BUILDING BETTER KNOWLEDGE GRAPHS THROUGH SOCIAL COMPUTING

Elena Simperl

University of Southampton, UK

@esimperl

Erasmus University Rotterdam
November 16th 2018

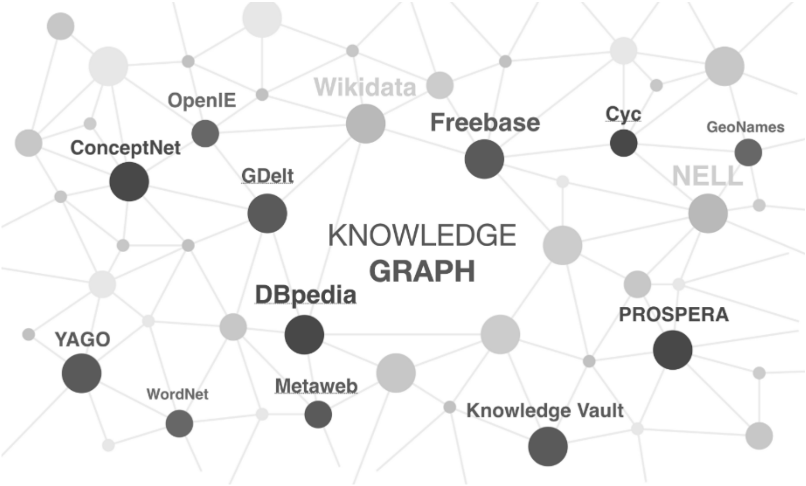
OVERVIEW

Knowledge graphs have become a critical AI resource

We study them as socio-technical constructs

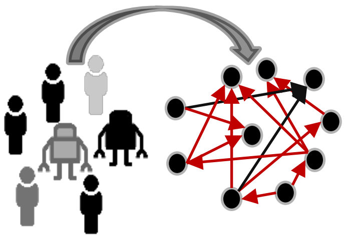
Our research

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

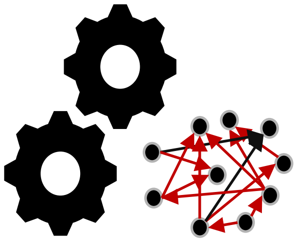


Picture from https://medium.com/@sderymail/challenges-of-knowledge-graph-part-1-d9ffe9e35214

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

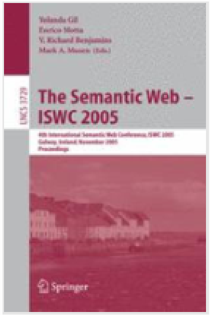
Wikipedia project creating a knowledge graph collaboratively

20k active users

52M items, no ‘explicit’ ontology

Open licence

RDF exports, connected to Linked Open Data Cloud



International Semantic Web Conference
ISWC 2005: The Semantic Web – ISWC 2005 pp 522-536 | Cite as

Ontologies Are Us: A Unified Model of Social Networks and Semantics

Authors   Authors and affiliations

Peter Mika

'ONTOLOGIES ARE US'

Piscopo, A., Phethean, C., & Simperl, E. (2017). What Makes a Good Collaborative Knowledge Graph: Group Composition and Quality in Wikidata. *International Conference on Social Informatics,* 305-322, Springer.

Piscopo, A., & Simperl, E. (2018). Who Models the World?: Collaborative Ontology Creation and User Roles in Wikidata. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 141.

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

Chen, J., Ren, Y., Riedl, J.: The effects of diversity on group productivity and member withdrawal in online volunteer groups. In: Proceedings of the 28th international conference on human factors in computing systems - CHI '10. p. 821. ACM Press, New York, USA (2010)

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

	Activity		Outcome	
H1	Bots edits	↑	Item quality	↑
H2	Bot-human interaction	↑	Item quality	↑
H3	Anonymous edits	↑	Item quality	↓
H4	Tenure diversity	↑	Item quality	↑
H5	Interest diversity	↑	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

		Model 1			Model 2			Model 3			Model 4		
		Coef.	SE	P	Coef.	SE	P	Coef.	SE	P	Coef.	SE	P
Label > = D		−.0715	.0609		−1.3024	.1037	***	−1.1739	.1779	***	−2.6487	.2125	***
Label > = C		−1.2553	.0642	***	−2.5499	.1081	***	−2.3874	.1815	***	−4.1062	.2175	***
Label > = B		−4.4452	.1028	***	−5.7677	.1361	***	−5.8900	.2145	***	−7.5732	.2450	***
Label > = A		−6.2173	.1320	***	−7.6024	.1628	***	−7.4843	.2262	***	−9.2759	.2573	***
Item age		.0003	.0001	***	.0001	.0001		.0002	.0001		−.0008	.0001	***
Group size		.0279	.0014	***	.0330	.0015	***	.0152	.0015	***	.0248	.0016	***
# Edits		.0029	.0003	***	.0033	.0003	***	.0039	.0003	***	.0040	.0003	***
p Bot edits	H1				1.4005	.1029	***				2.4695	.1237	***
Bot X Human	H2				4.6909	.3377	***				3.7688	.3618	***
p Anonymous edits	H3				−3.8258	1.2218	**				−3.6628	1.2403	
Tenure diversity	H4							1.5502	.1104	***	2.8043	.1166	***
Interest diversity	H5							1.0104	.1972	***	1.1004	.1999	***

SUMMARY AND IMPLICATIONS

01	02	03	04
The more is not always the merrier	Bot edits are key for quality, but bots and humans are better	Registered editors have a positive impact	Diversity matters

01	02	03	04
Encourage registration	Identify further areas for bot editing	Design effective human-bot workflows	Suggest items to edit based on tenure and interests

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

ONTOLOGY QUALITY: METRICS

Based on 7 ontology evaluation frameworks

Compiled structural metrics that can be determined from the dumps

noi	Number of instances	*ap; mp*	Average and median population
noc	Number of classes	*rr*	Relationship richness
norc	Number of root classes	*ir, mr*	Inheritance and median richness
nolc	Number of leaf classes	*cr*	Class richness
nop	Number of properties	*ad, md, maxd*	Average, median, and max explicit depth

Sicilia, M. A., Rodríguez, D., García-Barriocanal, E., & Sánchez-Alonso, S. (2012). Empirical findings on ontology metrics. *Expert Systems with Applications*, *39*(8), 6706-6711.

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)

EDITOR ROLES: METHODS

K-means, features based on previous studies

Analysis by yearly cohort

# edits	Total number of edits per month.	*# property edits*	Total number of edits on Properties in a month.
# ontology edits	Number of edits on classes.	*# taxonomy edits*	Number of edits on P31 and P279 statements.
# discussion edits	Number of edits on talk pages.	*p batch edits*	Number of edits done through automated tools.
# modifying edits	Number of revisions on previously existing statements.	*item diversity*	Proportion between number of edits and number of items edited.
admin	True if user in an admin user group, false otherwise.	*lower admin*	True if user in a user group with enhanced user rights, false otherwise.

EDITOR ROLES: RESULTS

190,765 unique editors over 55 months (783k total)

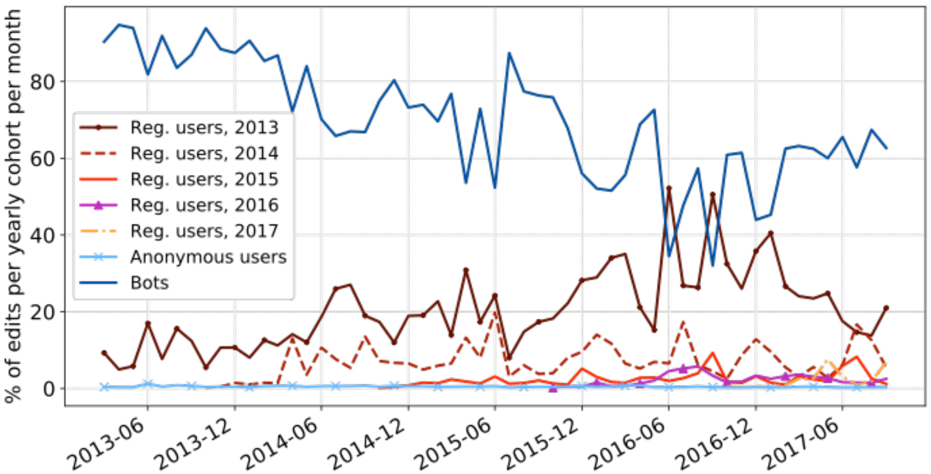
18k editors active for 10+ months

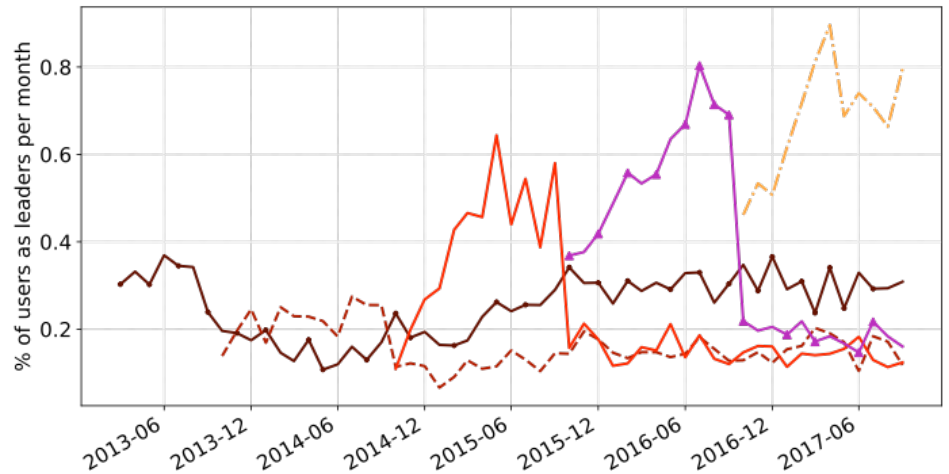
2 clusters, obtained using *gap statistic* (tested $2 \geq k \geq 8$)

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

EDITOR ROLES: RESULTS





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

Acosta, M., Zaveri, A., Simperl, E., Kontokostas, D., Flöck, F., & Lehmann, J. (2016). Detecting Linked Data quality issues via crowdsourcing: A DBpedia study. *Semantic Web* Journal, 1-34.

BACKGROUND

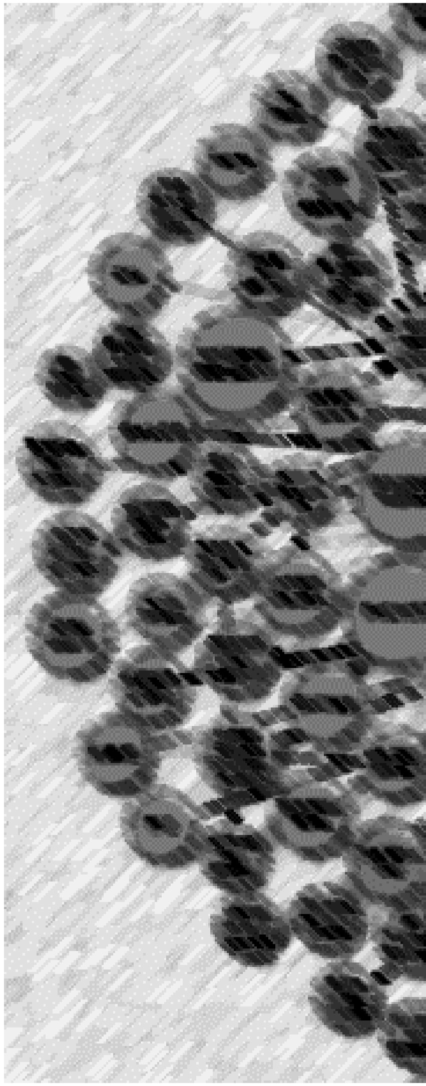
Varying quality of Linked Data sources

```
dbpedia:Dave_Dobbyn dbprop:dateOfBirth “3”.
```

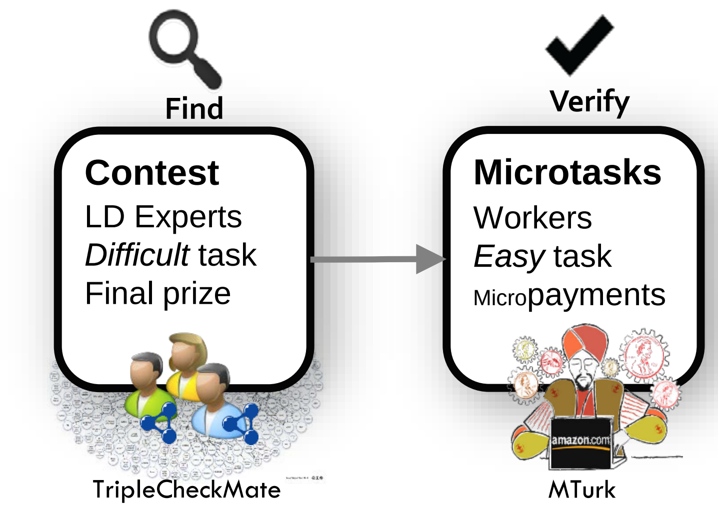
Detecting and correcting errors may require manual inspection

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

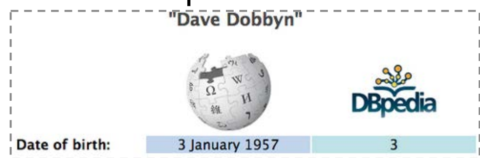


Approach



MTurk interfaces

Incorrect object



Incorrect data type

Incorrect outlink

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

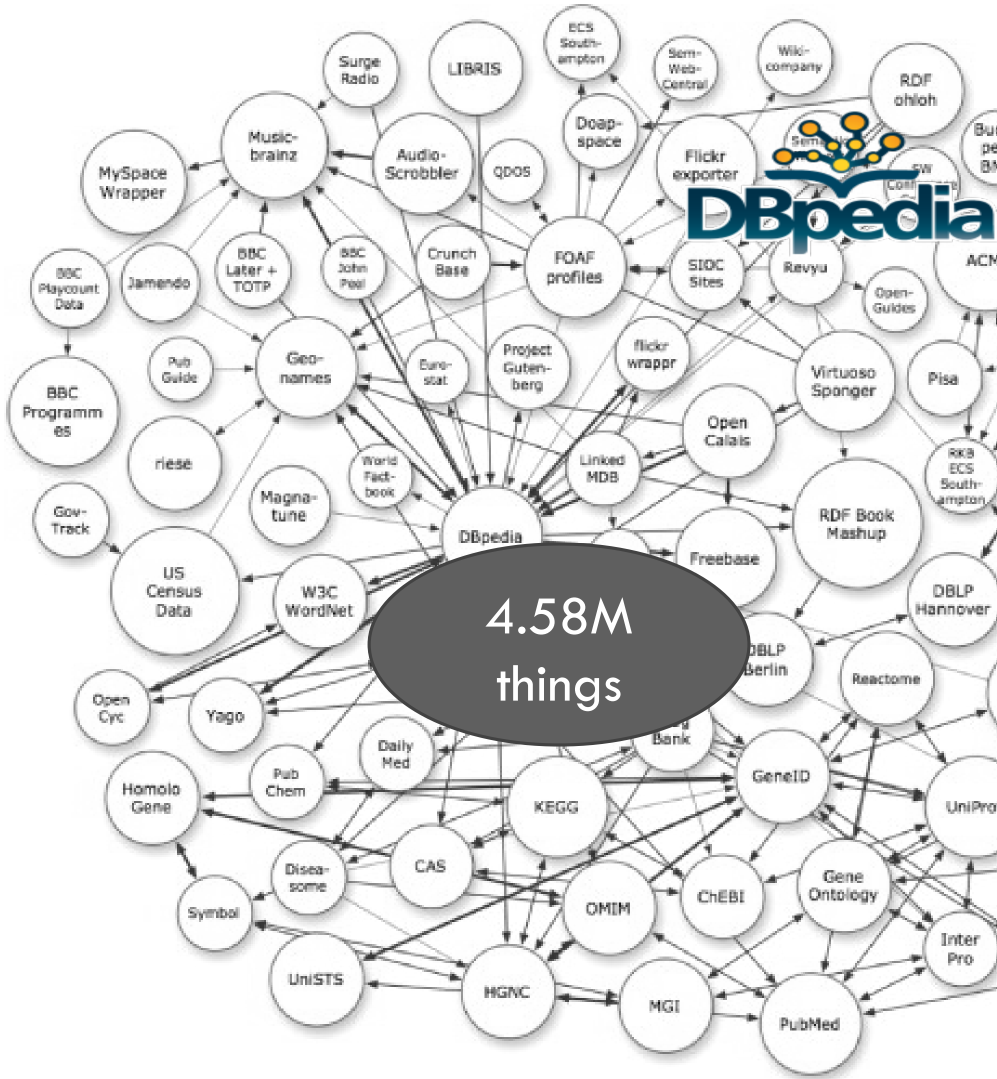
Results: Precision

	Object values	Data types	Interlinks
Linked Data experts	0.7151	0.8270	0.1525
MTurk (majority voting)	0.8977	0.4752	0.9412

ALL ROADS LEAD TO ROME

Bu, Q., Simperl, E., Zerr, S., & Li, Y. (2016). Using microtasks to crowdsource DBpedia entity classification: A study in workflow design. *Semantic Web Journal*, 1-18

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

BUILDING BETTER KNOWLEDGE GRAPHS THROUGH SOCIAL COMPUTING

- Bu, Q., Simperl, E., Zerr, S., & Li, Y. (2016). Using microtasks to crowdsource DBpedia entity classification: A study in workflow design. *Semantic Web Journal*, 1-18
- Acosta, M., Zaveri, A., Simperl, E., Kontokostas, D., Flöck, F., & Lehmann, J. (2016). Detecting Linked Data quality issues via crowdsourcing: A DBpedia study. *Semantic Web* Journal, 1-34.
- Piscopo, A., Phethean, C., & Simperl, E. (2017). What Makes a Good Collaborative Knowledge Graph: Group Composition and Quality in Wikidata. *International Conference on Social Informatics,* 305-322, Springer.
- Piscopo, A., & Simperl, E. (2018). Who Models the World?: Collaborative Ontology Creation and User Roles in Wikidata. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 141.