**TA**

Below are some comments for your project proposal submission. We have provided some questions or topics for you to think about in relation to your project. If we asked you to talk to us, please talk to one of the course staff in person.

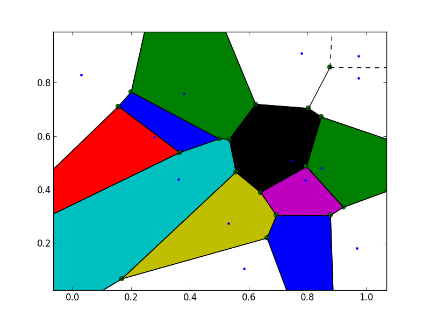
Comments:

This project is very interesting. You may want to look into maptools, OpenStreetMaps, etc. to plot the points on a map. If you’d like to compare bike time to other modes of transportation, do you have a plan for getting data about other forms of transportation? For example, maybe using Voronoi diagrams around bike stations with walking distances calculated by the Google maps API? We’d recommend getting started with looking into all the options soon as working with APIs can be non-trivial.

**PROFESSORS**

Assume: everyone will walk to the closest bike station. Is there a hard cutoff for how far people would walk?

Voronoi Diagram – use this to draw the boundaries for which bike station people will walk to



Maybe each person is not necessarily associated to just one, but the closest two stations.

Start with Clear Questions:

Plan out the questions. Distribute the questions among group members.

We can have two threads (two big questions)

We can start with visualization, clustering, analysis.

Neighborhoods.

One analysis is a broad scale analysis… And then a more specific in depth analysis for a specific route.

Any sort of public transportation routes adjustment recommendation should be made at the end as a policy/interpretation.

1. Network of common routes, with weights of how many people go between them. Find communities in the network, where clustering defined where there is a lot of traffic between the two. Perhaps there are some that are not well connected.
2. Cluster them by types of usage (short trip/long trip) … bike stations that are used primarily for work-residence transport.

Clusters based on neighborhood characteristics.

Analyze the same question but with different methods to see how sensitive the results are to your assumptions.

Measuring uncertainty typically with variance, but we could look at using Entropy. We would have to estimate all of these because we have empirical data.

It would be nice to meaningfully combine Two different data sets (eg. Transport time between stations given in blue bike data and population densities)