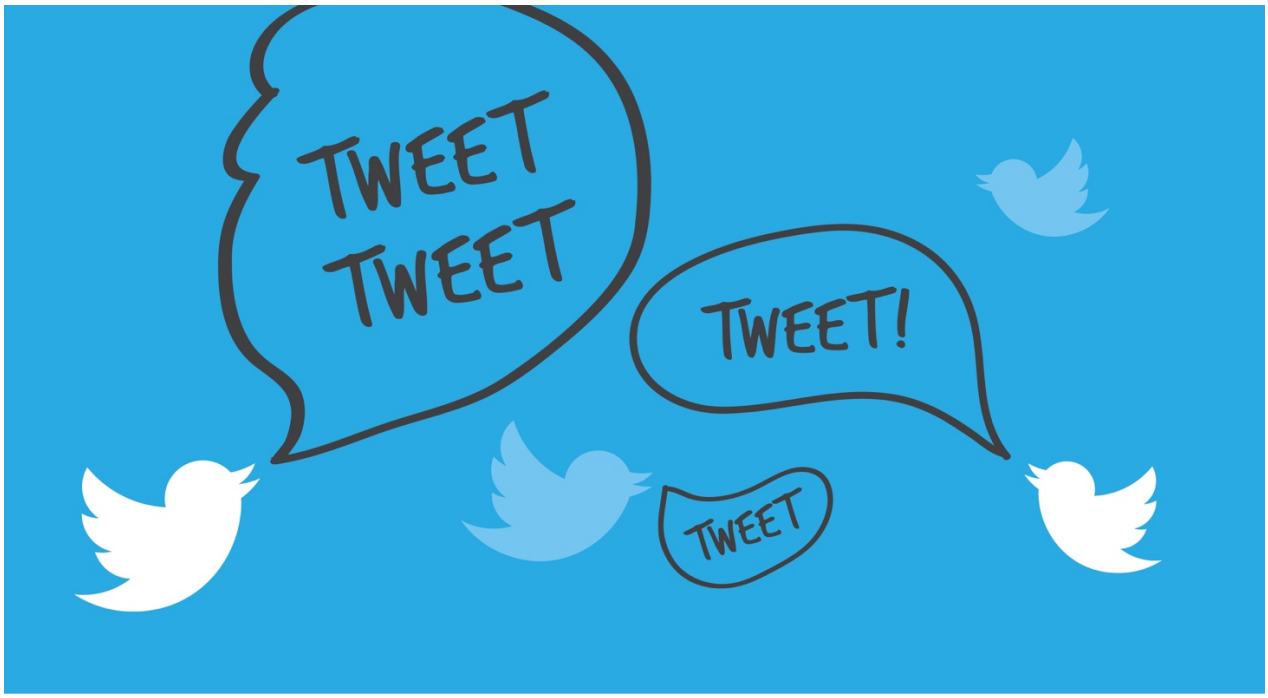


HOMEWORK 2

FROM RAW DATA TO TEMPORAL GRAPH STRUCTURE EXPLORATION

ELEFTHERIA APOSTOLAKI

COURSE: SOCIAL NETWORK ANALYSIS

INTRODUCTION

To begin with, it is worth mentioning that Python programming language was initially used to process the text file with the posted Tweets so as to create the requested csv files of this task. After generating these files, R programming language was used to subsequently plot the Twitter graphs and calculate anything required in the later questions.

PART 1: TWITTER MENTION GRAPH

An overview of the network directed graphs of 1st, 2nd, 3rd, 4th and 5th of July respectively can be seen below.

```
IGRAPH 1ebc18a DNW- 480069 526344 --
+ attr: name (v/c), weight (e/n), lty (e/c)
+ edges from 1ebc18a (vertex names):
 [1] rosietx1       ->chipcoffey      wackrepko      ->whatsupmartha   iammr2phly     ->ms_sugakane
 [4] leo02          ->katarnett       camronomari    ->brii_renee      cuesa          ->eatrealfest
 [7] egomonics      ->ploked          jennmae        ->shallomj        ltlv613        ->WayneGarciaKPTV
[10] studiofunhouse ->barchard        alyssaavz      ->carriieee       alyssaavz      ->melissa_makuch
[13] alyssaavz      ->mitchgeorge     alyssaavz      ->asiansteve      alyssaavz      ->themasonmusso
[16] brooke2404     ->Cynthia3200     carolinagrl91  ->AsiaMCekke      iambendavis    ->AddToAny
[19] istalker       ->QuelMarth       jetwhine       ->RunwayGirl      joeyineire     ->dharma_smokie
[22] justia         ->ErikJHeels      meganchichester->ViralURL        rinoua         ->happiijenny
+ ... omitted several edges
```

```
IGRAPH 3b1e11f DNW- 386181 424316 --
+ attr: name (v/c), weight (e/n), lty (e/c)
+ edges from 3b1e11f (vertex names):
 [1] kimn1214       ->thesounds       reternus       ->samuelchaney    aleara         ->extralife
 [4] anikaramirez   ->GeekTyrant      delfin54       ->globovision     sthomas1127    ->InsideHoops
 [7] blksheep       ->PhillyD         drkoob         ->MVPinBoynton    graceor        ->stephaniepratt
[10] helifreak      ->SavvyAuntie     kyralovemusic  ->eonline         marcea14       ->taylorswift13
[13] migreme        ->comunicadores   mzz_fantabulous->SongzYuuup      cesya          ->sammity
[16] dalgrev        ->emevieyra       enlightenedvi  ->alliejay        judyltang      ->travelzoo
[19] khanisha       ->candykizzes24   primuniz       ->foforks         teambudd1      ->adamsconsulting
[22] teambudd1      ->autonomy14      benniferrostock->CATv0nB         dudebrochill   ->XRockstarBabiiX
+ ... omitted several edges
```

```
IGRAPH 1127d94 DNW- 282674 376611 --
+ attr: name (v/c), weight (e/n)
+ edges from 1127d94 (vertex names):
 [1] hwhitehead     ->FoodFrontCoop   ini1911        ->netzpolitik     ini1911        ->sentral
 [4] livingrheum    ->christianfea    madness_hamster->Grater_Good     madness_hamster->Madness_Hamster
 [7] me2everyone88  ->davidcheyne     orangegirl1    ->pighair         orangegirl1    ->sweetchel2
[10] orangegirl1    ->caizer          popky          ->BridgetAyers    popky          ->applemacbookpro
[13] siharnik       ->kakaqui         siharnik       ->Inep_Imprensa   wannanoelisa   ->bowwow614
[16] ahzing         ->timbond         ahzing         ->MarkLeMunyon    audrey_gs      ->EqualityAmerica
[19] haifalicious   ->souljaboytellem mini_b_minaj    ->souljaboytellem mzdee718       ->MrMoto2nyce
[22] mzdee718       ->ConPhlict       sheririley     ->ESharpAgency    solemnankh     ->PISMOSHOU
+ ... omitted several edges
```

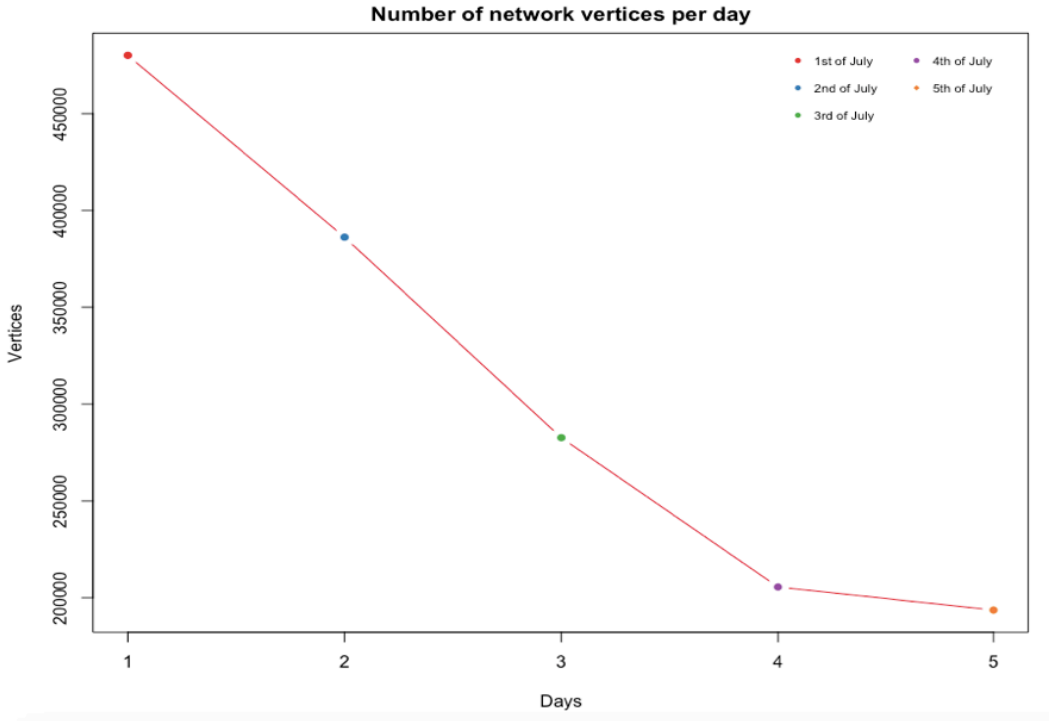
```
IGRAPH 32b4660 DNW- 205575 240552 --
+ attr: name (v/c), weight (e/n), lty (e/c)
+ edges from 32b4660 (vertex names):
 [1] dcompanyau     ->JezebellXOXO   gazatech       ->theebayk1d     gazatech       ->firebucket
 [4] disc_links     ->ayharano       hotboxmovies   ->JezebellXOXO   beverhausen    ->ElNacionalWeb
 [7] luizgustavoleme->mariasisa      cwhite24       ->youngnsaucy    tamtamdoll     ->blurtit
[10] tamtamdoll     ->tamtamdoll     astroranger    ->FunnyJoker     cbsandrewcohen ->latimes
[13] janvrsinsky    ->digitalps      jimmylep       ->CRAPIC_Design  jimmylep       ->smashingmag
[16] lynnemarian    ->paddymusttweet missjaclynrose ->teamlegion     penthouseink   ->JezebellXOXO
[19] dianabenedikt  ->bayareabites   disc_links     ->gabriegeraldo  karlieluv22    ->adamlambert
[22] steph996       ->jobroslover123 xaireel        ->delamarRX931   berryluv       ->addthis
+ ... omitted several edges
```

```
IGRAPH a6271ea DNW- 193674 222771 --
+ attr: name (v/c), weight (e/n), lty (e/c)
+ edges from a6271ea (vertex names):
 [1] emanman        ->FreddyAdu11    heartcures     ->wordoffaith    intuited       ->successsecret
 [4] jimbando       ->whozits        twiprodigy007  ->mjconnor66     twiprodigy007  ->LazyTourist
 [7] twiprodigy007  ->Greystoke1     twiprodigy007  ->Purkulater     de_merde       ->GeekspaceFR
[10] dudebrochill   ->ChiefHava      dudebrochill   ->alnpd          emilywelliver  ->ploked
[13] rimaj_muaah    ->lilduval       rimaj_muaah    ->drydickdonnie  wootboot       ->missjaclynrose
[16] lyndons        ->nambor         dudebrochill   ->maljackson     dudebrochill   ->joeymagnumryan
[19] mikecos23      ->ShoreTweet     mikecos23      ->BoughtInAP     wootboot       ->moontyger
[22] dudebrochill   ->BROADstBANKS   wootboot       ->kkinnee        clione         ->mituhime
+ ... omitted several edges
```

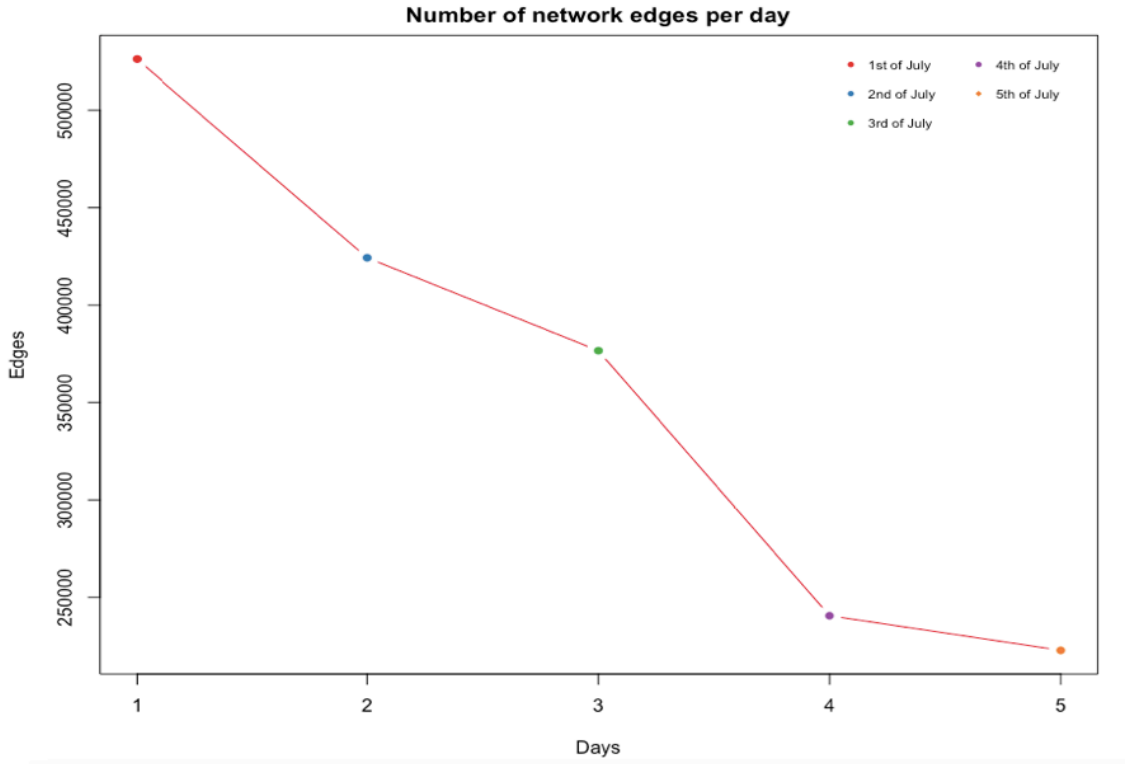
PART 2: AVERAGE DEGREE OVER TIME

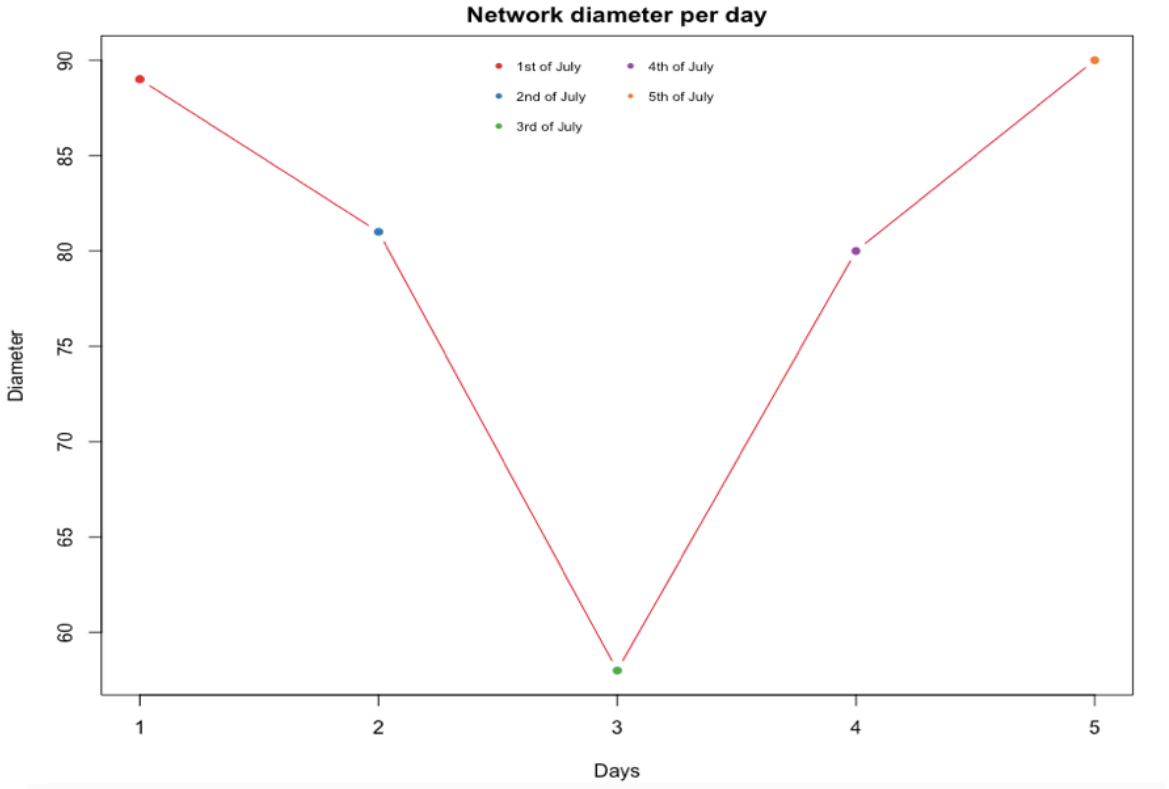
In this part, we are willing to examine how the network metrics presented below were changed over the first 5 days of July.

To begin with, it seems that the network of the first day is the one with the maximum number of vertices and edges.

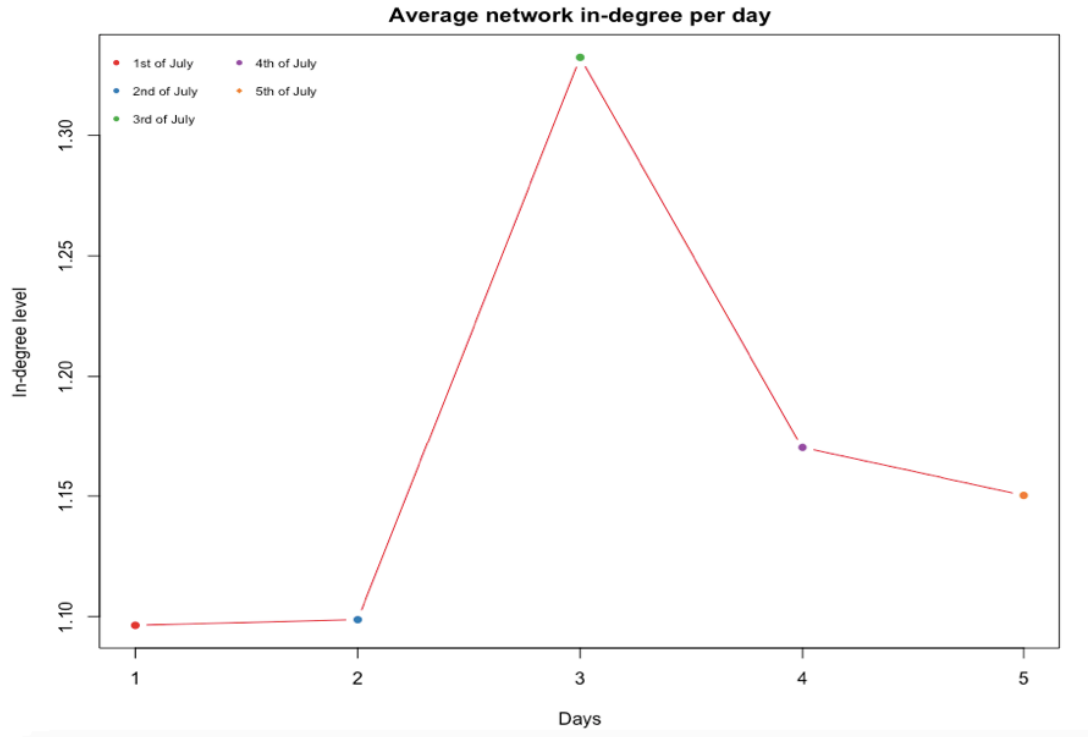


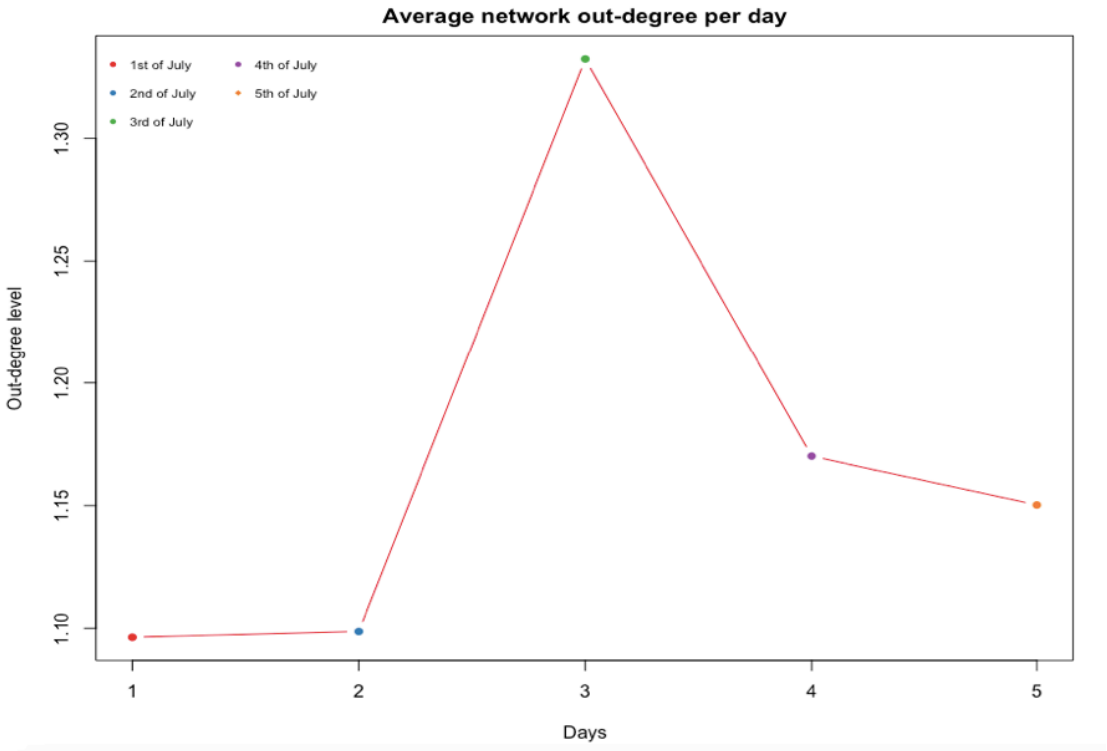
It is also noticeable that as time goes by, the number of vertices and edges is decreased indicating that the network of the fifth day includes in fact the least number of Twitter users and relations among them. However, this network is the one with the greatest distance between any pair of Twitter users.





On top of that, another conclusion is that the network created the first day had, in average, the least number of edges that point toward any vertex of interest as well as the least number of edges that any vertex of interest points toward. On the contrary, the network of the third day scores the highest as far as these two metrics are concerned.





PART 3: IMPORTANT NODES

Regarding the in-degree metric, it seems that only 20% of the users (i.e. ***tweetmeme*** and ***mashable***) are the same in the list of the top 10 Twitter users for the 5 consecutive days that are examined. These users have different ranking positions per day, though. Moreover, except ***addthis*** and ***BreakingNews*** users who seem to appear in 4 out of 5 days, the rest generally vary from one day to another.

```
> print(top_10_inDegr1)
    tweetmeme      mashable       addthis   smashingmag    mileycyrus  BreakingNews           cnn   GuyKawasaki       aplusk rafinhabastos
         2522          1627          1212           965           778           763           746           679           669           629
> print(top_10_inDegr2)
      tweetmeme       ddlovato       mashable        cnnbrk           cnn       addthis souljaboytellem    OfficialTila    officialtila
           2478           2242           1996           1300           1219           1118            898            748            738
     mileycyrus
            680
> print(top_10_inDegr3)
      tweetmeme souljaboytellem        addthis       mashable  BreakingNews         cnnbrk       moontweet        lilduval         PhillyD
           1826           1379           1002            940            874            856            720            428            365
    adamlambert
            362
> print(top_10_inDegr4)
   BreakingNews        addthis      tweetmeme       iamdiddy     mileycyrus         cnnbrk       mashable       lilduval souljaboytellem
            949            816            762            543            535            516            456            454            443
       TheOnion
            350
> print(top_10_inDegr5)
  davidmmasters       iamdiddy        addthis      tweetmeme       mashable  BreakingNews       moontweet     mileycyrus     rainnwilson
           1914           1147            861            746            550            490            360            353            339
AKGovSarahPalin
            332
```

Regarding the outdegree metric, we have almost the same proportion as we had before. More specifically, only the users ***wootboot*** and ***dudebrochill*** are included in the top 10 Twitter users list for all days whereas the users ***the_sims_3*** and ***failbus*** appear in this list in a percentage of 60% (3/5 days). Again, the rest of the users generally vary from one day to another.

```
> print(top_10_outDegr1)
   dudebrochill         failbus   tsliquidators      the_sims_3        wootboot vaguetweetstest         lmaobot        drharvey       luvorhate
            245             215             215             202             200             193             165             142             119
     help_echo
            106
> print(top_10_outDegr2)
 dudebrochill       wootboot        failbus     the_sims_3         dvdbot    takeyourpin     teamqivana     luvorhate modelsupplies    rt_thursday
          279            240            185            166            158            147            143            127            125            119
> print(top_10_outDegr3)
  drejones71    deana1981 killah360dhh       imbeeyo      java4two     ohmichael        nachhi dudebrochill      wootboot     medic_ray
         624          605          438          431          383          347          340          305          277          271
> print(top_10_outDegr4)
        swbot  dudebrochill       wootboot   fxxxyourlife    andreapuddu    azandiamjbb    hoboprophet        failbus     herpescure twiprodigy009
          830           391            353            257            246            244            240            239            216            202
> print(top_10_outDegr5)
        swbot twiprodigy008 twiprodigy005 twiprodigy007 twiprodigy009       wildingp   dudebrochill       wootboot    hoboprophet     the_sims_3
          876            808            672            644            588            339            331            319            255            225
```

	page.rank
tweetmeme	0.0017889774
mashable	0.0012591438
addthis	0.0011849832
smashingmag	0.0011813695
cnn	0.0007182473
mileycyrus	0.0007096070
KISSmetrics	0.0006783605
CourageCampaign	0.0006260832
aplusk	0.0005397417
rafinhabastos	0.0005195846

1ST DAY

	page.rank
ddlovato	0.0028156243
drew_taubenfeld	0.0023948782
mashable	0.0021490222
tweetmeme	0.0021307449
globalmanners	0.0018296943
cnn	0.0015276583
addthis	0.0013608868
souljaboytellem	0.0012128067
cnnbrk	0.0011659017
mileycyrus	0.0007575637

2ND DAY

	page.rank
tweetmeme	0.0024601761
souljaboytellem	0.0023079639
killerstartups	0.0020972462
addthis	0.0017641800
moontweet	0.0012360145
cnnbrk	0.0011727753
mashable	0.0011172055
BreakingNews	0.0010180932
PhillyD	0.0007181871
adamlambert	0.0006171114

3RD DAY

	page.rank
souljaboytellem	0.0056384024
addthis	0.0019969251
tweetmeme	0.0016726163
BreakingNews	0.0016722975
lilduval	0.0012235658
mileycyrus	0.0011959503
mashable	0.0011092977
iamdiddy	0.0010881197
cnnbrk	0.0010306820
garyvee	0.0009087917

4TH DAY

	page.rank
davidmmasters	0.0034208343
iamdiddy	0.0029246273
addthis	0.0022322762
aplusk	0.0021665916
tweetmeme	0.0016896555
mashable	0.0010695413
mrskutcher	0.0009199140
moontweet	0.0008531593
BreakingNews	0.0007359969
mileycyrus	0.0007279574

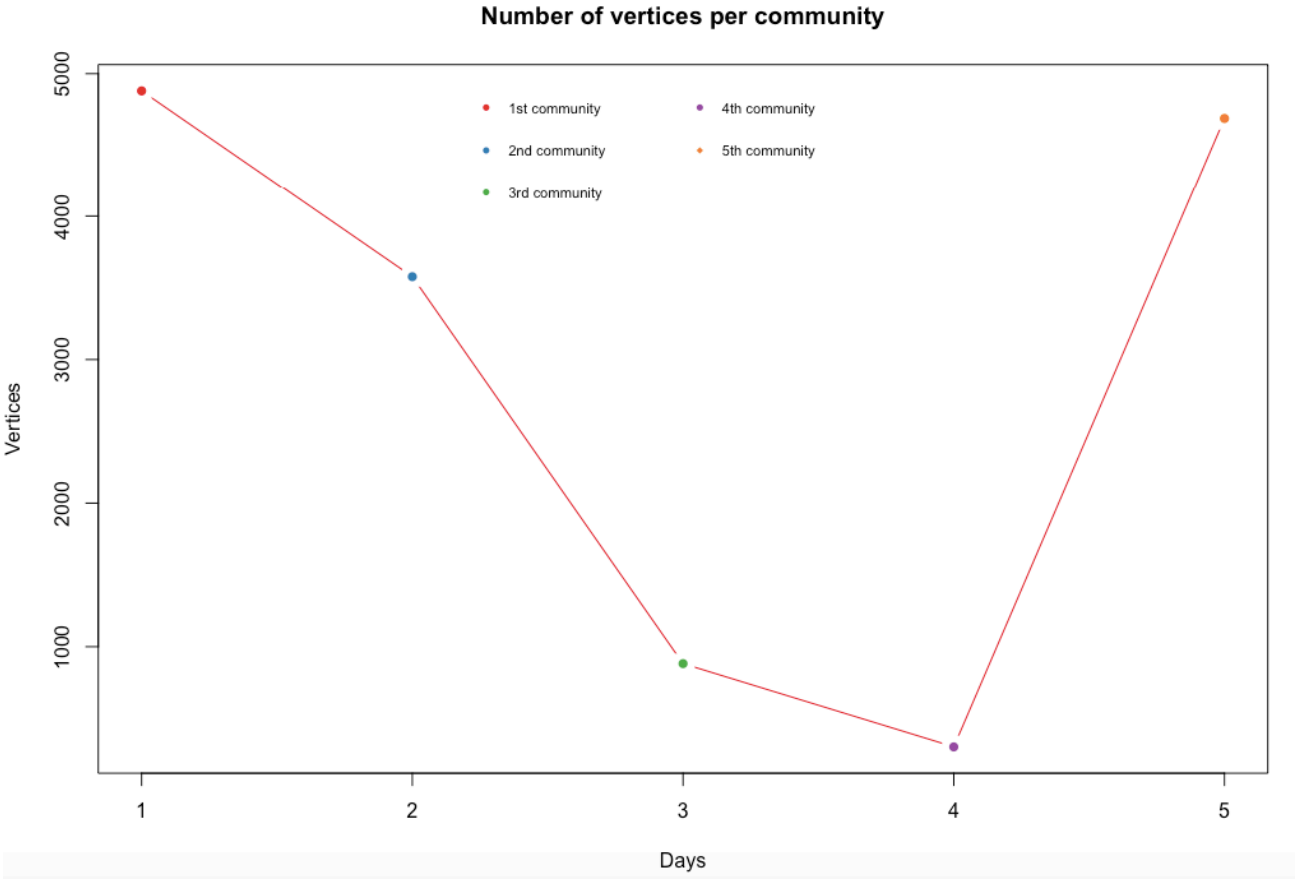
5TH DAY

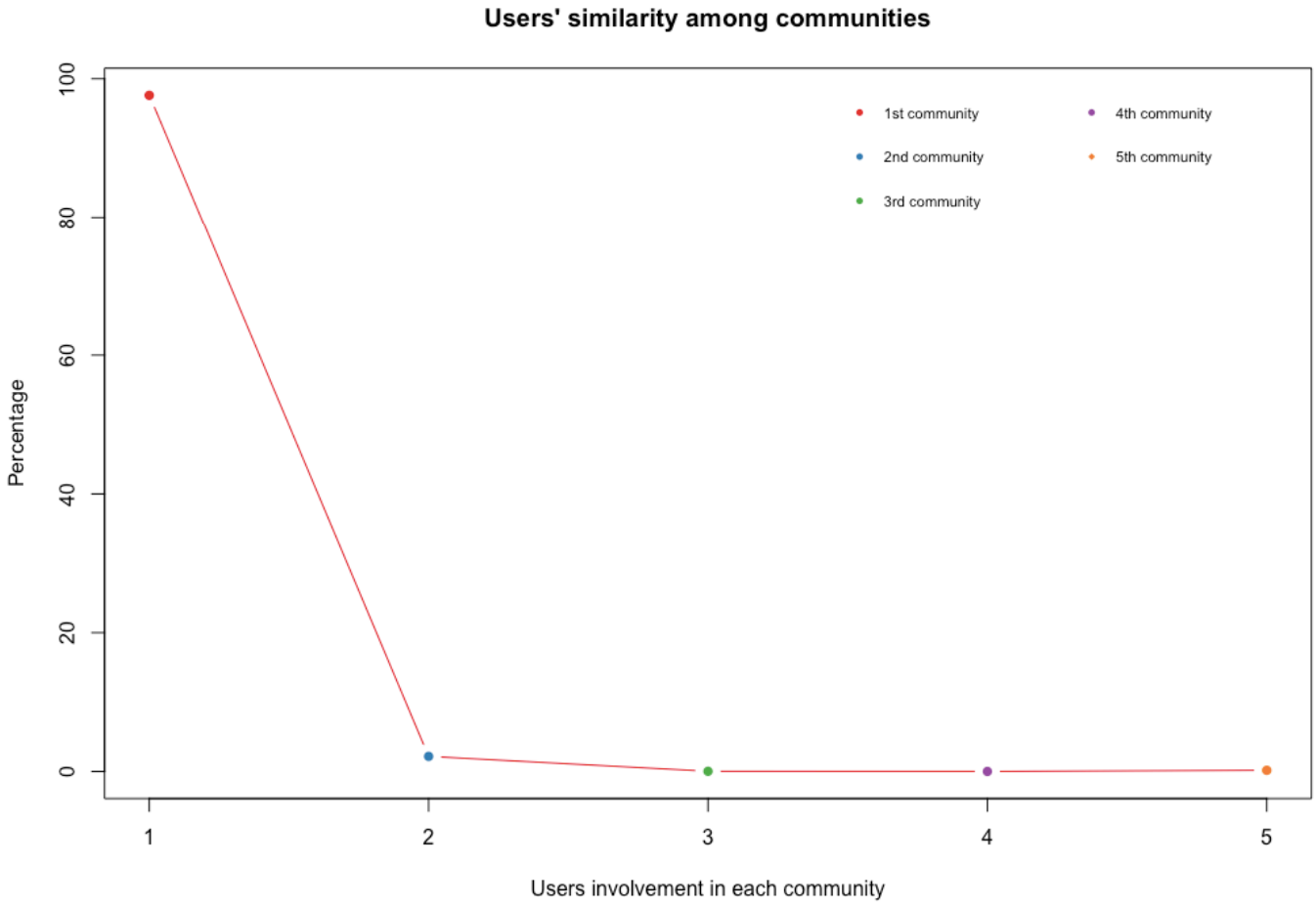
As far as PageRank is concerned, we understand from the above that the users ***tweetmeme***, ***addthis*** and ***mashable*** appear steadily in the top 10 Twitter- users list but in different positions each time. The remaining positions of this list are however randomly occupied by other users.

PART 4: COMMUNITIES

After trying to apply fast greedy clustering, infomap clustering and louvain clustering on the undirected versions of the 5 graphs, it was proven that the first two algorithms are highly inefficient since the computing time needed for the creation of the graph communities was extremely high and in some cases never completed. On the contrary, louvain clustering calculated the communities rapidly. Therefore, this is the algorithm selected in order to investigate the questions of this last part.

The Twitter user with name "***egomonics***" selected randomly in order to examine how the five communities in which this user participated were evolved during this 5-day period. The first criterion examined to assess this question is the difference in the number of edges in these communities while the second one is whether the same Twitter users comprised the communities the *user* "egomonics" belongs to. Therefore, let us now have a look at the findings presented in detail below.



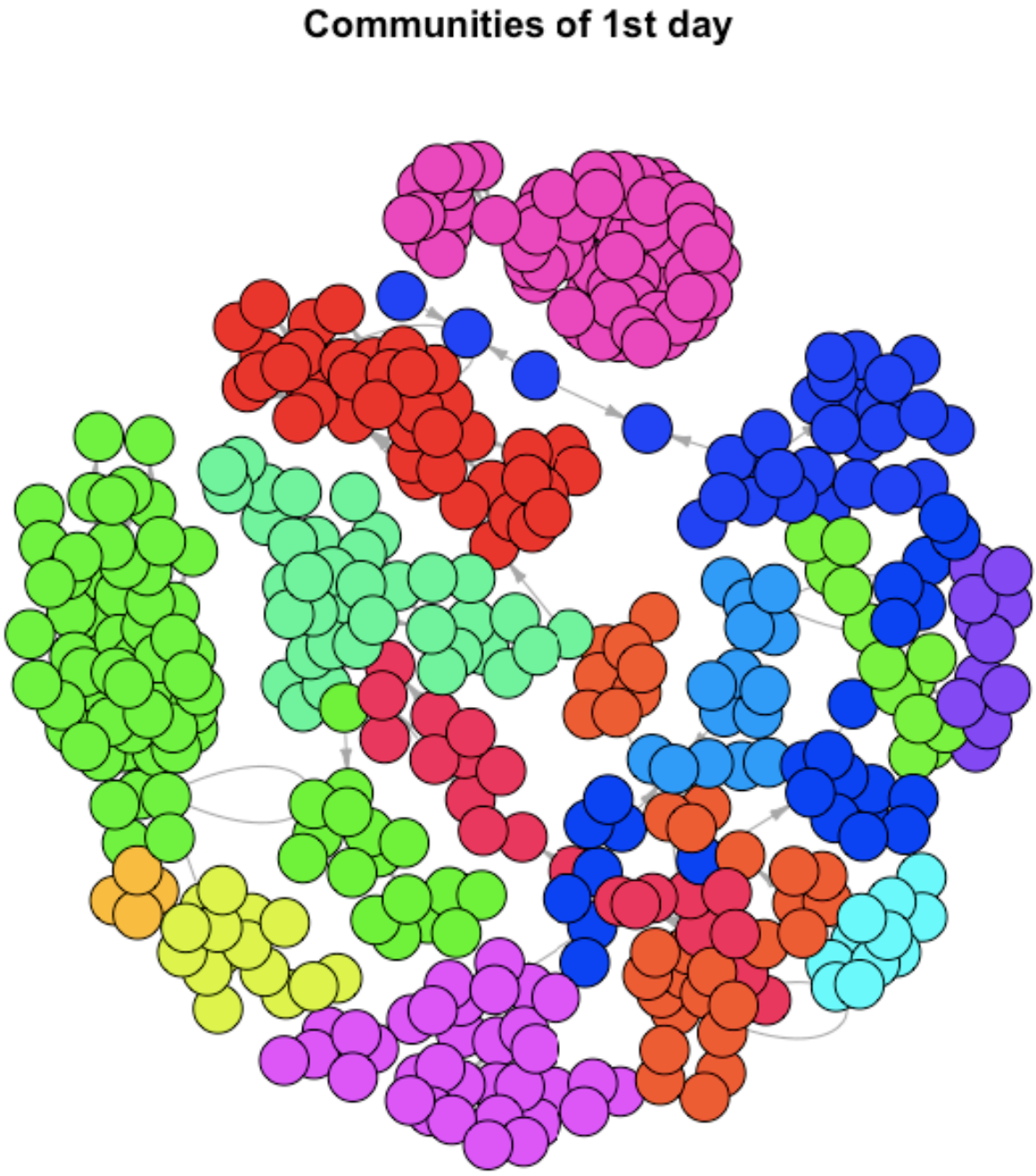


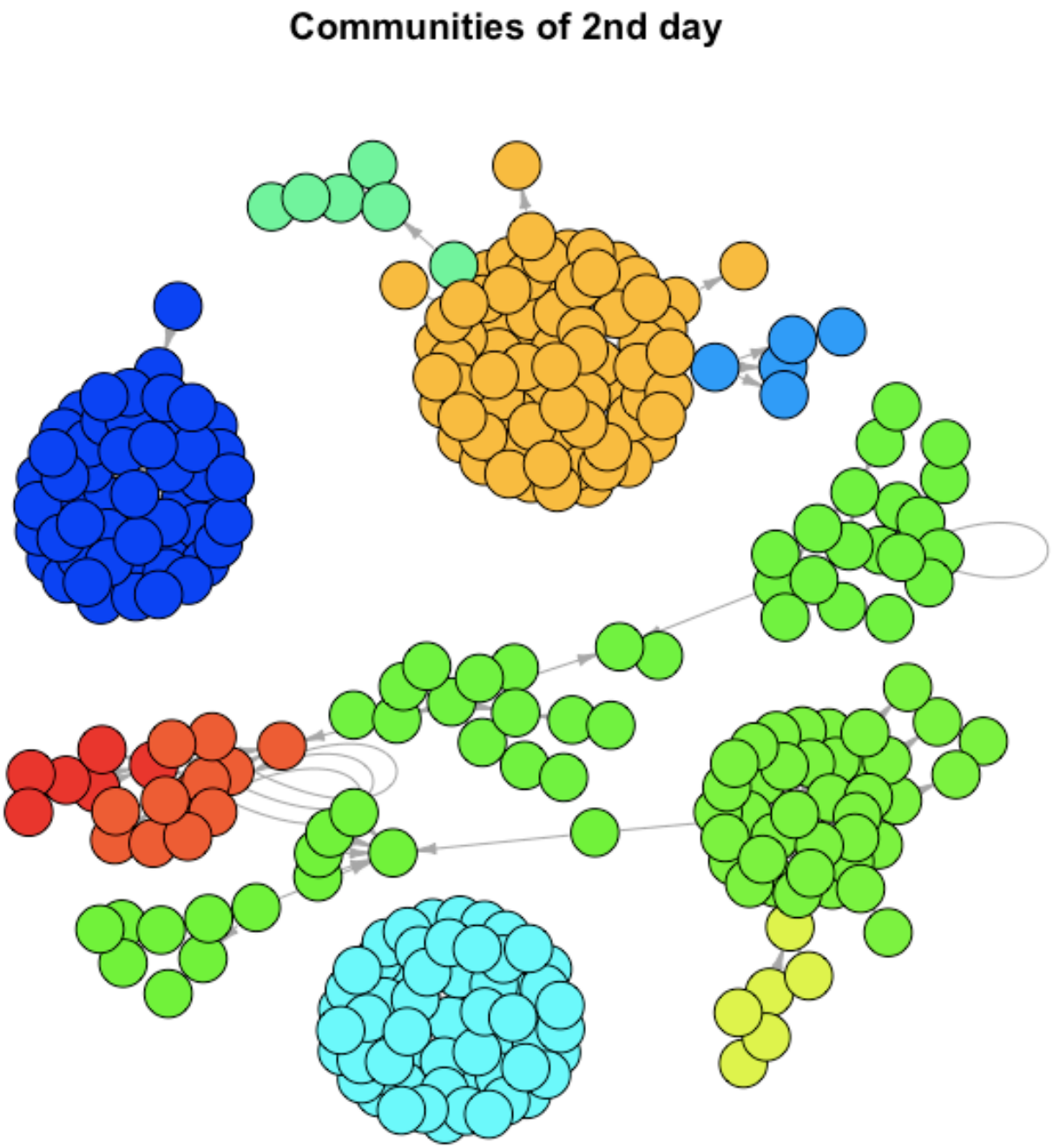
From the above plots, it is easily derived that the structure of the different communities of the user "egomonics" for these 5 days is not fixed in terms of number or identity of Twitter users.

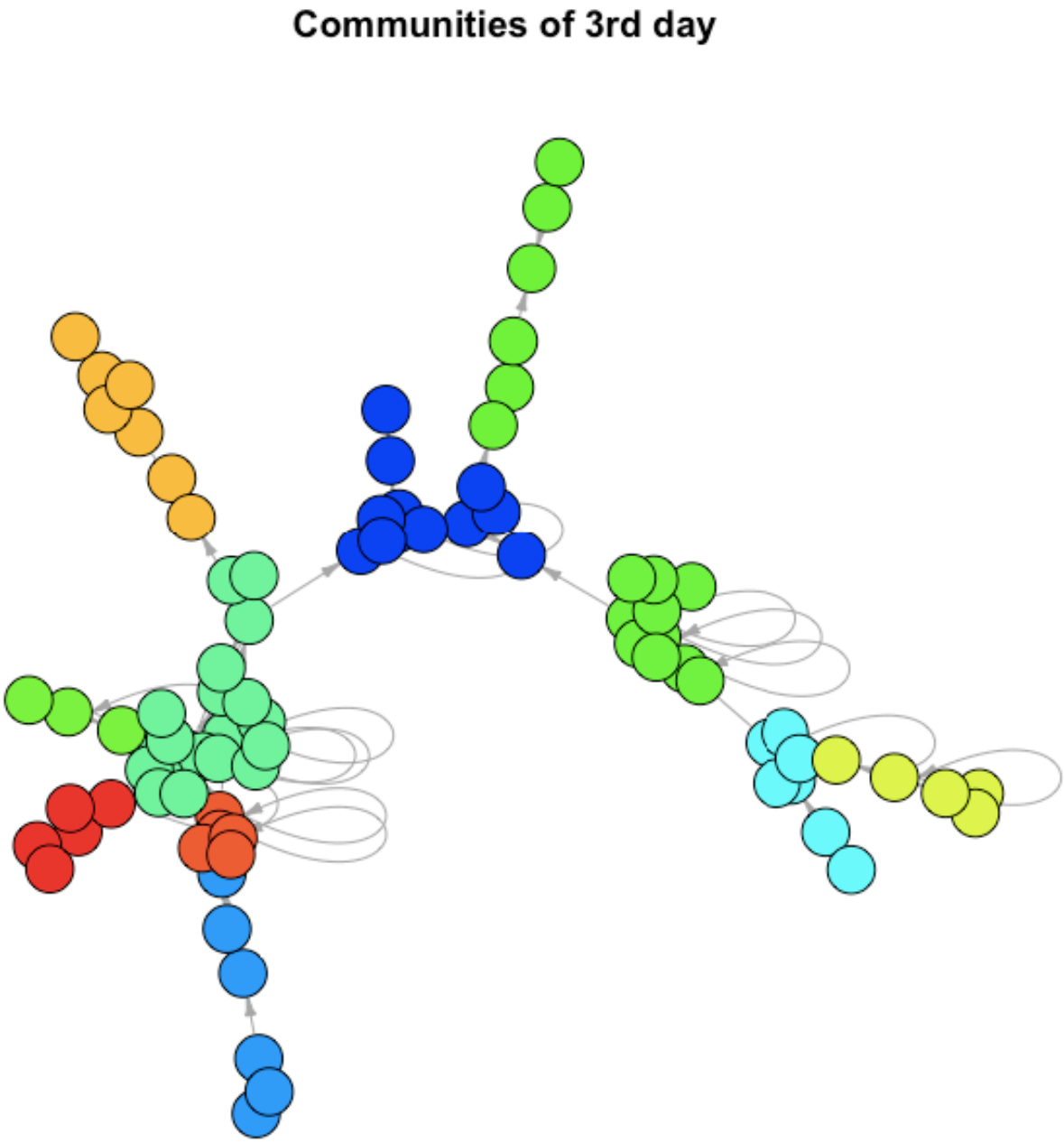
More specifically, the existence of Twitter users in the communities presents a downward trend day by day until the fifth day is reached when the number reaches its peak indicating high level of variability for these communities.

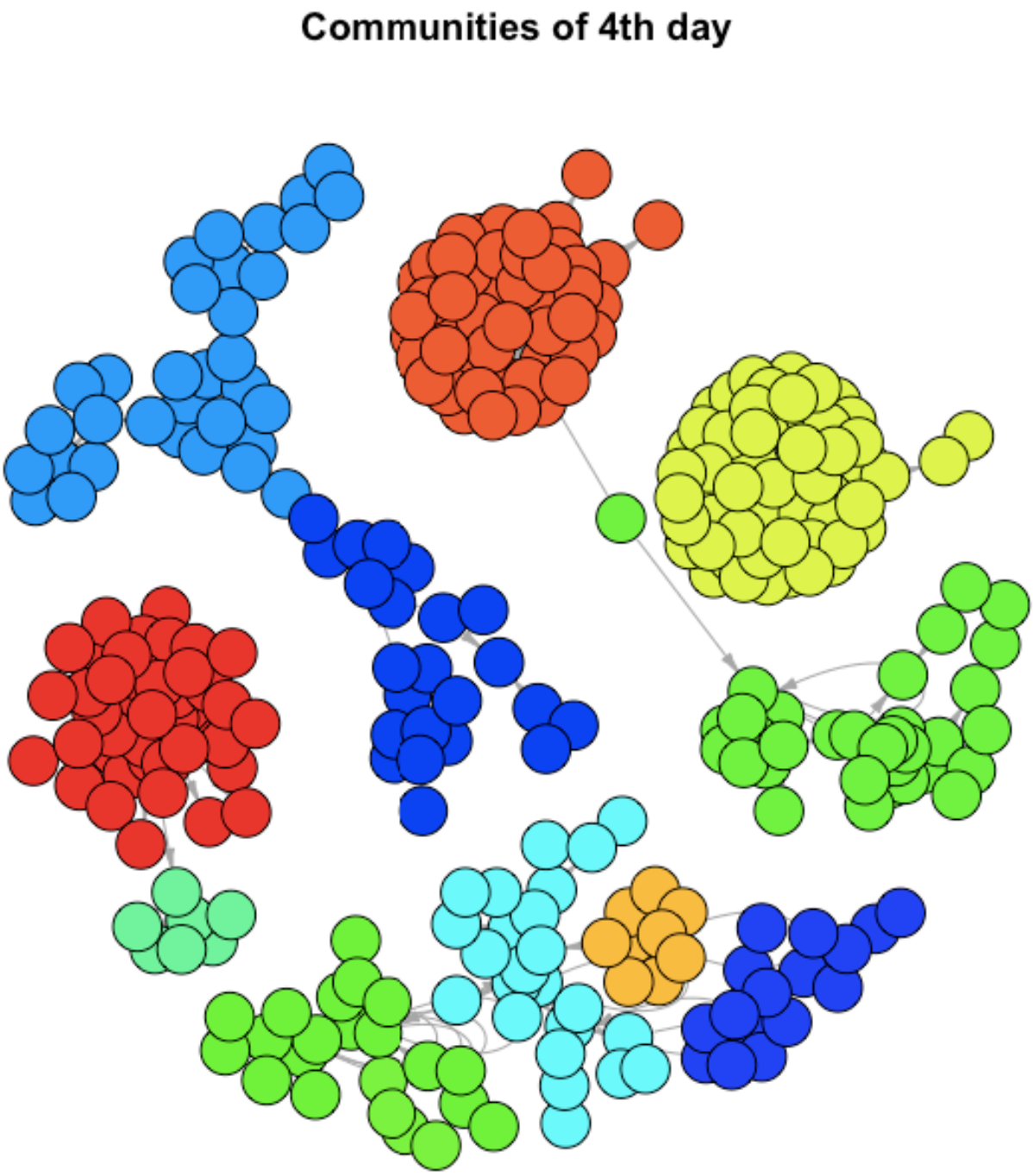
On top of that, it is proven that not alike users are included in the community of each day since much more than the majority of them (counted at **97,6%**) can be found only in 1 out of 5 communities in which the user "egomonics" is involved. The Twitter users that subsequently follow who are in 2 out of 5 communities amass a percentage of only **2,1%** implying that the formulation of the communities is not clearly the same in this 5-day period.

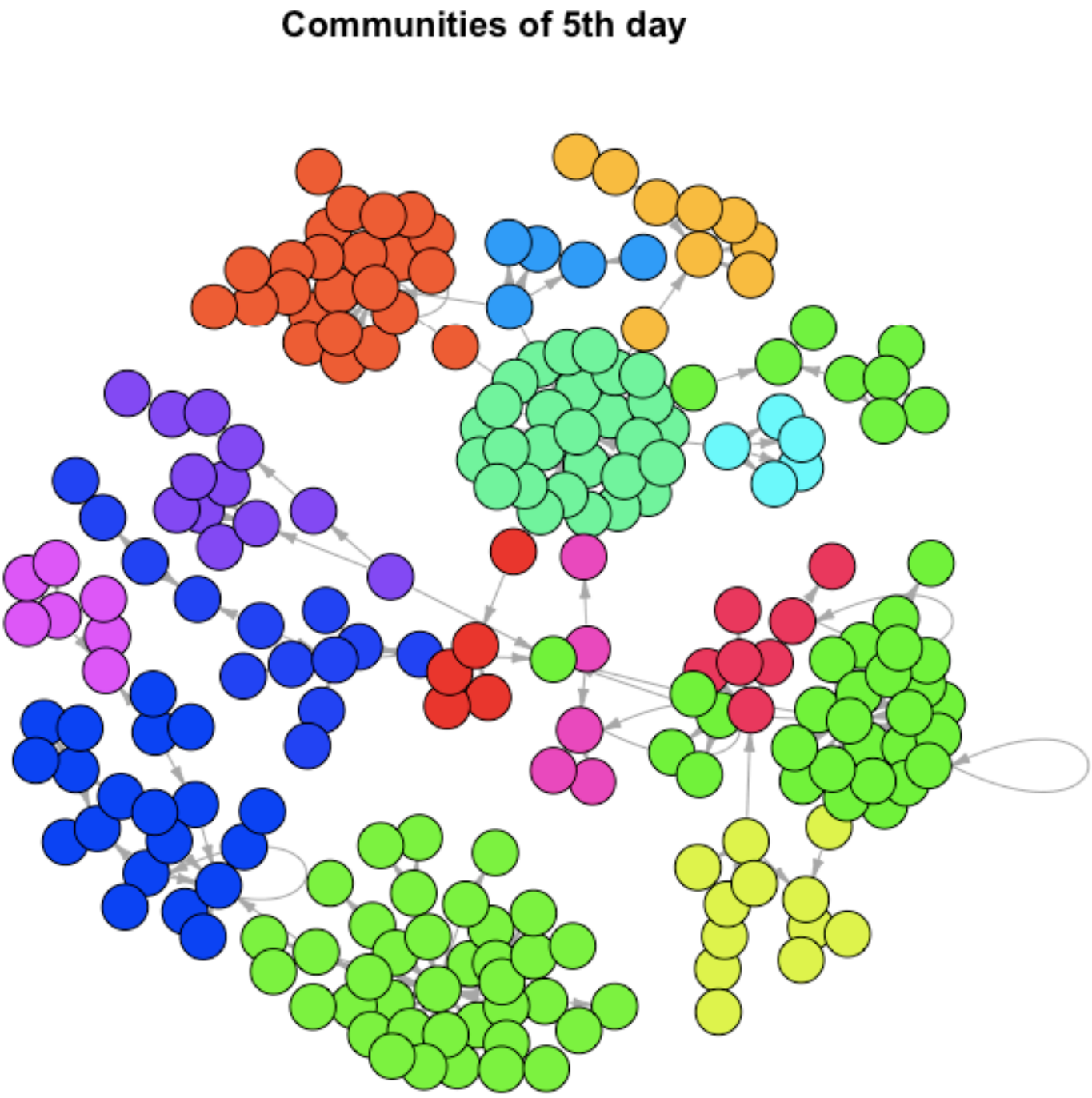
Finally, the requested graphs of each day depicting the different communities can be found below.


Communities of 1st day


Communities of 2nd day


Communities of 3rd day


Communities of 4th day


Communities of 5th day