Community Structures and Interactions in Vancouver's Bicycle Sharing Network *

Hong Jiang (Ethan) Sim
ethansim@outlook.com

April 10, 2020

1 Introduction

Vancouver is the most populous city and popular migration destination in Western Canada. As the population of Vancouver grows, it increases nuisance factors such as pollution, noise and congestion. The development of new sustainable mobility strategies has become an essential measure for city planning and growth. One measure taken by the City of Vancouver in 2016 was the launch of the bicycle sharing scheme (BSS), as means to sustainable short-distance travel. To ensure successful adoption of BSS in Vancouver, it is vital to understand the properties of urban mobility along with the intended usage and performance. The research work in this paper is dedicated to detect communities within the Vancouver's BSS network through clustering and inform the interactions between those communities. This can be very helpful in a variety of applications such as urban planning and viable business location scouting.

The dataset of this study was obtained from https://data.world and covered all recorded trips for the period May 1, 2018 - August 31, 2018. This period encompasses the entirety of the summer season of the year 2018. The dataset contained the following variables:

- Unique IDs of bicycles
- Names and unique IDs of origin and destination stations
- Start and end timestamps of each trip
- User subscription tier
- Distance traveled on each trip
- Time elapsed since start of rental for each trip

Some exploratory work addressed problems such as missing IDs, variable formatting and outliers such as maintenance stations. Weekends are disregarded, due to varying trip purposes. The final dataset comprised of 292,066 entries, distributed over 191 unique stations (see Figure 1) and was used to compile an Origin-Destination (OD) matrix by aggregating trips on station level. The OD matrix served as input for the derivation of graph object G, with each station representing node N in the network linked to every other station in the network by directed edges weighted by the aggregated number of trips observed.

*Research project report as part of the curriculum of DATA521 on Spring 2020 instructed by Professor Paramjit S. Gill

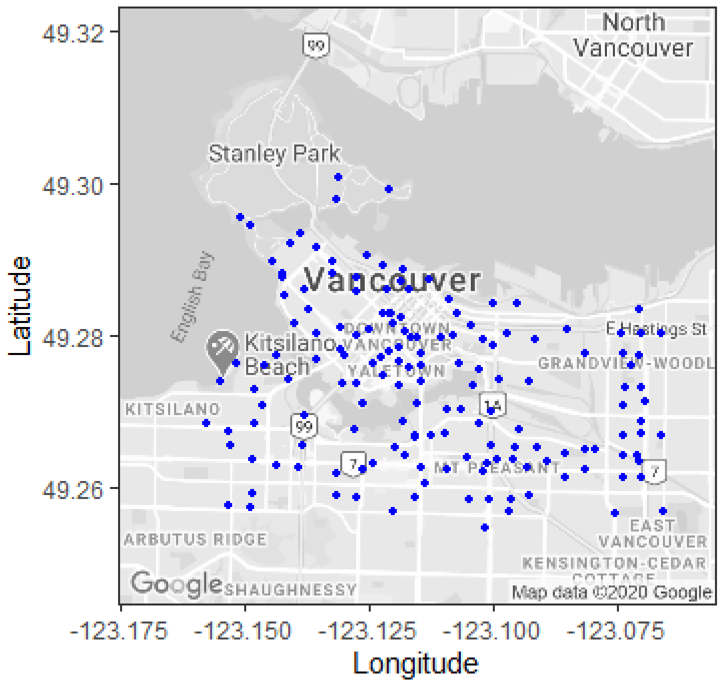

Figure 1: Location of bikesharing stations in Vancouver

Several measures are obtained to highlight properties of the network. Figure 2 depicts the marginal degree distribution of the network, with medians around 115. There is indication of multiple station hubs within the network, with possible correlation between incoming and outgoing station hubs. This network is also highly transitive, in terms of triplet stations, with a clustering coefficient of 0.81. It is moderately dense with an edge density of 0.56 and an edge connectivity measure of 2. Hence, the removal of two key edges between stations would partition the network.

The closeness and betweenness centrality measures help identify several key stations in the network. Based on Table 1, Pender & Gore and St George & Broadway are highly ranked on both measures, signifying them as bridges and shortest paths (on average) to other stations. 4th & Commercial stood out as the closest bike station to skytrain stations VCC-Clark and Commercial Broadway. Science World that is located just outside of the landmark of Vancouver - Telus Science World - serve as a bridge station along the shortest path to other stations.

Table 1: Top 5 stations by centrality measures

Closeness			Betweenness		
Station ID	**Station**	**Value**	**Station ID**	**Station**	**Value**
0213	Pender & Gore	0.0026	0213	Pender & Gore	631
0241	4th & Commercial	0.0026	0193	Science World	455
0166	Alder & 11th	0.0025	0246	Glen & Broadway	434
0248	St George & Broadway	0.0025	0074	10th & Main	429
0245	Woodland & 10th	0.0025	0248	St George & Broadway	415

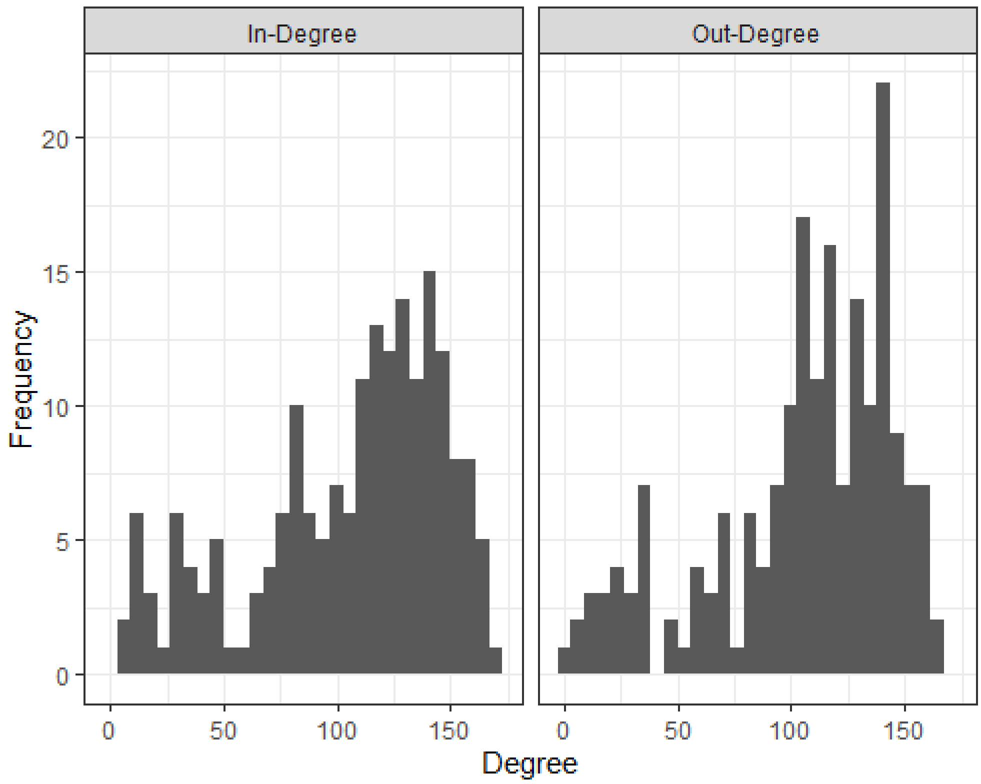

Figure 2: Histogram of in- and out-degrees of the network

2 Methods and Results

Several general statistics are obtained to highlight global trends of the network. Figure 3 reveals a notable trend where the total trips on weekdays are bimodal, with peaks around 9AM and 5PM, suggesting that users utilize this service frequently for commute to and from work. Most notable, the peak frequency around 5PM is higher than the 9AM peak on all days. It is possible that some users cycle recreationally after work hours.

It also helps to have some indicators on the usage behavior of the service. The resulting distributions of distance travelled and duration of trips are conveyed in Figure 4. Both distributions are right-skewed, with median distance of 2.2 km and 12 minutes. It is plausible users are using this service as a means to short distance connectivity.

2.1 Community detection

Given the large size of the network, clustering is used to reduce its complexity and enable better insight into mobility patterns of users. Identifying functional areas within the network would allow for proper adjustments of pricing policies and efficient redistribution of bicycles across the stations. This paper incorporates the Louvain methodology for community detection, which relies upon a heuristic for maximizing modularity in a random network. The heuristic consists of iterative steps of the following until global modularity maximum is achieved:

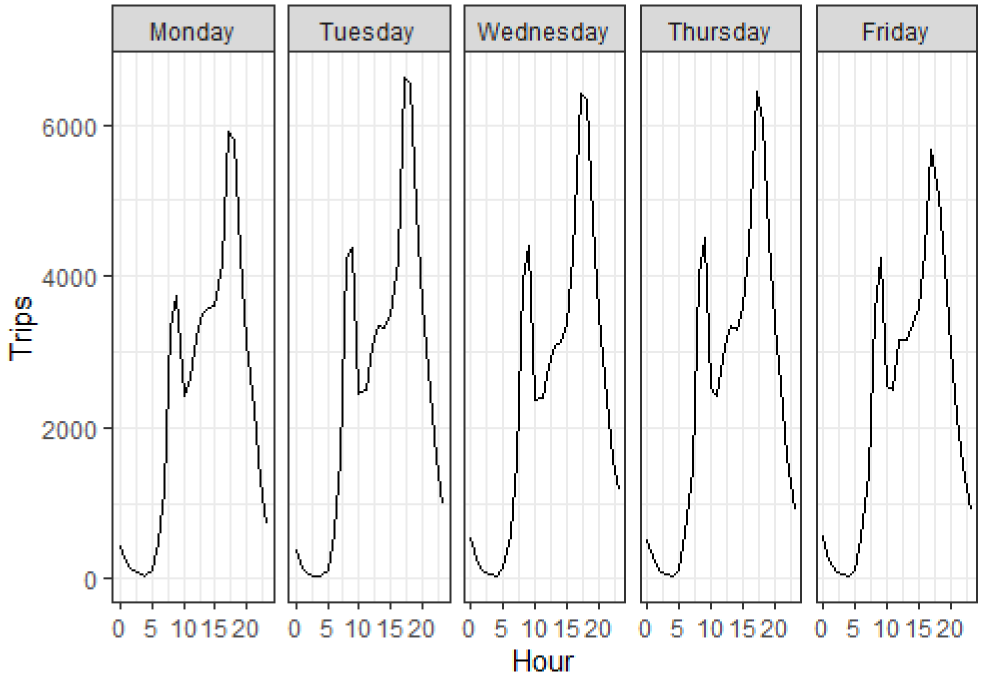

Figure 3: Number of trips per hour and day of the week for the summer season

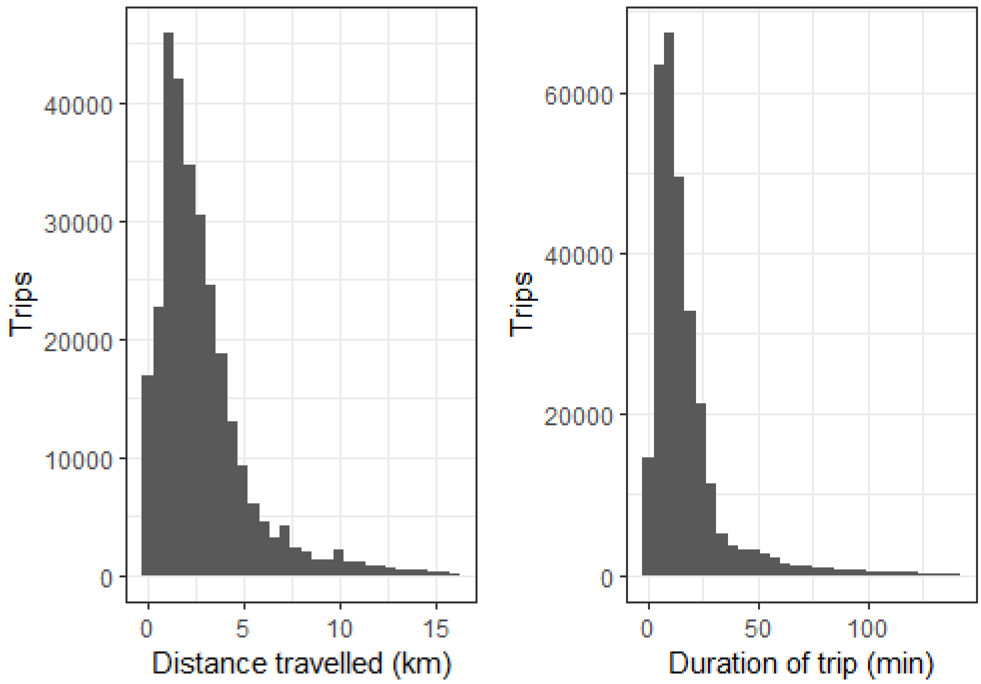

Figure 4: Histogram of trip length in kilometers (left) and trip duration in minutes (right)

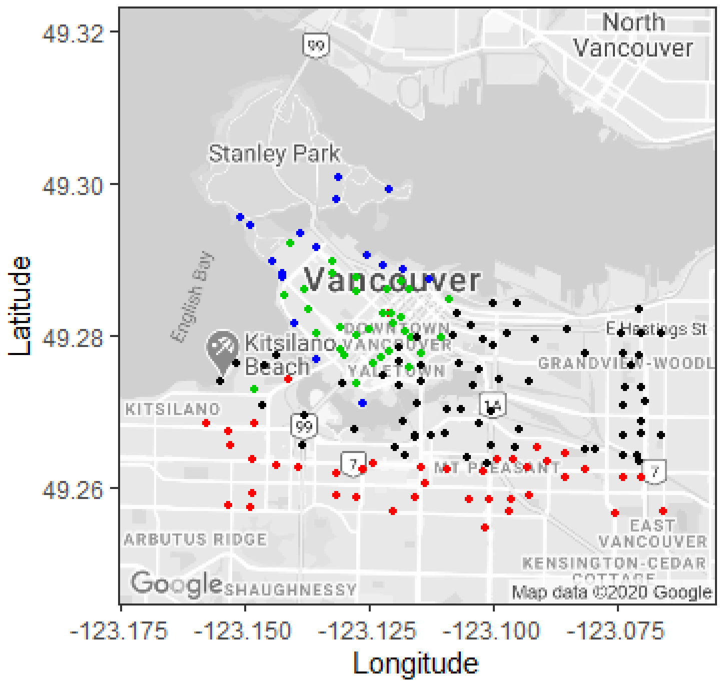

Figure 5: Stations colored according to respective community assignment (Louvain)

- A greedy assignment of nodes to communities, optimizing local modularity.
- Aggregation of nodes belonging to the same community and building of a new network for the respective communities.

The Louvain algorithm is ran on the dataset, returning an assigned cluster membership M for each station node N. A total of 4 communities were obtained and mapped out on Figure 5. The communities correspond to several functional areas in the city based on geography and demographics, indicating spatial dependencies based on usage. Below are some characteristics of the communities:

- Vancouver Peripheral (Blue) - encompasses the peripherals of downtown Vancouver, where beaches, harbors and parks are situated. There are likely more leisure and recreational based activities in this area.

- Vancouver Downtown (Green) - forms at the heart of downtown Vancouver, densely populated by business structures. Business and entertainment activities dominate this area.

- Vancouver Broadway (Red) - stretches through entirety of the major east–west thoroughfare of the city, where restaurants, shopping complexes and residential area meets. Possible activities include shopping and business.

- Vancouver East (Black) - encapsulates Skytrain terminus and connecting stations of VCC-Clark and Commercial-Broadway respectively. Consist of upstate commercial business areas, stadium and restaurants.

2.2 Community interaction

Interactions between identified communities contextualize understanding of spatial usage patterns to better enhance bicycle redistribution. Table 2 summarizes the community interactions between and within each other. Around 60% of all recorded trips start and end within the same community. The remaining 40% would mean operational intervention to ensure bicycle inventory during peak hours. It is helpful to characterize community interactions by destination and origin attractiveness.

Table 2: Summary of community interactions

Community	Stations	Community trips	Trips within	Trips out	Trips in
Vancouver Peripheral	17	65,098	45,588	19,510	25,050
Vancouver Downtown	43	95,456	51,872	43,584	32,977
Vancouver Broadway	51	38,482	17,738	20,744	11,037
Vancouver East	79	93,030	60,708	32,322	47,084

Vancouver Peripheral exhibit high origin attractiveness where 70% of its trip destinations are within its own community. Though this may seem like an indication of short trips inside of the community, observing the station locations on the map reveals that this community does not have a closed form. Hence, there might be more bicycles situated at the end stations (nodes) of the community.

Vancouver Downtown and Broadway behaves rather similar, with an approximately 50-50 split between trips within and out, suggesting that half the bicycle inventory for those communities would end up elsewhere in other communities. In particular, the Vancouver East community has the highest volume of trips in, suggesting high destination attractiveness. As Vancouver East encapsulates key skytrain stations, it is possible bicycle trips towards this community serve as last-mile connectivity. Notably, further research work is required to infer causal factors such as last-mile connectivity and temporal trends.

3 Conclusion

This paper presents the incorporation of the Louvain community detection on the Vancouver's bicycle sharing network. Four distinct communities were detected, reflecting known spatial structures of the city such as parks, beaches, business areas and transit stations. Inference regarding usage behavior and geographical boundaries of the city enable informed decision of bicycle redistribution policies and expansion.

As such, usage profiles do coincide with temporal trends as revealed on Figure 3. It helps to understand inventory demands based on the timing of the day along spatial structures. Such temporal trends could possibly be analysed using models such as the seasonal ARIMA and exponential smoothing. These are areas for future work.

References

Data and code. (2020). `https://github.com/ethaneas/bikesharing`.

Munoz-Mendez, F., Klemmer, K., Han, K., & Jarvis, S. (2018). Community Structures, Interactions and Dynamics in London's Bicycle Sharing Network. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*.