Exploratory cluster analysis of urban mobility patterns to identify neighborhood boundaries

Myeong Lee[1], Grant McKenzie[2], and Rajat Aghi[1]

[1]College of Information Studies, University of Maryland, College Park, USA
[2]Department of Geographical Sciences, University of Maryland, College Park, USA
myeong@umd.edu, gmck@umd.edu, raghi@umd.edu

Abstract. Defining neighborhood boundaries within a city is a complex and often subjective task. Neighborhoods boundaries are defined by the people that visit and live in the region, and activities that occur within those boundaries. Depending on the individual or group activity being conducted, these boundaries can change substantially. Transportation and human mobility patterns offer a novel basis on which to explore and delineate neighborhoods. In this work we take a first, exploratory step in capturing dynamically changing neighborhoods based on two different types of urban mobility data. Through clustering temporal urban mobility signatures of alternative transportation users in Washington, D.C., this work provides implications about the characteristics of different types of mobility data and research directions.

Keywords: urban mobility, transportation, temporal signature, cluster analysis, neighborhood boundaries

1 Introduction

The structure of a city is defined by the "memories and meaning" of the people that inhabit it [7]. Within this structure, various regions, often referred to as *neighborhoods*, emerge based on the socio-demographic make-up and activity patterns of the region's inhabitants and visitors. Defining these neighborhoods and their boundaries has been the focus of numerous studies, each approaching the task in a slightly different way. Existing work has focused on defining neighborhood boundaries via geosocial check-ins [3], housing prices [1] or the socio-demographics of the inhabitants [6]. Often, social media and census data were used for these studies. In recent years, access to new datasets produced via mobile application users as well as web-based analytics companies has increased. Alternative transportation (e.g., Uber) and shared economy companies (e.g., AirBnB) have offered access to their data via anonymized data dumps or application programing interfaces (APIs). Even cities and government bureaus are discovering the benefits of opening access to their previously private data stores. The ability to analyze these data allow researchers the opportunity to explore human mobility and transportation behavior and uncover patterns in the data that were not seen previously.

In this work we examine two alternative transportation datasets with the purpose of better understanding the role that human mobility plays in differentiating regions or neighborhoods within a city. Specifically, we compare car share (Car2Go) and bike share (Capital Bikeshare) datasets by examining the spatiotemporal usage patterns (signatures) within Washington, District of Columbia. Through cluster analysis, we group regions of the city that behave similarly in their temporal mobility patterns. Finally, we discuss the similarities and differences in how each mode of transportation delineates regions and hypothesize as to what might influence the construction of spatiotemporal neighborhood regions from these data.

2 Data & Methods

In this work, we used data from *Car2Go*,[1] an one-way car rental service, and *Capital Bikeshare*,[2] a shared bike rental service, both of which have been highlighted as alternative transportation methods of the sharing economy [2]. Car2Go data was collected from their public website every 5 minute for approximately one month from November 8th to December 4th in 2016 (about 4,962K records). Capital Bikeshare data was downloaded from the Capital Bikeshare website for a 2-month period between July 1st to August 31st (about 1,068K records). Notably different from other car rental services (e.g., ZipCar), Car2Go cars can be picked up and parked anywhere within the city. On the other hand, bicycles from Capital Bikeshare are available only at designated stations in Washington D.C (440 stations).

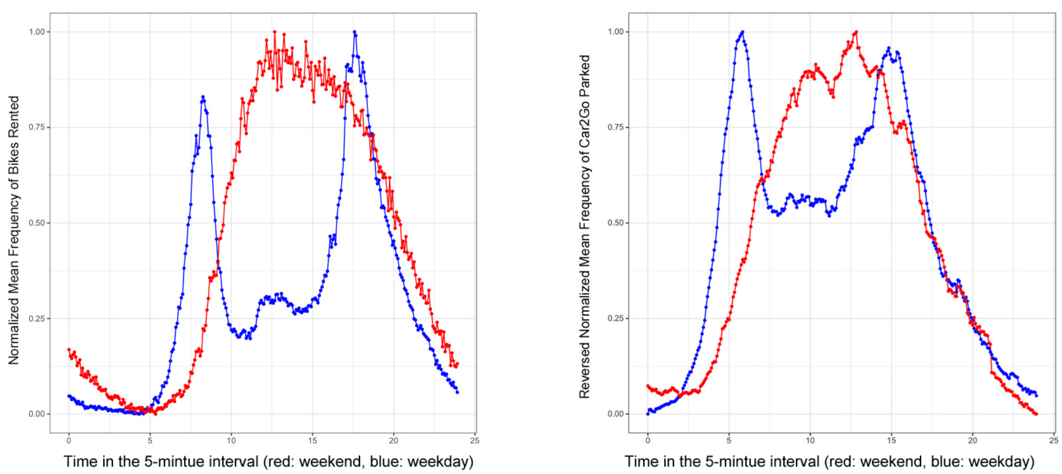

Fig. 1. City-wide temporal signatures for bicycle rentals (left) and car rentals (right).

One primary difference between the two datasets is at the nature of records: while Car2Go records the presence of each car on a street (i.e., the higher the number of records at a given time, the lower the number of cars in use), Bikeshare shows starting and ending times of bicycle rentals (i.e., the higher the number of records at a given time, the higher the number of in-use bikes). Given these characteristics, we created city-wide temporal signatures based on 5-minute intervals throughout a typical day. By aggregating the number of records every five minutes, it is possible to show the temporal patterns of each transportation users in Washington D.C. (Fig. 1). To make it easy to compare the two graphs, the Car2Go signatures were inverted to represents the amount of activity (rather than amount of stagnant cars).

1 http://www.car2go.com
2 https://www.capitalbikeshare.com/

Since people's life patterns on weekdays are often significantly different from weekends, we plotted signatures separately for weekdays and weekends.

In order to characterize each part of the city without using municipal boundaries (construct our own boundaries from the data), we created a spatial layer of 255 hexagons gridding up the city of Washington D.C. where each hexagon covers approximately 0.7 km^2, hinted by a study about identifying land uses in a city [5]. By counting the numbers of cars parked and bikes rented from within each hexagon region, we modeled each polygon as a normalized vector of 576 temporal features (values every 5 minutes over 48 hours) which represents a combination of weekday and weekend signatures. Then, we grouped hexagons into five clusters using K-means clustering based on these temporal vectors. To determine the number of clusters, $K = 5$, the Davies-Bouldin (DB) index was used where a local minimum in the graph of DB indices indicates a strong candidate for K[4]. The result of this clustering analysis provides spatiotemporal activity-based regional boundaries constructed from two different types of urban mobility.

3 Results & Discussion

The clustering results are depicted in Fig. 2. The green hexagons in the Bikeshare map and the blue ones in the Car2Go map imply that the center, or downtown, of Washington D.C. has a particular mobility pattern that is highly likely to be work places during weekdays as people tend to rent a car in the evening (Fig. 3). Also, the southern part of the city shows a group of hexagons that represent an area with a lack of data. While some crowded and vacant regions show similar patterns, it is obvious that the two mobility patterns indicate or may affect different types of neighborhood boundaries. The Car2Go map imply that a residential area in green hexagons is widely spread in the north-east region of the city while Bikeshare data shows the residential area (yellow hexagons) in a smaller region in the middle of the city. It is possible that the reason is due to the demographics of users and aims for using each transportation service.

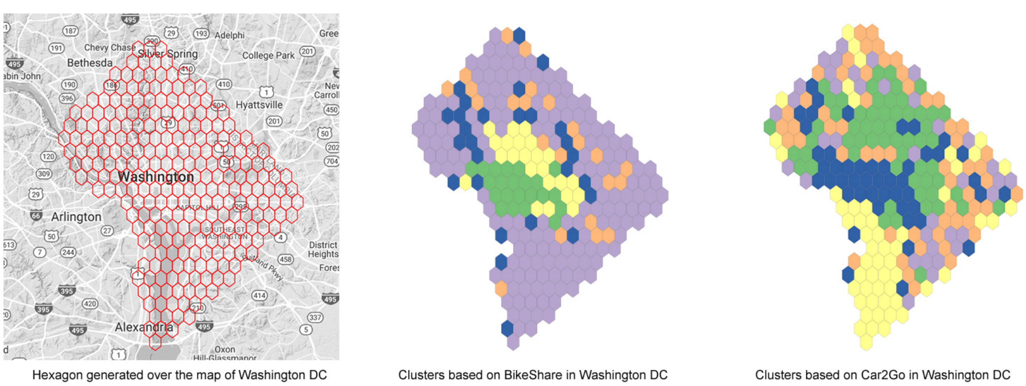

Fig. 2. Hexagons generated over Washington D.C. and clustering results based on temporal signatures of Car2Go and Bikeshare data.

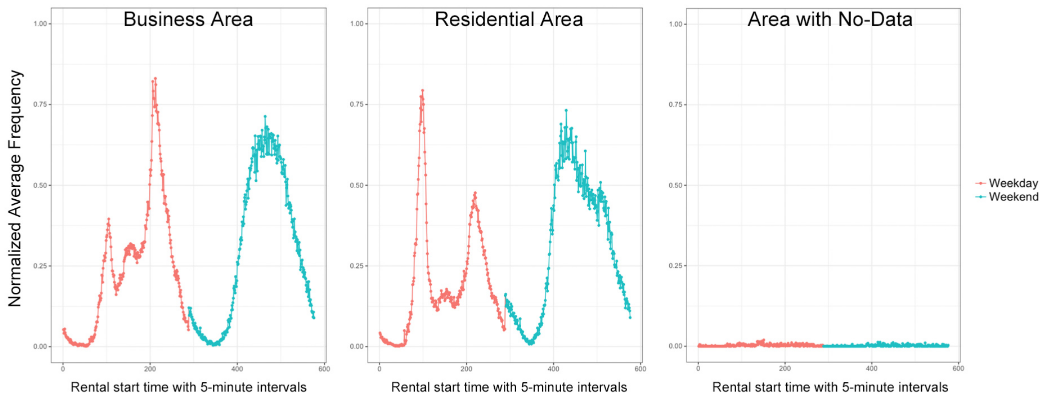

Fig. 3. Examples of cluster-wide temporal signatures for Bikeshare (red: weekday, blue: weekend).

There are biases in these datasets that may constrain the process of identifying neighborhood boundaries. Since Bikeshare data is available only at designated spots, the locations of bike stations likely affects the shaping of boundaries. Similarly, the purple cluster on the Bikeshare-based map does not distinguish a low-activity region from a no-data region. This will be addressed through an opacity value in future work. The number of clusters derived from the DB-index approach may not be feasible for different types of mobility data since clusters with geographically scattered hexagons provide limited understanding about neighborhoods.

Another important aspect of this approach is understanding the temporal and spatial resolution in units of analysis. We used 5 minutes for the unit of temporal signatures, and a 0.7 km^2 hexagon as the unit of spatial analysis. As some graphs show, however, arbitrarily fixed resolution of temporal and spatial resolution can affect the typicality of activity patterns for each region, which may influence the credibility and usability of dynamic neighborhood boundaries identified through the analysis.

These observations raise several hypotheses concerning users and their urban mobility in general, and dynamically changing neighborhood boundaries and characteristics of mobility data in particular: (1) users of Bikeshare and Car2Go are of different socio-demographics, life-style and economic status during their working hours; (2) different types of mobility data indicates different aspects of neighborhoods (e.g., business-oriented neighborhood boundaries on weekdays vs. leisure-oriented neighborhood boundaries on weekends); (3) Bikeshare data would have more variations in people's usage patterns than Car2Go data due to its dependence on weather and other environmental factors; and (4) temporal resolution of signatures and spatial resolution of hexagons (i.e., unit of analysis) influence the shape of neighborhood boundaries.

4 Conclusions & Future Work

Urban mobility patterns offer a unique perspective from which to explore the structure of a city. The way that inhabitants and visitors move about their urban environment often helps to define regions or neighborhoods within the city. In this preliminary work, we explore the spatiotemporal behavior of car rental and bike rental users within Washington D.C.

Our initial findings suggest that extracting temporal patterns within these data can help to delineate regions within a city. Our next steps will focus on adjusting the spatial and temporal resolution of the spatiotemporal signatures to determine their influence on defining regional boundaries. We will also compare these data to existing boundaries from a variety of sources as well as typical car and public transport commuting behavior within a city.

References

[1] Ayse Can. "The measurement of neighborhood dynamics in urban house prices". In: *Economic geography* 66.3 (1990), pp. 254–272.

[2] Boyd Cohen and Jan Kietzmann. "Ride on! Mobility business models for the sharing economy". In: *Organization & Environment* 27.3 (2014), pp. 279–296.

[3] Justin Cranshaw et al. "The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City". In: *Sixth International AAAI Conference on Weblogs and Social Media*. 2012.

[4] David L Davies and Donald W Bouldin. "A cluster separation measure". In: *IEEE transactions on pattern analysis and machine intelligence* 2 (1979), pp. 224–227.

[5] Vanessa Frias-Martinez et al. "Characterizing urban landscapes using geolocated tweets". In: *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE. 2012, pp. 239–248.

[6] John R Logan and Hyoung-jin Shin. "Birds of a feather: social bases of neighborhood formation in Newark, New Jersey, 1880". In: *Demography* 53.4 (2016), pp. 1085–1108.

[7] Kevin Lynch. *The image of the city*. Vol. 11. MIT press, 1960.