

*Privacy:  
A Natural Resource to Be Preserved*

Cynthia Dwork, Microsoft

# *The Promise of Data*

---

- ▶ Parking spot near favorite restaurant; parking/driving violations
- ▶ Maintenance
  - ▶ Power management, climate control
  - ▶ safer flying, oil refining
- ▶ Locating the defibrillator
- ▶ My car adapting to your erratic braking technique
- ▶ Advertising based on click stream analysis
- ▶ New professional contacts
- ▶ Medical Applications
  - ▶ Saving your life in the ER by getting your medical information
  - ▶ Learning medical facts, eg genotype/phenotype correlations
- ▶ Allocation of resources
  - ▶ Utilization of spare parts / employees
  - ▶ Usage of public funds, representation in congress



# *The Threat of Data*

---

- ▶ Parking spot near favorite restaurant; parking/driving violations
- ▶ Maintenance
  - ▶ Power management, climate control
  - ▶ safer flying, oil refining
- ▶ Locating the defibrillator
- ▶ My car adapting to your erratic braking technique
- ▶ Advertising based on click stream analysis
- ▶ New professional contacts
- ▶ Medical Applications
  - ▶ Saving your life in the ER by getting your medical information
  - ▶ Learning medical facts, eg genotype/phenotype correlations
- ▶ Allocation of resources
  - ▶ Utilization of spare parts / employees
  - ▶ Usage of public funds, representation in congress



# *The Threat of Data*

---

- ▶ Parking spot near favorite restaurant; parking/driving violations
  - ▶ Maintenance
    - ▶ Power management, climate control
    - ▶ safer flying, oil refining
  - ▶ Locating the defibrillator
  - ▶ My car adapting to your erratic braking technique
  - ▶ Advertising based on click stream analysis
  - ▶ New professional contacts
  - ▶ Medical Applications
    - ▶ Saving your life in the ER by getting your medical information
    - ▶ Learning medical facts, eg genotype/phenotype correlations
  - ▶ Allocation of resources
    - ▶ Utilization of spare parts / employees
    - ▶ Usage of public funds, representation in congress
  - ▶ **Andreas Weigend's entire talk**
- 



# *Our Focus: Trusted (and Trustworthy) Curator*

---

## Privacy-Preserving Analysis of Confidential Data

- ▶ Mathematical Definition of Privacy
- ▶ Finding Statistical Correlations
  - ▶ Analyzing medical data
  - ▶ Correlating cough outbreak with chemical plant malfunction
    - Can't be done with HIPAA safe-harbor sanitized data
- ▶ Noticing Events
  - ▶ Detecting spike in ER admissions for asthma
- ▶ Datamining Tasks
  - ▶ Clustering; learning association rules, decision trees, separators; principal component analysis
- ▶ Official Statistics
  - ▶ Contingency Table Release



# *Hasn't This Been Done Before?*

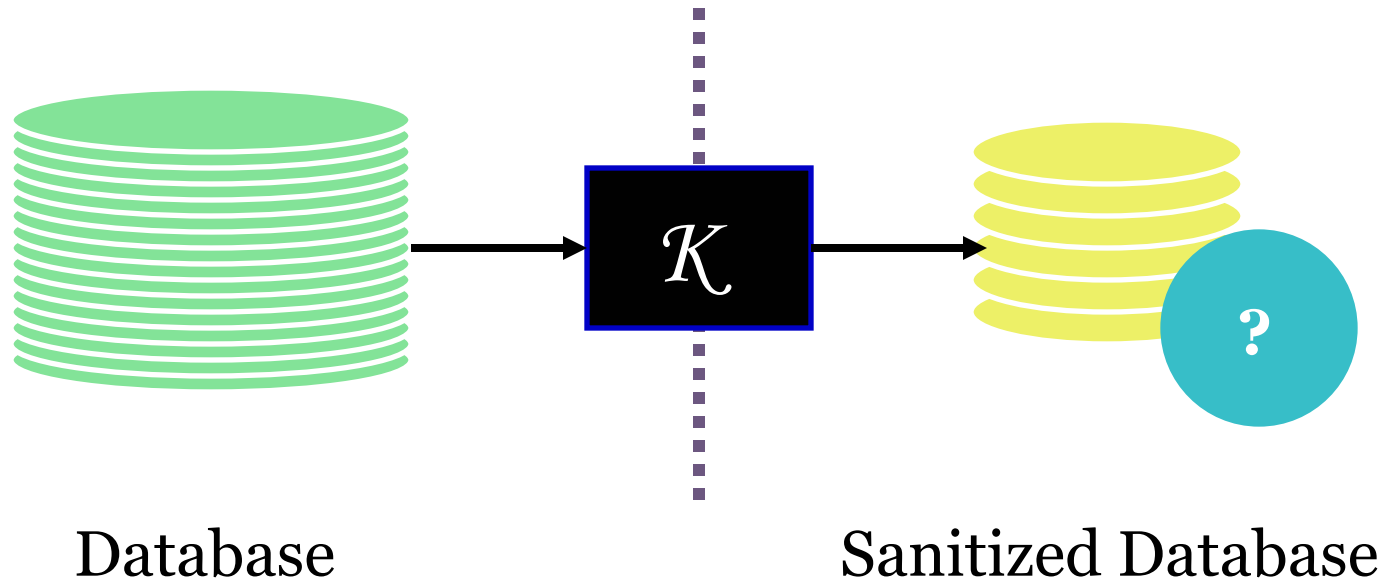
---

▶ Yes.



# Two Models

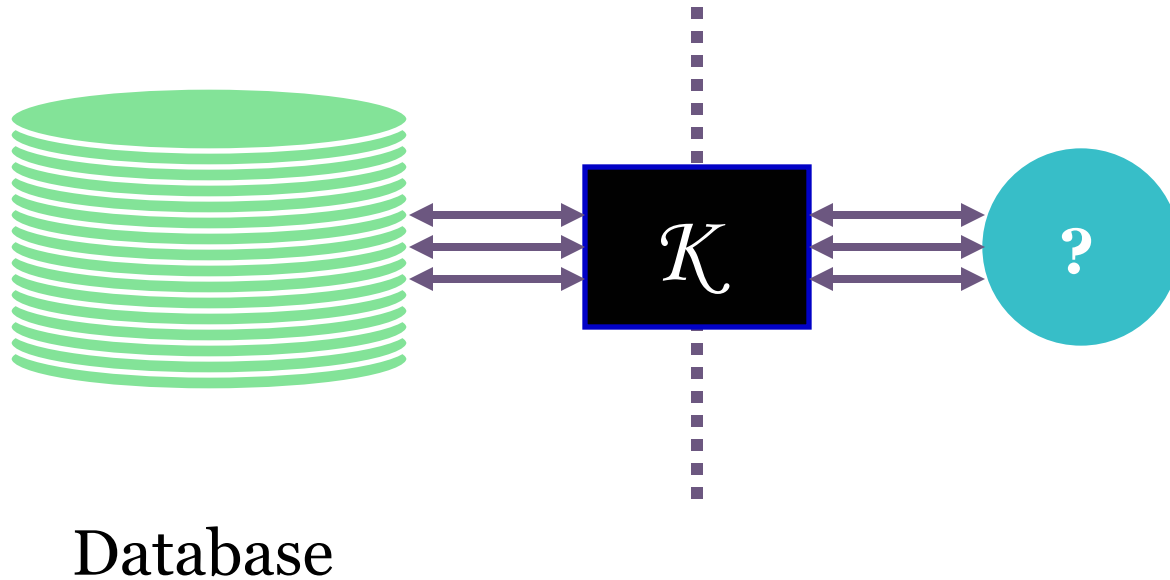
---



Non-Interactive: Data are sanitized and released

# Two Models

---



Interactive: Multiple Queries, Adaptively Chosen

# *Outline*

---

- ▶ Broken Privacy
- ▶ Wrong Privacy Promises
- ▶ A Right Privacy Promise: Differential Privacy
- ▶ Achieving Differential Privacy
- ▶ Limitations
- ▶ Summary and Open Questions



## *Linkage Attacks: A Special Case of Auxiliary Data*

---

- ▶ Using “innocuous” data in one dataset to identify a record in a different dataset containing both innocuous and sensitive data
- ▶ At the heart of the voluminous research on hiding small cell counts in tabular data



# NETFLIX

# Netflix Prize

Home Rules Leaderboard Register Update Submit Download

## NETFLIX

Browse Recommendations Friends Queue Buy DVDs  
Home Genres New Releases Previews Netflix Top 100 Crit

### Movies For You

Randy, the following movies were chosen based on your interest in:  
[Bowling for Columbine](#)  
[Carnivale: Season 1](#)  
[Fahrenheit 9/11](#)

**The Big One**

★★★★☆

...er subversive  
...ly from  
...n /  
...ichael

**Carnivale: Season 2**

Disc Series

★★★★☆

Daniel Kraus  
...rivingly cre  
...series cont  
...document t

**All**

...entures of a motley cre  
...nies who've made the C  
...stbowl their ... [Read Mo](#)

All Discs Guaranteed

### You really liked it...

Now own it for just \$5.99

Shop as low

titles

Original art

# Welcome!

The Netflix Prize seeks to substantially improve the accuracy of predictions about how much someone is going to love a movie based on their movie preferences. Improve it enough and you win one (or more) Prizes. Winning the Netflix Prize improves our ability to connect people to the movies they love.

Read the [Rules](#) to see what is required to win the Prizes. If you are interested in joining the quest, you should [register a team](#).

You should also read the [frequently-asked questions](#) about the Prize. And check out how various teams are doing on the [Leaderboard](#).

Good luck and thanks for helping!



# *The Netflix Prize*

---

- ▶ **Netflix Recommends Movies to its Subscribers**
  - ▶ Seeks improved recommendation system
  - ▶ Offers \$1,000,000 for 10% improvement
    - ▶ Not concerned here with how this is measured
  - ▶ Publishes training data



## *From the Netflix Prize Rules Page...*

---

- ▶ “The training data set consists of more than 100 million ratings from over 480 thousand randomly-chosen, anonymous customers on nearly 18 thousand movie titles.”
  - ▶ “The ratings are on a scale from 1 to 5 (integral) stars. To protect customer privacy, all personal information identifying individual customers has been removed and all customer ids have been replaced by randomly-assigned ids. The date of each rating and the title and year of release for each movie are provided.”
- 



## *From the Netflix Prize Rules Page...*

---

- ▶ “The training data set consists of more than 100 million ratings from over 480 thousand randomly-chosen, anonymous customers on nearly 18 thousand movie titles.”
  - ▶ “The ratings are on a scale from 1 to 5 (integral) stars. **To protect customer privacy, all personal information identifying individual customers has been removed and all customer ids have been replaced by randomly-assigned ids.** The date of each rating and the title and year of release for each movie are provided.”
- 



## *A Source of Auxiliary Information*

---

- ▶ **Internet Movie Database (IMDb)**
  - ▶ Individuals may register for an account and rate movies
  - ▶ Need not be anonymous
  - ▶ Visible material includes ratings, dates, comments



# *A Linkage Attack on the Netflix Prize Dataset*

Narayanan & Shmatikov 2006

---

- ▶ “With 8 movie ratings (of which we allow 2 to be completely wrong) and dates that may have a 3-day error, 96% of Netflix subscribers whose records have been released can be uniquely identified in the dataset.”
- ▶ “For 89%, 2 ratings and dates are enough to reduce the set of plausible records to 8 out of almost 500,000, which can then be inspected by a human for further deanonymization.”
- ▶ Watch what you say at the water cooler!
- ▶ Attack prosecuted successfully using the IMDb.
  - ▶ NS draw conclusions about user.
  - ▶ **May be wrong, may be right. User harmed either way.**
    - ▶ Gavison: Protection from being brought to the attention of others



## *Other Linkage Data Attacks in the Literature*

---

- ▶ HMO removes names, releases data (ZIP, Bdate, Gender)
- ▶ However, (Z,B,G) enough to uniquely ID most voters.
  - ▶ [S] observes, and responds! (k-anonymity)
  - ▶ Sweeney circa 1998
- ▶ Unfortunately, can still make inferences about secrets.
  - ▶ [MGK] observes, and responds! (l-diversity)
  - ▶ Machanavajjhala, Gehrke, and Kifer, ICDE 2006
- ▶ Unfortunately, multiple releases can compromise all.
  - ▶ [XT] observes, and responds! (m-invariance)
  - ▶ Xiao and Tao, SIGMOD 2007
- ▶ Next?



# *Analysis of Social Network Graphs*

---

- ▶ “Friendship” Graph
  - ▶ Nodes correspond to users
  - ▶ Users may list others as “friend,” creating an edge
    - ▶ Edges are annotated with directional information
- ▶ Hypothetical Research Question
  - ▶ How frequently is the “friend” designation reciprocated?



# *Anonymization of Social Networks*

---

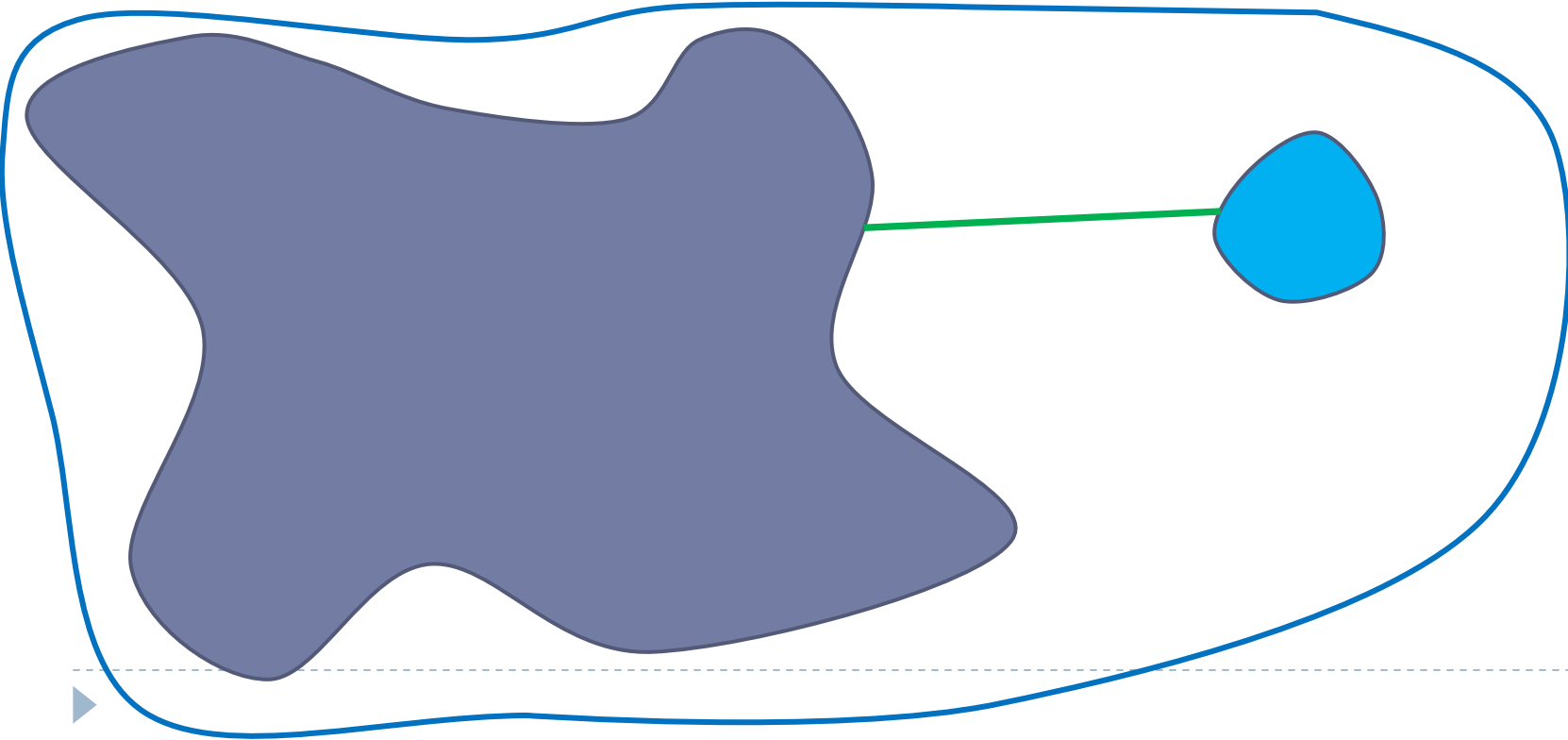
- ▶ Replace node names/labels with random identifiers
- ▶ Permits analysis of the structure of the graph
- ▶ Privacy hope: randomized identifiers make it hard/impossible to identify nodes with specific individuals, thereby hiding the privacy of who is connected to whom
- ▶ **Disastrous!**
  - ▶ Vulnerable to active and passive attacks
  - ▶ Backstrom, Dwork, Kleinberg 2007



## *Flavor of Active Attack*

---

- ▶ **Prior to release, create subgraph of special structure**
  - ▶ Very small: circa 12 nodes
  - ▶ Highly internally connected
  - ▶ Lightly connected to the rest of the graph

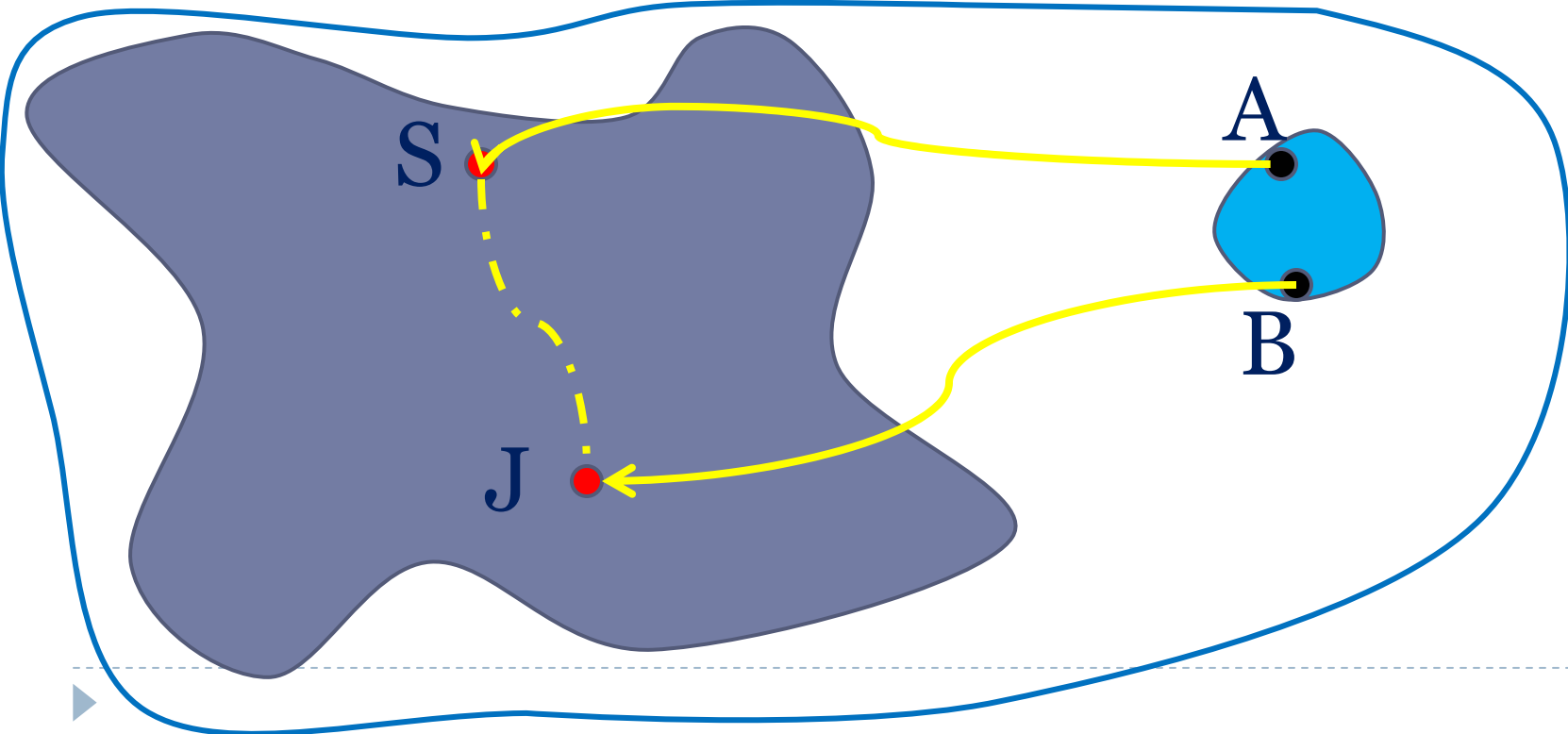


# *Flavor of Active Attack*

---

## ▶ Connections:

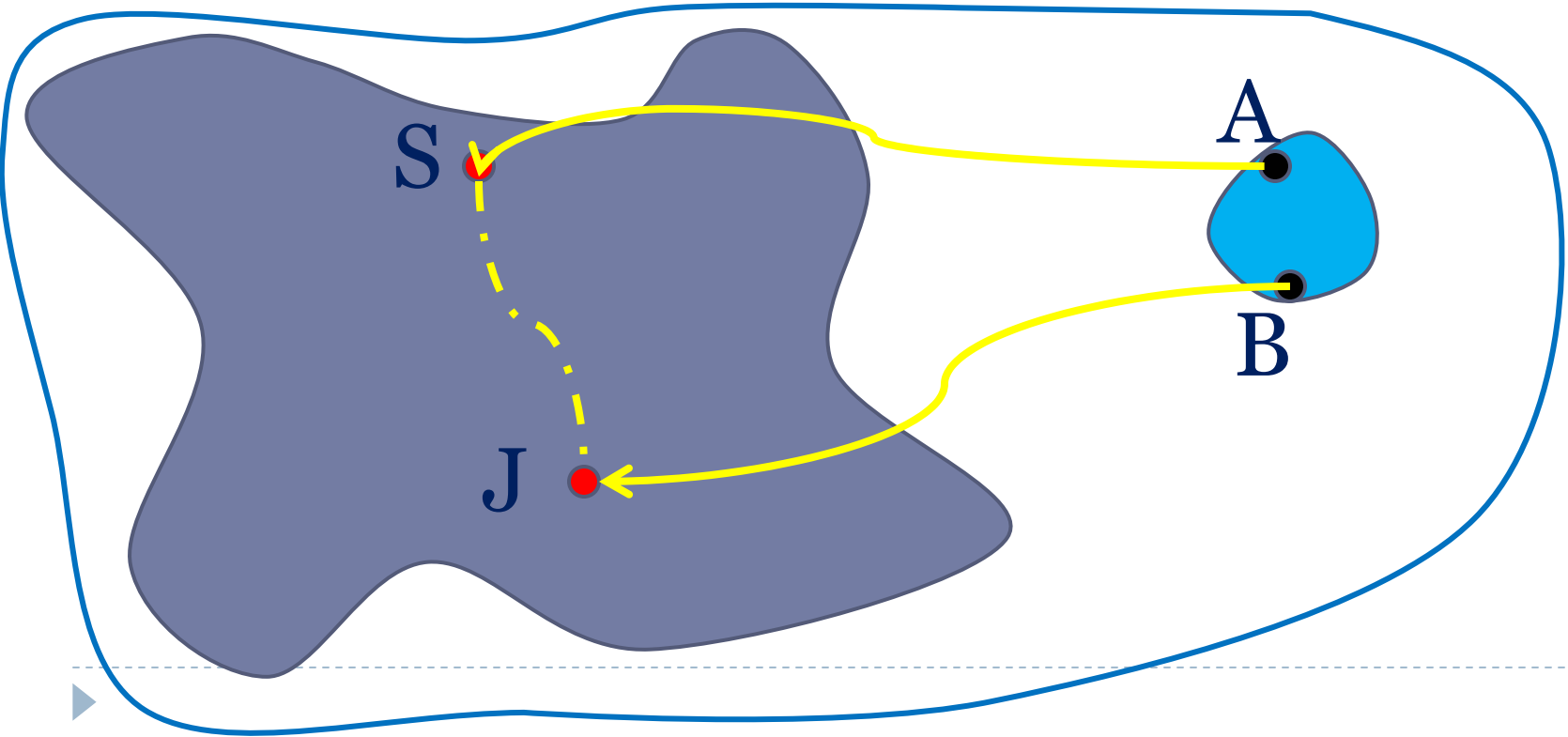
- ▶ Victims: Steve and Jerry
- ▶ Attack Contacts: A and B
- ▶ Finding A and B allows finding Steve and Jerry



# *Flavor of Active Attack*

---

- ▶ **Magic Step**
  - ▶ Isolate lightly linked-in subgraphs from rest of graph
  - ▶ Special structure of subgraph permits finding A, B



# *What is Going Wrong?*

---

- ▶ Guarantees are Syntactic, not Semantic
  - ▶ k, l, m
  - ▶ name replaced with random string
- ▶ Ad Hoc!
  - ▶ Privacy compromise defined to be a certain set of undesirable outcomes
  - ▶ No argument that this set is exhaustive or completely captures privacy
- ▶ Failure to account for auxiliary information
  - ▶ *In vitro vs in vivo*



# *Getting it Right in Cryptography:* *Semantic Security Against an Eavesdropper* [GM'82]

---

## ▶ Vocabulary

- ▶ Plaintext: the message to be transmitted
- ▶ Ciphertext: the encryption of the plaintext
- ▶ Auxiliary information: anything else known to attacker

▶ The ciphertext leaks no information about the plaintext.

## ▶ Formalization

Compare the ability of someone **seeing aux and ciphertext** to guess (anything about) the plaintext, to the ability of someone **seeing only aux** to do the same thing. Difference should be “tiny”.



# Statistical Databases

---

- ▶ Dalenius, 1977
  - ▶ Anything that can be learned about a respondent from the statistical database can be learned without access to the database
  - ▶ An ad omnia guarantee
- ▶ Happily, Formalizes to Semantic Security
  - ▶ Recall: Anything about the plaintext that can be learned from the ciphertext can be learned without the ciphertext
  - ▶ Popular Intuition: prior and posterior views about an individual shouldn't change "too much".
    - ▶ Clearly Silly
      - My (incorrect) prior is that everyone has 2 left feet.
      - Very popular in literature nevertheless
      - Definitional awkwardness even when used correctly

# *Semantic Security for Statistical Databases?*

---

- ▶ Dalenius, 1977
  - ▶ Anything that can be learned about a respondent from the statistical database can be learned without access to the database.
- ▶ Happily, Formalizes to Semantic Security
- ▶ Unhappily, Unachievable [Dwork and Naor 2006]
  - ▶ Both for not serious and serious reasons.

# *Semantic Security for Statistical Databases is Impossible*

---

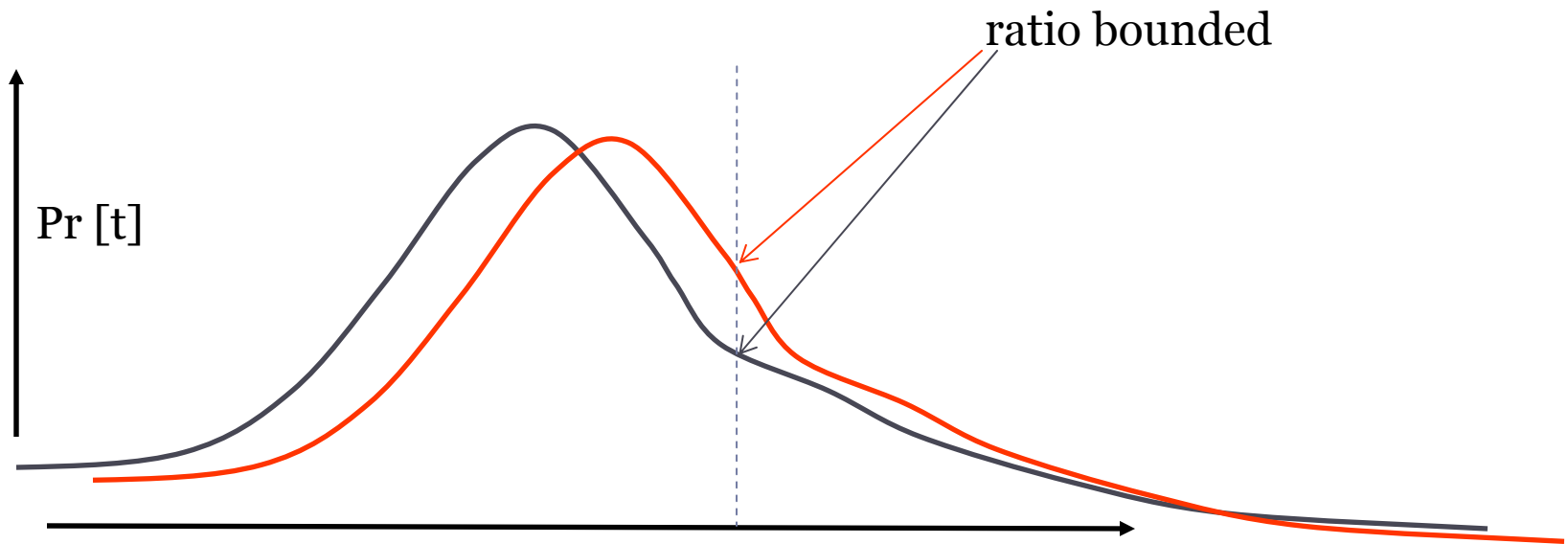
- ▶ **The Serious Reason (told as a parable)**
  - ▶ Database teaches average heights of population subgroups
  - ▶ “Terry Gross is 2 inches shorter than avg Swedish woman”
  - ▶ Access to DB teaches Terry’s height.
  - ▶ Terry’s height learnable from the DB, not learnable w/o.
- ▶ **Proof extends to “any” notion of privacy breach.**
- ▶ **Attack Works Even if Terry Not in DB!**
  - ▶ Suggests new notion of privacy: risk incurred by joining DB
    - ▶ “Differential Privacy”
  - ▶ Before/After interacting vs Risk when in/not in DB



# Differential Privacy

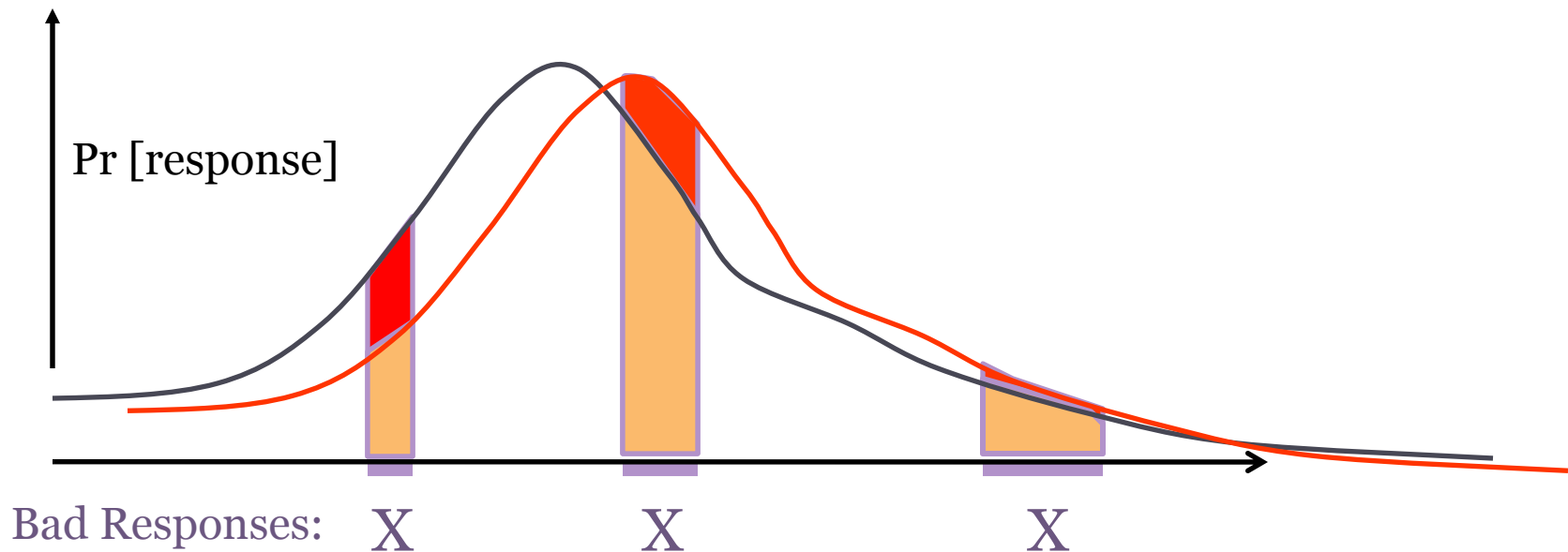
$\mathcal{K}$  gives  $\epsilon$ -differential privacy if for all values of DB and Me and all outputs t:

$$e^{-\epsilon} \leq \frac{\Pr[\mathcal{K}(\text{DB} + \text{Me}) = t]}{\Pr[\mathcal{K}(\text{DB} - \text{Me}) = t]} \leq e^{\epsilon} \quad (1+\epsilon)$$



# *Differential Privacy is an Ad Omnia Guarantee*

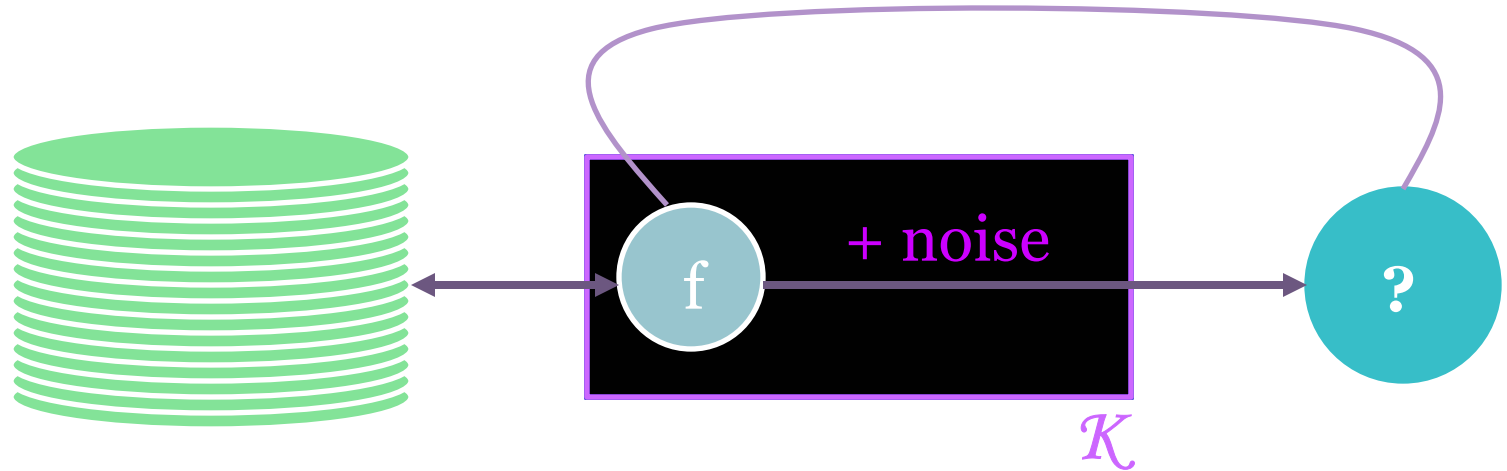
- ▶ No perceptible risk is incurred by joining DB.
- ▶ Anything adversary can do to me, it could do whether or not my data are in the dataset.



# *An Interactive Sanitizer: $\mathcal{K}$*

Dwork, McSherry, Nissim, Smith 2006

---



$f: \text{DB} \rightarrow \mathbb{R}$

$$\mathcal{K}(f, \text{DB}) = f(\text{DB}) + \text{Noise}$$

Eg,  $\text{Count}(P, \text{DB}) = \# \text{ rows in DB with Property P}$

---

## *Sensitivity of a Function $f$*

---

How Much Can  $f(\text{DB} + \text{Me})$  Exceed  $f(\text{DB} - \text{Me})$ ?

Recall:  $\mathcal{K}(f, \text{DB}) = f(\text{DB}) + \text{noise}$

Question Asks: What difference must noise obscure?

$$\Delta f = \max_{\text{DB}, \text{Me}} |f(\text{DB} + \text{Me}) - f(\text{DB} - \text{Me})|$$

eg,  $\Delta \text{Count} = 1$

# *Multiple /Complex Queries*

---

- ▶ **Noise magnitude increases**
  - ▶ Depends on the sensitivity of the query sequence
  - ▶ For certain classes of queries this is essential [Dinur-Nissim]
    - ▶ Even if a substantial fraction of queries may incur arbitrary noise [DMT, DY]
- ▶ **Speaks to the Non-Interactive Setting**
  - ▶ **Any non-interactive solution permitting “too accurate” answers to “too many” questions is vulnerable.**
- ▶ **Crucial in Arguing that Privacy has a Price**
  - ▶ There is no safe way to avoid increasing the noise as the number of queries increases



## *Summary: Achieved Much*

---

- ▶ **Defined Differential Privacy**
  - ▶ Proved classical goal (Dalenius, 1977) unachievable
  - ▶ “Ad Omnia” definition; independent of auxiliary information
- ▶ **General Approach; Rigorous Proof**
  - ▶ Relates degree of distortion to the (mathematical) sensitivity of the computation needed for the analysis
    - ▶ “How much” can the data of one person affect the outcome?
  - ▶ Cottage Industry: redesigning algorithms to be insensitive
- ▶ **Assorted Extensions**
  - ▶ When: noise makes no sense [MT], actual sensitivity is much less than worst-case [NRS], database is distributed [DKMMN], ...
- ▶ **Lower bounds on distortion**



## *Future Work*

---

- ▶ **Differential Privacy for Social Networks, Graphs**
  - ▶ What are the utility questions of interest?
- ▶ **Definitional Work for Other Settings**
  - ▶ “Differential” approach more broadly useful



---

*Privacy is a natural resource.  
It's non-renewable, and it's not yours.  
Conserve it.*

