



# k-Anonymität

Thomas Maier, Kai Sonnenwald, Tom Petersen

Universität Hamburg  
Fachbereich Informatik



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

# Agenda

---

1. Motivation & Abgrenzung
2. k-Anonymität
  - Generalisierung
  - Unterdrückung
3. Schwächen der k-Anonymität
4. l-Diversity
5. Schwächen der l-Diversity
6. t-Closeness
7. Literaturverzeichnis

# Motivation

Einführendes Beispiel, dass das Setting und in die Begriffe einführt

<b>Identifizier</b>	<b>Nicht-sensibel</b>			<b>Sensibel</b>
<b>Name</b>	<b>Geschlecht</b>	<b>PLZ</b>	<b>Geburtsdatum</b>	<b>Erkrankung</b>
Mia Schulz	w	21989	20.5.1944	Osteoporose
Elias Wagner	m	21727	25.8.1983	Gicht
Hanna Weber	w	20817	28.3.1953	Osteoporose
Leon Schulz	m	21220	28.10.1994	Bronchitis
Sofia Koch	w	20270	21.1.1965	Gicht
Leon Schmidt	m	20188	5.5.1958	Hepatitis
Hanna Schäfer	w	21462	11.2.1999	Epilepsie
Elias Schneider	m	20388	3.8.1971	Multiple Skle
Mia Fischer	w	21896	14.12.1999	Diabetes
Ben Meyer	m	21024	8.1.1982	Diabetes

## Anonym?

---

Sweeney - Beispiel [Swe02]

- 
- Group Insurance Commission veröffentlichte Patientendaten in anonymer Form
- Cambridge, Massachusetts voter registration list mit den öffentlichen Informationen der GIC abgeglichen
- Beispielsweise konnte Massachusetts governor William Weld eindeutig identifiziert werden.
- > [Swe00], [Gol06] Studien über die Eindeutigkeit von demographischen Faktoren in der U.S.-Bevölkerung **MAL REINSCHAUEN**

## Abgrenzung

---

Abgrenzung zu anderen Konzepten (statistische Datenbanken, Authentifikation, ...)

## k-Anonymität

---

Notizen

identifier, quasi-identifier, sensitive attributes

k-anonymity

# Begriffe

---

**Explicit identifier** Attribut, das ein Individuum (nahezu) eindeutig identifiziert. Bsp: Name, Adresse, Steuernummer, ...

**Sensitive attribute** Attribut, dessen Wert für ein Individuum in einer Datenmenge nicht herausgefunden werden darf.

**Quasi identifier** Attributmenge, die ein Individuum in Kombination identifizieren kann. *Formal in [Swe02] p. 7 auch [MKGV07] p. 3:* Eine Menge nicht-sensibler Attribute  $\{A_i, \dots, A_j\}$  einer Tabelle, deren Attribute mit einer externen Datenquelle verknüpft werden können, um mindestens ein Individuum der Gesamtmenge eindeutig zu identifizieren.

## k-Anonymität

**Informell:** Eine Tabelle (Datensatz?) erfüllt  $k$ -Anonymität, wenn jede Zeile (jeder Eintrag) ununterscheidbar von  $k - 1$  anderen Zeilen im Bezug auf jede “quasi identifizier“-Menge ist.

**Formal:** Sei  $T(A_1, \dots, A_n)$  eine Tabelle und  $Q_T = \{A_i, \dots, A_j\}$  der zugehörige quasi identifizier.  $T$  erfüllt  $k$ -Anonymität genau dann, wenn jede Belegung von Werten in  $T[Q_T]$  mindestens  $k$  mal auftritt, wobei  $T[Q_T]$  die duplikatenerhaltende Projektion von  $T$  auf die Attribute des quasi identifiziers beschreibt.



## Generalisierung

---

### **s. Samarati, Sweeney Kapitel 3**

domain, ground domain, generalization (partial ordering on domains)

generalized table

## Suppression

---

...

in combination with generalization

## Schwächen der $k$ -Anonymität

---

- Unsorted matching attack Veröffentlichung mehrerer  $k$ -anonymer Tabellen mit derselben Sortierung ausgehend von einer nicht-öffentlichen Tabelle. [Swe02] p.10
- Complementary release attack Veröffentlichung mehrerer  $k$ -anonymer Tabellen unterschiedlicher Generalisierung, die zusammengeführt die  $k$ -Anonymität verletzen. [Swe02] p.11
- Temporal attack Dynamische Tabellen können  $k$ -Anonymität verletzen. [Swe02] p.12
- Homogeneity attack Gleichheit der sensitive attributes einer Gruppe, die sich in den Werten des quasi identifiers gleicht, leakt das sensitive attribute eines Individuums. [MKG07] p. 2
- Background knowledge attack Nutzen von Hintergrundwissen, um mit hoher Wahrscheinlichkeit auf den Wert des sensitive attributes eines Individuums in einer Gruppe

# I-Diversity

---

## Schwächen der I-Diversity

---

Skewness attack  
similarity attack

# t-Closeness

---

## Literaturverzeichnis I

---



GOLLE, Philippe:

Revisiting the uniqueness of simple demographics in the US population.

In: *Proceedings of the 5th ACM workshop on Privacy in electronic society* ACM, 2006, S. 77–80




MACHANAVAJJHALA, Ashwin ; KIFER, Daniel ; GEHRKE, Johannes ; VENKITASUBRAMANIAM, Muthuramakrishnan:  
l-diversity: Privacy beyond k-anonymity.


In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1 (2007), Nr. 1, S. 3


## Literaturverzeichnis II

---

 SAMARATI, Pierangela ; SWEENEY, Latanya:  
Protecting privacy when disclosing information: k-anonymity and  
its enforcement through generalization and suppression /  
Technical report, SRI International.  
1998. —

Forschungsbericht

 SWEENEY, Latanya:  
Simple Demographics Often Identify People Uniquely.  
(2000)

 SWEENEY, Latanya:  
k-anonymity: A model for protecting privacy.  
In: *International Journal of Uncertainty, Fuzziness and  
Knowledge-Based Systems* 10 (2002), Nr. 05, S. 557–570