Undergraduate Independent Research Proposal

Predicting Bus Occupancy

Daniel Bordak, Erin Corrado, Revan Sopher, and Ashley Weaver

January 31, 2014

Motivation

Due to the periodic nature of class schedules, students at Rutgers tend to move between campuses at similar times during the day. This leads to predictable migratory patterns. However, the bus schedule does a poor job of taking this into account: on Livingston campus, for example, the College Ave/Livingston buses circulate consecutively yet are nearly empty, while the Busch/Livingston buses are less frequent and often packed to capacity. The aim of this project is to develop a system for better predicting demand for buses, such that an empirical argument for the optimization of the bus schedule might be made.

Approach

In order to facilitate simultaneous contributions from all four team members, we approach the problem in two manners: by collecting occupancy data ourselves, and by using transit times. Two students will work on each approach; if both these approaches prove fruitful, we will use a combination of the two to provide our final prediction.

Promiscuous WiFi

In this approach, we use the inherent correlation between amount of smart phones on the bus and amount of people on the bus. We would assemble a tiny computer system that we carry on the bus, programmed to use promiscuous WiFi to actively search for other WiFi enabled devices in the area. The computer will also log GPS coordinates, which combined with current time will provide a comprehensive overview of when and where the Rutgers buses are most used. We would log this data, repeating the experiment throughout several days.

Stop Length Inference

In this approach, we will use the logical correlation between the amount of time a bus spends at a stop and the amount of people riding: the idea is that as a bus nears capacity, stops will be longer due to the larger volume of riders entering and exiting. Every Rutgers bus periodically updates an internet source with its position, velocity, and destination. We would first have to measure the strength of this correlation; if we can prove a strong positive correlation, then we will use the publicly available bus travel data to measure length of stops, and therefore occupancy. Although this method inherently depends on a weaker correlation than the number of smartphones, by running the analysis on a log of travel data we could predict occupancy across a significantly larger domain.