Self-Balancing Bikes: Locally Re-routing Users to Improve the Flow of Bike Share Programs

Jahaziel Guzman
Brooklyn College
ja.xi.es@hotmail.com

Donald Hanson II
Adelphi University
donaldhanson@mail.adelphi.edu

Franky Rodriguez
St. Joseph's College
frodriguez@student.sjcny.edu

Briana Vecchione
Pace University
briana.vecchione@gmail.com

1. INTRODUCTION

In recent years, bike sharing programs have surged in popularity, doubling worldwide since 2012. New York City's bike-sharing program, commonly referred to as Citi Bike, has quickly become the largest bike-sharing scheme in the U.S. with 90,000 annual users and 20-40,000 trips per day. Because most bike shares are located in cities with concentrated business and residential areas, their systems suffer from load imbalance issues. To combat this issue, NYC Bike Share (NYCBS) currently employs vans to redistribute bicycles. Though moderately effective, this method is costly, accounting for up to 15% of gross revenue. Through the following research, we discover whether or not it would be reasonable to make small changes in rider routes to help balance the system.

In our trip data, there were over 5.5 million trips taken from July 1, 2013 to February 28, 2014. With this data, we found that the mean distance traveled was 1.81 km, and the median distance traveled was 1.42 km. This tells us that 50% of trips were less than 1 mile in distance. We considered all these trips when determining vehicle transports and the average availability throughout the system. We limited our data to weekdays between July 1 - November 31 for assessing the concentration of bikes at stations and for our trip simulations. If we divide the total number of docks in the system by the number of bikes, we observe that each station would be 57% full. If we had a balanced system, the concentration of bikes at each station would not deviate much from this number; however, we find that this is not the case.

We used two methods to assess the current status of the system. First, we classified stations if they were "congested" or "starved" at each time interval in our data. To do this, we calculated each station's availability by dividing the number of available bikes by the station's capacity for each interval throughout the day. Congested stations had concentrations over 80% and starved stations had under 20%. These metrics allowed us to assess two problems in our system: If a station is congested, users may not be able to park their bikes at that station. If a station is starved, users will not be able to start a trip from that station.

For each interval throughout the day, we averaged the number of congested and starved stations over all the days in our data to obtain a characteristic weekday. We observed that starvation is more common than congestion by almost 10% throughout the system. We also observed increases in both congestion and starvation between 8 and 10 AM. We can attribute the second observation to network flow that ensues during the morning rush hour.

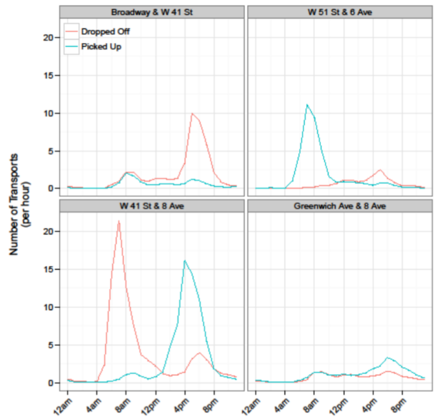

Figure 1: Vehicle transports averaged over all weekday for four representative stations

A significant portion of NYCBS's revenue goes to rebalancing the bicycles. NYCBS releases monthly reports that include the number of bicycles rebalanced each month, but do not report specific details, like when and where. We then attempted to find patterns in the workers' movements in order to figure out a way to remedy the imbalance. To do this, we found every instance in the data where a bicycle would start at a station that it did not end at. We assumed this occurrence was due to NYCBS workers rebalancing the bikes, despite the fact that some may have been attributed to faulty docks or taken for repairs.

Through this method, we were able to find that bicycles are picked up throughout the city and dropped off in the center of Manhattan. In most cases, these bikes are deposited near major transit hubs, including Port Authority, Penn Station, and Grand Central Station. Interestingly, there are stations where there are a high number of both pick-ups and a drop offs. From this, we were able to distinguish four unique station patterns. Figure 1 exhibits an example of each pattern, from right to left, in the following order: High pickup rate at a specific time, high drop-off rate at a specific time, high pickup and drop-off rate at different times, and little activity.

Figure 1 proves that van transports have a considerable influence on station activity. In order to discover just how much of an influence these vehicles have, we simulated a world in which there would be no van transports and eliminated any changes in station availability that would have occurred from vehicle rebalancing. All trips that occurred within our data that could not have been possible without rebalancing were classified as trip failures. Through our simulation, we found that 5.6% of trips could have not occurred without the vehicle transports. This amounts to around half of the amount of bicycles rebalanced.

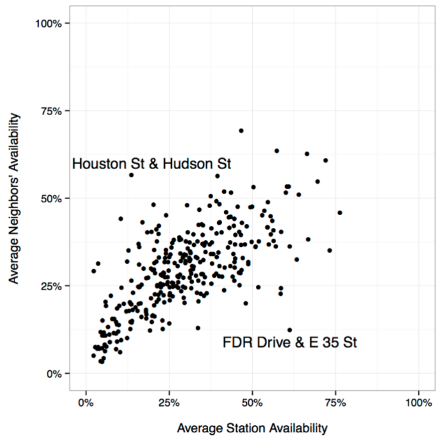

Figure 2: Average bike availability vs. average for three nearest neighbor stations for 6:00-6:15pm

To discover whether or not there was local re-routing potential, we computed some basic statistics about station proximities. We found that 91% of stations had another station within 300 meters, which equates to about the same length as a single Manhattan avenue. Once we distinguished the relative distance, we averaged each station's availability and compared it to the average of its three nearest neighbors over a 24-hour time period over all weekdays. Figure 2 shows this comparison at 6pm. From this, we were able to conclude that persistent differences exist between a station and its nearest neighbors for the same interval. Because the figure above displays averages, we can expect that within any given day, the differences would be much larger.

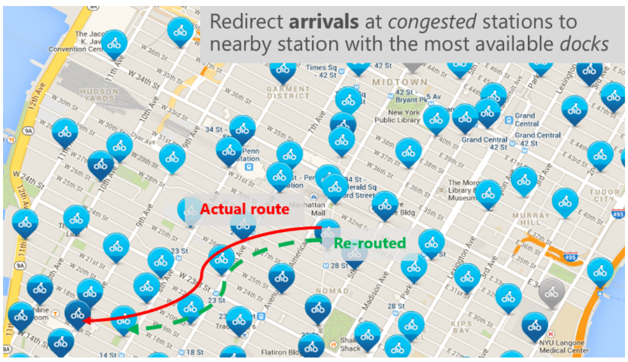

Figure 3: Schematic example of the re-routing algorithm

Noting our disparities between station availabilities, we then implemented a re-routing algorithm which acted as follows: It re-routed departures at starved stations to a nearby station with the most available bikes, and re-routed arrivals at congested stations to the nearby station with the most available docks. The algorithm would choose a station based on the availabilities of itself and its nearest neighbors. Figure 3 is a demonstration of the latter scenario, where the rider attempts to dock their bike at a station that is almost at capacity. In the simulation, the rider is redirected to another station within an avenue's distance with substantially lower availability.

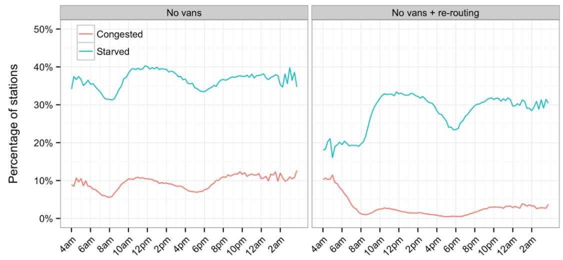

Figure 4: Comparison of starvation and congestion in the no van world vs. the local re-routing

We then assessed the system health using our re-routing scheme. Figure 3 depicts the performance of each of our simulations over a 24 hour period. As is evident, the simulation using rider re-routing performs significantly better than the simulation eliminating vans. It is important to note that we did not compare the rerouting scheme to the status quo because the vans tend to the demand of starved stations, and therefore look largely similar to the first simulation. From these results, we find that local re-routing drastically reduces congestion, but global patterns continue to drive most of starvation. When comparing trip failures, we found that local re-routing reduced the trip failure rate from 5.6% to 0.8%.

We take away three main findings from our research: First, we find that the New York City's bike sharing program suffers from both local and global imbalance. Second, we can conclude that local re-routing via a simple greedy algorithm appears to be a promising way to improve bike availability. Lastly, we can suggest that incentives be offered to riders to adopt re-routing schemes. Such incentives vary from introducing a point system for routes that rebalance the system to app development and/or improvements. For example, Citi Bike's mobile application could improve congestion by offering riders multiple routes to choose from when using their service.

2. ACKNOWLEDGMENTS

We would like to thank the Microsoft Data Science Summer School for financial support and Sharad Goel, Jake Hofman, Justin Rao and Hannah Wallach for mentorship.