

Twitter Analysis

Text processing & Network Analysis

Web Analytics
DATA620 Summer 2018

Team5: Christopher Estevez, Meaghan Burke, Rickidon Singh,
Ritesh Lohiya, Rose Koh

Contents

1. Introduction
2. Workflow
3. Data Acquisition and Transformation
4. Sentiment Analysis
5. Network Analysis

1. Introduction

- Twitter
- Hypothesis
- Data source

2. Workflow

- Data Acquisition and Transformation
- Frequency Analysis
- Wordcloud visualization
- Sentiment Analysis
- Network Analysis

3. Data Acquisition & Transformation

- Library {Tweepy; jsonpickle; json; pandas}
- Total tweets: 3,222
- Transformation
 - Extract username
 - Remove URL, RT
 - Remove Punctuations
 - Cast to lowercase
 - Remove double space

```
tweets = pd.DataFrame() #create pandas dataframe
tweets['TimeStamp'] = list(map(lambda tweet: tweet['created_at'], tweets_data))
tweets['Name'] = list(map(lambda tweet: tweet['user']['screen_name'], tweets_data))
tweets['text'] = list(map(lambda tweet: tweet['text'], tweets_data))
tweets['lang'] = list(map(lambda tweet: tweet['lang'], tweets_data))
tweets['tweet_id'] = list(map(lambda tweet: tweet['id'], tweets_data))
tweets['Favorites'] = list(map(lambda tweet: tweet['favorite_count'], tweets_data))
tweets['Retweet'] = list(map(lambda tweet: tweet['retweet_count'], tweets_data))
tweets['in_reply_to_status_id_str'] = list(map(lambda tweet: tweet['in_reply_to_status_id_str'], tweets_data))
tweets['in_reply_to_screen_name'] = list(map(lambda tweet: tweet['in_reply_to_screen_name'], tweets_data))
```

tweets.head(1)

	TimeStamp	Name	text	lang	tweet_id	Favorites	Retweet	in_reply_to_status_id_str	in_reply_to_screen_name
0	Thu Jul 19 12:12:54 +0000 2018	jimmyfallon	Google Translate Songs: #MammaMia Edition with...	en	1019917997127979008	209	34	None	None

```
tweets['text'][0]
```

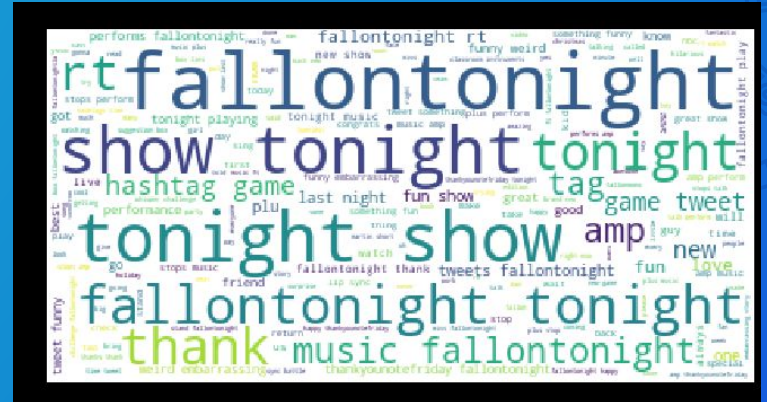
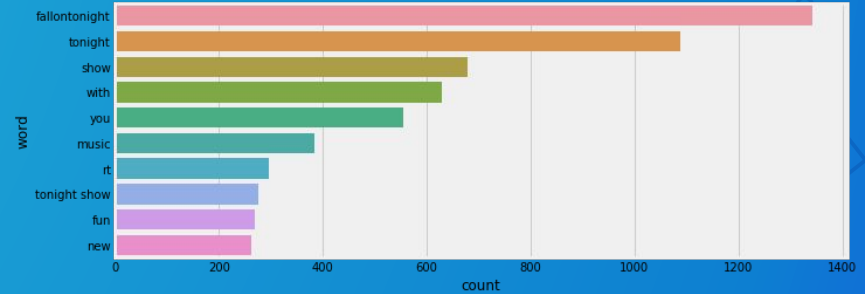
u'Google Translate Songs: #MammaMia Edition with #FallonTonight'

```
#delete certain punctuations, put the text in lower case and delete the double space with the function apply
import re
tweets['text'] = tweets['text'].apply(lambda x: re.sub('[!@#$.:.,?&]', '', x.lower()))
tweets['text'] = tweets['text'].apply(lambda x: re.sub(' ', ' ', x))
tweets['text'][0]
```

u'google translate songs mammamia edition with fallontonight'

4. Text Frequency

- Library
`{Sklearn.CountVectorizer;
Wordcloud}`
- Visualization



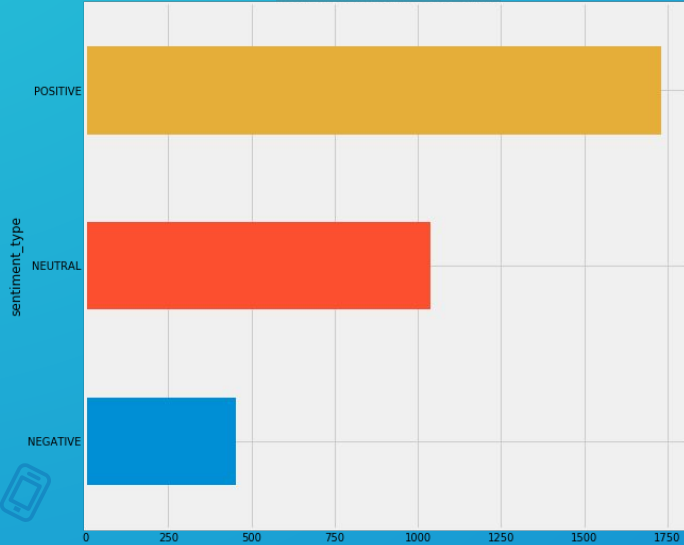
5. Sentiment Analysis

- Library {`nltk`, `sklearn.feature_extraction.text.TfidfVectorizer`}

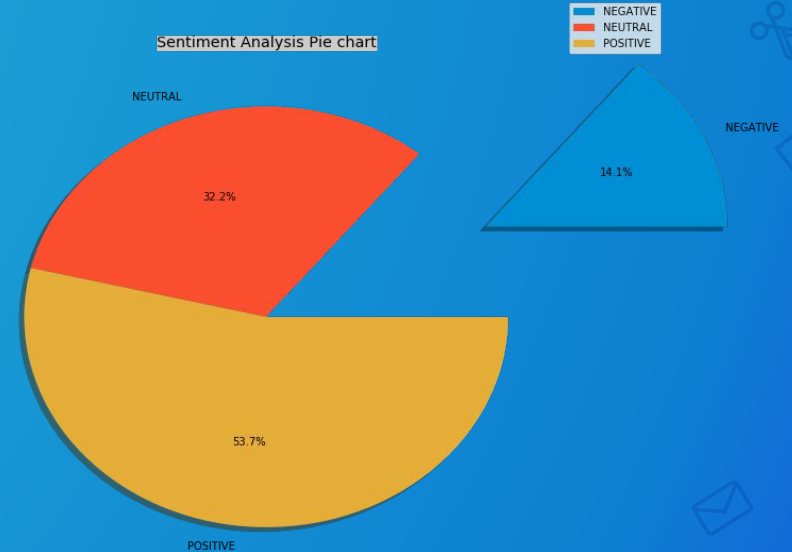
	sentiment_type	Retweet	Favorites	
sentiment_type				
	POSITIVE	1732	842.726905	3075.757506
	NEUTRAL	1037	788.453230	2601.179364
	NEGATIVE	454	1096.546256	2746.123348

5. Sentiment Analysis

Sentiment Analysis Bar chart



Sentiment Analysis Pie chart

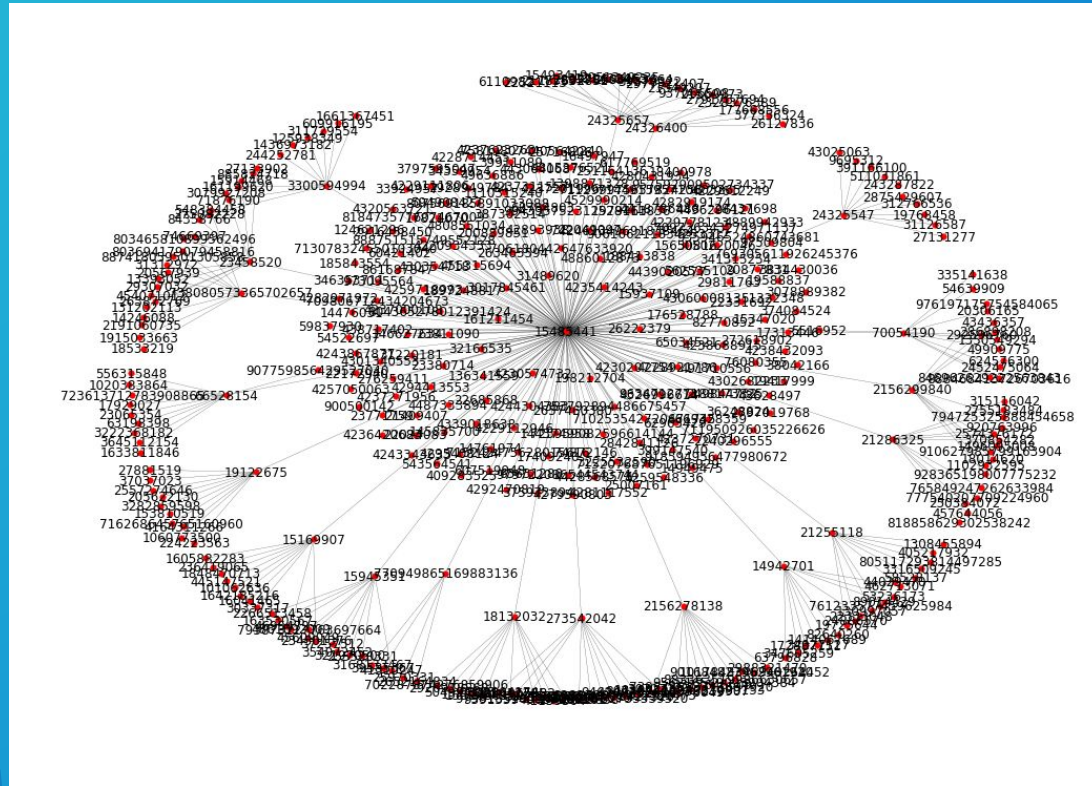


6. Network Analysis

- `api.friends_ids()`
- `api.get_user()`
- Combine:
 - Following list of JimmyFallon
 - JimmyFallon's Followers

	ID	Name	Screen Name	Followers Count	Favorites Count	Followed_By
54	2156278138	Grayson Dolan	GraysonDolan	6480632	10430	15485441
55	2156299840	Ethan Dolan	EthanDolan	6339582	20937	15485441
60	19122675	Alex Ovechkin	ovi8	2681183	14	15485441
0	15945351	Charlie Puth	charlieputh	2529730	2931	15485441
12	18132032	Dale Earnhardt Jr.	DaleJr	2465238	17766	15485441

6. Network Analysis



6. Network Analysis

```
combined_df[(combined_df["ID"] == 70054190) | #https://twitter.com/FrankieJGrande  
(combined_df["ID"] == 66528154) | #https://twitter.com/mikeshinoda  
(combined_df["ID"] == 5516952) | #https://twitter.com/lauren_hoggs  
(combined_df["ID"] == 907759856429527040)] # https://twitter.com/underyou
```

	ID	Name	Screen Name	Followers Count	Favorites Count	Followed_By
180	70054190	Frankie James Grande	FrankieJGrande	1076612	19607	15485441
49	66528154	Mike Shinoda	mikeshinoda	1035725	11913	15485441
50	907759856429527040	Lauren Hogg	lauren_hoggs	59147	2878	15485441
192	5516952	✨	underyou	35024	7630	15485441
96	907759856429527040	Lauren Hogg	lauren_hoggs	59151	2878	66528154
87	5516952	✨	underyou	35024	7630	70054190



Thanks!

Github Repository:

https://github.com/silverrainb/web-analytics/tree/master/final_project

Youtube Link:

<https://youtu.be/-K0e00Xw70Y>