International Journal of Research in Engineering, Science and Management
Volume 6, Issue 12, December 2023
https://www.ijresm.com | ISSN (Online): 2581-5792

Privacy loss in Online Social Network

Md. Afzal Ahmad[1*], Mritunjay Kumar[2]
[1,2]M.Tech. Scholar, Department of Computer Science, Central University of South Bihar, Gaya, India

***Abstract*: In the digital age, we are intricately connected through various social networks. Social networking platforms serve as avenues for communication, allowing individuals to interact with others through their profiles on specific networking platforms. Enormous amounts of data are generated every minute, driven by the vast user base of these platforms. Internet of Things (IoT)-based social networking platforms enable individuals to share public or private information, though users are cautious due to the critical importance of certain data. Despite the potential for information sharing in today's global internet era, social networks pose a serious threat to user privacy. Users may hesitate to share certain information, as the involvement of multiple users in sharing a data item puts the privacy of individuals at risk. While restrictions exist for users seeking access to others' data, these restrictions may not apply to posts, which are integral components of the social networking experience. In current online social networks, restrictions do not extend to the sharing of co-owned data. Each user holds their own opinion on who can access their data, and the posting of data is influenced by the opinions of the associated users involved. This thesis report explores how privacy loss is calculated for users involved in a group, considering different values of sensitivity and reputation. The computation of reputation is crucial in determining privacy loss, offering insights into the dynamics of information sharing within social networks.**

***Keywords*: Online Social Networks (OSNs), Privacy loss, Page rank algorithms, Sensitivity, Reputation.**

1. Introduction

Online Social Networks (OSNs) [1] are web-based services that provide a platform for individuals to share information and communicate with other network members through links. Various forms of social communication, such as sharing objects, organizing online events, and creating groups, are facilitated on OSNs. These platforms offer a digital environment for people to interact, with approximately 80% of active internet users currently visiting OSNs. A fundamental feature of OSNs is the ability to create and share personal profiles. Users post data in the form of messages, photos, or videos on OSNs, and some of this data is crucial and sensitive. Unauthorized access to such data can result in privacy violations. Over the recent decades, the popularity of OSNs has surged, providing users with virtual space to include profile information like gender, birthday, interests, education, contact details, and more. In multiparty resources where multiple users are involved, group photos are referred to as multiparty resources. An effective solution is required for the access control of multiparty resources to ensure privacy and security.

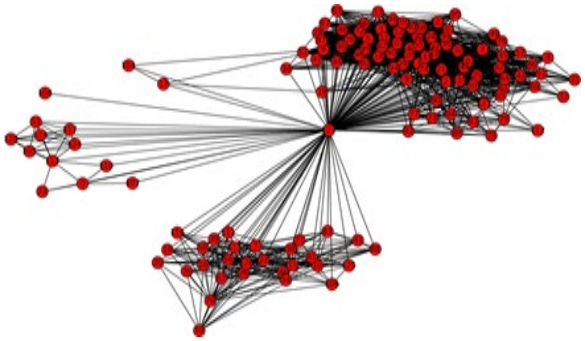

Fig. 1. Social graph

When a user uploads a group photo and tags other users present in the photo, the tagged users cannot restrict who can see the photo. Consider a scenario where users A and B are together in a photo. If user A posts this photo on a social network, and it is accessed by user A's colleagues, user B may find the photo sensitive. If user B is not familiar with user A's colleagues, it results in a privacy violation, a common issue in OSNs. Resolving such privacy concerns requires collaboration among all involved users. In cases where both A and B are owners of a photo, conflicts often cannot be entirely eliminated, leading to privacy loss for some users [2]. The owner seeks the opinions of all involved users before posting data, and posting is allowed only when certain conditions are satisfied. Reputation values among users are not fixed. If the owner's decision conflicts with co-owners' decisions, there is a reputation loss; conversely, if the decisions align, there is a reputation gain. If any user's privacy is affected due to another user, the reputation decreases concerning that user. This thesis report explores how privacy loss is calculated for users in a group when their data is shared on OSNs. The reputation of each user is calculated before and after data posting. Changes in the owner's reputation with respect to co-owners are analyzed: an increase may indicate a higher chance of privacy loss, while a decrease suggests a lower chance. Each node's page rank value represents their reputation. Sensitivity varies among users, and data may be less sensitive for some and more sensitive for others. Privacy loss is computed by considering different values of sensitivity and reputation.

OSNs, such as Facebook and Twitter, are closely associated with heightened security and privacy concerns. Users invest time in enhancing the content of their profiles [6], making privacy and confidentiality ongoing areas of significant concern for these users. Users are cautious about sharing every piece of

*Corresponding author: md.afzal8336@gmail.com

information, particularly with acquaintances rather than strangers [7]. In some instances, a user's content may involve the privacy of other users [8]. User privacy preferences may vary based on the type of information being shared. Employers often scour social networking sites to vet potential candidates before hiring them [9]. Presently, many OSNs restrict the level of information provided. An access control policy, also known as a privacy policy, outlines which users are permitted to access a user's data. Generally, a user privacy policy is represented by a set of users with whom the user wishes to share their data.

Jose et al. [11] proposed a mechanism for detecting and resolving privacy conflicts in social networks. The mediator seeks individual privacy policies from all users in a group and employs a mechanism to merge multiple users' privacy preferences into a single policy. This approach addresses the issue by prohibiting the sharing of an item if the privacy of any involved user is violated. Conversely, the item is shared only when privacy remains unviolated for all users involved.

Gulsum Akkuzu et al. [13] proposed a model to balance co-owned data sharing and user privacy. When a user shares content involving others, opinions are gathered before data sharing. Unfriending a user who leaks information serves as a form of punishment. User reputation is calculated based on feedback from connected users, where positive feedback reflects happiness and negative feedback indicates unhappiness. Owner reputation changes if shared data affects co-owners, depending on the type of data shared. Reputation systems in OSNs aim to assist users in deciding whom to trust and befriend, and in determining data availability. These systems provide insights into a peer's future actions based on their past behavior, with reputation values ranging from 0 to 1. A reputation value of 0 corresponds to a data sensitivity value of 0, indicating no concern, while a value of 1 signifies that all co-owners are concerned about the data's security, highlighting its importance to all users.

Liu et al. [15] demonstrated a method to measure privacy loss using two key parameters: i) sensitivity and ii) visibility. Sensitivity reflects the level of privacy risk associated with the data, with highly sensitive items being shared less frequently due to increased privacy loss chances. Visibility, on the other hand, gauges information spread; if information reaches a large audience, the likelihood of privacy loss increases.

2. Proposed Methodology

A. The Concept of Privacy

Privacy derives from the term "privatus," meaning separated from the rest. It lacks a universally specific definition applicable to all contexts. The risk to users' privacy intensifies when private data is entrusted to another party within an Online Social Network (OSN). The value of privacy in a given context hinges on the social importance within the network [20]. Privacy is often characterized as the right of individuals to determine for themselves. Sharing personal information publicly diminishes its meaning and intrinsic value. In OSN platforms, social relationships among individuals foster the erosion of privacy, making it challenging to safeguard on platforms explicitly designed for sharing. This chapter initially explores the general concept of privacy, delves into data privacy within an OSN, and examines the privacy loss associated with posting data on social networks.

B. System Model

The whole social networking site can be modelled as a directed graph G= (V, E) Where, V denotes the set of vertices and a vertex represents a user register with the OSN's.

$V= \{V_1, V_2, V_3, V_4, V_5, \dots, V_n\}$ denotes set of users.

E is the set of edges that denotes relationship between OSN's users.

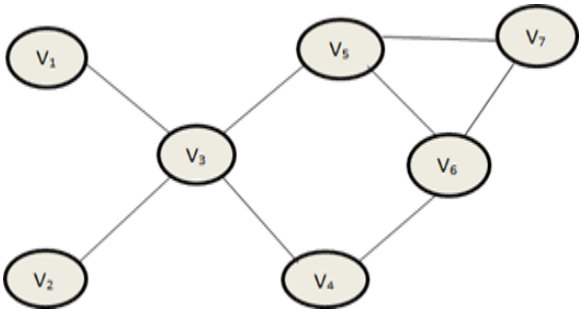

Fig. 2. System model graph

C. Sensitivity

Here we introduce sensitivity level (S_{level}) for each of the resources. More sensitive information is assigned with higher value of $S_{level} \in [0, 1]$ sensitivity decided by the user.

Each of the resource have sensitive level it is represented by a tuple (r_{id},S_{level}).

Some data will be more sensitive for some user but for another user it may less sensitive. If we take an example of data as picture consisting of 10 members for some members the loss of privacy may be low or high or very high. Hence, it can be said that for some members shared data might be not sensitive (no privacy loss) or sensitive (high privacy loss). Hence, based on this discussion we can say that we can divide sensitivity of data ranging between 0 and 1 into three parts.

Typical range of different sensitivity level

0.1-0.4 is kind of moderately sensitive data
0.41-0.7 is kind of sensitive data
0.71-1.0 is kind of highly sensitive data

D. Reputation

Reputation offers insights into users' behavior, predicting future actions and reducing interaction risks. Those with poor reputations are often avoided in group settings. User reputation in a network is influenced by neighboring connections, calculated through averaging appraisals from different sources. In photo sharing on Online Social Networks (OSNs), the owner's reputation is shaped by co-owners' decisions. Highly reputable users [24] face a greater risk of privacy loss due to their increased followership. Reputation is vital in deciding whether to interact with a user on OSNs, guiding trust, data accessibility, and respect for others' decisions in the data-

sharing process. Reputation creates a "shadow of the future" for interactions, with values fluctuating based on user activities. Conflicting decisions with co-owners result in reputation loss, while alignment leads to reputation gain. Respecting majority decisions signals the completion of the data-sharing process with majority satisfaction. User reputation evolves with their behavior.

Below is a figure depicting the sharing of data in a social network.

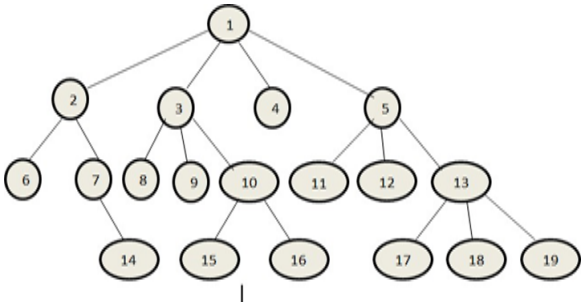

Fig. 3. Different level of social graph

In social networking, an increased level indicates more users interacting with the data. With a higher number of users interacting, the chances of privacy loss also increase. I aim to calculate privacy loss when sharing any data on a social network.

Privacy loss = Reputation x Sensitivity

Let's assume each node's PageRank value represents its reputation value.

E. Data Shared Model

A robust data sharing model is crucial to safeguard user data on online social networks. An effective model should be capable of adjusting the nuanced variations in intimacy between a user and their connections. For instance, if a female user prefers not to share her birthday group photo with all friends, common issues in OSNs, various solutions are employed to address such challenges.

Data shared model represented by following tuple:

- Owner: The user who wants to post the photo in social networks that user is known as the owner of those resources.
- Co-owner: Those users who are present in a photo that has to be posted on social network except owner all are called co-owner.
- Shared condition: This condition decides whether data can be shared or not shared.

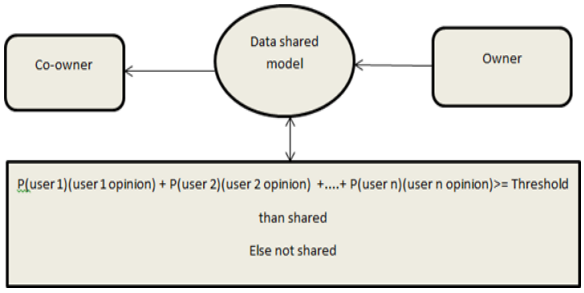

Fig. 4. Data shared model

Owner wants to post the data in social networks, but before posting the data, owner takes the opinion of all involved users and check the access condition for sharing the data.

Here in above figure:

P(user 1) is the page rank of user 1

P(user 2) is the page rank of user 2

.....

P(user n) is the page rank of user n

If their opinion will be yes, then take 1

If their opinion will be no, then take -1

If their opinion will be not any then we assume implicitly agree for sharing and take 1

Threshold is the average of all involved user page rank.

If satisfied above condition then owner post the data otherwise don't post the data in social network.

F. Pagerank

PageRank is an algorithm that measures the transitive influence or connectivity of nodes. The PageRank [25] algorithm to determine a page's importance. The algorithm assigns each page a relative numeric score of importance and authority by estimating the quantity of the links it contains. The algorithm is calculated using a simple iterative algorithm. It can be computed by distributing one node's rank over its neighbors. Cal- culating PageRank is quite time and memory consuming.

PageRank is defined as follows,

$$PR(P_i) = \frac{1-d}{N} + d(\sum_{pj \in M(P_i)} \frac{PR(P_j)}{l(P_j)})$$

Where,

N: Total number of nodes present in the graph

l(Pj): source of the incoming edges counts the total number of outgoing edges

PR(Pj): probability of the incoming nodes

d: damping factor usually let d=0.85

Damping factor [26] controls the convergence speed of Page Rank algorithm.

Therefore, a page that has many other pages linking to it is more important, and have a high PageRank. The PageRank algorithm works by giving individual PageRank, determined by the number of links that are pointed towards the page. In links to increase PageRank. In PageRank accurate values are obtained through many iterations. The value of the PageRank lies between 0 and 1. The PageRank value of individual node in a graph depends on the PageRank value of all the nodes which connect to it. In short PageRank is a "vote", by other pages in a graph.

Condition for algorithm exit check Margin of error

$|PR(t+1)-PR(t)| < \varepsilon$

$\varepsilon = 0.09$ (assume)

1) Damping factor

PageRank is one of the popular and widely used ranking method. The behavior of Page Rank also depends on the function of the damping factor [26]. Damping factor α ∈ [0, 1] used in the computation of PageRank. The damping factor parameter state that how much time random web surfer follow hyperlink structure than teleporting. Damping factor controls the proportion of time.

This was initially set to 85% or 0.85. Hence the damping factor is mostly considered to be 0.85. Many researchers observed that damping factor controls the convergence speed of PageRank algorithm. If we take α0.85 then we can say that out of total time 85% of time is taken by the web surfer to follow the hyperlink structure and 15% time they teleport to new web pages randomly.

2) Reputation calculated used Page Rank algorithms as a procedure

Algorithm-1 Page Rank Algorithms [25]

1. Enter the graph with the links
2. Initially set Page Rank=1/Total no of nodes
3. Counts the outbounds l(Pj) for each page i from 1 to N
4. Calculate

$$PR(P_i) = \frac{1-d}{N} + d\left(\sum_{pj \in M(P_i)} \frac{PR(P_j)}{l(P_j)}\right)$$

for each page i from 1 to N, where M(Pi) includes those pages that, have a link to page i.

5. Update all PR(Pi) for each page i from 1 to N
6. Repeat step 3 till changes to Page Rank (PR) are insignificant.

3) Updation of reputation after data sharing

We can say that which user has more rank that will be more reputed. Let 1, 2, 3, 4,.....n users have a group photo.

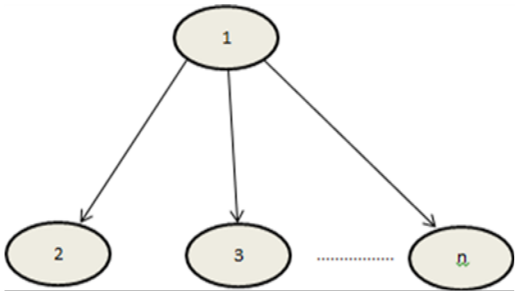

Fig. 5. Group photo of n user

User 1 wants to share this picture in social networks, at that time he is owner but, before sharing he wants to take the opinions of co-owner.

Let P(1) is the rank of node 1 in the graph. Let P(2) is the rank of node 2 in the graph. Let P(3) is the rank of node 3 in the graph.

Let P(n) is the rank of node n in the graph. If their opinion will be Yes then take 1

If their opinion will be No then take -1.

If their opinion will be not any then we assume implicitly agree for sharing and take 1.

We have to set a threshold.

If

P(1)(user 1 opinion) + P(2)(user 2 opinion) + P(3) (user3 opinion) +. + P(n)(user n

opinion)≥ Threshold then Share Otherwise not shared

Threshold = Averages of all the involved users rank

Co-owner 2's and n's opinion is not post the photo and 3's opinion is post the photo.

After post the photo in social network 1's reputation with respect to 2, 3 and n

Reputation of 1 with respect to 2 = P12 − (S2r ∗ P(2))
Reputation of 1 with respect to 3 = P13 + (1 − S3r) ∗ P(3)

.....

Reputation of 1 with respect to n =P1n - (Snr*P(n)) Where,

P12=P13=. =P1n=P(1)

S2r Sensitivity of user 2 with respect to resource.

S3r Sensitivity of user 3 with respect to resource.

Snr Sensitivity of user n with respect to resource.

4) Steps of data sharing and updation

Step 1. Let X={xi|1≤i≤n} be 'n' users in the group.

Pi represent the page rank of it h users in the group G.

Step 2. Let Opinion ∈{-1,1}

Opinion = 1 if user reply yes

Opinion = -1 if user reply no

Step 3. Compute Threshold= $\frac{1}{n}\sum_{i=1}^{n} P_i$

Step 4. If $\sum_{i=1}^{n} P_i\ O_i$Threshold then sharing permitted else do not shared

Step 5. Reputation updation of users

Let out of x1, x2, x3 ,.....,xn users xi share the data

R(Pi j) = Pi-(Sjr * Pj) if Oj= -1

=Pi+((1-Sjr) * Pj) if Oj= 1

Where R(Pi j) is the reputation of ith user w.r.to jth user

Pij is the rank of ith user, Sjr is the sensitivity of r for jth user.

Oj is the opinion of jth user.

5) Flow chart for calculation of privacy loss

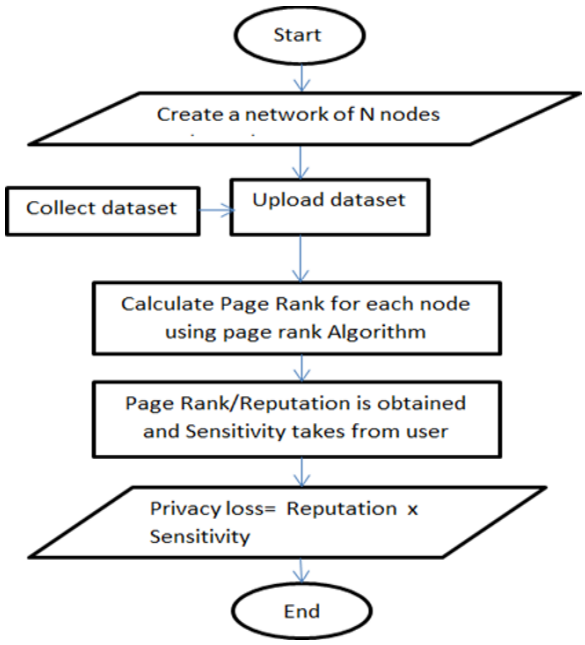

Fig. 6. Flow chart for calculation of privacy loss

The process of the calculation of privacy loss is as the following:

Step 1: Create a network of N number of nodes

Step 2: Collect the data set where you want

Step 3: Upload the data set in the network

Step 4: Calculate PageRank of every node in the network using PageRank algorithm

Step 5: When PageRank obtained for each node means Reputation obtained for each node

Step 6: Sensitivity takes from each user according to him how much data sensitive

Step 7: Find the privacy loss of each user using the formula

Privacy loss= Reputation x Sensitivity

3. Result & Discussion

A. Data Set Description

For our experiment, we utilized the Stanford Large Network Dataset Collection, specifically the musae-facebook dataset, as a proxy for the Facebook application to gather users' connectivity information. This dataset illustrates the connections between various nodes, where nodes represent users and links denote in-links and out-links. Node features are derived from site descriptions provided by page owners, summarizing the site's purpose. However, access to such information is contingent on individual privacy settings. The dataset, available through musae-facebook, is in .csv format. For our experiment, we focused on data related to social activity and profile attributes. This dataset comprises 22,470 nodes and 171,002 edges. Facebook, a social networking site, facilitates easy connections and sharing with family and friends. Users can share pictures, music, videos, articles, as well as their thoughts and opinions with a wide audience.

B. Graph Plotted on Sensitivity and Privacy Loss Values

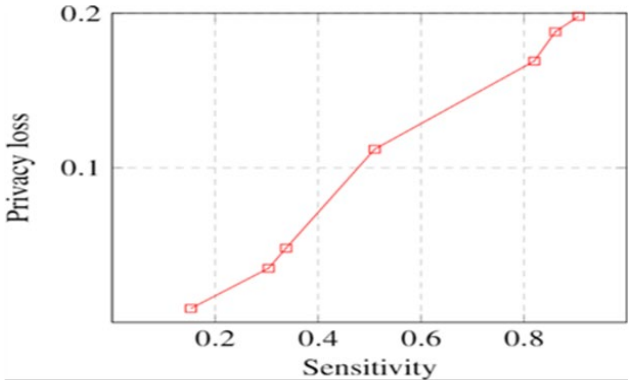

Fig. 7. Graph based on Sensitivity and Privacy loss values

Table 1
Sensitivity and their corresponding privacy loss values

Sensitivity	Privacy loss
0.153	0.009
0.304	0.035
0.338	0.048
0.51	0.112
0.82	0.169
0.861	0.188
0.906	0.198

Table 2
Difference of average privacy loss value and privacy loss

Sensitivity	Average Privacy loss- Privacy loss
0.153	0.0915
0.304	0.0655
0.338	0.0525
0.51	-0.0155
0.82	-0.0685
0.861	-0.0875
0.906	-0.0975

Average privacy loss = 0.1005

Here in the above table 1 shows the privacy loss value for different sensitivity values. For calculating privacy loss used the formula

Privacy loss = Reputation x Sensitivity

I have taken different sensitivity values between 0 and 1 and the reputation value automatically taken by the user between 0 and 1 and calculate the privacy loss. Using sensitivity and their privacy loss value draw the graph.

Using table 1 privacy loss values calculate the average privacy loss.

In table 2 subtract average privacy loss and privacy loss for different sensitive values and observe what will happen. When data sensitivity values less than 0.5 then difference of average privacy loss and privacy loss is large and When data sensitivity values greater than 0.5 then different of average privacy loss and privacy loss is low.

From above figures we concluded that

- When data sensitivity value is less than 0.5 then their privacy loss is less than average privacy loss.
- When data sensitivity value is greater than 0.5 then their privacy loss is greater than average privacy loss.
- When data sensitivity value is less than 0.5 then Difference of Average Privacy loss value and Privacy loss will be large.
- When data sensitivity value is greater than 0.5 then Difference of Average Privacy loss value and Privacy loss will be less.
- When data Sensitivity value increases then privacy loss chances will be more.

C. Graph Plotted on Reputation and Privacy Loss Values

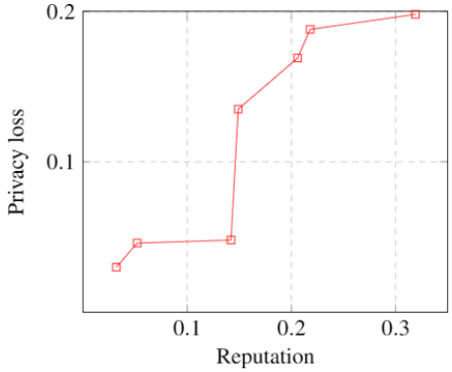

Fig. 8. Graph based on reputation and privacy loss values

Here in the below table 3 shows the privacy loss value for different reputation values.

For calculating privacy loss used the formula,

Privacy loss = Reputation x Sensitivity

Table 3
Reputation and their corresponding Privacy loss values

Reputation	Privacy loss
0.032	0.030
0.052	0.046
0.142	0.048
0.149	0.135
0.206	0.169
0.218	0.188
0.319	0.198

I have taken different reputation values between 0 and 1 and sensitivity value automatically taken by the user between 0 and 1 and calculate the privacy loss. Using reputation and their privacy loss value draw the graph.

From above figure we concluded that when Reputation value increases then privacy loss chances will be more.

4. Conclusion

This study explores privacy loss in Online Social Networks (OSNs) when co-owned data is shared. Before posting data, the owner seeks the opinion of co-owners, and data can be posted only if specific conditions are met. If co-owner privacy is violated, the owner's trust diminishes in relation to the co-owner. We aim to determine the owner's reputation before and after data posting, comparing it with other co-owners involved in that data. An increase in overall reputation after data posting indicates higher privacy loss chances, while a decrease suggests lower chances. Each node's PageRank value represents its reputation. Privacy loss is computed by considering different values of sensitivity and reputation. Increased data sensitivity and reputation values correspond to higher privacy loss chances.

References

[1] Lei Xu, Chunxiao Jiang, Nengqiang He, Zhu Han and Abderrahim Benslimane, "Trust-based Collaborative Privacy Management in Online Social Networks," 2018.

[2] V. K. Tuunainen, O. Pitk¨anen, and M. Hovi, "Users awareness of privacy on online social networking sites-case facebook," Bled 2009 Proceedings, p. 42, 2009

[3] G. Hogben, "Security issues and recommendations for online social networks," ENISA position paper, vol. 1, pp. 1–36, 2007

[4] D. Rosenblum "What anyone can know: The privacy risks of social networking sites," IEEE Security Privacy, no. 3, pp. 40–49, 2007

[5] Shaukat Ali, Naveed Islam, Azhar Rauf, Ikram Ud Din and Mohsen Guizani, "Privacy and Security Issues in Online Social Networks".

[6] M. Sahlabadi, R. C. Muniyandi, and Z. Shukur, "Detecting abnormal behavior in social network websites by using a process mining technique," Journal of Computer Science, vol. 10, no. 3, pp. 393–402, 2014.

[7] [7] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in Proceedings of the 2005 ACM workshop on Privacy in the electronic society, pp. 71–80, ACM, 2005.

[8] F. Wang and C. Yang, "An approach for setting access control rule based on multiparty collaborative in online social networks," in Information Science and Technology (ICIST), 2013 International Conference on, pp. 580–585, IEEE, 2013.

[9] J. Wortham, "More employers use social networks to check out applicants," The New York Times, vol. 20, 2009.

[10] H. Hu, G.J. Ahn, Z. Zhao, and D. Yang, "Game theoretic analysis of multiparty access control in online social networks," in Proceedings of the 19th ACM Symposium on Access Control Models and Technologies, New York, NY, June 2014, pp. 93–102.

[11] Jose M. Such, Member, IEEE, Natalia Criado, "Resolving Multi-party Privacy Conflicts in Social Media"

[12] H. Hu, G. J. Ahn, and J. Jorgensen, "Multiparty access control for online social networks: Model and mechanisms," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 7, pp. 1614–1627, July 2013.

[13] Gulsum Akkuzu, Benjamin Aziz, Mo Adda, "Advantages of Having Users' Trust and Reputation Values on Data Sharing Process in Online Social Networks".

[14] J. Pang and Y. Zhang, "A new access control scheme for facebook-style social networks".

[15] Kun Liu and Evimaria Terzi, "A framework for computing the privacy scores of users in online social networks," Ninth IEEE International Conference on, pages 288–297. IEEE, 2009.

[16] B. Carminati, E. Ferrari, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Semantic web-based social network access control," computers and security, vol.30, no. 2, pp. 108–115, 2011.

[17] H. Dubey and B. N. Roy, "An improved page rank algorithm based on optimized normalization technique," International Journal of Computer Science and Information Technologies, vol. 2(5), pp. 2183-2188, 2011.

[18] Wenpu Xing and A. Ghorbani, "Weighted PageRank algorithm," Communication Networks and Services Research, IEEE, pp. 305-314, 2004.

[19] Mary J Culnan and Pamela K Armstrong, "Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation," Organization science, 10(1):104–115, 1999

[20] Daniel J Solove, "Conceptualizing privacy," California Law Review, JSTOR, 2002.

[21] Samuel D Warren and Louis D Brandeis, "The right to privacy," Harvard law review, JSTOR, 4(5):193–220, 1890.

[22] Alan FWestin and Louis Blom-Cooper, "Privacy and freedom," volume 67, Atheneum New York, 197

[23] Ruth Gavison, "Privacy and the limits of law," Yale law journal, pp. 421–471,1980.

[24] Gulsum Akkuzu, Benjamin Aziz, Mo Adda, "Advantages of Having Users' Trust and Reputation Values on Data Sharing Process in Online Social Networks," United Kingdom, 2019.

[25] Julia Heidemann, Mathias Klier, Florian Probst, "Identifying key Users in online Social Networks: A PageRank Based Approach".

[26] Atul kumar Srivastava, Rakhi Garg, P.K. Mishra, "Discussion on Damping FactorValue in PageRank Computation," 2017.

[27] Ritu Sachdeva Research Scholar Head, Dr. Sachin Gupta Computer Science Dept. MVNU, Palwal, "A Literature Survey on Page Rank Algorithm".