

CS 432: Homework Number Four

Due on March 9, 2019 at 4:20 PM

Alexander Nwala

Tim Bruce

Problem 1

Determine if the friendship paradox holds for my Facebook account.* Compute the mean, standard deviation, and median of the number of friends that my friends have. Create a graph of the number of friends (y-axis) and the friends (x-axis) themselves, sorted by number of friends (y-axis). (The friends don't need to be labeled on the x-axis: just f1, f2, f3, ... fn.) Do include me in the graph and label me accordingly.

* = This used to be more interesting when you could more easily download your friend's friends data from Facebook. Facebook now requires each friend to approve this operation, effectively making it impossible.

I will upload a csv file of my 2014 friends list on the #assignment-4 slack channel

Solution

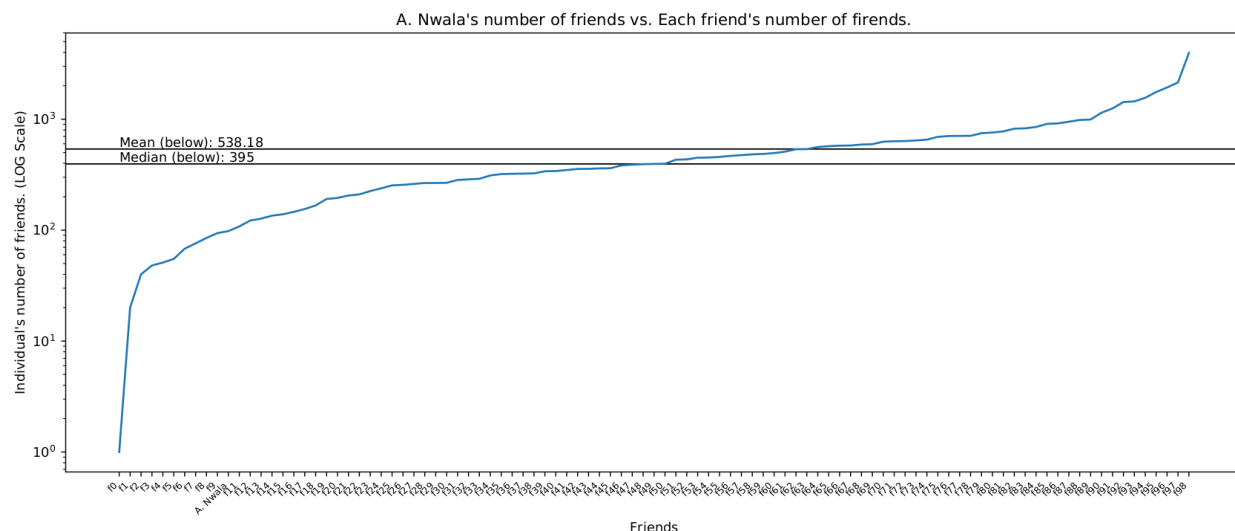
The CSV was downloaded from the assignment-4 channel in slack. It has two columns 'user' and 'friend-count'. The first line is skipped and then all subsequent lines are split at the ", ". Finally, all of the friends counts are read into code. Right before being sent to the graphing function, the length of the data is added to the data set as A. Nwala's number of friends.

To generate the mean, median, and standard deviation, I used a library called statistics, which can take a list of numeric types and output statistics data about it. Looking ahead to problem two (spoilers: It's similar to question one) I made this part into its own function. The function calculates and prints the data, and then calls the graphing function.

In the graphing function, the program sorts all of the data by number of friends, and then plots them on a line graph with a logarithmic scale. This question is fairly simple, so no code has been shown up to this point. I would however like to highlight a fancy thing I did on the matplotlib chart.

Using the line function and the inline text, I was able to label the mean and median on the chart. The final chart is shown below.

Image 1: The chart produced using the Python Matplotlib program.



Problem 2

Determine if the friendship paradox holds for your Twitter account. Since Twitter is a directed graph, use "followers" as value you measure (i.e., "do your followers have more followers than you?").

Generate the same graph as in question #1, and calculate the same mean, standard deviation, and median values.

If you do not have followers on Twitter (or don't have more than 50), then use my twitter account "acnwala".

Solution

Apparently I don't have enough followers on Twitter for you... or at least for a useful data set. That is okay because we can use the provided Twitter handle. I got the data from Twitter utilizing a bit of code that I found on stackoverflow. I basically made a function that takes the username as an input and completes the question. I ran into significant difficulties dealing with the Twitter API on account of its tendency towards difficult to understand policies. My solution was extreme caution:

Listing 1: Cautious Twitter API Access.

```

1 for user in followers:           # (above) For each user, get followers
2     ff = []                       # FF stands for "follower's followers"
3     for page in tweepy.Cursor(api.followers_ids, user_id=user).pages():
4         ff.extend(page)          # (above) actually gets the followers.
5         if len(page) > 5000:      # If it is a long page,
6             time.sleep(60)        # Treat it as a new call.
7     data.append(len(ff))
8     print(data[-1])
9     time.sleep(60)               # Next user.

```

By waiting for 60 seconds for each call, I did not exhaust my uses of the API. But, just in case, I also found a useful bit of code from the tweepy documentation.

Listing 2: API Waiting Code.

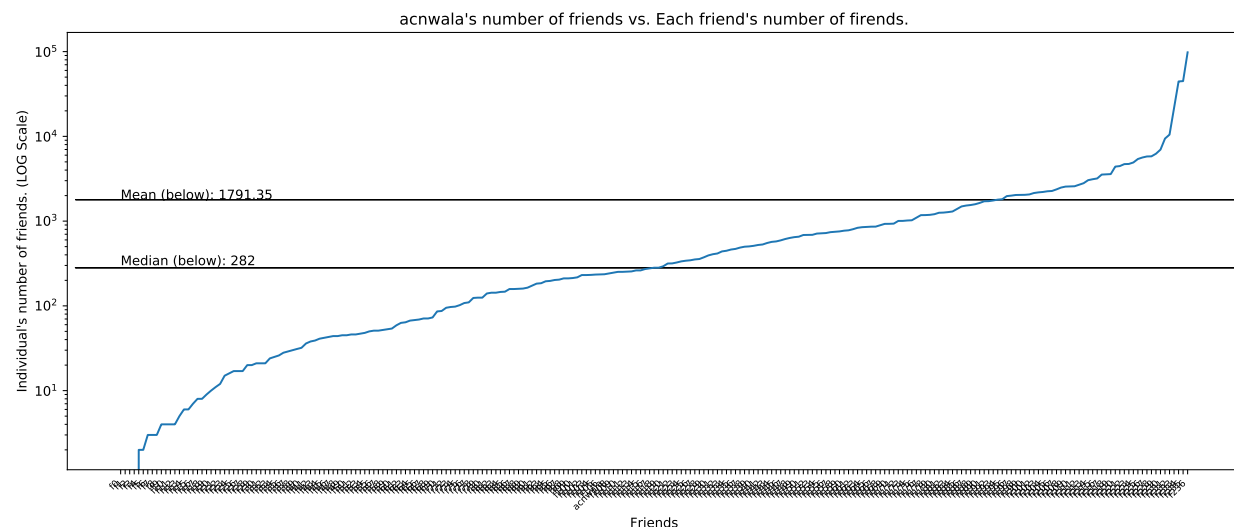
```

1 api = tweepy.API(auth, wait_on_rate_limit=True)

```

This "wait_on_rate_limit=True" makes the tweepy API wait when there is an api limiter placed on your account. It prevents overly long waiting, and ripping out of one's hair. Re-using the graphing code from the last question, I got the following result for Alexander Nwala's Twitter account.

Image 2: The data from user acnwala's Twitter Account.



Looking at the data collected. The friendship paradox definitely holds true for the average friend or follower, as A. Nwala's followers have on average more followers than he does by a factor of ten, but it also appears to hold true for the median friend or follower given these admittedly limited data sets.