



**AnoMed**



**Finanziert von der  
Europäischen Union**  
NextGenerationEU



GEFÖRDERT VOM

**Bundesministerium  
für Bildung  
und Forschung**

# Annual AnoMed Meeting

## WP 3.6 – Data Synthesis via Probabilistic Relational Models

Malte Luttermann

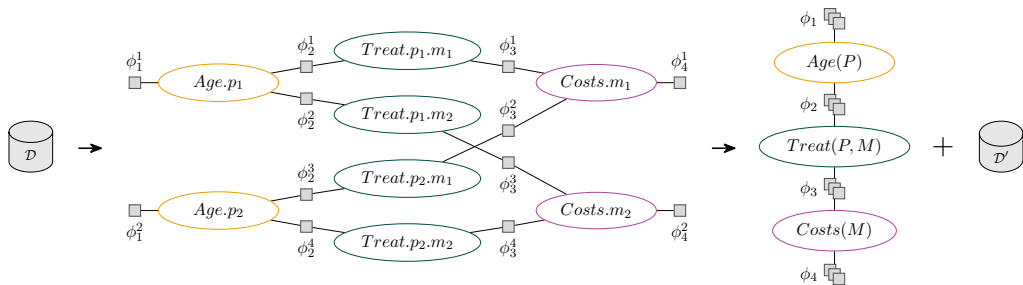
WP Partners: DFKI StarAI, UzL PrivSec

September 20, 2024

# Approach

Goal: Synthesise data to publish it without revealing sensitive information

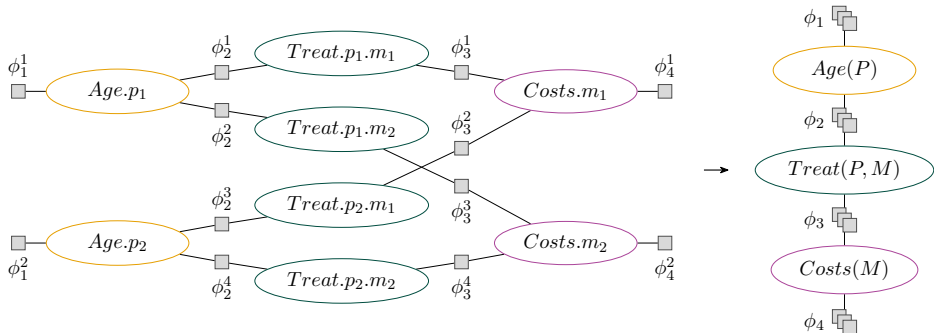
- ▶ Learn a differentially private probabilistic relational model (DP PRM)
- ▶ Reason over cohorts of patients using a lifted representation
- ▶ Sample from the DP PRM to create new publishable datasets



# Constructing a Lifted Model

(Luttermann, Braun, et al., 2024)

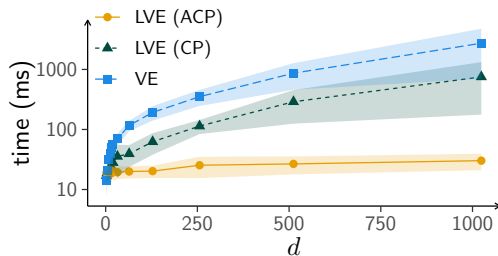
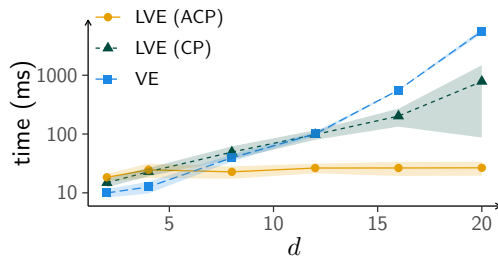
- ▶ Start with a propositional probabilistic model
- ▶ Apply a colour passing procedure to detect symmetries
- ▶ Lift the model by grouping symmetric subgraphs



# Constructing a Lifted Model

(Luttermann, Braun, et al., 2024)

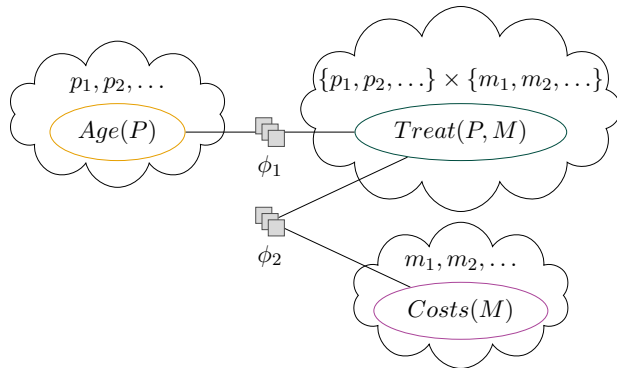
- ▶ Lifting enables reasoning over cohorts
- ▶ Groups of indistinguishable individuals hide sensitive information
- ▶ Reasoning over cohorts speeds up probabilistic inference



# Preserving Privacy

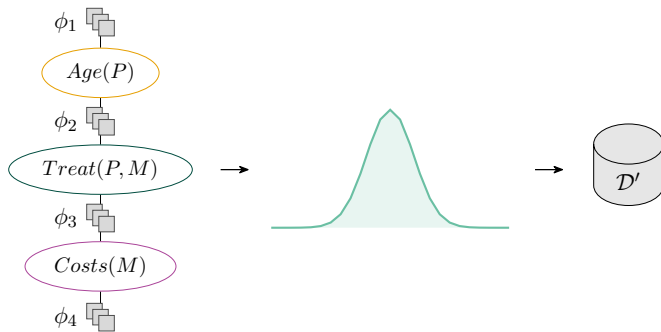
(Gehrke et al., 2024; Liebenow et al., 2024)

- ▶ New (sensitive) events over time must be included in the model
- ▶ Cluster events based on cohorts as they are expected to behave rather identically
- ▶ Combine cohorts over time if they behave more and more similar







## Generating New Synthetic Data Points





- ▶ A (DP) PRM encodes a probability distribution
- ▶ Sample from the distributions of the cohorts
- ▶ Release data sets for further use without privacy leakage



## List of Publications Related to WP 3.6 I

-  Marcel Gehrke, Johannes Liebenow, Esfandiar Mohammadi, and Tanya Braun (2024). »Lifting in Support of Privacy-Preserving Probabilistic Inference«. *German Journal of Artificial Intelligence*.
-  Johannes Liebenow, Yara Schütt, Tanya Braun, Marcel Gehrke, Florian Thaeter, and Esfandiar Mohammadi (2024). »DPM: Clustering Sensitive Data through Separation«. *To appear in: Proceedings of the Thirty-First ACM Conference on Computer and Communications Security (CCS-2024)*. ACM Press.
-  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, 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.

## List of Publications Related to WP 3.6 II

-  Malte Luttermann, Johann Machemer, and Marcel Gehrke (2024a). »Efficient Detection of Commutative Factors in Factor Graphs«. *Proceedings of the Twelfth International Conference on Probabilistic Graphical Models (PGM-2024)*. PMLR, pp. 38–56.
-  — (2024b). »Efficient Detection of Exchangeable Factors in Factor Graphs«. *Proceedings of the Thirty-Seventh International Florida Artificial Intelligence Research Society Conference (FLAIRS-2024)*. Florida Online Journals.
-  Malte Luttermann, Ralf Möller, and Marcel Gehrke (2023). »Lifting Factor Graphs with Some Unknown Factors«. *Proceedings of the Seventeenth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2023)*. Springer, pp. 337–347.
-  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.