4-4-2023

Comparing the data quality of crowdsourced maps - The case of Waze and OpenStreetMap in Miami

Flora Beleznay
Florida International University, fbele002@fiu.edu

Levente Juhasz
Florida International University, ljuhasz@fiu.edu

Recommended Citation

Beleznay, F. & Juhasz, L. (2023). Comparing the data quality of crowdsourced maps - The case of Waze and OpenStreetMap in Miami. FIU Undergraduate Research Conference. April 4, 2023. Florida International University. Miami, FL

Comparing the data quality of crowdsourced maps

The case of Waze and OpenStreetMap in Miami

Flora Beleznay and Levente Juhász
Geographic Information Systems (GIS) Center, Florida International University, Miami, FL

Introduction

Background

Technological advancements, such as smartphones with GPS receivers, cheap devices, and wide internet availability, have advanced the generation of geospatial content by non-professionals (or volunteers) resulting in large open geographic databases. This has been referred to as Volunteered Geographic Information (VGI). Waze and OpenStreetMap (OSM) are two VGI mapping applications that are updated and maintained by open collaboration. VGI has many advantages over traditional map data, such as reference maps generated by mapping agencies. These advantages include timely and cheap map updates. However, VGI generally lacks traditional quality assurance, since anyone with an internet connection can edit these maps and the data is not generated by these authoritative mapping agencies. Therefore, its quality can vary greatly and might not be suitable for some applications that require the most accurate and reliable data.

waze is a community-based traffic and navigation app, that provides real-time traffic updates (i.e. accidents, speed traps, etc.) and travel times to its users. In addition, Waze crowdsources map updates from its community by letting them add, delete and modify road data.

OpenStreetMap is often called the "*Wikipedia of maps*" since it is the largest and most successful examples of VGI that aims to build a freely accessible map database of the whole world.

Volunteered Geographic Information users can edit maps in multiple ways, most notably by adding new or missing streets. The pride of place has been identified as one of the motivation factors behind such contributions, where adding information about one's own group or community may be good for public relations, tourism, economic development, or simply demonstrating that one's own street or establishment is "on the map".

Motivation & Objectives

This project increases our understanding of VGI data quality. Most VGI studies analyze OSM, but this project extends this idea by adding Waze and exploring how similar different VGI datasets are in terms of quality and completeness. Since Waze is a commercial application, its base data is not readily available for download, therefore we must manually extract the data. OSM, on the other hand, is a free open geographic database making data extraction more simple. With this data we can quantitatively and qualitatively measure similarities and differences between the two VGI systems.

The objectives of this project can be summarized as follows:

- Create a tool to access Waze data for analysis
- Design & test a flexible methodology and framework that calculates quality metrics for different data sources
- Run exploratory analysis on derived quality metrics

Methodology

To investigate the positional or locational accuracy and the completeness, we wrote a custom software solution in Python that extracts Waze data for Miami-Dade County, FL. We also acquired OSM data for the same area. The data was imported it into a spatially enabled PostgreSQL database. A regular 45 by 50 area grid with a cell size of 0.01 decimal degree (DD) was superimposed over the study location, then OSM and Waze were compared for each analysis grid independently. Our analysis focused on positional accuracy and completeness. The road data was individually segmented to fit every cell in order for calculations of road length frequency and symmetric difference index could be completed. After segmentation was complete, we utilized RStudio to write a code calculating both the number of road segments and total length of said segments in each cell for both OSM and Waze data. With these calculations, we were able to utilize the symmetrical difference index formula (fig.3), setting any values of 0 in the denominator to null. We then connected these new features to the originally built PostgreSQL database. Finally, we uploaded this data into QGIS by making it into a new layer. This data was merged with the original grid.

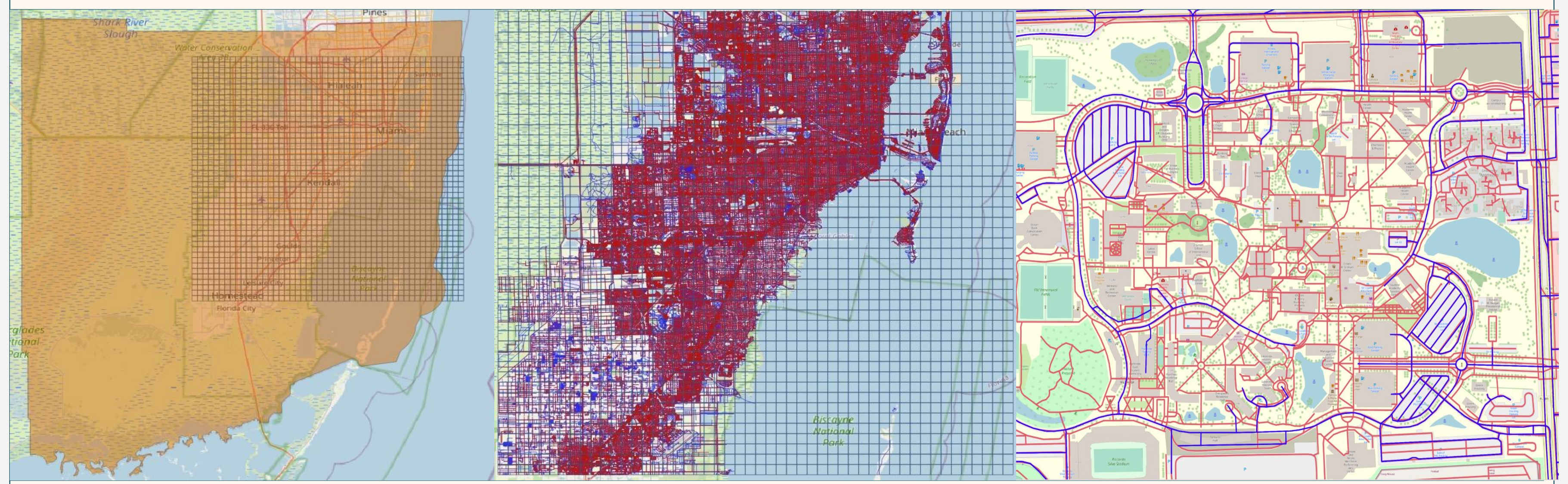

Figure 1. The location of study area within Miami-Dade County (a), a subset of analysis grid cells (b) and visual differences between the Waze (red) and OpenStreetMap (blue) datasets in FIU's MMC Campus (c)

Results

Symmetrical difference index

$$d_{(OSM,Waze)} = \frac{OSM - Waze}{max(OSM, Waze)}$$

Figure 3. Symmetrical difference index formula between OSM and Waze, where *OSM* and *Waze* are the length of OSM and Waze roads within an analysis cell

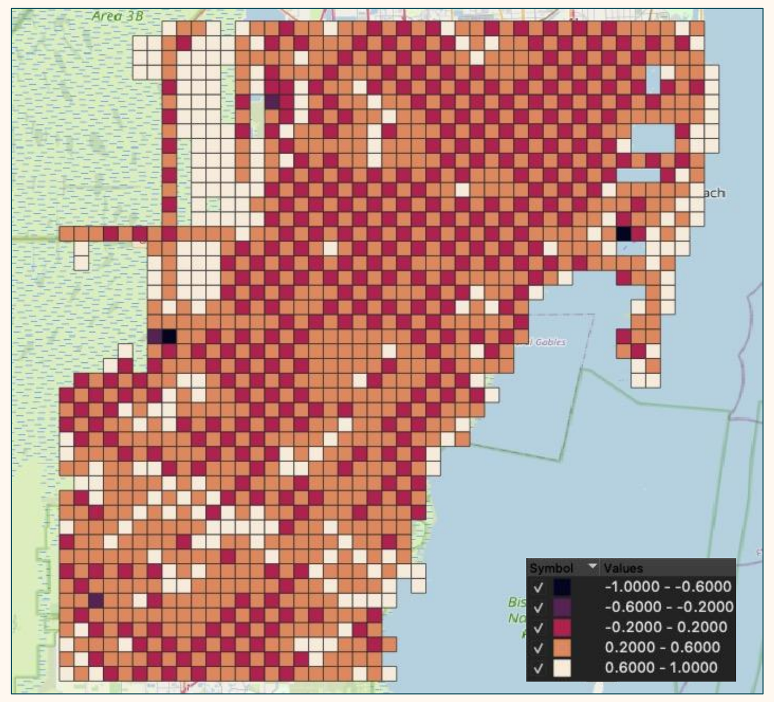

Figure 4. Visual representation of study Symmetrical difference index over study area. White cells depict areas with mostly OSM data present, while black cells show the same for Waze data.

The symmetrical difference index (SDI) between OSM and Waze is determined with the formula shown in fig. 3 where the data used is the total road length measured in m. This index value ranges between -1 and 1, where a value of -1 means that only Waze roads are present in the specified cell and vice versa for OSM. A value of 0 means that the total length of roads in the cell are equal for both Waze and OSM. Fig.4 depicts a cell that encompasses a large section of the FIU MMC campus. For this particular grid, the SDI is roughly 0.63, meaning that there are more roads documented by OSM in the cell than Waze. This is visually noticeable in Fig.4 (blue line segments show OSM data, red segments show Waze data).

Analysis of road frequency

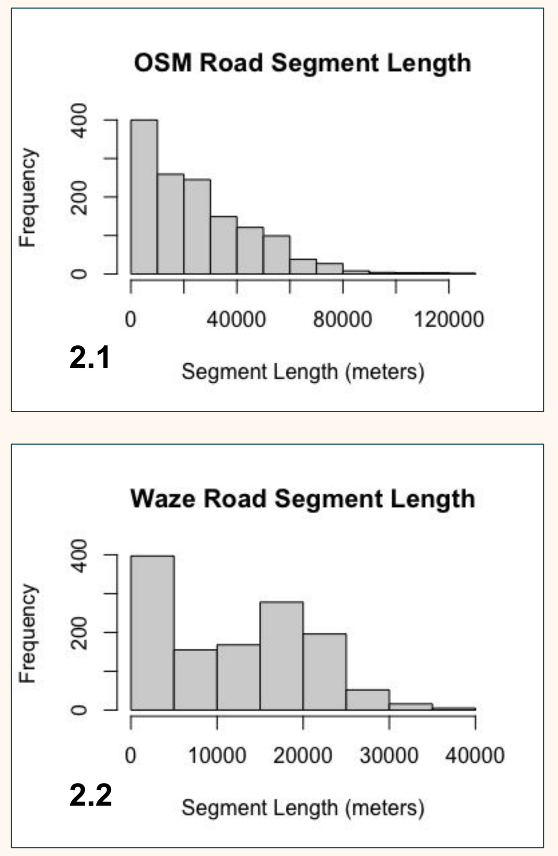

Figure 2. Frequency distribution of road segments. **2.1** Data from OSM sample **2.2** Data from Wazs sample

The frequency distributions depicted in fig. 2.1 and 2.2 compare the frequency of segments to the length of the segments in roads from both OSM (2.1) and Waze (2.2) data. The primary similarities in the two histograms are that the majority of segments are shorter in length, while the longer segments occur less often. The Waze road segment frequency however, has a smaller range than the OSM frequency.

Conclusion & Future work

The results of this study show that there is significant positional and locational accuracy differences between data taken from OSM and Waze. OSM visibly illustrated more detailed paths and side roads, while Waze focused primarily on main roads. This is supported by the road frequency analysis of both data sets. Furthermore, there were more cells containing a value of 1 over -1 in the symmetrical difference index, which means that there were more areas containing only OSM roads. This further suggests that OSM provides a more detailed resolution mapping system in comparison to Waze. For future work, we plan to design a more comprehensive analysis using both intrinsic and extrinsic quality assurance techniques as well as to extend the geographic scope to other study areas.

Contact Information

Flora Beleznay
fbele002@fiu.edu

Levente Juhász
ljuhasz@fiu.edu
GL 275-D