Modeling Trust for Rating Prediction in Recommender Systems

Anahita Davoudi* Mainak Chatterjee†

Abstract

Traditional recommender systems usually ignore the social interactions between users in a social network and assume that users are independent and identically distributed. This assumption hinders the users to have access to personalized recommendations based on their circle of trusted friends. To model the recommender systems more accurately and realistically, we propose a social trust model and use the probabilistic matrix factorization method to predict user rating for products based on user-item rating matrix. The effect of users friends tastes is modeled using a real-valued trust which is defined based on importance and similarity between users. Similarity is modeled using a rating-based (Vector Space Similarity algorithm) and connection-based methods; centrality is quantified using degree and eigen-vector centralities. To validate the proposed method, rating estimation is performed on the Epinions dataset. Experiments show that our method provides better prediction when using trust relationship based on centrality and similarity rather than using the binary values. Also, degree centrality is shown to be more effective compared to the eigen-vector centrality. In addition, trust model using connection-based similarity is observed to have better performance compared to the ones that use rating-based similarity.

1 Introduction

Recommender systems help users with item selection and purchasing decisions based on users' tastes and preferences using a variety of information gathering techniques. Such information is gathered either explicitly by mining user's ratings, or implicitly by monitoring user's behavior. These systems offer a personalized experience based on social interactions or user preferences are considered as a fantastic opportunity for retailers in e-commerce businesses. Many recommendation techniques have been studied [10, 20] and have been well adapted to commercial websites such as Amazon, Netflix, etc. Such commercial websites offer a vast number of products for users with different tastes.

Despite the fact that many studies have been done on similar problems, there is still great potential in using the social relationships in furnishing and harnessing the recommender systems. Traditional recommender systems assume that users are independent and identically distributed which results in ignoring the social interactions and trust relationships between users. However, user's social relationships play an important role in the behavior of users regarding future ratings. Since most of the similarities within a network are caused by the influence and interactions of its users, it is reasonable to develop a social recommender system based on the user connections and interactions. Social recommender systems focus on easing information and interaction burden by applying different methods that present the most relevant information to the users. But retailing platforms usually do not consider social factors such as relationships and trust among the users and the power of social influence is not exploited. On the other hand, social networking platforms generally do not consider online shopping related factors such as purchase history and product rating.

In addition to social connections, trust relationships also influence one's decisions and ought to be considered for recommendations. In a social network, trust relationships and social relationships are two different concepts. Two socially connected users would not necessary trust each other. Also, multiple connections of a user would not have equal impact on user's opinions and decisions. In addition to trust relationships, users with similar taste in purchasing would show similar behavior when rating a product as well.

In this paper, we combine the features of social networks and e-commerce platforms to design a social recommender mechanism to increase the prediction accuracy of product recommendations in e-commerce by considering the factors of similarity, user importance in the network, and social trust relationships. The proposed model could be practically applied to new emerging social commerce platforms. We argue that users are influenced by social interactions, in particular, by the set of trusted friends and their respective importance. To that end, we combine social trust connections and user-item matrix to predict the rating that a user would assign to a product. We use matrix factorization to factor user-item rating matrix into two low-dimensional matrices consisting of user latent matrix and item latent matrix. For the social connections, we consider both user importance and user similarity to build the social trust model between users. We use vector space similarity to obtain the similarity between users. Using degree and eigen-vector centrality, we quantify the importance of users in the network. We use a linear combination of similarity and centrality to model the trust parameter between users. The proposed method captures the balance between user taste and her friends' taste and adjusts the share of centrality and similarity in the trust values using two parameters. The low-dimensional latent user-specific and item-specific matrices are estimated by performing gradient descent on the objective function. We use a dataset from Epinions to validate the proposed model. We estimate the accuracy of the proposed method in terms of the mean ab-

*Department of Computer Science, University of Central Florida
†Department of Computer Science, University of Central Florida

solute error by comparing the predicted and the actual user ratings of products. Results reveal that there is a high correlation between the predicted and the actual ratings. The proposed method is also compared using binary trust values as well as considering the eigen-vector and degree centrality. In summary, our experiment results show that the proposed model could enhance recommendation accuracy.

2 Related Work

Let us discuss the various aspects such as trust, similarity, and user preference that that are relevant for this paper.

2.1 Recommender Systems Different types of recommender techniques has been developed: collaborative filtering, content-based or hybrid. Content-based systems use items' characteristics and the ratings that users have given to generate recommendations. Collaborative systems identify similar users and analyze their preferences to generate recommendations. In [2] the users' purchase patterns are derived by sequential pattern analysis to collaboratively recommend items to the users. There are many studies of the combination of content-based and collaborative-based systems [12]. User-based and item-based approaches are combined in [12] to build a hybrid recommendation of movies in P2P networks.

Collaborative filtering methods are divided into three further categories of memory-based, model-based and hybrid of both. Memory-based methods utilize users' past behavior and recommend products that other users with similar interests have selected in the past [20]. They have been widely used in commercial recommender systems [11, 19]. Memory-based algorithms are either user-based [1, 7] or item-based [11, 20]. User-based algorithms predict rating given by a user to an item based on the ratings by similar users, whereas, item-based algorithms estimate the rating based on the ratings of similar items previously chosen by the user. These methods find similar users [1, 7] or similar items [4, 11, 20] for providing accurate predictions. These methods use Pearson Correlation Coefficient (PCC) algorithm [19] or Vector Space Similarity (VSS) algorithm [1] to compute the similarity for finding the similar users or items. Methods used in traditional recommender systems are mostly based on user-item rating matrix. These algorithms usually fail to find similar users since density of ratings in user-item rating matrix is often less than 1 percent [11]. Model-based methods utilize available data to train a predefined model for rating prediction. Some of these methods are: clustering model [9] and the Matrix Factorization model [13]. Model-based approaches can handle problems with limited data using hierarchical clustering to enhance the accuracy of the prediction [9]. Matrix factorization is another model-based method which factorizes the user-item rating matrix using low-rank representation [13]. Although model-based methods mitigate the sparsity problem, handling users who have never rated any item is a challenging problem in both memory-based and model-based approaches.

2.2 Trust Models Trust has a significant impact on users' online purchasing behavior. Therefore, trust plays a critical role in e-commerce experience. Many trust-based models have been introduced such as TrustWalker [8] which is a combination of both trust-based and item-based recommendations, TidalTrust [5] which finds all the raters with shortest distance from the source user and aggregates their ratings. Also, in [18] trust-aware recommendation is used to increase recommendation accuracy.

2.3 Similarity The similarity between two users has been modeled by similarity measures such as Vector Space Similarity (VSS) and Pearson Correlation Coefficient (PCC). Both have been incorporated in social recommender systems [1]. Based on the similarity concept, the trust relations are bidirectional and equal in both directions. However, this is not true in real world relationships where trust relationships are non-transitive. Also, a user trust relationship value is affected by the importance of that user. Users usually tend to follow an important friend regardless of the similarity between them. The trust relationship enforced by the importance of user is asymmetric since every user have their unique importance.

2.4 User Preference Model To provide personalized recommendation, there are two ways to capture users' preferences [6]: implicit and explicit. The implicit method gathers users' behavior to obtain their preferences [2]. Matrix factorization models built in [10] use implicit feedback from system. The explicit method filters and analyzes interactions and feedback to obtain users' specification [21]. In [13] a user-item matrix is considered with users' social trust graph to build a latent low-dimensional matrix for providing a better recommendation. Users opinion is modeled based on her own and her friends' opinions which reflect real life social interactions [15]. The similarity between users is incorporated in social recommender systems [16]. Also social recommendation algorithms with social regularization terms is used in [14] to constrain matrix factorization objective functions. In addition, using trust values in recommender systems would help to predict the behavior of those users who have rated fewer products [16].

Here, we capture the effect of users' similarity in trust values. We also argue that the importance of a user must be taken into consideration for finding the trust values for predicting the rating of products.

3 System Model

We consider a social recommender system for a social network that is represented as a weighted directed graph of users where edges represent the social trust relationship between users. Users rate items (products) on a scale 1 to 5. The adjacency matrix $A_{N\times N}$ represents the social connections between users. Also user-item rating matrix shows the rating given by each user to each item. The user-item rating matrix $R_{M\times N}$ represents the ratings that each user assigns to each item. The existence of a social connection between two users would not necessarily reflect their level of trust in each other. The method presented here is based on the assumption that the trust between users is impacted by similarity and importance of users.

Problem Statement: In a given recommender system, how can we predict the rating that user i would assign to product j, when the social relationship graph and the user-item rating matrix are given.

3.1 Similarity Enforced Trust

3.1.1 Rating Similarity There are several users' characteristics that affect the value of trust between users. Similarity between users is one of the most important ones since two users with the same taste are more likely to trust each other. The effect of similarity has been incorporated in social recommender systems for predicting user rating. Vector Space Similarity (VSS) and Pearson Correlation Coefficient (PCC) [1] are the two most popular methods used for similarity estimation. Here we apply the VSS algorithm to identify the similarity between users utilizes the common items that have been rated by both users i and f to compute similarity which is given by:

$$Sim(i,f) = \frac{\sum\limits_{j\in I(i)\cap I(f)} R_{i,j}\cdot R_{f,j}}{\sqrt{\sum\limits_{j\in I(i)\cap I(f)} R_{i,j}^2}\cdot\sqrt{\sum\limits_{j\in I(i)\cap I(f)} R_{f,j}^2}} \tag{3.1}$$

where j is an item that both users i and f have rated and $R_{i,j}$ is the rating that user i assigned to item j. $I(i)$ represents the set of items rated by user i. VSS is defined in $[0,1]$; larger value implies more similarity between user i and user f. The trust values enforced by similarity can be modeled by weighted average rating of the users using the similarity scores as the weights. Consequently, a connection with high similarity will have more impact on the user's rating.

3.1.2 Connection Similarity The similarity between two users can also be determined by the connection between these two users. The similarity between two users can be measured by the mutual connection they have in common. This can be done using the each user list of connections. For each edge we get the list of connections for both users and then list of mutual connection on both sides. The larger the value would be, it could be an indication of users having more similarity which shows that their connection is more valid in shaping the trust. The list of friends for each user i is defined $F(i)$. The proportion of mutual friends to list of friends for the starting node of relationship is defined as follows:

$$Sim(i,f) = \frac{F(i)\cap F(f)}{F(i)} \tag{3.2}$$

3.2 Centrality Enforced Trust

Although similarity is a major driving force for trust between users, there are other aspects as well. A user with high importance (i.e., high impact) is more likely to be followed by her friends regardless of their similarities. This aspect of trust relationship is modeled by considering the importance of users. The importance of the users in a social network can be quantified using centrality measures such as degree centrality, betweenness, closeness, eigen-vector centrality and pagerank [17]. To obtain the importance of users, we use degree centrality and eigen-vector centrality.

Degree centrality is the simplest centrality measure. It shows the degree of a node, representing how many nodes are connected to it. Eigen-vector centrality gives each node a value which is proportional to the sum of values of its neighbors. Eigen-vector centrality has a property: it can be large either because a node has many neighbors or because it has important neighbors (or both). Pagerank and Katz are similar to eigen-vector centrality except that they add a small free centrality value to each node. Closeness measures the mean distance from a node to other node. Betweenness centrality measures the extent to which a node lies on paths between other nodes. We choose eigen-vector and degree centrality since they consider the connections and also the importance of each connection. Other measure either gives free initial centrality or capturing the path and distances between nodes which we are not interested in.

Degree centrality is used as the basic indication of a user's importance which can be defined as the *number of connections*. In our case, it is the number of incoming edges (in-degree) in the social graph. We define the degree centrality C_l of a user l as:

$$C_l = \sum_{\forall m, l\neq m} A_{l,m} \tag{3.3}$$

where $A_{l,m}$ is the element of the adjacency matrix which represents the connection between user l and user m. Thus, with all connections treated equally, a user with more incoming edges has higher importance in the network.

Eigen-vector centrality of user l at time t is the defined as *sum of the centrality* of all connections of l which is given as:

$$C_l(t) = \sum_{\forall m} A_{l,m}(t) \times C_l(t-1) \tag{3.4}$$

where $C_l(t-1)$ is the centrality of user l at time $t-1$. In contrast to the degree centrality, the eigen-vector centrality considers both the number of incoming edges and also the centrality of the neighboring users. The eigen-vector centrality is computed iteratively by setting all initial values to 1 i.e., $C_l(0) = 1$ for all user l.

3.3 Combined Similarity and Centrality Trust

We use a linear combination of similarity and centrality to model the trust of user i in user k as [3]:

$$\Gamma_{i,k} = \beta \frac{Sim(i,k)}{\sum_{l \in \mathcal{T}(i)} Sim(i,l)} + (1-\beta) \frac{C_k}{\sum_{l \in \mathcal{T}(i)} C_l} \tag{3.5}$$

Here, β is the parameter that defines the contribution of similarity and centrality to the overall trust. $\beta = 0$ implies purely centrality enforced trust while $\beta = 1$ refers pure similarity-based trust values. $\mathcal{T}(i)$ refers to the set of trusted friends of user i. C_k refers to the centrality (i.e., measured using either degree or eigen-vector centrality) of user k.

4 Trust Model for Matrix Factorization

Matrix factorization has been widely used to develop social recommender systems [10, 13, 15]. Generally, matrix factorization helps to estimate either the user-item rating or user-trust matrix [15] using low-dimensional representative latent matrices. Here matrix factorization for social recommendation proposed by [15] is employed to examine the performance of the proposed trust relationship. The user-item rating matrix is factorized to learn two $l-$dimensional feature representation of users U and items V.

The user-item rating matrix R consists of m users and n items with rating values in range $[0, 1]$. U_i and V_j represent the $l-$dimensional user-specific and item-specific feature vectors of user i and item j. The conditional distribution for R, given Γ, U, V and σ_R^2 is defined as [13]:

$$p(R|\Gamma, U, V, \sigma_R^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} [\mathcal{N}(R_{ij}|g(\sum_{k \in \mathcal{T}(i)} \Gamma_{ik} U_k^T V_j), \sigma_\Gamma^2)]^{I_{ij}^R} \tag{4.6}$$

where $\mathcal{N}(R_{i,j}|\mu, \sigma_\Gamma^2)$ is probability density function of the Gaussian distribution with mean μ and variance σ_Γ^2. Here, Γ is the proposed trust parameter given by Eq. (3.5), $\Gamma_{i,k}$ is the trust value between users i and k. $R_{i,j}$ is the rating given to item j by user i, and σ_R^2 is the rating variance. I_{ij}^R is an indicator function representing whether user i rated item j. Based on the Bayesian inference and assuming Γ is independent of U and V, the conditional probability of U and V, given R, Γ, σ_R^2 , σ_U^2, and σ_V^2, is defined as:

$$p(U, V|R, \Gamma, \sigma_\Gamma^2, \sigma_U^2, \sigma_V^2) = \prod_{i=1}^{m} \prod_{j=1}^{n} [\mathcal{N}(R_{i,j}|g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j), \sigma_\Gamma^2)]^{I_{i,j}^R} \times \prod_{i=1}^{m} \mathcal{N}(U_i|0, \sigma_U^2 \mathbf{I}) \times \prod_{i=1}^{m} \mathcal{N}(V_j|0, \sigma_V^2 \mathbf{I}) \tag{4.7}$$

where σ_U^2 and σ_V^2 are the variance of user and item feature matrices. $\mathbf{I}$ is the identity matrix. The function $g(x) = 1/(1 + exp(-x))$ is a mapping function whose range is within $[0, 1]$. The set $\mathcal{T}(i)$ contains user i's trusted friends. The proposed social recommender system is based on the idea that user's ratings are impacted by her own taste and her immediate friends' tastes. The parameter α is used to balance between these two factors. The term $U_i^T V_j$ represents the estimated taste of user i of item j, while $\sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j$ term reflects her immediate friends' taste, given as the weighted average of their taste using the trust value as weights.

4.1 User-Specific and Item-Specific Matrices

In order to find the optimal values of U and V, the log of the posterior distribution given in Eq. (4.7) should be maximized. Equivalently, U and V can be derived by minimizing the sum-of-squared-errors given in the following equation:

$$\mathcal{L}(R, \Gamma, U, V) = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{i,j}^R (R_{i,j} - g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j))^2 + \frac{\lambda_U}{2} ||U||_F^2 + \frac{\lambda_V}{2} ||V||_F^2 \tag{4.8}$$

where $||.||_F^2$ is the Frobenius norm. λ_U and λ_V are user and item latent variance ratios.

The gradient decent approach can be used to solve the minimization problem given in Eq. (4.8) for finding U and V. Gradient decent is a local optimization method based on the partial derivative of the objective function with respect to the decision variables (i.e., U and V). The partial derivatives of $\mathcal{L}$ with respect to U and V are given in Eqs. (4.9) and (4.10).

(4.9)

$$\frac{\partial \mathcal{L}}{\partial U_i} = \alpha \sum_{j=1}^{n} I_{i,j}^R g'(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j) V_j \times (g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j - R_{i,j})$$

$$+ (1-\alpha) \sum_{p \in \phi(i)} \sum_{j=1}^{n} I_{p,j}^R g'(\alpha U_p^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(p)} \Gamma_{p,k} U_k^T V_j) \times (g(\alpha U_p^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(p)} \Gamma_{p,k} U_k^T V_j) - R_{p,j}) \Gamma_{p,i} V_j + \lambda_U U_i$$

(4.10)

$$\frac{\partial \mathcal{L}}{\partial V_j} = \sum_{i=1}^{m} I_{i,j}^R g'(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j) \times (g(\alpha U_i^T V_j + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T V_j - R_{i,j}) \times (\alpha U_i + (1-\alpha) \sum_{k \in \mathcal{T}(i)} \Gamma_{i,k} U_k^T) + \lambda_V V_j$$

Here $g'(x)$ is the derivative of logistic function where $g'(x) = exp(x)/(1 + exp(x))^2$. $\phi(i)$ is the set of the users who trust user i [15].

5 Simulation Model and Results

In order to test the validity and accuracy of the proposed rate prediction framework, we conduct extensive simulation experiments with data from Epinions [22].

5.1 Dataset Description Epinions is a review and rating website which allows users to rate items by giving an integer number between 1 and 5. The users can also form social connections by adding other users as their trusted friends. The social connections in this dataset are binary values and do not represent the actual trust values. The dataset includes 22166 users and 355754 social connections, leading to 0.0724 percent density in the user social relationship matrix. The total number of items is 296277, with a total of 922267 ratings, which results in a very sparse item-rating matrix with 0.0140 percent density.

As a result, the user-item rating matrix is also relatively sparse. On average, users have 16.05 trusted friends. The maximum number of friends for a user is 1551 and the most trusted user has 2023 other users trusting her.

5.2 Evaluation Metrics Evaluation measures for recommender systems are usually divided into three categories: 1) Predictive Accuracy Measures (such as MAE, RMSE) which evaluate how close the recommender system is in predicting actual rating values, 2) Classification Accuracy Measures (such as Precision, Recall, F1) which measure the frequency with which a recommender system makes correct/incorrect decisions regarding items based on the relevancy of the recommended items, and 3) Rank Accuracy Measures (such as Discounted cumulative gain(DCG) and Mean Average Precision (MAP)) which evaluate the correctness of the ordering of items performed by the recommendation system.

Precision is a measure of exactness and determines the fraction of relevant items retrieved out of all items (e.g., the proportion of recommended movies that are actually good). Recall is a measure of completeness and determines the fraction of relevant items retrieved out of all relevant items (e.g. the proportion of all good movies recommended). The F1 Metric attempts to combine Precision and Recall into a single value for comparison purposes so it may be used to gain a more balanced view of performance. The precision is the fraction of all recommended items that are relevant and recall is the fraction of all relevant items that were recommended. F-measure is a single value combining different facets of accuracy (precision and recall). Ranking accuracy measure ranks all items for user such that higher-ranked recommendations are more likely to be relevant to users. Since our proposed model focus on the error in the rating prediction, we use the metrics in the first category which evaluate the prediction accuracy of the recommender system. The other two categories are used for classification and ranking.

5.3 Predictive Accuracy Measures Let us formally define that error matrix that we would use.

Mean Absolute Error (MAE): This metric measures the average variation in the predicted rating vs. the actual rating. Let $R_{i,j}^{pre}$ be the predicted rating and $R_{i,j}^{act}$ be the actual rating given by the user i to the product j. The MAE is defined as follows:

$$MAE = \frac{\sum_{i,j} |R_{i,j}^{pre} - R_{i,j}^{act}|}{N} \quad (5.11)$$

Mean Squared Error (MSE): This metric punishes big errors more severely and is defined as follows:

$$MSE = \frac{\sum_{i,j} |R_{i,j}^{pre} - R_{i,j}^{act}|^2}{N} \quad (5.12)$$

Root Mean Squared Error (RMSE): This metric is a variant of mean square error and is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i,j} |R_{i,j}^{pre} - R_{i,j}^{act}|^2}{N}} \quad (5.13)$$

All these metrics measure the accuracy of the actual predictions and are easy to compute efficiently. Moreover, MAE and MAE-based error estimates have well known statistical properties. These characteristics MAE and RMSE good representative of error metrics to analyze the accuracy of the proposed model.

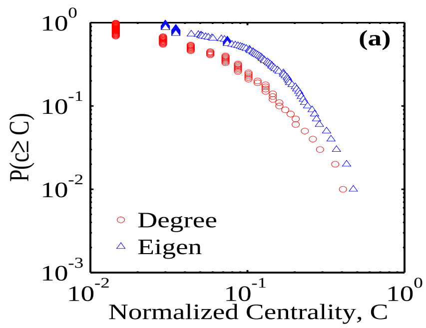

Figure 1: Distribution of centrality

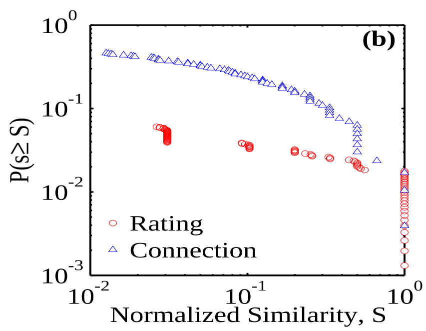

Figure 2: Distribution of similarity

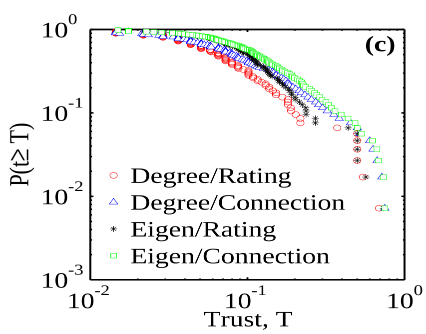

Figure 3: Distribution of trust values

5.4 Simulation Results As mentioned before, the trust relationships between users are defined based on centrality and similarity measures. Our objective is to capture the probability density function of centrality, normalized similarity, and trust. These distributions reveal what and how much impact each of these parameters have for various values of the parameter in question. Fig. 1 shows the distribution of degree and eigen-vector centrality. In Fig. 2, the distribution of rating-based and connection-based similarity are shown. The rating-based similarity has a relatively sparse distribution due to the lack of mutual rated products for two friends in many cases. The trust values are calculated as the weighted summation of centrality of similarity using the weight constant β. Fig. 3 shows the distribution of trust values using $\beta = 0.5$. Based on the different centrality and similarity measures, there are four types of trust values as illustrated in Fig. 3.

The proposed trust model is used to predict users rating based on the discussed matrix factorization technique using 75 percent of the data as the training set. According to Eq. (4.7), a user's opinion about a particular product would be a linear function of her connections' taste and her own taste using a weighting factor α. Smaller values of α is an indication of less impact from neighbors. As previously defined in Eq. (3.5), the trust model is presented as the linear combination of centrality and similarity using the weighting factor β. Higher values of β indicate higher impact of similarity rather than centrality on the trust values. Here, user and item latent variance ratio (λ_U and λ_V) are set to 0.001. The latent size is $L = 4$, $\alpha = 0.4$, and the number of iterations is 300. The performance of the proposed trust model for different values of β in terms of MAE and RMSE is shown in Figs. 4 and 5. Compared to the binary trust model (dashed black lines), the proposed trust model has better performance. Comparing different definitions of trust reveals that degree centrality is the better measure to model trust using eigen-vector centrality. The same is true for connection-based similarity compared to rating-based. An interesting point is that, although including centrality in trust model enhances the recommendation performance compared to the binary trust model, the trust models solely based on similarity (i.e.,$\beta = 1$) show the best performance for the studied network.

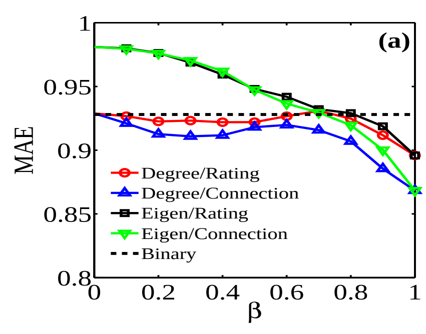

Figure 4: MAE using binary trust and the proposed trust model

The performance of the trust model (the definition which had the best performance in Figs. 4) and 5) for different latent sizes and training percentages are shown in Fig. 6 and Fig. 7. Generally, increasing the latent size as well as using more training data enhance the performance of the recommender system.

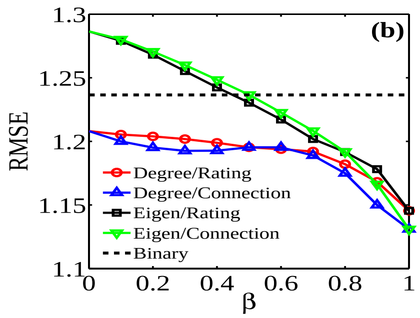

Figure 5: RMSE using binary trust and the proposed trust model

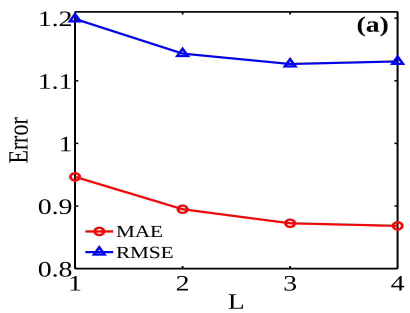

Figure 6: Errors for different latent sizes using degree centrality and connection-based similarity to define trust

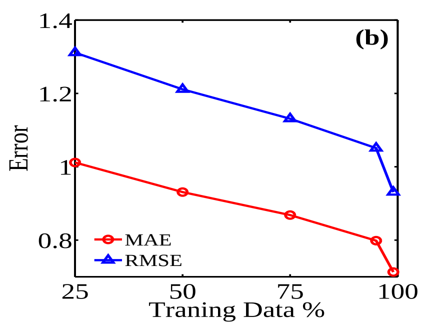

Figure 7: Errors for various training set sizes using degree centrality and connection-based similarity to define

The probability distribution of rating estimation error (i.e., estimated rating minus actual rating) for the binary trust and proposed trust model is shown in Fig. 8. Both probability distributions are a little right skewed, implying over-estimation. However, the proposed trust model seems to have relatively better performance especially for errors between 0.5 and 2, since it estimates more between 0.5 and 1 and less between 1 and 2 compared to the binary model. The probability distribution of absolute error ratio (i.e., absolute error divided by the actual rating) is shown in Fig. 9. The proposed trust model leads to less error ratio between 1 and 2 and more between 0 and 1 which implies relatively better performance.

In Figs. 10 and 11 the estimated versus actual ratings

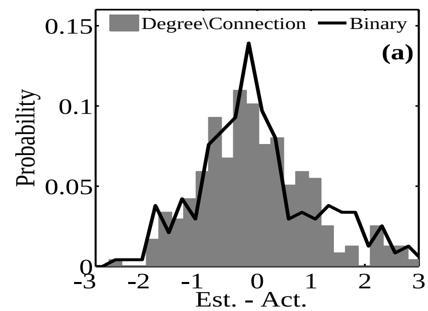

Figure 8: The probability distribution of error for rating estimation using binary trust and the proposed trust model

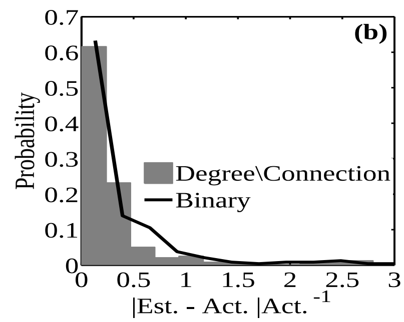

Figure 9: Absolute error ratio for rating estimation using binary trust and the proposed trust model

are shown for the proposed and the binary trust models. The boxes illustrate the lower, upper and inter quartiles, while the redline is the medium. The height of the boxes represents the variation of the estimated rating. Comparing Figs. 10 and 11, it is observed that the proposed trust model produces better estimations for low ratings (1 and 2) by slightly undermining the estimation. In addition, for high ratings, the proposed trust model reduces the variation of estimations, i.e., the height of the quartile boxes.

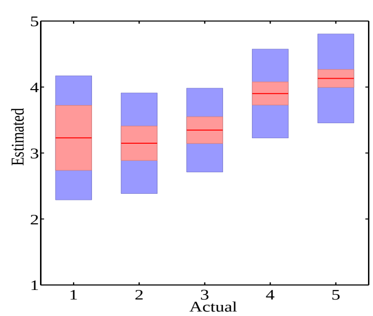

Figure 10: The quartile plot of actual versus estimated rating for the proposed trust model

6 Conclusions

With emerging applications of social networks and considering the role of social interactions in our daily life deci-

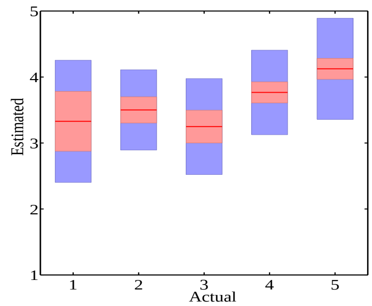

Figure 11: The quartile plot of actual versus estimated rating for the binary model.

sions, extracting information from user's social relationships is becoming a popular method for predicting user's behavior. We capture the trust relationships between users considering users with similar profile and their importance. The main assumption is that the users with more similarity would trust each other more; also users with higher importance would be trusted more. Similarity is quantified by a rating-based approach and a connection-based method. The importance is modeled by degree centrality and eigen-vector centrality. We define trust as a linear combination of similarity and centrality using a weighting parameter. The proposed framework is validated using real data from Epinions. Our result indicates that the proposed trust model produces better rating estimation in terms of the mean absolute error (MAE), the root mean squared error (RMSE) and error distribution, compared to the traditional binary trust model which is widely used in recommender systems. Trust enforced by degree centrality shows better performance compared to eigen-vector centrality. The same conclusion is valid for connection-based similarity compared to rating-based. The trust relationships are also observed to be more dependent on the similarity rather than centrality. Our proposed method is applied to the case when the graph of the social relationships is given. We only incorporate user's trust information since distrust information, although adding it may enhance the prediction accuracy, has a different feature space. Also the information propagation between users is ignored which eventually reduce the accuracy of prediction.

References

[1] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", Uncertainty in Artificial Intelligence, pp. 43–52, 1998.

[2] K. Choi, D. Yoo, G. Kim, Y. Suh, "A hybrid online-product recommendation system: combining implicit rating-based collaborative filtering and sequential pattern analysis", Electronic Commerce Research and Applications, vol. 11, pp. 309-317, 2012.

[3] A. Davoudi, M. Chatterjee, "Product Rating Prediction Using Trust Relationships in Social Networks", In Proc. IEEE CCNC, pp. 122-125, 2016.

[4] M. Deshpande and G. Karypis, "Item-Based Top N-Recommendation, ACM Transaction on Information Systems, Vol. 22, pp. 143–177, 2004.

[5] J. Golbeck, "Computing and Applying Trust in Web-based Social Networks", PhD thesis, University of Maryland College Park, 2005.

[6] U. Hanani, B. Shapira, P. Shoval, "Information filtering: overview of issues, research and systems", User Modeling and User-Adapted Interaction, Vol. 11, pp. 203-259, 2001.

[7] J. Herlocker, J. Konstan J., A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering", ACM SIGIR Conference, pp. 230–237 1999.

[8] M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation", ACM SIGKDD, pp. 397–406, 2009.

[9] A. Kohrs and B. Merialdo, "Clustering for Collaborative Filtering Applications", In proc. of the International conference on Computational Intelligence for Modeling Control and Automation, 1999.

[10] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques For Recommender Systems", Computer vol. 8, pp. 30–37, 2009.

[11] G. Linden, B. Smith, and J. York,"Amazon.com recommendations: Item-to-item collaborative filtering", IEEE Internet Computing, pp.76–80, 2003.

[12] Z.B. Liu, W.Y. Qu, H.T. Li, and C.S. Xie, "A hybrid collaborative filtering recommendation mechanism for P2P networks", Future Generation Computer Systems, vol. 26, pp. 1409-1417, 2010.

[13] H. Ma, H. Yang, M. R. Lyu and I. King, "SoRec: Social Recommendation Using Probabilistic Matrix Factorization", In proc. of ACM CIKM, pp. 931–940, 2008.

[14] H. Ma, D. Zhou, C. Liu, M. R. Lyu and I. King, "Recommender systems with social regularization", In proc. of ACM WSDM, pp. 287–296, 2011.

[15] H. Ma, I. King and M. R. Lyu, "Learning to Recommend with Explicit and Implicit Social Relations", ACM Transaction Intelligent Systems Technology, Vol. 2, 2011.

[16] H. Ma, "On measuring social friend interest similarities in recommender systems", In proc. of ACM SIGIR, pp. 465–474, 2014.

[17] M. Newman, "Networks:an introduction", Oxford University Press, 2010.

[18] J. O'Donovan and B. Smyth, "Trust in Recommender Systems", In Proc. of IUI, pp. 167–174, 2005.

[19] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture collaborative filtering of netnews", ACM CSCW, pp. 175–186, 1994.

[20] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms", World Wide Web conference, pp. 285–295, 2001.

[21] J.B. Schafer, J.A. Konstan, J. Riedl, "E-commerce recommendation applications", Data Mining and Knowledge Discovery, vol. 5, pp. 115–153, 2001.

[22] J. Tang. [online]. Available: www.public.asu.edu/ jtang20/datasetcode/truststudy.htm