BUILDING BETTER KNOWLEDGE GRAPHS THROUGH SOCIAL COMPUTING

Elena Simperl

University of Southampton, UK

@esimperl

Erasmus University Rotterdam
November 16th 2018
OVERVIEW

Knowledge graphs have become a critical AI resource
We study them as socio-technical constructs

Our research
- Explores the links between social and technical qualities of knowledge graphs
- Proposes methods and tools to make knowledge graphs better

Picture from https://medium.com/@sderymail/challenges-of-knowledge-graph-part-1-d9ffe9e35214
IN THIS TALK

Effects of editing behaviour and community make-up on the quality of knowledge graph

Crowdsourcing methods to enhance knowledge graphs
EXAMPLE: DBPEDIA

Community project, extracts structured data from Wikipedia

Consistent, centrally defined ontology; support for 125 languages; represents 4.5M items

Open licence

RDF exports, connected to Linked Open Data Cloud
EXAMPLE: WIKIDATA

Wikipedia project creating a knowledge graph collaboratively

20k active users

52M items, no ‘explicit’ ontology

Open licence

RDF exports, connected to Linked Open Data Cloud
Ontologies Are Us: A Unified Model of Social Networks and Semantics


BACKGROUND

Wikidata editors have varied tenure and interests

Editors and editing behaviour impact outcomes

- Group composition can have multiple effects
- Tenure and interest diversity can increase outcome quality and group productivity
- Different editors groups focus on different types of activities

FIRST STUDY: ITEM QUALITY

Analysed the **edit history of items**

- Corpus of 5k items, whose quality has been manually assessed (5 levels)*
- Edit history focused on community make-up
  - Community is defined as set of editors of item
  - Considered features from group diversity literature and Wikidata-specific aspects

*https://www.wikidata.org/wiki/Wikidata:Item_quality
## RESEARCH HYPOTHESES

<table>
<thead>
<tr>
<th>Activity</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Bots edits</td>
<td>Item quality</td>
</tr>
<tr>
<td>H2 Bot-human interaction</td>
<td>Item quality</td>
</tr>
<tr>
<td>H3 Anonymous edits</td>
<td>Item quality</td>
</tr>
<tr>
<td>H4 Tenure diversity</td>
<td>Item quality</td>
</tr>
<tr>
<td>H5 Interest diversity</td>
<td>Item quality</td>
</tr>
</tbody>
</table>

- **H3 Anonymous edits** is the only activity with a downward trend in the outcome of Item quality.
DATA AND METHODS

Ordinal regression analysis, trained four models

**Dependent variable:** 5k labelled Wikidata items

**Independent variables**
- Proportion of bot edits
- Bot human edit proportion
- Proportion of anonymous edits
- Tenure diversity: Coefficient of variation
- Interest diversity: User editing matrix

**Control variables:** group size, item age
## RESULTS

**ALL HYPOTHESES SUPPORTED**

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>SE</td>
<td>P</td>
<td>Coef.</td>
</tr>
<tr>
<td>Label &gt; = D</td>
<td>-.0715</td>
<td>.0609</td>
<td>***</td>
<td>-1.3024</td>
</tr>
<tr>
<td>Label &gt; = C</td>
<td>-1.2553</td>
<td>.0642</td>
<td>***</td>
<td>-2.5499</td>
</tr>
<tr>
<td>Label &gt; = B</td>
<td>-4.4452</td>
<td>.1028</td>
<td>***</td>
<td>-5.7677</td>
</tr>
<tr>
<td>Item age</td>
<td>.0003</td>
<td>.0001</td>
<td>***</td>
<td>.0001</td>
</tr>
<tr>
<td>Group size</td>
<td>.0279</td>
<td>.0014</td>
<td>***</td>
<td>.0330</td>
</tr>
<tr>
<td># Edits</td>
<td>.0029</td>
<td>.0003</td>
<td>***</td>
<td>.0033</td>
</tr>
<tr>
<td>p Bot edits</td>
<td></td>
<td></td>
<td></td>
<td>1.4005</td>
</tr>
<tr>
<td>Bot X Human</td>
<td></td>
<td></td>
<td></td>
<td>4.6909</td>
</tr>
<tr>
<td>p Anonymous edits</td>
<td></td>
<td></td>
<td></td>
<td>-3.8258</td>
</tr>
<tr>
<td>Tenure diversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest diversity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**H1**

**H2**

**H3**

**H4**

**H5**
## SUMMARY AND IMPLICATIONS

<table>
<thead>
<tr>
<th>01</th>
<th>02</th>
<th>03</th>
<th>04</th>
</tr>
</thead>
<tbody>
<tr>
<td>The more is not always the merrier</td>
<td>Bot edits are key for quality, but bots and humans are better</td>
<td>Registered editors have a positive impact</td>
<td>Diversity matters</td>
</tr>
<tr>
<td>Encourage registration</td>
<td>Identify further areas for bot editing</td>
<td>Design effective human-bot workflows</td>
<td>Suggest items to edit based on tenure and interests</td>
</tr>
</tbody>
</table>
SECOND STUDY: ONTOLOGY QUALITY

Analysed the **Wikidata ontology** and its **edit context**

- Defined as the graph of all items linked through **P31** (instance of) & **P279** (subclass of)
- Calculated evolution of quality metrics and editing activity over time and the links between them
  - Based on features from literature on ontology evaluation and community-driven ontology engineering
DATA AND METHODS

Wikidata dumps from **March 2013** (creation of **P279**)
to **September 2017**

- Analysed data in 55 monthly time frames

Literature survey to defined Wikidata ontology quality framework

Clustering to identify ontology editor roles

Lagged multiple regression to link roles and ontology features

- **Dependent variable**: Changes in ontology quality across time
- **Independent variables**: number of edits by different roles
- **Control variables**: Bot and anonymous edits
**ONTOMETRY QUALITY: METRICS**

Based on 7 ontology evaluation frameworks

Compiled structural metrics that can be determined from the dumps

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Description</th>
<th>Indicator</th>
<th>Description</th>
<th>Indicator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>noi</em></td>
<td>Number of instances</td>
<td><em>ap; mp</em></td>
<td>Average and median</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>population</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>noc</em></td>
<td>Number of classes</td>
<td><em>rr</em></td>
<td>Relationship richness</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>norc</em></td>
<td>Number of root classes</td>
<td><em>ir, mr</em></td>
<td>Inheritance and median</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>richness</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>nolc</em></td>
<td>Number of leaf classes</td>
<td><em>cr</em></td>
<td>Class richness</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>nop</em></td>
<td>Number of properties</td>
<td><em>ad, md, maxd</em></td>
<td>Average, median, and max explicit depth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ONTOLOGY QUALITY: RESULTS
LARGE ONTOLOGY, UNEVEN QUALITY

>1.5M classes, ~4000 properties
No of classes increases at same rate as overall no of items, likely due to users incorrectly using P31 & P279
ap and cr decrease over time (several classes are either without instances or sub-classes or both)
ir & maxd increase over time (part of the Wikidata ontology is distributed vertically)
K-means, features based on previous studies

Analysis by yearly cohort

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td># edits</td>
<td>Total number of edits per month.</td>
<td># property edits</td>
<td>Total number of edits on Properties in a month.</td>
</tr>
<tr>
<td># ontology edits</td>
<td>Number of edits on classes.</td>
<td># taxonomy edits</td>
<td>Number of edits on P31 and P279 statements.</td>
</tr>
<tr>
<td># discussion edits</td>
<td>Number of edits on talk pages.</td>
<td># batch edits</td>
<td>Number of edits done through automated tools.</td>
</tr>
<tr>
<td># modifying edits</td>
<td>Number of revisions on previously existing statements.</td>
<td>item diversity</td>
<td>Proportion between number of edits and number of items edited.</td>
</tr>
<tr>
<td>admin</td>
<td>True if user in an admin user group, false otherwise.</td>
<td>lower admin</td>
<td>True if user in a user group with enhanced user rights, false otherwise.</td>
</tr>
</tbody>
</table>
EDITOR ROLES: RESULTS

190,765 unique editors over 55 months (783k total)
18k editors active for 10+ months
2 clusters, obtained using gap statistic (tested $2 \leq k \leq 8$)

**Leaders**: more active minority (~1%), higher number of contributions to ontology, engaged within the community

**Contributors**: less active, lower number of contributions to ontology and lower proportion of batch edits
People who joined the project early tend to be more active & are more likely to become leaders.

Levels of activity of leaders decrease over time (alternatively, people move on to different tasks).
RESEARCH HYPOTHESES

**H1** Higher levels of leader activity are negatively correlated to number of classes \((noc)\), number of root classes \((norc)\), and number of leaf classes \((nolc)\)

**H2** Higher levels of leader activity are positively correlated to inheritance richness \((ir)\), average population \((ap)\), and average depth \((ad)\)
ROLES & ONTOLOGY: RESULTS

**H1** not supported

**H2** partially supported

Only *inheritance richness* (*ir*) and *average depth* (*ad*) related significantly with leader edits ($p<0.01$)

Bot edits significantly and positively affect the number of subclasses and instances per class (*ir* & *ap*) ($p<0.05$)
SUMMARY AND IMPLICATIONS

Creating ontologies still a challenging task

Size of the ontology renders existing automatic quality assessment methods unfeasible

Broader curation efforts are needed: large number of empty classes

Editor roles less well articulated than in other ontology engineering projects

Possible decline in motivation after several months
NOBODY KNOWS EVERYTHING, BUT EVERYBODY KNOWS SOMETHING

BACKGROUND

Varying quality of Linked Data sources

dbpedia:Dave_Dobbyn dbprop:dateOfBirth "3".

Detecting and correcting errors may require manual inspection

Different crowds are more or less motivated (or skilled) to undertake specific aspects of this work

We propose a scalable way to carry out this work
**Approach**

Contest
LD Experts
*Difficult* task
Final prize

**Microtasks**
Workers
*Easy* task
Micropayments

**MTurk interfaces**

Incorrect object
“Dave Dobbyn”

Date of birth: 3 January 1957

Incorrect data type
Kyoto University

Given the property "name", is the value "DBpedia" of type "english"?

Incorrect outlink
John Two-Hawks

**Findings**

Use the right crowd for the right task

Experts detect a range of issues, but will not invest additional effort

Turkers can carry out the three tasks and are exceptionally good at data comparisons

**Results: Precision**

<table>
<thead>
<tr>
<th></th>
<th>Object values</th>
<th>Data types</th>
<th>Interlinks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linked Data experts</td>
<td>0.7151</td>
<td>0.8270</td>
<td>0.1525</td>
</tr>
<tr>
<td>MTurk (majority voting)</td>
<td>0.8977</td>
<td>0.4752</td>
<td>0.9412</td>
</tr>
</tbody>
</table>
ALL ROADS LEAD TO ROME

THREE WORKFLOWS TO ADD MISSING ITEM TYPES

Free associations
Validating the machine
Exploring the DBpedia ontology

Findings

- **Shortlists are easy & fast**
  - Popular classes are not enough
  - Alternative ways to explore the taxonomy
- **Freedom comes with a price**
  - Unclassified entities might be unclassifiable
  - Different human data interfaces
- **Working at the basic level of abstraction achieves greatest precision**
  - But when given the freedom to choose, users suggest more specific classes
SUMMARY OF FINDINGS

Social computing offer a useful lens to study knowledge graphs

Social fabric of graphs affect quality

Crowdsourcing methods can be used to curate and enhance knowledge graphs