

OLD DOMINION UNIVERSITY

Assignment 4

David Bayard

March 16, 2019

QUESTION 1.

Determine if the friendship paradox holds for my Facebook account.

Solution:

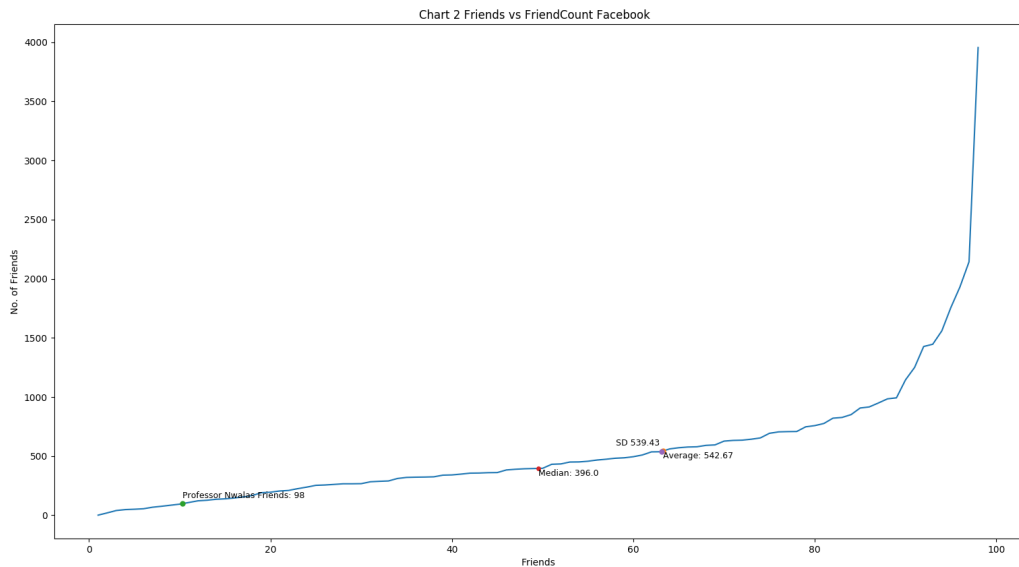
In order to determine if the friendship paradox is enforced, the mean, median, and standard deviation of professor Nwalas friends was calculated. This was done by reading the .csv file provided by professor Nwala and using built in python functions to calculate each data point, for example, the mean was calculated as shown:

```
1 average = statistics.mean(listValues)
2 roundedAverage = round(average,2)
```

Listing 1: Calculating Mean

The code above displays using the Statistics library to calculate the mean, the same was done for the median and standard deviation. A simple counter was used to keep track of the number of friends that professor Nwala has. Once the mean, median, and friend count were calculated, it was evident that the friendship paradox does, in fact, hold.

It is evident that the friendship paradox holds because professor Nwalas friend count is significantly smaller than the mean friend count. This means that professor Nwala has less friends than his friends have, on average. This point is illustrated on the graph below, which displays the distance between professor Nwalas friend count and the average friend count.



(a) Facebook Graph

QUESTION 2

Determine if the friendship paradox holds for your Twitter account. Since Twitter is a directed graph, use "followers" as value you measure

Solution:

In order to determine if the friendship paradox holds for a Twitter account, the number of followers for each follower had to be captured. In this case, professor Nwala's Twitter handle "acnwala" was used, and the number of followers for each of his followers was extracted using the Tweepy API. The Tweepy API provides a method to search for followers, which returns User objects, providing each follower for that user, and their number of followers. The usage of this API is demonstrated below.

```
1 try:
2     for user in tweepy.Cursor(api.followers, screen_name="acnwala").items():
3         followers.append(user)
4         counter += 1
```

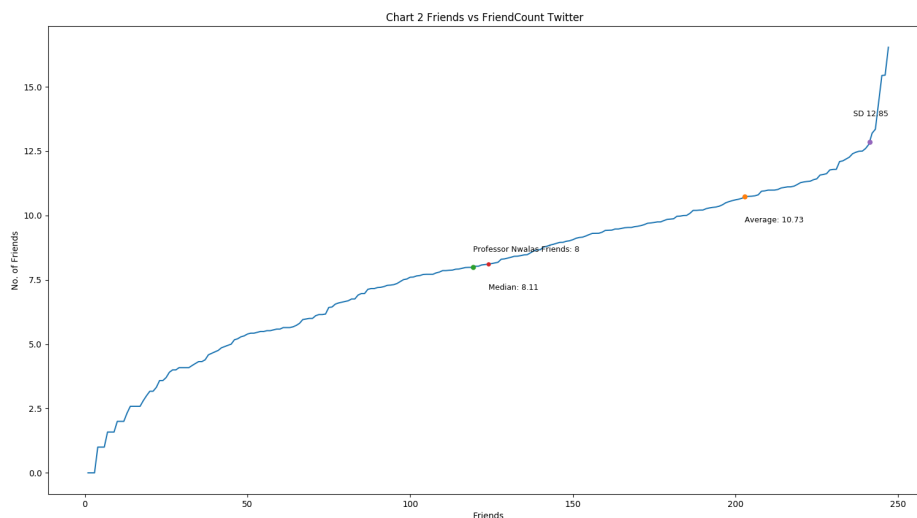
Listing 2: Getting User Objects

In the above code, the Tweepy API is used to extract User objects, and these objects are stored in the followers list. The counter variable keeps track of the total number of followers that professor Nwala has. Next, it is necessary to extract the number of followers that each follower has, and this is demonstrated below.

```
1 for follower in followers:
2     writer.writerow({'User': follower.name, 'FriendCount': follower.followers_count})
```

Listing 3: Extracting Follower Count

As demonstrated above, each User object provides methods that can be used to extract the name of each follower, along with their follower count. With this information, we can calculate the mean, median, and standard deviation of each of the followers just as was done from question 1. The graph below illustrates the information that was gathered from Twitter.



(a) Twitter Graph

The figure in page 2 was created using a logarithmic function with a base of 2, for each of the y values. This was done because there were several outliers, ranging from 0 to 90000 friends. This severely skewed the graph, and using a logarithmic function reduced this skew and demonstrates an accurate representation of the data.

The values for the mean, median, standard deviation, and personal count before the logarithmic function was supplied are listed below.

Statistics Before Logarithmic Function

Calculations			
Mean	Median	Standard Deviation	Personal Friend Count
1702.68	276	7404.11	247

From the information supplied by the table above, and from the graph on page 2, it is evident that the friendship paradox is upheld. This is depicted by a much higher average friend count, when compared to the personal friend count. This information basically dictates that on average, each friend of professor Nwala, has more friends (or followers) than professor Nwala has.

QUESTION 4. EXTRA CREDIT

Repeat question 2, but change "followers" to "following"? In other words, are the people I am following following more people?

Solution:

Using the Tweepy API, a list of professor Nwala's followers was obtained, and then each of these friends had their total number of friends extracted. The extraction of professor Nwala's friends is shown below:

```
1 # Returns an array of integers for each friend's id
2 following = tweepy.Cursor(api.friends_ids, screen_name="acnwala").items()
```

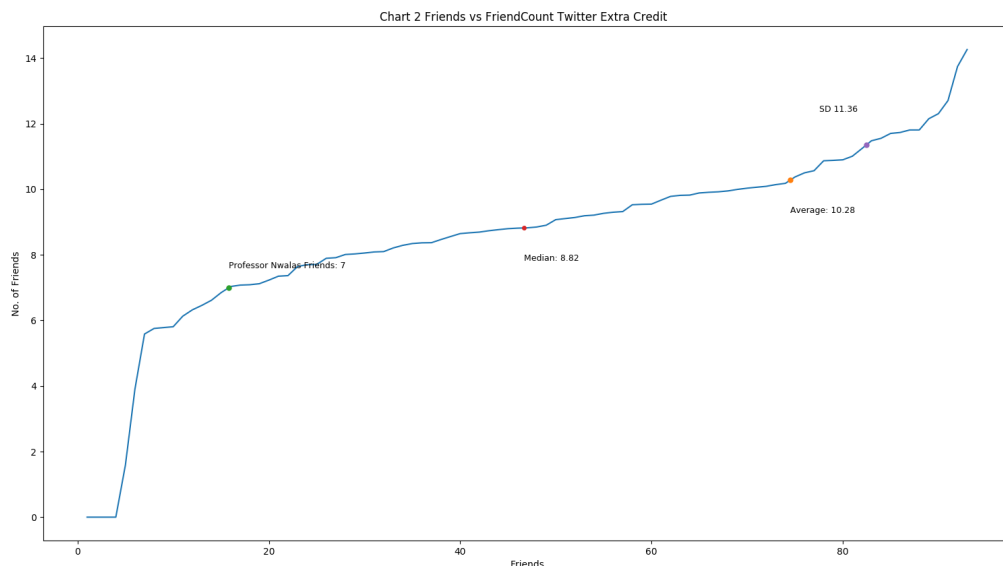
Listing 4: Extracting Friend Count

After running the code above, an array of ids exists, and using this array, it is possible to get the friend count for each user, whom is located based on their id.

```
1 user = api.get_user(user_id=item)
2 formatFollowing = {'Friend_Name' : user.screen_name, 'Number_Following' : user.
    friends_count}
```

Listing 5: Extracting Friend Count

The code above provides a user object for each user id, in this case held in the variable item, and these objects provide a screen name and friend count for later manipulation. The graph below displays the data after the mean, median, and standard deviation have been calculated. The logarithmic function has been applied to the data before graphing. To see the original data, visit "FollowerRecordsExtraCredit" csv file.



(a) Extra Credit Graph

It is evident that the friendship paradox holds in this situation as well. This is true because professor Nwala once again stands a substantial amount below average, with regards to the friend count. This dictates that most of professor Nwala's friends have more friends than he does, and thus the friendship paradox holds.