factors according to their potential tables:

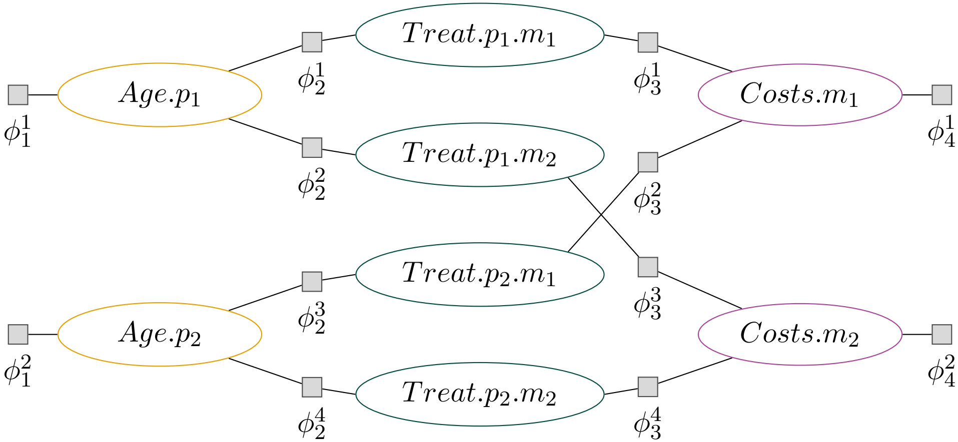
- $\phi_1^1(Age.p_1 \geq 18) = \phi_1^2(Age.p_2 \geq 18)$, $\phi_1^1(Age.p_1 < 18) = \phi_1^2(Age.p_2 < 18)$
- ...

$Age.p_1$	ϕ_1^1
≥ 18	$\varphi_1 \in \mathbb{R}^+$
< 18	$\varphi_2 \in \mathbb{R}^+$

$Age.p_2$	ϕ_1^2
≥ 18	$\varphi_1 \in \mathbb{R}^+$
< 18	$\varphi_2 \in \mathbb{R}^+$

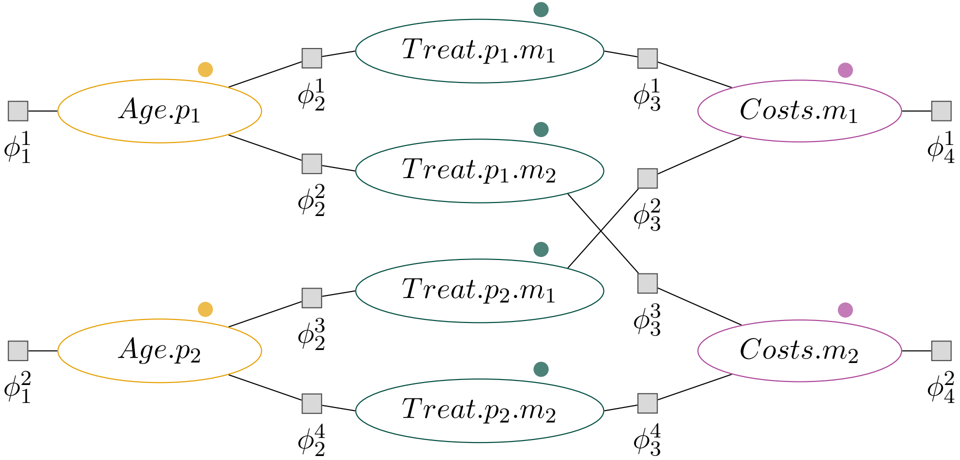
Constructing a Lifted Model

Colour Passing Procedure



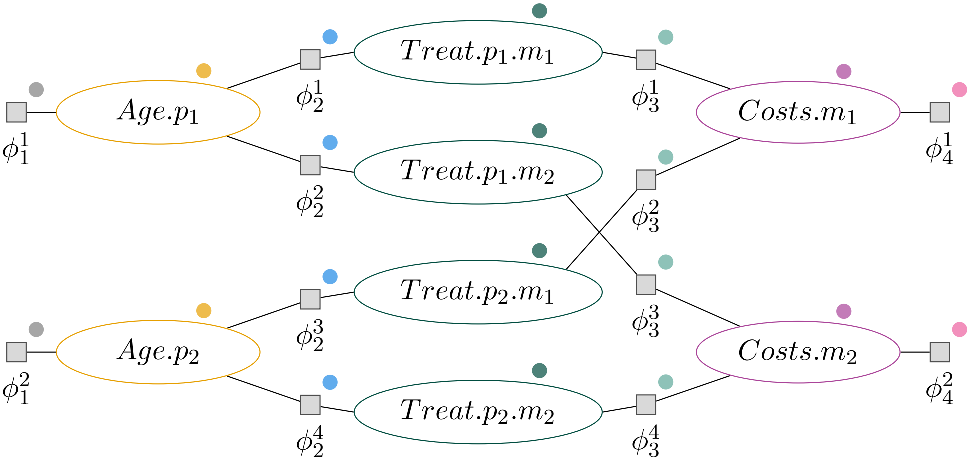
Constructing a Lifted Model

Colour Passing Procedure – Initial Colour Assignments to Random Variables



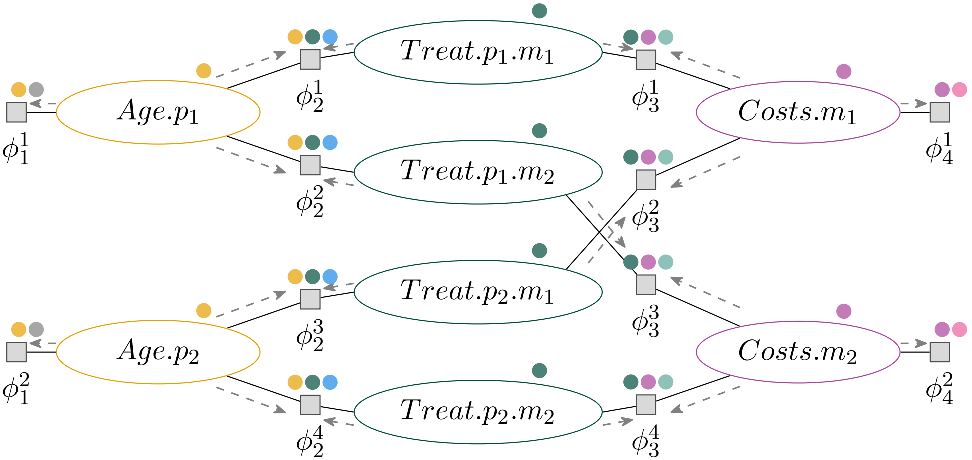
Constructing a Lifted Model

Colour Passing Procedure – Initial Colour Assignments to Factors



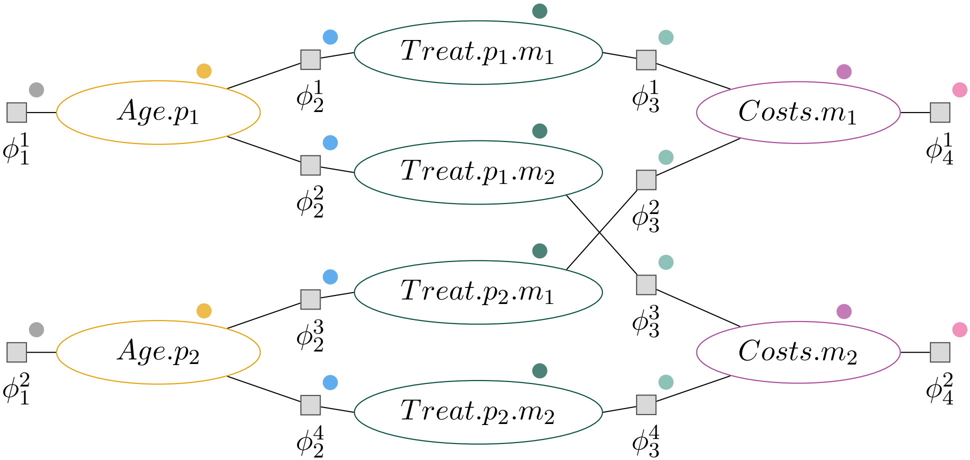
Constructing a Lifted Model

Colour Passing Procedure – Pass Colours from Random Variables to Factors



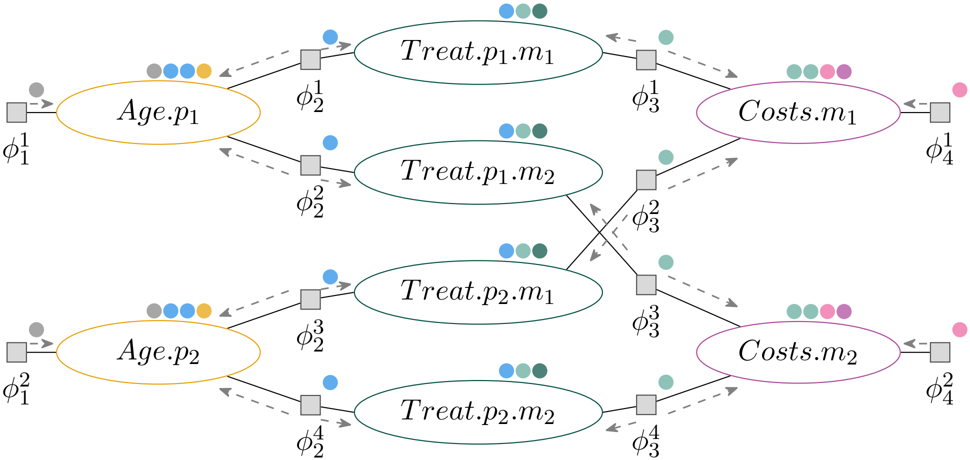
Constructing a Lifted Model

Colour Passing Procedure – Recolouring of Factors



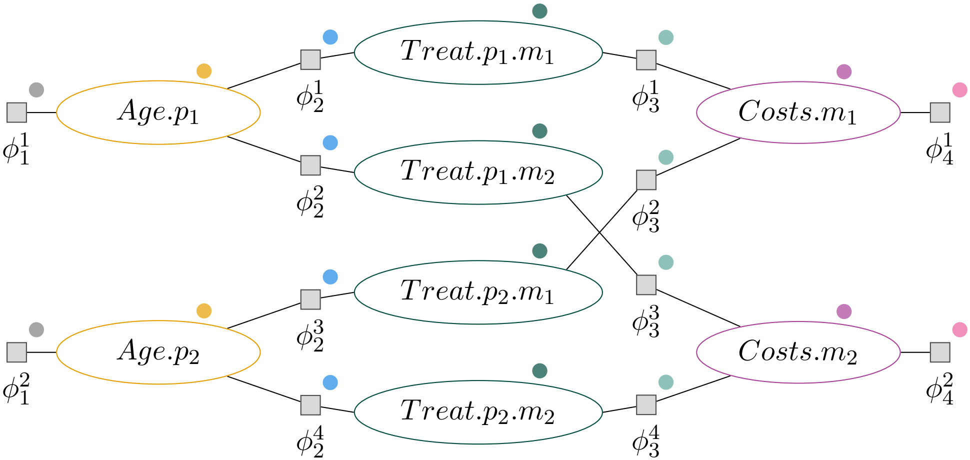
Constructing a Lifted Model

Colour Passing Procedure – Pass Colours from Factors to Random Variables



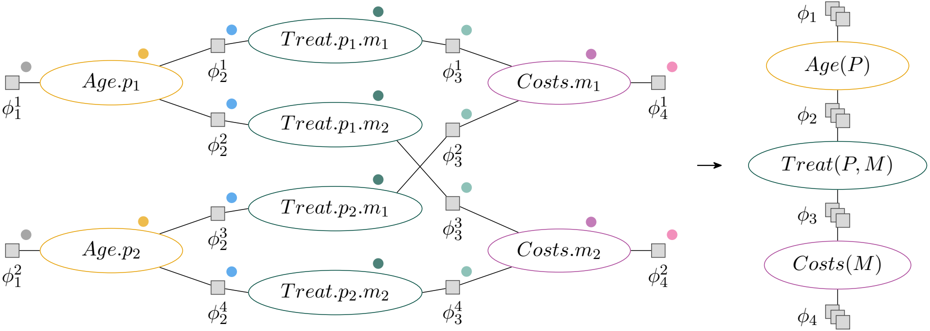
Constructing a Lifted Model

Colour Passing Procedure – Recolouring of Random Variables



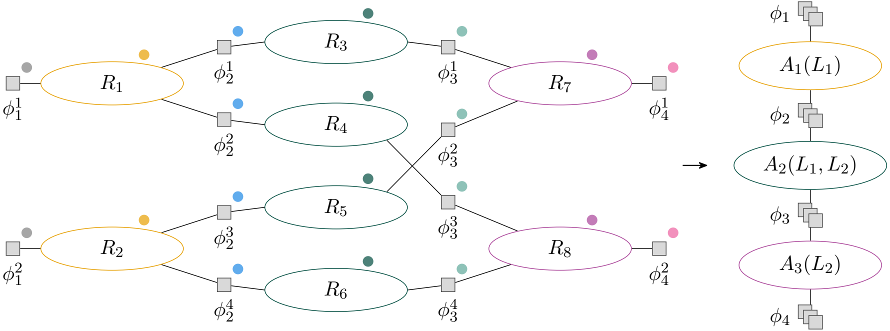
Constructing a Lifted Model

Construction of the Parametric Factor Graph



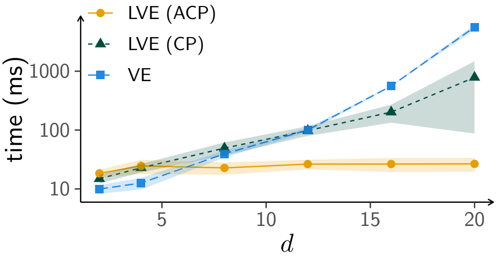
Constructing a Lifted Model

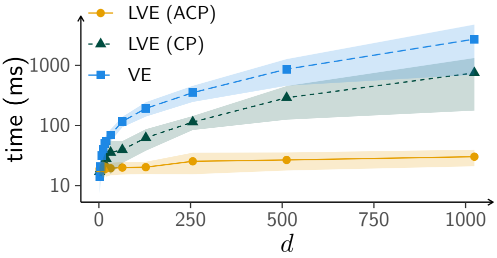
Construction of the Parametric Factor Graph



Lifted Probabilistic Inference

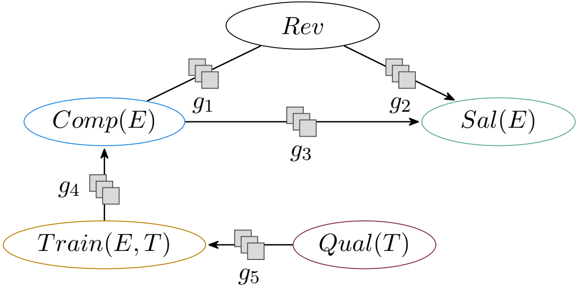
- Compute marginal distributions of random variables given observations for other random variables
- E.g., $P(Treat.p_1.m_1 \mid Costs.m_1 = \text{high})$ (»What is the probability that patient p_1 is treated with medication m_1 given that m_1 is expensive?«)





Lifted Causal Inference

- Compute causal effects and the effect of actions
  - »Is it worth the costs to send an employee to a training course?«
  - »What effect has sending all employees to a training course on the revenue?«



References I

Marcel Gehrke, Johannes Liebenow, Esfandiar Mohammadi, and Tanya Braun (2024). »Lifting in Support of Privacy-Preserving Probabilistic Inference«. *German Journal of Artificial Intelligence*.

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

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.