Citi Bike Usage Balancing through Congestion Pricing

Paul Griffioen
Dept. of Electrical and Computer Engineering
Carnegie Mellon University
pgriffi1@andrew.cmu.edu

Anthony Jin
Dept. of Electrical and Computer Engineering
Carnegie Mellon University
xiaoxiaj@andrew.cmu.edu

1. Abstract

The issue of congestion in bicycle sharing systems is prevalent in many cities. This project proposes an incentive-based pricing scheme for New York City's Citi Bike system that minimizes bicycle congestion. It approaches bicycle sharing systems from a network perspective, analyzing and optimizing the network by providing monetary incentives to users with the intent of modifying network properties. These monetary incentives are part of a much larger optimization problem designed to minimize both user dissatisfaction and bicycle congestion. To demonstrate the effectiveness of this approach, the Citi Bike system is modeled and simulated using MATLAB, where monetary incentives are given to users such that user dissatisfaction and bicycle congestion are minimized.

2. Introduction and Motivation

The Citi Bike system, located in New York City, is one of many bicycle sharing systems that exist in the world. Since users unlock a bike at one station and drop it off at another station, bikes tend to follow unidirectional flows at various times of the day. As a result, some stations have few to no bikes at them while others are completely full.

This unidirectional flow of bikes frustrates users who would like to pick up a bike from an empty station or drop off a bike at a full station. As a result, this project seeks to address the issue of system congestion through a pricing scheme based on the optimization of network structure and dynamics. In addition, alteration of the network topology to reduce congestion will be explored through optimal placement of nodes (stations). Looking at the problem from a network perspective is a unique approach that has not been considered with previous work. In the end, this project seeks to analyze the Citi Bike network and minimize congestion by modifying its structure through changing people's incentives.

3. Previous Work

Previous work has taken many viewpoints to solve the bike sharing congestion problem. The work in [1] proposes a dynamic incentives system that involves future traffic prediction, truck-based re-positioning policies, and schema to compute monetary incentives for users. The authors in [2] use a similar approach, combining staff-based dynamic vehicle distribution and real-time pricing incentives for customers. In contrast, [4] and [2] look at the inventory management of individual bike-sharing stations and formulate appropriate convex optimization problems. Similar to the aforementioned approaches, this project will address the congestion problem as a convex optimization problem and introduce a congestion-based pricing scheme to modify user incentives. Unlike prior work, however, it takes a network standpoint and models the Citi Bike system as a dynamic network.

4. Approach

The Citi Bike network is built using available data from Citi Bike's website [5]. The nodes represent different bike stations and the weighted links represent the number of trips made from one station to another in a certain time period t. A few node and edge attributes are also of interest. Each node (bike station) has an identification number, latitude and longitude, maximum capacity, and current number of stored bikes. Meanwhile, each edge (bike trip) has a bike identification number, user age, and subscription type. These attributes are used as inputs in the convex optimization problem for the incentive-based pricing scheme.

The trip data in [5] does not include explicit information about the addition or removal of bikes to the system or the transfer of bikes by truck from one station to another. Consequently, this project can only analyze bike trips from one station to another. In addition, there is no information about the number of bikes at particular stations at specific points in time. As a result, an initialization procedure is used to infer this information (see Section 5.2).

4.1 Congestion Prediction Model

In the Citi Bike network, a node's congestion level at any given time is predicted using its in-degree and out-degree along with other known node attributes such as maximum capacity and current number of stored bikes. Additionally, the optimal congestion level of a node is to have the same amount of stored bikes and empty slots, allowing future users to either pick up or drop off a bike. This situation can be described by a simple math model such as equation 4 in Section 4.2.

4.2 Sub-Functions and Parameters

To provide accurate monetary incentives to users, a convex optimization problem is formulated that seeks to minimize user dissatisfaction. Each term used in the formulation of this problem is applied in a certain time interval t. $u = \{u_1, u_2, ..., u_n\}$ represents the set of n users in the Citi Bike network, A_u represents the age of user u, and r_u represents the maximum distance user u is willing to walk. The authors in [1] conducted a survey on users of a real-world Bike Sharing System in a European city to study the maximum walking distance that a user is willing to take. While roughly 20% of the response were "unwilling to walk" or "unwilling to participate at any cost", most of the other participants are only willing to walk less or equal to one mile. Additionally, the results of the survey suggests that the relation between the cumulative willingness to walk and the maximum walking distance is a decaying exponential. In order to facilitate the optimization problem formulation (i.e. keep all variables continuous), we model r_u as a decaying exponential ranging from $r_u(min)$=0.1 and $A_u(max)$=60 to $r_u(max)$=1 and $A_u(min)$=18. In other words, an elderly person will be offered with closer candidate stations and a young person will be offered with candidate stations that are further away. If a user is older than 60 years old, he or she will be assigned with r_u(min). Similarly, a biker younger than 18 years old is assigned with r_u(max). The equation is found to be

$$r_u = \begin{cases} 1 & 0 \leq A_u \leq 18 \\ 2.683e^{-0.055A_u} & 18 \leq A_u \leq 60 \\ 0.1 & A_u \geq 60 \end{cases} \quad (1)$$

$s = \{s_1, s_2, ..., s_m\}$ is the set of m bike stations in the Citi Bike network, and s_{pu} and s_{du} represent user u's intended pick-up and drop-off stations, respectively. As a result, $d_{pu} = \{d^1_{pu}, d^2_{pu}, ..., d^m_{pu}\}$ and $d_{du} = \{d^1_{du}, d^2_{du}, ..., d^m_{du}\}$ denote the set of distances between the set of all stations s and station s_{pu} or s_{du}, respectively.

$$dist(s_1, s_2) = 7922 \arctan \frac{\sqrt{a}}{\sqrt{1-a}} \quad (2)$$

$$a = (\sin \frac{1}{2} d_{lat})^2 + (\cos l_{lat1})(\cos l_{lat2})(\sin \frac{1}{2} d_{lon})^2 \quad (3)$$

$$d_{lon} = l_{lon2} - l_{lon1} \quad (4)$$

$$d_{lat} = l_{lat2} - l_{lat1} \quad (5)$$

$$d_{pu} = dist(s_{pu}, s) \quad (6)$$

$$d_{du} = dist(s_{du}, s) \quad (7)$$

G_s represents the congestion level at station s, where C_s denotes the maximum capacity at station s, v_s denotes the current number of bikes at station s, and o_s and i_s denote the out-degree and in-degree of station s, respectively. k is a constant that weights the impact of the out-degree and in-degree on the congestion level.

$$G_s = |\frac{1}{2}C_s - v_s - k(i_s - o_s)| \quad (8)$$

Next, $s_{puc} = \{s^1_{puc}, s^2_{puc}, ...\}$ and $s_{duc} = \{s^1_{duc}, s^2_{duc}, ...\}$ represent user u's set of candidate stations for pick-up and drop-off, respectively.

$$s_{puc} = dist(G_s, d_{pu}, r_u) \quad (9)$$

$$s_{duc} = dist(G_s, d_{du}, r_u) \quad (10)$$

$d_{puc} = \{d^1_{puc}, d^2_{puc}, ...\}$ and $d_{duc} = \{d^1_{duc}, d^2_{duc}, ...\}$ denote the set of distances between the set of stations s_{puc} and station s_{pu} or between the set of stations s_{duc} and station s_{du}, respectively.

$$d_{puc} = dist(s_{pu}, s_{puc}) \quad (11)$$

$$d_{duc} = dist(s_{du}, s_{duc}) \quad (12)$$

In addition, $p_u = \{p^1_u, p^2_u, ..., p_{max}\}$ represents the ordered set of payments offered to user u, and B represents the current budget for the entire Citi Bike network. D_u denotes the dissatisfaction of user u, and w_1, w_2, and w_3 are constants that weights the importance of a station's congestion level and a user's age and the payment he or she receives on the user's dissatisfaction. Additionally, $d_c = \{d^1_c, d^2_c, ...\}$ represents all possible combinations of candidate stations d_{puc} and d_{duc}, while a_j is a constant that weights the likelihood of user u taking an alternative pair of pick up and drop off stations and choosing the corresponding payment option p_{uj}. Since a user is less likely to take a longer alternative route, a_j is biased in favor of the candidate stations that are closest to user u's intended stations. In order to aggressively decongest the network, w_1 is assigned with a much higher value than w_2 and w_3.

$$D_u = w_1 G_s + w_2 A_u a_j (d^j_c) - w_3 \sum_{j=1}^{m} a_j p^j_u \quad (13)$$

These sub-functions are primarily used to compute the congestion level at a particular station and a particular user's dissatisfaction. The congestion level is simply equal to the absolute value of half of the station's capacity ($\frac{1}{2}C_s$) minus the current number of bikes at the station (v_s) minus an estimate as to how many docks will be filled in the near future ($k(i_s - o_s)$). As a result, the congestion level describes the amount of fluctuation around a station's optimal capacity, which occurs when the station is half full. The user dissatisfaction is designed so that dissatisfaction increases with congestion level (G_s), the distance the user has to walk to modify his or her original route ($c_3 A_u(c_1 d_{puc} + c_2 d_{duc})$), and the user's age ($A_u$). However, user dissatisfaction decreases with respect to the monetary incentive the user receives ($\sum_{j=1}^{m} a_j p_{uj}$).

4.3 Objective Function

The objective function is designed such that individual user dissatisfaction is minimized.

$$\min D_u \tag{14}$$

4.4 Constraints

The objective function has a few constraints. First, the distance (d_{puc} and d_{duc}) between each candidate station and a user's intended station must not exceed the maximum distance the user is willing to walk (r_u). Second, each station's capacity (C_s) cannot be exceeded by the station's number of current bikes (v_s). Third, the budget available to each user (p_u) in the network for time interval t is constrained by the total budget (B) and the number of users in the system (n).

$$d_{puc} \leq r_u \tag{15}$$

$$d_{duc} \leq r_u \tag{16}$$

$$v_s \leq C_s \tag{17}$$

$$p_u \leq B/n \tag{18}$$

$$0 \leq \text{all parameters except } D_u \tag{19}$$

5. Experimental Setup and Practical Results

5.1 Network Analysis

Based on the network definition presented in Section 3, a sample dynamic network is constructed to represent bike usage from Thursday to Saturday in a normal week. First, Citi Bike data from the first three days of September 2016 is extracted from the original data set in [5] to generate a node table with 583 nodes and an edge list with more than 130,000 edges. Second, the node table and edge list are imported to Gephi as a dynamic network. The Gephi plugin *Geo Layout* is used to arrange the nodes according to their geographic locations, while network properties are dynamically changed by filtering out edges outside a particular time window. In this analysis, the time window for edge filtering is set to one hour.

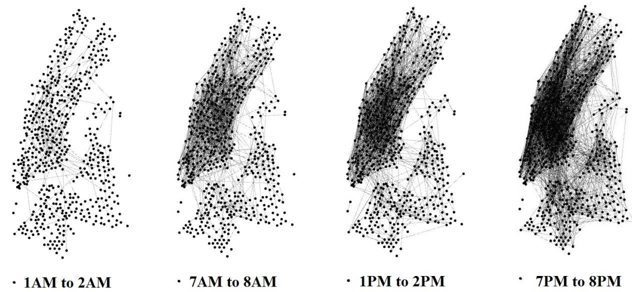

Figure 1: Citi Bike Network on September 1, 2016

Figure 1 shows the network at four different time intervals on September 1, 2016. From left to right, the figures represent the network between 1AM and 2AM, 7AM and 8AM, 1PM and 2PM, and 7PM and 8PM. Figure 2 illustrates how some of the network properties change over time. Every day, the maximum average node degree is observed between 4PM and 7PM, corresponding to the evening rush hour. On the two weekdays, there is also a secondary peak between 7AM and 8AM, reflecting the morning rush hour. As expected, this secondary peak is not present on September 3 since it is a Saturday. On the contrary, the network's average path length stays relatively consistent during the day and in the early evening. Bikers seem to take longer trips around midnight and in the morning (between 6AM and 7AM), while the trips between midnight and 6AM tend to be much shorter.

As for community structure, the network's modularity and formation of communities exhibit a periodic pattern. During the day, the network has high modularity and the nodes can be grouped into a few communities. At night, the network is less modular and only a handful of nodes can form communities with others. This observation is in line with fact that bike movement patterns tend to be unidirectional during the day (when bikers go to work) and disperse at night (when bikers go home in different directions).

This network setup has three limitations. First, the congestion level at each station (i.e. maximum capacity versus current number of stored bikes) is not taken into account. Second, the distance between each pair of nodes is assumed to be the same while a more realistic analysis should also take into account the physical distance between stations. Third, only a very small subset of the available data is exploited. These issues are addressed in Section 5.3.

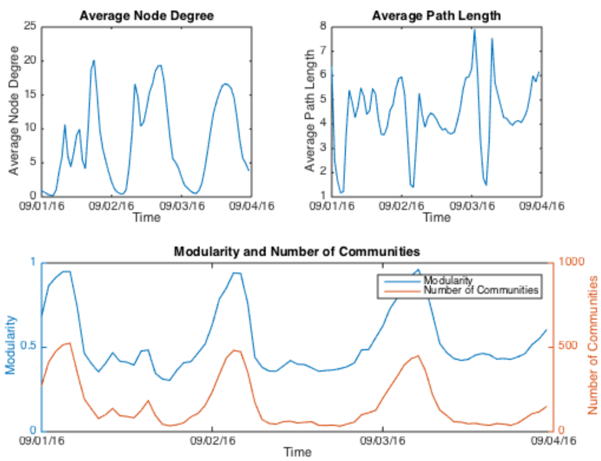

Figure 2: Citi Bike Network Properties

5.2 Initialization Procedure

The objective of the initialization procedure is to determine an initial state of the Citi Bike system at any point in time. The procedure starts by identifying a start time and assuming that no bikes are initially in the system. Over a specified time period, bikes are added to the system using incoming trip data for each station and are tracked by their identification numbers to determine which stations they are at. At the end of this period, it is assumed that all bikes have been added to the system, and each station is assigned with an initial number of bikes. Future trip data can now be used to determine the number of bikes at particular stations at any time after the initialization procedure.

5.3 Citi Bike Network Simulation

In order to demonstrate whether or not the aforementioned optimization problem results in decongesting the Citi Bike network, a simulator is created using MATLAB. The simulator first uses an entire month's trip data to implement the initialization procedure and estimate how many bikes are at each station. Beginning with the next month, trip data entries are stepped through for a time period specified by the user. Each entry has a start station and an end station corresponding to the biker's intended pick-up and drop-off stations, which are defined in the optimization problem. If the congestion level is high for one or both of these intended stations (too empty for pick-up or too crowded for drop-off), the biker is given a list of alternative pick-up and drop-off stations. Currently, the simulator assumes that the biker will choose a crowded pick-up station and a relatively empty drop-off station from the list of alternative stations. This assumption is made in order to see how the system's network properties are changed when each biker actively tries to decongest the network. Once the optimization problem is solved, this model will be replaced with monetary incentives and probabilities that model the biker's choice of alternate stations.

m2/no_incentives_before.png m2/no_incentives_after.png

Figure 3: Citi Bike Network with No Incentives

m2/incentives_before.png m2/incentives_after.png

Figure 4: Citi Bike Network with Optimal Station Choices

After the biker chooses start and end stations, the state of the Citi Bike system is updated with respect to the number of bikes at each station. Figure 3 shows the state of the Citi Bike system at 7AM and 10AM on March 3, 2016 when bikers are provided with no monetary incentives. Figure 4 shows the state of the system at the same times when bikers always choose pick-up and drop-off stations that will decongest the network. In each figure, the dots represent the stations, and the size of a dot is proportional to how many bikes are currently at that station. As shown, the network stays relatively congested when bikers are not provided with any incentives, but the network quickly decongests itself when all bikers choose optimal alternative stations (i.e. least congested stations within a one mile radius from the bikers' intended pick-up and drop-off stations). In light of this, we are confident that successfully solving and implementing the optimization problem will decongest the Citi Bike network.

Although the simulator overcomes the limitations of the basic network setup described in Section 5.1, it has a few other limitations. First, there is no information about the maximum capacity for approximately 10% of the bike stations, so an estimate must be used. Second, there is no data about bikes being moved to other stations by trucks, so this flow of bikes is unaccounted for. Because of this lack of information, the initial state of the Citi Bike system is slightly inaccurate.

6. Conclusion and Short-Term Plans

This work proposes a network-based framework to decongest the Citi Bike system via convex optimization. After constructing the system as a dynamic network, a MATLAB simulator is used to modify the network's properties using monetary incentives. In addition, the objective function and constraints are defined for the convex optimization problem. Currently, the framework assumes bikers choose crowded pick-up stations and relatively empty drop-off stations in order to illustrate how an optimal choice of incentives will affect the network. However, this framework will be replaced by accurate incentives and probability distributions once the convex optimization problem is solved. In this work, Paul Griffioen focused on developing the MATLAB simulator and Anthony Jin focused on formulating the convex optimization problem.

Future work of this project will focus on solving the proposed optimization problem and integrating the results with the present framework. By Milestone 3, we will finish implementing the decongestion framework with convex optimization and propose the related pricing scheme.

References

[1] A. Singla, M. Santoni, G. Bartok, P. Mukerji, M. Meenen, and A. Krause. "Incentivizing Users for Balancing Bike Sharing Systems." *AAAI*, pp. 723-729, 2015.

[2] J. Pfrommer, J. Warrington, G. Schildbach, and M. Morari. "Dynamic Vehicle Redistribution and Online Price Incentives in Shared Mobility Systems." *Intelligent Transportation Systems, IEEE Transactions on*, 15(4): 1567-1578, 2014.

[3] M. Rainer-Harbach, P. Papazek, B. Hu, and G. R. Raidl. "Balancing Bicycle Sharing Systems: A Variable Neighborhood Search Approach." *Springer*, 2013.

[4] T. Raviv and O. Kolka. "Optimal Inventory Management of a Bike-Sharing Station." *IIE Transactions*, 45(10): 1077-1093, 2013.

[5] Citi Bike, "System Data." Motivate International, Inc. Accessed 22 September 2016. [Online]. Available: https://www.citibikenyc.com/system-data.