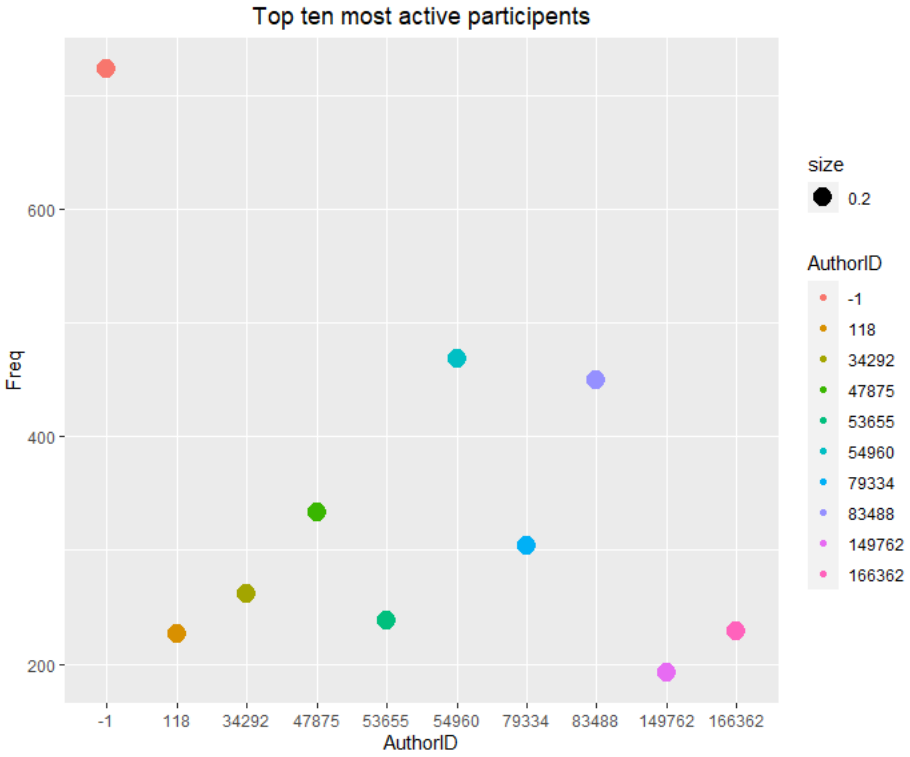
By looking at basic summary statistics, we notice that there is an author with AuthorID of -1 (the minimum of AuthorIDs (Graph (1.0))). Further investigation is needed as all the other authors have been assigned a positive ID number (integer). The number of posts by most active authors may shed light on this irregularity.

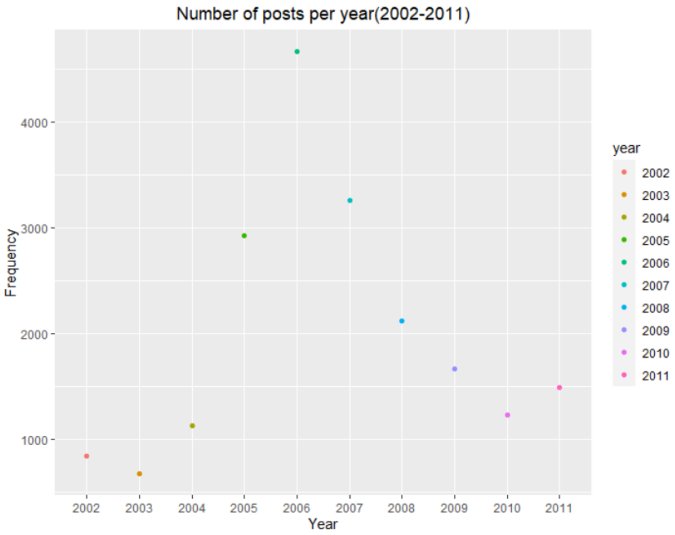
```
    AuthorID
Min.   :    -1
1st Qu.: 38226
Median : 79334
Mean   : 82158
3rd Qu.:116333
Max.   :252144
```

Graph (1.0)

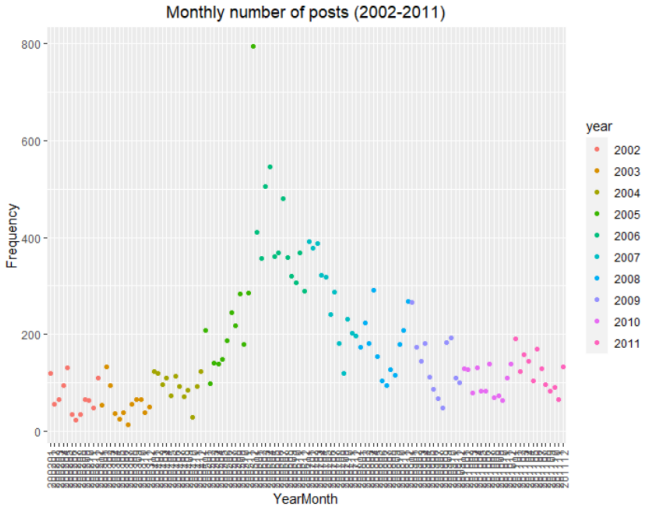


Graph (1.1)

The mysterious Author -1 has posted 200 more posts than the next most active author (Graph (1.1)). We may assume that the Author -1 represents authors who have wished to post anonymously.

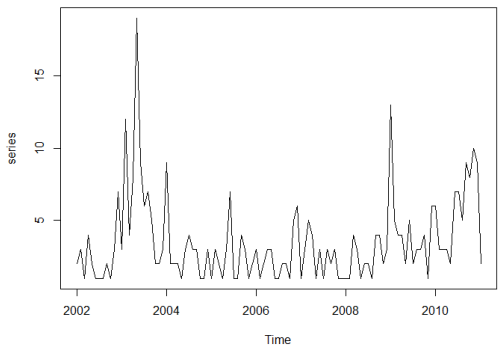


Graph (1.2)

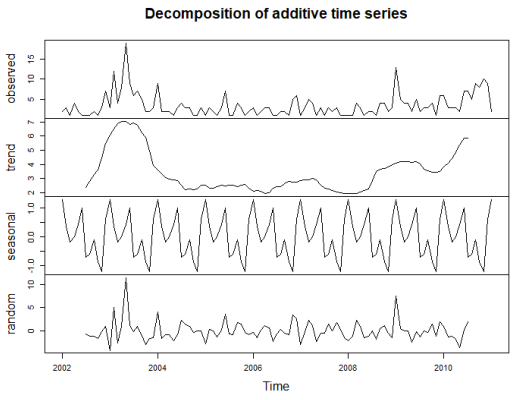


Graph (1.3)

The activity on the forum started with less than one thousand posts per year back in 2002 which is three posts on average per day. Then, it quadrupled in next four years and reached a high of more than four thousand and five hundred posts in 2006. Despite this steady upward trend, the activity on the forum gradually dived to the 2004 level again. By investigating the language used in the posts from 2004 to 2006, we may shed light on the reason/s behind this surge in activity in this period and whether it can be replicated in the future.

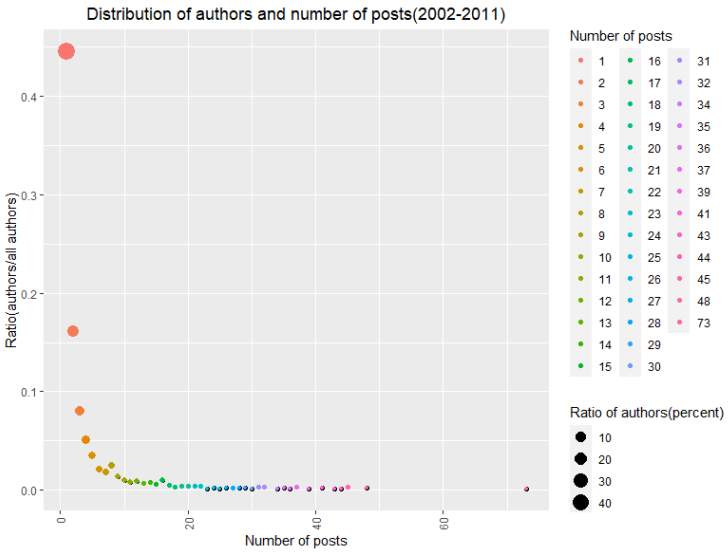


Graph (1.4)

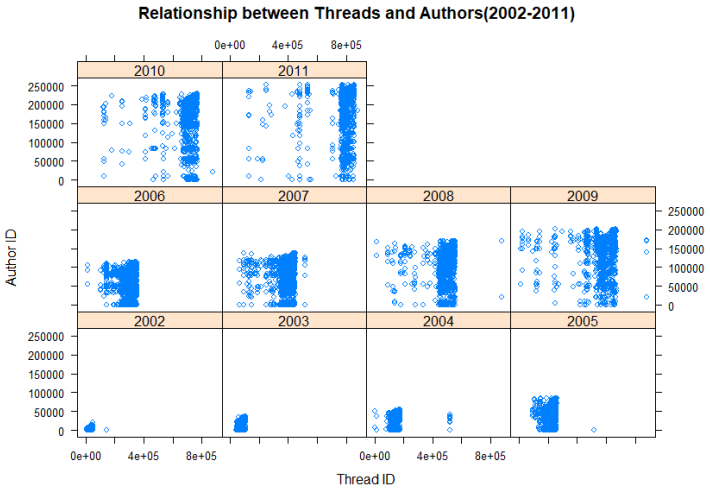


Graph (1.5)

The above time series graphs manifest the similar trend already observed from graph (1.2) and graph (1.3).



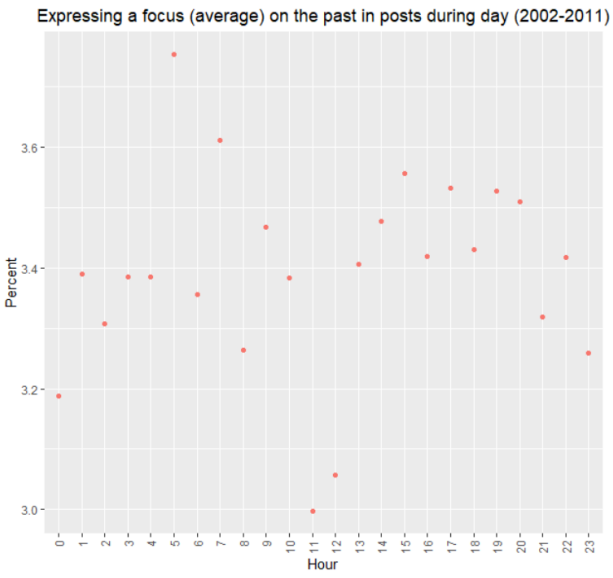
Graph (1.6)



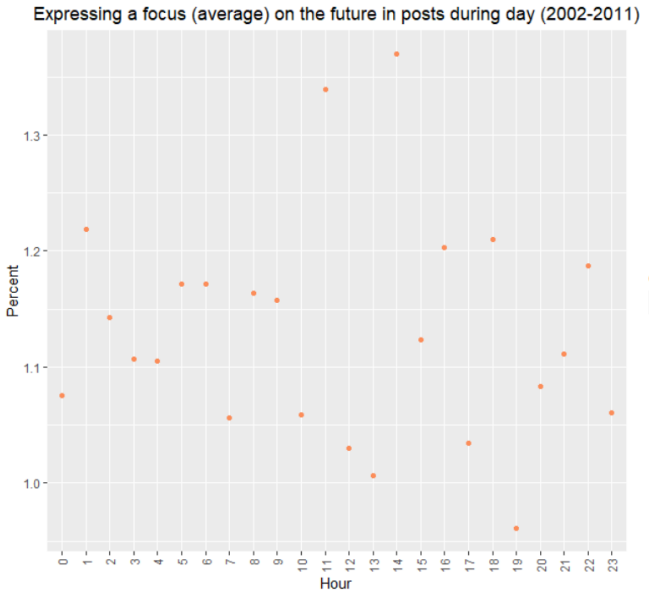
Graph (1.7)

The number of active participants in forums (i.e., those who have posted more than ten posts in ten-year period) is only a small share of all authors who contributed to this forum (Graph 1.6). However, this metric alone cannot be used to gauge the overall level of activities in the forum. Since many people who have visited this forum may have seen that other authors have already expressed ideas similar to them succinctly enough. Hence, they did not feel the need to repeat what has already been expressed in the forum.

The majority of activities in the forum are concentrated on newer threads than old ones (graph 1.7). Also, the forum has been successful in gaining new participants but not so much in maintaining their level of engagement in the form of posting (Graph 1.7). This forum seems to host threads which most of them are time sensitive as activity in the old threads drop significantly over time. In general, most social networks serve as town square which draws public attention by capitalising on the fear of missing out (FoMO) (Alutaybi, Al-Thani, McAlaney & Ali, 2020). That's why majority of traffic on the network is concentrated and threads are short-lived.

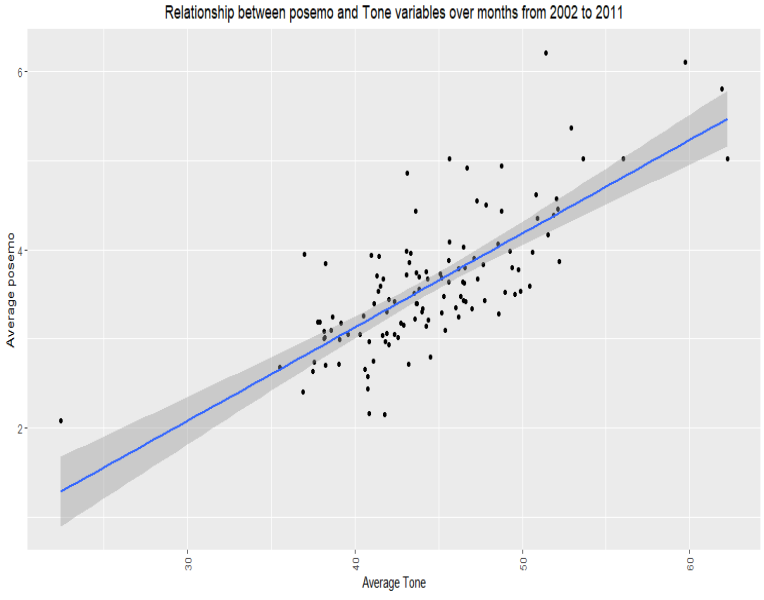


Graph (1.8)

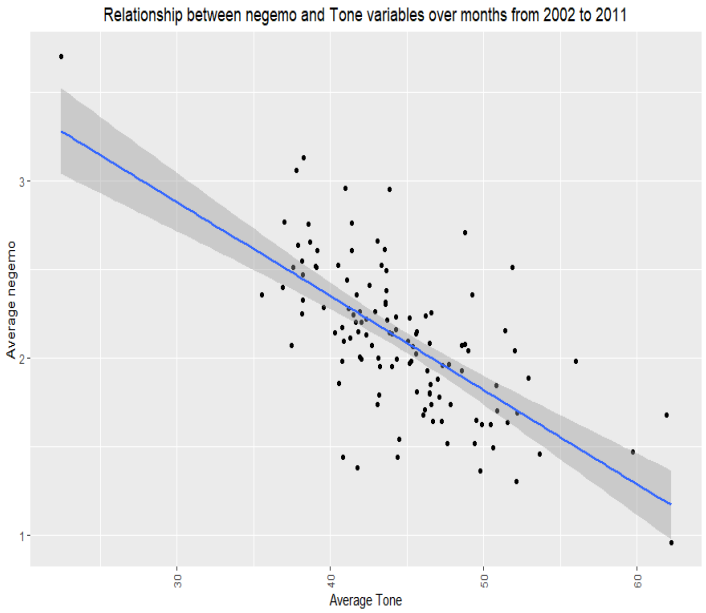


Graph (1.9)

It seems focus on the past decreases and focus on the future increases at around lunch break which can be attributed to release of dopamine after eating a tasty food (Graph 1.8 & 1.9). It might be a good idea to have controversial news be broken to public after lunch time to dampen the public outrage in the social networks as they have more optimistic view towards life.

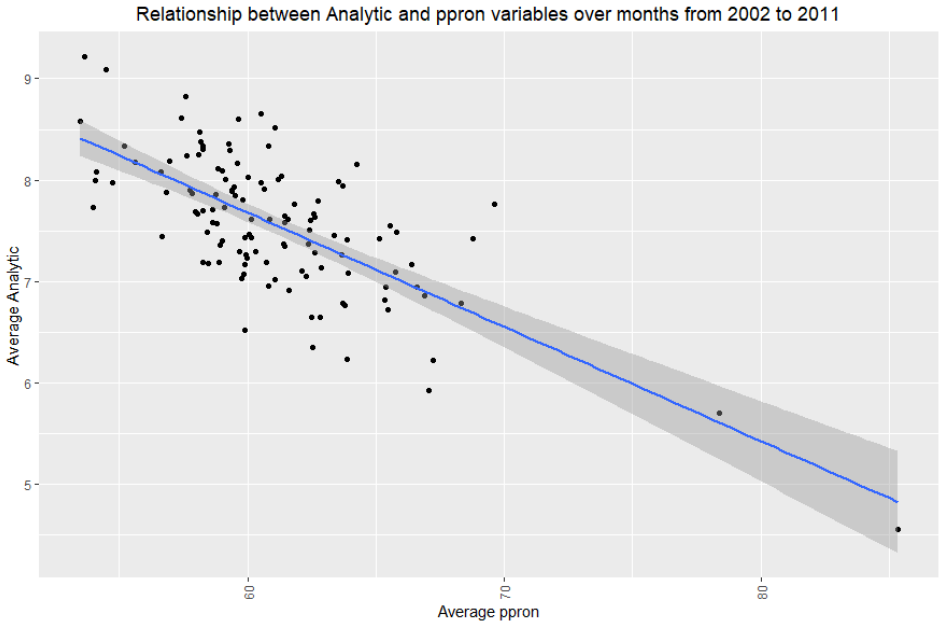


Graph (2.0)



Graph (2.1)

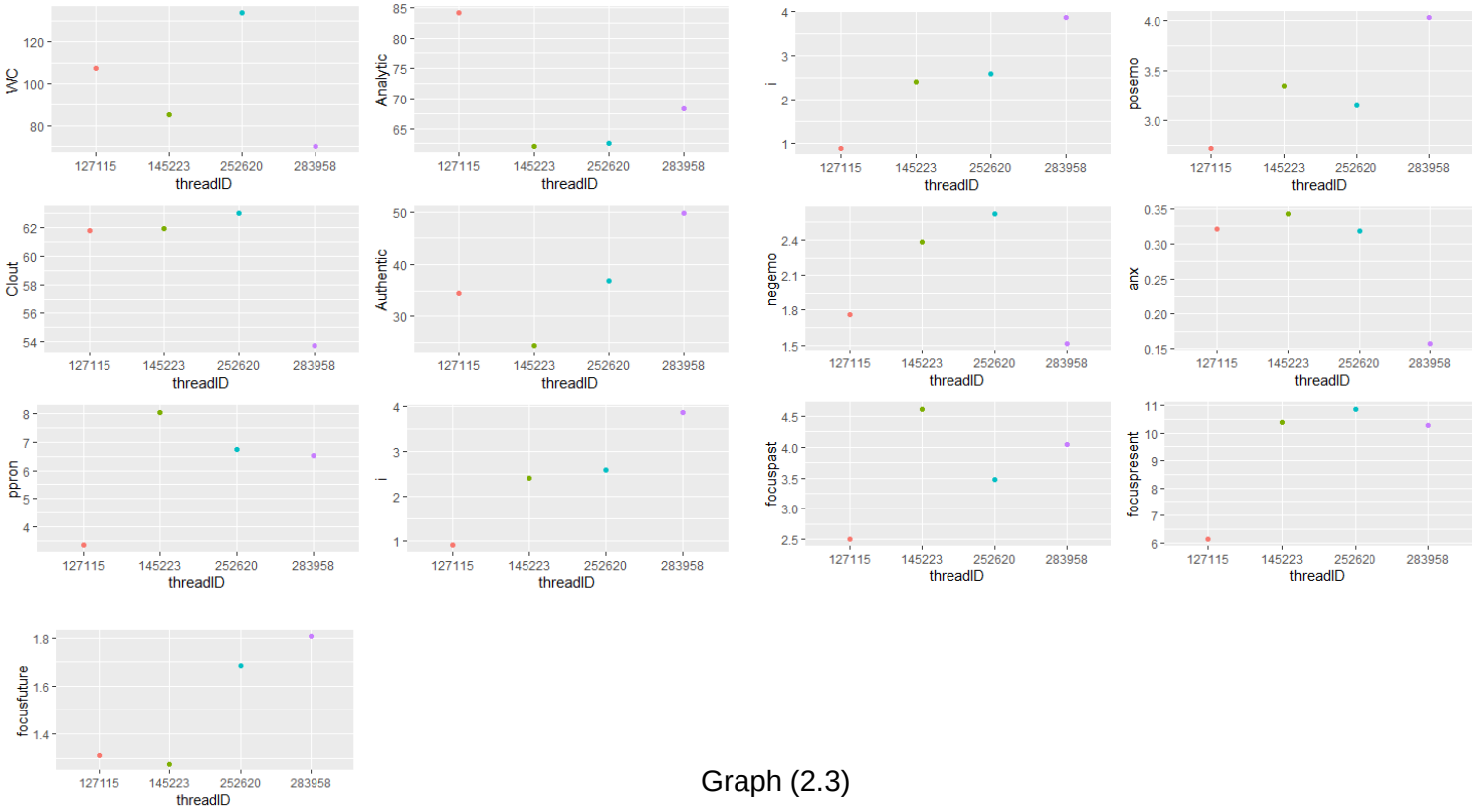
The emotional tone is positively correlated with positive emotions (posemo) and negatively correlated with negative emotions (negemo) (Graph (2.0) & Graph (2.1)). Also, it can be observed that linear relationship between Tone and posemo is much stronger than that of negemo.



Graph (2.2)

The increase in use of personal pronouns (ppron) decreases the use of Analytical language in posts (Graph 2.2) which is expected as personal pronouns are used mainly in Blogs, Expressive writing, and natural speech.

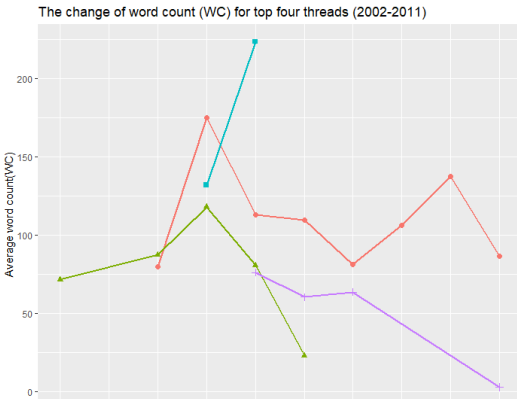
Since we already have established that activities are concentrated in the small number of threads at a time, we will look at the top 4 threads in terms of number of posts posted to gain insight regarding the relationship of linguistic variables and their levels over the ten-year period.

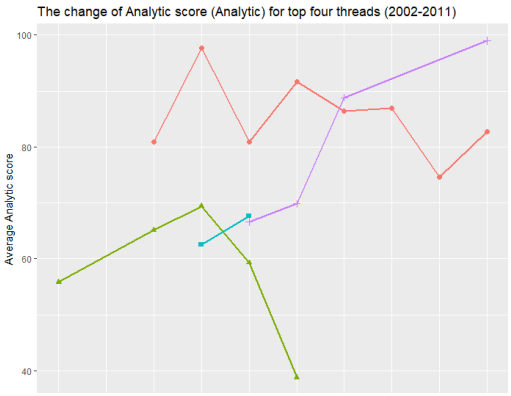


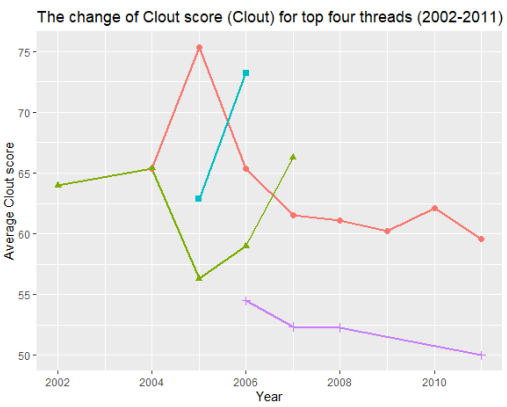
Graph (2.3)

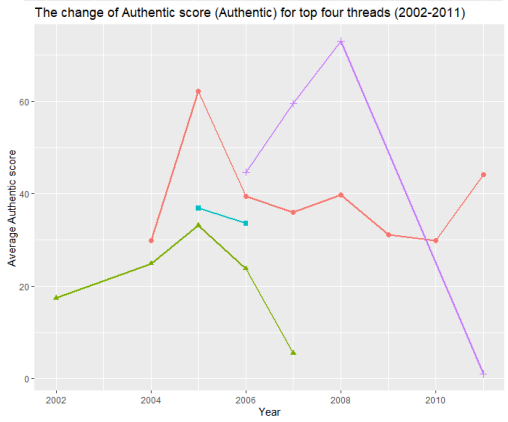
The words we use in our daily communications enables us to share our inner thoughts and desires. For instance, we can express sadness by using a range of words that are associated with feeling of sadness. In fact, by analysing the ratio of words belonging to each mental state category like happiness, sadness, anger, etc. can paint a general picture of what a message is about. A thread is also made of a collection of words contributed by multiple of authors which by analysing their occurrences through the thread give us an overall picture of what that thread is about. Although we use structure and tone of voice to decrypt a lot of messages which is part of evolving Natural language processing (NLP) field, analysing these statistics still provides an acceptable estimation of posts' message.

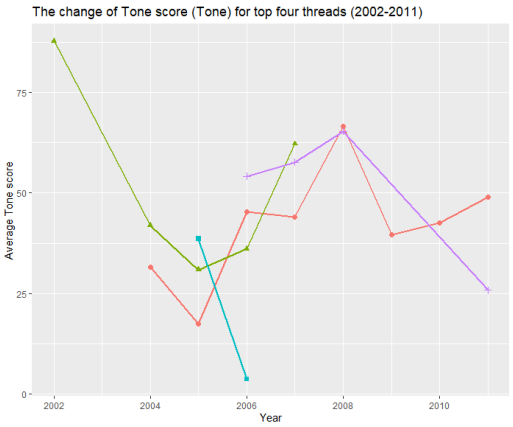
Levels of Linguistic variables give a signature of the thread. For instance, Thread 127115 (graph 2.3) has moderate average word count (WC) for its posts and low use of personal pronouns and "i" which coupled with high use of analytical words make this thread about business or money management. On there hand, threads 145223 and 252620 have close linguistic variables which can be clustered together. They both are more focused on past and present with low use of analytic language and authentic words. Moreover, they contain high levels of anxiety which makes them with high probability threads about topic that cause anxiety and negative emotions like police brutality and racial tensions. Finally, we look at the thread 283958 which has the highest "i" value which makes it more authentic and about topics that which focus on positive personal experiences like happiest memory in life which in turn lacks in clout.

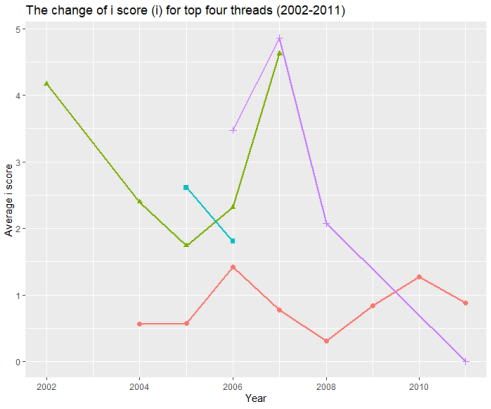


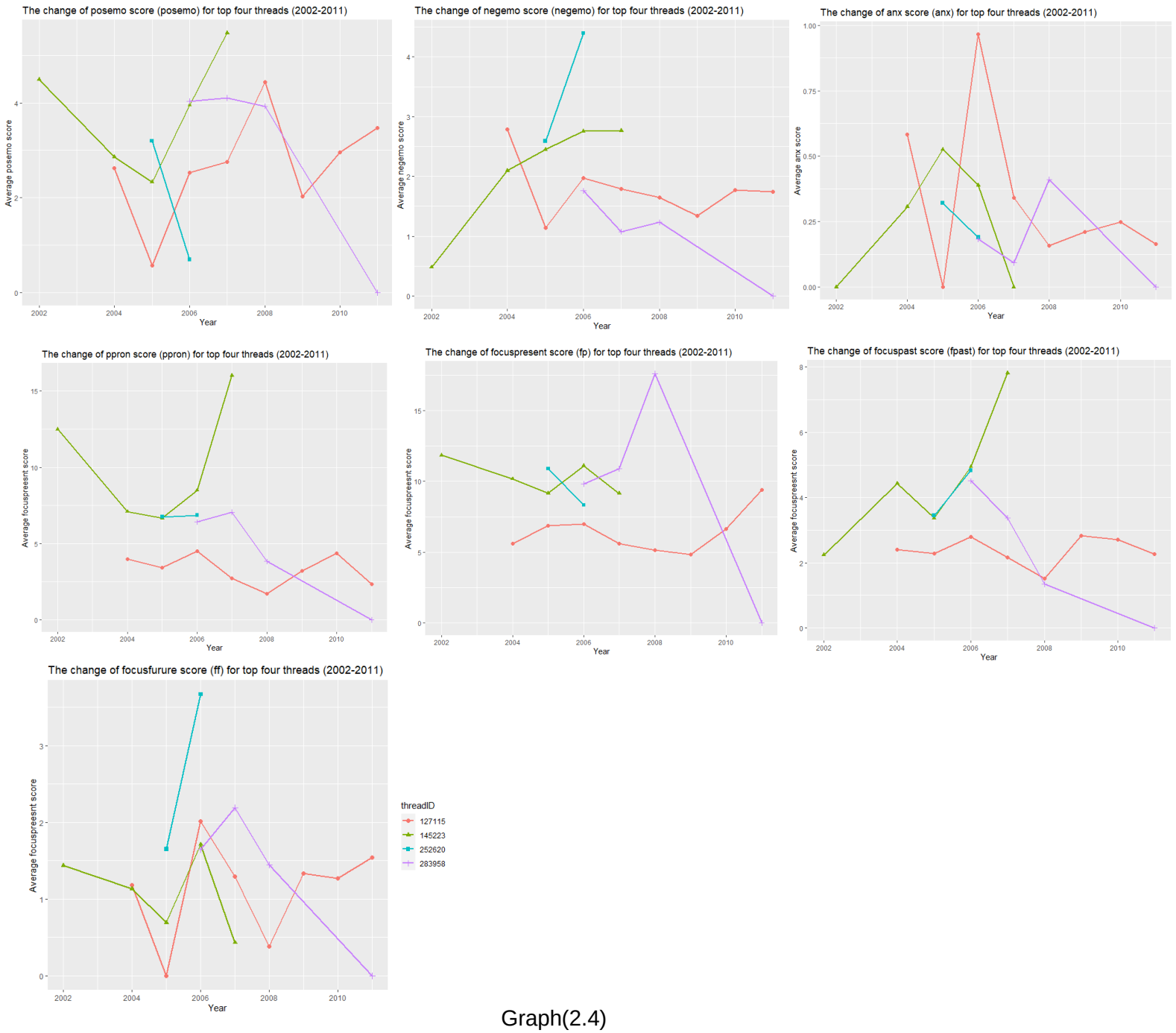












Graph(2.4)

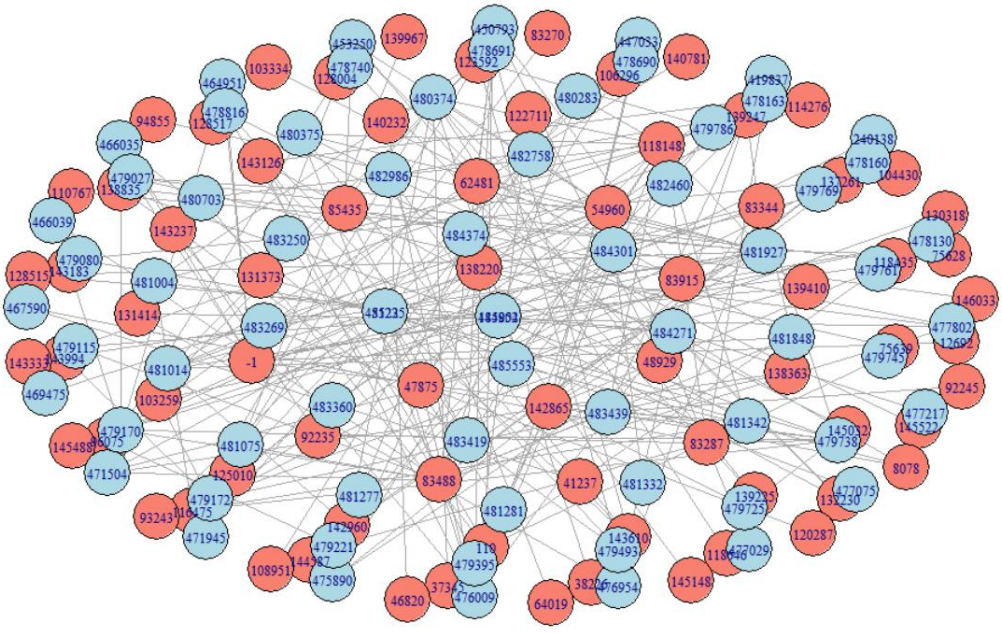
The word count (WC) in thread 145223 (gold line) increase gradually from 2002 and peaks at around 2005 then it plummets and go even lower than where it started (graph 2.4). Some other interesting metric about this thread is that the posemeo dropped sharply in this period while negemo and anx increased significantly. This thread is about topics that probably caused public backlash and outrage like start of a new war like aftermath of 2001. Also, we can see that use of analytic language increased until 2005 which was around the time United states was deeply entrenched into wars in the middle east and it can be explained by the fact that people tried to rationalise and separate facts form fearmongering. However, this upward trend subdued while use of emotional language became prominent in this thread which coincides with the time when public

opinion has become overwhelmingly polarised in the United States around this issue (ROSENTIEL, 2008). Also, focus past increased in this thread over time which can be attributed to people drawing comparison between Vietnam war and the war on terror. Interestingly, towards the end of this threads' activities over span of 2002 to 2007 the attention shifts more on the future than past which combined with sway towards more positive tone and decrease in the overall activity in the forum suggests that those who were against the war lost interest in the forum and stop participating and those who were for the war were overrepresented and showed positive attitude toward what was called the end of war by media.

Another pivotal event in this ten-year period was global financial crisis (GFC) of 2008 which we can further analyse by studying the two most active threads which were active around that time 283958 and 127115 (graph 2.4). As we have already established that 283958 is about topics which focus on positive personal experiences like happiest memory in life and lacks in clout, but during financial crisis the language in this forum completely swayed towards more analytical one. Moreover, posemo also dived significantly during this period for this specific forum. This shows that participants become more rational as things become more uncertain or as Daniel Kahneman states that System 2 thinking which is more conscious and logical mode of thinking takes over system 1 which is more intuitive and requires little effort.

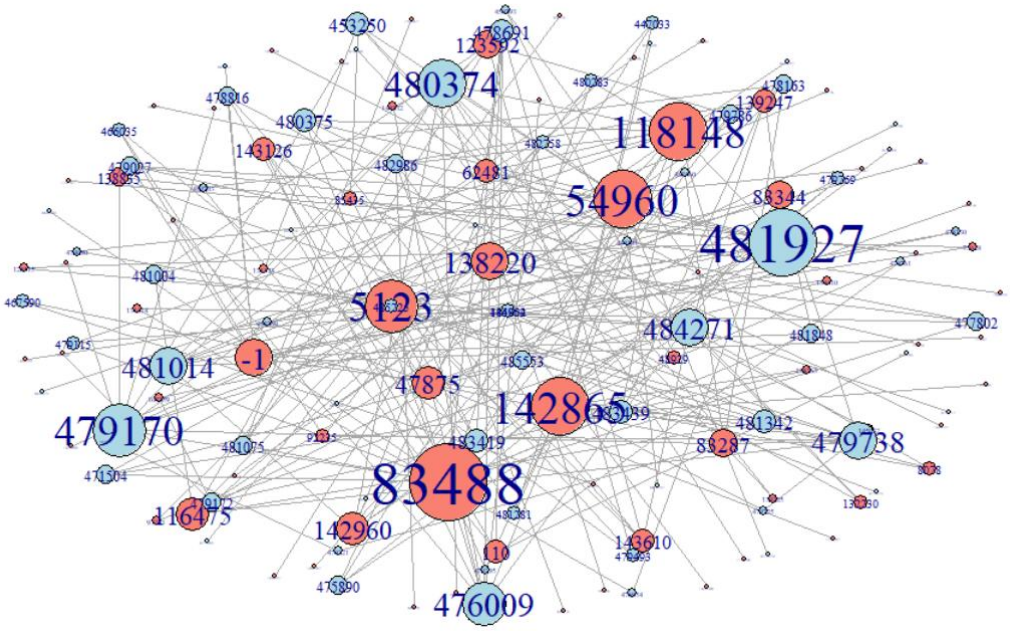
However, we cannot say the same for the thread 127115. One possible explanation is that this thread is for an age group who are not deeply concerned about the GFC like teenagers. In fact, majority of linguistic variables in this thread keep oscillating as if they are not affected by outside world events.

The Two Mode Network Of Authors Participating at April,2008



Graph 2.5: Red labels represents AuthorIDs and Lightblue labels represent the ThreadID

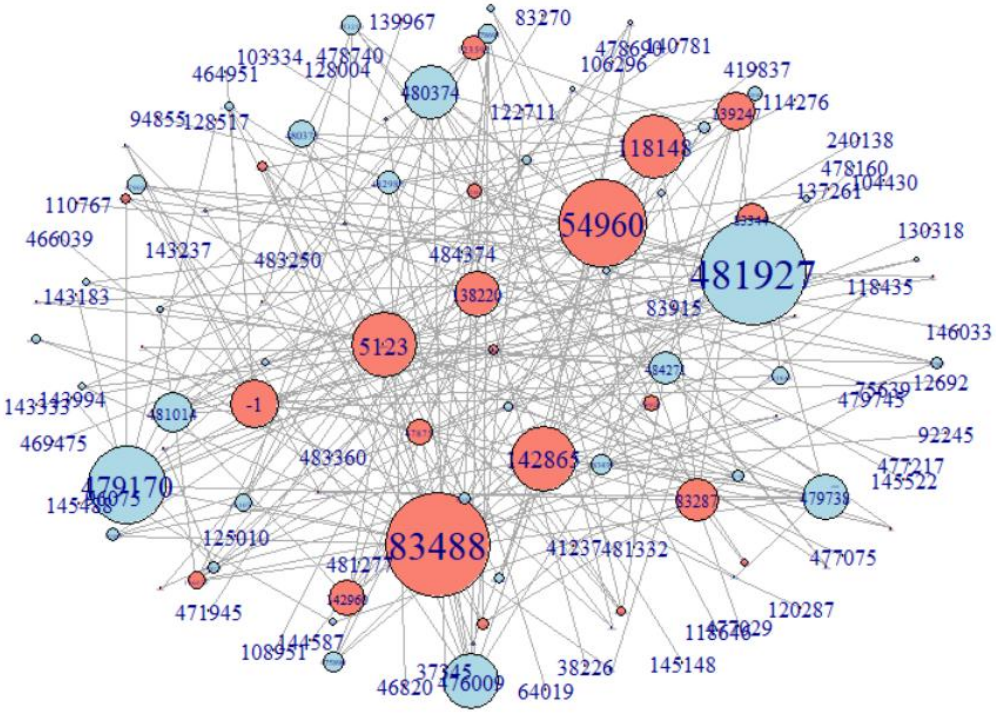
The Two Mode Network Of Authors Participating at April,2008



Graph 2.6: The size of labels correlates with degree of each vertix

(Note: red labels have been give higher coefficent to better stand out)

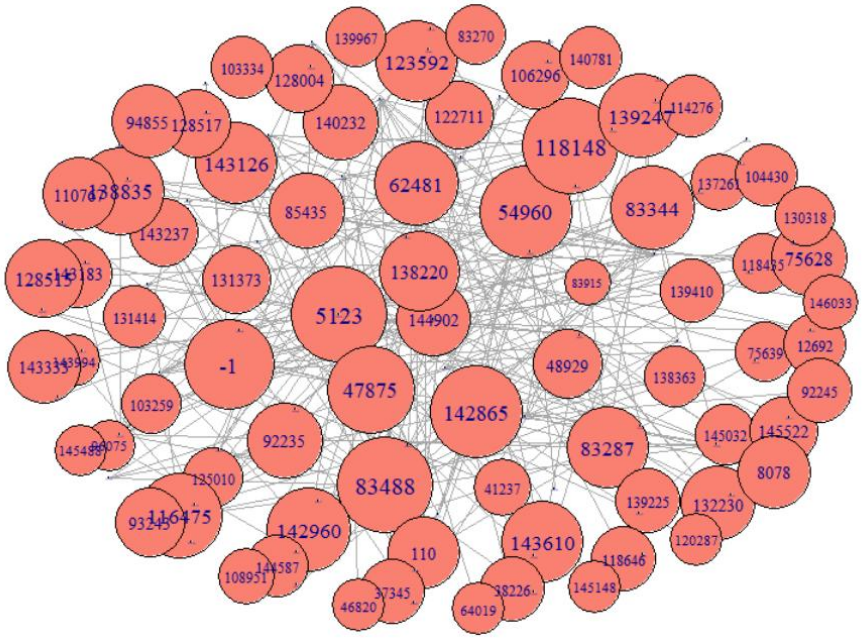
The Two Mode Network Of Authors Participating at April,2008



Graph 2.7: The size of labels correlates with Betweenness centrality of each vertix

(Note: red labels have been given higher coefficent to better stand out)

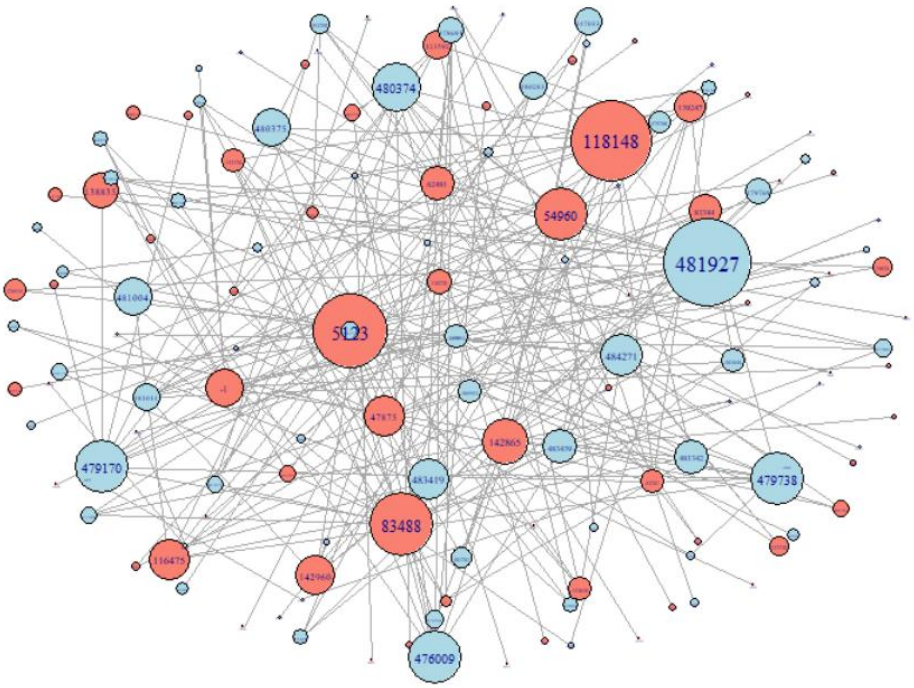
The Two Mode Network Of Authors Participating at April,2008



Graph 2.8: The size of labels correlates with Closeness centrality of each vertix

(Note: red labels represent Authors)

The Two Mode Network Of Authors Participating at April,2008

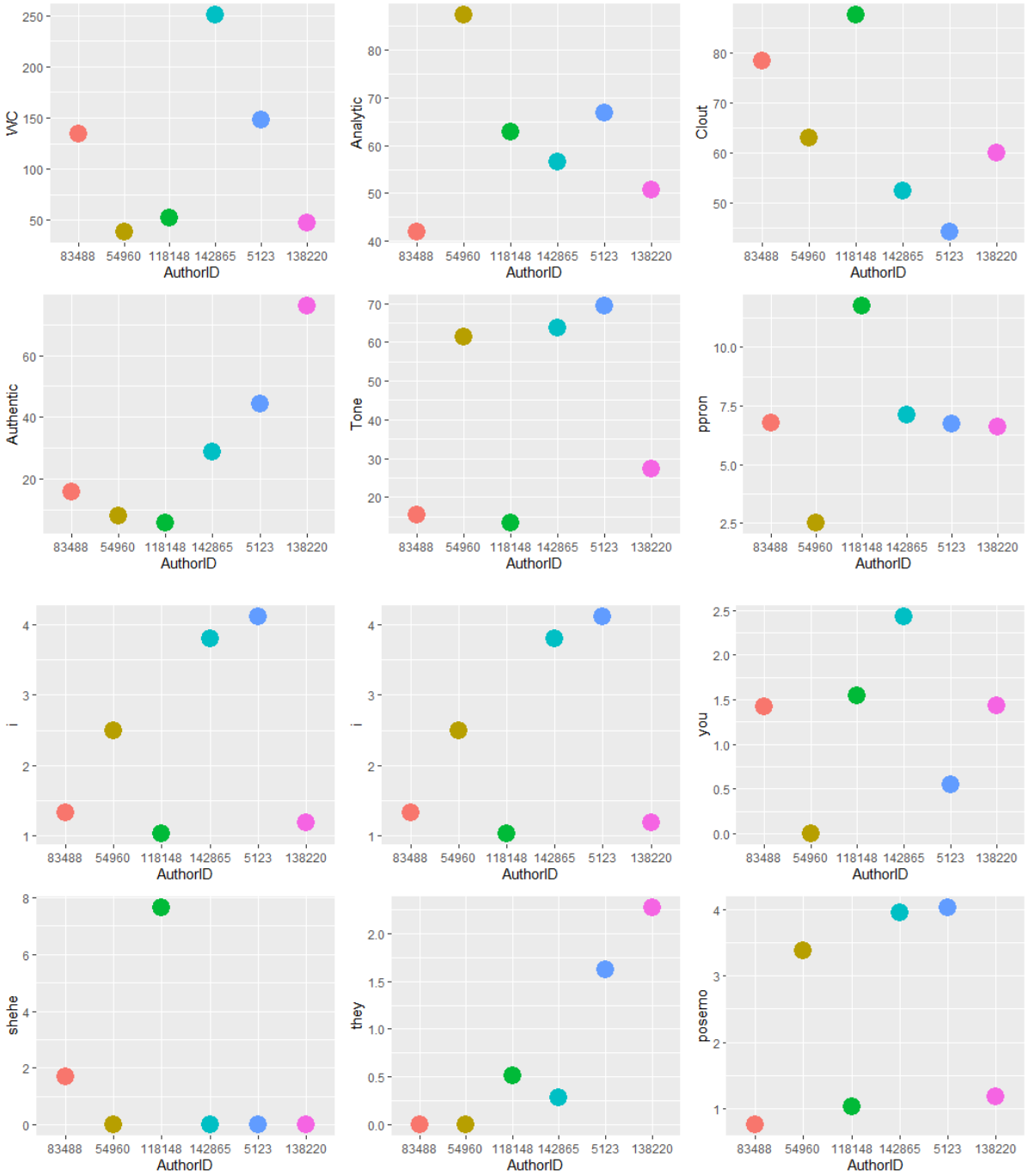


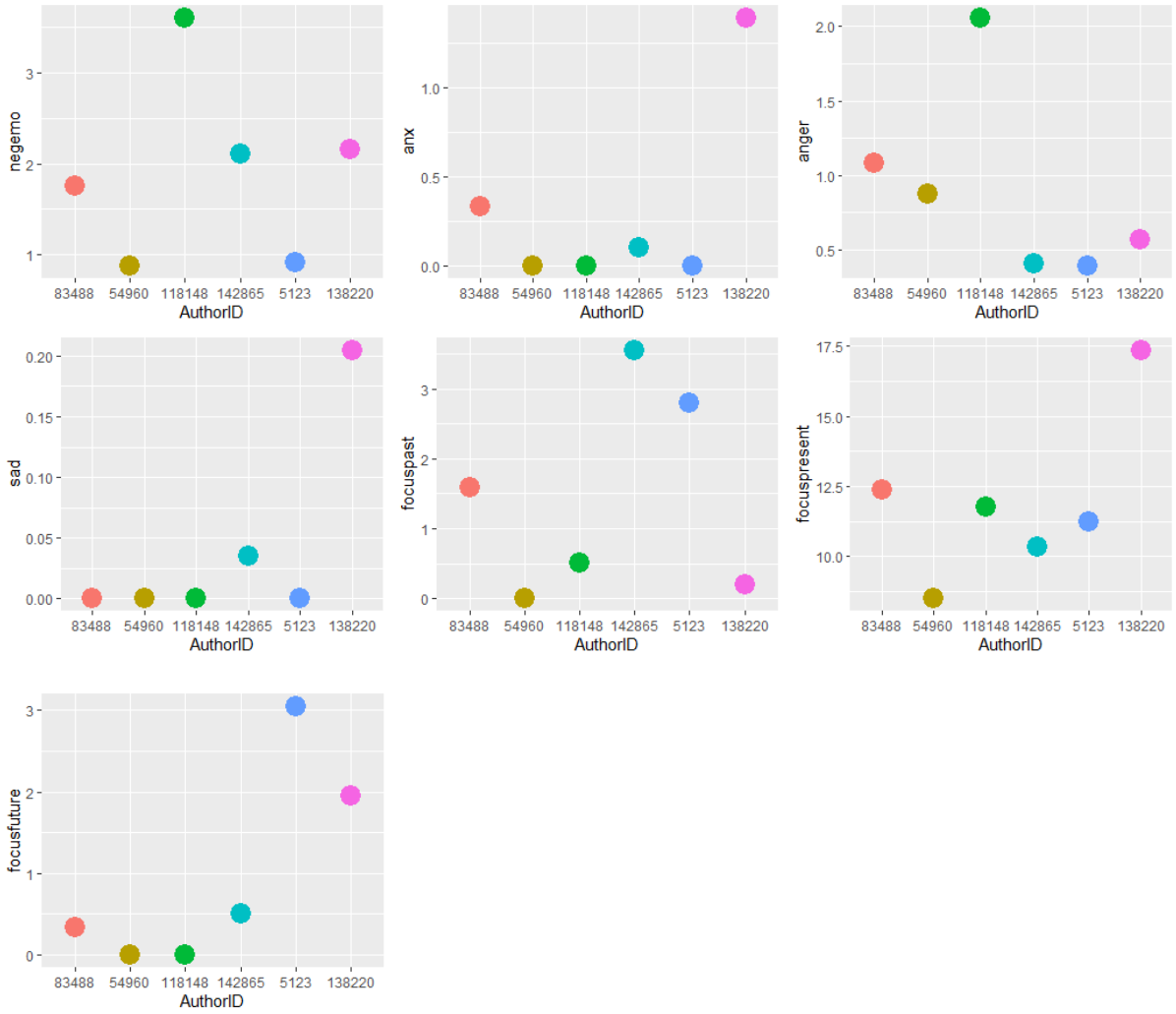
Graph 2.9: The size of labels correlates with Eigencentrality of each vertix

(Note: red labels represent Authors)

We can see from above graphs that author with ID 83488 is most influential if we take all the centrality mesures into account. The next five most influential authors are 54960, 118148, 142865, 5123, 138220 respectively. Now, we graph the level of language variables for each of the authors who are part of this network.

Average of language variables for top 6 most influential authors(April, 2008)





Graph (3.0)

Author 83488 has not posted especially long posts (~150 words per post) in comparison to less influential authors in this network (graph 3.0). This shows that posts from this author are concise and not too long to risk losing the readers' attention or interest. In fact, if we look carefully at rest of language variables for this author, a pattern emerges apart from few exceptions that the variables are somehow used in moderation. And, the language is not the most authentic, positive, or forceful one but a balanced level of all language variables.

This suggests that influential communicators use language variables in moderation and let the audience to gauge their emotional response towards what they have been represented to. Since audiences have their own worldview and way of making sense of ideas. This strategy becomes increasingly important as influencer wishes to appeal to larger audiences with diverse background.

References:

1. Alutaybi, A., Al-Thani, D., McAlaney, J., & Ali, R. (2020). Combating Fear of Missing Out (FoMO) on Social Media: The FoMO-R Method. International Journal Of Environmental Research And Public Health, 17(17), 6128. doi: 10.3390/ijerph17176128

2. Daniel Kahneman, D. (2022). System 1 and System 2 Thinking - The Decision Lab. The Decision Lab. Retrieved 6 May 2022, from https://thedecisionlab.com/reference-guide/philosophy/system-1-and-system-2-thinking.

3. ROSENTIEL, T. (2008). Public Attitudes Toward the War in Iraq: 2003-2008. Pew Research Center. Retrieved 20 April 2021, from https://www.pewresearch.org/2008/03/19/public-attitudes-toward-the-war-in-iraq-20032008/.

Appendix:

```
#library(ggplot2)


#options(ggplot2.continuous.colour="BuPu")


#rm(list = ls())
# set.seed(3084) # XXXXXXXX = your student ID
# webforum <- read.csv("webforum.csv")
# webforum <- webforum [sample(nrow(webforum), 20000), ] # 20000 rows


# store sample in new file
#write.csv(webforum, "webforum2.csv")


webforum = read.csv("webforum2.csv")


attach(webforum)


head(webforum)
```

```
# some basic statistics
summary(webforum)
str(webforum)

# investigate AuthorID -1
# the number of observations grouped by AuthorID
Author_count = as.data.frame(as.table(by(webforum, AuthorID, nrow)))

# sort the Author_count
Author_count = Author_count[order(Author_count$Freq, decreasing = TRUE),]

# we plot the first 10 rows (Graph 1.1)
suspicious_author = Author_count[1:10, ]

suspicious_author = suspicious_author[order(suspicious_author$Freq,decreasing =
TRUE),]
qplot(AuthorID, Freq, data = suspicious_author, color = AuthorID, main = "Top ten
most active participents", cex=0.2)+
  theme(plot.title = element_text(hjust = 0.5))



# investigate the activity of participants over time
str(Date) # Date column is of type character

# convert Date column from Character class to Date
Date = as.Date(Date,tryFormats = c("%Y-%m-%d", "%Y/%m/%d"))

# extract year from Date column
webforum$year = as.numeric(format(Date,'%Y'))
attach(webforum)



# group data frame by year column
```

```
posts_year= as.data.frame(as.table(by(webforum, year, nrow)))


# plot the posts_year(Graph 1.2)

qplot(year, Freq,data = posts_year, main = "Number of posts per year(2002-2011)", xlab = "Year", ylab ="Frequency", color = year) + theme(plot.title = element_text(hjust = 0.5))



# now we group based on year and month

# extract month from Date column

webforum$month = as.numeric(format(Date,'%m'))

attach(webforum)


# convert to character year and month column

webforum$year = as.character(format(Date,'%Y'))

webforum$month = as.character(format(Date,'%m'))


# concatenate year and month column

webforum = transform(webforum,yearmonth=paste0(year, month))


# convert to character year and month column

webforum$year = as.character(format(Date,'%Y'))

webforum$month = as.character(format(Date,'%m'))


# concatenate year and month column

webforum = transform(webforum,yearmonth=paste0(year, month))


# Group webforum by yearmonth column

posts_year_month = as.data.frame(as.table(by(webforum, webforum$yearmonth, nrow)))


# change the name of columns

colnames(posts_year_month) = c("YearMonth", "Freq")
```

```
posts_year_month$year = substr(posts_year_month$YearMonth, start = 1, stop = 4)


# we plot (Graph 1.3)

qplot(YearMonth, Freq,data = posts_year_month, xlab = "YearMonth",
ylab="Frequency", main = "Monthly number of posts (2002-2011)", color = year) +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ theme(plot.title =
element_text(hjust = 0.5))



# creating time series (Graph 1.4)

posts_time = as.data.frame(as.table(by(webforum, webforum$Date, nrow)))

series = ts(posts_time[,2], frequency = 12, start = c(2002), end = c(2011))

plot(series)


# decompose the series (Graph 1.5)

decomposed_series = decompose(series)

plot(decomposed_series)



# Group Author_count based on Freq

AuthorID_dist=as.data.frame(as.table(by(Author_count, Author_count$Freq, nrow)))


# change columns to numeric values

AuthorID_dist$Author_count.Freq = as.numeric(AuthorID_dist$Author_count.Freq)

AuthorID_dist$Freq = as.numeric(AuthorID_dist$Freq)


#Change column names

colnames(AuthorID_dist) = c("posts", "freq")


# number of Authors

num_Authors = length(unique(webforum$AuthorID))


# add a ratio column(freq /num_Authors)
```

```
AuthorID_dist$ratio = AuthorID_dist$freq/num_Authors


# get rid of outliers

AuthorID_dist = AuthorID_dist[order(AuthorID_dist$ratio, decreasing = TRUE),]

AuthorID_dist = AuthorID_dist[AuthorID_dist$ratio >= 0.001,]


# plot the whole thing (Graph 1.6)

qplot(AuthorID_dist$posts, AuthorID_dist$ratio, xlab = "Number of posts", ylab =
"Ratio(authors/all authors)", main = "Distribution of authors and number of
posts(2002-2011)")+ theme(plot.title = element_text(hjust = 0.5))+
geom_point(aes(size=AuthorID_dist$ratio*100, color =
as.factor(AuthorID_dist$posts))) + labs(size="Ratio of authors(percent)",
color="Number of posts")+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5,
hjust=1))


# investigating ThreadID and AuthorID correlation

# get rid of non-numeric columns

numeric_webforum = webforum[, c(-4,-5)]


#convert non-numeric columns to numeric

numeric_webforum$WC = as.numeric(numeric_webforum$WC)

numeric_webforum$year = as.numeric(numeric_webforum$year)

numeric_webforum$month = as.numeric(numeric_webforum$month)

numeric_webforum$X = as.numeric(numeric_webforum$X)

numeric_webforum$yearmonth = as.numeric(numeric_webforum$yearmonth)



library(lattice)

# Graph 1.7

xyplot(AuthorID ~ ThreadID | as.character(year), data = numeric_webforum, main =
"Relationship between Threads and Authors(2002-2011)", ylab = "Author ID", xlab =
"Thread ID")
```

```
# change of focuspast and future during the day

#extract hour from time

webforum$hour = substr(Time, start = 1, stop = 2)

webforum$hour = as.numeric(webforum$hour)

time_focusfuture = as.data.frame(as.table(by(webforum$focusfuture, webforum$hour, mean)))

colnames(time_focusfuture) = c("hour", "Freq")


# Graph (1.8)

qplot(hour,Freq,data =time_focusfuture, ylab= "Percent", xlab = "Hour", main = "Expressing a focus (average) on the future in posts during day (2002-2011)", color = "green") + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ theme(plot.title = element_text(hjust = 0.5)) + scale_color_brewer(palette = "Spectral")


time_focuspast = as.data.frame(as.table(by(webforum$focuspast, webforum$hour, mean)))

colnames(time_focuspast) = c("hour", "Freq")


# Graph (1.9)

qplot(hour,Freq,data =time_focuspast, ylab= "Percent", xlab = "Hour", main = "Expressing a focus (average) on the past in posts during day (2002-2011)", color = "green") + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ theme(plot.title = element_text(hjust = 0.5))




# correlation matrix

library(reshape2)

corr = melt(round(cor(numeric_webforum), digits = 3))


corr$Y1 <- cut(corr$value, breaks = c(-Inf, -1, -0.48, 0, 0.48, 1, Inf))


g = ggplot(data = corr, aes(x=Var1, y=Var2, fill=value))

g = g + geom_tile(aes(fill = Y1))+ scale_fill_manual(breaks=c("(-Inf,-1]", "(-1,-0.48]", "(-0.48,0]", "(0,0.48]", "(0.48, 1]", "(1,Inf)"), values = c("red", "pink", "lightblue",
```

```
"darkblue", "orange", "white")) + theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))

g




# investigate posemo and Tone

# add yearmonth column from webforum to numeric_webforum

numeric_webforum$yearmonth = webforum$yearmonth




# average of Tone per month ( 120 obs)

cor_tone = as.data.frame(as.table(by(numeric_webforum$Tone,numeric_webforum$yearmonth, mean)))


#change column names

colnames(cor_tone) = c("year", "freqtone")


# extract year from year month

cor_tone$year = substr(cor_tone$year, start = 1, stop = 4)


# convert year to numeric variable

cor_tone$year = as.numeric(cor_tone$year)




# average of posemo per month ( 120 obs)

cor_pemo= as.data.frame(as.table(by(numeric_webforum$posemo,numeric_webforum$yearmonth , mean)))

colnames(cor_pemo) = c("year", "freqpemo")




# extract year from year month
```

```
cor_pemo$year = substr(cor_pemo$year, start = 1, stop = 4)

cor_pemo$year = as.numeric(cor_pemo$year)


# combine the two data frame

cor_tone_pemo = cbind(cor_tone, cor_pemo)

cor_tone_pemo[,3]= NULL


#plot the whole thing, Graph (2.0)

qplot(freqtone, freqpemo, data= cor_tone_pemo, main = "Relationship between
posemo and Tone variables over months from 2002 to 2011" , xlab = "Average Tone",
ylab = "Average posemo")+ geom_smooth(method='lm')+ theme(axis.text.x =
element_text(angle = 90, vjust = 0.5, hjust=1))+ theme(plot.title = element_text(hjust =
0.5))



# investigate negemo and Tone

# average of negemo per month ( 120 obs)

cor_negemo =
as.data.frame(as.table(by(numeric_webforum$negemo,numeric_webforum$yearmont
h , mean)))

colnames(cor_negemo) = c("year", "freqnegemo")


# extract year from year month

cor_negemo$year = substr(cor_negemo$year, start = 1, stop = 4)

cor_negemo$year = as.numeric(cor_negemo$year)


# combine the two data frame

cor_tone_negemo = cbind(cor_tone, cor_negemo)

cor_tone_negemo[,3]= NULL


#plot the whole thing, Graph (2.1)

qplot(freqtone, freqnegemo, data= cor_tone_negemo, main = "Relationship between
negemo and Tone variables over months from 2002 to 2011" , xlab = "Average Tone",
ylab = "Average negemo")+
```

```
  geom_smooth(method='lm')+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ theme(plot.title = element_text(hjust = 0.5))


# investigate ppron and Analytic


# average of ppron per month ( 120 obs)

cor_ppron = as.data.frame(as.table(by(numeric_webforum$ppron,numeric_webforum$yearmonth, mean)))


#change column names

colnames(cor_ppron) = c("year", "freqppron")


# extract year from year month

cor_ppron$year = substr(cor_ppron$year, start = 1, stop = 4)


# convert year to numeric variable

cor_ppron$year = as.numeric(cor_ppron$year)




# average of Analytic per month ( 120 obs)

cor_analytic= as.data.frame(as.table(by(numeric_webforum$Analytic,numeric_webforum$yearmonth , mean)))

colnames(cor_analytic) = c("year", "freqanalytic")




# extract year from year month

cor_analytic$year = substr(cor_analytic$year, start = 1, stop = 4)

cor_analytic$year = as.numeric(cor_analytic$year)


# combine the two data frame

cor_analytic_ppron = cbind(cor_ppron, cor_analytic)

cor_analytic_ppron[,3]= NULL
```

```
#plot the whole thing, Graph (2.2)

qplot(freqanalytic, freqppron, data= cor_analytic_ppron, main = "Relationship between Analytic and ppron variables over months from 2002 to 2011" , xlab = "Average ppron", ylab = "Average Analytic")+ geom_smooth(method='lm')+ theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))+ theme(plot.title = element_text(hjust = 0.5))




#Graph (2.3)

# group webforum by threadId


group_thread = as.data.frame(as.table(by(webforum,as.integer(webforum$ThreadID) , nrow)))


colnames(group_thread) = c("threadID", "freq")


group_thread$threadID = as.character(group_thread$threadID)


group_thread$threadID = as.numeric(group_thread$threadID)


# get ThreadIDs with more than one hundred posts


big_threads = group_thread[group_thread$freq >100, ]

big_threads$threadID = as.numeric(big_threads$threadID) # [127115, 145223, 252620, 283958]


# extract the observation with these threadIDs

main_group = webforum[webforum$ThreadID %in% big_threads$threadID,]


# threadID = 127115

first_group = main_group[main_group$ThreadID == 127115, ]
```

```
first_mean = as.data.frame(sapply(first_group[,6:24], mean))

first_mean =t(first_mean)

rownames(first_mean) = NULL

colnames(first_mean) = c("WC", "Analytic", "Clout", "Authentic", "Tone", "ppron", "i",
"we", "you", "shehe", "they", "posemo", "negemo", "anx",
                         "anger", "sad", "focuspast", "focuspresent", "focusfuture")




# summary(first_group)
#
# str(first_group)
#
# plot(first_group$WC~first_group$AuthorID )

# threadID = 145223

second_group = main_group[main_group$ThreadID == 145223, ]

second_mean = as.data.frame(sapply(second_group[,6:24], mean))

second_mean =t(second_mean)

rownames(second_mean) = NULL

colnames(second_mean) = c("WC", "Analytic", "Clout", "Authentic", "Tone", "ppron",
"i", "we", "you", "shehe", "they", "posemo", "negemo", "anx",
                          "anger", "sad", "focuspast", "focuspresent", "focusfuture")
```

```
# threadID = 252620
third_group = main_group[main_group$ThreadID == 252620, ]

third_mean = as.data.frame(sapply(third_group[,6:24], mean))

third_mean =t(third_mean)

rownames(third_mean) = NULL

colnames(third_mean) = c("WC", "Analytic", "Clout", "Authentic", "Tone", "ppron",
"i", "we", "you", "shehe", "they", "posemo", "negemo", "anx",
                         "anger", "sad", "focuspast", "focuspresent", "focusfuture")




# threadID = 283958
fourth_group = main_group[main_group$ThreadID == 283958, ]

fourth_mean = as.data.frame(sapply(fourth_group[,6:24], mean))

fourth_mean =t(fourth_mean)

rownames(fourth_mean) = NULL

colnames(fourth_mean) = c("WC", "Analytic", "Clout", "Authentic", "Tone", "ppron",
"i", "we", "you", "shehe", "they", "posemo", "negemo", "anx",
                          "anger", "sad", "focuspast", "focuspresent", "focusfuture")



# combine all the means
```

```
mean_groups = rbind(first_mean, second_mean, third_mean, fourth_mean)

mean_groups <- cbind(mean_groups, data.frame(threadID = c(127115, 145223, 252620,
283958)))


#mean_groups = mean_groups[, c(1:3, 5:6, 10:13, 16:18)]

#plot the whole thing

#install.packages("gridExtra")

library(gridExtra)


mean_groups$threadID = as.factor(mean_groups$threadID)

# Graph 2.3

# + scale_color_brewer(palette = "Spectral")

plot1 = qplot(threadID, WC, data = mean_groups, color = threadID) +
theme(legend.position = "none")

plot2 = qplot(threadID, Analytic, data = mean_groups, color = threadID)+
theme(legend.position = "none")

plot3 = qplot(threadID, Clout, data = mean_groups, color = threadID)+
theme(legend.position = "none")

plot4 = qplot(threadID, Authentic, data = mean_groups, color = threadID)+
theme(legend.position = "none")

# plot5 = qplot(threadID, Tone, data = mean_groups, color = threadID )+
theme(legend.position = "none")

plot6 = qplot(threadID, ppron, data = mean_groups, color = threadID)+
theme(legend.position = "none")

plot7 = qplot(threadID,i, data = mean_groups, color = threadID) +
theme(legend.position = "none")


grid.arrange(plot1, plot2, plot3, plot4,plot6, plot7, nrow=3, ncol = 2)




# plot8 = qplot(threadID, we, data = mean_groups, color = threadID)+
theme(legend.position = "none")

# plot9 = qplot(threadID, you, data = mean_groups, color = threadID)+
theme(legend.position = "none")

# plot10 = qplot(threadID, they, data = mean_groups, color = threadID)+
theme(legend.position = "none")
```

```
plot11 = qplot(threadID, posemo, data = mean_groups, color = threadID )+
theme(legend.position = "none")

plot12 = qplot(threadID, negemo, data = mean_groups, color = threadID)+
theme(legend.position = "none")



plot13 = qplot(threadID, anx , data = mean_groups, color = threadID)+
theme(legend.position = "none")

# plot14 = qplot(threadID, anger, data = mean_groups, color = threadID )+
theme(legend.position = "none")

# plot15 = qplot(threadID, sad, data = mean_groups, color = threadID)+
theme(legend.position = "none")

plot16 = qplot(threadID, focuspast, data = mean_groups, color = threadID )+
theme(legend.position = "none")

plot17 = qplot(threadID, focuspresent, data = mean_groups, color = threadID )+
theme(legend.position = "none")

plot18 = qplot(threadID, focusfuture, data = mean_groups, color = threadID )+
theme(legend.position = "none")

grid.arrange(plot7,plot11, plot12, plot13, plot16, plot17, nrow=3, ncol = 2)

grid.arrange(plot18, nrow=3, ncol =2)



# Graph 2.4

# change of language over time

# change of wc

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_wc= as.data.frame(as.table(by(first_group$WC,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_wc) = c("threadID","year", "freqwc")


# 145223
```

```
second_wc=
as.data.frame(as.table(by(second_group$WC,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_wc) = c("threadID","year", "freqwc")


# 252620

third_wc= as.data.frame(as.table(by(third_group$WC,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_wc) = c("threadID","year", "freqwc")


# 283958

fourth_wc= as.data.frame(as.table(by(fourth_group$WC,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_wc) = c("threadID","year", "freqwc")




# combine all the data frames

year_wc = rbind(first_wc, second_wc, third_wc, fourth_wc)

year_wc$year = as.numeric(as.character(year_wc$year))

wc_plot = ggplot(year_wc, aes(year, freqwc, color = threadID))

wc_plot = wc_plot + xlab ("Year") + ylab ("Average word count(WC)") + ggtitle("The
change of word count (WC) for top four threads (2002-2011)")

wc_plot = wc_plot + geom_line(size=1)

wc_plot = wc_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2) +
theme(legend.position = "none")

wc_plot




# change of Analytic

# threadID = c(127115, 145223, 252620, 283958)


# 127115
```

```
first_Analytic=
as.data.frame(as.table(by(first_group$Analytic,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_Analytic) = c("threadID","year", "freqAnalytic")


# 145223

second_Analytic=
as.data.frame(as.table(by(second_group$Analytic,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_Analytic) = c("threadID","year", "freqAnalytic")


# 252620

third_Analytic=
as.data.frame(as.table(by(third_group$Analytic,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_Analytic) = c("threadID","year", "freqAnalytic")


# 283958

fourth_Analytic=
as.data.frame(as.table(by(fourth_group$Analytic,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_Analytic) = c("threadID","year", "freqAnalytic")




# combine all the data frames

year_Analytic = rbind(first_Analytic, second_Analytic, third_Analytic, fourth_Analytic)

year_Analytic$year = as.numeric(as.character(year_Analytic$year))

Analytic_plot = ggplot(year_Analytic, aes(year, freqAnalytic, color = threadID))

Analytic_plot = Analytic_plot + xlab ("Year") + ylab ("Average Analytic score") +
ggtitle("The change of Analytic score (Analytic) for top four threads (2002-2011)")

Analytic_plot = Analytic_plot + geom_line(size=1)

Analytic_plot = Analytic_plot + geom_point(aes(colour = threadID, shape = threadID),
size = 2) + theme(legend.position = "none")

Analytic_plot
```

```
# change of Clout

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_Clout= as.data.frame(as.table(by(first_group$Clout,list(first_group$ThreadID, first_group$year), mean)))

colnames(first_Clout) = c("threadID","year", "freqClout")


# 145223

second_Clout= as.data.frame(as.table(by(second_group$Clout,list(second_group$ThreadID, second_group$year), mean)))

colnames(second_Clout) = c("threadID","year", "freqClout")


# 252620

third_Clout= as.data.frame(as.table(by(third_group$Clout,list(third_group$ThreadID, third_group$year), mean)))

colnames(third_Clout) = c("threadID","year", "freqClout")


# 283958

fourth_Clout= as.data.frame(as.table(by(fourth_group$Clout,list(fourth_group$ThreadID, fourth_group$year), mean)))

colnames(fourth_Clout) = c("threadID","year", "freqClout")


# combine all the data frames

year_Clout = rbind(first_Clout, second_Clout, third_Clout, fourth_Clout)

year_Clout$year = as.numeric(as.character(year_Clout$year))

Clout_plot = ggplot(year_Clout, aes(year, freqClout, color = threadID))

Clout_plot = Clout_plot + xlab ("Year") + ylab ("Average Clout score") + ggtitle("The change of Clout score (Clout) for top four threads (2002-2011)")

Clout_plot = Clout_plot + geom_line(size=1)

Clout_plot = Clout_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2)+ theme(legend.position = "none")
```

Clout_plot

 # change of Authentic

 # threadID = c(127115, 145223, 252620, 283958)

 # 127115

first_Authentic=
as.data.frame(as.table(by(first_group$Authentic,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_Authentic) = c("threadID","year", "freqAuthentic")

 # 145223

second_Authentic=
as.data.frame(as.table(by(second_group$Authentic,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_Authentic) = c("threadID","year", "freqAuthentic")

 # 252620

third_Authentic=
as.data.frame(as.table(by(third_group$Authentic,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_Authentic) = c("threadID","year", "freqAuthentic")

 # 283958

fourth_Authentic=
as.data.frame(as.table(by(fourth_group$Authentic,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_Authentic) = c("threadID","year", "freqAuthentic")

 # combine all the data frames

year_Authentic = rbind(first_Authentic, second_Authentic, third_Authentic,
fourth_Authentic)

```
year_Authentic$year = as.numeric(as.character(year_Authentic$year))

Authentic_plot = ggplot(year_Authentic, aes(year, freqAuthentic, color = threadID))

Authentic_plot = Authentic_plot + xlab ("Year") + ylab ("Average Authentic score") +
ggtitle("The change of Authentic score (Authentic) for top four threads (2002-2011)")

Authentic_plot = Authentic_plot + geom_line(size=1)

Authentic_plot = Authentic_plot + geom_point(aes(colour = threadID, shape =
threadID), size = 2) + theme(legend.position = "none")

Authentic_plot


# change of Tone

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_Tone= as.data.frame(as.table(by(first_group$Tone,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_Tone) = c("threadID","year", "freqTone")


# 145223

second_Tone=
as.data.frame(as.table(by(second_group$Tone,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_Tone) = c("threadID","year", "freqTone")


# 252620

third_Tone= as.data.frame(as.table(by(third_group$Tone,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_Tone) = c("threadID","year", "freqTone")


# 283958

fourth_Tone=
as.data.frame(as.table(by(fourth_group$Tone,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_Tone) = c("threadID","year", "freqTone")
```

```
# combine all the data frames

year_Tone = rbind(first_Tone, second_Tone, third_Tone, fourth_Tone)

year_Tone$year = as.numeric(as.character(year_Tone$year))

Tone_plot = ggplot(year_Tone, aes(year, freqTone, color = threadID))

Tone_plot = Tone_plot + xlab ("Year") + ylab ("Average Tone score") + ggtitle("The change of Tone score (Tone) for top four threads (2002-2011)")

Tone_plot = Tone_plot + geom_line(size=1)

Tone_plot = Tone_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2) + theme(legend.position = "none")

Tone_plot


# change of i

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_i= as.data.frame(as.table(by(first_group$i,list(first_group$ThreadID, first_group$year), mean)))

colnames(first_i) = c("threadID","year", "freqi")


# 145223

second_i= as.data.frame(as.table(by(second_group$i,list(second_group$ThreadID, second_group$year), mean)))

colnames(second_i) = c("threadID","year", "freqi")


# 252620

third_i= as.data.frame(as.table(by(third_group$i,list(third_group$ThreadID, third_group$year), mean)))

colnames(third_i) = c("threadID","year", "freqi")


# 283958

fourth_i= as.data.frame(as.table(by(fourth_group$i,list(fourth_group$ThreadID, fourth_group$year), mean)))

colnames(fourth_i) = c("threadID","year", "freqi")
```

```
# combine all the data frames

year_i = rbind(first_i, second_i, third_i, fourth_i)

year_i$year = as.numeric(as.character(year_i$year))

i_plot = ggplot(year_i, aes(year, freqi, color = threadID))

i_plot = i_plot + xlab ("Year") + ylab ("Average i score") + ggtitle("The change of i score (i) for top four threads (2002-2011)")

i_plot = i_plot + geom_line(size=1)

i_plot = i_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2)+ theme(legend.position = "none")

i_plot


# change of posemo

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_posemo=
as.data.frame(as.table(by(first_group$posemo,list(first_group$ThreadID, first_group$year), mean)))

colnames(first_posemo) = c("threadID","year", "freqposemo")


# 145223

second_posemo=
as.data.frame(as.table(by(second_group$posemo,list(second_group$ThreadID, second_group$year), mean)))

colnames(second_posemo) = c("threadID","year", "freqposemo")


# 252620

third_posemo=
as.data.frame(as.table(by(third_group$posemo,list(third_group$ThreadID, third_group$year), mean)))

colnames(third_posemo) = c("threadID","year", "freqposemo")


# 283958

fourth_posemo=
as.data.frame(as.table(by(fourth_group$posemo,list(fourth_group$ThreadID, fourth_group$year), mean)))
```

```
colnames(fourth_posemo) = c("threadID","year", "freqposemo")



# combine all the data frames

year_posemo = rbind(first_posemo, second_posemo, third_posemo, fourth_posemo)

year_posemo$year = as.numeric(as.character(year_posemo$year))

posemo_plot = ggplot(year_posemo, aes(year, freqposemo, color = threadID))

posemo_plot = posemo_plot + xlab ("Year") + ylab ("Average posemo score") +
ggtitle("The change of posemo score (posemo) for top four threads (2002-2011)")

posemo_plot = posemo_plot + geom_line(size=1)

posemo_plot = posemo_plot + geom_point(aes(colour = threadID, shape = threadID),
size = 2) + theme(legend.position = "none")

posemo_plot



# change of negemo

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_negemo=
as.data.frame(as.table(by(first_group$negemo,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_negemo) = c("threadID","year", "freqnegemo")


# 145223

second_negemo=
as.data.frame(as.table(by(second_group$negemo,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_negemo) = c("threadID","year", "freqnegemo")

# 252620

third_negemo=
as.data.frame(as.table(by(third_group$negemo,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_negemo) = c("threadID","year", "freqnegemo")
```

```
# 283958

fourth_negemo=
as.data.frame(as.table(by(fourth_group$negemo,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_negemo) = c("threadID","year", "freqnegemo")



# combine all the data frames

year_negemo = rbind(first_negemo, second_negemo, third_negemo, fourth_negemo)

year_negemo$year = as.numeric(as.character(year_negemo$year))

negemo_plot = ggplot(year_negemo, aes(year, freqnegemo, color = threadID))

negemo_plot = negemo_plot + xlab ("Year") + ylab ("Average negemo score") +
ggtitle("The change of negemo score (negemo) for top four threads (2002-2011)")

negemo_plot = negemo_plot + geom_line(size=1)

negemo_plot = negemo_plot + geom_point(aes(colour = threadID, shape = threadID),
size = 2)+ theme(legend.position = "none")

negemo_plot



# change of anx

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_anx= as.data.frame(as.table(by(first_group$anx,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_anx) = c("threadID","year", "freqanx")


# 145223

second_anx=
as.data.frame(as.table(by(second_group$anx,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_anx) = c("threadID","year", "freqanx")


# 252620
```

```
third_anx= as.data.frame(as.table(by(third_group$anx,list(third_group$ThreadID, third_group$year), mean)))

colnames(third_anx) = c("threadID","year", "freqanx")


# 283958

fourth_anx= as.data.frame(as.table(by(fourth_group$anx,list(fourth_group$ThreadID, fourth_group$year), mean)))

colnames(fourth_anx) = c("threadID","year", "freqanx")



# combine all the data frames

year_anx = rbind(first_anx, second_anx, third_anx, fourth_anx)

year_anx$year = as.numeric(as.character(year_anx$year))

anx_plot = ggplot(year_anx, aes(year, freqanx, color = threadID))

anx_plot = anx_plot + xlab ("Year") + ylab ("Average anx score") + ggtitle("The change of anx score (anx) for top four threads (2002-2011)")

anx_plot = anx_plot + geom_line(size=1)

anx_plot = anx_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2)+ theme(legend.position = "none")

anx_plot



# change of fp

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_fp= as.data.frame(as.table(by(first_group$focuspresent,list(first_group$ThreadID, first_group$year), mean)))

colnames(first_fp) = c("threadID","year", "freqfp")


# 145223

second_fp= as.data.frame(as.table(by(second_group$focuspresent,list(second_group$ThreadID, second_group$year), mean)))
```

```
colnames(second_fp) = c("threadID","year", "freqfp")


# 252620

third_fp=
as.data.frame(as.table(by(third_group$focuspresent,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_fp) = c("threadID","year", "freqfp")


# 283958

fourth_fp=
as.data.frame(as.table(by(fourth_group$focuspresent,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_fp) = c("threadID","year", "freqfp")




# combine all the data frames

year_fp = rbind(first_fp, second_fp, third_fp, fourth_fp)

year_fp$year = as.numeric(as.character(year_fp$year))

fp_plot = ggplot(year_fp, aes(year, freqfp, color = threadID))

fp_plot = fp_plot + xlab ("Year") + ylab ("Average focuspreesnt score") + ggtitle("The
change of focuspresent score (fp) for top four threads (2002-2011)")

fp_plot = fp_plot + geom_line(size=1)

fp_plot = fp_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2)+
theme(legend.position = "none")

fp_plot



# change of focuspast

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_fpast=
as.data.frame(as.table(by(first_group$focuspast,list(first_group$ThreadID,
first_group$year), mean)))

colnames(first_fpast) = c("threadID","year", "freqfpast")
```

```
# 145223

second_fpast=
as.data.frame(as.table(by(second_group$focuspast,list(second_group$ThreadID,
second_group$year), mean)))

colnames(second_fpast) = c("threadID","year", "freqfpast")


# 252620

third_fpast=
as.data.frame(as.table(by(third_group$focuspast,list(third_group$ThreadID,
third_group$year), mean)))

colnames(third_fpast) = c("threadID","year", "freqfpast")


# 283958

fourth_fpast=
as.data.frame(as.table(by(fourth_group$focuspast,list(fourth_group$ThreadID,
fourth_group$year), mean)))

colnames(fourth_fpast) = c("threadID","year", "freqfpast")




# combine all the data frames

year_fpast = rbind(first_fpast, second_fpast, third_fpast, fourth_fpast)

year_fpast$year = as.numeric(as.character(year_fpast$year))

fpast_plot = ggplot(year_fpast, aes(year, freqfpast, color = threadID))

fpast_plot = fpast_plot + xlab ("Year") + ylab ("Average focuspreesnt score") +
ggtitle("The change of focuspast score (fpast) for top four threads (2002-2011)")

fpast_plot = fpast_plot + geom_line(size=1)

fpast_plot = fpast_plot + geom_point(aes(colour = threadID, shape = threadID), size =
2)+ theme(legend.position = "none")

fpast_plot




# change of focusfururе

# threadID = c(127115, 145223, 252620, 283958)
```

```
# 127115

first_ff= as.data.frame(as.table(by(first_group$focusfuture,list(first_group$ThreadID, first_group$year), mean)))

colnames(first_ff) = c("threadID","year", "freqff")


# 145223

second_ff= as.data.frame(as.table(by(second_group$focusfuture,list(second_group$ThreadID, second_group$year), mean)))

colnames(second_ff) = c("threadID","year", "freqff")


# 252620

third_ff= as.data.frame(as.table(by(third_group$focusfuture,list(third_group$ThreadID, third_group$year), mean)))

colnames(third_ff) = c("threadID","year", "freqff")


# 283958

fourth_ff= as.data.frame(as.table(by(fourth_group$focusfuture,list(fourth_group$ThreadID, fourth_group$year), mean)))

colnames(fourth_ff) = c("threadID","year", "freqff")



# combine all the data frames

year_ff = rbind(first_ff, second_ff, third_ff, fourth_ff)

year_ff$year = as.numeric(as.character(year_ff$year))

ff_plot = ggplot(year_ff, aes(year, freqff, color = threadID))

ff_plot = ff_plot + xlab ("Year") + ylab ("Average focuspreesnt score") + ggtitle("The change of focusfurure score (ff) for top four threads (2002-2011)")

ff_plot = ff_plot + geom_line(size=1)

ff_plot = ff_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2)

ff_plot
```

```
# change of ppron

# threadID = c(127115, 145223, 252620, 283958)


# 127115

first_ppron= as.data.frame(as.table(by(first_group$ppron,list(first_group$ThreadID, first_group$year), mean)))

colnames(first_ppron) = c("threadID","year", "freqppron")


# 145223

second_ppron= as.data.frame(as.table(by(second_group$ppron,list(second_group$ThreadID, second_group$year), mean)))

colnames(second_ppron) = c("threadID","year", "freqppron")


# 252620

third_ppron= as.data.frame(as.table(by(third_group$ppron,list(third_group$ThreadID, third_group$year), mean)))

colnames(third_ppron) = c("threadID","year", "freqppron")


# 283958

fourth_ppron= as.data.frame(as.table(by(fourth_group$ppron,list(fourth_group$ThreadID, fourth_group$year), mean)))

colnames(fourth_ppron) = c("threadID","year", "freqppron")



# combine all the data frames

year_ppron = rbind(first_ppron, second_ppron, third_ppron, fourth_ppron)

year_ppron$year = as.numeric(as.character(year_ppron$year))

ppron_plot = ggplot(year_ppron, aes(year, freqppron, color = threadID))

ppron_plot = ppron_plot + xlab ("Year") + ylab ("Average focuspreesnt score") + ggtitle("The change of ppron score (ppron) for top four threads (2002-2011)")

ppron_plot = ppron_plot + geom_line(size=1)
```

```
ppron_plot = ppron_plot + geom_point(aes(colour = threadID, shape = threadID), size = 2)+ theme(legend.position = "none")

ppron_plot


grid.arrange(wc_plot, Analytic_plot, Authentic_plot, nrow=3, ncol =1)









library(igraph)

library(igraphdata)


# convert year and month to numeric

webforum$year = as.integer(webforum$year)

webforum$month = as.integer(webforum$month)


# Graph all the 120 monthly social network in the forum(to find the graph with more than 30 nodes)

for(i in 2002:2011){

  for(j in 1:12){

    sixth_group = webforum[webforum$year == i & webforum$month == j, ]

    # edges dataframe

    links = sixth_group[,2:3]

    colnames(links) = c("from", "to")

    # nodes data frame

    nodes =as.data.frame(unique(sixth_group$ThreadID))

    colnames(nodes) = c("Nodes")

    # Authors data frame

    Authors = as.data.frame(unique(sixth_group$AuthorID))

    colnames(Authors) = c("Nodes")

    nodes = rbind(nodes, Authors)

    net6 <- graph_from_data_frame(d=links, vertices=nodes, directed=F)
```

```
    # change vertices apperence
    V(net6)$label <- NA
    V(net6)$size <- 4
    V(net6)$color <- "red"



    #plot(net)
    plot(net6, layout = layout.fruchterman.reingold, main=paste("The network of participants (", i, ",",j, ")"))
  }
}



sixth_group = webforum[webforum$year == 2008 & webforum$month == 4, ]

# edges dataframe
links = sixth_group[,2:3]
colnames(links) = c("from", "to")

# nodes data frame
nodes =as.data.frame(unique(sixth_group$ThreadID))
colnames(nodes) = c("Nodes")

# Authors data frame
Authors = as.data.frame(unique(sixth_group$AuthorID))
colnames(Authors) = c("Nodes")
nodes = rbind(nodes, Authors)
net6 <- graph_from_data_frame(d=links, vertices=nodes, directed=F)

# change vertices appearence
V(net6)$label <- NA
V(net6)$size <- 4
V(net6)$color <- "red"
```

```
#plot(net)(Graph 2.5)

plot(net6, layout = layout.fruchterman.reingold, main=paste("The network of participants (", 2008, ",",4, ")"))



#install.packages("tidyverse")

library("tidyverse")


# Create a 1 x 1 plotting matrix

par(mfrow = c(1, 1))


# grouping by AuthorID + ThreadID

p = aggregate(month ~ AuthorID + ThreadID, data = sixth_group, FUN = mean, na.rm = TRUE)


p = p[,1:2]


# check number of authors who have posted in more than 2 different threads at the specific month

q = as.data.frame(as.table(by(p, p$AuthorID, nrow)))


nrow(q[q$Freq>=2,])



# plot a basic graph


graph_1 <- graph.data.frame(p, directed=FALSE)


graph_1


plot(graph_1)
```

```
# it looks congested, tell R that it is a two-mode network
bipartite.mapping(graph_1)

V(graph_1)$type <- bipartite_mapping(graph_1)$type

plot(graph_1)



plot(graph_1, vertex.label.cex = 0.8, vertex.label.color = "black")

# distinguish between authors and threads by changing colours and size of nodes
V(graph_1)$color <- ifelse(V(graph_1)$type, "lightblue", "salmon")
V(graph_1)$shape <- ifelse(V(graph_1)$type, "circle", "square")
E(graph_1)$color <- "lightgray"

# try differnt layouts
plot(graph_1, vertex.label.cex = 0.8, vertex.label.color = "black")

plot(graph_1, layout=layout.bipartite, vertex.size=7, vertex.label.cex=0.6)




V(graph_1)$label.color <- "black" ##ifelse(V(graph_1)$type, "black", "white")

## V(graph_1)$label.font <-  2
V(graph_1)$label.cex <- 1 ##ifelse(V(graph_1)$type, 0.8, 1.2)

## V(graph_1)$label.dist <-0
V(graph_1)$frame.color <-  "gray"
V(graph_1)$size <- 10
```

```
plot(graph_1, layout = layout_with_graphopt)

par(mar=c(0,0,0,0))


# the graph is not connected need to choose one subgrapgh for further analysis
clu <- components(graph_1)
groups(clu)


dg <- decompose.graph(graph_1) # returns a list of three graphs

# we choose subgraph with index "1"
final_graph = dg[[1]]

V(final_graph)$size <- 10

V(final_graph)$label.cex <- 0.8

# graph 2.5
plot(final_graph, layout = layout.sphere(final_graph),  asp=0)


# now produce table of network's centrality measures
types <- V(final_graph)$type
deg <- degree(final_graph)
bet <- betweenness(final_graph)
clos <- closeness(final_graph)
eig <- eigen_centrality(final_graph)$vector

# put all in a dataframe
cent_df <- data.frame(types, deg, bet, clos, eig)
```

```
cent_df[order(cent_df$type, decreasing = TRUE),]


# Sizing Vertices by degree
V(final_graph)$size <- degree(final_graph)
V(final_graph)$label.cex <- degree(final_graph) * 0.2

#Graph 2.6
plot(final_graph, layout = layout.sphere(final_graph),  asp=0)

# # Sizing Vertices by Betweenness
# reset size
V(final_graph)$size <- 10

V(final_graph)$label.cex <- 0.8

max(cent_df$bet) # 1844.902

min(cent_df$bet) # 0


V(final_graph)$size <- betweenness(final_graph)/80
V(final_graph)$label.cex <- betweenness(final_graph)/1000

#graph 2.7
plot(final_graph, layout = layout.sphere(final_graph),  asp=0)


# Sizing Vertices by closeness
# reset size
V(final_graph)$size <- 10
```

```
V(final_graph)$label.cex <- 0.8


max(cent_df$clos) # 0.002538071


min(cent_df$clos) # 0.001191895



V(final_graph)$size <- ifelse(V(final_graph)$type,closeness(final_graph)*10,
closeness(final_graph)*10000)

V(final_graph)$label.cex <- ifelse(V(final_graph)$type,closeness(final_graph)*10,
closeness(final_graph)*500)


#graph 2.8

plot(final_graph, layout = layout.sphere(final_graph),  asp=0)



# Sizing Vertices by Eigenvector

# reset size

V(final_graph)$size <- 10


V(final_graph)$label.cex <- 0.8


max(cent_df$eig) # 1


min(cent_df$eig) # 0.002106946


V(final_graph)$size <- evcent(final_graph)$vector*20

V(final_graph)$label.cex <- evcent(final_graph)$vector*1
```

```
#graph 2.9
plot(final_graph, layout = layout.sphere(final_graph), asp=0)



# most influential authors
seventh_group = sixth_group[sixth_group$AuthorID == c(83488, 54960, 118148, 142865, 5123, 138220),]


by(sixth_group[,6:24], sixth_group$AuthorID, sum)


top_6_authors = aggregate(seventh_group[, 6:24], list(seventh_group$AuthorID), mean)


col_names = colnames(top_6_authors)


colnames(top_6_authors) = c("AuthorID", col_names[2:20])


# convert AuthorID to factor


top_6_authors$AuthorID = as.factor(top_6_authors$AuthorID)


library(tidyverse)
library(gridExtra)



top_6_authors$AuthorID <- factor(top_6_authors$AuthorID, levels = c(83488, 54960, 118148, 142865, 5123, 138220))
top_6_authors$AuthorID # the changed order of factor levels


plot1 = qplot(AuthorID, WC, data = top_6_authors, color = AuthorID, size = I(6)) + theme(legend.position = "none")
```

```
plot2 = qplot(AuthorID, Analytic, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot3 = qplot(AuthorID, Clout, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot4 = qplot(AuthorID, Authentic, data = top_6_authors, color = AuthorID, size = I(6))
+ theme(legend.position = "none")


plot5 = qplot(AuthorID, Tone, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot6 = qplot(AuthorID, ppron, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


#grpah 3.0

grid.arrange(plot1, plot2, plot3, plot4, plot5, plot6,  nrow=2, ncol = 3)



plot7 = qplot(AuthorID, i, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot8 = qplot(AuthorID, we, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot9 = qplot(AuthorID, you, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot10 = qplot(AuthorID, shehe, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot11 = qplot(AuthorID, they, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot12 = qplot(AuthorID, posemo, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")
```

```
#graph 3.0

grid.arrange(plot7, plot7, plot9, plot10, plot11, plot12,  nrow=2, ncol = 3)



plot13 = qplot(AuthorID, negemo, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot14 = qplot(AuthorID, anx, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot15 = qplot(AuthorID, anger, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot16 = qplot(AuthorID, sad, data = top_6_authors, color = AuthorID, size = I(6)) +
theme(legend.position = "none")


plot17 = qplot(AuthorID, focuspast, data = top_6_authors, color = AuthorID, size = I(6))
+ theme(legend.position = "none")


plot18 = qplot(AuthorID, focuspresent, data = top_6_authors, color = AuthorID, size =
I(6)) + theme(legend.position = "none")


#graph 3.0

grid.arrange(plot13, plot14, plot15, plot16, plot17, plot18,  nrow=2, ncol = 3)


#graph 3.0

plot19 = qplot(AuthorID, focusfuture, data = top_6_authors, color = AuthorID, size =
I(6)) + theme(legend.position = "none")


grid.arrange(plot19,  nrow=2, ncol = 3)
```