

# Developing a Capacity Correction Factor through an Analysis of Trips in Boston's Bike-Sharing System

Talal Al-Mulla, Catherine Cheng

---

## Abstract

Bike-sharing systems have transformed the transportation decisions of travelers, becoming one of the fastest growing modes of transportation. We examine the Boston Hubway system as a network, developing a capacity correction factor to assess the possible redistribution problem faced by this network. Our results show that Hubway has, to a first order, strategically placed their stations to avoid segmenting its network. We outline areas of future analysis that can offer increased insight into bike-sharing systems.

## Introduction

Hubway, Boston's first bike-sharing system, was first launched in June 2011. Implemented as a component of the Boston Bikes Program, Hubway attempts to increase and strengthen Boston's cycling community by promoting a lively and healthy lifestyle. Through its growth, Hubway has developed into a larger network outside of Boston that also includes Cambridge, Brookline, and Somerville. As a new form of transportation in the metro-Boston area, Hubway is transforming the transportation decisions of travelers as it offers, as some people claim, a healthier, more time efficient, and more convenient travel option. However, one of the issues that the Hubway system currently faces is the uneven distribution of bikes at its various stations. Since users are free to take bikes from any station to another, the travel needs and routes of users primarily control the distribution of bikes throughout the Hubway system.

In an attempt to better understand the Hubway system, we analyze this bike-sharing system as a network. In the Hubway network, the nodes are the Hubway stations and the edges are represented by the user trips. The edges are weighted based on the number of trips made between the two stations. We analyze this network as having directed edges, as it is

significant where the bikes are coming from and going to.

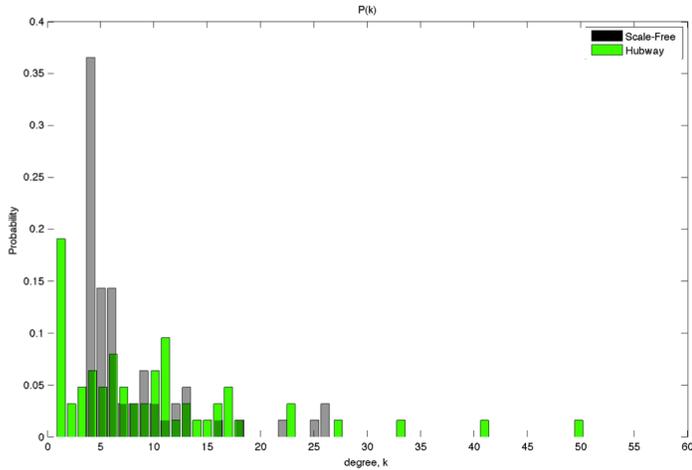
In this paper, we examine the Hubway network as a weighted network, taking into account the frequency of trips made between two stations. We investigate the connectivity between stations in relation to the number of trips made. In addition, we identify whether or not Hubway has strategically placed their stations to avoid having stations that are too frequently or rarely visited. We also identify stations that suggest a redistribution issue by analyzing the number of trips made in and out of each station.

## Method and Results

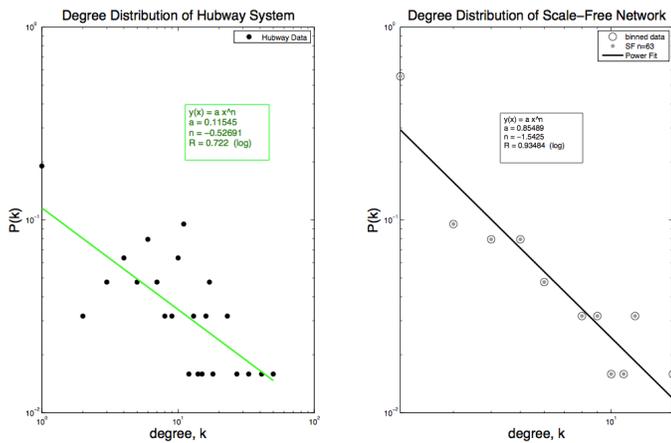
Hubway released the trip history data for over half a million trips in the form of a data visualization challenge at the end of 2012. This data included information such as the date, time, origin and destination stations, bike number, etc. of trips made. After examining the number of trips made throughout the day, we decided to analyze the hours of 6-9am, which experienced the most Hubway usage in the morning. We filtered the Hubway data to include only weekday trips from this time frame.

In order to model a network that best represents the Hubway system, we exclude the edges between stations that had a weight of less than 50. This means that over the entire lifetime of Hubway operations, fewer than 50 trips were made between these two stations, indicating that those routes were uncommon and not representative of the typical travel flow of riders.

To better understand what type of network the Hubway system most resembles, random world networks, small world networks, and scale-free networks of the same number of nodes ( $n=63$ ) and average degree ( $\langle k \rangle = 4.857$ ) as the Hubway network were created. Because the Hubway network contains such highly connected stations, the degree distribution by no means resembles a random network. Having nodes with a small degree and high connectivity, the Hubway network most closely compared to a scale-free network. However, it is important to note that the maximum degree of the Hubway network is much greater than that of the scale-free network (Figure 1).

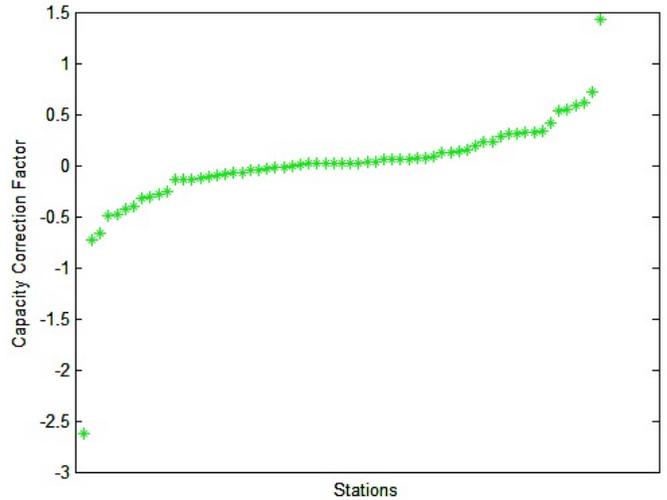


**Figure 1(a): Degree Distribution of Hubway Network (in comparison to scale-free network)**



**Figure 1(b): Power Fit for the Degree Distribution of Hubway and Scale-Free Networks**

In an attempt to better understand the distribution of bikes throughout the Hubway network as users travel from station to station, we examine the difference of the in-coming and out-going strength of each node and divide this by the capacity of the node and the number of days that data consists of. Since every station capacity varies, the number of bikes coming in and out of the station must be represented relative to its capacity. By dividing by the number of weekdays this data contained (184 days), we obtain a capacity correction factor. Stations with a correction capacity factor (CCF) of 0 have an equal balance of in-coming and out-going bikes. The magnitude of the CCF represents the ratio of the predicted capacity increase to the original station capacity. Positive CCF indicates stations that have a capacity shortage for bikes that



**Figure 2: Capacity Correction Factor for Hubway Stations**

are coming into the station. These stations require only an increase in empty racks. Negative CCF indicates stations that require an increase in the number of bikes at the station to accommodate a larger demand of out-going bikes. The capacity increase in these stations requires racks with bikes.

Figure 2 displays the CCF for each Hubway station, sorted from lowest to highest. The CCF for the stations has a mean of 0.0073 and a standard deviation of 0.4750. Our results revealed that 52 out of 63 Hubway stations analyzed are within one standard deviation of the mean; this indicates that most stations are naturally achieving a state of equilibrium. However, some stations have a redistribution problem that results in either a shortage of bikes because of a disproportionately higher number of out-going trips or a shortage of rack capacity because of a disproportionately higher number of in-coming trips. The station with the lowest CCF is at Washington St. at Rutland St. This station has the greatest need for increased rack capacity. Conversely, the station with the highest CCF is at Stuart St. at Charles St. This station has the reverse problem as it has the greatest need for extra bikes.

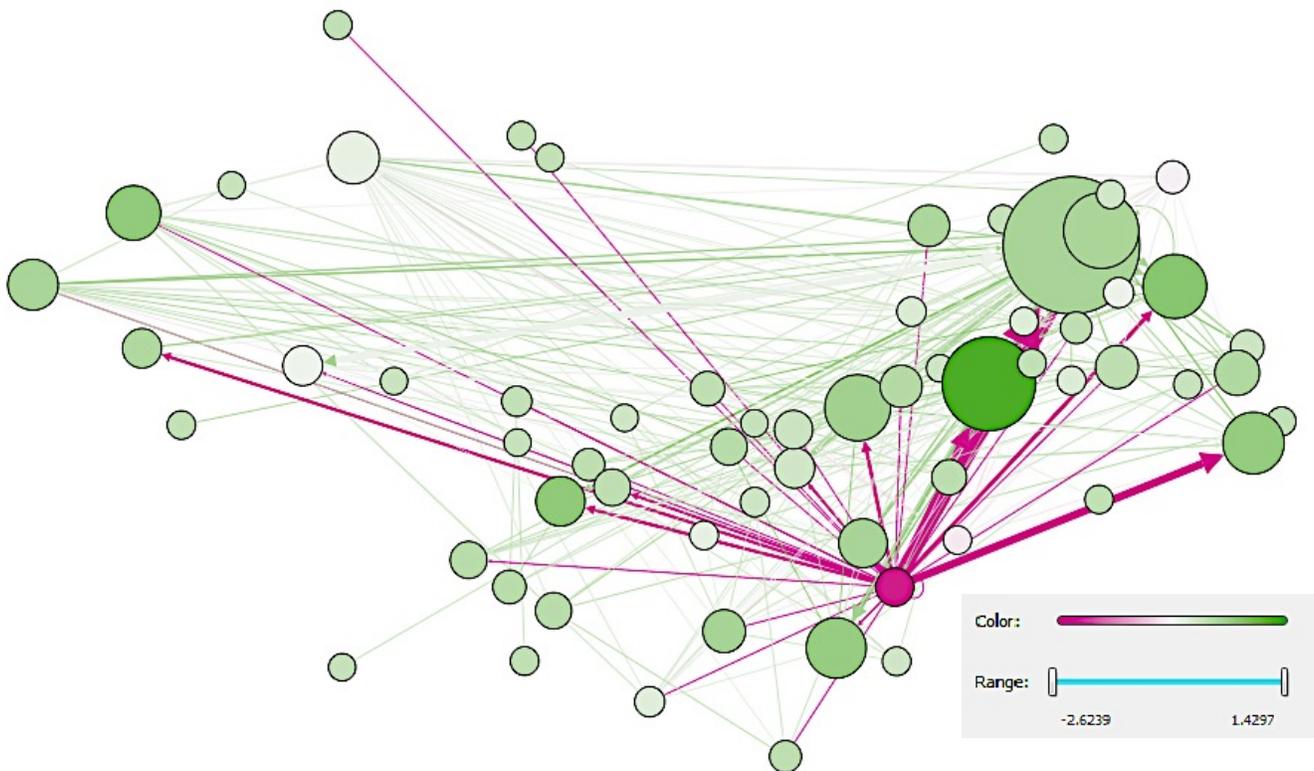
**Table 1: Capacity Correction Factor Distribution**

| CCF Range       | Percentage of Stations |
|-----------------|------------------------|
| [-2.62, -0.468) | 7.94%                  |
| [-0.468, 0.482] | 82.54%                 |
| (0.482, 1.43]   | 9.52%                  |

We show the CCF of each station in the Hubway network in the visualization in Figure 3. As the CCF distribution in Table 1 shows, many stations have a CCF close to the mean. Only a handful of stations are more than a standard deviation away; in fact, only two stations have a magnitude of CCF greater than 1. The visualization in Figure X depicts this using a color gradient. The size gradient of the stations in the visualization shows the relative incoming strength of each station with higher incoming strength stations having a bigger size. It is not surprising that stations

with high CCF (greener in visualization) have high incoming strength (bigger size in visualization). Examining the station with the least CCF, it is apparent its incoming strength is disproportionately higher compared to its CCF.

To better understand how Hubway stations are connected to one another, we examine the assortativity in the Hubway network. For each node, we calculated the average weighted in-degree of its outgoing neighbors ( $\langle k_{nn} \rangle$ ) and compared this to the in-degree of the source node (Figure 4). By examining this relation, we were able to see that nodes with a small degree are connected with other small degree nodes up to a certain threshold, in which nodes with very high degrees are connected with nodes of both low and high degrees.



**Figure 3: Visualization of Hubway Network with sizes of nodes ranked by in-coming strength in descending order (biggest node has highest in-coming strength) and CCF depicted with color gradient.**

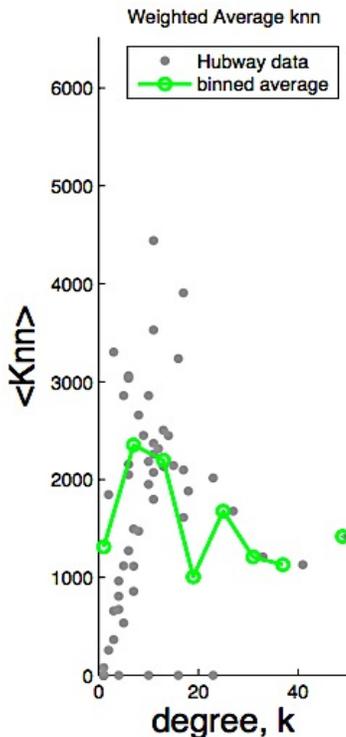


Figure 4: Weighted Average knn vs. k

In the Hubway network, we can interpret the connectivity of the stations by examining the  $\langle k_{nn} \rangle$  vs.  $k$  relationship. Stations that are less connected tend to be connected with similarly traveled stations. This can be explained by the fact that these stations, which are located further away from the “hubs” that exist within Cambridge, Boston, Brookline, and Somerville, typically have users that are not traveling to hubs. However this assortative relation

does not remain true for high degree stations; these highly connected stations, which can be considered “hubs” in the Hubway system, receive trips from stations that are both further away from the hubs as well as other hubs.

### A Comparison to Flows in London’s Subway System

We compare the Hubway network with London’s Subway system. Roth et al treat the London subway system as a network by taking the stations as nodes and trips between them as edges, as we do with our treatment of Hubway as a network.

While Hubway in the metro-Boston area is not as large as London’s metropolitan subway system, we find that similar to the London subway system, edge weights in the Hubway network follow a power law distribution (Figure 5) (Roth et al, 2011). Roth et al also confirms heterogeneity in the London subway network by comparing the first and second moments of the edge weight distribution. We compute the first and second

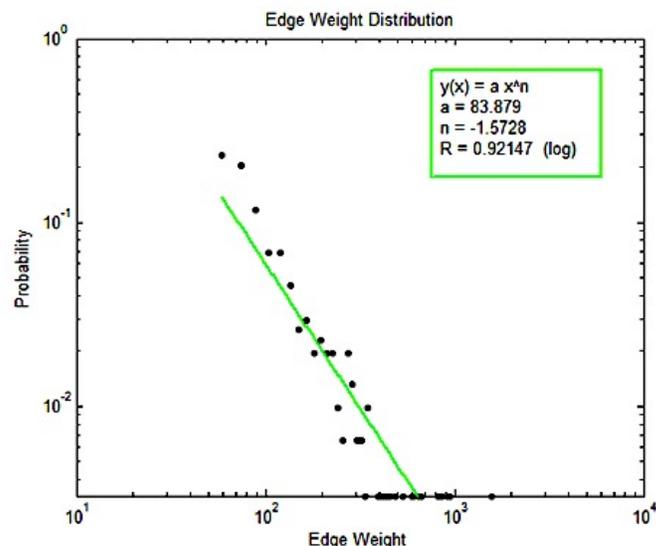


Figure 5: Hubway Network Edge Weight Distribution

moments of the edge weights of the Hubway network to obtain a ratio  $\langle w^2 \rangle / \langle w \rangle^2 = 2.1973$ , implying heterogeneity exists. It is worth noting, however, that the ratio was approximately 15.0 in the London subway network, meaning that network has more heterogeneous edge weights than the Hubway network. This is not surprising, however, since London’s subway network is orders of magnitude larger than the Hubway network.

### Conclusions and Future Work

Based on our analysis of the Hubway system, we make several conclusions. Hubway most closely resembles a scale-free network. However, there are several distinctions due to the high connectivity and small number of nodes in the Hubway network. Based on the analysis of the trips between stations, we have found that Hubway has, for the most part, strategically placed its stations.

Some challenges that were faced when analyzing the Hubway network included the fact that the Hubway network consists of very few nodes. As more stations are constructed, more data will be available to analyze the larger network. Future work that would offer valuable insight and information regarding bike-sharing systems include analyzing other existing and future

systems. Paris's large-scale public bike sharing system, Vélib', currently consists of over 1,200 stations. Analyzing this larger network may provide more valuable insight into the characteristics of a bike-sharing network.

In this paper we examined Hubway trips in the time period 6-9am. Future work could extend this to look at Hubway across different or longer time periods.

A network-based rigorous economic analysis of Hubway can be performed by examining the evolution of assortativity in the Hubway station network. Our hypothesis is there was assortativity when Hubway first started out, but then new stations were added strategically to decrease assortativity in order to avoid creating too many hubs that would detract from traffic at other stations. Hubway is a business after all and does not want to compete with itself. With more time and data several other economic questions can arise that can be answered about Hubway using a network based approach.

## References

Roth C, Kang SM, Batty M, Barthélemy M (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. PLoS ONE 6(1): e15923. doi:10.1371/journal.pone.0015923