

TRANSPORT FINDINGS

Identifying Optimum Bike Station Initial Conditions using Markov Chain Modeling

Mohammed Almannaa*, Mohammed Elhenawy†, Hesham Rakha‡

Keywords: optimization, bike share system, markov chain

<https://doi.org/10.32866/6801>

Transport Findings

Bike sharing systems (BSSs) are being deployed in many cities because of their environmental, social, and health benefits. To maintain low rental costs, rebalancing costs must be kept minimal. In this paper, we use BSS data collected from the San Francisco Bay Area to build a Markov chain model for each bike station. The models are then used to simulate the BSS to determine the optimal station-specific initial number of bikes for a typical day to ensure that the probability of the station becoming empty or full is minimal and hence minimizing the rebalancing cost.

RESEARCH QUESTION AND HYPOTHESIS

Bike sharing systems (BSSs) suffer from a central recurring problem: imbalance. Many bike stations become either empty or full during their daily operation. We hypothesize that we can reduce the cost of balancing bike stations by optimizing the number of bikes at each station at the start of the day, thus reducing the need for a dynamic balancing system (Schuijbroek, Hampshire, and van Hoes 2017; Raviv and Kolka 2013; Lu 2016). We formulate our hypothesis by modeling each station using a Markov chain.

METHODS AND DATA

This study uses Ford GoBike's BSS docking station data collected from August 2013 to August 2015 in the San Francisco Bay Area, as shown in Figure 1 (Bay Area Bikeshare 2016). The data provide the number of bikes at each station in one-minute intervals.

We used the discrete time-homogeneous Markov chain on a finite state space to model the system. We defined the state space as being all the possible states a station could be in. That is to say, if station s had N_s docks, then the number of states for that station would be $N_s + 1$, where the "empty station" is counted as one possible state.

* Virginia Polytechnic Institute and State University (VA); Civil and Environmental Engineering

† Queensland University of Technology; Centre for Accident Research and Road Safety

‡ Virginia Polytechnic Institute and State University (VA); Charles E. Via, Jr. Dept. of Civil and Environmental Engineering

ORCID iD: 0000-0002-5845-2929 Link: <https://sites.google.com/a/vt.edu/hrakha/>

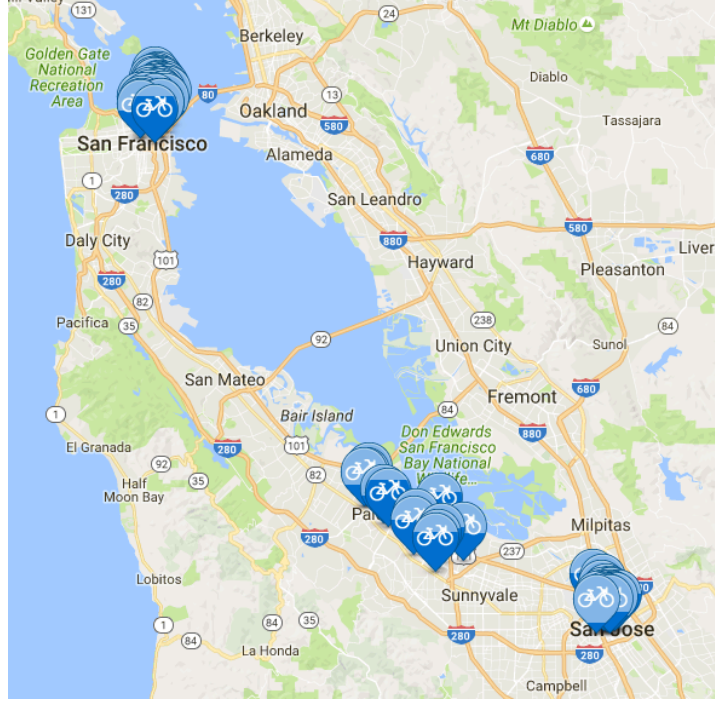


Figure 1: Locations of the 70 Stations Covering Five California Cities: San Francisco, Palo Alto, Mountain View, Redwood City, and San Jose (Bay Area Bikes, 2016).

A matrix $X_{s,d,h}$ was constructed for each station, $s \in S$, day of the week, $d \in 7$, and hour of the day, $h \in 24$ (i.e., a total of $S \times 7 \times 24$ X matrices were constructed of size $(N_s + 1) \times (N_s + 1)$). Using a specific X matrix, the transition frequency matrix was created by computing the elements f_{ij} , where $i, j \in \{1, \dots, N_s + 1\}$. The elements f_{ij} represent the number of times a transition occurred from state i to state j over a one-minute interval at a specific station, for a specific day of the week, and within a specific hour of the day. The transition probability matrix for a specific station, s , hour of the day, h , and day of the week, d , was then computed as $p_{ij} = \frac{f_{ij}}{\sum_{j=1}^{N_s+1} f_{ij}}$. The calculated transition matrices above are the one-step transition matrices for a specific station, day of the week, and hour of the day. Each transition (i.e., the time tick) is conducted per minute, making the movement between states as smooth as possible throughout the hour.

The probability distribution of available bikes at the end of the day at a particular station is shown in Equation 1.

$$\begin{aligned}
 &P(x^{\text{end of the day}} = q | x^{\text{start of the day}} = m) \\
 &= (P(x^{\text{end of the first hour}} = q | x^{\text{start of the first hour}} = m)) * \prod_{h=2}^{\text{last hour of the day}} P_h
 \end{aligned}
 \tag{1}$$

Here, P_h is the 60-minute transition matrix obtained from simulating the corresponding one-step transition matrix.

Equation 1 finds the probability distribution of the available bikes at the end of the day given that the station started the day with m bikes. We count all possible paths from m at the very first hour of the day to all possible values of m at the end of the day. We use the corresponding transition matrix to simulate the Markov chains in order to produce a probability distribution that describes the likelihood of a particular state at the end of the hour. This leads to the creation of a probability distribution of available bikes at the end of the first hour. After that, we can use this probability distribution as the initial state probabilities for the following hour and create the next probability distribution, which is the next 60-step transition matrix. This procedure is repeated until we reach our target hour and draw the final probability distribution as a function of each initial condition.

When running the Markov chain, our objective function was to find the best initial conditions to maximize the probability of the station operating at a bike-to-capacity ratio (number of bikes relative to the capacity of the station) within the range of 0.25 to 0.75 at the end of each hour, as shown in Equation 2.

$$\max_i \sum_{h=1}^{24} W_h \sum_{j=N_{\min}}^{N_{\max}} P_{ijh} \quad (2)$$

where i is the initial condition of station s , h is the hour of the day (considered only the hours from 6:00 a.m. to 8:00 p.m. in our case), W_h is the weight assigned to hour h (assumed to be 1.0), j is the expected state of the station at the end of the hour, N_{\min} and N_{\max} are the upper and lower desired bounds of the station status (in our case: $N_{\min}=0.25 \times (N_s + 1)$ and $N_{\max}=0.75 \times (N_s + 1)$), N_s is the capacity of station s , and P_{ijh} is the probability of having an i initial state and a resulting j state at the end of hour h .

FINDINGS

We used the BSS data to build the Markov chain for each station and day of the week combination to investigate the daily imbalances and identify the optimal inventory level that would minimize the probability of a station reaching an empty or full state. When analyzing the results, we first looked at all 70 stations, considering different initial conditions to identify stations that would benefit most from optimizing the initial station state. We grouped stations into three categories:

- (1) Those that have an imbalance issue but a small probability ($\leq 10\%$) for 25% of the initial conditions;
- (2) Those that have an imbalance issue with a medium probability (11% to 25%) for 25% to 45% of initial conditions; and

Table 1: Percentage of Each Category for Each City

City	Category		
	(1) Imbalance probability of $\leq 10\%$ for 25% of initial conditions	(2) Imbalance probability of 11–25% for 25 to 45% of initial conditions	(3) Imbalance probability $> 25\%$ for $> 45\%$ of the initial conditions
San Jose	43.75	12.50	43.75
Redwood City	57.14	28.57	14.29
Mountain View	14.29	57.14	28.57
Palo Alto	80.00	20.00	0.00
San Francisco	0.00	20.00	80.00

(3) Those that have an imbalance issue with a large probability ($> 25\%$) for $> 45\%$ of the initial conditions.

In Table 1, we present each category's percentage separately by city, as a previous study showed that there were close to no trips between the five cities (Ashqar et al. 2017).

As shown in Table 1, San Francisco has the highest percentage of category 3 stations, followed by San Jose. This demonstrates that San Francisco BSSs experience high bike demands, and thus are more likely to have an imbalance problem during the day. Our proposed approach would be less effective for the San Francisco BSSs and more effective for the other cities given that the daily evolution of states for San Francisco varies considerably.

Our analysis shows that the optimal initial conditions vary from day to day at the same station, and thus we present the optimal initial conditions for each day of the week for one selected station in Mountain View and one in San Francisco. Note that we made two assumptions when choosing the optimal initial conditions: (1) the bikes are taken from an infinite pool, meaning we have no constraints on the available inventory, and (2) there is no interaction between stations. The optimal station state is assumed to occur when the bike-to-capacity ratio ranges between 0.25 and 0.75 over the entire day, thus minimizing the probability of reaching either an empty or full state. Table 2 presents the optimum three initial states for stations 26 and 59 that result in the highest probability of maintaining a bike-to-capacity ratio ranging between 0.25 and 0.75 for the entire day. As was demonstrated earlier, the results of Table 2 demonstrate that there is a lower probability of being able to maintain the San Francisco station in the optimum range over the entire day, as was discussed earlier.

ACKNOWLEDGMENTS

This effort was funded by the University Mobility and Equity Center (UMEC).

Table 2: The Optimal Initial Conditions for Stations 26 and 59^a

	Station #26 Mountain View			Station #59 San Francisco		
	1st	2nd	3rd	1st	2nd	3rd
Saturday	6 (0.74)	5 (0.74)	7 (0.74)	7 (0.65)	8 (0.63)	9 0.63
Sunday	6 (0.74)	5 (0.74)	4 (0.73)	7 (0.62)	8 (0.62)	9 (0.62)
Monday	4 (0.70)	5 (0.69)	3 (0.69)	8 (0.42)	9 (0.42)	10 (0.41)
Tuesday	4 (0.71)	3 (0.70)	5 (0.70)	7 (0.42)	8 (0.41)	9 (0.41)
Wednesday	5 (0.71)	4 (0.71)	6 (0.69)	7 (0.38)	9 (0.37)	9 (0.37)
Thursday	4 (0.70)	5 (0.70)	6 (0.68)	7 (0.42)	8 (0.41)	9 (0.41)
Friday	5 (0.71)	4 (0.70)	6 (0.69)	7 (0.42)	8 (0.41)	10 (0.41)

^aOptimum number of initial bikes and probability of achieving the desired bike-to-capacity ratio.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

REFERENCES

- Ashqar, H.I. et al. 2017. "Modeling Bike Availability in a Bike-Sharing System Using Machine Learning." In *Models and Technologies for Intelligent Transportation Systems (MT-ITS), 2017 5th IEEE International Conference*.
- Bay Area Bikeshare. 2016. "Introducing Bay Area Bike Share, Your New Regional Transit System." 2016. <http://www.bayareabikeshare.com/faq#BikeShare101>.
- Lu, C.C. 2016. "Robust Multi-Period Fleet Allocation Models for Bike-Sharing Systems." *Networks and Spatial Economics* 16: 61–82.
- Raviv, T., and O. Kolka. 2013. "Optimal Inventory Management of a Bike-Sharing Station." *IIE Transactions* 45: 1077–93.
- Schuijbroek, J., R.C. Hampshire, and W.J. van Hoes. 2017. "Inventory Rebalancing and Vehicle Routing in Bike Sharing Systems." *European Journal of Operational Research* 257: 992–1004.