

Volunteered Geographic Information Users Contributions Pattern and its Impact on Information Quality

Amir Masoud Forati ¹, and Rina Ghose ^{1,*}

¹ aforati@uwm.edu

* rghose@uwm.edu

Keywords: VGI; Crowdsourcing; data quality; OSM; Volunteered geographic information; user participation; contribution pattern; OpenStreetMap

1. Introduction

The quality of user generated data through crowdsourcing activities has profound impacts on online mapping projects and VGI based platforms [4,5]. Spatial data quality can be ascertained by examining it through a set of criteria such as positional accuracy, thematic accuracy, timing precision, logical consistency, and completeness. Using these measures, multiple studies have been conducted to assess the quality of VGI data [6-9]. The voluntary nature of VGI contributions by individuals who may be untrained in cartography/GIS raises questions regarding the quality and reliability of these data [10]. Thus, it is necessary to design and implement an effective evaluation system to assess VGI data quality. In this study, we examine VGI participation level in OpenStreetMap through three case study neighborhoods located in three major cities: Tehran, London, and Los Angeles.

3. Research Questions, Data and Methodology

In this study, we use three metrics to examine (a) the level of participation in an area (b) local knowledge contribution pattern in OpenStreetMap.

1. The number of tags per object: This metric shows the level of completeness of map features. It implicitly conveys the level of local knowledge in the dataset, as tags are mostly provided by local users. The higher the number of tags per object, the higher the level of detail in a map.
2. The number of objects per user: This shows the level of users' participation and contribution pattern. The goal is to determine how many users provided information on how many objects. This helps us to examine how and to what degree local knowledge exists in the dataset. The higher number of objects per user, the less local knowledge in the dataset.
3. The number of users in a square kilometer: It explicitly conveys the participation level in the area, where a higher number of users in a square kilometer is an indication of both greater participation and a higher quality dataset.

Using these metrics, we compare the level of participation and knowledge contribution patterns in three cities: London (U.K), Tehran (Iran), and Los Angeles (USA). In each city, we selected a representative neighborhood as our study site. These study sites are West Kensington (U.K.), Westwood (USA), and Shekoufeh (Iran). We then obtained the dataset for each neighborhood from OSM on May 22, 2020, and analyzed the data based on the proposed metrics. The findings are displayed in Table 1.

Table 1. OSM participation in three neighborhoods

City	Neighborhood	objects	tags	users	tags/objects	objects/users	users per km ²
LA	Westwood	57063	174	148	0.003	385.5608108	45.96
London	West Kensington	16306	289	168	0.017	97.05952381	63.63

Tehran	Shekoufeh	7888	87	83	0.011	95.03614458	22.80
--------	-----------	------	----	----	-------	-------------	-------

Among the three cities, Westwood (L.A.) reports the highest number of objects, 57063 objects. In contrast, West Kensington (London) has 16306 objects, and Shekoufeh (Tehran) has 7888 objects. The high number in Westwood's case is likely caused by the City Hub LA project, through which detailed data sets have been incorporated in OSM. This includes Los Angeles County's excellent building footprints data set from L.A. County GIS Portal, and the Los Angeles County Land Use dataset. Thus, the LA OSM dataset shows a robust spatial dataset at an outstanding level of completeness. Next, we assess participation and the use of local knowledge through an examination of tags per feature. Tags depict detailed information about a particular map feature, therefore features with more tags contain more detailed and complete spatial knowledge. Creating tags requires detailed experiential and local knowledge that may be absent in the authoritative database. The higher number of tags in a feature is an indication of more detailed information about that feature in the dataset, and each new tag is a step towards improving the quality of the feature. In terms of number of tags per object, we find West Kensington to rank first (0.017), followed by Shekoufeh (0.011), while Westwood (0.003) ranks a distant third. This indicates that the London OSM map is rich in local knowledge and is, therefore, more detailed and complete. Tehran's OSM map has an acceptable level of tags per object in comparison with London, which means it is making good progress in incorporating local knowledge. Westwood in Los Angeles shows a surprisingly low number of tags, as 100 objects on its map have only three tags. Thus, while the LA OSM map has benefited from the City Hub LA project in incorporating official data sets, it lacks the local knowledge needed to produce a robust spatial database. We next examine participation in terms of the number of users per area, in West Kensington (London), 168 users created the whole data set. Further, 19.54% percent of users provided 90% of the entire data. In the case of Shekoufeh (Tehran), 83 users created 7888 objects, but three users created 70% of the entire dataset. Westwood, with 148 users, shows a similar pattern. The number of objects per user metric implies the level of user's participation and to what degree local knowledge exists in the dataset. Our findings show that West Kensington (97.06 objects/user) and Shekoufeh (95.03 objects/user) rank well, while Westwood (L.A.) is a distant third (385.56 objects/user). Further, London and Tehran have a similar level of local knowledge in their OSM dataset; in contrast, the L.A. OSM dataset lacks detailed local knowledge due to the lower participation of users. We find that London ranks first with 63 users/square km, followed by LA with 45 users/square kilometer. Tehran ranks third, with 22 users/square km, indicating that the level of participation in Tehran is very low in comparison to London or L.A. This may be caused by the lack of OSM based services in Iran, leading to its low level of popularity among citizens.

3. Discussion

According to our statistical investigation, 5.50% of users in Westwood (L.A.) have created 90% of the whole dataset. Similarly, about 20% of users in West Kensington (London), and Shekoufeh (Tehran) have produced 90% of the entire dataset, with three Shekoufeh contributors creating 70% of the dataset. Nonetheless, the contribution of the rest is significant in terms of local knowledge integration. In Shekoufeh, 47 users have created local knowledge-based features (such as a short alley, local bank name, pharmacy, dentistry, and a car shop), while 21 of whom have done only one edit in the system. West Kensington (London) and Westwood (L.A.) follow a similar pattern. In West Kensington, eight users have provided 80% percent of the whole dataset. Nevertheless, the rest play a significant role in the provision of local knowledge; 54 users have provided features which need local knowledge, and 30 users have done only one edit in the system. Yet, these 30 users provided details on features such as Exhibition Houses, bookstore, car sharing, supermarket, and London underground station name. Similarly, in LA, 90% of the whole dataset was provided by seven users, but 82 users provided local knowledge-based features, and 36 of them have created only one feature/tag such as high school, supermarket, café shop, convenience shop, and gas station. These contributions define critical public places from a local or experiential perspective that significantly enriches the database. This makes the crowdsourced information unique, UpToDate, and complete;

however, existing methodologies cannot assess the quality of these features; due to limitations of available VGI data quality assessment methodologies, quality evaluation of very important and defining VGI features which have less than 15 versions and created by the user with no performance history is impossible.

The advantage of VGI over authoritative maps provided by the federal and municipal agencies is that the former incorporates local knowledge while the latter does not. Because the available methodologies are not able to evaluate the quality of all OSM features in these areas, there is a need to design new methods or advance current methods so that we can determine the quality of local knowledge. This is also necessary as the number of monthly new registered OSM users is increasing, and many new contributors are registering and starting to create, edit, and use OSM. As of May 2020, up to 27,000 new contributors are registering per month. Yet, the dominant VGI quality assessment methods are not capable of assessing the quality of information created or edited by these new users or users with limited history, since existing methods assess VGI quality through users' reputation which in turn relies on users' contributions history. The inability of existing approaches to evaluate the quality of VGI with a low level of participation (limited number of versions), as well as the VGI generated by new users, underscore the need to provide an alternative to overcome these limitations.

Such challenges can be overcome through new modes of inquiry. By including these characteristics in the process of VGI quality assessment, the limitation of existing methods can be, to some extent, covered [32, 37]. Still, there is a need for further research on quality assessment of VGI, so that the VGI system with a low level of participation and information generated by new users or users with limited history is properly assessed.

References

1. Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221. <https://doi.org/10.1007/s10708-007-9111-y>
2. Wilson, M., & Graham, M. (2013). Neogeography and volunteered geographic information: A conversation with Michael Goodchild and Andrew Turner. *Environment and Planning A*, 45(1), 10-18. <https://doi.org/10.1068/a44483>
3. Elwood, S., Goodchild, M. F., & Sui, D. Z. (2012). Researching volunteered geographic information: Spatial data, geographic research, and new social practice. *Annals of the association of American geographers*, 102(3), 571-590. <https://doi.org/10.1080/00045608.2011.595657>
4. Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. *World Medical & Health Policy*, 2(2), 7-33. <https://doi.org/10.2202/1948-4682.1069>
5. Estima, J., & Painho, M. (2019). Photo based Volunteered Geographic Information initiatives: A comparative study of their suitability for helping quality control of Corine Land Cover. In *Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 1124-1142). IGI Global. <https://doi.org/10.4018/978-1-5225-8054-6.ch048>
6. Fonte, C. C., Antoniou, V., Bastin, L., Estima, J., Arsanjani, J. J., Laso Bayas, J-C., See, L., & Vatsava, R. (2017). Assessing VGI Data Quality. In G. Foody, L. See, S. Fritz, P. Mooney, A-M. Olteanu-Raimond, C. Costa Fonte, & V. Antoniou (Eds.), *Mapping and the Citizen Sensor* (pp. 137-163). Ubiquity Press Ltd.. <https://doi.org/10.5334/bbf>
7. Koukoletsos, T., Haklay, M., & Ellul, C. (2011, July). An automated method to assess data completeness and positional accuracy of OpenStreetMap. In *GeoComputation* (Vol. 3, pp. 236-241).
8. Helbich, M., Amelunxen, C., Neis, P., & Zipf, A. (2012). Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. *Proceedings of GI_Forum*, 4, 24.
9. Dorn, H., Törnros, T., & Zipf, A. (2015). Quality evaluation of VGI using authoritative data—A comparison with land use data in Southern Germany. *ISPRS International Journal of Geo-Information*, 4(3), 1657-1671. <https://doi.org/10.3390/ijgi4031657>
10. Flanagan, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3-4), 137-148. <https://doi.org/10.1007/s10708-008-9188-y>

11. Ostermann, F. O., & Spinsanti, L. (2011, April). A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management. In *Proceedings of AGILE* (Vol. 2011, pp. 1-6).
12. Haworth, B. T., Bruce, E., Whittaker, J., & Read, R. (2018). The good, the bad, and the uncertain: Contributions of volunteered geographic information to community disaster resilience. *Frontiers in Earth Science*, 6, 183. <https://doi.org/10.3389/feart.2018.00183>
13. Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of producers. *International journal of spatial data infrastructures research*, 4(4), 332-358. <https://doi.org/10.2902/1725-0463.2009.04.art16>
14. Van Exel, M., Dias, E., & Fruijtjer, S. (2010, September). The impact of crowdsourcing on spatial data quality indicators. In *Proceedings of the GIScience 2010 Doctoral Colloquium*, Zurich, Switzerland (pp. 14-17).
15. 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), 146-165. <https://doi.org/10.3390/ijgi1020146>
16. Rehrl, K., Gröechenig, S., Hochmair, H., Leitinger, S., Steinmann, R., & Wagner, A. (2013). A conceptual model for analyzing contribution patterns in the context of VGI. In *progress in location-based services* (pp. 373-388). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34203-5_21
17. Mooney, P., & Corcoran, P. (2014). Analysis of Interaction and Co-editing Patterns amongst Open Street Map Contributors. *Transactions in GIS*, 18(5), 633-659. <https://doi.org/10.1111/tgis.12051>
18. Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, 18(6), 877-895. <https://doi.org/10.1111/tgis.12073>
19. Adler, B. T., & De Alfaro, L. (2007, May). A content-driven reputation system for the Wikipedia. In *Proceedings of the 16th international conference on World Wide Web* (pp. 261-270).
20. Mooney, P., & Corcoran, P. (2012). Characteristics of heavily edited objects in OpenStreetMap. *Future Internet*, 4(1), 285-305. <https://doi.org/10.3390/fi4010285>
21. Nielsen, J. 2006. Participation inequality: Encouraging more users to contribute. *Jakob Nielsen's alertbox*, 9: 2006.
22. Hill, W. C., Hollan, J. D., Wroblewski, D., & McCandless, T. 1992. Edit wear and read wear. In: *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. ACM: pp. 3-9. <https://doi.org/10.1145/142750.142751>
23. Haklay, M. E. (2016). *Why is participation inequality important?* Ubiquity Press.
24. Bégin, D., Devillers, R., & Roche, S. (2013, May). Assessing volunteered geographic information (VGI) quality based on contributors' mapping behaviours. In *Proceedings of the 8th international symposium on spatial data quality ISSDQ* (pp. 149-154). <https://doi.org/10.5194/isprsarchives-XL-2-W1-149-2013>
25. Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge* (pp. 105-122). Springer, Dordrecht. https://doi.org/10.1007/978-94-007-4587-2_7
26. OSMstats. Available online: <https://wiki.openstreetmap.org/wiki/Stats> (accessed on 25/5/2020).
27. Girres, J. F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4), 435-459. <https://doi.org/10.1111/j.1467-9671.2010.01203.x>
28. 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. *The cartographic journal*, 47(4), 315-322. <https://doi.org/10.1179/000870410X12911304958827>
29. Fram, C., Chistopoulou, K., & Ellul, C. (2015). Assessing the quality of OpenStreetMap building data and searching for a proxy variable to estimate OSM building data completeness. *Proceedings of the GIS Research U.K. (GISRUK)*, 195-205.
30. Forghani, M., & Delavar, M. R. (2014). A quality study of the OpenStreetMap dataset for Tehran. *ISPRS International Journal of Geo-Information*, 3(2), 750-763. <https://doi.org/10.3390/ijgi3020750>
31. Fogliaroni, P., D'Antonio, F., & Clementini, E. (2018). Data trustworthiness and user reputation as indicators of VGI quality. *Geo-spatial Information Science*, 21(3), 213-233. <https://doi.org/10.1080/10095020.2018.1496556>
32. Forati, A. M., & Karimipour, F. (2016). A VGI quality assessment method for VGI based on trustworthiness. *GI_Forum 2016*, 4, 3-11.
33. Zhao, Y., Zhou, X., Li, G., & Xing, H. (2016). A spatio-temporal VGI model considering trust-related information. *ISPRS International Journal of Geo-Information*, 5(2), 10. <https://doi.org/10.3390/ijgi5020010>

34. Bishr, M., & Kuhn, W. (2013, September). Trust and reputation models for quality assessment of human sensor observations. In *International conference on spatial information theory* (pp. 53-73). Springer, Cham. https://doi.org/10.1007/978-3-319-01790-7_4
35. Moreri, K. K., Fairbairn, D., & James, P. (2018). Volunteered geographic information quality assessment using trust and reputation modelling in land administration systems in developing countries. *International Journal of Geographical Information Science*, 32(5), 931-959. <https://doi.org/10.1080/13658816.2017.1409353>
36. Keßler, C., & De Groot, R. T. A. (2013). Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. In *Geographic information science at the heart of Europe* (pp. 21-37). Springer, Cham. https://doi.org/10.1007/978-3-319-00615-4_2
37. Forati, A., Soleimani, S., Karimipour, F., & Malek, M. R. (2015). Including Users' Semantics in Evaluating the Credibility of Crowdsourced Landscape Descriptions. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40(3), 35. <https://doi.org/10.5194/isprsarchives-XL-3-W3-35-2015>
38. Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31(1), 139-167. <http://dx.doi.org/10.1080/13658816.2016.1189556>



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).