Quantifying Social Influence in Epinions

Akshay Patil Golnaz Ghasemiesfeh Roozbeh Ebrahimi Jie Gao
Department of Computer Science, Stony Brook University, NY
\{akshay, gghasemiesfe, rebrahimi, jgao\}@cs.stonybrook.edu

ABSTRACT

Many eCommerce websites and consumer review websites provide both reviews of products and a social network structure among the reviewers. Users can review products they purchased as well as the reviews others wrote. Users can also rate each other as trusted or untrusted relationships. By studying a data set from Epinions, we examine and quantify the correlation between trust/distrust relationships among the users and their ratings of the reviews. We discover that there is a strong alignment between the opinions of one’s friends and his/her ratings. We use this finding in conjunction with other features to build a model to predict one’s rating tendencies. Our prediction results demonstrate that rating dynamics can be predicted with good accuracy. Our findings also suggest that there is a strong alignment between the collective opinion of a user’s friends and the formation of his/her future relationships.

I INTRODUCTION

Consumer ratings, reports, surveys and polls existed long before the age of the Internet. They are revitalized due to the advent of the Internet and numerous eCommerce websites. Consumer ratings have become much more diverse, popular and accessible. While it was not hard to get reports and ratings about a new vehicle before the age of Internet, it might have been hard to find ratings about a toothbrush, a searing pan or a local restaurant.

Many eCommerce platforms such as Amazon and Ebay actively engage their users in providing their opinions or ratings of the products they have bought. Consumer review websites such as Epinions, Yelp, Angie’s List, etc allow users to review products and also rate reviews written by other users. These websites have come to play an important role in guiding people’s opinions on products and in many cases also influence people’s decisions in buying or not buying the product. As shown in the global Nielsen survey of 26, 486 Internet users in 47 markets, consumer recommendations are the most credible form of advertising among 78% of the study’s respondents [1]. A recommendation from a trusted source goes a long way in shaping others’ opinions.

Some websites incorporated social network structures into their rating systems. Users can rate each other as trusted or distrusted relationships. In such environments, we are dealing with two types of data: “rating system” data and “social structure” data. The incorporation of the social network structure into the rating system leads to an interplay between the two that affects both of them. Users may rate each other as trusted/distrusted based on reviews they write. The social structure might influence users in rating the reviews of others. Furthermore, the reviews in the system and the trust/distrust relationships based on those reviews might influence the formation of new relationships.

Our course of study in this paper can be summarized as follows: (1) We investigate whether there is correlation between the opinions of a user’s current trustees/friends and formation of his/her future relationships; (2) We also investigate whether there is correlation between the opinions of a user A’s current trustees/friends regarding another user B and the score of rating that A would assign to reviews written by B; and (3) We then use our findings in building a predictor model that is able to predict with good accuracy the rating a user is likely to assign to a review.

To carry out such an investigation we use data from a widely used consumer review website Epinions [2,3]. Users write reviews on products or services that may earn them money or recognition. Users also rate reviews written by other users and can also express trust or distrust towards other users. While trust relationships are publicly visible, distrust relations are not visible (except to the user making the relationship). When rating reviews, users have an option of making their rating public or keeping it private.

The trust and distrust relationships for a user are combined to determine the web of trust for that
user; the user is shown reviews written by people within this web of trust more prominently as opposed to other reviews.

1 PROBLEMS AND RESULTS

We define the problems in the context of the Epinions as follows,

1. We investigate the existence of a correlation between the opinions of a user’s current trustees/friends and formation of his/her future relationships. Specifically, if user A’s friends collectively have an opinion (trust/distrust) about user B, would user A’s future relation with B have a correlation with the collective opinion of his/her friends regarding B? For example, if more of A’s friends tend to trust B (rather than distrust B), would A be more likely to trust B? In how many cases would A make a decision in contrast to what his friends think of B? How are the choices made by A related to trustworthiness of both A and B? To carry out such an investigation, we must be very careful about issues like innate differences in trustworthiness of different users.

Results: Our findings suggest that there is a strong alignment between the collective opinion of a user’s friends and the formation of his/her future relationships, after we factor out the innate biases of trustworthiness of different users.

2. We also investigate the existence of a correlation between the opinions of a user A’s current trustees/friends regarding another user B and the score of rating that A would assign to reviews written by B. We are specifically interested in the case where A and B do not have a direct trust/distrust relationship but rather an indirect one (A is not friend/foe of B, but one of A’s friends/foes is a friend/foe of B). We refer to this problem as studying friend-of-friend dynamics. If there is no correlation between friend-of-friend dynamics and a user’s ratings, then we can say that the social structure does NOT provide any advantage to the ecology of the rating system. On the other hand, if there exists such a correlation, we could use this finding to recommend better content to the user. It would also imply that the social structure supports or improves the overall user experience by helping him/her identify relevant content.

Again, we take into consideration the innate differences in rating habits of different users.

Results: Our analysis leads us to conclude that in cases where user A’s friends have expressed approval or disapproval of another user B, there exists an alignment between the A’s rating of B’s reviews and his friends opinions regarding B. On the other hand, approval or disapproval expressed by foes seem to have no correlation with user’s rating whatsoever.

3. We use the FoF dynamics information in the form of features to build a predictor of ratings. We apply the predictor on the ratings assigned by a user to another user’s review and gauge its accuracy. The model that we develop starts from raw Epinions data and extracts a diverse set of features. These features help us in predicting the rating a user is likely to assign to a review. Our predictor model achieves an accuracy in excess of 76% with a ROC area (AUC) of 0.91.

II BACKGROUND AND RELATED WORK

The interplay of eCommerce and social networks has been studied in a number of prior work. A bulk of the research has been devoted to identifying, propagating and predicting trust (and distrust in some cases) relationships. Incidentally, Epinions has served as a platform for many of the studies conducted in this domain. In the following we briefly review the relevant work and point out the differences with what is presented in this paper.

1 SIGNED RELATIONSHIPS

A relationship between two individuals can be signed, i.e., positive or negative. A positive relationship means friends or a trusted relationship. A negative relationship means foes or a distrusted relationship. The study of signed relationships in social network can be traced back to studies in social psychology. The first of these theories known as the structural balance theory was formulated by Heider in 1946 and generalized by Cartwright & Harary in 1956. The theory considers the possible ways in which triangles on three individuals can be signed and states that triangles with exactly odd (one & three) number of positive signs are more prevalent in real networks as opposed to triangles with even (zero & two) numbers.

\[^3\text{While Leskovec et al. in [3] have looked at triadic closure: a relationship with an additional nodes X, we take an aggregate view of the relationship formation process; i.e. we intend to understand the collective role played by all of A’s friends in guiding his future relations.}\]
ber of positive signs. In other words, a triangle with three positive edges means ‘the friend of your friend is your friend’. A triangle with one positive edge and two negative edges means ‘the enemy of your enemy is your friend’. Both are commonly observed in real networks while the other two cases with three negative edges or exactly one negative edge are not likely. This is also termed the strong structural balance. Davis’s theory on weak structural balance [7] is a variant of the structural balance theory that states that all types of triangles except the ones with exactly two positive signs are plausible in real networks. Signed relationships and structural balance theory have been widely used in psychology and international relationships.

Leskovec et al. [1] studied a number of data sets from online review networks to see if they are consistent with the structural balance theory. However, one must notice that in Epinions the relationships are directional, in the sense that one user rates another user as trusted or distrusted, while the structural balance theory only considers mutual relationships. By ignoring the directed nature of links in their datasets, the authors find alignment with Davis’s theory on weak structural balance; their findings though are at odds with Heider’s structural balance theory in two out of the three datasets. On the other hand, when considering the directed nature of the trust/distrust links the authors show that structural balance theory is inadequate in explaining the online structures that come to develop in the datasets. Instead the authors develop the status theory which states that both the sign and direction of the link convey a notion of status. For example a positively directed link is interpreted as indicating that the creator of the link is conferring higher status to the recipient. On the other hand a negatively directed link would indicate that the creator views the recipient as having a lower status. The authors are able to show that the theory of status is a better fit to describe the structures that develop in the real-world datasets. In their follow up work on signed link prediction [8], the authors further shed light on these two theories; they find evidence that the theory on balance operates at a local level rather than the global level whereas the theory on status is shown to operate at a local level as well as lead to an approximate global ordering on nodes.

We do not intend to examine the alignment or disalignment of the data sets with previous theories. Instead, we aim to see whether there exists a correlation between the signed relationships and how users rate each other’s reviews/opinions.

2 PREDICTING SIGNED RELATIONSHIPS

In real data sets it is often the case that only some of the edges are explicitly labeled with a sign. Many other edges may be implicitly signed but not demonstrated in the collected data sets. Thus one interesting problem studied in the literature is to predict the sign of any two individuals.

The first set of papers consider propagation of trust or distrust in a network, thus predict the sign of an edge from the signs of other (neighboring) edges. [9] is amongst the first studies to study the problem of propagation of both trust and distrust in online eCommerce websites. They build on the “path-algebra” model of trust propagation put forward by Richardson et al. [10]. Given a small number of expressed trusts/distrusts per person, they are able to predict trust/distrust between any two individuals with good accuracy of about 85%. Predicting trust/distrust between two arbitrary people allows the system to filter/recommend content more effectively based on their inferred likes/dislikes.

[3, 11] build further on this idea. In [3], Massa & Avesani, experimenting with the same Epinions dataset that we intend to use, come to conclusion that local trust metrics computed via trust propagation are more effective in recommending “suitable” content as compared to global trust metrics and collaborative filtering techniques. In an earlier study by Massa & Avesani [2], they show the efficacy of local trust metrics over global trust metrics for controversial users (i.e. users who are liked and disliked in equal measure or users displaying a polarizing streak). Similarly, O’Donovan & Smyth in [11] demonstrate the usefulness of incorporating trust values into a collaborative filtering algorithm. They experiment with the MovieLens dataset [12] and are able to get a 22% improvement in prediction error rates by using trust values. We refer the reader to [13–16] for works on propagating trust in other related settings.

A separate line of research considered a diverse set of features, at both a network level and an individual level, for “signed” link prediction in online social networks. Liu et al. in [17] construct a diverse set of features from both user attributes and user interactions to predict “trust” links in Epinions. They find that adding interaction features improves accuracy of prediction as opposed to using user features alone for predicting “trust”. A drawback of their study lies in
Table 1: Overall Statistics for Epinions

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Persons</td>
<td>131,828</td>
</tr>
<tr>
<td>Number of Reviews</td>
<td>1,197,816</td>
</tr>
<tr>
<td>Number of Trust Edges</td>
<td>717,667</td>
</tr>
<tr>
<td>Number of Distrust Edges</td>
<td>123,705</td>
</tr>
<tr>
<td>Number of Ratings</td>
<td>12,943,546</td>
</tr>
<tr>
<td>Time Range</td>
<td>Jan-2001 to Aug-2003</td>
</tr>
</tbody>
</table>

the fact that they do not use “distrust” link information and thus are able to only predict “trust” or the lack of it in their dataset. On the other hand Leskovec et al. [8] expand the horizon by being able to predict both positive and negative links in online social networks. The experiments are conducted on datasets from Epinions, Slashdot & Wikipedia used in previous studies [9, 18, 19]. They demonstrate high accuracy (in excess of 90% for 2 out the 3 datasets) in predicting links and their signs based on localized features; in fact they achieve higher accuracy for predicting links that have a higher embeddedness i.e. those that belong to a greater number of triangles. More recent work by DuBois et al. [20] is an extension of the link prediction problem studied in [8]. They develop a spring-embedding algorithm that they use in conjunction with their path-probability technique [21] to infer trust or distrust. They are able to achieve similar accuracy levels as in [8] when predicting undirected signed links but the use of a spring-embedding algorithm limits the application of their technique to directed networks.

All the above work are concerned about detecting the signs of current relationships that are not shown in the (incomplete) data set. Part of our study involves predicting future relationships. But the thrust of our work is to look at whether the opinions of our friends/foes have any correlation with the formation of new relationships.

3 DETECTING DECEPTIVE/FAKE REVIEWS

Since consumers are increasingly relying on user-generated online reviews for making their purchase decisions [22], there has been increasing evidence of deceptive opinion spam - fictitious opinions that have been written to sound authentic. These deceptive opinions distort the perceived quality of the product and reduce the trustworthiness of online opinions in general. In recent years, a substantial body of work has been devoted to identifying and filtering out opinion spam, mostly using natural language processing and machine learning techniques. [23] identifies duplicate reviews, which are invariably fake in nature (since no two products/users would share/assign an identical review). Ott et al. in [24] designing a classifier using a given reference dataset as ground-truth. Feng et al. in [25] study the distributions of review rating scores to identify fake reviews. More recently, people start to look at structural information. Akoglu et al. in [26] exploit the connectivity structure between review data and the reviewers. [27] studies the prevalence of deception spam across different online consumer review websites.

We remark that our work is along a different direction from all the previous work. We aim to identify and quantify the level of alignment between one’s friends/foes opinions and the way he/she decides toward other users. Being able to quantify this correlation/alignment helps us build better models for recommending content as well as understand the interplay between the rating system and the social structure that supports it.

III DATASET

We conduct our experiments on the community of users of the popular consumer review website Epinions. The website allows users to write reviews about products and services. It also allows users to rate reviews written by other users. In our dataset, each review can be rated on an integer scale of [1 – 5] with a rating of 5 signifying a very good review. Epinions also allows users to define their web of trust, i.e. “reviewers whose reviews and ratings have been consistently found to be valuable” and their block list, i.e. “reviewers whose reviews are found to be consistently offensive, inaccurate or not valuable”. The web of trust and the corresponding block list for each user can be used to construct the directed, signed social network of trust and distrust.

Table 1 provides statistics for the Epinions dataset.
Figure 1: Progression of percentage of “trust” edges & R5 ratings over time.

Figure 2: Distributions for Epinions Data. The plots (from left to right, top to bottom) indicate, (1) Degree Distribution of Trust Edges, (2) Degree Distribution of Distrust Edges, (3) Distribution of the Number of Reviews Written per Person, (4) Distribution of the Number of Reviews Rated per Person, (5) Distribution of the Number of Ratings per Reviews, and (6) Distribution of Ratings in the Dataset.
The dataset contains 132k users who issued 840k statements of trust and distrust, wrote 1.2 million reviews and assigned 13 million ratings to the reviews. The dataset spans 2.5 years and it also contains timestamp information on trust/distrust edge formation as well as time of rating a review. About 85k users have received at least one statement of trust or distrust in the dataset. Trust (positive) edges dominate the social network, i.e. over 85% of the edges are trust ones.

Figure 1 shows the progression in the percentage of trust edges over time. As it is depicted, there is a drop in the overall percentage of trust edges indicating that users get more evolved into using distrust edges over time. In Epinions, when a user distrusts another user, reviews that are generated by the distrusted user will become hidden for him/her. Therefore, an increase in the usage of distrust edges would let users to see more of the reviews they like. This would increase the probability of assigning a rating of score 5 (R5) to a review by each user, which is shown in Figure 1 as well.

Figure 2 details various distributions that arise in the dataset. Summarizing the findings, we observe:

- A power-law degree distribution for both trust and distrust edges, i.e. which indicates that most users do not issue more than a handful of trust/distrust statements.
- A large number of users are passive users, i.e. users who write few or no reviews and rarely rate reviews.
- The bulk of content comes from a handful of users who take efforts on writing and rating reviews. We refer to these users as active users.
- A rating of 5 is by far the most prevalent rating in the dataset. Over 78% of all ratings have a score of 5; on the other hand only 0.01% and 2.13% of ratings have scores of 1 and 2 respectively.

Since the dataset contains a handful of active users, we further conduct experiments to test the overlap between them. To do so, we measure the overlap among top-k active users by ranking them in two parts:

1. Rank active users based on the number of trust statements received and the number of distrust statements received.
2. Rank active users based on the number of reviews written and the number of reviews rated.

Measuring the overlap based on part (1), would help us detect any significant overlap between the most
trusted active users and the most distrusted ones. While we would expect a non-zero overlap since leaders often polarize opinions, a large overlap would be detrimental to the objectivity of the system. Similarly, measuring the overlap based on part (2) would help us in observing any significant overlap between the top content generator (people who write reviews) active users and the top content consumer (people who rate reviews) active users. Again, a large overlap would be a disadvantage for the system since that would imply the existence of a closed group of content generators and consumers.

Figure 3 shows the overlap curves for both of the overlap measures. As it is depicted, there is a non-zero, small to moderate overlap for both of the measures for the top 1 – 10% of the active users. Furthermore, although there is a small presence of controversial users (users who are liked and disliked in equal measure; refer to [2] for a comprehensive analysis of such users), a large majority of the top trusted active users are distinct from the top distrusted ones.

IV RELATIONSHIP FORMATION

We investigate the correlation between the opinions of current friends and future relation formations under a simple scenario (see Figure 4). User A is about to trust or distrust user B at time t. A has n friends/users, F1, F2 . . . Fn, whom he trusts. Among these n friends/users, b of them trust B and r of them distrust him. It is important to note that distrust relationships are hidden, i.e. A does not know that “r” of his friends distrust B.

We summarize here the results of this section:

1. First, we present raw observations based on the above scenario. Though informative, these observations should pass some processing to be statistically meaningful.

2. Then, we employ a random shuffling approach utilized in [4] to gauge whether the cases we observe are over or under-represented in the data relative to mere chance. Therefore, we compare our observed results with results achieved after randomly shuffling the edge signs (trust/distrust). Besides over/under-representation, this approach allows us to determine statistical significance of our observed results.

3. Finally, our analysis should consider the fact that users in the data set exhibit a diverse spectrum of linking habits. Again, we use the approach put forward by [4]. We capture and gauge trustfulness and trustworthiness of users and employ them to get a better picture of the correlation between relationship formation and the opinions of friends.

1 RELATIONSHIP FORMATION SCENARIO AND RAW OBSERVATIONS

The following definition captures all the scenarios arising from Figure 4 in more detailed terminology. Depending on values of b and r, there are four distinct possible cases.

Definition 1. We categorize the cases as follows:

- Case 1: b = r
- Case 2: b < r
- Case 3: b > r
- Case 4: b = 0, r = n
1. $b > r$ i.e. A’s friends collectively trust user B.
2. $b < r$ i.e. A’s friends collectively distrust user B.
3. $0 < b = r$ i.e. A’s friends are neutral on whether to trust or distrust $B$.
4. $0 = b = r$ i.e. A’s friends are unopinionated about user B.

We say that A’s friends are opinionated about his decision (trust/distrust) toward $B$ if either (1) or (2) happens. On the other hand, if either (3) or (4) happens, we say that A’s friends are neutral. Depending on the cases, we can figure out if A agrees/disagrees with his friends in trusting $B$.

In 40% of the cases, A’s friends are unopinionated/neutral. The data show that a case of neutral (3) is very rare (≈1%). We do not seek to test any correlation in the unopinionated/neutral cases.

Table 3 summarizes the data in Table 2 in terms of agreement rate. If A’s friends are opinionated about another user, then there is a strong correlation between A’s friends opinions and his/her decision (93.85%); both the collectively trust and collectively distrust categories show very high agreement rates (95.09% and 85.93% respectively). The collectively distrust case is specifically interesting because distrust links are private to other users.

2 STATISTICAL SIGNIFICANCE

We use a similar approach as in [4] to test the statistical significance of raw observations. In such an analysis we intend to gauge whether the cases we observe are over or under-represented in the data relative to mere chance. Therefore, we compare our observed results with results achieved after randomly shuffling of the edge signs (trust/distrust). In other

---

In our dataset, we observe that about 69% of trust/distrust edges are observed on the first day of the crawl, i.e. they bear the same timestamp. Since there is no way of resolving the order in which these 69% of the edges were formed, we treat the network induced by these edges as the snapshot of the network at time $t_0$. We conducted our relationship formation analysis for the remaining subset of edges (31%) that are formed at time $t > t_0$. 

---

Table 2: Evaluation of role played by friends in determining relationships relative to the random-shuffling model. Cases marked in green ($\text{surprise} > 0$) indicate over-representation and cases marked in yellow ($\text{surprise} < 0$) indicate under-representation.

<table>
<thead>
<tr>
<th>Category</th>
<th>Case</th>
<th>Friends Choose To</th>
<th>A’s Decision</th>
<th>Count</th>
<th>Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreed with friends</td>
<td>$b &gt; r$</td>
<td>Collectively Trust</td>
<td>Trust</td>
<td>129,229</td>
<td>11.72</td>
</tr>
<tr>
<td></td>
<td>$r &gt; b$</td>
<td>Collectively Distrust</td>
<td>Distrust</td>
<td>18,420</td>
<td>457.88</td>
</tr>
<tr>
<td>Disagreed with friends</td>
<td>$b &gt; r$</td>
<td>Collectively Trust</td>
<td>Distrust</td>
<td>6,656</td>
<td>-101.98</td>
</tr>
<tr>
<td></td>
<td>$r &gt; b$</td>
<td>Collectively Distrust</td>
<td>Trust</td>
<td>3,016</td>
<td>-40.59</td>
</tr>
<tr>
<td>No role by friends</td>
<td>$0 &lt; b = r$</td>
<td>Neutral</td>
<td>Trust</td>
<td>1,852</td>
<td>-53.56</td>
</tr>
<tr>
<td></td>
<td>$0 = b = r$</td>
<td>Unopinionated</td>
<td>Distrust</td>
<td>878</td>
<td>-9.49</td>
</tr>
</tbody>
</table>

Table 3: The agreement rate of users in forming trust/distrust relationships in scenarios where their friends are opinionated.
Table 4: Generative & receptive surprise values for all 4 scenarios.

<table>
<thead>
<tr>
<th>Case</th>
<th>Percentage of A → B Trust Edges</th>
<th>Generative Surprise</th>
<th>Receptive Surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collectively Trust</td>
<td>95.10%</td>
<td>96.76</td>
<td>34.99</td>
</tr>
<tr>
<td>Collectively Distrust</td>
<td>14.07%</td>
<td>-104.15</td>
<td>-56.31</td>
</tr>
<tr>
<td>Neutral</td>
<td>67.54%</td>
<td>-6.73</td>
<td>-17.38</td>
</tr>
<tr>
<td>Unopinionated</td>
<td>74.89%</td>
<td>-26.87</td>
<td>-27.94</td>
</tr>
</tbody>
</table>

We would like to see whether our observations (in previous subsections) hold even when we take these different habits into consideration. Again, we use the approach put forward by Leskovec et al. in [4].

Definition 3 (From [4]). We define generative and receptive baselines for a user as the fraction of positive links created and received by the user respectively. The baselines for a group of users are derived by summation of individual users’ baselines in the group.

The generative baseline for user A is a measure of A’s trustfulness towards other users. Whereas, the receptive baseline for user B is a measure of B’s trustworthiness.

We also need the notions of generative/receptive surprise values in order to check the validity of observations with respect to linking habits.

Definition 4 (From [4]). The generative surprise for a case is the number of standard deviations by which the actual number of positive A → B edges in the data differs above or below the expected number. The receptive surprise is defined along similar lines. If there was no correlation between friend of A’s opinions and his decision to trust or distrust B, we would expect surprise values of 0.

If A made a decision in forming a relation with B solely based on his linking habits (which is given by generative baseline), then generative surprise = 0. If B is trusted/distrusted by people based on his receptive baseline, then receptive surprise= 0.

Table 4 provides the generative and receptive surprise values for all the cases of Definition 1. The observations show that A’s relationship with B has a strong correlation with his friends opinions regarding B. This alignment might be due to homophily/heterophobia or social influence and our data set can not conclusively show which is the case. However, the existence of such alignments has consequences and applications for websites like Epinions.

In our experiments, we consistently encounter large surprise values, thereby making all our observations statistically significant.
In the collectively trust case, both the generative and receptive surprise values are significantly positive. This means that A exceeds generative baseline in trusting B (which can be attributed to the fact that his friends have collectively trusted B or just homophily). Also, B exceeds it’s receptive baseline in being trusted by A.

In the collectively distrust case, both the generative and receptive surprise values are significantly negative. This means that A falls behind his generative baseline in distrusting B (which can be attributed to the fact that his friends have collectively distrusted B or heterophobia). B falls behind it’s receptive baseline in being distrusted by A.

V RATINGS AND FRIEND OF FRIEND DYNAMICS

We extend the analysis to investigate the existence of any correlation between the opinions of the web of trust of a user and the opinions he/she expresses towards other users. As mentioned before, users of Epinions rate reviews of other users. We seek to determine if there is any correlation between the friendship dynamics and the ratings of a user. We will refer to the problem as analyzing friend-of-friend (FoF) dynamics, though it will also include the remaining three cases, namely, friend-of-enemy (FoE), enemy-of-friend (EoF) and enemy-of-enemy (EoE).

Definition 5 (friend-of-friend (FoF) dynamics). We define the problem using Figure 5. Consider three individuals A, B and C. A has expressed trust/distrust of B at time $t_1$ and B has expressed trust/distrust of C at time $t_2$ (without loss of generality we can assume that $t_1 < t_2$). A chooses to rate a review written by C at time $t_3$ ($t_1 < t_2 < t_3$).

We would like to quantify the correlation between A’s rating and the opinion of B regarding C. Notice that A has NOT expressed any opinion of trust or distrust about C at time $t_3$.

The existence of a correlation might be used to improve recommendation systems (i.e. show content that would be useful to the user). There are four possible scenarios based on Figure 5. These four scenarios are depicted in Figure 6: FoF, EoF, FoE and EoE dynamics. In our dataset, 55% ($\approx 7.2$ million) of all ratings fall under at least one of these four scenarios (a rating can fall under multiple scenarios depending on A’s friends/foes being friends or foes with C). Table 5 provides the raw observations for each scenario. As illustrated in the table, the FoF scenario dominates (over 92%) the other three scenarios, while the EoE (under 1%) scenario being the most rare scenario.

Again we investigate the statistical significance of raw observations by using random shuffling model. We also consider the different rating/rate-ability habits of users by incorporating the generative/receptive baselines approach of [4].

1. We shuffle the edge signs and compute the surprise values for each of the four scenarios. The fourth row in Table 5 illustrates the observations.

It is important to point out that this analysis differs from the one carried out by Leskovec et al. in [4]. They investigate signed triads in online social networks and analyze the over or under-representation of various triadic cases by applying theories from social psychology. The reviews and subsequent ratings by users do not appear in their analysis.
These values indicate that the FoF scenario is over-represented in the dataset relative to chance, whereas the other three scenarios (EoF, FoE, EoE) are under-represented.

2. We keep the edge signs intact while shuffling the ratings around (keeping the same distribution of the ratings shown in figure 2). The random shuffling of the ratings allows us to study the over or under-representation of each rating in each scenario relative to chance (i.e. rating being assigned at random). Table 5 provides the raw counts, percentages and surprise values associated with ratings (as numbers in \{1, 2, 3, 4, 5\}) under each scenario.

3. To get a measure of the rating habits of a user, we extend the definitions of generative and receptive baselines defined in definition 3. Each user is associated with five generative/receptive rating baselines based on five categories of ratings. A generative rating baseline of 0.80 associated with a user A for rating of score 5 would imply that the user A assigns a rating of score 5 in 80\% of times. Similarly, a receptive rating baseline of 0.20 associated with a user C for rating of score 4 would suggest that user C receives a rating of score 4 for 20\% of his reviews. The generative rating baselines for user A is a measure of A’s opinions towards reviews written by other users, whereas the receptive rating baselines for user C are a measure of other users’s opinions of reviews written by C. We can now compute the generative rating surprise and receptive rating surprise values associated with each rating under each scenario. Table 6 provides the generative rating surprise and receptive rating surprise values associated with each rating under each scenario.

1 FINDINGS

We find the following by computing the under/over representation of each rating in each scenario (please refer to Tables 5 & 6 for the surprise values): i) Low ratings (ratings of scores 1 and 2) are over-represented in all scenarios except the FoF one; ii) Rating of score 4 is under-represented across all scenarios; iii) Rating of score 5 (the highest and most frequent rating) is over-represented in all scenarios except the EoF one (in which it is under-represented).

Based on our findings, we observe the following trends:
1. We see a clear trend of alignment between ratings and opinions of friends/foes in two scenarios:

**FoF**: We see a shift towards higher ratings (rating of score 5 is over-represented and ratings of scores 1 and 2 are under-represented in this scenario). These results suggest that user A is more likely to assign higher ratings to user C’s reviews when C happens to be a friend of a friend of A. The generative and receptive surprise values also indicate such a correlation.

**EoF**: We see a trend towards assigning lower ratings (ratings of scores 1 and 2 are over-represented whereas all other ratings are under-represented). This indicates that A would be more likely to assign lower ratings to C’s reviews when C happens to be an enemy of a friend of A. Again, the generative and receptive surprise values support such a correlation.

2. In the remaining two scenarios of FoE and EoE, we see a divided picture. Both low ratings (scores of 1 and 2) and high rating (score of 5) are over-represented whereas the ratings in the middle (scores of 3 and 4) are under-represented. For these two scenarios, we are not able to conclusively show signs of correlation from this analysis.

**VI BUILDING A PREDICTOR MODEL**

In order to further substantiate our findings, we use FoF Dynamics to build a predictor model and apply it to predict the rating assigned by a user to another user’s review. Such a model could also serve as a recommender system by recommending content that is likely to be rated well by a particular user. Table 7 details the features that we use to build such a model. The features can be classified into three distinct categories: (1) **Trust Features** that measure the trustfulness and trustworthiness of the users involved; (2) **Rating Features** which capture the rating habits of the users; and (3) **FoF Dynamics Features** which intends to utilize the correlation between a user’s rating and his friends opinions.

Since the dataset did not contain the actual reviews, we do not use any content features that could analyze the actual content of the reviews.

**1 CORRELATION COEFFICIENTS OF FEATURES**

First, we calculate the Spearman’s rank correlation coefficient between the features and the actual rating.
2 PREDICTING RATINGS

Having described the utilized features, we now describe the techniques used to build the predictor model and also list the performance of these techniques. First, instead of directly using the actual rating as the class label, we define the following three classes of ratings: (1) High ratings which include ratings of scores 4 & 5; (2) Medium ratings which includes ratings of score 3; and (3) Low ratings which include ratings of scores 1 & 2. Given the feature samples, our predictors will attempt to predict these class labels.

Due to the skewed nature of the ratings towards high ratings, if we randomly select training samples, the predictive model will overfit to the high ratings class, and may fail to predict other classes. To avoid over-fitting, we take a balanced sampling approach and generate a training set containing roughly an equal amount of samples from each class. In our experiments, we select 25,000 samples from each of the three classes.

In addition to the correlation coefficient analysis, we also determine the relative importance of the features by computing the mutual information between a feature and the class label. Table 8 lists the results of this analysis. Qualitatively, tables 7 & 8 are in rough agreement.

Finally, to predict the rating classes, we experimented with a number of classification techniques us-

---

Table 6: Rating Habits Analysis. Surprise values indicate over/under representation relative to the rating habits of the users. Quantities \( s_g \) & \( s_r \) indicate the generative/ receptive rating surprise values respectively.

We again observe a shift towards assigning higher ratings (score 5) in the FoF scenario and a shift towards assigning lower ratings (scores 1, 2 & 3) in the EoF scenario.

The correlation coefficient allows us to determine the direction of association between our features and the actual rating. A value of 0 for a feature would indicate that there is no relation between that feature and the actual rating. On the other hand, a positive correlation coefficient would indicate that an increase in the feature value is accompanied by an increase in the actual rating (and vice-versa). The magnitude of the correlation indicates the strength of this association.

Table 7 shows the correlation coefficients for the different features. The findings are along expected lines. They are as follows:

1. The trustworthiness of user C is positively correlated with the actual rating received. Thus a trustworthy user is likely to receive higher ratings for his reviews.

2. The rating tendency of user A is positively correlated with the actual rating i.e. a user is likely to rate current content based on his past rating trends. Similarly, the average rating received by C for his past reviews is a good indicator of the rating that he is likely to receive for his future reviews.

3. Along with the trust & rating features, we observe significant correlations for the FoF & EoF features. These results are in line with our earlier findings in section [V].
<table>
<thead>
<tr>
<th>Feature Class</th>
<th>Feature</th>
<th>Meaning</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust</td>
<td>A’s Generative Baseline</td>
<td>The generative baseline for a user is a measure of his trustfulness</td>
<td>-0.0036</td>
</tr>
<tr>
<td></td>
<td>C’s Generative Baseline</td>
<td>towards other users. The receptive baseline for a user is a measure</td>
<td>-0.1840</td>
</tr>
<tr>
<td></td>
<td>A’s Receptive Baseline</td>
<td>of his trustworthiness in other users eyes.</td>
<td>0.0236</td>
</tr>
<tr>
<td></td>
<td>C’s Receptive Baseline</td>
<td></td>
<td>0.0771</td>
</tr>
<tr>
<td>Rating</td>
<td>Avg. Rating given by A</td>
<td>The average rating given by a user captures his rating tendencies. The</td>
<td>0.2555</td>
</tr>
<tr>
<td></td>
<td>Avg. Rating given by C</td>
<td>average rating received by the reviews written by a user captures the</td>
<td>0.0431</td>
</tr>
<tr>
<td></td>
<td>Avg. Rating received by A’s</td>
<td>usefulness of the content generated by the user.</td>
<td>0.1286</td>
</tr>
<tr>
<td></td>
<td>reviews</td>
<td></td>
<td>0.3606</td>
</tr>
<tr>
<td></td>
<td>Avg. Rating received by C’s</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>reviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FoF Dynamics</td>
<td>Number of FoF Paths</td>
<td>These features capture the number of FoF, EoF, FoE &amp; EoE paths between</td>
<td>0.1112</td>
</tr>
<tr>
<td></td>
<td>Number of EoF Paths</td>
<td>two users. The features are useful in providing context in the absence</td>
<td>-0.0918</td>
</tr>
<tr>
<td></td>
<td>Number of FoE Paths</td>
<td>of a direct link.</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>Number of EoE Paths</td>
<td></td>
<td>-0.0001</td>
</tr>
</tbody>
</table>

Table 7: Spearman’s Rank Correlation coefficient between feature values and actual ratings. Correlation coefficients with significant amplitude (> 0.05) are highlighted in colors, red for negative correlation, and blue for positive correlation. Here user A is assigning a rating to a review written by user C.

<table>
<thead>
<tr>
<th>Class</th>
<th>Feature</th>
<th>Information Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trust</td>
<td>A’s Generative Baseline</td>
<td>0.1595</td>
</tr>
<tr>
<td></td>
<td>C’s Generative Baseline</td>
<td>0.2291</td>
</tr>
<tr>
<td></td>
<td>A’s Receptive Baseline</td>
<td>0.1943</td>
</tr>
<tr>
<td></td>
<td>C’s Receptive Baseline</td>
<td>0.4496</td>
</tr>
<tr>
<td>Rating</td>
<td>Avg. Rating given by A</td>
<td>0.3316</td>
</tr>
<tr>
<td></td>
<td>Avg. Rating given by C</td>
<td>0.3776</td>
</tr>
<tr>
<td></td>
<td>Avg. Rating received by A’s</td>
<td>0.2453</td>
</tr>
<tr>
<td></td>
<td>reviews</td>
<td>0.5362</td>
</tr>
<tr>
<td></td>
<td>Avg. Rating received by C’s</td>
<td></td>
</tr>
<tr>
<td></td>
<td>reviews</td>
<td></td>
</tr>
<tr>
<td>FoF Dynamics</td>
<td>Number of FoF Paths</td>
<td>0.3813</td>
</tr>
<tr>
<td></td>
<td>Number of EoF Paths</td>
<td>0.1894</td>
</tr>
<tr>
<td></td>
<td>Number of FoE Paths</td>
<td>0.0119</td>
</tr>
<tr>
<td></td>
<td>Number of EoE Paths</td>
<td>0.0198</td>
</tr>
</tbody>
</table>

Table 8: Mutual information between features and the rating classes (high, medium & low). Top-5 features are highlighted in blue.

Figure 7: Prediction results using different classification techniques. The figures plot precision, recall & f-measure for the different classifiers (left) and the detailed results for Bagging across the three rating classes (right). Bagging [30] is an ensemble method which improves the classification accuracy through sampling and model averaging. Bagging provides an accuracy in excess of 76% with an ROC area (AUC) of 0.91.
ing Weka \[31\]. Overall, tree-based classification techniques have worked well in our case. Figure 7 reports the prediction accuracy for the different techniques. The overall prediction accuracy is around 76%. The ROC area (AUC) is an impressive 0.91. This is significantly better than random guessing (with 33% accuracy). This result indicates that our feature set (that includes the FoF Dynamics Features) is predictive of the rating class, and the classifier can be used to predict ratings reliably. In fact, when ranked based on information gain, the FoF feature is among the top-3 features.

We also formulated rating prediction as a regression problem, meaning that we constructed a mapping from the feature set to the actual rating value. Regression performance is measured using common metrics such as mean absolute error (MAE) and mean squared error (MSE). The correlation coefficient between the predicted value and the true target value can also be considered as a metric of accuracy, with high value indicating good performance. In Epinions dataset, we achieve a MAE of 0.5541, a MSE of 0.7893 and a correlation coefficient of 0.7646. This indicates that the regression accuracy is also quite good.

We conclude that the feature set combining individual trust, rating & social dynamics features can support rating prediction, and that our approach leads to good prediction and classification accuracy.

VII CONCLUSION

In this paper, we looked at two distinct problems of alignment/correlations in the interplay of eCommerce and social networks. Firstly, we looked at the correlation between one’s current friends and his/her future relationships. Our findings suggest that users are more likely to decide future relationships in alignment to the opinions of their current friends. The interesting observation was that this alignment exists not only in choosing future friends, but also in choosing future foes.

In our second analysis, we studied the alignment between the FoF dynamics and a user’s rating of others content. We concluded that users are more likely to rate the content of a third person in alignment with the opinions of his/her friend regarding the third person. Our findings also show that the opinions of foes have little or no correlation with ratings of a user.

We also built a model that can predict the rating assigned by a user to a review. This model is able to achieve a good accuracy of 76.34% and an AUC of 0.91. Our predictor could also be used to recommend content to Epinions users.

Studying correlation and correctly measuring is very important in design and analysis of rating systems. We have showed that such a study can be utilized to improve the user experience in similar systems or websites. From the sociology point of view, both of the above alignments can be explained by social influence or homophily/heterophobia. However, we should mention that our analysis is not conclusive in favor of either explanation.

ACKNOWLEDGEMENTS

The authors acknowledge supports from DARPA through DARPA/ADAMS program under contract W911NF-11-C-0216, and from NSF through DMS-1221339, CNS-1217823 and CNS-1116881. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the government funding agencies.

References


