

# SOCIAL STRUCTURED AND SEMANTIC SEARCH

Raphaël Bonaque, Bogdan Cautis, François Goasdoué, Ioana Manolescu

## General Framework

A model, **S3**, that we introduce, to include:

- ▷ Structured documents, such as JSON and XML
- ▷ with Semantic annotations, in RDF
- ▷ within a Social network with usual interactions: relations between user, posting, commenting and tagging

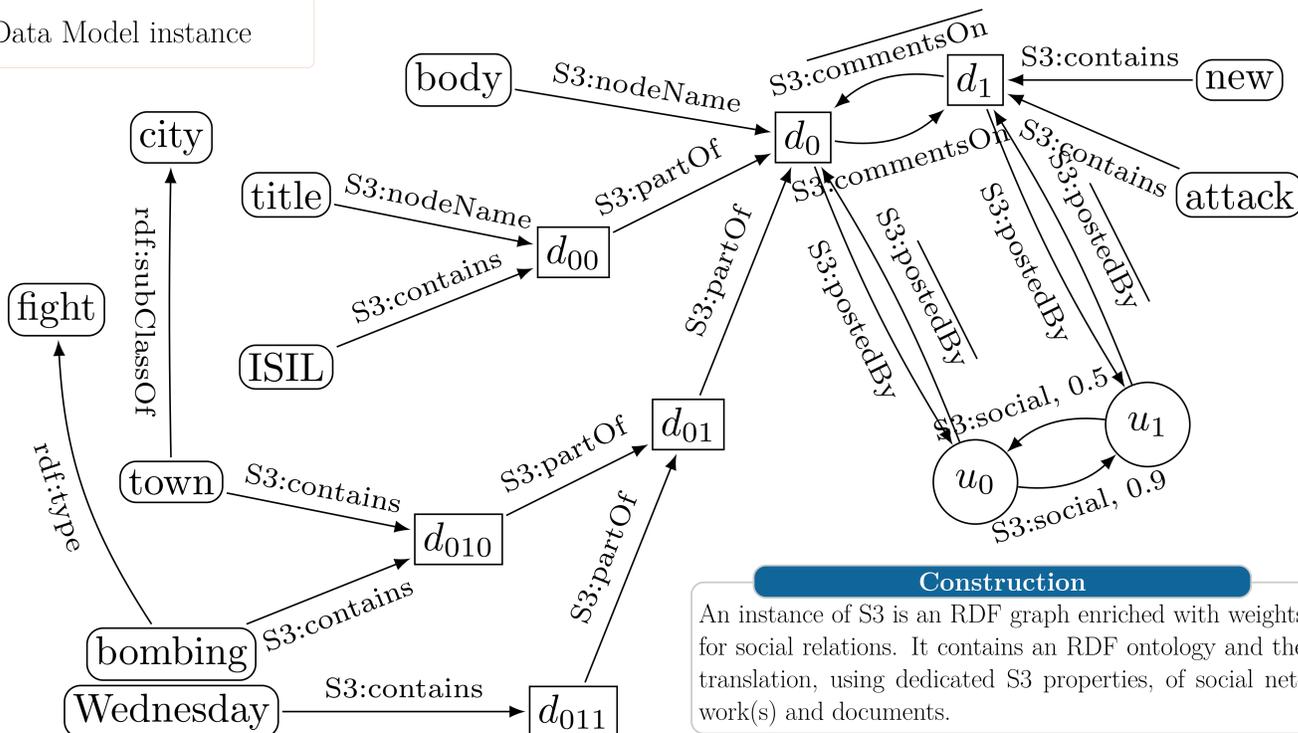
Several goals to achieve in this framework:

- ▷ Top-k search on this model: find the k best documents for a given user query
- ▷ Efficient and practical implementation of the search
- ▷ Proof of correctness

## State of the Art

- ▷ Top-k in social networks, unstructured document, no semantics:
  - Y. Li, Z. Bao, G. Li, and K.-L. Tan. Real time personalized search on social networks. In ICDE, 2015
  - M. Curtiss, I. Becker, T. Bosman, S. Doroshenko, L. Grijincu, T. Jackson, S. Kunnamur, S. Lassen, P. Pronin, S. Sankar, et al. Unicorn: A system for searching the social graph. PVLDB, 2013
  - Silviu Maniu, Bogdan Cautis: Taagle: efficient, personalized search in collaborative tagging networks, SIGMOD 2012
- ▷ Semantics, structured or unstructured documents, without social aspects:
  - François Goasdoué, Julien Leblay, Yannis Katsis, Ioana Manolescu, Stamatis Zampetakis. Growing Triples on Trees, VLDB Journal 2013
  - M. Parades, S. Malaika, J. Siméon, S. Khatchadourian, and K.-U. Sattler. Entity matching for semistructured data in the cloud. In ACM SAC, 2012
  - T. Tran, H. Wang, S. Rudolph, and P. Cimiano. Top-k exploration of query candidates for efficient keyword search on graph-shaped (RDF) data. In ICDE, 2009
  - M. Theobald, R. Schenkel, and G. Weikum. Efficient and self-tuning incremental query expansion for top-k query processing, SIGIR 2005
- ▷ Structured documents, without semantic nor social aspects:
  - C. Aksoy, A. Dimitriou, and D. Theodoratos. Reasoning with patterns to effectively answer XML keyword queries. The VLDB Journal, 2015
  - L. J. Chen and Y. Papakonstantinou. Supporting top-K keyword search in XML databases, ICDE 2010
  - M. Theobald, H. Bast, D. Majumdar, R. Schenkel, and G. Weikum. TopX: efficient and versatile top-k query processing for semistructured data. The VLDB Journal, 17(1), 2008.

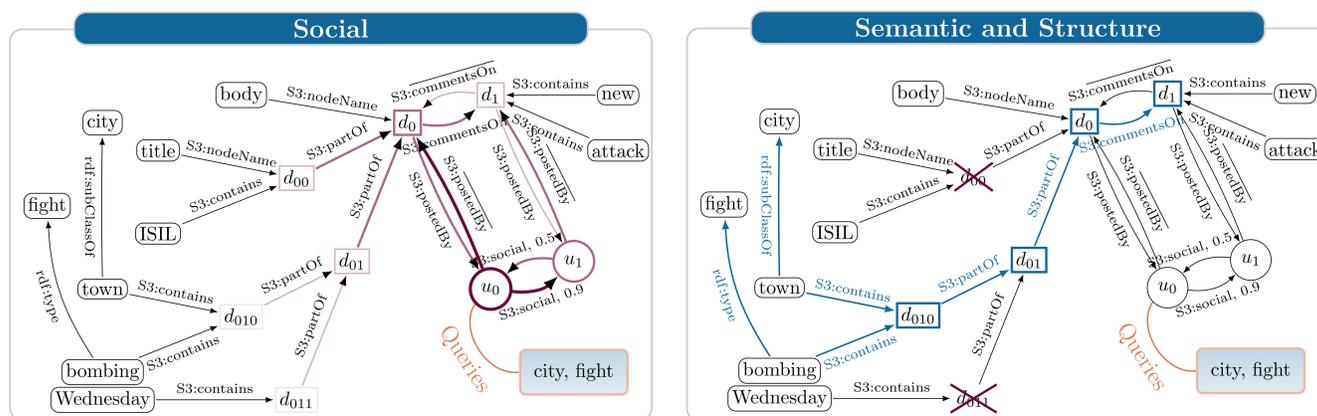
Data Model instance



## Contributions

- ▷ Formalised data and query model
- ▷ Implemented top-k search algorithm
- ▷ Proof of the correctness of the top-k search algorithm for a wide range of general scores
- ▷ Evaluation on datasets using real world data: Twitter (a subset of 2.8M nodes), Vodkaster, and Yelp!

## Our Top-k Approach



We propose a top-k algorithm working on customizable scores : the score of a document for a query must depends on the distance from the user making the query to the sources of relevant keywords for the query.

- ▷ Relevant keywords are derived from the query keywords by inference in RDF
- ▷ Keyword sources are users posting documents or tags containing them
- ▷ The distance between users depends on every social paths between them: paths following social interactions and going through ancestor relations in documents

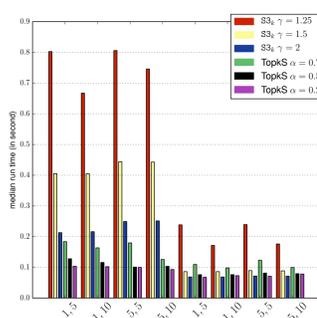
## Results

Our implementation was tested on datasets created from several social networks and knowledge bases with a score function generalising standard social and structural scores.

Compared to a state of the art algorithm working on social data, we:

- ▷ capture ~ 34% more results from the social interactions
- ▷ capture ~ 15% more results from the semantic inheritance

Comparing the runtime of our algorithm,  $S3_k$ , with an algorithm running only on social data and bag of words documents,  $TopkS$ , for more or less frequent keywords, and different sizes of queries and expected answers.



## Perspectives

- ▷ Personalizing query results based on per user ontology developed from the social links
- ▷ Developing a comprehensive way to query heterogeneous data models, not only social and structured, through extended RDF queries
- ▷ Accepted for publication at EDBT 2016 as "Structured, Social and Semantic Search"

PhD of Raphaël Bonaque, under the direction of Ioana Manolescu, Bogdan Cautis and François Goasdoué.