An Ad Omnia Approach to Defining and Achieving Private Data Analysis

by Cynthia Dwork
presented by Shaun Deaton

Examines two "ad omnia" notions of privacy in terms of statistical databases.

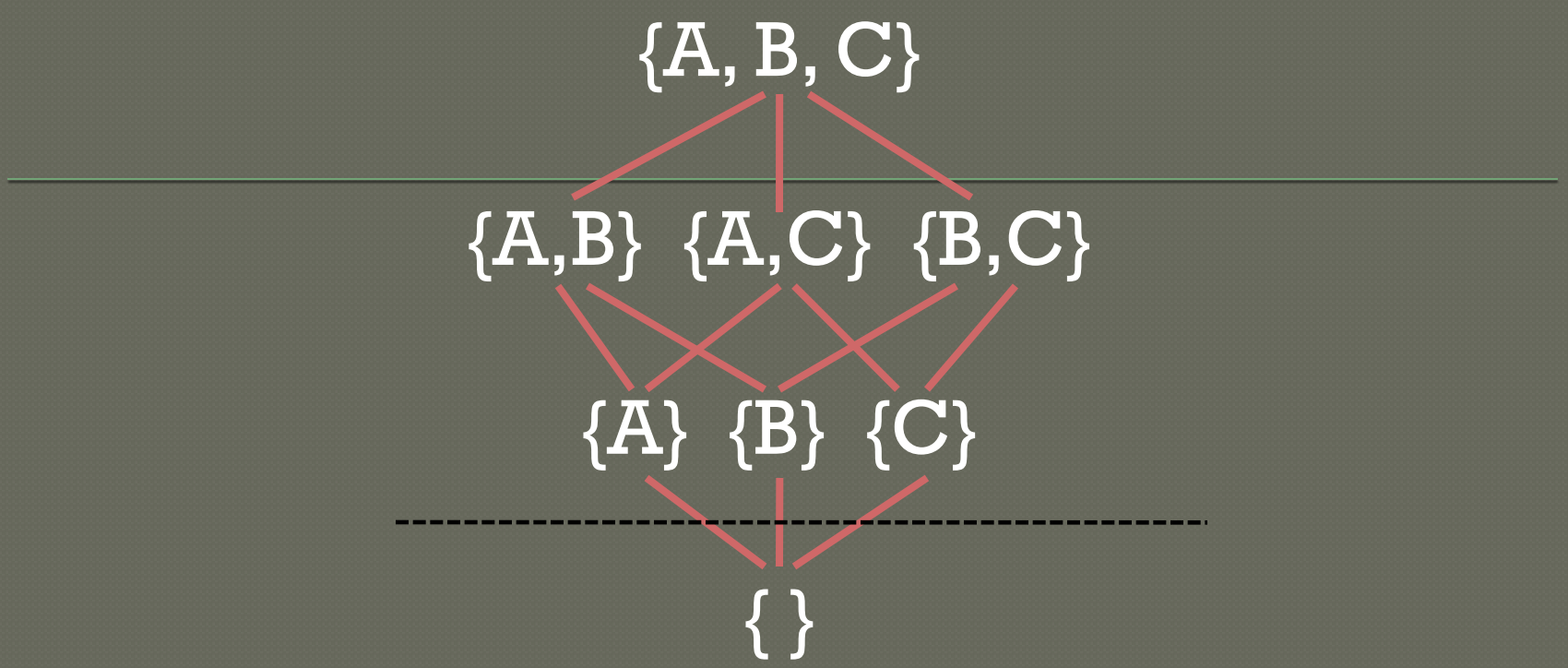
- Perfect semantic privacy
  - Impossible if require utility > 0.
- Differential privacy: minimize the increased risk incurred by joining or leaving a database

A definition for differential privacy; requires randomization mechanism $\mathcal{K}$ that takes databases as inputs to anonymize.

- $\mathcal{K}$ gives ε-differntial privacy if For All data sets D_1 & D_2 differing on at most one element, & For All $S \subseteq \text{Range}(\mathcal{K})$,

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\varepsilon) \Pr[\mathcal{K}(D_2) \in S],$$

where …



$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\varepsilon) \Pr[\mathcal{K}(D_2) \in S]$

A, B, and C are all possible "transcripts".

Queries are defined to be mappings from databases to vectors of real numbers, where $\mathcal{K}$ then adds "appropriate" noise to get the "response".

- Sensitivity of a query determines the spread in the noise required to normalize it over all databases differing by one.
- Global sensitivity & local insensitivity.
  - Global is between databases; overall what is the largest difference in outputs
  - Local is about a database; does changing an entry change the output (No → insensitive)

Questions/Discussion