PRIVACY CONCERNS IN SOCIAL NETWORKS

Renuka Devi¹, Swathi R², Bharath Narayanan³

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY, CHENNAI
johnpaulreddy875@gmail.com

Abstract—The current trend circulates around posting personal photos and other information on social network sites. Although this seems fun and is used by numerous youngsters, photos and information can be easily misused. In order to protect privacy and to have control over it, we introduce an algorithm. The algorithm first checks if the user has a valid original profile, then accesses information on the viewer and then displays the information which is related between the two. A higher level of privacy can be enforced by only displaying this common information between the two profiles.

Keywords- IDE – Integrated development environment, API-Application Programming Interface, COMPA – statistical tool

1.INTRODUCTION

Online social networking sites have been a boon and a bane. Even though it helps connect people of various age groups from all over the world, there is a tendency for people to misuse the photos or the information provided on social media. Numerous fake profiles prevail which sometimes make us wonder if its Facebook or 'Fake'book. Unwanted photo editing which is sexually abusive has caused a lot of problems leading to situations where the victims go till the verge of taking their own lives.

This scenario needs to be changed as people in social media have to be aware of the profiles they come across. They need to know if a profile is legitimate or not, or how they are related to the person before they accept the requests on Facebook. For this purpose, we introduce this application which will check for authenticity and relatedness of the profile.

This application, however, has age constraints. The algorithm that this application is built upon is only entitled to the age group 17-24. This is because parameters such as the number of photos, number of comments, maximum likes and comments the profile has, number of friends, face recognition technique is all considered to build this algorithm. These parameters are significant only in this age group. Depending on this algorithm, a significant percentage of authenticity is displayed and how the profile is related to the user is also displayed. This makes it easier for the user to identify if the profile is authentic and saves up the time of actually going to a person's profile and figuring out how they are related to each other. Using this, we believe one can be secured with the content they post on social media, as the authenticity level will give a medium of trust through which they can decide for themselves.

2.LITERATURE SURVEY

We surveyed the literature from various papers which included some of the algorithms and concepts that we wish to incorporate in our model.

Paper [1] talks about the observed privacy behaviours of the user. It talks about the 'only-friends' privacy scheme in Facebook and how it affects other users from viewing content on such profiles. 'Only-friends' mode proves to be the most privacy securing out of all the other sharing modes in Facebook. Paper[1] also talks about the association between network composition, expectancy violations, interpersonal privacy practices and gives a detailed view about enhancing an individual's privacy.
Paper[2] focuses on presenting an approach to detect compromised user accounts in social media. Often the behavioural pattern of few accounts changes and it is difficult to differentiate if it’s a legitimate user or a compromised account. Thus statistical approach is used to come up with a pragmatic solution for this problem, which led to the innovation of a tool, COMPA that implements their algorithm. It has been tested on a huge data set of 1.4 billion publicly available accounts.

Paper[3] also talks about finding out malicious profiles in social networks. For this purpose, they come up with a new algorithm where a classifier is constructed. With a help of these classifiers, they figure examples for positive and negative users. Then numerous parameters like the user degree, the user's connected communities number, the number of connection between a user's friends, the average number of friends inside connected communities are calculated. The classifier is later then constructed based on the parameters and evaluation takes place. The result is calculated in two different ways.

Paper[4] features the result of the study that was conducted on 4000 students at Carnegie Melon University. Paper[4] is based upon the information revelation in social media. The amount of information provided by a particular user and their respective privacy setting is analyzed. Based on the study paper[4] talks about how the user can attack a particular profile, the ways in which it can be approached and the effective way in which a user should use the privacy setting.

Paper[5] gives us a different approach to the leading fake profiles problem. Paper[5] states that if a person is not actually registered on Facebook, then the chances this person has a fake profile is investigated. A graphical approach is considered for the detection of fake profiles. Initially, an adversarial model is considered where the data sets are taken, then a graph is plotted for the evolution over time of the number of friends. Finally, a graph structure is plotted which includes the k-means clustering technique and the result is focused upon the average degree of the nodes and number of singleton friends. Based on this, an algorithm is built and fake profiles are analyzed.

Paper[6] sheds light on the privacy policy followed by college students and adolescents. Paper[6] includes studying the pattern of these age group students on the basis of them having either a public account or a private account. On concluding from the data set and related observations they come up with a conclusion that students are more likely to have a private profile if their friends and roommates have them, women are more concerned about their privacy, so they have more private accounts compared to that of men. The young adults have their unique set of cultural preferences in choosing between a private profile or a public profile.

Paper[7] highlights the emerging problem of fake profiles on Facebook. Paper[7] comes up with a solution by introducing a new software called 'Social Privacy Protector'. This software consists of three different protection layers for improving the privacy and security of a user's profile. The first layer identifies a user's friend who might have the potential to pose a threat then later restricts that particular friend to the user's personal information. The second layer is the expansion of privacy setting based on the various types of accounts. At the final layer, a lot of third-party applications get access to personal information through Facebook, the software then alerts the user about it.

Paper[8] talks about the advantages of social media like Facebook and then later comes up with privacy policies. It discusses the privacy settings like ‘only me’, ‘friends of friends’, ‘friends only’ and ‘public’. It also talks about the gender differences in revealing personal information like relationship status, phone number, email IDs, home address, hometown etc.

Paper[9] discusses the need in privacy for Facebook and the need to eradicate the fake profiles. For this purpose Paper[9] implements the software of Paper[7] and then uses machine learning algorithms such as various clustering techniques to build a classifier. This classifier helps in identifying if the particular user is a fake profile or not.

Paper[10] takes up how one single actor can create numerous malicious profiles. In order to identify this, they use a machine learning pipeline technique in order to classify an entire cluster of accounts as malicious or legitimate. They use parameters such as name, email id, company or university etc. It checks if all the emails share a common letter or a digit pattern and comparison of text frequencies across the entire user base.

3. PRESENT SYSTEM

The present system uses different techniques to detect fake profiles, but it does not provide a personalized level of authenticity or relatedness for each viewer. Unfortunately, it also does not provide
any check on the authenticity of friend requests and viewers of a personal account. It does not take into account mutual friends or relatedness in terms of a common workplace, education to display that information.

4. PROPOSED SYSTEM

We propose a system where not all users are given access to the same information but a rather viewer specific level of content. It takes into account the total number of friends, photos, maximum likes and comments each user has. It also has a unique algorithm that then suggests an authenticity level and displays the relatedness. It also checks the authenticity of accounts using facial recognition of photos to check if it is of the owner of the viewer account.

5. APPLICATION ARCHITECTURE

To get the authenticity and relatedness concept working, we have developed the application’s architecture diagram. Viewing this diagram will give a better insight into the working and implementation of the application.

It is divided into 4 modules. a) The first is the user interface which explains about the front end of the application.

b) The data extraction phase
c) The implementation of face recognition
d) Algorithm

![Fig 1- Architecture Diagram](image)

USER INTERFACE: We intend to showcase the working of this whole software through an application design. For this purpose, we use the sketch app, with which an appealing front end of the application is created. This app will increase accessibility along with the NetBeans Java program.
DATA EXTRACTION: Since Facebook does not disclose user information easily due to security issues, data extraction of users from Facebook has been a daunting task. For this purpose, we extracted data by three means:

- Initially, we manually collected data by visiting 40 profiles. We collected data such as the number of friends the user has, the number of photos, number of tagged photos, the maximum number of likes and the maximum number of comments the person has got.

A software called Facepager was used to extract the above parameters from public accounts on Facebook.

FACE RECOGNITION: We aim at detecting fake profiles in Facebook. For this purpose, a third party face recognition is used. Face recognition basically views the user’s pictures, checks the face which is repeated and stores it in its records. It then compares the record with the face that occurs more in tagged photos. Using the total number of tagged and user photos, the authenticity of the account are enhanced.

ALGORITHM: The algorithm basically checks two main things: -Authenticity, -Relatedness
The authenticity of the algorithm is checked by the number of friends, posts etc the user has in his account. If the user is legitimate, then the user should have some posts or should be tagged by their friends in other posts. This idea is incorporated into a code that acts on the SQL database. An appropriate percentage of authenticity is displayed in the output.

The relatedness is calculated by comparing the details of the user to the details of the other members of Facebook. Details such as the work place, check in places, places they are from, the school or college they attended are incorporated and compared to each other’s details. Then their relatedness is displayed.

Fig-5 ER diagram

6. TRADE OFF BETWEEN EXISTING AND CURRENT SYSTEM

The proposed system is more efficient than the existing system. Proposed system combines detecting fake profiles and displaying the relatedness before the two profiles. It comes up with a unique algorithm after studying various accounts in the age group of 17-24. The proposed system is also much efficient in fake profile detection by using concepts such as face-recognition that isn’t available in the existing system.

7. SCOPE

The scope of this application is limited. This application is only applicable to the profiles with age group 17-24. It is considered as such because the parameters that are considered so far can be significantly analyzed only with this particular age group. High school and college going student’s profiles were studied under a spotlight and their behavior was noticed. Depending on it, the age constraint was established.

8. HARDWARE AND SOFTWARE REQUIREMENTS

The hardware required will be a system with a processor of Pentium dual-core, hard disk of 120GB storage, a monitor of any size, input devices such as keyboard and mouse, and a ram of minimum 1GB

The software required will be operating system- windows 7, coding language - JAVA/J2EE, tools such as NetBeans IDE 7.2.1 and database- MySQL

Web Services – CSS and PHP.

9. CONCLUSION

In this paper, we present the application which aims to identify the fake profiles in Facebook using a set of parameters along with face recognition of user photos. to check the authenticity of the profile.
Then how the profile is related to the user is studied by checking the mutual places, education, work, places they have checked in.

The study presented in this paper is a work in progress with many available future directions. Using the existing idea, machine learning algorithms can later be infused to build a classifier that might be more efficient.

In the near future, we are going to work on considering a lot of other parameters to make the application work efficiently for all age groups.

10. REFERENCES

[1] Friends only: Examining a privacy-enhancing behavior in Facebook, Fred Stutzman and Jacob Kramer-Duffield.

[2] COMPA: Detecting Compromised Accounts on Social Networks, Manuel Egele, Gianluca Stringhini, Christopher Kruegel and Giovanni Vigna, University of California, Santa Barbara, CA.

[3] Strangers Intrusion Detection - Detecting Spammers and Fake Profiles in Social Networks Based on Topology Anomalies, Michael Fire, Gilad Katz, Yuval Elovici Telekom Innovation Laboratories and Information Systems Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

[4] Information Revelation and Privacy in Online Social Networks, Ralph Gross and Alessandro Acquisti, Data Privacy Laboratory School of Computer Science Carnegie Mellon University Pittsburgh, PA 15213

[5] FakeBook: Detecting Fake Profiles in On-line Social Networks, Mauro Conti, Radha Poovendran and Marco Secchiero, University of Padua, Italy.

[6] The taste for privacy: an analysis of college student privacy settings in an online social network. Kevin Lewis Jason, Kauffman Nicholas.

[7] Social Privacy Protector - Protecting Users’ Privacy in Social Networks Michael Fire, Dima Kagan, Aviad Elyashar, and Yuval Elovici Telekom Innovation Laboratories and Information Systems Engineering Department, Ben-Gurion University of the Negev, Beer-Sheva, Israel.

[8] Disclosure of personal and contact information by young people in social networking. An analysis of Facebook profiles as an example.

[9] Friend or Foe? Fake Profile Identification in Online Social Networks Michael Fire, Dima Kagan, Aviad Elyashar, and Yuval Elovici Telekom Innovation Laboratories at Ben-Gurion University of the Negev Department of Information Systems Engineering, Ben Gurion University

[10] Detecting Clusters of Fake Accounts in Online Social Networks Cao Xiao University of Washington and David Mandell Freeman Theodore Hwa