Christoph Berkholz
Maximilian Merz

Probabilistic Databases Under Updates

HIGHLIGHTS 2021
Aachen / Online

PROBABILISTIC DB's ...

tuple-independent db. representation

R	x	P
	a	1/2
	b	1/3

possible worlds

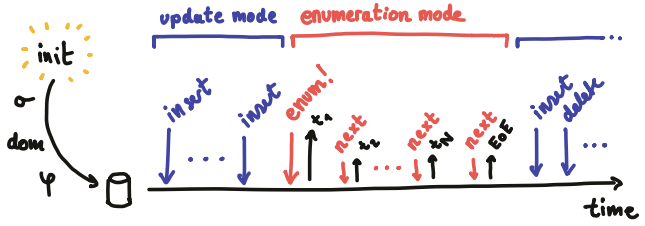
W_1: R(a), R(b) — $P_1 = 1/2 \cdot 1/3 = 1/6$ — $\varphi(W_1)$ ✓
W_2: R(a) — $P_2 = 1/2 \cdot \overline{2/3} = 2/6$ — $\varphi(W_2)$ ✓
W_3: R(b) — $P_3 = \overline{1/2} \cdot 1/3 = 1/6$ — $\varphi(W_3)$ ✓
W_4: ∅ — $P_4 = \overline{1/2} \cdot \overline{2/3} = 2/6$ — $\varphi(W_4)$ ×

query evaluation
$\varphi = \exists x : Rx$

$P = \sum_{W \models \varphi} P(W) = 1/6 + 2/6 + 1/6 = 4/6$

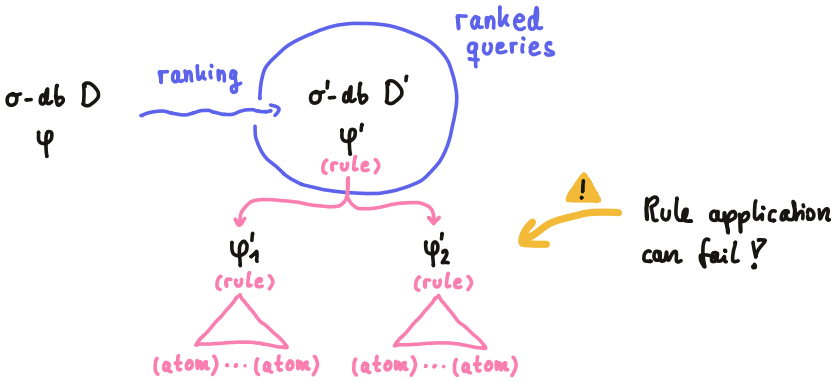
... UNDER UPDATES

fix db schema $\sigma = \{R_1, \dots, R_s\}$
fix finite domain $dom = \{a, b, c, \dots\}$
fix FO query $\varphi(\bar{x})$



BOOLEAN QUERY EVALUATION ...

For Boolean UCQ's φ:



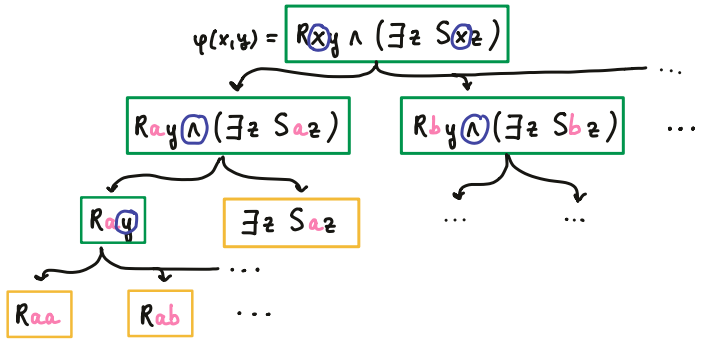
Dalvi, Suciu: EITHER △✓, THEN φ in P; OR ⚠, THEN: φ in #P

We showed:
1) We can pass updates through the ranking transformation.
2) Every single update affects only constant number of nodes.
3) Every single node can be updated in constant time.

↳ The rules by Dalvi & Suciu can be made dynamic!

... AND RANKED ENUMERATION

- Lifting Rules for Non-Boolean Queries:



- Non-repeating CQ's on probabilistic DB's.
  If the OMv-conjecture holds:

query is q-hierarchical ⟷ Dynamic Ranked Enumeration with $O(\log n)$ delay and $O(\log n)$ updates