Synopsis of

Enhancing Privacy in Online Social Networks using Data Analysis

THESIS

Submitted in partial fulfilment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

AGRIMA SRIVASTAVA
ID. No. 2011PHXF413H

Under the supervision of

Dr. G. Geethakumari

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI

2015

1 Introduction

Online Social Networks (OSNs) are one of the biggest advancements that have happened in the past decade. Some of the popular OSNs are *Facebook, Twitter, LinkedIn, Pinterst, MySpace* etc. [1]. In comparison to the real world where information is ephemeral, the information on the web remains for an infinite time thereby posing a great risk on the privacy of online users. Most of the time users are unaware of the potential risks involved when they are sharing sensitive information online. Whenever and wherever the Personal Identifiable Information (PII) is shared and stored, privacy concerns are bound to arise. Hence, it is difficult to preserve privacy in a domain like OSN which is inherently designed for sharing [2].

Leak in sensitive data could result in lawsuits, loss of customers' confidence, brand damage, erosion of privacy, bad press, loss of revenue etc. Photos and videos from the profiles could be morphed and used for threatening, blackmailing and defaming individuals. Likes and interests reveal a lot about a person and can lead to the formation of controversial opinions. Using address, the schedule of a person could be known which can result into a criminal attack or burglary. Social Security Number (SSN) of individuals could be determined using a combination of address, date of birth and gender resulting in ID theft or impersonalization [3]. E-mails and phone numbers could be misused for targeted advertising leading to unnecessary interruptions and spam. Study of privacy if viewed from the prism of privacy enhancing algorithms has a lot of missing links and is a topic of high relevance. Hence, it is a great challenge to protect the confidential and sensitive data from unauthorized users and ensure that the actual data is available to the legitimate users as well [4].

2 Motivation

The main entities responsible for the users' information disclosure are *the users themselves, their online connections and the service providers*. The details of each are stated as follows:

- **Users:** The users themselves could be responsible for sharing a lot of sensitive information about them. This usually happens when they are unaware of the huge privacy risks that follow after carelessly sharing information online or if they face difficulty in managing the privacy settings of their profiles.

- **Online Connections:** The users' online connections can further spread their information in the network. This implies that privacy of users alone does not depend upon them but is also linked with its connections. This could also be considered as a case of an *interdependent privacy*.
- **OSN Service Providers:** OSN service providers can share users' data to the third parties for mining and getting deeper insights in order to improve their business solutions. The untrusted third parties could try to infer other sensitive attributes even if the released data set is anonymized. This can harm the privacy of individuals significantly and the service provider could be held guilty. This would eventually make the customers lose faith in the service providers thereby degrading their market reputation.

The main aim of the research work is to propose privacy enhancing algorithms and address the data privacy issues from the perspective of all the three entities.

3 Objectives of Research

The three sub problems to be addressed are enlisted below:

- Measuring users' OSN data privacy
- Enabling selective sharing of sensitive OSN data
- Protecting sensitive OSN data from inference attacks

4 Scope and Problem Definition

OSN users in general tend to be unaware of the risks that follow after sharing sensitive information. Hence, there is a need to measure data privacy of the user with respect to their connections. Using a privacy measuring scale the individuals could measure their sharing proportions with respect to others in the network and reduce the risk of privacy breach by managing their sharing behaviors. The degree of privacy for an information also depends on how it is likely to spread in the network. In a way, privacy of a node depends upon the sharing behaviour and its topology in the network. Therefore, it is also important to

analyze the network topology of the node in order to measure how privacy preserving it could be. Adopting a coarse grained privacy mechanism such as sharing information to a group of "close friends" or the strong ties of the network is one solution to minimize the risk of unwanted disclosure. However, this does not fully contribute in the process of protecting privacy. There is a high probability for an unwanted information disclosure even if the information is shared only with the strong ties in the network. Hence, it is important to quantify the online trust of users before sharing sensitive information to them.

In order to improve their business solutions data holders often release the social network data and its structure to the third party. Before its release, the data undergo node and attribute anonymization. This does not prevent the users from inference attacks which an untrusted third party or an adversary could carry out by analyzing the structure of the graph. Therefore, there is an utmost necessity to not only anonymize the nodes and their attributes but also to anonymize the edge set in the released social network graph. Where anonymization preserves privacy it also reduces the utility of the datasets. Finding an efficient utility based privacy preserving solution to prevent third party inference attacks for an online social network graph is a very important challenge which has been addressed in our work.

5 Description of the research Work

5.1 Measuring Users' OSN Data Privacy

In order to make the users aware of the data privacy concerns their information sharing behaviour along with the sharing behavior of their connections should be measured. Measuring this abstract and unobservable trait is a challenging task and is possible only if there is a proper metric of privacy. We used Classical Test Theory (CTT), a psychometric model to measure data privacy of OSN users [5]. We carried out an extensive survey and analyzed the collected data to gather the privacy perceptions of people in an OSN. The questions in the survey were designed to measure users' data privacy and the users' response was captured in the dichotomous response matrix which can either take the value of 0 or 1.

5.2 Calculation of Privacy Quotient using Classical Test Theory Model

Privacy Quotient (PQ) is the score given to the OSN user by analyzing their sharing behaviors. It is the potential risk that is caused by the users' participation in the network. If β_i is the sensitivity of the profile item *i* and $V_{(i,j)}$ is the visibility of the profile item *i* for a user *j* then the privacy quotient $PQ_{(j)}$ can be calculated using equation 5.1

$$PQ_{(j)} = \sum_{i}^{n} PQ_{(i,j)} = \sum_{i}^{n} \beta_i * V_{(i,j)} \tag{5.1}$$

where the range of items i.e. *i* varies from $1 \leq \text{i} \leq \text{n}$.

On the basis of the data collected from 482 respondents we have calculated the sensitivity of 11 profile items using equation 5.1. In Figure 1 we can clearly see that *address* is the most sensitive attribute followed by *political views, contact number, religious views* and *relationship status*. Whereas *birthdate, current town* and *hometown* details are shared by most of the users and are comparatively less sensitive. Figure 2 shows a bar graph

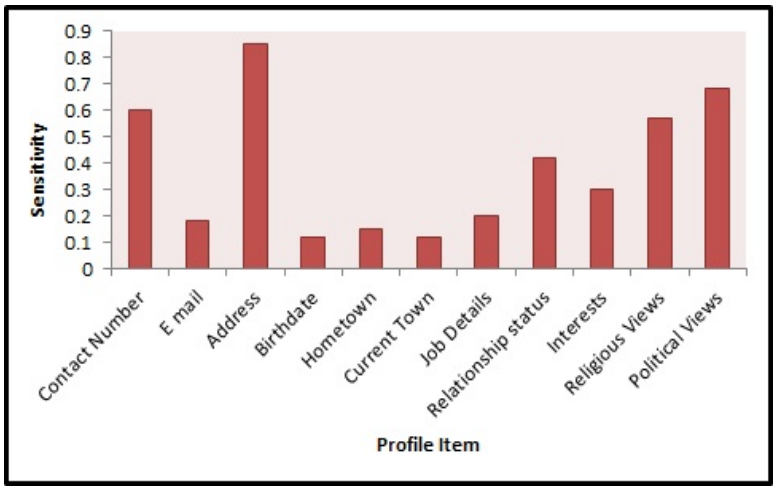

Figure 1: Sensitivity of the various profile items.

representing the number of users within a specific range of privacy quotient. In contrast to CTT the *Item Response Theory (IRT)* models are considered as *strong models* because their assumptions are stringent and do not meet the test data so easily. The IRT model is a nonlinear model, is greatly flexible and permits the calculation of one's probability of answering a question or sharing an item correctly. It fits the observed data well and computes intuitive values of *ability, sensitivity and visibility* [6].

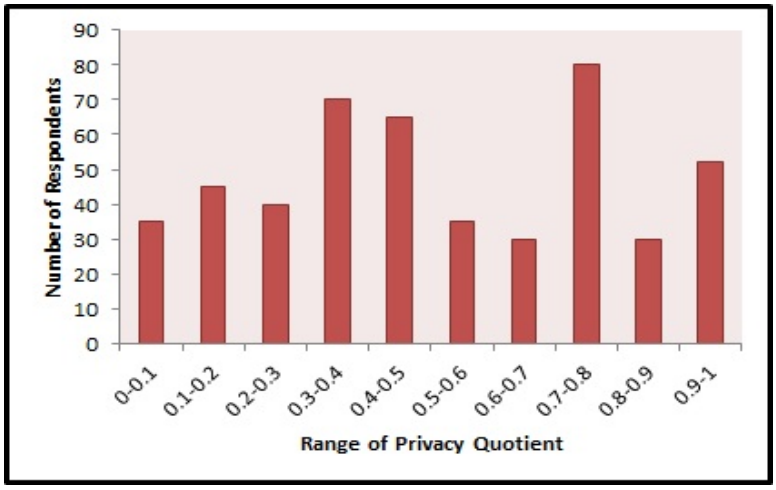

Figure 2: Number of users and the range of privacy quotient.

5.3 Calculation of Privacy Strength Using the Item Response Theory Model

Figure 3 represent the *item characteristic curve* for the *unconstrained and constrained one parameter logistic model*. The curves numbered from 1 to 11 represent 11 profile items namely : *Contact Number, Email, Address, Birthdate, Home Town, Current Town, Job Details, Relationship Status, Interests, Religious Views and Political Views*. Figure 4

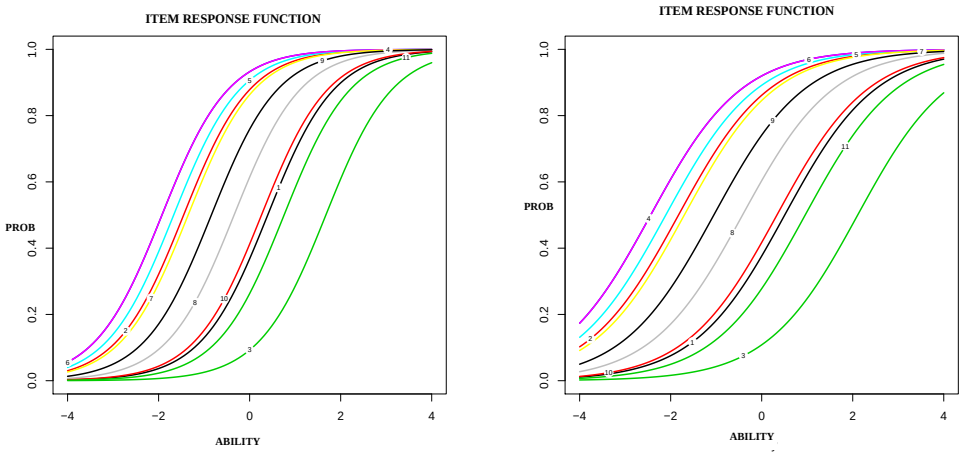

Figure 3: Item characteristic curve for the unconstrained and the constrained one parameter logistic model

represent the *item characteristic curve* for the *two parameter logistic model*. We see that the *political view* (curve no 11) has got the highest discrimination value and *address* (curve no 3) has got the lowest of all.

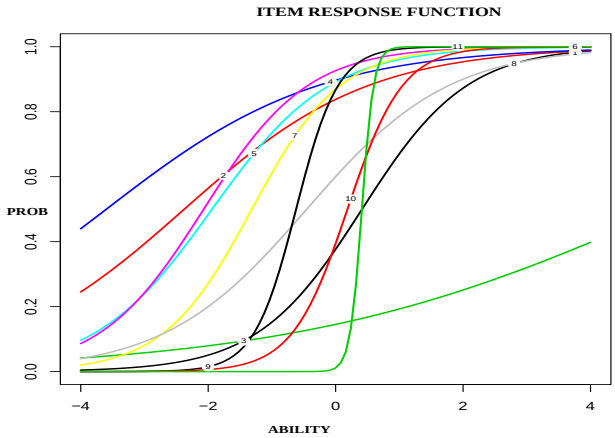

Figure 4: Item characteristic curve for two parameter logistic model.

We select the model which minimizes the loss of information. We use the concept of *Akaike Information Criterion (AIC)* and *Bayesian Information Criterion (BIC)* [7] [8] for model selection. The calculated PQs were grouped into five clusters of privacy strengths and were labelled as *High, Good, Average, Below Average and Poor*. Figure 5 represents the general statistics of the users' profile privacy strength. The framework outputs the list of users whose profiles have better privacy strengths than the user and also displays their sharing patterns. It allows the user to view the sensitivities of various profile items so that they can manage and enhance their privacy quotient and improve the privacy strengths of their OSN profile.

OSNs permit users to use privacy controls for the information that they want to share but managing and configuring privacy settings for every asset is tedious. Hence, there is a need for a system which could not only compare the users' privacy with their connections but could also recommend appropriate privacy settings for dynamic objects to them. Therefore, in the next subsection we will discuss the proposed privacy settings recommender system.

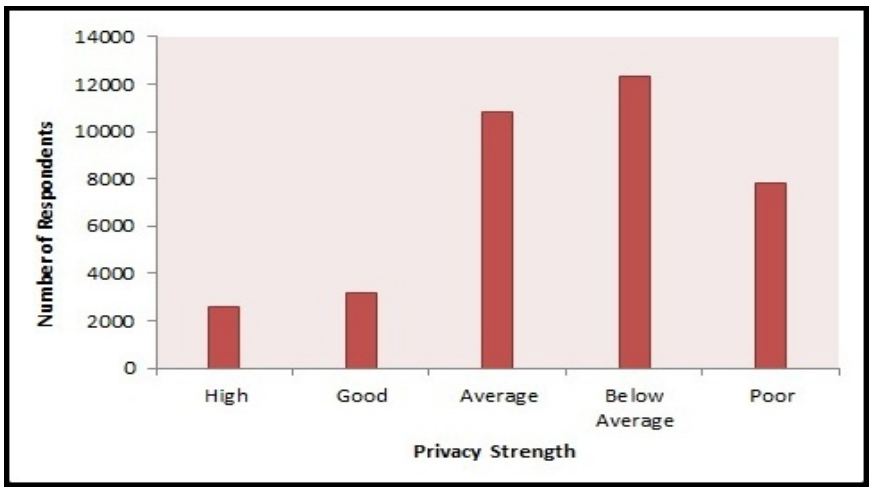

Figure 5: Number of respondents and their privacy strengths.

5.4 Privacy Settings Recommender System for Online Social Networks

Figure 6 shows the proposed privacy settings recommender system which recommends an optimal and meaningful privacy settings to the target user [9]. The solution goes through

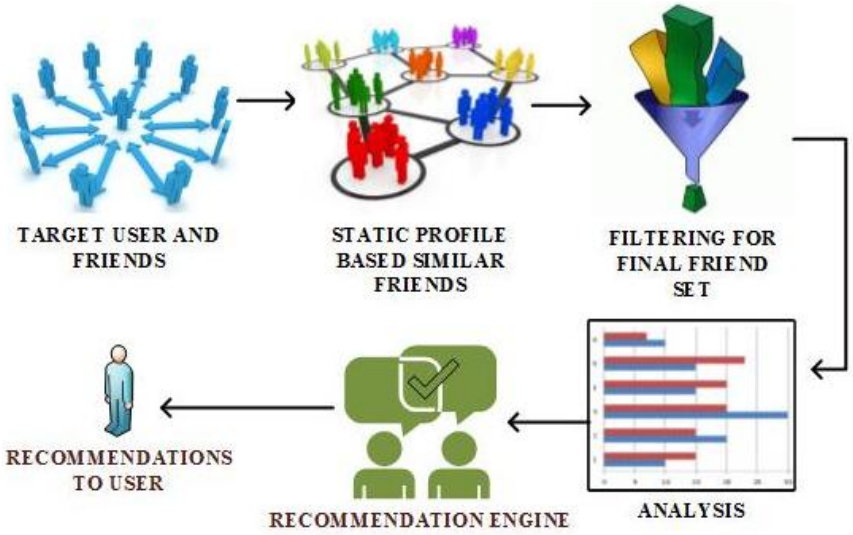

Figure 6: OSN privacy settings recommender system.

various stages and the output of one stage becomes the input for the next stage. Various stages of the privacy settings recommender system are as follows:

- Data collection of target users and their friends

- Formation of a user object matrix for dynamic contents
- Static profile based similarity and filtering
- Forming a refined Friend Set.
- A context based personalized privacy settings recommendation to the target user.

We selected target users from *Facebook* and applied our solution on their list of friends. We tested the solution on 150 *Facebook* users where each user had different number of connections. Figure 7 shows the average percentage of objects shared by the *Friend Set* and

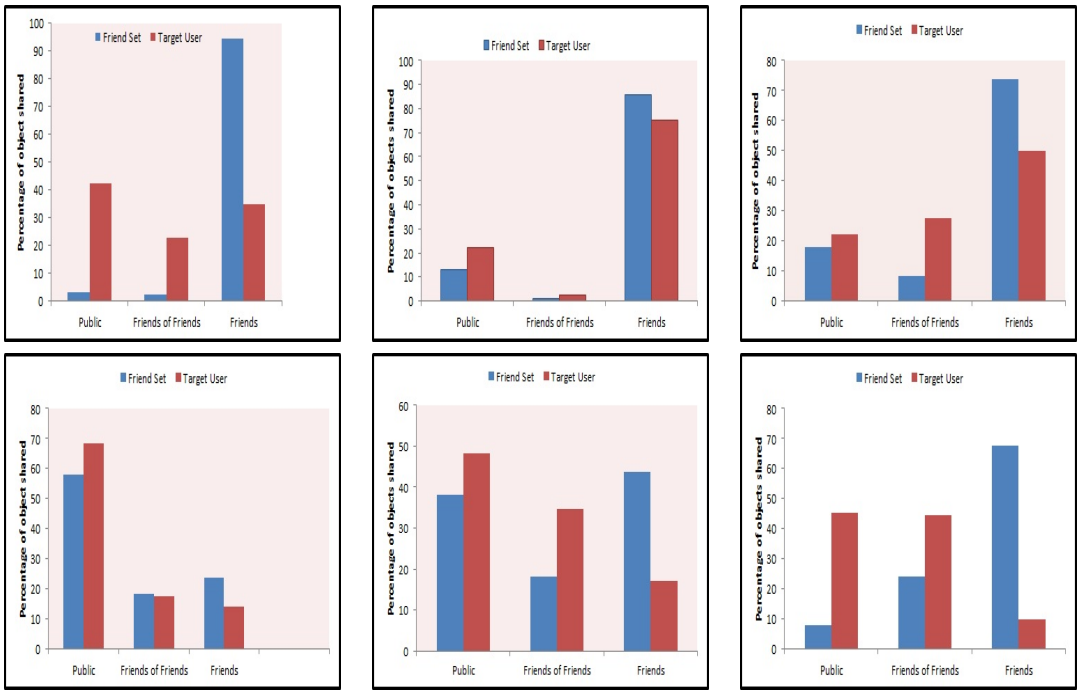

Figure 7: Comparison between the sharing behavior of target users and users in the Friend Set

compares it with the average percentage of objects shared by the target user for different visibility levels. The *Friend Set* is a group of users who have a better privacy quotient than the target user hence, an appropriate recommendation goes to the target user for modifying their privacy settings such that they can improve their privacy scores and can reduce the risk of information disclosure.

6 Enabling Selective sharing of sensitive data in an OSN

In this objective we aim to analyze the spread of the users' sensitive information to the rest of the network and prove that the nodes' sensitive information spread depends on its sharing behavior as well as its topology [10]. Each node has a *PQ, B and C* value. Here *PQ, B and C* represent the privacy quotient, betweenness centrality and closeness centrality respectively. Each property could either have a High (H) or a Low (L) value. Therefore, to identify the nodes having a high probability for spreading sensitive information we ran simulations and recorded the private information spread each time when the information was prevented from the nodes having properties as $\{\{L,L,L\},\{L,L,H\},\{L,H,L\},\{L,H,H\},\{H,L,L\},\{H,L,H\},\ \{H,H,L\},\{H,H,H\}\}$ respectively. Figure 8 shows the graph between the total number of nodes and the number of nodes knowing about the sensitive information of the source node [11].

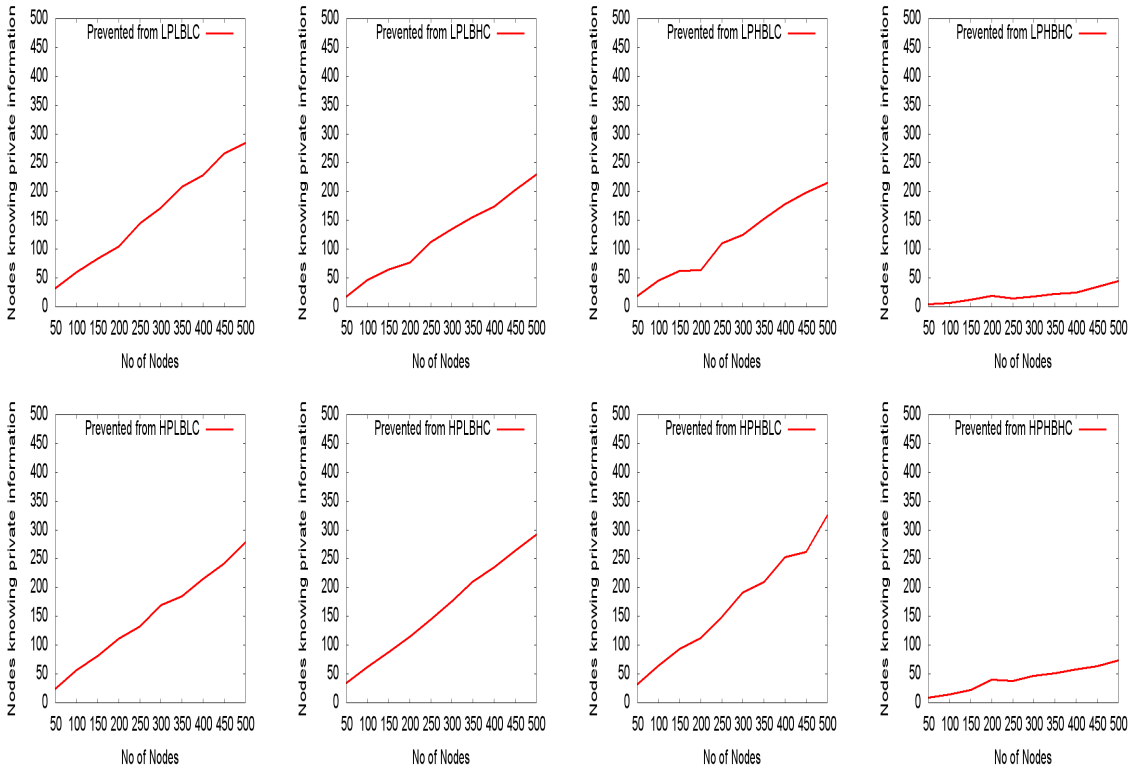

Figure 8: The spread of sensitive information after simulation

Observation: We observed that in all the cases the maximum sensitive information

flow is through nodes having the property of $\{L,H,H\}$ and $\{H,H,H\}$. An average spread of sensitive information was observed by the nodes having $\{L,H,L\}$ and $\{L,L,H\}$ property. Minimum sensitive information spread was observed through nodes having the property of $\{H,L,L\}$, $\{L,L,L\}$, $\{H,L,H\}$ and $\{H,H,L\}$. Table 1 summarizes the relation amongst the sensitivity of the shared profile item, property of the node, its group and the sharing suggestions given by the proposed privacy preserving algorithm.

Table 1: Sensitivity of the item, property of nodes, their groups and the sharing suggestions by the algorithm

Sensitivity of profile item	Property of node	Group	Sharing Suggestions
High	{L,H,H},{H,H,H}, {L,L,H},{L,H,L}	Group 1	Do not share sensitive information
Average	{L,H,H}	Group 2	Do not share sensitive information
Less	Any property	-	Share sensitive information following ICM.

In Figure 9, *LSI* indicates the Less Sensitive Information spread, *ASI(WM), ASI(M), HSI(WM) and HSI(M)* indicates the *Average Sensitive Information spread* and *High Sensitive Information spread*. Here *WM* and *M* indicates Without Model and with the implementation of Model respectively.

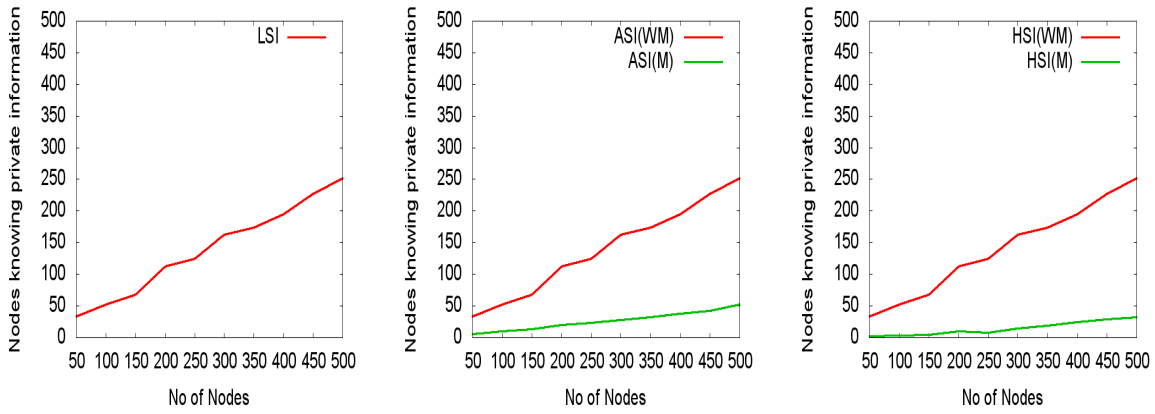

Figure 9: Simulation results for the spread of less sensitive information, average sensitive information and highly sensitive information

6.1 Quantifying Direct Trust for Sensitive Information Sharing in OSN

Direct trust is the trust value obtained with the direct relationship [12]. If *s* and *t* are the two nodes then the direct trust between *s* and *t* can be calculated as

$$DT(s,t) = \frac{W(R_i)}{N_s} * pfactor(s,t) \tag{6.1}$$

Here $W(R_i)$ is the weight of the relationship between *s* and *t*. A strong tie interacts more with the target user in comparison with a weak one hence, the strong tie will have a higher value of $W(R_i)$ than the weak tie.
N_s are the total number of edges a node *s* has.
pfactor of a node *s* for node *t* can be defined as

$$pfactor(s,t) = PQ(t) * (1 - PQ(s)) \tag{6.2}$$

6.2 Filtering out the Unconcerned Users from the Trusted Community

Using the *direct trust (DT)* calculated from equation 6.1 we categorize all the strong ties as *unconcerned, pragmatics and fundamentalist* [13]. The fundamentalist are the people who are cautious and worried about their privacy. They are distrustful of the organizations and choose privacy over any form of consumer service benefits. The pragmatics are the ones who believe that business organizations should earn the trust rather than assuming that they have it. They look at the benefits provided to them in comparison with the degree of intrusiveness of their personal information. The unconcerned people trust the organizations with the collection of their private information and are ready to use the customer service benefits in exchange of their personal information. In Figure 10 the *blue (rhombus)* line indicates the spread of sensitive information including the unconcerned nodes and the *red line (square)* indicates the spread of sensitive information without including the *unconcerned nodes*. The graphs show the percentage of nodes knowing the sensitive information with every iteration.

Observation: It was observed that while sharing the sensitive information with the trusted strong ties if the unconcerned users are excluded from the trusted community of strong ties the sensitive information-aware nodes are reduced to a great extent.

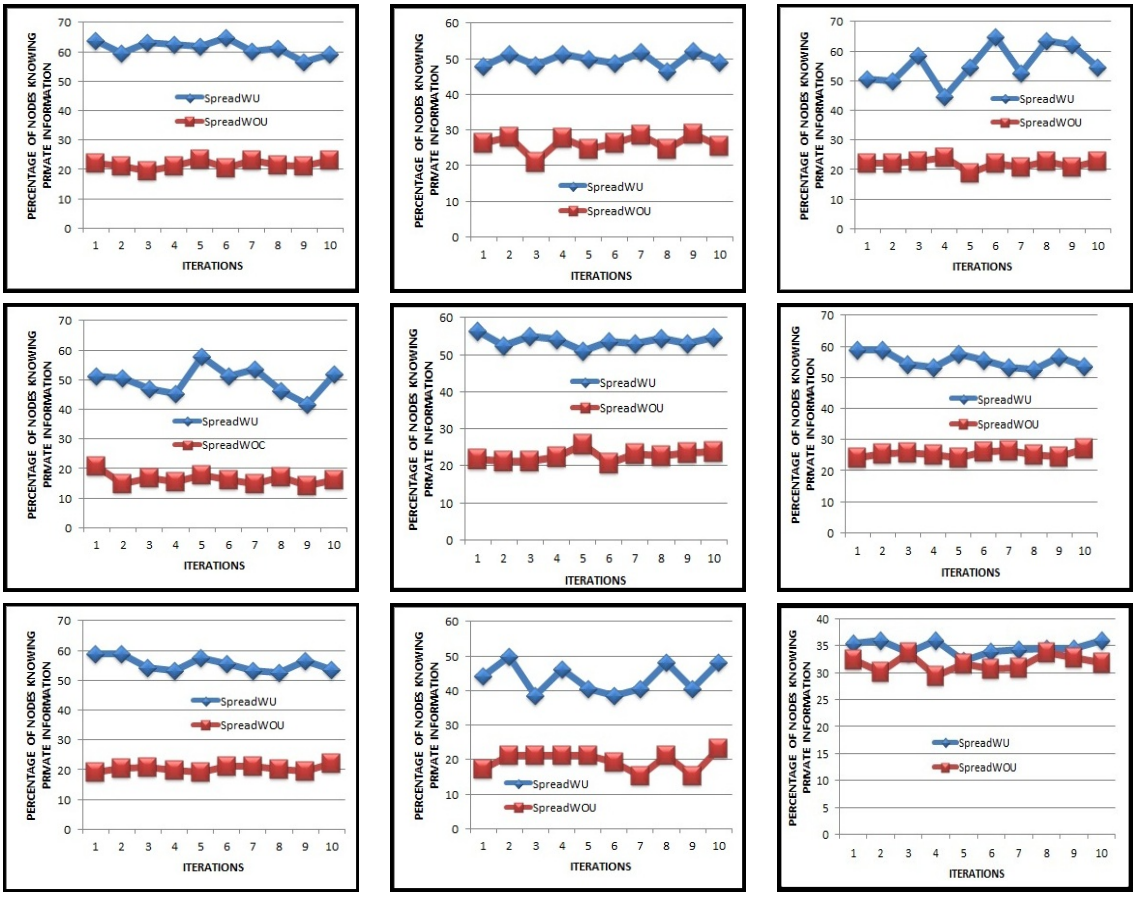

Figure 10: Comparison of sensitive information spread with and without including the unconcerned strong ties

7 Protecting Sensitive Data from Inference Attacks

In order to study social networks extensively the *data holders* share data and structural information to the third party data *recipients*. The data recipient then analyzes the interactions between nodes, looks at the structural patterns and concludes certain findings out of the same. They analyze the data for text analysis, search, image analysis, detecting patterns, target advertising etc. The data holder releases the attribute list and edge list to the third parties. The un-trusted third party could use network classification algorithms that involve the use of *local classifiers, relational classifiers and the inference algorithm* to infer private sensitive attributes from the available set of public and private attributes [14]. A node can possibly have four associations between them i.e. *a private node linking to a private node,*

a private node linking to a public node, a public node linking to a public node and a public node linking to another private node. In Algorithm 7.1 the complete graph *(G)* with a set of nodes *(V)* and the set of edges *(E)* will go as an input. This graph will have a set of *X* nodes i.e. the nodes whose sensitive attribute is set to public. It will also have a set of *Y* nodes i.e. the nodes whose sensitive attribute is hidden and is private.

Algorithm 7.1: RELEASING PARTIAL EDGE SET(V, E)

for all nodes $V_i \epsilon$ *V*

{ Label V_i as Private if $V_i \epsilon Y_i$.
{ Label V_i as Public if $V_i \epsilon X_i$.

for all edges $E_i \epsilon$ *E* formed by the nodes V_i and V_j make a labeled edge set E_{lab}

{ Label E_i as *Private-Public* if V_i is *Private* and V_j is *Public*
{ Label E_i as *Private-Private* if V_i is *Private* and V_j is *Private*
{ Label E_i as *Public-Private* if V_i is *Public* and V_j is *Private*
{ Label E_i as *Public-Public* if V_i is *Public* and V_j is *Public*

for all edges in $E_{lab}(i) \epsilon$ E_{lab}

{ Do not include the edge in E_{final} if E_i is labeled as *Private-Public*
{ Include the edge in E_{final} if E_i is labeled as *Private-Private*
{ Do not include the edge in E_{final} if E_i is labeled as *Public-Private*
{ Include the edge in E_{final} if E_i is labeled as *Public-Public*

Release E_{final}

After applying the network classification algorithm on the CoRA data set [15] the inference accuracy was recorded. We used the *Receiver Operating Characteristics (ROC) curves* which are an important visual tool to compare the accuracy of the classification model. For a given model an *ROC* curve shows a trade-off between the *true positive rate (TPR)* and the *false positive rate (FPR)*. Here the red line represents the accuracy of the network clas-

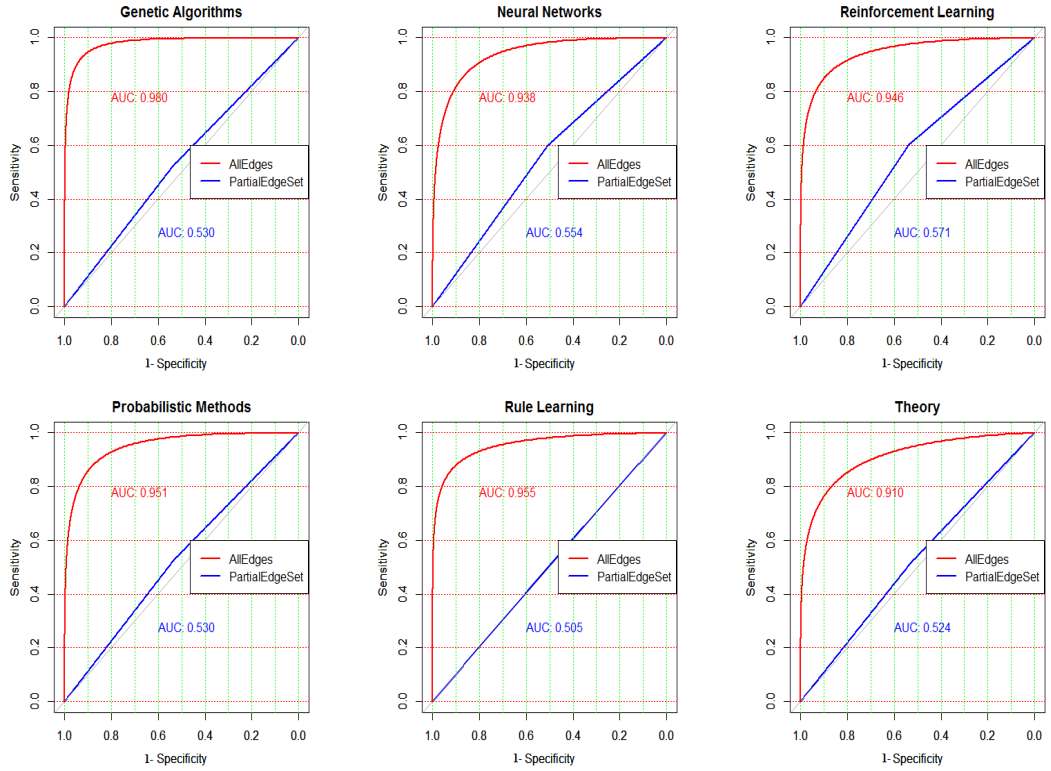

Figure 11: ROC curves for different labels of the CoRA data set.

sification algorithms when all the edges are shared and the blue line indicates the accuracy of the algorithm using the *partial edge set algorithm*.

Overall Deduction: The use of *partial edge set algorithm* could reduce the classification accuracy and could increase the misclassification rate significantly.

7.1 Proposed Privacy Utility Trade off Algorithm

A combination of assets which form a portfolio is referred to as "efficient" if it can give the best return for a certain level of risk. All the optimum portfolios should lie on the curve. Any portfolio that lies under the region represents a less than ideal investment. For our work we map utility with the rate of return and privacy with the *risk tolerance*. Each mechanism perturbs the data set D and changes it to the new data set D_A. This data set D_A will have a utility u and privacy p. In the proposed Algorithm 7.2, V and E are the set of nodes and edges in the actual graph. V_A and E_A are the set of nodes and the set of edges in the

anonymized graph. *NM* and *EM* are the mechanisms using which the nodes and edges in the actual graph are anonymized respectively. *C* is any collective classification algorithm and *TList* is the list of thresholds.

Algorithm 7.2: TRADE-OFF ALGORITHM($V, E, NM, EM, C, TList, UReq$)

for all nodes $V_i \epsilon$ *V*
Anonymize each node using the node anonymization mechanism *NM*
for each $EM_i \epsilon$ *EM*
{
  for each threshold $TList_i \epsilon$ *TList*
  {
    Anonymize edges using the edge anonymization mechanism
    EM_i and threshold $TList_i$
    Using the collective classification algorithm *C*, calculate the accuracy of inference
    Privacy = 1 - Accuracy
  }
  N_{G_1} = Neighborhood of all the vertices in graph G_1
  N_{G_2} = neighborhood of all the vertices in graph G_2
  Find out the similarity between the two graphs G_1 and G_2
  Utility = Similarity (G_1, G_2)
  Plot the utility, privacy pair on the graph
}
Determine the points on the efficient frontier
Output the best privacy preserving *EM* for a given value of *UReq*

Given different mechanisms and their respective *p* and *u* values the *data holders* should be able to select the mechanism which would ensure maximum utility and minimum privacy loss helping them to decide on the trade-off between privacy and utility.

8 Conclusion

Privacy Quotient (PQ) which is a score given to the users' on the basis of their sharing behavior was used to measure data privacy in OSNs. At first we carried out an extensive survey with a total of 600 respondents and used the CTT model to evaluate PQ for each subject. We also used the IRT model to implement a framework to measure the privacy strength of the users' profile for static profile items. We proposed and built a privacy settings recommender system which would compare the users' privacy quotient with their connections and also would recommend appropriate privacy settings for the dynamic profile items.

We proved that the privacy of a node not only depends on its sharing behavior but also on its topology in the network. A relationship between centrality and private information spread was drawn and it was observed that the nodes having low privacy quotient, high betweenness and high closeness centrality cause the maximum private information spread in the network. An information sharing algorithm was proposed which could prevent sensitive information from passing through those nodes which have a high probability of spreading information. The association between privacy and trust was closely studied and a trust enhancing model was proposed to refine the existing trusted community of strong ties after considering their online sharing behaviour. Data privacy from the perspective of inference attacks was studied thoroughly and a partial edge set anonymization algorithm was proposed that could reduce the accuracy with which the attackers could infer the sensitive attributes. Data perturbation results in the loss of utility. Hence, we measured the utility loss in the released dataset considering different structural metrics.

We proposed a utility privacy trade-off algorithm for the data sets anonymized using the edge anonymization algorithms. We used portfolio theory and the concept of efficient frontier to determine an appropriate trade-off such that the released datasets would have maximum utility and minimum privacy loss. With this work we hope to motivate advanced research in the field of data privacy specifically resolving issues related to measuring user privacy, enabling selective sharing of sensitive data and protecting sensitive data from inference attacks in OSNs. With our work we envision a number of privacy preserving applications which could be built using the proposed algorithms to make Online Social Networks a secure platform for sharing and exchanging information.

List of Publications

Peer Reviewed Journals

[Pub1] Agrima Srivastava and G Geethakumari (2014), "Privacy Landscape in Online Social Networks", Published in the *International Journal of Trust Management in Computing and Communications, Inderscience Publishers*, ISSN online: 2048-8386, ISSN print: 2048-8378, 2014.

[Pub2] Agrima Srivastava and G Geethakumari (2015), "Privacy preserving solution to prevent Inference Attacks in Online Social Networks", *International Journal of Computational Science and Engineering, Inderscience publishers*, 2015.

[Pub3] KP Krishna Kumar, Agrima Srivastava and G Geethakumari (2014), "A Psychometric Analysis of Information Propagation in Online Social Networks using Latent Trait Theory", Published in the Journal of 'Computing', Springer publishers, 2014.

International Conferences

[Pub4] Agrima Srivastava and G Geethakumari (2013), "Measuring Privacy Leaks in Online Social Networks", *Proceedings of the IEEE ICACCI-2013: Proceedings of the International Symposium on Women in Computing and Informatics (WCI-2013)*, August 22 - 25, Mysore, India, 2013, pp 95-100.

[Pub5] Agrima Srivastava and G Geethakumari (2013), "A Framework to Customize Privacy Settings of Online Social Network Users", *Proceedings of the IEEE International*

Conference on Recent Advances in Intelligent Computational Systems (RAICS) 2013, December 19 - 21, 2013, India, pp 187-192.

[Pub6] Agrima Srivastava and G Geethakumari (2014), "A Privacy Settings Recommender System for Online Social Networks", *Proceedings of the IEEE International Conference on Recent Advances and Innovations in Engineering - (ICRAIE-2014)*, May 09-11, 2014, India, pp 1-6.

[Pub7] Agrima Srivastava, K P Krishna Kumar and G Geethakumari (2014), "Preserving privacy in online social networks using the graph structural analysis", *Proceedings of the International Conference on Advances in Computing, Communications and Information Science, ACCIS - 14*, June 26 - 28, 2014, India. Proceedings in Elsevier India, pp 219-228.

[Pub8] Agrima Srivastava and G Geethakumari (2014), "Quantifying direct trust for private information sharing in an Online Social Network", *Proceedings of the International Symposium on Intelligent Informatics (ISI14)*, September 24-27, 2014, India; Proceedings in the Journal Advances in Intelligent and Soft Computing (Springer) Series, pp 21-30.

[Pub9] KP Krishna Kumar, Agrima Srivastava and G. Geethakumari (2014). "Preventing Disinformation Cascades using Behavioural Trust in Online Social Networks", *Proceedings of the International Conference on Advances in Computing, Communications and Information Science, ACCIS - 14* ISBN: 9789351072478, pp 113-123.

[Pub10] Agrima Srivastava and G Geethakumari (2015), "An efficient privacy utility approach for Online Social Network data publishing", accepted at the *IEEE 12th International Conference INDICON*, December 17-20,2015, India.

[Pub11] Agrima Srivastava and G Geethakumari (2015), "A framework for improving privacy strength of Online Social Network user profile", accepted at the *Grace Hopper Celebration of Women in Computing, India* 2015, December 2-4, 2015.

Bibliography

[1] Nicole B Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, Wiley Online Library, volume*, 13(1):210–230, 2007.

[2] Sabine Trepte and Leonard Reinecke. The social web as a shelter for privacy and authentic living. In *Privacy Online*, pages 61–73. Springer, 2011.

[3] Alessandro Acquisti and Ralph Gross. Predicting social security numbers from public data. *Proceedings of the National academy of sciences, National Acad Sciences, volume*, 106(27):10975–10980, 2009.

[4] Chi Zhang, Jinyuan Sun, Xiaoyan Zhu, and Yuguang Fang. Privacy and security for online social networks: challenges and opportunities. *Network, IEEE, volume*, 24(4):13–18, 2010.

[5] Ross E Traub. Classical test theory in historical perspective. *Educational Measurement: Issues and Practice, Wiley Online Library, volume*, 16(4):8–14, 1997.

[6] Xitao Fan. Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological measurement, Sage Publications, volume*, 58(3):357–381, 1998.

[7] Shuhua Hu. Akaike information criterion. *Center for Research in Scientific Computation*, 2007.

[8] David L Weakliem. A critique of the bayesian information criterion for model selection. *Sociological Methods & Research*, 27(3):359–397, 1999.

[9] J Ben Schafer, Joseph Konstan, and John Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166. ACM, 1999.

[10] Leucio Antonio Cutillo, Refik Molva, and Melek Onen. Analysis of privacy in online social networks from the graph theory perspective. In *Global Telecommunications Conference (GLOBECOM)*, pages 1–5. IEEE, 2011.

[11] Julian J McAuley and Jure Leskovec. Learning to discover social circles in ego networks. In *NIPS*, volume 272, pages 548–556, 2012.

[12] Na Li, Maryam Najafian Razavi, and Denis Gillet. Trust-aware privacy control for social media. In *CHI Extended Abstracts on Human Factors in Computing Systems*, pages 1597–1602. ACM, 2011.

[13] Alan F Westin and Louis Blom-Cooper. *Privacy and freedom*, volume 67. Atheneum New York, 1970.

[14] Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI magazine*, 29(3):93, 2008.

[15] Sofus A Macskassy, Foster Provost, and Foster Provost. Netkit-srl: A toolkit for network learning and inference. In *Proceeding of the NAACSOS Conference*, 2005.