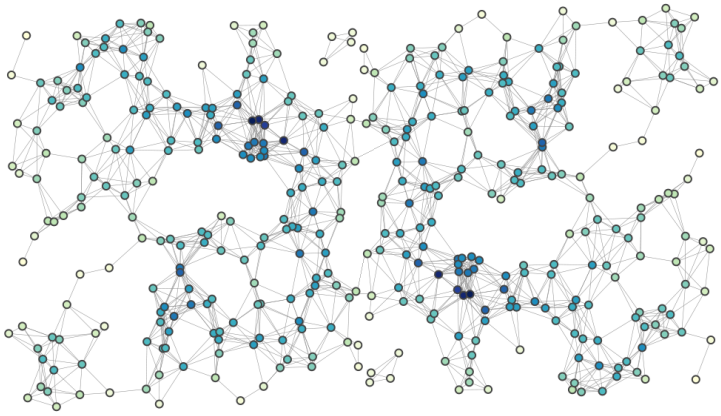


Old Dominion University

CS 432 Web Science

Assignment Four

Derek Goddeau

Professor

Michael L. Nelson

March 3, 2017

1 Facebook Friendship Paradox

In order to represent the data from the GraphML file I used the R `igraph` library. After reading it from the file I created a dataframe with all of the vertex attributes from which I subset only the needed data, names and friend counts. The `na.omit()` call removes 11 friends with no friend data from the dataset.

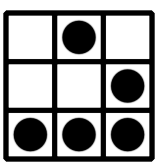
```
library(igraph)
g <- read_graph('../data/mln.graphml', format = c('graphml'))
df <- vertex_attr(g)


df.self <- data.frame(friend_count = 234, name = 'Michael L. Nelson')
df.friends <- data.frame(friend_count = df$friend_count, name = df$name)
df.all <- rbind(df.friends, df.self)
df.sorted <- na.omit(df.all[with(df.all, order(friend_count)), ])
```

The R `summary()` and `sd()` methods were used to calculate the mean, standard deviation, and median values.

Mean	Median	Standard Deviation
359	267	372

Table 1: Facebook Friend Statistics

As shown in figure 1 on page 2 the friendship paradox holds with both the median and mean. There are a small amount of outliers that skew the data slightly but not enough to question the outcome.



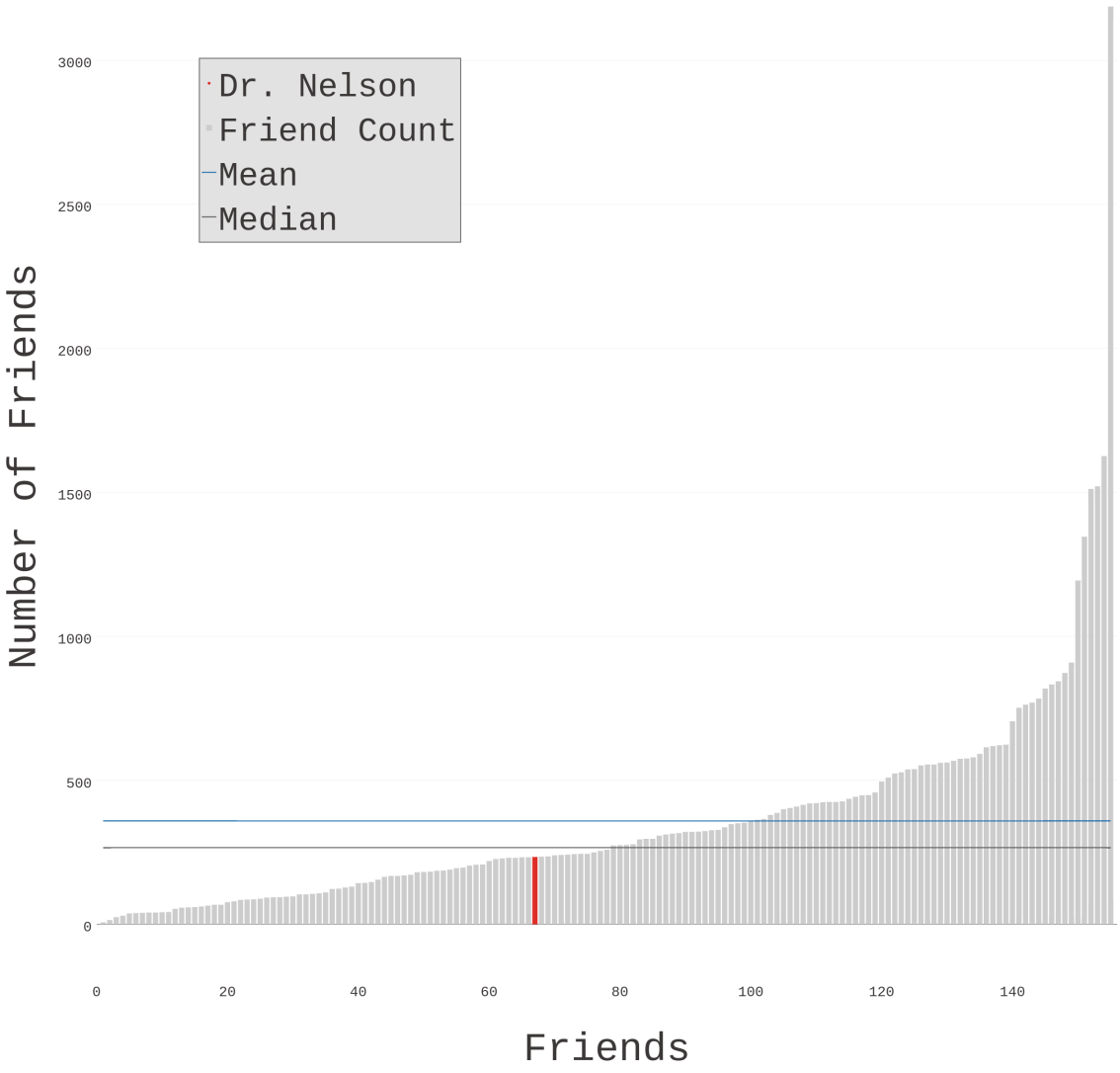


Figure 1: Dr. Nelson's Facebook Friends

Derek Goddeau (Datenstrom)
Old Dominion University

2 Twitter Friendship Paradox

To get the twitter followers I used Python but this time `python-twitter` instead of `tweepy`. This allowed for much simpler fetching of the followers as shown below it is reduced to a one liner after authenticating, but provides no way to get the data out of their `User` object, so I wrote just the data I needed to a CSV file for R to read in.

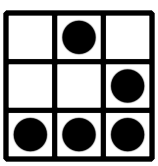
```
followers = self.api.GetFollowers(self.username)
```

From there it is the same exact process as with the Facebook data in R. Again I used `summary()` and `sd()` to get the statistics.

Mean	Median	Standard Deviation
1508	310	10143

Table 2: Twitter Follower Statistics

With this data, as shown in figure 2 on page 4 there are some extreme outliers which will skew the average and in fact the average is quite a bit larger than the mean. In figure 2 on page 5 Dr. Nelson's follower count is well above the median but also well below the average. In this case by the mean value the friendship paradox holds, but the median is a better value to judge the quality and therefore the friendship paradox does not hold. With clever use of statistics and graphs either argument could be made about the model.



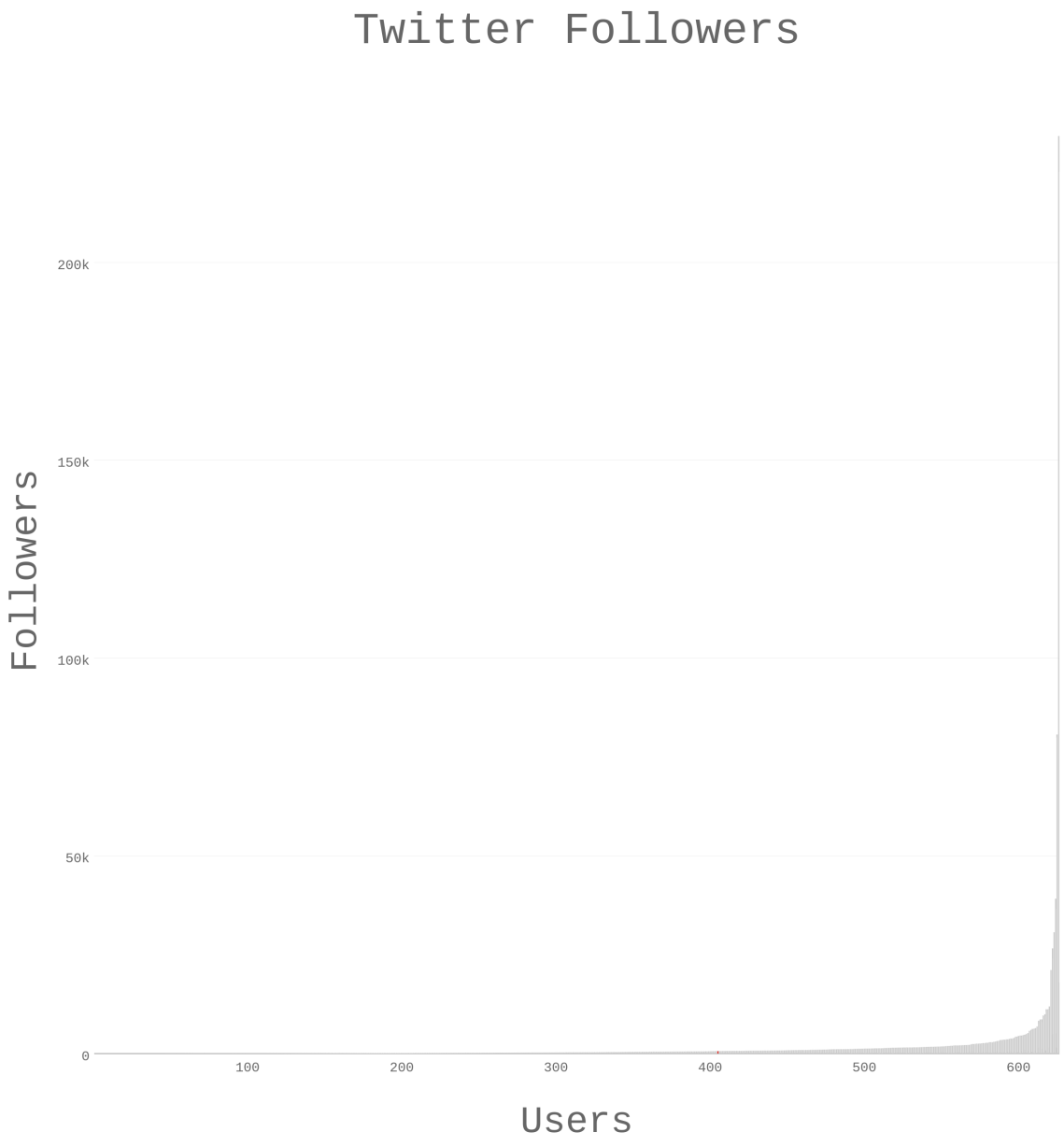


Figure 2: Dr. Nelson's twitter followers

Derek Goddeau (Datenstrom)
Old Dominion University

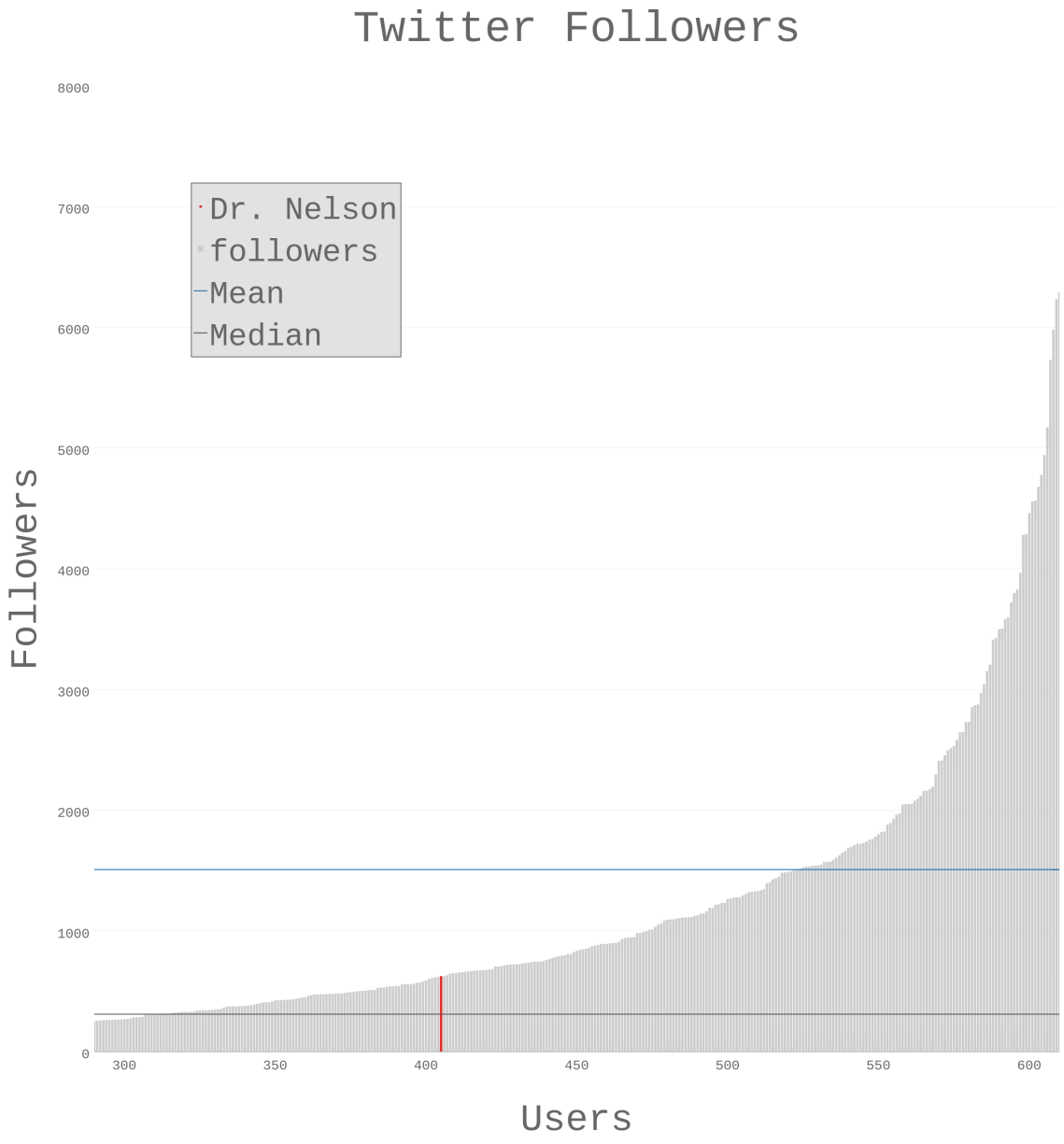


Figure 3: Dr. Nelson's twitter followers

Derek Goddeau (Datenstrom)
Old Dominion University