Data Privacy and Security

Project Ideas

Asst. Prof. M. Emre Gürsoy
www.memregursoy.com

General Advice

- Choose something that you'll enjoy
  - You're devoting a long time to it
  - Your course grade depends heavily on it (40%+5%)
- Choose something that'll be useful for you
  - Think about whether you want to have a «Github portfolio» or «research experience»
  - If you're already doing research, think about how your project may benefit from your domain expertise
- If you want a long-term outcome (publication, app, research credits in future semesters, ...), let me know
  - You're already putting a lot of work into your project
  - Can decide to go the extra mile (depends)
  - But your project should be shaped accordingly

General Advice

- Make weekly progress throughout the semester
  - Impossible to do 3 months of work in 3 weeks
  - If something doesn't go according to plan, you can change sooner rather than later
- Think about how much help you'll need from me
  - I have more knowledge and expertise in some topics compared to others
  - If you choose these topics, I can help more
  - E.g.: project numbers 2, 4, 6, 8, 9, 10, 11, 12
  - Instead, if you choose topics in which I don't have as much expertise, I can't help with the technical details (eg: acoustics, genomics) – you're on your own

Acoustic Keyloggers

1

- We type our passwords using a keyboard
  - Assume a shared computer with a keyboard
- Keylogger: software or hardware that logs keystrokes
  - Each key on your keyboard makes a slightly different sound -> acoustics
  - Humans have average typing speed (or motion)
  - Sounds+speed can be used to create a keylogger

Record key voices from shared keyboard at attack training time

ML model

Victim's key acoustics are recorded

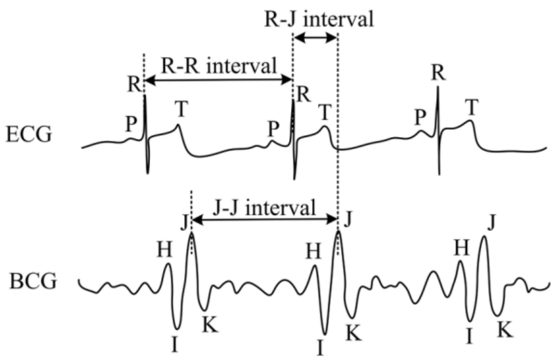
ML model outputs which keys the victim typed based on the acoustics

Acoustic Keyloggers

- Suitable for students who have knowledge/interest in acoustics, machine learning, signal processing.
  - Need to extract relevant acoustic and motion / keystroke frequency features from signals.
- You can record key sounds on your keyboard + your group members' keyboards (proof-of-concept).
  - One ML model for each keyboard
- Even more interesting:
  - Apple Magic keyboard
  - Portable keyboards
  - New gaming keyboards

Biometric Authentication

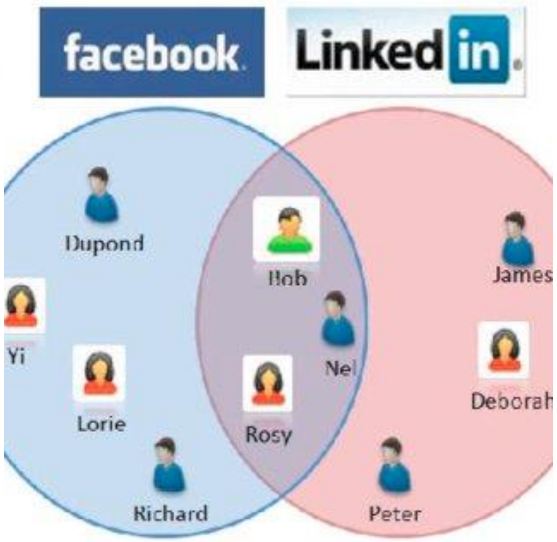
- Human body emits physiological signals
  - BCG, ECG, SCG, ...
  - Often unique from human-to-human



- Can biological signals be used as biometrics?
  - Applications: smartwatch, fitbit, fitness devices
  - Signal processing + machine learning
- Data sources:
  - UnoViS: https://www.medit.hia.rwth-aachen.de/publikationen/unovis/
  - PhysioNet: https://physionet.org/
  - WISDM: **https://archive.ics.uci.edu/ml/datasets/WISDM+Smartphone+and+Smartwatch+Activity+and+Biometrics+Dataset+**

Profile Matching

- A user has profiles on multiple social media sites:
  - A professional Facebook account with their name
  - An anonymous Twitter account (casual/activist)
- Profile matching: Match the user's anonymous Twitter account with their non-anonymous Facebook account
- How?
  - Photos
  - Workplace/education
  - Cross-posted content
  - Friend lists
  - Writing style

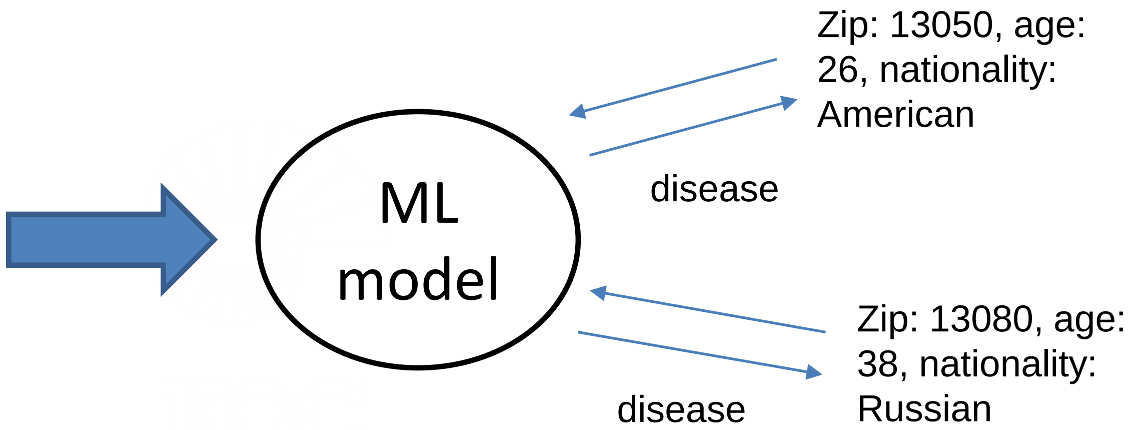


ML w/ Anonymized Data

- Typical ML pipeline:

Zip	Age	Nationality	Disease
13053	28	Russian	Heart
13068	29	American	Heart
13068	21	Japanese	Flu
13053	23	American	Flu
14853	50	Indian	Cancer
14853	55	Russian	Heart
14850	47	American	Flu
14850	59	American	Flu

Training data



- Anonymization generalizes the training data:

Zip	Age	Nationality	Disease
13053	28	Russian	Heart
13068	29	American	Heart
13068	21	Japanese	Flu
13053	23	American	Flu
14853	50	Indian	Cancer
14853	55	Russian	Heart
14850	47	American	Flu
14850	59	American	Flu

Anonymization

Zip	Age	Nationality	Disease
130**	<30	*	Heart
130**	<30	*	Heart
130**	<30	*	Flu
130**	<30	*	Flu
1485*	>40	*	Cancer
1485*	>40	*	Heart
1485*	>40	*	Flu
1485*	>40	*	Flu

ML w/ Anonymized Data

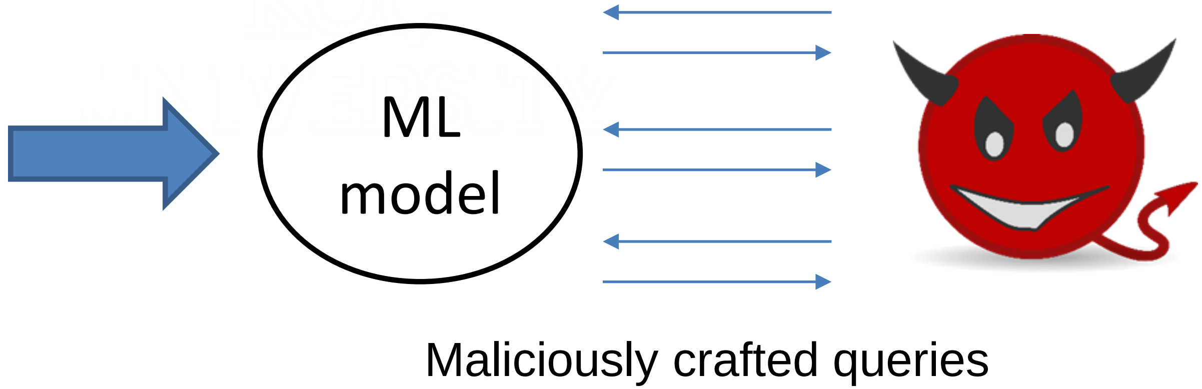
- How can we use the anonymized data for ML?
  - Some values are generalized: 13053 → 130**
  - Some values are suppressed (nationality)
- Should we generalize test data as well?
- Should we «re-construct» anonymized data?
- What design decisions/assumptions do we need to make?

- What is the accuracy impact of training ML models on anonymized data vs non-anonymized data?
  - Usually generalization and suppression cause information loss, thus accuracy is reduced

Anonymization vs ML Attacks

- ML attacks via carefully crafted test queries:
  - Membership inference attacks – was Alice's data used in training the ML model?
  - Model inversion attacks – reconstruct the original database from query answers

Zip	Age	Nationality	Disease
13053	28	Russian	Heart
13068	29	American	Heart
13068	21	Japanese	Flu
13053	23	American	Flu
14853	50	Indian	Cancer
14853	55	Russian	Heart
14850	47	American	Flu
14850	59	American	Flu



Can we thwart these attacks by building the ML model on anonymized data (rather than the original training data)?

Poisoning Attacks

- Training-time attack on ML
  - Training data contains malicious records
  - If we directly train a ML model on malicious data, we may end up with low accuracy (or bad behavior)
  - Remember the Gmail spam filter example?

Skewing Gmail's spam filter using fake spam/non-spam reports

A data poisoning (data pollution) attack

Poisoning Attacks

- Medical domain is considering ML-powered solutions
- Three sources of poisoning:
  - (1) Malicious adversaries
  - (2) Erroneous/imprecise measurements
  - (3) Inherent errors in medical testing – type-I and type-II
- These all cause medical data to be imperfect
- What happens when imperfect data is used for training a ML model in the medical domain?
  - Study different causes of imperfection (1-2-3 above)
  - Different ML models (DT, NB, DNN, SVM, kNN, ...)
  - Different datasets and classification tasks

Poisoning Attacks

- Create new poisoning attack strategies for:
  - Association rule learning (ARL)
  - Recommender systems
  - Time-series analytics/forecasting algorithms
  - ...
- You should study the popular algorithms for whichever task you choose (eg: FP-growth, Apriori for ARL)
- Then determine how you can «fool» the algorithms with as few data insertions/deletions/modifications as possible (ie: as little poisoning as possible)
- Do better than the naive baseline of adding/deleting whatever record that contains your target rule

Poisoning Attacks

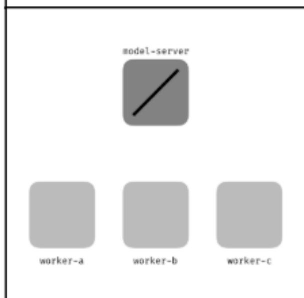
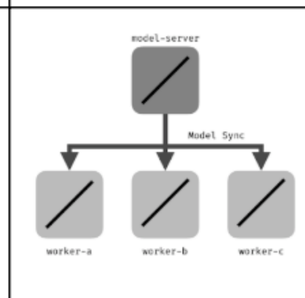
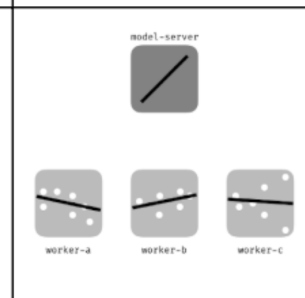
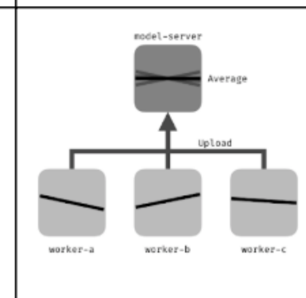
- Naive baseline attack:
  - I want to reduce conf(bread => butter)
  - I remove t_4 – reduces supp(bread, butter)
  - Or I add many transactions like t_5 – increases supp(bread)

conf(bread => butter)
= supp(bread,butter) / supp(bread)
= 0.2 / 0.6
= 1/3

Trans. ID	Bought Items
t_1	milk, bread
t_2	butter
t_3	beer, diapers
t_4	milk, bread, butter
t_5	bread

Federated Poisoning

- Federated learning (FL):
  - https://en.wikipedia.org/wiki/Federated_learning

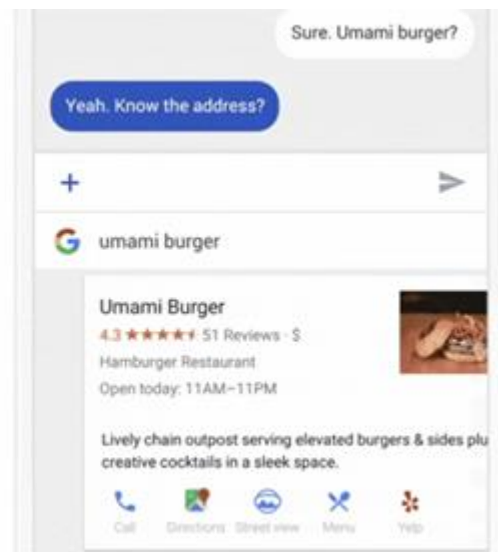
Step 1	Step 2	Step 3	Step 4
			
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

What if one or more nodes are malicious, and they train on poisoned data? → FEDERATED POISONING

«Data poisoning attacks against federated learning systems» – V. Tolpegin et al. Georgia Tech 2nd year undergrad student, year-long research project, received undergrad research award @ Georgia Tech, final paper published in A-level conference in September 2020.

Federated Poisoning

- In our prior work, we showed the effectiveness of federated poisoning on image classification
  - Popular task for neural networks these days
- But FL is / will be used in many other areas:
  - Gboard on Android – Google Keyboard
  - Digital health
  - IoT and Industry 4.0
  - NLP, sentiment analysis



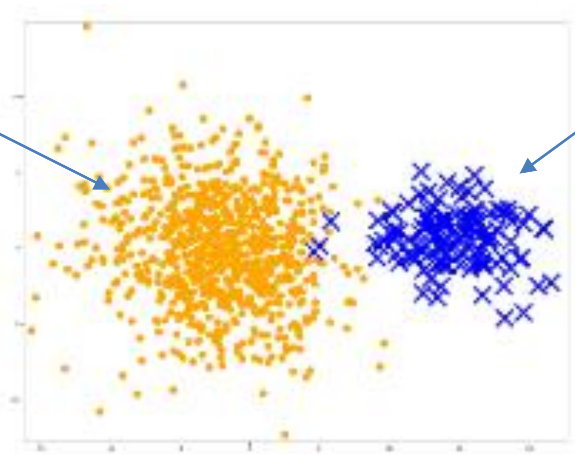
(1) Investigate the impact of federated poisoning in new application areas

(2) Find more effective federated poisoning attack strategies (or application-specific strategies)

Federated Poisoning

- Find defenses against federated poisoning
  - Can we separate malicious clients' updates from non-malicious clients' updates?
  - Can we use strategies from data privacy literature (eg: differential privacy) or distributed systems literature? (eg: Byzantine fault tolerance)

Honest participants' updates



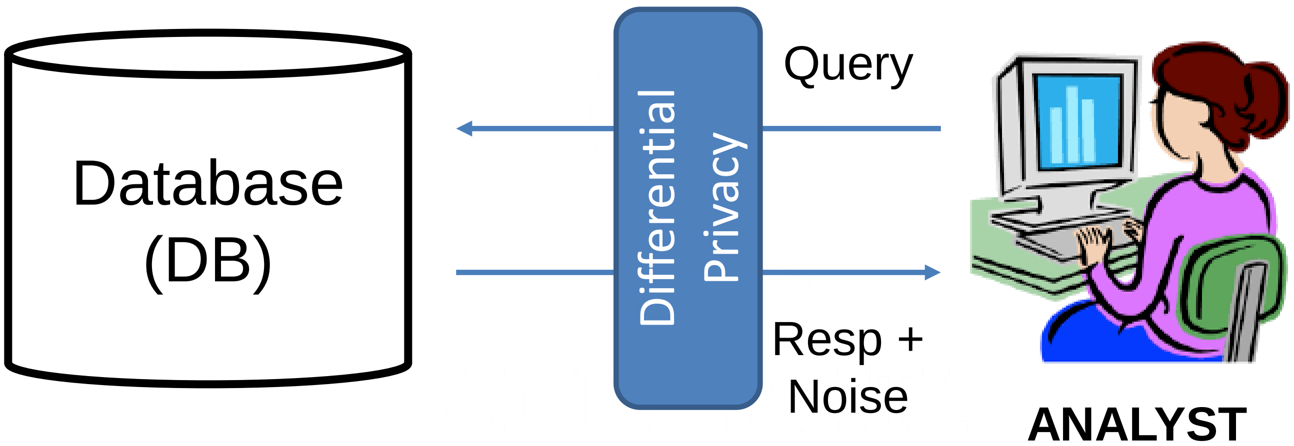
Malicious participants' updates

Separability results

Note: Federated learning experiments rely on neural networks, which take a long time to simulate on commodity hardware.

DP System Design

- Differential privacy (DP) is a popular privacy definition for privacy-preserving analysis of sensitive data:



- You can implement a system with this architecture for many kinds of data: medical data, genomics, education data, web statistics, location data...
  - Most of these are relevant especially due to COVID!

DP System Design

- Advice:
  - (1) Make sure you have large enough datasets
    - Tens or hundreds of thousands of records
    - Better if you have 2-3 datasets, each becomes a different experiment
    - Look for datasets online
  - (2) You can build your system using existing libraries or code them on your own
    - https://github.com/IBM/differential-privacy-library (IBM)
    - https://opendifferentialprivacy.github.io/ (Microsoft+Harvard)
    - https://github.com/google/differential-privacy (Google)
  - (3) Think about what statistics + services you may offer to analysts with differential privacy
    - Use them as proof-of-concept demonstrations for your system
    - Non-private accuracy vs private accuracy

Local DP (LDP)

- Local Differential Privacy (LDP) is used by major tech companies to collect user data.
  - Apple iOS devices – MacOS, iPhone, ...
  - Microsoft Windows
  - Google Chrome
  - ...

Apple's 'Differential Privacy' Is About Collecting Your Data—But Not Your Data

Microsoft Research Blog

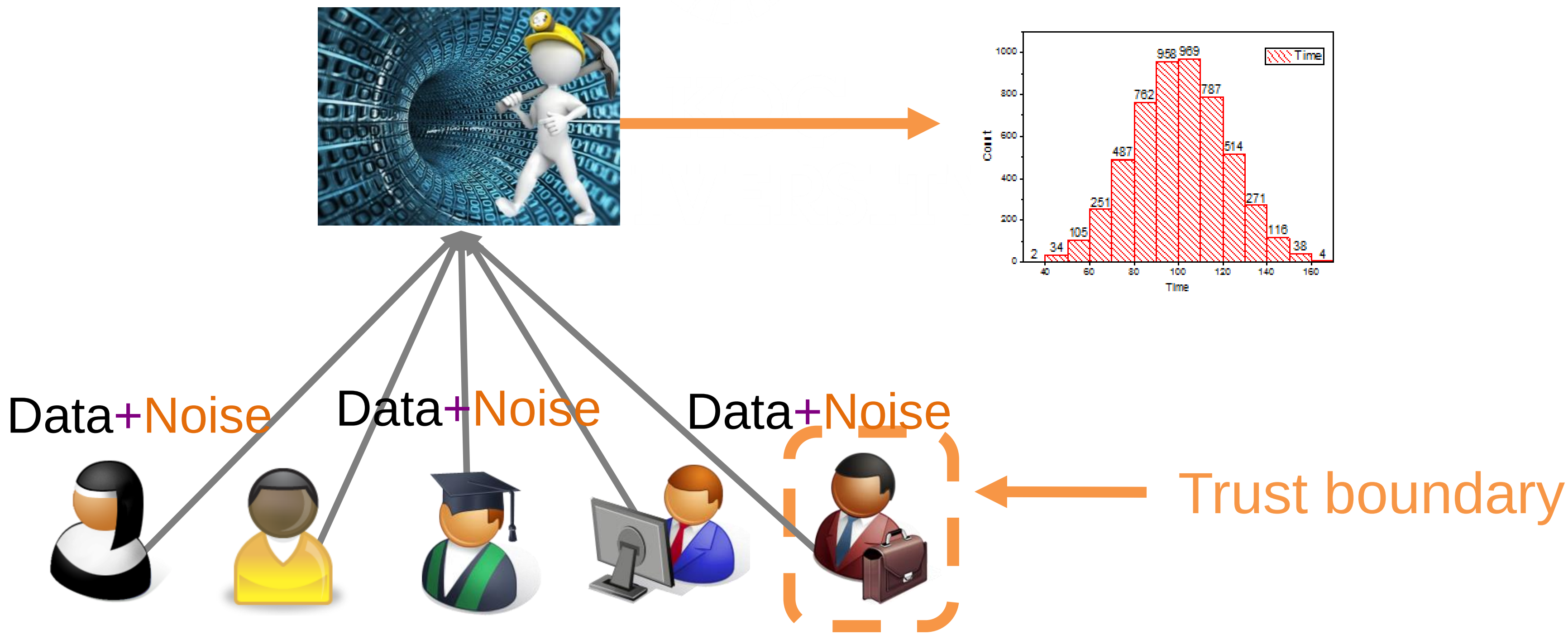
Collecting telemetry data privately

December 8, 2017 | By Bolin Ding, Researcher; Jana Kulkarni, Researcher; Sergey Yekhanin, Sr Principal Researcher

Windows 10

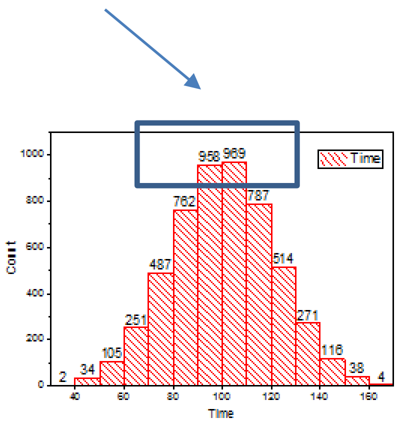
Local DP (LDP)

- Frequency estimation: a fundamental primitive
  - Each user has one or more items from domain D
  - Estimate the frequency (supp) of each item in D



Local DP (LDP)

- There are many protocols for finding heavy hitters (top-k popular items) with LDP

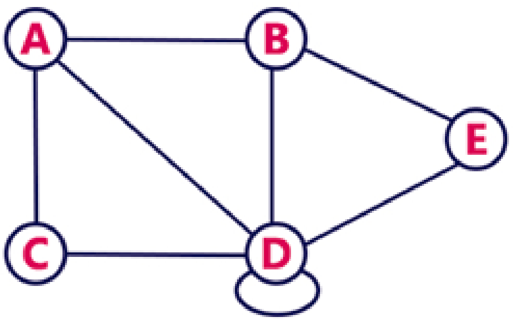


- But how about low-frequency items?
  - The «opposite» of heavy hitters
  - Discovering unused items
  - Which series/movies do Netflix users NOT prefer to watch?

- Sample problem formulations:
  - Finding bottom-k items
  - Finding items w/ frequency lower than threshold T

LDP on Graphs

- Graphs often encode relationships which are sensitive
- Perturb neighbor lists with LDP to hide relationships
  - A's perturbed neighbor list says she's not friends w/ B
  - But how about B's perturbed neighbor list??



	actual		perturbed
A:	0 1 1 1 0	$\xrightarrow{\text{LDP}}$	0 0 1 0 1
B:	1 0 0 1 1	$\xrightarrow{\text{LDP}}$	1 0 0 1 0

- Another privacy leakage: **A** must know which users exist in the network in order to construct her neighbor list

Bias + Fairness

- Tangentially related topic to privacy + security
  - Acceptable as course project
- As AI/ML becomes pervasive, algorithmic bias and dataset bias become important
  - Criminal justice: should a defendent receive bail?
  - Black persons are more likely to be wrongly labeled as «high-risk» and be denied bail

	White	Black
Wrongly Labeled High-Risk	23.5%	44.9%
Wrongly Labeled Low-Risk	47.7%	28.0%

`https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing`

Bias + Fairness

- Many metrics to measure bias and fairness
- Many methods to make «biased» ML methods «unbiased»
- Suitable for benchmarking or finding new applications:
  - https://github.com/Trusted-AI/AIF360
  - https://github.com/dssg/aequitas
  - ... many more

Q & A