# A Systematic review of Techniques to Spot Spammers on Twitter

**Pankaj Verma[1], Dr. Sunita Mahajan[2]**
Research Scholar, Assistant Professor
Department of Computer Science,
Arni University Kathgarh Indora, Himachal Pradesh.

## ABSTRACT

Social networking sites' rapid growth as platforms for communication, information sharing, storage, and management draws hackers who use the Internet to find security holes and take advantage of them for illegal gain. New fake internet accounts are created every day. On online social networks (OSNs), impersonators, phishers, scammers, and spammers are more prevalent and challenging to identify. Spammers are those who send a lot of unsolicited messages with the intention of advertising a product, tricking recipients into clicking on dangerous links, or infecting their computers in order to gain money. It has been extensively studied how to find junk profiles in OSNs. In this study, we examined the methods currently in use to identify spam users on the Twitter social network. Features for spam detection may be based on the user, the content, or both. The current study provides a summary of the techniques, attributes used, success rate, and limitations (if any) for identifying spam accounts, particularly on Twitter.

**Keywords:** Twitter, legitimate users, online social networks (OSNs), and spammers.

## INTRODUCTION

A social networking site allows users to: (a) build an account; (b) friend a list of other users; and (c) look through and explore their individual and other users' buddy lists, according to Boyd et al. Through the use of Web 2.0 technology, these online social networks (OSNs) enable user interaction. These social networking sites are expanding quickly and altering how individuals communicate with one another. In less than 8 years, these websites have evolved from a niche sector of surfing to a phenomenon that draws billions of internet users. Individuals with similar interests can connect with one another more easily thanks to online groups that bring them together.

Sixdegrees.com was the first social networking site to launch in 1997, and makeoutclub.com followed in 2000. Sixdegrees.com and similar websites had a short lifespan and quickly faded, while new websites like MySpace, LinkedIn, Bebo, Orkut, Twitter, etc. found success. Facebook, a very well-known website, was introduced in 2004 [5] and rapidly rose to fame throughout the globe. OSNs' greater user numbers make them more appealing targets for spammers and malevolent users. On social media websites, spam can take many forms and is difficult to identify. Everybody who has used the Internet has encountered spam of some kind, whether it is in emails, forums, newsgroups, etc. Spam [18] is defined as the practice of sending unsolicited bulk messages over electronic messaging systems. OSNs have grown in popularity and are now used as a platform for spam distribution. Spammers want to send product advertisements to users who are not connected to them. Some spammers post links that lead to phishing websites where users' sensitive information is stolen.

The identification of spam accounts in OSNs has been the subject of numerous papers. However, no review paper that consolidates the available research has yet been published in this sector. The purpose of our paper is to examine the academic research and work that has been done in this area by various scholars and to highlight the potential directions for future research. The methods for identifying spammers on Twitter have been analyzed and compared in this research, along with their presentations. The structure of this paper is as follows: The approach utilized to conduct this review is described in Section 2; subsequently, security vulnerabilities in OSNs were described in Section 3; spammers are defined in Section 4 along with their motivations; Sections 5 and 6 offer an introduction to Twitter and its dangers; Part 7 discusses the motivation for this survey paper, and Section 8 discusses the attributes that can be used to aid in detection. A comparative examination of the research produced by various researchers is reviewed in Part 8; new researchers are given research recommendations in Section 10; and the review is concluded in Section 11.

## 1. METHODOLOGY

The methods for detecting spam profiles in OSNs were surveyed after conducting a systematic review with a principled approach and searching major computer science research databases like IEEE Xplore, ACM Digital Library, SpringerLink, Google Scholar, and ScienceDirect for pertinent topics. We concentrated exclusively on articles published after 2009 since social networks were not conceptualized until 1997 [1] and only afterwards did they gain widespread acceptance. Later, in 2004 [1], Facebook was introduced, and it quickly gained popularity. However, it took some time for people to become accustomed to using these networks for communication, which is why they were the target of attacks.

Around 60 papers were found after searching the five major databases mentioned above. After reviewing all of the paper titles and abstracts, the papers that will be reviewed for this survey were chosen. Only papers that were deemed appropriate for the current investigation were selected. 21 papers in total have been chosen for examination after publications with titles and abstracts relating to spam message detection and other unrelated areas were eliminated. The majority of the criteria used to identify spammers have been used to categorize the publications. We are attempting to create a list of social networking papers on Twitter spam profile identification that we have read throughnd abstracts, the papers that will be reviewed for this survey were chosen. Only papers that were deemed appropriate for the current investigation were selected. 21 papers in total have been chosen for examination after publications with titles and abstracts relating to spam message detection and other unrelated areas were eliminated. The majority of the criteria used to identify spammers have been used to categorize the publications. We are attempting to create a list of social networking papers on Twitter spam profile identification that we have read through. The list may be lacking some items, but it helps to clarify the current research on identifying social network spammers. After reading this survey study, new researchers will find it easy to assess what research has been done, when it was done, and how the current body of work may be expanded to improve spam detection. When applicable, we have provided specifics on the approach utilized, the dataset used, the features for spammer detection, and the efficacy of the methods employed by different authors.

The papers discuss, in particular, the ramifications of spammers' interactions with members of social networks as well as current methods for identifying them.

## 2. OSN SECURITY PROBLEMS

Online social networking sites (OSNs) are susceptible to security and privacy problems due to the volume of user data that these sites process daily. Social networking site users are vulnerable to a range of attacks:

1) Viruses: spammers utilize social networking sites as a distribution channel [19] for dangerous files to infect users' systems.

2) Phishing attacks: By pretending to be a reliable third party, users' sensitive information is obtained [30].

3) Spammers send unwanted emails to social media users [11].

4) Sybil (fake) attack: To undermine the reputation of trustworthy network users, the attacker creates many false identities and poses as the real deal in the system [20].

5) Social bots: a group of fictitious personas made to harvest user information [32].

6) Cloning and identity theft, attacks in which perpetrators establish fake profiles of users who already exist on the same network or on different networks in an effort to deceive the cloned user's friends [23]. Attackers will gain access to victims' information if they accept the friend requests provided by these cloned identities. Users and systems are overextended by these attacks.

## 4. CLASSIFICATION OF SPAMMERS

The information provided by legitimate users is contaminated by spammers, who represent a threat to the safety and confidentiality of social networks. the following subsequent categories best describes spammers [22]:

1. Phishers are people who pose as regular users in order to obtain personal information from other legitimate users.
2. Fake Users: These are persons who spoof real users' profiles in order to distribute spam to their friends or other network users.
3. Promoters: those who transmit harmful advertisements or other links to promotions to other people in an effort to get their personal information.

**Spammers' Motives:**

a) Spread pornographic material

b) Transmit viruses

c) Phishing Attacks

d) Affect the reputation of the system

## 5. TWITTER AS A SOCIAL NETWORK

### 5.1 Introduction

As of right now, 500 million people utilize the social networking site Twitter to communicate information. It was first introduced on March 21, 2006 [14]. The name Twitter comes from the fact that Twitter's logo is a chirping bird. Users can use it to retrieve frequently exchanged information known as "tweets," which are public communications of up to 140 characters.

These tweets are automatically public by default, making them available to anybody who is following the tweeter. Users share these tweets, which may contain news, comments, photographs, videos, links, and messages. The following are some commonly used and relevant terms for our work on Twitter:

Tweets [3]: Twitter messages are limited to 140 characters.

Followers and Followings [3]: Users who a particular user follows are known as followers, whereas users who a user follows are known as followings.

Retweet [3]: a tweet that an individual has shared again with all of their followers.

Hashtag [3]: To make specific topics or keywords in a tweet more easily searchable, the # sign is used to tag them.

Mention [3]: By using the @ sign in front of other users' usernames, you can include their replies and mentions in tweets.

Lists [3]: Twitter offers a tool for grouping the people you follow into lists.

Direct Message [3]: Often known as a "DM," this designates the system used by Twitter for direct messaging between users.

According to Twitter policy [16], signs of spam profiles include metrics like following a lot of users quickly, having a post that is just links, using popular hashtags (#) when providing unrelated content, and consistently posting other users' tweets as your own. By tweeting to @spam, users have the option to report spammy profiles to Twitter. However, the Twitter policy [16] does not make it clear whether managers utilize user reports or automated processes to look for these circumstances, despite the fact that it is assumed that both approaches are used.

### a. Threats on Twitter

1. Spammed Tweets [13]: Twitter users are only permitted to publish tweets with a maximum of 140 characters, but despite this limitation, cybercriminals have found a method to take advantage of it by generating concise yet attractive tweets that include links to promos for free vouchers, job advertisements, or other promotions.

2. Downloads of malware [13]: Cybercriminals have used Twitter to disseminate tweets with links to websites where malware can be downloaded. The Twitter worms that transmitted direct messages and even malware that infected both Windows and Mac operating systems include FAKEAV and backdoor [13] programs. KOOBFACE [13], a piece of social media virus that attacked both Facebook and Twitter, has the worst reputation.

3. Twitter bots: Online criminals frequently utilize Twitter to manage and command botnets [13]. These botnets threaten the security and privacy of the users by controlling their accounts.

## 6. The Social Consequences of OSNs

In addition to the typical issues that social networking sites bring for users, such as spamming, phishing assaults, malware infestations, social bots, viruses, etc., the biggest challenge is maintaining the security and confidentiality of private data.

Social networking websites are created with the intention of making information readily available and accessible to others. But tragically, cybercriminals exploit this information, which is readily accessible, to launch focused assaults. Attackers can easily find a means to gain access to one user's account so that they can use that information to access other user accounts and the accounts of their friends.

## 7. MOTIVATION FOR REVIEW

Social networks have been a target for spammers due to the simplicity of information sharing and the ability to stay up-to-date on current subjects. It can be challenging to identify such fraudulent individuals in OSNs because spammers are well-aware of the methods available to identify them. For the purpose of collecting money, spammers can utilize OSNs as the ideal platform to pose as legitimate users and attempt to convince innocent users to click on harmful posts. The most crucial area being researched by numerous experts is how to identify such people in order to safeguard the network and protect users' private information. In order to quickly evaluate the work that has been done in this field, researchers will find this paper to be of great assistance.

## 8. TWITTER FEATURES DISTINGUISHING SPAMMERS AND NON-SPAMMERS

The papers analyzed in this study are shown in Table 1, along with the types of features that were utilized to identify spam Twitter profiles. Spam and non-spam profiles can be distinguished by either user-based or content-based characteristics. In any social network, user-based features are the characteristics of the user's profile and behavior, whereas content-based features are the characteristics of the text that users publish.

Table 1 lists the features for spotting spam profiles.

| **Characteristics used to identify spammer profiles** |
|---|
| User-based features: These comprise demographic information such as a user's profile information, number of followers and followings, followers-to-followers ratio, reputation, account age, average time between tweets, posting habits, idle hours, tweet frequency, etc. [33,12,34,3,26] |
| Content-based features: which include the quantity of hashtags (#), the quantity of URLs in tweets, @mentions, retweets, spam terms, and trending topics; duplicate tweets; HTTP links; etc. [33,7,11,25] |
| Both user- and content-based [1, 22, 24, 27, 29, 2, 4] |
| Any additional features, such as graph connectedness or pictorial distance: Graph-based features, neighbor-based features, interaction-based features, social links, social activities, and the Markov clustering method [21,9,28,33,23,6] |

Function of the aforementioned features in identifying spam profiles in accordance with Twitter rules [16]:

1. Numbers: of followers- Spammers have fewer followers.

2. The number of followers—Spammers frequently follow a lot of users.

3. Followers/Following Ratio: Spammers have a ratio of less than 1.

4. The ratio of followers to the total of followers and followings is referred to as reputation. Spammers are well-known.

5. Age of account-Current date and account creation date are used to determine the age of the account. While new accounts are more common among spammers, this feature is less useful to them.

6. The average amount of time between posts – spammers send out more tweets quickly to attract attention.

7. Posting time behavior: Spammers frequently post at set times, whether it's early in the morning or late at night when real users aren't using social networking sites.

8. Idle hours - spammers continue to send messages to reduce their idle hours.

9. Tweet frequency: To attract other users' attention, spammers tweet more frequently and at unusual hours.

10. The number of hashtags (#) used by spammers to entice genuine users to read their tweets by posting numerous unrelated updates to the most popular topics on Twitter.

11. Number of URL's - A huge number of URLs to harmful websites are included in spammers' tweets.

12. @mentions-to avoid being discovered, spammers utilize as many @usernames of unknown persons as possible in their tweets.

13. Retweets-are replies to any tweet that include the @RT symbol, and spammers use the @RT symbol the most in their tweets.

14. Spam Terms – The majority of spammers' tweets contain spam words.

15. HTTP links - Tweets produced by spammers contain the highest number of www or http://.

16. Duplicate tweets: Spammers frequently use many @usernames in their tweets to post identical tweets.

## 9. EXISTING METHODS FOR DETECTING SPAM PROFILES ON TWITTER

Researchers have employed a variety of strategies to identify the spam profile within distinct OSNs. Although Twitter is not only a social communication platform but is also used to share and disseminate information on hot subjects in real time, we are concentrating exclusively on the work that has been done to detect spammers there. An overview of the papers that were looked at about the identification of spammers on Twitter is shown in Table 2.

Table 2. A description of the methods used to identify spammers

| Author | Metrics Used | Methodology Used | Dataset Used | Results |
|---|---|---|---|---|
| Benevenuto et.al.[7] | User based and Content based | SVM | Validated on 1065 Twitter users | Accuracy 87.6% (with user based and content based features) and accuracy-84.5% (with only user based features) |
| Alex Hai Wang [1] | Graph Based and Content based | Compared NB, NN, SVM and Decision Tree | Validated on 500 Twitter users with 20 recent tweets | Naive Bayesian giving highest accuracy- 93.5% |
| Lee et. al. [22] | User based | Compared Decorate, SimpleLogistic, FT, LogiBoost, RandomSubSpace, Bagging, J48, LibSVM | Validated on 1000 Twitter users | Decorate giving highest accuracy- 88.98% |
| Gee et. al. [12] | User based | Compared NB, SVM | Validated on 450 Twitter users with 200 recent tweets | Accuracy-89.6% |
| McCord et. al. [24] | User based and content based | Compared RF, SVM, NB, K-NN | Validated on 1000 Twitter users with 100 recent tweets | Radom Forest giving highest accuracy- 95.7% |
| Lin et. al. [28] | URL rate, interaction rate | J48 | Validated on 400 Twitter users | Precision-86% |

| | | | | |
|---|---|---|---|---|
| Amit A. et. al. [2] | Introduce d 15 new features | Compared Random Forest, Decision Tree, Decorate, Naive Bayesian | Validated on 31,808 Twitter users | Accuracy-93.6% |
| Chakrabor ty et. al. [4] giving highest | User based; Content based | Compared Random Forest, SVM, Naïve Bayes, Decision | Tree Trained on 5000 Twitter users with 200 recent tweets | SVM accuracy-89% |
| Yang et. al. [6] | 18 features (8-existing & 10 new features introduce d) | Compared Random Forest, Decision Tree, Decorate, Naive Bayesian | Validated on two datasets-5000 users and then 3500 users with 40 recent tweets | Bayesian giving highest accuracy-88.6% |

In 2010, Alex Hai Wang [1] made substantial advancements in the field of spam profile detection utilizing both users- and content-based features. An early version of the spam identification system has been shown to locate questionable Twitter users. A directed social graph model has been proposed to examine the "follower" and "friend" relationships. In compliance with Twitter's spam policy, spam detection has been made simpler by using a Bayesian method of classification along with user- and content-based features. Traditional classification methods such as decision trees, support vector machines (SVM), naïve Bayesian, and neural networks have all had their performance examined using common assessment metrics, and it has been determined that the Bayesian classifier performs the best of all of them. Between the 500 people in the test dataset and the 2,000 users in the crawl dataset, the system achieved 93.5% accuracy and 89% precision. The drawback of this approach is that it was initially evaluated on a tiny dataset of 500 individuals while considering their 20 recent tweets.

Using social honeypots made up of real profiles, Lee et al [22].'s bot gathered proof of spam by browsing the profile of the user sending the unsolicited friend requests and URLs on MySpace and Twitter. Spammers have been identified using characteristics of profiles such as their posting habits, content, and friend information to build machine learning classifiers. After investigation, profiles of users who contacted these social honeypots on Twitter and MySpace via unsolicited friend requests have been gathered. For identifying spammers, the LIBSVM classifier has been utilized. One positive aspect of the approach is that it has been tested on two different dataset combinations: 10% spammers+90% non-spammers and 10% non-spammers+90% spammers. The approach has a drawback in that not as much data has been used for validation.

Based on the content of tweets and user-based attributes, Benevenuto et al. [7] identified spammers. The following tweet content attributes are used: the quantity of hashtags per word, the quantity of URLs per word, the quantity of words per tweet, the quantity of characters per tweet, the quantity of hashtags per tweet, the quantity of numeric characters in the text, the quantity of users mentioned in each tweet, and the quantity of points in time the twitter post has been retweeted. The features that set spammers apart from non-spammers include the percentage of tweets that contain URLs, the percentage of tweets that contain spam words, and the average number of words that are hashtags on the tweets. On Twitter, a dataset of 54 million users has been crawled, and 1065 users have been manually classified as spammers and non-spammers. Spammers and non-spammers have been separated using supervised machine learning, or SVM, classifiers. The system's detection accuracy is 87.6%, with only 3.6% of non-spammers incorrectly classified.

By sending a message to "@spam," Twitter makes it easy for users to report spam accounts to them. Gee et al. [12] took advantage of this property and used a classification technique to find spam profiles. With the use of the Twitter API, both legitimate user profiles and spam accounts have been compiled. The collected data was first represented in JSON before being provided in CSV format as a matrix. Users are rows in the matrix, and features are columns. Then, CSV data were trained using the SVM algorithm with a 10% error rate after first utilizing the Naive Bayes technique with a 27% error rate. 89.3% of spam profiles can be accurately detected. It has been stated that aggressive system deployment should only be done if precision is greater than 99%. This approach's limitation is that not very specialized features have been used for detection, and precision is also low, at 89.3%.

McCord et al. [24] employed content-based features such as the quantity of links, replies or mentions, retweets, and hashtags as well as user-based features like the quantity of friends and followers. Spam profiles on Twitter have been identified using classifiers including Random Forest, Support Vector Machine (SVM), Naive Bayesian, and K-Nearest Neighbor. The Random Forest classifier, which yields the best results after the SMO, Naive Bayesian, and K-NN classifiers, has been validated on 1000 users with 95.7% precision and 95.7% accuracy. As a result of the unbalanced dataset used and the fact that Random Forest is typically used in cases of unbalanced datasets, this approach's limitation is that the reputation feature has been giving incorrect results for the considered dataset, failing to distinguish between spammers and non-spammers. Finally, the approach has only been validated with a small sample size.

Using two distinct features—URL rate and interaction rate—Lin et al. [28] identified persistent spam accounts on Twitter. Many different indicators, including the number of followers, number of followings, followers-to-following ratio, tweet content, number of hashtags, URL links, etc., have been utilized by the majority of publications to identify spam accounts. However, according to this study, all of these criteria are not very useful for identifying spammers, so only two straightforward yet useful features—URL rate and interaction rate—have been applied. The ratio of tweets with URLs to all tweets is known as the URL rate, while the ratio of tweets that interact with one another is known as the interaction rate. The Twitter API was used to crawl 26,758 accounts, and J48 classifier analysis was performed on 816 long-surviving accounts with an accuracy rate of 86%. The approach's limitation is that only two variables were utilized to detect spam profiles; hence, if spammers maintain low URL rates and low interaction rates, the system will not function as planned.

There are two different kinds of spammer detection systems, according to Amit A. et al. [2]: one is URL-centric, which relies on identifying fraudulent URLs, and the other is user-centric, which is based on features relating to people such as followers and following ratios. The method used in this research is a hybrid one that takes into account both of the properties listed above. Together with an alert system to identify spam tweets, 15 new features have been proposed to catch spammers. Spammers' tweet campaigns and methods have also been researched. A dataset from Twitter with 500K users and another with 110,789 individuals were both used. Bait-oriented features, which highlight the strategies used by spammers to get victims to click on harmful links, include mentions of non-followers, trend hijacking, and trend intersection with well-known trends. Tweet interval variation, tweet volume variation, the ratio of tweet interval variation to tweet volume variation, and tweeting sources are examples of behavioral characteristics. Duplicate URLs, duplicate domain names, and an IP/domain ratio are examples of URL characteristics. Dissimilarity of tweet content, similarity of tweets, and URL and tweet similarity are all examples of content entropy properties. Follower/following ratio and the profile's description language dissimilarity are aspects of the profile. Then, using the Weka tool, all of these features were gathered from both malicious and benign users and fed into four supervised learning algorithms: decision tree, random forest, bayes network, and decorate. Using Decorate's classifier, which produces the best results, 93.6% of spammers have been found. It has been demonstrated that this method performs better than Twitter's spammer detection strategy. However, this method has only been tested on 31,808 individuals, whereas Twitter is taking into account millions of users.

A technique to identify abusive users that publish offensive content, including dangerous URLs, pornographic URLs, and phishing links, drive regular users from social networks, and violate their privacy has been presented by Chakraborty et al. [4]. The algorithm has two steps: the first checks a user's profile for offensive content before sending a friend request to another user, and the second checks the similarities between two profiles. The system is designed to advise the user whether or not to accept a friend request after these first two steps. It has been tested with a 5000-user Twitter dataset that was gathered using the REST API. Timing, content, and profile-based criteria are all taken into account when determining how to distinguish between abusive and non-abusive users. There have been SVM, Decision Tree, Random Forest, and Naive Bayesian classifiers employed. All classifiers are outperformed by SVM, and the model is operating at an accuracy of 89%.

Yang et al. [6] used new features to identify spammers on Twitter. There have been discussions about a number of evasion strategies used by spammers. Ten new detection features have been proposed, such as three graph-based features, three neighbor-based features, three automation-based features, and one timing-based feature. These features are expensive and difficult to get around because they are based on techniques that spammers don't use to avoid detection and require more time, money, and resources. With the help of classifiers like Random Forest, Decision Tree, Decorate, and Bayesian Network, 18 features—eight already existing and ten new—have been examined for detection purposes. A Bayesian classifier's accuracy of 88.6% is the best. This method has a limitation in that very little data has been crawled and only a specific sort of spammer is being found with a low detection rate, which is the minimum number of spammers found in the dataset.

## 10. RESEARCH DIRECTIONS

During the survey, it became pretty clear that there has been a lot of work done to identify spam profiles in various OSNs. Even so, the detection rate can be improved by switching up the method and using more substantial features as the determining factor. The following are a few findings from the survey:

1. Since Twitter has a billion active users and this number is constantly growing, almost all of the authors employed incredibly small testing datasets to assess the efficacy of their strategy. In order to evaluate the effectiveness of any strategy, it is necessary to expand the testing dataset.
2. A multivariate model needs to be created, second.
3. There is a need to create a technique that can identify various spammers.
4. The methods need to be tested on various mixtures of spammers and non-spammers.

## 11. CONCLUSION

Numerous methods have been developed and used by researchers to find spammers on different social networks. As may be concluded from the publications reviewed, the majority of the work has been done using classification techniques like SVM, Decision Tree, Naive Bayesian, and Random Forest. User-based features, content-based features, or a combination of both have been used for detection. A few authors additionally added new detecting features. A relatively small dataset was used to validate each method, and diverse combinations of spammers and non-spammers were not even tried. Numerous methods have been developed and used by researchers to find spammers on different social networks. As may be concluded from the publications reviewed, the majority of the work has been done using classification techniques like SVM, Decision Tree, Naive Bayesian, and Random Forest.

## 12. REFERENCES

[1] Wang, A. H. (2010, July). Don't follow me: Spam detection in twitter. In *2010 international conference on security and cryptography (SECRYPT)* (pp. 1-10). IEEE.

[2] Amleshwaram, A. A., Reddy, N., Yadav, S., Gu, G., & Yang, C. (2013, January). Cats: Characterizing automation of twitter spammers. In *2013 Fifth International Conference on Communication Systems and Networks (COMSNETS)* (pp. 1-10). IEEE.

[3] Malhotra, A., Totti, L., Meira Jr, W., Kumaraguru, P., & Almeida, V. (2012, August). Studying user footprints in different online social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1065-1070). IEEE.

[4] Chakraborty, A., Sundi, J., & Satapathy, S. (2012). SPAM: a framework for social profile abuse monitoring. *CSE508 report, Stony Brook University, stony brook, NY*.

[5] Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, *13*(1), 210-230.

[6] Yang, C., Harkreader, R., & Gu, G. (2013). Empirical evaluation and new design for fighting evolving twitter spammers. *IEEE Transactions on Information Forensics and Security*, *8*(8), 1280-1293.

[7] Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010, July). Detecting spammers on twitter. In *Collaboration, electronic messaging, anti-abuse and spam conference (CEAS)* (Vol. 6, No. 2010, p. 12).

[8] Clearinghouse, P. R. (2014). Fact Sheet 35: Social networking privacy: How to be safe, secure and social.

[9] Ahmed, F., & Abulaish, M. (2012, June). An mcl-based approach for spam profile detection in online social networks. In *2012 IEEE 11th international conference on trust, security and privacy in computing and communications* (pp. 602-608). IEEE.

[10] Kontaxis, G., Polakis, I., Ioannidis, S., & Markatos, E. P. (2011, March). Detecting social network profile cloning. In *2011 IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops)* (pp. 295-300). IEEE.

[11] Stringhini, G., Kruegel, C., & Vigna, G. (2010, December). Detecting spammers on social networks. In *Proceedings of the 26th annual computer security applications conference* (pp. 1-9).

[12] Gee, G., & Teh, H. (2010). Twitter spammer profile detection. *Available online: cs229. stanford. edu/proj2010/GeeTeh-Twitter Spammer Profile Detection. pdf*.

[13] http://about-threats.trendmicro.com/us/webattackInformation regarding Twitter threats.

[14] http://en.wikipedia.org/wiki/Twitter-Information of Twitter.

[15] http://expandedramblings.com/index.php/march-2013-bythe-numbers-a-few-amazing-twitter-stats-Regarding statistics of Twitter.

[16] http://help.twitter.com/forums/26257/entries/1831- The Twitter Rules.

[17] http://twittnotes.com/2009/03/, 2000-following-limit-ontwitter.html-The 2000 Following Limit Policy on Twitter.

[18] http://www.spamhaus.org/consumer/definition-Spam Definition.

[19] Baltazar, J., Costoya, J., & Flores, R. (2009). The real face of koobface: The largest web 2.0 botnet explained. Trend Micro Research, 5(9), 10.

[20] Douceur, J. R. (2002). The sybil attack. In *Peer-to-Peer Systems: First InternationalWorkshop, IPTPS 2002 Cambridge, MA, USA, March 7–8, 2002 Revised Papers 1* (pp. 251-260). Springer Berlin Heidelberg.

[21] Song, J., Lee, S., & Kim, J. (2011). Spam filtering in twitter using sender-receiver relationship. In *Recent Advances in Intrusion Detection: 14th International Symposium, RAID 2011, Menlo Park, CA, USA, September 20-21, 2011. Proceedings 14* (pp. 301-317). Springer Berlin Heidelberg.

[22] Lee, K., Caverlee, J., & Webb, S. (2010, July). Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 435-442).

[23] Bilge, L., Strufe, T., Balzarotti, D., & Kirda, E. (2009, April). All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th international conference on World wide web* (pp. 551-560).

[24] Mccord, M., & Chuah, M. (2011). Spam detection on twitter using traditional classifiers. In *Autonomic and Trusted Computing: 8th International Conference, ATC 2011, Banff, Canada, September 2-4, 2011. Proceedings 8* (pp. 175-186). Springer Berlin Heidelberg.

[25] Egele, M., Stringhini, G., Kruegel, C., & Vigna, G. (2013, February). Compa: Detecting compromised accounts on social networks. In *NDSS* (Vol. 13, pp. 83-91).

[26] Flores, M., & Kuzmanovic, A. (2013). Searching for spam: detecting fraudulent accounts via web search. In *Passive and Active Measurement: 14th International Conference, PAM 2013, Hong Kong, China, March 18-19, 2013. Proceedings 14* (pp. 208-217). Springer Berlin Heidelberg.

[27] Conti, M., Poovendran, R., & Secchiero, M. (2012, August). Fakebook: Detecting fake profiles in on-line social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 1071-1078). IEEE.

[28] Lin, P. C., & Huang, P. M. (2013, January). A study of effective features for detecting long-surviving Twitter spam accounts. In *2013 15th international conference on advanced communications technology (ICACT)* (pp. 841-846). IEEE.

[29] Lee, S., & Kim, J. (2012, February). Warningbird: Detecting suspicious urls in twitter stream. In *Ndss* (Vol. 12, pp. 1-13).

[30] Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, *50*(10), 94-100.

[31] Balasubramaniyan, V. A., Maheswaran, A., Mahalingam, V., Ahamad, M., & Venkateswaran, H. (2010). *A crow or a blackbird?: Using true social network and tweeting behavior to detect malicious entities in twitter*. Georgia Institute of Technology.

[32] Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2011, December). The socialbot network: when bots socialize for fame and money. In *Proceedings of the 27th annual computer security applications conference* (pp. 93-102).

[33] Zhu, Y., Wang, X., Zhong, E., Liu, N., Li, H., & Yang, Q. (2012). Discovering spammers in social networks. In *proceedings of the AAAI conference on artificial intelligence* (Vol. 26, No. 1, pp. 171-177).

[34] Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B. Y., & Dai, Y. (2014). Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *8*(1), 1-29.

[35] Verma, P., Khanday, A. M. U. D., Rabani, S. T., Mir, M. H., & Jamwal, S. (2019). Twitter sentiment analysis on Indian government project using R. *Int J Recent Technol Eng*, *8*(3), 8338-41.

[36] Verma, P., & Jamwal, S. (2020). Mining public opinion on Indian Government policies using R. *Int. J. Innov. Technol. Explor. Eng.(IJITEE)*, *9*(3).

[37] Thakur, M., & Verma, P. A Review of Computer Network Topology and Analysis Examples.

[38] Kumar, A., Guleria, A., & Verma, P. Internet of Things (IoT) and Its Applications: A Survey Paper.

[39] Thakur, N., Choudhary, A., & Verma, P. Machine Learning Algorithms-A Systematic Review.