#JNU Crisis Study

Social Media Analysis of JNU Crisis and its Impact on Well-being

Akshay Agarwal , Sarthak Ghosh, Anjjan Narayan

Abstract

The Jawaharlal Nehru University (JNU) has experienced a series of clashes, protests, media outrage, and controversies. The JNU crisis started out as a peaceful protest that was organized by students against the hanging of a terrorist named Afzal Guru. The students felt that the judicial killing was unjust. The protest soon turned out violent and caught immense media attention. The protest ignited political outrage and also divided the entire nation based on opinions. There were also reports about violent clashes caused due to this situation.

Our project tries to track this movement, observe how the sentiments of people change over time and correlate it to the developments that took place along the way, especially in terms of police and political action. We aim to see if we can find patterns in online activity around violent clashes, provide feedback on political and police actions, and study how well-being of those involved along with those on the periphery affected.

Also, this movement has changed color due to the events involved, and the people commenting about this, we thus want to track in which direction the movement has spiralled of, and the nature of change.

Introduction

On February 9, 2016, the left wing Democratic Student Union (DSU) from JNU held a protest against capital punishment for Afzal Guru who was convicted for his role in 2001 Parliament attacks. There were reports that anti-india slogans were raised at the protest. These anti-India slogans led to the arrest of the JNU Students' Union President Kanhaiya Kumar and Umar Khalid on charges of sedition. The support for Afzal Guru and the arrests of students caused a huge nationwide controversy. The students were then academically debarred. Although Kanhaiya Kumar was granted six month interim bail by the Delhi High Court, the controversy prolonged for a certain period of time. On return from jail, Kanhaiya Kumar went on to give a speech at JNU that created further uproar.

The controversy affected people from all walks of life. Politicians, Celebrities, Professors, Students etc started having different opinions on this situation. The media outrage added to the situation and also stirred various clashes and protests all around India. There were situation where celebrities have come out and expressed their opinions on the social media. The controversy also led many people to express their anger using social media.

We believe that it is of paramount importance to study such movements as they affect the wellbeing of the people involved, and the population in general in the many ways. Students from Kashmir who were part of the initial demonstration felt that they were being targeted, branded as terrorists for expressing their sentiments, and this affects them beyond just their academic performance. People from Kashmir might feel anti-India sentiments, and such protests and clashes might lead them to lose faith in the Indian government, and their concern about Kashmiri populations. Violent clashes can be brutal, and affect the physical well-being of people involved.

In our study we try to observe

- How Indian student movements can be tracked through twitter data, considering the fact that Twitter is not very widely used throughout India.
- What traces of reactive emotions are left on a social media platform like Twitter by people, following the outbreak of student movements.
- How sentiments of people change over time and correlate it to the developments that took place along the way, especially in terms of police and political action.
- The role of media and well-known public personnels in guiding public perception

Related Work

Social media has been extensively used by people to express their views and opinions on various matters. During the time of controversy and crisis, people tend to use social media aggressively to voice their opinions. Researchers have used social media like twitter and Facebook to understand how people react in a situation of crisis.

'This Protest Will be Tweeted' by Earl et al. is a research that examined the use of twitter during G20 meeting at Pittsburg. The authors analyzed twitter data to understand the moods and opinions of people. They examined 30,000 tweets over a 9 day period to build a conclusion that protesters frequently used Twitter to share information, including information about protest locations, and the location and actions of police. Twitter played a major role in supporting these protests. Although this work analyzes how twitter was used by protesters to carry out their protests successfully, our work tries to analyze the change in people's behavior over time as situation changes.

Ahmed and Jaidka [2] have produced an interesting paper that also deals with something similar to our proposed work. They studied media organization, NGO's, and individual user’s twitter data during and after the mass protests against the Delhi gang rape incident. They tried to understand how individual users disseminate information during times of controversy and protests.

Although the mentioned papers do work that is similar to our proposed work, we see that there is a lot of scope in understanding the depth of the situation using social media data. We also feel that the work by Shagun et al [3]. lays a solid groundwork on analyzing protracted nature of events impact people’s expression on twitter. Both the situation of Black lives matter movement and the current JNU controversy have lots of points in common. In our work, we would want to analyze details such as tweet count over a time period of significant event, people’s behavior towards the issue and much more.

Data Collection

For data collection, we used Twitter search and streaming API. We created different python scripts that search for the given keyword and generate a JSON file that gives the details of tweets such as tweet text, re-tweet count, tweet date, tweet time, tweet user details etc. For setting up the streaming API, we used inbuilt python library called tweepy. For setting up the search API we have used the twarc python library. The data was collected from February 23, 2016 to April 9, 2016. Due to technical errors/server failure we missed one week of data starting from the 1st of April.
Overall we gathered 800 thousand tweets with a file size of approximately 5.6GB.

The hashtags were chosen based on the the events and movements that were concurrently happening in India. We kept a watch on trending hashtags and added them dynamically to our searching script as the movement progressed. For example, on March 23, Kanhayia Kumar of JNU decided to go to Hyderabad to strengthen a seemingly detached student protest. Their common demand was the resignation of the Indian HRD minister. This gave rise to a lot of media attention and the two movements, #JNU and #HCU (Hyderabad Central University) merged. This led us to include a few more search keywords and hashtags like #RohithVermulla, #HCUCrackdown, etc.

Streaming API Hashtags	Search API Hashtags
1. `'#JNURow',`	1. `'#JNU',`
2. `'#JNUCrackdown',`	2. `'#JNURow',`
3. `'#JNU',`	3. `'#JNUCrackdown',`

4. '#jnustandoff',	4. '#JNUStories',
5. '#JNUStories',	5. '#jnustandoff',
6. '#JNURow',	6. '#JNUStandoff',
7. '#StandWithJNU',	7. ’#StandWithJNU',
8. '#SaveDemocracy',	8. '#SaveDemocracy',
9. '#ProtectJNU',	9. '#ProtectJNU',
10. '#WeAreJNU’ .	10. '#WeAreJnu',
11. '#HCU',	12. '#KanhaiyaKumar',
	13. ’#RohithVemula',
	14. '#NITSrinagar',
	15. '#CleanUp',
	16. '#ZeroTolerance4AntiNationals',
	17. ’#CleanUpJNU',
	18. '#HCU',
	19. '#NITSrinagar',
	20. '#DalitLivesMatter',
	21. '#HCUcrackdown',

We also searched for some keywords that are relevant to the topic. These keywords include

1. 'Afzal Guru',
2. 'Kashmir',
3. 'sedition',
4. 'Kanhaiya Kumar',
5. 'Shilpi Tiwari',
6. 'Umar Khalid',
7. 'Anti-national',
8. 'Judicial Killing',
9. 'ABVP' .

Implementation

Data Cleaning

In order to clean the data, we separated the tweets into english language tweets and other language tweets, this information is relayed by twitter itself. We removed all retweets. We then processed the english language tweets, removed all urls and created a list of urls, we further tokenized each tweet and removed the stop words. We also created a list of tweets with at least one hashtag from the above list.

Data Processing

In order to perform data analysis, tableau and python were used. We used nltk library in order to perform LIWC analysis, and n-gram analysis on the data. We relied on google places API in order to extract format location information of user, so that we could identify the tweeting behavior based on location. We performed classification on retweet count using values of the 5 LIWC categories (positive affect, negative affect, anger, anxiety and sadness) as feature vectors.

To understand what kind of external material shared on Twitter got most public attention, we extracted all urls from the tweets and used urllib library to open the original weblinks. Following this, we created a frequency distribution of the urls shared.

LIWC analysis was performed in order to gain better perspective about the sentiment expressed and variation over time in the collected tweets. N-gram analysis was done to gather insights about the topics that were being discussed, and understand what people were tweeting about.

Results

This section describes the results and interpretation that we have obtained based on the experiments and methodologies we have used. The visualizations were done using Tableau. As shown in Figure 1, we first created a visualization of how the number of tweets varies over each day. For this, we used all the tweets (english and non-english). We then mapped the peak events with the events that has been reported on the press. We were able to see that as there is a major event that has been reported by the media, there is sharp increase in the number of tweets. For example, on March 11, there was a sharp increase in the number of tweets. We then correlated it to the event that occurred on March 11 where the students of JNU who were arrested were allowed to take classes again at JNU.

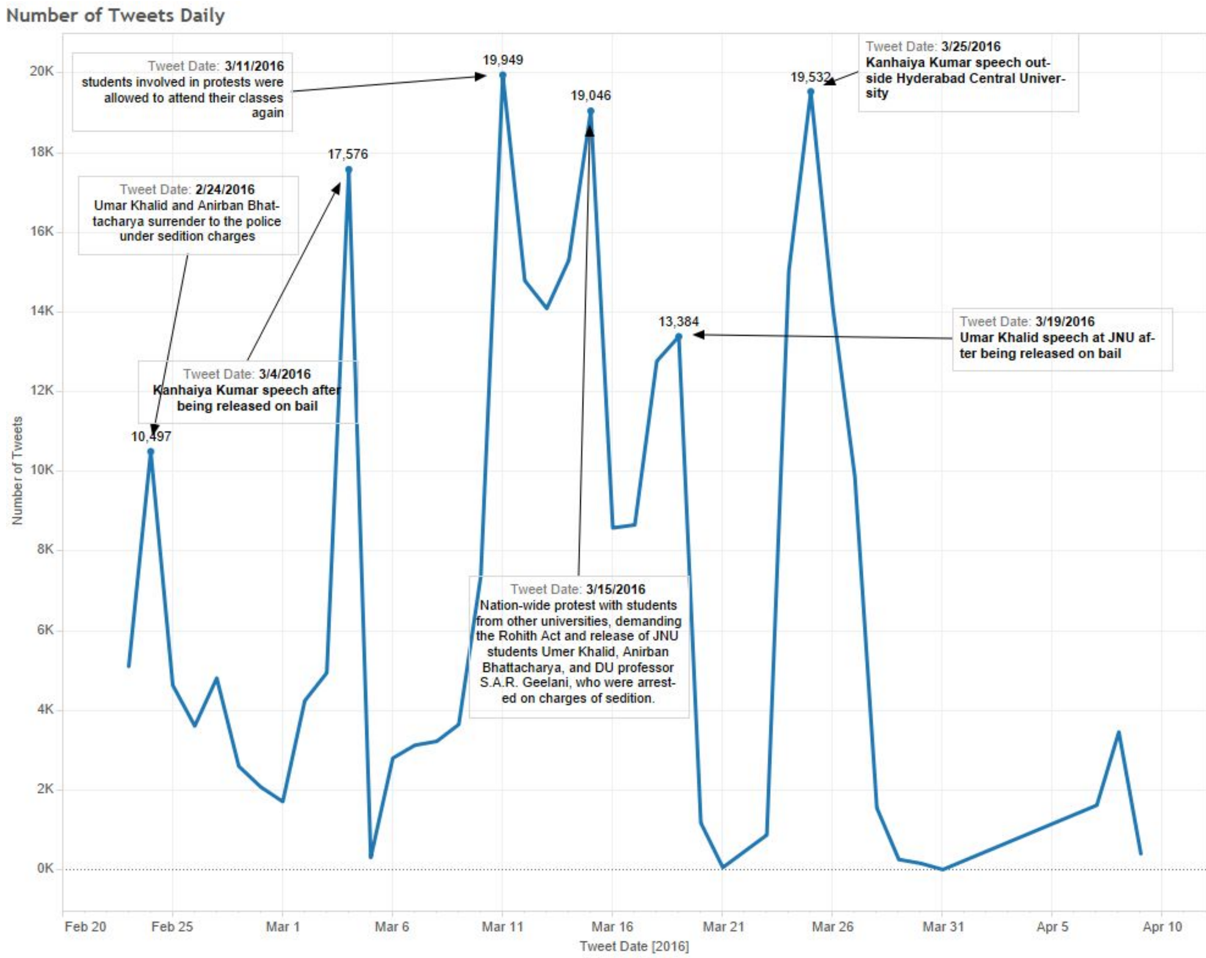


Figure 1. Time Series event mapping based on number of tweets

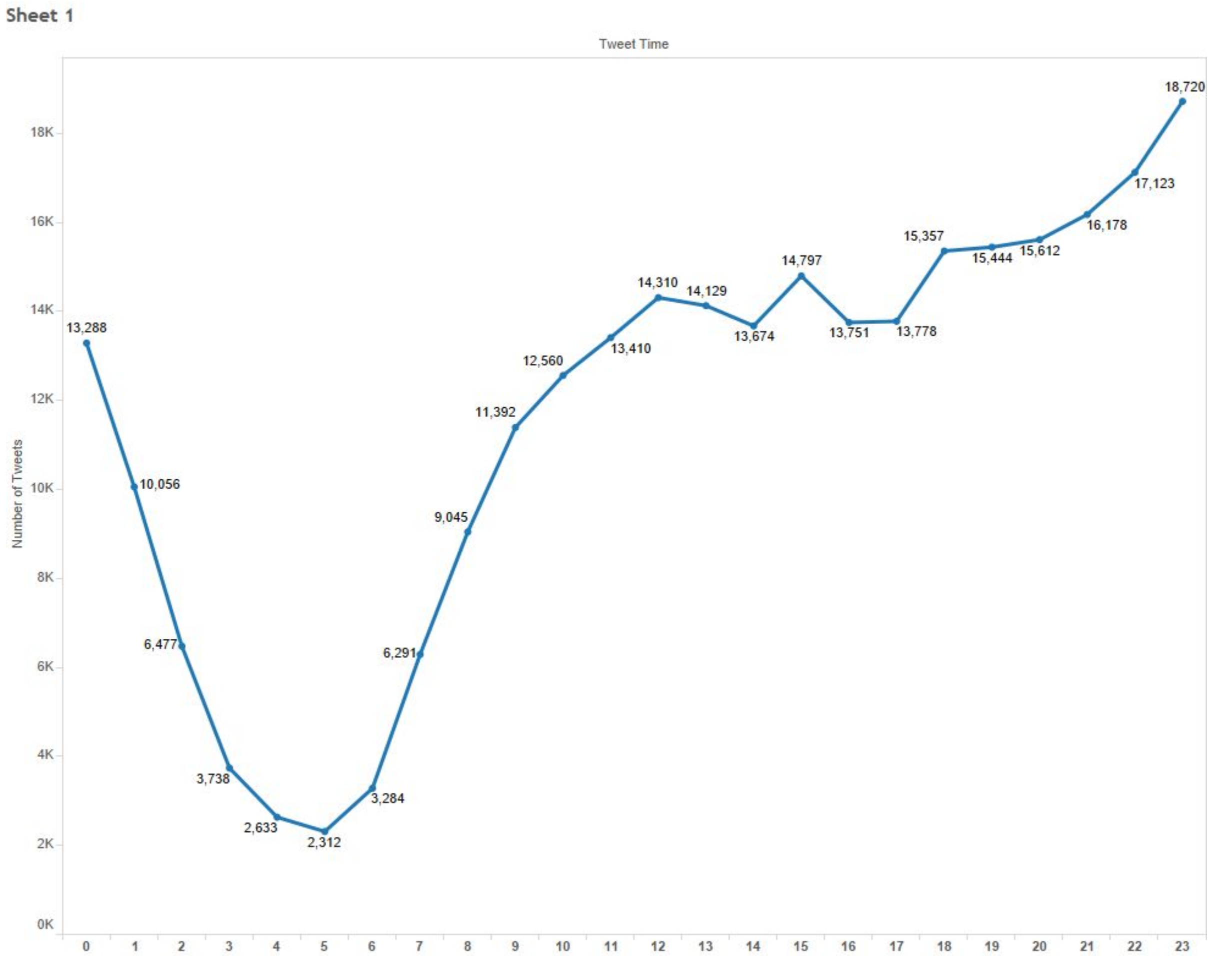


Figure 2. Time period of tweets over the day

In Figure 2 we try to see the time period when people are most likely to tweet about the crisis. We can see that majority of the tweets were posted during night time and especially during midnight. The time stamp we used was Indian Standard Time (IST).

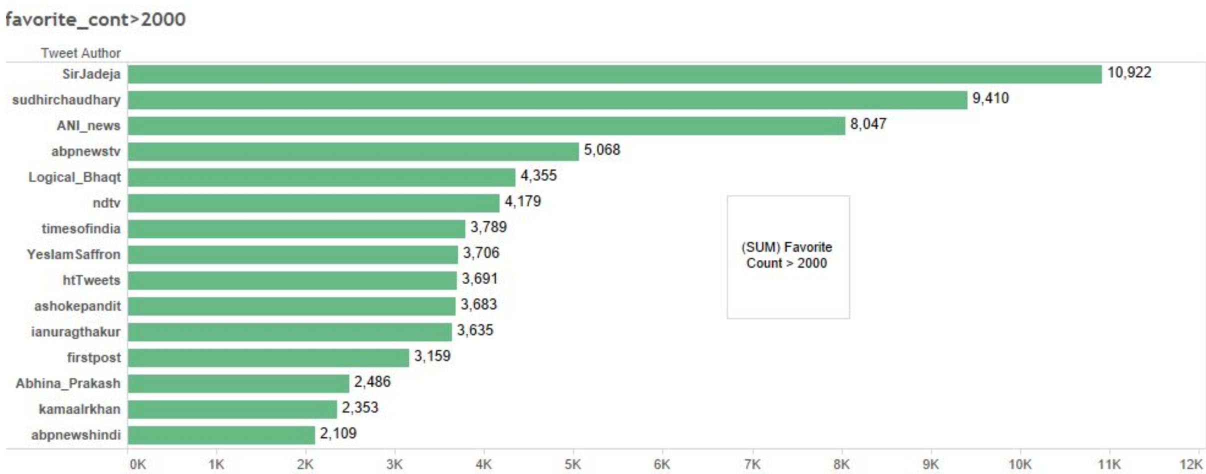


Figure 3. Users whose tweets have been favorited the most

Since the controversy that we are covering has involved people from all walks of life to share their opinion, we thought it would be a good idea to identify users whose tweets have made an impact on the overall users. As shown in Figure 3, we tried to find users whose tweets have been favorited most number of times. To make the study more concrete, we also considered the retweets. Figure 4 shows the users whose tweets have been retweeted atleast 400 times. We see that ANI_news which is an Indian news agency twitter handle has had the most impact on users. Following this, there are twitter handles such as SirJadega, Logical_Bhaqt, SudhirChoudhry who have played a major role in influencing other users. Using the retweet count as a measure of influencing power over Twitter, we see that among the highest influencers, most are journalists, or activists. It is rather interesting to note that the highest influence is achieved by a parody account holder by the name of Sir Ravindra Jadeja (name of an Indian Cricketer).

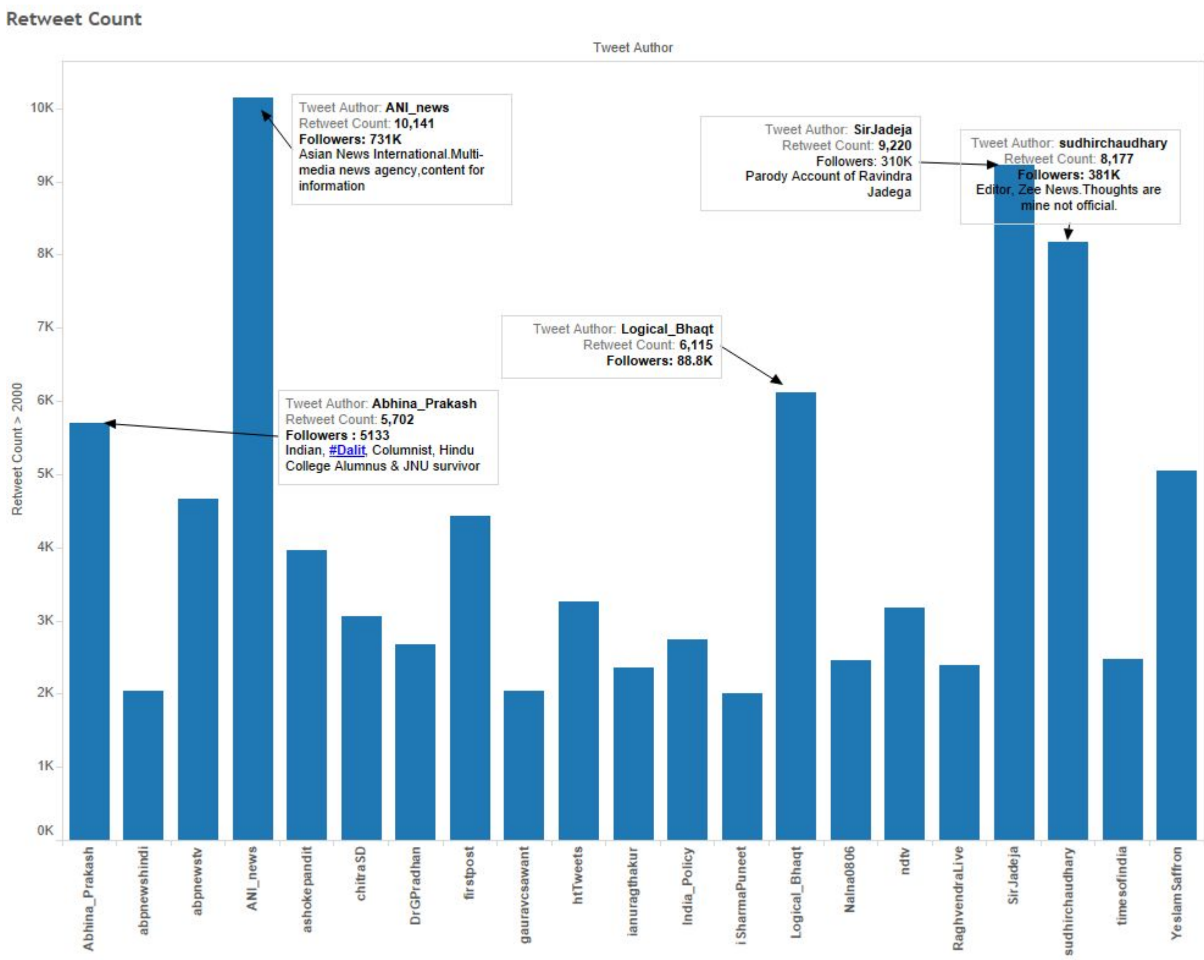


Figure 4. Users whose tweets have been retweeted the most(greater than 400)

Further, we have taken these users whose tweets have been retweeted most number of times and tried to do LIWC analysis on their tweets. We see that some users tweet content has more negative affect when compared to others. We see that the tweets that have been posted by ANI_News has Anger, Positive and Negative affect (mostly negative affect). It’s interesting to see that none of the top retweeted users have any swear analysis on their tweets. Figure 5 depicts these characteristics. This analysis tries to provide a good insight on the emotions expressed by users whose tweets influence many other users. This is important since we are trying to study a crucial controversy that has gripped the entire nation.

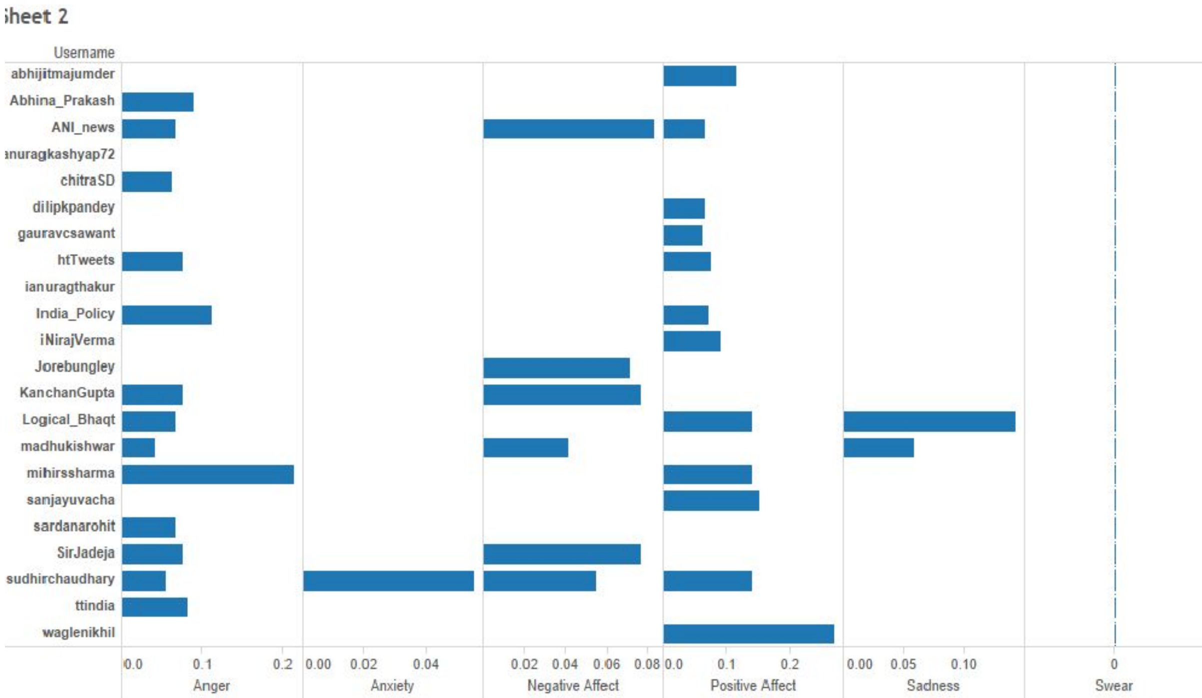
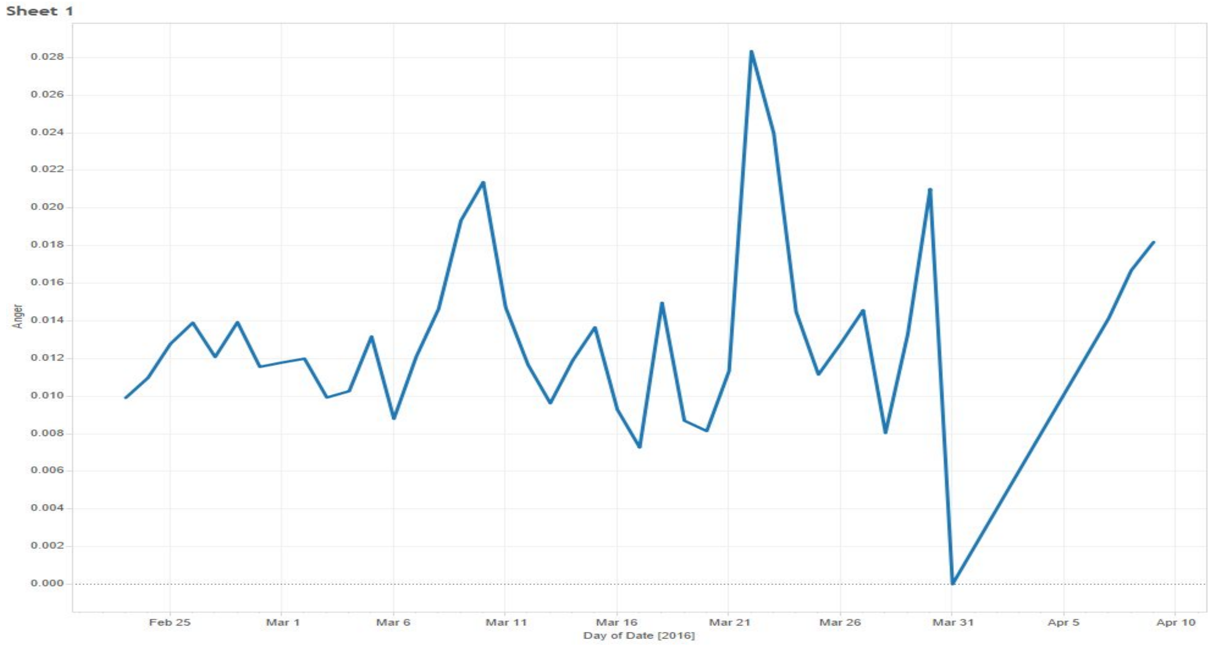


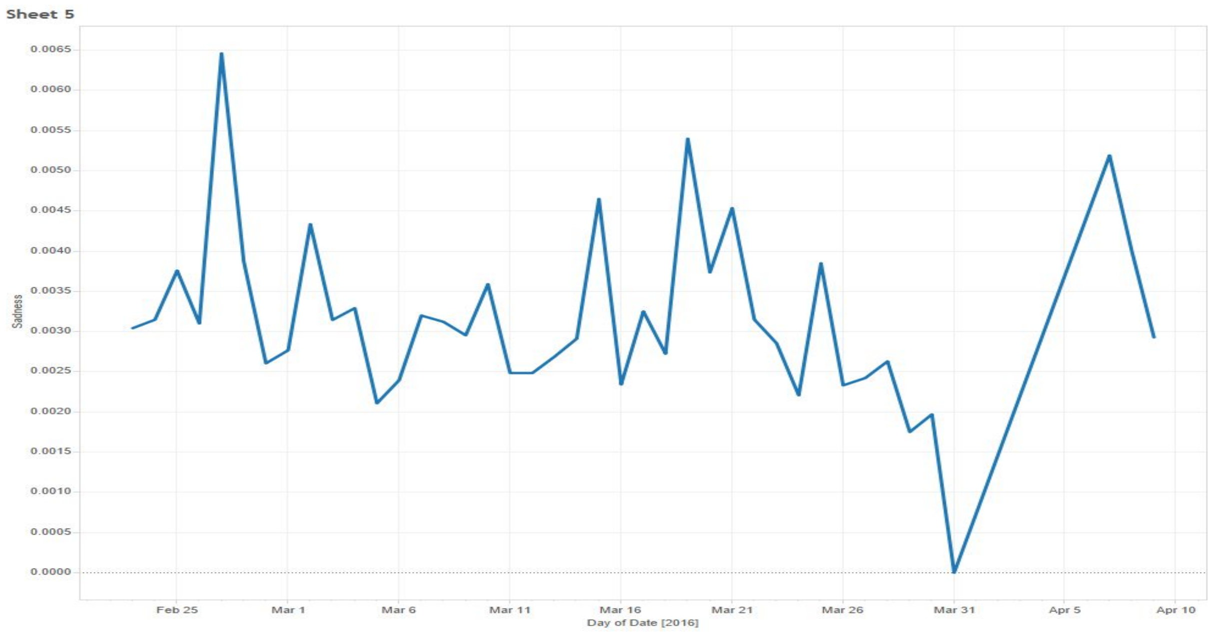
Figure 5. LIWC analysis of tweets of users whose tweets have been retweeted the most (greater than 400)

Now that we have used LIWC analysis on the users who have most number of retweet counts, we wanted to analyze the trend in emotions in all the tweets. For this, we performed LIWC analysis on tweets on a daily basis. Figure 6 shows this analysis. We can correlate the change in emotions expressed in the tweets based on the event mapping that we have already performed. For example, the anger level in the tweets peaked during March 21. This can be explained due to the fact that on March 19, Umar Khalid gave a speech after being released on bail.

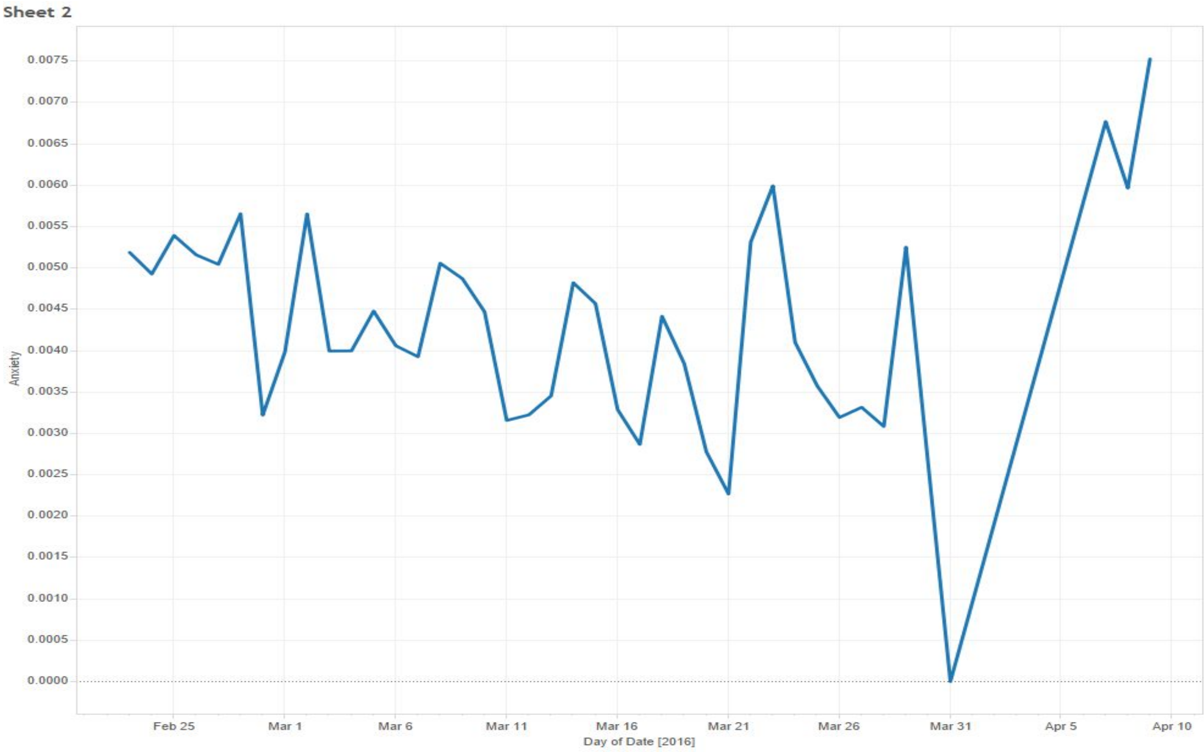
We also built a classifier using LIWC categories as feature vector set for predicting popularity based on retweet count. We assigned tweets which were retweeted more than 500 times a value of 1, and below that a value of 0, and then used a SVM classifier. The accuracy score was 98%. We further broke the retweet count into further categories and still achieved the same accuracy, however the accuracy with regressor was low.



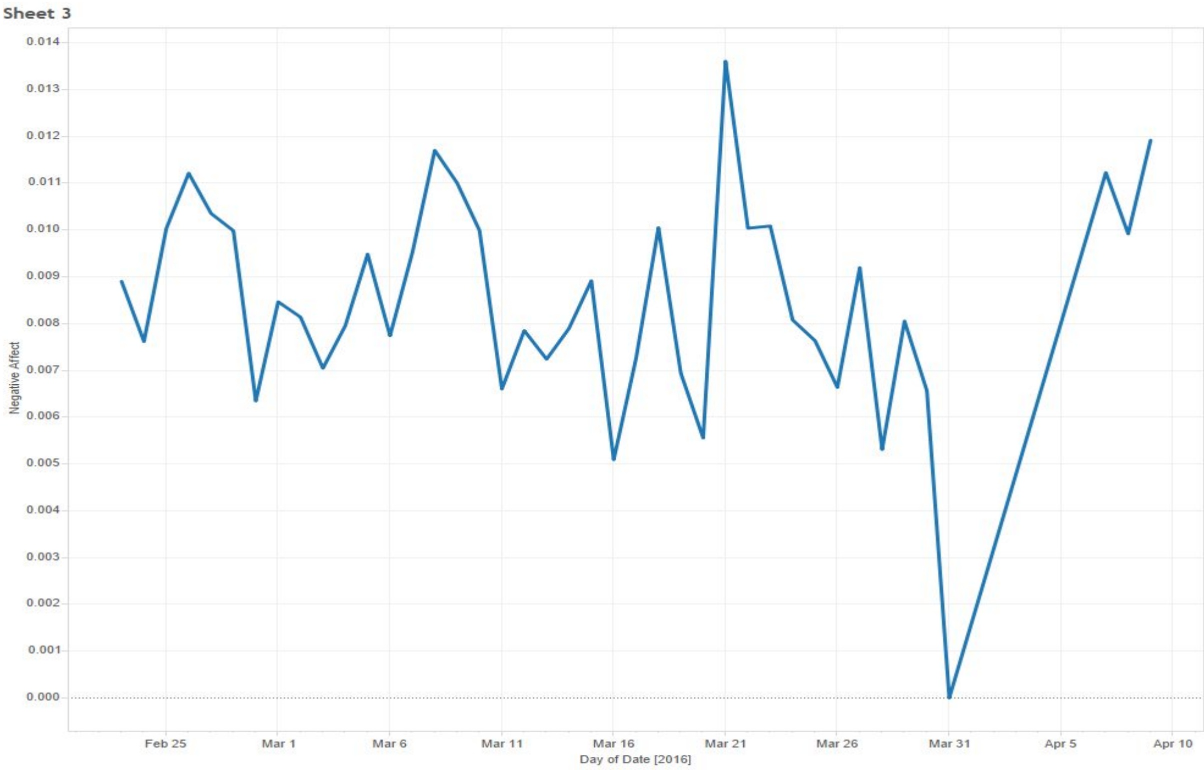
Anger



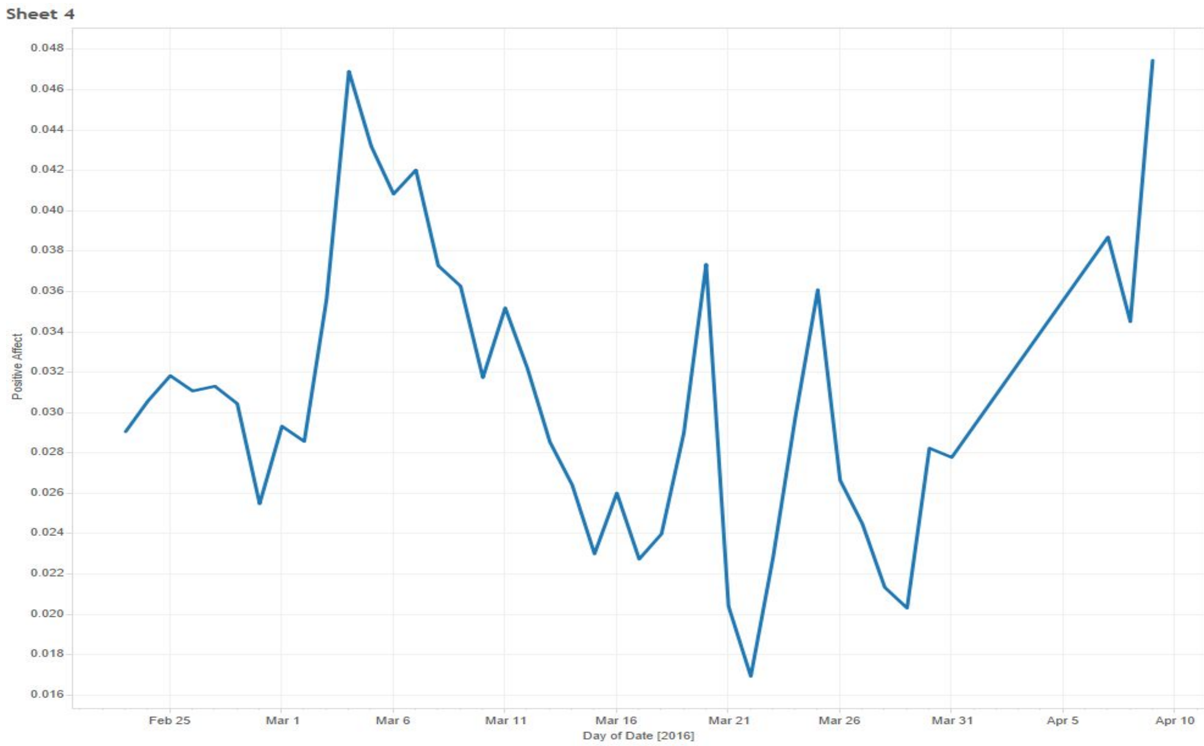
Sadness



Anxiety

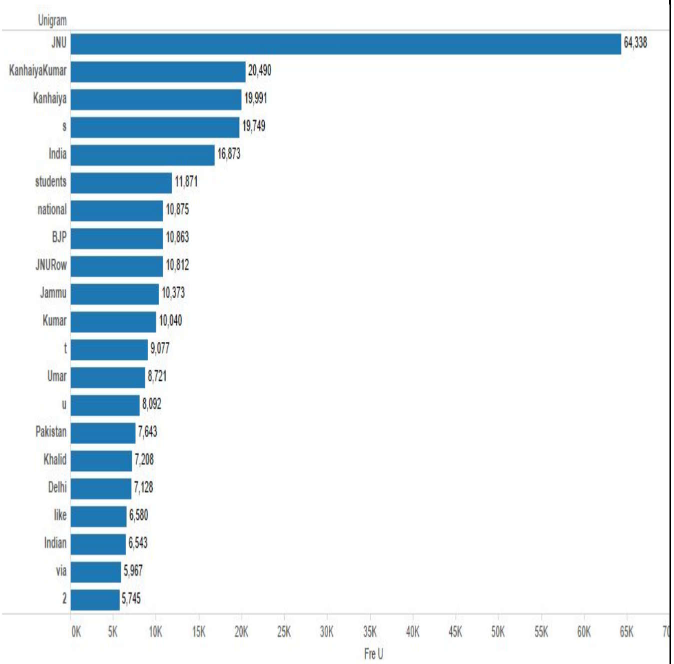


Negative Affect

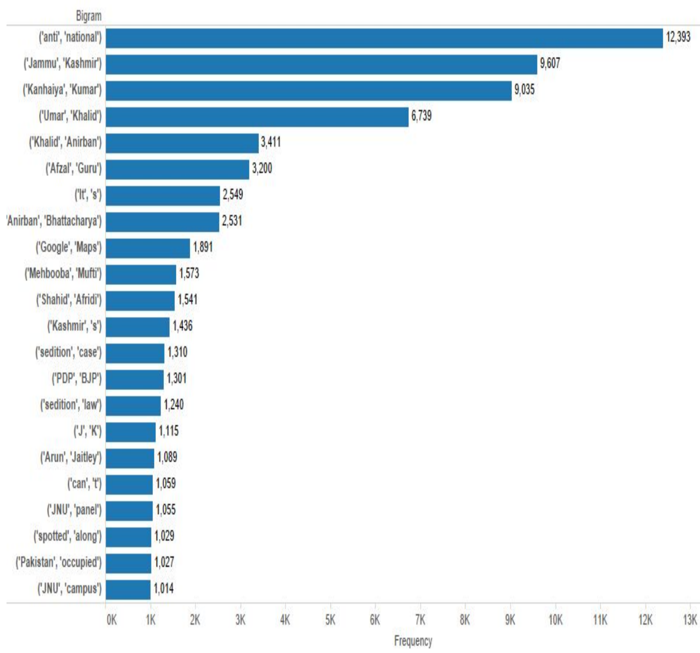


Positive affect

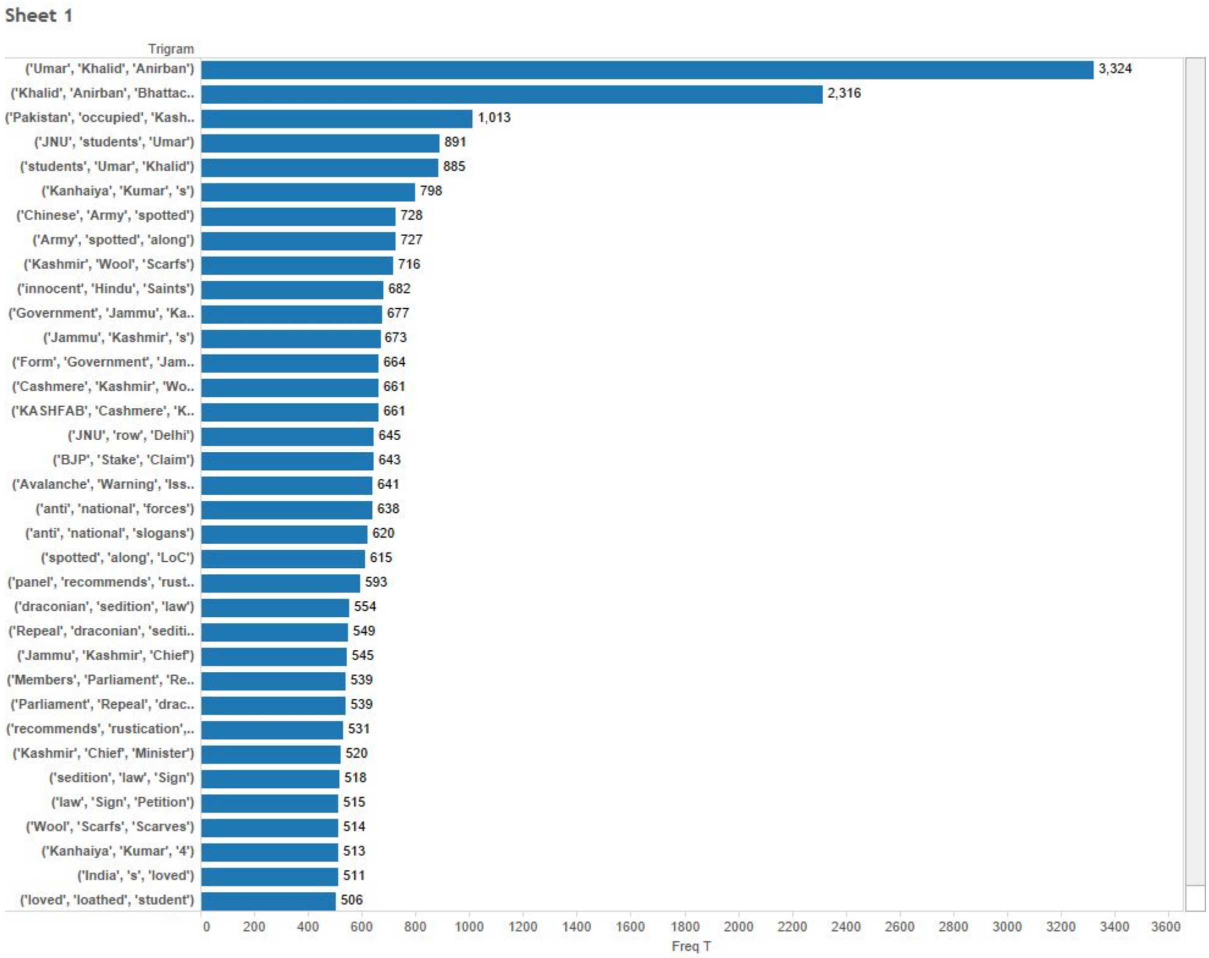
Figure 6. LIWC Analysis of tweets over whole time period



Unigrams



Bigrams



Trigram Frequency

Figure 7. Unigram, Bigram and Trigram Frequencies.

We identified the links that were shared at least 200 times. Based on this we could see that the video where Kanhaiya Kumar is released and his speech at JNU was shared almost 14,000 times. We could also find various articles and cartoons being shared during the time of controversy.

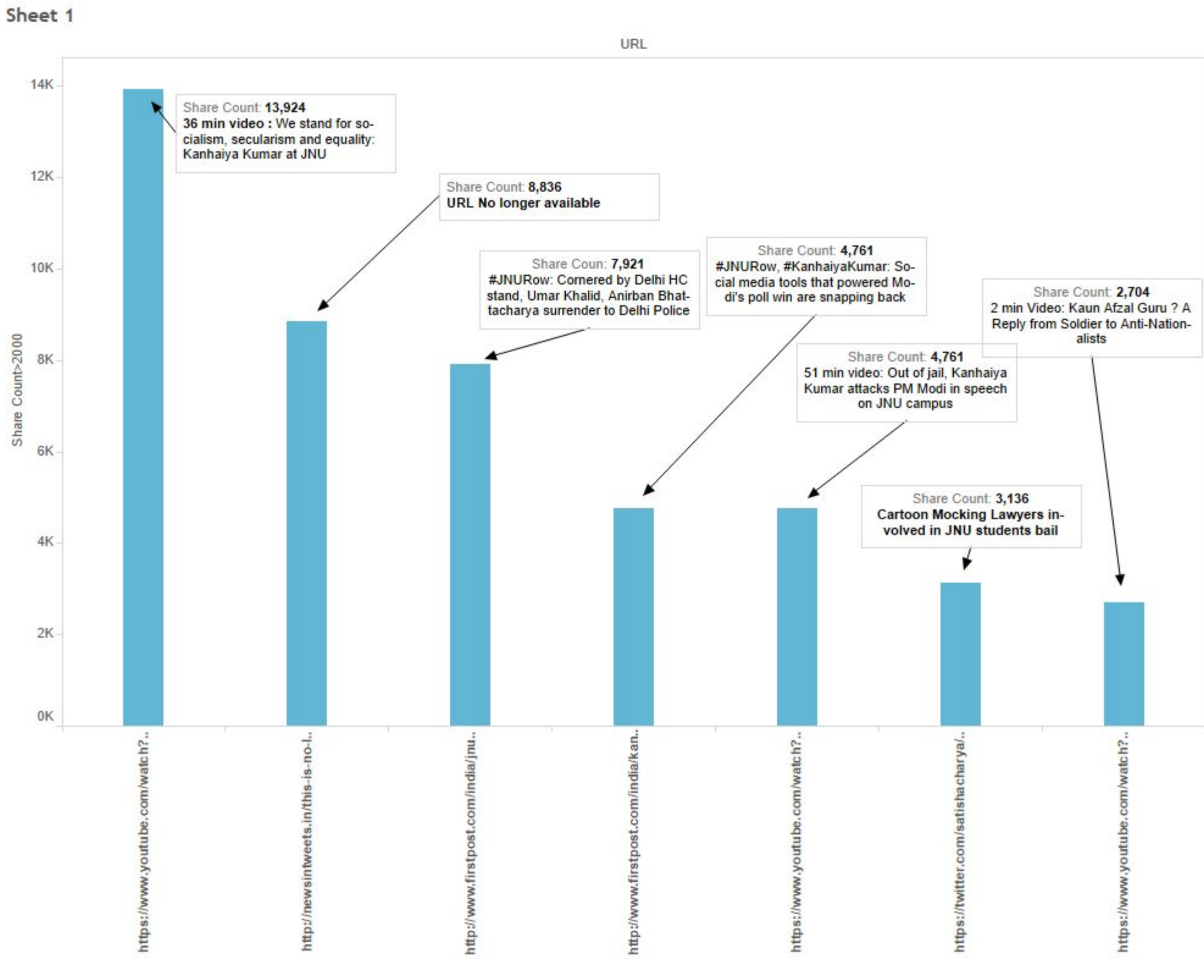


Figure 8. Most Shared URLs on tweets

Hashtag Analysis: To check how the nature of trending hashtags change over time, we created a frequency distribution of the top 20 most frequently used hashtags over the entire time period of our study- Figure 9 to Figure 13. In Figure 9 we see that the top 5 hashtags are #JNU, #Kanhayia and similar terms. This is highly predictable given that we specifically searched for these hashtags. However, it is still interesting to see that the frequency of these hashtags go dwindling down to single digits within the span of one month. This can be because of two reasons

1) The movement dies down, the government takes measures, or curbs it down or people forget about it.
2) The movement metamorphs into something different and gives rise to another movement. This would then lead to much different trends in hashtags as the movement changes its nature.

Hence we tried to observe what are the associated hashtags that have come along with the top few searched ones. Figure 10 to Figure 13 all show a variety of different hashtags which hint at the fact that the movement was politicised and changed from its original nature. We see hashtags like #IndiaPakistan, #BJP, #KashmirCries surfacing. Moreover the # JNU movement started getting associated with any other student movement in other parts of

India. This led to the surfacing of #HCU and #RohithVermula as associated hashtags with #JNU.

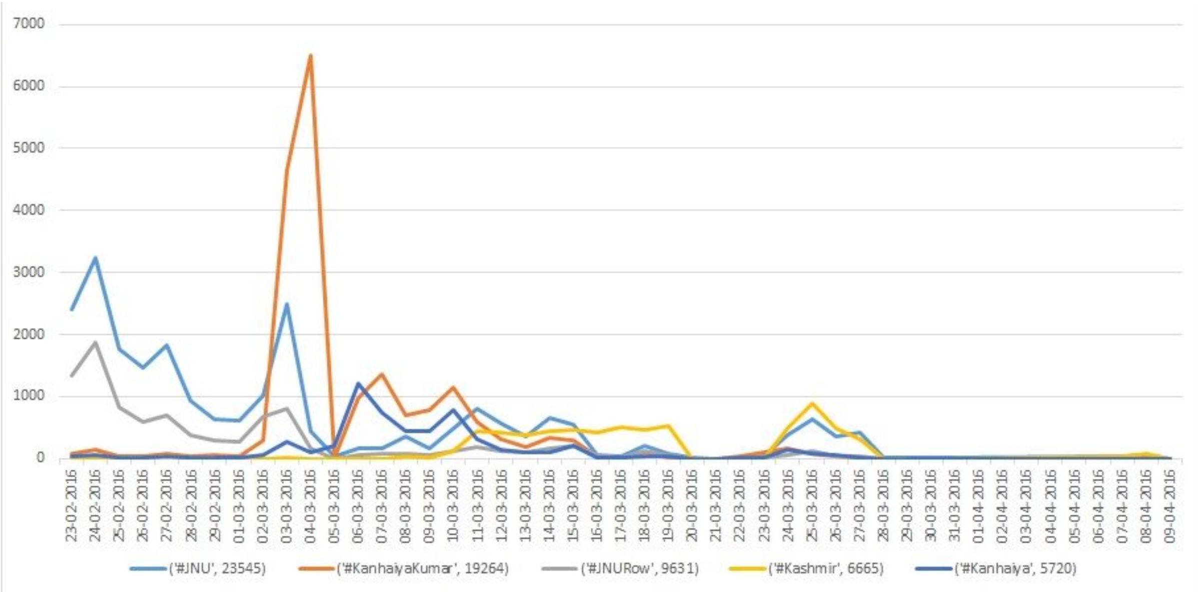


Figure: 9 . Day-wise use of hashtags (top 5 most frequently used hastags)

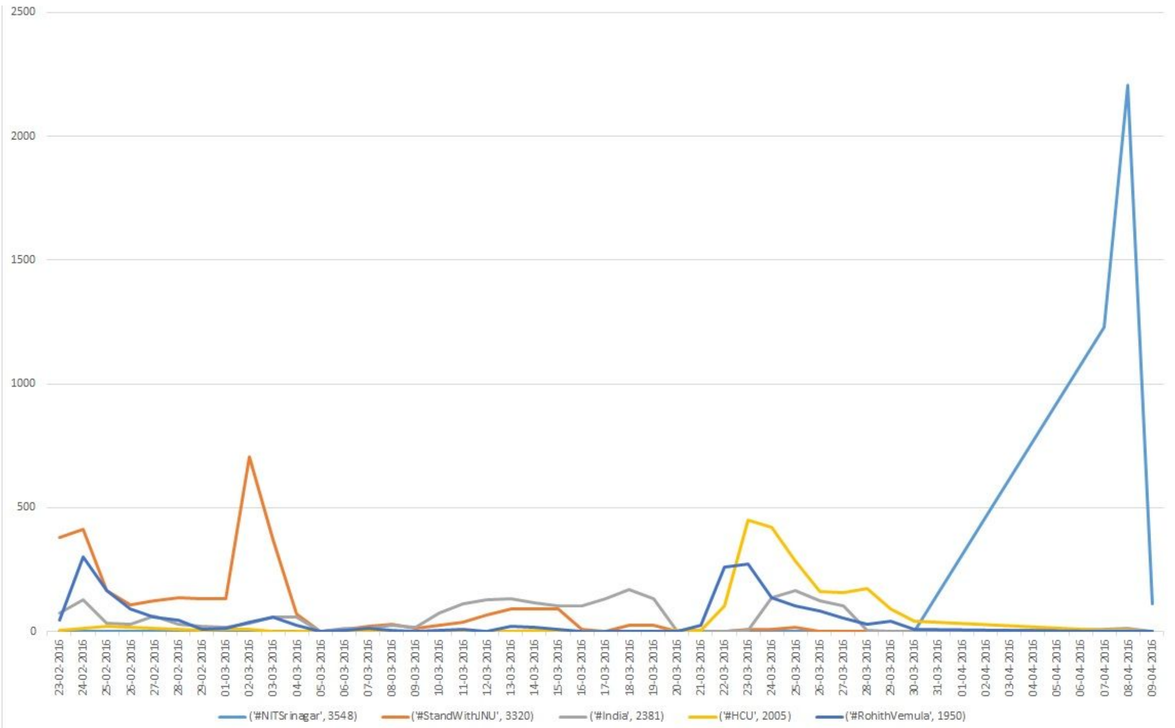


Figure 10. Day-wise use of hashtags (5th-10th most frequently used hashtags)

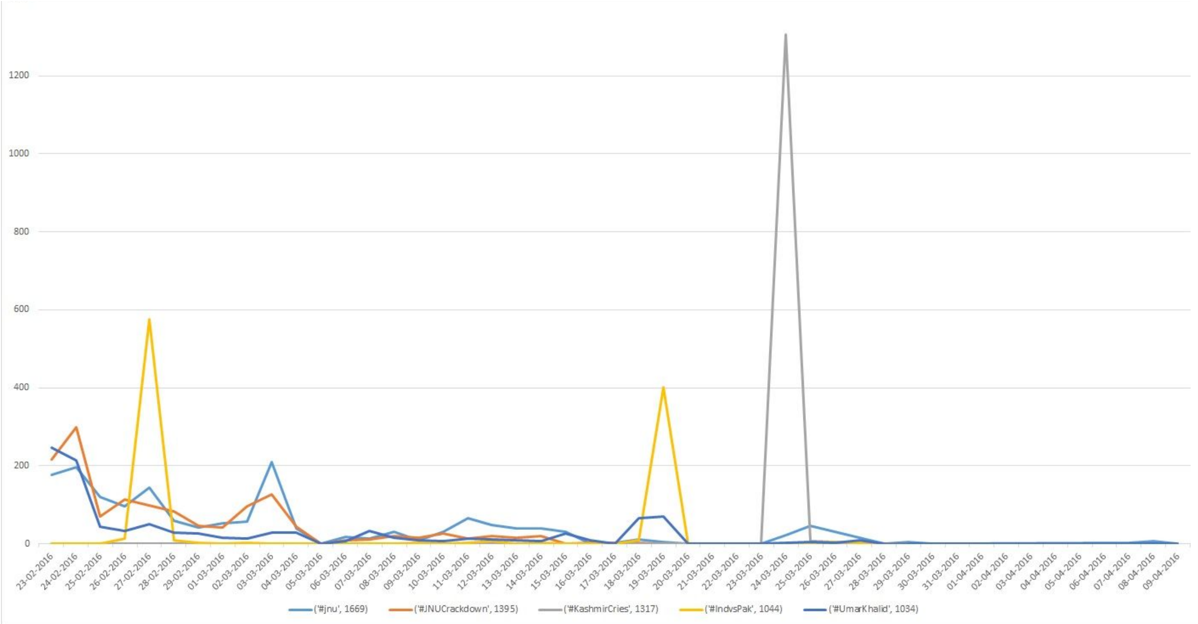


Figure 11. Day-wise use of hashtags (10th-15th most frequently used hashtags)

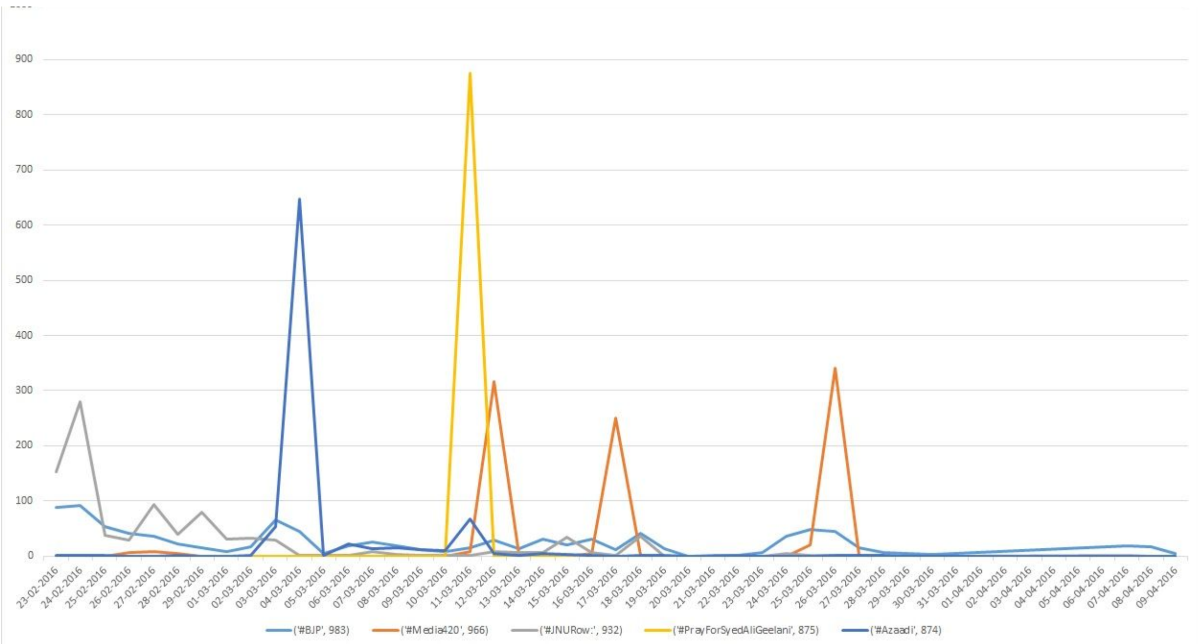


Figure 12: Day-wise use of hashtags (15th-20th most frequently used hashtags)

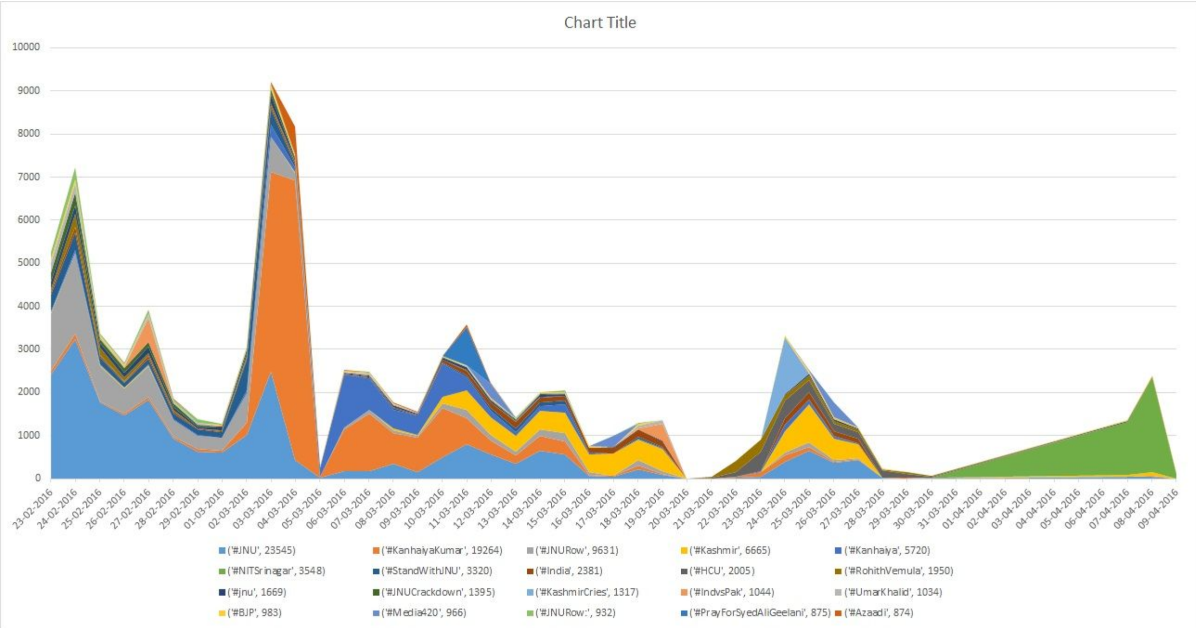


Figure 13. Day-wise use of hashtags (top 20 most frequently used hashtags)

N-Gram Analysis

Unigram	Bigrams	Trigrams	Start Date	End Date
[('JNU', 12146), ('JNURow', 5789), ('students', 1818)]	[(('JNU', 'students'), 892), (('Umar', 'Khalid'), 621), (('anti', 'national'), 424)]	[(('Umar', 'Khalid', 'Anirban'), 363), (('Khalid', 'Anirban', 'Bhattacharya'), 302), (('Anirban', 'Bhattacharya', 'surrender'), 155)]	2/23/16	2/29/16
[('KanhaiyaKumar', 12879), ('JNU', 5445), ('Kanhaiya', 2797)]	[(('KanhaiyaKumar', 'JNU'), 482), (('KanhaiyaKumar', 'speech'), 467), (('don', 't'), 435)]	[(('hey', 'follow', 'watch'), 420), (('follow', 'watch', 'followback'), 420), (('followback', 'Kanhaiya', 'Messi'), 420)]	3/1/16	3/6/16
[('JNU', 15882), ('Kashmir', 11443), ('Kanhaiya',	[(('Kanhaiya', 'Kumar'), 2503), (('anti', 'national'), 1631), (('Jammu',	[(('Pakistan', 'occupied', 'Kashmir'), 751), (('Chinese', 'Army', 'spotted'), 698),	3/7/16	3/13/16

6367)]	'Kashmir'), 1298)]	(('Army', 'spotted', 'along'), 695)]		
[('JNU', 18888), ('Kashmir', 14404), ('Kanhaiya', 6810)]	[(('Kanhaiya', 'Kumar'), 2810), (('anti', 'national'), 2039), (('Jammu', 'Kashmir'), 1883)]	[(('Pakistan', 'occupied', 'Kashmir'), 897), (('Chinese', 'Army', 'spotted'), 722), (('Army', 'spotted', 'along'), 720)]	3/8/16	3/14/16
[('Kashmir', 12247), ('Umar', 5727), ('JNU', 5386)]	[(('Umar', 'Khalid'), 4622), (('Jammu', 'Kashmir'), 2688), (('Khalid', 'Anirban'), 2281)]	[(('Umar', 'Khalid', 'Anirban'), 2274), (('Khalid', 'Anirban', 'Bhattacharya'), 1492), (('students', 'Umar', 'Khalid'), 717)]	3/15/16	3/21/16
[('Kashmir', 19683), ('JNU', 11850), ('BJP', 5044)]	[(('Jammu', 'Kashmir'), 4713), (('anti', 'national'), 3159), (('Kanhaiya', 'Kumar'), 2329)]	[(('Government', 'Jammu', 'Kashmir'), 666), (('Form', 'Government', 'Jammu'), 664), (('BJP', 'Stake', 'Claim'), 643)]	3/22/16	3/30/16

In the above n-grams we can clearly see that there is a shift to Kashmir as the most popular unigram, and that the entire movement has been colored in anti national sentiments, and how politics has played a major role in doing the same. We also observe, that any one pro JNU movement has been called as anti national.

Conclusions

We have presented an analysis of twitter users behavior during the event of a crisis that has gripped the entire nation. We could see that many users have gone on to express their views on the JNU controversy. We used the tweets to identify users whose tweets have had an influence on the people. This was done by using the retweets count as the indicator. We then performed LIWC analysis on their tweets that has had most impact on users and tried to analyze the behavior of people during these controversies. We also perform LIWC analysis on the complete data period to understand the effects of tweets. We have manually mapped some of the important dates with the events that have been gathered from press and media.

While trying to understand people's behavior during a national controversy was the [main idea, our observation showed a rise in negative affect and anger during some events in certain time period. This shows how people have used social network to display their opinions and views on a sensitive issue like the JNU controversy. The work also shows how media and other influential users play a major role in changing people's opinion towards a situation. We have provided a brief discussion on how emotions of people have changed when the media and other users have expressed events. It will be interesting to see how the work can be extended for bigger movements and events around the world. We could easily trace the impact of media and influential social media users on the views and emotions of normal users and people.

Using n-grams and hashtag analysis we can easily see the shift in the topics being discussed, and can track the movement. The tweeting behavior is consistent with the events and speeches by movement celebs. We were also able to develop intriguing insights into LIWC values of most famous tweets, and urls which were most shared.

There are many ways in which this work can be extended. An interesting observation would be to see people who are on the periphery, and what they think, how the sentiment spreads i.e. if any event has lead to these people in the periphery getting involved. Also, we can analyze the difference in opinion and shift of users from one opinion group to another. The data we gathered is for a limited time. We believe that we could have better analysis provided the data was large and spanned over a longer period of time.

Acknowledgement

We would like to thank Professor Munmun De Choudhury for her comments and ideas during the completion of this project.

References :

[1] Munmun De Choudhury, Andrés Monroy-Hernández, and Gloria Mark. 2014. "Narco" emotions: affect and desensitization in social media during the mexican drug war. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14). ACM, New York, NY, USA, 3563-3572. DOI=http://dx.doi.org/10.1145/2556288.2557197.

[2] Ahmed, Saifuddin and Kokil Jaidka. 2013. "The Common Man: An Examination of Content Creation and Information Dissemination on Twitter during the 2012 New Delhi Gang-Rape Protest" *Digital Libraries: Social Media and Community Networks* 82(79):117-126.

[3] *'Characterizing the dynamics of a protracted activist movement via social media'* by Shagun Jhaver, Munmun De Choudhury and Benjamin Sugar.