



AnoMed



Finanziert von der
Europäischen Union
NextGenerationEU

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

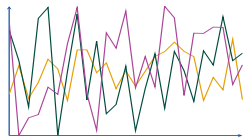
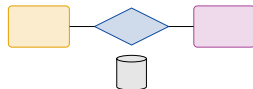
AnoMed – Anonymisierung für medizinische Anwendungen

Vorstellung Projektergebnisse

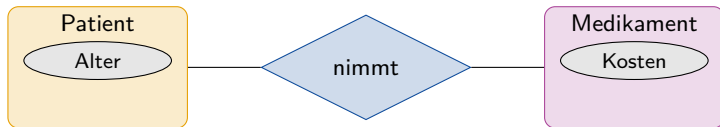
Malte Luttermann

3. Juli 2025

- ▶ Nutzung sensibler (personenbezogener) medizinischer Daten
 - ▶ Relationale Daten
 - ▶ Zeitreihen
 - ▶ Bilder
 - ▶ Videos
- ▶ Anonymisierende Datensynthese
 - ▶ Schutz der Privatsphäre (z.B. vor Rekonstruktionsangriffen, Membership Inference, ...)
 - ▶ Erhaltung der Datenqualität



Beispiel für relationale Daten



p_id	Alter	...
<i>alice</i>	45	...
<i>bob</i>	50	...
<i>charlie</i>	55	...
<i>dave</i>	17	...
<i>eve</i>	16	...
...	...	...

p_id	m_id	...
<i>alice</i>	<i>m</i> ₁	...
<i>alice</i>	<i>m</i> ₂	...
<i>bob</i>	<i>m</i> ₃	...
<i>charlie</i>	<i>m</i> ₄	...
<i>eve</i>	<i>m</i> ₅	...
...	...	...

m_id	Kosten	...
<i>m</i> ₁	45	...
<i>m</i> ₂	55	...
<i>m</i> ₃	15	...
<i>m</i> ₄	15	...
<i>m</i> ₅	50	...
...	...	...

Beispiel für synthetische relationale Daten

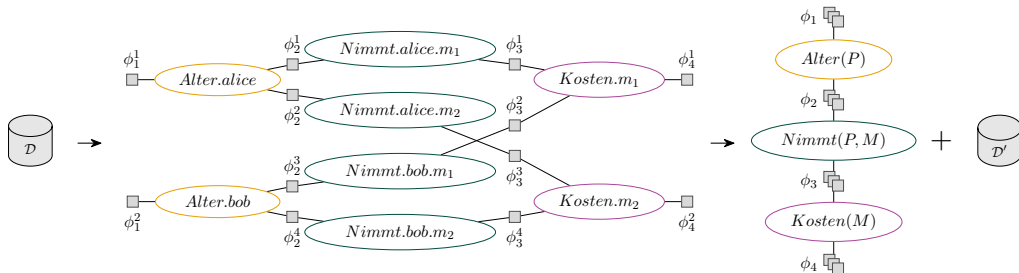
Originaldatenbank:

p_id	Alter	...	p_id	m_id	...	m_id	Kosten	...
<i>alice</i>	45	...	<i>alice</i>	<i>m</i> ₁	...	<i>m</i> ₁	45	...
<i>bob</i>	50	...	<i>alice</i>	<i>m</i> ₂	...	<i>m</i> ₂	55	...
<i>charlie</i>	55	...	<i>bob</i>	<i>m</i> ₃	...	<i>m</i> ₃	15	...
<i>dave</i>	17	...	<i>charlie</i>	<i>m</i> ₄	...	<i>m</i> ₄	15	...
<i>eve</i>	16	...	<i>eve</i>	<i>m</i> ₅	...	<i>m</i> ₅	50	...
...	...	...	...	...	...	...	...	...

Synthetische Datenbank:

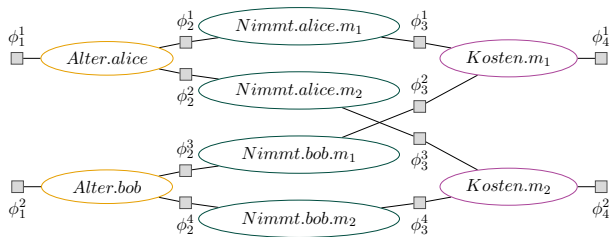
p_id	Alter	...	p_id	m_id	...	m_id	Kosten	...
<i>carol</i>	46	...	<i>carol</i>	<i>m</i> ₆	...	<i>m</i> ₆	40	...
<i>ted</i>	47	...	<i>carol</i>	<i>m</i> ₇	...	<i>m</i> ₇	45	...
<i>fred</i>	51	...	<i>ted</i>	<i>m</i> ₈	...	<i>m</i> ₈	19	...
<i>frank</i>	17	...	<i>fred</i>	<i>m</i> ₉	...	<i>m</i> ₉	20	...
<i>grace</i>	17	...	<i>grace</i>	<i>m</i> ₁₀	...	<i>m</i> ₁₀	50	...
...	...	...	...	...	...	...	...	...

- ▶ Idee: Verwendung eines probabilistischen relationalen Modells (PRM)
 - ▶ Nutzung eines PRM Lernverfahrens mit Differential Privacy (DP)
 - ▶ Zusammenfassen von Individuen zu Kohorten
 - ▶ Sampling des DP PRMs zur Generierung neuer Datenpunkte



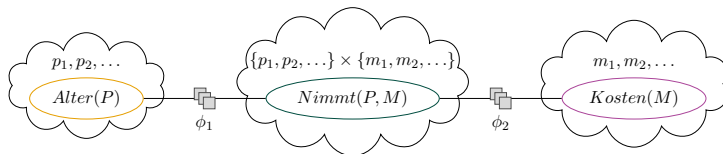
Ansatz zur anonymisierenden Datensynthese (Beispiel)

p_id	Alter	...	p_id	m_id	...	m_id	Kosten	...
<i>alice</i>	45	...	<i>alice</i>	<i>m₁</i>	...	<i>m₁</i>	45	...
<i>bob</i>	50	...	<i>alice</i>	<i>m₂</i>	...	<i>m₂</i>	55	...
<i>charlie</i>	55	...	<i>bob</i>	<i>m₃</i>	...	<i>m₃</i>	15	...
<i>dave</i>	17	...	<i>charlie</i>	<i>m₄</i>	...	<i>m₄</i>	15	...
<i>eve</i>	16	...	<i>eve</i>	<i>m₅</i>	...	<i>m₅</i>	50	...
...	...	...	...	...	...	...	...	...

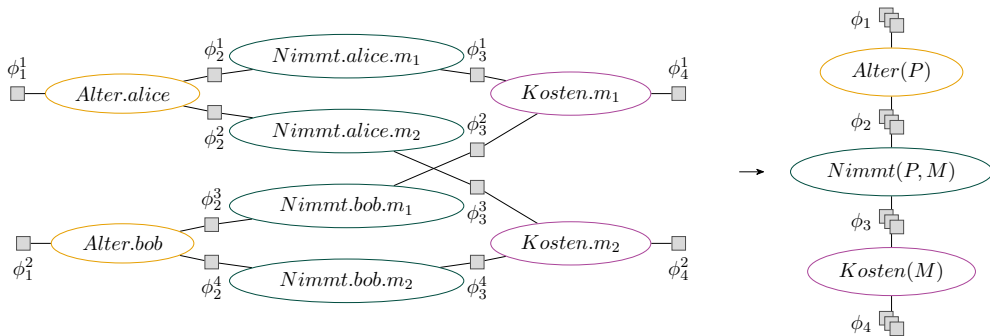


Ansatz zur anonymisierenden Datensynthese (Beispiel)

p_id	Alter	...	p_id	m_id	...	m_id	Kosten	...
alice	45	...	alice	m ₁	...	m ₁	45	...
bob	50	...	alice	m ₂	...	m ₂	55	...
charlie	55	...	bob	m ₃	...	m ₃	15	...
dave	17	...	charlie	m ₄	...	m ₄	15	...
eve	16	...	eve	m ₅	...	m ₅	50	...
...	...	...	...	...	...	...	...	...



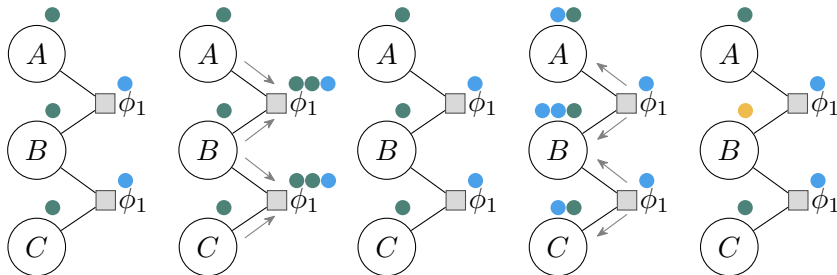
- Erkennung symmetrischer Subgraphen, um Kohorten zu bilden
- Einführung von logischen Variablen, um Kohorten zu repräsentieren



Advanced Colour Passing Algorithmus

(Luttermann, Braun u. a., 2024)

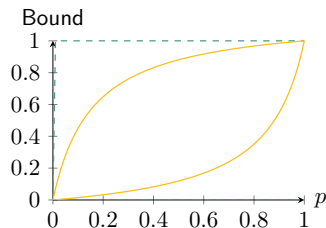
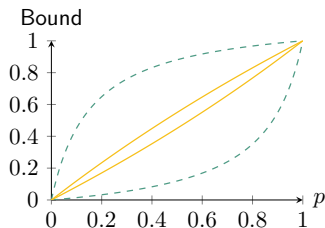
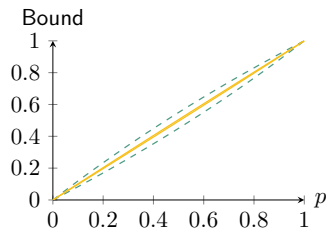
- ▶ Weitergeben von Farben zur Erkennung symmetrischer Subgraphen
- ▶ Gruppierung symmetrischer Subgraphen



Theoretische Schranken

(Luttermann, Speller u. a., 2025)

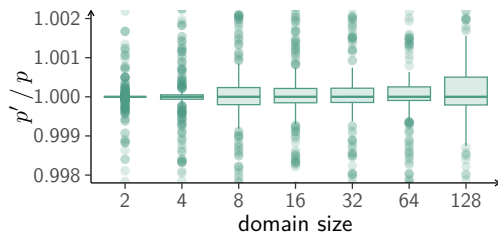
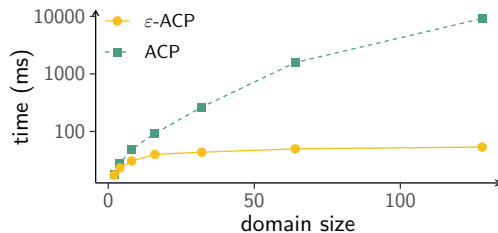
- ▶ Schranken für $m = 10$ (links), $m = 100$ (Mitte), and $m = 1000$ (rechts) Faktoren
- ▶ Gestrichelte Linie: $\varepsilon = 0.01$, durchgezogene Linie: $\varepsilon = 0.001$
- ▶ x-Achse: Echte Wahrscheinlichkeit p
- ▶ y-Achse: Schranke für die Abweichung von p



Advanced Colour Passing Algorithmus – Experimente

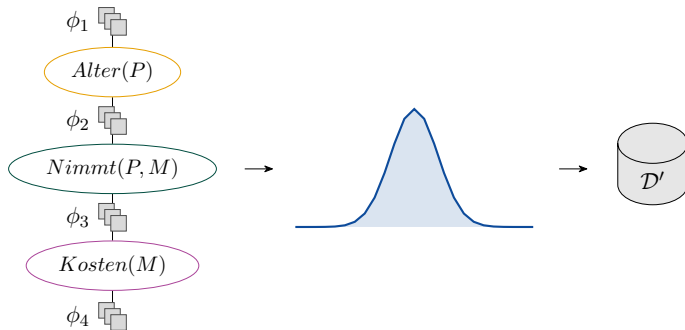
(Luttermann, Speller u. a., 2025)

- ▶ Links: Vergleich von Laufzeiten für Inferenz (Indikator für die Größe des Modells)
- ▶ Rechts: Quotienten von Anfrageergebnissen p' im komprimierten Modell und originalen Anfrageergebnissen p

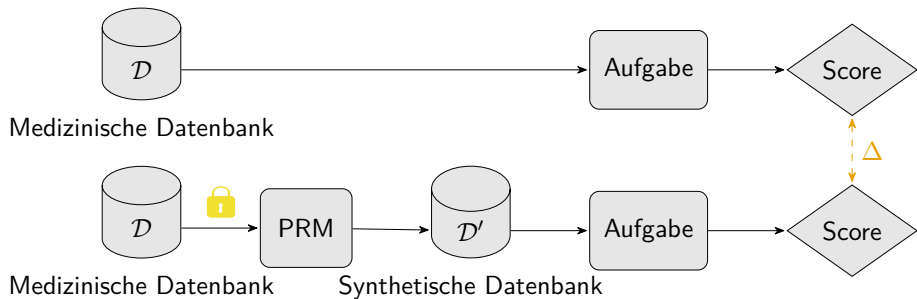


Generierung neuer synthetischer Datenpunkte

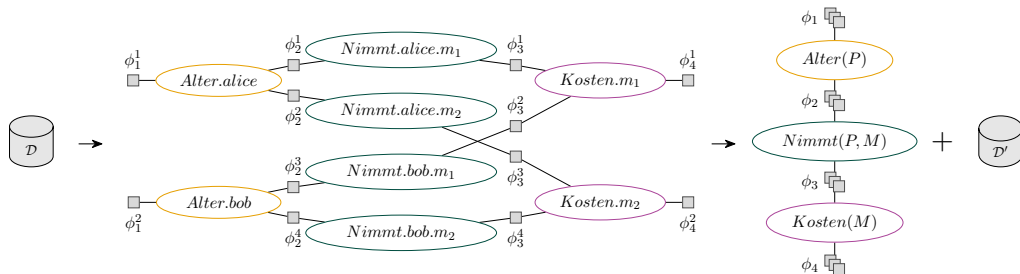
- ▶ Ein (DP) PRM kodiert eine Wahrscheinlichkeitsverteilung
- ▶ Sampling aus dieser Verteilung liefert neue Datenpunkte
- ▶ Veröffentlichung der Datenpunkte möglich



- ▶ Ziel: Generierung synthetischer relationaler Daten
 - ▶ Zur Veröffentlichung
 - ▶ Ohne Preisgabe sensibler Informationen



- ▶ Erste Ergebnisse zum Lernen probabilistischer relationaler Modelle
 - ▶ Advanced Colour Passing Algorithmus
 - ▶ Theoretische Schranken für die Abweichung von Anfragen



Nächste Schritte:

- ▶ DP Lernverfahren von PRMs
- ▶ Sampling von PRMs sowie Qualitätsbewertung synthetischer Daten

-  Malte Luttermann, Tanya Braun, Ralf Möller und 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, S. 20500–20507.
-  Malte Luttermann, Jan Speller, Marcel Gehrke, Tanya Braun, Ralf Möller und Mattis Hartwig (2025). »Approximate Lifted Model Construction«. *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-2025)*. IJCAI Organization.