Discovering spatiotemporal patterns on data quality assessment in collaborative mapping: A preliminary study in an area of Brazil

Elias Nasr Naim Elias *, Fabricio Rosa Amorim, Marcio Augusto Reolon Schmidt, Silvana Philippi Camboim

Federal University of Parana - Brazil, elias.naim@ufpr.br,fabricioamorimeac@hotmail.com, marcio.schmidt@ufu.br, silvanacamboim @ ufpr.br

* Corresponding author

Keywords: VGI, OpenStreetMap, Spatiotemporal Patterns, Geospatial Data Quality

Abstract:

Technological and computational advances have enabled geospatial data to be obtained and updated daily in recent decades. This aspect characterizes big data, where data flow in digital media comes from different sources, such as numerical modelling, use of smartphones and internet access (YANG et al., 2017). Given the aspects above, it is essential to consider the emergence of new data sources and methodologies for obtaining geospatial data, different from the classical approach associated with topographic mapping. As discussed by Brovelli, et al., 2019 there are currently different methods for obtaining geospatial data, ranging from topographic and aerial surveys to the use of Volunteered Geographic Information (VGI). The authors emphasize that it is essential to establish methodologies to assess quality, extract and integrate relevant information from different data sources.

Quality is an important aspect to be considered in geospatial data, as it allows determining its suitability for use for specific purposes. In VGI, this issue becomes even more mitigating because the data have heterogeneous aspects, as they can vary according to the location, characteristics of collaborations and collaborators profile. Therefore, quality assessment procedures in the VGI are described from extrinsic approaches, through parameters established by the International Organization for Standardization 19.157 (ISO, 2013) (BROVELLI and ZAMBONI, 2018; ZHANG and MALCZEWSKI, 2017; HAKLAY, 2010) intrinsic (SEHRA, SINGH and RAI, 2017), characterized by the history of editions, the number of contributors and contributions; or even by combining extrinsic and intrinsic parameters (NASIRI et al., 2018).

The assessment of intrinsic parameters is even more relevant in developing countries, where the chronic lack of resources for cartography often results in a lack of up-to-date data to provide a comparison. In addition, this situation makes data coming from VGI even more necessary to complement existing reference mappings (CAMBOIM, BRAVO and SLUTER, 2015). For this, understanding the quality of data in these regions is essential. Therefore, this research aims to go beyond the usual intrinsic parameters, such as the number of contributions and contributors, but to understand how the evolution of these parameters over time can be characterized as an additional quality measure.

Research on this topic has sought to understand how the dynamics of spatiotemporal patterns in the history of contributions can be measured and modelled to extract relevant information. Grinberger et al. (2021) addressed this aspect in work, who measured amounts of contributions over time to detect events on the OpenStreetMap (OSM) platform. Furthermore, Brückner et al. (2021) estimated the completeness of OSM retail stores. Both works were based on the accumulated amount of contributions over time and used the logistic regression model in their data. This model is described from an "S"-shaped trajectory and may be associated with the pattern of contributions in a given area, which starts with few contributions, shows growth and tends to stabilize over time.

As there is a finite number of features in a region, we proposed to study the evolution of the contributions in a homogeneous cell, looking for patterns that vary according to the location of this cell, explicitly working with data in large urban centres in Brazil. The hypothesis is that if we know the existing patterns, we will be able in the future to determine at what stage a given cell is and, therefore, how close it is to have its complete mapping. This information is beneficial for using data that are already more robust and encouraging mapping in regions that are still poorly mapped, working to reduce heterogeneities, especially in the poorest and most peripheral areas.

Given the questions presented, it is questioned whether the answers obtained regarding the spatiotemporal patterns can be related to the quality of the data in the VGI and whether the correlations regarding the regression models can be measured in the same study area and their variations can be attributed to spatial patterns. Furthermore, it starts from the hypothesis that obtaining quality in VGI platforms in developing countries, such as Brazil, can be a guide to finding parameters for their integration concerning topographic mapping. This work aimed to obtain the spatiotemporal patterns

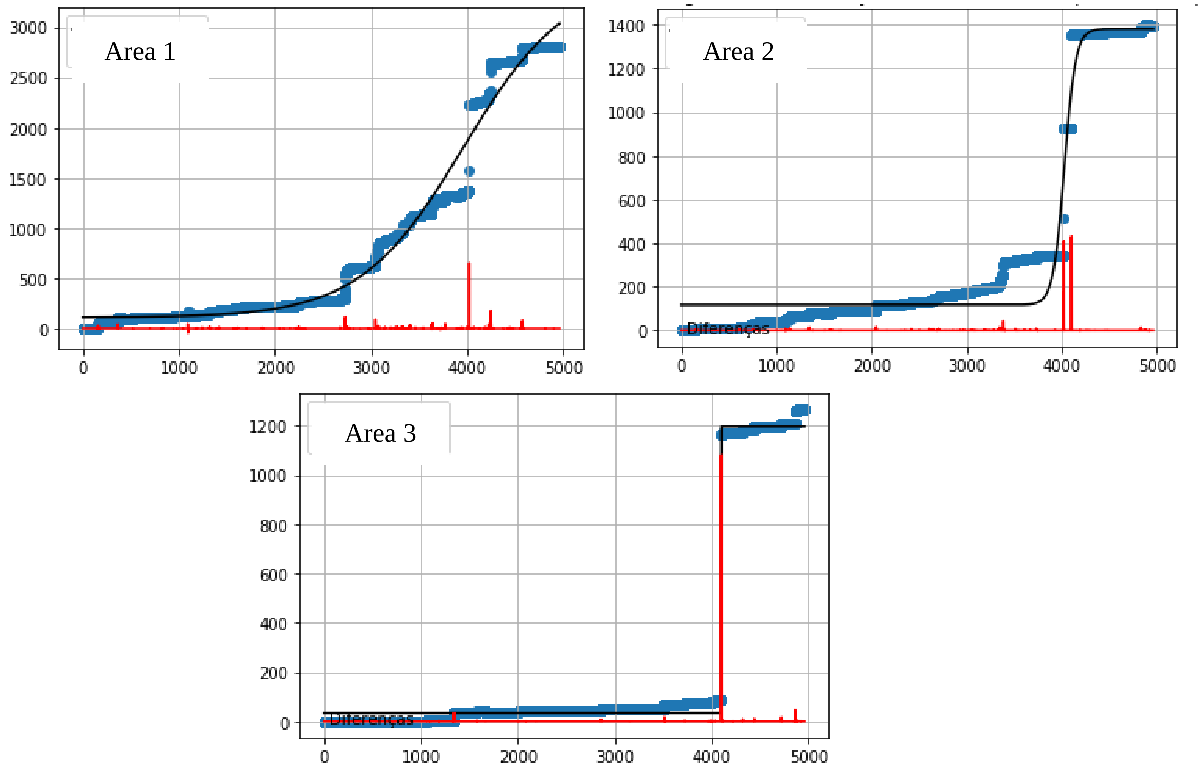
Abstracts of the International Cartographic Association, 3, 2021.
30th International Cartographic Conference (ICC 2021), 14–18 December 2021, Florence, Italy.
https://doi.org/10.5194/ica-abs-3-73-2021 | © Author(s) 2021. CC BY 4.0 License.

of the history of editions of the OSM platform in different parts of the municipality of Curitiba-Parana-Brazil. The adequacy of the logistic model was verified for the accumulated amount of punctual, linear and polygonal features mapped in a given area and the measurement of the smallest possible size of the surrounding rectangle. Furthermore, the feasibility of identifying the saturation of features in an area from the particularities and outliers of the data was identified.

The methodological procedures included obtaining historical data from the OSM from the OHSOME Application Programming Interface (API) (https://heigit.org/bigspatial-data-analytics-en/ohsome/) and the adaptation of scripts in Python language provided by researchers connected to OHSOME to obtain logistic regressions and their parameters. A Python interface was implemented to obtain the results. The input data are the geographic coordinates of the place to be evaluated, the surrounding rectangle's minimum size, and the intended time interval for the analyses. The output data are the tables with the number of accumulated features and the successive differences, the residuals for the logistic regression and the graphs containing the accumulated contributions and the performed regression.

The test was carried out in the central region of Curitiba-PR, between 01-01-2008 and 01-08-2021. It was assumed that since it is a region with a more significant amount of contributions, the interaction of users would be more significant, representing the intended logistical model. Furthermore, from the regression analysis in the region, it was noted that a 1x1 km enveloping rectangle would be sufficient to verify the behaviour of the growth of features in an area. To illustrate the aspects found for the evaluated municipality, Figure 1 presents the regression graphs for three different areas of Curitiba-PR. The first (area 1) is characterized by the central region and the other two (area 2 and area 3) by regions gradually distant from the first region.

Figure 1. Regressions in different regions of Curitiba-PR



As shown in Figure 1, in the blue and black colours are the accumulated growth of features in the OSM and the logistic regression and the number of daily contributions in red. The X axis represents the period, represented in days and in Y the number of features inserted or removed. A critical aspect observed is a more significant iteration of contributions in the central region, exhibiting gradual growth. In the other regions, in addition to less contributions, the interaction was reduced to the point that many contributions described growth in a single period in almost its entirety. Such aspects may be related to imports of data or unusual activities dated in a single period that interfere with the behaviour of the data. This aspect is evidenced in the graph of region 3, in which the contributions were concentrated on a single date.

The approach allowed preliminary estimates to verify the growth of features inserted in the OSM in the evaluated region. The study was applied from the accumulated amount of punctual, linear and polygonal features over time, and the behaviour of logistic regressions showed that the dynamics of contributions influence the quality of the model. In addition, it was noted that it is possible to estimate the saturation of the number of features inserted in the OSM by comparing the levels of completeness in subareas of the same region. This study is the first part of a larger project, including the specialization of residues and responses of regression models and their comparison with responses obtained concerning data quality evaluation procedures to obtain a tool for visualizing the framing of cells according to the spatiotemporal pattern of contributions.

https://doi.org/10.5194/ica-abs-3-73-2021 |

References:

Camboim, S. P., Bravo, J. V. M., Sluter, C. R. (2015). An Investigation into the Completeness of, and the Updates to, OpenStreetMap Data in a Heterogeneous Area in Brazil. ISPRS International Journal of Geo-Information, 4(3), 1366-1388.

Brovelli, M. A., Boccardo, P., Bordogna, G., Pepe, A., Crespi, M., Munafò, M., & Pirotti, F. (2019). Urban Geo Big Data. In 2019 Free and Open Source Software for Geospatial, FOSS4G 2019 (Vol. 42, No. 4, pp. 23-30). International Society for Photogrammetry and Remote Sensing.

Brovelli, M. A., Zamboni, G. (2018). A new method for the assessment of spatial accuracy and completeness of OpenStreetMap building footprints. ISPRS International Journal of Geo-Information, 7(8), 289.

Brückner, J., Schott, M., Zipf, A., Lautenbach, S. (2021). Assessing shop completeness in OpenStreetMap for two federal states in Germany, AGILE GIScience Ser., 2, 20.

Grinberger, A. Y., Schott, M., Raifer, M., Zipf, A. (2021). An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. Transactions in GIS, 25(2), 622-641.

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.

ISO 19.157. (2013). Geographic Information - Data Quality. International Organization for Standarization,

Nasiri, A., Ali Abbaspour, R., Chehreghan, A., Jokar Arsanjani, J. (2018). Improving the quality of citizen contributed geodata through their historical contributions: The case of the road network in OpenStreetMap. ISPRS International Journal of Geo-Information, 7(7), 253.

Sehra, S. S., Singh, J., Rai, H. S. (2017). Assessing OpenStreetMap data using intrinsic quality indicators: an extension to the QGIS processing toolbox. Future Internet, 9(2), 15.

Yang, C., Yu, M., Hu, F., Jiang, Y., Li, Y. (2017). Utilizing cloud computing to address big geospatial data challenges. Computers, environment and urban systems, 61, 120-128.

Zhang, H., Malczewski, J. (2017). Accuracy evaluation of the Canadian OpenStreetMap road networks. International Journal of Geospatial and Environmental Research, 5(2).

https://doi.org/10.5194/ica-abs-3-73-2021 |