Dynamic Credibility Threshold Assignment in Trust and Reputation Mechanisms Using PID Controller

A Thesis Submitted to the

College of Graduate Studies and Research

in Partial Fulfillment of the Requirements

for the degree of Master of Science

in the Department of Computer Science

University of Saskatchewan

Saskatoon

By

Mohsen Mohkami

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science

176 Thorvaldson Building

110 Science Place

University of Saskatchewan

Saskatoon, Saskatchewan

Canada

S7N 5C9

ABSTRACT

In online shopping buyers do not have enough information about sellers and cannot inspect the products before purchasing them. To help buyers find reliable sellers, online marketplaces deploy Trust and Reputation Management (TRM) systems. These systems aggregate buyers feedback about the sellers they have interacted with and about the products they have purchased, to inform users within the marketplace about the sellers and products before making purchases. Thus positive customer feedback has become a valuable asset for each seller in order to attract more business. This naturally creates incentives for cheating, in terms of introducing fake positive feedback. Therefore, an important responsibility of TRM systems is to aid buyers find genuine feedback (reviews) about different sellers.

Recent TRM systems achieve this goal by selecting and assigning credible advisers to any new customer/buyer. These advisers are selected among the buyers who have had experience with a number of sellers and have provided feedback for their services and goods. As people differ in their tastes, the buyer feedback that would be most useful should come from advisers with similar tastes and values. In addition, the advisers should be honest, i.e. provide truthful reviews and ratings, and not malicious, i.e. not collude with sellers to favour them or with other buyers to badmouth some sellers.

Defining the boundary between dishonest and honest advisers is very important. However, currently, there is no systematic approach for setting the honesty threshold which divides benevolent advisers from the malicious ones. The thesis addresses this problem and proposes a market-adaptive honesty threshold management mechanism. In this mechanism the TRM system forms a feedback system which monitors the current status of the e-marketplace. According to the status of the e-marketplace the feedback system improves the performance utilizing PID controller from the field of control systems. The responsibility of this controller is to set the the suitable value of honesty threshold. The results of experiments, using simulation and real-world dataset show that the market-adaptive honesty threshold allows to optimize the performance of the marketplace with respect to throughput and buyer satisfaction.

ACKNOWLEDGEMENTS

First of all, I wish to express my deepest gratitude to my supervisor, Dr. Julita Vassileva, for her continuous support and advice throughout my graduate studies at the University of Saskatchewan. Her guidance helped me in all the time of research and writing of several publications and this thesis. Her guidance has been priceless and I consider myself very fortunate to have her as my supervisor.

I also would like to thank Dr. Ralph Deters, Dr. Chanchal Roy, and my external examiner Dr. Li Chen for devoting a portion of their valuable time to serve as my thesis committee and providing me with constructive suggestions.

I am thankful as well for having the chance to work with Dr. Zeinab Noorian. She has helped me with her invaluable advice on my research and co-authoring several papers.

Finally I wish to express my heartfelt thanks to Gwen Lancaster and other staffs at the department of Computer Science and my colleagues from Multi-Agent, Distributed, Mobile and Ubiquitous Computing (MADMUC) lab for their helps throughout my graduate studies.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TRM	Trust and Reputation Management
P2P	Peer to Peer
WSN	Wireless Sensor Network
MANET	Mobile Ad hoc Network
PID	Proportional Integral Derivative

CHAPTER 1

INTRODUCTION

Open electronic marketplaces are uncertain places. They consist autonomous self-interested participants that act and interact flexibly and intelligently[26]. All kinds of people can interact in these marketplaces hiding their actual identities. We can expect that malicious sellers may impersonate benevolent ones. Online marketplaces provide consumers with the ability to leave feedback about sellers. However, buyers may be deceived by the number of positive reviews a seller has received, if the reviews are fake. Validating testimonies and reviews of different products provided by buyers is challenging, since participants neither feel obliged to provide truthful reputation information nor even to share their experiences with their peers.

The performance of an e-commerce systems is a function of a large number of parameters, such as information quality with respect to the accuracy of provided feedback [3], buyers satisfaction and their net benefits [5], service quality of providers, and assurance including the adopted credibility and security measures [33].

Designing reputation systems for open marketplaces seems to be an effective approach to ensure that only participants with satisfactory qualities can prosper [14, 39]. Trust and Reputation Management (TRM) systems in e-marketplaces help users find valuable reviews. In other words, TRM systems have the responsibility to extract honest and useful reviews from all reviews given to a seller. To do so, reputation systems assist buyers in their decision making process by providing them with trustworthiness assessment techniques to thoroughly evaluate the credibility of other buyers (advisers), considering various parameters and environmental circumstances.

Different trust and reputation mechanism have been proposed in the literature which model the trustworthiness of participants via different approaches such as socio-cognitive [37], game theoretical [18, 40], and probabilistic models [42, 36, 10]. Although these and other trust models [30] have shown very promising results in accurately modeling the trust of participants, there are certainly opportunities for further optimization with respect to the accuracy of the model. More specifically, existing reputation systems perform under the assumption of the existence of a credibility threshold to retain only trustworthy advisers. The credibility threshold sets a decision boundary on the behavioral model of advisers and characterized them as honest and malicious. These systems suffer from a lack of a systematic approach for adjusting the honesty threshold to the dynamic environmental conditions[1].

---

[1]Throughout the paper, the terms *credibility threshold* and *honesty threshold* are used interchangeably.

Trust and reputation systems use different criteria calculating trust values for buyers and assigning them as advisors for other buyers. While the objective of TRM systems is to find reliable sellers and reviews about a seller, the goal of Recommender Systems (RecSys) is to find products in which users may have interest. Assigning advisors for users has also been an approach in the area of the recommender system to recommend suitable items, but the focus is mostly on the similarities that exist between user (e.g. in interests, behavior traces, etc.). In order to differentiate between benevolent and malicious advisors, TRM systems specify a boundary as honesty threshold. Typically this boundary is defined as a fixed threshold which is independent of the situation of the e-marketplace.

Defining the threshold for acceptable level of honesty of advisers is very important. Having a fixed predefined trustworthiness threshold[2] makes a TRM system vulnerable in different environments. The foremost drawback of having a static honesty threshold is that an inappropriately set threshold would filter away possibly good advice, or the opposite - allow malicious buyers to badmouth good services. A low threshold will result in a plenty of possible advisers, but the quality of advice may be low. In this situation, deceitful advisers who maintain a minimum level of trustworthiness remain undetected and could actively contribute into a buyer's decision making process. On the other hand, a higher credibility threshold leads to the contribution of a smaller number of advisers and can make it impossible to find advisers. Clearly, adjusting a threshold value is a trade-off between the number of credible advisers and the risk of being misled by deceptive peers.

In order to mitigate this drawback, this thesis proposes a dynamic threshold management system that adopts the idea of feedback controllers from the field of control systems to optimize the market performance. We built a controller that monitors the quality of e-marketplace and uses a PID feedback controller technique [32] to determine new values for the honesty threshold. Buyers then dynamically re-evaluate their network of trustworthy advisers according to the new recommended value.

Our approach was validated experimentally by integrating our PID-based honesty threshold controller into a simulated e-marketplace with different population tendency. Experimental results show that adaptively tuning the honesty threshold to the market performance enables honest buyers to obtain higher quality of services and more accurately detect malicious advisers in comparison with the static threshold values defined based on designer intuition that are used in previous work.

This work situates well within the context that Jøsang and Golbeck (2009) outline as the central concern for future research in multi-agent trust modeling. According to [15], artificially adjusting the credibility threshold value might be inappropriate since one of the ways in which trust models can fail to be robust is in relying on a set of untrustworthy advisers. The methods we outline in this paper seek to address this concern.

A credibility evaluation mechanism guided by the PID-based threshold adjustment creates the opportunity of de-

---

[2]Throughout this thesis, the terms *credibility threshold* and *trustworthiness threshold* are used interchangeably.

signing self-improving trust and reputation systems which learn from the state of the e-marketplace promoting the acceptance of web-based agent-oriented e-commerce by human users.

The structure of this thesis is as follows. Chapter 2 presents background and literature survey of this study. In Chapter 3, the proposed approaches that can overcome the problems exist in the literature are described. In the first part of Chapter 4 describes the experiments carried out, and the results of these experiments are demonstrated. Finally, Chapter 5 summarizes the work done in this thesis and lays out the directions for future work.

CHAPTER 2

BACKGROUND AND MOTIVATION

Marsh's PhD dissertation [23] was one of the first studies that introduced the trust notion to the computer science domain. Although trust is a subjective term and different agents will have different ideas of trust, assigning a real value for trust makes it more easily understood and formalized. Recently, trust and reputation management systems have demonstrated their importance in online marketplaces more than at any other time. In online marketplaces, e.g. eBay, trust is an important part of the system, since buyers have to be sure of the trustworthiness of sellers and vice versa, otherwise they would never commit their money in online transactions.

2.1   Trust

As different studies illustrate, trust is the degree to which a person relies on somebody else [17][48]. Trust has different characteristics which are as follows:

1. Trust is effective in situations where the environment is nondeterministic and individuals should rely on each other to achieve goals and complete tasks. This characteristic is evident in online e-marketplaces where participants using that website have missions: buying what they want and selling their commodities. However, to achieve their goals, users have to rely on other participants whom they do not know well, yet have to rely on.

2. Trust mostly depends on the context. That is, if we trust the advice of a physician about healthy eating, it does not necessarily mean that we have to rely on him for his suggestions on how to deal with our car problems.

3. Trust is subjective. Not all people have the same opinion about one thing. The same can be said about whether someone should be considered trustworthy or not.

4. Trust is neither symmetric nor transitive. Based on characteristics 1-3, we can conclude that trust cannot be symmetric or transitive. Trust is not symmetric, because if person A trusts person B, it does not mean that person B trusts A. And trust is not transitive because if person A trusts person B and person B trusts person C, it does not necessarily imply that person A trusts person C.

2.2 Reputation

Although the terms *trust* and *reputation* have been used with similar meanings in different studies, they have a slight, yet critical, difference. Trust is a subjective opinion one agent can have about another, while reputation is what is generally thought about a particular agent by a group of other agents. The distinction between these two terms can be depicted through the following statements by Jøsang et al. [17]:

"I trust you because of your good reputation."

"I trust you despite your bad reputation."

As can be seen, one may trust another person because he is reputable, i.e. generally believed to be trustworthy. Or a person may trust another person although he does not have a good reputation in his community. The difference between these two is rooted in the subjectivity of trust, in contrast to reputation which is considered to be more objective, as it summarizes the trust of multiple agents. Naturally, the level of objectivity depends on how specific the context is in which the reputation is evaluated and on the number of agents that form the community computing the reputation.

Amazon and eBay benefit from reputation metrics instead of trust metrics. That is, they store a global value for each user based on the feedback of all other users who have had transactions with them. Many studies such as [17][48] mention that reputation is also referred to as *global trust*, and the subjective trust of one agent in another is referred to as *local trust*.

Most studies in the field of trust and reputation consider trust management systems that receive feedback from participants of the system in the form of both positive and negative feedback. Some researchers, however, believe that in some circumstances collecting complaints about users is sufficient (as it is more likely that users will have motivation to provide feedback if they feel cheated, than when they are satisfied). One of the earliest studies which only takes into account complaints (negative feedback) from peers in a P2P network is by Aberer and Despotovic [1]. Another study that uses negative feedback to build reputation is by Maiden et al. [22]. The authors make the assumption that almost everyone is trustworthy and being untrustworthy (malicious) is an exception. Obviously, this form of trust collection is lightweight both on the system side, since the traffic and processing required to process the trust feedback is reduced when we do not need to have positive or neutral feedback, and on the user side, as less input from the user is required.

2.3 Characteristics of Trust and Reputation

2.3.1 Application Areas

Trust and reputation management systems are applicable to different areas, forms of networks, and application domains. Yu et al. describe different aspects of applying trust and reputation to Wireless Sensor Networks (WSN) and Mobile

Ad hoc Networks (MANET) [48]. This phenomenon has been used in social networks for many years and researchers such as Massa and Avesani in [24] and Nojoumian and Lethbridge in [27] investigate the use of trust in this area.

Beside the traditional electronic marketplaces such as Amazon and eBay that were discussed in the paragraphs above, trust has recently been brought to a new marketplace called the data marketplace, where the products being sold are the information of users and other systems. In such systems, privacy and trust are combined to put users' minds at ease. One of the few studies in this specific area is carried out by Yassine et al. where different agents work on behalf of the users to process transactions [46]. In this study, the trustworthiness of consumers is evaluated based on a simple certification check. Certificates are given out by a trustworthy third party and each consumer's trust is measured based on the certificates they have. Without a doubt, the trust of consumers should be calculated and even updated based on the feedback of the users as well. Otherwise, it could lead us to major problems caused by malicious consumers that merely obtained specific certificates.

2.3.2 Architecture

Regarding the structure of the networks, trust management systems can be either centralized or decentralized. eBay, for example, has a centralized reputation management system where users send their feedback to a centralized server, through a website, and also receive information about different participants from the same server. On the other hand, other types of networks such as Peer to Peer (P2P) and Wireless Sensor Networks (WSN), where no centralized monitoring point exists, benefit from decentralized trust mechanisms. Wang, Zhang and Vassileva in [44] propose a super-agent based reputation mechanism to help consumers in web service selection to find trustworthy service providers. This case stands somewhere between centralized and decentralized trust management systems. Super-agents' responsibilities are to collect feedback from consumers about services and provide consumers with information about the trust value of specific services. For example, the consumer C wants to have a service S. It asks a super-agent SA about service providers. After receiving information from SA and completing a transaction with S, C updates its trust for both SA and C based on its satisfaction with the service. As a reward, super agents receive virtual points from service providers each time they introduce them to consumers. These points can be used to redeem services from providers.

Although Amazon and eBay have centralized systems and make use of reputation metrics, a trust and reputation management system can make use of reputation (global opinion about a certain item) metrics regardless of the architecture of the system, be it centralized or decentralized. This has been proved in several studies. For instance, Aberer and Despotovic in [1] use a pure decentralized network, P2P. Nevertheless, they end up having a global trust metric in their system. They utilize a decentralized storage method named P-Grid to store evidence throughout the system. After each transaction, the two participants can submit their complaints about each other. Complaints about each specific

peer in the network are stored in another particular peer which has been assigned by the P-Grid method. Consequently, an overall idea about a peer can be found in the system. On the other hand, Massa and Avesani in one study and Mohkami and Shakeri in another study argued that local trust metrics outperform reputation metrics, especially for controversial users [24] [25] . Massa and Avesani in [24] propose their idea and test it on a real online social network called Epinions. These studies show that although there is a central server for the social network, we can have a local trust metric for the trust and reputation management of that system. This issue, having a local or global trust metric, completely depends on the functionality of the system.

2.3.3 Dynamics

Dynamics of trust is another aspect of trust which most studies do not consider. In many research studies and applications, trust is updated in the same way for every node in a network in different stages of its trust evolution. For example, in many cases, trust metrics just add the previous trust value of a node with a fixed value or they just take the average of all the trust values available for the user to come up with a current trust value to predict the future behavior of that node. In contrast, Urbano, Rocha, and Oliveira in [43] and Nojoumian and Lethbridge in [27] consider dynamics of trust and define trust evolution functions and trust update functions. Urbano, Rocha, and Oliveira define a trigonometric formula to update trust that takes into account the trust values when updating them. For instance, the rate of the increase of trust for users whose trust is between 0.3 and 0.6 is higher than that of users with trust more than 0.6. Similarly, Nojoumian and Lethbridge introduce a trust update function which consists of two functions, one for calculating trust for cooperation and the other for defection of the users.

2.3.4 Advisor Network Formation

Generally speaking, in most current e-commerce websites such as eBay and Amazon all of the users act as advisors for each other. That is, the trust and reputation management systems within these websites calculate the global reputation of entities, sellers and products. As a result of these cumulative ratings for a seller or item, the users of that system decide whether to buy that item (e.g. in Amazon), or conduct business with a seller (e.g. in eBay).

A global reputation is not the best approach for these marketplaces. It is not rational to consider everyone as our advisors without thinking about who these users may be. This is not logical for two main reasons:

1. In any e-marketplace, even in a marketplace in the real world, there inevitably exist users who intentionally want to deceive others through their advice or ratings. For example, it is widely known and a lot of the ratings on TripAdvisor are fake, introduced by either malicious users who collude with restaurant or hotel owners to raise their reputation or to badmouth the competitors. They mostly entrap victims to lead them to do their desired actions, which is purchasing desirable items from their favorite sellers. These malicious users usually try to take

advantage of the market system's limitations to achieve their goals. Considering these circumstances, we do not want to be misled by invalid ratings.

2. Even if we ideally think of a malicious-free society, there are other factors that we should take into account. Not all users are alike and think the same way. The members of a marketplace can give feedback and rate products and sellers based on their interests, preferences, priorities. Naturally, we prefer to use those users' ratings who have the same interests as we do, (i.e. our taste-mates).

Consequently, a way to help the users of a community is by allowing them to make decisions based on the feedback of users similar to them. Thus, they only receive advice from those users who are trustworthy enough to overcome the first issue, and have a similar way of thinking to address the second problem.

This idea is the basis of the collaborative filtering method developed in the Recommender Systems area. As Herlocker, Konstan, and Riedl in [12] lay out, neighborhood-based collaborative filtering methods have three steps:

1. calculating the similarity of all users with the active user

2. choosing a subset of similar users as advisors

3. computing a prediction for items based on advisors' ratings

However, generally, collaborative filtering systems do not involve the second step mentioned in the list above. They take into account all users that have had the same experience as the active user to be advisors [19][20][35]. Herlocker, Konstan, and Riedl also mention that it is expensive in terms of computation to calculate the similarity of all users that can be considered as advisors. Moreover, it is not useful and effective as there may be only a few users who share the same taste with the active users. As a result, to help the system achieve better performance in terms of accurate predictions, several studies have considered limiting the number of advisors in a system [7][8][12]. In order to do so, they use two techniques to determine the number of advisors for each user. The first technique is setting an absolute correlation threshold. This threshold decides which users are qualified to be assigned as an advisor for the active user. For the second technique, they limit the number of advisors to the best n number of users who have the highest values in terms of similarity.

Another thing we can learn from the study carried out by Herlocker, Konstan, and Riedl is that there is a conspicuous inverse relationship between the correlation threshold and coverage in the results [12]. In other words, if the system chooses a higher correlation threshold, fewer advisors for a user will remain. Therefore, the user can predict the rating value of fewer items due to the lack of information and advice from its advisors. On the other hand, the value of the correlation threshold has a direct impact on the accuracy of the system. The user can make sure the remaining advisors have a minimum proven predicting value and that they provide a more accurate rating than when there are unreliable

advisors helping the user as well. As can be seen, there is a trade-off between having high coverage and high accuracy by choosing a correlation threshold value. In one study Gorner attempts to limit the number of advisors in a trust and reputation management system [8]. As Gorner mentions, we have to find an appropriate "sweet spot" that brings us the most reliable results. Therefore, he tries to find the parameters for the threshold value and the maximum number of neighbors that provide more accurate trust-modeling results. Based on their experimental results, he recommends the value 0.55 for the threshold value and 30 as the best value for the maximum number of advisors.

2.3.5   Trust and Reputation Models

As a Trust and Reputation model for Agent-based Virtual Organizations (*TRAVOS*), Teacy et al. propose two methods to evaluate the trustworthiness of agents [41]. As time passes, agents interact with each other and trust is built based on their experiences. Therefore, in the future agents can judge their target's trustworthiness considering their previous behavior, and decide whether to continue to interact with them. Although this scenario is unavoidable, there are numerous cases where the two parties of a business carried out in a marketplace are unknown to each other. In such cases, an agent cannot rely solely on their own judgment about the other party because it did not have any interaction with the other party. As a result, it has to ask for others' opinions about the other party's trustworthiness. The agent interprets reputation as the collective opinion that others have about the other party.

As discussed in the previous paragraph, this study considers two sources of trust for coming up with a final trust value for different agents. First, the agent (trustor) assesses the trust value in the trustee agent based on its own previous experiences through direct interaction with the trustee. Second, in case the trustor does not have any experience with the trustee, it asks others about their trust values in trustee and aggregates these values as the trusee's reputation.

This study considers the outcome of an interaction as a binary variable. An interaction is successful if the provider agent performs its objectives. Accumulating these interaction outcomes as experiences, agents use a probabilistic method to model other agents' trust values. However, using probability alone, the trust model cannot distinguish between cases where the agent has enough information and those where it has not. Therefore, the authors define another metric to overcome this shortage: *confidence* is the metric that each agent uses to measure if it has enough confidence in the trust value calculated or not. The more evidence one has, the more confident it will be.

This model sets a predetermined minimum confidence level. This is the minimum value of confidence that an agent should have to rely on the calculated trust value. If it has less than the predetermined minimum level of confidence, then the agent asks others for the trustee's reputation value. Obviously, other agents may not necessarily behave honestly and report the actual experience value. That is, among all the agents asked for their feedback about a certain agent there exist agents which are malicious, and do not give valid feedback. In order to tackle this problem, this study refers to a study by Jøsang, Ismail and Boyd [17] to describe two approaches: *endogenous* and *exogenous*. The

endogenous technique assumes that malicious users are in the minority and their votes deviate from the majority of reputation providers' opinions. In contrast, the authors define their solution to this problem as exogenous, that is, the trustworthiness of a reputation provider is assessed based on its previous opinions about other agents. The closer a reputation provider's opinion to the trustor's opinion about others, the higher the accuracy of its opinion about the current trustee will be.

In *REGRET* model introduced by Sabater and Sierra [38], the experience that users have after each interaction with others is called an impression. The authors define reputation and its formation as follows:

"These individual impressions are the bricks that properly combined may be used to build the reputation of other individuals".

There are different dimensions taken into account to calculate the reputation value in this study: Individual, Social, and Ontological. The Individual dimension of reputation populates the impression database of each agent. These impressions form the subjective reputation of other agents in the eyes of an agent. The authors of this study are the first to talk about the social dimension of different parties in a community. Obviously, as we can find in different studies, fields and resources, we judge people by their friends. This well-known fact plays an important role in this trust model. Different groups within a society are known for particular features and they have different levels of reputation. For example, charity communities or organizations are generally reputable for wholeheartedly helping people in need; therefore, they generally have a high reputation value among people. These communities are opposed to Mafia gangs that do not look benevolent in the general public's point of view. Meanwhile one may say that they are reputable for carrying out their own objectives, whatever they may be. Therefore, it depends on different groups of people how to observe and judge to give reputation value to a certain group. This study takes these facts into account and for the social dimension it defines four different types of interaction:

1. The trustor's experience of interaction with the trustee.

2. The trustor's experience of interaction with the trustee's group members.

3. The trustor's group members' experience with the trustee.

4. The trustor's group members' experience with the trustee's group members.

Finally, they argue that being a good service provider depends on different aspects of an agent's point of view. For instance, for one user, a good seller is one which provides a good-quality product, and the price and delivery time of that product are not that important. Therefore, when calculating the reputation value of a seller, such a user gives more weight to the quality of the delivered product rather than the price and delivery time.

The *FIRE* model [13] does not confine itself only to one specific source of trust, which can be direct interaction or witness reputation. Huynh, Jennings, and Shadbolt in this study define four types of trust and reputation and combine

them to get an overall picture of the likely performance of an agent.

Similar to other trust and reputation models, especially REGRET, this study takes previous interactions and their outcomes between agents as one of their sources of calculating trust values of users, named Interaction Trust (IT). This component of trust in this system is based on the direct trust that is discussed in the REGRET model. The reputation of agents is another form of input to this trust and reputation management system. The authors make use of the referral system in [6] and apply a variant of this method. In this article, this source of trust is known as Witness Reputation (WR). Based on different relationships agents may have in the real world, this system assigns a trust value for their roles in that relationship, Role-based Trust (RT). In other words, this component indicates how much we should expect another agent to carry out an interaction with respect to the relationship between two parties. For instance, as the article mentions, if the relationship between the two parties of an interaction is a general buyer-seller relationship, the buyer may expect the seller to sell a product with a slightly lower quality. Finally, as the authors argue, the main and most important contribution of this paper is the last model they define: Certified Reputation (CR). CR is a kind of trust source that involves users who can vouch for a trustee. These vouchers are selected by the trustee itself. These agents who vouch for a specific trustee should have previously interacted with them. Therefore, in addition to the previous three sources: IT, WR, and RT, the trustor asks these vouchers for their recommendations about the target agent, the trustee.

Needless to say, a rational agent will only ask those agents who will report the trustee's best ratings, so we have to keep in mind that this source of trust may overestimate the actual agent's performance. A reference for an applicant when applying for a university or a job is a good example of this type of trust. As mentioned before, the study integrates these four types of trust and reputation and comes up with a single trust and reputation model to generate the overall expected performance of a certain agent. The most obvious disadvantage of such a trust model that considers vouchers, where the target agent decides who those references should be, is that they are prone to coalition attacks.

Dellarocas takes into account the flaws of reputation systems that only work with the overall reputations of members inside a marketplace [4]. This paper proposes a model where only buyers can rate sellers, and it also lays out issues that may arise if bidirectional ratings exist. Reliable estimation of reputation is not a simple task as it is difficult to accumulate the ratings of all users. The differences in the ratings of users may stem from the differences in their tastes. This fact brings us to personalized reputation ratings. Based on the differences of buyers' ratings, the author defines two scenarios which result in a biased estimation of reputation. These scenarios mostly occur when malicious users (buyers and/or sellers) intentionally want to "rig the system".

Ballot stuffing: This scenario happens when some buyers rate sellers higher than what they deserve to help attract more customers. This takes place because the reputation system calculates the users' reputation based on all of the ratings, and these ratings, sometimes excessively deviating from the majority of ratings, can have a great impact on the target seller's reputation in general.

Bad-mouthing: Unlike the previous scenario where those buyers give high ratings to specific sellers, buyers in this case intentionally rate some sellers lower. In ballot stuffing, the buyers and sellers involved work in collusion to game the system and benefit from it. However, in cases of bad-mouthing, these sellers are competitors of these buyers' allies.

As his first proposed approach, Dellarocas offers the mechanism of controlled anonymity. In this method, buyers and sellers may not have an idea about each other's actual identity. This approach can potentially address the problem of bad-mouthing because a buyer does not know who exactly the other party of an interaction is. Therefore, it cannot give a seller a high rating with the intention of ballot stuffing that seller as it may not be the allied seller who has to be promoted. In contrast, it cannot bad-mouth an unknown seller as it may be one of its allies. However, this approach has several drawbacks and cannot be applied to all marketplaces.

First, not all domains can employ this approach. There exist domains in which it is not possible to hide the identities of users. Hotels and businesses that buyers and sellers interact with in person have to be present physically and cannot become anonymous. Even in other domains such a mechanism entails engagement of the marketplace as an intermediary to act on behalf of the seller. The marketplace itself has to carry out the duty of shipment so that the identity of the seller remains concealed to buyers. Second, in different circumstances buyers and sellers who want to collude, although it may be difficult, can find a way to signify their existence, so they can achieve their goals.

A very complicated procedure is required to make this mechanism work because the complexity provides avenues for cheaters to game the system. As an alternative approach to the problem of biased reputation rating, the notion of collaborative filtering has been utilized. This technique finds the nearest neighbor set N for the buyer b. The distance in this case is based on the similarity of the users. In this study, ratings given by the neighbors of an individual are divided into two clusters. The lower cluster N_l contains ratings which are fair and the upper cluster N_u includes unfair ratings.

2.3.6 Summary

With the sheer number of transactions carried out over the Internet, users seek a reliable source of information about the other party of their business transaction. As discussed in this chapter, several e-marketplaces make use of global reputation to help users distinguish between benevolent and malicious users. However, this approach comes with a few shortcomings and as a solution researchers propose considering local trust and advisory networks. Nevertheless, existing trust and reputation management systems do not take into account the dynamics of environments and prescribe a fixed value to aid the system distinguish malicious from benevolent users. Therefore, this thesis proposes approaches that remedy these deficiencies by means of a flexible trustworthiness threshold and bring more satisfactory transactions in an e-marketplace.

CHAPTER 3

PROPOSED APPROACH

As discussed in the background section, current online marketplaces employ the notion of global reputation for every user. This approach, as has been mentioned in several studies, despite its successes, has its own flaws. In this case all users play the role of advisors for each other. However, among these users there may be malicious users from whom other users should not receive advice. These issues stem from having a global reputation mechanism as a means to help the users of a system to have an overall idea about other entities' trustworthiness within the same system. To address this issue researchers in the field of TRM systems have come up with methods to help users gain more helpful information about other entities in a system by finding their reliable advisors. However, these systems have focused on evaluating the trustworthiness value, but what should be the threshold for trustworthy users is not discussed in the literature. In this study a systematic approach is proposed that enhances these systems by helping them find a way to decide about the trustworthy threshold that separates trustworthy advisors from malicious advisors.

Every individual within a society, and more specifically a marketplace, has different preferences. Therefore, when people want to try something they have not tried yet, or interact with a seller with whom they have not yet interacted, they tend to ask their friends for their recommendations about that specific item or seller. These friends are usually a subset of all of their friends that have two characteristics. First and foremost, they should have common interests so they can get suitable recommendations about the target entity, be it a product or a seller. In other words, if they do not have similarities in their tastes or preferences they cannot help each other wisely. Second, these friends should have already tried that product or have experience interacting with the designated seller. Obviously, if they do not have any idea about those entities, even if they have the same preferences, they cannot provide any useful information.

Following this natural procedure, the users of a marketplace should follow the same steps to have successful interactions so that their sources of information are not only trustworthy but also similar in their preferences.

As presented in the previous chapter, trust between two agents in TRM is modelled as a value which is usually a real number in the range of zero to one. A user with a trust value of one is considered to be completely honest and therefore trustworthy. Needless to say, a user with the trust value of zero is utterly dishonest. Evidently, the lower trustworthiness value a user has in this spectrum between zero and one, the more fraudulent that user is expected to be. In a marketplace, buyers have to find as advisors other reliable buyers in order to rely on their recommendations. In

order to do so, they model other users' trustworthiness values. These values give them an estimation of those buyers' reliability. Therefore, each user keeps a network of friends, or neighbors, who are in fact advisors for that user. When modeling other users' trustworthiness, based on different criteria, the judging user gives the target user a trust value in the range of [0, 1]. To form the advisory network, it has to answer the two following questions:

1. Whose trustworthiness should be modeled?

2. Who is qualified to become an advisor?

1. Assume a marketplace such as eBay, where a user examines a seller's reputation by taking into account the feedback from all users within that marketplace. Should an individual model the trustworthiness of all of the buyers in the system to make them part of their advisory network? This approach, unsurprisingly, has several shortcomings such as having high computation and time expenses to compute trust values. Moreover, one may not need all of those users as its advisors, as not all of them have adequate information to share about the target seller. The most appropriate buyers to model their trust are those who have interacted with the same sellers in the past.

2. In TRM systems that use agents, each user has an agent that helps finding reliable sources of interaction. When an agent models other buyers' trustworthiness it assigns a trust value to each of them. Using the assigned value, it has to decide whether the subject agent is eligible to be its advisor or not. In other words, who is trustworthy enough to become an advisor? To answer this question should define a minimum level of trustworthiness; this way, we can compare each agent's trust value with the minimum value to find out if the agent is trustworthy or not. Do we consider a user with a value of 0.7 as trustworthy? What about 0.5? Or 0.3? These are not simple questions that can be answered instantly and in one word. In some cases, in TRM systems, it is not clear how we should distinguish honest users from dishonest ones. Some researchers explicitly designate a fixed value as the threshold between these two categories of users. For instance, the study by Massa and Avesani [24] regards users with a trust value of less than 0.6 as untrustworthy, so the model discards their recommendation about other items. As can clearly be seen, different trust and reputation models consider different thresholds. Moreover, if two different TRM systems with different environments adopt the same trust and reputation model, they will not necessarily achieve the same result. Having the same settings for a trust model in different domains and environments is a naïve approach to address the issue of detecting untrustworthy users in that environment. In other words, having a fixed threshold for distinguishing malicious users from benevolent ones may not lead to the same result in two different environments.

As can be seen, different environments entail different trustworthiness thresholds. However, even in a specific environment, how can we make sure that the designated threshold suits that environment?
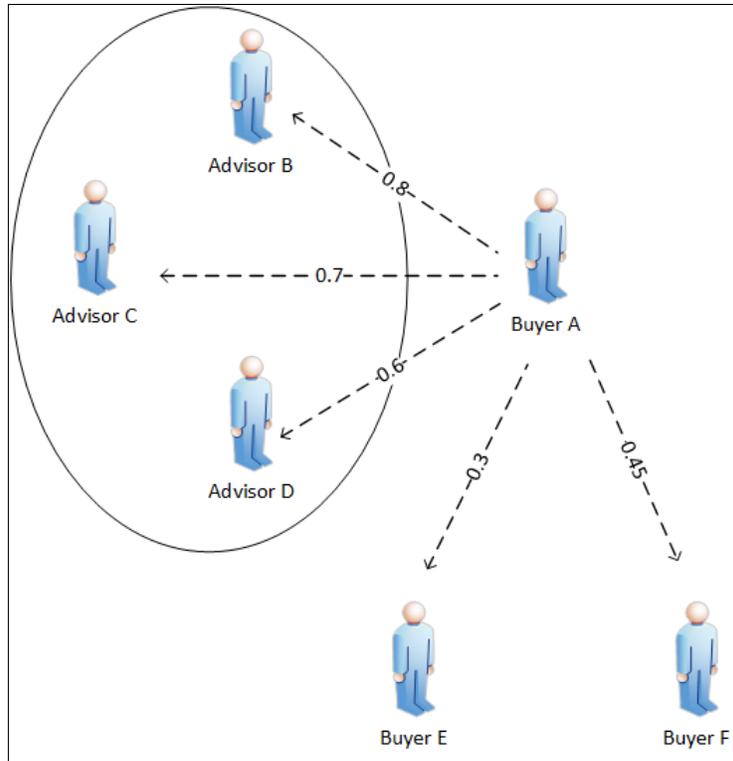
Figure 3.1: Advisory network with trustworthiness value of 0.55

As discussed before, users prefer to receive recommendations about other entities within the system from users who are similar to them and are trustworthy enough. Each of these networks is a community of users with similar perspectives and interests, and each user in a community asks for suggestions from their neighbors (i.e. advisors) in the same community.

As a result of these questions and answers, we can reach the conclusion that a marketplace needs to have a dynamic, rather than fixed, trustworthiness threshold to adapt itself to different environments. Inspired by the techniques used in the field of control systems, we propose the idea of feedback controllers.

In order to have a better idea about how a dynamic trustworthiness threshold works, consider Figure 3.1. Suppose Buyer A wants to form his network of advisors among the buyers B, C, D, E, and F. The numbers shown on the links between Buyer A and the other buyers, who are the potential advisors for Buyer A, show how much he trusts each of them. For instance, the trust of buyer in advisors B, C, and D is 0.8, 0.6, and 0.7, respectively. Obviously, the buyer's trust in Buyer E and F are 0.3 and 0.45, respectively. The reason why the Buyer A does not include Buyers E and F in the network of advisors is simple: he has a low value of trust in them. This means that the trust that a buyer has in other buyers needs to be higher than the threshold defined in the system for these buyers to be assigned as advisors for him. The value of the trustworthiness threshold in Figure 3.1 should be less than 0.6 and higher than 0.45, say 0.55, because it has included links with values higher than 0.5 and excluded the ones with lower than this value. In order
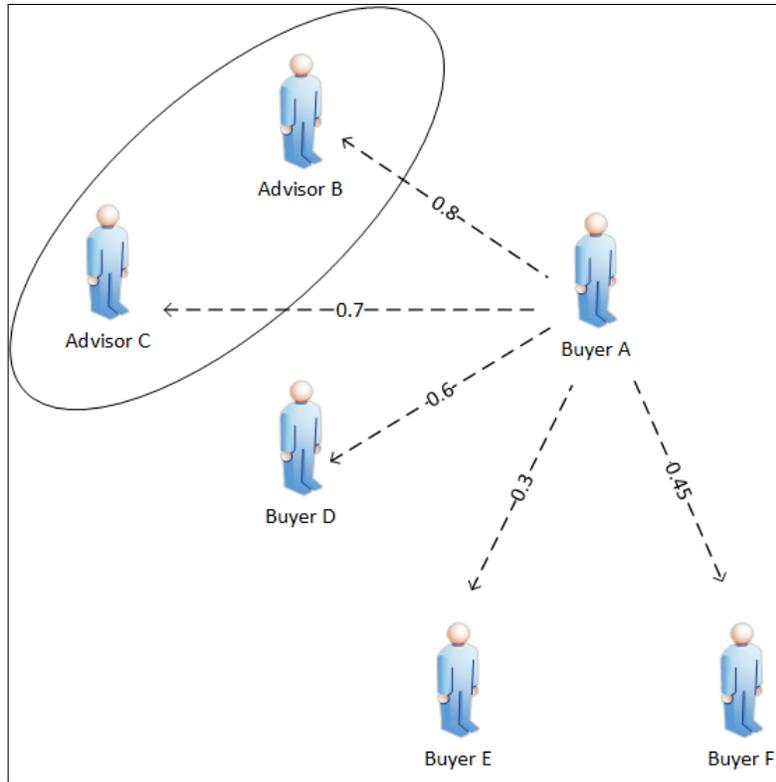
Figure 3.2: Advisory network with trustworthiness value of 0.65

to show the effect of change in the trustworthiness threshold, (see Figure 3.2) if the value of trustworthiness threshold increases from 0.55 to 0.65. Advisor B whom the Buyer A trusted only for 0.6 would be expelled from his advisory network. In other words, the advisor would not be considered trustworthy enough to advise the user anymore.

Figure 3.3 shows the structure of the place of the threshold computation layer proposed in this thesis in the architecture of an e-marketplace. Currently, the field of trust and reputation management systems is the only layer, interacting directly with data from the e-marketplace. However, this thesis proposes to add another layer on top of the TRM system: Feedback control system using Proportional-Integral-Derivative (PID) controller. As discussed in the examples above, different values of trustworthiness threshold cause buyers to have different number of advisors. The duty of the new added layer introduced in this thesis is to help the lower layer, the Trust and Reputation Management System, to choose a suitable trustworthiness threshold. This layer forms a feedback control system in the e-marketplace – it gauges the system's performance status and modifies the trustworthiness threshold used by the TRM system if needed.

3.1 PID Controller

Feedback control systems have been around for more than two millennia, as these systems have been used to adjust fluid levels and make these levels constant. The systems that utilize feedback control systems range from air conditioners to
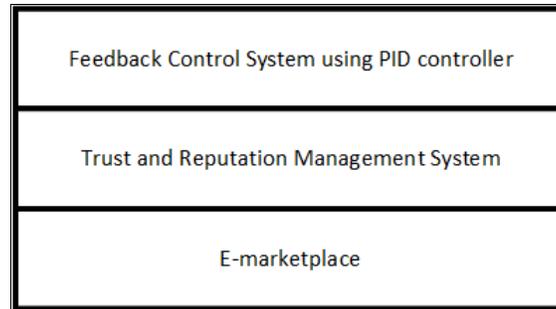
16

Figure 3.3: E-marketplace system layers

maintain the space temperature to aircrafts to maintain flight path and altitude. One of the most prominent examples of a feedback control system in everyday life are the termostat and the cruise control used in modern vehicles. The cruise controller, for example, has the goal to make the car drive at a desired reference speed.

The rationale behind feedback control systems is simple: they make use of feedback loops to gauge the system's current situation and change the relevant parameters within the system that help it reach and maintain the desired situation. These systems have a special component that modifies the process of the system and changes its functionality. This special component is called controller. In the literature of feedback control systems several controllers have been introduced and PID controller in one of them that this study utilizes to enhance existing TRM systems.

The idea of feedback control systems is observable in traditional coolers which use a floater to sense the level of water. According to the sensed level, the system lets in more water if this level is less than the desired one, and blocks the water's path to the container if the water has reached the designated level. These systems are rather intuitively designed. Almost two hundred years ago, engineers modernized this area as they introduced mathematically-defined control systems.

The Figure 3.4 shows the components and their relationships. Each feedback control system should have four components [11][31]: Reference (Setpoint), Controller, Process, and Feedback (Sensor).

As can be seen in Figure 3.4, the result of the Process component is shown as Output. Output is the current state of the system. In order to gauge this output we need to have a component which can sense the environment. Feedback is a part of the system which can measure the current state of the system's performance. Therefore, the input of this component would be the output of the system. This component is nothing but sensors which can sense the output of the system. When the output sensed by the Feedback component it will be compared with the Reference, which defines the goal of the system, and it will be the input of the Controller. In other words, Controller gets the difference between the desired goal (Reference) and the actual output (Feedback's output). Controller, as the name suggests, has to control the system. To do so, it has to control influential parameters in the system that change its output. The output of this unit is the input of the Process component. Using the suggestions of Controller, the Process functions with the changed parameters. Simply, the output of this component is the output of the system. This output is again measured by the
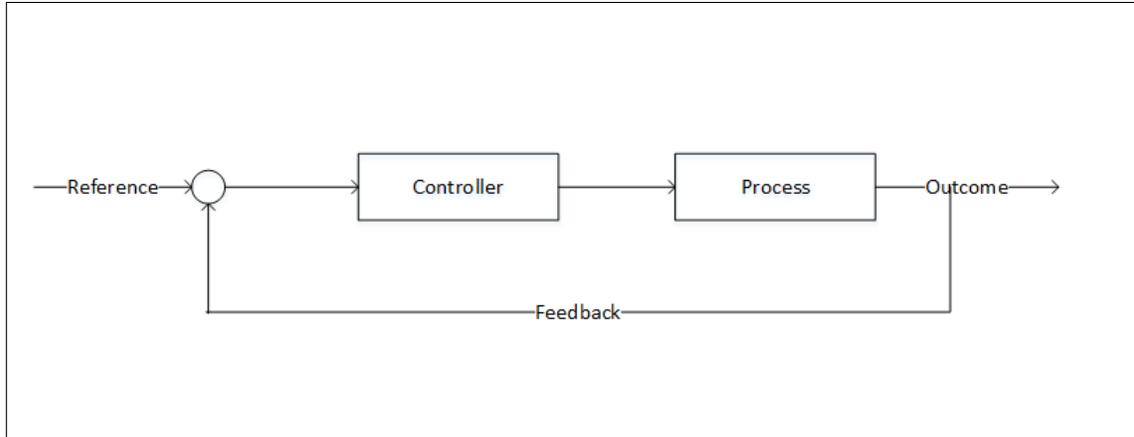
Figure 3.4: Feedback control system

Feedback unit and the cycle of procedures mentioned above is repeated until the sensed output equals the Reference value defined as the goal of the system.

Figure 3.5 takes a closer look at the relationships between different components in a feedback control system. As mentioned earlier, the starting point of the feedback control system is the reference which is defined as r in this figure. The output of the system is defined as the *Process Variable y*. The difference between process variable and reference is the *Error*, e in the system: $e = r - y$. As discussed above, the system has to refine itself using the feedback loop so that its output approaches the reference defined as its setpoint or goal. In other words, the larger the error, the farther the system is from its goal. In order to go towards its reference it has to control the system via parameters that affect the output. This parameter in these systems is called Control variable u. Modifying control variable, the process component functions differently and the output of the system changes compared with the time control variable had different values. The output of the system is again sensed by the feedback component and this value is compared with the reference and error is calculated. Based on the new value of error controller again changes the control variable and passes it to the process component.

As discussed earlier, a controller in a feedback control system gets the error in the system and tells the system how to react to the situation so that it can approach the value of reference. Different controllers have been introduced in the literature that act differently according to the error value. PID controller is one of the most widely used controllers used in industrial control systems. The Equation 3.1 shows how this controller takes into account the error in the system to suggest the corrective action.

$$u(t) = k_p e(t) + k_i \int_0^t e(\tau)d\tau + k_d \frac{de(t)}{dt} \tag{3.1}$$

Where k_p, k_i, and k_d are the coefficients that leverage the contribution of *Proportional* **P**, which captures the error $e(t)$ calculated in the time stamp t, *Integral* **I**, which accumulates all errors from the start of the e-marketplace, and
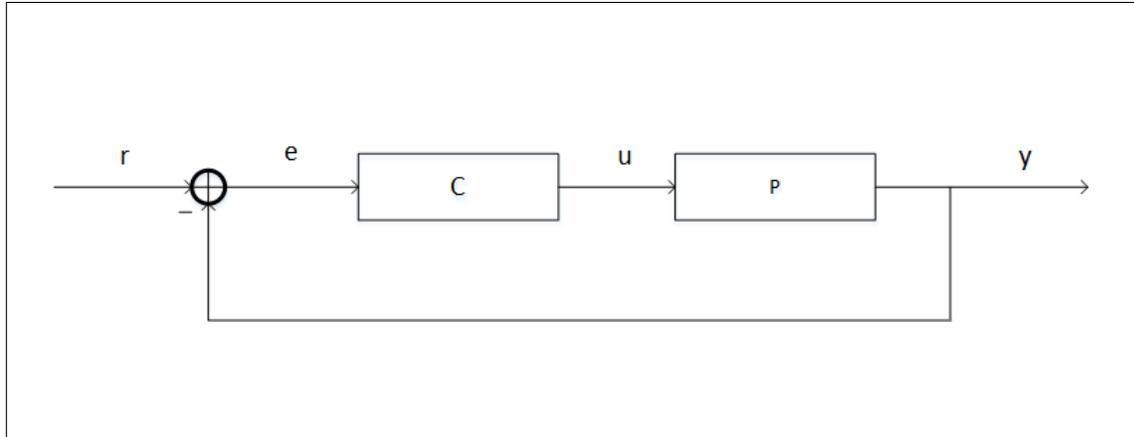
Figure 3.5: Feedback control system variables

Derivative **D**, which calculates the deviation of current error $e(t)$ from its previous value $e(t-1)$, respectively.

Each system that wants to adopt a feedback control system has to define the variables explained above. In a cruise controller system, the speed of the vehicle is the process variable or output of the system. Now if we want to find what would be the control variable we only have to think which parameter in a vehicle helps it change its speed: Pressing gas or brake pedal. The more we press the gas pedal the car gains more speed. Therefore, the control variable in this case would be pressing the gas/brake pedal. Suppose we set the cruise control's speed to 100 Kilometers per hour (km/h). So 100 km/h is the reference or r in the system. If the speed of the vehicle is 30 km/h, the e of the system will be 70 km/h since $e = (100 - 30)km/h$. And in this case the controller will advises the system to push the gas pedal more to reach the speed of 100 km/h.

In this study an e-marketplace that uses a TRM system is considered as a system which can be improved utilizing a feedback control system. In order to do so we have to have the components of such a system and define different variables each feedback control system entails. That is, we need to find which parameters in an e-marketplace can resemble the control and process variables and which parts can act as components defined in a feedback control system. The following paragraphs discuss the definition of each of these variables and components within an e-marketplace.

In order to adapt the trustworthiness threshold based on the global state of the marketplace, the model adopts the feedback controller that takes into account factors defined in [2][9][34] that contribute to the performance of an e-marketplace.

1. *Market liquidity* defines how successful a marketplace is in facilitating trading in the system such that the sellers do not reduce their products' prices noticeably enough to be able to sell them. Moreover, this factor takes into account buyers' ability to easily find their desired products. However, the existence of different products with competing features and a lack of honesty enforcement make buyers unsure about the best partner for the potential transactions. These factors affect the liquidity of a marketplace.

2. *Information asymmetry* describes whether a buyer has enough knowledge to make decisions about its purchases in the marketplace. It is imperative to have sufficient information about sellers and products in online marketplaces as they have more options than in the real world, and of course they have to deal with different situations. Here is where we can see the importance of having credible advisors to reduce the information asymmetry in a marketplace.

3. *Buyers' satisfaction* measures the extent to which a buyer is satisfied with its purchases within the marketplace. This factor can easily be measured by dividing all the successful transactions a buyer has had by the total number of transactions it has carried out.

If we consider $SuccessNumber(c)$ as the number of satisfactory transactions consumer c has achieved, and $Transaction-Number(c)$ as the total number of transactions of c, then the buyer's satisfaction is as follows, which we name Transaction Success Rate of c or $TSR(c)$ of a consumer:

$$TSR(c) = \frac{SuccessNumber(c)}{TransactionNumber(c)} \qquad (3.2)$$

Furthermore, we interpret $IntendedNumber(c)$ as the number of transactions that consumer c initially planned to conduct in the marketplace. Therefore, we introduce another variable as Transaction Rate of c or $TR(c)$ as follows:

$$TR(c) = \frac{TransactionNumber(c)}{IntendedNumber(c)} \qquad (3.3)$$

Throughout the users' existence in the system they should adjust their advisory network to have the appropriate recommendations. This procedure entails having a complete overview on the overall performance of the whole system. Suppose a marketplace does not perform well in respect to users' satisfaction. This situation shows that deceitful users could prosper and deceive other users so that users would not have satisfactory transactions. As a result, the central server asks users to examine their network of advisors and make sure their advisors qualify according to the credibility threshold. That is, users have to acquire more credible advisors by increasing their own credibility threshold. In this study we refer to credibility threshold as β.

To aid users' ability to modify their credibility threshold, the marketplace as a whole has to have a full overview of its performance. In order to gain this view, the central server asks consumers to report their TSRs and TRs. These two components together help the marketplace to determine the current state of its performance. Assume that users in a marketplace have high values of TRs. We should not conclude that the marketplace is at its peak. As mentioned in the equation 3.3, this value is simply the division of transaction numbers by the intended number of purchases, and does not tell us much about how satisfactory those transactions were. On the other hand, the TSR value alone may not provide enough information about the performance of the marketplace because the high value of this variable does not

necessarily mean that users in the system could easily find their desired products. It just indicates that out of all the transactions carried out, no matter how low this value may be, a high percentage have been successful. Therefore, the performance of the marketplace is calculated based on both of them as follows:

$$Q(t) = \frac{2 * TSR(t) * TR(t)}{TSR(t) + TR(t)} \tag{3.4}$$

Where $TSR(t) = \frac{\sum_{i=1}^{n} TSR(c_i)}{n}$ and $TR(t) = \frac{\sum_{i=1}^{n} TR(c_i)}{n}$ are the average of all $TSR(c)$ and $TR(c)$ shared by buyers at time stamp t, and Q(t) is the *harmonic mean* of the e-commerce quality metrics described above. Since the performance of the marketplace is a function of these quality metrics, we use a harmonic mean to balance them by mitigating the impact of the one with a larger value and aggravating the impact of the other with a lower value.

To adjust β accordingly, the server adopts the idea of feedback controller, specifically, PID controller [11][31].

In e-commerce systems, the ultimate goal is to maximize the performance of marketplaces in terms of buyers' satisfaction degree and market liquidity, achieving $Q(t) = 1$, so we initialize the goal r to $r = 1$. We designate error, $e(t)$, in the e-commerce system as the difference between the actual performance of the system $Q(t)$ and the goal r which is $e(t) = r - Q(t)$.

In the ideal e-commerce systems in which no malicious buyers exist $Q(t)$ could converge to one. However, in a realistic situation where the marketplace is populated with different participants with various behavioral dispositions, it is not reasonable to expect the perfect performance of the system; therefore, the system will have $Q(t) < 1$.

Given these values, the server calculates a new value for β that improves $Q(t)$ to reach the idealistic goal $r = 1$. To this end, the server incorporates PID controller to determine the extent that it has to change the value of β.

The new recommended value of β for the next time stamp $t + 1$ is formulated as follows:

$$\beta(t + 1) = \beta(t) + \beta_0(t + 1) \tag{3.5}$$

in which $\beta_0(t + 1)$ is formalized using the PID controller presented as,

$$\beta_0(t + 1) = k_p e(t) + k_i \int_0^t e(\tau)d\tau + k_d \frac{de(t)}{dt} \tag{3.6}$$

Where k_p, k_i, and k_d are the coefficients that leverage the contribution of *Proportional* **P**, which captures the error $e(t)$ calculated in the time stamp t, *Integral* **I**, which accumulates all errors from the start of the e-marketplace, and *Derivative* **D**, which calculates the deviation of current error $e(t)$ from its previous value $e(t - 1)$, respectively.

Since in the e-marketplace it is unrealistic to expect $Q(t)$ reaches the value of r (due to the activity of malicious participants), the server would stop adjusting β if $Q(t)$ reaches a stable point. More formally, server updates the value
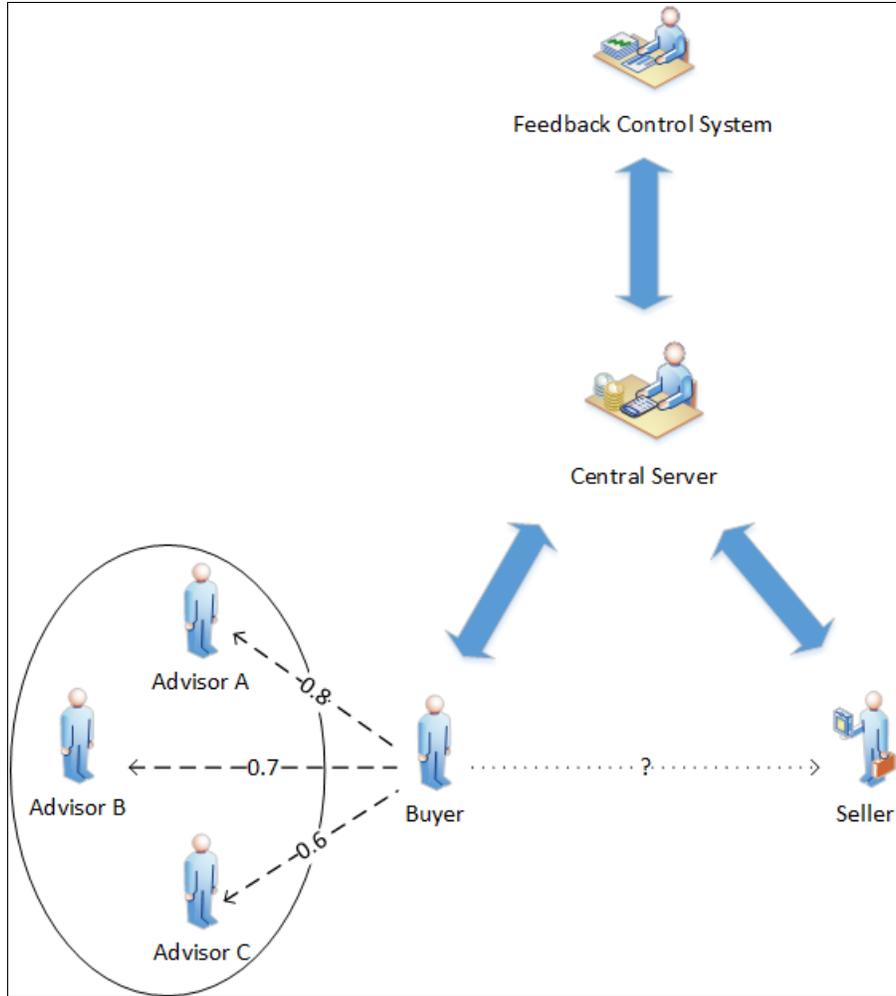
Figure 3.6: System Architecture

of β for the next time stamp $t + 1$, given the following conditions:

$$\beta(t + 1) = \begin{cases} \beta(t) + \beta_0(t + 1) & |Q(t) - Q(t - 1)| > \sigma \\ \beta(t) & \text{otherwise} \end{cases} \tag{3.7}$$

Where σ is a trigger threshold.

3.2   System Architecture

The proposed system is composed of the e-marketplace central server, consumer agents and provider (seller) agents. These agents have the responsibility of taking action on behalf of their owners. In other words, a consumer agent has to find the most trustworthy providers of a certain product and introduce them to the end user, the owner of the

agent. Agents communicate with one another and the central server to fulfill their needs within the system. The general architecture of the system can be seen in Figure 3.6.

Communications in the system are as follows:

1. Among consumer agents: This type of interaction is divided into two forms. These agents interact with each other at different times of their appearance in the system. First, in order to form their advisory network, each agent models other agents' trust values using a trust and reputation model. The other reason why these agents communicate with each other is the main rationale behind having a reputation system. They ask other agents about their opinions about the consumers they have had interactions with. These agents are neighbors (advisors) of a certain agent. As a result, an agent's advisors provide it with the trust values sellers have from their point of view.

2. Consumer agents and central server: A consumer agent sends a query to the server and requests information about consumers who have had interaction with a set of sellers. These requests and responses happen when an agent plans to form its advisory network. The set of sellers that it asks about is the set of sellers that the agent's owner has interaction experiences with. Therefore, agents introduced by the central server are those that have common previous providers with the inquirer agent. Other times, when these entities (the central server and consumer agents) interact it is for informational reasons. Consumer agents notify the server about their interaction outcomes that are used for further calculations. The server notifies agents about the current state of the marketplace environment and these messages are mostly cautionary.

3. Consumer agents and seller agents: When decided, a buyer communicates with a seller to purchase its product. Based on its quality of service, the seller responds to the request.

4. Seller agents and the central server: Seller agents notify the server about consumers with which they have had interactions recently.

In Figure 3.6, there is another component that we did not mention above which is the feedback control system. This part is an independent component that helps the central server improve the satisfaction from transactions in the marketplace. This part considers the factors defined earlier in this chapter and modifies the variable that influences the performance of the system: the trustworthiness threshold β. In this figure, the value of β is less than 0.6 since the Buyer's advisors have trust values more than 0.6. The feedback control system is the component that uses PID controller to help the TRM system find a more suitable β value. That is, if in this example buyers need more accurate recommendations they have to limit their advisory network by filtering out those who have lower values of trust. Therefore, if the value of β increases in the system and reaches 0.65, the Buyer in Figure 3.6 will expel Advisor B from the advisory network.

CHAPTER 4

EXPERIMENTS AND EVALUATION

In order to verify and evaluate the proposed approaches, two experiments are carried out. In the first experiment a simulated e-marketplace is used to evaluate the proposed approach which is populated by a number of sellers and buyers. In the second experiment, real-world data is used. In this case the dataset of an existing website, Yelp.com, is used on which users provide feedback about businesses.

4.1 Credibility Evaluation Mechanism

The TRM system used in the experiments adopts a variation of the *Prob-Cog* model [28], and formalizes the credibility degree of advisers in different steps.

Suppose that a buyer c sends a query to advisers requesting information about sellers $P = \{p_1, p_2, .., p_j, .., p_m\}$ on the outcomes of the interactions between the advisers and sellers occurring within a time threshold t (which diminishes the risk of changeability in sellers' behaviour). Adviser a_i responds by providing a rating vector R_{ij} for each seller, for example p_j. It contains a tuple $\langle r, s \rangle$, which indicates the number of successful (r) and unsuccessful (s) interaction outcomes with seller p_j respectively. Once the evidence is received, for each R_{ij}, buyer c calculates the expected value of the probability of a positive outcome ($P_r(R_{ij})$) for seller p_j based on a beta distribution [16] as follows:

$$P_r(R_{ij}) = \frac{r+1}{r+s+2} \tag{4.1}$$

Clearly, $0 < P_r(R_{ij}) < 1$ and as it approaches 0 or 1, it indicates *unanimity* in the body of evidence [45]. That is, particularly large values of s or r provide better intuition about an overall tendency and quality of sellers. In contrast, $P_r(R_{ij}) \approx 0.5$ (i.e. $r \approx s$) signifies the maximal conflict in gathered evidence, resulting in increasing the uncertainty in determining the quality of sellers. Based on these intuitions, we are able to calculate the degree of reliability and certainty of ratings provided by advisers. More formally, let x represent the probability of a successful outcome for a certain seller. Based on the Definitions (1) and (3) in [45], the *reliability degree* of each R_{ij} can be defined as follows:

$$Conf(R_{ij}) = \frac{1}{2} \int_0^1 \left| \frac{x^r(1-x)^s}{\int_0^1 x^r(1-x)^s \, dx} - 1 \right| \, dx \tag{4.2}$$

Theoretical analysis [45] demonstrates that, for a fixed ratio of positive and negative observations, the reliability increases as the number of observations increases. On the contrary, given a fixed number of observations, as the extent of conflict increases, the reliability of the provided observations decreases accordingly. That is, reliability is at a minimum when $P_r(R_{ij}) = 0.5$. As such, the less conflict in their ratings, the more reliable the advisers would be.

However, buyer c should not strictly judge the advisers with rather low reliability in their R_{ij} as deceptive advisers since this reliability factor could signify both the dishonesty of advisers and the dynamic and fraudulent behaviour of sellers reported by the advisers. For example, some malicious sellers may supply satisfactory quality of products in some situations when there is not much at stake and act conversely in other occasions associated with a large gain.

To address this ambiguity, buyer c computes $P_r(R_{cj})$ and $Conf(R_{cj})$ based on her personal experience, R_{cj}, with a set of sellers P with whom the advisers also have experience.[1] Through the comparison of advisers' metrics with the buyer's experience, the buyer would have more trust in those advisers with a similar rating pattern and satisfactory level of honesty. More formally, buyer c measures an average level of dishonesty of a_i by:

$$D_h(a_i) = \frac{\sum_{j=1}^{|P|} \mid P_r(R_{cj}) - P_r(R_{ij}) \mid}{|P|} \tag{4.3}$$

It may also happen that an honest adviser lacks experience with sellers. Thus, despite her inherent honesty, its reliability degree is low and it should not be highly trusted. To address this, we introduce an uncertainty function $U_n(a_i)$ to capture the intuition of information imbalance between c and a_i as follows:

$$U_n(a_i) = \frac{\sum_{j=1}^{|P|} \mid Conf(R_{cj}) - Conf(R_{ij}) \mid}{|P|} \tag{4.4}$$

Given the level of dishonesty of adviser a_i, the honesty of the adviser could be calculated as $1 - D_h(a_i)$. Similarly, given the uncertainty of adviser a_i, the certainty of the adviser would be $1 - U_n(a_i)$. Thus, a credible adviser should achieve higher honesty and certainty simultaneously. The *credibility degree* of adviser a_i is then calculated by reducing her honesty based on her certainty degree as follows:

$$CR(a_i) = (1 - D_h(a_i)) \times (1 - U_n(a_i)) \tag{4.5}$$

To retain only the most trustworthy advisers, an *credibility threshold*, β where $0 \leqslant \beta \leqslant 1$, is used to determine behavioral patterns of advisers. That is, if $CR(a_i) \geq \beta$, a_i will be counted as a *credible* adviser. In contrast, if $CR(a_i) < \beta$, a_i will be detected as *malicious* adviser and would be filtered out from the buyer c's advisers network.

---

[1]Here, we choose a set of sellers $P \subset \{p_1, ..., p_m\}$ with whom buyer c has sufficient experience, to make sure that the buyer has sufficient knowledge to judge the advisers.

4.2 Simulated e-marketplace

To show the efficacy of the approach proposed in this thesis, a sample of an online marketplace is simulated. The e-marketplace environment used for experiments is populated with self-interested buyers and sellers, and is operated for 20 days. The e-marketplace is initialized with 100 buyers in total, each of which has a maximum of 5 requests every day. For two reasons we used this number of users for this experiment. First, the purpose of this experiment is to prove the positive outcome of using the proposed approach. A small sample of a marketplace can demonstrate this fact. In addition, as has been mentioned throughout the thesis, the proposed approach performs as a layer on top of trust and reputation management systems in an e-marketplace and its responsibility is to improve these systems. Since TRM systems assign advisers to each user, and to enhance their performance these systems use PID controller to adjust the value of trustworthiness threshold for filtering these advisers networks. Therefore, number of users in the e-marketplace does not have any effect on the efficacy of utilizing PID controller in such systems. Also having this size of community for e-marketplace to prove the proposed approaches is known and used in the literature. For instance, this fact can be found in the works of several researchers such as Zhang in [49], Gorner in [7], and Noorian in [28].

Buyers (advisors) are divided into two groups: honest buyers (ones with high credibility), and dishonest buyers (ones with low credibility).

Honest advisors generate ratings that differs at most by 0.2 points from their actual ratings. In contrast, dishonest advisors generate ratings that differs at least by 0.2 points from the actual experience. For example, if the seller's Quality of Service (QoS) value was 0.9, then the honest advisor would generate a value between (0.7 and 0.9), and dishonest advisor would generate a value between (0.1 and 0.69).

It is assumed that there exist 80 sellers and 20 product types and every 4 of the sellers supply products with the same features. Sellers offer same price for products. It is further assumed the utility of each product is a value randomly distributed within [50, 70] for all sellers. Half of sellers, who supply the same kind of product, are high-performance with QoS values in the range (0.8-1.0). On the contrary, low-performance sellers generate QoS value in a range of (0-0.2). For example, if the seller's QoS is 0.3, the utility of its product is 60 and the price is 5, a buyer's actual profit of carrying out a transaction with that seller would be (0.3*60)-5=12.

A buyer, e.g. c, calculates the trustworthiness of sellers e.g. p_j through weighted aggregation of advisors ratings, $r_{(a_i)}$, with its own recent experiences $r_{(}(c))$, presented as follows:

$$\tau_{(p_j)} = \omega.r_{(c)} + (1 - \omega)\frac{\sum_{i=1}^{n} CR_{(a_i)} * r_{(a_i)}}{\sum_{k=i}^{n} CR_{(a_i)}} \tag{4.6}$$

Where $CR_{(x)}$ is the credibility of advisor x. Buyers subjectively decide to conduct a transaction if $\tau_{(p_j)} > T$ where T indicates the transaction threshold. The threshold T is set to be 0.6. Also, ω is the extent to which we want an agent rely on its own experience.

In this experiment, the credibility degree of advisors is calculated through the model presented in [29]. However, other credibility evaluation approaches can be used instead.

The experiments are conducted in two settings where different groups of buyers populate different percentage of e-marketplaces: 1) balanced environment where 50% of buyers are honest and 50% of them are malicious, and 2) dishonest majority where the number of dishonest buyers exceeds that of honest ones. The inequalities in buyer behaviors are set to be significant (a 75-25 ratio imbalance is used).

The e-marketplace operates for a period of 20 days and the reported results for each day are the average of 10 runs.

4.2.1 Experimental Results

In the following, comparative experiments are carried out to evaluate the performance of the e-marketplace in different environmental settings, adopting two different approaches: 1) the fixed β=0.5 approach and 2) the PID-based β approach.

The efficacy of two approaches in augmenting the performance of the electronic marketplace are examined across different dimensions:

- market liquidity

- information asymmetry

- buyers' satisfaction

First the market liquidity is measured by examining the transaction rate of different groups of buyers. Upon arrival, buyers randomly select sellers based on their promised utility (up to round 2). After acquiring sufficient experiences they establish their social network of trustworthy advisors, adopting different credibility threshold approaches: 1) the fixed β and 2) the PID-based β, which is initialized to 0.5. Given the initial setting of β, buyers have a similar transaction rate in initial days. However, as can be observed, as β increases, the transaction rate of the honest buyers increases while the transaction rate of dishonest advisors decreases Figure 4.1b.

From Figure 4.1, we notice that in both credibility threshold management approaches, honest buyers have higher transaction rates compared to the dishonest ones. However, comparative results indicate that in a PID-based β honest buyers have higher transaction rate than their counterparts in the fixed β approach. The adaptive approach of the server in adjusting β, based on the quality of marketplaces, results in 1) increase of honest buyers' transaction rate, and 2) detection and isolation of more dishonest advisors.

The adaptive adjustment of β is especially important when the majority of buyers are malicious. From Figure 4.2, we notice that in this environment, dishonest buyers in the case with fixed β have much higher transaction rate (Figure 4.2a) than their counterparts in the PID-based β case (Figure 4.2b). On the contrary, honest buyers have much
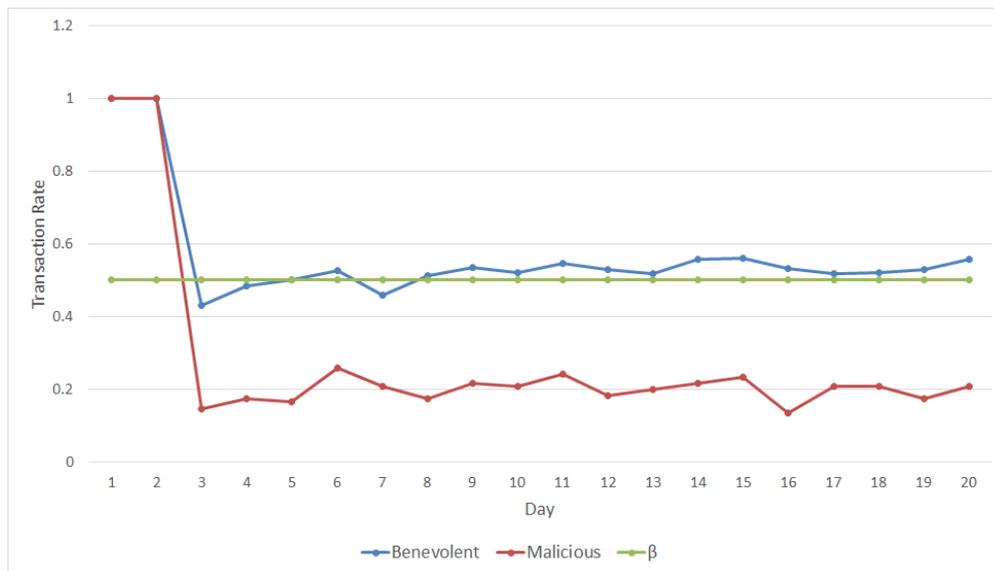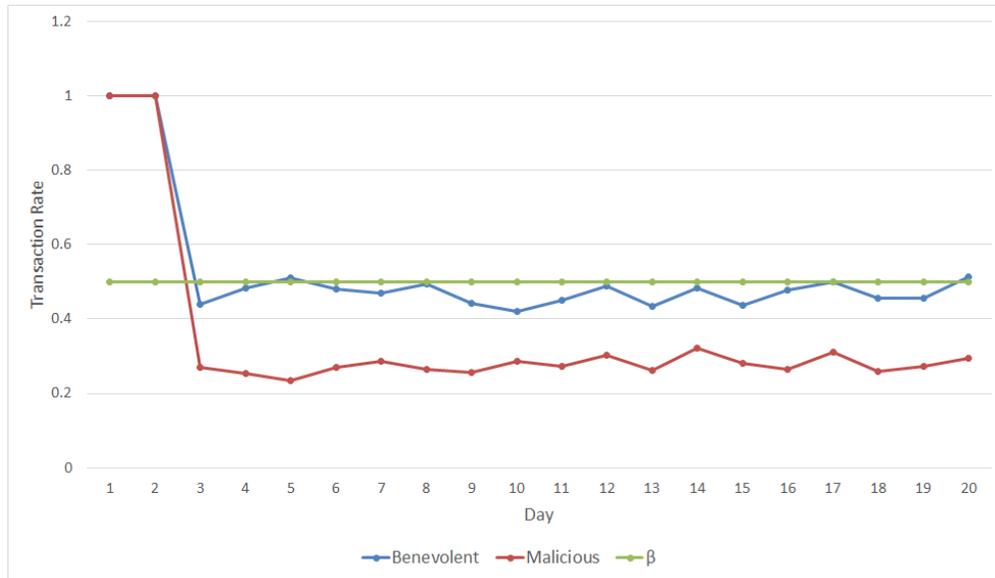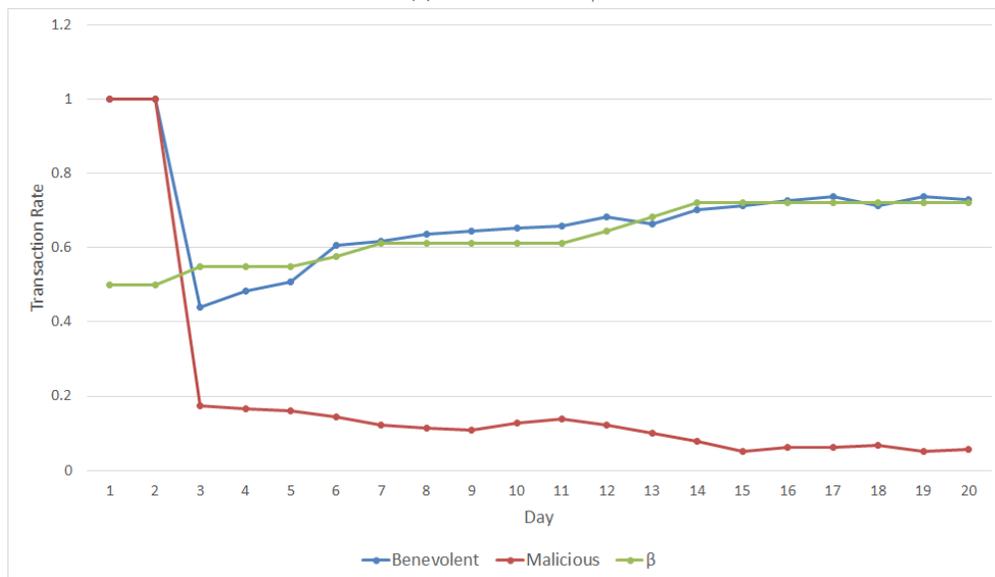
(a) Users use fixed β
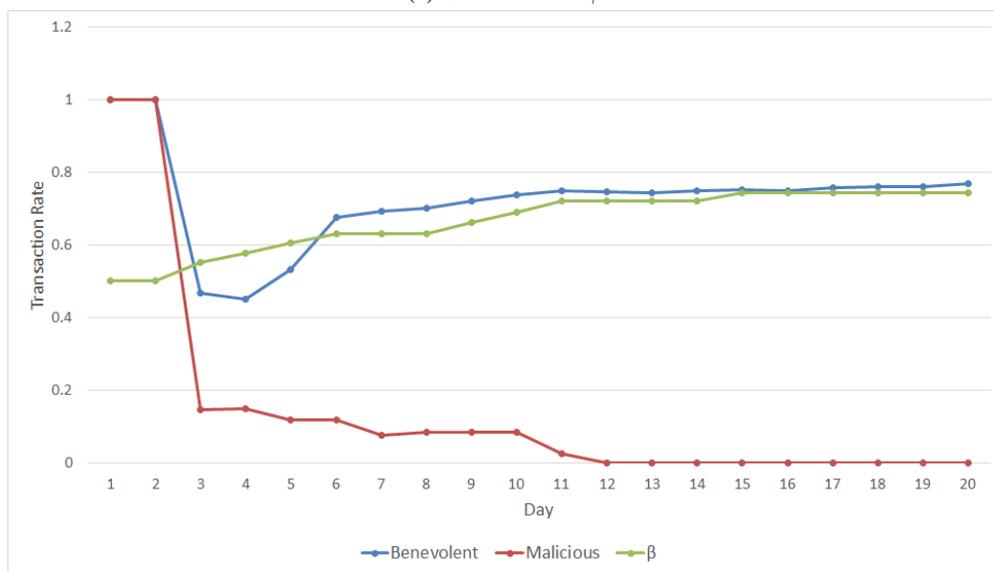
(b) Users use PID-baded β

Figure 4.1: The transaction rate of buyers in a balanced e-marketplace

(a) Users use fixed β

(b) Users use PID-baded β

Figure 4.2: The transaction rate of buyers when majority of buyers are dishonest

lower transaction rate in the fixed β than in the PID-based β case. The reason is that in the approach using fixed β a large number of dishonest advisors remain undetected and they continue to mislead buyers in their decision making process, impeding them in finding and conducting transactions with good sellers.

As can be seen in Figures (4.1b, 4.2b), even though the value of β gradually increases and the dishonest advisors are mostly filtered away, the honest buyers cannot conduct all the transactions they initially intended (i.e. $tr(c) < 1$). This is due to a lack of experience of buyers and advisors with the sellers that they intend to make transactions with.

The level of information asymmetry in the e-marketplace is measure by the accuracy of buyers in classifying their advisors. As shown in Figure 4.3, the accuracy of buyers in a PID-based β improves consistently and reaches the optimal value as they adaptively re-evaluate their network of advisors based on a new recommended value of β. On the other hand, although the static approach with fixed β shows good classification accuracy, it cannot further adapt to the incurring changes of the environment. This matter is clearly illustrated in Figure 4.3. When the majority of participants turns to be dishonest, the classification performance of buyers with fixed β drops considerably.

In addition, by comparing the metric of precision for buyers in different environmental conditions we notice that in the fixed β approach, as many dishonest advisors are inaccurately classified as honest ones, buyers significantly rely on dishonest advisors' feedback in their decision making.

The problem is aggravated when the number of dishonest buyers exceeds the number of honest ones in the marketplace (Figure 4.3b). The performance measures (i.e. precision) reflects the ineffectiveness of evaluation mechanism (with a fixed β) in detecting malicious advisors. Therefore buyers are better off to make a random decision on finding their transaction partners instead of relying on advisor's feedback identified by such credibility evaluation mechanisms. On the contrary, dynamically monitoring and tuning β enables buyers to achieve fairly good precision value hence undermining the impact of dishonest advisors.

Note that high precision and accuracy values describe the situation where buyers can access honest feedback, which implies the e-marketplace with low level of information asymmetry.

In the following paragraphs the buyers' satisfaction in the simulated e-marketplace is evaluated. In order to measure buyers satisfaction rate we compare the transaction rate and transaction success rate. We observe that in different environmental conditions, honest buyers provided with the PID-based β conduct more successful transactions (Figures 4.4 and 4.5).

Results of a balanced e-marketplace (Figure 4.4) suggest that in both credibility threshold management approaches, honest buyers have higher transaction success rates compared to the dishonest ones. However, comparative results indicate that in a PID-based β honest buyers conduct more satisfactory transaction, and dishonest buyers conduct less satisfactory transaction than their counterparts in the fixed β approach.

Alternatively, in the environment where the majority of buyers are malicious, dishonest buyers in the fixed β have much higher successful transaction (Figures 4.5a) than their counterparts in the PID-based β case (Figure 4.5b).
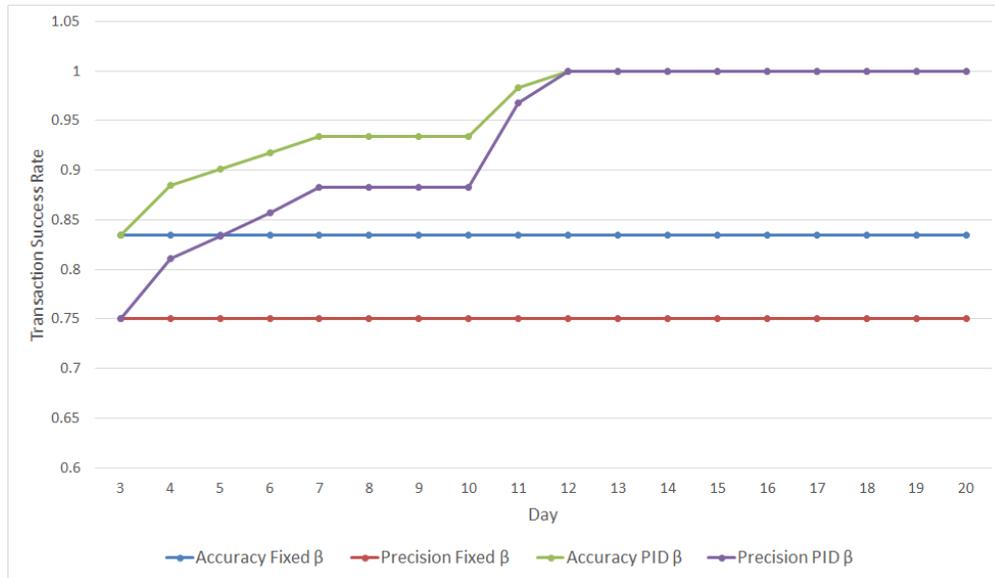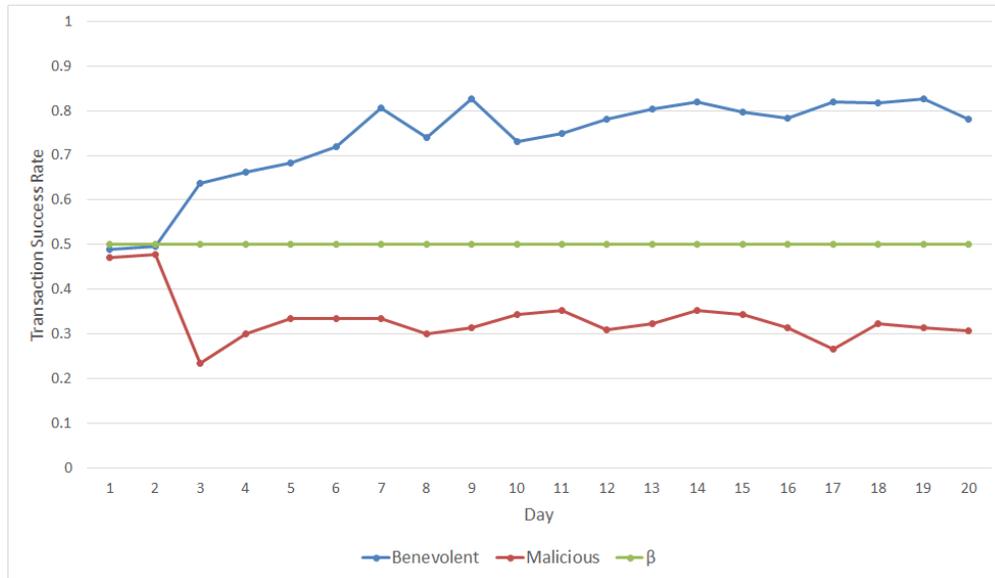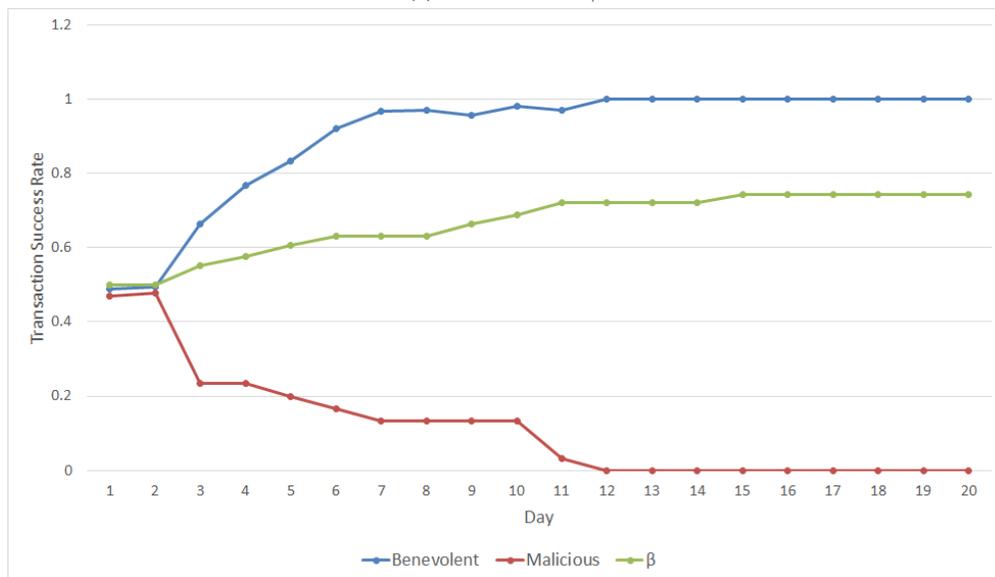
(a) Balanced e-marketplace
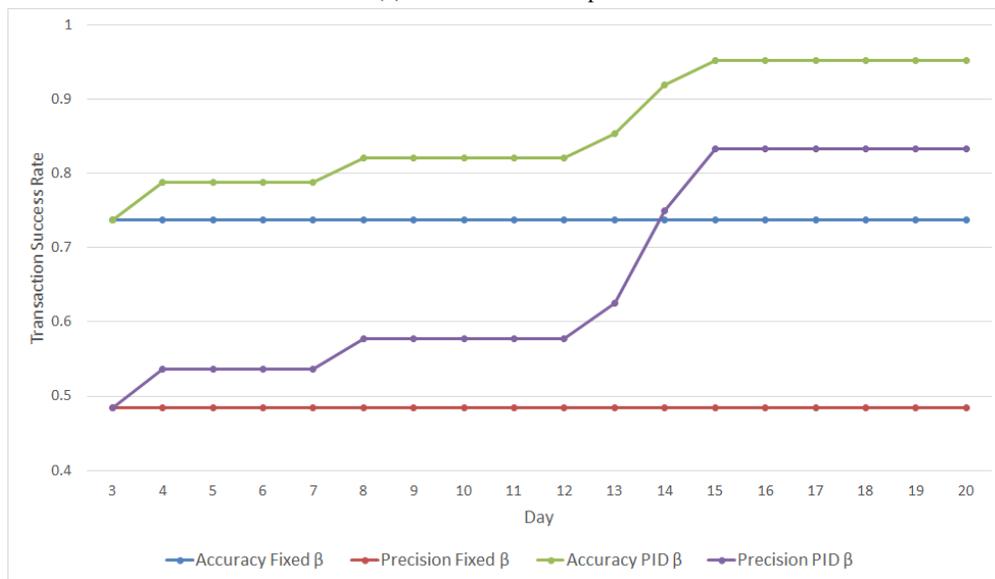
(b) With majority dishonest buyers

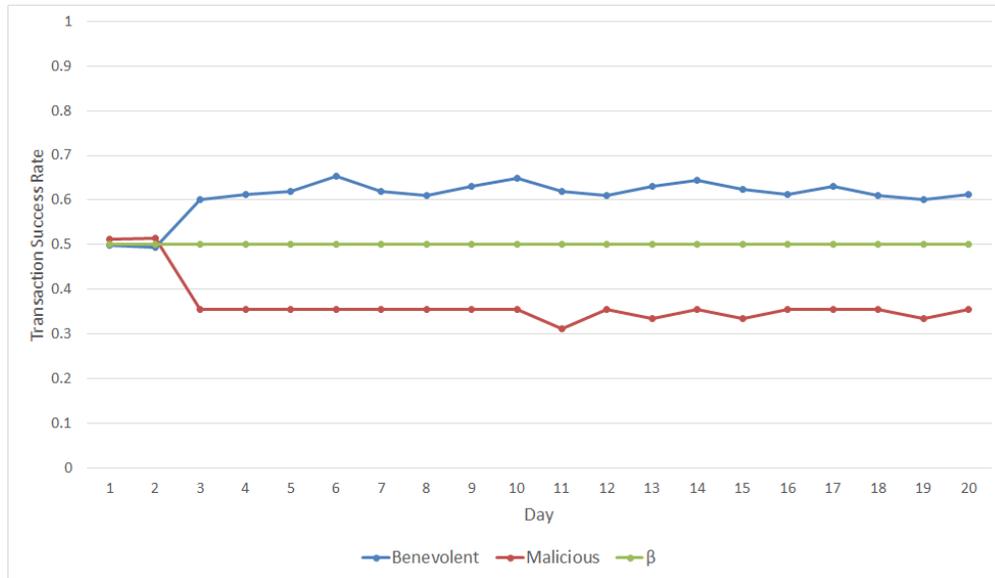**Figure 4.3:** The classification performance of different group of buyers
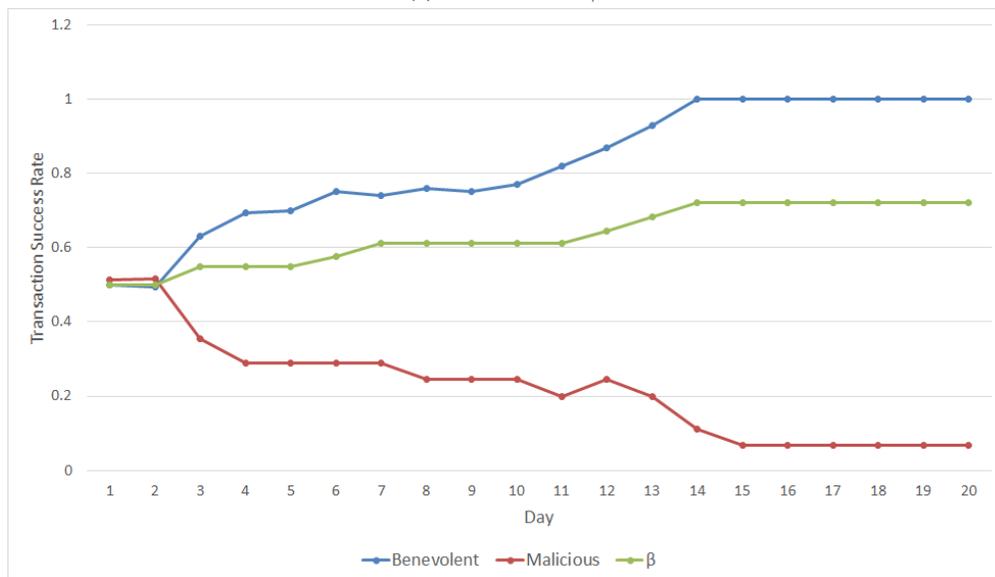
(a) Users use fixed β

(b) Users use PID-baded β

Figure 4.4: The transaction success rate of buyers in a balanced e-marketplace

(a) Users use fixed β

(b) Users use PID-baded β

Figure 4.5: The transaction success rate of buyers in an e-marketplace when majority of buyers are dishonest

Different population tendency in e-marketplace have negative impact on buyers transaction outcomes. Results obtained from the fixed β approach in different environmental settings indicate the inefficiency of static adaptation of threshold β in classifying different group of advisors. Specifically in dishonest majority e-marketplace, as a larger portion of dishonest buyers remains undetected, they have more influence in honest buyers' transaction decisions compared with the case in the balance environment, resulting in a fewer number of successful transaction outcomes (Figures 4.4a vs. 4.5a).

4.2.2 Summary

We have seen that using a principled method to evaluate the value of threshold provides improvement to the accuracy of credibility mechanisms in classifying different group of participants. We have also seen that the threshold parameter should be adjusted depending on the quality of marketplaces in terms of market liquidity and buyers satisfactions, allowing dishonest advisors to be filtering away.

With the presentation of this array of comparative results, the value of employing our PID-based mechanism for setting threshold becomes apparent. Our experiments demonstrate that employing a principled approach results in better overall trust modeling accuracy and market performance than the static approach in which the threshold value is selected arbitrarily by system designers.

The experimental results show how the use of the proposed self-adaptive threshold scheme can promote the honest buyers and deter the dishonest ones, more effectively than assuming the traditional threshold value of 0.5.

4.3 Real-world Scenario

To show that the proposed approach can function well in the real world the proposed approach is tested on a real-world dataset. The dataset gathered from the website Yelp.com has attributes that make it suitable to test the proposed approach.

The movement of crowdsourcing, where a large group of people produce reviews and convey their opinions about different products, has developed in recent years, especially since the advent of Web 2.0 around 2000. Thanks to the existence of websites such as Amazon and Trip Advisor, where users leave reviews about products and hotels, people no longer rely solely on businesses' advertisements anymore, which portray only the positive side of products and businesses. These crowdsourcing websites help people see a product or business from different angles. These reviews and ratings show both positive and negative aspects of the products or businesses, evaluated from different viewpoints.

Yelp is a consumer review site established in 2004 in San Francisco. Consumers use this website to provide their opinions about local businesses. These opinions are gathered in the forms of 5-star Likert scale ratings and raw text. Yelp call its members (users) Yelpers, and contains over 61 million local reviews [47]. This website has been successful

in attracting users and obtaining their reviews and in 2011 it had more than 10 million reviews [21]. Each review written by every user in this website is shown to the public. The rating given to a business affects its overall reputation in the system since the reputation of a business is simply the average of all the ratings that business has received.

Luca in [21] shows how the website Yelp.com has been a turning point for the independent restaurant industry. This study focuses on the business of restaurants in the city of Seattle. He leverages data from Yelp's website and the Washington State Department of Revenue to project the importance of this website's function. Using these data he found that from October 2009 until 2011 69% of restaurants in Seattle had been registered on the website. This article has three main findings that demonstrate the impact of reputation management systems. First, he comes up with an equation to calculate the effect a restaurant's reputation has on its revenue. If the reputation of a business in Yelp increases by just one star, its revenue will increase five to nine percent. Second, this study demonstrates that individual restaurants benefit more from this website compared to chain restaurants. Chain restaurants such as McDonald's usually have the same items on their menus and common food sources, so users' reviews do not change dramatically. Last but not least, the author shows that the market shares have decreased for chain restaurants since the time Yelp became popular. All of these facts show how a reputation on a website can have an impact on actual real-world business revenues.

Each year in Yelp a fraction of members are selected by the system as "elite". These users are special users in the system. When users want to find out more about a restaurant they can go to the web page that contains detailed information on that restaurant to have a closer look at the reviews. The reviews can be sorted based on several criteria. One is whether the reviewer is elite or not. As Luca shows in [21], restaurants with many reviews witness a larger response from the market, and reviews left by elite users have twice the impact of the reviews left by other users.

Currently, there exists a social network on Yelp's website. Users can follow and be friends with one another. However, there is no special ordering or mechanism that differentiates friends' reviews from others'. Moreover, consumers do not pay attention to the size of reviewers' friendship network [21].

To further evaluate the approach proposed in this thesis, it is tested against a real-world sample using the dataset of the Yelp.com website obtained on October 11, 2014 and containing data about five cities: Phoenix, Las Vegas, Madison, Waterloo and Edinburgh. Generally, Yelp has 22 categories for businesses, and 42,153 businesses have registered on the website. The number of businesses in each category can be seen in Figure 4.6. Almost 18,000 of the businesses registered on the website are under the Food and Restaurants categories, so these two categories include the majority of businesses on the website followed by shopping with more than 6,000 businesses. Note that each business can be categorized under more than one category. Almost 253,000 users exist in the system, providing it with approximately 1,125,500 reviews. Roughly 70% of these reviews have been given to businesses under the Restaurants and Foods categories, as can be seen in Figure 4.7. More than 75% of users have given at least one review in these two categories. In order to evaluate our proposed approach, we consider businesses under the Restaurants and Food categories that
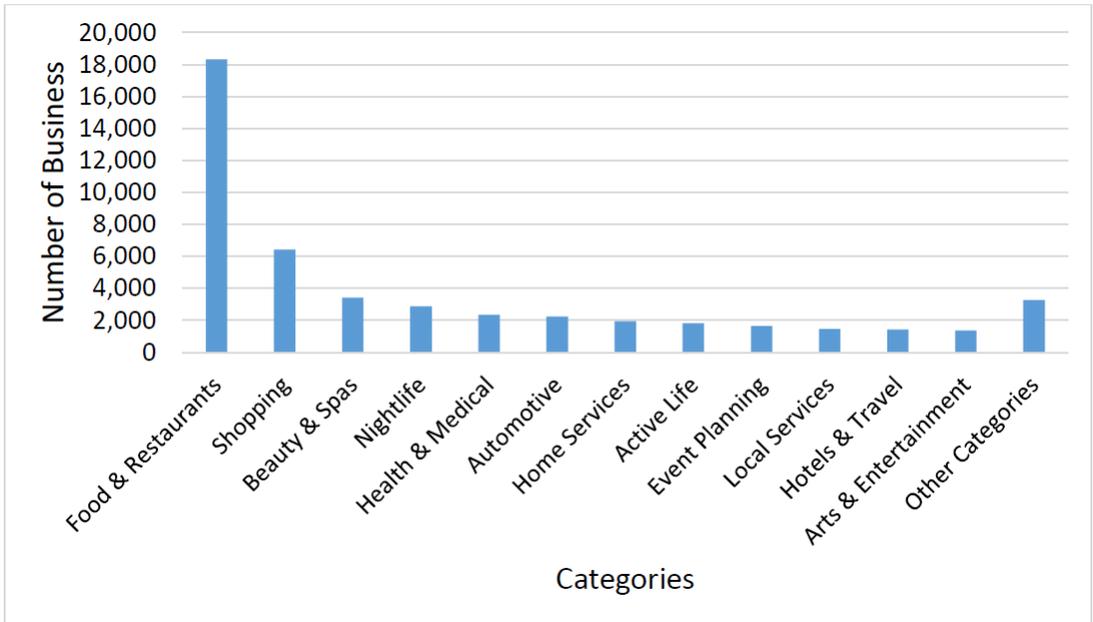
Figure 4.6: Number of businesses in each category

Figure 4.7: Number of reviews for categories

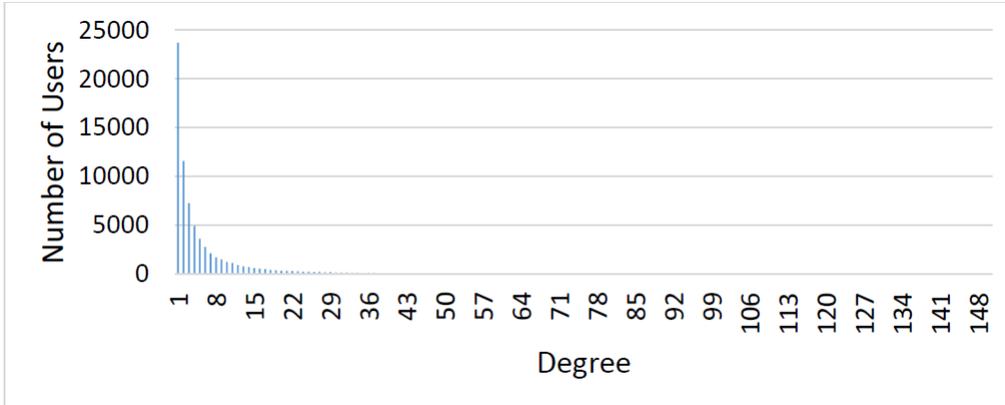**Figure 4.8:** Number of reviewers for categories

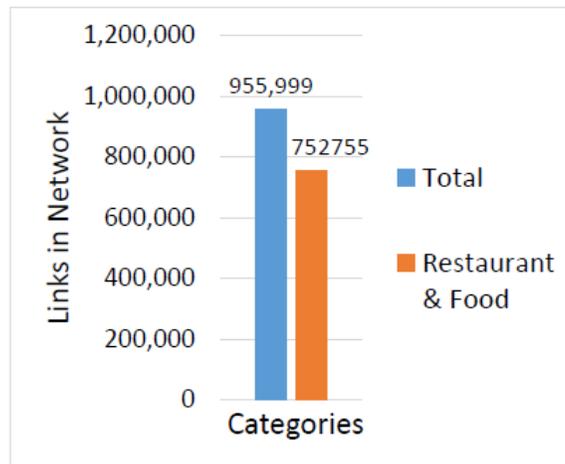**Figure 4.9:** Degree distribution of the friendship network

Figure 4.10: Number of links between users

include the majority of businesses.

In Yelp, users can become friends with one another. The network resulting from the friendships among users in the obtained dataset contains roughly 956,000 edges between users. If we consider users who have given at least one review to the Restaurant and Foods businesses (see Figure 4.10), the network will have 752,755 edges, which includes about 80% of the total edges in the system. The degrees distribution of this network can be seen in the Figure 4.9.

4.3.1 Experimental Settings

The Yelp is used to show how systematically determining the value of threshold can yield important gains in trust modeling in an e-marketplace. Each review written by every user in this website is shown to the public. The rating given to a business affects its overall reputation in the system since the reputation of a business in Yelp is simply the average of all the ratings that business has received. There are two main components that play a role on Yelp: users and businesses. In this experiment, users are considered as buyers, and in some cases they are referred as users of the system. On the other hand, we consider businesses as the sellers.

To adjust the dataset for our needs we consider each business as a seller and reviews as purchases that buyers(users) make.

In order to be able to measure the efficiency of the PID thresholding approach, the trust modeling has to be able to model the trust values in the system. As described in detail in previous chapters, the trust model used for this experiment works based on similar businesses that users have interacted with. Since in the dataset exist users who have given a few number of reviews for businesses and it affects the operation of the TRM system, we need to remove these users and keep users who have had interactions with similar businesses. Therefore the TRM system can model the trustworthiness values of users. As a result, in this experiment we only considered users in the system who have left at least 50 reviews. This limitation left us with acceptable numbers of users, businesses and reviews used for experiments in literature. This experiment is conducted with 1348 users, 15537 businesses and 128586 reviews. The date of reviews considered for the experiment is from 2011-01-01 to 2014-07-16.

The TRM system models the trust values that each user has in other users between the year 2011 and 2012. After modeling their trust values they are assigned as advisers for the user, if they are credible enough, i.e., having a trust value higher than the credibility threshold. The system run starts from year 2012. Subsequently, users seek to make transactions with businesses as exists in the original dataset which brings the distribution of transactions happening in the real word. The decision to initiate a transaction with a business is made by the user and it is entirely based on information about the business received from the user's advisers. The user calculates a reputation value for the business using the information he gathers from his advisers.

To evaluate the success of the transaction we use the following rule: the value calculated using advisers feedback

differs more than 0.02 from what the actual user experience with the business, we consider it as unsatisfactory transaction. The real transaction outcome (experience) exists in the dataset as a value between one and five (the users rating of the business), which we normalize between zero and one. At the end of each month, users report the number of their satisfactory and total number of interactions in the system to the central server. And TRM uses the Equations 3.2, 3.3, 3.4 to measure the performance of the system. After doing so, it employs the Equations 3.5, 3.1, 3.7 to calculate the new value of β for the next time slice.

We compare the outcome of the PID approach with the system which uses fixed credibility threshold of 0.55, the best credibility threshold introduced in the literature [7].

4.3.2 Experimental Results

The following graphs show the performance of the marketplace during the 30 month period, from 2012-01-01 until 2014-07-16.

Figure 4.11 illustrates the changes of β throughout the whole period of the experiment when the system uses the PID-based approach compared to when it uses fixed β which is set to 0.55. The credibility threshold of the PID approach starts from the pre-defined β as 0.55 and it almost reaches to 0.7 by the end of the experiment. In the following graphs we will see the effect of this increase of the credibility threshold.

First we measure the transaction rate of the system in both settings, where the TRM uses fixed and PID-based credibility threshold. As we can see in Figure 4.12a, initially the transaction rate for both cases are the same and this value is 0.788. After the seventh month we can observe a decrease in transaction of the marketplace which uses PID-based credibility threshold. This decrease continues until the end of the experiment period. The transaction rate of the marketplace when a fixed β is used stays at 0.775 while this value for PID-based β becomes 0.745. However, this decline was anticipated. As mentioned in the background section, as we increase the credibility threshold we are limiting the number of advisors. We can see this increase in β in Figure 4.11. In other words, users filter out more advisers. As a result, the user may not have enough information about a seller to carry out an intended transaction with that seller, which in turn decreases the transaction rate in the marketplace.

From Figure 4.12b we notice that the throughput of the marketplace increases when the TRM uses a PID-based β compared to the time it uses fixed β. Both cases initially have the same value of 0.228 for throughput. Again the changes happen after the seventh month. But this time the throughput increases more for the PID-based approach as it rises to 0.263 while this value for the fixed β approach increases only to 0.243.

As mentioned before, the system tries to improve the harmonic mean of transaction rate and throughput. As we can see in the following graph, the PID approach is successful at increasing this variable in comparison with the fixed approach. Figure 4.13 shows that this value for both approaches are 0.354 at the beginning of the experiment and
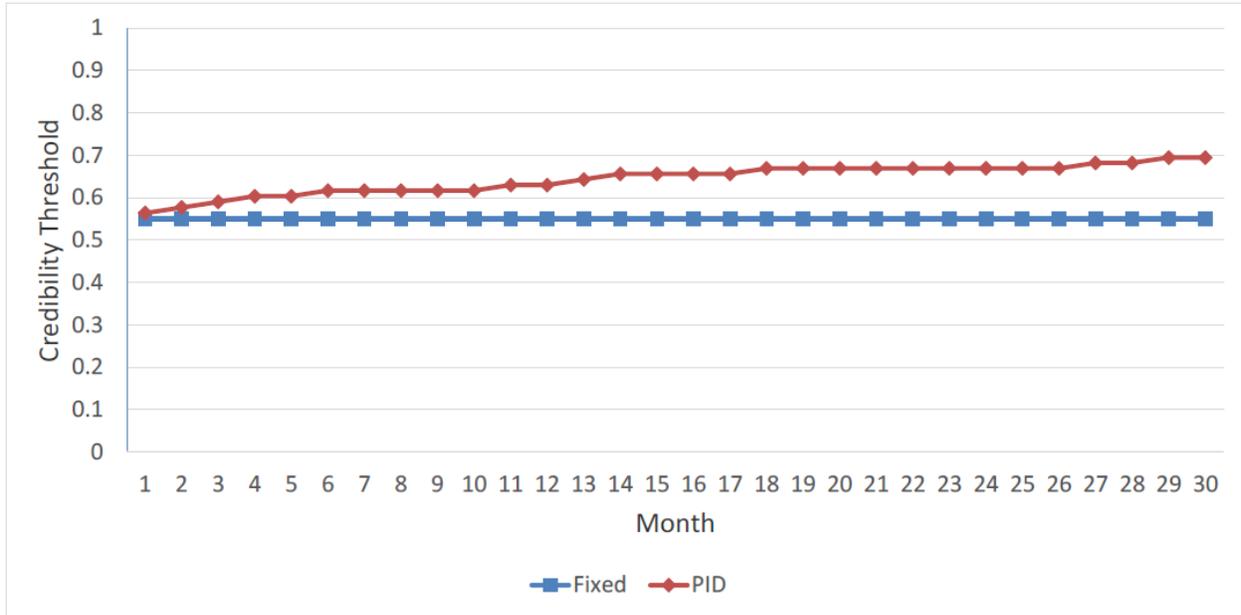
Figure 4.11: Credibility threshold value in two settings with fixed and PID-based β

then the PID approach starts to increase more that fixed one from the sixth month. By the end of the experiment, 30th month, the PID approach reaches the value of 0.389 while the fixed approach gains the harmonic mean of 0.370.

As can be seen, determining the credibility threshold using PID increases the accuracy of the model and the throughput of the system. On the other hand, it decreases the transaction rate. This was expected as each user filters out a fraction of its advisers by increasing the value of credibility threshold. And it results in less knowledge about a potential business. That is, in this case users do not know much about a business and they may not begin a transaction with that business which will result in lower value of coverage and transaction rate.

To measure the performance of the two approaches in addition to the metrics discussed above two other metrics are used.

1. Prediction Accuracy: Measured by the Mean Absolute Error (MAE) each approach generates.

2. Prediction Coverage: Number of reviews (items) predicted by the model out of all reviews in the dataset.

In order to measure the accuracy of both approaches we calculate the absolute error of the predictions for each approach. The lower the error, the higher the accuracy.

As we can notice from Table 4.1, PID approach outperforms fixed approach in terms of accuracy. However, it has a lower coverage. This was expected as each user filters out a fraction of its advisers by increasing the value of credibility threshold, which results in less knowledge available at the decision point about a potential business. That is, in this case users do not know much about a business and they may not begin a transaction with that business which will result in lower value of coverage and transaction rate.

As can be seen, having a dynamic thresholding management help users find more helpful advisers and therefore

40
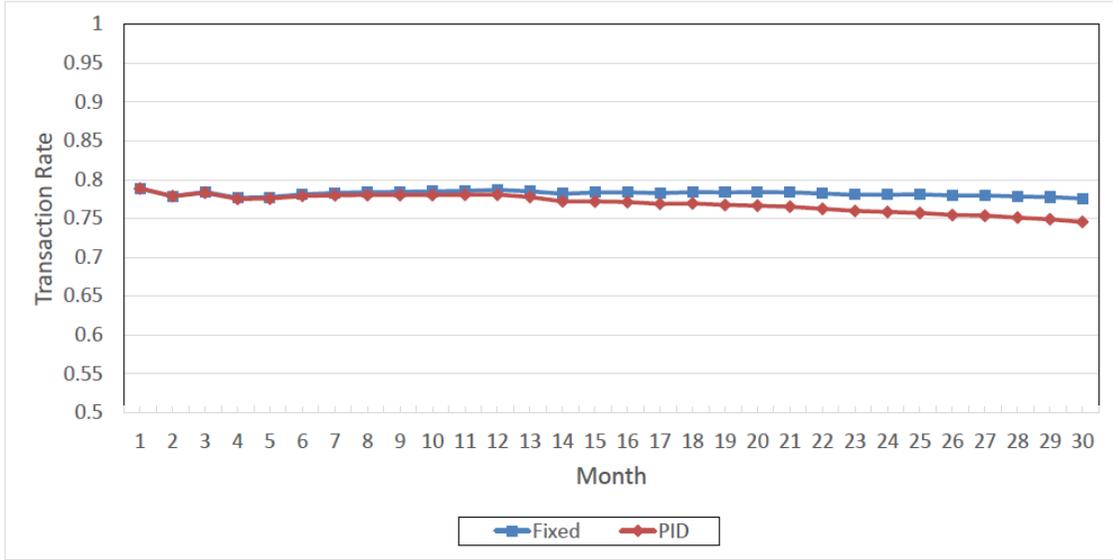
(a) Transaction rate of marketplaces when users adopt fixed or PID-based β
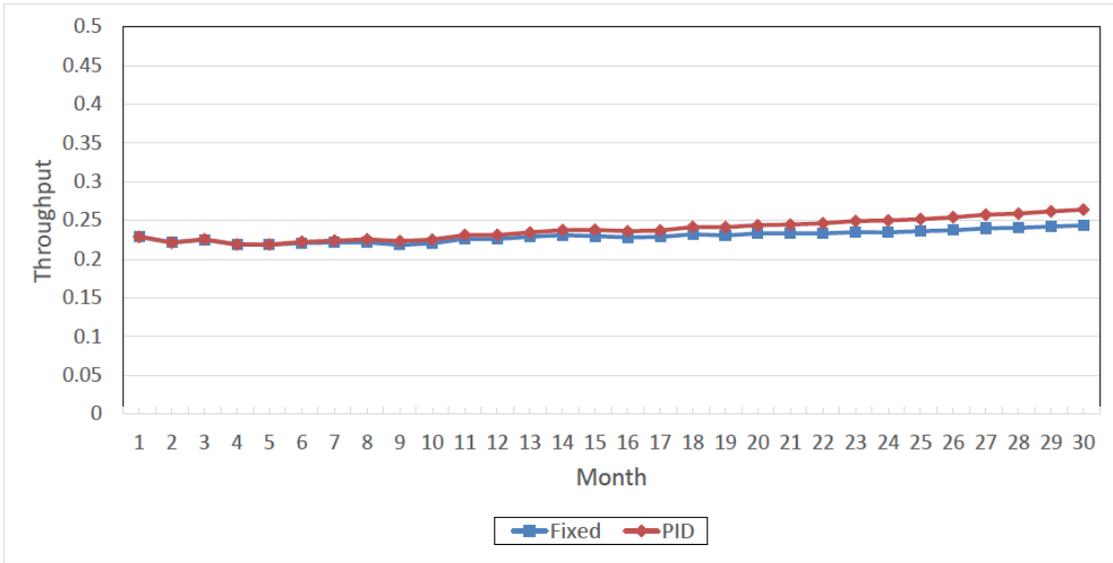
(b) Transaction success rate of marketplaces when users adopt fixed or PID-based β

Figure 4.12: Transaction rate and Throughput of marketplaces when users adopt fixed or PID-based β

	Fixed	PID
MAE (Mean Absolute Error)	0.1191	0.1150
Coverage	0.9986	0.9980

Table 4.1: Prediction accuracy of TRM systems when fixed or PID-based β is used
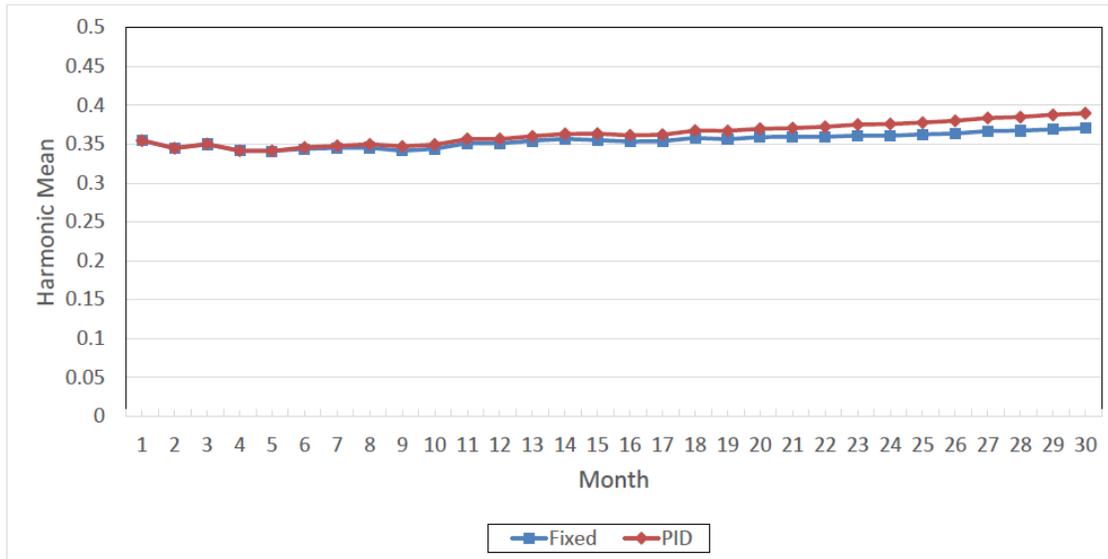
Figure 4.13: Harmonic mean of throughput and transaction rate in marketplaces when users adopt fixed or PID-based β

have more successful transactions. In addition, this approach results in more accurate predictions for the TRM system used in the marketplace.

4.4   Summary and Discussion

As shown in the previous sections, having a dynamic trustworthiness threshold has a positive effect on the e-marketplace performance. Different experiments with various settings show that having a feedback control system that uses PID controller result in more satisfactory transactions for users in the marketplace since a fixed trustworthiness threshold does not consider the dynamics within the environment.

Since the proposed approach is novel in the literature, first an experiment using simulation used to show the efficacy of the approach. In order to do so, an e-marketplace with buyers and sellers was simulated. This marketplace uses a TRM system to help buyers find sellers with high quality of service. The TRM system used in this marketplace, as introduced in the previous chapters, assigns advisors to buyers from other buyers in the marketplace. To show that having a threshold management system has a positive impact on the buyers' satisfaction, we designed the simulation to be run in two main settings. First the TRM system uses a threshold which is fixed throughout the simulation time. In the second setting the system is run as the TRM system uses a dynamic threshold that changes based on the marketplace's status to improve its performance in terms of its buyers' satisfaction. In the latter setting, the proposed PID-controller approach introduced in this thesis is used. The TRM system utilizes a feedback control system that uses PID controller to set an appropriate trustworthiness threshold to properly isolate malicious sellers and buyers. The results of this

experiment are positive since having a dynamic threshold help the system to aid benevolent users have more successful transactions while isolating malicious ones from providing bad advice to others. Comparing the results of the two settings, it can be seen that in the beginning both of them function the same as the trustworthiness thresholds used in them are close to each other. The results changes as time passes by. The TRM that uses a dynamic trustworthiness threshold it outperforms the one which uses a fixed threshold. This is due to the fact that the feedback control system senses the marketplace's situation, and if there is a need to modify the threshold, the controller advises TRM system to change it so that there would be improvements in the performance of the marketplace.

The experiment conducted on a real-world dataset from the crowd-sourcing website Yelp.com. Since the purpose of this experiment was to find the efficiency of utilizing the PID controller in a TRM system, the data in dataset was filtered in a way to remove the effect of other variables in the experiment. Therefore, the data about only two categories in the dataset was retained: Restaurants and Food. To counter the sparsity effect of the dataset that affects trust modeling by TRM system, the data was further filtered so that the functionality of the TRM could be compared in the two settings: using fixed and dynamic trustworthiness threshold. Although this filtering may help the TRM system to model the trustworthiness values of users in the e-marketplace, it may affect the results of experiment as we only consider those users who have left at least a certain number or reviews. According to the topics discussed in the Background section, however, we cannot conclude that filtering the dataset results in including only trustworthy users. This is due to the fact that current TRM system do not solely rely on global trust (i.e., reputation), so each user may not be trustworthy in the eye of other users only because he has reviewed many businesses. In addition to the performance of the marketplace in terms of users' transaction satisfaction, the prediction of the TRM system in two settings was compared using two measurements: Mean Absolute Error and Coverage.

As can be seen in the previous section, the value of the trustworthiness threshold is increased throughout the period of the experiment. This increase results in the increase of the rate of successful transactions in the system. This improvement can have a big impact on the communities where the number of transactions is large. Therefore, it will result in more satisfying transactions in the marketplace which in turn will result in attracting and retaining more users. Moreover, the error of the prediction decreases when we use dynamic trustworthiness threshold. However, this change limits the coverage of the prediction of the TRM system. As discussed earlier, since a higher trustworthiness threshold results in a small number of advisers for users the number of transaction decisions for which the system has enough information to predict a positive outcome decreases which means decrease in coverage.

CHAPTER 5

CONCLUSION

To conclude the work of this thesis, this chapter first summarizes the work and then lays out the directions for future work.

5.1 Summary

This paper pinpoints a common problem of existing trust and reputation systems in electronic commerce systems. Existing TRM systems have the assumption that a default honesty threshold exists, and most of the times the choice of this value for this "magic" threshold is left to the designers implementing a particular system. Moreover, the advances of the field of trust and reputation management systems have mostly been in the ways to determine the trust values of users in an environment. However, research about how to make a decision about trustworthiness has not been sufficient

Deciding the value of the credibility threshold used in social networks of buyers should consider the tradeoff between precision and coverage and should be determined in a dynamic manner and relative to the status of the environment. A low value of credibility threshold results in large number of advisers which might have little or even negative contribution to buyers decision making. On the other hand, having a high level of honesty threshold eliminates more advisers from the social network of buyers, which restricts buyer to acquire sufficient information about sellers.

In this study, an adaptive credibility threshold is introduced which uses the PID-controller that monitors the specific quality parameters of e-marketplace to determine new values for the honesty threshold to adapt to the changing marketplace. The proposed threshold adjustment works independently of the trust model used in the system, and this makes it well-suited to be incorporated with different credibility evaluation mechanisms and filtering models for electronic marketplaces.

Experimental results show the advantages of adaptive evaluation on the honesty threshold. In particular, we demonstrate that credibility evaluation mechanism guided by PID-based threshold management techniques can increase the performance of the marketplace in terms of satisfactory interactions. It also improves the TRM accuracy in terms of predicting the trust values in the system.

5.2   Future Work

Although the novelty of the proposed approach is to use the notion of feedback control systems to define a dynamic threshold in the field of trust and reputation management systems, it is not limited to this field and can be applied to other areas as well.

As mentioned in the Experiments and Evaluation section, based on the attributes of dataset it was filtered in order to remove the random variables that may affect experiments which may be resulted from the sparsity of dataset. With the advances in the field of TRM systems, it will be interesting to experiment the approach with the dataset of an e-marketplace which is available and the TRM system can produce the trustworthiness values regardless of sparsity of the dataset.

Another interesting direction for future work would be to improve the feedback controller method by adopting different dynamic performance metrics supported in the market microstructure literature, in addition to those considered here. Furthermore, since the buyers' contribution in providing feedback is an essential element in the performance monitoring of the marketplace, a useful direction for future work would be the incorporation of an incentive mechanism to promote more participation (in terms of providing honest feedback) from the buyers.

BIBLIOGRAPHY

[1] Karl Aberer and Zoran Despotovic. Managing trust in a peer-2-peer information system. In *Proceedings of the tenth international conference on Information and knowledge management*, pages 310–317. ACM, 2001.

[2] Rajiv D Banker and Iny Hwang. Importance of measures of past performance: Empirical evidence on quality of e-service providers*. *Contemporary Accounting Research*, 25(2):307–337, 2008.

[3] Stuart J Barnes and Richard Vidgen. Measuring web site quality improvements: a case study of the forum on strategic management knowledge exchange. *Industrial Management & Data Systems*, 103(5):297–309, 2003.

[4] Chrysanthos 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, 2000.

[5] William H Delone. The delone and mclean model of information systems success: a ten-year update. *Journal of management information systems*, 19(4):9–30, 2003.

[6] Ian Foster, Carl Kesselman, and Steven Tuecke. The anatomy of the grid: Enabling scalable virtual organizations. *International journal of high performance computing applications*, 15(3):200–222, 2001.

[7] Joshua Gorner, Jie Zhang, and Robin Cohen. Improving trust modeling through the limit of advisor network size and use of referrals. *Electronic Commerce Research and Applications*, 12(2):112–123, 2013.

[8] Joshua Mark Gorner. Advisor networks and referrals for improved trust modelling in multi-agent systems. Master's thesis, University of Waterloo, 2011.

[9] William Grey, Thomas Olavson, and Dailun Shi. The role of e-marketplaces in relationship-based supply chains: A survey. *IBM systems Journal*, 44(1):109–123, 2005.

[10] Yasaman Haghpanah and Marie desJardins. Prep: a probabilistic reputation model for biased societies. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, AAMAS '12, pages 315–322, 2012.

[11] Mark A Haidekker. *Linear Feedback Controls: The Essentials*. Newnes, 2013.

[12] Jon Herlocker, Joseph A Konstan, and John Riedl. An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms. *Information retrieval*, 5(4):287–310, 2002.

[13] Trung Dong Huynh, Nicholas R Jennings, and Nigel R Shadbolt. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 13(2):119–154, 2006.

[14] Athirai Aravazhi Irissappane, Siwei Jiang, and Jie Zhang. A framework to choose trust models for different e-marketplace environments. In *IJCAI*, 2013.

[15] Audun Jøsang and Jennifer Golbeck. Challenges for robust trust and reputation systems. In *Proceedings of the 5th International Workshop on Security and Trust Management (SMT 2009), Saint Malo, France*, 2009.

[16] Audun Josang and Roslan Ismail. The beta reputation system. In *In Proceedings of the 15th Bled Electronic Commerce Conference*, 2002.

[17] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision support systems*, 43(2):618–644, 2007.

[18] Radu Jurca and Boi Faltings. An incentive compatible reputation mechanism. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*, AAMAS '03, pages 1026–1027, New York, NY, USA, 2003. ACM.

[19] Xin Liu. Towards context-aware social recommendation via trust networks. In *Web Information Systems Engineering–WISE 2013*, pages 121–134. Springer, 2013.

[20] Xin Liu and Karl Aberer. Soco: a social network aided context-aware recommender system. In *Proceedings of the 22nd international conference on World Wide Web*, pages 781–802. International World Wide Web Conferences Steering Committee, 2013.

[21] Michael Luca. Reviews, reputation, and revenue: The case of yelp. com. Technical report, Harvard Business School, 2011.

[22] Wendy Maiden, Ioanna Dionysiou, Deborah Frincke, Glenn Fink, and David E Bakken. Dualtrust: A distributed trust model for swarm-based autonomic computing systems. In *Data Privacy Management and Autonomous Spontaneous Security*, pages 188–202. Springer, 2011.

[23] Stephen Marsh. *Formalising trust as a computational concept*. PhD thesis, University of Stirling, scotland, 1994.

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

[25] Mohsen Mohkami and Hassan Shakeri. Alleviating controversiality problem in trust estimation. In *Telecommunications (IST), 2012 Sixth International Symposium on*, pages 1103–1108. IEEE, 2012.

[26] Lik Mui, Mojdeh Mohtashemi, and Ari Halberstadt. Notions of reputation in multi-agents systems: a review. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*, AAMAS '02, pages 280–287, New York, NY, USA, 2002. ACM.

[27] Mehrdad Nojoumian and Timothy C Lethbridge. A new approach for the trust calculation in social networks. In *E-Business and Telecommunication Networks*, pages 64–77. Springer, 2008.

[28] Zeinab Noorian, Stephen Marsh, and Michael Fleming. Multi-layer cognitive filtering by behavioral modeling. In *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11)*. ACM, 2011.

[29] Zeinab Noorian, Mohsen Mohkami, and Julita Vassileva. Self-adaptive filtering using pid feedback controller in electronic commerce. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 267–272. ACM, 2014.

[30] Zeinab Noorian and Mihaela Ulieru. The state of the art in trust and reputation systems: a framework for comparison. *J. Theor. Appl. Electron. Commer. Res.*, 5:97–117, August 2010.

[31] Hitay Ozbay. *Introduction to feedback control theory*. CRC Press, 1999.

[32] Hitay Ozbay. *Introduction to feedback control theory*. CRC Press, 2000.

[33] Ananthanarayanan Parasuraman, Valarie A Zeithaml, and Arvind Malhotra. Es-qual a multiple-item scale for assessing electronic service quality. *Journal of service research*, 7(3):213–233, 2005.

[34] G Prem Premkumar. Perspectives of the e-marketplace by multiple stakeholders. *Communications of the ACM*, 46(12):279–288, 2003.

[35] Xueming Qian, He Feng, Guoshuai Zhao, and Tao Mei. Personalized recommendation combining user interest and social circle. *Knowledge and Data Engineering, IEEE Transactions on*, 26:1763–1777, 2014.

[36] Kevin Regan, Pascal Poupart, and Robin Cohen. Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 1206. AAAI Press, 2006.

[37] Cristiano Castelfranchi Rino Falcone. Socio-cognitive model of trust. In *Encyclopedia of Information Science and Technology (V)*. 2005.

[38] Jordi Sabater and Carles Sierra. Regret: A reputation model for gregarious societies. In *Fourth workshop on deception fraud and trust in agent societies*, volume 70, 2001.

[39] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.

[40] Michael J Smith et al. Learning to trust in the competence and commitment of agents. *Autonomous Agents and Multi-Agent Systems*, 18(1):36–82, 2009.

[41] WT Teacy, Jigar Patel, Nicholas R Jennings, and Michael Luck. Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 997–1004. ACM, 2005.

[42] WT Luke Teacy, Jigar Patel, Nicholas R Jennings, and Michael Luck. Travos: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

[43] Joana Urbano, Ana Paula Rocha, and Eugénio Oliveira. Computing confidence values: Does trust dynamics matter? In *Progress in Artificial Intelligence*, pages 520–531. Springer, 2009.

[44] Yao Wang, Jie Zhang, and Julita Vassileva. Super-agent based reputation management with a practical reward mechanism in decentralized systems. In *4th IFIP WG*, volume 11, pages 33–40, 2010.

[45] Yonghong Wang and Munindar P. Singh. Formal trust model for multiagent systems. In *IJCAI*, 2007.

[46] Abdulsalam Yassine, Ali Asghar Nazari Shirehjini, Shervin Shirmohammadi, and Thomas T Tran. Knowledge-empowered agent information system for privacy payoff in ecommerce. *Knowledge and information systems*, 32(2):445–473, 2012.

[47] Yelp. About yelp. `http://www.yelp.ca/about`. last checked: 12.12.2014.

[48] Han Yu, Zhiqi Shen, Chunyan Miao, Cyril Leung, and Dusit Niyato. A survey of trust and reputation management systems in wireless communications. *Proceedings of the IEEE*, 98(10):1755–1772, 2010.

[49] Jie Zhang and Robin Cohen. Evaluating the trustworthiness of advice about seller agents in e-marketplaces: A personalized approach. *Electronic Commerce Research and Applications*, 7(3):330–340, 2008.