

Research Article

Open Access

Mohammed Zia*, Ziyadin Cakir, and Dursun Zafer Seker

Turkey OpenStreetMap Dataset - Spatial Analysis of Development and Growth Proxies

<https://doi.org/10.1515/geo-2019-0012>

Received February 1, 2018; accepted April 16, 2018

Abstract: Number of studies covering major data aspects of OpenStreetMap (OSM) for developed cities and countries are available in scientific literature. However, this is not the case for developing ones mainly because of low data availability in OSM. This study presents a time-series spatial analysis of Turkey OSM dataset, a developing country, between the year 2007 and 2015 to understand how the dataset has developed with time and space. Five different socio-economic factors of the region are tested to find their relationship, if any, with dataset growth. An east-west spatial trend in data density is observed within the country. *Population Density* and *Literacy Level* of the region are found to be the factors controlling it. It has also been observed that the street network of the region has followed the Exploration and Densification evolutionary model. High participation inequality is found within the OSM mappers, with only 5 of them responsible for the country's 50% geo-data upload. Furthermore, it is found that these mappers use other Volunteered Geographic Information (VGI) and government open-dataset to feed into OSM. This study is believed to bring some high level insights of OSM for a developing country which would be useful for geographers, open-data policy makers, VGI projects planners and data-curators to structure and deploy similar future projects.

Keywords: OpenStreetMap, spatial development, socio-economic factors, Exploration and Densification

***Corresponding Author: Mohammed Zia:** GIScience Research Group, Institute of Geography, Heidelberg University, Heidelberg, Germany, E-mail: mohammed.zia33@gmail.com; <https://www.linkedin.com/in/zia33>

Ziyadin Cakir: National Innovation and Research Center for Geographical Information Technologies, Turkey and Department of Geology, Faculty of Mines, Istanbul Technical University, Turkey

Dursun Zafer Seker: Geomatics Engineering Department, Faculty of Civil Engineering, Istanbul Technical University, Turkey

1 Introduction

With the ease of online-data generation and dissemination by the advent of Web2.0 technology the narrowly-targeted goods and services are getting as economically attractive and lucrative as mainstream fare [1, 2]. For example, we now have Whatsapp, Skype, Viber etc. in Telecommunication sector, which was predominantly governed by Telecom industries for decades. Wikipedia holds a huge pile of online free-content and competes with other proprietary data sources like Encyclopedia Britannica. Recent free publication of classified information by WikiLeaks and OpenLeaks has challenged governments and other national intelligence agencies, who were regulating confidential information of citizens so far [3].

Similarly, the Volunteered Geographic Informations (VGI) [4, 5] or Crowdsourcing Geographic Data [6, 7] have evolved to hold and control the geographic data which was compiled and retained by National Mapping Agencies and other private cartographic companies in the past [8]. The reason for their sudden success is the ease of geo-data generation and circulation where human-beings act as a sensor [9]. This contemporary approach has allowed even naive cartographers/citizens with limited or no mapping experience to collect, map and upload geo-data of any region. One famous VGI project is called OpenStreetMap (OSM) which was started in the year 2004 with the objective to create a free and editable street map of the world [10]. Although there are many other VGI projects like Wikimapia, Wikiloc, Foursquare and Google Map Maker, the hype in popularity of OSM is because of flexible licensing, active community and advanced semantics. This classic example has recently gained immense popularity because of its big volume data, heterogeneity, abundance and free data access [11, 12], and therefore has attracted researchers from ranging domains [13].

By the end of 2015, OSM dataset has had an enormous amount of tagged geo-data in the form of approximately 5 billion GPS points, 3 billion nodes, 3 billion ways and 4 million relations, which were contributed by around 2.5 million registered users worldwide [14]. One of the many possible reasons for this popularity hype was the partial

pulling out of Google-Maps APIs in 2012 from their public domain [15]. This encouraged services like Apple iPhoto, FourSquare, Craigslist, Flickr [16], etc. to use OSM. This massive dataset has brought forth possibilities to investigate data aspects such as data accuracy, data exhaustiveness, time-series data evolution, motivation behind mapping, relationship with other VGI projects, etc. [17].

Dataset accuracy of OSM at regional and global scale has already been studied by researchers from different aspects; for examples, by comparing it with government and other dataset [18, 19], by using proxy approaches like Linus Law [20] or contributors count [21–23], by developing intrinsic quality assessment parameters and tools [24–26], by reviewing OSM changeset dump file [27] etc. OSM participation inequality is extensively studied by researchers as a proxy for data quality check-up [$data_quality \propto 1/participation_inequality$] [13, 22, 23]. However, this proxy was opposed by [28] by negating the idea of counting the number of contributors for quality assurance. [25] and [26] have tried to use the intrinsic parameters of OSM for data scrutiny and health, thus adding additional quality assurance mechanisms. It is reported that the early adopters of OSM project, like Germany and UK, contain dataset of high quality standards. Few regions of Germany contain dataset even more accurate and exhaustive than Google Maps. However, this quality standard is limited and many developing and under-developed nations still lag mapping of elementary features.

Detailed analysis on the coverage and quality estimation of OSM dataset has always been an issue because of strict licensing policy, limited usage/availability, high pricing etc. of government and other proprietary dataset that act as a reference dataset. For example, [29], [18], [19] and [30] have reported this limitation for Portugal, London, Kenya and Germany, respectively. Nonetheless, some notable scientific studies are available online that acted as a motivation for this study. [31] have studied the dataset of Germany using some proprietary dataset, [32] have compared the bicycle trail and lane dataset of OSM with data from local planning agencies for USA, [16] have mapped the street network dataset of USA by comparing it with TIGER/Line dataset etc. Few OSM use-case scenarios include measuring the street network evolution as a proxy for urban sprawl [34], developing Location-Based and Emergency Medical Services [35–37], generating interactive 3D City Models using Shuttle Radar Topography Mission height data [38], extracting Image-based road network [39] and multilane roads data [40], calculating shortest routes within urban cities [33], validating/reforming existing Land Use/Land Cover (LULC) data like Global Land Cover Maps [41] etc. It has also been used during nat-

ural calamities like the infamous Haiti earthquake of 2010 [42]. [9, 43, 44] have argued the personal satisfaction and community serving as two key motivating factors behind any crowd sourcing activity.

The aim of this study is to understand the time-series spatial evolutionary pattern of Turkey OSM dataset between the year 2007 and 2015 in order to help researchers, developers and policy makers to better structure current and future VGI projects, to identify best policies for open-data license, to predict data growth and quality of certain regions and more. Many researchers in the past have done similar studies for developed countries and cities. For example, [30] have reported how the online mapping was active in the year 2009 in Germany and tried to compare it with TeleAtlas MultiNet dataset, [45] have performed an interregional comparisons of European regions, [46] have analysed three urban areas in Ireland to understand the time-series street network evolutionary pattern in OSM and [47] and [48] have reported the densification and evolutionary model for Irish and Chinese street network dataset, respectively. Additionally, the effect of different socio-economic factors of the region, i.e. *Literacy Level* (number of graduate students), *Population Density*, *Tourism Activity*, *Internet Usage* [21] and *Human Development Index (HDI)*, are tested on this growth. A general commentary on the state of the OSM dataset quality and mappers involvement of the country is provided in the end. To the best of authors' knowledge, this kind of high resolution (at provincial level) OSM time-series statistical analysis of a developing country is first of its kind. We hope that this study will bring forth patterns, trend, proxy parameters and future research paradigms that are necessary to restructure existing and future VGI projects.

1.1 OSM in Turkey

No key study of Turkey OSM dataset regarding (a) spatial evolution with time, (b) different socio-economic factors governing it, (c) mappers involvement within the country, (d) dataset quality is available online. Current study, therefore, becomes essential to answer key elements of the ecosystem on spatial and temporal ground as has already been done by other researchers for developing nations but at a much coarser level. For example, [13], [19] and [48] have done similar spatial and temporal analysis for street networks of Beijing, Kenya and China, respectively. Turkey OSM provides a rich dataset with nearly 17 million points, 1.3 million edges and 0.4 million polygons [49]. In this study, the term edge is used for any line and

polyline feature. Authors hope that this will add-on some key findings of developing nations into the OSM literature.

1.2 Five aspects of dataset analysis

The studied five aspects of the analysis are as follows:

1. *Time-series spatial evolution of OSM dataset* - The spatial growth of the dataset within the region has been studied on an eight year of time-span (2007-2015). This advances the work of [13, 47, 48] for Ireland and China where authors have observed heavy-tailed pattern both for a certain time period and across time. It explains if the data density and inherent mapping activity by volunteered mappers is spatially biased or not.
2. *Effect of socio-economic factors on OSM spatial evolution* - Five socio-economic factors governing any VGI activity, as discussed by other researchers as well [21, 31], have been compared with corresponding OSM feature density for different regions. These factors are *Population Density, Literacy Level, Tourism Activity, Internet Usage*, and Human Development Index (*HDI*). It has been speculated that *Internet Usage* and *Tourism Activity* plays a vital role in VGI mapping in any region.
3. *Processes governing the evolution of road networks within the country* - The street network growth with time has been studied within the whole country by considering the Exploration and Densification mechanism of [50] which argues that any urban street mapping first follows an exploration of unmapped streets and then follows a densification of all nearby secondary and tertiary streets. This mechanism is observed in Ireland and Beijing as well [46, 47].
4. *Mapping behaviour of OSM mappers* - The density of the number of mappers has been measured to identify the maturity of the dataset of a region. This kind of proxy study has already been done by other researchers as well for different resolutions and regions, like in Germany by [30, 31], on a global scale by [3, 17], in England by [20], in UK and Ireland by [28], in Beijing, China by [13] and in USA by [16]. It has been reported that the number of active mappers, excluding bulk importers, show a strong correlation with data quality and density.
5. *Quality of the dataset* - Mappers' participation inequality and various data sources have been identified to provide a general commentary on the Turkey OSM dataset quality and accuracy. Participation inequality has been used as quality proxy by past researchers as well, for example for China [13], for major world cities

[21] and for whole Planet OSM dump file [22], where they have reported that it is inversely proportional to the dataset quality.

2 Methods

2.1 Source and format of the dataset used

Another reason for OSM popularity is its various data sources (Full Planet dump file [49], Geofabrik downloads [51], Overpass API [52]) and formats (ESRI-Shapefiles *.shp, Extensible Markup Language (XML) *.osm, Protocolbuffer Binary Format *.pbf), thus, causing high data interoperability. A switch from Creative Commons Attribution-ShareAlike 2.0 to Open Database License (OdbL) in 2012 [53] allowed users and developers to share, modify and use the dataset freely and per use-case [54]. Technological advancements and user requirements, further, caused the upgrade of its Editing API (latest v0.6) by bringing in many state-of-the-art editing functionalities [55]. The object history feature was introduced by Editing API v0.5 [55] and, therefore, one will not find any edits in Planet dump file prior to 2007 (section3.1). The 2012 event to change its license caused the loss of 1% dataset because of conflicting users' interest [53]. Nonetheless, the dump files are proven to be the best source of OSM dataset to study any time-series evolution [16, 27, 32], since other sources only reflect its latest snapshots for specialized regions.

Full Planet dump file (1.5 TB uncompressed) from September 2015 is downloaded from [49] for this study. The file is in a human readable XML format with three primitive data features, i.e. nodes or points; ways or polylines or polygons; and relations (logical combination of nodes, ways and/or other relations). These features are annotated by tags in a key-value structure of free format text fields [56]. Provincial data of five stated socio-economic factors of the region is downloaded from TUIK (Turkish Statistical Institute) website [57]. Specific downloaded columns from TUIK website are as follows: "Number of students for vocational training school and undergraduate programs of higher education institutions: Graduates / Total – Education – Higher Education" for *Literacy Level* (fig.3a), "Annual growth rate and population density of provinces – Population and migration – General Population Censuses" for *Population Density* (fig.3b), "Number of arriving foreigners by province of border gate and mode of transport : Air way – Tourism" for *Tourism Activity* (fig.3c), "The proportion of individuals regularly using the Inter-

net – Transportation and Communication” as *Internet Usage* (fig.3e), and “Per capita gross value added (GVA) : Per capita GVA (\$) – National Accounts” for *HDI* (fig.3d). Since the ill-famed Syrian refugee crisis has caused a heavy influx of foreigners into the country via roadways (South-East Anatolia, fig.3e), only airways mode of transport is considered for *Tourism Activity* (fig.3c). Data to create the road density map of the country (fig.3f) is downloaded from the General Directorate of Highways [58] (section3.1).

2.2 Steps in data processing

OSM XML file could be processed by a variety of command line tools, like *osmosis* (Java application for reading/writing databases [59]), *osmium* (multipurpose tool for data interoperability and time-series analysis [60]), *osm2pgsql* (tool to convert XML data to PostGIS-enabled PostgreSQL databases [61]), *osm2postgresql* (to simplify rendering with QGIS and other GIS/web servers [62]), *osm2pgrouting* (to import data file into pgRouting databases [63]) etc. For this study, an *osmium* based tool, called *osm-history-splitter* [64], is used to split the Planet dump file using the bounding-box of Turkey. Subsequently, the ESRI’s shapefile of the country is used to further divide the dataset into 81 provinces. Finally, the data was loaded into a PostGIS enabled PostgreSQL database.

It is observed that only 2% of all data points of the region have some sort of tags associated with them. Therefore, to study the spatial evolution of dataset it is divided into the following four categories: (a) Points(all) (all data-points from the dump file), (b) Points(tagged) (data-points with tags associated with them), (c) Edges (all highway features), and (d) Polygons (all building, landuse, and natural features). All developed scripts used to process the Planet dump file can be downloaded from [?], along with a comprehensive README file.

2.3 Data adjustment and normalization

Researchers, in the past, have already argued that feature count in a region can not be used to measure its density, especially for lines and polygons [18, 50]. In order to compare two elementary geometrical features, i.e. points, lines, or polygons, it is advised to count and compare the number of nodes constituting them [13] (fig.1). Same approach is used in this study to compare different features as well. Furthermore, each node count is divided by the area of the province to normalize it. In order to keep the point, line and polygon features comparable to each other the

dataset has not been normalized by the demography of the province. However, researchers like [50] and [47] have argued that to compare only the line features the dataset could be normalized by demography as well.

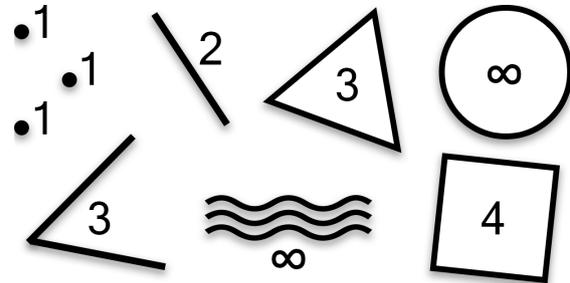


Figure 1: Different geometrical features along with corresponding nodes count.

Bulk import in OSM is defined as follows: Bulk import means more than a few hundred of nodes imported in a short period of time for a large area like a whole country [66]. This definition of bulk import is quite subjective and researchers have used their own definitions of threshold to remove it from dataset before any study [36]. The need to remove this kind of sudden import of dataset is to avoid unexpected data spikes or observations [27, 36]. In present study, authors have defined it as “25,000 nodes contributed by single mapper in a week” and have kept them out of the analysis. 10 mappers causing bulk import in the region are identified by this way and their contributions are removed for any spatial analysis.

In graph theory, the degree k_i of an i^{th} node is the number of nodes adjacent to it, i.e.

$$k_i = \sum_{j=1}^N a_{ij}$$

in terms of adjacency matrix [67]. In street network, the degree of a junction is the number of road segments originating from or terminating to it. Thus, frequency of different degree of junctions in a street network represents its topological structure. It explains how densely or sparsely the road segments are connected to each other. Degree distribution or fraction, which is defined as $P(k) = N(k)/N$ where $N(k)$ is the number of nodes with degree k and N is the total number of nodes in the network, is necessary to understand how evolved a network is at any given point in time. In other words, it explains how much the region has been explored and densified by mappers. High fraction of low degree junction in a city explains its Exploration phase as new roads are being identified and mapped, whereas, high fraction of high degree junction explains its Densifi-

cation phase [13]. These are two elementary mechanisms that govern the evolution of any street network [50].

Finally, per capita GVA (Gross Value Added) (\$) value of the region is used for *HDI* (Human Development Index) calculation. Previous researchers [21] have used per capita GNP (Gross National Product) value of a region to explain *HDI*. A detailed report of how to calculate *HDI* of a region can be found in the United Nations documentation [68, 69].

3 Results and Discussion

This section is divided into five aspects of current analysis.

3.1 Time-series spatial evolution of OSM dataset

Fig. 2 shows the time-series (2007-2015) evolution of different features, i.e. their normalized feature density, for 81 provinces of the country. It can be seen that all curves merge into the origin at around April 2007. It is because no object history feature was present in Editing APIs older than v0.5 which was introduced in 2007 [55], meaning no history data before that. Between 2007 and 2012 the slope of curves are mostly gentle (fig.2a,c,d). This is because of limited editing flexibility by old OSM license [53]. However, a sudden spike in slope can be seen from 2012 onwards because of the change of its license to ODbL [13]. This change caused sudden boost in data usage by different web-services, thus, encouraging developers and users to come back to the project to improve it [16]. High slope values for few provinces is because of the large number of active mappers in those regions [13]. It can be further seen that almost all the curves are exponential-step curves, rising each year at around summer time. It is believed that this pattern is because of the increased tourism and outdoor activity in the region during summer time, thus, causing mappers to plot what they observe and collect in the field.

Fig. 2 also shows feature density map of the country for three different time-slices (2009, 2012 and 2015). In general, it can be seen that the eastern and south-eastern parts of the country are less populated than the western and south-western parts. [48] has also reported similar spatial distribution for street networks in China-OSM. Since similar pattern, in general, can be seen for different socio-economic factors of the country fig.3, they are believed to be the driving forces for mapping activity. Fig. 3 shows the

Literacy Level, Population Density, Tourism Activity, HDI and Internet Usage in the country for the year 2014. Fig. 3f is a road density map of the country obtained from the General Directorate of Highways [58]. It can be stated that street network density is quite uniform within the country and whatever spatial bias one is observing in fig.2c is because of the non uniform mapping behaviour by OSM mappers [13, 46, 47]. Few provinces in the eastern and south-eastern parts are relatively dense. This is because of few senior mappers actively engaged in mapping activities in those regions. One such province is Batman (red box in fig.2) with its senior mapper from a university, table1. This mapper is responsible for the country's 3-4% of all geo-data upload. Since this mapper is currently a university student, only spike in the year 2015 can be seen.

3.2 Effect of socio-economic factors on OSM spatial evolution

Fig. 4 is a plot between *Population Density* and features density for three different time-slices (2008, 2011 and 2014). High R^2 value represents a strong correlation between the two axes. When comparing values based on human activities, values greater than 0.6 is considered to be highly correlated [45]. It can be seen that R^2 value is high for all features, especially for 2014. Similar observation is drawn for *Literacy Level* too, although it is not plotted to avoid redundancy and plot cluttering. No strong correlation between *Tourism Activity* – features density and *Internet Usage* – features density is found, in contrast to what [21] have reported. However, moderate correlation between *HDI* and features density is observed, similar to what has been reported by [21]. Overall, the *Population Density* and *Literacy Level* of the region are found to be the two strong proxies for features density in OSM, followed by *HDI*.

The R^2 value of *Population Density* for all different features increases with time (fig.4). For *Population Density* vs Points(all) the R^2 value for the year 2007 is high because of the scattered points lying close to the origin. It can be said that with time, overall, the correlation between socio-economic factors and OSM features density of a region gets stronger. Most of the data points for earlier years (2008) fall exactly on y-axis. This is because although the population of the region was high, people were unaware of the project and, therefore, mapping activity was quite limited.

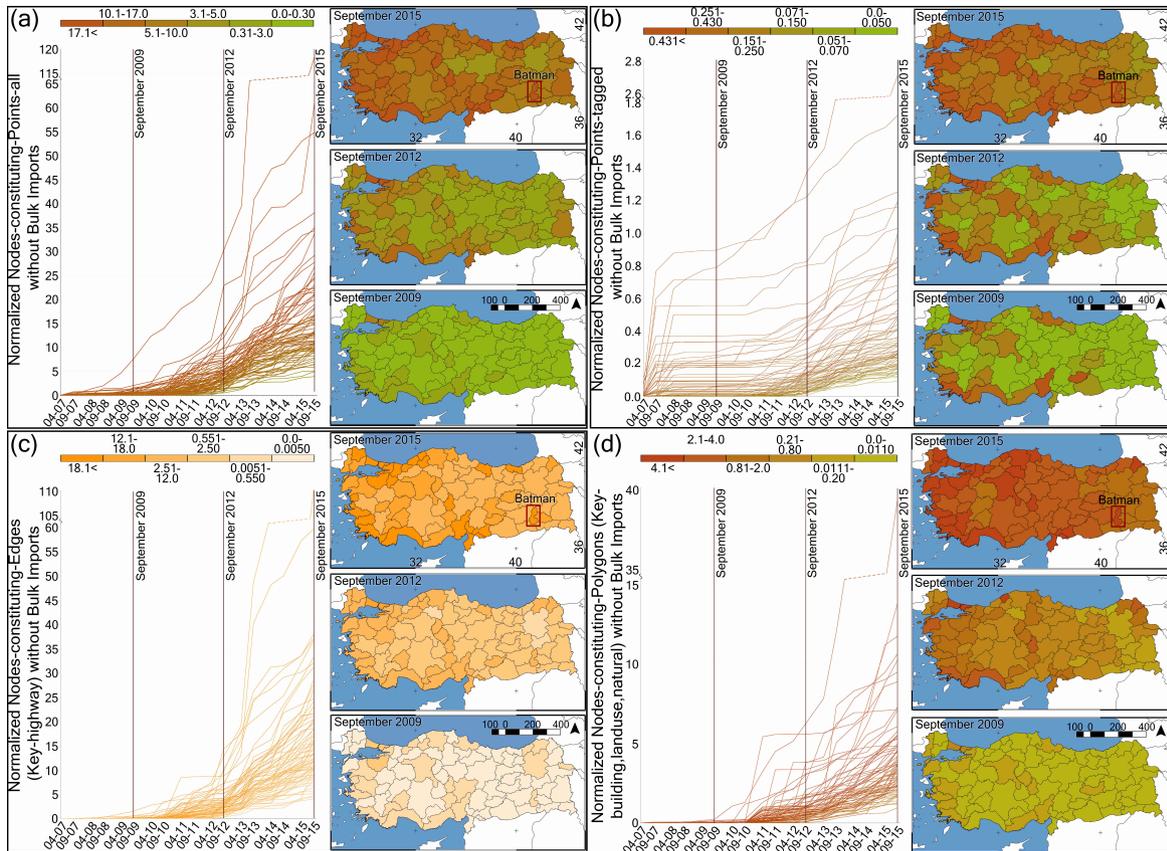


Figure 2: A plot between normalized features density and time. Maps represent density for three different time-slices (2009, 2012 and 2015).

3.3 Processes governing the evolution of road networks within the country

Elementary processes governing the evolution of any street network is explained by [50] by devising two processes, i.e. Exploration and Densification. According to this, any evolution first follows an exploration phase where new roads get discovered, followed by a densification phase where roads get connected to each other. This model is also observed in VGI projects [13]. Y-axis of fig.5 is the slope of the curve between degree distribution and time (section2.3). X-axis represents different provinces plotted from the west to east of the country. They are grouped together, according to fig.3e, to reduce plot cluttering. Note that by plotting data points on x-axis in this way, whatever trend we observe from left to right will also explain the west to east trend within the country. Authors have intentionally kept the degree of distribution to be maximum 6 as higher degree of street junctions are not possible in real world scenarios [13]. It can be seen that for 1 (blue line) and 3 (grey line) degree junctions the slope of the plot is positive and negative, respectively. This shows

a high slope value between the degree distribution and time curve for western provinces and low slope for eastern provinces. These high slopes explain two events in western provinces: (1) Heavy mapping in short time interval, (2) Shift from exploration to densification phase. For eastern provinces the opposite is true. Only 1 and 3 degree junctions are studied as only they exist in majority in developing countries [70] and represent organic street layout [71]. Turkey OSM has followed the Exploration and Densification mechanism of evolution in given time span and this behaviour could be used to predict its future trend. [13] and [47] have reported similar trend for Beijing, China and Ireland, respectively.

3.4 Mapping behaviour of OSM mappers

Fig. 6 is a normalized plot between the total number of contributors in a region and its feature density for three different time-slices (2009, 2012 and 2015). The R^2 value, again, is a measure of the correlation between the two axes. It can be said that with time the correlation between the number of contributors and features density within the country is

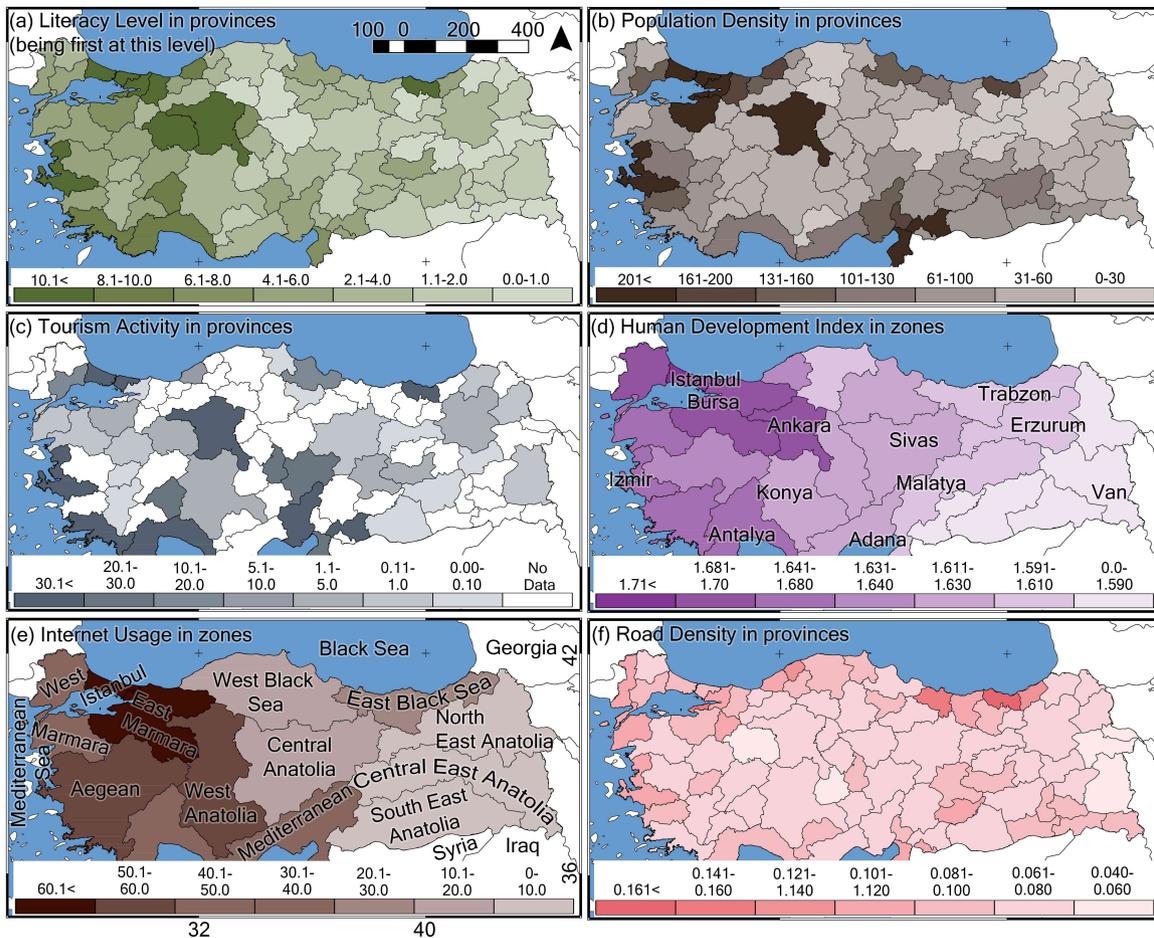


Figure 3: Maps showing the level of different socio-economic factors in the region. Color legends are meant for within map comparison and denotes normalized factors by area. The Road Density map shows a uniform distribution of road network within the country.

getting stronger. This is because of the recent popularity of OSM within geospatial community and change of its license to Obdl. For the year 2015, the correlation is quite strong. This proxy has also been discussed by other researchers who have reported a direct relationship between the total number of points and registered users in a region [21, 30]. It should be noted that bulk importers (section2.3) are not considered in this plot to keep the observation free from any individual bias. The R^2 value for Polygons for the year 2009 is quite high because of the cluttering of data points close to origin. This is because very limited contributors/mappers and features were there in OSM project in 2009 in Turkey. It is observed that although mappers do contribute few features as soon as they register to the project, usually they take a while to get started with serious contribution.

3.5 Quality of the dataset

Participation inequality by mappers is used as a proxy for dataset quality checkup [20]. Researchers have reported that bulk import by mappers make the dataset unsuitable for specialized use cases [21, 28]. Fig. 7 shows the fraction of bulk imports in Turkey-OSM dataset for different features. It can be seen that almost 75% of all geo-data upload corresponds to bulk imports by few mappers, by the end of 2015. There are reported 37 and 12 bulk importers for point/edge and polygon features, respectively. It should be noted that for points(tagged) no bulk importer is identified. This is because tagged geo-data is hard to generate or find open-source. The observed data quality of Turkey-OSM is poor and data correction and quality analysis are necessary before any specific use case.

Table 1 contains some basic information of top five bulk importers in the country. Nationality of these mappers gives an idea of their ground knowledge, Technical Background is important to know if they are versed

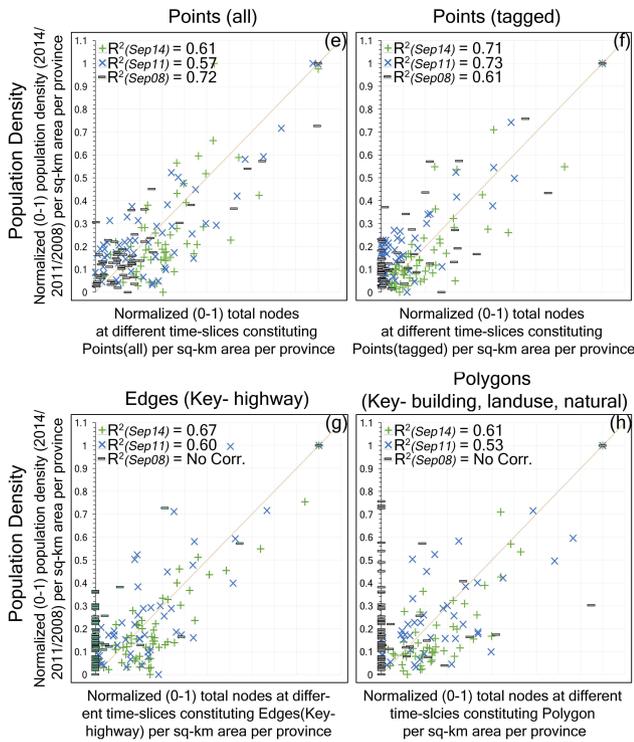


Figure 4: Graph between the Population Density and features density for different provinces at three different time-slices (2008, 2011 and 2014) showing an increasing correlation with time.

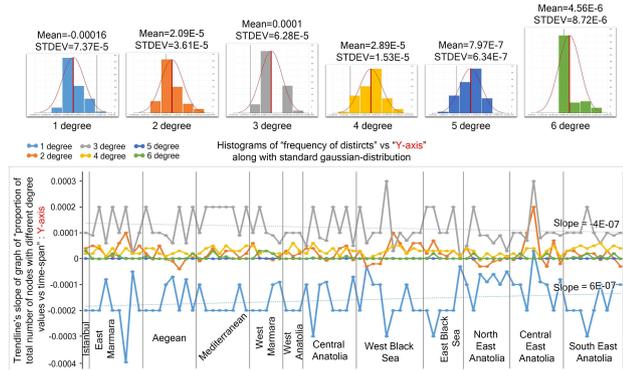


Figure 5: Graph showing the Exploration and Densification mechanism as followed by the street network dataset of the country.

with information and geospatial technology and Geo-data Sources to know their source of data. These information are collected by in-person communication with them. It should be noted that all of them are using satellite images of proprietary vendors, i.e. Google, Bing and Mapbox, for basemaps. They are also using geo-data from other VGI projects, like wikiloc and geofabrik, to upload into OSM. It shows an inter-dependency of different VGI projects for geo-data sources. It is, therefore, important to upload accurate data to VGI projects as it gets disseminated to

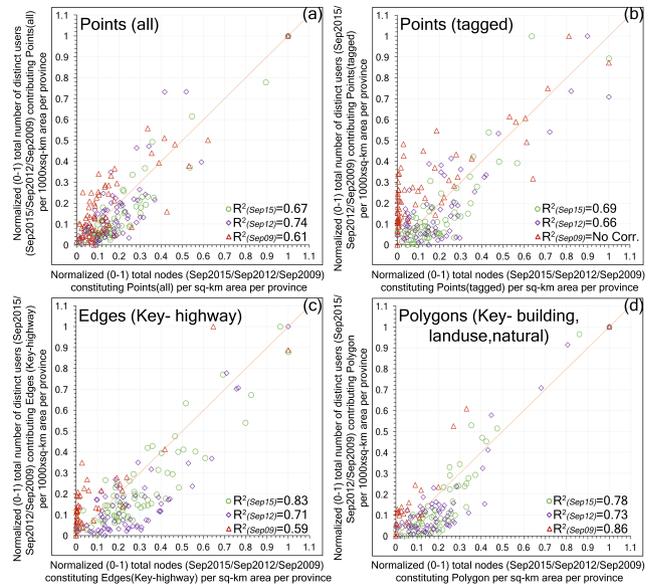


Figure 6: Graphs showing a good correlation between the total number of mappers in a region and its features density.

other projects as well, sometimes automatically. Two of the bulk mappers have used a more authentic geo-data source from local municipalities (Konya and Denizli Municipality). These two mappers are responsible for few data spikes in south-eastern province of the country, fig.2. In spite of having high participation inequality in the dataset, it is believed to be suitable for general use cases considering the varied data sources of these mappers.

4 Conclusion and Future Work

This study presents an analysis on the spatial evolution of Turkey-OSM dataset and its correlation with different socio-economic factors of the region, in an eight year time span (2007-2015). Five facets of the analysis are (a) how the dataset has spatially evolved in time?, (b) how the five identified socio-economic factors are related to this evolution?, (c) what evolutionary pattern does the street network dataset have followed?, (d) how the number of mappers are related to this evolution? and (e) how reliable the dataset is in terms of quality?

It is observed that the dump files of OSM do not contain any historical data earlier than 2007 because of the absence of object history feature in its Editing API. An exponential rise in data upload beyond 2012 is because of the change of its license from Creative Commons Attribution-ShareAlike 2.0 to Odbl, making it more open and flexible to end-users and developers. The exponential-step curve be-

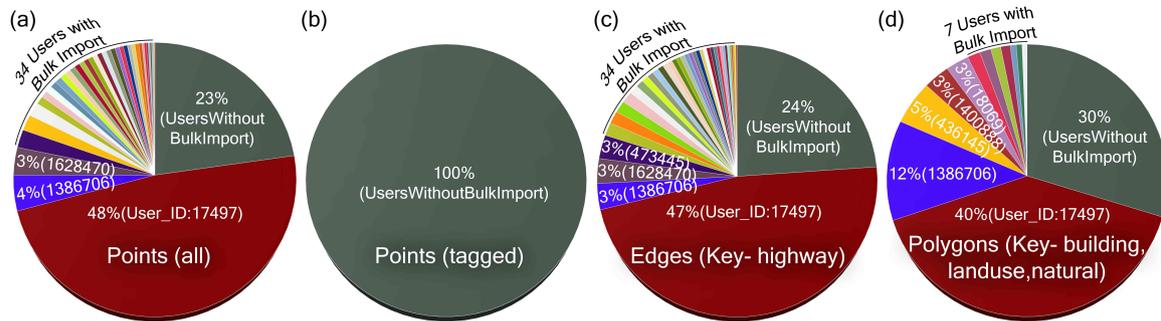


Figure 7: Pie charts showing the participation inequality in OSM data upload within the country, along with few bulk importers and their respective percentage contribution.

Table 1: General information about Turkey-OSM bulk importers

OSM User ID	OSM Username	Nationality	Technical Background	Geo-data Sources
17497	Roman	Germany	IBM System Programmer	- Bing-Maps and Mapbox satellite imageries to trace over. - .jpg images for boundary data. - .gpx files from Wikiloc.
1386706	Nesim Is	Turkey	Batman University Student	- Kentrehberi Konya Belediyesi. - Denizli Buyuksehir Belediyesi. - Google Street View. - In-person data collection. - Local community help. - Various literature sources. - Different websites.
18069	Claudius Henrichs	Germany		- Geofabrik Tools - OSM Inspector. - In-person data collection. - Mapped remotely using satellite images. - Publicly available data.
1400888	Summerson		GIS specialist	- Mostly polygons for buildings. - Bing satellite maps. - Online maps by municipalities. - Perform regular bulk imports.
436145	Penom	Turkey		- Regional Municipality for detailed mapping.

tween feature density and time is because of the increased mapping activity during summer each year. It is observed that, overall, there is an spatial bias in data upload from west to east of the country. *Population Density* and *Literacy Level* of the region are considered to be two good proxies for this bias, followed by *HDI*.

Country's OSM street network has followed the Exploration and Densification evolutionary model between the year 2007 and 2015. The network shows an organic street layout, which is typical for a developing nation. Along with the *Population Density* and *Literacy Level*, the number of mappers in a region is observed to be the third proxy for features density. It is observed that popularity rise and policies change play an important role in mappers behaviour toward mapping. Considerable participation inequality is observed in the dataset, with around 37 mappers responsible for country's 75% data upload, mainly

through bulk imports. It is found that different VGI projects share dataset with each other. It is, therefore, important to understand up to what extent different projects overlap to each other.

Future work might include identifying other strong socio-economic factors as proxies to predict the future of OSM in a region. This will help developers and project architects to better structure and plan policies for current and future VGI projects. Another study might be to identify the extent of overlap of different VGI projects and their possible source of geo-data error.

Acknowledgement: The research presented in this article is primarily funded by The Scientific and Technological Research Council of Turkey under the 2215 - Graduate Scholarship Programme for International Students (TUBITAK, URL: <https://www.tubitak.gov.tr/en>).

We also acknowledge financial support by Deutsche Forschungsgemeinschaft within the funding programme Open Access Publishing, by the Baden-Württemberg Ministry of Science, Research and the Arts and by Ruprecht-Karls-Universität Heidelberg.

References

- [1] Chris Anderson - The Long Tail, 2004 <http://www.longtail.com/about.html>
- [2] Coast S., Web and Wireless Geographical Information Systems - How OpenStreetMap Is Changing the World. Springer, 2011, 978-3-642-19172-5
- [3] Ma D., Sandberg M., Jiang B., Characterizing the Heterogeneity of the OpenStreetMap Data and Community. *ISPRS International Journal of Geo-Information*, 2015, 4(2), 535-550, DOI: 10.3390/ijgi4020535
- [4] Goodchild M. F., Geographic information systems and science: today and tomorrow. *Procedia Earth and Planetary Science*, 2009, 1(1), 1037-1043, DOI: 10.1016/j.proeps.2009.09.160
- [5] Li D., Qian X., A brief introduction of data management for volunteered geographic information. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/ Geomatics and Information Science of Wuhan University*, 2010, 35(4), 379-383
- [6] Heipke C., Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2010, 65(6), 550-557, DOI: 10.1016/j.isprsjprs.2010.06.005
- [7] Dodge M., Kitchin R., Crowdsourced cartography: mapping experience and knowledge. *Environment and Planning A*, 2013, 45, 19-36, DOI: 10.1068/a444484
- [8] Elwood S., Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS. *GeoJournal: An International Journal of Geography*, 2008, 72(3), 173-183, DOI: 10.1007/s10708-008-9186-0
- [9] Goodchild M. F., Citizens as sensors: the world of volunteered geography. *GeoJournal: An International Journal of Geography*, 2007, 69(4), 211-221, DOI: 10.1007/s10708-007-9111-y
- [10] Haklay M., Weber P., OpenStreetMap: User-Generated street Maps. *Pervasive Computing, IEEE*, 2008, 7(4), 12-18, DOI: 10.1109/MPRV.2008.80
- [11] Goodchild M. F., NeoGeography and the nature of geographic expertise. *Journal of Location Based Services*, 2009, 3, 82-96, DOI: 10.1080/17489720902950374
- [12] Haklay M., Singleton A., Parker C., Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, 2008, 2(6), 2011-2039, DOI: 10.1111/j.1749-8198.2008.00167.x
- [13] Zhao P., Jia T., Qin K., Shan J., Jiao C., Statistical analysis on the evolution of OpenStreetMap road networks in Beijing. *Physica A: Statistical Mechanics and its Applications*, 2015, 420, 59-72, DOI: 10.1016/j.physa.2014.10.076
- [14] OpenStreetMap stats report, 2016 http://www.openstreetmap.org/stats/data_stats.html
- [15] Introduction of usage limits to the Maps API, 2011 <http://googlegeodevelopers.blogspot.com.tr/2011/10/introduction-of-usage-limits-to-maps.html>
- [16] Zielstra D., Hochmair H. H., Neis P., Assessing the Effect of Data Imports on the Completeness of OpenStreetMap – A United States Case Study. *Transactions in GIS*, 2013, 17(3), 315-334, DOI: 10.1111/tgis.12037
- [17] Neis P., Zielstra D., Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap. *Future Internet*, 2014, 6(1), 76-106, DOI: 10.3390/fi6010076
- [18] Haklay M., How Good is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets. *Environment and Planning B: Planning and Design*, 2009, 37, 682-703, DOI: 10.1068/b35097
- [19] Leeuw J. D., Said M., Ortegah L., Nagda S., Georgiadou Y., DeBlois M., An Assessment of the Accuracy of Volunteered Road Map Production in Western Kenya. *Remote Sensing*, 2011, 3(2), 247-256, DOI: 10.3390/rs3020247
- [20] Haklay M., Basiouka S., Antoniou V., Ather A., How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, 2010, 47(4), 315-322, DOI: 10.1179/000870410X12911304958827
- [21] Neis P., Zielstra D., Zipf A., Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions. *Future Internet*, 2013, 5(2), 282-300, DOI: 10.3390/fi5020282
- [22] Neis P., Zipf A., Analyzing the Contributor Activity of a Volunteered Geographic Information Project – The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 2012, 1(2), 146-165, DOI: 10.3390/ijgi1020146
- [23] Mooney P., Corcoran P., Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS*, 2014, 18(5), 633-659, DOI: 10.1111/tgis.12051
- [24] Neis P., Goetz M., Zipf A., Towards Automatic Vandalism Detection in OpenStreetMap. *ISPRS International Journal of Geo-Information*, 2012, 1(3), 315-332, DOI: 10.3390/ijgi1030315
- [25] Goodchild M. F., Li L., Assuring the quality of volunteered geographic information. *Spatial Statistics*, 2012, 1, 110-120, DOI: 10.1016/j.spasta.2012.03.002
- [26] Girres J., Touya G., Quality Assessment of the French OpenStreetMap Dataset. *Transactions in GIS*, 2010, 14(4), 435-459, DOI: 10.1111/j.1467-9671.2010.01203.x
- [27] Barron C., Neis P., Zipf A., A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS*, 2014, 18(6), 877-895, DOI: 10.1111/tgis.12073
- [28] Mooney P., Corcoran P., Characteristics of Heavily Edited Objects in OpenStreetMap. *Future Internet*, 2012, 4, 285-305, DOI: 10.3390/fi4010285
- [29] Estima J., Painho M., Exploratory analysis of OpenStreetMap for land use classification. In: *GEOCROWD*, Orlando, Florida, USA, 2013
- [30] Zielstra D., Zipf A., A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In: *13th AGILE International Conference on Geographic Information Science 2010*, Guimarães, Portugal, 2010
- [31] Neis P., Zielstra D., Zipf A., The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007-2011. *Future Internet*, 2011, 4(1), 1-21, DOI: 10.3390/fi4010001
- [32] Hochmair H. H., Zielstra D., Neis P., Assessing the Completeness of Bicycle Trail and Lane Features in OpenStreetMap for the United States. *Transactions in GIS*, 2015, 19(1), 63-81, DOI: 10.1111/tgis.12081

- [33] Zielstra D., Hochmair H. H., Using Free and Proprietary Data to Compare Shortest-Path Lengths for Effective Pedestrian Routing in Street Networks. *Transportation Research Record: Journal of the Transportation Research Board*, 2012, 2299, 41-47, DOI: 10.3141/2299-05
- [34] Jia T., Jiang B., Measuring Urban Sprawl Based on Massive Street Nodes and the Novel Concept of Natural Cities. Cornell University Library, 2010
- [35] Mooney P., Corcoran P., Using OSM for LBS – An Analysis of Changes to Attributes of Spatial Objects: Advances in Location-Based Services. Springer, 2011, 978-3-642-24198-7
- [36] Amirian P., Basiri A., Gales G., Winstanley A., McDonald J., OpenStreetMap in GIScience - The Next Generation of Navigational Services Using OpenStreetMap Data: The Integration of Augmented Reality and Graph Databases. Springer, 2015
- [37] Azizan M. H., Lim C. S., Hatta W. A. L. W. M., Gan L. C., Application of OpenStreetMap Data in Ambulance Location Problem. In: 2012 Forth International Conference on Computational Intelligence, Communication Systems and Networks, Phuket, Thailand, 2012, 321-325, DOI: 10.1109/CICSyN.2012.66
- [38] Over M., Schilling A., Neubauer S., Zipf A., Generating web-based 3D City Models from OpenStreetMap: The current situation in Germany. *Computers, Environment and Urban Systems*, 2010, 34(6), 496-507, DOI: 10.1016/j.compenvurbsys.2010.05.001
- [39] Chen B., Sun W., Vodacek A., Improving Image-Based Characterization of Road Junctions, Widths, and Connectivity by Leveraging OpenStreetMap Vector Map. In: *Geoscience and Remote Sensing Symposium (IGARSS)*, Quebec City, Canada, 2014 IEEE International, 2014, DOI: 10.1109/IGARSS.2014.6947608
- [40] Li Q., Fan H., Luan X., Yang B., Liu L., Polygon-based approach for extracting multilane roads from OpenStreetMap urban road networks. *International Journal of Geographical Information Science*, 2014, 28(11), 2200-2219, DOI: 10.1080/13658816.2014.915401
- [41] Fritz S., McCallum I., Schill C., Perger C., Grillmayer R., Achard F., Kraxner F., Obersteiner M., Geo-Wiki.Org: The Use of Crowdsourcing to Improve Global Land Cover. *Remote Sensing*, 2009, 1(3), 345-354, DOI: 10.3390/rs1030345
- [42] Coast S., *Web and Wireless Geographical Information Systems - How OpenStreetMap Is Changing the World*. Springer, 2011, 978-3-642-19172-5
- [43] Budhathoki N. R., Haythornthwaite C., Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist*, 2012, 1-28, DOI: 10.1177/0002764212469364
- [44] Budhathoki N. R., Nedović -Budić Z., Bruce B., An interdisciplinary frame for understanding volunteered geographic information. *Geomatica*, 2010, 64, 11-26
- [45] Hagenauer J., Helbich M., Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *International Journal of Geographical Information Science*, 2012, 26(6), 963-982, DOI: 10.1080/13658816.2011.619501
- [46] Corcoran P., Mooney P., Characterising the metric and topological evolution of OpenStreetMap network representations. *The European Physical Journal*, 2013, 215(1), 109-122, DOI: 10.1140/epjst/e2013-01718-2
- [47] Corcoran P., Mooney P., Bertolotto M., Analysing the growth of OpenStreetMap networks. *Spatial Statistics*, 2013, 3, 21-32, DOI: 10.1016/j.spasta.2013.01.002
- [48] Zhang Y., Li X., Wang A., Bao T., Tian S., Density and diversity of OpenStreetMap road networks in China. *Journal of Urban Management*, 2015, 4(2), 135-146, DOI: 10.1016/j.jum.2015.10.001
- [49] Planet OSM, 2016 <http://planet.openstreetmap.org/planet/full-history>
- [50] Strano E., Nicosia V., Latora V., Porta S., Barthe M. I., Elementary processes governing the evolution of road networks. *Scientific Reports*, 2015, 2, DOI: 10.1038/srep00296
- [51] Geofabrik Downloads - OpenStreetMap Data Extracts, 2015 <http://download.geofabrik.de>
- [52] Overpass API - OpenStreetMap, 2016 <http://overpass-api.de>
- [53] OpenStreetMap Foundation - We are changing the license, 2012 https://wiki.osmfoundation.org/wiki/License/We_Are_Changing_The_License
- [54] Open Database License, 2015 http://wiki.openstreetmap.org/wiki/Open_Database_License
- [55] OpenStreetMap - API, 2015 <http://wiki.openstreetmap.org/wiki/API>
- [56] OpenStreetMap Elements, 2015 <http://wiki.openstreetmap.org/wiki/Elements>
- [57] TUIK (Turkish Statistical Institute), 2016 <https://biruni.tuik.gov.tr/bolgeselistatistik/sorguSayfa.do?target=degisken>
- [58] Republic of Turkey General Directorate of Highways - Inventory of State and Provincial Roads, 2015 <http://www.kgm.gov.tr/Sayfalar/KGM/SiteTr/Istatistikler/DevletYolEnvanteri.aspx>
- [59] Osmosis - A Java application for processing OSM data, 2015 <http://wiki.openstreetmap.org/wiki/Osmosis>
- [60] Osmium Tool - for OSM data processing, 2016 <http://osmcode.org/osmium>
- [61] Osm2pgsql - To convert OSM data to PostGIS-enabled PostgreSQL Database, 2015 <http://wiki.openstreetmap.org/wiki/Osm2pgsql>
- [62] Osm2postgresql - To simplify OSM data rendering with QGIS and other GIS/web servers, 2012 <http://wiki.openstreetmap.org/wiki/Osm2postgresql>
- [63] Osm2pgrouting - To import OSM data to pgRouting Database, 2016 <http://pgrouting.org/docs/tools/osm2pgrouting.html>
- [64] OpenStreetMap History Splitter, 2016 <https://github.com/MaZderMind/osm-history-splitter>
- [65] Github account containing python scripts, 2016 <https://github.com/Zia-/Turkey-OSM-Statistical-Analysis-Python-Scripts.git>
- [66] OpenStreetMap - Import/Catalogue, 2016 <http://wiki.openstreetmap.org/wiki/Import/Catalogue>
- [67] Cardillo A., Scellato S., Latora V., Porta S., Structural properties of planar graphs of urban street patterns. *Physical Review E*, 2006, 73(6), DOI: 10.1103/PhysRevE.73.066107
- [68] Human Development Report 2013 - The Rise of the South Human Progress in a Diverse World, 2013 http://hdr.undp.org/sites/default/files/hdr_2013_en_technotes.pdf
- [69] Lequiller F., Blades D., *Understanding National Accounts - Second Edition*. OECD Publishing, 2014

- [70] Which street pattern represents your continent?, 2013
<https://munsonscity.com/2013/10/09/>
- [71] Gersmehl P. J., The Language of Maps. Pathway in Geography Series. ERIC (Institute of Education Sciences), 1991