What are people doing with bike sharing system**?**
A study of Capital Bikeshare

Yinyue Hu

Final Paper for PUBP833

May.5, 2014

Introduction

Over the past few years, public bike sharing systems have attracted growing attention throughout the world. The aim of implementing the bike sharing systems is to encourage cycle usage to enhance sustainable mobility (Midgley 2009). Originated in Europe, bike sharing systems have been experiencing an explosive growth across the globe. As of April 2013, over 500 cities in 49 countries have launched bike sharing programs with a combined total of over 500,000 bicycles(Larsen 2013).

As a result of joint efforts between District of Columbia and Arlington County, Virginia, jurisdictions, Capital Bikeshare (CaBi) was launched in 2010. City of Alexandria (VA) and Montgomery County (MD) joined the program later in 2012 and 2013, respectively. The systems is currently operating throughout Washington, DC metropolitan area with over 300 stations and 2,600 bicycles(Capital Bikeshare 2014a), which makes it the second largest bike sharing systems in the United States. Like other bike sharing systems in the world, CaBi offers its users fast and easy access to cycle use and allows them to rent and return the bicycles at any station. Users can register for the CaBi services through long- (annual or monthly) or short- (1 or 3 day) term subscriptions, thus CaBi provides the users with great flexibility to use its services when in need.

The prevalence of bike sharing systems in the US and worldwide has brought up the question as to how to strategically deploy the systems to more urban areas and cover greater population, and thus triggers a series of work to explain the usage of existing bike sharing systems and assess the feasibility of the system design. This paper set out with a simple research question – can we identified clusters of ridership within CaBi network? And if so, are the clustering differ by different time of a year, days of a week and by user membership? A review of literature on the subject suggests that no clustering analysis has been conducted. Using an adapted Distance-based Spatial Clustering of Applications with Noise (DBSCAN), this paper hopes to contribute to the literature by uncovering the dynamic community structures of bike sharing systems with a novel analytical approach.

Bike Sharing Studies

Previous studies on CaBi have been focusing on regression models to explain and predict the ridership in relation to weather, availability of bike lanes, location of the stations and other demographic and socioeconomic factors. For instance, Gebhart and Noland (2013) examined bike usage in association with weather conditions from 2010 to 2011. They found that cold temperatures, rain, and high humidity levels are all negatively associated with the use of CaBi. Under conditions such as rain, darkness, cold temperature and high wind speed, the duration of each CaBi is also shorter, indicating that people may take other trip modes or give up longer trips.

Buck and Buehler (2012) studies how the availability of bike lanes affect the use of CaBi. They found that the presence and total length of bike lanes within a half mile buffer area of a CaBi station is significantly associated with the usage. Controlling for population, percentage of households with no access to an automobile and number of ABRA liquor license holders, their results suggested that one additional kilometer (0.62 miles) of bike lane within the half mile buffer is related to 0.855 additional CaBi rental per day at every station.

Daddio (2012) applied a more comprehensive model which includes income level, vehicle ownership, availability of alternative transportation, presence of retail, park an attractions and so forth. One notable finding from the study is that the location of CaBi stations is highly associated with usage. He found that the further away from the system center (weighted ridership center), the less the CaBi usage. Similarly, Rixey (2013) employed a regression model with socioeconomic and demographic factors and found population density, retail job density, median income levels, share of alternative commutes and non-white population to be significant determinants of CaBi use[1].

A review of literature suggests that the studies conducted on CaBi and other bike sharing systems mainly focus on predicting models of ridership. In contrast, the dynamic relationship among the interconnected stations within the bike sharing network is not fully studied. Rixey (2013) considered the networks effects by incorporating the number of stations within buffers with radii of 200, 400, 600, 800, 1200, 1600, 3200, 4000, 4800, 5600, and 6400 meters around each stations in his regression model and found that number of stations within 4800 meters has a strong, positive impacts on the ridership. Studying bike sharing networks in Europe, Froehlich et al. (2009; 2008) and Vogel et al. (2011) explored clustering methods to analyze stations' activity patterns but only limited their analysis to the similarity of stations' hourly pickup and return patterns.

One key factor that previous literature failed to consider is the interactions among the stations and how that could help to identify critical stations that serve as hubs to connect the rest of bike sharing stations. Thus this paper introduces a novel clustering method and hopes to fill in the research gaps on the subject.

Methodology

The Capital Bikeshare (2014b) provides trip history data from its website and welcomes the public to participate in analysis of CaBi ridership. The data is published quarterly and currently data is available from Q4 of 2010 to Q4 of 2013. This study is conducted on data acrss 2012 and 2013.

The trip history data lists information about any trip regarding its start and end time, date, station (name), bike ID, and the user membership (“member” for annual or monthly subscriptions and “casual” for 1 or 3 day membership). In addition to the trip history data, the CaBi website also provides information on the longitude and latitude coordinates of each station, which allows us to match with information from the trip history data and proceed with the spatial analysis. In total, the 2012 and 2013 dataset yield approximately 2,000,000 and 2,500,000 observations, respectively.

This study employs Distance-based Spatial Clustering of Applications with Noise (DBSCAN) to perform the clustering analysis. DBSCAN is a clustering algorithm developed by Ester et al. (1996) which is based on local connectivity and density functions of points. Compared to other clustering methods, DBSCAN has several advantages, as in particular to the purpose of this study: first, there is no need to specify the number of clusters a priori; second, it is able to identify clusters of arbitrary shapes; and third, it has a notion of noise (Figure 1) so that points with low connectivity will not be included into any cluster (Ester et al. 1996). This study uses DBSCAN under a few assumptions. First, the structures of the ridership communities vary under different conditions and the number and size of the clusters will be different. Second, the shape of the cluster may not be regular and it is possible to have some cluster located within other clusters. Third, a disparity in terms of the CaBi station usage is observed. In contrast to some well connected and highly utilized stations, some others are underused,

1 Rixey's study model was applied to data collected from three bike sharing systems- CaBi, Nice Ride in Minneapolis/ St. Paul, Minnesota, and Denver B-Cycle in Denver, Colorado and did not distinguish between the three systems except for including the system dummies.

so that it is not ideal to include the underused stations into any cluster. As a result, this study takes advantage of DBSCAN for the clustering analysis.

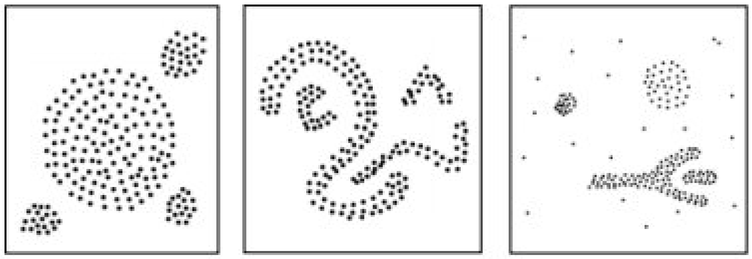

Figure 1: DBSCAN (Ester et al. 1996)

DBSCAN operationalizes the connectivity and density among points based on their distance with each other. And the DBSCAN algorithm requires an adjacent matrix of the distance between any two points. However, this deviates from what we want to measure in the study – the connectivity among CaBi stations based on the number of exchanges between any two. Thus an adjacent matrix of inverse number of exchanges per week between any two stations was created. Using the inverse number of exchange, we were able to fit into the DBSCAN algorithm – the more exchanges, the lower the inverse value, and thus the shorter "distance" between the two stations as in the DBSCAN algorithm.

The clustering analysis using DBSCAN was conducted on 2012 and 2013 data. To examine the temporal clustering patterns, 2013 data was further analyzed by quarter, weekday and weekend, and user membership. The analysis was implemented in R.

Results

Before proceed to the results of clustering analysis, Figure 2 to 4 present some descriptive statistics based on 2013 data. Figure 2 shows that the usage of CaBi stations is unevenly distributed. Whereas some stations experience almost 400 exchanges[2] each week, at about a third of the stations there are less than 10 exchanges per week. The ridership also varies across different time of the year and days of a week. Figure 3 is a heatmap of ridership in 2013 across all CaBi stations by month and day where darker color represents more rides. The graph is consistent with findings in previous literature about weather by showing that in winter months the usage is low. In contrast, the higher usage occurs in spring and summer months and the Saturdays of June experience the highest usage. Also, the variations are further complicated by different membership. As shown in Figure 4, whereas long-term CaBi members mostly use the service during weekdays, short-term casual users are most active over weekends. In sum, these descriptive statistics support our assumption about the variations in CaBi usage.

2 The value of exchange is calculated as the total of pickup and return volume at each station.

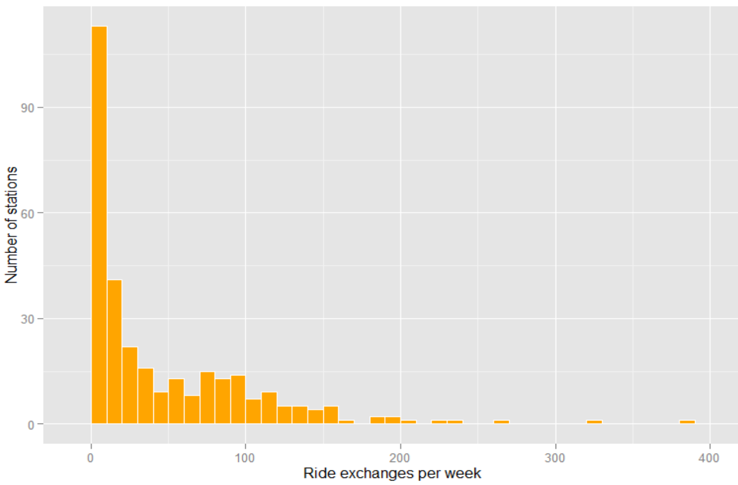

Figure 2: Ridership histogram, 2013

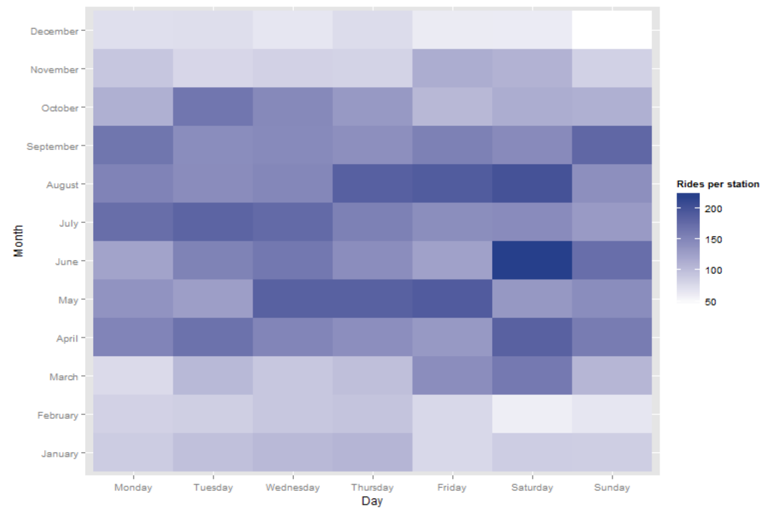

Figure 3: Ridership heatmap, 2013

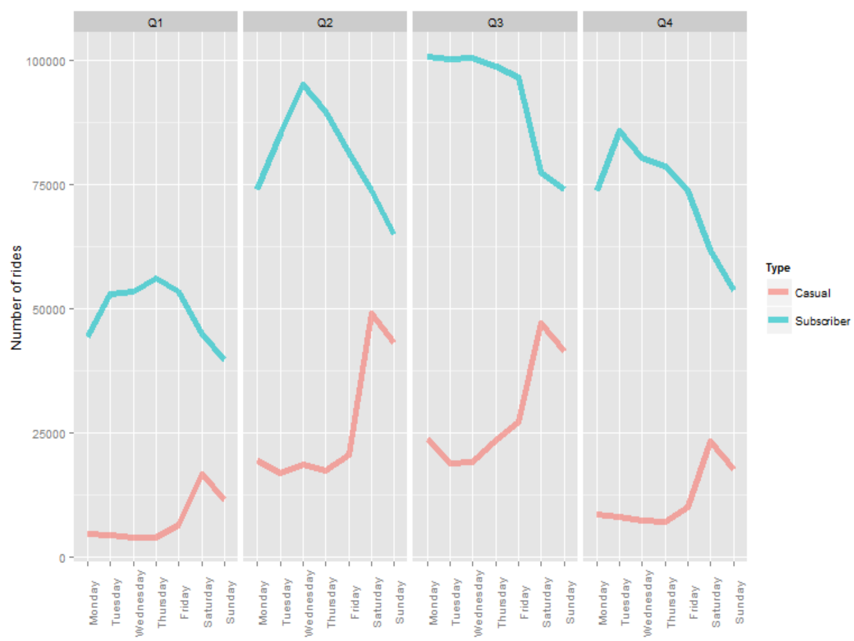

Figure 3: CaBi usage across membership, week and quarter

The clustering analysis performed on 2012 and 2013 data are shown in Figures 5 and 6. Points in the same color indicate that they belong to the same cluster, except for gray points which are noise stations that do not belong to any cluster. The size of the point indicate the amount of usage of the station – the bigger the point, the more the usage. The links between any two stations are plotted to indicate the exchange[3]. In the bottom right is the map of CaBi stations in the city of Alexandria. The clustering analysis of 2012 and 2013 each identified several clusters. The two clustering analyses both identified

3 The link is plotted if weekly average exchange between any two stations is greater than 8. After trying different threshold values, 8 is selected because the plot turns out not too sparse or crowded.

clusters in northwest DC and in Arlington from Pentagon City to Crystal City area. They also suggested clusters along the National Mall and in southeast DC, but the 2013 analysis differs from 2012 in that the southeast DC cluster covers a broader area and reaches as far as Lincoln. Thus we can see a small cluster along the National Mall that is located inside another cluster. Also the 2013 cluster identified another cluster of four stations in Columbia Height area.

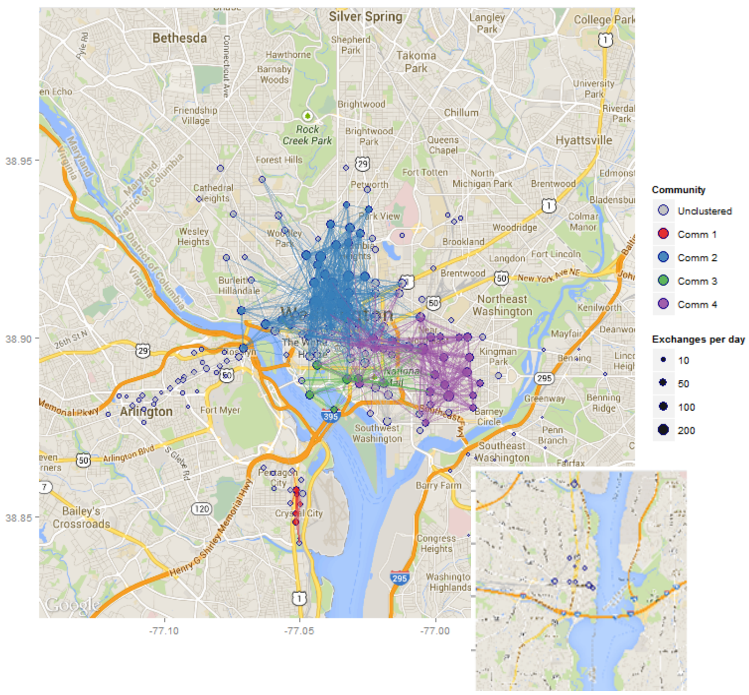

Figure 5: Clustering Analysis, 2012

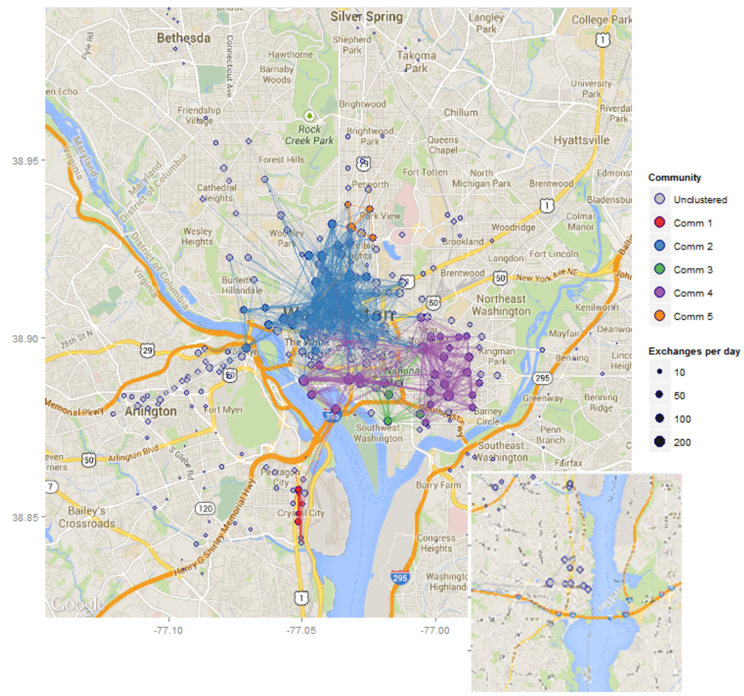

Figure 6: Clustering Analysis, 2013

The clustering analysis of 2012 and 2013 indeed suggests the existence of communities within the CaBi network. It also shows that stations in several areas (north Arlington, Alexandria and Montgomery County) are comparatively under-used and not well connected to other clusters. In addition, although stations in north Arlington do not form a cluster by themselves, the CaBi station at Rosslyn metro station in fact is part of the northwest DC cluster. This is indicative of the stations as a connector linking CaBi usage from north Arlington to DC area. Several critical stations – most exchanges with other stations, are also identified in the clustering analysis, such as Dupont Circle, Logan Circle, Union Station, Eastern Market, Pentagon City Mall and Smithsonian stations.

In addition to the analysis of 2012 and 2013 data, the next step is to look at the clustering difference by seasons, by doing the clustering analysis by quarter. Thus the four quarterly data of 2013 is analyzed and presented in Figures 7 to 10. The quarterly analysis does echo the findings from descriptive statistics about weather in that the first and fourth quarters show similar patterns whereas the second and third quarters yield comparable results. During spring and summer days, the usage is higher (as shown in the density of the links) but less communities are identified in DC – a small community in Cathedral Heights and a large one across DC. In contrast, during fall and winter days, the usage is lower and more communities are identified, which indicates that users took trips with shorter durations within a constrained area. This may due to the cold temperature and also a change in the purpose of using the CaBi services in the cold weathers. Whereas in warmer days users tends to ride for tourism or other leisure purposes, in colder days the use of CaBi is mainly for commutes to work or last-mile transport from public transits to their homes. In addition, CaBi stations at city of Alexandria do form a cluster by themselves in Q2 and Q3, suggesting that more rides have been made among the stations there over this period of time.

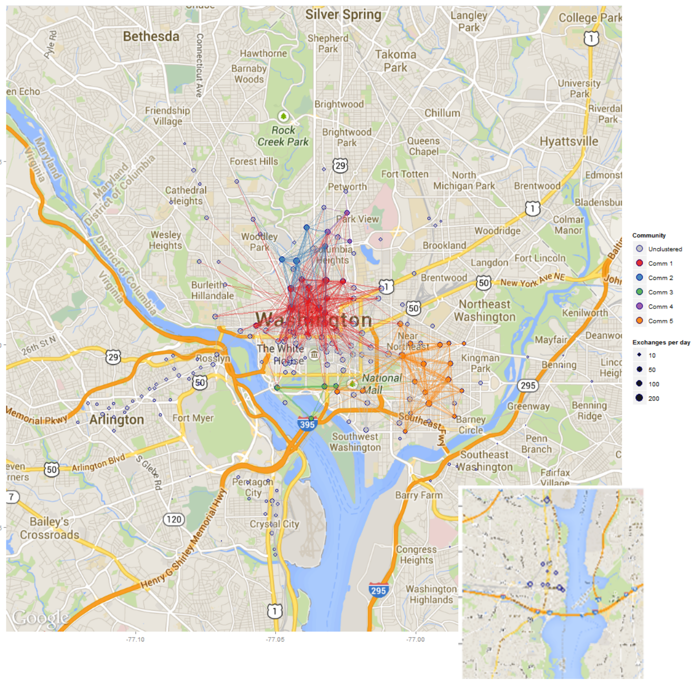

Figure 7: 2013, Q1

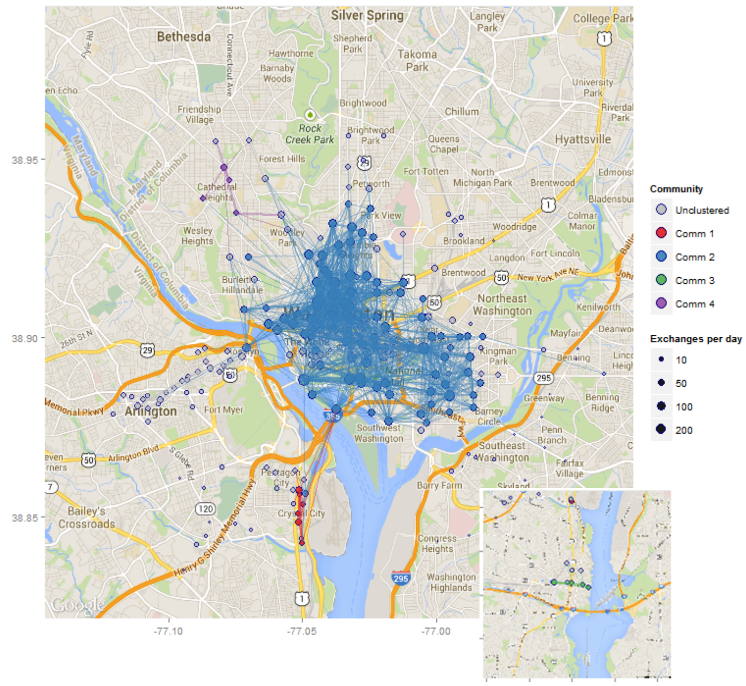

Figure 8: 2013, Q2

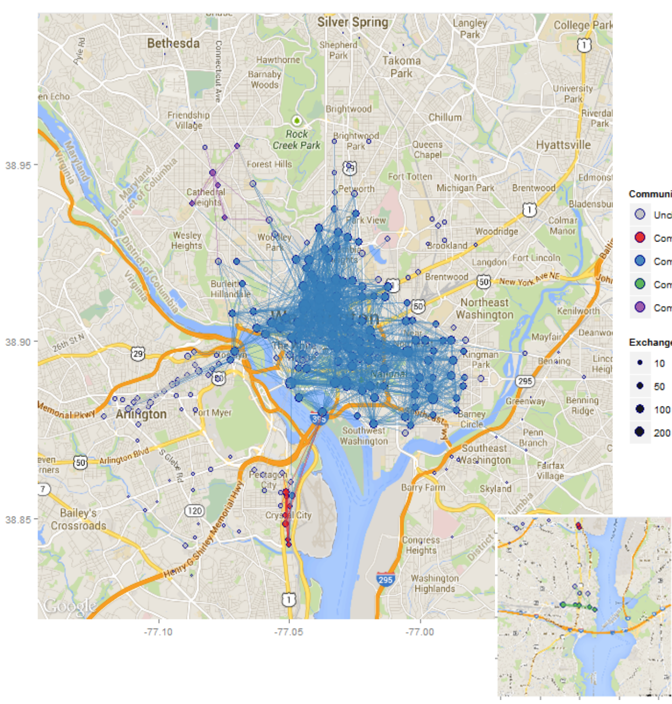

Figure 8: 2013, Q3

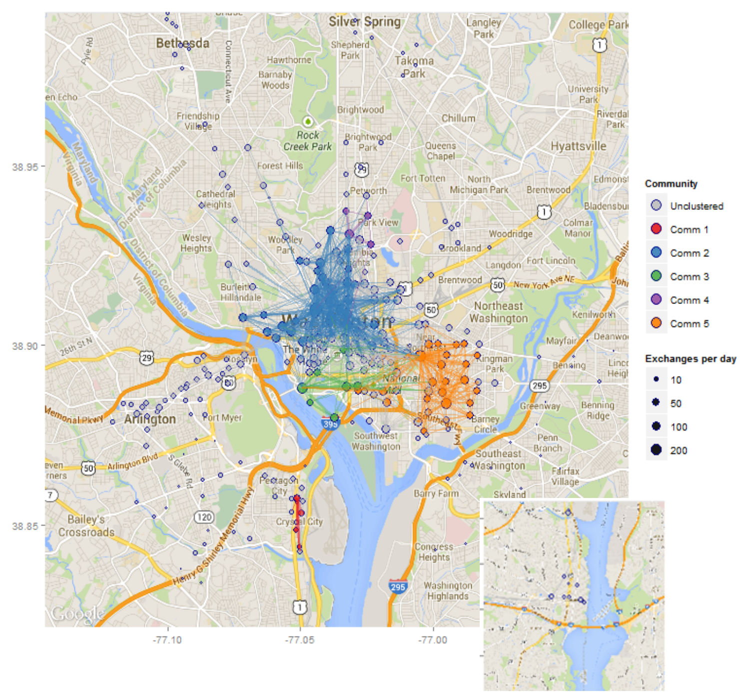

Figure 9: 2013, Q4

The clustering analysis is also conducted by distinguishing rides in weekdays and weekends. The results are presented in Figures 11 and 12. In order to further examine how weekday and weekend patterns vary with season, the clustering analysis is separately conducted on weekdays and weekends in the third and fourth quarters of 2013 (Figures 13 to 16). Overall, the results suggest that CaBi ridership in weekdays and weekends do differ with each other. In some areas a cluster can be identified in weekdays but not in weekends (Figures 11 and 12, Pentagon City-Crystal City cluster). And this difference is further complicated when analyzed at different time of the year. Third quarter weekday clusters are dense and large whereas fourth quarter weekday clusters are sparse and small, indicating that trips made during weekdays in the winter are short in distance and duration. Compared to weekdays, weekends' trips in Q4 are greater in number and longer in distance and duration. This further reinforces our earlier assumptions about the changes in the purpose of riding CaBi.

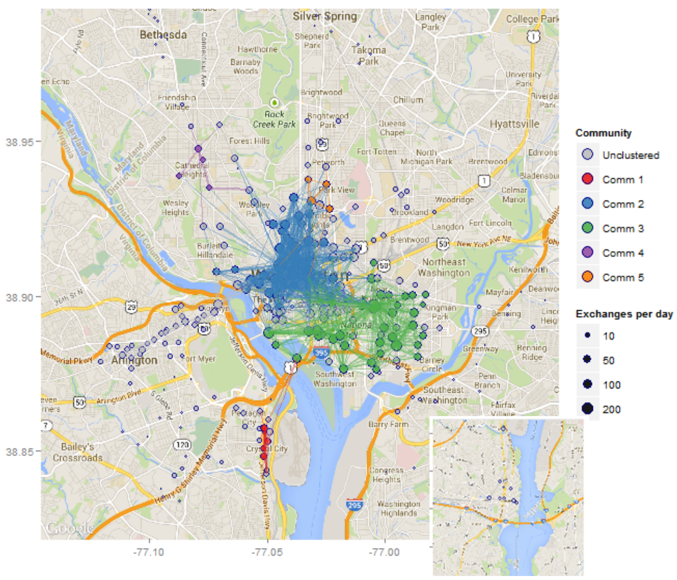

Figure 11: Weekdays, 2013

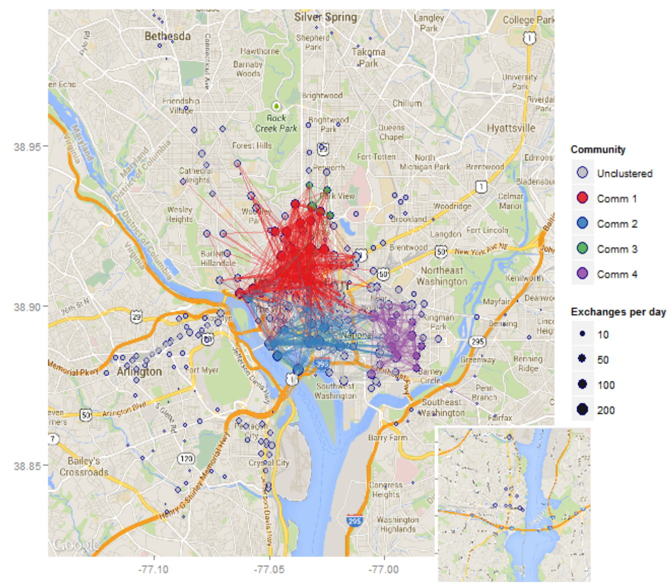

Figure 12: Weekends, 2013

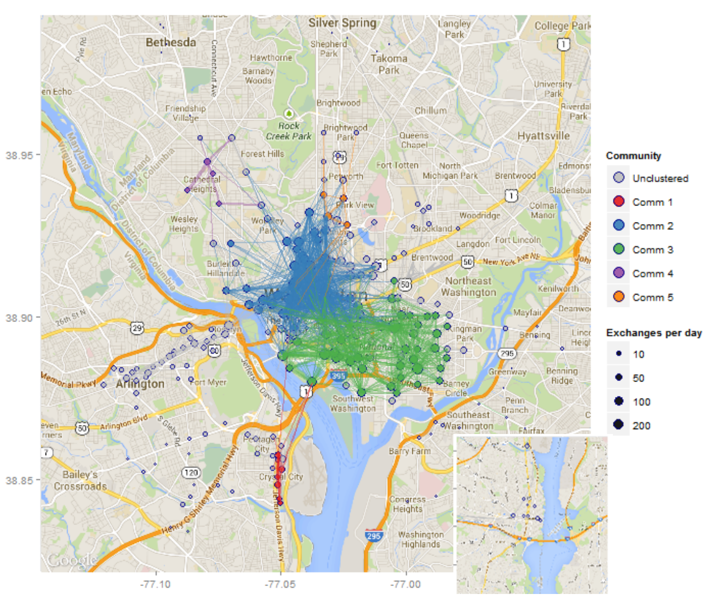

Figure 13: Weekdays, 2013 Q3

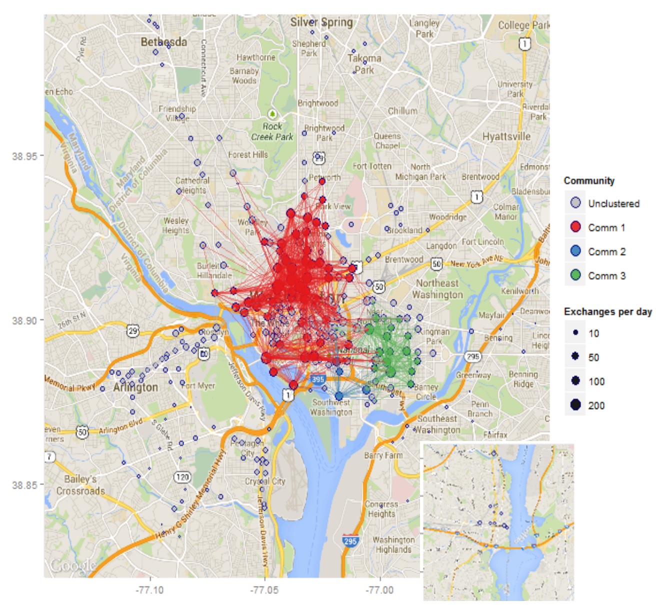

Figure 14: Weekends, 2013 Q3

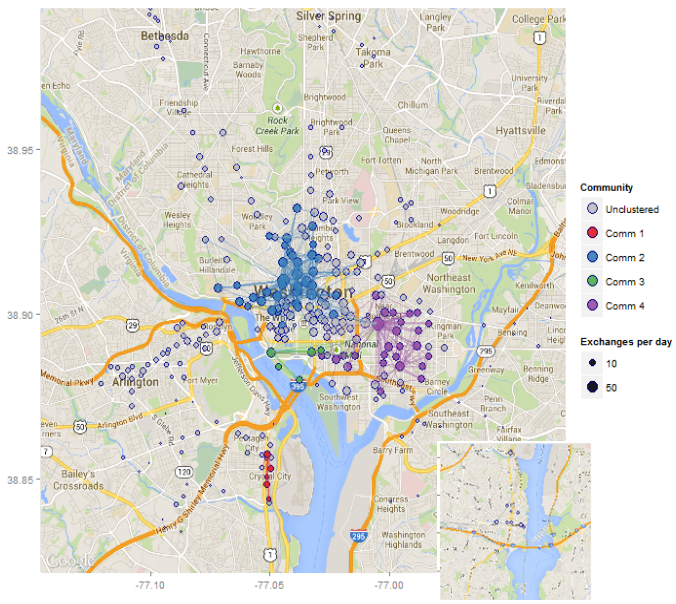

Figure 15: Weekdays, 2013 Q4

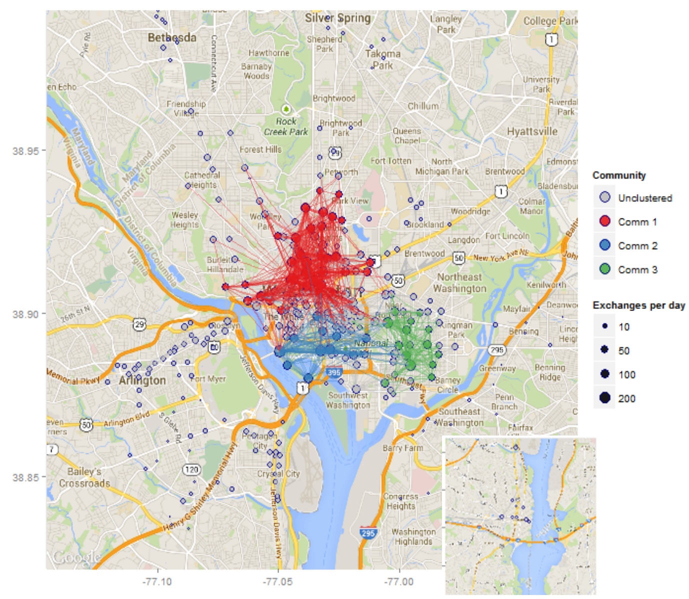

Figure 16: Weekends, 2013 Q4

Finally, the clustering analysis is performed separately on member and casual user data of 2013 (Figures 17 and 18). The results suggest less clusters in the study of casual users than of members. For casual users, the study only yields one cluster in Alexandria and one cluster across DC (and reaches as far as Rosslyn and Pentagon City). This is in contrast with the study of members, which yields three clusters in DC and one across Pentagon City – Crystal City. The DC cluster identified in the study of casual user, compared to order DC cluster such as in Figures 8 and 9, covers a small geographic area and is concentrated around Dupont Circle and along the National Mall. Many stations further away from these two areas (such as southeast and northwest DC) are considered as outliers. Compared to the clusters identified in the study of members, this indicates that casual users tend to ride CaBi within a certain area but the distance of the trip might be longer. In addition, casual users are more active to ride CaBi between Rosslyn and DC stations.

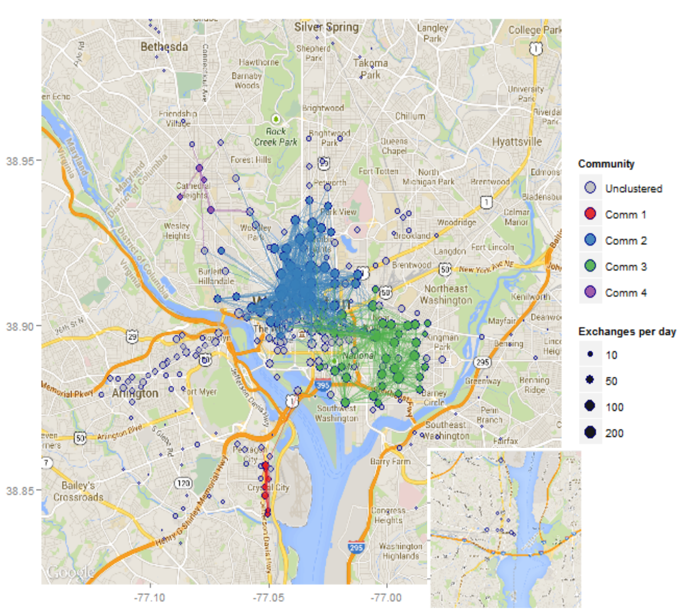

Figure 17: Member

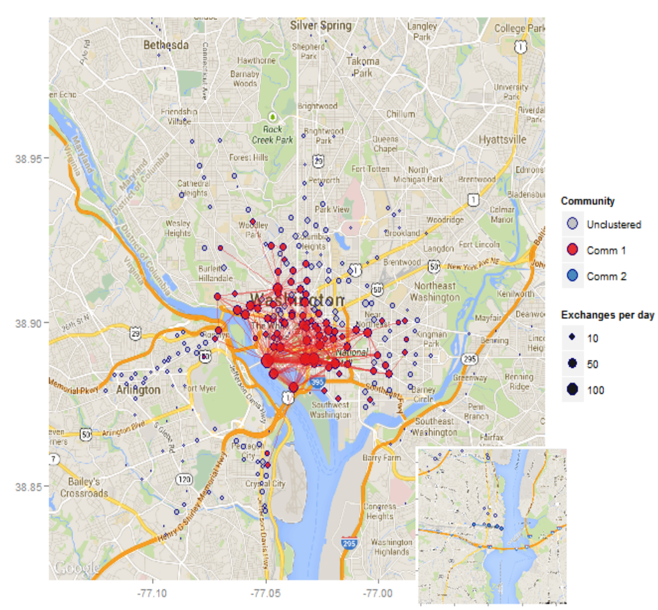

Figure 18: Casual User

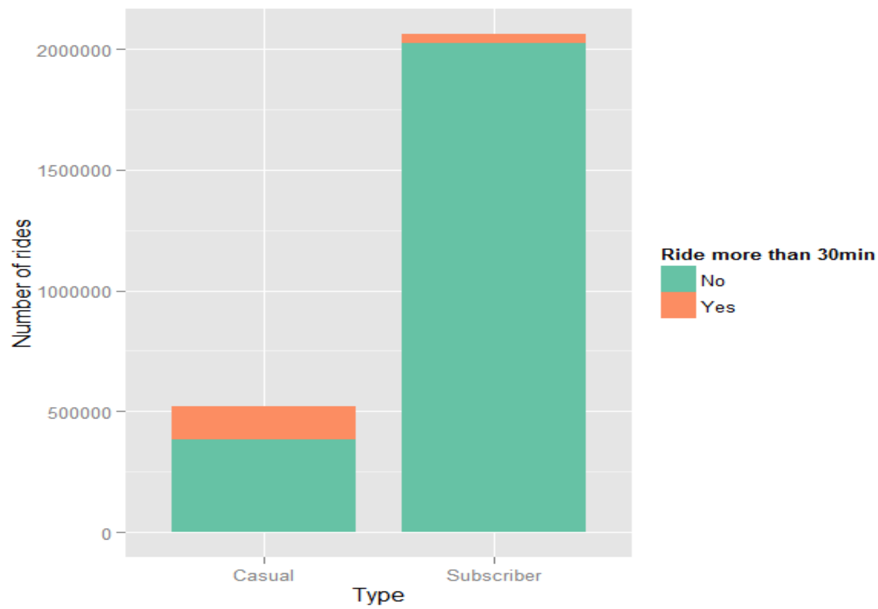

Figure 19: Short vs Long Ridership

These results are also echoed by a comparison between members and users by trip duration. Once a user has purchased a CaBi membership, whether long or short term, the use is free for the first thirty minutes. Beyond thirty minutes, users will have to pay an extra fee depending on the amount of overdue time. Figure 19 shows that casual users are more likely to ride over 30 minutes than members. Thus we may infer that casual users are more willing to pay extra free in order to take rides that are longer in distance and duration.

Conclusion

This paper conducted a clustering analysis of Capital Bikeshare's ridership data, using the algorithm adapted from Distance-based Spatial Clustering of Applications with Noise (DBSCAN). It was able to identify communities within the CaBi network. But the community structures vary across different time of the year, days of the week and membership. In addition, the clusters are more often identified in DC and Pentagon City-Crystal City area. In contrast, Rosslyn-Ballston, Montgonmery County and city of Alexandria stations are less used and connected, thus do not often fall into any cluster. The study also suggested several stations that are highly used and well connected, such as Dupont Circle, Logan Circle, Union Station, Eastern Market, Smithsonian.

The study has several policy implications and suggests future research directions. First, identifying the critical stations can provide information for replenishment strategies to maximize the usage of CaBi. Future studies can also incorporate history data and employ a combination of predicting models and clustering analysis for more dynamic suggestions. Second, the analyses throughout the study all point to the fact that the usage of CaBi vary across location, membership and season. Thus the marketing strategies to promote CaBi should target different populations at stations across the metro area.

References

Buck, Darren, and Ralph Buehler. 2012. “Bike Lanes and Other Determinants of Capital Bikeshare Trips”. 91st Transportation Research Board Annual Meeting.

Capital Bikeshare. 2014a. “Capital Bikeshare Dashboard.” http://cabidashboard.ddot.dc.gov/CaBiDashboard/#Home.

———. 2014b. “Capital Bikeshare.” http://capitalbikeshare.com/trip-history-data.

Daddio, David William. 2012. “Maximizing Bicycle Sharing: An Empirical Analysis of Capital Bikeshare Usage”. University of North Carolina.

Ester, Martin, Hans-peter Kriegel, Jörg S, and Xiaowei Xu. 1996. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.” In , 226–31. AAAI Press.

Froehlich, Jon, Joachim Neumann, and Nuria Oliver. 2009. “Sensing and Predicting the Pulse of the City Through Shared Bicycling.” In *Proceedings of the 21st International Jont Conference on Artifical Intelligence*, 1420–26. IJCAI’09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. http://dl.acm.org/citation.cfm?id=1661445.1661673.

Froehlich, Jon, Joachim Neumann, and N. Olivier. 2008. “Measuring the Pulse of the City through Shared Bicycle Programs.” International Workshop on Urban, Community, and Social Applications of Networked Sensing Systems- UrbanSense08.

Gebhart, Kyle, and Robert B. Noland. 2013. “The Impact of Weather Conditions on Capital Bikeshare Trips”. 92nd 25 Annual Meeting of the Transportation Research Board. http://phillymotu.files.wordpress.com/2013/04/the-impact-of-weather-conditions-on-capital-bikeshare-trips.pdf.

Larsen, Janet. 2013. “Bike-Sharing Programs Hit the Streets in Over 500 Cities Worldwide.” Plan B Updates. Earth Policy Institute. http://www.earth-policy.org/plan_b_updates/2013/update112.

Midgley, Peter. 2009. “The Role of Smart Bike-Sharing Systems in Urban Mobility.” *Journeys* 2: 23–31.

Rixey, R. Alexander. 2013. “Station-Level Forecasting of Bike Sharing Ridership: Station Network Effects in Three U.S. Systems”. 92nd 25 Annual Meeting of the Transportation Research Board.

Vogel, Patrick, Torsten Greiser, and Dirk Christian Mattfeld. 2011. “Understanding Bike-Sharing Systems Using Data Mining: Exploring Activity Patterns.” *Procedia - Social and Behavioral Sciences* 20. The State of the Art in the European Quantitative Oriented Transportation and Logistics Research – 14th Euro Working Group on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air Transport: 514–23. doi:10.1016/j.sbspro.2011.08.058.