Deep dive on politician impersonating accounts in social media

Koosha Zarei, Reza Farahbakhsh, Noel Crespi

Deep Dive on Politician Impersonating Accounts in Social Media

Koosha Zarei*, Reza Farahbakhsh*, Noël Crespi*
*Institut Mines-Télécom, Télécom SudParis, CNRS Lab UMR5157 Evry, France.
{Koosha.zarei, reza.farahbakhsh, noel.crespi}@telecom-sudparis.eu

***Abstract*—There is an ever-growing number of users who duplicate the social media accounts of celebrities or generally impersonate their presence on online social media as well as Instagram. Of course, this has led to an increasing interest in detecting fake profiles and investigating their behaviour. We begin this research by targeting a few famous politicians, including Donald J. Trump, Barack Obama, and Emmanuel Macron and collecting their activity for the period of 3 months using a specifically-designed crawler across Instagram. We then experimented with several profile characteristics such as username, display name, biography, and profile picture to identify impersonator among 1,5M unique users. Using publicly crawled data, our model was able to distinguish crowds of impersonators and political bots. We continued by providing an analysis of the characteristics and behaviour of these impersonators. Finally, we conclude the analysis by classifying impersonators into four different categories.**

***Index Terms*—Social Network Analysis; Instagram; Impersonation; Political Bot Detection; Fake profile; User profiling.**

I. Introduction

Online social networks (OSNs) are social connections with similar interests over the Internet. This world also contains thousands of accounts are impersonating real people, advertising commercial products, criticizing political candidates, deceptive social media campaigns [1], and sowing discord. These fake accounts spread false images and misinformation on many subjects (Figure 1). A New York Times study in 2018 found that many genuine accounts are copied and turned into automated bots sold by corporations. Many others are deployed in systematic information warfare campaigns conducted by governments [2], [3]. Investigating valuable hidden insights of OSNs [4] in regard to detecting imposters and understanding their behaviour is recognised as the hot research area. Instagram is an online social media platform for sharing visual media. According to Statistica [5], the mainly mobile sharing network is one of the most popular OSNs worldwide and had reached 1 billion active users monthly by 2018. As a result, numerous prominent public figures that are hugely active on Twitter uncovered their official accounts on Instagram. In such a scenario, the analysis of the user engagement and detecting bots are still an active domain of research. By considering genuine politician accounts on Instagram, some interesting questions will arise. What is the rate of user engagement in their shared media in the shape of liking and commenting? How many impersonators exist? and who are they? What is the activity of this group? Are they human or bot? In this study, "*Imposter*", "*Impostor*", and "*Impersonator*" terms are equal.

A. Impersonation and Social Media Profile Theft (SMPT)

An increasingly popular difficulty on OSNs that individuals are forced to deal with each day is the impersonation or Social Media Profile Theft (SMPT) [6]. SMPT takes place when an impostor sets up a fake profile on social media which mimics another user as a prank or to mock them. By using this account, they gain the trust of the original users followers for different purposes such as fake promotions, to generate followers, to gather information, spreads political views, supports or oppose actions etc. Some criminals are using this strategy to deceive the public and commit crimes. They attempt to establish relationships using false facts and then defraud unsuspecting targets. A fake social media account could result in legal action against the impersonator. On Instagram, it can be possible to report a fraud.

Fig. 1. Two samples of impersonators of Donald J. Trump on Instagram. The first snapshot belongs to the genuine account and the others are imposters.

The main contributions of this work are:

- We crawled and built a real dataset of Instagram posts of 3 months user activities including nearly 5M likes and 500K comments. Around 1,5M unique users are included.
- A novel process to identify impersonators on Instagram using profile features and users' activities is introduced.
- Detect a group of impersonator accounts of topmost politician figures on Instagram and categorize in 4 groups.
- Understand the motivation and activity of different impersonator accounts and an initial attempt to distinguish political bot generated impersonation.

The remaining of this study is as follows. Section II gives the relevant work. The dataset and explanation of data crawling are described in section III. The methodology of impostor identification is detailed in section IV. Next, we represent the user characteristics of imposters on section V and finally, section VI exposes future directions and concludes the study.

II. Related work

Recent research has discussed related problems and dedicated a fair amount of work to study OSNs [7]. Bots can alter the perception of social media influence, artificially enlarging the audience of some people, or they can ruin the reputation of a company [8]. The problem of rising social bots are discussed in [9]. There are various strategies to tackle the problem of bot detection. [10] suggested a profile-based approach and [11] proposed a novel framework on detecting spam content. Also, [12] presented a machine learning pipeline for detecting fake accounts.

On another line of research, the authors in [13] [14] look at the profile and behavioural patterns of a user and discussed existing challenges on different OSNs. By integrating semantic similarity and existing relationships between users, it is possible to match profiles across various OSNs [15] [16]. Also, [17] conducted a detailed investigation of user profiles and proposed a matching scheme. On Instagram, for the sake of mitigating impersonation attack, [18] explored fake behaviours and built an automated mechanism to detect fake activities.

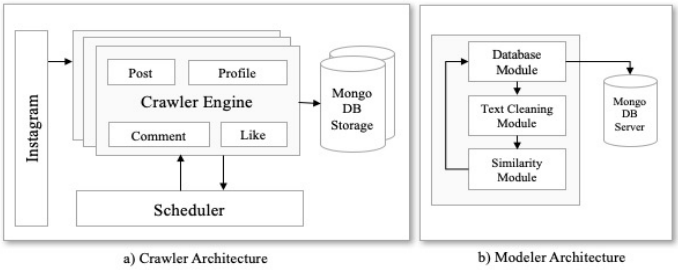
Political bots are automated accounts that are particularly active on public policy issues, elections, and political crises. The use of political bots during the UK referendum on EU membership is explained in [19] and also, [3] [20] described computational propaganda and define political bots designed to manipulate public opinion In the US context.

As far as our knowledge, the problem of finding impersonators of top politician figures on social media is not studied in the literature and this is the first study that analysis this phenomenon on Instagram.

III. Data Collection and Case Studies

A. Crawling

We designed a multi-threaded crawler to collect data from Instagram and store on a MongoDB server. The crawler connects to the Instagram API and receives data as a JSON file (Figure 2.a). The Instagram API Platform can be used to build non-automated, authentic, high-quality apps and services. The crawler consists of four modules, a) a post crawler, b) a comment crawler c) a like crawler, and d) a profile crawler. Each module executes the task associated with its module name. In line with Instagram policies and user privacy and ethical consideration defined by the community, we only gather publicly available data that are only obtainable from Instagram. The whole data collection process is designed exclusively for research purposes and the data is stored in an anonymized format.



a) Crawler Architecture b) Modeler Architecture

Fig. 2. General architecture of the implemented Crawler and Modeler.

B. Case Studies

Despite major investigations and the suspension of imposter by Instagram, imposters are still easy to find. We divided the profiles of genuine figures into different categories and we limited the scope of the study to the politicians. The logic for considering politicians is their large impact on OSNs. Almost everyone is interested in their shared media and the rate of their being targeted to be imitated also quite high. A vast majority of the imposter accounts from supporters to bots are trying to impersonate them in order to promote various goals. The other potential categories are news agencies and celebrities. We further narrowed our focus on two top popular political figures in the USA: Donald J. Trump (@realdonaldtrump), the current president of the USA (at the time of writing this paper), and Barack Obama (@barackobama), the former president. Additionally, we desired to see how the perceived findings are applicable for non-USA accounts so we also selected Emmanuel Macron, the current president of France (@emmanuelmacron). Particularly, these accounts are familiar in terms of the numbers of followers, received comments, and obtained likes.

C. Dataset

The collection process began on 30th December 2018 to past and stopped 1st October 2018. We collected all of the published posts in a three-month period. In keeping with our research purpose and Instagram policies, our principal targets were posts and the reactions based on them, so in the data collection period, around 80 Instagram posts, approximately 9M likes, 350K comments, and 1,5M profiles were crawled.

TABLE I
Characteristics of the dataset

	D. Trump	***B.Obama***	***E. Macron***
actual total #post	4,291	258	362
actual total #follower	11,1m	19,5m	1,3m
actual total #followee	8	14	81
#post crawled	37	11	27
total #comment	119.54K	176K	40.3K
total #like	3.3M	5M	888.6K
Total #mentions[a]	44.1K	3.238K	19.7K
avg #comment per post	3.2K	16K	1.4K
avg #like per post	89.3K	463K	32.9K
avg #mention per post	1.19K	4,5K	733
avg #comment per user	2.01	2.1	2.02
avg #like per user	3.0	3.4	2.67
crawled user #profiles[b]	501k	502k	501k

[a]profile mentioned in comments. [b] unique identified profiles.

Table I displays the characteristics of the dataset divided into several parts. The first part shows the actual number of posts, followers, and followees of the genuine account for the whole time-frame. The second part indicates the volume of data that we intended to study in the period of data collection. The dataset is thus a small part of the real data. While *Trump* shared 37 posts, which is more than others, surprisingly *Obama* received more reactions (like and comment). Besides, the number of users mentioned in the comments is included. The third part presents the average amount of reactions in each use-case. On average, *Obama* got the highest rate in all

features and *Macron* the least. It can be seen that in *Obama*'s case, with regards to the number of posts, users are more engaged. We presented the average like and comment reactions per unique user in the fourth part. Stats are from profiles who left comments or liked the post and we count them once in cases where they left several comments. For example, on average, while each user in Trump's case liked 3 posts, in *Macron* liked 2.67 posts.

To be able to perform analysis, for each use case, we randomly selected 500k unique users and crawled their profiles which are shown in the last part of the table. These users might be engaged in one or both reactions (like and comment). As a result, the total population contains nearly 1,5M profiles. As the process of crawling profiles is a time-consuming task, a proper pool of them is assessed for this study.

D. Instagram Characteristics

A profile on Instagram can be public or private. While in public profiles everything is visible, in private accounts few profile features are available. There are several ways to express reactions to a post: a) giving a like, b) posting comments, c) liking the comments, d) sharing in a shape of Direct (messaging feature), e) and bookmarking. We rely on the first two public actions (A and B) to identify the pool of users who are somehow reacting to the target profiles.

In general, our methodology is simple; make a pool of users who react to the crawled posts and then compare their public profiles to the target account.

IV. Identification of Impersonating Accounts

After crawling user profiles, here we identify impersonator accounts and arrange their classification. As a first step, we pick profile features and by applying the similarity measure, we can discover and extract similar users for further investigations.

A. Features

- **Username** is a string that individuals use on Instagram to define their profile address Composed of 30 symbols. Username, must contain only letters, numbers, periods and underscores. For example, the usernames of genuine accounts in this paper are @*realdonaldtrump*, @*barackobama*, and @*emmanuelmacron*.
- **Display Name** is what shows up on their Profile page, as well as next to a user's comments. "*President Donald J. Trump*", "*Barack Obama*", and "*Emmanuel Macron*" are the Display names of the genuine accounts.
- **Biography** is a section where users can include information about themselves and it is limited to 150 characters. The biography of our three genuine accounts are "*45th President of the United States*", "*Dad, husband, President, citizen*", and "*President de la Republique francaise*" respectively. All of these features are shown in Figure 1.
- **Profile Picture** on Instagram represents the account personage. This photo, whether a profile is public or private, is visible to everyone. Impersonator accounts copy the same (or a very similar) photo as their profile picture. All three of our politicians have their own clear face photo as their profile picture.
- **Shared Media** is the activity of a user in terms of publishing posts. For example, if someone is duplicating the post of the genuine account in their profile can be considered as an imposter.

B. Data Preparation and Cleaning

To make the model, we began with data preparation. We implemented a system including three main modules, a) a Database Module, b) a Text Cleaning Module and c) a Similarity Module (Figure 2. b). The Text Cleaning module removes duplicated words, punctuation, and emojis then filters the stop word (English) and tokenizes words containing at least three characters. These steps are necessary for all Username, Display Name, and Biography metrics.

The important factor is Username, which is a string without any space characteristic. Some examples are "*realdonaldtrump*" and "*itsdonaldtrumppp*" (The first one is the genuine Trump username and the other is an imposter). A useful approach is to extract sub-words from the username. Therefore, using the "wordninja" (github.com/keredson/wordninja), we broke down a single meaningless word into several meaningful words. Consequently, "*realdonaldtrump*" is converted to "*real donald trump*" and "*itsdonaldtrumppp*" to "*its donald trumppp*". Then, we can pass it to the Text Cleaning module. A final tokenized result would be "*donald trump*" for both. The Database module is responsible for handling interactions with the database. Finally, the data is ready for the similarity check and after measuring (Section IV-C), the Database module will update the documents with the calculated values.

C. Similarity Measures

Username, Display name, and Bio are all collections of strings, so to examine and rank the similarity, we can apply similarity measure algorithms. We employed two well-known similarity measures, Cosine and Euclidean, and found the Cosine results, in the Vector Space Model for texts, were better. Also, Cosine similarity is well-suited for use with very high dimensional data.

For matching profile photos, we leveraged a known library "Face Recognition" (github.com/ageitgey/facerecognition) built with deep learning that has an accuracy of 99.38% on face recognition. Accordingly, at first, each user's photo is examined whether a face exists in it or not. Next, the unique features of that face are identified and finally, the unique features of that face correlated the account holder's face are compared to recognize the person. We gave the actual profile photo of Donald J. Trump, Barack Obama, and Emmanuel Macron as the input to this library. The final output is delivered as a Boolean. "True" is the same person with equivalent facial features.

D. Modeling and Classification

We modelled the three first above-mentioned features in the Vector Space and applied TF-IDF transformation. This

TABLE II
TOTAL TARGET POPULATION, NUMBER OF IDENTIFIED SIMILAR ACCOUNTS BASED ON THE DEFINED SIMILARITY METRICS.

	D. Trump	*B. Obama*	*E. Macron*
#impersonator accounts	**108**	**38**	**21**
similar #username	37	6	4
similar #full_name	36	10	6
similar #bio	88	34	16
similar #photo	23	6	5
common in 1 metric	67	26	12
common in 2 metrics	13	4	4
common in 3 metrics	15	4	3
common in 4 metrics	13	4	2

made it possible to perform the calculations by employing components from the NLTK and scikit-learn packages. For obtaining similarity, we prepared a dictionary consisting of unique words of the Username, plus Display Name and Bio of each case. Trump's dictionary holds "*Donald J. Trump 45th President of the United States*", Obama consists "*Barack Obama 44th President of the United States*", and Macron contains "*Emmanuel Macron President de la Republique francaise*".

Succeeding, we checked and matched each user profile feature (Username, Display Name, and Biography) to this dictionary to find any similarity. We made a ground truth labelled data and after deep manual inspection on 150 similar profiles, we obtained the best threshold for each case as 0.3, 0.3, and 0.35 are for Trump, Obama, and Macron sequentially.

Based on the proposed model and assumed metrics, we count and cluster the impersonator users into few groups. Profiles that have similarity in a) 1 metric, b) two metrics, c) three metrics, d) and in 4 distinct metrics. Profiles with more similarity metrics are impersonating the genuine account more. The summary of findings is presented in Table II.

The "*#impersonator accounts*" in the table shows the cumulative amount of impersonators in each case. The number of Trump imposters is much larger than that of the other two. Each metric with its quantity is presented in the second part of the table. Throughout the crawled profiles, Trump has the largest numbers in all metrics, and Macron has the lowest. Interestingly, in all three use cases, the most similarity is in the biography feature. In the last part, we calculated the profiles that own similarity in more than one metric.

V. Characterization of Identified Accounts

A. Profile Characteristics

After discovering the imposters, we dive into their profile characteristics to recognize and compare their features. Figure 3 shows the number of imposters Followers and Followees.

The first observation is the highest density of dots is in the middle area (10^3 both followers and followees) which indicates most of the impersonators have a lot of followers and followees at the same time. It is a remarkable characteristic. Surprisingly, most of the imposters in all three cases have almost the same number of followers and followees. Furthermore, this trend continues as it increases, which indicates that they follow more as their followers grow. We can also see that

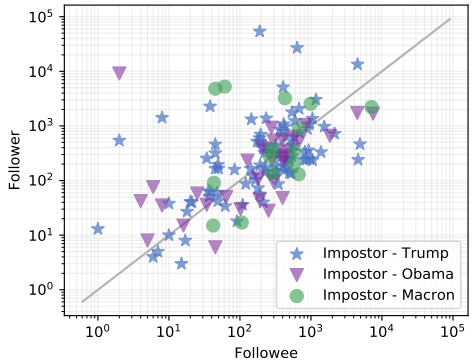


Fig. 3. Portion Follower vs. Followee in identified impersonator profiles.

some imposters have a higher number of followees rather than their followers, meaning they are following more.

Table III explores other features from imposters. The first section explains how many public profiles exist among impersonators and *Trump* owns the higher frequency. Moreover, the average number of the followers, followees, and media count are provided. In the case of the follower, while Trump and Macron imposters have a very high amount, *Obama*'s imposters have approximately half that amount. As for the number of followees, *Macron*'s imposters have even more. And for the amount of published media, Obama impersonators are in the lead.

TABLE III
IMPOSTERS PROFILE CHARACTERISTICS

	D. Trump	*B. Obama*	*E. Macron*
#public profile	46	14	4
avg #follower	1407	628	1369
avg #followee	491	652	812
avg #media_count	214	272	155
avg bio #length[a]	8.33	4.9	5.2
bio related #hashtag[b]	0.26	0	0.6
bio related #mentioned[c]	0.16	0.15	0.13

[a] in words. [b] Related hashtag in bio. [c] Related user mentioned in bio.

In the second part of Table III, we examined the impersonator's biography. On average, *Trump*'s imposters have longer biographies. Generally, in bio, there are four ways of promoting that are visible for everyone. Imposters are promoting by a) putting a hashtag starting with #, b) mentioning some user starting with @, c) expressing feeling by placing some words as text, d) or including URL address. For example, in the case of *Trump*, impersonators are promoting the upcoming 2020 US Presidential election using hashtag "*votefordonaldtrump2020*" or by mentioning (@realdonaldtrump). We count and display these in this part of the table (we do not consider URL addresses). On medium, while Macron's imposters hold more related hashtag, *Trump*'s impersonators are mentioning the genuine account higher.

B. Activity Characteristics

After identifying impersonators, now we cover their activities. In 3 months activity, the average comments posted by imposters are 2.79, 1, and 1 for Trump, Obama, and Macron cases. Clearly, Trump's imposters have more activities. Furthermore, the average like caught by imposters are 13.6, 1.7, and 6.7 respectively. This means imposters favoured to

do like rather than post a comment. Next, we did a manual check and investigated the post of the imposters to know if they are sharing the same posts (of the genuine account) or not. Interestingly, we discovered some are duplicating up to 60%. Among three cases, Trump's imposters duplicated more

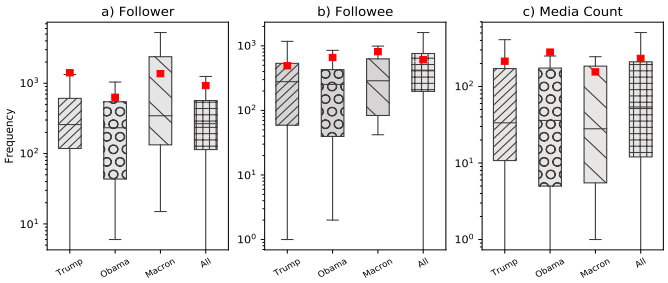


Fig. 4. Comparision of some Profile Features.

In Figure 4 by comparing the imposter's accounts in 3 use cases, we observe almost the same pattern with minor differences. In term of followers, Obama and Macron impersonators have the wider range than Trump. All three groups are following around 500 users in the median and 1000 on average. Surprisingly, the range and the median of Trump imposters are similar to all, but there is a tiny difference in average.

In term of followees, median and average numbers in all cases are approximately equal which means the impersonators do not have a strange number of followees. Furthermore, the range of 3 cases are wider than all users. And in shared media, normally all cases are posting between the range of 5 to 250. In this feature, again we can see the behaviour of Trump's imposters is highly similar to all users.

Next, we desire to know when impersonators are commenting? and compared to others, what is the rate? So, Figure 5 presents the age comments that are published. Plot a) is the cumulative distribution of the age of the comments (hour) which compared imposters to the whole dataset. For better presentation, we limit the figure in the x-axis. Nearly 30% of the comments (for both) are posted in the first hour. As it continues, imposters comment more, and in the first 10 hours, they posted 80% of their total comments, while this number is around 60% for others. This means imposters, in term of commenting, are really engaged in the first 10 hours.

The second Figure 5.b is the Boxplot representation of the age of the comments (minutes). As it can be seen, the range of all group is wider than imposters. Also Imposters, in the median commented by the minute 100, while others posted by the minute 250. The average point for both groups is large.

For better distinction of comments that are issued, we calculate the average published time of the comments per unique user and plot it on Figure 6. By considering the first hour, while on average 60% of the imposter's comments are issued, others are publishing less than 30% of their total. Furthermore, nearly 90% of the imposter's comments are published on the first day, but this number is around 80% for other. This means imposters are eager to comment really quick (abnormal activity). So, from the perspective of traffic

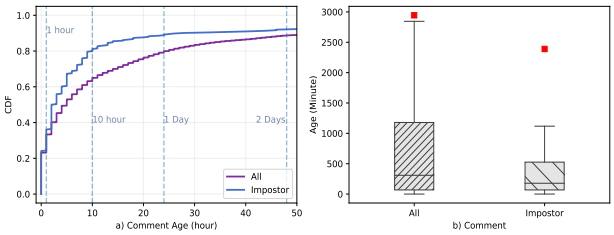


Fig. 5. Age of the posted comment earned by all users vs imposters.

management, they are producing huge network traffic and in a large-scale format that could contribute to traffic jams.

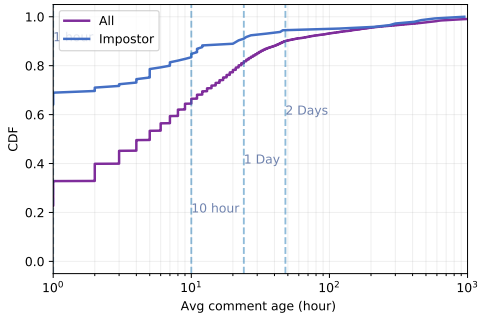


Fig. 6. CDF of the average of the comments posted by unique users.

C. Impersonator Categorization

Based on impersonator profile characteristics and their activities, four clusters as follow has been made that potentially have different motivations, aims, and behaviour (see Table IV):

- **Active Bot** are those accounts that have unique characteristics and are of high interest for us. Public profile, similarity at least in 2 metrics, similar photo, high rate of followers and followees are unique specialities. Active Bots duplicate more than 50% of the genuine account posts. For monitoring this, we did the manual check of all the imposters shared-post.
- **Inactive Bot** are generally considered another type of bots. The profile could be public or private, there is no published post, the rate of followers and followees are low and they have similarity at least in two metrics.
- **Fan / Opposition** generally supports or opposes that political figure. Features including public profile, similarity in at least 2 metrics, high number of the followers, low amount of followees, and similarity in profile photo are linked with this group. In some cases, they are using the flag of the country as the profile photo. Furthermore, they are duplicating roundly 30% of the posts.
- **Ordinary Users** also exist with a simple motivation in mind which is to support the favourite political figure. Generally, they show the support by mentioning the username (of genuine account) or adding hashtags (like makeamericagreatagain) or writing some words in bio. So, the only similarity is in just 1 metric. The range of followers and followees are normal.

D. Short dive on Bot activities

On Instagram, bots could have different aims and behaviours. Inactive Bots, as they have very weak profiles (low

TABLE IV
IMPOSTER CATEGORIZATION

	Profile Characteristics				Similarity						Activity
	Follower	Followee	Media Count	Public	Username	Display Name	Bio	Photo	Least Common Metrics	Most Common Metrics	Duplication Rate
Active Bot	>1k	>1k	>10	Y	Y/N	Y/N	Y/N	Y	2	4	>50%
Inactive Bots	<100	<100	<5	Y/N	Y/N	Y	Y	Y/N	2	4	-
Fan/Opposition	>500	<100	>50	Y	Y/N	Y/N	Y	Y	2	3	<30%
Ordinary Users	50<&<1k	50<& <1k	5<&<100	Y/N	N	N	Y	N	1	1	-

followee rate, no post, no profile picture) are not followed by the human. As we found, they are active to Like or Comment in favour of supporting or opposing someone. On the other hand, Active Bots hold better profiles (in term of features) and have a large number of followers and followees. Therefore they can be followed by humans or other bots (or a network of bots). Additionally, based on our manual inspection we identified that both Active and Inactive Bots are posting the same comment several times on different posts (in the first hour). It is an apparent behaviour of automated bots that are programmed to Comment. So, they are giving the fake visibility or fake popularity.

On other end, Fan/Opposition accounts have unique features. For example, they own a reasonable quantity of followers and do not follow a lot of pages. They are being followed by humans too. This sounds reasonable because at first, people like to follow the news of their favourite figure and in most cases, Fan Pages are covering the news better than the actual account. Secondly, sometimes the actual figure does not have any official page and as a result, fan pages will flourish. As we found, Fan Pages can be controlled by Machine or Human.

VI. CONCLUSION AND FUTURE WORK

In this paper, we did an analysis to discover impersonators based on the reactions on the published posts of top politician figures on Instagram. To our knowledge, this is the first paper that conducts such analysis on Instagram data. We have explained how imposters are identified and what features are considered. Based on user profiles and user activities, we achieve four different categories of imposters with different characteristics. We also showed impersonators and political bots are eager to comment in the early hours of a published post, and most of them have a normal number of followers, followees, and shared media. As a future direction, at first, we aim to handle the problem of fake account identification as a machine learning problem in a large-scale format and also, we intend to improve the detection by considering more features such as Stories (media that vanish after 24 hours) and IGTV (long-form video service) on Instagram. Next, we can move to shared-posts of impersonators to study their behaviours. Also, we can do cross users analysis and by extending this model we are able to identify aggressive political parties or people who aim to destroy the opposite parties by their activities.

REFERENCES

[1] New york times. https://www.nytimes.com/2016/11/18/technology/automated-pro-trump-bots-overwhelmed-pro-clinton-messages-researchers-say.html, 2019.

[2] New york times, https://www.nytimes.com/2018/02/20/technology/social-media-impostor-accounts.html, 2018.

[3] Philip N. Howard, Samuel Woolley, and Ryan Calo. Algorithms, bots, and political communication in the us 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2):81–93, 2018.

[4] Gohar F. Khan. *Seven Layers of Social Media Analytics: Mining Business Insights from Social Media Text, Actions, Networks, Hyperlinks, Apps, Search Engine, and Location Data*. 2015.

[5] Statista.com. Statista - the statistics portal for market data, market research and market studies, 2019.

[6] Bob Burg Bryan Kramer Jay Baer Kim Garst David Meerman Scott Mark Schaefer Sue Zimmerman Tyler J. Anderson Jon Mitchell Jackson, Chris Brogan. *The Ultimate Guide to Social Media For Business Owners, Professionals and Entrepreneurs*. 2018.

[7] P. Rajapaksha, R. Farahbakhsh, N. Crespi, and B. Defude. Inspecting interactions: Online news media synergies in social media. In *2018 IEEE/ACM ASONAM*, pages 535–539, Aug 2018.

[8] Z Gilani, R Farahbakhsh, G Tyson, and J Crowcroft. A large-scale behavioural analysis of bots and humans on twitter. *ACM Trans. Web*, 13(1):7:1–7:23, February 2019.

[9] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. The rise of social bots. *Commun. ACM*, 59(7), 2016.

[10] White J. S. Hudson B. Voter B. R. Matthews J. N. Gurajala, S. Profile characteristics of fake twitter accounts. 2016.

[11] S. Shehnepoor, M. Salehi, R. Farahbakhsh, and N. Crespi. Netspam: A network-based spam detection framework for reviews in online social media. *IEEE Transactions on Information Forensics and Security*, 12(7):1585–1595, July 2017.

[12] Cao Xiao, David Mandell Freeman, and Theodore Hwa. Detecting clusters of fake accounts in online social networks. In *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, AISec '15, pages 91–101. ACM, 2015.

[13] Francesco Buccafurri, Gianluca Lax, Serena Nicolazzo, and Antonino Nocera. Comparing twitter and facebook user behavior. *Comput. Hum. Behav.*, 52(C):87–95, November 2015.

[14] Bang Hui Lim, Dongyuan Lu, Tao Chen, and Min-Yen Kan. #mytweet via instagram: Exploring user behaviour across multiple social networks. IEEE/ACM ASONAM '15, pages 113–120. ACM, 2015.

[15] Ali Choumane, Zein Al Abidin Ibrahim, and Bilal Chebaro. Profiles matching in social networks based on semantic similarities and common relationships. In *Proceedings of the International Conference on Compute and Data Analysis*, ICCDA '17, pages 14–18. ACM, 2017.

[16] Katharina Krombholz, Dieter Merkl, and Edgar Weippl. Fake identities in social media: A case study on the sustainability of the facebook business model. *Journal of Service Science Research*, 4(2), Dec 2012.

[17] Oana Goga, Patrick Loiseau, Robin Sommer, Renata Teixeira, and Krishna P. Gummadi. On the reliability of profile matching across large online social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 1799–1808. ACM, 2015.

[18] Indira Sen, Anupama Aggarwal, Shiven Mian, Siddharth Singh, Ponnurangam Kumaraguru, and Anwitaman Datta. Worth its weight in likes: Towards detecting fake likes on instagram. In *Proceedings of the 10th ACM Conference on Web Science*, WebSci '18. ACM, 2018.

[19] Philip N. Howard and Bence Kollanyi. Bots, #strongerin, and #brexit: Computational propaganda during the UK-EU referendum. *CoRR*, abs/1606.06356, 2016.

[20] Jessica Baldwin-Philippi. The myths of data-driven campaigning. *Political Communication*, 34(4):627–633, 2017.