Exploring Maintenance Practices in Crowd-Mapping

Giovanni Quattrone
Dept. of Computer Science
University College London, UK
g.quattrone@cs.ucl.ac.uk

Martin Dittus
Dept. of Computer Science
University College London, UK
martin.dittus.11@ucl.ac.uk

Licia Capra
Dept. of Computer Science
University College London, UK
l.capra@ucl.ac.uk

ABSTRACT

Crowd-mapping is a form of collaborative work that empowers users to gather and share geographic knowledge. OpenStreetMap is one of the most successful examples of such paradigm, where the goal of building a global map of the world is collectively performed by over 2M contributors. Despite geographic information being intrinsically evolving, little research has so far gone into analysing maintenance practices in these domains. In this paper, we perform a preliminary exploration to quantitatively capture maintenance dynamics in geographic crowd-sourced datasets, in terms of: the extent to which different maintenance actions are taking place, the type of spatial information that is being maintained, and who engages in these practices. We apply this method to 117 countries in OSM, over one year of mapping activity. Our findings reveal that, although maintenance practices vary substantially from country to country in terms of how widespread they are, strong commonalities exist in terms of what metadata is being maintained and by whom.

Keywords

Maintenance work; collaboration practices; OpenStreetMap; crowd-sourcing; volunteered geographic information

1. INTRODUCTION

Crowd-sourcing has become a successful paradigm for knowledge gathering, where a crowd is mobilised to collect and maintain large repositories of information [11, 3]. The most successful example to date of this paradigm is Wikipedia, with its online community of editors that voluntarily contribute to build and maintain the whole body of knowledge. Another type of knowledge where crowd-sourcing has been widely applied is that of volunteered geographic information, with citizens becoming surveyors, in council-monitoring applications like FixMyStreet;[1] local reporters, as powered by Ushaidi's Crowdmap;[2] and cartographers, in geo-wikis like Cyclopath[3] and OpenStreetMap.[4] It is the latter type of knowledge that we are interested in this paper.

Research has developed methods to quantitatively analyse the accuracy [2, 5, 6, 7, 8, 9, 17], coverage [27, 8, 18], growth [25], and bias [27, 12, 8, 24] of volunteered geographic information. These methods have mainly been applied to OpenStreetMap, making this dataset probably the most widely and deeply investigated geographic information repository to date. However, there is currently a gap in terms of methods to quantitatively capture *maintenance practices* in geographic crowd-sourcing communities like OpenStreetMap. Geographic information is naturally volatile and always evolving (e.g., where a grocery store is today, a coffee shop might be tomorrow); indeed, companies like Google spend several billions of dollars each year just to maintain their proprietary maps up-to-date and to improve their accuracy.[5] Yet little is known about maintenance practices of geographic crowd-sourced information: whether maintenance takes place at all and, if so, about what, and by whom.

In order to analyse maintenance practices in spatial crowd-sourced datasets, we have developed a method that quantitatively captures: *(i)* the different types of maintenance actions that take place (i.e., enrichment vs. correction vs. removal of existing information), and how widespread they are; *(ii)* what type of spatial objects (e.g., schools, hospitals, restaurants) are being maintained; and *(iii)* who is mostly engaged in maintenance practices. We have applied this method to OpenStreetMap, analysing one year of mapping activity in 117 different countries. Our findings reveal that maintenance practices vary substantially from country to country, both in terms of their adoption and in terms of the type of spatial objects that are being maintained. However, there are strong commonalities in terms of the metadata that is being maintained and who engages in maintenance practice. Based on these quantitative findings we elaborate on the implications of relevance to those interested in crowd-sourcing spatial information.

2. RELATED WORK

Maintenance practices have been extensively studied in online self-organised communities, most especially Wikipedia. In this domain, researchers often refer to collaborative practices, rather than maintenance ones, intended as the editing activity performed by different editors on the same Wikipedia article, for example to update its content or to improve its quality.

In Wikipedia, collaboration has been studied from two main different perspectives: *(i)* the information that is being maintained;

[1] http://www.fixmystreet.com/
[2] http://www.ushahidi.com/products/crowdmap
[3] http://cyclopath.org/
[4] http://www.openstreetmap.org/
[5] http://www.wired.com/2014/12/google-maps-ground-truth

and *(ii)* who performs this practice. We review some of the works in each of these themes next.

What information is being maintained. Kaltenbrunner and Laniado analysed the evolution over time of maintenance practices on different topics in Wikipedia [13]. They found that Wikipedia is the most up-to-date encyclopaedia ever seen, and that maintenance is often triggered by external events, with ongoing events being often edited and discussed on Wikipedia nearly in real-time. On the other hand, articles about historical or scientific facts (i.e., those that are not on people's minds) may take years to reach similar levels of user attention.

Who engages in collaboration practices. A study conducted by Laniado and Tasso [15] described the evolution of the user collaboration network in Wikipedia. They found that there exists a nucleus of very active contributors, who seem to spread over the whole wiki, and who interact preferentially with inexperienced users. Other studies that focused on users and their collaborative practices found that the top Wikipedia editors are those who are more involved in article maintenance, revising already existing articles, using quality assurance systems, and invoking community norms [10, 23].

Maintenance/collaboration practices is an active research area also for volunteered geographic information (VGI); however, in this context, current research is mainly investigating how to design tools to facilitate collaboration practices in crisis mapping [1, 4, 14] and little research has gone into analysing these practices more generally. As we shift our attention from encyclopaedic knowledge to spatial knowledge, different collaboration practices may be adopted. In fact, geographic repositories differ from classic encyclopaedic ones in two fundamental ways: *space* and *time*. Specifically, geographic content has an intrinsic spatial dimension, and there is a relationship between the location of a contributor and the type of knowledge that she can offer. Furthermore, compared to the body of knowledge that repositories like Wikipedia maintain, most geographic content is intrinsically volatile and continuously evolving, as a result of natural processes, such as urbanisation. As the nature of content varies, so might the corresponding editing practices. Indeed, a study conducted a few years ago by Mashhadi et al. [19] showed that some properties that typically hold in encyclopaedic type of crowd-sourcing repositories like Wikipedia, do not hold in geographic ones such as OpenStreetMap; for example, it was found that, in the former, the quality of an article depends on how much editing experience its contributors had in the past, while no relationship was found between quality of the map and editing experience of mappers in OSM.

In this paper, we aim to cover this gap, by proposing a method to quantitatively capture maintenance practices in spatial crowd-sourced datasets. Before presenting the method itself, and reporting on the results obtained, we first briefly illustrate the dataset we chose for analysis, provide a working definition of maintenance over such dataset, and spell out the research questions our method aims to answer.

3. DATASET

We chose to apply our method to OpenStreetMap (hereafter OSM), as this is to date the most successful example of spatial crowd-sourced dataset, having been running since 2004, and comprising the largest (and most geographically widespread) user and content base. Furthermore, OSM has been subject to extensive research, so that we can relate our findings to previous studies.

The OSM dataset is freely available to download[6] and contains the history from 2006 of all edits (over 2.7 billions) performed by all users (over 2 millions) on all spatial objects. In OSM jargon, spatial objects can be one of three types: *nodes*, *ways*, and *relations*. Nodes are single geo-spatial points and typically represent Points-of-Interest (POIs); ways mostly represent roads (as well as streams, railway lines, and the like); finally, relations are used for grouping other objects together, based on logical (and usually local) relationships (e.g., bus routes).

We filter the data in a number of ways before we begin our analysis. Specifically, we restricted our attention to edits of POIs only, i.e., specific point locations described in OSM by latitude/longitude coordinates, plus a variety of attributes (or tags). By focusing on this subset of OSM objects (instead of ways and relations), we aim to capture the actions of a wide range of contributors, from casual mappers to highly-engaged ones; indeed, as Mooney and Corcoran describe: "*Editing or adding tags to objects in OSM is technically one of the simplest operations which contributors can perform as there is very good support in all of the software and web-based editors for this edit action*" [20]. In OSM, a POI edit is represented as a tuple:

$$\langle uid, changeset, tstamp, ver, lat, lon, taglist \rangle$$

where uid identifies the user who performed this edit, $changeset$ denotes the editing session within which this edit was performed; $tstamp$ is the timestamp of when this edit took place; ver is a sequential value indicating the edit version of this POI (i.e., $ver = 1$ indicates the POI has just been created, while $ver > 1$ indicates the current edit is an update (i.e., maintenance) of an already existing POI); lat and lon denote the geographic coordinates of the POI. Finally, $taglist$ contains an arbitrary list of attribute-value pairs that further describe the POI; examples of such attributes are 'name' (e.g., 'Hollywood Cafe'), 'amenity type' (used to distinguish between different categories of POIs, such as 'restaurant', 'pubs', 'school'), address details, opening hours, accessibility considerations, and so on. For the purpose of this study, we consider POIs to be all OSM nodes that have either a *name* or an *amenity* tag at any point in the relevant period.[7] Finally, we ignored the tag *created_by*, as this is added automatically by editing software and does not reflect user intent.

The second pruning step we performed was time-based. We wanted to avoid the initial phase of OSM, when almost all contributions are creations of new objects, with little to no maintenance work taking place. We thus extracted all POI edits from January 1st to December 31st 2014.

From the above dataset, we make an attempt to identify contributions by human editors, while discarding automated contributions representing bulk data imports. In some regions, a significant portion of OSM contributions are automated imports of public domain map data sets, often produced by national mapping organisations or derived from historic map data.[8] While such data can play an important role in filling gaps on the map, it was not produced by the OSM community of volunteer contributors, and is not representative of human maintenance practices, which is the subject of this study. Imports are not explicitly marked as such in the OSM dataset; we thus needed heuristics to identify them. We applied the same approach used in [24], and marked as imports those edits which came from a single user, in very large quantities (i.e., more than one thousand edits), in a short period of time (i.e., less than one hour), and that were spread over a large geographic area (i.e., in the scale of a whole city).

The final part of this pruning process is to select the geographical

[6]http://www.geofabrik.de/data/download.html

[7]http://wiki.openstreetmap.org/wiki/Map_Features

[8]http://wiki.openstreetmap.org/wiki/Import

areas of the world to analyse. We chose to study maintenance practices at country level. From the above sample, we discard countries with too little OSM editing activity to be meaningfully analysed (i.e., countries with less than one thousand contributions during the period of study). We ended up with a dataset having around 3.4M edits, of 2.7M POIs, done by 80k users, over the 117 countries highlighted in Figure 1.

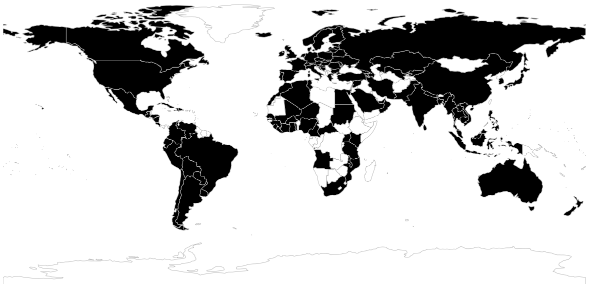

Figure 1: Map of the 117 Countries Under Analysis

4. FORMS OF MAINTENANCE

In order to quantitatively analyse maintenance of OSM information (and, more specifically, OSM POIs), we first need to automatically identify edits, in the OSM edit history, that are representative of such practice. We simply classify as maintenance actions all edits with $ver > 1$. Preliminary analysis shows that the time interval between two consecutive edits ($ver = n$ and $ver = n + 1$) is one week or longer for 90% of such edits, and that different users perform them.

We then distinguish three different forms of information maintenance, based on the type of *action* that took place since the POI previous version:

- *Add*, maintenance work where at least one new tag has been added to an existing POI (e.g., the tag 'opening_hours', along with its associated value, has been added to a restaurant already mapped in OSM).
- *Update*, maintenance work where the value of at least one of the already existing tags associated with a POI has been updated (e.g., the value of the tag 'amenity' has been changed from 'restaurant' to 'cafe', for a POI previously added to OSM).
- *Remove*, maintenance work where at least one tag has been deleted from an existing POI (e.g., the tag 'is_in', along with its associated value, has been removed from a POI present in OSM).

Note that the same edit may belong to different action classes (e.g., a single edit can both add a tag and update another). In our study, we will analyse them separately, as the drivers behind such actions can be quite different, and might thus result in different practices. In fact, intuitively, an add action can be seen as a sign of the user intent to *enrich* existing information, and it might be spurred by the emergence of novel location based services that require semantically richer POI information (e.g., opening hours, webpage). Conversely, an update action can be seen as a sign of the user intent to *correct* existing information; this may be the case for POIs that were last edited a long time ago, and thus now contain stale information (e.g., different business name or type), or the case for POIs whose name contains spelling mistakes. Finally, a remove action can be seen as a sign of the user intent to *polish* existing information; this may be the case for POIs that contain some deprecated tags.

5. RESEARCH QUESTIONS

In this work, we aim to explore the following research questions:

RQ1 (Spread) – How widespread is maintenance work? We begin our exploration by looking at the extent to which such practice is currently taking place across the 117 countries under exam.

RQ2 (What) – What information is being maintained? We then look more specifically at the type of information that is being maintained, to elicit POI information that is commonly maintained across all countries, if any, as well as potential regional differences.

RQ3 (Who) – Who is engaged in information maintenance? We finally shift our attention to the users performing maintenance edits, to understand whether this practice is evenly shared among editors, or whether it is undertaken by a select few.

To answer these questions we defined new metrics and conducted a large-scale quantitative analysis of maintenance practices of over 80k OSM mappers spread across 117 different countries.

6. METRICS AND RESULTS

RQ1 – How widespread is maintenance?

For each country under exam, we compute a *Maintenance Ratio* (MR), defined as the proportion of maintenance work that took place there, relative to the total number of edits (i.e., covering both creation and maintenance of POIs), for the period of study. Formally, let OSM_e be the set of OSM edits for a given country, and $OSM_m \subseteq OSM_e$ the set of OSM edits devoted to maintaining existing POIs. Then $MR = \frac{|OSM_m|}{|OSM_e|}$, $MR \in [0, 1]$. Intuitively, the closer this metric is to 1, the higher the proportion of maintenance work in that country (in 2014); vice versa, values close to zero indicate that almost all OSM editing activity is devoted to the creation of new POIs.

Table 1 shows the computed MR values in the 117 countries under exam, divided in quartiles. Maintenance practices vary widely: in a quarter of the analysed countries, maintenance is almost as frequent as the creation of new POIs ($MR > 0.42$), while there is another quarter of countries where maintenance is a much less widespread practice ($MR < 0.23$). There are also a few countries (e.g., Malawi, Mozambique, and Togo) where MR is almost zero, meaning that, in these countries, crowd workers are almost completely focused on the addition of new POIs, rather than in the maintenance of existing ones.

Min	*1st Qu.*	*Median*	*3rd Qu.*	*Max.*	*Freq. Distr*
0.02	0.23	0.33	0.42	0.77	

Table 1: Summary Statistics of Maintenance Ratio in the 117 Analysed Countries

To visualise where maintenance practices are taking place and to what extent, we report in Figure 2 a heatmap of MR values in the 117 analysed countries. Note that maintenance ratio is high

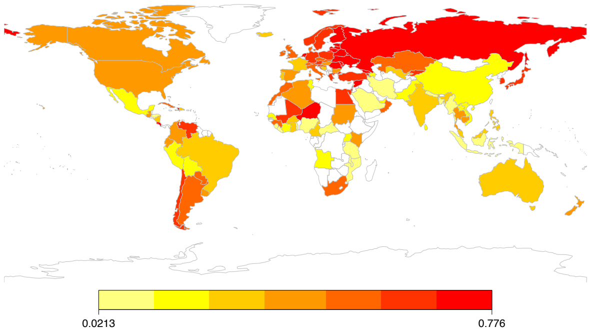

Figure 2: Maintenance Ratio in all Analysed Countries

both in countries with a long-standing OSM history (e.g., United Kingdom, Germany), and in countries where OSM activity started much later (e.g., Niger, Paraguay). Conversely, we find low values of MR in countries that joined OSM relatively recently (e.g., Peru, Angola), but also in countries that have had intense OSM activity for many years (e.g., France).

This result is somewhat surprising, as one might hypothesise that maintenance is a direct consequence of map maturity; that is, crowd workers first concentrate on adding information to a near empty map, and only later, as the map becomes richer and denser of information, crowd work starts to move towards maintaining what is already there. This phenomenon does hold, but only to a certain extent: we computed the number of OSM POIs mapped in a country, normalised by the area of the country, as proxy of OSM map maturity; we then computed the Spearman correlation [21] between such proxy and MR. As expected, we did obtain a positive correlation ($\rho = 0.39$, p-value < 0.001), meaning that maintenance practices are more widespread in countries where OSM map maturity is higher. However, the strength of the correlation is not very high. We hypothesise that a variety of local factors may play an important role in the rapid uptake of maintenance practice; for example, initiatives like the Humanitarian OpenStreetMap Team (HOT) may have fostered maintenance activities in countries with lower-than-average map maturity, in order to rapidly update the map following a natural disaster [22]. Also, rapid urbanisation in some regions may have caused faster than usual changes in the physical world, with consequent maintenance activity taking place on the digital map.

As different underlying phenomena may trigger maintenance work, we were interested in observing possibly different forms of maintenance (i.e., add, update or removal tags), when such practice does take place. To this purpose, we define a new metric, *Action Adoption* AA^{act} as the number of maintenance edits of type $act \in \{add, update, remove\}$ over the total number of maintenance edits that occurred in a country. Formally, let OSM_m be the set of OSM edits devoted to maintaining existing POIs, then $AA^{act} = \frac{|OSM_m^{act}|}{|OSM_m|}$, where $OSM_m^{act} \subseteq OSM_m$ is the set of OSM edits of action act over the initial set OSM_m. $AA^{act} \in [0, 1]$; a value of such metric close to 1 means that almost all maintenance edits are of action act, and vice versa.

Table 2 illustrates quartiles of the Action Adoption AA^{act} metric, $act \in \{add, update, remove\}$, binned over the 117 analysed countries. Results across all quartiles show that the add action (i.e., enriching existing information) is the most common one, usually more common than the update and remove actions combined (i.e.,

	Min	*1st Qu.*	*Median*	*3rd Qu.*	*Max.*	*Freq. Distr*
AA^{add}	0.14	0.42	0.52	0.59	0.88	
AA^{update}	0.07	0.24	0.32	0.38	0.75	
AA^{remove}	0.03	0.08	0.11	0.18	0.47	

Table 2: Summary Statistics of Action Adoption in the 117 Analysed Countries, divided by Adds, Updates, Removes

correcting existing information); furthermore, when correcting existing information, it is usually the case of updating an existing tag, rather than removing any. While this is the most common way of performing maintenance across the countries under exam, we also observe a big variance of the AA^{act} metric between the first and fourth quartiles, in each row (action) of Table 2 (i.e., AA^{add} ranges from 0.14 to 0.88, AA^{update} ranges from 0.07 to 0.75, and $AA^{removal}$ from 0.03 to 0.47). This suggests that there exist countries that do not follow the previously mentioned pattern. Indeed, we performed a manual investigation of some such cases and found that, for example, in Haiti, Turkey, and Niger the removal of tags is the most frequent maintenance practice performed; in Oman, Costa Rica and Azerbaijan, the updating of tags values is most frequent action instead.

RQ2 – What information is being maintained?

To understand if different types of POIs call for different levels of maintenance, we grouped POIs in each country according to their amenity type (e.g., restaurant, school, hospital, etc.). For each POI type, we then computed the corresponding Maintenance Ratio (MR) metric. We found a very skewed distribution in each country, with a minority of POI types (less than 10% of all types) being frequently maintained, and several hundreds of POI types receiving near zero maintenance instead. We then looked more closely into the frequently maintained ones, to see if there were commonalities among the analysed countries. Surprisingly, we found almost no overlap, with each country having a distinct set of POI types it maintains. For example, in the Netherlands, the most maintained POI types are ‘restaurant’, ‘cafe’ and ‘place_of_worship’, while in Russia the most maintained ones are ‘clinic’, ‘dentist’ and ‘public_building’. Although we do not know the cause, this result signals that different countries maintain distinct types of spatial information.

We then moved our attention from the types of spatial objects that are being maintained, to the set of *tags* that are being maintained, regardless of the POI type they refer to. To this purpose, we define *Tag Adoption* $TA_t^{act} = \frac{adoption_t^{act}}{|OSM_m^{act}|}$ as the ratio of the number of times tag t has been used for a certain action ($adoption_t^{act}$) over the total number of times action act occurred ($|OSM_m^{act}|$). $TA_t^{act} \in [0, 1]$; high values of TA_t^{act} indicate that tag t has been frequently used when act took place (e.g., tag ‘addr:street’ has been frequently used during an ‘add’ maintenance practice), and vice versa.

We computed Tag Adoption TA_t^{act} in each of the 117 countries under exams, for different maintenance actions $act \in \{$*add, update, remove*$\}$. We also computed this metric for the ‘creation’ action (i.e., when a POI is added to the map for the first time), to serve as a baseline. As for the case of POI types, in each country and for each action, we found a very skewed distribution, with only a minority of tags (less than 5%) being frequently edited. However, contrary to what we found for POI types, when zooming into this group of frequently edited tags, we found significant overlaps across countries. Table 3 reports both the name of the tags most frequently edited, and the number of countries in which such tags appeared in the list of the 5% most edited ones; the table further

Creation		*Maintenance*	
Adding a tag		*Adding a tag*	
Tag	*# Countries*	*Tag*	*# Countries*
name	117	name	108
amentiy	114	addr:street	44
place	78	addr:city	43
shop	64	wikipedia	31
addr:street	42	addr:housenumber	29
source	38	name:en	27
addr:city	31	addr:postcode	26
highway	22	source	24
addr:housenumber	21	operator	24
natural	16	name:ru	24
Maintenance			
Updating a tag		*Removing a tag*	
Tag	*# Countries*	*Tag*	*# Countries*
name	117	name	79
place	106	amenity	70
amenity	78	source	28
opening_hours	52	fixme	27
wikipedia	45	highway	26
shop	39	place	24
addr:street	31	building	21
source	30	note	18
name:en	27	is_in	16
website	24	wikipedia:en	16

Table 3: Top Ten Globally Adopted Tags for Each Action

distinguishes between a creation edit (top left of Table 3) and the three different types of maintenance edits (add, update, remove). For readability, only the top ten most globally adopted tags for each action are reported.

Although we cannot be sure of the rationale for these tags to be globally maintained, we can draw some interesting observations. As an example, let us consider the *addition* of tags – which, as seen before, is by far the most frequently performed maintenance practice worldwide. Aside from adding names to POIs that did not have one before, this practice seems to focus on address details (e.g, 'addr:street', 'addr:city', 'addr:housenumber', 'addr:postcode'). This corroborates the intuition that information maintenance is often subject to external drivers, such as the integration of location-based services over the base map,[9] which do require address information to operate effectively.

RQ3 – Who engages in information maintenance?

We finally move our attention from what information is being maintained to who takes charge of performing maintenance work. Previous studies of OSM have shown that there exists a small set of highly engaged (expert) users who are responsible for the majority of the mapping [24]; we wanted to investigate whether the very same users were also those taking charge of maintenance work. One might expect this to be the case for various reasons, ranging from motivation (i.e., the same drivers that make them map extensively may also drive them to maintain extensively), to knowledge (e.g., having previously contributed a lot of information, they might know what information is most stale and in need of updates), to skills (e.g., updating existing information may require users to have acquired a certain skill-set first, as was observed in other crowd sourcing communities like Wikipedia [16, 23, 26]).

To do so, we first grouped users within each country into five different classes of engagement (or expertise). We measured user's engagement using two alternative proxies: *(i)* $NumEdts$, that is their total number of OSM edits; and *(ii)* $ActDays$, that is the number of days during which they performed OSM editing activity.

[9]https://blog.openstreetmap.org/2015/02/16/routing-on-openstreetmap-org/

$NumEdts$		$ActDays$	
Class	*# Users*	*Class*	*# Users*
(0,1]	25,235	(0,1]	50,177
(1,10]	36,295	(1,5]	20,442
(10,100]	15,335	(5,10]	4,074
(100,1k]	3,927	(10,100]	6,100
(1k,10k]	606	(100,1k]	605

Table 4: Summary Statistics of Classes of Users

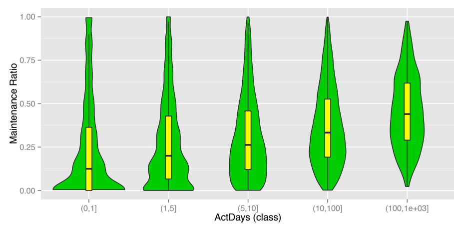

Figure 3: ActDays Vs. Maintenance Ratio

Summary statistics of the number of users per each class, across all countries, are reported in Table 4.

We computed the Maintenance Ratio (MR) metric defined before, but on a per user basis rather than on a per country basis. Figure 3 shows the quartiles (in yellow) and the frequency distributions (in green) of MR for each class of users according to the metric $ActDays$. Similar results were obtained when we grouped users according to the metric $NumEdts$. These results show that the more experienced the users are, the more effort they devote to maintaining existing POIs compared to the effort they spend to edit new ones.

7. CONCLUSION

In this paper, we have proposed a method to quantify maintenance work carried out in knowledge production communities, where knowledge has a distinct spatial nature, and where it naturally evolves over time. We have applied our method and metrics to OpenStreetMap in particular, one of the most successful examples of geographic crowd-sourced datasets. Our study of maintenance of OpenStreetMap POIs across 117 countries has revealed that such practice varies substantially from country to country, both in terms of its adoption, and in terms of the type of POIs that are being maintained. It has also revealed that, while the POI types that are being maintained differ, the tags that are being added/updated/removed are common across many countries. Our study also revealed that some maintenance actions, such as the addition of new tags to existing spatial objects, are more frequent than other actions, such as the updating or the removal of tags. At the moment, these maintenance actions are prevalently done by highly active users.

From a theoretical perspective, this work has presented a method to make visible the otherwise hidden maintenance practices of self-organised communities of practice interested in gathering and maintaining geographic knowledge. We have applied this method to a specific community and data type (OpenStreetMap and its POIs). However, we believe the same method can be used to study other data types within OpenStreetMap (e.g., ways and relations), as well as other crowd-mapping platforms, such as CrowdMap and FixMyStreet, for comparative studies. The method can also be reapplied to the same community and data type over time, in order to capture changes in behaviour, for example, as might be induced by major updates of the tools offered to support this practice.

8. REFERENCES

[1] J. Anderson, R. Soden, K. M. Anderson, M. Kogan, and L. Palen. EPIC-OSM: A Software Framework for OpenStreetMap Data Analytics. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 5468–5477. IEEE, 2016.

[2] J. Arsanjani, C. Barron, M. Bakillah, and M. Helbich. Assessing the Quality of OpenStreetMap Contributors together with their Contributions. In *Proc. of AGILE*, 2013.

[3] D. Brabham. *Crowdsourcing*. MIT Press, 2013.

[4] M. Dittus, G. Quattrone, and L. Capra. Analysing volunteer engagement in humanitarian mapping: building contributor communities at large scale. In *Proc. of CSCW*. ACM, 2016.

[5] H. Fana, A. Zipfa, Q. Fub, and P. Neisa. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science (IJGIS)*, 28(4):700âĂŞ–719, 2014.

[6] J. Girres and G. Touya. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4):435–459, 2010.

[7] M. Goodchild. Citizens as Sensors: the World of Volunteered Geography. *GeoJournal*, 69(4):211–221, 2007.

[8] M. Haklay. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, 37(4):682–703, 2010.

[9] M. Haklay, S. Basiouka, V. Antoniou, and A. Ather. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus Law to Volunteered Geographic Information. *The Cartographic Journal*, 47(4):315–322, 2010.

[10] A. Halfaker, R. Geiger, J. Morgan, and J. Riedl. The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist*, 57(5):664–688, 2013.

[11] J. Howe. The Rise of Crowdsourcing. *Wired*, 2006.

[12] K. Ishida. Geographical Bias on Social Media and Geo-local Contents System with Mobile Devices. In *Proc. of HICSS*, pages 1790–1796, 2012.

[13] A. Kaltenbrunner and D. Laniado. There is no deadline: time evolution of wikipedia discussions. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, page 6. ACM, 2012.

[14] M. Kogan, J. Anderson, R. Soden, K. M. Anderson, and L. Palen. Collaboration in OpenStreetMap: A Network Analysis of Two Humanitarian Events. In *Proc. of CHI*, 2016.

[15] D. Laniado and R. Tasso. Co-authorship 2.0: Patterns of collaboration in wikipedia. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, pages 201–210. ACM, 2011.

[16] J. Liu and S. Ram. Who does what: Collaboration patterns in the wikipedia and their impact on article quality. *ACM Transactions on Management Information Systems (TMIS)*, 2(2):11, 2011.

[17] I. Ludwig, A. Voss, and M. Krause-Traudes. A Comparison of the Street Networks of Navteq and OSM in Germany. *Advancing Geoinformation Science for a Changing World*, 1(2):65–84, 2011.

[18] A. Mashhadi, G. Quattrone, and L. Capra. Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 611–622. ACM, 2013.

[19] A. Mashhadi, G. Quattrone, L. Capra, and P. Mooney. On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases. In *Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, page 15. ACM, 2012.

[20] P. Mooney and P. Corcoran. Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS 2013*, 2013.

[21] J. Myers and A. Well. *Research Design and Statistical Analysis (2nd ed.)*. Routledge, 2003.

[22] L. Palen, R. Soden, T. J. Anderson, and M. Barrenechea. Success & scale in a data-producing organization: The socio-technical evolution of openstreetmap in response to humanitarian events. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pages 4113–4122. ACM, 2015.

[23] K. Panciera, A. Halfaker, and L. Terveen. Wikipedians are born, not made: a study of power editors on wikipedia. In *Proceedings of the ACM 2009 international conference on Supporting group work*, pages 51–60. ACM, 2009.

[24] G. Quattrone, L. Capra, and P. De Meo. There's no such thing as the perfect map: Quantifying bias in spatial crowd-sourcing datasets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1021–1032. ACM, 2015.

[25] G. Quattrone, A. Mashhadi, D. Quercia, C. Smith-Clarke, and L. Capra. Modelling growth of urban crowd-sourced information. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 563–572. ACM, 2014.

[26] D. M. Wilkinson and B. A. Huberman. Assessing the value of coooperation in wikipedia. *arXiv preprint cs/0702140*, 2007.

[27] D. Zielstra and A. Zipf. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In *Proc. of AGILE*, 2010.