

# Probabilistic Relational Reasoning for Differential Privacy

Gilles Barthe      Boris Köpf  
Federico Olmedo      Santiago Zanella Béguelin

IMDEA Software Institute, Madrid

POPL 2012

# Conflicting requirements!



User  
privacy

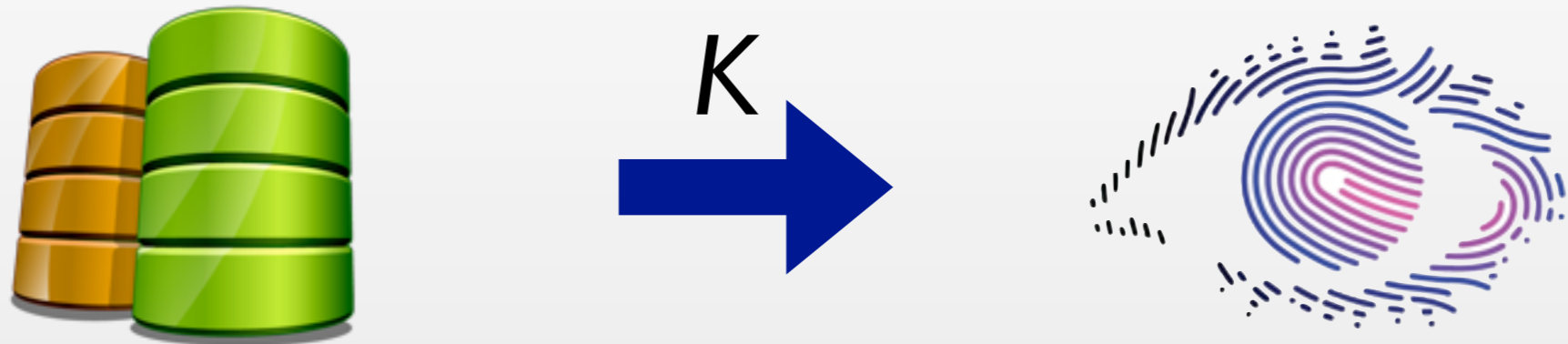
Utility of  
mining process



Need to achieve  
flexible balance

# Differential Privacy

Dwork [ICALP'06]



- Fix a (symmetric) adjacency relation  $\Phi$  on databases
- Fix a privacy budget  $\epsilon$

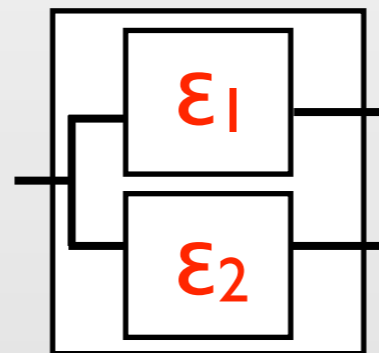
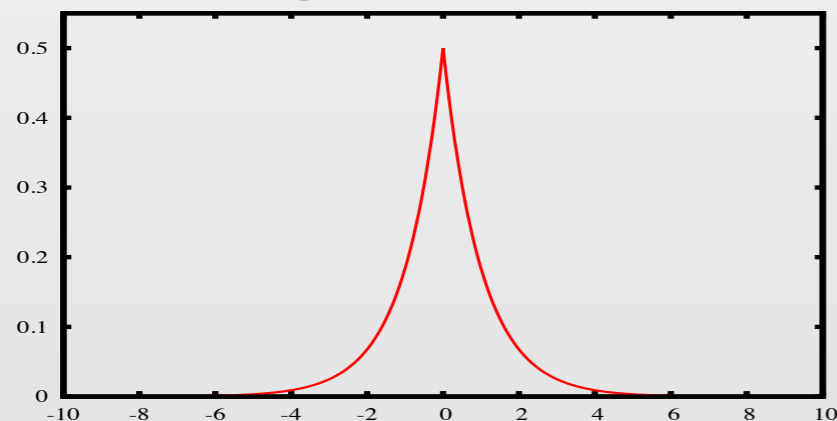
A randomized algorithm  $K$  is  **$\epsilon$ -differentially private** w.r.t.  $\Phi$  iff, for all databases  $D_1$  and  $D_2$ , and events  $S$

$$\Phi(D_1, D_2) \implies \Pr[K(D_1) \in S] \leq \exp(\epsilon) \times \Pr[K(D_2) \in S]$$

# Differential Privacy Primer

- Fundamentals

- Laplacian mechanism
- Composition theorems



$$\epsilon_1 + \epsilon_2$$

Language-based tool support available

- Expanding frontiers

- Mechanisms: exponential, median...
- Algorithms: streaming/graph/... algorithms
- Definitions: approximate differential privacy, pan privacy...

Increasingly complex, but not supported by existing tools!

# Our Contribution: CERTIPRIV

- Allows reasoning about approximate quantitative properties of randomized computations
  - Built from **first principles** and fully formalized in **Coq**
  - Machine-checked proofs of differential privacy
    - Correctness of Laplacian and Exponential **mechanisms**
    - State-of-art graph and streaming **algorithms**
- Generalizes CERTICRYPT and opens new applications to crypto

# Differential privacy as quantitative 2-safety

- $K$  is  $(\epsilon, \delta)$ -diff. private w.r.t.  $\Phi$  iff for all  $D_1$  and  $D_2$  and  $S$

$$\Phi(D_1, D_2) \implies \Pr[K(D_1) \in S] \leq \exp(\epsilon) \times \Pr[K(D_2) \in S] + \delta$$

Relational  
pre-condition

(Quantitative) relational  
post-condition

- We propose a **quantitative probabilistic relational Hoare Logic**

$$c_1 \sim_{\alpha, \delta} c_2 : \Phi \implies \Psi$$

such that  $c$  is  $(\epsilon, \delta)$ -diff. private w.r.t.  $\Phi$  iff

$$c \sim_{\exp(\epsilon), \delta} c : \Phi \implies \equiv$$

Needs to be lifted  
to distributions

# Characterizing differential privacy

$c_1 \sim_{\alpha, \delta} c_2 : \Phi \Rightarrow \Psi$  is valid iff for all  $D_1$  and  $D_2$

$$\Phi(D_1, D_2) \implies \text{lift}_{\alpha, \delta} \Psi (\llbracket c_1 \rrbracket D_1) (\llbracket c_2 \rrbracket D_2)$$

We define  **$\alpha$ -distance** such that:

- $c$  is  $(\epsilon, \delta)$ -diff. private w.r.t.  $\Phi$  iff for all  $D_1$  and  $D_2$

$$\Phi(D_1, D_2) \implies \Delta_{\alpha}(\llbracket c \rrbracket D_1, \llbracket c \rrbracket D_2) \leq \delta$$

- **Fundamental property of lifting**

$$\Delta_{\alpha}(\mu_1, \mu_2) \leq \delta \iff \text{lift}_{\alpha, \delta} \equiv \mu_1 \mu_2$$

# Lifting relations to distributions

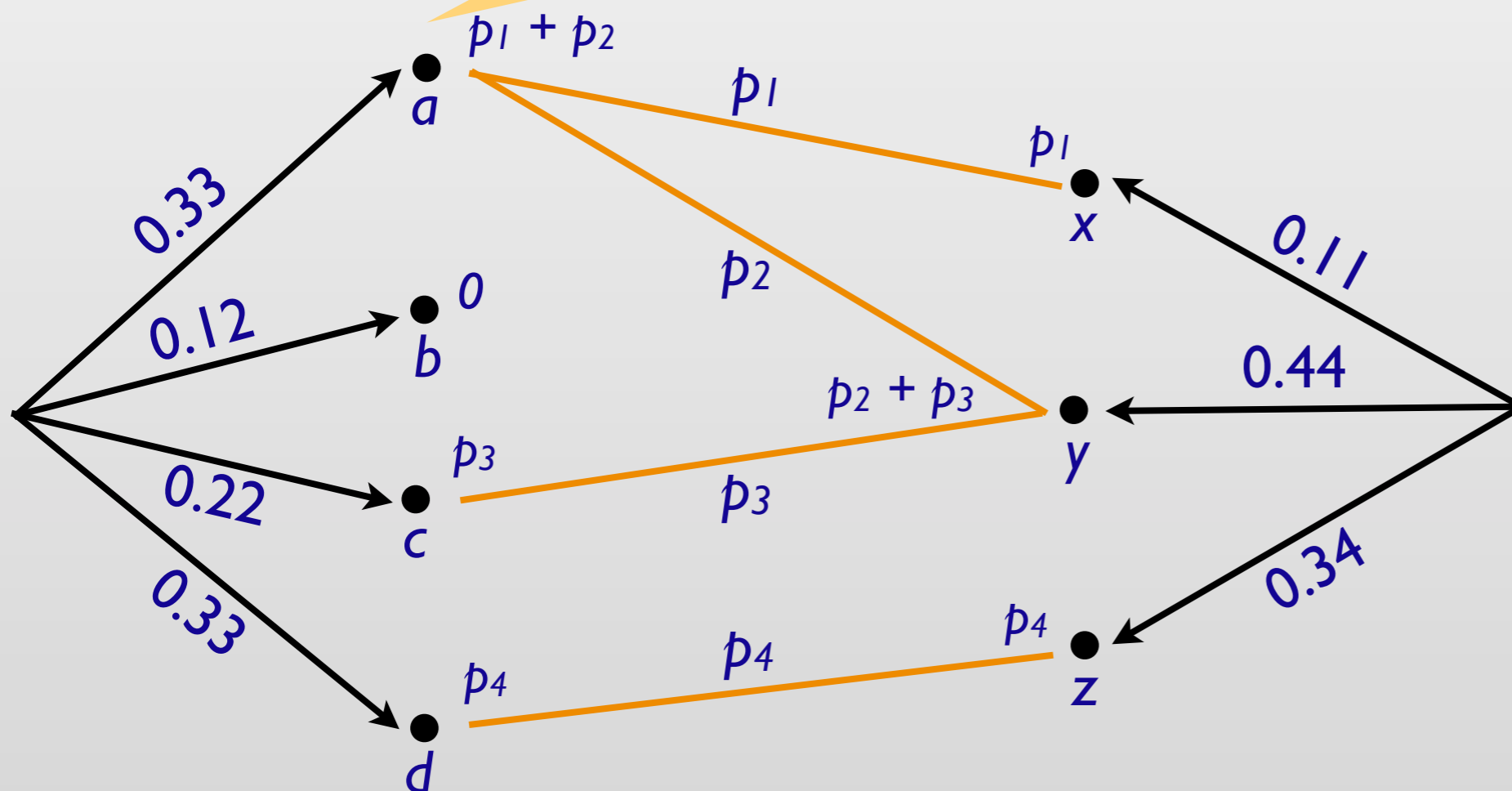
Given  $R = \{(a,x), (a,y), (c,y), (d,z)\}$ ,  $\alpha=1.1$  and  $\delta=0.01$

$$p_1 + p_2 \leq 0.33$$

$$\delta_a = \max\{0, 0.33 - \alpha (p_1 + p_2)\}$$

$$\delta_a + \delta_b + \delta_c + \delta_d \leq \delta$$

Witness distribution



$X \times Y$	$\mu(\cdot, \cdot)$
$(a, x)$	0.10
$(a, y)$	0.20
$(c, y)$	0.20
$(d, z)$	0.30
...	0

# Selected rules

## Sequential composition

$$\frac{\models c_1 \sim_{\alpha, \delta} c_2 : \Psi \Rightarrow \Phi' \quad \models c'_1 \sim_{\alpha', \delta'} c'_2 : \Phi' \Rightarrow \Phi}{\models c_1; c'_1 \sim_{\alpha\alpha', \delta+\delta'} c_2; c'_2 : \Psi \Rightarrow \Phi}$$

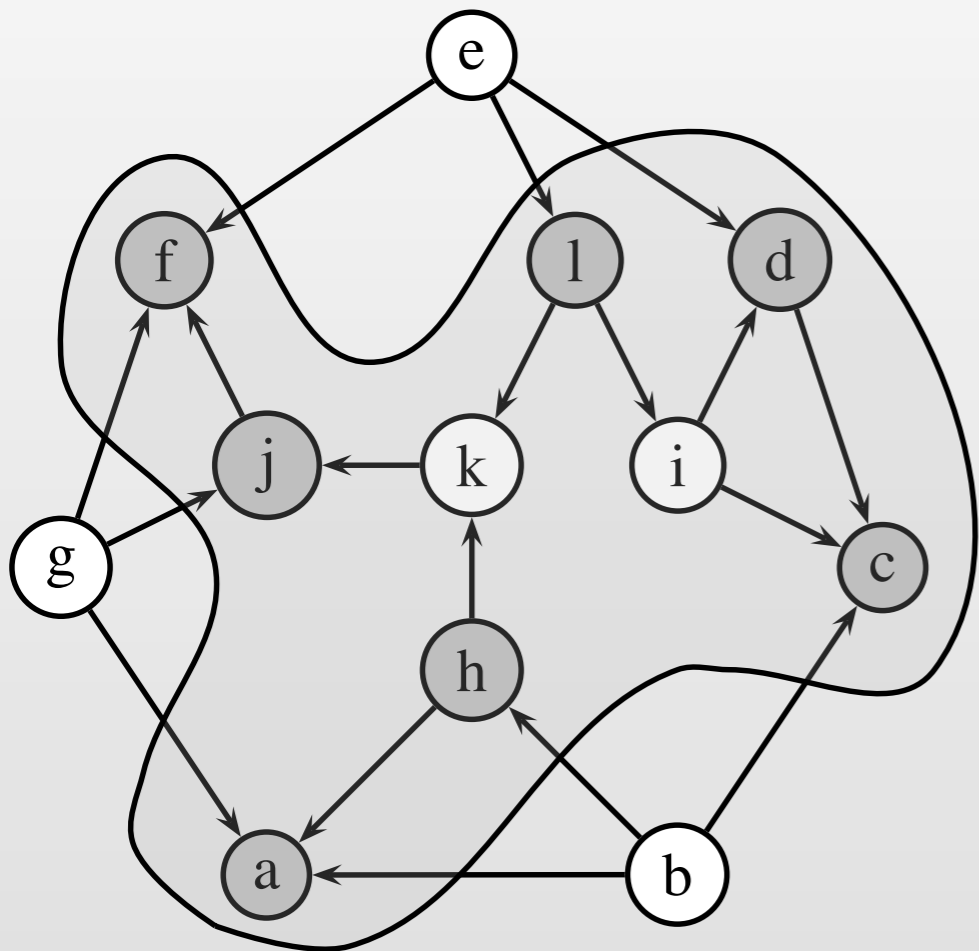
## Laplacian Mechanism

---

$$\models x \stackrel{\$}{\leftarrow} \mathcal{L}_\lambda(r) \sim_{\exp(\epsilon), 0} y \stackrel{\$}{\leftarrow} \mathcal{L}_\lambda(s) : |r\langle 1 \rangle - s\langle 2 \rangle| \leq \lambda\epsilon \Rightarrow x\langle 1 \rangle = y\langle 2 \rangle$$

# Application: Vertex Cover

Gupta et al. [SODA '10]



$\pi = [b, g, e, h, l, k,$

$\text{VertexCover}(V, E, \epsilon), \text{VertexCover}(V, E, \epsilon) :$

$$V\langle 1 \rangle = V\langle 2 \rangle \wedge E\langle 1 \rangle = E\langle 2 \rangle \cup \{(t, u)\} \implies \pi\langle 1 \rangle = \pi\langle 2 \rangle$$

$\text{VertexCover}(V, E, \epsilon)$

1  $\pi \leftarrow \text{nil}; n \leftarrow |V|; i \leftarrow 0;$

2 **while**  $E \neq \emptyset$  **do**

3  $v \leftarrow \text{pick}(V, E, \epsilon, n, i);$

4  $\pi \leftarrow v :: \pi;$

5  $V \leftarrow V \setminus \{v\}; E \leftarrow E \setminus (\{v\} \times V);$

6 **end**  $i \leftarrow i + 1$

7 **end**

$\text{pick}(V, E, \epsilon, n, i) \text{ deg}_E(v)$

Proven correct  
using CertiPriv

# Conclusions

- Framework for reasoning about quantitative relational properties of randomized computations
  - Laplacian and Exponential mechanisms
  - Differential privacy for streaming and graph algorithms
  - Asymmetric logic
- Further work:
  - Computational differential privacy
  - Hash functions onto elliptic curves and statistical zero-knowledge
- Challenge: logic for arbitrary quantitative relational properties

**Thanks for your attention!**

Define  $\alpha$ -distance as:

$$\Delta_\alpha(d_1, d_2) = \max_A(\max(d_1 \cdot 1_A - \alpha(d_2 \cdot 1_A), d_2 \cdot 1_A - \alpha(d_1 \cdot 1_A)))$$

$(\alpha, \delta)$ -lifting of relations to distributions:

$$\begin{aligned} \text{lift}_{\alpha, \delta} R (d_1 : \mathcal{D}_A) (d_2 : \mathcal{D}_B) &= \exists(d : \mathcal{D}_{A*B}), \\ &\pi_1(d) \leq d_1 \wedge \Delta_\alpha(\pi_1(d), d_1) \leq \delta \wedge \\ &\pi_2(d) \leq d_2 \wedge \Delta_\alpha(\pi_2(d), d_2) \leq \delta \wedge \text{range } R \ d \end{aligned}$$

Output perturbation makes numerical queries  $\epsilon$ -diff. private

- The  $\Phi$ -sensitivity of a query  $f : \mathcal{D} \rightarrow \mathbb{R}$  is defined as:

$$\Delta(f) = \max\{f(D_1) - f(D_2) \mid \Phi(D_1, D_2)\}$$

- The randomized computation

$$K(D) = f(D) + \text{Lap}(\Delta(f)/\epsilon)$$

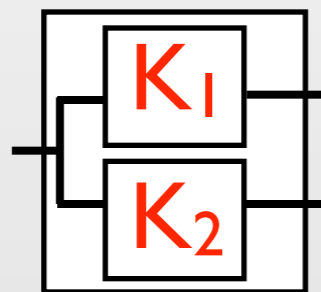
is  $\epsilon$ -differentially private

Density proportional to  
 $\exp(-\epsilon/\Delta(f))$

# Composition theorems

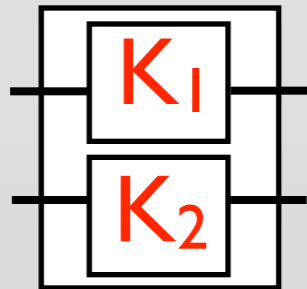
If  $K_1$  is  $(\epsilon_1, \delta_1)$ -diff. private and  $K_2$  is  $(\epsilon_2, \delta_2)$ -diff. private

- Sequential composition



$(\epsilon_1 + \epsilon_2, \delta_1 + \delta_2)$ -diff. private

- Parallel composition



$(\max\{\epsilon_1, \epsilon_2\}, \max\{\delta_1, \delta_2\})$ -diff. private

$K_1$  and  $K_2$  depend on disjoint parts of the database