Twitter API for Academic Research

Big data sources, crowdsourcing, crowdsensing

Michele Mazza

Install Selenium for Python

Go to:

https://developer.twitter.com/en/products/twitter-api/academic-research

and

APPLY

Describe the project

Over the past decade there has been a growing public fascination with the complex “connectedness” of modern society. This connectedness is found in many contexts: in the rapid growth of the Internet and the Web, in the ease with which global communication now takes place and in the ability of news and information to spread around the world with surprising speed and intensity. In this project we aim to analyze these phenomena involving networks and the aggregate behavior of groups of people; in a Social Network created from Twitter data.

During the European Football Championship the Belgium team started a trend: to go down on one knee and express solidarity with the Black Lives Matter movement. Some other teams decided to follow their example, others didn't. These events became public knowledge and every person in Europe had an opinion about them; the Italian team decided not to kneel and Italian citizens were divided among those who didn't agree with their decision and believed it was right for the team to kneel and others who expressed their solidarity with the decision to stand

Describe how you use the data

For the correct implementation and analysis of the network we will need to extract tweets from the Twitter API containing some hashtags, this is the query we built:

<<LIST OF HASHTAGS>

The hashtag are mostly in Italian, but we also check the tweet language with the attribute lang: it.
We will need these tweet attributes: user_id, mentions, retweet and quotes.
We will use them in this way: user_id are the nodes in the network, the nodes are linked to each other with the interactions that they do, technically speaking the edges will be mentions, retweets and quotes.
We won't use other attributes in this part of the project.
Later, to discover various communities within the network, we will need to classify the tweets and for that task we will check the hashtags.

Describe your methodology

For the correct implementation and analysis of the network we will do the following:
Coding with python;
Tweepy, in python, to interact with the twitter API;
https://docs.tweepy.org/en/stable/

Networkx, in python, to build the network, analyze and visualize it;
https://networkx.org/documentation/stable/index.html

We will parse each json object (tweet) to extract the user_id and add it to the network as a node, then we will add the edges using the various interactions we find inside the tweet object, iterating the procedure for the entire dataset extracted with the previous query. If you need more information about it, email me and I will try to send a piece of the code in order to better understand what is happening.

Describe how you will share the outcome

The project is for academic purposes only. It will not be used in any other way.
We will write a little essay to explain what we did and the data we used, we won't attach the data on it, just the code we write to build the network and analyze it.

I will discuss and look at the obtained network with my research team and professors in order to pass the exam and write the Master's thesis: this will be possible only with the academic API, otherwise the retrieved data won't be sufficient to implement a correct and real-like network. We won't look at the data themselves, just at the obtained network.