Privacy-preserving machine learning

Bo Liu, the HKUST
March, 1st, 2015.

Some slides extracted from

- Wang Yuxiang, Differential Privacy: a short tutorial.
- Cynthia Dwork, The Promise of Differential Privacy. A Tutorial on Algorithmic Techniques
- Christine Task, A Practical Beginners' Guide to Differential Privacy.
- Katrina Ligett, Tutorial on Differential Privacy.

Outlines

- Why privacy protection?
  - The attack model.
  - The privacy model.
  - The differential privacy*.
- How differential privacy?
  - Global sensitivity
  - Laplacian mechanism
  - Exponential mechanism
  - Sample and aggregate
- Differential privacy and machine learning.
  - Private machine learning
  - Non-interactive model.
  - Private Transfer Learning.
- Conclusion.

Outlines

- Why privacy protection?
  - The attack model.
  - The privacy model.
  - The differential privacy*.
- How differential privacy?
  - Global sensitivity
  - Laplacian mechanism
  - Exponential mechanism
  - Sample and aggregate
- Differential privacy and machine learning.
  - Private machine learning
  - Non-interactive model.
  - Private Transfer Learning.
- Conclusion.

Privacy Leakage Example

<u>Allowing public access or querying of sensitive database directly is vulnerable to privacy leakage.</u>

- The Netflix Dataset.
    - Cross-correlating volunteer public records from the IMDB identifies specific user in Netflix.
    - Very few background knowledge is required.
- The Facebook advertisement system.
    - The private information can be inferred by posing well designed ad for target user.
    - Very few public knowledge in open FB profile can lead to leakage.

Narayanan, A., & Shmatikov, V. (2008, May). Robust de-anonymization of large sparse datasets. In Security and Privacy, 2008. SP 2008. IEEE Symposium on (pp. 111-125). IEEE.
Korolova, A. (2010, December). Privacy violations using microtargeted ads: A case study. In Data Mining Workshops (ICDMW), 2010 IEEE International Conference on (pp. 474-482). IEEE.

Attack Models

- Record Linkage:
  - The attacker can confidently identify a small number of records in the released dataset.
  - $k\text{−anonymity}$ used for protection
  - Each published group contains at least k records.
- Table Linkage:
  - The attacker can identify whether the target is in the database or not.

Output only when >3 records exist.

3-anonymous

Job	Sex	Age	Disease (sensitive)
Professional	Male	[35-40]	HIV
Artist	Female	[30-35]	Flu

One published group

Fung, B., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. ACM Computing Surveys (CSUR), 42(4), 14.

Attack Models

- Attribute Linkage:
  - The attacker does not precisely identify the individual but can infer the sensitive attribute confidently.
  - $l-\mathrm{diversity}$ used for protection.
  - Each published group contains at least l distinct sensitive values.
- Probabilistic attack:
  - To ensure that the difference between the prior and posterior beliefs is small.
  - Differential Privacy for protection.

3-diversity

Job	Sex	Age	Disease (sensitive)
Professional	Male	[35-40]	HIV, Flu, Hepatitis
Artist	Female	[30-35]	HIV, Flu, Hepatitis

Group contains at least 3 kind of disease

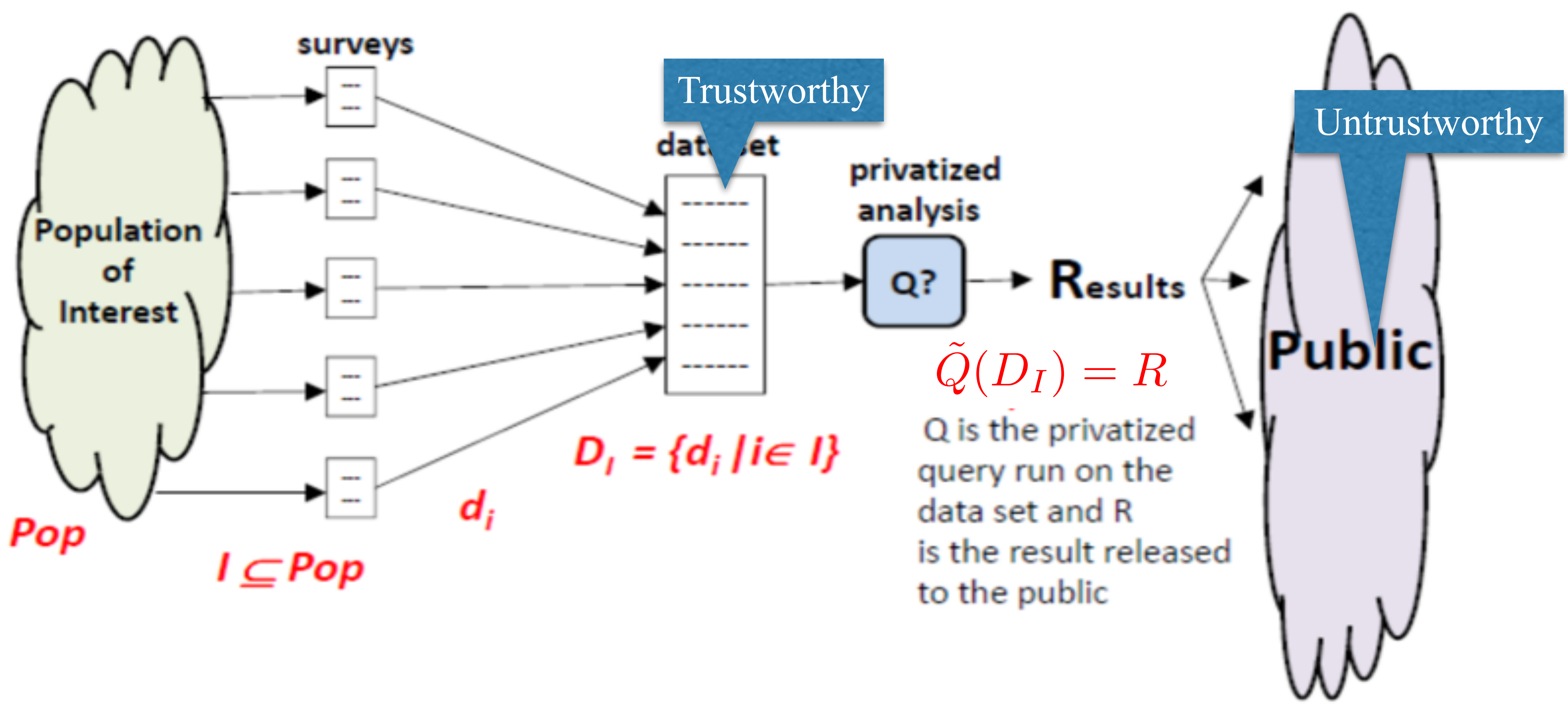
Fung, B., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. ACM Computing Surveys (CSUR), 42(4), 1

Differential Privacy(DP) Notations

Interactive Model:

➢ Untrustworthy data miners pose aggregate queries Q to database.

➢ Trustworthy database owner utilizes **mechanism** $\tilde{Q}$ to response the query privately.



Mohammed, N., Chen, R., Fung, B., & Yu, P. S. (2011, August). Differentially private data release for data mining. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 493-501). ACM.

DP Example

- Suppose hand in a survey about music taste.

  - Do you like music of Justin Bieber?
  - How many albums of Justin Bieber do you own?
  - Your age?
  - You gender?
  - Your job?

DP Example

- In what situations you will feel safe to hand in this survey.
  - The previous attack methods demonstrate that **anonymous publishing** is vulnerable to attack.
  - Your result has **no influence** on the query result.
    - Then your result enjoys no utility at all.
  - The attacker will learn **no new knowledge** about you by accessing the dataset.
    - Impossible! Your age will leakage if the attacker owns proper background knowledge(ex: you are 2 years older than average.)
    - No matter whether you hand in the survey or not.

DP

- DP guarantees that the query result **R** will be almost the same whether or not you hand in the survey.

- DP guarantees that the harm(privacy leakage) is almost the same whether or not you hand in the survey.

- **Ex:** The average age of survey taker is 21.31, whether or not you join the survey.

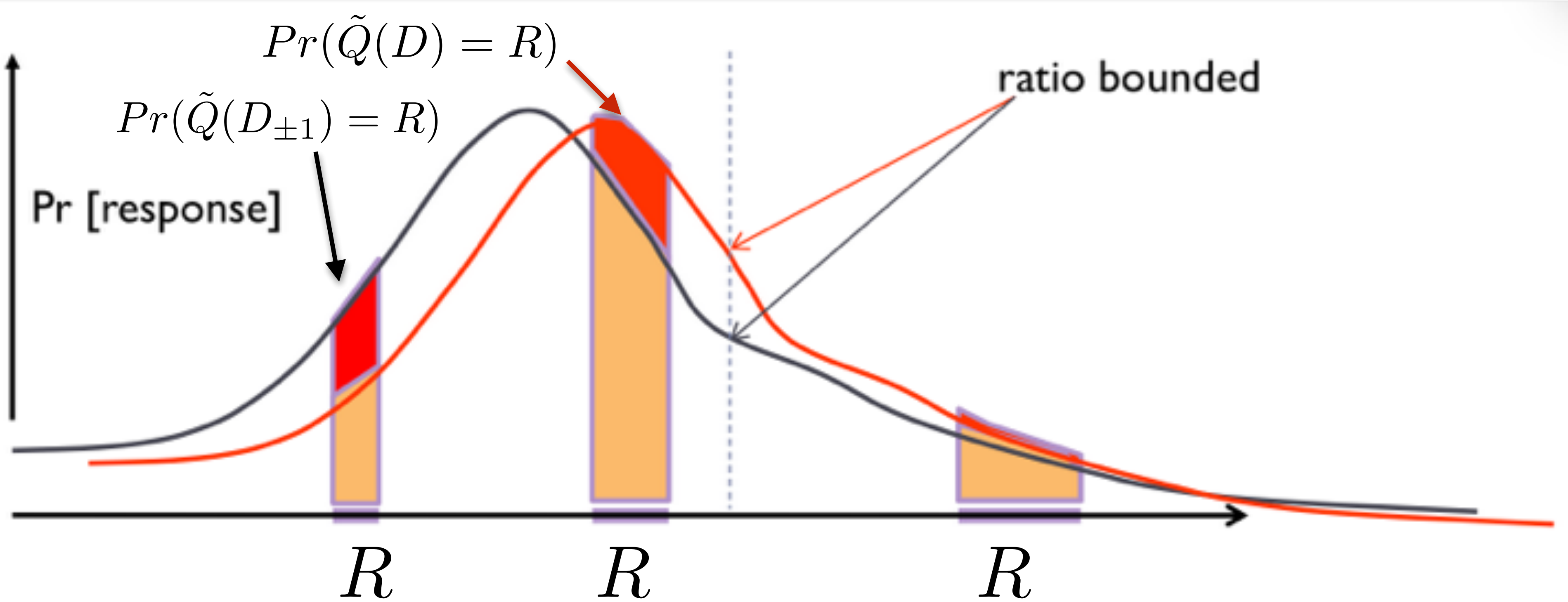
ϵ-DP definition

Mechanism $\tilde{Q}$ is dp, if its result R does not change much when one individual in the dataset changes (addition, deletion and modification).

Definition: ϵ-Differential Privacy, ϵ is called privacy budget.

$$\frac{Pr(\tilde{Q}(D) = R)}{Pr(\tilde{Q}(D_{\pm 1}) = R)} \leq e^{\epsilon}$$

For D and $D_{\pm 1}$ differs in 1 instance and any $R \in \mathrm{Range}(\tilde{Q})$



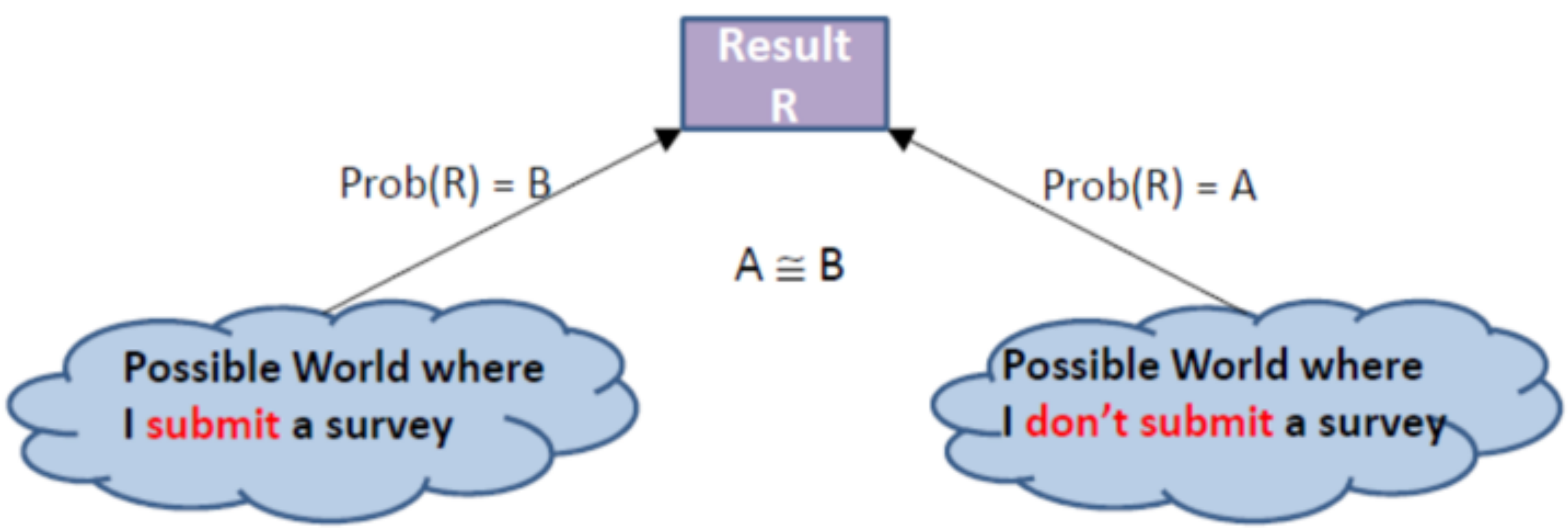
Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Theory of cryptography (pp. 265-284). Springer Berlin Heidelberg.

ϵ-DP definition

$\epsilon - dp$ mechanism returns a noisy result that
The average age of survey taker is 21.34.

The attackers have no idea whether I submit the survey or not.

Given R, how can anyone guess which possible world it came from?



k-fold composition of ϵ-dp.

- Utilizing $\epsilon - dp$ mechanism $\tilde{Q}$ k times for k queries is equivalent to $k\epsilon - dp$ mechanism.

- The protection of privacy is compromised. (Larger ϵ means less strict protection)

Advantages of DP

- DP serves as one of the most strict protection of privacy.
- DP makes no assumption about attackers' background knowledge.

Limits of DP

- DP does not protect the harm led by query result.
  - The attacker knows that Brody is 4cm higher than average.
  - Querying the average height will harm the privacy whether Brody join the survey or not.
- DP only protects the individual information rather than group information.
  - The attacker knows that you always act similar your 5 friends in the data.

Outlines

- Why privacy protection?
  - The attack model.
  - The privacy model.
  - The differential privacy*.
- How differential privacy?
  - Global sensitivity
  - Laplacian mechanism
  - Exponential mechanism
  - Sample and aggregate
- Differential privacy and machine learning.
  - Private machine learning
  - Non-interactive model.
  - Private Transfer Learning.
- Conclusion.

Global sensitivity(GS)

Definition:

The global sensitivity directly decides the noise magnitude added on data.

The maximum change of the result of query given a pair of neighbor dataset.

$$S = max_{D,D_{\pm 1}} |Q(D) - Q(D_{\pm 1})|$$

The global sensitivity for vector-value query.

$$S = max_{D,D_{\pm 1}} ||Q(D) - Q(D_{\pm 1})||_1$$

The global sensitivity is **query specific** rather than data specific.

Example of GS

- How many survey takers are female?
  - At most, one individual changes lead to #female changes by 1.
  - Thus, GS = 1.
- In total, how many Justin Bieber albums are bought by survey takers?
  - At most, the changed individual buys all 4 albums or buy no albums at all.
  - Thus, GS = 4.

How design dp mechanism?

- Laplace mechanism.
- Exponential mechanism.
- Sample and aggregate.

Laplace Mechanism

$$S = max_{D,D_{\pm 1}} |Q(D) - Q(D_{\pm 1})|$$

For Query Q whose result is R, the mechanism

$$\tilde{Q} = R + Lap(0, \frac{S}{\epsilon})$$

is $\epsilon - dp$. Lap denotes Laplace distribution.

To obtain $\epsilon - dp$ mechanism, the noise required only depends on global sensitivity of query Q and privacy budget ϵ.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Theory of cryptography (pp. 265-284). Springer Berlin Heidelberg.

Exponential Mechanism

- Suppose the query "The most frequent job that joins the survey?".
  - Add noisy on such categorical result is not applicable.

- Define utility function:
  For $R \in \text{Range(Q)}$, utility function $u(D, R)$ outputs a score.
  Ex, the frequency of job R in database D.
- Global Sensitivity: $S = max_{D, D_{\pm 1}} |u(D, R) - u(D_{\pm 1}, R)|$.

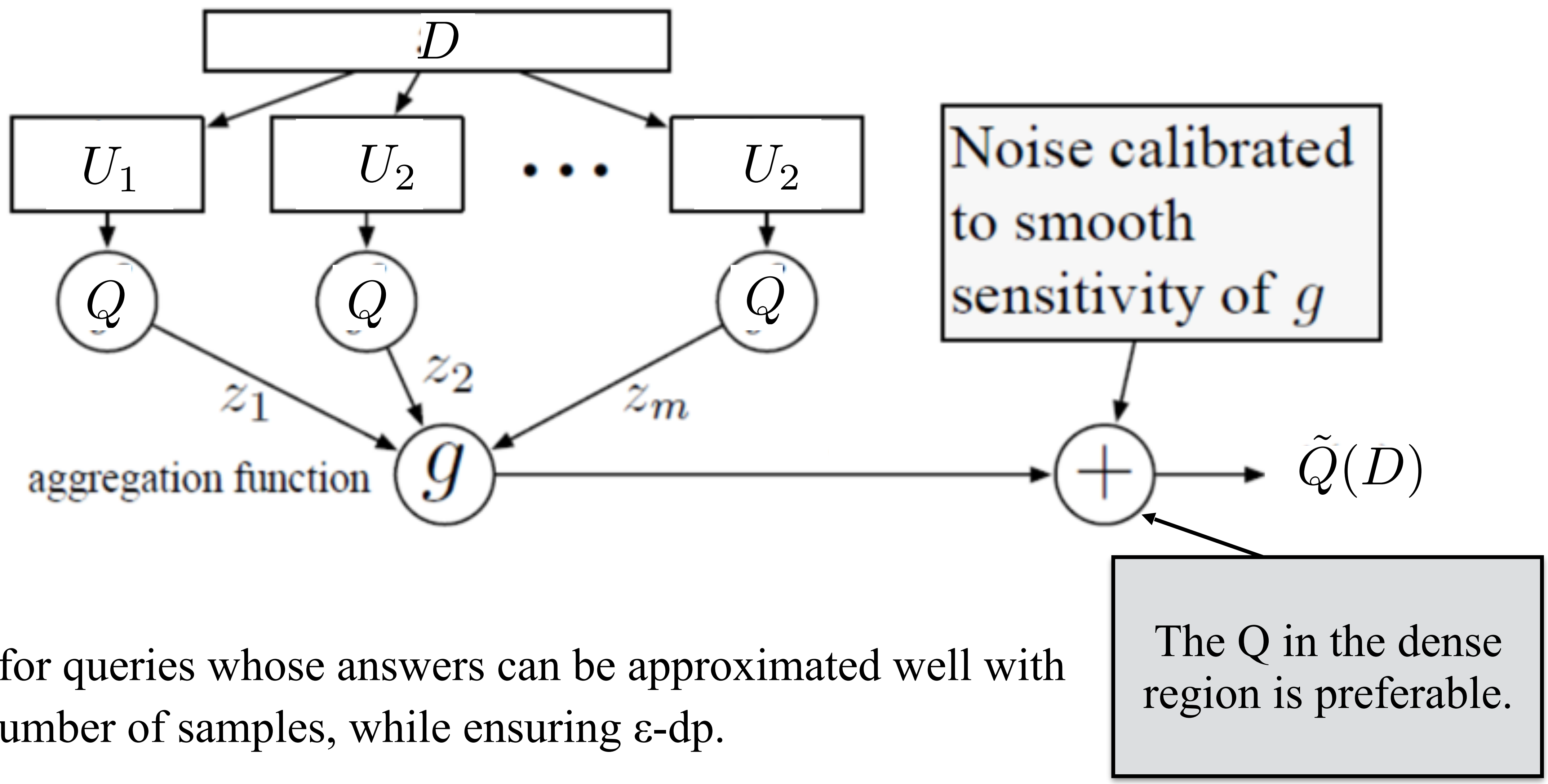
- Given dataset D, exponential mechanism select result R randomly.

$$Pr(R \text{ is selected}) \propto e^{\epsilon u(D,R)/2S}$$

McSherry, F., & Talwar, K. (2007, October). Mechanism design via differential privacy. In Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on (pp. 94-103). IEEE.

Sample and aggregate

$$\tilde{Q}(D) = g(Q(U_1), Q(U_2), \ldots, Q(U_m))$$



- Suitable for queries whose answers can be approximated well with a small number of samples, while ensuring ε-dp.
- Suitable for queries with large or unbounded sensitivity caused the sensitivity of g will be used.

Nissim, K., Raskhodnikova, S., & Smith, A. (2007, June). Smooth sensitivity and sampling in private data analysis. In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing (pp. 75-84). ACM.

Objective of DP Paper

- While keep the same level of DP, or keep privacy budget ϵ unchanged.
- Improve the utility of mechanism, or equivalently, minimize the noise added on result.

Outlines

- Why privacy protection?
  - The attack model.
  - The privacy model.
  - The differential privacy*.
- How differential privacy?
  - Global sensitivity
  - Laplacian mechanism
  - Exponential mechanism
  - Sample and aggregate
- Differential privacy and machine learning.
  - Private machine learning
  - Non-interactive model.
  - Private Transfer Learning.
- Conclusion.

Structure of a private machine learning paper

- Introduction
  - Why the data need to be learnt privately.
- Main contributions:
  - Propose a randomized algorithm.
  - Show this randomization need guarantee dp.
  - Show the sample complexity and usefulness under randomization.
- Evaluation:
  - Compare with brute force method (such Laplace mechanism)
  - Compare with non-private learning methods to show that the deterioration of performance is not significant. (The deterioration is the price of privacy.)

General method of private ML

- Output perturbation.
  - Learn the model from the clean data.
  - Utilize Laplacian mechanism or exponential mechanism to generate noisy model.
- Target perturbation.
  - Add the well designed perturbation item to the target function.
- Sample and aggregate.
  - For queries whose result can be approximated well using part of samples.

Differentially private logistic regression

The loss function/query for logistic regression(LR):

$$L(D, \lambda) = \frac{1}{n}\sum_{i=1}^{n} log(1 + e^{-y_i w^T x_i}) + \frac{1}{2}\lambda w^T w$$

$$Q(D) = w* = argmin_w L(D, \lambda)$$

$$(x_i, y_i) \in D$$

The logistic regression query expects the minimizer of loss function.

K. Chaudhuri, C. Monteleoni, Privacy-preserving logistic regression (NIPS'08)

Output perturbation(OP) LR

Given dataset D with sample size n and dimension d. The regularization parameter is λ.

Global sensitivity:

$$S = max_{D,D_{\pm 1}} |w^*(D) - w^*(D_{\pm 1})| \leq \frac{2}{n\lambda}$$

Output Perturbation, add Laplace noise on each dimension

$$z \sim Lap(0, \frac{2}{n\epsilon\lambda})$$

$$\tilde{Q}(D) = w^* + z$$

Intuition:

The larger regularization parameter, the more robust the model. Thus, the less noise is required to satisfy $\epsilon - dp$.

Target perturbation(TP) LR

Random sample a noise vector $\mathbf{b}$, each dimension $b_i \sim Lap(0, \frac{\epsilon}{2})$.

The perturbed objective function:

$$\tilde{L}(D, \lambda) = \frac{1}{n}\sum_{i=1}^{n} log(1 + e^{-y_i w^T x_i}) + \boxed{\frac{\mathbf{b}^T w}{n}} + \frac{1}{2} w^T w$$

$$\tilde{Q}(D) = argmin_w \tilde{L}(D, \lambda)$$

- OP and TP both guarantee ε-differential privacy.

The private model $\tilde{Q}_{TP}(D)$ by TP is guaranteed to be closer to clean model than OP.

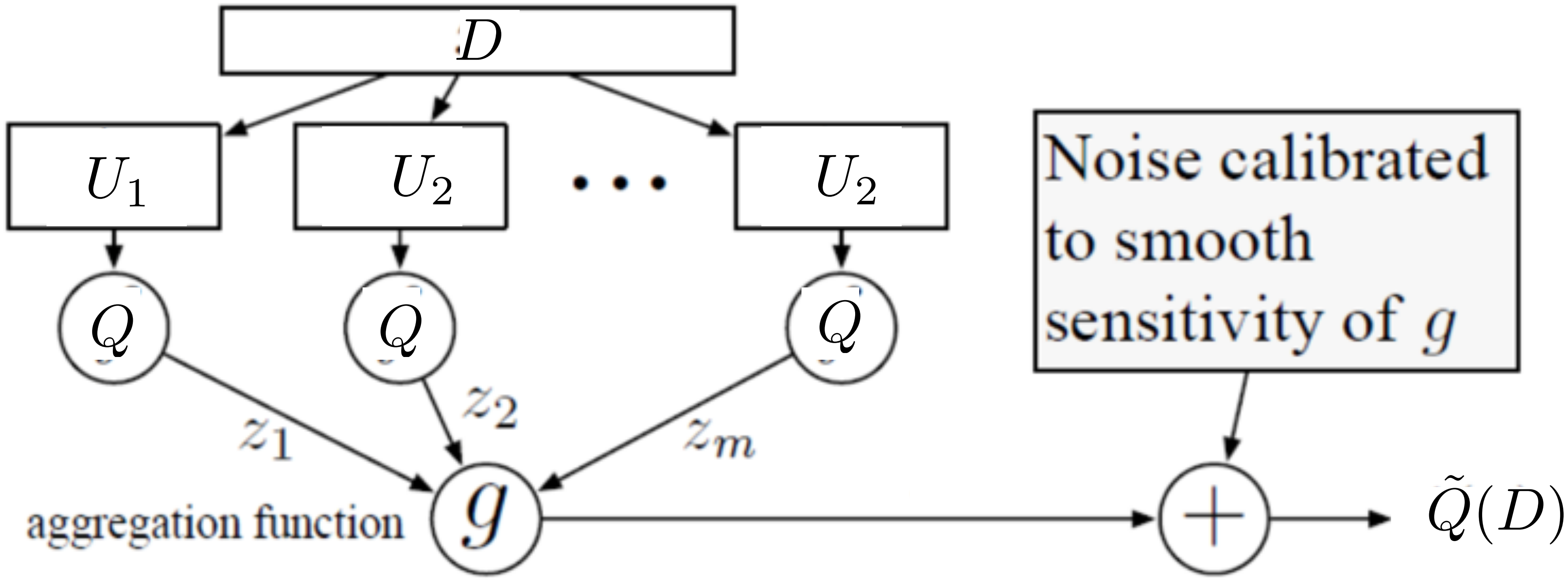
Differentially private k-means

$$cost_x(c_1, \ldots, c_k) = \frac{1}{n}\sum_{i=1}^{n} \min_j \left|\left|x_i - c_j\right|\right|_2^2$$

- Laplace mechanism based on global sensitivity overwhelms the result completely.
- Assume the data is "well-separated" that the cluster can be accurately estimated using a random subset.
  - Sample and aggregate framework works.

Nissim, K., Raskhodnikova, S., & Smith, A. (2007, June). Smooth sensitivity and sampling in private data analysis. In Proceedings of the thirty-ninth annual ACM symposium on Theory of computing (pp. 75-84). ACM.

Differentially private k-means



Private k-means:

Sample

1. Randomly split the training set as $(U_1, U_2, \ldots, U_m)$
2. Run non-private k-means method on each subset. And output cluster centers of each subset as $(z_1, z_2, \ldots, z_m)$.

Aggregate

3. Aggregate, $g(z_1, z_2, \ldots, z_m)$ outputs z_i in dense region. (Ex: z_i with minimum distance to t_{th} nearest neighbor).
4. Add Gaussian noise based on smooth sensitivity to guarantee $(\epsilon, \delta) - dp$.

Non-interactive Model

- Disadvantage of interactive model.
  - Assume m queries will be posed and $\epsilon - dp$ is required.
  - According to composition properties, $\frac{\epsilon}{m} - dp$ is required for each query.
  - Noise destroy the result.

- Non-interactive model.
  - Database owner publishes approximation of raw data.
  - Providing utility and privacy protection simultaneously.

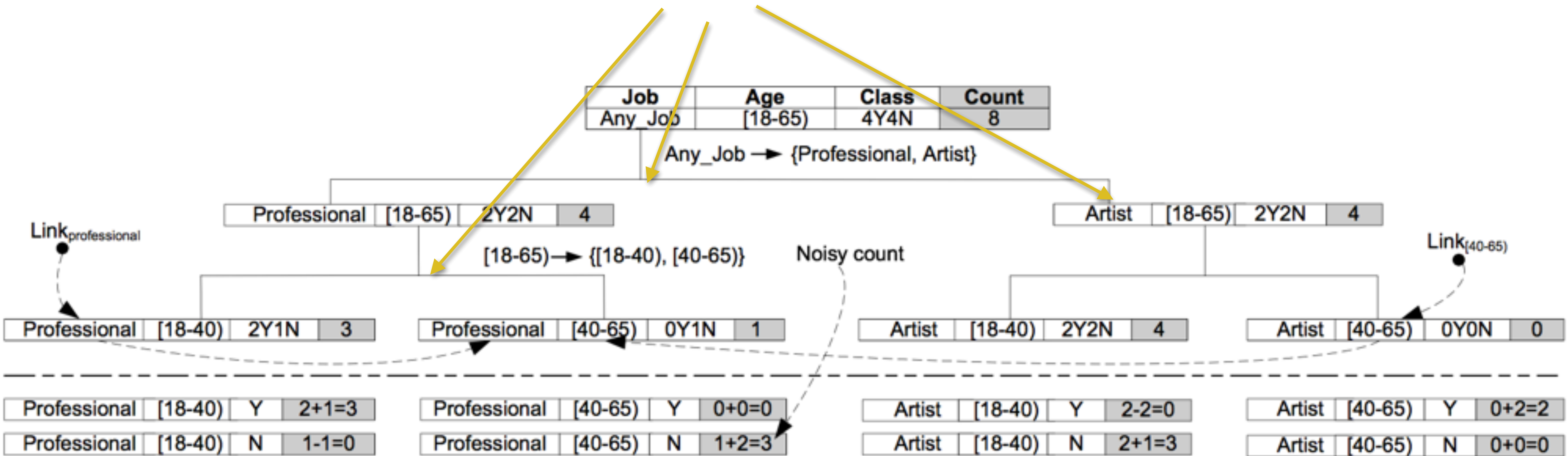
Non-interactive DP data release

- Partition based method.
- Model based method.
  - Specific designed for graph data.
- Public dataset based method.
  - If a similar structure public dataset is available.
  - Noisy reweighs the public instances so that the public and private dataset share similar marginal distribution(Domain Adaptation).
- Etc.

Ji, Z., & Elkan, C. (2013). Differential privacy based on importance weighting.*Machine learning*, *93*(1), 163-183.

Partition based data release

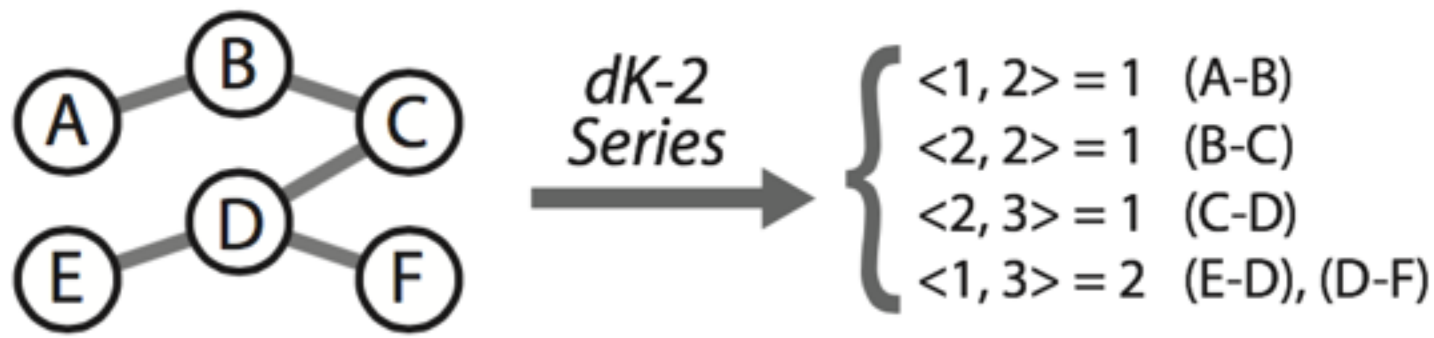
First step, partition the sample space similar to C4.5.



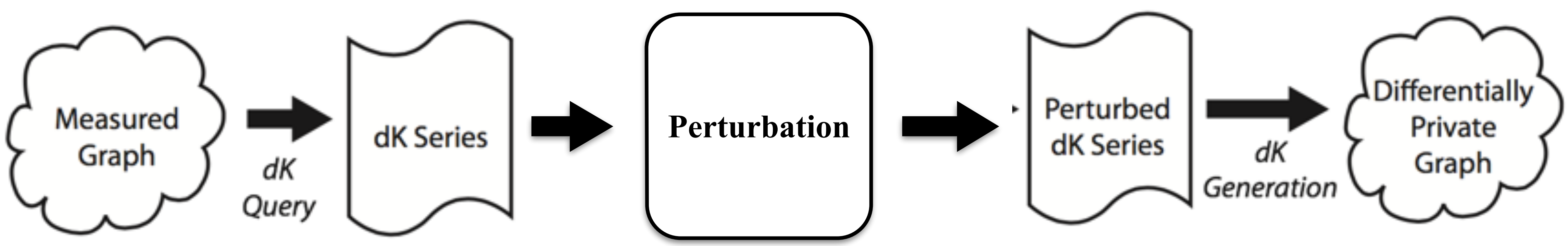
2nd step, publish noisy count of each partition(leaf in C4.5).

Noman Mohammed, Rui Chen, Benjamin C.M. Fung, and Philip S. Yu. 2011. Differentially private data release for data mining. KDD11.

Model based data release



- Dk-series serves as a generator model which captures topology of a graph.
- And a random graph can be generated to reproduce the raw graph.



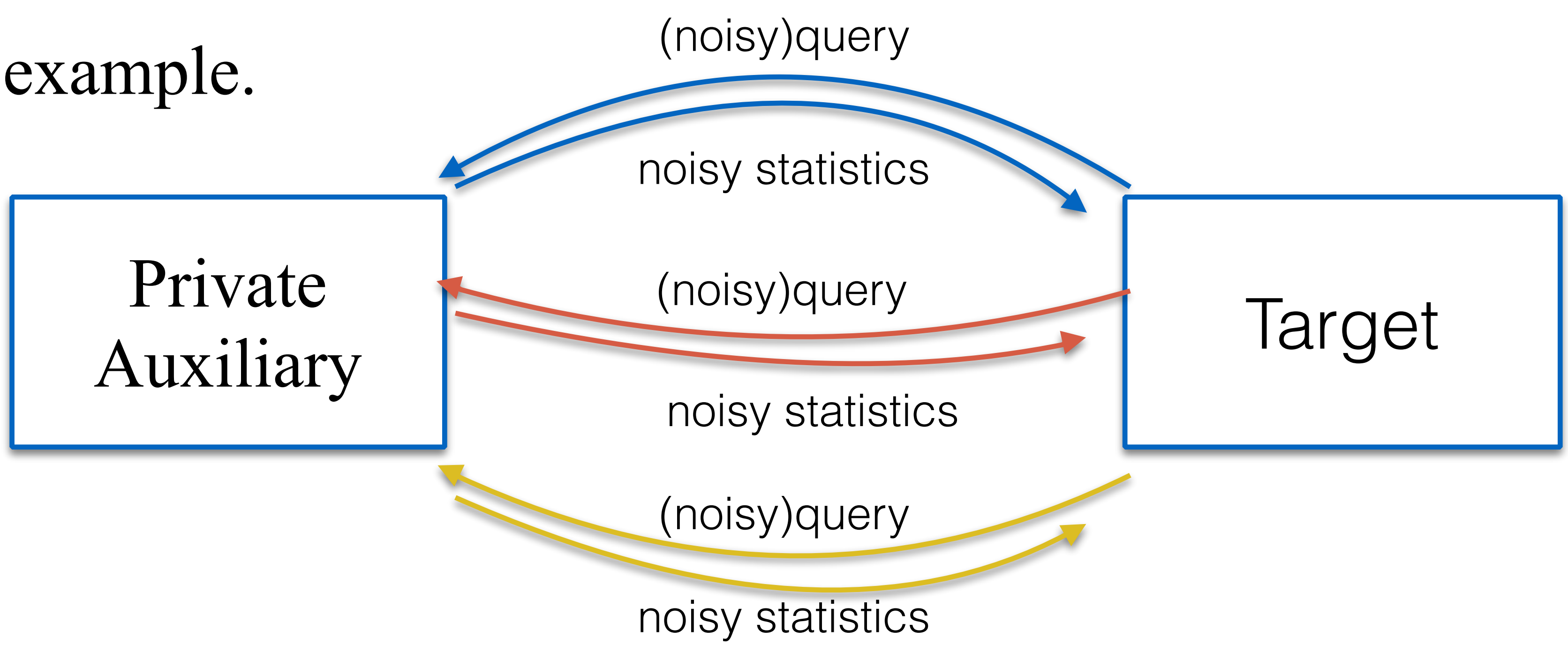
Such method publishes a noisy data generator model(dk-series) such that the raw graph can be reproduced.

Mahadevan, P., Krioukov, D., Fall, K., & Vahdat, A. (2006, September). Systematic topology analysis and generation using degree correlations. SIGCOMM11.
Sala, A., Zhao, X., Wilson, C., Zheng, H., & Zhao, B. Y. (2011, November). Sharing graphs using differentially private graph models. SIGCOMM11

Private transfer learning

- Take TrAdaBoost as example.

Interactive model



Private TrAdaBoost
For i=1,. . . ,N:
#Train a i_{th} base classifier.
For each iteration in training base classifier:
Target send the weight $\boldsymbol{w_i}$ to source.
Target queries the required statistics.
Private source returns the noisy result.

Disadvantages:

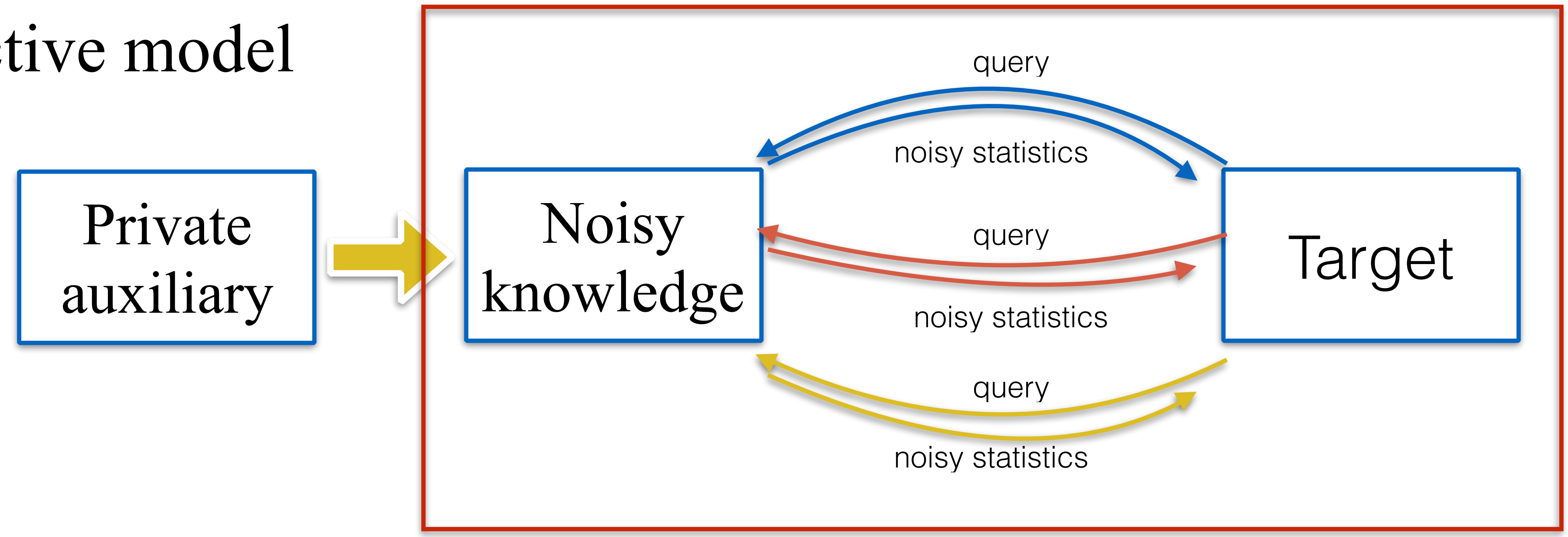
- Querying many times requires larger noise.
- Facing untrustworthy auxiliary, target has to protect privacy of target data as well.

Dai, W., Yang, Q., Xue, G. R., & Yu, Y. (2007, June). Boosting for transfer learning. In Proceedings of the 24th international conference on Machine learning (pp. 193-200). ACM.

Private transfer learning

Non-private transfer learning can be used directly.

Non-interactive model



- How to represent noisy knowledge?(What to transfer)
  - A noisy generator model or synthetic dataset.
  - A noisy classification model.
  - A noisy histogram.
  - etc.
- How target learns from noisy knowledge more efficiently and robust?(How to transfer)

Conclusion

- An introduction of privacy and popular protection differential privacy.
- An introduction of differentially private machine learning.
- Transfer learning:
  - Transfer learning from sensitive dataset faces privacy risks directly.
  - Obtaining a private knowledge representation facilitates private transfer learning.

Thank You!