Social fingerprinting: detecting spambot meetings by presenting DNA-enlivened behavior

Dharani R, PANIMALAR INSTITUTE OF TECHNOLOGY, Chennai

Aadithya R, PANIMALAR INSTITUTE OF TECHNOLOGY, Chennai

Bewin Marshel M, PANIMALAR INSTITUTE OF TECHNOLOGY, Chennai

Pugazhandhi S PANIMALAR INSTITUTE OF TECHNOLOGY, Chennai

***Abstract*: Online social network spambot detection is a valid test that takes into research and a discovery strategy capable of successfully identifying persistently improving spammers. A new influx of social spambots has recently emerged, spurred by human-like characteristics that enable them to evade detection even by the best-in-class algorithms. We show that effective spambot identification is feasible by conducting a thorough examination of their group's behavior and utilizing a sophisticated DNA method to reveal the actions of unauthorized community members. The conduct lifetime of an advanced record is contained in a collection of characters in the computerized DNA representation, brought to life by its organic counterpart. Then, we develop a similarity metric for such sophisticated DNA groups. We use client comparative studies and computerized DNA to depict both legitimate accounts and spambots.. We construct the Social Fingerprinting system using this representation, which, in both controlled and unsupervised versions, could tell spambots from real accounts. Finally, we assess the merits of social fingerprinting and contrast it with three state-of-the-art location estimation methods.. One of our method's peculiarities is the viability of using off-the-shelf DNA inspection methodologies to look into how people use the internet and to effectively rely on a small number of weak record qualities.**

Keywords: Spambot, online interpersonal organizations, Detection methods, Social spambots, Digital DNA, Supervised and Unsupervised design, Authenticity.

I. INTRODUCTION

Spambots have become a pervasive problem in online social networks, and traditional detection[3] methods are becoming increasingly ineffective against them. In this study, we suggest a novel strategy.to detecting spambot groups by utilizing digital DNA to model the behavior lifetime of online social network users. Our approach, called Social Fingerprinting, is inspired by the DNA sequencing technology used in genetics research. The digital DNA sequence represents the behavior lifetime of an online account, and we use a similarity measure to differentiate between genuine and spambot accounts.

By grouping accounts based on their digital DNA sequence, we can identify spambot groups and differentiate them from genuine accounts. We present both supervised and unsupervised versions of the Social Fingerprinting framework and evaluate its effectiveness against three cutting-edge spambot recognition methods. Our approach offers the potential for using off-the

behavior of online users and relies on a collection of simple account variables for effective spambot group identification with off DNA analysis techniques. Our results show that Social Fingerprinting is a promising approach to detecting spambots in online social networks.

II. RELATED WORKS

A novel architecture named FinEvent[2] is proposed from streaming social messages, for identifying hot social events, addressing the challenge of low accuracy and generalization caused by ambiguous features, dispersive content, and multiple languages. FinEvent models social messages into applications for weighted multi-relational graphs Reward-based Multi-Agent Learning and Contrastive Learning for incremental representation learning. A clustering model that is guided by deep reinforcement learning is also designed for social event detection. Experiments on Twitter[1][15] streams show significant improvement in performance, with increases ranging from 14-118% for offline, 8-170% for online, and 2-21% for cross-lingual tasks.

Spammers on social networks like Twitter and Sina Weibo spread illegal information to normal users. Current approaches use content and social following info to detect spammers[16] but these methods have limitations such as fake social following info and sparse feature modeling. To overcome we suggest a fresh approach dubbed CNMFSD to address these problems. that uses content and users' interaction relationships.

The proposed method has been tested on a real-world Twitter dataset and results show improved detection performance compared to existing methods.

Twitter bot detection is crucial, and current text-based techniques work well. New bots, however, get around them by faking dangerous material and exploiting stolen tweets. Semantic inconsistency characterizes these bots. Modern techniques using Twitter graph layout are competitive but a method fusing both text and graph modalities is lacking. Our paper proposes a novel model (BIC) that interacts and detects semantic inconsistency between text and graph modalities. Our framework outperforms competitive baselines and the interaction and semantic consistency [17]detection is proven effective.

Twitter is a well-liked microblogging website. with over 206 million daily active users. As its popularity grows, so does the number of Twitter bots. To combat misinformation, a unique entropy-based structure was proposed to detect associated bots based on user actions. User behavior is modeled as DNA sequences and the relative entropy between these sequences is used to determine the probability of an account being a bot. The proposed technique outperforms existing methods with a correctness of 0.9471, remember of 0.9682, F1 rating of 0.9511, and accuracy of 0.9457.

Efficient identification techniques are needed due to the existence of bots on Twitter. Existing methods face challenges with the large dataset size, constantly evolving bot accounts, and complexity of learning representations in Twitter attributed networks. This paper introduces ADNET, a novel framework that uses a topology-based approach to active learning with a deep auto - encoder to detect anomalies in Twitter attributed networks with less labeled data. Results show that ADNET outperforms previous methods and reduces annotation cost in detecting anomalous bot accounts in Twitter networks.

III. ESTABLISED TECHNIQUES

OSNs have become the perfect environment due to their widespread availability and ease of use, for the propagation of fraudulent and malicious accounts. However, there are dishonest scenarios in which social media accounts are created and maintained to disseminate unsolicited spam, promote illegal events and goods, sponsor public personalities[5], and eventually influence public opinion. While concealing one's true Identity is occasionally motivated by certain kind aspect of one's character. Social spambots are unique in that they change over time and employ complex strategies to circumvent early detection methods that are already well-established, such those that rely on the text of shared messages. even though their accuracy is lower, certain predictive models outperform others in certain circumstances.

To detect spambots in social media and online, several technologies are frequently used. Among such techniques are:

1. **Content-based methods:** To recognise spambots, these techniques analyse the content of the messages or posts. They utilise natural language processing (NLP) algorithms to identify dialect and patterns of behaviour that are indicative of spam.
2. **Network-based methods:** To identify bots, these techniques examine the aspects of the social system. They are looking for patterns of connectivity and behaviour patterns that imply spam.
3. **Machine learning-based methods:** Machine learning algorithms are used in these techniques . For example, patterns of behaviour that really are indicative of spambots. They may train models to classify accounts as spambots or legitimate users by combining content-based and network-based attributes, as well as other data such as account creation timeframes as well as IP addresses.
4. **Captcha-based methods:** To make a distinction between human users and spambots, these techniques employ a review system.
5. **Human-based methods:** To identify spambots, these techniques depend on individual moderators or crowdsourcing.

It is worth noting that spambots are constantly evolving, and detecting them can be a cat-and-mouse game between spammers and detection algorithms. Therefore, it's important to use a combination of these techniques and to constantly update and

improve detection methods to stay ahead of spammers.

DRAWBACKS

Decreased trust: The widespread presence of fictitious and malicious accounts can reduce the trust that users have in online social networks (OSNs), leading to a decrease in engagement and usage of these platforms.

Increased spam and unwanted content: Fictitious and malicious accounts can distribute spam and unwanted content, which can be annoying and time-consuming for users to deal with.

Biased public opinion: The proliferation of fake accounts can be used to manipulate public opinion and influence decision-making, which can have serious consequences for society.

Evolving techniques to evade detection: Fictitious accounts are often created and managed by individuals or groups with malicious intentions, who are constantly evolving their tactics to evade detection. This can make it difficult for OSNs to effectively address the problem.

Inaccuracy of predictive models: Even with the use of predictive models to detect fake accounts, some models may perform better than others, leading to a lower accuracy in identifying fictitious and malicious accounts. This can create a false sense of security and leave users vulnerable to deception and manipulation.

IV. DIGITAL DNA IN SOCIAL FINGER PRINTING

Digital DNA and social fingerprinting can also be used in spambot detection. Spambots are automated programs[18] that are designed to send large volumes of unsolicited messages, often for the purposes of advertising or spreading malware. One way to detect spambots is by analyzing the digital DNA and social fingerprint of the messages they send. Spambots often use similar patterns of language and behavior that can be identified through natural language processing (NLP) techniques. For example, they may use similar keywords or sentence structures, or they may send messages at similar times or intervals. Similarly, spambots often have a distinct social fingerprint that can be used to identify them. They may follow many users without being followed back, or they may have a high ratio of outgoing messages to incoming messages. They may also have many connections to other accounts that are known to be associated with spambots. By analyzing the digital DNA and social fingerprint of messages, it is possible to identify patterns that are consistent with spambot behavior and to flag these messages for further review or filtering. This can help to reduce the spread of spam and other unwanted messages on social media[8] platforms and other online services. However, it is important to note that spambots are constantly evolving, and new techniques may be needed to keep up with their changing behavior. As such, ongoing research and development is needed to stay ahead of the latest spambot threats.

V. HAM ALGORITHM

HAM (Hidden Action Markov) algorithm is a machine learning algorithm used for social fingerprinting in spambot detection. The HAM algorithm is based on the idea that spambots exhibit different patterns of behaviour than human users, and these patterns can be detected by analysing the sequence of actions they take on social media platforms. The algorithm works by first defining a set of "states" that correspond to different actions that a user can take on a social media platform, such as sending a message, following a user, or posting a comment. These states are then combined into a Hidden Markov Model (HMM), which is a probabilistic model that can be used to predict the likelihood of a particular sequence of actions given a set of observed data. To train the HAM algorithm, a dataset of known spambot and human user behaviour is used to estimate the probabilities of transitioning between different states. The algorithm then uses these probabilities to classify new users as either human or spambot based on their observed behaviour. One of the strengths of the HAM algorithm is its ability to adapt to changes in spambot behaviour over time. As spambots evolve and adopt new strategies, the algorithm can be retrained using new data to identify these new patterns of behaviour and update its classification rules accordingly. However, like all machine learning algorithms, the HAM algorithm is not perfect and can sometimes make mistakes in its classifications. Therefore, it is important to use the algorithm as

part of a larger suite of tools and techniques for spambot detection, and to continue to refine and improve the algorithm over time.

VI. CONTRAST OF THE NAIVE BAYES AND THE HAM ALGORTIHM

In machine learning, two approaches for classification are used: Naive Bayes as well as the Hamming algorithm. The Hamming algorithm is a way away approach that measures the similarity between the two vectors, whilst Naïve Bayes classifier is a probabilistic approach that relies on the Bayes theorem.

Naive Bayes is a simple and quick algorithm that accomplishes well with high-dimensional data. It computes the likelihood of each feature target class label and uses these odds to make predictions [4]. In some instances, Naive Bayes presumes that the characteristics are self - reliant of one another, which can be a limitation. Despite this assumption, Naive Bayes can perform well in a variety of classifier, along with spam filtering and text classification.

The Hamming method, on the other hand, is a distance-based way to determine the similarity between two vectors. It calculates the number of roles where the corresponding elements of two images differ and then uses the above value determines a vectors' similarity. The Hamming algorithm is frequently used in error-correcting codes, compression, and cryptography.

In contrast to Naive Bayes, the Hamming algorithm can handle non-independent features and works very well sparse data. However, it requires the features to be binary or categorical, which can be a restriction in some cases.

In overview, Naive Bayes and the Hamming algorithm are two different approaches to classification throughout machine learning. Naive Bayes is a simple and fast probability methodology that assumes feature independence, whereas the Pepping algorithm is a distance-based approach which can handle non-independent features but is[3] computationally expensive for large datasets. The algorithm chosen is determined by the specific requirements of the research at hand including the characteristics of the data being analysed.

VII. LCS: A DIGITAL SEQUENCE OF DNA RESEMBLANCE METRIC

A data format that represents a digital DNA sequence can precisely capture the behaviors of an only one OSN user. The analysis of focused in behalf of groupings over solo individuals, it may be helpful to manage and research diverse using digitized DNA sequences to describe the characteristics of the group. In this scenario, We examine group actions.by comparing the user's digital DNA to those other users in the same category. The M users' digital DNA sequences can be utilized to describe users of M = |A| in class A, specifically:

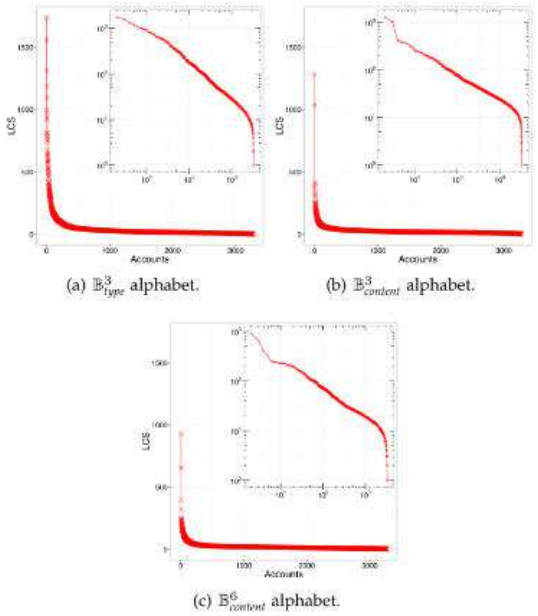
$$A = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{pmatrix} = \begin{pmatrix} (b_{1,1}, b_{1,2}, \ldots, b_{1,n}) \\ (b_{2,1}, b_{2,2}, \ldots, b_{2,m}) \\ \vdots \\ (b_{M,1}, b_{M,2}, \ldots, b_{M,p}) \end{pmatrix}$$

The aforementioned group A consists of M digital DNA sequences of varied lengths, one sequence for each group user, and is represented as a column vector.

In recent years, a wide range of approaches and algorithms for studying biology DNA or, more generally, threads have been developed. These techniques are mostly drawn from the fields of string mining and bioinformatics.

As a result, using a depiction of behavioral data using DNA strings allows you to take advantage of recent advances in these fields. Additionally, decades of development and research have led to effective and scalable algorithms that are quite suitable for organizing and evaluating OSNs data, which is massive and constantly developing by definition.

The k-common substring problem is the name of the subject. Finding the LCS that is shared by at least c of these strings for every string in this instance is the goal .2 c M, provided a vector A = (s1,..., sM) of M strings. Notably, it’s plausible to solve both the longest communal substring problem and the k-common substring by using the generic suffix to solve a problem in linear time and space tree and cutting-edge algorithms like the names featured in.



(a) $\mathbb{B}^3_{type}$ alphabet. (b) $\mathbb{B}^3_{content}$ alphabet.

(c) $\mathbb{B}^6_{content}$ alphabet.

Utilizing LCS c-common substring approach issues for each 2 c M, it is possible to generate an LCS curve It describes the correlation between the LCS's length and the quantity c of strings. The number of c accounts is depicted on the x axis.(corresponding to the digital DNA sequences, as well as k strings, used to determine LCS values), while the y axis shows the length of the LCS shared by at least k accounts. According to an LCS curve, each point is a fraction of the c accounts that shares the elongated substring (of length y) among all other possible subdivisions of the k accounts. The size of the LCS shared by all accounts decreases as the number of accounts c increases, which is a direct outcome of how the LCS is defined. As a result, LCS curves are nonincreasing monotonic functions:

$$LCS\,[c-1] \geq LCS[c]\ \forall\ 3 \leq c \leq M$$

This is also evident in the LCS curves shown in Figures (a), (b), and (c). As a result, a few accounts are more likely than large groups to have a long LCS.

VIII. ACCOUNT DNA CLASSIFICATION USING THE LCS CURVES

Account DNA classification using LCS bends serve transaction data from various accounts to determine their distinct behavioural traits. LCS (Longest Popular Subsequence) curves are a mathematical technique for comparing data sequence data and identifying differences and similarities.

To categorise accounts using LCS curves, you must collect transaction data for each consideration to be analysed. This information may include transaction date, transaction form, transaction amount, and account balances following the initial transaction.

Then, for each account, you would need to generate an LCS curve by comparing the transaction data to a reference curve. The reference curve could be created using historical data from similar accounts or by utilising an existing model of account behaviour.

After initiating the LCS curves, you can compare them to identify trends and anomalies [19]. Accounts with similar LCS curves are classified together, while accounting entries with distinct curves are classified individually.

This type of analysis can be helpful in detecting potential fraud or recognising accounts that need to be monitored more closely. It can also be used to identify account behaviour trends which can be used to inform business decisions or marketing strategies.

LCS graphs are an effective tool to analyse large amounts of transaction data and identifying patterns and anomalies. They can organization - wide make more informed decisions and enhance their job operations when mixed with the other techniques such as machine learning and deep visual analytics..

IX. METHODOLOGY RECOMMENDED

The recommended approach seeks to mimic online user behavior in order to detect social spambots. Digital DNA, which comprises of strings of characters that encode each activity done by the online account under inquiry, is used to model behaviors. DNA is a versatile model that may represent various activities across numerous social podiums and at different granularity levels. The approach has shown great performance in terms of standard classifier-based measures, proving its quality and viability. Twitter's spambot identification is a particular application on an unique social network, the suggested The Social Fingerprinting approach is platform and technology agnostic. As a result, the Social Fingerprinting technique can be applied to a variety of behavioral characterization tasks.

EFFECTIVENESS

The HAM algorithm has several advantages, including its ability to detect spambots by analysing the sequence of actions they take on social media platforms, its ability to adapt to changes in spambot behaviour over time, and its flexibility in modelling numerous activities made on different social media at various granularities. The proposed method of characterising online user behaviour using digital DNA and Social Fingerprinting can also provide various benefits in spambot identification and other behavioural characterization tasks. To begin, digital DNA is a versatile model that can represent numerous acts made on multiple social platforms, making it adaptable to a variety of use cases and scenarios. This means it can be used to detect spambots on many social networks, as well as for other behavioural characterization tasks like recognising bogus news or fraudulent actions. Second, the approach has shown great performance in terms of traditional classifier-based indicators, indicating that it is a useful tool for detecting spambots and other harmful activity. This implies it may be used to detect and remove spambots and other unwanted content from social media networks, hence improving the user experience for real users. Finally, because the Social Fingerprinting technique is platform and technology independent, it can be utilised for a wide range of behavioural characterization tasks other than spambot identification. As a result, it is a versatile tool for analysing and comprehending user behaviour across numerous online platforms and services.

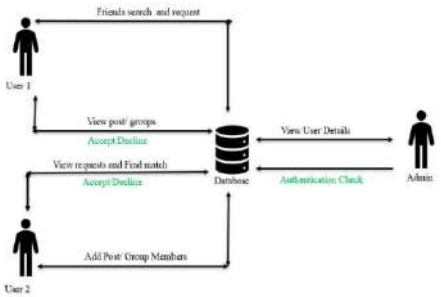


Figure 1: Architecture Diagram of proposed work.

X. DIGITAL GENOME

The genetic code contains the entirety of a person's genetic composition, which Nucleic acid (DNA) patterns represent. A DNA sequence is a sequence of characters that indicate the nucleotide sequence in a DNA molecule. The 4 potential characters are A, C, G, and T, representing the four nucleotide bases. Of a DNA Adenine, cytosine, guanine, and thymine are the building blocks of DNA. DNA from life forms contains the commands that control an organism's features and functions[7]. DNA sequences are being used on a global scale in biology, anthropology, forensics, and other scientific fields. Raw biological material can be used to read DNA sequences. DNA sequencing techniques. Two of the best known and extensively. Sequence alignment[12] and repetition/motif elicitation are two analytical methodologies used. The fundamental goal of these approaches is to identify trends and recurrence.in DNA sequences. In fact, a study of frequent sub-sequences and substrings can predict certain features of an individual, and correlations between different individuals can be established. We anticipate portraying OSNs users' interactions and behaviors using character strands that reflect the chronology of their actions, similar to biological DNA. Online behaviors Activities such as creating new content, responding to users, and following accounts can all be coded with different characters similar to how DNA sequences. The four DNA bases are represented by the letters A, C, G, and T. Well according to analogy, a user's behaviors act as the building blocks of His or her digital ancestry. On OSNs, there are numerous forms of user behavior; here, we take three threads based on time spent. They are as follows:

1. Addict
2. Prone to addict
3. Going to addict

XI. STEPS OF PROPOSED WORK

1. Signup & Login

In this, regular handlers who wish to intermingle with other users of the site must first complete the registration procedure by providing the site with the bare minimum of information, including a user name, password, mailing address, email address, and phone number. If a person wishes to access their account after registering, they must enter the right user name, email address, and password. If credentials are valid, the server will

permit access to the websites; otherwise, the server will generate a user name or password alert.

2. Update Details

The handler must wish to bring up-to-date their personal social portfolio after logging in because this is the essential step for all other system activities. On that screen, the user can add other details such as interests, education, the name of their college, and sons. After selecting the profile photo and clicking update profile, the server will be updated. with a profile key produced automatically. Sometimes individuals visit the profile update page to modify their profile photo. Choose a new profile picture on this screen, then click edit Profile. The server will then generate a new profile key and you can edit those data in the server.

3. TimeLine Add

In this module, the user posts some image content to express his or her feelings by way of sharing with peers. This entry will appear on his or her friend's list's timeline.

4. Friend Invite

In this section, the user types a portion of the text into the search bar before sending it as a request to the server. When a request of this nature is received, the server evaluates the likelihood of results before responding to the user. The only information in this reply is the names of the folks. A user must first choose some criteria before sending a friend request to anyone on this list.

5. Profile Congruence

When a handler makes a request for a friend, the server will execute this module. Server Obtain another handler's name and profile data from the database, as well as the profile details of the requested individuals. Profile matching occurs once the server compares both profiles with the provided five parameters using the profile matching algorithm. Finally, depending on five parameter matching, construct a single value. With this profile value, the user can read his or her friends' requests information. This handler might admit or deny the request based on this information.

6. View Secure Profile

Users in this module have a list of people's lists known as friends lists. However, these persons cannot read the profile details of others. If a user want to read the profile, he or she must first obtain the profile key from the profile owner and then view the profile information.

7. Collective Actions

Handlers can form groups in this module to share information with other users. If a group is created, the server will generate a group key. Only group activities are carried out using this key.

XII. SCREENSHOTS

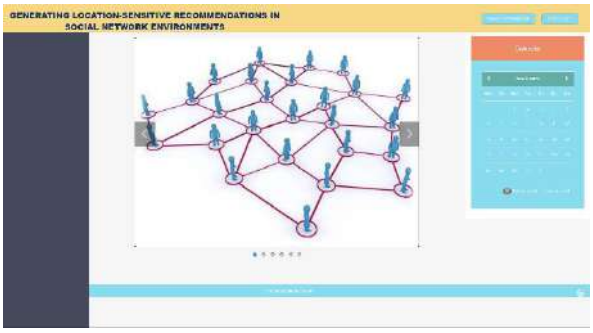


Figure 2: Home Page

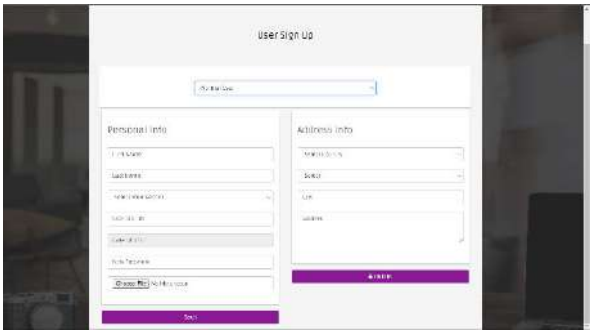


Figure 3: User Registration Page for Normal User

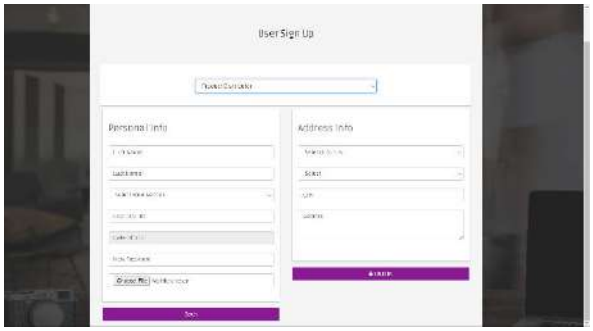


Figure 4: User Registration Page for Product Disturber

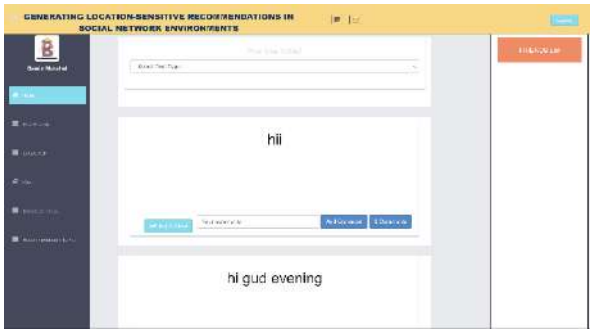


Figure 5: Social Feed of User Profile

Figure 6: Friend Request Page

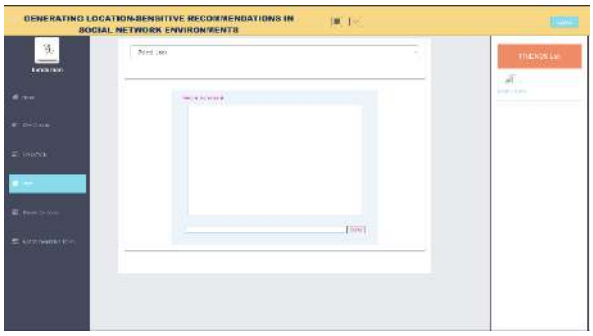


Figure 7: Chat page

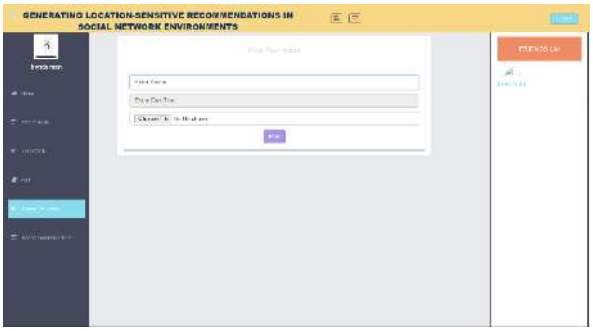


Figure 8: Event Page

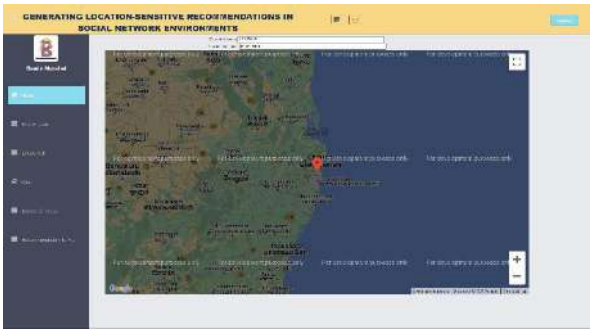


Figure 9: User Location Page

Figure 10: Admin Page

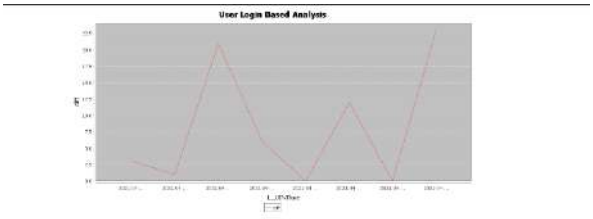
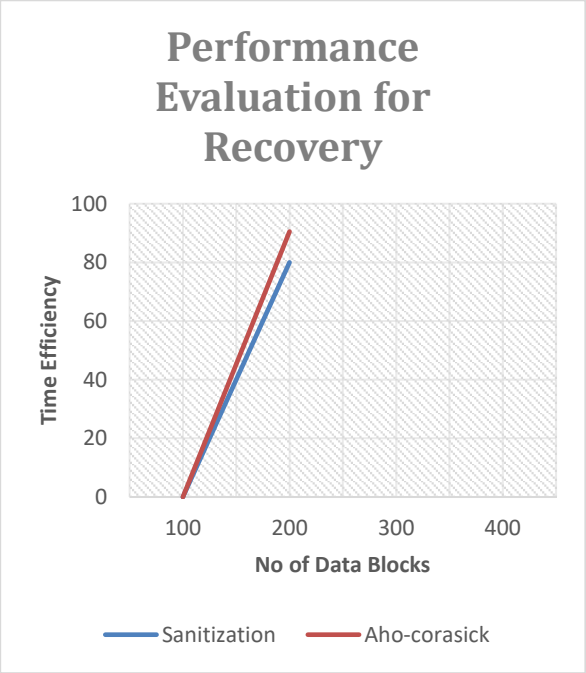


Figure 11: Generated Graph

XIII. RESULT AND DISSCUSSION

The performance is investigated by comparing the Sanitization Algorithm to our developed Aho-corasick algorithm. Assume that the total number

of blocks considered is 1000. The X axis of the graph compares the number of data blocks to the Y axis of time efficiency. The recovery time is 2.318 seconds. Thus, its time efficiency will be 89% if the formula Time Efficiency = (Time Required for Verification/Total Time of the Process) *100 is used. Because the recovery time in our proposed system is 1.99s, the time efficiency is calculated to be 90.5. Our system's performance has improved as time efficiency has been increased by examining the input text for the recovery process simultaneously.



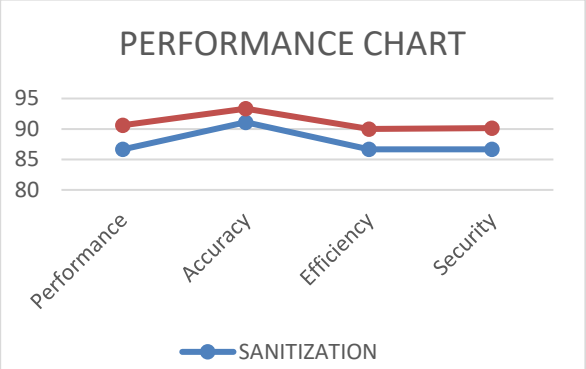


Figure 12: Performance Analysis and Performance Chart

XIV. CONCLUSION AND FUTURE SCOPE

We previously said that the constant influx of spambots was entirely designed to simulate the human behavior of OSN's legitimate clients. We also demonstrated that these novel types of spambots can circumvent cutting-edge computations designed to distinguish them. Following that, we proposed a more advanced DNA social showcasing strategy. We were able to test our working hypothesis using this method: there are still low-force signals that identify people from bots when studying clients on an aggregate rather than account-by-account basis. Our Social Fingerprinting identification technique and related algorithmic tool store generated from the domains of bioinformatics and string mining have showed exceptional discovery abilities for the majority of the most significant location metrics, exceeding cutting-edge arrangements.

We were able to achieve intriguing results when the features (topological features only) were considered. The findings show that the suggested method can be used to foresee crucial events that may occur in dynamic societies. While the model as described can produce intriguing findings, some aspects require more investigation in order to improve the method. For example, one shortcoming of our method is that we regard critical events equally. In some applications, however, various communities may have their own life cycles. Critical occurrences should be treated as such. One possible route is to modify the fluctuation computation by integrating a weighting mechanism in which the importance of events such as "appear" and "disappear" is not deemed equal but rather depends on community dynamics. Treating key events in accordance with the life cycles of communities is an important issue that, to our knowledge, has not yet been studied in the present literature. Further research in this area is required.

REFERENCES

[1] O. Ajao, J. Hong, and W. Liu. A survey of location inference techniques on twitter. Journal of Information Science, 1:1–10, 2015.

[2] E. Amig´ o, J. C. De Albornoz, I. Chugur, A. Corujo, J. Gonzalo, T. Mart´ın, E. Meij, M. De Rijke, and D. Spina. Overview of replab 2013: Evaluating online reputation monitoring systems. In Proceedings of CLEF, pages 333–352. Springer, 2013.

[3] F. Atefeh and W. Khreich. A survey of techniques for event detection in twitter. Computational Intelligence, 31(1):132–164, 2015.

[4] H. Bo, P. Cook, and T. Baldwin. Geolocation

prediction in social media data by finding location indicative words. In Proceedings of COLING, pages 1045–1062, 2012.
[5] J. Bollen, H. Mao, and A. Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In Proceedings of ICWSM, pages 450–453, 2011.
[6] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella. Discriminating gender on twitter. In Proceedings of EMNLP, pages 1301–1309, 2011.
[7] H.-w. Chang, D. Lee, M. Eltaher, and J. Lee. @ phillies tweeting from philly? predicting twitter user locations with spatial word usage. In Proceedings of ASONAM, pages 111–118, 2012.
[8] Y. Chen, J. Zhao, X. Hu, X. Zhang, Z. Li, and T.-S. Chua. From interest to function: Location estimation in social media. In Proceedings of AAAI, pages 180–186, 2013.
[9] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In Proceedings of CIKM, pages 759–768, 2010.
10] R. Compton, D. Jurgens, and D. Allen. Geotagging one hundred million twitter accounts with total variation minimization. In IEEE Big Data, pages 393–401, 2014.
[11] M. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonc¸alves, F. Menczer, and A. Flammini. Political polarization on twitter. In Proceedings of ICWSM, pages 89–96, 2011.
[12] M. D. Conover, B. Gonc¸alves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the political alignment of twitter users. In IEEE PASSAT/SocialCom, pages 192–199, 2011.
[13] D. Doran, S. Gokhale, and A. Dagnino. Accurate local estimation of geo-coordinates for social media posts. arXiv preprint arXiv:1410.4616, 2014.
[14] M. Dredze, M. Osborne, and P. Kambadur. Geolocation for twitter: Timing matters. In Proceedings of NAACL-HLT, pages 1064–1069, San Diego, California, 2016.
[15] Hao Peng, Ruitong Zhang, Shaoning Li, Yuwei Cao, Shirui Pan, Philip S. Yu, Fellow: Reinforced, Incremental and Cross-lingual Event Detection From Social Messages, IEEE JOURNAL OF IEEE 2022
[16] Hua Shen, Bangyu Wang, Xinyue Liu, And Xianchao Zhang Social Spammer Detection via Convex Nonnegative Matrix Factorization Journal Of IEEE (2022)
[17] Zhenyu Lei, Herun Wan, Wenqian Zhang,Shangbin Feng, Zilong Chen, Qinghua Zheng, Minnan Luo1 Association for the Advancement of Artificial Intelligence BIC: Twitter Bot Detection with Text-Graph Interaction and Semantic Consistency (2023)
[18]RosarioGilmary,AkilaVenkatesan,GovindasamyVaiyapuri, Deepikashini Balamurali Scientifc Reports | DNA influenced automated behavior detection on twitter through relative entropy (2022)
[19] Lulwah Alkulaib, Lei Zhang, Yanshen Sun, and Chang-Tien Lu Department of Computer Science, Virginia Tech, Falls Church, VA 22043 USA Department of Computer Science, Kuwait University, Kuwait {lalkulaib, zhanglei, yansh93, ctlu}@vt.edu Twitter Bot Identification: An Anomaly Detection Approach 2022 IEEE International Conference on Big Data (Big Data)
[20] Anisha P Rodrigues , Roshan Fernandes Aakash A, Abhishek B, Adarsh Shetty, Atul K,1 Kuruva Lakshmanna, and R. Mahammad Shafi Hindawi Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques Computational Intelligence and Neuroscience Volume 2022