

CS432 Web Science: Assignment 4

Finished on April 28, 2020

Dr. Michele C. Weigle

Hannah Holloway
hhollowa@odu.edu

Contents

Problem 1	2
Question	2
Answer	2
Problem 2	4
Question	4
Answer	4
Q3 - Extra Credit	6
Question	6
Answer	6

Listings

1	twitterPlotter.r	4
2	R script to generate Twitter Paradox Graph	8
3	R script to calculate mean, median, and standard deviation	8
4	Script to get my friends' friends count	8

List of Figures

1	Facebook Friendship Paradox	3
2	Twitter Friendship Paradox	5
3	My Twitter friends and their friends count	7

Problem 1

Question

The friendship paradox says that your friends have more friends than you do. Determine if the friendship paradox holds for a user's Facebook account. (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.)

HW4-friend-count.csv contains a user's friends' names and number of friends they each have.

Compute the mean, standard deviation, and median of the number of friends that the user's 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.) Include the user in the graph (count the number of their friends) and label as U.

Answer

I loaded the csv file into R and plotted the graph. The mean, median and standard deviation were also calculated in R.

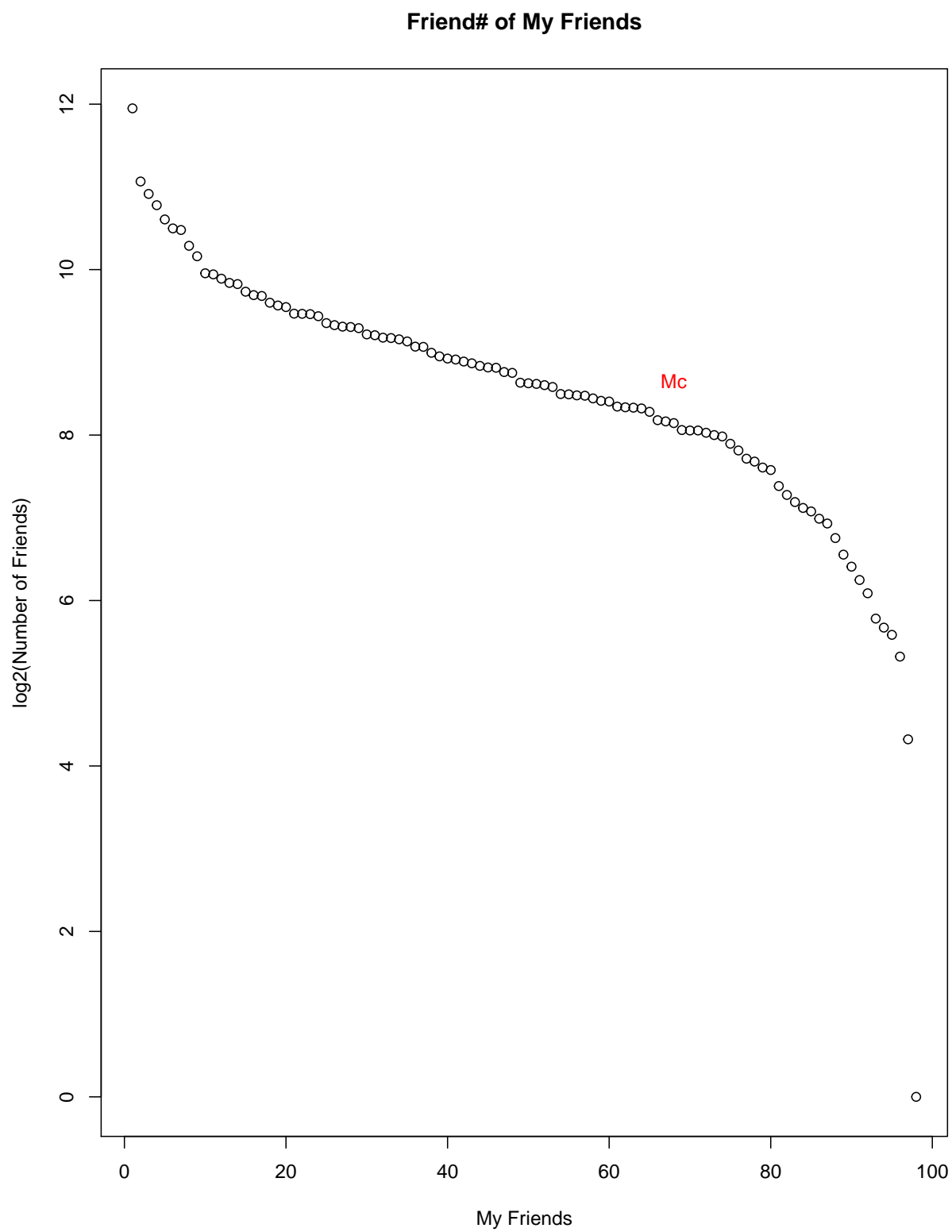


Figure 1: Facebook Friendship Paradox

Problem 2

Question

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

Generate the same graph as in Q1, and calculate the same mean, standard deviation, and median values.

For the Twitter 1.1 API to help gather this data, see:

<https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-list>

If you have less than 50 followers on Twitter, then you can use my Twitter account (weiglemc).

Answer

This approach was similar to the last, except I have to find the Twitter API and use it to get my follower count. I figured out the twitter API limits the number of requests, so I used the follower API to get 200 followers at a time. My Twitter's statistics: mean: 491.559 median: 228 standard deviation: 25609.87

```

1 pdf("twitterplotter.pdf")
  mydata=read.table("twitter.data",sep="\t",header = FALSE)
  num<-nrow(mydata)
  nmean<-mean(mydata$V1)
  nsd<-sd(mydata$V1)
6  nmedian<-median(mydata$V1)
  mlnIdx<-which(mydata$V2 == "Hannah")
  plot(sequence(num),mydata$V1, main = "Plot of number of followers on twitter",
        xlab = "friends name", ylab = "number of friends one have",xaxt='n',pch=16,cex=0.5)
  abline(v=mlnIdx,col="red",lwd=2,lty=2)
11 meanpos<-approx( x = mydata$V1,y = sequence(num), xout = nmean)
  medianpos<-approx( x = mydata$V1,y = sequence(num), xout = nmedian)
  sdpos<-approx( x = mydata$V1,y = sequence(num), xout = nsd)
  axis(1, at=c(1,num),
        lab=c("F1",paste("F",num)))
16 legend(x=.1,y=max(mydata$V1),c("Hannah"),
        col = c("red"),lty=c(2),
        cex=0.5)
dev.off()

```

Listing 1: twitterPlotter.r

The graph shows many of my followers have more followers than me. So the friendship paradox stands true on this case as well. The mean number of followers is more than my number of followers (330), but the median isn't (228).

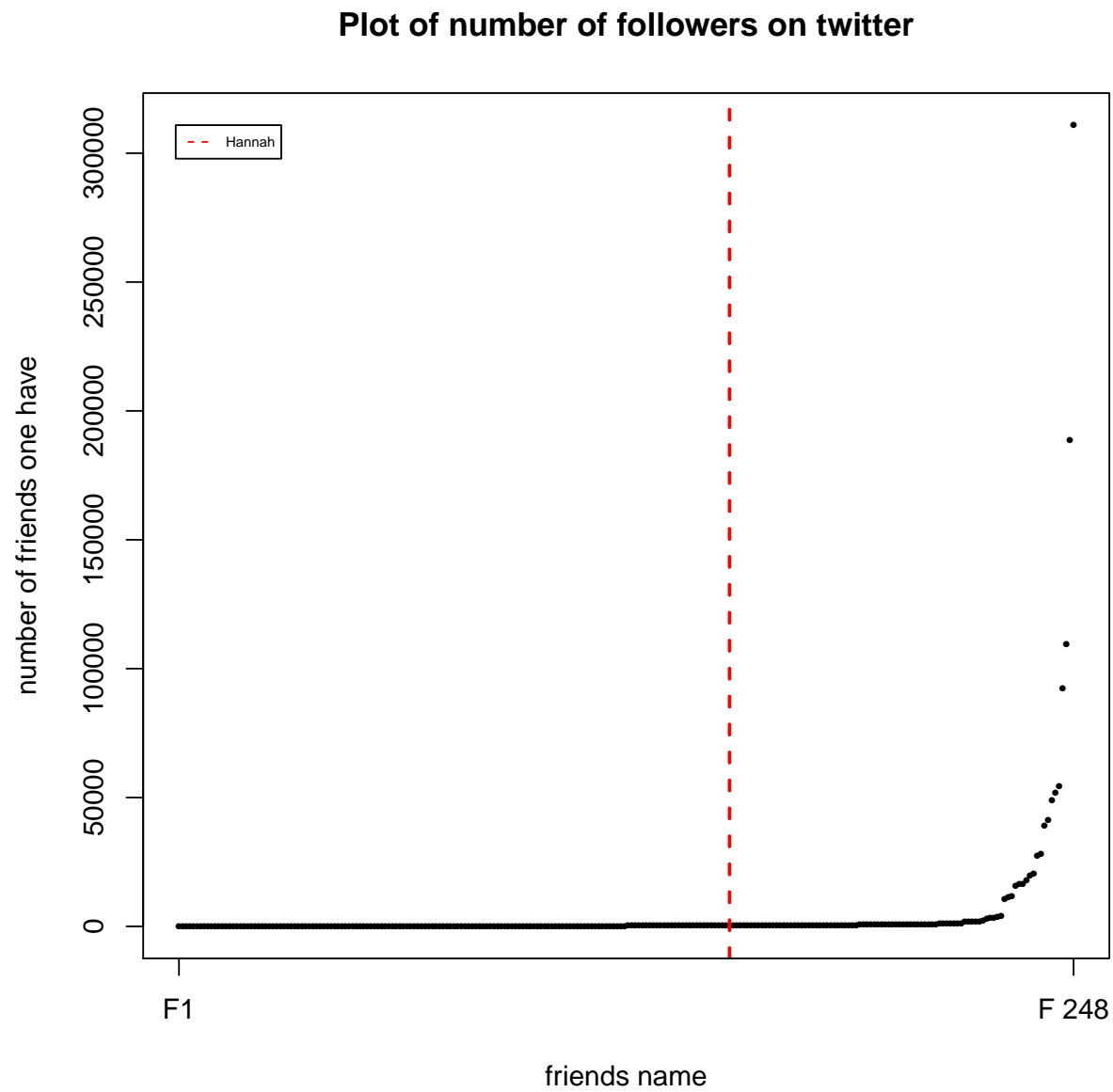


Figure 2: Twitter Friendship Paradox

Q3 - Extra Credit

Question

Repeat Q2, but change followers to following. In other words, are the people you are following following more people than you are?

For the Twitter 1.1 API to help gather this data, see:

<https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-friends-list>

Answer

The script I wrote to solve this task was similar to the script I wrote to solve the second question asked. I just replaced followers with following. I chose my account `_hollowayhannahm` since I have 273 followers. I used the term "friends" to represent the people I follow. The script to achieve this is displayed below. I first outputted a .csv file that shows each of my friends' screen names along with each of their friends count in addition to my twitter screen name and friends count. I outputted another .csv file that I will use for statistic purposes to get all of my friends' screen names with each of their friend counts but leaving out my screen name and friend count. The R script I used to calculate the statistics as well as the R script I used to generate the graph are displayed below.

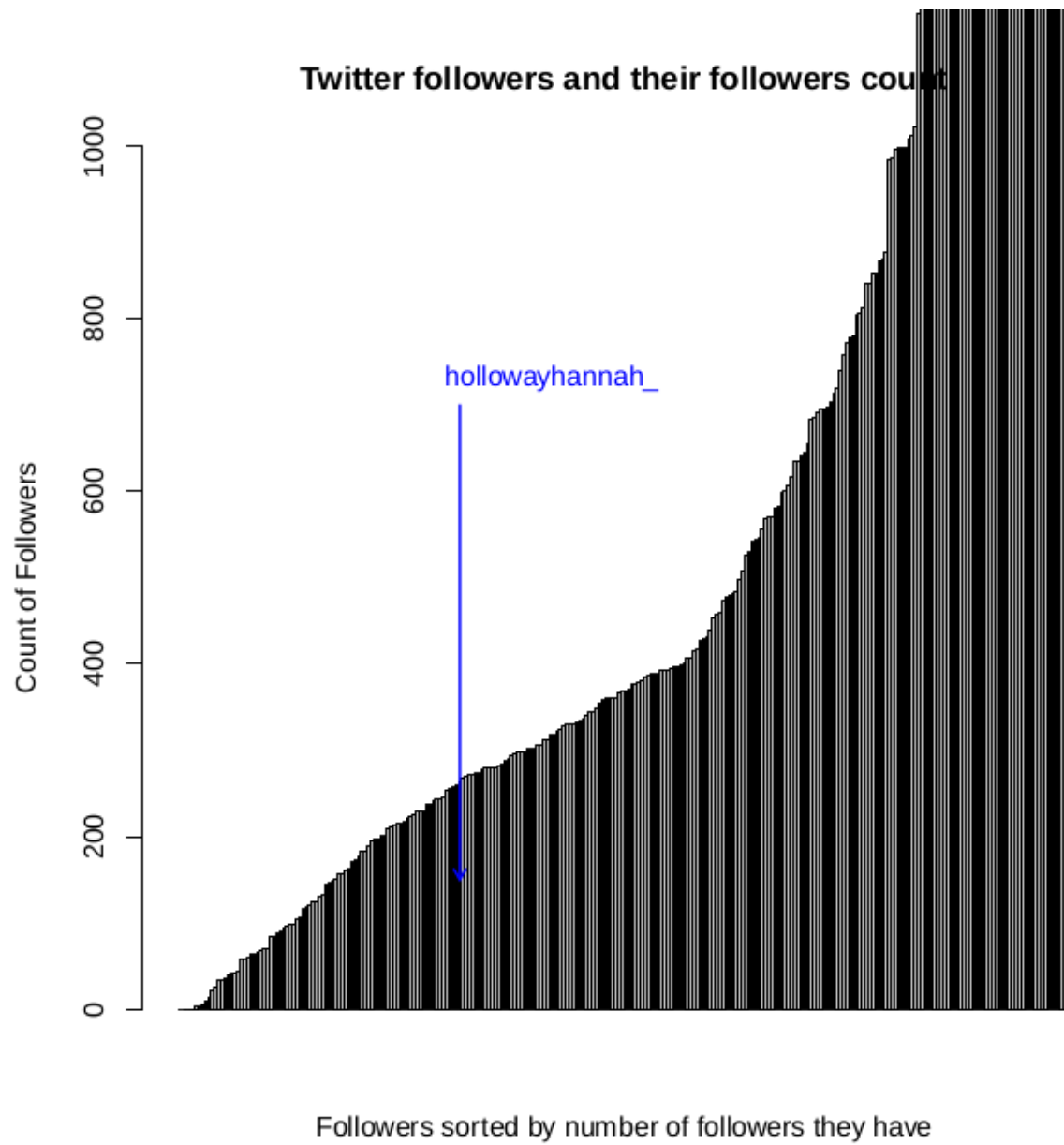


Figure 3: My Twitter friends and their friends count

The mean I calculated is 2953.198 and the the number of people I follow are 273. So, my answer to the question "Are the people I am following, follow more people?" is yes, since the number of users I follow is less than the mean. This proved the Friendship Paradox using my Twitter account's following list.

```

1 tmp <- read.csv("following_count.csv")
  following_count <- tmp[,2]
  barplot(following_count,main="Twitter followers and their followers count",xlab="Followers
    sorted by number of followers they have",ylab="Count of Followers",ylim=c(0,1000))
  arrows(x0=match(c(273),following_count)+12, y0=700, x1=match(c(273), following_count)+12, y1
    =150, length=0.07, lwd=1.5, col='blue')
  text(x=match(c(273), following_count),pos=4, y=730, labels="hollowayhannah-", col='blue')

```

Listing 2: R script to generate Twitter Paradox Graph

```

tmp <- read.csv("following_count_without_me.csv")
following_count <- tmp[,2]
> write(mean(following_count),stdout())
2953.198
5 > write(median(following_count),stdout())
369
> write(sd(following_count),stdout())
14123.16

```

Listing 3: R script to calculate mean, median, and standard deviation

```

import tweepy
2 import time
import csv

csv_writer = csv.writer(open("following_count.csv", "wb"))
csv_writer1 = csv.writer(open("following_count_without_me.csv", "wb"))
7 csv_writer.writerow(['Name','num_following'])
  csv_writer1.writerow(['Name','num_following'])
  ACCESS_TOKEN = '' # Variables that contains the user credentials to access Twitter API
  ACCESS_SECRET = ''
  CONSUMER_KEY = ''
12 CONSUMER_SECRET = ''
  count = 0

  auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET) #Authentication is handled by
    the tweepy.AuthHandler class
  auth.set_access_token(ACCESS_TOKEN, ACCESS_SECRET)
17

  api = tweepy.API(auth) # Construct the API instance

22 for user in tweepy.Cursor(api.friends, screen_name="hollowayhannah-").items():
  csv_writer.writerow([user.screen_name,int(user.friends_count)])
  csv_writer1.writerow([user.screen_name,int(user.friends_count)])
  count=count +1 # to count number of friends of mln
  csv_writer.writerow(['hollowayhannah-',count])

```

Listing 4: Script to get my friends' friends count

References

- [1] Getting Started with Twitter API: <http://pr.eyedomain.com/>