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.jahanbakhshmashhadi@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

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

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 challenged.

1. 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
Geographical information is fundamentally different from web-based information, both because of its temporal dynamicity (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; reporters, using micro-blogging services such as Twitter; and cartographers, with geo-wikis like Cyclopath and OpenStreetMap. 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 these 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 some 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. Volunteer 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 cyclists 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. Girges 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 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 encyclopedia [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](image)

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.

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 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.
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

4. URBAN CROWD-SOURCING DATASET

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.

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.

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.

![Distribution of edit counts for articles 240 weeks old](image)
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.

<table>
<thead>
<tr>
<th>City</th>
<th># Users</th>
<th># POIs</th>
<th># Edits</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>2,232</td>
<td>32,438</td>
<td>44,463</td>
</tr>
<tr>
<td>Rome</td>
<td>363</td>
<td>3,315</td>
<td>5,319</td>
</tr>
</tbody>
</table>

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.

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.

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: \(\text{poi} = (\text{name}, \text{amenity}, (\text{lat}, \text{lon}))\), where \text{name} is the POI’s name, \text{amenity} is its category (e.g., cafe, restaurant), and \((\text{lat}, \text{lon})\) are the coordinates defining its geographical position\(^5\). 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 \(\text{poi}_x\) be a single POI, and \(\text{POI}_x\) 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 \(\text{poi}_{osm}\) is equivalent to \(\text{poi}_{gt}\) (\(\text{poi}_{osm} \equiv \text{poi}_{gt}\)) if both their geographic and lexicographic differences are small. The geographic difference \(\text{geo}_{Err}(\cdot, \cdot)\) 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 \(\text{lexical}_{Err}(\cdot, \cdot)\) 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 0.35 \(^6\). Having defined these distance metrics, we can now quantify errors. Let \(\text{POI}_{osm} \subseteq \text{POI}_{osm}\) and \(\text{POI}_{gt} \subseteq \text{POI}_{gt}\) 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:

\[
\text{Geo}_{Err} = \text{AVG}(\text{geo}_{Err}(\text{poi}_{osm}, \text{poi}_{gt}))
\]
\[
\text{Lexical}_{Err} = \text{AVG}(\text{lexical}_{Err}(\text{poi}_{osm}, \text{poi}_{gt}))
\]
\[
\text{Amenity}_{Err} = \text{AVG}(\text{amenity}_{Err}(\text{poi}_{osm}, \text{poi}_{gt}))
\]

where: \(\text{Amenity}_{Err}(\cdot, \cdot)\) is defined equal to 0 if the two input POIs have the very same amenity type classification, 1 otherwise\(^7\); the average function \(\text{AVG}\) is computed on all \(\text{poi}_{osm} \equiv \text{poi}_{gt}, \text{poi}_{osm} \in \text{POI}_{osm}^\text{eq}, \text{poi}_{gt} \in \text{POI}_{gt}^\text{eq}\). The lower \(\text{Geo}_{Err}, \text{Lexical}_{Err},\) and \(\text{Amenity}_{Err}\) 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

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.

\(^5\) Note 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.

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

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

<table>
<thead>
<tr>
<th>City</th>
<th>GeoErr</th>
<th>LexicalErr</th>
<th>AmenityErr</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>24 m</td>
<td>0.13</td>
<td>0.36</td>
</tr>
<tr>
<td>Rome</td>
<td>34 m</td>
<td>0.07</td>
<td>0.45</td>
</tr>
</tbody>
</table>

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% 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.
tion in dynamic urban settings. Furthermore, this result has been observed in a system that is completely self-policied: 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 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 quality. In particular, we aim to test whether the same 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 environments.

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 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 three structural elements: name, amenity and position. The above property can thus be paraphrased in OSM so 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

Note that users who have edited only a handful of POIs are not considered local ‘experts’ according to our definition.
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 GeoErr, LexicalErr, and AmenityErr (all results are statistically significant, with $p$-value < 0.01).

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 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 study\(^9\) conducted in 2010, the adoption rate of smartphones 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 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.

\(^9\)Note 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.

\(^{10}\)http://tinyurl.com/y1j98vb

<table>
<thead>
<tr>
<th>City</th>
<th>GeoErr</th>
<th>LexicalErr</th>
<th>AmenityErr</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>-0.24</td>
<td>-0.14</td>
<td>-0.87</td>
</tr>
<tr>
<td>Rome</td>
<td>-0.39</td>
<td>-0.32</td>
<td>-0.89</td>
</tr>
</tbody>
</table>

Table 2: Pearson Correlation Coefficients $r$ Between Conscientiousness Degree and GeoErr, LexicalErr, and AmenityErr (all results are statistically significant, with $p$-value < 0.01).
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 as 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 Framework Programme (FP7/2007-2013) under the Grant Agreement n. 234239.

8. REFERENCES


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

<table>
<thead>
<tr>
<th>City</th>
<th>std($GeoErr$)</th>
<th>std($LexicalErr$)</th>
<th>std($AmenityErr$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>London</td>
<td>0.30</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>0.06</td>
<td>-0.01</td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>-0.06</td>
<td>0.38</td>
</tr>
<tr>
<td>Rome</td>
<td>0.28</td>
<td>0.21</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.25</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.19</td>
<td>0.45</td>
</tr>
</tbody>
</table>