

On the Accuracy of Urban Crowd-Sourcing for Maintaining Large-Scale Geospatial Databases*

Afra Mashhadi
Dept. of Computer Science
University College London
Gower Street
London WC1E 6BT, UK
a.jahanbakhsh Mashhadi@cs.ucl.ac.uk

Licia Capra
Dept. of Computer Science
University College London
Gower Street
London WC1E 6BT, UK
l.capra@cs.ucl.ac.uk

Giovanni Quattrone
Dept. of Computer Science
University College London
Gower Street
London WC1E 6BT, UK
g.quattrone@cs.ucl.ac.uk

Peter Mooney
Dept. of Computer Science
National University of Ireland
Maynooth
Co. Kildare, Ireland
peter.mooney@nuim.ie

ABSTRACT

The world is in the midst of an immense population shift from rural areas to cities. Urban elements, such as businesses, Points-of-Interest (POIs), transportation, and housing are continuously changing, and collecting and maintaining accurate information about these elements within spatial databases has become an incredibly onerous task. A solution made possible by the uptake of social media is crowd-sourcing, where user-generated content can be cultivated into meaningful and informative collections, as exemplified by sites like Wikipedia. This form of user-contributed content is no longer confined to the Web: equipped with powerful mobile devices, citizens have become cartographers too, volunteering geographic information (e.g., POIs) as exemplified by sites like OpenStreetMap. In this paper, we investigate the extent to which crowd-sourcing can be relied upon to build and maintain an accurate map of the changing world, by means of a thorough analysis and comparison between traditional web-based crowd-sourcing (as in Wikipedia) and urban crowd-sourcing (as in OpenStreetMap).

*Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
WikiSym '12, Aug 27–29, 2012, Linz, Austria. Copyright © 2012 ACM 978-1-4503-1605-7/12/08... \$15.00

Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications – Spatial Databases and GIS; H.5.3 [Group and Organization Interfaces]: Collaborative computing, computer-supported cooperative work

General Terms

Human Factors, Measurement, Reliability

1. INTRODUCTION

The share of the world's population living in cities has recently surpassed 50%, and it is expected that by 2025 another 1.2 billion people will be living in urban areas. Governments worldwide are supportive of this urbanisation process, as it is expected that economies of scale will make concentrated urban centres more productive than rural areas [2]. However, these benefits will be only realised if we are able to manage the increased complexity that comes with larger cities. Metropolitan cities are in fact very dynamic entities, with urban elements such as businesses, cultural and social Points-of-Interests (POIs), housing, transportation and the like, continuously changing. Collecting and maintaining accurate information about urban elements within spatial databases will thus become an incredibly onerous, yet essential task. Who is going to undertake it? At the moment, there exist commercial companies (such as Google and Navteq) that offer such service. However, the temporal accuracy of the information they provide has started to be challenged¹. A question arises as to whether a single company will ever be able to maintain comprehensive and accurate geographical information (both topographical and non topographical), especially as metropolitan cities, once just a handful few (i.e., New York, Tokyo and London), are now fast growing in number, size and dynamism.

In the context of Web 2.0, *crowd-sourcing* has emerged as a new paradigm that leverages community (or crowd) participation to effectively and efficiently accomplish a task traditionally undertaken by a few selected individuals. Wikipedia

¹http://www.sciuridae.co.uk/sat_nav_map_accuracy.htm

is probably the most successful example of web-based crowd-sourcing, with its online community of editors voluntarily contributing to gathering together the whole body of knowledge, a task once attributed to a selected few experts. A study reported in 2005 in *Nature* [5] found that Wikipedia articles had only slightly more errors than the *Encyclopaedia Britannica* for a sample of science-related articles, thus demonstrating the viability of crowd-sourcing as an effective and efficient way to gather knowledge on the web.

Geographical information is fundamentally different from web-based information, both because of its temporal dynamism (e.g., a new restaurant opening, while a previous business closes), and its spatial dimension (e.g., only someone who has physically been *there* has the knowledge to generate this information). Yet the recent popularity and widespread adoption of smart-phones has brought forward the crowd-sourcing paradigm as one worth considering in the urban domain too, with citizens becoming surveyors, using council-monitoring applications like *FixMyStreet*²; reporters, using micro-blogging services such as *Twitter*; and cartographers, with geo-wikis like *Cyclopath*³ and *OpenStreetMap*⁴. Indeed, these services offer mobile applications that users can deploy directly on their smart-phones to generate geo-tagged content *on the go*. Note that, in this paper, we focus on geographic information that requires explicit user input (*active* crowd-sourcing), as opposed to information automatically gathered by GPS sensors (*passive* crowd-sourcing) – a technique already used by both *Navteq* and *Google*. The question we then ask ourselves is the following: can *crowd-sourcing* be a viable way for maintaining *accurate* geographical information in *urban* environments?

To answer this question, we study some of the factors that are proven to have contributed to the success of *Wikipedia*. We group those factors as contributor-based and object-based properties and ask whether the same properties affecting accuracy are exhibited in urban crowd-sourcing. We note that although urban and web-based crowd-sourcing applications may serve different purposes, they both share same fundamental properties: firstly, they both rely on crowds (of registered users) to voluntarily contribute knowledge by following a non-monetary incentive model. Secondly, objects in both cases are created and modified by those users over time, resulting in a sustainable quality control mechanism. Given these similarities, one can expect the factors influencing quality of *Wikipedia* (e.g., the number of previous contributions of the editors) to similarly apply to urban crowd-sourcing too.

Therefore in this paper, we draw a comparative analysis between *Wikipedia* and *OpenStreetMaps* (OSM), a prime example of what we call *urban crowd-sourcing*. OSM currently boasts 547,270 users collectively building a free, openly accessible, editable map of the world. OSM is an example of crowd-sourcing with a strong urban component, both because of the spatio-temporal nature of the knowledge it gathers (map elements of the changing world), and because of the way such knowledge is contributed (i.e., OSM contributors

are asked not to use existing maps when editing OSM objects due to copyright issues; it can therefore be assumed that editing a urban element is done by a citizen who has actually visited that location).

The reminder of the paper is structured as follow: Section 2 presents an overview of the related literature. In Section 3, we summarise key research findings that relate *crowd behaviour* with *quality of content* in *Wikipedia*. In Section 4, we then undertake a detailed analysis of OSM to determine if the same correlation between crowd behaviour and quality of content exists. In Section 5, we first offer a definition of what quality means in urban crowd-sourcing settings; we then present the obtained results for OSM. In Section 6 we discuss the main findings of the results presented before. Finally we conclude this paper by presenting our ongoing and future work in Section 7.

2. RELATED WORK

The recent explosion of open content systems like *Wikipedia* has led to a new industry of online knowledge production and organization, carried out by distributed volunteers. Volunteered Geographical Information (VGI) [7] is only a special case of this larger Web phenomenon, with many examples beyond *OpenStreetMap*, including: *Google Map Maker* (<http://www.google.com/mapmaker>), a crowd-sourcing system designed to allow people to update the actual *Google Maps*; *Wikimapia* (<http://wikimapia.org/>), a system that combines *Google Maps* with a wiki system, allowing users to add information, in the form of a note, to any location on earth; and *Cyclopath*, a geographic wiki and route-finder for bicyclists developed by *GroupLens Research* at the University of Minnesota.

In order to offer services based on crowd-sourced geographic information, the quality of such information must be assessed first. Quality of VGI has been measured in comparison to traditional geographical datasets maintained by national mapping agencies (e.g., *Ordnance Survey*), as well as proprietary datasets maintained by commercial companies (e.g., *Navteq*). For example, *Haklay et al.* [8, 9] measured the *positional accuracy* of OSM *road* networks in the UK and found it to be very high (i.e., on average within 6 meters of the position recorded by the *Ordnance Survey*). The authors also investigated the impact of the number of contributors on positional accuracy, and estimated that high accuracy is achieved when there are at least 15 contributors per square kilometre. Works such as [6, 12] have confirmed these observations for countries like *France*, *Germany* and *Switzerland*. *Girres et al.* [6] also discovered a non-linear correlation between the number of OSM objects in an area, in relation to the number of contributors in the area (i.e., areas with up to three contributors per square kilometre had ten times more contributions than areas with only one contributor, and areas with more than three contributors had up to hundred times more contributions).

We note that the attention of the VGI community has focused on road networks only (i.e., the *way* objects in OSM). The contribution process associated with editing roads and that associated with editing POIs differ greatly: indeed, the former is typically done by users who have high expertise in both the geography of an area and the editing tools required

²<http://www.fixmystreet.com/>

³<http://cyclopath.org/>

⁴<http://www.openstreetmap.org/>

to digitally represent it, while the latter can be performed by any city dweller owning a smart-phone. The study we have presented in this paper is thus orthogonal to those conducted in the VGI community and, to the best of our knowledge, it is the first study to offer insights into the reliability of *urban crowd-sourcing* as a means to collect accurate POI data in urban environments.

Very relevant to our ongoing and future work is the study conducted in [16], where the authors analysed contributors' behaviour in the Cyclopath service. In particular, they compared Cyclopath with Wikipedia, to see if the users of the two systems exhibited similar behaviours when consuming, exploring, and editing data. The authors found that, like in Wikipedia (as we shall demonstrate for OSM), Cyclopath user edits exhibit a 'long tail', with a few users being responsible for the vast majority of contributions, while the rest contributing very little instead. They also studied users' behaviour as it changes over time, and found that power users edited much more than non power users immediately upon appearing, and that all editors' activity was characterized by an initial burst of intense activity followed by gradual decline to a fairly low, constant level. However, in this paper we only focus on the 'static' analysis of OSM quality and we leave the 'dynamic' analysis of quality over time for future work.

It is worth pointing out that the work that is being presented in this paper focuses on quality measured in terms of *accuracy*. A complimentary metric that is particularly relevant in the VGI community is *coverage*, that is, what portion of the physical world is being digitally mapped? A recent study by Zielstra et al. [25] investigated the issue of coverage in Germany, once again focusing on the OSM road network; their findings show a sharp decrease in coverage, compared to Navteq, as one moves away from city centres. However, we do not dwell into the issue of coverage in this paper.

3. WEB-BASED CROWD-SOURCING

Web-based crowd-sourcing refers to a paradigm where the online (web-based) community cooperates to accomplish a task traditionally assigned to a selected few. Wikipedia is probably the largest example of this new paradigm, with one of the biggest communities of Internet users, voluntarily editing over 2 million articles in more than 282 different language editions. Most importantly, the *quality* of Wikipedia articles has been found to be comparable to that of expertly-compiled encyclopaedia [11]. To understand the reasons behind this success, researchers have studied the *dynamics* of its contributors and of its content in depth, to see how these relate to the quality of Wikipedia content [18, 10, 19, 11]. We summarise the main results from the literature below.

Wikipedia Contributors. Despite an extremely large user-base, the vast majority of Wikipedia users (99%) are viewers-only and do not edit articles. Yet a great sense of collaboration and motivation exists amongst the remaining 1% [13]. Delving further into the characteristics of these Wikipedia's authors, it has been shown that authors' contributions follow a power-law distribution, with a few authors contributing a lot, and the

majority contributing only a few [21, 15, 14]. More precisely, analysis has shown that the top 10% of editors (by edit count) were credited with 86% of persistent word views (PWV), the top 1% with about 70%, and the top 0.1% (4200 users) were attributed 44% of PWVs [17]. Figure 1, taken from [21], illustrates this property by presenting *ranked authors* on the logarithmic x-axis and their corresponding *number of edits* on the logarithmic y-axis, for different language versions of Wikipedia. As shown, tens of thousands of authors edit the English version of Wikipedia less than 10 times in total, whereas the top 10 users contribute to the same version of Wikipedia in the order of 10,000 edits. It is also interesting to note that, the less popular the language version of Wikipedia, the lower the number of engaged authors and scale of edits.

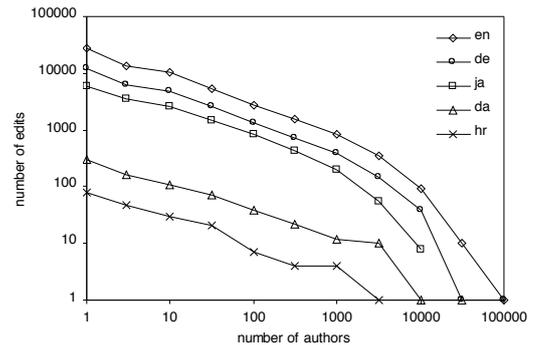


Figure 1: User Activity in Wikipedia

Wikipedia Articles. As [24] showed, not only the number of edits per author follows a power-law distribution, but also the number of edits per article, as illustrated in Figure 2. In other words, a few articles have received thousands of edits, while the majority of articles have received a few edits only instead.

Having observed these user and article dynamics, the question that naturally arises is whether (and how) these dynamics relate to the quality of Wikipedia articles. In this context, the research community has defined quality of an article in a variety of ways, including article popularity [3, 24], correctness of the grammar [22], and stability of the edits [1, 20]. Based on these definitions, the following four properties have been found to hold true:

PROPERTY 1. Users who edit a lot (also termed as power-users) provide better quality content. \square

A first study conducted on 'featured articles', that is, articles that have been identified as having the highest quality by Wikipedia's article assessment project, found that there exists a strong correlation between editors' activity levels and the quality of their edits [24]. This result has been independently confirmed by subsequent studies too [4, 15]. It is worth noting that, in the context of Wikipedia, a certain skill-set is required by authors to contribute good quality

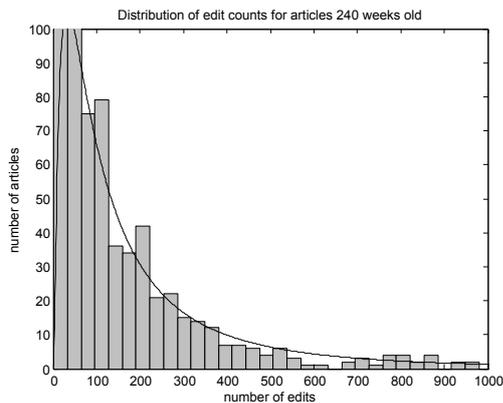


Figure 2: Distribution of Edits per Article in English Wikipedia

content (e.g., how to structure a document, how to create references, etc.). It is thus not surprising that the more edits an author makes, the more this skill-set develops and the quality of their contribution improves.

PROPERTY 2. The addition of a link, heading, or other *structural elements* to an article tends to indicate a higher quality of the editor. \square

A more fine-grained study was conducted by Druck et al., to understand if there existed a correlation between the type of edits performed, and the quality that the author contributed [4]. They found that users who make the effort to provide additional features during their edits often provide higher quality content.

PROPERTY 3. Users who previously contributed high (low) quality edits tend to continue to submit high (low) quality edits. \square

A further observation made in [4] and summarised above is that the quality of a contributor remains stable over time: users who have previously contributed good quality content are likely to continue doing so in the future. It is worth noting that this property may be a consequence of Wikipedia’s strong monitoring policy, which identifies inappropriate contributions and blacklists the responsible authors. By investing in this (resource-expensive) policy, Wikipedia is capable of limiting attacks caused by spammers, malicious users and bots.

PROPERTY 4. The higher the number of edits, the higher the quality of the article. \square

The last property moves the focus from editors to articles: as observed in [23], continuous editing of an article over time leads to it becoming more mature and increasing in quality. Note that this observation is agnostic of content type/topic, as the study looked at article quality purely in relation to the number of edits it received.

Having reviewed the factors (in terms of users’ and articles’ activity) that correlate with quality in web-based crowd-sourcing (and Wikipedia in particular), we next examine whether the same properties hold true in urban crowd-sourcing settings. To do so, we first introduce the dataset we have used for this investigation, that is, OpenStreetMap.

4. URBAN CROWD-SOURCING DATASET

Content gathered via urban crowd-sourcing differs from that collected via web-based crowd-sourcing in two fundamental ways: *space* and *time*. More precisely, urban content has an intrinsic spatial dimension, as it refers to some physical entity that exists in the real world (e.g., a building, a parking spot); such content is rather volatile if compared to the more static body of knowledge that an encyclopaedia collects, as urban elements dynamically change as cities grow and evolve. As the nature of content varies, so does the crowd that can gather it: while virtually anyone can contribute content in Wikipedia, in urban settings there exists an intrinsic correlation between the content to be gathered and who possesses the *knowledge* to gather it (that is, city dwellers that observe these urban elements at the right time and place). We thus question whether the properties found in web-based crowd-sourcing hold in urban settings too: that is, can urban crowd-sourcing be a suitable paradigm to create and maintain an accurate mapping between the digital and the physical world?

To answer this question, we have analysed OSM, the most famous example of Volunteered Geographic Information (VGI) publicly available today. OSM is a prime example of a large-scale, urban crowd-sourcing dataset, where registered users can contribute spatial content describing map features (such as roads and Points-of-Interest) to the global OSM database, thus collectively building a free, openly accessible, editable map of the world. Note that, while OSM is hosted on the Internet and is primarily a web-based system, it exhibits urban crowd-sourcing properties, both in the spatio-temporal nature of the knowledge it gathers, and in the way such knowledge is contributed: as OSM contributors are asked not to use existing maps when editing OSM objects (i.e., due to copyright issues), it can be assumed that editing an urban element is done by a citizen who has actually visited that location (e.g., via a mobile application running on their smart-phone and exploiting on-board GPS sensors).

The OSM dataset is freely available to download and contains the history of all edits (since 2006) on all spatial objects performed by all users. Spatial objects can be one of three types: *nodes*, *ways*, and *relations*. Nodes broadly refer to POIs, ways are representative of roads, and relations are used for grouping other objects together. For the purpose of this study, we restricted our attention to POIs only, and not roads; the rationale for this choice was to focus our attention on contributors (the urban crowd) from whom no specialised skill-set was required (as it is the case when editing roads instead). In particular, we focus on those nodes that represent urban elements commonly interpreted as POIs, such as cafes, restaurants, etc, but not elements of minor importance like letterboxes, telephone booths, etc. A node consists of three main attributes: a geographical position (latitude and longitude), a name, and an amenity type (e.g., hospital, cafe). While the geographical position is

compulsory, name and amenity are optional fields in OSM. Finally, we selected two cities to analyse: London, UK, as an example of a very well represented large metropolitan city in OSM, and Rome, Italy, as an example of a large city which is steadily increasing its spatial representation in OSM. To ensure we are considering genuine urban crowd-sourcing contributions, and not those made by bots, we have eliminated from the dataset those users who performed an excessive number of edits in a very short time (i.e., those who edited more than 40 POIs in a single *changeset* session in OSM); in this way we filtered out 71 users (out of 2302) from London and 13 users (out of 376) from Rome. The datasets we are left with are summarised below.

City	# Users	# POIs	# Edits
London	2,232	32,438	44,463
Rome	363	3,315	5,319

We have conducted a preliminary analysis of these datasets, to characterise both OSM contributors and OSM objects, as previously done for Wikipedia.

OSM Contributors. OSM contributors and Wikipedia contributors are very similar in terms of number of edits they contribute, both following a power-law distribution. Figure 3 illustrates this fact for OSM, with number of edits displayed in logarithmic scale: as shown, there exist a few users who are heavily engaged in editing POIs, while the majority of citizens edits just a few. A noticeable difference in terms of scale appears between London and Rome, with the latter having much fewer editors and edits (at least one order of magnitude less); indeed, our pre-analysis reveals that OSM is far less popular in Italy than in the UK (especially London where OSM was born). We will return to this important difference when analysing quality of contributed content in these cities.

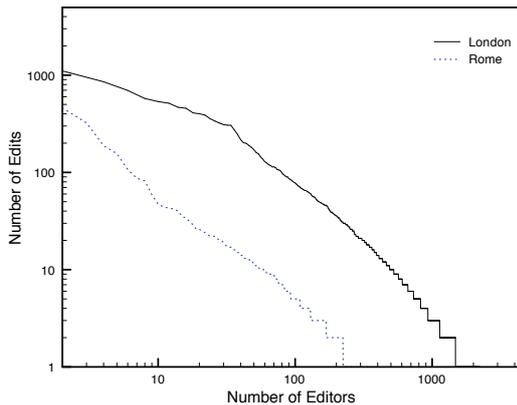


Figure 3: Distribution of Users' Edits in OSM

OSM Objects. As Figure 4 illustrates, OSM objects are edited following a power-law distribution, as was the

case for Wikipedia articles. However, the scale is considerably different. More precisely, in the case of Wikipedia, the majority of articles have on average 50 edits, whereas in OSM the majority of POIs have 1 or 2 edits only. Furthermore, while in Wikipedia the long tail stretches up to 1000 edits per article, the maximum observed number of edits per POI is approximately 50 in OSM. This is not surprising, as the nature of content is fundamentally different in the two scenarios: while Wikipedia articles can be edited many times to add more detailed information to them, POIs are much more limited in the number of editable features, these being position, name and amenity. As number of edits per article was an important parameter in Wikipedia to estimate the quality of content (Property 4), we will discuss this aspect of OSM quality in the next section.

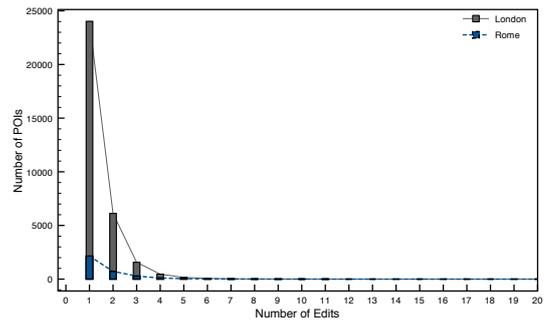


Figure 4: Distribution of Edits per POIs in OSM

5. URBAN CROWD-SOURCING ANALYSIS

Research conducted on Wikipedia has revealed that the quality of content is high, and that such quality is correlated with the editing characteristics of authors and content (Properties 1–4). We will now investigate whether the same can be said for OSM: are OSM objects of high quality? And if so, what characteristics of OSM editors and objects are good predictors of quality (do Properties 1–4 still hold)? By answering these questions, we aim to offer insights into the viability of crowd-sourcing as a means to gather and maintain accurate information about urban environments.

To begin with, we need to: (1) define benchmarks against which to compare quality; and (2) define quality metrics for OSM objects.

Benchmarks. We considered two different commercial geographic information systems, covering the same type of information (in terms of POIs) as OSM: *Navteq* and *Yelp*. Navteq (<http://www.navteq.com/>) is the leading global provider of maps and location data, covering not only roads but also millions of POIs of varying nature, from restaurants to hospitals and gas stations. Being a commercial service, Navteq's primary objective is to ensure the highest level of accuracy of its data (the information there contained is factually correct and up-to-date). Yelp (<http://www.yelp.com/>) focuses on business listings, from store-fronts (e.g., restaurants and shops) to services (e.g., doctors, hotels, and cultural venues). Once again, Yelp business model requires a high level of accuracy of its listings. We then built our

benchmark (or *ground truth* dataset) as the set-intersection of Navteq and Yelp data; in doing so, a POI in Navteq is considered to be the same POI in Yelp if the name is the same and the geographic distance is less than 20 meters (in the ground-truth dataset, the location of such POI is recorded as the point that is equidistant to the two locations recorded in their original datasets).

Metrics. In both OSM and in the ground-truth dataset, a POI is defined as a triple: $poi = \langle name, amenity, (lat, lon) \rangle$, where $name$ is the POI’s name, $amenity$ is its category (e.g., cafe, restaurant), and (lat, lon) are the coordinates defining its geographical position⁵. We then quantify *quality* of OSM data in terms of its *geographic error*, *lexicographic error*, and *amenity error* with respect to ground-truth data. More precisely, let poi_x be a single POI, and POI_x the set of all POIs, with x being either the OSM dataset or the ground-truth dataset (to which we will refer, for convenience, simply as gt). To be able to measure error we first need to relate POIs in OSM with the same POIs in the ground-truth dataset in an automatic way. We thus state that poi_{osm} is *equivalent* to poi_{gt} ($poi_{osm} \equiv poi_{gt}$) if both their geographic and lexicographic differences are small. The geographic difference $geo_{Err}(\cdot, \cdot)$ is computed as the Euclidean distance between the two points, and this is considered small if it is less than 100 meters. The lexicographic difference $lexical_{Err}(\cdot, \cdot)$ is computed as the Levenshtein distance between the POIs names, and this is considered small if the ratio between such distance and the maximum one for the pair is less than 0.35⁶. Having defined these distance metrics, we can now quantify errors. Let $POI_{osm}^{eq} \subseteq POI_{osm}$ and $POI_{gt}^{eq} \subseteq POI_{gt}$ be the sets of OSM POIs and ground-truth POIs that are deemed ‘equivalent’ to each other. Geographic, lexicographic and amenity errors are then computed as:

$$\begin{aligned} Geo_{Err} &= AVG(geo_{Err}(poi_{osm}, poi_{gt})) \\ Lexical_{Err} &= AVG(lexical_{Err}(poi_{osm}, poi_{gt})) \\ Amenity_{Err} &= AVG(amenity_{Err}(poi_{osm}, poi_{gt})) \end{aligned}$$

where: $Amenity_{Err}(\cdot, \cdot)$ is defined equal to 0 if the two input POIs have the very same amenity type classification, 1 otherwise⁷; the average function AVG is computed on all $poi_{osm} \equiv poi_{gt}, poi_{osm} \in POI_{osm}^{eq}, poi_{gt} \in POI_{gt}^{eq}$. The lower Geo_{Err} , $Lexical_{Err}$, and $Amenity_{Err}$ are, the more accurate the information stored in OSM.

Having defined quality metrics for OSM, we now present the main results of our analysis, which aims to answer two fundamental questions: first, what is the quality of OSM data? And second, what characteristics of OSM users and content are good predictors of quality?

5.1 Quality of OSM POIs

⁵Note that, while name and amenity are optional fields in OSM, they are mandatory in both Navteq and Yelp, thus always present in our ground-truth dataset.

⁶These values were chosen after manual inspection of a number of POIs jointly presents in the two datasets that we knew to be the same.

⁷Note that to compute this function we have manually matched all the amenities defined in OSM with those defined in the ground truth.

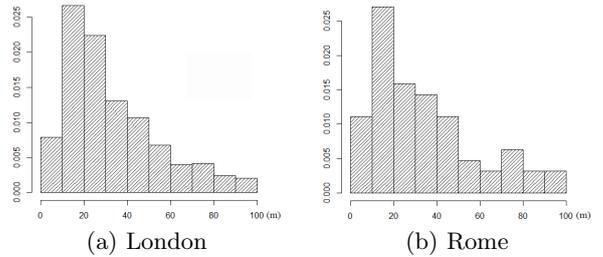


Figure 5: Normalised Density Distributions of Geographical Errors

Table 1 summarises the quality of OSM POIs for both London and Rome, where quality has been measured in terms of average geographic error, lexicographic error, and amenity error, as defined above. Figures 5 and 6 further show the normalised histograms (total area under the histogram equals 1) approximating the distributions of geographic and lexicographic errors. We have normalised the distributions to ease comparison between London and Rome. Note that we do not show the distribution of amenity errors because they can assume only one of two values, that is, either 0 or 1.

City	Geo _{Err}	Lexical _{Err}	Amenity _{Err}
London	24 m	0.13	0.36
Rome	34 m	0.07	0.45

Table 1: Quality of OSM POIs

The results indicate an overall *high quality* of information for OSM POIs: geographic errors are almost normally distributed and their average value is less than 25 meters in London (and less than 35 meters for Rome), thus revealing accurate positioning of POIs on a map with respect to the ground truth dataset. Lexicographic distances are almost all zero, thus revealing accuracy in spelling names of POIs. At first glance, there appears to be a high error in the classification of the amenity type instead; however, this quite high error rate may be due to the fact that the amenity field is optional in OSM, and could thus be null in many cases. We have investigated further in this direction, and Figure 7 separately shows the percentage of cases where there exists a match between OSM amenities and ground truth ones, the percentage of cases where there is a mismatch, and the percentage of cases where the amenity field is null in OSM, thus a match cannot be computed. As the results illustrate, despite the amenity field being optional in OSM, contributors take the burden of providing this information in the majority of cases (67% of cases in London, and 59% in Rome); in such cases, the information they provide is indeed correct, with less than 5% of mismatches, both in London and Rome.

This overall high level of quality in OSM is an extremely positive result, as it indeed suggests the suitability of urban crowd-sourcing as a means of collecting quality information in dynamic urban settings. Furthermore, this result has been observed in a system that is completely self-policed: in fact, unlike Wikipedia, OSM uses no quality assurance technique (i.e., there exist no moderators to monitor and remove incorrect content). Urban crowd-sourcing thus appears to be not only an accurate but also a very sustainable paradigm

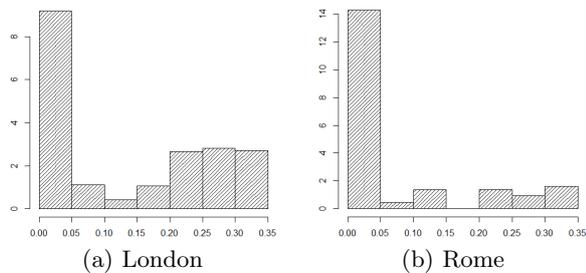


Figure 6: Normalised Density Distributions of Lexicographic Errors

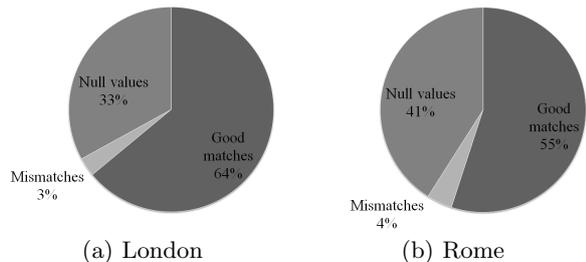


Figure 7: Distribution of Amenity Matches

to maintain digitised information of urban settings.

5.2 Factors Leading to Quality Data

Having observed the high level of quality of OSM POI data, we then proceed to unveil the factors contributing to such quality. In particular, we aim to test whether the same four properties that researchers have observed on Wikipedia (and thus web-based crowd-sourcing) hold true in the case of OSM (and thus urban crowd-sourcing) too.

Property 1. The first property observed in Wikipedia states that authors who edit a lot (i.e., power-users) contribute better quality content. To test whether this property holds in OSM too, we have investigated whether there exists any statistical correlation between error (be that geographic, lexicographic or amenity error) and the number of user edits. The results reveal no significant correlation between these parameters for either London or Rome; this would suggest that, unlike Wikipedia, *in OSM quality is independent of the number of user edits*. Although this result may seem surprising at first, it can actually be explained if we look at the fundamentally different nature of content in Wikipedia and OSM: articles in Wikipedia are much more complex than POIs in OSM. To be of high quality, the former must have, for example, a good structure and a logical flow, and it is thus expected that the more experienced a user is in editing Wikipedia’s articles, the higher the quality provided. The latter are very simple objects instead, with a rather limited number of editable features, each taking a very limited number of values. In this case, even a novice user can provide high quality information. This property is indeed highly desirable, as it means that *any citizen is a valuable contributor to maintain quality information about dynamic urban envi-*

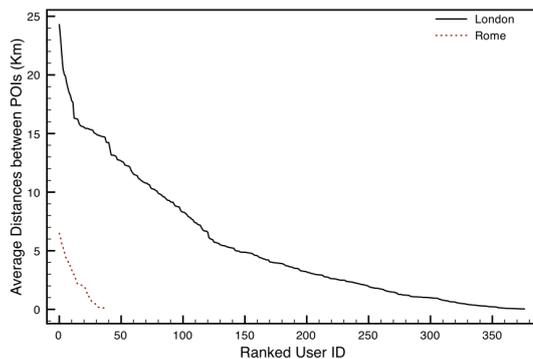


Figure 8: Local-Expert Distribution

ronments. In other words, urban crowd-sourcing does not rely on a small class of highly active and familiar users, but rather can take advantage of all citizens’ contributions to accurately map the urban world.

Seeing that quality is not related to the amount of activity of users in OSM, we have investigated whether other characteristics of users can be linked to quality instead. In particular, as content in urban crowd-sourcing is spatial in nature, one may expect the *local expertise* of editors to be directly linked to the quality of their edits. More precisely, we define ‘local experts’ those users who have edited a lot of POIs in a very confined geographic area; intuitively, such area may represent where they live or work, or in general an area where they spend a considerable amount of time, and of which they have very good knowledge. Figure 8 presents the distribution of the average distance between user edits, for those users who have edited at least 15 times⁸, with users ranked based on scale of contributions. We can see that in London there are some users who have edited POIs all over the city, with an average distance of up to 25 Km (the head of the distribution). However, the majority of the contributors only edited points less than 10 km apart, and the bottom 30% users (tail of the distribution) had restricted their contributions to 2 Km. The same distribution can be observed for Rome, but at a much smaller scale, corresponding to Rome being considerably smaller than London. Based on this new category of users (i.e., local experts), we then studied whether the quality of edits increases as does the local expertise of editors. Once again, our experimental analysis shows no statistical correlation between error and user activity (in terms of locality). This result reinforces the conclusion made before, that is, that *any citizen* (no matter whether a local expert or, for example, a tourist) is a valuable contributor to maintain high quality information about dynamic urban environments.

Property 2. We next examine Property 2 of Wikipedia, for further insights into who are the users that can provide higher quality content in urban crowd-sourcing settings. Recall that this property states that “the addition of a link, heading, or other structural element to an article tends to indicate higher quality editor”. In OSM, each POI has only

⁸Note that users who have edited only a handful of POIs are not considered local ‘experts’ according to our definition.

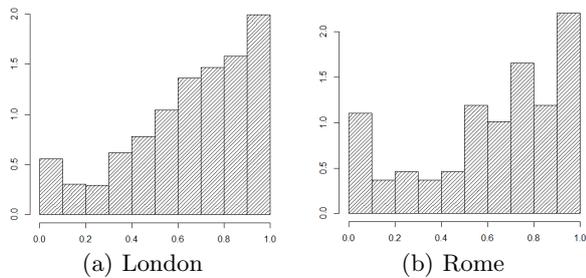


Figure 9: Normalised Density Distributions of Conscientiousness Degree

City	Geo_{Err}	$Lexical_{Err}$	$Amenity_{Err}$
London	-0.24	-0.14	-0.87
Rome	-0.39	-0.32	-0.89

Table 2: Pearson Correlation Coefficients r Between Conscientiousness Degree and Geo_{Err} , $Lexical_{Err}$ and $Amenity_{Err}$ (all results are statistically significant, with p -value < 0.01)

three structural elements: name, amenity and position. The above property can thus be paraphrased in OSM so to state that *conscientious* users, that is, users who provide complete information (i.e., both name and amenity) of the OSM POIs they edit, are also expected to be more accurate (they introduce lower errors). We have verified whether indeed this property holds in OSM; for the purpose of this analysis, we have defined the *conscientiousness degree* of a user as the portion of her edits which was complete (no missing values for name and amenity). Figure 9 shows the normalised density distributions of conscientiousness degree for both London and Rome users. Table 2 summarises the Pearson correlation coefficients r between conscientiousness degree and Geo_{Err} , $Lexical_{Err}$ and $Amenity_{Err}$. As shown, we found a negative correlation between conscientiousness degree of users and errors: the more conscientious a user is, the lower the error introduced by her edits⁹. This analysis reveals that in OSM *quality appears to come from users’ disposition* (e.g., how careful/picky they are). Importantly, *the majority of users under analysis have high values of conscientiousness degree* (see Figure 9), *thus suggesting that quality information is expected from most of OSM contributors*.

At this point, one may wonder whether conscientious users are always careful when editing lexicographic data, or whether they are so only when editing POIs in locations they most care about (e.g., the neighbourhood where they live or work). We do not have information about the ‘home’ (or ‘work’) location of OSM editors, but we have previously defined areas of local expertise for users, based on the geographic concentration of their contributions. We have then studied whether there exists a correlation between locality and conscientiousness, and found none: that is, OSM editors who carefully edit POIs do so no matter where the POIs are (i.e., they

⁹Note that the strong negative correlation between conscientiousness degree and amenity error can be explained by the fact that such degree measures the disposition of users to provide complete information (no empty values for name or amenity). This is in accordance with Figure 7 which shows that the main cause for amenity errors are empty values.

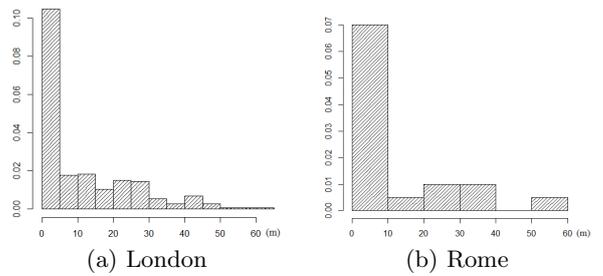


Figure 10: Normalised Density Distributions of Standard Deviation of Geo_{err}

care about factual accuracy of all urban information within their city, and not only some of its areas).

Property 3. The third property that has been observed in Wikipedia states that “users who previously contributed high (low) quality edits tend to continue to submit high (low) quality edits”. To verify whether this property holds in OSM, we have proceeded as follow: for each city under examination (London and Rome), we have defined users as ‘stable’ if they exhibited a consistent trend in their edits. A trend was defined consistent if the standard deviation of the error values associated to this user’s edits is low. Figures 10, 11, and 12 show the normalised density distributions of the standard deviation of errors for user edits, broken down per city and per error type. From the examination of these figures it is possible to see that there are very few ‘low stability’ users, that is, users either offer precise information about POIs or imprecise one, and they do so consistently over edits. This is valid for both geographic, lexicographic and amenity errors.

In a further analysis on *geographic* errors, we also found that, 72% of London editors offer consistently high quality information (low error); however, in the case of Rome, this is true for only 43% of users. Why are London editors consistently geographically accurate, while Rome editors are not? Although we cannot be certain of the reason, we may offer a plausible explanation: geographic accuracy is strongly related to the quality of the GPS sensor on board of the mobile device used for editing POIs; according to a comScore study¹⁰ conducted in 2010, the adoption rate of smart-phones in the UK was above 70%, while a meagre 11% in Italy. This would suggest that Londoners have higher-precision mobile devices from which to edit positioning information of POIs, and thus they are consistently offering higher quality geographic data, as opposed to their Italian counterpart. We thus hypothesise that *the characteristics of the input device have a strong impact on the quality of the edited information in urban crowd-sourcing settings*, whilst not the case in web-based crowd-sourcing.

We now focus our attention on *lexicographic* and *amenity* error results instead: this time, we observe a much stronger similarity between the two cities, with 67% and 62% of stable *and* accurate users, and a further 9% and 13% stable *and*

¹⁰<http://tinyurl.com/y1j9bvb>

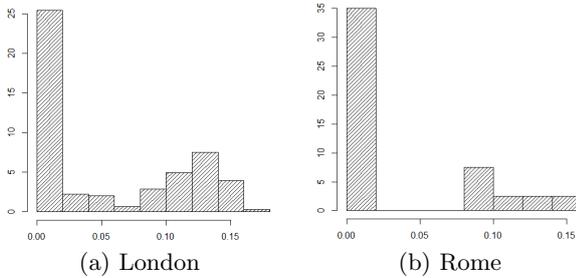


Figure 11: Normalised Density Distributions of Standard Deviation of $Lexical_{err}$

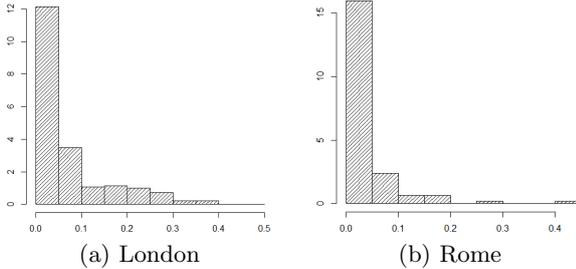


Figure 12: Normalised Density Distributions of Standard Deviation of $Amenity_{err}$

inaccurate users, in London and Rome respectively. Property 3 thus holds in OSM when we look at lexicographic and amenity error: in other words, if a user is careful to check that the input string correctly matches the POI name, very likely the same user will be still careful in the future, and this class of conscientious users represents the neat majority in both London and Rome, once again supporting the hypothesis of sustainability of using urban crowd-sourcing to maintain accurate urban data.

To conclude the investigation in this direction, we have examined if there is any correlation between stable user behaviour and quality. Table 3 shows the Pearson correlation between these two factors. Unsurprisingly, we found a positive correlation between ‘stable’ behaviour and quality: those people who offer high quality information also have a stable behaviour, so they very likely offer it consistently.

Property 4. Last but not least, we tested whether Property 4 of Wikipedia holds in OSM, that is, the higher the number of edits a POI has received, the higher its quality. We found that accuracy-related errors are independent of the number of edits per POI. We can explain this result with a reasoning that is similar to that presented when explaining why Property 1 does not hold either: OSM objects are very simple compared to Wikipedia articles. While a Wikipedia article can almost always be improved and expanded (i.e., the higher the number of edits, the higher the expected quality), a POI can be fully described (name, amenity type, position) in just one edit. If such information is entered correctly in the first place (as Properties 2 and 3 suggest), then there is no need for further updates, as these would not

improve the POI quality. This is a very attractive property: while Wikipedia articles must usually undergo several cycles of edits and improvements, before they reach a high enough quality to be useful, *the quality of information pertaining POIs is very high from the beginning*, and thus can be immediately leveraged upon, for example, by location-based services.

6. DISCUSSION

Is crowd-sourcing a viable way of maintaining quality information about urban settings? Based on the results presented before, the answer we offer is *yes*, for the following reasons: as observed in web-based crowd-sourcing, contributors in urban settings follow a power-law distribution, with a few users editing a lot, whilst the vast majority of users offering very few edits instead. However, while in Wikipedia quality information mostly comes from this very restricted set of so called power-users only (Property 1), in OSM the pool of quality contributors is much broader, as even novice users can offer the same level of quality of heavy editors. Indeed, in OSM, quality is more related to the natural disposition of the editors (their conscientiousness – Property 2), and in the two datasets analysed (London and Rome), the majority of users do fall into this category. Furthermore, such disposition is persistent (Property 3), so that quality remains consistently high in all edits performed by such users. Finally, as new urban elements are represented digitally, they can be immediately used and relied upon by users and applications, as the quality of such information is high from the very beginning (Property 4), a property that is particularly appealing in the case of dynamic urban settings.

7. CONCLUSION AND FUTURE WORK

In this paper, we have investigated crowd-sourcing as a means of maintaining accurate information about POIs in dynamic urban environments. We have done so for one example of urban crowd-sourcing dataset, that is, OpenStreetMap. First, our investigation has shown that POI data in OSM is highly accurate, where accuracy has been defined in terms of geographic (positioning) error, lexicographic error, and amenity error. Second, we have identified the factors (in relation to both users and content) that contribute to this high quality, and in so doing highlighted important differences with respect to web-based crowd-sourcing (and Wikipedia in particular): first, the *quality of contributors* in OSM is independent of the number of offered contributions; second, the *quality of contributions* in OSM is independent of the number of edits/revisions they have undergone. These two observations, coupled with the observation that the majority of OSM editors are *consistently conscientious* editors (and thus providers of quality information) does suggest of high potentials that urban crowd-sourcing has for maintaining accurate data about urban settings.

We plan to take this study further by moving our focus from accuracy to coverage, and investigate the socio-cultural factors that have an impact on the coverage of urban crowd-sourcing. In particular, we are looking into parameters such as wealth/poverty of an area.

Acknowledgment The research leading to these results has received funding from the European Community’s Seventh

City		avg(GeoErr)	avg(LexicalErr)	avg(AmenityErr)
London	std(GeoErr)	0.30	0.04	0.05
	std(LexicalErr)	-0.01	0.06	-0.01
	std(AmenityErr)	-0.01	-0.06	0.38
Rome	std(GeoErr)	0.28	0.21	0.07
	std(LexicalErr)	0.04	0.25	0.02
	std(AmenityErr)	0.03	0.05	0.45

Table 3: Pearson Correlation Coefficients r Between Stability and Errors (in bold are the statistically significant — p -value < 0.01)

Framework Programme (FP7/2007-2013) under the Grant Agreement n. 234239.

8. REFERENCES

- [1] B. Adler and L. De Alfaro. A content-driven reputation system for the wikipedia. In *Proceedings of the International Conference on World Wide Web*, pages 261–270, 2006.
- [2] L. Bettencourt and G. West. A unified theory of urban living. *Nature*, 467(7318):912–913, 2010.
- [3] T. Chesney. An empirical examination of wikipedia’s credibility. *Journal of First Monday*, 11(11), 2006.
- [4] G. Druck, G. Miklau, and A. McCallum. Learning to predict the quality of contributions to wikipedia. *Journal of WikiAI*, 8:7–12, 2008.
- [5] J. Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–901, 2005.
- [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. *Cartographic Journal, The*, 47(4):315–322, 2010.
- [10] A. Halfaker, A. Kittur, R. Kraut, and J. Riedl. A jury of your peers: quality, experience and ownership in wikipedia. In *Proceedings of the Interational Symposium on Wikis and Open Collaboration*, pages 1–10, 2009.
- [11] M. Hu, E. Lim, A. Sun, H. Lauw, and B. Vuong. Measuring article quality in wikipedia: models and evaluation. In *Proceedings of the conference on information and knowledge management*, pages 243–252, 2007.
- [12] 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.
- [13] J. Nielsen. Participation inequality: Encouraging more users to contribute. *Alertbox*, 2006.
- [14] F. Ortega and J. Gonzalez Barahona. Quantitative analysis of the wikipedia community of users. In *Proceedings of the International Symposium on Wikis*, pages 75–86, 2007.
- [15] K. Panciera, A. Halfaker, and L. Terveen. Wikipedians are born, not made: a study of power editors on wikipedia. In *Proceedings of the International Conference on Supporting Group Work*, pages 51–60, 2009.
- [16] K. Panciera, R. Priedhorsky, T. Erickson, and L. Terveen. Lurking? Cyclopaths?: A Quantitative Lifecycle Analysis of User Behavior in a Geowiki. In *Proceedings of Human Factors in Computing Systems*, pages 1917–1926, 2010.
- [17] R. Priedhorsky, J. Chen, S. Lam, K. Panciera, L. Terveen, and J. Riedl. Creating, destroying, and restoring value in wikipedia. In *Proceedings of the International Conference on Supporting group work*, pages 259–268, 2007.
- [18] D. Riehle. How and why wikipedia works: an interview with Angela Beesley, Elisabeth Bauer, and Kizu Naoko. In *Proceedings of the Interational Symposium on Wikis and Open Collaboration*, pages 3–8, 2006.
- [19] B. Stvilia, M. Twidale, L. Smith, and L. Gasser. Information quality work organization in wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6):983–1001, 2008.
- [20] F. Viégas, M. Wattenberg, and K. Dave. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the International Conference on Human factors in computing systems*, pages 575–582, 2004.
- [21] J. Voss. Measuring wikipedia. In *Proceedings of the 10th International Conference of the International Society for Scientometrics and Informetrics*, pages 24–28, 2005.
- [22] H. Welsler, D. Cosley, G. Kossinets, A. Lin, F. Dokshin, G. Gay, and M. Smith. Finding social roles in wikipedia. In *In Proceedings of the 2011 iConference*, pages 122–129, 2011.
- [23] D. Wilkinson and B. Huberman. Assessing the value of cooperation in wikipedia. *First Monday*, 12(4), 2007.
- [24] D. Wilkinson and B. Huberman. Cooperation and quality in wikipedia. In *Proceedings of the International Symposium on Wikis*, pages 157–164, 2007.
- [25] D. Zielstra and A. Zipf. A comparative study of proprietary geodata and volunteered geographic information for germany. In *Proceedings of the 13th International Conference on Geographic Information Science*, 2010.