TwitterProject Methods and Documentation:

Data Collection:

The source of all data was acquired through the tweepy library using the Twitter public api. A custom algorithm was implemented to extract the retweet data of selected tweets from both candidates, Jensen and Walz. Key data points included the username and tweet Id. The api call limit was 100 calls per tweet with no practical way to select additional starting points for additional user selection. With this in mind, the sample size was limited to 100 or less with a no floor. Since the retweet lists used for the api calls are unsorted, it can be inferred that the samples extracted are random, based on when the retweet was made. Ten tweets were selected between August and November, election season, between both candidates. Tweet selection was primarily based on number of total retweets so as to attain the largest sample possible, up to the 100 call limit.

With the samples taken, a unified list of all samples was generated and cleaned. This list of usernames was used to generate as “bot score” from Bot’o’meter, formerly bot or not from the University of Illinois.

An algorithm was developed that utilized the Bot’o’meter api to generated bot scores and probability scores for each user name. The api call limit was 400 calls per day. This created the need generate multiple user-score files that required post-generation analysis and cleaning.