Trust Propagation with Mixed-Effects Models

Jan Overgoor*, **Ellery Wulczyn*** and **Christopher Potts**
Stanford University
Stanford, CA 94305
{overgoor, ewulczyn, cgpotts}@stanford.edu

Abstract

Web-based social networks typically use public trust systems to facilitate interactions between strangers. These systems can be corrupted by misleading information spread under the cover of anonymity, or exhibit a strong bias towards positive feedback, originating from the fear of reciprocity. Trust propagation algorithms seek to overcome these shortcomings by inferring trust ratings between strangers from trust ratings between acquaintances and the structure of the network that connects them. We investigate a trust propagation algorithm that is based on user triads where the trust one user has in another is predicted based on an intermediary user. The propagation function can be applied iteratively to propagate trust along paths between a source user and a target user. We evaluate this approach using the trust network of the CouchSurfing community, which consists of 7.6M trust-valued edges between 1.1M users. We show that our model out-performs one that relies only on the trustworthiness of the target user (a kind of public trust system). In addition, we show that performance is significantly improved by bringing in user-level variability using mixed-effects regression models.

Introduction

A major goal for social networking sites it to facilitate interactions between individuals without prior experience with one another. The traditional approach to dealing with the uncertainty inherent in these interactions is to use public trust systems. A typical example would be eBay, where buyers and sellers evaluate each other based on trustworthiness and these evaluations are made public. Such public evaluations may fail to be informative for several reasons. The relative anonymity on the Web can lead some users to provide false information, to improve their own image or damage the image of others. Where the cover of anonymity is removed, users are often reluctant to give negative evaluations because they fear retribution (Adamic et al. 2011), which leads to misleading positive biases in the ratings. Furthermore, because these ratings depend on people's preferences and perspectives, users might not know how to interpret or value opinions provided by strangers, as opposed to opinions from acquaintances. This issue is especially important in situations where there is no agreement on the merit of the rated person or item (Massa and Avesani 2005; Golbeck 2005).

Trust propagation algorithms seek to overcome these shortcomings by inferring trust ratings between strangers from trust ratings between acquaintances and the structure of the network. Such approaches seek to capitalize on the strengths of existing ties when fostering new ones. In this paper, we investigate an algorithm of this sort that is based on user triads: the trust rating from user 1 to user 3 is predicted from the 1-to-2 and 2-to-3 trust ratings. The propagation function can be applied iteratively to propagate trust along paths between a source user and a target user. In addition, we show how mixed-effects linear regression models (Gelman and Hill 2007; Baayen, Davidson, and Bates 2008) can be used to bring different kinds of user-level variation into the model, so that predictions can be tailored to the different ways in which people interpret the concept of trust.

We evaluate our approach using data from CouchSurfing.org, a site that allows travelers to both provide and find temporary lodging. The site is an excellent testing ground for research into trust systems because its users provide a wide range of information about themselves, their couchsurfing experiences, and their connections with other users. Our primary data consists of 7.6M trust-valued edges between 1.1M users. We show that our model out-performs one that relies only on the global trustworthiness of the target user (a kind of public trust system). In addition, we show that performance is significantly improved by bringing in user- and network-level variability using mixed-effects regression models.

Related Work

A number of people have investigated the reputation system of the CouchSurfing network. Lauterbach et al. (2009) and Adamic et al. (2011) analyzed the effects that the public characteristics of the reference-, vouching- and friendship-networks have on the quality of the information they contain. Both argue that fear of reciprocity is an important factor in the extreme bias towards positive feedback. Furthermore, Bialski and Batorski (2007) tried to predict trust values based on external features of the users. In contrast, we

*The first two authors contributed equally to this study.

use only the network structure to predict trust values.

The task of predicting trust values from a trust network has been studied before. Guha et al. (2004) formulate a framework for propagating trust and distrust using matrix multiplication methods to predict the sign of a link in the Epinions network, and they discuss the issue of mapping predicted real values back to the relevant trust categories. Leskovec, Huttenlocher, and Kleinberg (2010) applied logistic regression, without propagation to predict the sign of a link in Epinions, Wikipedia, and Slashdot reputation systems using user features such as average incoming trust and frequencies of different types of user triads.

Due to the existence of users where there is no agreement on their trustworthiness in the Epinions network, Massa and Avesani (2005) argues for a local trust metric, personalized to each user as opposed to global trust systems. We propose a new algorithm for computing a local trust metric via trust propagation in a network with six trust categories instead of only two, and we explicitly model user differences.

CouchSurfing Dataset

CouchSurfing (CS; `http://www.couchsurfing.org/`) is an international hospitality network that allows travelers to find temporary hosts on their journeys. Travelers ("*surfers*") request hospitality from hosts who advertise their homes on the site. Due to the high-stakes of both letting a stranger into one's home and staying at a stranger's house, users are invested in a well-functioning trust system and take trust ratings seriously. As a result, data from the CS community is especially valuable for investigations into online trust. Users of the site maintain a public profile that contains personal information and a history of references. After a hospitality exchange, users leave each other public textual references, which are tagged as positive, neutral, or negative.

Furthermore, when users make an online friendship connection, they quantify their relationship along the dimensions of friendship degree and trust degree. The friendship degrees are visible to everyone on the site, but trust values are hidden. This allows users to confidentially assess each other's trustworthiness. A legend of the trust degrees ("*Site Val.*"), as they appear on the site, is displayed in Table 1. Note that the semantics for a trust value of 1 are closer to an unknown value than a very low score. In order to work with these values more effectively, and stay truer to their semantics, we mapped the occurrences of 1 to that of the mean of all trust values in the data set (4.26). The actual values we used are displayed in the *Used Val.* column of Table 1.

Site Val.	Descriptor	Freq.	Used Val.
1	*I don't know them enough to decide*	455K	4.26
2	*I don't trust this person*	21K	2
3	*I trust this person somewhat*	934K	3
4	*I generally trust this person*	2791K	4
5	*I highly trust this person*	2497K	5
6	*I would trust this person with my life*	912K	6

Table 1: Legend and frequency of trust values.

The goal of CS is to facilitate new connections between people, with its reputation system as a central component. It allows people to evaluate whether or not they want to interact with someone they have never met before. It highlights active and respectable users and reveals misbehavior. Since the interactions primarily take place offline, preventing bad ones is important. The public reference system is, however, not a fully reliable representation of the sentiment between users. For example, only 1.6% of references are non-positive, which suggests more positivity than is warranted (Adamic et al. 2011). Not everyone leaves a reference after an experience, and, if they do, they may be strongly biased towards a positive one out of a fear of retribution. The confidential trust values that users assign to each other are far less correlated ($p = 0.382$) than the public friendship levels ($p = 0.705$). This strongly suggests that the public rating features of the site are shaped in large part by fears of reciprocity.

The complete social network consists of 3.4M users, 1.1M of which have given or received at least one trust rating. In total, there are 7.6M connections with a trust rating ($\mu = 4.26, \sigma^2 = 1.19$).

Task

Our general task is to predict the trust one user has in another. In our approach, we base this estimation on the paths that connect them. We primarily investigate how trust propagates between people who are two steps apart. This can be thought of as closing a trust triad (see $\hat{t}_{13}$ in Figure 1). We define a trust propagation function between users 1 and 3 using the path via user 2 as

$$\hat{t}_{13} = \tau(t_{12}, t_{23})$$

where t_{ij} is the trust user i has in user j. Since there may be several paths of length two between two users (with set K of intermediate users), we can extend the model by letting

$$\hat{t}_{13} = g(\{\tau(t_{1k}, t_{k3}) | k \in K\})$$

where g aggregates all the different predictions associated with the different length two paths. Given such a one-step trust propagation function, one can predict trust on paths of length n by applying the function $n-1$ times, for example:

$$\hat{t}_{14} = \tau(\tau(t_{12}, t_{23}), t_{34}) = \tau(\hat{t}_{13} t_{34})$$

This too is illustrated in Figure 1. In this paper, our focus is on paths of lengths 2 and 3. We plan to investigate performance on paths of longer length in future work.

Trust Propagation Models

The *global* model predicts the trust user 1 has in user 3 to be the average of the trust scores user 3 has received. This is designed to approximate a public trust system, where people look at a set of public reviews and make their own assessment based on weighing the positive and negative opinions. The *global* model is independent of the trust triads between user 1 and user 2. In fact, it is independent of user 1, since all users will be assigned the same trust in user 3. Although this model cannot take into account differences in how much one

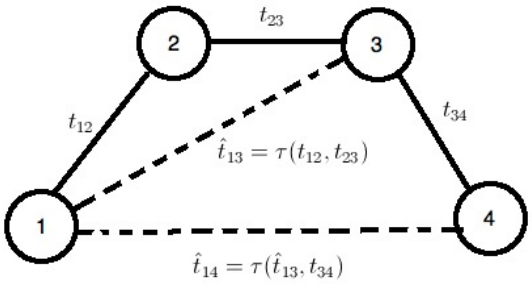

Figure 1: Trust propagation.

user trusts another, it has one advantage: the set of users who have reviewed user 3 is larger than the set of users that user 1 is acquainted with that have reviewed user 3, so there is typically more information available to *global* than a model that uses triads to make predictions, though this information is of lower quality because it is not specific to user 1.

In contrast to *global*, the models *lin*(t_{12}, t_{23}), *sp-avg*(t_{12}, t_{23}), and *lmer*(t_{12}, t_{23}) use triad structure to make predictions and can assign different trust values to a fixed user 3 for different values of user 1. *lin*(t_{12}, t_{23}) is a simple linear model of t_{13} using t_{12}, t_{23}, and the interaction term $t_{12} \cdot t_{23}$ as covariates. *sp-avg*(t_{12}, t_{23}) predicts the average value of t_{13} given t_{12} and t_{23}. In practice, this means binning the cases by their values of t_{12} and t_{23} and taking the average of the corresponding t_{13} values. If we consider t_{12}, t_{23}, and t_{13} to be random variables, then under certain probabilistic models for $p(t_{13} \mid t_{12}, t_{23})$, *sp-avg*$(t_{12}, t_{23})$ is equivalent to $E[t_3 \mid t_1, t_2]$, which achieves the lowest possible mean squared error (MSE) in predicting t_{13} given only t_{12} and t_{23}. In other words, with MSE as a model performance metric, *sp-avg* is better than any model that uses only t_{12} and t_{23} to predict t_{13}, including *lin*.

Given the same values of t_{12} and t_{23}, *lin* and *sp-avg* will predict the same trust for any user 1 and a fixed user 3. If there are differences between users in how trust propagates along a path with the same trust structure, it may be possible to make better predictions by devising models that take these differences into account. We therefore extend the simple linear fixed effects model *lin* to create a linear mixed effects model with user 1 as a random effect (*lmer*). This means that the degree to which t_{13} will be affected by t_{12}, t_{23}, and the intercept term will depend on user 1. In general, linear mixed effect models are appropriate for representing clustered, and therefore dependent, data. This arises, for example, when data is gathered over time on the same individuals, which is the case for the CS trust network.

We aggregate different predictions for the trust user 1 has in user 3 by letting the aggregation function g be the weighted average of the predictions, where the weights correspond the trust user 1 has in the intermediary user 2. This is motivated by the consideration that information coming from trusted friends is more reliable.

Model	Triad MSE	(δ)	All-Triad u1–u3 MSE	(δ)
mean	0.7498	(+0.027)	-	
global	0.7079	(+0.265)	-	
lin	0.6501	(+0.109)	0.6079	(+0.145)
sp-avg	0.6427	(+0.116)	0.5988	(+0.154)
lmer	0.5327	(+0.226)	0.5167	(+0.238)

Table 2: Performance on the two-step propagation prediction task. δ is the difference in MSE between applying the model to a shuffled versus a non-shuffled network.

Experiments

To test the merit of the different models, we extracted 14.4M triads and 180M tetrads from the CS trust network. For the mixed-effects model *lmer*, we used the `lme4` package in R (Bates 2005). Model performance is measured in MSE between the predicted value and the actual value. Each model was tested using 5-fold cross-validation, where we train on a random set of 300K training examples and tested on disjoint random set of 5K test examples for each fold.

We test our models on simple triad closure ("*Triad*") and triad closure using the weighed average of all triad closures between user 1 and user 3 ("*All-Triads u1-u3*"). Next we test the models on their merit for propagation. Since there is an error associated with our two-step trust propagation functions, we expect repeated application of these functions to lead to larger error terms. To evaluate our model of trust propagation, we would ideally investigate how the error increases for every path length up to the maximum in the system. In this paper, we only examine paths of length three, by applying the trust propagation function twice.

We evaluate our models against two baseline models. The model *mean* always predicts the mean of all trust values in the data set. We chose this as a baseline since 81.7% of scores are within one standard deviation of the mean. As a result, simply predicting the average trust in the system, which is 4.26, has a low MSE. Furthermore, in order to disentangle the influence of the distribution of trust scores from the structure of trust triads on model performance, we created a second data set by shuffling all the trust values and then tested our models on this new data set (Golbeck 2005). The comparison between a model's performance on a shuffled versus a true data set indicates how well it captures the structure of the trust network.

The results for two-step propagation are displayed in Table 2, and two-step propagation is displayed in Table 3. δ denotes the improvement over the MSE of the model on the shuffled data set.

Discussion

In the simple triad closure setting, all models perform better than our baselines. This indicates that the models discover some structure in the data despite the fact that the distribution of scores is so heavily clustered around the mean. The models that are some function of t_{12} and t_{23} all outperform the models that are not (i.e., *mean* and *global*). This demon-

Model	MSE (δ)
mean	0.7595 (+0.000)
global	0.8322 (+0.000)
lin	0.8144 (−0.148)
sp-avg	0.7887 (−0.122)
lmer	0.5392 (+0.126)

Table 3: Performance on the three-step propagation prediction task. δ is the difference in MSE between applying the model to a shuffled versus a non-shuffled network.

strates that there is structure in the trust triads that can be used to improve predictive power over a global metric.

The model *lin* performs slightly worse than *sp-avg*, which is to be expected since it is a less flexible model. *lin*'s fitted coefficients are nonetheless informative:

$$t_{13} = 2.72 + 0.2t_{12} - 0.04t_{23} + 0.04t_{12}t_{23}$$

The effect of coefficient for t_{12} is an order of magnitude higher than the coefficient for t_{23}. This might seem surprising since we are predicting the trust user 1 has in user 3 and t_{23} is the only information we have about user 3. Instead, what weighs most heavily into the prediction is how much user 1 trusts user 2, an indicator of user 1's trust behavior. It may be that how much a user trusts a stranger does not depend as much on the trustworthiness of the stranger, as judged by other people, as on that user's trust behavior. This underlines the subjectivity of trust and further motivates modeling user differences in trust behavior.

The strong performance of *lmer* shows that there are differences between users in trust behavior and in how trust propagates along a path with the same trust-valued edges. Modeling these differences improves accuracy. Finally, we observe using the average prediction over all trust triads between the source and target users leads to greater accuracy (except for *global* and *mean*, whose predictions are independent of the the trust triads).

For the task of propagating trust along three-step paths, most models perform worse than when predicting trust over two-step paths. We find, however, that *lmer* performs approximately the same on three-step paths as over two-step paths. This indicates that a model that uses trust triad structure and user differences is able to propagate trust well. As mentioned above, we would want to test the performance of *lmer* on paths of length larger than 3 in future work.

Conclusion and Future Work

We investigated a trust propagation algorithm that uses trust ratings between acquaintances to infer trust ratings between strangers in a social network. We argued that such algorithms have advantages over the public trust systems in wide use today. We then evaluated a family of trust propagation algorithms using data from CouchSurfing.org, a large social networking site with a wide range of trust ratings and other social metadata. The results provide initial experimental evidence in favor of trust propagation. In addition, we showed that using mixed-effects models with user-specific hierarchical effects can substantially improve performance. This is achieved by bringing into the models the substantial amount of variability users display in their use of these rating systems as well as differences in how trusting they are.

We plan to build on these insights by exploring how accurately our method of trust propagation can estimate trust beyond three steps. It will also be interesting to investigate the effect that user features like sex, age and origin have on propagation. One potential application of inferred trust scores is to provide personalized trust assessments between users in web-based social networks. Another is to investigate how an approximation of trust that a user has in a source can be used to weigh the information from that source when aggregating information from many sources.

Acknowledgments

We thank George Zisiadis from CouchSurfing and Bogdan State for their assistance in providing the data and making the project possible. This research was supported in part by ONR grant No. N00014-10-1-0109.

References

Adamic, L.; Lauterbach, D.; Teng, C.; and Ackerman, M. 2011. Rating friends without making enemies. In *Fifth International AAAI Conference on Weblogs and Social Media*.

Baayen, R.; Davidson, D.; and Bates, D. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4):390–412.

Bates, D. M. 2005. Fitting linear mixed models in R. *R News* 5(1):27–30.

Bialski, P., and Batorski, D. 2007. Trust networks: Analyzing the structure and function of trust. In *Sunbelt conference poster*.

Gelman, A., and Hill, J. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Golbeck, J. 2005. *Computing and applying trust in web-based social networks*. Ph.D. Dissertation, University of Maryland.

Guha, R.; Kumar, R.; Raghavan, P.; and Tomkins, A. 2004. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, 403–412. ACM.

Lauterbach, D.; Truong, H.; Shah, T.; and Adamic, L. 2009. Surfing a web of trust: Reputation and Reciprocity on CouchSurfing. com. *2009 International Conference on Computational Science and Engineering* 346–353.

Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010. Predicting positive and negative links in online social networks. *Proceedings of the 19th international conference on World wide web* 641–650.

Massa, P., and Avesani, P. 2005. Controversial users demand local trust metrics: An experimental study on epinions.com community. In *Proceedings of the National Conference on Artificial Intelligence*, volume 20, 121. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.