DBA5104: Introduction to Network Science & Analytics

Mini Project:
Analyzing Resilience of Bike-Sharing Networks

Written by: **Group 10**

Name	Matriculation Number
Manan Lohia	A0262838U
Hsu Chu-Jing (Andrew)	A0262744B
Lin Yi Chia (Jerry)	A0174754W
Toshinori Yoshiyasu	A0262738W
Lam Ivan Chuen Chun	A0262765W

Task 1: Network Resilience (10 marks)

(1) Implement the two types of node deletion strategies on your network, and plot the associated changes in average shortest path length (i.e., average distance) & diameter as a function of the fraction of nodes removed. Do the resulting plots resemble the Fig. 2 on Pg. 12 of the Albert et al. (2000) paper above? Why do you (/don't you) see this pattern?

Nodes: Stations ID │ Edges: A user trip between the nodes │ Type 1: Failure/Loss │ Type 2: Attack

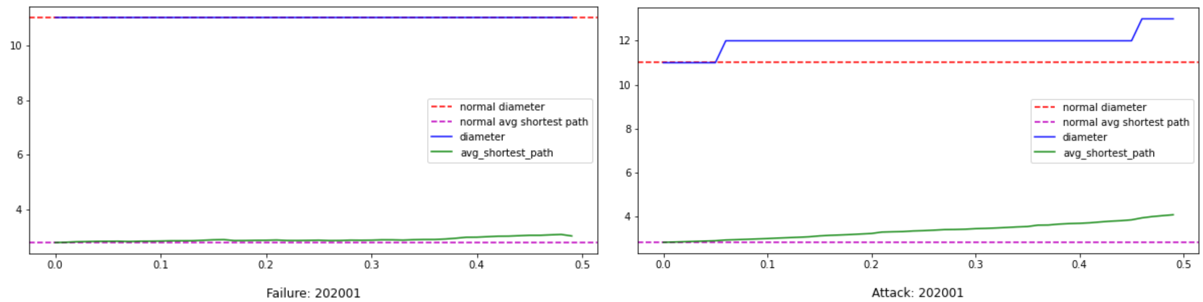


Figure 1: Diameter and Average Shortest Path After Type 1 and 2 Node Removals for January

As observed above, during Type 2 (attack) removals, both Diameter and Average Shortest Path Length (ASPL) tends to initially increase as nodes that are more connected are removed which is inline with the findings of Albert et al. There are also certain points where the Diameter suddenly increases and this may be due to certain nodes that are more connected being attacked, causing the network to split into disconnected components. While the increase in ASPL indicates that the network's communication efficiency has been disrupted. For Type 1 (failure) removals, nodes fail at random thus it seems to be removing nodes randomly as the nodes removed in January does not seem to significantly affect the Diameter and ASPL.

(2) Can you think of any other important graph-level metric, apart from average distance or diameter, that can potentially measure structural resilience of the network to such errors and attacks? Compute this metric, and test if this metric also shows significant variation in response to these two types of node deletions.

Largest Connected Component size (LCC) was the other metric chosen as this metric could help determine the connectivity and robustness of the network.

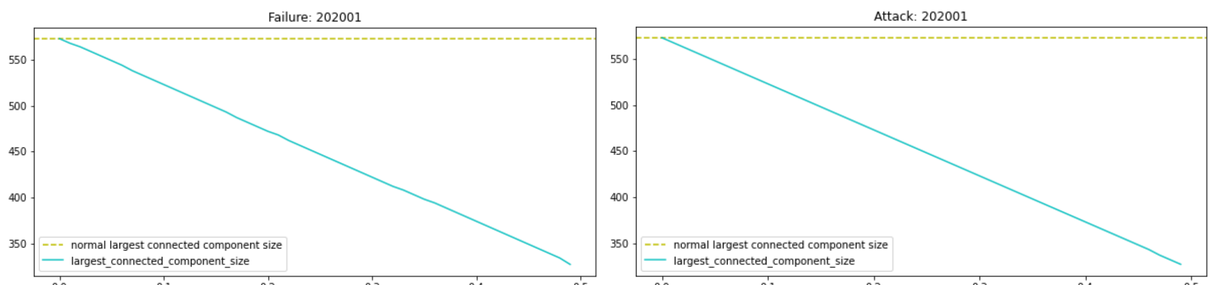


Figure 2: Largest Connected Component Size After Type 1 and 2 Node Removals for January

By observing the change in LCC after removing different portions of the nodes, it seems that the network is resilient to breakdowns in January as the size decreases gradually but no sudden drop, and that both removal strategies reduce LCC at the same rate. This indicates that although attack removals may be more effective at disrupting the network's communication pathways, both removal strategies similarly affect the network's overall connectivity and robustness.

In the plots in the subsequent months, there are sudden drops in the LLC where the diameter drops are caused by connected nodes being removed thus leaving leftover components. This could mean that within this bike network there are not many well-connected nodes that act as hubs as it takes the removal of specific nodes to decrease the LCC significantly. Another potential reason could be that bike networks may not be interconnected very well as there may be components that are only connected by a few bridge nodes thus if any node removal is not directed at these bridge nodes then the effect on LCC is more gradual.

(3) Repeat the above set of analyses for the network over time (e.g., for each month or quarter or year). Do you see any change in the resilience of the network over time? What patterns (if any) do you see from this temporal analysis?

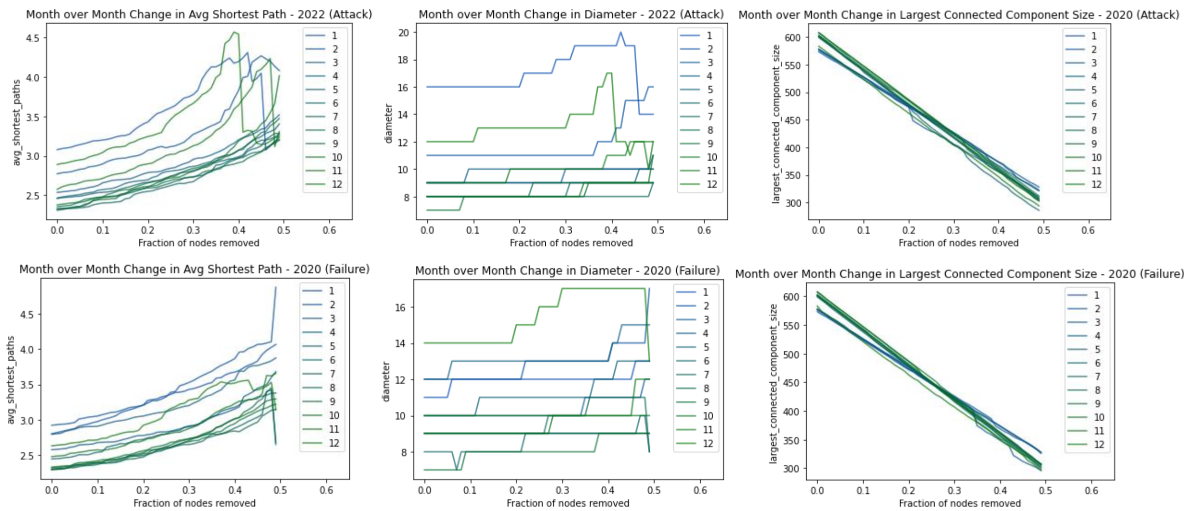


Figure 3: Diameter, ASPL, and LCC after Type 1 and 2 Node Removals Over 12 Months

As the original state of nodes changes each month as the desired routes by users can differ, however the trend for all three of our measurements seems to follow a similar pattern in both Type 1 and Type 2 node removals. Diameter, ASPL and LCC seem to move in correlation with each other and when more central nodes are removed, the Diameter and ASPL increases while LCC decreases.

There are changes in the resilience of the network over time as it depends on the network structure during each month as new nodes may be added which would mean new routes are taken by users. It can be observed that for certain months the LCC drops significantly instead of gradually during both types of node removals while in other months the LCC only decreases gradually. Also, for certain months, the LCC and ASPL suddenly drop after increases which may mean that the network consists of components that are connected by bridge nodes and once these bridge nodes are removed the LCC and ASPL decreases as the network is left with a series of unconnected components.

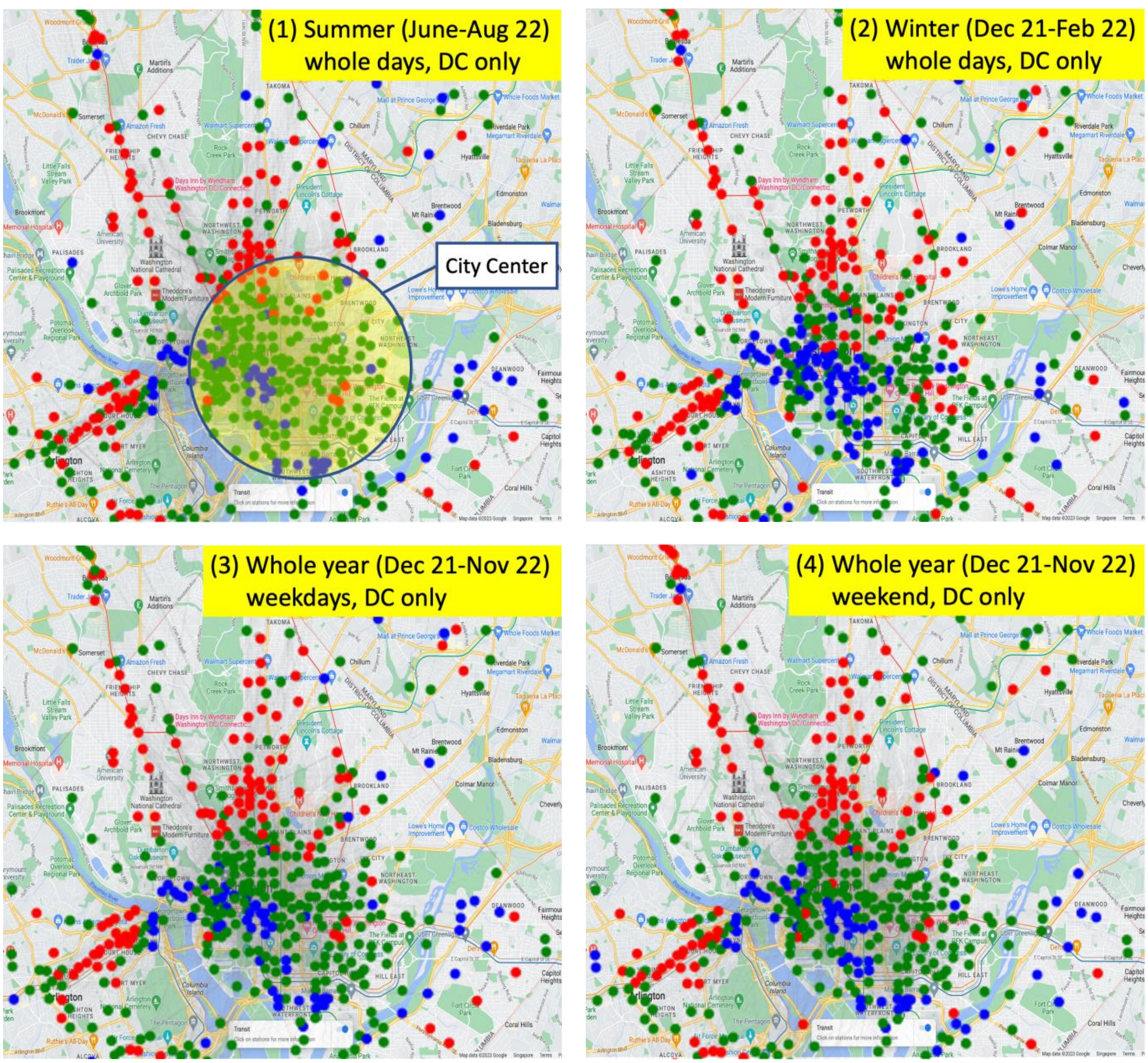
It is also interesting to note that this network exhibits the behavior of a scale-limited network as opposed to a scale-free network, which would be what we originally assumed the network to be as there would be more central nodes that are connected to many others and act as hubs while some only have a few connections that would mimic the structure of bike stations in more busy areas connected to stations in less busy areas.

Task 2: Additional Analyses (10 marks)

(1) Additional analyses. Connect/generalize some findings from this task to relate to what other bike-sharing companies in Singapore and around the world are facing.

2016 Member Survey Final Report (Capital Bikeshare, 2016) reveals two interesting facts. Firstly, Figure 41, p49 shows that customer satisfaction is the lowest with availability, which are “Availability of open docks” (23% answered “Poor”) and “Availability of bikes at docks” (24% answered “Poor”). Secondly, p54 indicates that the improvement of availability is demanded by heavy users, who usually bring a large profit to the company. Therefore, availability is analyzed in this part.

It is assumed that availability of open docks and bikes at docks deteriorate when the balance between the number of bikes coming to the station (call this “inbound trips”) and leaving the station (“outbound trips”') is disrupted. In the figure below blue nodes are stations with too many inbound trips, which would be likely to face “low availability of open docks”, red nodes are stations with too many outbound trips, which are considered likely to have “low availability of bikes at docks”, and green nodes are stations with healthy balance. It is classified as “too many” if the daily average of outbound/inbound trips is more than inbound/outbound trips by over5 10%. All edges are undirected and edge weights are the daily average of the total (inbound & outbound) trips between the two nodes. Only the DC area (Dec 21-Nov 22) is shown.



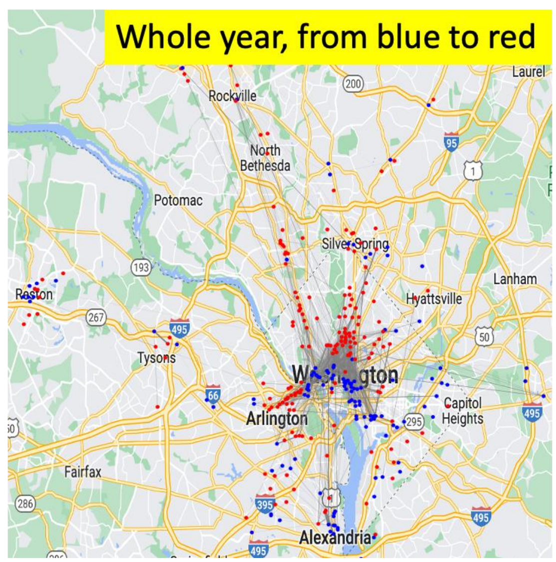
Findings and insights (new hypotheses) gained from the network graphs are summarized as follows.

Findings	Insights (new hypotheses)
(1) and (2): There is a less dense network with more imbalanced nodes in the city center in winter (Nov-Feb) comparing to in summer (June-Aug)	Temperature would affect user's behaviors. Perhaps severe weather conditions in winter such as low-temperature or snow prevent round trips by bike (increase one-way trips), causing more imbalanced nodes.
(3) and (4): Notable differences in terms of color, edges, weights are not observed between weekdays and weekends	User's behaviors are not so different between weekdays and weekends. (This is against our original hypothesis)
ALL: More blue nodes are in the city center, while more red nodes are in the outskirts areas especially on subway lines. Nodes on subway lines are strongly connected to nodes in the city center, not to nodes in other outskirts areas.	Some possible reasons are as follows: (1) Users who cannot catch the train (eg. operational troubles) in the morning rent a bike at a subway station to commute, but do not use a bike when returning home as they have more time. (2) Users who want to travel around the outskirts area come to the outskirts area by subway and rent a bike there but they come to the city center by bike, returning the bike there.

Singapore or other large cities tend to have a clear boundary between CBD and residential area and similar forms of public transportation network as Washington does. Therefore, it is possible that those cities could have a similar characteristic that is 'excessive bikes in the center, insufficient bikes in the outer areas'. If it is confirmed with other cities' samples too, the common characteristic would be a useful assumption especially for organizations that will launch a new bike sharing service in similar cities.

(2) Recommendations to bike-sharing companies

This figure below shows only trips (directed edges) from blue nodes (with excessive inbound trips) to red nodes (with excessive outbound trips). It is confirmed that there is a certain demand for these routes especially in the DC area.



It is recommended that Capital Bikeshare offer discounts for the routes from blue nodes to red nodes to promote use in the routes. This initiative will allow them to make customers move bikes from stations with excessive bikes to stations with insufficient bikes, optimizing bike allocation while earning revenue. It will not only improve bike and dock availability but also reduce labor costs If they have manually moved bikes from stations to stations. However, if the discount does not realize an increase in the number of trips, it will just end up with a revenue decrease. Therefore, It is necessary to perform additional price elasticity analysis to set the appropriate discount rate such that the total benefit gained from new demand and reduced labor costs exceeds the total costs of discount (amount of discount used by the existing customers).

[Reference] Capital Bikeshare. (2016). 2016 Member Survey Final Report. Retrieved from
https://d21xlh2maitm24.cloudfront.net/wdc/Capital-Bikeshare_2016MemberSurvey_Final-Report.pdf?mtime=20170303165531