RESEARCH

Studying Twitter User Accounts: Spotting Suspicious Social Bot Behavior

Marcelo J. Rovai

Correspondence: mrovair@udd.cl
Data Science Institute, Faculty of Engineering, Universidad del Desarrollo, Santiago, Chile
Full list of author information is available at the end of the article

Abstract

Using original tweets published during the first round of the 2017 Chilean presidential elections, this work aims to study the bot behavior of Twitter users by specific patterns retrieved from their tweets, such as the user's metadata, number of friends, followers, content, network, and time series. Each pattern is studied both individually and across different subsets of users, such as the number of tweets per account per day, newly created accounts, and so-called simple bots. Networking and timing related features proved to be critical in bot detection. Twitter users considered to "behave" like bots are compared with web applications (apps) used for bot detection. This work explores the visual analysis of groups of users with similar characteristics (clusters), suggesting that a bot behavior can be visually detected using dimensional reduction techniques such as Uniform Manifold Approximation and Projection (UMAP). The methodology used in this work can be applied to identify social bot behaviors in any set of tweets captured in a specific time frame.

Keywords: Social Bot; Twitter; Fake News; Cyborg; Botnet; Data Science; Visualization; UMAP

1 Introduction

In October 2019, massive protests took to the streets in Chile to protest against the highest inequality among OECD countries. Despite its legitimate demands, Chile's infrastructure had also suffered significant fires and destruction. Were these acts spontaneous? According to the U.S. State Department Acting Assistant Secretary Michael Kozak,"foreign actor" could have influenced protests in Chile once fake social media accounts pretending to be Chilean, was identified as emanating from Russia [1]. Are these true accusations? Could these fake accounts be called social bots? A year ago, in October 2018, Brazil elected as president Jair Bolsonaro who is a near-unknown and radical former army captain that served very poorly the Congress for nearly 30 years. Surprisingly, voters decided to migrate massively to the right side of the political spectrum after more than 24 years of center-left governments. Did social media and bots play a significant role in this presidential election [2]?

Many theories have emerged around the world, trying to explain these recent events both in Chile and Brazil. Furthermore, similar situations happened in Latin America (2017 Chilean presidential election [3], in Europe Brexit process [4]), and the U.S. (Trump election) with social bots being heavily used to spread low-credibility content and fake news [5].

Nowadays, we are observing a massive spread of digital fake news and misinformation, which constitutes a significant threat to modern democracies. This situation indicates that social bots play a disproportionate role in spreading articles from low-credibility sources on the so-called Social Media [6]. The phenomenon is, in fact, far broader than just social media. According to an independent survey conducted by Imperva [7] in 2016, bots drove nearly 51.8 % of internet traffic. In the 2018 survey [8], this number dropped to around 40 %, which may sound like good news; however, those bots are getting more sophisticated (nearly 75 % of bots are rated moderate or even sophisticated), what complicate its detection.

1.1 What are the Social Bots?

Social bots are bots found on social media as Twitter. But what is a bot? The term bot is not well defined, and different scientific disciplines (and scholars) diverge on a standard definition. In order to detect if an account is a social bot (or, a bot), two tasks known as automation and orchestration can be used. Automation and orchestration workflow are the common methods used in creating the account's content, as shown in Figure 1 [9]. The orchestration is related to networking, where users (bots or humans) interact among themselves to increase content's spread.

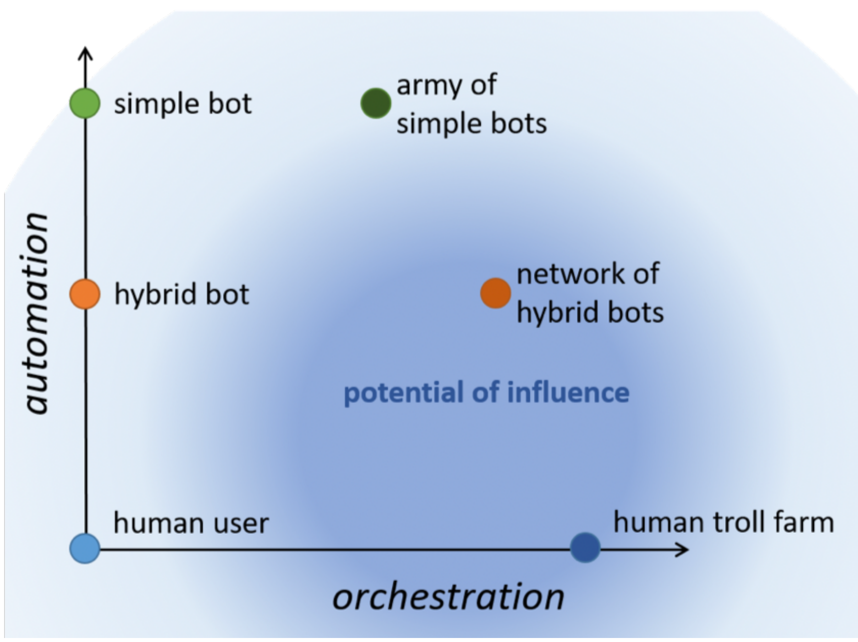

Figure 1: Social Bots - Automation and orchestration (GRIMME et al. 2017)

In terms of automation, a bot can be classified from a simple bot to a human user, going through a hybrid or cyborg bot (fully automated account mixing with a human user). Simple bots (generically called "bots") are automated user accounts that interact with the application through an API [10]. Regardless of its level of automation or orchestration, social bots can be programmed to perform tasks typically associated with human interaction, such as following other users, favoring tweets, sending direct messages, and, most importantly, posting content. Additionally, social bots often retweet anything published by a specific set of users or promoting a specific hashtag.

Many Twitter posts generated by bots can be used to perform essential functions, such as serving as part of an alert system which will post digital messages as seen in an earthquake event, those are the “good bots”. However, not all bot activities are good. in the case of a campaign (e.g. political or otherwise), bots are often used to generate high interest in specific content by spreading messages at speed not possible for human users [5] and here is where is possible to find the so called “bad bots.

Bad social bots usually share low-credibility content, having as a target users with many followers, and using mainly original tweets and retweets with rare replies. However, a single account can generate posts with the same content hundreds and thousands of times [6]. This work will concentrate on those “bad social bots”, that from now on will be called generically as “bots”.

Twitter is an important source for study because online interaction is part of day life of billions of people around the world, where social media users post their feelings, thoughts and opinions about every aspect of life [11]. This work aims to identify user accounts that could be potential bots based on their behavior. The data source of this study is an unstructured tweet dataset captured during the last Chilean presidential election (2017), the *Raw Tweets*.

1.2 Different types of Bots

Bots can be [12] [13]:

- Fake Followers
- Retweet Frauds
- Hashtag Promotion
- URL Spamming

To spot a simple bot is not complicated, once its typical characteristics are [6]:

- Relatively recent account creation date
- Account name containing numbers
- More Retweets than generating original content
- Very high frequency on posting tweets
- A high number of tweets since account’s start, but few followers
- Several different tweets with the same content
- Short replays
- Often No-Bio and No-Photo

According to [14], it is possible also to consider simple bots, accounts that have as characteristics:

- A high number of following accounts
- Tweet content with a high score in lexicon diversity (unique words per total words used)

And from [15]:

- A low number of mentions
- Few replays
- Username with long string of characters

In short, it is safe to say that meta-data extracted from tweets are considered to be among the most predictive and interpretive features. However, due to the low cost of creating a bot [9], its constant evolution, sophistication and mixed human techniques ("cyborgs"), it is currently impossible to detect all types of bots only by feature-based systems [15].

This work does not pretend to achieve a method or model to automatically spot a bot, but to study accounts from a set of selected features, in order to find users with a "bot behavior". One promising technique explored is the grouping of accounts into a behavioral cluster. Clusters of users with similar behavior, suggests that a bot can be spotted visually [16], and for that, this work explores a novelty technique using Uniform Manifold Approximation and Projection (UMAP).

2 Related Work

Several different approaches are found in the literature regarding bot detection. The most common approach is the use of classification models based on supervised machine learning, but not necessarily with good results [17][18].

It is possible to find websites where a user account can be tested. The most important and cited project is Botometer [19], an online tool to classify Twitter accounts as human or bot. Botometer is part of the OSoMe (Observatory on Social Media) [20], which is a joint project of the Network Science Institute (IUNI), the Center for Complex Networks and Systems Research (CNetS) at SICE, and the Media School at Indiana University. The base of Botometer is a Random Forest classification model, using more than 1,000 features extracted from tweets metadata, interaction patterns, and content. Features are grouped into six main classes: Network, User, Friends, Temporal, Content, and Sentiment.

A not well known, but an attractive model still under development, is TweetBotOrNot [21]. It uses a machine-learning algorithm (Gradient Boosted) that was trained on thousands of real bot and non-bot Twitter accounts. Downloading the last 100 tweets from users, TweetBotOrNot extracts over one hundred different features from user-level attributes (such as bio, location, number of followers and friends), tweets-level (such as number of hashtags and mentions) and, text-based patterns (such as number of hashtags, mentions, and links, length of tweets, punctuation and word complexity).

In Brazil, researchers from ITS Rio (Institute for Technology and Society of Rio de Janeiro) [22] and the Institute of Equity & Technology, developed PEGABOT [23][24], a platform where is possible to check the activity of a Twitter account to discover the probability of a profile being a bot. PEGABOT is not a machine learning model trained from real accounts, but a kind of score calculator based on user account features. The bot score is weighted according to the importance of the feature and also to the context. Features are divided among User, Friend, Network, and Temporal, being the last two, the more weighted due to their importance on bot spotting.

Among other models cited in the literature to spot more sophisticated bots, but not available for testing, are:

- BotWalk, an unsupervised near-real-time adaptive Twitter exploration algorithm [17]

- DNA Inspired Model, a strikingly novel, simple, and practical approach to model online user behavior, where researches extract and analyze digital (DNA) sequences from online user actions, using Twitter as a benchmark to test the model [25]

3 Data

The data used on this work (*Raw Tweets*) correspond to tweets posted during the 2017 Chilean presidential election [1]. The data covers 30 days (full November 2017) of tweets posted before and after the first round of 2017's elections (November 19, 2017).

3.1 *Tweet Dataset* General Exploration

The *Raw Tweet* dataset with a size of 7.5Gb, is a text file imported in a JSON format. Once cleaned and transformed, a *Tweet Dataset* is created as shown in Figure 2.

User	user_id	1601643 non-null	int64
	user	1601643 non-null	object
	name	1601576 non-null	object
	description	1601643 non-null	object
	verified	1601643 non-null	bool
	protected	1601643 non-null	bool
	location	1601643 non-null	object
	lang	1601643 non-null	object
	followers	1601643 non-null	int64
	following	1601643 non-null	int64
	favourites	1601643 non-null	int64
	lists	1601643 non-null	int64
	tweets_cnt	1601643 non-null	int64
	acc_creation	1601643 non-null	datetime64[ns, UTC]
	default_profile	1601643 non-null	bool
	default_prof_image	1601643 non-null	bool
	image	1601643 non-null	object
Tweet	text	1601643 non-null	object
	tweet_id	1601643 non-null	int64
	created_at	1601643 non-null	datetime64[ns, UTC]
	lenght	1601643 non-null	int64
	retweet_cnt	1601643 non-null	int64
	favorite_cnt	1601643 non-null	int64
	reply_to_twt_id	303655 non-null	float64
	reply_to_user	327627 non-null	object
	reply_to_user_id	327627 non-null	float64
	device	1601643 non-null	object
	RT	1601643 non-null	bool
	Reply	327627 non-null	object
	retweet_from	1601570 non-null	object
	num_hashtags	1601643 non-null	int64
	num_mentions	1601643 non-null	int64
	num_urls	1601643 non-null	int64
	clean_text	1601643 non-null	object
	day	1601643 non-null	int64

Figure 2: *Tweet Dataset* Features

[1]Data collected by Eduardo Graells-Garrido (UDD and BSC)

Each one of the individuals tweets has numerous information embedded in it (tweet object), that are a long list of 'root-level' attributes, such as id, created_at, and text. Tweet objects are also the "parent" object to several child objects. Tweet child objects include user, entities, and extended_entities [26].

Description and Location are information manually inserted by the user when Twitter's account is created. Usually, simple bots do not include that info once it is not mandatory, but if missed does not mean that the user is a bot. On *Tweet Dataset*, non-description or non-location (NaN data), were replaced by a null string ("").

Coordinates data found on *Raw Tweets* are a small number because usually, users do not provide information about where the tweet is generated. This info is not considered on the *Tweet Dataset*.

In general terms, the *Tweet Dataset* includes:

- Number of tweets: 1.6 million
- Number of unique user accounts: 91,500
- Average number of tweets per user: 17.5

The percentage of tweets that are:

- Retweets (RT): 31%
- Replies: 20%
- Plain Text: 26% (No RT or Mentions)

The percentage of tweets that includes:

- Hashtags (#): 28%
- Mentions (@): 74% (including RT)
- Mentions (@): 44% (not including RT)
- URLs (http or https): 37% (1/4 of them with two or more mentions)

Figure 3 shows the *Tweet Dataset* split in its main categories.

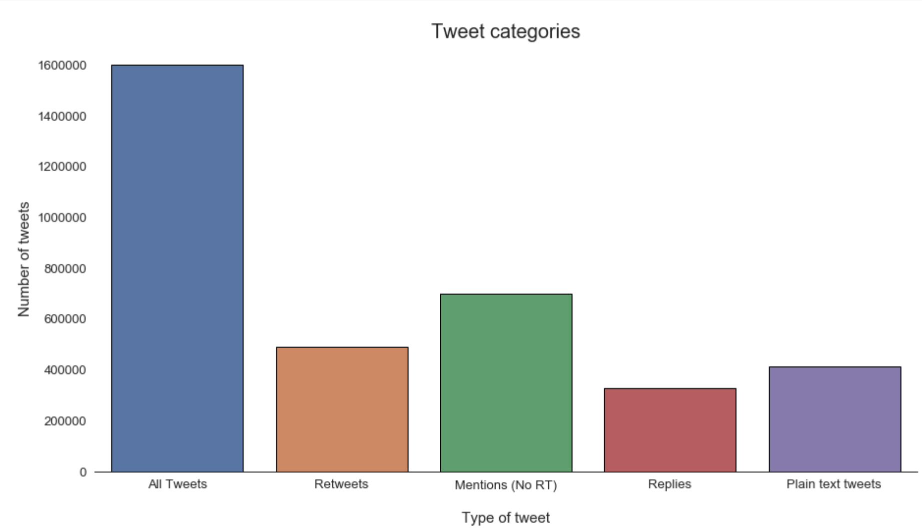

Figure 3: *Tweet Dataset* Summary Categories

It is possible to observe that mentioning an account (by its user screen-name, as @xxx) is a common practice on 75% of all tweets. This characteristic, when used by bots, increases the network and the effectivity of spreading fake news.

3.2 *Tweet Dataset* - Timeline and devices

Regarding its timeline, tweets spread from November 1 to November 30, 2017. November 19 is the day with more posts, which coincides with the first round of elections. Also, it is noted that the number of tweets increased significantly after November 19, as shown in Figure 4.

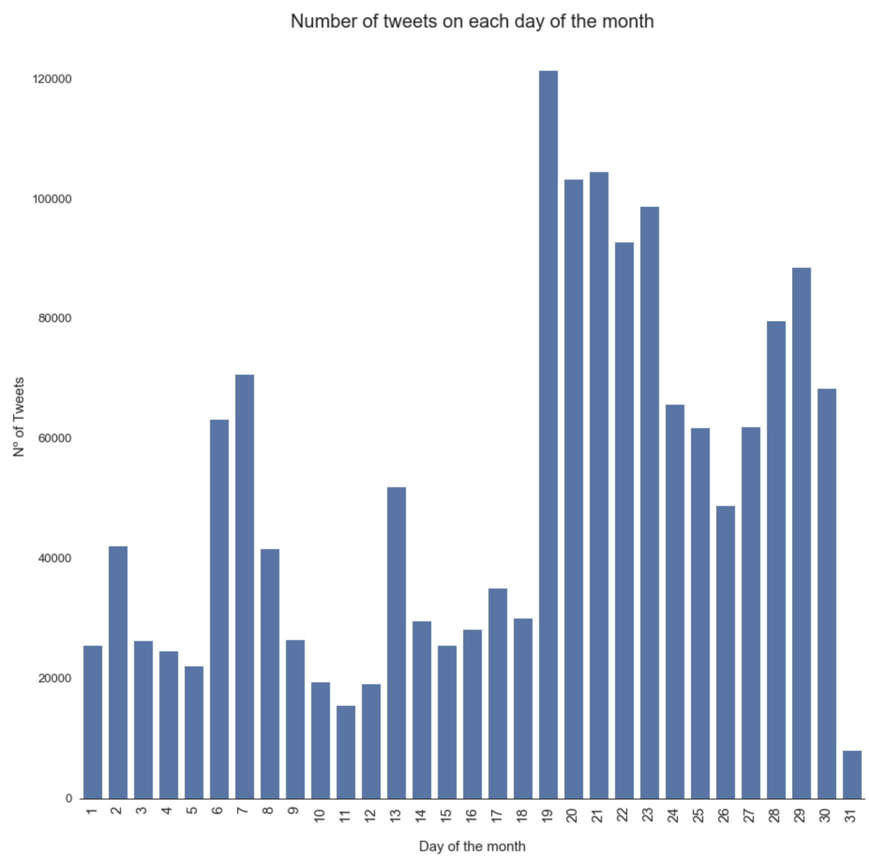

Figure 4: Tweets per Day

Figure 5 shows that around 1.1 million tweets (69% of total tweets) were generated by smartphones (Android or iPhone).

Also, some devices are classified officially by Twitter as "Bot" (1,330 tweets) from 89 unique user accounts. Looking deeper, only a few users generate more than 100 tweets during the month (4 users). One user that has the highest number of tweets (273) is @Ivonomas, a reporter that explicitly states on his profile that he retweets all tweets that mention him. Maybe this action could be done automatically, but it is not confirmed. Anyway this account is labeled officially as a bot by Twitter.

3.3 *Tweet Dataset* - Content

In terms of content, the tweets are heavily related with political subjects as shown in Figure 6. This is important because bots have more probability to be used on political events.

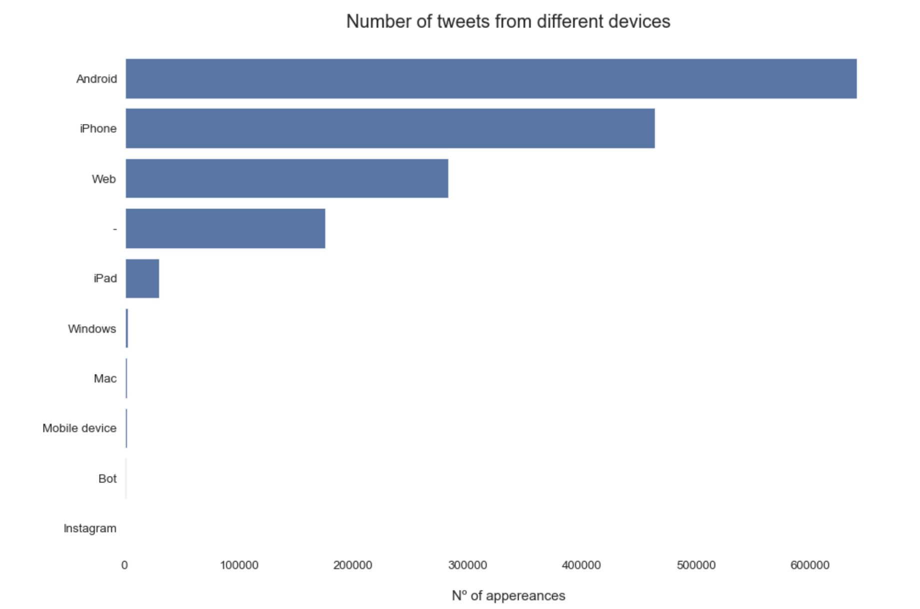

Figure 5: Tweets per Devices

Figure 6: *Tweet Dataset* Wordcloud

3.4 Deep diving on *Tweet Dataset* main features

Table 1 describes the essential numerical features of tweets through their primary statistical data.

3.4.1 Tweet Dataset - *Text Length*

Regarding text length, as expected, the average is 164 characters, being the median (50%) 139. Some tweets have more than 240 characters (15%), which seems rare, but this situation happens due to how Twitter applies its text limit (not exactly by the number of characters). For example, tweets with hundreds of characters were found, where dozens of mentions were posted with very few words. Even when mention to

	mean	std	min	25%	50%	75%	max
followers	124,937.8	472,966.1	0	292	1,623	11,466	2,889,231
following	15,217.4	76,771.9	0	342	1,091	3,448	761,106
favourites	11,644.4	29,737.5	0	399	2,203	9,316	897,789
lists	423.4	1,538.8	0	2	13	72	10,641
tweets_cnt	62,869.6	132,104.9	1	3,350	14,351	56,741	1,397,706
lenght	164.3	85.5	13	127	139	189	1,017
retweet_cnt	86.0	298.6	0	0	0	30	4,629
favorite_cnt	101.5	394.8	0	0	0	29	6,618
num_hashtags	0.4	0.9	0	0	0	1	28
num_mentions	2.2	4.8	0	0	1	2	54
num_urls	0.4	0.6	0	0	0	1	4

Table 1: *Tweet Dataset* Numeric Features

another account (@xxx) can be split into several characters, Twitter counts it more flexibly [27]. Bots could be who generate those rare tweets.

3.4.2 Followers / Following

Sometimes metrics as Followers and Following can bring confusion, where some authors also use “Friends”, that can be found associated with one or another entity. This work uses the official Twitter definition [28]. Below some examples of using Following and Follower:

User A following a User B means:

- User A is subscribing to User B tweets as a follower
- User B updates appear in User A’s home timeline
- User B can send User A Direct Messages

User A Followers are all users who receive User A tweets. If someone follows User A:

- They show up in User A followers list
- They see User A tweets in their home timeline whenever they log in to Twitter
- User A can start a private conversation with them

Table 1 shows that, on average, users are following fewer accounts (15,000) than are follow by others or be followers (125,000), what at first glance seems strange, because regular and ordinary people usually follow more people such as celebrities and politicians than are followed. This distortion can be explained by heavy users and also for bots that try to acquire as many followers is possible to gain popularity/influence and so evade detection by Twitter’s defense [29]. When looking for specific accounts, the relation between those metrics is essential to spot a bot.

It is not possible to reach any conclusion about a user be a bot or not, only looking at the data from a tweet point of view because every tweet repeats its user metadata. Metrics other than text length should be acquired from a user point of view (*User Behavior Dataset*).

4 Methodology

As explained in the *Tweet Dataset* General Exploration section, the *Tweet Dataset* is a huge file, where each row is a single tweet. From this dataset, a *User Behavior*

Dataset is constructed where each row is a single user account, with its associated features. This *User Behavior Dataset* is used to analyze accounts, spotting Social Bot behavior as shown in Figure 7

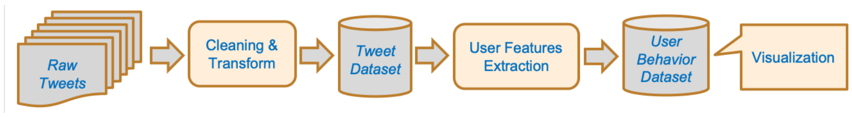

Figure 7: From *Raw Tweets* to *User Behavior Dataset*

The numeric features extracted from *User Behavior Dataset* are studied individually and compared in different subsets of data, as high frequent tweet user versus low-frequency tweet users and different clusters obtained from applying Uniform Manifold Approximation and Projection (UMAP) technics. Clusters of users with similar behavior are spot visually using UMAP [30], which is a dimensional reduction technique that aims to preserve the essential high-dimensional structure and present it in a low-dimensional representation. The resultant analysis can be used to spot social bots' behavior from user accounts.

4.1 *User Behavior Dataset*

Feature	mean	std	min	25%	50%	75%	max
default_prof_image	0.1	0.3	0	0.0	0.0	0.0	1.0
acc_verif	0.0	0.0	0	0.0	0.0	0.0	1.0
local_bol	0.6	0.5	0	0.0	1.0	1.0	1.0
default_profile	0.5	0.5	0	0.0	1.0	1.0	1.0
descr_bol	0.7	0.5	0	0.0	1.0	1.0	1.0
active_acc	0.9	0.3	0	1.0	1.0	1.0	1.0
tws_cnt	9,198.7	28,376.3	1	231.0	1,406.0	6,627.5	1,386,920.0
frs_cnt	753.6	4,236.8	0	106.0	302.0	775.0	761,106.0
fols_cnt	1,599.4	29,394.8	0	33.0	141.0	499.8	2,882,527.0
fav_cnt	2,315.7	8,656.2	0	55.0	337.0	1,556.8	897,789.0
tweets_analysed	16.4	165.0	1	1.0	2.0	7.0	25,900.0
account_age_days	1,889.8	1,058.6	0	978.0	2,241.0	2,819.0	4,065.0
ave_acc_tw_day	5.0	22.4	0	0.0	1.0	3.0	1,246.0
ave_recent_tw_day	1.7	12.9	0	0.0	1.0	1.0	1,669.0
max_tweet_day	5.8	53.8	0	1.0	1.0	3.0	8,044.0
ment_tweets_cnt	12.4	96.0	0	1.0	2.0	6.0	15,603.0
hash_tweets_cnt	4.4	53.6	0	0.0	0.0	1.0	7,140.0
unique_mentions_cnt	10.2	28.6	0	1.0	3.0	8.0	1,342.0
unique_hashtags_cnt	1.7	5.5	0	0.0	0.0	1.0	220.0
ment_idx	1.6	1.5	0	1.0	1.2	2.0	49.0
hash_idx	0.5	0.7	0	0.0	0.0	1.0	16.0
rt_ratio	48.5	44.4	0	0.0	50.0	100.0	100.0
mean_urls	0.2	0.4	0	0.0	0.0	0.3	3.0
mean_lenght	134.9	45.4	14	116.0	132.0	140.0	987.0
recent_age_days	8.2	9.6	1	1.0	1.0	16.0	30.0
user_name_len	11.1	2.7	3	9.0	11.0	13.0	15.0
name_len	13.2	4.8	0	10.0	14.0	17.0	50.0
user_name_len_num	0.8	1.8	0	0.0	0.0	1.0	14.0
simil_name	63.1	27.6	0	41.7	63.6	88.9	100.0
descr_len	58.5	56.2	0	0.0	46.0	108.0	179.0
fols_frs_ratio	3.1	120.4	0	0.2	0.5	0.9	22,797.0

Table 2: *User Behavior Dataset* Features - Statistical data

User behavior features as listed on first column of Table 2 are extracted from tweets. Initially, the tweets are grouped by user, and several metrics are generated from the grouped tweets of each user account. As shown in Figure 2, each individual tweet also includes user-specific metrics that provide relevant information about the user at the very moment when the tweet was posted. Information such as how many followers the user had, how many likes, his most current profile photo, description, and number of tweets. So, for each user, the most updated account information is retrieved from the last tweet posted. For example, for a user that has posted during all November 2017 (from day 1 to 30), user-specific metrics represent his situation on November 30.

New metrics are calculated (feature engineering) from the user's grouped tweets such as maximum tweets per day, average tweets per month, number of unique hashtags, mention rate, and retweet rate.

For each user, it is verified if the account is still active in 2019. This test is used to verify if by any chance Twitter eliminated the account in the present day. If one account has tweets included on *Raw Tweet* but today is not a valid Twitter's account, this account could be a bot.

Table 2 describes the essential numerical features of *User Behavior Dataset* through their primary statistical data.

4.2 *User Behavior Dataset* Feature Description and Analysis

In this section, each feature is described and analyzed individually. Having a single feature considered as a bot behavior does not mean that an account is a bot, only that it can be considered "suspicious". As many bot behavior features have an account more probability of being a bot.

default_prof_image - (Boolean variable): where 1 means that a Twitter's default user image was used (sometimes known as "Egg"). Using Default Image was common with simple bots. Mean = 0.1 means that the vast majority of the accounts do not use the default image.

acc_verif - (Boolean variable): If 1, means that Twitter has checked and considered the user, a legitimate one. Twitter usually verifies accounts that belong to celebrities and famous personalities who might be subject to identity theft on social networks. A verified account could show a bot-like behavior, once is not uncommon such accounts as such celebrities or politicians to post a high number of tweets per day [31]. When spotting bad social bots, verified accounts can be put aside. The vast majority of users are not verified.

local_bol - (Boolean variable): If 1, means that the user introduced "some text" as location. More than 50% of accounts have a "location", however, it does not mean that they are a valid location. Simple bots usually does not insert a location [6], but legit users could also not include a valid location, once it is an optional information.

default_profile - (Boolean variable): where 1 means that a Twitter's default user profile was used. Half of the account uses a default profile.

descr_bol - (Boolean variable): If 1, means that the user introduced "some text" as description. The majority of accounts has some text as description. Simple bots usually are created without description [6], but legit users could also not include a description, once it is an optional information.

descr_len: Number of characters found in the account description. Accounts with short or even no description can be suspicious to be a simple bot. Of course, a lack of description does not mean that the account is a bot but should be investigated.

active_acc - (Boolean variable): If 1, means that the account is currently active and not terminated by Twitter. Only a few active accounts in 2017 was confirmed not active ("0") in 2019 and if were posting a high frequency of tweets back in 2017, they are candidate to be a bot. If the account still exists but is blocked, this feature is set to "1".

account_age_days: Age in days, starting when the account was created until the last tweet posted by the user. Looking at Figure 8 is possible to observe two very marked group of accounts. One aged around 90 days and another with a little less than 3,000 days. Those accounts with less than 90 days should be analyzed because simple bots usually are created shortly before being used [32].

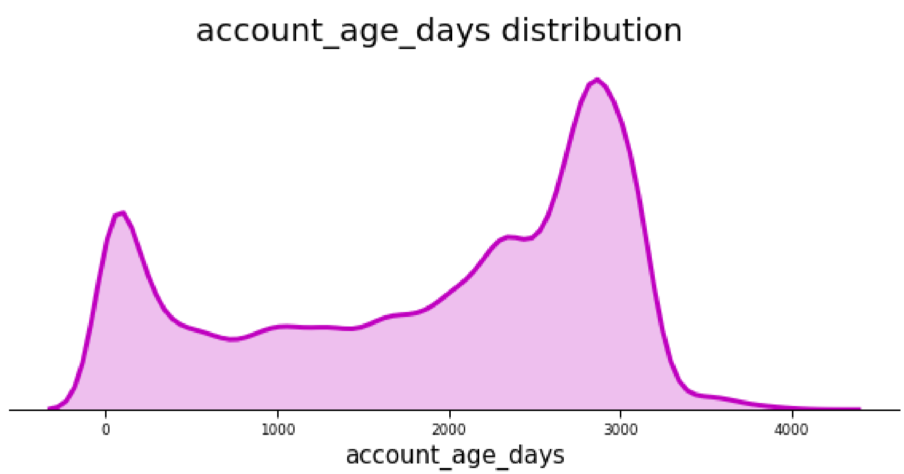

Figure 8: Account Age in days

tws_cnt: Total number of tweets post by user since the account was created. User accounts on *User Behavior Dataset* are, on average, 5 years old (see "account_age_days"), which leads to a high number of tweets (on average, 9,200 tweets).

frs_cnt: Number of Following users. 50% of accounts follow less than 300 users, 75% with less than 700, which is very common for normal humans' accounts. However, there are accounts following hundreds of thousands of users. Those accounts should be analyzed with care because smart bots try to follow as many accounts is possible to increase network and avoid detection [17].

fols_cnt: Number of Followers that follow this account. Human accounts tend to follow more accounts, having fewer followers. On *User Behavior Dataset*, 50% of accounts have less than 140 followers (or almost half of the accounts that they are following). Note that on average, followers are higher than following, this anomaly is because some outliers have a considerable number of followers. It could be celebrities (if the account is verified) or could be a bot [31].

fols_frs_ratio: Shows the relation between the number of followers that follow this account, divided by the following users. Humans usually have this ratio, low (more following accounts than followers) as can be verified with 75% of users that have this feature less than 0.9, but some outliers have huge ratios, what "push" the average to more than 3. Skewed ratios is one of the ways to identify suspicious accounts [29].

fav_cnt: Number of times that a user favorites a tweet (gives a like). 75% of accounts have around 1,600 likes and 2,800 days of existence, which results in less than 2 likes a day, which seems reasonable. However, some accounts reach more than 900,000 likes, and no account has more than 4,100 days in age, which means that those users give hundreds of likes daily, which is very unlikely for humans. This situation is suspicious and should be analyzed together with other suspicious behavior.

tweets_analysed: This feature captures the number of tweets posted by a user during the *User Behavior Dataset* timeline window.

max_tweet_day: Maximum number of tweets posted on a single day during November 2017. 75% of users posted less than three tweets /day, but around 356 non-verified users posted, on a single day, more than 144 tweets. Those accounts are responsible for almost 25% of all tweets during November 2017. This feature is one of the essential metrics to spot a bot [33] [32].

rt_ratio: Ratio in percentage between the number of tweets that are retweets and the total number of tweets during the period. Looking at the statistics and based on this feature's mean, it seems that users usually retweet 50% of the times, but this is not true. Looking in more detail, Figure 9 shows that there are two groups of very distinct users, the first with low rt_ratio, post new tweets, and the second around 100%, shows users that do not create new content and only retweet an existent tweet. Usually, simple bots tend to retweet [32].

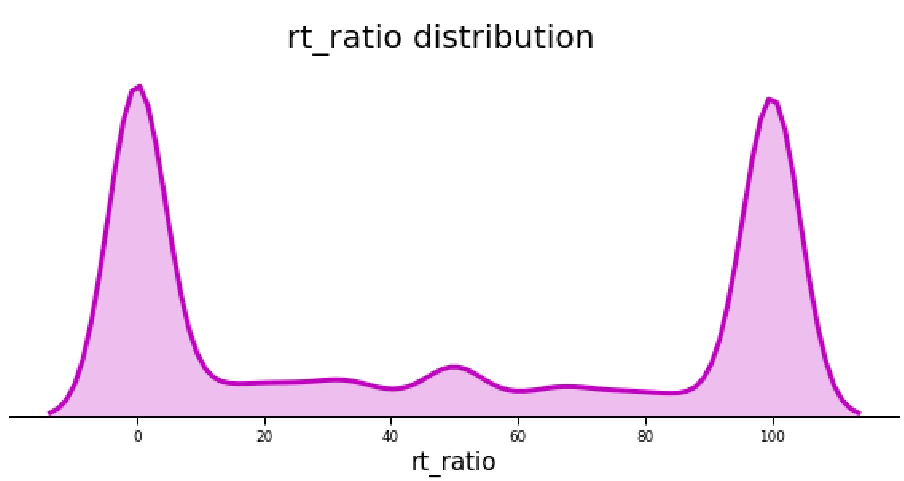

Figure 9: Retweet Ratio

ave_acc_tw_day: Average number of tweets per day. Take total tweets posted by user since the account was created (tws_cnt) divided by account age in days (account_age_days). 75% of accounts post less than three tweets a day. The global average in the *User Behavior Dataset* is 5. It is because there are accounts that post a massive number of tweets a day (more than 1,000). Those accounts are most probably automated (bots) [5] [32].

recent_age_days: Number of days between the last tweet posted and the oldest ones in the same *User Behavior Dataset* timeline window. This feature is used to calculate the ave_recent_tw_day feature.

ave_recent_tw_day: Average number of tweets per day, posted during the *User Behavior Dataset* timeline window. The calculation takes the number of tweets posted by user and divides by the number of days between first and last tweet posted during same timeline window. 75% of accounts post one tweet a day or none. The average in the *User Behavior Dataset* is 1.7 day, also because are accounts that post a massive number of tweets a day (more than 1,600). Those accounts are most probably automated (bots) [5] [32].

ment_tweets_cnt: Number of times that another user (via its screen-name, "@xxxx") is mentioned on a tweet. Half of the tweets include two or fewer mentions, which is typical for humans, but exceptions, can reach thousands of mentions and must be analyzed (author's hypothesis).

hash_tweets_cnt: Number of hashtags included on a tweet. 75% of users add at most one hashtag when posting. Exceptions of hundreds, even thousands of hashtags, are found. This anomaly can be related to automatized accounts (author's hypothesis).

unique_mentions_cnt: Total number of unique user screen-names (mentions) found on total tweets analyzed for that particular user at the timeline window. 75% of all users mention fewer than eight accounts.

unique_hashtags_cnt: Total number of unique hashtags found on total tweets analyzed for that particular user at the timeline window.

ment_idx: Total number of unique mentions (unique_mentions_cnt) divided by the total number of mentions (ment_tweets_cnt). Humans, usually vary subject and mention different users on their posts. Low index means that a unique user is mention several times, usually in different tweets and could indicate a bot that want to increase content's spread ("amplification") [33].

hash_idx: The total number of unique hashtags (unique_hashtags_cnt) divided by the total number of hashtags (hash_tweets_cnt). Humans, has a normal distribution of content, posting different subjects, with diverse hashtags. Low index means that a unique hashtag is used several times, usually in different tweets and could indicate a bot that want to increase content's spread ("amplification") [33].

mean_urls: The average number of URLs presents on each tweet. The majority of users do not include URLs on tweets. There is one suspicious group of users that includes URLs on all tweets and should be better analyzed (author's hypothesis).

mean_lenght: Tweet's average number of characters. The average number of characters goes around 135 characters, which makes sense due to Twitter's historical rules. The surprise is the high number of outliers with hundreds of characters, that is due tweets with a large number of hashtags and mentions. This behavior is suspicious and should be better analyzed (author's hypothesis).

user_name_len: Number of characters present on screen-name. This feature varies from 3 to 15 characters. The average is 11 characters and human users tend a keep it short and similar to their full name [33].

name_len: Number of Characters present on the name that the user chooses to be identified. 75% of accounts present less than 17 characters, but some users have big names. Sometimes is due to the field for the name is used to include erroneously, descriptions.

user_name_len_num: Screen-name's numeric characters. This feature shows how many of Screen-name's characters are numeric. Table 2 shows that 75% of users have one or none numeric characters on screen-name. Two or Four numeric characters are not uncommon, where age or year of birth can be added to scree-name by human users, but when several numbers are added, the screen-name became suspicious to be a bot [33].

simil_name: Shows the similarity between name and user name (screen-name), calculated by having Jaccard similarity between the set of letters of both features. A human user tends to have his screen-name (user_name) similar to his name and with few (or none) numeric characters [33]. For example, with a name: Marcelo Rovai and a screen-name: @mjrovai, the Jaccard Similarity is 0.6, which is OK. Low values of Jaccard Similarity, like 0 or 10%, could be suspicious and some users show this behavior as shown in Figure 10.

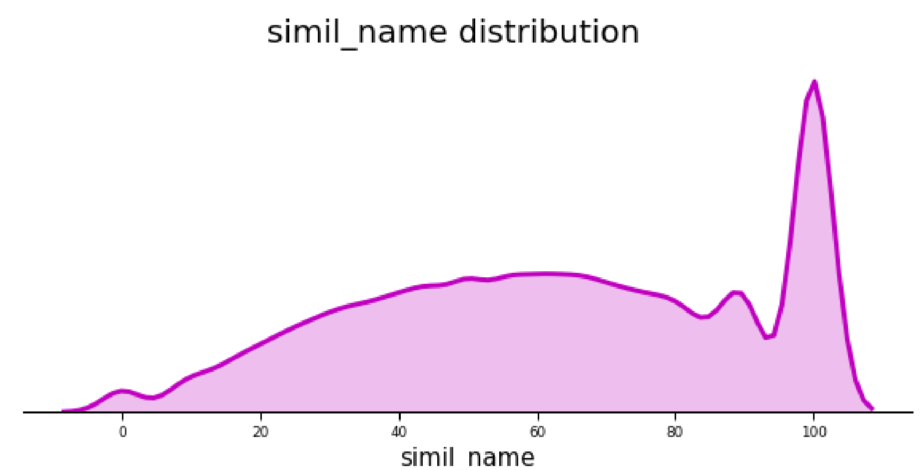

Figure 10: Jaccard Similarity Name and User-Name

5 Account study by frequency of publication

One critical metrics, when analyzing bot's behavior, is the frequency of tweets. By an academic study developed during 2016 U.S. Presidential election, accounts that post more than 50 tweets a day with the same specific hashtag is defined as having a high level of automation [5]. For DFRLab (Digital Forensic Research Lab) [33], accounts that post 72 tweets per day (one every ten minutes for twelve hours at a stretch) is suspicious to be a bot and over 144 tweets per day as highly suspicious. First Draft, a global non-profit organization that "supports journalists, academics and technologists working to address challenges relating to trust and truth in the digital age", considers a minimum of 100 tweets a day as a general rule for flagging an account as suspicious of automation [32].

The *User Behavior Dataset* has a total of 91,000 distinct accounts, being that around 780 of them posted more than 72 tweets on a single day. This number of accounts is less than 0.09% of all accounts that posted at least one tweet during the timeline window, but they are responsible for almost 42% of total tweets.

5.1 Filtering High and Low-Frequency Accounts

Selecting only the users that according to DFRLab are highly suspicious of being a bot (maximum number of tweets a day higher than 144), and also, not considering the verified accounts where celebrities could also behave as bots [31], the number of suspicious accounts is reduced to 356 users. However, this small number of suspicious accounts is responsible for posting an impressive number of 364,942 tweets; around 25% of the total tweets posted during the analyzed period (November 2017). Let us call this group of accounts, high-frequency users.

On the other hand, filtering accounts with shallow activity (less than 50 tweets/day), we get almost 90,000 accounts, that are responsible for 775,000 tweets (around 50% of total). Let us call this group, low-frequency users.

From total accounts, around 300 accounts (responsible for around 25% of total tweets) are considered “Gray Area”, let us call them mid-frequency users.

Looking at its distribution, 75% of high-frequency users, posted as a maximum, between 145 and 470 tweets and the low-frequency users, less than three tweets a day.

When exploring the *User Behavior Dataset*, some features presented a strange distribution such as desc_len , account_age_days, rt_ratio, recent_age_days, which can be better explained depending on how frequent a user posts.

desc_len: Low-frequency accounts description length shows a distinct group of users where short (or even NoN) descriptions are used. Instead, high-frequency users have a concentration in longer descriptions, as seen in Figure 11. The behavior shown is a surprise, once bots are expected to be more spotted among high-frequency accounts.

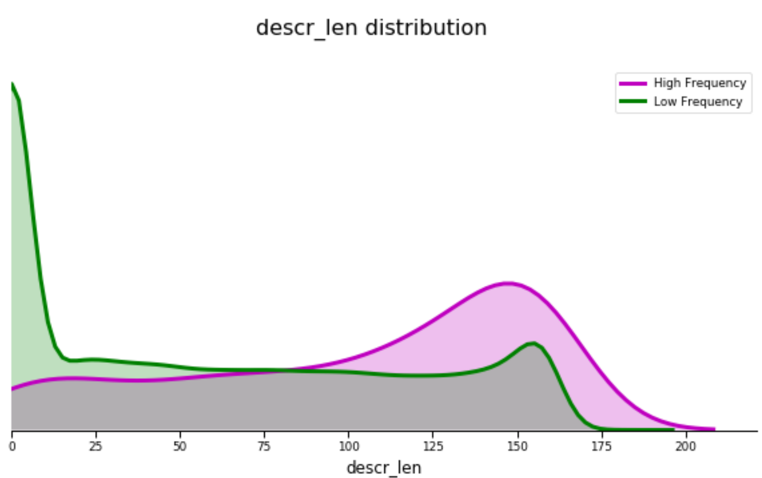

Figure 11: Account Description text length

account_age_days: When splitting into low and high-frequency users, it is possible to realize that low-frequency users have younger accounts, as shown in Figure 12.

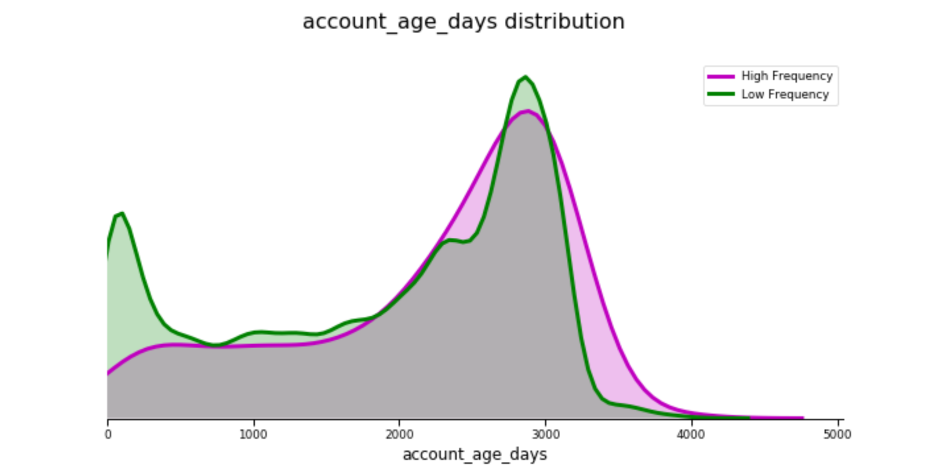

Figure 12: Account Age in days

recent_age_days: This feature shows that high-frequency accounts usually posted all month (Figure 13).

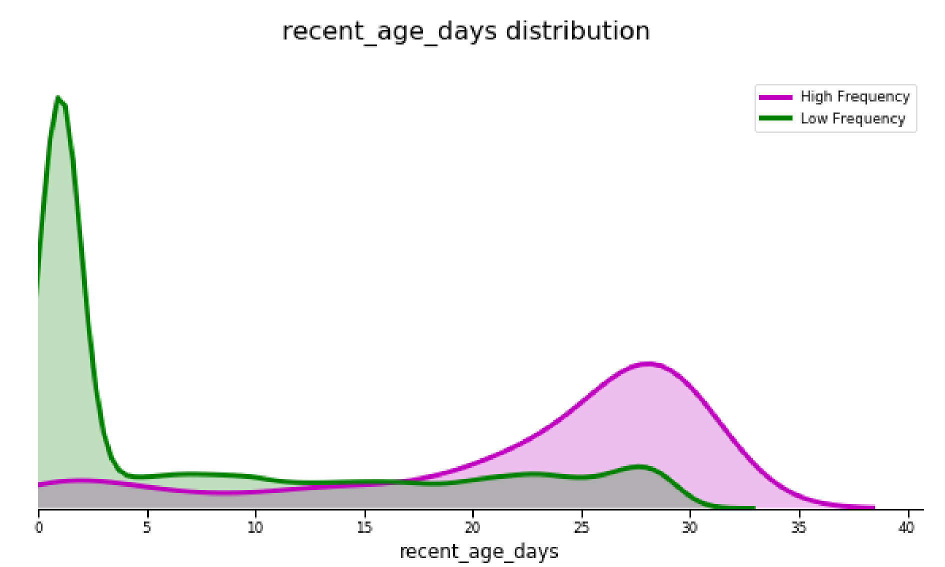

Figure 13: Recent Age Days

rt_ratio: Analyzing Figure 14 is possible to see that low-frequency accounts present two distinct groups of ratios, low and high. However, high-frequency users retweet less. This feature is calculated, dividing rt_tweets_cnt per tweets_analyzed and multiplying the result by 100.

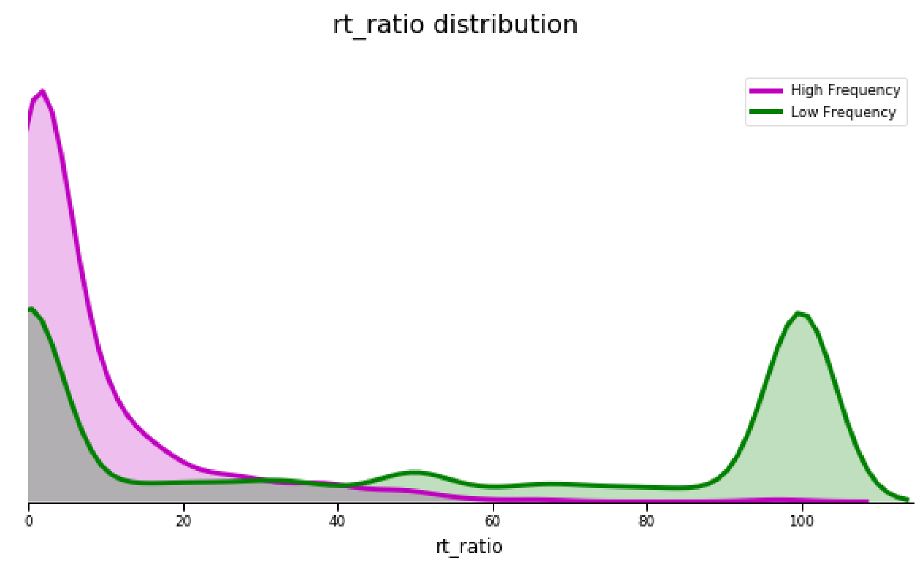

Figure 14: Retweet Ratio

6 Possible approaches to spot suspicious bot behavior

Once account metrics are collected in the *User Behavior Dataset*, the first natural approach is to develop an algorithm (or model) to spot automatically if an user presents a bot behavior. This approach is precisely the most common line of research found among academics, but while many of the metrics overlap, no two models are the same [34].

Botometer [35] and TweetBotOrNot [36] try, based on supervised machine learning models, to calculate automatically the probability of an account to be a bot or not, while PEGABOT [24] gives to each feature a score and, based on the sum of them, to define if an account is a bot or not. From the tests done on this work, PEGABOT did a better job, confirming how difficult it is to spot a bot using machine learning algorithms, most probably due to the labeled bot accounts used for model training, usually older than the most sophisticated bots found nowadays.

To spot a bot is, in fact, a hard task due to the combination of human and machine behavior found on more sophisticated approaches used nowadays on bot's creation.

None of those features, if taken individually, is enough to conclusively define whether an account is a bot or not. For example, could be perfectly reasonable for some social media users, such as celebrities or politicians post hundreds of tweets a day or to spot a lack of photo or description on a new Twitter user.

On the other hand, the more suspicious characteristics an account displays, the more likely it is to be automated and so, be a bot, cyborg or be part of an orchestration as a botnet. Even if a particular user shows multiples features that could spot him or her as a bot, this conclusion should be confirmed manually [32].

There is no right way to select the features that combined will spot a bot. The first attempt to spot a bot is by selecting a group of critical features and to filter suspicious users that share them. The confirmation should be done manually, account by account looking at the full set of features as a simple dashboard, for each individual account as shown in 9.

For this work three groups of accounts that share suspicious features are analyzed more deeply. They are:

- Simple Bots (SB)
- High-Frequency users (HF)
- Younger Mid-Frequency accounts (YMF)

6.1 Spotting Simple Bots

Classical simple bot behavior can be defined as:

- Account not verified
- Image Profile is the default
- Account description is missing
- 72 tweets posted in at least one day during the period

Filtering the *User Behavior Dataset* for users that math those parameters, three accounts were spotted:

- @fedoraletelier (full data in Appendix A)
- @Aliciacarafipl3 (full data in Appendix B)
- @Dolores09072598 (full data in Appendix C)

Each one of those users are analysed in more detail.

6.1.1 @fedoraletelier

Looking at the full data from @fedoraletelier in detail, it is possible to realize that besides the standard features used for selection, this user is most probably a social bot due:

- Several tweets with the precisely same content.
- A high number of unique mentions
- A high number of historical tweets (161/day)
- A high number of likes (120/day)

Figure 15 shows its tweets and Figure 16 its actual page, where it is possible to realize that this account posted around 100,000 tweets/year from 2017 to 2019, increasing the number of both followers and friends significantly. For an almost "anonymous" account, those numbers make no sense. Once this account is still active is also possible to verify that Botometer [35] classifies it more like a human, with a 30% possibility of being a bot (or 1.5/5, as shown in Figure 17), what is far from reality (Note that Temporal: 2.3 and Network: 3.0 are the highest individual scores).

For PEGABOT [24], the probability goes to 100%, confirming that this account behaves as a social bot, as shown in Figure 17. Note that User (Usuario: 100%), Frequency (Frequência:178%), and Network (Rede: 83%) are highlighted due to its high values. The bot behavior is also confirmed with TweetBotOrNot [36], with 99.3% probability, as shown in Figure 19.

6.1.2 @Aliciacarafipl3

In the case of @Aliciacarafipl3, it is possible to highlight:

- Recent average tweets per day are not high (26/day), but historically was (69/day)

- Timeline.png

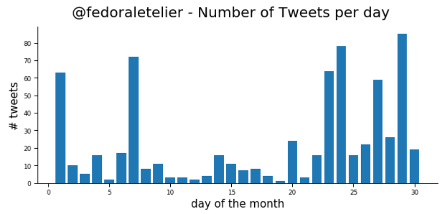

Figure 15: @fedoraletelier tweets/day

Figure 16: @fedoraletelier Twitter home page

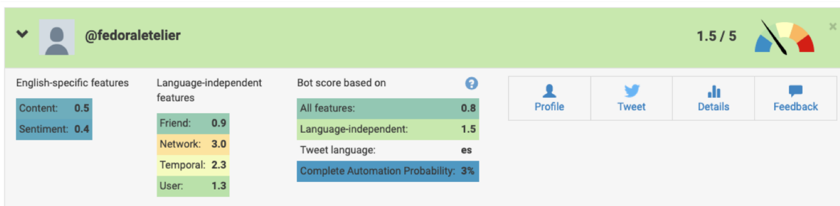

Figure 17: @fedoraletelier Botometer

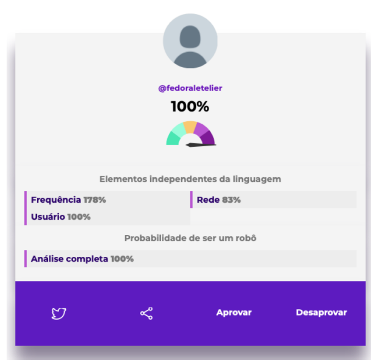

Figure 18: @fedoraletelier PEGABOT

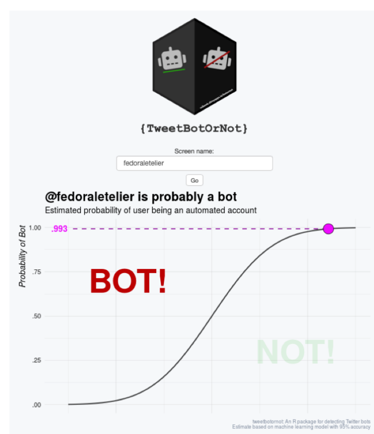

Figure 19: @fedoraletelier TweetBotOrNot

- Retweet Rate is very high (96%), which means that almost all posts are retweets
- 36 likes /day
- All tweets mention another account
- tweets repeated (similar text)

This account is still active, and it is possible to observe that in 2 years, the number of tweets jumped from 25,000 to 204,000, having the number of followers and friends also grown exponentially. So, on a simple analysis, this account most probably is a bot. This conclusion is also shared by PEGABOT (Figure 20) and TwitterBotOrNot (Figure 21), with respectively 100% and 99.7% probability of being a bot. Again, Botometer (Figure 22) does not consider this account a bot, indicating only a 25% probability score (1.3/5).

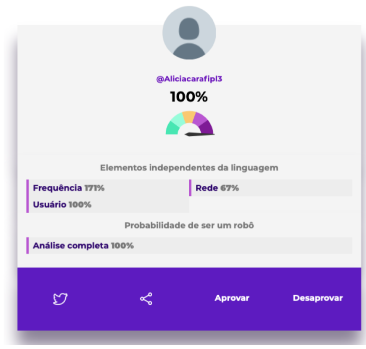

Figure 20: @Aliciacarafipl3 PEGABOT

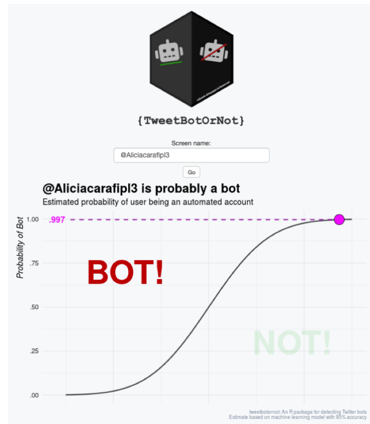

Figure 21: @Aliciacarafipl3 TweetBorOrNot

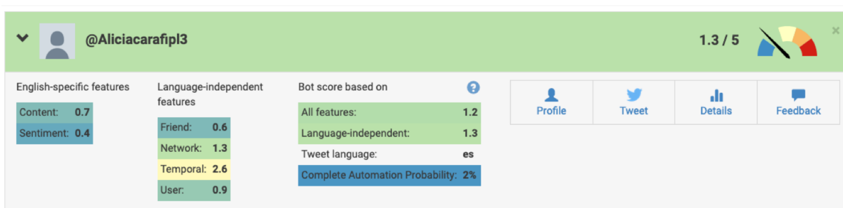

Figure 22: @Aliciacarafipl3 Botometer

6.1.3 @Dolores09072598

- Recent average tweets per day are not high (113/day)
- Followers/Following_index very low (Following 100 per each follower)
- 100% of tweets are pure retweets
- All tweets mention another account

This account is still active but protected. In the last two years, this account posted 37,000 tweets, increasing by five times the number of friends, but not followers, keeping Followers/Following_index almost the same (from 0.09 to 0.12). Once it is protected, web applications can not be used to measure its probability of being a bot.

6.1.4 SB - Conclusion

Accounts that are not verified and neither present an image profile or a description, but still, post a relatively high volume of tweets on a single day, most probably are bots. Also, it is possible to confirm that those accounts are Following many users and have a high number of Followers. Those simple analyses confirm that such kind of bot behavior (SB) is easy to spot. Still, only a very few users are spotted because it is also easy (and cheap) to create more sophisticated bots to avoid detection.

6.2 Spotting High-Frequency Users

Another manual attempt to select possible Social Bots is filtering not-verified high-frequency users and proceeding with a manual analysis as done with simple bots.

Three users with very high frequency of daily posts are analysed in more detail. They are:

- @AlbertoMayol (full data in Appendix D)
- @andres20ad (full data in Appendix F)
- @NelsonCL28 (full data in Appendix G)

6.2.1 @AlbertoMayol

This is a cyborg account. The user is human, but its behavior sometimes is like a bot (mechanized), as we can see on his post's timeline during November 2017 (Figure 23). He has the highest number of tweets on a single day among all users (5,163), posting only two days in the period with a total of 6,930 posts. On each one of the days, a unique text content was posted. His recent activity at Twitter is very reasonable; what confirms that account is a cyborg. The web bot apps that only analyse recent activity, do not consider this account as a bot, being the scores: Botometer: 3%; PEGABOT: 30% and TweetBotOrNot: 10%.

6.2.2 @andres20ad

In Figure 24, it is shown the timeline of @andres20ad, another active account that most certainly is handled by an human that sometimes has a bot behavior (cyborg). On 2017 November, 18, one day previous second round, this account posted 1,600 tweets, being the content split in only two different texts, as shown in Figure 25.

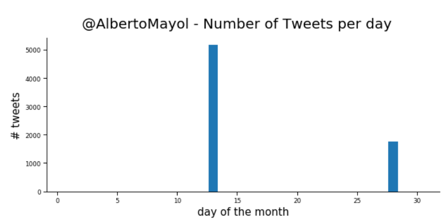

Figure 23: @AlbertoMayol tweets/day

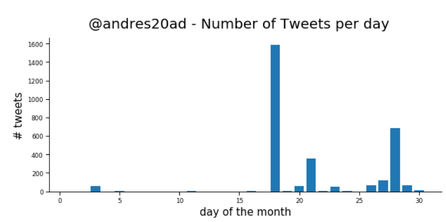

Figure 24: @andres20ad tweets/day

```
579503     En 1988 millones de chilenos se enfrentaron a ...
579527     En 1988 millones de chilenos se enfrentaron a ...
579556     En 1988 millones de chilenos se enfrentaron a ...
579579     El gobierno de Piñera fue el más corrupto de t...
579593     El gobierno de Piñera fue el más corrupto de t...
579597     El gobierno de Piñera fue el más corrupto de t...
579627     El gobierno de Piñera fue el más corrupto de t...
```

Figure 25: @andres20ad tweets on 11/19/17

6.2.3 @NelsonCL28

Figure 26 shows the timeline of another user with more than 1,600 on a single day. This account is still valid today, but its last post was this single one on November 19, 2017, day of Chilean Presidential election's 1st round. This account has very few followers (26), but this single tweet was retweeted 1,900 times and liked by

567 other users, which suggests that it must probably as part of a botnet of Fake-News. Figure 27 shows the actual post. Again, the web bot apps do not consider this account as a bot, being the scores: Botometer: 50%; PEGABOT: 27% and TweetBotOrNot: 18%. Interesting that in this case, Botometer shows the highest score among the users analyzed in this section. This result could probably due the long time of inactivity of this account, which alone, not means that a dormant user is a robot. A straightforward test was done with the user @ilzarovai, that it is not a bot, but has its account inactive by around one month. Botometer considers that this account has 69% of the probability of being a bot, a false positive. PEGABOT gives it 41% and TweetBotOrNot, 19%.

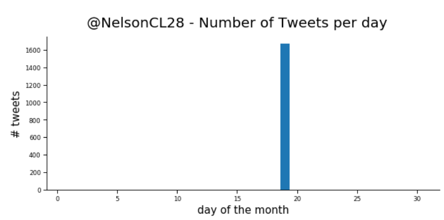

Figure 26: @NelsonCL28 tweets/day

Figure 27: Original single tweet

6.2.4 HF - Conclusion

As discussed in previous sections, non-verified HF users have a high probability of being a bot, which could be confirmed with those three examples. Indeed, those accounts are examples of more elaborated bots, where their changing behavior helps them to avoid detection by classical apps.

6.3 Spotting Young Mid-Frequency Accounts

In this group, accounts are filtered with not verified users posting more than 72 tweets/day and with less than 90 days since creation. Applying above filtering criteria resulted in 10 accounts spotted:

- **@EncuestaExpress** - (Appendix I) - [Bot] - On a single day (20/Nov), the user posted more than 700 tweets, having the great majority of them the same text. PEGABOT: 76%; Botometer: 78%
- **@RResponsablecl** - (Appendix J) - [Possible a Cyborg]: PEGABOT: 67%; Botometer: 20%
- **@cazadorandino90** - (Appendix X) - [Bot]: around 30,000 tweets in one year and no more tweets after October 2018. PEGABOT: 53%; Botometer: 24%
- **@Piagutierrezs** - (Appendix K) - [Bot]. Majority of tweets with the same text. This account is still active nowadays, with a high number of followers and friends. PEGABOT: 54%; Botometer: 18%
- **@NathalySeplved3** - (Appendix L) - [Possible a Cyborg]: Two days during the timeline window with more than 100 tweets (same text). Account is active nowadays, with high number of followers and friends. PEGABOT: 71%; Botometer: 8%

- **@ElCentinelaMPE** - (Appendix M) - [Could be human]:Only used during November 2017. Still active but not posting. PEGABOT: 45%; Botometer: 28%
- **@AShumman** - (Appendix O) - [Bot] Suspended Account
- **@PamelaSoler3** - (Appendix N) - [Bot] Not active nowadays
- **@Sumate_Guillier** - (Appendix P) - [Bot] Not active nowadays
- **@jav_ast** - (Appendix Q) - [Bot] Single day with 104 same text tweets. Account still active but not posting. PEGABOT: 69%; Botometer: 78%

6.3.1 YMF - Conclusion

Combining only three critical features as "not verified", posting more than "72 tweets/day" and with less than "90 days" since creation resulted in 9 of 10 accounts to be a bot, but not all of them spotted by apps. This result confirms that combining critical features and manual analysis is an effective method to spot a bot.

6.4 User Behavior - Manual spotting Results

Manually analyzing if a user is a bot or cyborg is a tedious and time-consuming task but is the only way to confirm a bot behavior. Many of the works that try to apply supervised machine learning techniques to identify if an account is a bot or not use humans (usually more than one to judge each account) before label it. With some of them, as Botometer, the result observed is not great.

7 Spotting User Behavior Clusters

This work experiments with an entirely new approach to spot a social bot behavior, visually exploring clusters of users with similar behavior, using a Uniform Manifold Approximation and Projection (UMAP) technic.

7.1 UMAP Overview

UMAP is a dimension reduction technique that can be used for visualization similarly to t-SNE [37]. The algorithm is founded on three assumptions about the data:

- The data is uniformly distributed on a Riemannian manifold;
- The Riemannian metric is locally constant (or can be approximated as such);
- The manifold is locally connected.

From these assumptions, it is possible to model the manifold with a fuzzy topological structure. The embedding is found by searching for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure.

The details for the underlying mathematics can be found in [30], McInnes, L, Healy, J, UMAP: Uniform Manifold Approximation, and Projection for Dimension Reduction.

7.2 Model Preparation

To the *User Behavior Dataset*, with its 32 features (user screen-name and 31 numeric metrics), was add a new feature named label. The idea is to label each of the users according to its frequency-post behavior, helping to spot users on the final UMAP visualization, but not to be used on calculations. Below how users are labeled:

- Low-Frequency: less than 50 tweets/day (89,959 users)

- Mid-Frequency: between 50 and 144 tweets/day (798 users)
- High-Frequency: More than 144 tweets/day (397 users)

As a model input, one Array with dimension (91154, 31) is created from the *User Behavior Dataset*, whre the "user_name" that is the dataset's index and the recently created "label" are taken out. The data do not need to be normalized, since according to [30], this is one of the advantages of UMAP.

7.3 Hyper-Parameters Tuning

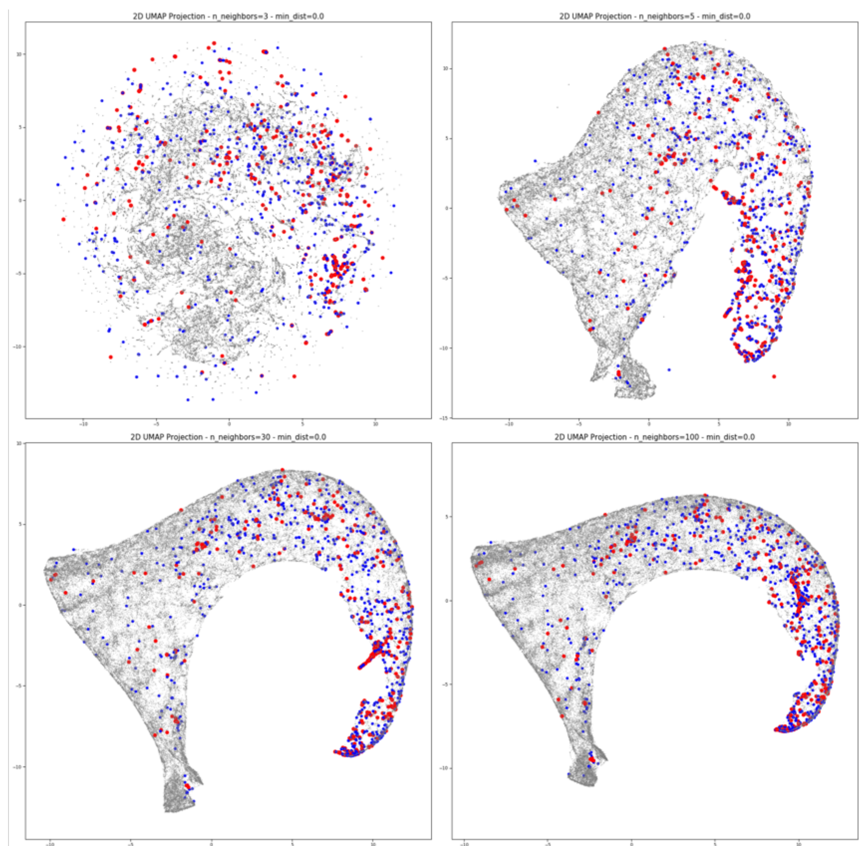

Figure 28: UMAP 2D User Dataset - Hyper Parameters selection

The three most essential hyper-Parameters of UMAP model [30] are:

- n_components: The target embedding dimension such as 2 for 2D and 3 for 3D.
- min_dist: In essence, this parameter determines how closely points can be packed together in the low dimensional representation. Low values on min_dist result in potentially densely packed regions, but will likely more faithfully represent the manifold structure. Increasing the value of min_dist forces the embedding to spread points out more, assisting visualization (and avoiding potential overplotting issues). This experience uses 0 for this parameter.
- n_neighbors: That represents some degree of trade-off between fine-grained and broad-scale manifold features. Smaller values ensure the detailed manifold structure is accurately captured (at a loss of the "big picture" view of the

manifold), while larger values capture large scale manifold structures, but at a loss of fine detail structure which gets averaged out in the local approximations, as shown in Figure 28. A good compromise seems to be reached with n_neighbors = 30.

Figure 29 shows the same parameters variation, but using n_components=3, what will result on a 3D visualization. Each one of resultant visualization shown in both figures are very fast to be computed, going from 1 to 4 minutes on a MacBook, 2.9GHz Core i7, 16GB RAM, where red Dots represent High-Frequency users; the blue: Mid-Frequency users and the Gray: Low-Frequency users.

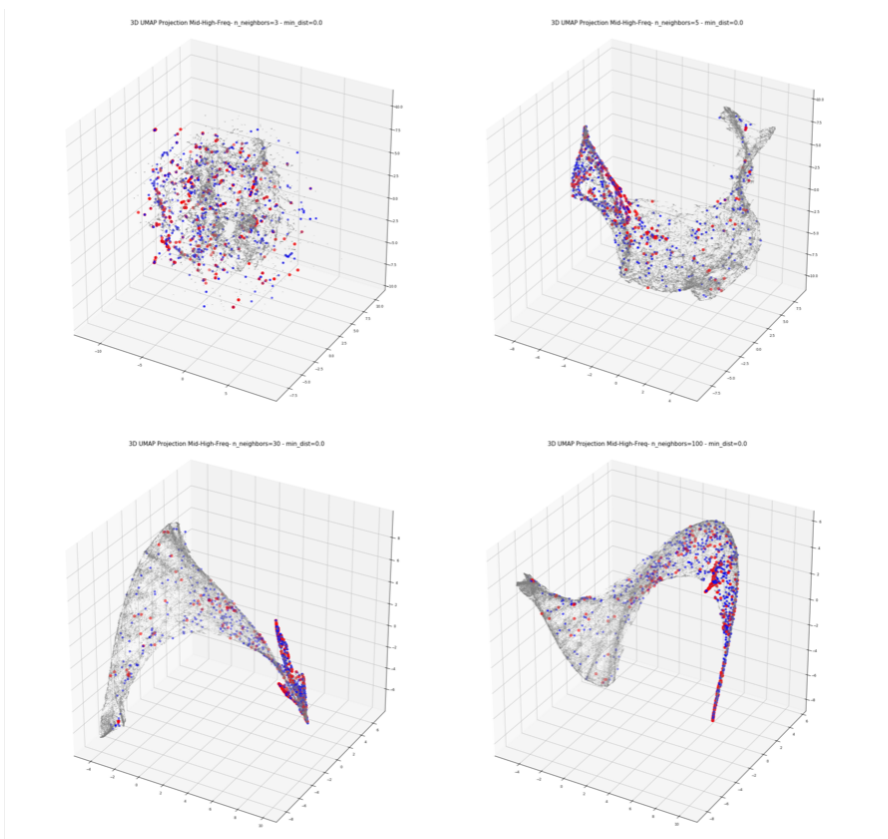

Figure 29: UMAP 3D User Dataset - Hyper Parameters selection

7.4 Creating clusters of users with similar behavior

Analysis can be done with both, 2D or 3D projection, but the ideal on 3D is taken three 2D projections (x-y, y-z and x-z).

On a first attempt and using as hyper-parameters: n_component = 2 (2D); n_neighbors = 30; and min_dist = 0, all 31 dimensions were reduced to only two, and used to plot the 2D figure. These two new dimensions "x" and "y" are add to the original *User Behavior Dataset* for exploring the regions of resultant visualization.

Intuitively, it is possible to observe from figure 28 that bigger x and smaller y means more concentration of high/mid-frequency users and possible bots. The next approach is to select a few significant areas of the figure to analyzed them in more detail as shown in Figure 30.

The three areas marked in Figure 30 are possible clusters, and three sub-datasets were created to be analyzed.

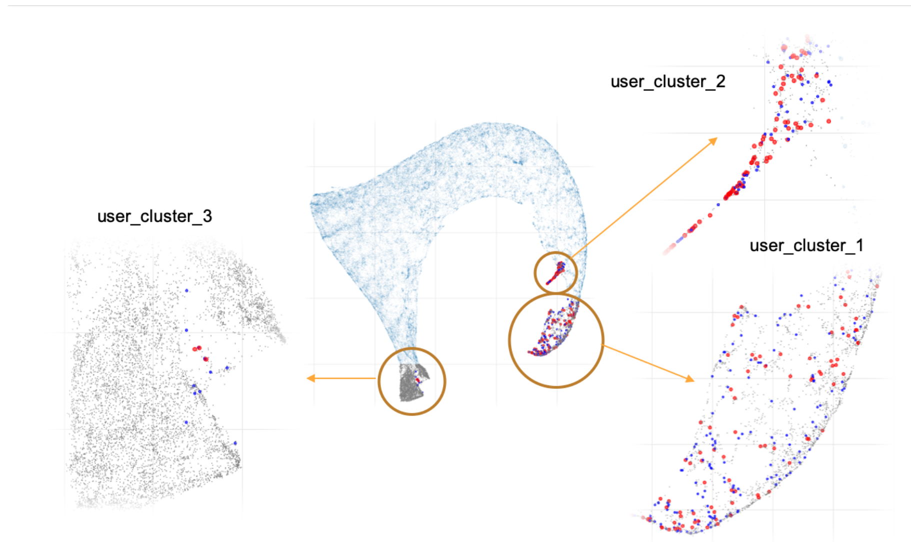

Figure 30: UMAP 2D - user selected clusters

7.4.1 Example - Comparing user_cluster_1 with user_cluster_3

Grouping the *User Behavior Dataset*, by the new X and Y features, is possible to analyse in details the features through their primary statistical data, as such:

- user_cluster_1 has a more significant concentration of High/Mid-frequency users
- user_cluster_1 has shorter screen_names
- user_cluster_1 retweet more than user_cluster_3, which is a good indication of social bots, as shown in Figure 31.
- user_cluster_3 has younger accounts
- user_cluster_3 is formed almost with low-frequency users
- user_cluster_3 is more populated with no or short account description
- user_cluster_3 show a more "human" similarity between name and screen_name (feature over 60%) as shown on Figure 32.

Only looking those metrics seems that accounts on user_cluster_1 (more to the right), tweet more frequently, are older, with higher retweet rate, have a full description and, do not have similarity on names (name and user-name could be inventions). At first sight, social bots could be concentrated on this area of graphics, however more work should be done here.

7.5 Spotting known bots on UMAP

Another interesting analysis is to spot where the known bots spotted on previous section are in the UMAP visualization.

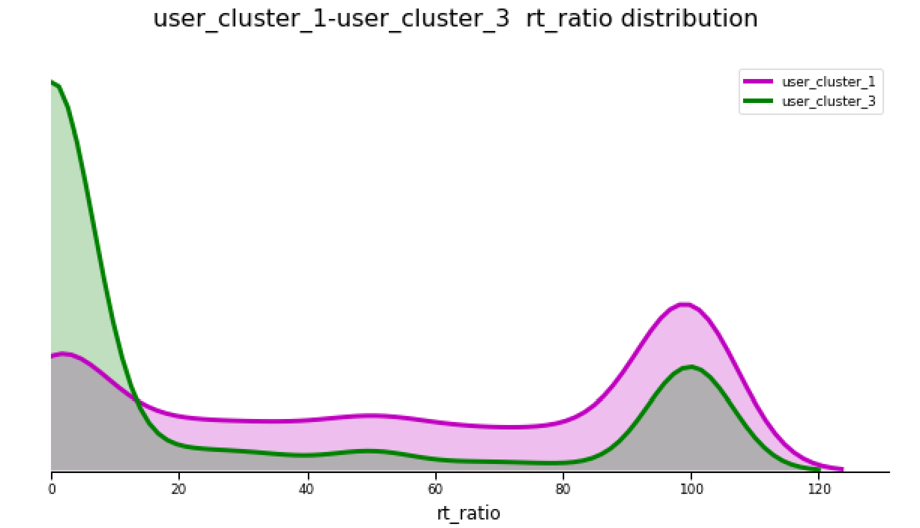

Figure 31: Retweet ratio by cluster

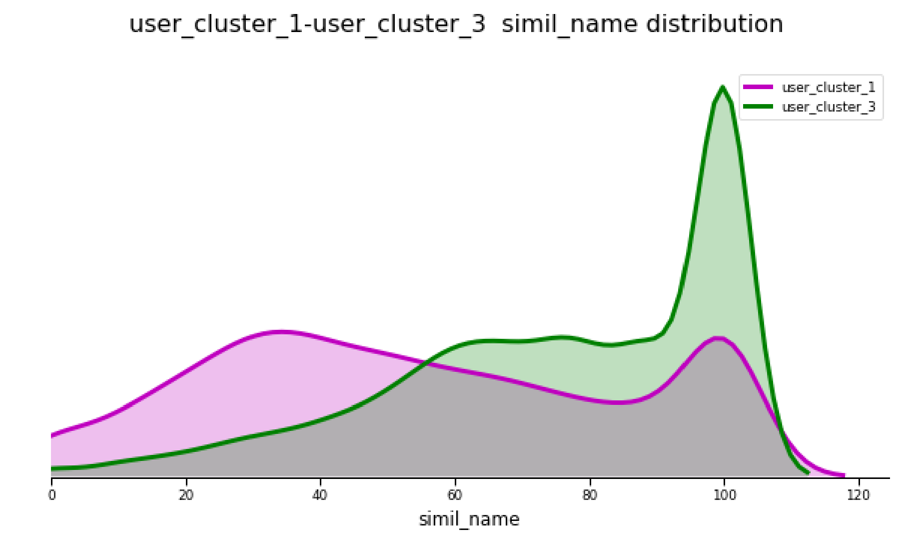

Figure 32: Similarity Name/User-Name by cluster

Known bots:

- @fedoraletelier (SB)
- @Aliciacarafipl3 (SB)
- @Dolores09072598 (SB)
- @AlbertoMayol (HF)
- @Tomaskovacic (HF)
- @NelsonCL28 (HF)
- @EncuestaExpress (YMF)
- @RResponsablecl (YMF)
- @cazadorandino90 (YMF)
- @Piagutierrezs (YMF)

- @NathalySeplved3 (YMF)
- @ElCentinelaMPE (YMF)
- @AShumman (YMF)
- @PamelaSoler3 (YMF)
- @Sumate_Guillier (YMF)
- @jav_ast (YMF)

Figure 33 shows where the known bots are located on a 2D UMAP visualization. Unfortunately, those suspicious users are spread all over the visualization without be grouped on a clear area of bots.

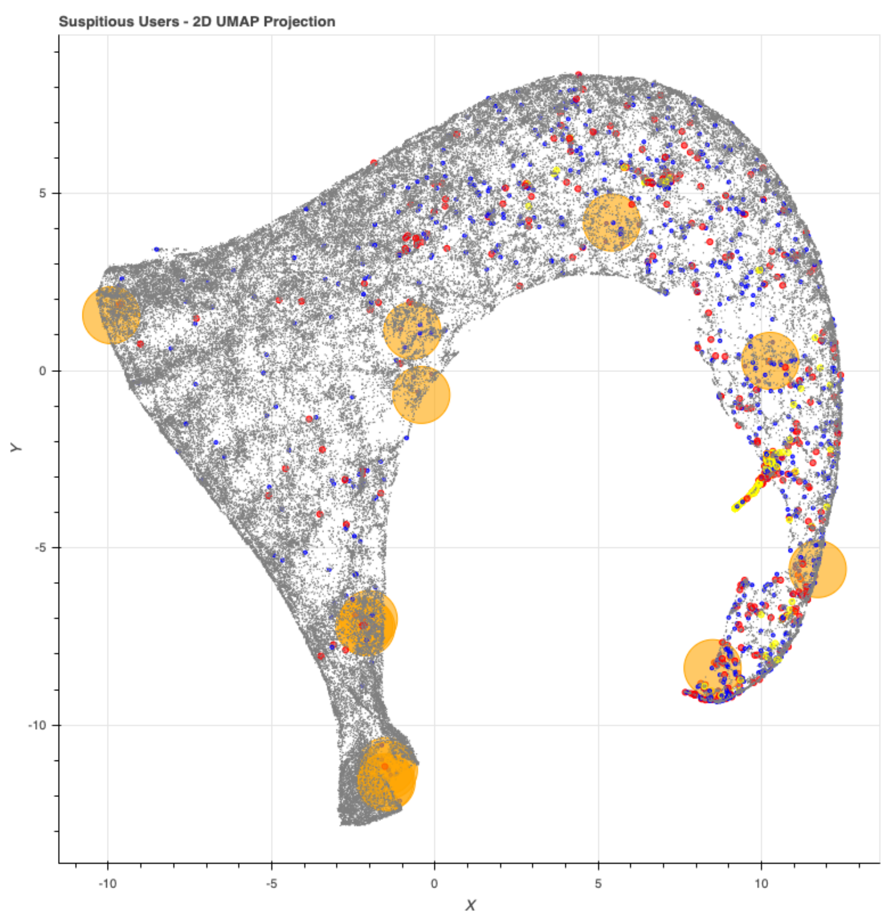

Figure 33: Suspicious Users

7.6 Filtering Mid/High-Frequency users

The next experiment is to apply UMAP to a filtered dataset, populated only with users that have posted more than 50 tweets/day in at least one occasion during the timeline window. In total, 1,195 accounts are considered, including verified users. From those, 798 accounts are Mid-Frequency (from 50 up to 144 tweets/day), and 397 are High-Frequency accounts (more than 144 tweets/day).

All users are colored in blue, except for verified users that are in yellow.

In this case, testing with the same hyper-parameters as done with the complete dataset seems that several clusters can be defined using n_neighbors=5, as shown in Figure 34.

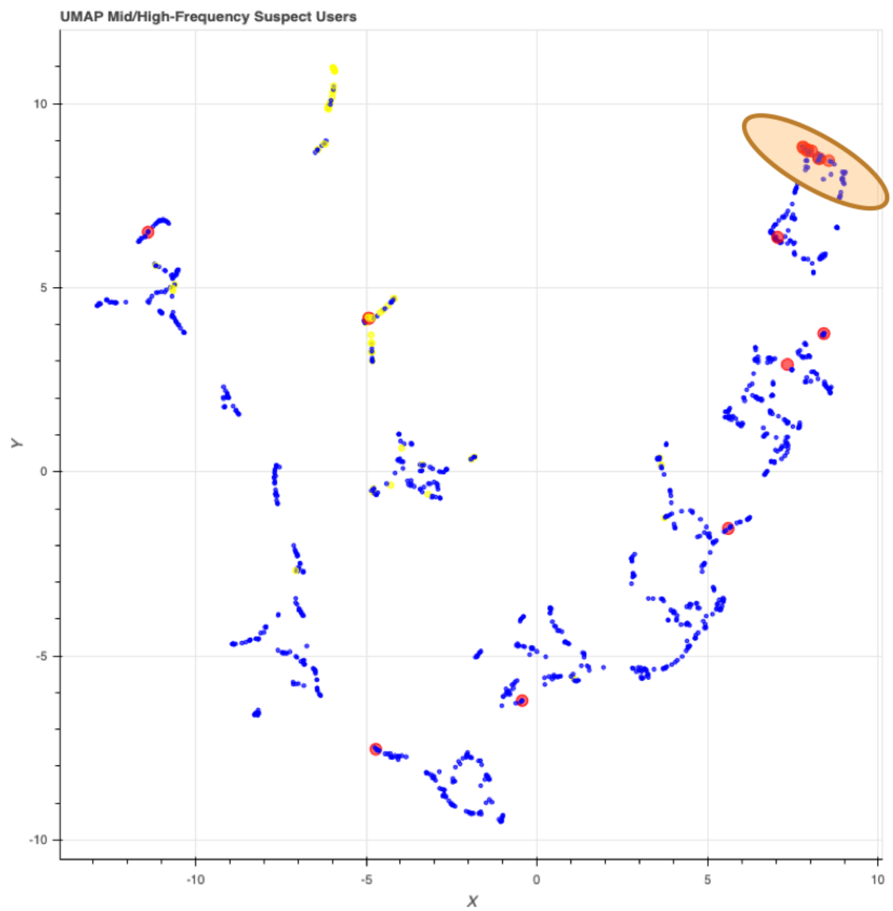

Figure 34: Suspect Users

Now, taking the known bots spotted on previous sections and now colored in red, the result appears that the cluster formed with both x and y greater than 5, concentrate the majority of them and more precisely at its border. For reference, this area is called Cluster 0.

Using Bokeh [38], and doing zooming, it is possible to execute an exploration on the users that are part of the cluster 0 (Figure 35). Once captured their user names, a manual analysis is done on their features in the same way that was done in previous sections.

When possible, the users are also tested with Botometer, PEGABOT, and Tweet-BotOrNot. Botometer missed almost all users, and PEGABOT is more assertive in finding bots. When working, TweetBotOrNot seems better when compared with Botometer, but it is precarious and usually presented error ("Server out").

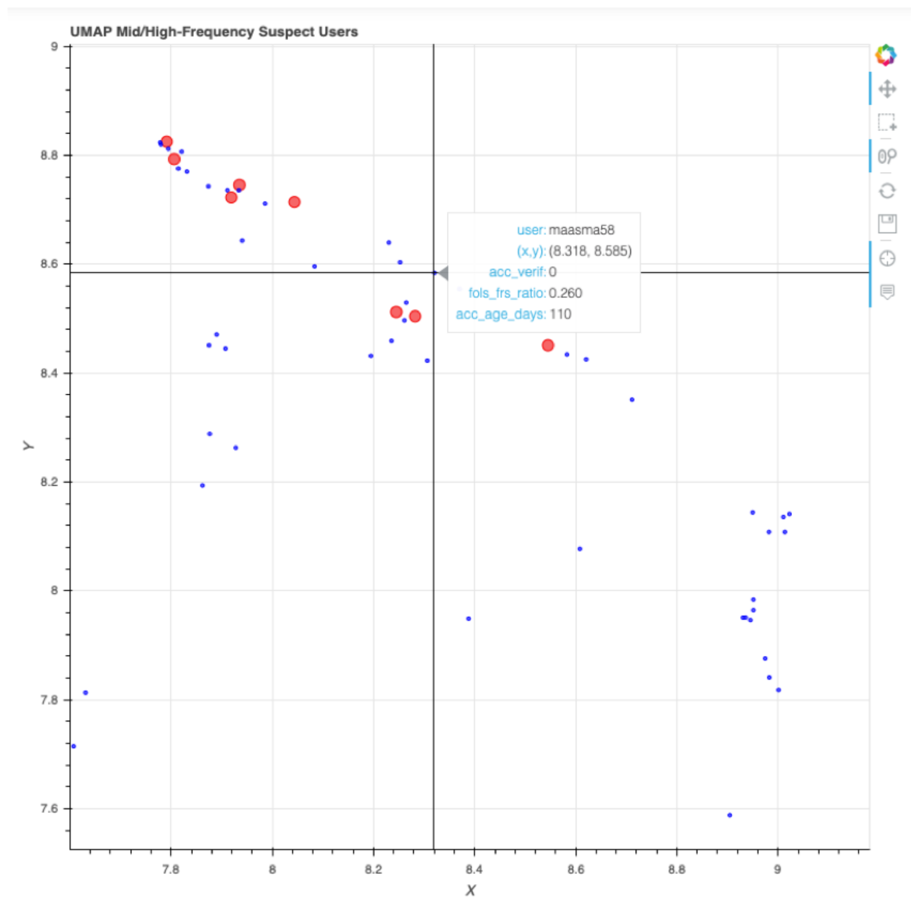

Figure 35: Cluster Exploration

7.6.1 Spotting Bots at "Cluster 0"

On the sequence, selected users are listed with the main conclusion about their behavior.

@mrgrtgautier = BOT

- 100% RT
- 85% of all tweets have hashtags
- Almost all tweets (53/55) at a single day
- Young account: 12 days
- Active account - Last tweet dec/2017
- Pegabot: 8%
- Botometer: 22%

@ChristianPolo7 = BOT

- 87% RT
- 97% of all tweets have mentions
- Very active account during elections and few months after
- Young account: 15 days
- Active account - Last tweet march/2018
- Botometer: No (22%)

- PegaBot: 41% (Network: 85%)

@Conapro_jjcc = BOT

- 0% RT
- 85% of all tweets have a single mention (@Guillier)
- Only posting on 2 days same tweet/day
- Young account: 22 days
- Active account
- Last tweet march/2018
- Botometer: Bot (92%)
- PegaBot: No (16%)

@Atletadelgol32 = BOT

- 97% RT
- 100% of all tweets have mentions
- Only posting on 2 days
- Young account: 3 days
- Not Active account

@ElCentinelaMPE = BOT

- 1% RT
- almost 100% of all tweets have hashtags, being #Elecciones on 96 of them
- Only posting on 5 days (81% of all tweets (114 on the same day)
- Young account: created start of the month and did not post after nov/2017
- Active account
- Botometer: No (28%)
- Pegabot: No find

@JuanManuelCorn5 = BOT

- 1% RT
- 100% of all tweets have mentions: a lot of them!
- Average text length per tweet: 654 (@XXX)
- Only posting a few days after elections
- Young account: 92 days and do not post after dec/2017
- Active account
- Pegabot: 47% (Network 58%)
- Botometer: No (28%)

@Santiag87306226 = BOT

- 5% RT
- 99% of all tweets have mentions
- 89% have hashtags
- 70% of all tweets on the same day (6/nov), almost all with same text
- Young account: 44 days
- Blocked account

@SFelipeAlegreJ = Possible BOT or Cyborg

- 0% RT
- 100% of all tweets have mentions
- 88% have hashtags
- Almost all tweets on the same day (23/nov), almost all with same text
- Young account: 93 days

- Active account
- Botometer: no (34%)
- Pegabot: No found

@Trab_vXguillier = Possible BOT or Cyborg

- 1 day with the same tweet, having almost half of analyzed tweets

@Sumate_Guillier = BOT

- 3% RT
- 195 tweets on the same day (23/nov), almost all with two different texts
- Young account: created after 1st round election
- Not Active account

@viejofasho = Possible BOT or cyborg

- 0% RT
- 100% mentions, with multiples mentions on a single tweet
- Average text length per tweet: 344
- Young account: created end of October/17
- 5,725 in 9 months
- Active account, but w/o tweet since mid/18.
- PegaBot: 42% (77%:Network)
- Botometer: No (18%)
- TweetBotOrNot: Error

@RResponsablecl = Possible BOT or cyborg

- One day (26/nov) with 162 tweets, 2 different texts.
- young account, created same month (nov/17)
- still active, but seems normal nowadays
- Pegabot Yes (66%)
- Botometer: No (24%)

@BassaRiveros = Possible BOT or cyborg

- one day (26/nov) with 162 tweets, 2 different texts.
- young account, created same month (nov/17)
- still active, but seems normal nowadays
- Pegabot Yes (100%)
- Botometer: No (36%)

@arqmneira = Possible NOT BOT

- PegaBot: 81%
- Botometer: no (10%)
- TweetBotOrNot: Error

@Ignacio90415476 = Possible a cyborg

- Account created 90 days before the start of nov/17
- Today is active and posted 6,000 tweets/year since 2017
- Botometer: No
- PegaBot: Yes (76%)

@maasma58 = Possible bot

- Account created 90 days before the start of nov/17
- Today Blocked

@Aptimate = Possible cyborg

- Account created 10 days before the start of nov/17

- Today active
- Botometer: 14%
- PegaBot: 36%

@mas_estudiantil = Possible bot

- Few activities and one day with the same tweet repeated
- Today active but w/o post
- Botometer: Not find
- PegaBot: Not Find

7.6.2 "Cluster 0" Conclusion

Manually exploring each one of those accounts selected from Cluster 0 proves that the vast majority of them have, in fact, a bot (or cyborg) behavior. The next step is to also plot those bots on a 3D visualization, but using again the complete *User Behavior Dataset*.

7.7 User Behavior Clustering Results

When plotting the Cluster O spotted bots, on a Y/Z 3D UMAP visualization, they also seem to be concentrate on one of its borders, as shown in Figure 36.

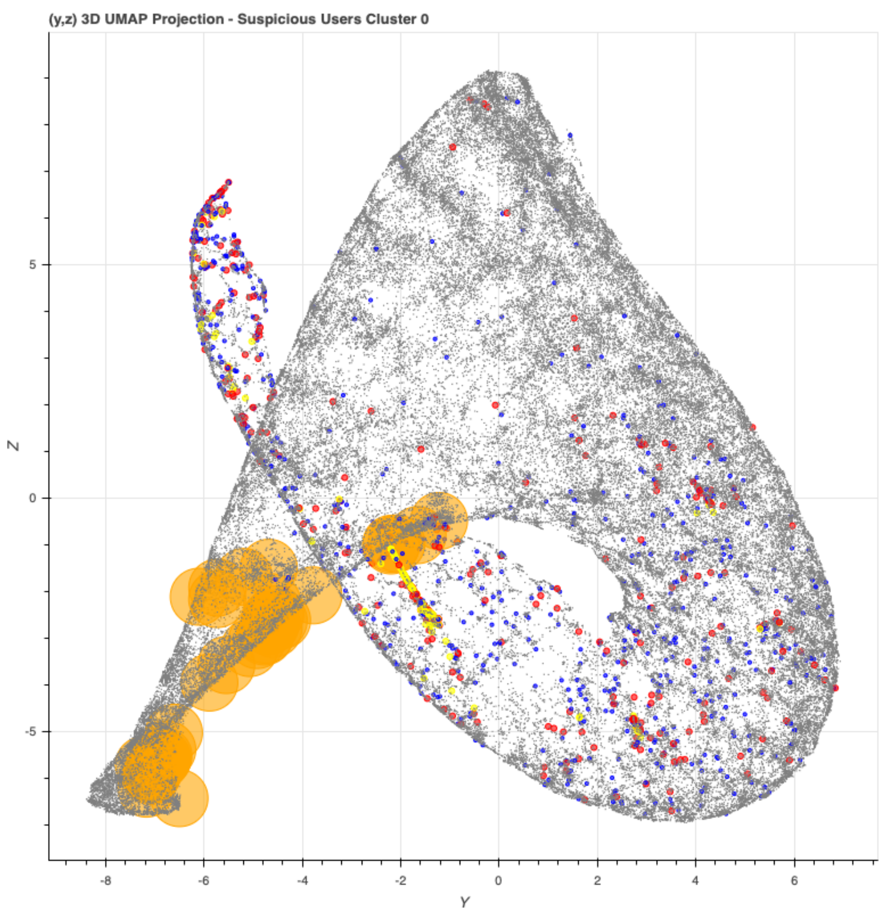

Figure 36: (y,z) plan of 3D UMAP Proj. - Cluster 0

This result is auspicious and is in line with [16], where the researches identify 10 distinct clusters, as shown in Figure 37. In that work, researchers create a 2-dimensional projection of users also obtained by a dimensionality reduction technique but using t-SNE [37] instead of UMAP. Varol et al. representation uses 100 features extracted from user accounts, which should be more representative than the 31 used on this work. From the 10 observed clusters, three of them, namely C0–C2, have high average bot scores, which was manually verified.

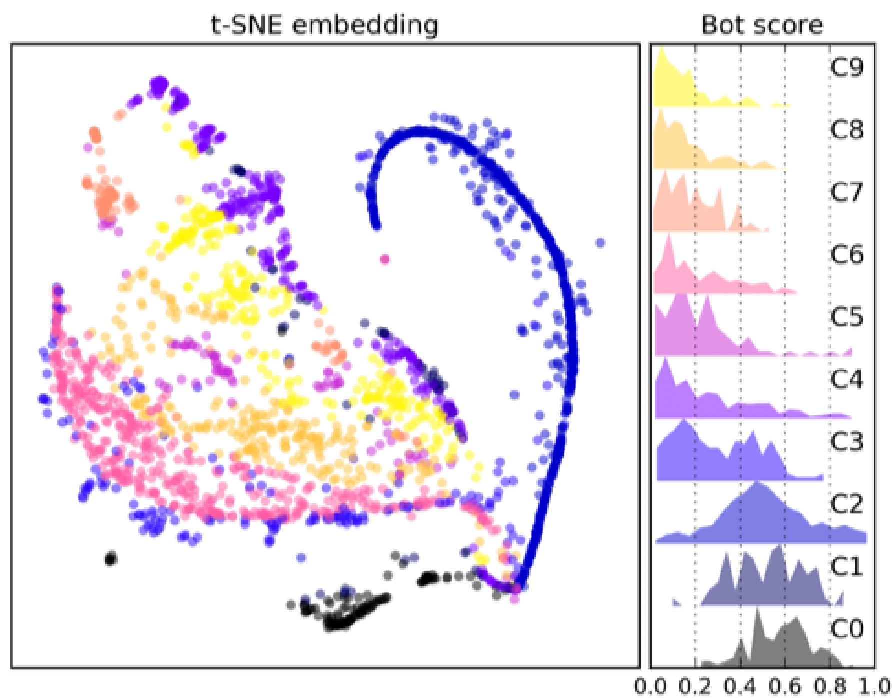

Figure 37: t-SNE Clusters - Varol et al. 2017

8 Future Work

The similarity between users in this work is calculated based on their 31-dimensional representation in the feature space and is still challenging to spot a clear cluster of similar social bot behavior more automatically. Future work should capture more features, focusing on timing and network and UMAP should be tested with different hyper-parameters.

Once a significant part of the analysis is done manually, an API similar to [39] should be developed to explore suspicious accounts, extracting features from its more current tweets.

Also, the methodology developed in this study should be applied to different tweet datasets.

9 Conclusion

Today, the traditional definition of fully automated bots operating in the social media space can no longer be applied as it, once it is much more common for humans to control these accounts by alternating their activities between "human" and "bot", the so-called "cyborgs". This constant behavioral shift makes it much more difficult to detect a bot, especially by Supervised Machine Learning models.

Fully automated bots (the so-called " Simple Bots ") are becoming increasingly difficult to find on social media, as nowadays the cost of developing a sophisticated bot is less than the cost of detecting it.

Modern bots are therefore more sophisticated, they work in an orchestrated manner by alternating their behavior between humans and machine and the best way to detect them is by studying the frequency of their tweets ("timing") and how they relate to each other ("networking").

In this work, a framework was presented in which unstructured data such as text extracted from millions of tweets are converted into a *User Behavior Dataset* of engineering features, to be used to spot a possible bot behavior, not importing if the account is fully automated or managed by a human.

Depending on the value of each *User Behavior Dataset* feature, it can be considered as a flag of a "suspicious behavior" (such as frequency of post over 144 tweets/day). However, none of those features, if taken individually, is enough to conclusively define whether an account is a bot or not.

This work verified that the more suspicious features an account displays, the more likely it is to be automated and so, be a bot, cyborg or be part of orchestration such as a botnet. This verification can be done manually, account by account looking at the full set of features and for that, a simple dashboard was developed to easily support this task.

This work also experimented with an entirely new approach, trying to visually spot a social bot behavior, by exploring clusters of users with similar behavior, using dimension reduction technique. When plotting a group of known bots, on a 3D visualization, a concentrating of accounts appeared on one of its borders, suggesting a possible region where bots could be much more easily spotted.

Acknowledgments
The analysis was performed in Python using Jupyter Notebooks, jointly with the Scikit-Learn, Pandas, Seaborn, Bokeh, Plotly, UMAP, Tweepy, and Botometer. It is essential to recognize the valuable advice, considerations, and share of experiences provided by Eduardo Graells-Garrido (UDD/BSC), Loreto Bravo (UDD), Leo Ferres (UDD), Maristela Rovai (SDSU), Ilza Rovai and Mauricio Pinto.

Source Code
Source code can be found at https://zenodo.org/record/3576148#.XfUbsi2ZNTY

Cite as
Rovai, Marcelo Jose. (2019, December 14). Studying Twitter User Accounts: Spotting Suspicious Social Bot Behavior (Version 2). Zenodo. http://doi.org/10.5281/zenodo.3576148

References

1. The Russians Are Meddling Again, this Time in Chile, Warns US Diplomat. https://www.rt.com/news/471945-russian-interference-chile-protests/ Accessed 2019-11-16
2. Bots in Brazil: The Activity of Social Media Bots in Brazilian Elections. https://www.wilsoncenter.org/blog-post/bots-brazil-the-activity-social-media-bots-brazilian-elections Accessed 2019-11-20
3. Santander, P., Elórtegui, C., González, C., Allende-Cid, H., Palma, W.: Redes sociales, inteligencia computacional y predicción electoral: el caso de las primarias presidenciales de chile 2017 **0**(41), 41–5656. doi:10.7764/cdi.41.1218. Accessed 2019-08-30
4. Study: Russian Twitter Bots Sent 45k Brexit Tweets Close to Vote. https://techcrunch.com/2017/11/15/study-russian-twitter-bots-sent-45k-brexit-tweets-close-to-vote/ Accessed 2019-11-20
5. Bots and Automation over Twitter During the U.S. Election. https://comprop.oii.ox.ac.uk/research/working-papers/bots-and-automation-over-twitter-during-the-u-s-election/ Accessed 2019-11-16

6. Shao, C., Ciampaglia, G.L., Varol, O., Yang, K.-C., Flammini, A., Menczer, F.: The spread of low-credibility content by social bots **9**(1), 4787. doi:10.1038/s41467-018-06930-7. Accessed 2019-06-13
7. Bot Traffic Report 2016. https://www.imperva.com/blog/bot-traffic-report-2016/ Accessed 2019-11-16
8. Bad Bot Report 2019: The Bot Arms Race Continues. https://www.imperva.com/resources/resource-library/reports/bad-bot-report-2019-the-bot-arms-race-continues/ Accessed 2019-11-16
9. Grimme, C., Preuss, M., Adam, L., Trautmann, H.: Social bots: Human-like by means of human control? Big Data **5** (2017). doi:10.1089/big.2017.0044
10. How to Spot a Twitter Bot. https://www.symantec.com/blogs/election-security/spot-twitter-bot Accessed 2019-11-16
11. Ahmed, W., Bath, P.A., Demartini, G.: Chapter 4: Using twitter as a data source: An overview of ethical, legal, and methodological challenges. In: Woodfield, K. (ed.) Advances in Research Ethics and Integrity vol. 2, pp. 79–107. Emerald Publishing Limited. doi:10.1108/S2398-601820180000002004. http://www.emeraldinsight.com/doi/10.1108/S2398-601820180000002004 Accessed 2019-10-10
12. The Follower Factory. https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html Accessed 2019-11-16
13. Platform Manipulation and Spam Policy. https://help.twitter.com/en/rules-and-policies/platform-manipulation Accessed 2019-11-16
14. Bot or Not: an End-to-end Data Analysis in Python. http://www.erinshellman.com/bot-or-not/ Accessed 2019-11-16
15. Ferrara, E., Varol, O., Davis, C., Menczer, F., Flammini, A.: The rise of social bots **59**(7), 96–104. doi:10.1145/2818717. 1407.5225. Accessed 2019-06-13
16. Varol, O., Ferrara, E., Davis, C.A., Menczer, F., Flammini, A.: Online human-bot interactions: Detection, estimation, and characterization. 1703.03107. Accessed 2019-06-13
17. Minnich, A., Chavoshi, N., Koutra, D., Mueen, A.: BotWalk: Efficient adaptive exploration of twitter bot networks. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17, pp. 467–474. ACM Press. doi:10.1145/3110025.3110163. http://dl.acm.org/citation.cfm?doid=3110025.3110163 Accessed 2019-07-06
18. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. In: Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion, pp. 963–972. ACM Press. doi:10.1145/3041021.3055135. http://dl.acm.org/citation.cfm?doid=3041021.3055135 Accessed 2019-07-01
19. Davis, C.A., Varol, O., Ferrara, E., Flammini, A., Menczer, F.: Botornot: A system to evaluate social bots, 273–274. doi:10.1145/2872518.2889302. 1602.00975. Accessed 2019-06-13
20. Observatory on Social Media. https://osome.iuni.iu.edu/ Accessed 2019-11-20
21. Kearney, M.W.: tweetbotornot: Detecting Twitter bots (web app: https://mikewk.shinyapps.io/botornot/). Unpublished. type: dataset (2018). doi:10.13140/RG.2.2.10732.82562. http://rgdoi.net/10.13140/RG.2.2.10732.82562 Accessed 2019-11-20
22. The Institute for Technology Society of Rio. https://itsrio.org/en/en-home/ Accessed 2019-11-20
23. Spottingbot GitHub - Analyzing Profile on Twitter for Detect Behavior of a Spamming Bot. https://github.com/AppCivico/spottingbot Accessed 2019-11-20
24. PEGABOT. https://pegabot.com.br Accessed 2019-11-19
25. Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., Tesconi, M.: DNA-inspired online behavioral modeling and its application to spambot detection. 1602.00110. Accessed 2019-07-02
26. Tweet Object. https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object Accessed 2019-12-09
27. Counting Characters. https://developer.twitter.com/en/docs/basics/counting-characters Accessed 2019-11-16
28. Following FAQs. https://help.twitter.com/en/using-twitter/following-faqs Accessed 2019-11-16
29. Freitas, C.A., Benevenuto, F., Ghosh, S., Veloso, A.: Reverse engineering socialbot infiltration strategies in twitter. 1405.4927. Accessed 2019-11-12
30. McInnes, L., Healy, J., Melville, J.: UMAP: Uniform manifold approximation and projection for dimension reduction. 1802.03426. Accessed 2019-11-11
31. Gilani, Z., Farahbakhsh, R., Tyson, G., Wang, L., Crowcroft, J.: Of bots and humans (on twitter). In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017 - ASONAM '17, pp. 349–354. ACM Press. doi:10.1145/3110025.3110090. http://dl.acm.org/citation.cfm?doid=3110025.3110090 Accessed 2019-11-13
32. How to Spot a Bot (or Not): The Main Indicators of Online Automation, Co-ordination and Inauthentic Activity. https://firstdraftnews.org/latest/how-to-spot-a-bot-or-not-the-main-indicators-of-online-automation-co-ordination-and-inauthentic-activity/ Accessed 2019-12-10
33. BotSpot: Twelve Ways to Spot a Bot. https://medium.com/dfrlab/botspot-twelve-ways-to-spot-a-bot-aedc7d9c110c Accessed 2019-11-16
34. The Not-so-simple Science of Social Media 'bots'. https://firstdraftnews.org/latest/the-not-so-simple-science-of-social-media-bots/ Accessed 2019-12-09
35. Botometer An OSoMe Project (bot•o•meter). https://botometer.iuni.iu.edu/!/ Accessed 2019-11-19
36. TweetBotOrNot. https://mikewk.shinyapps.io/botornot/ Accessed 2019-11-19
37. T-distributed Stochastic Neighbor Embedding. https://en.wikipedia.org/wiki/T-distributed$_{s}tochastic_{n}eighbor_{e}mbeddingAccessed2019-11-16$
38. Bokeh. https://docs.bokeh.org/en/latest/ Accessed 2019-11-16
39. Udd SpotBot Producto de Datos. https://github.com/Mjrovai/UDD_Master_Data_Science/tree/master/PD_SpotBot_Final_Project Accessed 2019-12-14

Additional Files
Appendices - User data sample

Appendix A

Data for Suspicious users.jpg

```
Info Data for @fedoraletelier

Name:                        fedoraletelier
Name Length:                 14
@user number Numeric Chars:  0
@user and name similarity:   0
User Id:                     2465345439
Default_Photo:               True
Photo URL:                   http://abs.twimg.com/sticky/default_profile_im
ages/default_profile_normal.png
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             False
Description:
Description length:          0
Account Active:              Active

Number of Tweets analyzed:          690
Last Tweet:                         2017-11-29T19:27:32.000000000
Older Tweet Analyzed:               2017-10-31T23:07:51.000000000
Account Active age (in days)        1,333.0
Average recent tweets per day:      24.64
Maximum Recent Tweets Per Day:      85 at day 29
Average tweets per day since start: 161.34
Total Tweets since start:           215,069
Followers_count:                    2,242
Friends_following_count             1,886
Followers/Following_index:          1.19
Favourites (Like) count:            156,764
Retweet_Like_index:                 0.0
Retweet Ratio:                      43%
unique mentions cnt:                325
Mention Ratio:                      70%
Mention index:                      0.16
unique hashtags cnt:                50
Hashtag Ratio:                      49%
Hashtag Index:                      0.1
Average URLs per tweet:             0.45
Average text length per tweet:      137.0

       [Tweets Sample at Peak Day]
 1429686    RT @Carmen_muller: @aldocardinali Guillier ha ...
1431240    RT @supernovaf_1000: #EstonoesPlazaItalia @Mov...
1433787    RT @Carmen_muller: @PadresPartSubv @mbachelet ...
1433812    RT @SoniaSHernandez: @PadresPartSubv @mbachele...
1435809    RT @Bella_DiGenaro: Extra! Extra! Me acaban de...
1435834    RT @klaudiaunik: @Bella_DiGenaro @guillier @je...
1435873    RT @Sirius4321: @Bella_DiGenaro @Makeka @guill...
1436465    RT @supernovaf_1000: #Cadem @guillier MIENTE N...
1439928    RT @alecabanast: @Commander_SB @anzunza @sebas...
1461708    RT @aeristac: @armandoredondo1 @aggrezgmailcom...
```

Appendix B

Data for Suspicious users-2.jpg

```
Info Data for @Aliciacarafipl3

Name:                        Aliciacarafiplanella
Name Length:                 20
@user number Numeric Chars: 1
@user and name similarity:  0
User Id:                     801131674402701312
Default_Photo:               True
Photo URL:                   http://abs.twimg.com/sticky/default_profile_im
ages/default_profile_normal.png
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             False
Description:
Description length:          0
Account Active:              Active

Number of Tweets analyzed:          747
Last Tweet:                         2017-11-30T22:22:11.000000000
Older Tweet Analyzed:               2017-11-01T21:52:14.000000000
Account Active age (in days)        373.0
Average recent tweets per day:      25.76
Maximum Recent Tweets Per Day:      78 at day 6
Average tweets per day since start: 69.12
Total Tweets since start:           25,781
Followers_count:                    200
Friends_following_count             113
Followers/Following_index:          1.77
Favourites (Like) count:            13,595
Retweet_Like_index:                 0.0
Retweet Ratio:                      96%
unique mentions cnt:                622
Mention Ratio:                      100%
Mention index:                      0.27
unique hashtags cnt:                96
Hashtag Ratio:                      19%
Hashtag Index:                      0.5
Average URLs per tweet:             0.15
Average text length per tweet:      140.0

       [Tweets Sample at Peak Day]
 144348    RT @LuisJValdivia: El problema de Guillier no ...
149179    RT @cla141966: #Tolerancia0 Guillier "estoy en...
149241    RT @pablolirar: Villegas se está paseando a Gu...
149269    RT @joseotero78: Guillier no tiene respaldo. E...
149374    RT @pablolirar: Jajajaja @Guillier dice que su...
149443    RT @joseotero78: Guillier dice que Chile crece...
149530    RT @pablolirar: .@Guillier:"Chile es el país n...
149716    RT @joseotero78: Mas de lo mismo! Guillier es ...
149776    RT @Regia_Pam: @Claunubed @guillier Esta noche...
149827    RT @joseantoniokast: .@guillier dice que soy u...
```

Appendix C

Data for Suspicious users-3.jpg

```
Info Data for @Dolores09072598

Name:                        Dolores Nunez
Name Length:                 13
@user number Numeric Chars: 8
@user and name similarity:  0
User Id:                     874119200205418497
Default_Photo:               True
Photo URL:                   http://abs.twimg.com/sticky/default_profile_im
ages/default_profile_normal.png
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             False
Description:
Description length:          0
Account Active:              Active

Number of Tweets analyzed:          431
Last Tweet:                         2017-11-30T18:54:12.000000000
Older Tweet Analyzed:               2017-11-03T04:13:46.000000000
Account Active age (in days)        171.0
Average recent tweets per day:      15.96
Maximum Recent Tweets Per Day:      113 at day 23
Average tweets per day since start: 0.56
Total Tweets since start:           96
Followers_count:                    28
Friends_following_count             300
Followers/Following_index:          0.09
Favourites (Like) count:            26
Retweet_Like_index:                 0.0
Retweet Ratio:                      100%
unique mentions cnt:                384
Mention Ratio:                      100%
Mention index:                      0.44
unique hashtags cnt:                79
Hashtag Ratio:                      28%
Hashtag Index:                      0.49
Average URLs per tweet:             0.14
Average text length per tweet:      129.0

       [Tweets Sample at Peak Day]
 1017117    RT @KiltroKaniechna: Quiero que gane Guillier,...
1017327    RT @pupi_oyanedel: Solo quiero a @guillier Pre...
1018146    RT @simonaxm: #LasCarasDeLaMoneda\n#GuillierGi...
1018208    RT @T13: 🔶 #LasCarasdelaMoneda | @guillier: "L...
1018312    RT @JovenesGuillier: Razones sobran, #SúmateAG...
1018366    RT @BelaAnjali: @aguilo_sergio @danieljadue @c...
1018399    RT @lufernavi: @sebasantanders @camila_vallejo...
1019101    RT @ChristianPolo7: #LasCarasDeLaMoneda Guilli...
1019631    RT @lorpalomita: @cron1411 @Coke92 Por 💌 a @mb...
1020345    RT @CARLOUREA: Vamos Alejandro Guillier. Todo...
```

Appendix D

Data for Suspicious users-4.jpg

```
Info Data for @AlbertoMayol

Name:                        Alberto Mayol
Name Length:                 13
@user number Numeric Chars: 0
@user and name similarity:  0
User Id:                     337542243
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/8927804748
29078530/4VeDYUuZ_normal.jpg
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             True
Description:                 Papá de Alessandro, esposo de Claudia, militan
te del Frente Amplio y de la U. Con el corazón a la izquierda.
Description length:          108
Account Active:              Active

Number of Tweets analyzed:          6930
Last Tweet:                         2017-11-28T15:55:56.000000000
Older Tweet Analyzed:               2017-11-13T01:51:06.000000000
Account Active age (in days)        2,325.0
Average recent tweets per day:      462.0
Maximum Recent Tweets Per Day:      5163 at day 13
Average tweets per day since start: 2.26
Total Tweets since start:           5,265
Followers_count:                    168,523
Friends_following_count             479
Followers/Following_index:          351.82
Favourites (Like) count:            443
Retweet_Like_index:                 0.0
Retweet Ratio:                      0%
unique mentions cnt:                6
Mention Ratio:                      75%
Mention index:                      0.0
unique hashtags cnt:                2
Hashtag Ratio:                      20%
Hashtag Index:                      0.0
Average URLs per tweet:             0.0
Average text length per tweet:      185.0

       [Tweets Sample at Peak Day]
 403036    Impresionante que @sebastianpinera sea incapaz...
403061    Impresionante que @sebastianpinera sea incapaz...
403075    Impresionante que @sebastianpinera sea incapaz...
403078    Impresionante que @sebastianpinera sea incapaz...
403083    Impresionante que @sebastianpinera sea incapaz...
403085    Impresionante que @sebastianpinera sea incapaz...
403095    Impresionante que @sebastianpinera sea incapaz...
403097    Impresionante que @sebastianpinera sea incapaz...
403098    Impresionante que @sebastianpinera sea incapaz...
403100    Impresionante que @sebastianpinera sea incapaz...
```

Appendix E

Data for Suspicious users-5.jpg

```
Info Data for @Tomaskovacic

Name:                        Tomas Iturbe Covacic
Name Length:                 20
@user number Numeric Chars:  0
@user and name similarity:   0
User Id:                     94687772
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/8925741438
40796672/j37U6wrm_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    Viña del Mar
default_profile:             False
Has Description:             True
Description:                 Ingeniero Comercial | Economista en formación
| Coordinador Nacional de @LaBrujulaCL | Viñamarino y Cruzado
Description length:          107
Account Active:              Active

Number of Tweets analyzed:          2062
Last Tweet:                         2017-11-30T16:06:34.000000000
Older Tweet Analyzed:               2017-11-01T15:58:23.000000000
Account Active age (in days)        2,917.0
Average recent tweets per day:      71.1
Maximum Recent Tweets Per Day:      1962 at day 21
Average tweets per day since start: 19.98
Total Tweets since start:           58,284
Followers_count:                    2,302
Friends_following_count             510
Followers/Following_index:          4.51
Favourites (Like) count:            619
Retweet_Like_index:                 0.0
Retweet Ratio:                      1%
unique mentions cnt:                58
Mention Ratio:                      5%
Mention index:                      0.32
unique hashtags cnt:                13
Hashtag Ratio:                      1%
Hashtag Index:                      0.31
Average URLs per tweet:             0.0
Average text length per tweet:      243.0

       [Tweets Sample at Peak Day]
 819662    @HalunkeValnor Compa, Bitar lleva más de dos m...
839341    RT @betojandro: De verdad un saco de wea tuite...
842770    RT @BeaSanchezYTu: "Yo no quiero un gobierno d...
851582    RT @RedGuillier: "El país habló fuerte y claro...
854504    @olivares3891 @NuevaMayoriacl @guillier Te fal...
875591    Ya, lean los cambios del comando nacional de G...
895915    1988: "Si gana el NO seremos Cuba"\n1989: "Si ...
896126    1988: "Si gana el NO seremos Cuba"\n1989: "Si ...
896135    1988: "Si gana el NO seremos Cuba"\n1989: "Si ...
```

Appendix F

Data for Suspicious users-6.jpg

```
Info Data for @andres20ad

Name:                        Andrés
Name Length:                 6
@user number Numeric Chars:  2
@user and name similarity:   0
User Id:                     762803779737616384
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/7628061062
84515328/y4AvLMUb_normal.jpg
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             True
Description:                 Uno de muchos, que luchan por mejorar este mun
do, con mayor justicia, solidaridad e igualdad
Description length:          92
Account Active:              Active

Number of Tweets analyzed:          3060
Last Tweet:                         2017-11-30T23:43:09.000000000
Older Tweet Analyzed:               2017-11-03T02:54:41.000000000
Account Active age (in days)        478.0
Average recent tweets per day:      113.33
Maximum Recent Tweets Per Day:      1583 at day 18
Average tweets per day since start: 18.93
Total Tweets since start:           9,049
Followers_count:                    6,464
Friends_following_count             6,208
Followers/Following_index:          1.04
Favourites (Like) count:            8,939
Retweet_Like_index:                 0.0
Retweet Ratio:                      0%
unique mentions cnt:                69
Mention Ratio:                      5%
Mention index:                      0.27
unique hashtags cnt:                10
Hashtag Ratio:                      75%
Hashtag Index:                      0.0
Average URLs per tweet:             0.09
Average text length per tweet:      271.0

       [Tweets Sample at Peak Day]
 579384    En 1988 millones de chilenos se enfrentaron a ...
579419    En 1988 millones de chilenos se enfrentaron a ...
579476    En 1988 millones de chilenos se enfrentaron a ...
579478    En 1988 millones de chilenos se enfrentaron a ...
579492    En 1988 millones de chilenos se enfrentaron a ...
579503    En 1988 millones de chilenos se enfrentaron a ...
579527    En 1988 millones de chilenos se enfrentaron a ...
579556    En 1988 millones de chilenos se enfrentaron a ...
579579    El gobierno de Piñera fue el más corrupto de t...
579593    El gobierno de Piñera fue el más corrupto de t...
```

Appendix G

Data for Suspicious users-7.jpg

```
Info Data for @NelsonCL28

Name:                         Nelson Sánchez G.
Name Length:                  17
@user number Numeric Chars:   2
@user and name similarity:    0
User Id:                      175420136
Default_Photo:                False
Photo URL:                    http://pbs.twimg.com/profile_images/9323402979
16387328/uUSFt7NI_normal.jpg
Acc Verified:                 False
Has Location:                 True
Location:                     Valparaíso, Chile
default_profile:              False
Has Description:              True
Description:                  Praise Kek!
Description length:           11
Account Active:               Active

Number of Tweets analyzed:           1669
Last Tweet:                          2017-11-19T20:20:33.000000000
Older Tweet Analyzed:                2017-11-19T20:20:33.000000000
Account Active age (in days)         2,662.0
Average recent tweets per day:       inf
Maximum Recent Tweets Per Day:       1669 at day 19
Average tweets per day since start:  0.02
Total Tweets since start:            50
Followers_count:                     26
Friends_following_count              22
Followers/Following_index:           1.18
Favourites (Like) count:             36
Retweet_Like_index:                  0.0
Retweet Ratio:                       0%
unique mentions cnt:                 0
Mention Ratio:                       0%
Mention index:                       0.0
unique hashtags cnt:                 4
Hashtag Ratio:                       100%
Hashtag Index:                       0.0
Average URLs per tweet:              1.0
Average text length per tweet:       279.0

       [Tweets Sample at Peak Day]
 663850    Impresentable suplantación de identidad! Concu...
663892    Impresentable suplantación de identidad! Concu...
663919    Impresentable suplantación de identidad! Concu...
664029    Impresentable suplantación de identidad! Concu...
664114    Impresentable suplantación de identidad! Concu...
664266    Impresentable suplantación de identidad! Concu...
664318    Impresentable suplantación de identidad! Concu...
664340    Impresentable suplantación de identidad! Concu...
664422    Impresentable suplantación de identidad! Concu...
664528    Impresentable suplantación de identidad! Concu...
```

Appendix H

Data for Suspicious users-8.jpg

```
Info Data for @Ivonomas

Name:                        Ivo Barrientos
Name Length:                 14
@user number Numeric Chars: 0
@user and name similarity:  0
User Id:                     1677545442
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/3788000003
11704613/57dba5d1d296053000c81f1acbf9cd60_normal.jpeg
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             True
Description:                 periodista sin título que busca solamente dese
nmascarar la demagogia que abunda en este espacio. Si me mencionas te hago
RT.
Description length:          124
Account Active:              Active

Number of Tweets analyzed:          299
Last Tweet:                         2017-11-30T21:03:14.000000000
Older Tweet Analyzed:               2017-11-01T07:09:21.000000000
Account Active age (in days)        1,566.0
Average recent tweets per day:      10.31
Maximum Recent Tweets Per Day:      40 at day 30
Average tweets per day since start: 4.86
Total Tweets since start:           7,618
Followers_count:                    2,149
Friends_following_count             3,254
Followers/Following_index:          0.66
Favourites (Like) count:            528
Retweet_Like_index:                 0.0
Retweet Ratio:                      1%
unique mentions cnt:                3
Mention Ratio:                      1%
Mention index:                      1.0
unique hashtags cnt:                9
Hashtag Ratio:                      14%
Hashtag Index:                      0.21
Average URLs per tweet:             0.9
Average text length per tweet:      137.0

       [Tweets Sample at Peak Day]
 1515720    Ex ministro de Piñera compartió Photoshop de M...
1528400    .PamJiles por eventual apoyo del FA a Guillier...
1528401    Felipe Kast hace llamado a Alejandro Guillier ...
1528420    RD no se sumará a campaña de Guillier: "Querem...
1533205    VIDEO | Guillier: "El Frente Amplio quiere no ...
1536375    "Cuánta hipocresía en una foto": la imagen que...
1538726    Luis Mesina pide a Alejandro Guillier aclarar ...
1540992    Pamela Jiles, diputada electa del Frente Ampli...
1543988    "No+AFP" se moviliza en la Alameda antes de cr...
```

Appendix I

Data for Suspicious users-9.jpg

```
Info Data for @EncuestaExpress

Name:                        #EncuestaExpressCL
Name Length:                 18
@user number Numeric Chars: 0
@user and name similarity:  0
User Id:                     921483211514970114
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9240128801
64540417/cLUUNNEv_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    Chile
default_profile:             False
Has Description:             True
Description:                 Medio Independiente.
#NoTeQuedesSinVoz!!
Si no votas ¿Con qué cara pides cambios en la sociedad?
#ElPoderEstaEnTuVoto
Description length:          117
Account Active:              Active

Number of Tweets analyzed:          1314
Last Tweet:                         2017-11-26T16:50:34.000000000
Older Tweet Analyzed:               2017-11-01T00:19:55.000000000
Account Active age (in days)        36.0
Average recent tweets per day:      52.56
Maximum Recent Tweets Per Day:      748 at day 20
Average tweets per day since start: 9.61
Total Tweets since start:           346
Followers_count:                    32
Friends_following_count             45
Followers/Following_index:          0.71
Favourites (Like) count:            83
Retweet_Like_index:                 0.0
Retweet Ratio:                      3%
unique mentions cnt:                43
Mention Ratio:                      90%
Mention index:                      0.02
unique hashtags cnt:                34
Hashtag Ratio:                      81%
Hashtag Index:                      0.02
Average URLs per tweet:             0.08
Average text length per tweet:      138.0

       [Tweets Sample at Pick Day]
 761035    Esta elección se define en Segunda:\n\n¿POR QU...
761110    Esta elección se define en Segunda:\n\n¿POR QU...
761414    [ACARREO DE VOTOS]\n¿Cuántos de los votos de @...
761415    Esta elección se define en Segunda:\n\n¿POR QU...
761460    [ACARREO DE VOTOS]\n¿Cuántos de los votos de @...
761466    Esta elección se define en Segunda:\n\n¿POR QU...
761493    Esta elección se define en Segunda:\n\n¿POR QU...
761516    Esta elección se define en Segunda:\n\n¿POR QU...
```

Appendix J

Data for Suspicious users-10.jpg

```
Info Data for @RResponsablecl

Name:                        Reg Responsable Cl
Name Length:                 18
@user number Numeric Chars: 0
@user and name similarity:  0
User Id:                     924982201434898433
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9249862491
43836674/DKYiJprG_normal.jpg
Acc Verified:                False
Has Location:                False
Location:
default_profile:             False
Has Description:             True
Description:                 Buscamos transformar la política de drogas de
Chile para regular democráticamente las vías de acceso al Cannabis medicin
al y para el uso adulto.
Description length:          144
Account Active:              Active

Number of Tweets analyzed:          422
Last Tweet:                         2017-11-26T12:57:43.000000000
Older Tweet Analyzed:               2017-11-01T01:06:03.000000000
Account Active age (in days)        27.0
Average recent tweets per day:      16.88
Maximum Recent Tweets Per Day:      162 at day 26
Average tweets per day since start: 1.37
Total Tweets since start:           37
Followers_count:                    34
Friends_following_count             345
Followers/Following_index:          0.1
Favourites (Like) count:            2
Retweet_Like_index:                 0.0
Retweet Ratio:                      3%
unique mentions cnt:                51
Mention Ratio:                      72%
Mention index:                      0.06
unique hashtags cnt:                24
Hashtag Ratio:                      53%
Hashtag Index:                      0.04
Average URLs per tweet:             1.27
Average text length per tweet:      223.0

       [Tweets Sample at Pick Day]
 1251424    El Candidato Alejandro Guillier reafirma su co...
1251944    Alejandro @Guillier reafirma su compromiso por...
1253536    Alejandro @Guillier reafirma su compromiso por...
1253544    Alejandro @Guillier reafirma su compromiso por...
1253608    Alejandro @Guillier reafirma su compromiso por...
1253681    Alejandro @Guillier reafirma su compromiso por...
1253699    Alejandro @Guillier reafirma su compromiso por...
1253717    Alejandro @Guillier reafirma su compromiso por...
1253839    Alejandro @Guillier reafirma su compromiso por...
```

Appendix K

Data for Suspicious users-11.jpg

```
Info Data for @Piagutierrezs

Name:                        Pilar Gutierrez Soto
Name Length:                 20
@user number Numeric Chars:  0
@user and name similarity:   0
User Id:                     903026378903343110
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9121626151
61495553/JvaFizmK_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    Puerto Varas, Chile
default_profile:             True
Has Description:             True
Description:                 Dirigente social
Description length:          16
Account Active:              Active

Number of Tweets analyzed:          419
Last Tweet:                         2017-11-28T22:36:15.000000000
Older Tweet Analyzed:               2017-11-01T03:08:48.000000000
Account Active age (in days)        89.0
Average recent tweets per day:      15.52
Maximum Recent Tweets Per Day:      131 at day 1
Average tweets per day since start: 19.24
Total Tweets since start:           1,712
Followers_count:                    1,030
Friends_following_count             2,539
Followers/Following_index:          0.41
Favourites (Like) count:            1,829
Retweet_Like_index:                 0.0
Retweet Ratio:                      5%
unique mentions cnt:                57
Mention Ratio:                      66%
Mention index:                      0.08
unique hashtags cnt:                6
Hashtag Ratio:                      7%
Hashtag Index:                      0.15
Average URLs per tweet:             0.9
Average text length per tweet:      154.0

       [Tweets Sample at Pick Day]
 6201    Piñera miente el país no se cae a pedazos GRAN...
6251    Piñera miente el país no se cae a pedazos GRAN...
6272    Piñera miente el país no se cae a pedazos GRAN...
6629    Piñera miente el país no se cae a pedazos GRAN...
6746    Piñera miente el país no se cae a pedazos GRAN...
6781    Piñera miente el país no se cae a pedazos GRAN...
6833    Piñera miente el país no se cae a pedazos GRAN...
6868    Piñera miente el país no se cae a pedazos GRAN...
6946    Piñera miente el país no se cae a pedazos GRAN...
7055    Piñera miente el país no se cae a pedazos GRAN...
```

Appendix L

Data for Suspicious users-12.jpg

```
Info Data for @NathalySeplved3

Name:                        Sepúlveda Natha
Name Length:                 15
@user number Numeric Chars: 1
@user and name similarity:  0
User Id:                     922302808791150592
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9223118740
74509313/GO9ffztI_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    Metropolitana de Santiago, Chi
default_profile:             True
Has Description:             True
Description:                 Más zurda que ayer!!!

Mi otra cuenta @nathalyseplved2   está sin funcionar,. nos leemos por acá.
Description length:          97
Account Active:              Active

Number of Tweets analyzed:          634
Last Tweet:                         2017-11-30T21:14:41.000000000
Older Tweet Analyzed:               2017-11-01T17:44:05.000000000
Account Active age (in days)        38.0
Average recent tweets per day:      21.86
Maximum Recent Tweets Per Day:      165 at day 20
Average tweets per day since start: 14.61
Total Tweets since start:           555
Followers_count:                    294
Friends_following_count             520
Followers/Following_index:          0.57
Favourites (Like) count:            315
Retweet_Like_index:                 0.0
Retweet Ratio:                      10%
unique mentions cnt:                94
Mention Ratio:                      87%
Mention index:                      0.13
unique hashtags cnt:                53
Hashtag Ratio:                      84%
Hashtag Index:                      0.06
Average URLs per tweet:             0.2
Average text length per tweet:      176.0

       [Tweets Sample at Pick Day]
 729046    RT @mtnancyts: @guillier @BeaSanchezYTu @marco...
730568    Los Zurdos de verdad deben unirse y dar el eje...
731302    Los Zurdos de verdad deben unirse y dar el eje...
732178    Los Zurdos de verdad deben unirse y dar el eje...
734723    Los Zurdos de verdad deben unirse y dar el eje...
734861    Los Zurdos de verdad deben unirse y dar el eje...
735795    Los Zurdos de verdad deben unirse y dar el eje...
735890    Los Zurdos de verdad deben unirse y dar el eje...
736092    Los Zurdos de verdad deben unirse y dar el eje...
```

Appendix M

Data for Suspicious users-13.jpg

```
Info Data for @ElCentinelaMPE

Name:                        El Centinela
Name Length:                 12
@user number Numeric Chars:  0
@user and name similarity:   0
User Id:                     921445622678478848
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9214464398
46383616/VJVnqA3F_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    Santiago, Chile
default_profile:             False
Has Description:             True
Description:                 Noticias y actualidad desde el Magíster en Per
iodismo UC-El Mercurio.
Description length:          69
Account Active:              Active

Number of Tweets analyzed:          141
Last Tweet:                         2017-11-20T00:44:21.000000000
Older Tweet Analyzed:               2017-11-02T17:00:04.000000000
Account Active age (in days)        30.0
Average recent tweets per day:      8.29
Maximum Recent Tweets Per Day:      114 at day 19
Average tweets per day since start: 0.47
Total Tweets since start:           14
Followers_count:                    9
Friends_following_count             25
Followers/Following_index:          0.36
Favourites (Like) count:            1
Retweet_Like_index:                 0.0
Retweet Ratio:                      1%
unique mentions cnt:                13
Mention Ratio:                      30%
Mention index:                      0.15
unique hashtags cnt:                13
Hashtag Ratio:                      90%
Hashtag Index:                      0.08
Average URLs per tweet:             0.74
Average text length per tweet:      166.0

       [Tweets Sample at Pick Day]
 620134    En las mesas de China y Hong Kong:\nPiñera 88 ...
624504    Rodeado de prensa y adherentes, Sebastián Piñe...
626233    Piñera sale resguardado tras votar. Afuera de ...
626646    En Viña del Mar, a esta hora vota el senador F...
626815    En Viña del Mar, a esta hora vota el senador F...
626863    En Viña del Mar, a esta hora vota el senador F...
626997    En Viña del Mar, a esta hora vota el senador F...
627004    En Viña del Mar, a esta hora vota el senador F...
627317    En Antofagasta, acaba de votar Alejandro Guill...
627339    En Viña del Mar, a esta hora vota el senador F...
```

Appendix N

Data for Suspicious users-14.jpg

```
Info Data for @PamelaSoler3

Name:                        gitanilla andaluza
Name Length:                 18
@user number Numeric Chars:  1
@user and name similarity:   0
User Id:                     909141300221501443
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9264312552
89266177/CWlqixNW_normal.jpg
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             True
Description:                 chilena , mamá y esposa ,feliz con mi vida , l
ibre pensante ..aprendiz de bruja y por sobre todo de IZQUIERDA y este es
el medio que comparto mis ideas!
Description length:          152
Account Active:              Not Active

Number of Tweets analyzed:          136
Last Tweet:                         2017-11-30T22:36:22.000000000
Older Tweet Analyzed:               2017-11-07T01:43:58.000000000
Account Active age (in days)        75.0
Average recent tweets per day:      5.91
Maximum Recent Tweets Per Day:      82 at day 23
Average tweets per day since start: 8.33
Total Tweets since start:           625
Followers_count:                    354
Friends_following_count             353
Followers/Following_index:          1.0
Favourites (Like) count:            2,131
Retweet_Like_index:                 0.0
Retweet Ratio:                      1%
unique mentions cnt:                31
Mention Ratio:                      65%
Mention index:                      0.14
unique hashtags cnt:                7
Hashtag Ratio:                      59%
Hashtag Index:                      0.08
Average URLs per tweet:             0.17
Average text length per tweet:      194.0

       [Tweets Sample at Pick Day]
 1060250    Que diferencia entre una familia que se ama po...
1060439    Que diferencia entre una familia que se ama po...
1061100    Que diferencia entre una familia que se ama po...
1061219    Que diferencia entre una familia que se ama po...
1061342    Que diferencia entre una familia que se ama po...
1061622    Que diferencia entre una familia que se ama po...
1061640    Que diferencia entre una familia que se ama po...
1062502    Que diferencia entre una familia que se ama po...
1062644    Que diferencia entre una familia que se ama po...
```

Appendix O

Data for Suspicious users-15.jpg

```
Info Data for @AShumman

Name:                        Alexander Shumman
Name Length:                 17
@user number Numeric Chars:  0
@user and name similarity:   0
User Id:                     914950957712146432
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9150016574
06173185/XHQaAxZ0_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    ogazparan
default_profile:             True
Has Description:             True
Description:                 autodidacta
Description length:          11
Account Active:              Active

Number of Tweets analyzed:          371
Last Tweet:                         2017-11-30T03:46:18.000000000
Older Tweet Analyzed:               2017-11-02T19:51:44.000000000
Account Active age (in days)        58.0
Average recent tweets per day:      13.74
Maximum Recent Tweets Per Day:      122 at day 23
Average tweets per day since start: 115.31
Total Tweets since start:           6,688
Followers_count:                    308
Friends_following_count             4,164
Followers/Following_index:          0.07
Favourites (Like) count:            6,407
Retweet_Like_index:                 0.0
Retweet Ratio:                      99%
unique mentions cnt:                276
Mention Ratio:                      100%
Mention index:                      0.41
unique hashtags cnt:                63
Hashtag Ratio:                      41%
Hashtag Index:                      0.3
Average URLs per tweet:             0.17
Average text length per tweet:      127.0

       [Tweets Sample at Pick Day]
 1017505    RT @Tinnnto: Acá esperando a que empiece #LasC...
1017565    RT @JParadaHoyl: A punto de partir #LasCarasDe...
1017574    RT @JMAS_Region: ¡En breve! Nuestro candidato ...
1017597    RT @MPialfaro: Esperando la entrevista de nues...
1017630    RT @pellegrini_jj: #LasCarasDeLaMoneda esta fo...
1017653    RT @Sumate_Guillier: Hoy 22.30 hrs! El candida...
1017669    RT @_Garrettt_: #LasCarasDeLaMoneda lo únicos ...
1018664    RT @canal13: Amigos recuerden, hoy después de ...
1018998    RT @CiudadanosxG: #Ahora #ElPresidenteDeLaGent...
1019105    RT @Sumate_Guillier: 🆙 Hoy 22:30 !!!!! El cand...
```

Appendix P

Data for Suspicious users-16.jpg

```
Info Data for @Sumate_Guillier

Name:                        #SúmateAGuillier
Name Length:                 16
@user number Numeric Chars: 0
@user and name similarity:  0
User Id:                     933384768804110336
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9333854407
94632192/MwSY39xo_normal.jpg
Acc Verified:                False
Has Location:                True
Location:                    Chile
default_profile:             True
Has Description:             True
Description:                 Súmate a apoyar la candidatura del próximo Pre
sidente de Chile, Alejandro Guillier.
Description length:          83
Account Active:              Not Active

Number of Tweets analyzed:          530
Last Tweet:                         2017-11-29T13:47:07.000000000
Older Tweet Analyzed:               2017-11-22T18:17:45.000000000
Account Active age (in days)        6.0
Average recent tweets per day:      88.33
Maximum Recent Tweets Per Day:      195 at day 23
Average tweets per day since start: 0.33
Total Tweets since start:           2
Followers_count:                    18
Friends_following_count             32
Followers/Following_index:          0.56
Favourites (Like) count:            1
Retweet_Like_index:                 17.0
Retweet Ratio:                      3%
unique mentions cnt:                15
Mention Ratio:                      53%
Mention index:                      0.03
unique hashtags cnt:                11
Hashtag Ratio:                      79%
Hashtag Index:                      0.02
Average URLs per tweet:             0.55
Average text length per tweet:      159.0

       [Tweets Sample at Pick Day]
 1011443    Hoy 22.30 hrs! El candidato de la gente @guill...
1011591    Hoy 22.30 hrs! El candidato de la gente @guill...
1011640    Hoy 22.30 hrs! El candidato de la gente @guill...
1011706    Hoy 22.30 hrs! El candidato de la gente @guill...
1012063    Hoy 22.30 hrs! El candidato de la gente @guill...
1012223    Hoy 22.30 hrs! El candidato de la gente @guill...
1012258    Hoy 22.30 hrs! El candidato de la gente @guill...
1012924    Hoy 22.30 hrs! El candidato de la gente @guill...
1013534    Hoy 22.30 hrs! El candidato de la gente @guill...
1013561    Hoy 22.30 hrs! El candidato de la gente @guill...
```

Appendix Q

Data for Suspicious users-17.jpg

```
Info Data for @jav_ast

Name:                        Javier Astudillo
Name Length:                 16
@user number Numeric Chars: 0
@user and name similarity:  0
User Id:                     934771590583279616
Default_Photo:               False
Photo URL:                   http://pbs.twimg.com/profile_images/9347807861
58592000/dj9mU4xY_normal.jpg
Acc Verified:                False
Has Location:                False
Location:
default_profile:             True
Has Description:             True
Description:                 Estudiante de Medicina UCH - Salud Artes y Cul
tura Comida Política y Gobierno
Description length:          77
Account Active:              Active

Number of Tweets analyzed:          104
Last Tweet:                         2017-11-27T12:03:15.000000000
Older Tweet Analyzed:               2017-11-27T11:38:45.000000000
Account Active age (in days)        0.0
Average recent tweets per day:      inf
Maximum Recent Tweets Per Day:      104 at day 27
Average tweets per day since start: inf
Total Tweets since start:           7
Followers_count:                    42
Friends_following_count             104
Followers/Following_index:          0.4
Favourites (Like) count:            0
Retweet_Like_index:                 0
Retweet Ratio:                      2%
unique mentions cnt:                7
Mention Ratio:                      100%
Mention index:                      0.04
unique hashtags cnt:                3
Hashtag Ratio:                      98%
Hashtag Index:                      0.01
Average URLs per tweet:             0.0
Average text length per tweet:      155.0

       [Tweets Sample at Pick Day]
 1309981    RT @kitaalarcon: @sebastianpinera Porque mejor...
1310246    #ESPUCh analiza las propuestas de los candidat...
1310282    #ESPUCh analiza las propuestas de los candidat...
1310294    #ESPUCh analiza las propuestas de los candidat...
1310321    #ESPUCh analiza las propuestas de los candidat...
1310331    #ESPUCh analiza las propuestas de los candidat...
1310349    #ESPUCh analiza las propuestas de los candidat...
1310354    #ESPUCh analiza las propuestas de los candidat...
1310380    #ESPUCh analiza las propuestas de los candidat...
1310467    #ESPUCh analiza las propuestas de los candidat...
```