Trust Based Fake Node Identification in Social Networking Sites

Spoorthy A S¹, Somnath sinha ²

¹ Department of Computer Science, Amrita School of Arts and Sciences, Mysore, India, spoorthysdechu@gmail.com
² Department of Computer Science, Amrita School of Arts and Sciences, Mysore, India, ssin.mca@gmail.com

Abstract. In recent years Social Networking is very commonly used for communications. It is the platform where users are free to join and share their data between groups of people. Due to the openness of this network any one can use multiple fake identifications and fraud the people into its group. These types of fake profiles are required to be identified and wipe out from the network. Some of the Social Networking Sites (SNS) that are commonly used are Twitter, Facebook, Instagram, WhatsApp, WeChat etc. These SNSs are already have some mechanism to identify fake profiles and prevent them from such type of malicious activities but are not sufficiently efficient. This paper represents comprehensive literature review for identification mechanism of fake account on Facebook and proposes a new mechanism to stop this type of harmful activities by using graph based activity diagram.

Keywords: Face book, Fake node, Online Social Networking Site (OSNs), Social networking site (SNS)

1. Introduction
SNS are the applications that are commonly used by the individual mainly for the communication purpose. The most commonly used ones are Twitter, Facebook, Instagram, WeChat, WhatsApp and so on. These are the networks where we can create multiple accounts and communicate with the users. In SNS a user has the ability to create multiple accounts, either fake or genuine. Fake profiles are used to fraud the users whereas the genuine profiles are used by the individuals in their day to day life. The SNS has many strength as well as shortcomings, strength like we can contact a person who is very far away from us where as the shortcomings is that people are getting fraud by the fake accounts through sharing their information to the unknown accounts. The statistic for the social networking site users according to the recent studies is shown below (Table 1.1)

Table 1.1: Statics of the Social Networking Sites as per 2019.

| Rank | Network  | Number of Daily Users | Monthly Visits |
|------|----------|-----------------------|----------------|
| 1    | Facebook | 1.6 billion           | 2.4 billion    |
| 2    | Google+  | 2 billion             | 395 million    |
| 3    | Twitter  | 134 million           | 330 million    |
| 4    | Linked in| 5.3 million           | 3.3 million    |
| 5    | Pinterest| 1 billion             | 291 million    |
This particular paper concentrates on one such SNS that is Facebook. Facebook was first introduced on February 4, 2004 by Mark Zuckerberg. It was mainly invented to help the college students to connect with each other. But recently it is used by every individual across the world to communicate with each other and to perform more activities like uploading photos, videos, tagging their friends and many more. Since it is open to every individual, recently there are many attacks that are taking place in this Facebook by getting the details of genuine user by a non-legitimate user.

There are many researches taking place in this area. Some of the common papers are by [1] Mohammadreza Mohammadrezaei proposed some of the methods like graph analysis and the classification algorithm to improve the efficiency of detecting the Fake profiles in the SNS, [2] Priyanka Kumari proposed method that identifies the fake profiles present in the users list using a “Social Media” application [3] Hani M. Ibrahim the proposed method uses a Novel Detection Technique that is Fake profile recognizer to verify the identity of the fake profiles. These papers give some of the ideas of eradicating the fake profile from the SNS. The main objective of the particular paper is to find these non-legitimate or fake nodes and completely eradicate its presence from the Facebook Network.

2. LITERATURE REVIEW

Facebook is one of the best SNS. Its benefits are many, may be in Business fields, entertainment field, in case of Journalism field, communication field and so on. Till date there is no proper procedure or a method that is found to eradicate the fake nodes from the Facebook. Some of the Existing systems are listed below to identify the Fake node.

The security challenges that occur in the social networking sites and about the fake nodes are discussed in by the authors Mohammadrezaei, M., Shiri, M. E., & Rahmani, A. M in the paper [1]. Some of the similarity measures such as common friends, cosine, and Jaccard, L1-measure and weight similarity are calculated. The proposed model was evaluated using Twitter dataset. Finally they have used Medium Gaussian SVM algorithm to predict the fake accounts. The paper covers some of the very important concepts like Graph Analysis and Similarity types, Introducing of Resampling, Principle Component Analysis and Machine Learning. The efficiency of the paper was expertise by Machine Learning method. The ratio of Fake accounts were identified in twitter data is 1:100 the main weakness of the proposed method is that fake account had to work in a network and from there they were able to find a legitimate or a fake account by analyzing their friends account. The future work of the paper was to present a method to identify the fake or legitimate account before performing any activities by the user account or to find the fake account during the time of the registration process.

Fake accounts are identified based on the activities and the account information is proposed by the authors Kumari, P., & Rathore, N. C in the paper [2]. To show the efficiency of the model, they developed an application that is similar to Facebook called “SocialMedia” to eliminate the Fake accounts that are found in the users friend list. The effectiveness of the proposed concept was given by this app. Online evaluation was made to identify the usage of such app in the network and the accuracy was found to be 87.5%. In the proposed system the parameters used was number of mutual friends, is family, age group, average like and average comment. After getting the parameters the trust weight calculation was done between the user’s friends. If the friends are having low trust weight then there is a high possibility of the account being a fake one. The future work was to collect other attributes of SNS users and to find the fake account with its help.

A Novel detection technique called Fake Profile Recognizer (FPR) which verify the profiles identity, and detect the fake profiles in Online Social Networking Sites is discussed by the author Meligy, A. M in the paper [3]. The detection of fake profile is based on utilizing Regular Expression (RE) and Deterministic Finite Automaton (DFA) methods. The proposed technique was evaluated on three datasets of Online Social Networking Sites: Facebook, Google+, and Twitter. The results reconnoitered
high Precision, Recall, accuracy, and low False Positive Rates (FPR) of detecting Fake Profiles in the three datasets. In twitter dataset there was a problem where Support Vector machine dint work correctly and which was later replaced by Social Honey pot.

The OSNs which are tremendously increasing the way in which people communicate and divide their personal, professional and political information is discussed by the author Conti, M., Poovendran, R., & Secchiero, M. in the paper [4]. As per the cyberspace the OSNs are attracting the Interest are trying to utilize this weakness. Recently the security and privacy issues are increasing day by day, the researchers are trying to find these and erase. The paper created an analysis of social network graphs[5] from background of privacy threats. In the future work the plan will be extended to characterize the online interaction with the user and strengthen the link and the quality of the mechanism.

The potential attacker can present the automated crawling and identity thefts that occur in SNS to get access to personal information of the user is discussed by the authors Leyla Bilge, Thorsten Strufe, Davide Balzarotti, Engin Kirda Eurecom in the paper [6]. The attack that was presented as identity theft of existing user profiles automatically and sending of friend requests to the contacts of the duplicated victim. The acceptance of friend request is the hope from the attackers. By beginning a relationship as a friend with the contacts of a victim, the attacker is able to access all the personal information provided by them.

The other method[7] was to present a valuable and feasible automated launch of cross-site profile cloning attack. So that while attacking they are able to create a fake profile in a network where the victim has not registered but contact the victim’s friend who registered on both the network.

The threats that occur in SNS and they have gone through some of the targets about how the attackers does the attack and how they perform it is discussed by the author Weimin Luo, Jingbo Liu, Jing Liu in the paper[8]. The paper is divided into 2 parts: User and Social Networking Sites. And there are some discussions on countermeasures against the attacks in Social Network and also the paper proposes a security framework towards the end. The disadvantage of the paper is towards the end they have designed a framework where the user must have some ideas on the security of their account without this the calculation of this stage becomes difficult.

The result of a research towards data security factors and wakefulness of some of the risk of SNS, and the most notable of identity theft is discussed by the author Lang, M., Devitt, J., Kelly, S., Kinneen, A., O'Malley, J., & Prunty, D in the paper[9]. The research was conducted for the people of 18-24 age groups residing in Ireland. Some of the findings were that many people had a informal outlook towards data backup and password protection, some were not aware about the attacks that was taking place through the virus transmitted by the portable device. The main disadvantage of this paper is the researcher have taken just 120 accounts to find the statistical correlation which not an effective one. Future study to prepare the questionnaire for larger number of people worldwide and to extend the study.

The SNS, fake account creation are becoming more harm than a cyber crime. Some of the algorithms and methods are being proposed for detection of the fake profile is discussed by the author Ramalingam, D., & Chinnaiah, V in the paper [10]. The paper presents some of the existing techniques to detect the fake profile. The disadvantages of the paper is that the prevention of fake account should be done rather than detecting them this method is not proposed in the paper

An attack so called automated Social engineering which explains how the SNS can be used for social engineering is discussed by the authors Huber, M., Kowalski, S., Nohlbe,rg, M., & Tjoa, S in the paper [11]. The particular research takes engineering to one step advance by automating tasks that are time intensive. The implementation was based on prototypical implementation (ASE bot), which had two
experiments that was conducted. The first experiment [12] which have examined the gathering information capabilities of the bot. In the second experiment prototype performs a Turing test. Making Facebook with restrictive security and policy may become destructive for the account this is the main disadvantage. The future work of the paper is to implement the original ASE bot to get a valuable insight.

Some of the security threats and some of the defense mechanisms regarding popular OSNs is discussed by the authors Gao, H., Hu, J., Huang, T., Wang, J., & Chen, Y. in the paper [13]. It tells about wide range of attacks and some of the defense mechanism for those attacks. The author categorize these attacks into four categories - privacy breaches, viral marketing, network structural attacks and malware attacks and focus mainly on privacy concerns. The future work is to create a SNS that has which lots of entertainment services and to secure the data of the user without any issues.

After going through these entire articles there is no perfect mechanism till date to eradicate the fake accounts from Facebook. And this becomes the main advantage of the particular article to eradicate the fake account from the Facebook using a trust based method.

3. FACEBOOK
The most popular SNS is the Facebook. And it has almost 2.41 billion active users as per the survey made in the year 2019. Facebook is not only used for entertainment purpose it has many activities in it. And it consists about 28 activities that are used by the user. Some of the activities that are performed by the Facebook are: Timeline, post view, post update, like, notification, login, log out, gaming, reading an article, Movies links, videos, profile update, comment, share posts, sending friend request, message, memories, history, tag photo, markets and many more.

Out of these activities some are more frequently used by the user are mentioned below:

a. **Post count**: Number of post uploaded by the user in his/her timeline. The Fake accounts are predictable to have lesser count of posts in his/her account. The count can be collected from the timeline of the users account.

b. **Comment count**: The average number of comments received for each post. The fake accounts are expected to have unwanted post with lesser number of comments. The counts of the comments are collected from the user timeline.

c. **Friendsversary**: The post that celebrates the day when they became friends. The Fake account does not care about celebrating anniversary. The counts of Friendversary are collected from the user timeline.

d. **Followers**: The count of the follower following the user. The Fake accounts will have a less followers count. The counts of the followers are taken from the timeline.

e. **Events**: The Number of events the user have attended. The Fake accounts do not update any events that they have attended. The number of event attended is collected from the user timeline.

f. **Friend request**: The number of friend requests accepted will be more. But in rear cases the request may be less. The Friend Request count is taken from the user’s timeline.

**CHALLENGES OF FACEBOOK:**
Some of the most popular challenges that the Facebook is facing is listed below:

1. **Battling Election Interference and Misreport:**
   Limits the spread of misreport and removes millions of the fake accounts daily. Uses some of the technology and some findings by the experts to remove bad actors. Some of the government partners, other companies, NGO’s helps in increasing the security.

2. **Strengthening the privacy and protection the information of the people:**
   It is mandatory to secure the data that is being given to the Facebook by the account holders. More than 400 apps were suspended who violated the policy. Rolled out some of the
GDPR style all over the world and tried making some of the privacy tools much easier to use. Established some of the Data use team for the maintenance purpose.

3. Prioritize the users safety and well being:
   Over 30,000 people are working for the privacy control. While working for the privacy its been detected that for about 99% of the terrorist related content is taking place in Facebook. And for about 97% of violence and the graphical contents were found. For this reason Facebook have launched an appeal to remove all these contents.

4. Giving more information to the user about the ads they see:
   It’s common that the number of ads increases in the Facebook and many among this are fake and many are the originals. Recently a new technique was introduced where it allows the user to watch those ads and report if they are fake ones. And there was a new creation of the political ads which was very transparent and it was started at US. In some of the countries the ads where posted by paying the Facebook.

5. Seeking the effective regulation:
   Facebook is working with lawmakers to make an effective and smart legislation to strike the balance between the competing tensions. It is mainly working with government for the security purpose of the platform.

4. Proposed Methodology
   The main objective of the research is to identify the suspicious account in Facebook. Here two stage mechanisms are used for identifying malicious account or node. The stages are described in the subsequent sections.

4.1 Stage 1: Measuring the authenticity of the account.
   In Facebook Most of the people are used to like the post that is posted by individual where the total number of people “likes” for the post goes to about 80%, and about 54% of people are watching “videos” for entrainment as well as knowledge purpose and some are interested to “read the article” where the count for this particular category lies for about 47% and in order to build social connections with beloved one’s, people do “messages” and the survey gives data about 47% and about 46% of people use Facebook to “news updates”. The graph of the activities is given in the Figure 4.1

![Figure 4.1: Popular activities in Facebook.](image-url)

Out of all these five categories we are just considering “Like” and “Videos”. Only these 2 parameters are considered because these both are most commonly used by every individual and the usage accuracy is also above 50%. Rest of the three that is reading an article, messages, news updates are below 50% hence these parameters cannot be considered.

Considering the Conditional probability model the probability of the of the like and videos are calculated where the probability of watching video is divided as regular and suspicious nodes where the
A regular node is further classified as watching videos and not watching videos. Similarly, the suspicious node is also divided as watching videos and not watching videos. Here the average count of users liking the post is .8% and those who don’t like the post are .2%. And the average count of watching video is .54% and the users who don’t watch video is about .46%.

The probability calculation is shown below (Figure 4.2)

$$P(S \cap NV) = 0.2 \times 0.46 = 0.092$$

$$P(R \cap NV) = 0.8 \times 0.46 = 0.368$$

$$P(S \cap V) = 0.2 \times 0.54 = 0.108$$

$$P(R \cap V) = 0.8 \times 0.54 = 0.432$$

From the above observation it is clearly explained that the probability $P(S \cap NV)$ is a highly suspicious account, $P(R \cap NV)$ and $P(S \cap V)$ are moderately suspicious accounts and $P(R \cap V)$ are the genuine accounts. All the genuine accounts are removed from this stage and the highly suspicious and moderately suspicious accounts are carried forward to the Stage 2. Here we are just categorizing some of the accounts as the highly suspicious and moderately suspicious ones. Complete identification of the fake nodes is done in the stage 2.

4.2 Stage 2: Detection of the fake nodes using direct graph.

The detection of fake node is done using the direct graph where the moderately suspicious and highly suspicious accounts are added to the direct graph where the connectivity of nodes is shown below (Figure 4.3).
Figure 4.3: Direct graph for Fake account detection

The fake nodes are detected by calculating the affinity[12]. Consider the 2 nodes as i and j between which the affinity value is calculated as

\[
A_{ij} = \frac{(T_{ij} - 2L_{ij})}{N} \frac{t_{ij} + l_{ij}}{N}
\]  

(1)

Where,

- \(T_{ij}\) = Number of similarity between event i & j.
- \(L_{ij}\) = Number of dissimilarity between event i & j.
- \(N\) = Total number of events

After calculating the affinity \((A_{ij})\) using equation(1) if the value is \(\leq 0\) then it is a suspicious account and if the value is \(>0\) then it is a genuine account. And if the value occurs in terms of Positive result then it has high affinity and if the value is in terms of negative then it has lesser affinity or no similarity between the nodes. If the more number of connections with other node is present and they are similar then the node is detected as a fake one and it is eliminated from the network else it is considered as a genuine account.

![Flowchart of the identifying fake accounts](image)

Figure 4.4: Flowchart of the identifying fake accounts

Algorithm:

Step 1: Gets the details of the user fake account.

Step 2: Identifies the suspicious accounts using the conditional probability model and categorize the account as suspicious or highly suspicious.

Step 3: The suspicious and highly suspicious accounts are checked for similarity testing by using the graphical relation.
Step 4: The accounts having more connectivity with the other account are taken as malicious account.

5. CONCLUSION AND FUTURE WORK
Throughout the paper the detection of fake nodes that are present in the Facebook is done. The literature review suggests some of the existing methods and their disadvantages in detecting the fake nodes. The paper has two stage mechanisms where the fake node can be detected. In the first stage some of the attributes of the accounts chosen and probability is calculated for those activities and in the second stage the detected fake nodes are done through direct graph by calculating the affinity. If more number of clones is present in multiple accounts then it is considered as a fake node. The paper detects the fake nodes where these fake nodes can be eliminated easily from the SNS. The Future work for the paper is to increase the accuracy by adding more general activities, and when it is detected further prevention technique can be added using some of the soft computing approach.

REFERENCES
[1] M. Mohammadrezaei, M. E. Shiri, and A. M. Rahmani, “Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms”, Hindawi Security and Communication Networks, 2018.
[2] P. Kumari and N. C. Rathore, “Fake Profile Identification on Facebook Through Social Media App,” Information and Communication Technology for Sustainable Development pp. 83–90, 2018.
[3] A. M. Meligy, “Identity Verification Mechanism for Detecting Fake Profiles in Online Social Networks,” no. January 2017, 2018.
[4] M. Secchiero, “FakeBook: Detecting Fake Profiles in On-line Social Networks,” International Conference on Advances in Social Networks Analysis and Mining 2012.
[5] K. Sreelakshmi, S. Anand, and S. Sinha, “Black Hole Attack in Mobile Ad Hoc Network – Analysis and Detection,” International Journal of Recent Technology and Engineering (IJRTE) no. February, 2019.
[6] L. Bilge, T. Strufe, D. Balzarotti, E. Kirda, and S. Antipolis, “All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks,” International World Wide Web Conference Committee (IW3C2) 2009.
[7] S. Sinha, A. Paul, S. Pal, and C. Engineering, “THE SYBIL ATTACK IN MOBILE ADHOC NETWORK: ANALYSIS AND,” “International Conference on Recent Trends in Communication and Computer Networks - ComNet 2013” organized by ACCEE,
[8] W. Luo, J. Liu, J. Liu, and C. Fan, “An Analysis of Security in Social Networks,” IEEE International Conference on Dependable, Autonomic and Secure Computing pp. 648–651, 2009.
[9] M. Lang, J. Devitt, S. Kelly, A. Kinneen, J. O. Malley, and D. Prunty, “Social Networking and Personal Data Security: A Study of Attitudes and Public Awareness in Ireland,” International Conference on Management of e-Commerce and e-Governmentno. January 2009, 2014.
[10] D. Ramalingam and V. Chinnaiyah, “Fake profile detection techniques in large-scale online social networks: A comprehensive review,” Comput. Electr. Eng., vol. 65, pp. 165–177, Jan. 2018.
[11] M. Huber, S. Kowalski, M. Nohlberg, and S. Tjoa, “Towards Automating Social Engineering Using Social Networking Sites,” pp. 117–124, 2009.
[12] R. Kulsum, S. Anand, and S. Sinha, “A Soft Computing Approach to Analyse Aodv Routing Protocol,” International Journal of Innovative Technology and Exploring Engineering (IJITEE) no. 8, pp. 1443–1446, 2019.
[13] H. Gao, J. Hu, T. Huang, J. Wang, Y. Chen, and A. Osns, “Security Issues in Online Social Networks,” Published by the IEEE Computer Society 2011.
[14] Facebook dataset| kaggle, 2018 [online]. Available from: https://www.kaggle.com/sheenabatra/facebook-data [Accessed on: Jul 15, 2018]
[15] Dustin scout, 2019 [online]. Available from: https://dustinstout.com/social-media-statistics/#google-plus-stats [Accessed on: July 2019]
[16] Facebook, 2019 [online]. Available from: https://newsroom.fb.com/news/2019/02/addressing-challenges/ [Accessed on: Feb 4, 2019]
[17] Statista Research Department. Statista, 2016 [Online]. Available from: https://www.statista.com/statistics/420714/top-facebook-activities-worldwide/ [Accessed on: Apr 16, 2016]