Communications of the IIMA

Volume 12 | Issue 4 Article 1

2012

Trusting the Trusted: An Empirical Analysis

Sanjog Ray
Indian Institute of Management

Ambuj Mahanti
Indian Institute of Management

Recommended Citation
Ray, Sanjog and Mahanti, Ambuj (2012) "Trusting the Trusted: An Empirical Analysis," *Communications of the IIMA*: Vol. 12: Iss. 4, Article 1.
DOI: https://doi.org/10.58729/1941-6687.1196
Available at: https://scholarworks.lib.csusb.edu/ciima/vol12/iss4/1

Trusting the Trusted: An Empirical Analysis

Sanjog Ray
Indian Institute of Management, Indore, India
sanjogr@iimidr.ac.in

Ambuj Mahanti
Indian Institute of Management, Calcutta, India
am@iimcal.ac.in

ABSTRACT

One aspect of research in the area of bootstrapping of cold start users in trust-aware recommender systems is on guiding new users find trusted users in the trust network. In this paper we examine the question "Which type of users do cold start users actually trust?" We analyze the users trusted by cold start users in the Epinions dataset (Massa & Avesani, 2006a). We examine the set of trusted users on the basis of four critical parameters: number of outgoing links, number of incoming links, number of items rated and a hybrid parameter defined as preference score. We also experimentally evaluate the impact of the four parameters on prediction accuracy. Our analysis shows that cold start users will more likely trust a user with higher number of incoming links as a trusted user even though trusting users with high preference score would result in better prediction accuracy.

Keywords: Information filtering trust aware recommender systems, personalization, Internet, electronic commerce, cold start users.

INTRODUCTION

The approach used for solving the cold start problem in a trust-aware recommender system is to ask a cold start user to select few existing users in the system that they trust. Using the ratings of the items rated by the selected trusted users, recommendations are generated for the cold start user. An implicit assumption made in this solution is that cold start users or users new to the system have knowledge of the existing users in the system. Moreover, even if the cold start user has knowledge of the existing users, it is assumed that he is aware of the ratings given by the existing users to different items. To overcome the shortcomings of the proposed approach, recent research have started focusing on the problem: to whom should the new users connect to? (Victor, Cornelis, De Cock, & Teredesai, 2008). The suggested approach is to identify key figures in the trust network based on their impact on coverage and accuracy of recommendations made and then ask the cold start user to choose few of these users presented before them as trusted users.

In this paper, we analyze the trust connections of the cold start users in the dataset from Massa and Avesani (2006b), a real life recommender system dataset, to examine "What are the

characteristics of those users that cold start users actually trust?" The objective of our work is to provide insights on the type of users that should be preferred while creating a list of prospective trusted users to be presented before a cold start user. Our paper first discusses the existing parameters that form the basis of selecting prospective trusted users and the reasoning behind it. We then analyze the set of users trusted by cold start users using four parameters: number of outgoing links, number of incoming links, number of items rated, and a hybrid parameter defined as preference score. Using rank correlation coefficient as the metric we show that in the Epinions dataset (Massa & Avesani, 2006a), users with higher number of incoming links are preferred as trusted users as compared to other three parameters. We also examine which parameter gives the best predictive accuracy when selected as trusted user. Our results show that those cold start users that trust users with high preference scores have the benefit of getting more accurate recommendations.

TRUST IN WEB BASED SOCIAL NETWORKS

Web based social networks (Boyd & Ellison, 2007; Golbeck, 2005) can be defined as follows: Services that are accessible over the web and allow individuals to (1) construct a public or semi-public profile within a system designed specifically to support social network connections, (2) articulate a list of other users with whom they share a connection, and (3) browse their list of connections and those made by others within the system.

The study of web based social networks (WBSN) primarily focuses on connections between people. Scholars have examined the behavior of people in social network, how connections are formed and their evolution. WBSNs are a source of rich behavioral data. Data in WBSNs is extracted from user profiles and explicitly made connections between users. In WBSNs, much of the products, services, and features are based around social connections. Social connections are based on trust. Trust-aware recommender systems exploit these social connections or trust data as well profile data of users to create intelligent applications that provide personalized recommendations to users in WBSNs that are relevant and trustworthy.

Trust is a social phenomenon. As a social concept, the many facets and influences of social trust has been extensively addressed in sociology and social psychology literature. It is a complex notion as a result it has many subtly different definitions. In Luhmann (1979), trust has been described as a tool for complexity reduction. Mayer, Davis and Schoorman (1995), proposed a trust model that is designed to focus on trust in an organizational setting. In Lewis and Weigert (1985), trust is described as follows: *"trust which undergirds our everyday lives is a pure social construction which answers to our need for security by seeming to be a fact when it is always a projected assumption."* One of the widely accepted definition of trust has been given by the sociologist Gambetta. Gambetta (1990) defines trust as *"Trust (or, symmetrically, distrust) is a particular level of the subjective probability with which an agent assesses that another agent or a group will perform a particular action, both before he can monitor such action (or independently of his capacity of ever to be able to monitor it) and in a context in which it affects his own action."*

Social trust is hard to model in a computational system as factors affecting trust are many like past experiences, psychological factors impacted by social conditioning, influence by others opinions, motives etc. Modeling of trust in social networks is essential so that it can be used in computation that will benefit users. Gambetta definition of trust as a probability formed the foundation on which new trust models were proposed to formalize trust as a computational concept. Few of the popular trust models that have successfully modeled trust into a useful computational notion have been explained below.

One of the widely cited trust model is one proposed by Marsh (1994). In this model, trust has been classified into three categories, namely:

a) Basic trust: Basic trust refers to the amount of trust an agent will exhibit with someone he has not interacted previously. Basic trust is associated with every agent and gives the impression of an agent's general disposition.
b) General trust: this refers to the trust one agent exhibits for someone he knows.
c) Situational trust: this kind of trust is situation specific, an agent may trust some other agent only in specific situations.

In Golbeck (2005), the Marsh model has been critiqued as being highly theoretical and difficult to implement, and that it is particularly inappropriate for use in web based social networks as the information requirements of Marsh model are not captured in most WBSNs. From the many trust models available, we adopt the trust model used by Golbeck (2005). Golbeck's model is appropriate for social networks where users add simple expressions of trust, and this appropriately describes the way trust statements are passed in the Epinions dataset (Massa & Avensani, 2006a), i.e. the dataset we use in this work. In the Epinions data set, trust statements are single value statements explicitly provided by the users.

The definition of trust in Golbeck model is: *trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome*. The action and commitment does not have to be significant. Given this view of trust, the important characteristics of trust that emerge from the model and are specifically adopted in this work as the trust model are described below.

a) Transitive: The notion of transitivity of trust implies that if a user Alice trusts another user Bob, and Bob trusts Carol, then we can imply that Alice also trusts Carol to some extent. It must be noted, that amount of trust Alice will have on Carol will be of a much lesser degree as compared to the degree of trust she has with Bob. In other words, with transitivity, intensity of trust decreases. Trust is not perfectly transitive. Examples of transitivity in real life situations are aplenty, for example when we are deciding about visiting a doctor, we ask a trusted friend for recommendations. And in most cases we go to a doctor suggested by our trusted friends. We trust the doctor to some extent because our friends trust him.
b) Context-dependent: The importance we give to the opinions of a trusted friend on different subjects varies. While Alice may value Bob opinion on movies, she may not value Bob's opinion on books.
c) Asymmetrical: Trust is asymmetrical. The degree to which Alice trusts Bob may not be of the same degree as Bob trusts Alice. There may be a possibility that Bob does not trust Alice at all, while Alice trusts Bob to a certain degree.

d) Subjective: Alice may be trusted by both Bob and Carol, but the reasons why they trust Alice may be different. Trust is a personal, subjective opinion given by one user on another.
e) Weighted: this property is an extension of the property subjective. Different degrees of trust are represented on a scale of 0 to 1, where 1 signifies maximum trust and 0 signifies distrust.

TRUST-AWARE RECOMMENDER SYSTEMS

With increasing popularity of recommender systems in ecommerce sites they have become susceptible to attacks by malicious users who try to influence the systems by inserting biased data into the system (Mobasher, Burke, Bhaumik, & Williams, 2007). Recent research on trust-aware recommender systems (Massa & Avesani, 2004, 2006a, 2007a) is gaining popularity as they are found to be more robust against shilling attacks.

Trust aware recommender systems also use rating data for making predictions but in addition to rating data they also utilize trust data. Trust data are trust statements made by one user about another. In the Epinions data set used (Massa & Avesani, 2006a), trust statement given by a user A to another user B represents explicit score provided by user A expressing how much value user A attaches to the ratings and reviews given to different items by user B. Trust statements are weighted, subjective, context dependent, and asymmetric (Massa & Avesani, 2004). For making predictions to an active user, trust-aware systems use the ratings made on different items by users trusted by the active user. While in collaborative filtering, ratings made by users similar to active user are used for making predictions. One of the major weaknesses of collaborative filtering system is their inability to calculate similarity between two users when numbers of co-rated items by both the users are few. Trust based systems overcome this drawback by using the concept of trust propagation (Massa & Avesani, 2004). Using the concept of trust propagation the system predicts the trust value between two users even if it has not been explicitly stated. The predicted trust value is dependent on the trust metric used. Trust metric can be global or local. Trust metrics in recommender system is one of the important research areas (Massa & Avesani, 2007b).

Prediction generation in trust-aware systems depend on the trust weightage between active user and other users connected to it, the propagation distance k and the ratings given by the trusted users to the item for which prediction is to be made. Users connected directly to the active user are said to be connected to the active user at trust propagation distance k=1or in other words users are present in web of trust level 1 for the active user. Users who are directly connected to trusted users of the active user at k=1 form the set of trusted users at trust propagation level 2 for the active user. As users at propagation distance k=2 are not directly connected to the active user, their trust value with active user is calculated using a trust metric. Final prediction for an item is made using the following formula

$$r_{ui}=\frac{\sum_{v\in N} T_{uv}\, r_{vi}}{\sum_{v\in N} T_{uv}} \quad (1)$$

Here r_{ui} denotes predicted rating of item i by user u., predicted rating for item i is the average of the rating given to i by those users who are connected to u at propagation distance k weighted by

their trust value with user u. T_{uv} is the trust value between user u and those users who have rated i and are present in the network N i.e. users connected to u at trust propagation distance k.

COLD START PROBLEM

Cold start problem for new user refers to the problem of making recommendations to users who are new to the system. Cold start problem is associated with new systems and new items also. In case of a new system it refers to the problem faced by a newly launched system in generating recommendations. Similarly, in case of items, it is hard to recommend newly introduced items to users as no ratings are available for them. This paper focuses on the cold start problem for new user. Cold start users constitute a large percentage of the total users. While solving the cold start problem for new user, two conflicting objectives are to be met. Firstly, accurate recommendations should be made to the user. Secondly, the preference elicitation process for a new user should not be bothersome to the new user. In this paper we primarily focus on the first objective of helping cold start user get better recommendations by studying the characteristics of users a cold start user trusts.

In traditional collaborative filtering based recommender systems, a user is initially asked to rate a few items, on the basis of which recommendations are made. Traditional systems are unable to make accurate recommendations for new users because to make accurate recommendations the system requires the user to rate a significantly large number of items. Recent work in trust aware recommender systems (Massa & Avesani, 2007a) has established that asking the new user to select a few trusted users from the network can result in more accurate recommendations than asking the cold start user to rate a few items. It has been shown that effort required from the cold start user during preference elicitation process is much less and number of recommendations that can be made are higher and more accurate as compared to traditional systems.

KEY PARAMETERS

Trust aware recommender systems approach for generating recommendations for a new user by inviting the user to select a few trusted users in the network is based on the assumption that a new user is aware of existing users in the system. This may not hold true in most cases. Even if the new user selects a few users as trusted users, the accuracy of recommendations that can be made and number of items for which predictions can be generated for the new user is dependent on the number of items rated by the chosen trusted users. Recommendation quality is also dependent on how connected the trusted users are in the trust network. Suppose a new user selects a trusted user who has rated only a few movies and is not connected to any other user in the network then the number and quality of recommendations made to the new user will be impacted negatively.

Prior research (Victor et al., 2008) has been shown that the accuracy of recommendations made to a new user is largely impacted by his web of trust. Web of trust of a user represent the set of users trusted by him. Therefore, it is important that new users are guided towards finding trusted users that are beneficial to him. In Victor, Cornelis, De Cock, and Teredesai (2008), they suggest providing a random list of key figures to the new user from which he can choose his set of

trusted users. There are three key figures defined in their approach, namely: mavens, frequent raters, and connectors. Mavens are people who write lot of reviews. Frequent raters are users who evaluate a large number of reviews. Connectors are a set of users with large number of outgoing links. Connectors are those users who have made trust statements on many other users. They show that a cold start user will be benefited if he connects to one of the key figures. The benefit of selecting a key figure as trusted user is measured by metrics like mean absolute error (MAE) i.e. difference of actual rating and predicted rating, and coverage i.e. defined as the number of items for whom predictions can be made for the user.

Key parameters selected by us to characterize the users trusted by cold start users were done in the context of a proper trust-aware recommender system environment. We use the dataset from Massa and Avesani (2006b), which has only user, item, rating and trust statement as the variables. As a result, the concept of mavens (Victor et al., 2008) is not applicable. The four key parameters that describe a user in a trust-aware network are as follows:

Number of Ratings

Number of items rated by a user has an impact on the accuracy and coverage of predictions made. A recommender system will definitely be able to make predictions for a large number of items if a cold start user selects a user who is a frequent rater of items. Frequent raters provide more information about their preferences, evaluating them will be much easier for a cold start user as compared to another user who has rated a few items. Ease of evaluation may make frequent raters more trustable than those who have rated few items.

Number of Outgoing Links

Number of outgoing links a user has represents how well the user is connected in the network. Cold start user connecting to a Connector, i.e. a user with large numbers of outgoing links, will result in more coverage. Because when a new user connects to a user with a high number of outgoing links at propagation distance 1, it gets connected to a large number of users at propagation distance 2. As propagation level increases, more users become part of the new user network. But trust value diminishes as it is propagated through the trust network [5], for better recommendations, a new user should be connected to as many users as possible at a shorter propagation level. Cold start users who want to be part of larger network may prefer to trust connectors.

Number of Incoming Links

In a trust based social network, users which are popular or more trusted are likely to have more number of incoming arrows than outgoing arrows. Users with large number of incoming links are the trusted users. Cold start user trusting a trusted user may not result in better accuracy or coverage. Because the way recommendations are generated in trust-aware recommender systems, users connected to frequent raters and connectors will get better coverage and prediction accuracy. New user may trust the trusted as the trusted user have a reputation of being trusted by many. Another motivation may be to be agreeable to the majority belief.

Hybrid Parameter

The final parameter we have used is a hybrid parameter that generates a preference score for a user using the three important elements of a trust based social network; number of incoming links, number of outgoing links, and number of items rated. The reasoning being a cold start user may prefer to trust users who are balanced in the three key parameters. We describe below our approach of calculating the preference score of a user.

Let I(u), O(u), and R(u) for a user u be the number of incoming links towards him, the number of outgoing links from him and the number of items rated by him. Let α, β, and γ be the yardsticks (minimum expected values) for the number of incoming links, outgoing links and the number of items rated. Let $I_{\alpha}(u)$, $O_{\beta}(u)$, and $R_{\gamma}(u)$ denote the scores for incoming links, outgoing links, and number of ratings respectively, which are computed as follows:

$I_{\alpha}(u)=int(I(u)/\alpha)$, $O_{\beta}(u)=int(O(u)/\beta)$, and $R_{\gamma}(u)=int(R(u)/\gamma)$, where int stands for integer division. Let σ(u) denote the preference score of a user u, then:

$$\sigma(u)=I_{\alpha}(u)*O_{\beta}(u)*R_{\gamma}(u) \quad (2)$$

For σ(u), we are using a multiplicative function for eliminating those users from consideration who have zero score for incoming links, outgoing links, or the number of ratings. Moreover, the preference score is designed in such a way that users who have balanced (almost equal) scores for incoming links, outgoing links, and number of ratings will have a better score over those users that have values skewed toward only one of the parameter.

EXPERIMENTAL SETUP AND METRICS

We performed the empirical evaluation of the four parameters used for creating the list of users on the Massa and Avesani (2006b) data set. This is one of the popularly used recommender system dataset and the only publicly available data set that has trust data. It consists of 50,000 users and 140,000 items. Total number of ratings is 660,000 and number of trust statements made is 490,000.Trust value is always one. Majority of users [53%] in dataset have rated less than 5 items. We consider users with less than 5 ratings as cold start users. A detailed analysis of the data set can be found at (Massa & Avesani, 2004, 2007a). From among the users trusted by the cold start users, we selected the most trusted 50 users, which we will refer as TU. We ranked the 50 users in set TU based on how many trust statements have been passed on them by cold start users i.e. user ranked 1 in set TU signifies that he been trusted the most number of times by the cold start users. Then, based on our four parameters incoming links, outgoing links, number of ratings, and preference score we created four ranking lists IR,OR,RR, and PR respectively. IR represents the ranking of the 50 users in set TU on the basis of the number of incomings links connected to them. Rank 1 in IR list suggests that the user has the highest number of incoming links among all the 50 users in set TU. Accordingly, OR represents the ranking list based on number of outgoing links, RR represents the ranking list based on number of movies rated and PR represents the ranking list based on the preference score. For finding the most important parameter among the four, we calculated the spearman rank-order correlation coefficient (rs)

between each of the four rank lists IR, OR, RR, and PR, and the TU rank list. The formula used for spearman rank-order correlation coefficient is as follows

$$r_s = 1 - \frac{6\sum D^2}{N\,(N^2 - 1)} \quad (2)$$

D^2 represents the square of difference between each pair of ranks and N represents number of values taken. As number of tied ranks are not very large the values of spearman correlation coefficient and Pearson correlation coefficient is same for all the four cases. For test of significance of the correlation coefficients obtained we used a test statistics α. The formula used for test statistics α with degrees of freedom N-2 is as follows

$$\alpha = \frac{r_s}{\sqrt{1 - r_s^2 / N-2}} \quad (3)$$

For comparing the four parameters on the basis of their impact on prediction accuracy of the, we conducted tests on the set of regular users that is those users who have rated at least 5 items. We first created four ranked lists, one for each of the four parameters. Each list consists of 450 users based on a particular parameter. For example the list for outgoing parameter will have the top 450 users ranked on the basis of number of outgoing links. Higher the number of outgoing links, higher the rank. We selected 450 top users as it represents 10% of all users in the data set that have been trusted by at least 20 users. We then select four test groups representing each of the four parameters based on a selection criterion. The following example describes our selection criteria. Let for user u_a the set of users trusted by him be represented as tt_a .Let the set of trusted users in tt_a that belong to the top 450 list of users based on outgoing links be represented as o_a , set of trusted users in tt_a that belong to the top 450 list of users based on incoming links be represented as i_a , set of trusted users in tt_a that belong to the top 450 list of users based on number of movies rated be represented as r_a, and set of trusted users in tt_a that belong to the top 450 list of users based on preference score be represented as s_a If the function n() represents number of users in a set, then user u_a will be categorized as a user with preference for trusting users with higher number of outgoing links only if

n(oa)/(n(oa)+n(ia)+n(ra)+n(sa))>0.35 (4)

n(oa)/n(tta)>0.30 (5)

n(oa)/(n(oa)+n(ia)+n(ra)+n(sa))>n(ia)/(n(oa)+n(ia)+n(ra)+n(sa)) (6)

n(oa)/(n(oa)+n(ia)+n(ra)+n(sa))>n(ra)/(n(oa)+n(ia)+n(ra)+n(sa)) (7)

n(oa)/(n(oa)+n(ia)+n(ra)+n(sa))>n(sa)/(n(oa)+n(ia)+n(ra)+n(sa)) (8)

By exchanging the positions of n(o_a) and n(i_a) in equation 4,5,6,7 and 8 , we get the conditions that need to be fulfilled by user u_a to be a categorized as a user preferring to trust users with large number of incoming links . Similarly, exchanging the positions of n(o_a) and n(r_a) and exchanging the positions of n(o_a) and n(s_a) in equations 4,5,6,7 and 8 we get the conditions that need to be fulfilled by user u_a for being categorized as a user preferring to trust users those are frequent

raters and those with high preference score respectively. The weightage of 0.35 and 0.30 used by us is specific to this dataset and will be different for each dataset.

After clustering the users on the basis of conditions discussed above, we created four test groups of users to represent the four parameters incoming links, outgoing links, movies rated, and preference score respectively. Each group consists of 10 users. In addition to comparing accuracy among the four parameters, we also examined the difference in accuracy by comparing the prediction value generated using only the set of trusted users representing a parameter with the predicted value obtained by considering all trusted users. For example, a user u_a who has been categorized as a user with a preference for trusting users with large number of outgoing links has 20 trusted users. Out of the 20 trusted users, 10 users are in the list of 450 top users with highest number of outgoing links. So, while testing accuracy for user u_a by predicting rating for an item i_t , we first calculate the prediction for i_t using only the 10 users which represent the parameter outgoing links and then again generate prediction for i_t using all the 20 trusted users. For the purpose of measuring the effectiveness of a parameter we choose the widely used metric mean absolute error (MAE) [2]. For a particular group representing a parameter, the MAE is calculated by averaging the MAE of each of the 10 user present in the group. We measured the MAE for different groups of users at trust propagation distance of 1. Trust value used was 1.

RESULTS AND DISCUSSION

In Table 1 below we report the spearman rank-order correlation values obtained. Scatter plots between the rank list in TU set and rank lists in set IR, OR, RR, and PR are shown in Figure 1, Figure 2, Figure 3, and Figure 4 respectively.

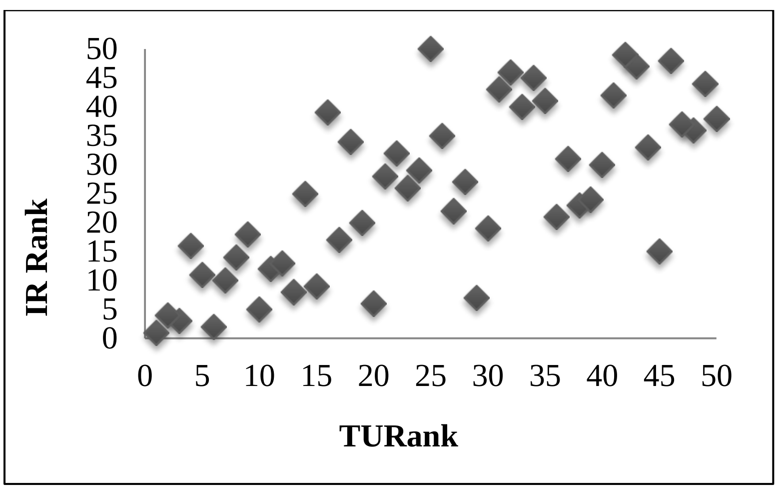

Figure 1: Scatter Plot of Rank in TU Set Plotted Against Rank in IR Set of Users Trusted by Cold Start Users.

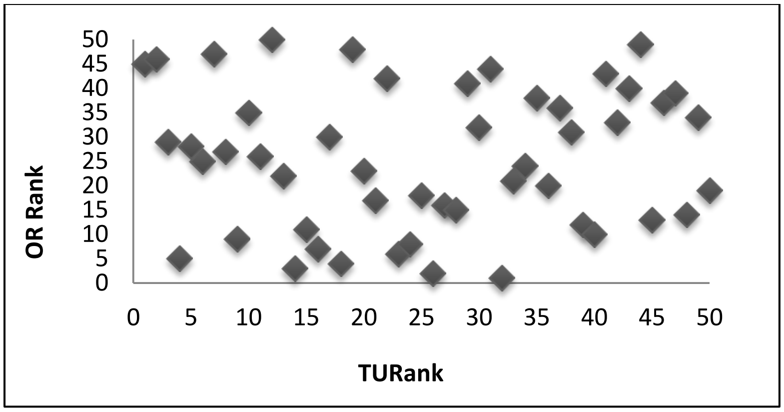

Figure 2. Scatter Plot of Rank in TU Set Plotted Against Rank in OR Set.

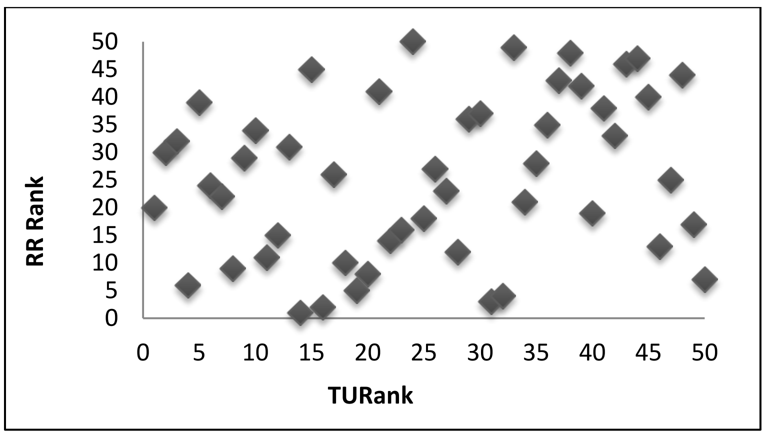

Figure 3. Scatter Plot of Rank in TU Set Plotted Against Rank in RR Set of Users Trusted by Cold Start Users.

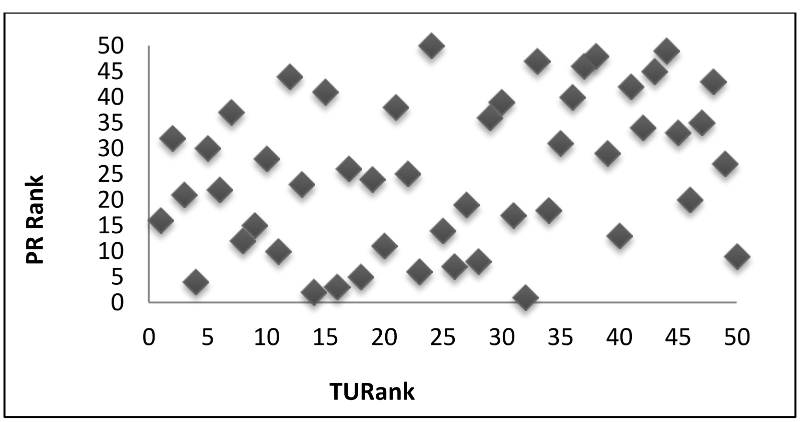

Figure 4. Scatter Plot of Rank in TU Set Plotted Against Rank in PR Set of Users Trusted by Cold Start Users.

IR	OR	RR	PR
0.71678*	0.30972*	0.25147**	0.31265*

* Correlation coefficient significant at 0.01 confidence level.
** Correlation coefficient significant at 0.05 confidence level.

These values of significance are based on one tailed significance test.

Table 1: Rank Correlation Coefficients between TU and the Rank Lists Based on the Four Parameters.

	Outgoing Links	**Incoming Links**	**Movies Rated**	**Preference Score**
MAE 1*	1.3464	1.1202	0.8452	**0.7615**
MAE 2**	1.3609	0.9876	0.8021	**0.7352**

***** Represents prediction error based on prediction made using only those trusted users that represent the parameter

** Represents prediction error based on prediction made using all the trusted users of the test user

Table 2: MAE Values of the Four Parameters.

Figure 1 clearly shows high positive correlation between rank of users trusted by cold start users and the rank of these users based on number of incoming links. The correlation score of 0.7168 between IR and TU is the only correlation score above 0.5 and is also significant at 99% confidence level. We can conclude from the results that among the set of users most trusted by cold start users, those users with high number of incoming links are more likely to be trusted than those users that score better in the other three parameters. Cold start users are more likely expected to trust the trusted as compared to the frequent raters and connectors.

The results indicating the importance of number of incoming links as the most critical parameter that cold start users consider while evaluating trustworthiness of users in the trust network is quite surprising, because previous research in guiding new user find trusted users in the trust network never considered incoming links as a key attribute. The reasoning being, in earlier research the criteria for selecting key parameters were based on the logic of improving the accuracy and coverage of recommendations made to a cold start user, and selecting user with high number of incoming links does not necessarily impact coverage or accuracy in a positive way. Results of our experiment to evaluate the four parameters on the criteria of prediction accuracy are shown in Table 2.

Those users which are categorized as users preferring to trust users with high preference score show the best predictive accuracy. While users which preferred to trust user with higher number of outgoing links i.e. connectors have the worst MAE values. We also analyzed the difference in

predictive accuracy between the predicted value obtained using all the trusted users for a test users and the predicted value obtained using only those trusted users representing the parameters. The difference is very minimal, the reason may be that in most of the test users the set of trusted users representing a particular parameter have a majority i.e. they constitute more than 50% of all users trusted by the test user. So very few test items have been rated by trusted users that do not represent a parameter.

We believe the reasons behind cold start users preferring trusted users over frequent raters and connectors cannot be explained through data. Research in electronic commerce has found that trust is strongly related to information disclosure (Metzger, 2004), this should have resulted in users with higher number of movies being preferred as trusted users. On the contrary, new users prefer trusting users who are already trusted by many, even if their taste does not match with the trusted user. This behavior can be better explained with sociological concept of trust. From sociological perspective, trust has three distinctive analytical dimensions: cognitive, emotional, and behavioral. Unlike research stated earlier that in ecommerce environment information disclosure leads to trust, from sociological perspective, no matter how much additional knowledge of an object one may gain, additional knowledge alone can never cause us to trust someone (Lewis & Weigert, 1985). The manifestation of trust on the cognitive level of experience is reached when their trust on an object is no longer affected with further information. The user or social actor no longer wants or needs further information or reasoning for trusting the object. The cognitive element in trust is characterized by a cognitive leap that cannot be explained by reason and experience alone. This cognitive base of trust lies in trust in trust (Luhmann, 1979). Each trusts on the assumption that others trust. As in real life, people are more likely expected to trust those who are trusted by others, the same concept of trust has diffused into a user decision-making process while selecting trusted users in social network. In the Epinions site (Massa & Avesani, 2006b), a trust statement signifies a similarity in ratings of both users for common items, but actual data shows a divergence in the way trust is understood by the user. Analyzing the Massa and Avesani dataset, we observe that for majority of trust statements made, correlation between the source user and target user can be calculated only for a small fraction. Out of 487,183 trust statements made, only 2.91% have a similarity value of greater than 0.5 and at least 4 co-rated items between source user and target user. This observation that majority of the trust statements passed between users cannot be seen as a reflection of similarity between the two users leads us to believe on the sociological concept of cognitive leap based on the logic of trust in trust. One limitation of our work is that we have made the analysis on only one dataset i.e. Epinions dataset., it would be interesting to see the results of analyzing the set of users trusted by cold start users in a different recommender system data set. Particularly, in data sets like Allconsuming.net, as in (Ziegler & Lausen, 2004) they have empirically shown correlation between trust and similarity in the Allconsuming.net dataset.

It is evident that users prefer to trust users that are trusted i.e. users with higher number of incoming links. However, trusting existing trusted users may not help cold start users get better recommendations as we have shown trusting users with high preference score will lead to better accuracy. Therefore, to make better recommendations as measured on accuracy, users need to be encouraged to pass trust statements only on those with whom their preferences actually match or those user with high preference score. Our research recommends that sites like Epinions should use terminology like similar or same interests instead of asking users to trust another user

because users perceive the terminology of trust as social trust as a result they behave accordingly. Getting users to connect to users that are similar to them will help in providing better recommendations. This can be achieved only by educating the user on concept of trust as used in the particular social network.

CONCLUSIONS

In this paper we explore the characteristics of those users that cold start users or new users actually trust. We conclude that as in real life, people are more likely expected to trust those who are trusted by others, the same concept of trust has diffused into a user decision making process while selecting trusted users in social network even though the concept of trust in web based social network is site specific and is different from social trust. In Epinions website network, a trust statement signifies a similarity in ratings of both users for common items, but actual data from the website shows a divergence in the way trust is understood by the user. We analyze the trust connections in Epinions dataset of the cold start users on the basis of four parameters: number of outgoing links, number of incoming links, number of items rated and a hybrid parameter defined as preference score. Our analysis shows that cold start users will more likely select a user with higher number of incoming links as a trusted user as compared to frequent raters and connectors. We also show that trusting users with high preference scores will give more accurate recommendations. We recommend that web based social networks should be careful while using the term trust in their websites, particularly when the concept of trust used in their website is different from the concept of social trust.

One limitation of our work is that we have made the analysis on only one dataset i.e. Epinions dataset., it would be interesting to see the results of analyzing the set of users trusted by cold start users in a different recommender system data set. We also want to explore in future whether the cold start user is influenced by distrust statements when selecting users to trust.

REFERENCES

Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication,13*(1). Retrieved from http://jcmc. indiana.edu/vol13/issue1/boyd.ellison.html

Gambetta, D. (1990). *Trust: Making and breaking cooperative relations*. Oxford, UK: Blackwell.

Golbeck, J. A. (2005). *Computing and applying trust in web-based social networks* (Doctoral dissertation). Retrieved from http://drum.lib.umd.edu/bitstream/1903/2384/1/umi-umd-2244.pdf

Lewis, J. D., & Weigert, A. (1985). Trust as a social reality. *Social Forces, 63*, 967-985. Doi: doi: 10.1093/sf/63.4.967

Luhmann, N. (1979). *Trust and power*. Hoboken, NJ: Willey and Sons.

Marsh, S. P. (1994). *Formalising trust as a computational concept* [Doctoral dissertation]. Retrieved from http://www.cs.stir.ac.uk/~kjt/techreps/pdf/TR133.pdf

Massa, P., & Avesani, P. (2004). Trust-aware collaborative filtering for recommender systems. *Proceedings of the Federated International Conference on the Move to Meaningful Internet*, 492-508.

Massa, P., & Avesani, P. (2006a). Trust-aware bootstrapping of recommender systems. *Proceedings of 2006 ECAI Workshop on Recommender Systems*, 29-33.

Massa, P., & Avesani, P. (2006b). Trust-aware bootstrapping of recommender systems [Data file]. *Proceedings of 2006 ECAI Workshop on Recommender Systems*, 29-33. Retrieved from http://www.trustlet.org/wiki/Downloaded_Epinions_dataset

Massa, P., & Avesani, P. (2007a). Trust-aware recommender systems. *Proceedings of the 2007 ACM Conference on Recommender Systems*,17-24. Retrieved from http://brettb.net/project/papers/2007%20Trust-aware%20recommender%20systems.pdf

Massa, P., & Avesani, P. (2007b). Trust metrics on controversial users: Balancing between tyranny of the majority and echo chambers. *International Journal on Semantic Web and Information Systems, 3*(1), 39-64. doi: 10.4018/jswis.2007010103

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*, 709-734.

Metzger, M. J. (2004). Privacy, trust, and disclosure: Exploring barriers to electronic commerce. *Journal of Computer-Mediated Communication*, *9*(4). Retrieved from http://jcmc.indiana.edu/index.html

Mobasher, B., Burke, R., Bhaumik, R., & Williams, C. (2007). Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. *ACM Transactions on Internet Technology*, *7*(4). doi: 10.1145/1278366.1278372

Victor, P., Cornelis, C., De Cock, M., & Teredesai, A. M. (2008). Key figure impact in trust-enhanced recommender systems. *AI Communications, 21*(2-3), 127-143.

Ziegler, C. -N., & Lausen, G. (2004). Analyzing correlation between trust and user similarity in online communities. *Proceedings of the Second International Conference on Trust Management*, 251-265.