RESEARCH ARTICLE OPEN ACCESS

CFPRS: Collaborative Filtering Privacy Recommender System for Online Social Networks

Bisan A. Alsalibi *, Nasriah Zakaria **
Department of Computer Science, Universiti Sains Malaysia, Penang, Malaysia

ABSTRACT
Social-networking sites (SNSs) are known to be among the most prevalent methods of online communication. Owing to their increasing popularity, online privacy has become a critical issue for these sites. The tools presently being utilized for privacy settings are too ambiguous for ordinary users to understand and the specified policies are too complicated. In this paper, a collaborative filtering privacy recommender system is proposed. The implementation of the system was initiated by examining the users' attitudes toward privacy; whereby the most significant factors impacting users' attitudes towards privacy were determined to be location, religion and gender. The next step involved the classification of the users into various groups on the basis of the above factors. The paper presents a method of integrating the identified factors into the collaborative filtering algorithm to improve the filtering process. The evaluation of results reflects the accuracy of recommendations and proves that the use of the clustering model assisted the CF recommender in its creation of appropriate recommendations for each user.

***Keywords* -** Social Networks, Privacy, Collaborative Filtering, Recommender Systems, Privacy Factors.

I. INTRODUCTION

Social-networking sites such as Facebook and Twitter have recently become one of the most remarkable modes of communicating online. The common purposes of such sites are to connect with new friends who share common interests, find old friends, find new job opportunities, receive and provide recommendations, and much more. The potential exposure of posting personal information on such sites can lead to different types of privacy risks, including identity theft and stalking [1]. The use of SNSs has continued to grow remarkably quickly, coupled with considerable shifts in the way users interact with them. SNSs were, in the past, used simply as a tool for communication and entertainment, but now they are incorporated into every aspect of the daily lives of millions of users, affecting the way they communicate with each other and do business. The increasing popularity of online social networks and people's extensive adoption of them has raised a variety of privacy concerns.

The first privacy problem is that social-networking sites do not adequately inform their users about the risks of revealing their private information online. Although the issue of online privacy has become a subject of discussion in the media, these privacy issues are still not considered significant for many users [2]. Users of SNSs are apparently unwilling to consider that they might encounter risks as a result of their activities on SNSs. Even if they want to protect their privacy, with too much data and too many friends, it is very difficult for them to control who can see the activities on their profile pages. The second problem is that even though users of SNSs can control access to their own profile, they cannot control what others view. It is possible to pass on information inadvertently, or for personal information to be posted without one's permission. For example, a user can upload an embarrassing photo of a friend; this photo can also be tagged directly to a friend's profile. Moreover, SNS service providers have unlimited access to users' data. With this enormous amount of information, there are many commercial opportunities for SNSs, such as selling personal data to third parties. The third problem is that privacy tools in SNSs are not flexible enough to protect user's data properly. For example, the current Facebook privacy setting GUI is considered to be too complex for many users [2].

Although some solutions have been proposed, the privacy problems cannot entirely be fixed. Therefore, this paper attempts to propose a privacy recommender system using a collaborative filtering technique that users can employ to scan their privacy level and provide them with recommendations and guidance to minimize privacy violations. Palestine, as a part of the Arab world and with a relatively homogenous culture and religion was chosen as the scope of this project. The main focus of this paper is to provide a mechanism for analyzing Palestinians' group patterns on Facebook based on the factors that affect their privacy settings, such as culture, religion, age and gender, and to, propose a privacy recommender system based on the identified factors and users' patterns.

This paper is organized as follows. Section 2 discusses privacy in SNSs in general. Section 3 presents related work. Section 4 concerns the collaborative filtering privacy recommender system, how it works and its implementation phases. Section 5 presents the results and evaluation of CFPRS. Finally, section 6 concludes the paper.

II. PRIVACY IN SOCIAL NETWORKS

The main reason for privacy issues in SNSs is users' indifference to, or lack of knowledge about, adjusting their privacy settings. A number of studies have demonstrated that, although a majority of SNS users seem to be aware of the availability of privacy settings and control, they rarely make use of such controls or change their default settings. Govani and Pashley examined users' awareness of privacy issues and of the existence of privacy settings and controls provided by Facebook. They found that the vast majority of users are aware of the possible risks in making their private information visible to the public (e.g., identity theft). Nevertheless, users feel comfortable enough to disclose their private information. Although almost all of them know how to limit the visibility of their private information, they do not take any action to do so [4]. SNSs must inform users about the consequences of their various activities while using SNSs and about which part of their information is accessible and for whom [3]. Furthermore, users need to have powerful and easy-to-use tools that enable them to manage the way other people can access their information in a simple and flexible way that does not require a lot of time and effort [3].

III. RELATED WORK

A number of solutions have been introduced to tackle the problem of user's online privacy. One example is Privacy Suites, proposed by Bonneau [5]. This tool is designed for Facebook users and allows them to share and adopt another user's privacy settings. This approach saves users time and effort, but the privacy settings for one user may not be suitable for another user because privacy preferences depend on a range of factors such as the age, gender, job and cultural background. Govani and Pashley conducted a survey which revealed that awareness is not enough to guarantee privacy protection [4]. Although the users who participated in their survey cited stalking and identity theft as their main privacy concerns, they still reported posting their mobile numbers and real names on their profiles. This kind of survey may increase user awareness regarding privacy risks. However, surveys do not have a perceptible effect on user' behavior towards privacy issues. Fang et al. proposed a privacy recommendation wizard for Facebook privacy settings [6]. This tool employs data mining and machine learning methods, including active learning, to provide feedback regarding existing privacy settings. It classifies a user's friends into groups based on their degree of interconnection, then chooses a sample friend from each group and asks the user to determine what items he would like that person to be able to see. Based on the user's answers to these and other questions about what the user would choose to reveal to different "friends," the wizard suggests personalized privacy settings. This recommendation tool helps users adjust their privacy settings to match their actual privacy preferences, especially those who face some difficulties in dealing with Facebook privacy settings control. The learning process in this tool takes as input the answers given by the user about how he would set privacy for various friends and uses this information to predict how the user would set privacy settings for his remaining friends' list. The learning process is based on classifying the active user's friends into community members who are connected to each other. The drawback is that this classification mechanism cannot predict the recommendations for a new user who does not yet have any friend.

In another study, researchers observed that it is necessary to propose tools that can help SNS users understand the results of various privacy settings. Lipford et al. proposed and evaluated an "audience view" tool, which allows a user to view his profile as it appears to each of his friends [7]. This interface has been recently adopted by Facebook. However, although the audience view helps users understand and evaluate the correctness of their existing privacy settings, it does not help them in determining how they should adjust their settings in order to achieve a safe configuration. One approach to suggesting privacy settings to new users is proposed in [8], and the importance of good initial settings (due to users' tendency to keep them) is noted. This study presented a review of how to employ machine learning to recommend primitive privacy settings that are more likely to be useful for users. Another helpful aid for better privacy protection is to provide some informative metrics by which they can obtain accurate information. A model which employs data mining and AI tools can be used to test the level of difficulty experienced when attempting to access a user's information and to inform users about their privacy risk [9].

IV. COLLABORATIVE FILTERING PRIVACY RECOMMENDER SYSTEM (CFPRS)

One of the most widely used techniques in recommendation systems is collaborative filtering. Collaborative filtering (CF) is the process of providing predictions and recommendations (filtering) by collecting preferences from many users (collaborating). In order to provide users with appropriate recommendations that they can use to protect their privacy, we propose a privacy recommender system that employs a user-based collaborative filtering algorithm. User based CF uses a neighborhood-based algorithm, so called because the system deals with the user as if he belongs to a group of users sharing the same interests or the same items. These users are called neighbors. The system uses information about these neighbors to make its predictions about the user. The system begins by finding other users who share the same interests or items. Then it gathers all those users' ratings and begins computing predictions [12]. Our system is based on the assumption that similar users are more likely to have similar adjustments to their privacy settings. The determination of similarity is obviously a critical step in collaborative filtering. A Demographic-based CF algorithm is used to find users who share similar profile data, such as location, gender, religion, age, and education level. This is the filtering portion of collaborative filtering. In this paper, we enhanced the filtering process of the CF algorithm by identifying the most significant privacy factors those which have the strongest effect on users' behaviors with regards to privacy. Those identified factors will be used as a basis for determining the level of similarity between users. The implementation of the system has three parts: (1) identifying the main privacy factors that affect the user's behavior regarding privacy settings; (2) classifying users into clusters on the basis of the identified factors; and (3) integrating the identified factors and clusters into the filtering process of collaborative filtering to enhance the recommendations.

4.1 Privacy Factors

The behavior of users towards privacy settings for SNSs is affected by a number of factors. These factors can be classified into two categories: individual-level factors (e.g., age, gender, and educational level) and collective level factors (e.g., culture, religion). A number of studies have tested to what extent individual and collective factors affect the behavior of SNS users towards privacy settings. To identify the most influential factors among Palestinian users, a survey was conducted to collect data about how Palestinian users select their privacy settings in Facebook. The survey is divided into two sections. The first section solicits demographic information and personal data about the user. The second section concerns general attitudes towards risk and asks the user to select a preferred privacy setting for each piece of information in section one. Section one consists of 11 items: name, user name, age, location, gender, education, religion, email, mobile, languages and user' profile picture. Section two asks about privacy settings based on the four levels of privacy available in Facebook: Public, Friends, Friends of Friends, and Only Me. Public means that the user's profile can be viewed by every SNS user. Friends means that only people that the user has accepted as Friends can see his/her profile. Friends of Friends means that the user's Friends and people that their friends have identified as Friends can see his/her profile. Only Me means that the user's profile can be viewed by no one but the user him/herself. Each item in section one is then assigned a particular privacy level in section two. For instance, once the user chooses his/her gender in section one, he chooses the corresponding privacy level from section two according to who he would like to be able to view his gender. Survey results are based on online responses from 477 participants from four different locations: South Gaza, North Gaza, Ramallah and Bethlehem. Participants in the survey consisted of 239 males and 238 females of different ages, religions and education levels. In order to analyse the data; collected by the questionnaire and identify the key influencing factors, the Key Influencers detection process was used; this determined the key influencer factors in the output data by applying the naive Bayes algorithm to the questionnaire data. The naive Bayes algorithm is a classification method which is mainly based on the Bayes rule of conditional probability [10]. This method works by considering that all factors are equally independent and important for each other. It then starts to analyze them individually, gradually weighting each one. The results show that the most influential factors on privacy settings are location, religion and gender. These three identified factors were then used as inputs to the Microsoft clustering algorithm as discussed in the next section.

4.2 Clustering Process

The behavior pattern of users connected via social networks can help to predict the dynamics of the network system [11]. Understanding the behavior of Palestinian users towards privacy settings is a key requirement for the design of the collaborative filtering recommender system. In order to better understand users' attitudes and identify privacy settings' patterns among the users surveyed, the Microsoft clustering algorithm was used. The input data for the clustering algorithm came from the questionnaire described in Section 4.1. The resulting clusters form around the most influential factors, thereby identifying them. Eight different clusters emerged as a result of the clustering process, as shown in Table 1.

Table 1: The resulted clusters

Cluster	*Location*	*Religion*	*Gende r*
Cluster	North Gaza	Muslim	Femal
Cluster	North Gaza	Muslim	Male
Cluster	South Gaza	Muslim	Male
Cluster	South Gaza	Muslim	Femal
Cluster	North Gaza	Christian	Femal
Cluster	North Gaza	Christian	Male
Cluster	Ramallah	Muslim	Male
Cluster 8	Ramallah, Bethlehem	Christian	All

Based upon analysis of the survey responses, privacy trends and Palestinian users' behavior toward privacy settings can be clearly seen. Results show that the vast majority of users tend to hide their email address and mobile number by adjusting them to the Only Me privacy level. Furthermore, results show that the proposed influential privacy factors (location, gender and religion) have different effects on the way users adjusting their privacy settings. Location has a noticeable impact on privacy settings in the sense that; users who live in South Gaza are more concerned about privacy than those who live in North Gaza, Bethlehem, and Ramallah. South Gaza users tend not to disclose most of their private information, in contrast to users from other locations who tend to disclose most of their profile information. Gender has a significant effect on privacy settings in that female users are more likely to be concerned about their online privacy than male users. The effect of gender can be illustrated by comparing Cluster Three and Cluster Four, since the two clusters have the same location and religion but different genders, as shown in Fig. 1 and Fig. 2. According to the results, we found that 51% of male users in Cluster 3 (South Gaza Muslim males) adjust their mobile privacy to Only Me, whereas 83% of female users from Cluster 4 (South Gaza Muslim females) adjust it to Only Me.

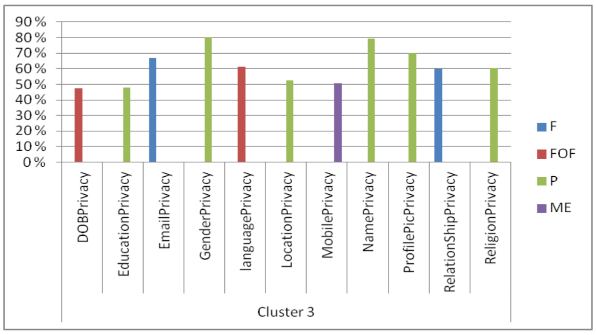

Fig. 1 Cluster 3 trend results

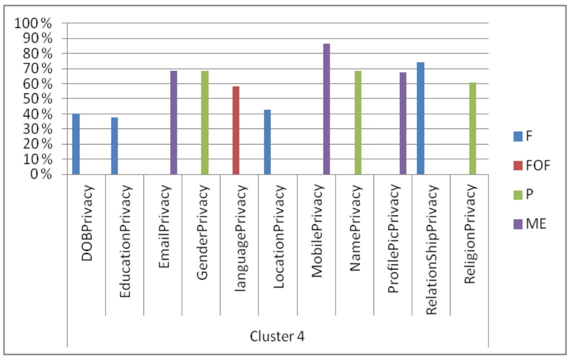

Fig. 2 Cluster 4 trend results

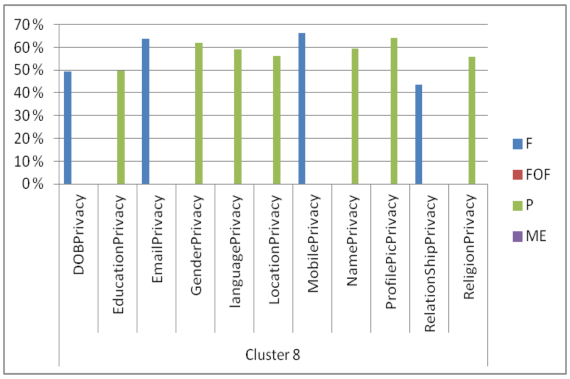

Fig. 3 Cluster 8 trend results

Results show that religion is another factor that influences the way people adjusting their privacy settings, in the sense that Muslim users are more concerned about their privacy than Christians, as seen in Fig.3. Fig. 3 shows the privacy preferences of Christian users (male and female) who live in Ramallah and Bethlehem, with regard to sensitive information such as mobile number, email, date of birth, and relationship status ; these items have been set to Friends privacy level by the majority of users. The majority of users set their remaining profile information to Public privacy level. It is notable that, there is no cluster either for Christians who live in South Gaza or Muslims who live in Ramallah. The recommender system will not be able to make recommendations to a new user if he belongs to a non-existent cluster. To overcome this limitation, the recommending system should also be able to provide recommendations for users based on expert opinions. In our system, we use expert recommendations based on the privacy framework proposed by [2].

4.3 The Impact of Privacy Factors on Clustering Results

As noted above, the behavior of users toward privacy settings is mainly affected by three factors: location, gender and religion.

First, to isolate the effect of gender on user behavior toward privacy settings, the location and religion factors must be constant. Fig. 4 shows a comparison between Cluster One (female Muslims who live in North Gaza) and Cluster Two (male Muslims who live in North Gaza) in terms of privacy settings. The comparison considered the most sensitive profile information items, which are mobile number, email address, date of birth and relationship status. From Fig. 4, we can see that 54% of users belonging to Cluster One set email privacy to Only Me, whereas 55% of Cluster Two users set it to Friends. Cluster One users, 68% set mobile privacy to Only Me, whereas 52% of Cluster Two users set it to Friends. Cluster One users, 55% set date of birth privacy to Friends, whereas 32% of Cluster Two users set it to public. Among Cluster One users, 65% set relationship status privacy to Friends, whereas only 41% of Cluster Two users set it to Friends. All of these results demonstrate that female users are more concerned about privacy than male users. Female users tend to hide almost all their sensitive information while using SNS. Secondly, to illustrate the effect of religion on privacy settings, Fig. 5 shows a comparison between Cluster One (female Muslims who live in North Gaza) and Cluster Seven (female Christians who live in Ramallah and Bethlehem) in terms of how they adjust their privacy settings. As shown, about 55% of Muslims from Cluster One set email address to Only Me, whereas 56% of Christians from Cluster Seven set it to Friends. In Cluster One, 68% of users set the mobile privacy to Only Me privacy level, whereas 50%

of Cluster Seven set it to Friends. Less sensitive information such as date of birth and relationship status has a slight difference, as shown in Fig. 5. We can conclude that Palestinian Muslim users have more private and closed profiles than their Christian counterparts.

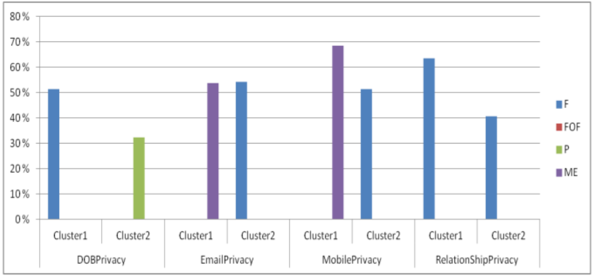

Fig. 4 Comparison between Cluster 1 and Cluster 2 in terms of gender effect

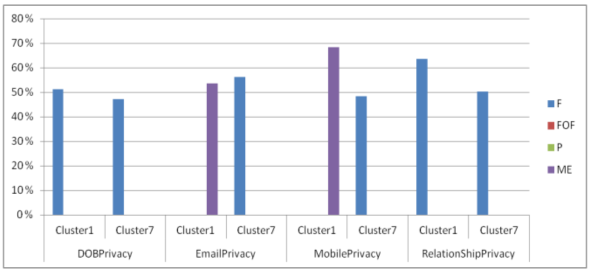

Fig. 5 Comparison between Cluster1 and Cluster 7 in terms of religion effect

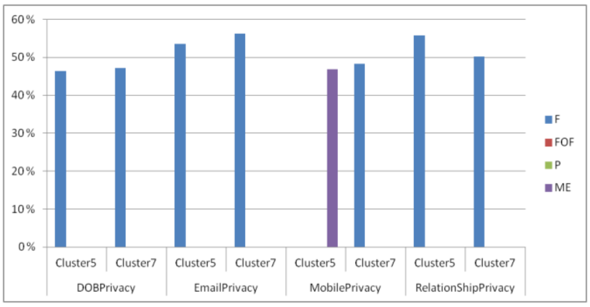

Fig. 6 Comparison between Cluster 5 and Cluster 7 in terms of location

Finally, to illustrate the effect of location on privacy, Fig. 6 shows a comparison between the users of Cluster Five (female Christians who live in North Gaza) and Cluster Seven (female Christians who live in Ramallah and Bethlehem) in terms of how they adjust their privacy settings. As shown, about 47% of users from Cluster 5 set mobile privacy to Only Me, whereas 48% of users from Cluster 7 set it to Friends priv. Less sensitive information such as email, date of birth and relationship status has a slight difference, as shown in Fig. 6. It is notable that location affects the behavior of users who have the same religion and the same gender, implying that people's privacy preferences are strongly affected by the culture and traditions of their surrounding environment.

4.4 System Design and Implementation

To clarify how this recommender would work, suppose a social-networking site has a database full of the private information and privacy settings of its users. When a new user subscribes to this SNS, he first has to create a profile. After he fills in his personal information in the profile form, the system invokes the user-CF algorithm and starts looking for users who have profile information similar to him. The similarity computation is based on the proposed factors that were identified in Section 4.1:

- Location and culture: The collaborative filtering algorithm starts by finding users who have the same location in their profiles as the new user's original location not where he currently living, but where he grew up, and where his family, friends and traditions exist.
- Religion: Within this first group, the collaborative filtering algorithm next finds users whose profiles list the same religion as the new user.
- Gender: Finally, within those who list the same religion and location, the algorithm finds users who have the same gender.

After completing these stages, the filtering results are generated a list of users who have similar profile information as the new user. The system retrieves the privacy settings of this group of users in order to start the second stage. In the second stage, the system analyzes these users' settings to determine the best privacy settings for the new user, and gives them back to him as recommendations. For example, suppose that the system needs to recommend privacy setting for the email address of the new user. The system counts the number of filtered users who set their email address to Public and stores it in variable A; users who set their email address to Friends of Friends and stores it in variable B. users who set their email address to Friends, and stores it in variable C; and users who set their email address to Only Me, and stores it in variable D. After that After that, the system compares A, B, C and D, and chooses the maximum value – that is, the system recommends the user set his email address to the same privacy level as the majority of the filtered users. The system repeats this stage for all the remaining profile items, and then gives the user the final privacy recommendations.

A major problem limiting the effectiveness of user based collaborative filtering is the "cold start problem," meaning that the algorithm is unable to return results when the database is empty or almost empty. In our case, the privacy recommender system would fail if the SNS database had few or no existing users to analyze. In this case, we solved the problem by using expert opinion according to the privacy framework proposed by [2], as shown in Table 2.

Table 2: Recommended privacy based on expert opinions

Information/ Item	*Information Type*	*Sensitivity*	*Recommended Privacy setting*
Name	Identity	Poisonous	Custom (Best Friends)
User Name	Identity	Healthy	All Friends
Gender	Demographic	Harmless	All Friends
Age	Demographic	Harmless	All Friends
Location	Demographic	Poisonous	Custom (Best Friends)
Education	Demographic	Harmless	All Friends
Religion	Demographic	Harmless	All Friends
Relationship	Demographic	Harmless	Custom (Best, Normal and casual Friends)
Email	Identity	Poisonous	Custom (Best Friends)
Mobile	Identity	Poisonous	Custom (Best Friends)
Language	Demographic	Harmless	All Friends
Profile Picture	Identity	Harmful	Custom (Best, Normal and casual Friends)

V. CFPRS EVALUATION

Collaborative filtering evaluation measures how well and accurately the CF recommender system is accomplishing the task of providing appropriate recommendations. Evaluating the recommender system's algorithms is often difficult because researchers within this field tend to use different metrics when evaluating the results of their experiments. This lack of standardization makes it difficult to compare the various available algorithms and to determine the most suitable algorithm for a particular purpose. This problem was investigated by Herlocker et al. in his study about collaborative filtering evaluation metrics [13]. He said that choosing a suitable dataset for a particular algorithm is one of the most difficult issues when evaluating collaborative filtering algorithms because different algorithms may behave better or worse on different datasets. The effectiveness of the CF recommender system for privacy settings can be measured by the extent to which users are satisfied with the recommendations they are given, i.e. whether they feel that the recommended privacy settings are suitable for their religion, gender and culture. In addition to user satisfaction metrics, the performance of the CF privacy recommender system can be measured by its accuracy, or how close its recommended ranking is to the actual ranking given by the user to that item, the accuracy can be measured using the Mean Absolute Error MAE as illustrated in Eq. (1). The third metric which can be used in evaluating a CF recommender system is its precision percentage, which is the percentage of the correct predictions recommended by the system as shown in Eq. (2).

$$MAE = \sum_{i=1}^{N} \frac{|Pi-Ri|}{N} \qquad (1)$$

$$P = \frac{N_{rr}}{N_r} \qquad (2)$$

5.1 Data Collection

The performance of a CF recommender system can be measured by splitting a user-rating dataset into two sets, a training dataset and a testing dataset. The error is measured on the testing dataset predictions after the algorithm has been fed with the training ratings. For this study, the training dataset consists of 477 users and the testing dataset consists of 80 users. Ten users per cluster were asked to evaluate the system. In our study, we used convenience sampling. Convenience sampling is a method in which the researcher chooses samples because of accessibility [14].

To collect data, we placed posts on Facebook seeking help with our study; the posts included a link to our online questionnaire. We also emailed friends of different locations, gender and religions asking them to complete our questionnaire. We used non-probabilistic sampling, which does increase the level of sampling error. However, we argue that this is an inevitable problem that researchers face when doing research within a short time-frame, as demonstrated by the fact that numerous researchers have used this method in published articles in peer-reviewed journals, as previously exemplified.

5.2 Evaluation Results

The experimental results of evaluating our CFPRS are shown in Table 3. The table includes the evaluation results of each Cluster, and the final row gives the average evaluation results for the whole system. Recall that the CFPRS has 4 privacy levels (Public, Friends of Friends, Friends, and Only Me) and 11 personal information items. Each privacy level has its own associated rank, from 1 to 4. In this case, according to Eq. (1), the maximum MAE is 3. The lower the MAE, the more accurately the recommendation engine predicts user privacy settings. The average mean absolute error of the CF recommender system is 0.217045463 out of 3, which reflects a high accuracy of the privacy settings recommended by the system. The average user rate of the CF recommender system is 4.5875 out of 5; this reflects to what extent the users are satisfied with the recommendations given by the system. The average precision of the system is around 92.15, which indicates the ability of the CF recommender system to generate appropriate privacy setting recommendations which are close to users' desired settings. It is noticeable from Table 3 that Cluster Eight (Christians who live in Ramallah and Bethlehem) has the lowest MAE and the highest precision. In contrast, Cluster Five (female Christians who live in North Gaza) has the highest MAE and the lowest precision.

Table 3: Evaluation results

N0.	*MAE*	*User Rate*	*Precision*
Cluster 1	0.1636	4.7	92.7272
Cluster 2	0.2181	4.6	93.6363
Cluster 3	0.3181	4.5	90.9090
Cluster 4	0.1181	4.8	95.4545
Cluster 5	0.3363	4.4	85.4545
Cluster 6	0.2090	4.5	93.6363
Cluster 7	0.2727	4.4	89.0909
Cluster 8	0.1	4.8	96.3636
Overall	0.2170	4.5875	92.15909

It can be clearly seen from Table 3 that we obtained MAE values very close to 0 (ranging from 0.1-0.33). This means that the CF recommender system is very accurate and provides appropriate recommendations to the users. Although the MAE of some clusters, such as Three and Five, could be considered a little high compared to other clusters, this does not necessarily mean that the recommendations given to users in those clusters are not good. The MAE is an inadequate measure in some cases since it focuses exclusively on the accuracy of the predictions without consideration for their influence on the user's decisions [13]. To overcome this limitation, precision and user satisfaction metrics are used.

It is evident from Table 3 that Cluster Eight (Christians who live in Ramallah and Bethlehem) and Cluster Four (female Muslims who live in South Gaza) have the lowest mean absolute errors (0.1 and 0.118 respectively) and the highest precision and user satisfaction (96.36%, 4.8 and 95.45%, 4.8 respectively). The high accuracy of the recommendations given to users of Clusters 8 and 4 is due to the privacy trend results for these clusters. The cluster results demonstrate that the vast majority of users belonging to Clusters 8 and 4 have similar privacy setting trends. As we can see in Fig. 7 and Fig. 2 respectively, about 65% of users have similar ways of adjusting their privacy settings. This similarity increases the probability that the active user (the user for whom the system is generating recommendations) will have a high degree of similarity to his cluster trend. The reason why Cluster Eight users have similar privacy setting opinions is due to the location and religion factors, in the sense that Bethlehem is full of Christians who have the same pure culture which has not been affected by the Muslim culture, in contrast to what happened to Christians in other parts of Palestine.

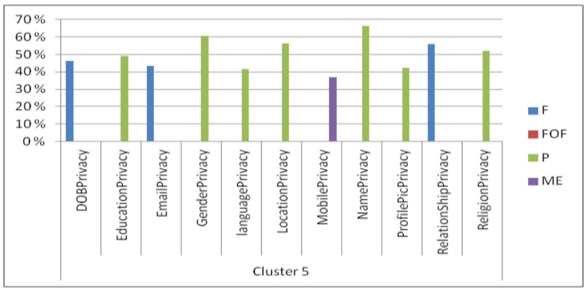

Fig. 7 Cluster 8 trend results

The same reasoning can be applied to Cluster 4, Muslim users who live in South Gaza, where female Muslims from South Gaza have many rules and religious restrictions that they all have to follow. In contrast to Cluster 8 and Cluster 4, Cluster 5 (female Christians who live in North Gaza) has the highest mean absolute error (0.336) and the lowest precision and user satisfaction (85.45% and 4.4 respectively). The low accuracy of the recommendations suggested by the CF recommender system for Cluster 5 users is due to the fact that Islamic cultural norms have a large impact on Muslims and the Christian minority alike.

The vast majority of North Gaza populations are Muslims which means that Islam is the dominant religion there. Despite being of a different religion, some Christians may be more likely to express greater observance to Islamic cultural norms due to their deferential social status in Gaza and some of them may refuse to accept these Islamic norms. This difference in the attitude of Christians has led to the contrast in behavior towards privacy settings.

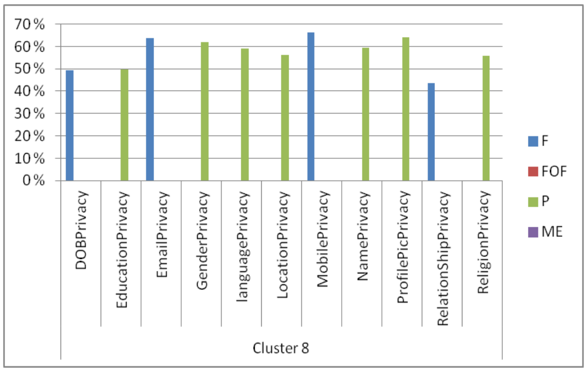

Fig. 8 Cluster 5 trend results

It is noticeable from Fig. 8 that users of this Cluster vary in the way they adjust their privacy settings, whereby less than 50% of users have the same settings for most of their information; this increases the probability that the active user will have dissimilar settings to what their Cluster suggests.

VI. CONCLUSION AND FUTURE WORK

Online privacy has emerged as an important problem in online social-networking sites. While these sites have recently become one of the most prevalent modes of online communication, the tools available for privacy settings are still difficult for ordinary users to understand and use. This paper proposed a CF privacy recommender system that would help every user to adjust his privacy settings according to the group of people and culture he belongs to. Users who are not aware of the importance of protecting their online privacy, as well as those who are aware but cannot use the complicated privacy tools, can both benefit from using the privacy recommender system.

The implementation phase of the system was initiated by understanding users' reactions towards privacy settings. During this stage, the factors that have the most impact on Palestinian users' attitudes towards privacy were identified. We identified three major factors, which are location, religion and gender. The next step was to classify Palestinian users into different groups based on the three proposed factors. Users from each group have similar ways of adjusting their privacy settings. Finally, we have presented an approach to integrate these identified factors and similarities into a collaborative filtering algorithm to implement a recommender system. The main goal of the integration step is to enhance the filtering process of collaborative filtering so that it can classify users efficiently according to their location, religion and gender.

To evaluate the CF privacy recommender system, it was tested by ten users from each Cluster. The evaluation results reflect the accuracy of the recommendations and prove that using the clustering model helps the CF recommender to generate appropriate recommendations suitable for each user. Such a CF privacy recommender system has the potential to become a powerful technology that can be used by many people across a wide range of SNSs.

In the future, we plan to conduct more studies on the factors that affect privacy to understand more deeply the ways in which each group of users adjust their privacy settings. For example, it would be interesting to enhance the location factor by extending the study to more locations in Palestine. Since the scope in this research was centered on Palestinian users, the study could be expanded in the future to include (for example) all Asian users. The generalization could be done by following the same procedure as in this study. The first step would be to conduct a survey to collect data and then identify the main privacy factors for those users. The second step would be to use the identified factors as the basis for the classification process in order to identify user patterns. And finally, the classification results would be integrated with the collaborative filtering algorithm to implement the privacy recommender system.

In our approach, we chose a user-based CF wherein we seek a community of similar users who share the same location, religion and gender as the active user. The recommendations in our technique are based on the opinion of the majority of similar users. We are aware that the opinion of the majority is not true for everyone, because it is based on user background and not given by trusted users. In the future, we plan to integrate an expert-based CF with the user-based CF in order to improve the accuracy of the recommendations. A weighting scheme will be needed to balance the two methods. In this case, the selection of similar users will not only depend on similar demographic characteristics as the active user, but will also include expert users from the same community as the active user. The recommendations for that community should become more precise by virtue of these users being more knowledgeable about the best privacy settings for users in each community.

VII. ACKNOWLEDGMENTS

B. Alsalibi would like to acknowledge and extend her heartfelt gratitude to Eng. Abdulla El-Madhoun who has made the completion of this paper possible and for assisting in the implementation of the algorithm.

REFERENCES

[1] A. Acquisti, and R. Gross, "Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook", Privacy Enhancing Technologies, Volume 4258, 2006, pp. 36-58.

[2] A. Ho., A. Maiga, and E. Aïmeur, "Privacy Protection Issues in Social Networking Sites", *in 7th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-09)*, 2009, pp. 271-278.

[3] M. Soryani and B. Minaei, "Social Networks Research Aspects: A Vast and Fast Survey Focused on the Issue of Privacy in Social Network Sites", International Journal of Computer Science Issues, Vol. 8, Issue 6, No 3, November 2011, pp. 363-373.

[4] T. Govani, and H. Pashley, "Student Awareness of the Privacy Implications when Using Facebook", 2005.

[5] J. Bonneau, J. Anderson, and L. Church, "Privacy suites: Shared Privacy for Social Networks", *in international conf. proc. of the 5th Symposium on Usable Privacy and Security*, ACM, 2009, pp. 1-2.

[6] L. Fang, H. Kim, K. LeFevre, and A. Tami, "A privacy recommendation wizard for users

of social networking sites” *in CCS '10 Proceedings of the 17th ACM conference on Computer and communications security*, 2010, pp. 630-632.

[7] H. Lipford, A. Besmer, and J. Watson, “Understanding privacy settings in Facebook with an audience view”, *in Proceedings of the 1st Conference on Usability , Psychology, and Security*, 2008.

[8] E. Aimeur, S. Gambs, and A. Ho, “UPP: User Privacy Policy for Social Networking Sites”, *in Proceedings of the 2009 Fourth International Conference on Internet and Web Applications and Services (ICIW '09)*, IEEE Computer Society, Washington, DC, USA, 267-272.

[9] J. Staddon, “Finding "hidden" connections on LinkedIn An argument for more pragmatic social network privacy”, *in Proceedings of the 2nd ACM workshop on Security and artificial intelligence (AISec '09)*, 2009, ACM, New York, NY, USA, pp. 11-14.

[10] R. Caruana and A. Niculescu-Mizil, “An empirical comparison of supervised learning algorithms”, *in Proceedings of the 23rd International Conference on Machine Learning*, 2006.

[11] S. Phulari, S. Khamitkar , N. Deshmukh, P. Bhalchandra , S. Lokhande and A. Shinde, “Understanding Formulation of Social Capital in Online Social Network Sites (SNS)”, *International Journal of Computer Science Issues*, Vol. 7, Issue 1, No. 3, January 2010, pp. 92-96.

[12] T. Khoshgoftaar and X. Su, “A Survey of Collaborative Filtering Techniques”, *Advances in Artificial Intelligence*, 2009, pp. 1-20.

[13] J. Herlocker, J. Konstan, L. Terveen and J. Riedl, “Evaluating Collaborative Filtering Recommender Systems:, ACM Transactions on Information Systems (TOIS), 2004, 22(1), pp. 5-53.

[14] D. Anderson, D. Sweeney, T. Williams, J. Freeman and E. Shoesmith, “Statistics for Business and Economics. London: South-Western Cengage Learning”, 2009.