Journal of Universal Computer Science, vol. 15, no. 9 (2009), 1886-1906
submitted: 15/8/08, accepted: 24/5/09, appeared: 1/5/09 © J.UCS

Applying Reputation Mechanisms in Communities of Practice: A Case Study

Claudia C. P. Cruz
(Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
claudia.paranhos@gmail.com)

Claudia L. R. Motta
(Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
claudiam@nce.ufrj.br)

Flávia Maria Santoro
(State Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
flavia.santoro@uniriotec.br)

Marcos Elia
(Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
melia@nce.ufrj.br)

Abstract: Communities of Practice (CoP) are groups of people sharing practices, interests, and work objectives. In the virtual world, however, it is difficult to trust contributions from people we do not know, because of the variety of information sources and behaviors. In this context, a Reputation Model for CoP (ReCoP) was developed. This paper presents a case study that evaluates one ReCoP mechanism: the degree of agreement among members in evaluating artifacts shared within the community. Data was extracted from a real scenario, and the results provide useful feedback for further studies and improvements in implementing the model.

Keywords: Community of Practice, Recommender Systems, Reputation Systems
Categories: L.6.1, L.6.2, L.6.0

1 Introduction

Knowledge is becoming increasingly important for achieving competitive advantage. The creation and dissemination of organizational knowledge is a social process by which knowledge is shared among members of an organization [Senge 90]. Thus, companies have adopted communities of practice to stimulate organizational learning and knowledge sharing [Novak and Wurst 2004].

A Community of Practice (CoP) can be defined as a group of people who interact regularly to share work experiences, interests or objectives [Wenger et al. 02]. The Internet has mushroomed as a fast, flexible and low-cost medium, contributing to the dissemination of virtual communities of practice within organizations. Hence, geographically distant members of groups of professionals can work together and exchange practical solutions to common problems [Droschl 04].

In this context, one of the most critical points for CoP is the credibility of shared information. A professional is only able to feel safe in using information or motivated

to reuse a design solution presented and debated within the scope of the community if a process of trust is well established among its members at that moment [Preece 04].

Part of the problem can be solved using Recommender Systems [Resnick and Varian 97], by applying the Collaborative Filtering technique [Konstan et al. 97]. The system can reduce problems of information overload and time wastage by recommending only relevant information properly evaluated by individuals with similar tastes. However, if the recommendations do not address users' needs, the system loses credibility, people will probably stop participating and the community ceases to be seen as a Community of Practice. Within organizations, this inhibits the exchange of experience and the creation of a solid knowledge base for future access.

Currently, Reputation Systems have been adopted as a solution by e-commerce, news-sharing and expert sites, in which trust in individuals is important for establishing interactions. ReCoP, a reputation model for CoP, was developed from solutions found in virtual environments and the literature. The model has been implemented within ActivUFRJ, a collaborative environment connecting professionals and students of Federal University of Rio de Janeiro in various academic-interest communities of practice. The model seeks to assist users in creating their trust networks in such a way as to receive artifacts recommended by people they trust [Cruz et al. 07] [Cruz et al. 08].

This paper presents a case study that evaluates one ReCoP mechanism: the degree of agreement among members in evaluating artifacts shared within the community. The study was conducted with data drawn from a real community-of-practice setting, and the results obtained provide useful information for application in further case studies and in improving implementation of the model.

The paper is organized as follows: Section 2 describes related work on Reputation Systems; Section 3 presents the ReCoP model; Section 4 reports on the case study; Section 5 offers some final remarks on the study; and Section 6 identifies the references cited.

2 Related Work

According to [Jøsang et al. 06], Reputation System performance is based on two main concepts: reputation and trust. Reputation is what is generally said or believed about a person's character, while trust reveals personal and subjective opinion. In virtual environments that apply Reputation Systems, users can decide whether or not to trust an individual from the moment he/she establishes his/her reputation.

Reputation reflects a general opinion people have of someone or something. Generally, this opinion is built up from information provided by members of a community on past experiences with the entity [Josang et al. 06]. According to [Resnick et al. 00], Reputation Systems represent an option to help users identify reliable relationships in the Internet, allowing them to evaluate individuals' actions, see community opinion-based reputations, and create their trust networks. In general, these systems aggregate values to calculate individuals' reputations based on feedback provided by other people after their interactions.

According to [Dellarocas 04], Reputation Systems also need to develop immunization mechanisms against the actions of individuals who make use of dishonest evaluations to enhance their reputation and diminish other people's

reputations so as to benefit from the services provided. Such immunization mechanisms include: causing the reputation estimate to be less vulnerable to fraudulent behavior by users; and preventing the use of anonymity, or allowing its controlled use, in order to protect users from possible dishonest evaluations.

Currently, some reputation and immunization mechanisms are used in auction sites, e-commerce, news sharing, expert sites, and other services that need to motivate trust among users in order to ensure that more people use them. The semantic web, social networking, virtual communities and, of course, Recommender Systems [O'Donovan and Smith 05] are all examples of research areas where issues of trust, reputation and reliability are becoming increasingly important, especially as we see related work progress.

2.1 Reputation model in electronic marketplaces

Consumer-to-consumer electronic transaction systems like eBay [Ebay 95] and OnSale Exchange [OnSale 87] create online marketplaces that bring together users who do not know each other. OnSale Exchange is an online auction specializing in computer products, consumer electronics, sporting goods, and auction classifieds, where sellers can list their items for sale and buyers compete in an auction-like bidding system to buy the posted items. Ebay collects user feedback on the transactions performed. The buyer/seller evaluates his/her counterpart positively (+1) when the negotiation meets his/her expectation; negatively (-1) when the negotiation does not meet his/her expectation; and neutrally (0) when he/she does not feel able to perform an evaluation, e.g., on withdrawing from a deal.

According to [Zacharia at al. 99], these kinds of online marketplace introduce two major issues of trust among the users of the system: a) the potential buyer has no physical access to the product of interest while he/she bids or negotiates. Therefore the seller could misrepresent the condition or the quality of his/her product in order to get more money; and b) the seller or buyer may decide not to abide by the agreement reached at the electronic marketplace, and ask at some later time to renegotiate the price or even refuse to complete the transaction. To solve these problems, [Zacharia et al. 99] proposes a reputation-brokering mechanism, so that each user can actually customize his/her pricing strategies according to the risk entailed by the reputation values of the potential counterparts.

2.2 Reputation model in e-commerce

In the e-commerce sites Amazon [Amazon 96] and Epinions [Epinions 99], users evaluate products available for purchase through ratings and comments. These evaluations are used to recommend similar products for the users themselves or for other users with similar preferences. In order to ensure the credibility of recommendations, the Reputation System collects users' opinions indicating whether or not an evaluation was useful for their purchase decision. Thus, reviewers gain reputation points for each positive return and lose points for negative returns.

The recommendation system assigns priority to products evaluated by high-reputation reviewers. In Epinions, users can add reviewers to their "Web of Trust". The system also allows users to block reviewers whose opinions they do not trust. Thus, the system can make customized recommendations based on users' trust

networks and avoid recommending items evaluated by people they have blocked.

2.3 Reputation model in news sharing

In the news-sharing site Slashdot [Slashdot 97], users post and comment on news. Comments can be rated by all other users (moderators), which count as positive or negative points towards the individual making the comment. A second evaluation layer was added in order to minimize the action of unfair or dishonest moderators, in which meta-moderators judge the evaluations.

The meta-moderators form part of a select group of individuals who have been registered in the system for a long time. Users viewed as unfair or dishonest by meta-moderators lose reputation points or are banished from the system, depending on the seriousness of their unfair behavior.

2.4 Reputation model in expert sites

In AllExperts [AllExperts 98], users sign up as volunteers to answer questions in certain categories of knowledge. The service only accepts as experts individuals who actually demonstrate having the skills necessary to answer the questions.

Users evaluate the specialist services by assigning grades from 0 to 10 on various criteria, such as: knowledge, cordiality, clarity of answer and response time. The specialists accrue points for the grades received, and those with the highest scores make up the list of the best specialists in a particular category of knowledge. The system also enables users to look up the history of the most recent evaluations, and the detailed scores obtained over time on each assessment criterion.

2.5 Trust and Reputation Model in Peer-to-Peer Networks

It is important to enable peers to represent and update their trust in other peers in open networks for sharing files, and even more so services. [Wang and Vassileva 03] proposes a Bayesian network-based trust model and a method for building reputation based on recommendations in peer-to-peer networks. Since trust is multi-faceted, peers need to develop differentiated trust in different aspects of other peers' capabilities. The peer's needs differ in different situations. Depending on the situation, a peer may need to trust in a specific aspect of another peer's capability or in multiple aspects. According to [Wang and Vassileva 03], Bayesian networks provide a flexible method for presenting differentiated trust and combining different aspects of trust. Evaluation of the model using a simulation shows that the system where peers communicate their experiences (recommendations) outperforms the system where peers do not share recommendations with each other, and that differentiated trust adds to performance on the percentage of successful interactions.

2.6 Trust and Reputation Model in Recommender Systems

Recommender systems have proven to be an important response to the problem of information overload, by providing users with more proactive and personalized information services. Also collaborative filtering techniques have proven to be a vital component of many such recommender systems, as they facilitate the generation of high-quality recommendations by leveraging the preferences of communities of

similar users. According to [O'Donovan and Smith 05], the traditional emphasis on user similarity may be overstated. Additional factors have an important role to play in guiding recommendation. Specifically the trustworthiness of users must be an important consideration. [O'Donovan and Smith 05] proposes two computational models of trust based on the past rating behavior of individual profiles, and shows how they can be readily incorporated into standard collaborative filtering frameworks in a variety of ways. These models operate at the profile level (average trust for the profile overall) and at the profile-item level (average trust for a particular profile when it comes to recommending a specific item). This trust information can be incorporated into the recommendation process and has a positive impact on recommendation quality.

2.7 Trust and Reputation Model in Virtual Communities

[Abdul-Rahman and Hailes 00] proposes a model that deals exclusively with beliefs about the trustworthiness of agents based on experience and reputational information. According to [Abdul-Rahman and Hailes 00], an experience is the result of a) evaluating an experience with an agent or b) relying on a recommendation from an agent. Informally, this is a model for determining trustworthiness of agents based on the agent's collected statistics on 1) direct experiences and 2) recommendations from other agents. Agents do not maintain a database of specific trust statements in the form of "a trusts b with respect to context c". Instead, at any given time, the trustworthiness of a particular agent is obtained by summarizing the relevant subset of recorded experiences.

2.8 Trust Model in Web-Based Social Networks

In real life we use information about reputations differently from what normally occurs in virtual interaction environments. What services we choose depends strongly on our friends' recommendations. Recommendations made explicitly by a user's "peers" or inferred implicitly through a social network have a significant influence on that user's decision-making process. Thus, the social context given by recommendations from a friend, or a friend's friend, is increasingly being investigated in virtual interaction environments, as can be seen from the work of [Goldbeck and Hendler 06] and [Abdul-Rahman and Hailes 00].

Reputations based on trust networks can be estimated from feedback from the people that users choose, making for estimates that are more reliable and have strong social relevance for users. The disadvantages, however, are that they entail high development costs to spread trust and they raise issues of privacy. In addition, if users choose unreliable people to join their network, they impair the results of the reputation estimation and expose themselves to the risk of receiving undesirable recommendations. Some systems allow users to indicate levels of trust in other users, without that trust being made visible to the members of the community. The FilmTrust system [Goldbeck and Heldler 06] requires users to rank (give trust marks to) friends they add to their social network. Users are warned in advance that they have to give marks from 0 to 10 for how much they trust a friend to recommend films.

In the next section, we present our proposal for implementing a reputation model

in the context of a CoP focusing on interactions directed to learning and knowledge sharing. Communities of practice need a different kind of reputation service, because the communities are much smaller than in e-commerce sites, and the activity is intellectual, not commercial. The system can leverage more information about the users, because it serves a known community (university or company).

3 ReCoP: A Reputation Model For CoP

Learning in CoP rests on the situated learning theory developed by [Lave and Wenger 91], which encompasses lasting interpersonal relations formed around shared practices. Social interaction is a critical component of situated learning, because learners are involved in a community that involves certain beliefs and behaviors to be learned. For there to be interaction among participants, they must be situated within a culture of collaboration and trust.

Development of the Reputation Model for Communities of Practice (ReCoP) addressed issues raised in related work by proposing a set of mechanisms to collect information on reputation and trust, with a view to conducting studies of reputation estimation appropriate to the community-of-practice setting. The model was partly implemented in the Collaborative Environment for Integrated Virtual Work (ActivUFRJ) system at Rio de Janeiro Federal University (UFRJ), which is designed to connect practitioners, researchers and students interested in similar subjects, via virtual communities of practice [Cruz et al. 07] [Cruz et al. 08], where exchanges of experience take place by way of notes and commentaries as part of activities to evaluate artifacts produced by members.

One of the aims of ReCoP is to assist users in creating their own trust networks, in order to receive recommendations on artifacts evaluated by members they trust. The model is based on two kinds of component: reputation mechanisms and immunization mechanisms (Figure 1). Each mechanism has a specific function that can boost the performance of the model as a whole, but which can also be implemented and applied independently of a community-of-practice setting. The mechanisms are associated with three dimensions of the model: Participant Profile, Reputation in CoP, and Recommendations in CoP.

- Participant Profile: mechanisms that use information relating to members' CoP registration data (Initial Reputation and Identity Control);
- Reputation in CoP: mechanisms that aggregate information (Degree of Agreement, Meta-Evaluation and Weight of Meta-Evaluator) about participants' contributions in order to calculate their reputation in communities of practice;
- Recommendations in CoP: mechanisms that aggregate information (History of Participation, Trust Networks) about recent contributions and trust relationships in order to assist in determining recommendations on artifacts shared by community members.

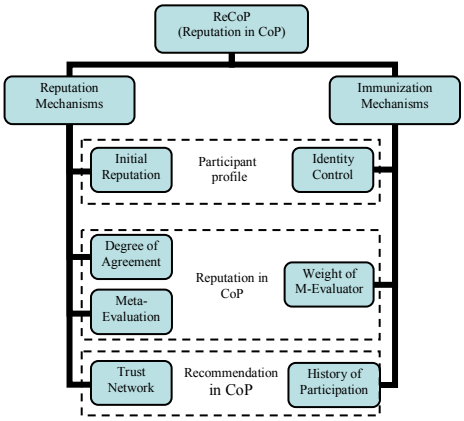


Figure 1: ReCoP – Reputation Model for CoP

3.1 Reputation Mechanisms

Reputation Mechanisms collect information about CoP participants in order to determine members' reputations and assist users in creating trust relationships. The reputation mechanisms comprise: Initial Reputation, Degree of Agreement, Meta-Evaluation and Trust Network.

- Initial Reputation: In related studies, new users begin their participation in the environments with no initial reputation value until they receive feedback from the community on their interactions. In CoP it is common for individuals to have built up a reputation on the subject of interest through their professional experience. The Initial Reputation mechanism is designed to create an expectation with regard to the new participant's behavior, by extracting information on his or her profile relating to the community's area of interest: publications, participation in projects, events, supervision etc. This information is associated with a ranking mechanism that determines an Initial Reputation value for new participants, categorizing them as expert, intermediate or beginner on the subject.
- Degree of Agreement: In communities of practice it is common for participants to analyze various people's opinions on an artifact before consulting it. People who hold an opinion that is consensual with the majority are more generally regarded as more trustworthy by the community than those who hold a completely divergent opinion. The Degree of Agreement mechanism seeks to identify the consensus between a participant and his or her peers (people who have evaluated the same artifacts) by way of the ratings they award. The degree of agreement in evaluations is classified into three levels that influence the participant's reputation: dissent (completely disagreed with his or her peers); partial consensus (partly agreed with his or her peers); consensus (totally agreed with his or her peers).
- Meta-Evaluation: In a community of practice, participants can play the roles of evaluator and of meta-evaluator. A participant is considered an

"evaluator" when evaluating artifacts, and a "meta-evaluator" when providing feedback on an evaluation. The meta-evaluation mechanism enables the community to provide the following kinds of feedback on evaluations of artifacts: helpful/unhelpful, agree/disagree, and relevant/irrelevant. Each type of feedback is associated with a comment by the meta-evaluator and with positive (+1), negative (-1), or neutral (0) marks, which are aggregated in order to generate the evaluator's reputation.

- Trust Network: Generally, trust is propagated within a CoP to the extent that the members themselves find people with similar interests who are recognized in the community for having a good reputation. The Trust Network mechanism enables users to form a social network of people they trust. The purpose of the trust network is to make it easier for people to receive recommendations on artifacts evaluated or suggested by members of the network as a yardstick.

3.2 Immunization Mechanisms

These mechanisms are intended to help protect the environment from possible fraud, thus enhancing its credibility and encouraging more people to use it. The immunization mechanisms that form part of the model are: Identity Control, History of Participation and Weight of Meta-Evaluator's Reputation.

- Identity Control: In the CoP, participant identity is a source of valuable information in emergency situations, and can be impaired if false profiles are created. The Identity Control mechanism links participant identification to the individual's personal registration data in the organization. In this way, it ensures that users cannot set up fictitious profiles in order to take fraudulent action to promote their own reputations or denigrate other participants.
- History of Participation: In the mature phase, the CoP gains support and recognition from the organization and begins to have a sustainable number of members. This, however, leads to a need to organize information flows better in order to facilitate community access to the most recent information. The History of Participation makes it possible to assist the process of recommending the artifacts most recently evaluated by the community, and to identify how recent reputation information is, and thus recognize current trends in participant behavior.
- Reputation Weight of the Meta-Evaluator: Users' reputations determine how trustworthy the community considers their opinions. In this way, meta-evaluators' reputations are treated as a weighting in estimating evaluators' reputations, thus guaranteeing that feedback from meta-evaluators with stronger reputations will be given more importance than feedback from meta-evaluators with weaker reputations.

3.3 ReCoP Implementation

In the ActivUFRJ environment, members participate in the CoP by sharing artifacts of common interest, and the system recommends artifacts evaluated by community members. On consulting artifact evaluations, users can supply feedback through meta-evaluation (Figure 2). The comments posted, as well as any answers to meta-

evaluators' comments, remain available for consultation by other members.

We attempt to construct a scale of reputation based on three dichotomous items regarding users' opinions about artifact evaluations:

- Agreement with evaluator opinions about an artifact: "I agree with you", "I disagree with you";
- Usefulness of an evaluation to the meta-evaluator: "You helped me"; "You did not help me";
- Relevance to the subject of the CoP: "Your evaluation is relevant", "Your evaluation is irrelevant".

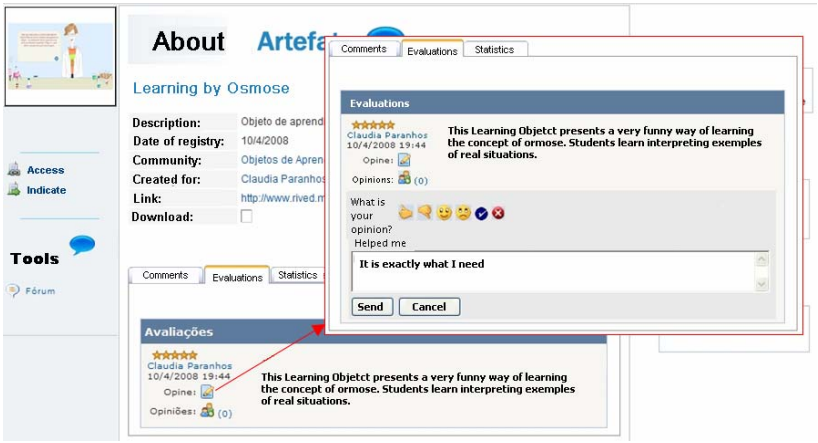


Figure 2: ActivUFRJ - Meta-evaluation Form

These items represent meta-evaluation options connected to rating values, which will generate a feedback to the community about the individual's reputation (Table 1). Users obtain reputation points when they receive positive meta-evaluations ("I agree with you", "You helped me", "Relevant"), which is regarded as indicating that the user contributed to the community. Users do not gain/lose points when they receive contrary meta-evaluations ("I disagree with you") because in CoPs individuals also learn through conflicts of opinion. This kind of feedback indicates that the user diverges from other community members. Users lose reputation points only when they receive negative meta-evaluations ("You did not help me", "Irrelevant"). This kind of negative feedback indicates situations where the user did not contribute to the community.

Var	*Icon*	*Meaning*	*Ratings*	*Feedback to community*
Agreement		I agree with you	+1	Positive)
		I disagree with you	0	Conflict
Usefulness		You helped me	+1	Positive
		You didn't help me	-1	Negative
Relevance		Relevant	+1	Positive
		Irrelevant	-1	Negative

Table 1: Meta-evaluation options

Consulting a member's profile displays a summary of their reputation in each CoP they belong to, showing the number of positive, contrary, and negative ratings. At this point, users can decide whether or not to add this member to their "trust network". Users can indicate the areas of interest where a specific member's opinion is trustworthy (Figure 3). This strategy is intended to help users categorize members of their trust networks by area of interest. This supplies input for the system to generate more customized and reliable recommendations on artifacts evaluated by members the user trusts.

Figure 3: Trust Network Form

4 Case Study

A case study was conducted to validate one of the reputation mechanisms (Degree of Agreement) proposed in the ReCoP model. It was performed using data from evaluations of papers submitted to the 14th Brazilian Symposium on Informatics in

Education (*XIV Simpósio Brasileiro de Informática na Educação*, SBIE 2003). This setting was considered appropriate, because it had features of a CoP which are implicit in the ReCoP model. It was a: "community of professionals with an interest in a specific subject (Informatics in Education), shared artifacts (scientific papers) produced by members of the community, and experts (Program Committee) evaluating the available artifacts with a view to recommending the best of them to the other community members." The setting comprised the following information:

- 33 reviewers involved in the process of evaluating papers;
- 428 papers submitted to SBIE 2003;
- Final result (published/not published) determined by the program committee for each paper submitted to the event;
- Points awarded by each reviewer for the various paper evaluation criteria: "Originality", "Technical Merit", "Readability", "Relevance";
- Reviewer's final recommendation on each paper evaluated (Recommended Action): "Accept"(5), "Accept Weakly"(3), "Reject Weakly"(2),"Reject" (1).

The reviewers' activity consisted in filling out an evaluation form reflecting the issues – "Originality", "Technical Merit", "Readability", and "Relevance" – that the community considers important in the scientific production in the papers. Analysis and evaluation of these items provided basic guidelines for the reviewers' final recommendation (accept or reject). Each paper was evaluated by three different reviewers, each of them awarding it recommendation points. At the end of the evaluation process, all the papers were listed in decreasing order by recommendation score obtained by adding the reviewers' final recommendation points. The papers were then separated into two groups: those for publication (with recommendation scores ≥ 12) and those not for publication (with recommendation scores < 12).

In this context, it was discovered that the degree of agreement among the reviewers on what action to recommend (accept or reject) is a truth criterion that influences the final outcome of whether or not the paper is published. Accordingly, reviewers' decisions to accept or reject a given paper were admitted as defining "recommendation", while the community could agree or disagree with those individual recommendations. Reviewers' reputations, however, would be determined by the degree of agreement between their recommendations and those of their peers, i.e., a given reviewer would have a good reputation if, in most cases, his or her opinion (Recommended Action) agreed with the opinions of the reviewers who evaluated the same papers (Final Result). The shaded area in Table 2 indicates situations of agreement between reviewer and peers in the paper recommendation process. Each combined RA-FR score in the form {accept-published (11); reject-not published (00)} is considered a positive point towards the reviewer's reputation.

Recommended Action (RA)		**Final Result (FR)**	
Levels		Accepted (1)	Rejected (0)
Accepted	1	11	10
Weakly accepted			
Weakly rejected	0	01	00
Rejected			

Table 2: Agreement between reviewer and peers in final recommendation

Operationally then, a measure of a reviewer's reputation is thus established as follows: one positive point is added for each agreement between the action recommended by that reviewer and the final result (Truth Criterion). However, in most CoP the final result of a recommendation is not known *a priori* and it would thus be necessary to construct another scale that would actually be operational and would have to correlate strongly with this true scale. It was thus endeavored to construct this other scale applying the same concept of the Degree of Agreement among the reviewers as regards the points awarded for items on the evaluation form (Originality, Technical Merit, Readability and Relevance) for a given paper, so that when the points of one reviewer (target-reviewer) are compared with those of the other two reviewers, three possible situations can arise (Table 3).

	Rev 1	**Rev 2**	**Rev 3**	**Rev1 reputation**
Dissents	0	1	1	0
Partial consensus	0	0	1	1
Full consensus	0	0	0	2

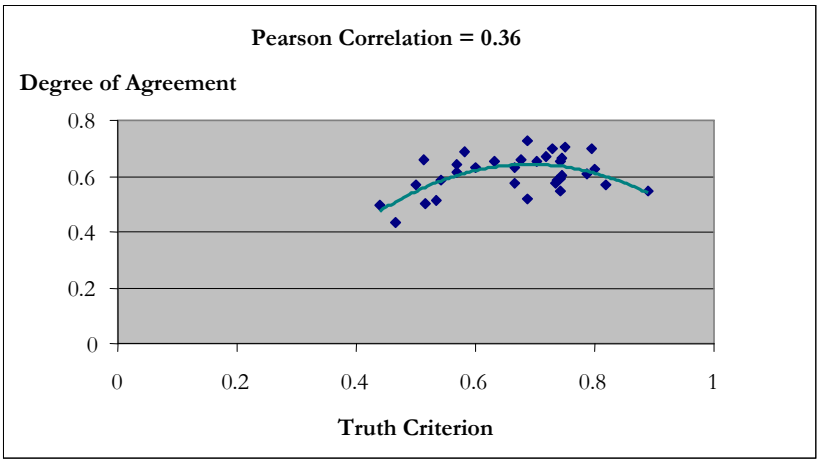
Table 3: Basis for calculating the Degree of Agreement scale

From Table 3 it can be seen that dissent occurs when reviewers (2) and (3) disagree with the target-reviewer (1), partial consensus occurs when one of the two agrees with the target reviewer, and total consensus, when all agree on the points. The target reviewer's possible reputation values for each item on the evaluation form are shown in the last column.

Later, using the scores obtained by all the reviewers, the internal consistency among the four items on the evaluation form was analyzed using Cronbach's Alpha coefficient (α) [Gliem and Gliem 03] to determine whether they in fact constituted a unidimensional scale. The item Readability was found to contribute to reducing the consistency among the items. A single reputation scale was thus generated from the mean among the points obtained only for the items "Originality", "Technical Merit" and "Relevance" ($\alpha = 73\%$). Lastly, the degree of correlation between this second scale and the first was investigated in order to validate the concept of reputation that is attributed to it. This will be described next.

4.1 Correlation Study

In order to evaluate the reputation estimate generated indirectly by the Degree of Agreement mechanism, the correlation between the Truth Criterion and Degree of Agreement reputation scales was analyzed. The Pearson correlation coefficient was calculated as 0.36, and the scales were found not to have a linear relationship, as shown in Graph 1.

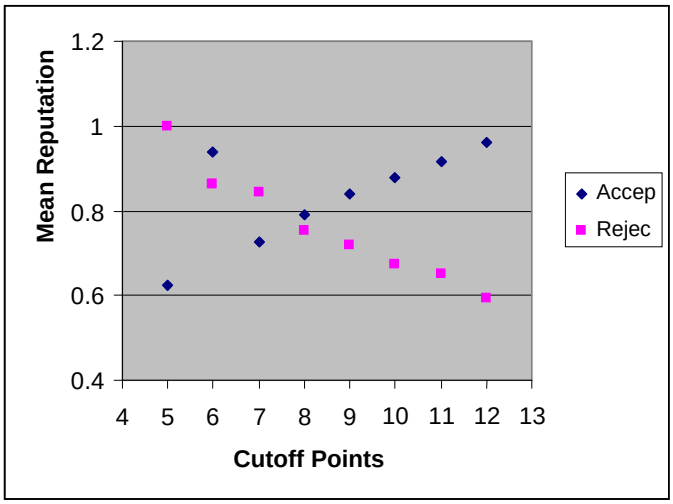


Graph 1: Pearson Correlation –Truth Criterion and Degree of Agreement

When the results were analyzed to determine what could have influenced them, it was found that, when the standard set by the Program Committee is very high, the reviewers tend to agree more on accepting than on rejecting papers. On the other hand, when the standard is low, they tend to agree more on rejecting than on accepting papers, as seen in Table 4 and confirmed by Graph 2. This demonstrates that when reputation is analyzed separately as "agreement on accepting" and "agreement on rejecting" these tend in opposite directions for the various cutoff points. The shaded rank in Table 4 represents the standard required by SBIE 2003 (recommendation score >=12).

Recommended Action	Score
(A)(A)(A)	5+5+5 = 15
(A)(A)(WA)	5+5+3 = 13
(A)(A)(WR)	5+5+2 = 12
(A)(WA)(WA) or (A)(A)(R)	5+3+3 or 5+5+1 = 11
(A)(WA)(WR)	5+3+2 = 10
Etc.	...

Table 4: Score options for paper recommendations



Graph 2: Reputation in accepting and rejecting by cutoff point

These results suggest that the items considered by the reviewers (and/or their respective weights) for accepting the papers may be different from the items considered for rejecting them. It can thus be concluded that reputation seen as agreement between the subject and the community in evaluations of scientific papers is a non-linear concept, which requires further study in order to be defined operationally. The non-linearity observed here will also occur in analyzing the reputations of other artifacts.

However, other factors may have caused the low correlation encountered. For example, in the above study a reputation scale by degree of agreement was constructed by comparing the responses given by different reviewers on the evaluation form items (Originality, Technical Merit and Relevance). On this classical approach, however, the score a reviewer gives on an item depends both on their implicit proficiency, which really does bear on their reputation, and on the degree of difficulty of the item being evaluated. With a view to separating these two factors, a further stage was performed using the Item Response Theory (IRT) [Baker 01].

4.2 Study with Item Response Theory

The Item Response Theory (IRT) enables comparisons to be made among individuals in the same population who have been subjected to partly differing instruments (tests, questionnaires). Respondents' success in a test can be forecast by their ability to respond correctly to the items that make it up. Each individual evaluated responds to the items according to their proficiency. The set of these item responses is a direct expression of the individual's latent proficiency. One of the key features of the IRT is that it is the items, rather than the instrument as a whole, that are the central element, making it possible to analyze individuals more specifically, separating proficiencies by the characteristics of the items.

In the previous stage of the case study, reviewers' reputations were determined,

on the one hand, as being a direct measure of their ability to make recommendations for and against artifacts (in this case, scientific papers), providing they agreed with the opinions of their peers. This results in papers published that have received recommendations for acceptance (or only one recommendation of "weakly rejection"), and not published that have received at least one recommendation of rejection. For a Reviewer j who evaluated Nj papers, where Nj = Nj(accepts) + Nj(rejects), there are thus the following empirical probabilities of that reviewer's deciding "correctly" in terms of the final decision (Equations 1 and 2).

$$Pb_{Acceptance} = \frac{N_j(accepts)}{N_j}$$

Equation 1. Probability of "correct" decision in accepting papers.

$$Pb_{Rejection} = \frac{N_j(rejects)}{N_j}$$

Equation 2. Probability of "correct decision" in rejecting papers.

Thus a reputation for Reviewer j determined by the truth criterion (direct measure) would be given by the mean number of total "correct" decisions in recommending scientific papers, calculated as the product of the probabilities of each of the two – supposedly independent – events occurring together, and the number of papers evaluated by j (Nj), as expressed by Equation 3.

$$MeanNumber_{Correct} = Pb_{Acceptance} * Pb_{Rejection} * N_j$$

Equation 3: Mean number of total "correct" decisions in accepting and rejecting

On the other hand, reviewers' reputations were also constructed, as an indirect measure, from the responses they gave to the items on the evaluation form (Originality, Technical Merit and Relevance). The fact that the construct for this proficiency proved to have high internal consistency among the items gave rise to the study and application of the IRT in this study, because this is a necessary condition of the model. Accordingly, reviewers' indirect reputations were considered as being their latent proficiency (θ) in agreeing totally (score 2), partly (score 1) or disagreeing (score 0) with their peers on the items evaluated (see Table 3).

The model applied was the IRT/GRM (Gradual Rated Model) [Baker 01] using the WINGEN software [Han 07] [Han and Hambleton 07], a Windows program that generates parameters and item responses of an IRT model that best fits the distribution data of the latent scale, θ, for each reviewer. The input data are: (a) Number of cases (quantity of papers evaluated); (b) Form of distribution (normal); (c) Mean and standard deviation of the Degrees of Agreement on the papers evaluated, calculated via the Degree of Agreement mechanism; (d) Number (=3) of items evaluated ("Technical merit, Originality, Relevance"); and (e) Number (=3) of respective response options ("0, 1, 2").

Once configured, one of the outputs of the WINGEN program is the mean number of "correct answers" in the "test" (in this case, the latent θ scale) from the

probability curves for "correct" decisions on the items. This mean value then constitutes the indirect measure of reputation, estimated by the Degree of Agreement mechanism applying the IRT to each reviewer, to be compared with the respective mean value of the direct measures produced in this case study (Equation 3).

In order to determine the efficiency of the reputation estimation by Degree of Agreement, the null hypothesis (H_0) formulated was that there is no difference between the reputation determined by the truth criterion and reputation estimated by the Degree of Agreement mechanism applying the IRT to each reviewer. To test the null hypothesis H_0, the Z statistic was used at a two-sided significance level of 5% ($Z=1.96$). The null hypothesis (H_0) test was approved for 94% of the cases and was rejected for only two reviewers. This result suggests strongly that the estimate of reputation by agreement using the IRT can be taken as a reliable measure.

In order to illustrate the good descriptive and predictive quality of the Item Characteristic Curve - ICC/IRT analyses obtained with the data on SBIE 2003 evaluations, three reviewers with the following evaluation profiles were selected from the sample:

- Reviewer 49: Atypical profile, different from the majority. The only reviewer to reject all the papers reviewed.
- Reviewers 53 e 67: who have been rejected by the null hypothesis (H_0) test.

Figure 4 shows the expected behavior for probability of "correct" decision in accepting papers, separately in each score: dissent (0), partial consensus (1) and consensus (2); previously defined (Table 3) for each item ("Originality", "Technical Merit" and "Relevance") that, combined, compose the ability of a reviewer to agree/to disagree with his/her pairs in papers evaluated. Each item contributes to form the judgment value on the reputation of the target-reviewer.

The horizontal scale represents the ability (θ) in standard deviation score average = 0 and shunting line standard deviation = 1. The vertical scale represents the probability of "correct" decision in accepting papers between (0,1) and the probability curve has the "S" form since it tends to zero and to one (extended from - infinite + infinite, without never touching the axle) in the limits of the scale, while it grows quickly in the middle way of the scale.

The shapes of the curves are associated with the reviewers' profiles and with the technical properties of the item: difficulty corresponds to the value on the latent θ scale for which there is 50% probability that the attribute will/will not be represented by the option/item, while discrimination is given by the slope of the probability curve (1^{st} derivative) for the option/item.

Since the scores dissent (0), partial consensus (1) and consensus (2) are dependent among each other by construction, we notice that, in the ideal situation depicted in Figure 4: (i) "S" curves of scores (0) and (2) would be symmetrical between themselves in the average point of the ability ($\theta=0$); (ii) "S" curve of score (1), that corresponds to the partial dissent, also reflects this symmetry around the average point ($\theta=0$), as for, by construction, this score is a combination of the other two. We also notice that in the idealized case, the average point ($\theta=0$) would correspond to the degree of difficulty of the three scores.

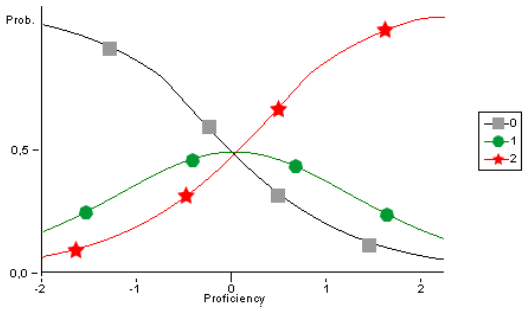


Figure 4: ICC standard behavior according to reviewer capacity

Using figure 4 as a standard, we can analyze how close or how far the actual experimental characteristic curves (ICCs) behave. Figures 5 - 13 show the Item Characteristic Curves (ICCs) – for the item as a whole and separately for each response option, and the Test Characteristic Curves (TCC) - which is the functional relation between the true score (average value of items scores) and the ability scale – generated by the IRT/GRM using the WINGEN software for the five reviewers selected from the sample. It is worthwhile noticing that the reputation estimated by the Degree of Agreement mechanism, as mentioned before, is based on the TCC curve calculations.

The curves suggest that the main reasons why Reviewer 49 - who rejected all the papers reviewed - differs from his peers have to do more with the relevance issue and less with the originality and technical merit ones. That is the case because the ICC curve is almost very flat over ability θ scale for the first indicating that he/she could agree, partially or totally disagree whatever the reputation ability may he/she have.

Analyzing the ICC curves for both reviewers (53 and 67) - who contributed for the null hypothesis H_0 rejection - we can see two possible misbehaviors in relation to our standard curve. For both reviewers, the relevance ICC curve tend to flatness while the originality and technical merit ICC curves tend to be steeper and shift to lower values of the ability scale.

5 Final Conclusions and Future Work

This paper presented a case study conducted to validate a reputation mechanism proposed in the ReCoP model: A Model for Reputation in Communities of Practice. A case study was carried out in two stages: the first led to results regarding the complexity both of the concept and estimating reputation; the second produced more conclusive results as to the validity of the estimation proposed by the model.

The main contributions drawn from this case study were the discovery that reputation by agreement among evaluations by members of the CoP is a non-linear concept when what is involved is measuring agreement in recommending an artifact and agreement in rejecting an artifact. Also, the various profiles of the professionals involved and the implicit difficulties of each in evaluating items according to their

latent proficiencies all have implications for estimation of their reputations.

Nonetheless, applying Item Response Theory to the results of this case study is regarded as showing signs that the estimation of reputation proposed by the ReCoP model can be applied in the context of a community of practice. Due to the limitations intrinsic to the setting chosen as the object of study, it was not possible, in this first case study, to validate the other mechanisms proposed in the model (initial reputation, meta-evaluation, weight of the meta-evaluator, identity control, and history of participation). For future work, the intention is to continue this research, including further case studies and validation of the proposed model as a whole.

Reviewer 49 – Atypical profile

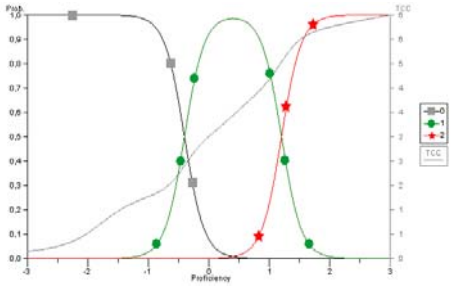


Figure 5: ICC- Originality

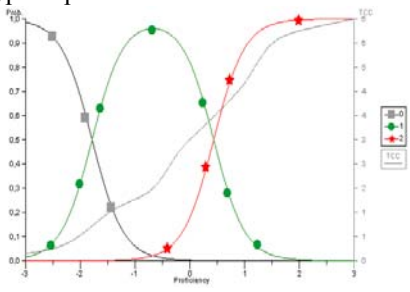


Figure 6: ICC - Technical Merit

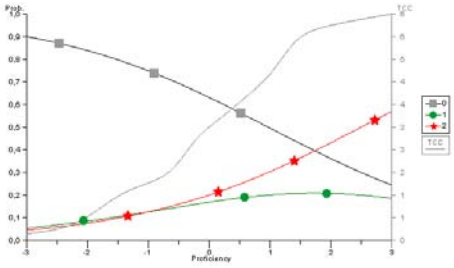


Figure 7: ICC – Relevance

Reviewer 53 - rejected by the null hypothesis (H_0) test

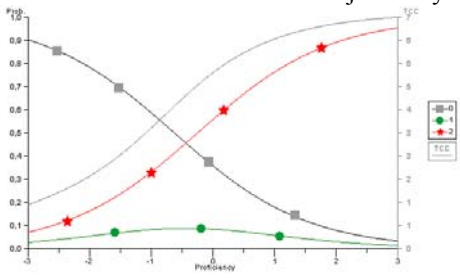


Figure 8: ICC-Originality

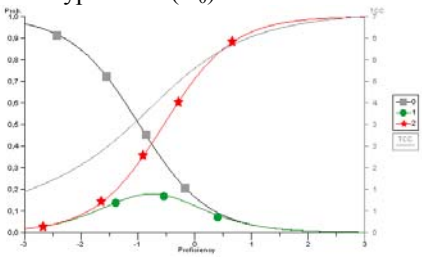


Figure 9: ICC-Technical Merit

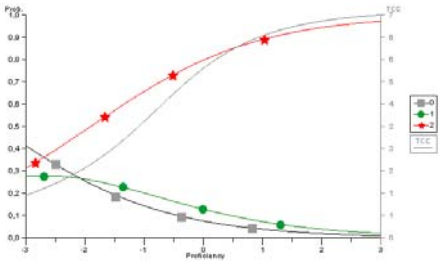


Figure 10: ICC-Relevance

Reviewer 67 - rejected by the null hypothesis (H0) test

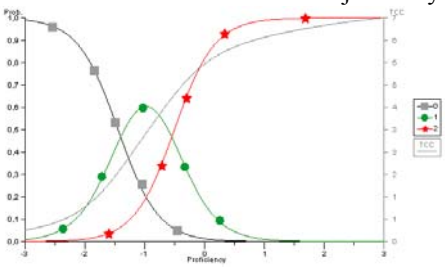


Figure 11: ICC-Originality

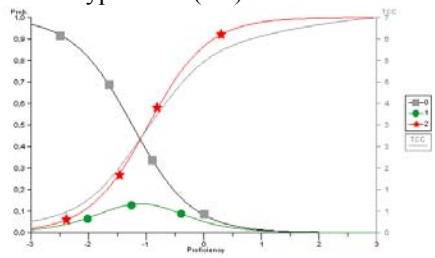


Figure 12: ICC -Technical Merit

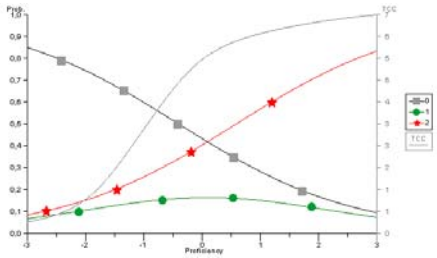


Figure 13: ICC – Relevance

Acknowledgements

This research was partially supported by Brazil's National Council for Scientific and Technological Development (CNPq) (www.cnpq.br/english/cnpq/index.htm).

References

[Abdul-Rahman and Hailes 00] Abdul-Rahman, A., Hailes, S. Supporting Trust in Virtual Communities. Proceedings of the 33rd Hawaii International Conference on System Sciences HICSS'00, vol. 6, p. 6007, 2000.

[AllExperts, 98] AllExperts, 1996, http://www.allexperts.com

[Amazon, 96] Amazon, 1996, http://www.amazon.com

[Baker 01] Baker, Frank "The Basics of Item Response Theory". ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD, 2001. Access: http://edres.org/irt/baker/

[Cruz et al. 07] Cruz, C.C.P.; Motta, C,L.R.; Santoro, F.M. Applying Reputation Mechanisms in Communities of Practice. In: IX Webmidia, Gramado, 2007.

[Cruz et al. 08] Cruz, C. C. P., Motta, C. L. R., Santoro, F. M., Elia M. Reputation Model in Communities of Practice: A Case Study. In Proceedings of the 2008 12th International Conference on Computer Supported Cooperative Work in Design – CSCWD 2008.

[Dellarocas 04] Dellarocas, C. S. Building Trust Online: The Design of Robust Reputation Reporting Mechanisms for Online Trading Communities. Idea Group Inc., chapter VII, p. 95-113, 2004.

[Droschl 04] Droschl, G. Communities of Practice: An Integrated Technology Perspective. Journal of Universal Computer Science, vol. 10, 2004 n°3, p284-293.

[Ebay 95] Ebay, 1995, http://www.ebay.com/

[Epinions 99] Epinions, 1999, http://www.epinions.com

[Golbeck and Hendler 06] Golbeck, J., Hendler, J. Inferring Binary Trust Relationships in Web-based Social Networks. ACM Transactions on Internet Technology, vol. 6, n. 4, p. 497–529, November, 2006.

[Gliem and Gleim 03] Gliem, J. A., Gliem, R. R. "Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales". Midwest Research to Practice Conference in Adult, Continuing, and Community Education, 2003.

[Han 07] Han, K. T. WinGen: Windows software that generates IRT parameters and item responses. Applied Psychological Measurement, 31(5), p. 457-459, 2007.

[Han and Hambleton 07] Han, K. T., Hambleton, R. K. User's Manual: WinGen (Center for Educational Assessment Report No. 642). Amherst, MA: University of Massachusetts, School of Education, 2007.

[Josang et al. 06] Josang, A., Ismail, R., Boyd, C. "A Survey of Trust and Reputation Systems for Online Service Provision". Distributed Systems Technology Centre and Information Security Research Centre, Queensland University of Technology Brisbane Qld 4001, Australia, 2006.

[Konstan et al 97] Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R, Riedl, J. "GroupLens: applying collaborative filtering to Usenet news." Vol. 40, p. 77-87, 1997.

[Lave and Wenger 91] Lave, J.; Wenger, E. Situated Learning: Legitimate Peripheral Participation. New York, NY: Cambridge University Press, 1991.

[Novak and Wurst, 2004] Novak, J.; Wurst, M. Supporting Knowledge Creation and Sharing in Communities based on Mapping Implicit Knowledge. Journal of Universal Computer Science, vol. 10, 2004, n°3, p235-251.

[O'Donovan and Smith 05] O'Donovan, J., Smyth, B. Trust in Recommender Systems. In: Proceedings of the 10th International Conference on Intelligent User Interfaces - IUI'05, January 9–12, 2005, San Diego, California, USA.

[OnSale 87] OnSale, 1987, http://www.onsale.com/

[Preece 04] Preece, J. Etiquette, empathy and trust in communities of practice: Steppingstones to social capital. Journal of Universal Computer Science, vol. 10, 2004, n°3, p294.

[Resnick and Varian 97] Resnick, P., Varian, H. R. Recommender Systems. In: Communications of the ACM, 40(3), 56-58, 1997.

[Resnick et al 00] Resnick, P., Zeckhauser, R., Friedman, E., & Kuwabara, K. Reputation Systems. In: Communications of the ACM, 43(12), 45-48, 2000.

[Senge 90] Senge, P. M., "The Fifth Discipline - Art and Practice of the Organization that Learns", New York, Doubleday, 1990.

[Slashdot 97] Slashdot, 1997, http://www.slashdot.org/

[Wang and Vassileva 03] Wang Y. Vassileva J. Trust and Reputation Model in Peer-to-Peer Networks, Proc. of IEEE Conference on P2P Computing, Linkoeping, Sweden, September 2003.

[Wenger et al. 02] Wenger, E. C.; Snyder, W. M.; Richard McDermott, R., "Cultivating Communities of Practice - A Guide to Managing Knowledge", Harvard Business School Press, Cambridge, MA 2002.

[Zacharia et al. 99] Zacharia, G.; Moukas, A.; Maes, P. Collaborative Reputation Mechanisms in Electronic Marketplaces. In: Proceedings of the 32nd Hawaii International Conference on System Sciences. 1999.