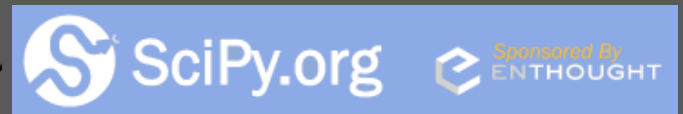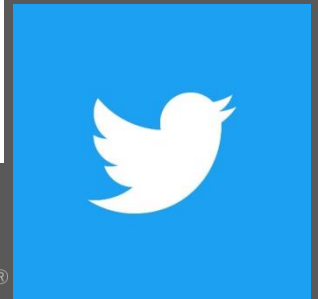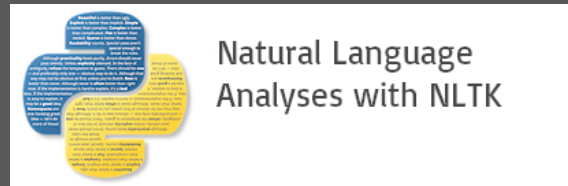Explore and Understand
the Amount of Fabrication
of a Topic in Twitter

BotBoosted

Brian Balagot
brian.balagot@gmail.com
https://www.linkedin.com/in/briancbalagot
https://github.com/brityboy

scikit
learn

SciPy.org

Sponsored By
ENTHOUGHT

Natural Language
Analyses with NLTK

mongoDB.

RAPPLER — News Video Business Newsbreak MovePH Views Life & Style Entertainment Sports Tech Br

INVESTIGATIVE

Fake accounts, manufactured reality on social media

Suspicious and fake Facebook accounts share common practices, forming a nest that spreads and repeats lies and propaganda

Chay F. Hofileña
@chayhofilena
Published 10:04 AM, October 09, 2016
Updated 9:55 AM, October 15, 2016

56
17K
Twitter

ILLUSTRATION BY EMIL MERCADO

SECTIONS — HOME — SEARCH — The New York Times

Fake News in U.S. Election? Elsewhere, That's Nothing New

By PAUL MOZUR and MARK SCOTT NOV. 17, 2016

An Internet cafe in Manila. The Philippines is one of Facebook's fastest-growing markets.
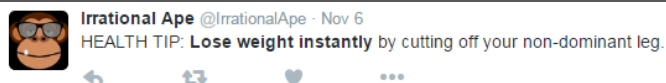Jay Directo/Agence France-Presse — Getty Images

An **app that can help** me **understand how much** about a topic on an online social network **is fabricated or not**, and **what the conversations are** about

To **explore** and **understand** the amount of **fabrication** on **Twitter**...

*A tweet is **fabricated** if it was not made by a bona-fide human or genuine news source*
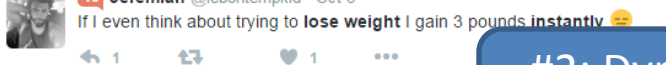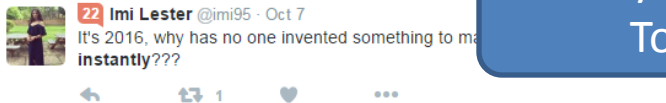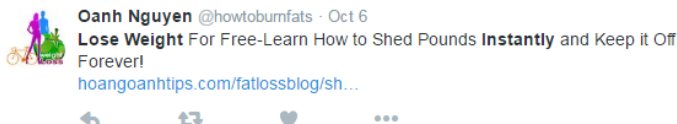
Search: lose weight instantly

REAL

Irrational Ape @IrrationalApe · Nov 6
HEALTH TIP: **Lose weight instantly** by cutting off your non-dominant leg.

#1: Lightweight Supervised Classifier

Nov 5
lose weight, follow his advice and

REAL

@lebonlempkiu · Oct
If I even think about trying to **lose weight** I gain 3 pounds **instantly**

REAL

Imi Lester @imi95 · Oct 7
It's 2016, why has no one invented something to ma
instantly???

#2: Dynamic Lightweight Topic Modeler

FAKE

Oanh Nguyen @howtoburnfats · Oct 6
Lose Weight For Free-Learn How to Shed Pounds **Instantly** and Keep it Off Forever!
hoangoanhtips.com/fatlossblog/sh...

REAL

Jamaicans Dot Com @jamaicansdotcom · Oct 6
VIDEO: Courgette Smoothie **Lose Weight** And Burn Belly Fat **Instantly**:
Courgette Smoothie **Lose Weight** And Burn B... bit.ly/2dudODB

"think invent try lose weight"

100% real

Imi Lester @imi95 · Oct 7
It's 2016, why has no one invented something to make you **lose weight instantly**???
1

REAL

#3: Corpus Summarizer

Oanh Nguyen @howtoburnfats · Oct 6
Lose Weight For Free-Learn How to Shed Pounds **Instantly** and Keep it Off Forever!
hoangoanhtips.com/fatlossblog/sh...

FAKE

Jamaicans Dot Com @jamaicansdotcom · Oct 6
VIDEO: Courgette Smoothie **Lose Weight** And Burn Belly Fat **Instantly**:
Courgette Smoothie **Lose Weight** And Burn B... bit.ly/2dudODB

REAL

"learn shed pounds burn belly fat smoothie"
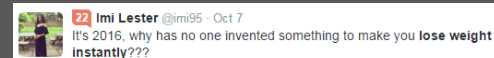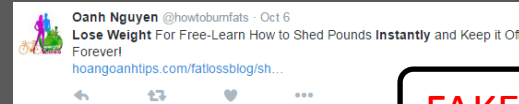
50% real

My Data Strategy / Pipeline

Labeled Tweets

"Genuine" Tweets

"Fabricated" Tweets

User Information (Prediction Features)

Topic Modeler

Topic Modeler

Supervised Classifier

Labeled dataset

Topics

Topics

Topics

Corpus Summarizer

Exemplary Tweets

Representative Words

9M tweets
9k users
75% fake, 25% real
@TheFakeProject

http://mib.projects.iit.cnr.it/dataset.html
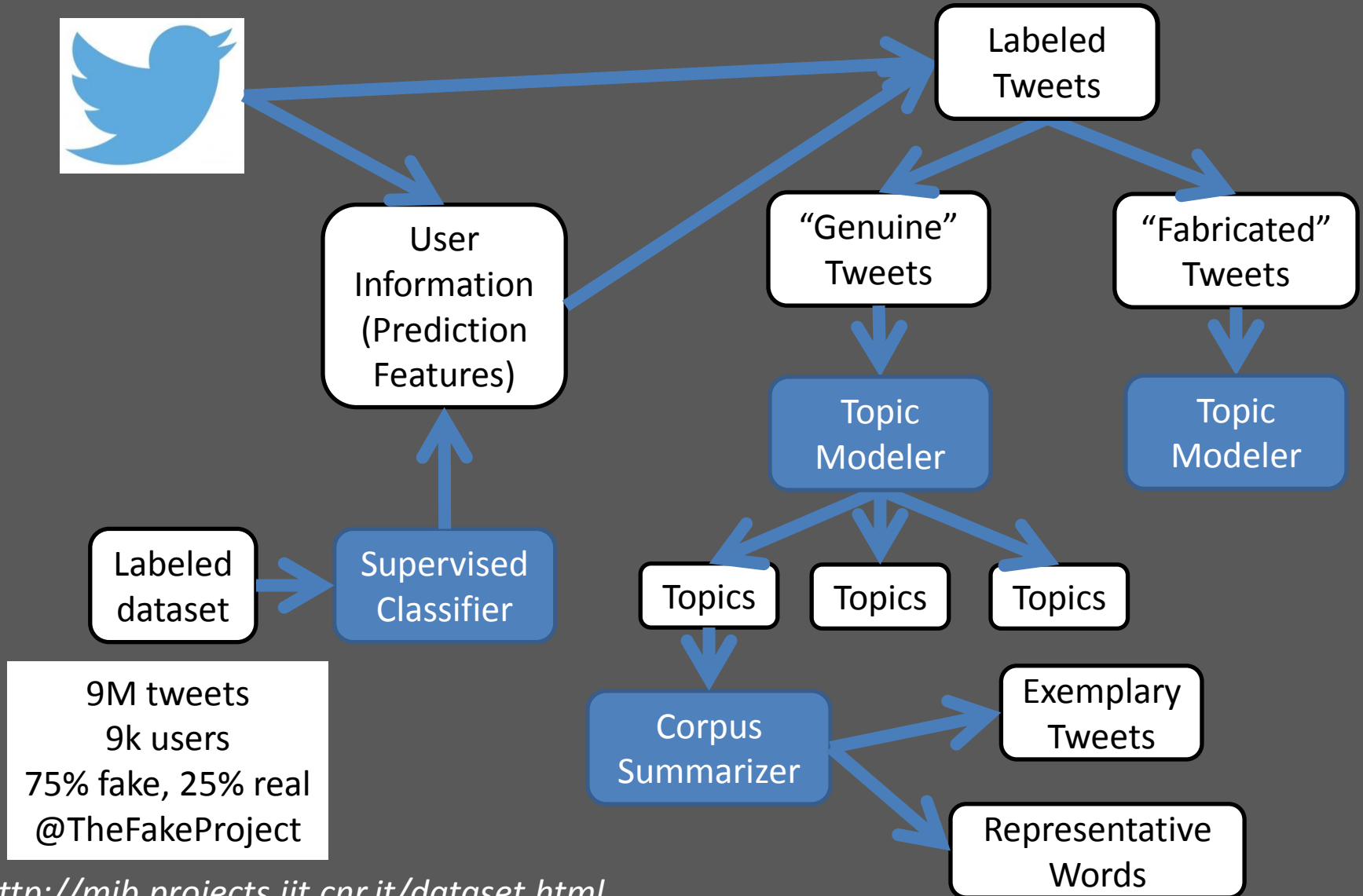
1: Lightweight Supervised Classifier

The GOAL: *Classify a user, based on their most recent tweet, as real or fake*

Account History: followers_count, friends_count, total_tweets, total_likes

Relative_Volume: likes_friends, tweets_friends

Behavior Rate: likes_day, tweets_day, friends_day, likes_friends_day

10FCV Acc: 97% → Account History → Random Forest

10FCV Acc: 95% → Behavior Rate → Random Forest
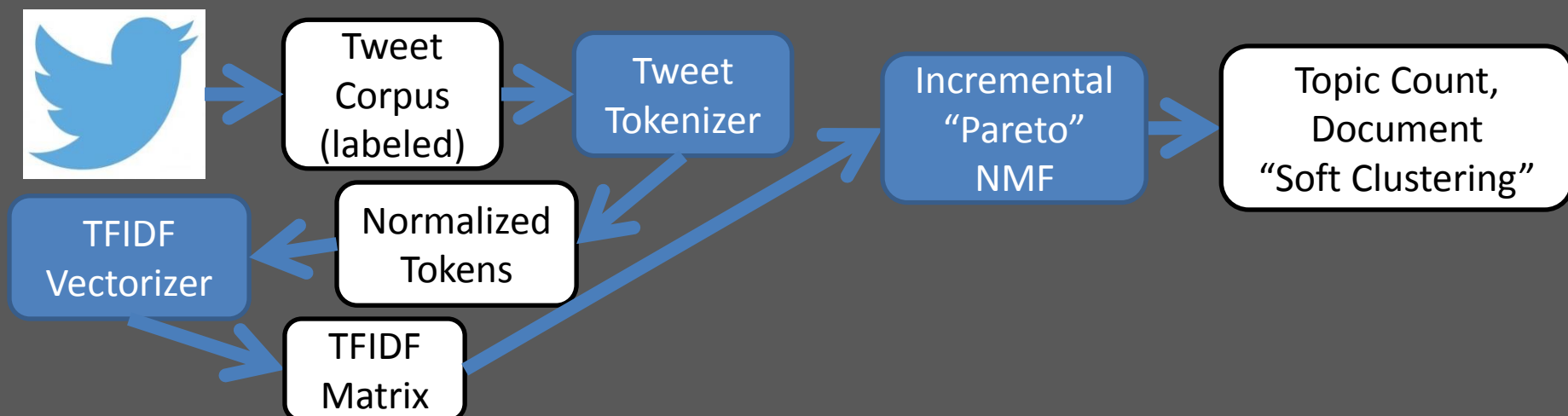
Account History
Behavior Rate
History Pred %
Rate Pred %
→ Random Forest → 10FCV Acc: 98%

2: Dynamic Lightweight Topic Modeler

The GOAL: *Heuristically determine the topic count of a corpus, on the fly*


```
initializing evaluation...
extracting 2 topics...
[9361 2543] is the topic distribution
extracting 4 topics...
[6683 2719 1933  569] is the topic distribution
extracting 6 topics...
[4284 2266 2111 1616 1105  522] is the topic distribution
extracting 8 topics...
[3289 2244 1918 1592 1093 1061  517  190] is the topic distribution
extracting 10 topics...
[2232 1721 1634 1583 1162 1135 1116  515  468  338] is the topic distribution
heuristic topic count is 8
```

3: Corpus Summarizer

The GOAL: *extract the most important tweet that captures the subtopic*
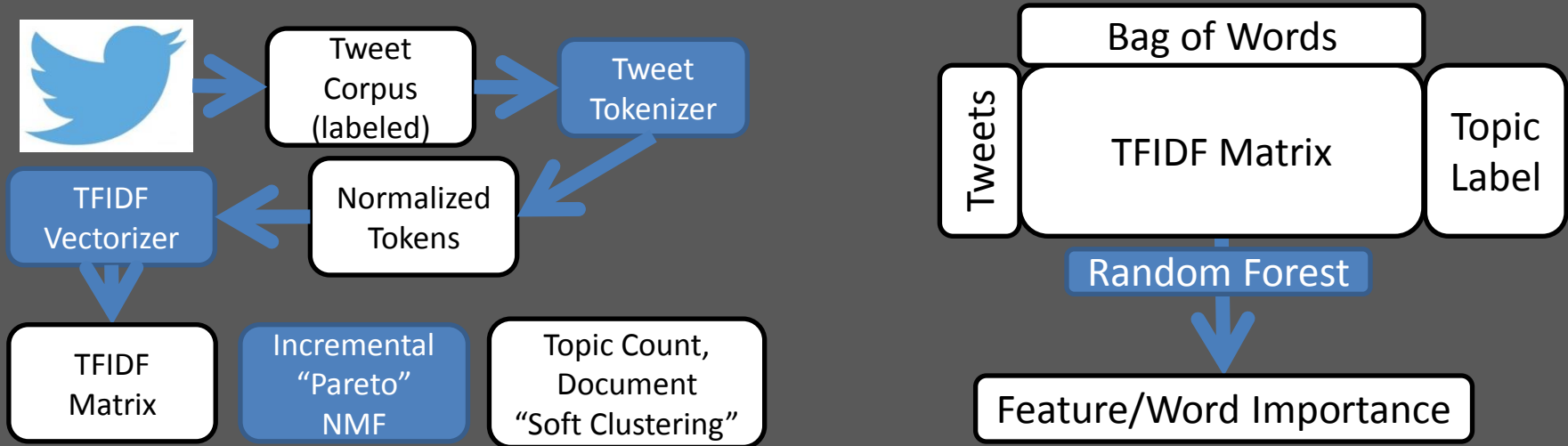
A tweet is important if it is made up of important words

A word is important if it is helpful in "bucketing" tweets into topics

Word Importance = **Term Frequency** X **Inverse Document Frequency** X **Feature/Word Importance**

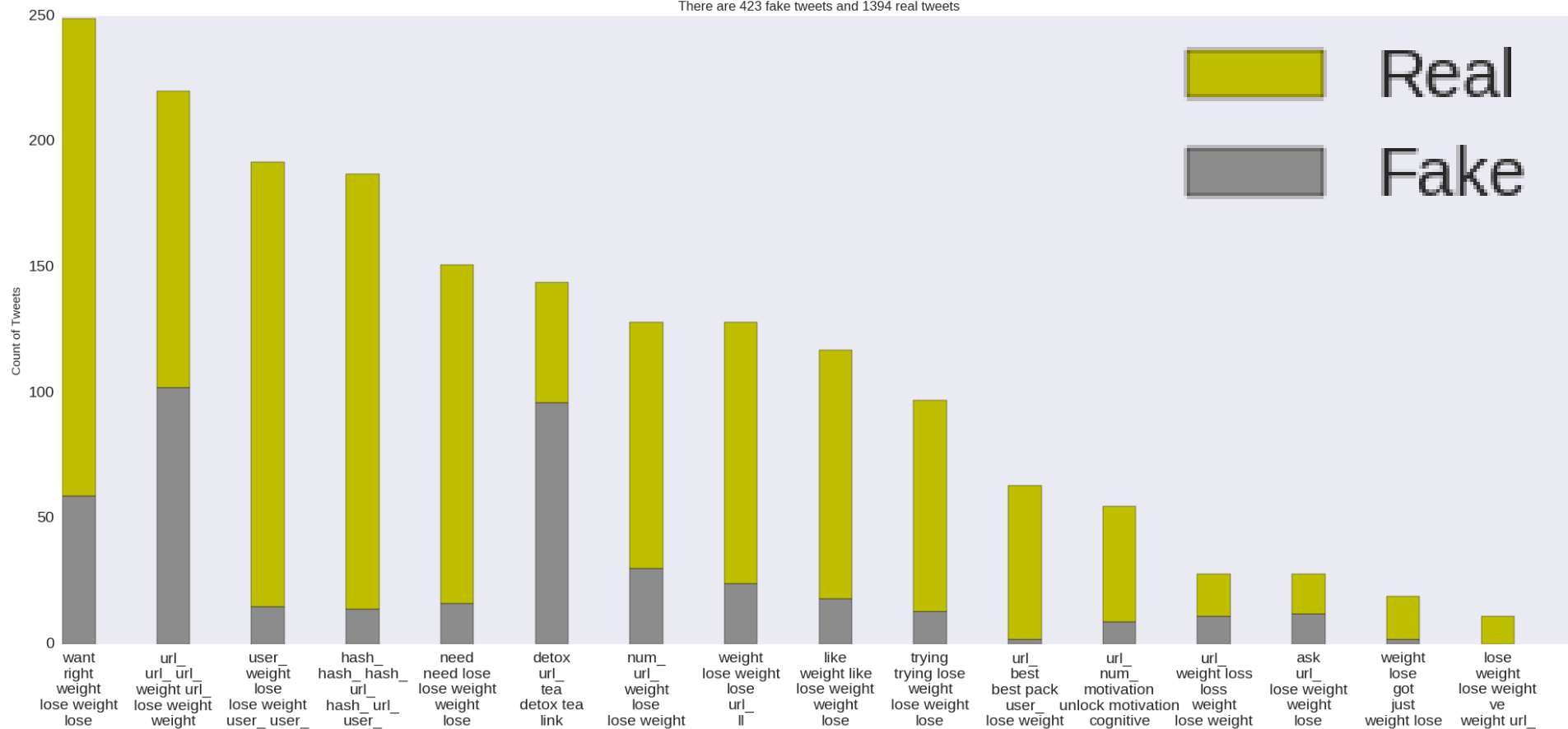**Tweet Importance** = **Highest Total Word Importance**

Tweets by Topic, by Real/Fake, for: lose weight now

There are 423 fake tweets and 1394 real tweets

Fake:Real Tweets / Topic Percent for lose weight now

On Average, 23.0 Percent are Fake, 77.0 Percent are Real

BotBoosted
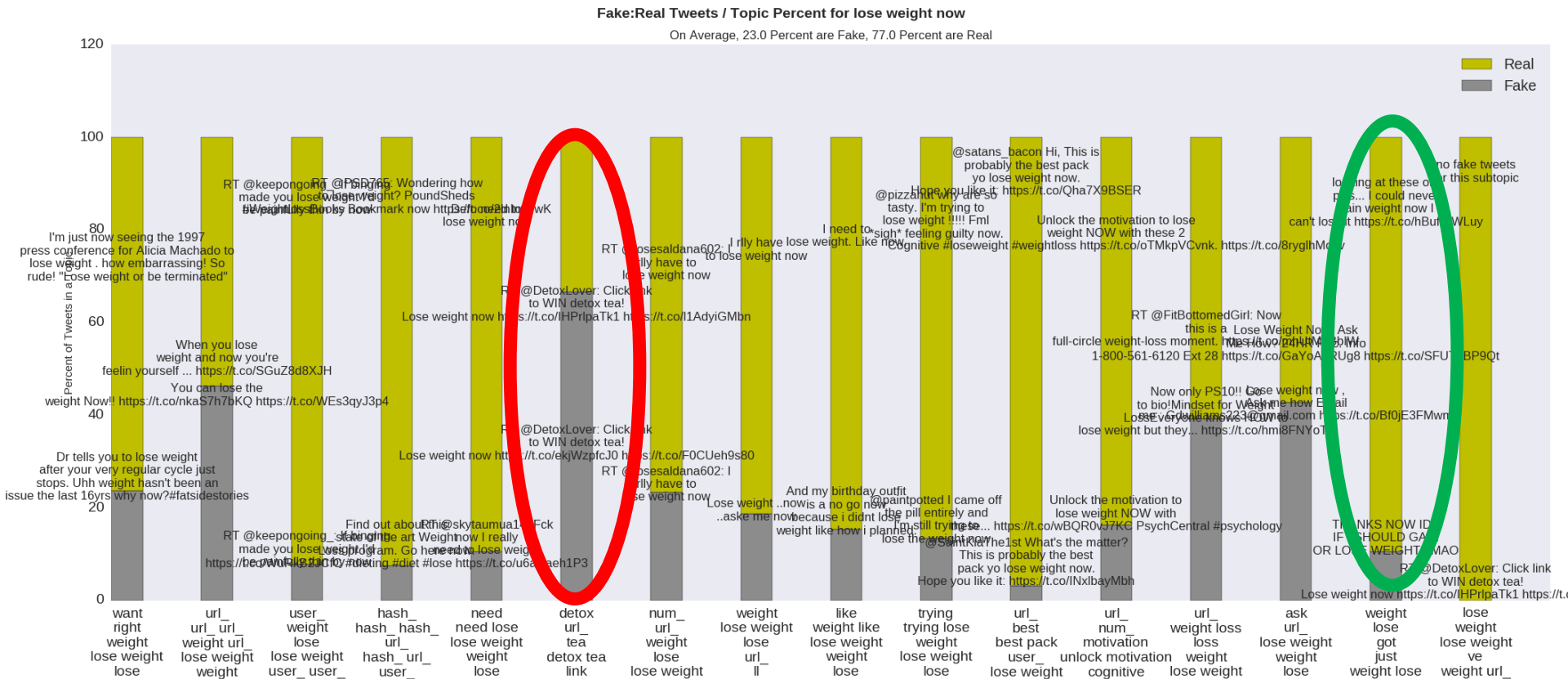
We can **see** which **conversations** are **boosted by bots**

Limitations & Future Work

- Twitter is <u>very aggressive at suspending spammers</u> *(77% suspended 1 day after 1st tweet)* (BotBoosted detects Twitter's False Negatives)

- Twitter's FREE API is not a statistically representative sample → Garden Hose API

- Strengthen the model by continuously training it with twitter's false negatives

- Benchmark Incremental "Pareto" NMF vs HDP and LDA to further improve the heuristic algorithm

- Improve corpus summarization with graph theory via tweet "centrality" based on word co-occurrence

BotBoosted

Explore and Understand the Amount of Fabrication of a Topic in Twitter

Thank You!

Major References:

- Aritter. Twitter NLP. (2016). Github repository https://github.com/aritter/twitter_nlp
- Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M. (2015). Fame for Sale: efficient detection of fake Twitter followers. Retrieved from Cornell University Library. (arXiv:1509.04098)
- Joseph, K., Landwehr, P., Carley, K. (2014). Two 1%s don't make a whole: Comparing simultaneous samples from Twitter's Streaming API.
- Karambelkar, B. (2015, Jan 5). How to use Twitter's Search REST API most effectively. Retrieved from http://www.karambelkar.info/2015/01/how-to-use-twitters-search-rest-api-most-effectively./
- Kharde, V., Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. International Journal of Computer Applications. Volum 139, No 11, April 2016.
- Kontaxis, G., Polakis, I., Ioannidis, S., Markatos, E. (2011). Detecting Social Network Profile Cloning. Retrieved from SysSec Consortium. (n.d.).
- Mori, T., Kikuchi, M., Yoshia K. (2001). Term Weighting Method based on Information Gain Ratio for Summarizing Documents retrieved by IR systems. Retrieved from http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings2/morisumm.pdf
- Thomas, K., Grier, C., Paxson, V., Song, D. (2011). Suspended Accounts in Retrospect: An Analysis of Twitter Spam. Retrieved from http://www.icir.org/vern/papers/twitter-susp-accounts.imc2011.pdf

Brian Balagot
brian.balagot@gmail.com
https://www.linkedin.com/in/briancbalagot
https://github.com/brityboy