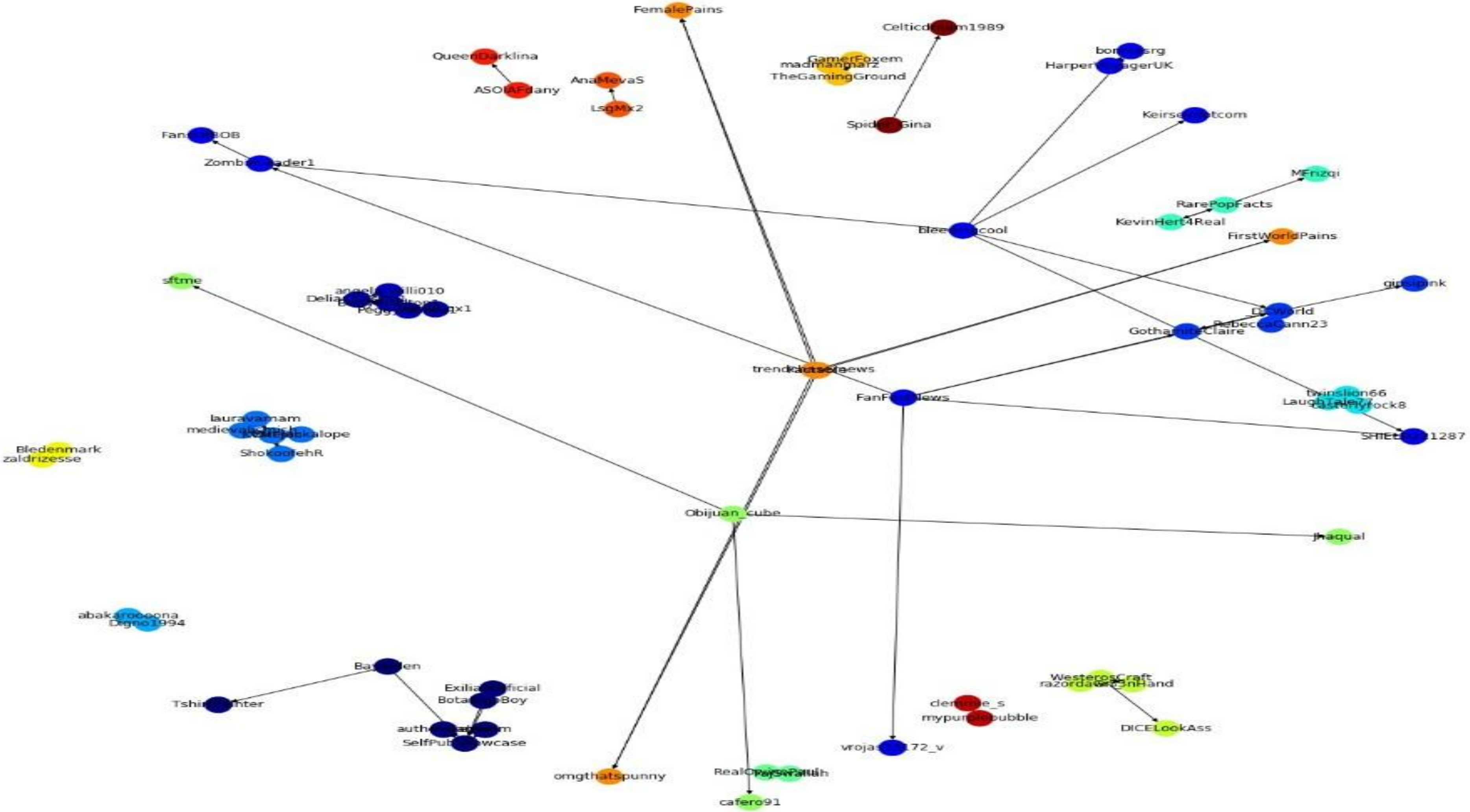
Network Science for Web(DS608)-Final Project

Group Members

1. Lokesh Garg(MT2018053)
2. Sagar Jhunjhunwala(MT2018097)
3. Shubham Darokar(MT2018113)
4. Vivek Gupta(MT2018134)
5. Apurva Kulkarni(PH2018020)

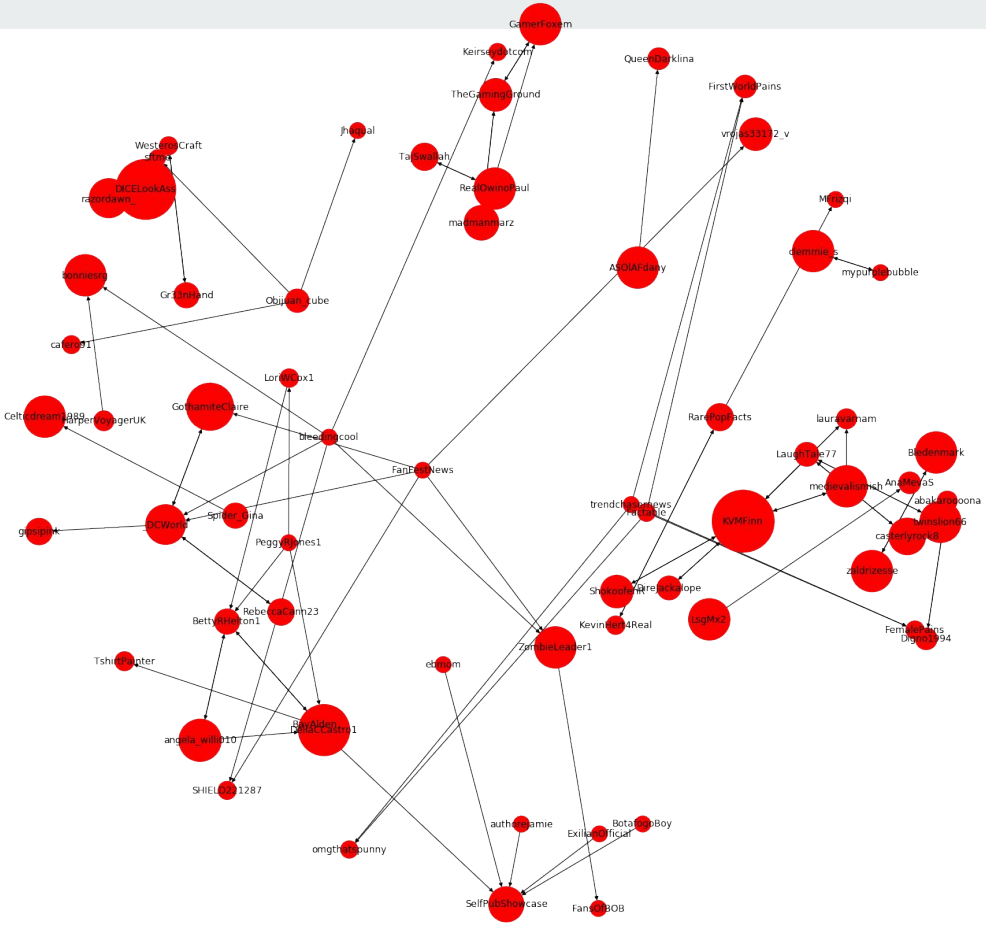
Follower-Followee Network for GOT

- Number of nodes: 66
- Number of edges: 81


FemalePains
QueenDarklina
AnaMevaS
LsgMx2
TheGamingGround
RarePopFacts
KevinHert4Real
FirstWorldPains
sftme
Bledenmark
zaldrizesse
ShokoofehR
Obijuan_cube
jhaqual
cafero91
omgthatspunny
WesterosCraft
DICELookAss

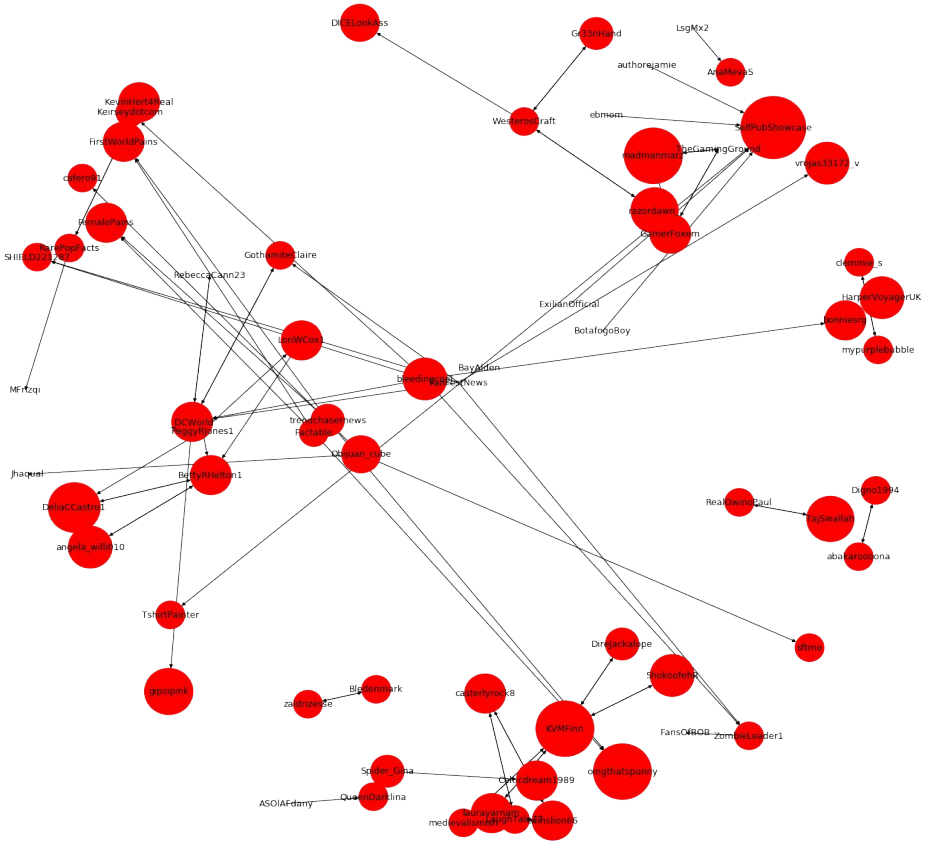
Page Rank:

PageRank works by counting the number and quality of edges to a node to determine a rough estimate of how important the node is. The underlying assumption is that more important nodes are likely to receive more links from other nodes.



Closeness:

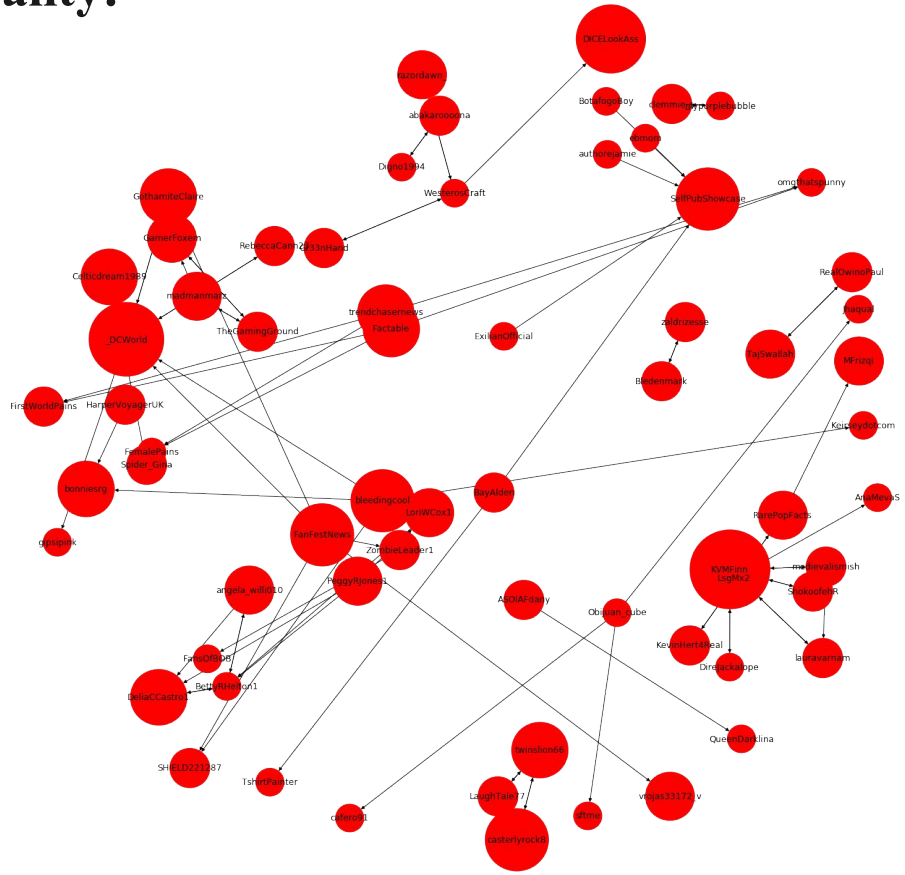
In a connected graph, closeness centrality (or closeness) of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the *closer* it is to all other nodes.



Betweenness:

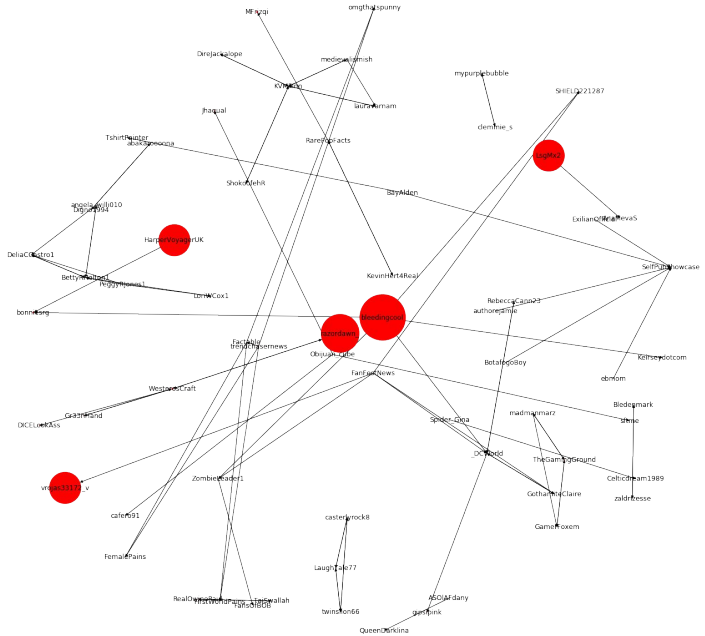
Betweenness is a centrality measure of a vertex within a graph. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

Degree Centrality:



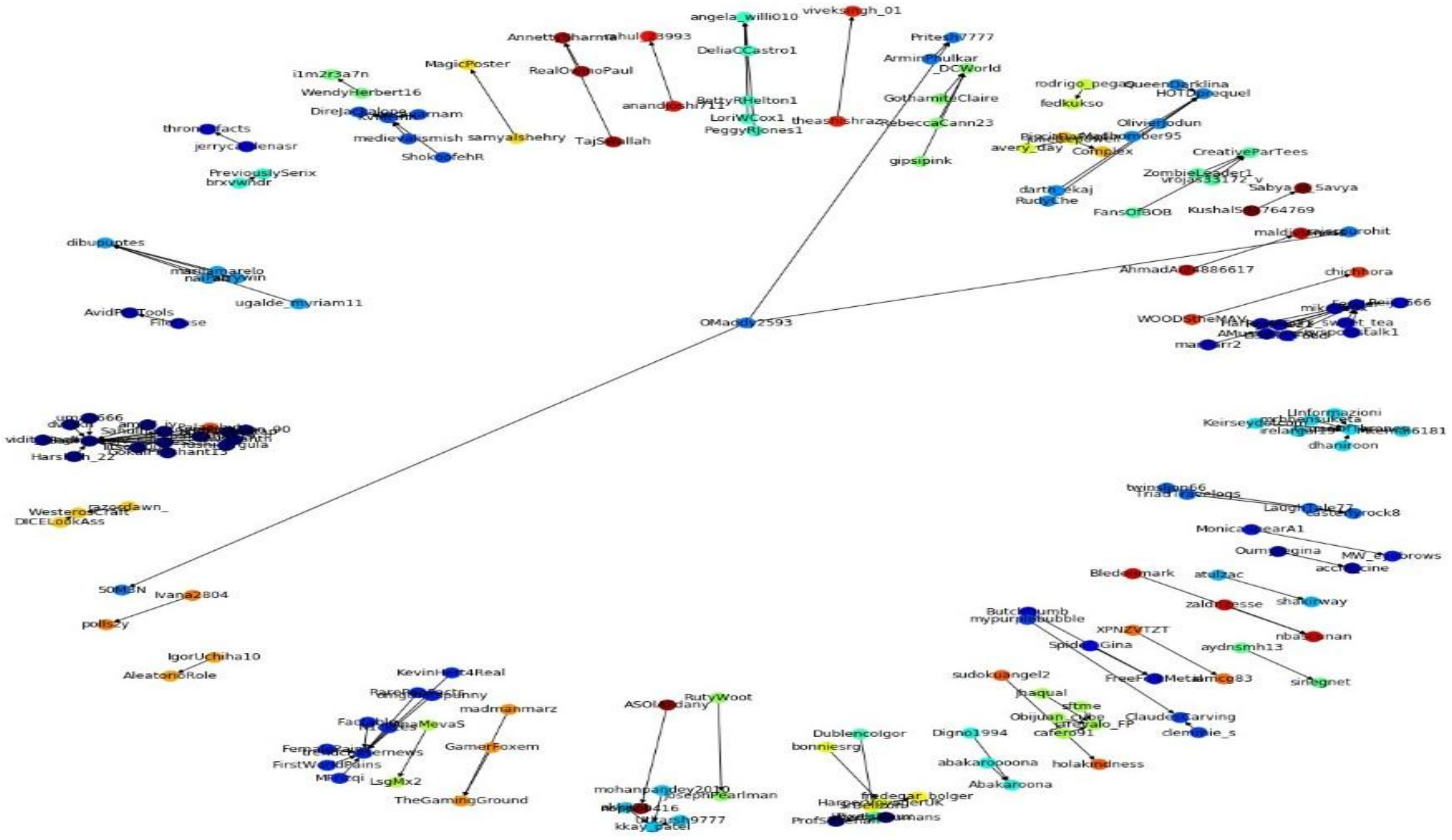
Eigen Vector:

Eigenvector centrality is a measure of the influence a node has on a network. If a node is pointed to by many nodes then that node will have high eigenvector centrality.



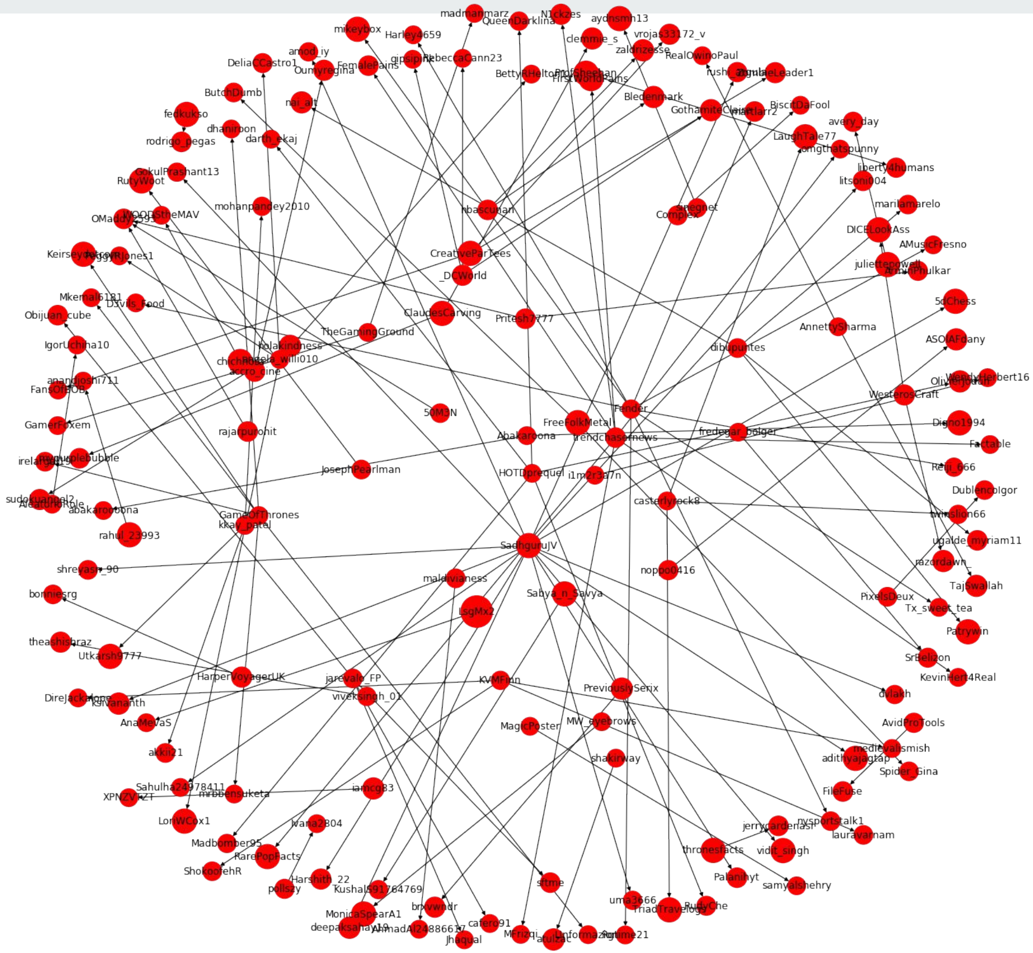
Retweet Graph for GOT:

- Number of nodes: 166
- Number of edges: 117



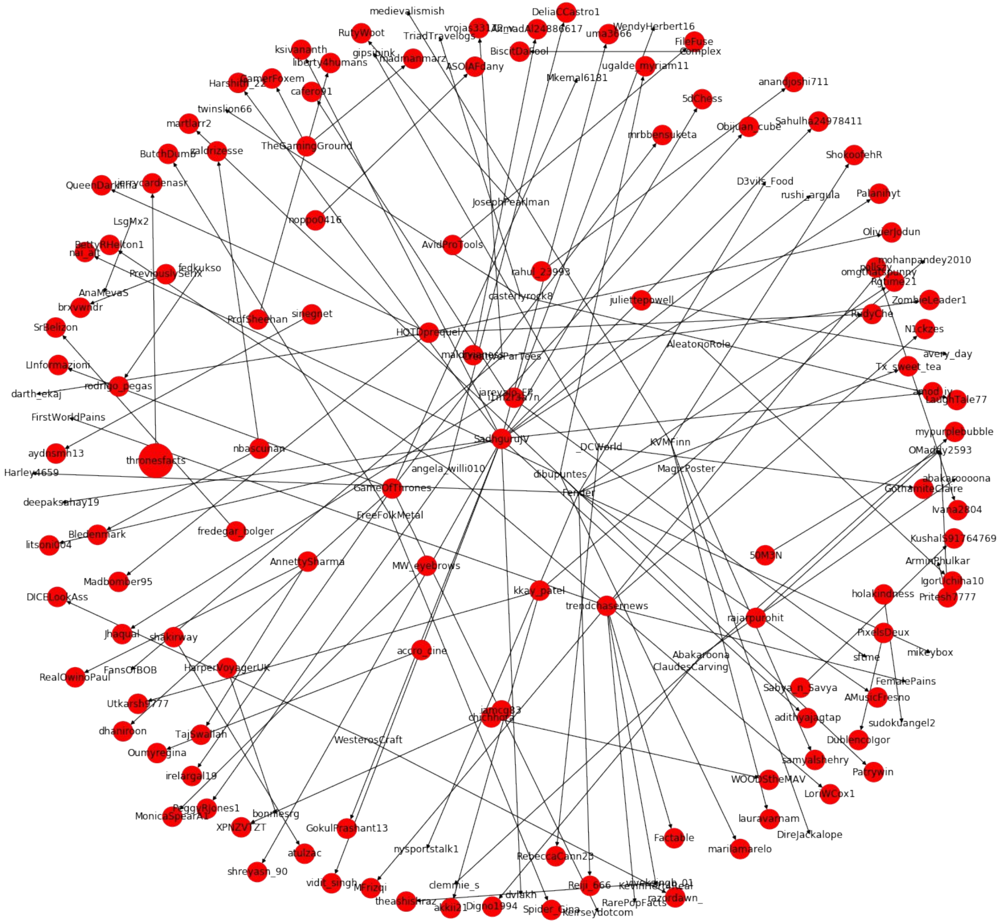
angela_willi010
DeliaCCastro1
LoriWCox1
PeggyRJones1
MagicPoster
i1m2r3a7n
WendyHerbert16
samyalshehry
TajSallah
ArminPhulkar
_DCWorld
GothamiteClaire
RebeccaCann23
gipsipink
rodrigo_pega
fedkukso
QueenDarklina
HOTDprequel
Olivierjodun
avery_day
Complex
CreativeParTees
FansOfBOB
Sabya_Savya
RudyChe
PreviouslySerix
brxvwhdr
dibupuntes
ugalde_myriam11
OMaddy2593
AhmadA4886617
chichhora
WOODStheMAV
Unformazioni
dhaniroon
Harsh_22
Westeros
DICELookAss
Ivana2804
pollszy
IgorUchiha10
AleatorioRole
KevinHeart4Real
madmanmarz
GamerFoxem
TheGamingGround
LsgMx2
FirstWorldPains
RutyWoot
mohanpandey2010
kkay_patel
Dublencolgor
bonniesrg
Digno1994
abakaropoona
Abakaroona
holakindness
sudokuangel2
Obijuan_cube
cafero91
sftme
Claude_carving
XPNZVTZT
SpiderGina
FreeFallMetal
Butchdumb
aydnsmh13
sirregnet
atulzac
shakirway
nbasenan
Bledermark
MonicaBearA1
MW_eyebrows
OumRegina
acchicine
LaughTale77
eastertyrock8
twinsdon66

Page Rank:



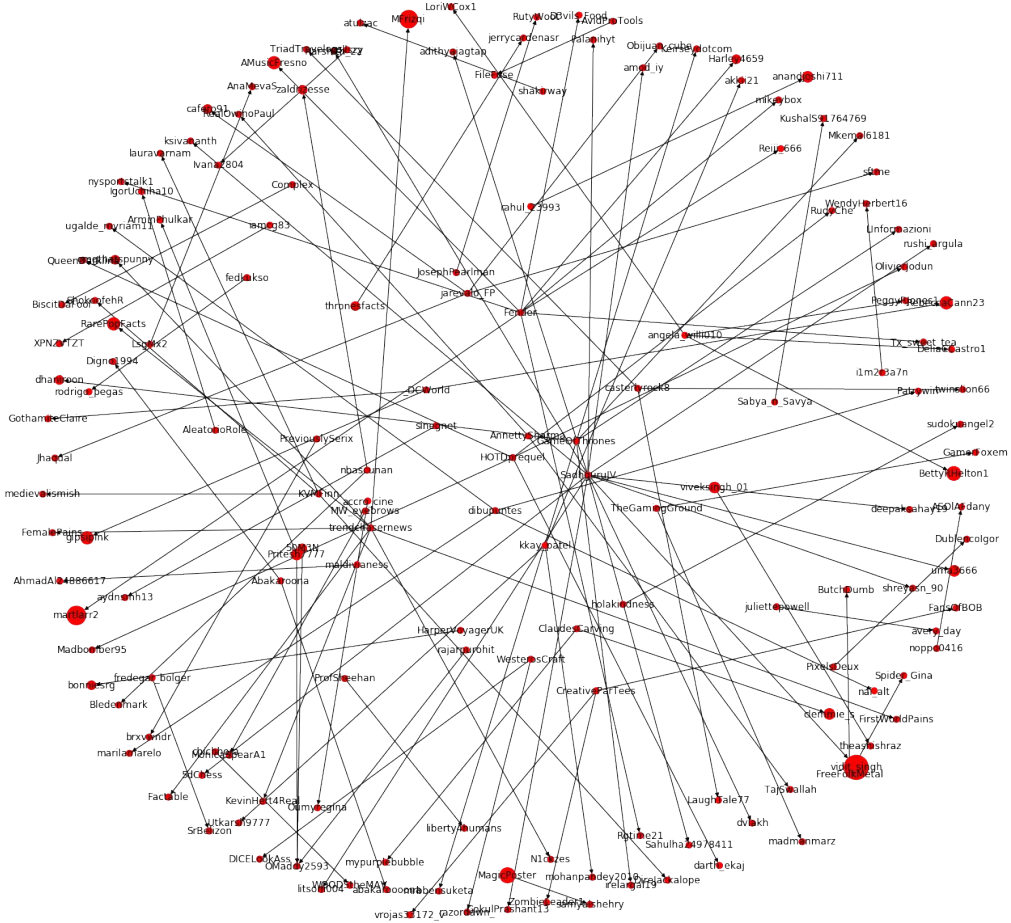
madmanmarz
QueenDarklina
N1ckzes
aydnsmh13
mikeybox
Harley4659
clemmie_s
vrojas33172_v
amod_iy
zaldrizesse
RealOwinoPaul
DeliaCCastro1
ButchDumb
nai_alt
Bledenmark
BiscitDaFool
fedkukso
rodrigo_pegas
dhanirpon
darth_ekaj
avery_day
LaughTale77
omgthatspunny
RutyWoot
GokulPrashant13
liberty4humans
litsoni004
mohanpandey2010
nbascunan
Complex
marilamarelo
DICELookAss
AMusicFresno
CreativeParTees
_DCWorld
Mkemal5181
D3vils_Food
Obijuan_cube
ClaudesCarving
Pritesh7777
5dChess
AnnettySharma
TheGamingGround
IgorUchiha10
ASOIAFdany
dibujuntes
FansOfBOB
GamerFoxem
WesterosCraft
50M3N
rajatpurohit
Abakaroona
FreeFolkMetal
fredegar_bolger
Digno1994
Factable
trendchasernews
JosephPearlman
HOTDprequel
Reiji_666
Dublencolgor
sudokuangel2
abakaroobona
casterlyrock8
rahul_23993
twinslion66
ugalde_myriam11
GameOfThrones
kkay_patel
SadhguruJV
shreyash_90
razordawn_
noppo0416
maldivianess
bonniesrg
Sabya_n_Savya
TajSwallah
PixelsDeux
Tx_sweet_tea
LsgMx2
theashishbraz
Patrywin
Utkarsh9777
SrBelizon
KevinHeft4Real
HarperVoyagerUK
jarevalo_FP
KVMFinn
PreviouslySerix
DireJackalope
Ksivananth
viveksingh_01
dvlakh
AnaMeliaS
MagicPoster
MW_eyebrows
AvidProTools
akkii21
medievalsmish
adithyajagtap
shakinway
Spider_Gina
Sahulha24978411
XPNZVRZT
mrbbensuketa
iamcg83
FileFuse
LordWCox1
nysportstalk1
lauravarnam
jerrygardenarsi
Ivana2804
Madbomber95
thronesfacts
vidit_singh
RarePopFacts
ShokoofehR
Palanihyt
Harshith_22
samyalshehry
pollsay
Kushal59176476
sttme
uma3666
MonicaSpearA1
brawwndr
TriadTravelogs
RudyChe
deepaksahay10
AhmadAl24886617
cafero91
Jhaqual
MFrizqi
atulzac
Infoformazing
Spime21

Closeness:

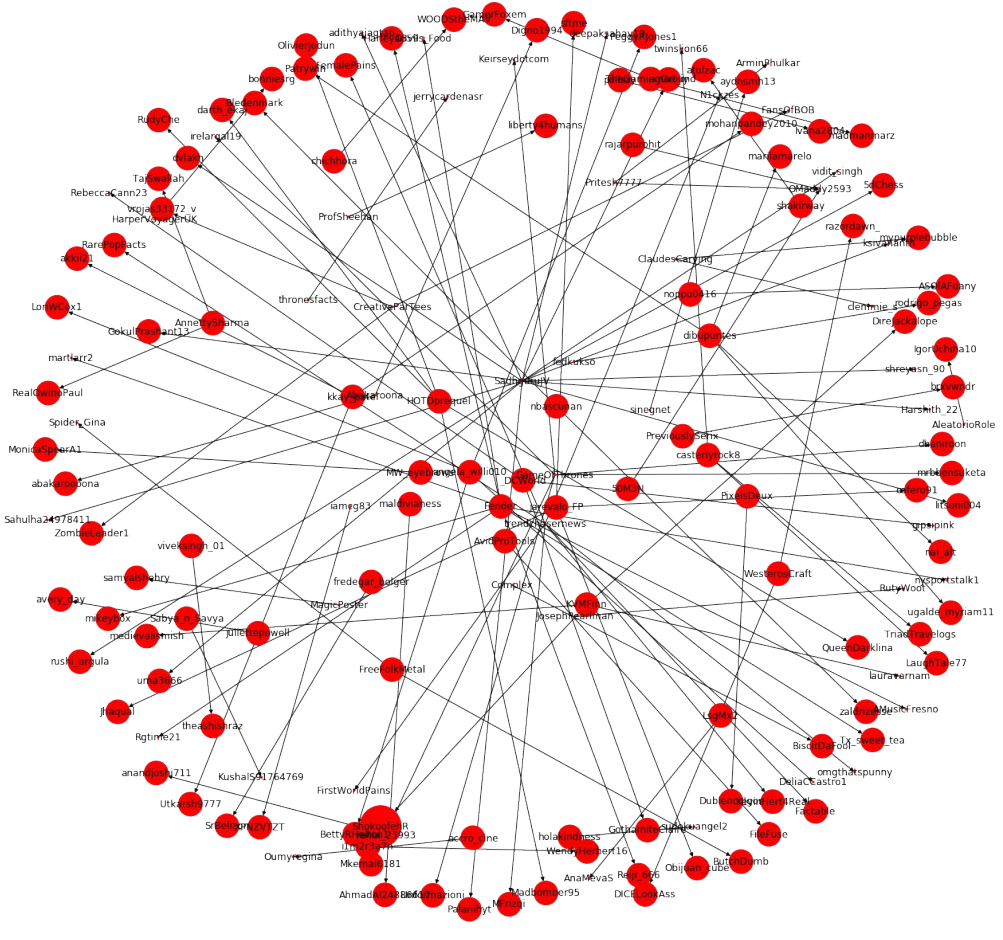


medievalismish
DeliaCCastro1
WendyHerbert16
RutyWoot
FileFuse
Complex
ksivananth
liberty4humans
ugalde_myriam11
Mkemal6181
anandjoshi711
GamerFoxem
5dChess
twinslion66
martlarr2
Obijuan_cube
mrbbensuketa
TheGamingGround
ShokoofehR
D3vils_Food
LsgMx2
AvidProTools
OlivierJodun
mohanpandey2010
fedkukso
AnaMevaS
casterlyrock8
juliettepowell
ZombieLeader1
sinegnet
ProfSheehan
SrBelizon
RudyChe
N1ckzes
HOTDprequel
avery_day
LInformazioni
Tx_sweet_tea
rodrigo_pegas
LaughTale77
FirstWorldPains
mypurplebubble
SadhguruJV
_DCWorld
KVMFinn
OMaddy2593
aydnsmith13
thronesfacts
nbascunan
angela_willi010
MagicPoster
Harley4659
GameOfThrones
dibupuntes
GothamiteClaire
Fender
deepaksahay19
Ivana2804
FreeFolkMetal
fredegar_bolger
Bledenmark
Kushal591764769
50M3N
ArminPhulkar
AnnettySharma
Madbomber95
MW_eyebrows
IgorUchiha10
Pritesh7777
DICELookAss
kkay_patel
holakindness
trendchasernews
rajatpurohit
Jhaqual
shakirway
PixelsDeux
accro_cine
mikeybox
FansOfBOB
HarperVoyagerUK
Abakaroona
ClaudesCarving
RealOwinoPaul
FemalePains
Sabya_n_Savya
AMusicFresno
Utkarsh7777
adithyajagtap
sudokuangel2
WesterosCraft
samyalshehry
WOODStheMAV
Patrywin
LoriWCox1
PeggyRJones1
bonniesrg
lauravarnam
XPNZVLZT
GokulPrashant13
DireJackalope
Factable
marilamarelo
atulzac
nysportstalk1
RebeccaCann23
shreyash_90
clemmie_s
vidit_singh
MFrizqi
Reiji_666
theashishraz
dvlakh
Digno1994
RarePopFacts

Degree Centrality:

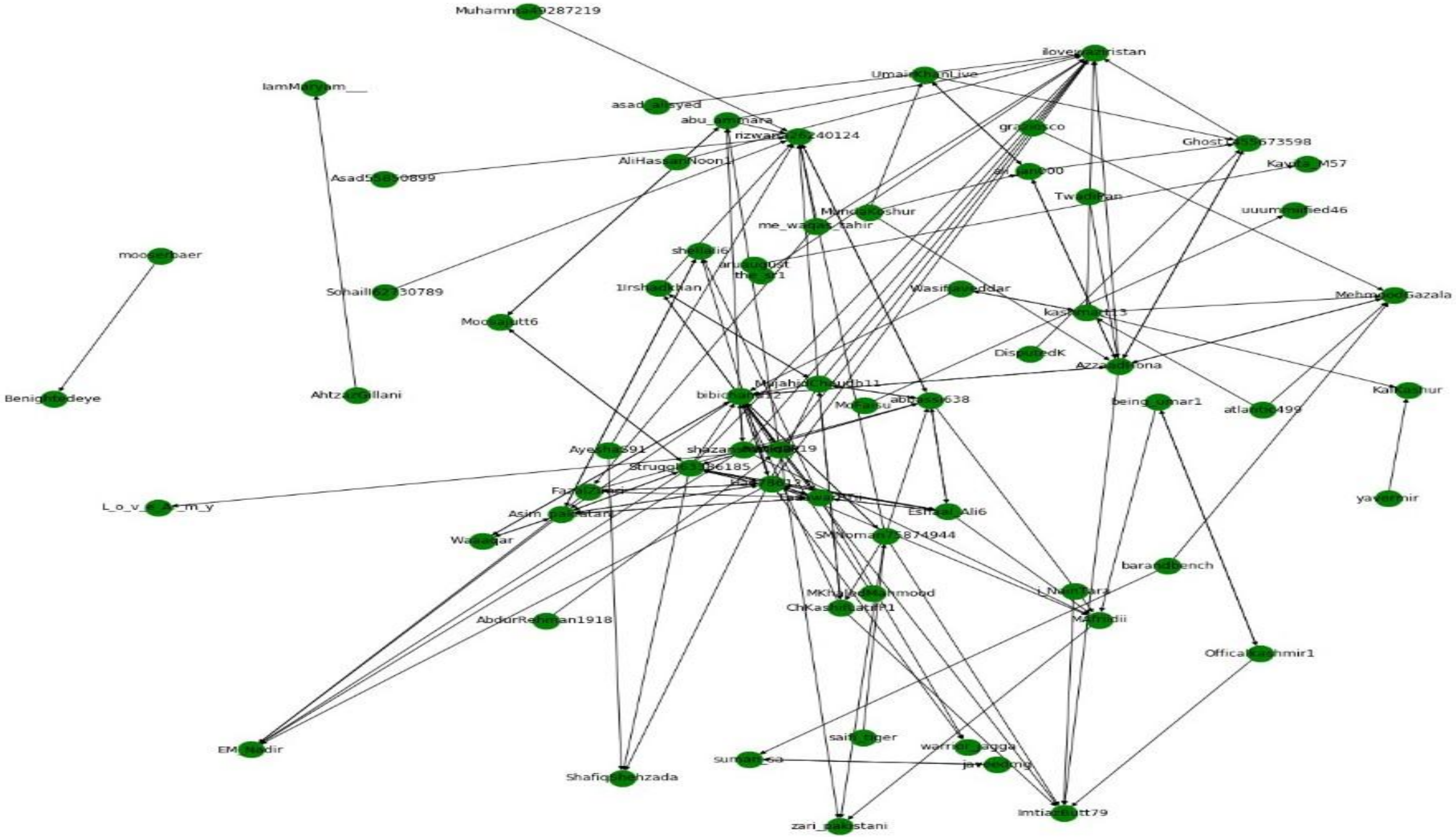


Eigen Vector:



Follower Followee for Kashmir:

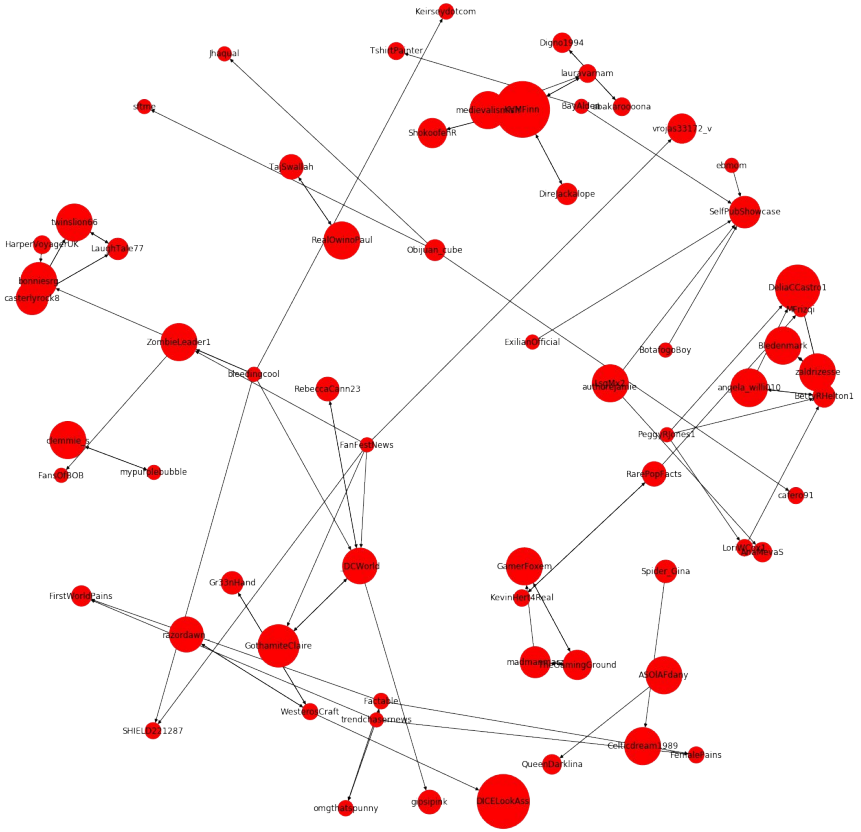
- Number of nodes: 66
- Number of edges: 152



lamMaryam___
abu_ammara
rizwan
AliHassanNoon1
UmairKhanLive
Kavita_M57
me_waqas_tahir
1Irshadkhan
Moosabutt6
MujahidChaudh11
DisputedK
being_umar1
atlantic499
yavarmir
Waqaqar
SMNoman75874944
MKhalidMahmood
barandbench
AbdurRehman1918
OfficalKashmir1
saifi_tiger
zari_pakistani
ImtiazButt79
Shafiqshehzada
MehmoodGazala

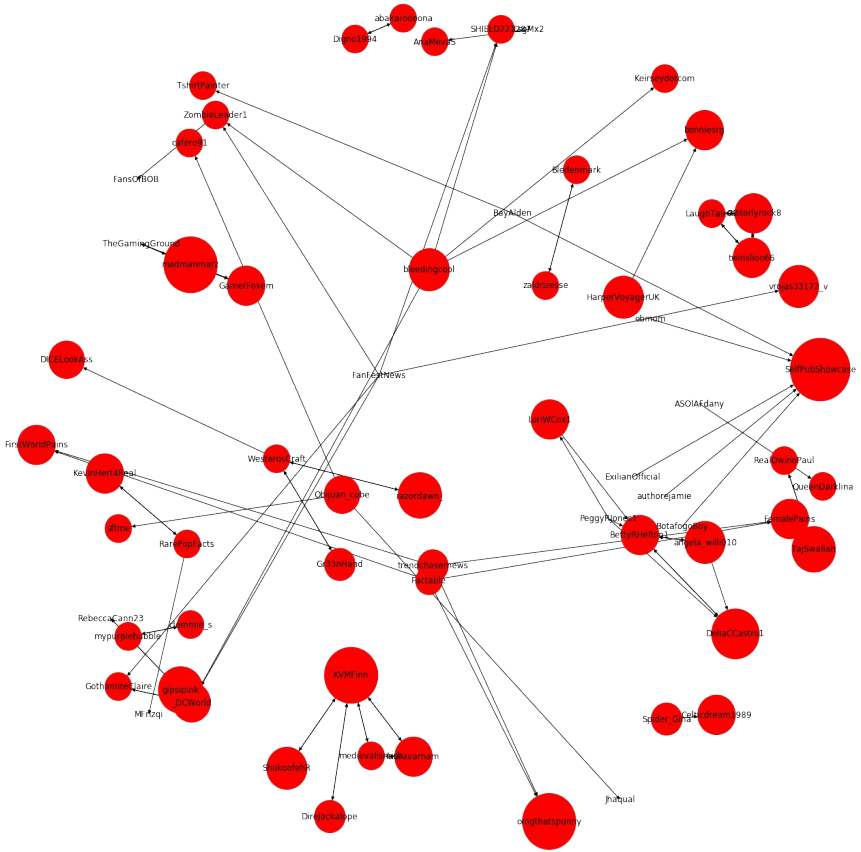
Page Rank:

PageRank works by counting the number and quality of edges to a node to determine a rough estimate of how important the node is. The underlying assumption is that more important nodes are likely to receive more links from other nodes.



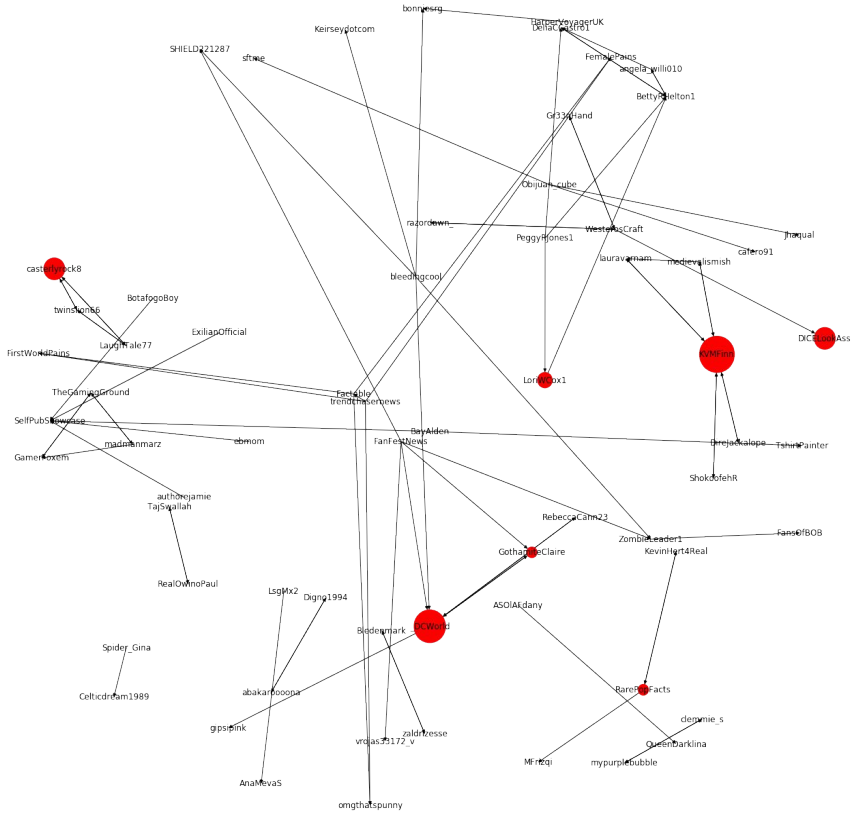
Closeness:

In a connected graph, closeness centrality (or closeness) of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus, the more central a node is, the *closer* it is to all other nodes.

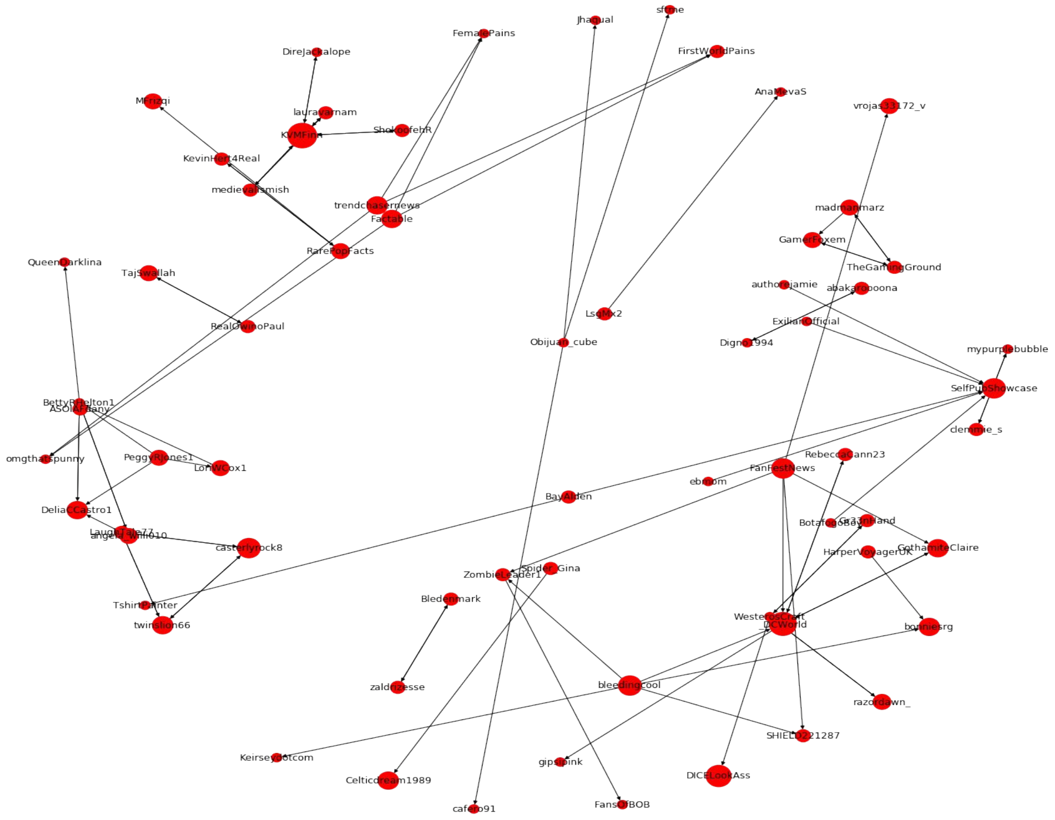


Betweenness:

Betweenness is a centrality measure of a vertex within a graph. Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.



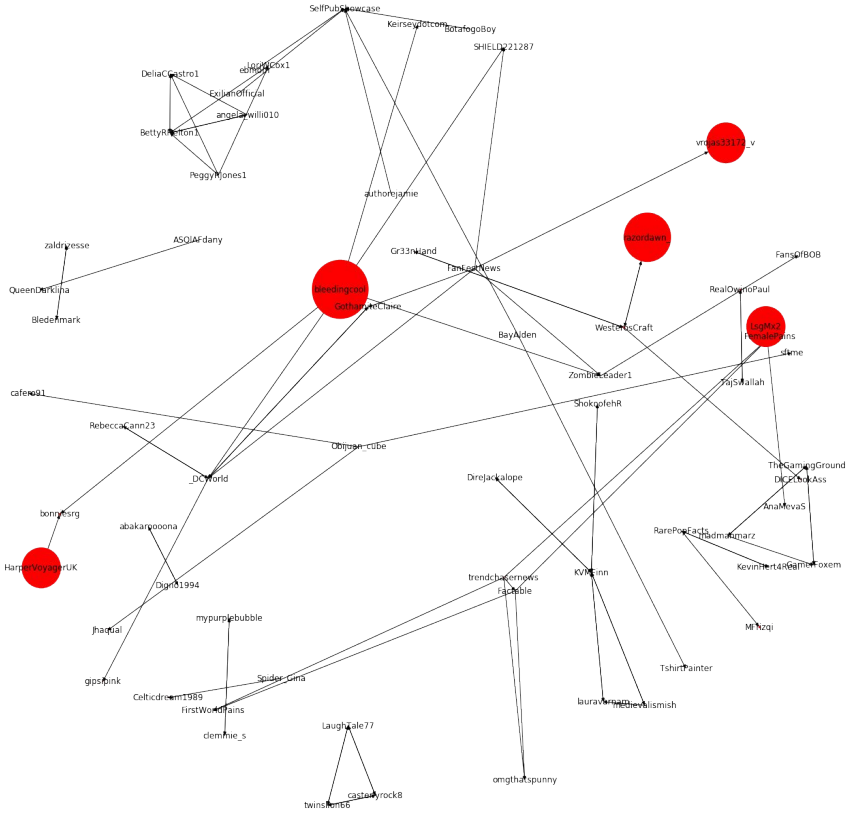
Degree:



sftme
Jhaqual
FemalePains
FirstWorldPains
DireJackalope
AnaMevaS
vrojas33172_v
MFrizqi
lauravarnam
KVMFinn
ShokoofehR
KevinHart4Real
medievalsmish
trendchasernews
Factable
madmanmarz
GamerFoxem
TheGamingGround
RareFopFacts
QueenDarklina
TajSwallah
authorjamie
abakarooona
LsqMx2
RealOwinoPaul
ExilianOfficial
Obijuan_cube
Digno1994
mypurplebubble
SelfPubShowcase
BettyPhelton1
ASOIAFfany
clemmie_s
omgthatspunny
PeggyRJones1
LoriWCox1
RebeccaCann23
FanFestNews
ebmom
BayAlden
DeliaCCastro1
BotafogoBeth
GrayinHand
LaughTale77
angela_whit010
casterlyrock8
HarperVoyagerUK
GothamiteClaire
ZombieLeader1
Spider_Gina
TshirtPainter
Bledenmark
WesterosCraft
_DCWorld
twinslion66
bonniesrg
zaldrizesse
bleedingcool
razordawn_
SHIELD221287
Keirseydotcom
gipsypink
DICELookAss
Celticdream1989
cafero91
FansOfBOB

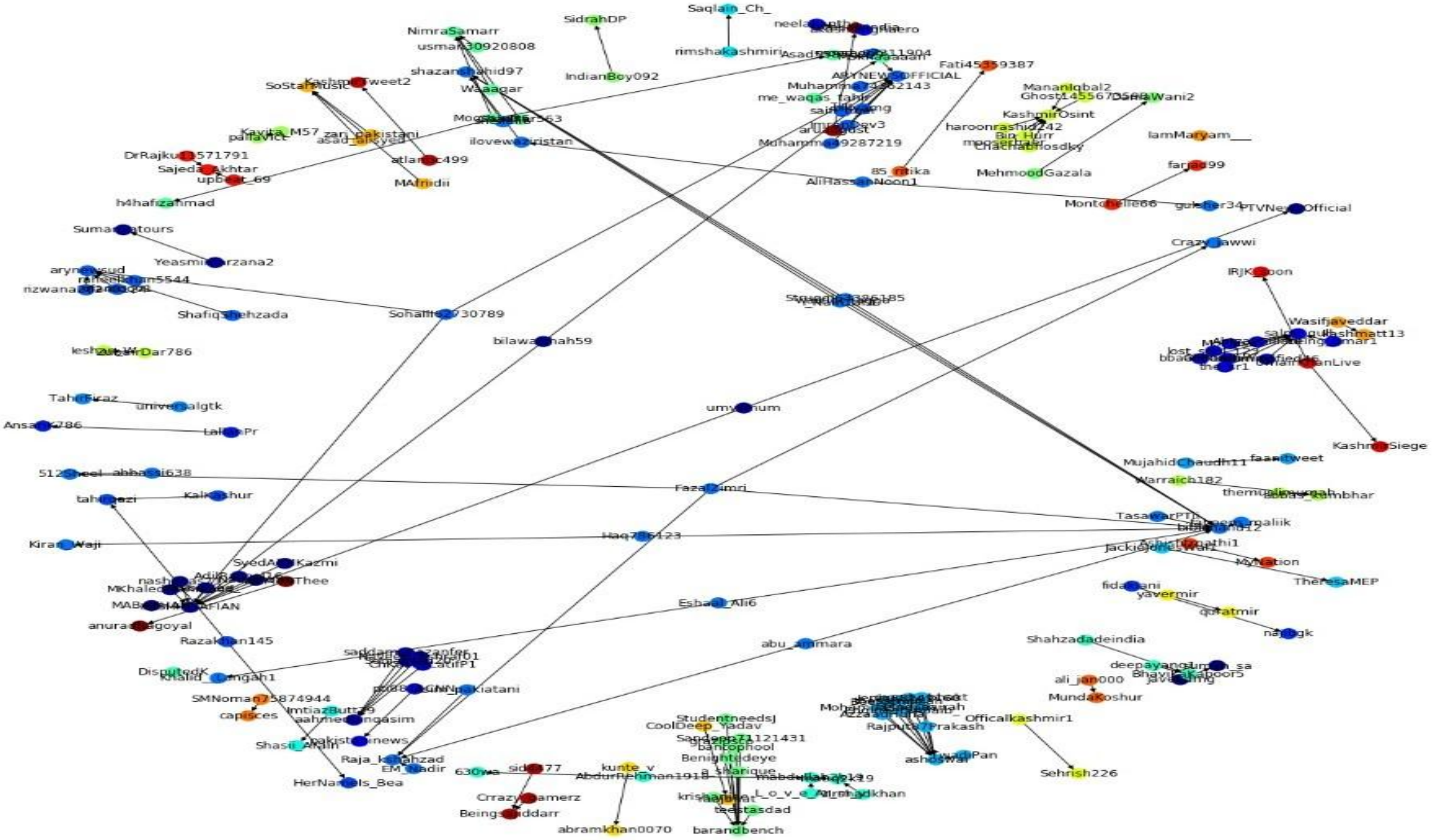
Eigen Vector:

Eigenvector centrality is a measure of the influence a node has on a network. If a node is pointed to by many nodes then that node will have high eigenvector centrality.

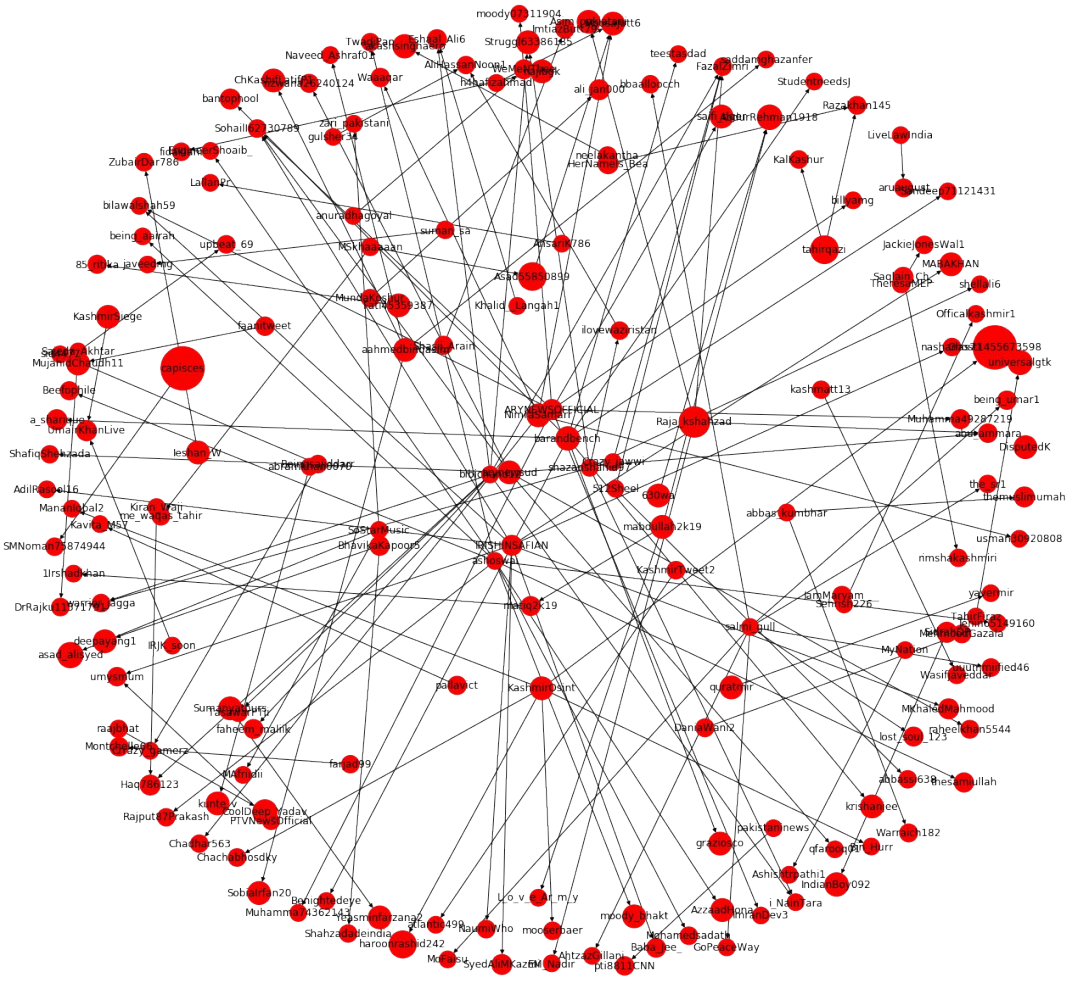


Retweet Graph for Kashmir:

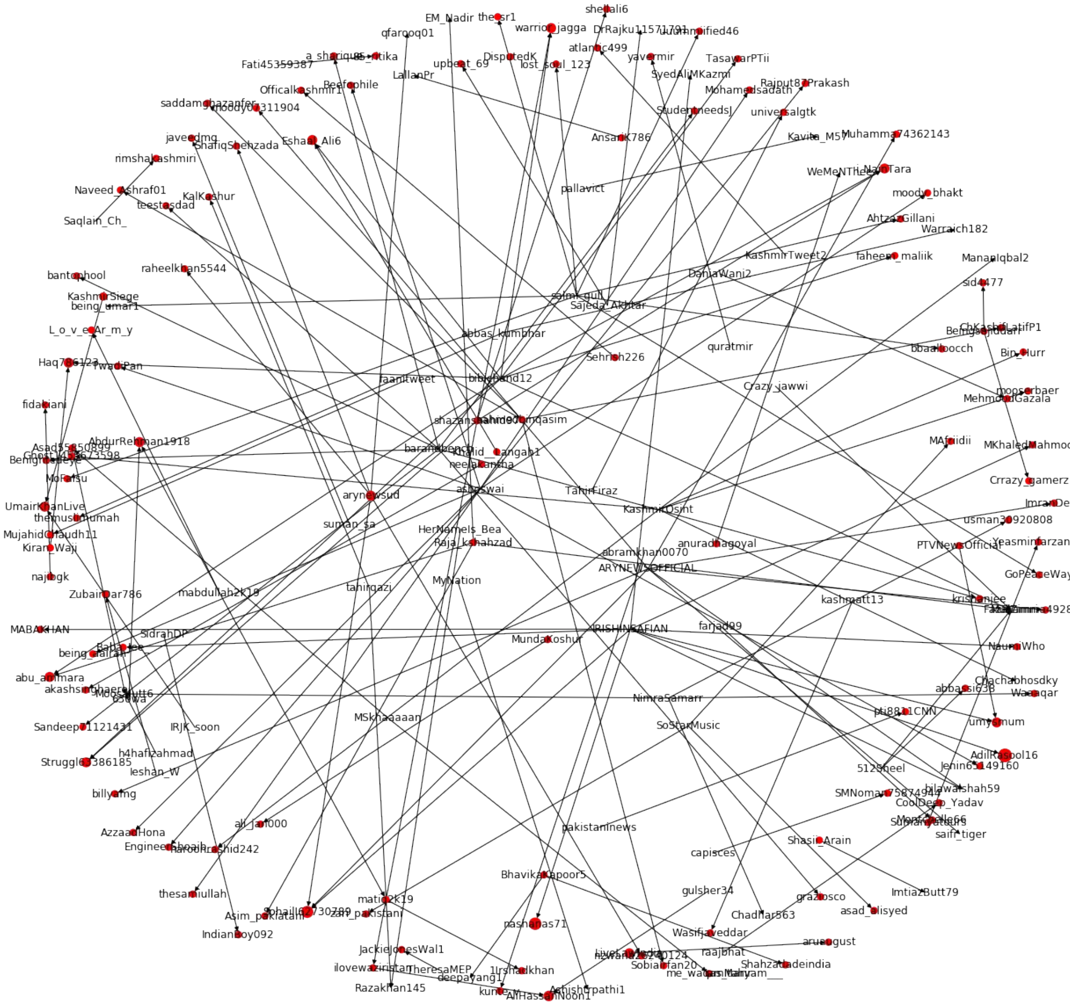
- Number of Nodes : 184
- Number of edges : 145


Saqlain_Ch_
SidrahDP
NimraSamarr
usman30920808
IndianBoy092
rimshakashmiri
shazanshahid97
Fati45359387
ARYNEWSOFFICIAL
KashmirTweet2
Waaagar
MananIqbal2
KashmirOsint
ilovewaziristan
IamMaryam___
atlantic499
DrRajkum11571791
Sajeda_Akhtar
MAfridii
85_ritika
MehmoodGazala
h4hafizahmad
AliHassanNoon1
Montchelle66
farjad99
PTVNewsOfficial
Yeasmin_Farzana2
Shafiqshehzada
Sohail62730789
IRJKNoon
bilawalshah59
Wasifjaveddar
umyanum
KashmirSiege
MujahidChaudh11
Warraich182
Fazal2imri
TasawarPTI
Haq786123
Kiran_Waji
MyNation
TheresaMEP
Eshaal_Ali6
anuradhagoyal
Razakhan145
abu_ammara
Shahzadadeindia
DisputedK
SMNoman75874944
capisces
ali_jan000
MundaKoshur
StudentneedsJ
CoolDeep_Yadav
Officalkashmir1
Shasii_Amin
Raja_kshahzad
EM_Nadir
HerNamels_Bea
kunte_v
abramkhan0070
Sehrish226
barandbench
teestasdad

Page Rank:

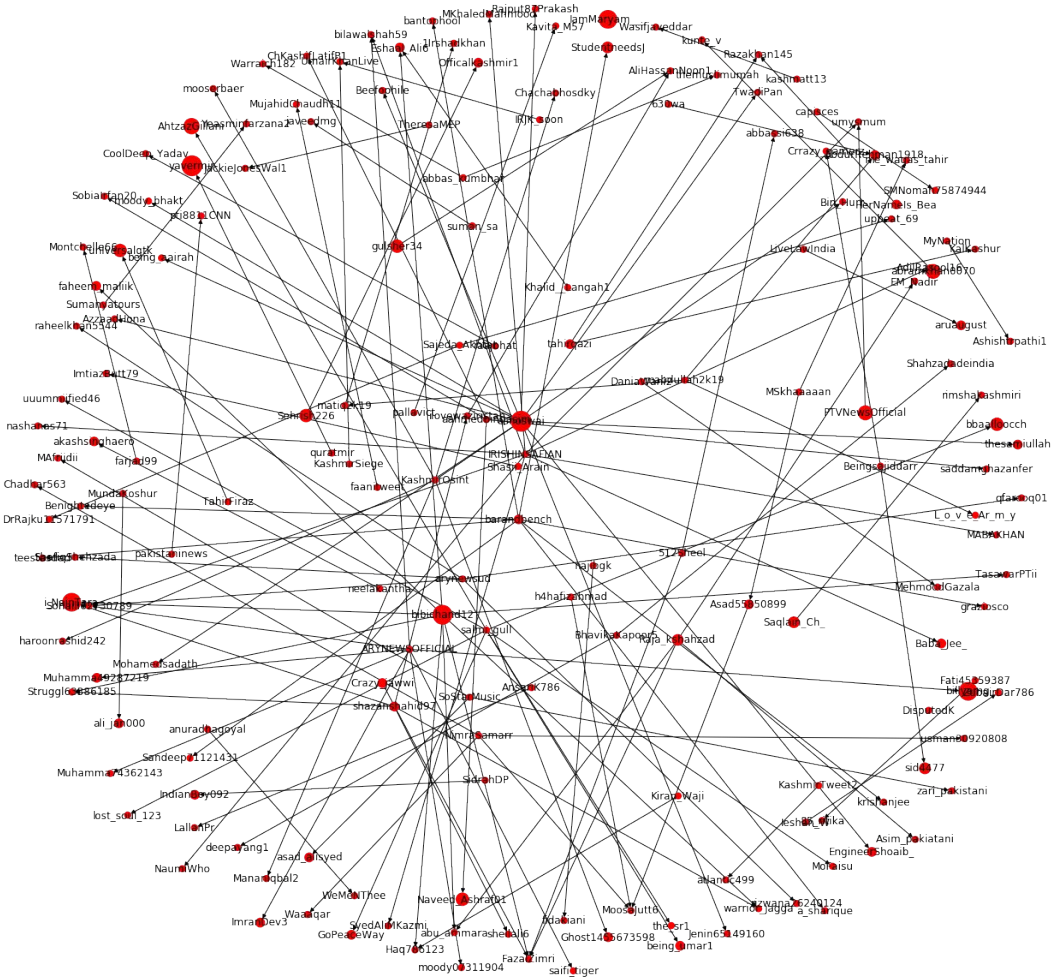


Closeness:

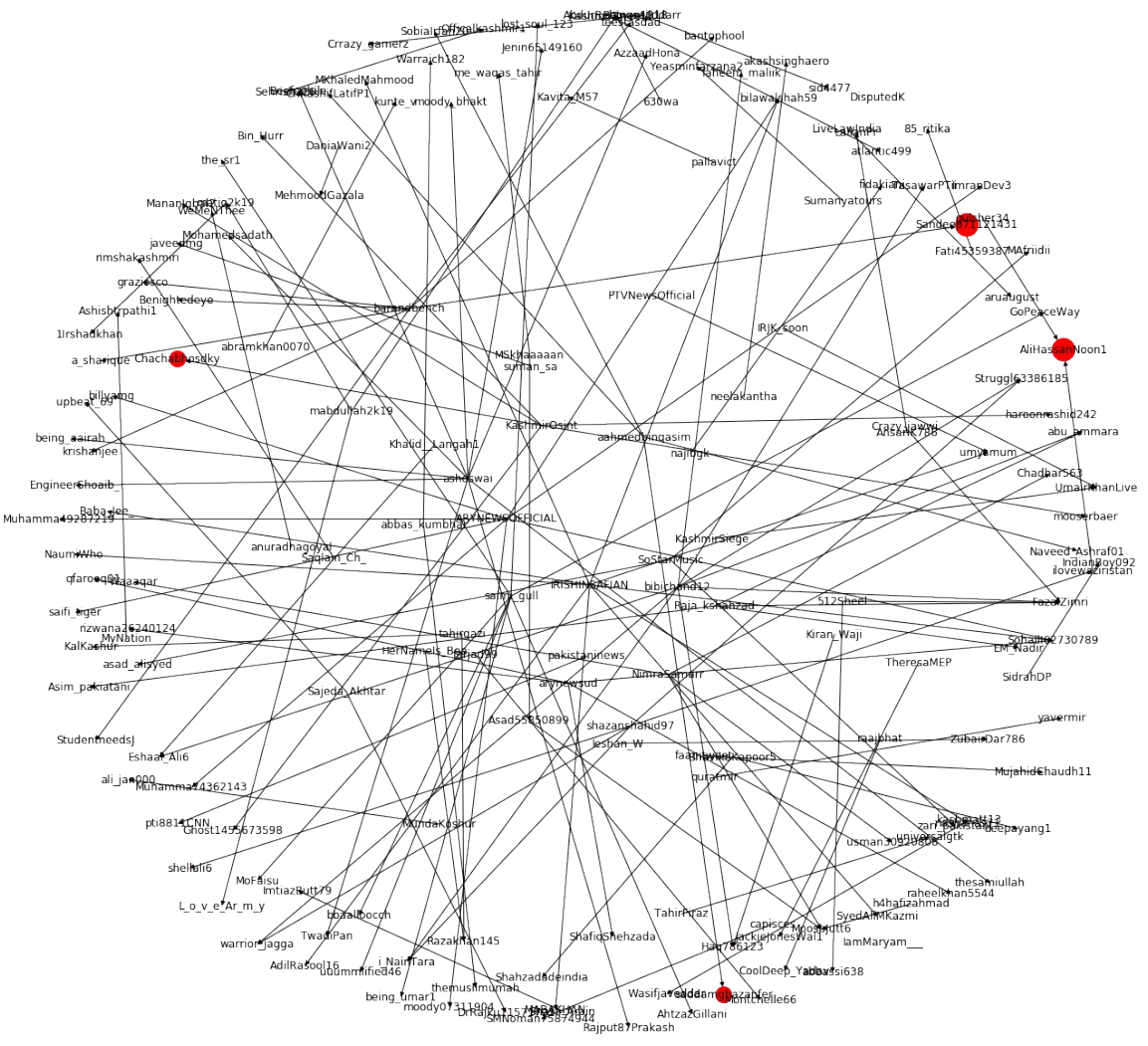


EM_Nadir
qfarooq01
shehali6
TasawarPTii
Rajput87Prakash
universalgtk
AnsarK786
Muhammad74362143
moody_bhakt
AhtzazGillani
Warraich182
ManarIqbal2
sid4477
Bin_Hurr
pallavict
quratmir
Sehrish226
Crazy_jawwi
KashmirTweet2
faheem_maliik
abbas_kumbhar
raahitweet
javeedmq
ShafiqShehzada
Naveed_Ashraf01
Saqlain_Ch_
KalKashur
raheelkhan5544
bantaphool
L_o_v_e_Ar_m_y
fidahiani
AbdurRehman1918
MoFarisu
UmairKhanLive
Kiran_Waji
najibgk
Zubairzar786
mabdullah2k19
MABAKHAN
SidrahDP
abu_ammara
Sandeep1121431
IRJK_soon
h4hafizahmad
leshan_W
billyamg
AzzaadHona
ali_ja1000
thesamiullah
Indianboy092
matin2k19
JackieJonesWal1
Razakhan145
AliHassanNoon1
Ashishtripathi1
pakistaninews
BhavikaKapoor5
capisces
gulsher34
grazosco
Chadhar563
ImtiazButt79
asad_alisyed
Shasii_Arain
Wasifjaveddar
aruengust
raajbhat
Shahzadadeindia
MSkhaaaaan
SoStarMusic
NimraSamarr
MundaKoshur
farjad99
TahirFiraz
KashmirOsint
anuradhagoyal
kashmatt13
HerNamels_Bea
Raja_kShahzad
abramkhan0070
ARYNEWSOFFICIAL
MyNation
tahirqazi
suman_sa
arynewsud
NauraWho
Waqqar
512Sheel
Jenin65149160
saifi_tiger
usman30920808
PTVNewsOfficial
GoPeaceWay
MAfridii
Crrazy_gamerz

Degree:



Eigen Vector:

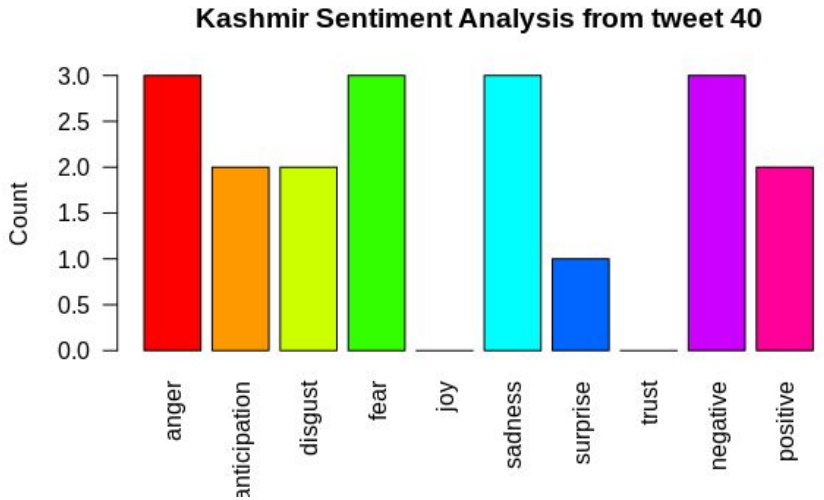


Part 2: Is there is shift in the sentiment over time?

- Kashmir Dataset

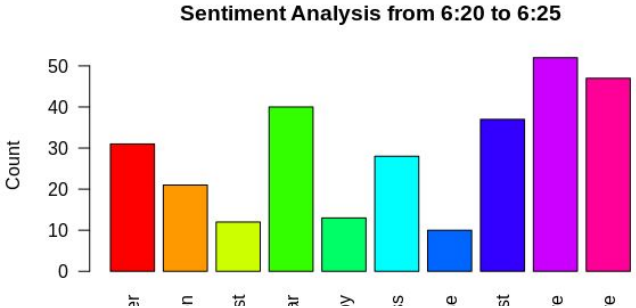
#Kashmir—Day 95

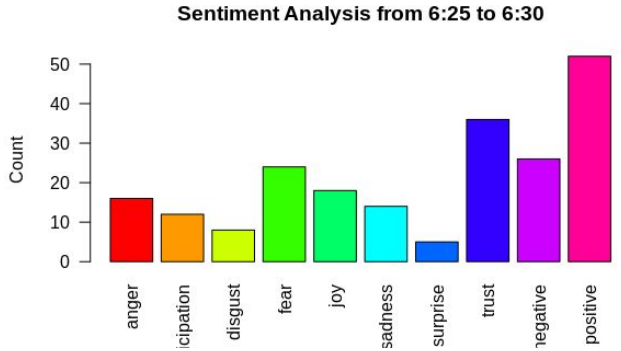
-Internet blocked
-Pre-paid mobile services blocked
-Businesses partially-closed, Rs 11k crore loss
-Schools/Colleges deserted
-Public transport off roads
-IT sector dead
-Tourism down by 85%
-100k job losses
-1000s in jails, untraceable
-Mental distress on rise

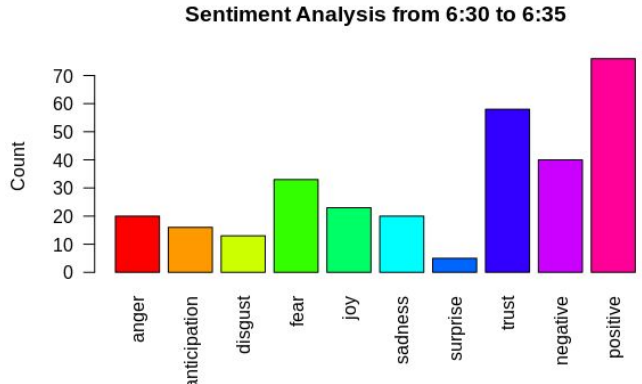


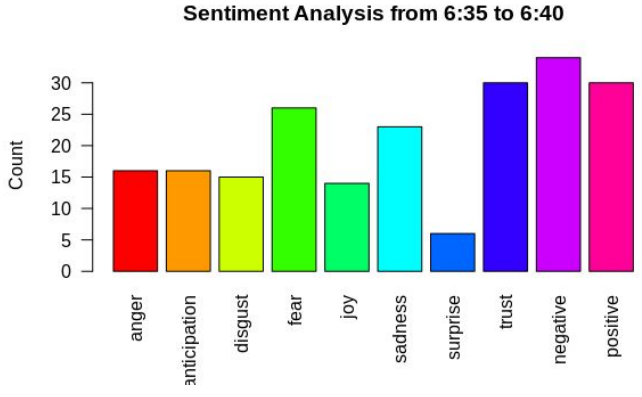
Part 2: Is there is shift in the sentiment over time?

- Kashmir Dataset





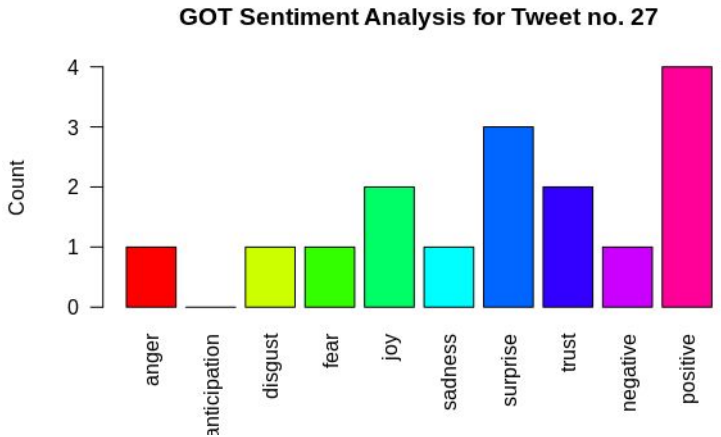




Part 2: Is there is shift in the sentiment over time?

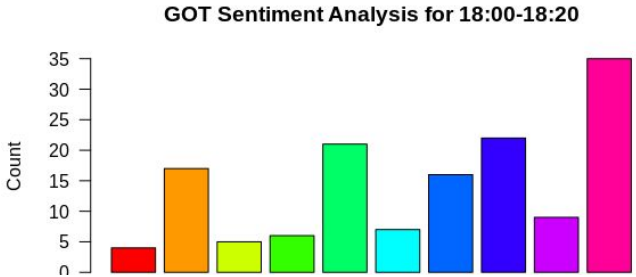
- GameOfThrones Dataset

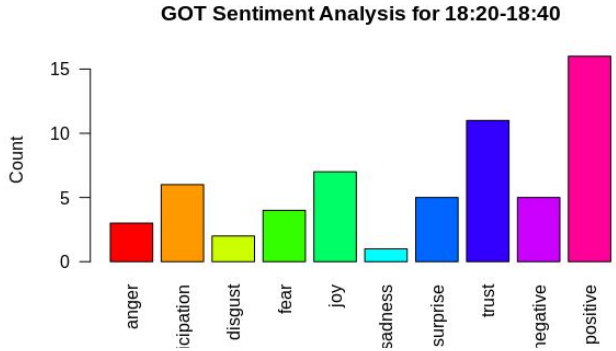
"Do u love #Sales? We got a huge 1 for ya! If u r a fan of #TheWalkingDead #GameOfThrones #StrangerThings #Arrow #Flash #Marvel #HarryPotter #Horror #DoctorWho and more, don’t miss our inventory reduction sale on designs inspired by our favs in #POPCulture #SupportSmallBusiness https://t.co/eCTNl26y2F"

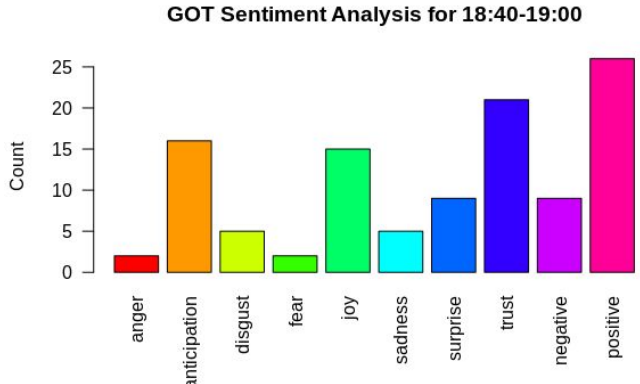


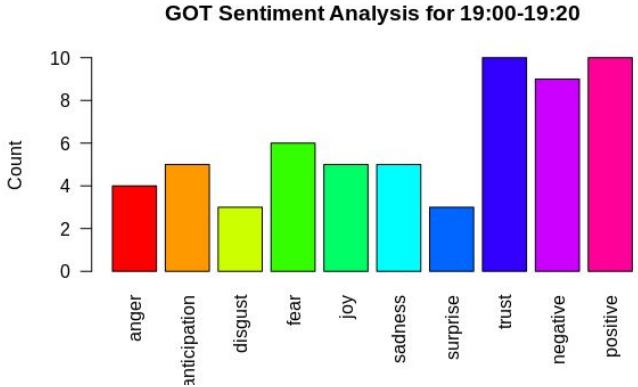
Part 2: Is there is shift in the sentiment over time?

- GameOfThrones Dataset



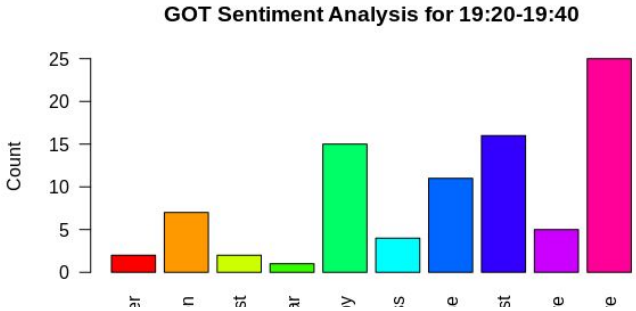


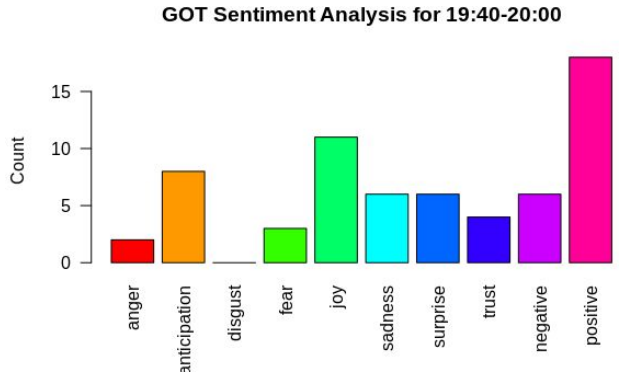


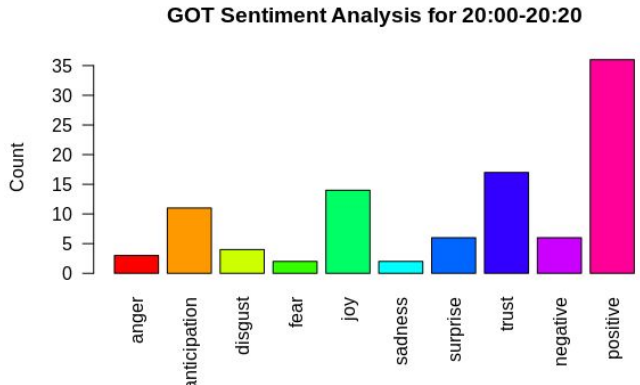


Part 2: Is there is shift in the sentiment over time?

- GameOfThrones Dataset







Is there any observable social learning over sentiments expressed ?

Degroot model: Take a society of agents where everybody has an opinion on a subject, represented by a vector of probabilities . Agents obtain no new information based on which they can update their opinions but they communicate with other agents. Links between agents (who knows whom) and the weight they put on each other's opinions is represented by a trust matrix T. Formally, the beliefs are updated in each period as

$$p(t) = Tp(t-1)$$

Results for Kashmir dataset:

- Initial opnions:

```
array([-1.25000000e-01,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  1.34615385e-01,  0.00000000e+00,
       -5.35714286e-02,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        1.27500000e-01,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
```

- Change in opinions:

```
[ 0.          0.05338542  0.015       0.21        0.21        0.
 -0.02678571  0.13461538  0.13461538 -0.03526187 -0.03513099  0.
 -0.03517316 -0.03521826 -0.03521924  0.         -0.01978175 -0.01590909
 -0.01464286 -0.03528627 -0.03510685 -0.03528627 -0.5         0.04761905
 -0.23148148  0.          0.         -0.5        -0.05       -0.00535714
```

Results for GOT dataset:

- Initial opinion:

```
array([ 0.        ,  0.1875    ,  0.2       ,  0.5       ,  0.1875    ,
        0.5       ,  0.5       ,  0.1875    ,  0.        ,  0.        ,
        0.        ,  0.1875    ,  0.03116883,  0.06597403,  0.10416667,
        0.25      ,  0.56818182,  0.        ,  0.        ,  0.        ,
        0.        ,  0.        ,  0.        ,  0.        , -0.2       ,
       -0.05      ,  0.        ,  0.        ,  0.23333333,  0.        ,
        0.5       ,  0.        ,  0.23333333,  0.        ,  0.        ,
        0.5       ,  0.        ,  0.23333333, -0.23181818, -0.23181818,
```

- Change in opinions:

```
[ 1.87500000e-01  0.00000000e+00  5.00000000e-01  2.33333333e-01
  5.00000000e-01  1.87500000e-01  5.00000000e-01  0.00000000e+00
  6.82575758e-02  0.00000000e+00  5.00000000e-01  3.11688312e-02
  1.87500000e-01  1.04166667e-01  7.89935065e-02  1.04166667e-01
  0.00000000e+00  2.08131313e-01  0.00000000e+00  0.00000000e+00
  1.04166667e-01  1.04166667e-01  0.00000000e+00 -4.37500000e-02
 -5.00000000e-02  1.25000000e-02  0.00000000e+00  0.00000000e+00
  0.00000000e+00  3.66666667e-01  0.00000000e+00  2.33333333e-01
```

Bot Detection:

1. The data we have does not contain labels defining which account is bot and which is not. So we have used unsupervised learning model in this part to detect bots.
2. We are using K-Means to detect bots in the dataset.
3. We pass important features to K-Means and divide the screen_names into 2 clusters, either of which can be bot class.
4. Then we again apply filters to the data to determine the possible bots in the dataset.
5. The features that we pass to the model is as follows:

Table 2. Bot Classification Variables By Area of Analysis

Area of Analysis	Variable
Profile	Absence of id
	Absence of a profile picture
	Absence of a screen name
	Has less than 30 followers
	Not geo-located
	Language not set to English
	Description contains a link
	Has sent less than 50 tweets
	2:1 friends/followers ratio
	Has over 1,000 followers
	Has the default profile image
	Has never tweeted
	50:1 friends/followers ratio
	100:1 friends/followers ratio
	Absence of a description
Text Analysis	Levenshtein distance between user's tweets is less than 30

	user_id	screen_name	lang	followers_count	friends_count	profile_url	description	location	ratio	cluster
0	1	1	1	0	1	1	1	0	0.0	1
1	1	1	1	1	0	1	1	1	0.0	1
2	1	1	0	0	1	0	1	1	0.0	1
3	1	1	0	0	1	0	1	1	0.0	1
4	1	1	1	1	1	0	1	1	0.0	0
...	...	...	...	...	...	...	...	...	...	...
195	1	1	1	1	1	0	1	1	1.0	0
196	1	1	1	0	1	0	1	0	0.0	1
197	1	1	1	1	1	0	0	0	0.0	0
198	1	1	1	0	1	0	0	1	0.0	1
199	1	1	0	1	1	1	1	1	0.0	0

Results for GOT dataset:

Possible Bots

1. Ksivananth
2. rushi_argula
3. Harshith_22
4. Sahulha24978411

```
if(df0['location'][i]==0):
    #print("inside ratio")
    if(df0['description'][i]==0):
        #print("inside loc")
        if(df0['followers_count'][i]==1):
```

Results for Kashmir dataset:

Possible bots:

1. Jenin65149160
2. usman30920808
3. IndianBoy092

```
if(df0['ratio'][i]==1):
    #print("inside ratio")
    if(df0['location'][i]==0):
        #print("inside loc")
        if(df0['followers_count'][i]==1):
            if(df0['description'][i]==1):
```

Stance Analysis : GOT dataset

Procedure :

1. Pre process dataset.
2. Identify polarity and subjectivity of tweets
3. Cluster users based on their modularity and polarity.
4. For each cluster :
    a. Identify polarity by max count
    b. Identify subjectivity by max count
    c. Calculate polarity
    d. Calculate subjectivity

Plot the result.