A Comparative Analysis of Reputation Mechanisms for Social Networks

Kate Anderson*

Department of Economics, University of Michigan, Ann Arbor, Michigan 48109

Jonathan Lafky†

Department of Economics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Andrew Stout‡

Department of Computer Science, University of Massachussets - Amherst, Amherst, Massachussets 01003

Jack Waddell§

Department of Physics, University of Michigan, Ann Arbor, Michigan 48109

This paper compares the ability of two synthetic reputation voting systems, one democratic and one meritocratic, to correctly identify high and low ability individuals connected via a network. We find that in the long run, both systems do well at identifying ability level. However, the speed of convergence is different for each system.

*Electronic address: andersok@umich.edu

†Electronic address: jm144@pitt.edu

‡Electronic address: stout@cs.umass.edu

§Electronic address: seoc@umich.edu

I. INTRODUCTION

Reputation systems have become a major component of many electronic environments. From online auctions, to social networking sites and message boards, reputations and the systems that track them play a vital role in ensuring high-quality interactions between users. What, then, is the best way to build a reputation system?

Several different systems have been implemented, with mixed results. Perhaps the most straightforward system is to calculate a simple weighted average of review scores, as is used by eBay, Amazon, and other major online companies. In this system each time two users interact one or both of them rates their experience. This can be either bi-directional, such as with eBay's ratings for both buyers and sellers, or mono-directional, as with Amazon's ratings for sellers only. This system uses global, centrally stored reputations. Any single individual has exactly the same reputation visible to all other people in the system. In this sense this means of tracking reputation can be considered to be objective.

However, in a more conventional model of reputation, there is no central scorekeeper maintaining a tally on each person's actions. Instead, every person stores their own opinion of other individuals, perhaps in consultation with common neighbors, making this a more subjective measure of reputation.

Furthermore, there are distinctions in reputation systems in which the reputation of the judges themselves affect the degree to which their judgment is considered, and those in which every observer has an equal ability to vote for the reputation of other members. These systems we call meritocratic and democratic voting, respectively. Systems such as the aforementioned eBay rating system are democratic, but others are more similar to Slashdot's comment rating feature, in which authors of highly rated comments are granted greater power to influence the ratings of future comments.

Given the numbers of members on such large systems, it is impossible for a single user to build a relationship or history with all other users, and therefore they rely on reputation systems in their judgment of whether to trust a given member, either as a potential partner for a transaction or as a source of information. Therefore it is important to test reputation voting systems to determine whether some can more accurately represent the community's judgment of the skill or trustworthiness of an individual.

II. RELATED LITERATURE

Network-based reputation mechanisms have been studied extensively in several fields, including computer science, political science and economics. The majority of such work has been done in computer science, where it is dominated by research on peer-to-peer networks.

Gupta et al. (2003) examine reputation systems on peer-to-peer networks. They address the challenge of maintaining accurate reputations for the quality of users' upload/download habits without the use of a central authority. They develop a system in which reputations are constructed and stored by each user individually, but using universal and objective metrics for assessing behavior. Their primary focus in analyzing the system is one of cost and efficiency of implementation.

Hogg and Adamic (2004) look at reputation mechanisms in online social networks. They compare the performance and implementation costs of an extension of the PageRank algorithm, the Sporas reputation mechanism and the Regret system for online social networks. Their analysis has an eye towards preventing, or at least mitigating the effects of collusion amongst agents. To this end they carry out a field experiment on BuddyZoo, testing their ability to carry out undetected collusion. They find that, in general, small-scale collusion in the presence of social network-based reputation mechanism is difficult to impossible, depending on the network configuration.

Baliga and Sjöström (2001) examine the optimal design of peer-review and self assessment schemes in the context of a workplace. The paper addresses the need to encourage employees to share their opinions of one another's quality, in an attempt to determine which projects to fund and which to cut. They examine mechanisms by which employees are incentivized to truthfully reveal other's qualities. Ultimately they conclude that none of the peer-review mechanisms outperforms self-assessment, demonstrating that in some instances networks are not helpful for ascertaining accurate reputations.

III. AN OBJECTIVE REPUTATION SYSTEM

In the objective system we examine the ability levels of a population of agents. We assume that each agent, i, has an ability level $\omega \in [0,1]$. Individuals with $\omega > .5$ we call high ability and those with $\omega < .5$ we call low ability. Each individual has a reputation level $R_i(t)$ which is based on his neighbors' evaluation of his ability level. Each of i's neighbors receive a signal of i's ability level,

$$S_{ij}(\omega_i, \omega_j) = \omega_i + \epsilon(\omega_j)).$$

The noise level in the signal, $\epsilon(\omega)$, is a decreasing function of the ability level of the observer, so that $\epsilon(\omega) \sim N(0, \sigma^2(\omega_j))$ with $\sigma^2(\omega_j) = (1 - \omega_j) \cdot \zeta$. ζ is the noise variance parameter.

Each neighbor then uses this signal, his past opinions of i and i's past reputation to form his current opinion of i, given by:

$$O_{ij}(t) = \beta_h O_{ij}(t-1) + \beta_r R_i(t-1) + (1 - \beta_h - \beta_r) S_{ij}(t),$$

where O_{ij} is j's opinion of i's skill, R_i is i's reputation, β_h is the historical weight of j's previous opinion of i, and β_r is the reputation weight for the conformity of j's opinion to the previous reputation of i. Note that the condition $\beta_h + \beta_r < 1$ must be maintained.

Each neighbor then places a vote, $V_{ij}(t) \in \{0, 1\}$ for individual i based on this opinion. $V_{ij}(t) = 1$ if $O_{ij}(t) > .5$ and $V_{ij}(t) = 0$ otherwise. $V_{ij} = 0$ is interpreted as vote for low ability and $V_{ij} = 1$ is a vote for high ability. The weighted average of all of i's neighbors' votes is then used to calculate i's reputation current reputation,

$$R_i(t) = \frac{\sum_{j=1}^{d_i} \alpha_j(t-1) V_{ij}(t)}{\sum_{j=1}^{d_i} \alpha_j(t-1)},$$

where α_j is voter weight:

$$\alpha_j(t-1) = \begin{cases} 1 & \text{for democratic voting} \\ 0.1 + 0.9 \cdot R_j(t-1) & \text{for meritocratic voting} \end{cases}$$

In democratic voting, every member has an equal influence over the reputation vote. In meritocratic voting, an individual's voting capacity is linearly related to its own reputation; a constant non-zero offset ensures that even the least reputable maintain some influence and that a community cannot spuriously get trapped in a zero-reputation loop.

IV. A SUBJECTIVE REPUTATION SYSTEM

In the decentralized subjective reputation system, an individual's reputation is in fact an implicit notion, rather than a quantity explicitly and centrally stored as above. That is, each individual stores an opinion of each of its neighbors, and updates those opinions as will be described below; a given individual's (global) reputation would of course be some kind of aggregation of all opinions of that individual, but such an aggregation is never performed. Thus, for example, an individual may have a positive (local) reputation within one community of the network, but a negative (local) reputation within another.

At each step, each individual updates its opinion of all of its neighbors as follows. For each neighbor i, the evaluating individual j forms a preliminary opinion:

$$O'_{ij}(t) = \beta_h O_{ij}(t-1) + (1-\beta_h) S_{ij}(t),$$

where β_h, O_{ij}, and S_{ij} are as defined above. The evaluating individual j then polls all of its common neighbors with i, and incorporates their opinions of i weighted by its own opinion of them:

$$O''_{ij}(t) = \frac{1}{N_{ij}} \sum_{k:k\sim i,k\sim j} O_{kj}(t-1) O_{ik}(t-1) + (1 - O_{kj}(t-1)) O'_{ij}(t),$$

where the normalizing factor $N_{ij} = |\{k : k \sim i, k \sim j\}|$ is simply the number of common neighbors of i and j. (Intuitively, O''_{ij} can be seen as an average of the common neighbor's opinions of i, where when j has a low opinion of a common neighbor, j weights its own preliminary opinion of i more heavily than the common neighbor's.) Finally, j's opinion of i is:

$$O_{ij}(t) = \beta_r O''_{ij}(t) + (1 - \beta_r) O'_{ij}(t),$$

where β_r is as defined above, and $O_{ij}(t)$ is truncated to be within the range $[0, 1]$.

Thus, instead of a global reputation, the evaluating agent gives consideration to a local reputation of the agent being evaluated, by "asking around" and incorporating others' opinions selectively based on the evaluating agent's opinion of those sources of information—one gives more consideration to the opinion of someone of whom one thinks highly than to that of someone of whom one has a low opinion.

The vote and reputation are calculated as for the objective system.

V. RESULTS

A. Objective Measure

We first explore the objective system of reputation for both democratic and meritocratic voting. The simulations are performed on a random network between $N = 50$ agents, with a wiring probability $p = 0.3$, unless otherwise noted. The variance parameter of the normally distributed signal noise was set to $\zeta = 1$ unless otherwise stated. The average percentage of correct classifications (i.e. reputation of high or low matches skill classification of high or low) is calculated for the entire network over the last 20% of each run, then averaged between runs.

Overall, we found that agents using an objective measure were very good at classifying agents' skill level. Over the parameter space (β_h, β_r) the agents' success rate was over 95%. There are a few rough trends that can be seen. Systems that put more weight on previous experience (high β_h) and moderate weight on previous reputation values (moderate β_r) tend to be more accurate in classifying individuals. Also meritocratic voting appears to be performing better than democratic voting, in general. However, none of these differences are statistically significant.

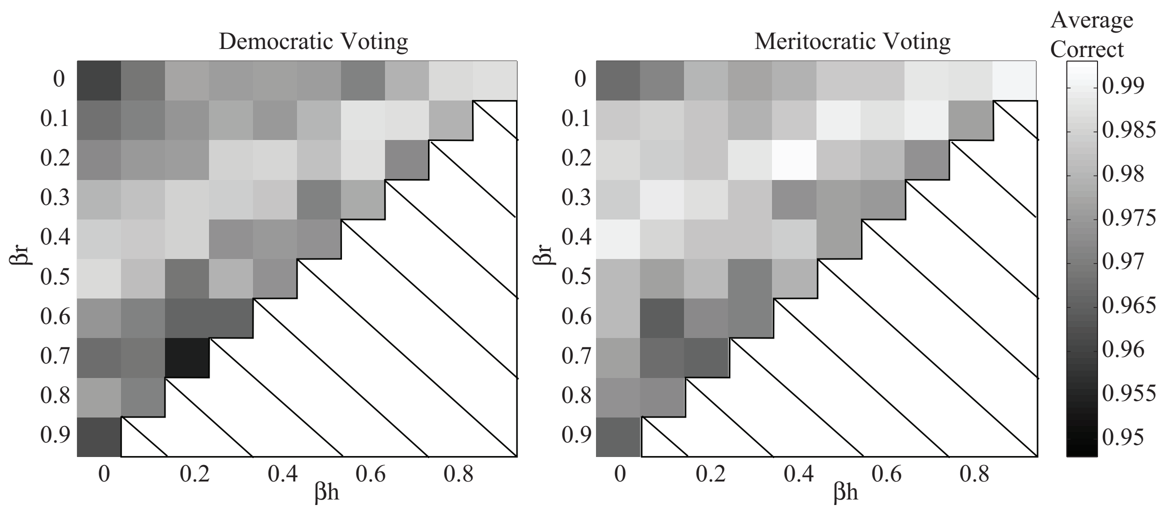

Figure 1: Average correct classification for democratic and meritocratic voting over the parameter space.

However, some significant differences were observed in the length of time it took individuals to correctly classify their neighbors—i.e., the degree of convergence of the classifications. The in-run standard deviation of the percent correct classification describes the degree to which opinions and classifications are still changing in the network. A large region of parameter space allows complete convergence of reputation. This convergence is more sensitive to reputation weighting than to history weighting: conformity has the stronger power to elicit a stable opinion. This is also seen in the difference between democratic and meritocratic voting. Meritocratic voting allows convergence with a small reputation weight because the more talented opinion-formers drive the conformity of the group.

In order to determine the dependence on number of agents and degree, a network of $N = 100$ agents was used with $p = \{0.15, 0.3\}$. The first value of p preserves the mean degree, while the second is the value used in the previous cases. A couple of trends can be noted. First, there is rough improvement with higher β_r. Secondly, higher connectedness p

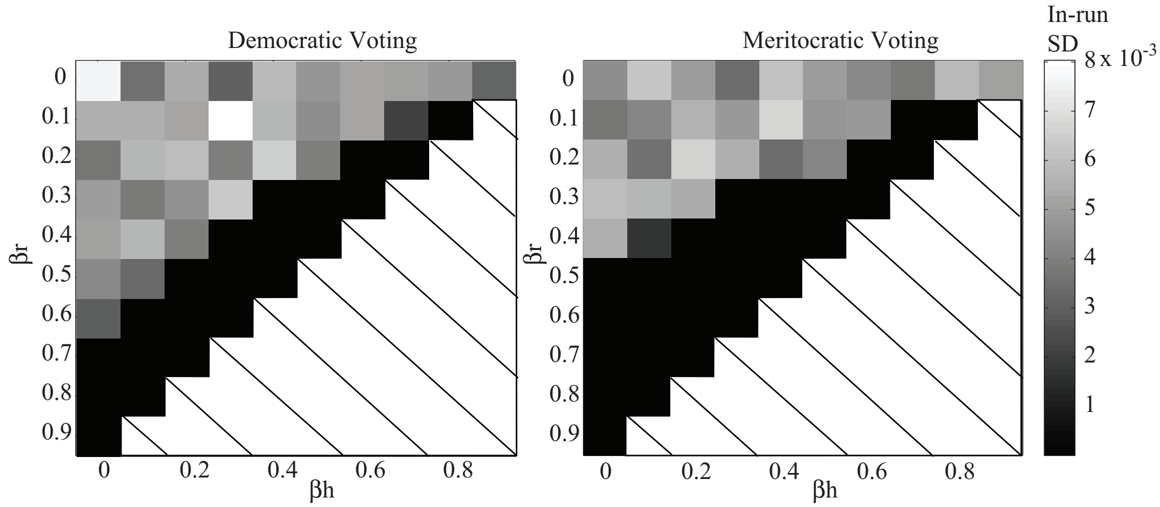

Figure 2: Average of the standard deviations of correctness over the last 20% of each run, for democratic and meritocratic voting over the parameter space.

results in better classification performance.

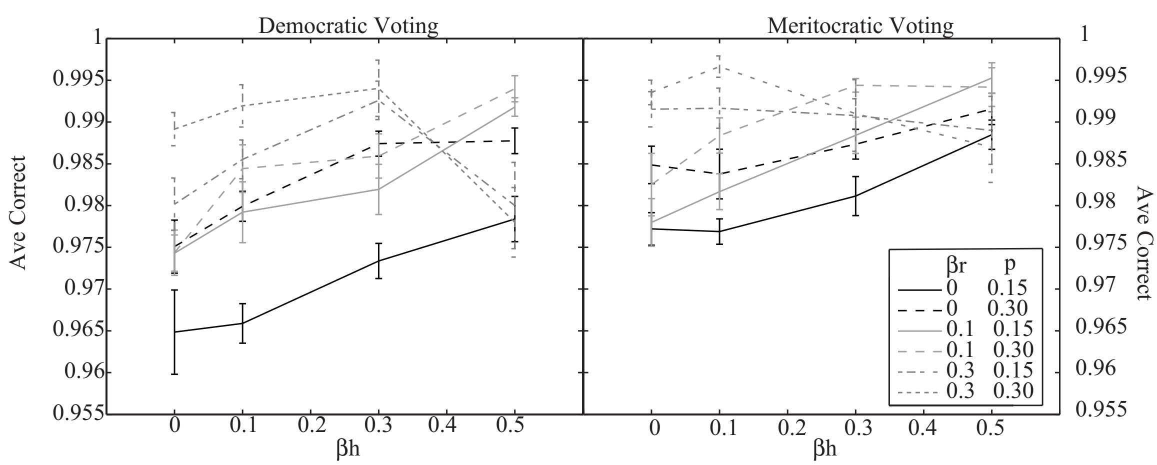

Figure 3: Average correctness for democratic and meritocratic voting for various reputation and history weights, and for wiring probability $p = \{0.15, 0.3\}$.

It is clear that the failure in classification occurs with agent i's true skill is near enough the 0.5 threshold and the signal noise of j large enough that the observation lies on the wrong side of the threshold compared to the true value. We briefly demonstrate the effect of increasing the variance parameter of the normally distributed noise, for β_r =0.1 and various values of history weight. As expected, increased noise variance leads to poorer classification by the network.

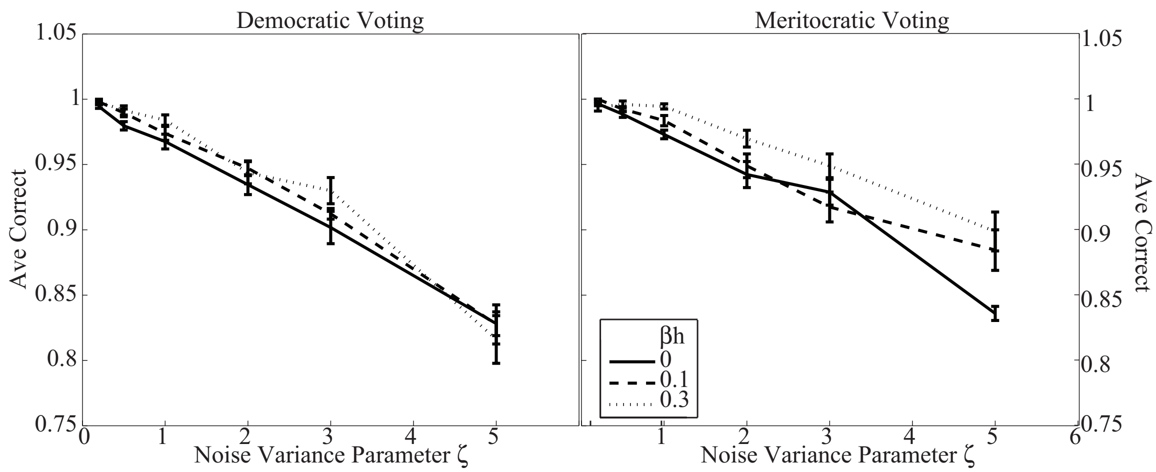

Figure 4: Average correctness for $\beta_{h=0.1}, \beta_r = \{0, 0.1, 0.3\}$ for various values of the noise variance parameter.

B. Subjective Measure

The system is modified by removing the central record of reputation, so that agent j must communicate with neighbors shared with i to determine a reputation. Again, there is a modest increase in correct classifications with increasing β_h, but only some of this increase is significant, particularly that of the democratic voting. Increased β_r has no clear effect (Fig 5).

VI. CONCLUSION AND FUTURE WORK

We have found there to be very little difference in effectiveness between the subjective and objective reputation systems, and only slight differences in converge between democratic

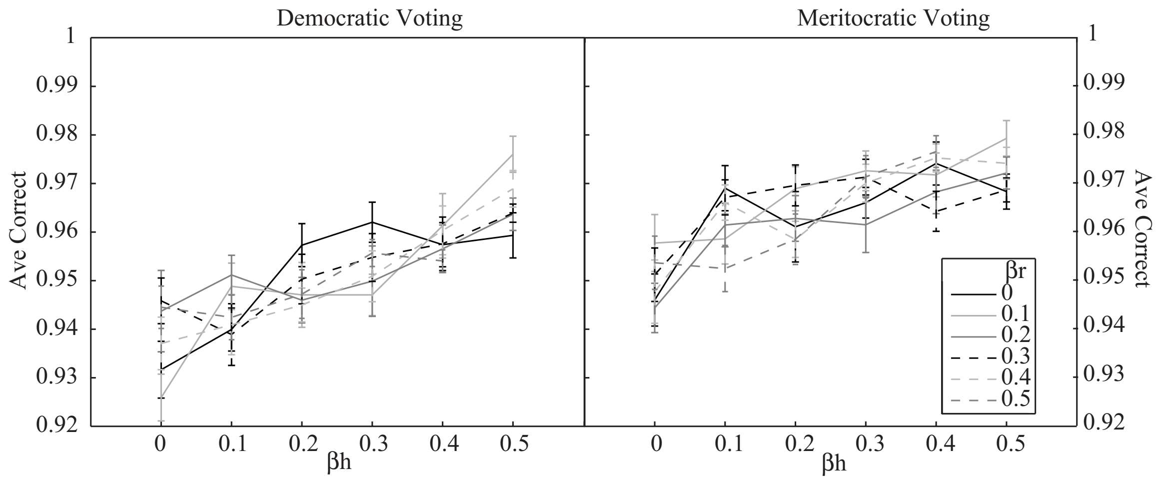

Figure 5: Average correctness for subjective measurement of reputation

and meritocratic voting systems. This suggests that either system can be safely implemented in place of the other, depending on the environment. The subjective system is likely better suited to decentralized situations such as Peer-to-Peer systems, whereas the objective system may be better for centralized online communities, such as message boards or marketplaces. Meritocratic voting may be useful in systems in which quick convergence may be helpful.

Examining the effectiveness of objective and subjective mechanisms, with either meritocratic or democratic vote weighting, with different networks structures and ability distributions may reveal more meaningful differences between the systems.

Acknowledgments

This research was done with the support of NSF and the NIH (JW). The authors would like to thank the Santa Fe Institute for their support for this project. The authors would also like to thank the Center for the Study of Complex Systems at the University of Michigan for the use of their computer systems for this investigation.

VII. BIBLIOGRAPHY

Baliga and Sjostrom “Optimal Design of Peer Review and Self-Assessment Schemes” RAND Journal of Economics, Vol. 32, No. 1. (2001) pp. 27-51

Cerulo “To Err Is Social: Network Prominence and Its Effects on Self-Estimation” Sociological Forum, Vol. 5 No. 4 (1990) pp. 619-63

Gupta, Judge and Ammar, “A Reputation System for Peer-to-Peer Networks” ACM NOSSDAV (2003)

Hogg and Adamic, “Enhancing Reputation Mechanisms via Online Social Networks” ACM EC (2004)