An exploration of future patterns of the contributions to OpenStreetMap and development of a Contribution Index

Abstract

OpenStreetMap (OSM) represents one of the most well-known examples of a collaborative mapping project. Major research efforts have so far dealt with data quality analysis but the modality of OSM's evolution across space and time has been barely noted. This study aims to analyze spatio-temporal patterns of contributions in OSM by proposing a contribution index (CI) in order to investigate the dynamism of OSM. The CI is based on a per cell analysis of the node quantity, interactivity, semantics, and attractivity (the ability to attract contributors). Additionally this research explores whether OSM has been constantly attracting new users and contributions or if OSM has experienced a decline in its ability to attract continued contributions. Using the Stuttgart region of Germany as a case study the empirical findings of the CI over time confirm that since 2007, OSM has been constantly attracting new users, who create new features, edit the existing spatial objects, and enrich them with attributes. This rate has been dramatically growing since 2011. The utilization of a Cellular Automata-Markov (CA-Markov) model provides evidence that by the end of 2016 and 2020, the rise of CI will spread out over the study area and only a few cells without OSM features will remain.

Keywords: OpenStreetMap, spatio-temporal analysis, collaborative contributing, Cellular Automata-Markov model

1. Introduction

Due to the interest, motivation, and efforts of interested volunteers, OpenStreetMap (OSM) has become an alternative source of geodata in both online and offline applications (Mondzech & Sester, 2011; Neis *et al*., 2013). While OSM received only minor attention in the first few years after its launch it now receives a substantial amount of contributions from across the world. 20 million nodes were provided by 100,000 users until 2008, rising to over 2.5 billion nodes provided by almost 1.7 million users by August 2014. This revolutionary data-gathering process continues to rise due to the rapid and wide penetration of smartphones, GPS-enabled devices, and the general awareness of citizen science projects among the population (Georgiadou *et al*., 2013; Mooney *et al*., 2013).

The entire mapping process in OSM is structured in a democratic manner in which anyone can: (a) sign up and join; (b) create/edit/delete spatial objects; (c) access the entire dataset; and, finally, (d) retrieve the entire dataset history free of charge so that every action can be retraced (Ramm *et al*., 2010). Additionally, OSM represents a rising network of volunteers shaping a community which intends to correct the inaccurate or erroneous contributions of others, and thus improve the entire data quality of OSM in a systematic way (Mooney & Corcoran, 2012b; Jokar Arsanjani *et al*., 2014). In doing so, the OSM community is actively managing feedback from participants to enhance the performance of the OSM database, the respective image libraries, editing tools, and other software functionalities. To guarantee data quality, active members of the community closely observe and report destructive and harmful activities through the wiki pages as well as online discussion lists (Ramm *et al*., 2010).

The OSM community has been intensively monitoring data contributions in order to help guide the efforts of all volunteers in the right direction (Goodchild, 2007; Corcoran *et al*., 2013; Jokar Arsanjani *et al*., 2014). The immense benefit of such crowdsourced projects like OSM can be considered from two complementing views. First, from the end-users' perspective, the free availability of geodata which are not restricted through data privacy regulations is essential. Resource intensive cumbersome data acquisition and/or product ordering processes are substantially reduced which, in turn, improves overall access to the data. More importantly, traditional geodata are often not very up-to-date and therefore data uncertainty concerns arise (Pourabdollah *et al*., 2013). Finally, dealing with cross-national studies, researchers are faced with language difficulties, varying object definitions, semantic interoperability, internal infrastructure organization, and different data handling processes, among others. Second, from a data provider's perspective, both commercial and non-profit enterprises have to deal with the costs and time required for data collection and attaching metadata to spatial entities regardless of whether they are extracted from high-resolution images or in-field surveying elaborated with local knowledge of objects (Haklay *et al.*, 2010). However, due to the availability of voluntarily provided geodata the situation has changed radically for both parties. OSM provides large amounts of simply accessible geodata at a high level of confidence in its data quality whilst being provided at low financial and time costs (Haklay, 2010; Hagenauer & Helbich, 2012, Jokar Arsanjani et al., 2015). Despite some volunteers' minor knowledge of mapping and data collection, the gathered information from them comprises new spatial objects and attributes that may never have existed in traditional databases, as empirically proven by Haklay *et al*., (2010), Neis and Zipf (2012), and Neis *et al*., (2013), among other studies. The importance of collecting VGI in developing countries is particularly important. As seen in the response to Haiti's earthquake and the Philippines' Typhoon (Zook *et al*., 2010; Yates & Paquette 2011; Roche *et al.*, 2011) timely up-to-date and geographically complete data coverage is available very quickly.

Although considerable research has been carried out on the topic of OSM data quality issues (Haklay & Weber, 2008; Haklay, 2010; Girres & Touya, 2010 ; Helbich *et al*., 2012; Barron *et al*., 2013; Jokar Arsanjani *et al*., 2013a), less attention (Neis *et al*., 2013; Corcoran *et al*., 2013; Jokar Arsanjani *et al*., 2014) has been paid to the spatio-temporal evolution of OSM. This is of great importance because if the degree of evolution in a specific area is high it is more likely to receive more reliable information. This is because more users are involved in the mapping process and, therefore, increase the control mechanism in the sense of Linus' Law (Haklay *et al*., 2010; Hardy *et al.,* 2012) which exist in volunteered geographic information (VGI) data collection. Linus' Law expects that the more edits contributed by mappers on OSM features the larger the increase in the data quality. Additionally, it is vital for the existing OSM community to know in which direction OSM is headed. Is it failing to maintain people's interest in contributing or is it continuing to attract more contributors and contributions leading to a richer and more accurate dataset.

The evolution of OSM in space and time is highly relevant as it can provide knowledge of how OSM might emerge in the future. It allows estimations to be made about the future data quality for certain areas which is of interest to OSM-dependent applications (e.g., OpenRouteService, OSM-3D, OpenMapSurfer) which this may impact. These estimations can obscure insights into how OSM, as a dynamic human-based system, functions and where and when OSM attracts people to contribute and which spatial features attract people's attention. A few investigations on monitoring the spatio-temporal evolution of OSM network have been carried out. For example, while Neis *et al*. (2012; 2013) consider the amount of nodes, ways, and relations to measure the development of OSM in a simplistic descriptive manner, Corcoran *et al*. (2013) propose two concepts i.e., *exploration* and *densification* for distinguishing between the types of on-going activities in OSM. In contrast to Neis *et al.* (2012; 2013), Jokar Arsanjani *et al*. (2014) project the geometry of contributions into a cellular grid and apply a cellular automata approach to monitor the spatiotemporal evolutionary patterns of OSM in a case-study area in Germany. A more practical and effective approach, which considers other criteria in addition to geometry and quantity of the contributions to OSM, is urgently required to allow for better quantitative and qualitative indications of activities in OSM to be determined.

In the current research, following Jokar Arsanjani *et al*. (2014), collaborative contribution to a project such as OSM is considered as a spatio-temporally explicit continuous and dynamic process. Thus, the OSM contributors are the actors, who are interactively contributing their information to the community. Based on the identified research gaps, the main objective of this paper is therefore to develop a contribution index (CI) for exploring OSM developments so that instead of the abovementioned approaches, an index is used to monitor the patterns of contributions. Additionally, this index is coupled with a CA-Markov approach in order to predict future OSM states over a representative study area. More precisely, in order to leverage the understanding of the spatio-temporal evolution of OSM, future states of the CI are monitored to predict upcoming OSM contributions until 2020 in space and time. The following research questions will be addressed:

- Which parameters besides the quantity of contributions do we need to take into consideration in order to design a CI?
- What does the spatio-temporal evolutionary pattern of the CI in the selected study area actually look like?
- How well does the CA-Markov model perform in predicting the future forms of OSM contributions?
- In which areas are more contributions received? Is there any spatial correspondence between the CI and land cover characteristics?

The remainder of the paper is structured as follows. Section 2 introduces the materials, Section 3 explains the methods used. Next, Section 4 discusses the empirical results and finally, Section 5 highlights major conclusions and outlines recommendations for future research.

2. Materials

2.1 Study area and data

The metropolitan area of Stuttgart, the capital of the Baden-Wuerttemberg state of Germany, and its surrounding areas are chosen as the study area (see Figure 1). The reasons for choosing this area are twofold: firstly, the Stuttgart region has been a dynamic area in receiving a large record of contributions according to the

OSMatrix (Roick *et al*., 2011); secondly, this area consists of a variety of land cover types i.e., artificial surfaces, agricultural areas, and forests.

The data used in this investigation are the OSM features extracted from the OSM planet file in July 2013. The OSM planet file represents every node that has been hitherto contributed and shared in OSM. It must be noted that these nodes represent the configuring nodes of every point, polyline, and polygon feature. The extracted dataset contains a variety of tags including the attributes "*osmtimestamp*", "*osmversion*", "*osmuser*", "*osmuid*", and "*osmid*" of objects. Furthermore, the CORINE land cover map of the study area provided by the European Environment Agency serves as a second data set, representing the latest update of land cover types prepared in 2006 at a 100 m spatial resolution (European Environment Agency, 2013). Land cover features permit us to compute the associations with the CI (e.g., urban areas contain more points of interest and objects rather than agricultural areas).

Figure 1: The geographical extent of the study site and corresponding land cover

3. Methods

As outlined in Figure 2, the workflow consists of two parts. While the first part introduces the CI, the second part is comprised of the CA-Markov model to predict future OSM states.

Figure 2: Schematic representation of the workflow

Subsequent analysis is based on the premise that the whole collaborative contributing process in OSM possesses the properties of a spatio-temporal dynamic phenomenon as it started in time and retains its dynamism across time and space, i.e., it spreads out across space and over time. This process presumably has emerged since 2004 and ever since has evolved spatio-temporally. For instance, once an object is created/edited/deleted, one contribution is made by a user and it is more likely that: (a) he/she comes back and continues his/her contribution and the phenomenon spreads across the network; (b) a new contributor gets involved and starts completing the contribution of the previous contributor(s) by creating/editing/deleting the adjacent objects; or finally, (c) the chances of receiving contributions from nearby cells are higher than from ones farther apart, because the process of contributing is continuous in time and space. Evidently, contributions are given at different rates in each area, so the rate of contributions depends on the quantity of existing objects and number of involved users. This is a function of the resident population apart from global mapping calls for humanitarian aids as seen in Haiti and Philippines. Since the degree of dynamism of this phenomenon differs over time and location the collaborative mapping process is considered as a space-time dependent dynamic phenomenon. This phenomenon can be modeled by means of a CA-Markov model. For the application of this model, the shared nodes as contributions must be transferred to a grid representation to obtain a better abstraction of the data. Additionally, the aggregation to cells reduces the data size considerably. The attributes of cells were combined and several classes for each variable were defined. Accordingly, the subsequently introduced CI is defined to have a better translation of contributions in terms of quantity, given attributes, number of involved users in mapping, and how many times an object has been edited.

3.1 OSM Data Abstraction Using Cellular Grids

In order to use a CA-Markov approach, the shared contributions must be transferred to a grid representation to have an appropriate abstraction of the data. Initially, a quantitative analysis of the amount of contributions is done to determine: (a) how and where the collaborative contributing has emerged and evolved; (b) the rate at which this phenomenon has disseminated; and (c) to investigate how land cover types play a role in receiving more contributed cells. Subsequently, the data are transferred to a grid representation with a spatial resolution of 100 m. This resolution, which is compatible with the CORINE land cover map, is selected to keep the computational tasks feasible while ensuring that the morphological pattern of features are retained. For the aggregation process a location-based join analysis is applied to transfer the attributes of OSM nodes to the cellular grid. Such representations are prepared for six timestamps from 2007 until 2012 (indicating the contribution by the end of each year) with 1-year sequences.

3.2 The Development of a Contribution Index

While previous investigations only measured the degree of activities in OSM by simply counting the number of nodes, roads, users, relations, and attributes separately (e.g., Neis *et al*., 2013), this study developed a CI which holistically quantifies the activities in OSM. The CI is based on the assumption that the amount of contributions per cell i is a function f of some existing measures such as the total number of given nodes (TNN), the mean number of the attributes (MNA), the number of contributing users (NCU), and the mean version number of nodes (MVN). Accordingly, four variables (i.e., “quantity”, “interactivity”, “semantic”, and “attractivity” (the ability to attract contributors)) are derived to categorize the contributions into categories. While quantity counts the number of nodes given in each cell, interactivity averages “*osmversion”* per cell which determines how many times a node has been edited. *Semantic* indicates how well the nodes within each cell are attributed, i.e., how many nodes are given attributes per individual cell on average. Attractivity is based on the number of users that have edited the nodes within each cell.

$$\mathrm{CI}_i = f\,(\mathit{nodes\_count}_i\,,\ \mathit{osmversion}_i\,,\ \mathit{number\ of\ attributes}_i\,,\ \mathit{number\ of\ osmusers}_i)$$

It is assumed that the degree of contributing within a cell is higher if certain conditions are met such as: (a) contributions are given semantic information i.e., the nodes are given attributes, so if the mean number of attributes per cell is 100%, every feature possesses at least an attribute. Cells with values of 0% contain no attributes for the contained features, therefore contributions, which are not given any attributes to describe them, lack of sufficient semantic description; (b) high quantity of contributions i.e., the number of nodes per cell identifies how densely the objects represented; (c) high attractivity per each cell i.e., number of “*osmusers”* attracted per cell determines that how many users have been contributing in each cell, so the more users are involved, the more reliable the contributions within a cell are likely to be; and (d) likewise, high interactivity within each cell i.e., “*osmversion”* indicates how many times each individual node has been edited, so higher "*osmversion*" numbers show that frequent activities have taken place in each particular cell.

In order to calculate the CI on the basis of the aforementioned variables, a characterization of the map patterns are required. A widely used statistic to detect locations of high and low values, among others, is the local G*-statistic (Getis & Ord, 1995). In accordance to Getis and Ord (1995), it is referred to a hot/cold spot as cells with high/low values which are surrounded by cells with high/low values, respectively. In addition to hot and cold spots, two other categories are considered that include cells placed between cold and hot spots and also the cells that have received zero contributions in the four following categories:

- A dead cell (DC) is a cell that no contribution is given to; therefore all variables have no value.
- A barely contributed cell (BCC) is a cell that falls into the category of cold spots, which means it contains the lowest bands of each variable i.e., minimum number of nodes with low values of interactivity, semantic, and attractivity.
- A fairly contributed cell (FCC) is a cell that falls neither into a cold nor hot spot category. This means that it contains a moderate amount of nodes (less than 100 nodes) with average values of interactivity, semantic, and attractivity.
- A highly contributed cell (HCC) is a cell that is highlighted as a hot spot and contains the highest values for each variable, i.e., above average number of nodes (above 100 nodes) with high values of interactivity, semantic, and attractivity.

3.3 The Cellular Automata-Markov approach

Finally, the cell-based CI is projected to future years through a CA-Markov model representing a frequently employed predictive modeling technique (e.g., Batty 1999; Jokar *et al.,* 2011; Spicer *et al*., 2012). It benefits from a multi-criteria evaluation function which combines cellular automata (CA) and Markov Chain models (Eastman 2012). While a Markov chain model quantifies transition probabilities of multiple classes of thematic maps, the CA model allocates the predicted quantity of fluctuations over the space for a certain period of time through the probabilistic measures. Since the Markov chain model itself does not generate spatial outputs, the model must be combined with a spatially explicit approach (Peterson *et al*., 2009; Guan *et al*., 2011). Due to the conceptual simplicity of the CA, it has been utilized for modeling a variety of dynamic phenomena, including land-use/land-cover changes (e.g., Mitsova *et al*., 2011), fire spread (e.g., Stambaugh & Guyette, 2008), disease dissemination (e.g., González *et al*., 2013), and social changes and dynamics (e.g., Dabbaghian *et al*., 2011).

The advantages of both models are integrated into a single and robust modeling technique called the CA-Markov model by quantifying the probabilities of phenomenon dynamism via the Markov chain model and allocating the estimated changes through CA to predict the future evolution (Zhou *et al*., 2012). The CA-Markov model is founded on an initial distribution of the dynamic phenomenon and a transition matrix, assuming that past driving forces will also operate in the future (Mondal & Southworth, 2010). Several empirical studies have confirmed the power of CA-Markov models (Kamusoko *et al*., 2009; Jokar Arsanjani *et al*., 2013c). Both aspects make CA-Markov modeling suitable for simulating the future evolutions of OSM contributions assuming no change in the form of intervention so that the contribution rate and number of involved users will continue to grow at the same rate as it has to date.

4. Results and discussions

4.1 Spatio-temporal Mapping of OSM contributions

Descriptive mapping permits us to visualize the location and number of nodes over time. As shown in Figure 3, the early contributions were delivered in 2007 and gradually began to grow out and spread over the whole area until the end of 2012. Spatially overlaying these maps with the CORINE land cover data reveals that the early contributions were received mainly in artificial surfaces (54%). This means that areas with a high number of nodes delineate the artificial surfaces. Agricultural areas rank second for receiving contributions (35%) followed by forest and semi natural areas (12%). Hot spots on the contribution maps of 2011 and 2012 roughly delineate the developed areas. From these hot spots residential areas (urban and rural) and road networks are easily detectable.

Figure 3: Spatio-temporal pattern of OSM contributions from 2007-2012

Moreover, visual analysis of Figure 3 in combination with land cover overlays reveal that the density of contributions is also increasing. This causes cells with higher number of nodes delineate residential areas such as urban and rural areas. Similarly, Crandall *et al*. (2009) and Li *et al*. (2013) remark that spatial patterns of Tweets from Twitter as well as Flickr photos' primarily delineate administrative boundaries of the United States and major roads.nterestingly, the number of involved users is also increasing so these trends indicate that more users will likely become involved (see Table 1) Interestingly the number of users involved is also increasing. These trends indicate that more users are likely to become involved in the future (see Table 1). The mean number of attributes identifies how many nodes are given attributes and they can identify objects. This measure also shows an increasing trend of additional attributes related to the contributed objects. The mean OSM version number shows how many times on average a node has been edited. As noted by Mooney and Corcoran (2014), a higher number of *osmversion* describes that the object has been modified more than once and therefore the uncertainty on the location and attribute of the object decreases.

Table 1: Descriptive statistics of the contributions and contributors from 2005–2012

Figure 4: Regression results of selected temporal OSM trends

Using univariate regression, Figure 4 relates selected OSM characteristics (e.g., the number of nodes) to the yearly timestamps. The coefficients of determination (R^2s) show a fairly good fit, in particular the variables "number of contributing users" as well as "number of contributed cells" achieve R^2s of 93 and 95%, respectively. In combination with Table 1 and Figure 4, the following conclusions can be drawn:

- Number of nodes: Although there were few nodes until 2008, the number has been constantly increasing. The regression depicts a sharply increasing rate of receiving contributions and its trend has been exponentially increasing from 2010 onwards.
- Number of contributed cells: Supported by the significant regression parameters, the cellular abstraction of contributions also reveals that the rate of receiving contributions in both forms – nodes and cells – has been increasing.
- Analysis of attributes (mean and standard deviation) also proves that over time, objects receive more attributes than before. Although this dropped sharply in 2011, thereafter it started to increase. This means that the contributions in 2011 had a reduced number of attributes relative to the other years and this could be due to a new wave of users that did not add attributes to their contributions.
- Version of contributions: the mean version of contributions increased from 1 (only edited by one user) in 2007 to almost 2 in 2012. This means that on average the objects were edited either by 2 users or within 2 editing sessions by a single user.
- Figure 4 (bottom) also displays an increasing rate of involved users involved in the mapping. The relatively high R^2 of approximately 95% demonstrates that despite a slow rate of gaining users the number of involved users is constantly rising indicating that OSM will continue to grow for the foreseeable future. Furthermore, along with the increase in number of users, the rate of nodes per person has been constantly increasing. Over time number of nodes, number of users and their share in mapping has been increasing, i.e., the users are mapping more than

before. This might be interpreted as the more users map the more skilled they become; however this assumption must be practically tested and is beyond the scope of this paper.

- A close observation of the contributing users between 2010 and 2012 shows that those 884 users from 2011 continued their contributions along with the 284 new users in 2012. Similarly, the 622 users of 2010 continued their contributions along with the 262 new users in 2011.

To sum up, the statistical analysis indicates a promising outlook in terms of OSM receiving further contributions. In other words, OSM is becoming more popular amongst people and it is very likely to continue its success into the future.

4.2 Contribution Index Analysis

In order to translate each category of the CI into the four indicators (i.e., quantity, interactivity, semantic, and attractivity), the fluctuations of each variable per each CI category is calculated through a zonal statistic. Table 2 depicts the variations for each variable per CI category. The mentioned indicators and thresholds can be used as a rough explanation for calculation of CI for other areas.

Table 2: Quantification of OSM contributions into four indicators

Figure 5: Patterns of CI from 2007 to 2012

The resulting CI maps are shown in Figure 5 displaying that the majority of cells were either dead or barely contributed cells. Since 2009, fairly and highly contributed cells appeared and then began to spread out over the study area. A major change to the cells in the study area is evident from 2010. Between 2011 and 2012, a large number of dead cell were converted to fairly and highly contributed cells so that only a small number of dead cells remain. This could possibly be the aftermath of mapping parties' calls that were sent out in 2011 and 2012 (see http://wiki.openstreetmap.org/wiki/Stuttgart/Stammtisch#). Likewise, these findings are supported by Figure 6 which indicates that the number of dead cells has been dramatically decreasing in favor of other categories, whilst fairly and highly contributed cells have taken bigger proportions since 2011.

Figure 6: Development trend of CI over time in terms of number of cells in each CI category

4.3 CA-Markov: implementation, validation and prediction

In accordance with Figure 2, to predict future CI patterns the CA-Markov model is set up. To determine the most appropriate transition rules, neighborhood definition, and kernel size as well as to evaluate the model's performance, the model is applied using the past data for the years 2010, 2011, and 2012 in an iterative manner until the associated transition rules resulted in highly correlated outputs compared to the actual reality represented through the latest available timestamp. The CI maps of 2010 and 2011 are imported into the CA-Markov model in order to simulate one timestamp after that, 2012. The actual CI map of 2012 is used to evaluate the performance of the CA-Markov model using the Kappa statistic. While several modeling parameters were tested, the most optimal match was achieved at iteration number of 42 using a 3×3 kernel size and a von Neumann neighborhood definition which yielded an overall 68.3% Kappa index of agreement. This indicates a substantial level of agreement between the simulated map and actual map according to Landis and Koch (1997). Figure 7 displays the resultant predicted map of 2012 as well as the actual map of 2012 for a better visual comparison.

Figure 7: The actual (left) and the predicted map (right) of the CI for the year 2012 by CA-Markov model

Finally, the characteristics of the best calibrated model in terms of kernel size, neighborhood function, and number of iterations are employed to simulate the upcoming OSM contributions for the years 2016 and 2020. Figure 8 illustrates the spatial pattern of the predicted CI maps. The predicted maps of CI in 2016 and 2020 disclose that the fairly and highly contributed classes with 35% of areal coverage will distinctly take over the dead cells and barely contributed cells with 75% and 90% coverage, respectively which cover artificial surfaces as well as forest and agricultural areas. The remaining barely contributed areas will cover partially the forest areas on the south-west part of the study area. This could be either because this land use type is not interesting enough to receive enough contribution or there are not many objects in these cells requiring mapping. Moreover, since OSM nodes are taken as footprints of contributions, only the edges of farming lands (i.e., features represented as polygons) are considered. The areas covered by polygons are not considered.

Figure 8: Predicted maps of CI for 2016 (left) and 2020 (right)

It should be mentioned that the resultant prediction patterns are based on the temporal growth of OSM and no potential driving factor is considered. Considering additional factors will most likely result in both different outputs and prediction patterns. Such a prediction approach permits researchers to gain an overall impression on the possible future patterns of OSM dissemination. The patterns of predicted maps show a converged form which is due to essence of CA-Markov model as noted by Eastman (2012).

5. Conclusions

Multiple objectives were considered in this research to (i) evaluate the trend of collaborative contributions to the OSM project over time and space on the basis of a grid representation within a sample study area, (ii) to develop a CI for indicating several aspects of contributions to OSM such as quantity, attractivity (how many users are active in a cell), semantic, and interactivity, (iii) predict the future status of contributions based on the CI in order to gain some insight regarding which direction the OSM project is heading in the future. The greater urban area of Stuttgart, Germany, which contains both urban and rural areas, was selected as a case-study. In order to develop a CI, in addition to number of nodes per cell, other variables such as average “osmversion”, average number of users and number of attributes within each cell were considered. The combination of these four variables as well as applying G*-statistics has allowed us to define four different categories of CI. These four categories are named as follows: a) dead cells in which no nodes exist; b) barely contributed cells in which the number of nodes is relatively low (2-10 nodes) and the contributions have been edited a few times and shared by a few users with minimum attributes; c) fairly contributed cells are those which contain up to 100 nodes contributed by a number of users and edited a number of times; while d) highly contributed cells are those which contain the most number of nodes (above 100 nodes) and are edited frequently amongst a high number of users.

The projection of CI in a spatio-temporal framework allows us to study the past contribution trends and also to simulate the future OSM contribution patterns of the CI through a CA-Markov model. The results reveal that the rate at which OSM is receiving contributions from users has been constantly increasing and is continuing to grow. Furthermore, the number of users and the number of given attributes have also been growing. This includes an increase in the number of contributions. The CI maps for historic timestamps also confirm our claim that these cells are being more actively contributed to. The simulated maps of 2016 and 2020 in addition to the qualitative measures of the CI indicate that a considerable amount of cells (up to 90%) will turn to fairly- and highly-contributed by these times. This could provide us with better data quality measures by minimizing the “long tail” effect. In other words contributions will be edited by a larger number of OSM users and shall subsequently benefit from the strength of the collaborative mapping efforts of the OSM community.

Currently there is a lack of empirical studies which investigate the characteristics of future contributions to OSM. In this regard the findings of our study certainly enhance the literature on OSM in a number of ways. Firstly, as shown in the results section, increasing numbers of people have been drawn gradually to OSM. Within the first three years of the launch of OSM very few contributions were provided. However since then an exponential rate of contributions have been received. Secondly, a spatial and temporal dependency between the contributions' characteristics (e.g., object type, quantity, number of involved users, version number) and physical characteristics exist. This has been demonstrated by considering the CORINE land cover map. In general, artificial surfaces are mapped earlier and in greater frequency than agricultural areas and forest/semi-natural areas. There are a number of reasons for this including: (a) the objects in such land-types are not always evenly distributed across regions; (b) less people are interactively involved with these feature types; (c) these land types change very slowly over a short period of time; and (d) many contributors might not know very much about these objects as they are not public places and often only the routes through them are mapped by users. The findings and results of the 2016 and 2020 simulation maps reveal that more users will contribute by creating/editing more objects containing an increasing number of attributes. It has been shown that these objects will then be revised by more users resulting in very few cells remaining unmapped or barely contributed. A valuable research finding from this work for OSM communities and OSM end-users is that in the next few years there will be many more contributions to OSM. Many more users will become involved and their contributions will have more attributes which will be revised and edited by a greater number of users.

Although our empirical findings are extracted from a specific case study this research has demonstrated that the characteristics of contributions are related to socio-economic and physical factors. It is of great research importance to investigate their relationship to how OSM is disseminated. Other issues including how bulk import of official data integrate with OSM data will allow for the development of a more extensive contribution index as part of our future work. It must be noted that the individuals, as mappers, have substantial influence over most OSM contributions. As OSM and similar projects are growing and being disseminated remarkably quickly further studies on understanding the behavior of these individuals in collaborative projects are required. These studies will need to investigate technological developments which make these projects more attractive and user-friendly. Therefore, as individuals are the main actors in such phenomenon, individuals-based modeling techniques such as agent-based modeling might be an alternative technique to simulate user contributions. Furthermore, considering contributions based on which object they represent e.g., buildings, roads, etc. may be beneficial to study the spatiotemporal patterns of OSM contributions in a more effective manner.

Acknowledgments:

The authors acknowledge the constructive comments of the three anonymous reviewers and the editor-in-chief all of which have helped to improve the article. Jamal Jokar Arsanjani was funded by the Alexander von Humboldt foundation.

References:

Balzter, H., Braun, P. W., & Köhler, W. (1998). Cellular automata models for vegetation dynamics. Ecological Modelling, 107(2-3), 113-125.

Barron, C., Neis, P. and Zipf, A. (2013). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. Transactions in GIS. doi: 10.1111/tgis.12073

Batty, M., Xie, Y., and Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. Computers, Environment and Urban Systems, 23(3), 205–233.

Chilton, S. (2009). Crowdsourcing is radically changing the geodata landscape: case study of OpenStreetMap. In Proceedings of the UK 24th International Cartography Conference, Santiago, Chile, 15–21 November 2009.

Corcoran, P., Mooney, P., & Bertolotto, M. (2013). Analysing the growth of OpenStreetMap networks. Spatial Statistics, 3(0), 21–32. doi:http://dx.doi.org/10.1016/j.spasta.2013.01.002

Dabbaghian, V., Spicer, V., Singh, S. K., Borwein, P., & Brantingham, P. (2011). The social impact in a high-risk community: A cellular automata model. Journal of Computational Science, 2(3), 238–246. doi:http://dx.doi.org/10.1016/j.jocs.2011.05.008

Eastman, J. R. (2012). IDRISI Selva, guide to GIS and remote processing. Guide to GIS and Remote Processing. Worcester: Clark University.

European Environment Agency (2013), Mapping Guide for a European Urban Atlas, Retrieved from http://www.eea.europa.eu/data-and-maps/data/urban-atlas.

Georgiadou, Y., Lungo, J. H., & Richter, C. (2013). Citizen sensors or extreme publics? Transparency and accountability interventions on the mobile geoweb. International Journal of Digital Earth, 1–18.

Girres, J.-F. and Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. Transactions in GIS, 14, (4), pp. 435-459.

González, R. E. R., Coutinho, S., Dos Santos, R. M. Z., & De Figueirêdo, P. H. (2013). Dynamics of the {HIV} infection under antiretroviral therapy: A cellular automata approach. Physica A: Statistical Mechanics and its Applications, 392(19), 4701–4716.

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

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. Spatial statistics, 1, 110-120.

Goodchild, M. F., and Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth, 3(3), 231–241.

Guan, D., Li, H., Inohae, T., Su, W., Nagaie, T., & Hokao, K. (2011). Modeling urban land use change by the integration of cellular automaton and Markov model. Ecological Modelling, 222(20-22), 3761–3772. doi:10.1016/j.ecolmodel.2011.09.009

Hagenauer, J. and Helbich, M. (2012): Mining urban land use patterns from volunteered geographic information using genetic algorithms and artificial neural networks. International Journal of Geographical Information Science (IJGIS). 26(6), 963-982. DOI:10.1080/13658816.2011.619501.

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), 682-703.

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

Haklay, M. M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. Cartographic Journal, 47(4), 315-322.

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

Hardy, D., Frew, J., Goodchild, M.F (2012). Volunteered Geographic Information Production as a Spatial Process. International Journal of Geographical Information Science 26(7):1191–1212.

Helbich, M., Amelunxen, C., Neis, P., & Zipf, A., (2012). Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. In Jekel T, Car A, Strobl J and Griesebner G: Geospatial Crossroads @ GI_Forum '12. Proceedings of the Geoinformatics Forum Salzburg. Wichmann: Heidelberg, pp. 24-33 (http://gispoint.de/fileadmin/user_upload/paper_gis_open/537521013.pdf)

Huang, B., Zhang, L., and Wu, B. (2009). Spatiotemporal analysis of rural–urban land conversion. International Journal of Geographical Information Science, 23(3), 379–398.

Jokar Arsanjani, J., Barron, C., Bakillah, M., Helbich, M., (2013a). Assessing the Quality of OpenStreetMap Contributors together with their Contributions. In proceedings of the AGILE 2013, Leuven, Belgium.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., Hagenauer, J., & Zipf, A. (2013b). Toward mapping land-use patterns from volunteered geographic information. International Journal of Geographical Information Science, 27(12), 963-982. DOI:10.1080/13658816.2013.800871.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., & Loos, L. (2014). The emergence and evolution of OpenStreetMap: A cellular automata approach. International Journal of Digital Earth, doi:10.1080/17538947.2013.847125

Jokar Arsanjani, J., Helbich, M., Kainz, W., & Bloorani, A. D. (2013c). Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. International Journal of Applied Earth Observations and Geoinformation. doi:10.1016/j.jag.2011.12.014

Jokar Arsanjani, J., Bakillah, M., (2015): Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap, International Journal of Digital Earth. DOI:10.1080/17538947.2013.847125

Jokar Arsanjani, Jamal, and Eric Vaz. 2015. An Assessment of a Collaborative Mapping Approach for Exploring Land Use Patterns for Several European Metropolises. International Journal of Applied Earth Observation and Geoinformation 35(0):329–37.

Kamusoko, C., Aniya, M., Adi, B., & Manjoro, M. (2009). Rural sustainability under threat in Zimbabwe – Simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Applied Geography, 29(3), 435–447. doi:http://dx.doi.org/10.1016/j.apgeog.2008.10.002

Landis, J.R., Koch, G.G., (1997). The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Mitsova, D., Shuster, W., & Wang, X. (2011). A cellular automata model of land cover change to integrate urban growth with open space conservation. Landscape and Urban Planning, 99(2), 141–153. doi:http://dx.doi.org/10.1016/j.landurbplan.2010.10.001

Mondzech, J., Sester, M. (2011). Quality analysis of OpenStreetMap data based on application needs. Cartographica, 46, 115–125.

Mooney, P., & Corcoran, P. (2014). Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS*,*18*(5), 633–659.

Mooney, P. and Corcoran, P. (2012b). Characteristics of Heavily Edited Objects in OpenStreetMap. Future Internet, 4(1), 285-305.

Mooney, P., & Corcoran, P. (2013). Has OpenStreetMap a role in Digital Earth applications? International Journal of Digital Earth, 1–20. DOI: 10.1080/17538947.2013.781688

Mooney, P., Corcoran, P. and Ciepluch, B. (2012). The Potential for using Volunteered Geographic Information in Pervasive Health Computing

Applications. Journal of Ambient Intelligence and Humanized Computing - special issue on Pervasive Health Computing.

Neis, P., and Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project – The Case of OpenStreetMap, ISPRS Int. J. Geo-Inf. 2012, 1, 146-165.

Neis, P., Zielstra, D., Zipf, A., (2013). Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. Future Internet. vol5, pp. 282-300.

Neis, P., Zielstra, D., (2014). Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. Future Internet, 6, 76-106.

OpenStreetMap (2013). OSMstats (retrieved from http://osmstats.altogetherlost.com/)

Pan, Y., Roth, A., Yu, Z., & Doluschitz, R. (2010). The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. Computers, Environment and Urban Systems, 34(5), 400-408.

Peterson, L. K., Bergen, K. M., Brown, D. G., Vashchuk, L., & Blam, Y. (2009). Forested land-cover patterns and trends over changing forest management eras in the Siberian Baikal region. Forest Ecology and Management, 257(3), 911–922. doi:10.1016/j.foreco.2008.10.037

Pourabdollah, A., Morley, J., Feldman, S., Jackson, M., (2013). Towards an Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps’ Road Network. ISPRS Int. J. Geo-Inf. 2(3), 704-728.

Ramm, F., Topf, J., & Chilton, S. (2010). OpenStreetMap: Using and Enhancing the Free Map of the World. Uit Cambridge.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2011). GeoWeb and crisis management: issues and perspectives of volunteered geographic information. GeoJournal, 78(1), 21–40. doi:10.1007/s10708-011-9423-9

Roick, O., Hagenauer, J., & Zipf, A. (2011). OSMatrix–grid-based analysis and visualization of OpenStreetMap. In Proceedings of the 1st European State of the Map.

Spicer, V., Reid, A. a., Ginther, J., Seifi, H., and Dabbaghian, V., (2012). Bars on blocks: A cellular automata model of crime and liquor licensed establishment density. Computers, Environment and Urban Systems, 36(5), 412–422.

Stambaugh, M. C., and Guyette, R. P. (2008). Predicting spatio-temporal variability in fire return intervals using a topographic roughness index. Forest Ecology and Management, 254(3), 463–473.

Wang, Y., and Li, S. (2011). Simulating multiple class urban land-use/cover changes by RBFN-based CA model. Computers and Geosciences, 37(2), 111–121.

Yates, D., & Paquette, S. (2011). Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake. International Journal of Information Management, 31(1), 6–13. doi:10.1016/j.ijinfomgt.2010.10.001.

Zhou, D., Lin, Z., & Liu, L. (2012). Regional land salinization assessment and simulation through cellular automaton-Markov modeling and spatial pattern analysis. Science of the Total Environment, 439, 260–274. doi:10.1016/j.scitotenv.2012.09.013

Zielstra, D. and Zipf, A., (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany. In Proceedings of 13th AGILE international conference on geographic information science, 10-14 May 2010, Guimaraes, Portugal. Available from: http://agile2010.dsi.uminho.pt/pen/ShortPapers_PDF/142_DOC.pdf

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered Geographic Information and Crowdsourcing Disaster Relief: A Case Study of the Haitian Earthquake, 2, 7–33. doi:10.2202/1948-4682.1069.