



# AVOID CAMOUFLAGED USERS ON A SOCIAL NETWORKS USING URGA

<sup>1</sup> DURGA SHANKAR BAGGAM, <sup>2</sup> SALINA ADINARAYANA, <sup>3</sup> RAGHURAMAN.K

<sup>1</sup> Assistant Professor-II/CSE ,SRC,SASTRA University, Kumbakonam, Tamilnadu, India

<sup>2</sup> Associate Professor/IT,Shri Vishnu Engg College Women,Bhimavaram,AndhraPradesh,India

<sup>3</sup> Assistant Professor-I/CSE ,SRC,SASTRA University, Kumbakonam, Tamilnadu, India

E-mail: <sup>1</sup>[baggam.cse@hotmail.com](mailto:baggam.cse@hotmail.com), <sup>2</sup>[s\\_adi\\_2k@yahoo.com](mailto:s_adi_2k@yahoo.com), <sup>3</sup>[raghunit@gmail.com](mailto:raghunit@gmail.com)

## ABSTRACT

Social network Sites (SNS) are the major source for social capital[6], now a day's dynamic mining [2] is a part of its activity for many organizations. Interestingly on the other end we can see that most of the users are having more than one registration on social networks [6] has been associated with two fold effects in positive and negative shades. Many young users are using these networks for adolescents [10] and for downloading free entertainment stuff, which may leads to lot of fake user registrations. Just receiving invitation to their emails, creates some of the user profiles and it then pulls the userdata of the mail account. The profile created may not be correct and this was copied [11] may lead to false or redundant information on the network domain. This became a common problem for all organizations, websites who are associated with the SNS to find many facts like number of users registered and tracking any person also sometime it may leads to ambiguity problem. As one may contain more than one profile created by her/him intentionally or unknowingly generated by anyone way specified above. Recent issues on SNS says that most of the Popular Indian film actress were defamed with fake user profiles which were created in Twitter by unknown persons and posted some morphing photos and cheap comments. These acts attracted most researchers to concentrate on providing solutions. The best feasible solution is the Centralized database, which may not be feasible for heterogeneous databases. so to overcome this we are proposing a technique in this paper known as URGA(User Registration Gateway Algorithm).

**Keywords:** *Social Capitals, RP(Registration page), DRS(Data Retrieval System), LD(Logical Database), ADSM(Adaptive Database Service Manager), URGA(User Registration Gateway Algorithm), UR-POLICY, RL(Record linkage)*

## 1. INTRODUCTION

Record linkage (RL) refers to the task of finding records in a dataset that refer to the same entity across different data sources. RL for remote database is a common situation from many years for implementation of many decision support systems, which is used to analyze remote data supporting a variety of decisions. RL is highly sensitive to the quality of the data being linked, so all data sets under consideration should ideally undergo a data quality assessment prior to record linkage. Many key identifiers for the same entity can be presented quite differently within data sets, which can greatly complicate record linkage. The data needed (local) master file called local Database; these techniques compare the common data to support these decisions that are often scattered as heterogeneous nonkey attribute values

of the two records to derive distributed databases. In such cases, it may be similarity measure typically the probability of a match necessary to link records in multiple databases. If the similarity measure can be consolidate and use the data pertaining to the same real-world entity. If the databases design is same, RL techniques can be easily done using the primary real-world situations. However, since these heterogeneous databases are usually designed and managed [2] where all the records are available either locally or remote by the organizations. They can be accessed by using some common key attributes like email-address or social-security-numbers [8] although; it may be possible by transferring the entire remote relation as a matching data from various sources in a batch. The databases exhibiting entity heterogeneity are distributed, and it is not possible to create and maintain a linkage. Finding records from a remote



database and linking them is traditional, maybe impractical for several reasons like updating, ownership cost allocation failure leads to staleness of data [2]. The systems must agree to transmit incremental changes to the data warehouse on a real-time basis. Even though they are having many limitations. This is because the warehouse must maintain a linkage table for each pair of sites, and must update them every time for the associated databases changes. The participating system allows controlled sharing of portions of their databases using standard database queries, but they do not allow the processing of scripts, stored procedures, or other application programs from another organization. Here our intention is not to discuss on ability of any existing systems but to suggest a system that is more efficient for record linking and comparing user profiles to maintain Data cleaned Social capitals.

## 1.1 MOTIVATION

The situation that motivated me to concentrate on this work is when I am searching for one popular personality on facebook, I identified problems like: (1) one user having more than one account's (2) genuinely active identity conflict. These things encouraged me to identify a system that can determine user genuinity. And also I understood that most of the users are not ready to project there correct information onto SNS and many young users may have more than one account for different purposes, measures of Facebook use and perceptions of social capital was presented by Jessica vitak.et.al [6]. So how do we get the exact userdata? This gave a way to propose a solution of this kind where both the end users will be satisfied. As a part of this process I had come across many similar problems [1] [2] [6] [7] [8] [9] [10] [11] and understood that there is a need for better system to handle userdata.

We have organized our work into following sections where **Section-2:** Related work **Section-3:**URGA System Architecture **Section-4:** Algorithmic approach **Section-5:** Results **Section-6:** Conclusion and Future work.

## 2. RELATED WORK:

This problem can be addressed as two major parts: Record linking and comparing profiles. We identified that Debabrata et.al [2] is one of the best solution for linking and then that pulled data must

be maintained in a logical database (LD) and then compare the details by using Pasquale et.al [9]. Here the above two solutions are independent entities; they should be integrated into a system to respond to the situation. This system is Designed and proposed as a User Registration Gateway Algorithm (URGA).

## 2.1 PROBLEM FORMULATION

Let a be an enquiry record generated on local site and b be a record on remote database R. Then records a and b can be linked by using a matching pattern U.

applying conditional probability linkage[2]:

$$P(U) \geq \alpha \text{ where } \alpha \in [0,1]$$

Branching probability can be realized by using  $\pi$ .

Let V be the set of attributes acquired into our local database then matching probability is  $P|V$ .

Possible realization boundary values are:

$$P^L|V \leq P|V \leq P^U|V$$

where L denotes lower and U denotes upper.

Always  $\alpha$  should be above  $M=1$ .

Character based comparison is adopted to match Similarities between two strings.

$$\Sigma (T_1, T_2) = 1/\mu \Sigma_{i=1..m} I_{T1[i]=T2[i]}$$

Where i is the character of the string, and  $\mu$  is the length of the string.

Global Similarities among the users are calculated as:

$$\sigma_{Ai}(U_x, U_y) = \sigma_{Ai}(U_y, U_x)$$

This returns values in [0, 1]

This calculation can be represented by adjacency matrix of A where i is the user details. If the matrix  $A[x,y]=1$  if there is an edge between the nodes representing  $U_x, U_y$  and otherwise 0 and this can be evaluated by Katz similarity coefficient[9]. This classification is called Binary Classification Problem(BCP).

## 3. URGA SYSTEM ARCHITECTURE

This was designed to avoid fake users on SNS and to have reliable services through policy-based actions. This architecture was introduced in such a way that it acts dynamically and responds to the environment. The UR-policy was designed to analyze the user inputs by sending to **Database Retrieval System (DRS)** and this gathers a set of records pertaining to user from various webservers using effective record linkage procedure. **Adaptive Database Service Manager (ADSM)** Analyze and further refine to check user genuinity. The Data Retrieval System (DRS) observes the input parameters associated with the

**Registration Page (RP)** dynamically. The Adaptive Database Service Manager (ADSM) takes corrective actions based on the results that are pulled from various Databases or Webservers over the Internet into a **logical database (LD)**, through this system.

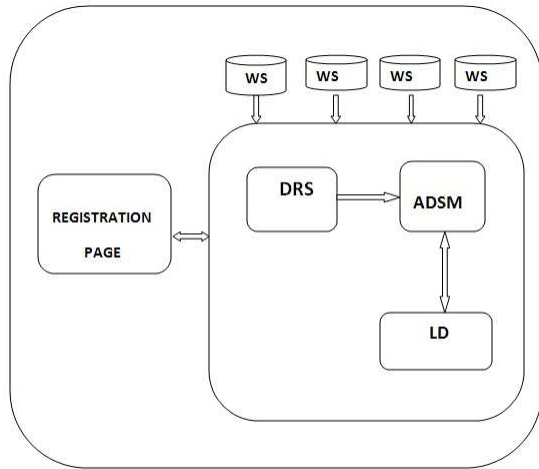


FIGURE-1: URGA ARCHITECTURE

### 3.1 DATA RETRIVAL SYSTEM:

The end user will be giving their data in Registration page. The keyed information will be monitored dynamically by our DRS. Now it shortlists the data to make a set of values used to pull some more information pertaining to user from different web servers and pushes to LD

### 3.2 ADAPTIVE DATABASE SERVICE MANAGER:

Here the LD will be processed by ADSM to analyze the user's reliability. This Analysis is done by URGA. This leads to generate a consolidated report which the DRS will use to get the accurate inputs from the user.

## 4. ALGORITHMIC APPROACH

Algorithm for retrieving and comparing data

```

Algorithm retrieve ()
{
//Let G be the T1
// cr be the current db for retrieval ((i.e.,) MID db
given by the user)
If(mid DB ==cr db)
    exit();
else
    
```

```

Call compare ()
}
    
```

Algorithm compare ()

```

{
// sequential RL technique
//q be the query string for retrieving data
// N be the variable for username, Fn for father's
name, Mid for Mail id
// ssn for social security Number, dob for date of
birth.
//i is the possible chances of retrieving data based
on various attributes.
Switch case(i)
// retrieving data based on MID.
Q= retrieve N, FN, MID,SSN from cr where
CR.MID =(retrieve DB2.MID from db2 where
db2.MID =(retrieve db3.MID from db3 where
db3.MID=( retrieve G.MID FROM G)
// retrieving data based on SSN.
Q= retrieve N, FN, MID,SSN from cr where ssn
=(retrieve db2.ssn from db2 where db2.ssn
=(retrieve db3.ssn from db3 where db3.ssn=(
retrieve G.SSN FROM G)
// retrieving data based dob .
q= retrieve N, FN, MID,SSN from cr where dob
=(retrieve db2.ssn from db2 where db2.dob
=(retrieve db3.ssn from db3 where db3.dob=(
retrieve G.dob FROM G)
// Likewise comparisons are made with enough
values of the attributes.
    
```

## 5. RESULTS:

To support this problem with solution, a system was implemented using dotnet technologies. Creating a dataset as different heterogeneous databases interpreting as mail servers with user data. On the frontend Registration page is created for user to register to the website using mail-id for a new SNS account. During this activity background process like DRS and ADSM will perform **Data Cleaning** action and that is shown in the below figures.

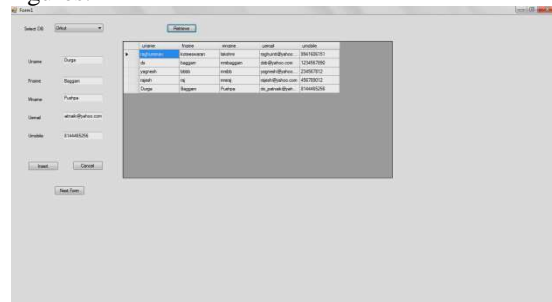


Fig-2: DRS Process Result

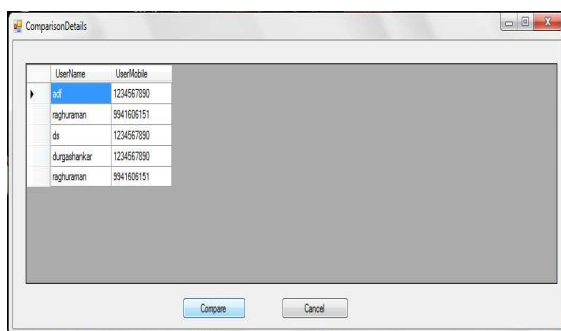


Fig-3 :ADSM process Result

## 6. CONCLUSION AND FUTURE WORK:

social networks are attracting more researcher's because it is dynamic in nature and depends upon the behavior of large set of people with different age groups from different origins of the world and their acts, result in different areas like Database size, complexity in cluster analysis, redundancy and many more DB issues like complexity in similarity join computation. One of the most interesting factors that motivated me is to identify a genuine userdata and avoid unauthorized user entry for adolescent action.

This is our preliminary work to have a better understanding of the system. As a result we had studied many papers related to it and accessed to many SNS's by registering on entering fake data. Interestingly found that most of the sites are not processing any verification other than email-id. This can be created by anyone as they like by giving details (invalid). We proposed a method URGA which acts as a gateway for any user registration. We implemented this algorithm in an stand alone application and tested it with sample data and found that it yielded better results as shown above in 2.4. The complexity here is all registration pages should undergo this process. Future enhancement for this technique is to implement a webserver with some programs like e-mail; SNS's on a LAN and implement the best algorithm.

## ACKNOWLEDGMENT:

We would like to thank all our colleagues who encouraged us to fulfill this work. This work can't be completed without the support of our families. We also extend our special thanks to Dr.Chandrashekar Rao Professor in Mathematics, SASTRA University for his suggestions.

## REFERENCES:

- [1] Christopher Sibona et.al,"Unfriending on Facebook:Friend Request and Online/Offline Behavior Analysis" Proceedings of the 44<sup>th</sup> Hawaii International Conference on System Sciences – 2011.
- [2] Debabrata Dey et.al "Efficient techniques for online record linkage" IEEE Transaction of knowledge and Data Engineering Vol-23, No-3, Mar 2011 Pg: 373-387.
- [3] [http://en.wikipedia.org/wiki/Criticism\\_of\\_Facebook](http://en.wikipedia.org/wiki/Criticism_of_Facebook).
- [4] <http://emailverify.net/>
- [5] <http://powershell.com/cs/media/p/2933.aspx> "pull data using email".
- [6] Jessica Vitak et.al," The Ties That Bond:Re-Examining the Relationship between Facebook Use and Bonding Social Capital" in the Proceedings of the 44th Hawaii International Conference on System Sciences – 2011 Pg:1-10.
- [7] Jitesh et.al "The Enron Email Dataset Database Schema and Brief Statistical Report1".
- [8] Min-Feng Wang et.al,"Enterprise Email Classification Based on Social Network Features" published In ICASNAM 2011 Pg: 532-536.
- [9] Pasquale De Meo et.al, "Finding similar users inFacebook"at <http://www.facebook.com/press/info.php?statistics>.
- [10] Robert LaRose et.al, "FACEBOOK FRIENDS: COMPULSIVE SOCIAL NETWORKING AND ADJUSTMENT TO COLLEGE"
- [11] Yingzi jin et.al,"Mining dynamic social networks from public news articles for company value prediction" at SNAM published at Springerlink.com. NOV- 2011.
- [12] <http://dev.mysql.com/doc/refman/5.5/en/index-btree-hash.html>.
- [13] <http://www.statsoft.com/textbook/>