The Impact of Covid-19 on the Chicago Public Bike Share Network Integration

David Xu (dxu7@uchicago.edu)
William Zhu (wzhu4@uchicago.edu)

Introduction

Started in March 2020, the covid-19 pandemic has greatly impacted people's travel patterns. Public transportation like the subway or buses became less desirable due to high risks of covid exposure. Instead, researchers found that people are opting for public bike share service as an alternative for their commute. Hu et al (2021) showed that in Chicago, compared to other forms of public transportations including buses and subways, public bike share usage declined less drastically during the pandemic because it has the lowest risk of exposure. Padmanabhan et al (2021) found that the average bike trip duration increased during the pandemic in New York, Boston, and Chicago.

Previous researchers have found interesting patterns of public bike share data by adopting a network science perspective. Bargar et al (2014) leverages the Louvian modularity community detection algorithm to compare the community patterns between Washington DC's capital bike share network and Boston's Blue bike share network. They found that the bike network cliques of DC are more spread out than those of Boston, mainly due to differences in cycling road infrastructure and city layout.

This project aims to detect the impact of covid-19 pandemic on the public bike share travel patterns of Divvy bike share service in Chicago from a network science perspective. In particular, we developed two competing hypotheses on how divvy bike travel network in June 2020 may differ from that of June 2019:

- Network Disintegration Hypothesis (H1): It’s likely that during the pandemic, people tend to bike to nearby zip codes rather than to downtown or other distant zip codes. Therefore, the divvy bike network disintegrates into smaller clusters.
- Network Integration Hypothesis (H2): It is likely that during the pandemic, people tend to bike to downtown or other distant zip codes as a substitute for other public transportation. Therefore, the divvy bike network integrates and forms larger communities.

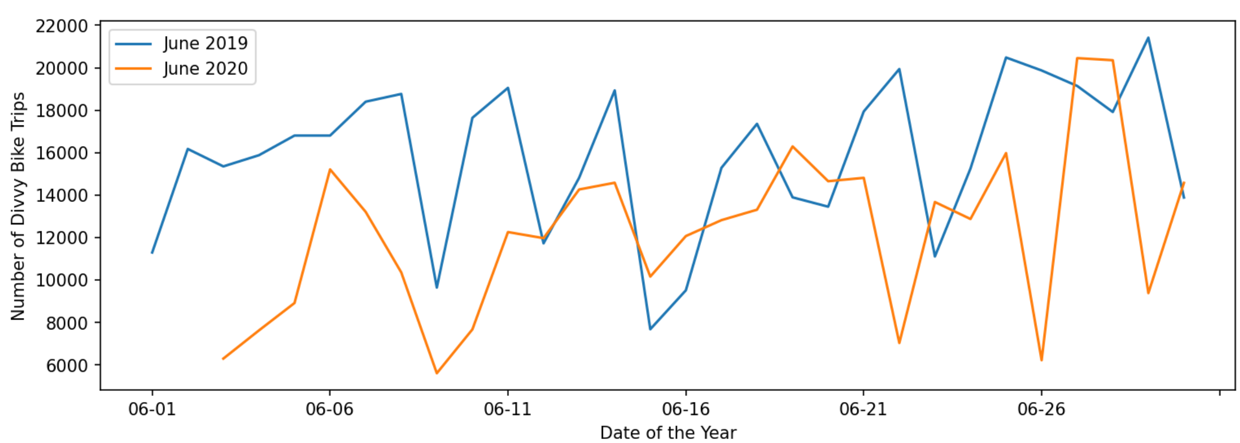
Using network measures including clustering coefficients, degree, modularity measures, and techniques of community detection algorithms, we found evidence to support Network integration hypothesis (H2). During the pandemic in 2020, the public bike share network in Chicago became more integrated and formed larger community clusters than those before the pandemic (in 2019).

Data and Methods

Divvy bikes is a public bike sharing service in Chicago operated by Lyft for the Chicago Department of Transportation. Launched in 2013, Divvy bike share currently has 681 bike stations distributed in 46 zip codes in the Chicago metro area. Riders can use their mobile phones to check out and return bikes at bike stations.

In this project, we compared the Divvy bike trip data of June 2019 and June 2020. The dataset is collected by Lyft and publicly accessible from here: https://www.divvybikes.com/system-data. Trips taken by service staff or lasted below 60 seconds were removed by the data provider. Both datasets only include normal bikes that must be checked-out and returned at bike stations installed across the Chicago metro area[1]. The two monthly datasets are stored in csv format. Each row represents a bike trip, with features including the starting and ending coordinates as well as timestamps. Figure 1 shows the number of divvy bike trips by day in June 2019 versus June 2020. We can see that on most days, the total number of trips decline in June 2020 compared to that of June 2019[2]. This is expected given the effect of the pandemic of human mobility.

Figure 1: Comparing the Number of Divvy Bike Trips (June 2019 vs June 2020)



[1] Divvy bike share service introduced the E-bikes in July 2021. Before then, only normal bikes were available for check-out.
[2] For some unknown reason, the June 2020 dataset provided by Lyft excluded trip data on June 1st and 2nd. It does not affect results presented in the rest of this paper

We converted the bike share data of the entire Chicago metro area into an undirected weighted network. Our network consists of 46 nodes, with each node represents a zip code in the Chicago metro area[3]. We aggregated the bike trips from the biking station into zip code, and only kept biking trips that travelled across zip codes. Based on the starting and ending coordinates of the bike stations, we identified the starting and ending zip code of each trip using the *uszipcode* Python package. In 2019, our network has 638 edges, and in 2020 the network has 690 edges.

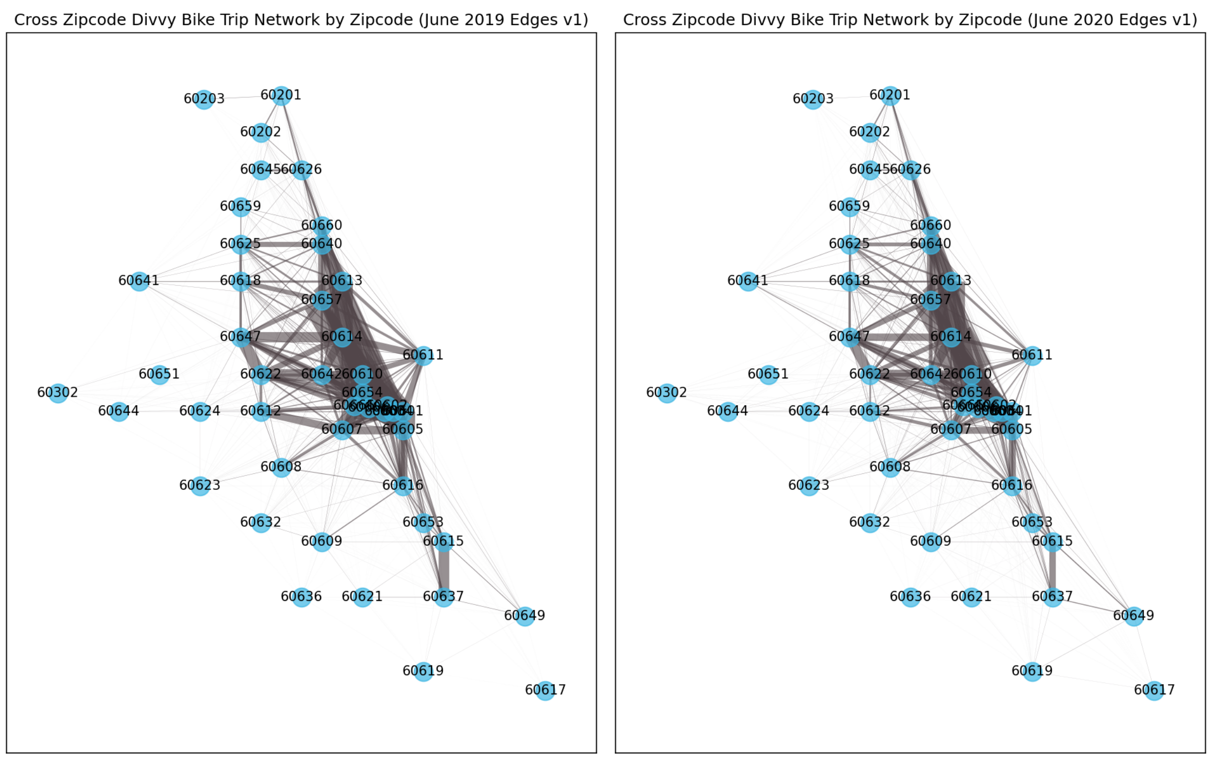
We defined the edge weights of the network in two ways: In the first way, we measured edge weights between two zip code nodes based on the number of cross zip code bike trips in the month. Figure 2 shows our network in 2019 and 2020 using the first edge measure. In both graphs, cross zip code bike trip volumes were higher in downtown and north of Chicago than the west and south. We can also see that the average edge width in the June 2020 network was smaller than that of June 2019. It means that the total number of cross zip code bike trips declined in June 2020 relative to June 2019, likely due to the effect of covid19.

While this method of measuring edge weights enjoys simplicity and clarity in interpretation, it has two limitations: (a) bike trips between two small zip codes (in terms of geographical size) with short average distance are way more frequent than trips between two large zip codes with long average distance. Therefore, the first way of

[3] We considered treating every bike station as a node. It poses two major challenges: (1) there are over 600 divvy bike stations. Visualizing all of them in a network and interpreting them will be difficult. (2) From June 2019 to June 2020, more divvy bike stations were installed. So the comparison will not be consistent.

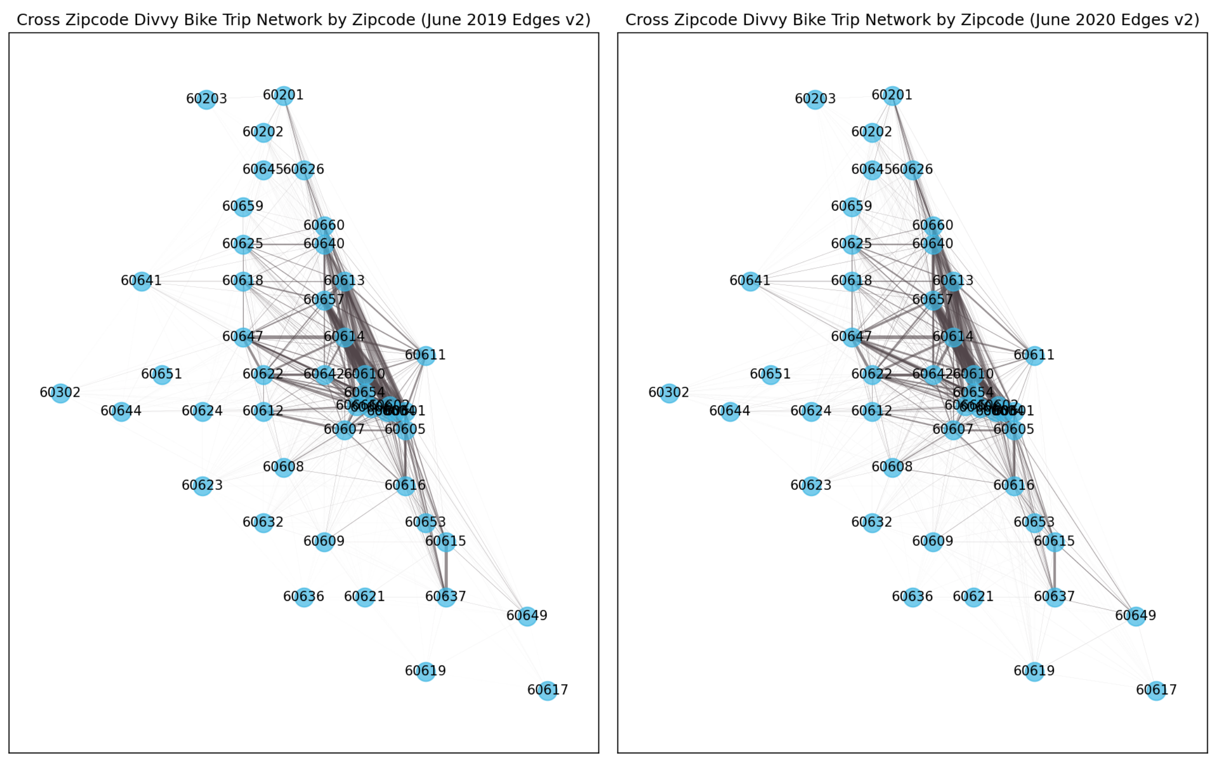
measuring edge weights is biased toward trips across small zip codes (especially in downtown) and undervalues trips across large zip codes. (b) We also want to explore how the tendency to bike across certain zip codes change in 2020, controlling for the total number of trips. The first way of measuring edge weights is unable to show us how edge weights between certain zip codes change ***relative to other zip codes***.

Figure 2: Cross Zip Code Divvy Bike Trip (Edge Weight v1 = Number of Cross Zip Code Trips)



To overcome these two limitations, we devised a second way to measure edge weights: *edge weights between two zip codes = the number of trips across the two zip codes * distance between the center of the two zip codes / the total number of trips in the month*. This edge weight approximates the average distance of the trips associated with a zip code, while accounting for the effect of the pandemic on overall human mobility. Figure 3 shows our network in 2019 and 2020 using the second edge measure. The main difference between the two measures could be seen on the edges across in the central region. By scaling the edges with distance, the edges across zip codes in the center city of Chicago become more balanced relative to the rest of the network.

Figure 3: Cross Zip Code Divvy Bike Trip (Edge Weight v2 = # of Cross Zip Code Trips * Distance Between Zip Codes / # of total trips in the month (“average distance”))



Based on the revised version of edge weight measure, Figure 3 shows that the edge patterns in the north, west, and south did not change much in June 2019 vs June 2020. It means that the pandemic did not disproportionately affect the divvy bike patterns of certain regions in Chicago over others. Furthermore, The relative connectedness of zip code 60611 declined in June 2020 compared to that of June 2019. It means that people were less likely to ride divvy bikes to get to or from 60611 during the pandemic. It makes sense because 60611 is Chicago’s major tourist destination, with attractions including the Navy Pier and the Magnificent Mile.

We applied both versions of the edge weights measure to test our network integration vs disintegration hypotheses using the following methods:

- **Weighted Degree**. In a weighted undirected network, the weighted degree of a node is the sum of the weight of edges to which the node is connected. In the first version where edge weight is the number of trips, the weighted degree represents the total number of trips that is connected to that zip code. In the second version where edge weight is an approximation of average distance, the weighted degree represents total distance by all bike trips connected to that zip code. If the network integration hypothesis (H2) is correct, we expect to see more nodes exhibit an increase in weighted degree, in both versions of edge weights.
- **Density**. Density is the proportion of observed ties in a network relative to the maximum number of possible ties. If the network disintegration hypothesis (H1), we expect to see a decrease in network density. Alternatively, if the network integration hypothesis (H2) is correct, we expect an increase in network density.
- **Average Clustering Coefficient**. The Average Clustering Coefficient of the network measures the degree of integration of the whole network. If the network integration hypothesis (H2) is correct, we expect to find an improvement in the average clustering coefficient for the June 2020 Divvy bike network compared to that of June 2019.

- **Community Detection Algorithms**. In network science, a community refers to a subgroup of nodes that are more densely connected to each other relative to other nodes in the network. Community detection algorithms, including Clauset-Newman-Moore Greedy algorithms and Louvain, are methods to identify communities in a network. If the network disintegration hypothesis (H1) is correct, we expect to find a greater number of smaller communities in the June 2020 Divvy bike network than the June 2019 Divvy bike network, and vice versa.
- **Modularity Measure**. Modularity measure refers to the ease in identifying communities in a network. A network with high modularity means that nodes share dense edges with other nodes in the same community, and weak edges with nodes in other communities. Modularity is biased toward detecting large communities, and has limitations in measuring small communities. Therefore, if the network integration hypothesis (H2) is correct, we expect to find an improvement in modularity measures for the June 2020 Divvy bike network compared to that of June 2019.

Results

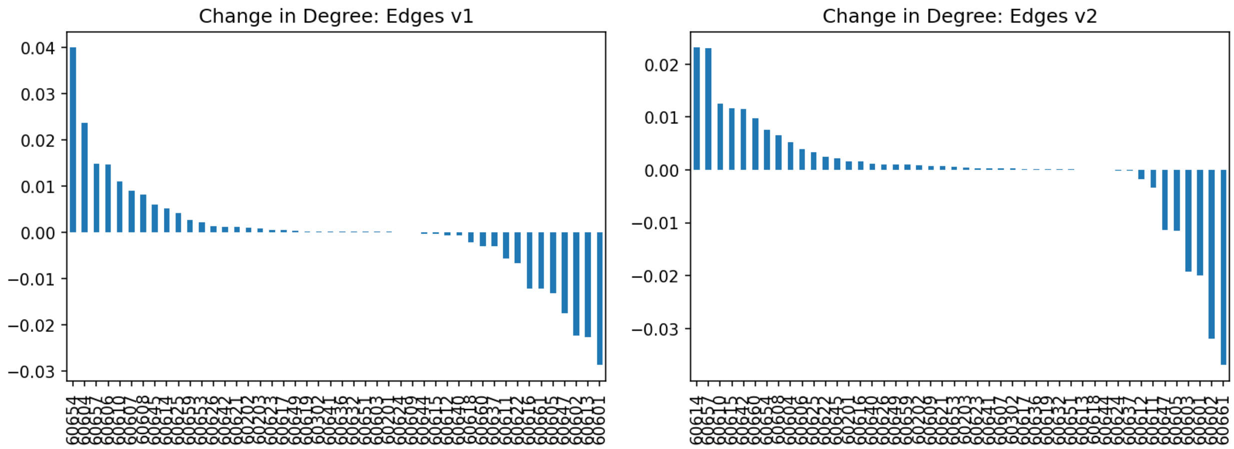
We started by examining three descriptive measures on the network: change in weighted degree, density, and clustering coefficient. Table 1 reports the density and clustering coefficient of our network. In 2019, the density of the network is 0.616, and in 2020 the density increases to 0.667. Considering the biking network as a whole, more bike trip connections have been made across zip codes, indicating an integration in the biking network across zip codes. The average clustering coefficient also increased from 0.824 in June 2019 to 0.838 in June 2020. The June 2020 network was more integrated and transitive. We also see the change in both versions of weighted degree for each node (zip code) in the network in Figure 4. First, we see more nodes experiencing increases in degrees (in networks of both edge weights) than decreases. In the network with the first edge measure, 11 nodes have decreases in degrees with the remaining 35 nodes enjoying increases in degrees. In the network with the second edge measure, the numbers are 16 and 30 respectively. Second, we see that center city zip codes[4] are the main zip codes that suffer from the decrease in degrees. This is likely correlated with the drop in tourism as a result of the covid pandemic. The nodes that experienced the highest increases in degrees are residential areas to the north of the loop area. These observations combine to suggest that the network integration hypothesis (H2) is more likely, and that local people do seem to increase bike as a transportation method in covid times.

Table 1. Density and Clustering Coefficient

	June 2019	June 2020
Density	0.616	0.667
Clustering Coefficient	0.824	0.838

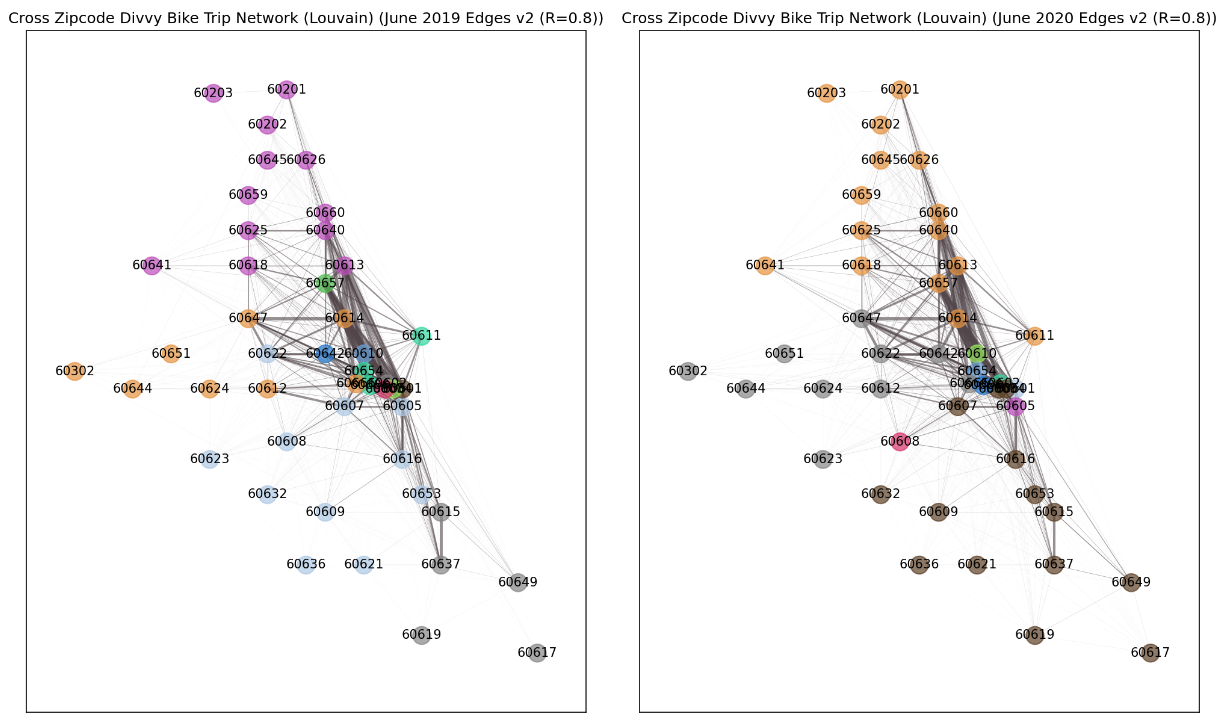
[4] The following center zip codes see relatively large decreases in degree: 60601, 60602, 60603, 60605, 60661. These zip codes cover most of the area of the loop, and a small area to the west of the loop (west of Chicago river).

Figure 4: Change in Weighted Degree



We further applied community detection algorithms to identify communities in the network, and examine if communities increased in size over time. Following Bargar et al (2014), we adopted the Louvain algorithm. Figure 5 shows the result of the Louvain algorithm on our network. Nodes that share the same color belong to the same “community” as discovered by the algorithm. In figure 5, comparing the network of 2019 and 2020, many smaller communities merge together to form larger ones. For example, in June 2019, the zip code nodes in the south side of Chicago are mainly divided into two communities. Those two communities merged into one in June 2020. In June 2019, zip codes like 60642, and 60657 did not belong to any communities. In June 2020, these zip codes were absorbed by the larger communities in the north and west side. As a result of the communities mergers, the average size of communities increased in June 2020. The integration of nodes and formation of larger communities confirm the network integration hypothesis (H2).

Figure 5: Louvain Community Detection (Edge Weights Version 2)



It is also interesting to note that in both graphs in Figure 5, the nodes in downtown Chicago did not form a coherent community, despite close proximity with each other[5]. It shows that people tend to use divvy bikes to commute to downtown zip codes from or to other regions, rather than travel within these downtown zip codes on divvy bikes.

We further quantify the results of community detection algorithms by looking at the change in modularity scores. In addition to Louvain, we also experimented with the Clauset-Newman-Moore greedy algorithm. Table 2 reports the modularity score for both

[5] We experimented by zooming in center city nodes alone, and found that these zip codes are either in one group or in one group each, when we varied the resolution parameter. This really speaks to the lack of structure of the center city nodes as opposed to the nodes in other areas. One possible reason is the interference of tourists in the area.

algorithms. In both cases, the modularity measures increased from June 2019 to June 2020. It means that nodes are more densely linked to other nodes in the same community in June 2020 than before the pandemic, and the communities are more well-defined while they increased in size. This observation further confirms the network integration hypothesis (H2).

Table 2: Modularity Measures (Revised Edge Weights)

Modularity	June 2019	June 2020
Clauset-Newman-Moore Greedy	0.0797	0.1430
Louvain	0.0870	0.1080

Conclusion

In this project, we applied community detection and clustering algorithms to test whether the cross zip code divvy bike network became more integrated or disintegrated as the result of the covid-19 pandemic. Our analysis confirmed the network integration hypothesis. It means that for cross zip code divvy bike trips during the pandemic, people increased their tendency to bike to downtown or other distant zip codes rather than neighboring zip codes.

This project can be improved in the following ways:

- Instead of treating every zip code as a node, it will be interesting to treat every bike station as a node. In this case, we will be able to include analysis of all divvy bike trips that start and end within the same zip code. The community detection visualization of over 600 nodes of bike stations will show some very interesting patterns.
- In this project, we only compared the trip patterns of two months in Chicago (June 2019 and June 2020). It will be useful to compare other months to see whether the network integration hypothesis remains robust in other months (e.g. March to October 2019 vs March to October 2020).
- In this project, we do not consider vertex attributes. Attributes such as average income level, or race composition, or covid-related variables could also be helpful in enriching the analysis.
- Future projects can also compare the public bike networks of different cities (e.g. Boston, DC, NYC, etc) to identify similarities and differences in the impact of covid-19 to biking networks.

References

Hu, Songhua, Chenfeng Xiong, Zhanqin Liu, and Lei Zhang. "Examining spatiotemporal changing patterns of bike-sharing usage during COVID-19 pandemic." Journal of transport geography 91 (2021): 102997

Padmanabhan, Vyas, Praveena Penmetsa, Xiaobing Li, Fatema Dhondia, Sakina Dhondia, and Allen Parrish. "COVID-19 effects on shared-biking in New York, Boston, and Chicago." Transportation research interdisciplinary perspectives 9 (2021): 100282.

Bargar, Alicia, Amrita Gupta, Srishti Gupta, and Ding Ma. "Interactive visual analytics for multi-city bikeshare data analysis." In The 3rd International Workshop on Urban Computing (UrbComp 2014), New York, USA, vol. 45. 2014.