

Who are the contributors to OpenStreetMap and what do they do?

Peter Mooney¹, Padraig Corcoran²

¹Department of Computer Science, National University of Ireland Maynooth, Co. Kildare. Ireland
Tel. +353-1-708 3847 Fax +353-1-708 3848

peter.mooney@nuim.ie

²School of Informatics and Computer Science, University College Dublin, Belfield, Dublin 4. Ireland

padraig.corcoran@ucd.ie

Summary: Social network analysis (SNA) is the mapping and measuring of relationships between people, groups, organizations, computers, URLs, and other connected information/knowledge entities.

OpenStreetMap offers a unique global collaborative generated and maintained spatial dataset. This paper presents some results of analysis of the social networking aspects of OpenStreetMap. The OSM history database for London is used for analysis. We find that the contributor network of OSM exhibits social network characteristics.

KEYWORDS: OpenStreetMap, History, Social Networks, Web GIS, VGI

1. Introduction

OpenStreetMap (OSM) is the most famous example of Volunteered Geographic Information (VGI) (Goodchild; 2007) on the Internet today. Using a collaborative crowd-sourced model for spatial data collection and management OSM has grown to become a truly global dataset (Mooney and Corcoran; 2011c). Our paper at GISRUUK 2011 (Mooney and Corcoran; 2011b) demonstrated how the annotation process in OSM can be investigated through analysis of historical OSM data. However this paper only investigated specific features (“heavily edited features”). We have extracted the entire history of OSM edits for London, UK. In this paper we present some characteristics of contributors to OSM London and apply some social network analysis techniques to these contributions.

2. OSM Contributors: Case-study of London

In this paper we have extracted the history of all OSM contributions to London extending to the M25 motorway. There are a total of 3, 811, 876 nodes and a total of 876, 743 ways (polygons or polylines). The history contains edits to London from April 2005 to October 2011. There are 2, 795 unique contributors to London OSM over this period. Figure 1 shows the spatial distribution of all edits by the top 20 contributors from January 2011 to October 2011 where each individual contributor is given a distinct node colour. It is interesting to observe that contributors work in geographic clusters, such as the orange in the south east and yellow, red, and purple in London city and along the Thames. In Table 1 we summarise the top 20 contributors to OSM in London over this period. Table 1 indicates the number of edits made by each of these contributors, the month and year of their first edit in OSM London, the number of changesets created, and the number of ways they created where they were the first contributor. These 20 contributors have made a total of 481, 401 edits (55% percent of all edits to ways). There are 419, 801 distinct ways in the London database. These 20 contributors were the creators of 255, 222 of these ways (over 61%). They represent a very important group of contributors in the OSM community. Over 72% of contributors made 20 edits or less to the London OSM.

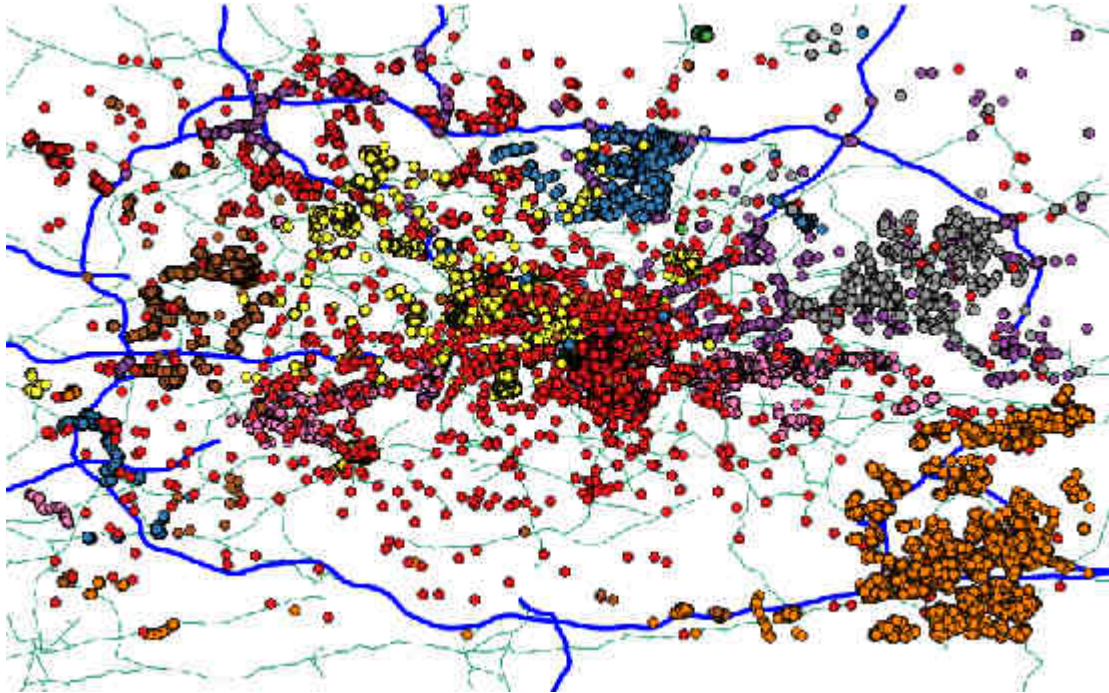


Figure 1: Spatial distribution of edits by the top 20 contributors to OSM London. Each contributor is given a different colour point.

3. Social Networking Analysis in OSM

No explicit social network representation is available for OSM. Contributors do not “follow” or “be friends” with other contributors as is common in social media such as Twitter or Facebook (Lewis et al.; 2008). If we suppose that the contributions to OSM were modelled as a graph $G = (V,E)$ where V is the set of all vertices and represent the OSM contributors and E is the set of edges connecting the vertices. As Kolaczyk (2009) suggests this as a problem of network topology inference where one must investigate other aspects of the data to infer edges between vertices. We used a simple concept of “co-edits” as a scoring method which we feel links well to the collaboration aspect of OSM. Two contributors i and j are connected if they have both edited the same (way) W , under some conditions. Then $S(i, j)$ returns the total number of ways (PW) that i and j have both edited. $S(i, j)$ is assigned as the cost of the edge $e(i,j)$ between i and j . There are some interesting graph statistics that are computed by researchers in social network analysis. A clique in an undirected graph $G = (V,E)$ is a subset of the vertex set C in V such that for every two vertices C_i and C_j in clique C there exists an edge connecting C_i and C_j . In social networking this can indicate a measure of “social cohesion” amongst groups where Falzon (2000) defines it as “the maximal subnetwork containing three or more actors all of whom are connected to each other”. Betweenness Centrality (BC) ranks nodes by how many shortest paths between other nodes they are on (van Duijn and Vermunt; 2006). High betweenness represents a single point of failure in a network but also an influence over what happens in a network. Eigenvalue Centrality (EC) gives high scores to nodes if they are connected to many other nodes that are themselves important in the network and is similar Google Page Rank and is a natural way of defining important clusters (van Duijn and Vermunt; 2006). We will now discuss three examples of social network graphs for the top 20 contributors to OSM London.

The top 20 contributors to OpenStreetMap in London ordered by total number of edits of ways.

Rank	Name	OSMID	Edits	Changesets	1st Edit	Created
0	TimSc	6809	74,161	1742	03 – 2007	42,879
1	Ed Avis	31257	46,346	7385	09 – 2008	27,332
2	Tom Chance	346	40,233	963	09 – 2006	23,320
3	Urban Ramble	88718	39,353	1467	01 – 2009	14,060
4	8on	1238	35,628	1291	12 – 2005	25,831
5	dpisping	28024	33,464	551	02 – 2008	7,240
6	Steve Chilton	736	32,537	1404	04 – 2006	19,153
7	Paul Todd	12503	28,138	150	09 – 2007	12,897
8	Welshie	508	22,013	1647	09 – 2005	12,685
9	Blumpy	64226	14,111	218	10 – 2008	9,114
10	SomePlace Mapper	31795	13,124	1630	03 – 2008	4,512
11	V-na-more	210115	13,008	515	12 – 2009	8,529
12	OJW	1337	12,313	491	05 – 2006	10,719
13	Joel	16703	12,063	924	10 – 2007	4,612
14	TomH	3980	11,431	488	09 – 2006	8,210
15	PeterITO	4951	11,237	963	04 – 2007	3,369
16	Derick Rethans	37137	11,158	429	08 – 2009	8,514
17	RandomJunk	4049	10,806	394	10 – 2006	7,819
18	Sladen	12671	10,143	111	10 – 2007	2,245
19	Wyndale	118262	10,134	530	04 – 2009	2,182

Table1: Rank ranges from 0 to 19 where 0 indicates first,1 indicates second etc

Example 1: No Threshold: $S(i, j) > 0$

Using any co-edit as inference of an edge the resultant graph G is shown in Figure 2. The graph has $V = 20$ vertices and $E = 189$ edges with average degree 18.9. G is one edge short of being fully connected. There are two cliques in the graph with average size 19 nodes. BC and EC are very similar for all nodes at 0.20 to 0.31 indicating that all contributors are co-editing together.

Example 2: Threshold: $S(i, j) \geq 100$

A more realistic concept of co-editing exists where two contributors i and j co-edit a large number of W above some threshold for $S(i, j)$. When this threshold is applied the network dramatically changes as illustrated in Figure 3(a) where $S(i, j) \geq 100$. Now $G = (V, E)$ has 20 nodes and 105 edges with an average degree of $G = (V, E)$ now 10.5. 13 cliques appear in the graph with the average clique size of 7. Contributors ranked 0, 1, and 15 are connected to 16 or more contributors. Contributor 0 and 1 have highest BC (0.157 and 0.11). The importance to the network of 0,1, and 15 are shown by EC values of 0.3, 0.297, and 0.29 respectively.

Example 3: Temporal Threshold

In Figure 3(b) a temporal threshold is applied so that $S(i, j)$ returns the number of W where i and j co-edited within 1 month of each other. This more realistically represents collaborative editing. Now $G = (V, E)$ has 20 nodes and 78 edges with an average degree of 7.8. 27 cliques now appear in the graph with an average clique size of 4. Contributors 9, 16, and 12 are connected to 3 nodes or less potentially indicating the edit their own areas only. Contributor 6 has highest BC (0.15) followed by 0 (0.11) and 10 (0.096). Contributor 0 continues to be an important node with highest EC (0.36) followed by contributor 1 (0.311) and 10 (0.31).

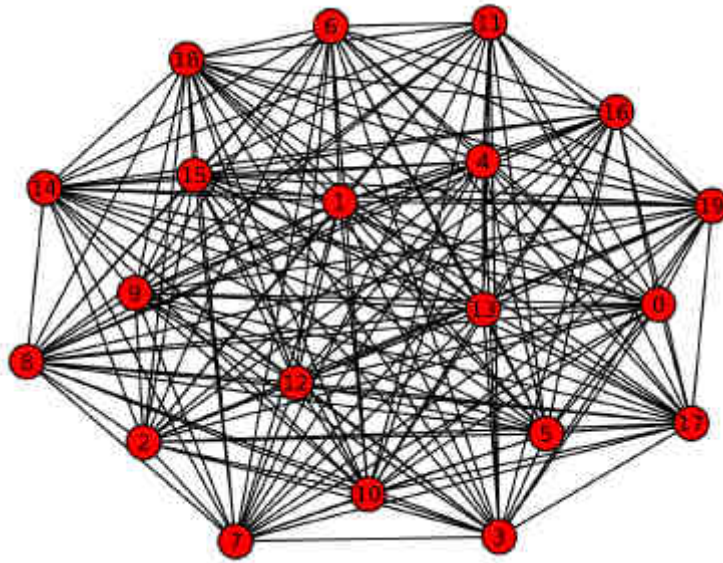


Figure 2: The co-edit social network of the top 20 contributors to OpenStreetMap in London. An edge between vertices (contributors) indicates that these contributors co-edited at least one way. No threshold is set for $S(i, j)$

4. Conclusions and Future Work

Fu et al. (2008) remarks that there is growing interest and concern regarding the topological structure of new online social networks. Despite their potentially large size they possess small-world and scale-free features. In their analysis of the social network in Wikipedia Iba et al. (2010) recommend focus on "prolific authors who start and build articles of high quality" from the thousands of other Wikipedia editors. We have focused on prolific contributors to OSM in London. As shown in Table 1, 20 contributors (less than 0.01% overall) are responsible for greater than 50% of all edits to London OSM over a 6 year period. However these 20 contributors created over 61% of London's ways yet are only responsible for the first time creation of 7, 204 Points of Interest (POI) nodes from a total of 59, 978 or 12%. No explicit social network is integrated into OSM. Our paper has highlighted social network characteristics amongst the top 20 contributors. By inferring links between contributors to form a graph metrics such as betweenness centrality and eigenvalue centrality illustrate the importance of certain contributors to the overall network. In our immediate future work we are currently investigating methods to carry out dynamic community discovery in terms of key life cycle events in the project (such as "TimSC" leaving (TimSC; 2011), or a large influx of new contributors). Do these events cause the expansion or contraction of "small world" OSM communities? Finally, our future work will look at both qualitative and quantitative approaches to investigation of whether there is an OSM equivalent of social cohesion as defined by direct social network links and what it actually means in practice?

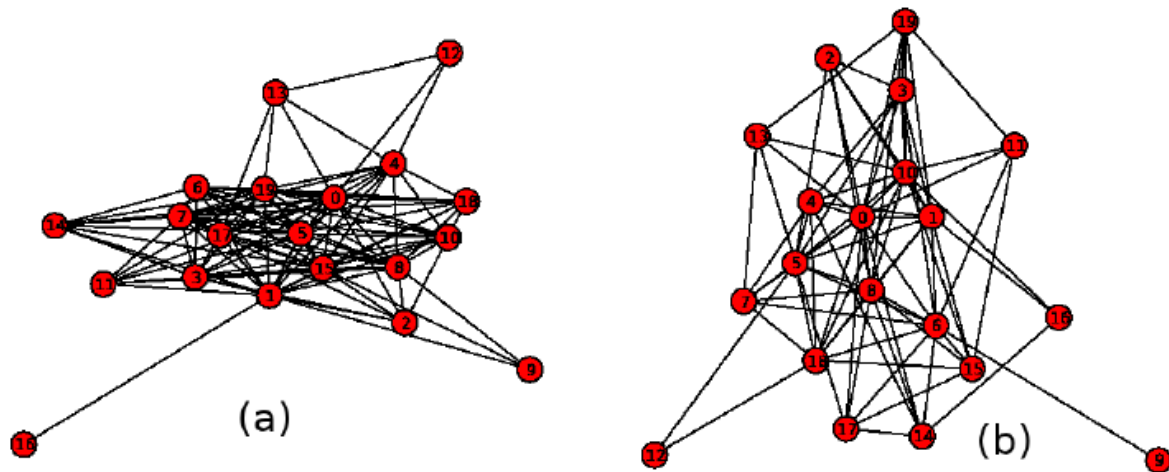


Figure 3: The co-edit social network of the top 20 contributors to OpenStreetMap in London. In Figure (a) an edge between vertices (contributors) indicates that these contributors co-edited at greater than or equal to the threshold for $S(i, j) = 100$ co-edits. In Figure (b) an edge between vertices (contributors) indicates they co-edited features within 1 month of each other

7. References

- AltogetherLost (2011). OpenStreetMap members, nodes, and ways statistics, Webservice <http://osmstats.altogetherlost.com/>.
- Budhathoki, N. R., Nedovic-Budic, Z. and Bruce, B. (2010). An interdisciplinary framework for understanding volunteered geographic information, *Geomatica, The Journal of Geospatial Information Science, Technology and Practice*. **64**(1): 14–29.
- Cristani, M., Bazzani, L., Paggetti, G., Fossati, A., Tosato, D., Bue, A. D., Menegaz, G. and Murino, V. (2011). Social interaction discovery by statistical analysis of f-formations, *Proceedings of the British Machine Vision Conference*, BMVA Press, pp. 23.1–23.12.
- Falzon, L. (2000). Determining groups from the clique structure in large social networks, *Social Networks* **22**(2): 159 – 172.
- Fu, F., Liu, L. and Wang, L. (2008). Empirical analysis of online social networks in the age of web 2.0, *Physica A: Statistical Mechanics and its Applications* **387**(2-3): 675 – 684.
- Goodchild, M. (2007). Citizens as sensors: the world of volunteered geography, *GeoJournal* **69**: 211–221.
- Haklay, M. (2010). How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets, *Environment and Planning B: Planning and Design* **37**(4): 682 –703.
- Iba, T., Nemoto, K., Peters, B. and Gloor, P. A. (2010). Analyzing the creative editing behavior of wikipedia editors: Through dynamic social network analysis, *Procedia - Social and Behavioral Sciences* **2**(4): 6441 – 6456.
- Kolaczyk, E. (2009). *Statistical Analysis of Network Data*, Springer Science+Business Media.
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. and Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using facebook.com, *Social Networks* **30**(4): 330 – 342.
- Mooney, P. and Corcoran, P. (2011b). Annotating spatial features in openstreetmap, *Proceedings of*

the 19th annual gis research uk (gisruk) 2011, Portsmouth, England.

Mooney, P. and Corcoran, P. (2011c). Can volunteered geographic information be a participant in eEnvironment and sdi?, in J. Hrebíček, G. Schimak and R. Denzer (eds), *Environmental Software Systems.Frameworks of eEnvironment*, Vol. 359 of *IFIP Advances in Information and Communication Technology*, Springer Boston, pp. 115–122.

TimSC (2011). Timsc - statement regarding leaving the openstreetmap project, User Wiki-pages on OpenStreetMap <http://wiki.openstreetmap.org/wiki/User:TimSC/Quit>.

van Duijn, M. A. and Vermunt, J. K. (2006). What is special about social network analysis?, *Methodology:European Journal of Research Methods for the Behavioral and Social Sciences* **2**(1): 2 – 6.

8. Biography

Dr. Peter Mooney is a research fellow at the Department of Computer Science NUI Maynooth and he is funded by the Irish Environmental Protection Agency STRIVE programme (grant 2008-FS-DM-14-S4). Dr. Padraig Corcoran is a lecturer and post-doctoral researcher at the School of Computing and Informatics at University College Dublin and is funded under the IRCSET EMBARK programme. The authors gratefully acknowledge these supports.