The ISC Int'l Journal of Information Security

July 2013, Volume 5, Number 2 (pp. 171–187)

http://www.isecure-journal.org

Prediction of User's Trustworthiness in Web-based Social Networks via Text Mining

Hossein Mohammadhassanzadeh [1,*], Hamid Reza Shahriari [1]

[1] *Amirkabir University of Technology, Department of Computer Engineering and Information Technology, Tehran, Iran*

ARTICLE INFO.

Article history:
Received: 31 December 2012
Revised: 19 August 2013
Accepted: 03 September 2013
Published Online: 20 March 2014

Keywords:
Trust, Reputation, Text Mining, User Similarity, Social Networks, Similarity Measure.

ABSTRACT

In Social networks, users need a proper estimation of trust in others to be able to initialize reliable relationships. Some trust evaluation mechanisms have been offered, which use direct ratings to calculate or propagate trust values. However, in some web-based social networks where users only have binary relationships, there is no direct rating available. Therefore, a new method is required to infer trust values in these networks. To bridge this gap, this paper aims to propose a new method which takes advantage of user similarity to predict trust values without any need for direct ratings. In this approach, which is based on socio-psychological studies, user similarity is calculated from the profile information and the texts shared by the users via text-mining techniques. Applying Ziegler ratios to our approach revealed that users are more than 50% more similar to their trusted agents than to arbitrary peers, which proves the validity of the original idea of the study about inferring trust from language similarity. In addition, comparing the real assigned ratings, gathered directly from users, with the experimental results indicated that the predicted trust values are sufficiently acceptable (with a precision of 61%). We have also studied the benefits of using context in inferring trust. In this regard, the analysis revealed that the precision of the predictions can be improved up to 72%. Besides the application of this approach in web-based social networks, the proposed technique can also be of much help in any direct rating mechanism to evaluate the correctness of trust values assigned by users, and increases the robustness of trust and reputation mechanisms against possible security threats.

1 Introduction

Nowadays, stars can be seen in a variety of websites, which are used to rank the entities in terms of their reputation. This reputation is calculated using the trust values or ratings assigned by others to an entity. Trust and reputation mechanisms have been basically introduced in virtual communities where some interactions take place among different users. These interactions usually include buying/selling goods, exchanging information, and service provision. In these virtual environments, in which something is being exchanged, there also exists the "Risk of Prior Performance" [1]. In other words, a consumer who accepts the terms of transactions may pay the fee without "squeezing the

* Corresponding author.
Email addresses: hassanzadeh@dal.ca (H. Mohammadhassanzadeh), shahriari@aut.ac.ir (H. R. Shahriari)

ISSN: 2008-2045

tomato" or directly checking the products, or assessing the quality of the appropriate service. Due to the lack of previous knowledge in online societies, participants are often misguided during the decision making process. In such a vulnerable situation, using trust and reputation mechanisms is inevitable. Moreover, the growing tendency to use social networks has brought about some changes in their application. Social networks were initially proposed to connect people and support their social relationships in virtual environments. Given the expanding applications of social networks, other interactions such as providing services, exchanging information, marketing, buying/selling, etc. have also emerged. As in the preceding virtual environments, trust and reputation play a major role in these recent applications of social networks as well. However, calculating trust and reputation in social networks is not totally free of problems. In this regard, Golbeck [2] has stated that one computational problem with trust is to determine how much the users in the network should trust the others to whom they are not connected. In the same vein, Golbeck [2] presents a method for inferring the trust value between two individuals who are not directly connected. This mechanism uses the paths which connect them in the social network, and infers the trust values along those paths.

Although this mechanism and other similar methods [3] are applicable to social networks, in all these approaches, some initial trust values are required, which are assigned directly by the users. However, in some social networks, it is not possible for the users to directly indicate the extent to which they can trust the other users in the network. Therefore, in these networks, trust values need to be inferred implicitly from some other resources.

In this respect, some of the preceding studies [4, 5] have found a positive correlation between trust and interest similarity. In these works, user similarity is regarded as the basis for generalizing the available trust values to unknown cases.

In addition, it has been previously shown [6] that there is also a strong correlation between user similarity and the words used by users in their speaking and writing. Taking these assumptions into account, in the present study, user similarity is extracted from the texts shared by the users in a social network. This similarity is calculated via some text mining techniques. We also consider profile similarity to compare trust values resulting from different facets of similarity.

Furthermore, the role of context in the calculation of trust and inferring context-aware trust values has also been investigated here, yielding the conclusion that considering context in these processes can bring about considerable improvement in the accuracy of the results.

The rest of this paper is organized as follows: Section 2 goes through the concepts and definitions of the terms used in this paper, followed by a review of the related works. In Section 3, the technique used for calculating the trust is explained. Section 4 describes the proposed method and explains the experimental results and the validation process. In Section 5, the advantages, applications as well as the flaws of the proposed method are being dealt with. And finally the paper is concluded in Section 6, where some suggestions are also offered for future works.

2 Background and Related Works

In this section, two major studies related to our own are being reviewed to shed more light on the correlation between user similarity and users' trust values. However, some of the key terms and concepts used in these studies are required to be explained beforehand.

2.1 Trust and Reputation

Considering the different fields in which trust and reputation are used, various definitions have been proposed for these concepts so far; however, there is no universal agreement on the definitions. Wang *et al.* for instance, [7] adopt the following working definitions for trust and reputation, which distinguish between these two concepts based on the information resources:
Trust - a peer's belief in another peer's capabilities, honesty and reliability, based on its own direct experiences;
Reputation - a peer's belief in another peer's capabilities, honesty and reliability based on recommendations received from other peers.

In a similar vein, Artz *et al.* [8] have extracted three general definitions from the existing research in order to pinpoint some different aspects of trust. The first definition, given by Mui *et al.* [9], is based on the past encounters, and may be thought of by some as the "reputation-based" trust. According to Mui *et al.* "fTrust is a subjective expectation an agent as about another's future behavior based on the history of their encounters."

Propounding the idea of context in trust calculation for the first time, the second definition considers the "competence" of a party instead of his/her abilities. Having been proposed by Grandison and Sloman [10], this perspective defines trust as "the firm belief in the competence of an entity to act dependably, securely, and reliably within a specified context." Yet, Olmedilla *et al.* [11] look at the concept from another aspect and define it based on actions. From their viewpoint,

"Trust of party A to party B for a service X is the measureable belief of A in that B behaves dependably for a specified period within a specified context (in relation to service X)."

All these definitions are regarded as the primary descriptions of trust and reputation, however, they focus on the applications of trust specifically in e-commerce systems. Although these definitions are applicable to some aspects of relationship in social networks, there is still a need to a definition which is more consistent with specific characteristics of social networks and covers other aspects of the network as well.

In older virtual communities, users built relationships only based on one or two transactions like selling/buying, marketing, advising, etc. Nevertheless, nowadays in social networks, there are various kinds of transactions existing simultaneously (e.g. friendship, family and working relationships in addition to marketing, etc.), which make it more complicated for the users to make reliable relationships.

In this regard, Golbeck [4] calls this kind of trust as "social trust" and believes that social trust depends on a host of factors which cannot be easily modeled in a computational system; Therefore, she proposes a new definition of trust, as "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 companionship of "social trust" with the earlier four definitions presents a comprehensive definition of trust in social networks, which completely covers all aspects of todays multi-dimensional social networks.

2.2 Social Networks

Social network is a society of people, which can be mapped to the structure of a graph. In this graph, the nodes and edges represent the users and their relationships respectively. The relationships between users could be friendship, common interest, financial exchange, etc.

Although this structure is inevitable in any social networks, Golbeck [2] differentiates between a social network and a web-based social network. In Golbeck's opinion, a social network could be derived from many possible ways, such as users connected through transactions in online auctions, and users who post within the same thread on a news group. Hence, Golbeck states four criteria to differentiate between a social network and a web-based social network: 1.the web-based social network is accessible over the web with a web browser, without any need to add-ons; 2.users must explicitly state their relationship with others; 3. the system must support users to make their explicit connections; 4. preserving the users privacy, relationships must be visible and browsable.

Given the present study, it is strongly believed that any social network which satisfies these criteria, can thoroughly provide the fundamental structure as well as the contextual information for the proposed approach. Thus, these criteria are adopted for the purpose of this paper. It has to be noted that web-based social networks are being referred to as social network in brief hereafter.

2.3 Trust Mechanisms in Social Networks

In addition to commercial transactions on the Internet, there are also some other kinds of communications where trust mechanisms are applicable and in some cases they are even vital. Social Network as an emerging phenomenon is one of these areas. It contains a wide variety of relationships, namely friendship, familial, professional, academic, and even commercial, each of which requires a different level of trustworthiness for a person. Thus, it is necessary for users to be able to assess their friends, assign trust values to them and discriminate between different friends.

In response to this need, some trust mechanisms have been proposed specifically for social networks. These mechanisms infer trust values in social networks using rating systems and the graph structure of social networks.

TidalTrust [2] is one of these mechanisms which takes advantage of trust ratings and path lengths in social networks to infer the degree of trust. Golbeck investigated the validity of her proposed mechanism on a movie social network named FilmTrust.

$$t_{is} = \frac{\sum_{j \in adj(i) \ni t_{ij} \geq max} t_{ij} t_{js}}{\sum_{j \in adj(i) \ni t_{ij} \geq max} t_{ij}} \qquad (1)$$

In TidalTrust, $t_{i,j}$ represents the trust rating from node i to node j. The inferred trust rating from node i to node s is obtained by (1). In this formula, the variable max represents the possible largest trust value in the path from i to s. This value can be used as a minimum threshold for every node such that a path can be found from source to sink.

The formula respects the fact that more trusted neighbors are generally more reliable to find the path. Figure 1 illustrates the process of determining the trust threshold. The label on each edge represents the trust rating between nodes. The label on each node indicates the maximum trust value on the path leading to that node s. The two nodes adjacent to the sink are of values of 9, thus 9 is regarded as the max

value. The bold edges indicate the paths which will ultimately be used in the calculation as they are at or above the max threshold.

MoleTrust is another trust mechanism which is applied on an online social network named MoleSkiing. In this network, users comment on the weather and the status of skiing paths and others assign trust values to both the comments and the user. Basically, MoleSkiing is a recommender system which works on the basis of trust values.

MoleTrust predicts trust scores of unknown users from the point of view of user m. This mechanism first arranges the users based on the shortest-path distance from user m. A parameter of MoleTrust is the Trust Propagation Horizon: trust is not propagated at distances greater than this horizon. The intuition is that the reliability of the propagated trust decreases with every new trust hop.

Moreover, this way, the number of nodes to be considered by the trust metric is reduced, which in turn results into a shorter computational time. At this point, MoleTrust removes all the trust edges between a user at a certain distance, and those with a lower or equal distance. For example, every edge from users at distance 3, to users at distance 1, 2 or 3 is removed from the social network. The first step ends here. The social network is now a directed acyclic graph where trust flows from m to other users and never flows back, i.e. there are no cycles [3].

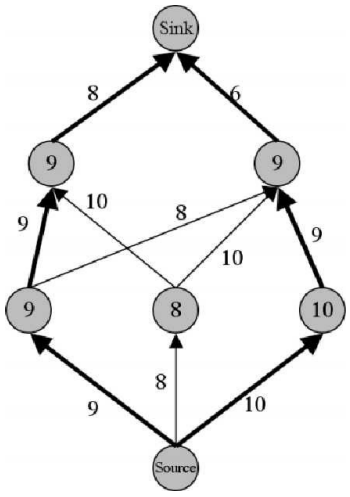


Figure 1. The process of determining the trust threshold (variable *max*)

The second step is a simple graph walking over the modified social network, starting from user m, whose trust score is the maximum by definition. MoleTrust first computes the trust score for all the users at distance 1, followed by the same value for all the users at distance 2, 3, etc. The trust score of one user at distance x merely depends on the trust scores of users at distance *x-1*, which are already computed and definitive.

In order to predict trust score for a user, MoleTrust analyzes incoming trust edges. However, only trust edges coming from users with a predicted trust score greater than 0.6 are considered. The other users are regarded as not trustworthy and their trust statements should simply be ignored [3].

Based on (2), the predicted trust score for user B is the average of all the incoming trust edge values $(t_{i,b})$, weighted by the trust score t_i of the user who has issued this trust statement.

$$t_B = \frac{\sum_{i \in \{i|(i,B) \in E\}} t_i t_{(i,b)}}{\sum_{i \in \{i|(i,B) \in E\}} t_i} \tag{2}$$

2.4 Context-aware Trust

Context is a source of information not being used frequently in trust mechanisms, and includes any information that can be used to characterize the situation of an entity. Dependency of trust mechanisms on users' information is a strong motivation to utilize context in these mechanisms, to introduce context-aware trust mechanisms and to investigate the performance of context in trust calculations.

A brief review of the literature [12] reveals that in extending a trust model, taking into the role of context consideration can

- Reduce complexity in management of trust relationships [13]
- Improve the recommendation process [14]
- Help to infer trust information in context hierarchies [15]
- Improve performance [16]
- Help to learn policies/norms at runtime [16] [17]
- Provide protection against changes of identity and first time offenders [16] [18]

In this regard, Neisse *et al.* [13] proposed the idea of using the abstraction of context-aware domains to reduce the complexity in the management of trust relationships. According to Neisse *et al.* it is also possible to use context information to improve the recommendation process [13].

It has also been suggested [14] that context can often be structured hierarchically. Gujral *et al.* [19] view context as a multi-dimensional trust modeling for agents when goal requirements are multi-dimensional. Similarly, Rehak [15] defines a set of reference contexts in a metric space and associates it with the truthfulness of the data.

The abovementioned works along with some other

studies carried out in this regard all suggest that, despite the fact that the majority of trust models only consider the history of relationships and recommendations in order to estimate the value of trust in the next interaction, context-aware trust mechanisms make use of some other factors to improve the estimations.

2.5 Related Works

Although many studies have been conducted by researchers on trust and reputation quantification, only a few studies have focused on applying user similarity to calculate the trust values.

In one of such studies, Ziegler *et al.* [5] have presented two frameworks for analyzing the correlation between interpersonal trust and interest similarity. In fact, they investigated the interaction between trust and similarity on the basis of evidence from socio-psychological research, stating that there is a positive interaction between friendship and attitude similarity. Burgess *et al.* [20], Newcomb [21], and Byrne [22] [**?**] have also provided evidence for the existence of this positive interaction in their works.

Having evaluated the frameworks in the case of a book and a movie social network, Ziegler *et al.* claimed that the results of their work indicate a strong positive interaction between interpersonal trust and interest similarity. Ziegler *et al.* introduced two ratios for evaluating their framework. First, they computed the average similarity score of trusted peers for user at and call it (z_i). They also calculated the average of similarity scores of all peers for user (a_i) , denoted by (s_i) Table 1.

A comparison of pairs (z_i, a_i) revealed that in 173 cases (66%), users were more similar to their trusted peers than to the arbitrary ones, while the opposite situation held true for only 88 agents (34%). Users had an average similarity score of 0.247 with respect to their trusted peers, while the same value for the similarity between the users and the whole society was found to be 0.163. In other words, users were more than 50% more similar to their trusted agents than arbitrary peers [5].

Table 1. Ziegler Ratios for evaluating his framework

Average similarity scores of trusted peers	$z_i = \frac{\sum_{a_j \in trust(a_i)} Sim(a_i, a_j)}{\lvert trust(a_i)\rvert}$
Average similarity scores of all peers	$s_i = \frac{\sum_{a_j \in A/\{a_j\}} Sim(a_i, a_j)}{\lvert A\rvert - 1}$

$trust(a_i)$: The collection of trusted peers of user a_i
A: The collection of all peers in society

Ziegler *et al.* [5] also mentioned that their findings are line with the research findings on in recommender systems and collaborative filtering, where people are suggested products based on their similarity with other customers.

Yet there has been another study to investigate different features of profile similarity and explored how these features influence assigning trust values. Golbeck [24] has computed trust by gathering information from friends and using the existing trust values in a social network. In this work, an unknown user's rating of a specific product is predicted with the help of the trusted friends of the user and their ratings of the same item.

In addition to the overall similarity, Golbeck has shown that a strong correlation exists between trust and some other facets of profile similarity such as the largest single difference and the agreement on movies which the source has given the maximum rating. Golbeck believes that these findings can be used in several areas such as trust-based recommender systems, refining trust inference algorithms, and trust estimation for intelligence systems.

In the same vein, Golbeck *et al.* [24] assume social media as a place where users present themselves to the world, revealing personal details about their lives. Focusing on personality, Golbeck et. al attempted to understand how some of this information can be utilized to improve the users' experiences with interfaces and with one another.

In addition to these basic studies on evaluating the relationship between similarity and trustworthiness, a more recent study [25] has been conducted to investigate the relationship between personality and implementation of social media. In this study, Quercia *et al.* [25] have set out to analyze the relationship between personality and different types of Twitter users, including popular users and influentials. Having used the information related to 335 users, the researchers found out that both popular users and influentials are extroverts and emotionally stable. They also realized that popular users are 'imaginative', while influentials tend to be 'organized'.

Quercia *et al.* also used three counts publicly available on users' profile (including following, followers, and listed counts) to present a way for accurately predicting a user's personality. They showed that by using these three quantities about an active user, one can predict the user's five personality traits with an acceptable accuracy [25].

Quercia *et al.* gathered data from a Facebook Application called myPersonality to associate personality scores with Twitter users. Using the dataset, they could predict the user's five personality traits on a

[1,5] scale.

In another study, Tang *et al.* [26] benefited from social theories to broaden their understanding of trust. To this end, they used the idea of homophily to explain why trust relations are established. The homophily effect suggests that similar users have a higher likelihood to establish trust relations.

Elsewhere in Tang's study [26], the researcher also investigated homophily in trust relations via studying the correlation between trust relations and users' similarity. They specifically sought to answer two main questions,, within the context of product review sites:

- Are users with trust relations more similar in terms of their ratings than those without?
- Are users with higher similarity more likely to establish trust relations than those with lower similarity?

In response to the first question, Tang employed the cosine similarity of users' rating vectors to measure their rating similarity. The evidence from calculations suggested that it is highly probable for the users with trust relations to have higher rating similarities than those without [26].

In the same study, the second research question was aimed to find if users with higher similarity at time t are more likely to establish trust relations at time *t+1*, than the ones with lower similarity. This issue was studied in the context of Epinions since it provides temporal information when ratings are created and when trust relations are established. The findings of the study indicated that users with higher rating similarity are more likely to establish trust relations than those with lower similarity [26].

Positive answers to both the above mentioned questions provided evidence for the existence of homophily in trust relations. Taking this into account, the present study has exploited the homophily effect for trust prediction, as being discussed below.

2.6 Our Proposed Method to Infer Trust

Considering what was suggested by the aforementioned works, it can be realized that there is a new trend to predict trust value when it is unknown. As mentioned above, in these two studies, there were some rating mechanisms in which users assign trust values to friends and express which items they appreciate. Yet sometimes, there is no possibility to implement a direct rating mechanism in social networks.

In popular social networks such as Facebook, users only have a binary relationship, including being friend or not, and no direct and explicit rating is available. Therefore, a method is required to infer the values of trust and user reputation in social networks without any need to trust values assigned by users. In this respect, we have earlier explored the interaction of trust and user similarity and showed that there exists a promising correlation between user similarity and user's trustworthiness (with an accuracy of 20% in exact values and near 40% in values with a difference of one unit) [27]. This finding thus has encouraged us to develop our proposed approach and expand the method to achieve better results in inferring trust values.

In the present paper, it will be explained that in the absence of direct rating, there are some other information resources which can be used to extract user similarity and infer trust values. These resources include user's personal information, user interest, and the posts and the comments shared by the user.

3 Measuring User Similarity

In this section, the procedure of measuring user similarity is being described, and then the continuous similarity values are generalized to form the basis for predicting the discrete trust values.

The method used to calculate similarity is schematically represented in Figure 2. As can be viewed in the figure, in this approach each user is represented by a text file. A higher similarity between two documents signifies more similarity between the corresponding users. As shown in Figure 2, there are six steps to be followed to extract user similarity, as being described below:

***1. Collecting Information*:** As mentioned before, the data is collected from three different sources, namely the users' profile containing personal information and users' favorites, the posts shared by users, and the notes annotated as comments. The information collected form each of these sources form a text file which would be mined to extract similarity. In computations, $D = \{d_1,..., d_n\}$ is a set of documents, in which d_i is the document related to user i, and $T = \{t_1, ..., t_m\}$ is the set of distinct terms occurring in D. Therefore, in the calculations, a document is presented with an m-dimensional vector $\boldsymbol{t_d}$.

***2. Removing Stop Words*:** although stop words like "a" and "the" are used frequently, they are neither descriptive nor important for the document's subject [28]. These words are thus omitted from the set of T, as they have no additive information to differentiate between users.

***3. Porter Stemming*:** In mining documents, there are lots of words which seem different but they are derived form the same unique stem. Therefore, these

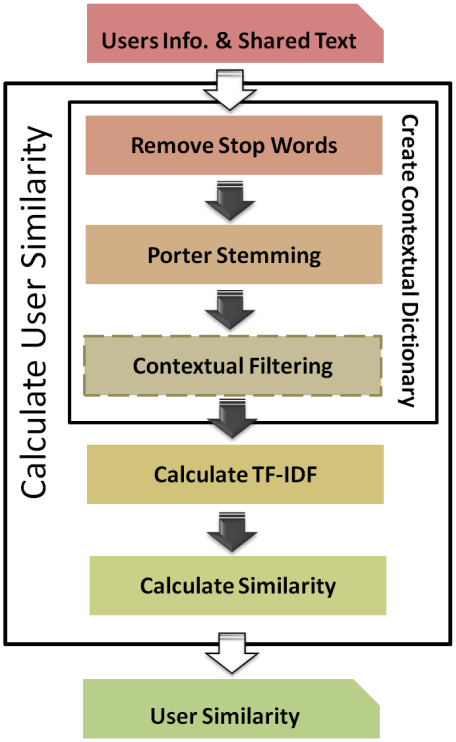


Figure 2. Steps of measuring user similarity [27]

words are treated as a single word as they are thematically similar [28]. Porter stemming [29] is an algorithm to distinguish these words by suffix-stripping. For example, in this algorithm *production*, *produce*, *produces* will be mapped to the stem *produc*. In this experiment, a bottom up approach [30] has been used to stem Persian words.

The algorithm used in the present study is rule based and bottom up, meaning that at first, it tries to find substrings of the word that are stems or morphemes deriving from a main stem, called *core*. Subsequently, each of the cores are joined to other elements of the word for generating that word according to the available rules. Finally, each core with at least one correct generation is considered as a correct core and its stem is the correct stem of the word. The algorithm thus includes three phases: 1. Substring tagging, 2. Rule matching, and 3. Anti rule matching [30].

In the substring tagging phase, morphological information for all possible substrings of the word are to be extracted. At the end of this phase, it is known that which substrings of the word are morphemes and which ones are not. Moreover, the clusters of which each morpheme is a member, are also distinguished. These clusters are then used in the rule matching phase. Accordingly, prior to the second phase, the cores in the word are known as well.

***4. Contextual Filtering*:** this step is optional and will only be considered if context-aware trust values

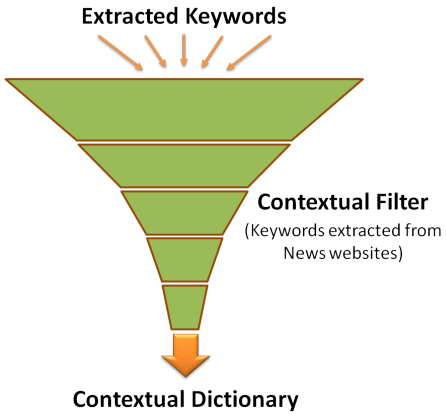


Figure 3. Contextual Filter and it's mechanism

are intended. However, it has been shown that considering filtration and inferring context-aware trust values considerably increases the accuracy.

In order to infer context-aware trust values, it is necessary to differentiate between words and the context in which they are used and consider only the words relevant to the desired context. To this end, a number of keywords in four different categories (Science, Arts, Sport and Politics) were gathered from different News websites. These keywords formed the contextual filters of our study.

When the keywords extracted from the text shared by users are passed through the contextual filters (Figure 3), the outcome is a contextual dictionary which contains only the words which exist in the very relevant context. For example when we want to examine the users' trust in each other in the field of Sport, the extracted keywords are passed through "Sport Filter" and only the words which exist in the filter and are related to sport will pass the filter. These words constitute the contextual dictionary and consequently will be considered in the calculations.

In fact, a contextual filter eliminates words irrelevant to the target context from the calculation and results in context-aware trust values with regard to the context.

***5. Calculating TF-IDF*:** $tf(d,t)$ stands for the frequency of the term $t \in T$ in document $d \in D$. Although more frequent words are assumed to be more important, this is not usually the case in practice [28]. Instead of frequent words, it is more informative to consider terms that appear frequently in a small number of documents but rarely in other documents. These terms tend to be more relevant and specific for that particular group of documents and therefore more useful in finding similar documents. $tfidf(d,t)$ is calculated through:

$$tfidf(d,t) = tf(d,t) \times \log(\frac{|D|}{df(t)}) \quad (3)$$

where *df(t)* is the number of documents in which term t appears. Now it can be said that the vector consists of:

$$\boldsymbol{t_d} = (tfidf(d,t_1), ..., tfidf(d,t_m)) \quad (4)$$

***6. Calculation of Similarity Measures*:** A similarity/distance measure reflects the degree of closeness or separation of the target objects. In general, similarity/distance measures map the distance or similarity between the symbolic descriptions of two objects into a single numeric value, which depends on two factors: the properties of the two objects and the measure itself [28].

In case of using *TF-IDF* in the calculation of document similarity, there are several measures which can be used. These measures are presented in Table 2 [28] [31]. $vect_1$ and $vect_b$ are the vectors containing the most important terms of documents a and b. The values resulting from these 4 measures fall into different ranges.

In cosine similarity, for instance, the obtained values range between -1 and +1. Nonetheless, in the case of information extraction, due to the positive values of *tfidf* vector, cosine is also positive, and thus the similarity is limited to the values within the range of [0,1]. Given the Pearson Correlation, the values again range from -1 to +1. In the abovementioned cases, a value of -1 signifies a perfect negative linear relationship between variables, 0 indicates no linear relationship, and +1 implies a perfect positive relationship between users. In this study, negative relationships have been discarded and considered as 0. In the other two measures, i.e. Euclidean Distance and Jacquard Coefficient, the resulting values range from 0 to +1, as is also required in our work.

All these measures are applicable in text mining and information retrieval. However, based on the available data, the output may vary from a measure to another. Hence, it was worth studying the efficiency of each measure on our data in order to find the best measure fitting our purpose. The accuracy of these measures and the comparisons in this regard are presented in the experimental results section.

***7. User Similarity*:** Calculations of user similarity yielded a matrix with 1 as its main diagonal and similarity values as its other entries. The entry in the i-th row and j-th column indicates the similarity between nodes i and j. It is worth mentioning that the output of this phase constitutes of five similarity values for each couple of users.

Therefore, one similarity value as "Overall Similarity" (regardless of any context) and four other values regarding to each four different context. We then discretize these similarity values to the range of [0,5] as trust values between users in each context.

4 Experimental Results

It has been suggested by many studies in the field of psychology that people who have more interests in common, are more similar [20] [21] [22] [23], and have more confidence in each other [4] [5]. Moreover, it seems that people who use the same words in writing and have similar utterances, would be more similar in terms of their personality. Thus, the correlation between similarity and trust, can provide an estimate of trust values.

To evaluate our proposed approach, we gathered all the available information about a user and then tried to figure out which resources yield better results.To this end, as can be found in every online social network, all these information were divided into three categories of profile information, user post and user comments (Figure 4).

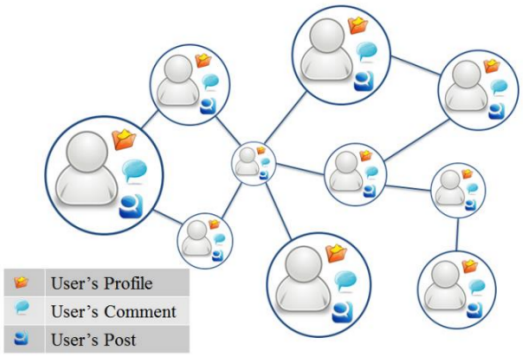


Figure 4. A sample of social network. The size of each user is proportional with the volume of information he/she has exposed earlier in the networks.

As mentioned before, the texts shared by the users are used to calculate their similarity. These texts, either written by an individual user him/herself, or only shared by them, implicitly represents some aspects of their personality. In the following subsection, the data which are used in validation process are being described in details.

4.1 Experimental Data

In online social networks, users are encouraged to present personal information. Yet, it is probable that users do not share this amount of information with all their friends in real world. The information that express personal feelings and interests are referred to as **profile** information.

Table 2. Similarity/Distance measures studied in this approach

Measure	Formula
Euclidean Distance	$Dis_E(\boldsymbol{t_a}, \boldsymbol{t_b}) = (\sum_{i=1}^{m} \lvert w_{t,a} - w_{t,b} \rvert^2)^{\frac{1}{2}}$
Cosine Similarity	$Sim_C(\boldsymbol{t_a}, \boldsymbol{t_b}) = Cos(\theta) = \frac{\boldsymbol{t_a}.\boldsymbol{t_b}}{\lvert \boldsymbol{t_a} \rvert \times \lvert \boldsymbol{t_b} \rvert}$
Jacquard Coefficient	$Sim_J(\boldsymbol{t_a}, \boldsymbol{t_b}) = \frac{\boldsymbol{t_a}.\boldsymbol{t_b}}{\lvert \boldsymbol{t_a} \rvert^2 + \lvert \boldsymbol{t_b} \rvert^2 - \boldsymbol{t_a}.\boldsymbol{t_b}}$
Pearson Correlation Coefficient	$Sim_P(\boldsymbol{t_a}, \boldsymbol{t_b}) = \frac{m \sum_{i=1}^{m} w_{t,a} - w_{t,b} - TF_a \times TF_b}{\sqrt{(m \sum_{i=1}^{m} (w_{t,a})^2 - (TF_a)^2) \times (m \sum_{i=1}^{m} (w_{t,b})^2 - (TF_b)^2))}}$

$T = \{t_1, t_2, .., t_n\}, w_{t,a} = tfidf(d_a, t), \lvert \boldsymbol{t_a} \rvert = \sqrt{(p_1)^2 - ... - (p_2)^2}$

Table 3. Statistical information about experimental users

a. Population	
count	153 users
b. Gender	
Male	101 (66%)
Female	52 (34%)
c. Age	
≤20	0 (0%)
> 20 and ≤ 25	45(29%)
> 25 and ≤ 30	83(54%)
> 30 and ≤ 40	21(14%)
> 40 and ≤ 65	4(3%)
d. Education	
Diploma	10 (7%)
Bachelor	57 (37%)
M.Sc.	77 (50%)
Ph.D.	9 (6%)

On the other hand, users **post** their thoughts, expressions, and tweets directly on the network and share it publicly or customize its visibility just to a list of friends. Moreover, users in online societies, just as in real world, join conversations, state their opinions and explain their views. These conversations are labeled as **comments**. These three types of resources - profile, post, comment- comprise the datasets required to evaluate the proposed approach. This information is gathered through our experimental society of Facebook users who are the first author's friends. This society contains the author's classmates, his colleagues, and childhood friends, and also his family members, which properly represents a real society. The information about the experimental users is summarized in Table 3.

The first dataset, i.e. the profile information, includes 14 features. Birth date and sex were used as the primary data in the calculations. However, it was found out that these features reduce the accuracy of the method. Therefore, they were discarded in final dataset. This dataset is used to compare the efficiency of trust values resulting from interest similarity with those obtained form the shared text similarity. It should be noticed that since the method used to calculate similarity was the same for all datasets, the comparisons between the results are creditable.

Some details of each dataset are presented in Table 4. In this table, the volume of each dataset after stemming and creating dictionaries is provided. As can be seen, the profile dataset (section a) is the smallest one, as it is only made up of some keywords in different categories, but the two other datasets (sections b and c) are composed of usual sentences, which obviously contain various words, resulting in a larger number of unique words.

Table 4 also contains a sample of the words in the dictionaries to present the content and structure of these dictionaries. All the texts are gathered during a period of three months from September 2011 to November 2011 via Facebook Developer.

To evaluate the accuracy of the inferred trust values, we asked the users to indicate the trustworthiness of their friends on 5-point scale (from completely untrustworthy to completely trustworthy) via an instrument developed by first author.

In addition to the overall (context-less) trust values, in order to assess the context-aware inferred trust values, the users were also asked to express the trustworthiness of their friends in each of the four contexts.

We also mapped the continues trust values resulted from user similarity to the discrete range of (0 to 5) and comparisons were made between these values and the ratings assigned by the users to study the precision of each aspect of similarity -similarity values resulting from three different datasets- in the prediction of trust values.

Furthermore, the accuracy of the results has been measured using the percentage of trust values resulting

Table 4. Samples of Datasets including the volume of each dataset

a. Profile Dataset
(Total Keywords: 6056, Unique Keywords: 777)

Word	Profile Parameter	Frequency of Occurrence
Tehran	Home town	6.2 E-03
Amirkabir University of Technology	College	1.81 E-03
Celine Dion	Music	1.65 E-03
Swimming	Sport	1.32 E-03
Sleeping	Activities/Interests	9.90 E-04
Animal Farm	Book	8.25 E-04
Reza Yazdani	Music	6.60 E-04
The Green Mile	Movie	6.60 E-04

b. Post Dataset
(Total Keywords: 52351, Unique Keywords: 16930)

Word before Stemming (Persian)	English Meaning	Frequency of Occurrence
ایران	Iran	2.08 E-03
زندگی	Life	1.94 E-03
سال	Year	1.87 E-03
کار	Work	1.50 E-03
خدا	God	1.47 E-03
بازی	Game	1.20 E-03
فیسبوک	Facebook	1.06 E-03
عشق	Love	8.79E-04

c. Comment Dataset
(Total Keywords: 66248, Unique Keywords: 18933)

Word before Stemming (Persian)	English Meaning	Frequency of Occurrence
اطلاع	Information	6.03 E-05
دانلود	Download	6.03 E-05
فارغ	Graduate	6.03 E-05
خوشبختی	Happiness	4.52 E-05
سختی	Difficulty	4.52 E-05
احترام	Respect	4.52 E-05
وطن	fatherland	3.01 E-05
عجله	Hurry	3.01 E-05

from user similarity which has the least variance - trust values with no difference (exact x indicated by user) or differ just in one unit (x-1 or x+1)- from the expected value assigned by the users. The less obtained value varies from the expected one, the more precise it believed to be.

4.2 Ziegler Ratios and the validity of our approach

Prior to evaluating the accuracy of proposed approach in inferring trust values, it is necessary to make sure about the basic acclaim. It is stated earlier that the people who talk similarly may have similar personalities and consequently have more trust in each other. Therefore, the validity of the basic idea of the approach needs to be well-proved in order to accurately evaluate the inferred values.

To this end, Ziegler ratios were employed to ensure that users are more similar to their trusted peers rather than to the arbitrary ones. Table 5 and Table 6 compare the values resulting from both Ziegler framework and the approach proposed in this study.

A comparison of pairs (z_i, s_i) reveals that in 91 cases (60%), users were more similar to their trusted peers than to the whole society, whereas the opposite observation can be made for only 88 users (40%) (Table 6).

Table 5. Ziegler Ratios, validity of proposed approach

	Average of similarity scores of trusted peers(z)	Average of similarity scores of all peers(s)	$\frac{z}{s}$
Our proposed approach	0.130	0.084	1.548
Ziegler framework	0.247	0.163	1.515

Table 6. Users with s>z or z>s and corresponding ratios

	Users with $z > s$ (Ut)	Users with $s > z$ (Uv)	$\frac{Ut}{Uv}$
Our proposed approach	91	62	1.47
Ziegler framework	173	88	1.97

As it is observed in the Tables, users have an average similarity score of 0.130 with their trusted peers, while a value of 0.084 was found for the users similarity with the whole society (Table 5). Not unlike the Ziegler framework, more than 50% of the users were more similar to their trusted agents than to the arbitrary peers.

In comparison with Ziegler results, although the value of Ut/Uv is relatively smaller in our approach than the Ziegler framework, it can be viewed that z/s is slightly larger than the value gained from the Ziegler framework. Based on Zieglers acclaim and the similar results achieved by our approach, one can certainly make the claim that people, who talk similarly may have similar personalities and consequently have more confidence in each other.

4.3 Results

A comparison of the inferred values with real values indicates that there are 6 possibilities of difference:

0, 1, 2 through 5. The minimum value of difference (zero) happens when we predict the values exactly the same as the real ones, and the maximum difference of 5 happens when we infer a trust value as 0 (or 5) while the user has assigned the value of 5 (or 0) to the corresponding value.

The experimental results of the present study have shown that the largest number of inferred trust values belong to the categories of equal values and values with 1 unit difference. Therefore, apart form the equal values, we decided to consider the inferred values with a difference of 1 unit in the total precision, as we believe it depicts a better picture of the performance of our approach.

Figure 5 clearly illustrates this conception. Notice that the phrase "differs a unit(s)" in figure 5 means if a user rated a friend with a discrete value equal to "x", the proposed approach predicted the corresponding trust value as "x-a" or "x+a" (unless values don't exceed the boundaries).

For example, if the user's rating value is 3, and we state that the predicted value differs 1 unit, this predicted value should be either 2 or 4. It should be considered only the values resulting from Euclidean measure are depicted in figure 5, however, other measures have more or less the same results.

The basic results are presented in figure 6. This figure, including three charts, displays the accuracy of proposed approach in the prediction of trust values in comparison with the real values assigned by the users.

In figure 6, the precision of the inferred trust values resulting from four different similarity measures are represented. These values are divided into three categories, as follows:

(1) Equal: Inferred Values exactly the same as Real Values
(2) 1 unit difference: Inferred Values differing 1 unit from Real Values
(3) Total: Sum of (1) and (2)

Chart (a) is related to the precision of the inferred trust values resulting from profile dataset. These values are obtained from 4 different measures introduced earlier. Chart (b) and (c) represent the results yielded from Post and Comment datasets respectively.

Regarding figure 6, it can be stated that trust values resulting from the comment similarity are more accurate than values resulting from the two other datasets. It is clear that in the comment dataset (chart c) all similarity measures perform better than in the two other datasets (chart a and b). Moreover, the maximum precision of 26% is resulted from Euclidean Similarity.

In the case of values with 1 unit of difference, Euclidean Similarity and Cosine Similarity give almost the same results. Given the total precision though, Euclidean Similarity is always prominent (with 56% precision in Post and Comment dataset and 50% for Profile Dataset).

As a result, based on figure 6, it can be claimed that among all the 4 similarity measures, Euclidean Similarity outperforms the other three measures and thus fits our approach in the best manner.

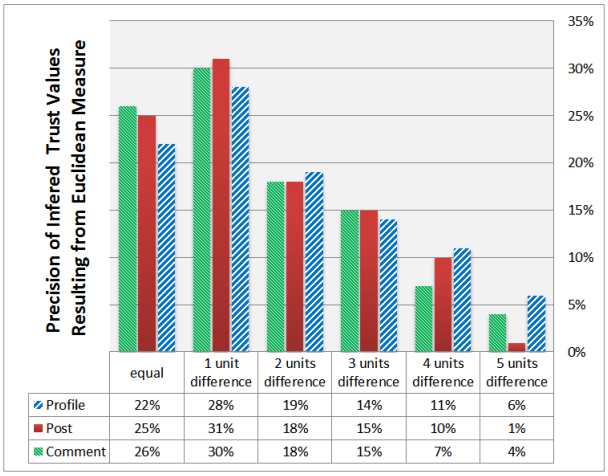


Figure 5. Precision from comparison of real rates with inferred ones. differs a unit(s) means if a user rated a friend x, inferred value would be x-a or x+a.

4.4 Coefficient Vector, the 1st step for improvement

We have also experimented the combination of values resulting from three datasets with different coefficients for each one. The resulting values are calculated using the following formula:

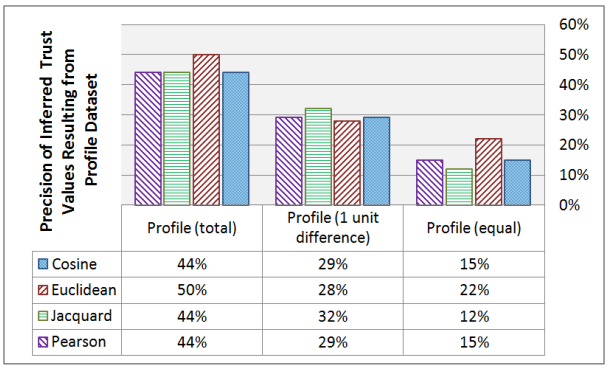
$$\begin{aligned} CombinationalValue = {} & \alpha * profilesimilarityvalue \\ & + \beta * postsimilarityvalue \\ & + \gamma * commentsimilarityvalue \end{aligned} \quad (5)$$

In equation 5, the vector (α, β, γ) is called **coefficient vector** and it is assumed that $\alpha + \beta + \gamma = 1$

Using the aforementioned formula, we attempted to examine four approaches for assigning values to coefficient vector variables, as described below:

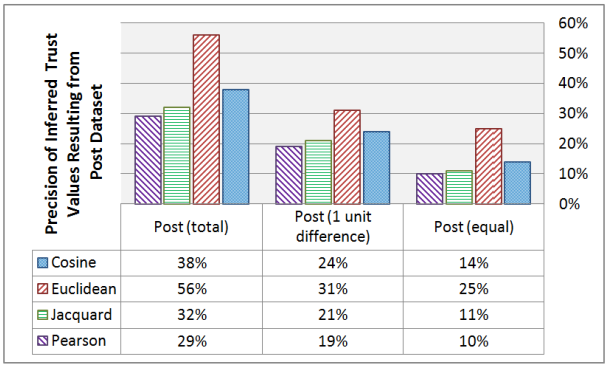
***a. Proportional Weightening*:** The first approach includes assigning values corresponding to the precision of each dataset. In other words, the more precise a dataset, the greater coefficient it will have.

For example, the precision of Cosine similarity resulting from the three datasets was 44% from Profile,



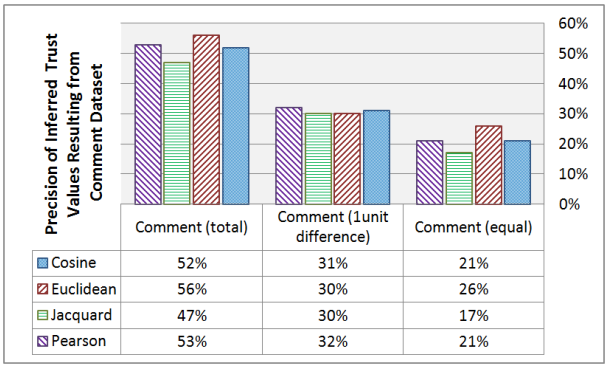
	Profile (total)	Profile (1 unit difference)	Profile (equal)
Cosine	44%	29%	15%
Euclidean	50%	28%	22%
Jacquard	44%	32%	12%
Pearson	44%	29%	15%

(a) Precision of trust values inferred from Profile dataset in comparison with real values assigned by users



	Post (total)	Post (1 unit difference)	Post (equal)
Cosine	38%	24%	14%
Euclidean	56%	31%	25%
Jacquard	32%	21%	11%
Pearson	29%	19%	10%

(b) Precision of trust values inferred from Post dataset in comparison with real values assigned by users



	Comment (total)	Comment (1unit difference)	Comment (equal)
Cosine	52%	31%	21%
Euclidean	56%	30%	26%
Jacquard	47%	30%	17%
Pearson	53%	32%	21%

(c) Precision of trust values inferred from Comment dataset in comparison with real values assigned by users

Figure 6. Comparing precision resulted from matching users' rates with the inferred ones from datasets.

38% from Post and 52% from Comment. Hence the corresponding coefficient vector will be $(0.33, 0.28, 0.39)$.

In this approach, if the values resulting from each dataset - Profile, Post and Comment - are 2, 4, and 2. The final value will be equal to: $0.33 * 2 + 0.28 * 4 + 0.39 * 2 = 2.56 \approx 3$. Figure 7 displays the precision of Weightening approach in prediction of trust values. The vector in front of each measure of similarity is the effective coefficient vector in the calculations. It can be inferred from figure 7 that, not unlike to the basic evaluation (figure 6), Euclidean similarity outperforms the other measures in this approach as well.

***b. Random Weightening*:** another approach used in this study is assigning random values to the coefficient vector. In this approach we generate random values for (α, β, γ) for thousand times and pick up the most precise vector from all the vectors generated randomly. Figure 8 depicts the precision resulting from Random Weightening.

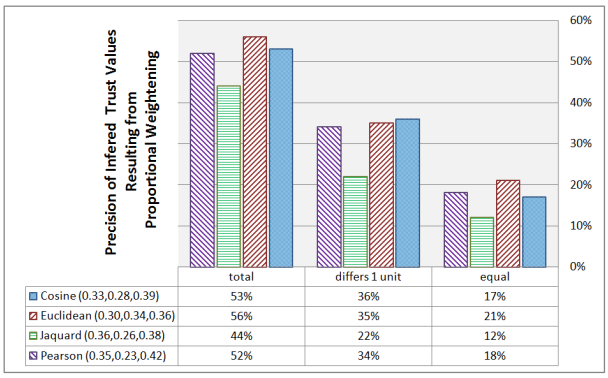
As can be viewed from the figure, in this approach Cosine Similarity has the highest precision (59%) followed by Euclidean Similarity (58%) which has only a slightly lower value.

It is noteworthy here that random Coefficient vectors are not limited to the vectors in front of each measure similarity in figure 8. However, the best precisions are resulted from these probable coefficient vectors.

***c. Simple Majority Voting*:** in addition to the coefficient vector, we also examined the Voting approach. In this approach, different datasets are considered as different classifiers, and the results of these classifiers will be gathered to gain one trust value.

The value which has the most consensuses will be chosen as the final value. In fact, one of the parameters of coefficient vector will be 1 and the two others will be zero. According to figure 9, the capability of Euclidean Similarity is again well-proved, but this approach fails to reach the maximum precision of Cosine Similarity obtained in Random Weightening approach.

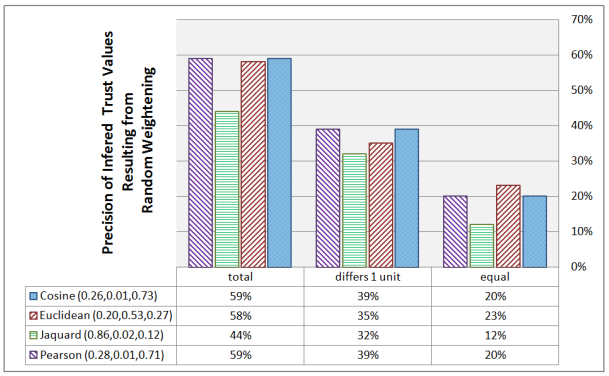
The interesting point here is that in most of the cases of using combined values, there is no trust value which differs 5 units from the expected value assigned by the users. Therefore, one van observe that using coefficient vectors is of another advantage in addition to the precision improvement.



	total	differs 1 unit	equal
Cosine (0.33,0.28,0.39)	53%	36%	17%
Euclidean (0.30,0.34,0.36)	56%	35%	21%
Jaquard (0.36,0.26,0.38)	44%	22%	12%
Pearson (0.35,0.23,0.42)	52%	34%	18%

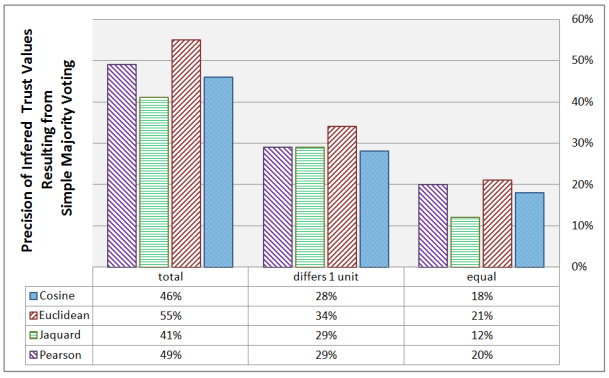
Figure 7. Precision resulted from combined values generated by Proportional Weightening approach

***d. Growing Coefficient*:** we have tested the effect of each coefficient by changing its value and divid-



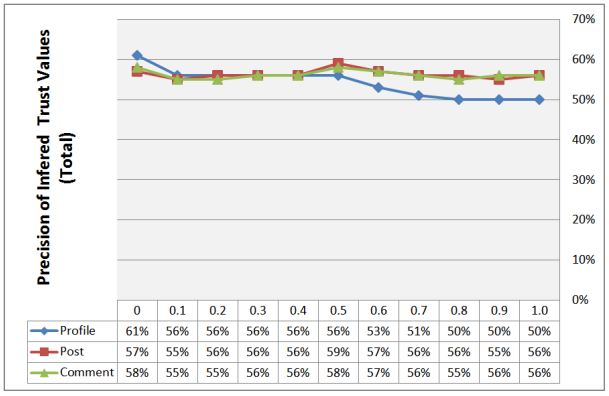
	total	differs 1 unit	equal
Cosine (0.26,0.01,0.73)	59%	39%	20%
Euclidean (0.20,0.53,0.27)	58%	35%	23%
Jaquard (0.86,0.02,0.12)	44%	32%	12%
Pearson (0.28,0.01,0.71)	59%	39%	20%

Figure 8. Precision resulted from combined values generated by Random Weightening approach



	total	differs 1 unit	equal
Cosine	46%	28%	18%
Euclidean	55%	34%	21%
Jaquard	41%	29%	12%
Pearson	49%	29%	20%

Figure 9. Precision resulted from combined values generated by Simple Majority Voting



	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Profile	61%	56%	56%	56%	56%	56%	53%	51%	50%	50%	50%
Post	57%	55%	56%	56%	56%	59%	57%	56%	56%	55%	56%
Comment	58%	55%	55%	56%	56%	58%	57%	56%	55%	56%	56%

Figure 10. Trend line of total precision of inferred trust values by changing the coefficients

ing the remaining value (1.0 - changing coefficient) equally between the two other coefficients. This way, the efficiency of a dataset can be traced by growing it's coefficient. In this study, only the Euclidean measure has been considered in the calculations.

Figure 10 illustrates the results of the analyses in this regard. In this figure, the horizontal vector represents the interval values of independent coefficient (the coefficient that the effect of whose value is under consideration). For each value of the independent coefficient, the remaining value from 1.0 is equally divided between the two other coefficients.

For example, the second value of the horizontal vector (0.1) signifies that the value of one of the coefficients is 0.1, and the remaining value (i.e. 0.9 = 1.0 - 0.1) is divided equally between the two other ones. Thus, if $\alpha = 0.1$, ($\beta = \gamma = 0.45$), the precision will be 56% (circle mark), if $\beta = 0.1$, ($\alpha = \gamma = 0.45$), the precision will drop to 55% (square mark), and finally if $\gamma = 0.1$, ($\alpha = \beta = 0.45$), the precision remains 55% (triangle mark).

Figure 10 shows that resulting precision ranges from 50% and 61% . Similar to the three previous approaches, this approach emphasizes the fact that changes in the coefficient vector can cause substantial variations in the precision of the results.

As can be viewed in Figure 10, the growth of α and β make no significant changes in the total precision, while greater γ results in less precision. The maximum precision (61%) through all four approaches is achieved here, when we set $\alpha = 0$ and completely ignore the Profile dataset.

4.5 Context-aware trust: 2nd step for improvement

As stated earlier, using context in trust management mechanisms is of multiple advantages, such as improving performance and reducing complexity, which have motivated us to investigate the benefits of using context in our proposed method.

As discussed in Section 3, we have used keywords extracted from News website in four different categories to organize our contextual filtering. These categories include Science, Arts, Sport and Politics. In this section, the efficiency of each context in inferring context-aware trust will be studied in details.

Table 7 summarizes the precision of our method after considering context in the calculations. As can be viewed, the slightest improvement is observable in the context of Science with the precision of 54% (Table 7.a). Moreover, the filtering text related to Sport and inferring trust in this context have also resulted in a significant improvement (Table 7.c). In fact, the best result is obtained in the context of Sport, with the precision of 72%. The precision of the two other contexts, i.e. Art and Politics, stand between these two boundary values. According to the data, the artistic and Politics trust values have improved with precision of 64% (Table 7.b) and 67% (Table 7.d) respectively

Table 7. Accuracy of context-aware trust

Dataset	Inferred trust equal to real values(%)	Inferred trust with one unit difference(%)	Total precision
profile	21%	24%	45%
post	20%	29%	49%
comment	21%	21%	42%
coefficient vector (0.37, 0.56, 0.27)	20%	34%	54%

a. context-aware trust: Science

Dataset	Inferred trust equal to real values(%)	Inferred trust with one unit difference(%)	Total precision
profile	26%	30%	56%
post	23%	35%	58%
comment	26%	28%	54%
coefficient vector (0.31, 0.55, 0.14)	26%	40%	66%

b. context-aware trust: Art

Dataset	Inferred trust equal to real values(%)	Inferred trust with one unit difference(%)	Total precision
profile	33%	28%	61%
post	25%	26%	51%
comment	36%	27%	63%
coefficient vector (0.37, 0.36, 0.27)	29%	43%	72%

c. context-aware trust: Sport

Dataset	**Inferred trust equal to real values(%)**	**Inferred trust with one unit difference(%)**	**Total precision**
profile	24%	30%	54%
post	24%	36%	60%
comment	27%	26%	53%
coefficient vector (0.39, 0.52, 0.09)	25%	41%	67%

d. context-aware trust: Politics

Although all the four contexts have experienced an improvement in their results, the amount of this improvement varies from context to context. Having studied the results of these differences, we found that the providing more texts for the context will result into more precise estimations. For example, most of the texts we gathered, including posts, comments and even profile, were about sports, indicating that experimental users talk about sports more often than the other three contexts. In contrast, the least available resources appeared to be those related to Science.

Not surprisingly, it was observed from the data that people, especially young friends, tend to talk about scientific issues less frequently than they do about every other topic such as sport, art and social events including politics and economic concerns. Thus, it can be concluded that more resources (text) will result in higher precision.

5 Discussion

In the previous sections, we introduced our proposed approach for inferring trust values regardless of direct rating., The experimental results were also presented in Section 4. In the present section, precisions resulting from four different similarity measures are depicted. It was observed from the analyses that in most cases, Euclidean Similarity results in better precision compared to the other three measures. In fact, this measure achieved an accuracy of 56% when applied to comment dataset regardless of context.

We also scrutinized the efficiency of combining values resulting from different datasets. In this regard, Coefficient Vector was introduced, which was found to aggregate trust values resulting from different datasets and offer a new value for the corresponding values. In addition, we studied the efficiency of this vector with four different approaches for weightening the coefficients. it was indicated that using the coefficient vector will improve the precision up to 61%. It has to be noted that this maximum value was achieved when we completely ignored the Profile dataset.

Furthermore, investigations were carried out to see the role of considering context in calculating similarity and inferring trust. This experience brought about promising results. Interestingly, it was observed that the precision of our approach increased to 72% in the context of Sport. Further studies on our datasets revealed that the more text available in one context, the more accurate the inferred trust values will be.

Above all and despite all the satisfactory results, there still remains an important point to be considered in this respect. In fact, in the first glance, it may be hard to believe that the similarity in terms of the words utilized by users can be used as a measure of evaluating trustworthiness, as one may assume that users with conflicting interests (and so with no trust in each other) may use similar words in their comments and posts.

Experimental results, however, indicate that this statement may only be true if a small amount of text is shared by users. In fact, if we consider only a small collection of users' texts, we would be misguided and it cannot lead us to accurate similarity values. In contrast, in a period of five months with a considerable amount of text, the probability that two users with different personalities talk similarly and share similar posts is really rare, as it can be found in real world as well.

In spite of the advantages of the prposed method, there were also some challenges faced with in the process of conducting this study. First, gathering the

required texts and creating the datasets through Facebook appeared to be a very demanding and time-consuming process. Besides, contacting the users to ask them to participate in the study was also a difficult task, which also took a long time.

6 Conclusion and Future Works

In the present paper, a new method was proposed to infer trust values from the similarities existing among users. The main advantage of the proposed approach is that it does not need direct ratings of users. Accordingly, it can be used in any social networks where some contextual information about the users is available. The information employed in this method mainly consists of the text shared by the users within social networks.

The results of the present study with regard to this new method have revealed that it can be used in a variety of fields besides its application in every social network. In fact, this approach can help recommender systems present recommendations more effectively. It also improves ratings confidentiality in the direct rating systems.

In our future studies, we are going to change the method so that it calculates the similarity for our inference engine. Beside TF-IDF, there are some other approaches, such as KL-Divergence, which give an estimate of text similarity. Further experiments are to be done in order to examine the efficiency of these methods in our proposed approach to find the most appropriate technique in this regard.

In addition to the applications of the proposed approach in social networks as well as the other areas as mentioned by Golbeck [4], the consistency of our approach with any community with a graph structure can make it an effective instrument to improve the security of trust and reputation mechanisms against various threats such as malicious peers and collectives, malicious spies and driving down the reputation of a reliable peer [32]. The method can be used on the network implicitly to make comparisons between the ratings assigned by users and the expected values. Every mismatch could be a biased value. This way, the efficiency of similar methods which use direct rating could be also studied.

After inferring the level of trust, these inferred values can be used to calculate the reputation. According to the literature [7], reputation is defined as a peer's belief in another peer's capabilities, honesty and reliability based on recommendations received from other peers. An assumption can be that a person's recommendation is directly related to the trust value which the person has. Therefore, it is believed that the inferred trust values lead us to reputation values. Especially when context is taken into consideration, context-aware reputation can be calculated easily as well.

Acknowledgements

This work is partially supported by Research Institute for Information and Communication Technology (ITRC) [17170/500]. In addition I warmly thank Mr. Hossein Homaei, PhD Candidate, for his valuable advices and friendly help.

References

[1] A. Josang, R. Ismail, and C. Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618 – 644, 2007. ISSN 0167-9236.

[2] J. Golbeck. Computing and applying trust in webbased social networks. In *Proceedings of Computational Intelligence and Industrial Application*, PACIIA '08, pages 342–347, 2008.

[3] P. Avesani, P. Massa, and R. Tiella. A trust-enhanced recommender system application: Moleskiing. In *Proceedings of the 2005 ACM Symposium on Applied Computing*, SAC '05, pages 1589–1593, New York, NY, USA, 2005. ACM. ISBN 1-58113-964-0.

[4] J. Golbeck. Trust and nuanced profile similarity in online social networks. *ACM Trans. Web*, 3 (4):12:1–12:33, September 2009. ISSN 1559-1131.

[5] CN. Ziegler and J. Golbeck. Investigating interactions of trust and interest similarity. *Decision Support Systems*, 43(2):460–475, 2007.

[6] M.A. Morid, A. Omidvar, and HR. Shahriari. An enhanced method for computation of similarity between the contexts in trust evaluation using weighted ontology. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on*, pages 721–725, 2011.

[7] Y. Wang and J. Vassileva. Trust and reputation model in peer-to-peer networks. In *Peer-to-Peer Computing*, pages 150–157. IEEE Computer Society, 2003. ISBN 0-7695-2023-5.

[8] D. Artz and Y. Gil. A survey of trust in computer science and the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):58 – 71, 2007. ISSN 1570-8268.

[9] L. Mui, M. Mohtashemi, and A. Halberstadt. A computational model of trust and reputation. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*,

2002.

[10] M. Grandison, T. Sloman. A survey of trust in internet applications. *Communications Surveys Tutorials, IEEE*, 3(4):2–16, 2000. ISSN 1553-877X.

[11] D. Olmedilla, OF. Rana, B. Matthews, and W. Nejdl. Security and trust issues in semantic grids. In *Proceedings of the Dagsthul Seminar, Semantic Grid: The Convergence of Technologies*, volume 5271, 2005.

[12] M. Tavakolifard, SJ. Knapskog, and P. Herrmann. Trust transferability among similar contexts. In *Q2SWinet*, pages 91–97. ACM, 2008. ISBN 978-1-60558-237-5.

[13] R. Neisse, M. Wegdam, and M. Van Sinderen. Context-aware trust domains. In *EuroSSC*, volume 4272 of *Lecture Notes in Computer Science*, pages 234–237. Springer, 2006. ISBN 3-540-47842-6.

[14] R. Neisse, M. Wegdam, M. Sinderen, and G. Lenzini. Trust management model and architecture for context-aware service platforms. In *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS*, volume 4804 of *Lecture Notes in Computer Science*, pages 1803–1820. Springer Berlin Heidelberg, 2007. ISBN 978-3-540-76835-7.

[15] S. Holtmanns and Z. Yan. Context-aware adaptive trust. In *Developing Ambient Intelligence*, pages 137–146. Springer Paris, 2006. ISBN 978-2-287-47469-9.

[16] M. Rehak, M. Gregor, M. Pechoucek, and J. Bradshaw. Representing context for multiagent trust modeling. In *IAT*, pages 737–746. IEEE Computer Society, 2006.

[17] S. Toivonen and G. Denker. The impact of context on the trustworthiness of communication: An ontological approach. In *ISWC Workshop on Trust, Security, and Reputation on the Semantic Web*, volume 127 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2004.

[18] M. Rehak and M. Pechoucek. Trust modeling with context representation and generalized identities. In *CIA*, volume 4676 of *Lecture Notes in Computer Science*, pages 298–312. Springer, 2007. ISBN 978-3-540-75118-2.

[19] N. Gujral, D. DeAngelis, K. Fullam, and K. Barber. Modeling multi-dimensional trust. In *Proceedings of the Workshop on Trust in Agent Societies*, 2008.

[20] EW. Burgess and P. Wallin. Homogamy in social characteristics. *American Journal of Sociology*, 49:pp. 109–124, 1943. ISSN 00029602.

[21] T. Newcomb. The acquaintance process. *Holt, Rinehart, and Winston*, 1961.

[22] D. Byrne. Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology*, 62:pp. 713–715, 1961.

[23] D. Byrne. The attraction paradigm. *Academic Press*, 1971.

[24] J. Golbeck, C. Robles, M. Edmondson, and K. Turner. Predicting personality from twitter. In *SocialCom/PASSAT*, pages 149–156. IEEE, 2011. ISBN 978-1-4577-1931-8.

[25] D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our twitter profiles, our selves: Predicting personality with twitter. In *SocialCom/PASSAT*, pages 180–185. IEEE, 2011. ISBN 978-1-4577-1931-8.

[26] J. Tang, H. Gao, X. Hu, and H. Liu. Exploiting homophily effect for trust prediction. In *WSDM*, pages 53–62. ACM, 2013. ISBN 978-1-4503-1869-3.

[27] H. Mohammadhassanzadeh and HR. Shahriari. Using user similarity to infer trust values in social networks regardless of direct ratings. In *Information Security and Cryptology (ISCISC), 2012 9th International ISC Conference on*, pages 66–72, 2012.

[28] A. Huang. Similarity measures for text document clustering. In *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC2008), Christchurch, New Zealand*, pages 49–56, 2008.

[29] M. Porter. An algorithm for suffix stripping. *Program: electronic library and information systems*, 14(3):130–137, 1980.

[30] A. Sharifloo and M. Shamsfard. A bottom up approach to persian stemming. In *IJCNLP*, 2008.

[31] A. Strehl, J. Ghosh, and R. Mooney. Impact of similarity measures on web-page clustering. In *Proceedings of the AAAI Workshop on AI for Web Search (AAAI 2000)*, pages 58–64, Austin, TX, USA, 2000.

[32] FG. Marmol and GM. Perez. Security Threats Scenarios in Trust and Reputation Models for Distributed Systems. *Elsevier Computers and Security*, 2009.

Hossein Mohammadhassanzadeh was born in 1987 in Iran. He got his B.Sc. (2009) and M.Sc. (2013) in Engineering of Information Technology, from University of Isfahan and Amirkabir University of Technology respectively. He started his career as a PhD student at the faculty of Computer Science in Dalhousie University, Halifax, Canada in 2013. His research interest includes Data Mining, Information Retrieval, Social Network Analysis, Financial Engineering, and recently Knowledge Management and Health Informatics Systems.

Hamid Reza Shahriari is currently an assistant professor at the Department of Computer Engineering and Information Technology in Amirkabir University of Technology in Tehran, Iran. He received his Ph.D. in computer science from Sharif University of Technology in 2007. His research interests include information security especially in vulnerability analysis, security in e-commerce, trust and reputation models, and database security.