Quantitative evaluation of volunteered geographic information paradigms: social location-based services case study

Talia Dror, Sagi Dalyot* and Yerach Doytsher

As of 2010, 90% of the data that existed in the world were created within the previous 2 years, while personal location data have been singled out as one of the five primary 'big data' streams in the 2011 McKinsey report. By 2020, the volume of existing data will increase by 50-fold, where a large percentage of this volume will be associated with geospatial data. One of the reasons for this is the existence of the Volunteered Geographic Information (VGI) paradigm, which encapsulates the idea of using the Internet (Web 2.0) to create, share, visualise, and analyse geographic information and knowledge. This neogeography revolution has started to fundamentally transform how geographic data are acquired, maintained, analysed, visualised, and consequently used. Thus, it has the potential to influence common practices, since it captures a broad knowledge of the environment we live in, in all aspects of life, encompassing new services to take place, applications and processes to be developed – all of which are location based. The diversity of applications and services that explore the potential of VGI argues for its current usability relevance: ranging from transportation network analysis, to air pollution and air quality, to natural disaster decision-making systems. This revolution has contributed to the development of two important working and knowledge paradigms: Crowdsourcing and Wisdom of the Crowd, widely used today within the mapping and geo-information discipline. Still, both terms are commonly misused and replaced. This paper aims at distinguishing between the terms via the quantitative and theoretical examination of four widely used social location-based services: OpenStreetMap (OSM), Moovit, Waze and Ushahidi. Eight primary characteristics that influence the paradigm of both Crowdsourcing and Wisdom of the Crowd are defined and examined, aiming to investigate and emphasise the differences between the four, namely: diversity, decentralisation, independency, aggregation, knowledge, activity, privacy and exploitation. It was found that OSM is an excellent example of a Crowdsourcing service, while though Ushahidi is considered as a Crowdsourcing service, its characteristics are coupled better with those of Wisdom of the Crowd. Moovit and Waze do not correspond to the Crowdsourcing paradigm, and thus are categorised as Wisdom of the Crowd services.

Keywords: Crowdsourcing, Wisdom of the crowd, Social location-based services

Introduction

For more than three decades now, we have experienced a significant development of online publishing tools, and particularly of the World Wide Web (WWW) (Beners-Lee, Cailliau, Groff and Pollermann, 1992). These developments have simplified information interaction between users, together with the discovery and obtaining of enormous amounts of data and information. The invention of the WWW was especially meaningful, mainly because of the development of user interface (Bowman *et al.* 1994), which among others enabled and simplified the visualisation of geographic information. Nowadays, new mapping applications, infrastructures and services are overwhelming the Internet; this trend is known as 'The Geographic World Wide Web' (The GeoWeb) (Haklay, Singleton and Parker, 2008).

During the mid 2000s, the GeoWeb transformed itself as the platform for novel and pioneering online geographical information systems. This inspired and enabled the mapping field to become not only an experts' domain (authoritative only), but also a public domain, modifying the framework from top-down to bottom-up. Users all

Mapping and Geo-Information Engineering, Technion – Israel Institute of Technology, Haifa, Israel

*Corresponding author, email dalyot@tx.technion.ac.il.

Received 2 February 2015; accepted 2 February 2015
DOI 10.1179/1752270615Y.0000000013

over the world are nowadays involved in data processing, mainly thanks to Web 2.0 technologies, allowing public mapping services to become widely and extensively used (Haklay, 2010). Nowadays, more and more mapping projects and services use groups of volunteers focusing on the collection and dissemination of geographic data (as opposed to authoritative mapping agencies only), making it possible to create and update geospatial information infrastructure and services, such as online maps and applications, having the aspiration to practically replace licensed surveyors, cartographers and geographer experts. This phenomenon (and some may say revolution) is known as neogeography, which has contributed significantly to the development of two important knowledge paradigms: Crowdsourcing and Wisdom of the Crowd.

Wisdom of the Crowd

The expression 'Wisdom of the Crowd' was coined and extensively investigated by Surowiecki (2004), mainly within the business community, claiming that 'Large groups of people are smarter than an elite few, no matter how brilliant – better at solving problems, fostering innovation, coming to wise decisions, even predicting the future'. Surowiecki described the crowd as 'any group of people' that 'can act collectively to make decisions and solve problems'. According to Surowiecki, big organisations, such as companies or government agencies; small groups, such as students; and even groups that are not aware of themselves as being groups, such as gamblers, might act as a crowd. However, to compose a 'wise' crowd, four primary attributes need to exist:

1. Diversity – where each individual contributes diverse pieces of data or information.
2. Decentralisation – pointing to the fact the crowd's input (e.g. answers) are not subjective or biased from the inner groups' hierarchy, i.e. the 'people above them' (e.g. boss or sponsor).
3. Independence – indicating the person's opinion is not influenced by people from the group who exist in his close vicinity, but from their own personal judgement.
4. Aggregation – facilitating a mechanism that combines the individual opinions received unto a single collective decision or conclusion.

Crowdsourcing

The term Crowdsourcing was first coined by Howe (2006a) and since then is increasingly expanding, receiving considerable attention since then in a variety of research fields, such as economics, computing, mapping, etc., and also among companies, non-profit organisations and academic communities (Zhao and Zhu, 2014). Originally, Howe's definitions referred to a new business model, which since then has expanded to other fields, mainly because of web innovations (Brabham, 2008). The two key definitions made by Howe are:

1. 'The White Paper Version: Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call'.
2. 'The Soundbyte Version: The application of open-source principles to fields outside of software'.

One of the pivot articles dealing with the term Crowdsourcing is Brabham (2008), who describes Crowdsourcing as 'a distributed problem-solving model, is not, however, open-source practice. Problems solved and products designed by the crowd become the property of companies, who turn large profits from this crowd labor. And the crowd knows this going in'. Moreover, the author claims that 'Crowdsourcing can be explained through a theory of crowd wisdom, an exercise of collective intelligence, but we should remain critical of the model for what it might do to people and how it may reinstitute long-standing mechanisms of oppression through new discourses ... It is a model capable of aggregating talent, leveraging ingenuity ... Crowdsourcing is enabled only through the technology of the web'. In addition, the Merriam-Webster online dictionary defines Crowdsourcing as 'the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers'.

Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012) attempted to embed an integrated (to all disciplines) definition of Crowdsourcing, by making the following definition:

> Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or a company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilise to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken.

Paper objective

As will be discussed in the Related Work section, Wisdom of the Crowd and Crowdsourcing are easily confused and sometimes incorrectly used. For example, does Wikipedia have the nature of the Wisdom of the Crowd or Crowdsourcing? Wu, Harrigan and Cunningham (2011), for instance, made the assertion that Wikipedia has the characteristics of the Wisdom of the Crowd, though Howe (2006b) and Huberman, Romero and Wu (2009) stated that Wikipedia is a clear example of Crowdsourcing. Moreover, there exist research that do not make any differentiation between the two paradigms (Vukovic 2009; Stranders, Ramchurn, Shi and Jennings, 2011). It is the authors' belief that such a distinction should be given, since both encompass (at least to some extent) different characterisations and arrangements.

Moreover, the aforementioned definitions of Crowdsourcing and Wisdom of the Crowd are focused on various fields and disciplines, but not on the mapping and geo-information discipline. It seems that there are no clear definitions of both terms with respect to the geospatial domain, e.g. services, applications and processes. Accordingly, the purpose of this research paper is

to analyse these terms and knowledge paradigms with respect to social location-based services, to emphasise the special and unique attributes and characterisations within the mapping and geo-information discipline. An effort is made to characterise the differences between the two terms, with the use and analysis of four services that are widely used today by tens of millions of users around the globe: OSM (© OpenStreetMap contributors), Ushahidi (© 2008–2014 Ushahidi), Waze (© 2009–2014 Waze Mobile), and Moovit (© 2014 MOOVIT).

This paper is structured as follows: The Related Work section provides a review of state-of-the-art and relevant research papers dealing with the characterisation of Crowdsourcing and Wisdom of the Crowd. This is followed by the Methodology section, which describes the methodology used in this research paper for choosing the indices to facilitate the quantitative examination of the four social location-based services, accompanied with a general introduction of the four services. Identification of Terms section analyses the eight indices with respect to the four services to provide a clear identification and quantitative evaluation to the two terms. The results of the analysis are presented in the Assessment and Result Analysis section, which is followed by the Conclusions section, presenting the discussion of the results and conclusions.

Related work

There exists research aiming to investigate and examine the term Crowdsourcing within the scope of its implementation. An effort to characterise the term was made in Hudson-Smith, Batty, Crooks and Milton (2009), describing Crowdsourcing by using principles, concepts and ideas stemmed from the term Wisdom of the Crowd. The authors presented examples regarding new approaches and systems of collecting, mapping and sharing geocoded data, by analysing neogeography through online mapping tools, such as GMapCreator and MapTube. This resulted in a definition of Crowdsourcing, which, according to the authors, resulted in very little difference between the two terms. Alonso and Lease 2011, also aimed at characterising Crowdsourcing through the term Wisdom of the Crowd. This was made through the presentation of services such as Mechanical Turk and Crowdflower. The authors introduce the motivation for volunteers to contribute, and explain the advantages and disadvantages of using the crowd for certain processes and applications. Still, no clear differentiation was made in this paper, where the authors basically relied only on Wisdom of the Crowd.

Bihr (2010), who made an effort to compare between the two terms, describes the general similarities, as well as the differences, that exist between the two, through the presentation of several examples. Perhaps, one of the authors more significant claims is that: 'Crowdsourcing can enable the Wisdom of the Crowd (but does not have to)', though it is not mandatory.

The current research tries to emphasise finding clear and comprehensive definitions of the term Crowdsourcing. Still, the review of these has made it clear, and to some extent surprising, that there is not a single and clear definition of Crowdsourcing, despite the attempts made to search for such a definition. For example, Schenk and Guittard (2011) made an effort to compare the term Crowdsourcing via the use of several similar concepts (such as Open Innovation, User Innovation and Free-Libre-Open Source software), highlighting existing dissimilarities. In addition, the authors focused on the definition of typology with respect to Crowdsourcing, with the analysis of two different views: (1) the integration of the crowd information, (2) the selection of a single answer from the crowd information. Accordingly, tasks that can be crowdsourced were introduced and divided into three main groups: (1) simple tasks (e.g. data collection), (2) complex tasks (e.g. problem solving), and (3) creative tasks (e.g. design). Finally, benefits (such as cost, quality, variability of data, motivations and incentives) and drawbacks (such as lack of contributors, clarity of task) of Crowdsourcing were presented.

Estelles-Arolas and Gonzalez-Ladron-de-Guevara (2012) presented with an integrated Crowdsourcing definition by trying to find as extensive a definition as possible that will cover as many Crowdsourcing processes and applications as possible (see the definition in the Crowdsourcing section). The authors' definition is a result of analysing 40 available definitions, which resulted with eight principal characteristics: (1) the defined crowd, (2) the task's clear goal, (3) clear motivations and incentives obtained, (4) identifying the potential crowd, (5) defined compensation (by crowdsourcer), (6) the type of process, (7) making public the call to participate, and (8) usage of medium. These eight characteristics were analysed through 11 known projects, such as Wikipedia, YouTube and more. According to these characteristics, the authors concluded that Wikipedia and YouTube, for example, are ambiguous when it comes to a clear Crowdsourcing definition. This is because characteristics 4, 5 and 7 do not exist in Wikipedia, while in YouTube, only characteristics 1 and 8 exist.

Another attempt was made by Zhao and Zhu (2014), who overviewed the current status of Crowdsourcing research, trying to present with a detailed examination of the visible and invisible substrate of Crowdsourcing research, and pointing to possible future research directions. Moreover, the paper aimed at distinguishing between Crowdsourcing and three related terms: Open Innovation, Outsourcing and Open Source. In addition, the authors presented a framework conceptualisation of Crowdsourcing based on four questions: (1) who is performing the task? (2) why do they perform it? (3) how is the task being performed? and (4) what is the relation to the ownership, and what is being accomplished?

Through the above examples, it is clear that the term Crowdsourcing does not have a comprehensive definition and strict characterisations. On the other hand, the term Wisdom of the Crowd does have a clear definition and processes, as presented earlier in the Wisdom of the Crowd section. However, there is no current effort found that aimed at analysing the term Wisdom of the Crowd with respect to new projects and applications introduced from new technological developments and current trends. Moreover, no research was found that tried to define – and perhaps, differentiate – the two terms, with the focus on the geospatial scientific discipline and geoservices.

Methodology

Following the review in the Related Work section, it is evident that Crowdsourcing and Wisdom of the Crowd are often terminologically intertwined and indefinite. This is perhaps because of the fact that the use of these terms is common and widespread in a diversity of fields

and disciplines – which are also very dynamic and changing, or because they are still not established enough, and thus continue to modify and change (mainly because of new technological developments). Consequently, this research paper chooses to make this distinguish via the comparison and evaluation of the two terms with respect to four popular social location-based services, which are characterised by the use of a crowd to provide the data. The services encompass processes having a geographic and geospatial nature and characterisation: OSM, Ushahidi, Waze and Moovit. All four have tens of millions of users worldwide, offering location-based services, in which volunteers are the fundamental core for creating these services via the data and information they collect and share.

Following the research presented in the Related Work section, with the examination of uses and definitions that appear in this context, it was decided that the evaluation and examination will be carried out with respect to the following eight indices, which characterise the processes of using the crowd. The evaluation will be carried out quantitatively, where each index will be analysed with respect to the two terms (i.e. knowledge paradigms) and then with respect to the social services analysed. The eight indices are: (1) diversity, (2) decentralisation, (3) independency, (4) aggregation, (5) knowledge, (6) activity, (7) privacy, and (8) exploitation (misuse).

The first four are explained in the Wisdom of the Crowd section mentioned above, while the last four are: knowledge – the previous familiarity or understanding volunteers are assumed to have to supply the data and information; activity – the measure of volunteers functionality, i.e. how diligent and actively contributing the volunteer should be, to initiate a beneficial task or mission; privacy – how the data and information collected from volunteers, and on volunteers, are used (and to what extent), and what options do volunteers have with regard to these issues; and exploitation – the deliberate misuse of the service and the damage that is caused in the case of fake or misleading data and information being input by the volunteers. Comprehensive explanations of every index with respect to the four services are given in the next section.

An overview of the four social location-based services is given here. This will help in understanding the background of the proposed quantitative analysis that is carried out in the Identification of Terms section.

OSM

The term 'OSM' refers to a collaborative social online project and an open-source editable vector map of the world, created and updated by volunteers (some authoritative data are contributed by mapping agencies, but this is to a much lesser extent). The project[1] aims to create a comprehensive online map of the world that is editable and free to use, especially in countries where geographic information is expensive or unreachable (i.e. does not exist) for individuals and small organisations, and also frequently changed (Haklay *et al*. 2008) (Fig. 1). Users can view and edit the underlying OSM data, upload GPX files (GPS traces) from hand-held GPS units (and also smartphones and tablets equipped with GPS chips), or correct errors in local areas according to satellite imagery and out-of-copyright maps, which are integrated into the mapping interface (Haklay and Weber 2008) (Fig. 2).

The OSM project is becoming more widely used and available to users worldwide, and nowadays, even competes with other alternative online mapping projects and platforms, such as the commercial Google Maps, Yahoo! Maps or Bing Maps. This is mainly because of the use of volunteers and the frequency of updates (event based), which presents increasing and rapid improvement of quality of data in terms of accuracy and completeness.

Waze

Waze is a community-and-GPS-based traffic and geographical navigation service. Drivers living and driving in the same area can share real-time traffic and road information with others. The data, collected automatically from the driver just by driving and navigating the area with an open Waze app, are based on the car direction, location and speed, all sent to Waze servers for further analysis and the dissemination to other users (Fig. 3).

Besides the automatic processing and dissemination of information, users can also actively report traffic jams, accidents, road dangers, fuel stations with the lowest fuel price along their route, speed and police control, hazards on the road, etc. Moreover, there also exists online map editor where a user can add new roads or update existing ones, can add landmarks, house numbers, and many more. The collected data are aggregated (a sort of validation, making sure that several events of the same feature exist), and then provided to the user community as alerts, traffic flow updates, etc (Fig. 4).

Moovit

Moovit is a real-time cooperative, community-driven service for public transportation and transit navigation based on GPS. The real-time data are collected from public transportation companies as well as from users travelling by bus, metro train, etc. GPS location-based information is sent to Moovit servers, which aggregate the data and information delivered, to provide other users of 'the community' with accurate and up-to-date transport times, suggested and planned routes, time schedules, and so on.

Users can get a live map that shows nearby transit stops, lines stopping there and arrival times of the next public transport vehicles. Users can plan their trip based on current available information, by choosing the most efficient and convenient option suggested by Moovit. Step-by-step directions are provided to the users (changes, stops), including walking segments, transit transfers, updates and alerts, until the user arrives to his/her desired destination. Moreover, the user can receive dynamic estimated time of arrival (ETA) to the desirable destination and can actively add new information to the system, such as new bus station, crowdedness in a bus or a stop, etc. (Fig. 5) – all to be available and used socially by other users.

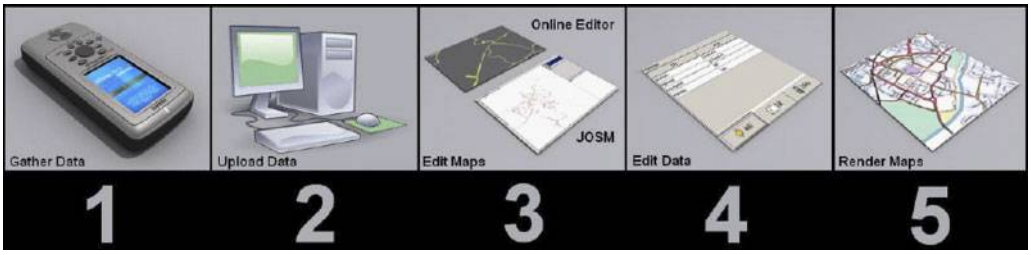
Ushahidi

Ushahidi[2] is a platform used for the creation of crisis maps on the basis of integrating data from multiple sources, such as eyewitness reports, emails, text

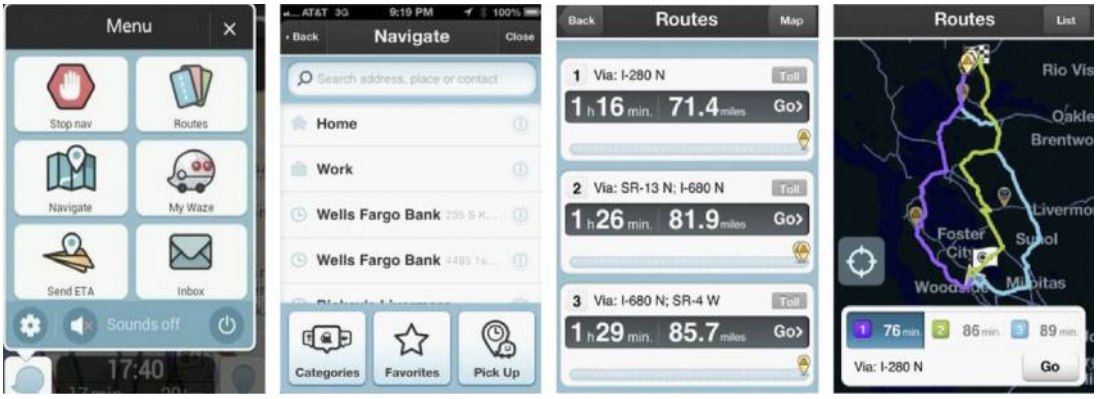
1. OSM might not be considered as a mapping service, since it has the characteristics of a project. However, data and information available in OSM maps serve the public and are used for navigation applications such that it is common for companies and organisations to facilitate these 'services' and make use of it; hence, it is considered as a service for the purpose of the analysis made in this paper.

2. In Swahili: testimony or witness.

1 An example of an online topographic OSM map – Technion area (Haifa, Israel) and its surroundings (*Source*: OpenStreetMap.com)



2 Schematic workflow for creating OSM maps (*Source*: Haklay and Budhathoki 2010)



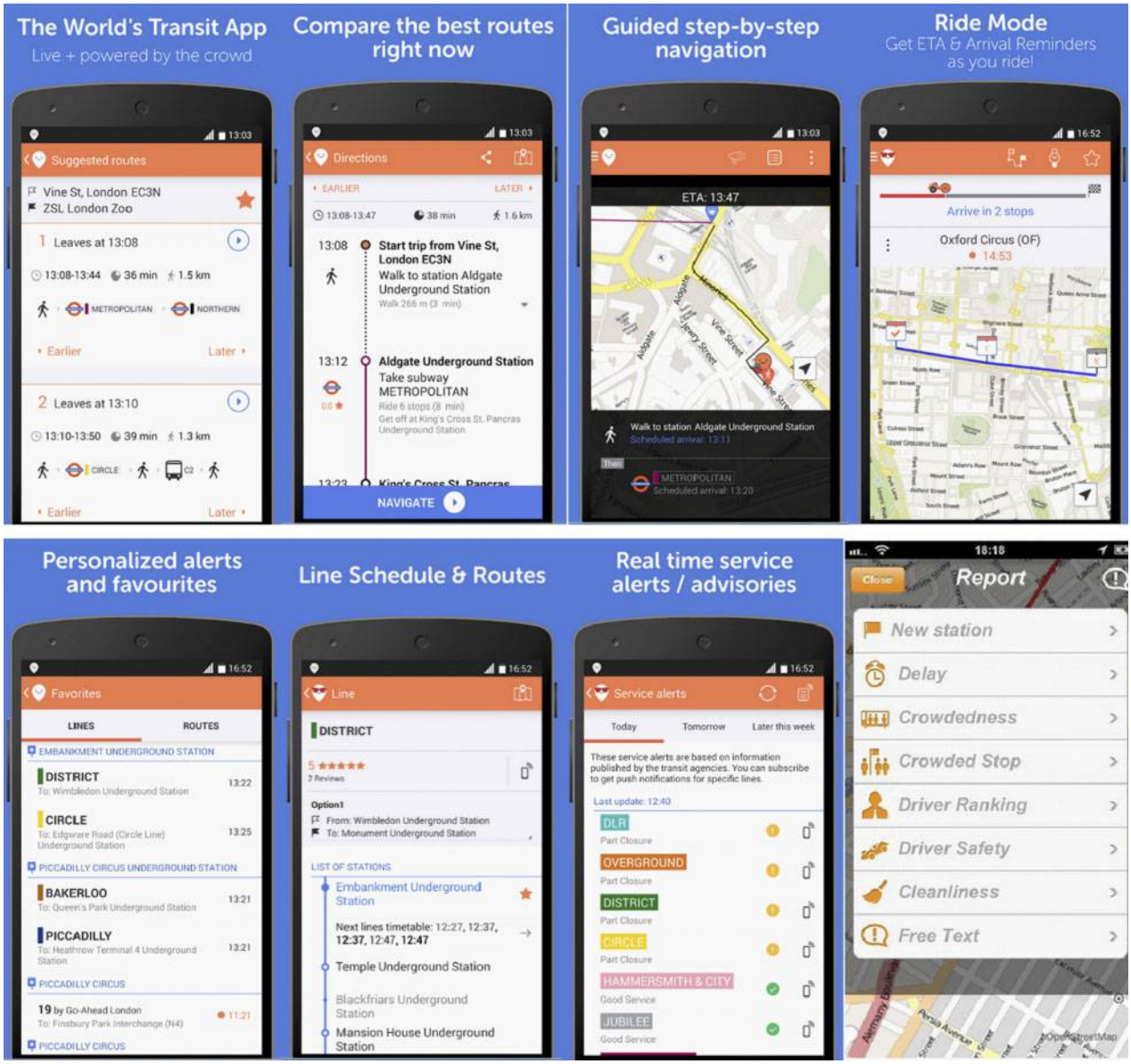
1 Main menu screen; *2* navigation screen; *3* estimated time of arrival (ETA) screen of different optional routes; *4* optional routes display (*Source*: Waze.com)

3 Waze screenshot examples (from left to right)

messages, and pictures, where the location is obtained via GPS-positioning or text message description. These are placed geographically (geo-tagged) on a Google Maps, OSM or Bing map. The platform became very popular and widely acknowledged during recent times of crises, such as the Kenyan post-election violence crisis (2008), the Haitian earthquake (2010), The Japanese tsunami (2011), and the Australian flooding (2011). The Ushahidi platform today is considered to serve as an 'activist mapping' manifest, enabling the fusing of social activism, citizen journalism and geospatial information for public use. Thanks to the rapid growth of the Internet and mobile devices and communication services, which enable real-time and accurate update of

1 Report menu screen; *2* traffic jam report screen; *3* visualisation of a traffic report; *4* closure report screen (*Source*: Waze.com)

4 Waze screenshot examples (from left to right)



5 Moovit screenshots examples, showing different available processes and services (*Source*: Moovitapp.com)

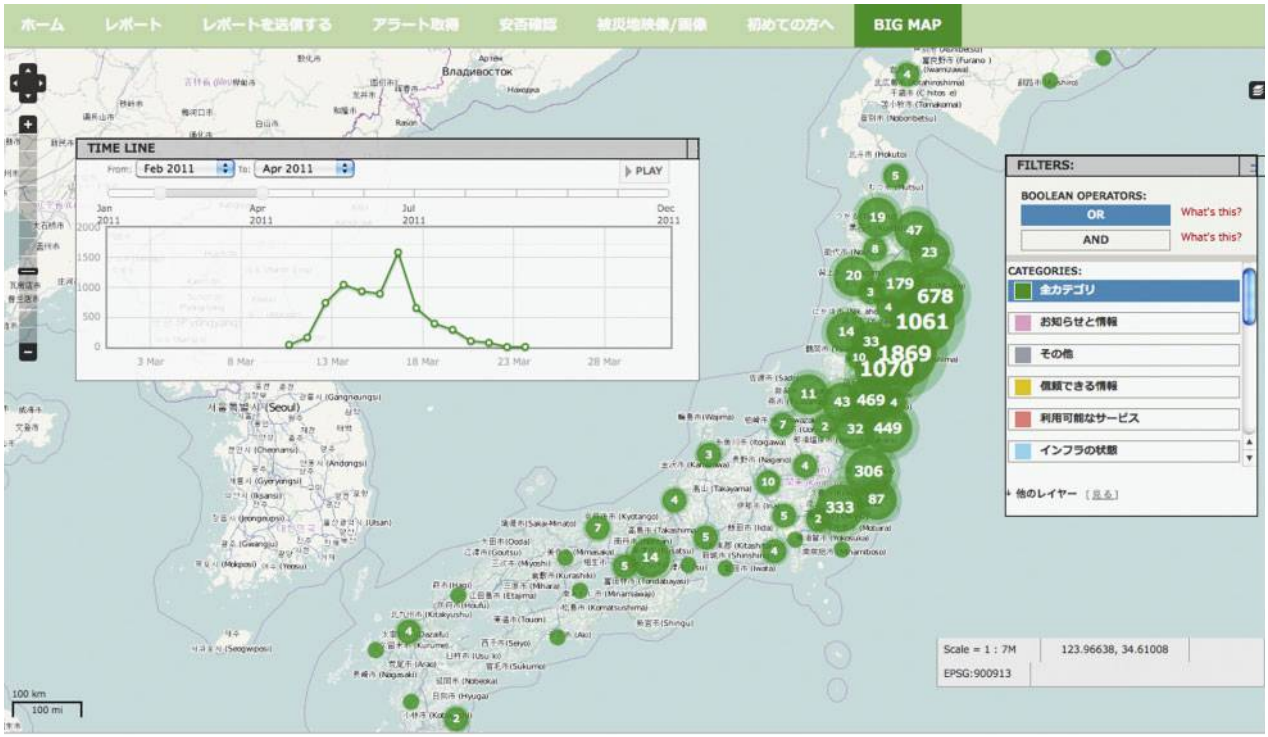
data, Ushahidi now offers various products and services, some of which are tailored to the user demands.

The main mapping tool behind Ushahidi is Crowdmap, which creates interactive maps for visualising location-based data on a map including a timeline. The common data report is translated geographically to a point-position on the map, while an aggregation process of these reports allows the formation of a 'heat-map' showing the current situation with places requiring the most urgent help with an emergency (Fig. 6). The tool collects data and information from the web, mobile phones (SMS, MMS), or Twitter using a form. The most notable advantage is the possibility to conveniently use mobile devices leading to onsite incident reporting (Fig. 7). Crowdmap can be installed and setup on demand, allowing users to set up their own deployments of Ushahidi without having to install it on a web server. Today, the service allows users to insert check-ins (similar to Foursquare), thus enabling an interactive (social) map that is structured location and detailed information. For example, during the Japanese Tsunami, the Crowdmap website was established in <12 h and was active for almost 1 year, receiving close to 13 000 reports dealing with trapped people, dangerous areas that should be avoided, and supplies of food and clean water (Fig. 6) (McDougall, 2012). This tool was found to serve as an informative source, augmenting other authoritative sources used by rescue teams and authorities.

Identification of terms

Diversity

As described in the Introduction section, volunteers participating in a process (task) defined as Wisdom of the Crowd must produce different – and diverse – pieces



6 An example of reports during the Japanese tsunami in 2011, showing the position and aggregation of reports, together with a timeline of number of reports (*Source*: www.sinsau.info)

7 Mobile interface of an Ushahidi report form (left); an example of geotagged Ushahidi reports during the Haitian earthquake (*Source*: Ushahidi Blog)

of data to be consequently transformed into information, usually of a specific element. Basically, this is also the case for a process configured as Crowdsourcing, where the volunteers should contribute diverse data, but here it has a more general and wider purpose (volunteers can be characterised as agents). Diversity encourages a variety of innovative ideas (Surowiecki, 2004), and in the mapping discipline, it helps to retrieve various pieces of information regarding a specific feature or to cover wide topographic areas while increasing the certainty and update-rate of the collected data.

All four social location-based services have been very successful because of the ease and wide variety of geospatial data and information volunteers can contribute. In OSM, volunteers can add buildings, roads, shops, schools and everything needed to complete missing information. In Ushahidi, volunteers can add different types of reports, including free text, required to describe an event that requires handling. Waze users (drivers) can add new roads, place of accidents, police traps, road dangers, or can map a petrol station with the lowest fuel prices. Moovit users can map a route and schedule of the bus ride by leaving the Moovit app open or actively map a new bus station. According to current quality standards and definitions with respect to crowdsourced information retrieved by volunteered geographical data (e.g. Haklay 2010), the existence of a wide range of contributors, i.e. heterogeneity, for these social services contributes to the overall process and task. This manifests itself in terms of the improvement of the geospatial and geometric completeness of information, together with temporal quality of the mapping infrastructure (which in case of Ushahidi the temporal aspect is critical). Consequently, this index is important with different magnitudes to both Crowdsourcing and in Wisdom of the Crowd, and as such is significant to the evaluation of the four social services analysed here.

Decentralization

Decentralisation is strongly correlated to the diversity index, because of the fact that the existence of a similarity factor among contributors influences and reduces the diverseness of a new product or task, i.e. the outcome is biased: '... the more similar the ideas they appreciate will be, and so the set of new products and concepts the rest of us see will be smaller than possible' (Surowiecki, 2004). Moreover, decentralised organisations have the same aspect: 'power does not fully reside in one central location, and many of the important decisions are made by individuals based on their own local and specific knowledge rather than by an omniscient or far seeing planner' (Surowiecki, 2004). Thus, the results derived from the Wisdom of the Crowd process will be more innovative and reliable when they are decentralised. For Crowdsourcing tasks the same advantages of decentralised sources exist, though to a lesser extent. Hence, the two terms should have a relatively high rate of decentralisation aspects within their scope, e.g. projects, services, applications, etc.

If commercial companies have an influence on the collected data, they are considered as users and, as such, can contribute data. For example, a petrol station offering the cheapest fuel can have an effect on the driver's chosen route (Waze), a bus company reporting that all buses are overcrowded and late, except for their own buses, can change bus users' choices (Moovit). Such inputs might change the users' choices that are directly derived from the service they acquire and consequently persuade users not to trust the information they receive from the services. Namely, projects based on 'the crowd' aspire to receive as true and accurate information as possible, and as such decentralisation helps to achieve this, especially with respect to a centralised process. It is assumed that public organisations should maintain objectivity (as in the case of Ushahidi), while private and commercial companies might be biased in favour of their interests. While these concerns exist in relation to major projects that are centralised by nature, it can be assumed that it does not occur here because of the decentralisation factor and the fact that vast numbers of users and contributors exist.

Independency

An independent answer is crucial in Wisdom of the Crowd: 'independence of opinion is both a crucial ingredient in collectively wise decisions and one of the hardest things to keep intact' (Surowiecki 2004). In Crowdsourcing, independent contribution (as in mapping) is important, but still not a crucial aspect, since contributors can be aided by other users (mentors, guide) and thus be affected by contributions made by other contributors – though still having no effect on the final product. Therefore, it can be seen that while independency is essential for Wisdom of the Crowd, it is less crucial with Crowdsourcing.

For example, if a volunteer sees that there is a good mapped area in OSM, possibly he/she will search for another less mapped area to map. In Ushahidi it is assumed that all reports are independent, since it is vital to acquire a reliable and informative status at all times (though one can encourage his or her neighbour to upload a report to amplify a certain danger or problem). Still, while doing the mapping task of the chosen area, he/she should independently map according to his own data and knowledge. In Waze, on the one hand, independent data are important, specifically all the drivers should supply their own driving route regardless of other users and report their own alerts. On the other hand, a driver's route is influenced by all the information gathered from other drivers. The same is true with Moovit, where every bus passenger should supply his own records of bus driving time or a station location. However, the passenger's public transportation route chosen is influenced by other users' experience, thus having a sort of a 'chicken or an egg' effect.

Aggregation

In Crowdsourcing, the volunteers serve as sensors or agents (especially in mapping projects) to provide the needed data (Goodchild, 2007). There is no aggregation during the process for a single piece of data (element, feature).[3] However, Wisdom of the Crowd takes place only if an aggregation process is implemented on the volunteers' contribution. 'If that same group, though, has a means of aggregating all those different opinions, the group's collective solution may well be smarter than

3. The online Miriam-Webster dictionary describes the term aggregation as 'the collecting of units or parts into a mass or whole'. Still, in this paper, we relate to aggregation as presented in section 'Wisdom of the Crowd'.

even the smartest person's solution' (Surowiecki, 2004). Such aggregation is vital for Wisdom of the Crowd.

In OSM, the environment is mapped by users, where the most current update is added to OSM and considered as the final version (while history is archived, but not used on the final map). Thus, there is no aggregation measure when making an OSM map. However, in aggregation is crucial, since authorities will act depending on the magnitude of the existing crises; thus, the more reports that exist – the more urgent handling is required. Similarly, in Waze, to be able to receive the most accurate and valid information regarding the location (and time) of a traffic jam, an aggregation of all drivers' 'reports' is crucial – without this, this service will not exist. In Moovit, to get the most accurate ETA, the data, collected from all users travelling by bus while using Moovit, are aggregated, and the precise time is delivered to the other users waiting for that bus or train.

Hence, aggregation is one of the most prominent indices that differ between the two terms, i.e. Wisdom of the Crowd must have an aggregation measure while Crowdsourcing does not. Moreover, Waze and Moovit services are more and more widespread because of their ability to aggregate data in real time and deliver it to their users as updates.

Knowledge

Another attribute that distinguishes between Wisdom of the Crowd and Crowdsourcing is accessible and obtainable knowledge of the participating crowd, i.e. the previous familiarity, understanding and professionalism volunteers are assumed to have in order to supply the data and information. Wisdom of the Crowd is based on the decision making of every individual that participates in a task (Surowiecki, 2004): 'The crowd also has to be able to distinguish the good solutions from the bad'. However, with Crowdsourcing projects and services, some do not rely on the existing knowledge and professionalism of participants; thus, this attribute is nice to have, but not essential.

For instance, to map a new road in OSM or Waze, the user does not have to know his way around, since he is just following a road. Thus, prior knowledge in mapping or about the vicinity is not required. The result will be exactly the same as that from a user who has good knowledge of the area. In addition, as with Moovit, one can receive a route of a bus without any prior knowledge. The assumption is that there is no requirement from users to have any knowledge – all is done automatically by tailored protocols and algorithms in the system. Nevertheless, there are mapping assignments, which can be regarded as active, that do require a certain amount of human intelligence. For example in OSM, mapping the geometry of a building requires some preliminary knowledge of mapping, while inserting the numbering of the building requires knowledge of the surroundings. In Ushahidi, though locations are provided using GPS receivers, much of the data provided by users are text, thus certain knowledge is presumed to be possessed by the user, mainly with respect to the vicinity the user is in, and the problem the user is reporting. Incorrect input might damage a reliable report and will result in authorities going to the wrong place or arriving unequipped.

Moreover, some researchers have tried to understand contributor patterns, such as the motivation and demographic factors of OSM volunteers. One of these factors is educational background, and it was found that $\sim$70% of OSM contributors had a college, university or higher education degree (Neis and Zielstra, 2014). This may seem like a surprising result, since it shows that although a participant does not have to have any mapping knowledge, most of the contributors are well educated, thus perhaps pointing to the fact that with Crowdsourced mapping prior knowledge is an advantage. Another explanation is the fact that people who are familiarised with OSM have heard about it during their studies or work. It is the belief of the authors that if the same study was made with Waze and Moovit contributors, the results would probably be different; these projects' starting point is that anybody can use these services and benefit from them.

Activity

This index explains whether a positive action of a person is required, or perhaps a passive one is sufficient; such as whether the contributor needs to be actively and specifically involved in the task of contributing data to facilitate a reliable process. In Wisdom of the Crowd, the volunteer must be active, e.g. answer a question, solve a problem, do a task, etc. Moreover, the aggregation measure should be used to get the best result (Lorenz, Rauhut, Schweitzer and Helbing, 2011). Therefore, an organisation that identified tasks or asked questions, needs to choose an appropriate aggregation measure, which is undoubtedly a positive action. In Crowdsourcing, the understanding of this index is more complex and thus will be explained through examples.

For mapping a new road in OSM, a positive action is needed: map via the use of the OSM interface website, or to physically go out with a hand-held GPS unit and measure. The same is in Ushahidi, which is known as 'activist mapping', users must be active to contribute with informative data and information; without this, the service will not exist. In Waze, users are involved in contributing their data only because they use the service (an active Waze app). Hence, no special effort is required by the driver to contribute to Waze thanks to the algorithmic concepts embedded in Waze. In addition, with open Moovit app one can receive a route of a bus or ETA without human intervention, as long as the passenger uses the service. Such that Moovit requires in-bus data for receiving dynamic ETA to the desirable destination. Accordingly, Crowdsourcing does not have to be active when there are tasks that do not require the contributor to make an active intervention, as in following a route; on the other hand, there are tasks, such as problem solving, that require only one answer to be chosen and as such require activity. However, all the mapping projects that are based on crowd information try to do their best to get as much active updating from the volunteers as possible, as this information has a great influence on a quality of the volunteered map. Still, the aspiration is that such updates will be done automatically, without too much being demand from and bothering of the user.

Privacy

Privacy issues are related to the way the volunteers' contributed information is used, and what a user needs to share about himself/herself (that is not directly related to the process), and to what extent. In Wisdom of the Crowd, a question or a task is given to a group of people while the group details has no weight or meaning (affect) on the

result, such that the information can be given by anonymous volunteers. In Crowdsourcing, there are two main types of tasks: (1) a task designated for a specific group of people with the aim of receiving a cheap and fast service, instead of an expensive professional service that often requires a long waiting time and (2) a task or a specific question for which a single solution is required, and as such it is directed to the general public in search of this solution. The need for the user details exists in both cases, where the outcome or the work is mostly a financial one and compensation or an incentive is at hand. However, if the tasks are done voluntarily, then there is no need for user information, and privacy can be maintained.

Owing to the fact that this research work is focused on voluntary services, this index score will be given without regard to tasks for which any compensation is given, since a Crowdsourcing process originally does not require an invasion of user privacy.

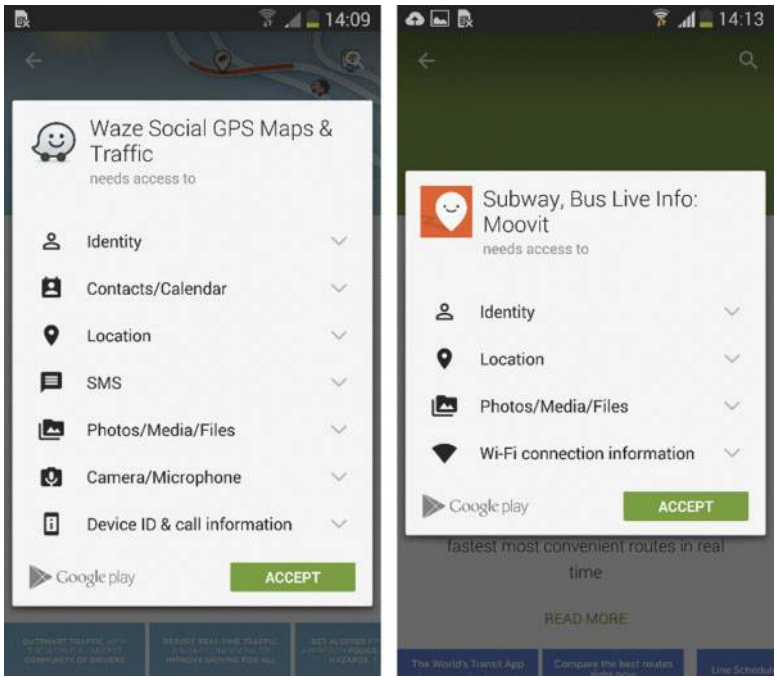
Consequently, privacy issues were inspected with respect to the four social services. To start editing in OSM the user is required to log in delivering an email address only and creating a password. In Moovit, to be able to download the application (before initial use), the user is required to sign an agreement that allows the service to access certain personal details, such as location, photos, WiFi connection and general personal information (identity) (Fig. 8). Waze requires even more accessibility to personal details to download the application, namely, location, personal photos, SMS, contacts and calendar data, identity, etc (Fig. 8). It is believed that since these services are focused on being social, the requirement for additional – and some may say – irrelevant access to personal data is made (as with Waze, their purchase by Google has intensified this requirement). Still, in Waze a user can stay anonymous and stay a temporary user and use the service without the need to create a user account and register (however, these are required to be able to edit a map, to advance in rank, etc.). Ushahidi is similar to OSM, in the way that no specific information about the user is required to contribute data. Still, if a user wants the authorities to know that he/she is experiencing a danger and wants them to know of his state and position, he/she will have to deliver more personal data (as shown in Fig. 7).

Still, a problem exists with the implicit information that can be retrieved and used by such services. Although a user does not have to 'identify' himself/herself as in Waze and Moovit, certain data are collected, while specific spatio-temporal data mining processes can extract information about patterns and behaviours of that user. Thus, information that is not explicitly entered can be retrieved and used (home address, place of work, etc.). This issue does not exist in OSM and Ushahidi.

Exploitation

The last index deals with the intended damage and harm an exploitation of social services may cause. In Wisdom of the Crowd, one cannot easily cause significant damage, because of the fact that the end result from all the available results of volunteers is obtained via an aggregation process, such that, in order to bias the results significantly, many users are required to operate together in a malicious way. However, in Crowdsourcing, generally a single input or solution is required (or a few responses). Therefore, the outcome can be easily misled if wrong answer(s) are given (certain processes can be incorporated to identify and filter bad or wrong answers, though this is the case where prior knowledge regarding the input and outcome is known).

A simple test showed that in its current state, OSM can be easily mapped incorrectly in most areas with numerous precision errors, such as in the case of buildings, roads, etc. These errors can be intentional or unintentional. If no active users contribute concurrently to a specific



8 Screenshots showing download requirements regarding user's privacy information (Waze, left and Moovit, right) (*Source*: Google Play)

area, these errors cannot be detected and filtered out and thus might exist in the map for a long period. In contrast, if there are many volunteers in an area that is familiar and is being mapped regularly, errors and inaccuracy can be detected and fixed quickly. A similar assumption is relevant also to Ushahidi, perhaps of greater magnitude, since it incorporates an aggregation process: the more incorrect and erroneous reports, the more inferior the updated map will be. However, since the Ushahidi serves as a social activism and citizen journalism tool and service, it could be assumed that malevolent exploitation of Ushahidi is rarely made, though again possible.

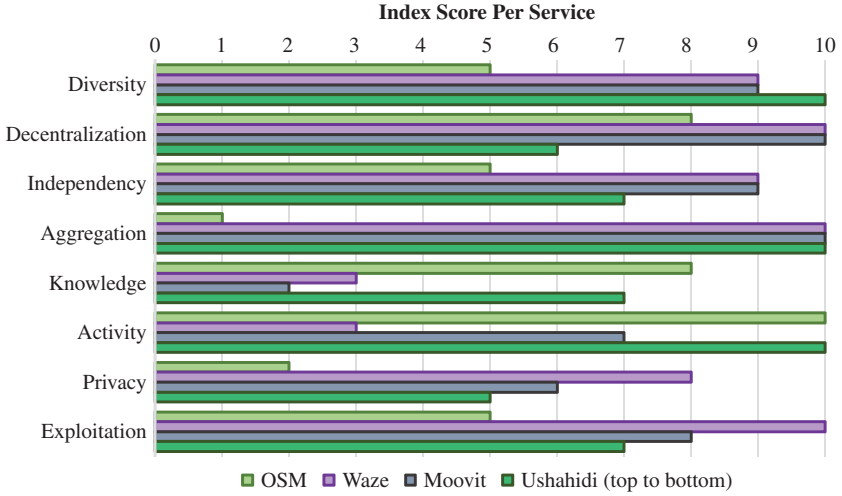
Sinai, Partush, Yadid and Yahav (2014) demonstrated a Sybil attack on Waze service, by creating a large number of fictitious identities, while using them to simulate traffic jams, and thus influencing the route presented to other users. Such exploitation of a social navigation service not only has financial implications, but also affects driver's safety. It should be noted that one cannot identify such an attack easily, since the number of fictitious users might be large, and also the fictitious user's behaviour is hard to be identified automatically.

In Moovit, initial information is based on the schedules of buses, trains, etc., supplied by the companies themselves. Apparently, one might think that this information cannot be damaged, since there exists a reliable reference, the official information. However, since transportation and transfer are dynamic and change frequently, schedules sometime become irrelevant (traffic jams, missing busses, etc.), and this is the main reason and motivation behind Moovit, to use the users to retrieve actual relevant information. The reliability of information received from the service is increasing as the number of users (volunteers) grows. Consequently, a Sybil attack similar to the one demonstrated on Waze could theoretically harm the Moovit service.

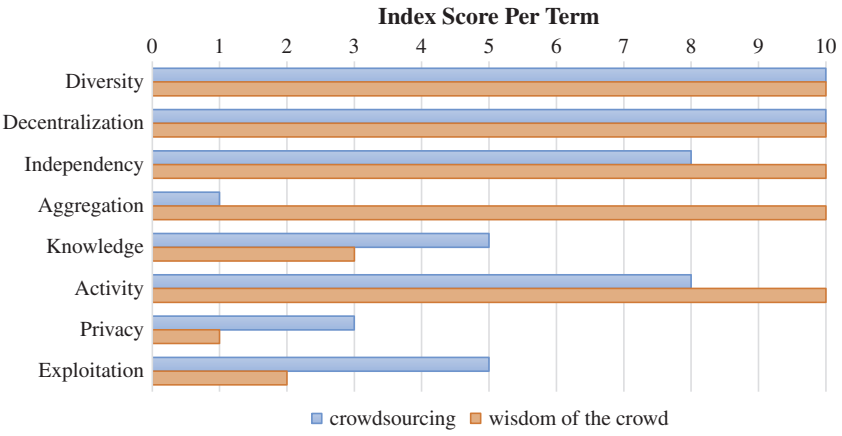
Assessment and result analysis

A system of scores was devised based on the above-mentioned explanations and concepts, on the scale (weight) of 1–10 based on each index. A score of 1 is the lowest, i.e. insignificant, unnecessary and irrelevant for the service, and 10 is the highest, i.e. very significant, necessary and relevant for the service. Figure 9 depicts the scores of all indexes for all the social location-based services analysed.

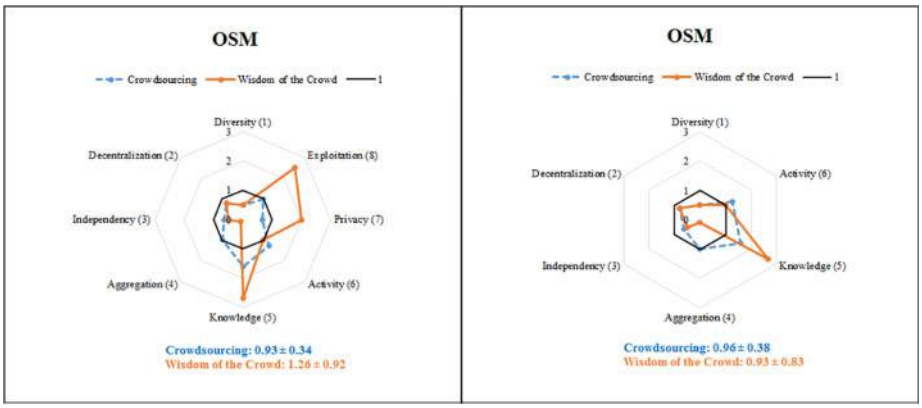
A similar grade (weight) system was given to each index with respect to the two terms, how significant or influential the index is to the term. Figure 10 depicts the grades given to Crowdsourcing and Wisdom of the Crowd.



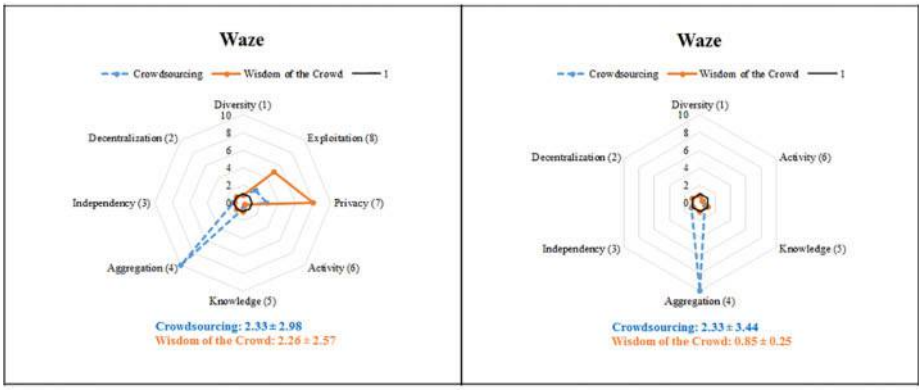
9 Colour bars represent the four services with respect to the eight indices. The grades are in 1–10 scale, where 1 represents the lowest influence of the index on the service and 10 the highest influence



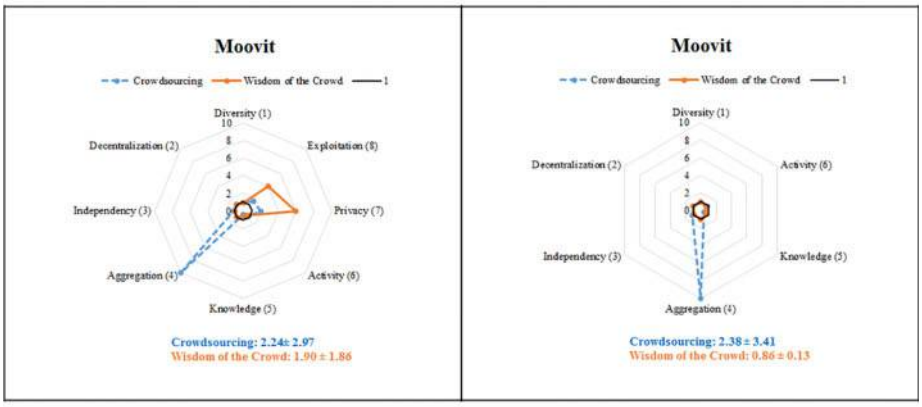
10 Colour bars represent the two terms with respect to the eight indices. The numbers are in 1–10 scale



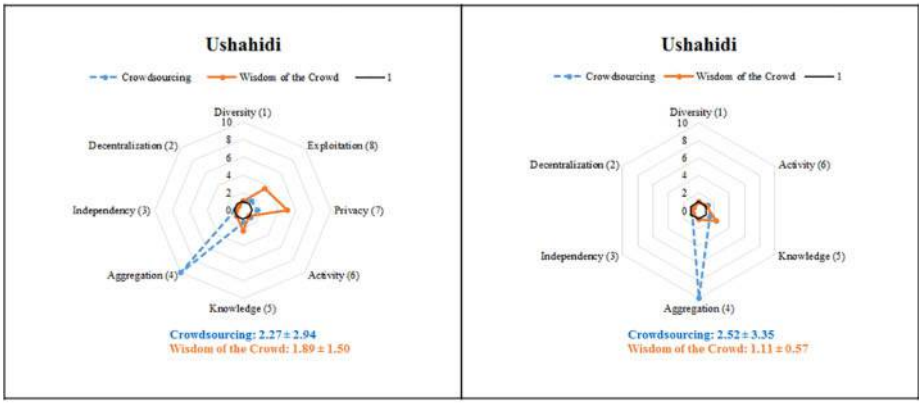
(a)



(b)



(c)



(d)

11 ***a–d*** **Radar representation of all services with respect to normalised scores of all indices: Crowdsourcing (dashed blue) and Wisdom of the Crowd (orange), with overall mean and average value, respectively (bottom). Left-hand side with respect to all eight indices and right-hand side with respect to the first six indices. Mean score of 1 and SD score of 0 represent absolute match to the term**

To analyse all, a quantitative formulisation was modelled defining the four services either as Crowdsourcing or as Wisdom of the Crowd. This was carried out in two stages:

1. Dividing the score of Fig. 9 by the score of Fig. 10, and thus obtaining a normalised score of the index for the social service.
2. Calculating the average normalised score and the standard deviation (SD) for each of the services for both terms.

Figure 11 depicts the results of this formulisation, where after normalisation, a score of 1 means that a service is highly correlated to the term representing it (either Crowdsourcing, blue, or Wisdom of the Crowd, orange). Score with a value that is <1 or >1 means that the service does not reflect in full either of the terms; the farthest the value from 1 is, the lower the correlation with that term. Also, because of a preliminary examination of values in Figs. 9 and 10 showing that the last two indices, namely Privacy and Exploitation, had for most services a reverse effect on the results (no correlation, i.e. high values in Fig. 9 as opposed to relatively small values in Fig. 10), Figure 11 presents the values obtained twice: first, as derived from all eight indices, and second, as derived by only from the first six indices.

Figure 11*a* clearly depicts that the average score of OSM for all analyses is close to 1, whereas for the radar representation for Crowdsourcing is much more compact with most values around 1 (black line). Even with the indices 'Privacy and Exploitation' for Crowdsourcing, the value is 0.93 having a small SD value of 0.34, implying that the service was highly correlated with this paradigm when compared to Wisdom of the Crowd having a value of 1.26 ± 0.92. After the last two indices were removed, the SD value representing Crowdsourcing remains smaller than the one of Wisdom of the Crowd: 0.38–0.83, respectively, with the overall value (0.96) that is closest to 1 received from all the analyses made.

When Waze (Fig. 11*b*) and Moovit (Fig. 11*c*) were analysed with respect to all eight indices, no clear conclusion could be made. All values were ~ 2, with very high SD value that is sometimes >3. One of the reasons for this is that these services aim to be social, and hence private information of users is collected. When Privacy and Exploitation were removed from the analysis, both services showed a clearer resolution, having values closer to 1 with respect to Wisdom of the Crowd, with a relatively small SD value: 0.85 ± 0.25 for Waze, and 0.86 ± 0.13 for Moovit.

Ushahidi (Fig. 11*d*) also showed ambiguous results. When all indices were analyzed, no conclusion could be obtained: values about 2 for both terms having high SD values. When the Privacy and Exploitation were omitted from the analysis, Ushahidi still could not be characterised as Crowdsourcing as anticipated; surprisingly, according to the definitions made in this paper, its Wisdom of the Crowd value was close to 1: 1.11 ± 0.57. This is surprising since Ushahidi is mostly associated as a Crowdsourcing service,[4] though this analysis showed the opposite, that it is more characterised as Wisdom of the Crowd. This is mainly the result of an aggregation process that is handled in Ushahidi.

4. https://www.udemy.com/crowdsource-mapping/.

Conclusion

Volunteered geographic information paradigms related to the terms Crowdsourcing and Wisdom of the Crowd were investigated and analysed with respect to four social location-based services. Since both terms are tightly inter-related and do not have a clear and uniform definition, when social location-based services are at hand, a new measuring analysis system was needed and hence developed and modelled for this research paper. The model advised was based on quantitative measures and grades that are based on eight different indexes that characterise such paradigms. This model was then evaluated on four different commonly used social location-based services. Analysis showed that (as assumed) OSM is strongly correlated with Crowdsourcing service, even when privacy and exploitation measures were inserted into the quantitative analysis. Ushahidi, on the other hand, although considered as a Crowdsourcing service, showed the characteristics of a Wisdom of the Crowd, with grade values corresponding to those of this paradigm. In contrast, based on the definitions made in this paper, Waze and Moovit showed less characteristics of Crowdsourcing. After omitting Privacy and Exploitation from the analysis, both were found to be correlated with Wisdom of the Crowd, perhaps mainly because of the fact that their core service is based on an aggregation process; without this, such service could not exist and hence could not serve with the adequate and expected service. Still, when the last two indices, namely Privacy and Exploitation, were incorporated into the analysis, the overall value of both could not be correlated to either of the paradigms. This is because of the fact that the service itself requires certain information that could be regarded as private, mainly because both aspire to be social and offer the experience of a social network (and as such do not 'follow' the Wisdom of the Crowd paradigm). Consequently, these data can also be used maliciously for other purposes (either by the service provider itself or by a third party that might hack the system).

Furthermore, the analysis showed that a process characterised as Crowdsourcing could be transformed to be a Wisdom of the Crowd one. This occurs when volunteers continue to update and insert data to services while embedding a certain degree of aggregation (as in Ushahidi). Still, in the case that volunteers' answers and solutions are collected, and only one, or a relatively small number, is chosen, this is close to being a complete Crowdsourcing service, and in this case, the aggregation process is not implemented. Future experiments with more indices and services could provide better clarification and comprehension of the two terms, in terms of characterisation and workflow, i.e. the related processes they encompass. However, because of rapid technological developments and services available, such a clear definition is perhaps hard to accomplish, since it seems that both paradigms are in principle flexible and dynamic, which also change because of current social trends. Moreover, the services themselves might not conform to the terms, rubrics and characteristics, since they themselves continue to evolve, adding continuously new features, attributes, and possibilities.

References

Alonso, O. and Lease, M. 2011. Crowdsourcing 101: putting the WSDM of crowds to work for you. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM), Hong Kong, China*, Available at: http://ir.ischool.utexas.edu/wsdm2011_tutorial.pdf. [Accessed 2 October 2014], pp. 1–2.

Berners-Lee, T., Cailliau, R., Groff, J. and Pollermann, B. 1992. World-wide web: the information universe. *Internet Research*, 2(1), pp. 52–58.

Bihr, P. 2010. Crowdsourcing 101: crowdsourcing vs wisdom of the crowd slides. Available at: http://www.slideshare.net/thewavingcat/101-crowdsourcing-vs-wisdom-of-the-crowd. [Accessed 2 October 2014].

Bowman, C. M., Danzig, P. B., Hardy, D. R., Manber, U. and Schwartz, M. F. 1994. *Harvest: a scalable, customizable discovery and access system*. Technical rept., Colorado Univ.: Boulder Dept of Computer Science.

Brabham, D. C. 2008. Crowdsourcing as a model for problem solving: an introduction and cases. *Convergence: The International Journal of Research into New Media Technologies*, 14(1), pp. 75–90.

Estelles-Arolas, E. and Gonzalez-Ladron-de-Guevara, F. 2012. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), pp. 189–200.

Goodchild, M. 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), pp. 211–221.

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

Haklay, M. and Budhathoki, N. R. 2010. OpenStreetMap – overview and motivational factors horizon infrastructure challenge theme day. University of Nottingham, UK.

Haklay, M., Singleton, A. and Parker, C. 2008. Web mapping 2.0: the neogeography of the geoweb. *Geography Compass*, 2(6), pp. 2011–2039.

Haklay, M. and Weber, P. 2008. OpenStreetMap: user-generated street maps. *IEEE Pervasive Computing*, 7(4), pp. 12–18.

Howe, J. 2006a. Crowdsourcing: a definition (weblog, 2 June). Available at: <http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html>. [Accessed 2 July 2014].

Howe, J. 2006b. The rise of crowdsourcing. *Wired Magazine*, 6(14): p. 183.

Huberman, B. A., Romero, D. M. and Wu, F. 2009. Crowdsourcing, attention and productivity. *Journal of Information Science*, 35(6), pp. 758–765.

Hudson-Smith, A., Batty, M., Crooks, A. and Milton, R. 2009. Mapping for the masses: accessing web 2.0 through crowdsourcing. *Social Science Computer Review*, 27(4), pp. 524–538.

Lorenz, J., Rauhut, H., Schweitzer, F. and Helbing, D. 2011. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences USA*, 108(2), pp. 9020–9025.

McDougall, K. 2012. An assessment of the contribution of volunteered geographic information during recent natural disasters. EPrints, University of Southern Queensland, Australia, pp. 201–214.

Neis, P. and Zielstra, D. 2014. Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap. *Future Internet*, 6(1), pp. 76–106.

Schenk, E. and Guittard, C. 2011. Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics*, 7(1), pp. 93–107.

Sinai, M. B., Partush, N., Yadid, S. and Yahav, E. 2014. Exploiting social navigation, arXiv, Ithaca, NY. [arXiv:1410.0151].

Stranders, R., Ramchurn, S. D., Shi, B., Jennings, N. R., 2011. CollabMap: Augmenting maps using the wisdom of crowds. In *Proceedings of the AAAI Workshop on Human Computation*, San Francisco, CA, USA, 8 August, pp. 111–112.

Surowiecki, J. 2004. *The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*, Doubleday books, New York.

Vukovic, M. M. 2009. Crowdsourcing for enterprises, SERVICES '09 *Proceedings of the 2009 Congress on Services – I*, IEEE Computer Society, Washington, DC, USA, pp. 686–692.

Wu, G., Harrigan, M. and Cunningham, P. 2011. Characterizing Wikipedia pages using edit network motif profiles. In *Proceedings of the 3rd international workshop on Search and mining user-generated contents*, October, pp. 45–52, ACM.

Zhao, Y. and Zhu, Q. 2014. Evaluation on crowdsourcing research: current status and future direction. *Information Systems Frontiers*, 6(3), pp. 417–434.