A novel application of models of species abundance to better understand OpenStreetMap community structure and interactions

Peter Mooney[1,*]

[1] Department of Computer Science, Maynooth University, Maynooth, Ireland; peter.mooney@mu.ie

* Author to whom correspondence should be addressed.
This abstract was accepted to the Academic Track of the State of the Map 2019 Conference in Heidelberg after peer-review.

Society is built around people coming together into social groups called communities. Communities identify different groups and very often the bond within communities is a set of shared goals and the division/sharing of labour, skills and other resources. The feeling of contributing positively to our own communities is one of the most fundamental feelings of satisfaction in life [1]. All of these traits are visible within the millions of contributors to OSM which form the OSM Community. Attempts to understand how the OSM community works have appeared in the academic literature, such as in [2]. There is a curiosity and fascination about the OSM community given the global extent of OSM: crossing cultures, geographical boundaries and languages; the altruistic nature of its members; and OSM's success as a primarily Internet-based community. In this paper we argue that the model of community required for OSM is more nuanced that many of the current quantitative approaches.

Of many suggested contribution models, OSM has been shown to loosely adhere to the 90-9-1 rule of contribution. In [3] the authors highlight that about 90% of the members of community-based projects only consume the collected information, while 9% occasionally contribute and only 1% demonstrate a very active pattern of contribution. As argued in [4], "characterizing VGI data requires understanding contributors' behaviour and many typologies of contributors are proposed in an attempt to link VGI contributors with the nature of the data they provide". In [5] the authors identify different phases of contributor life cycle from a temporal perspective as a contributor's lifespan is a 'university metric'. In a more computationally complex approach [6], a multigraph approach with data mining is developed to characterise individuals and identify behavioural groups. We consider a very novel approach to community identification and understanding by borrowing concepts and methodologies from theories and models of species abundance to the individual contributors of the OSM community. This is a novel approach for VGI but well established within Ecological Sciences.

Mooney (2019). A novel application of models of species abundance to better understand OpenStreetMap Community structure and interactions
In: Minghini, M., Grinberger, A.Y., Juhász, L., Yeboah, G., Mooney, P. (Eds.). Proceedings of the Academic Track at the State of the Map 2019, 5-6. Heidelberg, Germany, September 21-23, 2019. Available at https://zenodo.org/communities/sotm-2019
DOI: 10.5281/zenodo.3386559

Hughes [7] argues that “in animal and plant communities most of the individuals belong to a small number of abundant species, whereas most of the species are represented by a small number of individuals”. In OSM we see that most individual contributors make a small number of contributions. However, from the global OSM community, a small number of species (groupings) are represented by a small number of contributors (for example very high frequency contributors). We use the OSM Planet History data for a number of selected regions to consider OSM contribution history in those regions. We then develop and apply the Community Level Models (CLMs) from authors in [8] as well as others. We define different types of OSM community member species. For example, we may create species which are differentiated by the number of OSM relations they have created/edited. More sophisticated species can then be developed and integrated into different models. CLMs allow the creation of species co-occurrence matrices to environmental variables (such as quantity of edits, types of tagging, etc.) allowing prediction of the community structure and the distributions of species. CLMS can predict species distributions and changes in the community composition more accurately than other methodologies [8]. Assuming OSM contributors exist in isolation and do not influence one another's editing potentially limits understanding of contribution patterns [9]. This innovative approach potentially offers better understanding of interactions between contributors: for example, between experienced contributors and new entrants or editing interactions during events such as mapping parties. Comparison and analysis of CLMs applied to different countries and regions is also of great value.

References

[1] Proctor, N. (2013). Crowdsourcing—An Introduction: From Public Goods to Public Good. *Curator: The Museum Journal, 56(*1), 105-106.

[2] 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.

[3] Carron-Arthur, B., Cunningham, J. A., & Griffiths, K. M. (2014). Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. *Internet Interventions*, *1*(4), 165-168.

[4] Bégin, D., Devillers, R., & Roche, S. (2017). Contributors’ Withdrawal from Online Collaborative Communities: The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *6*(11), 340.

[5] Bégin, D., Devillers, R., & Roche, S. (2017). Contributors’ enrollment in collaborative online communities: the case of OpenStreetMap. *Geo-spatial Information Science*, *20*(3), 282-295.

[6] Truong, Q. T., De Runz, C., & Touya, G. (2019). Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *International Journal of Geographical Information Science*, *33*(8), 1651-1682.

[7] Hughes, R. G. (1986). Theories and models of species abundance. *The American Naturalist*, *128*(6), 879-899.

[8] Maguire, K. C., Nieto-Lugilde, D., Blois, J. L., Fitzpatrick, M. C., Williams, J. W., Ferrier, S., & Lorenz, D. J. (2016). Controlled comparison of species-and community-level models across novel climates and communities. In *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826), 20152817.

[9] Mooney, P., & Corcoran, P. (2014). Analysis of Interaction and Co-editing Patterns amongst OpenStreetMap Contributors. *Transactions in GIS*, *18*(5), 633-659.