

Deriving Spatio-Temporal Geographies from mobile GPS data: A data-driven approach

Nombuyiselo Murage¹, Alessia Calafiore¹, Franck Lenoir², Irem Ozen²

¹University of Liverpool, ²Tamoco

Project Background

This study leverages on the opportunities presented by individual-level GPS data to study the relations between places and people as they move through the city space. The study develops a GPS-based methodology to understand the Spatio-temporal characteristics of human movement in urban spaces by modelling the collective experiences of city dwellers using network graphs. The result is regions - Spatio-temporal geographies - that have a shared common experience and whose boundaries make up the *living boundaries* described in this study.

Data and Methods

The dataset used in this study consists of **~1.19 billion** mobile phone GPS traces (Unique ID, Latitude, Longitude and Timestamp) that have been generated by around **6.8 million unique users** in New York City for October – December 2019. This study proposed data-binning as an alternative technique of location obfuscation to address geoprivacy concerns. Each GPS datapoint was converted into Uber H3 geometries at a coarser resolution, obfuscating the datapoints individually, thus, preserving the spatial pattern of the user's movement.

The study's methodology consists of three main steps. The first step, data-preprocessing, transformed the dataset from a collection of mobile phone GPS traces into user trajectories (i.e. stops and trips). Two algorithms - Outlier detection and H3 compression - were developed to clean and compress the data before finally extracting stop. Next, the extracted stops were then classified into either home, work or leisure stop points and subsequently filtered so that only leisure stop points remained.

Finally, community detection was accomplished by feeding the filtered stop points into a network graph and running a spatially augmented community detection algorithm on the resulting graph. The research extends previous studies on spatially augmented community detection by introducing custom weights (modelled on user-to-place interactions) to the network graph before running the community detection process. Additionally, we explore the temporal characteristics of human behaviour (the user-to-

place interactions) by filtering the data at various temporal frequencies (i.e. month, intra-month (weekday, weekend) and intra-day (morning, midday, afternoon, evening)).

Key Findings

Data preprocessing was found to be an essential step in accurately determining the input data for our analysis. The dataset reduced to **1.09 billion** after cleaning out erroneous data and **80.38 million** after compression. **20.12 million** stops were extracted based on a 15-minute stay threshold, and filtered leisure stops totalled to **approximately 4.2 million**.

This study demonstrated that the process of community detection is improved by introducing the role of spatial-temporal behaviour of people through the custom weights. A visual inspection of the boundaries shows that the community detection algorithm is sensitive to the addition or subtraction of each user-to-place interaction. Additionally, the research demonstrated that the boundaries of NYC are more stable at a coarser temporal frequency (month) and become more unstable at more refined temporal frequencies (day).

Value of Research

This study demonstrated that data-binning is a viable method for addressing geo-privacy through location obfuscation that still maintains the integrity of the trajectory analysis.

The study developed a data cleaning solution (Outlier Detection Algorithm) to clean GPS data at scale as well as a novel technique of compressing GPS data and extract stops – H3 Trajectory Compression Algorithm (implemented in both Python and SQL). The algorithm (SQL) performed well on a large dataset and achieved a 92.7% counts reduction (from **~1.09 billion data points to a relatively usable 80.38 million data points**).

Finally, this paper presents a methodology of modelling and understanding people's movement collectively and at scale by combining graph theory and spatial analysis, which can be applied to future research work.