The Workshops of the Thirtieth AAAI Conference on Artificial Intelligence
Incentives and Trust in Electronic Communities: Technical Report WS-16-09

User Participation and Honesty in Online Rating Systems: What a Social Network Can Do

Alan Davoust and **Babak Esfandiari**
Department of Systems and Computer Engineering
Carleton University, Ottawa, Canada

Abstract

An important problem with online communities in general, and online rating systems in particular, is uncooperative behavior: lack of user participation, dishonest contributions.

This may be due to an incentive structure akin to a Prisoners' Dilemma (PD). We show that introducing an explicit social network to PD games fosters cooperative behavior, and use this insight to design a new aggregation technique for online rating systems.

Using a dataset of ratings from Yelp, we show that our aggregation technique outperforms Yelp's proprietary filter, as well as baseline techniques from recommender systems.

Introduction

Users of online rating systems contribute ratings and reviews of products (e.g. on Amazon), movies (IMDB, Movielens), or businesses (Tripadvisor, Yelp), and base their decisions to consume these items on aggregate ratings computed from other users' contributions. The intended effect is that the rating shown to a user is *predictive* of the user's future experience with the rated item, and will help the user make an informed decision whether to consume it.

For such systems to be successful, their users need to *cooperate*, by contributing honest ratings. However, contributing these ratings is costly and only benefits others: this implies that participation in such systems has an incentive structure akin to a Prisoners' Dilemma (Harper et al. 2005): the rational behavior is therefore to *free-ride*, i.e. to use the ratings provided by others without contributing. Other online communities, such as peer-to-peer file-sharing networks and recommender systems face the same problem, and a number of incentives schemes have been proposed to address it (Feldman et al. 2004; Ling et al. 2005).

In addition, in the case of rating systems, since positive ratings boost product sales and drive traffic to restaurants and hotels, there is also an incentive for the businesses being rated (or the suppliers of products for sale, etc.) to manipulate the ratings, for example by buying positive ratings and reviews from users. Therefore, a market has flourished for fake reviews (and ratings) on Yelp and other sites (Ashton 2012; Orland 2011).

A major difficulty in dealing with this kind of noncooperative behavior is to distinguish dishonest manipulation from normal disagreement between users. In this context, the users may have a *subjective* perception of whether others are being cooperative or not, and the classic PD model therefore poorly describes the situation.

In this paper, we show how explicit social networks can foster cooperative behavior in the PD game, and investigate how this result transfers to online rating systems, where uncooperative behavior includes dishonest contributions. Our key contribution is a rating filtering scheme that assumes very simple strategic behavior on the users' part, and outperforms several other methods, including Yelp's existing filtering system. We also propose a performance evaluation method which does not require knowing which data is malicious: its rationale is that the rating aggregation's *predictive accuracy* should improve for the users that the scheme *assumes honest*.

The rest of this paper is organized as follows. After a discussion of related work, we present a social network-based variant of the Prisoners' Dilemma, and simulation results showing the dominance of cooperative behavior in this game. We then adapt this game model to an online rating system, where the ratings are shared through an explicit social network. Finally, we present a rating filtering method based on the users' strategic management of their social connections, and compare this approach with Yelp's proprietary filter and baseline techniques from recommender systems.

Related Work

Incentives for Participation

A number of incentive schemes have been proposed to encourage cooperative behavior in online communities, ranging from psychological incentives (Ling et al. 2005) to monetary rewards (Bhattacharjee, Goel, and Kollias 2009). However, these schemes mostly do not address the issue of *honest* contribution, unless honest contributions can be positively identified. An exception is the technique proposed in (Miller, Resnick, and Zeckhauser 2005). However, this technique assumes that ratings by different users are noisy approximations of some "true" quality, rather than a subjective one.

Filtering and Aggregation Techniques

An entirely different approach is to filter contributions deemed malicious, or to aggregate ratings in attack-resistant ways, i.e. "neutralizing" these dishonest contributions.

Yelp implements a proprietary filter which removes approximately 20% of ratings and reviews (considered "unreliable"), which are hidden and do not count towards aggregate ratings. This filter is controversial[1], and its efficiency has not been evaluated directly, for lack of a "ground truth".

Instead of explicit filtering, several manipulation-resistant techniques have been proposed for rating aggregation. In order to address *vote-buying* (reputable users selling high ratings), the Iolaus system (Molavi Kakhki, Kliman-Silver, and Mislove 2013) orders and re-centers ratings before aggregating them, so that users who give mostly extreme ratings will see these become an average rating. This is shown to mitigate the effect of vote-buying, in the sense that the rankings of businesses are quite stable when large amounts of extreme votes are introduced. However, the resulting rating aggregation turns out to be *less* predictive of the users' future experience (than a simple average rating). While the attack-resistant property is valuable, ultimately the purpose of the system is to provide useful ratings to the users.

Trust Networks

Several studies have proposed to use social networks as trust networks (e.g. SumUp (Nguyen et al. 2009) and again Iolaus (Molavi Kakhki, Kliman-Silver, and Mislove 2013)), weighing the aggregated ratings according to topological properties of the social network (e.g. the flow between the users). Such schemes are mainly designed to counter *Sybil* attacks, where the attacker creates multiple identities that all rate the products in order to manipulate the aggregate ratings.

More generally, if the users do not have existing social trust relationships, reputation algorithms (e.g. Eigentrust(Kamvar, Schlosser, and Garcia-Molina 2003) or AppleSeed (Ziegler and Lausen 2005)) can help users identify trustworthy peers based on feedback on their past behavior. However, implementing such algorithms would require the users to rate each others' ratings, adding a layer of complexity to the interface, whereas social networks are often already present (as in Yelp) without being used in the rating aggregation.

Cooperative Behavior in n-player Prisoners' Dilemma games

The Prisoners' Dilemma and its Variations

Formally, the Prisoners' Dilemma (PD) is a symmetric 2-player game with two strategies, known as "cooperate" and "defect". The payoffs are as follows: if one player cooperates and the other defects, then the defector gets a higher payoff, traditionally noted T (for "temptation"), and the cooperator gets the lesser payoff S (for "sucker"). If both cooperate, they obtain the payoff R ("reward"), and if both defect they obtain the payoff P ("punishment"). The payoff values must respect the ordering $T > R > P > S$, which makes *defect* the dominant strategy.

In our situation of interest – online communities, and rating systems in particular – the game definition must be extended to n players, who play repeatedly and remember past results: the *n-player iterated Prisoners' Dilemma*. At each turn, each player chooses a strategy, and receives the sum of payoffs from each pairwise interaction with another player.

In the iterated 2-player game, *adaptive* strategies, such as the well-known "tit-for-tat", where a user first cooperates, then systematically mimics the opponent's previous move, can lead to sustained cooperation (Axelrod and Hamilton 1981). However, for n players, since the players select a single strategy at each turn, they cannot selectively retaliate against defectors without everyone ending up defecting (Yao and Darwen 1994; Ohtsuki et al. 2006). In this model, the rational behavior in an online community is therefore to defect, i.e. free-ride.

An interesting variation of the game is one where players may be organized in a social network, and repeatedly play (n-player) PD games with their immediate neighbours (Ohtsuki et al. 2006; Santos, Pacheco, and Lenaerts 2006). In this setting, the players may also be able to strategically modify their social connections to improve their utility. In this case, evolutionary simulations showed cooperative behaviours generally emerging rapidly (Zimmermann et al. 2000; Santos, Pacheco, and Lenaerts 2006).

This result suggests that the users' ability to modify their social connections can be a determining factor in eliciting cooperative behaviour between the participants. The question now is whether this result can be transferred to a real system, where behaviors are more complex than simply cooperating and defecting.

However, the above results were obtained in evolutionary simulation, with complex rules that may not be directly adaptable to a rating system. In addition, they require the users to interact only with their direct neighbours. Translating this idea to a rating system, where cooperation means contributing and sharing useful ratings, this implies that the users only share ratings with their social connections. As we will see further on, this is problematic in a setting where the data is sparse: the users will only have access to a very small amount of data. We therefore proceed in two steps. We first design a game with simple rules, which can be adapted in a rating system, then extend it to give the users access to more data. We simulate both games to ensure that cooperation is a dominant strategy.

Simulation of Social Network-Based PD Games

First, we specify a simple social network-based n-player Prisoners' Dilemma game, and compare several cooperative and non-cooperative strategies, in order to confirm that the cooperative strategies are dominant. Later, this game specification and dominant strategy will guide our proposal of a rating aggregation scheme based on a social network.

[1]the filter criteria are kept secret, but many business owners believe it to be an extortion scheme whereby fake reviews are filtered from businesses who buy advertizing from Yelp, whereas good reviews are filtered from businesses who don't. This has been the object of several (unsuccessful) lawsuits.

A simple n-player social network-based PD game can be defined as follows:

- The players of the game are connected in a social graph.
- At each turn, each player chooses a single move of the PD game (cooperate or defect).
- In addition, at each turn a player may unilaterally create an edge to another arbitrary node, remove any number of adjacent edges or leave the social network unchanged[2].
- For each player, the payoff for a round is as in a n-player PD game played with its direct neighbours: each player's payoff is the sum of the payoffs that this player gets from the interactions with its different neighbours.
- The game is iterated: the players remember the results of previous rounds and can modify connections accordingly.

Strategies We implemented this game with the Netlogo simulation tool (Tisue and Wilensky 2004), and simulated three simple strategies.

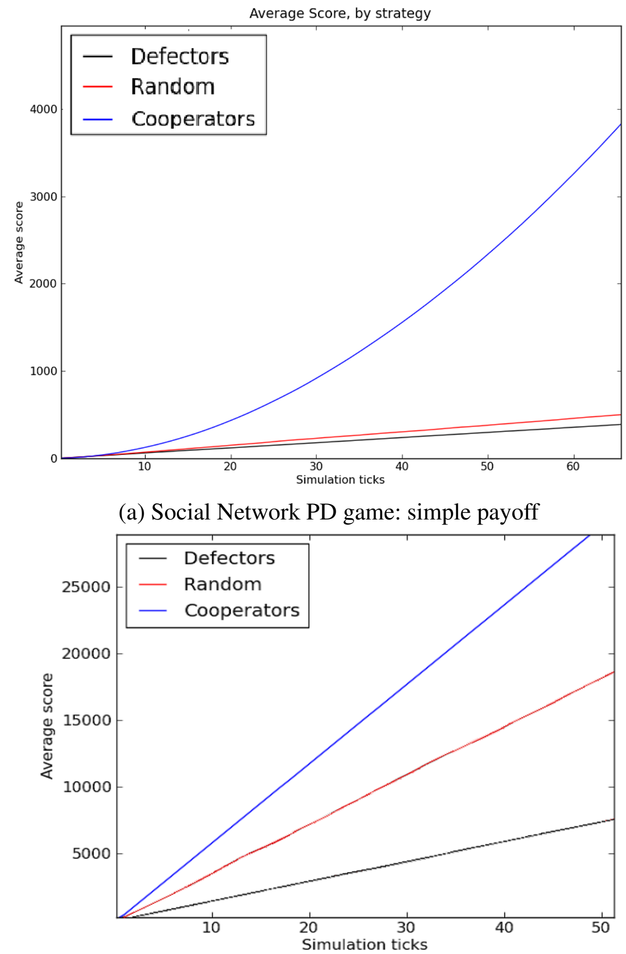
With respect to the prisoners' dilemma, the first strategy was to systematically cooperate, the second to systematically defect, and the third to act randomly. Then, in all strategies, the players modified their connections according to the outcome of past interactions. They created links to (random) other players, and removed links to players who had defected.

Evolution of the Social Graph Simulating these strategies, we observe that the cooperators quickly establish a connected network, while the defectors and random players are isolated due to their propensity to defect. The subgraph of cooperators densifies until becoming complete. The defectors create new links at every round, but these links are systematically removed, due to them defecting. As a result, the nodes remain isolated. The random players create links at each round, and these links last until one of the connected players defects. We observe that random players (here with a cooperation probability of 0.5) have small degrees, with an average value around 0.87.

As a result, the largest connected component of the graph is made up of all the cooperative players, plus a number of random players, namely all those who have established (direct or indirect) links with cooperators since they last defected. Empirically, we observe that in a simulation run with 100 players implementing each strategy, the size of the GCC hovers around 140 players.

Payoffs Due to their much higher degrees, the cooperators immediately acquire much higher payoffs than the players implementing other strategies (figure 1a). As the graph of cooperators densifies, their *per round* payoffs increase (convex curve), then stabilize once their degrees stabilize.

[2]allowing only one edge creation per turn vs. unlimited edge removal is intended to avoid a "race" situation between one user trying to establish an edge vs. the target user removing it: this "race" would have no useful interpretation in a real system which should have safeguards against unwanted connections. We also note that in previous studies the players were unable to unilaterally remove connections.



(a) Social Network PD game: simple payoff

(b) Social Network PD game: Connected Components payoff

Figure 1: Social Network PD game simulation: mean scores of players, by strategy

Connected Component Payoffs

As discussed above, we now extend this game so that the players can interact not only with their direct connections, but with a wider community. For this purpose, we introduce an alternative payoff function, inspired by data sharing mechanisms in peer-to-peer networks (in particular, so-called "friend-to-friend networks" (Galuba 2009; Popescu 2006)). Instead of sharing data only with one's immediate neighbors, data is shared transitively in the entire network, i.e. the entire connected component that the user is part of. This can be viewed as interpreting social trust as transitive[3].

The new payoff function is therefore calculated by playing a n-player PD game with all the players in the same connected component of the social graph.

We modified our Netlogo model to include this scoring method. In this game, once all the cooperators form a connected network (almost immediately), their payoffs per

[3]We leave the analysis of partial transitivity to future work.

round become constant (linear increase in fig. 1b): at this point, there is no longer any benefit in creating additional connections.

However, the defectors can also benefit from this scoring method: if they can establish a single connection with a cooperator, they also become part of the connected component with the other cooperators, and can derive a very high payoff even while having very few connections.

A realistic way for cooperators to handle this would be to "blacklist" users so that once a connection has been removed once, it can no longer be re-created. We modified the players' connection management strategy to systematically blacklist users after dropping a connection. This makes it impossible for defectors to re-establish unwanted connections.

Conclusions

Through this social network-based PD game, we showed that when the users *strategically rewire their connections*, the PD strategy of *cooperating* dominates the strategy of *defecting*, as well as the *random* strategy.

More sophisticated behaviours may show opportunities to "game the system" (e.g. the *whitewashing* and *sybil* attacks), but we leave this analysis to future work: for now, these results intuitively illustrate the importance of the social network to n-player PD games.

However, it is now important to validate that this strategy is consistent with useful behaviour in an online rating system, where the payoff essentially results from the users having access to useful ratings.

A Social Network-based Rating System

In the social network-based PD game, a simple connections management strategy enabled the cooperative users to form a connected network and isolate defectors. In this case study, we adapt the game definition and strategies to a rating system, and validate them with a dataset of Yelp ratings[4].

Modeling the users' interaction by a PD represents the following simplifications: (i) a user's payoff is the number of ratings that this user consumes and (ii) every user consumes all of the available ratings at each turn. Furthermore, the social network-based user interaction can be interpreted as follows. The participants are organized in an explicit social network, they can establish new connections anytime (provided the target user agrees, as per the idea of "blacklisting"), and unilaterally remove connections that they are unhappy with. The users share their ratings through the social network, and conversely, when a user views the aggregate rating of an item, this aggregate is computed from ratings by users of her network, including direct and indirect connections.

For now, we consider a simple average as an aggregation mechanism, and focus on the filtering of the ratings that the social network provides.

Payoff and Cooperative Behavior

According to the notion that cooperative behavior is beneficial to others, cooperation and defection in a rating system are subjective. For example, a user Alice may find that some restaurant on Yelp has a high average rating, and decide to go to this restaurant, only to be disappointed by her experience. In such a situation, the users who contributed the high ratings of that restaurant (including for example some user Bob) were, in Alice's perspective, *defecting*, whereas those who contributed low ratings were *cooperating*. However, some other user Charlie may go to the same restaurant and enjoy his experience because the food matches his tastes. In Charlie's perspective, Bob is a cooperator. In addition, Alice and Charlie could very well have contributed mutually useful ratings on some other restaurant, making them perceive each other as cooperators.

What we are interested in discovering is the result, at a global level, of the users applying *local* decisions following the simple connections management strategy discussed above, which we introduced with clearly partitioned cooperator/defector populations. In this case, since there is no "objective" labelling of cooperators/defectors, we cannot compare the utilities of these two strategies. Instead, we can let all the users apply this strategy, and evaluate the predictive accuracy of the ratings that they will have access to.

Agreement Graph

Our first question is the following: if these users apply the strategy of maintaining connections with (subjectively) cooperative users, then what are the properties of the social network that will emerge from these local decisions? One possibility would be that the strategy might allow users to form small clusters of like-minded users, with intra-cluster agreement and inter-cluster disagreement. Our analysis will show that this is not the case.

We can construct this social network using the Yelp dataset. We first construct the graph of users who have rated at least two businesses in common, in order to have a reasonable basis to evaluate agreement. This graph includes 25,936 users, and 1.93M edges. Of these users, 25,813 (99.5%) form a giant connected component (GCC) in the graph. The remaining 55,000 users have at most 1 rated business in common with any other user, making it difficult to evaluate their agreement with other users[5].

The graph has a scale-free topology, meaning its connectivity is robust to the removal of many (random) nodes and edges (Albert, Jeong, and Barabási 2000).

We also find that removing the edges that represent *disagreement* does not break up the graph either. In other words, the users do not form clusters with internal agreement. Instead, we observe that most of the users find high level of agreement with a small number of other users. The transitive closure of these local agreements forms a GCC

[4]The dataset was collected and shared by the authors of (Molavi Kakhki, Kliman-Silver, and Mislove 2013). It includes around 280,000 ratings by 80,000 users on 9,000 businesses listed in Boston, USA, as well as the social links between the users. For each rating, the dataset indicates whether it was flagged by Yelp's filter.

[5]This also shows the extreme sparseness of the dataset.

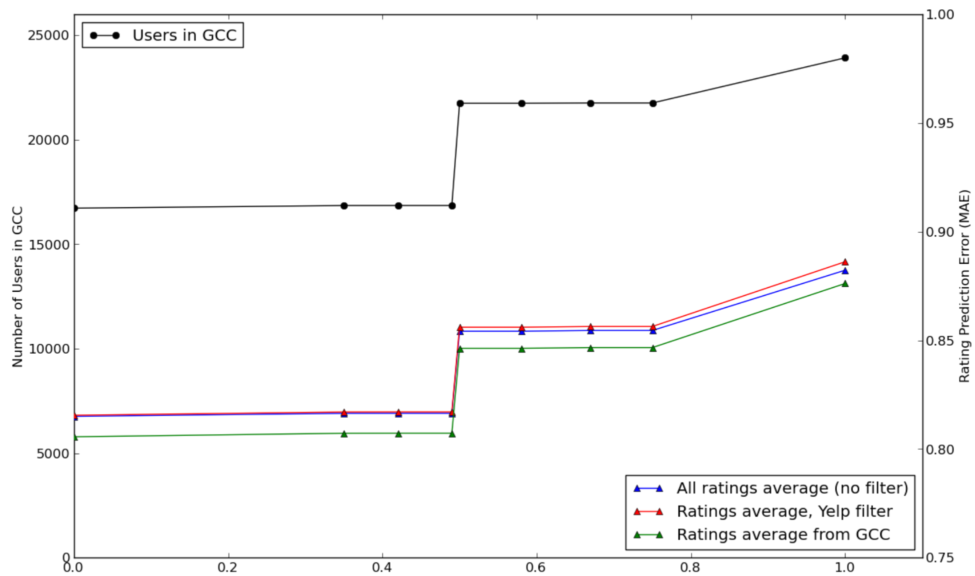


Figure 2: Evaluation of the agreement graph as a ratings filter, against Yelp's filter. For different threshold values (x axis), the top plot indicates the number of users in the GCC, and its y scale is on the left axis. The other plots indicate the MAE for different filtering criteria, and their scale is on the right axis.

with a very large number of users, and those who are not in the GCC are mostly isolated.

With an agreement threshold of 1, the graph is left with 1.28M edges (66% of the full graph), but still has a giant connected component comprising 23,920 users, or 92% of the total (Fig. 2, top plot, with y scale at left).

Our conclusion is that this strategy does not allow the users to form small communities of like-minded users. Instead it appears to isolate some users who agree with nobody else (or have no ratings in common with others). As we show in the next section, isolating these users is nonetheless beneficial to the users in the connected component.

Evaluation

We now propose a method to evaluate this strategy with respect to the users' utility. Beyond any specific goal of attack-resistance, ideally these different mechanisms should objectively make the system more useful, in the sense that aggregate ratings shown to honest users should be better predictions of their future experience. As the honest users are not known, we can at least evaluate the improvement to the users that the scheme *assumes* honest. In this sense, we can compare different filtering and aggregation techniques.

Applying this method to the agreement graph, the main question is: for the users inside[6] the GCC of the agreement graph, is the data available in the GCC more valuable than the full collection of ratings?

As discussed previously, the relevance of a set of ratings to a user can be measured by the *predictive accuracy* of this set of ratings with respect to the users' own ratings. This predictive accuracy can be compared with the baseline (all ratings) and with the Yelp filter (fig. 2). As an accuracy metric, we use the mean absolute error (MAE) between the true rating and the rating prediction, averaged over all users.

We observe that the agreement graph, for all threshold levels, provides valuable filtering. Although the absolute error reduction is quite small (note the y scale on the right), we find that the error is significantly lower with the agreement graph than with Yelp's filter (paired t-test, $p < 0.0001$.).

We note that the baseline and the performance of Yelp's filter appear to vary depending on the (unrelated) threshold used in the agreement graph. This is simply because we evaluate the performance of the rating aggregation, for all three filtering mechanisms, on the same set of users, i.e. those in the GCC of the agreement graph, which is different for each threshold (fig. 2, top plot).

As the error increases with the size of the GCC, this analysis does not lend itself to determining the best value to use as a threshold. We address this in the next section, in a more general comparison between filtering methods.

Social Network The users of Yelp can establish "friend" relationships and form an undirected social graph. Presumably, this social network was not established following the exact strategy that we described above. However, people tend to associate with people of similar tastes and interests (McPherson, Smith-Lovin, and Cook 2001), and it is worth evaluating this social network as a filtering mechanism, and comparing it with the "agreement-based" strategy.

Among the 80,000 users of our dataset, around 25,000 have at least one friend. They form another scale-free graph with a giant connected component of 23,538 users. Of these, only around 13,000 are also found in the GCC of the largest agreement graph, which makes the two graphs rather com-

[6] For the users outside of the GCC, who are mostly isolated and have access to little or no data, the agreement graph is clearly not helpful.

plementary.

Repeating the experiment of the agreement graph, we find again that the predictive accuracy of the ratings from the GCC of the social network is significantly better than the baseline and Yelp's filter.

In addition, the success of these two graph-based techniques suggests an additional approach: since the nodes in the social graph's GCC only have a limited overlap with those of the agreement graphs, it would be valuable to merge these graphs, which could potentially provide the observed improvements to more users.

Using a threshold of 1 for the agreement graph, we merge the two graphs and obtain a giant connected component with 34,599 nodes, for which the MAE values are: baseline 0.9091, Yelp filter 0.9194, network GCC 0.9039. Again, the network GCC provides a statistically significant improvement, with $p < 0.0001$.

Similarity-based Filtering The success of the "agreement graph" approach suggests that the users could simply take into consideration the ratings from the k most similar users, a technique first used in the collaborative filtering system Grouplens (Resnick et al. 1994).

In this technique, a rating by a user u on an item i is predicted by aggregating the ratings on i of other users similar to u. The similarity between two users is measured by how closely these users have agreed on past ratings of items they both rated.

One difficulty is that the Yelp ratings are very sparse, meaning that in many cases (i.e. for many specific ratings) the dataset simply does not have k other ratings that can be used as a basis for prediction (i.e. ratings from users with non-zero similarity). When k neighbours are not available for prediction, we fall back to the baseline approach (aggregating all available ratings). This allows us to apply the technique for any user, and better compare with the alternative techniques.

In general, we observe that similarity-based filtering does not perform well, since it is generally worse or comparable to the baseline (all ratings being used).

Overall Comparison These different techniques were individually compared with the baseline, and on different subsets of users. We now compare them all together, on a set of users for whom all the techniques are applicable, namely the intersection of the social graph's GCC with the most restrictive agreement graph's GCC (with a threshold of 0).

In table 1, we compare the different filtering techniques, and indicate for each technique what proportion of the original ratings are removed.

We observe that on this population, the Yelp filter outperforms the baseline. The agreement graph techniques outperform the Yelp filter (statistically significant improvement, $p < 0.0001$). Among the different thresholds for the agreement graph, the difference between threshold values 0 and 0.5 are not statistically significant, but they are between 0.5 and 1. The social network performs significantly better than the agreement graph. When the two graphs are combined, performance drops slightly (i.e. the value goes up, but lower is better; the difference is statistically significant for $p < 0.01$) with an agreement threshold of 1. However, as mentioned previously, with the combined graph the filter applies to more users.

Table 1: Comparison of different ratings filtering methods

Method		**Ratings removed**	**MAE**
baseline		0%	0.7925
Similarity		N/A	0.7925
Yelp Filter		19%	0.7911
Agreement	t=1	25%	0.7834
Graph (AG)	t=0.5	28%	0.7826
	t=0	33%	0.7826
Social Network (SN)		38%	0.7806
SN + AG	t=1	21%	0.7820
SN + AG	t=0	25%	0.7805

Conclusions

In this paper, we have proposed the use of an explicit social network as a trust network, in an online community where ratings are shared. Using a general game-theoretic model, we have shown that such a mechanism introduces incentives for cooperative behavior: the key mechanism is that users should manage their social connections *strategically*, i.e. drop connections to uncooperative users. In this sense, the trust management shifts from the system to the users.

In practice, this suggests that users of a rating system should maintain connections with the users whom they trust and / or agree with. If the users act rationally, then uncooperative users should end up being isolated.

While we could not evaluate this directly, for lack of "ground truth" knowledge of the users' honesty, we evaluated the strategy using a dataset where Yelp considers approximately 20% of the ratings to be "unreliable".

Using an objective performance metric, we showed that the filtering produced by our agreement graph is more useful than Yelp's filter, and more useful than a traditional similarity-based approach, in the sense that aggregating the remaining ratings produces higher predictive accuracy.

Interestingly, Yelp's existing social network has similar properties, and even generally outperforms the agreement graph. Since these two graphs only have partial overlap, they could be combined to provide many users with a valuable data filtering mechanism.

In practice, Yelp could implement this filter based on the existing social network, then make "friend" suggestions to users excluded from this network, based on past agreement. Such a filter could advantageously replace Yelp's existing filter, which is viewed with suspicion by businesses, and has caused several lawsuits.

Future Work In future work, it would make sense to explore the *partial* transitivity of trust, and investigate whether it may be optimal to propagate ratings to a limited social distance, instead of the entire connected component.

References

Albert, R.; Jeong, H.; and Barabási, A.-L. 2000. Error and attack tolerance of complex networks. *Nature* 406(6794):378–382.

Ashton, K. 2012. The not so secret business of fake yelp reviews. http://www.dailydealmedia.com/657the-not-so-secret-business-of-fake-yelp-reviews/.

Axelrod, R., and Hamilton, W. D. 1981. The evolution of cooperation. *Science* 211(4489):1390–1396.

Bhattacharjee, R.; Goel, A.; and Kollias, K. 2009. An incentive-based architecture for social recommendations. In *Proceedings of the Third ACM Conference on Recommender Systems*, RecSys '09, 229–232. New York, NY, USA: ACM.

Feldman, M.; Lai, K.; Stoica, I.; and Chuang, J. 2004. Robust incentive techniques for peer-to-peer networks. In *Proceedings of the 5th ACM conference on Electronic commerce*, EC '04, 102–111. New York, NY, USA: ACM.

Galuba, W. 2009. Friend-to-friend computing: Building the social web at the internet edges. Technical Report LSIR-REPORT-2009-003, EPFL.

Harper, F.; Li, X.; Chen, Y.; and Konstan, J. 2005. An economic model of user rating in an online recommender system. In Ardissono, L.; Brna, P.; and Mitrovic, A., eds., *User Modeling 2005*, volume 3538 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 307–316.

Kamvar, S. D.; Schlosser, M. T.; and Garcia-Molina, H. 2003. The Eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the 12th international conference on World Wide Web*, WWW '03, 640–651. New York, NY, USA: ACM.

Ling, K.; Beenen, G.; Ludford, P.; Wang, X.; Chang, K.; Li, X.; Cosley, D.; Frankowski, D.; Terveen, L.; Rashid, A. M.; Resnick, P.; and Kraut, R. 2005. Using social psychology to motivate contributions to online communities. *Journal of Computer-Mediated Communication* 10(4).

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 415–444.

Miller, N.; Resnick, P.; and Zeckhauser, R. 2005. Eliciting informative feedback: The peer-prediction method. *Management Science* 51(9):1359–1373.

Molavi Kakhki, A.; Kliman-Silver, C.; and Mislove, A. 2013. Iolaus: Securing online content rating systems. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, 919–930. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee.

Nguyen, T. D.; Bonan, M.; Jinyang, L.; and Lakshminarayanan, S. 2009. Sybil-resilient online content voting. In *NSDI*, volume 9, 15–28.

Ohtsuki, H.; Hauert, C.; Lieberman, E.; and Nowak, M. A. 2006. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* 441(7092):502–505.

Orland, K. 2011. Microsoft investigating claims of ratings manipulation in xbox live indie games.

Popescu, B. 2006. Safe and private data sharing with turtle: Friends team-up and beat the system (transcript of discussion). In Christianson, B.; Crispo, B.; Malcolm, J.; and Roe, M., eds., *Security Protocols*, volume 3957 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 221–230.

Resnick, P.; Iacovou, N.; Suchak, M.; Bergstrom, P.; and Riedl, J. 1994. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, 175–186. ACM.

Santos, F. C.; Pacheco, J. M.; and Lenaerts, T. 2006. Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology* 2(10):e140.

Tisue, S., and Wilensky, U. 2004. Netlogo: A simple environment for modeling complexity. In *in International Conference on Complex Systems*, 16–21.

Yao, X., and Darwen, P. J. 1994. An experimental study of n-person iterated prisoners dilemma games. *Informatica* 18(4):435–450.

Ziegler, C.-N., and Lausen, G. 2005. Propagation models for trust and distrust in social networks. *Information Systems Frontiers* 7(4-5):337–358.

Zimmermann, M. G.; Eguíluz, V. M.; San Miguel, M.; and Spadaro, A. 2000. Cooperation in an adaptive network. *Advances in Complex Systems* 3(01n04):283–297.