

# Bot-or-Not: social bot detection

Onur Varol, Clayton Davis, Prashant Shiralkar, Emilio Ferrara, Filippo Menczer, and Alessandro Flammini



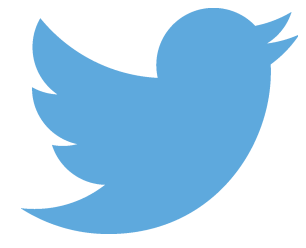
INDIANA UNIVERSITY  
BLOOMINGTON

## I. Detecting social bots

Can we verify the authenticity of users who participate in discussions and produce content in social media?

Our system detects with **94%** accuracy (measured by AUC) *whether a user is a bot or not*, by exploiting over 800 features that capture **user meta-data, social contacts, diffusion networks, content, sentiment, and temporal patterns.**

## II. Dataset



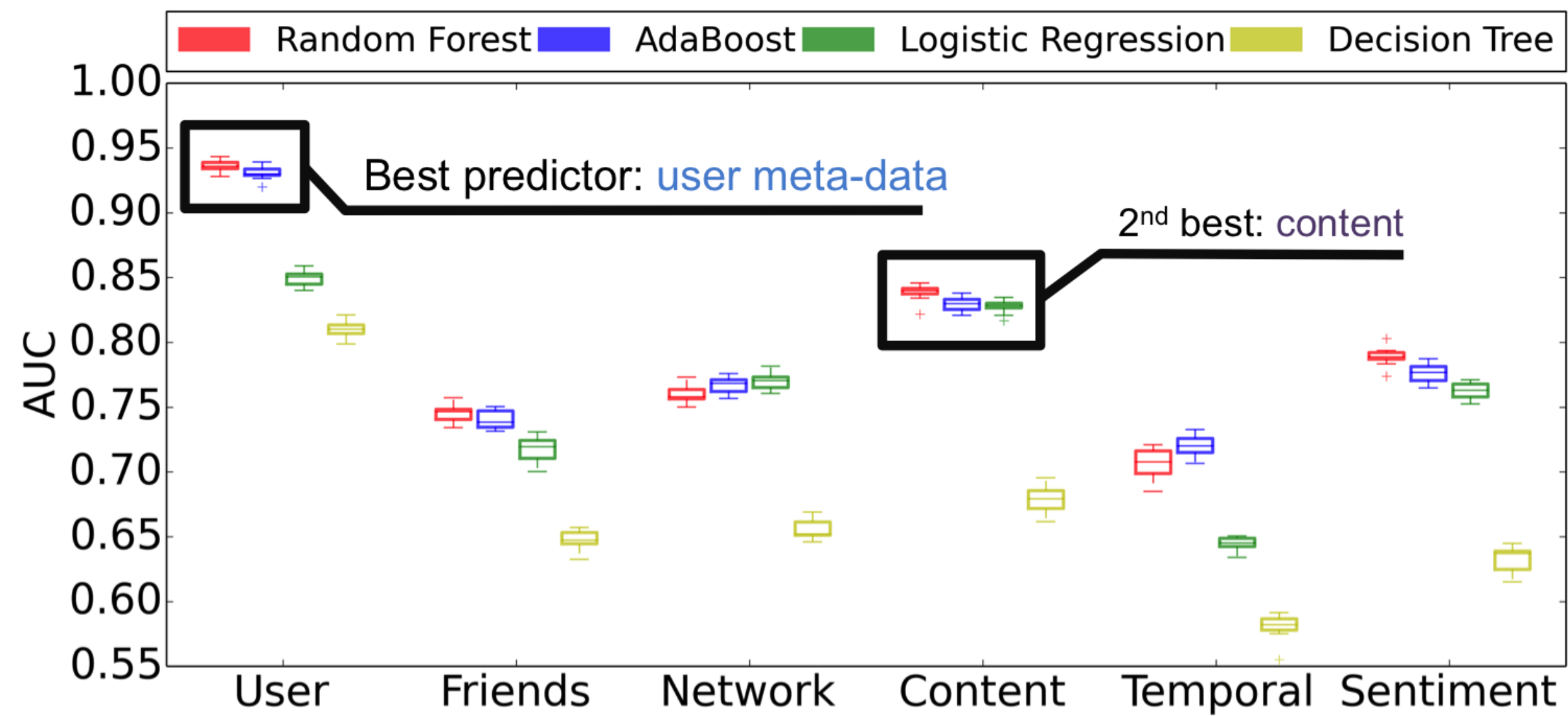
14.8K Legitimate users  
15.8K Social bot users

Lee, Kyumin, Brian David Eoff, and James Caverlee. "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter." ICWSM. 2011.

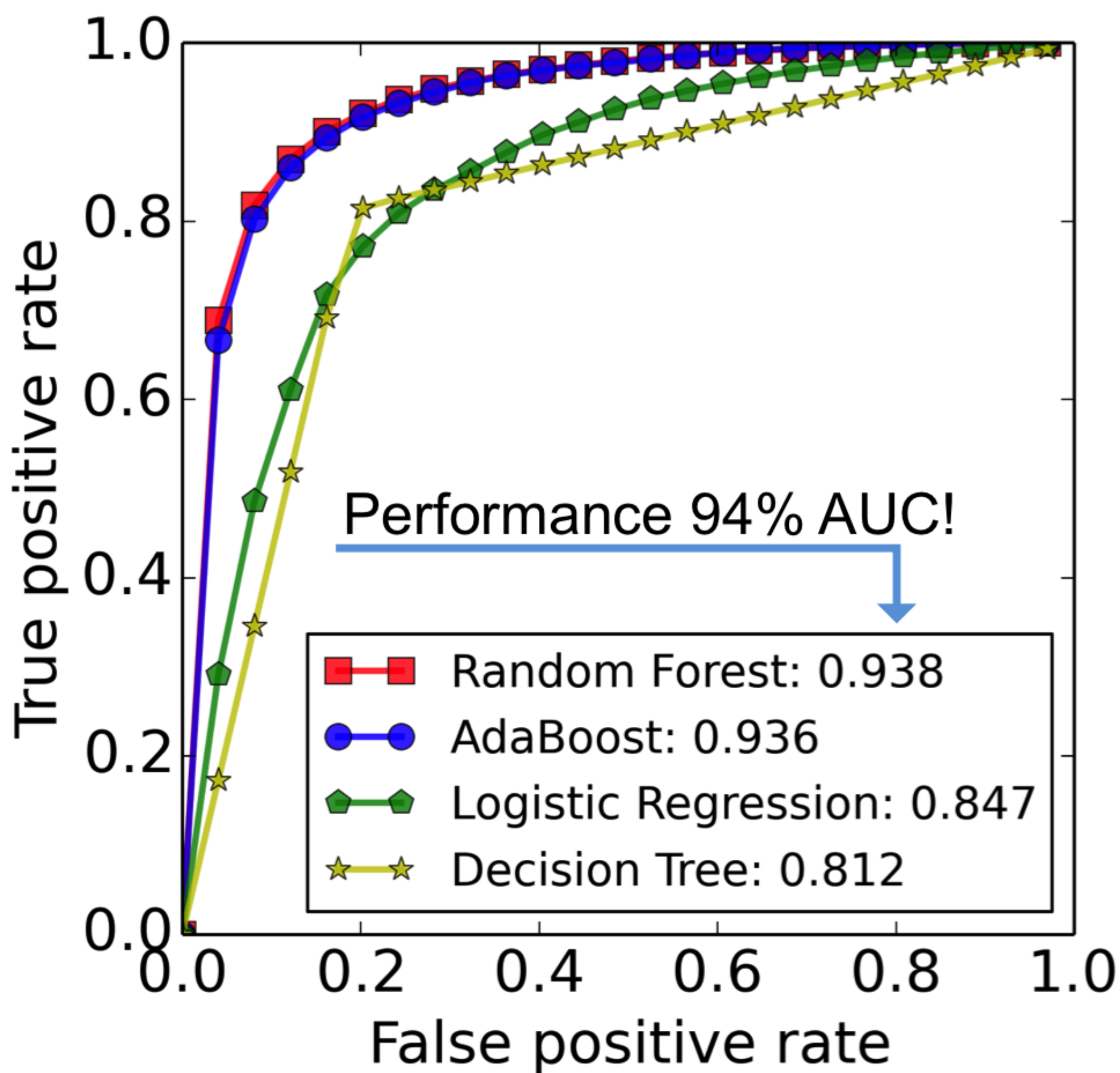
## III. Features and feature classes

Class (#features)	Description and Examples
Network (101)	Retweet, Mention, HT-co-occurrence networks
User (56)	User meta-data.
Friend (208)	Contacts of a given user.
Timing (16)	Temporal info.
Content (274)	Part-of-Speech (POS) tagging & Meme information
Sentiment (226)	Features representing sentiments.

## IV. Feature selection



## V. Performance evaluation: ROC



## VI. Demo

Front-end: **HTML5/Javascript** — Back-end: **Python/Flask** — Interface: **D3.js** — Authentication: **Twitter**

Bot or Not? A Truthy project

truthyatindiana Check User

Hello Clayton Allen Davis and welcome to BotOrNot. Here is the report you requested.

- Retrieving basic user data from Twitter ✓
- Retrieving user timeline from Twitter ✓
- Retrieving retweets/mentions from Twitter ✓
- Analyzing data with Truthy ✓

Classification

62%

Content 46% Temporal 32% Sentiment 45% Network 43%

Friend 71% User 56%

<http://truthy.indiana.edu/botornot>

## VII. User information

Customizations

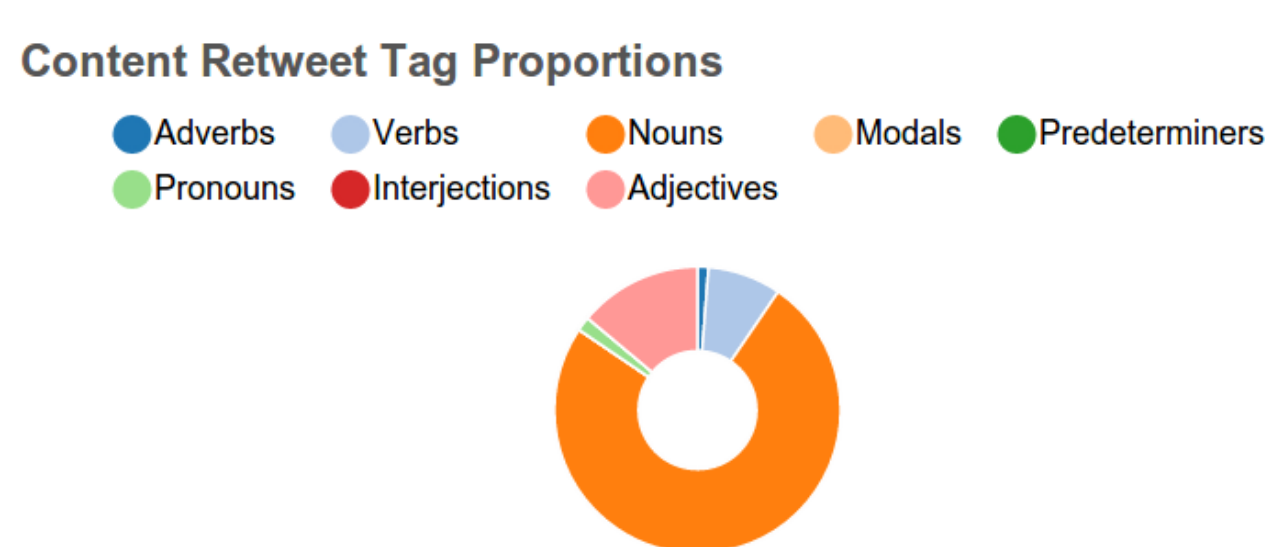
Screen name

Description text

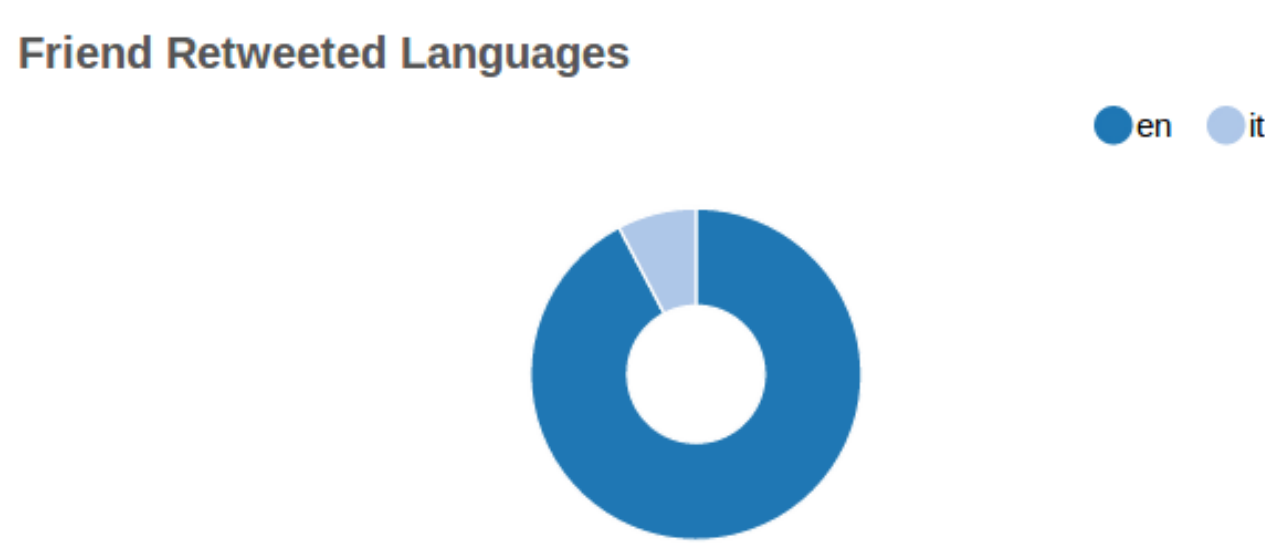
Temporal changes

TWEETS 1,834 FOLLOWING 3,666 FOLLOWERS 67K

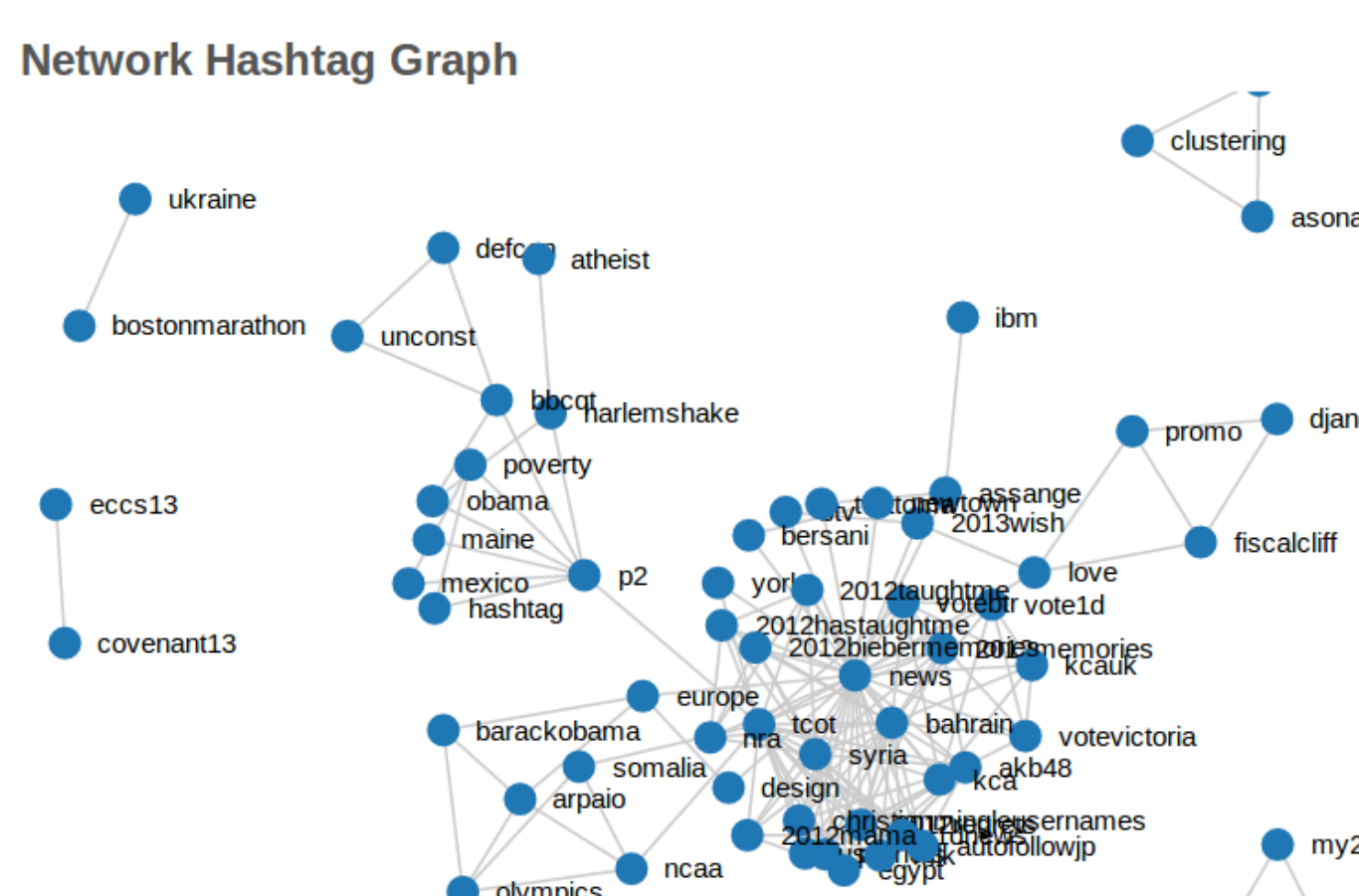
## VIII. Content POS tags



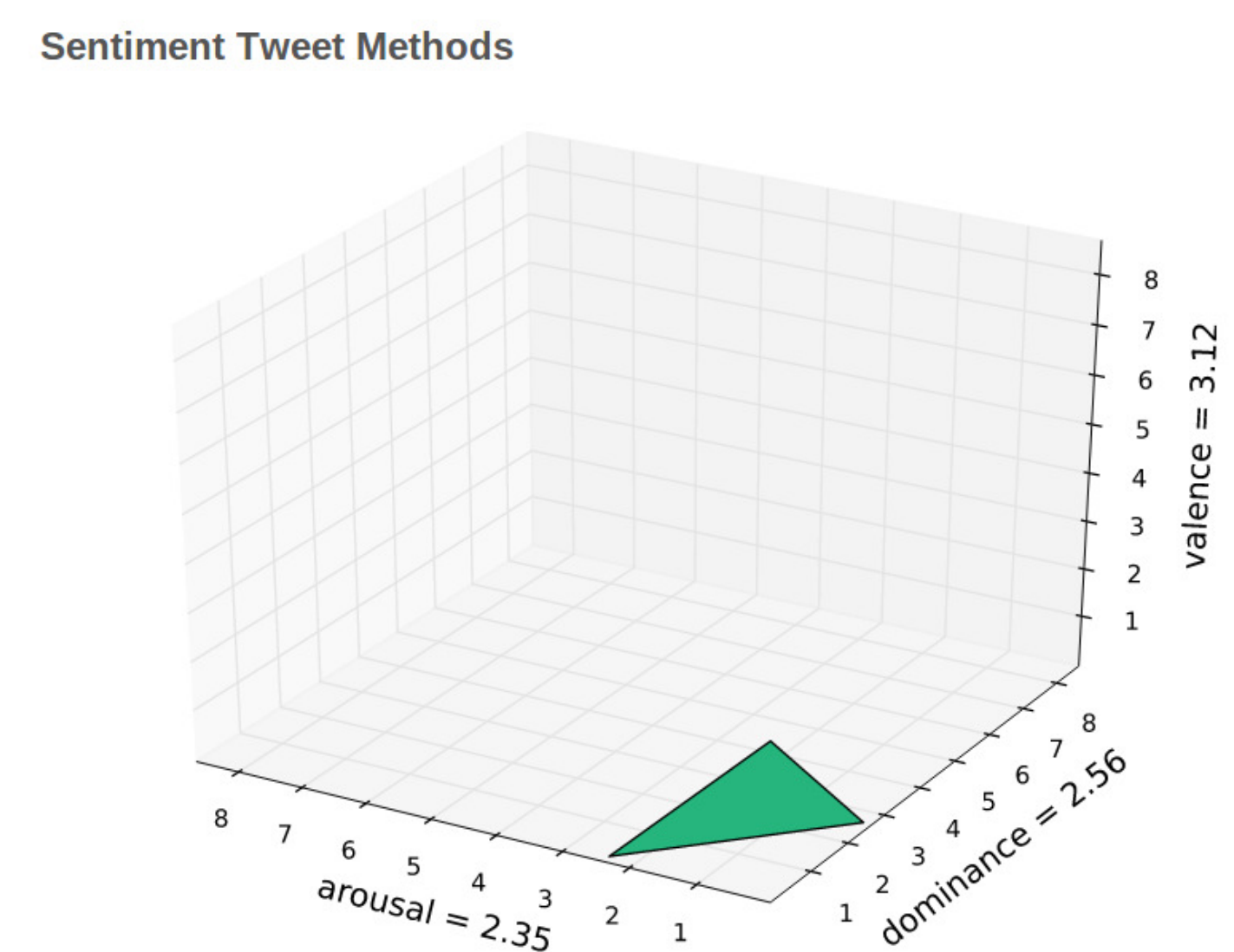
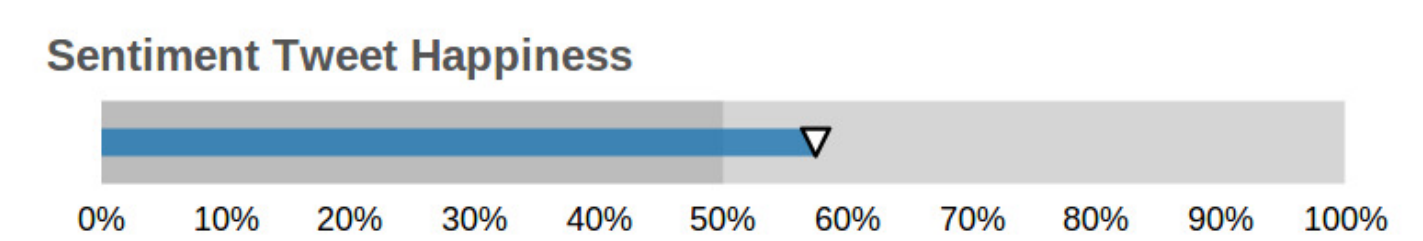
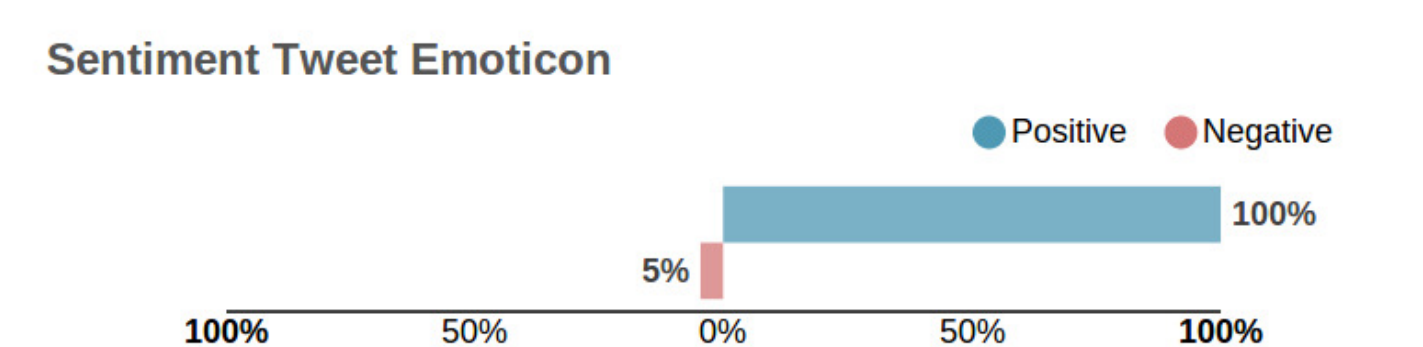
## IX. Contact languages



## X. Network



## XI. Sentiment Analysis



## XII. Temporal Signals

