

# IntRank: Interaction Ranking-Based Trustworthy Friend Recommendation

Lizi Zhang   Hui Fang   Wee Keong Ng   Jie Zhang  
*School of Computer Engineering*  
*Nanyang Technological University, Singapore*  
*Email: {y080077, hfang1, awkng, zhangj}@ntu.edu.sg*

**Abstract**—Social networks are fundamental to virtual communities (e.g., forums, blogs) and virtual communities benefit from well-established social networks. As making friends with other members is a common way to establish social relationships and people need to decide whom they should trust when making friends, friend recommendation has received considerable attention in virtual communities. Towards this goal, we first formulate hypotheses on factors that influence trust and the probability of establishing friendships from various interaction attributes in virtual communities. Through experiments on real interaction and friendship data, we validate the proposed hypotheses and propose a novel interaction ranking-based trustworthy friend recommendation model called IntRank for recommending trustworthy friends to community members. Different from traditional friend recommendation mechanisms, IntRank is built on the foundation of carefully verified interaction attributes that influence trust and friendship probability in virtual communities. It is able to effectively recommend trustworthy friends as confirmed by the performance evaluation results.

**Keywords**—Trust; Social network mining; Friend recommendation; Interaction attributes; Logistic regression

## I. INTRODUCTION

Virtual communities (also known as online communities such as forums, blogs, newsgroups, social networks, etc.) have brought different people together to interact with one another. People share information, express opinions, exchange ideas, make friends, and therefore form social networks in virtual communities. According to Garton *et al.*'s definition, a social network is "a set of people [...] connected by a set of social relationships such as friendship, co-working or information exchange" [9]. Among various virtual communities, *Slashdot*<sup>1</sup> introduced the *Zoo* feature in January 2002. This new feature allows its users to tag other community members as either friends or foes. A user declares other users as friends if he approves of their comments and finds them trustworthy. Members in a user's foe list are those whose comments he disapproves and distrusts [16], [2]. Previous studies on *Slashdot* showed that nearly 80% of the relationships are friendships [16], [14]. This indicates that people tend to form positive relationships (friendships) with other community members. *Slashdot Zoo* is considered to be one of the earliest examples of virtual communities with a social networking component that encourages friendships (by declaring friends) among members

[14]. Other popular virtual communities, to name a few, *Facebook*, *MySpace*, *Essembly* and *Friendster*, all encourage friend connections and have nurtured rich social networks within their communities.

Previous studies have shown that virtual communities benefit from well-established social networks in various aspects. These can be summarized as follows. First of all, well-established social networks within a community are key to stimulating users' interests in contents and therefore their contributions in knowledge supply. Well-established social networks help to form a well-connected virtual community as the social networking component allows people to connect one another and influence their discovery of relevant contents and products [4]. A virtual community also sustains well by having more participants with strong commitment driven by the social networks inside [2]. Zadeh *et al.* pointed out that well-established social networks within a well-connected virtual community make people feel useful and a sense of belonging to the community; thus, they are able to make real contribution to the community [25].

Second, a community benefits from its well-established social network financially. A study by Bernier and Ganley has shown that encouraging a more supportive and better connected virtual community through social networking mechanisms help users feel committed to the community and more willing to pay subscription fees [2]. This study also suggested that users' financial investments on their virtual community may subsequently lead to improved participation quality [2], where a virtuous cycle of activities, social networks and subscriptions can be formed. In addition, Ganley and Lampe [8] further pointed out that since the business model of *Slashdot* and many other Web 2.0 sites are based on advertising, a critical determinant of revenue potential is the amount of activities on the site. They also suggested that the amount of activities is best tracked by the quantity and quality of membership, while high quantity and quality of membership is reflected by well-established social networks.

Third, well-established social networks can be further applied to recommender systems. Driven by the intuition that "if I like that person I might also be interested in his content", Sinha and Swearingen have proven that users' friends consistently provided better recommendations than traditional recommender systems which rely on collaborative

<sup>1</sup><http://www slashdot.org>

filtering [22]. Brzozowski *et al.* have further found that “the closer a user is to recommending friends, the stronger the persuasion is likely to be” [4]. Thus, beyond developing better recommender algorithms, one could look into leveraging people recommendations for recommending contents [6].

In fact, the continuously growing size of community members and information with widely varying quality in virtual communities has also raised a critical issue among community members when they make friends with others: Whom should I trust? Therefore, on top of the above benefits that well-established social networks bring to virtual communities, each community member also benefits from effective automatic trustworthy friend recommendation mechanisms since they need to know whom they should trust when making friends. In addition, a survey found that 93% users on *Slashdot* are non-subscribers and they can only declare a limited number of friends [2]. To make use of the limited friend list size, people have to spend a lot of time to manually pick the most trustworthy friends in their own opinions. Obviously, such manual decisions are not scalable as the size of a community can be very large. Hence, an automatic trustworthy friend recommendation mechanism helps users save time and maximize their benefits correspondingly.

As such, the aforementioned motivations provide impetus to recommend the most trustworthy friends to community members which in turn encourages friendship formation and social networking. This has received considerable attention in recent years. First, traditional reputation systems compute users’ trust and reputation values and only recommend those that have high trust and reputation values [1], [24]. However, they fail to incorporate useful user interaction information which we shall examine in this work. Traditional recommender systems predict user preferences based on ratings and have been proven to be vulnerable to unfair ratings, discrimination and other problems. Besides, link predictions and other mechanisms [14], [17] based on a “friend-of-a-friend” approach predict friendships in terms of existing social networks. However, link prediction has many limitations in most real virtual community environment and “friend-of-a-friend” approach may lead to negative network effects like “rich getting richer”. A more detailed review of these friend recommendation mechanisms will be given in Section 2.

Bernier and Ganley suggested that a social network should be both “a precursor to and a manifestation of trust” [2]. In other words, users declare trustworthy people as their friends. Mui *et al.* defined trust as a “subjective expectation an agent has about another’s future behavior based on the history of their encounters” [19]. It emphasizes that trust should arise from direct interactions. In view of various limitations within previously proposed mechanisms, we want to recommend trustworthy friends to a particular user  $U$  from the other users with whom he has direct interactions in the past. From these interactions, we evaluate the trust  $U$  has

in these users. This gives the probability that  $U$  would like to establish friendships with them. We want to find out people with whom  $U$  will most likely make friends and recommend them to  $U$  as trustworthy friends. Towards this goal, we first investigate various interaction attributes and propose hypotheses on how they influence trust and friendship probability in virtual communities. Using real interaction and friendship data collected from *Slashdot*, we perform logistic regression analysis to evaluate the correlation between the proposed interaction attributes and friendship probability. Based on the validated hypotheses, we propose a novel interaction ranking-based recommendation model called IntRank for recommending trustworthy friends to community members, and evaluate its performance on the collected data.

Our investigation reveals four interaction attributes that influence trust and friendship probability in virtual communities, and reinforces the fact that subjective manual ratings are not accurate for evaluating trust and friendship probability among community members. As a result, we demonstrate four new perspectives of interactions influencing trust and friendship probability among members in virtual communities: interaction frequency, interaction quality, seriousness in interactions, and common interest. Different from traditional friend recommendation mechanisms, the proposed model IntRank is built on the foundation of carefully verified interaction attributes that influence trust and friendship probability in virtual communities. It is able to effectively recommend trustworthy friends as confirmed by the performance evaluation results.

## II. RELATED WORK

Traditional reputation systems for e-commerce allow users to rate one another or rate products or services based on their quality [11]. From these ratings, reputation systems compute users’ trust or reputation values, and users with high trust or reputation will be recommended. Typically, interactions, especially communications between members take place frequently in virtual communities. In this work, we show four different types of interaction attributes influence friendship probability in virtual communities. Without incorporating members’ rich interaction information (*e.g.* commenting on others’ postings, the length of users’ comments, and the time when the comments are provided), rating-based reputation systems working in e-commerce are not able to reflect a virtual community member’s trustworthiness objectively or accurately.

Major types of recommender systems have also been studied [24]. These recommender systems heavily rely on ratings given by users in order to predict their preferences [1]. For example, collaborative filtering systems require users to rate items to express their preferences. Then the systems find users with similar preferences according to different algorithms such as cosine similarity and Pearson correlation, and generate recommendations for users based

on their similarity. Chen *et al.* argued that these techniques have potentials to be used for recommending people in virtual communities [6]. However, since these systems heavily rely on subjective rating values, they are vulnerable to unfair ratings, discrimination and other problems [11]. On top of that, traditional recommender systems may suffer from the “sparsity problem”, *i.e.*, in reality two users are unlikely to have many similarities and therefore the range of recommendation partners is usually limited [20].

Researchers in the graph mining area have applied link prediction for social networks to recommend friends in virtual communities. Liben-Nowell and Kleinberg defined the link prediction problem as “Given a snapshot of a social network at time  $t$ , we seek to accurately predict the edges that will be added to the network during the interval from time  $t$  to a given future time  $t'$ ” and discussed various link prediction methods [17]. As verified by Kunegis *et al.*, *Slashdot Zoo* is a small world network where the measured average distance is less than the average distance in a random graph [14]. Some link prediction methods such as the basic graph distance predictor are not effective in virtual communities with small world networks [17]. Besides, Popescul *et al.* also pointed out that social network data are extremely noisy and the characteristics useful for link prediction are not readily available [21].

To facilitate the friend recommendation process, *Facebook* has launched a feature called “People you may know” to recommend friends based on a “friend-of-a-friend” approach which has been commonly used in many different virtual communities. However, a recent work by Daly *et al.* demonstrated that this mechanism leads to the “rich getting richer” problem and thus may decrease the value of the social network [7]. In other words, under this mechanism a user with a wider social circle has a higher probability of friendship overlap and therefore may be recommended frequently to many different users. This bias to already well connected individuals will result in the recommended users becoming even more connected.

Massa and Avesani proved that due to the significant proportion of controversial users (are trusted and distrusted by many), a global agreement of the trust value of these users cannot exist [18]. They prefer a local view in the prediction of the trustworthiness of a user in a personalized manner; that is, user B develops trust for A and chooses A as his friend based only on B’s personal view of his direct interactions with A. Compared with those previous mechanisms discussed above, our proposed model IntRank is based on the direct interaction information; thus, it objectively and accurately evaluates trust and the probability to establish a friendship, and is able to recommend trustworthy friends from people with whom a user has interacted in the past.

### III. HYPOTHESES

In virtual communities, choosing friends could be influenced by many factors. Considering friendships manifest

trust and trust arises from interactions, we explore five interaction attributes between a user pair: reply frequency, comment length, comment score, time difference between a posting and a reply, and domain similarity of user A’s replies to B. Using these interaction attributes, we formulate five hypotheses of how these interaction attributes influence B’s trust in A and the probability that B would like to establish a friendship with A.

#### A. Reply Frequency

In the context of virtual communities, we note that having a continuous supply of knowledge from members is the biggest challenge; many virtual communities failed due to members’ low willingness to share knowledge with others [5]. Following the interpretation of trust in virtual communities [23]—which is “openness to discussion and willingness to share data”, we use the reply frequency  $X_f^{\Delta T}(A, B)$  from user A to B as an indicator of A’s trust from the point of view of B during period  $\Delta T$ . It is calculated as:

$$X_f^{\Delta T}(A, B) = \begin{cases} N^{\Delta T}(A, B)/\Delta T' & N^{\Delta T}(A, B) > 1; \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

where  $N^{\Delta T}(A, B)$  is the number of replies from A to B during time period  $\Delta T$  and  $\Delta T'$  is the time difference between A’s first and last reply to B. Within a certain time frame, if user A produces more frequent direct replies than C to B, then from the point of view of B, A is more open to discussion and more willing to share opinions, which also suggests more efforts in the “continuous supply of knowledge” [5]; thus, A is more trustworthy in terms of B and B is more probable to choose A as his friend. Hence, we formulate the first hypothesis as follows:

- H1 During time period  $\Delta T$ , the frequency of A’s replies to B’s postings  $X_f^{\Delta T}(A, B)$  influences B’s trust in A and the probability that B would like to establish a friendship with A; the higher the reply frequency from A, the higher the probability that B would like to choose A as his friend.

#### B. Comment Length and Score

An interaction between users A and B is said to be of high quality if A provides useful and insightful information to B during the interaction, and *vice versa*. The higher interaction quality a user has with another, the more he can be trusted and the higher probability he will be chosen as a friend. Generally speaking, in virtual communities, interaction quality between two users can be reflected by the length of a comment from one user to another and the score the comment received. The longer a user’s comment is, the higher the probability that the comment contains rich and helpful information. A user who usually writes long comments is perceived to be more trustworthy as the recipient is more likely to receive valuable information. Thus, the long-comment writer is more probable to be

chosen by the recipient as his friend. On the score or rating of a comment received, a user whose comments always receive high scores tend to be perceived as more trustworthy, and likewise, more probable to be chosen by the reply recipient as his friend. Hence, we propose the following two hypotheses:

- H2 During time period  $\Delta T$ , the average length of comments  $X_\ell^{\Delta T}(A, B)$  provided by user A to B influences B's trust in A and the probability that B would like to establish a friendship with A; the longer the comments from A, the higher the probability that B would like to choose A as his friend.
- H3 During time period  $\Delta T$ , the average score of comments  $X_s^{\Delta T}(A, B)$  provided by user A to B influences B's trust in A and the probability that B would like to establish a friendship with A; the higher the score of comments, the higher the probability that B would like to choose A as his friend.

### C. Time Difference

Previous studies have analyzed temporal patterns of time difference between a user's posting and the comments (replies) provided by other users to the posting [12], [13]. Although it was not verified, Skopik *et al.* suggested that the time intervals between a posting and its replies may be useful in trust evaluation [23].

Consider the following scenario: When user B posts something, user C replies more quickly than user A to B's posting. From B's point of view, who is more trustworthy and more worthwhile to be his friend? There are two possibilities. Clearly, user C shows more eagerness to reply—this may suggest C's active involvement. On the other hand, an immediate reply from C may also indicate that C has not properly digested the posting from B; thus, user C can be seen to be exhibiting a casual attitude. Replies that came later could be due to more care and attention given to properly comprehend the posting from B. Thus, A could be seen to be more serious in the interaction, more trustworthy, and more probable to be chosen as B's friend. Based on these two possibilities, we propose the third hypothesis:

- H4 During time period  $\Delta T$ , the average time difference  $X_t^{\Delta T}(A, B)$  between user A's replies to user B's postings influences B's trust in A and the probability that B would like to establish a friendship with A; the longer the time difference, the higher the probability that B would like to choose A as his friend.

### D. Domain Similarity

The basic idea about the friend recommendation algorithm based on content matching comes from the intuition that if two people both post content on similar topics, they might be interested in getting to know each other. A previous survey found that 74.4% users said common content of interest would make them more likely to connect with others [6].

In other words, people tend to trust and make friends with those who have common interest with them. We argue that the greater the proportion of user A's replies to B's postings falling within A's favorite domain (*i.e.*, the domain of most of A's postings or replies), the higher the probability that A is chosen as B's friend. This can be further explained as follows. Previous work has shown that in this case user B plays the activator role [23] whose postings are worthy of discussion. Since B's postings attract replies from others (user A in this case), B is also considered to possess the expertise or competencies under this domain; this indicates that this domain is probable of B's interest too. Therefore, if most of A's interactions with B take place in A's favorite domain (which is also of interest to B), we can conclude that they share common interest; thus, A is trustworthy in terms of B and B tends to make friends with A.

In this paper, we use domain similarity to express similarity in users' interest. The domain similarity of user A from B's viewpoint is determined as follows:

$$X_d^{\Delta T}(A, B) = s^{\Delta T}(A, B)/N^{\Delta T}(A, B) \quad (2)$$

where  $s^{\Delta T}(A, B)$  is the number of replies (that are in the domain where A posts the most in the whole community) from A to B during time period  $\Delta T$  and  $N^{\Delta T}(A, B)$  is the number of replies from A to B during  $\Delta T$ . Clearly,  $X_d^{\Delta T}(A, B) = 1$  means that all of A's replies to B belong to both A and B's common domain of interest while  $X_d^{\Delta T}(A, B) = 0$  means that there is maximum diversity between A and B's interest. As such, the fifth hypothesis is formulated as follows:

- H5 During time period  $\Delta T$ , the domain similarity  $X_d^{\Delta T}(A, B)$  of all of user A's replies to user B influences B's trust in A and the probability that B would like to establish a friendship with A; the larger the domain similarity, the higher the probability that B would like to choose A as his friend.

## IV. HYPOTHESIS EVALUATION

This section evaluates the five hypotheses proposed in the previous section using interaction and friendship data from *Slashdot*. We first describe the data that are collected from the website. We then perform logistic regression analysis to validate the hypotheses and analyze the results.

### A. Data Preparation

*Slashdot* is a forum for posting news and comments with a distinct, technology-centric culture. Once news has been posted, anyone may provide comments to the news or to other users' comments. Each news or comment can be posted in a various domain such as games, hardware, mobiles, stories, book reviews, and so on. *Slashdot* has introduced a moderation system to maintain the quality of postings. This consists of two layers where M1 is for moderating comments to news, and M2 is for moderating

M1 moderators. In fact, *Slashdot* has been chosen as an ideal example in various studies on virtual communities in recent years [16], [14], [23], [11], [2], [8]. To validate the proposed hypotheses, we focus on the correlation between various interaction attributes and friendship probability in *Slashdot*. We extract data from *Slashdot* containing 102,199 comments written by 11,117 users from December 20, 2003 to February 23, 2011 across all domains by randomly sampling users. The postings of each user are continuous with respect to certain time interval. The comment contents, comment scores, comment posted time, comment domains, and the user pairs (comment sender and comment recipient) are all included. After extracting these users' friend list data from their profile pages respectively, we add 77,050 friend pairs into the database (user A and B form a friend pair if A is in B's friend list, and *vice versa*). To improve the quality and representativeness of the data, we filter out anonymous comments, replies to anonymous comments, and self-replies.

### B. Experiments

We notice there exist comment records with extremely diversified interaction attribute values. For example, due to the existence of extremely long or short comments, there may be outliers that seem most influential in determining the degree of influence on friendship probability. However, they appear incidentally and often mislead the estimation of the coefficient. Hence, without distorting the overall distribution of data, we filter out the top and bottom 0.1% records ranked by reply frequency, comment length and comment time difference where extremely diversified interaction attribute values may appear.

After pre-processing the data as illustrated above, we continue to normalize data to ensure that they are in the same order of magnitude. To evaluate the influence of each interaction attribute on friendship probability based on the proposed hypotheses, we establish the following logistic regression model:

$$\begin{aligned} \text{logit}(p) = \ln[p/(1-p)] = & \beta_0 + \beta_f X_f^{\Delta T}(A, B) \\ & + \beta_\ell X_\ell^{\Delta T}(A, B) + \beta_s X_s^{\Delta T}(A, B) \\ & + \beta_t X_t^{\Delta T}(A, B) + \beta_d X_d^{\Delta T}(A, B) + \varepsilon \end{aligned} \quad (3)$$

In this model,  $p$  stands for the probability that user B chooses A as his friend. We use 1 to indicate the fact that A is in B's friend list (in the captured 77,050 friend pairs) while 0 means A is not chosen by B as his friend. Accordingly, here  $p$  measures the probability that the value 1 appears. The symbols  $\beta_f$ ,  $\beta_\ell$ ,  $\beta_s$ ,  $\beta_t$ , and  $\beta_d$ , are the coefficients of each factor (interaction attribute). The symbol  $\beta_0$  is a constant and  $\varepsilon$  is an error term representing factors which cannot be directly observed or easily quantified.

### C. Result Analysis

The logistic regression results of the established model (Eq. 3) are presented in Table 1. As we choose alpha to be 0.05 with two-tailed, any factor with the  $P$  value

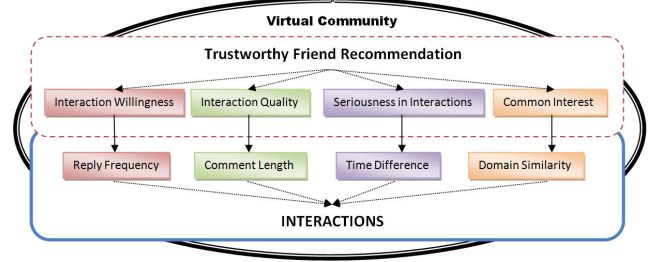


Figure 1. Recommending trustworthy friends from interactions in virtual communities.

Table I  
RESULTS OF THE LOGISTIC REGRESSION MODEL

Iteration 0:	Log Likelihood	-1540.3229	
Iteration 1:	Log Likelihood	-1524.4577	
Iteration 2:	Log Likelihood	-1513.804	
Iteration 3:	Log Likelihood	-1513.6922	
Iteration 4:	Log Likelihood	-1513.6921	
Number of observations		41,975	
LR $\chi^2(5)$		53.26	
Prob > $\chi^2$		0.0000	
	Coefficient	$z$	$P >  z $
<b>Reply Frequency</b>	2.961	2.33	0.020
<b>Comment Length</b>	1.646	3.76	0.000
<b>Comment Score</b>	-0.009	-0.03	0.979
<b>Time Difference</b>	1.762	2.30	0.022
<b>Domain Similarity</b>	0.736	5.70	0.000
<b>Constant</b>	-5.617	-26.93	0.000

greater than 0.05 will be rejected since this indicates a non-significant correlation. Thus, all the proposed interaction attributes except for comment score in the model have significant influence on friendship probability. In the following, we evaluate each of them based on the proposed hypotheses in Section 3.

1) *Reply Frequency*: From Table 1, the positive coefficient (2.961) suggests that the reply frequency is positively correlated with friendship probability. This validates the first hypothesis (H1). As an indicator of openness to discussion and willingness to share data, reply frequency determines whether a user can be trusted and the probability that he would be chosen as a friend by others. Our result shows that if user A always frequently replies to user B's postings in the past, B will be more probable to trust A and choose A as his friend since A is willing to share opinions and supply knowledge.

2) *Comment Length*: Given the positive coefficient (1.646), the comment length has a positive influence on friendship probability. This supports the second hypothesis (H2). As an indicator of interaction quality, comment length determines whether a user is trustworthy and the probability to be chosen as a friend. Our result shows that if user A always gives long comments (replies) to user B's postings in the past, B will be more probable to trust and make friends

with A since from interactions with A, it is highly possible for B to gain a lot of valuable information.

3) *Comment Score*: Contrary to Hypothesis H3, comment score does not show significant influence on friendship probability ( $P$  value = 0.979). Previous studies showed low score comments may be hidden and comments with a score of 1 or 2 may not have been rated by many moderators [23], [15]. Hence, we re-test our model after removing all low scored ( $-1$  and  $0$ ) comments and filtering out potentially non-rated comments by sampling. From the experiments, the  $P$  value decreases to 0.603; this still suggests the rejection of H3.

In fact, many previous studies have pointed out the limitations of the *Slashdot* moderation system. According to Brennan *et al.*, classifying a comment into a specific score may involve much noise and the benefits of classifying a comment as 4 instead of 5 are negligible towards improving the interaction quality [3]. The study by Lampe and Resnick showed that low score comments, non top-level comments, or late posted comments are likely to be overlooked by moderators [15]. In other words, this reveals the existence of buried treasures; *i.e.*, comments that should have high scores but did not, also causes some trash to be surfaced. Our experiment further reinforces that subjective manual ratings, and thus, models relying on them may not be accurate or reliable.

4) *Time Difference*: The result of experiment proves H4, as suggested by the positive coefficient 1.762. We hence argue that a late reply suggests longer time incurred to digest comments and provide new insights. The serious attitude shown in late replies suggests their providers are trustworthy and people prefer to make friends with them.

5) *Domain Similarity*: The positive correlation coefficient of 0.736 validates H5. We may conclude that people with common interest tend to trust each other and become friends. As interest also indicates rich experiences or expertise, opinions on certain domains from people with common interest can be favored and considered as more trustworthy and valuable; this increases the probability that they are trusted and chosen as friends by other people with common interest.

## V. INTERACTION RANKING-BASED TRUSTWORTHY FRIEND RECOMMENDATION

The previous two sections have identified and validated that reply frequency (interaction willingness), length of comments (interaction quality), time difference (seriousness in interactions), and domain similarity (common interest) influence trust of user A from the viewpoint of user B and the probability that A is chosen by B as a friend. These factors provide additional inputs for evaluating the probability that one would like to establish friendships with other community members from their interactions in virtual communities (Fig. 1). In reality, among all the users who replied to B, only a very small portion are chosen by B as friends. Due to the

extremely imbalanced distribution of friendship and non-friendship among all the users that have interactions with B, most machine learning approaches are not effective to recommend a friend to B by predicting whether a user is B's friend using the four interaction attributes. For instance, we find the Hidden Naive Bayes model tends to "blindly" claim that any user is not B's friend. With these, we propose an interaction ranking-based recommendation model called **IntRank** in Alg. 1, to recommend trustworthy friends in virtual communities.

### Alg. 1 : Interaction Ranking-Based Trustworthy Friend Recommendation Model: IntRank

**Input** :  $U$ , the user to whom IntRank recommends friends;  
 $\vec{S}$ , all the users who replied to  $U$  in the past;  
 $t$ , the timestamp when IntRank recommends friends to  $U$ ;  
 $k$ , the number of friends recommended to  $U$ ;  
**Output**:  $\vec{S}^k$ , the  $k$  most trustworthy friends recommended to  $U$  by IntRank;

$p^{\Delta T}(U, \vec{S}) = \emptyset$ ;

$(w_f, w_\ell, w_t, w_d, w_c) \leftarrow \text{learnCoefficients}(t)$ ;

$\Delta T \leftarrow \text{getTimePeriod}(t)$ ;

**foreach**  $S_i$  in  $\vec{S}$  **do**

$$\chi^{\Delta T}(U, S_i) = w_f X_f^{\Delta T}(U, S_i) + w_\ell X_\ell^{\Delta T}(U, S_i) + w_t X_t^{\Delta T}(U, S_i) + w_d X_d^{\Delta T}(U, S_i) + w_c; \quad (4)$$

$$p^{\Delta T}(U, S_i) = 1 - 1/(1 + e^{\chi^{\Delta T}(U, S_i)}); \quad (5)$$

add  $p^{\Delta T}(U, S_i)$  into  $p^{\Delta T}(U, \vec{S})$ ;

sort( $\vec{S}, p^{\Delta T}(U, \vec{S})$ );

return  $\vec{S}^k$ : the  $k$  recommended trustworthy friends;

In this algorithm, the weighting values  $w_f$  (reply frequency),  $w_\ell$  (comment length),  $w_t$  (time difference),  $w_d$  (domain similarity),  $w_c$  (constant) are obtained from the logistic regression results at time  $t$ . This process of logistic regression is similar to the one illustrated in the previous section except that now it is based on factors excluding comment score since hypothesis H3 has been rejected.  $\Delta T$  denotes the time period during which  $U$  received comments from all the users who replied to him,  $\vec{S}$ ; thus, it should be the period from the time  $U$  received the first reply from  $\vec{S}$  to the timestamp  $t$ . Eq. 5 first computes the probability  $p^{\Delta T}(U, S_i)$  that user  $U$  would like to choose every user  $S_i \in \vec{S}$  as his friend, which can be derived from Eq. 3. Here, the comment score is excluded and the error term  $\varepsilon$  is ignored.

We name  $p^{\Delta T}(U, S_i)$  as **Interaction Index**; this indicator shows  $U$ 's degree of interactions with  $S_i$  from various

perspectives and therefore suggests the probability that  $U$  would like to establish a friendship with  $S_i$ . Then, based on the results from Eq. 5, the users in  $\bar{S}$  are sorted in descending order. A number of the most trustworthy users with the highest interaction index values (ranked top) will be recommended to the user  $U$  as friends.

The proposed IntRank demonstrates several advantages over traditional friend recommendation mechanisms. First, it is clear that IntRank provides a more comprehensive view of evaluating the trust of community members and the probability that other members would like to make friends with them by integrating different carefully verified interaction attributes. With a focus on direct interactions, IntRank is able to recommend trustworthy friends to community members. Second, by evaluating the trust and the probability of establishing friendships, the interaction ranking-based recommendation from IntRank does not cause the “rich getting richer” problem and is able to maintain the social network as a “manifestation of trust” [2]. This will increase the value of the social network and benefit virtual communities. Third, without considering subjective manual ratings as what the traditional recommender systems usually do, IntRank is much more objective and therefore more reliable.

## VI. PERFORMANCE EVALUATION

For the evaluation of IntRank, we experiment on the same preprocessed *Slashdot* users’ interaction and friendship data described in Section 4. We compare the output of the proposed algorithm with users’ friend lists.

The trustworthy friends recommended by IntRank rely on rich interaction information. To reduce the possibility that a user “blindly” adds another user as a friend without having many interactions with him, we selected the top 150 users who received the most replies. We use IntRank to recommend the most trustworthy friends to these 150 users from all the 3,307 users who replied to them. It was observed that some users in the 150 users’ friend lists never replied to them according to our collected data set. This is possibly because their replies were not sampled in Section 4. Thus, we continue to trim these 150 users’ friend lists by removing those who never replied to them according to our collected data set. After trimming their friend lists, only 2 users have more than 10 friends in their friend lists. We then compare the recommended results with the 150 users’ friend lists to evaluate the accuracy of IntRank recommendation results.

To begin with, the weighting values in Eq. 4 ( $w_f, w_\ell, w_t, w_d, w_c$ ) are learned from the logistic regression results of the preprocessed *Slashdot* interaction and friendship data. For each user  $U$ , after computing the interaction index of all the other users who replied to him in the past, IntRank sorts these users and the most trustworthy friends to be recommended are ranked top. We evaluate the accuracy of the recommendation results when the number of friends

recommended by IntRank is set to be between 1 and 10. For each user  $U$ , the accuracy of IntRank recommendation results is calculated as follows:

$$Accuracy = \begin{cases} N_F/N_R & N_R < N_L; \\ N_F/N_L & N_R \geq N_L. \end{cases} \quad (6)$$

where  $N_R$  is the number of friends recommended by IntRank,  $N_L$  is the size of  $U$ ’s friend list, and  $N_F$  is the number of recommended friends that are in  $U$ ’s friend list.

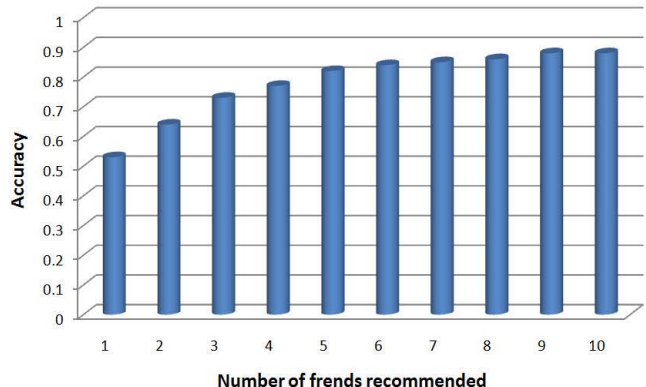


Figure 2. The accuracy of IntRank recommendation results.

Fig. 2 shows the average accuracy of IntRank recommendation results for the selected 150 users when the number of friends recommended by IntRank is set to be of different sizes. Obviously, the more friends recommended by IntRank, the higher accuracy achieved. To illustrate, assuming a user  $U$  has only added  $S$  in his friend list, when IntRank recommends only 1 friend to  $U$ , the chance to rank  $S$  at the top is low. If IntRank ranks  $S$  at the second place, when its recommendation size is 1, the accuracy is 0. However, when IntRank’s friend recommendation size is set to be larger than 1, the accuracy becomes 1. Since most of the users have fewer than 10 friends in our data set, in general when IntRank recommends more friends, the chance to get most of the friends in the friend list becomes higher. It is also noted that when IntRank recommends more than 4 friends to a user, its recommendation accuracy rises to over 0.8 and becomes more stable with the increase in the number of friends recommended. We conclude that for the given data set, IntRank is able to recommend trustworthy friends to a user when he has rich interactions with others in the virtual community.

## VII. CONCLUSIONS AND FUTURE WORK

Users in virtual communities are known to benefit more and more from well-established social networks in many aspects, and one of the most important features is the effective automatic recommendation of friends. In this paper we studied the influence of interaction attributes on trust and friendship probability in virtual communities and proposed a novel interaction ranking-based model for recommending trustworthy friends to community members.

In relation to the state-of-the-art research in literature, we make three unique contributions. First and foremost, we identified and validated four new interaction attributes: reply frequency, comment length, time difference, and domain similarity that influence trust and friendship probability in virtual communities by performing logistic regression analysis on real and large interaction and friendship data from *Slashdot*. We have further verified that subjective manual ratings (comment score) are not accurate, and thus cannot be directly relied on to recommend trustworthy friends. Second, we demonstrated one's view of choosing friends in virtual communities from four perspectives: interaction willingness, interaction quality, seriousness in interactions, and common interest. With the consideration of these new perspectives, members' trust and the probability to establish friendships in virtual communities can be evaluated in a more comprehensive way. Third, we proposed an interaction ranking-based trustworthy friend recommendation model IntRank in the context of virtual communities. From the performance evaluation results, we conclude that IntRank is able to recommend trustworthy friends effectively. This novel model has several advantages over many traditional friend recommendation mechanisms. Specifically, IntRank provides a more comprehensive approach to recommend trustworthy friends with whom the user has intensively interacted before, maintains the social network as a "manifestation of trust" and achieves more objective and reliable results.

Since IntRank relies on direct interactions, for future work, we will continue to evaluate the influence of indirect interactions on trust and friendship probability in virtual communities. We also plan to evaluate the proposed hypotheses and IntRank using data from other virtual communities.

#### REFERENCES

- [1] Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transaction on Knowledge and Data Engineering* **17**(6) 734–749 (2005)
- [2] Bernier, M., Ganley, D. Will networking nerds pay for stuff that matters? The relationship between social networks and subscriptions in virtual communities. *International Conference on Computational Science and Engineering*. (2009)
- [3] Brennan, M., Wrazien, S., Greenstadt, R.: Using machine learning to augment collaborative filtering of community discussions. *9th International Conference on Autonomous Agents and Multiagent Systems*. (2010)
- [4] Brzozowski, M., Hogg, T. and Szabo, G. Friends and foes: ideological social networking. *26th Conference on Human Factors in Computing Systems*. (2008)
- [5] Chen, C., Hung, S.: To give or to receive? Factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Information and Management* **47**(4) (2010)
- [6] Chen, J., Geyer, W., Dugan, C., Muller, M., Guy, I. "Make new friends, but keep the old" — recommending people on social networking sites. *27th International Conference on Human Factors in Computing Systems*. (2009)
- [7] Daly, E. M., Geyer, W., Millen, D. R.: The network effects of recommending social connections. *4th ACM Conference on Recommender Systems* (2010)
- [8] Ganley, D., Lampe, C. The ties that bind: social network principles in online communities. *Decision Support Systems* **47** (2009)
- [9] Garton, L., Haythornthwaite, C. and Wellman, B. Studying online social networks. *Journal of Computer-Mediated Communication* **3**(1) (1997)
- [10] Grandison, T., Sloman, M.: A survey of trust in internet applications. *IEEE Communications Surveys and Tutorials* **3** (2000)
- [11] Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. *Decision Support Systems* **43**(2) (2007) 618–644
- [12] Kaltenbrunner, A., Gómez, V., López, V.: Description and prediction of Slashdot activity. *5th Latin American Web Congress*. (2007)
- [13] Kaltenbrunner, A., Gómez, V., Moghnieh, A., Meza, R., Blat, J., López, V.: Homogeneous temporal activity patterns in a large online communication space. *IADIS International Journal on WWW/Internet* **6**(1) (2007) 61–76
- [14] Kunegis, K., Lommatzsch, A. and Bauchhage, C. The Slashdot zoo: mining a social network with negative edges. *18th International World Wide Web Conference*. (2009)
- [15] Lampe, C., Resnick, P.: Slash(dot) and burn: distributed moderation in a large online conversation space. *Conference on Human Factors in Computing Systems*. (2004)
- [16] Leskovec, J., Huttenlocher, D., Kleinberg, J.: Prediction positive and negative links in online social networks. *19th International World Wide Web Conference*. (2010)
- [17] Liben-Nowell, D., Kleinberg J. The link prediction problem for social networks. *12th International Conference on Information and Knowledge Management*. (2003)
- [18] Massa, P., Avesani, P.: Controversial users demand local trust metrics: an experimental study on epinions.com community. *20th National Conference on Artificial Intelligence*. (2005)
- [19] Mui, L., Mohtashemi, M., Halberstadt, A.: A computational model of trust and reputation. *35th Hawaii International Conference on System Sciences*. (2002)
- [20] O'Donovan, J., Smyth, B.: Trust in recommender systems. *10th International Conference on Intelligent User Interfaces*. (2005)
- [21] Popescul, A., Ungar, L. Statistical relational learning for link prediction. *Workshop on Learning Statistical Models from Relational Data at IJCAI-2003*. (2003)
- [22] Sinha, R., Swearingen, K. Comparing recommendations made by online systems and friends. *DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*. (2001)
- [23] Skopik, F., Truong, H.L., Dustdar, S.: Trust and reputation mining in professional virtual communities. *9th International Conference on Web Engineering*. (2009)
- [24] Terveen, L., Hill, W.: Beyond recommender systems: helping people help each other. *Human Computer Interaction in the New Millennium* (2001)
- [25] Zadeh, B.S., Ahmad, N., Abdullah, S., Abdullah, H. The social capacity to develop a community. *Current Research Journal of Social Science* **2**(2) (2010)