Identifying Botnets by Analysing Twitter Traffic during the Super Bowl

Salah Safi[1], Huthaifa Jawazneh[1], Antonio Mora[1], Pablo García[1], Hossam Faris[2] and Pedro A. Castillo[1]

[1]*School of Informatics and Telecommunications Engineering, University of Granada, Spain*

[2]*The University of Jordan, Amman, Jordan*

Keywords: Social Network Analysis, Twitter Data, Classification, Illegal Content Broadcasting, Botnets.

Abstract: Detecting accounts broadcasting illegal contents at sporting events in social networks is an important problem of difficult solution, since the traditional intrusion detection systems are not effective in online social networks due to the speed with which these kind of messages and contents spread. Thus, there is an increasing need for an adequate and efficient detection system of the so-called botnets used for the distribution of illegal contents on online social networks. In this paper we propose using well-known classification methods to analyse the activity of Twitter accounts in order to identify botnets. We have analysed the Twitter conversations that include hashtags related to the Super Bowl LIII (February 3, 2019). The objective is to identify the behaviour of various types of profiles with automatic and non-standard spamming activities. In order to do so, a dataset from public data available on Twitter that includes content published by human-managed accounts and also by bots that used hashtags related to the event has been collected. This dataset has been manually labelled to create a training set with tweets posted by humans and bots active in Twitter. As a result, several types of profiles with non standard activities have been identified. Also, some groups of accounts have been identified as botnets that were activated during the Super Bowl LIII (2019).

1 INTRODUCTION

Nowadays, millions of users on Twitter use the social network to share opinions, interests and content of all kinds, to keep in touch with each other, or to report news about events (Liu et al., 2016; Compton et al., 2013).

Among those accounts, also millions of software-controlled accounts (social bots), that generate content and interactions, operate in the online social network (OSN). While some of those bots perform useful and legal functions, such as disseminating news (Lokot and Diakopoulos, 2016; Haustein et al., 2016; Savage et al., 2016), others form networks of compromised computers, connected in a coordinated way and controlled by a third party (botnets) that use social media platforms as a communication way to achieve illegitimate objectives, such as influencing public opinion, promoting terrorist propaganda, perform distributed denial-of-service attacks, steal data, or simply spamming (Berger and Morgan, 2015; Abokhodair et al., 2015; Ferrara et al., 2016a; Ferrara et al., 2016b). Furthermore, a high percentage of the traffic of social networks spam contains links to malicious contents.

Given the high traffic that sports evens generate on OSNs, attracting the attention of millions of people, some organizations take advantage of the OSN structure of Twitter to host different botnets aiming to disseminate illegal content about those events by including connection links in tweets.

This work has as main objective to propose a method for the detection of botnets that disseminate illegal contents by analysing Twitter traffic using standard classification techniques in order to identify botnets related to the illegal diffusion of sporting events.

To this end, a comprehensive analysis using classification and data visualization methods has been carried out on a dataset collected considering the hashtags defined by users in Twitter during the Super Bowl LIII (February 3, 2019). After an intensive experimentation it has been verified the characteristic behavior of accounts managed by software, as reported in (Das et al., 2016; Varol et al., 2017), verifying the validity of the experiments.

The remainder of this paper is organized as follows: Section 2 presents the state of the art on the botnet detection problem. Section 3 describes the data collecting process using the Twitter API as well as the datasets used in this paper. Section 4 explains the

Safi, S., Jawazneh, H., Mora, A., García, P., Faris, H. and Castillo, P.
Identifying Botnets by Analysing Twitter Traffic during the Super Bowl.
DOI: 10.5220/0010022301470154
In *Proceedings of the 12th International Joint Conference on Computational Intelligence (IJCCI 2020)*, pages 147-154
ISBN: 978-989-758-475-6

proposed approach to identify botnets broadcasting illegal content. Experiments and obtained results are presented in Section 5, followed by the conclusions reached and future lines of work commented in Section 6.

2 RELATED WORK

Traditionally approaches to detect botnets (networks of compromised computers, connected in a coordinated way and controlled by a third party for malicious purposes) have been focused on data and sentiment analysis (Takeichi et al., 2015; Mathews et al., 2018). For instance intrusion analysis systems (IDS) have commonly been based on inspecting flow anomalies or protocols, and server blacklists.

However, recently some organizations have used botnets to coordinate illegal activities using OSN platforms such as Twitter and Facebook. In those activities, some accounts are used as a control tool to propagate commands used to activate other nodes in the botnet (Romera, 2010; Baltazar, 2010). This way, they can dominate discussions and manipulate opinions about topics such us commercial products (Clark et al., 2016), elections (Bessi and Ferrara, 2016), advertising (Echeverría and Zhou, 2017), or political issues (Abokhodair et al., 2015).

Different methods have been proposed as detection mechanisms, usually based on monitoring the network (Goebel and Holz, 2007), clustering taking into account information about similar attacks (Gorman, 2009) or monitoring changes in DNS-records (Campbell et al., 2011). However, these techniques are not effective in detecting traffic generated by botnets that exploit OSN websites.

Thus, taking into account the importance of this problem, several machine-learning methods have recently been used for the detection of malicious profiles, such as neural networks (Alsaleh et al., 2014) or support vector machine models (Zheng et al., 2015). Other authors propose applying clustering methods (Wang et al., 2013) or classification techniques (Yang et al., 2014) to large datasets collected from the OSN to identify large-scale attacks.

In any case, since the problem of the detection of botnets that broadcast illegal content related to sporting events has not been addressed previously using classification methods, a new method for collecting and analysing OSN data becomes necessary. Thus, considering the kind of problem we address in this paper, in our proposal the published tweets relative to the Super Bowl LIII (2019) have been collected (the dataset has been collected during an important sporting event, which is another point to remark). Next, two datasets have been manually created with suspicious and non-suspicious tweets to train a classifier that has been used later on the complete set of tweets. Finally, after a deep data-analysis, the accounts that may be part of content broadcast botnets have been identified.

3 DATASET: COLLECTING DATA RELATED TO THE SUPER BOWL LIII (2019)

As stated above, the proposed methodology has been applied to a Twitter dataset collected taking into account the hashtag `#SuperBowl`, used by Twitter users during the Super Bowl LIII (February 3, 2019).

Thus, taking that hashtag, the Twitter REST API[1] and TwitteR[2] package were used to download the corresponding tweets that were published one week before the match (since January 26th, 2019), during the match, and one week after the event (until February 10, 2019).

Collected data were stored in a database to be processed later using scripts developed in R[3]. Aiming for reproducibility, both the collected data and the R scripts to process the information have been published in an open repository in GitHub[4]. This can be also considered as a contribution of this work.

In this work, just the information contained in the text of the tweet has been used. After collecting the complete set of tweets, we manually selected a subset and labelled them to form two subsets: one formed with 500 suspicious tweets (those that include "live stream" text plus a link to a illegal streaming web service) and the second one formed with 500 non-suspicious tweets. The procedure for tagging such tweets was based on the tweets including links to illegal content and the inclusion of words to attract the attention of other users to click on such links. Table 1 describes the collected datasets in terms of number of tweets, while Figure 1 shows the number of tweets collected per day using a log scale.

Then, controlled datasets were divided into two disjoint parts, for training the classification methods and testing their accuracy. Thus, 75% of tweets were first used to train the model, and then, 25% of tweets were used to test the accuracy of this model.

[1] https://developer.twitter.com

[2] https://cran.r-project.org/web/packages/twitteR

[3] https://www.r-project.org/

[4] https://github.com/pacastillo/datasetSB2019/

Table 1: List of datasets and details in terms of number of tweets collected. Suspicious and non-suspicious datasets were formed by selecting a number of tweets from the complete set and then they were manually labelled as suspicious (of being a botnet) or not.

Dataset	Number of tweets
complete	1.796.506
suspicious	500
non-suspicious	500

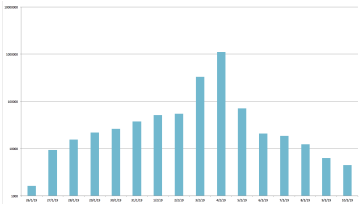

Figure 1: Number of tweets per day (log scale). Figure shows number of tweets since January 26th, 2019 until February 10, 2019. As it can be seen, during the event and the days immediately following, the number of tweets was very high, but decreasing (with public attention) along time.

4 METHODOLOGY

The objective of the proposed method is to identify the behaviour of different types of profiles with automatic and non standard spamming activities on Twitter, broadcasting illegal content related to sporting events. In order to do so, we propose taking the following steps (see Figure 2):

1. Data collection by downloading tweets using the Twitter API.
2. Process obtained data in order to create two datasets, one with suspicious tweets and another one with non-suspicious tweets.
3. Use standard classification methods to create classification models trained with the training dataset.
4. Use the best classification model obtained to identify those accounts that have posted suspicious tweets.
5. Analyse the publication pattern and retweets (RT) of those accounts.
6. Visualize the networks of followers and friends to characterize the botnets according to their structure.

Regarding the classification methods to apply to the collected tweets to separate them in suspicious or non-suspicious, the following algorithms are proposed:

- *Tree model* (Breiman et al., 1984): A decision tree consists of answer-nodes, that indicate a class, and decision-nodes, that contain an attribute name and branches to other sub-decision trees. Building a decision tree can be done using many algorithms, i.e. ID3 and C4.5 (Quinlan, 1986). The construction of a tree model is similar to that of classical decision trees: The process divides the input space of the training data through decision points (nodes) assigning a linear model suitable for that sub-area of the input space. This process may lead to a complex tree model. Tree models can learn and tackle tasks with high dimensionality, even up to hundreds of attributes, and generate reproducible and comprehensible representations. The final model consists of the collection of predicted values at the answer-nodes along with the path from the root to that node given as a result.

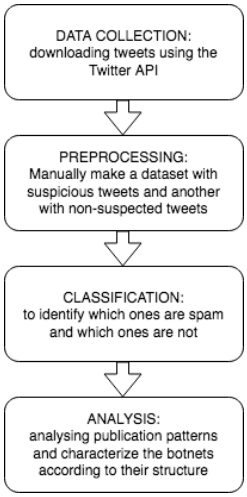

Figure 2: Summary of steps followed in the proposed methodology.

- *Random Forest (RF)*: This method is based on the construction of a set of decision trees using a stochastic process over the basis of C4.5 algorithm. It was proposed by Breiman (Breiman, 2001; Liaw and Wiener, 2001) and aims to create independent and uncorrelated trees based on different and random input vectors, following the same distribution. The result is the average of the generated trees during the process.
- *Support Vector Machine (SVM)*: Support vector machine (SVM) (Cortes and Vapnik, 1995; Min and Lee, 2005) method is based on statistical learning theory. It has been successfully used in classification and regression problems (Jari et al., 2008). In classification problems the algorithm searches for an optimal hyperplane that separates every two classes, maximising the margin between them. The hyperplane is defined by a subset of training set samples, called support vectors.

The developed classification models will be evaluated and assessed based on standard evaluation criteria, i.e. the *accuracy, precision, recall* and *F-measure*, based on the confusion matrix shown in Table 2, and defined according to Equations 1 to 4.

Table 2: Confusion matrix for a binary classifier (TP=true positive; FN=false negative; FP=false positive; TN=true negative).

Actual class	Predicted class	
	suspicious	non-suspicious
suspicious	TP	FN
non-suspicious	FP	TN

- Accuracy (correctly classified cases over the total amount):

$$A = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Precision (percentage of actual classifications that are correct):

$$P = \frac{TP}{TP+FP} \quad (2)$$

- Recall (percentage of possible classifications which are correct):

$$R = \frac{TP}{TP+FN} \quad (3)$$

- F-measure (combination of recall and precision):

$$F = \frac{2x(PxR)}{P+R} \quad (4)$$

In the experiments performed, the objective is designing models that maximise A and F measurements.

5 EXPERIMENTS AND OBTAINED RESULTS

In this section, the setup environment for the experiments conducted over the dataset is described in detail, following the steps detailed in Section 4.

In most data mining and machine learning algorithms, the values of some parameters have a high influence on its performance. After intensive experimentation, the best performance using RF was achieved with a number of trees equal to 50. As far as the TREE model is concerned, the batch size was set to 100 and the minimum number of instances allowed at a leaf node is set to 2. For SVM, a Radial Basis Function kernel is selected due to its reliable performance (Chao and Horng, 2015), and the Cost (C) and Gamma (γ) parameters set to 100 and 0.01 respectively.

All the experiments have been performed using the standard 10-fold cross validation (Kohavi, 1995) in order to estimate how accurately the predictive model will perform in practise, limiting, at the same time, the overfitting problem. In addition, due to the stochastic component present in some of the methods, 30 repetitions of each experiment have been done per method, reporting the average and standard deviation.

After performing all the classification experiments using the proposed methods on the controlled dataset (752 tweets for training and 250 for testing), the obtained results are shown in Table 3.

As it can be seen, the random-forest (RF) classifier performs slightly better than the other methods, achieving 99,78% accuracy, 99,1% precision, 99,1% recall and 99,1% F-measure. In any case, on this problem all classifiers achieve relatively high classification performance.

5.1 Identifying Suspicious Twitter Accounts

Using the best classifier model obtained, the whole dataset was classified between suspicious and non-suspicious tweets. As a result of this process, a total of 3132 tweets were identified as suspected of broadcasting illegal contents. However, a manual inspection was carried out, resulting that a total of 319 tweets were found to be false positive, as those tweets did not disseminate illegal content. The classifier might have found words such as “streaming”, but they do not include links to illegal contents (see Table 4). Finally, a total of 2813 correspond to true positive, as those tweets include links to illegal contents and they have been correctly classified.

Table 4 shows examples of tweets where links to illegal contents related to the football match (above) are broadcasted, also a couple of examples of tweets that do not include them (in the middle), and finally, several false positives (below).

Once the suspicious tweets were identified, the operating patterns of accounts that published them have been analysed in order to establish the structure of the botnets, searching for information about what accounts have done RTs, how often, and what applications have been used. Details on the activity of the **accounts identified as suspicious** are shown below:

- “main” accounts that post tweets with links then receive exactly 5 RTs almost simultaneously. In all cases, each of these accounts performs only one publication during the match, using the “Twitter Web Client” application.
- RTs that are almost simultaneous, using the application “Twitter Web Client”, and around thirty

Table 3: Classification performance of the proposed models on the controlled dataset.

Model	Accuracy	Precision	Recall	F-measure
SVM	$99,62 \pm 0,01$	$98,00 \pm 0,71$	$98,00 \pm 0,71$	$97,6 \pm 0,5$
RF	$99,78 \pm 0,05$	$99,1 \pm 0,2$	$99,1 \pm 0,2$	$99,1 \pm 0,2$
TREE	$99,50 \pm 0,01$	$99,5 \pm 0,0$	$99,5 \pm 0,0$	99 ± 0

Table 4: Examples of tweets identified as suspicious (top), non-suspicious (middle), and false-positive (bottom).

Client	User	Text
Suspicious tweets		
Twitter Web Client	JennyZica	*How to watch 2019 SUPER BOWL: live stream Rams vs Patriots Live HD stream Link Live Stream https://t.co/G0ochlCg0t*
Twitter Web Client	RonaldR28645340	*Watch Live: Patriots Championship Parade 2019 Live Stream ACCESS 100% FREE Patriots Super Bowl LIII victory parade: https://t.co/DiSKB8dkj5*
Non-suspicious tweets		
Twitter for iPhone	dougrutherford	*So, I guess this will be known forever as the Boring Bowl. This is time wasted that I will never get back. #superbowl*
Twitter for iPhone	Pamcakes336	*Only here to watch the @pewdiepie Super Bowl commercial tbh #superbowl*
False-positive tweets		
Twitter Web Client	stockerblog	*#Amazon Takes to #SuperBowl to Boost New Streaming Series #Hanna https://t.co/JnS0yJz8v5 via @variety $AMZN*
Twitter for iPhone	themimicmoment	*#SuperBowl has huge mainstream appeal as stories flood the news wires about betting odds and action. But in that fleeting moment*

seconds after the original tweet is published.

- Main accounts that post tweets with links that then receive exactly 16 RTs almost simultaneously. In all cases, each of these accounts performs only one publication during the match, using the "Twitter Web Client" application.
- The "secondary" accounts that make up the following RT pattern: RTs that are almost simultaneous, using the application "Twitter Web Client", and around one minute after the original tweet is published.
- Main accounts that post tweets with links that then receive exactly five RTs almost simultaneously. In all cases, each of these accounts performs only one publication during the match.
- RTs that were found to be simultaneous using the application "Twitter Web Client", and around one minute after the original tweet is published.
- Main accounts that post tweets with links that then receive exactly 7 RTs almost simultaneously. In all cases, each of these accounts performs only one publication during the match.
- Main accounts that post tweets with links that then receive exactly 7 RTs almost simultaneously. In all cases, each of these accounts performs only one publication during the match.
- Main accounts post tweets with links that then receive exactly 7 RTs almost simultaneously. In all cases, each of these accounts performs only one publication during the match.
- The "secondary" accounts that make up the following RT pattern: RTs found to be almost simultaneous, using the application "Twitter Web Client", and around one minute after the original tweet is published.

Thus, from the analysis of Twitter data presented above, around 82 bot profiles that have been active during the Super Bowl LIII (2019) have been identified. It is important to note that much of the accounts identified as suspicious using the proposed method have been suspended and removed from Twitter to this day: *ciinthiaromero, BS3Sports, dammamy11, JannaHofmann, LulyOficial, errafik76, Mohammmad_11, kristy1445, sportstream1234, Kristenb420, MariaJones424, TinaCring, maax_moon, av910281028120, fightnightlives, onlinestream24, foxxsports24, iva_12segov.*

As stated above, some tweets were classified as suspicious while they do not contain links to illegal content. Once verified, they correspond to accounts

of TV channels whose publication pattern and RT obtained has generated these false-positives. Below are details about the activity of these accounts identified as suspicious but which actually turned out to be **false positives**:

- *wbz, wbzsports, wbznewsradio*: WBZ-TV is a news and sports CBS-Boston-News' channel (cbsboston.com). These accounts posted several tweets and many users made RT using different applications ("Twitter for iPhone", "Twitter Web Client", "Twitter for Android", "TweetDeck", "Twitter for iPad", "Twitter Lite").
- *LiveStream777*: Live Stream Football is a sports English channel (streamlive7.com) that only made two tweets on the day of the match, having received 29 RT from users who used the applications "Twitter for iPhone" and "Twitter for Android".

Paying attention to the previous analysis, it can be seen that those accounts that used botnets to disseminate links to illegal contents about the match follow a very similar pattern (and even use the same set of "secondary" accounts to retweet the main accounts' publications); thus, after a single publication, several RTs are received in a few seconds almost simultaneously. It has been proved that the RT pattern is very important for finding tweets with suspicious content (links to streaming websites) and thus, to identify accounts that use botnets to spam or broadcast illegal contents.

At the same time, some accounts that make legal publications and receive many RTs generated these false-positives. However, in these cases, there was not a specific time pattern obtaining RTs (as in the previous cases). Moreover, not only the time pattern of those RTs is very different, but also the set of accounts that retweeted those tweets are very different (account names do not seem to follow a pattern); so it can be deduced that no botnets have been used in these cases.

As an additional processing, a visualization step has been carried out just to validate the obtained results, verifying that the identified suspicious accounts are part of botnets, as well as to obtain a representation of their structure. The way to identify botnets is by analysing the structure of relationships between profiles. To do this, the followers and friends of accounts identified as disseminators of illegal contents have been obtained, and then the relationships networks are plotted, characterizing the botnets according to their structure: active accounts in a botnet are usually all interconnected, while in a conversation between real people, a few dominant accounts make publications and other secondary-accounts interact with them (elevenpaths, 2018).

Thus, Figure 3 shows the structure of two botnets activated during the Super Bowl LIII. As it can be seen, they have a very regular structure, in which the profiles are interconnected to each other, showing a similar activity.

On the other hand, Figure 4 shows the relationships of accounts that do not belong to botnets. In these cases there is no interconnection between all the accounts, so it can be seen some dominant profiles that make a few publications with many secondary accounts that do RT. They are accounts with many followers in which the main account publishes and the others spread its messages.

6 CONCLUSIONS AND FUTURE WORKS

A new method for detecting botnets by classifying traffic on Twitter has been presented in this work. The proposed method and its effectiveness have been tested using a dataset collected directly from the Twitter social network during the Super Bowl LIII (2019) sporting event.

Experimentation has shown that botnets that spread illegal content can be identified using standard classification methods. Our objective was finding the most appropriate classification method to separate the downloaded tweets (complete dataset) in order to identify which ones are spam and which ones are not (suspicious or non-suspicious), minimising the false positives.

It has been shown that those accounts that broadcast illegal contents about sporting events follow a similar behavioral pattern, making use of the same set of "secondary" accounts that RT tweets published by very specific accounts almost simultaneously in a few seconds.

Thus, detecting that interaction pattern between accounts might be very useful to detect tweets with suspicious content (links to streaming platforms) and thus to identify botnets. Also, after analysing the network structure of the botnets, taking into account friends and followers relations, it has been concluded that in general their shape follows a similar pattern.

Taking into account the obtained results from the analysis of data extracted from Twitter, it has been demonstrated that the proposed method can be very useful to the administrators of an OSN for controlling the broadcasting of illegal contents using botnets. Finally, as a last thought, it should be noted that most

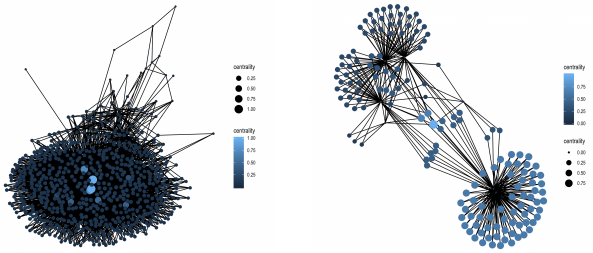

Figure 3: Examples of botnets identified: It can be seen that all the accounts are interconnected with each other, and all of them exhibit similar activity in terms of intensity.

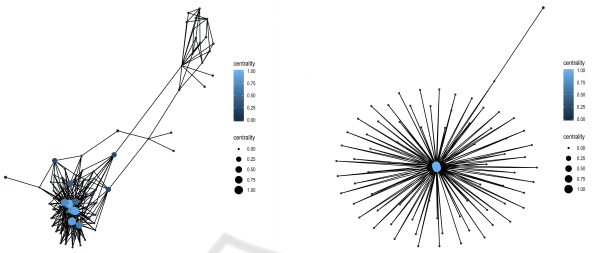

Figure 4: Examples of dominant profiles with a lot of followers that spread their tweets: In these cases there is no interconnection between all the accounts.

of the accounts identified in this work have been suspended and removed from Twitter to this day.

As future work we will focus on the analysis of Twitter data taking into account a greater number of sporting events, e.g. soccer or tennis. As stated above, in this work just the information contained in the text of the tweet has been used. However, for future work we plan to use additional information regarding tweets and the accounts that publish them, such as RT number, number of followers, etc. We also propose the future use of this methodology to identify new botnets, integrating the developed programs into a system to work in real time, instead of using post-processing, and even, using more advanced classification methods, such as deep-learning models.

ACKNOWLEDGMENTS

This work has been supported by the Ministerio español de Economía y Competitividad under project TIN2017-85727-C4-2-P (UGR-DeepBio) and TEC2015-68752 (also funded by FEDER). It has also been supported in part by projects B-TIC-402-UGR18 (FEDER and Junta de Andalucía), and RTI2018-102002-A-I00 (Ministerio Español de Ciencia, Innovación y Universidades).

REFERENCES

Abokhodair, N., Yoo, D., and McDonald, D. W. (2015). Dissecting a social botnet: Growth, content and influence in twitter. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pages 839–851, New York, NY, USA. ACM.

Alsaleh, M., Alarifi, A., Al-Salman, A. M., Alfayez, M., and Almuhaysin, A. (2014). Tsd: Detecting sybil accounts in twitter. In *Proceedings of the 2014 13th International Conference on Machine Learning and Applications*, ICMLA '14, pages 463–469, Washington, DC, USA. IEEE Computer Society.

Baltazar, J. (2010). Web 2.0 botnet evolution – koobface revisited. *Trend Micro*.

Berger, J. and Morgan, J. (2015). *The isis twitter census: Defining and describing the population of isis supporters on twitter*, volume 3. The Brookings Project on US Relations with the Islamic World, 20 edition.

Bessi, A. and Ferrara, E. (2016). Social bots distort the 2016 u.s. presidential election online discussion. *First Monday*, 21(11).

Breiman, L. (2001). Random forests. *Mach Learn*, 45(1):5–32.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.

Campbell, S., Chan, S., and Lee, J. R. (2011). Detection of fast flux service networks. In *Proceedings of the Ninth Australasian Information Security Conference - Vol-*

ume 116, AISC '11, pages 57–66, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

Chao, C.-F. and Horng, M.-H. (2015). The construction of support vector machine classifier using the firefly algorithm. *Computat Intell Neurosci*.

Clark, E. M., Jones, C. A., Williams, J. R., Kurti, A. N., Norotsky, M. C., Danforth, C. M., and Dodds, P. S. (2016). Vaporous marketing: Uncovering pervasive electronic cigarette advertisements on twitter. *PLOS ONE*, 11(7):1–14.

Compton, R., Lee, C., Lu, T.-C., Silva, L. D., and Macy, M. (2013). Detecting future social unrest in unprocessed twitter data: "emerging phenomena and big data". *2013 IEEE International Conference on Intelligence and Security Informatics*, pages 56–60.

Cortes, C. and Vapnik, V. (1995). Support Vector Networks. *Mach Learn*, 20:273–297.

Das, A., Gollapudi, S., Kiciman, E., and Varol, O. (2016). Information dissemination in heterogeneous-intent networks. In *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16, pages 259–268, New York, NY, USA. ACM.

Echeverría, J. and Zhou, S. (2017). The 'star wars' botnet with ¿350k twitter bots. *CoRR*, abs/1701.02405 arXiv preprint arXiv:1701.02405.

elevenpaths (2018). Twitter botnets detection in sports event. https://blog.en.elevenpaths.com/2018/12/detection-botnet-twitter-cybersecurity.html. (visited January 2019).

Ferrara, E., Varol, O., Menczer, F., and Flammini, A. (2016a). Detection of promoted social media campaigns. In *Proceedings of the Tenth International Conference on Web and Social Media, Cologne, Germany, May 17-20, 2016.*, pages 563–566.

Ferrara, E., Wang, W., Varol, O., Flammini, A., and Galstyan, A. (2016b). Predicting online extremism, content adopters, and interaction reciprocity. In *Social Informatics - 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part II*, pages 22–39.

Goebel, J. and Holz, T. (2007). Rishi: Identify bot contaminated hosts by irc nickname evaluation. In *Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets*, HotBots'07, pages 8–8, Berkeley, CA, USA. USENIX Association.

Gorman, G. (2009). Google groups trojan. http://www.symantec.com/connect/blogs/google-groups-trojan. (visited January 2019).

Haustein, S., Bowman, T. D., Holmberg, K., Tsou, A., Sugimoto, C. R., and Larivière, V. (2016). Tweets as impact indicators: Examining the implications of automated "bot" accounts on twitter. *J. Assoc. Inf. Sci. Technol.*, 67(1):232–238.

Jari, C., Wen, L., and Kang, L. (2008). Support vector machine classification for large data set vis minmum enclosing ball clustering. *Neurocomputing*, 71:611–619.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes*, pages 1137–1145. Morgan Kaufmann.

Liaw, A. and Wiener, M. (2001). Classification and regression by randomforest. *Forest*, 23.

Liu, X., Li, Q., Nourbakhsh, A., Fang, R., Thomas, M., Anderson, K., Kociuba, R., Vedder, M., Pomerville, S., Wudali, R., Martin, R., Duprey, J., Vachher, A., Keenan, W., and Shah, S. (2016). Reuters tracer: A large scale system of detecting & verifying real-time news events from twitter. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 207–216, New York, NY, USA. ACM.

Lokot, T. and Diakopoulos, N. (2016). News bots: Automating news and information dissemination on twitter. *Digital Journalism*, 4(6):682–699.

Mathews, P., Gray, C., Mitchell, L., Nguyen, G. T., and Bean, N. G. (2018). SMERC: social media event response clustering using textual and temporal information. *CoRR*, abs/1811.05063.

Min, J. H. and Lee, Y. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert Syst. Appl.*, 28(4):603–614.

Quinlan, J. (1986). Induction of decision trees. *Mach Learn*, 1:181–186.

Romera, R. (2010). Discerning relationships: The mexican botnet connection. *Trend Micro*.

Savage, S., Monroy-Hernandez, A., and Höllerer, T. (2016). Botivist: Calling volunteers to action using online bots. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 813–822, New York, NY, USA. ACM.

Takeichi, Y., Sasahara, K., Suzuki, R., and Arita, T. (2015). Concurrent bursty behavior of social sensors in sporting events. *PloS one*, 10 12:e0144646.

Varol, O., Ferrara, E., Davis, C. A., Menczer, F., and Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. In *International AAAI Conference on Web and Social Media*, page 280–289. AAAI, AAAI.

Wang, G., Konolige, T., Wilson, C., Wang, X., Zheng, H., and Zhao, B. Y. (2013). You are how you click: Clickstream analysis for sybil detection. In *Proceedings of the 22Nd USENIX Conference on Security*, SEC'13, pages 241–256, Berkeley, CA, USA. USENIX Association.

Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B. Y., and Dai, Y. (2014). Uncovering social network sybils in the wild. *ACM Trans. Knowl. Discov. Data*, 8(1):2:1–2:29.

Zheng, X., Zeng, Z., Chen, Z., Yu, Y., and Rong, C. (2015). Detecting spammers on social networks. *Neurocomputing*, 159:27 – 34.