AnoMed

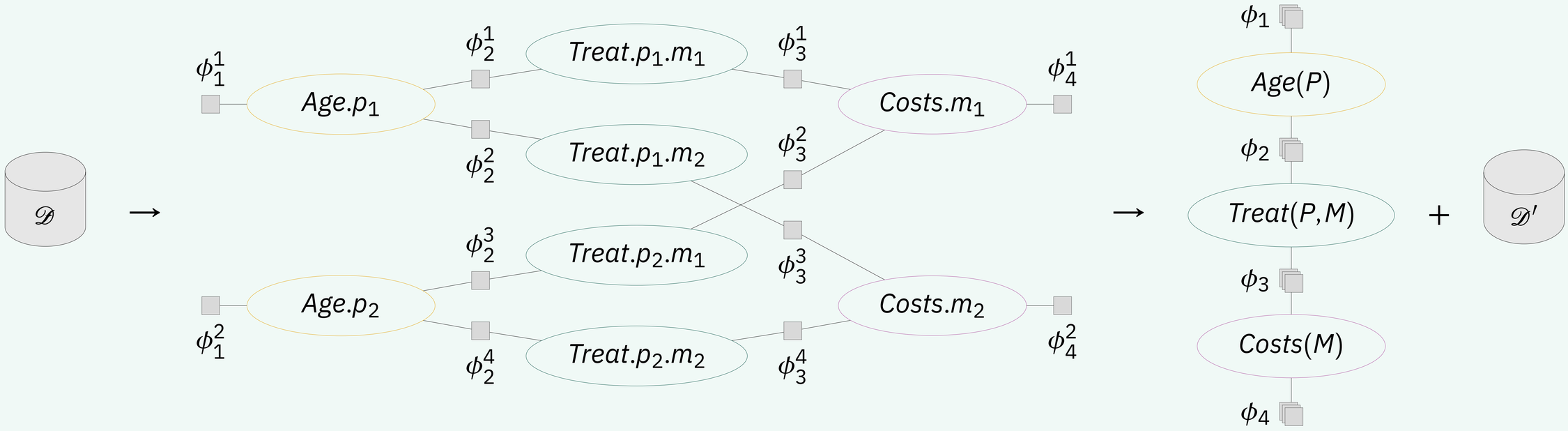
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
Finanziert von der Europäischen Union
NextGenerationEU

WP 3.6 – Data Synthesis via Probabilistic Relational Models

DFKI StarAI, UzL PrivSec

1. Motivation and Overview

Goal: Synthesise data to make it openly available without revealing sensitive information



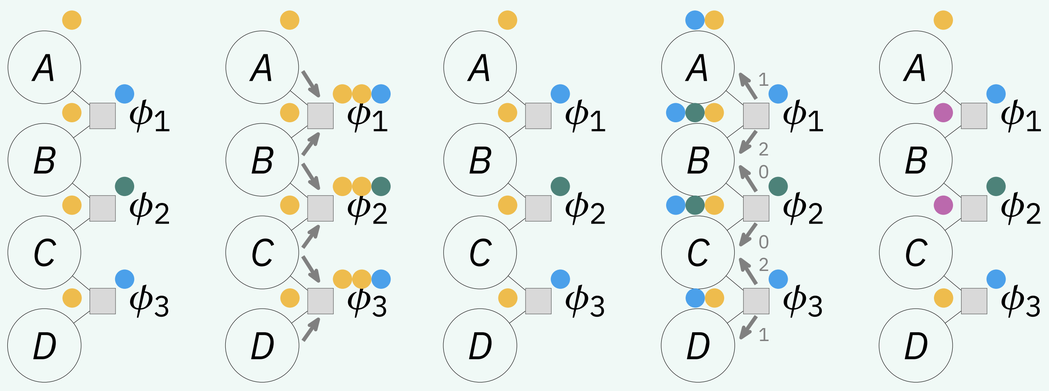
2. Approach

Learn a differentially private probabilistic relational model (DP PRM), keep it DP over time, and sample from it:

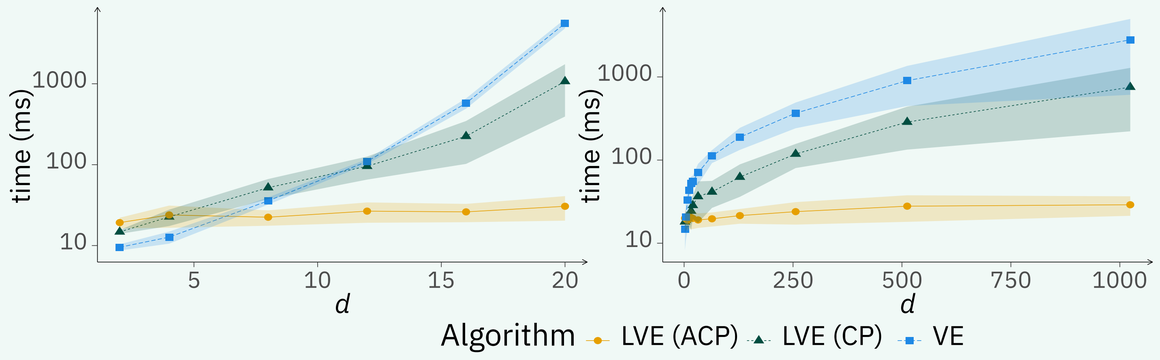
(1) Learn a propositional probabilistic model from a given database
(2) Lift the model to obtain a DP PRM and reason over cohorts
(3) Sample from the DP PRM to generate new synthetic data points

3. Constructing a Lifted Model

- Advanced Colour Passing (ACP) to lift a propositional model
- Pass colours around to detect symmetries in a graph

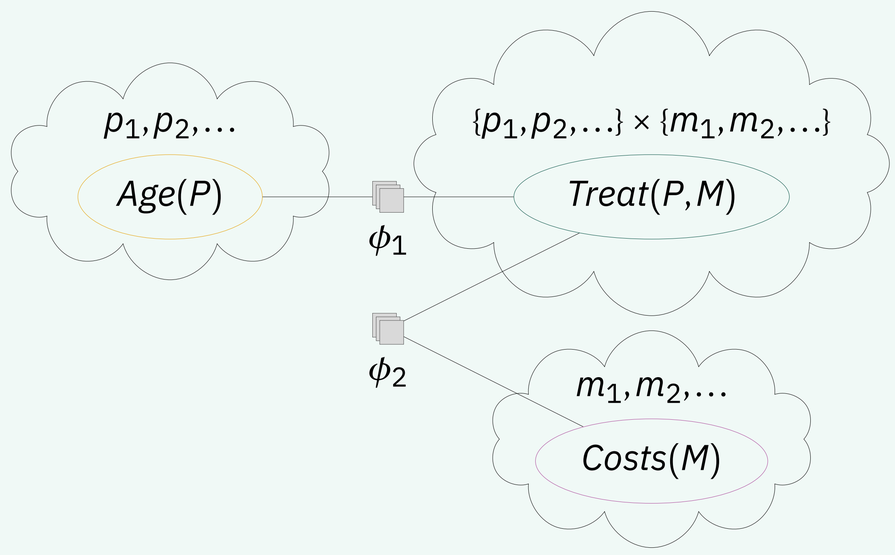


- Reasoning over cohorts also speeds up probabilistic inference



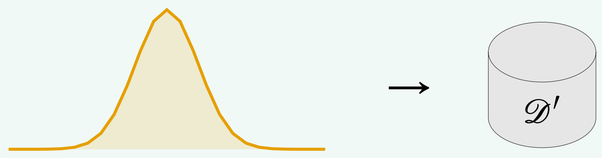
4. Preserving Privacy in a DP PRM

- Events (e.g., new lab results) over time are sensitive and must be included in the model
- Use DP clustering to avoid leakage of sensitive data
- Cluster events based on cohorts as they are expected to behave rather identically
- Combine cohorts over time if they behave strongly similar



5. Generating Synthetic Data

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



6. Related Publications

1. Marcel Gehrke, Johannes Liebenow, Esfandiar Mohammadi, and Tanya Braun (2024). »Lifting in Support of Privacy-Preserving Probabilistic Inference«. *German Journal of Artificial Intelligence*
2. Marcel Gehrke, Ralf Möller, and Tanya Braun (2020). »Taming Reasoning in Temporal Probabilistic Relational Models«. *Proceedings of the 24th European Conference on Artificial Intelligence (ECAI-2020)*. IOS Press
3. 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 31th ACM Conference on Computer and Communications Security (CCS-2024)*. ACM Press
4. Malte Luttermann, Tanya Braun, Ralf Möller, and Marcel Gehrke (2024). »Colour Passing Revisited: Lifted Model Construction with Commutative Factors«. *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI-2024)*. AAAI Press
5. Malte Luttermann, Mattis Hartwig, Tanya Braun, Ralf Möller, and Marcel Gehrke (2024). »Lifted Causal Inference in Relational Domains«. *Proceedings of the 3rd Conference on Causal Learning and Reasoning (CLeaR-2024)*. PMLR
6. Malte Luttermann, Johann Machemer, and Marcel Gehrke (2024a). »Efficient Detection of Commutative Factors in Factor Graphs«. *Proceedings of the 12th International Conference on Probabilistic Graphical Models (PGM-2024)*. PMLR
7. Malte Luttermann, Johann Machemer, and Marcel Gehrke (2024b). »Efficient Detection of Exchangeable Factors in Factor Graphs«. *Proceedings of the 37th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2024)*. Florida Online Journals
8. Malte Luttermann, Ralf Möller, and Mattis Hartwig (2024). »Towards Privacy-Preserving Relational Data Synthesis via Probabilistic Relational Models«. *Proceedings of the 47th German Conference on Artificial Intelligence (KI-2024)*. Springer
9. Malte Luttermann, Ralf Möller, and Marcel Gehrke (2023). »Lifting Factor Graphs with Some Unknown Factors«. *Proceedings of the 17th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2023)*. Springer
10. Simon Schiff, Marcel Gehrke, and Ralf Möller (2018). »Efficient Enriching of Synthesized Relational Patient Data with Time Series Data«. *Proceedings of the 8th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2018)*. Elsevier