Cover Sheet

By including this statement, we, all the students listed in the table below, declare that:

- We hold a copy of this assignment if the original is lost or damaged.

- We hereby certify that no part of this assignment has been copied from any other student's work or from any other source except where due acknowledgement is made in the assignment.

- No part of the assignment has been written for us by any other person except where collaboration has been authorised by the unit coordinator.

- We are aware that this work may be reproduced and submitted to plagiarism detection software programs for the purpose of detecting possible plagiarism; this software may retain a copy on its database for future plagiarism checking.

- We hereby certify that no part of this assignment or product has been submitted by any of us in another (previous or current) assessment, except where appropriately referenced, and with prior permission from the unit coordinator for this unit.

- We hereby certify that we have read and understand what the University considers to be academic misconduct, and that we are aware of the penalties that may be imposed for academic misconduct.

Name	Student Number	Contribution (%)
Rachel Hardie	18820821	25%
Dylan Wang	18998014	25%
Dylan Yoo	18640377	25%
Sreenath Ramachandran	18878716	25%

```
library("tm")
library("rtweet")
library("twitteR")
library("igraph")
library("knitr")
tweets=read.csv("SWAProject.csv")
```

We then constructed a document-term matrix using TFIDF weighting to describe the frequency of terms in the tweets we had collected.

```
corpus=Corpus(VectorSource(tweets$text))
corpus = tm_map(corpus, function(x) iconv(x, to = 'UTF8', sub = 'byte'))
corpus = tm_map(corpus, function(x) iconv(x, to = 'ASCII', sub = ' '))
corpus = tm_map(corpus,removeNumbers)
corpus = tm_map(corpus, removeWords,c(stopwords(),"J.K. Rowling","https", "t.co"))
corpus = tm_map(corpus,removePunctuation)
corpus = tm_map(corpus,stripWhitespace)
corpus = tm_map(corpus, tolower)
corpus = tm_map(corpus, stemDocument)
corpus = tm_map(corpus, removeWords,c(stopwords(),"jk","https", "t.co","rowl","jkrowl",
                                 "will","like","just","doe","far","quot","look",
                                 "take","make","jaqkcpmtg","ygorfremo","ask","peopl",
                                 "next","twitter","peopl","new","write","charact","got",
                                 "ohhgmljqrm","say","defend","respond", "vrmgsnrsa",
                                 "bbepgeik","nkuwxgsn","novrijqr","nlrbgguhxt"))
dtm = DocumentTermMatrix(corpus)
tweet.wdtm = weightTfIdf(dtm)
tweet.matrix = as.matrix(tweet.wdtm)
```

Finding the similarity index

To find the cosine similarity matrix we first find the normalised tweet matrix.

```
#Find similarity matrix
S=tweet.matrix%*%t(tweet.matrix)
#to find the cosine similarity matrix we need  the normalised tweet matrix.
norm.tweet.matrix = diag(1/sqrt(rowSums(tweet.matrix^2))) %*% tweet.matrix
#cosine similarity
CS=norm.tweet.matrix%*%t(norm.tweet.matrix)

head(S)[1:20]
```

```
##  [1] 2.78426986 0.00000000 0.00000000 0.00000000 0.00000000 0.23041424
##  [7] 0.00000000 0.49056435 0.05247791 0.10696634 0.49056435 0.00000000
## [13] 0.00000000 0.05247791 1.30694847 0.00000000 0.05247791 0.00000000
## [19] 0.00000000 0.10696634
```
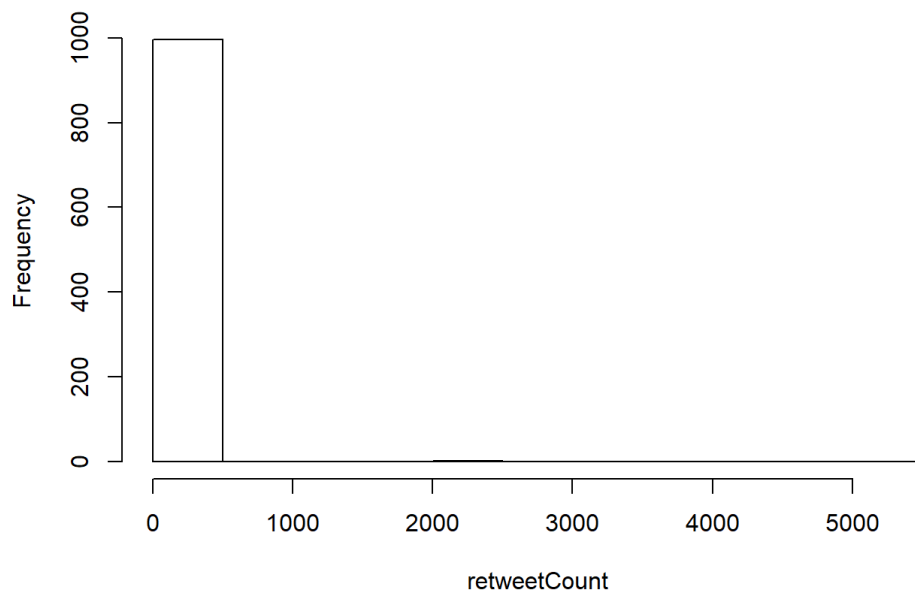
```
head(CS)[1:20]
```

```
##  [1] 1.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.07619747
##  [7] 0.00000000 1.00000000 0.06553893 0.06550168 1.00000000 0.00000000
## [13] 0.00000000 0.06553893 1.00000000 0.00000000 0.06553893 0.00000000
## [19] 0.00000000 0.06550168
```

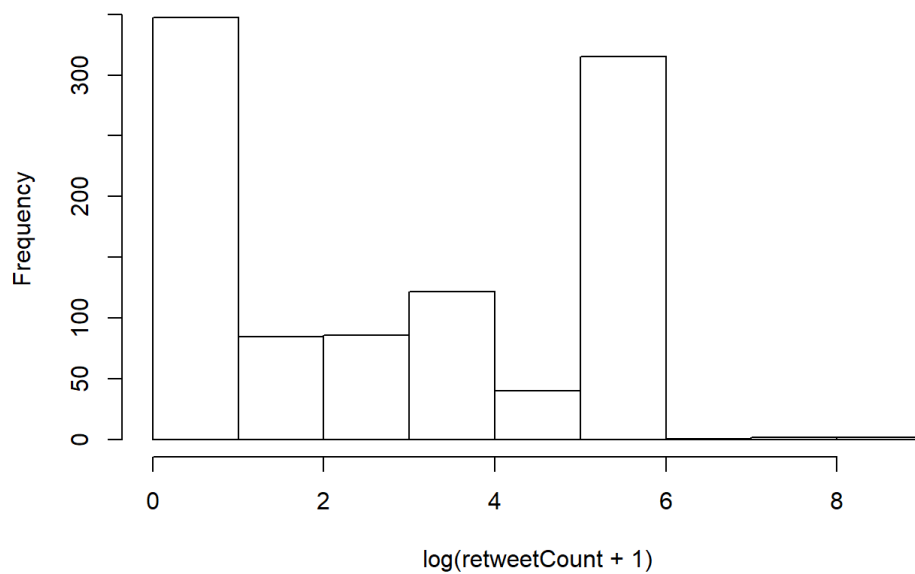# Difference between Cosine similarity and S

Cosine similarity gives the angle between the documents. The similarity matrix is more like the magnitude of the difference between the documents whereas the cosine similarity matrix gives the direction.Cosine similarity of the documents will be between 0 and 1. ##8.4.2 As the histogram of retweet count is exponenetial we take logarithm of retweet counts we take the log of retweet count+1 so as to eliminate log 0 from the data

```
#8.4.2
retweetCount=tweets$retweet_count

hist(retweetCount)
```
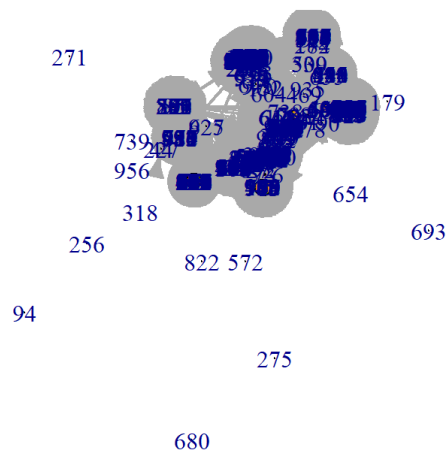
Histogram of retweetCount


```
hist(log(retweetCount+1))
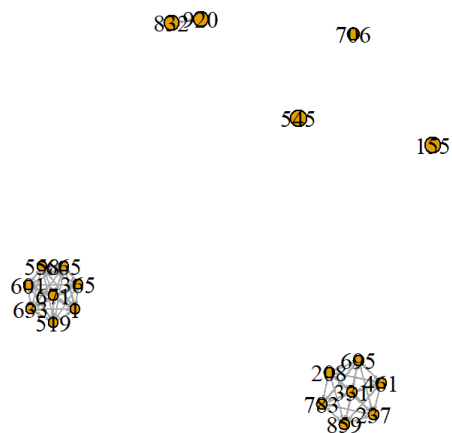```

### Histogram of log(retweetCount + 1)



# 8.4.3

To make the graph simpler we are taking the top 20 retweeted tweets.

```
graph=graph.adjacency(S,weighted = TRUE)
plot(simplify(graph),layout = layout.drl,vertex.size=log(retweetCount+1))
```

271

179

739
956
625
318
654
256
693
822 572
94
275
680

```
# to make the graph simpler we are taking the top 20 retweeted tweets.
id=order(log(retweetCount+1),decreasing = TRUE)[1:20]
tweets.imp=tweets[id,]
retweetCount.imp=tweets.imp$retweet_count
#retweetCount.imp
tweet.imp=tweet.matrix[id,]
s.imp=tweet.imp%*%t(tweet.imp)
set.seed(1500)
graph1=graph.adjacency(s.imp,weighted = TRUE)
plot(simplify(graph1),layout=layout.auto,
     vertex.size=log(retweetCount.imp+1),
     edge.arrow.size=0.1,vertex.label.size=.5,vertex.label.color="black")
```

706

545

105

Results

Connected tweets mostly are the same tweets retweeted by different users.

8.5.1

Degree centrality

```
degCentrality=degree(graph)
topTweetsDegree=order(degree(graph),decreasing = TRUE)[1:10]
tweets$text[topTweetsDegree]
```

```
##  [1] Macaulay Culkin tweets at J.K. Rowling asking for a job - https://t.co/BYCO6Q5Rjd \n\nMacaulay Culkin wants to be pa
rt of J.K. Rowlingâ\200\230sÂ Harry Potter universe.\nAfter receiving some backlash for casting an Asian actress as a villai
nâ\200\231s slave in the upcoming â\200□Fantastic Beast...
##  [2] Harry Potter fans really want Macaulay Culkin to get a part in Fantastic Beasts after his tweet to J.K. Rowling\nhtt
ps://t.co/MZnaUpyggm https://t.co/CrYlJKjx26

##  [3] #1 (and only) reason I will never watch Harry Potter movies (including Fantastic Beasts) or read any of the books?
J.K. Rowling.

##  [4] Elijah, I also see that Macaulay has talked to J.K. Rowling about being cast in the Fantastic Beast movie.. I'm glad
to know he's back into the movies.. Feel free to write me messages on Twitter or E-Mail me at some time if you want to!!!!

##  [5] @hermioneswife @jahuraaa @piccolosturban @luciusobsessed Yikes. Tbh does J.K. Rowling even get to chose the casting
100% just from the Harry Potter series alone there werenâ\200\231t poc at all. Or the few poc described in the book werenâ\2
00\231t in the film or mentioned in the film
##  [6] J.K. Rowling confirms massive fan conspiracy in 'Harry Potter' universe The final trailer for the highly anticipated
second installment of the "Fantastic Beasts" franchise, "The Crimes of... https://t.co/igWMjjgREQ

##  [7] I don't really have an issue with Nagini being cast as a Korean woman. No problem there.\n\nI do, however, have a pr
oblem with J.K. Rowling pulling random out of her ass about the Wizarding World.\n\nI get it: she wrote it, she can do what
she wants. But it still comes across as lazy
##  [8] I've read fantastic beasts and where to find me but there isn't a book about the Crimes of Grindelwald, that I know.
\n\nBut I know F.B is an extention of the wizarding world that J.K Rowling recently started. https://t.co/yKKNjU5fKj

##  [9] Macaulay Culkin sent his CV to J.K. Rowling for a part in Fantastic Beasts - now people really want him to be cast -
https://t.co/sNCq6qT5AS https://t.co/3hlb8ulVbk

## [10] J K Rowling being J K Rowling. Seven books and not a single peep. In deathly hallows, Nagini is an old woman ready t
o jump harry. But Maledictus turn into beasts PERMANENTLY <U+0001F926><U+0001F3FE><U+200D><U+2642><U+FE0F><U+0001F926><U+000
1F3FE><U+200D><U+2642><U+FE0F><U+0001F926><U+0001F3FE><U+200D><U+2642><U+FE0F> https://t.co/l4Um3gIHVV
## 412 Levels: 'A racist, misogynist disaster': J.K. Rowling responds as fans slam Fantastic Beasts casting\nhttps://t.co/u1
aQMhhkDV ...
```

Closeness centrality

```
closenessCentrality=closeness(graph)
```

```
## Warning in closeness(graph): At centrality.c:2617 :closeness centrality is
## not well-defined for disconnected graphs
```

```
topTweetsCloseness=order(closeness(graph),decreasing = TRUE)[1:10]
```

```
## Warning in closeness(graph): At centrality.c:2617 :closeness centrality is
## not well-defined for disconnected graphs
```

```
tweets$text[topTweetsCloseness]
```

```
##  [1] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [2] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [3] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [4] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [5] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [6] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [7] Me in 2016: "Oh boy, more movies set in the Harry Potter universe, written by J.K. Rowling herself! This should be f
ascinati--"\n\n2018: "Abusey McSparrowHatter is our bad guy and Nagini was a Korean lady the whole time."\n\nMe: https://t.c
o/mXO3IXzDSM
##  [8] @jk_rowling @J_A_Moulton Nagini is an Indian word and â\200\230Nagaâ\200\231 people are Indian mythology, which it s
eems J K didnâ\200\231t even think of naming as a part of Indonesian ethnic groups. \nThatâ\200\231s because she had already
done Parvati so didnâ\200\231t want to double up and cause too much representation.
##  [9] @jk_rowling @J_A_Moulton Nagini is an Indian word and â\200\230Nagaâ\200\231 people are Indian mythology, which it s
eems J K didnâ\200\231t even think of naming as a part of Indonesian ethnic groups. \nThatâ\200\231s because she had already
done Parvati so didnâ\200\231t want to double up and cause too much representation.
## [10] @jk_rowling @J_A_Moulton Nagini is an Indian word and â\200\230Nagaâ\200\231 people are Indian mythology, which it s
eems J K didnâ\200\231t even think of naming as a part of Indonesian ethnic groups. \nThatâ\200\231s because she had already
done Parvati so didnâ\200\231t want to double up and cause too much representation.
## 412 Levels: 'A racist, misogynist disaster': J.K. Rowling responds as fans slam Fantastic Beasts casting\nhttps://t.co/u1
aQMhhkDV ...
```

Betweenness centrality

```
betweennessCentrality=betweenness(graph)
topTweetsBetweenness=order(betweenness(graph),decreasing = TRUE)[1:10]
tweets$text[topTweetsBetweenness]
```

```
##  [1] J K Rowling being J K Rowling. Seven books and not a single peep. In deathly hallows, Nagini is an old woman ready t
o jump harry. But Maledictus turn into beasts PERMANENTLY <U+0001F926><U+0001F3FE><U+200D><U+2642><U+FE0F><U+0001F926><U+000
1F3FE><U+200D><U+2642><U+FE0F><U+0001F926><U+0001F3FE><U+200D><U+2642><U+FE0F> https://t.co/l4Um3gIHVV
##  [2] @hermioneswife @jahuraaa @piccolosturban @luciusobsessed Yikes. Tbh does J.K. Rowling even get to chose the casting
100% just from the Harry Potter series alone there werenâ\200\231t poc at all. Or the few poc described in the book werenâ\2
00\231t in the film or mentioned in the film
##  [3] So J.K. Rowling just revealed that Nagini was a human?\nLike\nShe wrote HP treating her like a pet, every character
and the narrator treated her like a pet. It would make her being a slave till her death. Thatâ\200\231s gross. \n\nRowling,
you just shouldâ\200\231ve added LGBTs. Far less complex.
##  [4] @saltybisque @BeastsMovieUK @jk_rowling Exactly. So, why would a survivor allow and bask his casting? J.K Rowling is
a survivor and if there was any truth to his case then I donâ\200\231t think she wouldâ\200\231ve been okay with it. Remembe
r we donâ\200\231t know what truly happened but you can do some (actual) research
##  [5] Elijah, I also see that Macaulay has talked to J.K. Rowling about being cast in the Fantastic Beast movie.. I'm glad
to know he's back into the movies.. Feel free to write me messages on Twitter or E-Mail me at some time if you want to!!!!

##  [6] Major Bombshell: J.K. Rowling Has Revealed Daniel Radcliffe Is Chinese https://t.co/DKKxmNZ3Kw https://t.co/Ve03iAC1
aT

##  [7] Major Bombshell: J.K. Rowling Has Revealed Daniel Radcliffe Is Chinese https://t.co/DKKxmNZ3Kw https://t.co/Ve03iAC1
aT

##  [8] Don't tell J.K. Rowling. The Japanese American National Museum is the place to find fantastic beasts. Dive into the
colorful world of toy designer Mark Nagata in the @jamuseum 's latest exhibit.  https://t.co/9IT90ba67z

##  [9] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie: Macaulay Culkin wants to make ano
ther big splash in Hollywood -- and he's shamelessly, and publicly, lobbying J.K. Rowling to help him do it. The 38-year-old
'Home Alone'â\200¦ https://t.co/1GtyysXvIB https://t.co/9a3RObly2z9
## [10] @jk_rowling @J_A_Moulton Nagini is an Indian word and â\200\230Nagaâ\200\231 people are Indian mythology, which it s
eems J K didnâ\200\231t even think of naming as a part of Indonesian ethnic groups. \nThatâ\200\231s because she had already
done Parvati so didnâ\200\231t want to double up and cause too much representation.

## 412 Levels: 'A racist, misogynist disaster': J.K. Rowling responds as fans slam Fantastic Beasts casting\nhttps://t.co/u1
aQMhhkDV ...
```

8.5.2 Page rank to calculate the most influential tweets

```
adjacency.to.probability = function(A) {
  cols = ncol(A)
  for (a in 1:cols) {
    A[, a] = normalise(A[, a])
  }
  return(A)
}
normalise=function(x){
  if(sum(x)!=0){
    x=x/sum(x)
  }else{
    x=x
  }
  return(x)
}
T = adjacency.to.probability(S)
J = matrix(rep(1/1000, 1000 * 1000), 1000, 1000)
alpha=0.8
M = alpha * T + (1 - alpha) * J
M = adjacency.to.probability(M)
#Power method to find stationary distribution
#function to calculate Euclidean distance between two vectors.
differenceEuc=function(x,y){
  return(sqrt(sum((x-y)^2)))
}
stationary.distribution = function(T) {
  # first create the initial state distribution
  n = ncol(T)
  p = rep(0, n)
  p[1] = 1

  # now take a random walk until the state distribution reaches the
  # stationary distribution.
  p.old = rep(0, n)
  while (differenceEuc(p, p.old) > 1e-06) {
    p.old = p
    p = T %*% p.old
  }
  return(p)
}

p = stationary.distribution(M)
influentialTweets=order(p,decreasing = TRUE)[1:11]
#took 11 to get 10 distinct users as one user was repeat in top10.
tweets$text[influentialTweets]
```

```
##  [1] 'Harry Potter' (1997 - 2007) was a book series by J.K. Rowling.
##  [2] Macaulay Culkin Wants J.K. Rowling to Write Him Into 'Fantastic Beasts'
##  [3] J.K. Rowling is racist. don't @ me
##  [4] We don't deserve J.K. Rowling.
##  [5] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/nKU39wXgSn
##  [6] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/nKU39wXgSn
##  [7] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/nKU39wXgSn
##  [8] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/nKU39wXgSn
##  [9] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/nKU39wXgSn
## [10] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/vrmGsnr6SA
## [11] Macaulay Culkin Asks J.K. Rowling to Write Him into Next 'Fantastic Beasts' Movie https://t.co/nKU39wXgSn
## 412 Levels: 'A racist, misogynist disaster': J.K. Rowling responds as fans slam Fantastic Beasts casting\nhttps://t.co/u1
aQMhhkDV ...
```

8.5.3 Influential Users

```
users=tweets$screen_name[influentialTweets]
users
```

```
##  [1] culturalkumite  MekoStarr        meka_activated  KaylinGie14
##  [5] SHOT97KRABDADDY BroGod4Life      KMJonAIR        RedRoman2
##  [9] FrankWHmag      FrankWHmag       Hard2bAhDiamond
## 961 Levels: __CAR0USEL__ __nocturnale _carol_clarinet ... zwolftenaugust
```

8.5.4 influence ratio

```
InfluenceRatio= tweets$followers_count[influentialTweets]/
                tweets$friends_count[influentialTweets]
influence <- data.frame(users,InfluenceRatio)
names(influence) <- c("User Names","Influence Ratio")
influence
```

```
##          User Names Influence Ratio
## 1     culturalkumite      61.0000000
## 2          MekoStarr       0.4434639
## 3     meka_activated       3.6000000
## 4         KaylinGie14       4.5000000
## 5   SHOT97KRABDADDY       0.8822614
## 6         BroGod4Life       0.4813763
## 7           KMJonAIR       6.0286659
## 8          RedRoman2       0.5573470
## 9          FrankWHmag       0.8508475
## 10         FrankWHmag       0.8508475
## 11 Hard2bAhDiamond       0.3292683
```
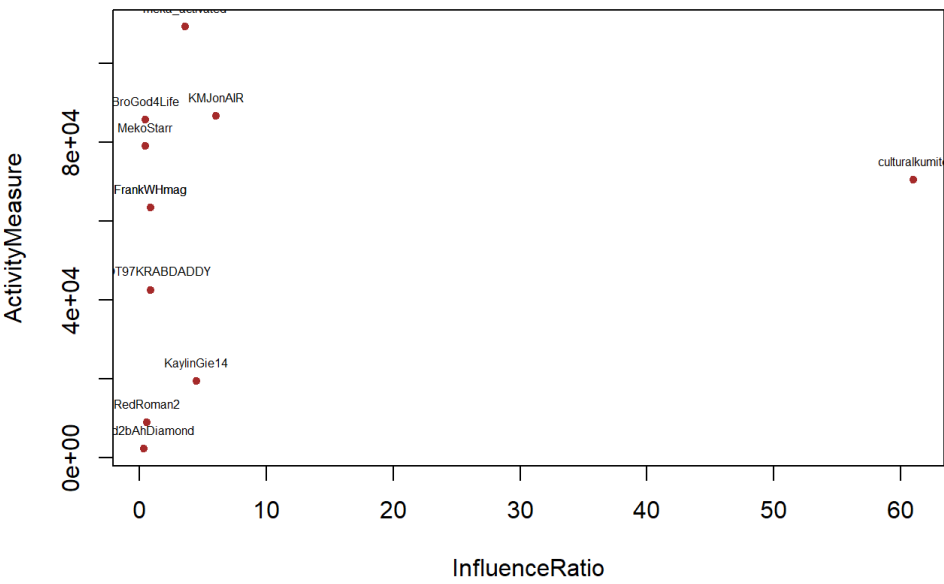
8.5.4 Activity measures

```
ActivityMeasure=tweets$statuses_count[influentialTweets]
activity <- data.frame(users,ActivityMeasure)
names(activity) <- c("User Names","Activity Measure")
activity
```

```
##          User Names Activity Measure
## 1     culturalkumite            70521
## 2          MekoStarr            79068
## 3     meka_activated           109267
## 4         KaylinGie14            19296
## 5   SHOT97KRABDADDY            42525
## 6         BroGod4Life            85739
## 7           KMJonAIR            86606
## 8          RedRoman2             8933
## 9          FrankWHmag            63404
## 10         FrankWHmag            63404
## 11 Hard2bAhDiamond             2202
```

8.5.5

```
plot(InfluenceRatio,ActivityMeasure,col="brown",pch=20,
     text(InfluenceRatio,ActivityMeasure,labels = users,cex = 0.5,pos=3))
```

Findings

The most influential user 'culturalkumite' is a bot. The high influential ratio is as it has 61 followers and it follows only 1 user.This may be due due to the fact that it tweets about authors, film and art in general.People might be following the bot to get latest news in those fields. Other users of influential tweets do not have a very high influence ratio and are most likely fans of J.K.Rowling's work.

Conclusion

J.K Rowling is a very popular author. Tweets about her often changes. Of the tweets we have downloaded most of them were about her creations. Most of the influential tweets were about the movie Fantastic beasts. There were some users who were accusing Rowling of Racism in the charecters she created, particularly for the 'nagini' role in the film Fantastic Beasts.