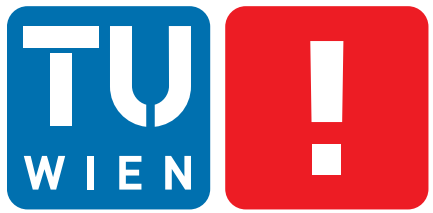
Contrasting RDF Stream Processing Semantics

Minh Dao-Tran Harald Beck Thomas Eiter

The 5th Join International Semantic Technology Conference

November 11-13, 2015



FAKULTÄT
FÜR !NFORMATIK

Faculty of Informatics

Outline

Illustrating Scenario

Divergence in RDF Stream Processing (RSP)

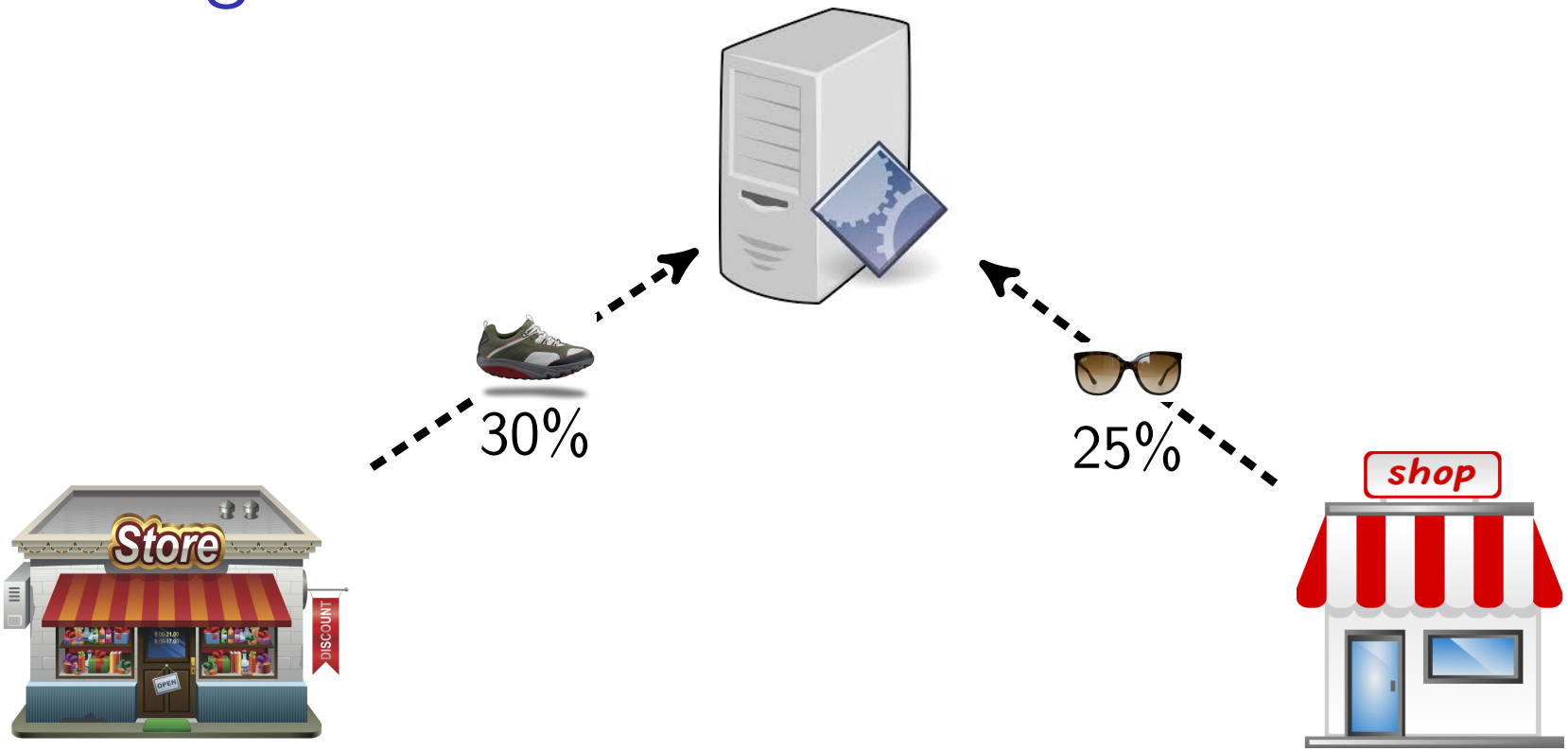
A **L**ogic for **A**nalyzing **R**easoning over **S**treams (LARS [Beck *et al.*, 2015])

Capturing and Comparing C-SPARQL and CQELS with LARS

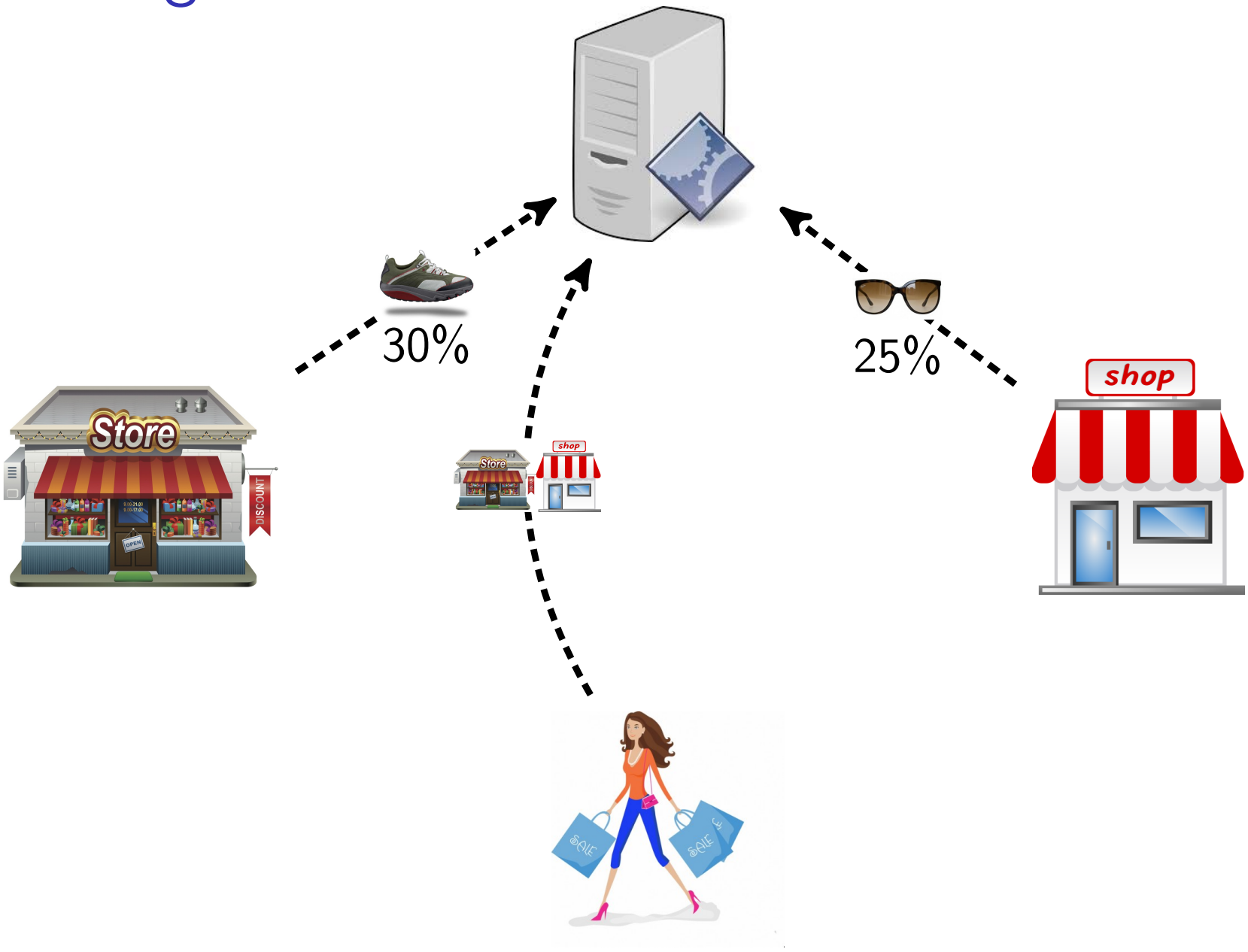
Illustrating Scenario

Illustrating Scenario

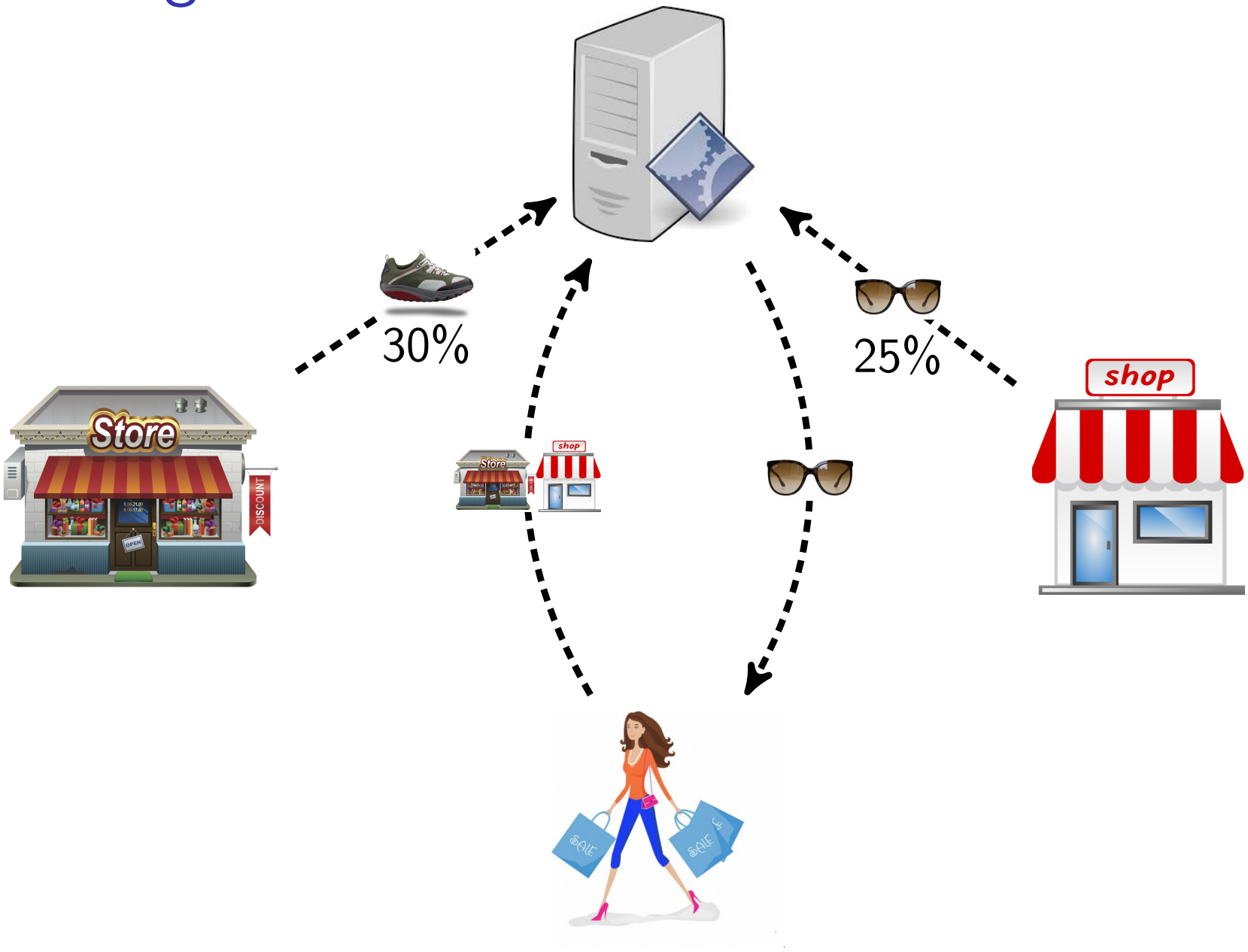
Illustrating Scenario



Illustrating Scenario



Illustrating Scenario



RDF Stream Processing at a Glance

Extend RDF data model with the temporal aspect

- RDF streams = RDF triples/graphs + timestamps

Extend SPARQL with window operators

- Adapt CQL [Arasu *et al.*, 2006] to SPARQL

Recent Developments in RSP

RSP Query Engines:

- C-SPARQL [Barbieri *et al.*, 2010]
- CQELS [Phuoc *et al.*, 2011]
- SPARQL_{Stream} [Calbimonte *et al.*, 2010]
- ...

Benchmarking systems:

- LSBench [Phuoc *et al.*, 2012]
- SRBench [Zhang *et al.*, 2012]
- CSRBench [Dell'Aglio *et al.*, 2013]
- YABench [Kolchin and Wetz, 2015]

Recent Developments in RSP

RSP Query Engines:

- C-SPARQL [Barbieri *et al.*, 2010]
- CQELS [Phuoc *et al.*, 2011]
- SPARQL_{Stream} [Calbimonte *et al.*, 2010]
- ...

Benchmarking systems:

- LSBench [Phuoc *et al.*, 2012]
- SRBench [Zhang *et al.*, 2012]
- CSRBench [Dell'Aglio *et al.*, 2013]
- YABench [Kolchin and Wetz, 2015]

All comparison at the operational level!

RSP Divergence

```
SELECT ?shop ?product ?percent
  FROM <products>
  STREAM <coupons>   [RANGE 30m]
  STREAM <locations> [RANGE 5m]
WHERE {
  ?shop    :offers     ?coupon.  ?coupon :reduce ?percent.
  ?coupon  :on         ?product. ?user   :isNear ?shop.
  ?product :g_classify ?gender.
  FILTER
  (?percent >= 20 && ?gender != 1)}
```

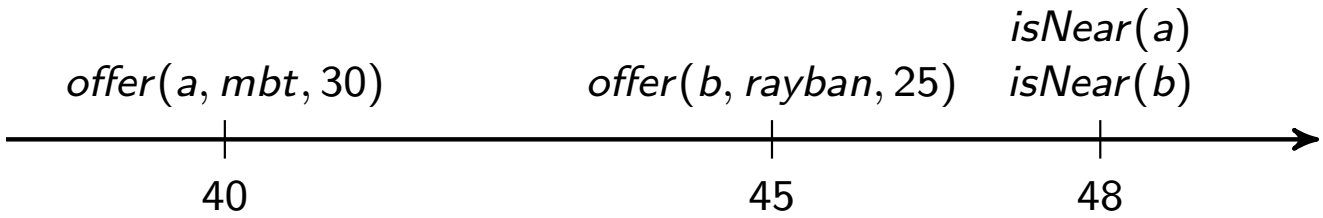
```
SELECT ?shop ?product ?percent
FROM <products>
WHERE {
  STREAM <coupons> [RANGE 30m] {
    ?shop    :offers   ?coupon.
    ?coupon  :reduce   ?percent.
    ?coupon  :on       ?product. }
  STREAM <locations> [RANGE 5m] {
    ?user    :isNear   ?shop. }
  ?product :g_classify ?gender.
  FILTER
  (?percent >= 20 && ?gender != 1)}
```

Key Differences between C-SPARQL and CQELS

	C-SPARQL	CQELS
create snapshot	merge patterns on input streams into the default graph	apply patterns on input streams
execution mode	pull-based	push-based

LARS in a Nutshell: Stream Representation

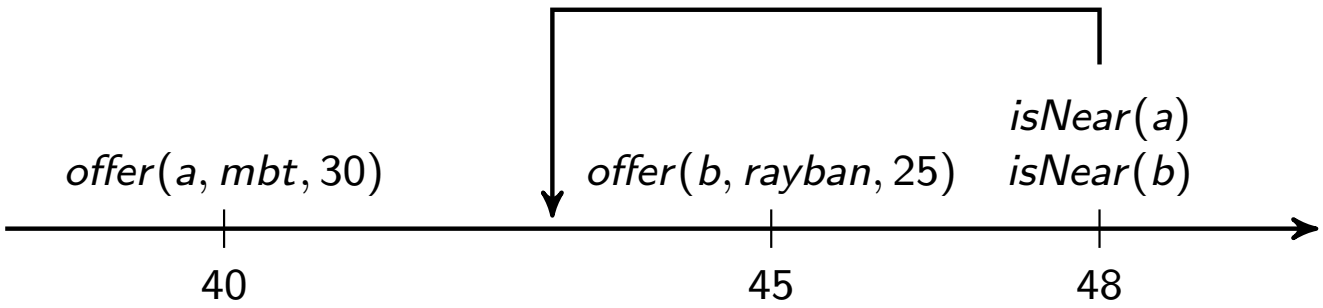
$S = (T, \upsilon)$



$T = [0, 50]$

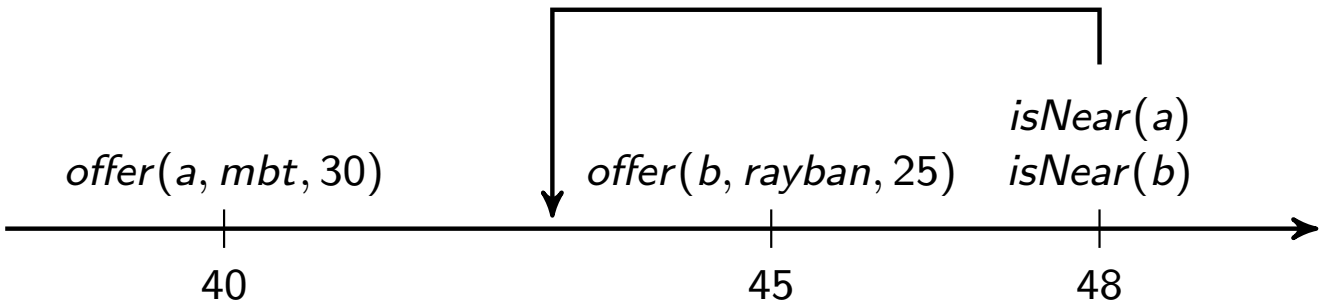
$$\upsilon = \left\{ \begin{array}{l} 40 \mapsto \{offer(a, mbt, 30)\}, 45 \mapsto \{offer(b, rayban, 25)\} \\ 48 \mapsto \{isNear(a), isNear(b)\} \end{array} \right\}$$

LARS in a Nutshell: Window functions



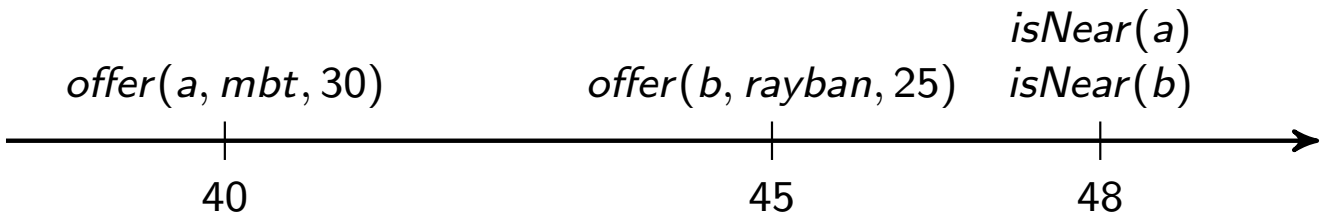
$$S' = w_\iota(S, t, \vec{x})$$

LARS in a Nutshell: Window functions



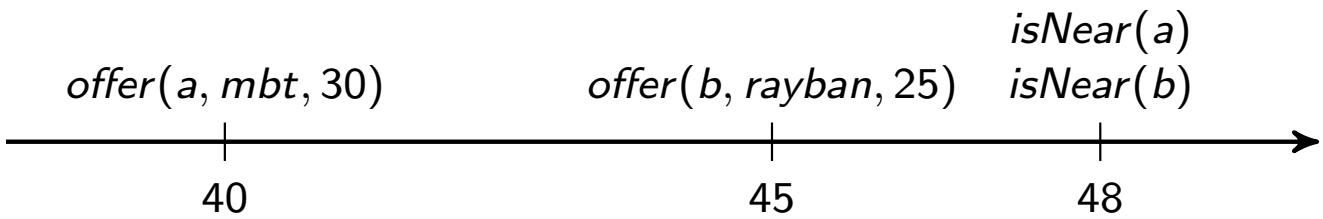
$$S' = w(S, 48, (5, 0, 1)) = ([43, 48], \left\{ \begin{array}{l} 45 \mapsto \{offer(b, rayban, 25)\} \\ 48 \mapsto \{isNear(a), isNear(b)\} \end{array} \right\})$$

LARS in a Nutshell: LARS formulas



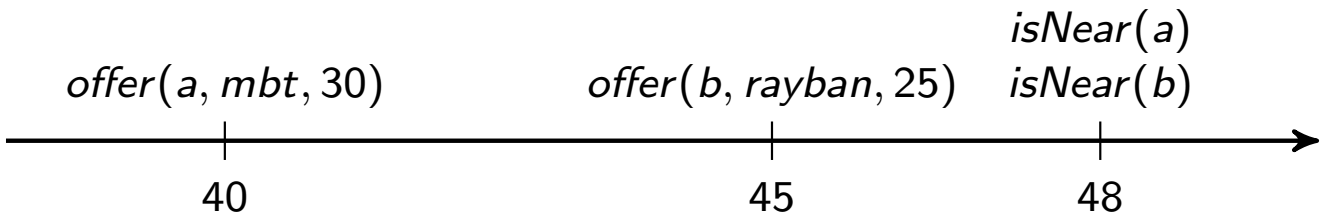
$\alpha ::=$

LARS in a Nutshell: LARS formulas



$$\alpha ::= a \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha$$

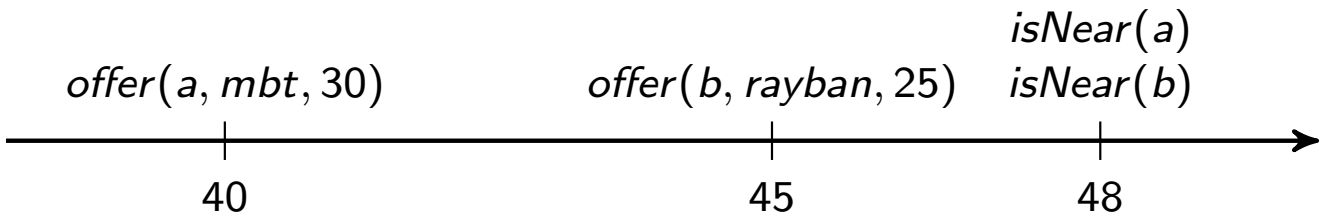
LARS in a Nutshell: LARS formulas



$$\alpha ::= a \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Diamond\alpha \mid \Box\alpha \mid @_t\alpha$$

- various ways for time references

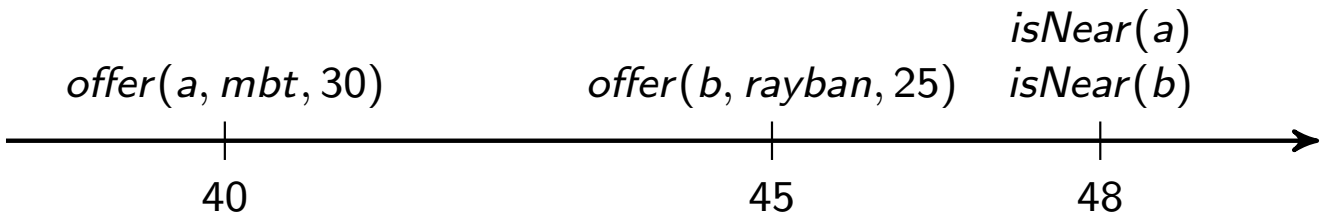
LARS in a Nutshell: LARS formulas



$$\alpha ::= a \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Diamond\alpha \mid \Box\alpha \mid @_t\alpha \mid \boxplus_{\iota}^{\vec{x}}\alpha$$

- various ways for time references
- window operators with possibility to nest

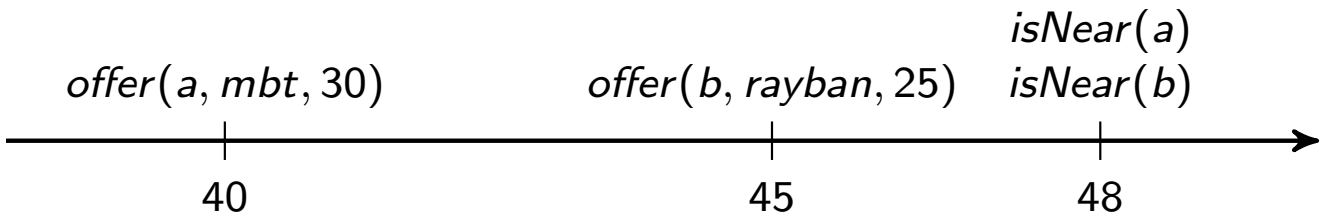
LARS in a Nutshell: LARS formulas



$$\alpha ::= a \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Diamond\alpha \mid \Box\alpha \mid @_t\alpha \mid \boxplus_{\iota}^{\vec{x}}\alpha$$

- various ways for time references
- window operators with possibility to nest
- $\boxplus^{10}\Diamond offer(Sh, Pr, Pe)$

LARS in a Nutshell: LARS formulas



$$\alpha ::= a \mid \neg\alpha \mid \alpha \wedge \alpha \mid \alpha \vee \alpha \mid \alpha \rightarrow \alpha \mid \Diamond\alpha \mid \Box\alpha \mid @_t\alpha \mid \boxplus_{\iota}^{\vec{x}}\alpha$$

- various ways for time references
- window operators with possibility to nest
- $\boxplus^{10}\Diamond offer(Sh, Pr, Pe)$
- $\boxplus^{5}\Box isNear(a)$

LARS in a Nutshell: LARS rules/programs

$$\alpha \leftarrow \beta_1, \ldots, \beta_j, \text{not } \beta_{j+1}, \ldots, \text{not } \beta_n.$$

LARS in a Nutshell: LARS rules/programs

$$\alpha \leftarrow \beta_1, \ldots, \beta_j, \text{not } \beta_{j+1}, \ldots, \text{not } \beta_n.$$

$$\begin{aligned} ans(Sh, Pr, Pe) \leftarrow\ & \boxplus^{30} \Diamond offer(Sh, Pr, Pe), \boxplus^{5} \Diamond isNear(Sh), \\ & g\_classify(Pr, Ge), Pe \geq 20, Ge \neq 1. \end{aligned}$$

LARS in a Nutshell: Semantics

- extend Answer Set Programming semantics
- answer streams: input stream + intentional facts + satisfaction + minimality

LARS to Analyze RSP Queries

Translations from RSP queries to LARS programs:

- offer the correspondence between query answers and answer streams
- capture differences in RSP queries:
  - creating snapshots
  - execution modes

LARS to Analyze RSP Queries (cont'd)

Comparison based on LARS:

- Difference in creating snapshots: can be remedied in practice
- Difference in execution modes: agreement can only be possible interval-wise, on non-dense input streams.

Conclusions

Identify crucial divergences in RSP engines' semantics

Introduce a common model for RSP queries (details in paper)

Capture and analyze the difference of C-SPARQL and CQELS with LARS

References I


Arvind Arasu, Shivnath Babu, and Jennifer Widom.
The CQL continuous query language: semantic foundations and query execution.
VLDB J., 15(2):121–142, 2006.


Davide Francesco Barbieri, Daniele Braga, Stefano Ceri, Emanuele Della Valle, and Michael Grossniklaus.
C-SPARQL: a continuous query language for rdf data streams.
Int. J. Semantic Computing, 4(1):3–25, 2010.


Harald Beck, Minh Dao-Tran, Thomas Eiter, and Michael Fink.
LARS: A logic-based framework for analyzing reasoning over streams.
In *AAAI*, 2015.

References II


Jean-Paul Calbimonte, Óscar Corcho, and Alasdair J. G. Gray.
Enabling ontology-based access to streaming data sources.

In *ISWC (1)*, pages 96–111, 2010.


Daniele Dell'Aglio, Jean-Paul Calbimonte, Marco Balduini, Óscar Corcho, and Emanuele Della Valle.
On Correctness in RDF Stream Processor Benchmarking.
In *ISWC 2013*, pages 326–342, 2013.


Maxim Kolchin and Peter Wetz.
Demo: YABench - Yet Another RDF Stream Processing Benchmark.
In *RSP Workshop*, 2015.

References III


Danh Le Phuoc, Minh Dao-Tran, Josiane Xavier Parreira, and Manfred Hauswirth.
A native and adaptive approach for unified processing of linked streams and linked data.
In *ISWC (1)*, pages 370–388, 2011.


Danh Le Phuoc, Minh Dao-Tran, Minh-Duc Pham, Peter Boncz, Thomas Eiter, and Michael Fink.
Linked stream data processing engines: Facts and figures.
In *ISWC - ET*, pages 300–312, 2012.


Ying Zhang, P. Minh Duc, O. Corcho, and J. P. Calbimonte.
SRBench: A Streaming RDF/SPARQL Benchmark.
In *ISWC*, pages 641–657, 2012.