Course – Social Network Analysis (BIA 658)

Group E:

Team Members – Naman Jain, Dmitry Rassakhatsky, Ruihan Lu, Yijing Mao, Tengjiao Zhu

Guided By - Prof. Hani Safadi

Date: December 20, 2014

**Deepwater Horizon Oil Spill**

**Twitter Analysis**

**SOME FACTS ABOUT THE BP OIL SPILL**

1. In the BP Oil Spill, more than 200 million gallons of crude oil was pumped into the Gulf of Mexico for a total of 87 days, making it the biggest oil spill in U.S. history.

2. 16,000 total miles of coastline have been affected, including the coasts of Texas, Louisiana, Mississippi, Alabama, and Florida.

3. Even though the gushing well was capped in July 2010, oil is still washing up on shores, which might cause long-term damages to people living in the area.

4. The initial oil rig explosion killed 11 people and injured 17 others.

5. President Obama announced that his administration would create a $20 billion spill response fund.

6. Responders used 5.5 million feet of boom, a barrier placed in water, to collect and absorb oil.

7. Of the 400 miles of Louisiana coast, approximately 125 miles have been polluted by the oil spill.

8. A method of treating the oil spill is "in-situ burning" or burning oil in a contained area on the surface of the water, which has negative effects on the environment.

9. Over 8,000 animals (birds, turtles, mammals) were reported dead just 6 months after the spill, including many that were already on the endangered species list.

10. BP is responsible for close to $40 billion in fines, clean-up costs, and settlements as a result of the oil spill in 2010, with an additional $16 billion due to the Clean Water Act.

11. Over 30,000 people responded to the spill in the Gulf Coast working to collect oil, clean up beaches, take care of animals and perform various other duties. As of 2012, the Gulf was still polluted with oil.

**ANALYSIS**

The original dataset consists of four tables viz. HASHTAG, MENTION, TWEET, TWEETER.

Bird’s eye view of the data set:

|  |  |  |  |
| --- | --- | --- | --- |
| Tables | Columns | No. of Rows | Unique |
| HASHTAG | tweetid , hashtag | 3,37,441 | hashtag – 12,968  tweetid – 1,39,790 |
| MENTION | tweetid , mention | 1,94,606 | mention - 18,384  tweetid – 1,41,683 |
| TWEET | tweetid, tweeter, content, date, retweet, score, firstpost | 72,740 | tweeter – 13,662  tweetid - 72,740 |
| TWEETER | tweeter , type | 675 | tweeter - 675 ; type - 7 |

On initial analysis we found out that the data was not clean, as can be seen above.

The query used to find the Unique fields is:-

**HASHTAG:**

Select count(distinct hashtag) from HASHTAG;

Select count(distinct tweetid) from HASHTAG;

**MENTION:**

Select count(distinct mention) from MENTION;

Select count(distinct tweetid) from MENTION;

**TWEET:**

Select count(distinct tweeter) from TWEET;

Select count(distinct tweetid) from TWEET;

**DATA PREPROCESSING**

**MENTION NETWORK**

The MENTION table contains 194,606 rows, which is a very large dataset for analysis and also it does not contain all the data required for analysis. So the relevant data containing tweeters and mention was extracted for proper analysis.

*Firstly*, we matched TWEET and MENTION table on the basis of *tweetid*(Inner Join) to extract the tweeters and tweetid that were only present in the tweet table. This query gave us 68,046 rows, which is still large for analysis purposes.

Select TWEET.tweetid,TWEET.tweeter,MENTION.mention from TWEET,MENTION where TWEET.tweetid = MENTION.tweetid and TWEET.date between '2010-04-22 00:00:00' and '2010-07-14 19:55:43'; 🡪 Name this table as TWEETER\_MENTION

*Secondly*, on analysing this data in NODEXL we found out that there are many self-loops that does not make any sense in the Mention network. So we deleted the Retweets from the table by the below query. This got us to around 52,000 rows.

SELECT \* from TWEETER\_MENTION where TWEETER\_MENTION.tweeter != TWEETER\_MENTION.mention;

🡪 Name this table as TWEETER\_MENTION\_CLEAN

*Thirdly*, as the data is too large, so we decided to only include the mentions which have been mentioned more than once. The data now contains 5,100 rows. This method would allow us to only concentrate our analysis on important tweeters.

Select TWEETER\_MENTION\_CLEAN.tweetid, TWEETER\_MENTION\_CLEAN.tweeter,TWEETER\_MENTION\_CLEAN.mention,count(mention) from TWEETER\_MENTION\_CLEAN GROUP BY tweeter , mention having count(mention)>1 order by count(mention) desc; 🡪 Name this table as MENT

*Lastly*, as we needed type of tweeter too for analysis, so we matched above table with TWEETER with the below query,

Select MEN.tweetid,MEN.tweeter,MEN.mention,TWEETER.type from MEN left join TWEETER on MEN.tweeter = TWEETER.tweeter;

Finally we have the data for Mention network analysis. The data now contains 3,548 rows, which is perfect for analysis.

**HASHTAG and AFFILIATION NETWORK**

The HASHTAG table provided initially had 337,441 rows, as can be seen from the table above. As this data is too large to perform analysis, so we filtered the data.

The sequence of queries performed to extract the required data are,

First we extracted data from ‘TWEET & HASHTAG’ table on the basis of tweetid.

Description of query**:**

1. In the first query we matched tweetid from TWEET and HASHTAG table. This provided us 67,484 rows.

2. In this query we just included one more column ‘type’ from the TWEETER table.

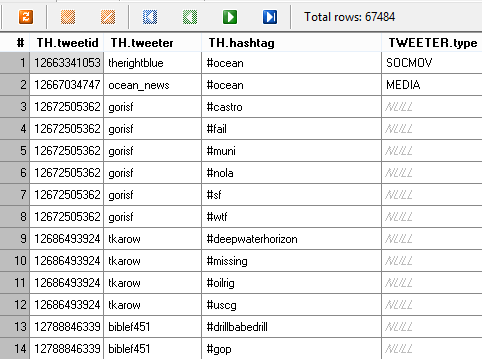
3. As the data set has too many records to make sense, the data was further reduced by only including records that contains no. of hashtag greater than 200. Thus we got top 37 hashtags.

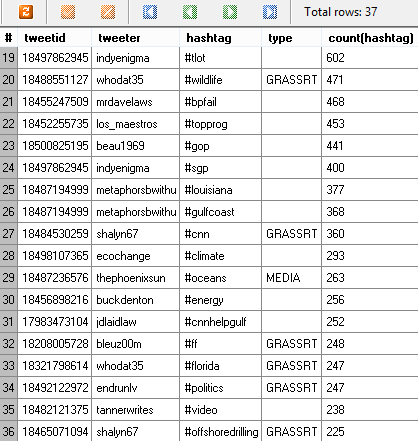
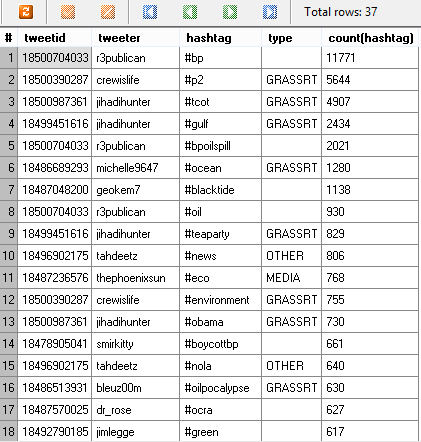
4. Including other fields with top hashtags. Here we reduced the no. of records from 67,484 to 43,613.

5. Still the dataset is large for analysis. Also, for better visualization purposes the dataset was further reduced to include only hashtags used more than once. This brought down no. of records to 4,415. This no. finally seemed fine for analysis.

Select TWEET.tweetid , TWEET.tweeter , HASHTAG.hashtag from TWEET,HASHTAG where TWEET.tweetid = HASHTAG.tweetid; 🡪 Name this table as TH

SELECT TH.tweetid , TH.tweeter , TH.hashtag , TWEETER.type from TH left join TWEETER on TH.tweeter = TWEETER.tweeter; 🡪 Name this table as TWTHSH



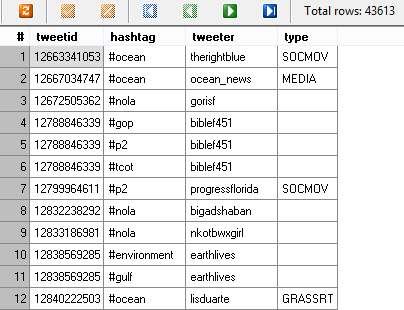
select tweetid , tweeter , hashtag , type , count(hashtag) from TWTHSH group by hashtag having count(hashtag)>200 order by count(hashtag); 

select tweetid, hashtag , tweeter, type from TWTHSH where hashtag in (

select hashtag from TWTHSH 🡪 Name this table as TOP\_HASH

group by hashtag

having count(hashtag) > 200);



select distinct hashtag , tweeter , type , count(hashtag)

from TOP\_HASH

group by hashtag , tweeter 🡪 Name this table as TOP\_HASH\_CLEAN

having count(hashtag) > 1

order by count(hashtag) desc;



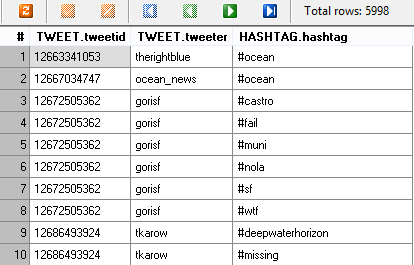
Query**:**

*First 15 days of HASHTAG network:*

Select TWEET.tweetid , TWEET.tweeter , HASHTAG.hashtag from TWEET,HASHTAG where TWEET.tweetid = HASHTAG.tweetid and TWEET.date between '2010-04-22 18:11:55' and '2010-05-07 23:59:59';

*Last 15 days of HASHTAG network:*

Select TWEET.tweetid , TWEET.tweeter , HASHTAG.hashtag from TWEET,HASHTAG where TWEET.tweetid = HASHTAG.tweetid and TWEET.date between '2010-06-31 18:11:55' and '2010-07-14 19:55:43';



**MENTION NETWORK**

1. **Attributes and network structure.**

The mention network is directed since the relationship that one mentions another is not necessarily reciprocated. The weight of edges are the frequency a tweeter was mentioned by the same tweeter in different tweet. For example, tweeter A may mention tweeter B in three different tweets, so the weight of edge between A and B is three. In this mention network, nodes are tweeters, edges represent mention relationship. There are reciprocal exchanges in the mention network. For example, “newshour” mentioned “annashoup” reciprocally, “annashoup” mentioned “newshour”.

1. **Network structure.**

A network exhibits a small world structure if it has high average clustering coefficient and short average shortest path length. In our case, the network does not exhibit small world structure because average clustering coefficient is 0.11 and average closeness centrality is 0.005, which means it has low clustering coefficient and relatively high shortest path length. Therefore, the network does not exhibit small world structure.

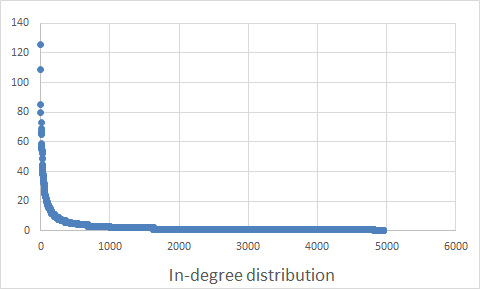


Figure 1 In-degree distribution

The maximum in-degree is 126. The minimum in-degree is 0 and the average in-degree is 2.67. It means that the network has few very highly connected nodes and large number of nodes with low degree. The figure above shows a long tail feature.

Figure 2 Clustering Distribution

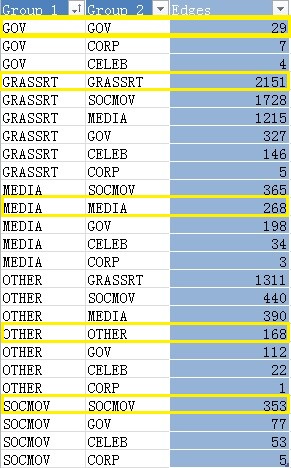
Since the clustering coefficient decreases as the node degree increases, there is evidence of power-law distribution.

1. **Influential Tweeters Analysis.**

Eigenvector centrality measures a node’s influence. It identifies the most influential tweeters by measuring the highest eigenvector centrality. In the mention network, CORP has the highest average eigenvector centrality, so CORP is more influential on aggregate. It tells the truth that when the crisis happened, corporations and celebrities played an important role in online media. They are mentioned frequently and made this a hot topic on twitter. The oil spill thing has complex relationship with oil and carrier corporations, so it is normal that these corporations were active in social media. The most influential tweeters from each group are shown below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **The most influential tweeter** | **The tweeter’s eigenvector centrality** | **The average eigenvector centrality/group** |
| CELEB | zaibatsu | 0.030 | 0.009 |
| CORP | bp\_america | 0.083 | 0.023 |
| GOV | gohsep | 0.027 | 0.002 |
| GRASSRT | seachele420 | 0.01 | 0.001 |
| MEDIA | thephoenixsun | 0.055 | 0.002 |
| OTHER | learnfromnature | 0.037 | 0.001 |
| SOCMOV | therightblue | 0.042 | 0.002 |

1. **Group Communication.**



The above table which we obtained from NodeXL gives us the overall group communication. On aggregate, edges flowing within group or communications within group is *2969*, while average communications across group is *6443*. As the number for across group is more than the within group, this means that communication across the group is much more than within groups. Hence, communication is more concentrated across groups.

1. **Communities and predefined groups.**



Figure 3 Predefined groups by type

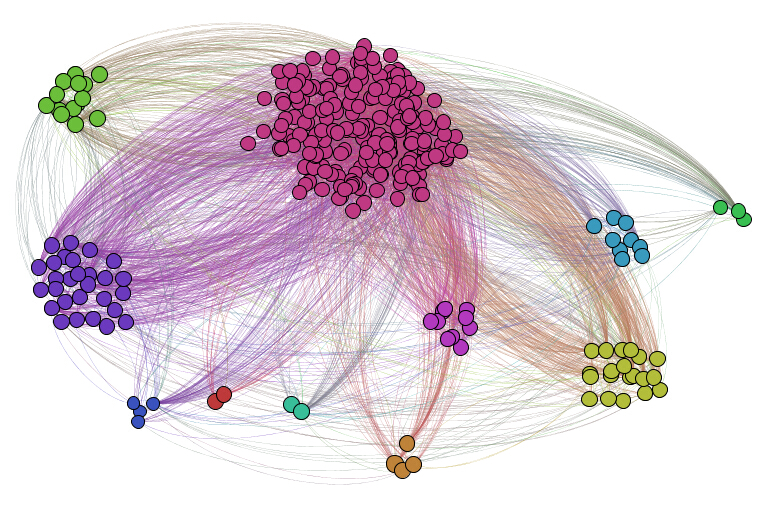
****

Figure 4 Communities

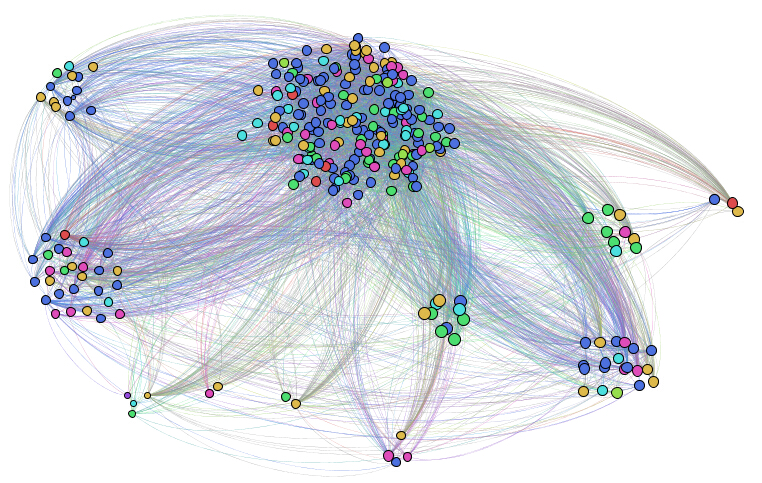
****

Figure 5 Nodes of predefined groups in communities

*Figure 1* shows 7 predefined groups and one additional group named NONE, which are random tweeter users that don’t belong to any group.

If we do not consider NONE group, then GRASSRT group plays an important role in the oil spill disaster, which means that tweeters in this group are kind of opinion leaders and very much active in social network.

In *figure 4* there are 11 communities based on communication pattern. Different color stands for different communities. Comparing it to the predefined groups, the *figure 5* is the nodes from predefined groups, but in 11 communities. 1The color of nodes here represents the seven types of predefined groups. Dark blue represents GRASSRT, red is GOV, purple is CORP, light blue is OTHER, lime is SOCMOV, lawn green is CELEB, medium red is MEDIA and orange is none. These nodes randomly sit in the 11 communities. The 11 communities are not correspond to the predefined groups.

1. **Mention network sentiment.**

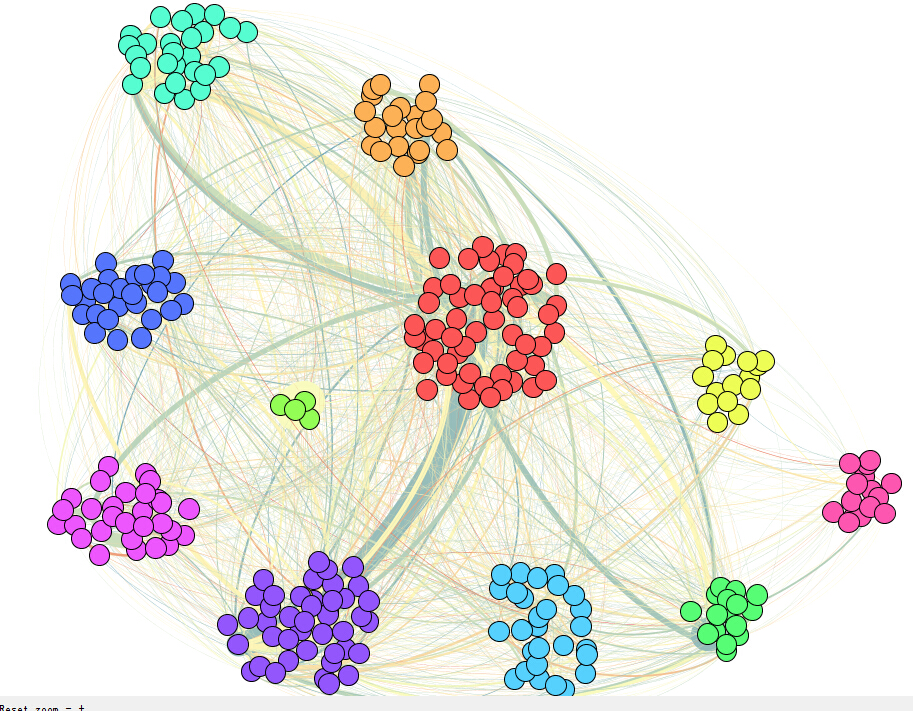


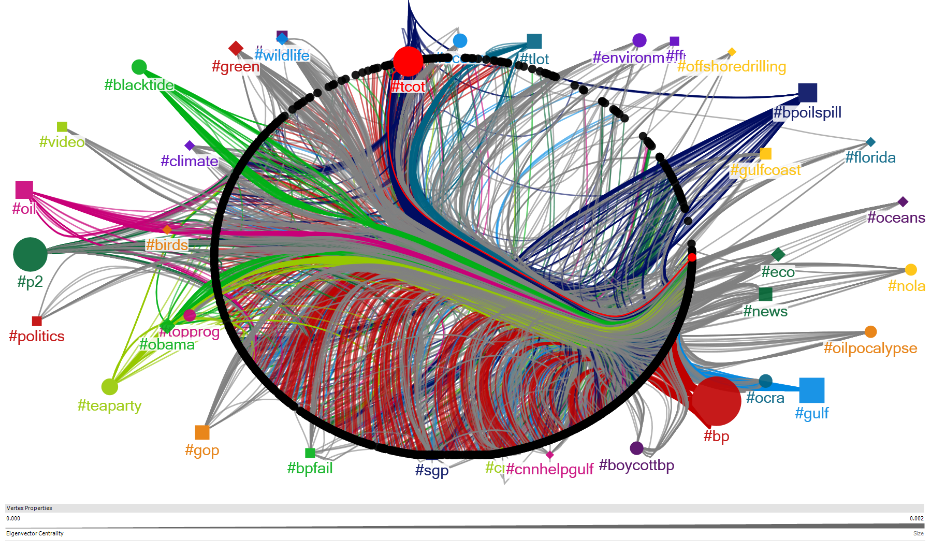
Figure 6 Mention network sentiment

The sentimental value range from negative to positive, the colour follows it rang from blue to yellow. The above graph shows the 11 communities’ sentiment on the crisis. The blue is more obvious than yellow and sum of all the sentiment value is negative. It means the opinion trend to negative on this event. It makes sense that oil spill must have negative influence on both economy and politics, so social network discussion probably worry about its directed and side effects.

**HASHTAG NETWORK**

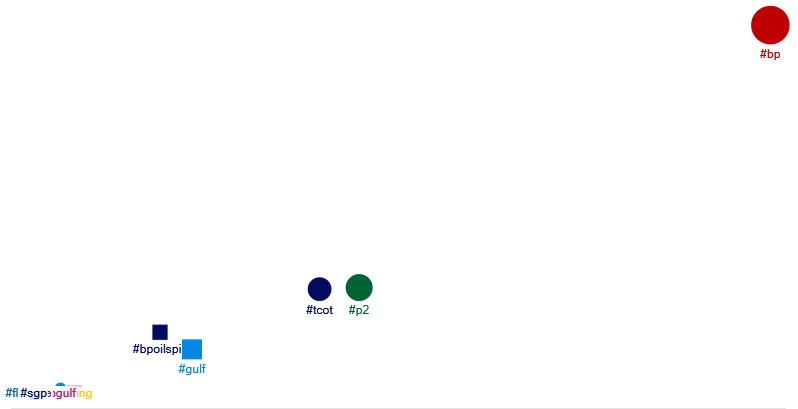
1. **Hashtag Network.**

The below graph represents the undirected hashtag network between tweeters and the hashtags they used. The black circle in the centre represents tweeters, and the coloured representations are different hashtags. The top 10 hashtags’ edges are coloured for better visibility, rest are grey. Furthermore, the size of the nodes of the hashtags are based on eigenvector centrality. The larger the eigenvector centrality, the bigger the hashtag. Moreover, to make the graph more appealing, dynamic filter of degree 2 was used.



1. **Most frequent hashtags.**

As can be seen from the graph below, the most frequent hashtags are #bp>#p2>#tcot>#gulf>#bpoilspill>… and so on. Moreover it can be seen from the screenshots in the analysis part that #bp was used 11,771 times, #p2 was used 5,644, #tcot – 4,907 times, #gulf – 2,434 times and #bpoilspill – 2,021 times respectively.

 Degree ------------->

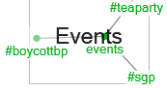
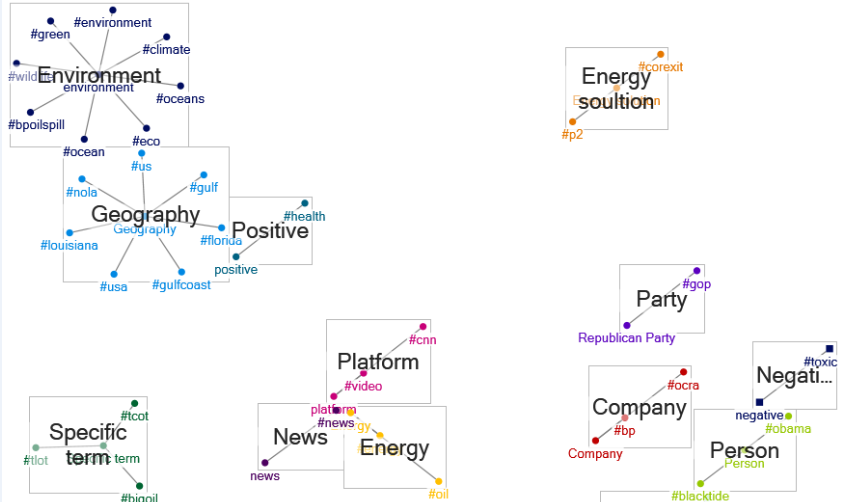
X-Axis -------- Degree Y-Axis------- Betweenness Centrality

The table below shows thirty-five hashtags that have highest degree and also contains the themes they have been classified to.

|  |  |  |
| --- | --- | --- |
| Node | Theme | Degree |
| #bp | Company | 3102 |
| #gulf | Geography | 1406 |
| #p2 | Energy solution | 1076 |
| #tcot | Specific term  (Top Conservationist on Twitter) | 955 |
| #oil | Energy | 836 |
| #blacktide | Person | 759 |
| #obama | Person | 656 |
| #bpoilspill | environment | 574 |
| #ocean | environment | 482 |
| #environment | environment | 470 |
| #eco | environment | 462 |
| #green | environment | 443 |
| #teaparty | events | 388 |
| #news | news | 378 |
| #corexit | Energy solution | 360 |
| #gop | Republican Party | 346 |
| #tlot | Specific term  (Top Libertarians on Twitter) | 346 |
| #nola | Geography | 335 |
| #wildlife | environment | 301 |
| #boycottbp | events | 291 |
| #florida | Geography | 289 |
| #cnn | platform | 257 |
| #energy | Energy | 239 |
| #louisiana | Geography | 231 |
| #sgp | events | 227 |
| #ocra | Company | 227 |
| #usa | Geography | 226 |
| #climate | environment | 223 |
| #video | platform | 219 |
| #gulfcoast | Geography | 208 |
| #health | positive | 205 |
| #us | Geography | 193 |
| #oceans | environment | 193 |
| #toxic | negative | 185 |
| #bigoil | Specific term | 175 |

1. **Group by themes.**

The team picked up high degree hashtags into new file and filtered some hashtags in which degree centrality is low to draw the group picture.



The group majors in company, environment, geography, specific term, energy solution, energy, news and person.

1. **First and last 15 days.**

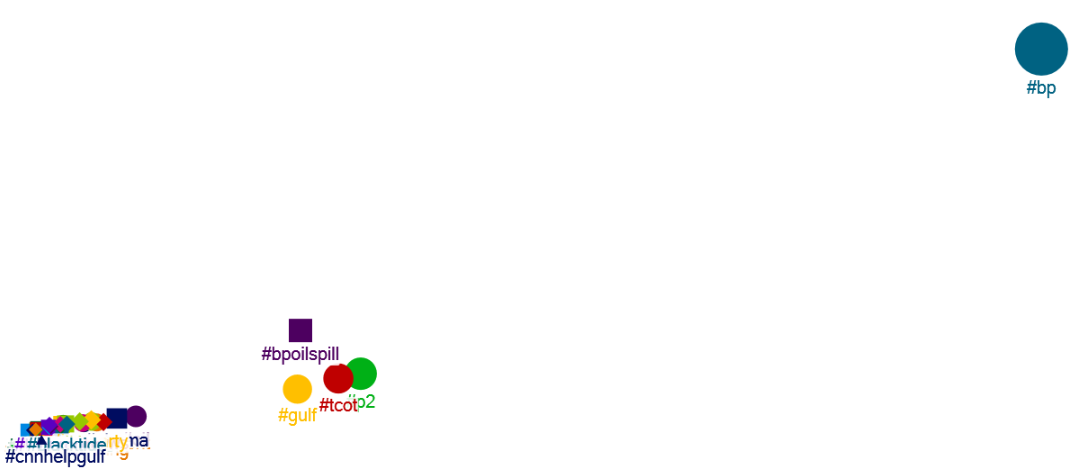


Figure 7 First 15 days – hashtags

X-Axis -------- Degree Y-Axis------- Betweenness Centrality

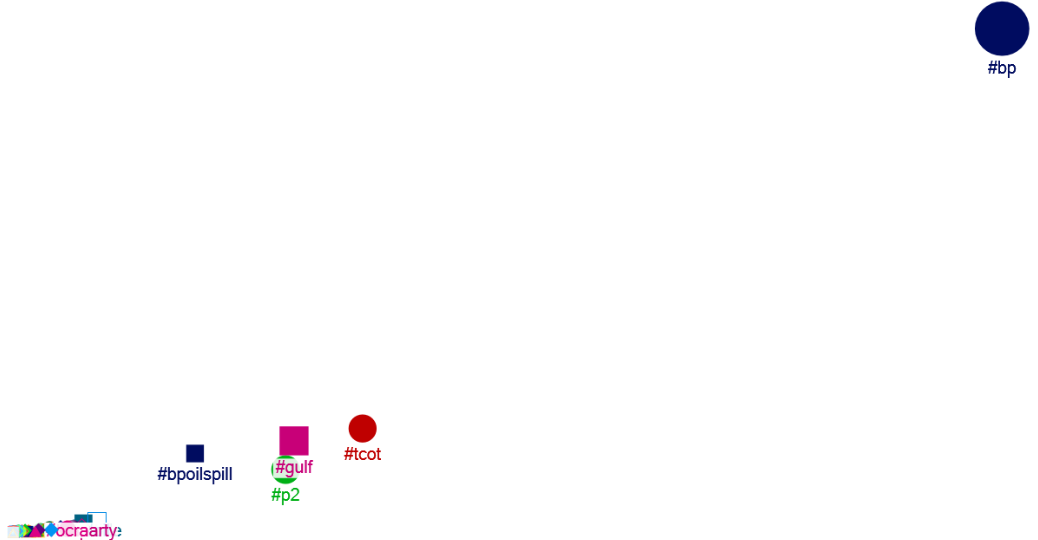


Figure 8 Last 15 days – hashtags

X-Axis -------- Degree Y-Axis------- Betweenness Centrality

Comparing the above graphs it can be seen that hashtag #bp is dominating in both the first and last 15 days. It shows that even after nearly 3 months people are equally concerned with oil spill. People also used other hashtags like #tcot, #p2, #bpoilspill, #gulf and so on, more frequently in the first 15 days as they are closely concentrated in the graph. This can be attributed to the fact that as the event had just occurred people are tweeting more about these hashtags to show their concern for the environment and spread the awareness. However in the last 15 days there is quite a spread between them. This shows that as the time passed people seem to be less concerned about the oil spill.

**AFFILIATION NETWORK**

The Affiliation network models the association between tweeters and the hashtags they use in their tweets. This network would represent between various stakeholders and their main concerns or positions in the crisis.

The nodes represents tweeters(classified in 8 groups) and hashtags. For better visualization purposes only top 8 hashtags are shown. Edge represents tweeters of a particular who used particular hashtag in their tweets. Edges have been coloured with the same colour as the tweeter group for better visualization purpose. The size of the hashtag represents its frequency of occurrence.

On examining the network it was found out that there are 8 groups instead of given 7, with the 8th group classified as NONE. This group is named NONE because tweeters in this group doesn’t belong to any of the given 7 groups: GRASSRT,OTHER,MEDIA,SOCMOV,GOV,CELEB,CORP. So these group could belong to normal people.

On further analysing the graph it can be seen that group NONE has the largest number of tweeters, followed by GRASSRT>OTHER>MEDIA>SOCMOV, with GOV, CELEB and CORP having very few tweeters.

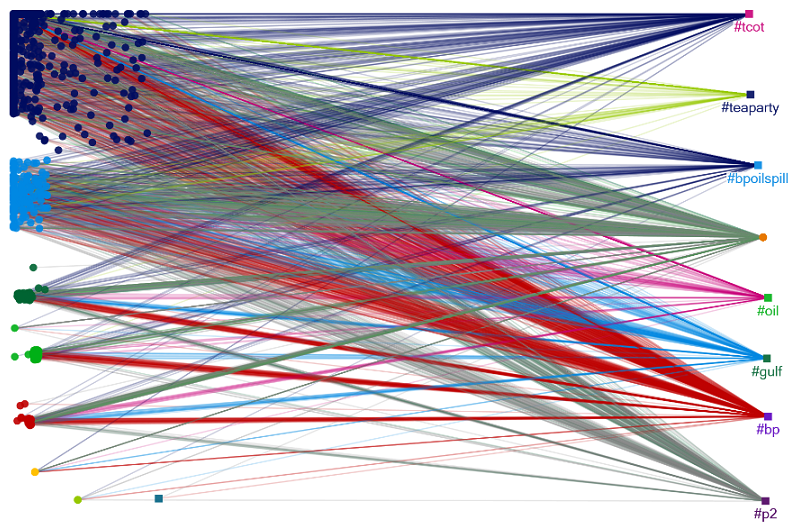
Furthermore, on analysing the hashtags we can see that *‘#bp’* was tweeted by tweeters from all groups with the largest from NONE, and then GRASSRT.

‘*#tcot’* was tweeted by 5 groups and not tweeted by CELEB and CORP. It was used more by NONE, GRASSRT and OTHER groups. By doing some research on this hashtag we found out that it is an acronym for “Top Conservatives on Twitter”. When used in the hashtag form, the term provides a way for conservatives in particular and Republicans in general to locate and follow the tweets of their like-minded brethren.

‘*#p2’* was tweeted by NONE,GRASSRT,OTHER,MEDIA, SOCMOV, and not tweeted by GOV and CELEB. .This hashtag stands for “Progressive hashtag on twitter”.

**Some Interesting Insights on ‘*#tcot’* and ‘*#p2’*** - TCOT and P2 folks are people who construct and disseminate the political memes of the day – perhaps by minute. Political pundits, mainstream media and the cable news channels religiously follow them. One more pretty interesting thing we found out is that **‘*#p2’*** is used by Conservatives with negative comments to Liberals.

*‘#bpoilspill’* was also tweeted by many groups, but from predefined groups it was mostly tweeted by GRASSRT and OTHERS.

****

H

A

S

H

T

A

G

S

S

CORP

CELEB

GOV

SOCMOV

MEDIA

OTHER

GRASSRT

NONE

Figure 9 Affiliation network

## SENTIMENTAL ANALYSIS

To determine the emotion coloring of the BP oil spill in Tweeter the sentiment analysis for each tweet was made. The analysis was made using AFINN dataset, which contains a list of words that rated for valence with an integer between minus five (negative) and plus five (positive). However, the dataset contains only sentiment levels for individual words. So, for the calculation of the sentiment level for each tweet message the proposed method was used. Calculation of sentiment level per each tweet was made using Python script ([link to script](http://cl.ly/code/2C3Y44300O3V/SentimentAnalysis.py)). For further use of the obtained data, all results were stored in the dictionary object.

All tweeter users in the dataset were divided into the following groups:

* SOCMOV – social movement organizations;
* MEDIA – media companies;
* CELEB – celebrities;
* CORP – corporations;
* GOV – government;
* GRASSRT – grassroots;
* OTHER – twitter accounts for stakeholders who have an established identity outside of twitter;

However, during the data analysis it was revealed that the majority of users do not classified. For this analysis those user were forced into the OTHER group. See average sentiment level across the groups below.

Table 1 Average sentiment level across the groups

|  |  |  |
| --- | --- | --- |
| **Group** | **Avg. Sentiment Level** | **Number of Tweets** |
| MEDIA | -0.03754815704400687 | 5686 |
| GRASSRT | -0.06599101947108077 | 18910 |
| CELEB | 0.020745799301055284 | 263 |
| OTHER | -0.05136092580821749 | 40728 |
| CORP | 0.0715488375170081 | 286 |
| GOV | 0.03512297618834427 | 887 |
| SOCMOV | -0.05743605823622334 | 5980 |

Figure 10 Average sentiment level across the groups

The analysis of the level of sentiment in groups shows that the sentiment level in general for each group was pretty neutral. The GRASSRT group has the lowest sentiment level, however it is so close to zero that can be treated as neutral. The same pattern can be found in other groups. However, it is only means that tweets in general were not very emotional.

The next step of the sentiment analysis was the sentiment levels across different hashtags. The analysis of the original data showed that some tweeter users do not use space between hashtags, so the hashtags #stevens and #bia could be also written as #stevens#bia. In order to take into account each hashtag correctly in this case that hashtags were treated as two different hashtags. In addition, the HASHTAG table does not contain all hashtags that were used in tweets, that is why for this analysis a new hashtag map was created. See the graph plot with top 25 hashtags below. Hashtags are sorted by popularity.

Macintosh HD:Users:rassakhatsky:Downloads:click_to_enter_plot_title.pdf

Figure 11 Average sentiment levels across the hashtags – Top 25

The graph plot shows that sentiment level for most hashtags is slightly below zero, however some hashtags have a positive sentiment level. The #bp hashtag is the most widely used, which was expected. For this reason it is quite neutral. The most positive hashtag is #cnnhelpsgulf that was used as an advertisement of Larry King show on CNN ([link](http://larrykinglive.blogs.cnn.com/2010/06/17/lkl-monday-2-hour-disaster-in-the-gulf-how-you-can-help-special/)). As we can see, in advertising campaigns media try to use words with positive connotations. The tweet data analysis shows that most tweeter users just retweet a message with the #cnnhelpsgulf hashtag. The #boycottbp was the most negative among the top hashtags.

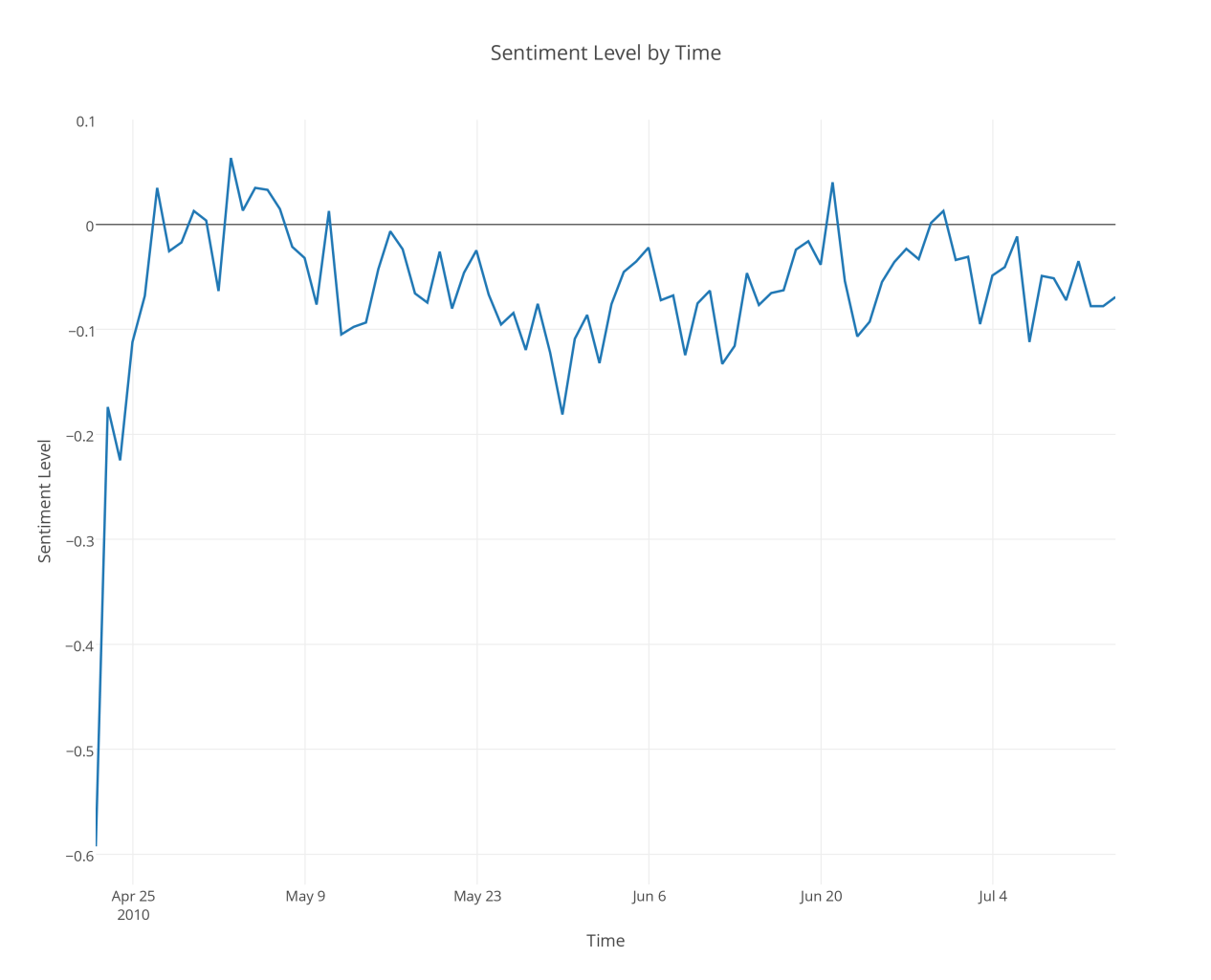


Figure 12 Sentiment Level by Time

The graph above shows us how the average sentiment level changed by time. It is clear that overtime they are getting more positive. The data frame is from April 22 until July 14. The lowest sentiment value was reached on April 22, when the information comes out. In addition, during the first month, high fluctuation of sentiment’s level can be notice, over time they disappeared, and average sentiment level became more stable.

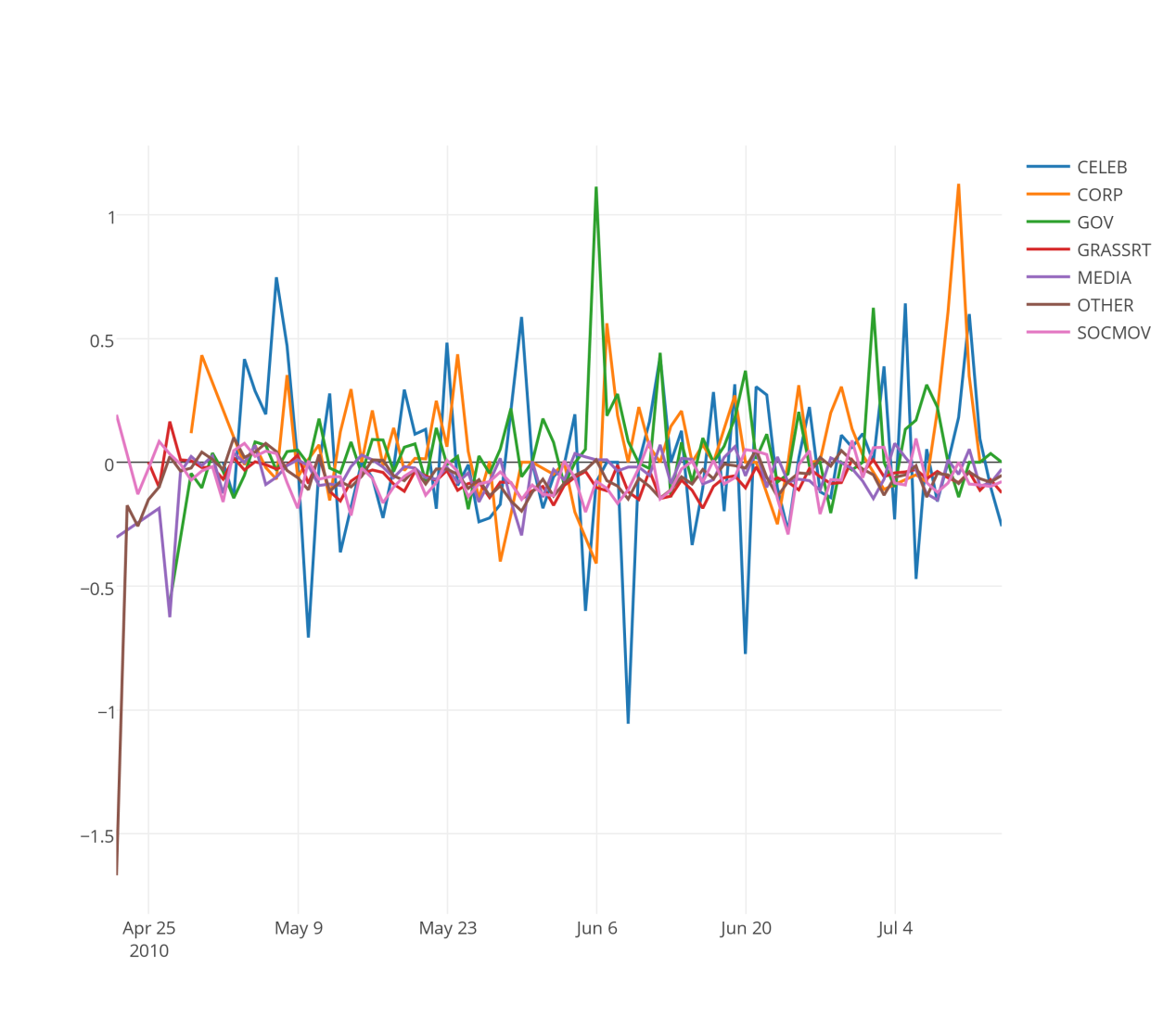


Figure 13 Group Sentiment Level by Time

More information can be extracted from the graph when we analyze sentiment changing by group type. Other group was very negative initially, however after a while it started fluctuate around 0. Corporation and Government groups were positive almost all the time. Celebrities were the most volatile group, moreover the average sentiment level for them was only extremely positive or extremely negative. All other groups were fluctuated around 0 and didn’t have any significant fluctuations.

## IMPLICATIONS

From the above analysis, we can conclude that social media is an effective tool for public to engage in social events. In the oil spill dataset, tweeters of different groups communicated with each other. Communication is more concentrated across groups than within groups. Group CORP is the most influential group in the mention network. Tweeter “pbhnetwork” (The PHB Network) which is a media company has been mentioned most in this dataset. That is because the PHB Network has posted a lot of information on Twitter and has been an active media in this oil spill event. When we analyzed the hashtag in the network, “#bp” has been the most frequently used hashtag.

We also learned that the data in the real world is not clean and needs to be pre-processed for analysis. This means that we need to prepare from the initial raw data the final data set that is to be used for subsequent phases. In our data set there were four tables having different data that needed to be correlated to get the required data for proper analysis. Also, the most important part is the Business Understanding phase where we enunciate the project objectives and requirements clearly in terms of business, translate these objectives, and then prepare a preliminary strategy for achieving these objectives.

Moreover, on seeing other teams’ presentations we also learned that there are numerous ways to extract and analyse the same dataset, and it all boils down to what the business wants.

**Most Influential Actors**

For finding out the most influential actors in the network, we calculated the eigenvector centrality. The higher the eigenvector centrality, the more influential the vertex is. The top ten is shown in Table 2. Tweeter ‘*whodat35’* has the highest eigenvector centrality, thus the most influential actor in this network. Typically, among these ten most influential actors, six are from group GRASSRT. They are active accounts on Twitter but do not correspond to established stakeholders in the oil spill event.

Table 2 Top ten eigenvector centrality

|  |  |  |
| --- | --- | --- |
| **Vertex** | **Eigenvector Centrality** | **Group** |
| Whodat35 | 0.009 | GRASSRT |
| Seachele420 | 0.009 | GRASSRT |
| winterthur | 0.007 | GRASSRT |
| therightblue | 0.006 | SOCMOV |
| nwf | 0.006 | SOCMOV |
| ibrrc | 0.006 | OTHER |
| oceanshaman | 0.006 | GRASSRT |
| supermanhotmale | 0.005 | GRASSRT |
| soloann | 0.005 | GRASSRT |
| smdindustries | 0.005 | OTHER |

Twitter is both a new medium for media companies to disseminate information and a platform for masses to express their ideas. Media companies use twitter to disseminate information faster and at the same time it helps them reach the wider audience across the globe. Also as many people on twitter follow these media companies, so they are frequently updated of the latest news.

Furthermore, it is also a platform for masses to openly express their ideas. And it is evident from the findings in the above question where we identified ‘*whodat35’* as the most influential tweeter in this oil spill disaster. On further analysis about this tweeter we found out that this person is a nature enthusiast.

Also, to further strengthen our analysis, we have shown a pie chart showing the distribution of tweets from different groups, including media companies which accounts for 9.9% and tweets from masses including celebrities, social movement organizations, government, corporations, grassroots and so on accounts for 90.1%.

Figure 14 Classifications of Tweets

**Social Movement Organizations – Twitter Call for Action**

We already know that Twitter acts like a platform for masses to express their ideas. We can continue checking whether Twitter has helped masses in real life. Taking SOCMOV for example, we suppose that when tweets from SOCMOV tweeters have been retweeted, that means public are reacting to social movement organizations. Among all the tweets of SOCMOV tweeters, 42 tweets have been retweeted once and 100 tweets haven’t been retweeted (Figure 14). Thus, 29.58% of the tweets have been retweeted, indicating that social movement organizations have 29.58% probability of benefiting from Twitter to call for action. In other words, social movement organizations have benefitted from Twitter to call for action in the oil spill event.

Figure 15 Retweets of SOCMOV

**Most Active Political Party**

For this oil spill event, we also tried to find out the most active political party. We calculated the centrality for all the GOV tweeters. Results are shown as below (Figure 15). Among these GOV tweeters, *‘senate\_gops’* (Senate Republicans) and *‘davidvitter’* (David Vitter) are Republican Party members. *‘Senatormenendez’* (Robert Menendez) is Democratic Party member. *‘Senatorsanders’* (Bernie Sanders) is independent. Others are all agencies. And among the political parties, *‘senate\_gops’* (Senate Republican) and *‘davidvitter’* (David Vitter) have the highest degree centrality as well as eigenvector centrality. Therefore, we can conclude that Republican Party is most active in this oil spill event.

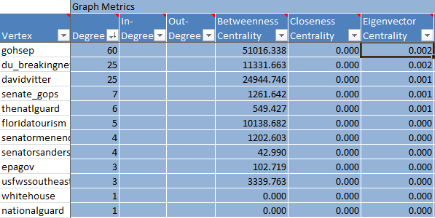


Figure 15 Centrality of Government Tweeter

**Political Parties’ Engaging Pattern Analysis**

Since we have found out the most active political party, we can continue analyzing the engaging pattern of political parties. We suppose that if political parties retweet a tweet, than they are engaging in a debate online not just relaying their pre-existing frames. As mentioned above, the only political party members are *‘senate\_gops’* (Senate Republications) and *‘davidvitter’* (David Vitter) which represent Republican Party and *‘senatormenendez’* (Robert Menendez) which represents Democratic Party. Thus, we calculated the retweet of these tweeters. Only ‘*du\_breakingnews’* and *‘davidvitter’* (David Vitter) has retweeted once. Also, we analyzed the structure of retweets (Figure 16). From this picture, we can see that GOV tweeters only account for 0.35%, indicating that political parties mainly relay on their pre-existing frames instead of engaging in online debate.

Figure 17 Classification of Retweets

**Other Findings**

Besides what we have analyzed, we can also see that the majority of the most influential actors in this network are grassroots tweeters rather than the established stakeholders in real life.