ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

# SCIENTIFIC WAY TO IDENTIFY DUPLICATE PROFILE IN SOCIAL NETWORKING WEBSITES

**PANKAJ**

**NET (COMPUTER SCIENCE & APPLICATIONS)**

**B.TECH.( IT)**

**MCA**

**MBA ( IT & MARKETING)**

**MAHARSHI DAYANAND UNIVERSITY, ROHTAK**

**HARYANA**

**INDIA**

**Abstract**

In present scenario an Individual user will have multiple social network accounts to stay involved with friends in several social networking sites. Online social network users aren't attentive to the varied security attacks like privacy violation, fraud, etc. Different on-line social users can assume it as real users and that they could be capable them that aren't truly the real user.It is estimated 1.96 billion user are active on social networking sites. Only Facebook have 1.87 billion active user in a month.. During this research paper, I had tried to analyze social network knowledge supported attributes similarity. The planned system will cite as several similar social network profiles as a potential and analyze them so as to seek out whether or not it belongs to same or totally different persons. It makes different user straightforward to speak with one another during a safe and secure manner.

**Keywords—Social Network Analysis, Social Engineering Attack; Duplicate profiles; world profile info, Profile attributes matching, Suspicious profiles**

## I Introduction

Social network makes our digital life become straightforward to own social relationship with friends on-line and social network websites like FacebookLinkedin,Twitter, Google+ for

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

connecting people, YouTube for video sharing, LinkedIn for skilled identity, speaker for change daily tweets (messages) of some event etc, are getting in style social network web site used among peoples of all ages particularly among youths. These social network sites square measure celebrated among net users and users square measure interconnected to every different via on-line social relationship called friendly relationship. Varied social network sites has been developed to realize their attraction among folks wherever any use will get will membership from the top of a straightforward registration type. A private user will build multiple accounts together with his same attributes like E mail id or mobile range in several social networks. [1] An existing user will have several numbers of social network profiles among identical or totally different network. It makes social network prone to attack by mistreatment somebody similar attributes. Most of the cyber crimes square measure happening in social network sites. [2] Any user will build profiles with other attributes such a same name, college, age, profile pictures, etc. faux profiles square measure being created altogether the social sites and victim personal info is turning into a lot of and a lot of prone to attack by the offender in varied ways in which. In the several cases Name are often same for several users, however the opposite attributes like profile image, qualification, address and mobile numbers all can't be same for multiple social network users [1] [2]. The most plan behind the planned approach is to seek out out as several social network users that have similar attributes and to seek out out the originality or real users from it. In a recent analysis reveals that just about eightieth of profiles in face book square measure faux one. Any user will build some faux accounts by mistreatment different attributes to fool other users. [3] Several users' prefer to disclose their personal info like phone no., date of birth, address etc in their profiles. Avails and revealing of such personal info could be the sources of profile knowledge that the offender is attempting to urge access to form similar profiles. The other user will produce faux profiles within the name of the actualuser thereupon personal info and check out to launch varied attacks like cussing and posting irreverent messages and tries to fool others to urge the direction. [4] The planned approach will cite as several similar social network profiles as potential supported the similarity of profile attributes and analyze them so as to seek out whether or not it belongs to same or totally different persons. The in public on the market profiles and their attributes square measure extracted is so stored them in world info in profiles in order that their existence are often checked in several social networks. It helps the others used to own on-line social communication with one another during a safe and economical manner. II.

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

**II Connected Works**

Many of cybercrimes square measures owing to the facilities to form unlimited numbers of profiles by identical person among identical, or totally different networks WHO try and act as real users and violating the principles. It makes different users tough to spot WHO square measure the $65000 users and WHO square measure fake. So social network analysis involves the situation that may be a new space of analysis that involves analyzing the network structure for the advantage of society in order that online users will have social affiliation and voice communication during a safe means with different user round the globe. In recent eventualities it becomes a heavy downside and lots of analyses have begun their research in distinguishing the $65000 user's identity.

**III. LIMITATIONS of the prevailing system**

**A. Protection of users' privacy limits knowledge collections**

Despite the fact that social media appears to be an awfully open house, each social networking site have their own privacy settings. E.g. Facebook, Linkedin, Twitter and its provide a facility to possess users their own privacy settings wherever they will show restricted range of knowledge and conceal their personal details while not confirming them. That's the real profile detail isn't perpetually doable. [3] B. Social network profiles square measure at risk of numerous attacks: The users within the social network square measure associated with every others within the kind of relationship. Social users square measure at risk of numerous attack like identity clone attack wherever the intruders produce some faux profiles with the name of some existing users, and he will produce the profiles with the precisely same as some existing user that appears almost like the important user and also the assailant would possibly attempt to launch numerous attacks. [7] C. Social networks square measure additional advanced to analysis: The users within the social network square measure associated with every others within the kind of social relationship and that they will be portrayed in graphical format and these networks square measure advanced to investigate. D. Finding the importance user's identity is tough to detect: A user will produce any range of profiles at intervals a similar or alternative social network with false identities to fool alternative users. Therefore the complete social network becomes additional advanced and confuses alternative users with multiple identities. [5]

**IV. Projected approach:**

Going by literature review, we've pointed out that the social network has advanced

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

structure and to investigate their relationship and patterns is advanced in nature. What is {more} the prevailing processes need more procedure time and involve several complexities. In Social network terribly giant amounts of private data square measure being shared and denote on-line daily. So AN anonymous user will retrieve the non-public details of people and pretend or false profiles that appear to be just like the real users. Currently a day's social users square measure additional prone to varied social engineering attacks like Identity Clone attack, faux profiles creation, hacking, etc. owing to their personal attributes square measure simply accessible. Hackers square measure invariably making an attempt to search out loopholes within the existing system. On-line social network permits its users to make infinite numbers of profiles to attach with social relationship with others. A user will produce several numbers of profiles inside constant or completely different|completely different} social network with different identities. The social insurance is additionally a significant issues related to this. A user will create as several profiles as he wishes. Another user will produce profiles with constant name of already existing users with the intent to fool others user and to urge the non-public data. It became troublesome to cite that one is real and that is faux one. The projected system to be developed is an application through that the user to possess the power to possess a Web based user profiles search mechanism. This facility may well be wanting to hunt for individual users in a very variety of social networks and manufacture a consolidated output in conjunction with an outline of with duplicated user profiles. Initial we tend to create our own social network profiles to research completely different attainable structures and attributes. Another application can be a "meta" social network website that encompasses a single surroundings for user through that they'll search and access different profile details. Users might connect their accounts with different social networks and therefore the Meta social network web site would consolidate all their data and friends' networks. It provides the user to possess a straightforward and effective thanks to communicate and to stay up-to date with their friends' activities across all the social networks from one surroundings. A social network could be a place wherever we tend to trust every and each user supported their on-line identities solely. However, most of the individuals out there don't seem to be real account, however the faux individuals with false identities World Health Organization try to try to to some malicious activity called a shaper. A shopper will represent himself as a true user to urge the non-public details of different users. Even he will chat on-line with different users, build trust among potential users by posting distinctive and original concepts with one

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

another, steal cash belongings or maybe life as a result of cyber criminals square measures targeting social networking sites to steal cash. The most purpose of the projected work is to produce a mechanism to resolve these problems with the assistance of victimization world information of profiles wherever the attributes of varied social network profiles square measure hold on. The projected tool will sight duplicate profiles exist in social domains, and conduct a case study. Initial is that the knowledge

Collection part wherever we tend to collect some in public accessible user profile knowledge sets that has got to be extracted and place those detail data into a information known as world information so as to search out the duplicate profiles. Then we tend to analyze and study the pattern of the in public accessible profile data to spot the $65000 user's identity from an outsized dataset.

## V. Project style approach

During this section we've printed the planning approach of the projected add a diagrammatical kind of distinctive duplicate profiles and checking the existence of comparable profiles on several social networks and substantiating the real profiles. The projected model includes of 3 elements and therefore the following section describes it one by one. Within the following diagram the projected work has been bestowed wherever the full workflow is split into three processes. • Profile identification method. • Profile analysis to envision the existence in Social networks • Profile verification method to envision whether or not the profile belongs to constant user of some completely different user profiles.

The main objective of the projected approach is to search out the actual user's identity across numerous social networks with the assistance of victimization world information for profiles wherever the attributes of varied social network profiles square measure hold on. AN algorithmic rule that detects similar profiles in numerous social networks and extracting their profile attributes from numerous suspicious profiles that helps to spot the actual users among them and realize any duplicate profile existence inside constant or completely different social networks by victimization their profile similarity among them. The projected profile similarity technique in social networks helps to spot a specific on-line user World Health Organization has multiple social networking accounts inside constant or {many completely different|many various|many alternative} social network sites and map and compare his/her profile's attribute values with different similar on-line user within the same or different social network to try to to on-line search easier and to enhance the web search by victimization the world

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

information of profiles. the first advantage of the projected approach: one. Since all the profiles details square measure hold on within the world information rather than storing them into completely different information of profiles. 2. Here the common profile attributes of users that square measure common in several social network square measure extracted and stores those attributes within the world information in order that the projected mechanism will simply realize the duplicate profiles and to form search method quicker and check the existence of same profiles inside constant or several social networks. 3. rather than checking duplicate profile existence in several social network separately, the world information of profiles is employed wherever all profiles details square measure hold on. 4. Memory consumption may also be reduced and therefore the web search may also be improved by utilizing social network / user generated content to enhance search. 5. to prevent and stop the speedily increasing faux profile creation round the world.

## VI. Projected style approach description:

There square measure increasing numbers of social network users day by day. This increase of social network users create the full network prone to attack since there's no restriction of making profiles and anyone will produce accounts with constant name of others profiles with the intent to fool others or to post some irreverent personal data of some already existing users while not the intention of the actualusers.

### 1. Profile identification process:

Step1. Assortment of profiles from several social networks: this can be the data gathering step from completely different social networks.

Step2. Distinctive similar attributes among them: to form a relationship between 2 or additional people across social networks.

Step3: Choosing solely the suspicious profiles: b. Evaluating profiles to validate the existence of user profiles in several Social Networks: Step one. Check and Matching similar attributes among the profiles (matching profile fields): Here the 2 profiles for his or her similarity supported their hypertext markup language structures of profiles square measure checked and place them in a very common information of profiles. It ends up in the subsequent 2 outcomes: actual matching: the primary class of matching analyze the user profiles by victimization some matching performs like string comparison to envision whether or not there exist any 2 knowledge fields that square measure specifically similar. Matching functions of this kind manufacture a Boolean result. E.g. the precise field

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

matching performs to match attributes like "usernames".

**2. Partial matching:**

The second class of matching functions analyzes the user profiles that match the elements of connected profile attributes. They're additional helpful in cases wherever the user profile knowledge has several redundant values like several abbreviations, misspellings or some missing values (e.g., address). Some perform can be wont to alter the matching of comparable knowledge values that square measure part associated with one another. AN example of such a perform is address or location matching. Step2: Extracting the common attributes supported similarity scores: Step3: check the existence of social sites of the profiles (e.g. face book, tweeter, YouTube etc.) c. Profile Verification to spot the actualsocial User: Whenever a user search and kind for any user profile, several profiles show up with the similar name out of that one are going to be the actualuser that the supposed user is checking out. however users cannot guarantee that whether or not profile we tend to square measure looking out is faux or real. this existing approach verifies those profiles data by manual method and checks for whether or not the name could be a documented person or it checks to match with different similar user. If that profile is a few famous person likes celebrities or any politician then it'll check whether or not the profile is connected with any official sites or govt. approved pages or if they're connected to any TV shows, Interview etc. If the profile looks to be uncertain then they'll be asking proof sort of a faxed ID. the most disadvantage with this current approach is that it's terribly manual, takes ton of your time to method it so overwhelming. Verification ways square measure additional typically wont to realize the first profiles and to evidence of the actualuser identities in social networks. It helps the opposite users to urge the real, real users and trustworthy data and find out high-quality sources of knowledge and to keep up trust that the legitimate sources of knowledge square measure real. The projected method verifies the users within the following ways:

Step1. Checking the buddies network (Mutual friends relationship)

Step2. Maintaining trust or distrust among social networks users

Step3. Verification victimization form and validate of answers. Step4. raise users to transfer any government ID proof like PAN card, adhaar card copy etc. Step5. confirmatory those Govt. ID card data victimization matching techniques with the user profile attributes Step6. higher cognitive process to urge the ultimate result (same person or completely different user).

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

**VII. Experiments**

During this following section, the experimental ways and techniques square measure bestowed for the projected approach to validate the faux profile detection method. As face book is that the most well-liked social network. Face book user profiles data is extracted and therefore the numerous users' attributes like Profile Name, Address, Interests, User Home Page, Likes and Friend relationship square measure more to the world information. Then we tend to relate those attributes to different social network to envision whether or not there's the other similar user profile exists. then the actualuser detection method has been evaluated to validate them.

Software analysis: For the projected approach, the Neo4j a pair of.2.5 tool is employed for analysing the profiles dataset in graphical format as shown within the figure on top of. The projected detection approach is initial enforced on in public accessible datasets. Computer code like PHP (Hypertext pre Processor) is employed for describing the hypertext markup language structure of profiles within the net and store those detail attributes of profiles in My SQL information that is termed as world information. Then compare their attributes of profiles and extract those similar user profile knowledge to represent them in social graph and then it'll valid victimization the projected detection model.

**VIII. Future Plan:** Social network analysis deals with advanced network structure of profiles and that square measure speedily ever-changing over time. This propose work ANalyze those in public accessible social network knowledge and gift an approach for locating duplicate profiles by profile similarity technique and store them within the world information and supply mechanism to envision the actualuser identity. identity verification may also be enforced to validate the actualuser identities supported finger prints, signatures, facial structure, person's voice etc. that is exclusive for every user. This work done will be improved by Future analysis by implementing profiles matching algorithmic rule to verify suspicious profiles and therefore the user verification method will be improved by victimization identity verification system to validate the suspicious profile. {data mining|data methoding} techniques may also be wont to improve the duplicate profiles detection process and to investigate them to envision in what number social networks those similar profiles square measure existed.

**IX. Conclusion:**

Our approach identifies the actual user profile on social networks. This projected approach initially establish the common

ISSN :2455-0108 Impact Factor: 2015-3.462
IJO–Science
INTERNATIONAL JOURNAL ONLINE OF SCIENCE
Volume II Issue II April 2016

attributes to envision within the similar profile is exist in tin and to envision the actual user.Then the profile matching mechanism has been enforced by scrutiny the similarity between the attributes and supply a choice algorithmic rule that justify our approach. so it solves the matter of identifying fake user profile and detects the actual real user.

**References:**

http://ijcsit.com/docs/Volume%207/vol7issue2/ijcsit2016070213.pdf

http://ieeexplore.ieee.org/document/5272173/?reload=true