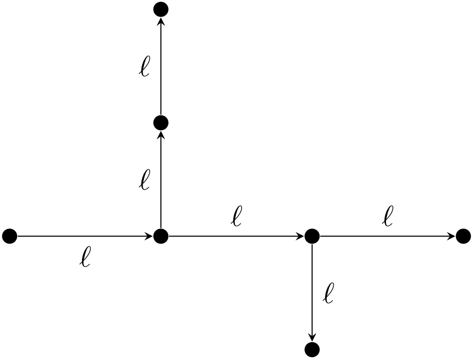
Nesting Depth of Operators in Graph Database Queries: Expressiveness vs. Evaluation Complexity

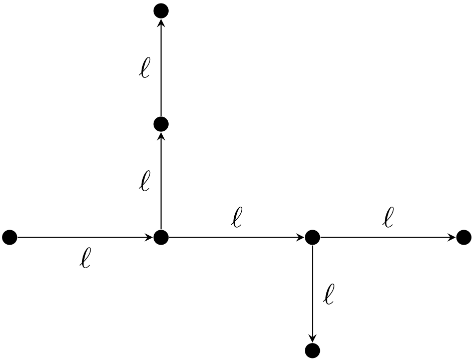
M. Praveen and B. Srivathsan

Chennai Mathematical Institute, India

Regular Path Queries

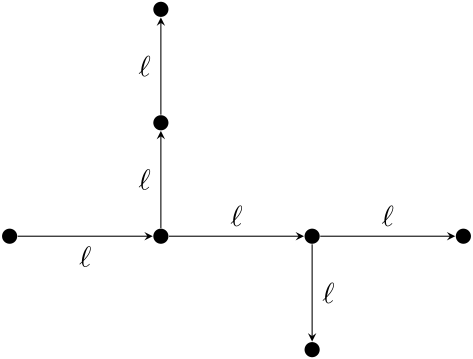


Regular Path Queries



A regular path query: ℓ^*

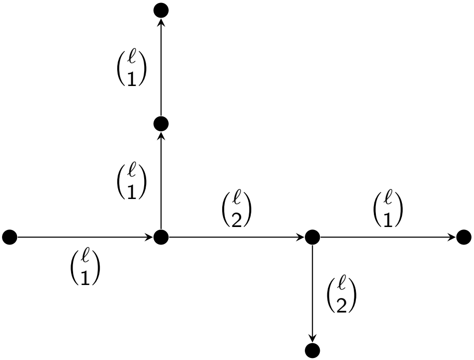
Regular Path Queries



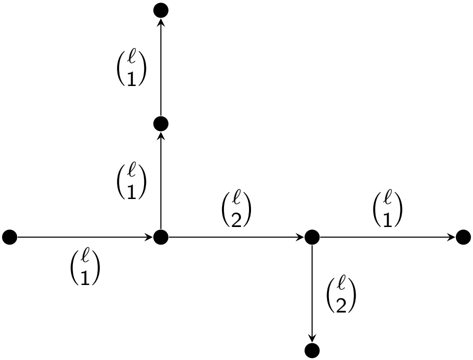
A regular path query: ℓ^*

$Q(G) = \{(u, v) \mid v \text{ is reachable from } u\}$

Regular Data Path Queries

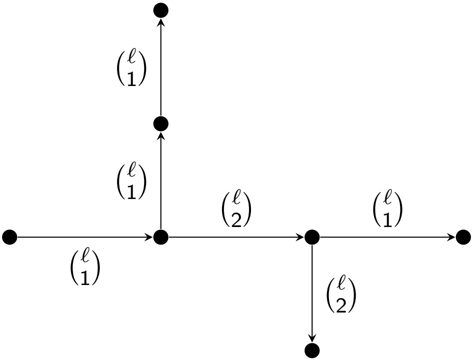


Regular Data Path Queries



[Barceló, Libkin, Reutter 2011] A parameterized RE: x^*

Regular Data Path Queries



[Barceló, Libkin, Reutter 2011] A parameterized RE: x^*

$$Q(G) = \{(u, v) \mid$$
v is reachable from u using links of common priority$\}$

Iterated Bindings

$$\ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ )$$

Iterated Bindings

$$\ell \downarrow_{x_1} (\; (\ell[x_1^=])^* \;)$$

$$\binom{\ell}{1}\binom{\ell}{1}\binom{\ell}{1}, \binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}$$

Iterated Bindings

$$\ell \downarrow_{x_1} (\ (\ell[x_1^=])^* \ )$$

$$\binom{\ell}{1}\binom{\ell}{1}\binom{\ell}{1}, \binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}$$

$$\ell \downarrow_{x_2} \ (\ \ell \downarrow_{x_1} (\ (\ell[x_1^=])^* \ ) \cdot \ell[x_2^=] \ )$$

Iterated Bindings

$$\ell \downarrow_{x_1} \left(\; (\ell[x_1^=])^* \; \right)$$

$$\binom{\ell}{1}\binom{\ell}{1}\binom{\ell}{1}, \binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}$$

$$\ell \downarrow_{x_2} \left(\; \ell \downarrow_{x_1} \left(\; (\ell[x_1^=])^* \; \right) \cdot \ell[x_2^=] \; \right)$$

$$\binom{\ell}{1}\boxed{\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}}\binom{\ell}{1}$$

Iterated Bindings

$$\ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ )$$

$$\binom{\ell}{1}\binom{\ell}{1}\binom{\ell}{1},\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}$$

$$\ell \downarrow_{x_2}\ (\ \ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ ) \cdot \ell[x_2^=]\ )$$

$$\binom{\ell}{1}\boxed{\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}}\binom{\ell}{1}$$

$$\ell \downarrow_{x_2}\ \{\ [\ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ )]^* \cdot \ell[x_2^=]\ \}$$

Iterated Bindings

$$\ell \downarrow_{x_1} \left((\ell[x_1^{=}])^* \right)$$

$$\binom{\ell}{1}\binom{\ell}{1}\binom{\ell}{1}, \binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}$$

$$\ell \downarrow_{x_2} \left(\ell \downarrow_{x_1} \left((\ell[x_1^{=}])^* \right) \cdot \ell[x_2^{=}] \right)$$

$$\binom{\ell}{1}\boxed{\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}}\binom{\ell}{1}$$

$$\ell \downarrow_{x_2} \left\{ [\ell \downarrow_{x_1} \left((\ell[x_1^{=}])^* \right)]^* \cdot \ell[x_2^{=}] \right\}$$

$$\binom{\ell}{1}\ \boxed{\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}}\ \boxed{\binom{\ell}{3}\binom{\ell}{3}\binom{\ell}{3}\binom{\ell}{3}}\binom{\ell}{1}$$

Iterated Bindings

$$\ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ )$$

$$\binom{\ell}{1}\binom{\ell}{1}\binom{\ell}{1}, \binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}$$

$$\ell \downarrow_{x_2}\ (\ \ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ ) \cdot \ell[x_2^=]\ )$$

$$\binom{\ell}{1}\boxed{\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}}\binom{\ell}{1}$$

$$\ell \downarrow_{x_2}\ \{\ [\ell \downarrow_{x_1} (\ (\ell[x_1^=])^*\ )]^* \cdot \ell[x_2^=]\ \}$$

$$\binom{\ell}{1}\ \boxed{\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}\binom{\ell}{2}}\ \boxed{\binom{\ell}{3}\binom{\ell}{3}\binom{\ell}{3}\binom{\ell}{3}}\binom{\ell}{1}$$

Regular Expressions With Binding [Libkin, Tan, Vrgoč 2013]

$$r ::= \epsilon \mid a \mid a[c] \mid r + r \mid r \cdot r \mid r^* \mid a \downarrow_x (r)$$

Evaluating Queries With Iterated Bindings is Harder

- Query evaluation problem: is $(u, v) \in Q(G)$?

Evaluating Queries With Iterated Bindings is Harder

- Query evaluation problem: is $(u, v) \in Q(G)$?

- Only with bindings that are never iterated: NP-complete.

Evaluating Queries With Iterated Bindings is Harder

- Query evaluation problem: is $(u, v) \in Q(G)$?

- Only with bindings that are never iterated: NP-complete.

- With bindings that are inside the scope of iterations,

Evaluating Queries With Iterated Bindings is Harder

- Query evaluation problem: is $(u, v) \in Q(G)$?
- Only with bindings that are never iterated: NP-complete.
- With bindings that are inside the scope of iterations, that are inside the scope of bindings,

Evaluating Queries With Iterated Bindings is Harder

- Query evaluation problem: is $(u, v) \in Q(G)$?
- Only with bindings that are never iterated: NP-complete.
- With bindings that are inside the scope of iterations, that are inside the scope of bindings, that are inside the scope of ...: $\textsc{Pspace}$-complete.

Nesting Depth of Iterated Bindings

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Nesting Depth of Iterated Bindings

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

$\mathbf{E_1}$: $a \downarrow_{x_1} (a[x_1^=]^*)$

Nesting Depth of Iterated Bindings

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

$\mathbf{E_1}$: $a \downarrow_{x_1} (a[x_1^{=}]^*),\ b \downarrow_{x_2} (a \downarrow_{x_1} (a[x_1^{=}]^*))$

Nesting Depth of Iterated Bindings

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

$$\mathbf{E_1} : \; a \downarrow_{x_1} (a[x_1^=]^*), \; b \downarrow_{x_2} (a \downarrow_{x_1} (a[x_1^=]^*))$$

$$\mathbf{F_1} : \; (a \downarrow_{x_1} (a[x_1^=]^*))^*$$

Nesting Depth of Iterated Bindings

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

$$\mathbf{E_1} : \; a \downarrow_{x_1} (a[x_1^=]^*), \; b \downarrow_{x_2} (a \downarrow_{x_1} (a[x_1^=]^*))$$

$$\mathbf{F_1} : \; (a \downarrow_{x_1} (a[x_1^=]^*))^*$$

$$\mathbf{E_2} : \; b \downarrow_{x_2} ((a \downarrow_{x_1} (a[x_1^=]^*))^*)$$

Results

1. The resulting hierarchy of data languages is strict, and so is the expressiveness of queries.

Results

1. The resulting hierarchy of data languages is strict, and so is the expressiveness of queries.

2. It is undecidable to check if a given REWB has a language equivalent one at lower levels.

Results

1. The resulting hierarchy of data languages is strict, and so is the expressiveness of queries.

2. It is undecidable to check if a given REWB has a language equivalent one at lower levels.

3. An REWB query in level i can be evaluated in Σ_i in the polynomial time hierarchy.

Results

1. The resulting hierarchy of data languages is strict, and so is the expressiveness of queries.
2. It is undecidable to check if a given REWB has a language equivalent one at lower levels.
3. An REWB query in level i can be evaluated in Σ_i in the polynomial time hierarchy.
4. Controlled reductions from satisfiability of quantified Boolean formulas to query evaluation.

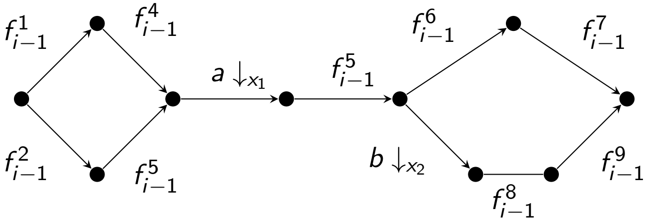
Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

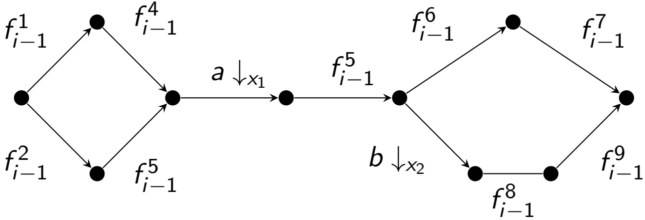
Automaton for e_i



Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for e_i

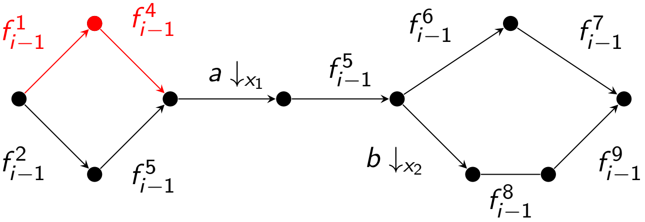


$$(f_{i-1}^1 f_{i-1}^4 + f_{i-1}^2 f_{i-1}^5) a \downarrow_{x_1} (f_{i-1}^5)(f_{i-1}^6 f_{i-1}^7 + b \downarrow_{x_2} (f_{i-1}^8 f_{i-1}^9))$$

Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for e_i

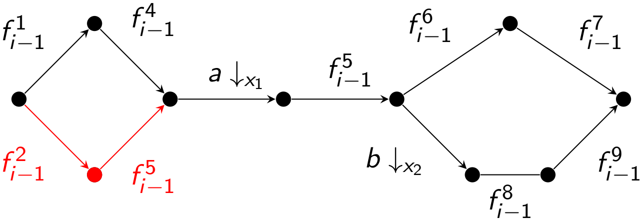


$$(f^1_{i-1} f^4_{i-1} + f^2_{i-1} f^5_{i-1}) a \downarrow_{x_1} (f^5_{i-1})(f^6_{i-1} f^7_{i-1} + b \downarrow_{x_2} (f^8_{i-1} f^9_{i-1}))$$

Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for e_i

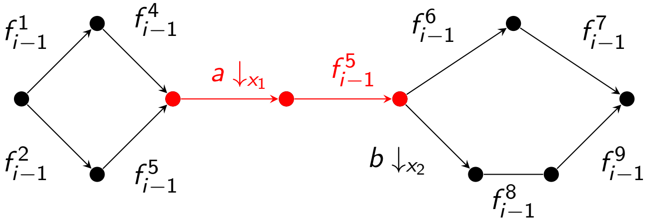


$$(f^1_{i-1} f^4_{i-1} + f^2_{i-1} f^5_{i-1}) a \downarrow_{x_1} (f^5_{i-1})(f^6_{i-1} f^7_{i-1} + b \downarrow_{x_2} (f^8_{i-1} f^9_{i-1}))$$

Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for e_i

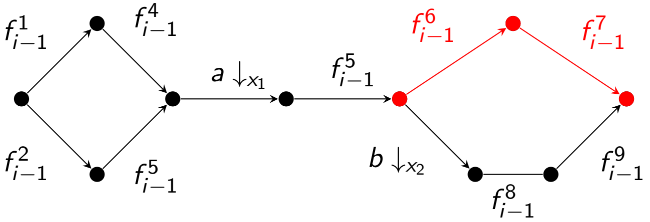


$$(f^1_{i-1} f^4_{i-1} + f^2_{i-1} f^5_{i-1}) a \downarrow_{x_1} (f^5_{i-1}) (f^6_{i-1} f^7_{i-1} + b \downarrow_{x_2} (f^8_{i-1} f^9_{i-1}))$$

Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for e_i



$$(f^1_{i-1} f^4_{i-1} + f^2_{i-1} f^5_{i-1}) a \downarrow_{x_1} (f^5_{i-1})(f^6_{i-1} f^7_{i-1} + b \downarrow_{x_2} (f^8_{i-1} f^9_{i-1}))$$

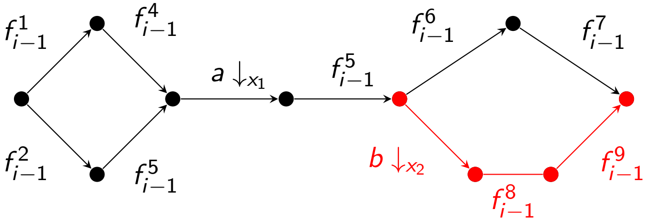
Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$

$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$

$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for e_i

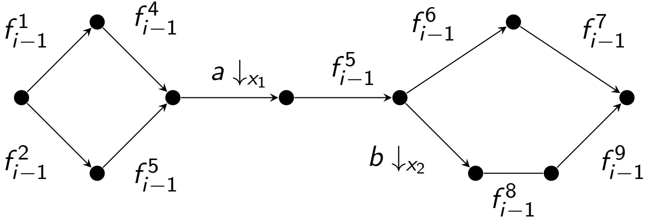


$$(f_{i-1}^1 f_{i-1}^4 + f_{i-1}^2 f_{i-1}^5) a \downarrow_{x_1} (f_{i-1}^5)(f_{i-1}^6 f_{i-1}^7 + b \downarrow_{x_2} (f_{i-1}^8 f_{i-1}^9))$$

Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

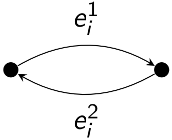
Automaton for e_i



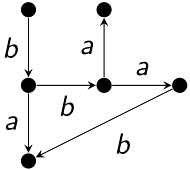
Automata View of Controlled REWBs

$$F_0 ::= \epsilon \mid a \mid a[c] \mid F_0 + F_0 \mid F_0 \cdot F_0 \mid F_0^*$$
$$E_i ::= F_{i-1} \mid E_i + E_i \mid E_i \cdot E_i \mid a \downarrow_{x_j} (E_i)$$
$$F_i ::= E_i \mid F_i + F_i \mid F_i \cdot F_i \mid F_i^*$$

Automaton for f_i

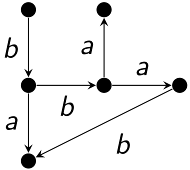


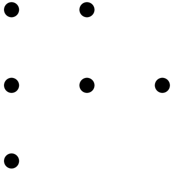
Evaluating F_i Queries Using Oracle for E_i



$$(e_i^1 e_i^2)^*$$

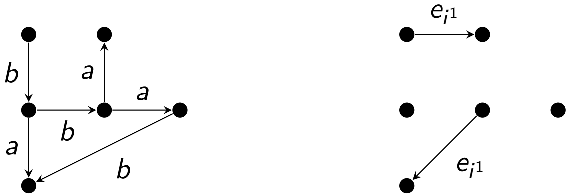
Evaluating F_i Queries Using Oracle for E_i





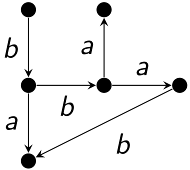
$(e_i^1 e_i^2)^*$

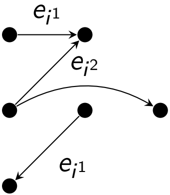
Evaluating F_i Queries Using Oracle for E_i



$(e_i^1 e_i^2)^*$

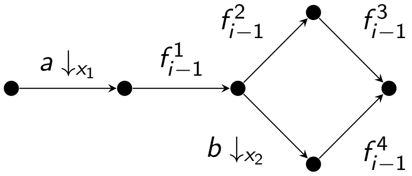
Evaluating F_i Queries Using Oracle for E_i

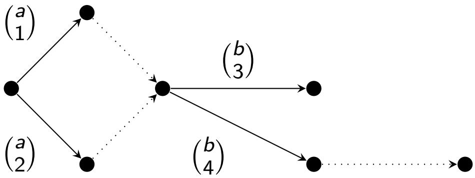




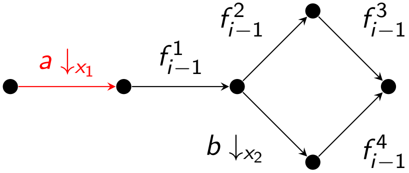
$$(e_i^1 e_i^2)^*$$

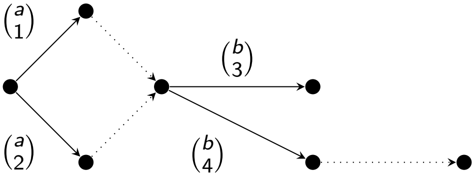
Evaluating E_i Queries Using Non-determinism



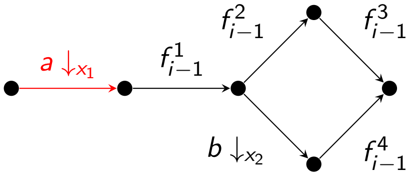


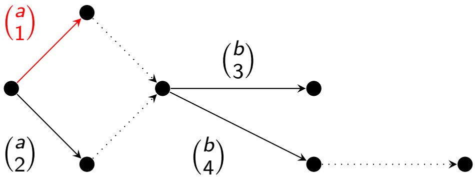
Evaluating E_i Queries Using Non-determinism



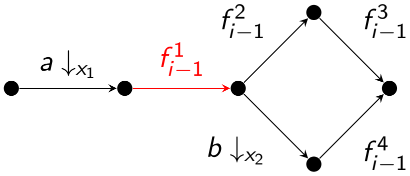


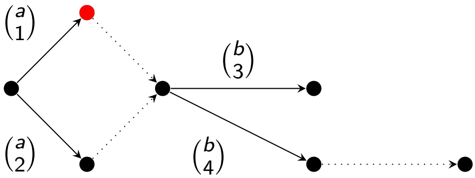
Evaluating E_i Queries Using Non-determinism



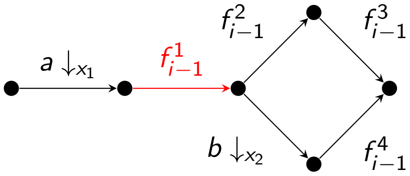


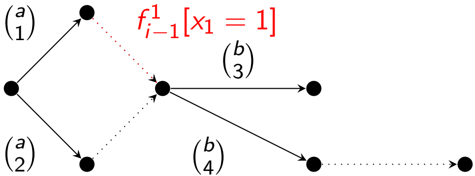
Evaluating E_i Queries Using Non-determinism



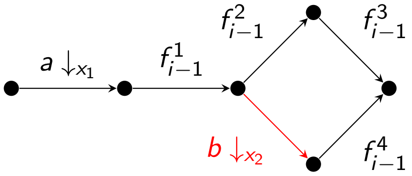


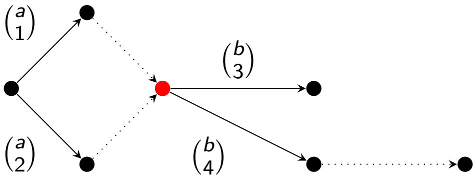
Evaluating E_i Queries Using Non-determinism



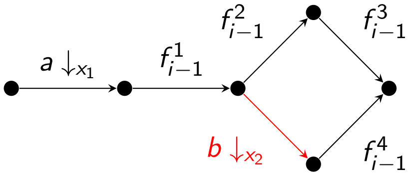


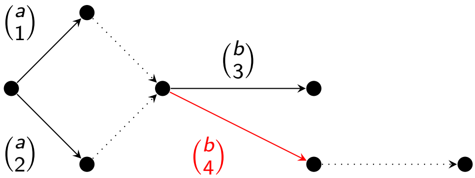
Evaluating E_i Queries Using Non-determinism



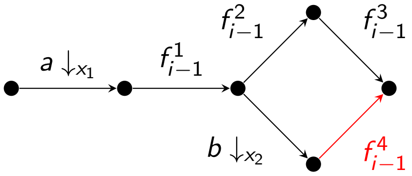


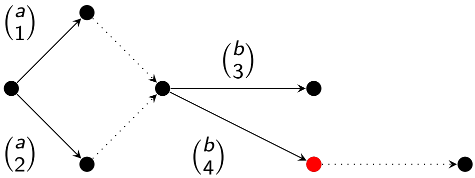
Evaluating E_i Queries Using Non-determinism



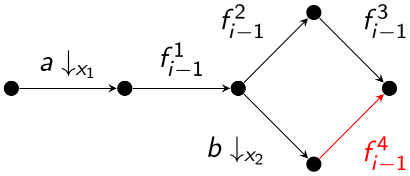


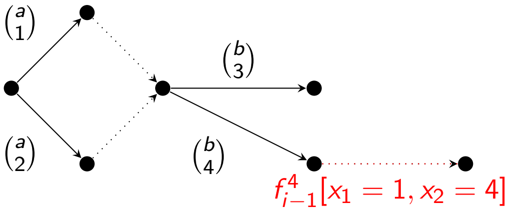
Evaluating E_i Queries Using Non-determinism





Evaluating E_i Queries Using Non-determinism





Complexity of Query Evaluation

- Queries in E_i can be evaluated in Σ_i.

Complexity of Query Evaluation

- Queries in E_i can be evaluated in Σ_i.
- What about lower bounds?

Gap in Lower Bounds

- Satisfiability of $\exists\ \overline{x_1}\ \forall\ \overline{x_2}\ \phi$,

Gap in Lower Bounds

- Satisfiability of $\exists\ \overline{x_1}\ \forall\ \overline{x_2}\ \phi$, can be reduced to query evaluation,

Gap in Lower Bounds

- Satisfiability of $\exists\ \overline{x_1}\ \forall\ \overline{x_2}\ \phi$, can be reduced to query evaluation, where the number of alternations in the query

Gap in Lower Bounds

- Satisfiability of $\exists \; \overline{x_1} \; \forall \; \overline{x_2} \; \phi$, can be reduced to query evaluation, where the number of alternations in the query is the number of variables in $\overline{x_2}$.

Gap in Lower Bounds

- Satisfiability of $\exists\ \overline{x_1}\ \forall\ \overline{x_2}\ \phi$, can be reduced to query evaluation, where the number of alternations in the query is the number of variables in $\overline{x_2}$.

- Satisfiability of $\exists^{k_1/n_1}\ \forall^{k_2/n_2}\ \exists^{k_3/n_3} \ldots \phi$

Gap in Lower Bounds

- Satisfiability of $\exists \, \overline{x_1} \, \forall \, \overline{x_2} \, \phi$, can be reduced to query evaluation, where the number of alternations in the query is the number of variables in $\overline{x_2}$.

- Satisfiability of $\exists^{k_1/n_1} \, \forall^{k_2/n_2} \, \exists^{k_3/n_3} \cdots \phi$ can be reduced to query evaluation where the number of alternations in the query

Gap in Lower Bounds

- Satisfiability of $\exists\ \overline{x_1}\ \forall\ \overline{x_2}\ \phi$, can be reduced to query evaluation, where the number of alternations in the query is the number of variables in $\overline{x_2}$.

- Satisfiability of $\exists^{k_1/n_1}\ \forall^{k_2/n_2}\ \exists^{k_3/n_3} \cdots \phi$ can be reduced to query evaluation where the number of alternations in the query is $k_2 + k_4 + \cdots$.

Summary

- Nesting iterated bindings increase expressiveness.

Summary

- Nesting iterated bindings increase expressiveness.
- Undecidable to check if a given expression has a language equivalent one at lower levels.

Summary

- Nesting iterated bindings increase expressiveness.
- Undecidable to check if a given expression has a language equivalent one at lower levels.

Query level	Evaluation	Parameterized by query size
E_1	NP-c	W[SAT] - W[P]
E_i, $i > 1$	? - Σ_i	?
Unbounded	PSPACE-c	AW[SAT] - uniform-XNL

Summary

- Nesting iterated bindings increase expressiveness.
- Undecidable to check if a given expression has a language equivalent one at lower levels.

Query level	Evaluation	Parameterized by query size
E_1	NP-c	W[SAT] - W[P]
E_i, $i > 1$	? - Σ_i	?
Unbounded	PSPACE-c	AW[SAT] - uniform-XNL

Thank You