FO queries strongly distributing over components

F. Di Cosmo

CILC
Trieste, 20 June 2019

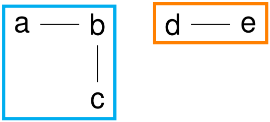
Connected Components

Connected Components

$$D = \{P(a, b),\\ R(b, b),\\ P(b, c),\\ P(d, e)\}$$

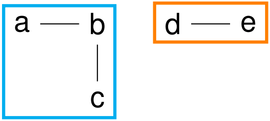
Connected Components

$$D = \{P(a,b),\\ R(b,b),\\ P(b,c),\\ P(d,e)\}$$



Connected Components

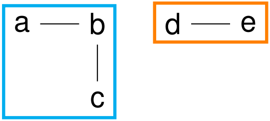
$$D = \{P(a,b),\ R(b,b),\ P(b,c),\ P(d,e)\}$$



$$C_1 = \{P(a,b), R(b,b), P(b,c)\} \quad C_2 = \{P(d,e)\}$$
$$\mathrm{cc}\,(D) = \{C_1, C_2\}$$

Connected Components

$$D = \{P(a,b), R(b,b), P(b,c), P(d,e)\}$$



$$C_1 = \{P(a,b), R(b,b), P(b,c)\} \quad C_2 = \{P(d,e)\}$$

$$\text{cc}\,(D) = \{C_1, C_2\}$$

Definition

Two facts are **directly connected**, $\sim_d$, if they share some individual.
The **connection relation**, $\sim$, is the transitive closure of $\sim_d$.
The **set of connected components** of D, **cc**(D), is the set of equivalence classes of $\sim$.

Distributed strategy

Distributed strategy

1) Each node hosts a local database;

Distributed strategy

1) Each node hosts a local database;
2) Local databases are closed by connectivity: each node hosts some connected components, say one;

Distributed strategy

1) Each node hosts a local database;
2) Local databases are closed by connectivity: each node hosts some connected components, say one;
3) The query is computed on each node;

Distributed strategy

1) Each node hosts a local database;
2) Local databases are closed by connectivity: each node hosts some connected components, say one;
3) The query is computed on each node;
4) The answers sets are merged, obtaining $\bigcup_{C \in \mathrm{cc}(D)} q(C)$.

Distributed strategy

1) Each node hosts a local database;
2) Local databases are closed by connectivity: each node hosts some connected components, say one;
3) The query is computed on each node;
4) The answers sets are merged, obtaining $\bigcup_{C \in \mathbf{CC}(D)} q(C)$.

Definition

A query q **distributes** if, for each database D:

$$q(D) = \bigcup_{C \in \mathbf{CC}(D)} q(C)$$

Main question

Which queries distribute?

Which queries distribute?

Results for Datalog$\neg$ (Ameloot 2017), exploiting recursion:

Which queries distribute?

Results for Datalog¬ (Ameloot 2017), exploiting recursion:

Syntactic Characterization

A query distributes iff it has a connected specification.

$E(x,z) \leftarrow R(x,y), P(z,w), \neg R(x,z)$ is not connected.

$E(x,z) \leftarrow R(x,y), P(y,z), \quad R(x,z)$ is connected.

Which queries distribute?

Results for Datalog¬ (Ameloot 2017), exploiting recursion:

Syntactic Characterization

A query distributes iff it has a connected specification.

$E(x,z) \leftarrow R(x,y), P(z,w), \neg R(x,z)$ is not connected.

$E(x,z) \leftarrow R(x,y), P(y,z), \quad R(x,z)$ is connected.

Decision Problem

To establish whether a given query distributes is undecidable.

Which queries distribute?

Results for Datalog¬ (Ameloot 2017), exploiting recursion:

Syntactic Characterization

A query distributes iff it has a connected specification.

$E(x,z) \leftarrow R(x,y), P(z,w), \neg R(x,z)$ is not connected.

$E(x,z) \leftarrow R(x,y), P(y,z), \quad R(x,z)$ is connected.

Decision Problem

To establish whether a given query distributes is undecidable.

What if recursion is omitted?

Preliminaries

Preliminaries

Language: NR Datalog$\neg$ $\rightsquigarrow$ FO relational queries.

Given $\varphi(\mathbf{x})$ and S, $q_\varphi(S) = \{\mathbf{a} | S \models \varphi(\mathbf{a})\}$

Preliminaries

Language: NR Datalog$\neg \rightsquigarrow$ FO relational queries.

Given $\varphi(\mathbf{x})$ and S, $q_\varphi(S) = \{\mathbf{a} | S \models \varphi(\mathbf{a})\}$

Structures: Databases $\rightsquigarrow$ FO relational structures.

$$\mathcal{L} = \{P/2, R/2\} \quad S = \left(\mathrm{dom}(S), P^S, R^S\right)$$

$$\mathrm{dom}(S) = \{a, b, c\}, P^S = \{(a, b), (b, c), (d, e)\}, R^A = \{(d, e)\}$$

Preliminaries

Language: NR Datalog$\neg \rightsquigarrow$ FO relational queries.

Given $\varphi(\mathbf{x})$ and S, $q_\varphi(S) = \{\mathbf{a} | S \models \varphi(\mathbf{a})\}$

Structures: Databases $\rightsquigarrow$ FO relational structures.

$$\mathcal{L} = \{P/2, R/2\} \quad S = \left(\text{dom}(S), P^S, R^S\right)$$

$$\text{dom}(S) = \{a, b, c\}, P^S = \{(a, b), (b, c), (d, e)\}, R^A = \{(d, e)\}$$

Problem: Distribution $\rightsquigarrow$ Strong distribution.

$$q(S) = \bigcup_{S' \subset S} q(S') = \bigcup_{C \in \mathbf{CC}(S)} q(C)$$

Preliminaries

Language: NR Datalog$\neg \rightsquigarrow$ FO relational queries.

Given $\varphi(\mathbf{x})$ and S, $q_\varphi(S) = \{\mathbf{a} | S \models \varphi(\mathbf{a})\}$

Structures: Databases $\rightsquigarrow$ FO relational structures.

$$\mathcal{L} = \{P/2, R/2\} \quad S = \left(\text{dom}(S), P^S, R^S\right)$$

$$\text{dom}(S) = \{a, b, c\}, P^S = \{(a,b),(b,c),(d,e)\}, R^A = \{(d,e)\}$$

Problem: Distribution $\rightsquigarrow$ Strong distribution.

$$q(S) = \bigcup_{S' \subset S} q(S') = \bigcup_{C \in \mathbf{CC}(S)} q(C)$$

- Monotony: $S' \subset S \Rightarrow q(S') \subset q(S)$
- Locality: $\mathbf{a} \in q(S) \Rightarrow \exists C \in \text{cc}(S)\ \mathbf{a} \in q(C)$

Monotony

Monotony

Theorem (Łoś - Tarski)

An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).

Monotony

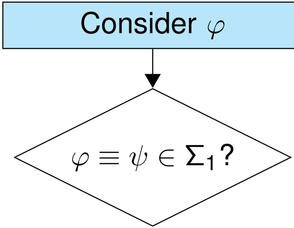
Theorem (Łoś - Tarski)

An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).

Consider φ

Theorem (Łoś - Tarski)

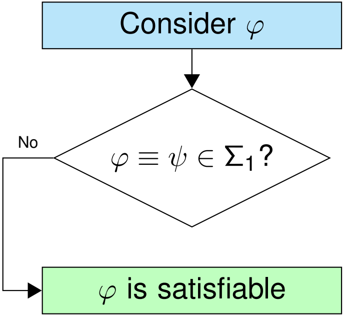
An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).



Monotony

Theorem (Łoś - Tarski)

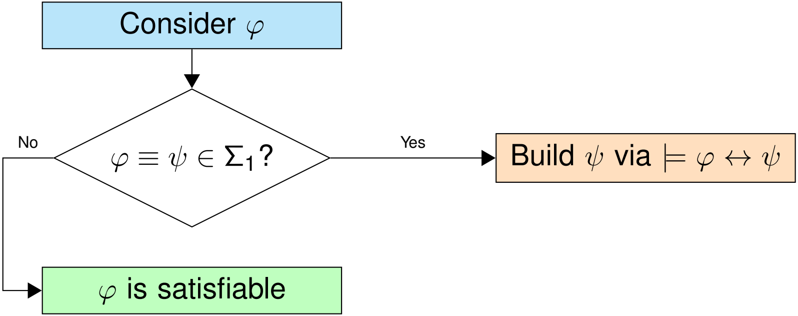
An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).



Monotony

Theorem (Łoś - Tarski)

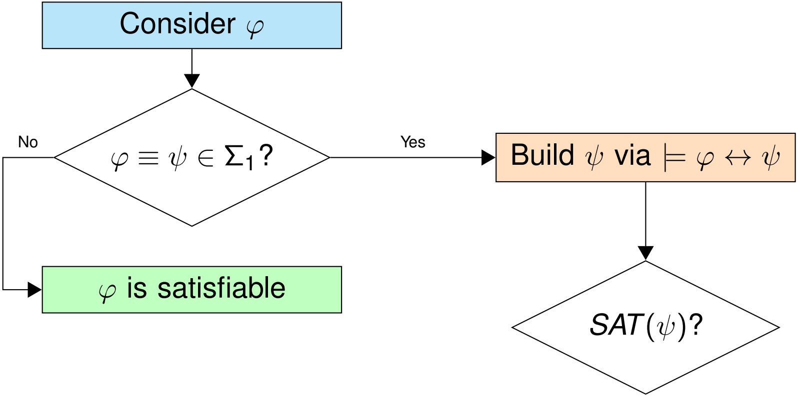
An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).



Monotony

Theorem (Łoś - Tarski)

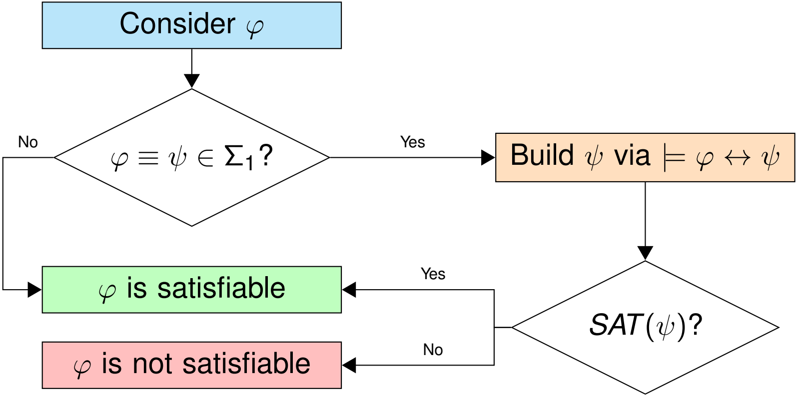
An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).



Monotony

Theorem (Łoś - Tarski)

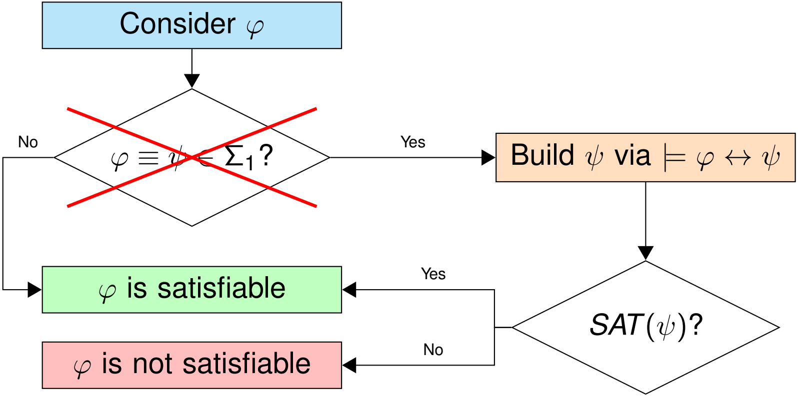
An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).



Monotony

Theorem (Łoś - Tarski)

An FO formula φ is preserved under superstructures (q_φ is monotone) iff φ is equivalent to a $\psi \in \Sigma_1$ (existential formulas).



Locality

Locality

No safeness... no locality.

Locality

No safeness... no locality.

Theorem

A $\bigwedge_{i \in I} L_i$ *is local iff:*

- *it is contradictory or*
- *all literals are negated and there is only one variable or*
- *each varaible in a negated literal occurs also in an asserted one (safeness).*

No safeness... no locality.

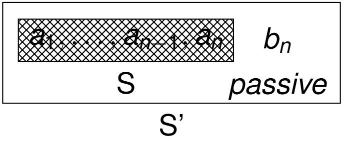
Theorem

A $\bigwedge_{i \in I} L_i$ is local iff:

- *it is contradictory or*
- *all literals are negated and there is only one variable or*
- *each varaible in a negated literal occurs also in an asserted one (safeness).*

$\varphi(x_1, \ldots, x_n)$

x_n *only in* $\neg A$



Conclusions

- Undecidability.
- Safeness (*locality*).
- Strong distribution vs *Simple* distribution.
- Constraints on structures.

Essential bibliography

Ameelot, Tom J. et al.
Datalog queries distributing over components
ACM TOCL, 2017.

Berger, Gerald and Pieris, Andreas
Ontology mediated queries distributing over components
IJCAI, 2016.