Evolution of OSM data after massive imports : comparison between France and Germany

Laboratoire / Laboratory : Laboratoire LASTIG, Université Gustave Eiffel, IGN-ENSG

Equipe / Team : MEIG (Médiation et Enrichissement de l'Information Géographique)

Partenaire International envisagé / International partner : Université de Heidelberg (Allemagne) [Pr. Alexander Zipf, Geographisches Institut Heidelberg]

Tuteur / Tutor : Arnaud Le Guilcher (MEIG, LASTIG, Université Gustave Eiffel, IGN-ENSG), arnaud.le-guilcher@ign.fr

Filière : Datascience et intelligence artificielle, informatique, artificial intelligence and cybersecurity

General presentation of the project :

The OpenStreetMap (OSM) project was built in 2004 as a commun good producing free geographic data (e.g. roads, buildings, points of interests), created and edited by citizens, as opposed to geographic data s produced by public institutions and companies which were not freely available. Editions made by citizens, and the growing size of the OSM community, has enabled OSM platform to have high quality and up-to-date data in many regions of the world. In the last decade, institutional data have been made freely available in several countries, including France, and the OSM database has started mixing these institutional data, which were often added in the database by massive data imports, with data from individual contributions. Following [Hayat 2019], an important question is how individual contributions interact with massive imports to and how good the resulting data are.

Project goals :

The goal of the project is to build a reproducible framework to evaluate the evolution of OSM data coming from massive imports, and in particular to determine how much individual contributors intervene to check and modify these objects, and whether these modifications improve their quality.

A first study on this topic has shown that significant results can be found ([Le Guilcher et al., 2022]), and that they depend on the type of data (buildings, roads,..) under study. To improve on these first results and give them a more general scope, we identify more specific goals:

1) Propose a method to identify massive imports. To identify massive imports, we use the information on the source of data when it is available, but this information is not always available (for example, in [Le Guilcher et al. 2022], this information is available for only 25% of the objects of type 'road'). So a first goal is to conceive a learning method (with a bayesian classifier, or random forests for example) that could detect these massive imports when no source is given.

Information on the sources of data for three datasets in the Occitanie Region
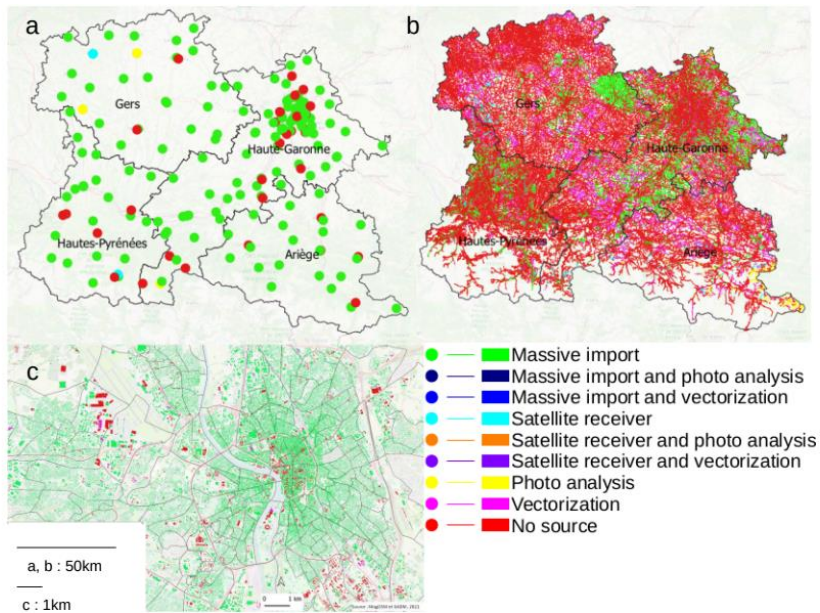
2)Propose a learning method to determine if modifications done by individual contributors actually improve the quality of the data. This method will rely both on intrinsic criteria for quality (for example the granularity ([Barron et al. 2014])) and on a comparison with external sources, using appropriate similarity measures ([Maidaneh Abdi et al. 2020]).

3) Propose a method to group and compare objects on the basis of their evolution patterns, to identify how much the data type and the cultural context influence this evolution

4) Apply this framework on data of different types and on different study areas in France and Germany to evaluate the transferability of the learning techniques, and identify the characteristics of the dataset that determine the evolution of the data.

Bibliography

Main reference

[Le Guilcher et al., 2022] : Le Guilcher, A., Olteanu-Raimond, A. M., & Balde, M. B. (2022). ANALYSIS OF MASSIVE IMPORTS OF OPEN DATA IN OPENSTREETMAP DATABASE: A STUDY CASE FOR FRANCE. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*, 99-106.

Other references

[Barron et al. 2014] Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS*, *18*(6), 877-895.

[Grinberger et al. 2021] : 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.

[Hayat 2019] : Hayat, F. (2019). *Production des biens communs numériques et usages cartographiques* (Doctoral dissertation, Université de Paris).

[Maidaneh Abdi et al. 2020] : Maidaneh Abdi, I., Le Guilcher, A., & Olteanu-Raimond, A. M. (2020). A REGRESSION MODEL OF SPATIAL ACCURACY PREDICTION FOR

OPENSTREETMAP BUILDINGS. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(4).

[Schott et al. 2021] :Schott, M., Grinberger, A. Y., Lautenbach, S., & Zipf, A. (2021). The Impact of Community Happenings in OpenStreetMap—Establishing a Framework for Online Community Member Activity Analyses. *ISPRS International Journal of Geo-Information*, *10*(3), 164.

[Witt et al. 2021] :Witt, R., Loos, L., & Zipf, A. (2021). Analysing the Impact of Large Data Imports in OpenStreetMap. *ISPRS International Journal of Geo-Information*, *10*(8), 528.