An Efficient Model to Detect Fake and Cloned Profiles through Similarity Analysis

[1]Vineetha Venugopal, [2]Saravanan A.
[1]Research Scholar,Department of Computer Science,
Sree Saraswathi Thyagaraja College,Pollachi
vineethavenugopal2320@gmail.com
[2]Associate Professor,Department of Computer Science,
Sree Saraswathi Thyagaraja College,Pollachi

Abstract: Individuals' use of online social networks has grown critical in the modern day. Due to amplified habit of usage of these Social Networks, the security threats have also enlarged. There are many attackers who create fake profiles to attract innocent people and abuse them in various ways. They either create fake profiles or they do identity theft from other users and create profiles using this identity. The latter method is called profile cloning. In this paper, a method has been detected to identify cloned profiles by using feature extraction method. Fake profiles are detected by using IP addressing method. For detecting cloned profiles, first the features are extracted using Web Scraper Tool. Using the search engine similar profiles are searched and features are extracted. Then a similarity measure is performed on these profiles. Based on the similarity, profiles can be verified whether they are cloned or not. After identifying cloned profiles, the users' locations are detected through the IP address for verification of fake profiles. In this work, an attempt has been made to propose a method to detect fake and cloned profiles. The method was implemented on two hundred profiles in Face book and it showed accurate results.

Keywords: Online social network (OSN) -Fake Profile -Profile Cloning- Fraud Detection Identity cloning

1. Introduction

Online social networks such as Facebook, Twitter, and Instagram have grown in popularity among internet users in recent years [1],[2],[3],[4],. Social network members exchange a variety of forms of information with their contacts. They may even share their personal data through online connections whom they have no connection in reality. Such sort of sharing becomes the major reason for being exposed to many kinds of security dangers, such stealing personal information, developing credibility in online social networks via creating phony accounts, undertaking dishonest actions online, and sending email spam.

A severe problem in cyberspace is that the exact identity of users who do scam and deceitful activities remain unknown to the users. They create false identities. They make relationships with the innocent people of these social networks using their false identities. Therefore, it is a challenging factor to maintain the actual identity of the user. Just like false identities, stealing other users' identity is also a common problem in cyberspace. It is using somebody's identity decisively without his consent. It has been regularly observed that malicious users create an account by performing identity theft from other users in the same social network or in different social networks and target victims for performing online scams. This phenomenon is known as Profile Cloning [5-10].

Profile cloning [10] may be done in two ways:- Same-site profile cloning and Cross-site profile cloning[11].

In same-site profile cloning, credentials of a user are acquired from the same online social network and a clone of the profile is made on the same network. On cross-site profile cloning, credentials of a registered user are acquired from one online social network and a replica profile is generated in some other online social network in which that individual has not enrolled. Maximum quantity of information is extracted from the user's profile to offer a sense of genuine

profile to other members of the social network [10],[11]. It is really tricky to discover such sort of profiles. As to the Norton research, in 2021, approximately 27 million Indian adults faced identity theft in the preceding 12 months [12]. So it has become extremely necessary to preserve the private information of a user. Also, a big number of internet users are ignorant as they do not have the understanding about the privacy settings. Maximum number of users do not adjust their default privacy settings for different fields as they are not aware of how this might be manipulated by the adversaries and fake users and give birth to profile clones[8],[10]. In this work, an effort has been made to explore and give technique by which we may find out whether the profile is fake or cloned.

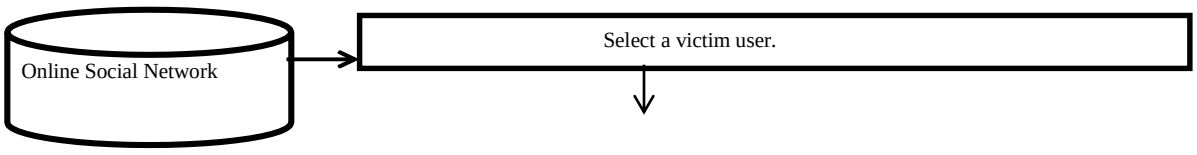
The authors presented a technique based on the similarity [13]index parameter computation. Here the user's profile attributes is given a hypothetical weight based on its significance in computing the Similarity [13] Index and the profile picture is given a weight larger than all other criteria. Profile Cloning should be identified with the user's name. Here, the user name is also used as parameter and it is checked by verification.

In this paper, a method to detect cloned and fake profiles in Online Social Networks is implemented. In this method, initially cloned profiles are detected. The detection method uses Cosine Similarity Measurements along with a threshold value. It is an efficient method to detect cloned profiles. After finding the set of cloned profiles, the method identifies fake profiles among them. Sometimes, the same user can create multiple accounts with the same data. Such profiles need not be fake profiles. Fake profile users provide fake data and pretend the victims to be genuine users. They provide wrong location in their profile. They also may not be ready to share the geo-location facilities provided in Facebook. Such users can be tracked through their IP address. Through the IP detection, their exact location can be tracked. In this work, fake users are identified through detecting their IP addresses amongst the set of cloned profiles. In this paper, an

attempt has been made to identify fake profiles and cloned profiles in Online Social Networks. It can be helpful for normal users of Online Social Networks if they find any suspicious friend requests or if they find any suspicious friends who fake them through Facebook Messenger Chats or Messages[14-19].

2. Method

In this proposed system, cloned profiles are identified by using feature similarity measures [13],[20]. Initially, we extract profile features of a user profile using Web Scraper tool[1]. Then similar profiles are searched in Facebook using the same Web Scraper tool[1]. Here a same site profile cloning method is implemented to identify cloned profiles in the same website. After searching for similar profiles in Facebook, a similarity index is calculated. Here Cosine similarity index [15] is used to find the similarity measures [11-13] Based on the similarity index [15], we verify for cloned profiles and genuine profiles manually. If the similarity index is greater than or equal to the threshold value [15], [16] then the profile can be considered as a cloned profile. If the similarity index is less than or equal to the threshold value [15], [16] it can be genuine profile. Also, a method to identify fake profiles [14],[23],[24] can be implemented through which the adversary's location can be tracked through the IP[17] address tracking method.



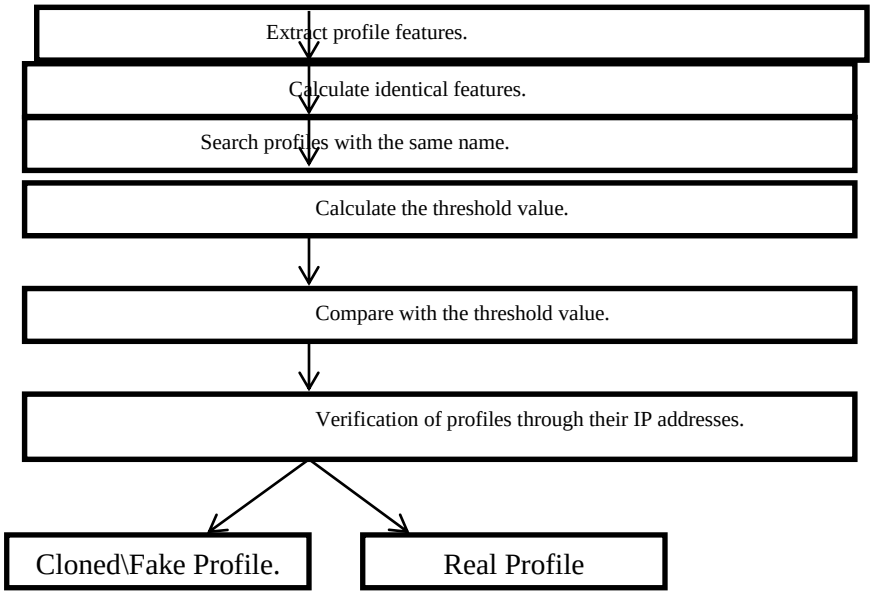


Figure 1. Architecture of same site-profile cloning identification method

Algorithm 1:Similarity Checking Algorithm

Input: Profiles P1 and P2 of two users

P1(A).Ai and P2(A).Ai are the text values of a feature A in P1 and P2

Output: Result :Similar / Not SimilarBegin

For each Ai in (P1 P2) do

CI=CosineSimilarity Index

T=Threshold

If CI ≥ T then Result=Similar Else

Result= Not similar.End

Return ResultEnd.

The cloned and fake profile detection system can be divided into 4 steps:-

1. Feature extraction Phase

2. Feature matching Phase

3. Identical Feature Measurement Phase.

4. Find the threshold.

5. Verification of profiles through their IP address.

Step 1- Feature extraction phase

In this phase, a profile is searched in Facebook with a particular username and features [20] included in the About section of that user profile like Places Lived, Work and Education, Contact info etc. is extracted from Facebook using Web Scraper tool[1]. They are imported to Microsoft Excel database [22].

Step 2- Feature matching phase

In this phase, the profiles with the same username is searched in Facebook and the features[20] included in the About section of those users is extracted from Facebook using the same Web Scraper Tool[1]. These features[20] are again imported to Microsoft Excel. Then the features[20] imported in Microsoft Excel are compared for similarity[13] between these profiles in Facebook.

Step 3-Identical feature measurement phase

In this phase, all the features[20] included in the About section of the user like Work and Education, Places Lived, Contact info etc. are measured for similarity[13]. Profile features[20] with the same username are compared and a similarity index value is calculated. The similarity[13] is calculated based on Cosine Similarity [15]. The similarity takes the common values between the two social users. This considers the similarity [13] between their features[20].

Two social users are more similar, in case if the users have more general text values [4].

The Cosine similarity index [15](also termed the Cosine similarity coefficient) analyses members for two sets to identify which members are common and which are separate. It's a measure of similarity[13] for the two sets of data . The greater the proportion, the more similar the two populations [15].

The formula to determine the Index is:

$$\text{Cosine Index} = \cos \text{Ø} = \frac{A.B}{\|A\| \, \|B\|} \quad (1)$$

Cosine similarity[15] index, the steps are:-

1. Determine the vector for A. A is a particular profile feature[20] of a particular user.

2. Determine the vector for B.B is the same profile feature[20] of another user.

3. Find the product of A and B and divide it by the magnitude of vectors A and B .

Example:- "Thrissur" to "Thrissur,Palghat"

P1(A)="Thrissur"

P2(A)="Thrissur,Palghat"

Similarity Score=1.4

Step 4-Finding and comparing with the threshold value

After getting the similarity score among different users, it is compared with a threshold value. If the similarity score is greater than or equal to the threshold value, then the profiles are similar and there are chances that it could be cloned. If it is lesser than the threshold value, then the profiles are not similar and it may not be cloned.

The formula to find the threshold is :

$$TSH = \frac{\text{Cosine Index}(P1(A),P2(A)) + \text{Cosine Index}(P1(A),P3(A)) + \ldots \text{Cosine Index}(P1(A),Pn(A))}{\text{Total Similarity Index Value}} \quad (2)$$

P1(A) is the attribute A for Profile 1.

P2(A) is the attribute A for Profile 2.

Pn(A) is the attribute A for Profile n.

If Cosine Index>=TSH, then the profiles are similar, else not similar [16].Finally the profiles can be verified for accuracy of results.

Step 5-Verification of profiles through their IP addresses

For verifying the profiles, IP [17],[8],[19]address of the user can be detected. Fake Profile users will not reveal their original location or they will not share their location provided in Facebook Share Location Services. So their location details can be verified by detecting their IP address[19]. If the profile is fake, all their location details provided will be proven incorrect and such profiles can be considered as suspicious to be fake profiles[23],[24].

3. Results and Discussion

3.1 Data extraction from Facebook account

To retrieve data from Facebook, a web scarper tool can be used. It is a primary way to get data from any social network like Facebook, Twitter, Instagram etc. The web scraper tool[1] can be linked to Microsoft Excel, Google Drive, Google Spreadsheet, or any other database tools. We can import the extracted data to Microsoft Excel Spreadsheet. Different features[20] from Facebook included in the About Section like Name, Work and Education, Places Lived etc. can be extracted from Facebook using the Web Scraper tool[1]. The data received through Web Scraper tool[1] are imported to Microsoft Excel Worksheet. Then based on the similarity measurement we can identify the profile as cloned profile or genuine profile [15].

Name	Work and Education	Places Lived	Contact and Basic info	Family and relationships	Life Events
Vineetha Venugopal	Studied at Noorul Islam University	IJK	vineetha72017@gmail.com		Graduated from Noorul Islam University - Noorul Islam Centre for Higher Education
Vineetha Venugopal	Studied at govt medical college Trivandrum			Smrithi Daughter,Sruthi Daughter,Saraladevi Mother,Damodaran Father	

Figure 2 Experimentation using Web Scraper Tool

Using the above described approaches we may get data from Facebook. By utilizing keyword search engine, we may search profiles that are comparable to user profile with the same

name. Using feature[20]similarity measures[11] we may examine if the profile is cloned or not and the verification is done using IP [17],[18],[19]address detection method for the validation of findings.

3.2 Result Analysis

To determine the work's efficacy, a total of two hundred users or participants were analyzed. Each of the selected attributes is utilized to generate the similarity index between the two profiles in question. The Cosine similarity index [15] technique is used to compute the similarity index since it is a very simple and efficient approach. If the similarity index is more than the threshold value **[**16], the profile might be termed cloned. Figure 3 depicts the graph derived from Facebook's same-site profile copying of ten profiles. The analysis is performed on two hundred subjects, and their profiles are compared to those of users with the same username. A similarity index greater than threshold value [16], indicates that the profile may be cloned. Verification of profiles is done through detecting the IP [17] ,[18], [19]addresses of the users. A cloned profile is anyway a fake profile. So through the detected IP [17], [18], [19]addresses from the list of cloned profiles, the profiles who pretend to be genuine can be considered suspicious to be fake profiles[23],[24]. Table II and Table III [25] shows the performance evaluation of results performed on two hundred subjects. Out of two hundred subjects, five profiles were detected as cloned profiles. And for verifying these five subjects, their IP[17],[18],[19] addresses were tracked to find their location. The IP[17],[18],[19] address detection method provided accuracy for detection of fake profiles[23],[24] from the list of cloned profiles detected through similarity[13] measurement score. But detection of IP[17],[18],[19] addresses has its own limitations too. Suppose the adversary is registered from a company located at a particular place, the IP[17],[18],[19] address detected will display results of that particular location only. Due to

this reason, they IP[8],[17] ,[19]address tracking method may not produce 100% accuracy results always. But fake profile users will never share their original location with the person whom he is contacting or the adversary will never use the Share Your Location facility provided through Facebook. In such a case, the profile user who is already considered as suspicious, can be detected through the IP[17],[18],[19] addressing detection method, which can be an efficient method for detecting fake profiles[23],[24].

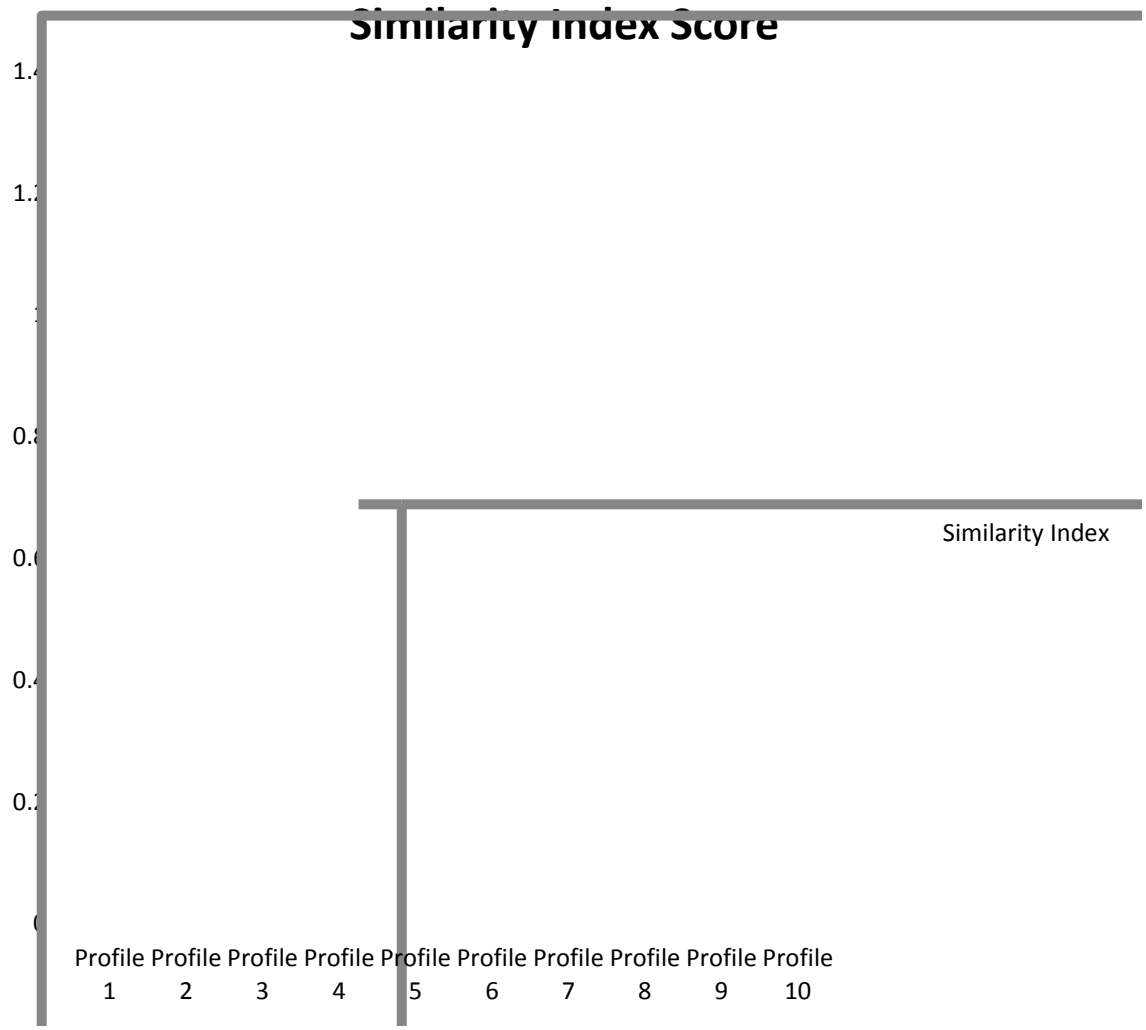


Figure3 Similarity Index Score

Table II Evaluation Of Performance After Detection Of Cloned Profiles

TOTAL NUMBER OF RECORDS CHECKED	**200**
Number of real profiles identified by the system asreal(TN) Number of normal profiles identified by the system ascloned (FN) Number of cloned profiles identified by the system as real (FP) Number of cloned profiles identified by system ascloned (TP)	192zz 3 0 5

Table III Evaluation Of Performance After Detection Of Fake Profiles

TOTAL NUMBER OF RECORDS CHECKED	5
Number of real profiles detected by the system as real(TN) Number of real profiles detected by the system as fake(FN) Number of cloned profiles detected by the system as real(FP) Number of cloned profiles detected by the system as fake(TP)	3 0 0 2

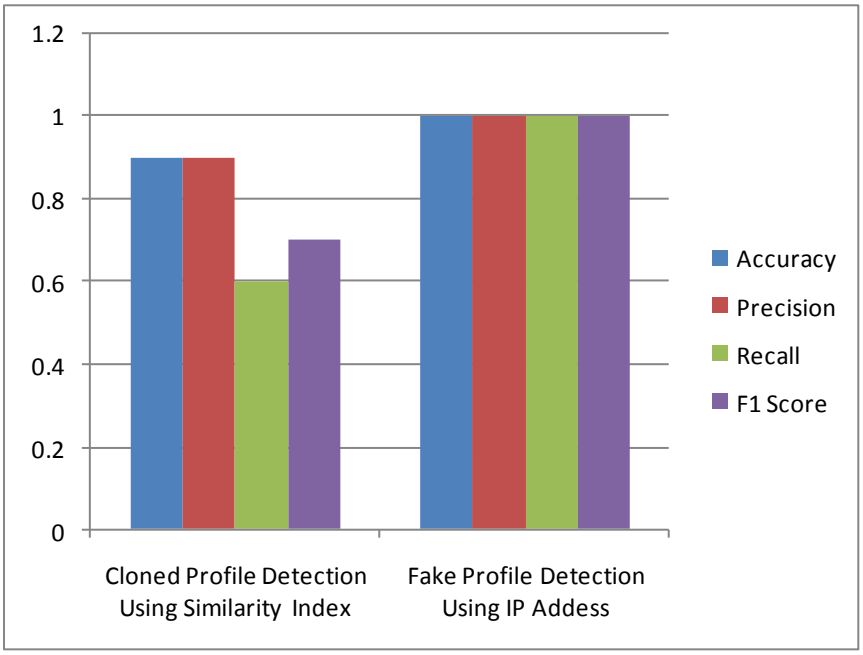


Figure 3 Performance Evaluation Results

Results of Table II and Table III [25]shows that 5 out of 200 profiles were cloned. Cloned profiles were detected using similarity measures [11],[13]using Cosine index Similarity[15] and comparing them a threshold value [16]. The cloned profiles were verified using IP[17],[18],[19] address detection method. Out of 5 cloned profiles, 2 profiles were found to be fake since their location mentioned, and their location tracked through IP [17],[18],[19]address were different.

Since this research has been implemented with a new methodology, it has been difficult to find research references and adjust it with the above experiment. So, this research is unable to compare the performance with other research references.

4. Conclusion

Due to the amplified tradition of using online social networks, the safekeeping traits of these users have to be taken care of. There are malevolent users who use these social networks just to spread spam activities. They collect the particulars of honest users and use them for performing their deceitful activities. Profile cloning is one such activity where details of users are collected and fake profiles are created using those details. Their intention is to attract innocent victims and perform destructive activities. In this work, an attempt has been made to detect cloned and fake profiles. Due to the huge volume of data and profiles used in social networks, the manual detection method using IP address becomes a workload and cannot be implemented efficiently for large volume of data. Also, detection using the IP address has its own limitations. The web scraper tool used in this method for feature extraction of data is combined with similarity measurement techniques using Cosine Similarity makes detection of cloned profiles an easy and efficient mechanism. We would also like to apply Big Data technique in future considering the hug

5. References:

[1] SP Maniraj, G Harie Krishnan, T Surya, R Pranav, "Fake Account Detection using Machine Learning and Data Science," International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 1, pp. 583–585, Nov. 2019, doi: 10.35940/ijitee.a4437.119119.

[2] S. R. Sahoo and B. B. Gupta, "Fake profile detection in multimedia big data on online social networks," International Journal of Information and Computer Security, vol. 12, no. 2/3, p. 303, 2020, doi: 10.1504/ijics.2020.105181.

[3] Suriakala, M., Revathi, S., "Privacy Protected System for Vulnerable Users and Cloning Profile Detection Using Data Mining Approaches," Tenth International Conference on

Advanced Computing,2018,pp. 124–132, doi:10.1109/icoac44903.2018.8939100

[4] S. Garg, K. Kaur, N. Kumar, and J. J. Rodrigues, "Hybrid deep-learning-based anomaly detection scheme for suspicious flow detection in SDN: A Social Multimedia Perspective," IEEE Transactions on Multimedia, vol. 21, no. 3, pp. 566–578, 2019, doi:10.1109/tmm.2019.2893549d.

[5] Punkamol and R. Marukatat, "Detection of account cloning in online social networks," in 8th International Electrical Engineering Congress (iEECON), pp. 1-4,2020, doi: 10.1109/ieecon48109.2020.229558.

[6] M. K. S. Paul, D. S. Saheb, K. V. Kumar, S. Vinod, and U. G. Student234, "Detection of fake and clone accounts in twitter using classification and distance measure algorithms," Journal of Resource Management and Technology.

[7] Conti, M., Poovendran, R., & Secchiero, M.,"Fakebook: Detecting fake profiles in on-line social networks," In IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,2012,pp. 1071-1078, doi: 10.1109/asonam.2012.185

[8] M. Zare, S. H. Khasteh, and S. Ghafouri, "Automatic ICA detection in online social networks with PageRank," Peer Peer Netw. Appl., vol. 13, no. 5, pp. 1297–1311, 2020, doi:10.1007/s12083-020-00894-6

[9] Kontaxis, G., Polakis, I., Ioannidis, S., & Markatos, E. P. ,"Detecting social network profile cloning," in IEEE international conference on pervasive computing and communications workshops (PERCOM Workshops),2011,pp. 295-300, doi: 10.1109/percomw.2011.5766886

[10] Anand, K., Kumar, J., & Anand, K.,"Anomaly detection in online social network: A survey", in International Conference on Inventive Communication and Computational Technologies (ICICCT),March 2017,pp. 456-459), doi: 10.1109/icicct.2017.7975239.

[11] L. Jin, H. Takabi, and J. B. D. Joshi, "Towards active detection of identity clone attacks on

online social networks," in Proceedings of the first ACM conference on Data and application security and privacy - CODASPY '11, 2011,doi:10.1145/1943513.1943520.

[12] S. Dogra, "Over 27 million Indian adults experienced identity theft in the past 12 months, says Norton report," India Today. www.indiatoday.in/technology/news/story/over-27-million-indian-adults-experienced-identity-theft-in-the-past-12-months-says-norton-report-1792553-2021-04-19 (Accessed: 17-Sep-2022).

[13] Kumar, N., & Dabas, P.,"Detection and Prevention of Profile Cloning in Online Social Networks," in 5th International Conference on Signal Processing,Computing and Control (ISPCC),2019,pp. 287-291,doi: 10.1109/ispcc48220.2019.8988394

[14] U. Can and B. Alatas, "A new direction in social network analysis: Online social network analysis problems and applications," Physica A, vol. 535, no. 122372, p. 122372, 2019, doi:10.1016/j.physa.2019.122372.

[15] Sowmya and M. Chatterjee, "Detection of fake and cloned profiles in online social networks," SSRN Electron. J., 2019, doi:10.2139/ssrn.3349673.

[16] V. A. Dabeeru, "User profile relationships using string similarity metrics in social networks," arXiv [cs.SI], 2014,doi:10.48550/arXiv.1408.3154.

[17] C. Xiao, D. M. Freeman, and T. Hwa, "Detecting clusters of fake accounts in online social networks," in Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security, 2015. pp. 91-101,doi:10.1145/2808769.2808779

[18] Q. Gong et al., "DeepScan: Exploiting deep learning for malicious account detection in location-based social networks," IEEE Commun. Mag., vol. 56, no. 11, pp. 21–27, 2018, doi: 10.1109/mcom.2018.1700575.

[19] J. Lloyd, "How to Trace an IP Address," wikiHow. https://www.wikihow.com/Trace-an-IP-Address (Accessed: 18-Sep-2022).

[20] S. Jia, X. Zhang, X. Wang, and Y. Liu, "Fake reviews detection based on LDA," in 2018 4th International Conference on Information Management (ICIM), 2018,doi: 10.1109/infoman.2018.8392850.

[21] N. A. Patel and R. Patel, "A survey on fake review detection using machine learning techniques," in 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-6,doi:10.1109/CCAA.2018.8777594

[22] N. Shoeibi, N. Shoeibi, P. Chamoso, Z. Alizadehsani, and J. M. Corchado, "A hybrid model for the measurement of the similarity between Twitter profiles," Sustainability, vol. 14, no. 9, p. 4909, 2022, doi:10.3390/su14094909.

[23] D. Ramalingam and V. Chinnaiah, "Fake profile detection techniques in large-scale online social networks: A comprehensive review," Comput. Electr. Eng., vol. 65, pp. 165–177, 2018, doi:10.1016/j.compeleceng.2017.05.020

[24] Romanov, A. Semenov, O. Mazhelis, and J. Veijalainen, "Detection of fake profiles in social media-Literature review," in International Conference on Web Information Systems and Technologies, vol. 2, SCITEPRESS, 2017, pp. 363–369.

[25] Sowmya, P., & Chatterjee, M.,"Detection of fake and clone accounts in twitter using classification and distance measure algorithms," in 2020 International Conference on Communication and Signal Processing (ICCSP),July 2020, pp. 0067-0070, doi: 10.1109/iccsp48568.2020.9182353.