

---

## A New Approach for Trust Prediction by using collaborative filtering based of Pareto dominance in Social Networks

Elnaz Ghodousi<sup>1\*</sup>, Ali Hamzeh<sup>2</sup>

<sup>1</sup>Computer Science and Engineering Dept, Shiraz University, Shiraz, Iran, M. Sc. student  
[e.ghodousi@shirazu.ac.ir](mailto:e.ghodousi@shirazu.ac.ir)

<sup>2</sup>Computer Science and Engineering Dept, Shiraz University, Shiraz, Iran, Ph. D  
[ali@cse.shirazu.ac.ir](mailto:ali@cse.shirazu.ac.ir)

### Abstract

*Along with the increasing popularity of social web sites, users rely more on the trustworthiness information for many online activities among users.[24] However, such social network data often suffers from two problems, (1)severe data sparsity and are not able to provide users with enough information, (2)dataset's is very large. Therefore, trust prediction has emerged as an important topic in social network research. In this paper we proposed a new approach by using collaborative filtering method and the concept of Pareto dominance. We uses Pareto dominance to perform a pre-filtering process eliminating less representative users from the k-neighbour selection process while retaining the most promising ones. The results from experiments performed on FilmTrust dataset and Epinions dataset.*

**Keywords:** pareto dominance; trust; trust prediction; collaborating filtering; social network..

---

## 1 Introduction

Along with the increasing growth of social networks, the definition of social network services is improved as well. The domain of social networks from collaborative based sites like: facebook to propagation-based sites like: twitter, youtube, is increased. YouTube reports 24 hours of new video uploaded to their site every minute, and 2 billion videos watched every day. Facebook has over 600 million users who upload 2.5 billion photos per month, plus status updates, comments, videos, questions, and discussion posts. Twitter has over 200 million users creating over 90 million new tweets a day. The numbers continue for review sites, blogs and blog comments, more specialized social networks, and so on. With so much user interaction and content created, the question of whom and what to trust has become an increasingly important challenge on the web. A user is likely to encounter dozens if not hundreds of pieces of user-generated content each day, and some of it will need to be evaluated for trustworthiness. Fig.1 [22]

Trust is a pervasive concern in human and computational interactions [1, 2]. We must trust the services we receive and we rely on for our work and daily life, ranging from the use of cars and transports to the use of internet services, or the interaction between computational agents performing searches or other delegated tasks. Trust in automation [3] has been identified as a major concern in the development of human centered computing, proposing active evaluation of trust strategies to correct unjustified trust or mistrust, and to assess their consequences

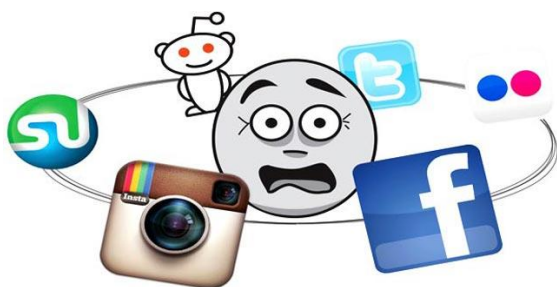


Figure 1 A demonstration for scenarios that users make trust decision in various forms, such as accepting others' adding requests from facebook, believing news from other twitter users, reading messages sent by others

Some philosophical definitions of Trust are:

- “the degree of subjective belief about the behaviors of (information from) a particular entity” [4]
- “the quantified belief by a truster with respect to the competence, honesty, security, and dependability of a trustee within a specified context” [1]
- “a particular level of the subjective probability with which an agent will perform a particular action,

both before [we] can monitor such action (or independently of his capacity of ever be able to monitor it) and in a context in which it affects [our] own action”.

Trust information can help a user make decisions, sort and filter information, receive recommendations, and develop a context within a community with respect to whom to trust and why.

Most interactions between two users in online social networks can be broken down into the scenario seen in Fig. 2: Alice is a service requester, and Bob is a service provider. Bob is the target whose trust is to be evaluated along with the topic of one of his services, and Alice's question, “Can I trust Bob on this service?” An effective trust evaluation algorithm is expected to provide a proper answer for Alice.

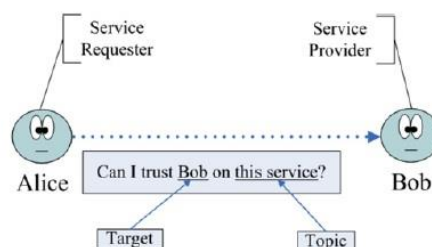


Figure 2 A scenario of trust evaluation.

To make a decision about whether or not to trust Bob, it is natural for Alice to ask her neighbors for suggestions. Next, her neighbors will ask their own neighbors. They will continue to repeat this asking process until they connect with someone who knows Bob. [5]

## 2 Related Work

Jennifer Golbeck was one of the first pioneers to research the problem of trust prediction from a computer science perspective. In Golbeck [6], she discusses various properties of trust, such as transitivity, composability, and asymmetry, and proposes algorithms for inferring binary and continuous trust values from trust networks, based on trust propagation. Kuter and Golbeck [7] suggest another trust inference algorithm called Sunny. The algorithm uses a probabilistic sampling technique to estimate the confidence in the trust information from some designated sources.

An efficient trust propagation algorithm is suggested by Massa and Avesani [8]. To compute the trust rating for a particular sink, its neighbours are first filtered to exclude untrustworthy members whose trust ratings are less than a threshold. Then, a weighted average is computed and assigned to the sink. The algorithm starts from a source and recursively computes the trust ratings for the nodes it discovers until the rating for the sink is computed.

Sherchan et al. [9] suggest a temporal Hidden Markov Model for reputation prediction. The model has five states and each state is represented by four hidden factors. To incorporate temporal sensitivity into a basic model, the authors suggest to remove older data and add more recent one in each iteration.

Liu et al. [10] develop a taxonomy of user relationships for the Epinions dataset. This taxonomy is used to obtain an extensive set of simple features which is in turn employed for training Naive Bayes and Support Vector Machines (SVM) classifiers. However, one should note that it is not always feasible to employ the overwhelmingly large number of features suggested.

Borzemek et al. [11] suggest a set of five user-similarity features. The first three features are graph-based and capture the incoming and outgoing edges for a pair of users. The last two features capture the number of reviews of a prospective trustee, and the number of the rated.

Noor and Sheng [12] and Sinclaire et al. [13] focus on the trustworthiness of prediction and the impact it has on consumers. Noor and Sheng [12] suggest to compute the trustworthiness as a sum of feedbacks weighed by their trust credibilities. The trust credibilities are in turn computed from two major components: Feedback Density and Majority Consensus. Majority Consensus measures how well the feedbacks by a particular set of consumers are aligned with the feedback majority. Feedback Density penalizes services that receive their feedbacks from a smaller number of unique consumers.

B.Huang , A. Kimmig et al present a flexible framework for probabilistic modeling of social networks that allows one to represent these different models and more. The framework, probabilistic soft logic (PSL), is particularly well-suited for this domain, as it combines a declarative, first-order logic-based syntax for describing relational models with a soft-logic representation, which maps naturally to the non-discrete strength of social trust. They demonstrate the flexibility and effectiveness of PSL for trust prediction using two different approaches: a structural balance model based on social triangles, and a social status model based on a consistent status hierarchy. [14]

X. Zheng et al. propose a new trust prediction model based on trust decomposition and matrix factorization, considering all the above influential factors and differentiating both personal and interpersonal properties. In this model, we first decompose trust into trust tendency and tendency-reduced trust. Then, based on tendency reduced trust ratings, matrix factorization with a regularization term is leveraged to predict the tendency reduced values of missing trust ratings, incorporating both propagated trust and the similarity of users' rating habits.[15][23]

In this paper, we propose a new trust prediction model based on rating matrix. The main contributions of our work are summarized as follows:

1. Find the k users most similar to the active user (the k-neighbours of the active user). This phase has the most significant impact on the quality of the prediction. The method proposed in this paper provides a novel approach for obtaining a suitable set of neighbours to the active user.
2. Predict the rating that the active user would give to other users they have not yet rated, by observing the ratings of their k-neighbours. When trying to predict an user's value, there will normally be a significant number of neighbours who have not rated the item; therefore, mechanisms must be defined that enable the k-neighbours' ratings to be combined satisfactorily.

### 3 The Proposed Approach

In this section, first we will discuss about collaborating filtering method Then explain the concept of Pareto dominance, Finally, we explain the proposed method.

#### 3-1 Collaborating Filtering

Collaborative filtering is one of the widely used technologies in the e-commerce recommender systems. It can predict the interests of a user based on the rating information of many other users. One of the main problems faced by CF is the high degree of sparsity in the rating databases, which arises from the small percentage of available items for which a given user generally provides ratings. Thus, when we want to calculate the similarity between each pair of users, we must do so by only considering the items that both users have rated in common. [25]

Traditional metrics display a marked tendency to show high similarity between users based on the similarity of their ratings on a very small set of items. These metrics can assign maximum similarity to two users who have each rated hundreds of items but who have only rated three items in common. Using the k-nearest neighbour's (KNN) algorithm, it is common to find active users with a significant number of inadequate neighbours (neighbours who have little in format ion in common with the active user). Our hypothesis is that it is possible to improve the quality of the recommendations of a CFRS if we use the Pareto dominance concept, which eliminates the less representative users from the k-neighbours selection process and keeps the most promising ones. [21]

### 3-2 The concept of dominance

The solution provided here is to use the Pareto Dominance concept used in multi objective optimisation problems [16] to identify those users who correctly represent the user and who therefore must be considered to be candidate neighbours. In a multi objective optimisation problem, it is said that a solution  $x'$  is Pareto-optimal, efficient or non-dominated (under the minimisation hypothesis) if no other feasible solution exists (we will call the set of all feasible solutions  $X$ ), which takes a lower value in some objective without causing a simultaneous increase in at least one other one. Formally, we are faced with a problem of optimisation in which we must find the vector  $x = (x_1, \dots, x_n)^T \in X$  that optimises the multi objective function  $f(x) = (f_1(x), \dots, f_n(x))^T$ . [21]

### 3-3 Explanation of new method

#### 3-3-1 Introduction

The following sets are defined:

$U = \{u \in \mathbb{N} \mid 1 \leq u \leq l\}$ , set of users.

$I = \{i \in \mathbb{N} \mid 1 \leq i \leq p\}$ , set of items.

$V = \{v \in \mathbb{N} \mid \min \leq v \leq \max\} \cup \{0\}$ , set of possible ratings.

$R_u = \{(i, v) \mid i \in I, v \in V\}$ , ratings of user  $u$ .

We define the rating average of the user  $u$  as  $\bar{r}_u$ .

#### 3-3-2 Selecting the candidate neighbours (non-dominated users)

We determine the set of users who are candidate neighbours of the active user or, more formally, the set of non-dominated users with respect to the active user.

Let  $d(r_{x,i}, r_{y,i})$  be the absolute difference between the ratings given by user  $x$  and user  $y$  to the item  $i$ .

$$d(r_{x,i}, r_{y,i}) = \begin{cases} |r_{x,i} - r_{y,i}| & r_{y,i} \neq 0 \\ \infty & r_{y,i} = 0 \end{cases} \quad (6)$$

We say that user  $x$  dominates user  $y$  with respect to another user  $u$  (denoted as  $x >_u y$ ), if the following expression (7) is satisfied.

$$x >_u y \forall i \in I_u \Leftrightarrow d(r_{u,i}, r_{x,i}) \leq d(r_{u,i}, r_{y,i}) \wedge \exists j \in I_u \mid d(r_{u,j}, r_{x,j}) < d(r_{u,j}, r_{y,j}) \quad (7)$$

We define  $C_u$  as the set of users who are candidate neighbours to the user  $u$  (non-dominated users). The following expression must be satisfied:

Let  $D_u$  be the set of users who are dominated by at least one user with respect to user  $u$ .

$$\begin{aligned} C_u &\subset U, u \notin C_u, C_u = U - (D_u \cup \{u\}), \forall y \in \\ D_u, \exists x \in C_u \mid x >_u y \end{aligned} \quad (8)$$

#### 3-3-3 Finding the k-neighbours

To find the active user's k-neighbours we must complete the following steps:

1. Calculate the similarity of the active user to each of the users who are candidate neighbours ( $C_u$ ).

2. Find the k users with the highest similarities to the active user. [21]

To perform the first step we will use Pearson Correlation (10) similarity measure.

Let  $A_{x,y} = \{i \in I \mid r_{x,i} \neq 0 \wedge r_{y,i} \neq 0\}$  be the set of items rated by both users  $x$  and  $y$ .

Correlation  $(x, y) =$

$$\frac{\sum_{i \in A_{x,y}} (r_{x,i} - \bar{r}_x)(r_{y,i} - \bar{r}_y)}{\sqrt{\sum_{i \in A_{x,y}} (r_{x,i} - \bar{r}_x)^2 \cdot \sum_{i \in A_{x,y}} (r_{y,i} - \bar{r}_y)^2}} \quad (10)$$

#### 3-3-4 Trust prediction

Determine the trust predictions for the active user based on ratings made by the set of k-neighbours.

To complete the first step (prediction), we must define the way in which the ratings of the k-neighbours are combined. Therefore we use deviation from the mean (12) for trust prediction.

Let  $G_{u,u'} = \{n \in K_u \mid t_{n,u'} \neq 0\}$  be the set of neighbours who have trusted user  $u'$ . We define  $K_u$  as the set of k-neighbours of the active user.

$$p_{u,u'} = \bar{t}_u + \frac{\sum_{n \in G_{u,u'}} \text{sim}(u,n)(t_{n,u'} - \bar{t}_n)}{\sum_{n \in G_{u,u'}} \text{sim}(u,n)} \Leftrightarrow G_{u,u'} \neq \emptyset \quad (12)$$

$t_{n,u'}$  Trust between the user 1 and user 2

## 4 Experiment And Result

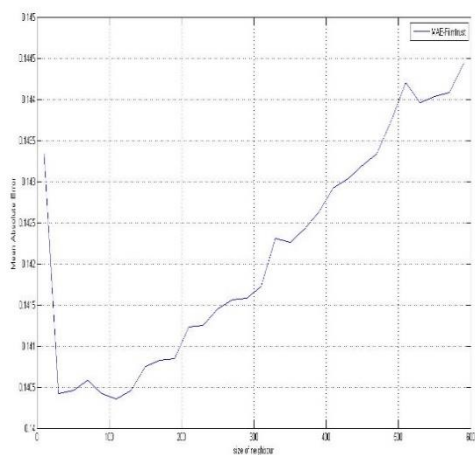
In this paper, we want to compare the prediction performance with other methods in trust prediction.

### 4-1 Data set

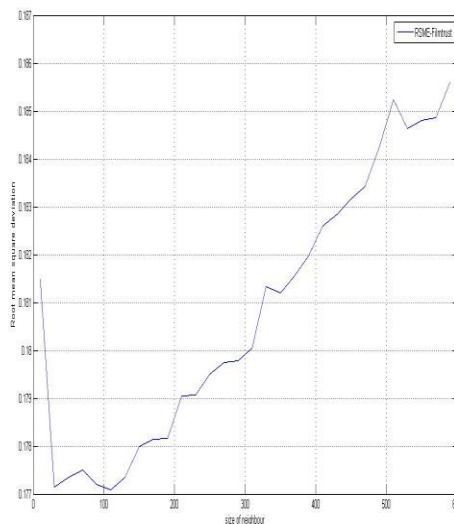
A real-world datasets are used in our experiments, namely FilmTrust and Epinion. The FilmTrust data set is provided by Guo et al. [17], containing 35,497 movie ratings given by 1508 users ranging from 0.5 to 4.0 with step 0.5. Users can share their ratings and specify other users as trustworthy. In total, there are 1853 trust statements.

Epinions dataset was collected by Paolo Massa in a five-week crawl from Epinions.com. It consists of two parts: one is the ratings part; the other is the trust part. The Epinions dataset consists of 49,290 users, 139,738 items, 664,824 reviews from users to items, and 487,181 trust statement between users.

In the case of FilmTrust and Epinion we use 80% of each user's data and item's data for training and the remaining data use as test data.

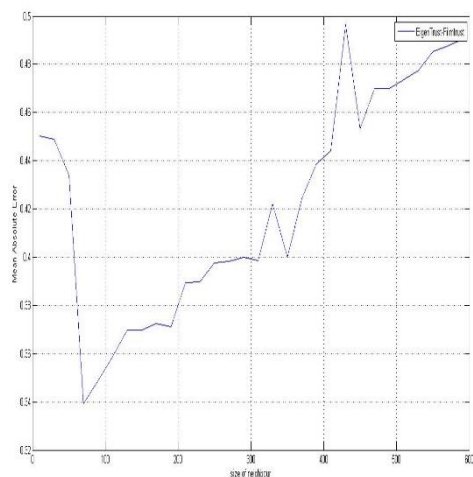


**(A) Our method**

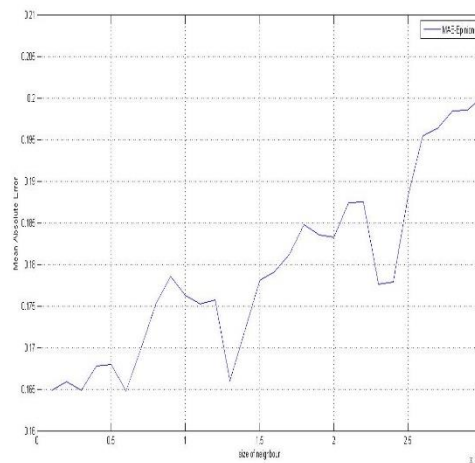


**(C) Our method**

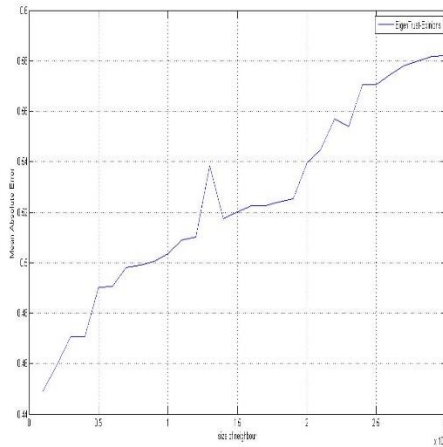
Figure 3 Improvements produced in quality measures results using the proposed method on filmtrust dataset: (A)MAE-our method, (B)MAE-EigenTrust, (C)RMSE-ourmethod



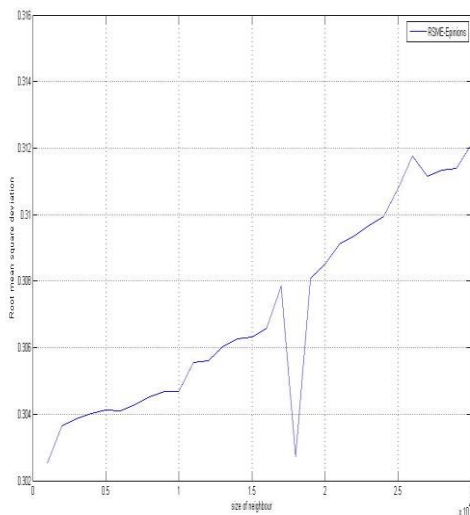
**(B) EigenTrust**



**(D) Our method**



(E) EigenTrust



(F) Our method

Figure 4 Improvements produced in quality measures results using the proposed method on Epinions dataset: (A)MAE-our method, (B)MAE-EigenTrust, (C)RMSE-ourmethod

#### 4-2 Evaluation metrics

We use MAE, RMSE, Kendall's  $\tau$  statistic, and Spearman's rank correlation  $\rho$  to evaluate the prediction precision of algorithm, which are computed by measuring the deviation between the prediction rating and the actual rating, while  $\tau$  and  $\rho$  measure ranking performance. Obviously, the smaller MAE (or RMSE) is, the higher the prediction precision of algorithm is. MAE and RMSE can be computed as:[18]

$$MAE = \frac{1}{\#M_u} \sum_{u' \in M_u} |r_{u,u'} - p_{u,u'}| \quad (13)$$

$$RMSE = \sqrt{\frac{1}{\#M_u} \sum_{u' \in M_u} (r_{u,u'} - p_{u,u'})^2} \quad (14)$$

Let  $M_u = \{u' \in U | r_{u,u'} \neq 0 \wedge p_{u,u'} \neq 0\}$  be the set of users trusted by user  $u$  where a prediction can be obtained.

#### 4-3 Evaluation results

The results demonstrate that compared with the previous approach, the accuracy of the proposed approach significantly is improved. the previous method such as : EigenTrust, TidalTrust, PSL.

EigenTrust [19] is a global metric analogous to PageRank [20] that computes a trust value for each node by finding the left principle eigenvector of a normalized trust matrix.

TidalTrust [6] is a graph-based algorithm that propagates trust values through neighbors by recursively using the weighted average of neighbor trust to decide a node's trust for another. In contrast to EigenTrust, TidalTrust predicts distinct trust values per link, rather than a single global trust value per node.

Probabilistic soft logic (PSL) [14] is a system for probabilistic modeling using first-order logic syntax. PSL uses soft truth values, relaxing truth to the interval  $[0; 1]$  and adapting logical connectives accordingly.

The results shown in Table1, Fig3, Fig4. As shown my proposed method in better than another methods.

Table 1: MAE, Kendalltau statistic  $\tau$ , and Spearman's rank correlation  $\rho$  evaluation on filmtrust dataset

Method	MAE	$\tau$	$\rho$
EigenTrust	0.339	-0.054	-0.074
TidalTrust	0.236	0.089	0.117
PSL-Balance	0.193	0.235	0.314
PSL-Status	0.230	0.205	0.277
Proposed method	<b>0.1402</b>	<b>0.4682</b>	<b>0.5041</b>

As shown in Table 1, our propose method is statistically tied for the best-performing method on all three metrics.

#### 5- Conclusion

This paper proposes the use of collaborating filtering and the concept of Pareto dominance submitted a new approach to perform trust prediction with the rating matrix. This idea was implemented as follows. We first Calculate the similarity of the active user to each of the users who are candidate



neighbours that selected by concept of Pareto dominance, then Find the k users with the highest similarities to the active user. Finally Predict the rating that the active user would give to other users they have not yet rated.

We conducted a series of experiments to compare traditional methods with our proposed methods. According to the results of the experiments, our proposed methods considerably outperformed traditional methods in, MAE, RMSE, Kendall's  $\tau$  statistic, and Spearman's rank correlation and coverage, especially when the user similarity is calculated using the Pearson correlation coefficient. Experiments with two datasets from a real online social network. It would be possible to use other methods in collaborating filtering and new methods of similarities in future works.

## References

- [1] B. Bhargava, L. Lilien, A. Rosenthal, M. Winslett, M. Sloman, T.S. Dillon, E. Chang, F.K. Hussain, W. Nejd, D. Olmedilla, V. Kashyap, "The pudding of trust". *IEEE Intell. Syst.* **19**(5), 74–88 (2004)
- [2] N. Shadbolt, "A matter of trust". *IEEE Intell. Syst.* **17**(1), 2–3 (2002).
- [3] R.R. Hoffman, M. Johnson, J.M. Bradshaw, A. "Underbrink, Trust in automation". *IEEE Intell. Syst.* **28**(1), 84–88 (2013).
- [4] K.S. Cook (ed.), "Trust in Society. Russell Sage Foundation Series on Trust", vol. 2. (Russell Sage Foundation, New York, 2003).
- [5] W .Jiang, G .Wang, J.Wu, "Generating trusted graphs for trust evaluation in online social networks ".*Future Generation Computer Systems.* Vol 31. (2012).
- [6] J.Golbeck, Computing and applying trust in web-based social networks. PhD thesis . (2005).
- [7] RV .Guha, R .Kumar, P.Raghavan, Tomkins, "A Propagation of trust and distrust". In: WWW, pp 403–412. (2004).
- [8] P.Massa, P.Avesani, "Controversial users demand local trust metrics: An experimental study on epinions".com community. In:AAAI, pp 121–126 (2005).
- [9] Sherchan W, Nepal S, Bouguettaya A, "A trust prediction model for service web". In: TrustCom, pp 258–265. (2011).
- [10] Liu H, Lim EP, Lauw HW, Le MT, Sun A, Srivastava J, Kim YA,"Predicting trusts among users of online communities: an epinions case study". In: ACM conference on electronic commerce, pp 310–319. (2008).
- [11] P.Borzemek, M.Sydow, A.Wierzbicki, "Enriching trust prediction model in social network with user rating similarity". In: CASoN, pp 40–47. (2009).
- [12] Noor TH, Sheng QZ, "Credibility-based trust management for services in cloud environments". In: ICSOC, pp 328–343. (2011).
- [13] J.Sinclair, J.Simon, R.Wilkes , "A prediction model for initial trust formation in electronic commerce". In: International Business Research, pp 17–27. (2010).
- [14] B.Huang, A.Kimmig, L.Getoor, J.Golbeck, "A Flexible Framework for Probabilistic Models of Social Trust". *Social Computing, Behavioral-Cultural Modeling and Prediction Lecture Notes in Computer Science* Volume 7812, 2013, pp 265-273. (2013).
- [15] X .Zheng., Y.Wang , A .Mehmet, Y.Zhong, G.Liu. "Trust Prediction with Propagation and Similarity Regularization", AAAI Publications, (2014).
- [16] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the Strength Pareto Evolutionary Algorithm, Technical Report 103, Computer Engineering and Networks Laboratory (TIK), Swiss Federal Institute of Technology (ETH) Zurich, (2001).
- [17] G. Guo, J. Zhang, and N. Yorke-Smith. A novel Bayesian similarity measure for recommender systems. In Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), (2013).
- [18] Cacheda, V. Carneiro, D. Fernandez, V. Formoso, "Comparison of Collaborative Filtering Algorithms: Limitations of Current Techniques and Proposals for Scalable, High-performance Recommender System", *ACM Transactions on the Web*, 5(1),article 2, (2011).
- [19] S. Kamvar, M. Schlosser, and H. Garcia-Molina." The Eigentrust algorithm for reputation management in P2P networks". In International Conference on World Wide Web (WWW), (2003).
- [20] S .Page, R .Brin, Motwani, and T. Winograd. "The PageRank citation ranking: Bringing order to the web". Technical Report 1999-66, Stanford InfoLab, (1999).
- [21] F. Ortega, J.L. Sanchez, J. Bobadilla, A. Gutierrez. "Improving collaborative filtering based recommender systems results using Pareto dominance". *Information Sciences: an International Journal* Volume 239, Pages 50-61 (2013).
- [22] T.DuBois, J. Golbeck, A.Srinivasan. "Predicting trust and distrust in social networks". In Proceedings of the third IEEE International Conference on Social Computing (SocialCom) (2011).
- [23] N. Korovaiko, A. Thomo. "Trust prediction from user item ratings". *Social Netw. Analys. Mining*, 3(3):749–759, (2013).
- [24] J.Huang, F.Nie, H.Huang, Y. Tu. "Trust prediction via aggregating heterogeneous social networks". In 21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 November 02, 2012, pages 17741778, (2012).
- [25] D.Jia, F.Zhang, S.Liu. "A Robust Collaborative Filtering Recommendation Algorithm Based on Multidimensional Trust Model". *Journal of Software*, Vol 8, No 1 (2013).