Mind the Map: The Impact of Culture and Economic Affluence on Crowd-Mapping Behaviours

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

Afra Mashhadi
Bell Labs
Alcatel Lucent
Dublin, Ireland
afra.mashhadi@alcatel-lucent.com

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

ABSTRACT
Crowd-mapping is a form of collaborative work that empowers citizens to collect and share geographic knowledge. OpenStreetMap (OSM) is a successful example of such paradigm, where the goal of building and maintaining an accurate *global* map of the changing world is being accomplished by means of *local* contributions made by over 1.2M citizens. While OSM has been subject to many country-specific studies, the relationship between national culture and economic affluence *and* users' participation has been so far unexplored. In this work, we systematically study the link between them: we characterise OSM users in terms of who they are, how they contribute, during what period of time, and across what geographic areas. We find strong correlations between these characteristics and national culture factors (e.g., power distance, individualism, pace of life, self expression), and well as Gross Domestic Product per capita. Based on these findings, we discuss design issues that developers of crowd-mapping services should consider to account for cross-cultural differences.

Author Keywords
OSM; 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 has become a popular paradigm to harvest user-generated content into meaningful and informative collections. A specific domain where this paradigm has been widely applied to is that of volunteered geographic information: citizens have become surveyors, with council-monitoring applications like FixMyStreet;[1] local reporters, as powered by Ushaidi's Crowdmap;[2] and cartographers, with geo-wikis like Cyclopath[3] and OpenStreetMap.[4] We refer to this instantiation of the crowd-sourcing paradigm as crowd-mapping.

OpenStreetMap (OSM) is perhaps one of the most successful examples of crowd-mapping, with currently over 1M users, collectively building a free, openly accessible, editable map of the world. Various country-specific studies of OSM have been conducted, mainly to quantify the accuracy (e.g., in the UK [8], France [7], and Germany[21]) and coverage (e.g., in London, UK [23]) of its information, with respect to centrally-maintained geographic datasets. However, an important aspect that has so far been neglected is the relationship between national culture and economic affluence *and* this form of collaborative work.

Previous research has shown that cultural values shape our thinking, behaviours and social procedures [17]. Such cultural differences have visible manifestations not only in the actions people undertake in the physical world, but also in the behaviour they manifest in the online world (e.g., [28, 6]). We thus hypothesize that users' participation in computer-mediated crowd-mapping activities like OSM varies significantly between countries. As such initiatives heavily rely on local volunteers to accomplish their goals, quantifying users' engagement in countries around the world, through the lens of national culture and economic affluence, should highlight important design considerations that engineers of crowd-mapping systems might ponder, to better account for national diversities.

In this paper, we study the behaviour of nearly 165,000 OSM users spread across 35 countries, who have collectively made over 13M contributions to OSM to date. We describe OSM crowd-mapping characteristics, on a per country basis, in terms of: *who* the contributors are (power vs. normal users); *how many* contributions they make, and over *how long* a period of time; *how detailed* contributions are; *when* edits are

CSCW'14, February 15–19, 2014, Baltimore, Maryland, USA.
Copyright 2014 ACM 978-1-4503-2540-0/14/02...$15.00.
http://dx.doi.org/10.1145/2531602.2531713

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

made (both to capture speed of OSM adoption, and burstiness of activity); and, finally, *where* in the country edits are done (how geographically spread they are). We then correlate these OSM factors with cultural and economic factors that anthropologists have found to be most accountable for cross-country differences: *power distance*, that is, the extent to which people (especially those less powerful) accept power to be unequally distributed [13]; *individualism vs. collectivism*, that is, the extent to which social relationships are loose (e.g., people look after themselves), as opposed to relationships integrated in strong and cohesive groups [13]; *pace of life*, that is, the speed at which we move, the high-pressure at which we work, and the density of experiences and activities we undertake daily [18]; *self-expression vs. survival*, that is, the extent to which subjective well-being and quality of life take precedence over economic and physical security, as captured by the World Value Survey [15]; and, finally, *Gross Domestic Product*, to capture a country's standard of living. We find that GDP per capita and power distance are strongly correlated with users' engagement (how many OSM users exist in a country, to what extent they contribute, and over what period of time); self-expression and pace of life are strongly correlated with the quality of the edits (how detailed they are, and how geographically spread); finally, GDP per capita is strongly correlated with the speed of technology adoption and burstiness of contributions. We discuss the implications of these findings in assessing the long-term sustainability of OSM across cultures, and suggest design questions that engineers of crowd-mapping initiatives should consider, to account for cross-cultural differences. We finally conclude the paper with a discussion of known limitations and future work.

RELATED WORK

The relationship between cultural values and users' social behaviours has been the subject of extensive research in the domain of online communities, following seminal work of social psychologist Geert Hofstede [13]. Hofstede administered opinion surveys to IBM employees in over 70 countries. After analysing over 100,000 questionnaires, he discovered significant differences between cultures; he was then able to attribute most of the variance in the data to five cultural factors: individualism vs. collectivism (i.e., the extent to which individual goals take precedence over collective ones), power distance (i.e., the extent to which social hierarchies are accepted), masculinity vs. femininity (i.e., the importance of values such as competitiveness, ambition and power as opposed to relationships and quality of life), uncertainty avoidance (i.e., a society's propensity to accept unusual changes and the resulting uncertainty), and long-term orientation (i.e., a disposition to favour long-term rewards vs. short-term ones). Various cultural anthropologists and psychologists have subsequently proposed other (related) cultural factors, that help understand people's differing behaviours when faced with the same situation in the real world. For example, Hall [9] proposes the concept of polychronic versus monochronic time orientation, to differentiate between cultures that give higher importance to human interactions rather than strict time management and punctuality; Levine and Norenzayan [18] suggest the concept of pace of life, measured as a combination of people's walking speed, the work speed of postal clerks, and the accuracy of bank clocks, to analyse behavioural difference across cultures; and Inglehart and Baker [15] refer to self-expression vs. survival values, that is, the extent to which subjective well-being and quality of life take precedence over economic and physical security (which tend to be taken for granted in more industrialised countries), as captured by the World Value Survey.

These theories have been used as a framework to analyse people's differing behaviour not only in the physical world, but also in their manifested actions in online communities. A very thorough and recent survey of such studies has been compiled by Gallagher and Savage [5]. A common aim behind these works is to help improve our understanding of how cultural variations impact online community dynamics, so to more effectively manage them. Studies span a variety of aspects: from understanding the impact of cultural differences on people's motivation to use an online community [22], to the impact of culture on behavioural dynamics, both in social networking websites [4, 16, 6] and more recently within online group decision-making tools [28].

Hofstede's cultural attributes (power distance and individualism vs. collectivism in particular) have been used to understand online crowd-sourcing behaviours too, more specifically in the context of Wikipedia (e.g., [26, 10]). In this case, language has often been chosen as unit of analysis. For example, Pfeil et al. [26] have analysed Wikipedians' editorial actions in light of cultural factors, and found strong negative correlation between power distance and 'delete' actions, strong positive correlation between collectivism and 'expand' actions, and strong positive correlation between individualism and 'modify' actions. They went on to explain these findings as follow: Wikipedians from countries where subordinate relationships are well accepted (high power distance), are more reluctant to adopt subversive (i.e., delete) actions, as they accept they do not have the right or privilege to do so. Wikipedians from collectivist countries are more likely to add information to expand already existing Wikipedia pages, while those from countries with high individualism tend to have their own opinions prevail over group consensus, as exhibited by the higher number of modifications (i.e., corrections) to existing articles. Findings about editing behaviours in Wikipedia cannot be directly transferred to the crowd-mapping domain, as the unit of cultural analysis they used (i.e., the language edition of Wikipedia) is too coarse grained. More precisely, in the crowd-mapping domain exemplified by OSM, where the task to be accomplished (i.e., mapping geographic information) requires contributors to have detailed local knowledge, a finer (country) level of granularity is required.

Studies conducted to date in the domain of crowd-mapping can be broadly grouped in two streams: on one hand, analyses of the quality of the volunteered geographic information; on the other hand, investigations into the motivation behind users' engagement with the mapping tasks. Quality of information has been measured in OpenStreetMap in terms of accuracy (e.g., in the UK [8], France [7], and Germany[21]), and more recently coverage (e.g., in Germany [30] and in London, UK [23]). While the former has been found to be consistently high, coverage is presently low, especially as one moves away from urban areas to more rural ones. This finding raises concerns in terms of the long-term sustainability of crowd-mapping: is coverage going to spontaneously grow, or are there going to be areas that will continue to be neglected? Studies so far have been limited in that they quantify coverage at one specific point in time, and in so doing they cannot disclose the factors that contribute to crowd-mapping participation over time; furthermore, comparative studies between different countries are lacking.

A second stream of research has looked into users' motivation behind the mapping task. OpenStreetMap contributors' motivation has been studied by means of an online questionnaire [3]: 444 contributors of differing age, education, and past OSM experience took part. Among them, ideology (i.e., the willingness to contribute to the community that is behind the OSM project) emerged as the single highest motivational factor; they also found geography to be a community-wide motivator, with both casual and serious mappers concerned with building a complete map of their local territory. These motivational factors emerged also in studies of Cyclopath, another successful geo-wiki that is being used to digitally map route information for cyclists in Minneapolis. Indeed, researchers found that, after being attracted to volunteer information in response to a call to fix a problem, contributors remain engaged to aid the local cycling community [25, 27]. However, findings from these studies cannot be generalised to differing countries, as even in [3] over 90% of respondents where from either the UK or North America.

A challenging question that has so far been left unexplored is thus the extent to which people's engagement with a crowd-mapping initiative varies with respect to national characteristics: are there countries where a crowd-mapping task will succeed more than in others? Can we identify them, both to raise awareness (of importance, for example, to those building location-based services on top of such maps), as well as to intervene with purpose-planned corrective actions? As a first step to provide answer to these questions, we present in this paper a large-scale, quantitative analysis of the extent that national *cultural factors* and *economic affluence* correlate with users' crowd-mapping behaviour in OpenStreetMap.

DATA DESCRIPTION

OpenStreetMap

We begin our study with a detailed description of the crowd-sourcing dataset at hand, that is, OpenStreetMap. The dataset is freely available to download[5] and contains the history (since 2006) of all edits (over 2 billions) performed by all users (over 1M) 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) (e.g., cafes, restaurants, hospitals, schools); 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).

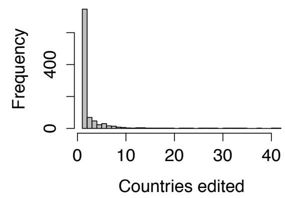

Figure 1. Distribution of number of edited countries per user

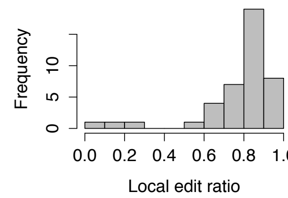

Figure 2. Distribution of local edits per country

In order to reduce the OSM dataset to a more manageable size, we restricted our attention to nodes only. A node consists of three main attributes: a geographical position (latitude and longitude), a name, and an optional amenity type (e.g., hospital, cafe). Furthermore, we used stratified sampling and selected 43 countries, out of the 133 presently mapped in OSM, so to have representatives across the whole range of cultural and economic values we focus on in this study (further details about these values in the next section). From this sample, we removed all contributions that appeared to be 'bulk imports' (i.e., over 1K edits, performed by a single user, within a single session, and spread over a large geographic area). We chose to discard these edits as not representatives of 'human' crowd-sourcing activity, but rather of 'data donations' given to the wider OSM community by external organisations. These imports represent less than 30% of the total edits considered.

Last but not least, if we are to examine crowd-mapping behaviour in relation to national cultural factors, we need 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. 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; however, we can grow confidence in our analysis by performing the following processing: first, we infer 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, as shown in Figure 1. We then compute, for each of the 43 pre-selected countries, the ratio of edits done by home users, over all edits done in that country; Figure 2 illustrates the resulting distribution. Finally, we remove from our analysis countries with less than 70% of contributions made by (inferred) locals. Eight countries were removed during this processing, including four (Canada, Libya, Nigeria and Peru) where the majority of edits are indeed 'foreign'. In the end, we are left with a dataset comprising nearly 165,000 users, who have collectively made over 13M contributions to OSM to date, across the 35 countries coloured in Figure 3. We next describe the cultural and economic factors we chose to investigate in our analysis of cross-country crowd-mapping behaviours.

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

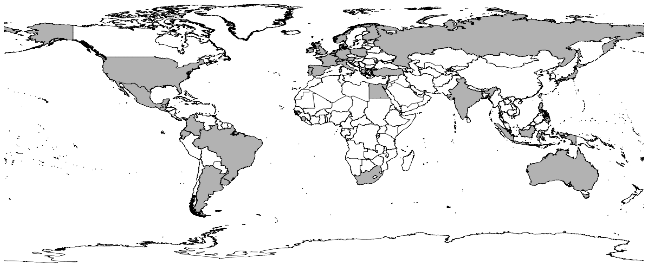

Figure 3. Map of final 35 countries under analysis.

Cultural and Economic Factors

Based on previous cross-cultural studies of users' behaviours in online communities [5], we have selected the following five cultural and economic factors to analyse manifested crowd-mapping behaviours in OSM.

Power Distance Index (PDI). The first cultural feature we consider is Hofstede's Power Distance Index [13]. This feature 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. Pfeil et al. [26] have studied the relationship between this feature and editing activity in different language editions of Wikipedia, and found that the higher the PDI, the lower the number of 'delete' actions on each other's work (as if people did not feel they had the power or the right to make the decision of deleting somebody-else's work). In the context of OSM, we expect PDI to be strongly correlated with *who* OSM editors are (power vs. normal users), and *how* they map (using what actions).

Individualism vs Collectivism (IDV). The second cultural feature we consider is Hofstede's Individualism index [13]. This feature 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. Western cultures tend to have high IDV, as opposed to Eastern ones, that are more collectivist instead. This feature has also been studied in the context of Wikipedia, where it has been observed that: the lower the IDV (the higher the collectivism), the higher the number of 'addition' to Wikipedia articles; while the higher the IDV (the higher the individualism), the higher the number of 'modification' to Wikipedia articles [26]. In other words, 'modifications' are signals of individual opinions prevailing over group consensus, while 'additions' are signals of cross-group collaborations towards the growth and development of Wikipedia. As for the power distance index, we expect IDV to correlate both with *who* OSM editors are (power vs. normal users), and with *how* they map (using what actions).

Gross Domestic Product Per Capita (GDP-pc). The third metric we use is Gross Domestic Product per capita, as obtained from the CIA World Factbook.[6] This index is often considered a good indicator of a country's standard of living. To participate in crowd-mapping, citizens must be able to afford the technology (i.e., a desktop computer and an Internet connection) in the first place, so we expect GDP-pc to manifest itself in relation to *who* the OSM contributors of a country are (how many of them, with respect to the country population), and *where* (how spread) edits are across the whole territory (the poorer the country, the more geographically clustered the edits are likely to be, around a few centers of relative wealthier status).

Self-Expression Index (SEI). As our fourth factor, we use the Self-Expression Index measured by Inglehart and Baker [15] using data collected by the World Value Survey [1]. This survey provides a comprehensive country-level measurement of major areas of human concern, from religion to politics, from economics to social life. After analysing the data, Inglehart and Baker found that most of the cross-national variance in the survey data could be explained using the Self-Expression Index. Intuitively, the higher the score, the greater the extent to which subjective well-being and quality of life take precedence over basic survival values, such as economic and physical security (which tend to be taken for granted in most post-industrial societies). We expect our analysis to highlight a significant correlation between SEI and *how much* OSM users contribute.

Pace of Life Index (PLI). The fifth and final cultural factor we consider is Pace of Life Index, originally defined by Levine et. al [19] as a combination of walking speed, work speed, and accuracy of public clocks. The index was computed for 31 countries, so that the higher the pace, the lower the PLI value (to be interpreted as number of seconds). Many Western Europe countries (e.g., Switzerland, Germany, Sweden) as well as Japan exhibit very fast pace of life (very low PLI), while Latin American ones have a much slower paced lifestyle (high PLI). Reinecke et. al. have analysed the relationship between pace of life and users' participation in group decision making [28], and indeed found that, in countries with faster pace, the number of Doodle polls completed per Internet user is higher. PLI has been widely analysed in cross-cultural studies of users behaviours in online communities [5]; we have thus chosen to include it in our analysis, despite the fact we have this value for only 16 of the selected 35 countries. Even in this restricted sample, we expect PLI to manifest itself in a variety of ways: from *how much* OSM users edit, to *when/where* edits enter the map.

Figure 4 illustrates how the five cultural/economic factors above vary among the countries under exam – the darker the shade, the higher the corresponding factor. As it emerges from these maps, these factors are not independent of each other: for example, countries with higher GDP per capita also exhibit higher individualism index, while countries with lower GDP per capita tend to have a slower pace of life (higher PLI). Table 1 details the dependency (measured as Pearson Correlation coefficients) between these five factors. We note that, as often the case when performing cross-cultural analysis, we risk stereotyping the people within a country based on the country's dominant cultural values [5];

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

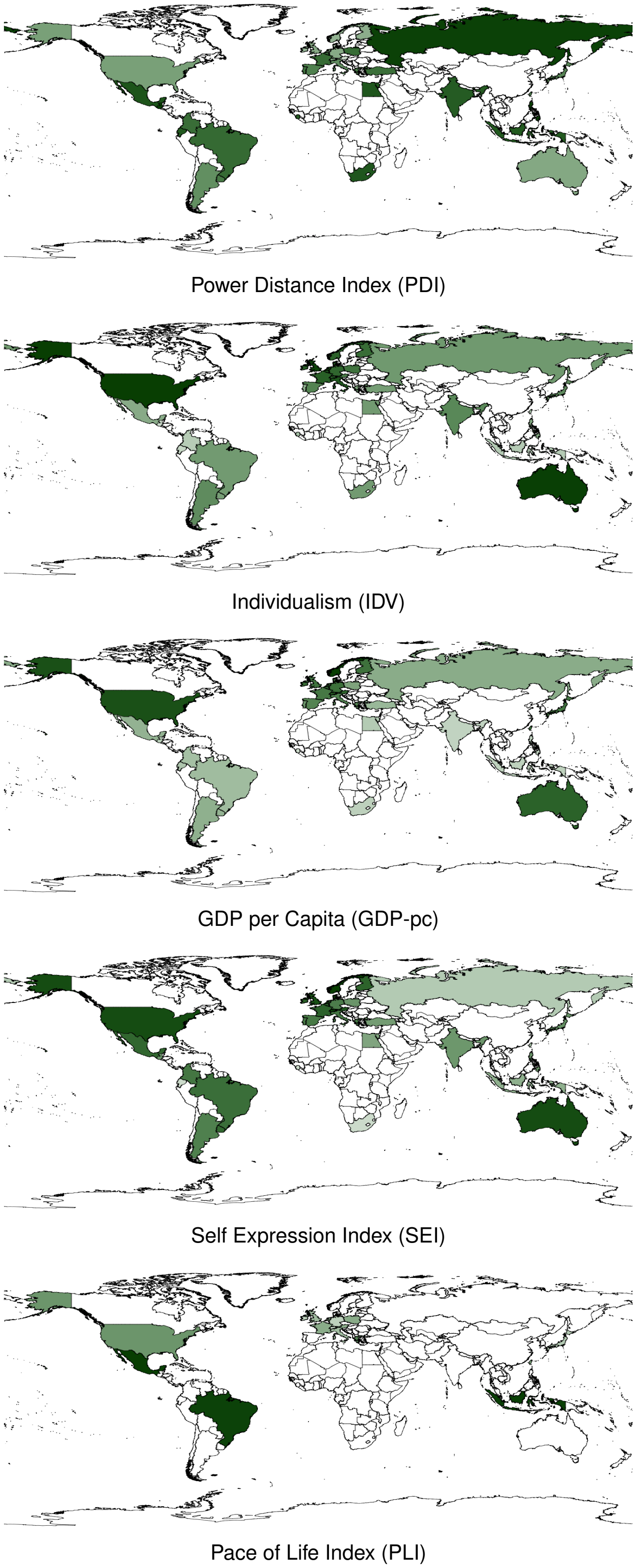

Figure 4. Choropleth Maps of Countries' Cultural/Economic factors – the darker the shade, the higher the corresponding index. White areas mark countries that have not been included in the study.

however, the purpose of the study is not to capture within-country cultural differences, but rather to analyse contrasting behavioural tendencies. As OSM users have been observed to be a rather homogeneous group of people (i.e., young, male, educated, in full-time employment and with Internet access) [3], we may actually assume they have a comparable sub-culture, so that potential differences in crowd-mapping behaviour between countries are indeed most likely due to the influence of cultural and economic factors.

	PDI	IDV	GDP-pc	SEI	PLI
PDI	—	-0.69 ***	-0.72 ***	-0.71 ***	0.72 ***
IDV	-0.69 ***	—	0.69 ***	0.67 ***	-0.71 ***
GDP-pc	-0.72 ***	0.69 ***	—	0.65 ***	-0.87 ***
SEI	-0.71 ***	0.67 ***	0.65 ***	—	-0.29
PLI	0.72 ***	-0.71 ***	-0.87 ***	-0.29	—

Table 1. Correlations among the cultural/economic factors with p-values codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 ' ' 1**

METHODOLOGY

In order to understand to what extent crowd-mapping behaviour in OSM correlates with the cultural and economic factors described in the previous section, we adopt the following methodology. First, we describe the manifested OSM behaviour *for each country* under exam along the following dimensions.

Who the OSM Users are. We compute both the number of OSM contributors per country population, as well as the ratio of OSM users over a country Internet population. We expect the former to be strongly correlated to GDP-pc, while with the second we are partly controlling for individual economic wealth. A consistent finding in the study of collaborative-work platforms has been that a small fraction of users is responsible for the vast majority of contributions (e.g., [2, 14]). In these cases, the 80-20 rule-of-thumb is frequently found to hold: that is, less than 20% of users (often called 'power users'), are responsible for more than 80% of the content. A substantial body of research has indeed studied motivation and commitment of such participants in user-generated content systems, acknowledging the central role they play in the sustainability of such systems. In this study, we thus also compute the ratio of OSM power users over a country's population/Internet population/OSM users population. To do so, we first need to define what a OSM power user is. We proceeded in steps: first, we computed the frequency distribution of OSM user edits on a per country basis, and found them to be very similar among each other (Figure 5 shows the frequency distribution plots – in log scale – for two sample countries, France and United Kingdom). We then ordered OSM contributors per country (from the most to the least active), computed their cumulative ratio of contributions, and finally considered 'power users' the minimum number of contributors required to cover 80% of the country's OSM content. If we look at Figure 6, where the cumulative contribution plots are illustrated for the case of France and United Kingdom, we find that power users are those that make more than 100 edits. We used this threshold across all 35 countries; in all countries, the OSM power user population was less than 10% of the OSM population.

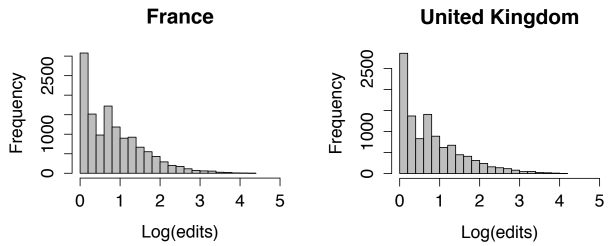

Figure 5. Frequency distribution of OSM user edits

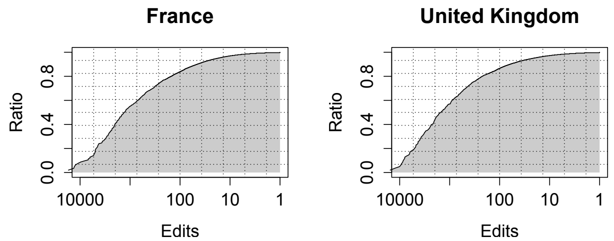

Figure 6. Cumulative ratio of contributions of OSM users

How long/how much OSM users edit. We describe the 'average OSM contributors' in each country in terms of: longevity, that is, the time passed between their first and last edit; and the average number of edits their perform during their OSM lifetime. We also consider the 'average OSM power contributors' separately, and compute how long/how much they edit on average. As these values follow a power law distribution, we compute their geometric mean.

How the edits themselves are performed. In studies of Wikipedia, cultural factors have been linked to the proportion of add/modify/delete actions performed on articles. In OSM, the objects being edited are of a much simpler nature than Wikipedia articles, so we look at the following two editorial characteristics instead: first, the (in)completeness of the information provided. In OSM, nodes only require to have a name and spatial (latitude/longitude) information, while the amenity type (e.g., cafe, school, etc) is optional. Previous studies have shown that the less 'complete' the information being provided about a POI, the lower its spatial positional accuracy [24]. We thus compute the proportion of OSM POI edits within a country where the amenity type field is left blank. Second, we compute the number of edits that modify information previously entered.

When the edits are performed. We capture two different aspects: first, we look at how quickly the OSM user base grows in the first year of OSM adoption in that country; second, we capture how consistent the influx of contributions is over time. To quantify the former, we compute the average monthly contributors that OSM attracted during its first year of usage across each country. To quantify the latter, we count how many edits are performed in each month, then compute the ratio of the Simpson diversity index[7] on such vector of values over the number of considered months (the higher the *normalised temporal* diversity, the lower the burstiness of contributions over time).

Where the edits are performed. Finally, we compute how spatially clustered edits are in a country. To do so, we divided each country into 5Km by 5Km grid cells, computed how many edits fall within each cell, then calculated the Simpson diversity index on such vector of values (the higher the *spatial* diversity, the more geographically distributed edits are over the whole country territory). Note that, depending on the country's territory, there might be areas that have no urbanisation at all (e.g., deserts), and would thus require no edits. To account for this, we normalise the computed value by the size of the country's urbanised areas.

Table 2 summarises all the above mentioned parameters, together with a plot of their frequency distribution. We also list three countries following the top/bottom values.

Once parameters were determined, we computed the Pearson Correlation between the above OSM characteristics and the previously selected cultural/economic factors (for parameters that were moderately/strongly skewed, we first computed the square root/log value); a complete set of results is shown in Table 3. As Table 1 previously highlighted, cultural/economic factors are correlated with each other. To better understand what cultural/economic factors are most strongly linked to what OSM characteristics, we proceeded as follow: for each OSM characteristic, we built a series of regressions, starting from the simplest (using as single dependent variable the cultural/economic factor with the highest correlation value as per Table 3), then adding one by one the other factors. Each time, we checked the adjusted R^2 value of the more complex models, to see if the newly added cultural/economic factors where indeed improving the model, or whether they were not adding any information w.r.t. the OSM characteristic under exam. We elaborate on these results in the next section.

RESULTS

Who are OSM users?

To begin with, we look at the penetration of OSM in a country, in terms of the number of OSM users, both relative to the country population, and relative to the number of Internet users in that country. We expect the former ratio to have strong correlation with GDP-pc: citizens must be able to afford the technology in the first place, so that richer countries are more likely to have a larger user base. However, with the second ratio we partly control for wealth, so that we should be able to observe what other cultural values are correlated with OSM engagement. The first two rows of Table 3 show the result of these correlations. As expected, GDP-pc has a very strong correlation (0.78) with the former, but less so with the latter (0.56). It is interesting to observe that at the top of the scale, countries such as Austria, Germany and Switzerland (characterised by a high GDP-pc), exhibit the highest penetration of OSM users, both relative to the country population and relative to the number of Internet users in that country. This is however not the case when looking at countries scoring lowest on this OSM factor. To understand if other cultural

[7]The Simpson diversity index is a measure that reflects how many different entries there are in a data set and the value is maximized when all entries are equally high [29].

	Parameter	Min	Distribution	Max	Bottom 3 countries	Top 3 countries
Who	OSM users per population	2.0e-06		8.2e-04	India, Egypt, Indonesia	Austria, Germany, Switzerland
	OSM users per million internet users	22		1097	Japan, Egypt, Brazil	Austria, Switzerland, Germany
	OSM power users per population	1.1e-07		6.1e-05	India, Egypt, Sierra Leone	Austria, Germany, Switzerland
	OSM power users per million internet users	0.80		82	Egypt, Sierra Leone, Taiwan	Austria, Germany, Switzerland
	OSM power users per OSM users	0.023		0.11	Uruguay, Egypt, Sierra Leone	Japan, Germany, Finland
How much	Geometric mean edits per user	2.4		7.9	Sierra Leone, Uruguay, Egypt	Germany, Portugal, Finland
	Geometric mean edits per power user	204		571	Portugal, Indonesia, Taiwan	Sierra Leone, Russia, UK
How long	Geometric mean user lifetime (days)	2.1		7.8	India, Mexico, Sierra Leone	Germany, Switzerland, Austria
	Geometric mean power user lifetime (days)	7.0		316	India, Guatemala, Colombia	Sierra Leone, UK, Germany
How	Null amenity ratio	0.43		0.77	UK, Guatemala, Philippines	Sierra Leone, Turkey, Egypt
	Overwrite ratio	0.19		0.50	Sierra Leone, US, Turkey	Finland, Austria, Germany
When	Speed of adoption (active users per month)	0.25		16	Guatemala, Ecuador, Sierra Leone	Finland, UK, Spain
	Normalised temporal diversity	0.054		0.89	Egypt, Sierra Leone, Guatemala	UK, Germany, Switzerland
Where	Normalised spatial diversity	2.5e-05		5.5e-03	Australia, Brazil, Russia	Belgium, Switzerland, Netherlands

Table 2. Description of our parameters

		PDI		IDV		GDP-pc		SEI		PLI	
Who	OSM users per population	-0.71	***	0.60	***	0.78	***	0.64	***	-0.77	***
	OSM users per Internet users	-0.53	***	0.46	**	0.56	***	0.52	**	-0.57	*
	OSM power users per population	-0.74	***	0.63	***	0.78	***	0.62	***	-0.75	***
	OSM power users per Internet users	-0.59	***	0.51	***	0.61	***	0.50	**	-0.63	**
	OSM power users per OSM users	-0.37	*	0.46	**	0.48	**	0.19		-0.57	*
How much	Geometric mean edits per user	-0.23		0.45	**	0.34	*	0.04		-0.67	*
	Geometric mean edits per power user	-0.04		0.13		-0.04		-0.09		-0.63	**
How long	Geometric mean user lifetime	-0.62	***	0.69	***	0.66	***	0.44	*	-0.68	**
	Geometric mean power user lifetime	-0.29		0.39	*	0.21		0.26		-0.53	*
How	Null amenity ratio	0.37	*	-0.41	*	-0.38	*	-0.42	*	0.12	
	Overwrite ratio	-0.44	**	0.40	*	0.40	*	0.29		-0.67	**
When	Speed of adoption	-0.40	*	0.21		0.46	**	0.40	*	-0.40	
	Normalised temporal diversity	-0.59	***	0.68	***	0.69	***	0.62	***	-0.62	*
Where	Normalised spatial diversity	-0.38	*	0.33		0.57	***	0.30		-0.85	***

Table 3. Pearson Correlation with p-values codes (0 '*' 0.001 '**' 0.01 '*' 0.05 ' ' 1)**

factors are correlated with OSM adoption, we computed a series of multiple linear regressions: we first added Pace of Life Index (second highest correlation in Table 3) to GDP-pc, but found the model not improving over the basic one with GDP-pc only; we then discovered that by adding Power Distance Index we improved the model the most (adjusted R^2 growing from 0.29 to 0.36). Indeed, PDI has a strong negative correlation (-0.53) with OSM users per Internet population. Although there might well be hidden factors that this correlation analysis does not capture, we might speculate that countries where people do not perceive a strong hierarchical structure and division of power (low PDI) are more likely to take part in collaborative work, as represented by OSM.

To illustrate this further, we plot in Figure 7 the number of OSM contributors per full population (left) and the number of OSM contributors per Internet population (right) as they vary with GDP per capita (x axis) and Power Distance Index (size of the circle). As shown, when we consider the correlation between GDP per capita and the number of OSM contributors per country population (left), there are very few outliers, reflecting the strong correlation between the two parameters. However, when considering the correlation between GDP per capita and the number of OSM contributors per Internet population (right), the number of outliers increases significantly, signalling that GDP per capita alone is no longer able to capture OSM user engagement. If we look at the outlier countries, we observe that many countries with low Power Distance Index cluster above the fitted line (and viceversa).

We then repeated the same analysis as above, but focusing on OSM *power* users (rows 3 and 4 of Table 3): results are very similar to those obtained when looking at the broader OSM population, confirming the above observations.

Finally, we looked at the number of OSM power users over the OSM population itself, to see if the emerging contributors' roles in OSM vary between countries. We observe a strong negative correlation with PLI (-0.57): that is, the more 'relaxed' the pace of the country (high PLI), the lower the number of power users within the OSM user base, as if people would not add 'OSM mapping' to their schedules as a frequent activity to engage in. Rather, most OSM contributors engage equally in these countries. For countries where we do not have PLI values (recall we have PLI for only 16 out of 35), the other factor that counts the most is GDP-pc instead: in this case, wealthier countries also have a higher number of power users.

We have so far analysed the composition of the OSM user base, but have not yet looked into how much they contribute when they do engage. We address this next.

How much do they edit?

We have computed the average number of OSM contributions made every month. The higher the average, the higher the continuous engagement with the system, the higher the OSM long-term sustainability overall (a recent study conducted for the city London, UK, has found that OSM users who do not edit for 3 to 6 months tend to leave the system entirely [14]). We first consider the whole OSM user base, and computed the geometric mean of monthly edits. As shown in Table 3, PLI is again negatively correlated with this OSM characteristic (-0.67): not only countries with a slower pace have fewer power users, they also overall make much fewer contributions

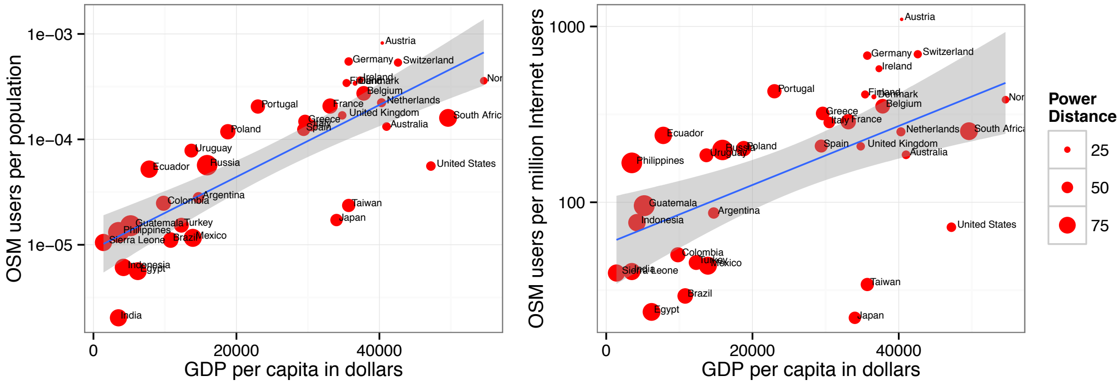

Figure 7. Engagement of OSM user around the world

than countries where people tend to fit many different activities in their daily life. Using linear regressions, we found that power distance was also useful to further understand this OSM characteristic: that is, the higher the power distance in the country, the fewer the average number of contributions per OSM user. Editors from countries such as Uruguay, Egypt and Sierra Leone, characterised by a slow pace of life and high power distance, tend to edit significantly less than contributors from Germany, Portugal and Finland (i.e., countries with faster pace of life and lower power distance). Based on this finding, we might speculate that in countries where citizens expect the few in power to 'do something', they are less apt to take the initiative to contribute to crowd-sourcing platforms like OSM, where users are all equally expected to contribute.

We repeated this analysis focusing on power users only. As observed when looking at the whole OSM user base, we found PLI to be negative correlated with this OSM characteristic (-0.63): both normal and power users from countries with a slower pace make much fewer contributions than those from countries with more active lifestyles. Using linear regression, we then tried to see if other cultural factors (e.g., power distance) were related to how much power users edit: in contrast to what observed with the whole OSM user base, we found no other cultural factors to be significant.

How long do they remain engaged?

We next analyse OSM user engagement in terms of how long users keep contributing to the map (that is, the time passed between the first and last edit of each user). We first consider all OSM users, and compute the geometric mean of their lifespan. As shown in Table 3, this factor has the strongest correlation with Individualism (0.69): we might speculate that editors from countries where one's goals and desires are highly valued, tend to remain engaged for longer. For example, OSM users from countries such as Germany, Switzerland and Austria, which are characterised by high individualism, are those who are engaged for the longest time; this is in contrast to users from countries such as India, Mexico, and Sierra Leone, where individualism is low instead, and so is OSM users' longevity. The only other factor that significantly contributes to this OSM characteristic, when performing multiple linear regression, is power distance (correlation -0.62): OSM users tend to disengage with collaborative-work platforms quicker, if they come from countries where the norm is to have the few in power to take charge of common-good tasks.

When we look at power users' longevity, Individualism is confirmed as an important cultural factor; however, pace of life bears even stronger relevance (-0.53): power users remain engaged longer in countries with active and full lifestyles (low PLI).

How are the edits performed?

Having characterised OSM users, we now look at their edits. The first aspect we are interested in is the 'null amenity ratio': previous studies have shown that OSM objects with a non-null amenity field also have higher spatial accuracy [24]. Are there cultural/economic factors that would enable us to expect lower null amenity ratio (thus higher accuracy) for the countries under exam? As Table 3 shows, Self-Expression Index has the highest (negative) correlation (-0.42): we might speculate that in countries that particularly care about quality of life and subjective well-being, OSM edits tend to be conscientiously done (lower null amenity ratio). Adding other factors(e.g., PDI, GDP-pc, IDV) to a regression model does not improve the model further.

The second characteristic we look at is the overwrite ratio, or simply the number of 'modify/delete' actions over the total number of edits. When performing a mapping task, we need to be aware of the saturation effect: if the digital map is near complete, modify actions are likely to be more frequent than add actions, relative to a country whose digital map is still very sparse. We do not know the level of saturation of OSM in the countries under exam; according to a recent study [23], the OSM map in London, UK, is still rather sparse in terms of Points-of-Interest, and the UK is the country where OSM was born and where it has been edited the most. We could then speculate that OSM is not yet at a stage where saturation emerges. In this case, two cultural factors are most strongly (negatively) correlated with the overwrite ratio: first, the Pace of Life Index; that is, the more 'relaxed' the pace of the country (high PLI), the less engaged OSM users are in overwrites. Second, the Power Distance Index: in line with findings from Pfeil et al. [26] for Wikipedia, people from countries with high power distance are less likely to modify/delete others' work, as if they do not feel to have the power or right to do so.

When are contributions made?

We are next interested in looking at the speed of adoption of OSM. To do so, we look at the first year of use of OSM in each selected country, and compute the monthly average number of unique users that contribute to the map. As Table 3 shows, the factor that most strongly correlates with this metric is GDP per capita; that is, OSM penetration is faster in richer countries. There are a number of factors we cannot control for and that might bias results (e.g., local events organised by OSM to recruit more participants). We re-computed the correlation, after discarding countries where we knew the OSM organisation to be particularly active in terms of recruitment campaigns (i.e., UK, US, Germany, France); even in this case, results were confirmed, with GDP-pc being the single most relevant factor in explaining OSM speed of adoption.

We also correlated cultural/economic factors with a metric aimed at capturing the 'burstiness' of contributions; we do so as we expect burstiness to be low (i.e., Simpson temporal diversity to be high), in countries where OSM has become a mature and stably-adopted technology, with a solid and loyal user base whose contributions keep coming in over time (and predictably so). Viceversa, burstiness may be high in countries where the OSM user base is very volatile, with large contributions sometimes coming in, followed by periods with near no activity, so that the long-term sustainability of the map is much more difficult to predict. Regression analysis shows that two factors are most accountable for high diversity (low burstiness): GDP-pc and Self-Expression, that is, the higher the wealth and well-being of the country, the more organic the growth of the map. Countries such as United Kingdom, Germany and Switzerland, that are characterised by high GDP-pc and high Self-Expression, are in the top three position for low 'burstiness' of contributions.

Where are contributions made?

The last OSM factor we measure is the geographic spread of the edits made on a per country basis. The lower the spatial diversity, the higher the geographic clustering of OSM user edits in a few urbanised areas; this might be problematic, as it signals that some areas are being mapped, while many others are being severely neglected. As shown in the last row of Table 3, we observe a very strong negative correlation with PLI (-0.85): that is, the faster the pace of life and work in the country (low PLI), the higher the spatial diversity; at the same time, the higher the economic affluence of a country (GDP-pc), the higher the spatial diversity. This may suggest that countries with low PLI and high GDP-pc, such as Belgium, Switzerland and Netherlands, are more uniformly developed and better connected internally, so that OSM edits cover a much wider area (relative to the country's urbanised areas).

DISCUSSION

Implications

The previous results have highlighted strong correlations between the culture and economic affluence of a country, and the way OpenStreetMap is being used within it. Although cultural and economic characteristics of a country do not capture all factors that may contribute to the observed OSM users' behaviours, they help highlighting important national differences that engineers of crowd-mapping initiatives should consider to account for cross-cultural differences. We elaborate on them below.

On User Engagement. Our analysis has revealed that GDP per capita is negatively correlated with the size of the OSM user base severely. Once we control for that (by computing the ratio of OSM users over Internet users), power distance is negatively correlated with the size of the participating crowd. A design question we may then ask is how to enlarge the user base in this context. Word-of-mouth dissemination is unlikely to attract new contributors, as the task at hand (i.e., building a map of the world) is possibly perceived as one to be undertaken by 'those in power', rather than by all citizens. Public recruitment campaigns, backed up by people or institutions that are highly regarded within countries with high PDI, may have a better chance of succeed in this goal. Having the financial means to acquire the technology needed remains a hindrance, but mapping does not necessarily have to be done with personal devices. Indeed, schools could make use of OSM to teach subjects such as geography, while engaging new users in the mapping task.

Power distance has a strong negative correlation not only with the size of the user base, but also with how much they contribute. As mentioned above, being asked to contribute by those in power may act as good motivator to start engaging users more. Further research is needed to evaluate to what extent we can rely on intrinsic motivation to contribute (as OSM does now), and to what extent reward systems should be brought in instead, to stimulate and to improve the user engagement within the collaborative work platform. In countries with high power distance index, rank in a society is usually highly regarded, so that leveraging a (virtual) social status within the OSM community could be used as reward system to foster user engagement: for example, from new user, to mayor of an area, to curator of content (as in Wikipedia). This may also help address another issue for which Power Distance seemed 'culpable', that is, the short lifetime of OSM users.

Apart from PDI, Pace of Life is correlated with the amount of contributions people make: countries with a slow pace do not seem to add crowd-mapping to the range of activities they wish to allocate their time to. If this is indeed the case, a possible venue for exploration in this case is implicit crowd-sourcing, that is, embedding the crowd-mapping task within (or disguised as) another task they would normally do anyway (e.g., by children in schools, as part of teaching a geography class).

On Task Quality. A common concern in crowd-sourcing activities is related to the quality of the task, once performed by skilled employees, and now done by any volunteer. Quality in OSM can be measured in a variety of ways, including accuracy and coverage. We did not explicitly measure these two metrics as part of this work, but we can consider 'null amenity ratio' as proxy for the former, and spatial diversity ratio as proxy for the latter. Null amenity ratio is lowest (expected highest accuracy) when Self-Expression Index is highest. What interventions can we plan in countries where survival needs dominate over subjective well-being instead? We may speculate that OSM users in such countries do not have the time to carefully edit OSM objects, pressed by other more basic needs. This is a case where a different design may easily help: for example, using a voting mechanism so that a new edit is not immediately added to the map, but first needs to be checked and approved by others for correctness. Voting may also be seen as a less strong action than a delete or modify, and might thus help in those countries where high power distance results in a very low overwrite ratio.

As for spatial diversity, we found that countries with lower economic affluence, as well as countries with slower pace of life, tend to have their edits clustered around few areas, with many other urban areas being largely neglected. To remedy this, a form of more 'directed' crowd-sourcing could be designed, similar to what Cyclopath first tried [25, 27]: rather than letting the OSM task be completely self-driven, attention of the OSM community could be explicitly driven towards more neglected areas. This could be subtle (e.g., highlighting neglected areas on the map), or be explicitly linked to the reward system mentioned above.

On Technology Adoption. Finally, we look at the speed of penetration of OSM and the level of burstiness of its use. In both cases, the factor that mostly correlates with these OSM characteristics is GDP per capita: the richer the country, the faster the technology adoption and the stabilisation of its use. When similar crowd-mapping initiatives are launched in countries of low GDP-pc, it may be worth planning low-cost recruitment campaigns coupled with shared access to mapping technologies, so to speed up the penetration of the crowd-mapping initiatives. Where there is a simple way of linking the collaborative task to educational institutions and their practices, this may represent a happy combination.

Limitations

The results presented in this study suffer from three main limitations: first, we acknowledge that the results presented in this paper do not imply any causality between a country's cultural and economic factors *and* its OSM user characteristics. There might indeed exist other underlying factors that contribute to the above correlations and that our analysis does not fully capture. For example, it might be that OSM is better known in Western countries than in the Eastern ones, and these Western countries also happen to have higher GDP-pc and individualism. Despite this limitation, we believe this study offers valuable insights on the differing crowd-mapping behaviours that emerge around the world; cultural and economic characteristics of the analysed countries represent a useful lens through which to present and quantify results.

Second, one may argue that a user could live in a country while editing mostly another, so that the inference scheme we use to associate a user to its home country (with related cultural and economic factors) would fail as an effect of the globalisation of online communities. However, there have been quite a few studies now, especially on Wikipedia, showing that most contributors to this online collaborative platform actually exhibit strong geographic locality in the articles they edit [11, 20, 12]. When the task undertaken has an even stronger geographic component than Wikipedia (i.e., mapping POIs), we may assume geographic distance to matter even more.

Third, assuming that the geographic boundaries of a country act as a sort of containment for it cultural and economic values (that we then equally apply to all its citizens), may attract criticism. This issue has indeed been acknowledged to be a limitation in any cross-cultural studies [13]; however, despite the fact that geographical countries may not be the best way to study culture, these are often the only kind of units available for comparison.

Future Work

The study presented in this paper has looked at users' engagement with OSM, while remaining agnostic of *what* people edit in different countries. A preliminary investigation on the types of OSM objects that people map around the world revealed that the range is indeed huge: from topographic information (e.g., roads, rivers), to basic services (e.g., schools, hospitals), from leisure offerings (e.g, restaurants, cafes), to possibly 'unusual' and very 'personal value' information (e.g., trees, benches, etc.). A future direction of research is to analyse what (combination of) categories of OSM objects are being edited around the world, possibly revealing interesting correlations between national characteristics of the examined countries and the type of information maps being created. A second direction of research we intend to pursue is to understand the within-country characteristics of crowd-mapping: in particular, we wish to quantify how spatially concentrated the contributions of OSM users are, whether this concentration depends on countries' characteristics (e.g., GDP-pc), and/or whether it varies with the type of information being mapped (i.e., some categories having very localised value only).

Acknowledgment

This work was partly supported by the Knowledge and Innovation Community EIT ICT Labs, and by the Intel Collaborative Research Institute on Sustainable Connected Cities.

REFERENCES

1. W. V. S. Association et al. World values survey 1981-2008 official aggregate v. 20090901. *Madrid: ASEP/JDS*, 2009.

2. S. L. Bryant, A. Forte, and A. Bruckman. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *Proceedings of the 2005 international ACM SIGGROUP conference on Supporting group work*, GROUP '05, pages 1–10, 2005.

3. N. Budhathoki and C. Haythornthwaite. Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist*, 2012.

4. C. N. Chapman and M. Lahav. International ethnographic observation of social networking sites. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pages 3123–3128. ACM, 2008.

5. S. E. Gallagher and T. Savage. Cross-cultural analysis in online community research: A literature review. *Computers in Human Behavior*, 2012.

6. R. Garcia-Gavilanes, D. Quercia, and A. Jaimes. Cultural Dimensions in Twitter: Time, Individualism, and Power. In *AAAI ICWSM*, July 2013.

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

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

9. E. Hall. *The Dance of Life: The Other Dimension of Time*. Anchor Books. Knopf Doubleday Publishing Group, 1984.

10. N. Hara, P. Shachaf, and K. F. Hew. Cross-cultural analysis of the wikipedia community. *Journal of the American Society for Information Science and Technology*, 61(10):2097–2108, 2010.

11. D. Hardy, J. Frew, and M. F. Goodchild. Volunteered geographic information production as a spatial process. *International Journal of Geographical Information Science*, 26(7):1191–1212, 2012.

12. B. J. Hecht and D. Gergle. On the localness of user-generated content. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 229–232. ACM, 2010.

13. G. H. Hofstede and G. Hofstede. *Culture's consequences*. SAGE Publications, Incorporated, 2001.

14. D. Hristova, G. Quattrone, A. Mashhadi, and L. Capra. The Life of the Party: Impact of Social Mapping in OpenStreetMap. In *AAAI ICWSM*, July 2013.

15. R. Inglehart and W. E. Baker. Modernization, Cultural Change, and the Persistence of Traditional Values. *American Sociological Review*, 65(1):19–51, 2000.

16. K. Karl, J. Peluchette, and C. Schlaegel. Who's posting facebook faux pas? a cross-cultural examination of personality differences. *International Journal of Selection and Assessment*, 18(2):174–186, 2010.

17. D. E. Leidner and T. Kayworth. A Review of Culture in Information Systems Research: Toward a Theory of Information Technology Culture Conflict. *Management Information Systems Quarterly*, 3(2):357–399, 2006.

18. R. Levine. *A Geography of Time: The Temporal Misadventures of a Social Psychologist, Or how Every Culture Keeps Time Just a Little Bit Differently*. BasicBooks, 1997.

19. R. V. Levine and A. Norenzayan. The pace of life in 31 countries. *Journal of cross-cultural psychology*, 30(2):178–205, 1999.

20. M. D. Lieberman and J. Lin. You are where you edit: Locating wikipedia contributors through edit histories. In *3rd International AAAI Conference on Weblogs and Social Media*, 2009.

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

22. V. Madupu and D. O. Cooley. Cross-cultural differences in online brand communities: an exploratory study of indian and american online brand communities. *Journal of International Consumer Marketing*, 22(4):363–375, 2010.

23. A. Mashhadi, G. Quattrone, and L. Capra. Putting Ubiquitous Crowd-sourcing into Context. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pages 611–622, 2013.

24. A. Mashhadi, G. Quattrone, L. Capra, and P. Mooney. On the accuracy of urban crowd-sourcing for maintaining large-scale geospatial databases. In *Proceedings of the 8th International Symposium on Wikis and Open Collaboration*, WikiSym '12, pages 15:1–15:10, 2012.

25. K. Panciera, M. Masli, and L. Terveen. "How should i go from ___ to ___ without getting killed?:" motivation and benefits in open collaboration. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 183–192, 2011.

26. U. Pfeil, P. Zaphiris, and C. S. Ang. Cultural differences in collaborative authoring of wikipedia. *Journal of Computer-Mediated Communication*, 12(1):88–113, 2006.

27. R. Priedhorsky, M. Masli, and L. Terveen. Eliciting and focusing geographic volunteer work. In *Proceedings of the International Conference on Supporting Group Work*. ACM, 2010.

28. K. Reinecke, M. K. Nguyen, A. Bernstein, M. Näf, and K. Z. Gajos. Doodle around the world: online scheduling behavior reflects cultural differences in time perception and group decision-making. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, pages 45–54, 2013.

29. E. H. Simpson. Measurement of diversity. *Nature*, 163:688–688, 1949.

30. D. Zielstra and A. Zipf. A comparative study of proprietary geodata and volunteered geographic information for germany. In *Proceedings of the 13th International Conference on Geographic Information Science*, 2010.