ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

A Secure Framework for Protection of Social Networks from Information Stealing Attacks

Dr. M.BalaGanesh [1] Mrs.V.Sathya[2] S.Vinoth Kumar[3]

Professor, Department of CSE, Sembodai Rukmani Varatharajan Engg College, Sembodai - 614 280, TamilNadu, India[1]

Asst. Professor, Department of CSE, Sembodai Rukmani Varatharajan Engg College, Sembodai - 614 280, TamilNadu, India[2]

M. E., Department of CSE, Sembodai Rukmani Varatharajan Engg College, Sembodai - 614 280, TamilNadu, India[3]

ABSTRACT: Social networks are online applications that allow their users to connect by means of various link types. Since these sites gather extensive personal information, there is a promising chance for leakage of personal information and inference attacks. To prevent the social networks from such attacks, a secure framework is proposed here. As first step, the social network is modeled as a connected graph where nodes and edges represent users of network and relationships among them respectively. Then three kinds of learning methods are applied for modeling the inference attacks. The sensitive attributes of each person's record is gathered by applying Naive Bayes Classification and each sensitive attribute is classified into a class set. This is known as local classification scheme. In relational classification scheme, relationship between nodes that is persons are examined and link information is inferred. Collective inference scheme attempts to use both the local and relational classifiers in a precise manner to attempt to increase the classification accuracy of nodes in the network. After modeling the network attacks, the sensitive attributes and friendship links are either removed or modified. Thus the social network is sanitized and removing details and links reduce the classification accuracy of classifiers. Thus the proposed approach effectively maintains confidentiality of the data set even after its release so that the attackers have no chance to infer sensitive information of users.

Keywords: Social Network Analysis; Data Mining; Social Network Privacy

I. INTRODUCTION

The main aim of the project is to prevent the social network from private information stealing attacks. Model the social network as a connected social graph. Implement local classifier, relational classifier and collective classifier on the graph .Modify the personal information of users. The scope of the project is to prevent the social network user privacy data, by converting the dataset into graph and generalization and suppression techniques are applied in order to provide security to the user's data.

Algorithm or techniques used in this research work are listed below. They are Local Classifier used for removing personal details. Relational Classifier used for removing link structure details. Collective inference classifier used for removing both personal details and link structure details. Social networks are online applications where a number of users share their favorite activities, professional life and etc. Since these networks gather huge amount of personal information of the user, the privacy of users have become a major concern.

The advertisers, marketing experts make use of this information for their marketing operations. The problem here is, they try to infer personal information from the user profile and their friend's profile. The social networks are subject to confidential information inference attacks. The attackers infer the private information of users by analyzing the attribute values of user profiles, and thus it affects the privacy of the users.This work focuses on the problem of private information leakage for individuals as a direct result of their actions as being part of an online social network. We model an attack

Scenario as follows: Suppose the famous social network site Facebook wishes to release data to electronic arts for their use in advertising games to interested people. However, once electronic arts have this data, they want to identify

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

the political affiliation of users in their data for lobbying efforts. Because they would not only use the names of those individuals who explicitly list their affiliation, but also through inference could determine the affiliation of other users in their data, this would obviously be a privacy violation of hidden details. To the best of our knowledge,
this work that discusses the problem of sanitizing a social network to prevent inference of social network data and then examines the effectiveness of those approaches on a real-world data set.

In order to protect privacy, we sanitize both details and the underlying link structure of the graph. That is, we delete some information from a user's profile and remove some links between friends. We also examine the effects of generalizing detail values to more generic values. We then study the effect these methods have on combating possible inference attacks and how they may be used to guide sanitization. In our proposed system we implement the below implementation procedure, we collect username and password credentials of social network users.

Download the profiles of users from social network site and store them as HTML files and then we will design a HTML crawler that parses HTML files and collects attribute values of user profiles. Store the results in database. After completion of this process we will export the records in database into **.CSV** file. The file is then converted into graph. In graph nodes represent persons and edges represent friendship links. Design Naïve Bayes Local Classifier, Weighted-vote relational neighbor (wvRN) Relational Classifier and Relaxation Labeling Network Classifier (it combines both local and relational classifiers). After completing this we will identify private detail attributes and define possible values for each detail attribute.

Each possible value is said to be a class, then we apply relaxation labeling on the graph and classify the persons into any of the predefined classes. Thus, initially perform inference attack on the dataset Then we perform the sanitization process for this we will have three steps, the below steps will show the proper sanitization process.

II. RELATED WORK

In this work, we touch on many areas of research that have been heavily studied. The area of privacy inside a social network encompasses a large breadth, based on how privacy is defined. Other papers have tried to infer private information inside social networks. In [4] He .J, Chu .W, and Liu .V, "Inferring Privacy Information from Social Networks," Proc. Intelligence and Security Informatics, 2006. He et al. consider ways to infer private information via friendship links by creating a Bayesian network from the links inside a social network. While they crawl a real social network, Live Journal, they use hypothetical attributes to analyze their learning algorithm.

Name of value	Variable
Node numbered i in the graph	n_i
All details of node n_i	D_i
Details j of node n_i	D_i^j
Friendship link between person n_i and n_k	$F_{i,k}$
The weight of a friend link from n_i to n_j	$W_{i.j}$
All friends of n_i	N_i
n_i is in class m	C_m^i

Table 1 Common Notation Used in the Work

For further clarity, in Table 1, we have a reference for many frequently used notations.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

It provide techniques that can help with choosing the most effective details or links that need to be removed for protecting privacy. Finally, we explore the effect of collective inference techniques in possible inference attacks. In [2] Gross .R, Acquisti .A, and Heinz .J.H, "Information Revelation and Privacy in Online Social Networks," Proc. ACM Workshop Privacy in the Electronic Soc. (WPES '05), 2005 Gross et al. examine specific usage instances at Carnegie Mellon. They also note potential attacks, such as node reidentification or stalking, that easily accessible data on Facebook could assist with. They further note that while privacy controls may exist on the user's end of the social networking site, many individuals do not take advantage of this tool. This finding coincides very well with the amount of data that we were able to crawl using a very simple crawler on a Facebook network. It extends on their work by experimentally examining the accuracy of some types of the demographic reidentification that they propose before and after sanitization**.**

Diameter of the largest component	16
Number of nodes in the graph	167,390
Number of friendship links in the graph	3,342,009
Total number of listed details in the graph	4,493,436
Total number of unique details in the graph	110,407
Number of components in the graph	18

Table 2 General Information about the Data

In Table 2, we provide some general statistics of our Facebook data set, including the diameter mentioned above. Common knowledge leads us to expect a small diameter in social networks [12]. Note that, although popular, not every person in society has a Facebook account and even those who do still do not have friendship links to every person they know. Additionally, given the limited scope of our crawl, it is possible that some connecting individuals may be outside the Dallas/Fort Worth area. This consideration allows us to reconcile the information presented in [12] and our observed network diameter.

The Facebook platform's data has been considered in some other research as well. In [7] Jones .H and Soltren J.H, "Facebook: Threats to Privacy," technical report, Massachusetts Inst. of Technology, 2005., Jones and Soltren crawl Facebook's data and analyze usage trends among Facebook users, employing both profile postings and survey information. However, their paper focuses mostly on faults inside the Facebook platform. They do not discuss attempting to learn unrevealed details of Facebook users, and do no analysis of the details of Facebook users.

The crawl consisted of around 70,000 Facebook accounts. In [10] E. Zheleva and L. Getoor, "To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private user Profiles," Technical Report CS-TR-4926, Univ. of Maryland, College Park, July 2008., Zheleva and Getoor attempt to predict the private attributes of users in four real world data sets: Facebook, Flickr, Dogster, and BibSonomy. They do not attempt to actually anonymize or sanitize any graph data. Instead, their focus is on how specific types of data, namely, that of declared and inferred group membership, may be used as a way to boost local and relational classification accuracy.

The defined method of group-based (as opposed to details-based or link-based) classification is an inherent part of our details-based classification, as we treat the group membership data as another detail, as we do favorite books or movies.

In fact, Zheleva and Getoor work provides a substantial motivation for the need of the solution proposed in our work. Finally, in [8] Lindamood .J, Heatherly .R, Kantarcioglu .M, and Thuraisingham .B, "Inferring Private Information Using Social Network Data," Proc. 18th Int'l Conf. World Wide Web (WWW), 2009, we do preliminary work on the effectiveness of our Details, Links, and Average classifiers and examine their effectiveness after removing some details from the graph. Here, we expand further by evaluating their effectiveness after removing details and links.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

III. PROPOSED WORK

The proposed system solves the problem of private information leakage in social network. First the social network data is collected and then converted into social graph. Three types of classifiers are applied on the graph to infer sensitive information of users and classify them into set of classes. Based on this classification model, the system identifies what are all personal attributes to be removed or modified. The system manipulates the relationship among the users of network and decides whether the friendship links should be removed or not.

Collective Inference classification accuracy increased. Removing the links between user from network achieve only minimal exceptions when compared with other learning methods. The proposed technique is effective at reducing the classification of networks for user details which that are classified as sensitive. Sanitization methods effectively handle the utility of user details (delete and maintain).

A. *Social Network Dataset Collection:*

To evaluate the effect that changing a person's details has on their privacy, we needed to first create a learning method that could predict a person's private details (for the sake of example, we assume that political affiliation is unspecified for some subset of our population).

In Figure. 1, Username and password details of users in social network such as Facebook are collected. Log in to user accounts and download their profiles as .html files. Now apply html parser to that parses HTML files and collects attribute values of user profiles. Store the results in database. The records in database are exported into .csv format file for network classification. Model the dataset file as network graph.

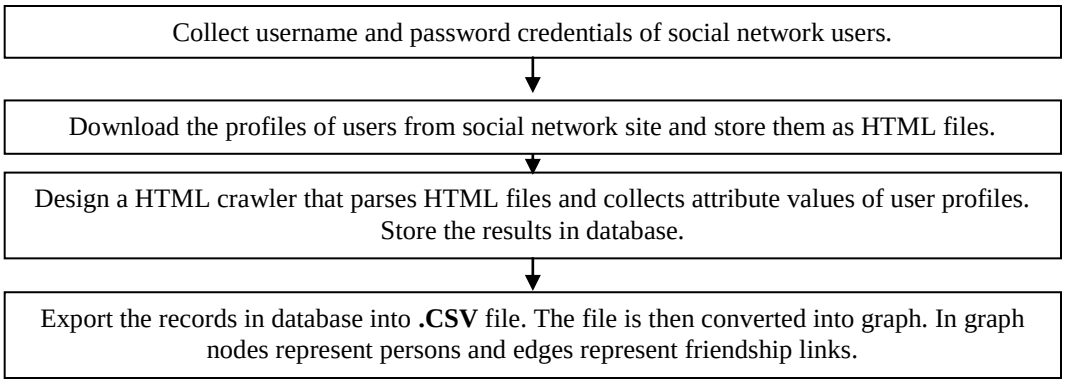

Fig. 1. Social Network Dataset Collection

B. *Inference Attack Modeling:*

Design naïve bayes local classifier, wvRN classifier and relaxation labeling network classifier. Relaxation labeling method utilizes both naïve bayes and wvRN classifier. Now apply Relaxation labeling on the network graph and thus model inference attacks as shows in Fig. 2.

Using naı¨ve Bayes as our learning algorithm allowed us to easily scale our implementation to the large size and diverseness of the Facebook data set.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

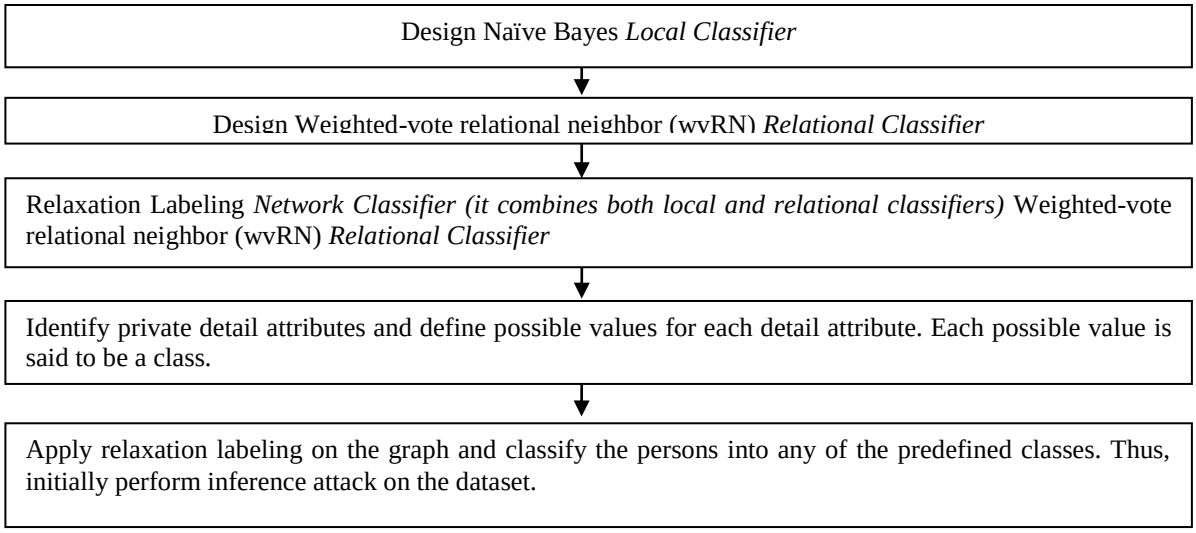

Fig. 2. Inference Attack Modeling

C.*Sanitizing Social Network Dataset:*

In sanitizing procedure, first the private details of nodes i.e persons are removed, as shown in Figure. 3. Then links that are more sensitive to inference attacks are removed. Finally values of some detail type attributes are generalized by replacing original values with some reference values.

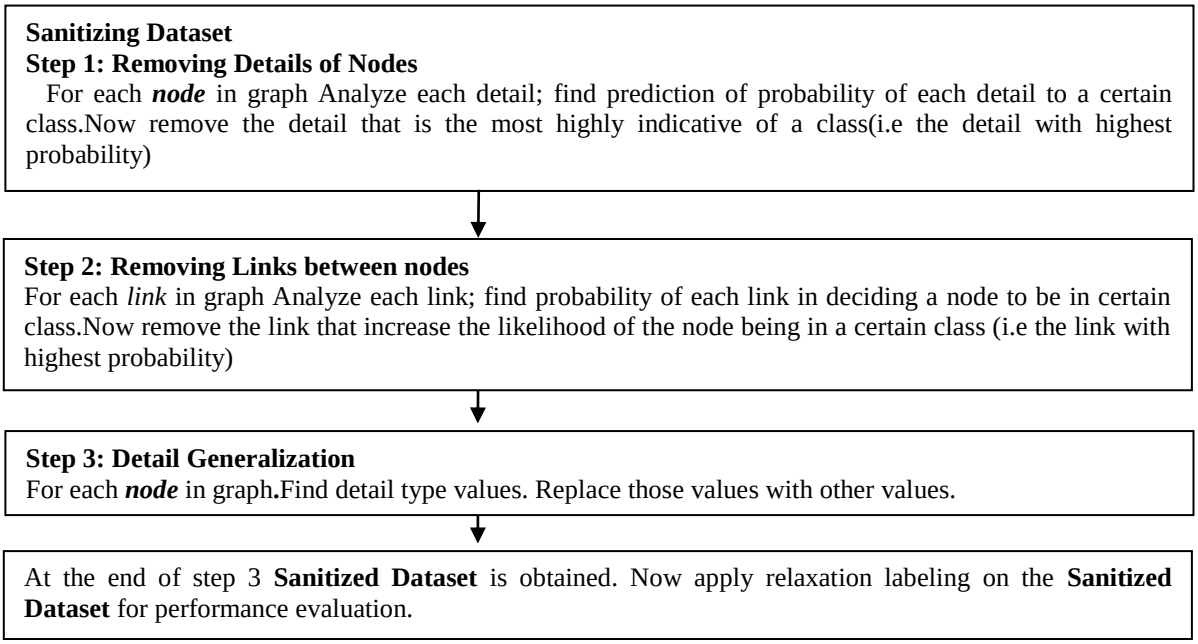

Fig. 3. Sanitizing Social Network Dataset

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

D.*Constructing Secured Social Network Database:*

To protect social network database from attacks the plain text values are first generalized. In generalization, values of numerical type attributes and categorical attributes are replaced with generalized values. In suppression, parts of the attribute values are replaced with some special characters.

Finally SQL aware AES encryption technique is used to encrypt the generalized and suppressed records. Thus secured database is constructed as shown in Figure. 4.

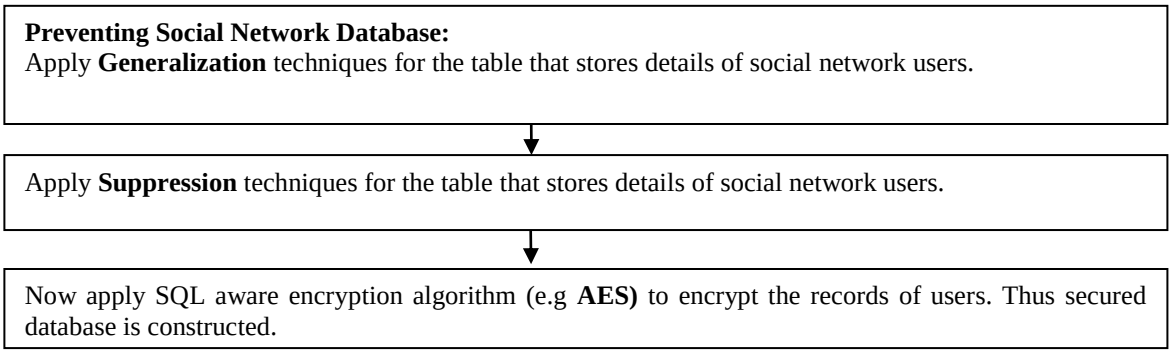

Fig. 4. Constructing Secured Social Network Database

IV. RESULTS

We note that in the Facebook data, there are a limited number of "groups" that are highly indicative of an individual's political affiliation. When removing details, these are the first that are removed. We assume that conducting the collective inference classifiers after remov-ing only one detail may generate results that are specific for the particular detail we classify for. For that reason, we continue to consider only the removal of 0 details and 10 details, the other lowest point on the classification accuracy. We also continue to consider the removal of 0 links and 10 links due to the marginal difference between the ½6; 7& region and removing 10 links.

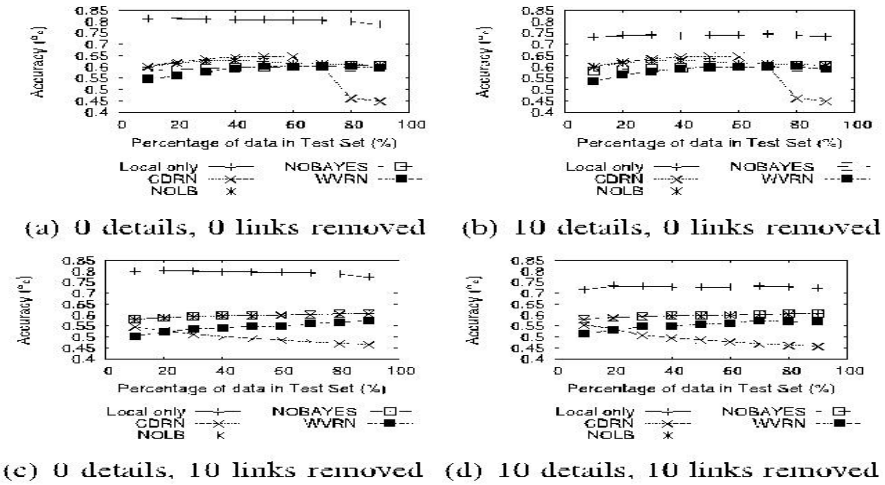

Fig. 5. Prediction accuracy of relaxation labeling using the Average local classifier (political affiliation).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

For the experiments using relaxation labeling, we took the same varied ratio sets generated previously. For each, we store the predictions made by the details only, links only, and average classifiers and use those as the priors for the NetKit toolkit. For each of those priors, we test the final accuracy of the cdRN, wvRN, nLB, and nBC classifiers. We do this for each of the five sets generated for each of the four points of interest. We then take the average of their accuracies for the final accuracy.

Fig. 5 shows the results of our experiments using relaxation labeling. The effects of collective inference on four real-world data sets: IMDB, CORA, WebKB, and SEC filings. While they do not discuss the difference in the local classifier and iterative classification steps of their experiments, their experiments indicate that Relaxation Labeling almost always performs better than merely predicting the most frequent class. Generally, it performs at near 80 percent accuracy, which is an increase of approximately 30 percent in their data sets. However, in our experiments, Relaxation Labeling typically performed no more than approximately 5 percent better than predicting the majority class for political affiliation. This is also substantially less accurate than using only our local classifier. We believe that this performance is at least partially because our data set is not densely connected.Our results in Fig. 5 indicate that there is very little significant difference in the collective inference classifiers except for cdRN, which performs significantly worse on data sets where there is a small training set. These results also indicate that our Average classifier consistently out-performs relaxation labeling on the pre- and postanony-mized data sets.

Additionally, if we compare Figs. 5a and 5b and Figs. 5c and 5d, we see that while the local classifier's accuracy is directly affected by the removal of details and/or links, this relationship is not shown by using relaxation labeling with the local classifiers as a prior. For each pair of the figures mentioned, the relational classifier portion of the graph remains constant, only the local classifier accuracy changes. From these, we see that the most "anonymous" graph, meaning the graph structure that has the lowest predictive accuracy, is achieved when we remove both details and links from the graph.

V. CONCLUSION AND FUTURE WORK

We addressed various issues related to private information leakage in social networks. We proved that using both friendship links and details together gives better predictability than details alone. In addition, we explored the effect of removing details and links in preventing sensitive information leakage. In the process, we discovered situations in which collective inferencing does not improve on using a simple local classification method to identify nodes.When we combine the results from the collective inference implications with the individual results, we begin to see that removing details and friendship links together is the best way to reduce classifier accuracy. This is probably infeasible in maintaining the use of social networks.

However, we also show that by removing only details, we greatly reduce the accuracy of local classifiers, which give us the maximum accuracy that we were able to achieve through any combination of classifiers. We also assumed full use of the graph information when deciding which details to hide. Useful research could be done on how individuals with limited access to the network could pick which details to hide.Future work could be conducted in identifying key nodes of the graph structure to see if removing or altering these nodes can decrease information leakage. To prevent the database of social network, generalization and suppression methods are applied. Then the encrypted versions of generalized and suppressed records are stored in the database**.**

REFERENCES

1.Backstrom.L, Dwork.C, and Kleinberg.J, "Wherefore Art Thou r3579x?: Anonymized Social Networks,Hidden Patterns, and Structural Steganography,", Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 181-190, 2007.
2.Gross.R, Acquisti.A, and Heinz.J.H, "Information Revelation and Privacy in Online Social Networks", Proc. ACM Workshop Privacy in the Electronic Soc. (WPES '05), pp. 71-80, http://
dx.doi.org/10.1145/1102199.1102214, 2005.
3.Hay.M, Miklau.G, Jensen.D, Weis.P, and Srivastava.S, "Anonymizing Social Networks", Technical Report 07-19, Univ. of Massachusetts Amherst, 2007.
4.He.J, Chu.W, and Liu.V ,"Inferring Privacy Information from Social Networks", Proc. Intelligence and Security Informatics, 2006.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 2, Issue 1, January 2014

5.Heussner.K.M,"Gaydar' n Facebook: Can Your Friends Reveal Sexual Orientation?", ABC News, http://abcnews.go.com/Technology/gaydar-facebook-friends/story?id= 8633224#.UZ939UqheOs, Sept 2009.
6. Johnson.C, "Project Gaydar", the Boston Globe, Sept. 2009.
7.Jones .H and Soltren J.H, "Facebook: Threats to Privacy", Technical report: Massachusetts Inst. of Technology, 2005.
8.Lindamood .J, Heatherly .R, Kantarcioglu .M, and Thuraisingham .B, 'Inferring Private Information Using Social Network Data', Proc. 18th Int'l Conf. World Wide Web (WWW), 2009.
9. Liu.K and Terzi.E, 'Towards Identity Anonymization on Graphs', Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), pp. 93-106, 2008.
10. Macskassy.S.A and Provost.F, 'Classification in Networked Data: A Toolkit and a Univariate Case Study', J. Machine Learning Research, vol. 8, pp. 935-983, 2007.
11. Zheleva.E and Getoor.L , 'Preserving the Privacy of Sensitive Relationships in Graph Data', Proc. First ACM SIGKDD Int'l Conf. Privacy, Security, and Trust in KDD, pp. 153-171, 2008.
12.D.J. Watts and S.H. Strogatz, "Collective Dynamics of Small- World Networks," Nature, vol. 393, no.6684,pp. 440-442, June 1998.