Differential Privacy and Risk Ratios: The semantics of privacy

CompSci 590.03

Instructor: Ashwin Machanavajjhala

Differential Privacy

[Dwork ICALP 2006]

For every pair of inputs that differ in one row

For every output …

D_1 D_2 O

Adversary should not be able to distinguish between any D_1 and D_2 based on any O

$$\log\left(\frac{\Pr[A(D_1) = O]}{\Pr[A(D_2) = O]}\right) < \varepsilon \quad (\varepsilon > 0)$$

Privacy Desiderata

- Privacy of an individual is some measure of information leaked by A(D) in comparison to A(*D without that individual*)

- Privacy should be ensured even if adversary has background knowledge

- Privacy mechanisms should compose (and not degrade under postprocessing)

- Privacy should not be achieved by obscurity

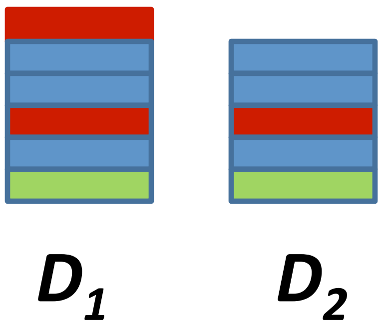
Does differential privacy satisfy all these desiderata?

Privacy Desiderata

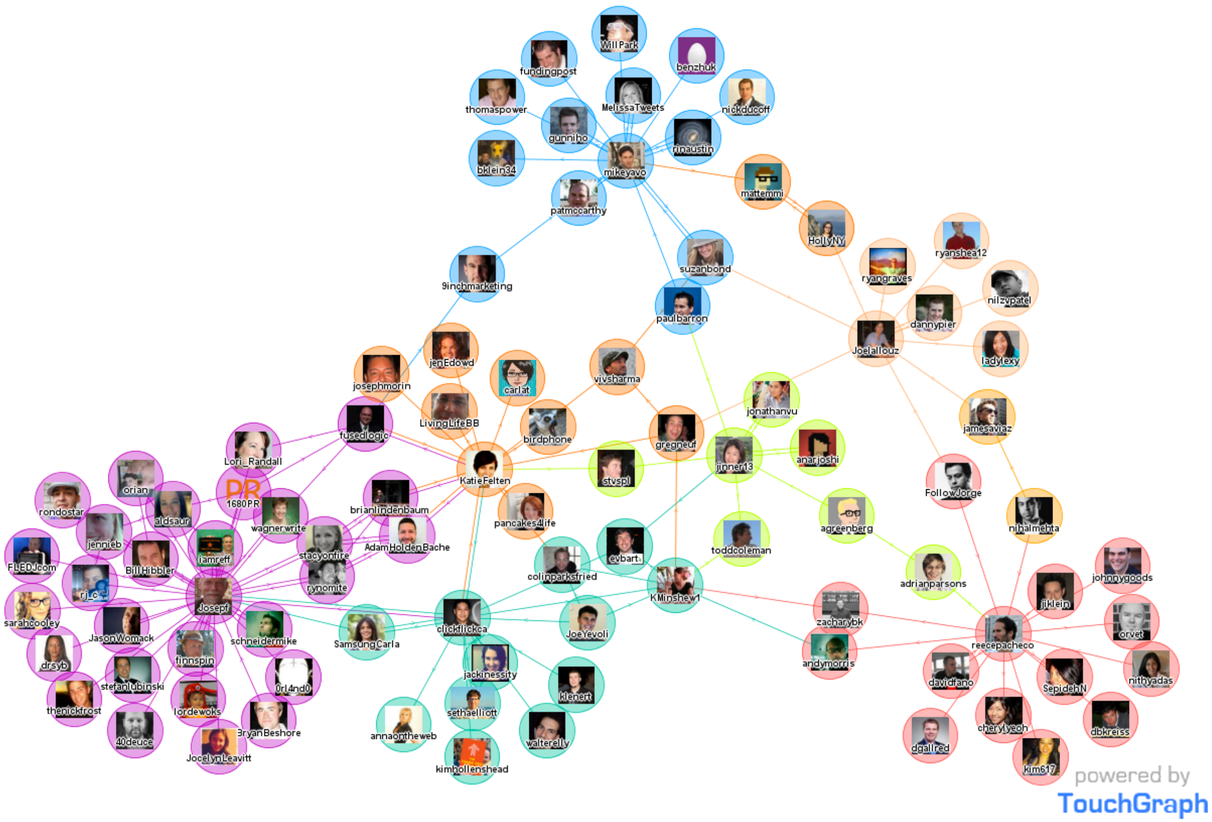
- Privacy of an individual is some measure of information leaked by A(D) in comparison to A(*D without that individual*)
- Privacy should be ensured even if adversary has background knowledge
- Privacy mechanisms should compose (and not degrade under postprocessing) ✓
- Privacy should not be achieved by obscurity ✓

Neighboring databases

For every pair of inputs
that differ in one row



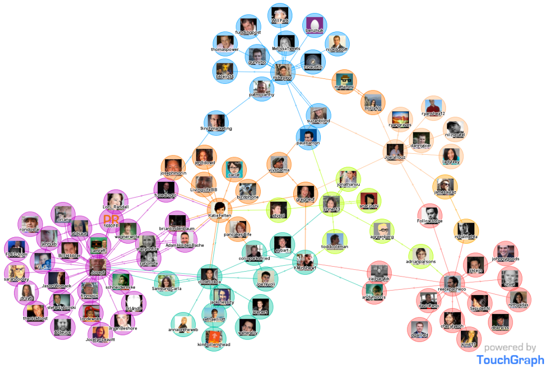
What are neighboring databases for … ?



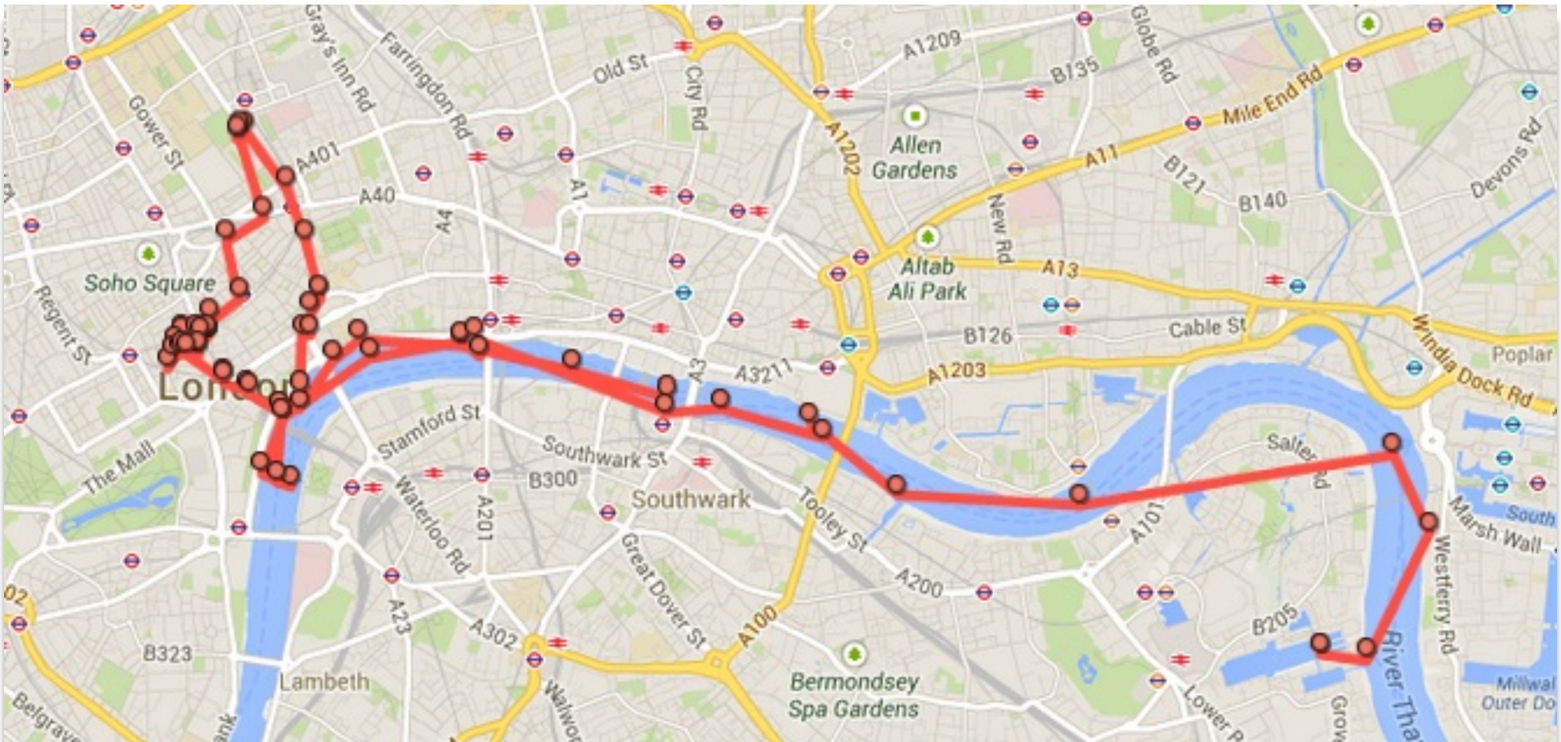
Neighboring Databases …

… differ in one record.

- In graphs, a record can be:
  - An edge (u,v)
  - The adjacency list of node u



What are neighboring databases for …

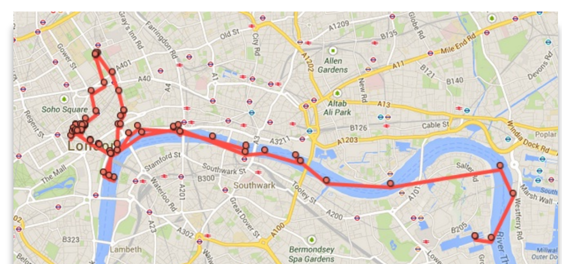


Duke UNIVERSITY

Neighboring Databases …

… differ in one record.

- In location trajectories, a record can be:
  - Each location in the trajectory
  - A sequence of locations spanning a window of time
  - The entire trajectory



What do different neighbor definitions mean?

The semantics of privacy

- Suppose we did not want an adversary to tell whether or not an individual record was in or out of the table.

- Formally,

  Let $\theta(r)$ be adversary's prior over whether record r is in the table
  Let X denote the domain of record r

Single Record Computation Case

- Let A be a computation on the single record r
- Let y = A(r) be the output of the computation.

- Does not make sense for a computation to work on no records.

$$\max_{x_1, x_2 \in X} \max_{y \in range(A)} \frac{\Pr[A(x_1) = y]}{\Pr[A(x_2) = y]} \leq e^{\varepsilon}$$

That is, given any output, one can't distinguish between any two possible values that the record can take.

Duke UNIVERSITY

Adversary's odds

- Do not want an adversary to be able to tell whether or not a record satisfies any property (male vs female, red vs blue, etc).

- Any property of a record can be captured by a set of values S
- The adversary's odds that record r has a value in S is:

$$\frac{\Pr[r \in S \mid A(r) = y]}{\Pr[r \notin S \mid A(r) = y]}$$

Posterior Odds

$$\frac{\Pr[r \in S]}{\Pr[r \notin S]}$$

Prior Odds

Bayes Risk Ratio

- Do not want an adversary to be able to tell whether or not a record satisfies any property (male vs female, red vs blue, etc).

- Bayes Risk Ratio:

$$\max_{S \subset X} \max_{y \in range(A)} \frac{\Pr[r \in S \mid A(r) = y] / \Pr[r \in S]}{\Pr[r \notin S \mid A(r) = y] / \Pr[r \notin S]} \leq e^{\varepsilon}$$

That is, the ratio of the adversary's posterior odds that r is in S versus r is not in S and his prior odds is bounded for all S and for all outputs y.

An equivalence?

A satisfies ε-differential privacy

if and only if

A has Bayes risk bounded by exp(ε)

Independent of the adversary's prior!

DP => Bounded Bayes Risk

$$\frac{\Pr[r \in S \mid A(r) = y]/\Pr[r \in S]}{\Pr[r \notin S \mid A(r) = y]/\Pr[r \notin S]}$$

$$= \frac{\sum_{x \in S} \Pr[r = x \mid A(r) = y]/\Pr[r \in S]}{\sum_{x \notin S} \Pr[r = x \mid A(r) = y]/\Pr[r \in S]}$$

$$= \frac{\sum_{x \in S} \Pr[A(x) = y]\Pr[r = x]/\Pr[A(r) = y]\Pr[r \in S]}{\sum_{x \notin S} \Pr[A(x) = y]\Pr[r = x]/\Pr[A(r) = y]\Pr[r \in S]}$$

$$\leq \max_{x1,x2 \in X} \frac{\Pr[A(x1) = y]}{\Pr[A(x2) = y]} \frac{\sum_{x \in S} \Pr[r = x]/\Pr[A(r) = y]\Pr[r \in S]}{\sum_{x \notin S} \Pr[r = x]/\Pr[A(r) = y]\Pr[r \in S]}$$

$$= e^{\varepsilon}$$

Bounded by DP

Cancels out

Bounded Bayes Risk => DP

- For every pair of values x1, x2 in X, consider an adversary whose prior is: Pr[r = x1] = p and Pr[r = x2] = 1-p

- Let S = {x1}, then

$$\frac{\Pr[r \in S \mid A(r) = y]/\Pr[r \in S]}{\Pr[r \notin S \mid A(r) = y]/\Pr[r \notin S]}$$
$$= \frac{\Pr[r = x1 \mid A(r) = y]/\Pr[r = x1]}{\Pr[r = x2 \mid A(r) = y]/\Pr[r = x2]}$$
$$= \frac{\Pr\left[A(r) = y \middle| r = x1\right]}{\Pr\left[A(r) = y \middle| r = x2\right]} = \frac{\Pr\left[A(x1) = y\right]}{\Pr\left[A(x2) = y\right]}$$

- Since Bayes Risk is bounded, DP is ensured.

Extending to databases

- Suppose we did not want an adversary to tell whether or not an individual record was in or out of the table.

- Formally,

  Let θ be adversary's prior over *the entire database*
  Let X denote the domain of each record r in the database

Bayes risk

- Let A be a computation on the entire database D
- Let y = A(D) be the output of the computation.

- Bayes Risk:

$$\max_{r \in X, D} \max_{y \in range(A)} \frac{\Pr[r \in D \mid A(D) = y] / \Pr[r \in D]}{\Pr[r \notin D \mid A(D) = y] / \Pr[r \notin D]} \leq e^{\varepsilon}$$

An equivalence

An algorithm A satisfies ε-differential privacy

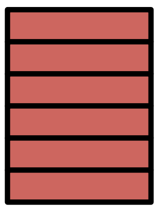
if and only if

A has Bayes risk bounded by exp(ε)

NO

Example

- Adversary thinks there are only two databases with equal probability



- But adversary can tell whether a record is red or blue after seeing output of algorithm that uses Laplace mechanism to release number of red records.

An equivalence

An algorithm A satisfies ε-differential privacy

if and only if

A has Bayes risk bounded by exp(ε)

For an adversary who thinks the records are independent!

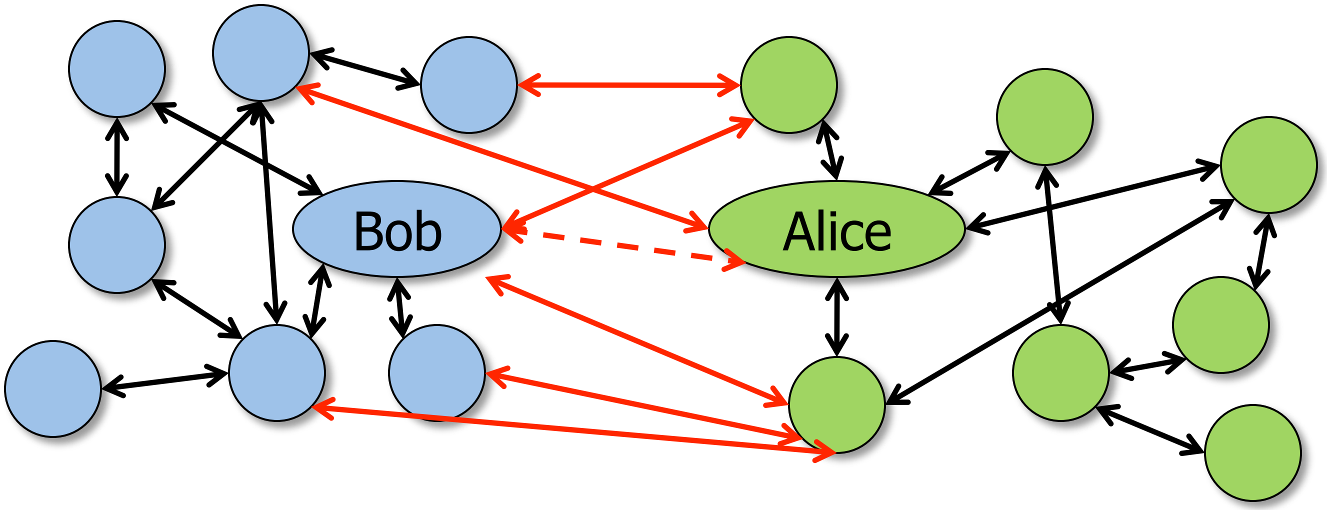
Consequences

1. Choose what is a record carefully. The privacy guarantee is about the record.

2. Is there a better definition than differential privacy that protects against all adversaries in terms of Bayes Risk?

3. Is the independence assumption valid?

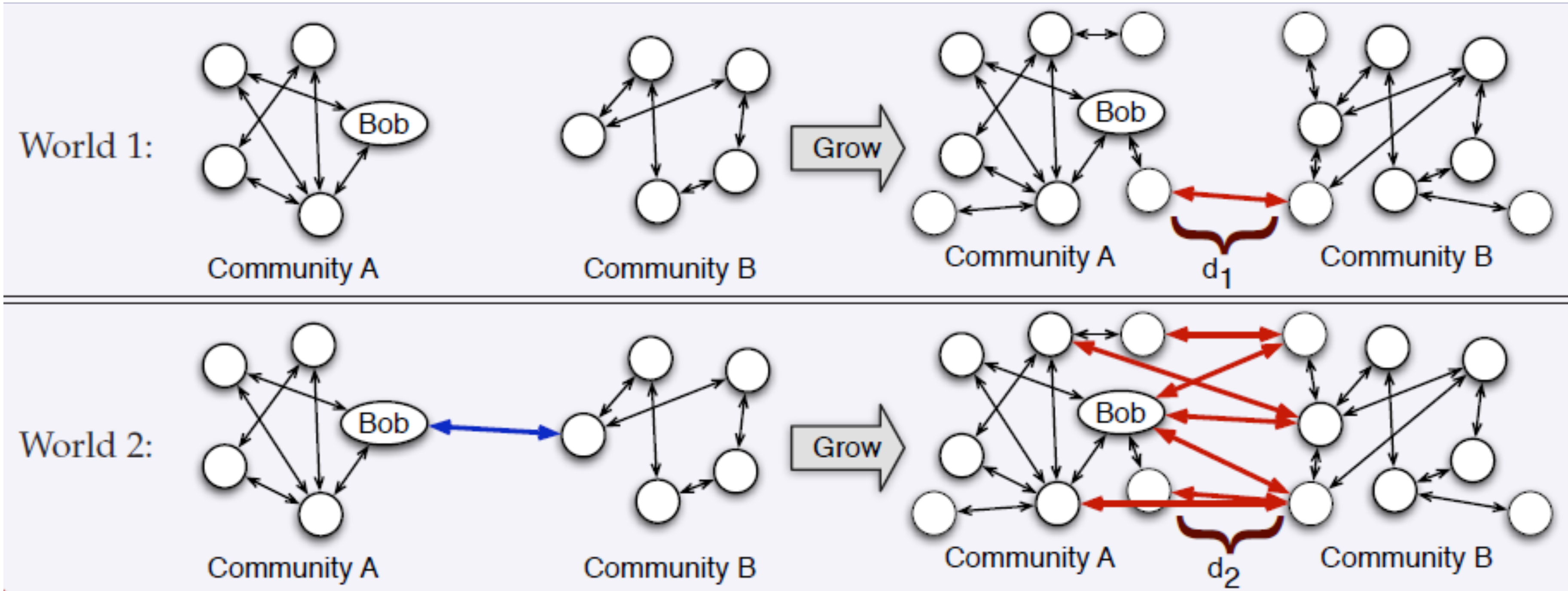
Correlations and DP



- Want to release the number of edges between **blue** and **green** communities.
- Should not disclose the presence/absence of Bob-Alice edge.

[KM11]

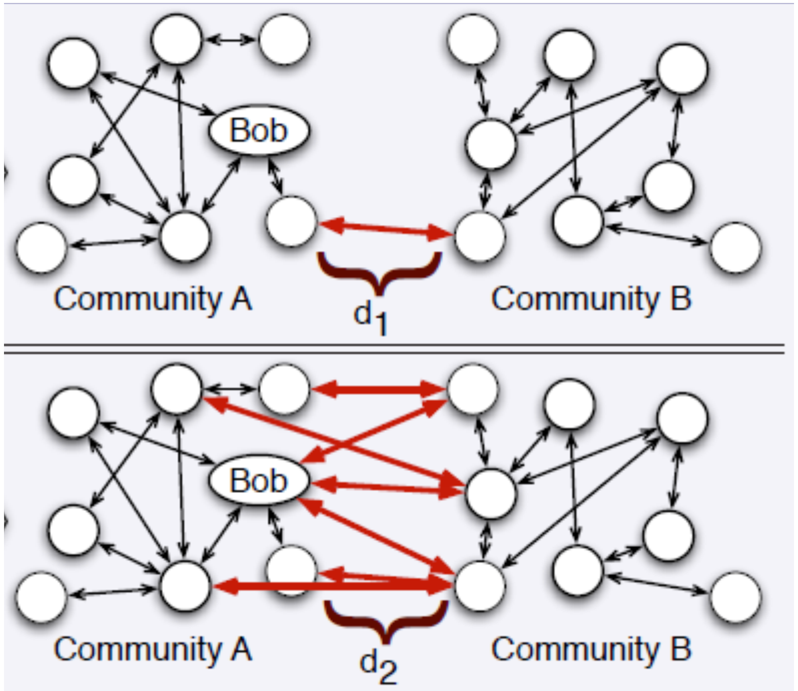
Adversary knows how social networks evolve



Depending on the social network evolution model, *(d_2-d_1)* is *linear* or even *super-linear* in the size of the network.

[KM11]

Differential privacy fails to avoid breach



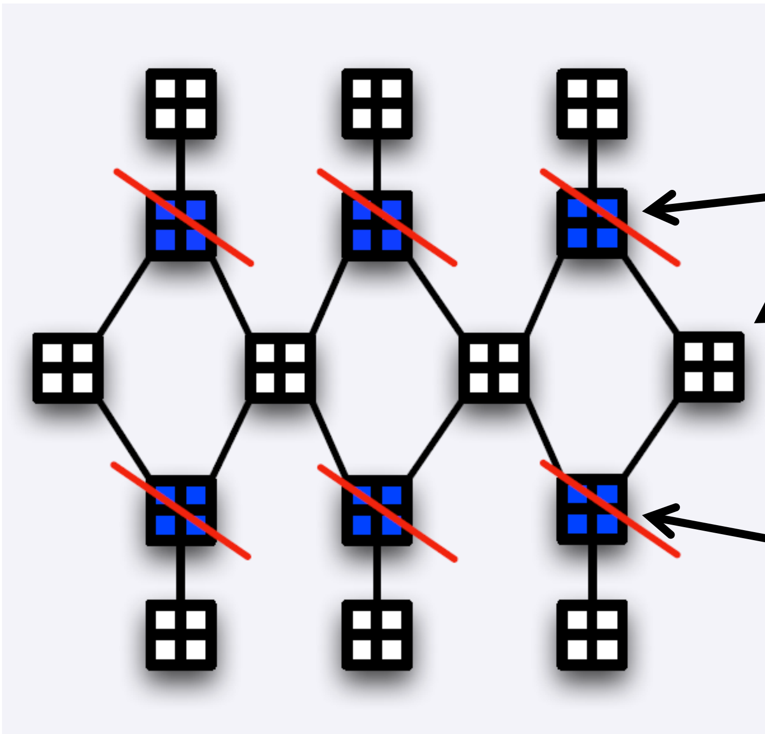
Output $(d_1 + \delta)$

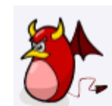
$\delta \sim$ Laplace$(1/\varepsilon)$

Output $(d_2 + \delta)$

Adversary can distinguish between the two worlds if $d_2 - d_1$ is large.

Reason for Privacy Breach



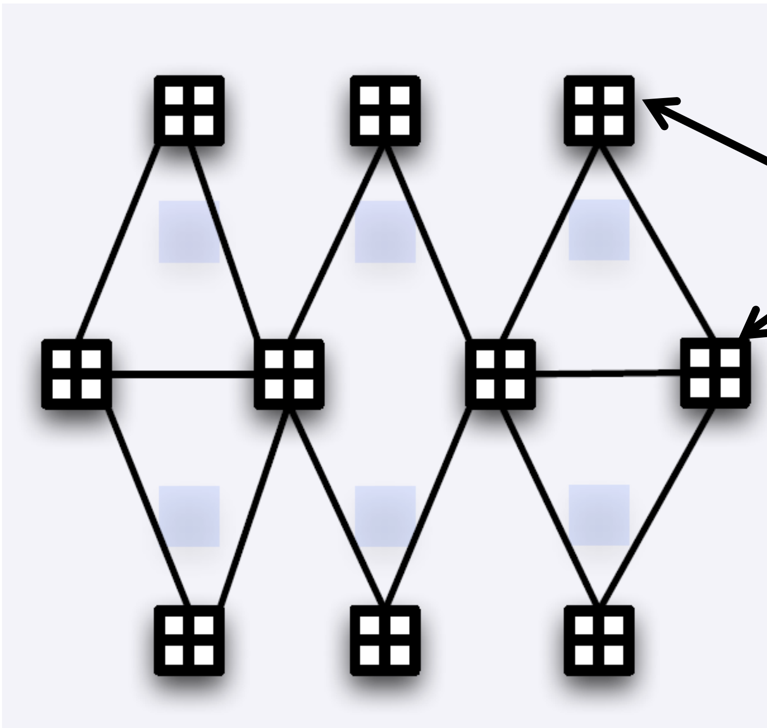
- Pairs of tables that differ in one tuple
-  cannot distinguish them

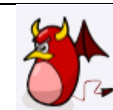
Tables that do not satisfy background knowledge

Space of all possible tables

Duke UNIVERSITY

Reason for Privacy Breach



 can distinguish between every pair of these tables based on the output

Space of all possible tables

No Free Lunch Theorem

It is not possible to guarantee *any* utility in addition to privacy, *without making assumptions about*

- the data generating distribution
- the background knowledge available to an adversary

[KM11]

[DN 10]