Quanteda and Twitter

Rodrigo Esteves de Lima-Lopes
State University of Campinas
rll307@unicamp.br

Contents

1 Introduction

Quanteda is a package for managing and analyse text quantitatively. It is quite easy to use and will bring us a number of interesting functions.

1.1 You will need:

1. The package `Quanteda`, `quanteda.textplots` and `quanteda.textstats` which can be installed using RStudio
2. The package `rtweet`, we installed last tutorial.
3. Package `DT` for viewing the KWIC inside R.

2 Scraping Tweets

We are going to use the same data we have used in the previous tutorials.

3 Doing some analysis

3.1 Creating a network of hashtags for each president

First we will do the magic for Lulla. We are going to:

1. Select the hashtags using the command `dfm_select`
2. Select the 50 more frequent using `topfeatures` command

```
tag.LI <- dfm_select(LI.dfm, pattern = ("#*"))
toptag.LI <- names(topfeatures(ttag.LI, 50))
```

Now let us construct a feature-occurrence matrix for the hashtags

```
tag.fcm.LI <- fcm(tag.LI)
```

Let us make a FCM only with the top hashtags

```
top.plot.LI <- fcm_select(tag.fcm.LI, pattern = toptag.LI)
```

And then we make our network

```
textplot_network(top.plot.LI,
                 min_freq = 0.1,
                 edge_alpha = 0.8,
                 edge_size = 5)
```

Let us see how is the final product

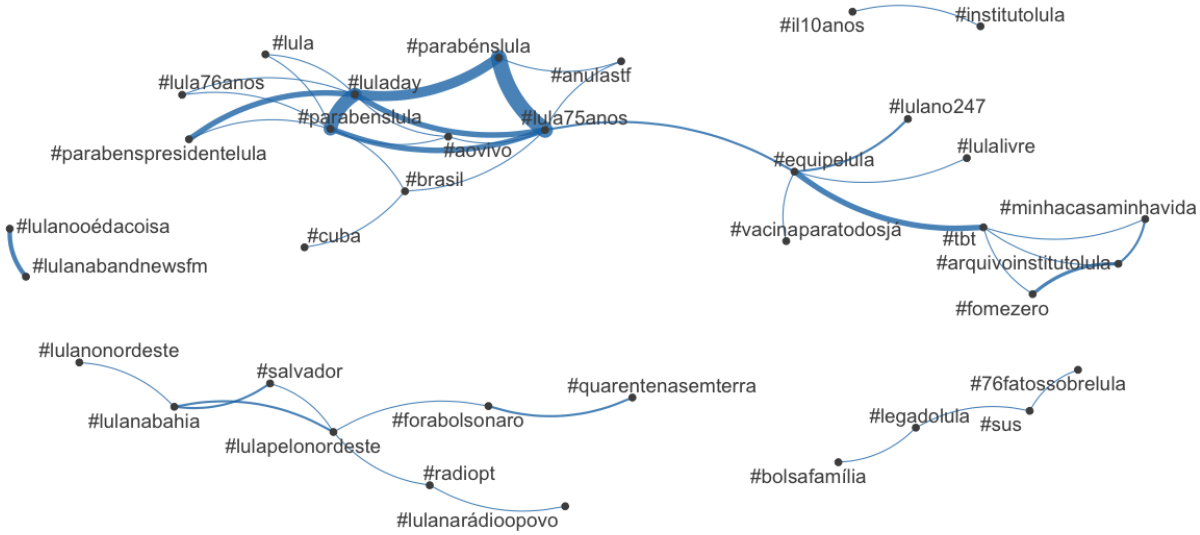


Figure 1: Lula's network of hashtags

Now let us see how it works for JB, all in a single batch of commands:

```
tag.JB <- dfm_select(JB.dfm, pattern = ("#*"))
toptag.JB <- names(topfeatures(tag.JB, 50))
tag.fcm.JB <- fcm(tag.JB)
top.plot.JB <- fcm_select(tag.fcm.JB, pattern = toptag.JB)
textplot_network(top.plot.JB,
                 min_freq = 0.1,
                 edge_alpha = 0.8,
                 edge_size = 5,
                  edge_color = "orange")
```

3.2 Analysing user interactions

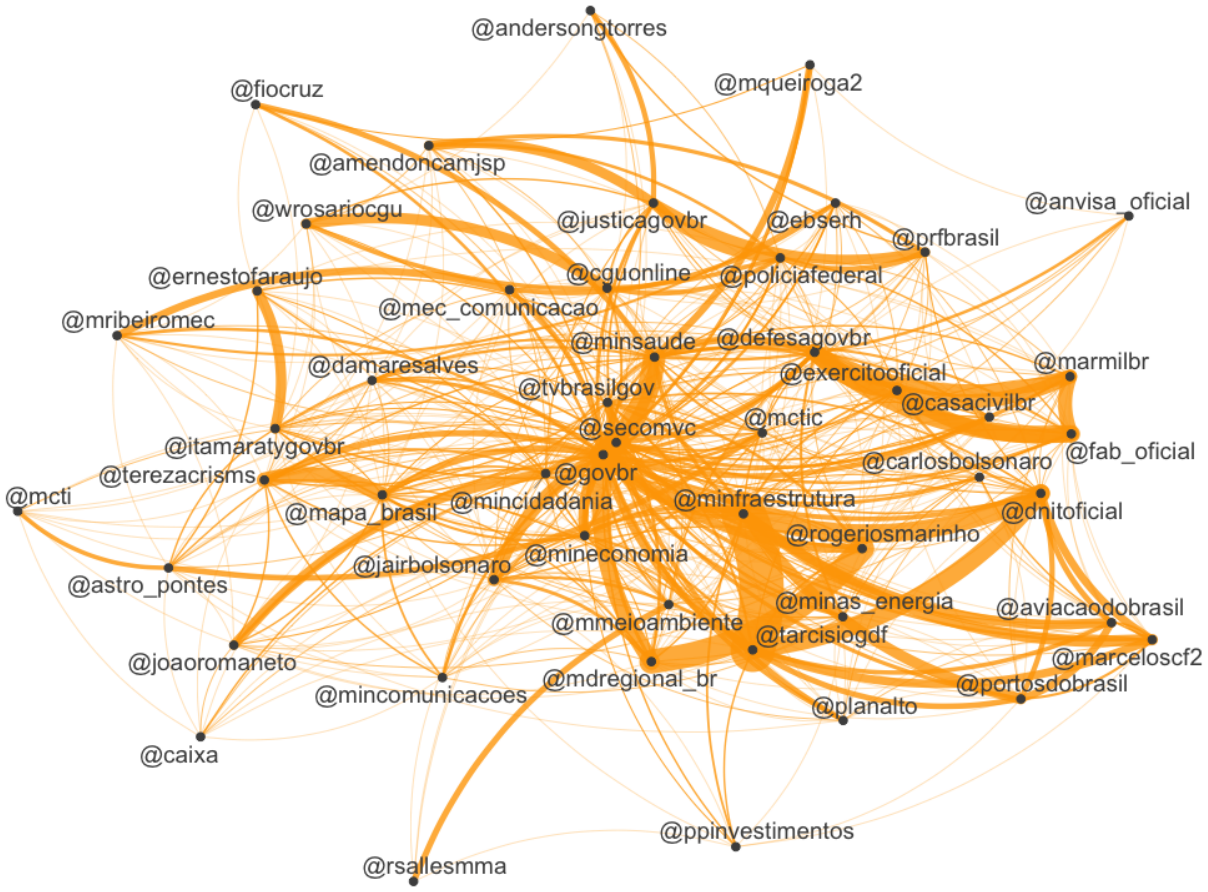
We can use the same methodology to study users interaction. The difference we are going to change the search for `*#` to `*@`. Let us start by Lula, but in a single command:

```
user.LI <- dfm_select(LI.dfm, pattern = ("@*"))
topuser.LI <- names(topfeatures(user.LI, 50))
user.fcm.LI <- fcm(user.LI)
```

Figure 2: Covas' network of hashtags

```
user.plot.LI <- fcm_select(user.fcm.LI, pattern = topuser.LI)
textplot_network(user.plot.LI,
                 min_freq = 0.1,
                 edge_alpha = 0.8,
                 edge_size = 5,
                 edge_color = "red")
```

The result would be something similar to it:



Now let us do the same for JB

```
covas.user.dfm <- dfm_select(covas.dfm, pattern = "@*")
topuser.covas <- names(topfeatures(covas.user.dfm, 50))
head(topuser.covas)

covas.user.fcm <- fcm(covas.user.dfm)
View(covas.user.fcm)

covas.user.plot <- fcm_select(covas.user.fcm,
                              pattern = topuser.covas)
textplot_network(covas.user.plot,
                 min_freq = 0.1,
                 edge_color = "pink3",
                 edge_alpha = 0.8,
                 edge_size = 5)
```

The result should be something similar to:

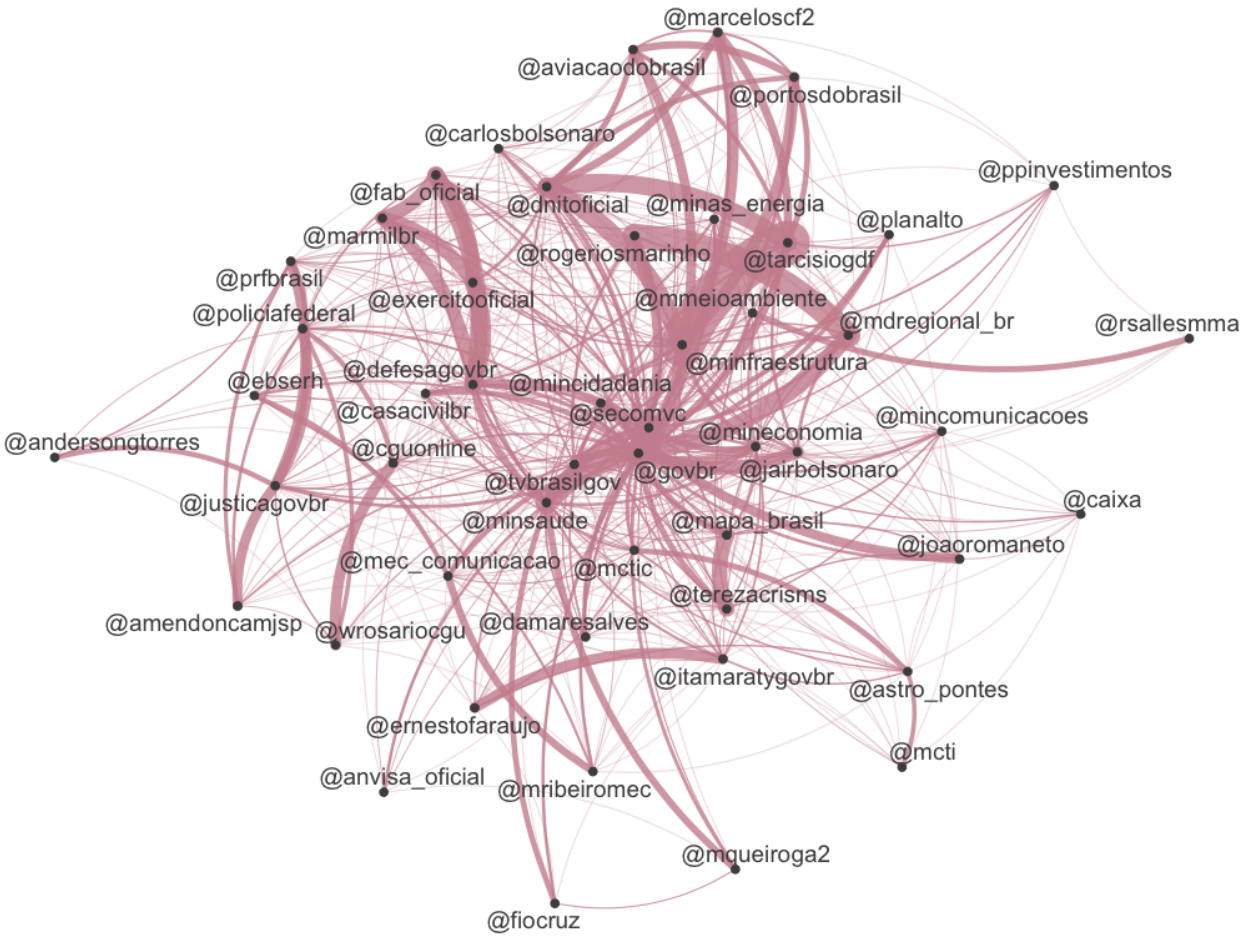


Figure 3: JB's users