The Life Cycle of Volunteered Geographic Information (VGI) Contributors: the OpenStreetMap Example

D. Bégin[1], R. Devillers[1,2], S. Roche[2]

[1] Department of Geography, Memorial University, St. John's (NL), Canada
Email:{d.begin; rdeville}@mun.ca

[2] Centre de recherche en géomatique, Université Laval, Québec (QC), Canada
Email: stephane.roche@scg.ulaval.ca

1. Introduction

The Web 2.0 changed the way Internet users interact with knowledge (Gore 1998; Goodchild 2007) by allowing knowledge sharing through various online systems (e.g. Wikipedia). In GIScience, Volunteered Geographic Information (VGI) has attracted the attention of scholars due to its ability to crowdsource geographic information potentially useful in many contexts (Haklay 2014; Arsanjani *et al.* 2015).

Classifications of VGI contributors have been proposed, based on users' motivation (Coleman *et al*. 2009) or on the volume of their contributions (Panciera *et al.* 2010; Neis and Zipf 2012). Existing studies show that the nature of the contributions broadens with the time spent in a project (Kim 2000; Panciera *et al.* 2009) but none clearly linked them to the timespans of the different stages in the life cycle of contributors. This paper presents the first detailed analysis of the time over which contributors participate to a VGI project by using OpenStreetMap (OSM) data, identifying sets of contributors that share similar temporal patterns of contributions, and discussing the potential impacts on contributions.

2. Contributors' Timespan Distribution

While OSM data can be accessed by anyone, only registered users can edit the database. Once registered, no mechanism identifies users that stop contributing to the project. We define a 'registered user' as someone that created an OSM account, while a 'contributor' is a registered user that started at least one editing session (i.e. a changeset). 'Contributors' timespan' refers to the timespan between a contributor's first and last edit.

All the transactions made in OSM until September 1, 2014, were extracted and loaded into a PostgreSQL 9.3 database. Statistical analyses and visualizations were performed using the R 3.2.1 software.

A first analysis compared cumulative OSM registered users with actual contributors, creating daily Contributors/Registered Users ratios (Figure 1). Ratios reveal wide variations over time, ranging from 6% to 47%, for an average of 30.9%. Results support Neis and Zipf (2012) findings that only a third of registered users eventually become contributors.

A complementary cumulative distribution function (CCDF) of contributors' timespan was also generated (Figure 2). It represents the proportion of contributors who edited the database for a similar period of time or longer. Five pivotal points were identified based on this figure and on additional analyses.

A first pivotal point is found at about one hour of contributions, where 15% of participants stopped contributing in a matter of seconds. This abrupt break in the curve represents new contributors that made only a few edits, or even none, before the OSM API automatically closes their one and only editing session left idle for an hour.

The proportion of OSM users who contributed data keeps decreasing rapidly for about an hour then it slows down until it reaches our second pivotal point after 24 hours (one day). Analyses show that 60% of contributors did not edit data beyond this point, a proportion

similar to what was found in Wikipedia (Panciera *et al.* 2009).

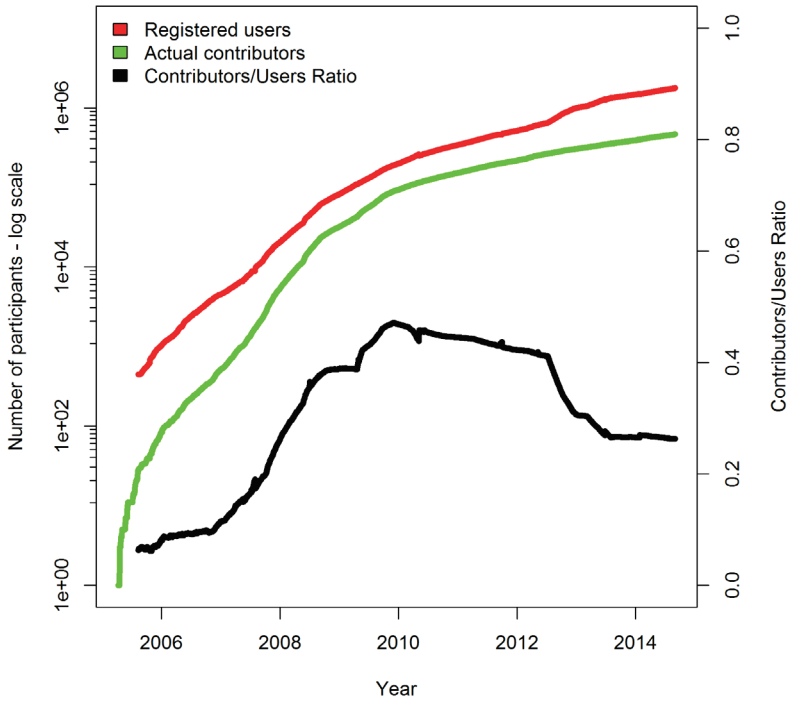

Figure 1: Participation in the OSM project over time.

The proportion of contributors then follows a regular (log based) decline for about a year after which it slows down significantly. Fewer than 20% of participants contributed beyond a first year, our third pivotal point. The fourth point is about five years later and concerns less than 1% of participants.

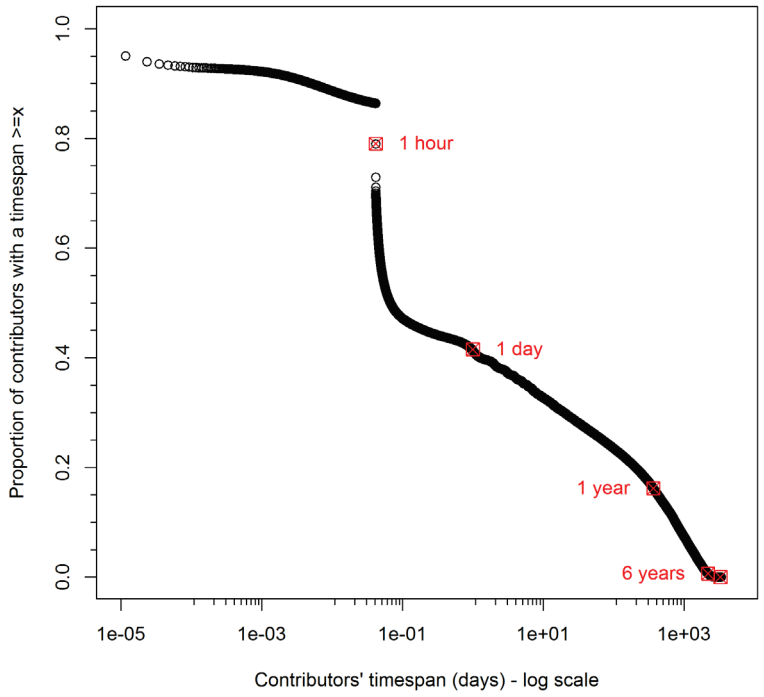

Figure 2: Complementary cumulative distribution function (CCDF) of contributors' timespan showing the proportion of contributors by timespan (in days).

The fifth and last point is located at nine years, a time at which only a few dozen accounts were active. Many of OSM developers and innovators (Beal and Bohlen 1957) have then kept editing the database over the years.

3. Evolution of Edits Over the Life Cycle of Contributors

The contributions made at each stage of the life cycle are described in Table 1. The table shows the proportion of participants who reached each stage among the members of their cohort ('Users'), the volume of edits made for the 10^{th}, 50^{th} and 90^{th} percentiles (P*), the average number of days actually involved in contributing ('Days'), and the average number of days spent between contributions ('Idle').

Table 1: Life cycle stages of OSM registered users and contribution metrics.

Name	Starts after	Lasts for	Users (%)	P10	P50	P90	Days	Idle
Lurkers	NA	NA	69.1	NA	NA	NA	NA	NA
Contributors	*See breakdown below*		30.9	0	16	1.7e3	11	17
Contributors' stages	…	…	…	…	…	…	…	…
Visitors	0 second	1 hour	21.0	0	2	42	1	NA
Novices	1 hour	23 hours	37.8	0	6	93	1	NA
Amateurs	1 day	364 days	25.1	4	107	2.6e3	6	13
Adherents	1 year	5 years	20.0	15	716	3.7e4	46	21
Veterans	6 years	3 years	13.2	373	2.3e4	4.4e5	253	8
Elders	9 years	NA	33.3	3.8e3	4.6e4	3.8e5	320	9

The *lurkers* represent the registered users that have not contributed to OSM yet (69.1%). The remaining 30.9% of registered users that have at least started an editing session are *contributors*. Contributors' life cycle stages are described above.

Visitors and *Novices* contribute little to OSM, with a median contribution (P50) around five edits. Further analyses of the data indicate that the few most productive visitors and novices accounts were dedicated to large data imports, with tens of thousands of edits done within this short period of time.

Amateurs and *Adherents* have similar contribution profiles. The volume of edits from the lower half (P50) of contributors increases proportionally to the time they spent in the project but this proportion doubles for the most active ones (P90).

Veterans and *Elders* show different contribution profiles, with larger and more evenly distributed contributions, a behaviors also observed in Wikipedia (Geiger and Halfaker 2013). Furthermore, these participants contribute on average twice more often than those from all other stages, with an average of one day out of nine (idle).

Overall, the proportion of participants who keep contributing within a cohort is slightly higher than what was found within Wikipedia (Zhang et al. 2012).

4. Conclusion

This paper reported on an analysis of the time over which OSM users contribute to the project. Pivotal points found on the cumulative complementary distribution function show six different phases in the life cycle of contributors.

By assessing each phase, we found that contributions are more frequent, larger, and evenly distributed against participants that spent more than five years in the project. We also found that for each phase, about 20% of participants keep contributing and move to the next phase. Since these core contributors produce most of the data over a long period of time, we

expect that the data quality either improves or becomes increasingly uniform over the years.

Further analysis should assess if core contributors can be identified from their early behaviors in order to keep them engaged by designing crowdsourcing activities adequately.

Acknowledgements

The authors would like to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) for their financial support through the Discovery Grant of the second author.

References

Arsanjani, J.J., Zipf, A., Mooney, P., Helbich, M., 2015. OpenStreetMap in GIScience, In: Lecture Notes in Geoinformation and Cartography, Cartwright, W., Gartner, G., Meng, L. and Peterson, M.P. (Eds.), Springer, Switzerland, pp. 324.

Beal, G.M., Bohlen, J.M., 1957. The diffusion process, Iowa State College, Ames (USA).

Coleman, D.J., Georgiadou, Y., Labonté, J., 2009. Volunteered geographic information: The nature and motivation of producers. International Journal of Spatial Data Infrastructures Research, 4, pp. 332-358.

Geiger, R.S., Halfaker, A., 2013. Using edit sessions to measure participation in Wikipedia, In: Proceedings of the 2013 conference on Computer supported cooperative work, pp. 861-870.

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

Gore, A., 1998. The Digital Earth: Understanding our planet in the 21st Century. Australian surveyor, 43(2), pp. 89-91.

Haklay, M., 2014. OpenStreetMap studies (and why VGI not equal OSM). Po Ve Sham – Muki Haklay's personal blog (August 14), pp.1-2.

Kim, A.J., 2000. Community building on the web: Secret strategies for successful online communities, Addison-Wesley Longman Publishing,.

Neis, P., Zipf, A., 2012. Analyzing the Contributor Activity of a Volunteered Geographic Information Project—The Case of OpenStreetMap. ISPRS International Journal of Geo-Information, 1(2), pp. 146-165.

Panciera, K., Halfaker, A., Terveen, L., 2009. Wikipedians are born, not made: a study of power editors on Wikipedia, In: Proceedings of the ACM 2009 international conference on Supporting group work, pp. 51-60.

Panciera, K., Priedhorsky, R., Erickson, T., Terveen, L., 2010. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki, In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1917-1926.

Zhang, D., Prior, K., Levene, M., 2012. How long do Wikipedia editors keep active? In: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration, pp. 4.