**Tyler Baker and Rathish Parayil Sasidharan**

**Data 620**

**Final Project**

Your proposal should describe at a high level what you’re seeking to accomplish, and your motivation for performing this analysis.  A guiding question or hypothesis to test is one good way to start.  If you are going to work in a small group (encouraged!), you should also list your partners’ names.

You should briefly describe your data sources, plan for doing the work, and up front concerns.  If you are working in a group, please describe the roles and responsibilities of each group member.

We’ll treat this proposal as a planning document, not a blue print containing “firm, fixed requirements.”

**Goal**

For the Final Project, we will be using sentiment analysis and network analysis to examine the YELP review data.

This project has 2 parts.

**Part 1:**

We will build a bipartite network using the restaurant review data and project a user network do some centrality measures to identify the prominent reviewers.

We will be using island method of network reduction since our initial analysis shows that most of the reviewers are having degree < 5.

**Part 2:**

We will analyze the restaurant review data and perform rating prediction using classification.

**Libraries:**

**We will be primarily using NLTK, pandas, networkx, and matplotlib to accomplish this project.**

**• Requests**

**• Pandas**

**• Networkx**

**• Maplotlib's Pyplot**

**• Numpy**

**• NLTK**

**• Re (regular expressions)**

**Data source:**

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DMWCBT

business\_id : Restaurant id

user\_id : Reviewer id

text: Review text

Rating: 0 - 5

Useful: Number of people marked this review as useful