Community interactions in OSM editing

Dipto Sarkar[1,]* and Jennings Anderson[2]

[1] Department of Geography and Environmental Studies, Carleton University, Ottawa, Canada; diptosarkar@cunet.carleton.ca
[2] YetiGeoLabs, Montana, USA; jennings.anderson@gmail.com

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2021 Conference after peer-review.

OpenStreetMap (OSM) data is produced by a vibrant community of mappers. These geospatial data produsers (a portmanteau of “producer” and “user” commonly used when describing members of peer-production communities) represent a plethora of individuals with different motivations, methods of contribution, and usage [1, 2]. Thus, the 1.6M OSM contributors have been aptly described as a community of communities [3]. In recent years, corporate editing teams have introduced a new dynamic to the discussion of communities in OSM; editing teams hired by corporations, such as, Apple, Facebook, Microsoft, Uber, are capable of contributing thousands of changesets each day, outpacing the average volunteer contributor [4, 5]. Additionally, corporate editors (CEs) tend to focus their editing on particular types of map features. These two attributes of corporate editing can lead to CEs breaking off into a siloed group of their own with little or no interaction with the rest of the editors on the map.

Previous research on the OSM community using network analysis methods showed there was limited collaboration between editors with most objects being edited only a few times [6]. Senior editors in particular perform a majority of the mapping work on their own, but do indeed interact with others through co-editing—where subsequent mappers edit the same objects [7]. Since these studies were performed, the OSM community has grown significantly and the community dynamics have significantly evolved with more individual and organized participation, such as corporate editing.

Here, we use a data driven approach to characterize the interactions between the CEs and the rest of the OSM community. We define interactions through editing patterns, constructing a network of interactions where each node represents an editor, and two nodes are connected by an edge if the two mappers have edited the same map object. If the mapper represented by node A edits an object last edited by the mapper represented by node B, then the edge connecting these nodes is directed from A to B. Thus, the node’s degree is the number of co-editors of the node with the in-degree representing the mappers who edited objects after the user and the out-degree representing the number of mappers editing objects before the user. The network can have multiple disconnected components as not everyone co-edits with everyone else in the network. We utilized the OSM-Interactions tilesets to construct these networks [8]. OSM-Interactions vector tiles contain the editing history of all highway and building objects at zoom level 14. They include minor changes to

Sarkar, D. & Anderson, J. (2021). Community interactions in OSM editing
In: Minghini, M., Ludwing, C., Anderson, J., Mooney, P., Grinberger, A.Y. (Eds.). Proceedings of the Academic Track at the State of the Map 2021 Online Conference, July 09-11 2021, 6-8. Available at https://zenodo.org/communities/sotm-2021
DOI: 10.5281/zenodo.5112211

the geometry of objects in which only nodes are moved, but the parent way is left untouched. In this way, we are capturing the complete history of map objects in OSM, as opposed to just changes to the basic OSM elements (primarily nodes or ways).

In keeping with the objects which are primarily edited by CEs, our analysis focuses only on highway and building objects for construction of the network. The nodes in the network are further annotated with a binary category representing whether they are a CE or not. We classify a mapper as being a CE or not by comparing usernames in the network to the disclosed lists of usernames on a corporation's OSM wiki or Github page.

We focus on 4 locations: Egypt, Jamaica, Thailand, and Singapore. We create networks for each of these locations at 3 timepoints, 2015, 2017, and 2019 to characterize the changes between over time. Thus, we constructed and analyzed 12 networks. The locations were chosen as they all have different groups of CEs active.

Across all networks, the Largest Connected Component (LC) accounted for 94% of all nodes highlighting significant interactions amongst all mappers. The LC is the largest group of connected nodes, meaning there exists a path between any two users via other co-editors. The more prolifit a mapper edits, the more likely they are to be a part of the LC. Within the LC, the rate of growth of CE nodes exceeds the rate of growth of non-CE nodes at a rate of 11:1 between 2015 and 2019. However, both types of editors (CE and non-CE) have a comparable number of in and out degrees in most networks, indicating that they edit other people's work and have their work edited at a similar rate. Figure 1 shows specific characteristics of each network:

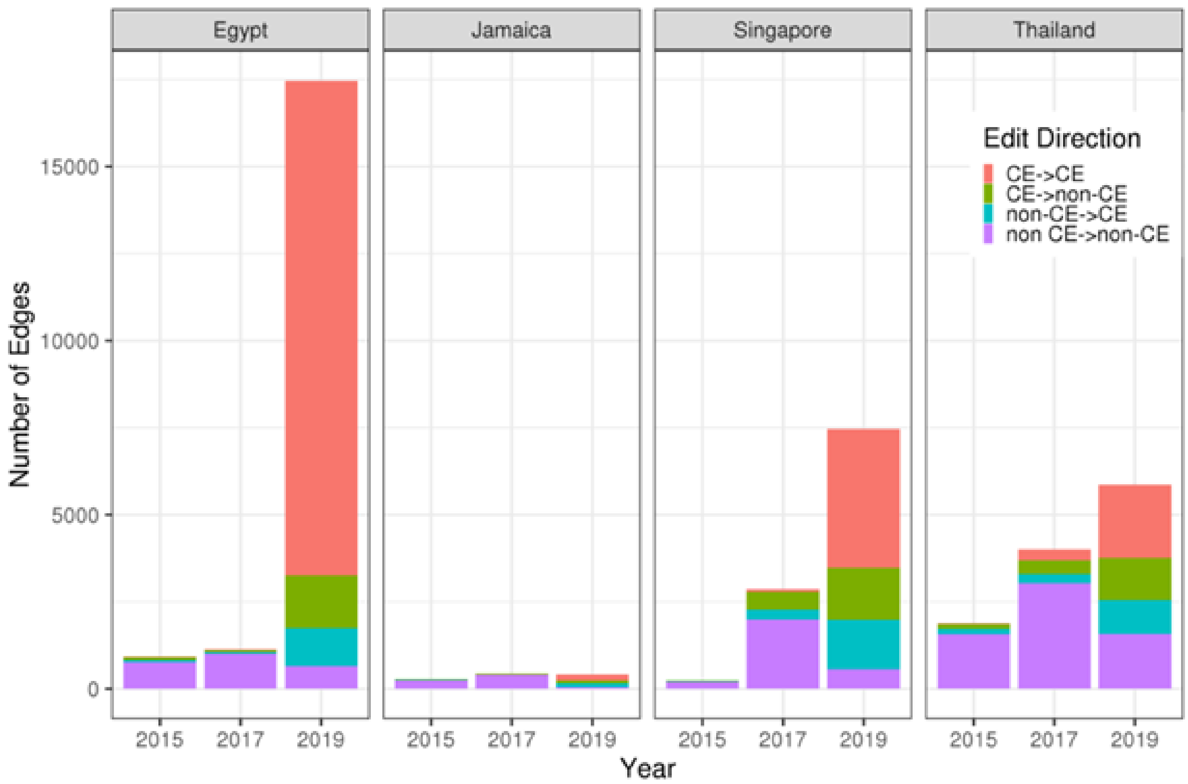

Figure 1. Number of edges in each network each year. Edges are colored by the type of source and target nodes. Orange represents corporate editors co-editing each other, green represents corporate editors editing the work of non-corporate editors, blue represents the work of non-corporate editors editing corporate editors, and purple represents volunteer co-editing activity.

In terms of who edits whose work, CE's edit other CE's work most often, but interaction between CEs and non-CEs have also grown through time, keeping the network connected (Figure 1). The massive growth in the orange edges in Figure 1 in 2019 represents many CEs editing each other. The increase in the green bars shows that corporate editors are also editing (at a much lesser rate) the existing work on the map—however, the blue area shows that volunteers are also editing corporate-sponsored work at a similarly increasing rate.

With regards to age of the mappers (calculated in terms of their enrollment date in OSM) and the volume of edits they perform, younger mappers in both groups tend to edit others' work at a higher rate than senior mappers, but there is more variation in these statistics for non-CE mappers. This is a finding contrary to previous research on editing interaction patterns mentioned above. Additionally, characterizing the time between edits shows that edits made by CE's persist for a slightly shorter duration than edits made by non-CE, primarily due to other CEs editing the same object soon after.

In conclusion, the editing networks highlight the vibrancy of data co-production. The volunteer editor and CEs are interacting with each other's edits to produce the map. The per-group interaction is nuanced and shows unique editing patterns which warrant further investigation. During the timespan of this study, the rate of growth of the CE community was faster than the non-CE community, but whether the pattern will hold over time and whether other locations exhibit the same pattern will require further research.

References

[1] Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for Open Collaboration: Crowd and Community Models and the Case of OpenStreetMap. *American Behavioral Scientist,* 57(5), 548-575.

[2] Coleman, D. D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures Research,* 4(1), 332-358.

[3] Solís, P., Anderson, J., and Rajagopalan, S. (2020). Open Geospatial Tools for Humanitarian Data Creation, Analysis, and Learning through the Global Lens of YouthMappers. *Journal of Geographical Systems*.

[4] Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate editors in the evolving landscape of OpenStreetMap. *ISPRS International Journal of Geo-Information*, 8(5), 232.

[5] Anderson, J., & Sarkar, D. (2020). Curious Cases of Corporations in OpenStreetMap. In: Minghini, M., Coetzee, S., Juhász, L., Yeboah, G., Mooney, P., & Grinberger, A. Y. (Eds.). *Proceedings of the Academic Track at the State of the Map 2020 Online Conference*, 13-14.

[6] Mooney, P., & Corcoran, P. (2012). How social is OpenStreetMap? In: *Proceedings of the 15th AGILE International Conference on Geographic Information Science*, 282-287.

[7] Mooney, P., & Corcoran, P. (2014). Analysis of interaction and co-editing patterns amongst openstreetmap contributors. *Transactions in GIS*, 18(5), 633-659.

[8] Anderson, J (2020). Analyzing OpenStreetMap Contributions at Scale: Introducing osm-interactions Tilesets. In: Abramowicz, W., & Klein, G. (Eds.) *Business Information Systems Workshops, BIS 2020 International Workshops Lecture Notes in Business Information Processing*, 394, 267-271.