‘Twitter Analysis’

28/6/2019

Assignment 2

Social Network Analysis

Athens University of Economics and Business

Brikena Kokalari

Contents

1. Introduction

The aim of this assignment is to create a weighted directed graph with igraph,1 using raw data from Twitter. We will use as a dataset a compressed file with Tweets posted during July 2009. (https://drive.google.com/open?id=1RjWUg-6KrVOjJPZHHQg-h_9gSSWZUPn-) and we will use only the first five days of July. We tried to extract 5 csv file using Python, one for each day, but the created files were not successful for further analysis. However, the code and the created file are included. For this assignment, the provided data are used in order to continue with the analysis.

Regarding the social network we are going to create plots that visualize the 5-day evolution of different metrics for the graph. In addition, we will write to code to create and print data frames for the 5-day evolution of the top-10 Twitter users regarding their in-degree, out-degree and PageRank.
Finally, we will try to perform community detection on the mention graphs, using three different algorithms, greedy clustering, infomap clustering and louvian clustering. Taking into account one random Twitter user, we will use one of the aforementioned algorithms to spot the evolution of the communities this user belongs to and will visualize each community.

2. Q1 & Q2

Firstly, we imported the dataset in R and created 5 dataframes, one for each day and created the directed igraph .
Then we calculated the metrics:

- Number of vertices

01/07/2009	02/07/2009	03/07/2009	04/07/2009	05/07/2009
477324	383820	280490	203939	192159

- Number of edges

01/07/2009	02/07/2009	03/07/2009	04/07/2009	05/07/2009
522477	420929	372787	238159	220652

- Diameter of the graph

01/07/2009	02/07/2009	03/07/2009	04/07/2009	05/07/2009
89	87	58	80	89

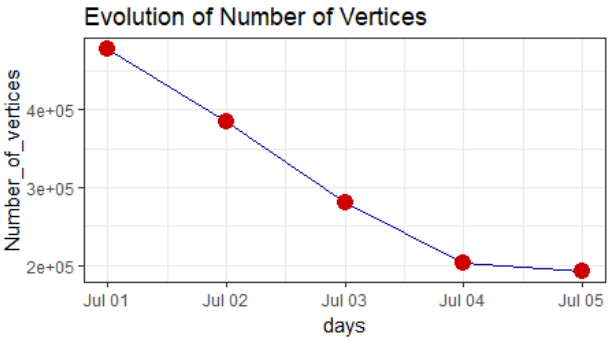
- Average in-degree

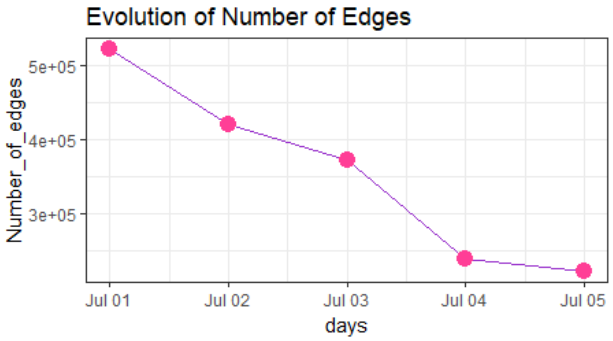
01/07/2009	02/07/2009	03/07/2009	04/07/2009	05/07/2009
1.09	1.10	1.33	1.17	1.15

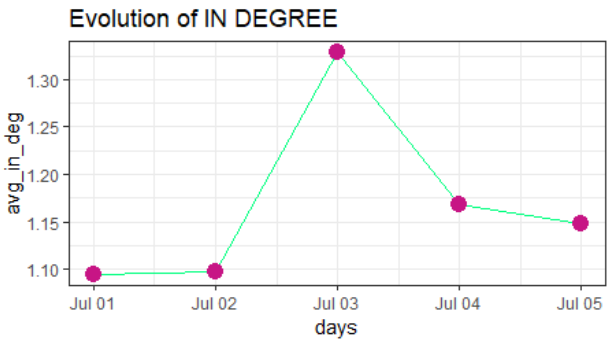
- Average out-degree

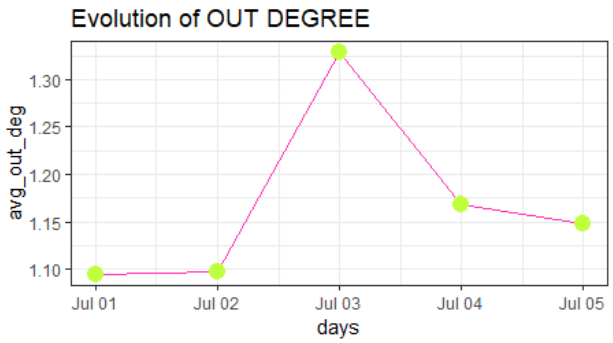
01/07/2009	02/07/2009	03/07/2009	04/07/2009	05/07/2009
1.09	1.10	1.33	1.17	1.15

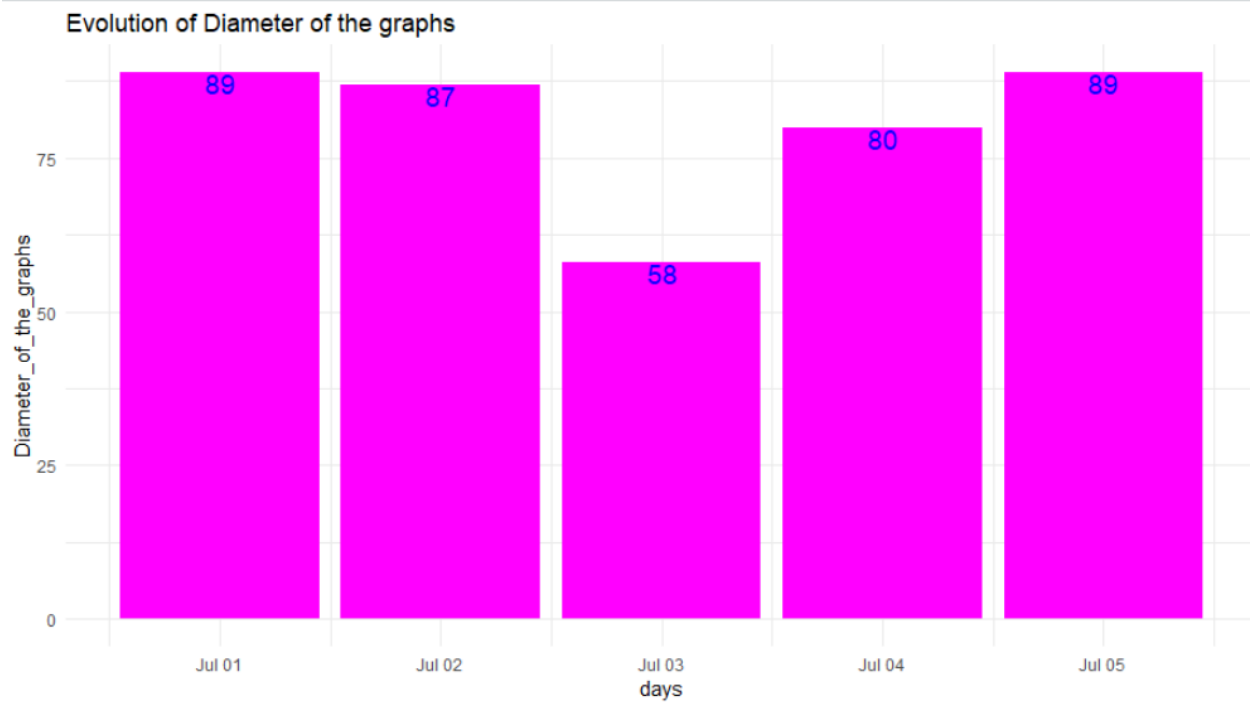
From the above metrics we notice that there is a gradual decrease in the number of vertices and edges during the days. The diameter, the longest shortest path connecting two vertices in a graph rose and fell during the days. The average in degree and average out degree remained the same which means that on average each user twitted as much as the “mentions” he/she received from other users.


Evolution of Number of Vertices
Number_of_vertices
4e+05
3e+05
2e+05
Jul 01
Jul 02
Jul 03
Jul 04
Jul 05
days


Evolution of Number of Edges
Number_of_edges
5e+05
4e+05
3e+05
Jul 01
Jul 02
Jul 03
Jul 04
Jul 05
days


Evolution of IN DEGREE
avg_in_deg
1.30
1.25
1.20
1.15
1.10
Jul 01
Jul 02
Jul 03
Jul 04
Jul 05
days


Evolution of OUT DEGREE
avg_out_deg
1.30
1.25
1.20
1.15
1.10
Jul 01
Jul 02
Jul 03
Jul 04
Jul 05
days


Evolution of Diameter of the graphs
Diameter_of_the_graphs
75
50
25
0
89
87
58
80
89
Jul 01
Jul 02
Jul 03
Jul 04
Jul 05
days

3. Q3

In this step we are going to create data frames for the 5-day evolution of the top-10 Twitter users with regard to:

- In-degree
- Out-degree
- PageRank

In-degree

- In degree on 01/07/2009

tweetmeme	mashable	addthis	smashingmag	mileycyrus
2522	1627	1213	965	776
BreakingNews	cnn	GuyKawasaki	aplusk	rafinhabastos
762	744	676	666	629

- In degree on 02/07/2009

tweetmeme	ddlovato	mashable	cnnbrk	cnn
2478	2242	1992	1297	1218
addthis	souljaboytellem	OfficialTila	officialtila	mileycyrus
1121	891	748	736	680

- In degree on 03/07/2009

tweetmeme	souljaboytellem	addthis	mashable	BreakingNews
1825	1377	1002	939	874
cnnbrk	moontweet	lilduval	PhillyD	adamlambert
856	719	428	365	362

- In degree on 04/07/2009

BreakingNews	addthis	tweetmeme	iamdiddy	mileycyrus
949	816	762	543	535
cnnbrk	mashable	lilduval	souljaboytellem	TheOnion
515	456	454	439	349

- In degree on 05/07/2009

davidmmasters	iamdiddy	addthis	tweetmeme	mashable
1914	1145	859	746	550
BreakingNews	moontweet	mileycyrus	rainnwilson	AKGovSarahPalin
490	360	352	338	331

Out-degree

- Out degree on 01/07/2009

```
   dudebrochill   tsliquidators         failbus      the_sims_3        wootboot
            245             215             215             202             200
vaguetweetstest         lmaobot        drharvey       luvorhate       help_echo
            193             165             142             119             106
```

- Out degree on 02/07/2009

```
dudebrochill      wootboot       failbus    the_sims_3        dvdbot   takeyourpin
         278           240           185           166           158           146
  teamqivana     luvorhate modelsupplies   rt_thursday
         143           127           125           118
```

- Out degree on 03/07/2009

```
drejones71    deana1981 killah360dhh      imbeeyo     java4two    ohmichael
       613           600           436           429           376           347
    nachhi dudebrochill     wootboot    medic_ray
       335           305           277           260
```

- Out degree on 04/07/2009

```
    swbot  dudebrochill      wootboot  fxxxyourlife   andreapuddu   azandiamjbb
      826           390           353           256           246           241
  failbus    herpescure   hoboprophet twiprodigy009
      237           216           216           202
```

- Out degree on 05/07/2009

```
       swbot twiprodigy008 twiprodigy005 twiprodigy007 twiprodigy009      wildingp
         876           808           672           640           588           335
dudebrochill      wootboot   hoboprophet    the_sims_3
         330           319           236           225
```

PageRank

- PageRank on 01/07/2009

	page.rank day1
tweetmeme	0.0018003979
mashable	0.0012672888
addthis	0.0011974775
smashingmag	0.0011883703
cnn	0.0007209619
mileycyrus	0.0007148954
KISSmetrics	0.0006823756
CourageCampaign	0.0006297084
aplusk	0.0005421232
rafinhabastos	0.0005239454

- PageRank on 02/07/2009

	page.rank day2
ddlovato	0.0028350622
drew_taubenfeld	0.0024114096
mashable	0.0021622758
tweetmeme	0.0021452328
globalmanners	0.0018409773
cnn	0.0015359171
addthis	0.0013739174
souljaboytellem	0.0012102347
cnnbrk	0.0011719464
mileycyrus	0.0007634394

- PageRank on 03/07/2009

	page.rank day3
tweetmeme	0.0024798174
souljaboytellem	0.0023219229
killerstartups	0.0021139838
addthis	0.0017776163
moontweet	0.0012445916
cnnbrk	0.0011826916
mashable	0.0011245064
BreakingNews	0.0010254502
PhillyD	0.0007245655
adamlambert	0.0006215651

- PageRank on 04/07/2009

	page.rank day4
souljaboytellem	0.0056501774
addthis	0.0020150001
BreakingNews	0.0016927288
tweetmeme	0.0016836489
lilduval	0.0012287297
mileycyrus	0.0012095750
mashable	0.0011181157
iamdiddy	0.0010981366
cnnbrk	0.0010346190
garyvee	0.0009158967

- PageRank on 05/07/2009

	page.rank day5
davidmmasters	0.0034511101
iamdiddy	0.0029484343
addthis	0.0022484888
aplusk	0.0021837304
tweetmeme	0.0017031700
mashable	0.0010784785
mrskutcher	0.0009274927
moontweet	0.0008598342
BreakingNews	0.0007432564
mileycyrus	0.0007321081

Again, we have fluctuations between the users over the days. There are users that appear only once in the top 10 during these days and users that appear more than once. Moreover, we can observe that the users that appear in the top 10 regarding their in degree centrality on a specific day, appear as well in the top 10 of PageRank of that day.

4.Q3

In out last step, we have to detect communities. Trying three different alorithms, fast greedy clustering, infomap clustering, and louvain clustering on the undirected versions of the 5 graphs, we could not obtain results from all. Interestingly, Louvain performed better than Infomap and fast greedy clustering in all the graphs in our study. It maximizes a modularity score for each community, where the modularity quantifies the quality of an assignment of nodes to communities by evaluating how much more densely connected the nodes within a community are, compared to how connected they would be in a random network. The Louvain algorithm is one of the fastest modularity-based algorithms and works well with large graphs.

Picking Louvain Clustering, we chose the user “Mashable” to detect the evolution in communities. The membership function gives a vector of a community ids for every node in the graph. So ,in our case Mashable belongs to community 45904 on 01/07/2009, the next day he belongs to community 18673 , and to 11345, 19096 and 21695 respectively the three other days. The community Mashable belongs to has changed during the days.

In the below graphs, we can se how the communities are formed from 01/07/2009-05/07/2009.