There's No Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-sourcing Datasets

Giovanni Quattrone
University College London
London, UK
g.quattrone@cs.ucl.ac.uk

Licia Capra
University College London
London, UK
l.capra@cs.ucl.ac.uk

Pasquale De Meo
University of Messina
Messina, Italy
demeo@unime.it

ABSTRACT
Crowd-sourcing has become a popular form of computer mediated collaborative work and OpenStreetMap represents one of the most successful crowd-sourcing systems, where the goal of building and maintaining an accurate global map of the world is being accomplished by means of contributions made by over 1.2M citizens. However, within this apparently large crowd, a tiny group of highly active users is responsible for the mapping of almost all the content. One may thus wonder to what extent the information being mapped is biased towards the interests and agenda of this group of users. In this paper, we present a method to quantitatively measure content bias in crowd-sourced geographic information. We then apply the method to quantify content bias across a three-year period of OpenStreetMap mapping in 40 countries. We find almost no content bias in terms of what is being mapped, but significant geographic bias; furthermore, we find that bias in terms of meticulousness varies with culture.

Author Keywords
Content bias, OpenStreetMap, crowd-sourcing, volunteered geographic information, cross-cultural

ACM Classification Keywords
H.2.8 Database Management: Database Applications—*Spatial Databases and GIS*; H.5.3 Group and Organization Interfaces: Computer-Supported Cooperative Work

INTRODUCTION
Crowd-sourcing is a form of online collaborative work practice, where a task that was traditionally performed by a selected number of skilled employees is now being performed by a large, self-selected group of volunteers in answer to an open call [21]. Crowd-sourcing has been used to accomplish a variety of tasks [3]: from distributed human intelligence tasking, as exemplified by the open calls usually appearing on Amazon Mechanical Turk, to scientific (objective) problem solving; from peer-vetted (subjective) creative production, as in Threadless,[1] to knowledge discovery and management. It is the latter type of crowd-sourcing tasks we are interested in this paper, whereby a crowd is mobilised to collect and maintain large repositories of information. The most successful example of this is Wikipedia, with its online community of editors voluntarily contributing to a repository of the whole body of knowledge. A specific type of knowledge where crowd-sourcing has also been widely applied is that of volunteered geographic information, where citizens have become surveyors, with council-monitoring applications like FixMyStreet;[2] local reporters, as powered by Ushaidi's Crowdmap;[3] and cartographers, with geo-wikis like Cyclopath[4] and OpenStreetMap.[5]

Crowd-sourcing platforms for knowledge discovery and management have been subject to scrutiny by the research community over the years; one of the first issues raised was that of information accuracy, as we take a task away from skilled employees and assign it to an undefined, self-selected crowd. Several studies have revealed that the emerging crowd-editing practises result in high quality data gathering overall (e.g., [23, 18, 15, 13, 28]). However, information accuracy is not the only concern that emerges as a result of this knowledge production paradigm shift; another important issue is that of *content bias*.

Although crowd-sourcing seeks contributions from potentially anyone, in practice previous work has shown that only a tiny group of contributors (often referred to as 'power users') produces most of the content [22, 25, 32, 33, 35, 43]. For this reason, one may wonder to what extent the gathered information is biased towards the interests of this small group of highly active contributors. In this paper, we investigate this question for crowd-sourced geographic information repositories – and OpenStreetMap in particular – by quantitatively measuring the extent to which the mapping of 'power users' is *mis*aligned with respect to what done by the whole OpenStreetMap contributors' base.

If such misalignment is high, the rationale behind the very existence of OpenStreetMap might be at stake. At a time when

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CSCW 2015, March 14–18, 2015, Vancouver, BC, Canada.
Copyright © 2015 ACM 978-1-4503-2922-4/15/03...$15.00.
http://dx.doi.org/10.1145/2675133.2675235

[1] `https://www.threadless.com/`
[2] `http://www.fixmystreet.com/`
[3] `http://www.ushahidi.com/products/crowdmap`
[4] `http://cyclopath.org/`
[5] `http://www.openstreetmap.org/`

maps were in the hands of a few selected companies, and thus their content inevitably reflected private agendas (e.g., of the military, as in the case of Ordnance Survey; of commercial profit, as in Google Maps and Bing Maps), OpenStreetMap offered a radically different approach, empowering anyone to freely edit its content. As Mikel Maron, Director of the GroundTruth Initiative[6] and Board Member of OpenStreetMap Foundation states, *"We really think it's important that the people who are creating the map data, or are involved with creating the map data, are the people who are living there, and the data that's created from their communities is something they have a stake in and is something that they can fully use to improve their own situation and advantage the agenda of the community"*.[7] Although the crowd-sourcing paradigm enables this goal *in principle*, actual adoption of the paradigm sees a rather small (and also demographically-biased) community core being responsible for the vast majority of the OpenStreetMap knowledge base, thus questioning whether the map that emerges in OpenStreetMap is biased towards the interests of its 'power users'. With a growing number of companies and non-for-profit organisations worldwide relying on OpenStreetMap to deliver their own location-based services,[8] it has become important to provide a method to quantitatively answer this question.

In this paper, we propose a method to quantify *content bias*, to be interpreted here as the extent to which the interests of the power users are *mis*aligned with respect to those of the rest of the OpenStreetMap contributors' base. We consider users' *interests* as captured by what OpenStreetMap users edit in the map; we subsequently measure bias in terms of three signals that mappers implicitly leave: *what* information is being mapped, *where* the information is being mapped, and finally *how* meticulous is the information being edited. For each of these dimensions, we virtually compare two maps: one that emerges from the actions of *power users*, and one that emerges from the actions of the very long tail of *occasional contributors*. We apply the method to OpenStreetMap, as this is the most popular crowd-sourced geographic information system in existence to date. In this context, we quantify bias for 40 countries around the world, examining all contributions made in a period of three years (from 01/2010 to 12/2012).

Our findings reveal that there is almost no bias in terms of what content is being produced: the type of spatial objects that power users edit is very well aligned with the information being mapped by the very many occasional mappers. On the contrary, there is higher bias in terms of where information is being mapped: the map that emerges from power contributors is very dense around big cities and near empty anywhere else, whereas the map that emerges from occasional contributors offers a more uniform spatial coverage, albeit with lower density. Finally, we found bias in how meticulously the map is being edited to vary with culture: where power distance is low and individualism is high, the crowd performs a more accurate work than power users, and viceversa.

The remainder of the paper is structured as follows: we first position this work in the context of related literature; we then describe the dataset used in this study, and provide a definition of bias within this context. We present the method we developed to quantitatively measure the previously defined forms of bias, and analyse the results obtained when applying the method to the dataset under exam. Finally, we summarise the main contributions of this work, we reflect on the theoretical and practical implications of the proposed method, we acknowledge current limitations, and elaborate on future directions of research.

RELATED WORK

The topic of bias, interpreted as lack of neutrality, has been extensively studied in collaborative knowledge production platforms such as Wikipedia. Indeed, although Wikipedia advocates a strict 'neutral point of view' (NPOV) policy, many studies that looked into measuring the extent to which this policy is adhered to in practice have found otherwise. For example, [5, 34] examined the extent to which content and perspectives vary across *cultures*, by comparing articles about famous people in the Polish and English editions of Wikipedia, and found substantial differences between the two monolingual versions. Lam et al. [27] have studied *gender* imbalance in Wikipedia, and reported on how in some domains, such as movies, articles of particular interest to females were substantially less covered than articles of specific interest to males – a direct consequence of having very few female editors relative to male ones. The same issue of gender gap also received media attention; for example, in 2011 The New York Times reported that, according to many surveys, less than 15 percent of Wikipedia contributors are women [8].

Other studies of Wikipedia have looked into the effect that varying levels of *user engagement* have on the resulting knowledge base. Wikipedia, as many online user-generated content platforms, exhibits a highly skewed user base, with a small fraction of less than 20% of users (often called *power users*) being responsible for the creation of 80% or more of available content [33, 25, 35]. A study by Kittur et al. [25] suggests that Wikipedia was indeed initially driven by the 'power of the few' top contributors, rather than being the result of an actual 'wisdom of the crowd'; this has however been shifting over time, with a larger variety of users committing contributions in recent years. Note that varying levels of engagement do not necessarily represent a problem: as found in previous research, the top Wikipedia editors are indeed those who use the quality assurance systems [17], those who invoke community norms, and those who revise available entries [33], making them both more accurate and complete, so that their work has a positive effect, both directly (in terms of quality of the content) and indirectly (in terms of community growth).

[6] `http://groundtruth.in/` – A new media and technology consulting company specialising in community-based participatory technologies, especially mapping and citizen journalism, in poor and marginalized regions throughout the world.

[7] `http://www.wired.co.uk/news/archive/2013-08/05/slum-mapping-google-maps-cartography`

[8] `http://wiki.openstreetmap.org/wiki/List_of_OSM_based_Services`

As we shift our attention to *geographic* knowledge repositories, new and different forms of bias may emerge, as there is an intrinsic relationship between the location of a contributor and the type of knowledge they can offer (whereas, in general-knowledge repositories such as Wikipedia, it is often the case that a user can contribute to any article, regardless of where they are geographically based). Few studies exist in this domain, with a notable exception being the recent work by Hetch and Stephens [19]: they examined three of the most common sources of volunteered geographic information (VGI [14]) - that is, Twitter, Flickr, and Foursquare - and showed that all of them are strongly biased towards urban areas, at the expense of rural ones (for instance, they found that there are 24.4 times more Foursquare users per capita in urban areas than rural ones).

OpenStreetMap is probably the VGI repository most studied to date, but research has so far focused on two other questions: accuracy and coverage. The former has been found to be very high [1, 10, 13, 14, 15, 16, 28] when compared to proprietary mapping datasets, in a variety of countries across Europe. The latter has been found to sharply decrease as we move away from city centres [44], or when we move in areas with low socio-economic status [15]; for this reason, some authors have labelled user-generated geo-tagged information as geographically biased, in the sense that it does not cover with equal interest urban and rural areas [24], and neither well-off and deprived areas [15].

Geographic bias is certainly important, but it is not the only form of bias worth investigating in geographic knowledge repositories. For example, Stephens [42] found that user adoption of volunteered geographic information in OpenStreetMap and Google MapMaker is not uniform, with an abundance of men compared to women. Furthermore, similar to what observed in Wikipedia, many geo-spatial communities manifest a highly skewed user base, with a tiny fraction of users being responsible for the majority of available content [32, 22, 43]. There is thus an open opportunity here to go beyond the geographic bias, and to explore the extent to which the mapped content is representative of the interests of small groups of users only. In this paper, we propose a method to quantify different forms of content bias in geographic crowd-sourced repositories, and apply it to the case of OpenStreetMap in particular, to measure how biased the mapped information is towards the interests of a minority of very active users. In the next section, we describe the dataset under exam, and the forms of bias our method aims to measure.

We note that different forms of bias have been studied in other domains. For example, bias intended as the expression of the personal views of the editors of a community has been investigated in Foursquare [39], Pinterest [7], YouTube [6] and Yelp [9]. However, these platforms specifically call for the personal expressions of individuals' experiences. As such, their goal is fundamentally different from that of knowledge gathering platforms like Wikipedia and OpenStreetMap, that aim for a certain ideal of neutrality and objectivity.

DATASET

Before delving into a detailed description of our proposed method, we describe the crowd-sourcing datasets to which it is best applied. First, the crowd-sourced content should have a geographic component; second, the content should be semantically varied and rich, so that different geographical objects can be mapped, and with a different degree of semantic richness; finally, different users' groups should be contributing to the crowd-sourcing task, so that their actions can be compared. Many crowd-sourcing datasets meet these criteria, including WikiMapia,[9] an open-content collaborative mapping project where volunteer users mark and describe all geographical objects on Earth; Google Map Maker,[10] a service launched by Google in June 2008 designed to allow Google Map users to update and improve the map of places that matter to them; and OpenStreetMap,[11] a collaborative project to create a free editable map of the world. In this paper, we chose to apply our method to OpenStreetMap (hereafter OSM), as of all suitable geographic crowd-sourcing projects and associated datasets, this is probably the most successful one, having been running since 2004, and comprising the largest (and most geographically widespread) user and content base. Furthermore, OSM is the one that has been researched the most in the literature, so that we can relate our findings to previous studies.

The OSM dataset is freely available to download[12] and it contains the history (since 2006) of all edits (over 2 billions) performed by all users (over 1.6M) on all spatial objects. In OSM jargon, spatial objects can be one of three types: *nodes*, *ways*, and *relations*. Nodes are single geospatial points and typically represent Points-of-Interest (POIs); ways mostly represent roads (as well as streams, railway lines, and the like); finally, relations are used for grouping other objects together, based on logical (and usually local) relationships (e.g., administrative boundaries, bus routes).

In order to reduce the OSM dataset to a more manageable size, we restricted our attention to nodes that describe Points-of-Interest (POI) only. A node consists of three main attributes: a geographical position (latitude and longitude), a name, and an amenity type used to describe the category the POI belongs to; examples of categories include: 'hospital', 'place of worship', 'restaurant', and 'school'. The amenity type field is optional, and although OSM guidelines suggest a common taxonomy to use,[13] in practice users are free to use whatever vocabulary their prefer. We further restricted our attention geographically; previous research has shown that OSM presents large differences in data contributions and community development in different areas of the world [31, 37]. To capture bias in these differing contexts, we adopted a stratified sampling approach, and reduced our analysis to 40 (out of 133) countries that have OSM presence, so to in-

[9] `http://en.wikipedia.org/wiki/WikiMapia`

[10] `http://en.wikipedia.org/wiki/Google_Map_Maker`

[11] `http://www.openstreetmap.org/`

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

[13] `http://wiki.openstreetmap.org/wiki/Map_Features#Amenity`

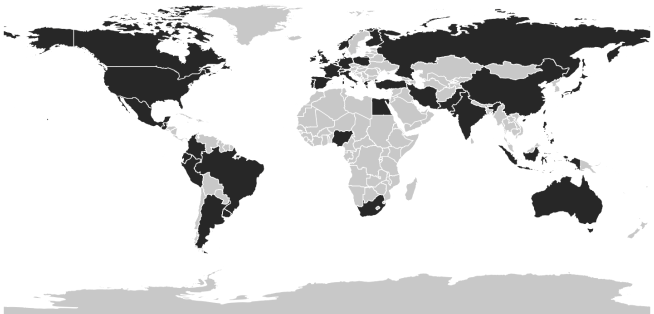

Figure 1. Map of final 40 countries under analysis (in alphabetical order: Argentina, Australia, Austria, Belgium, Brazil, Canada, China, Colombia, Denmark, Ecuador, Egypt, Finland, France, Germany, Greece, Guatemala, India, Indonesia, Iran, Ireland, Italy, Japan, Mexico, Netherlands, Nigeria, Norway, Pakistan, Peru, Philippines, Poland, Portugal, Russia, South Africa, Spain, Switzerland, Taiwan, Turkey, United Kingdom, United States, Uruguay)

	Min	*1st Qu.*	*Median*	*3rd Qu.*	*Max.*
OSM nodes	4,541	72,500	139,800	357,600	3,417,700
OSM users	161	1068	1,776	3,876	44,460
OSM nodes per km^2	0.01	0.04	0.35	1.55	9.57
OSM users per population	9.2e-07	1.3e-05	6.7e-05	2.1e-04	8.2e-04

Table 1. Summary Statistics of OSM Features in the 40 Analysed Countries

clude countries representative of different cultural values and economic status, as these will be used later in the study.

Two further pruning steps have been conducted: within the selected 40 countries, we chose to discard the very first years of OSM mapping, as only a few early-adopters were contributing at that point; the dataset would thus not have allowed us to quantify crowd (i.e., occasional workers) behaviour at that time. From 2010 onwards, all 40 countries under exam appear to have passed the cold-start phase, as an analysis of OSM users and activity showed.

Finally, from the above sample, we removed all contributions that appeared to be 'bulk imports',[14] as these are not representatives of 'human' crowd-sourcing activity, but rather of 'data donations' given to the wider OSM community by external organisations. Imports are not explicitly marked as such in the OSM dataset; we thus needed heuristics to identify them. We marked as imports those edits which came from a single user, in very large quantities (i.e., more than one thousand of edits), in a short period of time (i.e., less than one hour), and that were spread over a large geographic area (i.e., a whole city), using the same approach used in [37].

In the end, we are left with a dataset comprising nearly 120k users, who have collectively made over 13M contributions to OSM, across the 40 different countries that are highlighted in Figure 1. Summary features of this dataset are showed in Table 1.

METHOD

To understand whether the OSM map is biased towards the interests of the very active minority of 'power users', we proceed as follow. We first distinguish power users from rest of occasional mappers (which we will refer to as 'the crowd'); then we separate the OSM map in two virtual maps, one derived from the edits of power users and one derived from the edits of the the crowd; subsequently, we define which form of content bias we are interested to capture; finally, we define metrics to capture the content (mis)alignments of the virtual maps generated by power users and the crowd. We will discuss these three steps next.

Power Users vs. The Crowd

The term 'power user' is used informally in online collaborative-work platforms to indicate the top contributors of the community. Many authors refer to the 80-20 rule-of-thumb (or Pareto principle) and define 'power users' those users (typically 20% or less of the whole user base) that collectively produce more than 80% of the content [4, 33, 25]. We use the same definition in this work and, for each country under exam, we consider as power users the minimum number of users required to account for 80% of the edits in that country. Interestingly, we found that these users are less than 10% of the OSM population for each country.

One may wonder to what extent the above definition of power users impacts results, and what the effect of using different thresholds would be. To investigate this, in each country under exam, we first divided OSM users into five groups (rather than just power vs. crowd), each with an exponentially increasing number of OSM contributions (from $< 10^1$ to $> 10^4$). For each group and for each country, we then computed all our bias metrics and found that, in the vast majority of cases, results for the three least active groups, and results for the two most active groups, were very similar. We found also that the activity threshold separating these two groups is, in most cases, the same emerging when distinguishing power users vs. the crowd following the 80-20 rule of thumb. We thus chose to present results comparing the two broad classes of power users and the crowd only.

Forms of Content Bias

In this work, we are interested in quantitatively measuring the extent to which the interests of the power users and the interests of the crowd are *mis*aligned. We define users' *interests* of both power users and the crowd as captured by what they edit in their own virtual maps (note that lurkers – individuals who use OSM without contributing any content – are not considered in this study, as they do not leave any quantifiable trace about their interests).

We quantify *mis*alignment according to three dimensions: *(i) what* spatial objects are being edited, *(ii) where* these objects are physically located, and *(iii) how* meticulously they are edited.

- *What is being mapped?* The fist bias dimension we measure is the extent to which the POI amenity types mapped by power users are (mis)aligned with those mapped by the rest of the OSM community. Previous qualitative studies of OSM have revealed that power editors are a niche group prevalently formed by educated, young, white male, and that evidence of this niche group dominance in OSM is also reflected in the amenities that have been mapped [42].

[14] `http://wiki.openstreetmap.org/wiki/Import`

We thus aim to measure whether the POI amenity types that this group edits are of interest to them only, or whether the same amenity types are also being edited by the much larger crowd of occasional mappers.

- *Where is mapping taking place?* The second bias dimension we measure is the extent to which the geographic areas mapped by power users are (mis)aligned with respect to those mapped by the rest of OSM community. As power users exhibit rather peculiar demographics [42], one may wonder whether the map that emerges from their edits predominantly covers areas of interest to them (e.g., places where they live, work and mingle), and whether these areas do not align with those of interest to the much larger OSM crowd, thus leaving big holes in the emerging map.
- *How meticulous is the mapping?* The third and last bias dimension we measure is the level of accuracy that emerges from the two virtual maps. We do so as past work on Wikipedia has revealed that power editors are responsible for the high quality of Wikipedia articles, being engaged in moderating and quality control activities, such as enforcing community norms and reverting edits when they do not follow Wikipedia's policies [33]. If a similar separation of roles emerges in OSM too, then we would expect the virtual map created by power editors to signal higher accuracy than the one emerging from crowd mappers.

Metrics

Having distinguished power users from the rest of the crowd, and having defined the three forms of content bias that we want to capture, we can proceed to define metrics to measure them.

What is being edited

The first metric we define aims to measure to what extent the amenity types edited by power users are also being edited by the crowd. To do so, for each country, we compute two vectors, $\vec{a_p}$ and $\vec{a_c}$, where the i-th entry of $\vec{a_p}$ (resp., $\vec{a_c}$) indicates how many times the i-th amenity is mapped by power users (resp., the crowd). We then define our first metric *Amenity Misalignment* (AM) as:

$$AM = 1 - \theta(\vec{a_p}, \vec{a_c}) \tag{1}$$

where $\theta(\cdot, \cdot)$ is the Cosine similarity between the two vectors.[15] Since $\vec{a_p}$ and $\vec{a_p}$ cannot have negative entries, then their Cosine similarity ranges from 0 to 1 and so also AM does. Values of AM close to zero (i.e., $\theta(\vec{a_p}, \vec{a_c}) \approx 1$) indicate that power users and the crowd edit the same amenities (even if at different scale in terms of absolute number of edits), and we interpret this as an indication of very low misalignment; viceversa, values of AM close to the unit (i.e., $\theta(\vec{a_p}, \vec{a_c}) \approx 0$) signal that power users and the crowd have interest in completely different amenities (i.e., high misalignment).

[15]The Cosine similarity is a well known vector similarity measure and it is mostly used in high-dimensional positive spaces such as in Information Retrieval and Text Mining. It computes the cosine of the angle between two vectors [41], so it assumes values belonging to the real interval $[-1, 1]$ and is equal to the unit when the two vectors are perfectly aligned.

The AM metric alone is not sufficient to understand what kind of misalignment is taking place. For example, AM could be high because power users and the crowd have mapped completely different amenities; however, it could also be high because, for example, power users have mapped very few amenity types, while the crowd has mapped the same types plus many more. We thus consider, in conjunction with AM, a further metric that we call *Amenity Diversity Misalignment*, which aims to capture how *diverse* the range of amenities mapped by power users is vs. the range of those mapped by the crowd. Formally, we begin with the two vectors $\vec{a_p}$ and $\vec{a_c}$ as before; we then measure the diversity of amenities produced by power users (resp., the crowd) by computing the Shannon Diversity Index[16] on $\vec{a_p}$ (resp., $\vec{a_c}$). We refer to these quantities as $AmnDiv_p$ and $AmnDiv_c$. We then define our second metric *Amenity Diversity Misalignment* (ADM) as the relative difference between $AmnDiv_p$ and $AmnDiv_c$:

$$ADM = \frac{AmnDiv_p - AmnDiv_c}{\max(AmnDiv_p, AmnDiv_c)} \tag{2}$$

Because both $AmnDiv_p$ and $AmnDiv_c$ are equal to or greater than 0, then $ADM \in [-1, +1]$. Positive values of this metric mean that the virtual map generated by power users shows a broader set of amenities than the virtual map generated by the crowd; negative values point out the opposite behaviour; values closed to 0 mean that the virtual map generated by power users and the virtual map generated by the crowd show no misalignment in term of diversity of mapped amenities.

Where is mapping taking place

To quantify spatial misalignment, the first metric we define looks at the difference between the geographic areas mapped by power users and those mapped by the crowd. We proceeded as follow: we subdivided each country in grids of different sizes (i.e., 2km $\times$ 2km, 5km $\times$ 5km, 15km $\times$ 15km). We then computed two vectors, $\vec{g_p}$ and $\vec{g_c}$, where the i-th entry of $\vec{g_p}$ (resp., $\vec{g_c}$) indicates how many times the i-th cell of the grid has been mapped by power users (resp., the crowd). Similarly to what was done when defining Amenity Misalignment in Equation 1, we now define *Spatial Misalignment* (SM), as:

$$SM = 1 - \theta(\vec{g_p}, \vec{g_c}) \tag{3}$$

In the next section, we will report only results obtained when considering the 2km $\times$ 2km grid, as very similar results were obtained when considering the other grid sizes.

Similarly to what was done when defining Amenity Diversity Misalignment, we consider, in conjunction with SM, a further metric that we call *Spatial Diversity Misalignment*, which aims to capture how *diverse* the range of areas mapped by power users is vs. the range of those mapped by the crowd. As before, we begin with the two vectors $\vec{g_p}$ and $\vec{g_c}$; we then measure the diversity of areas mapped by power users (resp., the crowd) by computing the Shannon Diversity Index on

[16]The Shannon diversity index is a measure that reflects how many different entries there are in a data set and the value is maximised when all entries are equally high [40].

$\vec{g_p}$ (resp., $\vec{g_c}$). We refer to these quantities as $SpDiv_p$ and $SpDiv_c$. We then define the *Spatial Diversity Misalignment* (SDM) metric as the relative difference between them:

$$SDM = \frac{SpDiv_p - SpDiv_c}{\max\left(SpDiv_p, SpDiv_c\right)} \tag{4}$$

As for the SM metric, we will report results obtained on the 2km $\times$ 2km grid only, as these are consistent with those obtained when operating with other grid sizes.

How meticulous is the mapping

The previous metrics capture misalignment between the map that emerges from power users and the map that emerges from the crowd, in terms of what is being mapped and where. What is left open is to measure which of the two virtual maps is more accurate and more up to date. Previous research has shown that, in Wikipedia, high accuracy is mainly due to the actions performed by power editors [33]. We are thus interested in measuring whether a similar phenomenon is emerging also in OSM. To do so in an automatic way is extremely difficult, as we would need access to an idealistic ground truth map, against which to compare the two virtual ones emerging in OSM. As such ground truth is not ready available, we will leverage past findings to find good proxies instead.

Let us consider accuracy first. By comparing geographic information in OSM with that of proprietary mapping datasets (e.g., Ordnance Survey, Navteq, Yelp), for selected regions and for selected periods of time, past studies have found that OSM spatial data is accurate [15, 16, 30], both in terms of positioning of OSM nodes and ways, and in terms of lexicographic spelling of POI names. Furthermore, it has found that OSM nodes' accuracy (both spatial and lexicographic) is strongly correlated with the amenity type field being non-null (recall that this field is optional in OSM) [30]. We thus measure the proportion of OSM objects with a non-null amenity field and use it as proxy for accuracy. More precisely, we compute $NotNullRatio_p$ and $NotNullRatio_c$ as the ratio of edits with non-null amenity appearing in the virtual map generated by power users and in the virtual map generated by the crowd respectively. Similarly to what was done when defining Amenity Diversity Misalignment in Equation 2, we then define an *Accuracy Misalignment* (AcM) metric as the relative difference between them:

$$AcM = \frac{NotNullRatio_p - NotNullRatio_c}{\max\left(NotNullRatio_p, NotNullRatio_c\right)} \tag{5}$$

Let us now consider how to measure which of the two virtual maps is more up-to-date. As for accuracy, it is rather difficult to automatically quantify this, as we do not possess a ground truth dataset we can compare against. However, once again we can measure a plausible proxy by looking at the ratio of OSM POIs whose details have been updated at least once since first being added to the map; we consider this a signal that, for whatever reason (e.g., urbanisation of an area, migration, or unlikely natural disaster) stale information is removed as map updates are carried out. We thus compute $UpdRatio_p$ and $UpdRatio_c$ as the ratio of OSM objects that have been updated at least once in the power users' and in the crowd's virtual maps respectively. Similarly to what was done when defining Amenity Diversity Misalignment in Equation 2, we are interested to see the relative difference between the update ratio of the virtual map generated by power users and that generated by the crowd. For this reason, we define the *Update Misalignment* (UM) metric as:

$$UM = \frac{UpdRatio_p - UpdRatio_c}{\max\left(UpdRatio_p, UpdRatio_c\right)} \tag{6}$$

Dimension	*Metric*	*Min*	*1st Qu.*	*Median*	*3rd Qu.*	*Max.*
What	AM	0.01	0.03	0.05	0.08	0.15
	ADM	-0.31	-0.18	-0.01	0.07	0.24
Where	SM	0.20	0.35	0.45	0.56	0.87
	SDM	-0.84	-0.46	-0.27	-0.09	0.24
How	AcM	-0.57	-0.14	0.12	0.39	0.74
	UM	-0.38	-0.12	0.23	0.52	0.92

Table 2. Summary Statistics of Metrics

RESULTS

Quantifying Bias in the Map

We have computed the previously defined six metrics in each of the 40 countries under exam; summary statistics of the obtained values can be found in Table 2. We now elaborate on these results.

What

From the first block of rows in Table 2, we can observe that the virtual map generated by power users and the one produced by the crowd show very little misalignment in terms of what types of amenities are being mapped. In fact, AM values range between 0.01 and 0.15: that is, in all the analysed countries, the amenities mapped by power users are roughly the same (even if with different magnitude) as those mapped by the crowd. This low misalignment is further reflected in the diversity of amenities being mapped, with values of ADM in the range -0.18 to 0.24 across three quarter of the considered countries (from the first to the last quartile). These two results combined suggest that the misalignment on the 'what' dimension is low.

In order to give a graphical and intuitive visualisation of the low misalignment between the types of amenities mapped by power users and the crowd, Figure 2 plots how many times in Norway (the country with the highest AM value of 0.15) each amenity has been mapped by the crowd (x axis) and by power users (y axis). From this plot we can observe that, with the only exception of amenity type 'charging_station', that has been heavily edited by power users only, all other amenities have been commonly mapped by power users and by the crowd (albeit with different magnitude scale).

Where

From the second block of rows in Table 2, we can observe that spatial misalignment between the map generated by power users and the map generated by the crowd is higher than the 'what' dimension. In fact, SM varies between 0.35 and 0.87 in three quartiles, meaning that, in three quarter of the considered countries, the areas mapped by power users only partially overlap with those mapped by the crowd. Furthermore,

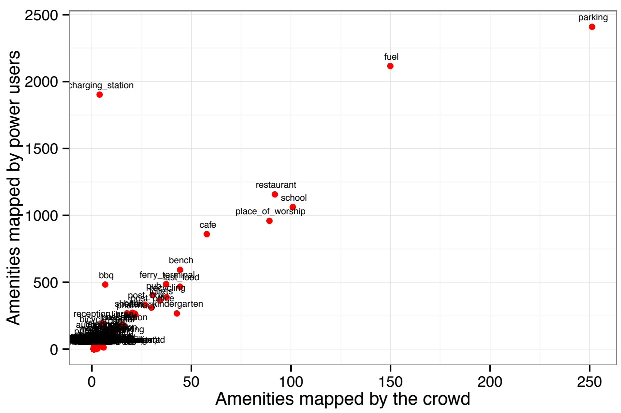

Figure 2. How many times in Norway each amenity has been mapped by the crowd (x-axis) and by power users (y-axis)

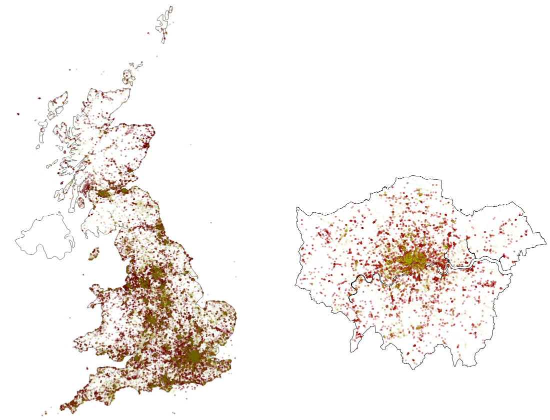

Figure 3. Edits performed by powers users (in yellow marks) and the crowd (in red marks) in United Kingdom (left) and London (right). This figure is best viewed in color.

Metric	*Sub-Metric*	*Min*	*1st Qu.*	*Median*	*3rd Qu.*	*Max.*
AcM	$NotNullRatio_p$	0.34	0.46	0.52	0.57	0.65
	$NotNullRatio_c$	0.14	0.44	0.46	0.69	0.94
UM	$UpdRatio_p$	0.27	0.39	0.41	0.45	0.48
	$UpdRatio_c$	0.02	0.11	0.30	0.44	0.74

Table 3. Summary Statistics of Sub-Metrics from which AcM and UM are Based

SDM is generally lower than zero, suggesting that the virtual map produced by power users concentrates on fewer areas than the one produced by the crowd. One may argue that this is simply a natural consequence of the fact that power users are much fewer than the crowd, and that, in geographic wikis, one can only map where one lives (e.g., knowledge is significantly constrained by geography, unlike what happens in Wikipedia). It is thus 'obvious' that there is high spatial misalignment between the two virtual maps of power users and the crowd. However, given how easily people adopt one map and trust its content, it is important to understand the impact of this finding.

To this end, Figure 3 overlays the two virtual maps, one generated by power users (in yellow marks) and one by the crowd (in red marks), both for the whole United Kingdom (left) and for the metropolitan city of London (right). The darker the color, the higher the relative concentration of edits. As shown, power users intensively map a very few selected areas, these being major urban areas in the UK (e.g., London, Liverpool and Manchester), and selected central areas within the city of London. Crowd mappers spread their activity over the whole country (and also over the whole city of London) much more uniformly instead. While we illustrate the case of the United Kingdom and London only, similar patterns have been observed in other countries we have examined. This signals an important problem: if power users concentrate their edits in a few, heavily urbanised areas (perhaps because that is where they live or mingle), who is going to map all the areas where there is absence of power users? And zooming in within urban areas: who is going to map areas within a city where power users do not exhibit an interest? We will elaborate on this point in the next section.

It is worth noting that, while spatial misalignment is high in three quarters of the countries examined, this is not the case for the remaining quarter, where SM is low and ranges between 0.20 and 0.35. These include Ecuador, Portugal, Philippines and South Africa, countries that have less than 1,000 OSM contributors each. It may thus be that these countries have not yet completely passed the cold start phase of OSM usage, with the user base being mainly composed of homogeneous early adopters, rather than a broader and more varied crowd.

How

We now turn our attention to the third and final block of rows in Table 2. Unlike what previously found for the 'what' and 'where' dimensions, here results are not clear-cut: both the AcM metric (proxy for accuracy) and UM metrics (proxy for up-to-dateness) record high absolute values across most countries, suggesting that crowd editing behaviour is substantially different from power users editing behaviour. However, in about half the countries these values are negative (i.e., crowd workers take charge of filling in details and of updating content), while in the other half these are positives (i.e., power users take charge of these editing activities). To shed light on this, we also examined the sub-metrics from which AcM and UM are derived – that is, $NotNullRatio_p$, $NotNullRatio_c$, $UpdRatio_p$ and $UpdRatio_c$. Summary statistics of the obtained values can be found in Table 3. An important observation can be drawn from these values: while the behaviour of power users does not vary much across countries (e.g., the ratio of edits with a non-null amenity filed goes from a minimum of 0.34 to a maximum of 0.65), there is a huge variation in the behaviour of the crowd (e.g., from a minimum $NotNullRatio_c$ of 0.14 up to 0.94, and from a $UpdRatio_c$ of 0.02 up to 0.74). We thus hypothesize that culture plays a role in the way the 'how' dimension of bias manifests itself. This hypothesis is corroborated by recent studies of various online collaborative work platforms, that have revealed how the manifested users' behaviour is shaped by social norms and values that vary between countries [38, 12, 11, 34, 37]. To verify whether this hypothesis is supported, we proceeded as follow.

First, we needed to grow confidence in the fact that OSM edits in a given country are indeed performed by users who are representative of the country's culture. Though we cannot know if OSM users were 'born and bred' in any given country or if, for instance, they were migrants who only recently moved there, we can grow confidence in our analysis after performing the following steps: we inferred a 'home country' for each OSM user with more than 5 edits (while marking 'homeless' those with less than that); there was no ambiguity in this inference, as all users had over 90% of their edits (and the vast majority exactly 100%) done in a single country. We then computed, for each examined country, the ratio of edits done by home users relative to all edits done in that country; finally, we removed from our analysis countries with less than 70% of contributions made by (inferred) locals; eight countries were removed during this processing, including Canada, Libya, Nigeria, and Peru. We then proceeded with our cultural analysis, focusing on the three dimensions that previous cross-cultural studies of users' behaviour in collaborative platforms have found to be most significant [11]:

- *Power Distance Index* (PDI). This is a Hofstede's cultural dimension [20] and it represents the extent to which a nation expects power to be unequally distributed. For example, in countries where PDI is higher, decision making is accepted to be in the hands of a few powerful people, rather than a broader citizen base. Among all the countries we have included in this study, Austria, Denmark and Ireland are those with lowest PDI values; conversely, Russia, Philippines and Guatemala have the highest ones.

- *Individualism vs Collectivism* (IDV). This also is a Hofstede's cultural dimension [20] and it measures the extent to which individuals are integrated into groups: the higher the individualistic score, the higher the importance given to personal achievements and individual rights (or of those within very close circles), as opposed to wider community interest. Among all the countries we have included in this study, Guatemala, Ecuador and Colombia have the lowest IDV values; conversely, United Kingdom, Australia and United States have the highest ones.

- *Gross Domestic Product Per Capita* (GDP-pc). This index, obtained from the CIA World Factbook,[17] is often considered a good indicator of a country's standard of living. Among all the countries we have included in this study, Pakistan, Nigeria and Philippines have the lowest values of GDP-pc; conversely, United States, South Africa and Norway have the highest ones.

We then computed the Spearman Correlation coefficients between cultural/economic factors and the sub-metrics from which AcM and UM are based (i.e., $NotNullRatio_p$, $NotNullRatio_c$, $UpdRatio_p$ and $UpdRatio_c$). Table 4 shows the results.

Let us consider the first block of rows (i.e., $NotNullRatio_p$ and $NotNullRatio_c$ metrics). Results suggest three interesting findings: first, there is limited correlation between any

Metric	*Sub-Metric*	*PDI*	*IDV*	*GDP-pc*
AcM	$NotNullRatio_p$	-0.19 *	0.21 *	0.12
	$NotNullRatio_c$	-0.39 *	0.44 **	0.17 *
UM	$UpdRatio_p$	-0.18 *	0.23 *	0.16
	$UpdRatio_c$	-0.49 **	0.31 *	0.19 *

Table 4. Spearman Correlation between our metrics regarding the dimension 'how' and cultural/economic factors with p-values codes (0 '' 0.01 '*' 0.05 ' ' 1)**

cultural dimension and the behaviour of power users; indeed, Table 3 had already shown that the ratio of non-null amenities by power users did not vary much among countries. Taken together, these results suggest that OSM power users form a rather homogeneous group of contributors across the globe. Second, we find a significant correlation between the behaviour of the crowd and both power distance (-0.39) and individualism (0.44). This suggests that the virtual map emerging from crowd edits in countries with high IDV/low PDI has a higher ratio of non-null amenities (our proxy for accuracy) than the one emerging from crowd edits in countries with low IDV/high PDI; this is in line with what was found in [37], where it was showed that people from countries with high IDV/low PDI tend to edit conscientiously (lower null amenity ratio). Furthermore, this level of non-null amenities is even higher than what achieved by power editors in the same countries; this explains the negative values of AcM in Table 2. Viceversa, in countries with low IDV/high PDI, the crowd-generated map achieves lower ratio of non-null amenities than the power users-generated map, thus accounting for the positive values of AcM in Table 2. Third and last, wealth as measured by GDP-pc bears a weak correlation with the accuracy with which contributors edit the map.

If we now turn our attention to the second block of rows in Table 4 (i.e., $UpdRatio_p$ and $UpdRatio_c$ metrics), we find very similar results. That is, the behaviour of power users is only weakly correlated with the culture. On the contrary, the behaviour of the crowd is significantly correlated with power distance (-0.49): the lower the PDI, the higher the ratio of updates performed by the crowd; in high PDI countries, this ratio is even higher than the update ratio of power users, thus accounting for the negative values of UM in Table 2. This result is in line with what it was already found for Wikipedia [34], where people from countries with high power distance are less likely to modify/delete others' work, as if they did not feel to have the power or right to do so. Finally, GDP-pc plays only a weak role in the way editors update the map. We elaborate on the practical and theoretical implications of these findings next.

DISCUSSION

Main Contributions

In this paper, we have proposed a method to quantitatively measure bias in crowd-sourced geographic information repositories. The method first requires to segment the self-selected crowd-base into two groups; for the purpose of this study, we have been looking at power users versus crowd mappers in OpenStreetMap. The method then defines metrics to measure the difference between two virtual maps, each one emerging from the sole contributions of one group, in terms

[17] **`https://www.cia.gov/library/publications/the-world-factbook`**, last access: 27/05/2013.

of: what spatial objects are being edited, where these objects are physically located, and how meticulously they are edited. We have thus quantified to what extent the emerging map in OSM is representative of the interests of the broad OSM contributors' base, despite the fact that a self-selected group of less than 20% power users is responsible for over 80% of OSM content.

Our findings reveal that the two user bases are well aligned in terms of what information is being mapped: the POIs of interest to power-users are of the same type being mapped by the crowd. However, we found strong bias in terms of the areas being mapped, with the map emerging from power users' activity being very dense around city centres, and near empty elsewhere, whereas the map constructed by the crowd covers the whole country, as well as intra-urban areas, much more uniformly. Finally, we found bias in how meticulously the map is being edited to vary with culture: where power distance is low and individualism is high, the crowd performs a more accurate work than power users (as captured by our proxies – non-null ratio and update ratio), and viceversa.

Practical Implications

Being able to quantify bias using the method presented in this paper has practical implications, to the benefit of both developers of crowd-mapping platforms (be them OSM, CrowdMap, FixMyStreet, etc.), and their potential end-users.

Developers of geographic crowd-sourcing platforms can use this method to measure the extent to which the ongoing mapping task is being representative of the broader contributors' user base. Despite the crowd-sourcing platform being open to anyone, in practice there are natural self-selection processes that cause a small subset of the user base to actually drive the majority of the mapping effort. This is not necessarily a problem: it may well be that this very active core is doing editing work that is well representative of the interests of the whole editors' base. However, up to now there has been no automatic way to measure at scale whether this was the case, and to what extent. The method we proposed offers a solution, with intuitive metrics that can reveal (mis)alignments between the two groups.

As the user base evolves over time, so does the bias they introduce. As an example, a recent study has revealed that, in 2008, power users on Amazon Mechanical Turk were primarily American, young, female, and well-educated; however, only one year after, a fundamental change had taken place, and the majority of the Turkers had then become male Indians [26]. By repeatedly applying the method we proposed over time, one can then keep track of how the mapping task is evolving, as a result of potential changes in the contributors' community.

Once bias has been made visible to developers of crowd-mapping initiatives, actions can be taken to reduce it, if deemed appropriate. What these actions might be is an open research question. For example, possible within-platform interventions include explicit cues for mapping certain types of spatial objects, and/or certain areas. At the moment, with the exception of Cyclopath, we are aware of no crowd-mapping platform that suggests what to map, where and how in any way; however, in Cyclopath, highlighting neglected areas of a map has been shown to be effective in directing mappers' work in those areas [36]. If a map happened to show 'what'/'where' forms of biases, similar visual cues could be used with the aim to drive the mapping efforts of power and occasional users alike, although research is required to see under what circumstances crowd-mappers would respond as Cyclopath users to such cues (the geographic spread of the mapping community probably being an important parameter). As evidenced by our analysis, some forms of bias vary with culture, suggesting that different interventions might be required in different countries: for example, where PDI is high, triggers that explicitly call the crowd to update possibly stale information may encourage them to conduct more map maintenance work than what they currently do; this is not necessary in countries with low PDI, as the crowd seems to already perform a more meticulous job than power users.

Being able to quantitatively measure bias in a map empowers tool developers to make it visible to map users (be them citizens, businesses and/or non-for-profit organisations): as Prof. Jerry Brottom clearly stated in a 2013 Wired article,[18] there is no such thing as a perfect map: *"No map can be perfect and there will always be gaps or lack of detail – whether it's the Scottish Island of Jura that Google somehow lost entirely, private roads in wealthy areas or informal settlements like slums. It's something that people often forget or get complacent about when they think they've found the best option though."*. While this is a basic principle that all geographers and cartographers know, people tend to forget that. If maps were to show their biases, users could make informed choices of which one to rely upon for their varying needs. The GIScience community has been actively researching methods to visualise uncertainty in geospatial data [29]; as we become capable of measuring bias too, similar lines of research open up, which may lead to the development of new and alternative approaches to map-data visualisation, and of the effects that they have on users' mapping behaviour.

Theoretical Implications

The ability to measure bias in online collaborative platforms is a powerful tool in the hands of researchers too. As online collaborative platforms have been proliferating, many large datasets have become available that afford us the ability to observe, measure and model human activity at a level of granularity unprecedented before. Indeed, a whole new research field called 'computational social science' has emerged, lying at the intersection of computer science, social science, and statistics, that claims to be able to validate old theories about human activity and relationships for the first time at scale, and possibly develop new ones.

The approach adopted by computational social scientists to understand human behaviour initially attracted criticism from traditional social scientists, who objected that the digital traces left in online collaborative platforms were not representative of the broader population, and thus their findings

[18] `http://http://www.wired.co.uk/news/archive/2013-08/05/slum-mapping-google-maps-cartography`

could not be generalised, despite being derived from much larger data samples than those used in previous social science studies [2]. Developing methods to quantify bias in online collaborative platforms offers researchers the ability to understand to what extent their findings are generalisable, and under what circumstances.

This paper makes one step in that direction: we have developed a method to quantitatively measure the bias introduced by power users versus occasional contributors in crowd-sourced maps. The comparison of power vs. occasional users has wide applications, as most online collaborative platforms exhibit the same engagement dichotomy. However, there are many other forms of bias that need quantification, including demographic bias (from gender to age to ethnicity and the like), knowledge bias (from amateur or hobbyists to professionals), and community-roles bias (from moderators to contributors to developers).

Limitations

One limitation of this work is that all results have been computed per country level. A more fine-grained study, for example at regional or city level, would be able to reveal biases emerging from more localised mapping dynamics than those we detected with this work. However, the method proposed in this paper could also be applied to more fine-grained studies of crowd-sourced geographic datasets.

Similar concerns apply to the temporal granularity of the study: here we studied the bias resulting from three years of cumulative activity; however, the group of power users may change over time, and so may their work. By selecting a more fine-grained time interval, and by repeating the same type of analysis over time, changes in bias, possibly resulting from changes in the power user base, would be captured.

It is also worth pointing out that, for scalability reasons, this work focuses on quantifying bias in OSM by looking at only one OSM data type, that is nodes. If other OSM data types were considered (e.g., ways and relations), different biases may emerge. We leave the investigation of this question to future work; however, we note that the applicability of the method proposed in this paper withstands.

We would also like to acknowledge the fact that the correlations found between the meticulousness of the emerging map and culture do not imply any causality. There might indeed exist other underlying factors that contribute to the above correlations and that our analysis does not capture. For example, map updates may take place for different reasons, such as urbanisation or migration processes, and these processes might be related to country's cultural and economic factors.

Last but not least, we are aware that the metrics we defined for the 'how' dimension are nothing but proxies; although we grounded these choices in findings from past works, there is need for further evidence that links our 'accuracy metric' and 'update metric' to what we indeed aim them to capture. Smaller scale qualitative studies could be conducted to shed light in this direction.

Finally, we would like to restate that this work focuses on quantifying bias in the *map* produced by *power users* versus that emerging from the activity of the much larger crowd of *occasional mappers*. Our method considers users' interest as captured by what OSM users edit in the map. We are aware of two main limitations of this method: first, what users map may only broadly, and not exactly, represent their interests. Second, while this method enables us to discuss how representative the emerging OSM map is of the broader OSM *contributors*' base, it does not afford us the ability to discuss its representativeness with respect to the interests of the very many lurkers that use the map without however leaving any digital trace of what their interests might be.

Future Work

As we previously hinted to, an important direction of future research is the analysis of the temporal evolution of bias in crowd-sourced geographic datasets. As time progresses, the user base keeps changing across a variety of dimensions (e.g., size, demographics, expertise); furthermore, as more and more information is being mapped, saturation effects may start to emerge, with potential impact on the user base itself (as some users may lose interest, once they see certain POI types have been completely mapped, or when they see 'sufficient' information being mapped). Developing a method that is capable of tracking users' evolution during their engagement with the crowd-mapping platform, and of measuring bias in relation to the stage of 'maturity' and 'adoption' of a crowd-sourcing platform, would offer platforms' developers a better understanding of how the community is evolving, and what types of interventions might be required and would be most effective over time.

Acknowledgment

This work was partly supported by the 3cixty EIT ICT Labs project.

REFERENCES

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

2. Boyd, D., and Crawford, K. Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society 15*, 5 (2012), 662–679.

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

4. Bryant, S., Forte, A., and Bruckman, A. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. In *Proc. of GROUP*, ACM (2005), 1–10.

5. Callahan, E., and Herring, S. Cultural Bias in Wikipedia Content on Famous Persons. *Journal of the American Society for Information Science and Technology 62*, 10 (2011), 1899–1915.

6. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y., and Moon, S. I Tube, You Tube, Everybody Tubes: Analyzing the

World's Largest User Generated Content Video System. In *Proc. of IMC*, ACM (2007), 1–14.

7. Chango, S., Kumar, V., Gilbert, E., and Terveen, L. Specialization, Homophily, and Gender in a Social Curation Site: Findings from Pinterest. In *Proc. of CSCW*, ACM (2014), 674–686.
8. Cohen, N. Define Gender Gap? Look Up Wikipedias Contributor List. *The New York Times* (January 2011).
9. Dai, W., Jin, G. Z., Lee, J., and Luca, M. Optimal Aggregation of Consumer Ratings: An Application to Yelp.com. Tech. rep., National Bureau of Economic Research, 2012.
10. Fana, H., Zipfa, A., Fub, Q., and Neisa, P. Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science (IJGIS) 28*, 4 (2014), 700–719.
11. Gallagher, S., and Savage, T. Cross-cultural analysis in online community research: A literature review. *Computers in Human Behavior* (2012).
12. Garcia-Gavilanes, R., Quercia, D., and Jaimes, A. Cultural Dimensions in Twitter: Time, Individualism, and Power. In *Proc. of ICWSM* (2013).
13. Girres, J., and Touya, G. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS 14*, 4 (2010), 435–459.
14. Goodchild, M. Citizens as Sensors: the World of Volunteered Geography. *GeoJournal 69*, 4 (2007), 211–221.
15. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design 37*, 4 (2010), 682–703.
16. Haklay, M., Basiouka, S., Antoniou, V., and Ather, A. How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus Law to Volunteered Geographic Information. *The Cartographic Journal 47*, 4 (2010), 315–322.
17. Halfaker, A., Geiger, R., Morgan, J., and Riedl, J. The Rise and Decline of an Open Collaboration System: How Wikipedia's reaction to sudden popularity is causing its decline. *American Behavioral Scientist 57*, 5 (2013), 664–688.
18. Halfaker, A., Kittur, A., Kraut, R., , and Riedl, J. A Jury of your Peers: Quality, Experience and Ownership in Wikipedia. In *Proc. of WikiSym*, ACM (2009).
19. Hecht, B., and Stephens, M. A Tale of Cities: Urban Biases in Volunteered Geographic Information. In *Proc. of ICWSM 2014* (2014).
20. Hofstede, G. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. SAGE Publications, 2001.
21. Howe, J. The Rise of Crowdsourcing. *Wired* (2006).
22. Hristova, D., Quattrone, G., Mashhadi, A., and Capra, L. The Life of the Party: Impact of Social Mapping in OpenStreetMap. In *Proc. of ICWSM* (2013).
23. Hu, M., Lim, E., Sun, A., Lauw, H., and Vuong, B. Measuring Article Quality in Wikipedia: Models and Evaluation. In *Proc. of CIKM*, ACM (2007).
24. Ishida, K. Geographical Bias on Social Media and Geo-local Contents System with Mobile Devices. In *Proc. of HICSS* (2012), 1790–1796.
25. Kittur, A., Chi, E., Pendleton, B., Suh, B., and Mytkowicz, T. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. In *Proc. of WWW* (2007).
26. L., J. R., Irani, Silberman, M., Zaldivar, A., and Tomlinson, B. Who are the Crowdworkers? Shifting Demographics in Mechanical Turk. In *Proc. of CHI* (2010).
27. Lam, S. T. K., Uduwage, A., Dong, Z., Sen, S., Musicant, D. R., Terveen, L., and Riedl, J. WP:Clubhouse? An Exploration of Wikipedias Gender Imbalance. In *Proc. of WikiSym*, ACM (2011), 1–10.
28. Ludwig, I., Voss, A., and Krause-Traudes, M. A Comparison of the Street Networks of Navteq and OSM in Germany. *Advancing Geoinformation Science for a Changing World 1*, 2 (2011), 65–84.
29. Maceachren, A. M., Robinson, A., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Information Science 32* (2005), 160.
30. Mashhadi, A., Quattrone, G., Capra, L., and Mooney, P. On the Accuracy of Urban Crowd-Sourcing for Maintaining Large-Scale Geospatial Databases. In *Proc. of WikiSym*, ACM (2012).
31. Neis, P., Zielstra, D., and Zipf, A. Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet 5*, 2 (2013), 282–300.
32. Neis, P., and Zipf, A. Analyzing the Contributor Activity of a Volunteered Geographic Information ProjectThe Case of OpenStreetMap. *ISPRS International Journal of Geo-Information 1*, 2 (2012), 146–165.
33. Panciera, K., Halfaker, A., and Terveen, L. Wikipedians Are Born, Not Made: a Study of Power Editors on Wikipedia. In *Proc. of GROUP*, ACM (2009), 51–60.
34. Pfeil, U., Zaphiris, P., and Ang, C. S. Cultural Differences in Collaborative Authoring of Wikipedia. *Journal of Computer-Mediated Communication 12*, 1 (2006), 88–113.
35. Priedhorsky, R., Lam, S., Panciera, K., Terveen, L., and Riedl, J. Creating, Destroying, and Restoring Value in Wikipedia. In *Proc. of GROUP*, ACM (2007), 259–268.

36. Priedhorsky, R., Masli, M., and Terveen, L. Eliciting and Focusing Geographic Volunteer Work. In *Proc. of CSCW*, ACM (2010), 61–70.

37. Quattrone, G., Mashhadi, A., and Capra, L. Mind the Map: The Impact of Culture and Economic Affluence on Crowd-Mapping Behaviours. In *Proc. of CSCW*, ACM (2014), 934–944.

38. Reinecke, K., Nguyen, M. K., Bernstein, A., Näf, M., and Gajos, K. Doodle Around the World: Online Scheduling Behavior Reflects Cultural Differences in Time Perception and Group Decision-Making. In *Proc. of CSCW*, ACM (2013), 45–54.

39. Rost, M., Barkhuus, L., Cramer, H., and Brown, B. Representation and Communication: Challenges in Interpreting Large Social Media Datasets. In *Proc. of CSCW*, ACM (2013), 357–362.

40. Shannon, C. A Mathematical Theory of Communication. *The Bell System Technical Journal 27* (1948), 379–423 and 623–656.

41. Singhal, A. Modern Information Retrieval: A Brief Overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering 24*, 4 (2001), 35–43.

42. Stephens, M. Gender and the GeoWeb: Divisions in the Production of User-Generated Cartographic Information. *GeoJournal* (2013), 1–16.

43. Vasconcelos, M., Ricci, S., Almeida, J., Benevenuto, F., and Almeida, V. Tips, Dones and ToDos: Uncovering User Profiles in FourSquare. In *Proc. of WSDM*, ACM (2012), 653–662.

44. Zielstra, D., and Zipf, A. A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In *Proc. of AGILE* (2010).