CS 532: Assignment 4

Dinesh Kumar Paladhi

Spring 2016

Contents

1 Problem 1

```
The ‘‘friendship paradox’’ (http://en.wikipedia.org/wiki/Friendship_paradox)
says that your friends have more friends than you do.

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 themselves, sorted by
number of friends (x-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 email to the list the XML file that contains my Facebook
friendship graph ca. Oct, 2013.  The interesting part of the file looks
like this (for 1 friend):

<node id=‘‘Johan_Bollen_1448621116’’>
        <data key=‘‘Label’’>Johan Bollen</data>
        <data key=‘‘uid’’><![CDATA[1448621116]]></data>
        <data key=‘‘name’’><![CDATA[Johan Bollen]]></data>
        <data key=‘‘mutual_friend_count’’><![CDATA[37]]></data>
        <data key=‘‘friend_count’’><![CDATA[420]]></data>
</node>

It is in GraphML format: http://graphml.graphdrawing.org/
```

1.1 Solution

1. I tried to understand “Friendship paradox” from the source provided in the question.
2. It is the phenomenon which states that your friends on an average will have more friends than you.
3. Based on this phenomenon this question is phrased and the solution proves it.
4. As facebook has increased its privacy settings, we cannot scrap friends details now and so we got XML file from our professor which was previously saved in 2013.
5. Now this file should be processed and I need to get the number of friends my friends have and it compare it with my count of friends.
6. I used minidom to parse the graphml file and Tag name “data” was used to get all the data elements with this tag name.
7. Later with the help of key value from the data attributes I managed to get the “Label” and “friend_count”.
8. So people with friend_count are only considered and all of them have been written into a csv file. Python code for this can be found in 1
9. One csv file consists of all the Labels of friends and their friends count but it does not “Michael Dr.Nelson’s” friends count. This is shown in 1
10. Other csv file consists of the same data as in previous file but “Dr.Michael Nelson’s” friend count is included here. This is shown in 2
11. The reason for creating to different files is that one with Dr.Nelson’s count is used to plot graph and other one without Dr.Nelson’s count is used to calculate mean,median and standard deviation.

12. Once I got these two files I wrote code in R to find out mean,median and standard deviation and they are shown in 2.

13. A bar plot is also drawn which shows friends on x-axis and friends count for each friend on y-axis. This can been seen in 3 and the code for it can be seen in 3

14. Friends count of Dr.Michael Nelson is indicated with a blue arrow mark in the graph so that we can visualize the friendship paradox here.

15. The Mean we calculated is 358.987 and the friends count of Dr.Michael Nelson is 154. Therefore friends count of Dr.Michael Nelson is less than the Mean which shows that Michael Nelson's friends have more friends than Michael Nelson.

16. This proves Friendship Paradox using facebook friends list.

1.2 Code Listing

Here is the Python program for acquiring friends and the number of friends they have in facebook.

```
from xml.dom.minidom import parse
import csv

all_Data = parse("mln.graphml")
data_tags=all_Data.getElementsByTagName("data") #searching for data tags
csv_writer = csv.writer(open("frnds_count.csv", "wb"))
csv_writer1 = csv.writer(open("frnds_count_without_mln.csv", "wb"))
csv_writer.writerow(['Name','num_frnds'])
csv_writer1.writerow(['Name','num_frnds'])
count = 0
for nod in data_tags:

        if nod.attributes['key'].value=='Label': #searching Label attributes and getting
            their value
                name = nod.childNodes[0].data
        if nod.attributes['key'].value=='friend_count': #searching friend_count attributes
            and getting their value
                frnd_count = nod.childNodes[0].data
                csv_writer.writerow([name,int(frnd_count)])
                csv_writer1.writerow([name,int(frnd_count)])
                count=count +1 # to count number of friends of mln
csv_writer.writerow(['Michael Nelson',count])
```

Listing 1: Python program for acquiring facebook friends and the count of their friends

1.3 Results

Sample list of Friends and their count without Nelson

Name	num_frnds
Doug Nelson	7
Brian K. Saunders	15
Winnie Elliott	25
Joan A. Smith	30
Bob Mathers	38
Thomas Allen	39
Lloyd Nelson	40
Scott W Laney	41
Mary McManus	41
Greg Szalkowski	42
William Glidden Jr	43
Lori Toms	54
Jeroen Bekaert	58
Brooks Childers	59
Frank C. Thames	60
Fred Moore	62
Bonnie Austin Allen	65
Blades Elliott	68
Matthew Kesten	68
Dale Andrews	77
Donna Lynne Nelsor	80
Dongwon Lee	85
Sebastian Parreirinh	86
Ron Larsen	87
Scott Schuetz	89
Richard Weissler	93
Steve Velotas	94
Joe Reisel	94
Ashley Song	96
Brian E McCallum	97
Tom Habing	104
Joel Carter	104
Mike Weisenborn	106
John McManus	108
Trey Arthur	111
Mike Koch	123

Figure 1: Sample list of Friends and their count without Nelson

Sample list of Friends and their count with Nelson

Steve Velotas	94
Joe Reisel	94
Ashley Song	96
Brian E McCallum	97
Tom Habing	104
Joel Carter	104
Mike Weisenborn	106
John McManus	108
Trey Arthur	111
Mike Koch	123
Kurt Severance	124
Sandy Esler Ring	128
Denise Kuftic Scearce	131
Fran Austin	143
Cynthia Suarez Eudy	144
Paul Ayris	147
Michael Nelson	154
Sally Jo Cunningham	155
Cathy Marshall	165
Julianne Allen	168
Lisa Owsley Rippy	168
Laura Garlow Yarborough	170
Brenda F. Nelson	172
Miriam Blake	181
Daryl Schoolar	182
Bill Lefurgy	183
Byron Marshall	186
Seth Littman	187
Stefan Gradmann	190
Leslie Carr	195
Megan Winget	197
Gillian Hurd Peterson	204
Michael Yager	207
Dan Berkery	208
Amanda Hall	220
George Buchanan	227
Frank Shipman	229
Michael R. Witty Jr.	231

Figure 2: Sample list of Friends and their count with Nelson

R code and results for Calculation of Mean,Median and Standard Deviation

```
1
2  R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3  Copyright (C) 2014 The R Foundation for Statistical Computing
4  Platform: i386-w64-mingw32/i386 (32-bit)
5
6  R is free software and comes with ABSOLUTELY NO WARRANTY.
7  You are welcome to redistribute it under certain conditions.
8  Type 'license()' or 'licence()' for distribution details.
9
10   Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 > tmp <- read.csv("frnds_count_without_mln.csv")
21 > num_frnds <- tmp[,2]
22 > write(mean(num_frnds),stdout())
23 358.987
24 > write(median(num_frnds),stdout())
25 266.5
26 > write(sd(num_frnds),stdout())
27 371.5853
28 >
```

Listing 2: R code for Mean,Median and Standard Deviation

R code and results for plotting graph between facebook friends and their count of friends

```
1
2  R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3  Copyright (C) 2014 The R Foundation for Statistical Computing
4  Platform: i386-w64-mingw32/i386 (32-bit)
5
6  R is free software and comes with ABSOLUTELY NO WARRANTY.
7  You are welcome to redistribute it under certain conditions.
8  Type 'license()' or 'licence()' for distribution details.
9
10   Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 > tmp <- read.csv("frnds_count.csv")
21 > num_frnds <- tmp[,2]
22 > my_position <- (num_frnds==154)
23 > barplot(num_frnds,main="Facebook friends and their number of friends",xlab="Friends sorted
       by increasing number of friends",ylab="Count of Friends",ylim=c(0,4000))
24 > arrows(x0=match(c(154),num_frnds)+12, y0=(max(num_frnds) - 80), x1=match(c(154), num_frnds
     )+12, y1=200, length=0.07, lwd=1.5, col='blue')
25 > text(x=match(c(154), num_frnds)+12, y=(max(num_frnds))+200, labels="Michael nelson", col='
     blue')
26 > text(x=match(c(154), num_frnds)+12, y=(max(num_frnds))+200, labels="Michael nelson", col='
     blue')
```

Listing 3: R code for plotting graph between facebook friends and their count of friends

Graph showing facebook friends and their count of friends

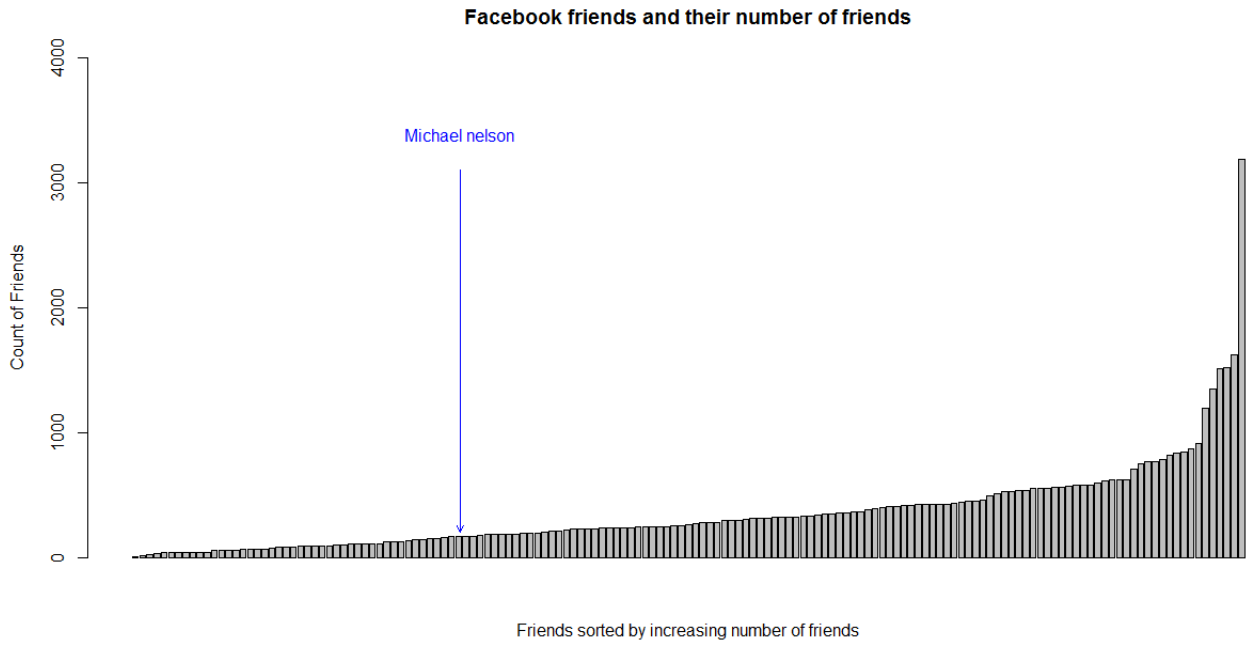

Figure 3: facebook friends and their count of friends

2 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 calcuate the same
mean, standard deviation, and median values.

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

https://dev.twitter.com/docs/api/1.1/get/followers/list

If you do not have followers on Twitter (or don’t have more than 50),
then use my twitter account ‘‘phonedude\_mln’’.
```

2.1 Solution

1. This question is about determining whether “Friendship paradox” holds for Twitter account. Here I took “followers ” as a value of measure to prove the paradox.
2. In order to prove this I need to have the twitter data,so I started to write a python code 4 to get data from twitter using twitter API. I choose the source account as my account(“dineshpaladhi”) because I have 52 followers.
3. From my previous assignments I got my Customer and Access tokens. I used tweepy library in order to get the follower’s screen name and the number of followers they have.
4. Screen_name and followers_count gives us the required data and they are stored in two csv files.
5. First csv file can be found here 4 and it contains the followers screen name and their followers count along with my followers count.
6. Second csv file can be found here 5 and it does not contains my followers count.
7. Files are saved in this way so that it would be easy for me to plot graphs and calculate mean,mode and standard deviation separately.
8. R code for calculating mean,mode and standard deviation can be seen here 5.
9. R code for bar plot can be seen here 6 and my followers count is indicated with a blue arrow mark in the graph so that we can visually see friendship paradox.This can been seen in 6.
10. The Mean I calculated is 71.365 and the followers count of “dineshpaladhi” is 52. Therefore followers count of “dineshpaladhi” is less than the Mean which shows that “dineshpaladhi’s” followers have more followers than “dineshpaladhi”.
11. This proves Friendship Paradox using Twitter’s follower list.

2.2 Code Listing

Python Code for getting Twitter followers and their count of followers

```
import tweepy
import time
import csv

csv_writer = csv.writer(open("followers_count.csv", "wb"))
csv_writer1 = csv.writer(open("followers_count_without_me.csv", "wb"))
csv_writer.writerow(['Name','num_followers'])
csv_writer1.writerow(['Name','num_followers'])
ACCESS_TOKEN = '118623489-QsSuqItzx8cnReRHI67ylffqpOPNs7z4Qp8hcOiI'  # Variables that
    contains the user credentials to access Twitter API
ACCESS_SECRET = 'PAPovgDO6QPy9QV8BbllM8p2MGWrcLLD8pesMHjXxTEMl'
CONSUMER_KEY = 'wxSZ8GSC7aRC7dAsM3m7UqgIg'
CONSUMER_SECRET = 'HuauQk780HuKWyQky9e4J6QM1DlwVxHXuvrLbgHGWhkmRXlvE4'
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)


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

for user in tweepy.Cursor(api.followers, screen_name="dineshpaladhi").items():
        csv_writer.writerow([user.screen_name,int(user.followers_count)])
        csv_writer1.writerow([user.screen_name,int(user.followers_count)])
        count=count +1 # to count number of friends of mln
csv_writer.writerow(['dineshpaladhi',count])
```

Listing 4: Python Code for getting Twitter followers and their count of followers

2.3 Results

Sample list of Twitter Followers and their count with "dineshpaladhi"

vineethchandr12	14
AkkineniAkshay	18
urs_dineshj	19
triskadieka	20
oncemanivamsi	20
bill_owns24x7	21
apoorvadasari16	22
Bharathmotha	24
maithri3	27
babisandeep	31
freinds4u2002	32
Harishkumar1772	34
joseph_udithraj	35
rlnsrlns	37
VasaviMIC	37
RamyaMattey	38
bhanupradeep1	39
likhit_kakollu	39
vaisham92	44
IA5194	45
hellboy8tg	51
dineshpaladhi	52
RithvikKranti	59
Paladhi	59
MommyOddenino	62
pranay2012	62
TCATIndia	70
maruthip25	88
DeonneLivingsto	102
aruna_teluguone	109
KlaraAdrinson	116
LatrinaMcRattig	117
OvalsSusie	166
mohanchandranp	371
alokraj68	587
telugufirst	971

Figure 4: Sample list of Twitter Followers and their count with source

Sample list of Twitter Followers and their count without "dineshpaladhi"

Name	num_followers
CsUdaykumar	3
KumarPaladhi	4
BvriteceA10	4
amrutpatil3210	4
Skumars317	5
Rajitha1829	5
majetisiri	7
JPravallika	7
vamsikrishna657	7
reddy1516	7
satvikgadam	7
papu100030	8
vineeth_sps	10
sunny24051018	11
santoshlagisety	11
Harish0123Te	12
destroyinangel	13
vineethchandr12	14
AkkineniAkshay	18
urs_dineshj	19
triskadieka	20
oncemanivamsi	20
bill_owns24x7	21
apoorvadasari16	22
Bharathmotha	24
maithri3	27
babisandeep	31
freinds4u2002	32
Harishkumar1772	34
joseph_udithraj	35
rlnsrlns	37
VasaviMIC	37
RamyaMattey	38
bhanupradeep1	39
likhit_kakollu	39
vaisham92	44

Figure 5: Sample list of Twitter Followers and their count without source

R code and results for Calculation of Mean,Median and Standard Deviation

```
1
2  R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3  Copyright (C) 2014 The R Foundation for Statistical Computing
4  Platform: i386-w64-mingw32/i386 (32-bit)
5
6  R is free software and comes with ABSOLUTELY NO WARRANTY.
7  You are welcome to redistribute it under certain conditions.
8  Type 'license()' or 'licence()' for distribution details.
9
10   Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 >  tmp <- read.csv("followers_count_without_me.csv")
21 >
22 > followers_count <- tmp[,2]
23 >
24 > write(mean(followers_count),stdout())
25 71.36538
26 >
27 > write(median(followers_count),stdout())
28 29
29 >
30 > write(sd(followers_count),stdout())
31 158.895
32 >
```

Listing 5: R code for Mean,Median and Standard Deviation

R code and results for plotting graph between Twitter followers and their count of followers

```
1
2  R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3  Copyright (C) 2014 The R Foundation for Statistical Computing
4  Platform: i386-w64-mingw32/i386 (32-bit)
5
6  R is free software and comes with ABSOLUTELY NO WARRANTY.
7  You are welcome to redistribute it under certain conditions.
8  Type 'license()' or 'licence()' for distribution details.
9
10   Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 > tmp <- read.csv("followers_count.csv")
21 > follower_count <- tmp[,2]
22 > barplot(follower_count,main="Twitter followers and Count of their followers",xlab="
       Followers sorted by the number of followers they have",ylab="Count of Followers",ylim=c
       (0,1000))
23 > arrows(x0=match(c(52),follower_count)+12, y0=700, x1=match(c(52), follower_count)+12, y1
       =70, length=0.07, lwd=1.5, col='blue')
24 > text(x=match(c(52), follower_count)+10,pos=4, y=730, labels="dineshpaladhi", col='blue')
25 >
```

Listing 6: R code for plotting graph between Twitter followers and their count of followers

Graph showing Twitter followers and their count of followers

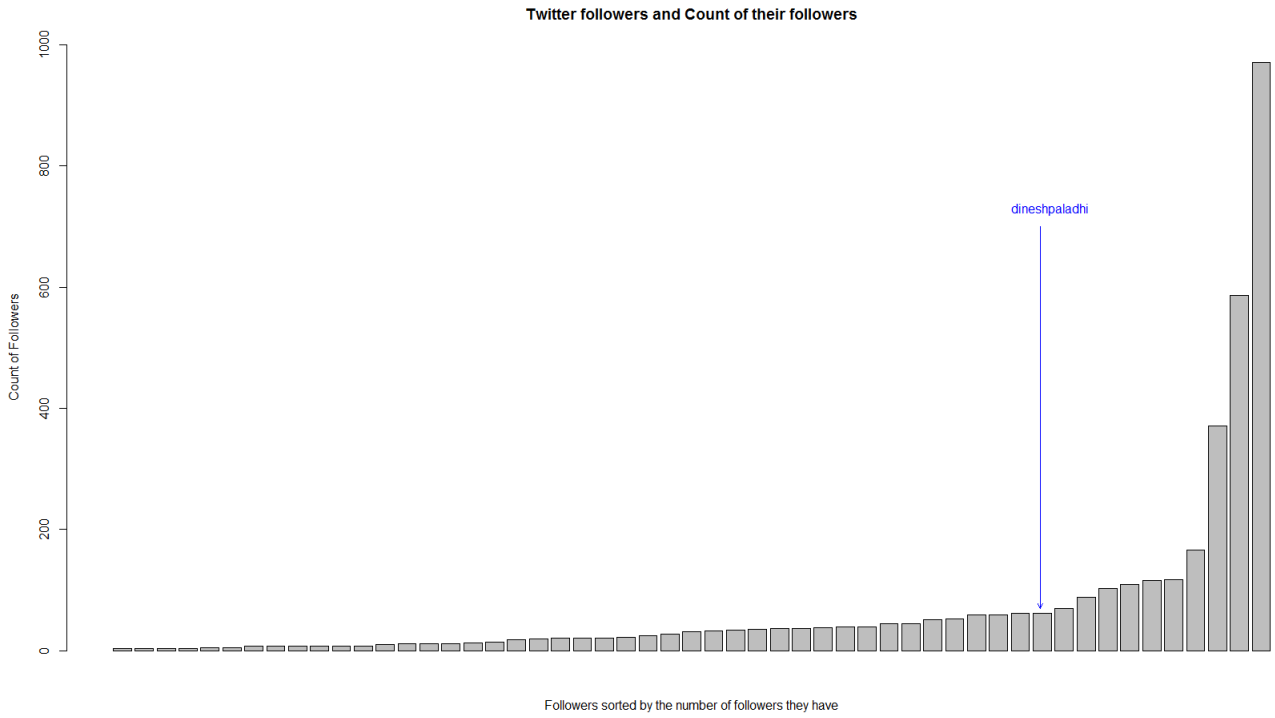

Figure 6: Twitter followers and their count of followers

3 Problem 3

```
Extra credit, 3 points:
Repeat question #1, but with your LinkedIn profile.
```

3.1 Solution

1. This question is about determining whether "Friendship paradox" holds for LinkedIn account. Here I took "connections " as a value of measure to prove the paradox.
2. This is similar to the question 1 and the process of getting LinkedIn data is different and rest of all is the same.
3. Now in order to get LinkedIn data I planned to not use API as it wasted a lot of my time and I was unsuccessful.
4. So, I thought to scrape the data using command prompt and the command to get all my connections is shown in 7
5. The sample output for this is shown in 8.
6. My plan was to strip unnecessary data from each of the my connection's link and take them into a csv file.
7. Later on I would run a for loop which takes each connection and get the number of connections they have and store in a file.
8. This gives me the required data and then the same graph,mean,median,mode can be calculated.
9. I was unable to complete this solution.

3.2 Commands

Comannd to get LinkedIn Connections

```
atria:~/Webscience/cs532-s16/Assignment 4/3> lynx -dump "http://www.cs.odu.edu/~dpaladhi/Contacts%20_LinkedI
n.html#?filter=recent" | grep -i "https://www.linkedin.com/contacts/view?id="
```

Figure 7: Comannd to get LinkedIn Connections

3.3 Results

Sample Connections

```
https://www.linkedin.com/contacts/view?id=li_308892058&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_321973817&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_321973817&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_333289816&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_333289816&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_352975695&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_352975695&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_355053674&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_355053674&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_358091196&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_358091196&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_359905605&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_359905605&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_367547946&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_367547946&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_368717207&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_368717207&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_369776711&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_369776711&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_378428801&trk=contacts-contacts-list-contact_name-0
https://www.linkedin.com/contacts/view?id=li_378428801&trk=contacts-contacts-list-contact_name-0
```

Figure 8: Page rank along with 10 URIs

4 Problem 4

```
Extra credit, 1 point:

Repeat question #2, but change ‘‘followers’’ to ‘‘following’’?  In
other words, are the people I am following following more people?
```

4.1 Solution

1. This question is similar to the second question. There I used “followers” and here I need to use “following”.
2. I choose the source account as my account(“dineshpaladhi”) because I have 186 people following me.
3. So, this should prove that “people I am following following more people?”
4. To make it more understandable, here I used “friends” to represent “following” people .
5. Python code is same for this except for the api object we use. Earlier I used api.followers in order get followers and now I used api.friends in order to get the list of people I am following. Python code for this can be seen here 7.
6. First csv file can be found here 9 and it contains the friends screen name and their friends count along with my friends count.
7. Second csv file can be found here 10 and it does not contains my friends count.
8. R code for calculating mean,mode and standard deviation can be seen here 8.
9. R code for bar plot can be seen here 9 and my friends count is indicated with a blue arrow mark in the graph so that we can visually see friendship paradox.This can been seen in 11.
10. The Mean I calculated is 8292.253 and the following people count of “dineshpaladhi” is 186. Therefore following people count of “dineshpaladhi” is less than the Mean which shows that “dineshpaladhi’s” friends have more friends than “dineshpaladhi”.
11. This proves Friendship Paradox using Twitter’s following list.

4.2 Code Listing

Python Code for getting Twitter friends and their count of friends

```
import tweepy
import time
import csv

csv_writer = csv.writer(open("following_count.csv", "wb"))
csv_writer1 = csv.writer(open("following_count_without_me.csv", "wb"))
csv_writer.writerow(['Name','num_following'])
csv_writer1.writerow(['Name','num_following'])
ACCESS_TOKEN = '118623489-QsSuqItzx8cnReRHI67ylffqpOPNs7z4Qp8hcOiI'  # Variables that
    contains the user credentials to access Twitter API
ACCESS_SECRET = 'PAPovgDO6QPy9QV8BbllM8p2MGWrcLLD8pesMHjXxTEMl'
CONSUMER_KEY = 'wxSZ8GSC7aRC7dAsM3m7UqgIg'
CONSUMER_SECRET = 'HuauQk780HuKWyQky9e4J6QM1DlwVxHXuvrLbgHGWhkmRXlvE4'
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)


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

for user in tweepy.Cursor(api.friends, screen_name="dineshpaladhi").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(['dineshpaladhi',count])
```

Listing 7: Python Code for getting Twitter friends and their count of friends

4.3 Results

Sample list of Twitter friends and their count with "dineshpaladhi"

vaisham92	125
vennelakishore	126
bomanirani	127
vishnu_dir	137
tamannaahspeal	138
ImZaheer	138
LeoDiCaprio	142
chetan_bhagat	142
bipsluvurself	142
ArvindKejriwal	143
Samanthaprabhu	148
hellboy8tg	149
harbhajan_singh	150
trishtrashers	150
BillGates	169
LakshmiManchu	175
TheVijayMallya	178
shakira	178
ActorMadhavan	185
dineshpaladhi	186
the_hindu	193
pranay2012	194
maruthip25	202
vivek_oberoi	210
thekiranbedi	217
RanaDaggubati	219
FarOutAkhtar	225
shrutihaasan	229
IndiaToday	231
udaychopra	237
TimesNow	238
ParineetiChopra	243
Riteishd	244
ImRaina	247
SunRisers	248
Actorjiiva	249
priyamani6	253

Figure 9: Sample list of Twitter friends and their count with source

Sample list of Twitter friends and their count without "dineshpaladhi"

Name	num_following
DalaiLama	0
RJPrateeka	0
arrahman	0
Virajitha_0805	1
iamnagarjuna	1
urstrulyMahesh	1
sonunigam	2
menongautham	2
SCGanguly	4
BvriteceA10	6
aamir_khan	6
amir_khan	6
mangeshkarlata	7
MSDmahi	8
vineeth_sps	8
ashabhosle	9
reddy1516	14
majetisiri	14
satvikgadam	14
sachin_rt	15
BeingSalmanKhan	16
KatrinaKaif4u	16
NatGeo	17
IA5194	17
akshaykumar	18
ssrajamouli	18
superstarrajini	19
AkhilAkkineni8	19
vijay_m92	19
virendersehwag	21
sai_yerneni	21
oncemanivamsi	21
ndtv	23
iSumanth	24
MadhuriDixit	25
Harish0123Te	29

Figure 10: Sample list of Twitter friends and their count without source

R code and results for Calculation of Mean,Median and Standard Deviation

```
1
2  R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3  Copyright (C) 2014 The R Foundation for Statistical Computing
4  Platform: i386-w64-mingw32/i386 (32-bit)
5
6  R is free software and comes with ABSOLUTELY NO WARRANTY.
7  You are welcome to redistribute it under certain conditions.
8  Type 'license()' or 'licence()' for distribution details.
9
10   Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 > tmp <- read.csv("following_count_without_me.csv")
21 > following_count <- tmp[,2]
22 >
23 > write(mean(following_count),stdout())
24 8292.253
25 >
26 > write(median(following_count),stdout())
27 104.5
28 >
29 > write(sd(following_count),stdout())
30 59357.03
31 >
```

Listing 8: R code for Mean,Median and Standard Deviation

R code and results for plotting graph between Twitter friends and their count of friends

```
1
2  R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
3  Copyright (C) 2014 The R Foundation for Statistical Computing
4  Platform: i386-w64-mingw32/i386 (32-bit)
5
6  R is free software and comes with ABSOLUTELY NO WARRANTY.
7  You are welcome to redistribute it under certain conditions.
8  Type 'license()' or 'licence()' for distribution details.
9
10   Natural language support but running in an English locale
11
12 R is a collaborative project with many contributors.
13 Type 'contributors()' for more information and
14 'citation()' on how to cite R or R packages in publications.
15
16 Type 'demo()' for some demos, 'help()' for on-line help, or
17 'help.start()' for an HTML browser interface to help.
18 Type 'q()' to quit R.
19
20 > tmp <- read.csv("following_count.csv")
21 > following_count <- tmp[,2]
22 > 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))
23 > arrows(x0=match(c(186),following_count)+12, y0=700, x1=match(c(186), following_count)+12,
     y1=150, length=0.07, lwd=1.5, col='blue')
24 > text(x=match(c(186), following_count),pos=4, y=730, labels="dineshpaladhi", col='blue')
25 >
```

Listing 9: R code for plotting graph between Twitter friends and their count of friends

Graph showing Twitter friends and their count of friends

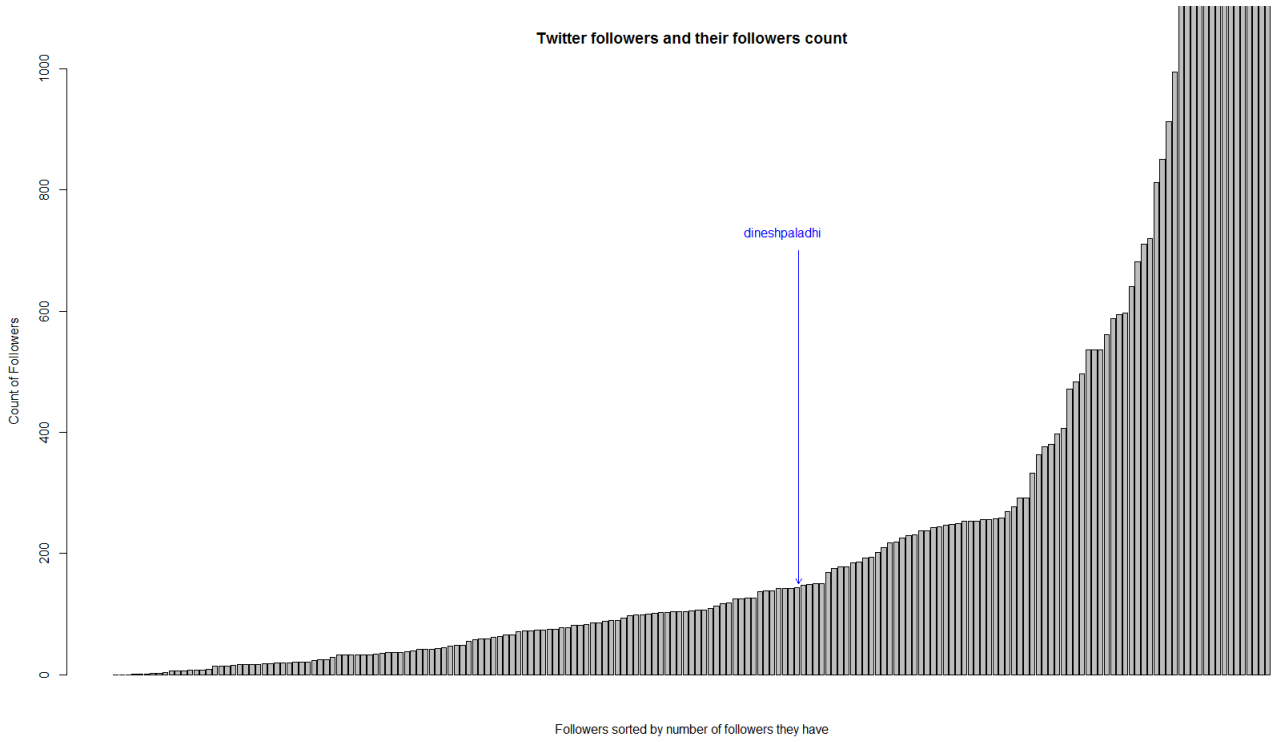

Figure 11: Twitter friends and their count of friends

Bibliography

[1] Friendship paradox : https://en.wikipedia.org/wiki/Friendship_paradox , Wikipedia, 2015

[2] MiniDom : https://wiki.python.org/moin/MiniDom, MartinvonLoewis, 2012

[3] Python Read XML file : http://www.mkyong.com/python/python-read-xml-file-dom-example/, mkyong, 2014

[4] Add Text to a Plot : https://stat.ethz.ch/R-manual/R-devel/library/graphics/html/text.html, The R Graphics Package, 2013

[5] Add Arrows to a Plot : https://docs.python.org/2/library/math.html, The R Graphics Package, 2013

[6] GET followers/list : https://dev.twitter.com/rest/reference/get/followers/list, Twitter Incorporation, 2015

[7] The best way to get a list of followers in Python with Tweepy : http://stackoverflow.com/questions/17455107/the-best-way-to-get-a-list-of-followers-in-python-with-tweepy, alecxe, 2013

[8] How to use linkedin API with python : http://stackoverflow.com/questions/17600244/how-to-use-linkedin-api-with-python, user1681664, 2013