Seunghyuk Baek

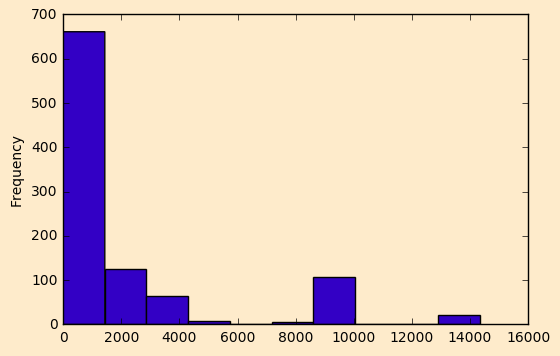
CS5664 spring 2018

HW3

For this assignment, I collected 1000 tweets that contain word ‘Donald Trump’ using tweepy’s offline tweet collect module.

1. diffusion of the topic

The propagation of the topic (‘Donald Trump’) can be visualized by histogram of retweeted tag.



It seems a lot of tweets are under 1000 retweet counts. However, some tweets are retweeted more than 8000 times.

1. Hashtag associated with the topic

The hashtag related with ‘Donald Trump’ is following:

[('', 942), (u'AmericaFirst', 18), (u'Trump', 12), (u'FakeNews', 8), (u'JulianAssange', 7), (u'DonaldTrump', 3), (u'BREAKING', 3), (u'LatinosForTrump', 2), (u'onpoli', 2), (u'Syria', 2)]

It seems the word ‘AmercaFirst’ is the most related to the ‘Donald Trump’

To perform sentiment analysis, the text 1000 tweets were tokenized and classified as positive or negative sentiment. Performing sentiment analysis was mainly adopted from marcobonzanini.com the blog of English data scientist Dr. Bonzanini. I downloaded his test set of positive and negative tweet samples and trained classifier with naïvebayesclassifier module. As a result, 1000 tweets containing ‘Donald Trump’ 466 positive tweets and 534 negative tweets.