2020

Application of the Benford's law to Social bots and Information Operations activities

Lale Madahali
University of Nebraska at Omaha, lmadahali@unomaha.edu

Margeret Hall
University of Nebraska at Omaha, mahall@unomaha.edu

Recommended Citation

L. Madahali and M. Hall, "Application of the Benford's law to Social bots and Information Operations activities," 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), Dublin, Ireland, 2020, pp. 1-8. doi: 10.1109/CyberSA49311.2020.9139709

Application of the Benford's law to Social bots and Information Operations activities

Lale Madahali
University of Nebraska at Omaha
Omaha, United States
lmadahali@unomaha.edu

Margeret Hall
University of Nebraska at Omaha
Omaha, United States
mahall@unomaha.edu

***Abstract*— Benford's Law shows the pattern of behavior in normal systems. It states that in natural systems digits' frequency has a certain pattern such that the frequency of numbers' first digits is not evenly distributed. In systems with natural behavior, numbers begin with a "1" are more common than numbers beginning with "9". It has many applications in forensic accounting, stock market, finding abnormal data in survey data, and natural science. It implies that if the distribution of first digits is outside of the expected distribution it can be indicative of fraud. In this paper, we investigate whether bots and Information Operations' activities conform to this law. Our results showed that Bots' behavior do not adhere to Benford's Law, suggesting that using this law helps in detecting malicious online accounts and activities on social media. However, activities related to Information Operations did not show consistency in regards to the number distribution. Our findings shed light on the importance of examining regular and anomalous online behavior to avoid malicious and contaminated content on social media. This is the first attempt in investigating automated accounts and Information Operations' activities in terms of adherence to the law of digital fraud.**

Keywords—social bots, Information Operations, Benford's law

I. INTRODUCTION

Benford's Law is an auditing tool for fraud detection. Fraud detection in social media is gaining increase attention and studies resolve the issue by machine learning techniques. Benford's law applies to accounting forensics [1], finance [2], survey data [3], natural science [4], and religion [5]. According to this Law, digits' frequency has a certain pattern such that the frequency of numbers' first digits is not evenly distributed. In systems with natural behavior, numbers begin with a "1" are more common (more than six times as frequent) than numbers beginning with "9" [6]. Therefore, the probability that the first digit of a number is one is about 30 percent, whereas the probability of nine being the first digit is 4.6 percent. Table 1 shows the probabilities of all digits 0 to 9 in each of the first four places in any number. In this study, we apply this law to the anomalous behavior in online social media including activities of bots and Information Operations (hereafter IO).

The word "bot" is derived from the word "robot", which is defined as automated agents that function on an online platform [7]. They are computer programs that are persistent, autonomous, reactive, run continuously, formulate decisions, and activated by itself [8]. There are various typologies known for bots including Webrobots, Chatbots, Spambots, Social bots, Sockpuppets, Trolls, and Cyborgs [9]. In this study, we use datasets containing Spambots, Social bots, and Trolls, because research shows that 15% of Twitter traffic are created by bots [10]. Webrobots are used for crawling the Internet and information retrieval. Chatbots are usually engaged in natural conversations with humans [9]. Therefore, the purpose of Webrobots and Chatbots incite different definition and operation for them, and they do not need to be detected. Whereas, other typologies have different definitions which causes them to be leveraged for malicious purposed, for example spreading misinformation or impersonating other people. Therefore, we run our analysis on datasets containing Spambots, Social bots, and Trolls.

IO run by influence campaigns try to manipulate online discussions are becoming a serious concern [11]. There are studies dedicated to analyze these campaigns in terms of the discourse they are involved in, their dynamics, and influence measurements [11][12].There is no prior study on examining IO activities in terms of following digital fraud patterns as they are understood to be malicious and malevolent activities.

In this study, we investigate whether the activities of automated agents and IO conform to the law of naturally-occurring systems. Table I provides the frequencies and probabilities of each digit for normal systems. To accomplish this, we conducted two sets of analyses of four datasets for social bots and seven datasets originating from state-backed information operations[1]. To the best of our knowledge, there is lack of research on applying Benford's Law to bots' activities as well as IO activities. It is important to know these activities because they are very common and are growing. We are trying to know whether social bots' and IO activities adhere or deviate from the law of normal systems and behaviors. Therefore, we address two research questions:

RQ1: Do Information Operations and social bots' activities adhere to Benford's Law?

TABLE 1. EXPECTED FREQUENCIES BASED ON BENFORD'S LAW

Digit	1st place	2nd place	3rd place	4th place
0		.11968	.10178	.10018
1	.30103	.11389	.10138	.10014
2	.17609	.19882	.10097	.10010
3	.12494	.10433	.10057	.10006
4	.09691	.10031	.10018	.10002
5	.07918	.09668	.09979	.09998
6	.06695	.09337	.09940	.09994
7	.05799	.09035	.09902	.09990
8	.05115	.08757	.09864	.09986
9	.04576	.08500	.09827	.09982

Our research objective is to see whether bots and IO characteristics follow the Benford's Law. Our study

[1] https://about.twitter.com/en_us/values/elections-integrity.html#data

contributes to the research by finding out whether malicious online activities adhere to a law which exists in natural systems. We found out that bots' activities show nonconformity to the Benford's law. However, trolls' activities do not show a consistent pattern.

II. MOTIVATION

Social networks have made the world smaller and people closer to each other. Services like Twitter, Facebook, and Instagram have provided a ubiquitous service to connect people. Nowadays most individuals use social media for entertainment, getting news, and other various purposes. Therefore, they are exposed to overwhelming information spread from both human and automated agents. Bots are responsible for significant proportions of online activity. While Subramanian and colleagues [13] found out that about 8.5% of all Twitter users were bots, other studies show 16% of Twitter traffic is generated by bots [14]. They also play a major role in disseminating hyperpartisan "fake news", particularly during special events like elections [15] [16]. Further, bots can impersonate real people and deceive the audience. Therefore, it is important to know how pattern of bots' activities look like.

IO is run with the purpose of interfering in foreign countries' affairs. Thus, there is a need to understand these activities more in order to be able to detect them and tackle their consequences. In this study, we assume that both these activities are fraudulent, and seek to understand them better in terms of their conformity to the law of normal behaviors. Further, this study contributes to the field of cyber security by understanding online spammers especially because both can be used for propaganda and public opinion manipulation.

A. Theoretical basis

Benford's Law is an auditing tool for fraud detection. In states that in systems with natural behavior, numbers begin with a "1" are more common (more than six times as frequent) than numbers beginning with "9" [1]. Therefore, the probability that the first digit of a number is 1 is about 30 percent, whereas the probability of 9 being the first digit is 4.6 percent. It shows that in systems with normal behaviors for a digit *d* the formula for predicting its frequency *p* is as follows:

$$\boldsymbol{P(d) = log_{10}(1 + \frac{1}{d})} \qquad (1)$$

According to this Law, digits' frequency has a certain pattern such that the frequency of numbers' first digits is not evenly distributed. In systems with natural behavior, numbers begin with a "1" are more common (more than six times as frequent) than numbers beginning with "9" [1]. Therefore, the probability that the first digit of a number is one is about 30 percent, whereas the probability of nine being the first digit is 4.6 percent. Table 1 shows the probabilities of all digits 0 to 9 in each of the first four places in any number. We start with the law of first significant digit for the first digit:

$$\boldsymbol{Prob(first\ significant\ digit = d) = \log(1 + d^{-1}) \quad d = 1, 2, \dots, 9} \qquad (2)$$

The law of second significant digit states that:

$$\boldsymbol{Prob(first\ significant\ digit = d) = \sum_{k=1}^{9} \log_{10}(1 + (10k + d)^{-1}) \quad d = 0, 1, 2, \dots, 9} \qquad (3)$$

Therefore, the general form of the law is:

$$\boldsymbol{Prob\left(mantissa \le \frac{t}{10}\right) = \log_{10} t \quad t \in [1, 10)} \qquad (4)$$

Mantissa of a positive real number is defined as the unique number r in $[\frac{1}{10}, 1)$ with $\boldsymbol{x = r * 10^n}$ for some integer n

$$\boldsymbol{Prob\left(mantissa\ (base\ b) \le \frac{t}{b}\right) = \log_b t \quad for\ all\ \ t \in [1, b)} \qquad (5)$$

Therefore, the general significant-digit law looks like:

$$\boldsymbol{Prob(D_1 = d_1, \dots, D_k = d_k) = \log_{10}[1 + (\sum_{i=1}^{k} d_i * 10^{k-i})^{-1}} \qquad (6)$$

for all positive integers k, all $\mathbf{d_1} \in \{1, 2, \dots, 9\}$ and $d_j \in \{0,1, \dots, 9\}$, j=2, ..., k

Using Benford's law for the first two digits is more common in fraud detection and is more reliable than considering only the first digit [17]. Following is the common formula for the law for the first two significant digits:

$$\boldsymbol{Prob(D_1D_2 = d_1d_2) = \log\left(1 + \frac{1}{d_1d_2}\right)} \qquad (7)$$

$$d_1d_2 \in [10,11, \dots, 98,99]$$

III. RELATED WORK

In this section, we give an overview of studies on social bots, trolls and IO, and Benford's law applications.

Bots. Studies on bot detection mainly leverage machine learning techniques to improve the performance of bot detection, understanding [18], characterizing [19], and classifying them [20] [21]. For example, bot detection studies have leveraged graph-based features [22] [23], contextual information [24], sentiment analysis [25], temporal features [10], and hybrid feature sets [26]. Additionally, there are studies on classifying and characterizing bots which introduce typologies and taxonomies of bots in order to provide a better understanding of them [27] [9]. Additionally, there are real-time bot detection platforms like Debot [10], BotOrNot [26], and BotWalk [28]

Trolls. Research in the area of trolls is more toward finding out what trolls are trying to disseminate, especially during election or other political events [16][29], and how they participate in online discourse to manipulate the public opinion [11]. Trolls are considered a specific type of

accounts called sockpuppets run mainly with political purposes [9]. This is the first attempt to examine troll's activities in terms of conformity to online fraud pattern.

Benford's Law. The first article published on Benford's Law was written by Simon Newcomb when he observed the beginning pages of the book were more frequently used and worn out compared to the last pages. He discovered that other scientists used tables with lower numbers than tables with higher numbers, concluding that numbers with lower digits are more frequent than numbers with higher digits. He came up with a probability formula for any non-zero first digit [30]:

$$P(d)=Log_{10}(1+\frac{1}{d})$$

Where p is the probability for digit d. For more empirical evidence and explanation about the Benford's law and its bases see [17]. Initially, Newcomb's article did not receive much attention, likely due to a lack of theoretical explanation [1]. After 50 years, physicist Frank Benford realized the same pattern that the first few pages of his books were more worn out than the last pages. He also asserted that the numbers beginning with low digits are more common. Thus, he came to the same conclusion and tested his hypothesis by collecting and analyzing more than 20,000 observations from various sources like areas of rivers and atomic weights of elements [31]. Therefore, Benford's law is sometimes referred to as Newcomb-Benford's Law [32]. Benford's Law has variety of applications in stock market [33], natural sciences (such as physics, astronomy, geophysics, chemistry, and so forth) [34], world religions [5], finance [2], tax returns [34], and social networks [35].

IV. DATA

We analyze the behavioral pattern of social bots and IO on Twitter because these accounts are widespread and play a major role in misinformation spread in online social media. In order to accomplish this, we collected datasets on bots and trolls from different resources to run our analysis on them. For examining social bots we collected 10 datasets including Texas A&M Social Honeypot [36], Socialspambot annotated by Cresci and colleagues [37], and Twitter bots datasets by [18].

For analyzing IO activities, we gathered 7 collections of Twitter integrity datasets released by Twitter on February 8th, 2019. Twitter believes that these accounts originate from state-back Information Operations and therefore, suspended the accounts. It contains 25,076,853 tweets from 2009 to 2018. There are 3613 accounts connected to Russia Internet Research Agency (RIRA), 770 accounts originate from Iran, 15 accounts come from Bangladesh, 2320 accounts in a separate collection originate from Iran, 416 accounts from Russia, 1196 accounts from Venezuela, 764 accounts in a separate set from Venezuela.

We run the analysis for both human accounts and bots accounts. We focus on following count, follower count, favorite count, reply count, retweet count in analyzing all datasets except for ZafarGilani's datasets because they have follower_following_ratio rather than separate following count and follower count. In ZafarGilani's dataset there is only follower to friend ratio, statuses count, favorites count, retweet count, and reply count and therefore, we exhibit the results for bots and humans across these features. We run the analysis for both human and bot accounts.

TABLE II. BOTS' DATASETS AND THEIR FEATURES

Datasets	Features
Socialhoneypot content polluters	Number of followers, number of followings, number of tweets
Socialhoneypot legitimate users	Number of followers, number of followings, number of tweets
Socialspambot fake followers, Socialspambot1, Socialspambot2, Socialspambot3, Traditionalspambot1, Traditionalspambot2, Traditionalspambot3, Traditionalspambot4	Statuses count, number of followers, number of following, number of favorites, number of listed
ZafarGilani	Statuses count, number of favorites, number of replied

TABLE III. TROLLS' DATASET. THESE ARE TWITTER ELECTIONS INTEGRITY DATASET. TWITTER BELIEVES THESE ACCOUNTS ARE PART OF STATE-BACKED INFORMATION OPERATIONS.

Dataset	Number of users	Number of tweets
Bangladesh	15	26214
Iran 2320	2311	4447056
Iran 770	771	1122936
RIRA	3608	8768633
Russia 416	416	765246
Venezuela 1196	1196	8961788
Venezuela 764	755	984980

V. ANALYSIS AND METHODS

For testing how the data fits the Benford's distribution (test of goodness-of-fit), we use Pearson's chi-squared test, Mantissa Arc test, Mean Absolute Deviation (MAD), and Distortion Factor to ensure the robustness of the results.

Pearson chi-squared test: the first step in this test is to calculate the chi-squared test statistic, χ^2. It shows a normalized sum of squared deviations between observed and theoretical frequencies. Second, we need to calculate the degrees of freedom of the statistic. Since we consider all of the 9 possible digits, the degrees of freedom will be 9-1=8. The null hypothesis is that the data follows the Benford's distribution [33]. The test statistic is calculated using this formula:

$$\chi^2(8) = \sum_{i=1}^{n} \frac{(N_{obs}-N_{ben})^2}{N_{ben}} \quad (8)$$

- Mantissa Arc test (MAT): for a positive real number x, Mantissa is the unique number r in $[\frac{1}{10},1)$ with $x = r * 10^n$ for some integer n [17]. More details about Mantissa is in the theory section (section A). In MAT, each data point x_i is mapped to a unit circle with the following coordinates:

$$x - coordinate = \cos(2\pi * (\log_{10}(x_i)\, mod 1)) \quad (9)$$

$$y - coordinate = \sin(2\pi * (log_{10}(x_i)\, mod 1)) \quad (10)$$

The reason behind having modulus 1 is that we only need to take the right side of the decimal point. The center of the unit circle is called mean vector and its coordinates are calculated using following equations (actually average of data points' coordinates):

$$x - coordinate = \frac{\sum_{i=1}^{N} \cos(2\pi * (\log_{10}(x_i)\, mod 1)}{N} \quad (11)$$

$$y - coordinate = \frac{\sum_{i=1}^{N} \sin(2\pi * (\log_{10}(x_i)\, mod 1)}{N} \quad (12)$$

If the mean vector (center of the circle) is at (0,0), mantissas of the data points are uniformly distributed on the unit circle. The length of the mean vector is shown by L^2, and therefore the test statistic for MAT is defined as [38]:

$$p - value = 1 - e^{-L^2 \times N} \quad (13)$$

The null hypothesis is that the data is uniformly distributed, and the degrees of freedom is 2. For further detail on this test see [38].

- Mean Absolute Deviation (MAD): this test calculates the average deviation of data points (OBS_i) from Benford's data points (BEN_i). The MAD test statistic is:

$$Mean\ Absolute\ Deviation = \frac{\sum_{i=1}^{k} |OBS_i - BEN_i|}{k} \quad (14)$$

As Nigrini suggests, the cutoff point for non-conformity to Benford's distribution is MAD ≥ 0.002, $MAD \in [0.0018, 0.0022]$ as marginally acceptable conformity, $MAD \in [0.0012, 0.0018]$ as acceptable conformity, and MAD ≤ 0.0012 as close conformity [38].

- Distortion Factor (DF): this test checks whether the data points are overstated (positive number) or understated (negative number). For example, the DF value of 0.0054 suggests that the numbers are overstated by 54 percent. Thus, it calculates the difference between the actual mean (AM) and expected mean (EM) [38]. Its test statistic is:

$$DF = \frac{AM - EM}{EM} \quad (15)$$

VI. RESULTS AND DISCUSSION

Bots' results. In Socialhoneypot [20] and Socialspambot [37] datasets we have the statistics for bots as well as genuine accounts. Therefore, we compare the results for both across the features. Fig 1, Fig 2, and Fig 3 illustrate the results for follower count, statuses count, and following count for bots and humans in Socialhoneypot dataset, suggesting that legitimate accounts conform to the Benford's law distribution, whereas bots' statistics do not. For Bots' account features all tests showed nonconformity (Table V). In order to examine the real accounts' activities, we tested two datasets and results suggests conformity for real accounts (Table IV).

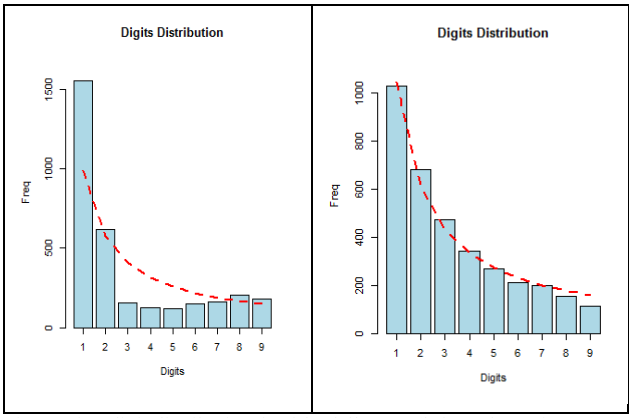

Fig. 1. First digit distribution for fake followers (left) and genuine accounts (right) for their followers. These accounts are from Crecsi 2017 dataset [37]. As can be seen, the distribution does not conform to the law.

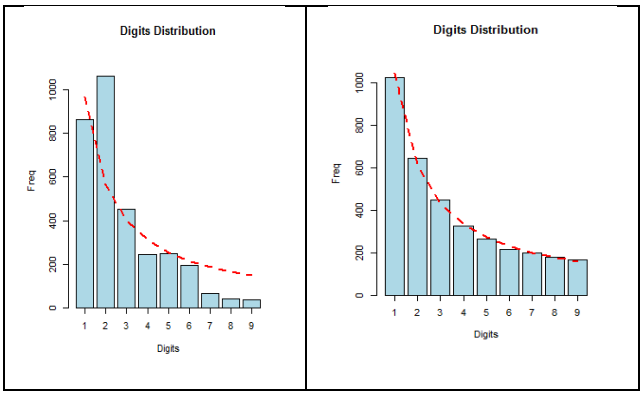

Fig. 2. First digit distribution for statuses counts for bots (left) versus genuine accounts (right)

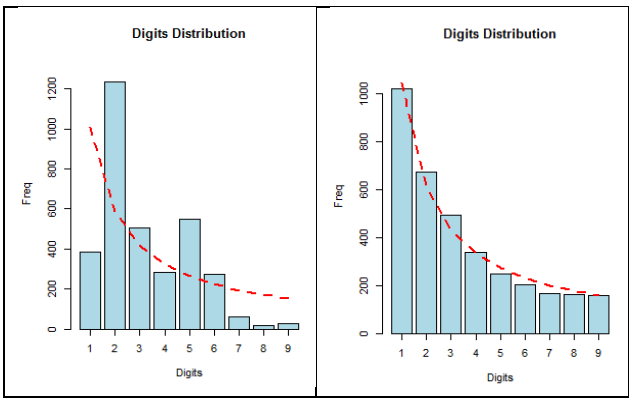

Fig. 3. First digit distribution for following count for bots (left) versus genuine accounts (right).

We use three tests on datasets and the results of these tests are not coherent sometimes. For example, in Socialspambot1 for the statuses count feature, two tests (Pearson's Chi-squared and Mean Absolute Devian) showed nonconformity and only MAT indicated conformity (Table VI). In these cases, we chose to use a voting strategy and decided based on the result of the majority of tests as well as looking at the visual results. Additionally, it is said

that in testing Benford's law we should not rely only on p-value because even the real data does not conform to Benford's distribution perfectly [38].

TABLE IV. REAL ACCOUNTS' FEATURES AND THEIR CONFORMITY TO THE BENFORD'S DISTRIBUTION

dataset	Features	Adherence to Benford's law
Legitimate users	Following	2 tests Conformity
	Follower	2 tests Conformity
	Statuses count	All tests conformity
	Favorites count	Does not exist for this dataset
Genuine accounts	Following	2 tests conformity
	Follower	2 tests conformity
	Statuses	All tests conformity
	Favorites count	All tests conformity

TABLE V. FEATURES OF BOTS AND THEIR CONFORMITY TO THE BENFORD'S LAW

Features	Adherence to Benford's law
following	Nonconformity
Follower	Nonconformity
Statuses count	Nonconformity
Favorites count	Nonconformity

TABLE VI. EXAMPLE OF DIFFERENT TEST RESULTS FOR STATUSES COUNT IN SOCIALSPAMBOT1 DATASET

Test	Result	Final decision on the test
Pearson's Chi-squared test	X-squared = 66.365, df = 8, p-value = 2.59 e-11	Nonconformity
Mantissa Arc Test	L2 = 0.001894, df = 2 , p-value = 0.1531	Conformity
Mean Absolute Devian	0.024	Nonconformity

For the bots collection by Zafar Gilani and colleagues [18], we tested follower to following ratio, statuses count, favorite count, retweet count, and reply count. Fig. 4, Fig. 5, and Fig. 6 show the results for Zafar Gilani's dataset. For human accounts, chi-square and MAD tests show nonconformity to the distribution, whereas MAT shows conformity for tweeting behavior and therefore, we use the voting strategy and mark it as nonconformity. Therefore, while tweeting, retweeting, replying, and follower_friend_ratio behavior does not represent conformity, favoriting shows conformity. For bot accounts, tweeting, replying, and favoriting show conformity, while retweeting and follower_friend_ratio exhibit nonconformity.

In fact, in this dataset, both bots and humans exhibit similar and unexpected behaviors. We argue that these results might be due to their data collection. In their data collection, Zafar Gilani manually categorize bots and humans using a team of four people and based on that data collection they get their results for the characterization of bots versus humans [39]. Another reason might be that they collected the tweet behavior of accounts and they consider tweeting, replying, and mentioning as a single action and therefore, their graph is more of a communication graph. In contrast, other datasets have a static snapshot of the accounts' features rather than communication graph features. Further research is needed to investigate the attributes of communications across these features for bots versus humans. Detailed results of tests for bots and humans in ZafarGilani's dataset are shown in Table VII.

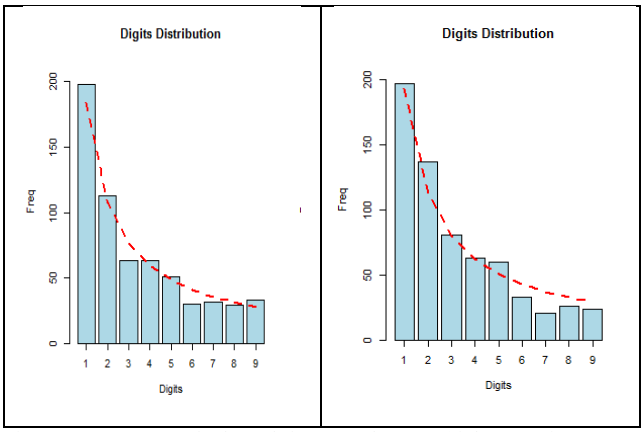

Fig. 4. Statuses count for bots (left) versus humans (right) tweeting behavior in Zafar Gilan's dataset

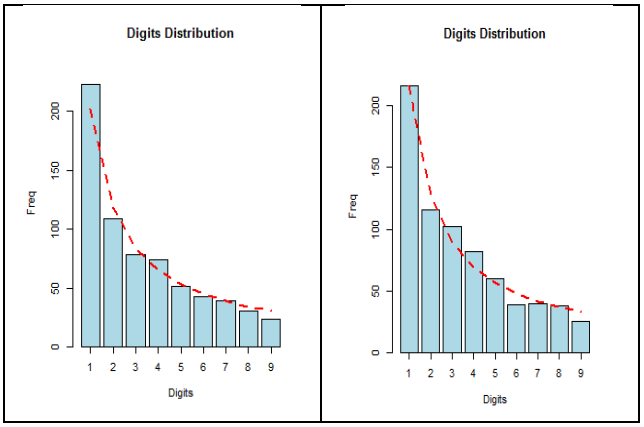

Fig. 5. Favoriting behavior for bots versus humans in Zafar Gilani's behavior

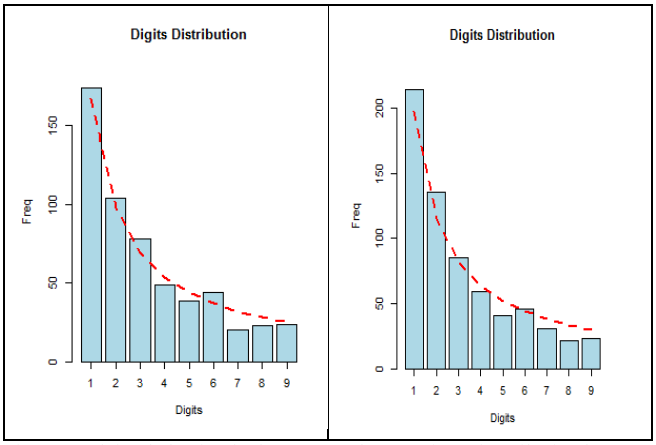

Fig. 6. Replying behavior for bots versus humans in Zafar Gilani's dataset

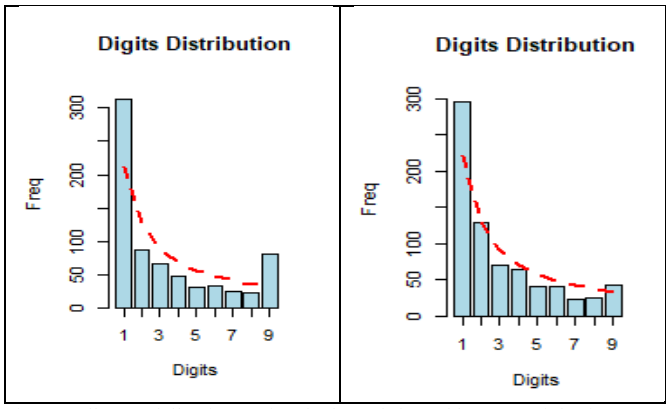

Fig. 7. Follower_following_ration for bots (left) and humans (right) in Zafar Gilani's dataset

Table VII. Test results for ZafarGilani's dataset

Accounts	Feature	Test result
Bots	Following	Does not exist
	Follower	Does not exist
	Statuses count	1 test nonconformity
	Favorites count	1 test nonconformity
	Retweet count	nonconformity
	Reply count	conformity
	Follower_following_ratio	nonconformity
Humans	Following	Does not exist
	Follower	Does not exist
	Statuses count	2 tests nonconformity
	Favorites count	All tests conformity
	Retweet count	nonconformity
	Reply count	2 tests conformity
	Follower_following_ratio	nonconformity

Our findings of IO activities show none of them conform to Benford's law for follower count. While following count for Venezuela, Iran770, and RIRA exhibit conformity to the law, the rest of them do not. Regarding like count, Russia416, Venezuela764, Iran770, Venezuela1196, Iran2320, and Bangladesh represent conformity. Reply count statistics for Russia416, Iran770, Venezuela1196, Iran2320 adhere the law. Retweet count behavior Russia416, Venezuela764, Iran770, Venezuela1196, Iran2320, conform to the law and only Bangladesh does not. In other words, we cannot say all trolls' activities deviate from the Benford's law distribution.

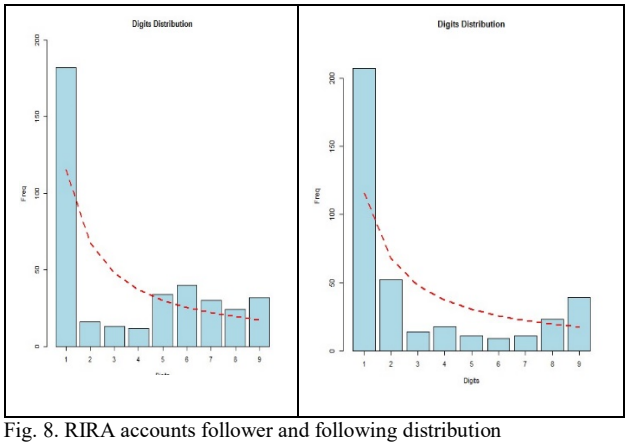

Fig. 8. RIRA accounts follower and following distribution

These findings are coherent with findings of [35] and give more specific insight about automated accounts and accounts run by foreign agents. One might be interested in investigating the use of this law as an underlying tool for auditing online world, rising this question "Are all activities not conforming to Benford's law classified into malicious activities?" or "Can Social media platforms use Benford's Law as a solely tool to detect and suspend the accounts?".

VII. CONCLUSION

We analyzed the online anomalous behavior in terms of conforming to the law of normal systems. Previous research shows that all normal behaviors have similar patterns in that the probability of occurrence of number 1 is more than number 2 and so on. This study was the first attempt to examine the activities of automated accounts and IO in their adherence to the law of normal systems. Our findings indictaed that bots' activities do not adhere to the Benford's law. Expectedly, the real accounts represented conformity to the law. These results are in line with the findings of previous research which found that Benford's law applies to online social networks. However, the IO activities did not show any consistent pattern in this regard. While some features showed conformity, some did not. These results have implications in better understanding the malicious online activities and contributes to the field of cyber security. Future research will consider examining other types of bots' activities such as Chatbots, Webrobots, and bots on other platforms.

REFERENCES

[1] C. Durtschi, W. Hillison, and C. Pacini, "The Effective Use of Benford ' s Law to Assist in Detecting Fraud in Accounting Data," *J. Forensic Account.*, vol. 99, no. 99, pp. 17–34, 2004, doi: DOI:

[2] P. Clippe and M. Ausloos, "Benford's law and Theil transform of financial data," *Phys. A Stat. Mech. its Appl.*, vol. 391, no. 24, pp. 6556–6567, 2012.

[3] G. Judge and L. Schechter, "Detecting problems in survey data using Benford's Law," *J. Hum. Resour.*, vol. 44, no. 1, pp. 1–24, 2009.

[4] M. Sambridge, H. Tkalčić, and A. Jackson, "Benford's law in the natural sciences," *Geophys. Res. Lett.*, vol. 37, no. 22, p. n/a-n/a, Nov. 2010, doi: 10.1029/2010GL044830.

[5] T. A. Mir, "The law of the leading digits and the world religions," *Phys. A Stat. Mech. its Appl.*, vol. 391, no. 3, pp. 792–798, 2012.

[6] C. Durtschi, W. Hillison, and C. Pacini, "The Effective Use of Benford ' s Law to Assist in Detecting Fraud in Accounting Data," *J. Forensic Account.*, vol. 99, no. 99, pp. 17–34, 2004, doi: DOI:

[7] S. Franklin and A. Graesser, "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents," in *International Workshop on Agent Theories, Architectures, and Languages*, 1996, pp. 21–35.

[8] M. Tsvetkova, R. García-Gavilanes, L. Floridi, and T. Yasseri, "Even good bots fight: The case of Wikipedia," *PLoS One*, vol. 12, no. 2, p. e0171774, Feb. 2017, doi: 10.1371/journal.pone.0171774.

[9] R. Gorwa and D. Guilbeault, "Unpacking the Social Media Bot: A Typology to Guide Research and Policy," *Policy and*

Internet, vol. 9999, no. 9999, pp. 1–24, 2018, doi: 10.1002/poi3.184.

[10] N. Chavoshi, H. Hamooni, and A. Mueen, "DeBot: Twitter bot detection via warped correlation," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, no. October 2017, pp. 817–822, 2017, doi: 10.1109/ICDM.2016.86.

[11] A. Arif, L. G. Stewart, and K. Starbird, "Acting the part: Examining information operations within# blacklivesmatter discourse," *Proc. ACM Human-Computer Interact.*, vol. 2, no. CSCW, p. 20, 2018.

[12] S. Zannettou, T. Caulfield, W. Setzer, M. Sirivianos, G. Stringhini, and J. Blackburn, "Who Let The Trolls Out?: Towards Understanding State-Sponsored Trolls," in *Proceedings of the 10th ACM Conference on Web Science*, 2019, pp. 353–362.

[13] V. S. Subrahmanian *et al.*, "The DARPA Twitter bot challenge," *Computer (Long. Beach. Calif).*, vol. 49, no. 6, pp. 38–46, 2016.

[14] C. M. Zhang and V. Paxson, "Detecting and analyzing automated activity on twitter," in *International Conference on Passive and Active Network Measurement*, 2011, pp. 102–111.

[15] B. Kollany, P. N. Howard, and S. C. Woolley, "Bots and automation over Twitter during the U.S. Elections," 2016.

[16] A. Badawy, E. Ferrara, and K. Lerman, "Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign," *Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2018*, pp. 258–265, 2018, doi: 10.1109/ASONAM.2018.8508646.

[17] T. P. Hill and others, "A statistical derivation of the significant-digit law," *Stat. Sci.*, vol. 10, no. 4, pp. 354–363, 1995.

[18] Z. Gilani, R. Farahbakhsh, G. Tyson, L. Wang, and J. Crowcroft, "Of bots and humans (on twitter)," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 349–354.

[19] Z. Gilani, R. Farahbakhsh, G. Tyson, L. Wang, and J. Crowcroft, "An in-depth characterisation of Bots and Humans on Twitter," pp. 1–18, 2017.

[20] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu, "A New Approach to Bot Detection: Striking the balance between Precision and Recall," *IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, pp. 533–540, 2016.

[21] S. Kudugunta and E. Ferrara, "Deep neural networks for bot detection," *Inf. Sci. (Ny).*, vol. 467, pp. 312–322, 2018, doi: 10.1016/j.ins.2018.08.019.

[22] Y. Boshmaf, K. Beznosov, and M. Ripeanu, "Graph-based Sybil detection in social and information systems," pp. 466–473, 2014, doi: 10.1145/2492517.2492568.

[23] D. M. Beskow and K. M. Carley, "Bot conversations are different: Leveraging network metrics for bot detection in Twitter," *Proc. 2018 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Mining, ASONAM 2018*, pp. 825–832, 2018, doi: 10.1109/ASONAM.2018.8508322.

[24] D. M. Beskow and K. M. Carley, "Its all in a name: detecting and labeling bots by their name," *Comput. Math. Organ. Theory*, vol. 25, no. 1, pp. 24–35, 2018, doi: 10.1007/s10588-018-09290-1.

[25] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on twitter: Are humans more opinionated than bots?," in *Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014, pp. 620–627.

[26] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *Proceedings of the 25th International Conference Companion on World Wide Web*, 2016, pp. 273–274.

[27] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, no. 7, pp. 96–104, 2016.

[28] A. Minnich, N. Chavoshi, D. Koutra, and A. Mueen, "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*, 2017, pp. 467–474.

[29] E. Ferrara, O. Varol, F. Menczer, and A. Flammini, "Detection of Promoted Social Media Campaigns," *Proc. Tenth Int. AAAI Conf. Web Soc. Media*, no. Icwsm, pp. 563–566, 2016.

[30] S. Newcomb, "Note on the Frequency of Use of the Different Digits in Natural Numbers," *Am. J. Math.*, vol. 4, no. 1/4, pp. 39–40, 1881, doi: 10.2307/2369148.

[31] F. Benford, "The law of anomalous numbers," *Am. Philos. Soc.*, pp. 551–572, 1938.

[32] L. Zhipeng, C. Lin, and W. Huajia, "Discussion on Benford's Law and its application." Cornell University Library, 2004.

[33] L. Zhipeng, C. Lin, and W. Huajia, "Discussion on Benford's Law and its application." Cornell University Library, 2004.

[34] M. Sambridge, H. Tkalčić, and A. Jackson, "Benford's law in the natural sciences," *Geophys. Res. Lett.*, vol. 37, no. 22, p. n/a-n/a, Nov. 2010, doi: 10.1029/2010GL044830.

[35] J. Golbeck, "Benford's law applies to online social networks," *PLoS One*, vol. 10, no. 8, pp. 1–11, 2015, doi: 10.1371/journal.pone.0135169.

[36] K. Lee, B. D. Eoff, and J. Caverlee, "Seven Months with the Devils: An in-depth characterisation of Bots and Humans on Twitter," *Proc. Fifth Int. AAAI Conf. Weblogs Soc. Media*, pp. 185–192, 2011, doi: 10.1145/3110025.3110090.

[37] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "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*, 2017, pp. 963–972.

[38] M. J. Nigrini, *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*, vol. 586. John Wiley & Sons, 2012.

[39] Z. Gilani, R. Farahbakhsh, G. Tyson, and J. Crowcroft, "A Large-scale Behavioural Analysis of Bots and Humans on Twitter," *ACM Trans. Web*, vol. 13, no. 1, p. 7, 2019.