Differentially Private Data Analysis of Social Networks via Restricted Sensitivity

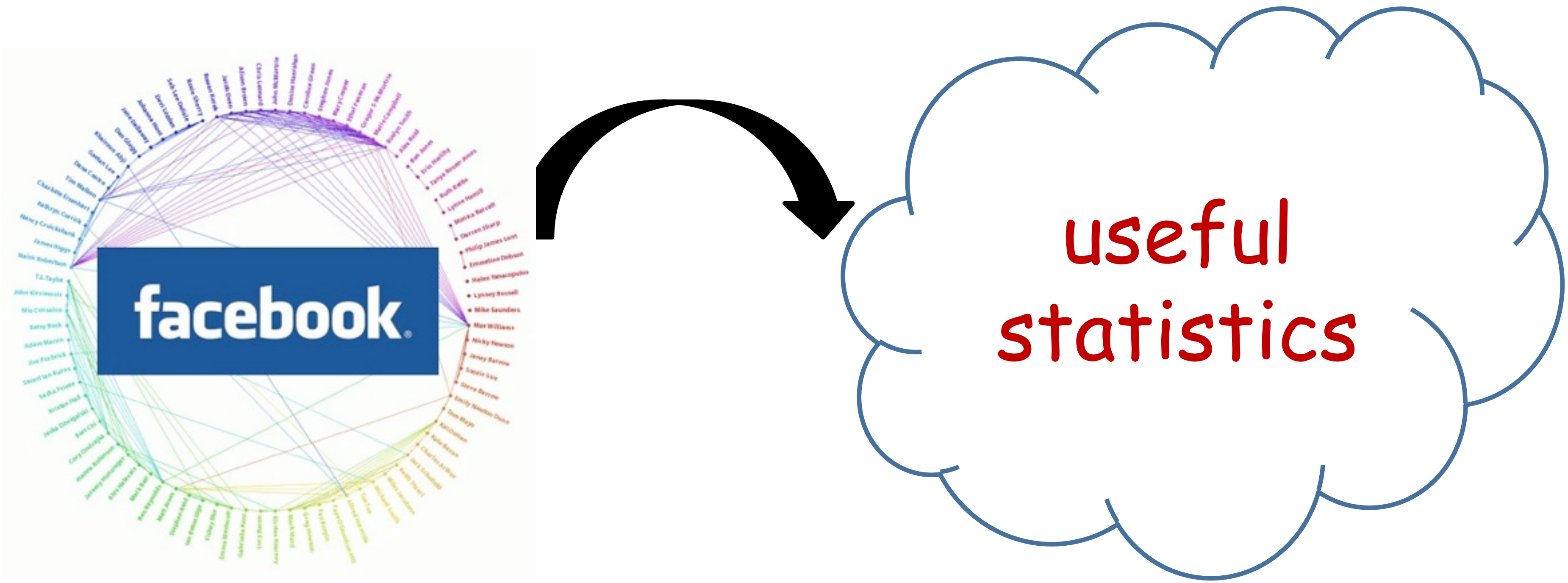
Jeremiah Blocki   Avrim Blum

Anupam Datta   Or Sheffet

Carnegie Mellon University

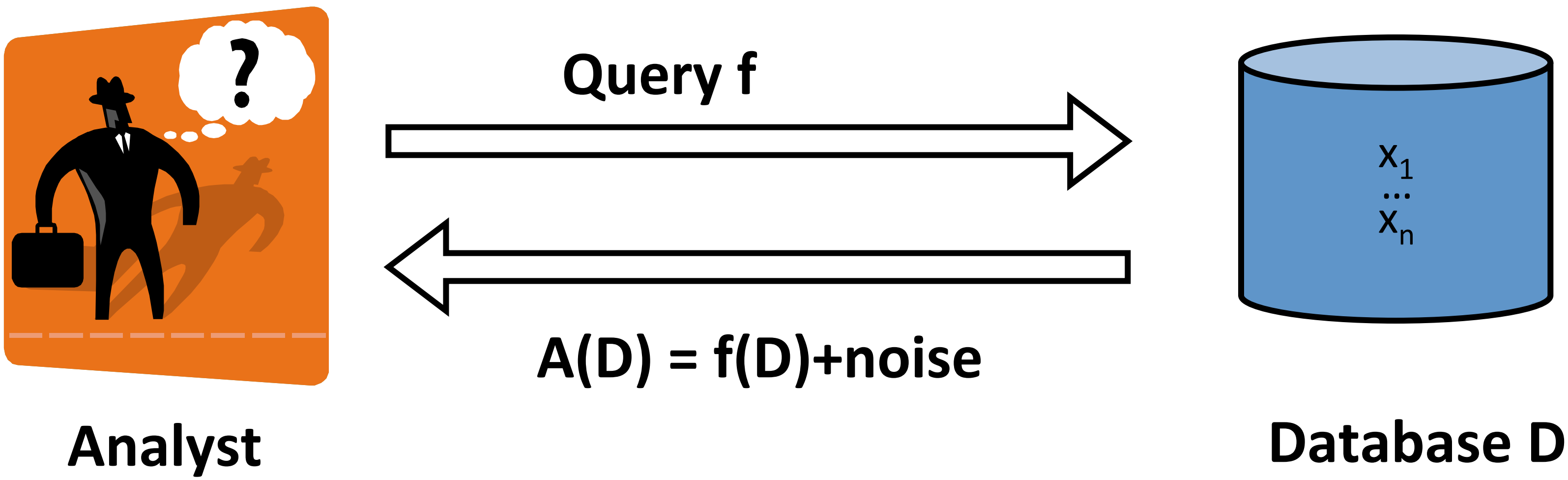
Foundations of Privacy, Fall 2014

Goal



Preserve Privacy and Release Useful Statistics

Usual Differential Privacy Setting



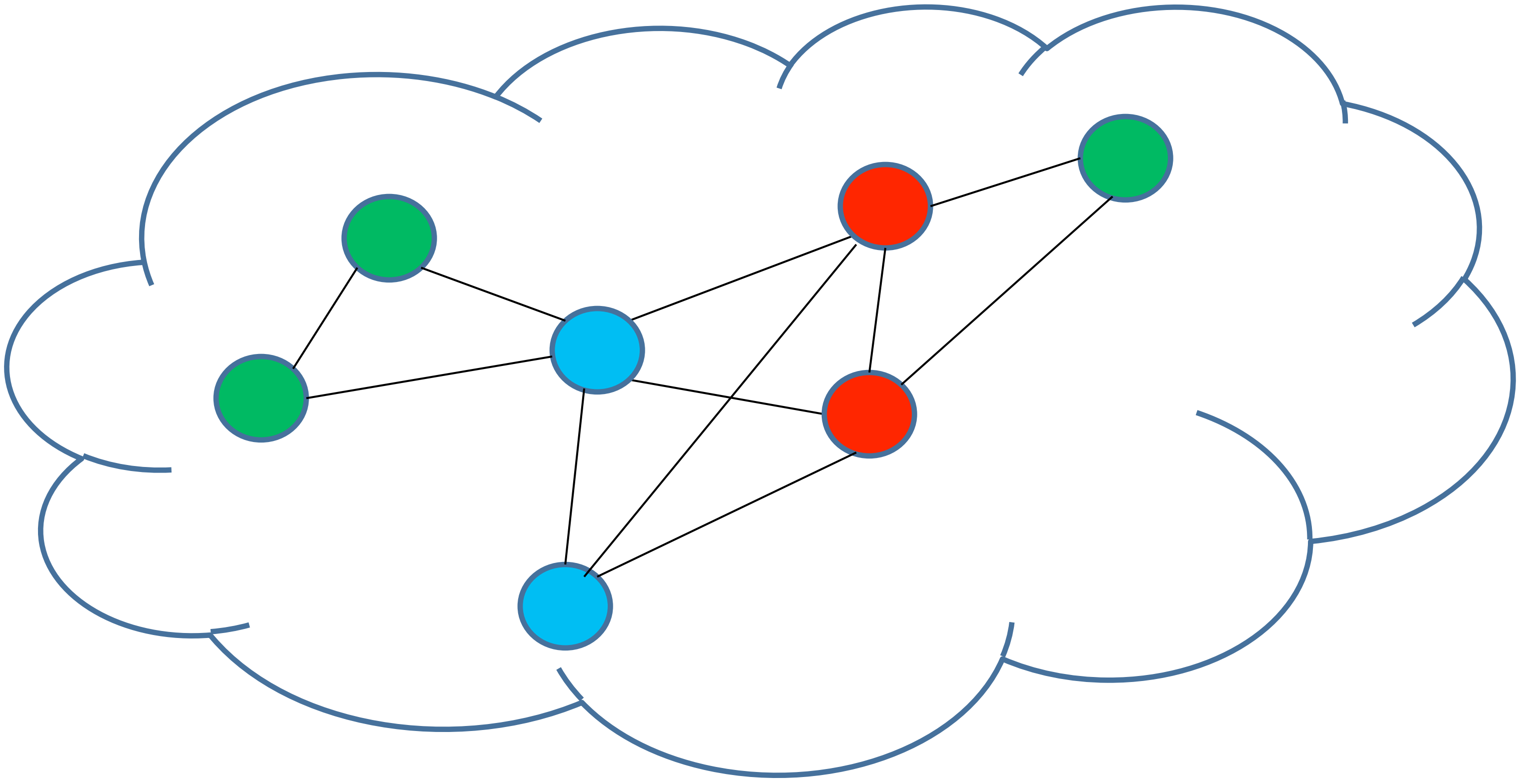
Us al:

- cur for al
- erentia cy

Outline

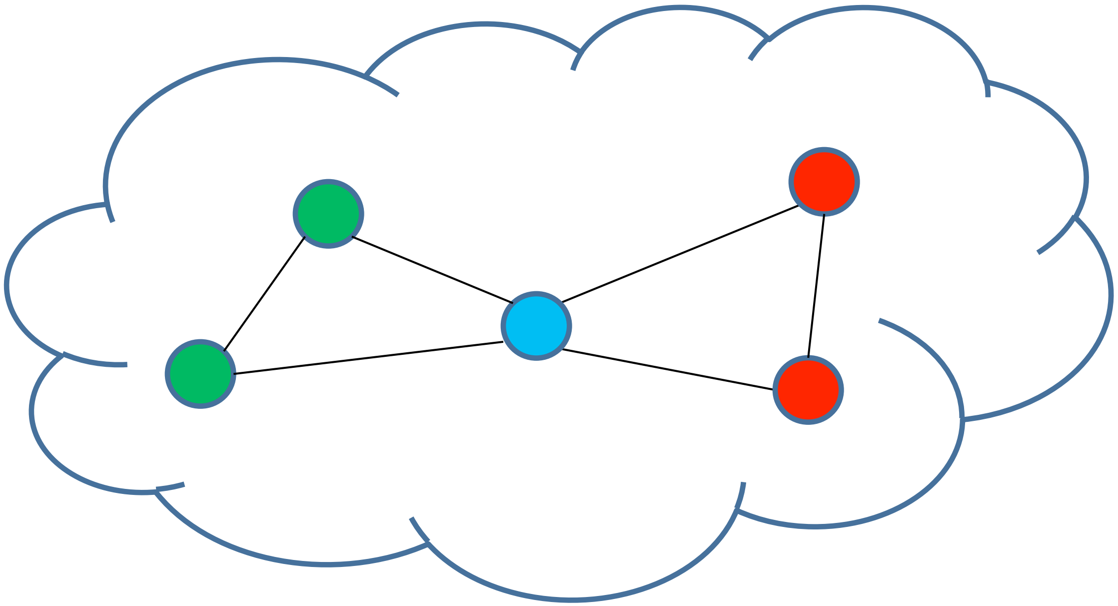
- **Background**
  - Social Networks
  - Differential Privacy
- The Problem
- Restricted Sensitivity
- Algorithms

Social Network



Vertices in a social network G [illegible] G are labeled (e.g., doctor, lawyer, professor).

Local Profile Query



How many people know 2 lawyers who know each other and 2 doctors who know each other, but the lawyers aren't friends with the doctors?

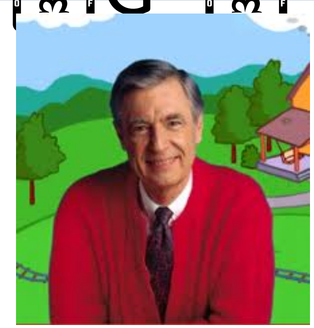
Differential Privacy (Dwork et al)

An algorithm A satisfies (□,□)-differential privacy for social networks if for any S □ Range(A)

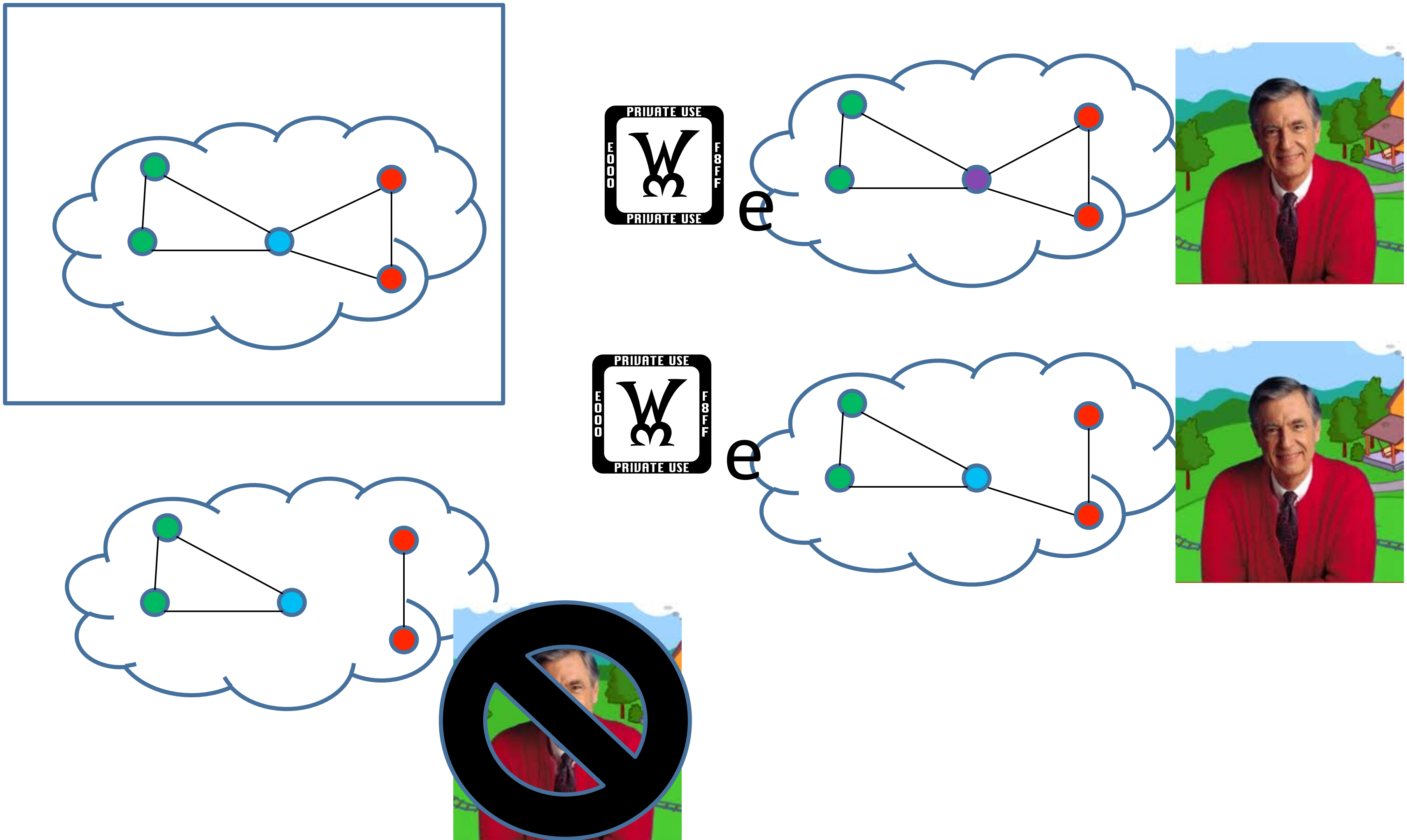
Pr[A(G)□S] ≤ e^□ Pr[A(G')□S] + □

for every *neighboring* social networks G□G' □ G.

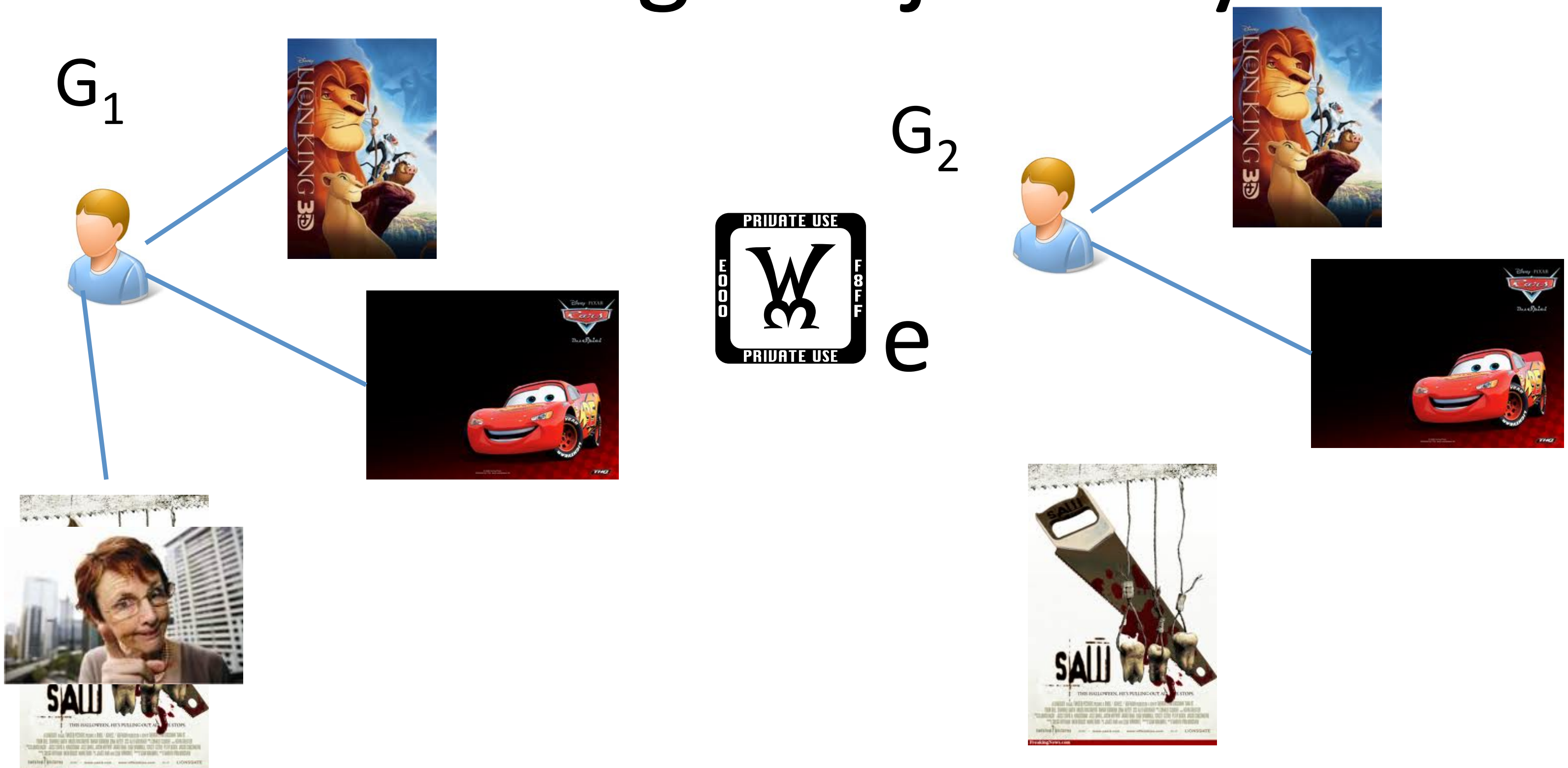
Important Question: When are G and G' are neighbors?



Edge Adjacency

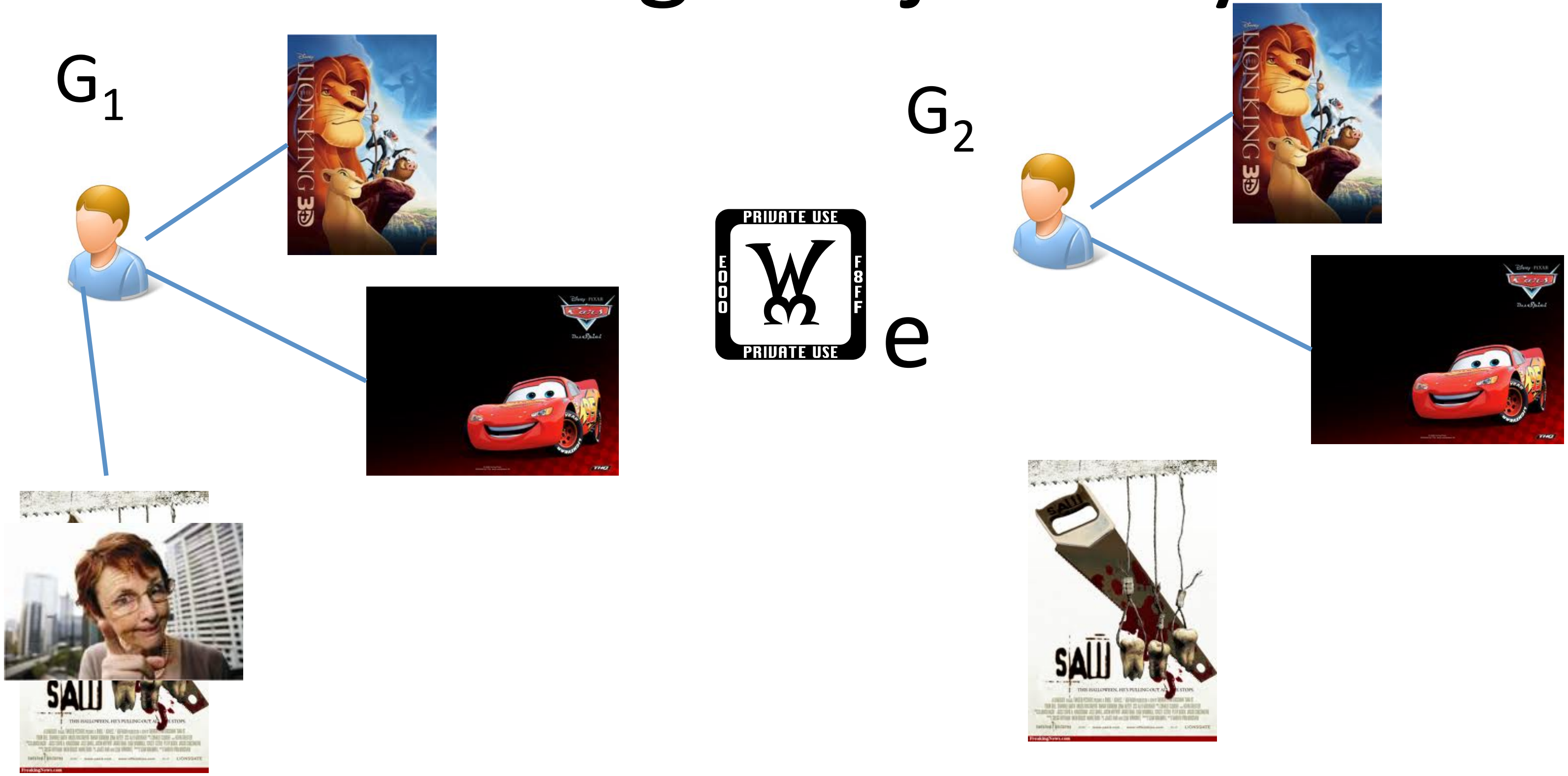


Edge Adjacency



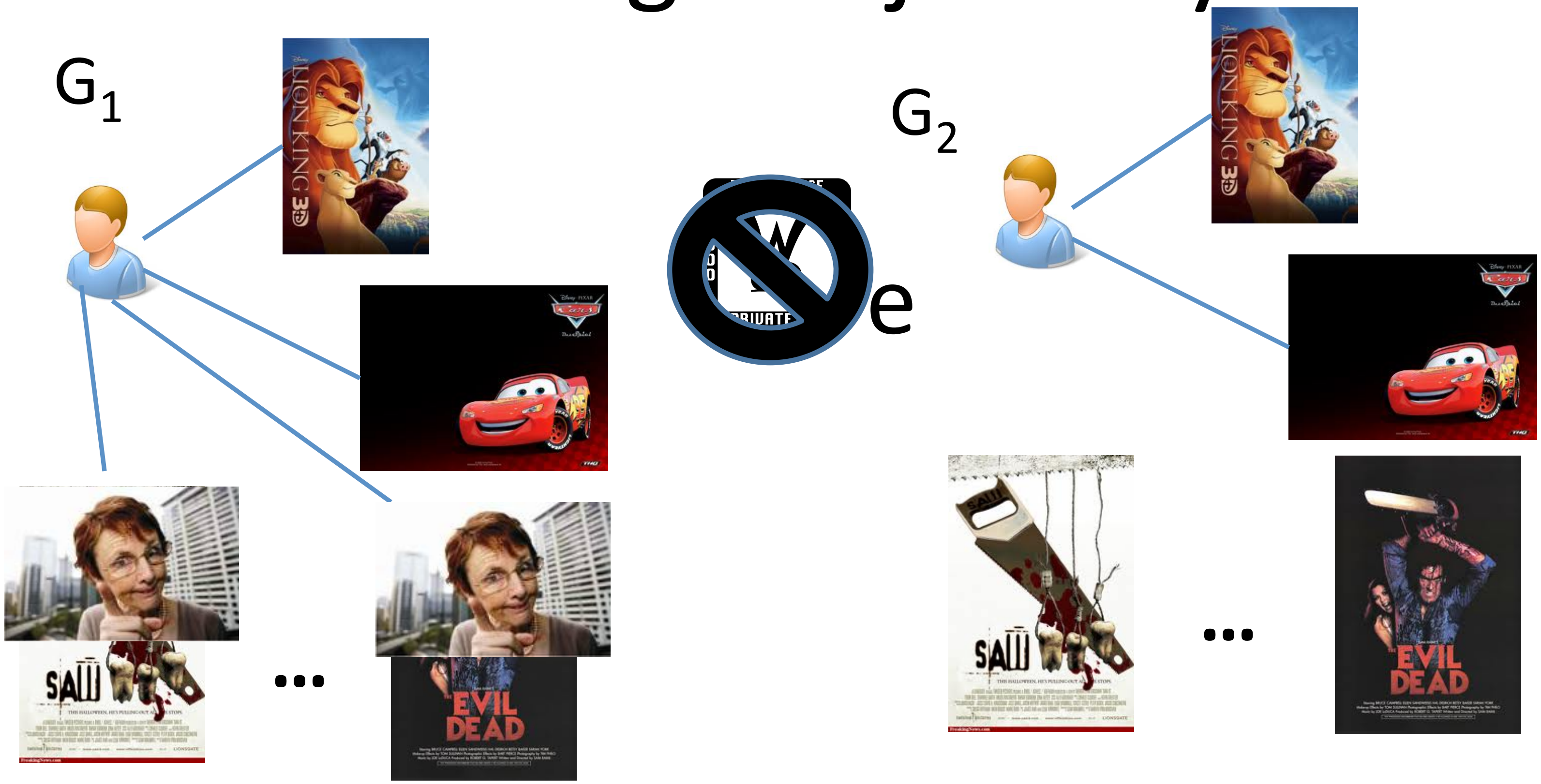
Pr[A(G1)= ] ≤e[PRIVATE USE]Pr[A(G2)= ] + [PRIVATE USE]

Edge Adjacency



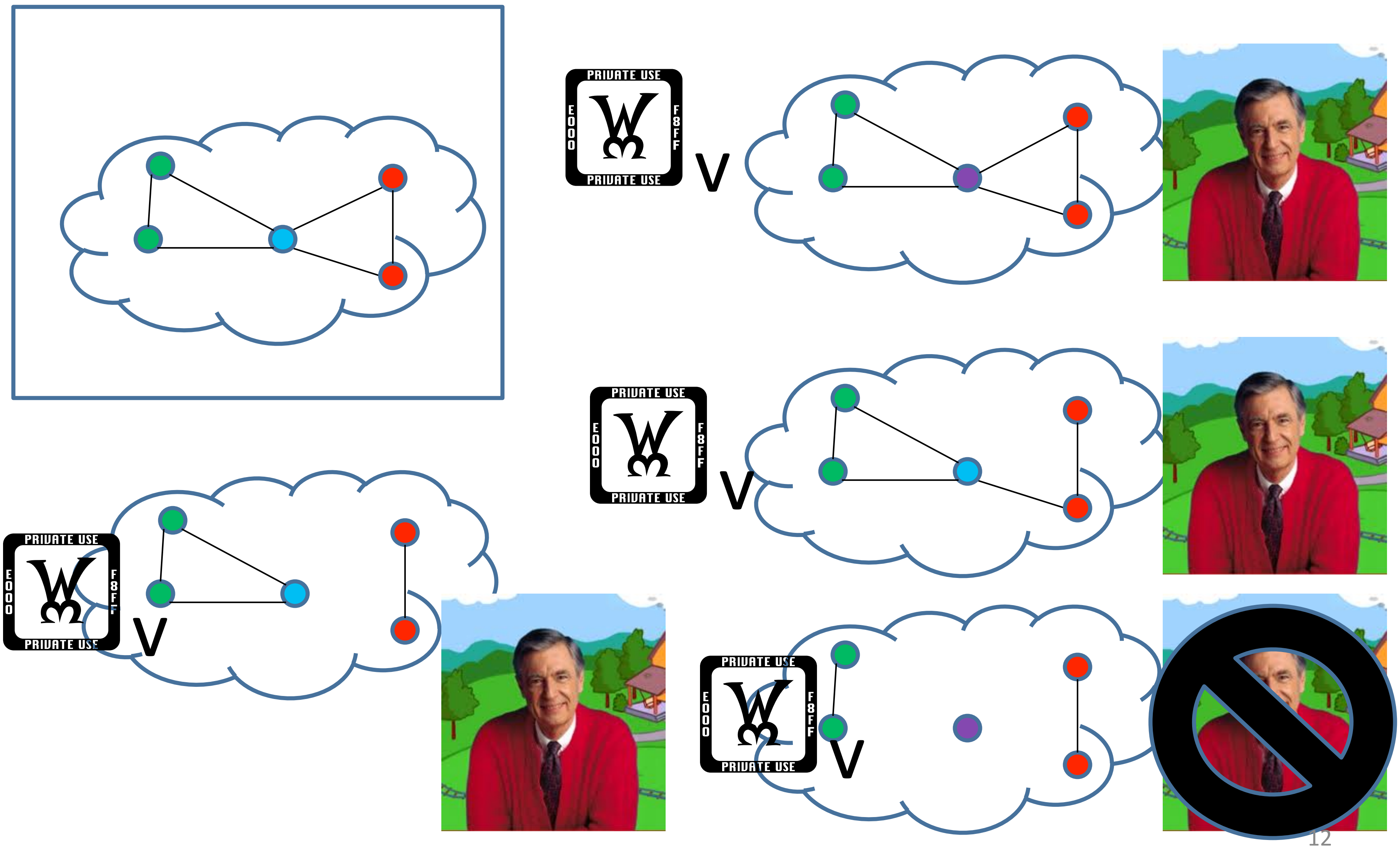
Johnny’s mom cannot tell if he watched Saw.

Edge Adjacency

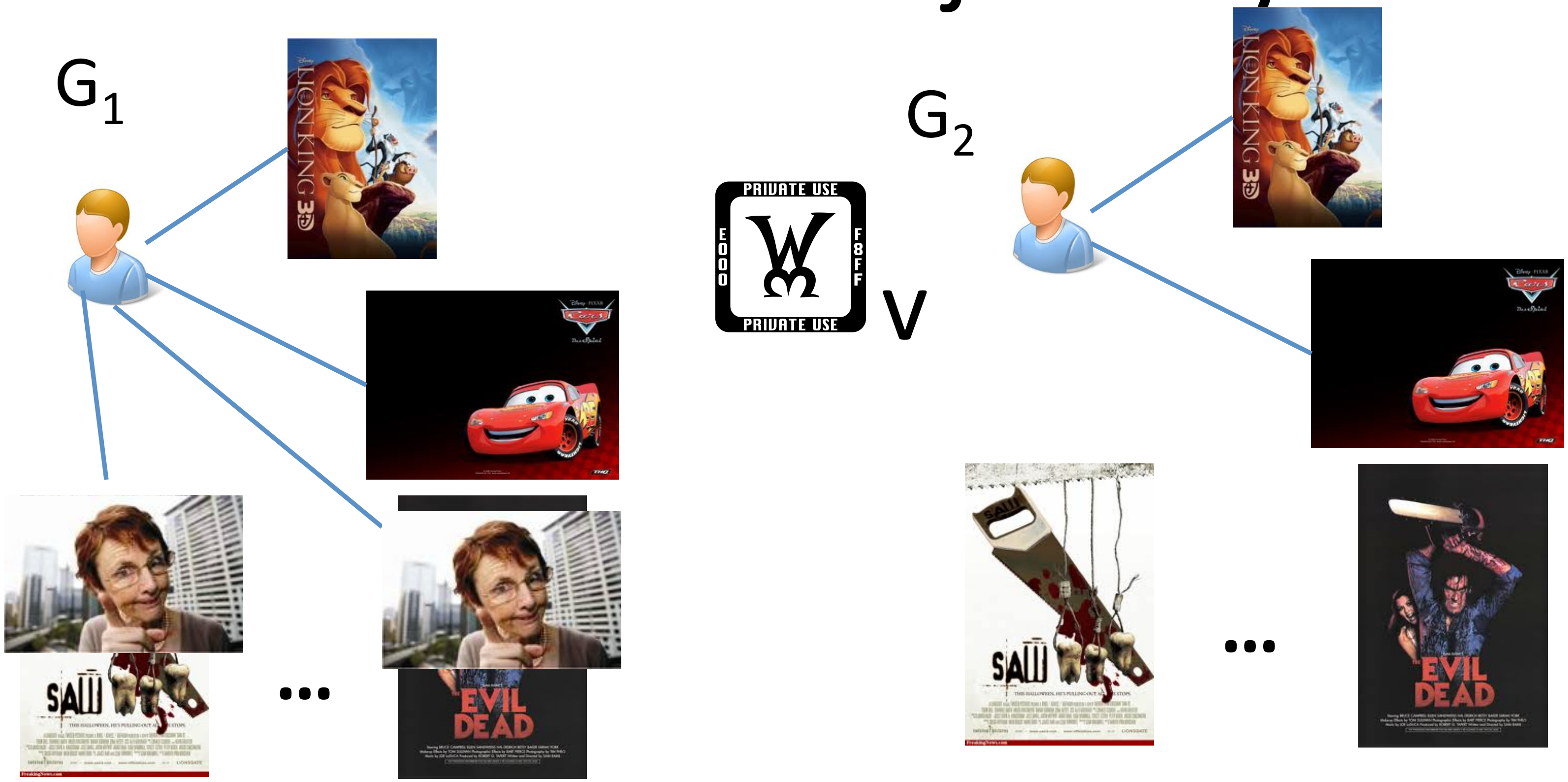


Johnny's mom may be able tell if he watches R-rated movies.

Vertex Adjacency

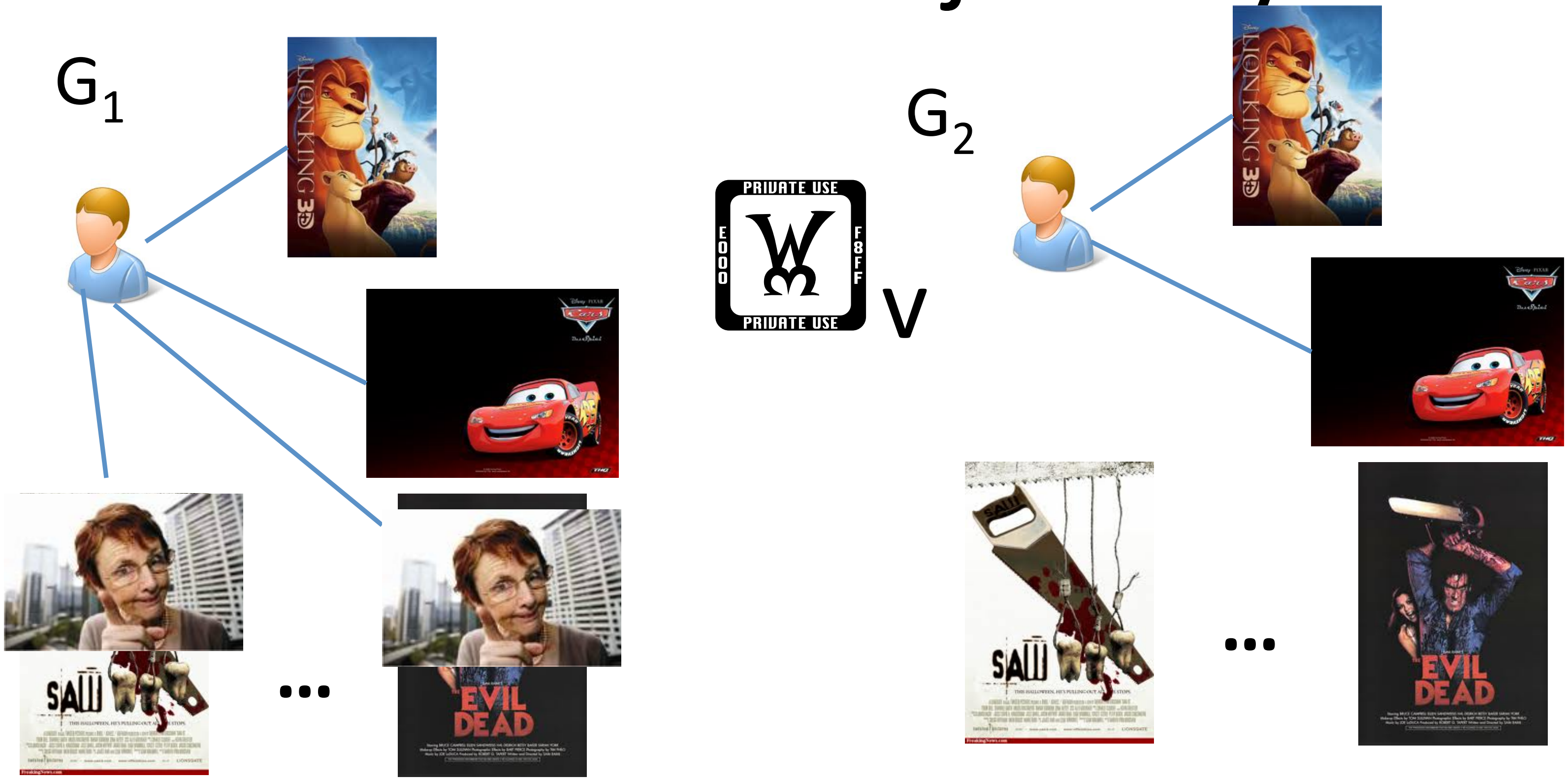


Vertex Adjacency



$\Pr[A(G_1)=$  $] \leq e^{[illegible]} \Pr[A(G_2)=$  $] + [illegible]$

Vertex Adjacency



Johnny’s mom cannot tell if he regularly watches R-rated movies.

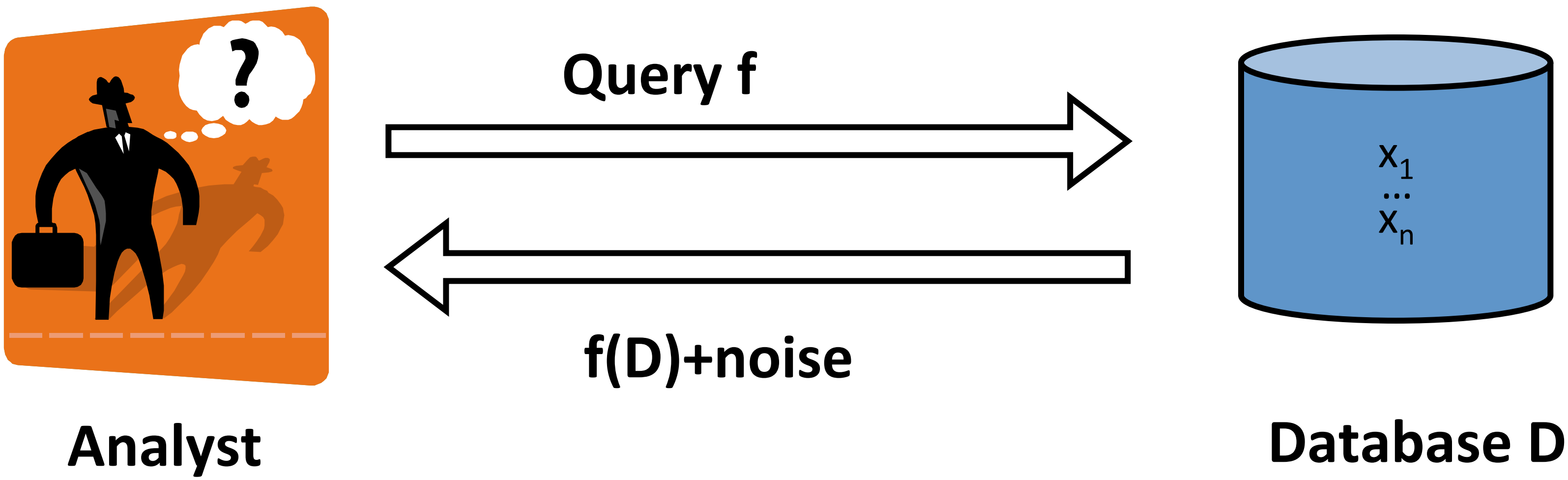
Differential Privacy (Dwork et al)

An algorithm A satisfies (ϵ, δ)-differential privacy for social networks if for any S $\subseteq$ Range(A)

$$\Pr[A(G) \in S] \leq e^{\epsilon} \Pr[A(G') \in S] + \delta$$

for every pair of *vertex adjacent* social networks G and G’

Usual Differential Privacy Setting

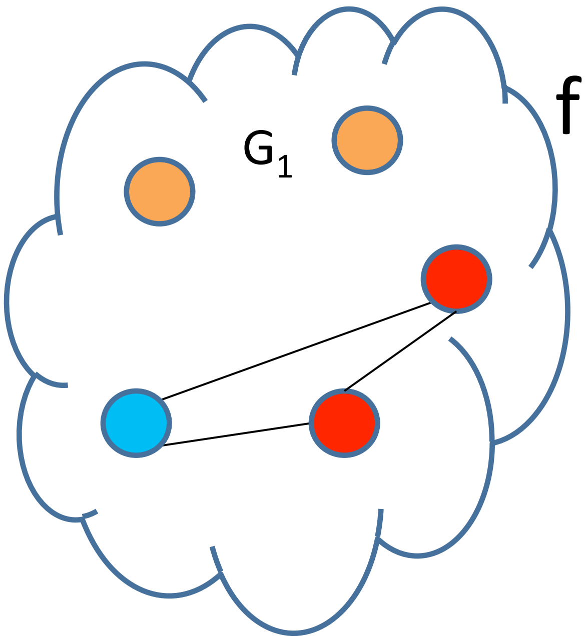


Us al:

- cur for al
- erentia cy

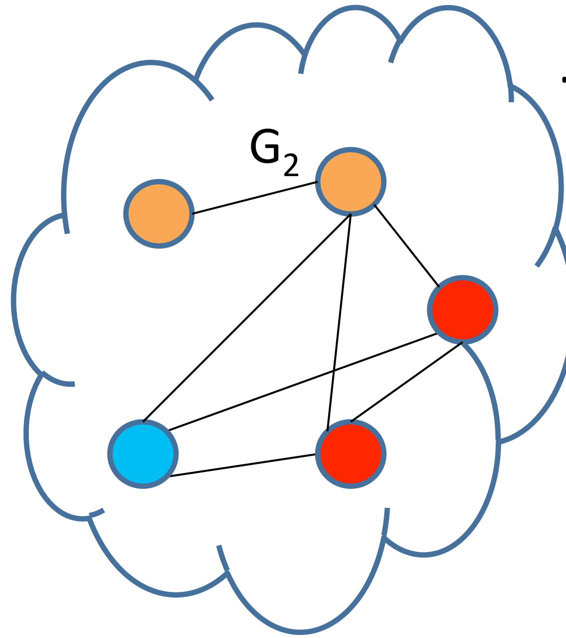
Challenge: High Sensitivity

f(G) = “how many people in G know a pianist?”



$f(G_1)=0$

v



$f(G_2)=n$

Previous Work

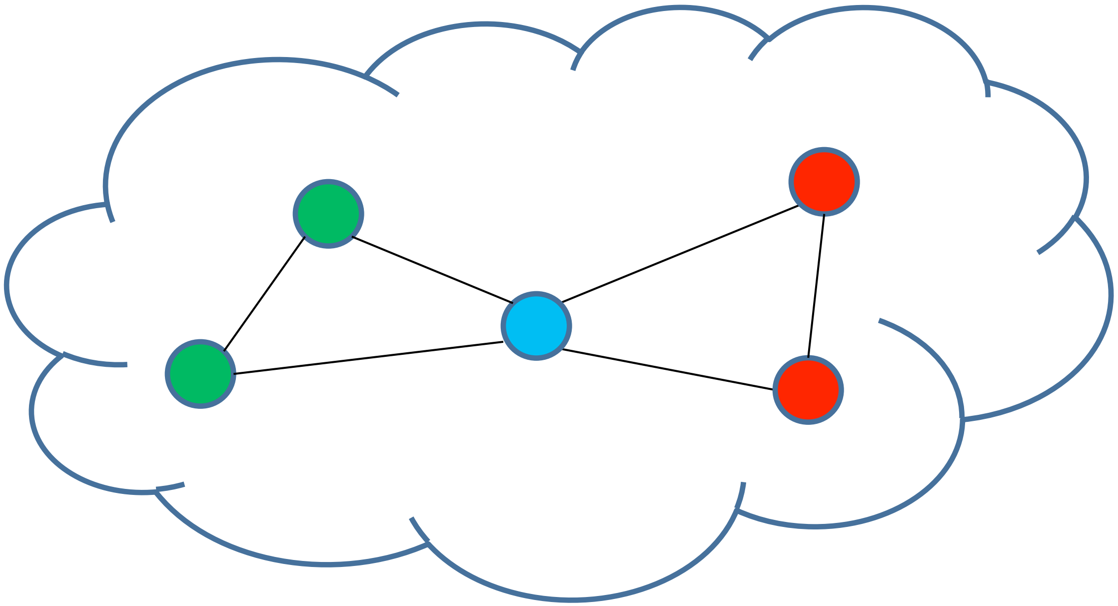
- Edge Adjacency Model
  - Degree Distribution [HLM'09]
  - Subgraph Counting [KRSY'11]
  - Cut Queries [GRU'12, BBDS'12]

- This Work: Vertex Adjacency
  - Node-level differential privacy [KNRS'12]
    (Concurrent independent results)

Outline

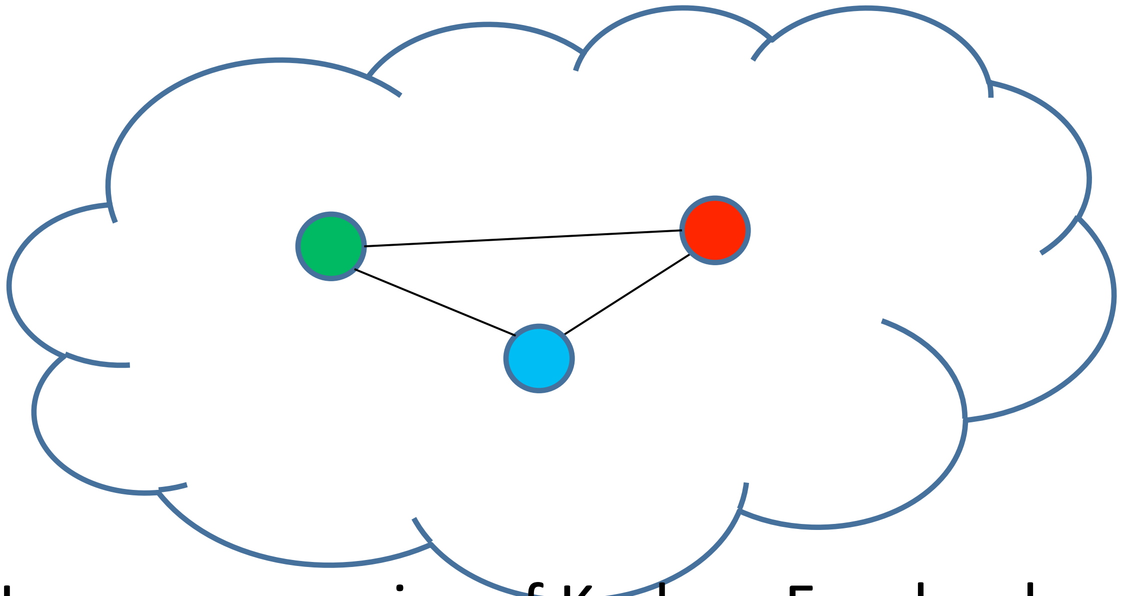
- Background
- **The Problem**
  - Interesting Queries
  - Sensitivity of a Query
  - Interesting Queries have High Sensitivity
- Restricted Sensitivity
- Algorithms

Local Profile Query



How many people know 2 lawyers who know each other and 2 doctors who know each other, but the lawyers aren't friends with the doctors?

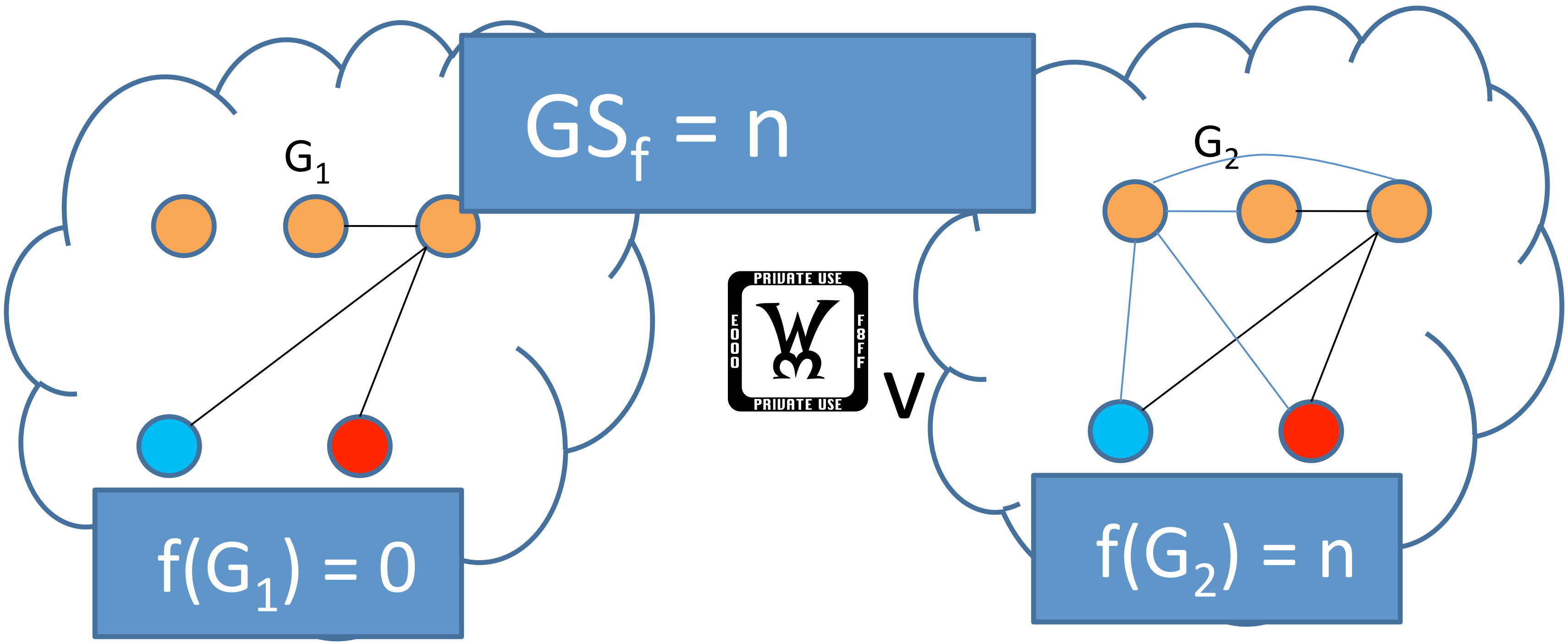
Subgraph Counting Queries



How many copies of K_3 does Facebook contain where one node is a doctor, one node is a professor and one node is a lawyer?

Global Sensitivity

f(G) = "how many people in G know two pianists?"



Global Sensitivity

Global Sensitivity of f:

$$GS_f = \max_{G_1, G_2} \left(\frac{\left| f(G_1) - f(G_2) \right|}{d(G_1, G_2)} \right)$$

Local Profile Queries (f):

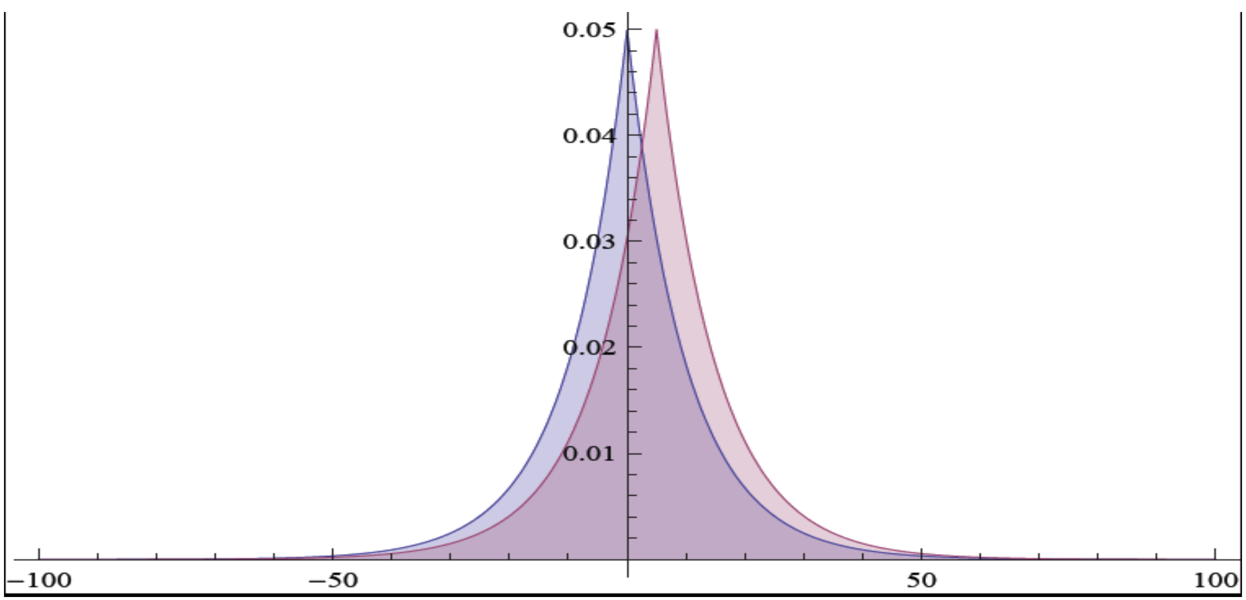
$$GS_f = n$$

Laplacian Mechanism

The mechanism

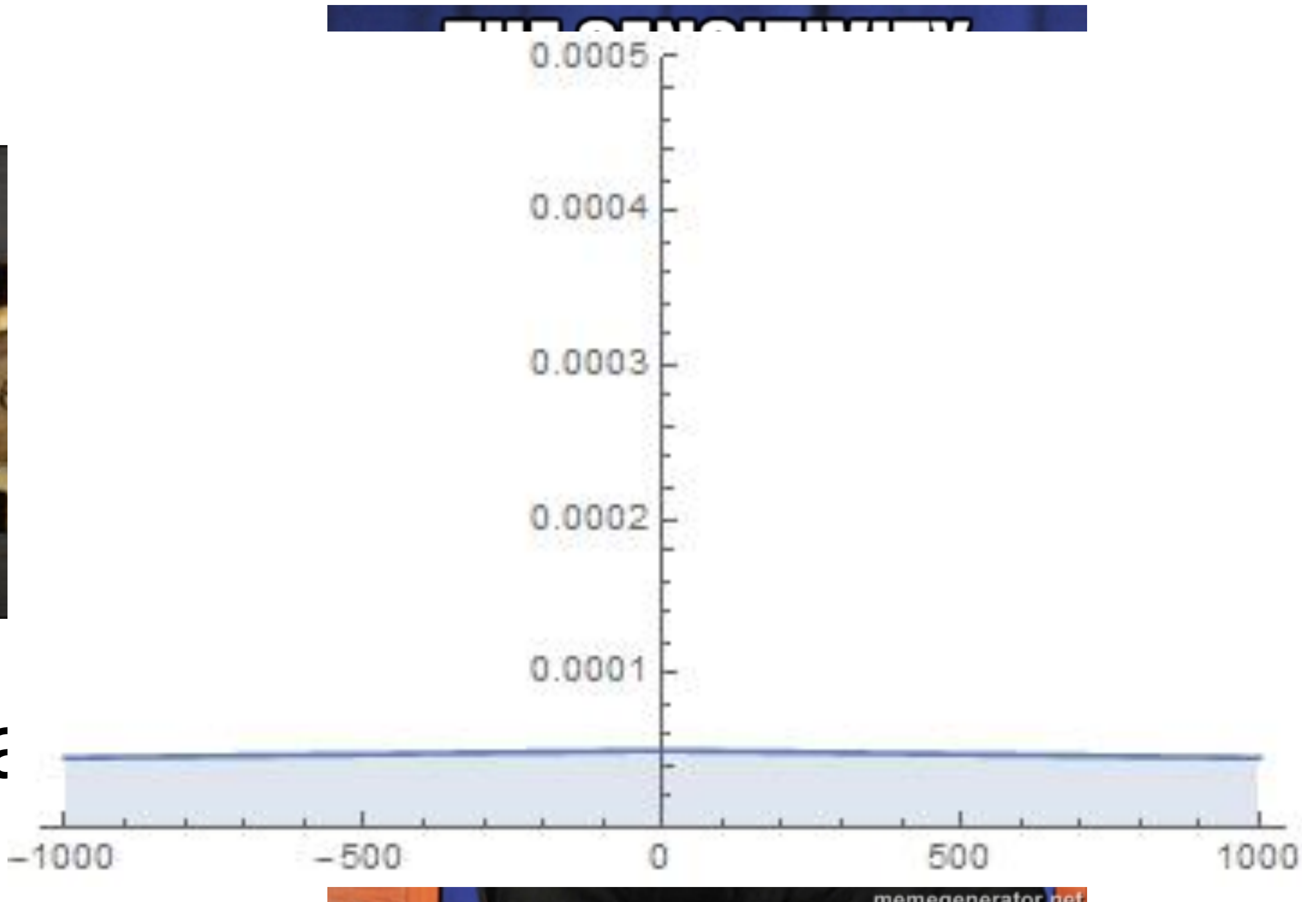
$$A(G) = f(G) + Lap(GS_f/\epsilon),$$

satisfies (ϵ,0)-differential privacy.



Privacy vs. Accuracy

$$A(G) = f(G) + \mathrm{Lap}(n/\epsilon)$$

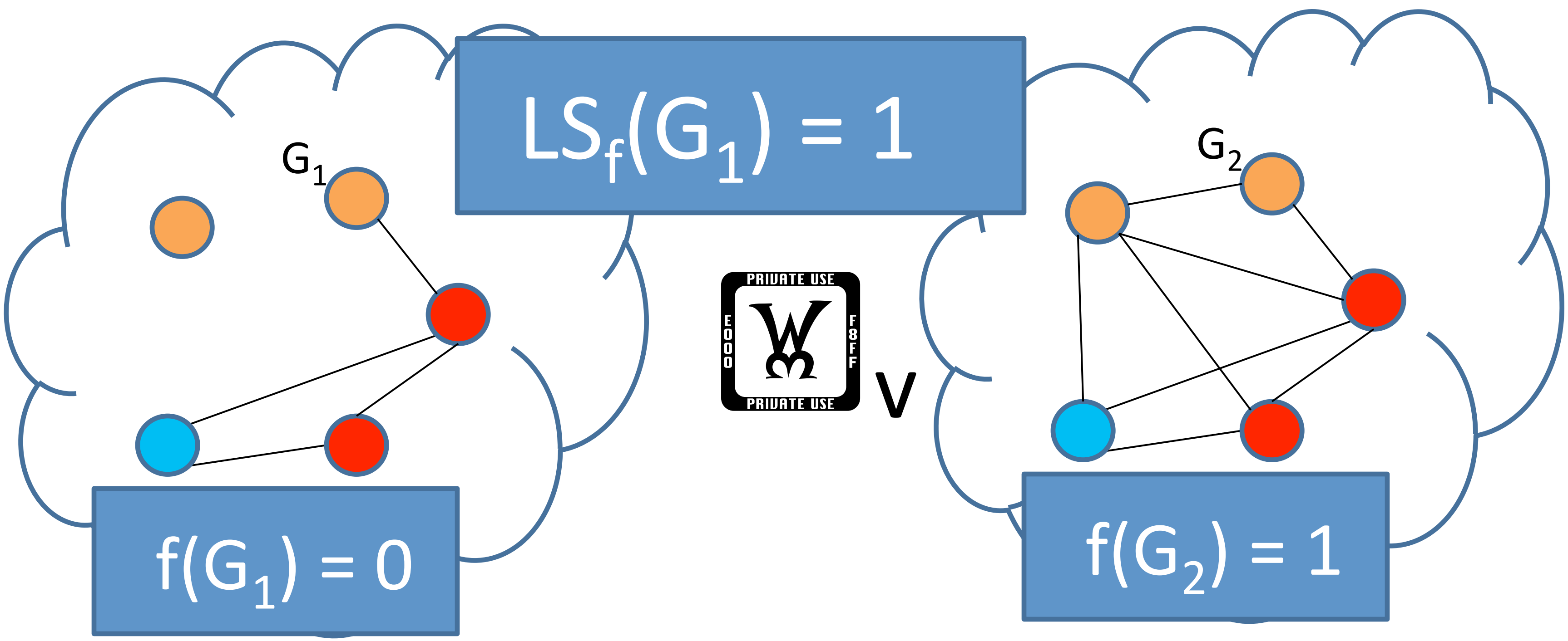


Priva

rate!

Local Sensitivity

f(G) = “how many people in G know two pianists?”



Local Sensitivity

Local Sensitivity of f at G:

$$LS_f(G) = \max_{G\sim G'} |f(G)-f(G')|$$

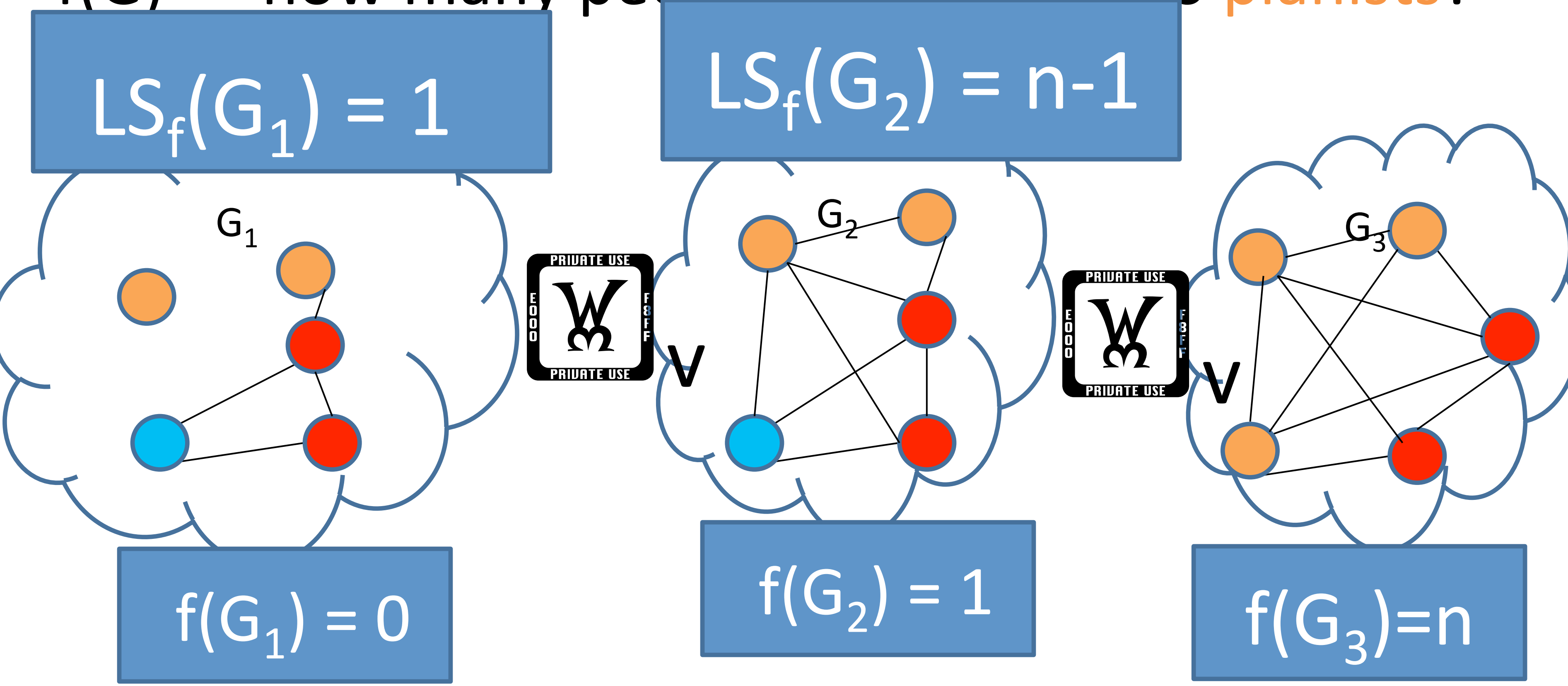
The mechanism

$$A(G) = f(G) + Lap(LS_f(G)/\epsilon)$$

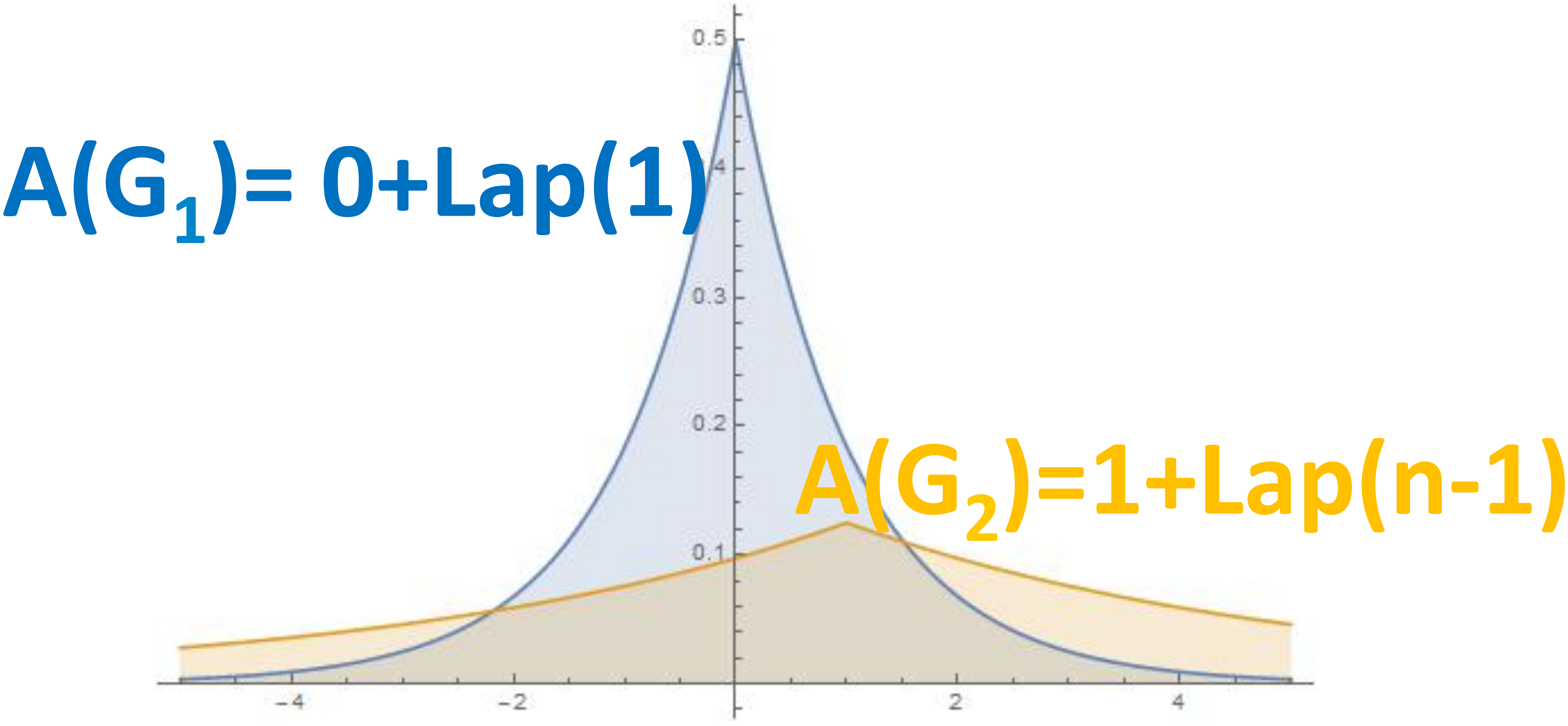
does not satisfy differential privacy.

The Problem

f(G) = “how many people in G know two pianists?”

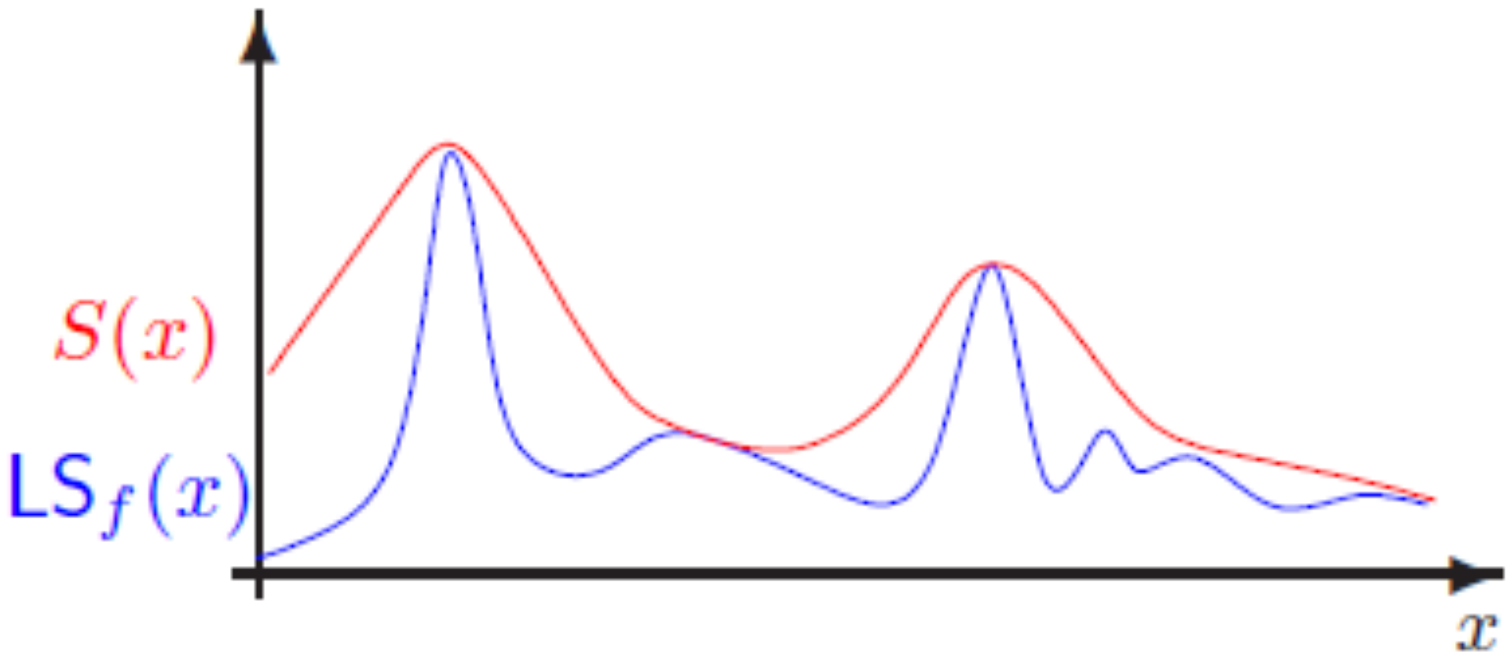


PDFs with $\varepsilon = 1$, n=5



Smooth Sensitivity (Nissim et al)

Problem: $LS_f(G)$ itself could be highly sensitive!



Smooth Sensitivity (Nissim et al)

A □-smooth upper bound on the local sensitivity $S_{f,□}$ satisfies

1. $S_{f,□}(G)$ □ $LS_f(G)$ for all G □ G.
2. $S_{f,□}(G) \leq e^{□} S_{f,□}(G')$ for all G'~G.

Theorem: The mechanism $A(G) = f(G) + Lap(2\, S_{f,□}(G) / □)$
satisfies (□, □)-differential privacy with □=-□/2 ln □.

Smooth Sensitivity

A β-smooth upper bound on the local sensitivity $S_{f,\beta}$ satisfies

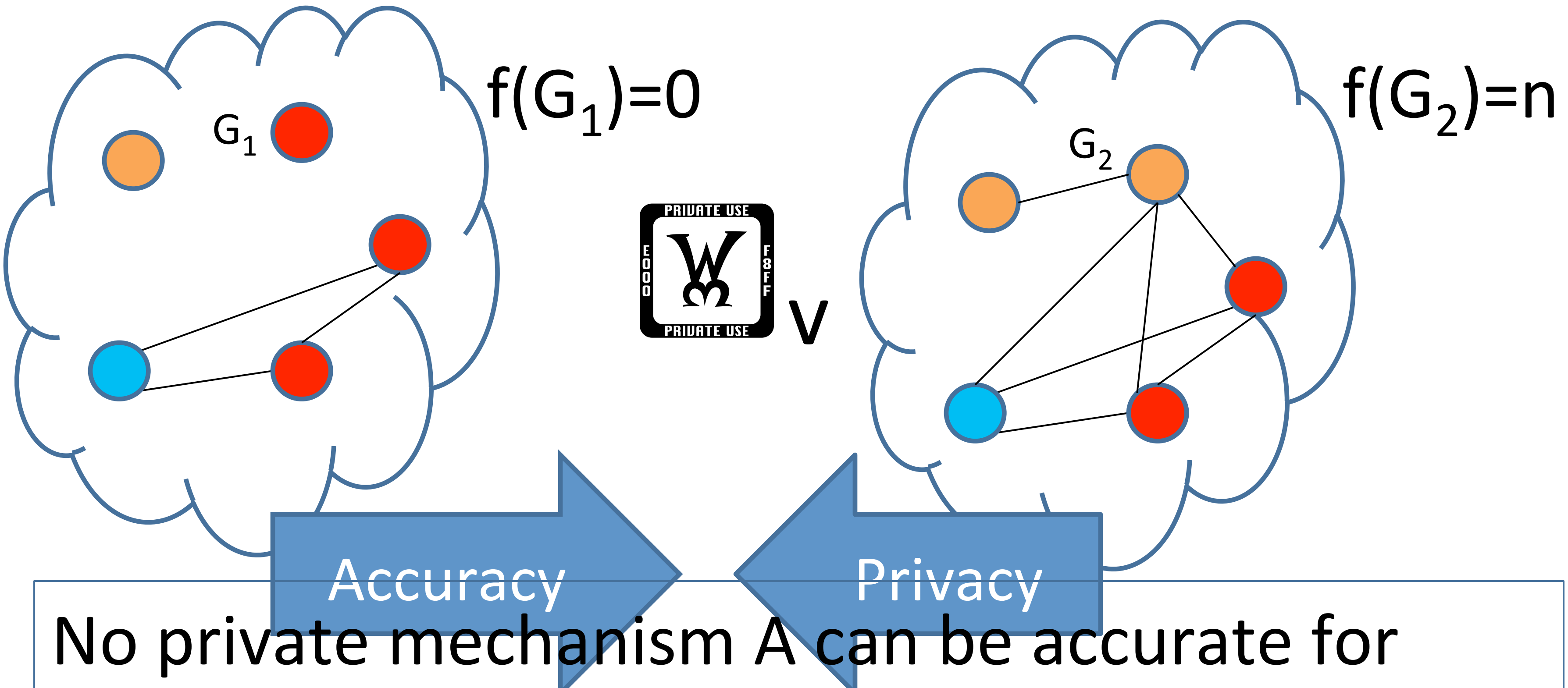
1. $S_{f,\beta}(G) \geq LS_f(G)$ for all $G \in \mathcal{G}$.
2. $S_{f,\beta}(G) \leq e^{\beta} S_{f,\beta}(G')$ for all $G' \sim G$.

When is $S_{f,\beta}(G)$ small?

1. $LS_f(G)$ must be small.
2. For any nearby graph G’, $LS_f(G')$ must also be small.

Even the Smooth Sensitivity is High!

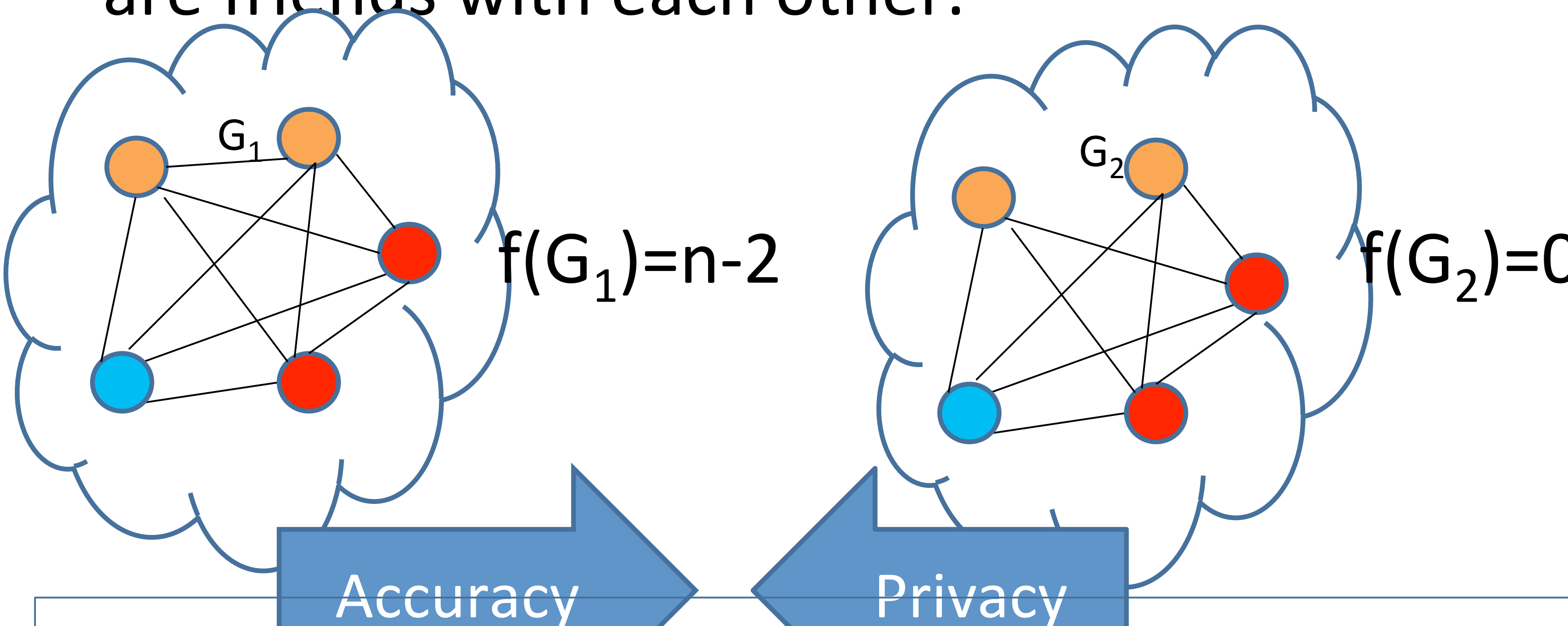
f(G) = "how many people in G know a pianist?"



No private mechanism A can be accurate for both G_1 and G_2!

Hopeless?

f(G) = "how many people know two pianists who are friends with each other."

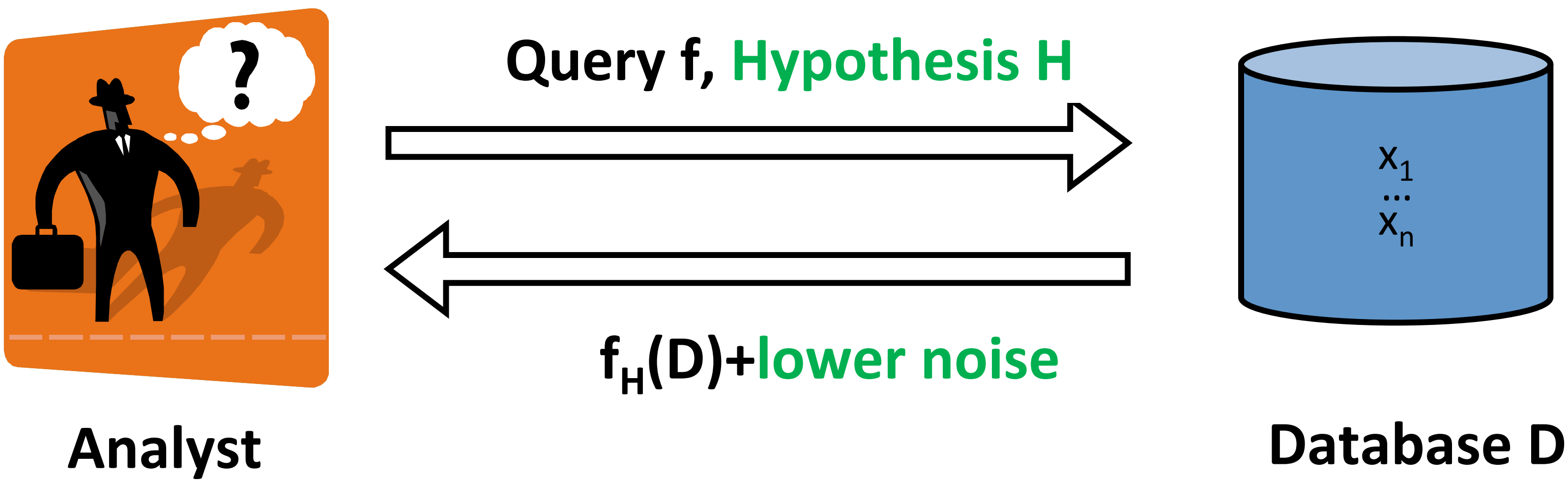


No private mechanism A can be accurate for both G_1 and G_2!

Outline

- Background
- The Problem
- **Restricted Sensitivity**
  - Lower Sensitivity for Interesting Queries
  - Challenges in Designing Mechanisms
  - Relaxed Accuracy Goal
- Algorithms

Our Setting

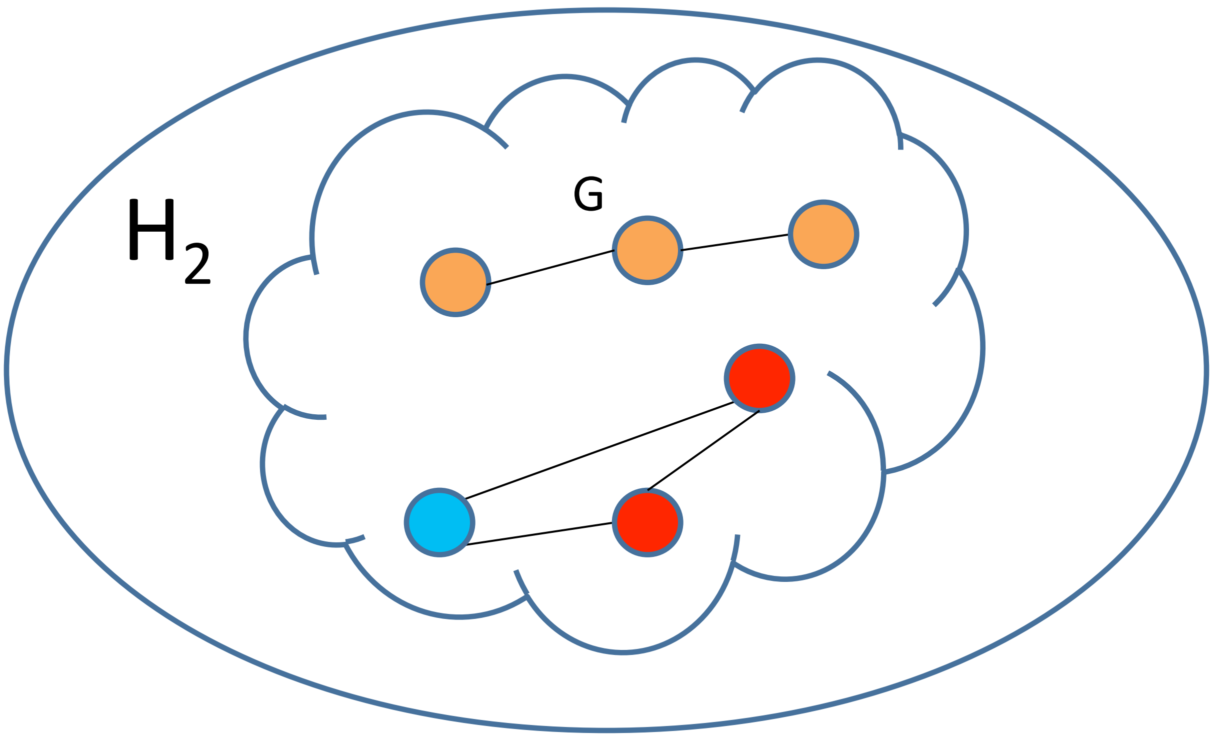


- Accurate for D in H
- Differential Privacy

Bounded Degree Hypothesis

Bounded Degree Hypothesis:

$H_k = \{ G \mid \max_{v \in V(G)} \deg(v) \leq k \}$

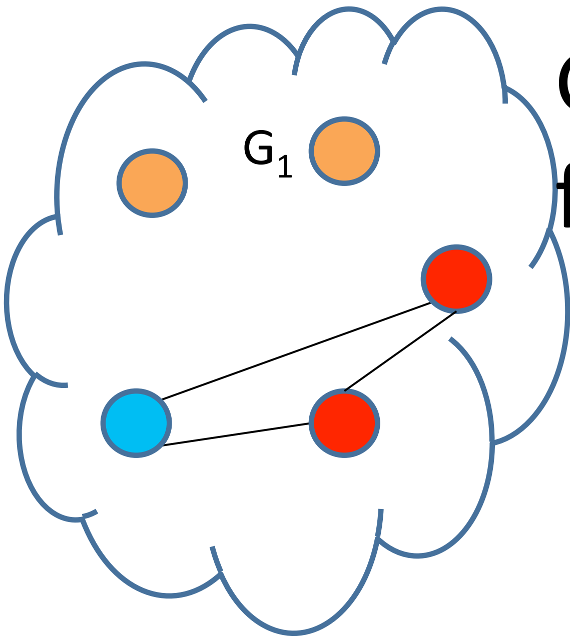


Typical:

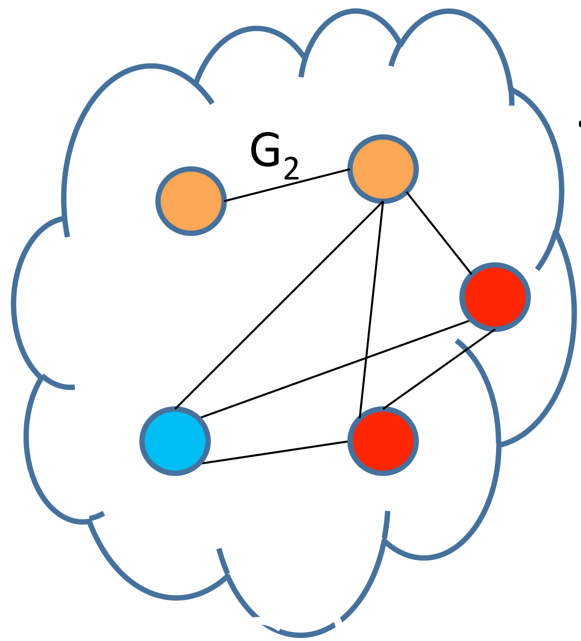
$k << n$

Challenge Revisited

f(G) = "how many people in G know a pianist?"



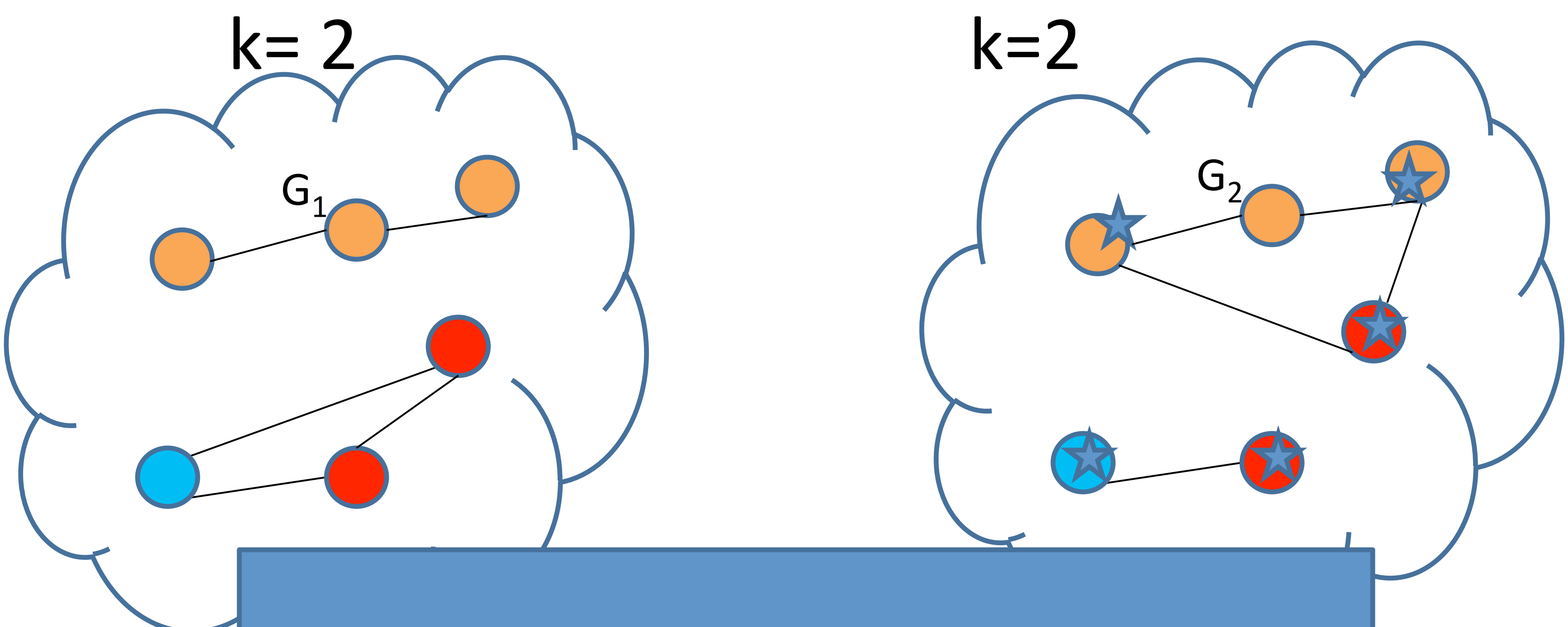
G_1 □ H_k.
$f(G_1) = 0$



G_2 □ H_k.
$f(G_2)=n$

Restricted Sensitivity $RS_f(H_k)$

Fact: For local profile queries f, $RS_f(H_k) \leq 2k+1$



$|f(G_1) - f(G_2)| \leq 5 = RS_f(H_2)$

Restricted Sensitivity

Hypothesis: H [illegible] G

$$GS_f = \max_{G_1, G_2} \frac{\left| f(G_1) - f(G_2) \right|}{d(G_1, G_2)}$$

$$RS_f(H) = \max_{G_1, G_2 \in H} \frac{\left| f(G_1) - f(G_2) \right|}{d(G_1, G_2)}$$

Sensitivity over H_k

	Local Profile Query		Subgraph Counting Query (P)	
Adjacency	**Smooth**	**Restricted**	**Smooth**	**Restricted**
Edge	k+1	k+1	$O(\|P\| k^{\|P\|-1})$	$O(\|P\| k^{\|P\|-1})$
Vertex	n	2k+1	$O(n^{\|P\|-1})$	$O(\|P\| k^{\|P\|-1})$

Two "Strawman"-Algorithms

A(G) = f(G)+	$Lap(GS_f/$[illegible]$)$	$Lap(RS_f(H_k)/$[illegible]$)$
Accurate on H_k?	NO	YES
Private?	YES	NO

Restricted Sensitivity

For $k \ll n$ the mechanism

$$A(G) = f(G) + Lap(RS_f(H_k)/\epsilon)$$

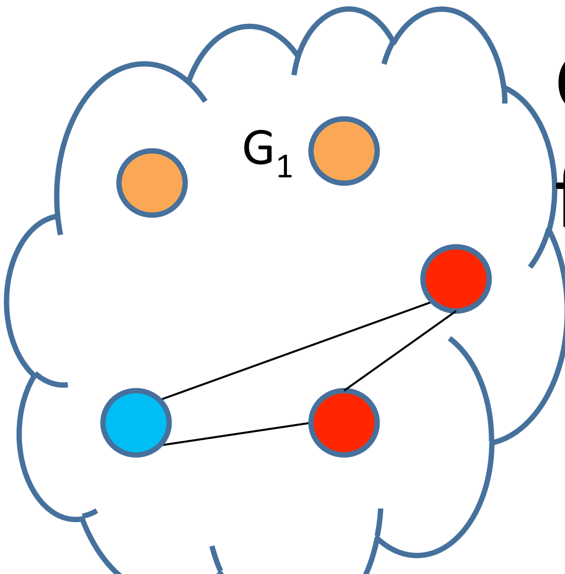
is accurate!

The mechanism

$$A(G) = f(G) + Lap(RS_f(H)/\epsilon)$$

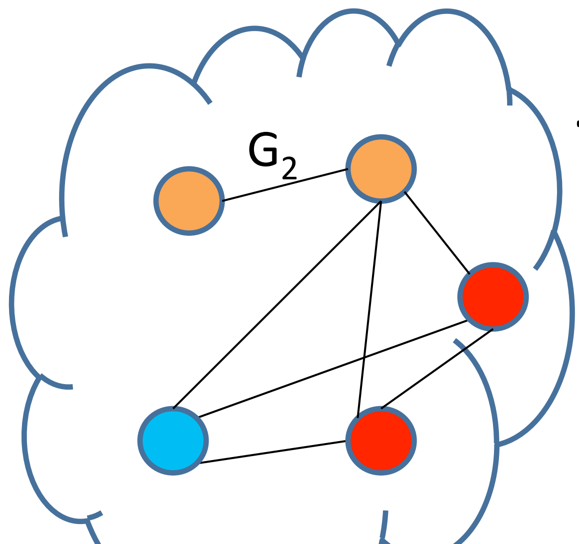
does not satisfy differential privacy for all G.

f(G) = "how many people in G know a pianist?"



G_1 [illegible] H_k.

$f(G_1) = 0$

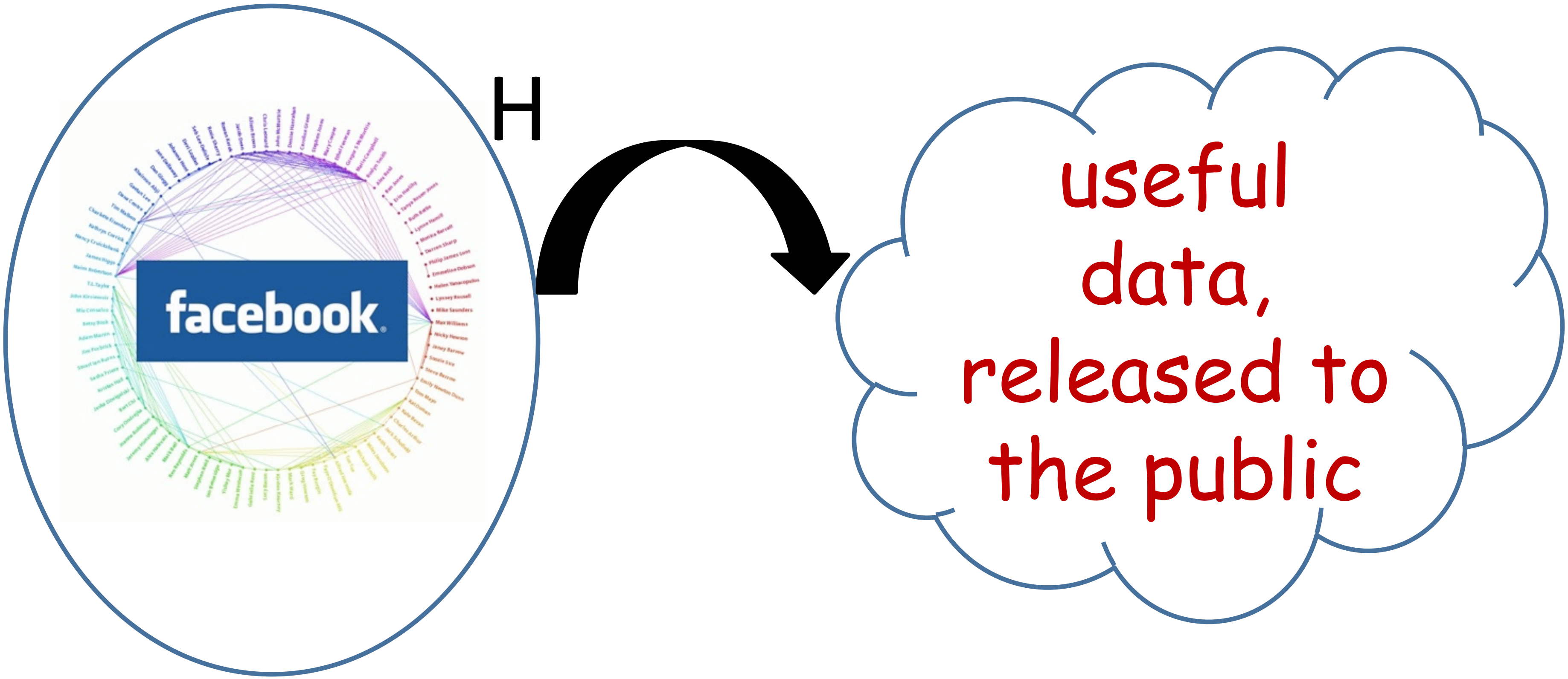


G_2 [illegible] H_k.

$f(G_2)=n$

Pr[A(G_1) [illegible] 0] ≤ e[illegible] Pr[A(G_2) [illegible] 0] +

Privacy for All, Accuracy for Some

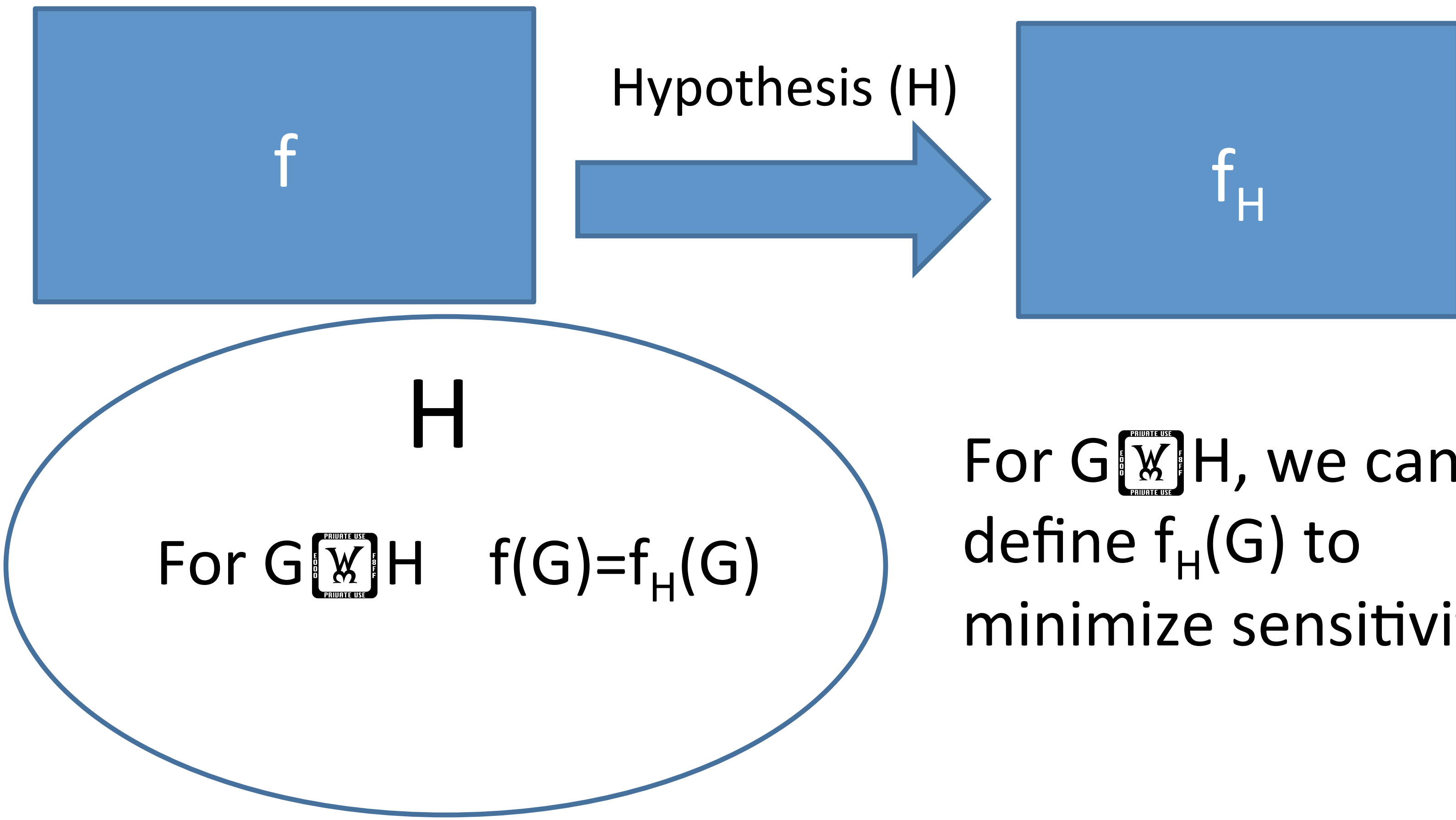


1. Privacy for all G
2. Accurate statistics for G [illegible] H

Outline

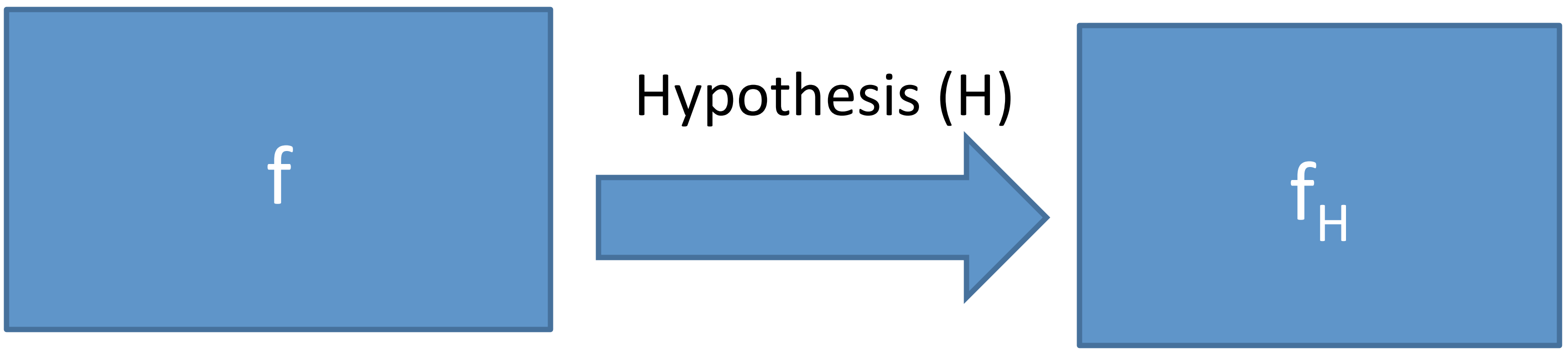
- Background
- The Problem
- Restricted Sensitivity
- **Algorithms**
  - **General Template**
  - Possibility: A General Inefficient Algorithm
  - Efficient Algorithms for H_k via Projections

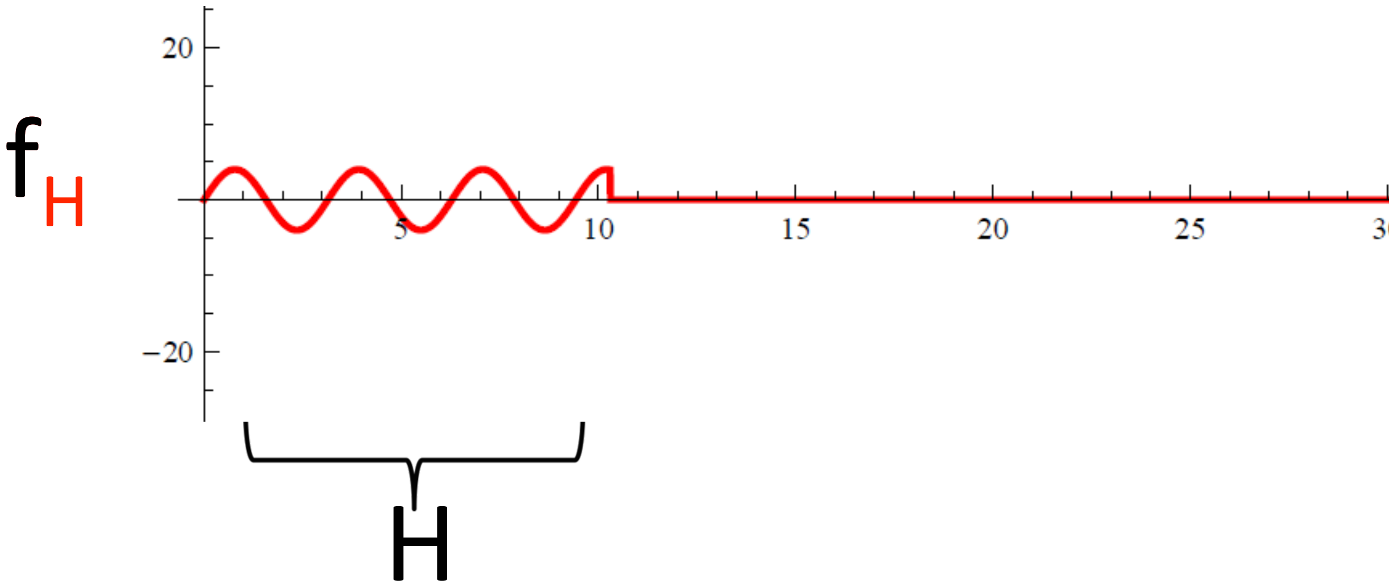
Accuracy for Some



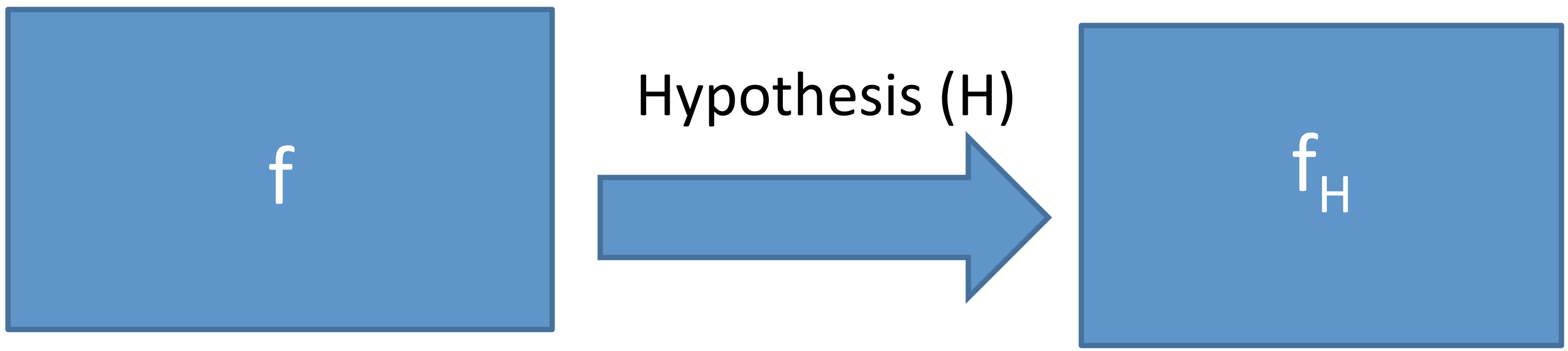
For $G \in H$, we can define $f_H(G)$ to minimize sensitivity.

Accuracy for Some





Privacy For All



Answer f_H in a differentially private manner.

Ideally, we would like to compute f_H efficiently.

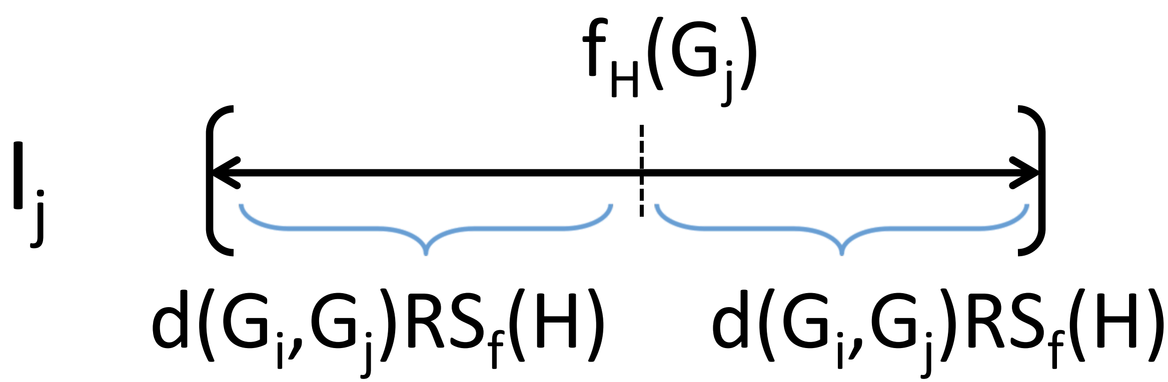
Outline

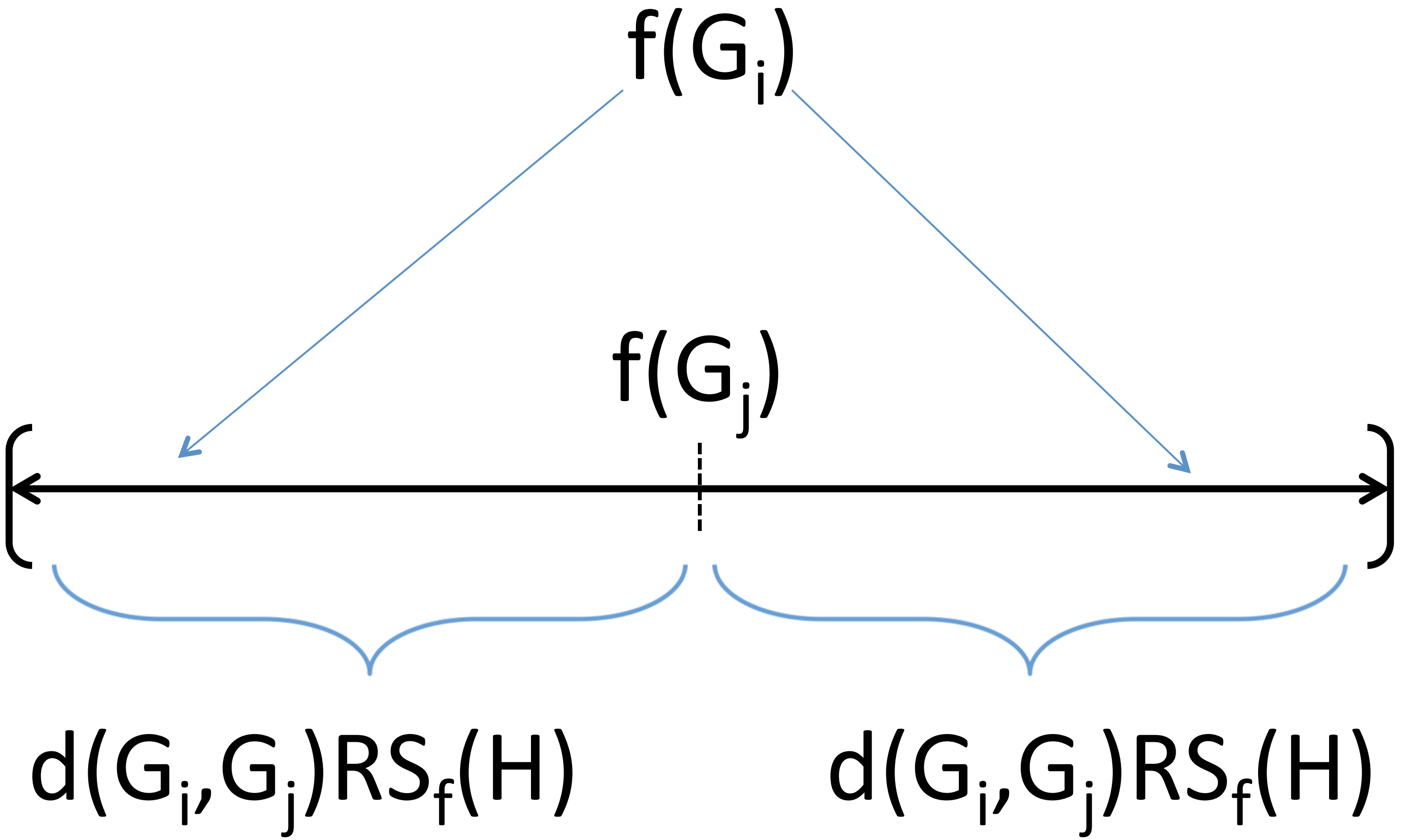
- Background
- The Problem
- Restricted Sensitivity
- **Algorithms**
  - General Template
  - **Possibility: A General Inefficient Algorithm**
  - Efficient Algorithm for H_k via Projections

General Construction of f_H

Consider a canonical ordering $G_1, G_2, \ldots$ over all social networks($\mathcal{G}$).

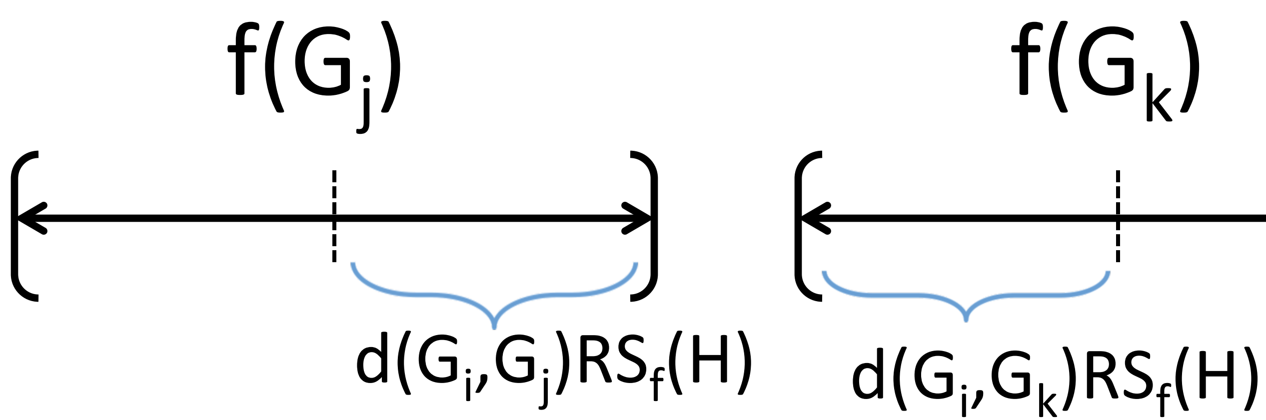
- For $G_i \in H$ set $f_H(G_i) = f(G_i)$
- Find $f_H(G_i) \in \bigcap_{j<i} I_j$




f(G_i)
f(G_j)
d(G_i,G_j)RS_f(H)
d(G_i,G_j)RS_f(H)

Suppose for contradiction that no value exists....

Can find two nonintersecting intervals I_j and I_k ($j,k < i$).



$$\frac{\left|f\left(G_j\right)-f\left(G_k\right)\right|}{d\left(G_k,G_i\right)+d\left(G_j,G_i\right)} > RS_f\left(H\right)$$

Can Find Non Intersecting Intervals

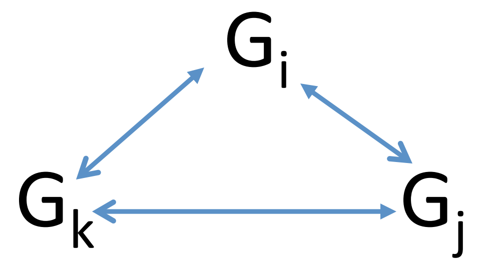
$$j = \arg\min_{t<i} \left[\max_{x \in I_t} x \right]$$

$$k = \arg\max_{t<i} \left[\min_{x \in I_t} x \right]$$

$$x \in I_j \cap I_k \Rightarrow \forall t < i, \quad x \in I_t$$

Suppose for contradiction that no value exists....

Triangle Inequality



$$\frac{\left|f\left(G_j\right)-f\left(G_k\right)\right|}{d\left(G_k,G_j\right)} \geq \frac{\left|f\left(G_j\right)-f\left(G_k\right)\right|}{d\left(G_k,G_i\right)+d\left(G_j,G_i\right)} > RS_f\left(H\right)$$

Contradiction of Inductive Hypothesis!

Suppose for contradiction that no value exists….

Then there must exist two intervals which don't intersect (j,k < i).

$$[f(G_j) - d(G_i, G_j) RS_f(H), f(G_j) + d(G_i, G_j) RS_f(H)]$$

$$[f(G_k) - d(G_i, G_k) RS_f(H), f(G_k) + d(G_i, G_k) RS_f(H)]$$

$$\frac{f(G_j) - f(G_k)}{d(G_k, G_i) + d(G_j, G_i)} > RS_f(H)$$

Suppose for contradiction that no value exists….

Triangle Inequality

$$\frac{f(G_j) - f(G_k)}{d(G_k, G_j)} \geq \frac{f(G_j) - f(G_k)}{d(G_k, G_i) + d(G_j, G_i)} > RS_f(H)$$

Contradiction of Inductive Hypothesis!

General Construction of f_H

Theorem: f_H satisfies

1. $f_H(G) = f(G)$ for all G [illegible] H
2. $GS_{f_H} = RS_f(H)$

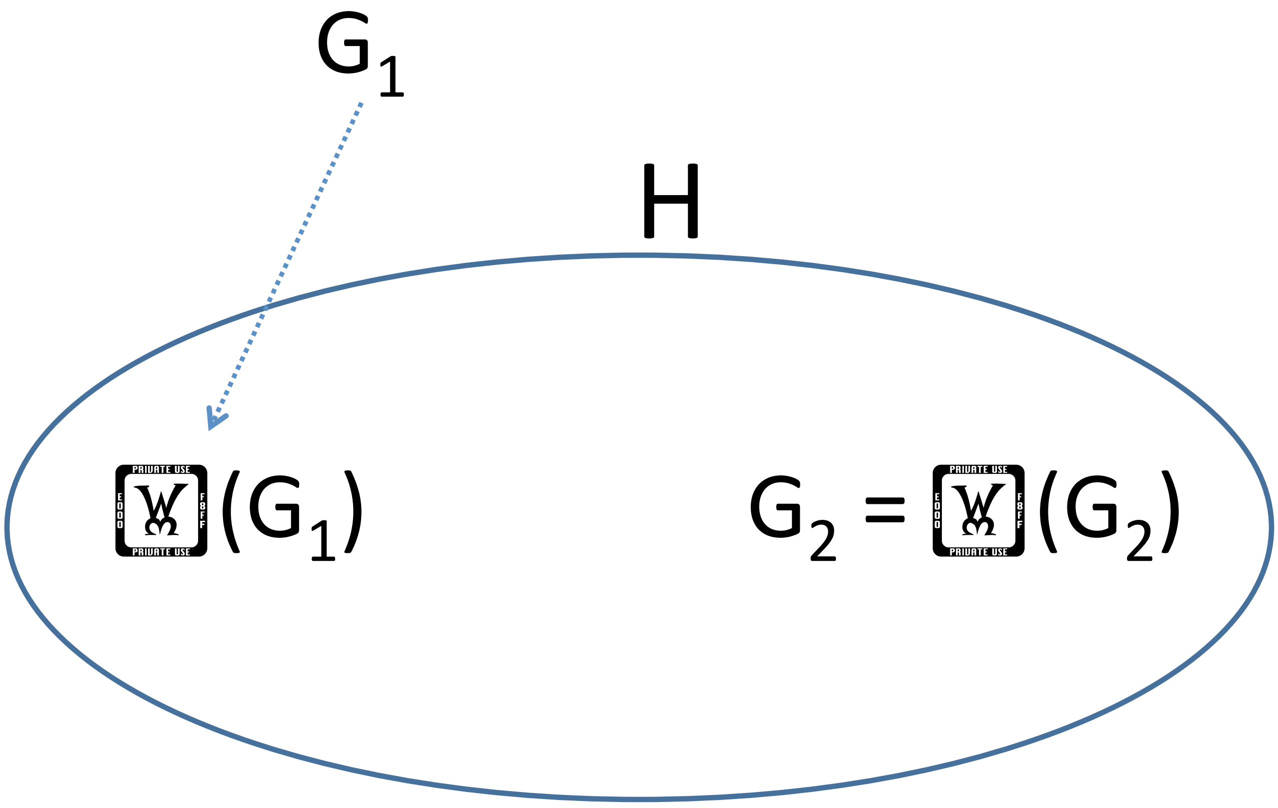
Disclaimer: The general construction of f_H is *not* efficient.

Outline

- Differential Privacy in Social Networks
- The Problem
- Restricted Sensitivity
- **Algorithms**
  – General Template
  – Possibility: A General Inefficient Algorithm
  – **Efficient Algorithms for H_k via Projections**
    - Edge Adjacency Model
    - Vertex Adjacency

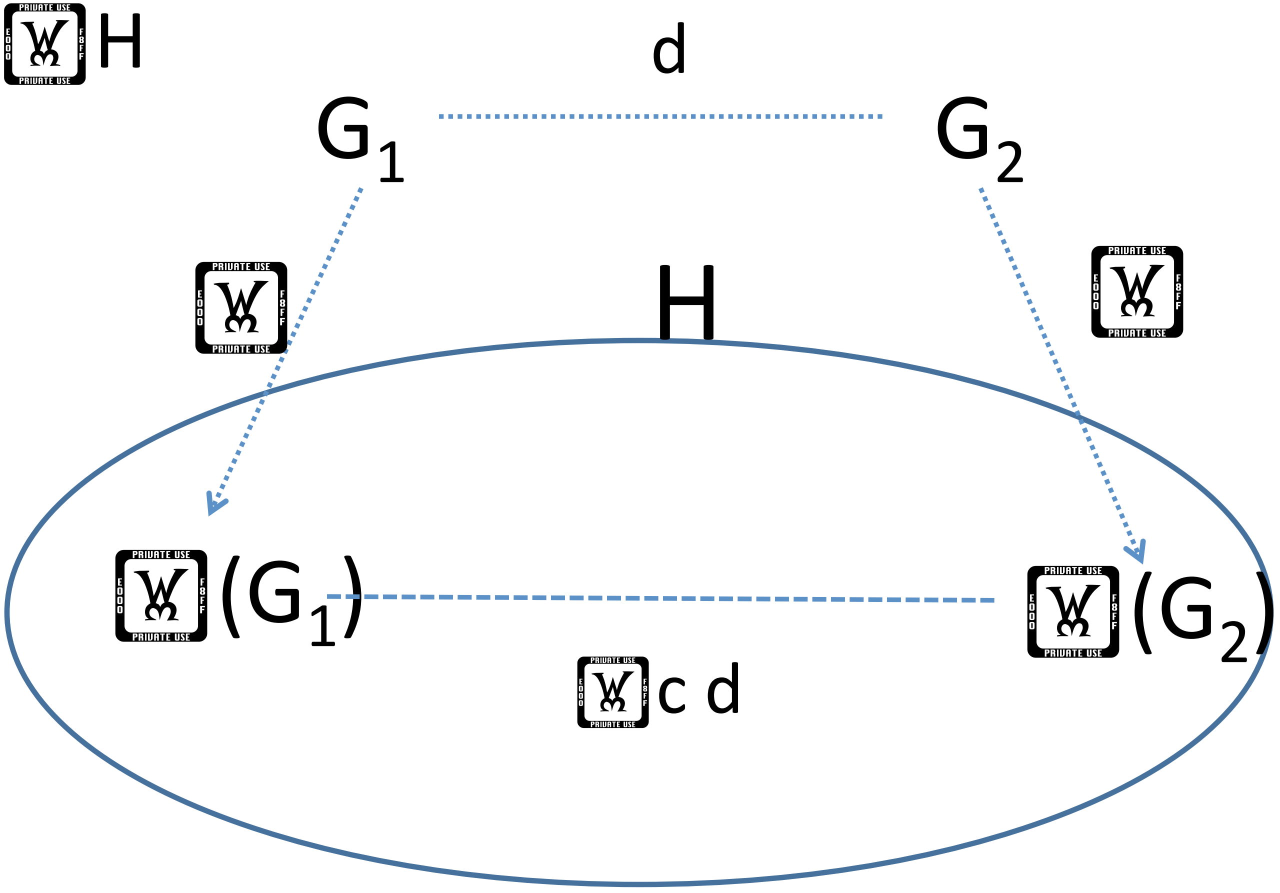
Projection onto H

: G H



c-smooth Projection

□: G □ H



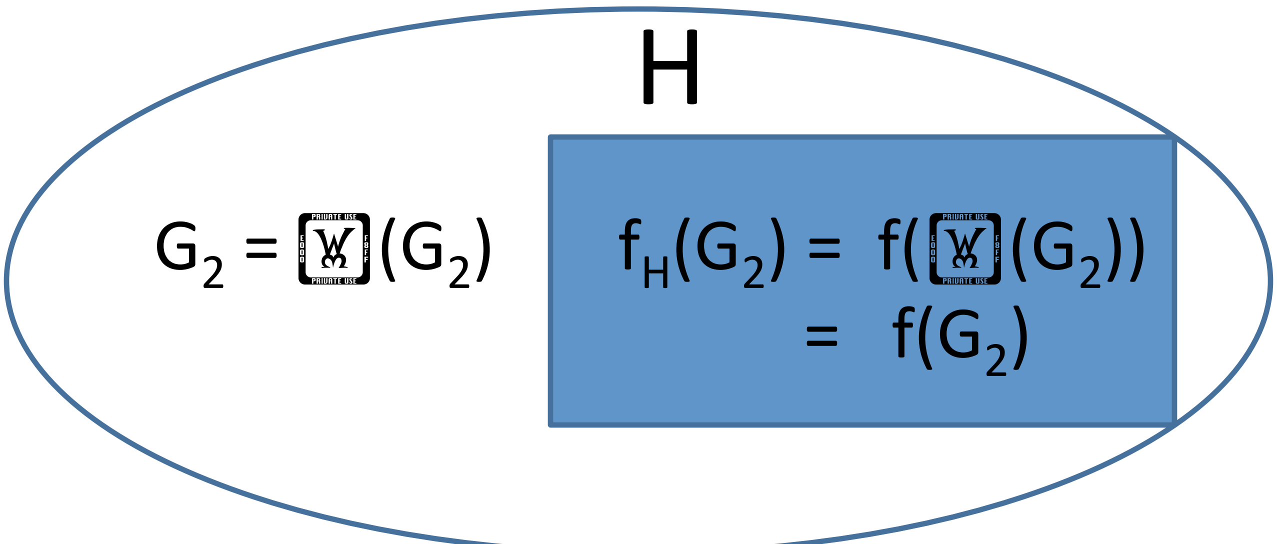
c-smooth Projection Lemma

Let : G H be a c-smooth projection then $f_H(G) = f((G))$ satisfies:

1. For G H, $f_H(G) = f(G)$, and
2. $GS_{f_H} \leq c\ RS_f(H)$.

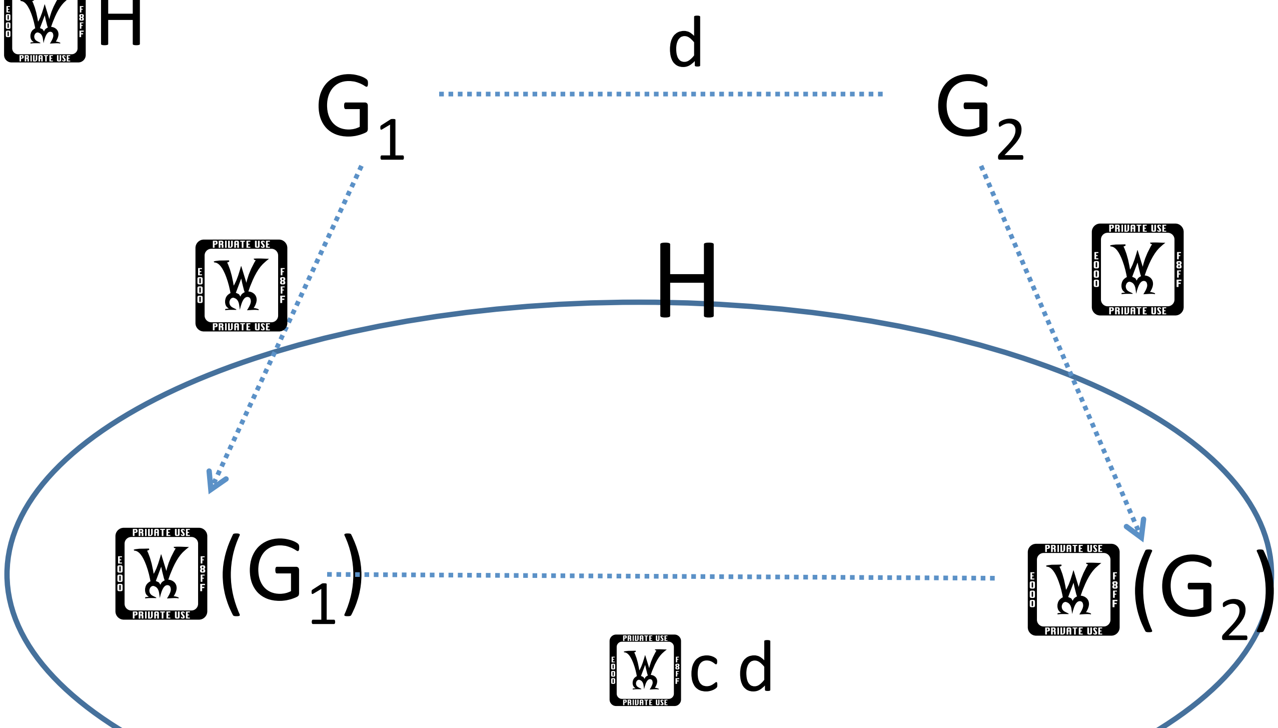
Proof of (1)

(1) For G [illegible] H, $f_H(G) = f(G)$,



Proof of (2)

: G H



$$| f_H(G_1) - f_H(G_2) | = | f(\text{}(G_1)) - f(\text{}(G_2)) |$$

$$\leq d(\text{}(G_1), \text{}(G_2))\, RS_f(H)/d$$

$$\leq c\, RS_f(H)$$

Outline

- Differential Privacy in Social Networks
- The Problem
- Restricted Sensitivity
- **Algorithms**
  – General Template
  – Possibility: A General Inefficient Algorithm
  – Efficient Algorithms for H_k via Projections
    - **Edge Adjacency Model**
    - Vertex Adjacency

Edge Adjacency (□: G □ H_k)

1. For each vertex v with deg(v) > k:
   - Let $(v,u_1),\ldots,(v, u_{deg(v)})$ denote the edges incident to v in canonical order
   - Mark the edge (v,u_i) for each i > k.
2. Delete all marked edges.
3. Call the resulting graph □(G).

Edge Adjacency (□: G □ H_k)

Claim: The efficiently computable mapping □(G) is a 3-smooth projection.

Proof:

(1) If G □ H_k then no edges are deleted so □(G)=G.

(2) WTS: For any G~G', d(□(G), □(G')) ≤ 3.

Edge Adjacency (□: G □ H_k)

(2) For any G~G’, d(□(G), □(G’)) ≤ 3.

Proof:

If E(G’) = E(G)+(u,v) then

i. □(G’) might still contain (u,v).
ii. At most one edge (u,w_1) □ E(□(G))\ E(G’).
iii. At most one edge (v,w_2) □ E(□(G))\ E(G’).

So d(□(G), □(G’)) ≤ 3.

(QED)

Outline

- Differential Privacy in Social Networks
- The Problem
- Restricted Sensitivity
- **Algorithms**
  – General Template
  – Possibility: A General Inefficient Algorithm
  – **Efficient Algorithms for H_k via Projections**

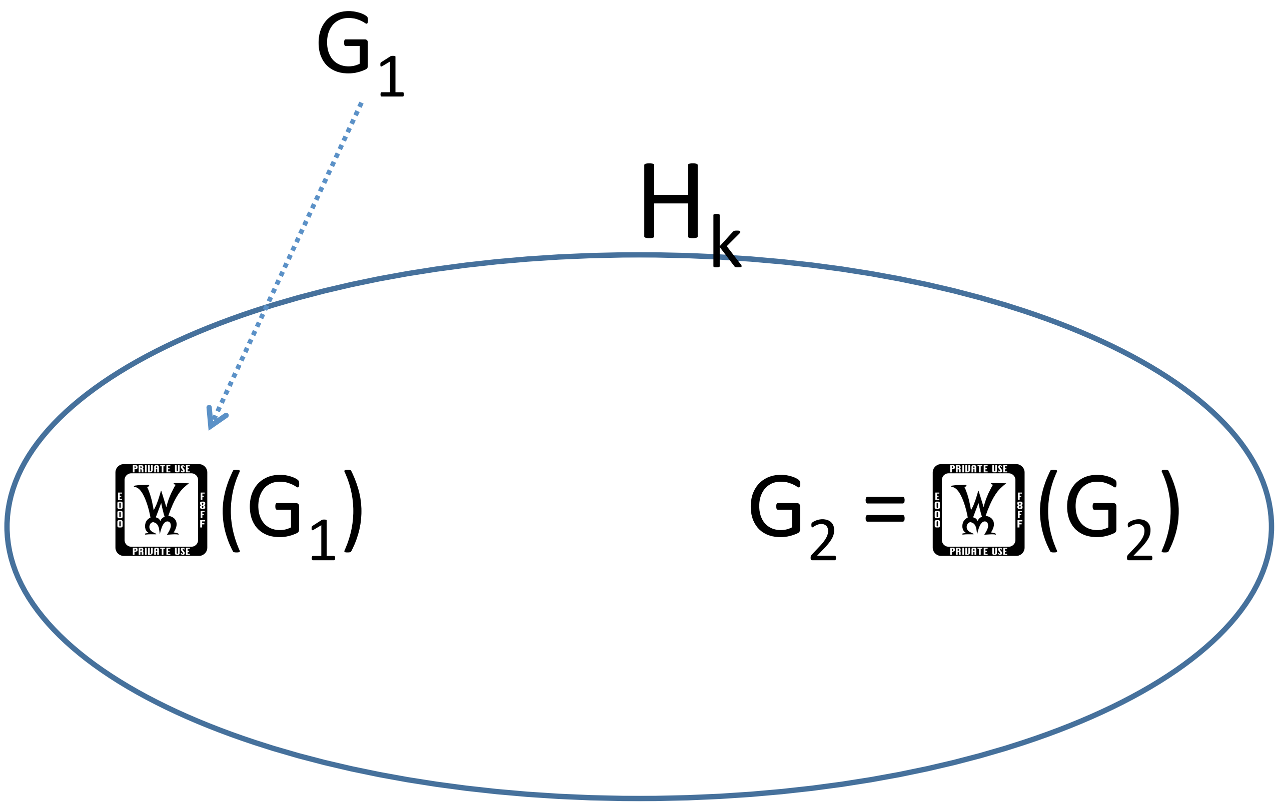
High Level Picture

- Concept: Smooth Distance Estimation
  - Project all G to H_{2k}
  - Lemma: Accuracy for G in H_k (leveraging smooth sensitivity)

- Constructing a Smooth Distance Estimator
  - LP Rounding

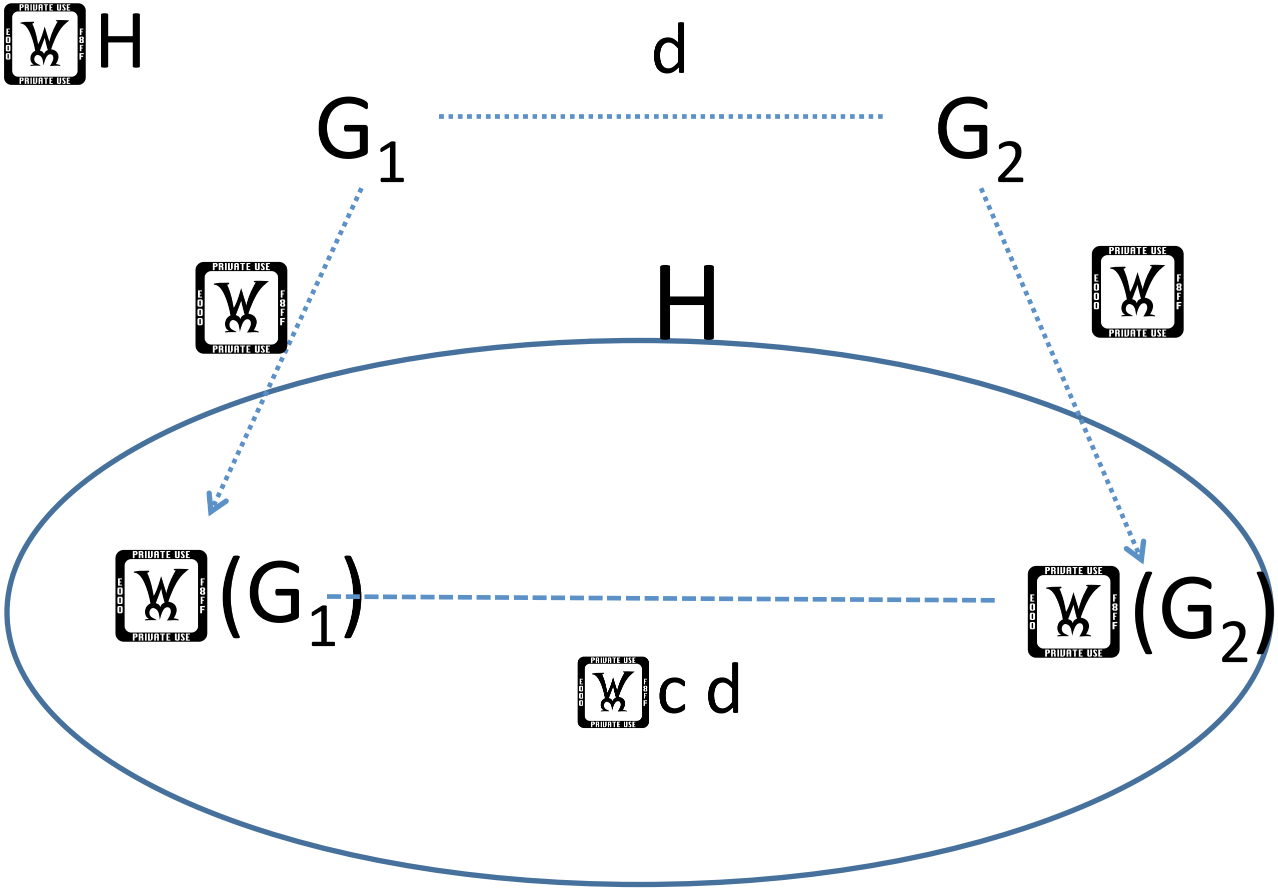
Naïve Projection

[illegible]: G [illegible] H



Map Close Graphs to Close Graphs

□: G □ H



Projection Lemma

Let □: G □H be a projection such that

1. For G □ H, □(G) = G, and
2. For G~G', d(□(G), □(G)) □ c

then $f_H(G) = f(□(G))$ satisfies:

1. For G □ H, $f_H(G) = f(G)$, and
2. $GS_{f_H} \leq c\ RS_f(H)$.

Construction is efficient in the edge adjacency model!

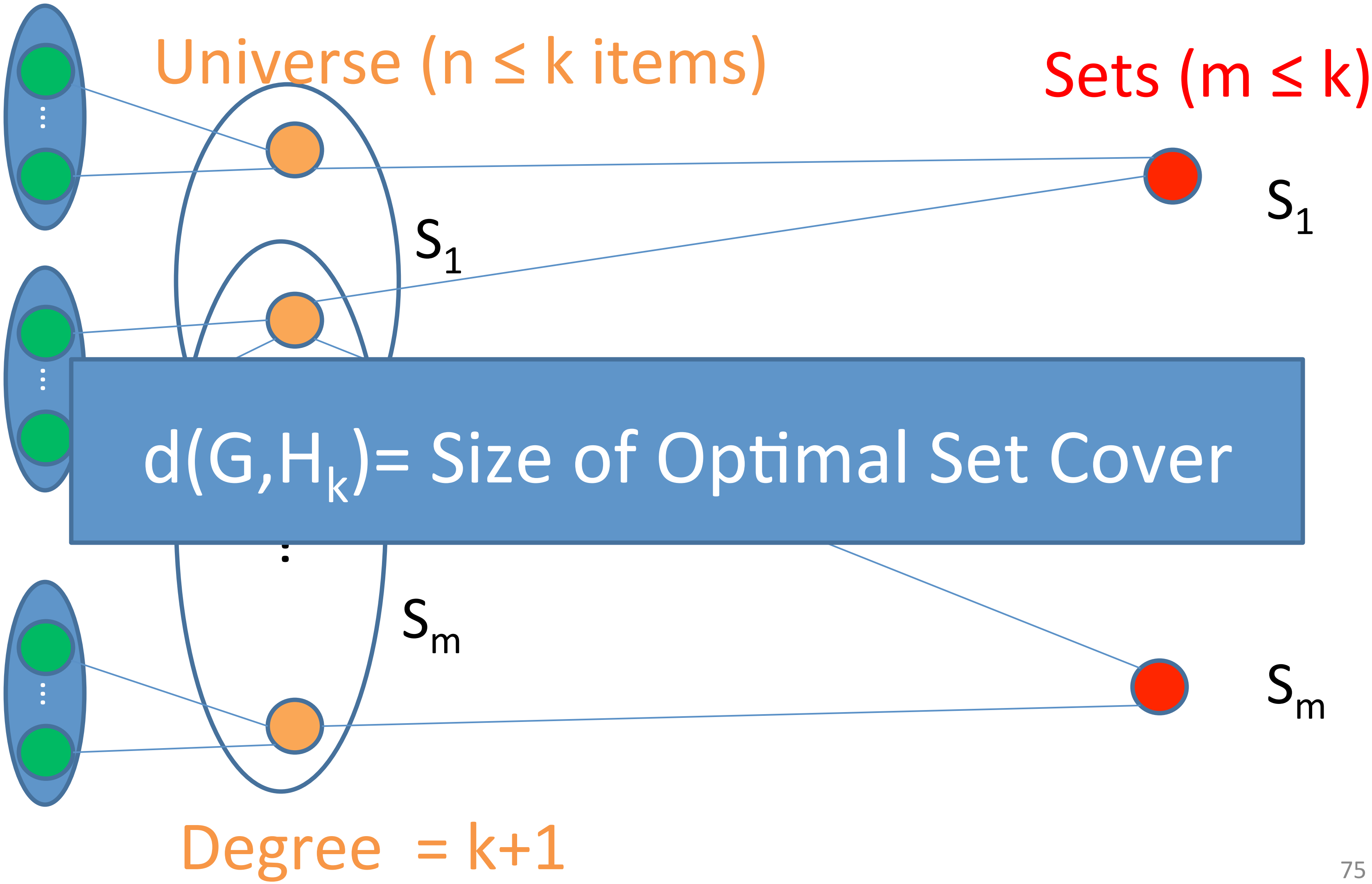
Vertex Adjacency?

- Map close graphs to close graphs in H_k?
  - Worked for Edge Adjacency Model

- Can't Work! ☹ Why?
- It would allow us to approximate $d(G,H_k)$.

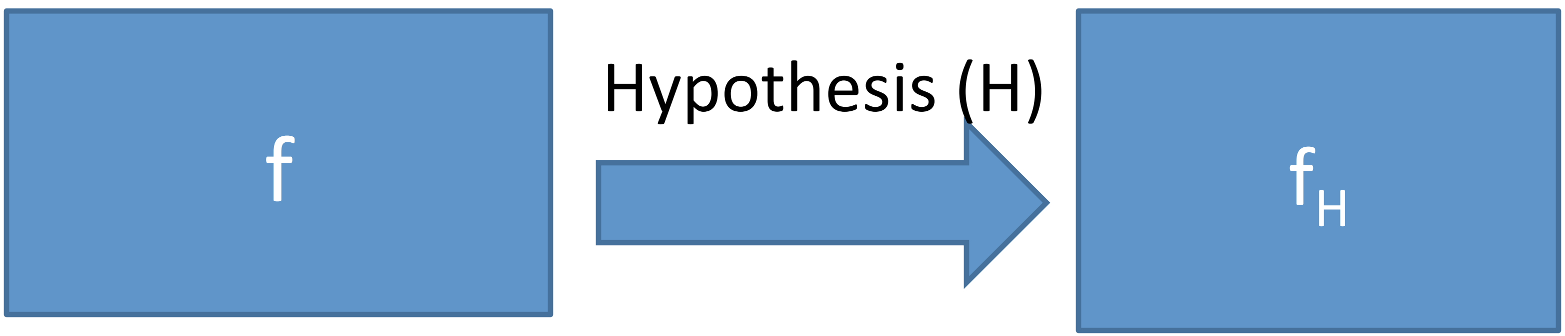
Claim: It is NP-hard to approximate $d(G,H_k)$ to within any constant factor (reduction from set cover)

Reduction from Set Cover



$d(G,H_k)$= Size of Optimal Set Cover

A New Approach



Leverage Smooth Sensitivity!

$$\forall G \in H,\ S_{f_H,\beta}(G) = O(RS_f(H))$$

$$GS_{f_H} = RS_f(H)$$

Privacy For All

Theorem: The mechanism

$$A(G) = f_{H_k}(G) + Lap\left(\frac{2S_{f_{H_k},\beta}(G)}{\varepsilon}\right)$$

preserves differential privacy for [illegible]=-[illegible]/2 ln [illegible].

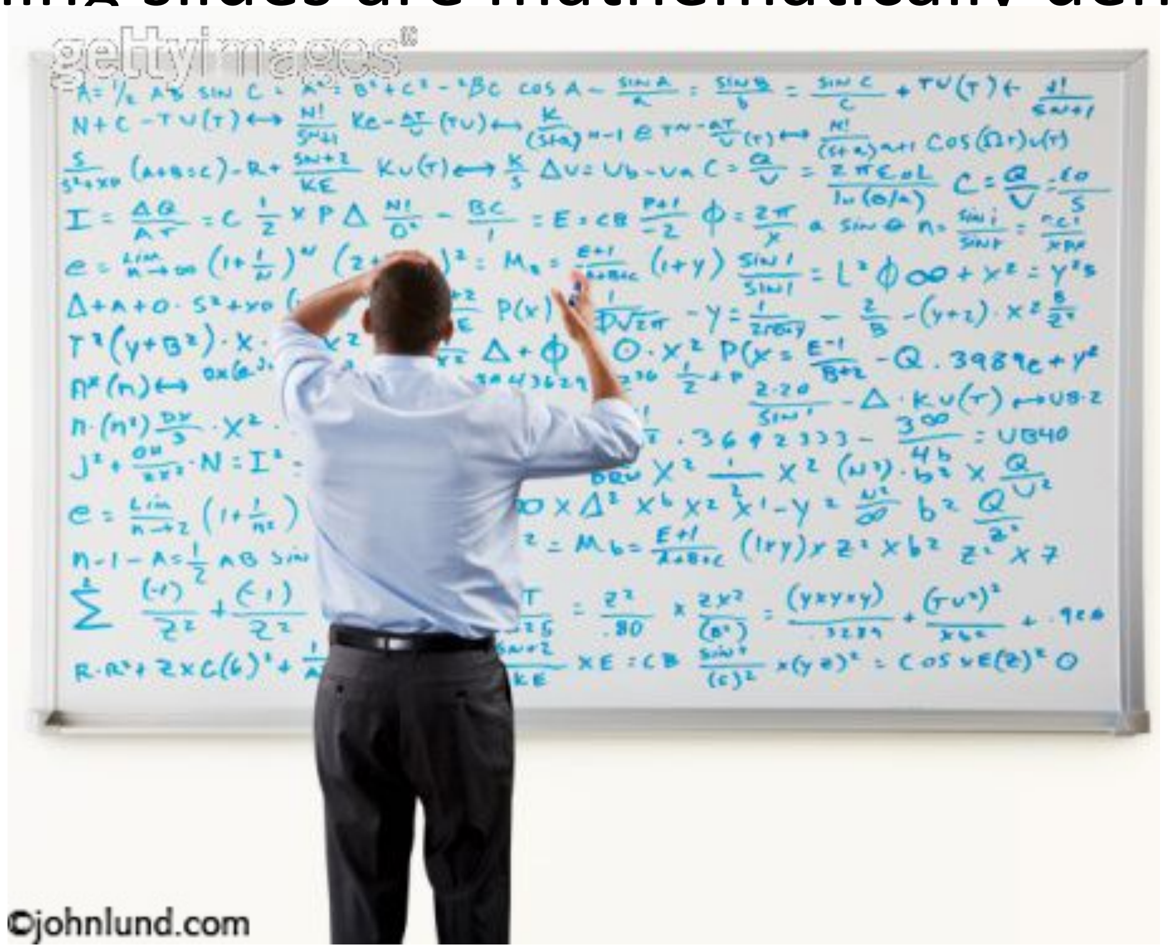
Goal: Accuracy For Some

$$\forall G \in H_k, S_{f_{H_k},\beta}(G) = O(k).$$

For Local Profile Queries

Warning!

Remaining slides are mathematically dense!



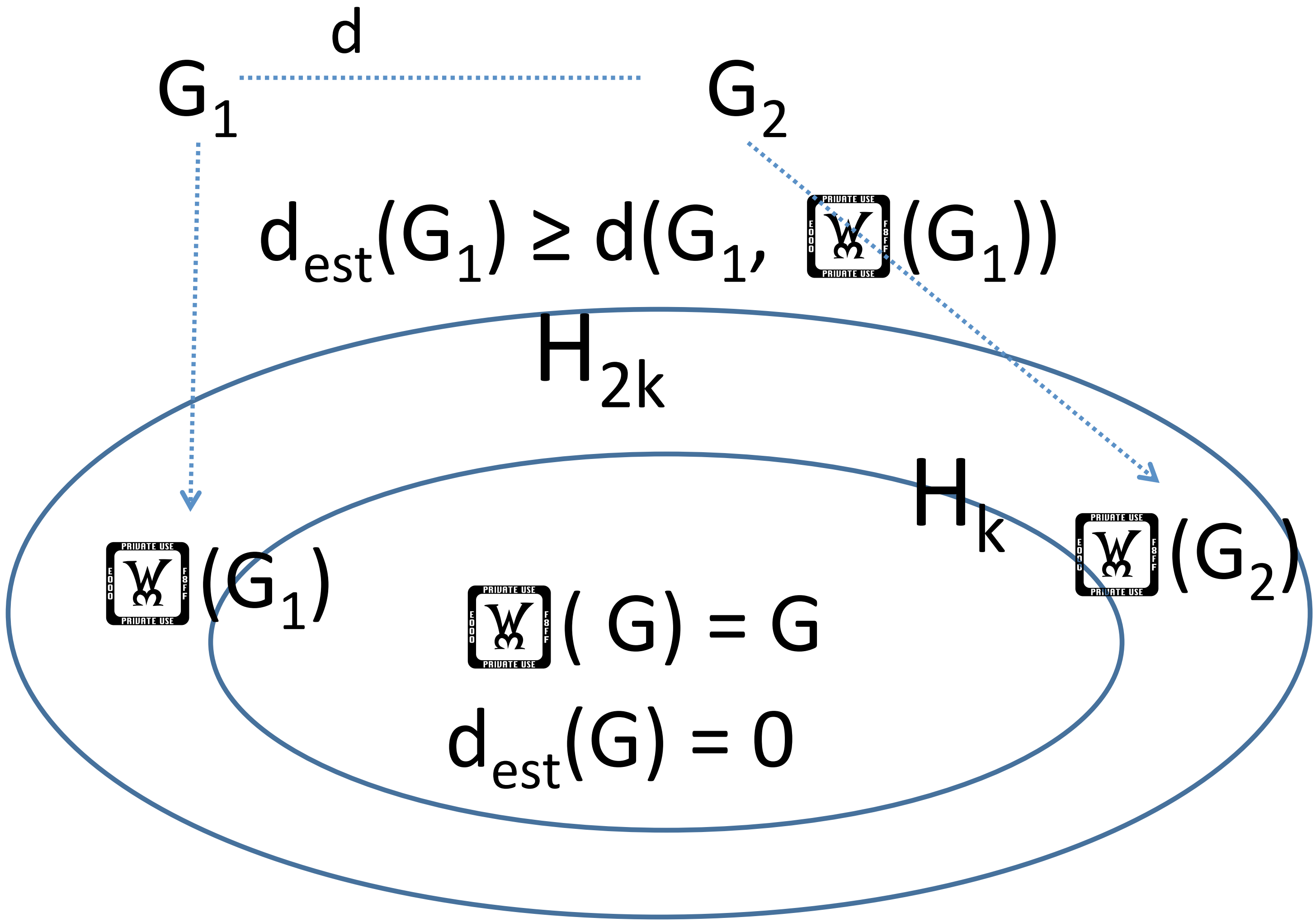
High Level Picture

- Concept: c-Smooth Distance Estimation

- Lemma
  - Accuracy for Some

- Constructing a 4-Smooth Distance Estimator
  - LP Rounding

c-Smooth Distance Estimation



C-smooth [illegible] $|d_{est}(G_1)- d_{est}(G_2)| \leq c\ d$

c-Smooth Distance Estimation

Definition: Let □: G □ H_{2k} be an efficiently computable projection and let d_{est} be an efficiently computable function which satisfies

1. For G □ H_k, $d_{est}(G) = 0$,
2. $d_{est}(G) \geq d(G,$ □$(G))$, and
3. $|d_{est}(G) - d_{est}(G')| \leq c$ for $G \sim G'$,

then d_{est} is a c-smooth distance estimator.

High Level Picture

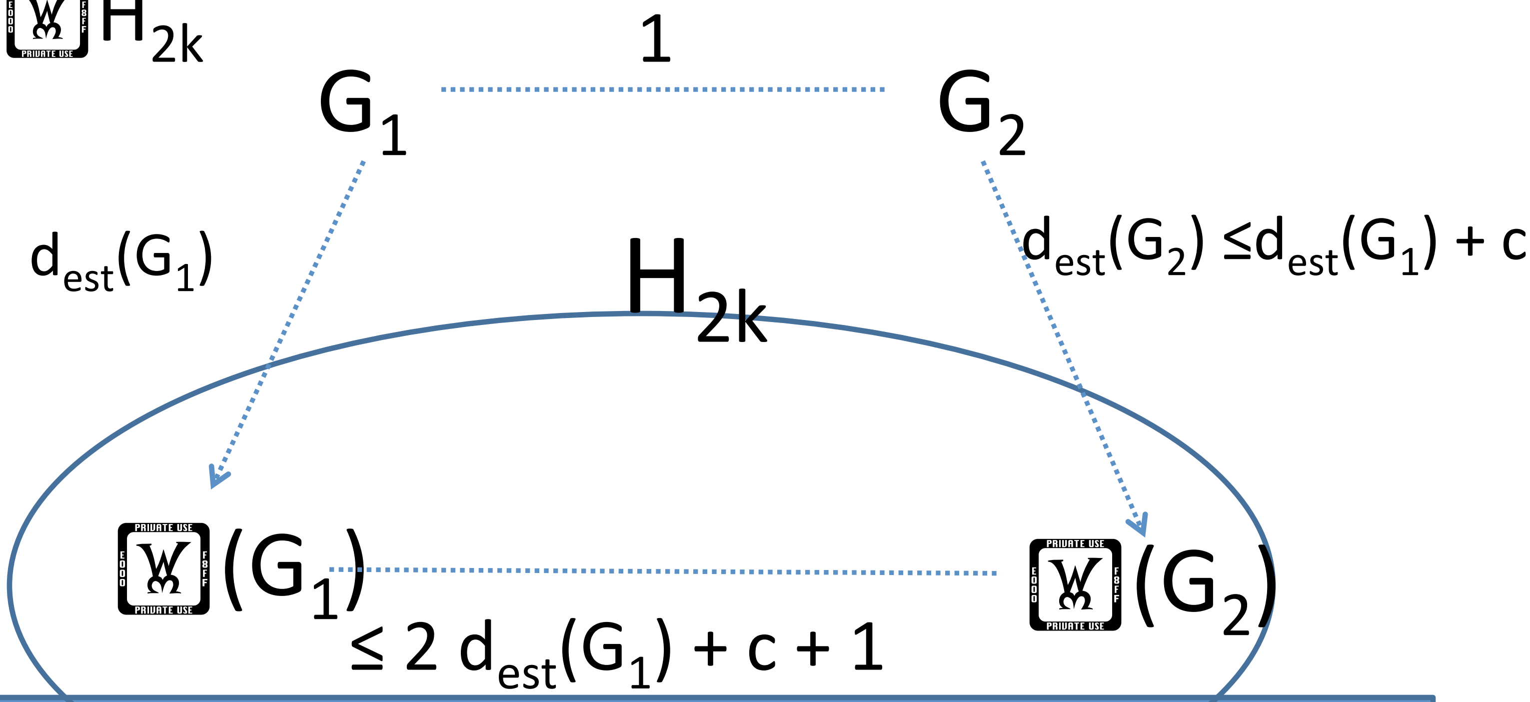
- Concept: c-Smooth Distance Estimation

- **Lemma**
  - **Privacy for All, Accuracy for Some**

- Constructing a 4-Smooth Distance Estimator
  - LP Rounding

c-Smooth Distance Estimation

: G H_{2k}



$$LS_{f_H}(G_1) \leq (2d_{est}(G_1) + c + 1) RS_f(H_{2k})$$

Smooth Sensitivity

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})$$

Fact 1: $S_{f_{H_k},\beta}(G) \geq LS_{f_{H_k}}(G)$

Fact 2: $S_{f_{H_k},\beta}(G)$ is [illegible]-smooth

Fact 3: $\forall G \in H_k, \quad S_{f_{H_k},\beta}(G) = O(k)$

c-Smooth Distance Estimation Lemma

Let [illegible]: G [illegible] H_{2k} be a projection with a c-smooth distance estimator d_{est}, and let

$$f_{H_k}(G) = f(\mu(G)).$$

Then for every G [illegible] H_k $S_{f_{H_k},\beta}(G) = O(k)$.

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})$$

Fact 1: $S_{f_{H_k},\beta}(G) \geq LS_{f_{H_k}}(G)$

$$S_{f_{H_k},\beta}(G) \geq \exp\left(-\frac{\beta}{c}\left(d_{est}(G) - d_{est}(G)\right)\right)(2d_{est}(G) + c + 1)RS_f(H_{2k})$$

$$\geq (2d_{est}(G) + c + 1)RS_f(H_{2k})$$

$$\geq (2d_{est}(G) + c + 1)\max_{G' \sim G}\left(\frac{f(\mu(G)) - f(\mu(G'))}{d(G, G')}\right)$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1) RS_f(H_{2k})$$

Fact 1: $S_{f_{H_k},\beta}(G) \geq LS_{f_{H_k}}(G)$

$$S_{f_{H_k},\beta}(G) \geq (2d_{est}(G) + c + 1) \max_{G' \sim G}\left(\frac{f(\mu(G)) - f(\mu(G'))}{2d_{est}(G) + c + 1}\right)$$

$$\geq \max_{G' \sim G}\left(f_{H_k}(G) - f_{H_k}(G')\right)$$

$$= LS_{f_{H_k}}(G)$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})$$

Fact 2: $S_{f_{H_k},\beta}(G)$ is [illegible]-smooth.

$$\frac{S_{f_{H_k},\beta}(G)}{S_{f_{H_k},\beta}(G')} = \frac{\max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})}{\max_{d \geq d_{est}(G')} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G')\right)\right)(2d + c + 1)RS_f(H_{2k})}$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})$$

Fact 2: $S_{f_{H_k},\beta}(G)$ is [illegible]-smooth.

$$\frac{S_{f_{H_k},\beta}(G)}{S_{f_{H_k},\beta}(G')} = \frac{\exp\left(-\frac{\beta}{c}\left(d^* - d_{est}(G)\right)\right)(2d^* + c + 1)RS_f(H_{2k})}{\max_{d \geq d_{est}(G')} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G')\right)\right)(2d + c + 1)RS_f(H_{2k})}$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1) RS_f(H_{2k})$$

Fact 2: $S_{f_{H_k},\beta}(G)$ is [illegible]-smooth.

$$\frac{S_{f_{H_k},\beta}(G)}{S_{f_{H_k},\beta}(G')} \leq \frac{\exp\left(-\frac{\beta}{c}\left(d^* - d_{est}(G)\right)\right)(2d^* + c + 1) RS_f(H_{2k})}{\exp\left(-\frac{\beta}{c}\left(d^* - d_{est}(G')\right)\right)(2d^* + c + 1) RS_f(H_{2k})}$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})$$

Fact 2: $S_{f_{H_k},\beta}(G)$ is [illegible]-smooth.

$$\frac{S_{f_{H_k},\beta}(G)}{S_{f_{H_k},\beta}(G')} \leq \frac{\exp\left(-\frac{\beta}{c}\left(d^* - d_{est}(G)\right)\right)}{\exp\left(-\frac{\beta}{c}\left(d^* - d_{est}(G')\right)\right)}$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1) RS_f\left(H_{2k}\right)$$

Fact 2: $S_{f_{H_k},\beta}(G)$ is [illegible]-smooth.

$$\frac{S_{f_{H_k},\beta}(G)}{S_{f_{H_k},\beta}(G')} \leq \exp\left(-\frac{\beta}{c}\left|d_{est}(G') - d_{est}(G)\right|\right)$$

$$\leq \exp(-\beta)$$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1) RS_f(H_{2k})$$

Fact 3: $\forall G \in H_k, \quad S_{f_{H_k},\beta}(G) = O(k)$

Calculus ➡ $S_{f_{H_k},\beta}(G) \leq g\left(\frac{\beta}{c}\right) \exp\left(\frac{\beta}{c} d_{est}(G)\right) RS_f(H_{2k})$

Where $g(x) = \begin{cases} c+1 & 0 \leq x \leq \frac{2}{c+1} \\ 2\frac{1}{x}\exp\left(-1 + \frac{c+1}{2}x\right) & x > \frac{2}{c+1} \end{cases}$

Smooth Upper Bound

$$S_{f_{H_k},\beta}(G) = \max_{d \geq d_{est}(G)} \exp\left(-\frac{\beta}{c}\left(d - d_{est}(G)\right)\right)(2d + c + 1)RS_f(H_{2k})$$

Fact 3: $\forall G \in H_k, \quad S_{f_{H_k},\beta}(G) = O(k)$

$$S_{f_{H_k},\beta}(G) \leq g\left(\frac{\beta}{c}\right)\exp\left(\frac{\beta}{c} d_{est}(G)\right) RS_f(H_{2k})$$

$$\forall G \in H_k, \quad S_{f_{H_k},\beta}(G) = g\left(\frac{\beta}{c}\right) RS_f(H_{2k}) = O(k)$$

High Level Picture

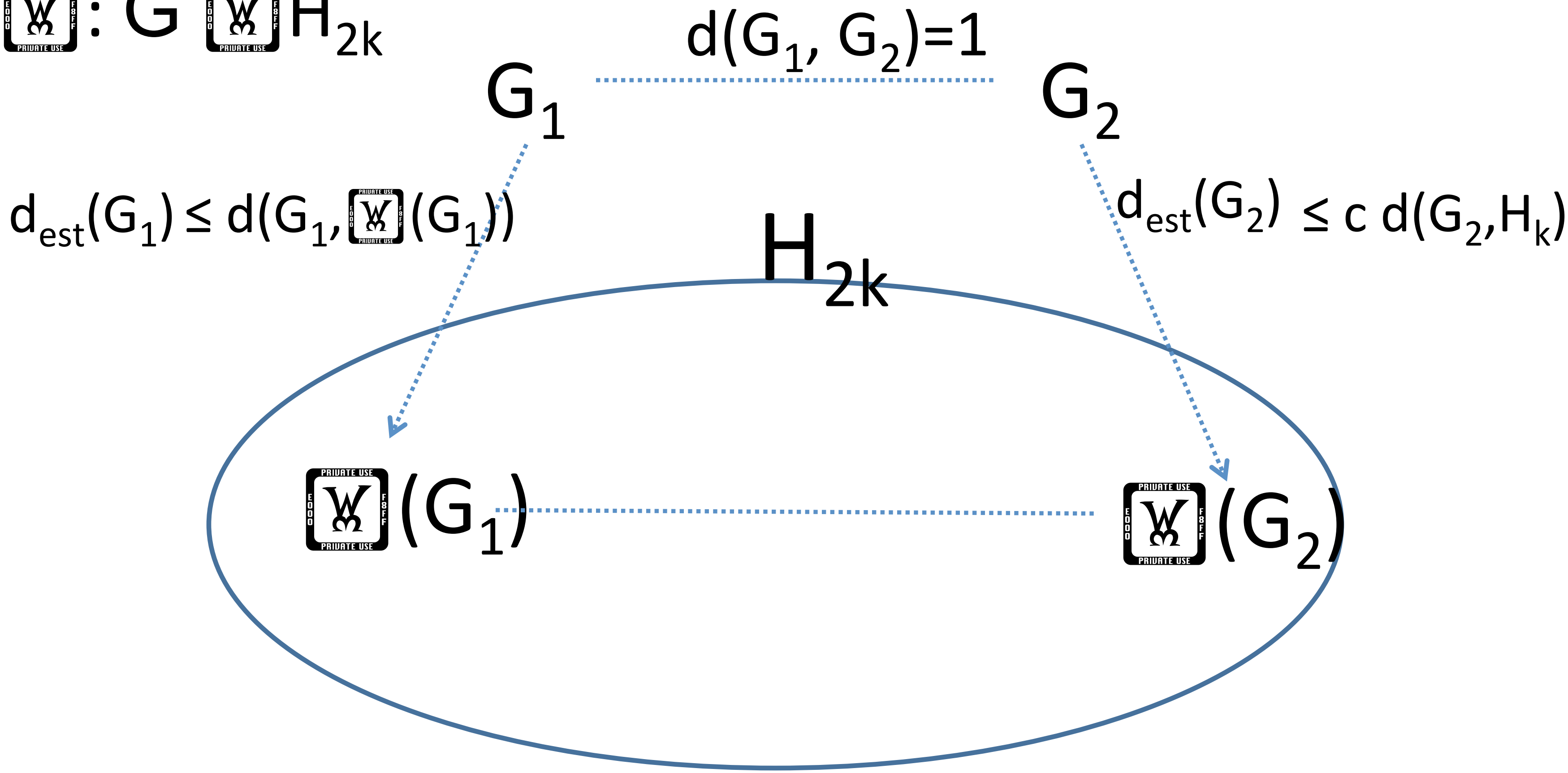
- Concept: c-Smooth Distance Estimation

- Lemma
  - Privacy for All, Accuracy for Some

- **Constructing a 4-Smooth Distance Estimator**
  - **LP Rounding**

c-Smooth Distance Estimation

□: G □ H_{2k}



C-smooth □ $|d_{est}(G_1) - d_{est}(G_2)| \leq c$

c-Smooth Distance Estimation via LP

$$\min \sum_{v \in V} x_v \quad s.t$$

Integral

Intuitions

∀v

∀u

∀(

∀u

or

Proof: Let v be the vertex such that G-v = G'-v.

1. Solve LP for G, and set
   - $x_v^* = 1$
   - $x_u^* = x_u$ for u v
2. Now x* is a feasible LP solution for G'.

$$\left| d_{est}(G) - d_{est}(G') \right| \le 4 x_v^*$$

(G) H_k

$$d_{est}(G) = 4 \sum_{v \in V} x_v$$

4-Smooth: $\left| d_{est}(G) - d_{est}(G') \right| \le 4$

Rounding the LP

$$\min \sum_{v \in V} x_v \quad s.t$$

$$\forall v, \quad 1 \geq x_v \geq 0$$

$$\forall u, v, \quad w_{u,v} \geq 0$$

$$\forall (u,v) \in E, \quad w_{u,v} \geq 1 - x_u - x_v$$

$$\forall u, \quad \sum_{v \neq u} w_{u,v} \leq k$$

$$y_v = \begin{cases} 1 & \text{If } x_v \geq 1/4 \\ 0 & \text{o.w.} \end{cases}$$

$$e_{u,v} = \begin{cases} 0 & \text{If } x_v \geq 1/4 \\ 0 & \text{If } x_u \geq 1/4 \\ 1 & \text{o.w.} \end{cases}$$

$e_{u,v}=1 \Rightarrow w_{u,v} \geq 1/2$

Rounding the LP

$$\sum_{v \in V} y_v \leq 4 \sum_{v \in V} x_v$$

$$\forall v, \quad 1 \geq y_v \geq 0$$

$$\forall u,v, \quad e_{u,v} \geq 0$$

$$\forall (u,v) \in E, \quad e_{u,v} \geq 1 - y_u - y_v$$

$$\forall u, \quad \sum_{v \neq u} e_{u,v} \leq 2k$$

$$y_v = \begin{cases} 1 & \text{If } x_v \geq ¼ \\ 0 & \text{o.w.} \end{cases}$$

$$e_{u,v} = \begin{cases} 0 & \text{If } x_v \geq ¼ \\ 0 & \text{If } x_u \geq ¼ \\ 1 & \text{o.w.} \end{cases}$$

$e_{u,v}=1$ ⇒ $w_{u,v} \geq ½$

Keep edge ⇔ $e_{u,v} = 1$

Call the resulting graph (G).

Rounding Facts

$$d\left(G,\mu(G)\right)\leq\sum_{v\in V}y_v\leq 4\sum_{v\in V}x_v=d_{est}(G)$$

$\forall G,$

So d_{est} is 4-smooth distance estimator for !

$$\forall G\in H_k,\quad \mu(G)=G$$

$$\forall G\in H_k,\quad d_{est}(G)=0$$

Reminder

Let □: G □ H_{2k} be a projection with a c-smooth distance estimator d_{est}, and let

$$f_{H_k}(G) = f(\mu(G)).$$

Then for every G □ H_k $S_{f_{H_k},\beta}(G) = O(k).$

Furthermore, $S_{f_{H_k},\beta}(G), \mu$ are both efficiently computable.

Goal Met

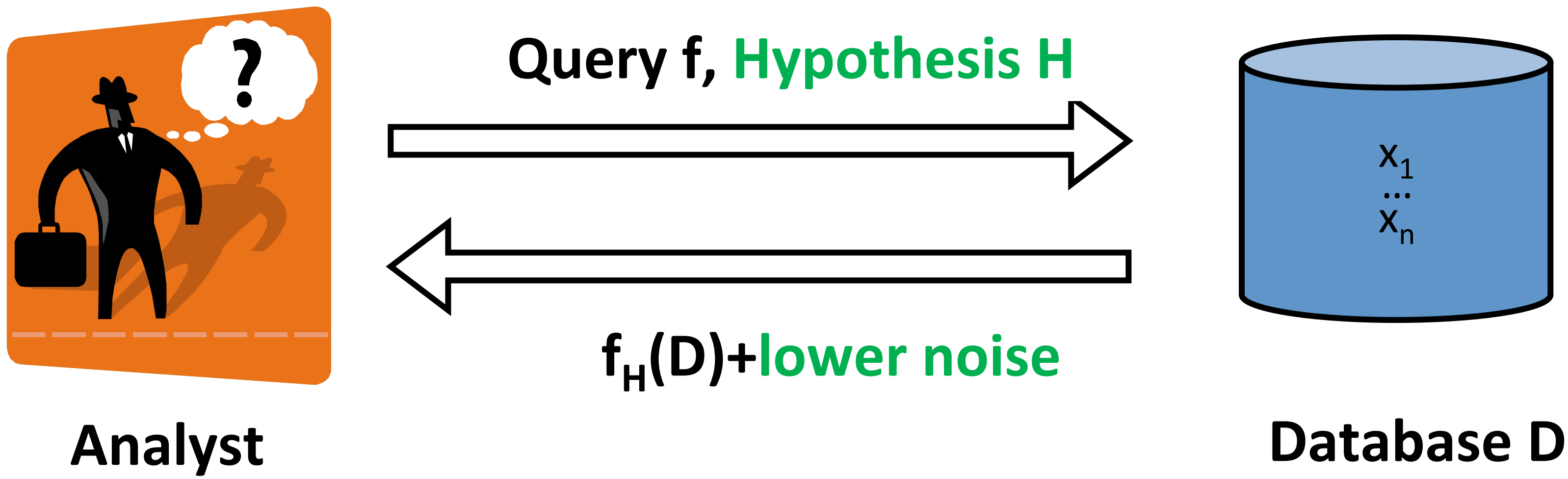
Theorem: The efficiently computable mechanism

$$A(G) = f_{H_k}(G) + Lap\left(\frac{2S_{f_{H_k},\beta}(G)}{\varepsilon}\right)$$

preserves differential privacy for [illegible]=-[illegible]/2 ln [illegible].

$$\forall G \in H_{k,} A(G) = f(G) + Lap\left(\frac{O(k)}{\varepsilon}\right).$$

Differential Privacy via Restricted Sensitivity



- Accurate <u>for D in H</u>
- Differential Privacy

Summary

	Adjacency	Hypothesis	Efficient?	Sensitivity of f_H
Alg 1	Any	Any	No	$RS_f(H)$
Alg 2	Edge	H_k	Yes	$3\ RS_f(H_k)$
Alg 3	Vertex	H_k	Yes	$O(RS_f(H_{2k}))$

	Local Profile Query		Subgraph Counting Query (P)	
Adjacency	**Smooth**	**Restricted**	**Smooth**	**Restricted**
Vertex	n	2k+1	$O(n^{\|P\|-1})$	$O(\|P\|\ k^{\|P\|-1})$

Open Questions

- Restricted sensitivity: Other relevant hypotheses H and associated constructions

- Social network privacy: Alternative to vertex adjacency
  - Too weak: Information about node A also in node B *influenced* by A

Thanks for Listening!

Questions?