Information Propagation on Twitter

Vandalur – Kelambakkam Road

Chennai – 600127

Authors

Shubham Kumar Lala - 18BCE1025,

Akshat Kumar - 18BCE1033,

N. Mahesh - 18BCE1256

1. Abstract

Twitter has been the site where all formal and informal discussion takes place. The social network formed on Twitter is one of the largest networks. Every individual form a network or his or her own group of individuals. The aim of this project is to analyse the impact, the tweets of some people have on other. This analysis is done on the basis of the reactions on the tweets that an actor posts, the comments and reposting of the tweet. The reposting of the tweets can be used to analyse the connectivity of the people across the globe and also lets us understand the structure of the network. The data that we get can be mapped to form the network graph. The ego graph could be formed from the followers of a particular person, doing so for multiple users gives us an intricate graph that helps in making our analysis more precise.

The main objective of this project is to find the flow of information on Twitter, one of the social network giants. The frequency graphs tell us the most used words and the trends that are being followed. Also the cluster graph shows us how many words are related to each other, and they themselves form a cluster. The frequency of the positive and negative words show the sentiment of the information being propagated throughout the social media. For this project we have chosen the hashtag #mondaymotivation and we have extracted all the tweets of people who have used that hashtag. This dataset has been formed with the help of the twitter's API. We have used twitter's API and R's Machine Learning libraries to form clusters and data frames of the words that are related to the hashtag #mondaymotivation.

2. Literature Review

1. ttps://academic.oup.com/eurpub/article/29/Supplement_4/ckz187.101/5623163?searchresult=1
2. https://academic.oup.com/ofid/article/6/Supplement_2/S249/5603832?searchresult=1
3. https://academic.oup.com/neuro-oncology/article-abstract/21/Supplement_6/vi131/5620057?redirectedFrom=fulltext
4. http://www.academia.edu/download/30738159/TwitterWriteup_Sadikov.pdf
5. https://link.springer.com/chapter/10.1007/978-3-642-16567-2_16
6. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1037&context=acis2011
7. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(17)32802-7/fulltext
8. https://www.sciencedirect.com/science/article/pii/S1389128619303548
9. https://www.journals.elsevier.com/social-networks
10. https://ieeexplore.ieee.org/abstract/document/7542505

3. Results and Discussion

• Text Analysis

3.1 Libraries Installed :-

- twitter
- graphics
- purr
- stringr
- tm
- syuzhet

3.2 Connecting to Twitter API :-

- Connection to Twitter API is made using a Twitter Developer Account.
- The API key provided by Twitter is used to connect to Twitter.

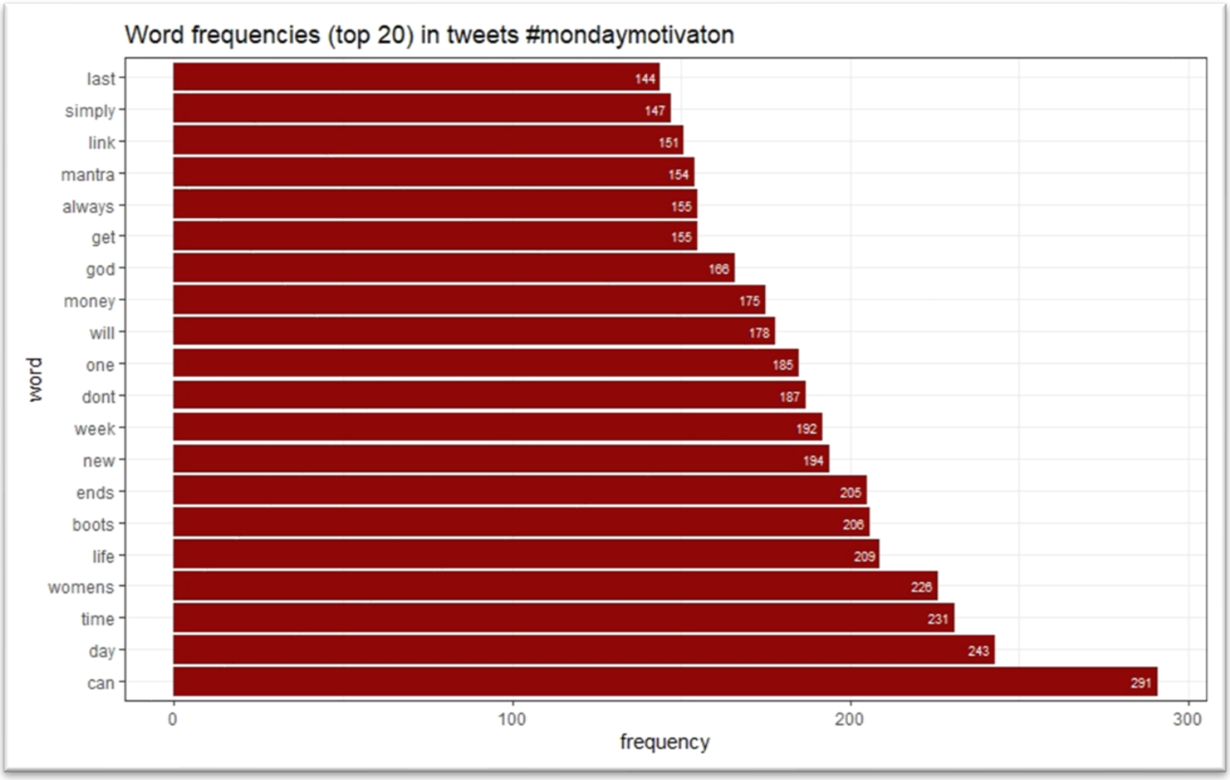
3.3 Checking for connection :-

```
> searchTwitter('analytics')
[[1]]
[1] "JustBeMentalist: RT @newsbytesph: Gov't eyes data science school as demand for data analytics peaks https://t.co/fLMT8U:
28B"

[[2]]
[1] "NH_SHAPIRO: Check out my Gig on Fiverr: analyze your website today if you use google analytics https://t.co/gE8SuBi8GL"

[[3]]
[1] "achyutha: RT @Soumyadipta: I had accidentally deleted the earlier tweet while checking the analytics. I am putting out t
he thread again.\nRavish Kumar…"

[[4]]
[1] "readingtweet: RT @reach2ratan: #Coronavirus: #Hackers are exploiting the COVID-19 outbreak to steal your information htt
ps://t.co/s9L23alAcP\n\n#CyberSecur…"

[[5]]
[1] "LarryBoyer: RT @LarryBoyer: Death of An Iconic Brand: RIP Toys R Us https://t.co/tfunycyKU5 #analytics https://t.co/RwY:
QjrKXO"

[[6]]
[1] "CyberSecurityN8: RT @reach2ratan: #Coronavirus: #Hackers are exploiting the COVID-19 outbreak to steal your information
 https://t.co/s9L23alAcP\n\n#CyberSecur…"
```

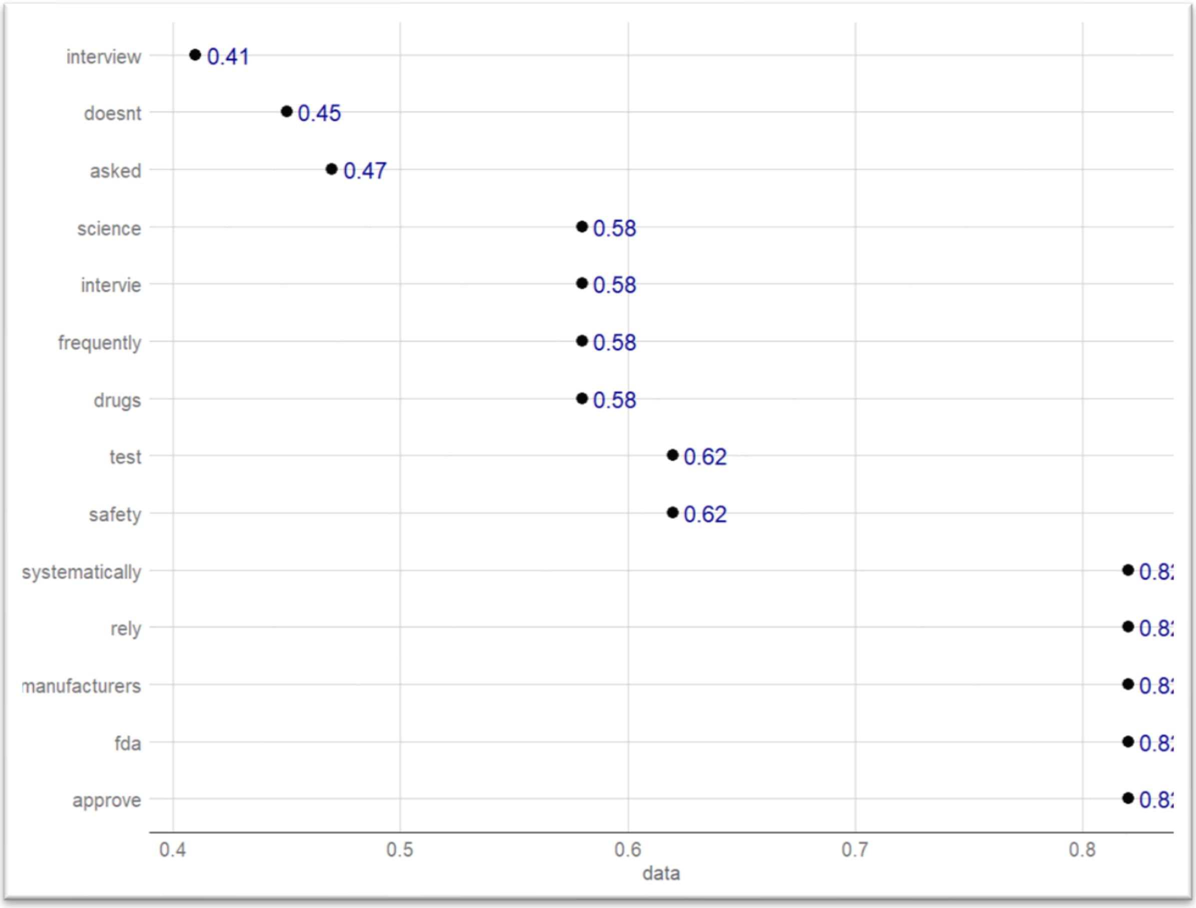
3.4 Searching Twitter for MondayMotivation related tweets and loading into a dataframe :-

```
> tdm <- TermDocumentMatrix(corpus)
> tdm
<<TermDocumentMatrix (terms: 5707, documents: 5000)>>
Non-/sparse entries: 35382/28499618
Sparsity           : 100%
```

3.5 Word Frequencies :-



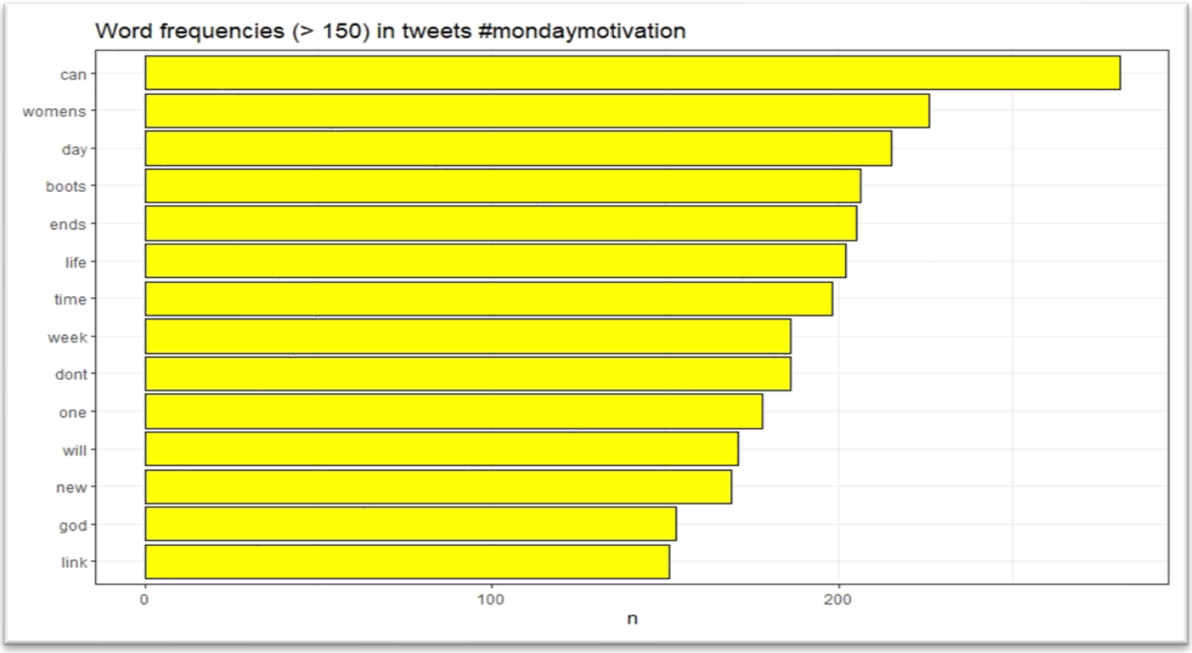
3.6 Word associations : -



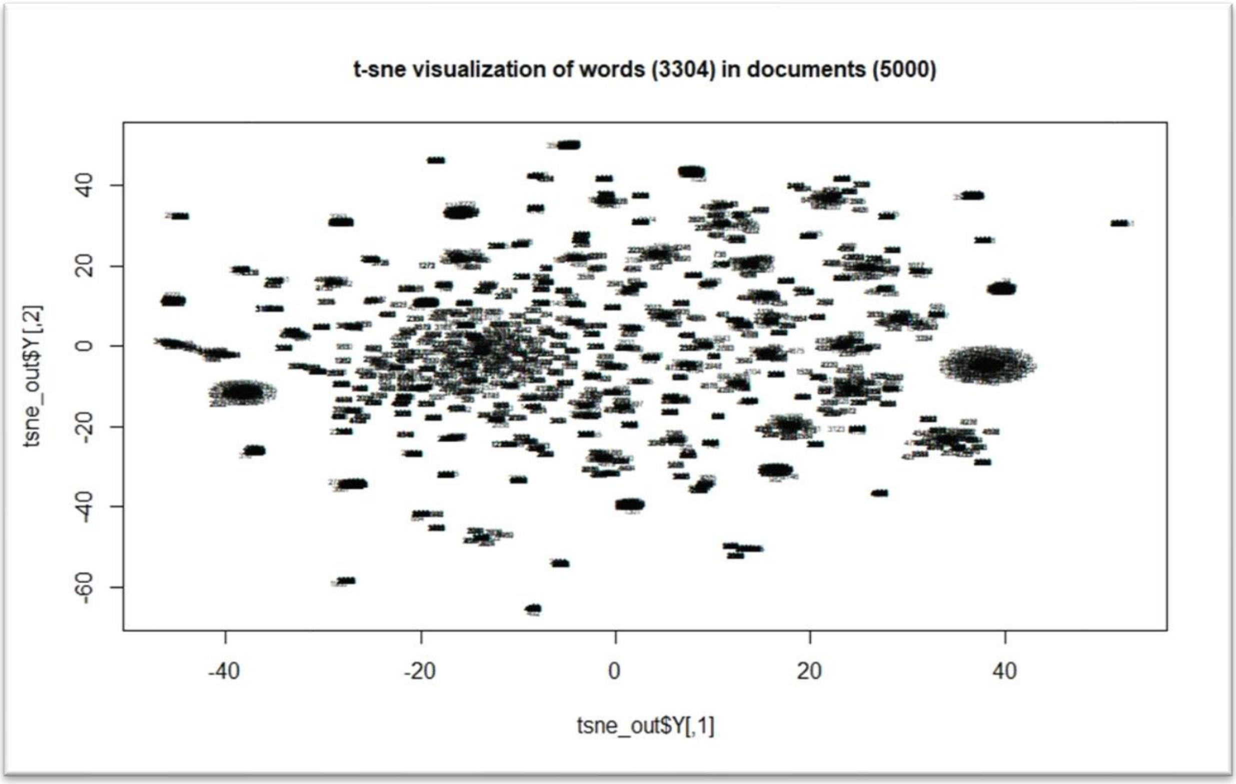
3.7 Finding words which have 100+ frequency :-

```
> findFreqTerms(tdm,100)
 [1] "boots"    "ends"     "womens"   "back"    "week"    "love"      "can"    "simply"  "people"
[10] "last"     "bhagavad" "enter"    "day"     "happy"   "god"       "mantra" "says"    "read"
[19] "via"      "one"      "morning"  "good"    "life"    "get"       "new"    "start"   "time"
[28] "dont"     "great"    "best"     "win"     "world"   "change"    "just"   "forget"  "will"
[37] "uufef"    "always"   "money"    "like"    "make"    "something" "never"  "link"    "fun"
[46] "date"
>
```

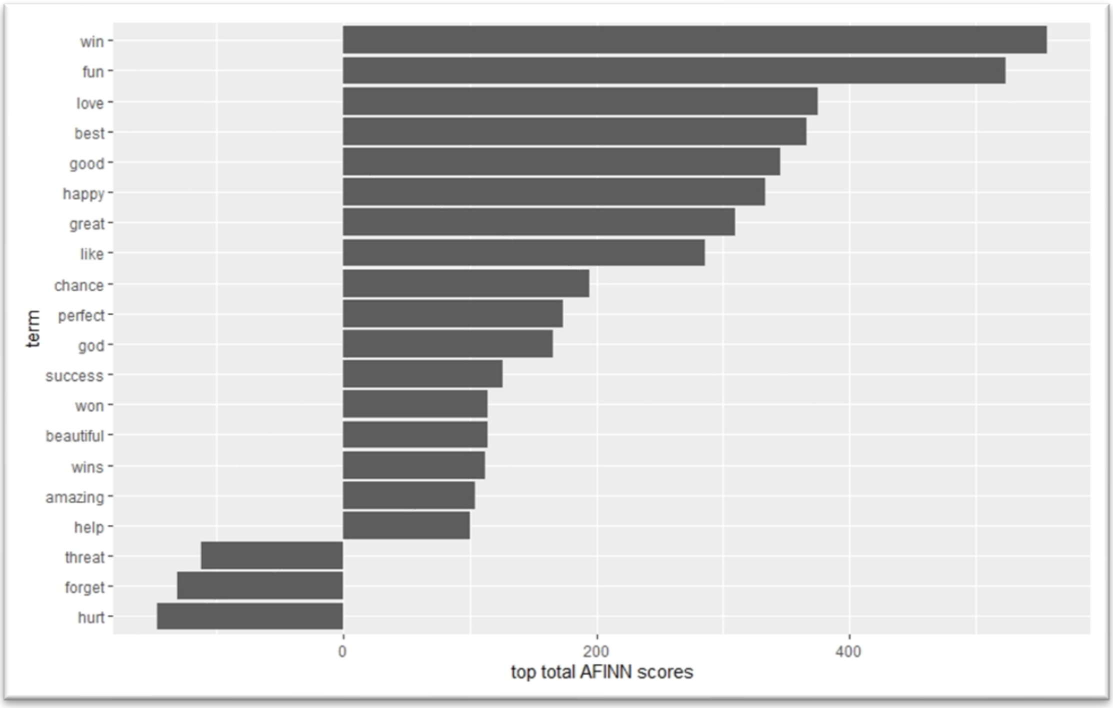
3.8 Plotting Word frequency (>150) in tweets #mondaymotivation :-



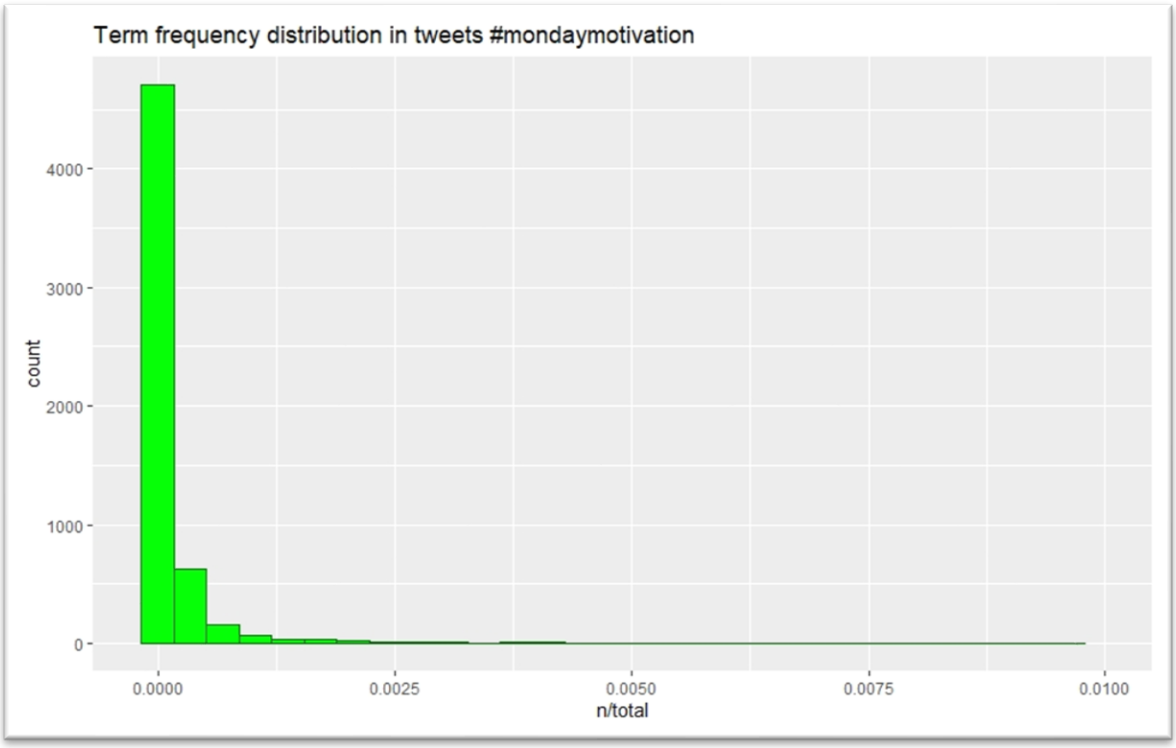
3.9 Dimension reduction with t-sne :-



3.10 AFINN Sentiments :-

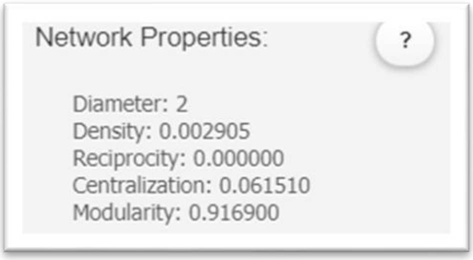


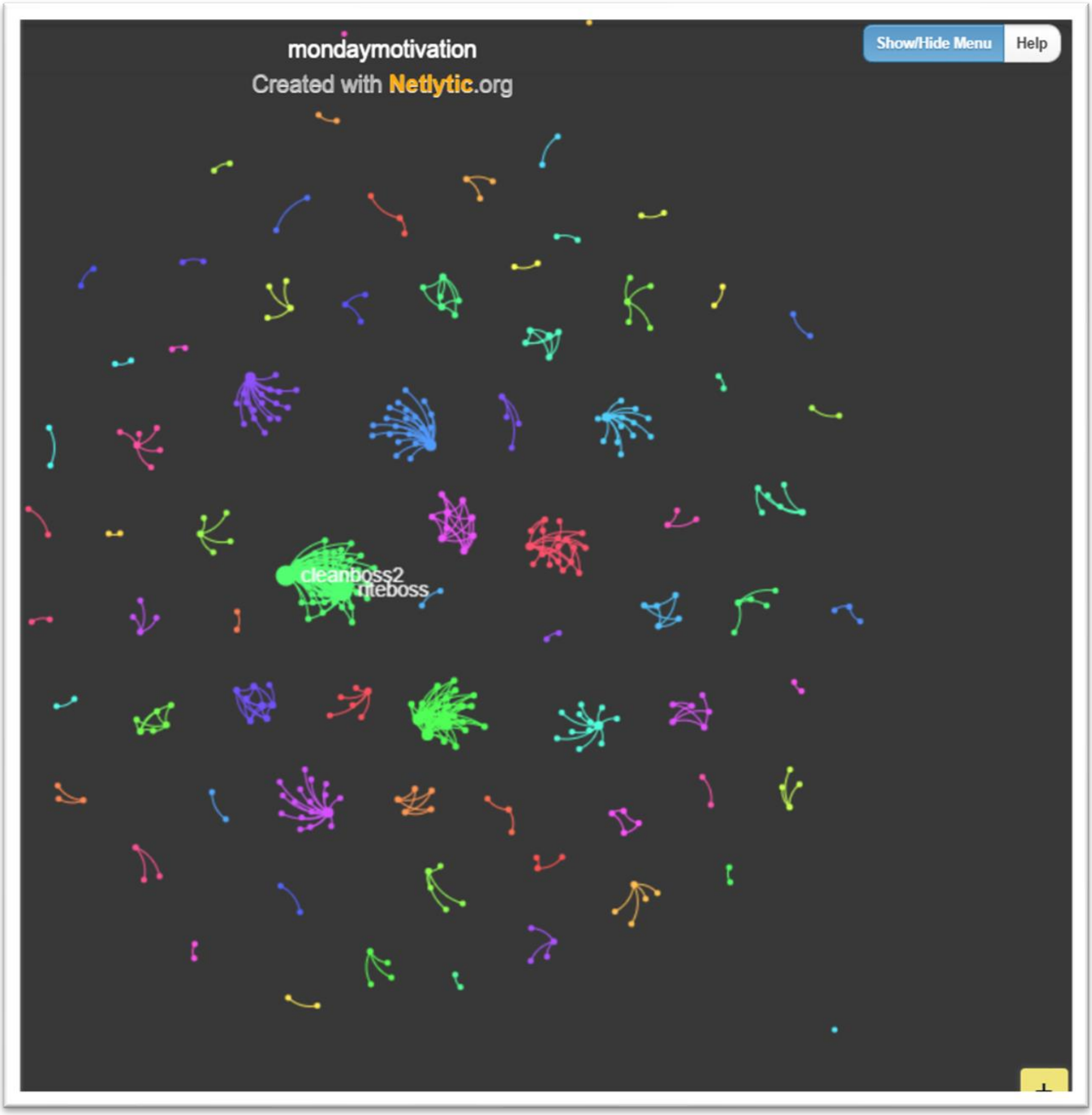
3.11 Term frequency distribution :-



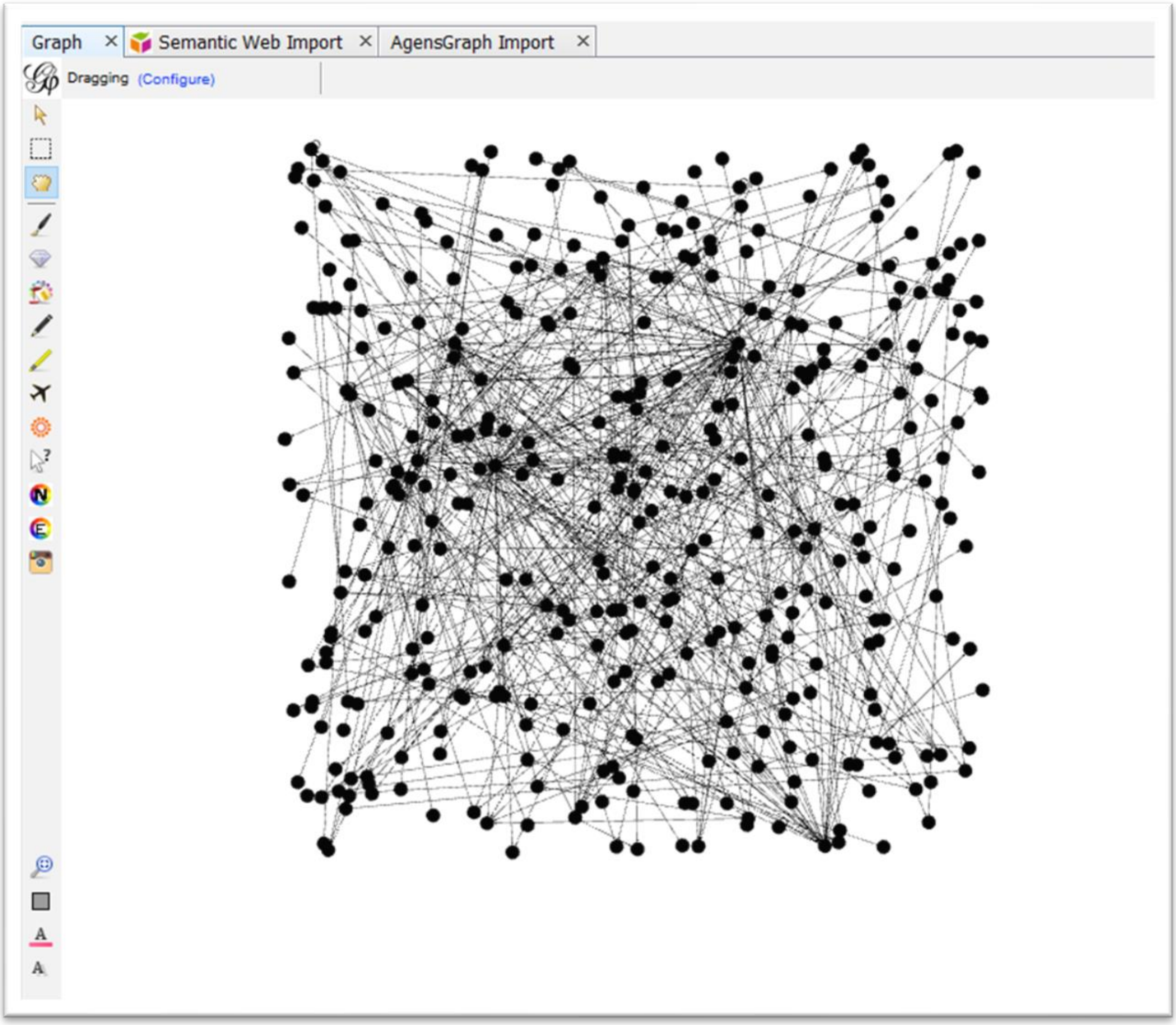
• Network Analysis

3.12 Network Analysis Using Netlytic



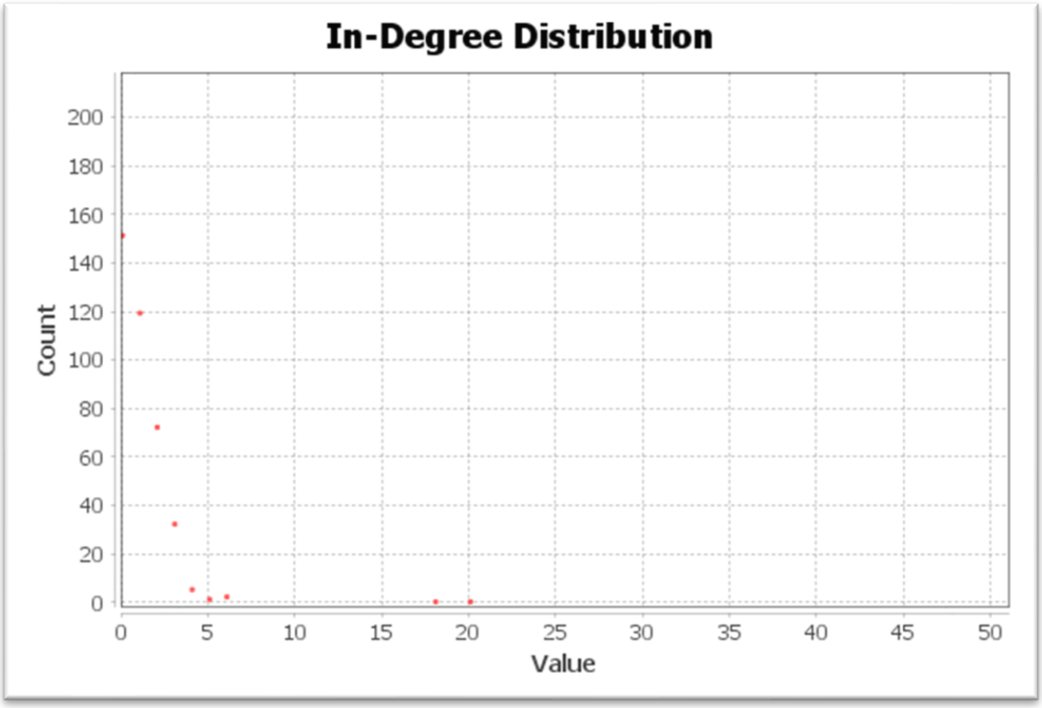


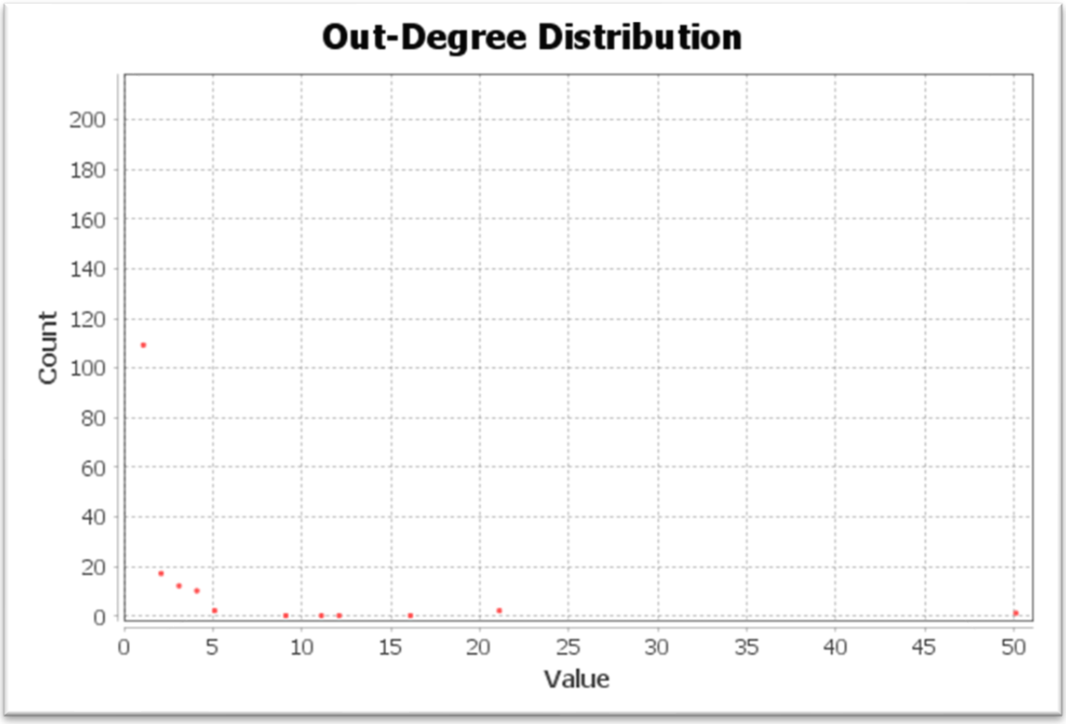
3.13 Graph in Gephi

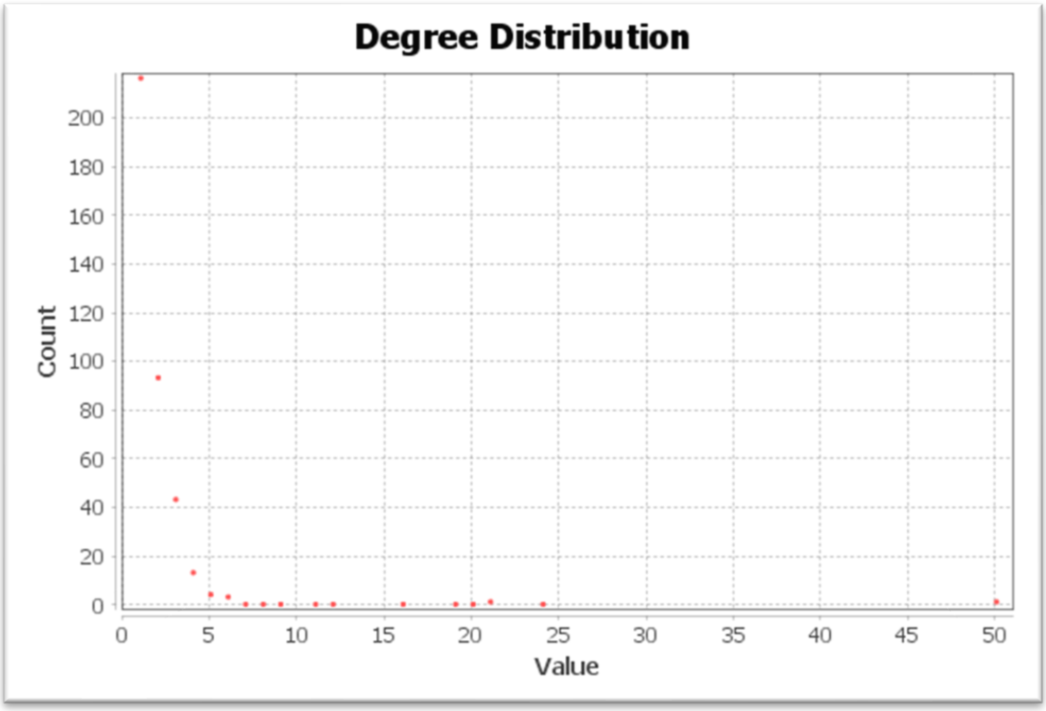


3.14 Reports :-

- Degree Report -



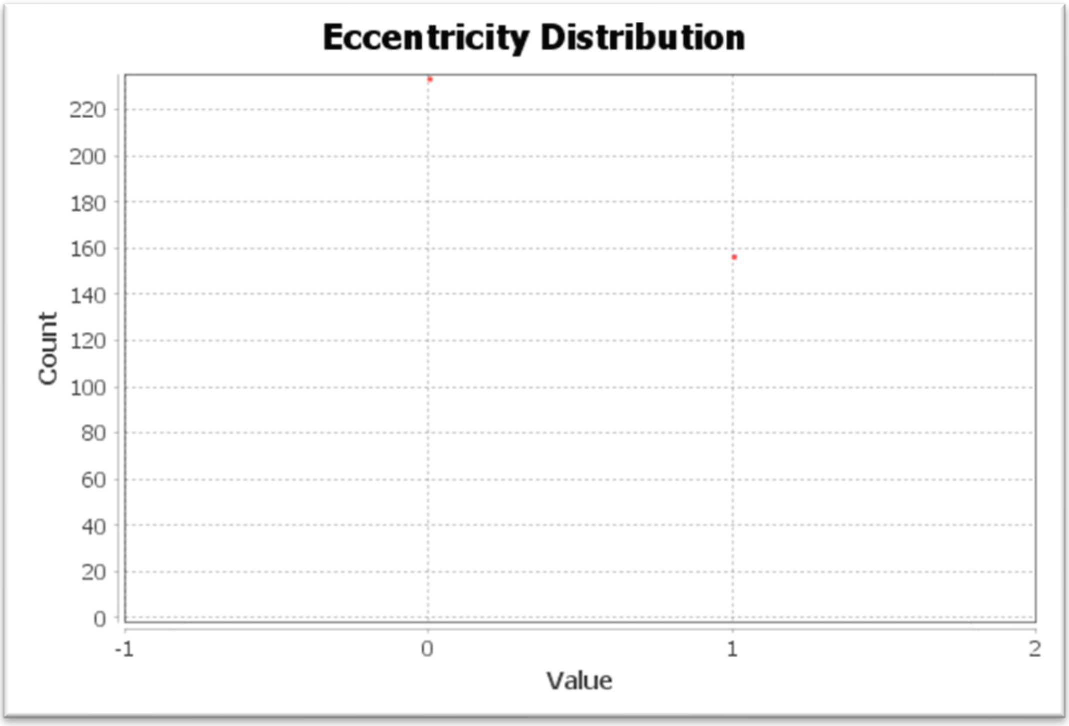


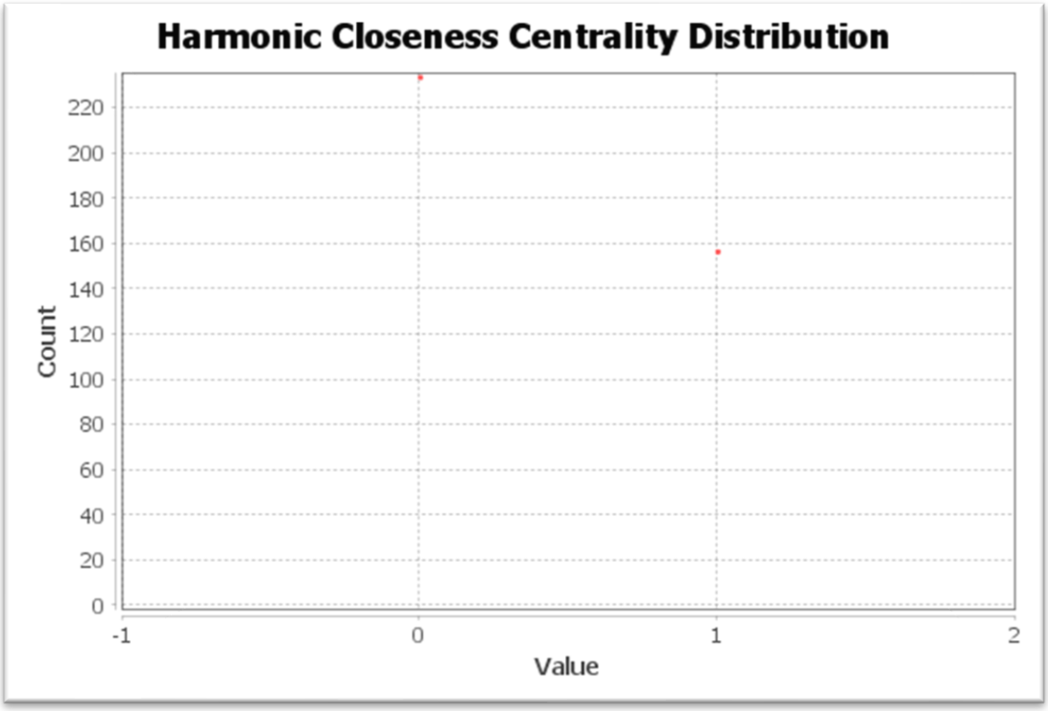


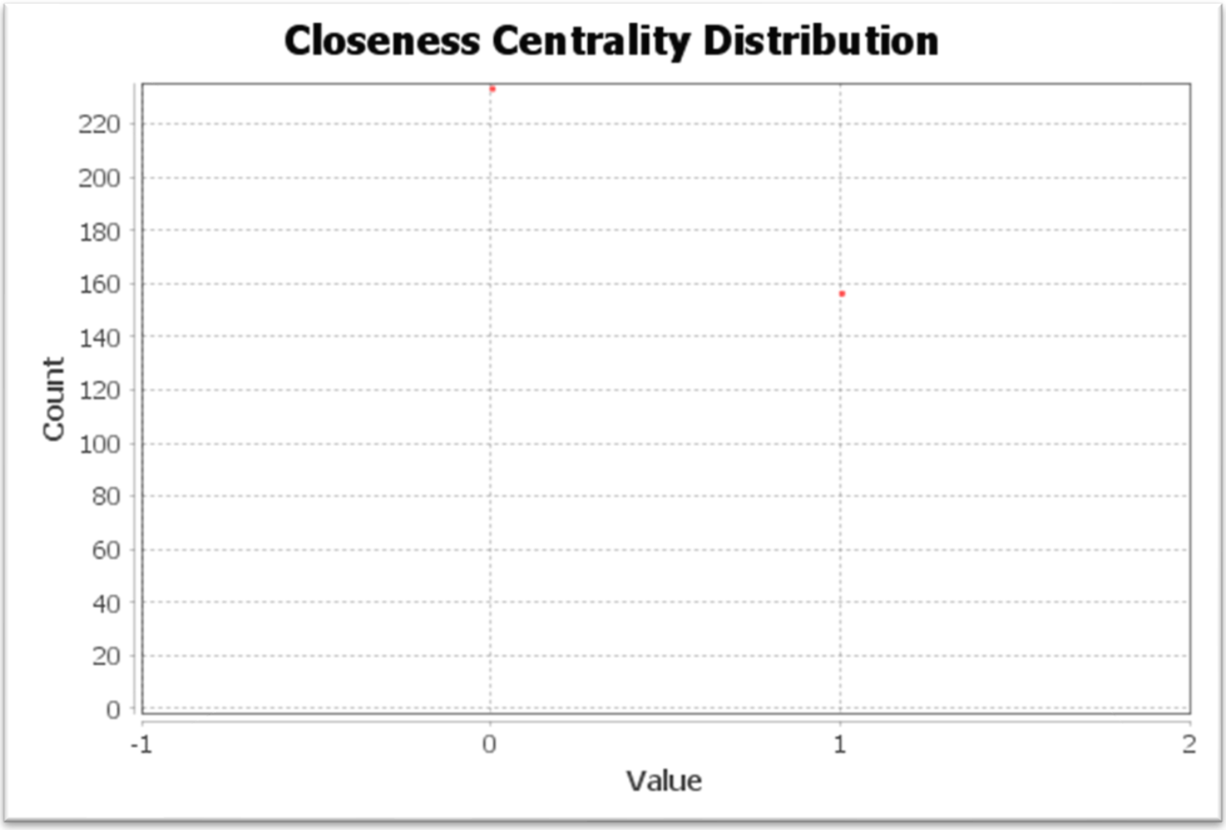
- Minimum Spanning Tree –

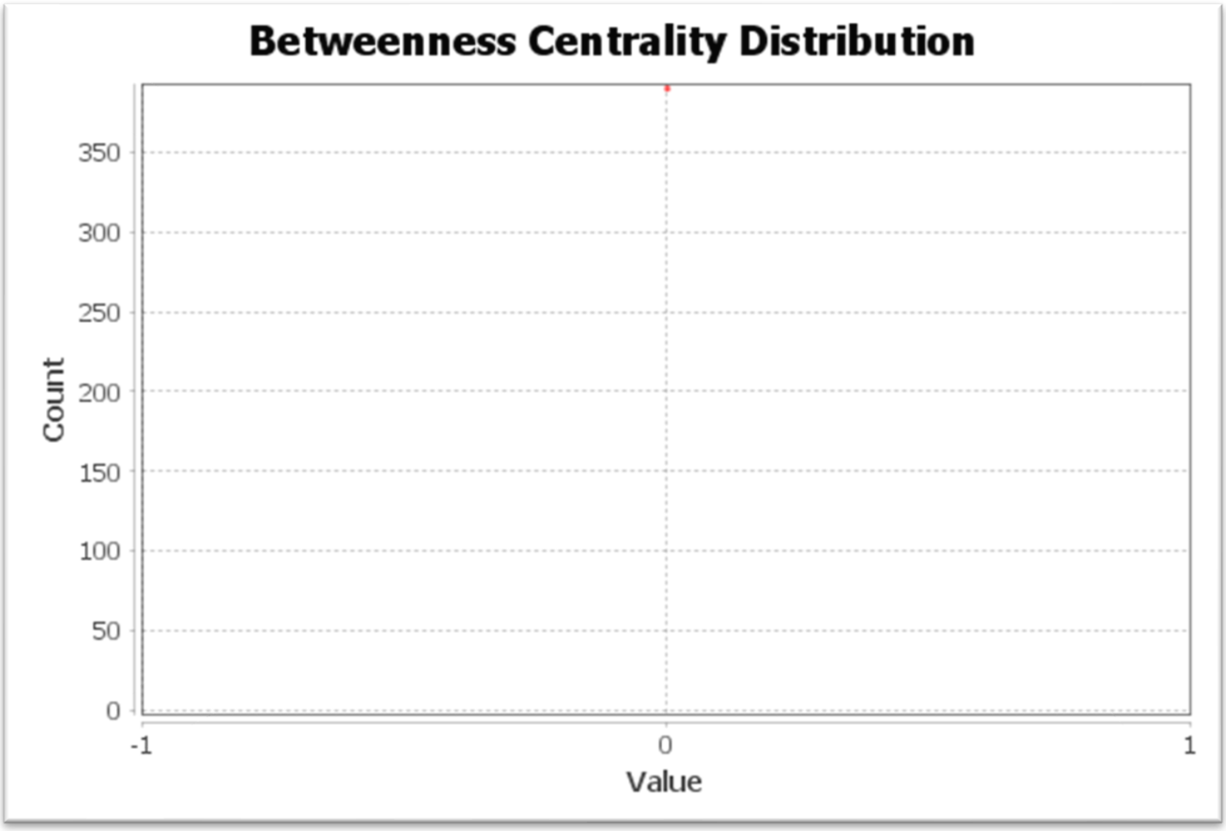
Number of edges in the minimum spanning tree: 314
Weight of the minimum spanning tree: 314.0

- Network Diameter -

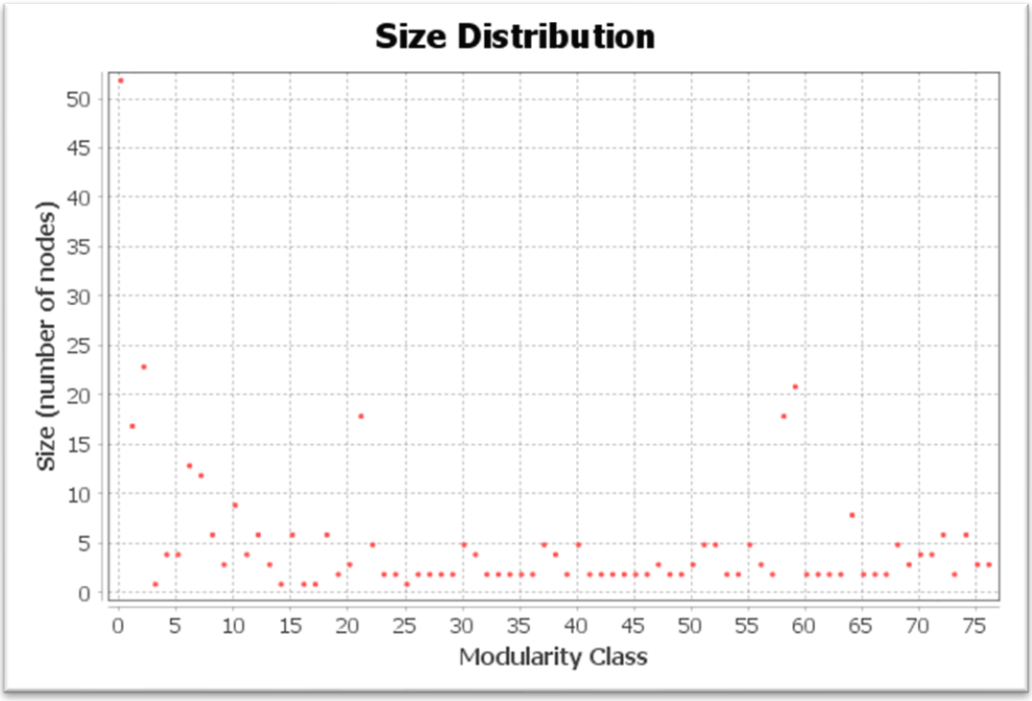




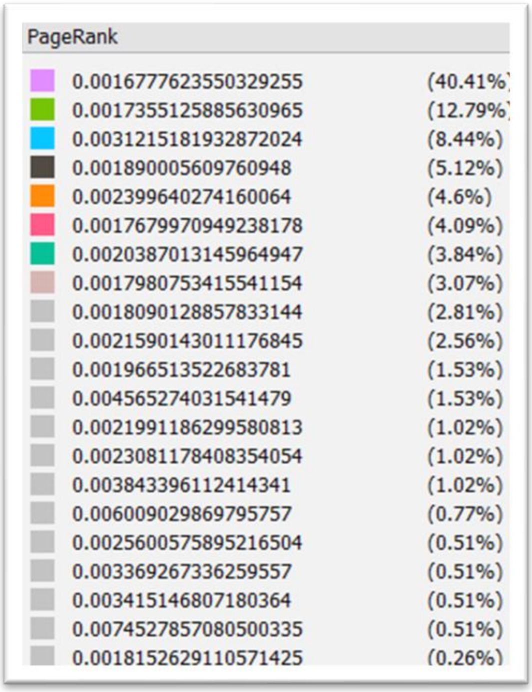

Closeness Centrality Distribution
Count
220
200
180
160
140
120
100
80
60
40
20
0
-1
0
1
2
Value


Betweenness Centrality Distribution
Count
350
300
250
200
150
100
50
0
-1
0
1
Value

- Modularity : -



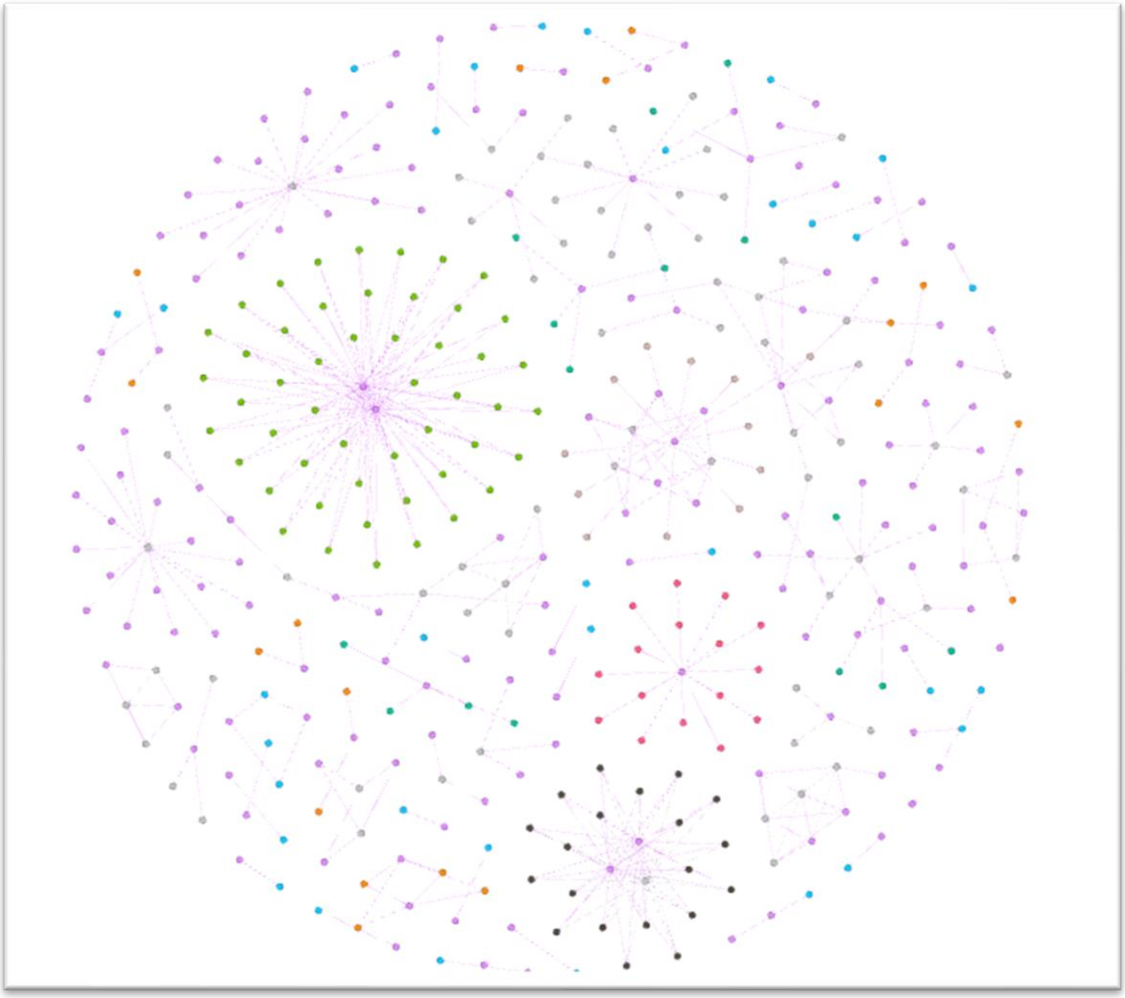
- PageRank : -



PageRank

0.0016777623550329255	(40.41%)
0.0017355125885630965	(12.79%)
0.0031215181932872024	(8.44%)
0.001890005609760948	(5.12%)
0.002399640274160064	(4.6%)
0.0017679970949238178	(4.09%)
0.0020387013145964947	(3.84%)
0.0017980753415541154	(3.07%)
0.0018090128857833144	(2.81%)
0.0021590143011176845	(2.56%)
0.001966513522683781	(1.53%)
0.004565274031541479	(1.53%)
0.0021991186299580813	(1.02%)
0.0023081178408354054	(1.02%)
0.003843396112414341	(1.02%)
0.006009029869795757	(0.77%)
0.0025600575895216504	(0.51%)
0.003369267336259557	(0.51%)
0.003415146807180364	(0.51%)
0.0074527857080500335	(0.51%)
0.0018152629110571425	(0.26%)

- USING Fruchterman-Reingold Layout : -



- Girvan – Newman :-

Parameters:

Respect edge type for shortest path betweeness:	yes
Respect parallel edges for shortest path betweeness:	yes
Respect edge type for modularity computation:	yes
Respect parallel edges for modularity computation:	yes

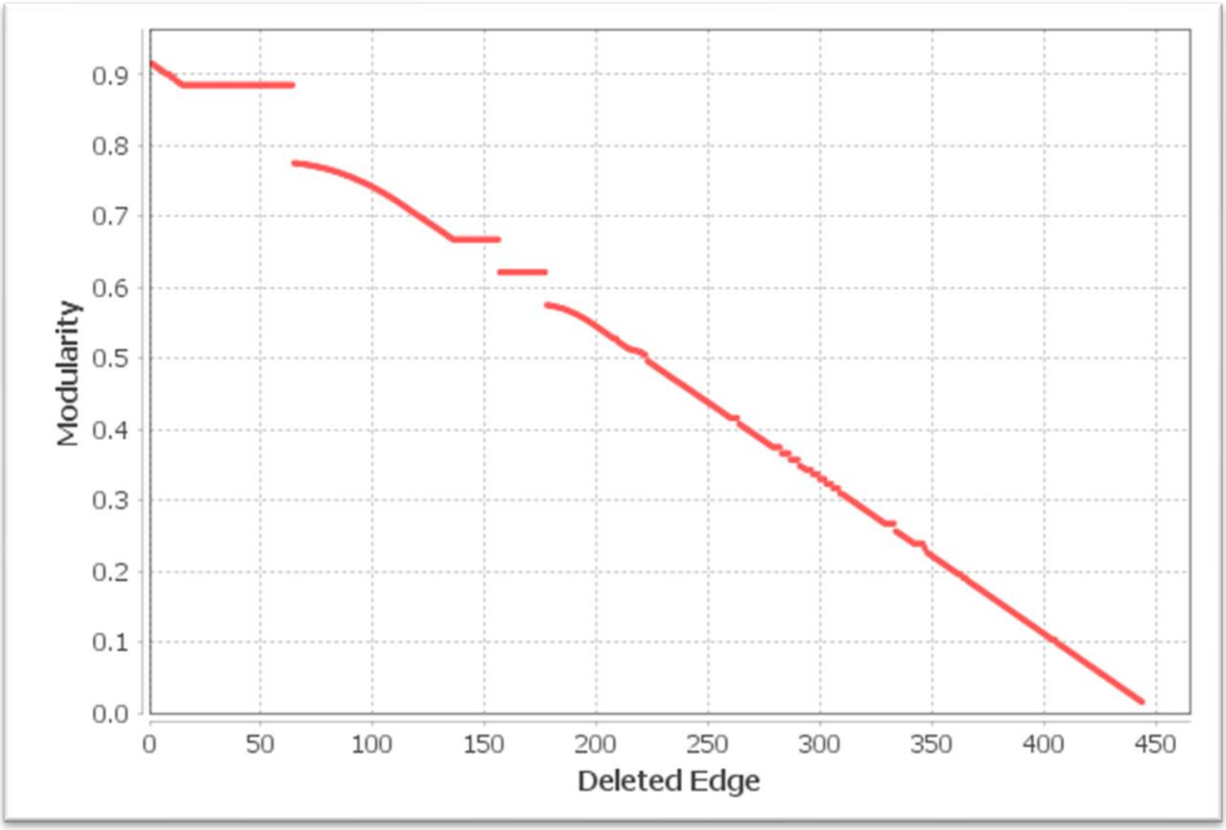
Processed Graph Data

Nodes: 391
Edges 455

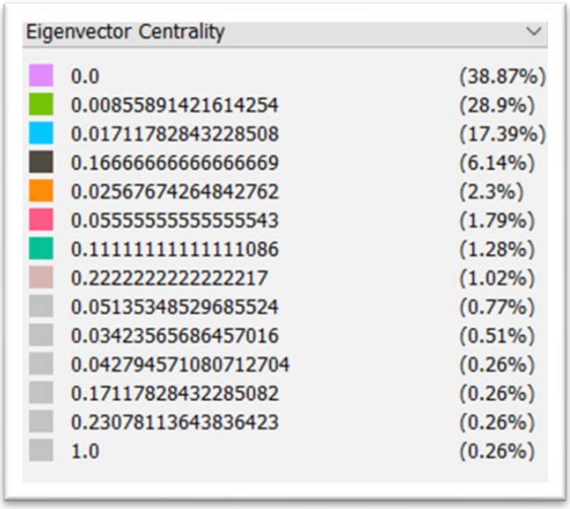
Processing time: 0.428 sec.

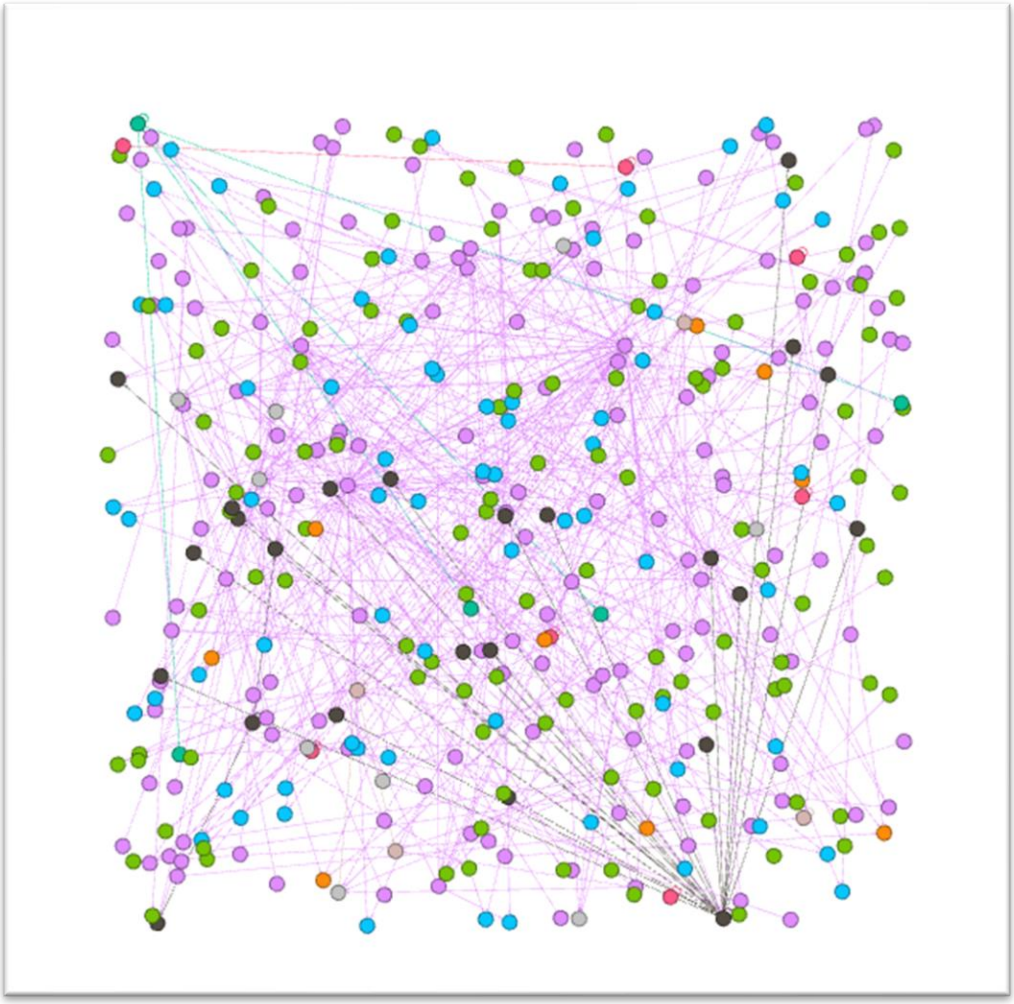
Communities

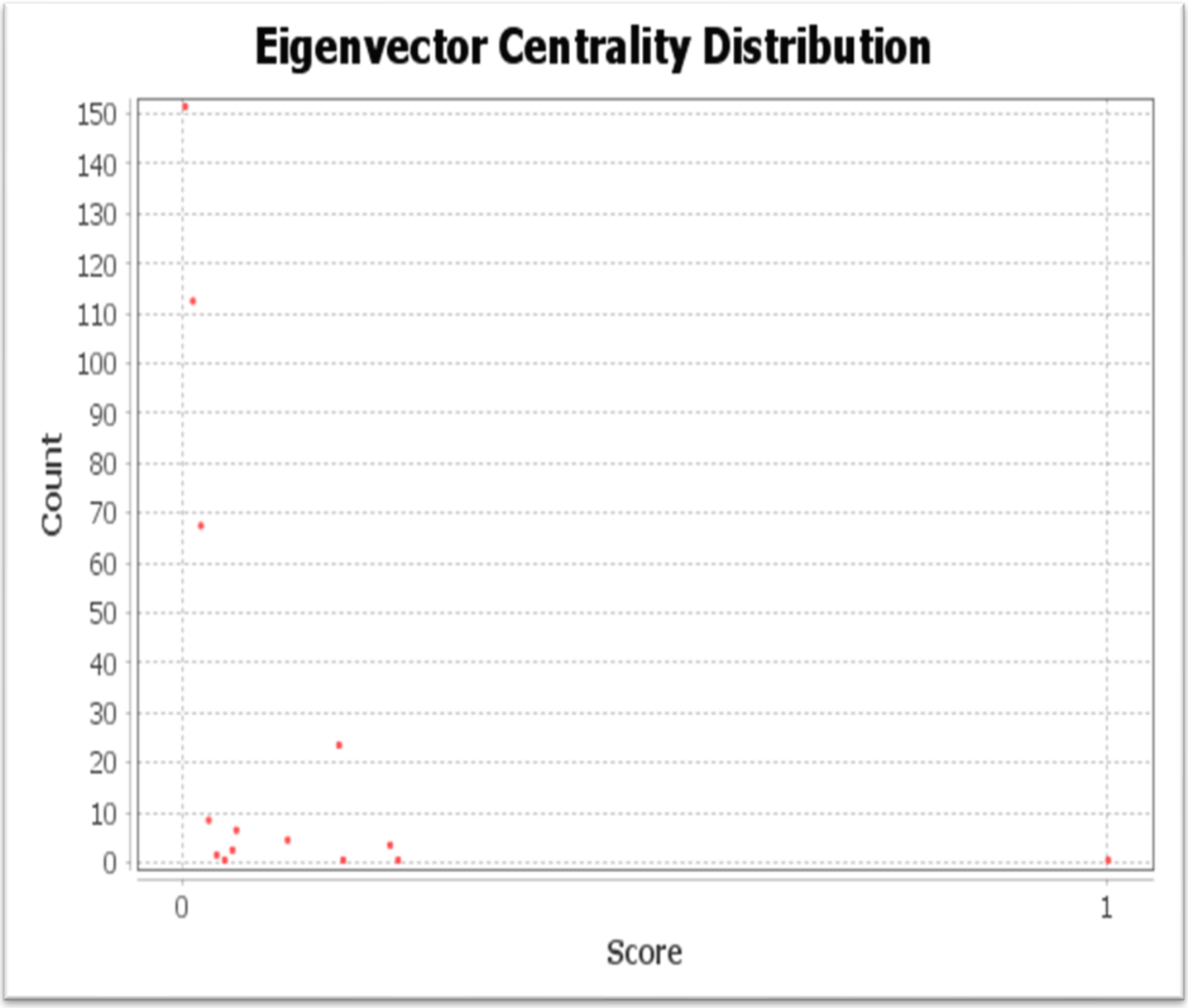
Number of communities: 77
Maximum found modularity: 0.9182273



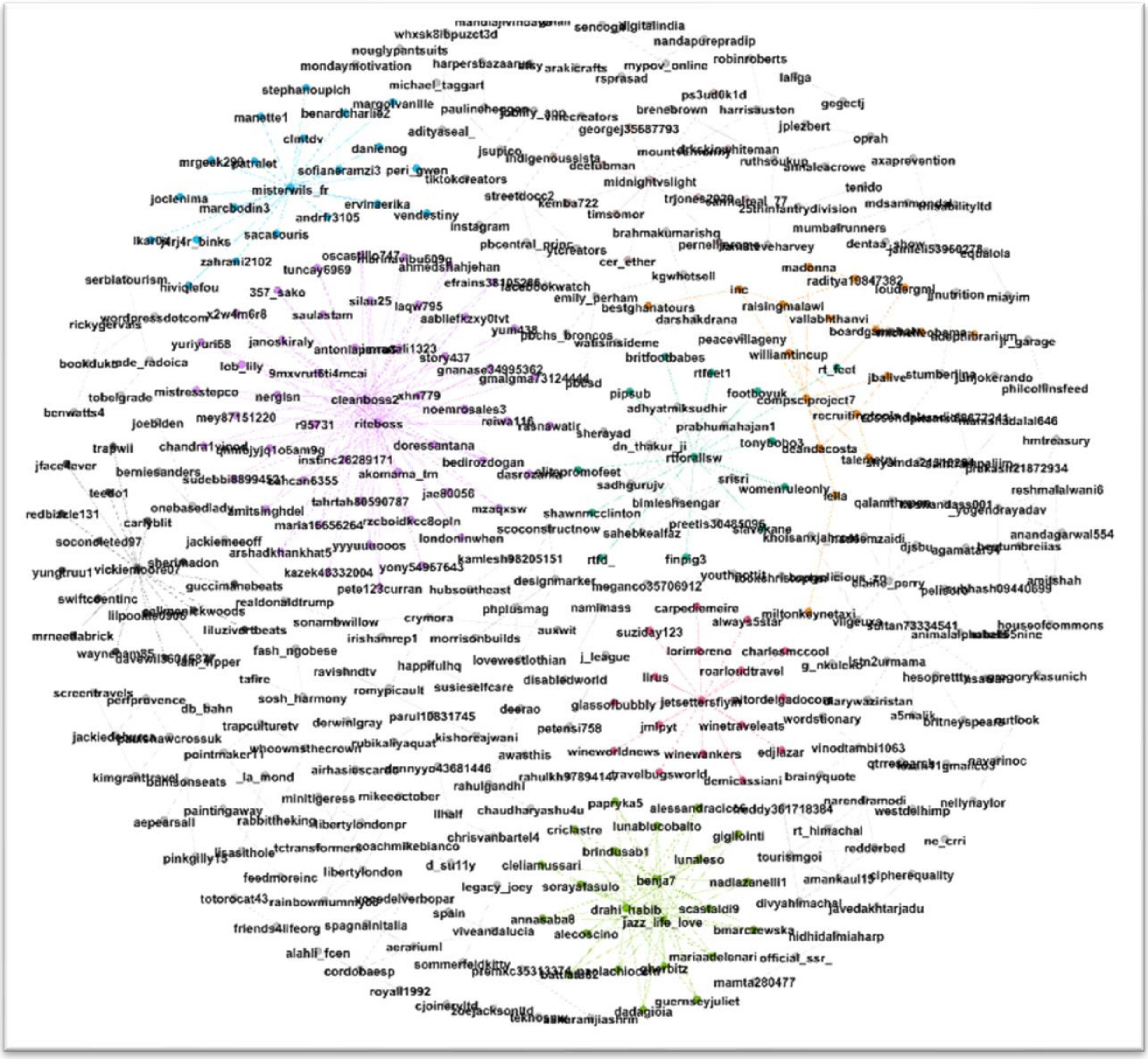
- Eigen Vector Centrality : -






Eigenvector Centrality Distribution
Count
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
0
1
Score

- Modularity Class With Usernames : -



- Test Results : -

Id	Label	indegree	outdegree	Degree	Eccentricity	closnesscentrality	betweenesscentrality	pageranks
n1	nandapurepradip	0	2	2	1	1	0	0.001678
n2	_lisasithole	0	1	1	1	1	0	0.001678
n3	sultan73334541	0	4	4	1	1	0	0.001678
n4	anandagarwal554	0	2	2	1	1	0	0.001678
n5	reshmalalwani6	0	2	2	1	1	0	0.001678
n6	subhash09440699	0	2	2	1	1	0	0.001678
n7	tenido	0	1	1	1	1	0	0.001678
n8	cleanboss2	0	50	50	1	1	0	0.001678
n9	riteboss	0	50	50	1	1	0	0.001678
n10	whxsk8ibpuzct3d	0	1	1	1	1	0	0.001678
n11	bumsonseats	0	3	3	1	1	0	0.001678
n12	jetsettersfiyin	0	16	16	1	1	0	0.001678
n13	royall1992	0	1	1	1	1	0	0.001678
n14	benja7	0	21	21	1	1	0	0.001678
n15	drahi_habib	0	21	21	1	1	0	0.001678
n16	jazz_life_love	3	21	24	1	1	0	0.001815
n17	lstn2urmama	0	1	1	1	1	0	0.001678
n18	mhall55nine	0	1	1	1	1	0	0.001678
n19	pelisoro	0	1	1	1	1	0	0.001678
n20	annaleacrowe	1	1	2	0	0	0	0.001678
n21	arakicrafts	1	2	3	1	1	0	0.001678
n22	premkc35313374	0	1	1	1	1	0	0.001678
n23	parul10831745	0	4	4	1	1	0	0.001678
n24	jamieli53960278	0	1	1	1	1	0	0.001678
n25	jackiemeeoff	0	2	2	1	1	0	0.001678
n26	onebasedlady	0	3	3	1	1	0	0.001678
n27	adeptlibrarium	0	1	1	1	1	0	0.001678

4. References

1. K. Awati, A gentle introduction to topic modeling using R, https://eight2late.wordpress.com/2015/09/29/agentle-introduction-to-topic-modeling-using-r , 29 september 2015.
2. Clara clustering http://www.sthda.com/english/articles/27-partitioningclustering-essentials/89-clara-clustering-large-applications.
3. T. Graham & R. Ackland, Topic modeling of tweets in R : A tutorial and methodology. https://www.academia.edu/19255535/Topic_Modeling_of_Tweets_in_R_A_Tutorial_and_Methodology