Indian Journal of Science and Technology, *Vol 9(30), DOI: 10.17485/ijst/2016/v9i30/99069, August 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

An Enhanced Design for Anonymization in Social Networks

C. K. Shyamala* and S. Hemaashri

Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham (University) Amrita Nagar P.O, Ettimadai, Coimbatore 641112, Tamil Nadu, India; ck_shyamala@cb.amrita.edu, hemaashris@gmail.com

Abstract

Objectives: Our primary objective is to safeguard the privacy of the users by anonymizing the sensitive data shared by the users in Social networking sites. **Methods/Statistical Analysis:** The user's friends are grouped dynamically into various categories such as best friends, normal friends and casual friends based on their closeness with the user. The most private sensitive data is solely disclosed to best friends, while the sensitive data is anonymized using generalization and revealed to the normal friends. The last category of friends is exposed to only the least private information using another level of generalization. **Findings**: Currently, Social networking sites have become the rapid and preferred way of communicating with each other to share information. On one hand the options and benefits expand constantly, while on the other data privacy risks and sensitivity issues accumulate, eventually the privacy of the user is at grave, our work addresses this issue. The proposed design is resilient to Sybil attacks, where it restricts the revelation of the sensitive data of the user by anonymizing the shared data among various categories of friends. **Application/Improvement:** The performance of the system is enhanced by restricting the access to the sensitive data among various categories of friends based on their closeness degree.

Keywords: Automated Grouping, Data Analysis, Generalization, Privacy Protection, Sensitive Sentiments, Social Networking Sites

1. Introduction

Social Networking Sites (SNS) aids to connect people online, facilitates easier communication and creates a bonding among people, to develop and renew online relationship. Social networking sites act as a platform for communication and information sharing among people. Ever since the social networking sites had come into existence, they have been trying to bring people together to establish and maintain online relationships. It influences the lives of the users where the count of the web users also keeps booming day by day. Consecutively, it leads to a significant increase in continuous usage of social networking sites as mentioned in[1]. The significant increase in online information sharing and communication in turn resorts to various privacy issues. Subsequently, a significant thread of research focuses on the presence of various privacy issues in social networking. LeBlanc, D., and Biddle, R has presented a study regarding the perception of the users regarding the risks associated with online activities in[2]. Their study suggests that the users are aware about the risk factors involved in financially related online transactions. But the users fail to analyse the risks involved in social networking sites, which simplifies the work of the attackers. The author of[3] have also analysed the lack of awareness of the users and their attitude towards privacy protection and the problems related to privacy in social networking sites. This work studies the level of awareness of the users towards privacy protection in social networking sites. Based on the analysis done in this work, it is inferred that level of awareness among users is inconsistent which has brought a huge gap between the current privacy practice and the user's awareness towards privacy. This variation in perceived privacy and practiced privacy

**Author for correspondence*

leads to privacy leaks. As the facilities and options for sharing the information gets increased in social networking sites and due to lack of awareness of the users' towards privacy protection, the users' may add strangers to their friends' circle. Consequently, the users may tend to share their sensitive data to friends of friends and strangers. They may not be necessarily a trustworthy node as suggested in[4]. This has given rise to wavering of trust. Trust plays a significant role in online information sharing and communication. In order to enhance and protect the privacy of the user, Zhou, L analyses various levels of trust in social networking sites, by obtaining the trust ratings from other users. The assessment of trust is mandatory to ensure security in social networking sites. Consequently, the methods and the problems that occur in quantifying trust is addressed in[5]. Their study discusses the issues related to modelling trust and trust management in social networking. As a step further to this, various types of threats in social networking sites are identified in[6], where trust and privacy related threats, social and identity theft related threats are addressed[7]. The focus of the underlying problem lies in analysing various methods to mitigate privacy threats and to improve trust and privacy protection. The authors of [7] have presented an overview of the privacy risks prevailing in the social networking sites and have illustrated suggestions to safeguard the privacy of the users.

Privacy of the user can be protected by creating groups to categorize friends based on the trust levels. The users are given the privacy to create groups and segregate their friends accordingly in various groups as mentioned in[8]. But this adds as an extra overhead to the user, where the user created groups remain dormant most of the time. A Recommendation method for grouping friends is suggested in[9] and various data mining methods are suggested to classify their significant friends based on their relationship and trust level in[10,11]. Though these techniques for classifying friends do exist, an automated grouping system is proposed in[12], where friends are classified automatically based on their trust levels as best friends, normal friends and casual friends. The system protects the privacy of the user by segregating friends as best friends, normal friends and casual friends providing varied access to the sensitive data of the user to each group of friends where the data is revealed to a particular group of friends based on their closeness level.

The privacy of the user is enhanced by forging a dynamic categorization of friends. A recommendation system has been proposed in[13] which is modelled to suit the social networking sites. New acquaintances may turn up to become malicious users where they tend to attack the sensitive data of the users. As an attempt to fix this issue, the system analyses the friend requests and provides suggestions to the user to accept or deny the friend request. The fake identities in the system as identified as Sybil nodes. They may turn up to be a serious threat to mitigate the privacy which may lead to the piracy of the sensitive data of the user. The Sybil nodes are identified and removed from the system as suggested in[14].

As a part of our initial research, we have done automatic grouping of friends, followed by a recommender system and have analysed the threats prevailing in the system. The sensitive data of the user is given varied level of access to his/her friends, based on their trust level. The levels of trust among the friends vary, this has led to the introduction of various anonymization techniques to protect the sensitive data. Greedy algorithm for anonymizing SNS data is discussed in[15] which provides structural anonymity with sensitive attribute protection. The sensitive user data is protected in[16] where the users are allowed to specify their privacy policy. A model is proposed in to protect the privacy of the user that uses cryptographic techniques to control privacy over shared data. The shared data is displayed in an encrypted format. The user gives keys as access to the friends whom he thinks he can trust but that is an overhead to the user.

Though cryptographic techniques do exist, privacy of the user can also be protected by anonymizing the user data using K-anonymity method. In k-anonymity, the data of the corresponding user remains undistinguished from rest of the users. Another method to protect the privacy of the user is discussed in[17], where the leakage of sensitive information may be prevented. It discusses the attacks that occur in K-anonymized dataset and proposes an approach called l-diversity to overcome it. The limitations of l-diversity are illustrated in[18] where the problem of attribute disclosure remains unsolved. As a step further, the study proposes a method of finding t-closeness, to safeguard the sensitive information of the user. The sensitive information of the user can also be protected by sensitive data analysis as illustrated in[19]. The analysis is done based on context sensitive tone lexicon learning mechanism for polarity identification. A system that performs sentimental analysis for text and emoticons as proposed in[20] is implemented using the user data which is posted. It gives results as neutral, positive or negative.

But as human language is versatile, a single meaning for a word is not a feasible study. A self-maintaining system for sentimental analysis is proposed in[21]. To enhance the privacy of the data, generalization system is proposed in[22] that desensitizes the sensitive text in a document. Another model to protect sensitive data by a method called generalization is suggested in[23,24] uses generalization to do sensitive data and geographical data analysis.

In this paper, a system to enhance the privacy of the user is proposed which is complimented by providing varied level of data revelation to his friends based on their level of closeness. The sensitive data of the user can be geographical data, data related to relationships or sentiments. The access to the user data varies for each of the user's friends based on their level of closeness with the user. Their closeness level is determined by performing automated grouping as discussed in, where they are classified based on their closeness levels as best friends, normal friends and casual friends. The proposed model uses sensitive data analysis to restrict the data access to various levels of friends. The best friends are given the benefit of complete access to the sensitive data which includes geographical, relationship and sentimental data. The geographical and sentimental data is generalized and revealed to normal friends. As a step further to this, the sensitive data is subjected to second level of generalization, where the restricted access to sensitive data is granted to casual friends as they are considered to remain as a mere acquaintance to the user.

2. Materials and Methods

2.1 Privacy Recommender and Implementer

A Privacy protection mechanism to restrict the sensitive data disclosure is introduced in the proposed scheme. The revelation of sensitive data of an individual is restricted as she/he prefers to hide the sensitive private data. As a part of our initial research work, we have proposed an automated dynamic grouping system, in an attempt to improve recommendation for social networks in[25]. The proposed scheme in[25] groups friends automatically as best friends, normal friends and casual friends based on the interaction history and closeness degree. The grouping system is interconnected to a recommender thereby providing a dynamic grouping system. The system detects and analyzes the activity of malicious users, the fake identities termed as Sybil nodes prevailing in the system are identified and removed from the social network[25] does not address the issues of data privacy, which is predominant in social networking.

In this paper, we have addressed the privacy issues of data sharing and have proposed a scheme to enhance the privacy of user. This is complimented by providing varied level of data revelation to his friends based on their level of closeness. When a SNS user uploads a data, in a social networking site, the data is processed by a natural language processor which segregates the words in the sentence. The words are analyzed to test the rate of sensitivity present in it. The sensitive data can be geographical data, factual data, data related to relationships or sentiments. The sensitive words are separated from the rest of the non-sensitive parts by sensitive data detector. It also identifies the sensitive words which are indirectly combined with other data. The sensitive data which is identified is stored separately, while the sensitive data of the user cannot be revealed to all its friends. They cannot be hidden or anonymized as well. Level of privacy is in demand. The focus of the work is to provide restricted access to the sensitive data. The sensitive data should be revealed to friends based on their closeness level. The friends are classified based on their relationship status and closeness level as best friends, normal friends and casual friends as mentioned in[25]. The close friends are given higher priority, whereas the casual friend is given the least priority in terms of privacy. As mentioned in Table 1 the best friends are given the privilege of complete access to the sensitive data which includes geographical, relationship, factual and sentimental data. The geographical, factual, relationship and sentimental data is generalized and revealed to normal friends. As a step further to this, the geographical and sentimental data which is generalized is subjected to another level of generalization and revealed to the casual friends. The restricted access to generalized, sensitive data

Table 1. Varied level of generalization for sensitive data

Best friends	Normal friends	Casual friends
Reached the movie theatre at Fun mall, Peelamedu	Reached the movie theatre at Peelamedu	Reached the movie theatre at Coimbatore
Having a yummy dinner at Sugam Hotels, R.S.Puram	Having yummy dinner at R.S.Puram	Having yummy dinner at Coimbatore

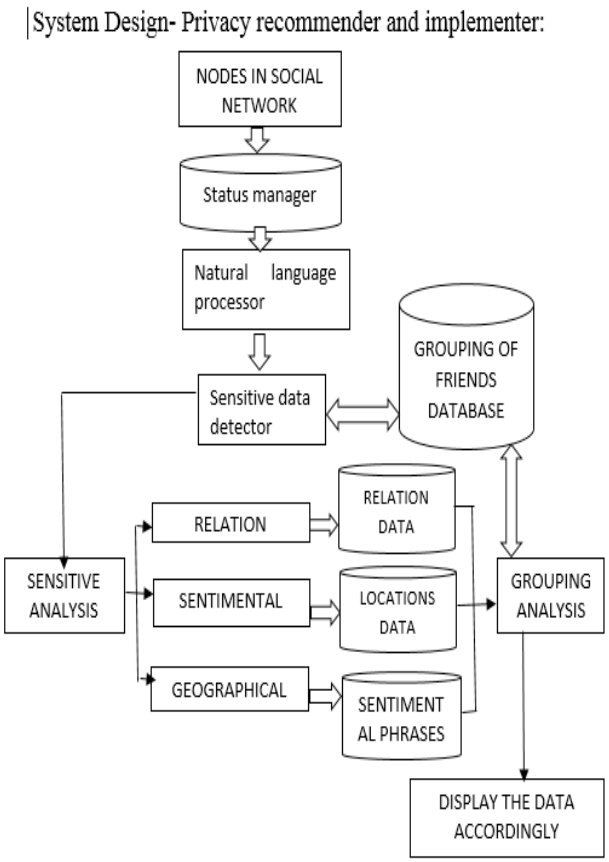

Figure 1. Privacy recommender and implementer

Algorithm for privacy recommender and implementer:

```
For each user
{
Step 1: Obtain the status updates from the interaction database
Step 2: Separate the words in the texts using natural language processor
Step 3: Analyse the sensitive words and separate it
Step 4: Obtain the sensitivity type of the separated words
Step 5: For every sensitive word
If (sensitive word is geographical data or sentimental data)
 {
        If (friend level == CDB)
                Display the sensitive data
        Else if (friend level == CDN)
                Generalise the sensitive data to first level
    Else if (friend level == CDC)
                Generalise the sensitive data to second level
 }
Else If (sensitive word is relationship data or factual figures)
 {
        If (friend level == CDB)
                Display the sensitive data
        Else if (friend level == CDN)
                Generalise the sensitive data to first level
    Else if (friend level == CDC)
                Generalise the sensitive data to second level
 }
```

Figure 2. Algorithm for Privacy recommender and implementer

CD_B- Closeness degree of best friend
CD_N- Closeness degree of normal friend
CD_C- Closeness degree of casual friend

Figure 3. Parameters of Privacy recommender and implementer

is granted to casual friends as they are mere acquaintances to the user. The system design for privacy recommender and implementer is illustrated in Figure 1. The algorithm and the parameters for privacy recommender and implementer are detailed in Figure 2 and 3 correspondingly.

3. Results and Discussion

Five hundred nodes in a social networking site is considered for simulation from[25]. As a part of our initial research work in[25], friends are grouped automatically based on the closeness degree, relationship status and interactions. The closeness degree is quantified for the friends where they are grouped as best friends, normal and casual friends. The automated grouping is enhanced by a recommender system.

In this paper, privacy of the data shared between a user and his/her friends is in focus. Restricted access of the sensitive data of the user is provided to normal and casual friends by the proposed privacy recommender and implementer. The sensitivity of the data is quantified, analysed and segregated as geographical data, factual data, relationship and sentimental data using a sensitive data detector. Based on the sensitivity type of the data and the closeness degree with the friends, data is revealed to best friends, normal and casual friends with increased privacy. The result of the simulation proves that the proposed privacy recommender system effectively reduces the sensitive data leakage that occurs in the social network. The decreased degree of sensitive data leakage proves the efficacy of the system.

3.1 Varied Level of Generalization for Geographical Data

The status updates uploaded by the SNS users in the social networking sites are obtained from the interaction database[25]. The status updates are segregated as various types based on their sensitivity level. As illustrated in Table 2 the geographical data which specifies the exact location at "Fun Mall" is solely revealed to the best friends, whereas it

is generalized to only reveal the area, "Peelamedu", where Fun Mall is located. This generalized location is displayed to normal friends. The location is further generalized to casual friends where "Peelamedu" is replaced with "Coimbatore".

3.2 Varied Level of Generalization for Relationship Data

The sensitive data which has the information regarding the personal details including the relationship details of the user is identified and segregated as relationship data. As mentioned in Table 3 the relationship data ("best friend") is wholly revealed to the best friends, whereas it is generalized as "friend" and revealed to normal and casual friends.

Table 2. Varied level of generalization for geographical data

Best friends	Normal friends	Casual friends
Chilling at an ice cream parlour with my best friend	Chilling at an ice cream parlour with my friend	Chilling at an ice cream parlour with my friend
Had the best time with my mom yesterday	Had the best time with my family yesterday	Had the best time with my family yesterday
Went outing with my college friends.	Went outing with my friends.	Went outing with my friends.

Table 3. Varied level of generalization for relationship data

Type of sensitive data	Best friends	Normal friends	Casual friends
Geographical	Zeroth level of generalization	First level of generalization	Second level of generalization
Relationship	Zeroth level of generalization	First level of generalization	First level of generalization
Sentiments	Zeroth level of generalization	First level of generalization	Second level of generalization
Factual figures	Zeroth level of generalization	First level of generalization	First level of generalization

3.3. Varied Level of Generalization for Sentimental Data

The sensitive data which specifies the personal details regarding emotions are segregated as sentimental data. As illustrated in Table 4, "crying" is replaced as "worrying", which is further generalized as "sad" as the user may not feel comfortable to reveal his/her personal emotions as such to his normal and casual friends.

3.4 Varied Level of Generalization for Factual Data

The personal data which precisely denotes the duration or the time period is generalized to reduce the level of sensitivity in it. As illustrated in Table 5 the precise time "6 pm" is made visible to best friends, where "6 pm" is generalized as "evening" and it is displayed to normal and casual friends.

Restricted revelation of geographical data is achieved, where the precise location is revealed to best friends, generalized location information is revealed to normal friends, the location information is further generalized and revealed to casual friends. Considering the relationship data, the information regarding first level relations are revealed to best friends. The relationship informa-

Table 4. Varied level of generalization for sentimental data

Best friends	Normal friends	Casual friends
I had been crying all day after watching the movie.	I had been worrying all day after watching the movie	I had been sad all day after watching the movie
I feel so blessed to receive this award.	I feel so glad to receive this award.	I feel so happy to receive this award.

Table 5. Varied level of generalization for factual data

Best friends	Normal friends	Casual friends
Waiting to attend the party at 6 pm	Waiting to attend the party in the evening	Waiting to attend the party in the evening.
The films released during 1960's always remain memorable.	The films released during old times always remain memorable.	The films released during old times always remain memorable.

User Id	Best friend	Normal friend	Casual friend
1	I feel excited	I feel delighted	I feel happy
25	Happy to be at ooty	Glad to be at Hill station	Satisfied to be at Nilgiris
73	Wow,party starts at 9 am	Good,party starts at night	Yes,party starts at night
100	I am deppressed now	I am worried now	I am sad now
170	Great to have you Ramya	Good to have you my friend	Happy to have you my friend
230	Last week,movie released	Currently,movie released	Currently,movie released
287	Finally I saw Ravi	Finally I saw my friend	Finally I saw my friend
345	I am overwelmed	I am excited	I am happy
376	Trip with my mom	Trip with my family	Trip with my family
438	He was murdered last week	He was murdered recently	He was murdered recently
492	My best friend is annoyed	My friend is irritated	My friend is sad

Figure 4. Status of data shared with privacy

tion is generalized where, they are revealed as second level relations to normal and casual friends. Considering the sentimental data, the intense sentimental words are revealed as such to close friends, the generalized sentiments are revealed to normal friends, the sentiments are further generalized and revealed to casual friends. Factual figures or numbers are also generalized and revealed to normal and casual friends.

Based on the obtained results, the status of the data shared among various levels of friends is depicted in Figure 4. The status updates shared by the users in SNS is generalized and revealed to normal friends and casual friends. From Figure 4 it is evident that the restricted access to the shared data is provided to the friends based on their closeness level.

4. Conclusion

In our initial research work in[25], an automated dynamic grouping system was proposed which group friends automatically based on their closeness levels and provides recommendations to the new acquaintances. As an attempt to enhance the privacy of the sensitive data shared in the automated grouping system, we have proposed a privacy recommender and implementer for social networking sites. In this paper, we propose an enhanced design for anonymizing the sensitive data of SNS users. Our primary objective is to safeguard the privacy of the users by restricting the revelation of sensitive data. The access to the sensitive data is restricted and revealed to various categories of friends based on their closeness degree. The simulation results indicate the efficiency of the system in terms of privacy of data shared in SNS.

5. Acknowledgment

I would like to thank my Guide Dr. C.K. Shyamala, Assistant Professor, Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham, Coimbatore for her guidance, reviews and valuable comments which helped a lot in writing this paper.

6. References

1. Li Y. Survey on situation of Chinese college students choosing to use social networking. In: Computer Research and Development (ICCRD), 2011 3rd International Conference; 2011 Mar. p. 344-348.
2. LeBlanc D, Biddle R. Risk perception of internet-related activities. In: Privacy, Security and Trust (PST), 2012 Tenth Annual International Conference; 2012. p. 88-95.
3. Zamzami IF, Olowolayemo A, Bakare KK, Kind DA. Sensitivity to online privacy in social networking sites. In: Information and Communication Technology for the Muslim World (ICT4M), 2010 International Conference; 2010 Dec. p. B-21-26 .
4. Nagle F, Singh L. Can friends be trusted? Exploring privacy in online social networks. In: Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances; 2009 Jul. p. 312-315.
5. Hoens TR, Blanton M, Chawla NV. A private and reliable recommendation system for social networks. In: Social Computing (SocialCom), 2010 IEEE Second International Conference, IEEE; 2010 Aug. p. 816-825.
6. Al Hasib A. Threats of online social networks. IJCSNS International Journal of Computer Science and Network Security. 2009 Nov; 9(11):1-6.
7. Fire M, Goldschmidt R, Elovici Y. Online social networks: threats and solutions. Communications Surveys and Tutorials, IEEE. 2014 Jan; 6(4):2019-2036.
8. Ur B, McGrath R. Grouping Friends for Access Control in Online Social Networks, p. 1-17.
9. Zhe Z, Li Z. A method of visualizing friend relations and recommending groups in online social network. In: Fuzzy Systems and Knowledge Discovery (FSKD), 2012 9th International Conference; 2012 May. p. 836-839.
10. Tanbeer SK, Jiang F, Leung CK, MacKinnon RK, Medina IJ. Finding groups of friends who are significant across multiple domains in social networks. In: Computational Aspects of Social Networks (CASoN), 2013 Fifth International Conference; 2013 Aug. p. 21-26.

11. Zhou L. Trust based recommendation system with social network analysis. In: 2009 International Conference on Information Engineering and Computer Science; 2009 Dec. p. 1 – 4.
12. Qian C, Xiao X, Chen S, Wang X. Grouping friends to improve privacy on Social Networking Sites. In: Conference Anthology; 2013 Jan. p. 1-6.
13. Can AB, Bhargava B. Sort: A self-organizing trust model for peer-to-peer systems. Dependable and Secure Computing, IEEE Transactions. 2013 Jan; 10(1): 14-27.
14. Samuel SJ, Dhivya B. An efficient technique to detect and prevent Sybil attacks in social network applications. In: Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference; 2015 Mar. p. 1-3
15. Shishodia MS, Jain S, Tripathy BK. GASNA: greedy algorithm for social network anonymization. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; 2013 Aug. p. 1161-1166.
16. Wang Z, Liao J, Cao Q, Qi H, Wang Z. Friend book: a semantic-based friend recommendation system for social networks. Mobile Computing, IEEE Transactions. 2015 Mar; 14(3):538-551.
17. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD). 2007 Mar; 1(1) .
18. Li N, Li T, Venkata Subramanian S. t-closeness: Privacy beyond k-anonymity and l-diversity. In: Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference; 2007 Apr. p. 106-115.
19. Babour A, Khan JI. Tweet Sentiment Analytics with Context Sensitive Tone-Word Lexicon. In: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE Computer Society; 2014 Aug. p. 392-399 .
20. Akaichi J. Social Networks' Facebook' Statutes Updates Mining for Sentiment Classification. In: Social Computing (SocialCom), 2013 International Conference; 2013 Sep. p. 886-891.
21. Bahrainian SA, Liwicki M, Dengel A. Fuzzy Subjective Sentiment Phrases: A Context Sensitive and Self-Maintaining Sentiment Lexicon. In: Proceedings of the 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), IEEE Computer Society; 2014 Aug 1. p. 361-368.
22. Anandan B, Clifton C, Jiang W, Murugesan M, Pastrana-Camacho P, Si L. t-Plausibility: Generalizing Words to Desensitize Text. Transactions on Data Privacy. 2012 Dec; 5(3):505-534.
23. Nguyen-Son HQ, Nguyen QB, Tran MT, Nguyen DT, Yoshiura H, Echizen I. Automatic anonymization of natural languages texts posted on social networking services and automatic detection of disclosure. In: Availability, Reliability and Security (ARES), 2012 Seventh International Conference; 2012 Aug. p. 358-364.
24. Kataoka H, Watanabe N, Mizutani K, Yoshiura H. DCNL: Disclosure Control of Natural Language Information to Enable Secure and Enjoyable E-Communications. In: U-and E-Service, Science and Technology; 2009 Dec. p. 131-140.
25. Shyamala CK, Hemaashri S, Swetha R. An improved recommendation system for social networks. In: International Journal of Control Theory and Applications. (ICSCS). IJCT A International Science Press. 2015; 8(5):1903-1910.