UNIVERSITY
OF TRENTO

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

38050 Povo – Trento (Italy), Via Sommarive 14
http://www.dit.unitn.it

DIGITAL REPUTATION FOR VIRTUAL COMMUNITIES

Anurag Garg, Roberto Battiti

October 2005

Technical Report # DIT-05-066

Digital Reputation for Virtual Communities

Anurag Garg Roberto Battiti

Dipartimento di Informatica e Telecomunicazioni,

Università di Trento, Via Sommarive 14, 38050 Povo (TN), Italy.

{garo,battiti}dit.unitn.it

Abstract

In this report we provide an overview of digital reputation management systems for virtual communities. We begin with a discussion on the how trust is created in the real world. We then highlight how the process of trust creation differs in the virtual world. Next, we talk about the different ways in which digital reputations are useful to virtual communities such as incentivizing cooperation and punishing malicious users, providing recommendations or as a form of distributed authentication. We then discuss how evolutionary biology and game theory have informed research in reputation systems. This is followed by a discussion of the various components of reputation such as context, the forms of evidence, first and second order reputation and architectural considerations including the role of distributed hash tables in decentralized reputation management systems. We conclude with a discussion of some reputation management systems such as complaints-based trust, EigenTrust, PeerTrust and ROCQ.

Keywords: Reputation management, virtual communities, peer-to-peer systems.

1 Introduction

A virtual community can be defined as a group of people sharing a common interest or goal who interact over a virtual medium, most commonly the Internet. Virtual communities are characterized by an absence of face-to-face interaction between participants which makes the task of measuring the trustworthiness of other participants harder than in non-virtual communities. This is due to the anonymity that the Internet provides, coupled with the loss of audiovisual cues that help in the establishment of trust. As a result, digital reputation management systems are an invaluable tool for measuring trust in virtual communities.

Trust is an important component of all human interactions whether they take place online or not. There is an implicit assumption of trust in each interaction that we participate in that involves some investment on our part. The Merriam-Webster dictionary defines trust as *assured reliance on the character, ability, strength, or truth of someone or something*. Even more pertinent to virtual communities is an alternative definition which states that trust is a *dependence on something future or contingent; reliance on future payment for property (as merchandise) delivered.* These definitions illustrate that the basis of trust is an expectation of future payment or reward and that the transaction partner will behave in an honest fashion and fulfill their obligations.

The processes behind the creation of trust can be classified in four broad categories:

1. **Personality-based Trust**: a person's innate disposition to trust others.

2. **Cognitive Trust**: through recognition of the characteristics of the transaction partner.

3. **Institution-based Trust**: a buyer's perception that effective (third-party) institutional mechanisms are in place to facilitate transaction success [49].

4. **Transaction-based Trust**: that relies on a participant's past behavior to assess their trustworthiness.

Personality-based trust does not have a significant use in virtual communities where decisions on trust are made algorithmically. Most cognitive factors that form trust in real-life interactions such as the manner of a person, their body language and the intonations of their speech are also absent in virtual communities. However, there are alternative forms of cognition that can be used instead. These are almost invariably based on the virtual identity of the community member. In most *purely* online contexts - as opposed to contexts where the online identity is linked to a non-online identity such as with online banking - there are virtually no costs to creating a new virtual identity. Obtaining an email address on any of the free web-based email services such as Yahoo, Hotmail, Gmail is trivial and with each one comes a new virtual identity. While these services are increasingly adopting strategies such as requiring image reading to counter automated registration spam programs, it is not difficult to create say a dozen accounts in 10 minutes. And as the e-Bay scam case [33] showed, that is all a malicious user needs to surmount simple feedback systems and commit online fraud.

Enforcing trust in virtual communities is hard not only because of the difficulties in recognizing the trustworthiness of participants but also because of the lack of adequate monitoring and appropriate sanctions for dishonest behavior. This lack of *institution-based trust* is because there are not enough trusted third-parties with the power to punish in order to ensure honesty of all the players [4]. This problem is exacerbated in decentralized virtual communities and electronic marketplaces. An organization like e-Bay with a centralized infrastructure can act as the trusted third party to at least store feedback information in a reliable fashion even though the information itself may not be reliable. The problem becomes much harder with decentralized systems such as peer-to-peer networks where the absence of a centralized authority is a defining feature of the system.

Hence, it appears that the best strategy for creating trust in virtual communities is through *transaction-based trust* where feedback on participants' past behavior is collected and aggregated to decide their trustworthiness. A transaction-based trust strategy is far from perfect and suffers from many of the same shortcomings as the other strategies. For instance, the lack of verifiable identities [18] can make a transaction-based system vulnerable to manipulation. However, as many recent proposals have shown [1,11,15,23,32,46] such strategies are not contingent upon a trusted third party and the community as a whole provides the institutional basis for trust. Hence, if the problem of identity can be solved, transaction-based systems are capable of providing the solution for virtual communities.

Transaction-based trust creation strategies are also commonly known as reputation-based trust management systems or just *reputation systems*. Some authors use the term reputation systems narrowly to include only those systems that monitor past transactions of participants to compute their trustworthiness. This information is then shared with other participants who decide whether or not to interact with the target participant on its basis. We use the term in a more general sense to include recommendations systems that recommend items as well as systems that perform distributed authentication by inferring trustworthiness. All these systems have one feature in common. The collection and aggregation of feedback in order to rate objects or people.

Before proceeding further, we would like make a quick note on terminology. The decision-making participants of a virtual community have been referred to in the research literature by many names. These include “individual” or “user” which usually refer to a real person, “agent” that refers to an automated program that decides on the basis of some rules and “peer”, “player” and “actor”, that may refer to both a person and a program etc. In this chapter, we shall use these terms interchangeably as our primary interest is in the decision-making unit and not whether it is human or machine. However, where necessary, we shall specify if the unit being referred to must necessarily be a human being or an automated entity. Research in digital reputation also owes a lot to economics. Hence the terms “game” for the interactions of the virtual community and “co-operation” and “defection” for behaving honestly and cheating are also commonly used in literature.

Reputation systems research lies at the intersection of several disciplines including evolutionary biology [19,24,45], economics (game theory and mechanism design) [3,6,29,34], sociology (social network theory [5,31,40]) and computer science (e-commerce, peer-to-peer systems, cryptography etc.). In this chapter, we survey the approaches taken by researchers from these disciplines to provide the context for digital reputation schemes. We begin by listing the various applications of digital reputation systems. This is followed by a discussion of what motivates the participants of a virtual community to cooperate with each other. We then look at what are the requirements of a good reputation system. This is followed by a more detailed analysis of the components of reputation systems. We classify different types of feedback and their role in constructing reputations. We discuss second-order reputation and the motivations for providing feedback. Reputation modifying factors such as transaction context are looked at next followed by a discussion on how reputations can be interpreted. We conclude by looking at several specific digital reputation systems that have been proposed.

2 Applications of Reputation Management

The main uses of digital reputation management systems in virtual communities:

1. Incentivizing cooperation

2. Identifying and excluding malicious entities

3. Authenticating users in a distributed manner

4. Providing recommendations

Resnick [39] further defines three requirements for a reputation system: 1) to help people decide whom to trust, 2) to encourage trustworthy behavior and appropriate effort and 3) to deter participation by those who are unskilled or dishonest. To this we can add the requirements that a reputation system 4) must preserve anonymity associating a peer's reputation with an opaque identifier and 5) have minimal overhead in terms of computation, storage and infrastructure.

The participation of an individual in a virtual community strongly depends on whether, and how much benefit they expect to derive from their participation. If an individual feels that they will not gain anything from joining the community they are unlikely to participate. Hence, a good distributed system must be designed such that it incentivizes cooperation by making it profitable for users to participate and withhold services from users that do not contribute their resources to the system.

An equally serious challenge to distributed systems and virtual communities comes from users who act in a malicious fashion with the intention of disrupting the system. Examples of such malicious behavior include users who pollute a file-sharing system with mislabeled or corrupted content, nodes that disrupt a peer-to-peer (P2P) routing system to take control of the network, nodes of a mobile ad hoc network (MANET) that misroute packets from other nodes [11] and even spammers [9] who are maliciously attacking the email community. Therefore, a central objective of all digital reputation schemes is to identify such users and punish them or exclude them from the community. This exclusion is usually achieved by allowing members to distinguish between trustworthy and untrustworthy members. An untrustworthy member will not be chosen for future interactions. In contrast with incentivizing cooperation that motivates truth from participants, this is based on punishing falsehoods.

Another use for digital reputation systems is for distributed authentication. Distributed authentication does not rely on strict hierarchies of trust such as those using certification authorities and a centralized public key infrastructure that underpin conventional authentication. Such networks capture trust relationships between entities. Trust is then propagated in the network so that the trust relationship between any two entities in the network can be inferred. PGP [48], GnuPGP and OpenPGP-compatible [13] systems all use webs of trust for authentication. Closely related are virtual social networks including the Friend of a Friend (FOAF) system, LinkedIn, Tribe.net, Orkut, and Friendster that use some or the other form of trust propagation.

Recommendation systems are another application that relies on similar principles as digital reputation systems. Instead of computing the trustworthiness of a participant, recommendation systems compute the recommendation score of objects based on the collective feedback from other users. These systems are in widespread commercial use and examples include the Amazon recommendation system and the IMDB movie recommendations. In recommendation systems, objects instead of members of a virtual community are rated. The members may then be rated on the quality of feedback they provide just like in reputation systems.

When the number of objects to be rated is usually large compared to the user-base, the collected data can be very sparse. Collaborative-filtering based recommendation systems use the collective feedback to try to compute the similarity in the ratings made by any two users and using this similarity to weigh the rating of one user when predicting the corresponding choice for the other user. There are many ways in which this similarity can be computed. One method uses the Pearson's correlation coefficient [38]:

$$w_{u,i} = \frac{\sum_j (v_{u,j} - \bar{v}_u)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{u,j} - \bar{v}_u)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}} \tag{1}$$

where $w_{u,i}$ computes the similarity between users u and i on the basis of their votes for all objects j that both have voted for and $\bar{v}_u$ and $\bar{v}_i$ denote respectively the average ratings given by users u and i.

Another method is to use vector similarity [10]:

$$w_{u,i} = \sum_j \frac{v_{u,j}}{\sqrt{\sum_{k \in I_u} v_{u,k}^2}} \frac{v_{i,j}}{\sqrt{\sum_{k \in I_i} v_{i,k}^2}} \tag{2}$$

where $v_{u,j}$ is the rating given by user u for object j as before and I_u is the set of objects for which user u has a rating. Other model-based methods exist that use Bayesian network models or clustering models to group users together and use group membership to predict what a user may vote for a given object. Clustering models have also been shown to be useful in reputation systems for eliminating spurious ratings [16] made by malicious users in an electronic marketplace.

3 Motivating Cooperation

Motivating co-operation among participants has been a subject of research since long before the first virtual communities were formed. As long as there are shared resources, there will be users who are tempted to use more than their fair share for their own benefit even if it is at the expense of the community as a whole. This conflict between individual interests and the common good is exemplified in the "tragedy of the commons" a term coined by Hardin [28] who used the overgrazing of the English "commons" (property that was shared by the peasants of a village) as an example. In reputation systems research such selfish users are termed as *free-riders* or *free-loaders*.

An example of free-riding can be found in mobile ad hoc networks (MANETs). MANETs function on the premise that nodes will forward packets for each other even though forwarding a packet consumes power. However, if there are free-riders that inject their own packets in the network but do not forward any packets from other nodes, they can exploit the cooperative nature of other users.

Another example of free-riding in virtual communities is found in early file-sharing systems where an individual is allowed to download files without being required to contribute any files for uploading. In their measurement study of Napster and Gnutella, Sariou et al. [43] reported that this resulted in significant fractions of the population indulging in client-like behavior. They reported that a vast majority of modem users were free-riders who did not share any files and only downloaded files offered by others. However, equally interesting is the fact that many users with high bandwidth connections indulged in server-like behavior and offered their own files for sharing but did not download any files at all. This unselfish behavior is the flip-side of free-riding where users in a virtual community indulge in *altruism*.

Altruism has been studied by evolutionary biologists [19, 24, 45] in humans and other primates. These authors have sought to explain altruistic behavior by arguing that it signals "evolutionary fitness". By behaving generously an individual signals that they have "plenty to spare" and is thus a good mating choice. In the context of virtual communities, altruistic behavior is motivated in part by a desire to signal that interacting with the individual is likely to be beneficial in other contexts as well.

Another strand of research comes from economists who have studied co-operation by setting up non-zero sum games and determining the equilibria that result both through analytic game theory and through simulation. In game theory, it is usually assumed that players are "rational" or "selfish", i.e., they are only interested in maximizing the benefit they obtain with no regard to the overall welfare of the community. This is distinct from an "irrational" player whose utility function depends on something more than just their own benefit and whose behavior cannot be predicted. An example of irrational players are "malicious" players who actively wish to harm other players or the game as a whole even if it means reducing their own personal benefit.

The "game" that is typically used for modeling the problem of cooperation is the Prisoner's Dilemma [21, 26]. The classic prisoner's dilemma concerns two suspects who are questioned separately by the police and given a chance to testify against the other. If only one prisoner betrays the other, the betrayer goes free while the loyal prisoner gets a long sentence. If both betray each other they get a medium sentence and if both stay silent they get a small sentence. Hence, if both prisoner's are "selfish" and have no regard for the other prisoner, we see the best strategy for a prisoner is always to betray the second prisoner, regardless of what the other prisoner chooses. If the second prisoner confesses, the first prisoner must confess too otherwise the first prisoner will get a long sentence. And if the second prisoner does not confess, the first prisoner can get off free by confessing as opposed to getting a short sentence by not confessing.

If the game is played only once, the solution is obvious. "Always Defect" is the dominant strategy[1]. Axelrod [6] studied an interesting extension to the classic problem which he called the iterated prisoner's dilemma. Here, members of a community play against each other repeatedly and retain memory of their past interactions. Hence, they have an opportunity to punish players for their past defections. Axelrod set up an experiment with various strategies submitted by fellow researchers playing against each other. He discovered that "greedy" strategies tended to do very poorly in the long run and were outperformed by more "altruistic" strategies, as judged by pure self-interest. This was true as long as cheating was not tolerated indefinitely. By analyzing the top-scoring strategies, Axelrod stated several conditions necessary for a strategy to be successful. A successful strategy should be 1) *Nice*: it will not defect before an opponent does. 2) *Retaliating*: it

[1] A dominant strategy always give better payoff than another strategy regardless of what other players are doing.

will always retaliate when cheated. A blind optimist strategy like “Always Cooperate” does not do well. 3) *Forgiving*: in order to stop endless cycles of revenge and counter-revenge a strategy will start cooperating if an opponent stops cheating. 4) *Non-envious*: a strategy will not try to outscore an opponent. Axelrod found that the best deterministic strategy was “tit-for-tat” in which a player behaved with its partner in the same way as the partner had behaved in the previous round.

The prisoner’s dilemma is a game with a specific reward function that encourages altruistic behavior and discourages selfish behavior. Let T be the temptation to defect, R be the reward for mutual cooperation, P be the Punishment for mutual defection and S be the Sucker’s punishment for cooperating while the other defects. Then, the following inequality must hold:

$$T > R > P > S \tag{3}$$

In the iterated game, yet another inequality must hold:

$$T + S < 2R \tag{4}$$

If this is not the case, then two players will gain more in the long run if one cooperates and the other defects alternately rather than when both cooperate. Hence, there will be no incentive for mutual cooperation. Recall that the objective of a rational player is to maximize their individual score and not score more than the opponent.

Other virtual communities may have different cost and reward functions. The game-theoretic approach is to devise strategies to maximize individual utility in a fixed game. However, if a community is designed so that its interests as a whole are not aligned with those of an individual, the system is will be “gamed” by the rational users leading to the ultimate failure of the virtual community. Mechanism design [29] deals with the design of systems such that players’ selfish behavior results in the desired system-wide goals. This is achieved by constructing cost and reward functions that encourage “correct” behavior.

Another example of incentivizing cooperation can be found in collaborative content distribution mechanisms [2,8,12] such as BitTorrent [14] where peers cooperate to download a file from a server by downloading different chunks in parallel and then exchanging them with each other to reconstruct the entire file. In fact, BitTorrent implements a “tit-for-tat” strategy to prevent free-riding. Basic reputation schemes have also been implemented in second generation file-sharing systems such as Kazaa that measure the participation level of a peer based on the number of files the peer has uploaded and downloaded. Peers with a higher participation level are given preference when downloading a file from the same source.

4 Design Requirements for a Reputation System

While mechanism design may help achieve overall system goals through proper incentivization, it is ineffective against deliberately malicious (or irrational) participants. To solve this problem a number of reputation systems have emerged. These operate by allowing a user to rate the transactions they have had with other users. This feedback is collected and aggregated to form a reputation value which denotes the trustworthiness of a user. This reputation information is made available to requesting users who then make their decisions on whether or not to interact with a given user based on its reputation. Several competing schemes for aggregating this feedback have been proposed [1, 15, 17, 23, 32, 46, 47].

The architectural and implementation details of the aggregation mechanism depend on the underlying network on which the virtual community is based. When the community is built on top of a traditional client-server network, a trusted third party exists which can be relied on to collect and aggregate opinions to form a global view. e-Bay feedback, Amazon customer review and the Slashdot distributed moderation systems are all examples where feedback from users is stored in a centralized trust database. The aggregation is performed in this centralized database and all users have access to the global reputations thus computed. When the community is built on top of a P2P network, the challenges of managing feedback become much harder. There is no centralized, trusted, reliable, always-on database and the collection, storage, aggregation and dispersal of trust information must be done in a distributed way. Relying on third parties for storage and dissemination also makes the system vulnerable to tampering and falsification of trust information in storage and transit. Moreover, the system must also provide redundancy because users may drop out of the network at any time.

Hence, there are two separate but interrelated challenges that must be overcome by any distributed trust management system. The first is the choice of an appropriate trust metric that accurately reflects the trustworthiness of users and is resistant to tampering and other attacks. This was recognized by Aberer and Despotovic as "the semantic question: which is the model that allows us to assess trust [of an agent based on that agents behavior and the global behavior standards]" [1]. The second is designing a system architecture that is robust against the challenges mentioned above or the "data management" problem.

At this point, it is useful to ask why a reputation system would work in a virtual community and what may cause it to fail. The reasons for potential success are: 1) The costs of providing and distributing reputations are negligible or zero. 2) The infrastructure for decentralized aggregation and dissemination already exists in the form DHT-based[2] routing systems [37, 44]. 3) It is easy to build redundancy in the reputation system. And the potential failings are: 1) Users may lie about the feedback they provide. 2) Users may not bother to give feedback. 3) Untrustworthy users may mask their behavior or retaliate against negative feedback by sending negative feedback about other users. 4) Users may try to game the system by building up their reputation by acting honestly over several small transactions followed by cheating in a large transaction in a process known as *milking*. 5) Users may form malicious groups that give false positive ratings to each other in order to boost each others reputations. 6) Users may re-enter the system with a new identity to conceal their past behavior. Hence, a good reputation system must try to overcome these potential failings.

[2] In systems using distributed hash tables or DHTs, objects are associated with a key (usually produced by hashing the object ID such as filename) and each node in the system is responsible for all objects associated with a key that falls in a certain range.

5 Analysis of Reputation System Components

5.1 Direct vs. Indirect Evidence

Two distinct types of evidence are usually combined by reputation systems to form the reputation of a user. These are 1) *direct evidence*, which consists of a user's first-hand experiences of another user's behavior and 2) *indirect evidence* which is second-hand information that is received from other nodes about their own experiences. If only first-hand information were used to decide the reputation of another peer, the sparseness of feedback would be a problem because in large virtual communities the interaction matrix is usually very sparse. A user would not be able to make a trust judgment on another user with whom they have never interacted before. Hence, users must rely on indirect evidence to compute the reputation of users that are new to them.

In a centralized system such as e-Bay all indirect evidence is collected at the central trust database and is made available to other users. In a decentralized system indirect evidence can be shared in a number of ways. The interested user may ask for indirect evidence from its neighbors or other users it trusts as in [15]. The indirect evidence may also be propagated to all users in the network using a recursive mechanism like in [32]. *Designated Agents*[3] may also be chosen from within the community to store indirect evidence which can then be furnished to requesting users [1,23,46]. In the latter schemes, designated agents responsible for specific users in the system are chosen using distributed hash tables. Multiple such agents are chosen to ensure redundancy in case an agent leaves the network or tries to falsify this information.

5.2 Second-Order Reputation

Using second-hand information in a decentralized system leads to another problem. How can a user know that the provider of the second-hand information is telling the truth?

[3] The term *Designated Agents* was coined by the authors and includes all systems where one or more users are made responsible for storing and sharing another user's reputation information by the system.

We can think of an individual's reputation for providing accurate reputation information on other as their second-order reputation information. Second-order reputation is often termed as *credibility* as it measures the truthfulness of an individual as a provider of information.

The problem was first addressed by Aberer and Despotovic [1] who were among the first to use a decentralized reputation storage system. They use the trustworthiness of an agent (first-order reputation), to decide whether to take its feedback into account. The feedback from agents who are deemed trustworthy are included and that from untrustworthy agents is excluded. Kamvar et al. [32] also use the same strategy of using first-order reputation as second-order reputation as well. However, while it is reasonable to assume that an individual who cheats in the main market cannot be relied upon for accurate feedback, the reverse is not necessarily true. An individual can act honestly in the main market and enjoy a high reputation and at the same time provide false feedback to lower the reputation of others in the community who are after all his/her competitors. An example where such a strategy would be advantageous is a hotel rating system where a hotel may provide very good service and thus have a high reputation but at the same time may give false feedback about its competitors to lower their reputation. Hence its credibility is very different from its reputation.

The solution is to recognize credibility as different from first-order reputation and compute it separately. Credibility can be computed in several different ways. Credibility values can be expressed explicitly and solicited separately from reputation values. However this leads to a problem of endless recursion. To ensure that users do not lie about other users credibility, we need third-order reputation to measure an individual's reputation for providing accurate credibility information and so on.

Credibility can also be computed implicitly. This can be done in two ways. In the first method, inspired by collaborative filtering, the similarity of user j's opinions to those of user i are used to compute the credibility of user j in the eyes of user i (C_{ij}). This can be done using Pearson's correlation coefficients as used by Resnick et al. [38] or by using vector similarity like in Breese et al. [10]. PeerTrust [46] uses yet another similarity metric that resembles the Pearson coefficient. In this method if user i's opinions are often at variance with those of user j, i will have a lower credibility value for j and vice versa. The second approach for implicit credibility computation measures the credibility of a user by taking into account the agreement between the feedback furnished by the user and the views of the community as a whole. The community average view can be represented by the reputation value computed from the feedback from all the reporting users. If a user gives wrong feedback about other users, that is his or her feedback is very different from the eventual reputation value for computed, its credibility rating is decreased and its subsequent reports have a reduced impact on the reputation of another user. This method of computing credibility has an advantage in that it is more scalable and it can operate in decentralized systems where a complete record of all opinions expressed by other individuals may not be available due to privacy concerns. However, there is a danger of "group-think" in such systems. There is strong encouragement for an individual to agree with the opinion of the group as a whole as disagreements are punished by lowering the credibility of the individual who disagrees.

5.3 Motivation for Providing Feedback

A closely related issue is how to motivate community members to provide feedback to others when providing feedback consumes their own resources. In some respects this problem is the same as that of motivating cooperation between community members as discussed above. However, there are some important differences. In a designated agent system, designated agents must use their own resources to store, compute and report reputation values for other members. Most literature assumes that agents will perform this task because they are "altruistic" or because they too will benefit from a designated agent system. However, in a large network, there is little incentive for a particular individual to expend resources to maintain the reputation system. How do we prevent such an individual from free-riding the reputation system itself? Moreover, from a game theoretic perspective, not reporting feedback may be advantageous to an agent in a competitive situation. By reporting feedback to other agents, it is sharing information with them which if kept to itself may have given it an advantage.

One strategy to encourage truthful reputation reports is to create a side-market for reputation where accurate reputation reports are rewarded with currency that can be traded for accurate reports from others [30]. Such a system needs to be structured in such a way that providing more feedback and providing honest feedback results in more credit. However this solution suffers from the recursion problem mentioned above as third-order market would need to be created and so on.

An alternative way to avoid the recursion problem is to incorporate the credibility of an individual in the reputation system as a separate variable as discussed in the previous section. If the credibility is computed through direct evidence only and is not shared with others as in [23] then the recursion problem can be avoided.

An individual with a high credibility will be listened to by other participants in the virtual community with more attention and can thus be said to have greater influence on the community. Hence, selfish individuals may adopt a dynamic strategy in which they adjust their behavior on the basis of the credibility (and thus influence) of the partner they are about to transact with. Here, an individual will behave honestly with another individual who has high credibility and will cheat with an individual with low influence. An adverse report from the low influence individual will carry less weight and not have an adverse impact on the reputation of the first individual, while the opposite is true of the high influence individual.

This provides an incentive for individuals to give truthful reputation reports for others. Accurate reports result in higher influence which in turn would result in a higher likelihood of cooperation by the transaction parters. This incentivization does not come for free. It depends on two conditions: 1) individuals will behave dynamically, i.e., users must cheat with a different probability based on the influence of their transaction partners, and 2) credibility information is available globally. When these conditions are not met, this subtle form of incentivization is absent.

An additional problem introduced when individuals adopt dynamic startegies. A single reputation value can no longer predict the behavior of an individual at all times as their behavior will depend on the credibility of their interaction partner. As a result, such scenarios have not been the subject of much research.

Therefore, there seems to be only one solution in that an individual's behavior in the main market and in the secondary reputation market should both influence a unified reputation value. And since this unified reputation value is public, it incentivizes truthful reporting. However, the question of how much relative weight to give to the two types of market transactions remains unresolved. If too much weight is given to the main market there is not enough disincentive to make false reports. On the other hand, if not enough weight is given to the main market, an individual can be a truthful reporter but cheat in the main market without being punished sufficiently.

5.4 Motivation is not a Problem: Dissenting Views

A number of authors do not agree that motivation (of cooperation and of sharing feedback) is a problem at all. Fehr and Gächter [19] develop a theory of "altruistic punishment". They designed an experiment that excluded all explanations for cooperation other than that of altruistic punishment. They designed a McCabe style investment game [7] where players could punish their partners if they wished. However, punishing was costly both to the punisher (1 point) and the punished (3 points). Punishment induced cooperation from potential non-cooperators thus increasing the benefit to the group as a whole even though it was costly to the punisher. For this reason punishment was an altruistic act. They concluded that negative emotions are the cause of altruistic punishment and that there is a tendency in humans to punish those that deviate from social norms.

Similarly, Gintis et al. [24] argue that behaving in an altruistic fashion sends a signal to other participants that interacting with that individual is likely to prove beneficial. This argument cuts at the heart of the game theoretic notion of how rational agents operate. Rational agent behavior now becomes probabilistic in that an agent may act in an altruistic fashion in the hope of future reward instead of only interacting when an immediate benefit is expected. This interpretation also depends on whether the individuals in the system are humans or are automated agents that do not share the “altruistic” characteristics and behave in a strictly rational sense.

This has been used as evidence that altruism serves an agent’s self-interest. It also explains why greedy strategy are outperformed by more altruistic strategies in Axelrod’s experiment. Hence, there is some theoretical basis to the claim that a virtual community can be be self-correcting and will exclude bad participants.

5.5 Other Design Issues

A number of other design choices must be made in a reputation system.

Reputation Context : There has been some research on whether reputation is contextual. It is often assumed that reputation is heavily dependent on context. This point of view was aptly expressed by Mui et al. [35]

> “Reputation is clearly a context-dependent quantity. For example, one’s reputation as a computer scientist should have no influence on his or her reputation as cook.”

However, in a contrary argument Gintis et al. [24] suggest that compartmentalizing reputation too strictly can have a negative effect. Their contention is that altruistic behavior is motivated in part by a desire to signal that interacting with the individual in question will be beneficial in other contexts as well. They argue that the notion of reputation in a real world is far more fuzzy and incorporates generosity as well. A generous participant is more likely to be honest as well. This gives participants an incentive to behave in an altruistic fashion in addition to behaving in an honest fashion.

Transaction Value : A reputation system must also be able to distinguish between small and large transactions. Buying a pencil online is not the same as buying a car. If all transactions are rated equally, an individual may exploit the system through a strategy called *milking* where they act honestly in a number of small transactions to build their reputation and then cheat in a large transaction without hurting their reputation too much. In PeerTrust [46] this problem is solved by incorporating a transaction context factor that weighs the feedback according to size, importance and the recency of the transaction.

Interpreting Reputation : Once the reputation of a user (or object in case of recommendation systems) has been computed, it can be used in several ways depending on the application context. In a file-sharing system [15] a peer may choose the peer with the highest reputation to download the file from. On Amazon, the recommendation system may prompt one to buy a book that one was not aware of before. GroupLens [38] helps decide which movie a user decides to watch.

Equally common are applications that demand a Boolean yes/no decision on “Should i trust j?”. There are several methods by which this translation from an arbitrary range of reputation values to a binary value can be achieved. These include 1) using a deterministic threshold (I will trust you if your reputation value exceeds 6 on a scale of 10), 2) relative ranking (I will trust you if your reputation is in the top 10% of community members), 3) probabilistic thresholds (the probability that I trust you is a monotonic function on the range of possible trust values) and 4) majority rounding (I will trust you if I trust a majority of people with reputation values close to your reputation).

Another approach [40, 42] is to include information that allows a user to decide how much faith it should place in the reputation value. On e-Bay this information is the number of feedbacks a user has received. A user with a high reputation and a large number of feedbacks is much more trustworthy than one with a low number of feedbacks.

Benefits of High Reputation : Many reputation systems particularly those proposed for file-sharing applications [1,15,32] do not consider the consequences of a peer having a high reputation. In file-sharing systems the most reputable peer in the network will be swamped with requests as all peers are going to want to download the resources from it. Hence peers with high reputations are “punished” for their high reputations by having to expend more of their resources to serve others instead of being benefited from their reputation. In this scenario the interests of the system are not aligned with that of individuals and the individual peers have no motivation to act honestly and thus increase their reputation.

In order to motivate individuals to try and acquire high reputations, there needs to be a mechanism by which nodes that have a high reputation are rewarded. In a file-sharing application this could be achieved through preferential access for higher reputation nodes to resources at other nodes.

Positive vs. Negative Feedback : A reputation system may be based solely on either positive feedback, negative feedback or a combination of both. The disadvantage of a negative feedback only system [1] is that each new entrant into the system has the highest possible reputation. Hence a misbehaving individual may create a new identity and shed their bad reputations to start afresh. Using only positive feedback, on the other hand, makes it hard to distinguish between a new user and a dishonest user. If old users choose not to interact with any users without a minimum level of positive feedback, a new user may thus find itself frzen out of the group and interacting only with malicious users.

Identities : A reputation system that allows unlimited creation of new identities is vulnerable to manipulation [18]. Hence there must be a cost of a new identities. However, at the same time this cost must not be so large as to discourage newcomers from joining. Friedman and Resnick argue that in social situations there is inevitably some form of initiation dues [20]. They also find that while these dues are inefficient, especially when there are many newcomers to a community, no other strategy does substantially better in terms of group utility.

6 Some Reputation Systems

We now look at some reputation management algorithms that have been proposed in recent years.

6.1 Complaints-Based Trust

One of the first reputation management algorithms for the peer-to-peer systems was proposed by Aberer and Despotovic in [1]. This system is based solely on negative feedback given by peers when they are not satisfied by the files received from another peer. The system works on the assumption that a low probability of cheating in a community makes it harder to hide malicious behavior.

Let P denote the set of all peers in the network and B be the behavioral data consisting of trust observations $t(q, p)$ made by a peer $q \in P$ when it interacts with a peer $p \in P$. We can assess the behavioral data of a specific peer p based on the set

$$B(p) = \{t(p, q) \textit{ or } t(q, p) \mid q \in P\} \tag{5}$$

In this manner, the behavior of a peer takes into account not only all reports made *about* p but also all reports made *by* p. In a decentralized system a peer q does not have access to the global data $B(p)$ and B. Hence it relies on direct evidence as well as indirect evidence from a limited number of witnesses $r \in W_q \subset P$:

$$B_q(p) = \{t(q, p) \mid t(q, p) \in B\} \tag{6}$$

$$W_q(p) = \{t(r, p) \mid t(r, p) \in B \wedge r \in P\} \tag{7}$$

However, the witness r itself may be malicious and give false evidence. Aberer and Despotovic assume that peers only lie to cover their own bad behavior. If peer p is malicious and cheats peer q, q will file a complaint against it. If p also files a complaint against q at the same time it could be difficult to find out who is the malicious peer. However, if p keeps acting maliciously it will become easy to detect it since there will be a lot of complaints filed from peer r about a set of good peers and a lot of complaints filed from these good peers all about peer p. Based on this, the reputation of p can be calculated as:

$$T(p) = |\{c(p,q) \mid q \in P\}| \times |\{c(q,p) \mid q \in P\}| \tag{8}$$

Aberer and Despotovic proposed a decentralized storage system called *P-Grid* to store reputation information. Each peer p can file a complaint about another peer q at any time as follows:

$$insert(a_i, c(p,q), key(p)) \; and \; insert(a_j, c(p,q), key(q))$$

where a_i and a_j are two arbitrary agents. Insertions are made on the keys of both p and q since the system stores complaints both by and about a given peer.

Assuming that an agent is malicious with probability π and that an error rate of ε is tolerable, then a peer p will need to receive r replicas of the same data satisfying $\pi^r < \varepsilon$ to ensure that the error rate is not exceeded. If a simple majority rule is employed, the total number of queries for each trust verification will never exceed $2r + 1$.

A peer p making s queries will obtain a set

$$W = \{(cr_i(q), cf_i(q), f_i, a_i) \mid i = 1 \ldots w\}$$

where w is the number of different witnesses found, a_i is the identifier of the i-th witness, f_i is the frequency with which witness a_i is found and $s = \sum_{i=1}^{w} f_i$. $cr_i(q)$ and $cf_i(q)$ are the complaints about q and filed by q respectively as reported by witness a_i. Different witnesses are found with different frequencies and the ones which are found less frequently have probably been found less frequently also when complaints were filed. So it is necessary to normalize $cr_i(q)$ and $cf_i(q)$ using frequency f_i in this way:[4]

$$cr_i^{norm}(q) = cr_i(q) \left(\frac{s - f_i}{s}\right)^s \tag{9}$$

$$cf_i^{norm}(q) = cf_i(q) \left(\frac{s - f_i}{s}\right)^s \tag{10}$$

[4] Note that the equation below corrects the one presented in the original paper

Each peer p can keep a statistics of the average number of complaints filed (cf_p^{avg}) and received (cr_p^{avg}) and can determine if a peer q is trustworthy basing on the information returned from an agent i using this algorithm:

Algorithm 1: Trust Assessment Using Complaints

```
$decide(cr_i^{norm}(q), cf_i^{norm}(q)) =$
if
    $cr_i^{norm}(q) cf_i^{norm}(q) \leq \left(\frac{1}{2} + \frac{4}{\sqrt{cr_p^{avg} cf_p^{avg}}}\right)^2 cr_p^{avg} cf_p^{avg}$
then return 1; else return -1.
```

This algorithm assumes that if the total number of complaints received exceeds the average number of complaints by a large amount, the agent must be malicious.

6.2 EigenTrust

Kamvar et al. [32] presented a distributed algorithm for the computation of the trust values of all peers in the network. Their algorithm is inspired by the PageRank algorithm used by Google and assumes that trust is transitive. A user weighs the trust ratings it receives from other users by the trust it places in the reporting users themselves. Global trust values are then computed in a distributed fashion by updating the trust vector at each peer using the trust vectors of neighboring peers. They show that trust values asymptotically approach the eigenvalue of the trust matrix, conditional on the presence of pre-trusted users that are always trusted.

A peer i may rate each transaction with peer j as positive ($tr(i,j) = 1$) or negative ($tr(i,j) = -1$).The local trust value at i for j can then be computed by summing the ratings of individual transactions:

$$s_{ij} = \sum tr_{(}i, j) \tag{11}$$

Local trust values are normalized as follows:

$$c_{ij} = \frac{\max(s_{ij}, 0)}{\sum_l \max(s_{il}, 0)} \tag{12}$$

to keep them between 0 and 1.

In order to aggregate normalized local trust values, trust values furnished by acquaintances are weighted by the trust a peer has in the acquaintances themselves:

$$t_{ik} = \sum_j c_{ij} c_{jk} \qquad \text{(note that } \textstyle\sum_j t_{ij} = 1\text{)} \tag{13}$$

where t_{ik} is the trust a peer i places in peer k based on the opinion of its acquaintances. If we define $\overrightarrow{t_i}$ to be the vector containing the values t_{ik} and C to be the matrix containing the values c_{ij}, we get $\overrightarrow{t_i} = C^T \ \overrightarrow{c_i}$. The preceding expression only takes into account opinions of a peer's acquaintances. In order to get a wider view a peer may ask its friends' friends and so on:

$$\overrightarrow{t} = \left(C^T\right)^n \overrightarrow{c_i} \tag{14}$$

Kamvar et al. show that if n is large, peer i can have a complete view of the network and the trust vector $\overrightarrow{t_i}$ will converge to the same vector for every peer i (the **left principal eigenvector of** C), conditional on C being irreducible and aperiodic.

They further add three practical issues to this simple algorithm. If there are peers that can be trusted *a priori* the algorithm can be modified to take advantage of this. They define p_i as $\frac{1}{|P|}$ (where P is the set containing the pre-trusted peers) if i is a pre-trusted peer, and 0 otherwise. In presence of malicious peers, using an initial trust vector of $\overrightarrow{p}$ instead of if replace $\overrightarrow{e}$ generally ensures faster convergence. If a peer i has never had any interaction with other peers, instead of being left undefined, c_{ij} can be defined as:

$$c_{ij} = \begin{cases} \frac{\max\left(local_{ij}, 0\right)}{\sum_l \max\left(local_{il}, 0\right)} & \text{if } \sum_l \max\left(local_{il}, 0\right) \neq 0 \\ p_j & \text{otherwise} \end{cases} \tag{15}$$

In order to prevent malicious collectives from subverting the system, Kamvar et al. further modify the trust vector to:

$$\overrightarrow{t}^{k+1} = (1-a) C^T \ \overrightarrow{t}^{k} + a \ \overrightarrow{p} \tag{16}$$

where a is some constant less than 1. This ensures that at each iteration some of the trust must be placed in the set of pre-trusted peers thus reducing the impact of the malicious collective.

Hence, given A_i the set of peers which have downloaded files from peer i and B_i the set of peers from which peer i has downloaded files, each peer i executes the following algorithm:

Algorithm 2: Distributed EigenTrust

```
Query all peers $j \in A_i$ for $c_{ji}t^{(0)} = c_{ji}p_j$;
repeat
$t_i^{(k+1)} = (1-a)\left(c_{1i}t_1^{(k)} + c_{2i}t_2^{(k)} + \ldots + c_{ni}t_n^{(k)}\right) + ap_i$;
send $c_{ij}t_i^{(k+1)}$ to all peers $j \in B_i$;
wait for $c_{ji}t_j^{(k+1)}$ from all peers $j \in A_i$;
$\delta = \left|t^{(k+1)} - t^{(k)}\right|$;
until $\delta < \varepsilon$;
```

Each peer thus obtains the same global trust value matrix for all other peers in the network. Kamvar et al. further describe a DHT-based solution to anonymously store multiple copies of the trust value for a given peer at several *score managers*. This eliminates the problem caused by malicious peers reporting false trust values for themselves to other peers in order to subvert the system.

6.3 PeerTrust

In PeerTrust [46], Xiong and Liu define five factors used to compute the trustworthiness of a peer. These are: 1) feedback obtained from other peers, 2) scope of feedback such as number of transactions, 3) credibility of feedback source, 4) transaction context factor to differentiate between mission-critical and non-critical transactions and 5) community context factor for addressing community related characteristics and vulnerabilities.

Given a recent time window, let $I(u,v)$ denote the total number of transactions between peer u and v, $I(u)$ denote the total number of transactions of peer u, $p(u,i)$ denote peer u's partner in its i^{th} transaction, $S(u,i)$ denote the normalized amount of satisfaction peer $p(u,i)$ receives from u in this transaction, $Cr(v)$ denote the credibility of peer v,

$TF(u,i)$ denote the adaptive transaction context factor for peer u's i^{th} transaction and $CF(u)$ denote u's community context factor. Then, the trust value of peer u denoted by T(u) is:

$$T(u) = \alpha * \sum_{i=1}^{I(u)} S(u,i) * Cr(p(u,i) * TF(u,i) + \beta * CF(u) \tag{17}$$

where α and β are weight factors.

In their experimental study, Xiong and Liu turn off the transaction context factor and the community context factor and use two credibility metrics. The first is based on the trust value of the reporting peer (similar to EigenTrust) while the second is based on the similarity between the reporting peer and the recipient of the trust information. They further propose using a PKI-based scheme and data replication to increase the security and reliability of their system.

6.4 ROCQ

Garg et al. [22, 23] proposed a scheme that combines local opinion, credibility of the reporter and the quality of feedback to compute the reputation of a peer in the system. In their scheme direct evidence in the form of local opinion is reported to score managers which are chosen using distributed hash tables.

The reputation R_{mj} of user j at score manager m is:

$$R_{mj} = \frac{\sum_i O_{ij}^{avg} \cdot C_{mi} \cdot Q_{ij}}{\sum_i C_{mi} \cdot Q_{ij}} \tag{18}$$

where C_{mi} is the credibility of user i according to user m, O_{ij}^{avg} is i's average opinion of j and Q_{ij} is the associated quality value reported by i.

The quality value of an opinion depends on the number of transactions on which the opinion is based and the consistency with which the transaction partner has acted. Thus an opinion is of greater quality when the number of observations on which it is based is larger and when the interactions have been consistent (resulting in a smaller variance). When the number of observations is high but they do not agree with each other, the quality value is lower.

The credibility of a user is based on direct evidence only and is not shared with other users. This prevents the recursion problem of calculating the third-order reputation and so on. The credibility is based upon the agreement of the reported opinion with the group consensus as reflected in the reputation value and is updated after every report received. The precise formula for adjusting the credibility of user i by user m is:

$$C_{mi}^{k+1} = \begin{cases} C_{mi}^{k} + \frac{(1-C_{mi}^{k})\cdot Q_{ij}}{2} \cdot (1 - \frac{|R_{mj} - O_{ij}^{avg}|}{s_{mj}}) \\ \qquad \text{if } |R_{mj} - O_{ij}^{avg}| < s_{mj} \\ C_{mi}^{k} - \frac{C_{mi}^{k}\cdot Q_{ij}}{2} \cdot (1 - \frac{s_{mj}}{|R_{mj} - O_{ij}^{avg}|}) \\ \qquad \text{if } |R_{mj} - O_{ij}^{avg}| > s_{mj} \end{cases} \tag{19}$$

where C_{mi}^{k} is the credibility of user i after k reports to user m, O_{ij}^{avg} is the opinion being currently reported by user i, Q_{ij} is the associated quality value, R_{mj} is the aggregated reputation value that user m computed for j and s_{mj} is the standard deviation of all the reported opinions about user j. In this way, if a reporting user is malicious, its credibility rating is gradually reduced since its opinion does not match that of the community as a whole.

The authors also propose combining both direct and indirect evidence to create reputation. They proposed a threshold number of interactions between two users below which users will rely on the global reputation of their prospective partner and above which they would rely on first hand evidence only. This eliminates the problem of sparsity of data while at the same time allowing for reputation to be tailored according to personal experience.

6.5 Propagation of Trust and Distrust

A number of mathematical approaches to propagating trust [25,41] and distrust [27] have been proposed. In particular, Guha et al. gave a number of models of atomic (single-step) propagation. Let B be a belief matrix whose ij^{th} element signifies i's belief in j. B can be composed of either the trust matrix (T_{ij} is i's trust in j) or both the trust and the distrust (D_{ij} is i's distrust in j) matrices (say $B_{ij} = T_{ij} - D_{ij}$). Then, atomic propagation of trust takes place by: 1) *direct propagation* (matrix operator[5]: **B**) : assumes that trust is *transitive* so if i trusts j and j trusts k then we can infer that i trusts k, 2) *co-citation* ($B^T B$): if both i and j trust k and if i trusts l, then j also trusts l, 3) *transpose trust* (B^T): assumes that trust is *reflexive* so that if i trusts j then trusting j should imply trusting i, and 4) *trust coupling* (BB^T): if i and j trust k, then trusting i should also imply trusting j. They then go on to propagate trust using a combined matrix that gives weights to the four propagation schemes:

$$C_{B,\alpha} = \alpha_1 B + \alpha_2 B^T B + \alpha_3 B^T + \alpha_4 B B^T \tag{20}$$

Thus, applying the atomic propagations a fixed number of times, new beliefs can be computed. In a limited set of experiments, they study the prediction performance of their algorithm on real data and find that the best performance comes with one-step propagation of distrust while trust can be propagated repeatedly.

7 Conclusions and Future Work

The area of reputation systems remains fertile for future research. Initial research in this area has focused on applying lessons from diverse fields such as evolutionary biology and game theory to computer science. In particular game-theoretic models such as the iterated prisoner's dilemma were adapted to virtual communities. However this analysis is limited by not considering the presence of irrational (malicious) players in the community.

Simultaneously, several new reputation schemes have been proposed. These schemes typically propose a new trust model followed by experimental simulation. However it is difficult to compare the schemes side-by-side as each scheme makes its own assumptions about the interaction model, modes of maliciousness and levels of collusion among malicious peers, not to mention widely varying experimental parameters such as the number of peers in the system and the proportion of malicious peers.

[5] The matrix operator when applied to a belief matrix would yield a new matrix indicating inferred trust.

More recently, there has been some work on analyzing systems in the presence of malicious peers. For instance, Mundinger and Le Boudec [36] analyze the robustness of their reputation system in the presence of liars and try to find the critical phase transition point where liars start impacting the system.

As the interest in virtual communities, particularly self-organizing communities grows, we are likely to see a lot more research on the various facets of this topic.

References

[1] K. Aberer and Z. Despotovic. Managing trust in a peer-2-peer information system. In *CIKM*, pages 310–317, 2001.

[2] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46(4):1204–1216, July 2000.

[3] G.A. Akerlof. The market for ’lemons’: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

[4] Y. Atif. Building trust in e-commerce. *IEEE Internet Computing*, 6(1), 2002.

[5] C. Avery, P. Resnick, and R. Zeckhauser. The market for evaluations. *American Economic Review*, 89(3):564–584, 1999.

[6] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[7] J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity and social history. *Games and Economic Behavior*, 10:122–42, 1995.

[8] E. Biersack, P. Rodriguez, and P. Felber. Performance analysis of peer-to-peer networks for file distribution. In *The Fifth International Workshop on Quality of Future Internet Services (QofIS’04)*, Barcelona, September 29 – October 1 2004.

[9] P.O. Boykin and V. Roychowdhury. Personal email networks: An effective anti-spam tool. *IEEE Computer*, 38(4):61–68, April 2005.

[10] J. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Uncertainty in Artificial Intelligence. Proceedings of the Fourteenth Conference (1998)*, pages 43–52, 1998.

[11] S. Buchegger and J-Y. Le Boudec. A robust reputation system for P2P and mobile ad-hoc networks. In *Proceedings of the Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.

[12] J. W. Byers, J. Considine, M. Mitzenmacher, and S. Rost. Informed content delivery across adaptive overlay networks. *IEEE/ACM Transactions on Networking*, 12(5):767–780, October 2004.

[13] J. Callas, L. Donnerhacke, H. Finney, and R. Thayer. RFC 2440 - OpenPGP message format, November 1998.

[14] B. Cohen. Incentives build robustness in BitTorrent. In *1st Workshop on Economics of Peer-to-Peer Systems*, Berkeley, CA, USA, June 5–6 2003.

[15] E. Damiani, S. De Capitani di Vimercati, S. Paraboschi, and P. Samarati. Managing and sharing servents' reputations in p2p systems. *IEEE Transactions on Data and Knowledge Engineering*, 15(4):840–854, Jul./Aug. 2003.

[16] C. Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *Proceedings of the 2nd ACM conference on Electronic commerce*, pages 150–157. ACM Press, 2000.

[17] C. Dellarocas. Mechanisms for coping with unfair ratings and discriminatory behavior in online reputation reporting systems. In *ICIS '00: Proceedings of the twenty first international conference on Information systems*, pages 520–525, Atlanta, GA, USA, 2000. Association for Information Systems.

[18] J. Douceur. The Sybil Attack. In *Proceedings of the 1st International Peer To Peer Systems Workshop (IPTPS 2002)*, Cambridge, Massachusetts, March 2002.

[19] E. Fehr and S. Gächter. Altruistic punishment in humans. *Nature*, 415(6868):137–40, January 2002.

[20] E. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(1), 2001.

[21] D. Fudenburg and J. Tirole. *Game Theory*. MIT Press, Cambridge, MA, 1991.

[22] A. Garg, R. Battiti, and R. Cascella. Reputation management: Experiments on the robustness of ROCQ. In *Proceedings of the 7th International Symposium on Autonomous Decentralized Systems (First International Workshop on Autonomic Communication for Evolvable Next Generation Networks)*, pages 725–730, Chengdu, China, April 2005.

[23] A. Garg, R. Battiti, and G. Costanzi. Dynamic self-management of autonomic systems: The reputation, quality and credibility (RQC) scheme. In *The 1st IFIP TC6 WG6.6 International Workshop on Autonomic Communication (WAC 2004)*, October 2004.

[24] H. Gintis, E.A. Smith, and S. Bowles. Costly signalling and cooperation. *Journal of Theoretical Biology*, 213:103–119, 2001.

[25] J. Golbeck, B. Parsia, and J. Hendler. Trust networks on the semantic web. In *Proceedings of Cooperative Intelligent Agents*, Helsinki, Finland, 2003.

[26] M. Granovetter. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, 91:481–510, 1985.

[27] R. Guha, P. Raghavan, R. Kumar, and A. Tomkins. Propagation of trust and distrust. In *Proceedings of WWW 2004*, pages 403–412, New York, NY, USA, May 17–22 2004.

[28] G. Hardin. The tragedy of the commons. *Science*, 162:1243–48, 1968.

[29] M. Jackson. Mechanism theory, 2000.

[30] R. Jurca and B. Faltings. An incentive compatible reputation mechanism. In *Proceedings of the IEEE Conference on E-Commerce*, June 24-27 2003.

[31] S.M. Kakade, M. Kearns, L.E. Ortiz, Robin Pemantle, and Siddharth Suri. Economic properties of social networks. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA, 2005.

[32] S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in P2P networks. In *Proceedings of the twelfth international conference on World Wide Web*, pages 640–651. ACM Press, 2003.

[33] S. Kirsner. Catch me if you can. *Fast Company*, http://www.fastcompany.com/magazine/73/kirsner.html, Retrieved August 13th, 2005.

[34] R. Morselli, J. Katz, and B. Bhattacharjee. A game-theoretic framework for analyzing trust-inference protocols. In *Proceedings of the Second Workshop on the Economics of Peer-to-Peer Systems*, 2004.

[35] L. Mui, M. Mohtashemi, and A. Halberstadt. Notions of reputation in multi-agents systems: A Review. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 280–287. ACM Press, 2002.

[36] J. Mundinger and J. Y. Le Boudec. The impact of liars on reputation in social networks. In *Proceedings of Social Network Analysis: Advances and Empirical Applications Forum*, Oxford, July 2005.

[37] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content addressable network. In *Proceedings of the 2001 ACM SIGCOMM Conference*, pages 161–172, San Diego, CA, August 2001.

[38] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of ACM 1994*

Conference on Computer Supported Cooperative Work, pages 175–186, Chapel Hill, North Carolina, 1994. ACM.

[39] P. Resnick, R. Zeckhauser, E. Friedman, and K. Kuwabara. Reputation systems: Facilitating trust in Internet interactions. *Communications of the ACM*, 43(12):45–48, 2000.

[40] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on eBay: A controlled experiment. Working paper originally presented at the ESA conference, June 2002.

[41] M. Richardson, R. Agarwal, and P. Domingos. Trust management for the semantic web. In *Proceedings of the Second International Semantic Web Conference*, pages 351–368, Florida, USA, 20–23 October 2003.

[42] M. Sabel, A. Garg, and R. Battiti. WikiRep: Digital reputations in collaborative applications. In *To appear in AICA Annual Congress 2005*, Udine, Italy, 5–7 October 2005.

[43] S. Saroiu, P. Gummadi, and S. Gribble. A measurement study of peer-to-peer file sharing systems, 2002.

[44] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proceedings of the 2001 ACM SIGCOMM Conference*, pages 149–160, San Diego, CA, August 2001.

[45] R. L. Trivers. The evolution of reciprocal altruism. *Quarterly Journal of Biology*, 46:35–57, 1971.

[46] L. Xiong and L. Liu. PeerTrust: Supporting reputation-based trust in peer-to-peer communities. *IEEE Transactions on Data and Knowledge Engineering, Special Issue on Peer-to-Peer Based Data Management*, 16(7):843–857, July 2004.

[47] G. Zacharia, A. Moukas, and P. Maes. Collaborative reputation mechanisms in electronic marketplaces. In *Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences-Volume 8*, page 8026. IEEE Computer Society, 1999.

[48] P. Zimmerman and R. Philip. *The official PGP user's guide*. MIT Press, Cambridge, MA, 1995.

[49] L. Zucker. Production of trust: Institutional sources of economic structure 1840–1920. *Res. Organ. Behavior*, 8(1):53–111, 1986.