Mining the Web of Data with Metaqueries

Francesca A. Lisi

University of Bari "Aldo Moro"
Department of Computer Science
Lab of Knowledge Acquisition and Machine Learning (LACAM)

FrancescaAlessandra.Lisi@uniba.it

ILP 2018

Introduction I

The Web of Data

- Feature: builds upon the WWW infrastructure to represent and interrelate data (aka *Linked Data*),
- Aim: transforming the Web from a distributed file system into a *distributed database system*.
- The foundational *standards* of the Web of Data include:

  URI used to identify resources
  RDF used to relate resources

Introduction II

RDF as a data model

- In RDF[a] data is represented in the form of triples $\langle subject\ predicate\ object \rangle$.
- The resulting collection of triples is a *directed, labeled graph* which can be accessed by posing SPARQL[b] queries.
- The link between RDF and *Description Logics* (DLs) allows several *entailment regimes* for query answering in SPARQL.

---

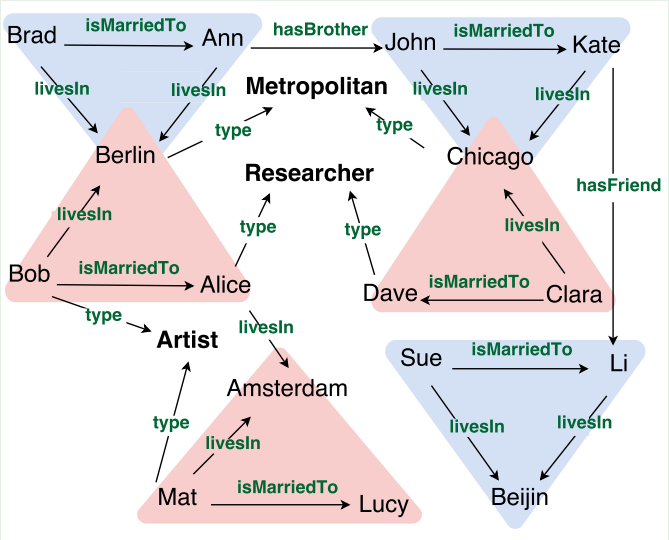
[a] `https://www.w3.org/RDF/`
[b] `https://www.w3.org/TR/rdf-sparql-query/`

Introduction III

Knowledge graphs (KGs)

- Huge RDF graphs, see, *e.g.*, DBpedia (`http://wiki.dbpedia.org/`)
- Automatically constructed by applying information extraction techniques

An example of KG [Tran et al., 2017]



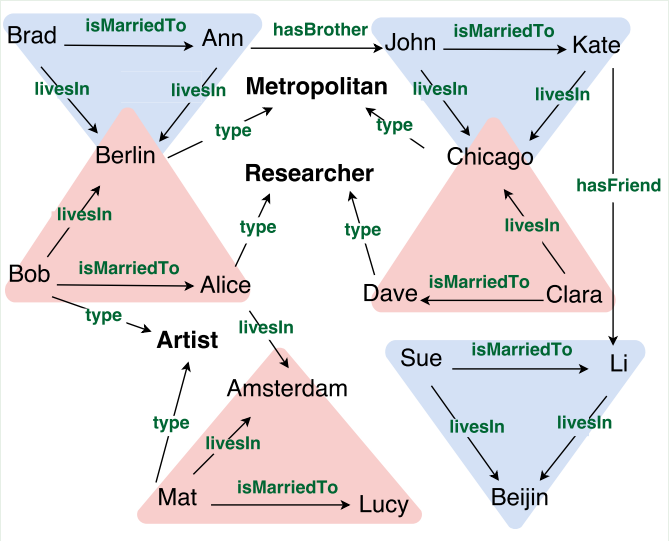
Introduction IV

The curation of KGs

- KGs are inherently *incomplete*.
- KGs particularly need to be curated by performing the task of *completion* (aka *link prediction*)
- Data mining algorithms can be exploited to automatically build rules able to make predictions on missing links.

Introduction V

An example of rule mining for KG completion [Tran et al., 2017]



New facts, *e.g.*, $livesIn(alice, berlin)$, $livesIn(dave, chicago)$ and $livesIn(lucy, amsterdam)$, can be derived from the following mined rule:

$$r1: \ isMarriedTo(x, y), livesIn(x, z) \Rightarrow livesIn(y, z) \qquad (1)$$

and used to complete the KG.

Introduction VI

Challenges of WoD Mining

- Size of KGs
- Open and distributed environment

Suggested solution (already sketched in [Lisi, 2017])

- Exploiting some useful meta-information about the KG in hand.
  - *e.g.*, domains, ranges and confidence values of relations inside the KG (*i.e.*, its schema)
- Adapting well-known data mining techniques that work at the meta-level
  - *e.g.*, *metaquerying* [Ben-Eliyahu-Zohary and Gudes, 1999]

Introduction VII

Metaquerying

- Technique for mining *frequent patterns* in relational databases
- A *metaquery* is a template that describes the type of pattern to be discovered in relational databases [Shen et al., 1996].
- Metaqueries are naturally expressed by means of a *second-order logic language*.

Contribution of the paper

1. Proposal of a metaquerying approach to WoD mining
2. Definition of a metaquery language for WoD mining
   - based on second-order DLs, but
   - implementable with SPARQL.
3. Preliminary analysis of mechanisms for metaquery answering

An example of metaquery for WoD mining

$$P(X, Y), Q(X, Z) \Rightarrow Q(Y, Z) \tag{2}$$

References I

Ben-Eliyahu-Zohary, R. and Gudes, E. (1999).

Towards efficient metaquerying.

In Dean, T., editor, *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, IJCAI 99, Stockholm, Sweden, July 31 - August 6, 1999. 2 Volumes, 1450 pages*, pages 800–805. Morgan Kaufmann.

Lisi, F. A. (2017).

Towards a metaquery language for mining the web of data.

In Calì, A., Wood, P. T., Martin, N. J., and Poulovassilis, A., editors, *Data Analytics - 31st British International Conference on Databases, BICOD 2017, London, UK, July 10-12, 2017, Proceedings*, volume 10365 of *Lecture Notes in Computer Science*, pages 90–93. Springer.

Shen, W., Ong, K., Mitbander, B. G., and Zaniolo, C. (1996).

Metaqueries for data mining.

In *Advances in Knowledge Discovery and Data Mining*, pages 375–398. AAAI/MIT Press.

Tran, H. D., Stepanova, D., Gad-Elrab, M. H., Lisi, F. A., and Weikum, G. (2017).

Towards nonmonotonic relational learning from knowledge graphs.

In Cussens, J. and Russo, A., editors, *Inductive Logic Programming - 26th International Conference, ILP 2016, London, UK, September 4-6, 2016, Revised Selected Papers*, volume 10326 of *Lecture Notes in Computer Science*, pages 94–107. Springer.