Prediction of Preferred Personality for Friend Recommendation in Social Networks using Artificial Neural Network

Nafis Neehal
Department of CSE
Islamic University of Technology
Gazipur, Dhaka
nafisneehal@iut-dhaka.edu

M. A. Mottalib
Department of CSE
BRAC University
Dhaka, Bangladesh
mottalib@bracu.ac.bd

***Abstract*—Social Networking sites have nowadays become the most common way to communicate over online for people around the world. For making friends in social network, there remains an underlying friend recommendation framework which suggests friends to the users. However, most of the existing friend recommendation frameworks consider only the number of mutual friends, geo-location, mutual interests etc. to recommend one person as a friend to another. But, in real life, people, who have similar personalities, tend to become friends to each other, application of which is completely missing in the modern friend recommendation frameworks. Hence, we have proposed a personality based friend recommendation framework in this paper, which consists of a 3-Layered Artificial Neural Network for friend preference classification and a distance-based sorted subset selection procedure for friend recommendation. Our model tends to achieve a fairly high precision, recall, f1-measure and accuracy of around 85%, 85%, 82% and 83% respectively in the friend choice classification task.**

***Index Terms*—friend recommendation, social network, artificial neural networks, classification**

I. Introduction

A. Overview

Social networks today made our life easier by helping us get connected quickly with people all around the world saving time and money [1]. However, there are many users in a social network who are facing lots of problems to interact with a large amount of data from the network and don't quite know what they are really interested about. A friend recommendation engine is hence needed to provide a good way to diminish this problem as well as satisfy user needs. A recommendation engine facilitates the users by helping them in making an informed decision based on the information they need, like item recommendations based on users' previous behaviour and the information on them collected earlier by the system.

Nowadays, in every major field playing pivotal role in the economy like e-commerce, entertainment etc. recommendation engines are being used [1]–[3]. Many of the rising social networks such as Facebook has their own recommendation engine for recommending friends to users. All though, both of the recommendation engines perform the same job of recommendation, the inherent nature of a recommendation engine used in social networks is way different than the one used in e-commerce or entertainment industry. [4].

However, there are many complex factors related while considering the relation between two person. Firstly, giving a description of items is much easier than describing a person's interest and hobbies. Secondly, with the course of time, a person's hobbies and interests can change. Thirdly, even today, a person adds another person as a friend in their social network account based on geo-location or mutual work/study experience and not based on having similar personality. This is how friend recommendation in social networks are influenced by the factors mentioned above. [5]

B. Research Goal and Contribution

The primary goal of a friend recommendation system in social network is to provide with the most relevant data to the user based on their requirement or demand. But now-a-days in social networks there are too much data leading to an overwhelming condition. For instance, if we take Facebook, it has a worldwide monthly active user of 1.26 Billion which is increasing by 15% every year [6].

Also there is the factor that Facebook or any other recommendation system does not allow users to choose the category of people they want to be friends with. In Facebook, friends are recommended based on people a user searches for, people who searched for the user, number of mutual friends, group affiliation and so on. And there is no way that users can personalize this recommendation criteria. So it becomes very hard to choose friends among this huge amount of people. Recommendation system typically helps people by narrowing down the choice domain. Different people have different agenda. Thats why developing a general recommendation system by adopting traditional methods to satisfy everyone is difficult [6], [7].

978-1-5386-9111-3/19/$31.00 ©2019 IEEE

Hence, in lieu of having quite a many methods for recommending friends in social networks, we have realized and addressed the lack of a proper recommendation method and proposed an idea of friend recommendation in social networks like Facebook based on similarity in personality between users. In our work we have preliminary proposed a framework for automated prediction of a user's probable choice of human personality for friendship. This predictive framework can be further extended to develop a friend recommendation system too.

Briefly, in our method, we have used a pre-computed survey which had classified human personality into 5 broad categories based on the answers to questions of a 50-item questionnaire about standard Big Five Personality from IPIP Pool. Besides the personality scores in each of the 5 classes for a sample user, the survey result also contained the type of personality the sample user wanted to be friend with among the aforementioned 5 human personality classes. We have trained a 3 layered Shallow Artificial Neural Network (ANN) to create the mapping between the user's personality trait to the user's choice of personality for friendship. After proper training, this trained model can be used on any unknown sample user's personality scores to predict his/her choice of friends and based on that a recommendation can be made.

II. Literature Review

A. Recommendation Systems

There are two methods for building a recommendation framework. [5].

1) Content based recommendation system: Content-based filtering method depends on a user's history of behaviour, such as the item ratings the user gave, the products the user browsed, and their history of purchasing goods. This model simultaneously establishes a model for user's behaviour and also a model for each item to describe its characteristics. After that, a cross-match is performed between the user model and item model by gauging how much similar they are to each other and then with enough similarity an item is recommended to the user.

2) Collaborative filtering based recommendation system: Collaborative filtering is different than content based recommendation. Collaborative filtering based recommendation recommends items for people based on users who are similar to them. In collaborative filtering, an item is recommended to a person if another user who is similar to him has recommended it. In collaborative filtering, we basically analyze relationships between dependencies among products and users and from that we try to identify new user-item.

B. Friend Recommendation Systems

In real life, people typically relies on the opinions or interests or recommendations that they get from their friends in social media before purchasing a product [13].

TABLE I
Literature Review

Type	Author	Contribution
Graph Based	Nowell and Kleinberg (2004)	recommended friends considering only the local features of a network graph [18]
	Symeonidis et al. (2010)	introduced transitive node similarity into the features of global graph [19]
	Patil (2009), Xie (2010), Scellato et al. (2011)	link prediction solved by considering the shared common interest among people [20]–[22]
CF Based Approach	Kautz et al. (1997)	expand user awareness by shared document [23]
	Chen et al. (2009)	compared social relationship based algorithms and content similarity for recommendation purpose [24]
	Guy et al. (2010)	aggregated social network information and highlighted three classes of social information [25]
	Ziegler and Golbeck (2009)	strong correlation of users similarities with social information [26]
	Liang and Li (2011)	hybrid system to recommend based on user interest and social information together [27]
	Agarwal and Bharadwaj (2011)	weighted measure of features that help users to connect. [13]

Hence, Social networking platforms can have the means to offer a shared platform for sharing interests, recommendations etc. and also can serve as a platform which provide proper incentives in marketing of products by modeling consumer behaviour (Chen and Qi 2011 [16]; Bonchi et al. 2011 [17]).

There are several mathematical models and methodologies available which shows how people interact with one another. [8]–[12]. In the following subsections, the contributions of various link prediction approaches are described which encompasses two major categories: CF-based approaches and graph-based approaches . A synopsis of some approaches is given in Table 1.

C. Personality Analysis

The Big Five personality traits, also known as the five factor model (FFM), is a taxonomy for personality traits [14]. The five factors are:

- Extraversion - Typically higher score in this trait means the user is outgoing/energetic in nature and solitary/reserved for vice versa.
- Neuroticism - Typically higher score in this trait means the user is sensitive/nervous in nature and secure/confident for vice versa.
- Agreeableness - Typically higher score in this trait means the user is friendly/compassionate in nature and challenging/detached for vice versa.

Test

Rating	I....	Rating	I.....
	1. Am the life of the party.		26. Have little to say.
	2. Feel little concern for others.		27. Have a soft heart.
	3. Am always prepared.		28. Often forget to put things back in their proper place.
	4. Get stressed out easily.		29. Get upset easily.
	5. Have a rich vocabulary.		30. Do not have a good imagination.
	6. Don't talk a lot.		31. Talk to a lot of different people at parties.
	7. Am interested in people.		32. Am not really interested in others.
	8. Leave my belongings around.		33. Like order.
	9. Am relaxed most of the time.		34. Change my mood a lot.
	10. Have difficulty understanding abstract ideas.		35. Am quick to understand things.
	11. Feel comfortable around people.		36. Don't like to draw attention to myself.
	12. Insult people.		37. Take time out for others.
	13. Pay attention to details.		38. Shirk my duties.
	14. Worry about things.		39. Have frequent mood swings.
	15. Have a vivid imagination.		40. Use difficult words.
	16. Keep in the background.		41. Don't mind being the center of attention.
	17. Sympathize with others' feelings.		42. Feel others' emotions.
	18. Make a mess of things.		43. Follow a schedule.
	19. Seldom feel blue.		44. Get irritated easily.
	20. Am not interested in abstract ideas.		45. Spend time reflecting on things.
	21. Start conversations.		46. Am quiet around strangers.
	22. Am not interested in other people's problems.		47. Make people feel at ease.
	23. Get chores done right away.		48. Am exacting in my work.
	24. Am easily disturbed.		49. Often feel blue.
	25. Have excellent ideas.		50. Am full of ideas.

Fig. 1. Questionnaire

- Conscientiousness - Typically higher score in this trait means the user is efficient/organized in nature and easy-going/careless for vice versa.
- Openness to experience - Typically higher score in this trait means the user is inventive/curious in nature and consistent/cautious for vice versa.

Scores in each of this 5 personality traits is calculated by the survey questionnaire presented in Figure 1. This questionnaire was collected from **International Personality Item Pool (IPIP)** which is considered as a standard for personality analysis. There are total 50 questions in this survey [15]. As an answer to each of this 50 questions, surveyee has to give a number between 1 to 5, while 1 means surveyee completely disagrees and 5 means complete agrees and 2,3 and 4 are in between. Although for our work we didn't perform the survey and used a pre-computed dataset instead, yet it is worth mentioning how it was done.

III. Methodology

The basic workflow diagram of our proposed framework is given in Fig 2. As already mentioned, we have deployed an Artificial Neural Network to predict the choice of friendship of social network users. As there are five classes of human personalities to choose from, one single user can choose to be friend with people from one single category or more than one. So eventually, the problem in hand gets transformed into a multi-label classification problem where the labels are the five human personalities. Once, a user's choice of friendship is predicted by the ANN, then users having those personality traits can be recommended as friends. The one big advantage of our system over the traditional friend recommendation methods used by giant social networks is users get to choose the type of people they want to be friend with, which is very much similar to what happens in our real life.

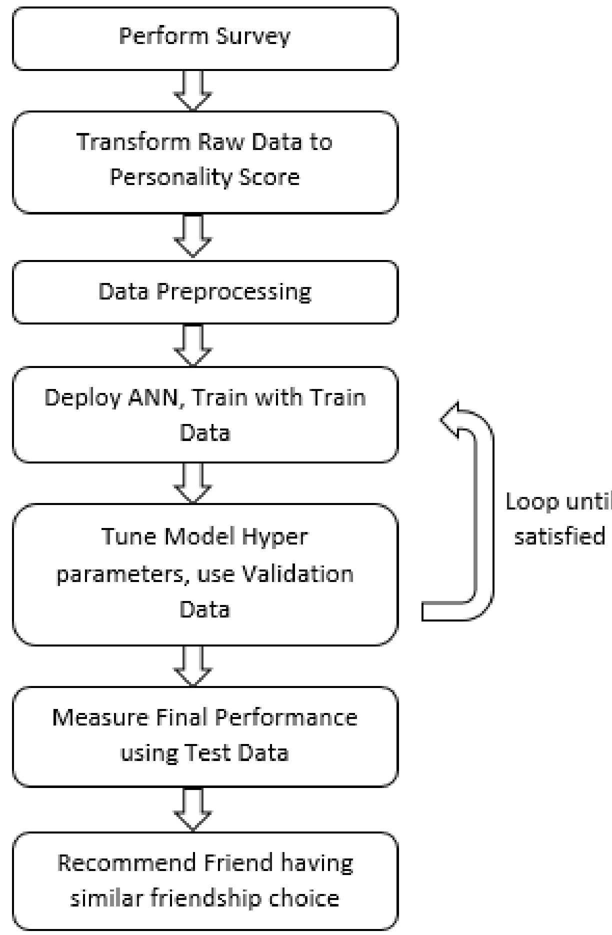

Fig. 2. Workflow Diagram

A. Data Collection and Description

In general, the first stage of our proposed method should be to perform a rigorous survey which typically consists of 50 standard question items in order to calculating the user's personality. From the answers to these 50 questions, we can calculate each user's personality score in each of the 5 categories. Besides the questions, the survey also contains a multiple-answer question about the choice of user's personality to be considered for friendship. The five personality categories are presented as options and the surveyee can choose either a single option or more than one options. In our work though, we haven't done any survey, nor did we calculate personality scores for each categories. Instead, we have used a pre-computed data set titled as "My Personality" data set (Released by faculties from University of Cambridge). This is the most popular social network users' personality data set available on

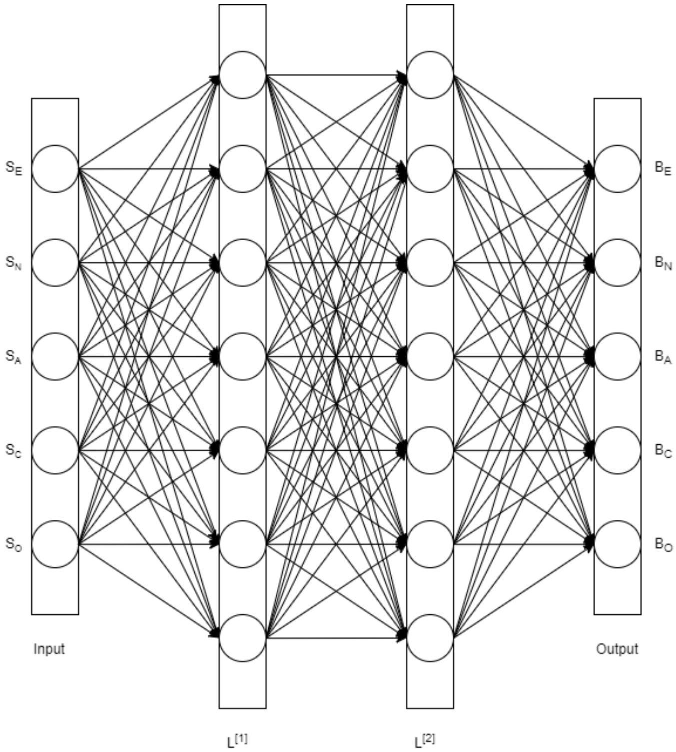

Fig. 3. Proposed 3-Layered ANN

the internet. The data we received had 9,917 samples, where each sample consist of personality scores in 5 personality domains, and choice of friendship of each user. It had total 5 features which are the 5 personality scores in each of the 5 personality categories and had total 5 boolean target variables that we have to predict for unknown data samples using our ANN, each having a value 1 if that sample user wanted to make friendship with people belonging to that category and 0 otherwise.

B. Data Preprocessing

Standard data pre-procesing like normalizing the features, find and fill out missing values etc were done on the pre-computed data set. Also, in the data set we received, the target variable values were presented as 'Y' standing for YES and 'N' standing for NO. We had to change those categorical values to numeric boolean values namely 0 for NO and 1 for YES.

C. Artificial Neural Network

Artificial Neural Networks, as we know it, is trivially a very powerful classifier. General structure of a neural network consists of several neurons stacked up in layers. There are 3 kinds of layers, namely - Input Layer, Hidden Layer and Output Layer.

In general, number of nodes or neurons in the input layer is equal to number of features used from the data and, number of nodes in the output layer is equal to number of target variables. The in-between layers are called hidden layers. Number of hidden layers and number of nodes in each hidden layer is generally chosen in an arbitrary manner, typically which is based on trial and error methods.

TABLE II
CHOICE OF NETWORK HYPERPARAMETERS

Number of Hidden Units	Overall Error	Convergence Time
2	3.2058	186 epochs
3	3.3084	894 epochs
4	3.2496	102 epochs
5	3.1726	96 epochs
6	3.0673	103 epochs
7	2.8847	117 epochs
8	2.8640	122 epochs
9	3.2831	102 epochs
10	3.2957	94 epochs
11	3.2543	96 epochs
12	3.2949	84 epochs
13	3.2699	77 epochs
14	3.2944	99 epochs
15	3.1887	87 epochs
16	3.2404	49 epochs

For our neural network, we chose to have 2 hidden layers, each having 7 nodes as this was the most optimal choice in terms of overall network error and convergence time (in epochs) as shown in Fig 3. This specific structure was chosen based on trial and error methods, result of which is shown in Table 2. Details of our network's structure is shown in Table 3.

TABLE III
NEURAL NETWORK STRUCTURE

Layer	Number of Nodes	Layer Dim	Weight Dim	Bias Dim	Activation
Input	5	5x1	None	None	None
Hidden 1	7	7x1	7x5	7x1	ReLU
Hidden 2	7	7x1	7x7	7x1	ReLU
Output	5	5x1	5x7	5x1	Sigmoid

In general, a neural network works in 3 stages - first the Feed-Forward stage where each sample is propagated through the network and the output is predicted, which is shown in Algorithm 1.

Algorithm 1 Forward Propagation in Neural Network

```
1: procedure FORWARDPROPAGATION(X)
2:     Z[1] = W[1].X + b[1]
3:     A[1] = ReLU(Z[1])
4:     Z[2] = W[2].A[1] + b[2]
5:     A[2] = ReLU(Z[2])
6:     Z[3] = W[3].A[2] + b[3]
7:     A[3] = Sigmoid(Z[3])
8: end procedure
```

The notations that have been used in forward propagation is described below:

- $Z^{[i]}$ = Output of i^{th} Layer of Neural Network
- $A^{[i]}$ = Activation of i^{th} Layer of Neural Network
- X = input

- $W^{[i]}$ = Weight from (i-1)th to ith Layer of Neural Network
- $b^{[i]}$ = Biases for ith Layer of Neural Network

Then the second stage - calculate the error of the whole network, which is returned by the loss function and described in Algorithm 2.

Algorithm 2 Loss Function in Neural Network

procedure LOSSFUNCTION(Y)
  RETURN $\frac{1}{m}\sum_{i=1}^{m}(-Y^{(i)}\log(\hat{Y^{(i)}})-(1-Y^{(i)})\log(1-\hat{Y^{(i)}})$
end procedure

Here, m is number of training samples, and $Y^{(i)}$ is the original output for ith sample and $\hat{Y^{(i)}}$ is the predicted output for ith sample.

And then the third stage is Back-Propagation which is done to update the weights accordingly to minimize the cost function or to reduce the error and described in Algorithm 3.

Algorithm 3 Back Propagation in Neural Network

```
procedure BACKPROPAGATION(X, Y)
    dZ[3] = A[3] - Y                                  ▷ Derivatives Calculation
    dW[3] = (1/m) dZ[3].A[2].T
    dB[3] = (1/m) np.sum(dZ[3], axis=1, Keepdims = True)
    dZ[2] = W[3].T.dZ[3] * DReLU(Z[2])
    dW[2] = (1/m) dZ[2].A[1].T
    dB[2] = (1/m) np.sum(dZ[2], axis=1, Keepdims = True)
    dZ[1] = W[2].T.dZ[2] * DReLU(Z[1])
    dW[1] = (1/m) dZ[1].XT
    dB[1] = (1/m) np.sum(dZ[1], axis=1, Keepdims = True)

    W[1] = W[1] - 0.01 * dW[1]                        ▷ Parameter Update
    b[1] = b[1] - 0.001 * dB[1]
    W[2] = W[1] - 0.01 * dW[2]
    b[2] = b[1] - 0.001 * dB[2]
    W[3] = W[1] - 0.01 * dW[3]
    b[3] = b[1] - 0.001 * dB[3]
end procedure
```

The notations that have been used in back propagation is described below:

- $dZ^{[i]}$ = Derivative of Z of ith Layer of Neural Network
- $A^{[i].T}$ = Transpose of $A^{[i]}$
- $dW^{[i]}$ = Derivative of W of ith Layer of Neural Network
- $dB^{[i]}$ = Derivative of B of ith Layer of Neural Network
- DReLU =

$$\begin{cases} 0 & x < 0 \\ 1 & X \geq 0 \end{cases}$$

D. Recommendation Procedure

Based on the output of the ANN, the recommendation is done. At first, for a single user U_i whose friendship's choice is already predicted by the ANN, $U_E \subset U$ is selected, where U_E is the set of all users having a similar choice of friendship trait as U_i and U is the set of all users. Then all elements of U_E are sorted in an ASCENDING order based on distance between U_i's score and each user in U_E's score for each personality category. For distance measurement, standard *Euclidean Distance* measure is used. The top *k* number of users by having the minimum distance and having the most similar personality then are recommended as friends.

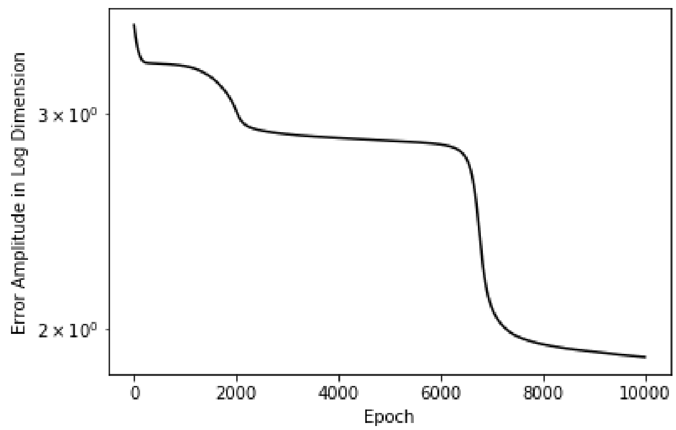

Fig. 4. Error VS Epoch

IV. PERFORMANCE EVALUATION

We evaluated the performance of the ANN we developed for predicting the personality choice of a user for friendship. The average training loss vs Epoch graph is given in Fig 4. This is a semilogy graph, which shows for a duration of 10,000 epoch, the loss gets reduced in a bit slow manner, although the general trend of the graph was always downwards with number of epochs increasing.

The average precision and recall of the model (shown in Fig 5.) - both were around 85%, and average F1-Measure was around 82% and average accuracy was around 83%. This performance was measured on our test dataset. The overall size of our dataset was 9917 samples, among which 9000 samples were used for training and validation purpose and the rest were used for final test purpose.

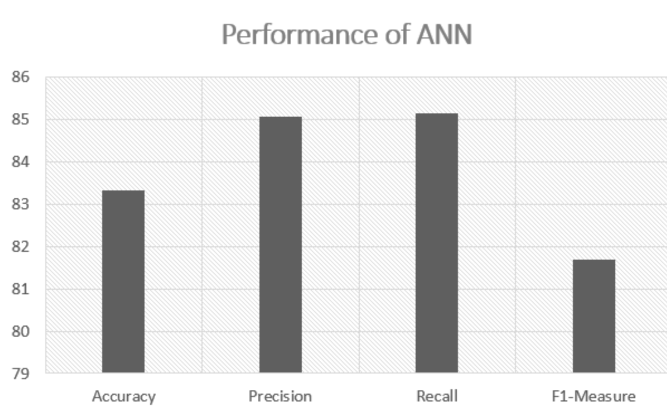

Fig. 5. Performance of the Neural Network

V. Future Work

In our work, the performance evaluation task was limited to only evaluating the performance of the ANN. We did not perform the performance of the recommender system itself as according to our understanding it would take a bigger comprehensive study with a probable duration of 3-6 months. For evaluating the performance of the recommender system, we might have to evaluate whether a user has actually become friend with any of the recommended friends from our proposed recommendation list, how much interaction is there for a certain period of time after becoming friend - a quantitative measurement of the success of friendship would be required to be precise. So a comprehensive study about the performance of that recommendation system along with any other probable optimization to our model remain to be implemented.

VI. Conclusion

In real life, for making friends we subconsciously give highest priority to the similarity or preference in human personality. To imitate that in social network is our main objective. We believe, proper implementation of this friend recommendation system do have a huge potential of being successful if implemented properly.

References

[1] C. Basu, H. Hirsh, and W. Cohen, Recommendation as classification: Using social and content-based information in recommendation, in Proceedings of the national conference on artificial intelligence. John Wiley & Sons LTD, 1998, pp. 714720.

[2] G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, Knowledge and Data Engineering, IEEE transactions on, vol. 17, no. 6, pp. 734749, 2005.

[3] Konstas, V. Stathopoulos, and J. M. Jose, On social networks and collaborative recommendation, in Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009, pp. 195202.

[4] Hotho, R. Jschke, C. Schmitz, and G. Stumme, Information retrieval in folksonomies: Search and ranking, in The semantic web: research and applications. Springer, 2006, pp. 411426.

[5] Song Chen, Samuel Owusu, Lina Zhou, Social Network Based Recommendation Systems: A Short Survey in IEEE Computer Society, 2013.

[6] Li bian, Henry Holtzman Online Friend recommendation through Personality Matching and Collaborative Filtering in The Fifth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM 2011.

[7] Namrata M. Eklaspur, Anand S. Pashupatinath, Friend recommender system for social networks by life style extraction using probabilistic method in International Journal of Computer Science Trends and Technology in June-2015.

[8] Lizi Zhang, Hui Fang, Wee Kiong Ng, Jie Zang Interaction Ranking-Based trust worthy Friend Recommendation in International Joint Conference of IEEE TrustCom-11/IEEE ICESS-11/FCST-11, in 2011.

[9] Xiwang Yang, Wang Guo, Young Liu Bayesian-Inference-Based Recommendation in Online Social Networks, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 4, in APRIL 2013.

[10] Wu Youyoua, Michal Kosinskib, and David Stillwella Computer-based personality judgments are more accurate than those made by humans in PNAS in January 2015.

[11] Zhiwei Deng, Bowei He, Chengchi Yu, and Yuxiang Chen, Personalized Friend Recommendation in Social Network Based on Clustering Method in ISICA, 2012, China.

[12] Jilian Tang, Xia Hu, Huan Liu, Social Recommendation: A Review in 2013, China.

[13] Vinti Agarwal, K. K. Varadwaj A collaborative filtering framework for friend recommendation in social networks based on interaction intensity and adaptive user similarity. In September, 2012.

[14] Jennifer Gobeck, Christina Robles, Karen Turner Predicting Personality with Social Media in 2011

[15] IPIP.ori.org standard questionnaire on Big Five Personality Analysis The Big Five Personality Test

[16] Chen, Yubo and Fay, Scott and Wang, Qi, The Role of Marketing in Social Media: How Online Consumer Reviews Evolve (January 11, 2011). Available at SSRN: https://ssrn.com/abstract=1710357 or http://dx.doi.org/10.2139/ssrn.1710357

[17] Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes. 2011. Social Network Analysis and Mining for Business Applications. ACM Trans. Intell. Syst. Technol. 2, 3, Article 22 (May 2011), 37 pages. DOI: https://doi.org/10.1145/1961189.1961194

[18] David Liben-Nowell and Jon Kleinberg. 2007. The link-prediction problem for social networks. J. Am. Soc. Inf. Sci. Technol. 58, 7 (May 2007), 1019-1031. DOI=http://dx.doi.org/10.1002/asi.v58:7

[19] Symeonidis, P., Ntempos, D., and Manolopoulos, Y. (n.d.). Recommender systems for location-based social networks.

[20] Grinde, B., Patil, G., Grinde, B., and Patil, G. G. (2009). Biophilia: Does Visual Contact with Nature Impact on Health and Well-Being? International Journal of Environmental Research and Public Health, 6(9), 23322343. http://doi.org/10.3390/ijerph6092332

[21] Xie, X., and Xing. (2010). Potential Friend Recommendation in Online Social Network. In 2010 IEEE/ACM Intl Conference on Green Computing and Communications Intl Conference on Cyber, Physical and Social Computing (pp. 831835). IEEE. http://doi.org/10.1109/GreenCom-CPSCom.2010.28

[22] Scellato, S., and Mascolo, C. (2011). Measuring user activity on an online location-based social network. In 2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 918923). IEEE. http://doi.org/10.1109/INFCOMW.2011.5928943

[23] Kautz, H., Selman, B., and Shah, M. (1997). Referral Web: combining social networks and collaborative filtering. Communications of the ACM, 40(3), 6365. http://doi.org/10.1145/245108.245123

[24] Hsiu-Chia Ko, and Tsun-Keng Chen. (2009). Understanding the continuous self-disclosure of bloggers from the cost-benefit perspective. In 2009 2nd Conference on Human System Interactions (pp. 520527). IEEE. http://doi.org/10.1109/HSI.2009.5091033

[25] Jacovi, M., Guy, I., Ronen, I., Perer, A., Uziel, E., and Maslenko, M. (2011). Digital Traces of Interest: Deriving Interest Relationships from Social Media Interactions. In Proceedings of the Twelfth European Conference on Computer Supported Cooperative Work, ECSCW 2011, 24-28 September 2011, Aarhus University, Denmark (pp. 2140). Retrieved from http://www.ecscw.org/2011/05- Jacovi et Al 21-40.pdf

[26] Golbeck, J., and Jennifer. (2009). Trust and nuanced profile similarity in online social networks. ACM Transactions on the Web, 3(4), 133. http://doi.org/10.1145/1594173.1594174

[27] Liang, T.-P., Ho, Y.-T., Li, Y.-W., and Turban, E. (2011). What Drives Social Commerce: The Role of Social Support and Relationship Quality. International Journal of Electronic Commerce, 16(2), 6990. http://doi.org/10.2753/JEC1086-4415160204