<https://drive.google.com/file/d/0B3e3qZpPtccsMFo5bk9Ib3VCc2c/view>

<https://www.kaggle.com/mrisdal/fake-news>

www.cs.cornell.edu/people/pabo/movie-review-data/

Gillin, J.: Politifact’s guide to fake news websites and what they peddle. PunditFact, 20 April 2017. http://bit.ly/2pHYKDV

Glader, P.: 10 journalism brands where you ﬁnd real facts rather than alternative facts. Forbes, 1 February 2017. <http://bit.ly/2sXPpvf>

Scrapy: An open source and collaborative framework for extracting the data you need from websites in a fast, simple, yet extensible way.

<https://scrapy.org/>

<http://fakenewswatch.com/> comprehensive resource on online misinformation

<http://www.mturk.com>

<http://www.twittermonitor.net/>

<http://mia.kaist.ac.kr/publications/rumor>

https://www.statista.com/topics/1164/social-networks/https://www.snopes.com/fact-check/600-murders-in-chicago/

<http://www.emergent.info/>

<https://www.snopes.com/>

<https://www.factcheck.org/>

Stanford large network dataset collection

<http://snap.stanford.edu/data/>

<https://tags.hawksey.info/>

<https://www.weibo.com/overseas>

<http://www-personal.umich.edu/~mejn/netdata/>

I’ve trained the model on a synthesis of five large datasets:

* the [Kaggle Fake News](https://www.kaggle.com/c/fake-news) set
* the [FakeNewsNet](https://github.com/KaiDMML/FakeNewsNet) set
* the [Kaggle Fake News Detection](https://www.kaggle.com/jruvika/fake-news-detection) set
* the [Fake Or Real News](https://github.com/GeorgeMcIntire/fake_real_news_dataset) set
* the [Kaggle Getting Real About Fake News](https://www.kaggle.com/mrisdal/fake-news) set

<http://trendistic.com/view/updates/>

<http://trendistic.com>

<https://www.dropbox.com/s/7ewzdrbelpmrnxu/rumdetect2017.zip?dl=0> (most important)

<http://alt.qcri.org/~wgao/data/> (most important)

<http://service.account.weibo.com> (Chinese website)

https://twitter.com/PolitiFact

http://www.politifact.com

PolitiFact website is operated by the Tampa Bay Times, where the reporters and editors can make fact-check regarding the statements.

<http://www.fakenewschallenge.org/>

# Software for complex networks: NetworkX <https://networkx.github.io/>

Twitter developer platform <https://developer.twitter.com/>

<https://www.weibo.com/login.php?url=http%3A%2F%2Fservice.account.weibo.com%2F>

**Reddit an online social media and discussion website identifies controbversial threads based on up-votes and down votes**

Link <https://www.reddit.com/controversial>

Reditt code is open source link <https://github.com/reddit-archive/reddit>

# **An Exhaustive Study of Twitter Users Across the World**

Link <http://www.beevolve.com/twitter-statistics/>

# **Implementation of Darwini graph generator**

Link <https://issues.apache.org/jira/browse/GIRAPH-1043>

Important links for getting datasets from twitter, Sina Webio and facebook API

<https://dev.twitter.com/docs>

<https://apps.twitter.com>

<http://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3/en>

<http://www.socialgist.com/>

<http://www.socialgist.com/press/socialgist-emerges-as-the-first-official-provider-of-social-data-from-chinese-microblogging-platform-sina-weibo/>

<https://developers.facebook.com/docs/>

<https://developers.facebook.com/docs/apps/register>

publicly available dataset is the PHEME dataset of rumours and non-rumours,11 which includes a collection of 1,972 rumours and 3,830 non-rumours associated with 5 breaking news stories (Zubiaga et al. 2016b)

<https://figshare.com/articles/PHEME_dataset_of_rumours_and_non-rumours/4010619>

The most widely used dataset for rumour tracking is that by Qazvinian et al. (2011), which includes over 10,000 tweets associated with 5 different rumours, each tweet annotated for relevance toward the rumour as related or unrelated.

While not specifically intended for rumour tracking, the dataset produced by Zubiaga et al.(2016) provides over 4,500 tweets categorised by rumour.

PHEME stance dataset which provides tweet-level annotations of stance (support, deny, query, comment) for tweets associated with nine events.

<https://figshare.com/articles/PHEME_rumour_scheme_dataset_journalism_use_case/2068650>

Other publicly available dataset is Ferreira and Vlachos (2016); this dataset does not provide social media data, but it may be leveraged for social media rumour classification as it contains 300 rumoured claims and 2,595 associated news articles, collected and labelled by journalists, along with an estimation of their veracity.

A dataset released for the Fake News Challenge is also annotated for stance (agrees, disagrees, discusses, unrelated). This dataset is, however, made of news articles instead of social media posts.

<http://www.fakenewschallenge.org/> (most important)

pre-trained word embeddings Google Newsword2vec embeddings code

<https://github.com/mmihaltz/word2vec-GoogleNews-vectors>

The dataset produced for RumourEval 2017 (Derczynski et al. 2017), a shared task that took place at SemEval 2017, includes over 300 rumours annotated for veracity as one of true, false, or unverified. Another dataset suitable for veracity classification is that released by Kwon et al. (2017), which includes 51 true rumours and 60 false rumours. Each rumour includes a stream of tweets associated with it.

**Web based Fact checking/rumour verification/fake news detection application links**

<http://www.pheme.eu/>

<http://www.emergent.info/>

<https://www.si.umich.edu/research/research-projects/rumorlens/>

<http://twittertrails.com/>

<http://hoaxy.iuni.iu.edu/>

<http://revealproject.eu/>

<http://www.invid-project.eu/>

<https://firstdraftnews.com/crosscheck-launches/>

<http://www.lemonde.fr/verification/>

<https://meedan.com/en/check/>

<http://idir-server2.uta.edu/claimbuster>

<http://www.unahakika.org/>

<http://www.nickdiakopoulos.com/2012/01/24/finding-news-sources-in-social-media/>

<http://twitdigest.iiitd.edu.in/TweetCred/>

<http://www.snopes.com/>

<https://knightcenter.utexas.edu/>

<http://ejc.net/>

<http://us2.campaign-archive1.com/?u=657b595bbd3c63e045787f019&id=2208e04aa6>

<http://us9.campaign-archive1.com/?u=79fa45ed20ff84851c3b9cd63&id=02624abd8b>

<http://www.politifact.com/>

<https://www.factcheck.org/>

<https://www.washingtonpost.com/news/fact-checker>

<https://www.snopes.com/>

<https://www.truthorfiction.com/>

<https://fullfact.org/>

<http://hoax-slayer.com/>

<https://reporterslab.org/fact-checking/>

<http://fiskkit.com/>

<https://blog.google/topics/journalism-news/labeling-fact-check-articles-google-news/>

<https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/>

<https://blog.twitter.com/2010/trust-and-safety>

<http://service.account.weibo.com/> (sign in required)

<http://www.axiapr.com/blog/elements-of-news> (fact-checking studies that have considered timeliness are to date limited, given the importance of timeliness in news)

The source url is accessed and the downloaded content parsed using the Newspaper library

Link <https://github.com/codelucas/newspaper/>

FakeNewsNet: <https://github.com/KaiDMML/FakeNewsNet> The dataset is constructed using an end-to-end system,FakeNewsTracker <http://blogtrackers.fulton.asu.edu:3000>

BuzzFeedNews: <https://github.com/BuzzFeedNews/2016-10-facebook-factcheck/tree/master/data>

LIAR https://www.cs.ucsb.edu/ william/software.html

BS Detector <https://github.com/bs-detector/bs-detector>

CREDBANK <http://compsocial.github.io/CREDBANK-data/>

BuzzFace <https://github.com/gsantia/BuzzFace>

FacebookHoax <https://github.com/gabll/some-like-it-hoax>

**BuzzFeedNews** : This dataset comprises a complete sample of news published in Facebook from 9 news agencies over a week close to the 2016 U.S. election from Septem- ber 19 to 23 and September 26 and 27. Every post and the linked article were fact-checked claim-by-claim by 5 Buz- zFeed journalists. It contains 1,627 articles 826 mainstream, 356 left-wing, and 545 right-wing articles.

**LIAR** : This dataset (Wang 2017) is collected from fact- checking website PolitiFact. It has 12.8 K human labelled short statements collected from PolitiFact and the statements are labeled into six categories ranging from completely false to completely true as pants on fire, false, barely-true, half- true, mostly true, and true.

**BS Detector** : This dataset is collected from a browserextension called BS detector developed for checking news veracity. It searches all links on a given web page for refer- ences to unreliable sources by checking against a manually compiled list of domains. The labels are the outputs of the BS detector, rather than human annotators.

**CREDBANK** : This is a large-scale crowd-sourceddataset (Mitra and Gilbert ) of around 60 million tweets that cover 96 days starting from Oct. 2015. The tweets are related to over 1,000 news events. Each event is assessed for credi- bilities by 30 annotators from Amazon Mechanical Turk.

**BuzzFace** : This dataset (Santia and Williams ) is collected by extending the BuzzFeed dataset with comments related to news articles on Facebook. The dataset contains 2263 news articles and 1.6 million comments discussing news content.

**FacebookHoax** : This dataset (Tacchini et al. 2017) comprises information related to posts from the facebook pages related to scientific news (non- hoax) and conspir- acy pages (hoax) collected using Facebook Graph API. The dataset contains 15,500 posts from 32 pages (14 conspiracy and 18 scientific) with more than 2,300,000 likes.

BuzzFeedNews only contains headlines and text for each news piece and covers news articles from very few news agencies. LIAR dataset contains mostly short statements instead of entire news articles with the meta at- tributes. BS Detector data is collected and annotated by using a developed news veracity checking tool, rather than using human expert annotators. CREDBANK dataset was originally collected for evaluating tweet credibilities and the tweets in the dataset are not related to the fake news articles and hence cannot be effectively used for fake news detection. BuzzFace dataset has basic news contents and social context information but it does not capture the temporal information. The FacebookHoax dataset consists very few instances about the conspiracy theories and scientific news.

<https://www.kaggle.com/kumudchauhan/fake-news-analysis-and-classification>

Fake news detection papers with code

<https://paperswithcode.com/task/fake-news-detection/latest>

<https://paperswithcode.com/task/fake-news-detection>

international workshop on semantic evaluation (SEMEVAL 2016,2017,2018,2019,2020)

refer papers and datasets of semeval for problems also

<http://alt.qcri.org/semeval2019/index.php?id=tasks>

<http://precog.iiitd.edu.in/resources.html> (important dataset link)