CS5562: Trustworthy Machine Learning

Part II Lecture 1 $\rightarrow$ What is privacy?

Reza Shokri[a]

Aug 2023

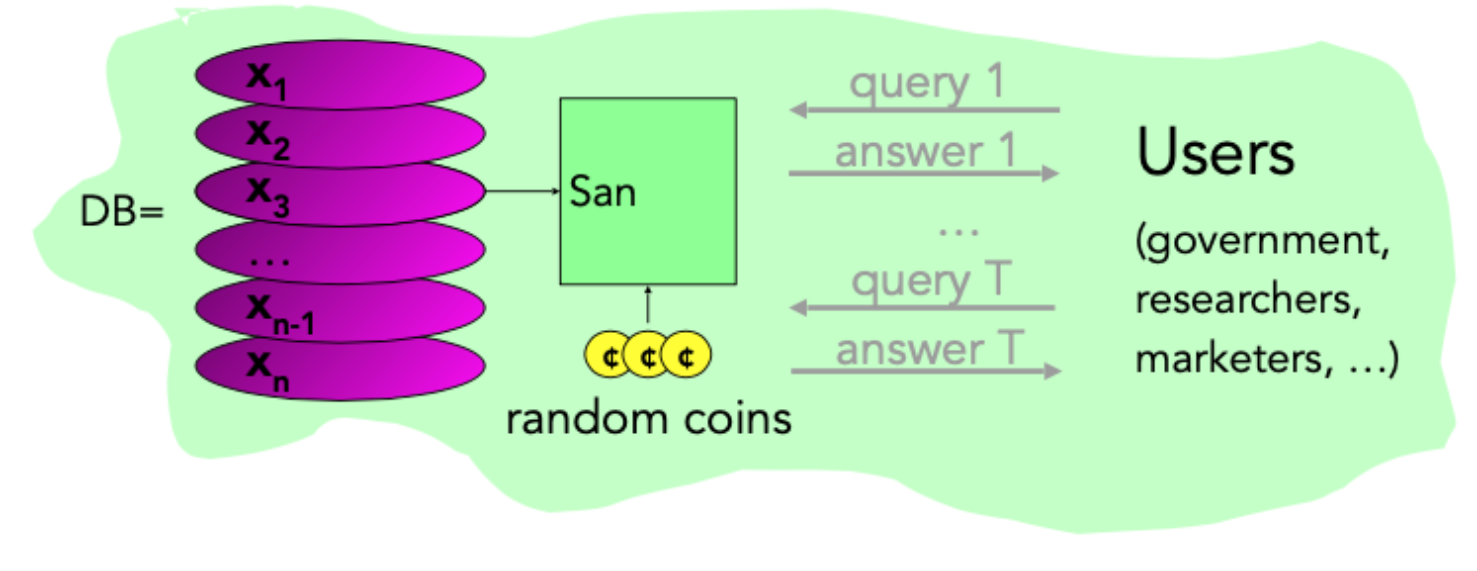
[a]Acknowledgment. The wonderful teaching assistants: Hongyan Chang, Martin Strobel, Jiashu Tao, Yao Tong, Jiayuan Ye

Public Data Conundrum [1]

- Health-care datasets
  - Clinical studies, hospital discharge databases ...
- Genetics datasets
  - $1000 genome, HapMap, DeCODE ...
- Demographic datasets
  - U.S. Census Bureau, sociology studies ...
- Search logs, recommender systems, social networks, blogs ...
  - AOL search data, online social networks, Netflix movie ratings, Amazon ...

[1] Acknowldegement. Slides partially follow lectures given by Vitaly Shmatikov.

Basic Setting



Release useful information about the dataset while preserving privacy.

Examples of Sanitization Methods

- Input perturbation
  - Add random noise to database, release
- Summary statistics
  - Means, variances
  - Marginal totals
  - Regression coefficients
- Output perturbation
  - Summary statistics with noise
- Interactive versions of the above methods
  - Auditor decides which queries are OK, type of noise

Data "Anonymization" and k-anonymity

Data "Anonymization"

- How?
- Remove "personally identifying information" (PII)
  - Name, Social Security number, phone number, email, address ... what else?
- Problem: PII has no technical meaning
  - Defined in disclosure notification laws
    - If certain information is lost, consumer must be notified
  - In privacy breaches, any information can be personally identifying
    - Examples: AOL dataset, Netflix Prize dataset
  - Not very well-defined for unstructured data, e.g., media and audio

Data "Anonymization"

- How?
- Remove "personally identifying information" (PII)
  - Name, Social Security number, phone number, email, address ... what else?
- Problem: PII has no technical meaning
  - Defined in disclosure notification laws
    - If certain information is lost, consumer must be notified
  - In privacy breaches, any information can be personally identifying
    - Examples: AOL dataset, Netflix Prize dataset
  - Not very well-defined for unstructured data, e.g., media and audio

Latanya Sweeney's Attack

- Massachusetts hospital discharge dataset released as anonymous

Medical Data released as anonymous

	SSN	Name	Ethnicity	Date Of Birth	Sex	ZIP	Marital Status	Problem
1			black	09/27/64	male	02139	divorced	obesity
2			black	09/30/64	male	02139	divorced	hypertension
3			black	04/18/64	male	02139	married	chest pain
4			black	04/15/64	male	02139	married	chest pain
• 5			black	09/15/64	male	02138	married	shortness of breath
6			caucasian	03/13/63	male	02141	married	hypertension
7			caucasian	03/18/63	male	02141	married	shortness of breath
8			caucasian	09/13/64	female	02138	married	shortness of breath
9			caucasian	09/07/64	female	02138	married	obesity
10			caucasian	05/14/61	female	02138	single	chest pain
11			caucasian	05/08/61	female	02138	single	obesity

- Public Voter dataset

Voter List

	Name	Address	City	ZIP	DOB	Sex	Race	Party	.....
	.....	.....	.....	.....	.....	.....	.....	.....	.....
	.....	.....	.....	.....	.....	.....	.....	.....	.....
•	Jim A. Cosby	570, Laurel St.	Cambridge	*02138*	*9/15/64*	*male*	*black*	democrat	.....
	.....	.....	.....	.....	.....	.....	.....	.....	.....

Source: [Samarati and Sweeney, 1998]

Observation 1: Dataset Joins

- Attacker learns sensitive data by joining two datasets on common attributes
  - Anonymized dataset with sensitive attributes
    - Example: age, race, symptoms
  - "Harmless" dataset with individual Identifiers
    - Example: name, address, age, race
- Demographic attributes (age, ZIP code, race, etc.) are very common in datasets with information about individuals

Observation 2: Quasi-Identifiers

- Sweeney's observation: (birthdate, ZIP code, gender) uniquely identifies 87% of US population
  - Side note: 63% [Golle, WPES '06]
- Publishing a record with a quasi-identifier is as bad as publishing it with an explicit identity
- Eliminating quasi-identifiers is not desirable
  - For example, users of the dataset may want to study distribution of diseases by age and ZIP code

k-anonymity

- Proposed by [Samarati and Sweeney, 1998]
- Hundreds of papers since then
  - Extremely popular in the database and data mining communities (SIGMOD, ICDE, KDD, VLDB)
- NP-hard in general, but there are many practically efficient k-anonymization algorithms
- Most based on generalization and suppression

Anonymization in a Nutshell

- Dataset is a relational table
- Attributes (columns) are divided into
  - quasi-identifiers (e.g. race, age)
  - and sensitive attributes (e.g. symptoms, medical history)
- Generalize/suppress quasi-identifiers, don't touch sensitive attributes (keep them "truthful")

Example of Generalization/Suppression

Race	Age	Symptoms	Medical History
Person	3*	fever	HIV positive
Person	4*	chest pain	Asthma
...	...	...	...

Anonymization in a Nutshell

- Dataset is a relational table
- Attributes (columns) are divided into
  - quasi-identifiers (e.g. race, age)
  - and sensitive attributes (e.g. symptoms, medical history)
- Generalize/suppress quasi-identifiers, don't touch sensitive attributes (keep them "truthful")

Example of Generalization/Suppression

Race	Age	Symptoms	Medical History
Person	3*	fever	HIV positive
Person	4*	chest pain	Asthma
...	...	...	...

k-anonymity: Definition

- Any (transformed) quasi-identifier must appear in at least k records in the anonymized dataset
  - k is chosen by the data owner (how?)
  - Example: any age-race combination from original database must appear at least 10 times in the anonymized database.
- Guarantees that any join on quasi-identifiers with the anonymized dataset will contain at least k records for each quasi-identifier.

Achieving k-anonymity: Curse of Dimensionality

- Generalization/suppression of quasi-identifiers fundamentally relies on spatial locality
  - Each record must have k close neighbors
- Real-world datasets are very sparse
  - Many attributes (dimensions)
    - Netflix Prize dataset: 17,000 dimensions
    - Amazon customer records: several million dimensions
  - "Nearest neighbor" is very far
- Projection to low dimensions loses all info $\Rightarrow$ k-anonymized datasets become useless

What does k-anonymity mean for privacy?

If an anonymized table is released under k-anonymity...

- **Membership disclosure**: can an attacker guess whether a given person is in the dataset?
- **Sensitive attribute disclosure**: can an attacker tell whether a given person has a certain sensitive attribute (e.g. HIV positive/negative)?
- **Identity disclosure**: can an attacker tell which data record corresponds to a given person?

What does k-anonymity mean for privacy?

If an anonymized table is released under k-anonymity...

- **Membership disclosure**: can an attacker guess whether a given person is in the dataset?
- **Sensitive attribute disclosure**: can an attacker tell whether a given person has a certain sensitive attribute (e.g. HIV positive/negative)?
- **Identity disclosure**: can an attacker tell which data record corresponds to a given person?

↑

This interpretation is correct, assuming the attacker does not know anything other than the quasi-identifiers
But this does not imply any privacy!
Example: k clinical records, all HIV+

A closer look at k-anonymity definition

"Any (transformed) quasi-identifier must appear in at least k records in the anonymized dataset"

- This definition does not mention sensitive attributes at all!
- Assumes that the attacker will be able to join only on quasi-identifiers
- Does not say anything about the computations that are to be done on the data

What does k-anonymity mean for membership disclosure?

- With large probability, quasi-identifier is unique in the population
- But generalizing/suppressing quasi-identifiers in the dataset does not affect their distribution in the population (obviously)!
  - Suppose anonymized dataset contains 10 records with a certain quasi-identifier ...
  - ... and there are only 10 people in the population who match this quasi-identifier
- k-anonymity may not hide whether a given person is in the dataset

What does k-anonymity mean for sensitive attribute?

- Intuitive reasoning:
  - k-anonymity prevents attacker from telling which record corresponds to which person
  - Therefore, attacker cannot tell that a certain person has a particular value of a sensitive attribute
- This reasoning is fallacious!

3-Anonymization

Caucas	78712	Flu
Asian	78705	Shingles
Caucas	78754	Flu
Asian	78705	Acne
AfrAm	78705	Acne
Caucas	78705	Flu

$\Rightarrow$

Caucas	787XX	Flu
Asian	78705	Shingles
Caucas	787XX	Flu
Asian	78705	Acne
AfrAm	78705	Acne
Caucas	787XX	Flu

This is 3-anonymous, right?

Joining With External Database

...	...	...
Rusty Shackleford	Caucas	78705
...	...	...

+

Caucas	787XX	Flu
Asian	78705	Shingles
Caucas	787XX	Flu
Asian	78705	Acne
AfrAm	78705	Acne
Caucas	787XX	Flu

Problem: sensitive attributes are not "diverse"

within each quasi-identifier group

Joining With External Database

...	...	...
Rusty Shackleford	Caucas	78705
...	...	...

+

Caucas	787XX	Flu
Asian	78705	Shingles
Caucas	787XX	Flu
Asian	78705	Acne
AfrAm	78705	Acne
Caucas	787XX	Flu

Problem: sensitive attributes are not "diverse"

within each quasi-identifier group

Another Attempt: ℓ-diversity

Caucas	787XX	Flu
Caucas	787XX	Shingles
Caucas	787XX	Acne
Caucas	787XX	Flu
Caucas	787XX	Acne
Caucas	787XX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Shingles
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Flu

Entropy of sensitive attribute within each quasi-identifier group must be at least L

Source: [Machanavajjhala et al., 2007]

Still Does Not Work

Raw database

...	Flu
...	Flu
...	Cancer
...	Flu
...	Flu
...	Flu
...	Flu
...	Cancer
...	Flu
...	Cancer
...	Flu
...	Flu

80% have Flu

Anonymization A

Q1	Flu
Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Cancer
Q1	Cancer
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu

50% cancer ⇒ quasi-identifier group is "diverse"

This leaks a lot information

Anonymization B

Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Flu
Q1	Flu
Q1	Flu
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Flu

80% flu ⇒ quasi-identifier group is not "diverse"

...yet does not leak much

Still Does Not Work

Raw database

...	Flu
...	Flu
...	Cancer
...	Flu
...	Flu
...	Flu
...	Flu
...	Cancer
...	Flu
...	Cancer
...	Flu
...	Flu

80% have Flu

Anonymization A

Q1	Flu
Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Cancer
Q1	Cancer
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu

50% cancer ⇒ quasi-identifier group is "diverse"

This leaks a lot information

Anonymization B

Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Flu
Q1	Flu
Q1	Flu
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Flu

80% flu ⇒ quasi-identifier group is not "diverse"

...yet does not leak much

Still Does Not Work

Raw database

...	Flu
...	Flu
...	Cancer
...	Flu
...	Flu
...	Flu
...	Flu
...	Cancer
...	Flu
...	Cancer
...	Flu
...	Flu

80% have Flu

Anonymization A

Q1	Flu
Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Cancer
Q1	Cancer
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu

50% cancer ⇒ quasi-identifier

group is "diverse"

This leaks a lot information

Anonymization B

Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Flu
Q1	Flu
Q1	Flu
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Flu

80% flu ⇒ quasi-identifier

group is not "diverse"

...yet does not leak much

Still Does Not Work

Raw database

...	Flu
...	Flu
...	Cancer
...	Flu
...	Flu
...	Flu
...	Flu
...	Cancer
...	Flu
...	Cancer
...	Flu
...	Flu

80% have Flu

Anonymization A

Q1	Flu
Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Cancer
Q1	Cancer
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu

50% cancer $\Rightarrow$ quasi-identifier

group is "diverse"

This leaks a lot information

Anonymization B

Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Flu
Q1	Flu
Q1	Flu
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Flu

80% flu $\Rightarrow$ quasi-identifier

group is not "diverse"

...yet does not leak much

Still Does Not Work

Raw database

...	Flu
...	Flu
...	Cancer
...	Flu
...	Flu
...	Flu
...	Flu
...	Cancer
...	Flu
...	Cancer
...	Flu
...	Flu

80% have Flu

Anonymization A

Q1	Flu
Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Cancer
Q1	Cancer
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu
Q2	Flu

50% cancer $\Rightarrow$ quasi-identifier group is "diverse"

This leaks a lot information

Anonymization B

Q1	Flu
Q1	Cancer
Q1	Flu
Q1	Flu
Q1	Flu
Q1	Flu
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Cancer
Q2	Flu
Q2	Flu

80% flu $\Rightarrow$ quasi-identifier group is not "diverse"

...yet does not leak much

Try Again: t-Closeness

Caucas	787XX	Flu
Caucas	787XX	Shingles
Caucas	787XX	Acne
Caucas	787XX	Flu
Caucas	787XX	Acne
Caucas	787XX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Shingles
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Flu

Distribution of sensitive attributes within each quasi-identifier group should be "close" to their distribution in the original database

Trick question: Why publishing quasi-identifiers at all??

Source: [Li et al., 2006]

Try Again: t-Closeness

Caucas	787XX	Flu
Caucas	787XX	Shingles
Caucas	787XX	Acne
Caucas	787XX	Flu
Caucas	787XX	Acne
Caucas	787XX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Flu
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Shingles
Asian/AfrAm	78XXX	Acne
Asian/AfrAm	78XXX	Flu

Distribution of sensitive attributes within each quasi-identifier group should be "close" to their distribution in the original database

Trick question: Why publishing quasi-identifiers at all??

Source: [Li et al., 2006]

Anonymized "t-close" Database

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

This is k-anonymous, ℓ-diverse and t-close...

...so secure, right?

What Does Attacker Know: join on sensitive attribute

Bob is white and I heard he was admitted to hospital with flu...

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

What Does Attacker Know: join on sensitive attribute

Bob is white and I heard he was admitted to hospital with flu...

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

What Does Attacker Know: join on sensitive attribute

Bob is white and I heard he was admitted to hospital with flu...

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

What Does Attacker Know: join on sensitive attribute

Bob is white and I heard he was admitted to hospital with flu...

This is against the rules! "flu" is not a quasi-identifier

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

What Does Attacker Know: join on sensitive attribute

Bob is white and I heard he was admitted to hospital with flu...

This is against the rules! "flu" is not a quasi-identifier

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

Yes...and this is yet another problem with k-anonymity

What Does Attacker Know: join on sensitive attribute

Bob is white and I heard he was admitted to hospital with flu...

This is against the rules! "flu" is not a quasi-identifier

Caucas	787XX	HIV+	Flu
Asian/AfrAm	787XX	HIV-	Flue
Asian/AfrAm	787XX	HIV+	Shingles
Caucas	787XX	HIV-	Acne
Caucas	787XX	HIV-	Shingles
Caucas	787XX	HIV-	Acne

Yes...and this is yet another problem with k-anonymity

What does the attacker know: link with incomplete public data

- Even under k-anonymity, side knowledge about an individual allows adversary to infer its other remaining sensitive attribute.
- But, is it really practical? What if precise side-information is not available?

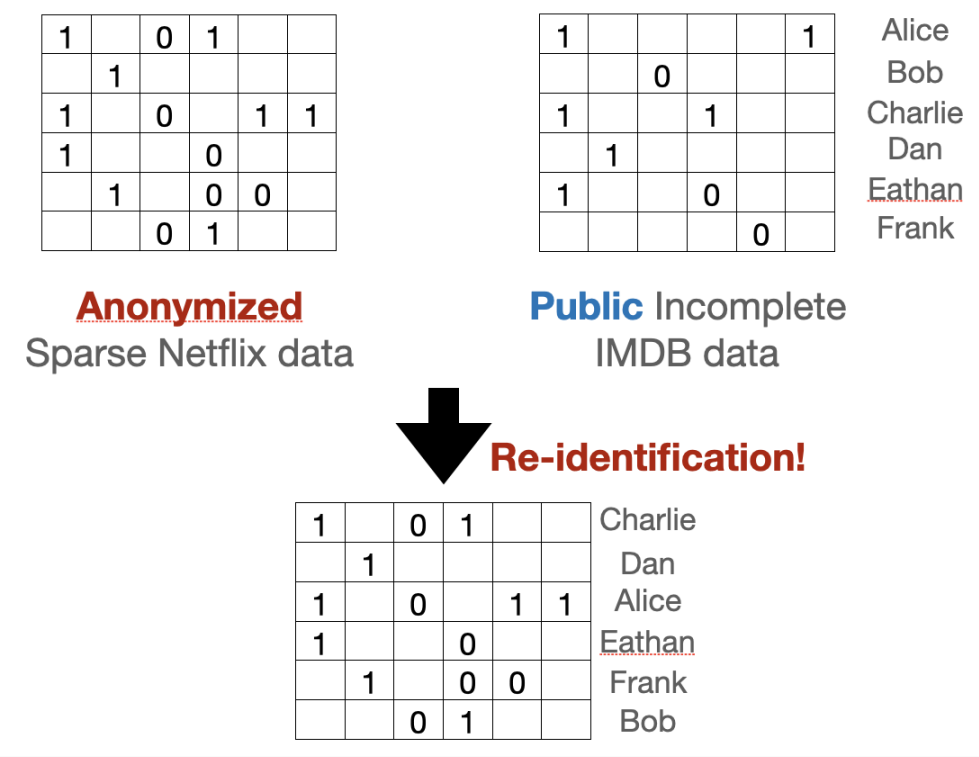
Even incomplete public data suffice for launching re-identification attack!

What does the attacker know: link with incomplete public data

- Even under k-anonymity, side knowledge about an individual allows adversary to infer its other remaining sensitive attribute.
- But, is it really practical? What if precise side-information is not available?

Even incomplete public data suffice for launching re-identification attack!

Netflix Challenge Attack

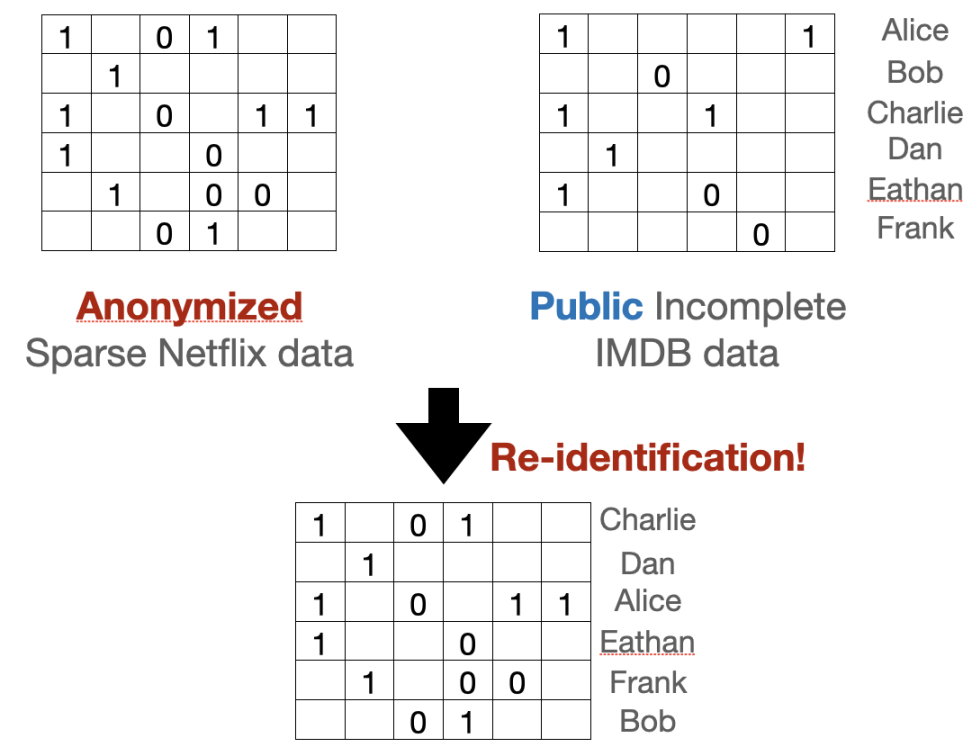


- Sparse Netflix data: not all movies are rated
- Incomplete Public data
- Attack: carefully design similarity score, and find the public record that is most similar to a target anonymized record

With four ratings, the first match between anonymized data and public data has a much higher similarity than second match $\Rightarrow$ Re-identified!

Source: [Narayanan and Shmatikov, 2008]

Netflix Challenge Attack



- Sparse Netflix data: not all movies are rated
- Incomplete Public data
- Attack: carefully design similarity score, and find the public record that is most similar to a target anonymized record

With four ratings, the first match between anonymized data and public data has a much higher similarity than second match $\Rightarrow$ Re-identified!

Source: [Narayanan and Shmatikov, 2008]

What does the attacker know: link with learned model

- Directly comparing an (incomplete) public record with a (complete) anonymized record may not always make sense
  - E.g. what if the two records have no overlapping attributes?
- Could we learn a model about patterns associated with individuals from the public record, and use it to predict other attributes associated with this individual?
- Compute the predictive probability that an anonymized record is from a target user. High probability $\Rightarrow$ Re-identification!

Location Traces

time	...	t	...
Alice	...	visited location	...
Bob	...	...	...
...	...	...	...

Table 1: Raw location traces

What is sensitive about releasing location traces?

- Mobility mode
- Context inference
- Contacts
- Co-traveler

Source: [Shokri et al., 2011]

De-identification of Anonymized Location Traces

time	...	t	...
record 1	...	visited location	...
record 2	...	...	...
...	...	...	...

Table 2: Anonymized location traces

Adversary's knowledge $\mathcal{M}_{person}$ about a known person (easy to obtain via partial observations)

- Mobility model
- Sample traces

$$Pr\left[\text{Identity}(\text{record}_1) = Alice | \mathcal{M}_{person}\right]$$
$$= \frac{Pr\left[\mathcal{M}(record_1) = \mathcal{M}_{Alice} | \text{record}_1 \text{ is Alice's}\right] \cdot Pr[\text{record}_1 \text{ is Alice's}]}{\sum_{person} Pr\left[\mathcal{M}(record_1) = \mathcal{M}_{person} | \text{record}_1 \text{ is person's}\right] \cdot Pr[\text{record}_1 \text{ is person's}]}$$

Source: [Shokri et al., 2011]

Figure 1: A few exposed Location (Work, home, ...)

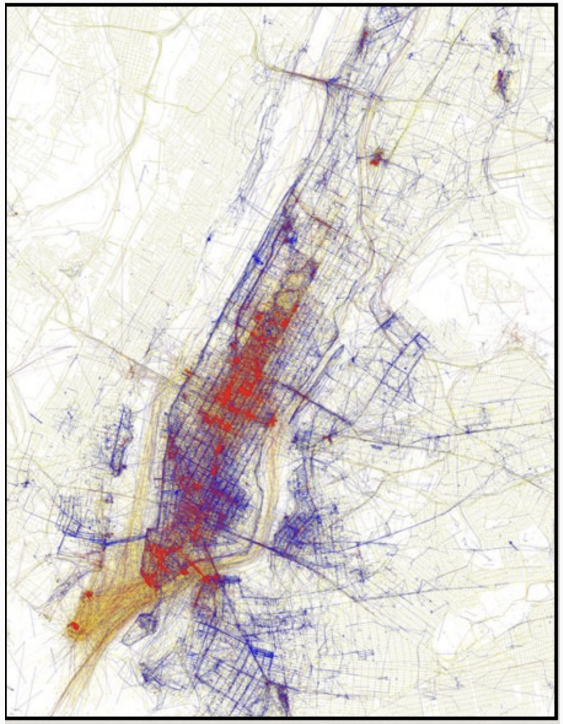
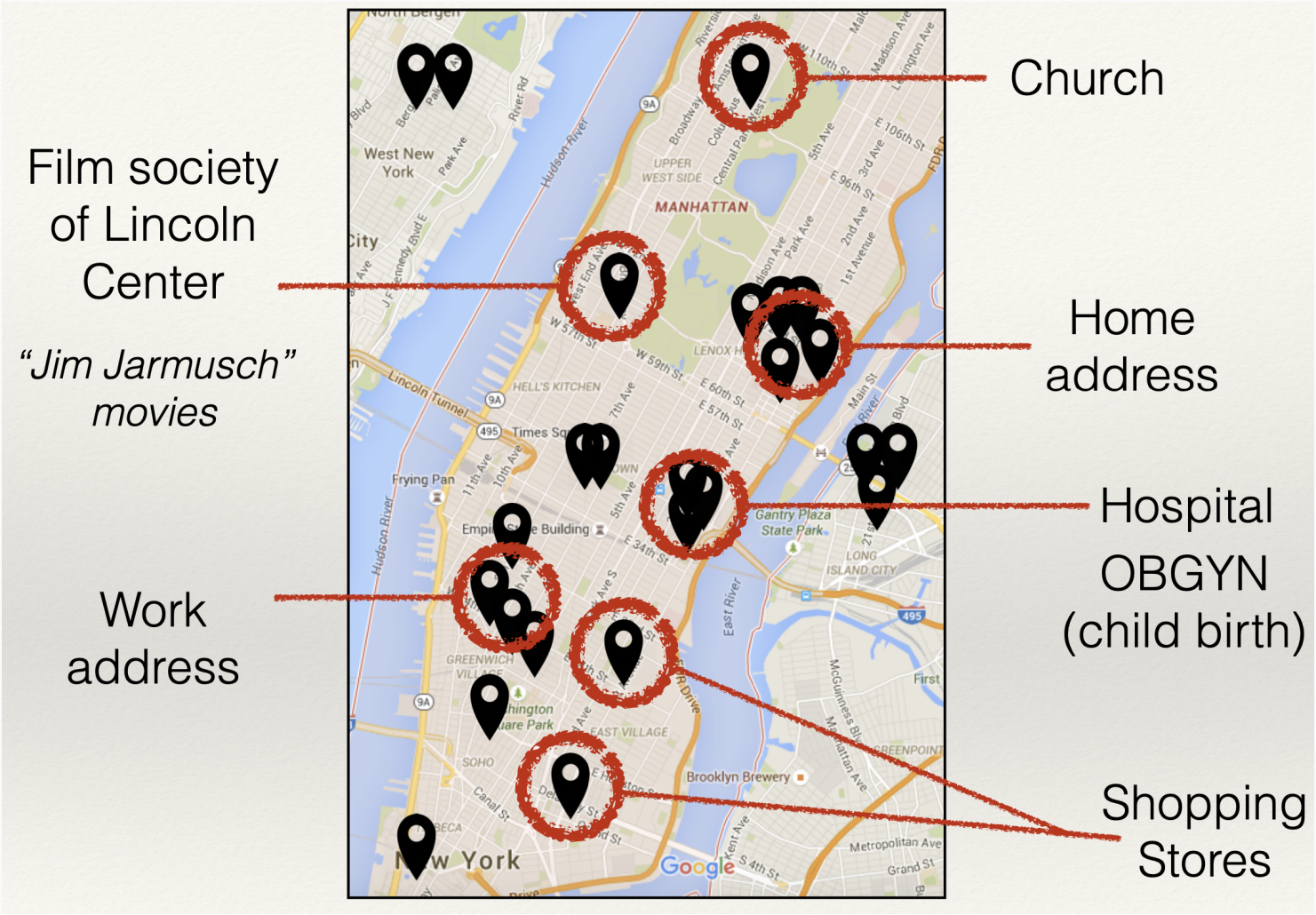


Figure 2: Anonymized Location Datasets

Figure 3: This person's full trajectory



Information leakage in re-identified trajectory: Work, Home, Religion, Life events, Interests, Wealth

What does the attacker know: Joining attack on released $k-$ anonymous tables

- Up to now: precise/approximate/abstract external knowledge suffice for launching re-identification attacks.
- What if the external knowledge is also k-anonymous?

Still extremely vulnerable to re-identification attack!

Composition of k-anonymous tables fails to be k-anonymous

Two hospitals release anonymized tables about their patients.

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

(b)

(a) satisfies 4-anonymity and (b) satisfies 6-anonymity.

If Alice's employer knows that she is 28, lives in zip code 13012 and visits both hospitals, what could the employer infer about Alice?

Source: [Ganta et al., 2008]

Composition of k-anonymous tables fails to be k-anonymous

Two hospitals release anonymized tables about their patients.

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

(a)

	Non-Sensitive			Sensitive
	Zip code	Age	Nationality	Condition
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

(b)

(a) satisfies 4-anonymity and (b) satisfies 6-anonymity.

If Alice's employer knows that she is 28, lives in zip code 13012 and visits both hospitals, what could the employer infer about Alice?

Source: [Ganta et al., 2008]

Failure of k-anonymity in practice

- Use LinkedIn.com to re-identify students in a k-anonymized dataset published by EdX.
- EdX data: demographics, engagement, and final grade. Released 476,532 students' records under k-anonymity with $k = 5$.
- Attack: search matching user on LinkedIn for 135 anonymized students' records. (People post resumes which could be read by a recruiter account costing $ 119.95 for a month.)
- All three re-identified students failed at least one course (which they did not reveal on LinkedIn), thus revealing private facts.

Source: [Cohen, 2022]

Summary

- Re-identification attacks are possible by exploiting external knowledge/previously released tables.
  - However, syntactic definitions such as k-anonymity do not preserve privacy under external knowledge/previously released tables.
  - Consequently, private information (membership, sensitive attribute) may still be leaked even when k-anonymity is achieved for all the (previously) released table.
- We need better methodology and definition for preserving privacy.

References i

Cohen, A. (2022).
Attacks on deidentification's defenses.
arXiv preprint arXiv:2202.13470.

Ganta, S. R., Kasiviswanathan, S. P., and Smith, A. (2008).
Composition attacks and auxiliary information in data privacy.
In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 265–273.

Li, N., Li, T., and Venkatasubramanian, S. (2006).
t-closeness: Privacy beyond k-anonymity and l-diversity.
In 2007 IEEE 23rd international conference on data engineering, pages 106–115. IEEE.

References ii

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007).

l-diversity: Privacy beyond k-anonymity.

ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1):3–es.

Narayanan, A. and Shmatikov, V. (2008).

Robust de-anonymization of large sparse datasets.

In 2008 IEEE Symposium on Security and Privacy (sp 2008), pages 111–125. IEEE.

References iii

Samarati, P. and Sweeney, L. (1998).

Generalizing data to provide anonymity when disclosing information.

In PODS, volume 98, pages 10–1145.

Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., and Hubaux, J.-P. (2011).

Quantifying location privacy.

In 2011 IEEE symposium on security and privacy, pages 247–262. IEEE.