

APPLICATION FOR MEASURING REPRESENTATIVE VARIABLES OF COLLECTIVE SPACE INTELLIGENCE

Caio dos Anjos Paiva¹, Raphael Gonçalves de Campos², Silvana Philippi Camboim³

Geodetic Science Graduate Program, Department of Geomatics, Federal University of Parana, Curitiba 81531970, Brazil; E-Mails: anjospaiva@gmail.com¹; raphageoc@gmail.com²; silvanacamboim@gmail.com³

These authors contributed equally to this work.

KEY WORDS: Collective Space Intelligence, Geospatial Data Quality, OpenStreetMap

ABSTRACT:

The scarcity of metrics for analysing the quality of Voluntary Geographic Information without direct comparisons with reference data makes it impossible to use this information in many areas of society. Especially in developing countries, where collaborative data can help fill the deficit of official data, studies on intrinsic parameters of quality become an alternative to conventional comparative methods for evaluating spatial data. A recurring parameter in related research is Collective Spatial Intelligence. Seeking to offer researchers on the subject a tool capable of measuring the Collective Spatial Intelligence in predefined areas, we developed a Python application that counts representative values of this intelligence in political-administrative limits. Considering that, in general, the quality of spatial data is inferred on these limits, research that seeks to explain the VGI quality without using official data as a reference can be facilitated.

1. INTRODUCTION

Whether from conventional mappings or collaborative platforms such as OpenStreetMap (OSM), the rational use of spatial data requires prior knowledge of its quality. Therefore, it is common to apply comparative methods to measure quality. Reference data considered more reliable are used to infer the quality of another product under investigation. In developing countries, where the lack of up-to-date data and restrictions on access to official reference data do not allow comparative methods, this process is complex.

Furthermore, if it is precisely the lack of official data that motivates the search for Voluntary Geographical Information (VGI), the classic approach defined by official metrics for evaluating spatial data quality is not adequate. Because of this, the study of intrinsic parameters proposes an alternative to classical methods (Hecht et al., 2013).

In general, these parameters are based on the so-called Law of Linus. Aimed at software development from the opening of codes to the community of collaborating programmers, the law states that once different individuals examine the codes of software, the presence of errors becomes evident and tends to decrease (Raymond, 2001). In the context of collaborative mapping, several studies link Lei and Linus to VGI quality through the call to Collective Spatial Intelligence (IEC) (Goodchild and Glennon, 2010; Haklay, 2010). According to these studies, the more an area is mapped, and the more collaborators carry out edits in this area, the greater the reliability of the local mapping.

Thinking about the advancement of research that follows this line of reasoning, we developed an application to measure Collective Spatial Intelligence in different areas. We aimed to provide user researchers with the possibility of verifying the variation of the IEC in different political-administrative contexts. To this end, we define two variables to represent Collective Spatial Intelligence. These variables are indirectly present in the history of elements mapped on the OpenStreetMap (OSM) platform and correspond to

the number of employees who worked in a given area and the number of edits carried out in it.

Our application differs from others available on the internet, as the query can be performed simultaneously in a set of areas, regardless of their geometric dimensions. Thus, the classification of areas according to Collective Spatial Intelligence can be performed in a GIS environment. Considering that the scarcity of related studies on VGI quality, especially in developing countries, has not yet confirmed the IEC as a quality parameter, our application is an efficient analysis tool. Since, in general, the inference of spatial data quality is based on delimitations by area, research that seeks to explain the VGI quality using the IEC as a parameter can be facilitated based on its political-administrative measurement.

2. APPLICATION DEVELOPMENT

In OSM, geometries are formed by nodes (points) and ways (lines and polygons). This information is the basis of our application. Each element has a unique identifier (`osm_id`) and a type identifier (`osm_type`) that defines whether the element is a node or a way. Our application was developed within QGIS software, through a script written in Python language. In QGIS, the user must load the layers with the OSM data and the geographic division necessary for their analysis.

The first task performed by the application is reading the OSM layers. For this, the user must load in QGIS, the OSM elements of the region to be investigated. Once this task is done, the script accesses the features' attributes to retrieve the `osm_id` and `osm_type` values of each feature. In this step, the requirement for the script to perform the reading is that vector-type features form the respective layers. When the `osm_id` column exists in the layer, the script interprets that it stores OSM data. In response, a list of `osm_id` is generated. This list is the input for the request process along with the history of features on the OSM platform. With the `osm_id` adequately listed, it is possible to access the OSM API and get its edit history. Each type of feature, node or way, requires a different path to access the history through the API. The `osm_type` defines this path.

With the path defined, a request is made in the OSM feature history, which returns the temporal information of the changes made to the elements. Data such as the name of the user who made the change and the change date are available in .xml format. This allows counting the number of contributors and edits made in a time interval through the query in the .xml file.

A feature can be modified multiple times by a user, so the script removes duplicate names from the user list in the request process. After the query, the processing retrieves the feature history, counting the number of non-repeated collaborators who worked in each geographic unit, as well as the number of edits performed on them. These values are listed in a .txt file, whose osm_id column is the binding key to the read values.

Next, the sum of the values read is associated with the area limits previously entered by the user. This is the subsequent step of the application. For the script to perform this task, the user must first name the layer holding the vector boundaries as "Grid".

The script initially considers that a way may belong to one or more area boundaries to count the number of collaborators and edits per area (Figure 1). This situation is recurrent, especially with elements that make up road systems, such as streets and roads. As the application intends to individually measure the Collective Spatial Intelligence of the user-defined limits, linking the same element to more than one area cannot occur directly since the dynamics of collaborations in the element can result in duplicate values in distinct areas.

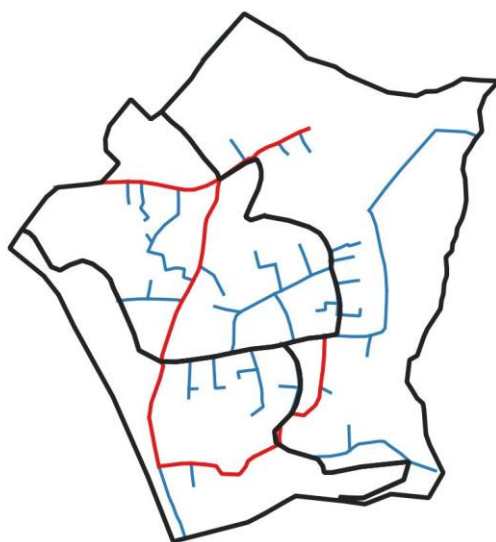


Figure 1. Example of ambiguous relationship

The implemented solution divides the number of contributors in a way by the number of geographic units that intersect with it, thus obtaining a weighted number by intersections. (Figure 2).

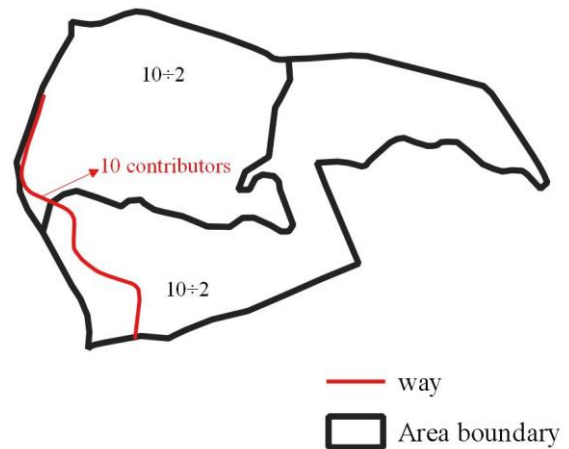


Figure 2. Solution for ambiguity relationship

As the nodes have an unambiguous location, determined through a pair of coordinates (latitude and longitude), the counting of those values in each area boundary is facilitated (Figure 3).

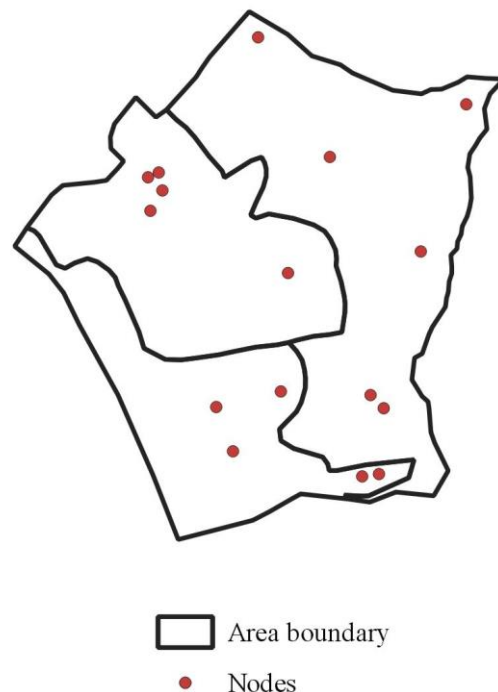


Figure 3. Example of unambiguous relationship

An internal function of the Qgis processing module is used to carry out the count, which has the Algorithm ID: "qgis:joinbylocationsummary". This function joins and summarizes information from a layer to a base layer, conditioned by a spatial relationship. This function has as main arguments:

- Base layer;
- Layer to be joined;
- List of columns to be joined;
- Spatial conditions for the union;
- Data summary functions.

For our case, the base layer corresponds to the area limits defined by the user. The layer to be merged is the OSM data, being the

union's focus, the columns that hold the count of the number of contributors and edits per feature. The spatial condition for summarization is the intersection between layers.

With these parameters, the function will identify the unambiguous spatial relationship between nodes and area boundaries and weight the values to be assigned to the boundaries when the spatial relationship takes place with a way that intersects more than one boundary, according to Figure 4.

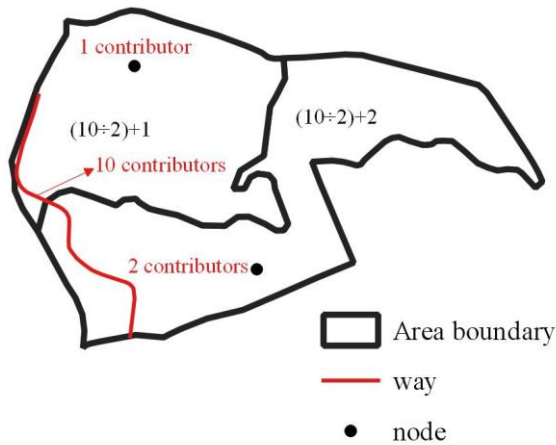


Figure 4. Example of Collective Spatial Intelligence value acquisition

The algorithm's output is a layer stored in memory, so, based on the identifiers of this output layer and the base layer, which are identical for each feature, the information on the number of collaborators is transported to the base layer. In order for the application to evaluate several layers, a column is previously created in the base layer to store the data to be transported, so each processed layer is added to the previous value of the field with the new value.

3. RESULTS

The main result of this work is the tool produced whose code is available at: <https://github.com/Labgeolive-UFP/OSM-Crowd-Intelligence-Tool>. There are the layers used in the development of the application, as well as the script itself. To better understand the process that the user will execute when using the tool, the following flowchart (Figure 5) illustrates the sequence of steps of the application. In blue, the steps of extracting the identifiers belonging to a mother location are shown; in green, the validation steps of the OSM elements, with a subsequent listing of the validated identifiers; in orange, the requisition steps in the OSM history, with the generation of the list of collaborators and editions by element; and in grey, the linkage of the sum of computed values to the area limits defined by the user.

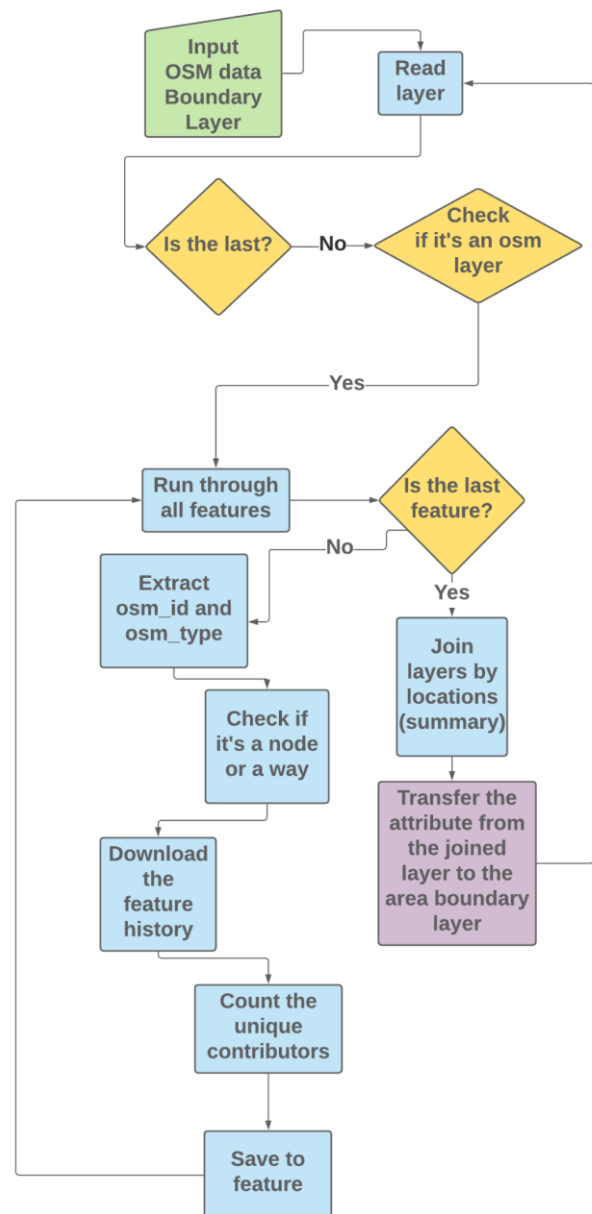


Figure 5. Application flowchart

4. CONCLUSION

Given the studies that relate Collective Spatial Intelligence as a parameter of quality of Voluntary Geographical Information, we sought, with the development of the application presented in this article, a way to measure this intelligence. These applications are particularly important for developing countries, where collaborative data can help solve the chronic lack of up-to-date cartographic information.

As the OpenStreetMap platform is currently the most used collaborative mapping platform globally, its data is the input to our application.

Unlike other similar features, our application differs from the others, as it allows the user to obtain collective intelligence values per area. These areas, in turn, are defined by him, and with

that, the measurement process takes place simultaneously in different areas, whether they are close or not.

Regardless of the required area size, the values read from the collaborative data histories are linked to blocks, which allows the end-user to identify and compare the variation of the IEC in different regions or periods. As the script was developed in QGIS, the representation of this measurement as choropleth maps can be performed within the same platform.

Considering the different scenarios of deficit of official data, observed mainly in developing countries, we consider this application a valuable contribution to studies that propose to infer the quality of collaborative data. Since the rational use of these data depends on prior knowledge of the reliability of voluntary information, tools such as this one are essential for collecting data that provide the researchers involved with inputs for their analyses.

5. REFERENCES

Goodchild, M. F. and Glennon, J. A. (2010) ‘Crowdsourcing geographic information for disaster response: A research frontier’, *International Journal of Digital Earth*, 3(3), pp. 231–241. doi: 10.1080/17538941003759255.

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), pp. 682–703. doi: 10.1068/b35097.

Hecht, R., Kunze, C. and Hahmann, S. (2013) ‘Measuring completeness of building footprints in OpenStreetMap over space and time’, *ISPRS International Journal of Geo-Information*, 2(4), pp. 1066–1091. doi: 10.3390/ijgi2041066.

Raymond, E. S. (2001) *The Cathedral and the Bazaar*. 1st edn. Newton, Massachusetts, USA: O’Reilly Media. doi: 10.1007/s12130-999-1026-0.