Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata

ALESSANDRO PISCOPO, University of Southampton, United Kindgdom
ELENA SIMPERL, University of Southampton, United Kingdom

Wikidata is a collaborative knowledge graph which is central to many academic and industry IT projects. Its users are responsible for maintaining the schema that organises this knowledge into classes, properties, and attributes, which together form the Wikidata 'ontology'. In this paper, we study the relationship between different Wikidata user roles and the quality of the Wikidata ontology. To do so we first propose a framework to evaluate the ontology as it evolves. We then cluster editing activities to identify user roles in monthly time frames. Finally, we explore how each role impacts the ontology. Our analysis shows that the Wikidata ontology has uneven breadth and depth. We identified two user roles: contributors and leaders. The second category is positively associated to ontology depth, with no significant effect on other features. Further work should investigate other dimensions to define user profiles and their influence on the knowledge graph.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**; **Wikis**; • **Information systems**;

Additional Key Words and Phrases: Collaborative knowledge engineering, Wikidata, user roles, ontologies

ACM Reference Format:
Alessandro Piscopo and Elena Simperl. 2018. Who Models the World? Collaborative Ontology Creation and User Roles in Wikidata. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW (November 2018), 18 pages. https://doi.org/10.1145/3274410

1 INTRODUCTION

Wikidata is the collaborative knowledge graph (KG) initiated by the Wikimedia Foundation in 2012 [45]. Since its launch, its community has grown up to more than 100 thousand registered editors, who have contributed knowledge about around 46 million entities.

Knowledge graphs are a technology used to add context and depth to anything from web search to product recommendations and intelligent assistants. They describe real-world entities, their relationships, and attributes [32]. A KG typically spans across several domains and is built on top of a conceptual schema, or *ontology*, which defines what types of entities (*classes*) are allowed in the graph, alongside the types of *properties* they can have.

Creating KGs is not trivial. It requires a mix of sophisticated machine algorithms and human input. DBpedia, for example, is automatically extracted from Wikipedia via mapping rules created by knowledge engineers and domain experts [24]. YAGO is extracted from Wikipedia as well, but relies on other resources to recognise the type of each entity in the KG [26]. Freebase, a Google project closed in 2015, was built by a team of experts supported by crowdsourcing [32]. While many of these KGs overlap from a content point of view, they differ in their sociotechnical fabric. Each of

Authors' addresses: Alessandro Piscopo, University of Southampton, Southampton, United Kindgdom, alessandro.piscopo@gmail.com; Elena Simperl, University of Southampton, Southampton, United Kingdom, E.Simperl@soton.ac.uk.

2573-0142/2018/11-ART $15.00
https://doi.org/10.1145/3274410

them automates some parts of the knowledge graph creation and maintenance process in slightly different ways. More importantly for our context, they vary in the way they use crowdsourcing—sometimes, KG platforms rely on experts for specific types of activities (Freebase, Yago), define rules for how and by whom some activities should be carried out (DBpedia, Yago), or provide tools to facilitate collaboration (DBpedia). In this respect, Wikidata is modelled after Wikipedia. It is built by its community of editors without any restrictions on what entities or properties to create first or how to group entities into classes and create a class hierarchy. It relies of a minimal set of predefined roles that come with a set of special rights or permissions and encourages discussions and collaborative decisions via social channels such as Talk pages.

A range of factors—its open, collaborative approach, the support from the Wikimedia Foundation and the community gathered around it, the insights from prior analogue projects—have created a fertile ground to make Wikidata one of the largest knowledge resources ever available in around six years. The Wikidata community is entirely responsible for editing and maintaining the knowledge graph, which is released under an open license and continuously updated—10 million edits monthly as of April 2018. This makes it an interesting subject of study for collaborative ontology engineering—besides its scale, in terms of community size, levels of participation and size of the KG, it is the only project of its kind that has approached an activity that used to be carried out by teams of trained professionals according to carefully crafted methodologies, using grassroots contributions by its community and minimal guidelines and instructions.

The tasks carried out by Wikidata users may differ considerably in terms of workflow and cognitive effort. They can perform simple, repetitive edits, e.g. adding statements that assert facts about a particular entity, such as *Ada Lovelace–place of birth–London*. They can also work on more complex tasks that require cognition of knowledge engineering concepts, such as the curation of the ontology, which forms the backbone of every KG. This means grouping similar entities into classes and declaring properties that may apply between entities from that class. In our example, the class would be *human* for *Ada Lovelace* and *city* or *capital* for *London*. This combination of tasks is peculiar to Wikidata and, in relation to CSCW, led researchers to place it at an intersection between peer-production systems and collaborative ontology engineering communities [27]. Wikidata editors seem to take roles that are somewhere on a line between these two categorisations, with some focusing more on adding and maintaining data, while others working on organising the underlying ontology [27]. Moreover, seasoned users tend to be more active and to participate in the life of the community [34]. They more likely undertake ontology-related activities, such as creating classes and properties and defining sub– and super-classes, feeling that these define their identity on the platform [34].

In this paper, we build upon these previous studies to understand the range of contributions users make to the Wikidata ontology and their impact on the ontology quality. We carry out a literature review to understand the different approaches and metrics that have been applied in the past to assess the quality of ontologies and define a framework suitable for Wikidata. We then cluster editing activities using k-means and features suggested in related work to identify characteristic user roles in monthly time frames. Finally, we explore the relationship between user roles and ontology quality in time by fitting several regression models.

Our paper fills a gap in collaborative knowledge engineering by analysing one of the most significant projects in this space. In the context of Wikidata, our contribution is threefold: (*i.*) we propose a quality framework for the Wikidata ontology, demonstrating its suitability as a tool to monitor changes to its quality; (*ii.*) we derive a set of user roles based on a broader range of activity patterns and contribute to contextualise Wikidata within the field of online collaboration; (*iii.*) we shed light on the links between collaborative processes and the outcomes of its community, by investigating how user roles influence the ontology quality.

2 WIKIDATA: A COLLABORATIVE KNOWLEDGE GRAPH

Items and *Properties* are the building blocks of Wikidata. Items refer to concrete or abstract entities, e.g. *Ada Lovelace* or the *Colosseum*, or to classes of entities, such as *humans* or *music genres*. Properties are used to state any relationship between entities. Items and Properties are identified by alphanumeric codes (URIs), in which a letter—*Q* for Items, *P* for Properties—is followed by a number (e.g. Q5 or P31). Editors can add human-readable labels and descriptions to Items and Properties in any of the 358 languages used in Wikidata [45].

Facts about Items and Properties are asserted by means of *statements* (see Figure 1). The core of a statement is the *claim*, a property-value pair that connects an Item or a literal, and possibly a *qualifier*, to the Item they refer to. Statements can be enriched with *references* or *ranks*. The knowledge graph is the set of all statements, which link Items to each other.

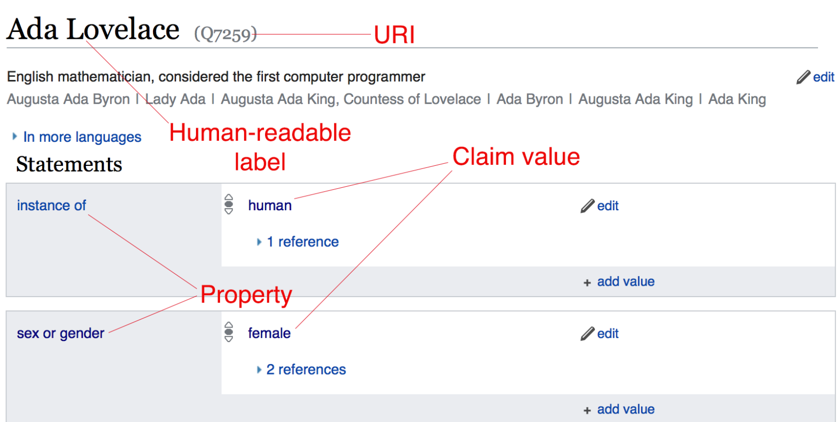

Fig. 1. An example of Wikidata Item, shown through its web interface, with its main elements highlighted. In the picture, references are collapsed, they can be shown by clicking on the related link.

Wikidata editors can be either *bots*, i.e. software programmed to execute specific tasks, or *humans*. Bots author most edits [38] and are used to carry out a wide range of routine tasks. They are initiated by human editors and approved by the community [49]. Users who operate bots are also responsible for their work and must check if they cause any damage to the knowledge graph.

Human editors can contribute either after registering or anonymously. The latter are responsible for a small percentage of revisions (0.5%), although they edit Properties to a larger extent than bots [27]. Anonymous edits are likely to be detrimental for Item quality [33], in line with studies on anonymous editing on Wikipedia, which often results in spam and vandalism [1]. Concerning registered editors, they may be granted different sets of rights if they take on higher-responsibility roles such as 'administrator', 'bureaucrat' or 'Property creator' [50]. These roles are assigned through a community vote or directly by another user in the same group. Regular registered editors can edit and create new Items, and use various types of tools to perform revisions in batch. Semi-automated editing tools, such as *QuickStatements* or *The Wikidata Game*, allow users to edit at a much higher rate than it would be possible through the web interface. We refer to edits made using these tools also as 'batch edits'. Regular editors cannot carry out any changes that might have a substantial influence on the entire graph, such as deleting pages, or creating new Properties. These are a prerogative of users with additional privileges, who may edit templates; block and remove users; and patrol the system for quality checking. Administrators and Property creators are the only users allowed to create new Properties, under the approval of the community.

2.1 The Wikidata ontology

Ontologies capture the classes of entities of interest in a given domain and their properties [37]. They are used to facilitate communication among people, as they standardise terminology and provide guidance for classification. They are equally useful to add context and depth to algorithms for anything from web search to recommender systems and conversational agents. Different communities have developed their own ontologies, for example in cultural heritage [10] or health and life sciences [3, 5]. Ontologies vary in size, breadth, and format. For example, the DBpedia ontology has 735 classes and around 2800 properties [32], whereas YAGO counts with more than 480 thousand classes and only 77 properties [32]. A domain-specific ontology such as the Gene Ontology has more than 40 thousand classes [3]. Some ontologies are encoded in so-called formal knowledge representation languages—these have a rigorous mathematical underpinning which allows one to use automatic reasoning to detect inconsistencies, classify new entities into classes based on their properties, and derive new knowledge.

Wikidata does not have a predefined formal ontology. The ontology is loosely defined by the relationships between the Items in the graph, similarly to Freebase [6]. This means, among others, that Wikidata does not formally distinguish between Items that are classes, for example *city*, and Items that are entities, for example *London*. This distinction is important for any AI that would use the knowledge graph to understand whether in a given context the entity *London* would refer to the largest city in the United Kingdom or would rather stand for the British government, which is an entity of a different class, *executive body*. Prior studies into the Wikidata ontology have used Properties describing taxonomic relations—*instance of* (P31) and *subclass of* (P279)—to identify those Items which take the function of classes. Erxleben et al. [11] exploited these relations to extract an OWL ontology from Wikidata[1]. Another study [8] used P31 and P279 to explore common issues in the Wikidata taxonomy, highlighting three main anti-patterns, attributable principally to the misuse of the two Properties. Following the same approach, we define the Wikidata ontology as the set of all its Properties and the Items that are used as classes, i.e. those that are subject or object of *subclass of* (P279), or object of *instance of* (P31). In our previous example, the ontology would entail the Property *place of birth* alongside all classes to which the two entities, *Ada Lovelace* and *London* are linked to: humans, cities, capitals, financial centres, etc.

A consequence of the informal, collaborative nature of Wikidata is that its ontology may change quickly, depending on edits on Properties and on the use of P31 and P279. An application looking up information in Wikidata via an API would receive very different results at different points in time, as these results very much depend on the ontological structure of the knowledge graph. While previous studies have started investigating how users edit the ontology (see next section), none of them has yet looked at editing activities over time, which is one of the aims of our paper.

2.2 User roles and collaboration in Wikidata

There is an extensive literature around user behaviour in online communities and specifically on the roles taken by different contributors. Distinctive roles are often identified from patterns spontaneously emerging from user activity [48]. Although they may coincide or overlap with formal or hierarchical roles, they are more generally determined by different levels of "expertise, identity, achievement, and community involvement" [2]. Several studies define user roles along a core-periphery axis within the community they belong to. Less active users are generally a majority and are situated towards the periphery, whereas a smaller core of contributors takes additional responsibilities and administrative duties [2]. Roles are not fixed and community members tend to move across roles, often from the periphery to the centre [9, 30], but not necessarily. An analysis

[1] The Web Ontology Language (OWL) is a formal language commonly used to describe logical relations in ontologies [46].

of functional roles in Wikipedia [2] has shown that users move in both directions between higher and lower responsibility roles, albeit with less editors getting to roles closer to the centre.

Müller-Birn et al. [27] studied Wikidata user roles. Their work draws upon peer-production systems and collaborative ontology engineering literature to profile editors based on their revisions' scope (e.g. claims, labels, references) and type (e.g. addition or deletion). Human users are categorised into six profiles. Two of them are primarily engaged in modifications to the ontology (*Property editors* and *Property engineers*), whereas the others focus on adding and revising other elements, such as references and common Items. Bots have similar profiles, with the exception of Property engineers. Our work is complementary—we look at user profiles and how their editing change the quality of the ontology in time. [34] qualitatively analysed the evolution of users as they become established members of the Wikidata community, through the lenses of activity theory [21] and legitimate peripheral participation [23]. Seasoned users are more likely to have a central role in the community, participating more often in discussions on community pages, and mentoring less experienced peers. Furthermore, users with different levels of experience are likely to carry out different tasks. More established editors focus more on quality control and on building and maintaining the ontology. Some of them voice their concerns about common errors in the construction of taxonomic hierarchies done by novices, echoing what was reported in [8]. These errors often involve the use of an incorrect Property—e.g. *Ada Lovelace–instance of–computer scientist*, instead of *occupation*, or *MSF Canada–subclass of–Médecins Sans Frontières*, instead of *part of*—or of an incorrect object Item, although to a lesser extent. Other issues cause redundancies, rather than inconsistencies, e.g. when an Item is a sub-class of two Items, one of which is an instance of the other [8]. Whereas [34] sheds some light on the tasks undertaken by users with different levels of experience in Wikidata and hints at the existence of various user roles within the platform, the research is qualitative and based on a small sample of interviews. The current paper could be seen as a quantitative follow-up of that study, identifying Wikidata user profiles over time and understanding their impact on the part of Wikidata that seems to rely on experienced contributors, its ontology.

3 DATA

The Wikimedia Foundation releases dumps containing the history of each page within Wikidata. Everything within the platform, including Items and Properties, has a corresponding web page, which is assigned to a particular namespace. For each version of a page, called *revision*, several metadata fields are available: the name and identifier of the user responsible for the revision, the timestamp of the revision, and a comment. We extracted from the dumps the data updated to the 1 *October* 2017. We imported each revision of every statement within Wikidata into a PostgreSQL database to be able to find all revisions at any point in time along the lifespan of Wikidata. Deleted statements were appropriately tagged. The ontology graph was built on the basis of the P31 (*instance of*) and P279 (*subclass of*) Properties. Qualifiers were not included in the analysis, similarly to [11]. Administrators and bots were identified through the lists of users by group made available by the Wikimedia Foundation, cross-checking these with data extracted from the *Request for Permission* pages in Wikidata. Edits made through semi-automated tools were identified by looking at the comments and tags attached to each revision [36]. To examine the evolution of user activity and of ontology quality over time, we extracted monthly slices of the data and collected all variables for each slice. P31 was created in early *February* 2013, while P279 dates back from early *March* 2013. Hence, the first slice in our dataset is *March* 2013, the last is *September* 2017, for a total of 55 slices. The code and part of the datasets produced are available at https://github.com/Aliossandro/who_models_the_world_submission_1161_cscw.git.

4 METHODS

We carried out three related studies, each with their own methods: **Study 1**, to define a quality framework suitable for Wikidata based on existing literature, which we applied to evaluate the quality of the ontology over time; **Study 2**, to identify user roles based on their activity patterns, using the data from Section 3; the results of Study 1—a measure of the Wikidata ontology quality over time—and Study 2—roles and editing patterns of each user for every month—were linked together in **Study 3**, where we explored how specific user roles trigger changes in the quality of the Wikidata ontology. The findings of the first two studies were used to define hypotheses for the third. In the following, we report on the methods used by each of them.

4.1 Study 1 - Ontology quality

Assessing the quality of an ontology can follow various approaches [7, 17, 44]. Choosing an approach depends on context, including the purpose of the assessment and the available data [7]. The current work aims to gain a quantitative understanding of the effects of different editing patterns on the quality of the Wikidata ontology. Therefore, we sought an approach that

R1 considers primarily factors that editors could potentially influence (as opposed to externalities around the use of the ontology by developers, its suitability for a particular task, etc.);

R2 is able to assess the ontology over time to observe its evolution;

R3 uses only the ontology for the assessment and does not require additional task-based evaluations (a counterexample would be aspects such as completeness or coverage, which need to take into account a reference model or a gold standard);

R4 includes indicators that could be implemented unambiguously and be computed automatically (for example, aspects such as the understandability of an ontology can be assessed in various ways).

A basic distinction can be made with regard to the type of evaluation: *ontology validation* checks whether the ontology is fit for purpose and meets its requirements, whereas *ontology verification* focuses on the process (i.e. whether the ontology was built correctly) [44]. We follow an ontology validation approach; the processes used by the Wikidata community are defined and assessed by the community itself, according to their principles and models of governance. Ontology validation can refer to various layers of an ontology: its labels (or vocabulary), its syntax (the correct use of a format), its semantics (whether the modelling is appropriate), and many more [44]. Our analysis focuses on structural quality, which can be computed from the data published by Wikidata and are directly controlled by Wikidata editors. Other aspects, e.g. vocabulary, syntax, context are related to applications using Wikidata or to its data model and as such not connected to editing activities.

We surveyed the extensive research around structural ontology metrics to inform the design of our framework. In order to identify relevant papers, we crossed the results from queries to widely used academic literature search engines (i.e. Google Scholar and Web of Science) with the references found in a number of ontology evaluation surveys [7, 17, 28, 44]. We used the following keywords: 'ontology metrics', 'ontology evaluation framework', and 'ontology evaluation'. From the results, we selected only papers including primarily structural metrics. We evaluated each framework in our selection against the requirements set above (Table 1). The approaches in [31] and [40] were not suitable for our purposes, as the first focused on measuring the level of connectedness between ontology pairs and the second used a system of subjective ratings which was not compatible with our requirements. The other frameworks shared a number of metrics related to breadth, depth, and fitness of use compared to the rest of the knowledge graph, e.g. number of classes, average sub-class hierarchy depth, and average number of instances per class. For the purpose of our analysis, we picked a set of 14 indicators, primarily from [37]. The metrics in this work covered a wider range of aspects than those in [51] and [52], while being suitable to be implemented on Wikidata, which

Framework	R1	R2	R3	R4
OntoMetric [40]	No	Partially	No	Partially
Gangemi et al. [14]	Partially	Yes	No	Partially
OntoQA [39]	Yes	Yes	Partially	Yes
Orme et al. [31]	No	Yes	No	Yes
Sicilia et al. [37]	Yes	Yes	Partially	Yes
Yang et al. [53]	Yes	Yes	Yes	Yes
Yao et al. [51]	Yes	Yes	Yes	Yes

Table 1. Ontology metric frameworks evaluation against the requirements set in the present study

Indicator	Description	Feature
Number of instances ($\|I\|$)	Items used as subject of P31 but not as an object and not connected to any other Items through P279.	*noi*
Number of classes ($\|C\|$)	Items connected to at least another Item through P31 (as an object) or P279 (as a subject or object).	*noc*
Number of root classes	Classes for which no super-class exists ($\|C_i\|, \neg\exists C_j \| C_i \not\subseteq C_j$).	*norc*
Number of leaf classes	Classes that have at least a super-class for which no sub-class exists ($\|C_i\|, \neg\exists C_j \| C_j \not\subseteq C_i$).	*nolc*
Number of Properties ($\|p\|$)	Possible relations between Items.	*nop*
Population	Number of instances per class.	*ap* (average)
		mp (median)
Class richness	Ratio between classes with instances and all instances ($\frac{\|C'\|}{\|C\|}$).	*cr*
Inheritance richness	Number of sub-classes per class.	*ir* (average)
		mir (median)
Relationship richness	Ratio between number of relations of classes except sub-class relations and all relations ($\frac{\|P\|}{\|SC\|+\|P\|}$).	*rr*
Class hierarchy depth	Explicit depth of class hierarchy. Class hierarchies are formed by chains of sub-class relations.	*ad* (average)
		md (median)
		maxd (max)

Table 2. Wikidata quality indicators [37] used in the present analysis

lacks the stringent logical definitions typical of formally-defined ontologies. We adjusted each of the metrics in our set to capture the ontology model of Wikidata.

Our Wikidata ontology quality framework consists of 14 features, shown in Table 2. The features counting the number of classes (*noc*, *norc*, *nolc*) and Properties (*nop*) measure the ontology size. *norc* is a count of the classes that have sub-classes, but no explicitly stated super-class. It can assume values between 1 and $|C|$, where higher values suggest more diverse knowledge in the ontology [37]. *nolc* refers to the classes with no sub-classes. *noc* and *nolc* increasing at a pace comparable to that of the total of the Items in the graph would be a sign of a growing number of Items erroneously treated as classes, confirming prior findings around the misuse of P31 and P279. A slower growth would indicate that the taxonomy is actively and successfully maintained. Experienced users focus on this task according to [34]. *noi* counts the Items for which type information is specified, i.e. which are subject of P31 statements, and ideally should be equal or close to the total of Items.

The average population of an ontology *ap* measures how instances are distributed across the classes of the ontology. Together with class richness (*cr*), it provides an indication of whether the information in the ontology is sufficient to describe the data [37]. Low *cr* and *ap* may mean that the instances in the KG do not represent all the knowledge in the ontology [39]. This seems unlikely in Wikidata, where the ontology is not built separately from the rest of the graph, and could hence be attributed to the misuse of taxonomic properties. Stable or increasing *cr* and *ap* may be a sign of successful efforts to maintain the ontology by a part of the Wikidata community, whereas decreasing values would point to empty classes continuously added without sufficient quality checks. Relationship Richness or *rr* is the ratio between the overall number of relations of the entities in an ontology, divided by the sum of the sub-class relations plus the overall relations. Values closer to 1 are characteristic of rich ontologies, whereas lower values indicate ontologies containing mostly taxonomic relations [37]. Inheritance richness (*ir*) is used to understand how knowledge is distributed across the different branches and levels of the ontology. It measures the average number of sub-classes per class. High *ir* values would correspond to a shallower ontology, with classes that tend to represent more general knowledge, whereas lower values typically reflect very specialised, vertical ontologies [37]. Hierarchy depth metrics (*ad*, *md*, and *maxd*) describe the length of (explicit, as opposed to automatically inferred) sub-class relations paths in the taxonomy [44]. Deeper ontologies are often seen as more reliable [13], although they may result in being less understandable and usable [14]. Continuously increasing ontology depth and *ir* may point to an ontology becoming too specialised and overly convoluted. We may expect the Wikidata ontology to have unequal depth, with low average depth and high *maxd* values increasing over time. The set in [37] included also the OntoRank metric. We did not include that, as it did not apply to Wikidata.

4.2 Study 2 - User roles

The k-means algorithm has been used in several studies to identify emerging user roles [12, 25, 27]. In our case, it helped us cluster human registered editors according to several features (Table 3). The choice of features was informed by prior studies about community dynamics in Wikidata and other platforms. According to what reported in [34], established users perform more revisions (*# edits*), undertake more often tasks related to the maintenance of the ontology (*# ontology edits*, *# taxonomy edits*, and *# Property edits*), patrol the graph to correct errors and uphold quality (*# modifying edits*), and interact more with the community (*# discussion edits*). Because of the varying levels of activity across months, we normalised these variables by their monthly totals, in order to make them comparable across different time frames. Experienced editors are also more likely to carry out more revisions through semi-automated tools, therefore we included the proportion of these types of edits (*p batch edits*). Moreover, the activity of Wikidatians vary also in terms of focusing on a larger or smaller range of Items [34]. This is taken into account by the *Item diversity* feature. Finally, formal roles have been connected to functional roles in other platforms such as Wikipedia [2]. We added two binary variables indicating whether a contributor belonged to an administrative user group (e.g. *bureaucrats, administrators, or stewards*) or to any other user group with extended user rights (e.g. *Property creators or rollbackers*) in a determined time frame. Many approaches exist to determine the appropriate k in unsupervised clustering algorithms. We used the *gap statistic*, which compares a measure of variance with a null reference distribution of the data [41]. This measure has outperformed several other approaches in the literature.

Feature	Description	Feature	Description
# edits	Total number of edits in a 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 semi-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.

Table 3. Features used to cluster users

4.3 Study 3 - Relationship between user roles and ontology quality

We formulated two research hypotheses linking user role features to the ontology quality metrics based on the findings from the first two studies. We looked at this relation in terms of how the activity of a determined user type, specifically leaders, results in changes in various ontology metrics. Our approach follows [19] and consists in applying a lagged multiple regression model to predict changes in an ontology metric considered between two points in time $metricT_n - metricT_{n-1}$, holding it constant at $metricT_{n-1}$. This approach has the advantage to control for regression towards the mean and to remove the influence of $metricT_{n-1}$ on the relation between predictors and the dependent variable [19]. The metrics chosen for the dependent variables, computed as $metricT_n - metricT_{n-1}$, represented ontology breadth (*noc*, *norc*, *nolc*), depth (*ad*), and distribution of instances and classes (*ir* and *ap*). The independent variables were the number of edits made by leaders and by contributors. As control variables, besides the initial value of the dependent variable ($metricT_{n-1}$) we added to each model the number of bot and anonymous edits, which have been shown to influence the quality of Wikidata [33]. Independent and control variables had different scales and were standardised to have mean 0.

5 RESULTS

5.1 Study 1 - Ontology quality

We implemented the indicators based on the specification from Table 2 and computed their values on the data from Section 3. Summary statistics about the current state of the graph are presented in Table 4. The Wikidata ontology has roughly 1.5 million classes, about 2000 times larger than the DBpedia ontology and three times than YAGO [32] (see also Section 2.1). With regard to Properties, Wikidata enables the expression of a vast number of relations, exceeding both that of DBpedia and YAGO (Section 2.1). *noi* is smaller than the total number of Items, meaning that a large number of entities have no defined type information. Concerning other indicators, the differences between mean and median values suggest that instances (*ap* and *mp*) and sub-classes (*ir* and *mir*) are unevenly distributed across the ontology. As regards the evolution of the ontology, all indicator series were tested for significance using a 1-sample t-test against the null hypotheses that no trends took place over the time span observed. All values in Figures 2, 3, and 4 presented a significant trend, except average population (*ap*), max depth (*maxd*), and the medians, whose value

Feature	Value	Feature	Value	Feature	Value
Total Items	38, 621, 989	*nop*	3589	*rr*	0.9
noi	32, 858, 649	*ap*	23.6	*cr*	0.04
noc	1, 465, 639	*mp*	0 (0; 0)	*ad*	46.3
nolc	1, 349, 963	*ir*	1.9	*md*	51 (35; 61)
norc	26, 265	*mir*	0 (0; 0)	*maxd*	96

Table 4. Ontology metrics figures at 1*October* 2017. In brackets, 25^{th} and 75^{th} percentiles.

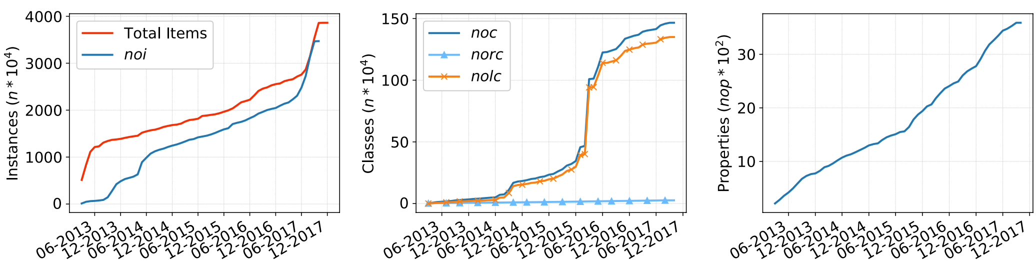

Fig. 2. Evolution of number of entities, classes, and Properties in Wikidata over time

was zero. A very large number of classes has no sub-classes at all. This trend is visible since the early months of Wikidata. The totals of classes (*noc*) and leaf classes (*nolc*) increase in a similar fashion, both presenting a spike at the beginning of 2016, and reaching comparable values. By contrast, root classes (*norc*) increase slowly but constantly, reaching a maximum value of $\sim$ 26 thousand (Figure 2). Most other indicators tell a similar story. Both population (*ap*) and class richness (*cr*) decrease in the time span considered, suggesting that several classes are either without instances or sub-classes or both (Table 2). The median number of sub-classes (*mir*) is stable around zero. If we look at variables related to the size of the taxonomic hierarchy, such as inheritance richness (*ir*) and max depth (*maxd*), it seems that a part of the Wikidata ontology is distributed vertically. *maxd* increases over time, reaching values higher than 80 (Figure 4), i.e. hierarchies with more than 80 levels. Finally, Wikidata classes seem to be well defined, with several relations besides P279, which can be seen in the rapid increase of relationship richness (*rr*).

5.2 Study 2 - User roles

We collected a total of 783, 604 user/timeframes, corresponding to 190, 765 unique human registered users working over 55 months. The average time people contribute to Wikidata was four months (median 1, interquartile range 1,3). 28, 600 users were active across more than five months and around 18, 000 across more than ten. 143 performed edits throughout the entire time span considered.

We tested values for k between 2 and 8. The maximum value was chosen in order to keep the number of clusters manageable. The gap statistic returned an optimal $k = 2$. An independent t-test across the two clusters showed that the mean of each variable differed significantly, with the exception of *admin* and *lower admin*. We called these roles *contributor* and *leader*. The large majority of user/timeframes (771, 044) fell in the first role, characterised by lower levels of activity for all variables. On average, contributors perform a lower number of edits, preferably without semi-automated editing tools, and participate less in community discussions. Leaders have more sustained editing activity and deeper engagement in community pages. They devote more effort to revising properties and taxonomic relations (P31 and P279 statements). Besides, they are more active in modifying previously added content and edit more often through semi-automated tools.

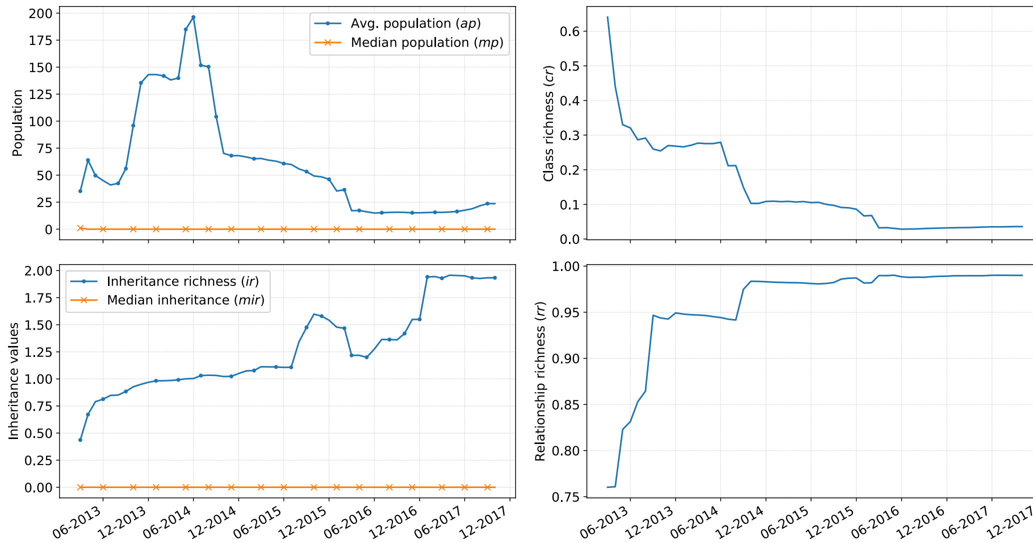

Fig. 3. Wikidata quality assessment

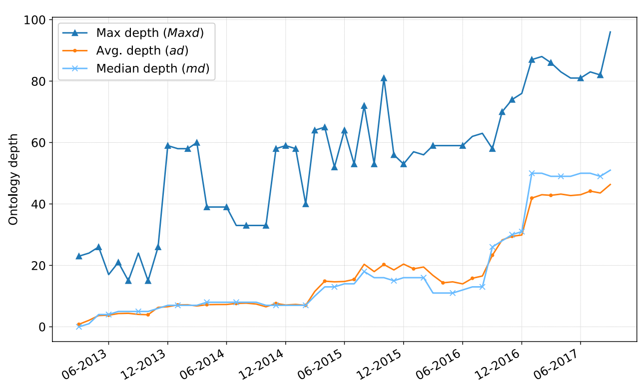

Fig. 4. Ontology depth

It is important to note users can morph into different roles over time. Editors with at least one month in the leader role had on average a higher activity lifespan on the platform (~ 9 months; interquartile range 1, 11) than editors without (~ 4 months; interquartile range 1, 2). A Mann-Whitney test confirmed that the difference between these two groups was significant ($p < 0.05$). Aggregate analyses over the entire history of online communities can lead to misrepresenting some of their underlying dynamics, due to Simpson's paradox [4]. Hence, following the approach in [4], we looked at user roles dynamics by year of joining the platform. We divided editors in five yearly cohorts, from Wikidata's launch (*October* 2012) to the end of our dataset (*September* 2017), each including users joining from October of one year to September of the following (e.g. *October* 2012 to *September* 2013, etc.). Users who joined earlier in the project (before *September* 2013) are more active over all time frames considered (Figure 5.c) and are the most likely to take a leader role (Table 5). For all cohorts, the percentage of leaders peaks initially, followed by a drop after around a year. This is less pronounced for 2012–2013 editors (Figure 5.d), but it must be considered that our analysis starts on *March* 2013, leaving out the first months of their activity.

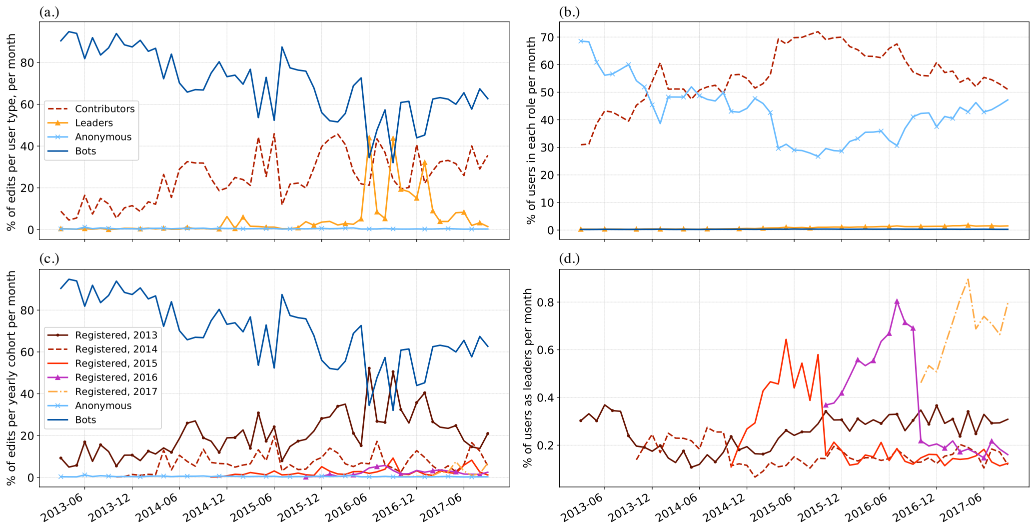

Fig. 5. Proportion of contributions per user type and by yearly cohort over time and percentage of users per type. The count of anonymous users is based on unique IP addresses, as these users are only known through them. Nothing prevents editors to connect from different addresses, though. Years in (c.) refer to the period between October of the previous year and September of the following (e.g. 2013 means Oct. 2012–Sep. 2013).

first joined	**# users**	**# leaders**	$p(leader \mid first\ joined)$
Oct. 2012–Sep. 2013	26,492	1540	0.058
Oct. 2013–Sep. 2014	38,679	1159	0.029
Oct. 2014–Sep. 2015	39,318	1430	0.036
Oct. 2015–Sep. 2016	43,714	1796	0.041
Oct. 2016–Sep. 2017	42,562	2187	0.051
Total	190,765	8112	0.042

Table 5. Breakdown of users and leaders by yearly cohort. By leader, we refer to anyone who has taken on a leader role in at least one time frame.

5.3 Study 3 - User profiles and ontology quality

The third contribution of this paper concerns the influence of each Wikidata user role on the quality of the Wikidata ontology. Prior work about Wikidata has highlighted different sets of users that take on quality control tasks concerning the conceptual structure of the graph. Experienced editors report in [34] to feel as part of their duties the maintenance and cleaning of the ontology, compared to when they were novices on the platform. Müller-Birn et al. [27] identify two roles that are mainly focused on creating and editing Properties. Similar roles have been identified also in collaborative ontology development projects [12], where some editors have been shown to work primarily on organisational and hierarchy-cleaning tasks.

In the second study, we have identified two roles that Wikidata editors may take: contributor and leader. Leaders perform more higher-responsibility tasks, including quality control and maintenance of the ontology. Several users move between the two roles over time, although only a minority ever shows a leader activity pattern. Based on the studies cited above, we assume that these users are more familiar with quality issues around the Wikidata ontology and their work on the ontology improves its quality, contrasting harmful behaviour from other users, deleting incorrect P31/P279

	noc			*norc*			*nolc*		
	Coef.	SE	*p*	Coef.	SE	*p*	Coef.	SE	*p*
Intercept	0.002	9927	$*$	477.5	38.50	$**$	0.002	9840	
Initial *noc*|*norc*|*nolc*	0.001	0.002		−96.06	98.21		0.001	0.002	
# contributor edits	−6054	0.001		42.69	93.55		−7533	0.001	
# leader edits	−2968	0.001		65.84	49.69		−4592	0.001	
# anonymous edits	0.002	0.001		−0.956	46.44		0.002	0.001	
# bot edits	−0.001	0.001		19.36	41.76		−0.001	0.001	
	ir			*ap*			*ad*		
	Coef.	SE	p	Coef.	SE	p	Coef.	SE	p
Intercept	0.035	0.012		−0.430	2.087		0.842	0.260	$***$
Initial *ir*|*ap*|*ad*	−0.004	0.024		1.826	2.760		0.270	0.552	
# contributor edits	−0.031	0.024		−3.316	2.767		−0.251	0.544	
# leader edits	0.038	0.13	$**$	−3.316	2.395		1.325	0.301	$***$
# anonymous edits	−0.021	0.014		−1.034	2.571		−0.400	0.318	
# bot edits	0.039	0.013	$**$	5.727	2.267	$*$	0.348	0.293	

$***p < .001$, $**p < .01$, $*p < .05$.

Table 6. Lagged regression analysis of proportion of activity of each user type on *noc*, *norc*, *nolc*, *ir*, *ap*, and *ad*

statements and working as 'ontology curators', e.g. creating new hierarchies and replacing *instance of* with *subclass of* relations (and vice versa) when needed. This results in two 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*).

The lagged regression analysis did not show any significant relation between *noc*, *norc*, *nolc*, and leader activity levels (Table 6). **H1** was thus **not supported**. **H2** was **partially supported**. Leader edits were significantly and positively associated with hierarchy depth (*ad*) and inheritance richness (*ir*). The latter, as well as average population (*ap*), were also positively related to bot edits.

6 DISCUSSION

6.1 Ontology Quality

Different dynamics emerge from our evaluation of the Wikidata ontology. It grows over time, in terms of numbers of classes (*noc*), leaf classes (*nolc*), and Properties (*nop*). It also appears to be unevenly maintained and more so as the knowledge graph grows. The combined dynamics of class richness (*cr*) and average population (*ap*), simultaneously dropping in *June* 2014 and both dwindling afterwards, together with the difference between *ap* and the median number of instances per class *mp* (Figure 3) suggest that an increasingly high number of empty classes exists next to a vastly populated core of the ontology with deep taxonomic hierarchies, as average (*ad*) and max depth (*maxd*) show. Whereas a share of the ontology growth may be attributed to erroneous revisions such as those described in Section 2.2 and discussed in [8, 34], these are not the only reason behind the trends observed. An examination of the peaks in Figures 2 and 3 revealed that a single bot (MicrobeBot) added ∼ 400 thousand classes in *March* 2016, causing the sudden rise of *noc* and *nolc*. Two bots (SuccuBot and PLBot) are responsible for *June* 2014's peak in *ap*, which subsequently falls due to an increase in the number of classes. These mass revisions are not necessarily incorrect. Bots are often programmed to carry out edits within a particular domain, creating new statements by extracting information from a source. MicrobeBot created numerous new classes by adding sub-classes statements to as many protein and gene Items. These edits are not formally incorrect, yet they are questionable under a knowledge engineering point of view, as it is highly unlikely that instances will ever added to these classes. Describing a concept as a class or an instance determines the properties and attributes that can be applied to it and it is an essential part of building an ontology [29]. Prior literature has discussed the tension between freedom and standardisation in projects collaboratively producing structured data [15]. The liberal policies of Wikidata put virtually no restrictions on the edits users can make. This liberal approach may potentially lead to imbalances in the conceptual knowledge of the graph such as those identified. Quality metrics such as those from our framework can address this issue, by monitoring the evolution of the ontology.

Quality metrics. The metrics used in this paper have been variously related to different characteristics of ontologies. Deeper ontologies—Wikidata's taxonomic tree has an average depth of 40 sub-class relation—are likely to be harder to understand and to be manipulated by users [14], although they have been associated with more reliable semantic content [13]. Others have found connections between ontology accuracy, inheritance (ir), and relationship richness (rr) [22]. Nevertheless, our quality framework is limited in expressiveness. Whereas the structural indicators considered work well with the observational data provided by Wikidata and are able to illustrate trends over time, they are hardly comparable across ontologies and do not provide any direct insight into the correctness of the conceptualisation [44]. To deal with the first, we could consider normalised indicators [44]. For the second, we could try to detect inconsistencies, either by inspecting samples of the class hierarchy [43] or by using reasoning software—however, the size of the ontology makes both tasks extremely challenging for state of the art tools. Further on, our framework does not consider multilingual aspects of Wikidata. A comparison between Wikipedia articles has found that different language versions overlap only to a very small extent, with regard to the topics they cover [16]. A similar issue was noted in Open Street Map [15], where users often found the knowledge organisations of editors with different cultural backgrounds not meeting their specific needs. Our framework could be extended to include indicators that consider structural changes for different language spaces as well. To sum up, the metrics computed provide only a partial picture of the quality of the Wikidata ontology. Yet, it is an important part. First, our findings may be a starting point for future studies that want to explore differences in quality between domains in Wikidata conceptual knowledge. Second, the information provided by our metrics may be used to test future design solutions. For example, measures to make the effects of any edit on the hierarchy may be adopted to address the misuse of taxonomic relations, following a suggestion regarding collaborative ontology development contexts in [42]. An analysis of the metrics selected in this work may be subsequently used to assess the success of such approach.

6.2 User Roles

We detected the existence of two distinct user roles in Wikidata: leaders and contributors. Compared to the latter, the first perform on average more edits on classes, carry out more maintenance work, and get more involved in community discussions. Editors move up and down between these roles. In each time frame of the period examined, leaders are only a small minority ($\sim 1\%$) and that their overall revisions are generally less than those from contributors. When this is not the case (*June* and *August* 2016, *January* 2017, see Figure 5.a), the usual proportions are reverted by single human editors who perform revisions at a high rate ($\sim$ 2M per month) through semi-automated tools. This reinforces the considerations made in Section 6.1 concerning the benefits of using our metrics to continuously monitor the quality of the Wikidata ontology. Early editors are more likely to take a leader role (Table 5), a behaviour that has been noted also among users of other platforms (e.g. Reddit [4]) and may be explained by a self-selection bias, i.e. users arriving to Wikidata in its early stage may be on average those who are more inclined to like it [4]. After a first drop, the percentage of early users taking a leader role remains higher than other cohorts (Figure 5.d), which may confirm this hypothesis. Moreover, each cohort shows an initial peak of users working as leaders a few months after joining, followed by a decline. We return on that in the next paragraphs.

Compared to prior studies on roles in Wikidata or in collaborative ontology engineering projects, the Wikidata community seems less structured. A direct comparison may not be appropriate though, since our conceptualisation is built along different dimensions than earlier ones. Whereas the analysis in [27] looks at edit types with more granularity, distinguishing between revisions on the various parts on an Item, we included features related to editors' interaction within the community and use of interfaces, i.e. semi-automated tools. Similar considerations can be done concerning research on collaborative ontology engineering roles, which relies mainly on structural aspects of the concepts edited, e.g. their depth or their function in the ontology hierarchy [12, 47].

The Reader-to-Leader (R2L) framework [35] describes people's behaviour in online platforms and has been largely influential in the CSCW community. Thanks to its generalisability, it has been applied and empirically validated on numerous collaborative platforms, e.g. Wikipedia; in addition, it explains transitions between roles. We discuss our findings through the lens of this framework and of previous empirical studies that rely on that, focusing on two aspects: (*i.*) the articulation of roles, also as regards administrative responsibilities; (*ii.*) the transitions between roles.

Regarding (*i.*), the R2L framework categorises users into four roles, *reader*, *contributor*, *collaborator*, *leader*, ordered according to their centrality along a core-periphery axis. Readership may prove difficult to detect in Wikidata, as the relevant data can be accessed through APIs and tools that are much less mature than those used in this work. Contributors in R2L move their first steps on the platform, performing small edits and slowly engage in relationships with the other users [35]. Collaborators are more established members of the community, who cooperate with other participants to achieve more complex outcomes. Wikidata contributors present a mix of the features of these two roles. Some editors perform a large number of revisions with semi-automated tools, but do not comment on talk pages; others are less active, but engage more often with the community; other users edit less, but stay active for longer times. This diversified behaviour, which characterises also leaders, stands out when compared to other studies, e.g. to Arazy et al.'s analysis of Wikipedia's participation dynamics [2]. In their work, underpinned by the R2L, the authors bring together organisational and functional features, showing that specific activity patterns correspond to each of the nine Wikipedia's formal roles. On the contrary, formal roles—administrators—were not distinguishing features for the roles detected in the current study. Wikidata is still an emerging project, therefore its community structure might still be evolving to more defined patterns of activity. However, there is evidence from our analysis that its evolution has so far been distinct from Wikipedia. Whereas in the online encyclopaedia the administrators' share of edits increased in the first years [18], this rise is yet to happen in Wikidata, if it will happen at all. Moreover, Wikidata has still a lower normative burden of rules and policies than Wikipedia [34]. A simpler bureaucracy may require fewer declared roles, which in turn changes how and how much people participate. Our findings demonstrate that Wikidata's sociotechnical fabric diverges from prior projects in both the areas of peer-production and collaborative knowledge engineering and may actually represent a new paradigm of collaborative system. Under a methodological viewpoint, the clustering algorithm used (k-means) may not perform well with some types of clusters, e.g. non-spherical or unevenly distributed ones. Different algorithms may give different results, although each has its strengths and weaknesses. Future studies should test other approaches to compare their performances.

For what concerns (*ii.*), i.e. role transitions, in R2L they happen in both directions from periphery to core, although mainly in a sequential fashion [2, 35]. The minority of Wikidata editors who ever work as leaders move up and down between roles, seemingly 'stepping up' at some point to provide greater support to the project. Several users follow the path to leadership in their first months, but yearly cohorts in Figure 5.d show that they often do not beat that path again. This may be a sign of declining participants' motivation—one challenge initiatives such as Wikidata and Wikipedia may face is the lack of shorter-term, tangible goals and achieving a specific editor status might act as a proxy to them [20]. Our analysis, alongside previous studies [27], could inform the definition of these 'badges' and help study their uptake and effects.

6.3 Ontology quality and user roles

Our results partially support the qualitative findings from [34]. We did not find any significant influence of users in the contributor role on any of the metrics analysed. Leader activity is significantly related to the average number of sub-classes per class (*ir*) and depth of the ontology (*ad*), which may be interpreted as a confirmation that more seasoned users concentrate their efforts on vertically extending and consolidating the ontology to ensure that all entities in the KG are optimally represented and organised into classes. Their efforts appear to have limited effect though. Appropriate tools may be designed to help them control Wikidata's quality. Bot edits are positively related to the average number of sub-classes per class (*ir*). To a certain extent this can be attributed to the introduction of several Wikidata bots, which e.g. add data from other sources into Wikidata automatically. What is surprising is the lack of any substantial influence on the total number of classes (*noc*), reinforcing the impression that Wikidata user dynamics differ from those observed in prior collaborative ontology engineering projects.

7 CONCLUSION

This work is the first to address the structural quality of the Wikidata ontology over time and the community dynamics that influence it. Its contribution is threefold. First, we have devised and applied a set of indicators to assess the Wikidata ontology from its early days until *September* 2017. Second, we have identified roles for Wikidata editors according to their activity patterns. Finally, we have explored how these roles influence the quality of the ontology.

The Wikidata ontology is large and messy, with numerous underpopulated classes and uneven depth. This confirms prior literature suggesting that several Wikidata contributors fail to use correctly the taxonomic relations P31 (*instance of*) and P279 (*subclass of*). On the other hand, we found evidence suggesting that parts of the ontology have higher depth and are likely to be curated by a core of expert users. We identified two activity patterns: *contributors*, i.e. users with lower number of edits and less engaged in community discussions, and *leaders*, who are more active in all of the features considered. Only a minority of users presents a leader activity pattern at any time during their interaction with the platform. Finally, whereas the activity of leaders seems to influence positively the depth of the ontology, no relation could be proven between any editor category and variables concerning the breadth of the ontology. Future work should explore what variables are at play in that regard.

The use of KGs to empower any kind of information-based system has become widespread in recent years and is likely to grow further. The communities behind them need appropriate tools and frameworks to produce high-quality data. For that purpose, it is important to understand how knowledge is created within these communities and by whom. This paper is a step in this direction.

ACKNOWLEDGMENTS

This project is supported by funding received from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 642795 (WDAqua ITN).

REFERENCES

[1] B. Thomas Adler and Luca de Alfaro. 2007. A content-driven reputation system for the Wikipedia. In *WWW*. ACM, 261–270.

[2] Ofer Arazy, Felipe Ortega, Oded Nov, M. Lisa Yeo, and Adam Balila. 2015. Functional Roles and Career Paths in Wikipedia. In *CSCW*. ACM, 1092–1105.

[3] Michael Ashburner, Catherine A Ball, Judith A Blake, David Botstein, Heather Butler, J Michael Cherry, Allan P Davis, Kara Dolinski, Selina S Dwight, Janan T Eppig, et al. 2000. Gene Ontology: tool for the unification of biology. *Nature genetics* 25, 1 (2000), 25.

[4] Samuel Barbosa, Dan Cosley, Amit Sharma, and Roberto M. Cesar Jr. 2016. Averaging Gone Wrong: Using Time-Aware Analyses to Better Understand Behavior. In *WWW*. ACM, 829–841.

[5] Jonathan BL Bard and Seung Y Rhee. 2004. Ontologies in biology: design, applications and future challenges. *Nature Reviews Genetics* 5, 3 (2004), 213.

[6] Kurt D. Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD Conference*. ACM, 1247–1250.

[7] Janez Brank, Marko Grobelnik, and Dunja Mladenić. 2005. A survey of ontology evaluation techniques. In *Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD)*.

[8] Freddy Brasileiro, João Paulo A. Almeida, Victorio Albani de Carvalho, and Giancarlo Guizzardi. 2016. Applying a Multi-Level Modeling Theory to Assess Taxonomic Hierarchies in Wikidata. In *WWW (Companion Volume)*. ACM, 975–980.

[9] Linus Dahlander and Siobhan O'Mahony. 2011. Progressing to the center: Coordinating project work. *Organization science* 22, 4 (2011), 961–979.

[10] Martin Doerr. 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine* 24, 3 (2003), 75.

[11] Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. 2014. Introducing Wikidata to the Linked Data Web. In *Semantic Web Conference (1) (Lecture Notes in Computer Science)*, Vol. 8796. Springer, 50–65.

[12] Sean M. Falconer, Tania Tudorache, and Natalya Fridman Noy. 2011. An analysis of collaborative patterns in large-scale ontology development projects. In *K-CAP*. ACM, 25–32.

[13] Miriam Fernández, Chwhynny Overbeeke, Marta Sabou, and Enrico Motta. 2009. What Makes a Good Ontology? A Case-Study in Fine-Grained Knowledge Reuse. In *ASWC (Lecture Notes in Computer Science)*, Vol. 5926. Springer, 61–75.

[14] Aldo Gangemi, Carola Catenacci, Massimiliano Ciaramita, and Jos Lehmann. 2006. Modelling Ontology Evaluation and Validation. In *ESWC (Lecture Notes in Computer Science)*, Vol. 4011. Springer, 140–154.

[15] Andrew Hall, Sarah McRoberts, Jacob Thebault-Spieker, Yilun Lin, Shilad Sen, Brent Hecht, and Loren Terveen. 2017. Freedom vs Standardization: Structured Data Generation in a Peer Production Community. In *CHI*. ACM, 6352–6362.

[16] Brent J. Hecht and Darren Gergle. 2010. The tower of Babel meets web 2.0: user-generated content and its applications in a multilingual context. In *CHI*. ACM, 291–300.

[17] Hlomani Hlomani and Deborah Stacey. 2014. Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal* 1, 5 (2014), 1–11.

[18] Aniket Kittur, Ed Chi, Bryan A Pendleton, Bongwon Suh, and Todd Mytkowicz. 2007. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. *World Wide Web* 1, 2 (2007), 19.

[19] Aniket Kittur and Robert E. Kraut. 2008. Harnessing the wisdom of crowds in Wikipedia: quality through coordination. In *CSCW*. ACM, 37–46.

[20] Robert E Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design.* Mit Press.

[21] Kari Kuutti. 1996. Activity theory as a potential framework for human-computer interaction research. *Context and consciousness: Activity theory and human-computer interaction* 1744 (1996).

[22] Birger Lantow. 2016. OntoMetrics: Application of On-line Ontology Metric Calculation. In *BIR Workshops (CEUR Workshop Proceedings)*, Vol. 1684. CEUR-WS.org.

[23] Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation.* Cambridge University Press.

[24] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. 2015. DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web* 6, 2 (2015), 167–195.

[25] Jun Liu and Sudha Ram. 2011. Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Trans. Management Inf. Syst.* 2, 2 (2011), 11:1–11:23.

[26] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M. Suchanek. 2015. YAGO3: A Knowledge Base from Multilingual Wikipedias. In *CIDR*. www.cidrdb.org.

[27] Claudia Müller-Birn, Benjamin Karran, Janette Lehmann, and Markus Luczak-Rösch. 2015. Peer-production system or collaborative ontology engineering effort: what is Wikidata?. In *OpenSym*. ACM, 20:1–20:10.

[28] Juan Francisco García Navarro, Francisco José García-Peñalvo, and Roberto Therón. 2010. A Survey on Ontology Metrics. In *WSKS (1) (Communications in Computer and Information Science)*, Vol. 111. Springer, 22–27.

[29] Natalya F Noy, Deborah L McGuinness, et al. 2001. Ontology development 101: A guide to creating your first ontology. (2001).

[30] Siobhán O'Mahony and Fabrizio Ferraro. 2007. The emergence of governance in an open source community. *Academy of Management Journal* 50, 5 (2007), 1079–1106.

[31] Anthony M. Orme, Haining Yao, and Letha H. Etzkorn. 2006. Coupling Metrics for Ontology-Based Systems. *IEEE Software* 23, 2 (Mar 2006), 102–108.

[32] Heiko Paulheim. 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8, 3 (2017), 489–508.

[33] Alessandro Piscopo, Christopher Phethean, and Elena Simperl. 2017. What Makes a Good Collaborative Knowledge Graph: Group Composition and Quality in Wikidata. In *SocInfo (1) (Lecture Notes in Computer Science)*, Vol. 10539. Springer, 305–322.

[34] Alessandro Piscopo, Christopher Phethean, and Elena Simperl. 2017. Wikidatians are Born: Paths to Full Participation in a Collaborative Structured Knowledge Base. In *HICSS*. AIS Electronic Library (AISeL).

[35] Jennifer Preece and Ben Shneiderman. 2009. The Reader-to-Leader framework: Motivating technology-mediated social participation. *AIS transactions on human-computer interaction* 1, 1 (2009), 5.

[36] Cristina Sarasua, Alessandro Checco, Gianluca Demartini, Djellel Difallah, Michael Feldman, and Lydia Pintscher. 2018. The Evolution of Power and Standard Wikidata Editors: Comparing Editing Behavior over Time to Predict Lifespan and Volume of Edits. *Journal of Computer Supported Cooperative Work* (2018).

[37] Miguel-Ángel Sicilia, Daniel Rodríguez, Elena García Barriocanal, and Salvador Sánchez Alonso. 2012. Empirical findings on ontology metrics. *Expert Syst. Appl.* 39, 8 (2012), 6706–6711.

[38] Thomas Steiner. 2014. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata. In *OpenSym*. ACM, 25:1–25:7.

[39] Samir Tartir and Ismailcem Budak Arpinar. 2007. Ontology Evaluation and Ranking using OntoQA. In *ICSC*. IEEE Computer Society, 185–192.

[40] Adolfo Lozano Tello and Asunción Gómez-Pérez. 2004. ONTOMETRIC: A Method to Choose the Appropriate Ontology. *J. Database Manag.* 15, 2 (2004), 1–18.

[41] Robert Tibshirani, Guenther Walther, and Trevor Hastie. 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63, 2 (2001), 411–423.

[42] Markel Vigo, Caroline Jay, and Robert Stevens. 2015. Constructing Conceptual Knowledge Artefacts: Activity Patterns in the Ontology Authoring Process. In *CHI*. ACM, 3385–3394.

[43] Johanna Völker, Denny Vrandečić, and York Sure. 2005. Automatic evaluation of ontologies (AEON). In *International Semantic Web Conference*. Springer, 716–731.

[44] Denny Vrandečić. 2010. *Ontology evaluation*. Ph.D. Dissertation. Karlsruhe Institute of Technology.

[45] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (2014), 78–85.

[46] W3C OWL Working Group. 2012. OWL 2 Web Ontology Language Document Overview (Second Edition) - W3C Recommendation 11 December 2012. (Dec. 2012). http://www.w3.org/TR/owl2-overview/

[47] Simon Walk, Philipp Singer, Markus Strohmaier, Tania Tudorache, Mark A. Musen, and Natalya Fridman Noy. 2014. Discovering Beaten Paths in Collaborative Ontology-Engineering Projects using Markov Chains. *Journal of Biomedical Informatics* 51 (2014), 254–271.

[48] Howard T. Welser, Eric Gleave, Danyel Fisher, and Marc A. Smith. 2007. Visualizing the Signatures of Social Roles in Online Discussion Groups. *Journal of Social Structure* 8 (2007).

[49] Wikidata. 2018. Wikidata:Bots. https://www.wikidata.org/wiki/Wikidata:Bots. (2018). [Online; accessed 08-April-2018].

[50] Wikidata. 2018. Wikidata:User access levels. https://www.wikidata.org/wiki/Wikidata:User_access_levels. (2018). [Online; accessed 08-April-2018].

[51] Haining Yao, Anthony Mark Orme, and Letha Etzkorn. 2005. Cohesion metrics for ontology design and application. *Journal of Computer Science* 1, 1 (2005), 107–113.

[52] Jonathan Yu, James A Thom, and Audrey Tam. 2007. Ontology evaluation using Wikipedia categories for browsing. In *Proceedings of the sixteenth ACM conference on Conference on Information and Knowledge Management*. ACM, 223–232.

[53] Yang Zhe, Dalu Zhang, and YE Chuan. 2006. Evaluation metrics for ontology complexity and evolution analysis. In *e-Business Engineering, 2006. ICEBE'06. IEEE International Conference on*. IEEE, 162–170.