

Trust and Nuanced Profile Similarity in Online Social Networks

JENNIFER GOLBECK

University of Maryland, College Park

Online social networks, where users maintain lists of friends and express their preferences for items like movies, music, or books, are very popular. The Web-based nature of this information makes it ideal for use in a variety of intelligent systems that can take advantage of the users' social and personal data. For those systems to be effective, however, it is important to understand the relationship between social and personal preferences. In this work we investigate features of profile similarity and how those relate to the way users determine trust. Through a controlled study, we isolate several profile features beyond overall similarity that affect how much subjects trust hypothetical users. We then use data from FilmTrust, a real social network where users rate movies, and show that the profile features discovered in the experiment allow us to more accurately predict trust than when using only overall similarity. In this article, we present these experimental results and discuss the potential implications for using trust in user interfaces.

Categories and Subject Descriptors: H.4.2 [Information Systems Applications]: Types of Systems—*Decision support*

General Terms: Human Factors

Additional Key Words and Phrases: Social networks, trust, recommender systems

ACM Reference Format:

Golbeck, J. 2009. Trust and nuanced profile similarity in online social networks. *ACM Trans. Web* 3, 4, Article 12 (September 2009), 33 pages.

DOI = 10.1145/1594173.1594174 <http://doi.acm.org/10.1145/1594173.1594174>

1. INTRODUCTION

Social notions of trust have become a topic of interest among the computer science community. Trust provides information about with whom we should share information, from whom we should accept information, and what consideration to give to information from people when aggregating or filtering data. At the same time, social networking has become a major movement on the Web, with hundreds of Web sites comprising hundreds of millions of users

Author's address: College of Information Studies, 2117K Hornbake Building, South Wing, University of Maryland, College Park, College Park, MD 20741; email: golbeck@cs.umd.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.
© 2009 ACM 1559-1131/2009/09-ART12 \$10.00

DOI 10.1145/1594173.1594174 <http://doi.acm.org/10.1145/1594173.1594174>

ACM Transactions on the Web, Vol. 3, No. 4, Article 12, Publication date: September 2009.

[Golbeck 2005]. Many social networks allow users to express information about their relationships, including how much they trust people. This provides a minimally invasive way to collect information about users' social contexts. By using information that users are already expressing, there is little overhead to gathering trust data which can then be analyzed and used to create socially intelligent systems.

To effectively use trust in systems it is important to understand what users mean when they say that they trust someone and how much they trust them. The sociological literature has studied reasons for trust extensively. However, when using Web-based social networks, most of the information that sociology considers important is not available (e.g., we do not know the history between people, the user's own background and how likely they are to trust in general, the familial/business/friend relationship between users). Thus, we must understand trust from only the available information.

In the case of trust in Web-based social networks, the information we have available is the social network, the trust values in it, and the profiles of users. Those profiles include personal information and frequently include the users' opinions and ratings of items. This information can be used to explain the magnitude of trust between people or to compute recommendations about how much one user should trust another. There are many algorithms that use only the social network and trust values to compute how much one user should trust another. These have been shown to give relatively accurate results, but are only effective when users are connected in the social network. Collaborative filtering (CF) algorithms, on the other hand, use only profile information when computing recommendations for users. CF algorithms generally compute the overall similarity or correlation between users, and use that as a weight when making recommendations. These algorithms are relatively good at making recommendations for users, and could be applied to computing trust recommendations, but they are only effective when users have a common set of rated items. When the ratings are sparse, it is difficult to compute similarity measures between users.

How profile similarity relates to trust is a relatively unexplored space. Trust-based recommender systems like Ziegler [2005], Massa and Bhattacharjee [2004], and Golbeck [2006b] assume trust captures similarity between users. The relationship between trust and overall similarity was shown in Ziegler and Golbeck [2006]. However, in previous work, we saw that trust-based recommendations *outperformed* collaborative filtering algorithms in certain cases [Golbeck 2005]. Because collaborative filtering algorithms use overall similarity of user profiles to make recommendations, those results suggested that when users assign trust, they are capturing more than just overall similarity.

In this work, we explore the relationship between trust and profile similarity. We will show through surveys and analysis of data in existing systems that when users express trust, they are capturing many facets of similarity with other users. We begin by presenting a definition of trust and related work on similarity, trust, and trust computation. We then present a study where users are given generated profiles for hypothetical users, and from the results we extract several features of profile similarity that correlate with trust. Those

results are then brought to data taken from our existing FilmTrust system, and we show that using that set of features to predict trust is better correlated with known trust values, and is more accurate than using overall similarity alone. Finally, we discuss the implications of this work for creating intelligent systems that respect users' social preferences.

2. BACKGROUND AND RELATED WORK

As a social concept, trust has many facets and influences that are beyond the scope of what can be captured in social networks or online systems. To compute with trust, it is important to know precisely what is being captured and what its features are. In this section, we present a definition of trust and related work on how people assign trust in general, and in online systems.

2.1 Web-Based Social Networks

Web-based social networks have grown dramatically in the last five years [Golbeck 2007]. There are hundreds of Web sites with over a billion user accounts among them.¹ All this publicly available data is a great resource for doing research, but it also has some challenges.

In particular, using online social networks provides only a limited snapshot of the users and their interactions. Typically, the networks only provide a list of friends for each user. Few show the last date that a given user was active, and none show more detailed information about a user's history of interaction with the Web site. No networks publicly share a history of interactions between users, either. Communications are kept private, and the formation or removal of relationships are not shown or recorded in a user's profile. Essentially, at a given point in time, only the current set of connections in the network are available, with little to no background information or history to support it. This prevents the application of more advanced social network models which rely on histories of users and interactions, and is an important consideration when developing algorithms to work in this environment.

These networks all share similar structural properties. They all have small world properties [Watts 2003; Kleinberg 2000] and their degree distribution follows the power law distribution [Barabasi and Albert 1999]. This means the graphs are sparse, but well connected. The average shortest path between any two nodes grows logarithmically with the size of the network, and clustering around a given node is high relative to what would be expected in a random graph.

2.2 A Definition of Trust

Social trust depends on a host of factors which cannot be easily modeled in a computational system. Past experience with a person and with their friends, opinions of the actions a person has taken, psychological factors impacted by a lifetime of history and events (most completely unrelated to the person we are deciding to trust or not trust), rumor, influence by others' opinions, and motives

¹See <http://trust.mindswap.org/cgi-bin/relationshipTable.cgi>.

to gain something extra by extending trust are just a few of these factors. For trust to be used as a rating between people in social networks, the definition must be focused and simplified.

Marsh [1994] addressed the issue of formalizing trust as a computational concept in his Ph.D. dissertation at the University of Stirling. His model is complex and based on social and psychological factors. Although this work is often cited, the model is highly theoretical and difficult to implement. It is particularly inappropriate for use in social networks because his focus was on interacting agents that could maintain information about history and observed behaviors. In all Web-based social networks that implement trust, users assign a single rating without explicit context or history to their neighbors and thus much of the information necessary for a system like Marsh's is missing. To capture the nuanced information required by an implementation of Marsh's system, users would need to log tremendous amounts of information, and even then the complexity of data required by Marsh's model would be unsatisfied. One of the stated goals of our work is to take advantage of social networks as they exist on the Web with users adding simple expression of trust as they determine it (rather than computing it for them). This differs from the agent system in Marsh's work, and that divergence in goals prevents an application of his model in this context.

Castelfranchi and Falcone [1998, 2002] present an excellent analysis of trust in multi-agent systems. They break trust down into its components, including the beliefs that must be held to develop trust, how that relates to previous experience, as well as giving descriptions of when trust is rational and irrational. Their work draws largely on the psychological literature, and includes many psychological factors in developing a model for trust in multi agent systems. In Web-based social networks, however, this information is not made explicit. Networks do not present (nor, generally, do they even have access to) background information, and unless a person owns the logs of a social network (which are never released to the public), there is no record of historical interactions. Thus, the information necessary to apply the results from this work is not available. In the research presented in this article, we want to show that useful results can be produced without this extensive background. Because the data that we have chosen as our focus is limited to single values, the data necessary to apply the more advanced and complex model of Castelfranchi and Falcone is not available.

Deutsch [1962] contains a frequently referenced definition of trust. He states that trusting behavior occurs when a person (say Alice) encounters a situation where she perceives an ambiguous path. The result of following the path can be good or bad, and the occurrence of the good or bad result is contingent on the action of another person (say Bob). Furthermore, the negative impact of the bad result is greater than the positive impact of the good result. This further motivates Alice to make the correct choice. If Alice chooses to go down the path, she has made a trusting choice. She trusts that Bob will take the steps necessary to ensure the good outcome. The requirement that the bad outcome must have greater negative implications than the good outcome has positive implications has been countered in other work [Golebmiewski and McConike 1975], which does not always require disparity.

Sztompka [1999] presents and justifies a simple, general definition of trust similar to that of Deutsch: “Trust is a bet about the future contingent actions of others.” There are two main components of this definition: belief and commitment. First, a person believes that the trusted person will act in a certain way. The belief alone, however, is not enough to say there is trust. Trust occurs when that belief is used as the foundation for making a commitment to a particular action. These two components are also present in the core of Deutsch’s definition: we commit to take the ambiguous path if we believe that the trusted person will take the action that will produce the good outcome.

We adopt this as the definition of trust for our work: trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome. The action and commitment does not have to be significant. We could say Alice trusts Bob regarding email if she chooses to read a message (commits to an action) that Bob sends her (based on her belief that Bob will not waste her time). Since trust is usually not a simple binary value (trust or no trust), we consider it to be similar to a probability that a person will commit to an action. When subjects in our experiments and users of our systems are asked to assign trust, they are given this definition.

2.3 Trust and Similarity

The sociology and social psychology literature extensively addresses factors in trust, but the coverage does not directly address how trust relates to similarity (overall or in more specific ways). There is, however, strong results linking similarity and friendship, or interpersonal attraction. A positive relationship between attitude similarity and friendship has been shown in Burgess and Wallin [1943], Newcomb [1961], Byrne [1961], and [Byrne 1971].

In the recommender systems literature, there have been several small studies that touch on similarity and trust. Sinha and Swearingen [2001] and Swearingen and Sinha [2001] have shown that users tend to prefer recommendations from friend than from systems. The authors suggest that this is because people have more trust for friends.

The connection between user similarity and trust was most strongly established in Ziegler and Golbeck [2006]. Through an analysis of two systems, this work showed that there was a strong and significant correlation between trust and user similarity; the more similar two people were, the greater the trust between them.

The correlation of trust and similarity is not surprising. However, does trust *only* capture overall similarity between people? Clearly it reflects other social factors. Is it possible that trust also captures more nuanced facets of correlation between users in online systems? In related work on the FilmTrust system we found results that suggest users’ trust ratings do capture more than just overall similarity. In Section 3 we present those results and describe how they have motivated this work, which looks further into how similarity affects trust.

2.4 Computing Trust from Trust

In a system that has a trust component, users will have made some direct statements about people they trust. These statements form a social network. In this article we are examining how trust corresponds to different features of profile similarity, but if we have direct trust information from the users, that can be a more effective means of predicting how much one person will trust another.

In fact, there is a large body of work on algorithms for inferring trust in social networks. While designed for peer-to-peer systems rather than social networks, one of the most widely cited trust algorithms is EigenTrust [Kamvar et al. 2004]. It considers trust as a function of corrupt vs. valid files that that a peer provides. A peer maintains information about the trustworthiness of peers with which it has interacted based on the proportion of good files it has received from that peer. For one peer to determine the trustworthiness of another with which it has not interacted, it needs to gather information from the network and infer the trustworthiness. The EigenTrust algorithm calculates trust with a variation on the PageRank algorithm [Page et al. 1998], used by Google for rating the relevance of Web pages to a search.

Advogato is a Web site, at <http://advogato.org>, that serves as a community discussion board and resource for free software developers. It also is the testbed for Raph Levin's trust metrics research [Levin and Aiken 1998]. Each user on the site has a single trust rating calculated from the perspective of designated seeds (authoritative nodes). Trust calculations are made using a network flow model. His metric composes certifications between members to determine the trust level of a person, and thus their membership within a group. Users can be certified at three levels: apprentice, journeyer, and master.

Ziegler and Lausen [2004] propose a trust algorithm called Appleseed. Like Advogato, it is a group trust metric. However, instead of using maximum flow, the basic intuition is motivated by spreading activation strategies. Like EigenTrust, Appleseed is based on finding the principal eigenvector. It is a local trust metric, and given a network and a source it returns a ranking of all the nodes in the network.

One problem that arises in algorithms that are based on finding the principal eigenvector, like Kamvar et al. [2004], Ziegler and Lausen [2004], and Richardson et al. [2003], is that trust must first be normalized to work within the matrix. This means that the normalized trust value from a person who has made many trust ratings will be lower than if only one or two people had been rated. However, socially, trust is not a finite resource; it is possible to have very high trust for a large number of people, and that trust is not any weaker than the trust held by a person who only trusts one or two others.

Golbeck [2005] introduces the TidalTrust algorithm for computing personalized trust values in social networks. Unlike the eigenvalue-based approaches, it outputs a trust recommendation in the same scale that users assign trust values.

These algorithms are generally quite effective at predicting trust and for improving recommendations in certain cases. However, there are drawbacks

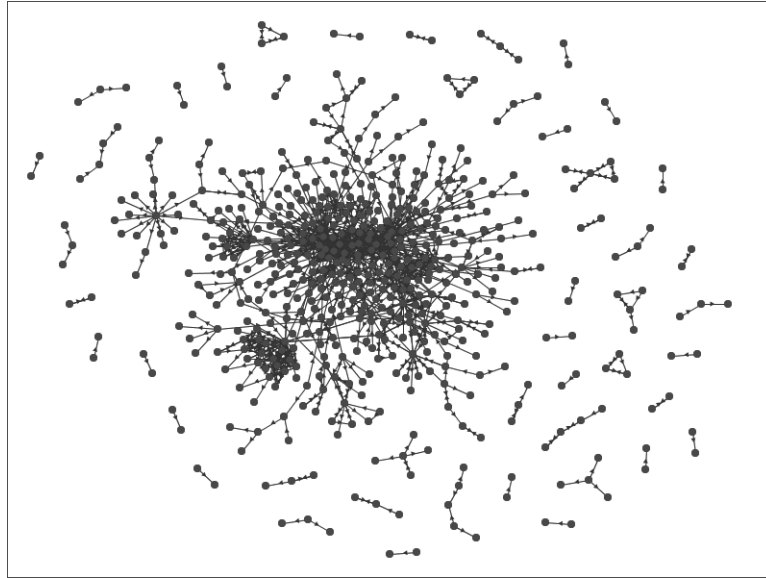


Fig. 1. The FilmTrust social network. Users with no friends are not shown, although they constitute 58% of the population. Other users are not connected into the main component and are shown in the small clusters scattered around the edges.

that make them inapplicable in many cases. Specifically, they can only be used when there are paths in the social network connecting individuals *and* when the trust values on those paths are accessible. In some cases, paths between two people will not exist in the network. In our own system, FilmTrust, over 58% of the users have *no* friends in the system, and over 60% are connected only into small clusters, detached from the main component (see Figure 1). Thus, for the majority of users, network-based trust algorithms will be ineffective. Also, while the FilmTrust system uses the trust values of any user to compute recommendations, privacy is always a concern. It is reasonable to assume that many users or networks will not allow a system to access trust values, directly or indirectly. In that case, the only trust information available is the set of values assigned directly by the user.

2.5 Collaborative Filtering

Collaborative filtering recommends items to users based on their similarity to other users and/or the similarity of the items they like to other items. The trust-based techniques we present in this article are an alternative embodiment of some of these systems, here, we present related work on the major types of collaborative filtering systems.

2.5.1 User-Based Collaborative Filtering. First introduced in the Tapestry system [Goldberg et al. 1992], user-based collaborative filtering generally computes the similarity or correlation between two users, and uses this measure to weight their ratings of items. One of the early applications was to filtering

Usenet News with the GroupLens system [Konstan et al. 1997a]. Herlocker et al. [1999] provided a framework for developing collaborative filtering algorithms, and investigated variations on these correlation and similarity measures, including the use of z-scores and alternative correlation measures (e.g., Spearman vs. Correlation). There have been countless variations and optimizations to improve the accuracy of these algorithms [Ma et al. 2007; Huang et al. 2004; Jin et al. 2004]. Other ways of improving systems to make them more satisfactory to the user rather than simply more accurate were also presented in Herlocker et al. [2004].

There have been a number of user-based collaborative filtering methods developed for movie recommendations in particular. MovieLens [Herlocker et al. 2000; Miller et al. 2003], Recommendz [Garden and Dudek 2005], and Film-Conseil [Perny and Zucker 2001] are just a few of the Web sites that implement recommender systems in the context of films.

2.5.2 Model-Based Collaborative Filtering. One of the drawbacks of user-item collaborative filtering is the sparsity of data. In large systems, it is often the case that users have few, if any, items rated in common. Scalability is also an issue; the computational power required to find similar users increases as the size of the system increases. To address these issues, a number of model-based collaborative filtering algorithms have been developed.

Machine learning techniques have been an effective way to model user preference. Clustering is one technique that has shown improved performance over simpler models. Clustering can be treated as a classification problem, where users are grouped into classes [Ungar and Foster 1998; Breese et al. 1998; Basu et al. 1998]. Coclustering methods that cluster both users and items have been shown to be effective and computationally efficient [George and Merugu 2005]. Bayesian network models [Breese et al. 1998; Chien and George 1999] build a probabilistic model of the users' preferences, generating expected values for user ratings of items.

Matrix-factorization methods build a ratings matrix and predict the missing values. These approaches include singular value decomposition (SVD) [Sarwar et al. 2000] and nonnegative matrix factorization (NNMF) [Srebro and Jaakkola 2003]. Similar to clustering methods and using matrix factorization is latent semantic analysis [Hofmann 2004], a technique translated from information retrieval to collaborative filtering.

While user-based collaborative filtering algorithms look for people most similar to the user, item-based collaborative filtering [Sarwar et al. 2001] works with the set of items. The algorithms look at items rated by the user, and then identify items similar to those to recommend. This approach has been shown to have good performance and quality. Finally, some work such as Xue et al. [2005] and Wang et al. [2006] has integrated both user and model based approaches.

3. FILMTRUST WEB SITE

In earlier work we developed the FilmTrust Web site, a trust-based social networking website where users can rate and review movies. We used trust in that system to produce predictive movie ratings, and the interesting results that

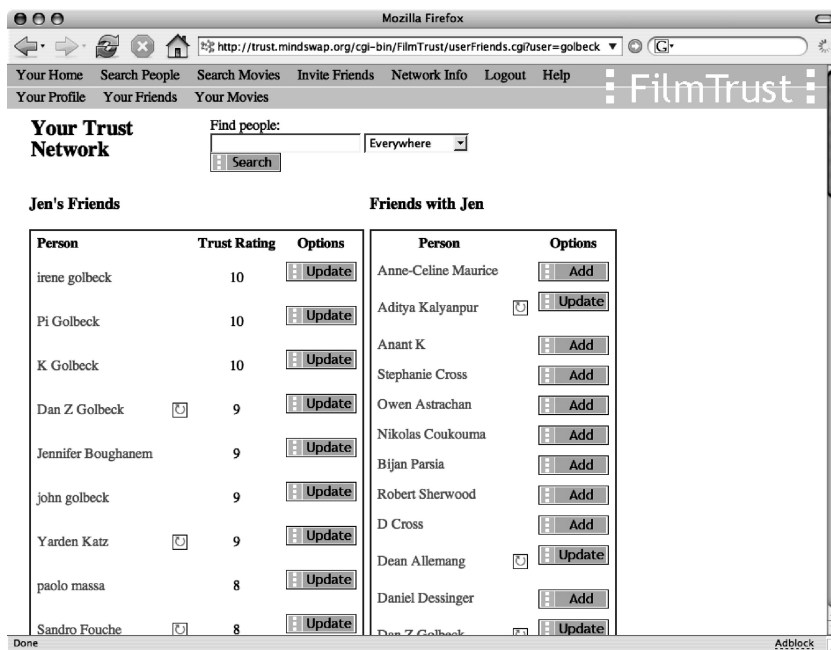


Fig. 2. A sample Friends page from FilmTrust. Friends of the user are on the left, with the user's trust ratings. People who have listed the user as a friend are on the right.

we obtained in our analysis motivated this work that further examines trust and user similarity. We will also use data from FilmTrust in the analysis in Section 6.

In this section, we introduce the FilmTrust Web site and present an overview of the relevant results. More thorough discussions of this system and the results can be found in Golbeck [2005], Golbeck [2006b], and Golbeck and Hendler [2006].

3.1 The FilmTrust Web Site

The FilmTrust Web site, at <http://trust.mindswap.org/FilmTrust>, is a social network with trust ratings. Users maintain the standard social network information, and also rate movies and write reviews.

In the social networking part of the Web site, users make lists of their friends. They also are required to assign a trust rating to each friend that indicates how much they trust that person's opinions about movies. Trust is asymmetric—people do not necessarily trust each other the same amount—and we believe friendship relationships are also asymmetric. Thus, relationships are not required to be mutual in our system. Users keep a list of friends, but the friends do not approve the relationship. Thus, on the user's "Friends" page (see Figure 2), there are two lists: the people the user has listed as friends (with trust ratings), and the people who have listed the user as a friend. Because of the sensitive nature of trust ratings, the user is the only person who can see the trust values they assign. These values are hidden from everyone else.

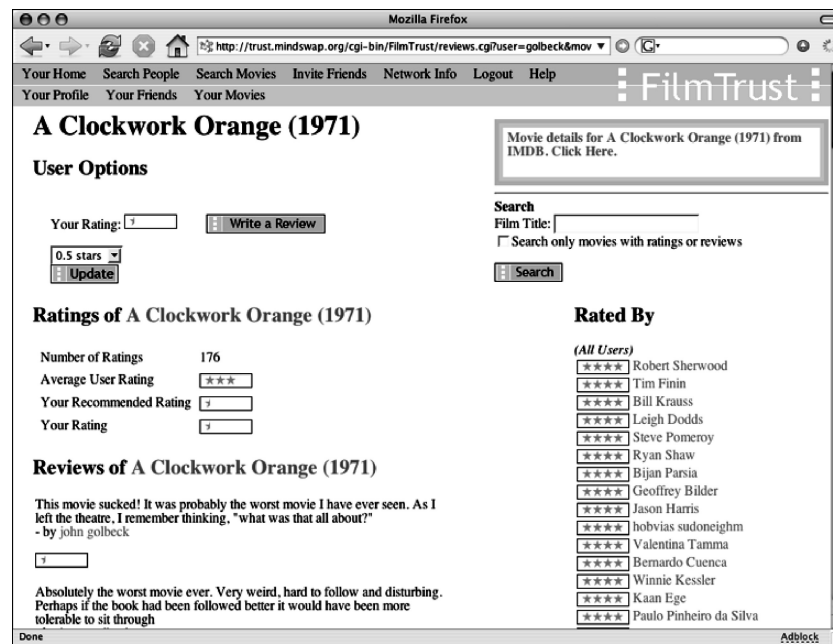


Fig. 3. The movie page on FilmTrust for the film *A Clockwork Orange*.

In the movies section of their space, users can rate and review films. Movie ratings are on a scale where a half star is the worst rating, and 4 stars is the best rating. Reviews are free text and are not limited in length.

The social network is combined with the movie data on the movie information pages within the site (see Figure 3). In the next section, we describe how the two are combined.

3.2 Using Trust in FilmTrust

On the movie information pages, we use trust to create predictive movie ratings, that is, we offer a rating that we believe will best match the user's opinion. This is a common task undertaken by collaborative filtering algorithms. While those systems use similarity measures between users to generate a rating, we rely on trust. The basic premise of our algorithm is to compute a weighted average of the ratings assigned to a film, using trust as a weight. If the user does not have a direct trust rating for a person who has rated the movie, we compute an inferred trust value using the TidalTrust algorithm. For further details, see Golbeck [2005].

3.2.1 Creating Predictive Movie Ratings. One of the features of the FilmTrust site that uses the social network is the "Recommended Rating" feature. As Figure 3 shows, users will see this in addition to the average rating for a movie.

The "Recommended Rating" is personalized using the trust values for the people who have rated the film (the raters). First, the system searches for raters

that the source knows directly. If there are no direct connections from the user to any raters, the system moves one step out to find connections from the user to raters of path length 2. This process repeats until a path is found. The opinion of all raters at that depth are considered. Then, using TidalTrust, the trust value is calculated for each rater at the given depth. Once every rater has been given an inferred trust value, only the ones with the highest ratings will be selected; this is done by simply finding the maximum trust value calculated for each of the raters at the selected depth, and choosing all of the raters for which that maximum value was calculated. Finally, once the raters have been selected, their ratings for the movie (in number of stars) are averaged. For the set of selected nodes S , the recommended rating r from node s to movie m is the average of the movie ratings from nodes in S weighted by the trust value t from s to each node:

$$r_{sm} = \frac{\sum_{j \in S} t_{sj} r_{jm}}{\sum_{i \in S} t_{si}}. \quad (1)$$

This average is rounded to the nearest half-star, and that value becomes the “Recommended Rating” that is personalized for each user.

As a simple example, consider the following:

- Alice trusts Bob 9.
- Alice trusts Chuck 3.
- Bob rates the movie “Jaws” with 4 stars.
- Chuck rates the movie “Jaws” with 2 stars.

Then Alice’s recommended rating for “Jaws” is calculated as follows:

$$\frac{t_{Alice \rightarrow Bob} * r_{Bob \rightarrow Jaws} + t_{Alice \rightarrow Chuck} * r_{Chuck \rightarrow Jaws}}{t_{Alice \rightarrow Bob} + t_{Alice \rightarrow Chuck}} = \frac{9 * 4 + 3 * 2}{9 + 3} = \frac{42}{12} = 3.5. \quad (2)$$

3.3 FilmTrust Evaluation

For each movie the user has rated, the recommended rating can be compared to the actual rating that the user assigned. In this analysis, we also compare the user’s rating with the average rating for the movie, and with a recommended rating generated by an automatic collaborative filtering (ACF) algorithm. There are many ACF algorithms, and one that has been well tested, and which is used here, is the classic user-to-user nearest neighbor prediction algorithm based on Pearson Correlation [Konstan et al. 1997b]. If the trust-based method of calculating ratings is best, the difference between the personalized rating and the user’s actual rating should be significantly smaller than the difference between the actual rating and the average rating.

On first analysis, it did not appear that the personalized ratings from trust had any benefit over the average. The difference between the user’s rating and the recommended rating (call this difference δ_r) was not statistically different from the difference between the user’s actual rating and the average rating

(call this δ_a). The difference between a user's actual rating of a film and the ACF calculated rating (δ_{cf}) also was not better than δ_a in the general case. A close look at the data suggested why. Most of the time, the majority of users' actual ratings are close to the average. This is most likely due to the fact that the users in the FilmTrust system had all rated the AFI Top 50 movies, which received disproportionately high ratings. A random sampling of movies showed that about 50% of all ratings were within the range of the mean \pm a half star (the smallest possible increment). For users who gave these near-mean ratings, a personalized rating could not offer much benefit over the average.

However, the point of the recommended rating is more to provide useful information to people who disagree with the average. In those cases, the personalized rating should give the user a better recommendation, because we expect the people they trust will have tastes similar to their own.

To see this effect, δ_a , δ_{cf} , and δ_r were calculated with various minimum thresholds on the δ_a value. If the recommended ratings do not offer a benefit over the average rating, the δ_r values will increase at the same rate the δ_a values do. The experiment was conducted by limiting δ_a in increments of 0.5. The first set of comparisons was taken with no threshold, where the difference between δ_a and δ_r was not significant. As the minimum δ_a value was raised it selected a smaller group of user-film pairs where the users made ratings that differed increasingly with the average. Obviously, we expect the average δ_a value will increase by about 0.5 at each increment, and that it will be somewhat higher than the minimum threshold. The real question is how the δ_r will be impacted. If it increases at the same rate, then the recommended ratings do not offer much benefit over the simple average. If it increases at a slower rate, that means that, as the user strays from the average, the recommended rating more closely reflects their opinions. Figure 4 illustrates the results of these comparisons.

Notice that the δ_a value increases about as expected. The δ_r , however, is clearly increasing at a slower rate than δ_a . At each step, as the lower threshold for δ_a is increased by 0.5, δ_r increases by an average of less than 0.1. A two-tailed t-test shows that at each step where the minimum δ_a threshold is greater than or equal to 0.5, the recommended rating is significantly closer to the actual rating than the average rating is, with $p < 0.01$. For about 25% of the ratings assigned, $\delta_a < 0.5$, and the user's ratings are about the same as the mean. For the other 75% of the ratings, $\delta_a > 0.5$, and the recommended rating significantly outperforms the average. As is shown in Figure 4, δ_{cf} closely follows δ_a . For $\delta_a < 1$, there was no significant difference between the accuracy of the ACF ratings and the trust-based recommended rating. However, when the gap between the actual rating and the average increases, for $\delta_a \geq 1$, the trust-based recommendation outperforms the ACF as well as the average, with $p < 0.01$. Because the ACF algorithm is only capturing overall correlation, it is tracking the average because most users' ratings are close to the average.

Figure 3 illustrates one of the examples where the recommended value reflects the user's tastes. *A Clockwork Orange* is one of the films in the database that have a strong collective of users who hated the movie, even though the average rating was 3 stars and many users gave it a full 4-star rating. For the user shown, $\delta_a = 2.5$ a very high value while the recommended rating exactly

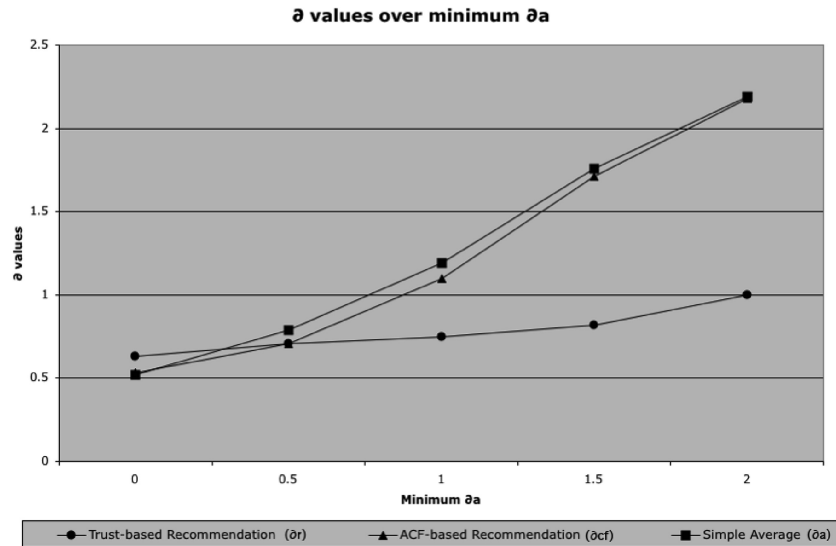


Fig. 4. Error rates from different predictive recommendation methods in FilmTrust. Using the simple average, a collaborative filtering algorithm, and a trust-based method all perform equally overall. However, as the threshold for the minimum difference between the user's rating and the average increases, the trust algorithm significantly outperforms both other methods.

matches the user's low rating of 0.5 stars. These are precisely the type of cases that the recommended rating is designed to address.

3.4 Motivation

These results suggest that when users assign trust ratings, they are capturing more than simply how similar they are to others. While we tested cases where the user is different from the average, our experiments actually captured *two* features at once. As the user gets further from the average, in most cases their ratings also enter the extreme ends of the spectrum. Thus, it is possible that the trust-based predictions work better when users agree on movies where their opinions are very high or very low. In either case, if trust were only a reflection of overall similarity, we would expect closer alignment of the performance from ACF algorithms and trust algorithms. What precisely is expressed in trust beyond overall similarity? That question motivates the experimental work in the rest of this article.

4. THE EXPERIMENT

In order to understand how similarity on different profile attributes affect the trust people assign to one another, we conducted a survey-based experiment. Subjects began by rating a large number of movies. We then generated profiles of hypothetical users that included ten to twenty films. In a profile, the subject was shown the hypothetical user's rating for each film, along with the subject's own rating and the average rating. Subjects were asked to say how much they would trust this hypothetical user based on the profile. In the

profiles we varied when and how much the hypothetical person agreed with the user.

4.1 Subject Background

Fifty-nine subjects participated in this study. They ranged in age from 20 to 52, with an average age of 32 (standard deviation of 8.5 years). Education levels were mixed among subjects who reported it: six subjects were current undergraduates, eleven had bachelor's degrees, twenty-three had master's degrees, and eleven had Ph.Ds.

On average, subjects reported watching movies about once a week, and looking at movie related media and Web sites every week or two.

Further details about subjects' movie rating patterns are contained in Section 4.2.2.

4.2 Stage One: Learning Subjects' Opinions

4.2.1 *Compiling a List of Movies.* Subjects began the study by rating a large collection of movies. In choosing movies to include as part of this study, there were several factors we considered important to create a fair coverage:

- (1) Movies most subjects would have seen—We wanted our list of movies to contain films that had reached a wide audience, so that subjects could rate a large number of films. To ensure that many popular movies were in the mix, we included the 100 worldwide top grossing films of all time.²
- (2) Cover a broad spectrum of genres—For subjects who had interests mostly in specific genres, we took the top 10 rated movies from each genre as listed in the Internet Movie Database (IMDB): Action, Adventure, Animation, Family, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Independent, Musical, Mystery, Romance, Science Fiction, Thriller, War, and Western.
- (3) Include bad movies—The nature of the first two considerations necessarily means that our movies would be skewed toward the good end. To be sure that plenty of bad films were on the list, we looked at the IMDB 100 worst rated movies. We took all the films on that list that had at least 1,000 ratings to ensure that these were bad movies that our subjects may have seen. Once the movies with under 1,000 ratings were removed, 65 bad movies were left to be included in our set.

After making these selections, some films were duplicated (e.g., *Lord of the Rings: Return of the King* was one of the top grossing movies as well as appearing in the top ten lists of the action, adventure, and fantasy genres). With duplicates removed, we were left with a total of 283 films. This made up the core set of movies that we used in the study.

In stage one of the experiment, we presented users with a list of these 283 films and asked them to assign a rating to each movie that they had seen and about which they had an opinion. Ratings were on a 1–10 scale (1 for bad movies,

²<http://imdb.com/boxoffice/alltimegross?region=world-wide>.

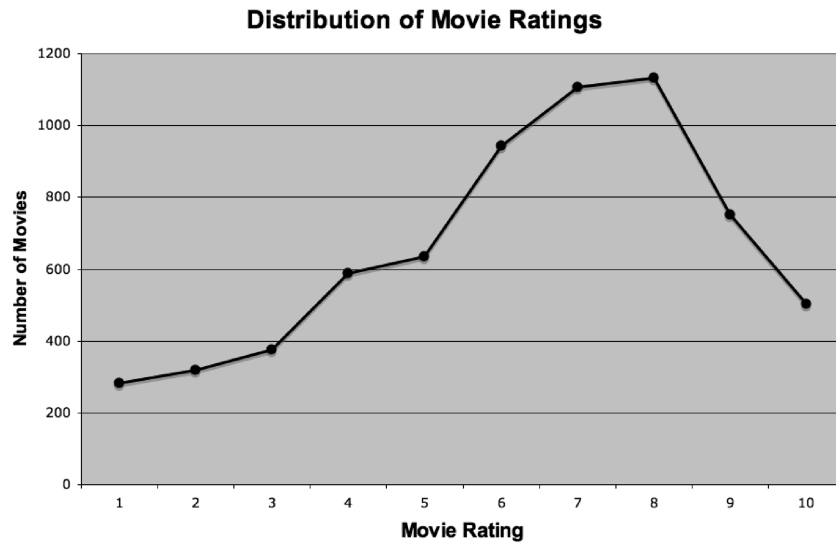


Fig. 5. Distribution of movie ratings assigned by subjects.

and 10 for good ones). This rating system matched the Internet Movie Database scale, which allowed us to use the IMDB average ratings in our analysis.

4.2.2 Movie Rating Statistics. Subjects rated an average of 136.4 movies (standard deviation 51.3). Ratings for movies were normally distributed around a mean of 6.3 (standard deviation 2.4) on a scale of 1–10. This was slightly skewed to the high end because many of our movies were pulled from top-rated lists. The distribution of ratings is shown in in Figure 5.

Most subjects ratings were distributed around the mean, however there were some outliers. For example, one subject assigned a rating of 1 to 81 of the 183 movies he or she rated while the average subject (excluding this outlier) only rated 3.3 movies with a 1. This subject alone was responsible for over 25% of all the ratings of 1.

Note that the selection of movies in this sample is not representative of the movies that would be found in a live working system. They were chosen to ensure that subjects had many of movies to rate, that the movies were diverse in genre and age, and that they were likely to receive high and low ratings. While the movie rating distribution is not necessarily reflective of what a user would assign in a real system, it is useful in conducting these controlled experiments. We show later in section 6 that the results obtained here with respect to trust can be integrated successfully into a real system.

4.3 Stage 2: Assigning Trust to Profiles

To measure how trust correlated to similarity on certain profile features, we generated profiles representing the preferences of hypothetical users and asked subjects to rate how much they trusted those people.

4.3.1 Generating Profiles. After rating the films, users were presented with a series of profiles. This work was conducted in two phases. These were designed to test different conditions.

4.3.1.1 Phase 1. In the first phase, subjects were presented with 28 profiles (P_1 – P_{28}). Each profile contained ten to sixteen films chosen from the set of movies the subject rated. For each film in the profile, subjects were shown the average rating for each film, their own rating, and a rating generated to represent the opinion of a hypothetical user.

Call the subject's rating of movie i r_{si} , the profile's rating of the same movie r_{pi} , and the average rating r_{ai} . Call the absolute difference between any two ratings Δ . There are three categories (C_1 – C_3) where we initially expected to see a correspondence of trust and specific profile features.

- C_1 : Extreme Ratings ($r_{si} \leq 2$ or $r_{si} \geq 9$). If the subject rated a movie 1, 2, 9, or 10, we considered the rating to be on the extreme ends of the scale.
- C_2 : Far from average ($\Delta_{r_{si}, r_{ai}} \geq 4$). Among our subjects, the average difference between the subject's rating and the average was 1.82, with a standard deviation of 1.78. Since the data was normally distributed, approximately 95% of the ratings were within ± 4 of the average. Thus, we looked for a difference ≥ 4 to qualify a rating as largely different from the average.
- C_3 : Extreme ratings that are far from the average ($\Delta_{r_{si}, r_{ai}} \geq 4$ and either $r_{si} \leq 3$ or $r_{si} \geq 9$). This occurs when the difference between the subject's rating and the average is ≥ 4 and the subject's rating is 1, 2, 9, or 10.

Each profile, P_j , was generated by choosing a set of movies, M_k from the movies rated by the subject. Four to six movies came from a category C_m and six to ten movies from its complement. For example, a profile could comprise four movies where the subject had given extreme ratings, and six movies where the subject's ratings were not extreme. Once the movies were selected, we generated the profile rating for each film.

We tested for small, medium, and large differences between the subject's rating and the profile's rating by randomly generating a Δ in the correct range:

- (1) small variations: $0 \leq \Delta_{r_{si}, r_{pi}} \leq 1.5$,
- (2) medium variations: $0 \leq \Delta_{r_{si}, r_{pi}} \leq 6.5$,
- (3) large variations: $0 \leq \Delta_{r_{si}, r_{pi}} \leq 9.5$.

For profile P_j , r_{pi} is created for each of the four to six movies selected from the category C_m , and for each of the six to ten movies chosen from its complement, $\overline{C_m}$. For example, profile P_3 was generated with small differences on movies with extreme ratings (C_1), and large differences on movies with nonextreme ratings ($\overline{C_1}$). The number of profiles with small (s), medium (m), and large (l) differences among the categories is shown in Table I.

Subjects were presented with a total of 28 profiles and asked to rate how much they trusted the person represented by the profile. Trust was assigned on a scale of 1 (low trust) to 10 (high trust). There was also space to comment on any specific reasons for the rating.

Stage 2: Rate Profiles

http://trust.mindswap.org/cgi-bin/FilmTrust/study/stage

Stage 2: Profile 20

(of 28)

Person 20

How much do you trust Person 20 's opinions about movies? [Continue with Stage 2-->](#)

If there are any specific reasons you gave this person the rating you did, please enter them here:

Movie Name	Average Rating	Your Rating	Person 20 's Rating	Difference Between Your Rating and Person 20 's Rating
Spice World (1997)	3	1	8	7
Gigli (2003)	2.3	1	6	5
The Mummy (1999)	6.6	6	5	1
Shrek (2001)	8	7	7	0
E.T. the Extra-Terrestrial (1982)	7.8	6	6	0
Cast Away (2000)	7.3	7	7	0
Top Gun (1986)	6.5	6	7	1
Jaws (1975)	8.2	10	1	9
Psycho (1960)	8.6	10	3	7
The Shawshank Redemption (1994)	9.1	7	7	0

Done [Adblock](#)

Fig. 6. A sample profile from the experiment.

4.3.1.2 Phase 2. Just as in Phase 1, the second phase of the study was conducted by presenting profiles with generated ratings to the subjects. Subjects were asked to rate how much they trusted the hypothetical user represented by the profile in the way way. However, instead of generating differences between the subject's rating and the profile rating that fell within a range, we used fixed sets of rating differences and chose to distribute them in different ways.

The variation among profiles in the first phase was useful because it provided data on many different features of profile similarity. However, a more controlled experiment was required to test specific hypotheses. In phase two, each profile consisted of ten movies. Exactly four were chosen from one of the categories previously listed, and six movies were chosen from its complement.

To control for overall difference, standard deviation, and other factors independent of the movies, we established three predefined sets of rating differences ($\Psi_1 - \Psi_3$) (see Table II). Each set contained ten values, $\psi_1 - \psi_{10}$. The Ψ_n values were applied to the movie ratings in each profile, such that for movie j and Ψ_n , $\Delta_{r_{sj}, r_{pj}} = \Psi_n$. This ensured that the average rating, the variance, the distribution of ratings, and the largest $\Delta_{r_{si}, r_{pi}}$ were the same among all profiles using a given Ψ . Thus, for a fixed Ψ , the only variation among profiles is which movies the Ψ_n are applied to, and this allows us to test if variation on certain categories of films is more important than others. An example profile generated using Ψ_3 , where the first four movies have extreme ratings from the user, is shown in Figure 7.

Table I.

This table shows the number of profiles presented to each subject in each category with the shown small, medium, and large differences. For example, (s,l) on $(C_1, \overline{C_1})$ would indicate a small difference between the profile and user ratings on movies from category C_1 , and a large difference between the profile and user ratings on movies in from the complement of category C_1 . (Note: No (s,s) profile was used for C_3 because the small differences were addressed in C_1 and C_2 , and C_3 is just the intersection of C_1 and C_2 .)

	$(C_1, \overline{C_1})$	$(C_2, \overline{C_2})$	$(C_3, \overline{C_3})$
(s,s)	2	2	0
(s,m)	4	4	1
(s,l)	2	2	1
(m,s)	2	2	1
(l,s)	2	2	1

Table II.

For each predefined set of rating differences (Ψ_i), ten individual differences (ψ_j) are enumerated

	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5	ψ_6	ψ_7	ψ_8	ψ_9	ψ_{10}
Ψ_1	1	1	1	1	0	0	0	0	0	0
Ψ_2	4	4	4	4	1	0	1	1	0	0
Ψ_3	6	6	6	6	1	0	1	1	0	0

Once the films were selected, the user's rating for each movie was paired with a ψ_j to produce the profile rating.

Let $\psi_{high} = \{\psi_1, \psi_2, \psi_3, \psi_4\}$ and $\psi_{low} = \{\psi_7, \psi_8, \psi_9, \psi_{10}\}$. For each $(C_m, \overline{C_m})$ we selected a set of ten movies, M_k . Four movies (m_1-m_4) were chosen from C_m and six (m_5-m_{10}) from $\overline{C_m}$. Profile ratings were generated by pairing m_1-m_4 with either ψ_{low} or ψ_{high} and m_5-m_{10} with the remaining ψ values. For each $(C, \Psi, \{high, low\})$ combination, subjects were shown three profiles, for a total of 54 profiles to be rated.

4.3.2 Profile Rating Statistics. Figures 8 and 9 show the overall distribution of trust ratings assigned to profiles in Phases 1 and 2 of the study. In Phase 1, the ratings have a normal distribution, skewed to the high end with an average of 6.90. This distribution and mean are expected based on the way profiles were generated; the Δ for each movie was randomly generated within a given range, and more profiles had small differences than big ones, leading to the high average trust rating. In Phase 2, the distribution of ratings is less obviously normal, which is explained by the use of the Ψ distributions to create profiles. The average trust rating was also lower in Phase 2, at 5.95.

Stage 2: Rate Profiles

Stage 2: Profile 32

(of at most 54)

Person 32

How much do you trust Person 32 's opinions about movies? Continue with Stage 2-->

If there are any specific reasons you gave this person the rating you did, please enter them here:

Movie Name	Average Rating	Your Rating	Person 32 's Rating	Difference Between Your Rating and Person 32 's Rating
The Usual Suspects (1995)	8.7	10	4	6
Titanic (1997)	6.9	2	8	6
Catch Me If You Can (2002)	7.6	9	3	6
Twister (1996)	5.9	5	5	0
Back to the Future (1985)	8.1	6	7	1
Grease (1978)	6.8	7	7	0
Armageddon (1998)	5.7	6	6	0
Gigli (2003)	2.3	1	7	6
Toy Story 2 (1999)	8.1	7	8	1
The Maltese Falcon (1941)	8.4	8	9	1

Fig. 7. A sample profile from the experiment where differences were generated using the differences in Ψ_3 .

5. EXPERIMENTAL RESULTS

In Phase 1, the profiles had a wide range of characteristics. We measured agreement on a profile P_j by the average $\Delta_{r_{si}, r_{pi}}$ for all movies in the profile. Call the average difference on any set of movies Θ , and let the overall average difference on all movies in P_j be $\bar{\Theta}_j$.

We are interested in identifying which features of profiles, if any, correlate more strongly with trust than overall agreement. For example, does agreement on movies where the user gave an extreme rating matter more than good overall agreement? Does the user's difference from average ratings matter more? In this section we analyze the data to identify how these features correlate with trust.

5.1 Correlation of Trust and Overall User Similarity

As mentioned in the previous work, Ziegler and Golbeck [2006] established a relationship between trust and user similarity. We began by testing the correlation of trust with overall similarity to show that same result on our data. Correlation was measured by computing the Pearson correlation coefficient. We first checked the correlation of trust assigned to a profile P_j (call this t_j), and $\bar{\Theta}_j$. That correlation was -0.65 , indicating a strong negative correlation; that is, as the average absolute difference increases, trust decreases.

To gain more insight into the correlation of trust and overall similarity, we looked at intervals of average difference. We took all of the profiles with

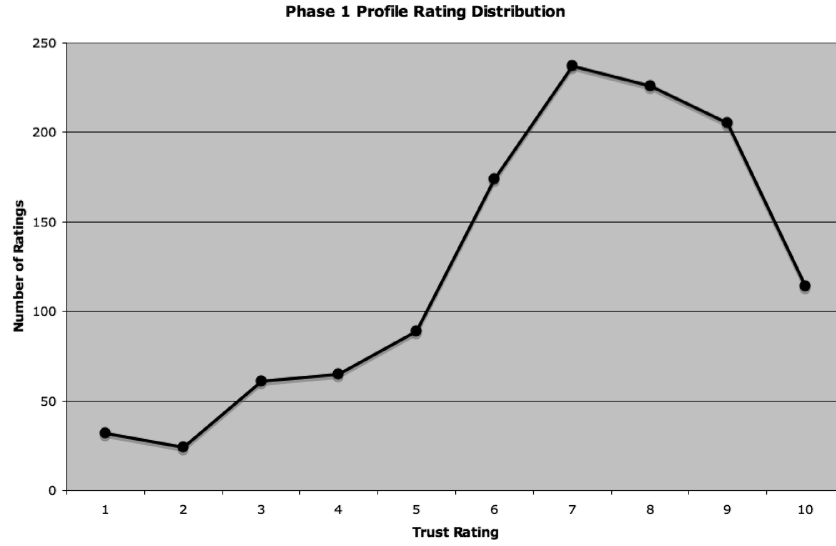


Fig. 8. Distribution of profile trust ratings assigned by subjects, Phase 1.



Fig. 9. Distribution of profile trust ratings assigned by subjects, Phase 2.

$0 \leq \bar{\Theta} < 1$ and grouped them together. We did the same for $1 \leq \bar{\Theta} < 2$, $2 \leq \bar{\Theta} < 3$, etc. Call these groupings $\bar{\Theta}^{[a,b]}$. With this grouping, we analyzed the trust value assigned to each profile in the group, and thus we could track how trust changed as $\bar{\Theta}$ varied. Table III shows the average trust value for each range of $\bar{\Theta}$ values. An ANOVA test on this data showed that the trust values were significantly different based on $\bar{\Theta}$ for $p < 0.001$. Using a standard 2-tailed t-test, we compared the trust values for each consecutive range of $\bar{\Theta}$ values (e.g.,

Table III.

Average trust values for ranges of average difference ($\bar{\Theta}$). Note that there were no profiles with $\bar{\Theta} \geq 8$. That would require all of the users ratings to be very high or very low to allow profile ratings to have such a large difference on the majority of movies. For similar reasons, there were less than ten movies in the range [6,7) and [7,8).

$\bar{\Theta}$ Range	Average Trust Rating
[0,1)	8.01
[1,2)	6.44
[2,3)	5.27
[3,4)	3.96
[4,5)	2.72
[5,6)	1.94
[6,7)	2.09
[7,8)	2.25

we compared the trust values for $\bar{\Theta}^{[0,1)}$ to those for $\bar{\Theta}^{[1,2)}$, $\bar{\Theta}^{[1,2)}$ to $\bar{\Theta}^{[2,3)}$, and so on). Our results showed a significant difference between each consecutive pair from $\bar{\Theta}^{[0,1)}$ to $\bar{\Theta}^{[5,6)}$ with $p < 0.05$. Beginning with the comparisons of $\bar{\Theta}^{[5,6)}$ to $\bar{\Theta}^{[6,7)}$, the differences were not statistically significant because there were very few datapoints. There were no profiles with $\bar{\Theta} > 8$, and only four profiles with $\bar{\Theta}^{[7,8)}$. To reach such high overall average differences, the user would need to have most of their ratings either very high or very low to allow for the possibility of such a big difference with most of the profile ratings.

These results support the previous research that trust correlates with user similarity. We see significant increases in trust values as the overall agreement increases.

5.2 Phase 1 Results: The Maximum Difference Effect

With the data in Phase 1, we began by looking at agreement on movies where the user had assigned extreme ratings ($r_{si} \leq 2$ or $r_{si} \geq 9$). It could be that extreme ratings indicate that the user is more passionate about the movie, and thus it would matter more to the user that the profile agrees on these films over those where the rating is in the middle range. Our first experiment to check this was to hold overall average difference ($\bar{\Theta}$) constant and then check the correlation between trust and agreement on extremes. In this case, if agreement on extremes increases, then agreement on nonextremes ($3 \leq r_{si} \leq 8$) must decrease.

When holding the overall average difference constant, as agreement on extremes increased, trust increased. This initially would indicate that trust is tied, in part, to agreement on extremes. However, when we held the overall average difference constant and increased agreement on *nonextremes* (and thus agreement on extremes decreased), trust *also increased*. These results indicated that

some factor other than extremes or nonextremes was involved and was affecting the trust values.

Consider what happens when the average difference between the user's ratings and the profile's ratings increases. There are two possibilities—the number of movies where there is disagreement must increase, or the size of the $\Delta_{r_{si}, r_{pi}}$ must increase. If the overall average difference is kept constant, and agreement increases in one category, that means agreement must *decrease* over the rest of the movies. Thus, on the set of movies where agreement decreases, the standard deviation among those movies may increase, or the differences between the user and profile ratings on movies will grow larger. It is possible that these changes affect trust more strongly than extreme or nonextreme values. To test how these factors correlate to trust, there are two tests we conducted: the correlation of trust with the standard deviation (σ) on a profile, and the correlation of trust with the maximum $\Delta_{r_{si}, r_{pi}}$ on a given profile. Call this single maximum $\Delta_{r_{si}, r_{pi}} \nabla$.

As the single maximum difference increases, standard deviation will also increase. To see the effect of the single maximum difference independent of the standard deviation, we looked at the correlation of the single maximum difference and trust with standard deviation and the overall average difference held in fixed ranges. We used integer ranges for standard deviation and the overall average difference (e.g., [0,1), [1,2). This created 100 smaller datasets where we could examine the effect of the single maximum difference on trust (e.g., one of these data sets is σ in the range (0,1] and $\bar{\Theta}$ in the range (2,3]). Within each $(\sigma, \bar{\Theta})$ pair, trust values were grouped by ∇ in integer ranges ($0 \leq \nabla < 1, 1 \leq \nabla < 2, \dots$). If ∇ is a factor in how users assign trust, we expect to see significant changes in trust when ∇ increases while σ and $\bar{\Theta}$ are held constant.

There were approximately 1,200 trust values assigned in Phase 1 of the study. Because our data was broken up into 100 smaller datasets, which in turn divide the data into ten groups, very few $(\sigma, \bar{\Theta})$ pairs had enough data to allow for a statistical analysis. We expect to have very little data in the higher ranges for σ and $\bar{\Theta}$, because profiles with such large differences on every movie were not considered. Similarly, with a low $\bar{\Theta}$, it is not mathematically possible to have high values of σ . In fact, approximately 80% of our data came from profiles where $\sigma < 2$ and $\bar{\Theta} < 2$.

We were able to perform a statistical analysis of the single maximum difference - trust relationship on two $(\sigma, \bar{\Theta})$ pairs:

- (1) $(\sigma, \bar{\Theta}) = ([0,1), [0,1))$: 416 trust ratings, about 35% of the total number of trust ratings made in Phase 1.
- (2) $(\sigma, \bar{\Theta}) = ([1,2), [1,2))$: 355 trust ratings, about 30% of the total number of trust ratings made in Phase 1.

In both of these cases, which together covered approximately 65% of all the trust ratings given in Phase 1, we see a very clear relationship between the single largest difference and trust.

Table IV shows the average trust values as the single maximum differences increases. In both cases, we can see the average trust rating decreases

Table IV.
For the two $(\sigma, \bar{\Theta})$ pairs that allowed a statistical analysis, this table shows the ranges of ∇ and the corresponding average trust value

$(\sigma, \bar{\Theta}) = ([0,1], [0,1])$				
∇ Value	0-1	1-2	2-3	3-4
Average Trust Value	9.11	8.58	8.16	7.52

$(\sigma, \bar{\Theta}) = ([1,2], [1,2])$				
∇ Value	2-3	3-4	4-5	5-6
Average Trust Value	7.30	6.61	6.22	5.49

as the single maximum differences increases. An ANOVA showed that there were significant differences within the population for $p < 0.01$ when $(\sigma, \bar{\Theta}) = ([0, 1], [0, 1])$ and also when $(\sigma, \bar{\Theta}) = ([1, 2], [1, 2])$. Pairwise comparisons using a 2-tailed t-test were made for each population $i \leq \nabla < i + 1$ and $i + 1 \leq \nabla < i + 2$. For the populations shown in Table IV, each pairwise comparison was statistically significant for $p < 0.05$. This result suggests that the maximum difference between the user's rating and profile's rating of a given movie, is a significant factor in how users assign trust. Specifically, for the movie with the largest difference between the user's rating and the profile's rating, as the difference grows, trust decreases. This is true even with overall agreement and standard deviation kept constant.

5.3 Phase 2 Results: Extreme Ratings

In Phase 1, we could not see if extreme ratings and trust had a relationship because the profiles had other factors that were impacting trust whereas, in Phase 2, the profiles were much more controlled. Profiles used one of three sets of rating differences (Ψ_n) , and the ψ values were distributed among a set of ten movies, selected for certain features. For example, if a profile were using Ψ_2 and the ψ_{high} values were being assigned to movies the user gave extreme ratings, a sample profile might look like what is shown in Table V.

Thus, for a given Ψ_n , the only variable was to which movies the ψ values were assigned. $\bar{\Theta}$, ∇ , σ , and other factors were all constant.

Our first hypothesis was that when the larger ψ values in Ψ_n were applied to movies where the user gave an extreme rating (C_1), that users would have lower trust than when the ψ_{high} values were used on \bar{C}_1 movies.

For each Ψ_n , we compared the average trust rating assigned to a profile when ψ_{high} were used on movies with extreme ratings vs. when ψ_{high} were applied to movies with nonextreme ratings. For Ψ_1 , there was no statistically significant difference. We believe this is because users consider a maximum difference of 1 on a 1–10 scale to be trivial. For Ψ_2 and Ψ_3 , the average trust ratings were significantly lower ($p < 0.05$) for profiles where movies with extreme ratings had the large differences.

Table VI shows the average trust values assigned to profiles for each Ψ_n based on which films received the ψ_{high} . When ψ_{high} were assigned to movies

Table V.
A sample Phase 2 profile using Ψ_2 with ψ_{high} on
movies with extreme ratings

Movie	User Rating	Profile Rating	ψ
Movie 1	9	5	4
Movie 2	10	6	4
Movie 3	1	5	4
Movie 4	2	6	4
Movie 5	6	7	1
Movie 6	4	4	0
Movie 7	8	7	1
Movie 8	3	2	1
Movie 9	5	5	0
Movie 10	7	7	0

Table VI.
Average trust ratings in phase 2 for
each Ψ_n with ψ_{high} distributed to
Extremes or Nonextremes

	ψ_{high} Assigned To	
	Extremes	Non-Extremes
Ψ_1	8.62	8.72
Ψ_2	5.23	5.83
Ψ_3	2.65	3.34

with extreme ratings, the average trust value for the profile averaged over 0.5 lower than when the movies with extreme ratings used ψ_{low} .

From this, we are able to conclude that agreement on movies to which the subject has assigned extreme ratings (Θ_χ) affects trust, independently of overall agreement. When agreement is closer on these movies, trust is higher.

5.4 Phase 2: Trust Predicting Trust

As discussed in the previous work, there are factors specific to each person and each relationship that affect trust independently of similarity. While most of that information is not available in the types of systems we are studying, we *do* have the user's own trust ratings, which provide some insight.

In a real system, like FilmTrust, users may make choices about who their friends are that impact the trust ratings they assign. For example, someone who only connects to and rates their closest friends may assign trust values with higher average than someone who rates a large group of acquaintances. Because of unknowable factors like this, it is difficult to analyze a real system and determine why some users assign, on average, higher trust values than others.

In Phase 2 of our experiment, though, all the subjects were rating the same population of hypothetical users. The profiles were carefully controlled, so the profiles rated by each subject all varied from the user in the same way. That is, Profile i in Phase 2 contained ten movies selected on the same criteria for all users (e.g., four movies where the user gave extreme ratings and six where

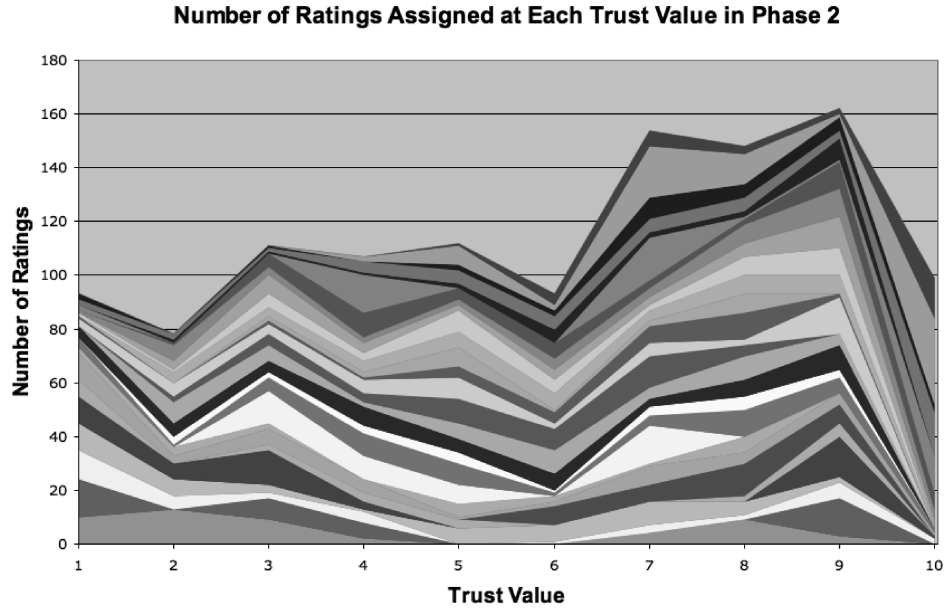


Fig. 10. Distribution of trust ratings assigned by 29 subjects in Phase 2.

the user gave nonextreme ratings) where the profile's rating differed from the user's rating by the same amount on each type of film. Thus, we can look the average trust value assigned by user s (call this τ_s) to estimate how "trusting" the user is.

Figure 10 shows the ratings assigned by 29 subjects in Phase 2. Each colored layer represents a single user, and the width of the layer indicates how many trust ratings that user assigned at the value indicated on the x-axis. The distribution of ratings varies greatly from user to user. For individuals, the average trust rating assigned by subjects to profiles in Phase 2 ranged from as low as 3.73 up to 8.48.

For example, the lavender layer second from the top indicates a very trusting subject. This user assigned 16 trust ratings at a level 10, one at level 9, six at level 8, and nine at level 7, and only five trust ratings lower than that. On the other hand, the bottom user represented by a blue layer, assigned much lower ratings. This subject only assigned six values above a level 4, with the remaining values all below that.

All subjects saw profiles that differed from them in the same way, so this range of τ among users indicates that some subjects are inclined to assign higher trust values than others. It could be that some subjects are more trusting than others, or that subjects interpret the scale of trust values differently. Whatever the reason, the average trust value has a wide range.

The τ for a given user informs what trust values we expect them to assign in the future. If a user is consistently giving low trust ratings we can anticipate that their future ratings will also be lower. Since we are only considering the

trust ratings the user has assigned, there are no issues of privacy or access control to worry about; the information is simply part of the user's profile. We include a τ in our analysis in Section 6.

5.5 Difference from Average

In both Phase 1 and Phase 2 of the experiments, we tested to see if agreement and disagreement in C_2 , where the user's rating was far from the average rating, affected trust. When considering this, we do not want to conflate these results with those of extreme ratings. Overall, 28% of the movies in the system were given extreme ratings (rated 1, 2, 9, or 10). In the set of movies where the user's rating was far from the average (a difference of at least 4), 50.4% of the movies had extreme ratings. Since the proportion of extreme ratings was significantly higher in this group, it would be possible for the impact of extreme ratings to make it appear as though difference from average was also significant. Thus, we controlled for extreme ratings in our analysis.

With the data from Phase 1, we tested to see if large differences on movies in C_2 caused trust to change significantly from when there were only small differences on movies in C_2 . We were unable to find any statistically significant results overall, nor when we eliminated movies with extreme ratings from the pool. We performed the same analysis in Phase 2. Again, overall and when we controlled for extreme ratings, there was no significant impact visible in our results.

C_3 considered both extreme ratings *and* difference from average. An analysis here did not indicate any results beyond what we saw in C_1 when looking at extreme ratings only. Since C_2 did not indicate significant differences in trust values, this is not surprising.

It was a somewhat surprising result that difference from average did not affect trust, since the initial results of the FilmTrust study described in section 3 were seen when the user's opinion grew apart from the average. While that analysis was checking the accuracy of recommended ratings, not the correlation of trust ratings, we expected to see a parallel result in this study. However, the earlier results included movies with extreme ratings in with the movies that were far from average. Thus, it is likely that we were seeing a manifestation of the extreme rating phenomenon in those results, rather than finding a point of conflict.

5.6 Experimental Conclusions

From the survey experiments conducted here, our results suggest that overall difference ($\bar{\Theta}$), the maximum single difference in movie ratings (∇), difference on movies to which the user assigned extreme ratings (Θ_x), and the user's average trust value (τ) all are factors in how users assign trust. However, it is difficult from these experiments to know how strongly each of those features affect trust. In the next section, we use these parameters in FilmTrust and show how they can be composed to predict trust more accurately and precisely than using overall agreement alone.

6. PREDICTING TRUST WITH PROFILE SIMILARITY

The study results are controlled, and many aspects of trust cannot be accounted for. In a real system, social features independent of similarity will certainly affect the trust ratings people assign to one another. These social features are, in fact, the more studied aspects of trust. A person's propensity to trust in general, the nature of the relationship with the person being trusted, their history outside of the system, and many other factors will impact the trust rating that is assigned, and those factors are largely independent of any profile similarity measures. The question we face is whether or not the aspects of profile similarity that correlate with trust can be used to predict trust in a real system.

To test this, we looked at data taken from the FilmTrust system. When two people are connected in the FilmTrust social network, call the person doing the trusting the *source* and the person being trusted the *sink*. For each ordered pair of users, we collected the trust value, the average trust value assigned by the source (τ), number of movies in common, overall error ($\bar{\Theta}$), overall error on extremes (Θ_χ), and the maximum difference (∇). Our goal was to see if a composition of the values could yield a relatively accurate estimate of trust. To ensure we had enough data to work with, we eliminated any pairs with less than 10 movies in common. That left us with 366 user-pairs to analyze.

It is important to note that, unlike the controlled environment of our study where certain variables could be held constant, these measures of error are all interrelated. The Pearson correlation coefficient for $\bar{\Theta}$ and ∇ is 0.42. $\bar{\Theta}$ and Θ_χ also had a correlation coefficient of 0.42. Θ_χ and ∇ had a high correlation of 0.70. For some pairs, Θ_χ did not exist because the sink had not rated any movies given an extreme rating by the source.

Using these $\bar{\Theta}$, Θ_χ , and ∇ values, along with the τ for the source to adjust our estimate up or down, we were able to predict trust rather accurately. Equation (3) gives the equation that uses a weighted average of each factor listed above to predict trust between user i and j (t_{ij}).

$$t_{ij} = \min \left(10, \begin{array}{ll} [w_{\bar{\Theta},1}\bar{\Theta}_{ij} + w_{\nabla,1}\nabla_{ij} + w_{\Theta_\chi}\Theta_{\chi,ij} + (\tau_i - \tau)] & \text{if } \Theta_\chi \text{ exists} \\ [w_{\bar{\Theta},2}\bar{\Theta}_{ij} + w_{\nabla,2}\nabla_{ij} + (\tau_i - \tau)] & \text{otherwise} \end{array} \right). \quad (3)$$

We estimated weights based on the impact each factor had in the initial study, with $(w_{\bar{\Theta},1}, w_{\nabla,1}, w_{\Theta_\chi}, w_{\bar{\Theta},2}, w_{\nabla,2})$ set to $(0.7, 0.2, 0.1, 0.8, 0.2)$. The optimal values for these weights will vary from system to system. They can be computed from the real data produced by users through techniques such as regression analysis or learning algorithms.

We computed the accuracy of trust values predicted by formula 3 by taking the mean absolute error (MAE) of the computed trust value and the known trust rating. This was compared to the accuracy of trust predicted using only overall agreement. As is shown in Table VII, our trust estimation using the set of profile similarity measures has an MAE of 1.22, a 36% increase in accuracy over trust prediction using overall similarity alone. A two-tailed t-test shows that these results are statistically significant for $p < 0.001$. The correlation was also quite high, with a Pearson Correlation Coefficient of 0.68 using this formula vs. 0.24 using overall agreement alone.

Table VII.
Statistics relating actual trust values and predicted trust values derived from
overall agreement and Equation (3)

	Overall	Equation 3
Correlation With Actual Trust Value	0.24	0.73
MAE of Computed and Known Trust Value	1.91	1.22
Standard Deviation of Average Difference	1.95	0.95

While Equation (3) and the values used in it are fine-tuned to the FilmTrust system, the large increase in accuracy afforded by incorporating these profile similarity measures is a good indicator that other systems can incorporate the measures to estimate the trust users have in one another. The weights in the formula may change, but these results from FilmTrust show that parameters derived from the survey can be used to predict trust in a working system.

These results from FilmTrust are a strong reinforcement to the results obtained in Section 5. Using overall error ($\bar{\Theta}$), overall error on extremes (Θ_χ), and the maximum difference (∇) along with the user's average trust rating (τ), we significantly improved the correlation, accuracy, and standard deviation of trust predictions from profile information. This supports our previous analysis that $\bar{\Theta}$, Θ_χ , ∇ all impact trust, since our optimized formulation included nonzero weights for each parameter. Furthermore, because these parameters are effective in a real system, it shows that our survey results are not merely the product of a tightly controlled experiment, but rather are more generally applicable set of principles regarding trust and profile similarity.

The greater impact of this analysis is that we can use these profile similarity measures to compute trust when there are no trust values in the system. When users are new, or in social networks that do not support trust ratings, the nuanced profile similarity measures developed in Section 5 can be combined to estimate trust. In systems where there are no trust values to use for normalization with the τ factor, the trust values are, in a sense, normalized by default; trust for everyone is computed in the same way, and individual propensities to trust or not trust will not be a factor. While this may lead to estimated trust values that are further from the trust value a user would assign on his or her own, a system with no explicit trust rating system would likely hide these computed values anyway, and simply use the values internally for functionality in the system. Thus, whether the social network does or does not support explicit trust ratings, this method of computing trust based on similarity can provide a reasonable estimate of trust to be used in recommender systems or other applications.

7. CONCLUSION

Up to now, the research has only shown a correlation between trust and overall similarity. Through the surveys conducted here, we have shown that in addition to overall similarity, there is also a correlation between trust and the largest single difference in ratings, and between trust and the agreement on movies the source has given extreme ratings. We have also shown that some sources tend to assign higher ratings than others when rating a population of sinks that vary from the source in the same way.

We then showed in the FilmTrust system that a composition of these measures can be used to predict trust with a stronger correlation, better accuracy, and less variation than when using overall agreement alone. This shows that the insights obtained in our study can be implemented in a real system to improve the accuracy of trust inferences.

8. DISCUSSION

When working in a system that has trust and allows users to rate items, there are two methods for determining how much a source should trust a sink: predicting trust using the social network, and predicting trust using profile similarity. Using trust values in a social network, trust can be computed by gathering information from friends, who they trust, and how much (as discussed in Section 2.4). As has been shown in previous work, this can be effective when the users are connected into a large social network where trust values are accessible. However, our other work has shown that in many Web-based social networks, a significant percentage of users are completely isolated from most others, and social network-based techniques are not applicable [Golbeck 2006a].

To compute trust in these cases where there are insufficient social connections, we only have access to the source's trust ratings, the source's profile information, and the sink's profile information. From the experimental survey results and their application in the FilmTrust system, we have shown that trust correlates more strongly with a measure that includes several facets of profile similarity than with simple overall similarity.

These results are useful for several applications:

- (1) **Trust-Based Recommender Systems**—There have been several studies on using trust in Web-based social networks in recommender systems [Golbeck 2005; Ziegler 2005; Massa and Bhattacharjee 2004]. The benefits of using trust from social networks is that much of the computation or data storage can be distributed. A new method for computing trust when social networking connections are too sparse, or for users who do not have any social connections, would allow these same techniques to be applied in cases where they currently cannot. This type of trust derived from social networking and profile similarity would also be useful to other recommender systems techniques that consider trust, such as O'Donovan and Smyth [2005].
- (2) **Refining Trust Inference Algorithms**—In our literature search, we found no algorithms that use trust relationships and connections in a social network to infer trust *as well as* profile similarity. When profiles and social network information is available, it may be possible to improve the accuracy of the trust inference algorithms by using all of the available information.
- (3) **Trust Estimation for Intelligent Systems**—In this work, the only application of trust computed from a social network that we present is FilmTrust, a recommender system. There are other applications of trust. In other work we have discussed the use of trust for priorities in default logics [Katz and Golbeck 2006], the basis for a trust-based policy system. We are currently also looking at using trust in other policy systems and for belief revision. In these applications, we need an accurate estimate of trust. As discussed

before, when there is a strongly connected component, social network-based algorithms can be used to infer a trust rating. However, when a large number of users are not connected, another mechanism is needed to compute trust. We have shown that trust can successfully be estimated using the profile features discovered in our survey, and that can fill in the gaps left network-based trust algorithms.

9. FUTURE WORK

We plan to extend our work in this space by investigating how these results can be incorporated into algorithms and intelligent systems. We are in the early stages of work for creating policy systems using trust. In these systems there is a social network backing, but using policy sets as profiles may also allow an application of these results such that trust could be estimated even when the social network does not exist.

One of our ongoing projects is to investigate how this type of profile analysis can be integrated with the network-based trust algorithms to improve the accuracy of the results. We are also looking at whether these similarity measures can be used with trust ratings as the basis instead of movie ratings. While the structure of social networks suggests this would only be effective for users within a small distance of the user, if finding maximum differences, agreement on extremes, and overall agreement on trust ratings assigned to common friends can improve the accuracy of predicted trust, it would be a worthwhile modification of the algorithm.

There are many similarities theoretically and practically between the way trust is computed in social networks and the way predictive recommendations are generated by collaborative filtering algorithms. The similarity measures used in this article to predict trust are essentially a modified version of user-user similarity collaborative filtering algorithms. An interesting space of future work inspired by the research presented in this article involves the direct application of some of the new, more advanced model-based collaborative filtering algorithms to predicting trust.

ACKNOWLEDGMENTS

Thanks to Jim Hendler, Alon Motro, John Kleint, and all the students from MINDSWAP who provided insight and comments on this work in various stages. Thanks also to the reviewers; their comments made this work significantly better.

REFERENCES

- BARABASI, A.-L. AND ALBERT, R. 1999. Emergence of scaling in random networks. *Science* 286, 509.
- BASU, C., HIRSH, H., AND COHEN, W. 1998. Recommendation as classification: Using social and content-based information in recommendation. In *Proceedings of the 15th National/10th Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence (AAAI'98/IAAI'98)*. American Association for Artificial Intelligence, Menlo Park, CA, 714–720.

- BREESE, J. S., HECKERMAN, D., AND KADIE, C. 1998. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*. 43–52.
- BURGESS, E. AND WALLIN, P. 1943. Homogamy in social characteristics. *Amer. J. Sociol.* 2, 49, 109–124.
- BYRNE, D. 1961. Interpersonal attraction and attitude similarity. *J. Abnor. Soc. Psych.* 62, 713–715.
- BYRNE, D. 1971. *The Attraction Paradigm*. Academic Press, New York, NY.
- CASTELFRANCHI, C. AND FALCONE, R. 1998. Principles of trust for mas: Cognitive anatomy, social importance, and quantification. In *Proceedings of the 3rd International Conference on Multi Agent Systems*.
- CASTELFRANCHI, C. AND FALCONE, R. 2002. Social trust: A cognitive approach. In *Trust and Deception in Virtual Societies*. Cristiano Castelfranchi and Yao-Hua Tan, Eds., Kluwer Academic Publishers.
- CHIEN, Y.-H. AND GEORGE, E. 1999. A Bayesian model for collaborative filtering. In *Proceedings of the 7th International Workshop on Artificial Intelligence and Statistics*.
- DEUTSCH, M. 1962. Cooperation and trust. Some theoretical notes M. R. Jones, Ed. *Nebraska Symposium on Motivation*. Nebraska University Press.
- GARDEN, M. AND DUDEK, G. 2005. Semantic feedback for hybrid recommendations in recommendz. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'05)*. IEEE Computer Society, Los Alamitos, CA, 754–759.
- GEORGE, T. AND MERUGU, S. 2005. A scalable collaborative filtering framework based on co-clustering. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM'05)*. IEEE Computer Society, Los Alamitos, CA, 625–628.
- GOLBECK, J. 2005. Computing and applying trust in Web-based social networks. Ph.D. thesis, University of Maryland, College Park, MD.
- GOLBECK, J. 2006a. The dynamics of web-based social networks: Membership, relationships, and change. *ACM Trans. Inform. Stud.*
- GOLBECK, J. 2006b. Generating predictive movie recommendations from trust in social networks. In *Proceedings of the 4th International Conference on Trust Management*.
- GOLBECK, J. 2007. The dynamics of Web-based social networks: Membership, relationships, and change. *First Monday* 12, 11.
- GOLBECK, J. AND HENDLER, J. 2006. Filmtrust: Movie recommendations using trust in Web-based social networks. In *Proceedings of the IEEE Consumer Communications and Networking Conference*.
- GOLDBERG, D., NICHOLS, D., OKI, B. M., AND TERRY, D. 1992. Using collaborative filtering to weave an information tapestry. *Comm. ACM* 35, 12, 61–70.
- GOLEBIEWSKI, R. AND MCCONIKE, M. 1975. The centrality of interpersonal in group processes. In *Theories of Group Processes*. Cary Cooper, Ed. Wiley, Hoboken, NJ.
- HERLOCKER, J. L., KONSTAN, J. A., BORCHERS, A., AND RIEDL, J. 1999. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'99)*. ACM, New York, NY, 230–237.
- HERLOCKER, J. L., KONSTAN, J. A., AND RIEDL, J. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW'00)*. ACM, New York, NY, 241–250.
- HERLOCKER, J. L., KONSTAN, J. A., TERVEEN, L. G., AND RIEDL, J. T. 2004. Evaluating collaborative filtering recommender systems. *ACM Trans. Inform. Syst.* 22, 1, 5–53.
- HOFMANN, T. 2004. Latent semantic models for collaborative filtering. *ACM Trans. Inform. Syst.* 22, 1, 89–115.
- HUANG, Z., CHEN, H., AND ZENG, D. 2004. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Trans. Inform. Syst.* 22, 1, 116–142.
- JIN, R., CHAI, J. Y., AND SI, L. 2004. An automatic weighting scheme for collaborative filtering. In *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'04)*. ACM, New York, NY, 337–344.

- KAMVAR, S. D., SCHLOSSER, M. T., AND GARCIA-MOLINA, H. 2004. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International World Wide Web Conference*.
- KATZ, Y. AND GOLBECK, J. 2006. Social network-based trust in prioritized default logic. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI'06)*.
- KLEINBERG, J. 2000. The small-world phenomenon: An algorithm perspective. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC'00)*. ACM, New York, NY, 163–170.
- KONSTAN, J. A., MILLER, B. N., MALTZ, D., HERLOCKER, J. L., GORDON, L. R., AND RIEDL, J. 1997a. GroupLens: Applying collaborative filtering to usenet news. *Comm. ACM* 40, 3, 77–87.
- KONSTAN, J. A., MILLER, B. N., MALTZ, D., HERLOCKER, J. L., GORDON, L. R., AND RIEDL, J. 1997b. GroupLens: Applying collaborative filtering to usenet news. *Comm. ACM* 40, 3, 77–87.
- LEVIN, R. AND AIKEN, A. 1998. Attack resistant trust metrics for public key certification. In *Proceedings of the 7th USENIX Security Symposium*.
- MA, H., KING, I., AND LYU, M. R. 2007. Effective missing data prediction for collaborative filtering. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07)*. ACM, New York, NY, 39–46.
- MARSH, S. 1994. Formalising trust as a computational concept. Ph.D. thesis, Department of Mathematics and Computer Science, University of Stirling.
- MASSA, P. AND BHATTACHARJEE, B. 2004. Using trust in recommender systems: An experimental analysis. In *Proceedings of the 2nd International Conference on Trust Management*, C. Jensen, S. Poslad, and T. Dimitrakos, Eds. Lecture Notes in Computer Science, vol. 2995. Springer-Verlag.
- MILLER, B. N., ALBERT, I., LAM, S. K., KONSTAN, J. A., AND RIEDL, J. 2003. Movielens unplugged: Experiences with an occasionally connected recommender system. In *Proceedings of the 8th International Conference on Intelligent User Interfaces (IUI'03)*. ACM, New York, NY, 263–266.
- NEWCOMB, T. 1961. *The Acquaintance Process*. Holt, Rinehart, and Winston, New York, NY.
- O'DONOVAN, J. AND SMYTH, B. 2005. Trust in recommender systems. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*.
- PERNY, P. AND ZUCKER, J. D. 2001. Preference-based search and machine learning for collaborative filtering: The “film-conseil” recommender system. *Inform. Interac. Intell.* 1, 1, 9–48.
- PAGE, L., BRIN, S., MOTWANI, R., AND WINOGRAD, T. 1998. The pagerank citation ranking: Bringing order to the Web. Tech. rep. 1998, Stanford University.
- RICHARDSON, M., AGRAWAL, R., AND DOMINGOS, P. 2003. Trust management for the semantic web. In *Proceedings of the 2nd International Semantic Web Conference*.
- SARWAR, B., KARYPIS, G., KONSTAN, J., AND REIDL, J. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web (WWW'01)*. ACM, New York, NY, 285–295.
- SARWAR, B., KARYPIS, G., KONSTAN, J., AND RIEDL, J. 2000. Application of dimensionality reduction in recommender systems: A case study. In *Proceedings of the ACM WebKDD Workshop*.
- SINHA, R. AND SWEARINGEN, K. 2001. Comparing recommendations made by online systems and friends. In *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*.
- SREBRO, N. AND JAAKKOLA, T. 2003. Weighted low rank approximation. In *Proceedings of the 20th International Conference on Machine Learning*.
- SWEARINGEN, K. AND SINHA, R. 2001. Beyond algorithms: An hci perspective on recommender systems. In *Proceedings of the ACM SIGIR Workshop on Recommender Systems*.
- SZTOMPKA, P. 1999. *Trust: A Sociological Theory*. Cambridge University Press, Cambridge, UK.
- UNGAR, L. AND FOSTER, D. 1998. Clustering methods for collaborative filtering. In *Proceedings of the Workshop on Recommendation Systems*. AAAI Press, Menlo Park, CA.
- WANG, J., DE VRIES, A. P., AND REINDERS, M. J. T. 2006. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'06)*. ACM, New York, NY, 501–508.
- WATTS, D. J. 2003. *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton Studies in Complexity). Princeton University Press.

- XUE, G.-R., LIN, C., YANG, Q., XI, W., ZENG, H.-J., YU, Y., AND CHEN, Z. 2005. Scalable collaborative filtering using cluster-based smoothing. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'05)*. ACM, New York, NY, 114–121.
- ZIEGLER, C.-N. 2005. Towards decentralized recommender systems. Ph.D. thesis, Albert-Ludwigs-Universität Freiburg, Freiburg i.Br., Germany.
- ZIEGLER, C.-N. AND GOLBECK, J. 2006. Investigating Correlations of Trust and Interest Similarity. *Decision Support Services*.
- ZIEGLER, C.-N. AND LAUSEN, G. 2004. Spreading activation models for trust propagation. In *Proceedings of the IEEE International Conference on e-Technology, e-Commerce, and e-Service*. IEEE Computer Society Press, Los Alamitos, CA.

Received January 2007; revised April 2008, March 2009; accepted April 2009