Understanding the Roles of Communities in Volunteered Geographic Information Projects

Peter Mooney and Padraig Corcoran

Abstract As a society we are now more connected than ever before. Citizens interact with each other and form virtual communities based on a common interest or being involved in certain cultural, political, intellectual, or other issues. Volunteered Geographic Information (VGI) is generated when citizens annotate content from social media and smart devices. OpenStreetMap is a famous example of a VGI project with a very large community of contributors. This chapter examines this community and investigates the types of contributors and interactions amongst members of OSM. Our results show that there are very small groups of individuals creating and editing over 85 % of all OSM objects in three case-study cities. Editing and contribution behaviour is mostly steady and consistent over time except during months where OSM 'mapping parties' occur or when freely available spatial data is bulk imported into OSM. The paper also provides results of analysis into the social interaction between contributors to OSM where we show that a very small number of contributors are actively editing and maintaining the data submitted by other contributors. This indicates that most contributors work exclusively on their own data and rarely edit the work of others. Overall this paper will be of interest to LBS practitioners considering using OSM as a source of spatial data for LBS applications.

P. Mooney (✉)
Department of Computer Science, National University of Ireland Maynooth (NUIM),
Co. Kildare, Ireland
e-mail: peter.mooney@nuim.ie; peter.mooney@tnuim.ie

P. Corcoran
School of Computer Science and Informatics, University College Dublin (UCD),
Dublin 4, Ireland

P. Mooney
Department of Computer Science, National University of Ireland Maynooth,
Maynooth, Ireland

1 Introduction

In today's information society citizens are more connected to each other than at any
other time in human history. The Internet, social media, and continued advances in
smart device technology and telecommunications have had a profound affect on
how citizens interact with each other. The Location-based Services (LBS) industry
has grown exponentially in recent years based on these types of interactions. These
factors have combined to see an unprecedented increase in the amount of user-
generated content (UGC) created and made available on the Internet today. Text
messaging, social media interactions, photos, video, blog entries, etc. are amongst
the most popular forms of UGC. Recently, with the widespread availability of
consumer GPS devices much of this UGC now contains spatial information
(embedded geographical coordinates, locational information, etc.).

A special form of UGC is Volunteered Geographic Information (VGI). Over-
views of VGI can be found in a number of recent papers (Dodge and Kitchin 2011;
Goodchild 2008; Mooney et al. 2010). VGI can range from geographical coordi-
nates automatically embedded in a digital photograph (Goodchild 2009) and made
available in some online repository to more complex forms of spatial data such as
annotated GPS tracks and trails (Mooney and Corcoran 2012; Neis et al. 2012).
OpenStreetMap (OSM) is a famous example of VGI on the Internet and in recent
years has been subject to analysis by many leading GIS researchers. Anyone can be
a contributor to OSM and these contributors form a very large community of
citizens collecting (and subsequently editing) spatial data. The OSM collaborative
model is not dissimilar to that of Wikipedia where a large online community
collaborates to collect knowledge and information to create an online encyclopedia.

This chapter focuses on examining the community of contributors in OSM using
an analysis of the historical database of contributions to OSM for three major cities
(Berlin, London, and Paris). The analysis in the chapter is based on providing
answers to the following research questions. (1) Are the contributors to OSM in
these cities actually interacting with each other in a collaborative manner to build
the OSM databases for their city? (2) If there is collaboration then how do we
quantify it and (3) what are the types of contributions members of OSM actually
make? In their analysis of editing behaviour in Wikipedia Iba et al. (2010) rec-
ommend that analysis of collaborative knowledge projects should focus on the most
prolific actors in these networks. Despite the potentially large size of collaboration
networks they still process small world properties. Our analysis primarily focuses
on the most frequent contributors to OSM in the three case-study cities.

The remainder of the chapter is organised as follows. Section 2 provides
an overview of the related literature on Volunteered Geographic Information.
Section 3 provides the experimental analysis of the OSM historical data for our

three cities. Section 3.1 describes the spatial data and contributor characteristics of these study areas. Section 3.2 then moves to investigate the overall editing behaviour of contributors in the study areas by analysing way creation and editing on a monthly basis over the entire history period. In Sect. 3.3 we describe how a social network data structure is extracted and built from the OSM history data using edit interactions between contributors to infer linkages between contributors. Section 3.4 attempts to classify the edit interaction behaviour of high ranking contributors in each of the cities. In Sect. 3.5 edit interactions are used to build and describe a collaborative social network for OSM in the cities. Section 4 is the final section in the paper where the key outcomes and findings are discussed in a review of the paper. The paper closes with Sect. 4.2 and a discussion of some of the issues that provide scope for future work on this topic.

2 Review of Related Literature

Volunteered Geographic Information (VGI), the term coined by Goodchild in (2008), is the recent empowerment of citizens in the collaborative collection of geographic information. OpenStreetMap (OSM) is a collaborative project to create a free editable map database of the world as is probably the most well known example of VGI. There are few reports published in the literature regarding the extraction of social network characteristics directly from VGI. In OSM there are consultation and collaborative discussions on Wikis and mailing lists (Budhathoki et al. 2010) and at "mapping parties" but these are not very easily quantifiable. From a Location-based Services viewpoint VGI offers opportunities for access to a vast array of data and information about the environment around us. The quality of the VGI is a major obstacle towards its more widespread adoption in application areas such as LBS. Goodchild (2009) argues that very different mechanisms will be required to ensure the quality of data volunteered by amateurs. Despite the absence of clear methodologies or approaches to ensuring the quality of VGI Goodchild remarks that since 2007 "there are now literally hundreds of Web services that collect, compile, index and distribute VGI content" (Goodchild 2009). For this reason that we are interested in exploring the social processes that occur in the collection of VGI. Coleman et al. (2010) argue that by understanding the motivations of contributors to VGI and their social interactions one might be able to better understand the decisions made around the quality of contributions. "The crowd" as a metaphor signifies the power that can emerge from a mass of individuals converging to tackle a set of tasks. In the virtual realm, a crowd can be drawn together across a widely distributed set of actors for little cost in order to tackle very large challenges (Dodge and Kitchin 2011) such as mapping the transportation networks, utility networks, built environment, and green spaces of an entire urban metropolis. From a social viewpoint McLaren (2011) suggests that quality assurance could be directly provided by members of the local communities who take direct responsibility for authenticity of data in their area.

Quality analysis of VGI requires a broad multifaceted approach (Mooney and Corcoran 2012a, b). While this paper does not focus on the geometric and semantic quality of geographic objects in OSM we feel that understanding the characteristics of contributors who generate OSM data is an important aspect towards making decisions regarding the quality of OSM data. Many researchers (Haklay et al. 2010; Over et al. 2010; Girres and Touya 2010; Mooney and Corcoran 2012) have shown that when a large number of contributors work on OSM for a specific area this usually leads to better quality data and a stable, well maintained, OSM dataset for that area. Previous studies have focussed on relationships between the number/quantity of contributors in a given area without focussing on (1) who the contributors are or (2) what are the dynamics of the community within which these contributors are working. Wikipedia offers the closest comparison to OSM in terms of a large crowd working on a collaborative knowledge project. Iba et al. (2010) analyses editing patterns of Wikipedia contributors using a social network analysis. They identify the most creative Wikipedia editors among the few thousand contributors who make most of the edits from a pool of millions of active Wikipedia editors. They identify the key category prolific authors who start and build new articles of high quality. Feldstein (2011) carried out an analysis of how Wikipedia articles are created. Felstein comments that common wisdom has it that the Wikipedia has been created by "the crowd". He argues that this does not hold at the level of article creation and "at least not in the sense that a large swarm of Wikipedia editors descends upon a blank topic page and, when the dust settles, a fully formed Wikipedia article appears". Felstein's pilot study suggests that the article creation process, at least, seems to more closely mirror the traditional writer/editor process than it does the "crowd as writer-editor".

Overall we can see that the social processes governing the creation and maintenance of spatial data in VGI are not well understood yet. The role, influence, and work performed by "crowd" as a whole and individually have not been quantified. The next section of the paper provides the analysis of the OSM history data for the three case-study cities (London, Berlin, Paris).

3 Experimental Analysis

The OSM history for three cities (London, Berlin, and Paris) was extracted from the complete OSM history file (http://wiki.openstreetmap.org/wiki/Planet.osm/full). This file is generated every 2–3 months and made available from the OSM website. The OSM history is encoded in OSM-XML format. We used polygons for London, Berlin, and Paris based on the EU Nomenclature of Territorial Units for Statistics (NUTS) coding to OSMOSIS and extracted the three cities. The resulting OSM-XML was then processed and stored in a Postgresql PostGIS database. The complete history for the three cities was extracted from early 2006 until March 2012.

Table 1 A summary of the overall spatial and contributor characteristics extracted from the historical analysis of OSM for London, Berlin, and Paris

Statistic	London	Berlin	Paris
Area (km^2)	1,572	891	1,368
Total contributors	2,795	2,848	1,168
Unique ways	419,801	347,556	280,765
Total way edits	876,743	897,610	5,21,637
Contributors ≤10 edits	1,738 (63 %)	1,571 (55 %)	774 (67 %)
Contributors ≤20 edits	2,002 (72 %)	1,852 (65 %)	865 (74 %)
Contributors ≥200 edits	250 (8 %)	333 (11 %)	101 (8 %)
Top 10 (ways created)	194,511 (46 %)	155,495 (44 %)	218,888 (77 %)
Top 10 (total edits)	365,984 (41 %)	425,905 (47 %)	367,228 (70 %)
Top 250 (ways created)	404,135 (96 %)	319,708 (91 %)	279,452 (99 %)
Top 250 (total edits)	762,766 (87 %)	822,591 (91 %)	516,501 (98 %)
Ways (1 vers)	228,222 (54 %)	176,665 (50 %)	155,724 (55 %)
Ways (vers ≥5)	33,230	48,152	12,745

3.1 Characteristics of the Study Areas

In Table 1 we provide a summary of some of the key characteristics from the database of all edits to OSM in Berlin, London, and Paris. There are over 419,000 unique ways in London, almost 350,000 unique ways in Berlin, and almost 300,000 ways in Paris.

In Berlin and London there are over 800,000 edits to ways (polygons and polylines) whilst there are just over 500,000 edits to ways in Paris. The number of distinct contributors to OSM in Berlin and London is almost 3,000 whilst the number for Paris is just over 1,000. We ranked all of the contributors in each city by the number of distinct edits to ways which they performed. In the three cities this resulted in a long-tailed distribution of a small number of people performing a large number of edits. We used the threshold of the top 250 ranked contributors (benchmarked from the London dataset). The top 250 in London performed exactly 200 or more edits. The top 250 in London are responsible for 87 % of edits, 91 % of edits in Berlin, and 98 % of edits in Paris. Within this group of the top 250 the top 10 contributors have performed 41 % of all edits in London, 47 % of edits in Berlin, and 70 % of edits in Paris. At the opposite end of the ranking the number of contributors performing only a small number of edits (≤10) is substantial. In London 63 % of contributors perform ten or less edits whilst this figure is 55 % in Berlin and 67 % in Paris. Overall, it is apparent that in each city a relatively small number of contributors to OSM are bearing the responsibility of the vast majority of all editing work. Focussing on the creation of ways (essentially the first representation of a given polygon or polyline in the OSM database) the statistics are very interesting. In London and Berlin the top 10 ranked contributors are responsible for the creation of over 40 % of all ways in those cities and over 75 % for Paris. When we look at the top 250 contributors, in terms of way

creation, they are responsible for 87, 91, and 99 % of all way creation in London, Berlin, and Paris respectively. This indicates that this group of contributors are of fundamental importance to OSM in these cities.

Each time a way is edited a new version is created in the OSM database. On initial viewing of Table 1 it might appear that all of the ways are being edited frequently. This is not the case. Over 50 % of all ways in the three cities are single version (ways are only created and never edited). These ways are created but are left untouched by their creator or other contributors. Edits can consist of modifications to geometry and/or editing of tagging and attribution contribution. Mooney and Corcoran 2012a, b) introduce the concept of 'heavily edited' OSM objects. Their analysis investigates the properties of those ways with ≥15 versions. In the table we see that a relatively small percentage of ways have 'high edit' characteristics. In this case we consider 'high edit' ways as those with ≥5 edits and at least two distinct contributors. We found that the threshold of ≥15 versions as outlined in Mooney and Corcoran 2012a, b) was too high and their analysis was performed on larger datasets. From this part of the analysis we can see a small subset of contributors performing the vast majority of editing work to OSM in the three cities.

3.2 Characteristics of Way Creation and Editing

In this section we will discuss the overall characteristics of the creation of new ways and the subsequent editing of existing ways in the three case-study cities. If one examines the current snapshot of OSM for any given area it is impossible to understand how the current map representation has evolved to it's current configuration.

In Fig. 1 we show a time series graph of total monthly edits and way creation in Paris over the entire period represented in the OSM history for the city. Up to the beginning of 2010 the pattern of collaborative mapping in Paris was one of low frequency way creation followed by slightly higher frequency editing. In the period of April 2010 to January 2012 the French National Cadastre dataset was imported into OSM. This resulted in thousands of way creations per-day and subsequent edits. OSM import rules mean that bulk imports such as this must be intelligently merged with existing data already in the OSM database. An editing campaign in late 2011 was instigated to fix some metadata/tagging problems with the import. In Fig. 2 we show a time series plot of the total monthly edits and way creation in Berlin. The London city time series is very similar. With the exception of a period of consecutive months in 2010 the amount of editing and maintenance of existing ways was greater than the monthly rate of way creation. There are two spikes in way creation in 2009 and 2011 most probably relating to smaller scale bulk imports—from ortho-photo tracing and the national post-code database. However the monthly rate of way creation is decreasing. This is to be expected as the number of 'new features' left for mapping diminishes rapidly. Berlin shows impressive and reasonably consistent total monthly rates for editing of existing
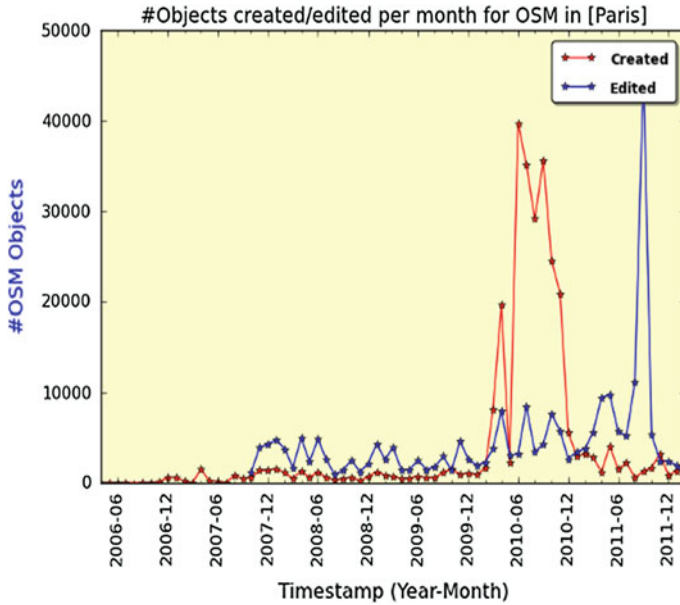
Fig. 1 A timeseries graph of total monthly edits and way (object) creation in Paris over the entire period of OSM history in the city

ways. This is very important as it can indicate that the OSM database for the city is being well maintained and kept up-to-date.

3.3 Creating a Social Network Structure in OSM from Edit Interactions

Usually, in social network analysis one infers linkages between nodes (actors, participants, etc.) by finding similarities in the characteristics of the nodes. If some measure of similarity is satisfied then there is a link or edge between those two nodes. For example in a Twitter conversation network there is a natural link between two Twitter users if they follow each other or retweet each other's tweets. In Flickr or Facebook networks linkages are most naturally between users who 'like' each other or are 'contacts' with others. A citation network can be considered as a specific form of a social network based on the citation patterns of academics. In Wikipedia where contributors collaborate to edit an article on a given subject a link is inferred if (1) two users edit the same article, (2) have conversation in the talk pages of the article wiki, or (3) edit each other's work on the chapter. In OSM it is not easy to build a social network data structure directly from the characteristics of contributors. Such information is not available for automated extraction. Contributors in OSM do not "follow" each other as in social
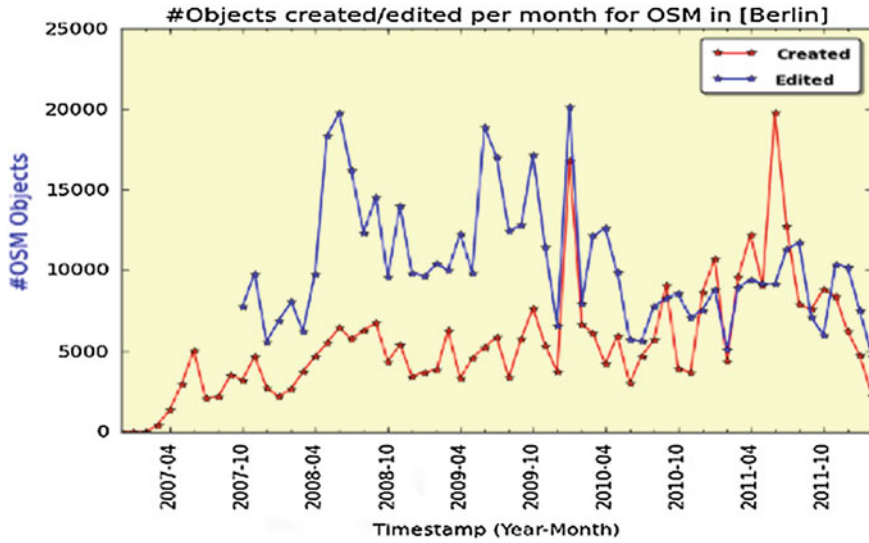
Fig. 2 A timeseries graph of total monthly edits and way (object) creation in Berlin over the entire period of OSM history in the city

networks such as Facebook. There are no explicit linking mechanisms between contributors in OSM expressed in any OSM datasets. Finally, unlike Wikipedia, there is no hierarchy in contribution to OSM. Some OSM contributors, after a long period of time contributing to OSM, become 'de-facto' moderators or administrators. However this does not bestow any special rights on them and any large scale changes or edits (deletions, reverts of potential vandalism, etc.) to the OSM database must be done with the consensus of the wider OSM community. In this analysis we infer a linkage between two contributors if they have *co-edited* a way.

We define co-editing as follows. The historical data for all of the contributions performed by a given contributor A are extracted from the database. Suppose a contributor A edits the work of contributor B on some way object O then $s(A, B, O) = 1$ where A creates the version of O labeled O_{t+1} where B created the version of O labeled O_t. Then O_t and O_{t+1} are consecutive versions of the same object O. We call this an Edit Interaction (EI). In the next section we use EI as a means of understanding the social network characteristics of the contributor network to OSM in the three case-study cities.

3.4 Edit Interaction Behaviour Amongst Contributors

We computed all of the EI for every contributor in the top 250 contributor list for each of the cities. The EI of the top 250 contributors for each of the three cities were summarised under four distinct classifications based on the types of edits

Table 2 Results of clustering and classification of edit interactions (EI) of the top 250 contributors to London, Berlin, and Paris

City	Only created	Geometry edits	Tagging only	Geometry and tagging	Unclassified
London	27	98	81	35	9
Berlin	18	68	28	120	16
Paris	13	90	37	101	9

performed namely: 'Only Created' where this contributor only created a way as an edit and performed no further editing. Inherent in the creation of a way is the association of tags with the first version of the way. Creation of a way is a special form of edit; 'Geometry Edit' is where a contributor potentially created a way but only performed geometry edits on that way. This also includes the case where $s(A, B, 0) = 1$ and A only edited or changed the geometry of O as created by B, 'Tagging only' is where a contributor only performed edits on the tags associated with a way object, and finally 'Geometry and Tagging' is where the edits are both related to changes in the geometry of the way and the associated tagging metadata. The scores for these four classes are stored in an EI description vector for each contributor. We applied k-means clustering to this dataset with k = 4. Bayesian classification was used to assign contributors to a specific classification. We carried out a process of manual verification to ensure that the classification was as expected. In Table 2 we tabulate the results of the clustering and classification of the EI of the top 250 contributors for each city.

In Table 2 the majority of the unclassified contributors are at the lower end of the top 250 contributors ranking where their EI cannot be easily classified as belonging to one of the specific four classes. There are a number of interesting observations from the results in Table 2. London has the largest number of contributors who 'Only Create' ways and those contributors whose edits are predominantly 'Tagging Only'. It could be argued that 'Tagging Only' requires the least amount of work in terms of contribution effort. Berlin has the largest number of contributors (almost 50 % of the top 250 contributors) whose EI are predominantly 'Geometry and Tagging'. Berlin also has the largest number of contributors (28) whose EI are predominantly 'Tagging Only'. Paris appears to have top 250 contributors EI very similar to that of Berlin.

3.5 Social Network Characteristics of the Edit Interactions

Using the EI data from the previous section we built a social network graph for the top 250 contributors for the three cities. An edge exists between two contributors where contributor A creates the version of O labelled O_{t+1} where contributor B had created the version of O labelled O_t. Table 3 tabulates the summary of the social network characteristics of the Edit Interactions (EI) for the top 250 contributors to each of the three cities in our case study.

Table 3 The social network characteristics of the edit interactions (EI) for contributors in our three case study cities

Network characteristic	London	Berlin	Paris
Nodes (contributors)	1,788	1,986	1,054
Percentage of all contributors	64 %	70 %	90 %
Edges	15,765	16,483	9,763
Strongly connected components	1,541	1,762	836
Mean betweeness centrality	0.12	0.15	0.08
Mean eigenvalue centrality	0.08	0.11	0.07
Mean degree	8.817	9.12	6.38
Observations	12 nodes with BC > 0.45	16 nodes with BC > 0.51, 8 nodes with EC > 0.25	3 nodes with BC > 0.61

In Table 3 we show the results of three important social network characteristics. Firstly, Eigenvalue Centrality (EC) is very similar conceptually to Google Page Rank. The EC metric returns a value [0...1] for any node in the network. EC is somewhat recursive in its definition as high scores are given to nodes if they are connected to other nodes which are also themselves important in the network. In terms of social interaction the concept is such that a contributor's influence in the network is proportional to the total influence of the contributors to whom they are connected. Secondly, Betweeness Centrality (BC) is a measure of a node's centrality in a network. Suppose for a node X then the BC of X is equal to the number of shortest paths from all nodes to all other nodes that pass through X. BC is a useful measure of a node's importance to the network overall rather than simply a measure of it's connectivity. In terms of social networking the concept of BC is that if two contributors do not know each other but have share a contributor who edits both their work then it is very likely that they will be connected to each other soon and edit each other's work directly. Finally, we compute the number of strongly connected components in the network. A strongly connected component is a subgraph of a network where there is a path from every node in the subgraph to every other node in the subgraph. Strongly connected components have been shown to correspond to smaller, tightly knit, groups of nodes in a social network—for example a group of friends with common interest or working on a specific task/job. BC, EC, and strongly connected components are well known network statistics and consequently are available in most leading graph handling packages.

In the case of each EI social network in Table 3 the number of nodes is greater than 250. The EI interactions for the top 250 contributors are computed and contributors not in this subset are included because they have had their work edited by one of the top 250.

Fig. 3 The edit interactions social network for the top 250 contributors to openstreetmap in Berlin. For visualisation clarity the number of EI must be ≥10 for a graph edge to be visible on the graph. The nodes are weighted corresponding to their betweeness centrality (*blue low–red high*)
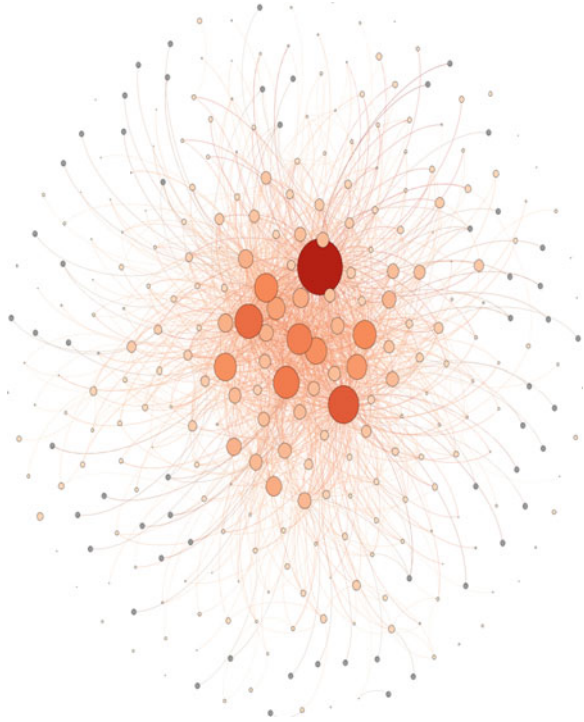

Figure 3 shows the social network graph of the EI for the top 250 contributors in Berlin. The size of the nodes are weighted by their betweeness centrality values (blue nodes have very low BC while red nodes have high BC). It is clearly evident that there are several very dominant contributors who are interacting with many other contributors in terms of EI. For visualisation purposes and to make the linkages between contributors with higher BC more apparent we have omitted edges in the network where the number of EI between two contributor nodes is less than 10.

4 Conclusions and Future Work

The analysis outlined in this chapter is based on working with the entire edit history of OSM for three cities. The type of retrospective analysis carried out would not be possible using only the current snapshot of OSM which is available in real-time for download. Overall there are strong indications that there is collaboration amongst the OSM community in the three cities studied. However it is difficult to ascertain at this moment if this collaboration is intentional or incidental. Is contributor A editing the work of contributor B through a process of mutual collaboration or are these edits performed without interaction between A and B?
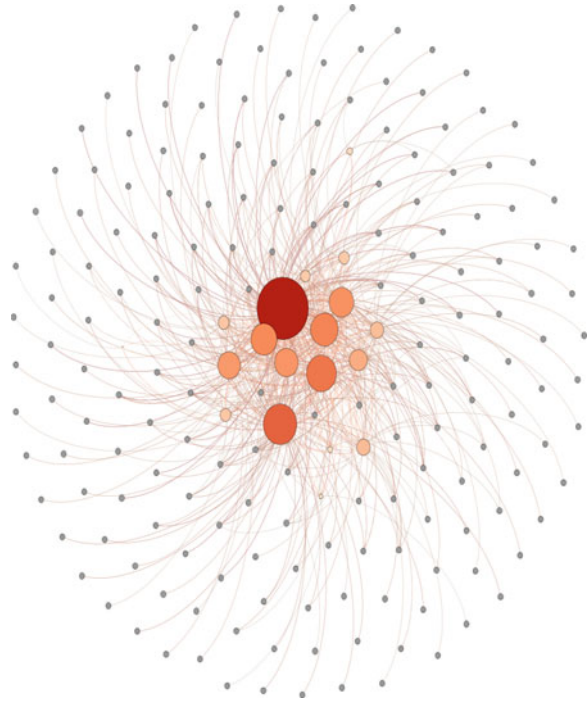
This analysis shows that a small number of dedicated contributors are responsible for a very large amount of the total OSM work in these cities over the entire period of their history. In the remainder of this section we shall review the key outcomes and findings from our work.

4.1 Review of Key Findings

Table 1 provided an overview of the key characteristics of the three cities investigated in our case study. The information in Table 1 is only accessible through analysis of the history of the OSM project over a long period of time. London and Berlin are very similar for all of the characteristics discussed. Paris displays some different types of characteristics—a small contributor base in comparison to the other two cities but these contributors appear to carry out more work in total. However Paris has been the subject to a significant bulk import. The three cities have very similar results for: the overall percentage of ways with only a single version, the overall percentage of contributors with ≤10 edits performed, and the percentage of ways created by the top 250 ranked contributors. Figures 1 and 2 shows the total monthly edits and way creation for Paris and Berlin respectively. In particular Paris shows a dramatic way creation phase in 2010 which then changes to a sustained period of way editing in 2011 (as explained by the bulk import of data). Berlin, on the other hand, displays increases and decreases in the editing and way creation processes in periodic fashion. This shows two important characteristics of VGI for use in LBS applications and services. VGI, such as OSM, can undergo periods of sustained editing or maintenance while also experiencing periods where there is a large influx of newly created data. In the case of Berlin and London (figure not shown) total monthly way creation is decreasing overall. There are less 'new things to map' and the job for OSM contributors in cities such as these changes from 'white space filling on the map' to maintenance of the OSM database in terms of: updating metadata for features and changes in environmental geometry (demolished buildings, changes in road/street network geometry, etc.).

  Table 2 shows that the principal contributors to all of the three cities are similar in terms of how they contribute and edit OSM. For Berlin and Paris there appears to be a preference for "geometry only" and "geometry and tagging" behaviours while in London there appears to be more focus from contributors on "way creation" and then subsequent "tagging" of ways. Table 3 provides a summary of the social network characteristics of the Edit Interactions of the top 250 contributors in the three cities. Visualisations of two of these characteristics are shown in Figs. 3 and 4. Crucially, this part of the analysis demonstrates that a small number of contributors (depicted by larger sized nodes in Figs. 3 and 4) are involved in collaborative editing work with a large number of other contributors. Their importance to the collaborative editing process in OSM is indicated by their high scores from the social network characteristics of BC and EC. BC is used to weight

Fig. 4 The edit interactions social network for the top 20 contributors (from the top 250 contributors) in London. The social network shows that this special subset of contributors are heavily involved in the editing of ways created by other contributors. In this network G = (197,693)

nodes in Figs. 3 and 4 respectively. For the three cities over 64 % of the total number of contributors are involved, to some degree, in the collaborative process of map editing as inferred by the EI of the top 250 contributors. Berlin and London contain a large number of strongly connected components within this social network data structure. The mean BC and EC for the three cities is low which is expected given the high percentage of contributors performing a small number of edits.

4.2 Issues for Further Work

The results outlined in the paper have successfully provided insightful answers to the research questions outlined at the beginning of the paper. However, there are still a number of areas where further work is required. Automated inference of links (edges) based on commonalities in contributor profiles (for example: contributor A lives in the same part of Berlin as contributor B) or contributor A is "friends" with contributor B is not currently possible as these types of information are not made available in the OSM datasets. We intend to conduct a survey of OSM contributors in London to find biographical details which can be used to

construct a social network data structure. The properties of these graphs could be compared to the graphs automatically generated from the OSM history.

We are currently undertaking research into extending our definition of co-edits to include a spatial component—that is contributors who co-edit ways in the same geographical area (council area, postcode, borough, etc.) are more likely to be linked. This could help us understand if local groups (strongly connected components or cliques) form when we weight co-edit linkages between contributors more heavily if they are spatially related. The influence of mapping parties (gatherings of small numbers of OSM contributors on a particular date to collaborative map an area) will be investigated in this analysis. We chose the top 250 ranked contributors as the threshold for analysis in this study. We feel that this is adequate for an analysis of this type for a city/urban area. As discussed earlier there are a very large percentage of contributors (just under 75 %) who perform a relatively small number of edits (≤20). We feel that by considering a greater number of contributors from the overall ranking may not provide us with any additional insight into the collaboration amongst the OSM community for city/urban areas. As immediate future work we shall be updating our analysis for other cities/regions. Rather that analyse a fixed subset of the contributor population we shall use a relative subset (top 5 %, top 10 %, etc.) to ensure that the results are consistent across different OSM contributor bases.

Whilst the quality of the data generated by these contributors is beyond the scope of this study a longitudinal study could investigate the quality (geometric, semantic, spatio-temporal, etc.) of the OSM data as the social network characteristics of the contributor network changes. Girres and Toura (2010) and Haklay et al. (2010) amongst others have investigated and commented on the effects of increased numbers of contributors on overall data quality in OSM. However these studies were performed on a static snapshot of OSM rather than an analysis over a longer period of time.

Finally we see in Figs. 1 and 2 that there is an overall decrease in the map editing and creation in OSM in Paris and Berlin (and London). This is coupled with the information in Table 1 where we see a small number of contributors taking responsibility for a very large amount of work. Is this actually sustainable going forward. The sustainability of VGI projects such as OSM is a critical issue for LBS developers if they choose VGI as a source of spatial data. McLaren (2011) asks that "with so many crowd sourced sites contending for the attention of the citizen, will fatigue and lack of interest over time make citizen contributions a scarce resource?" Cuff et al. (2008) warn that "today's exotic and disturbed data collection practices may appear banal 10 years hence". They emphasis that to maintain citizen interest in urban sensing projects and VGI the user interface to applications and management systems are crucial. They comment that "if individuals are motivated to participate user interfaces are critical to both data collection and interpretation". If it is too difficult or costly to contribute volunteered information people will simply avoid the hassle. While this is not directly a topic discussed in this paper the LBS community have a part to play in development of

better or improved user-interfaces to allow even greater participation, using mobile devices, in the collection and maintenance of VGI.

References

Budhathoki NR, Nedovic-Budic Z, Bruce B (2010) An interdisciplinary framework for understanding volunteered geographic information. Geomatica J Geospatial Inf Sci Technol Pract 64(1):14–29

Coleman D, Sabone B, Nkhwanana N (2010) Volunteering geographic information to authoritative databases: linking contributor motivations to program effectiveness. Geomatica 64(1):383–396

Cuff D, Hansen M, Kang J (2008) Urban sensing: out of the woods. Commun ACM 51(3):24–33

Dodge M, Kitchin R (2011) Mapping experience: crowdsourced cartography. Social Sciences Research Network 4:55–80

Feldstein A (2011) Deconstructing wikipedia: collaborative content creation in an open process platform. Procedia Soc Behav Sci 26:76–84, ISSN 1877-0428, doi:10.1016/j.sbspro.2011.10.564

Girres JF, Touya G (2010) Quality assessment of the French openstreetmap dataset. Trans GIS 14(4):435459

Goodchild MF (2008) Whither VGI? GeoJournal 6(72):239–244

Goodchild M (2009) NeoGeography and the nature of geographic expertise. J Location Based Serv 3(2):82–96

Haklay 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 J 47(4):315–322

Iba T, Nemoto K, Peters B, Gloor PA (2010) Analyzing the creative editing behavior of wikipedia editors: through dynamic social network analysis. Procedia Soc Behav Sci 2(4):6441–6456

Mclaren R (2011) Crowdsourcing support of land administration: a new, collaborative partnership between citizens and land professionals. Technical report, Royal Institution of Chartered Surveyors commissioned by RICS Land Group, RICS, Parliament Square, London SW1P 3AD United Kingdom

Mooney P, Corcoran P, Winstanley AC (2010) Towards quality metrics for openstreetmap. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, GIS'10, ACM, San Jose, pp 514–517

Mooney P, Corcoran P (2012a) Characteristics of heavily edited objects in openstreetmap. Future Internet 4(1):285–305

Mooney P, Corcoran P (2012b) The annotation process in openstreetmap. Trans GIS, http://dx.doi.org/10.1111/j.1467-9671.2012.01306.x

Neis P, Zielstra D, Zipf A (2012) The street network evolution of crowdsourced maps: openstreetmap in Germany 2007–2011. Future Internet 4(1):1–21

Over M, Schilling A, Neubauer S, Zipf A (2010) Generating web-based 3D city models from openstreetmap: the current situation in Germany. Comput Environ Urban Syst 34(6):496–507