CS-484 Project Proposal

Perontsi Eva, AM 1142

November 29, 2020

For this project, I decided to work on the Greek twitter graph and do a visualization of conversations over time. I decided to work on October 2020, when the twitter community was really active about a lot of different subjects, such as the upcoming US elections, covid-19, the conviction of Golden Dawn etc. More specifically, I have collected the majority ($> 80\%$) of mentions, retweets, replies and quotes that took place in the Greek community in twitter, in a daily basis, from the 1$^{\text{st}}$ of October 2020 until the end of the month.

1 The dataset

I got the data from the crawler that collects twitter data for the CARV lab at ICS-FORTH. For each day, I have four different .txt files that contain user IDs and a weight, for example in a "mention" file, that contains the majority of "mention" interactions in the Greek twitter, each line contains the mentioner-uid, the mentioned-uid, and the weight of their connection (how many times the "mention" interaction happened between them in a day).

Below there is an example of a file that will be used to construct the graph.

Figure 1: Example of a "mention" data file


```
74233 15116120 1
336113 18764318 3
646913 20926521 1
659833 255040592 1
659833 659833 7
659833 250058740 2
659833 255471924 2
659833 380648579 1
659833 1130817038723801090 1
```

2   Goals

My goals for this project are:

1. to create a time-lapse graph that will show how the community connected, talked and interacted in each of the four ways mentioned above

2. to measure some metrics about homophily, connection between groups

3. to make conclusions about what keeps the community busy by seeing the trends of these days, and maybe find connections between subjects and communities