Analyzing Movement Patterns Across Brown's Campus

Wesley Herts | Colby Tresness | Grant Gustafson | Cody Yu

Classifiers:

- Scikit-Learn Classes Used: BernoulliNB, SVC, SGDClassifier, DecisionTreeClassifier, and KNeighborsClassifier
- Graphs represent mean cross-validation accuracies withholding corresponding locations

Class Day vs Non-Class Day:

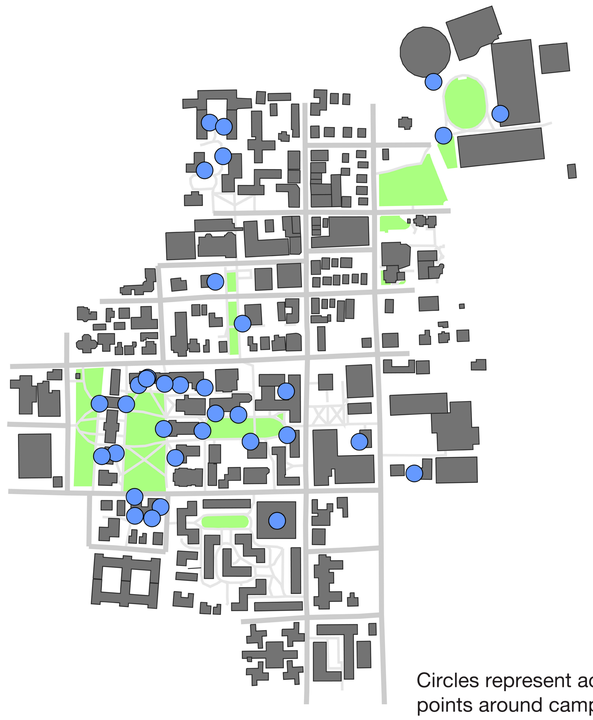
Precipitation vs No Precipitation:

Results:

- Pembroke and Mac-BH are most important for classifying a class day
- Pembroke and the Main Green are most important for indicating precipitation

Questions:

- How do people move?
- Do movement patterns indicate if class is in session?
- Do movement patterns indicate if it is precipitating?
- Which Access Points (APs) best predict behavior?



Circles represent access points around campus

Dataset:

- 3 million connection records from 49 Access Points in public areas
- Data is between Nov 1, 2015 and March 10, 2016
- Records include access point, connection & disconnection time, and unique device identifier
- Weather information: atmospheric conditions from wunderground.com
- Brown Academic Calendar: categorizing days into regular semester, reading period, finals, or recess

Challenges:

- Scale of data: Computations often expand data by orders of magnitudes
- Had to run computations on Brown's BSN cluster
- Framework to visualize campus: had to scrape data (from OpenStreetMap), parse, prune unnecessary information, and export in a format easily renderable by D3

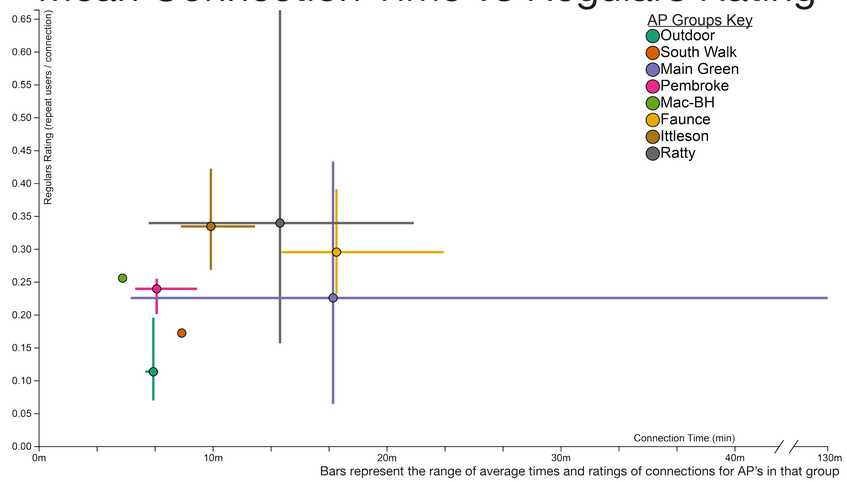
Tools and Technologies:

OpenStreetMap, python, {JSON}, D3, CSV, scikit learn, JS, Spark, WU WEATHER UNDERGROUND

Regulars Rating:

- The "regulars rating" expresses the number of repeated users per connection area on campus
- Regulars Rating = $\frac{\text{Total Connections - Distinct Users}}{\text{Total Connections}}$

Mean Connection Time vs Regulars Rating



- This graph shows the positive correlation between average connection time and regulars rating at a location
- People who consistently return to the same location tend to stay longer

Movement Prediction:

- Input: a list of a person's previously visited APs
- Creates edges by chaining every combination of input APs to every other AP on campus
- Calculates weighted frequency based on how often paths appear in the dataset, weighting more recent and direct paths heavier