CS432/532 Project 1-Page Proposal

1. Team members (in alphabetical order of the last name):

Tyler Gabriel, Jesse White

2. What problem you want to work on?

The problem we wish to work on is determining the similarity of two Twitter users based on shared likes, retweets, followers, users followed, and the similarity between tweet syntax. This will be represented as a score based on the weighted values of the previous categories.

3. Which NoSQL database you want to use?

We will use MongoDB for this project.

4. What data will you use (if you use a public data set)? If you want to generate your own data briefly explain why it is the best way. For example, you can say there is no existing public data set (make sure it's true).

While there are datasets for Twitter such as *awesome-twitter-data*, for the scope of this application we think it is best to gather and store our own data using a Python program. We believe that this will be the most up to date and relevant way for users to compare similarity.

5. Why is it an important/interesting problem?

This is an interesting problem because it can allow people to connect with other users similar to them. It can also allow for people to be grouped based on how similar they are. This can be used to gain insight on subsets of the Twitter community.

6. Why is it significant (non-trivial)?

This problem is non-trivial because it requires gathering data, storing data, performing analysis on data (such as syntax similarity), and finally, calculating similarity. Additionally, there will be a command line interface used to access the database as an end user.

7. Why is it doable in this semester? (Describe your plan, strength, familiarity with the language/tool of your choice, etc.)

This project is doable this semester because both partners have experience with Python and the Twitter API. These will be used to handle gathering and accessing data. We believe MongoDB will allow us to store our data efficiently as well.