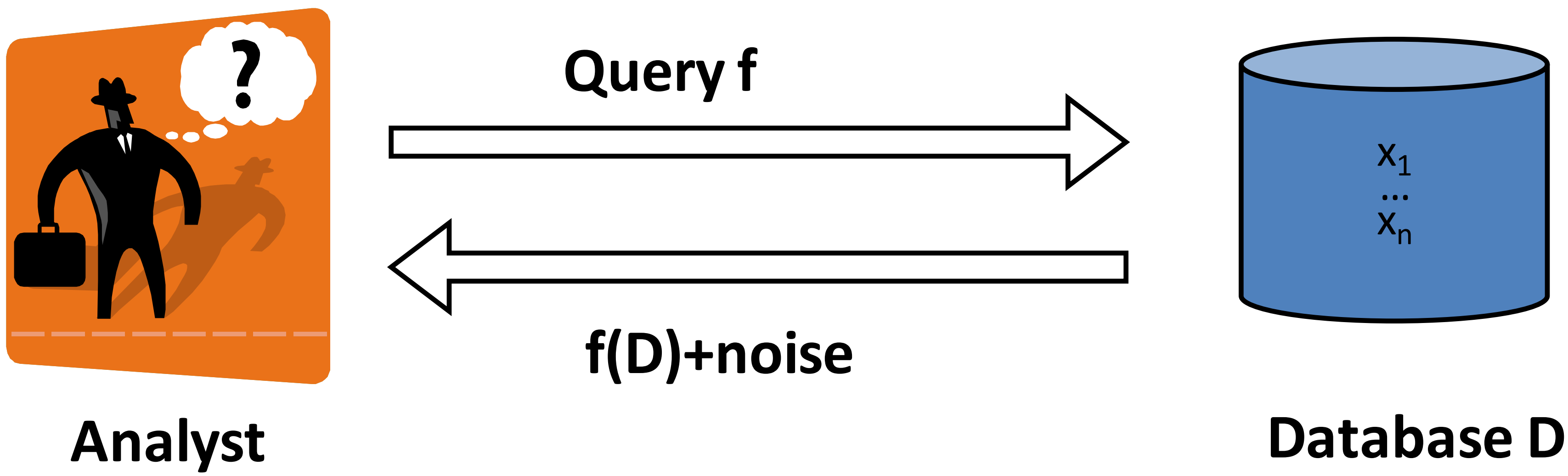
Global, Smooth, and Restricted Sensitivity in Differentially Private Data Analysis

Anupam Datta
Carnegie Mellon University

Fall 2016

Differential Privacy Setting



Usual goal:

- Accurate <u>for all D</u>
- Differential Privacy

Global Sensitivity

Global Sensitivity of f:

$$GS_f = \max_{D_1 \sim D_2} \| f(D_1) - f(D_2) \|$$

- Example Query, f = median

  GS_f is very high

- Issue: Global sensitivity depends only on function not on data set

Local Sensitivity

Local Sensitivity of f at D_1:

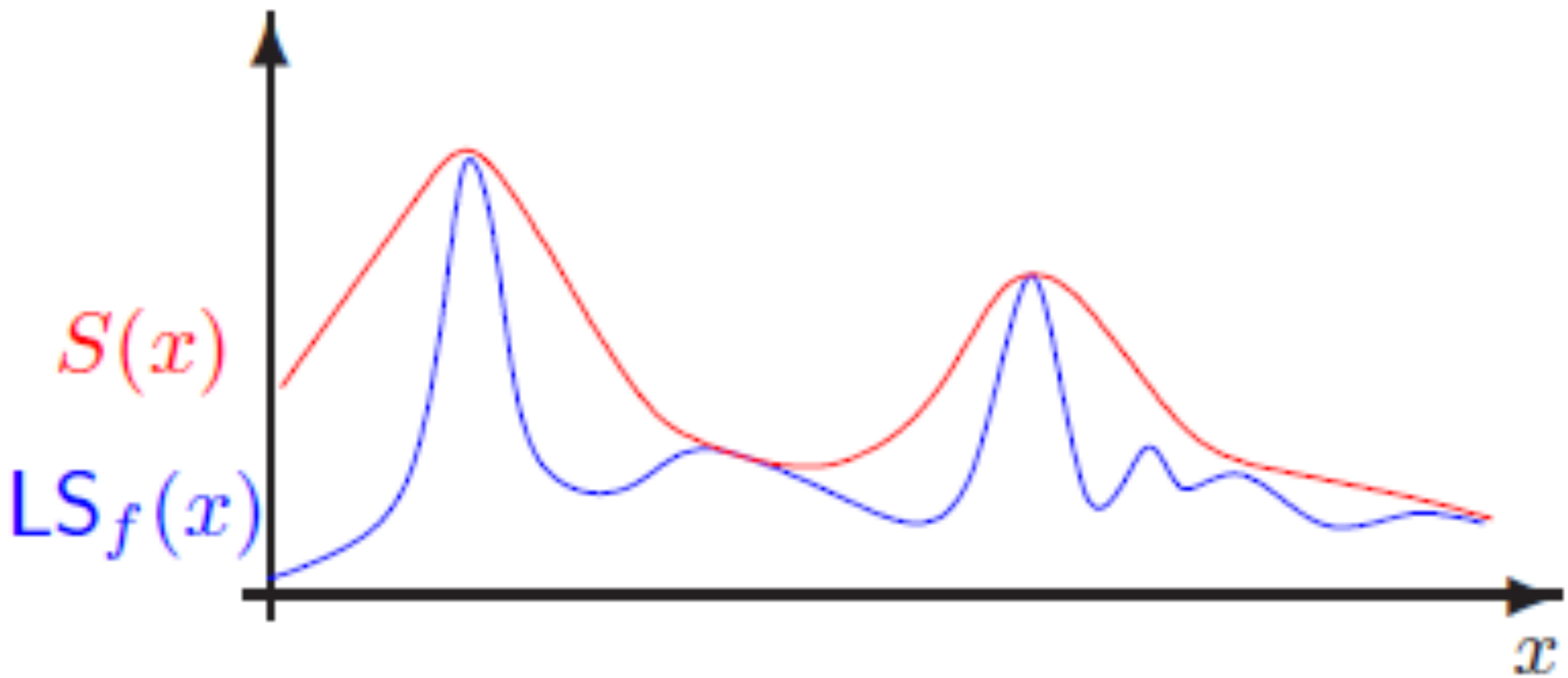
$$LS_f(D_1) = \max_{D_2 : D_1 \sim D_2} \| f(D_1) - f(D_2) \|$$

- Example Query, f = median

  $LS_f\ (D_1) << GS_f$

- Insight: Local sensitivity depends on function <u>and</u> data set

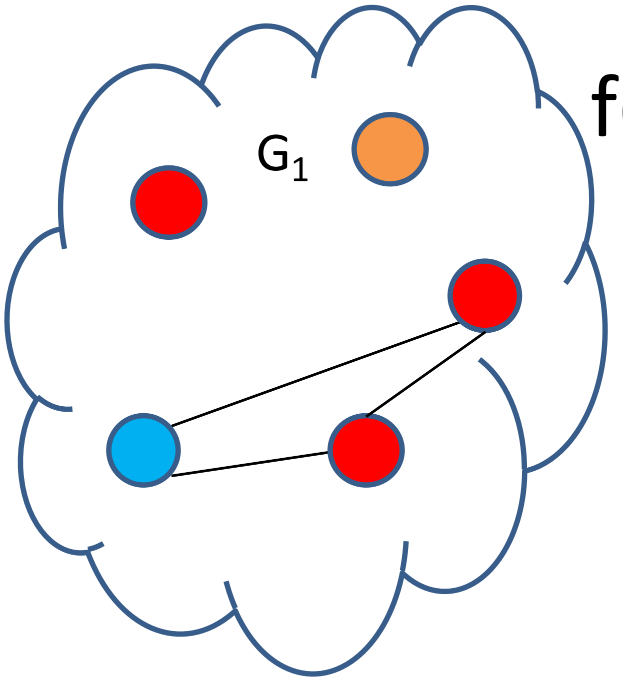
Smooth Sensitivity



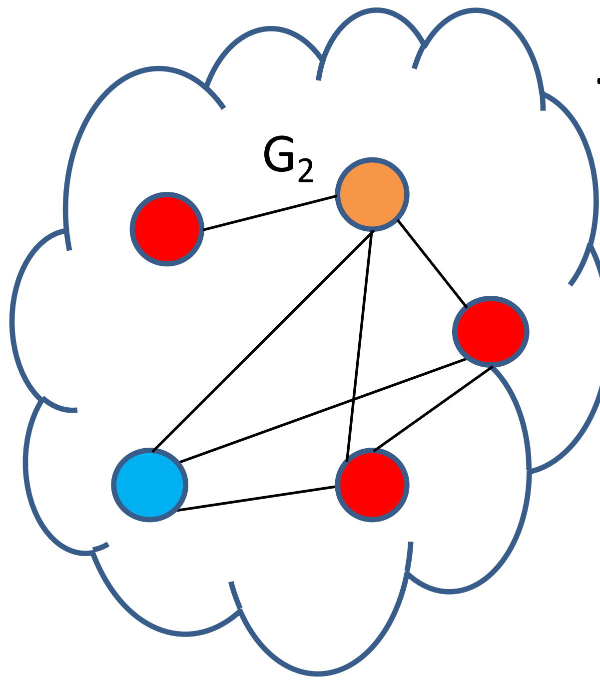
- Add noise proportional to smooth sensitivity rather than global sensitivity to satisfy differential privacy

Challenge: High Global and Smooth Sensitivity in Vertex Adjacency Model

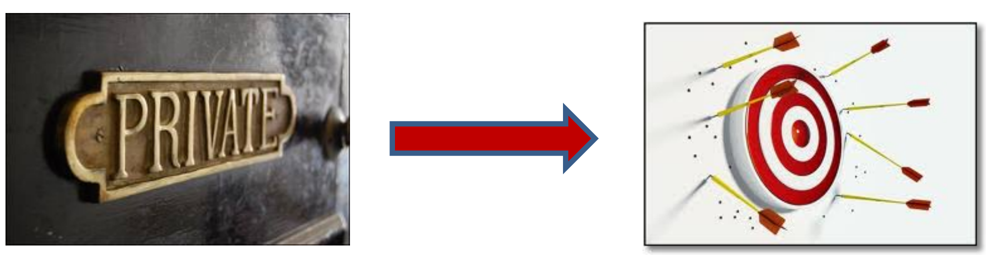
f(G) = “how many people in G know a pianist?”



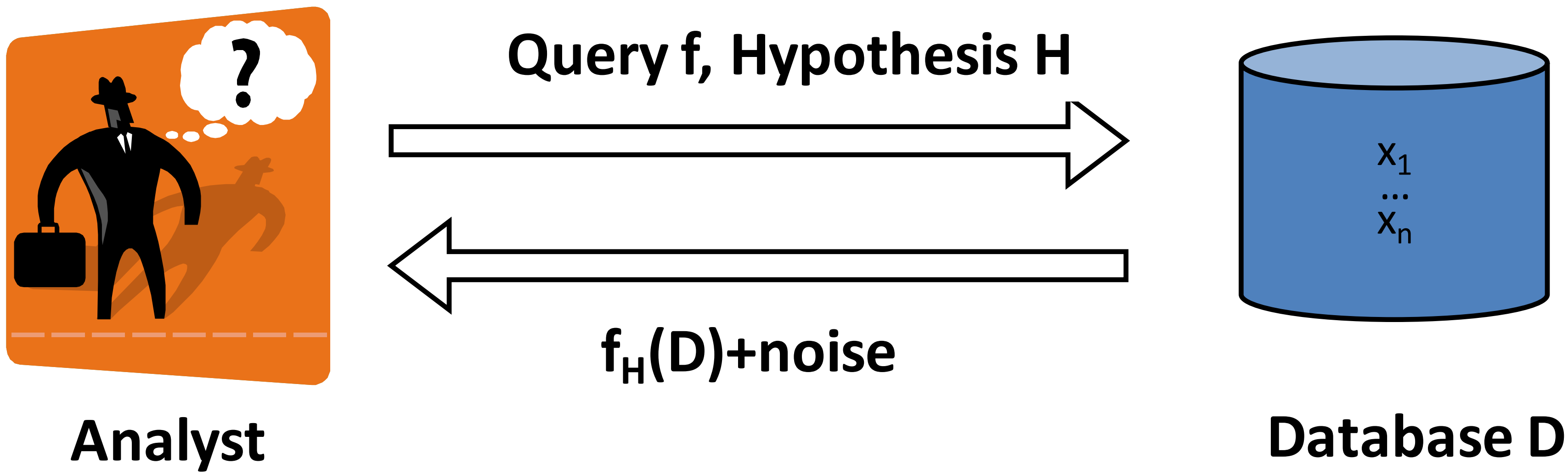
$f(G_1)=0$



$f(G_2)=n$



Restricted Sensitivity



- Accurate <u>for D in H</u> (lower noise)
- Differential Privacy

Restricted Sensitivity

<u>Hypothesis:</u> H subset of G

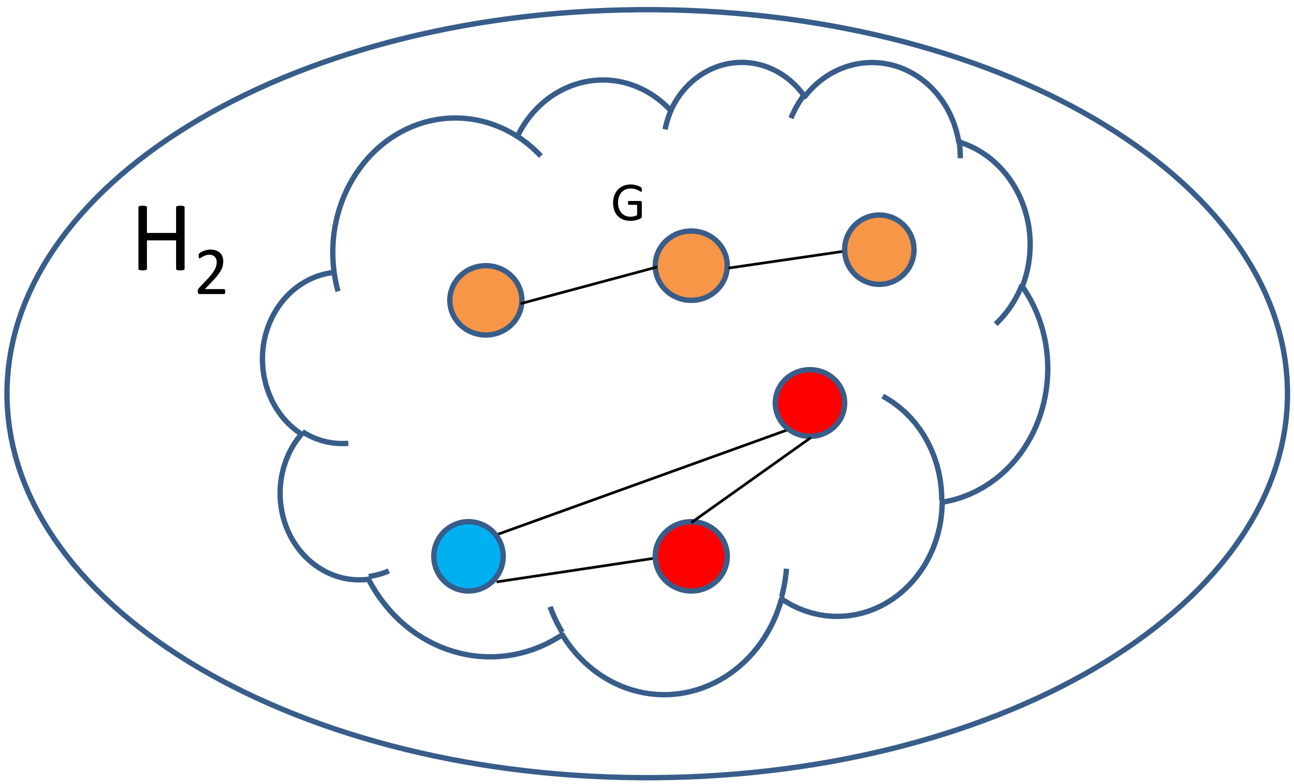
$$GS_f = \max_{G_1, G_2} \frac{|f(G_1) - f(G_2)|}{d(G_1, G_2)}$$

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

Bounded Degree Hypothesis

Bounded Degree Hypothesis:

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



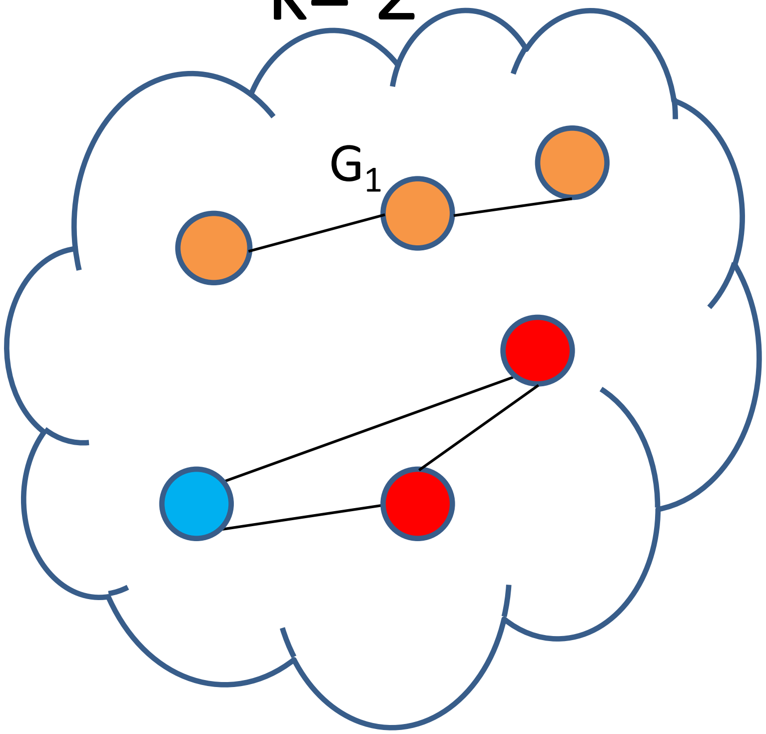
Typical:

$k << n$

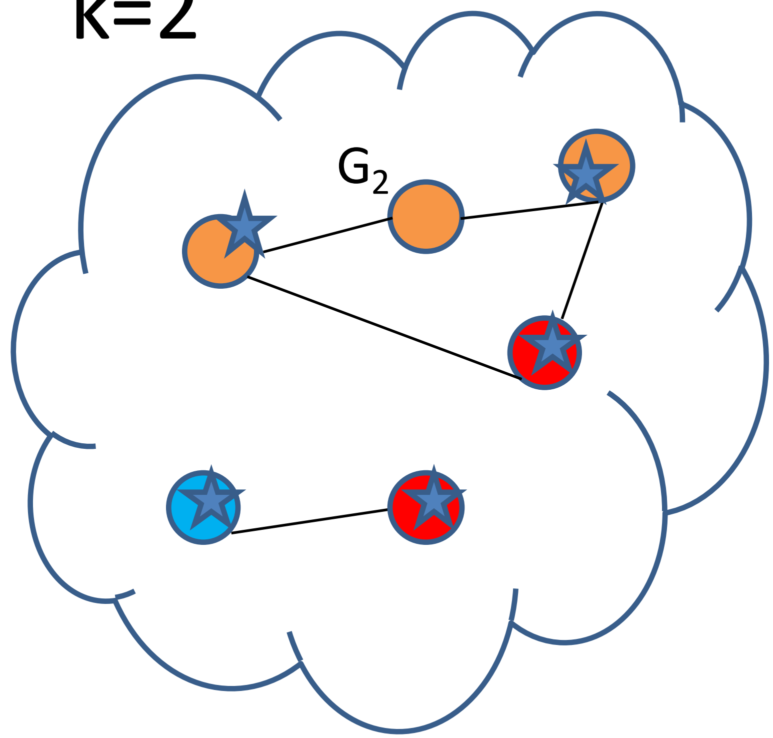
Restricted Sensitivity $RS_f(H_k)$

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

k= 2



k=2



Algorithms

- Efficient algorithms via projections

- Much higher accuracy for graphs (datasets) that satisfy hypotheses (e.g., degree bounded by k)

- Satisfies differential privacy