The Role of Local Content in Wikipedia: A Study on Reader and Editor Engagement[1]

Marc Miquel[2], David Laniado[3] and Andreas Kaltenbrunner[4]

Abstract. About a quarter of each Wikipedia language edition is dedicated to representing "local content", i.e. the corresponding cultural context —geographical places, historical events, political figures, among others—. To investigate the relevance of such content for users and communities, we present an analysis of reader and editor engagement in terms of pageviews and edits. The results, consistent across fifteen diverse language editions, show that these articles are more engaging for readers and especially for editors. The highest proportion of edits on cultural context content is generated by anonymous users, and also administrators engage proportionally more than plain registered editors; in fact, looking at the first week of activity of every editor in the community, administrators already engage correlatively more than other editors in content representing their cultural context. These findings indicate the relevance of this kind of content both for fulfilling readers' informational needs and stimulating the dynamics of the editing community.
Keywords: Wikipedia; Online Communities; User Engagement; Culture; Cultural Context; Online Collaboration

[es] El rol del contenido local en Wikipedia: un estudio sobre la participación de lectores y editores

Resumen. Aproximadamente una cuarta parte de cada Wikipedia está dedicada a representar "contenido local", es decir, el contexto cultural correspondiente de la lengua —lugares geográficos, hechos históricos, figuras políticas, entre otros—. Este estudio presenta un análisis que se centra en la participación del lector y el editor en contenido local en términos de páginas vistas y ediciones, y examina la proporción de ediciones en artículos relacionados con el contexto cultural. Los resultados muestran que estos artículos obtienen una mayor atención por parte de los lectores, y especialmente de los editores. La mayor proporción de ediciones sobre contenido de contexto cultural es generada por usuarios anónimos, mientras que los administradores se involucran proporcionalmente más que simples editores registrados; de hecho, al observar la primera semana de actividad de cada editor de

---

[2] Tecnocampus-Universitat Pompeu Fabra, Barcelona (España)
     E-mail: mmiquelr@tecnocampus.cat
     ORCID: https://orcid.org/0000-0002-2919-9469
[3] *Eurecat,* Centre Tecnològic de Catalunya (España)
     E-mail: david.laniado@eurecat.org
     ORCID: https://orcid.org/0000-0003-4574-0881
[4] ISI Foundation, Turin (Italia)
     E-mail: kaltenbrunner@gmail.com
     ORCID: https://orcid.org/0000-0002-2271-3066

la comunidad, se observa que los administradores ya se involucran correlativamente más que otros editores en el contenido que representa su contexto cultural. Estos hallazgos indican la relevancia de este contenido tanto para satisfacer las necesidades de información de los lectores como para estimular las dinámicas de la comunidad de editores.
Palabras Clave: Wikipedia; Comunidades en línea; Participación del usuario; Cultura; Contexto cultural; Colaboración online

Sumario. 1. Introduction. 2. Dataset description. 3. Analysis and results. 4. Conclusions. 5. Bibliography. 6. Appendix.

1. Introduction

Wikipedia is the most popular knowledge repository on the Internet. It is used in fact-checking, education, news source, among many other contexts (Okoli, 2014). It is important to note that this is achieved with no central authority. Editors are volunteers not directed towards specific topics and maintain editorial freedom. They align with the vision of collecting and giving free access to "the sum of human knowledge"[5].

The editors of each of the 300 Wikipedia language editions decide individually and sometimes in collaboration which are the topics that deserve more coverage, following personal interests and eventually considering which articles are being consulted at the moment[6]. Even though Wikipedia is defined as an encyclopaedia[7], there is evidence that a large part of its content does not strictly follow a balanced coverage of encyclopedic topics; computational topic analyses reflect an overrepresentation of biographies, popular culture and arts (Kittur, Chi, & Suh, 2009).

Editors aim at covering the readers' evolving informational needs. Wikipedia's coverage of news and current events drives editor activity and reader attention any given week (Keegan, Gergle, & Contractor, 2013). Collaborations to create these articles involve more editors and happen at a higher speed than any other type of articles.

However the most read articles do not necessarily correspond to those frequently edited, suggesting some degree of non-alignment between user reading preferences and author editing preferences (Lehmann, Müller-Birn, Laniado, Lalmas, & Kaltenbrunner, 2014). Warncke-Wang, Ranjan, Terveen, & Hecht (2015) analyzed four large language editions and showed that there is an extensive misalignment between the content created and the one that is consumed.

Editor topical preferences depend on factors such as their domain expertise (Halatchliyski, Moskaliuk, Kimmerle, & Cress, 2010; Yarovoy, Nagar, Minkov, & Arazy, 2020), political identity (Neff et al., 2013), among others. Rizoiu, Xie, Caeta-

---

[5]   https://en.wikipedia.org/wiki/Wikipedia:Prime_objective
[6]   https://weekly.hatnote.com/
[7]   https://en.wikipedia.org/wiki/Wikipedia:Wikipedia_is_an_encyclopedia

no, & Cebrian (2016) proved that through the analysis of the edits of an editor it is possible to detect his or her personal traits such as gender and sexual preferences, and even those which tend to vary with the context of the language edition such as or education level, political or religious affiliation.

Cultural contextualization. In fact, cultural and geographical context are key factors to understand the content created. Each Wikipedia language edition is culturally contextualized, meaning that the context "is the cause of some of the content diversity in multilingual Wikipedia" (Hecht, 2013, p. 23).

For example, the resulting link graph between the articles is very focused towards the articles of the territories where the language is spoken (Hecht & Gergle, 2009; Samoilenko, Karimi, Edler, Kunegis, & Strohmaier, 2016); editors tend to edit about the articles they have nearby (Hecht & Gergle, 2010); and the points of view contained in the different language versions of an article also differ greatly depending on the language edition and the topic (Callahan & Herring, 2011; Massa & Scrinzi, 2011; Pentzold et al, 2017).

To assess the extent of content representing the languages' geographical and cultural context in each Wikipedia language edition, in our previous work (Miquel-Ribé & Laniado, 2016) we proposed a method to collect all the articles that relate to the language, people and territories where the language is spoken. We called it Cultural Identity Related Articles (CIRA), and in later work Cultural Context Content (Miquel-Ribé & Laniado, 2018; Miquel-Ribé & Laniado, 2019). In the following we will refer to such content as Cultural Context Content (CCC).

On average, CCC takes a quarter of the first 40 language editions in number of articles, with cases in which it occupies over 44.2% (English) and others with as little as 9.0% (Dutch). Far from being a one-time event, the creation of CCC is a phenomenon sustained over time. Editors create it regularly and often call it "local content", in opposition to the articles that are expected to be in every Wikipedia language edition as notably global knowledge. A significant part of it tends to be unique or exclusive to a single language edition (Miquel-Ribé & Laniado, 2018).

It is not known to what extent the coverage of CCC appears as a need of the readers interpreted by editors, and to what extent it merely results from the editors' motivation to represent content they relate to or identify with, similarly as they do it in social media. The high coverage of CCC has sometimes been considered an "over-representation" or a systemic bias, especially when compared with some gaps or the insufficient coverage of topics that relate to specific areas of the world.

Wikipedians and participation. Even though the creation of CCC is generalized among all the Wikipedia language editions, no study has ever analyzed which types of editors create it, the regularity of the task and if it could constitute an essential trait of the Wikipedian. Editors are characterized by some specific traits, and they evaluate each other's' trustworthiness based on the quantity and the endurance of their edits (Krupa, Vercouter, Hübner, & Herzig, 2009), topical preferences and the functional role or flag they have (Arazy, Nov, & Ortega, 2014).

Editors with a flag have a functional role in the community: the rest of editors expect them to fulfil certain actions they have been entrusted with, often in the maintenance of content but also in decision-making. The process by which these editors receive the flag is either through a request or by being proposed by another editor. Participation is one of the determinant characteristics to obtain such a flag (Burke & Kraut, 2008).

Flags also present a progression ladder that editors may want to climb to achieve a more central position; for most of the communities, administrators (sysops and bureaucrats) take the most relevant attributions and responsibilities and are in the highest level, while registered editors are in the lowest rank (Arazy, Ortega, Nov, Yeo, & Balila, 2015). Editors can also act without logging in (anonymously), but then their IP is recorded instead of their username, and sometimes they are not allowed to edit specific pages.

All in all, while there is extensive research on the many facets of participation and the different types of editors, there is no comprehensive analysis of the extent of topical preferences in editing communities. Therefore, it is unknown whether local content or CCC is prevalently created by anonymous editors or by those who hold a central role in the community, and whether it is a matter of a few or an entire community endeavour to represent the geographical context they live in and its different cultural expressions.

Research questions. We thus believe that characterizing how CCC is created and consumed will help us understand both the dynamics in the Wikipedia communities and the role of the repository in society. We look for answers to the following research questions:

RQ1: Does cultural context content reflect a higher level of editor and reader engagement?

RQ2: Which types of editors in the community engage more in creating cultural context content?

Structure of the paper. In order to answer these questions, we explore how editors participate in the creation of CCC and readers consume it in 15 Wikipedia language editions. To this aim, we run a computational approach to retrieve articles that relate to the editors' cultural context and then analyze the editors' interactions through which they were created and the pageviews they received during a period of six months (Section 2). In Section 3 we present our analyses and findings in two empirical sections aimed at answering each of the research questions. In Section 4 we present conclusions and their implications, and discuss limitations and future work along with some suggestions for Wikipedia communities.

2. Dataset description

In this section we describe the dataset we use for the analyses.

CCC dataset and segments. For the identification of the content that can be considered "local" to each language, we relied on the dataset from a previous study (Miquel-Ribé & Laniado, 2016). As explained above, Cultural Context Content is the group of articles in a Wikipedia language edition that relates to its editors' geographical and cultural context (places, traditions, language, politics, agriculture, biographies, events, etc.)". CCC includes three segments: CCC Keywords, CCC Geolocated, and the general or rest of CCC.

The articles in CCC Keywords (CCC KW) contain in their title the language name, a name of the country or region where the language is spoken or the demonym of its inhabitants. CCC Keywords articles often present an overview of a topic from a cultural context perspective and tend to be comprehensive. A good example of CCC Keywords from the English Wikipedia CCC is 'English poetry', an article

which contains the word 'English' in its title, and whose text is dedicated to summarizing a broad topic (poetry from the United Kingdom written in the English language).

The articles in CCC Geolocated (CCC GL) are all the ones that contain a marker of geolocation in the territories where the language is spoken either legally or natively. Some of them are municipalities, events or monuments, but sometimes even biographies contain a geolocation tag to mark the place of birth or a specific creation. Geolocated articles are around 20% of CCC, while those containing keywords are less than 5%. The articles from the rest of CCC tend to be very specific topics.

In this study, we will compare these two different segments of CCC, considering an intersection between them as a third, different set called CCC Keywords-Geolocated (CCC KW-GL). In order to understand the differences in accumulated engagement, we will compare these three CCC segments with the rest of CCC (CCC rest) and the set of Wikipedia articles that are not included in CCC (WP rest).

Language selection. In line with related studies (Miquel-Ribé & Laniado, 2018; Miquel-Ribé & Laniado, 2019), in our dataset the percentage of CCC for the first forty languages in the number of articles is 23.1%, while for the entire set of 304 languages it is 12.5%; in 165 Wikipedia language editions, CCC occupies less than 10% of their content. 127 Wikipedia language editions contain less than 100 articles geolocated in their territories and 204 less than 100 articles including keywords such as the territory name or the language name on their title.

Therefore, there are many language editions not covering their cultural context sufficiently for informing current events or for basic encyclopedic purposes. This is in high contrast with language editions like German, with 2 million articles and a CCC of 31.2%, Hungarian, with 379,000 articles and a CCC of 14.8%, or Estonian, with 143,000 articles and a CCC of 20.1%. The size of CCC across languages is variable, but having at least minimal extent of it seems to be indicative of a functional project.

By taking a quick look at the size of Wikipedia language editions and their number of editors with administrator roles[8], we see that only 52 out of the entire list of Wikipedia language editions contain more than 100,000 articles, and have a median of 18 administrators. For language editions with less than 100,000 articles, the number of administrators is usually less than ten.

To select a group of languages to understand the creation and consumption of CCC, we consider that it is necessary to choose Wikipedia language editions from different geographical backgrounds, linguistic diversity and number of speakers and of different sizes but containing more than this threshold of 100,000 articles, to ensure a certain content completion and community activity.

We thus selected the following 15 Wikipedia language editions: Arabic, Basque, Catalan, English, German, Hebrew, Hungarian, Icelandic, Italian, Japanese, Macedonian, Romanian, Russian, Spanish, and Turkish. 7 of them have more than a million articles, and 8 surpass the 100,000 articles threshold. The median number of administrators of these languages is 29 with an average of 121 (English Wikipedia has 1122).

Editing history and pageviews. Once we have selected the set of languages, we retrieve the number of pageviews each of their articles received during a period of six months (from January to June 2016). Then we retrieve their edit history available

---

[8]   https://meta.wikimedia.org/wiki/List_of_Wikipedias

in the Wikimedia dumps[9] accounting the edits until June 2016, which include the edited page and the corresponding editor. This way, we store the edits made by registered editors, anonymous editors, i.e. non-registered, as well as those made by bots, which are automatic programs managed by registered editors.

For registered editors, we differentiate edits made by editors having or not having the flag of administrator. Among the different user flags, we only consider the administrator flag as it is the most usual and the one with more attributions (Arazy et al., 2015).

3. Analysis and results

3.1. RQ1. Reader and editor engagement in CCC

Our first research question (RQ1) asks whether the Cultural Context Content reflects a higher level of editor and reader engagement than the rest of Wikipedia.

3.1.1. Edits and pageviews in CCC segments

We quantify the engagement around this kind of content in terms of number of pageviews and edits. The different segments of CCC (CCC geolocated articles, CCC with keywords on title and the rest of CCC) represent different sorts of information with very different levels of popularity (e.g. a geolocated article on a city may be very different from a summary of a literature genre). Then, it would be interesting to compare the CCC segments to see how editing and reading popularity differs in them.

Taking this into account, we propose establishing a double-level comparison between reader engagement (using the number of pageviews for each article page) and editor engagement (using the number of edits during the entire history), in the different segments of CCC and in the rest of Wikipedia. In this regard, we propose using the Kruskal-Wallis test in order to verify the hypothesis that the different CCC segments have more edits and pageviews than the rest of the Wikipedia articles. This test is often referred to as the ranked version of ANOVA. Since it is a non-parametric test, it is used when samples do not follow a normal distribution. Hence, it is an extension of the Mann-Whitney U test. However, this test requires different groups to compare to have exclusive members.

Since the CCC segments of CCC Geolocated and CCC Keywords have some articles in common, we proposed dividing them into the smaller segments: CCC Keywords – CCC Geolocated as the intersection of both segments (CCC KW-GL), CCC Keywords without the previously selected (CCC KW), CCC Geolocated without the previously selected (CCC GL), and the rest of articles which compose CCC (CCC Rest). The remaining articles not in CCC constitute the WP Rest. The size of the segments is as mentioned from smallest to largest.

Table 1 shows the mean ranks of the number of edits and pageviews for the articles in each CCC segment. A higher mean rank for a set implies the number of edits or pageviews is higher as they are ranked in ascending order. Results of the Krus-

---

9   https://dumps.wikimedia.org/

kal-Wallis test confirm the hypothesis that generally the different segments of CCC have more edits and pageviews than the rest of the Wikipedia articles. Differences between segments are all significant (p < 0.001).

Language	Metric	CCC KW-GL	CCC KW	CCC GL	CCC REST	WP REST
Arabic	Edits	290,758	246,308	266,367	216,993	173,681
	PV	298,913	249,623	270,021	211,799	174,928
Basque	Edits	197,027	166,467	172,850	160,888	97,755
	PV	185,864	149,178	144,771	139,770	100,220
Catalan	Edits	382,121	349,292	259,115	334,169	221,303
	PV	372,024	344,451	221,513	320,783	226,112
English	Edits	2,498,244	2,623,270	2,562,090	2,795,003	2,211,878
	PV	2,621,479	2,714,416	2,566,498	2,799,549	2,203,191
German	Edits	1,145,545	1,003,962	1,025,320	1,027,417	853,503
	PV	1,191,907	980,990	1,051,316	1,027,248	850,468
Hebrew	Edits	129,083	112,869	125,163	102,682	78,834
	PV	121,954	103,719	127,252	94,977	82,149
Hungarian	Edits	279,864	193,815	261,092	210,794	150,715
	PV	267,768	196,386	266,990	202,851	152,022
Icelandic	Edits	32,655	26,267	28,703	20,042	19,240
	PV	32,894	24,117	29,157	15,083	21,209
Italian	Edits	1,055,822	833,045	888,808	805,224	553,683
	PV	1,005,778	755,232	839,191	775,189	562,063
Japanese	Edits	761,386	577,648	686,429	545,199	420,352
	PV	752,618	555,455	640,870	513,959	451,398
Macedonian	Edits	70,510	64,058	63,015	52,545	38,754
	PV	44,326	43,466	43,001	41,377	41,257
Romanian	Edits	261,977	249,560	230,016	230,892	145,402
	PV	237,757	250,070	213,301	210,767	150,158
Russian	Edits	583,347	804,657	515,294	754,215	591,888
	PV	618,330	768,412	513,234	710,884	604,667
Spanish	Edits	833,199	756,783	783,634	743,856	505,647
	PV	836,554	728,382	730,250	707,357	520,390
Turkish	Edits	158,900	129,431	166,207	150,433	110,832
	PV	155,726	136,848	128,604	136,489	118,873

Table 1. Mean ranks (ascending) for the number of edits and pageviews in different segments and intersections of CCC and the rest of Wikipedia. Darker colours represent higher mean ranks, indicating a higher number of edits and pageviews in that content type. CCC KW-GL: intersection of articles with keywords on title and geolocation, CCC KW: articles with keywords on title and without CCC KW-GL, CCC GL: articles with geolocation without the CCCKW-GL, CCC REST: articles in CCC not included in the previous selections, WP REST: the rest of Wikipedia without articles from CCC.

In most language editions, all the CCC segments have higher values for both edits and pageviews than the rest of the articles (RQ1). In Table 1, a pattern transition can be seen in the mean ranks in almost all languages: CCC Keywords-Geolocated ob-

tains the highest mean ranks, followed by either CCC Geolocated or CCC Keywords. In almost all cases, the rest of CCC is lower than the other segments but still higher than the rest of Wikipedia. Since the test has been applied to both edits and pageviews with the same population of articles, it is possible to compare the mean ranks obtained for two metrics for each group of articles. For almost all languages and in any of the CCC segments, mean ranks computed for edits are higher than for pageviews.

For illustrative purposes, we included Figure 1, which shows the average values for edits and pageviews for each segment of CCC (including the overlapped CCC Keywords-Geolocated). The figure shows the edits (as bars) and the pageviews (as horizontal lines) both depicted all over their range in the dual-axis Y. Beyond comparing engagement of readers and editors for different CCC segments, Figure 1 also allows one to observe the differences in absolute values for the Wikipedia language editions; for instance, while the range of pageviews for the Japanese Wikipedia is up to 40,000, for the Basque Wikipedia it is about 400.

Discussion. We have presented a comparison between the different segments of CCC at article level for participation and readership. In the first place, results confirm a higher level of participation in CCC than in the rest of Wikipedia, especially in articles from the group with both keywords in the title and geolocation tag, and the rest of articles with keywords on the title. This pattern is analogous for the readership, with a higher number of pageviews in the segments of CCC than in the rest of Wikipedia.

Results from the overall comparison for each of the CCC segments are more marked for edits. Hence, editors engage more in participating in CCC than readers viewing it. In other words, editors could be motivated to edit and grow CCC even over the possible demand by readers.

Kruskal-Wallis tests (p-values lower than 0.001 for all languages) confirm the hypothesis that the different segments of CCC have more edits and pageviews than the rest of the Wikipedia articles, apart for some specific segments in Catalan, Icelandic and Russian.

3.1.2. Edits and pageviews in CCC Geolocated articles

The comparison between CCC segments and the rest of Wikipedia revealed that cultural context content articles are a shared interest between readers and editors. In fact, they both show significantly higher values than the rest of Wikipedia content, and especially in the case of CCC Keywords or CCC Geolocated articles.

We pay special attention to the CCC Geolocated articles and represent them visually in a map as they can reveal the coincidence in interest between editors and readers, and the distribution of such interest among different kinds of concepts such as cities and physical geographical elements.

Figure 2 presents a map for the Catalan and English Wikipedia – the same figure is created for the rest of languages in Appendix Figures A.1 to A.9. In the map, each article is depicted with a dot. The size of the dot represents the number of pageviews, and the colour is the number of edits presented as a divergence continuum red-green where the middle point is 250 edits in beige. This way, it can be easily perceived when an article has been created through many edits, and at the same time, it is popular in terms of pageviews. Additionally, among the articles popular both among editors and readers, we show a few relevant ones for each language to provide details.
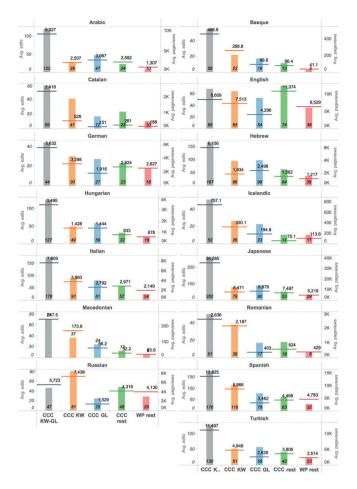
Figure 1. Reader and Editor Engagement in CCC segments compared to the rest of Wikipedia. Average absolute values for the number of edits (bar) and the number of pageviews (line) in the different article types: CCC Keywords-Geolocated (CCC KW-GL), CCC Keywords (CCC KW), CCC Geolocated (CCC GL), rest of CCC (CCC rest), rest of Wikipedia (WP rest). The graph presents a dual y-axis with edits and pageviews.

At first glance, a minority of articles obtain many more edits and pageviews than the others. These are usually the main cities from each territory or special monuments within them. For instance, in the Catalan Wikipedia the article 'Barcelona' is among the most edited with 1,833 edits, while for the English, 'New York City' and 'Buckingham Palace' are articles which exceed the 8,851 and 2,583 edits respectively. Each Wikipedia presents a different scale, but the imbalances are similar. Likewise, the map also shows that an important density of articles in the territory revolves around the biggest cities.

Since the urbanisation in the different countries varies a lot, some languages like German and Catalan present a very strong density, while others like Hebrew or Arabic exhibit many empty areas (Figure A.2). There are exceptions, but generally the disposition of points is somewhat similar to a population map. In fact, besides the articles

about cities, the rest of concepts range from a company to a historical event or monument, which usually take place or are located in urban fields. Since the big dots (which imply more pageviews) tend to be greener than the rest (which imply more edits), it is possible to state that editors and readers are both interested in these geolocated articles.

Discussion. Previous research showed there exists a misalignment between demand and supply (Warncke-Wang et al., 2015). The map visualizations allow hypothesizing about a coincidence between the most viewed and the most edited, usually important cities.
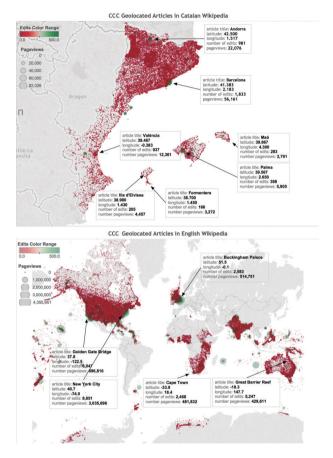

Figure 2. Editor and reader engagement in CCC Geolocated articles from the Catalan and English Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.

3.1.3. Edits and pageviews in CCC by editor type

We now quantify the engagement around this kind of content in terms of the number of pageviews and edits. As commonly done in previous research, we excluded bot

edits[10]. Bots' activity is directed by a small group of editors and focused on a reduced range of tasks. Therefore, we consider them to be less representative of the interests and preferences of the editors' communities.

Figure 3 shows the relative weight of CCC in each Wikipedia language edition in terms of the number of articles, pageviews, registered editors' edits and anonymous edits. One can notice a crescendo according to this order in almost all cases, indicating that CCC as a whole receives overall a greater attention than the rest of the content by readers, and even more so by editors. It is important to remark that anonymous editors devote almost half of their attention to CCC (mean 47.3%, median 45.2%, standard deviation 10.5).

To verify these results at the article level, we compare CCC to the rest of Wikipedia articles, both in terms of human edits (anonymous and registered editors' edits) and pageviews. We perform a Mann-Whitney[11] to test the hypothesis that in general there would be more human edits and pageviews in CCC articles than in the rest of Wikipedia articles.

The results (significant with p-values < 0.005 for all languages) of the test confirm our hypothesis, and are consistent with results from the previous section. They show a higher engagement in CCC than the rest of Wikipedia content, highlighting in most cases greater differences for edits than for pageviews (data not shown).

Discussion. The higher engagement of editors with CCC points out a preference towards representing their context that seems to go beyond a possible response to readers' information demand. This preference is particularly high for anonymous editors, which we assume to be mainly spontaneous editors, and tend to engage more in editing topics related to their nearby environment. This suggests that this type of editor engages in CCC for being more motivated by the content itself rather than by an interest for the entire Wikipedia project. In the next section, we deepen this result and study behaviour patterns according to editor types.

3.2. RQ2. Community participation in CCC

Our second research question (RQ2) asks which types of editors engage more in the creation of cultural context content. The subsections 3.2.1 and 3.2.2 will give an answer to it.

3.2.1. Proportion of edits in CCC by editor type

To find out which editor types engage more in creating content representing cultural context, we first look at the overall proportion of edits made by each editor type, taken as a group, in CCC and in all Wikipedia articles as two sets of articles (Figure 4). We differentiate bots from human editors, among which we distinguish three types: administrators, plain registered editors and anonymous editors.

---

[10]   Registered nicknames that operate as bots are granted a special flag and are stored in a Mediawiki table which is regularly made available as a dump.

[11]   This test is employed to compare differences between only two independent groups for a not normally distributed dependent variable. It is referred to as the ranks version of t-test because it uses ranks calculations in order to avoid the problems of absolute values in a non-normal distribution.
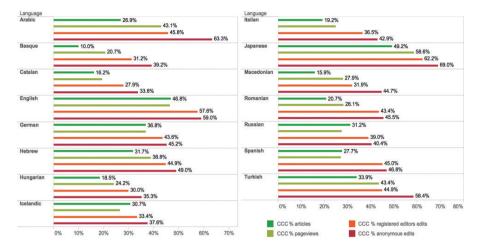
Figure 3. The relative weight of CCC in number of articles, number of pageviews, registered editor edits and anonymous edits.

In the top graph, we look at edits made by bots and find a lower percentage of edits in CCC compared to the percentage in the entire Wikipedia. This implies that CCC articles have a higher proportion of human edits than the entire Wikipedia.

In the bottom graph, we focus on human edits, and find that in almost all the languages considered, anonymous editors have a higher impact on CCC than on the entire encyclopaedia confirming results from Figure 4, while the opposite is the case for registered editors, and administrators present a more stable pattern (RQ2).
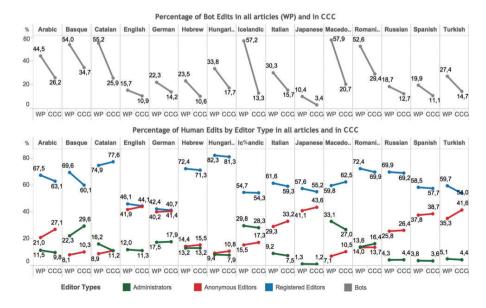

Figure 4. Percentage of edits by editor type in CCC vs all articles. Top: edits by bots. Bottom: edits by human editors.

Discussion. Figure 4 shows that the entire community participates in editing CCC, from the core to the periphery, with special engagement among the groups of administrators and anonymous editors. Anonymous editors appear to be especially triggered to edit about their cultural context, which may respond to following more their immediate interests rather than a stronger concern for the Wikipedia project and its content as a whole.

3.2.2. Administrators proportion of edits in CCC

To assess whether the differences observed between the impact of different user groups at the aggregated level are consistent also when considering individual editors separately (and not just produced by outliers), we compute the percentage of edits made in CCC by each editor and compare the distribution of this variable for different user groups. As it is not possible to study anonymous editors at the individual level, due to the fact that IP addresses cannot reliably and unequivocally identify users, for this and the next analyses we focus only on registered editors.

To focus on editors which are more likely to be local to a language edition, we furthermore only consider for each language edition the editors who have more edits in that language edition than in the other language editions. We test the hypothesis that administrators have a higher proportion of their edits in CCC than non-administrators. The distribution of the proportion of edits to CCC is not normal, so we perform a non-parametric test (we use a Mann-Whitney U test, which provides mean ranks as a result).

The results, shown in Table 2, validate the hypothesis that administrators devote a significantly higher proportion of their contributions to CCC for most of the languages. While in the Japanese, and although much less markedly in the English, we find the opposite result, which may be related to the larger size of CCC in these languages. Differences are not significant for the German, Hebrew, Romanian and Turkish Wikipedia.

Discussion. We could interpret that in general administrators have a higher proportion of edits made in CCC than plain registered editors in light of their tasks: they are responsible for ensuring content quality (Suh, Convertino, Chi, & Pirolli, 2009), and tend to interact preferentially with inexperienced editors (Laniado & Tasso, 2011; Laniado, Tasso, Volkovich, & Kaltenbrunner, 2011). Therefore, their proportionally higher activity on such central and more read content like cultural context representations is consistent with their position and role in the project. A comparison by edit count, reported in Miquel-Ribé (2017: 173, figure 42) also points to a generally higher proportion of edits in CCC for experienced editors.

3.2.3. Administrators proportion of edits in CCC during the first seven days

The very first week of activity has been considered especially meaningful to detect those editors who would become highly participative (Panciera, Halfaker, & Terveen, 2009) since in only a few days after registering they already exhibit different characteristics from the rest of editors. In this period of editor life, content choices may be free from other conditionings and community dynamics, and they may provide clues on the motivations that attracted a user to participate in the project. Hence, by means of a Mann-Whitney test, we assess whether the proportion of edits in CCC

during the first seven days differs for administrators with respect to the other registered editors. Our hypothesis is that administrators will have a higher proportion of edits in CCC than plain registered editors.

Proportion of Edits in CCC

Languages	Admins Mean Rank	Non-Admins Mean Rank	Number of Admins	Number of Non-Admin	Z	p-value	
Arabic	98,579	79,031	32	158,036	-2.743	0.006	*
Basque	2,052	1,502	8	2,998	-2.045	0.041	*
Catalan	23,065	17,928	24	35,861	-2.747	0.006	*
English	3,262,617	3,456,449	1,244	6,911,584	-3.630	0.000	*
German	358,752	340,695	241	681,160	-1.499	0.134	
Hebrew	41,080	36,841	38	73,648	-1.307	0.191	
Hungarian	43,268	33,070	35	66,114	-3.604	0.000	*
Icelandic	2,037	1,527	20	3,039	-2.905	0.004	*
Italian	111,171	98,895	100	197,702	-2.311	0.021	*
Japanese	96,961	146,441	47	292,819	-4.225	0.000	*
Macedonian	3,289	2,339	13	4,670	-2.997	0.003	*
Romanian	24,479	21,555	18	43,094	-1.092	0.275	
Russian	204,136	158,606	86	317,150	-5.143	0.000	*
Spanish	402,207	329,177	71	658,297	-3.610	0.000	*
Turkish	79,883	70,886	26	141,749	-1.227	0.220	

Table 2. The proportion of edits in CCC: Admins vs Non-Admins. The values are the Mann-Whitney U test resulting mean ranks (ascending) for every editor's proportion of its edits in CCC articles and the Z-scores (standardized scores) which is computed to obtain the p-values. Darker colours represent higher mean ranks, indicating a higher proportion of edits in CCC. Statistically significant results ($p<0.05$) are marked with a star. The results of the test confirm the hypothesis that administrators have a higher proportion of edits in CCC than non-administrators in all the tested language editions apart from Japanese and English.

The results of the Mann-Whitney test show that administrators have a higher mean rank than non-administrators for the proportion of edits in CCC in the first 7 days in Wikipedia after performing the first edit (Table 3). This confirms the hypothesis that administrators tend to have a higher proportion of participation in CCC than the rest of registered editors during the first seven days after registering.

Discussion. These results suggest that in many language editions editors that are more prone to get involved in the project such as administrators may be especially more interested in creating cultural context content in the first phase. In other words, we may conjecture that they care about how their most immediate environment is depicted on Wikipedia, and that a sense of belonging to the context could precede a sense of belonging to the Wikipedia community.

4. Conclusions

Wikipedia has become a fundamental tool to give access to free knowledge to all layers of the population in most places in all continents. The role of "local content"

or articles about the cultural context of each language edition has been essential for the project to become both a local and a general encyclopedia and fulfil the readers' needs. In this paper, we explored how Wikipedia editors participate in the creation of this content, and readers consume it in 15 diverse Wikipedia language editions.

Proportion of Edits in CCC in the first 7 days

Languages	Admins Mean Rank	Non-Admins Mean Rank	Number of Admins	Number of Non-Admi..	Z	p-value	
Arabic	95,603	79,031	32	158,036	-2.345	0.019	*
Basque	1,957	1,502	8	2,998	-1.729	0.084	
Catalan	23,051	17,928	24	35,837	-2.794	0.005	*
English	3,408,433	3,456,423	1,244	6,911,584	-0.907	0.364	
German	376,956	340,688	241	681,160	-3.057	0.002	*
Hebrew	38,706	36,843	38	73,648	-0.582	0.560	
Hungarian	41,465	33,071	35	66,114	-3.045	0.002	*
Icelandic	1,863	1,528	20	3,039	-1.941	0.052	
Italian	110,670	98,896	100	197,702	-2.262	0.024	*
Japanese	110,404	146,439	47	292,819	-3.122	0.002	*
Macedonian	3,170	2,340	13	4,670	-2.676	0.007	*
Romanian	26,196	21,555	18	43,094	-1.752	0.080	
Russian	196,782	158,608	86	317,150	-4.425	0.000	*
Spanish	387,798	329,178	71	658,297	-2.937	0.003	*
Turkish	79,624	70,886	26	141,749	-1.204	0.229	

Table 3. Comparison of the proportion of edits in CCC during the first seven days by administrator functional role. The values are the Mann-Whitney U test results and mean ranks (ascending) and the Z-scores (standardized scores) which are computed to obtain the p-values. Darker colors represent higher mean ranks, indicating a higher proportion of edits to CCC. Statistically significant results ($p<0.05$) are marked with a star. The results of the test confirm the hypothesis that administrators have a higher proportion of edits in CCC than non-administrators already in their first seven days after the first edit in all the tested language editions but Japanese.

We measured the number of edits devoted to these articles by different types of editors, and the pageviews they received from readers. The results suggest that Cultural Context Content (CCC) is a centre of interest in Wikipedia content; CCC is more popular than the rest of content for both readers and editors, and proportionally more for the latter, which answers our first research question (RQ1).

A comparative study between participation and readership has revealed that the edition of these contents transcends the demand by readers. Anonymous editors and administrators are the types of editors who present a higher engagement in its creation (RQ2). The impact or proportion of anonymous edits is higher than in the entire Wikipedia.

On an individual basis, administrators dedicate a higher proportion of edits to CCC than an average registered editor. While this could be due to the patrolling tasks, administrators already show a higher proportion of edits in CCC after their first days of registering in Wikipedia, suggesting that being motivated to represent one's cultural context could be a motivator to become a committed Wikipedian.

Based on these findings, we conclude that cultural contextualization is not a marginal or anecdotal phenomenon. Rather than being an undesired bias, we could see its relevance for the different types of users of Wikipedia. While the contextualization of user-generated content in projects like Wikipedia or Flickr has been the object of previous studies (Hecht, 2013), this is the first systematic analysis that investigates the relevance of this phenomenon for editors and readers across multiple Wikipedia language editions.

4.1. Limitations and future work

Firstly, through the selection of articles that compose the CCC of every language, we investigated different segments of it, those including keywords in their title, as well as those with a geolocation tag. These features identify specific types of articles, as every article containing the demonym, the language name or the country name tends to include a summary of a topic revolving the context, while geolocated articles are generally about places. The use of these segments has been useful in order to understand that the engagement in CCC is unequal. However, it would be interesting to have more fine-grained categories, such as the eleven topics used by Kittur et al. (2009) in the topical analysis of Wikipedia.

Secondly, the general community analysis taking into account the types of editors has been necessary to see who creates CCC. We could infer the importance of editing CCC for Wikipedians through their proportion of edits in it. The fact that administrators distinguish for a higher engagement in CCC already in their first seven days suggests that editing CCC could be a Wikipedian trait to value. To understand better the influence of the context on editors' motivation, it would be desirable to complement the current quantitative approach with self-reported methods like a survey or interview with Wikipedians.

Thirdly, our study focused on 15 language editions, aiming at the diversity of spread, number of speakers and geographical location, and putting as a requirement having more than 100,000 articles, to ensure a degree of completion and community activity. We observed that languages with fewer articles have a very scarce representation of CCC (or subsegments). This confirms our general assumption that "creating CCC is part of a healthy Wikipedia". Nonetheless, analyzing the edits made by anonymous editors in minor language editions could still be a good indicator of the potential for growing the community. Studying these languages, the representation of their CCC and the barriers of many kinds (cultural like the language standardization or social status, technical like limited Internet access, socio-economic like lack of welfare, etc.) could lead to a better understanding of the necessary factors for a Wikipedia community to grow.

4.2. Recommendations

Based on the research results from this study and our knowledge on the Wikimedia movement, we make the following four recommendations:

Readership should be considered before deleting CCC articles. The readership of CCC is very relevant as it allows us to reflect on the usefulness of this content. Even the CCC articles which are not CCC Geolocated or CCC Keywords have on average more pageviews than the rest of the Wikipedia articles. This is relevant because often

some CCC articles are considered not notable enough. In certain languages and contexts, local cultural expressions do not have the same number of sources. Since the current notability policy is a one-size-fits-all, the same criteria are applied to every topic without taking into account any contextual information like the number of sources available. We believe it is important at least to inform editors that local content is being consumed.

CCC Keywords articles should become a priority for every language community. CCC Keywords articles are valuable summaries of different topics (as explained before, they could be e.g. "Italian music", "Italian cuisine") and they are amongst the most viewed articles. This means that synthesizing and presenting a topic from a cultural context perspective tends to repay the effort. These articles should become a priority before other more specific articles.

Creating CCC should be used as an activity to attract newcomers. The fact that proportionally there are more edits made by anonymous editors in CCC than in the entire Wikipedia suggests that this content may attract new people to edit. This result is consistent in all the 15 Wikipedia language editions that have been analyzed. In a way, it is expected that any reader is more familiar with their cultural context and may be able to detect gaps or errors in articles about it. This should be used in favour of engaging newcomers into campaigns to create more CCC and teach them the rules and tools in order to become Wikipedians.

Specific topics in CCC should be coordinated by affiliates or groups of editors. Although representing cultural context seems to be a universal activity in Wikipedia that attracts anonymous editors, there are certain specific subtopics that need coordination. For example, those about galleries, libraries, archives and museums (known as GLAM[12]), monuments, among many other topics. Nonetheless, there are no affiliates (chapters or user groups) in the Wikimedia movement for every language edition. This means that in certain languages there is no capacity for coordination, and it is more difficult that some CCC topics are mapped and gaps are addressed. This would justify dedicating resources to support minimal infrastructure and ensure that the affiliates network reaches every context in the world.

4.3. Content diversity and strategy

We have demonstrated that creating CCC is essential for both readers and editors; for both filling their informative needs and for stimulating community dynamics. Even though, as stated in the limitations, we did not analyze the 304 language editions, we could see that many languages do not cover their context adequately: 165 Wikipedia language editions have less than 10% of their content dedicated to CCC and 127 have less than 100 articles geolocated in the territories where the language is spoken as indigenous or has official status.

The creation of CCC by smaller Wikipedias is essential in order to collect their heritage and to give vitality to the language. In this sense, we must bear in mind that 253 out of the 302 languages having a Wikipedia edition are languages that coexist in their territory with a language of a higher status, which typically implies a restriction to a smaller range of domains than the other language or other sorts of minoriti-

---

[12]   https://en.wikipedia.org/wiki/Wikipedia:GLAM

zation. We need to encourage these languages to represent their context when it does not happen spontaneously.

The Wikimedia movement has defined "knowledge equity"[13] as one of the most important goals for the Strategic direction towards 2030. It proposes focusing the "efforts on the knowledge and communities that have been left out by structures of power and privilege". This means that it is necessary that every community represents its context in its own language or languages, and for that, it might be necessary to break some social, political, and technical barriers for them to be able to do it.

Culture gaps between the different language editions occur when articles belonging to CCC have no equivalent in other language editions (Miquel-Ribé & Laniado, 2018). It may not be possible to bridge the gaps in their completion, but ensuring a minimal coverage of every culture seems a reasonable goal. Visualizations and tools[14] to help communities bridge the gaps between languages have been proposed in the Wikipedia Diversity Observatory project[15] (Miquel-Ribé & Laniado, 2020). But bridging the gaps between languages is as important as the coordination and the campaigns in order to create the content that represents the context in its native language in the first place. If Wikipedia as a multilingual project aims at being complete in terms of content diversity, it is essential that cultural contextualization is not only a phenomenon that occurs spontaneously in some languages, but a goal that is pursued with tools and guidance for it to happen consistently in every Wikipedia language edition.

5. Bibliography

Arazy, O., Nov, O., y Ortega, F. (2014). "The [Wikipedia] World is not flat: on the Organizational Structure of Online production Communities". Ecis 2014.

Arazy, O., Ortega, F., Nov, O., Yeo, L., y Balila, A. (2015). "Functional roles and career paths in Wikipedia." (pp. 1092–1105). *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. http://doi.org/10.1145/2675133.2675257

Burke, M., y Kraut, R. (2008). "Mopping up: modeling wikipedia promotion decisions." (pp. 27–36). *Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08)*. DOI: http://doi.org/10.1145/1460563.1460571

Callahan, E. S., y Herring, S. C. (2011). "Cultural Bias in Wikipedia Content on Famous Persons." *Journal of the American Society for Information Science and Technology*, vol. 62, núm. 10, 1899–1915. DOI: http://doi.org/10.1002/asi.21577

Halatchliyski, I., Moskaliuk, J., Kimmerle, J., y Cress, U. (2010). "Who integrates the networks of knowledge in Wikipedia?" *Int. Sym. Wikis*, 1. DOI: http://doi.org/10.1145/1832772.1832774

Hecht, B. J. (2013). *The Mining and Application of Diverse Cultural Perspectives in User-Generated Content*. United States.

Hecht, B. J., y Gergle, D. (2010). "On the localness of user-generated content" (pp. 229–232). *Proceedings of the 2010 ACM conference on Computer supported cooperative work (CSCW '10)*. DOI: http://doi.org/10.1145/1718918.1718962

---

[13] https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20
[14] https://wdo.wmcloud.org/
[15] https://meta.wikimedia.org/wiki/Wikipedia_Diversity_Observatory

Hecht, B., y Gergle, D. (2009). "Measuring self-focus bias in community-maintained knowledge repositories." *Proceedings of the fourth international conference on Communities and technologies (C&T '09)*. DOI: http://doi.org/10.1145/1556460.1556463

Keegan, B., Gergle, D., y Contractor, N. (2013). "Hot Off the Wiki: Structures and Dynamics of Wikipedia's Coverage of Breaking News Events." *American Behavioral Scientist*, vol. 57, núm. 5, 595–622. DOI: http://doi.org/10.1177/0002764212469367

Kittur, A., Chi, E. H., y Suh, B. (2009). "What's in Wikipedia?: mapping topics and conflict using socially annotated category structure." *Proceedings of the SIGCHI conference on human factors in computing systems (CHI '08)*. DOI: http://doi.org/10.1145/1518701.1518930

Krupa, Y., Vercouter, L., Hübner, J. F., y Herzig, A. (2009). "Trust Based Evaluation of Wikipedia's Contributors." En *Engineering Societies in the Agents World X*, *5881*(Chapter 13), 148–161. DOI: http://doi.org/10.1007/978-3-642-10203-5_13

Laniado, D., y Tasso, R. (2011). "Co-authorship 2.0." *Proceedings of the 22nd ACM conference on Hypertext and hypermedia (HT '11)*. DOI: http://doi.org/10.1145/1995966.1995994

Laniado, D., Tasso, R., Volkovich, Y., y Kaltenbrunner, A. (2011). "When the wikipedians talk: Network and tree structure of wikipedia discussion pages." *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*.

Lehmann, J., Müller-Birn, C., Laniado, D., Lalmas, M., y Kaltenbrunner, A. (2014). "Reader preferences and behavior on Wikipedia." *Proceedings of the 25th ACM conference on Hypertext and social media (HT '14)*. DOI: http://doi.org/10.1145/2631775.2631805

Massa, P., y Scrinzi, F. (2011). "Exploring linguistic points of view of Wikipedia." *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09)*. DOI: http://doi.org/10.1145/2038558.2038599

Miquel-Ribé, M., y Laniado, D. (2016). "Cultural identities in wikipedias." *Proceedings of the 7th 2016 International Conference on Social Media & Society*. DOI: https://doi.org/10.1145/2930971.2930996

Miquel-Ribé, M. (2017). *Identity-based motivation in digital engagement: the influence of community and cultural identity on participation in Wikipedia.* Doctoral dissertation, Universitat Pompeu Fabra.

Miquel-Ribé, M., y Laniado, D. (2018). "Wikipedia Culture Gap: Quantifying Content Imbalances Across 40 Language Editions." *Frontiers in Physics*, vol. 6, 234. DOI: http://doi.org/10.3389/fphy.2018.00054

Miquel-Ribé, M., y Laniado, D. (2019). "Wikipedia Cultural Diversity Dataset: A Complete Cartography for 300 Language Editions." *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM '19)*.

Miquel-Ribé, M., y Laniado, D. (2020). "The Wikipedia Diversity Observatory: A Project to Identify and Bridge Content Gaps in Wikipedia." *Proceedings of the 16th International Symposium on Open Collaboration (OpenSym '20)*. DOI: https://doi.org/10.1145/3412569.3412866

Neff, J. J., Laniado, D., Kappler, K. E., Volkovich, Y., Aragón, P., y Kaltenbrunner, A. (2013). "Jointly They Edit: Examining the Impact of Community Identification on Political Interaction in Wikipedia." *PloS One*, vol. 8, núm. 4, 60584. DOI: http://doi.org/10.1371/journal.pone.0060584

Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F. Å., y Lanamäki, A. (2014). "Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership." *Journal of the Association for Information Science and Technology*, vol. 65, núm. 12, 2381-2403.

Panciera, K., Halfaker, A., y Terveen, L. (2009). "Wikipedians are born, not made: a study of power editors on Wikipedia." *Proceedings of the ACM 2009 international conference on Supporting group work (GROUP '09)*. DOI: http://doi.org/10.1145/1531674.1531682

Pentzold, C., Weltevrede, E., Mauri, M., Laniado, D., Kaltenbrunner, A., y Borra, E. (2017). "Digging wikipedia: the online encyclopedia as a digital cultural heritage gateway and site." *Journal on Computing and Cultural Heritage*, vol. 10, núm. 1, 1-19. DOI: https://doi.org/10.1145/3012285

Rizoiu, M.-A., Xie, L., Caetano, T., y Cebrian, M. (2016). "Evolution of Privacy Loss in Wikipedia" (pp. 215–224). *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. DOI: http://doi.org/10.1145/2835776.2835798

Samoilenko, A., Karimi, F., Edler, D., Kunegis, J., y Strohmaier, M. (2016). "Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity." *EPJ Data Science*, vol. 5, núm. 1, 171–21. DOI: http://doi.org/10.1140/epjds/s13688-016-0070-8

Suh, B., Convertino, G., Chi, E. H., y Pirolli, P. (2009). "The singularity is not near: slowing growth of Wikipedia." *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09)*. DOI: http://doi.org/10.1145/1641309.1641322

Warncke-Wang, M., Ranjan, V., Terveen, L. G., y Hecht, B. (2015). "Misalignment Between Supply and Demand of Quality Content in Peer Production Communities." *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (w 2011)*.

Yarovoy, A., Nagar, Y., Minkov, E., y Arazy, O. (2020). "Assessing the Contribution of Subject-matter Experts to Wikipedia." *ACM Transactions on Social Computing*, vol. 3, núm.4,1-36. DOI: http://doi.org/10.1145/3416853
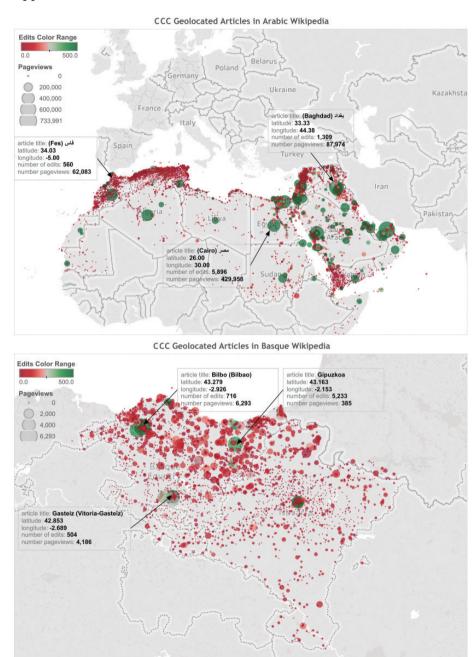
6. Appendix

Figure A1. Editor and reader engagement in CCC Geolocated articles from Arabic and Basque Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
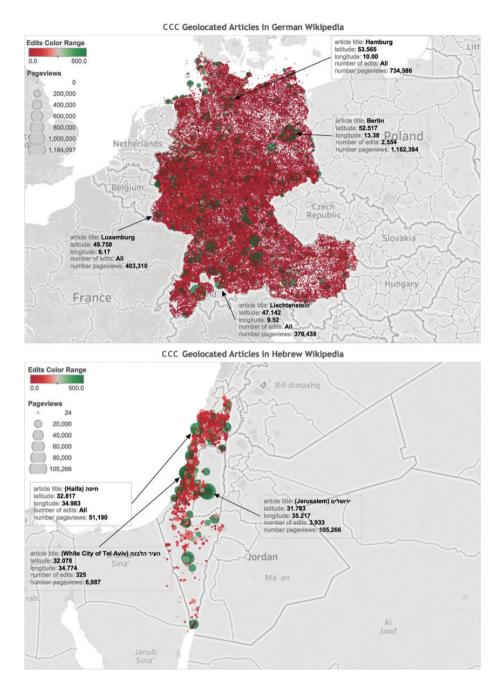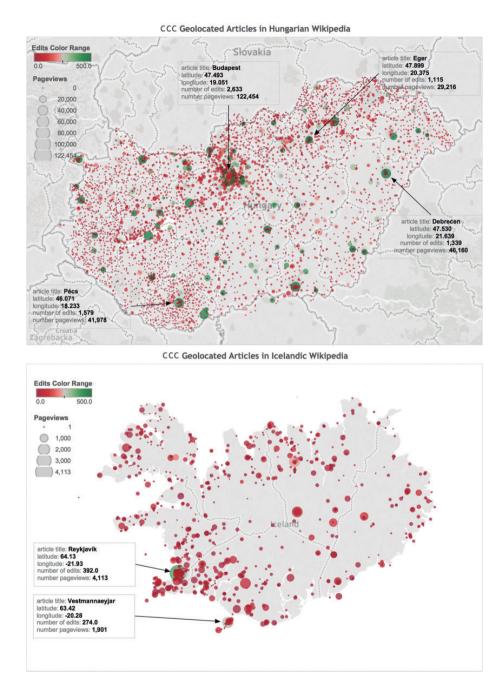
Figure A2. Editor and reader engagement in CCC Geolocated articles from the German and Hebrew Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
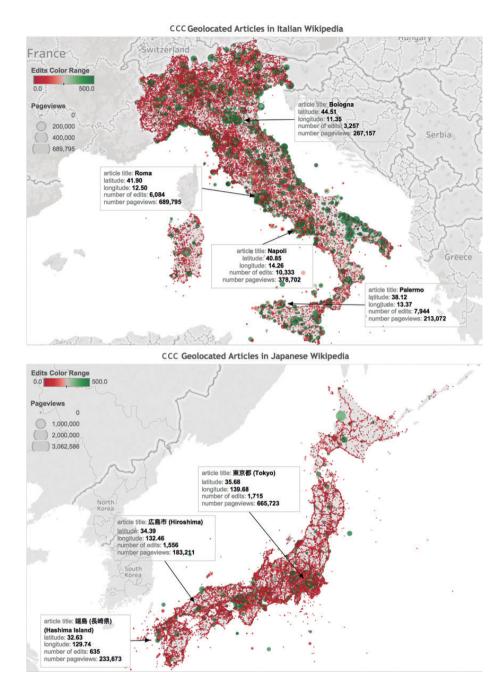
Figure A3. Editor and reader engagement in CCC Geolocated articles from Hungarian and Icelandic Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
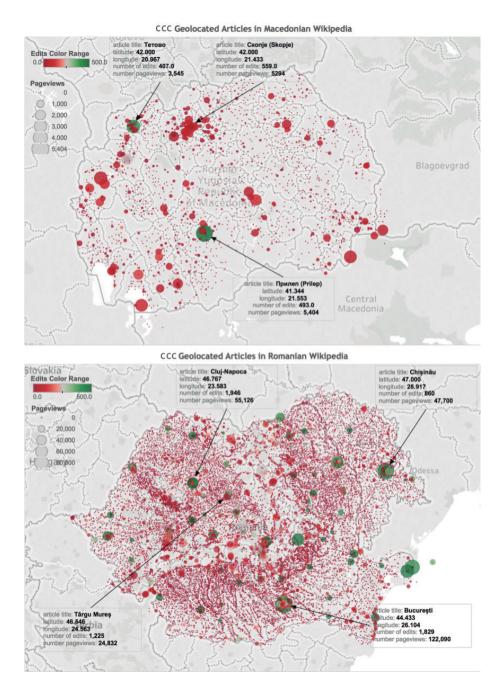
Figure A4. Editor and reader engagement in CCC Geolocated articles from Italian and Japanese Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
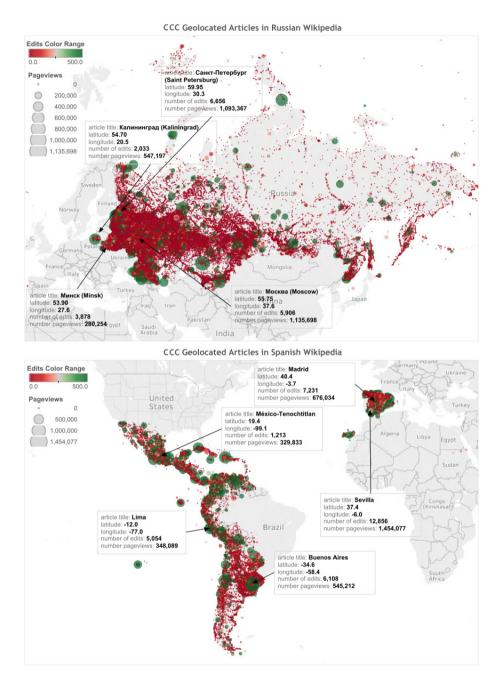
CCC Geolocated Articles in Macedonian Wikipedia

CCC Geolocated Articles in Romanian Wikipedia

Figure A5. Editor and reader engagement in CCC Geolocated articles from Macedonian and Romanian Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
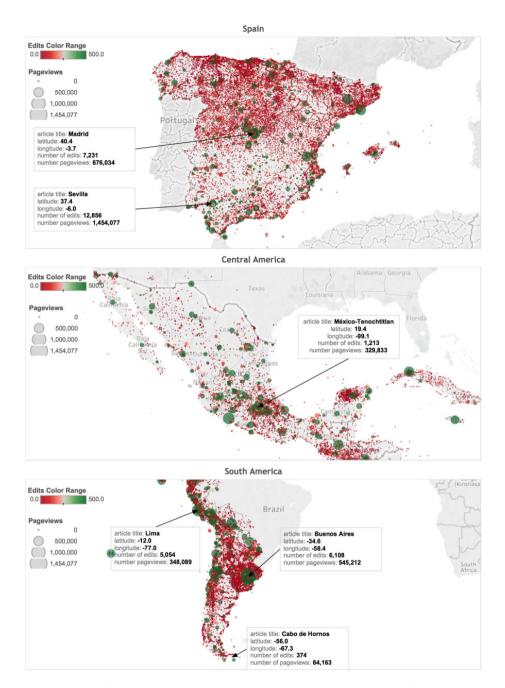
Figure A6. Editor and reader engagement in CCC Geolocated articles from Russian and Spanish Wikipedia (top and bottom, respectively). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.

Figure A7. Editor and reader engagement in CCC Geolocated articles from Spanish Wikipedia (zooms on Spain, Central America and South America). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
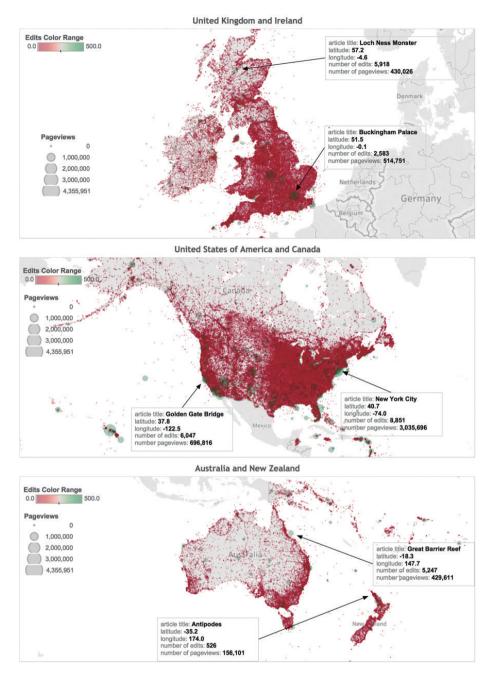
Figure A8. Editor and reader engagement in CCC Geolocated articles from English Wikipedia (zooms on the United Kingdom, United States of America and Canada, Australia and New Zealand). Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.
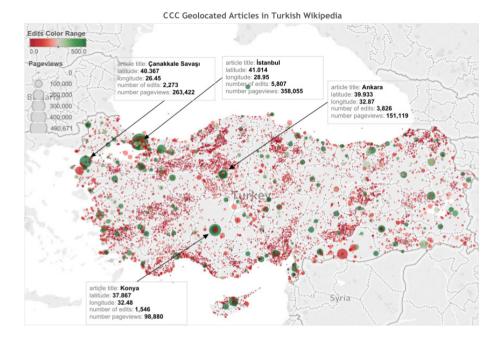
Figure A9. Editor and reader engagement in CCC Geolocated articles from Turkish Wikipedia. Each point is a CCC geolocated article. Colour represents the number of edits, depicted as a continuum from red to green with a middle point of 250 edits in colour beige. Size represents the number of pageviews. Important geolocated articles are marked with infoboxes.