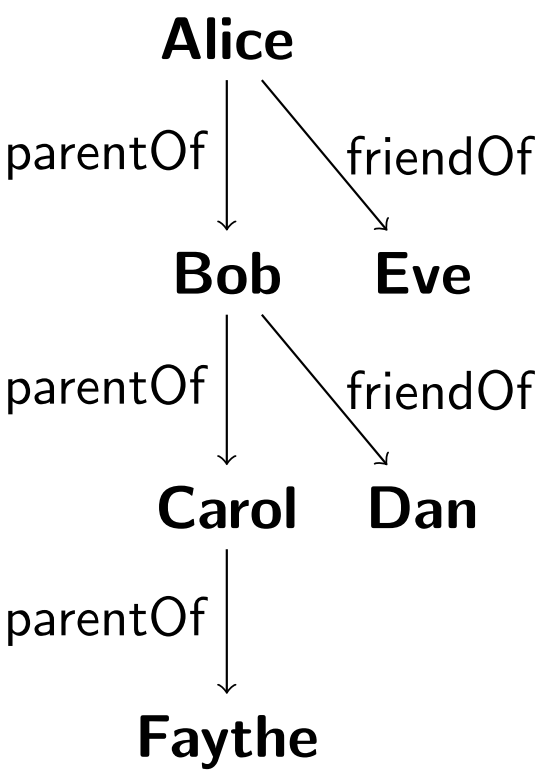
Graph query optimization using semi-join rewritings

Jelle Hellings[1]
jelle.hellins@uhasselt.be

Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium

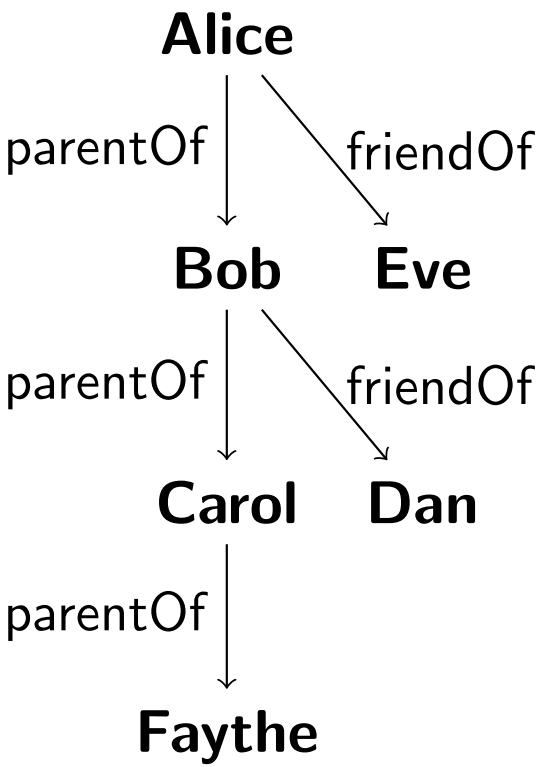
[1] Joint work with Catherine L. Pilachowski, Dirk Van Gucht, Marc Gyssens, and Yuqing Wu.

Graphs and Graph Querying



Query: 'Great-grandparents and their friends'

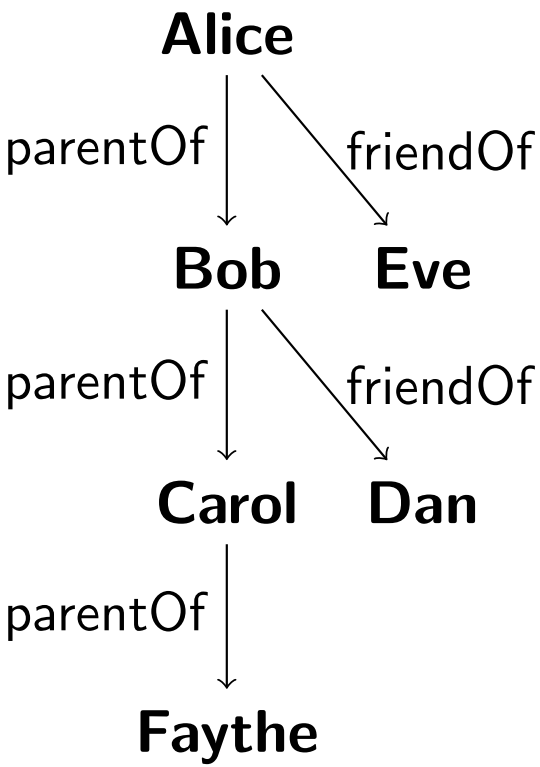
Graphs and Graph Querying



Query: 'Great-grandparents and their friends'

- (Great-grandparents, descendant):
  $\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}$

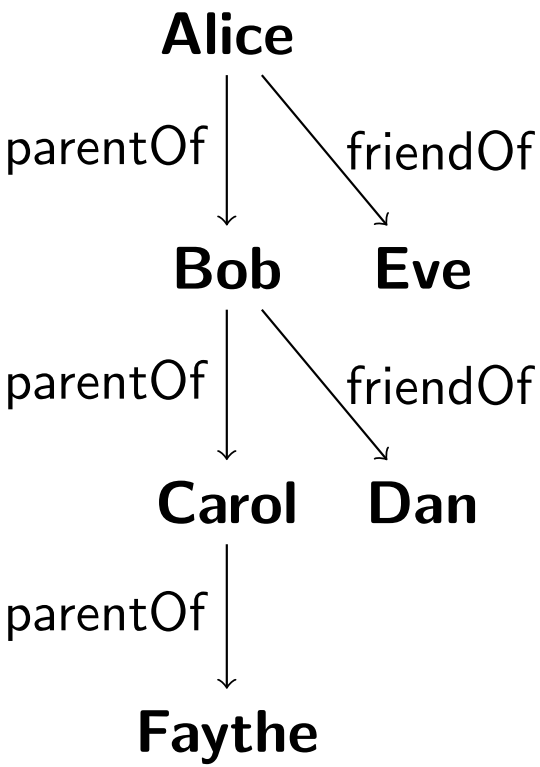
Graphs and Graph Querying



Query: 'Great-grandparents and their friends'

- (Great-grandparents, descendant):
  $\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}$
- Great-grandparents:
  $\pi_1[\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}]$

Graphs and Graph Querying



Query: 'Great-grandparents and their friends'

- (Great-grandparents, descendant):
  $\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}$
- Great-grandparents:
  $\pi_1[\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}]$
- Complete query:
  $\pi_1[\text{parentOf} \circ \cdots \circ \text{parentOf}] \circ \text{friendOf}$

Graph Query Language

$$\underline{\mathrm{id}} \mid \mathrm{di} \mid \underline{\ell} \mid \ell^{\frown} \mid \pi_j[e] \mid \overline{\pi}_j[e] \mid \underline{e \circ e} \mid \underline{e \cup e} \mid e \cap e \mid e - e \mid \underline{[e]^*}$$

- Regular Path Queries

Graph Query Language

$$\underline{\mathrm{id}} \mid \mathrm{di} \mid \underline{\ell} \mid \underline{\ell^{\frown}} \mid \underline{\pi_j[e]} \mid \overline{\pi}_j[e] \mid \underline{e \circ e} \mid \underline{e \cup e} \mid e \cap e \mid e - e \mid \underline{[e]^*}$$

- Regular Path Queries
- Nested Regular Path Queries

Graph Query Language

$$\underline{\mathrm{id}} \mid \mathrm{di} \mid \underline{\ell} \mid \underline{\ell^{\frown}} \mid \underline{\pi_j[e]} \mid \overline{\pi}_j[e] \mid \underline{e \circ e} \mid \underline{e \cup e} \mid e \cap e \mid e - e \mid \underline{[e]^*}$$

- Regular Path Queries
- Nested Regular Path Queries
- FO[3] augmented with transitive closure:

  graph-navigational core of XPath, GXPath, SPARQL, ...

Query Evaluation

$$\mathrm{id} \mid \mathrm{di} \mid \ell \mid \ell^{\frown} \mid \pi_j[e] \mid \overline{\pi}_j[e] \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^*$$

Query Evaluation

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^{\frown}} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid e \circ e \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid [e]^*$$

- 'Easy to evaluate'

Query Evaluation

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^{\frown}} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid \underline{e \circ e} \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid \underline{[e]^*}$$

- 'Easy to evaluate'
- 'Expensive to evaluate'

Query Evaluation

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^{\frown}} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid \underline{e \circ e} \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid \underline{[e]^*}$$

- 'Easy to evaluate'
- 'Expensive to evaluate'

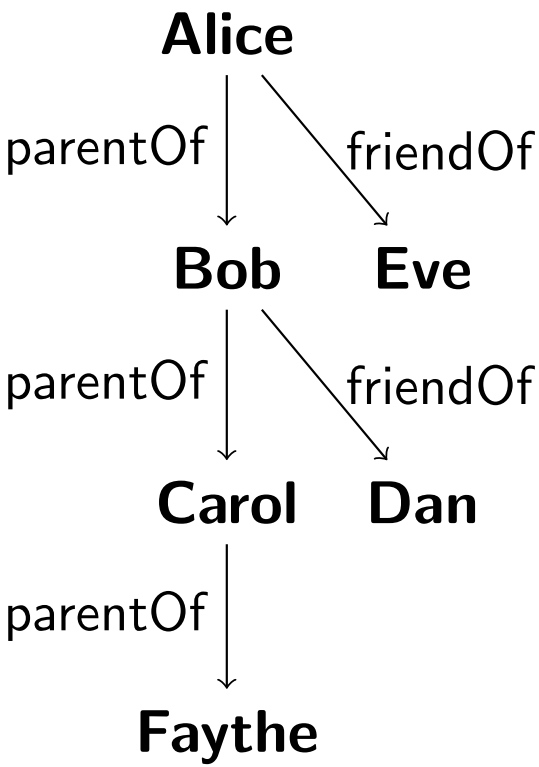
Idea: add partial alternatives for $\circ$ and $[\cdot]^*$

$$\pi_1[\mathsf{parentOf} \circ \mathsf{parentOf} \circ \mathsf{parentOf}] \circ \mathsf{friendOf}$$

can be rewritten into

$$\pi_1[\mathsf{parentOf} \ltimes (\mathsf{parentOf} \ltimes \mathsf{parentOf})] \rtimes \mathsf{friendOf}.$$

Query Optimization by rewriting?



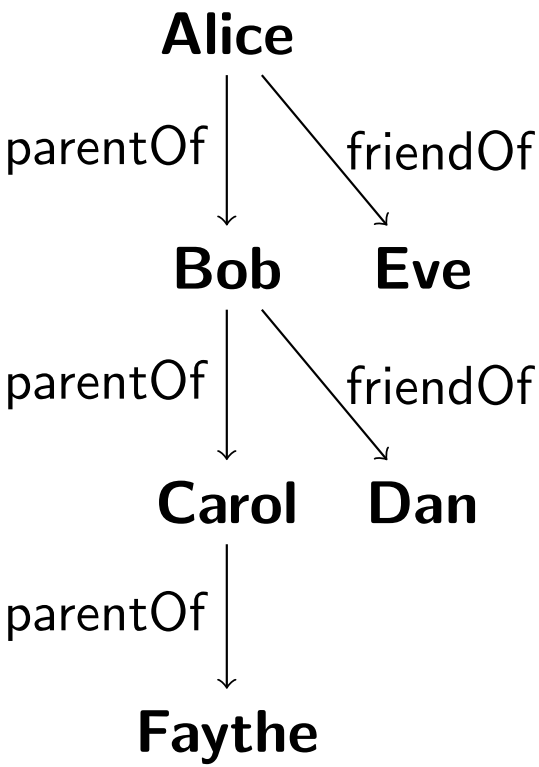
Problem

$$\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}$$

is not equivalent to

$$\text{parentOf} \ltimes (\text{parentOf} \ltimes \text{parentOf}).$$

Query Optimization by rewriting?



Problem

$$\text{parentOf} \circ \text{parentOf} \circ \text{parentOf}$$

is not equivalent to

$$\text{parentOf} \ltimes (\text{parentOf} \ltimes \text{parentOf}).$$

Solution

j-test-equivalent rewriting: we have $e_1 \equiv_j e_2$, if, for every graph $\mathcal{G}$,

$$\pi_j[e_1]\langle\mathcal{G}\rangle = \pi_j[e_2]\langle\mathcal{G}\rangle.$$

Rewrite composition and transitive closure

- Rewrite $\circ$ into $\ltimes$ and $\rtimes$
- Rewrite $[\cdot]^*$ into $\mathsf{fp}_{j,\mathfrak{N}}[\cdot;\ \cdot]$ (fixpoint iteration)

$$\mathrm{id} \mid \mathrm{di} \mid \ell \mid \ell^{\frown} \mid \pi_j[e] \mid \overline{\pi}_j[e] \mid e \circ e \mid e \cup e \mid e \cap e \mid e - e \mid [e]^* \mid$$
$$e \ltimes e \mid e \rtimes e \mid \mathfrak{N} \mid \mathsf{fp}_{j,\mathfrak{N}}[e;\ e]$$

Rewrite composition and transitive closure

- Rewrite $\circ$ into $\ltimes$ and $\rtimes$
- Rewrite $[\cdot]^*$ into $\mathsf{fp}_{j,\mathfrak{N}}[\cdot;\ \cdot]$ (fixpoint iteration)

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^\frown} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid e \circ e \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid [e]^* \mid$$
$$\underline{e \ltimes e} \mid \underline{e \rtimes e} \mid \mathfrak{N} \mid \mathsf{fp}_{j,\mathfrak{N}}[e;\ e]$$

Analysis

- FO[2]

Rewrite composition and transitive closure

- Rewrite $\circ$ into $\ltimes$ and $\rtimes$
- Rewrite $[\cdot]^*$ into $\mathsf{fp}_{j,\mathfrak{N}}[\cdot;\ \cdot]$ (fixpoint iteration)

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^\frown} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid e \circ e \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid [e]^* \mid$$
$$\underline{e \ltimes e} \mid \underline{e \rtimes e} \mid \underline{\mathfrak{N}} \mid \underline{\mathsf{fp}_{j,\mathfrak{N}}[e;\ e]}$$

Analysis

- FO[2] and FO[2]-like recursion
- For j-test-equivalent rewriting: only restrictions on $\cap$ and $-$

Rewrite composition and transitive closure

- Rewrite $\circ$ into $\ltimes$ and $\rtimes$
- Rewrite $[\cdot]^*$ into $\mathsf{fp}_{j,\mathfrak{N}}[\cdot;\ \cdot]$ (fixpoint iteration)

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^\frown} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid e \circ e \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid [e]^* \mid$$
$$\underline{e \ltimes e} \mid \underline{e \rtimes e} \mid \underline{\mathfrak{N}} \mid \underline{\mathsf{fp}_{j,\mathfrak{N}}[e;\ e]}$$

Analysis

- FO[2] and FO[2]-like recursion
- For j-test-equivalent rewriting: only restrictions on $\cap$ and $-$
- Rewriting is sound and 'complete'

Rewrite composition and transitive closure

- Rewrite $\circ$ into $\ltimes$ and $\rtimes$
- Rewrite $[\cdot]^*$ into $\mathsf{fp}_{j,\mathfrak{N}}[\cdot;\ \cdot]$ (fixpoint iteration)

$$\underline{\mathrm{id}} \mid \underline{\mathrm{di}} \mid \underline{\ell} \mid \underline{\ell^\frown} \mid \underline{\pi_j[e]} \mid \underline{\overline{\pi}_j[e]} \mid e \circ e \mid \underline{e \cup e} \mid \underline{e \cap e} \mid \underline{e - e} \mid [e]^* \mid$$
$$\underline{e \ltimes e} \mid \underline{e \rtimes e} \mid \underline{\mathfrak{N}} \mid \underline{\mathsf{fp}_{j,\mathfrak{N}}[e;\ e]}$$

Analysis

- FO[2] and FO[2]-like recursion
- For j-test-equivalent rewriting: only restrictions on $\cap$ and $-$
- Rewriting is sound and 'complete'
- Rewriting results in a 'small' query: number of steps needed to evaluate the result is twice the length of the original query

Future Work

- Study (small extensions of) FO[2] in more detail
- Further query optimization using information on the data
- Apply similar techniques to relational databases (SQL)

Fixpoints and transitive closure (example)

The transitive closure query

$$\pi_1[[\text{parentOf} \circ \overline{\pi}_1[\text{researcherAt}]]^* \circ \text{ownsPet}]$$

Fixpoints and transitive closure (example)

The transitive closure query

$$\pi_1[[\mathsf{parentOf} \circ \overline{\pi}_1[\mathsf{researcherAt}]]^* \circ \mathsf{ownsPet}]$$

is equivalent to the FO[2]-like query

$$\mathsf{fp}_{1,\mathfrak{N}}[\mathsf{parentOf} \ltimes \overline{\pi}_1[\mathsf{researcherAt}] \ltimes \mathfrak{N};\ \mathsf{ownsPet}].$$