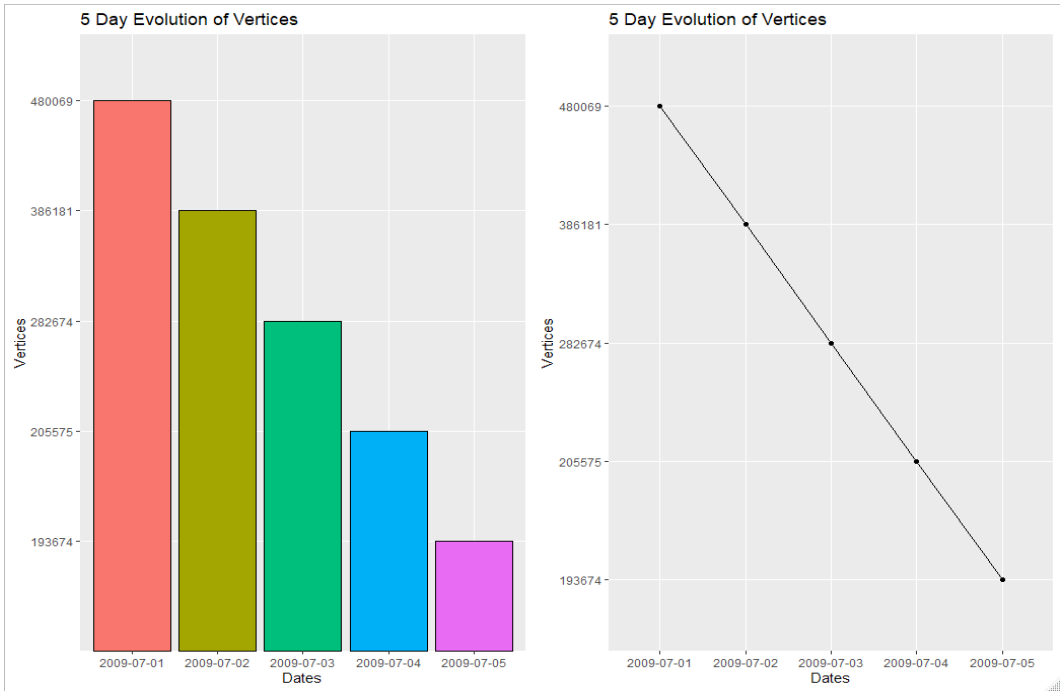
SNA Homework 2

George Vogiatzis (p2821827)

Part 1: Implemented by python script included in the zip archive.

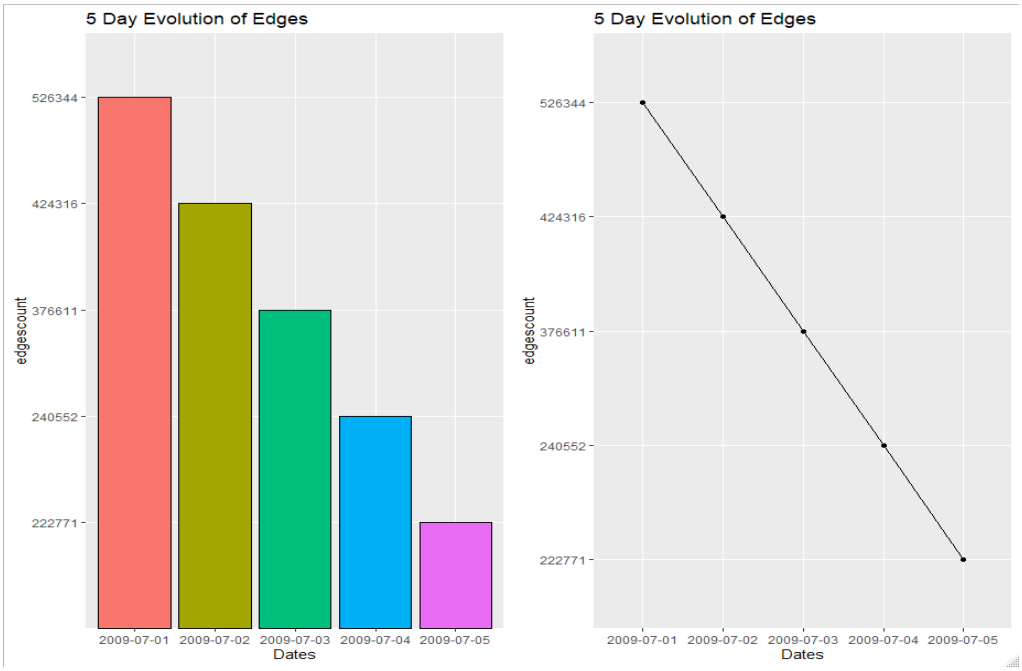
Part 2 Average Degree Over Time

5-day Evolution Nr of Vertices



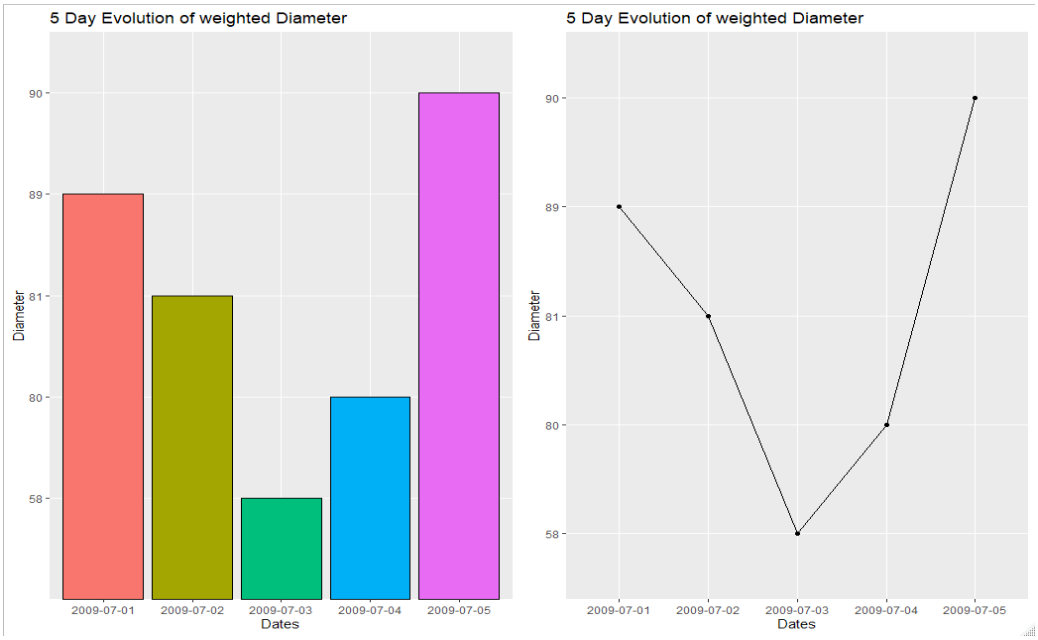
The number of vertices appears to be reducing as each day passes beginning July 1st. For example 480069 users posted in the first day while 193674 posted on the fifth of July.

5-day Evolution Nr of Edges



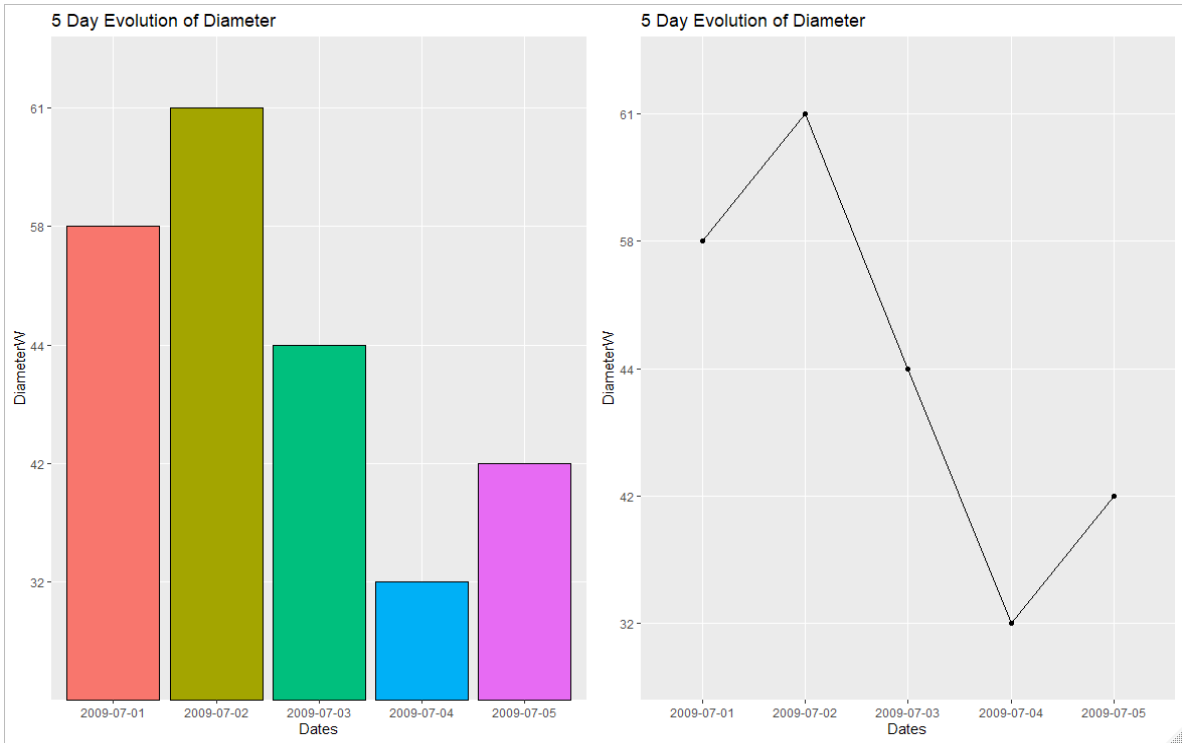
Normally as the number of nodes decline, the nr of edges declines as well, since there are fewer connection over the period of the first days of July. The drop was from 526344 edges to 222771 on the 5th of July.

5-day Evolution of Weighted Diameter



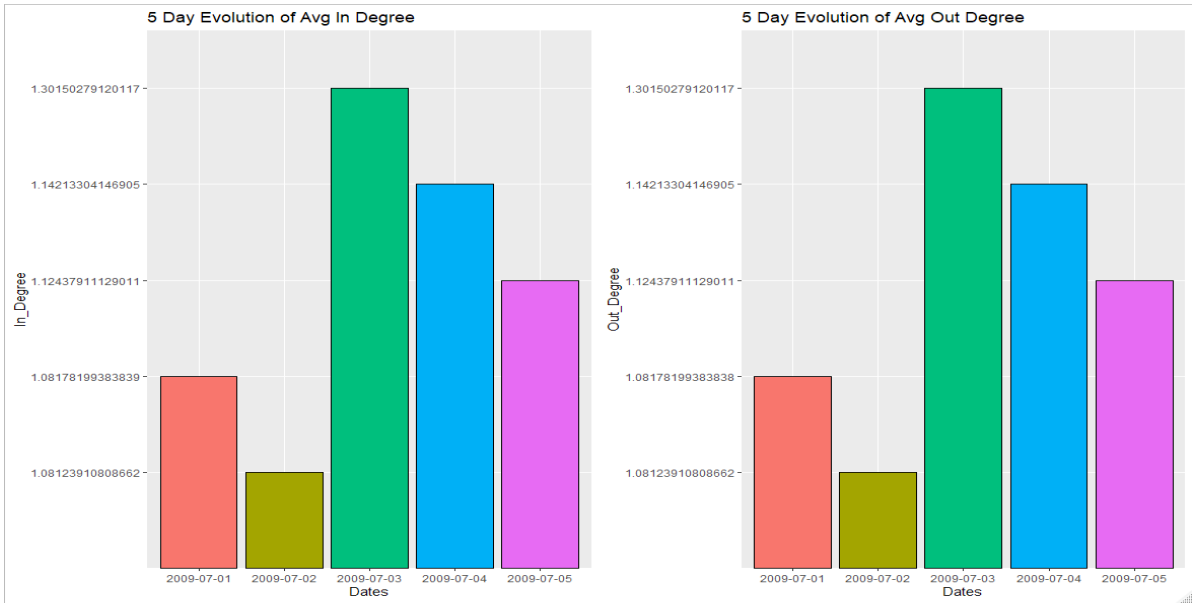
The weighted diameter was 89 in the first day,dropped to 58 in the 3rd of July and peaked again at 90 in the 5th of July. This means that the paths between the nodes became shorter in the mid period of the 5 days and reached the same level again on July 5th.

5-day Evolution Diameter



The diameter without the weight reveals that the paths between the most distant nodes which are 58, 61, 44, 32 and 42 respectively.

5-day Evolution In & Out Degree



In degree and out degree follow a similar pattern in the first 5 days of July which is normal because for every connection to a user we have a different connection from the same user.

Part3 Important nodes

In Degree Top10 July-1-2009

```
in_degree_1

rafinhabastos          705

GuyKawasaki            730

cnn                    761

aplusk                 852

BreakingNews           900

mileycyrus             948

addthis               1418

smashingmag           1670

mashable              1847

tweetmeme             3498
```

In Degree Top10 July-2-2009

```
      in_degree_2

mileycyrus               886

souljaboytellem         1074

cnn                     1234

addthis                 1267

cnnbrk                  1420

mashable                2227

officialtila            2236

OfficialTila            2517

tweetmeme               3380

ddlovato                4766
```

In Degree Top10 July-3-2009

```
in_degree_3
```

imeem	596
Jeepersmedia	715
cnnbrk	901
PhillyD	917
moontweet	920
BreakingNews	1059
mashable	1081
addthis	1221
souljaboytellem	2063
tweetmeme	2579

In Degree Top10 July-4-2009

in_degree_4	
cnnbrk	540
lilduval	556
mashable	616
souljaboytellem	678
mileycyrus	684
iamdiddy	723
songzyuuup	950
addthis	1029
BreakingNews	1199
tweetmeme	1343

In Degree Top10 July-5-2009

in_degree_5	
AKGovSarahPalin	440
imeem	461
moontweet	488
mileycyrus	545
mashable	636

```
BreakingNews             705

addthis                 1095

tweetmeme               1597

davidmmasters           2178

iamdiddy                2980
```

The top users in terms of In Degree appear to be changing over time and they appear to be people or users trending in the period analysed.

Out Degree Top10 July-1-2009

```
out_degree_1

juliesearser             166

lmaobot                  171

jamokie                  183

vaguetweetstest          196

wootboot                 201

the_sims_3               221

tsliquidators            224

failbus                  251

dudebrochill             259

teamqivana               412
```

Out Degree Top10 July-2-2009

```
out_degree_2

marcodane79              158

dvdbot                   161

pheasantphun             174

the_sims_3               183

thickdecadence           204

failbus                  206

modelsupplies            214

wootboot                 241

penishunter              273
```

dudebrochill	291

Out Degree Top10 July-3-2009

out_degree_3	
azandiamjbb	353
ohmichael	371
thickdecadence	427
nachhi	431
killah360dhh	451
andreapuddu	466
java4two	470
imbeeyo	537
deana1981	628
drejones71	794

Out Degree Top10 July-4-2009

out_degree_4	
nachhi	283
modelsupplies	293
dj_fresh	353
azandiamjbb	363
wootboot	365
itz_cookie	368
dudebrochill	396
hoboprophet	463
andreapuddu	487
swbot	863

Out Degree Top10 July-5-2009

out_degree_5	
dudebrochill	331
hoboprophet	472

```
apeeescape              536

bilbo232                572

wildingp                640

twiprodigy005           672

twiprodigy008           808

twiprodigy007           812

twiprodigy009           824

swbot                   951
```

The Top-10 users in term of Out Degree are changing more than those of the In Degree in that period of time and they do not appear in the same frequency, as those of the In Degree, in the following days of July until the 5th.

Page Rank Top10 July-1-2009

```
>Rank_day1
>rafinhabastos
>aplusk
>CourageCampaign
>KISSmetrics
>mileycyrus
>cnn
>smashingmag
>addthis
>mashable
>tweetmeme
```

Page Rank Top10 July-2-2009

```
>Rank_day2
>mileycyrus
>cnnbrk
>souljaboytellem
>addthis
>cnn
>globalmanners
>tweetmeme
>mashable
>drew_taubenfeld
>ddlovato
```

Page Rank Top10 July-3-2009

```
>Rank_day3
>adamlambert
>PhillyD
>BreakingNews
>mashable
```

```
>cnnbrk
>moontweet
>addthis
>killerstartups
>souljaboytellem
>tweetmeme
```

Page Rank Top10 July-4-2009

```
>Rank_day4
>garyvee
>cnnbrk
>iamdiddy
>mashable
>mileycyrus
>lilduval
>tweetmeme
>BreakingNews
>addthis
>souljaboytellem
```

Page Rank Top10 July-5-2009

```
>Rank_day5
>mileycyrus
>BreakingNews
>moontweet
>mrskutcher
>mashable
>tweetmeme
>aplusk
>addthis
>iamdiddy
>davidmmasters
```

The top 10 users in terms of PageRank appear to recycled during the period mentioned, which is normal since they are trending more than the rest of the users.

Part 4 Communities

Clustering Methods

I tried to apply fast greedy clustering, infomap clustering and louvain clustering on the undirected version of the graphs, however I obtained results using Louvain clustering only. Fast and greedy clustering & louvain clustering performance is as shown below. I ran fast and agreedy only for performance comparison purposes once as it took a lot of time to run. Could not obtain any results at all using infomap clustering.

```
system.time(communities_louvain3 <- cluster_louvain(jul3_undirected))

>                               user                system        elapsed
>    3.30    0.17    3.47
```

```
system.time(communities_fast_greedy3 <- cluster_fast_greedy(jul3_undire
cted))

>                               user                system        elapsed
>  1342.30    0.69  1356.43
```

Random User-Mashable Day 1 Membership

```
>                                                                  mashable
>    78078
```

Random User-Mashable Day 2 Membership

```
>                                                                  mashable
>    46103
```

Random User-Mashable Day 3 Membership

```
>                                                                  mashable
>    19196
```

Random User-Mashable Day 4 Membership

```
>                                                                  mashable
>    22068
```

Random User-Mashable Day 5 Membership

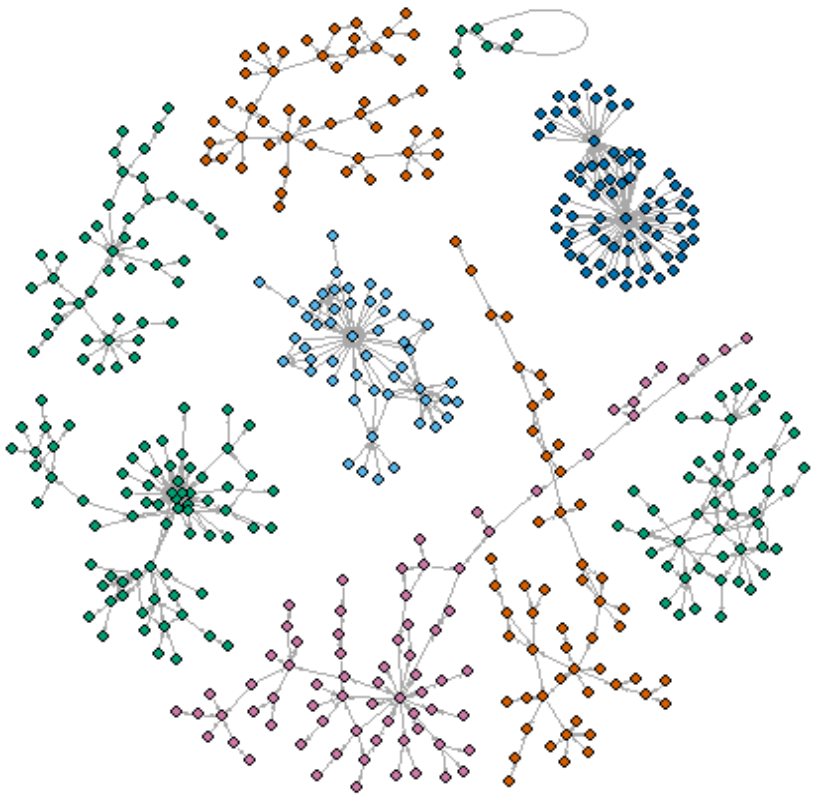
```
>                                                                  mashable
>    18279
```

Creating Community Sizes & Coloring different Communities

```
V(jul1)$color        <-        factor(membership(communities_louvain1))
V(jul2)$color        <-        factor(membership(communities_louvain2))
V(jul3)$color        <-        factor(membership(communities_louvain3))
V(jul4)$color        <-        factor(membership(communities_louvain4))
V(jul5)$color        <-        factor(membership(communities_louvain5))

community_size1             <-             sizes(communities_louvain1)
community_size2             <-             sizes(communities_louvain2)
community_size3             <-             sizes(communities_louvain3)
community_size4             <-             sizes(communities_louvain4)
community_size5 <- sizes(communities_louvain5)
```

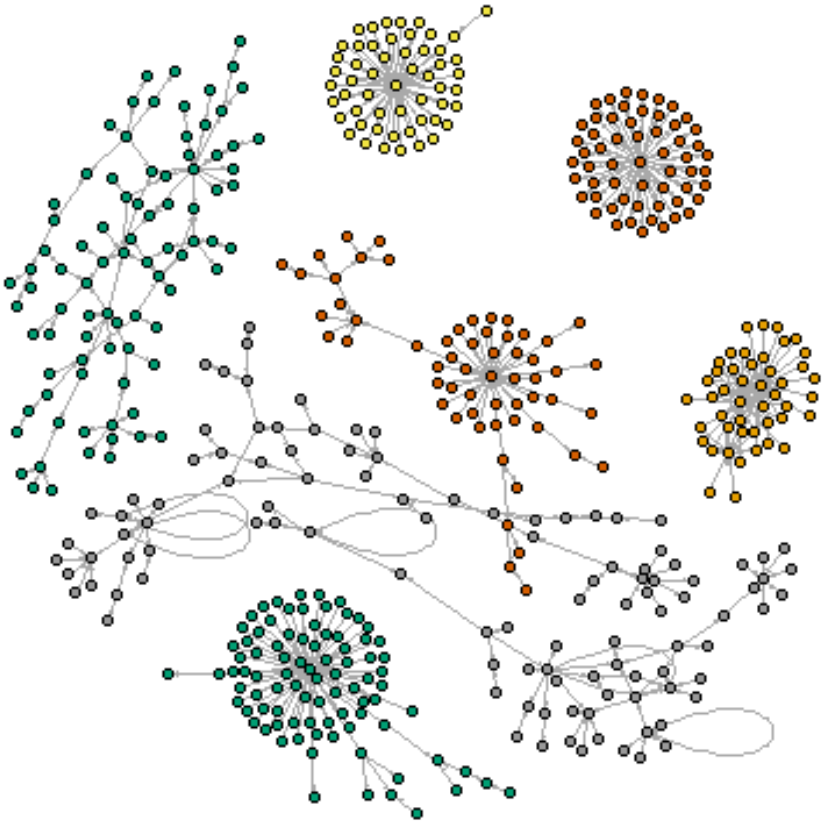
Graph Day 1 Communities

```
inMidCommunity1<-unlist(communities_louvain1[community_size1 > 40 & community_size1 < 80])

jul1Sbgrph <- induced.subgraph(jul1, inMidCommunity1)

plot(jul1Sbgrph, vertex.label = NA, edge.arrow.width = 0.8, edge.arrow.size = 0.2, coords = layout_with_fr(jul1Sbgrph), margin = 0, vertex.size = 3)
```



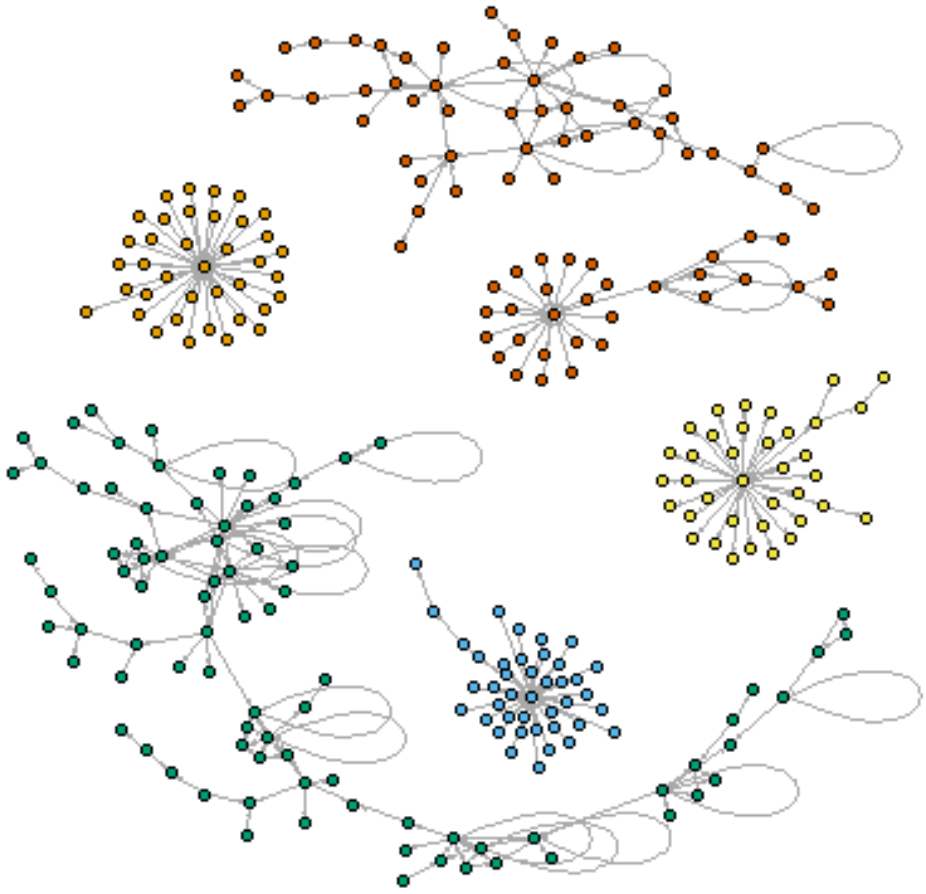
Graph Day 2 Communities

```
inMidCommunity2<-unlist(communities_louvain2[community_size2 > 40 & community_size2 < 90])

jul2Sbgrph<-induced.subgraph(jul2, inMidCommunity2)

plot(jul2Sbgrph, vertex.label = NA, edge.arrow.width = 0.8, edge.arrow.size = 0.2, coords = layout_with_fr(jul2Sbgrph), margin = 0, vertex.size = 3)
```



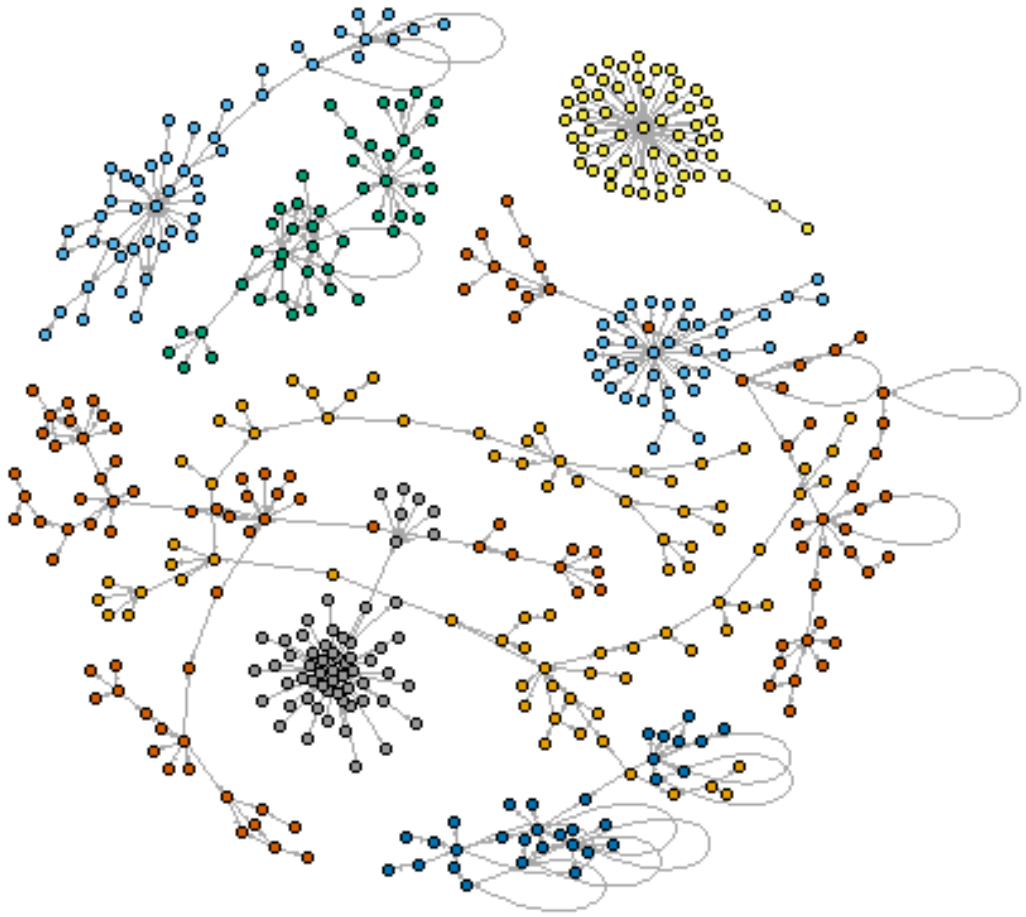
Graph Day 3 Communities

```
inMidCommunity3<-unlist(communities_louvain3[community_size3 > 30 & com
munity_size3 < 100])

jul3Sbgrph<-induced.subgraph(jul3, inMidCommunity3)

plot(jul3Sbgrph, vertex.label = NA, edge.arrow.width = 0.8, edge.arrow.
size = 0.2,coords = layout_with_fr(jul3Sbgrph), margin = 0, vertex.size
= 3)
```



Graph Day 4 Communities

```
inMidCommunity4<-unlist(communities_louvain4[community_size4 > 30 & com
munity_size4 < 100])

jul4Sbgrph<-induced.subgraph(jul4, inMidCommunity4)

plot(jul4Sbgrph, vertex.label = NA, edge.arrow.width = 0.8, edge.arrow.
size = 0.2,coords = layout_with_fr(jul4Sbgrph), margin = 0, vertex.size
= 3)
```



Graph Day 5 Communities

```
inMidCommunity5<-unlist(communities_louvain5[community_size5 > 40 & com
munity_size5 < 100])

jul5Sbgrph<-induced.subgraph(jul5, inMidCommunity5)

plot(jul5Sbgrph, vertex.label = NA, edge.arrow.width = 0.8, edge.arrow.
size = 0.2,coords = layout_with_fr(jul5Sbgrph), margin = 0, vertex.size
= 3)
```