DP & Relational Databases: A case study on Census Data

Ashwin Machanavajjhala

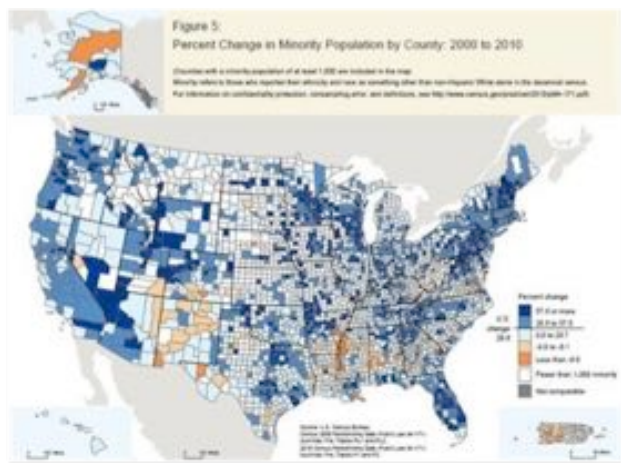
ashwin@cs.duke.edu

Aggregated Personal Data …

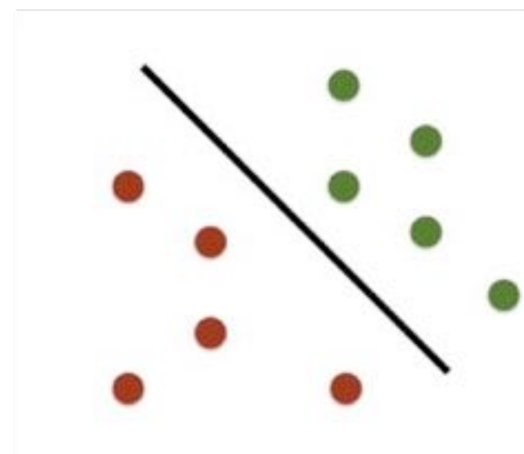
… is made publicly available in many forms.

De-identified records
(e.g., medical)

Statistics
(e.g., demographic)



Predictive models
(e.g., advertising)



… but privacy breaches abound

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr.
Published: August 9, 2006

SIGN IN TO E-MAIL THIS

NETFLIX

Why 'Anonymous' Data Sometimes Isn't

By Bruce Schneier 12.13.07

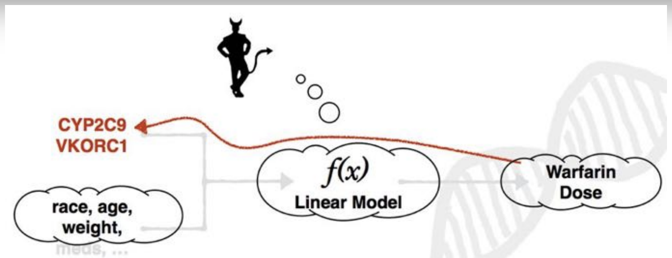
Last year, Netflix published 10 million movie rankings by 500,000 customers, as part of a challenge for people to come up with better recommendation systems than the one the company was using.

The Scientist » The Nutshell

"Anonymous" Genomes Identified

The names and addresses of people participating in the Personal Genome Project can be easily tracked down despite such data being left off their online profiles.

By Dan Cossins | May 3, 2013



Privacy in Pharmacogenetics: An End-to-End Case Study of Personalized Warfarin Dosing

Differential Privacy

[Dwork, McSherry, Nissim, Smith TCC 2006, Gödel Prize 2017]

The output of an algorithm should be *insensitive* to adding or removing a record from the database.

Think: Whether or not an individual is in the database

Differential Privacy

- Property of the privacy preserving computation.
  - Algorithms can't be reverse-engineered.
- Composition rules help reason about privacy leakage across multiple releases.
  - Maximize utility under a privacy budget.
- Individual's privacy risk is bounded despite prior knowledge about them from other sources *

A decade later ...

- A few important practical deployments …

OnTheMap *[ICDE 2008]*

[CCS 2014]

[Apple WWDC 2016]

- … but little adoption beyond that.
  - Deployments have needed teams of experts
  - Supporting technology is not transferrable
  - Virtually no systems/software support

This talk

Theory & Algorithms → **Practice** → **Systems**

No Free Lunch *[SIGMOD11]*
Pufferfish *[TODS14]*
Blowfish *[SIGMOD14,VLDB15]*

LODES *[SIGMOD17]*
2020 Census *[ongoing]*
IoT *[CCS17, ongoing]*

DPBench *[SIGMOD16]*
DPComp *[SIGMOD16]*
Pythia *[SIGMOD17]*
Ektelo *[ongoing]*
Private-SQL *[ongoing]*

This Talk

- Theory to Practice
  - Utility cost of provable privacy on Census Bureau data

- Practice to Systems
  - Ektelo: An operator based framework for describing differentially private computations

Part 1: Theory to Practice

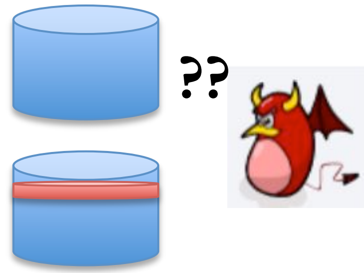
- Can traditional algorithms for data release and analysis be **replaced with provably private** algorithms while ensuring **little loss in utility**?

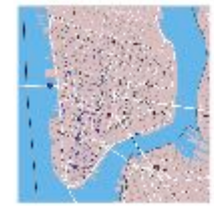
Yes we can … on US Census Bureau Data

The utility cost of provable privacy on US Census Bureau data

- Current algorithm for data release with *no provable guarantees* and parameters used have to be kept *secret*

The utility cost of provable privacy on US Census Bureau data



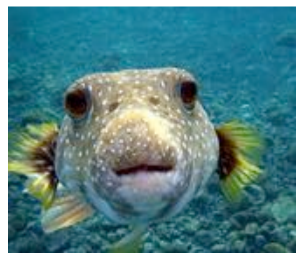


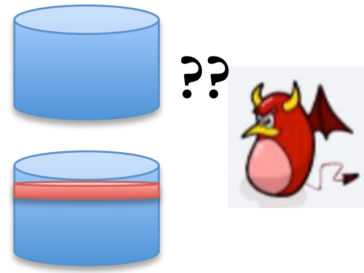
US Law: Title 13 Section 9 ≈ Pufferfish Privacy Requirements ← DP-like Privacy Definition ← Noisy Employer Statistics

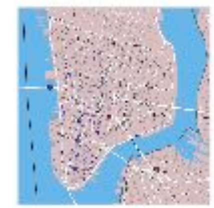
Comparable or lower error than current non-private methods

The utility cost of provable privacy on US Census Bureau data

SIGMOD 2017







US Law:
Title 13
Section 9

≈

Pufferfish
Privacy
Requirements

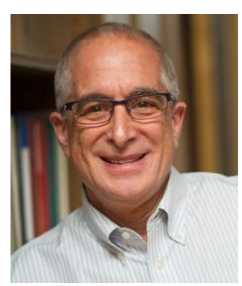
←

DP-like
Privacy
Definition

←

Noisy
Employer
Statistics

Sam Haney



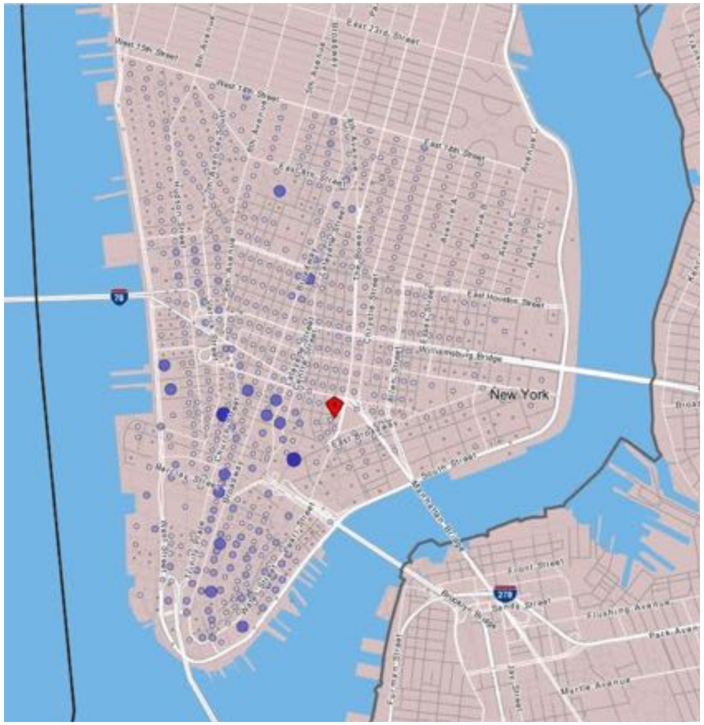
John Abowd

Matthew Grahan

Mark Kutzbach

Lars Vilhuber

US Census Bureau's OnTheMap

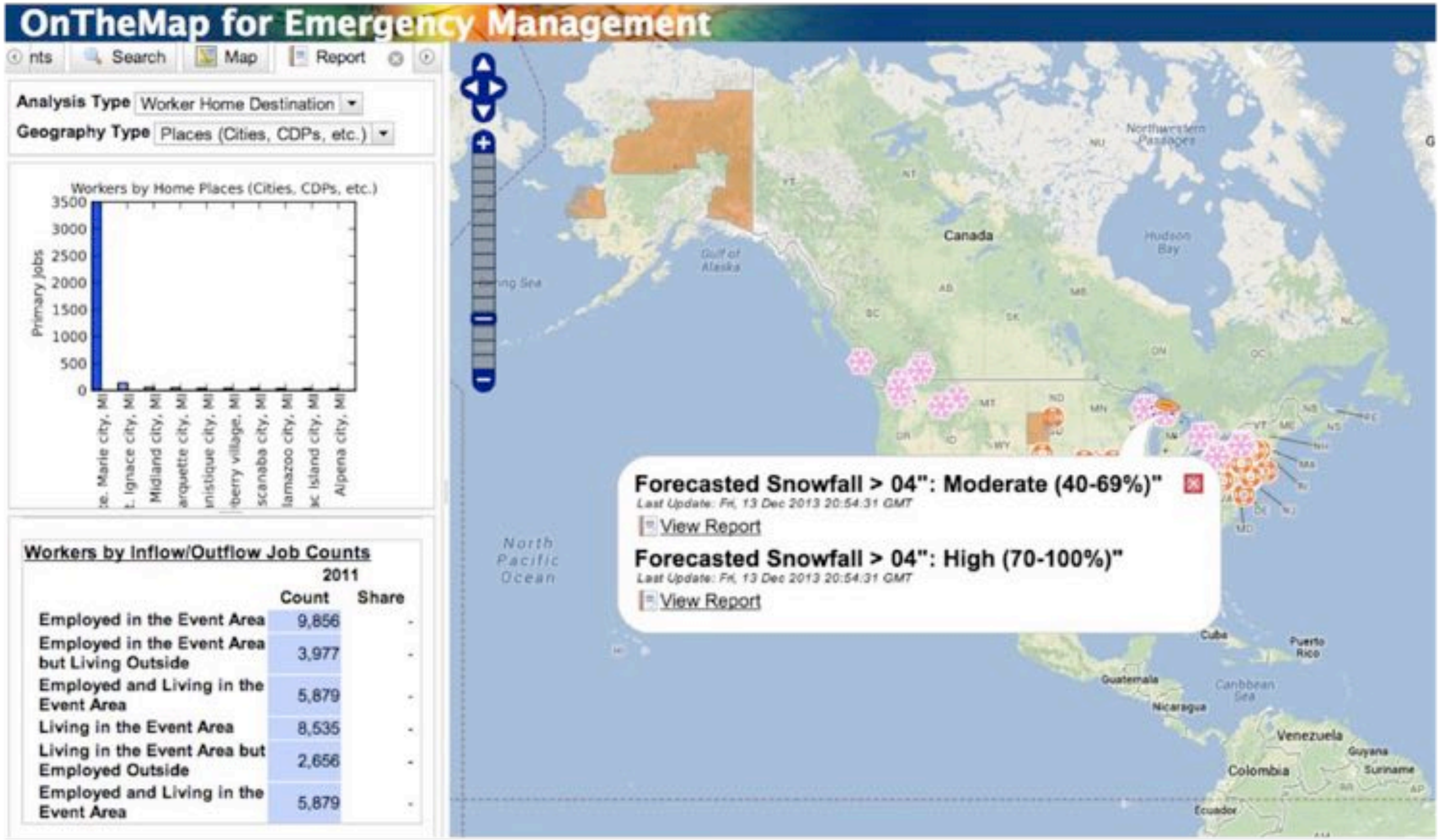


Employment in Lower Manhattan

Residences of Workers Employed in Lower Manhattan

Available at http://onthemap.ces.census.gov/.

OnTheMap



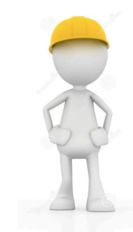
OnTheMap for Emergency Management
Search
Map
Report
Analysis Type
Worker Home Destination
Geography Type
Places (Cities, CDPs, etc.)
Workers by Home Places (Cities, CDPs, etc.)
Primary Jobs
3500
3000
2500
2000
1500
1000
500
0
Midland city, MI
Alpena city, MI
Workers by Inflow/Outflow Job Counts
2011
Count
Share
Employed in the Event Area 9,856
Employed in the Event Area but Living Outside 3,977
Employed and Living in the Event Area 5,879
Living in the Event Area 8,535
Living in the Event Area but Employed Outside 2,656
Employed and Living in the Event Area 5,879
Forecasted Snowfall > 04": Moderate (40-69%)"
Last Update: Fri, 13 Dec 2013 20:54:31 GMT
View Report
Forecasted Snowfall > 04": High (70-100%)"
Last Update: Fri, 13 Dec 2013 20:54:31 GMT
View Report
Canada
Gulf of Alaska
North Pacific Ocean
Hudson Bay
Northwestern Passages
Cuba
Puerto Rico
Guatemala
Nicaragua
Caribbean Sea
Venezuela
Guyana
Suriname
Colombia
Ecuador

Underlying Data: LODES

Employer
Employer ID
Location
Ownership
Industry

Jobs
Start Date
End Date
Worker ID
Employer ID

Worker
Worker ID
Age
Sex
Race/Ethnicity
Education
Home Location



Goal: Release Tabular Summaries

Counting Queries

- Count of jobs in NYC
- Count of jobs held by workers age 30 who work in Boston.

Marginal Queries

- Count of jobs held by workers age 30 by work location (aggregated to county)

Release of data about employers and employees is regulated by …

- Title 13 Section 9

Neither the secretary nor any officer or employee …

… make any publication whereby the data furnished by any particular establishment or individual under this title can be identified …

Current Interpretation

- The existence of a job held by a particular individual *must* not be disclosed.

No exact re-identification of employee records … by an informed attacker.

- The existence of an employer business as well as its type (or sector) and location is not confidential.

Can release exact numbers of employers

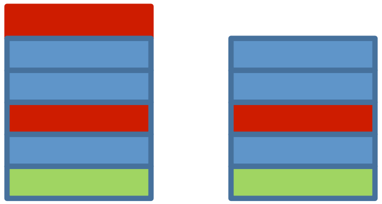
- The data on the operations of a particular business must be protected.

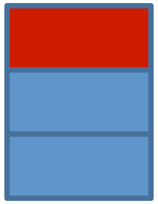
Informed attackers must have an uncertainty of up to a multiplicative factor $(1+\alpha)$ about the workforce of an employer

Can we use differential privacy (DP)?

For every pair of
Neighboring Tables

For every output …





D_1 D_2 O

Should not be able to distinguish whether O was generated by D_1 or D_2

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

Neighboring tables for LODES?

- Tables that differ in …
  - one employee?
  - one employer?
  - something else?

- And how does DP (and its variants) compare to the current interpretation of the law?
  - Who is the attacker? Is he/she informed?
  - What is secret and what is not?

The Pufferfish Framework

[TODS 14]

- What is being kept secret?

  A set of Discriminative Pairs (mutually exclusive pairs of secrets)

- Who are the adversaries?

  A set of Data evolution scenarios (adversary priors)

- What is privacy guarantee?

  Adversary can’t tell apart a pair of secrets any better by observing the output of the computation.

Pufferfish Privacy Guarantee

$\forall\, w \in range(M)$
$\forall\, (s, s') \in S_{pairs}$
$\forall \theta \in D, s.t. \quad P(s|D), P(s'|D) \neq 0$

$$e^{-\varepsilon} \leq \frac{P(s|M(\mathfrak{D}) = w, \theta)}{P(s'|M(\mathfrak{D}) = w, \theta)} \Big/ \frac{P(s|\theta)}{P(s'|\theta)} \leq e^{\varepsilon}$$

Posterior odds of s vs s'

Prior odds of s vs s'

Advantages of Pufferfish

- Gives a deeper understanding of the protections afforded by existing privacy definitions
  - Differential privacy is an instantiation

- Privacy defined more generally in terms of customizable secrets rather than records

- We can tailor the set of discriminative pairs, and the adversarial scenarios to specific applications
  - Fine grained knobs for tuning the privacy-utility tradeoff

Customized Privacy for LODES

- **Discriminative Secrets**:
  - (w works at E, or w works at E')
  - (w works at E, w does not work)
  - ($|E| = x$, $|E| = y$), for all $x < y < (1+\alpha)x$
  - …

- **Data evolution scenarios**:
  - All priors where employee records are independent of each other.

Example of a formal privacy requirement

DEFINITION 4.2 (EMPLOYER SIZE REQUIREMENT). *Let e be any establishment in $\mathcal{E}$. A randomized algorithm A protects establishment size against an informed attacker at privacy level (ϵ, α) if, for every informed attacker $\theta \in \Theta$, for every pair of numbers x, y, and for every output of the algorithm $\omega \in range(A)$,*

$$\left| \log \left(\frac{Pr_{\theta,A}[|e| = x | A(D) = \omega]}{Pr_{\theta,A}[|e| = y | A(D) = \omega]} \Big/ \frac{Pr_{\theta}[|e| = x]}{Pr_{\theta}[|e| = y]} \right) \right| \leq \epsilon \quad (4)$$

whenever $x \leq y \leq \lceil (1+\alpha)x \rceil$ *and* $Pr_{\theta}[w = x], Pr_{\theta}[w = y] > 0$.

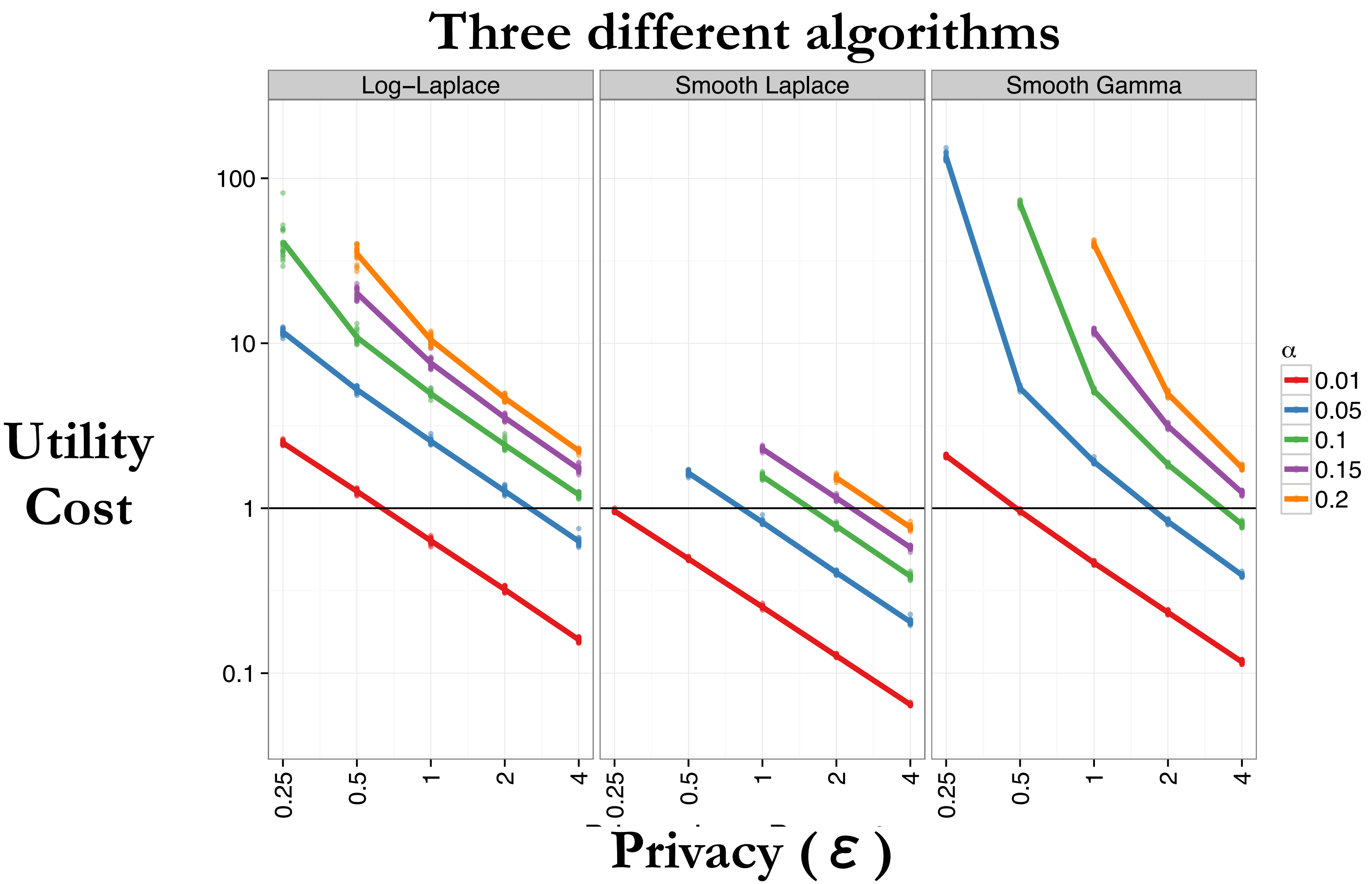
Customized Privacy for LODES

- Provides a differential privacy type privacy guarantee for all employees
  - Algorithm output is insensitive to addition or removal of one employee
- Appropriate privacy for establishments
  - Can learn whether an establishment is large or small, but not exact workforce counts.
- Satisfies sequential composition

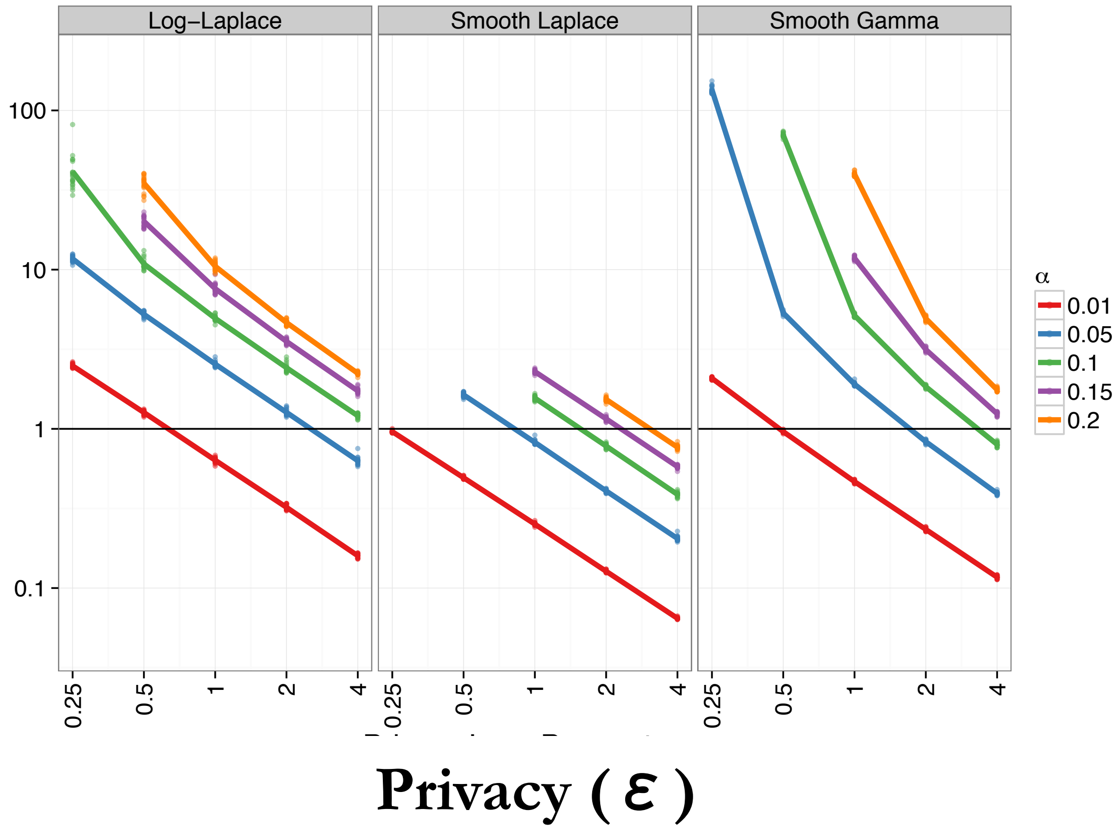
What is the utility cost?

- Sample constructed from 3 states in US
  - 10.9 million jobs and 527,000 establishments
- Q1: Marginal counts over all establishment characteristics
  - 33,000 counts are being released.
- Utility Cost: error (new alg.)/error (current alg.)

Utility Cost



Utility Cost



- For $\varepsilon \geq 1$, and $\alpha \leq 5\%$ utility cost is at most a factor of 3.
- Can design a DP algorithm that protects both employer and employee secrets.

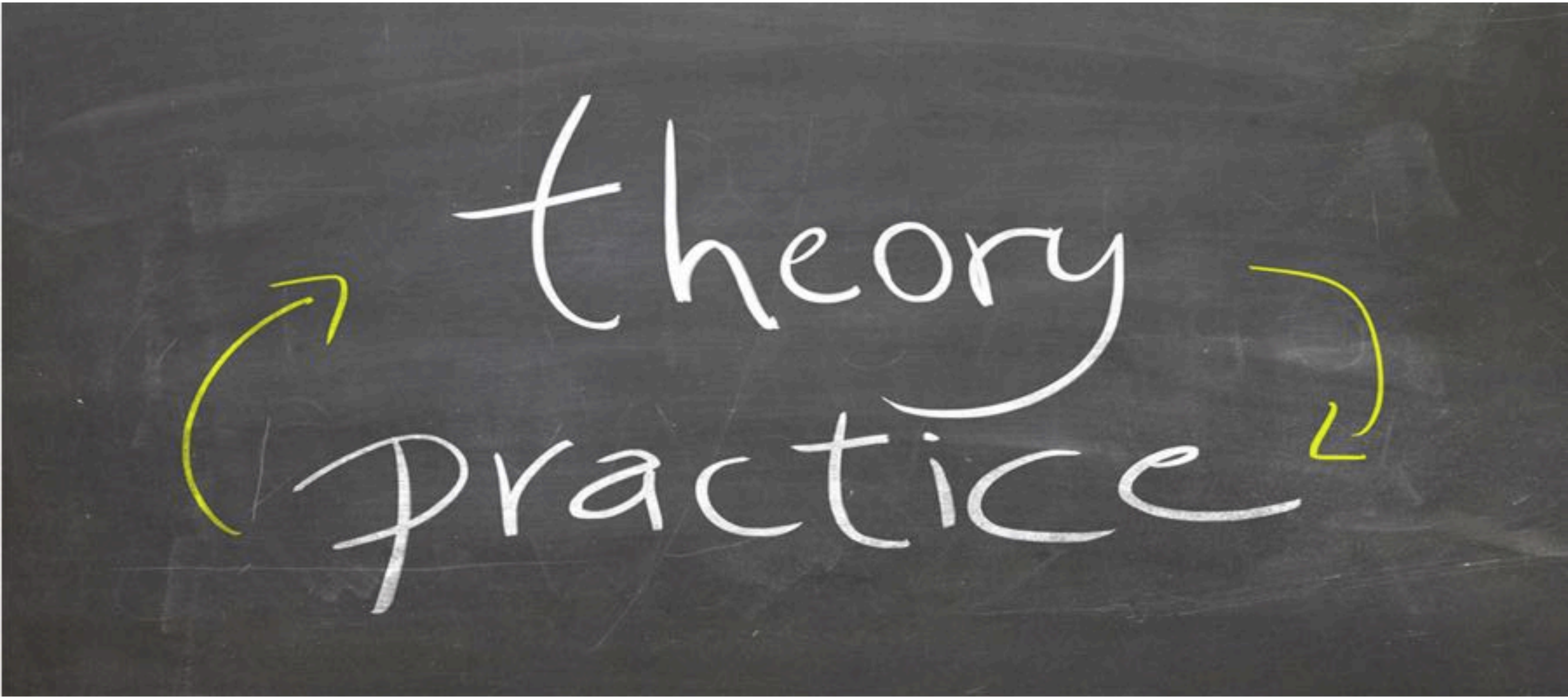
  It has uniformly high cost for all epsilon values.

Summary: Theory to Practice

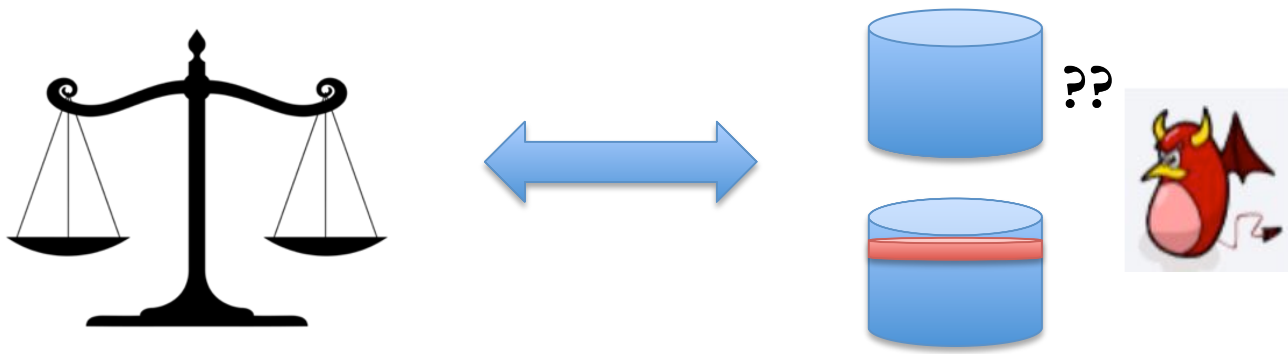
- Can traditional algorithms for data release and analysis be **replaced with provably private** algorithms while ensuring **little loss in utility**?

- Yes we can … on US Census Bureau Data
  - Can release tabular summaries with *comparable or better utility* than current techniques!

Takeaways



Challenge 1: Policy to Math



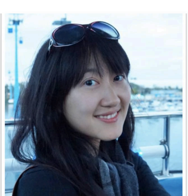
Challenge 2: Privacy for Relational Data

Employer
Employer ID
Location
Ownership
Industry

Jobs
Start Date
End Date
Worker ID
Employer ID

Worker
Worker ID
Age
Sex
Education

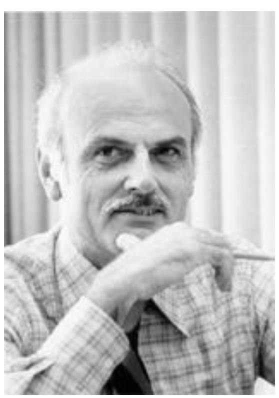
- Privacy for each entity
- Redefine neighbors
- Constraints
  - Keys
  - Foreign Keys
  - Inclusion dependencies
  - Functional Dependencies



Xi He

Challenge 3: Algorithm Design

"… without exception ad hoc, cumbersome, and difficult to use – they could really only be used by people having highly specialized technical skills …"



E. F. Codd on the state of databases in early 1970s

Part 2: Practice to Systems

- Can **provably private** data analysis algorithms with **state-of-the-art utility** be achieved by DP-**non-experts**?

Systems Vision

Given a task specified in a high level language,
and a privacy budget*

synthesize an algorithm to complete the task
with (near-)optimal accuracy,
and with differential privacy guarantees.

Systems Vision

Given a relational schema, a set of SQL queries,
and a privacy budget*

synthesize an algorithm to answer these queries
with (near-)optimal accuracy,
and with differential privacy guarantees.

State of the art

- Systems that answer SQL queries are **far from optimal in terms of utility**.
  - Answer one query at a time

- Sophisticated algorithms that achieve near-optimal error for specialized query types
  - **Linear queries** on “single” tables
  - Certain queries on graphs

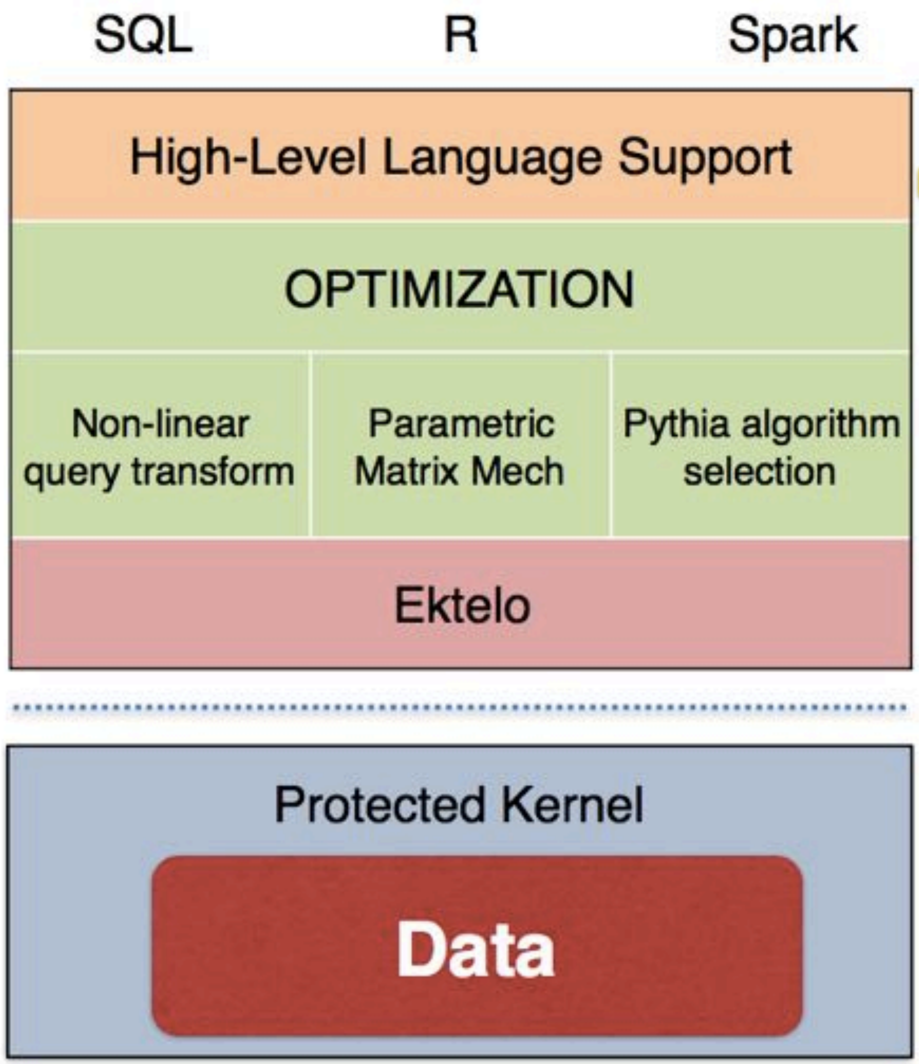
Challenges for a non-expert

- Need to cast problems in terms of specialized queries.
- Algorithms assume special representations of data
  - Possibly exponential size in the input
- No standard implementations of algorithms
- Algorithms achieving best utility can depend on the dataset and privacy parameters used

System-P Vision

Inputs

- Policy G = (ϵ, ρ)
- Analysis task
- Error tolerance
- Schema
- Constraints



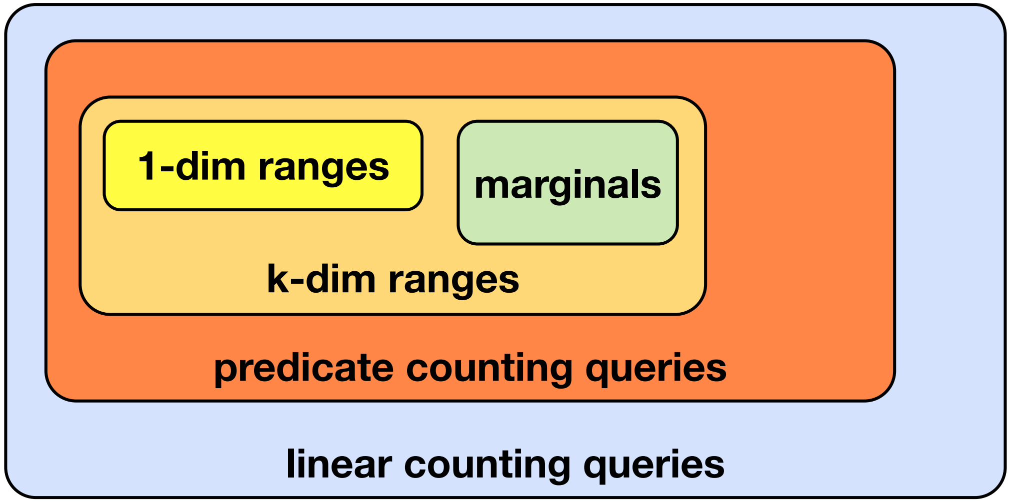
Outputs

- Answer
- Error bound(s)
- Budget rem. ϵ'

Gerome Miklau

Michael Hay

Linear queries



- **1-dimensional range queries**: intervals
- **Marginals / data cube queries / contingency tables**: aggregate over excluded dimensions.
- **k-dimensional range queries**: axis-aligned rectangles
- **Predicate counting queries**: only 0 or 1 coefficients
- **Linear counting queries**: arbitrary coefficients

Census Summary File (SF-1)

P3. RACE [8]

Universe: Total population

Total:	P0030001
White alone	P0030002
Black or African American alone	P0030003
American Indian and Alaska Native alone	P0030004
Asian alone	P0030005
Native Hawaiian and Other Pacific Islander alone	P0030006
Some Other Race alone	P0030007
Two or More Races	P0030008

P4. HISPANIC OR LATINO ORIGIN [3]

Universe: Total population

Total:	P0040001
Not Hispanic or Latino	P0040002
Hispanic or Latino	P0040003

P5. HISPANIC OR LATINO ORIGIN BY RACE [17]

Universe: Total population

Total:	P0050001
Not Hispanic or Latino:	P0050002
White alone	P0050003
Black or African American alone	P0050004
American Indian and Alaska Native alone	P0050005
Asian alone	P0050006
Native Hawaiian and Other Pacific Islander alone	P0050007
Some Other Race alone	P0050008
Two or More Races	P0050009
Hispanic or Latino:	P0050010
White alone	P0050011
Black or African American alone	P0050012
American Indian and Alaska Native alone	P0050013
Asian alone	P0050014
Native Hawaiian and Other Pacific Islander alone	P0050015
Some Other Race alone	P0050016
Two or More Races	P0050017

P20. HOUSEHOLDS BY PRESENCE OF PEOPLE UNDER 18 YEARS BY HOUSEHOLD TYPE BY AGE OF PEOPLE UNDER 18 YEARS [34]

Universe: Households

Total:	P0200001
Households with one or more people under 18 years:	P0200002
Family households:	P0200003
Husband-wife family:	P0200004
Under 6 years only	P0200005
Under 6 years and 6 to 17 years	P0200006
6 to 17 years only	P0200007
Other family:	P0200008
Male householder, no wife present:	P0200009
Under 6 years only	P0200010
Under 6 years and 6 to 17 years	P0200011
6 to 17 years only	P0200012

P28. HOUSEHOLD TYPE BY HOUSEHOLD SIZE [16]

Universe: Households

Total:	P0280001
Family households:	P0280002
2-person household	P0280003
3-person household	P0280004
4-person household	P0280005
5-person household	P0280006
6-person household	P0280007
7-or-more-person household	P0280008
Nonfamily households:	P0280009
1-person household	P0280010
2-person household	P0280011
3-person household	P0280012

Census Summary File (SF-1)

P3.	**RACE [8]**	
	Universe: Total population	
	Total:	P0030001
	White alone	P0030002
	Black or African American alone	P0030003
	American Indian and Alaska Native alone	P0030004
	Asian alone	P0030005
	Native Hawaiian and Other Pacific Islander alone	P0030006
	Some Other Race alone	P0030007
	Two or More Races	P0030008
P4.	**HISPANIC OR LATINO ORIGIN [3]**	
	Universe: Total population	
	Total:	P0040001
	Not Hispanic or Latino	P0040002
	Hispanic or Latino	P0040003
P5.	**HISPANIC OR LATINO ORIGIN BY RACE [17]**	
	Universe: Total population	
	Total:	P0050001
	Not Hispanic or Latino:	P0050002
	White alone	P0050003
	Black or African American alone	P0050004
	American Indian and Alaska Native alone	P0050005
	Asian alone	P0050006
	Native Hawaiian and Other Pacific Islander alone	P0050007
	Some Other Race alone	P0050008
	Two or More Races	P0050009
	Hispanic or Latino:	P0050010
	White alone	P0050011
	Black or African American alone	P0050012
	American Indian and Alaska Native alone	P0050013
	Asian alone	P0050014
	Native Hawaiian and Other Pacific Islander alone	P0050015
	Some Other Race alone	P0050016
	Two or More Races	P0050017

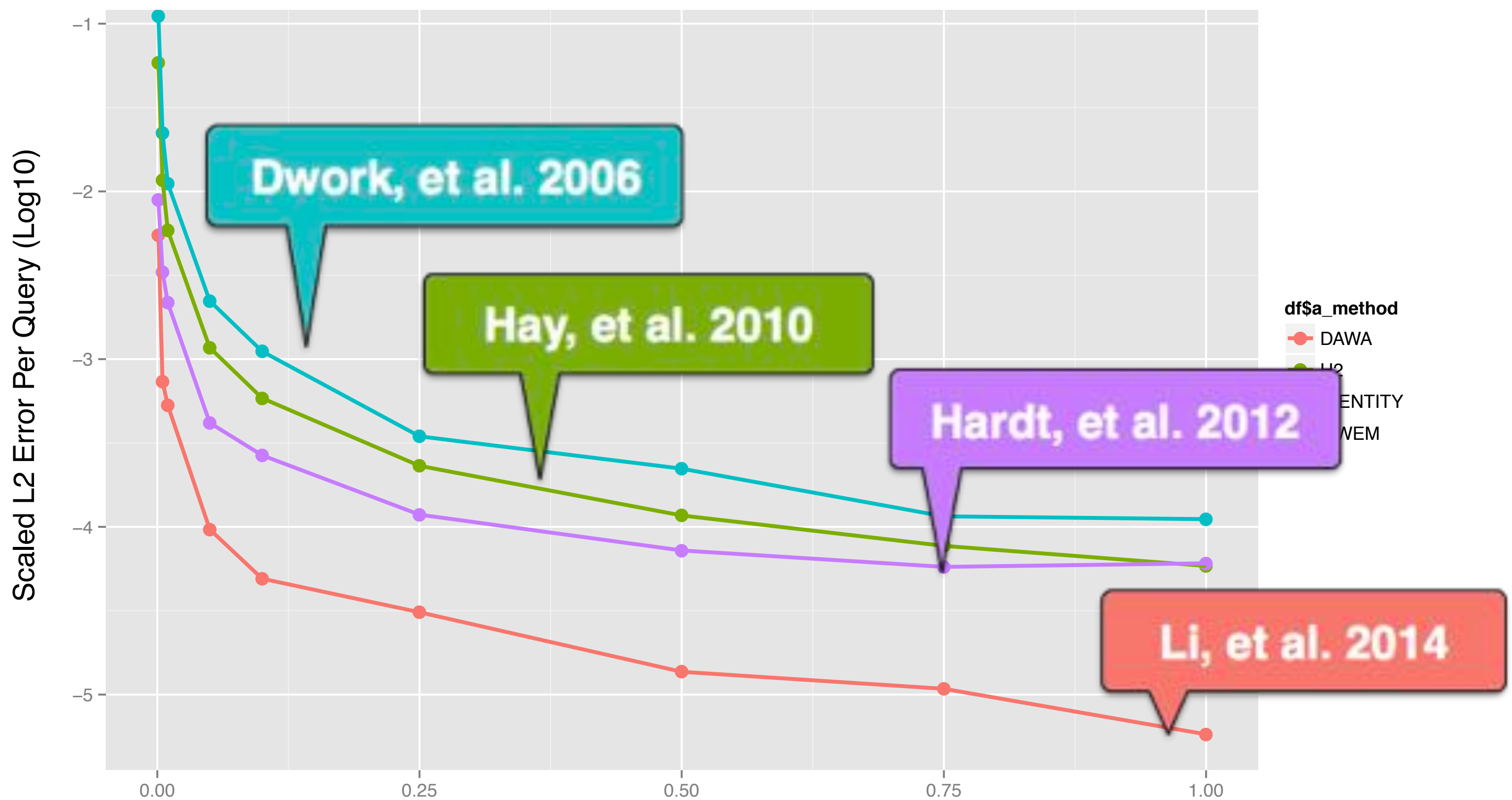
A large fraction of SF-1 are **linear queries** on persons

Algorithms for linear queries

Less Useful

Utility (Error)

More Useful

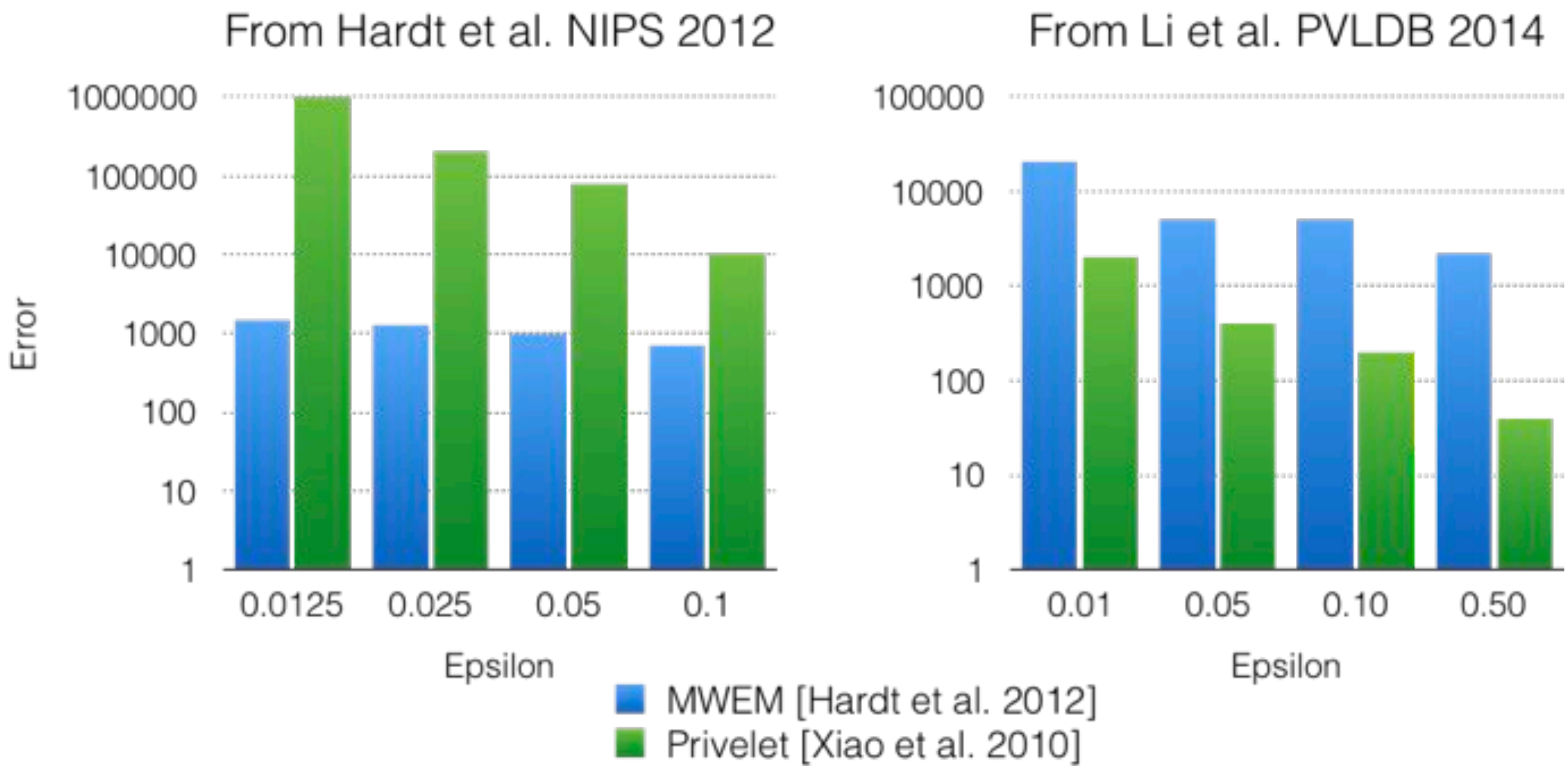


More Private

Privacy (Epsilon)

Less Private

But the story is more nuanced ...



Obstacle to adoption

- Practical performance of privacy algorithms is **opaque to users**.

- Literature has **conflicting evidence** on best algorithms

- Privacy non-experts default to the simplest algorithms like Laplace Mechanism.

DPBench

- A benchmark study of algorithms for answering linear counting queries in low dimensions
  - 15 published algorithms evaluated under
  - ~8,000 distinct experimental configurations

SIGMOD 2016

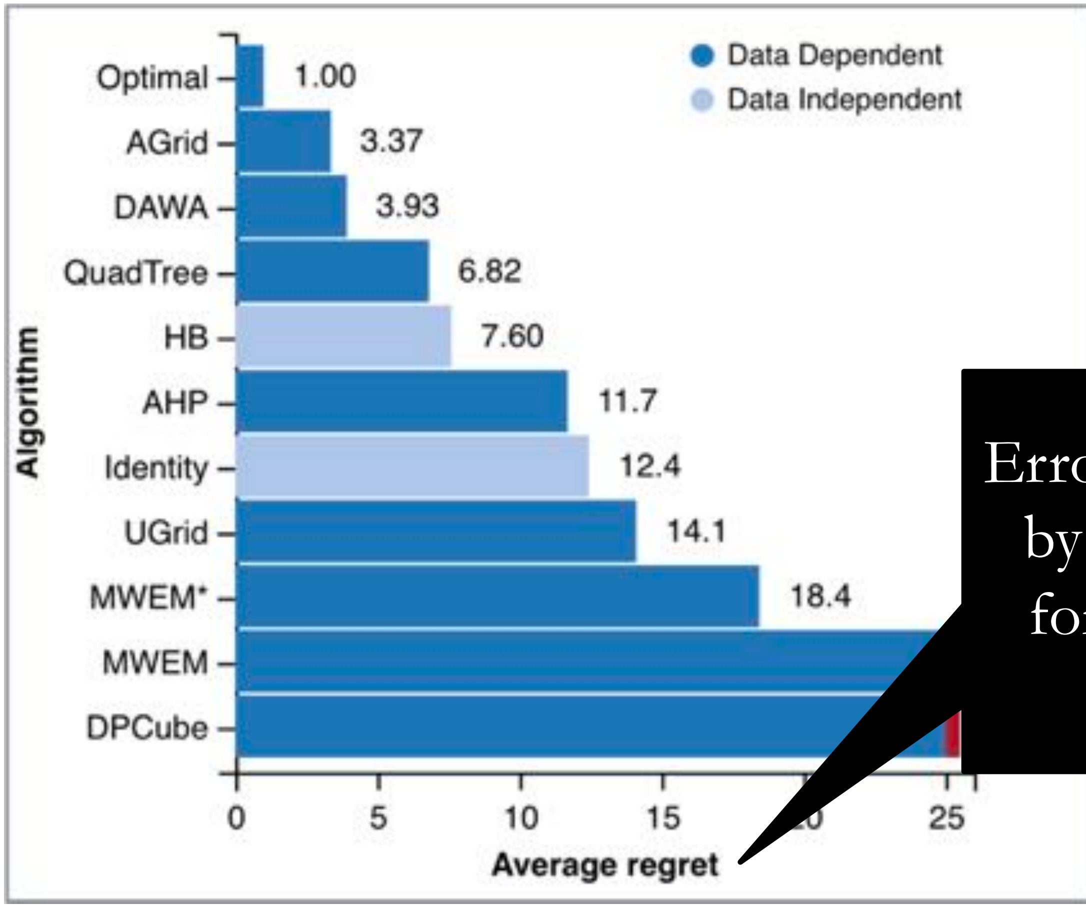

Yan Chen


Dan Zhang


Gerome Miklau

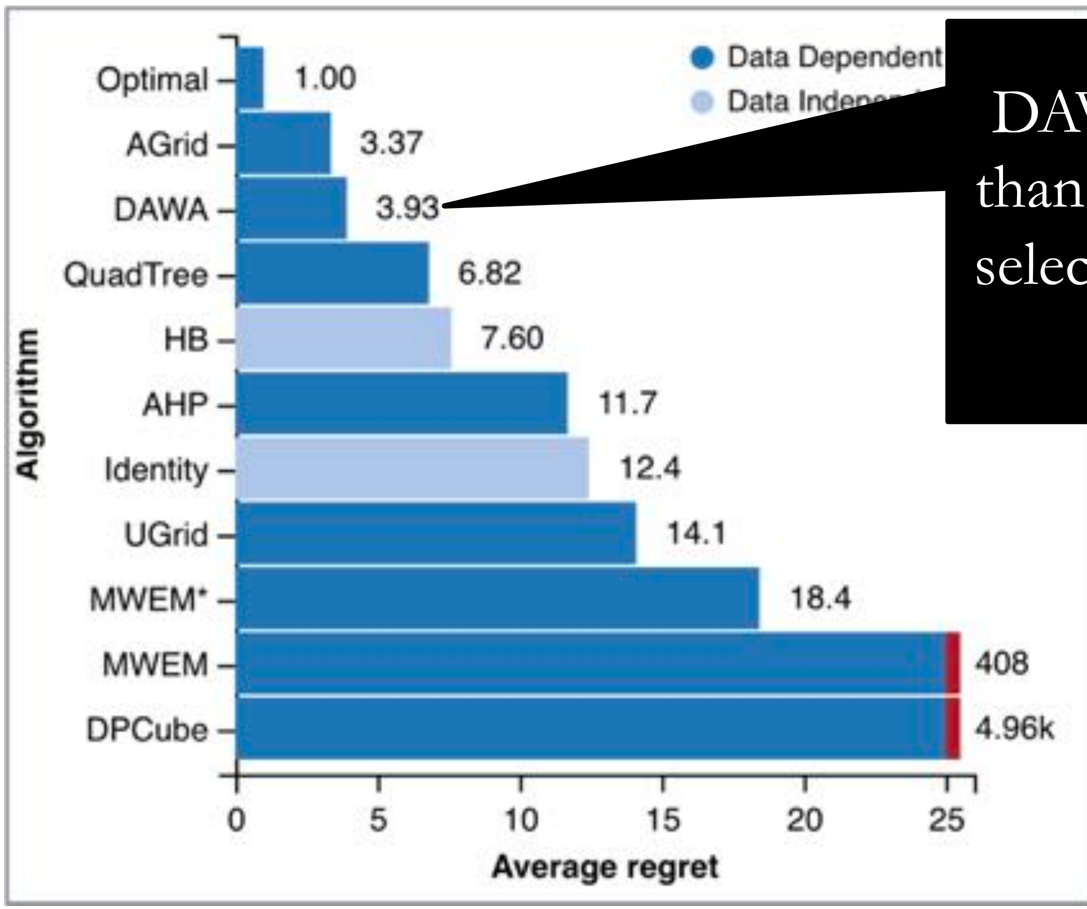

Michael Hay

Key Finding: No algorithm to rule them all



Error of algorithm A divided by error of best algorithm for given dataset averaged over 54 datasets

Key Finding: No algorithm to rule them all

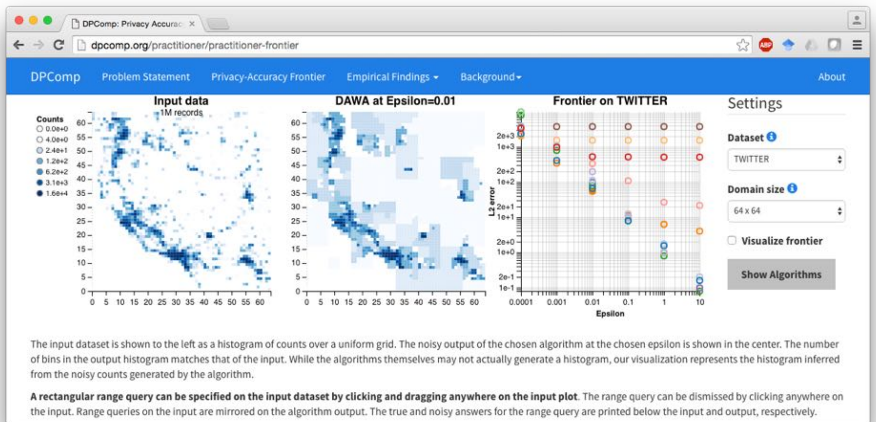


DAWA has ~4x more error than an oracle that somehow selects the best algorithm for each dataset

Visualizing the state of the art

DPComp

SIGMOD 2016



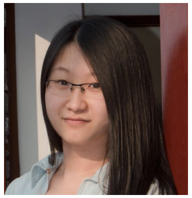
Colgate UNIVERSITY

Duke UNIVERSITY

UMASS AMHERST

Yan Chen

George Bissias



Dan Zhang

Gerome Miklau

Michael Hay

DPBench/DPComp

- Identifies the state-of-the art for low-dimensional counting queries …

- … but, algorithm design for a new task is still a challenge

Toward algorithm synthesis

```
D = ProtectedDataSource(source_uri)
D = D.filter(lambda row: row.sex == 'M'
            and row.age//10 == 3)
   .map(lambda row: row.salary)
x = D.vectorize(n=10**6)

Wpre = PrefixMeasurement(len(x))

R = DomainReductionDawa(x, epsilon/2)
x = x.reduce(R)
Wpre = Wpre.reduce(R)

M = GreedyHierarchyMeasurement(Wpre)
y = x.VectorLaplace(M, epsilon/2)
x_hat = LeastSquares(M, y)

return dot_product(Wpre, x_hat)
```

This algorithm computes CDF of salaries for males in 30s

Toward algorithm synthesis

```
D = ProtectedDataSource(source_uri)
D = D.filter(lambda row: row.sex == 'M'
             and row.age//10 == 3)
    .map(lambda row: row.salary)
x = D.vectorize(n=10**6)

Wpre = PrefixMeasurement(len(x))
```

Preprocessing & Input creation

```
R = DomainReductionDawa(x, epsilon/2)
x = x.reduce(R)
Wpre = Wpre.reduce(R)

M = GreedyHierarchyMeasurement(Wpre)
y = x.VectorLaplace(M, epsilon/2)
x_hat = LeastSquares(M, y)
```

DP Logic

```
return dot_product(Wpre, x_hat)
```

Algorithms to plans

```
D = ProtectedDataSource(source_uri)
D = D.filter(lambda row: row.sex == 'M'
            and row.age//10 == 3)
    .map(lambda row: row.salary)
x = D.vectorize(n=10**6)

Wpre = PrefixMeasurement(len(x))

R = DomainReductionDawa(x, epsilon/2)
x = x.reduce(R)
Wpre = Wpre.reduce(R)

M = GreedyHierarchyMeasurement(Wpre)

y = x.VectorLaplace(M, epsilon/2)

x_hat = LeastSquares(M, y)

return dot_product(Wpre, x_hat)
```

Data transformation

Data Reduction

Query Selection

Private Measurement

Inference

DAWA [VLDB 2014]

```
D = ProtectedDataSource(source_uri)
D = D.filter(lambda row: row.sex == 'M'
          and row.age//10 == 3)
   .map(lambda row: row.salary)
x = D.vectorize(n=10**6)

Wpre = PrefixMeasurement(len(x))
```

```
R = DomainReductionDawa(x, epsilon/2)
x = x.reduce(R)
Wpre = Wpre.reduce(R)
```

Data Reduction

```
M = GreedyHierarchyMeasurement(Wpre)
```

Query Selection

```
y = x.VectorLaplace(M, epsilon/2)
```

Private Measurement

```
x_hat = LeastSquares(M, y)
```

Inference

```
return dot_product(Wpre, x_hat)
```

AHP [SDM 2014]

```
D = ProtectedDataSource(source_uri)
D = D.filter(lambda row: row.sex == 'M'
             and row.age//10 == 3)
    .map(lambda row: row.salary)
x = D.vectorize(n=10**6)

Wpre = PrefixMeasurement(len(x))

R = ClusterAHP(x.VectorLaplace(Identity(len(x)), epsilon/2))
x = x.reduce(R)
Wpre = Wpre.reduce(R)

M = Identity(len(x))

y = x.VectorLaplace(M, epsilon/2)

x_hat = LeastSquares(M, y)

return dot_product(Wpre, x_hat)
```

Data Reduction

Query Selection

Private Measurement

Inference

Operator classes and instances

Transform	
TV	T-Vectorize
TP	V-Partition
TR	V-Reduce

Inference	
LS	Least squares
NLS	NN Least squares
MW	Mult Weights
HR	Thresholding

Reduction selection	
RA	AHPcluster
RG	Grid
RD	Dawa
RW	Workload-based
RS	Stripe(attr)

Query	
LM	Vector Laplace

Query selection	
SI	Identity
ST	Total
SP	Privelet
SH2	H2
SHB	HB
SG	Greedy-H
SU	UniformGrid
SA	AdaptiveGrids
SQ	Quadtree
SW	Worst-approx
SPB	PrivBayes select

- **Private** operators change the database, but have no output
- **Private → Public** operators release differentially private answers
- **Public** operators are postprocessing

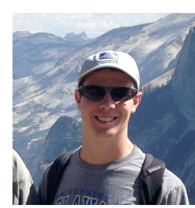
Ektelo

TPDP 2017

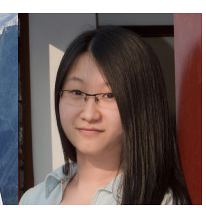
- A system for describing differentially private algorithms as plans composed of vetted operator implementations
  - Currently supports algorithms that answer sets of linear queries

- Any ektelo plan satisfies differential privacy
- Can express many state of the art algorithms
- Can create new algorithms by composing operator implementations

Ios Kotsogiannis



Ryan Mckenna



Dan Zhang

Gerome Miklau

Michael Hay

DP Algorithms in Ektelo

ID	Cite	Algorithm name	Plan signature
1	[8]	Identity	SI LM
2	[39]	Privelet	SP LM LS
3	[17]	Hierarchical (H2)	SH2 LM LS
4	[34]	Hierarchical Opt (HB)	SHB LM LS
5	[22]	Greedy-H	SG LM LS
6	-	Uniform	ST LM LS
7	[15]	MWEM	I:(SW LM MW)
8	[42]	AHP	SI LM HR RA TR SI LM LS
9	[22]	DAWA	RD TR SG ML LS
10	[6]	Quadtree	SQ LM LS
11	[33]	UniformGrid	SU LM LS
12	[33]	AdaptiveGrid	SU LM LS TP[SA LM] LS
13	NEW	DAWA-Striped	RS TP[RD TR SG LM] LS
14	NEW	HB-Striped	RS TP[SHB LM] LS
15	NEW	PrivBayesLS	SPB LM LS
16	NEW	MWEM variant b	I:(SW LM NLS)
17	NEW	MWEM variant c	I:(SW SH2 LM MW)
18	NEW	MWEM variant d	I:(SW SH2 LM NLS)

DPBench Algorithms (IDs 1–12)

New Algorithms (IDs 13–18)

ε ktelo

- Code reuse
  - **Unified 18 implementations** of the Laplace mechanism in DPBench algorithms
- Improved operator implementations
  - **10x runtime improvement** by using a general purpose inference method
- Plan rewrite rules
  - **5x** runtime improvement and **3x** accuracy improvement
- New algorithms by composing operators
  - **10x accuracy improvement** over the state-of-the-art

Summary

- Goal: Empower non-experts to analyze sensitive data **with provably private** algorithms while ensuring **little loss in utility**.

- Needs a shift from theory to systems oriented research

- Number of interesting theoretical and systems research challenges in the context of relational databases yet to be solved to make DP practical.

Thank you ☺

[SIGMOD 11] D. Kifer, A. Machanavajjhala, "No Free Lunch in Data Privacy"

[TODS 14] D. Kifer, A. Machanavajjhala, "Pufferfish"

[SIGMOD 14] X. He, A. Machanavajjhala, B. Ding, "Blowfish privacy"

[VLDB 15] S. Haney, A. Machanavajjhala, B. Ding, "Design of Policy-Aware DP Algs"

[ICDE 08] A. Machanavajjhala, D. Kifer, J. Gehrke, J. Abowd, L. Vilhuber, "Privacy: From theory to practice on the map"

[SIGMOD 17] S. Haney, A. Machanavajjhala, J. Abowd, M. Graham, M. Kutzbach, L. Vilhuber, "Utility Cost of Formal Privacy for Releasing National Employer-Employee Statistics"

[SIGMOD 16] M. Hay, A. Machanavajjhala, G. Miklau, Y. Chen, D. Zhang, "Principled evaluation of differentially private algorithms using DPBench"

[SIGMOD 17] I. Kotsogiannis, A. Machanavajjhala, M. Hay, G. Miklau, "Pythia"

[TPDP 17] D. Zhang, R. McKenna, I. Kotsogiannis, G. Miklau, M. Hay, A. Machanavajjhala " ε ktelo: A Framework for Defining DP Computations"