Graph Exploration: Taking the User into the Loop

Davide Mottin, Anja Jentzsch, Emmanuel Müller
Hasso Plattner Institute, Potsdam, Germany

2016/10/24
CIKM2016, Indianapolis, US
Where we are

Background (5 min)
Graph models, subgraph isomorphism, subgraph mining, graph clustering

Exploratory Graph Analysis (35 min)

Focused Graph Mining (35 min)

Refinement of Query Results (35 min)

Real World-Use Case (15min)
Linked Data graphs

Challenges and discussion
The Web of Data

- 1,019 datasets
- 84+ billion RDF triples
- 808+ million RDF links between datasets

http://lod-cloud.net
Vocabularies on the Web of Data

- The Web of Data is heterogeneous
  - Many vocabularies are in use (576 as of October 2016)
  - Many different ways to represent the same information

https://lov.okfn.org
RDF Data Model

ns:cikm2016 rdf:type ns:Event

rdfs:label ACM Conference on Information and Knowledge Management (CIKM2016)

ns:location dbpedia:Indianapolis
RDF Data Model

ns:cikm2016 rdf:type ns:Event

rdfs:label ACM Conference on Information and Knowledge Management (CIKM2016)

ns:location dbpedia:Indianapolis

dbpedia:populationTotal 820445

dbpedia:leaderName Joseph H. Hogsett
Linked Data exploration use cases

- Dataset exploration
- Graph mining
- Query formulation and refinement

But Linked Data is messy
Linked Data graph exploration challenges

- Nested graphs → Makes reasoning difficult
- Loose structure → Things have different property sets
- Incomplete → Missing property definitions
- Poorly formatted → Property types used inconsistently
- Inconsistent → Multiple representations claim opposite things
## Linked Data exploration systems timeline

<table>
<thead>
<tr>
<th>System category / year</th>
<th>2001...</th>
<th>...2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browsers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Textual</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faceted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>xx</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other paradigms</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recommenders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type/domain specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-type/domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undisclosed/commercial</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>View-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Algorithm-based</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Research focus

- **SemWeb in the Scientific American**
- **OWL is a W3C recom.**
- **LOD & DBpedia**
- **SPARQL is a W3C recom.**
DBpedia Mobile

- displays Wikipedia data on map
- aggregates different data sources

RelFinder

- visualization of paths between any 2 entities
- path identification on instance level

gFacet

- Schema exploration
- combines graph-based visualization and faceted filtering techniques

Graph layout is indexed with a spatial data structure, i.e., an R-tree, and stored in a database.

In runtime, user operations are translated into efficient spatial operations (i.e., window queries) in the backend.

LODeX

- Explore a Linked Dataset using a schema summary
- Pick graphical elements from it to create a visual query
- Browse the results
- Refine the query

Aemoo

Exploratory search system based on Encyclopedic Knowledge Patterns
- EKP are knowledge patterns that define the typical classes used to describe entities of a certain class

American Football Player
- Chris Hinton
- Eli Manning
- Marvin Harrison
- Joe Namath
- Joseph Addai
- Don Shula
- Andrew Luck
- Curtis Painter
- Peyton Manning
- Archie Manning

The Indianapolis Colts are an American football team based in Indianapolis, Indiana; they play their games in Lucas Oil Stadium. The team is a member of the South Division of the American Football Conference (AFC) in the National Football League (NFL). The Colts were members of the National Football League from their founding and were one of...(go to Wikipedia page)

Linked Jazz

- reveals the network of the social and professional relations within the American jazz community

Semantic Wonder Cloud

http://sisinflab.poliba.it/semantic-wonder-cloud/index/
inWalk

Castano, S., Ferrara, A. and Montanelli, S. inWalk: Interactive and Thematic Walks inside the Web of Data. EDBT, 2014
ProLOD++

Mining Graph Patterns on the Web of Data

ProLOD++

- Web framework for various data profiling and mining tasks on Linked Datasets
- Explorative research on Linked Dataset graphs to find
  - frequent graph patterns
  - common graph patterns for classes
  - general graph model for Linked Datasets

https://prolod.org
Definition of core set of frequent graph patterns in Linked Datasets based on satellite component analysis

- Path (44x)
- Star (21x)
- Star (17x)
- Star (16x)
- Star (14x)
- Star (13x)
- Star (12x)
- Star (10x)
- Star (9x)
- Star (9x)
- Star (7x)
- Star (6x)
- Star (6x)
- Caterpillar (5x)
- Caterpillar (5x)
ProLOD++
Graph patterns

- Group class-coloured graphs by their permutation groups [Luks82]
  - Permutation group: the set of all automorphisms of a graph

Loupe

Requirements for Linked Data exploratory search systems

- The system provides efficient overviews
- The system helps the user to understand the information space and to shape his mental model
- The user can explore multiple, heterogeneous results and browsing paths
- The system eases the memorization of relevant results
- The system inspires the user and shapes his information need
- The system provokes discoveries
Challenges

● Displaying the graph for exploration
  - E.g. by clustering of topical domains
  - Allowing the user to drill down

● Live graph exploration
  - E.g. via federated SPARQL queries
    ▪ Requires knowledge on endpoint URIs
    ▪ Slow in real-time

● Guiding the user to interesting parts of the graph
  - Usually done by entity inlinks
    ▪ Limited insights
Tutorial outline

Background (5 min)
Graph models, subgraph isomorphism, subgraph mining, graph clustering

Exploratory Graph Analysis (35 min)

Focused Graph Mining (35 min)

Refinement of Query Results (35 min)

Real World-Use Case (15 min)
Linked Data graphs

Challenges and discussion
Summary of Exploratory Graph Analysis

Approximate Queries
- User query is imprecise

By-Example methods
- User query is an example result

- Only need a partial knowledge on the data
- No need for complicate query languages (use examples, partial descriptions)
- The query adapts to user need
- Enable exploratory search by using small queries on the data
Challenges for Exploratory Graph Analysis

Database

• Unsupported in most of the current graph databases
• No ”universal” index to answer multiple type of queries
• Partitioning only for exact query answering

Data mining

• User interactivity in the exploration process
• No solutions for probabilistic graphs
• Respond to queries while the graph changes
• Find examples in streaming settings

Information retrieval

• Exploiting query logs for personalized query answering
• Retrieve results in form of documents converting the query structures
Summary of Focused Graph Mining

The focus on individual user interest
... as **Query** to the Graph Mining System
... as **Seed Node(s)** for Local Search
... as **Attributes** and **Weights**

- get or infer user interest
  → unexpected results
- interactive exploration
  → intuitive parametrization
- adaptive graph mining
  → individual local search
Challenges for Focused Graph Mining

User interactivity in the graph mining process
• unsupported in most of the current graph mining algorithms
• huge variety of user interactions possible
• feedback loop needs to be unified and become exchangeable

Revolution of formal models and search algorithms
• insufficient extensions of existing models and algorithms
• adaptive steering of algorithms vs. fixed parametrization
• evaluation of algorithms with user studies

Scalability of algorithms for real-time interaction
• NP-hard problems, heuristic algorithms, ..., still not scalable
• exploit the user interest for pruning the search space
Summary of Refinement of Query Results

Refinement
- The user query is too restrictive or too generic

Top-k Results
- Queries typically have inexact matches

Skyline Queries
- Find small set of interesting items with many dimensions and incremental updates

- The user might have a very generic idea of how to describe the structure of interest
- The system guides the user towards the answer with simple steps
- The results are explained with reformulations
- The query matches are inexact and interesting
Challenges for Refinement of Query Results

- Real time performance
- Profiling of queries for optimized performance

- Personalized reformulations and interactivity
- Facet search discovery in graphs

- Uncertain graph data
The missing tiles in graph exploration

- Interactivity
- Adaptivity
- Personalization
- Scalability
Slides: https://hpi.de//mueller/tutorials/graph-exploration.html