USE MULTIPLE MODELING APPROACHES TO STUDY SUSTAINED ONLINE COMMUNITIES

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, Saskatchewan

By
Yan Mao

PERMISSION TO USE

In presenting this thesis in partial fulfillment 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

University of Saskatchewan

Saskatoon, Saskatchewan, Canada

S7N5A9

ABSTRACT

In recent years, extensive studies of many interesting aspects of online community dynamics promoted a better understanding of this area. One of the most challenging problems facing builders of online communities is the design of incentive mechanisms that can ensure user participation. However, running online community experiments in the real world is expensive, and requires a great deal of motivation from users.

In this thesis two major approaches are explored: system dynamics modeling and agent-based modeling, to simulate the overall behaviours of participants in online communities. Although these models are developed by using two different methodologies, both of them can provide insights into the user motivation process, incentive mechanism evaluation and community development. The target online community for my study is called Comtella, which is used in several senior Computer Science classes in the Department of Computer Science, University of Saskatchewan. Simulation models for the Comtella online community have been developed and the simulation results are useful to provide future directions for incentive mechanism improvement.

ACKNOWLEDGEMENT

I would like to thank my supervisor Dr. Julita Vassileva and Dr. Winfried Grassmann for their guidance, advice and constant belief in my ability to finish the simulation models and successfully investigate the problem throughout my research. Their outstanding assistance in achieving the publications, sponsorship in pursuing this graduate program and support to attend the conferences are so much appreciated.

With great pleasure, I would also like to respectfully acknowledge Dr. Nathaniel Osgood for intelligently guiding and encouraging my research work and interests, as well as Dr. Ralph Deters for supporting my research.

The comments of the advisory committee have helped make this thesis better and I would like to thank all members: Dr. Simone Ludwig, and Dr. Chris Soteros.

Special thanks to Ms. Jan Thompson for her help all the time, MADMUC folks for supporting, and my friends who have been encouraging always. I would also express my loving appreciation to my parents for giving me life as well as their constantly understanding and support.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

cont.	continue
id	identifier
ABM	Agent-based Modeling
ABS	Agent-based Simulation
BBS	Bulletin Board System
CDF	Cumulative Distribution Function
CMC	Computer-Mediated Communication
DAI	Distributed Artificial Intelligence
IS	Information System
JADE	Java Agent DEvelopment Framework
OOCC	Open Online Collaboration Community
P2P	Peer-to-Peer
QPID	Qualitative Politicized Influence Diagrams
SD	System Dynamics
URL	Uniform Resource Locator

CHAPTER 1

INTRODUCTION

In Wikipedia (2007), an online community is defined as a group whose members are connected by means of information technologies, typically the Internet, rather than in person. People use the term "online community" (or "virtual community") frequently to refer to computer-mediated communication (CMC) groups. There is no accepted definition of online community, and in 1996 a group of academics held a workshop and identified the following key characteristics of an online community (Whittaker et al. 1997):

• People have a shared goal, interest, need, or activity which is the primary reason for belonging to the community;

• People engage in repeated, active participation, intense interactions, shared activities, and often feel in strong emotional ties;

• People have access to shared resources and the policies that determine the access to those resources;

• There is reciprocity of information exchange, support, and services among members;

• There is a shared context of social conventions, language, and protocols.

Online communities are important for many reasons. First, they help to replace the relationships lost as more and more informal public spaces disappear from our real lives (Rheingold 1993 and Schuler 1996). They allow people with similar interests to connect with each other and to gain benefits from the presence and activities of other people in online communities. Online communities provide not only information, resources, and conversations which people can use and participate in, but they also provide a way to form social relationships that allow people to do things together with others in a new way. This may, to some extent, help “increase involvement within people's face-to-face communities by increasing democratic participation and other community activism” (Blanchard 2004). Second, when people experience the feeling of belonging to an online community, the positive emotion becomes an intrinsic incentive for further participation in the community, which makes online communities self-sustained.

Since the early nineties the popularity of online communities has increased dramatically. A huge number of people join into the virtual environment (such as BBS, discussion groups, chat rooms) day and night to not only share papers, music/video files and other kinds of web-resources online, but also to interact with others, exchange opinions, publish news, debate issues, etc. This provides a great opportunity for knowledge exchange and helps people to connect across boundaries. At the same time, spurred by the rapid emergence of online communities, studying the complex dynamics involved in communities becomes an exciting new research area (Krichmar et al. 2005). One of the most challenging activities within this vision is to explore the factors that contribute to specific online behaviour (Vassileva et al. 2005), such as contributing new resources which can help to encourage and sustain the social engagement among members in the community (Kelly et al. 2002; Ludford et al. 2004). Under-contribution and lurking are phenomena that cause problems in online communities (Nonnecke and Preece 2000). For example, Adar and Huberman (2000) found that in Gnutella, two-thirds of users share no files and 20 percent provide 98 percent of all the music files available on Gnutella. In some open source development communities the situation may be even worse (Mockus et al. 2002), with an estimated 4 percent of developers contributing nearly 88 percent of new code and 66 percent of code fixes. In some particular instances these low levels of participation are not detrimental, e.g. in file-sharing communities, because of the nature of shared materials (shared music files do not expire, can only be multiplied). However, even such communities can only become sustainable after reaching a “critical mass” of contributions. Therefore, user motivation processes and incentives mechanisms are quite important to online communities in certain phases of their lifetime and are worth further study.

Previous work has dealt with factors that attract people to participate in online communities (Thompson et al. 2002; Leimeister et al. 2004). Many researchers tried to investigate the motivation of users by applying social psychology theories (Beenen 2004) such as building a social reputation system (Kollock et al. 1996; Jensen et al. 2002; Bretzke and Vassileva 2003) and introducing reward mechanisms (Cheng and Vassileva 2005). Other researchers also use the methods of improving the framework (Brook et al. 2003) and user interface (Webster and Vassileva 2006) to stimulate user participation. However, the knowledge of dynamic online behaviours and user motivation in the communities is still deficient. Running online community experiments in the real world is expensive, and requires a great deal of motivation from users. Besides, complex dynamics involved in this problem and bounded human judgment (Simon 2000) prevent us from fully understanding the problem.

Faced with the overwhelming complexity of the real world, time pressure, and limitations in information availability and processing capabilities, computer simulation modeling offers attractive and inexpensive means of investigating such phenomena without risk (Pidd 1993). For these reasons, simulation becomes the most promising tool to assist researchers in studying complex phenomena such as user participation in online communities and evaluating the effects of incentive mechanisms.

Several simulation methodologies exist that can be used to study online communities in different ways: system dynamics modeling, agent-based modeling and hybrid modeling. System dynamics (SD) modeling and agent-based simulation (ABS) are two major widely acknowledged modeling methodologies in the computational area. System dynamics is a quantitative simulation methodology that emphasizes the use of feedback loops to understand the basic structure of a system, and thus to understand the behaviour it can produce. On the other hand, agent-based simulation focuses on the behaviour of the individuals and studies the dynamics of a system resulting from the interactions of individual agents. For my study, both system dynamics modeling and agent-based modeling approaches are applied to study the target system -- Comtella system, which is an online community developed in the MADMUC lab (stands for Multi-Agent

Distributed Mobile and Ubiquity Computing Lab) at the University of Saskatchewan.

Comtella is a system which enables users in the community to share web resources, typically web links to academic papers. Like other small-scale online communities, under-contribution is a big problem of Comtella, and simulation models can help us get an insight into this problem without investing much time or resources. The main objective of the thesis is to study through computer simulation the effects of incentives mechanisms for participation in Comtella. The real data collected from the Comtella community is used for calibration and validation of the models. Users have given consent for the use of the data for research (the consent form is presented in Appendix A).

The rest of the thesis is organized as follows. Chapter 2 provides background information about the Comtella online community, presents some related work on modeling systems and elaborates the objectives of the study. Chapter 3 presents the system dynamics simulation models as well as the simulation results. Chapter 4 describes the agent-based simulation model and experimental results. The discussion and the directions for future work are presented in the last chapter.

The thesis has five appendices: Appendix A presents the consent form for the use of the data for research, Appendix B presents the structure of the Comtella database, Appendix C and D present the calculation details and the parameter list for the system dynamics models, and Appendix E presents the details of the calculations on membership level in the agent-based model.

CHAPTER 2

RELATED WORK AND OBJECTIVES

Under-contribution is a big problem faced by nearly all online communities. To deal with this problem, many communities deploy incentive mechanisms to reward participation. Simulation models are very helpful for studying the dynamics of online communities and also useful for measurement and further improvement of incentive mechanisms in online communities. Previous work has studied factors that attract people to participate in online communities. In this chapter, I give a brief introduction to an online community for sharing URLs of online articles, called Comtella, which is the main target community to be modeled.

This chapter also reviews work related to system dynamics modeling and agent-based modeling in the context of online communities. System dynamics simulation is a quantitative approach using feedback loops among stocks and flows for continuous processes, while agent-based simulation is an approach focusing on the basic interactions of individuals. Agent-based simulation works for both continuous and discrete systems. The main concepts of these different modeling approaches are described first, and several existing models developed by researchers are presented in order to show how they map real world problems into simulation models.

The main research goal of the thesis is presented last, including the details of the incentive mechanism to be modeled, the aspects to be considered, and how those expected results will help in the design and tuning of the incentive mechanisms in the real Comtella community.

2.1 Comtella: an online community for sharing papers

An educational online learning community called Comtella has been developed at the MADMUC Lab of University of Saskatchewan (Vassileva, 2002). It has been used in several senior Computer Science courses where students can share class-related digital resources on the web, such as bookmarks to articles, papers, etc. Another version of Comtella is used by research groups where researchers can combine their literature research efforts and create a digital library with low maintenance costs.

Normally a particular Comtella community is used for one particular course and hence the number of participants is small. As in any online community, there are users who do not share anything in the community (free-riders). Typically, they enter the community, search and download what they need, then log out. Especially for small-scale online communities like Comtella, free-riders might have an even more harmful impact compared to large-scale open source online communities. According to Dunbar (1996) and Shirky (2002), for a smaller social group, the quality of the connections is higher, because increasing the number of people in the group weakens communal connection. As a result, in small-scale communities like Comtella, users are better connected and both over-contributors and free-riders can affect the overall participation levels much faster and stronger than in large-scale open source online communities. Therefore, an efficient incentive mechanism is needed to motivate users.

The main incentive mechanism of Comtella (Cheng 2005 and Vassileva 2005) rewards contributions using hierarchical memberships in the community (gold, silver, bronze and common member) based on the user participation level. A user membership is determined by the activity points that are rewarded for each dimension of participation (e.g. contributing many links or good quality links, or participating in discussion, providing comments and ratings, etc.). In this way the incentive mechanism in Comtella provides a combined measure of user participation, which is quite understandable to users and effective. The expectation is that users will be more willing to readily engage in competition to achieve a higher level of membership than when competing along multiple dimensions of participation.

As mentioned in Cheng's thesis (2005), the Comtella system uses fixed thresholds for each membership level in order to classify users in the community. Once the number of activity points reaches the threshold, the user reaches the corresponding membership level. In this way, the users are strongly motivated to participate before they reach the highest membership level.

Since 2002, several versions of Comtella have been developed and deployed, which are basically used as class-supporting tools in several computer science courses. Two versions of Comtella were deployed for undergraduate courses in 2003-2004 and 2004-2005 winter sessions respectively, which are called respectively, 'the early version of Comtella' and 'the latest version of Comtella' in my thesis.

In both versions of Comtella, users are visualized as stars in a night sky (as shown in Figure 2.1), and the information on a particular user will be displayed when the mouse rests on the star. The size and brightness of a star is determined by the contribution level of the user represented by the star.

Higher-level memberships result in larger stars in the visualization, better interfaces and services (such as personalized messages), and more privileges or special rights. For example, participants with higher membership levels might get personalized messages showing the desired number of contributions for the current week, or the quality of their contributions and ratings in previous weeks.

In my study I try to gain insights into this user motivation process and incentives mechanisms by simulation in both system dynamics and multi-agent frameworks, and identify the important factors in this process such as the reward factor, thresholds for membership upgrading, etc. The next two sections explain the incentive mechanisms used in these two versions of the Comtella community.

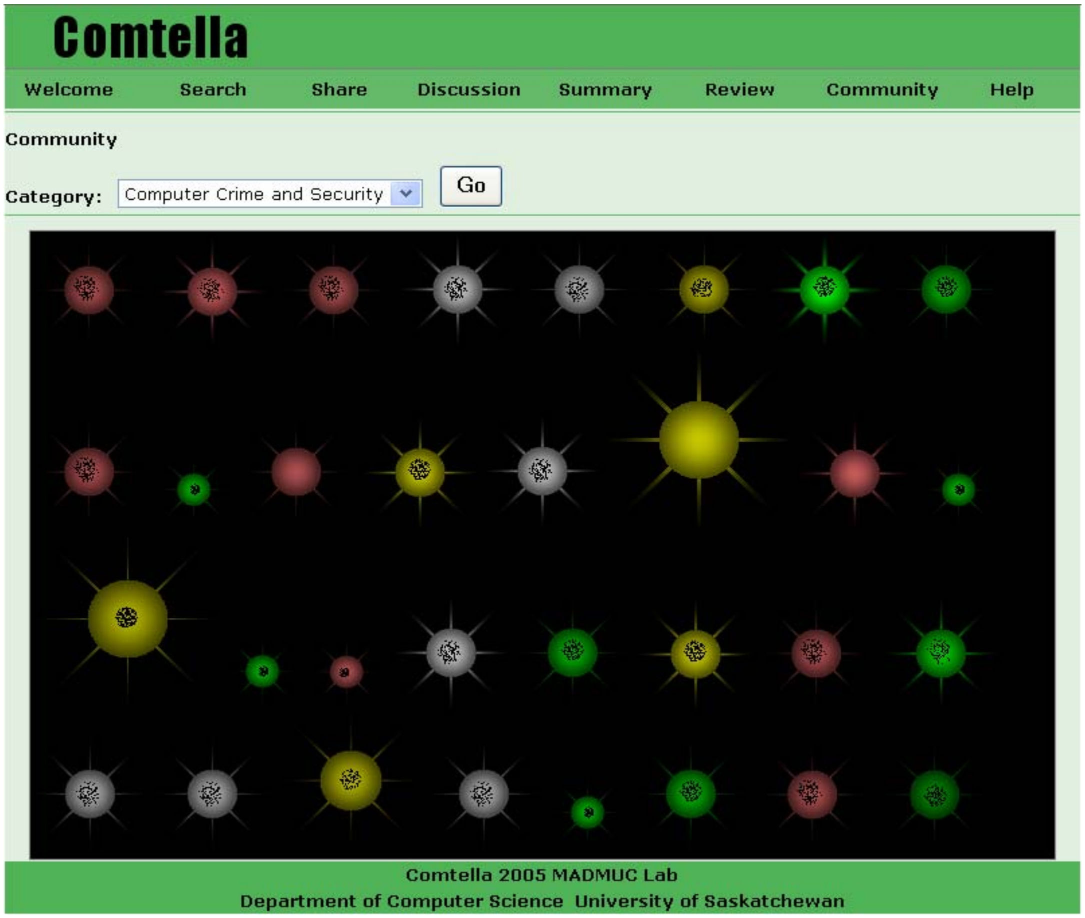


Figure 2.1. Visualization of the memberships of users in the Comtella community

2.1.1 Early version of Comtella incentive mechanism

For the early version of Comtella, the incentive mechanism in Comtella is presented in Figure 2.2, which shows the important relationships between different factors.

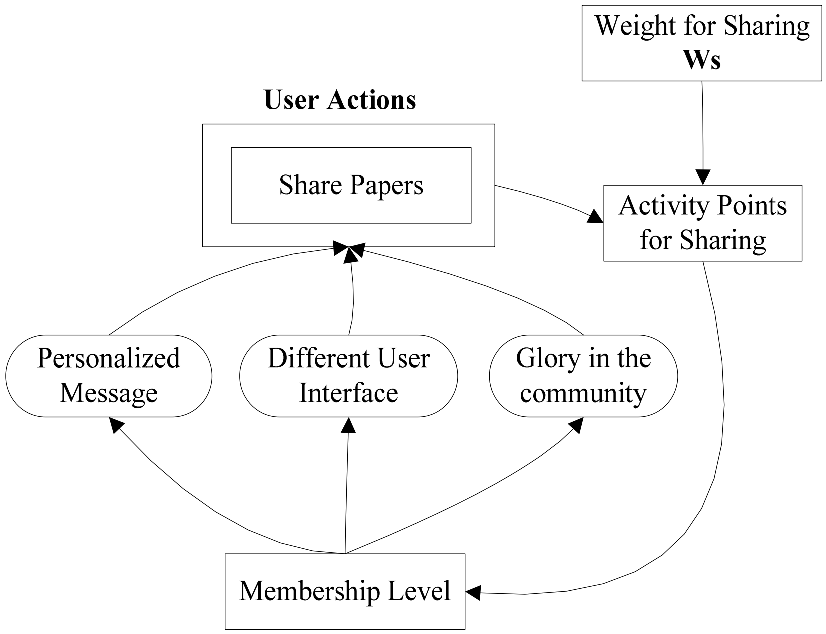


Figure 2.2. The incentive mechanism in Comtella

Users can get rewards (called activity points) by contributing new resources, and they can be motivated by these activity points because when the number of points becomes large enough, the user's membership level is upgraded to a higher one, and the user is shown with a brighter and larger star in the visualization and receives a user interface with corresponding color and services. The reward of each contribution is determined by the reward unit W_S (Weight of Sharing).

A model is developed for this version of the incentive mechanism, which is presented in Section 3.1 in Chapter 3.

2.1.2 Extended version of Comtella incentive mechanism

Figure 2.3 presents an overview of the extended version of the Comtella incentive mechanism. Compared to Figure 2.2, the extended version of the incentive mechanism contains a community model and an individual model, which are used to determine the value of individual adaptive reward units. There are two main actions, *share papers* and *rate papers*. Participants gain activity points, based on the reward units, when they share or rate papers. At the same time, participants will be rewarded a number of C-points when they rate papers, which can make their contributions more visible.

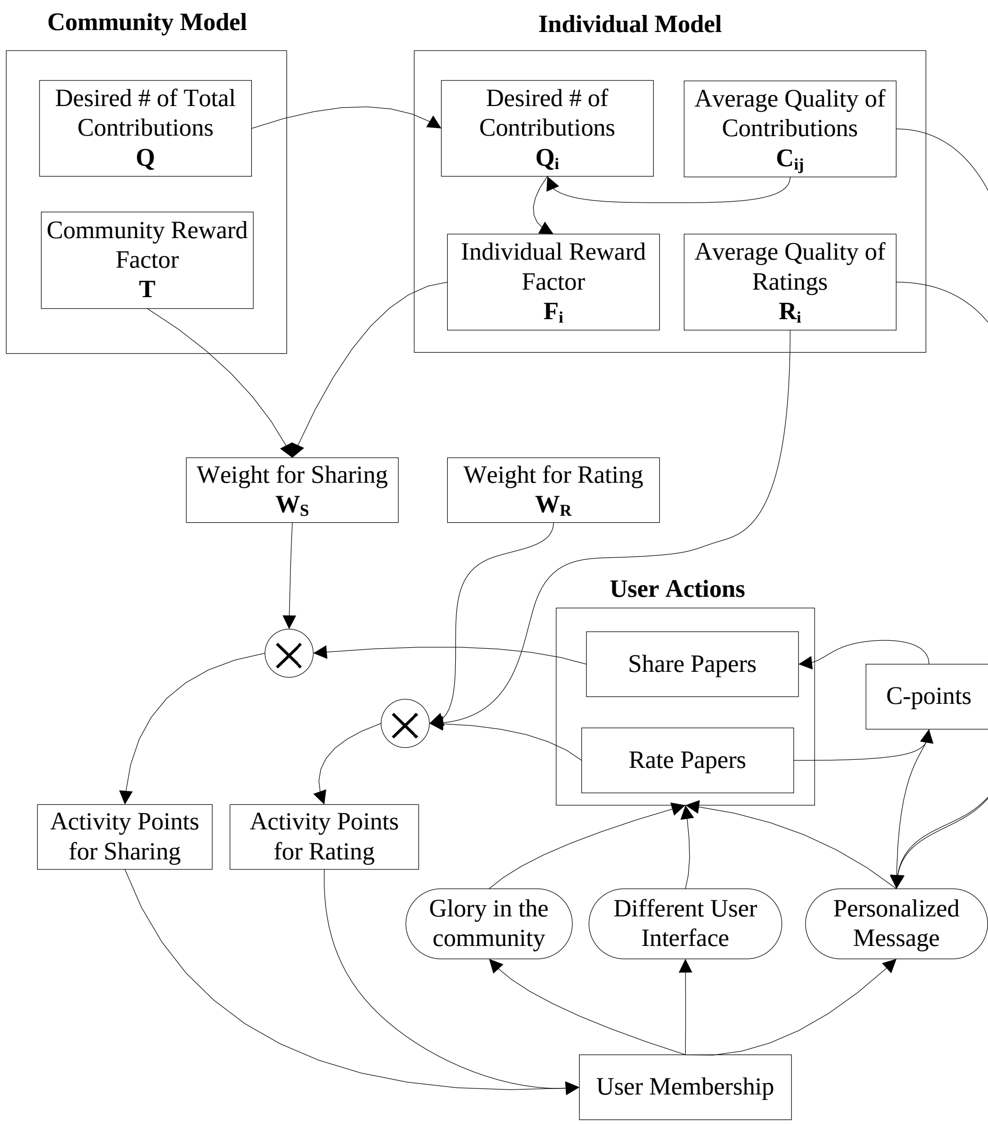


Figure 2.3. The extended version of the incentive mechanism in Comtella

There are three main differences between the two versions of Comtella: quality control, membership decay, individual adaptive reward units.

• Quality control

In order to make the community sustainable in the long term, quality control is needed. User ratings are used to measure the quality of the shared resources in the community. A virtual currency called "C-points" is introduced in the new Comtella incentive mechanism to motivate users to rate resources shared by others. Users can allocate a number of C-points they have earned when sharing web-resources in the community as an "ad fee" in order to make their own contributions more visible in the search result list. When users rate the web-resources, a number of C-points are given to them as reward.

• Membership decay

In the latest version of the mechanism, the number of points will decay over time, which encourages participants in the Comtella community to participate more in order to reach their desired number of points.

• Individual individual adaptive reward units

In the early version of the mechanism, the reward unit for sharing is constant, while the individual reward units for sharing and rating in the latest version of Comtella are adaptive. Two reward functions that can be adapted for a particular period of time are introduced into the motivational mechanism: community reward factor T and individual reward factor F_i, where the subscript stands for individual i. The community factor T reflects in the community how useful the newly shared resources are, and T depends on the time when the participant contributes the resources. It is a function of time, so it is also called the "time-function factor" and makes the rewards dynamic. The individual factor F_i defines the extent to which the newly-shared resources will be rewarded, which is based on the individual patterns of contributions and the need of the community at the moment. It is a function of the weekly contributions of the participant and the desired number of individual weekly contributions (Q_i). The variable Q_i depends on the reputations of individual participants, so the individual factor F_i makes the rewards personalized and is also called the "over-limit factor" for individual i.

In the new version of the incentive mechanism designed by Cheng (2005), the individual adaptive reward units for sharing (W_S) and rating (W_R) are applied to evaluate the engagements (both in quantity and quality) of the users in order to calculate their reputations for each week. Figure 2.4 shows the variation over time of the adaptive reward units, which are different for different individual users. From the figure we can see that the individual reward unit for rating (W_R) is constant, while the individual adaptive reward unit for sharing W_S varies over time. At the first half of the period the value of W_S is higher than W_R, since there is a strong demand for resources. After that period, the value of W_S decreases below the constant value of W_R, since at this point, quality control is needed to help filter out low-quality resources from the great amount of resources in the community.

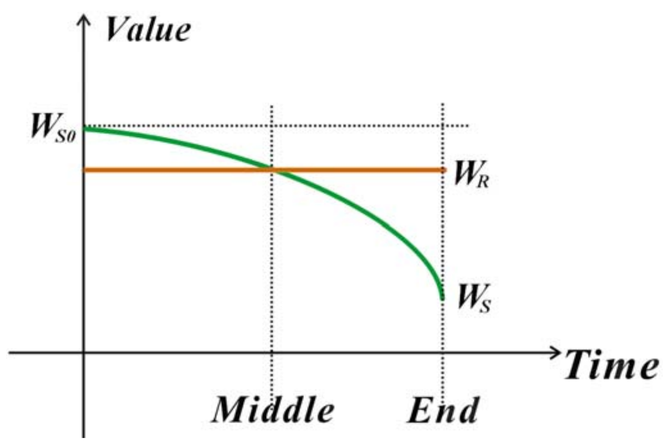


Figure 2.4. The individual adaptive reward units for sharing (W_S) and rating (W_R) (Copied from Cheng 2005 with permission)

In the beginning of the period it is important to encourage contributing new resources, and the reward functions are high. After the total number of contributions approaches the desired number, it is more important to encourage ratings which can help users in the community cope with information overload (Vassileva 2005). As a result, both of the two reward units gradually decrease with time according to the membership levels of the users. In most cases, the individual rewards for the users with low membership levels will decrease much faster than the users with high membership levels, and the individual reward unit will decrease faster after the middle of the period.

The value of W_S depends on three factors: the constant part W_{S0}, the time-function factor T, and the over-limit factor F_i (as shown in Figure 2.3). Equations are presented in Section 4.2.

In Cheng's design (2005), the values of the individual adaptive rewards depend on the individual reputations, which are measured by four different factors: the quantity of the contributions (PaperQuanCr), the quality of the contributions (PaperQualCr), the quantity of ratings (RatingQuanCr), and the quality of ratings given by the participant (RatingQualCr).

Here, the quantity of the contributions (PaperQuanCr) is determined by the reward unit for sharing (W_S) as well as the total number of contributions shared by the participant. Similarly, the quantity of the ratings is determined by the reward unit for rating (W_R), and the total number of ratings given by the participant.

The quality of the contributions (PaperQualCr) is determined by the weekly average ratings earned by the participant, and the quality of ratings (RatingQualCr) is calculated as the sum of the quality of each rating, which is measured by the difference between the value of the rating and the average of all the ratings that the resource gets. More details are provided in Section 4.2.

2.2 Modeling systems: agent-based approach vs. system dynamics approach

Agent-based simulation (ABS) and system dynamics (SD) are two widely used modeling methodologies. Because of the nature of the techniques, each has advantages over the other in particular applications. They can help to generate complementary insights and increase the researchers' understanding of the dynamics of systems and processes.

2.2.1 Agent-based modeling

Agent-based modeling is a powerful modeling technique. It is classified as micro-level simulation (Davidsson 2002) because it tries to reproduce the behaviours of each individual as opposed to macro-level simulations where mathematical models are used to describe the behaviour of the system as a whole. There has been a growing consensus that agent-based simulation is an efficient and useful approach to study different phenomena in social groups. Experts like E. Bonabeau (2002) argue that agent-based modeling is flexible because people can not only increase or decrease the number of agents in the system, but can also adjust the properties of the agents such as their behaviors, the method of communication, and the environment in which the agent resides. He suggests, "ABM can bring significant benefits when applied to human system".

The agent-based modeling approach was developed from the research of Distributed Artificial Intelligence (DAI) in the 1970's. It models the essential characteristics of the individual, as well as the rules for individual interactions and the consequence of the interactions. The basic building block of a system is the individual agent. An agent-based model consists of "a set of agents that encapsulate the behaviours of the various individuals that make up the system, and execution consists of emulating these behaviours" (Parunak et al. 1998).

When using an agent-based approach to model a system, the system is modeled as a collection of autonomous decision-making agents (Figure 2.5). Each agent evaluates its situation individually and makes decisions on the basis of a set of rules (Bonabeau 2002). Most of the actions performed by the users in any given workspace can be attributed to internal urge (to achieve incentives) or reaction to the environment. Agents are elegant metaphors to abstract both of these motives in pieces of software. By creating behaviours and clear communication protocols between agents and environment, elements of reactivity and characteristics of individual agents can be achieved. While creating a simulation, we do not have to incorporate all actions of an individual but only the actions that are relevant for the community. Thus it is possible to start from a very simple model exhibiting some basic behaviours of individuals and more complex actions can be introduced in future iterations.

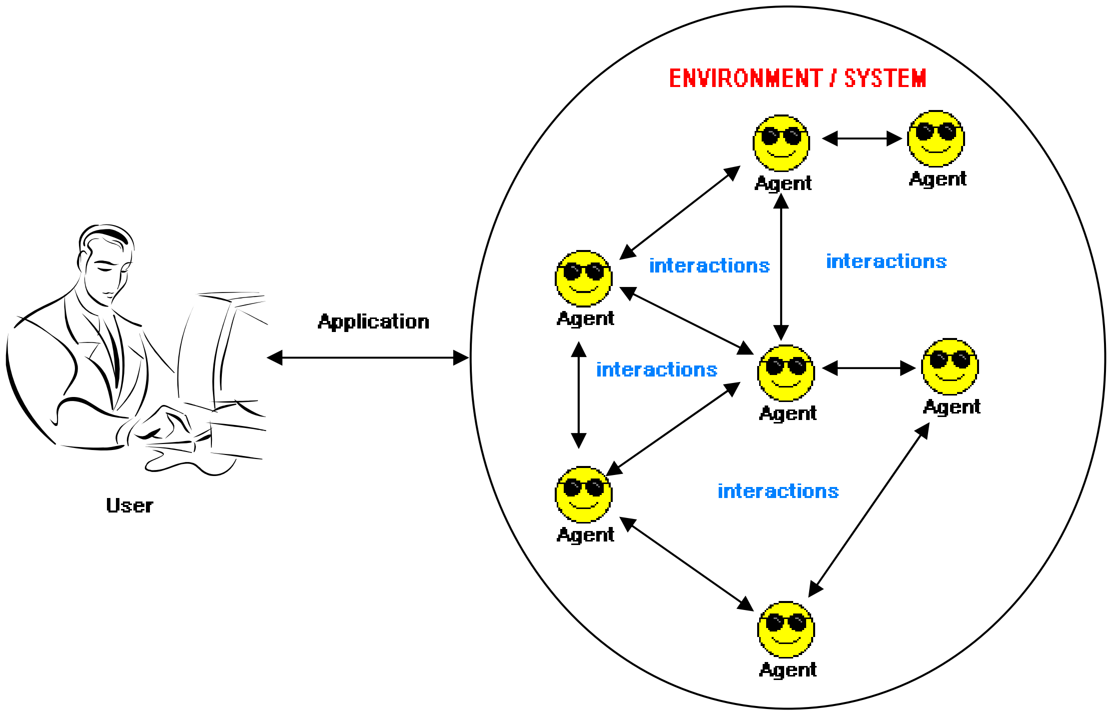


Figure 2.5. An example of multi-agent system

There are several virtual environments and frameworks for agent-based models. The most widely acknowledge ones are:

• StarLogo

Developed at MIT Media Lab, StarLogo is a programmable modeling environment designed for education. It can be used to model the behaviours of decentralized systems, which means systems without an organizer or coordinator, such as traffic jams and ant colonies. Compared with the traditional Logo, StarLogo enables people to control thousands of graphic "turtles" in parallel on the screen by giving commands or writing programs. The application package can be downloaded from the website http://education.mit.edu/starlogo/.

• JADE (Java Agent DEvelopment Framework)

Agent-based systems are intrinsically peer-to-peer. "Each agent is a peer that potentially needs to initiate a communication with any other agent as well as it is capable of providing capabilities to the rest of the agents" (Bellifemine 2003). JADE is an open-source software framework for the development and run-time execution of agent-based peer-to-peer applications. It can be considered as an agent middleware that is independent of the applications and deals with message transport, encoding and parsing, agent life-cycle, etc. Agents are implemented as one thread per agent, and the multi-thread solution will be offered directly by the Java language, where JADE supports the scheduling of cooperative behaviours and tasks in an effective way. More information can be found at the website http://jade.tilab.com/.

- AnyLogic

AnyLogic is a professional simulation tool for modeling complex hybrid, discrete and continuous systems. It is developed as a commercial multi-approach cross-domain tool that supports most approaches from discrete event to continuous modeling, which is quite valuable and efficient for agent-based modeling. Users can build the agent-based model graphically or by writing Java code, and can analyze the model output data by using the powerful data collection and statistical analysis tools of AnyLogic. The application package can be downloaded from the website http://www.xjtek.com/. Compared to JADE and StarLogo, AnyLogic provides more powerful analysis tools and seems to be the most efficient one, so I chose this software (AnyLogic, version 5) to build the agent-based model.

Agent-based models are quite useful in many situations. Since the mid-1990s, these models have been extensively used in solving a variety of social, political, and economic problems in social systems. Examples of applications include traffic and customer flow management, stock market and strategic simulations, operational risk and organizational design simulation, peer-to-peer systems, etc (Bonabeau 2002). Nyik San Ting (Ting and Deters 2003) presented a 3-level simulator on top of existing tools (e.g. the agent platform JADE) in order to study the effects of user behaviour in the performance of complex peer-to-peer networks. Upadrashta et al. (2005) studied specifically the effect of social networks on the performance of peer-to-peer systems by simulating such an environment with the JADE multi-agent system platform. In their model each agent maintained a "friends list" for each category of interest, and used it for searching files in the network. According to their results, "friends lists" reduced search time for queries as well as the number of messages circulating in the system, which provided a useful insight into optimizing search and quality of service in the P2P (peer-to-peer) environment.

With the rapid development of online communities in recent years, researchers also used agent-based models to simulate such systems with the belief that the agent-based approach is a powerful way to study human behaviours in online communities. Each participant in the online community is represented by an agent in the model, which can have several features. Zhang and Tanniru (2005) proposed an agent-based model for virtual learning communities and studied the individual behaviours of participants. In their model the characteristics of a participating agent included expertise level, activity level, sharing level, loyalty, intellectual gain, social gain, etc. They simulated the whole interaction process to better understand, forecast and manage the overall development of the virtual community. Kazuaki et al. (2004) discussed the agent-based simulation approach to analyze online community activities, and the design problem of the decision-making model of the agents that form multi-agent systems.

Although the agent-based approach is quite feasible and powerful, there are also some disadvantages. The extra complexity (time to build, difficulty of calibration and difficulty of formally analyzing) significantly increases the computational requirements and the agent-level detail becomes a cognitive burden of understanding model behaviour (Pavlov et al. 2004). Thus in recent years more and more researchers have modeled online communities using the system dynamics approach.

2.2.2 System dynamics modeling

Developed by Jay W. Forrester in the 1950s, the system dynamics approach is "the study of information-feedback characteristics of industrial activity to show how organizational structure, amplification (in politics), and time delays (in decisions and actions) interact to influence the success of the enterprise" (Forrester 1958; 1961). It emphasizes the use of stocks and flows as well as feedback structures to understand behaviour (Sterman 2000). What system dynamics attempts to do is to understand the basic structure of a system, and thus understand the behaviour it can produce. The model is "a set of equations and execution consists of evaluating them" (Parunak et al. 1998). In other words, a system dynamics model is a system of differential equations.

Stocks and flows are central ideas in dynamics which are formulated mathematically, and the dynamic system behaviours arise due to the flows into or out of the stocks. Figure 2.6 shows a typical system dynamics model with one stock and two flows (one inflow and one outflow).

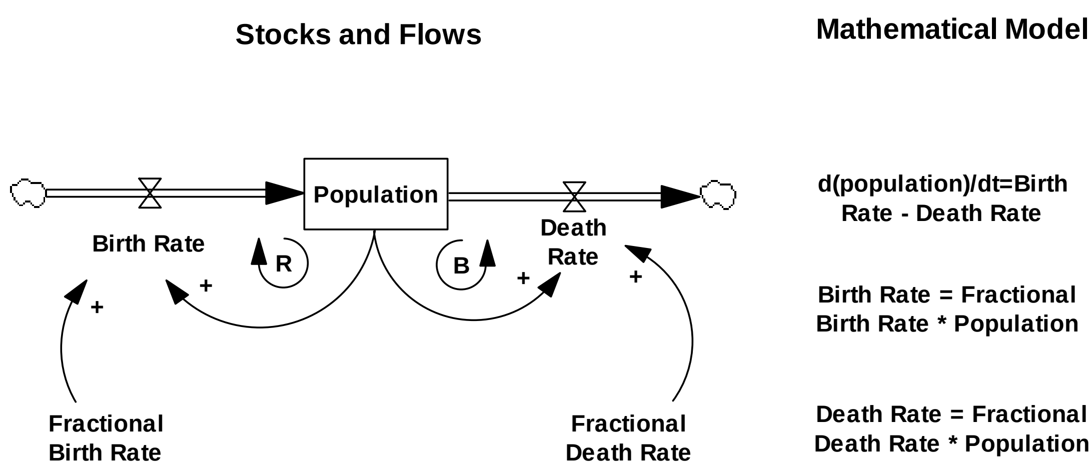


Figure 2.6. An example of typical system dynamics model

Stocks accumulate certain quantities over time, and their value represents the quantity of entities in the stock. As shown in the figure, stocks are represented by rectangles. Flows are represented by pipes pointing into or out of the stocks, and the values are given by the equations mathematically. The flows that point into and out of the stocks are called the inflow and outflow respectively, which determine the change rates of the stock (the inflow minus the outflow). Consequently, if the inflow exceeds the outflow, the value of the stock will increase. On the other hand, the value of the stock will decrease, when the outflow exceeds the inflow. The circles with letter “R” (or “B”) represent the reinforcing (or balancing) feedback loops.

In general there are two basic approaches to system dynamics (Harris & Williams 2005):

- Approaches that map the dynamic relationships in the real world.

The purpose of this kind of system dynamics model is to understand the possible consequences of those relationships or to develop theories about them. Examples include system dynamics models for cellular receptor dynamics (Wakeland et al 2004) and the qualitative politicized influence diagrams (QPID) developed by John Powell et al. (2003) in Bath University estimating behaviours from system structure in the social science area.

- Approaches that simulate the dynamic relationships in the real world.

The purpose of this kind of system dynamics model is to explore the dynamic consequences of different kinds of relationships. Examples include system dynamics models for software project management (Sterman 1992), supply chains management (Akkermans 2001), disease prevention and control (Homer et al. 2004), etc. In Bill Harris’s presentation he said, “The simulation isn’t intended to give you the ‘right’ answer; it’s intended to be another discussant in the room, blending its unique insights with those others provide. But it does help with an area that most of us don’t do well intuitively.” (Harris and Williams 2005).

There are several virtual environments for system dynamics models, such as Powersim Studio, iThink/Stella, NetLogo, etc. The most widely acknowledged ones are:

- Vensim

Vensim is a visual simulation tool designed for system dynamics modeling. It can be used to conceptualize, simulate, analyze, and optimize models of dynamic systems in a simple and flexible way. Models can be built either graphically or in a text editor. The application package can be downloaded from the website http://www.vensim.com/.

- AnyLogic

As mentioned in section 2.2.1, AnyLogic is a professional simulation tool that also supports system dynamics modeling. Users can graphically build the system dynamics model and can analyze the output data using the powerful data collection and statistical analysis tools of AnyLogic. The application package can be downloaded from the website http://www.xjtek.com/.

Table 2.1. General comparison of agent-based vs. system dynamics modeling approach

	Agent-based Modeling	**System Dynamics Modeling**
Focus	Rules of interaction among agents	System structure
Building block	Individual agent	Stocks and Flows
Level of Modeling	Micro/Individual	Macro/Aggregate (typically but not always)
System Structure	Not fixed	Fixed
Time	Discrete or continuous	Continuous

Table 2.1 compares the agent-based modeling and system dynamics modeling approach. From the table we can see that agent-based modeling needs more information because it models the micro-level individual agents. Compared to the agent-based approach, system dynamics modeling may be considered more conceptually descriptive,

and force the modeler to consider carefully the appropriate level of aggregation.

System dynamics is a methodology particularly suited to analyze such complex, large-scale, non-linear, partially quantitative dynamic systems (Sterman 2000). One important advantage of system dynamics is that the model structure leading to system behaviours is made transparent. However, one problem of system dynamics models is that the model structure has to be fixed before starting the simulation (Richardson 1991). The range of system dynamics applications includes the fields of education, politics, medicine, economics, biology, environmental studies, information science, society organization, etc. It can be applied whenever problems can be expressed involving variable behaviours through time, such as supply chains, pricing and capital investment, group dynamics, population dynamics, market share, and business cycles.

In recent years some researchers have developed system dynamics models of online communities. The design of community information systems required “much theoretical research to solve design problems” (De Moor 2005) and system dynamics performed as a much-needed research instrument for community research.

For example, Quentin Jones et al. (Jones et al. 2002) used the system dynamics approach to examine internet-based group communication as "mass interaction" in a virtual community. The purpose of their study is to explore the impact of systems effects in Usenet discourse. They described the non-linear feedback loops generated by user information overload, and comparatively analyzed various computer-mediated communication (CMC) technologies in terms of group-level usability.

In 2004, Diker (2004) reported developing a system dynamics model to study growth problems in a special kind of virtual community: open online collaboration communities (OOCC). He first developed a dynamic simulation model to simulate the OOCC as an example, and in the next phase interviewed a group of members of an actual OOCC in order to test the applicability of the simulation model. Combining results from the interviews and results from the simulation model, Diker explored the dynamic interactions among user motivation, participation, collaboration and the quality of products in the OOCC, which helps to explain the dynamics of growth in OOCCs.

De Moor (2005) showed how system dynamics simulation could play an important role in a systematic development of design theory for community information systems. He presented a meta-model of system development which studies the role of theory in community information system (IS) design, and showed a simple model of an online community, where the key concept of interest is ‘community spirit’. As a result he addressed the idea of practically embedding system dynamics modeling in the study of online communities.

2.3 Research objectives

The main objective of the thesis is to study through computer simulation the effects of incentives mechanisms for participation in this particular online community, Comtella.

Compared to previous research, this work is distinctive in several ways. First, instead of the general growth problems in online collaboration communities, the model proposed in this thesis will focus on the particular incentive mechanism used, and its effect on user participation in the Comtella community. There are several factors that need to be considered, such as the reward factor, thresholds for membership upgrading, the decay rate, etc. Second, our study focuses on a small-scale educational online community which maintains a certain number of users. This distinguishes it from other related studies on large open online communities by other researchers. Third, since Comtella is designed and developed at the MADMUC Lab of University of Saskatchewan, I have access to the Comtella database (presented in Appendix B) containing the data from the actual system deployment, which can be used for model evaluation. Fourth, to the best of my knowledge no one has applied and compared system dynamics and agent-based approaches to model the same phenomenon in the area of online communities. Last, the model also focuses on how to increase the percentage of the middle level participants, which is unique in the research on Comtella community. Cheng (2005) mentioned that for an effective incentive mechanism, most of the participants should be classified into the middle levels. These participants may be motivated more easily by the incentive mechanism than other participants, not only because they have gained some rewards and a higher membership is achievable for them, but also because they have fear of losing their current membership level and rewards. In addition, maintaining a large percentage of middle level participants will also stimulate participants with low membership levels. In Comtella the middle level participants are considered to have either bronze or silver membership levels. As a result, bronze and silver members are considered more active in the community and their contributions need to be maintained.

The results of this work are expected to help investigate the user participation problem, evaluate and improve the current incentive mechanism, and find out whether the factors in the current incentive mechanisms (membership thresholds, reward unit functions, etc.) are efficient enough to ensure user participation. The results may also suggest further guidelines for setting thresholds and reward units in the real Comtella system.

There are several factors that need to be considered:

- How to classify the participants into different groups in the community for study, and the proportions of different groups.
- How different groups impact the whole system.
- Which elements have the most impact on users (especially the middle level participants, bronze and silver members) and can effectively motivate them, and how to adjust these factors to make the community more sustainable.
- How users in the community behave according to the memberships based on the current incentive mechanism (in order to evaluate the incentive mechanism).
- How to optimize the system using the simulation models and suggest guidelines for setting thresholds and reward units.

The comparison of the system dynamics and agent-based methodology to simulate the system can lead to interesting findings regarding the suitability of each of these methods to study online communities.

2.4 Summary

This chapter introduced the two different versions of Comtella system, the research approach, as well as the objectives of the study. In Section 2.1, the incentive mechanism used in the early version of Comtella as well as the extended version of the incentive mechanism used in the latest version of Comtella were described. The related research on systems modeling using agent-based approach and system dynamics approach was reviewed in Section 2.2, to provide a general view of community research and system modeling. The main focus of the study and the difference from previous research efforts were presented as well. In Section 2.3, the research goal of this thesis was presented.

In the next chapter, two prototypical simulation models of the Comtella incentive mechanism of hierarchical memberships are presented. These models are system dynamics models which are developed by using Vensim software.

CHAPTER 3

SYSTEM DYNAMICS MODEL OF THE COMTELLA INCENTIVE MECHANISM

In this chapter, a system dynamics approach to model the incentive mechanisms in Comtella is presented. Section 3.1 presents the system dynamics model for the early version of the Comtella incentive mechanism, which exhibits the basic system behaviours. Although this model is prototypical and models an early version of Comtella incentive mechanism, it is still useful and provides insights into the dynamics of online communities. In Section 3.2, another system dynamics model for the extended version of the Comtella incentive mechanism is described. Compared to the early version of the incentive mechanism, the extended version considers the rating behaviours of participants, the quality control, the membership decay, and individual adaptive reward units. For brevity, in this chapter the two models are called the "first version model" and the "second version model" respectively. Comparison of the simulation results and the validation experiments are presented in Section 3.3.

3.1 First version model

The system dynamics simulation model of the Comtella community was built using the visual modeling tool Vensim (http://www.vensim.com/download.html). It provides a simple and flexible way of building simulation models from causal loop or stock and flow diagrams.

3.1.1 *Model description*

The causal loop diagram shown in Figure 3.1 presents the basic conceptual model of the early version of Comtella incentive mechanism.

As introduced earlier in Section 2.1.1, in the early version of the Comtella incentive mechanism, the only action of participants is contributing new resources. Quality control is not included in the incentive mechanisms, so the feedback structure is quite simple. Activity points are used to measure the membership levels. By contributing new resources in the community, the participants are rewarded a number of activity points based on the share rates (that is, the number of contributions shared per week) and the reward units (represented by variable W_S, as shown in Figure 2.2, page 9). As a result, the membership levels are upgraded to higher ones according to the membership upgrading thresholds, and the participants will get encouraged to share more resources.

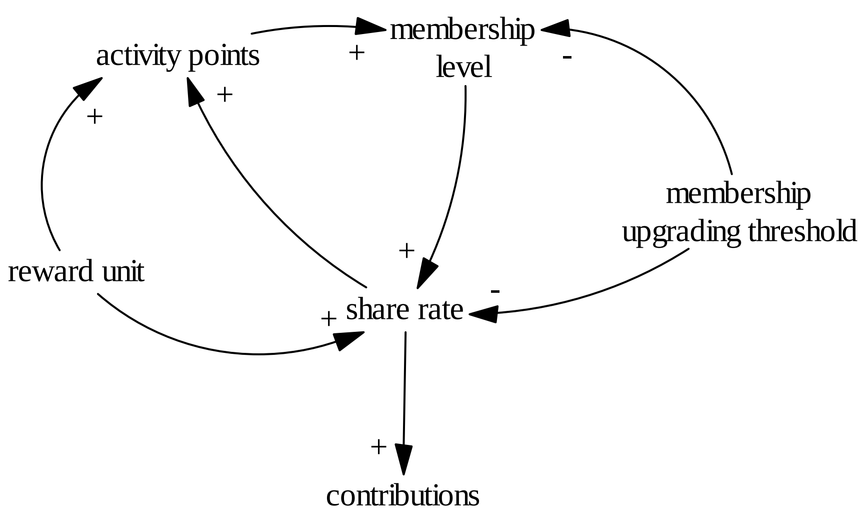


Figure 3.1. Concept model for the first version model (causal loop)

The most important system inputs in the model are:

- Reward unit (W_S): For each contribution, the number of activity points given to users as a reward.

- Membership upgrading threshold: The number of activity points (threshold) needed to reach the next membership level. For a fixed reward unit, if the membership upgrading threshold is too low, users in the community can reach a high level in the membership hierarchy easily and therefore stop participating and contributing new resources after they secure the highest level of membership. On the other hand, if the membership upgrading threshold is too high, users will gradually lose interest in sharing and become free-riders.

The most important system outputs in the model are:

- Membership levels: In our case the whole population in the Comtella community is divided into four user groups based on different membership levels. To measure how these user groups change over time, an aging chain is applied to the system dynamics model, a modeling pattern which is widely used to capture the demographic structure of a population. It includes a set of member groups (according to different memberships of Comtella users). Also, the rates of inflow and outflow of different user groups have to be determined, which cause the population of different user groups to change over time.
- Contributions: The total number of contributions is an important factor which measures whether the online community is successful or not. I want to measure the share rates (contributions shared per week) of the different user groups according to how much they contribute to the total number of contributions in the community.

Figure 3.2 further explains the concept model of Figure 3.1, and shows the main reinforcing loop in the model. It focuses on the feedback relationships among user memberships, activity points and share rates of different user groups. Thresholds have damping effects towards both memberships and share rates. Further details will be provided in Section 3.2.

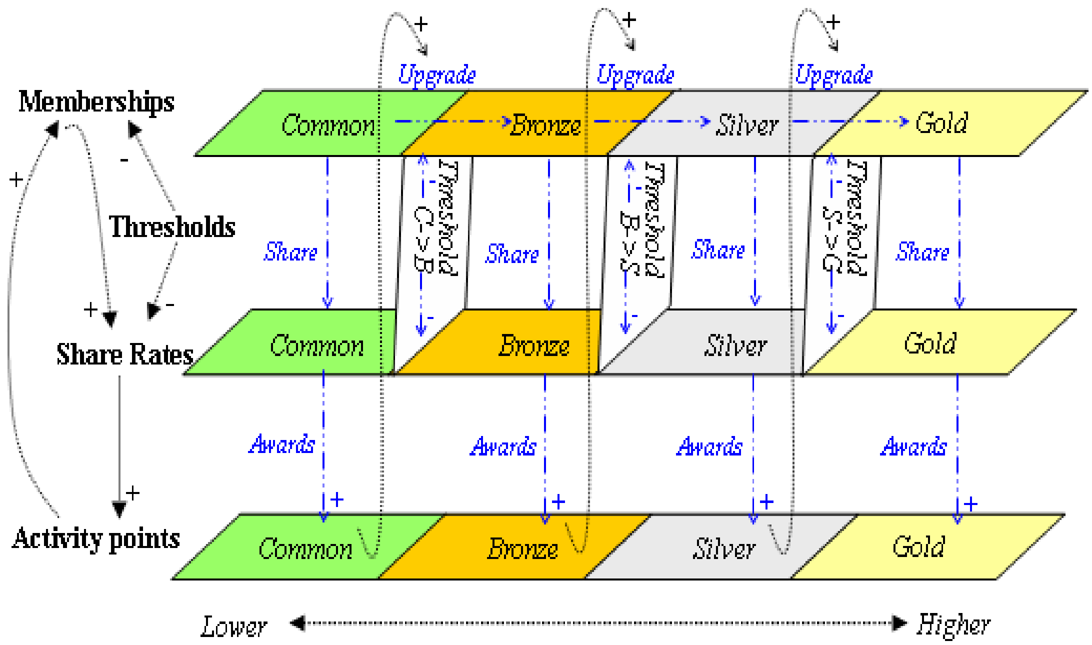


Figure 3.2. Concept model for the first version model (cont.)

3.1.2 Model structure

For the first version model, in order to measure the important factors, the model is divided into two sectors that will be discussed in detail in the next section:

1. The 'Population sector' uses an aging chain to represent the demographic structure of the population. The population of each user group is modeled as a stock.

2. The 'Share rate sector' models the share rate control for different user groups.

A description of the first version model has been published already in the proceeding of HICSS 2007 (Mao, Vassileva, & Grassmann, 2007).

3.1.2.1 Population sector

The 'Population sector' (the part in the dashed oval of Figure 3.3) models the demographic structure of the population. The time unit of this model is set to be one week, so all the parameters represent weekly quantities.

Here, the variables surrounded by angle brackets '< >'with light gray color are called shadowed variables, which means those variables are defined somewhere else already. Shadowed variables are used in most system dynamics models in the thesis, in order to reduce clutter and increase the clarity of the model structure.

In this sector, the population of the community is the sum of all the numbers of users in different user groups. Each user group is modeled as a stock, which changes weekly according to inflow and outflow:

$$\textit{Weekly change rate} = \textit{inflow} - \textit{outflow} \tag{3.1}$$

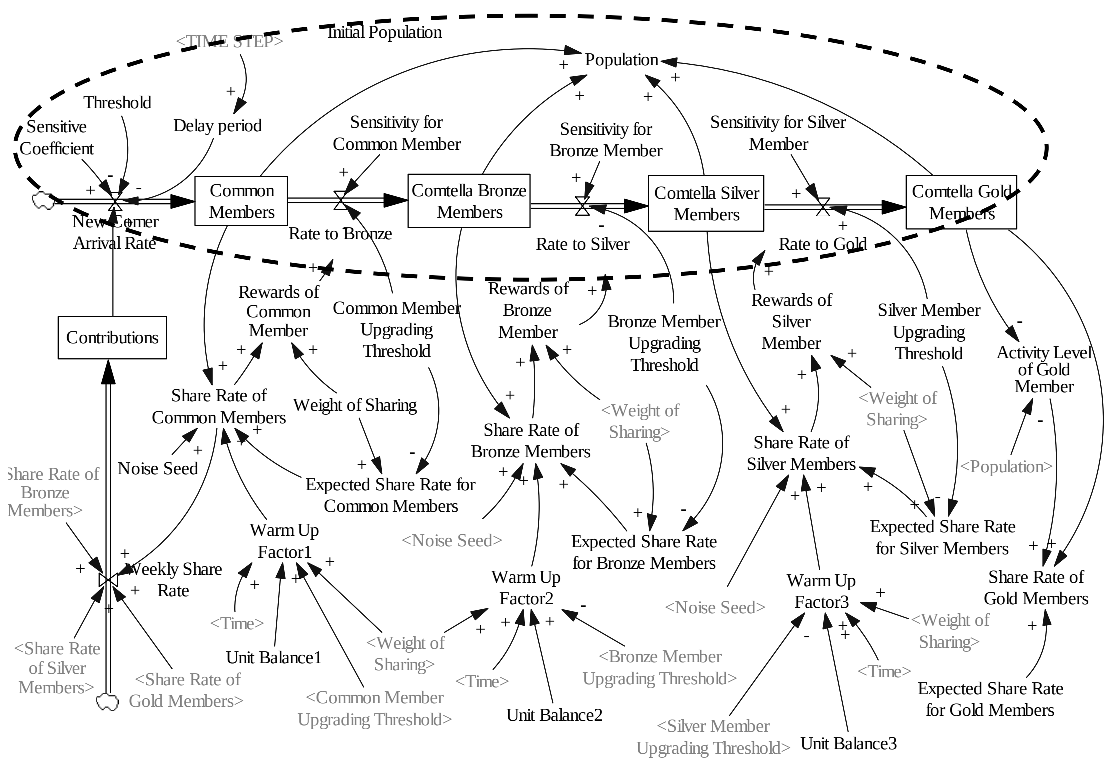


Figure 3.3. Population sector

For each stock that represents one user group, the equations for the inflows and outflows have to be formulated, in order to calculate the population size of the user group. Take the stock "*Comtella Bronze Members*" as an example. During the simulation there are some common members who upgrade their membership levels to bronze membership levels by sharing, and the number of these common members who upgrade their membership levels can be considered as the inflow of the stock "*Comtella Bronze Members*". Since the number of common members upgraded to bronze membership level in each week is represented by the outflow of the stock "*Common Members*", the inflow of the stock "*Comtella Bronze Members*" equals to the outflow of the stock "*Common Members*", which is called "*Rate to Bronze*" in the system dynamics model.

Similarly, part of the bronze members will upgrade their membership levels to silver membership levels, and the number of these bronze members can be considered as the outflow of the stock "*Comtella Bronze Members*". Since the number of bronze members whose membership level is upgraded to silver membership level in each week is represented as the inflow of the stock "*Comtella Silver Members*", the outflow of the stock "*Comtella Bronze Members*" equals to the inflow of the stock "*Comtella Silver Members*", which is called "*Rate to Silver*".

Consequently, the weekly change rate of bronze members equals to "*Rate to Bronze*" minus "*Rate to Silver*":

$$\frac{d}{dt}(Comtella\ \ Bronze\ \ Members) = Rate\ \ to\ \ Bronze - Rate\ \ to\ \ Silver \tag{3.2}$$

The formula for the change rate of common members to become bronze members is defined as:

$$Rate\ \ to\ \ Bronze = \frac{Sensitivity\ \ for\ \ Common\ \ Member * Rewards\ \ of\ \ Common\ \ Member}{Common\ \ Member\ \ Upgrading\ \ Threshold} \tag{3.3}$$

Here, the "*Sensitivity for Common Member*" is a dimensionless constant, and the variable "*Rewards of Common Member*" represents the total number of activity points that are rewarded to all common members in a week, which is calculated as the product of the reward unit (which is represented by the variable "*Weight of Sharing*") and the number of contributions that are shared by all common members:

$$\begin{aligned} &Rewards\ of\ \ Common\ Member = \\ &\qquad Share\ rate\ of\ \ Common\ Member * \ Weight\ of\ \ Sharing \end{aligned} \tag{3.4}$$

In order to make it more general, there is also a positive feedback in the model to represent the incoming rate of new users in the community, even though this positive feedback structure is not used in my model. There is a delay in this process because it is assumed that new participants will be attracted to the online community only when the total number of resources is high enough to motivate them and reaches a threshold. A variable called "*Threshold*" is used to represent this threshold, and it has the unit of Link (which stands for the URL of a paper / web-resource). The incoming rate for new participants is calculated as:

$$New\ Comer\ Arrival\ Rate = \frac{Sensitivity * Contributions}{(Threshold * Delay\ period)} \tag{3.5}$$

In the formulas the variable "*Sensitivity*" also works as a dimensionless constant. In case of Comtella where there is no inflow of new users, the scalar "*Sensitivity*" is set to be 0 because it is assumed that the number of students in a particular course is fixed. The delay period is represented by the variable "*Delay period*", which has a unit of Week.

3.1.2.2 Share rate sector

The 'Share rate sector' models the share rates of different user groups. The total number of contributions is modeled as a stock with weekly change rate "*Weekly Share Rate*":

$$\begin{aligned} Weekly\ Share\ Rate = {} & Share\ Rate\ of\ \ Bronze\ Members \\ & + Share\ Rate\ of\ \ Common\ Members \\ & + Share\ Rate\ of\ \ Gold\ Members \\ & + Share\ Rate\ of\ \ Silver\ Members \end{aligned} \tag{3.6}$$

The ‘Share rate sector’ was divided into two parts, one corresponding to non-gold members (the part in the dashed oval of Figure 3.4) and one corresponding to gold members (Figure 3.5).

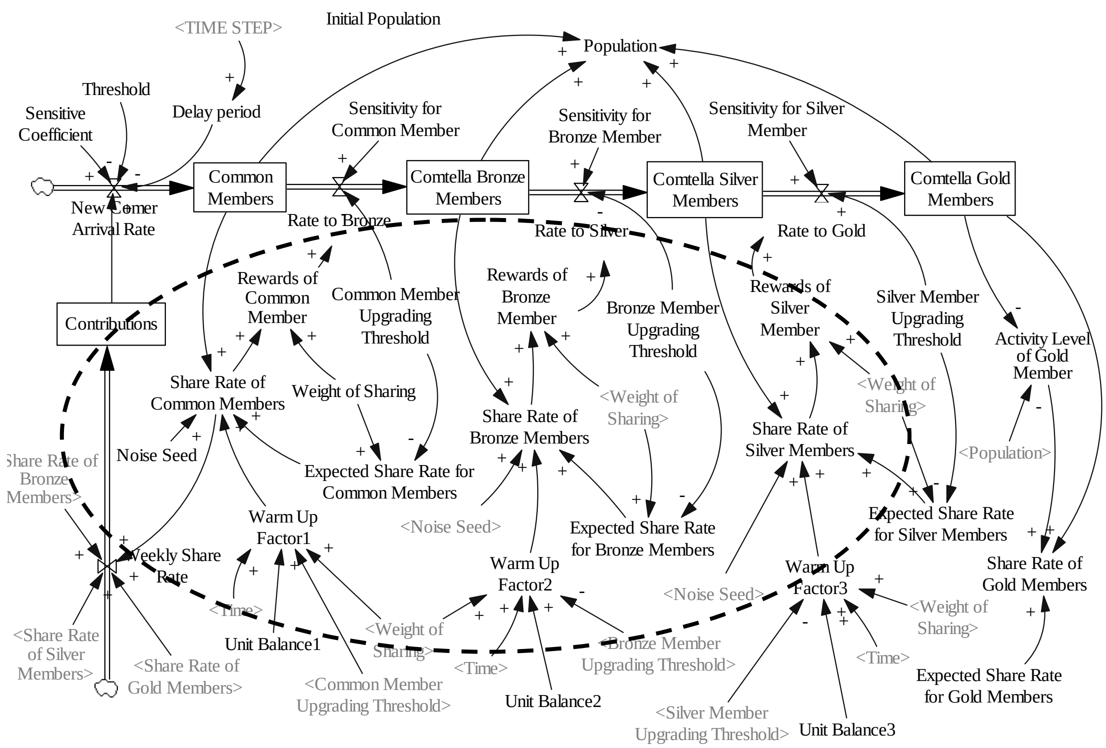


Figure 3.4. Share rates of common, bronze and silver members

Common, bronze and silver members can be motivated by the membership upgrading process, which is affected by the reward unit and membership upgrading thresholds. In the non-gold member case, the formulas of the share rates for common, bronze and silver members are similar and they have the potential applicability of subscripting, which means these state variables can be modeled as arrays. Users are motivated by the rewards. In order to get a higher membership level, they will contribute more when their membership level is upgraded.

Take “*Share Rate of Common Members*” as an example. Each common member can be considered as an entity in the stock "*Common Members*", and the number of these entities equals the value of the stock. Since the behaviours of these entities in the stock are driven by the same inflow and outflow, it is assumed that all the entities have the same behaviours. In other words, common members have the same share rate.

For each common member, the variable "*Expected Share Rate for Common Members*" reflects a desired number of weekly contributions shared by common members. The value of the desired number is calculated by dividing the membership upgrading threshold by the reward unit for sharing. According to the usage data from the deployment of Comtella, participants share less in the beginning, and gradually raise their weekly share rate to the desired number of weekly contributions. I called this process as the "warm-up stage", and a combined measure is used to measure how fast the weekly individual share rate can reach the desired share rate. For example, the combined measure of the warm-up stage for common members is called "Warm-Up Factor1".

In order to make the dynamic model more realistic, a stochastic control factor is introduced in the calculation of share rates, which is a random number (dimensionless). Since the uniform distribution is one of the most important distributions that are widely used for the generation of random numbers, the stochastic control factor is calculated as: *RANDOM UNIFORM (0.3, 1, Noise Seed),* where the variable "*Noise Seed*" is a dimensionless constant. The range of the random number [0.3, 1] and the value of the *Noise Seed* are chosen arbitrarily.

Considering all the factors mentioned above, the weekly individual share rate for each common member is measured by the product of the warm-up factor "*Warm Up Factor1*", the desired number of weekly contributions "*Expected Share Rate for Common Members*", and the stochastic control factor *RANDOM UNIFORM (0.3, 1, Noise Seed)* to make the weekly individual share rates. The value of weekly individual share rate will converge to a desired value gradually by assumption (with stochastic control), and the formula for the share rate of all common members equals the product of the population size of common members and the weekly individual share rate of common members:

$$\begin{aligned} \textit{Share Rate of Common Members} \\ &= \textit{Common Members} \\ &* \textit{RANDOM UNIFORM (0.3,1, Noise Seed)} \\ &* \textit{Expected Share Rate for Common Members} \\ &* \textit{Warm Up Factor1} \end{aligned} \tag{3.7}$$

In this formula the units of "*Common Members*", "*Expected Share Rate for Common Members*" and "*Warm Up Factor1*" are Person, Link/Person and 1/Week respectively, which implies that the *Share Rate* has the unit of Link/Week.

The following equation is used to measure the warm-up factor. This warm-up stage is considered to be affected by the reward unit which is represented by the variable "*Weight of Sharing*", the membership upgrade thresholds, and the time. This equation works properly only when the unit of time is Week, and might be changed if the time has another unit.

$$\begin{aligned} \textit{Warm Up Factor1} = \textit{Unit Balance1} * \\ \left(1 - \frac{1}{1 + \left(\dfrac{\textit{Weight of Sharing}}{\textit{Common Member Upgrading Threshold}}\right) * \textit{Time}}\right) \end{aligned} \tag{3.8}$$

For gold members there is no incentive to upgrade their membership, so the reward unit and thresholds have little impact on the share rate.

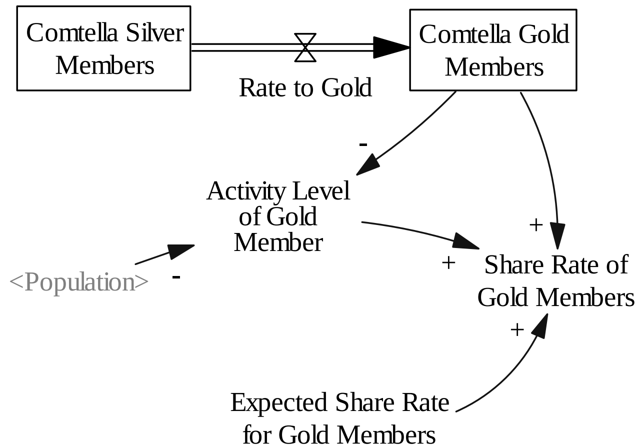


Figure 3.5. Share rate of gold members (gold member part of Figure 3.4)

Since it is not clear what motivates the gold members in Comtella, I assumed that gold members can be motivated by the average activity level of the whole user group. Here the activity level is a variable with range [0, 1] that is affected by the percentage of gold members in the population:

$$\mathit{Activity\,Level\,of\ Gold\,Member} = 1 - \frac{\mathit{Comtella\,Gold\,Members}}{\mathit{Population}} \tag{3.9}$$

Similar to the other three user groups, the formula for share rate is defined as the product of the population size of gold members and the weekly individual share rate of gold members. Here the weekly individual share rate of gold members is calculated by multiplying the activity level of gold members by the desired individual weekly share rate:

$$\begin{aligned} \mathit{Share\ Rate\ of\ Gold\ Members} &= \mathit{Expected\ Share\ Rate\ for\ Gold\ Members} \\ &\quad {*}\mathit{Activeness\ Level\ of\ Gold\ Member} \\ &\quad {*}\mathit{Comtella\ Gold\ Members} \end{aligned} \tag{3.10}$$

3.2 Second version model

3.2.1 Model description

After designing and evaluating the first version model, I developed the second system dynamics model for the extended version of incentive mechanisms in the Comtella community, and the model is called the “second version model” in the thesis. As mentioned in Section 2.1.2, the extended version of the Comtella incentive mechanism includes quality control, individual adaptive reward units, and membership decay. Compared to the first version model, there are some extensions:

- Both sharing and rating behaviours are modeled, and activity points are rewarded based on both of these activities. The participants in the community can get rewards by either sharing resources or rating shared resources in the community that can be used to upgrade their memberships and encourage them to participate more in the community.
- The virtual currency “C-points” was introduced in the second version model as rewards to motivate users to rate. Participants can use C-points to display their contributions more prominently, which helps them to get a higher chance of receiving ratings from the others. Therefore the participants are motivated to earn more C-points through rating if they want to get higher reputations.
- Activity points and C-points decay over time according to the membership level of users, which encourages participants in the community to share more resources or give more ratings in order to reach their desired number of points.

Based on the concept model for the first version model (Figure 3.1, page 26), Figure 3.6 shows the concept model for the second version model.

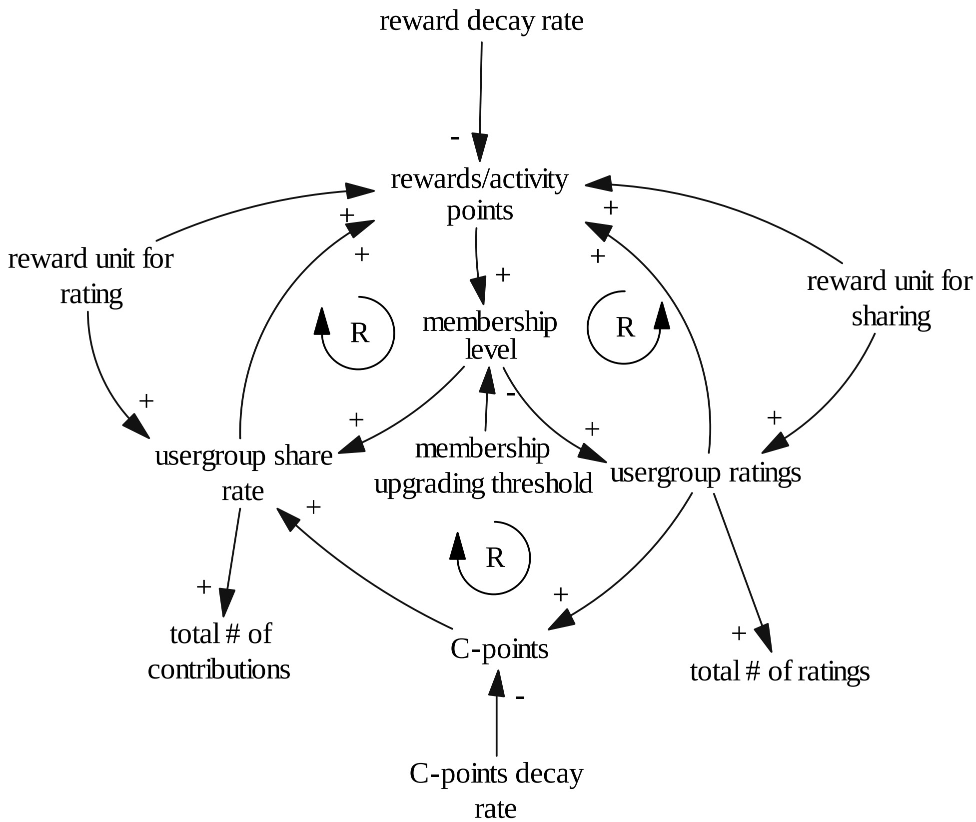


Figure 3.6. Concept model for the second version model (causal loop)

As demonstrated in Figure 3.6, there are two sources of activity points: the activity points obtained by contributing new resources, and the activity points obtained by giving ratings. The number of the activity points in these two parts depends on the individual adaptive reward units and the weekly share rates / ratings. When there are more resources that have been shared (or rated), more activity points will be awarded to the participants. As a result, upgrading the membership level becomes easier, which encourages the participant to increase their weekly share rates (weekly ratings). These reinforcing effects form a loop, which is represented as circle arrows with a letter 'R' in the middle, which means the loop has a positive feedback structure.

Activity points and C-points decay over time, with the rates called "*reward decay rate*" and "*C-points decay rate*" respectively. The number of C-points depends on the decay rate of C-points and the weekly ratings given by the participant. Since a number of C-points are awarded for rating, the number of C-points increases when the value of the weekly ratings increases.

3.2.2 *Model structure*

The second version model is divided into three sectors that will be discussed in detail in the following sections:

1. The ‘Contribution sector’ is based on the ‘Population sector’ and the ‘Share rate sector’ of the first version model. It uses an aging chain to represent the demographic structure of the population, and models the reinforcing loop of sharing behaviours.

2. The ‘Reward sector’ models the dynamics of activity points over time.

3. The ‘C-point sector’ models the reinforcing loop of rating behaviours and dynamics of C-points.

3.2.2.1 Contribution sector

Figure 3.7 shows the demographic structure of the population, as well as the dynamics of sharing behaviours. As in Figure 3.3, users are represented by stocks according to different membership levels. The membership upgrading process is the same as in the first version model; and the membership decay process is represented by the flows which go into the stocks that represent users with a lower level of membership.

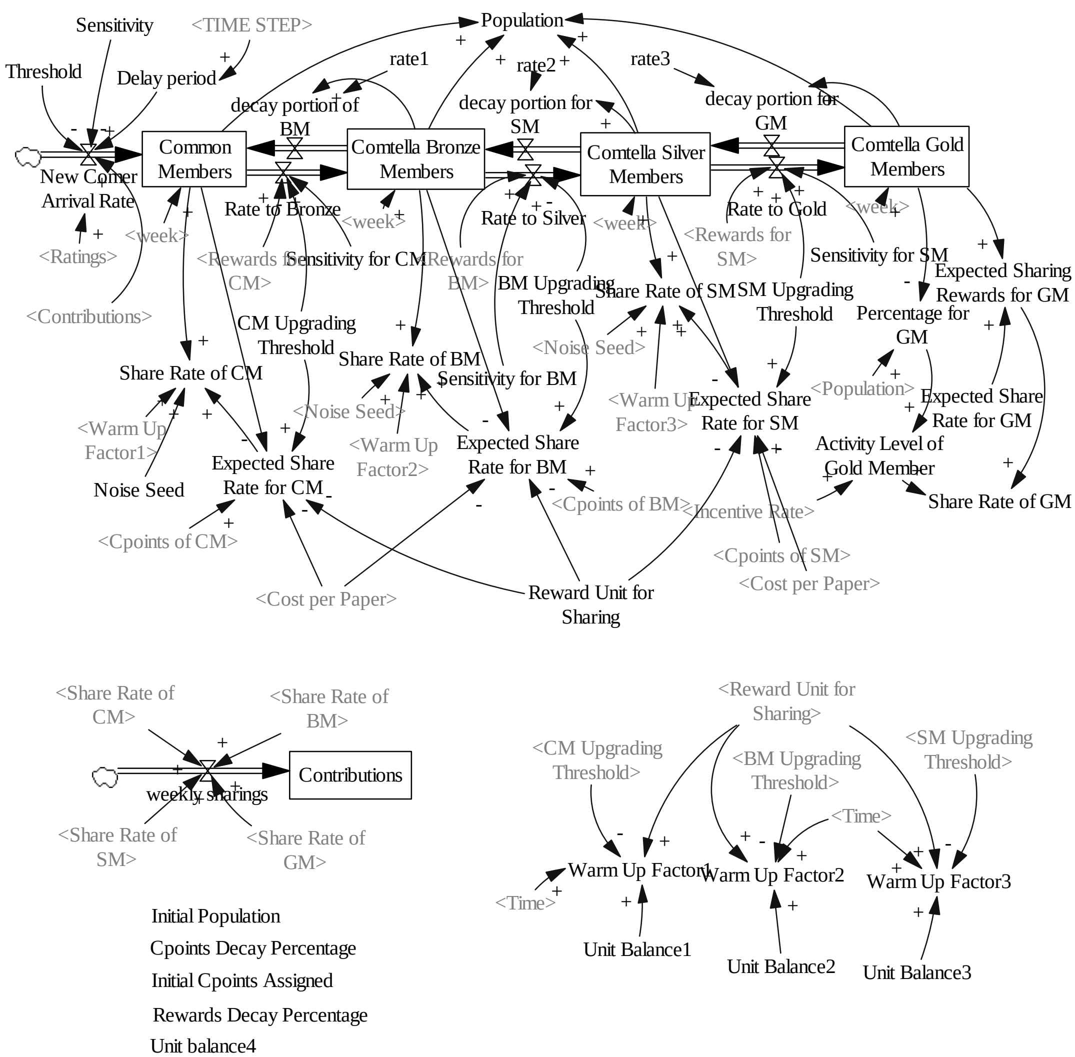


Figure 3.7. Contribution sector

The only difference is the membership decay. Take the stock "*Comtella Bronze Members*" as an example. Due to the adaptive incentive mechanisms, membership levels will decay over time. As a result there are flows coming from a higher membership level to a lower membership level. There are two inflows: the weekly change rate of common members to become bronze members, as well as the weekly change rate of silver members to become bronze members. They are represented by the variables "*Rate to Bronze*" and "*decay portion for SM*" respectively. Simarly, the outflows are the weekly change rate of bronze members to become silver members, as well as the weekly change rate of bronze members to become common members. They are represented by the variables "*Rate to Silver*" and "*decay portion for BM*" respectively. The differential equation for the stock "*Comtella Bronze Members*" is defined as:

$$\begin{aligned}
&\frac{d}{dt}(Comtella \quad Bronze \quad Members) = \\
&\qquad Rate \quad to \quad Bronze + decay \quad portion \quad for \quad SM \\
&\qquad - Rate \quad to \quad Silver - decay \quad portion \quad for \quad BM
\end{aligned} \tag{3.11}$$

The formulations of share rates in the sector are analogous to the first version model (Equation 3.7 and Equation 3.8). Taking "*Share Rate of CM*" as an example, the equation is:

$$\begin{aligned}
Share\ Rate\ of\ CM = \ & Comtella\ Members \\
& * Expected\ Share\ Rate\ for\ CM \\
& * Warm\ Up\ Factor1
\end{aligned} \tag{3.12}$$

Here the variable "*Warm Up Factor1*" is calculated by Equation 3.8.

Since C-points are involved in the dynamics of sharing behaviours now, the formula of "*Expected Share Rate for CM*" is revised. There are two main factors that can affect the desired weekly share rate. One factor is the number of new contributions that are needed to get enough activity points for the next membership level, and its value is calculated by dividing the membership upgrading threshold by the reward unit for sharing. The other factor is the total number of C-points that participants can spend to make their contribution more prominent. When participants have a large number of C-points, they have a high possibility to share more resources according to how many C-points they have. It is assumed that each participant in the community has equal probability to spend a number of C-points when the participant shares new resources. So a variable "*Cost per Paper*" is introduced to reflect the estimated average number of

C-points assigned to each new resource in order to increase its ranking in the search results, and the value of the variable "*Cost per Paper*" is obtained from the real data.

Taking the common members as an example, a stock called "*Cpoints of CM*" measures the total number of C-points of all common members. Then for each common member, the average number of C-points of the participant is calculated by dividing "*Cpoints of CM*" by the population size of common members. As a result, the maximum number of resources which each common member is able to share is calculated by dividing the average number of the C-points by the estimated average number of C-points assigned to each new resource ("*Cost per Paper*").

Taking these two factors mentioned above into consideration (the number of new contributions that are needed to get enough activity points for the next membership level, and the total number of C-points that participants can spend for new resources), the desired weekly share rate for common members is calculated as:

$$Expected\ Share\ Rate\ for\ CM = MAX\left(\frac{CM\ Upgrading\ Threshold}{Reward\ Unit\ for\ Sharing}, \frac{Cpoints\ of\ CM}{Cost\ per\ Paper * Common\ Members}\right) \tag{3.13}$$

The units of "*Cpoints of CM*", "*Cost per paper*" and "*Common Members*" are Point, Point/Link and Person respectively, which implies that the variable "*Expected Share Rate for CM*" has the unit of Link/Person.

In this way, the desired share rate is dynamic. It is not only decided by membership upgrading thresholds and the reward unit for sharing, but also is affected by the average number of C-points gained by participants in the membership group.

For the reason mentioned before in Section 3.1.2.2, gold members have separate formulas that are different from non-gold members. As mentioned before, it is assumed that gold members can be motivated by the average activity level of the whole user group. Compared to Equation 3.9, a variable "*Incentive Rate*" is introduced which approximately describes how much motivation is gained by gold members from rewards, or how much they can be motivated by the rewards. The basic idea is that if gold members get more rewards than in the previous week, they are motivated to share more resources.

The activity level is defined as:

$$\begin{aligned} &\textit{Activity Level of Gold Member} = \\ &\qquad \textit{Percentage for GM} * (1 + \textit{Incentive Rate}) \end{aligned} \tag{3.14}$$

Similar to the first version model, the variable "*Percentage for GM*" is a dimensionless variable with range [0, 1] that is affected by the percentage of gold members in the population. Equation 3.9 can be used to calculate the value of this percentage.

The following formula shows how to calculate the value of the variable "*Incentive Rate*", and there is further explanation in Appendix C:

$$\textit{Incentive Rate} = \frac{\textit{new GM rating rewards-decay rate for GM} + \textit{Awards for GM}}{\begin{pmatrix} \textit{Awards for GM-Reward Unit for Sharing} \\ \textit{*Expected Sharing Rewards for GM*Percentage for GM} \end{pmatrix}} \tag{3.15}$$

3.2.2.2 Reward sector

Figure 3.8 illustrates the dynamics of rewards. Participants can gain rewards by either sharing new resources, or giving new ratings, so the total number of rewards is defined as the sum of these two kinds of rewards.

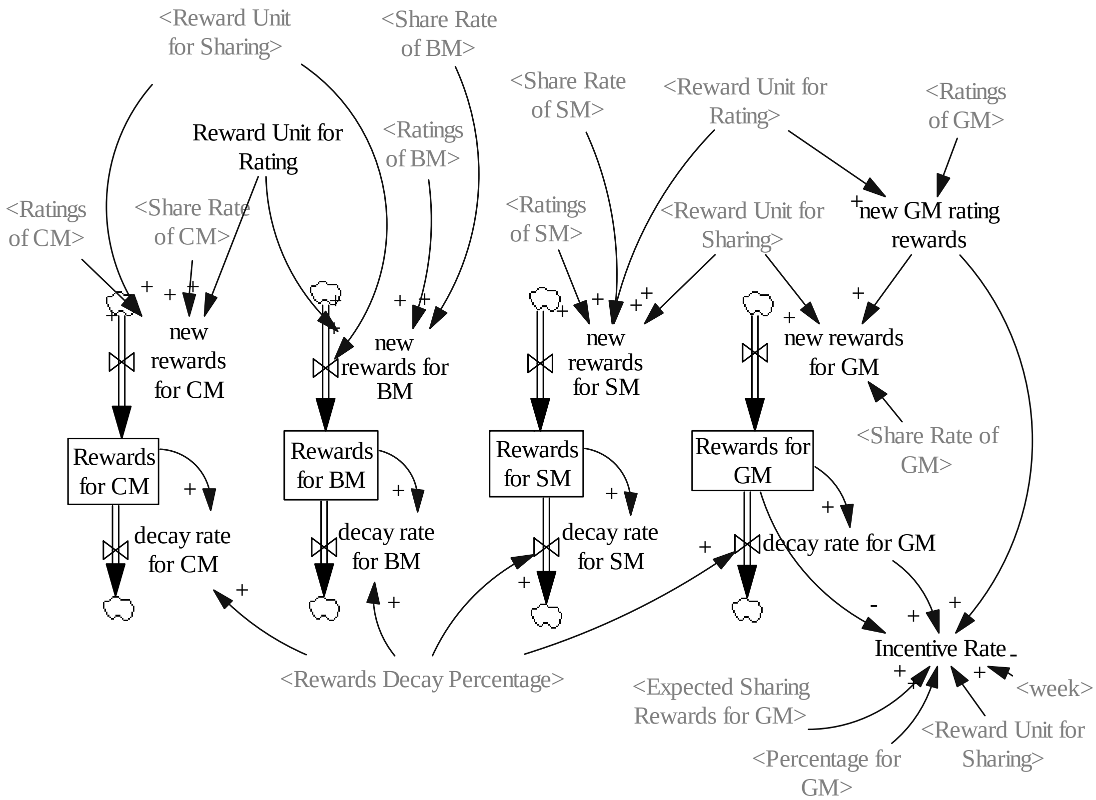


Figure 3.8. Reward sector

Take the flow rates of "*Rewards for BM*" as an example, the rewards has the unit of Point and the differential equation is defined as follows:

$$\frac{d}{dt}(Rewards\ \ for\ BM) = new\,rewards\ \ for\ BM\ \text{-}\ decay\,rate\ \ for\ BM \tag{3.16}$$

where:

$$\begin{aligned} &new\ \ rewards\ \ for\ BM\ = \\ &\qquad Reward\,Unit\ \ for\ Sharing * Share\,Rate\,of\ \ BM \\ &\qquad + Reward\,Unit\ \ for\ Rating * Ratings\,of\ \ BM \end{aligned} \tag{3.17}$$

As part of the extended version of the incentive mechanism, the value of rewards also decreases over time, which motivates users in the community to continue participating in order to reach their desired number of points or keep their membership status. Variable "*Rewards Decay Percentage*" has constant value with unit of 1/Week, and is used to control how much rewards are lost per time unit. The outflow rate of "*Rewards for BM*" is calculated as:

$$\begin{aligned} &decay\ rate\ \ for\ BM\ = \\ &\qquad Awards\ \ for\ BM * Rewards\ Decay\ Percentage \end{aligned} \tag{3.18}$$

3.2.2.3 C-point sector

The 'C-point sector' (Figure 3.9) is more complicated than the 'Reward sector', since the C-points also change according to the variation of the demographic structure of the population, not only decaying over time. For example, if the membership level of a particular participant is silver in the current week, but decreases to bronze membership in the following week, then the number of C-points gained by this participant will be "transferred" between two membership groups, silver and bronze, that is, from "*Cpoints of SM*" to "*Cpoints of BM*". In this sector, the number of C-points transferred from the silver member group to the bronze member group is modeled as a co-flow called "*coflow SB*". Similarly, the number of C-points transferred from the bronze member group to the silver member group is modeled as a co-flow called "*coflow BS*".

Take "*Cpoints of BM*" as an example. Participants can be rewarded a number of C-points only by rating resources in the community, though at the very beginning each participant will get a fixed amount of C-points. So the initial value of "*Cpoints of BM*" will be the product of the initial value of C-points assigned to each participant and the size of the bronze membership group.

The equation of the change rates is illustrated as:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{dt}}(\mathrm{Cpoints\ of\ \ BM}) &= \mathrm{earnings\ of\ \ BM + coflowSB + coflow\ CB} \\ &\qquad \mathrm{-\ costs\ of\ \ BM\ -\ coflow\ BS\ -\ coflow\ BC} \end{aligned} \tag{3.19}$$

For the inflows of stock "*Cpoints of BM*", there are three sources of C-points: C-points rewarded by rating resources (flow "*earnings of BM*"), C-points transferred from common members due to the membership upgrade (flow "*coflow CB*"), and the C-points transferred from silver members due to the membership decay (flow "*coflowSB*").

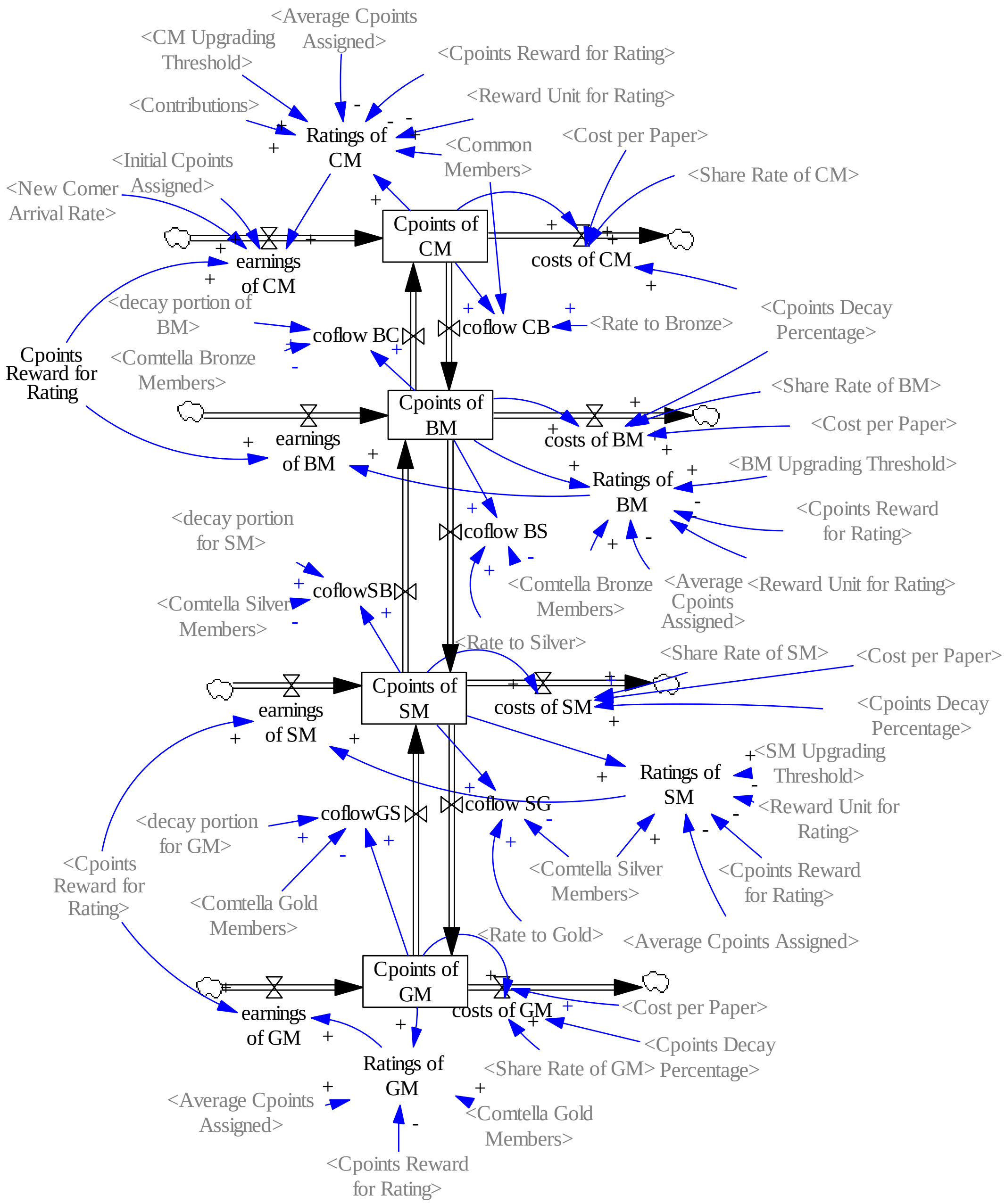


Figure 3.9. C-point sector

The part of C-points rewarded by rating resources is represented by the flow "*earnings of BM*", and is calculated by multiplying the number of ratings from bronze members (reflected by the variable "*Ratings of BM*") by the C-points reward unit for each rating (reflected by the variable "*Cpoints Reward for Rating*").

For the part of C-points transferred from common members due to the membership upgrade (represented by flow "*coflow CB*"), the value equals to the number of common members who upgrade their membership level to bronze membership level (reflected by the variable "*Rate to Bronze*") times the average number of C-points which each common member has.

Similarly, the part of C-points transferred from silver members due to the membership decay (represented by flow "*coflow SB*") is measured by multiplying the number of silver members whose membership level becomes the bronze membership level (reflected by the variable "*decay portion for SM*") by the average number of C-points which each silver member has.

Finally, the sum of the three inflows is calculated by:

$$\begin{aligned} &earnings\ of\ BM + coflowSB + coflow\ CB = \\ &\qquad Ratings\ of\ BM * Cpoints\ Reward\ for\ Rating \\ &\qquad + \frac{Rate\ to\ Bronze}{Common\ Members} * Cpoints\ of\ CM \\ &\qquad + \frac{decay\ portion\ for\ SM}{Comtella\ Silver\ Members} * Cpoints\ of\ SM \end{aligned} \tag{3.20}$$

For the outflow rate "*costs of BM*", there are four ways of losing C-points: C-points spent for promotion of new contributions, C-points expired after a particular period of time due to the decay (flow "*costs of BM*" calculates the sum of the first two parts), C-points transferred to common members due to the membership decay (flow "*coflow BC*"), as well as the C-points transferred to silver members due to the membership upgrade (flow "*coflow BS*"):

$$
\begin{aligned}
costs\ of\ BM + coflow\ BS + coflow\ BC = \\
& Cost\ per\ Paper*Share\ Rate\ of\ BM \\
& +Cpoints\ Decay\ Percentage*Cpoints\ of\ BM \\
& +\frac{decay\ portion\ of\ BM}{Comtella\ Bronze\ Members}*Cpoints\ of\ BM \\
& +\frac{Rate\ to\ Silver}{Comtella\ Bronze\ Members}*Cpoints\ of\ BM
\end{aligned} \tag{3.21}
$$

3.3 Results

The simulations are based on the assumption that users in the community are mainly motivated by activity points. The reason is that the current incentive mechanism in Comtella uses activity points as the only measure of user memberships because it provides a simple notion of competition.

First, I discuss the results of the simulation of the first version model. In order to test the simulation model and the behaviours it can produce, the parameters will be tested, and the results will also be compared with historical data of the real Comtella system deployment in the database to see whether the simulation model truly measures the behaviours of the actual Comtella community and captures the effect of incentive mechanism on the demographic structure of Comtella population. In the experiment, the impacts of membership upgrading thresholds on different user groups and total number of contributions are further investigated. Details of the Comtella database are provided in Appendix B.

3.3.1 Base run

Once the model is developed, it is possible to experiment with different parameters in order to analyze different scenarios. However, it is useful to have a base run first to validate the model behaviour.

3.3.1.1 First version model

For the early version of Comtella community, the real data was collected during one academic term (16 weeks) experiment with 32 fourth-year students of the Department of Computer Science while taking a course on Ethics and Information Technology in 2003-2004 winter sessions.

Similar to the real experiment, the length of the simulation is 16-weeks, and the parameters for the first version model are listed in Table 3.1. The values of the system parameters are obtained from the real data, and the custom parameters are determined by testing and analysis.

Table 3.1. Parameters for basic system dynamics model

System Parameters		
Name	**Unit**	**Value**
Initial Population	Person	32
Weight of Sharing	Point/Link	4
Common Member Upgrading Threshold	Point/Person	24
Bronze Member Upgrading Threshold	Point/Person	32
Silver Member Upgrading Threshold	Point/Person	40
Custom Parameters		
Name	**Unit**	**Value**
Noise Seed	Dimensionless	0.5
Sensitivity parameter for common members	Dimensionless	0.175

Sensitivity parameter for bronze members	Dimensionless	0.575
Sensitivity parameter for silver members	Dimensionless	0.29
Expected Share Rate for Gold Members	Link/Person/Week	3.5

Figure 3.10 presents the simulation results of the first version model.

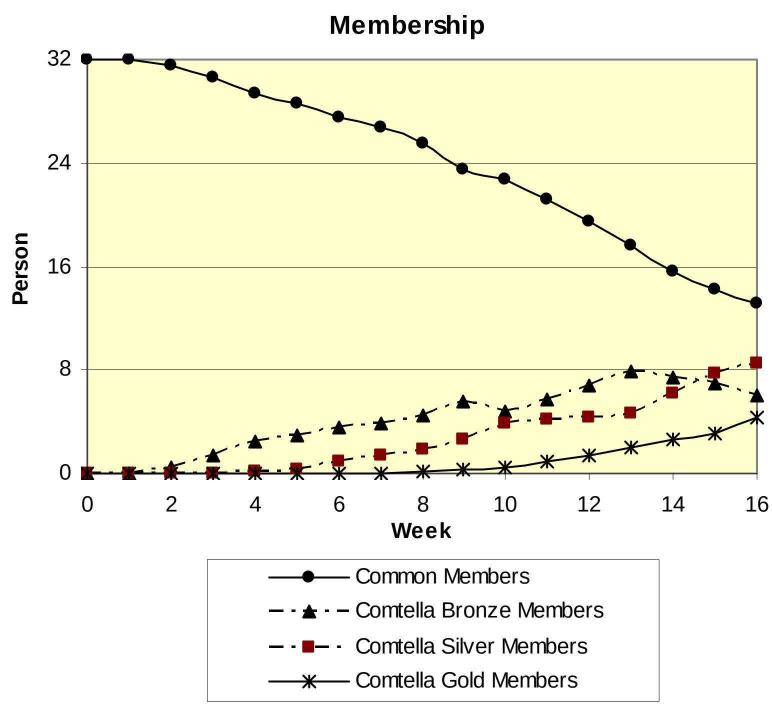


Figure 3.10. Base run results (first version model)

It can be seen from the figure that at the very beginning everyone is a common member, and as time goes on, their membership levels upgrade, and the population of common member group decreases to 40 percent of the whole population at last. After 9 weeks the population of gold members and silver members increase. At the same time the population of bronze members decreases slightly in the end as a result of the balancing effects among the growing rates of these user groups. After 16 weeks, the population of bronze, silver and gold member increases to 18.7, 25 and 12.5 percent respectively.

In order to validate the model, the simulation results are compared with the real data, which was collected during one academic term experiment.

Figure 3.11 shows a fairly good fit between the real data and the data generated by the simulation model. There is a slight deviation of the change rate of gold members, which is probably caused by the variety of the student behaviours.

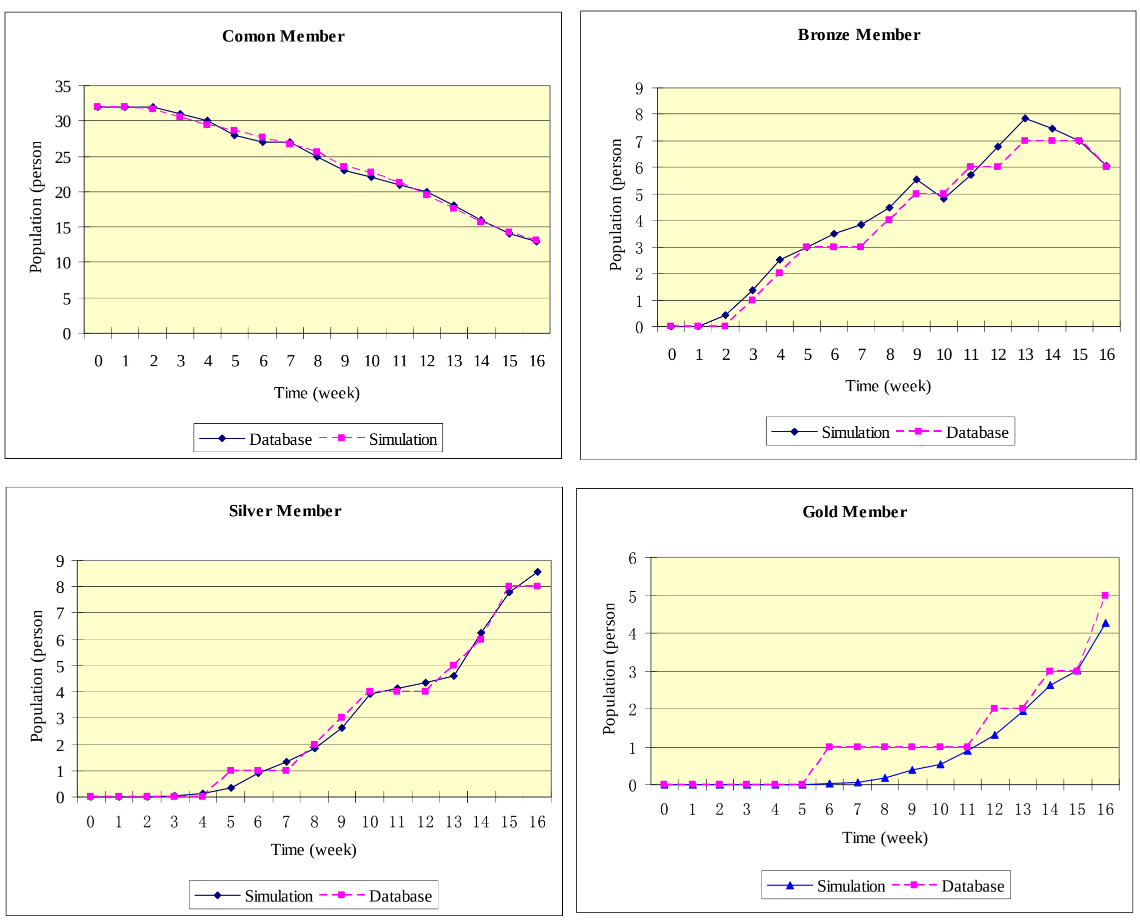


Figure 3.11. Validity test results (first version model)

3.3.1.2 Second version model

For the latest version of Comtella community, the real data was collected during one academic term experiment with two groups of fourth-year students of the Department of Computer Science while taking a course on Ethics and Information Technology in 2004-2005 winter sessions. Each group has 17 participants. The experiment was carried out for 11 weeks and participants are divided into two random-selected groups: the test group and the control group.

The test group used the latest version of Comtella community (with C-points, rating behaviors, membership decay and adaptive reward unit), and the control group used a similar-looking system, but with fixed reward unit and no C-points. Two separate systems with some different features were used to serve these two participant groups (Table 3.2).

Table 3.2. Differences between the two systems for the two groups (Copied from Cheng 2005 with permission)

Feature	System for Test Group	System for Control Group
Hierarchical Memberships	✓	✓
Showing Contribution Levels in Previous and Current Week	✓	✓
Interface for Rating Articles	✓	✓

C-points as Reward for Rating	✓	×
Adaptive Weights for Sharing and Rating	✓	×
Personalized Messages	✓	×

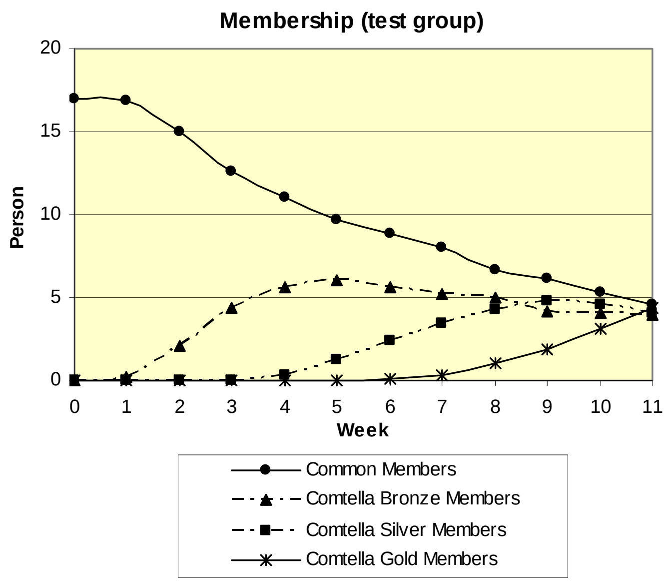
In order to do policy validations by comparing the simulation results with the real data through the base run, two sets of parameters and the units of the parameters are listed in Table 3.3 for the test group and the control group respectively. Here, the values of the system parameters are obtained from the real Comtella system, and the custom parameters are determined by analysis and testing. A full list of parameters is presented in Appendix D.

Table 3.3. Parameters for the extended system dynamics model

System Parameters			
Parameters	**Value (Test Group)**	**Value (Control Group)**	**Unit**
Initial Population	17	17	Person
Reward Unit for Sharing	4	4	Point/Link
Reward Unit for Rating	3	3	Point/Link
C-point rewards for rating	3	0	Point/Link
Initial Cpoints Assigned	20	0	Point/Person

Rewards Decay Percentage	0.7	0.7	1/Week
C-points Decay Percentage	0.5	0	1/Week
Possibility to assign C-points	0.458	0	Person*Week/Link
Average C-points assigned	5.618	5.618	Point/Person/Week
Common Member Upgrading Threshold	24	24	Point/Person
Bronze Member Upgrading Threshold	32	32	Point/Person
Silver Member Upgrading Threshold	40	40	Point/Person
Custom Parameters			
Sensitivity for Common Member	0.13	0.19	1/Week
Sensitivity for Bronze Member	0.16	0.19	1/Week
Sensitivity for Silver Member	0.21	0.14	1/Week
Decay rate of Bronze Member	0.25	0.15	1/Week
Decay rate of Silver Member	0.075	0.05	1/Week
Decay rate of Gold Member	0.125	0.125	1/Week

Figure 3.12 shows the simulation results for these two separate systems. Similar to the first version model, all participants have the lowest membership levels at the very beginning, and gradually upgrade their membership levels over time. The total population is 17 throughout the experiment.



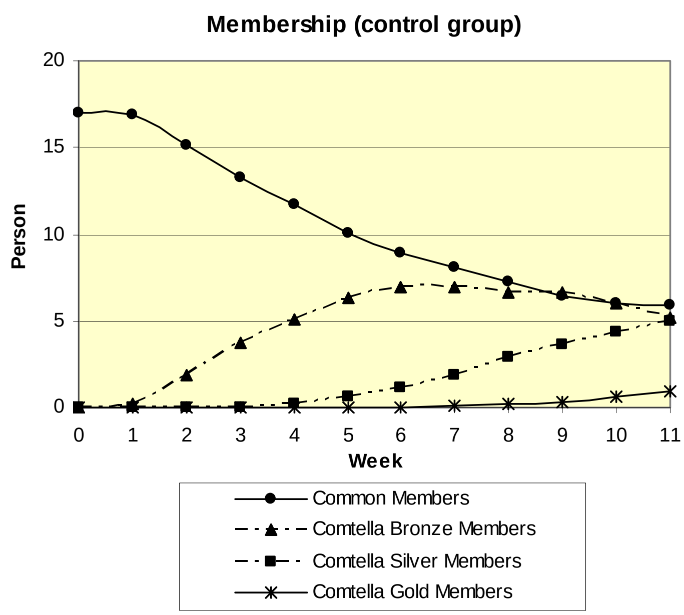


Figure 3.12. Simulation results (second version model)

For the test group, the population of common members decreases to 29 percent of the whole population at last. After 5 weeks the population of bronze members reaches its peak value, and decreases slightly afterwards while the silver member population starts to increase. At last the population of bronze, silver and gold member increases to 17.6, 29.4 and 23.5 percent respectively. For the control group, the population of common members decreases to 35 percent of the whole population at last, a little bit higher compared to the test group. The bronze member population reaches its peak value after 6 weeks, while silver and gold member populations increase quite slowly to 29.4 and 5.9 percent respectively in the last week. The results suggest that C-points and individual adaptive reward units can motivate participants in Comtella online community to some extent.

To validate the model, similarly to the first version model, the simulation results are compared with the real data. Curves for output variables with trend lines are shown in Figure 3.13. Compared to the Comtella database, the simulation results for common members, total number of contributions and ratings fit quite well with the real data. Besides that, the end values of outputs shown in the diagram have very good fits with the real data. However, the extended system dynamics model is not able to generate curves for bronze, silver or gold members that fit well with the real data. In the database it can be seen that the populations of bronze, silver and gold member abruptly rise after 5 weeks. The percentage of bronze member jumps from 0% to 41.2% suddenly in the 5th week, and similarly the percentage of silver member increases 29.4% in the 6th week.

There are several reasons for the deviation. First, the database lacks details to separate the data of 2005 reading week break from the whole set of data. As a result, data of the 6th week is collected from three weeks, February 7, 2005 to February 28, 2005. Second, my system dynamics models encountered difficulties to model anomalies. Ideally the differential equations in the system dynamics models can generate all kind of behaviors, and a perfect fit can be made if the number of variables in the model is large enough. However in my case the level of the simulation is carefully controlled and the structures of the system dynamics models are fixed before the simulation starts, without enough capabilities to capture the anomalies in the system.

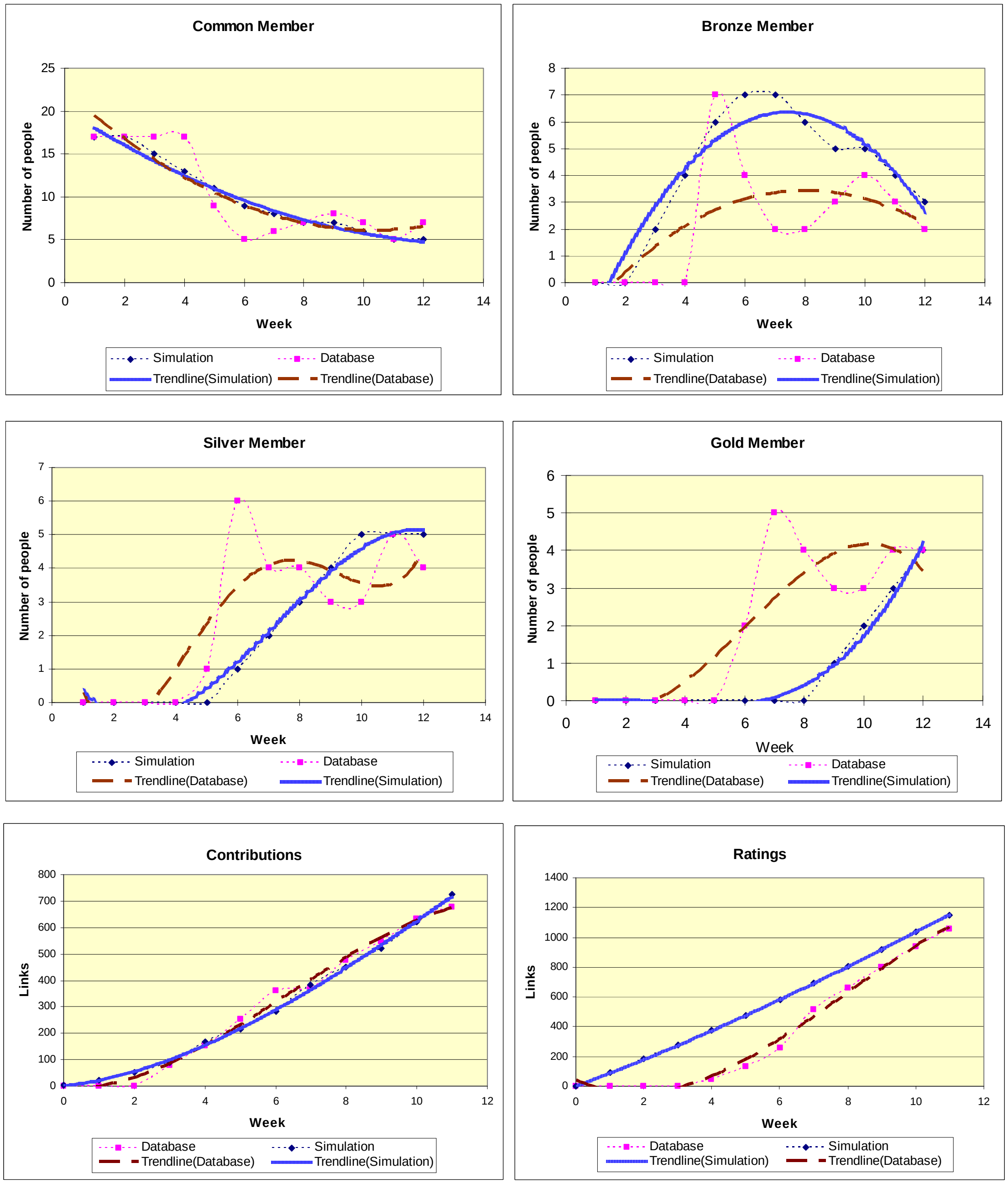


Figure 3.13. Validity test results (second version model)

For the real Comtella community, there might be several factors that result in sudden increases or decreases, such as the number of related web-resources online that can be shared, the popularity of the topic that was specified for a particular week, external factors, such as course work, etc. For example, popular topics in specific weeks will definitely result in higher user participation. These characteristics of the Comtella community as well as other small-scale educational online communities with fixed number of participants might result in rapidly emerging changes.

Moreover, the participants may be more motivated when there is membership decay, but we do not have enough details to model it. As a result, using the membership to measure user participation might not be good enough because the effectiveness of the incentive mechanism might be under-estimated.

3.3.2 Sensitivity tests

After developing the model, it is always important to know how sensitive the model is to different assumptions. Vensim has a sensitivity capability that makes it easy to run Monte-Carlo sensitivity simulations.

Monte Carlo simulation is used to explore the future possibilities and uncertainty of the selected output variables through hundreds or thousands of repeated simulations, representing unknowns as a pool of possible values from which values are drawn at random. Given the uncertainty of model parameters and stochastic control, confidence bounds are used to demonstrate the validity of the model.

Table 3.4 represents the range of values of several input variables that need to change in order to do the validity tests.

Table 3.4. The range of values of input variables (first version model)

Parameters	Value	Range
Common Member Upgrading Threshold	24	[18, 32]
Bronze Member Upgrading Threshold	32	[24, 40]
Silver Member Upgrading Threshold	40	[32, 48]
Sensitivity for Common Member	0.575	[0.4, 0.65]
Sensitivity for Bronze Member	0.175	[0.15, 0.2]
Sensitivity for Silver Member	0.29	[0.25, 0.32]

Through the Monte Carlo simulation for the first version model, the simulation-based confidence bounds (or the uncertainty) of Comtella memberships are found, as shown in Figure 3.14. The diagram shows the 50%, 75%, 95% and 100% confidence bounds estimated from a collection of 1000 simulations using the advanced version of the Vensim software, and a percentage of test cases in the Monte Carlo simulation falls in the confidence bounds with a particular percentage. As an example, 100% of the test cases locate in the 100% confidence bounds.

It can be seen that 50% of the values in each separate diagram lie inside the yellow area with the median in the middle which is quite close to the real value according to Figure 3.11 (page 50).

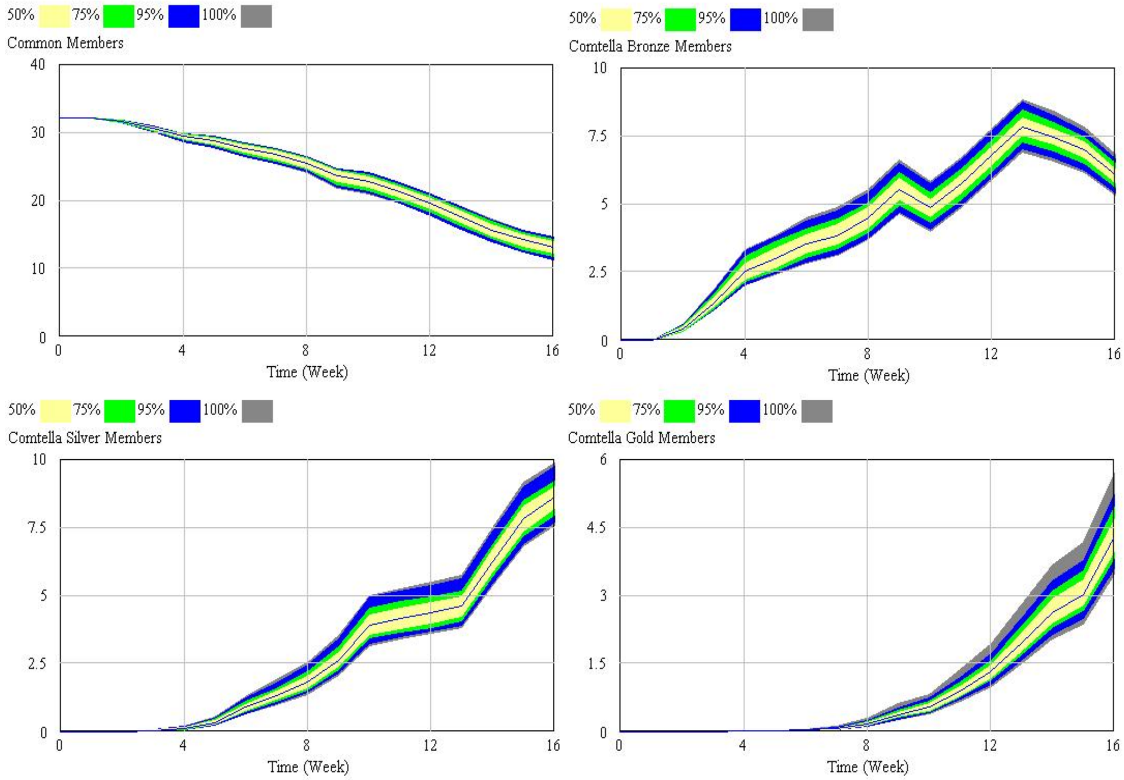


Figure 3.14. Sensitivity test results (first version model)

Similarly, Table 3.5 presents the range of values of several input variables that need to change in order to do the validity tests for the second version model. Results are shown in Figure 3.15.

Table 3.5. The range of values of input variables (second version model)

Parameters	Value	Range
Common Member Upgrading Threshold	24	[18, 32]
Bronze Member Upgrading Threshold	32	[24, 40]
Silver Member Upgrading Threshold	40	[32, 48]

Sensitivity for Common Member	0.13	[0.10, 0.15]
Sensitivity for Bronze Member	0.16	[0.12, 0.2]
Sensitivity for Silver Member	0.21	[0.15, 0.25]

Figure 3.15 shows the 50%, 75%, 95% and 100% confidence bounds of Comtella memberships estimated from a collection of 1000 simulations through the Monte Carlo simulation for second version model.

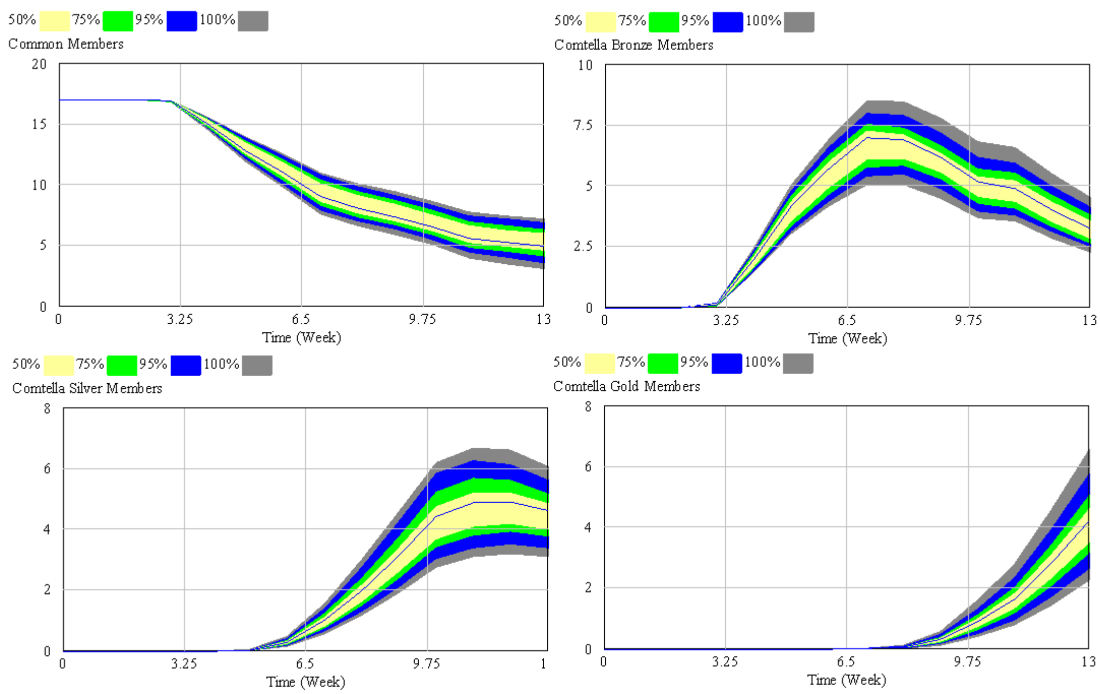


Figure 3.15. Sensitivity test results (second version model)

In the diagram there are narrow band of uncertainty at the start, and the bands grow larger and larger over time due to the positive feedback in the motivation process. On the other hand, the confidence bounds of Comtella bronze and silver members are no longer growing over time at last. After the 7^{th} week, the 50% confidence bound of Comtella bronze members starts to shrink over time and ranges from 3 to 4 persons in the end.

Although the 50% confidence bounds are a little bit wider than Figure 3.14 (page 59), they are still good enough with the median in the middle. Three-quarters of the values in each separate diagram lie inside the green area with a small range of only 1 or 2 persons. Thus, the system dynamics models are good enough to gain insights into the user motivation problem with tolerant error rate of numerical prediction.

3.3.3 Experiments

For the earlier version of the incentive mechanism, an experiment is conducted to analyze the impact of the three thresholds to the demographic structure of the Comtella population. I want to find the factors that can increase the number of middle level participants and control the number of gold members and common members. The reason for keeping these two populations small has been mentioned in Section 2.3. The free-riders in the common member group should also be reduced to a certain level, but on the other hand users should not be able to upgrade their memberships too fast, and the gold membership should be considered as a small exclusive club to attract users.

Table 3.6 presents the variation of different user group populations and the total number of contributions by changing the three thresholds respectively, where the membership upgrading threshold for common, bronze and silver members are called "*threshold1*", "*threshold2*" and "*threshold3*" for short. Besides that, the number of bronze and silver members is called "*B and S members*" for short. The unit of variation is 1 reward unit.

In order to decrease the percentage of free-riders, the size of the common member group should be small. However on the other hand, we have to prevent the population of gold members from increasing too fast and maintain both a large percentage of the middle level participants (that is, the bronze and silver members) and a large number of contributions, so there are tradeoffs among our three goals: keeping the size of common member small, keeping the large percentage of bronze and silver members high, and keeping the total number of contributions large.

Table 3.6. Impact of variation by changing different thresholds (first version model)

	Common Members	Bronze Members	Silver Members	Gold Members	B and S Members	Contributions
Increase threshold1	+5.99%	-1.84%	-4.45%	-6.82%	-3.37%	+5.38%
Decrease threshold1	-6.56%	+1.61%	+4.89%	+8.03%	+3.52%	-5.7%
Increase threshold2	-	+10.26%	-3.41%	-6.61%	+0.96%	+1.1%
Decrease threshold2	-	+5.20%	-1.70%	-3.41%	-1.03%	-1.1%
Increase threshold3	-	-	+3.33%	-6.34%	+0.92%	+0.82%
Decrease threshold3	-	-	+1.72%	-3.27%	-0.93%	-0.87%

In the table, all the values are calculated at the end of the simulation. From the results it can be seen that the common membership upgrading threshold (threshold1) has the greatest impact on the population of the middle level participants. To balance the three goals, the threshold for common members should slightly decrease in order to encourage the common members to share more resources, and the thresholds for bronze and silver members should increase to stimulate middle level participants.

Now let us move on to the experiments for the second version model.

For the extended version of the incentive mechanism, since the total number of contributions and the percentage of the bronze and silver members are the foremost output values to evaluate the system, the following experiments are launched for the test group to test system parameters that are important for evaluating and improving the total number of contributions. Quality control is also important, but currently there are not enough details to model it.

In total there are five experiments, and they test the impacts of the length of experiment, the population size, the reward decay rate, the reward unit for sharing, and the membership thresholds respectively for the test group. The values of these system parameters are changed, and the diagrams to compare the results are provided.

1. Change the length of experiment

This experiment focuses on the impact of the length of experiment on the total number of contributions and ratings, as well as the percentage of the middle level participants. Figure 3.16 shows the experiment results for a 16-week experiment. The total number of contributions and ratings still increases gradually after the 11th week. On the other hand, the percentage of bronze and silver members reaches the peak value in the 9th week, and decreases by 18% in 3 weeks after that. In the last 4 weeks the percentage of bronze and silver members only decreases by 5.4%, which means only 1 out of 17 participants.

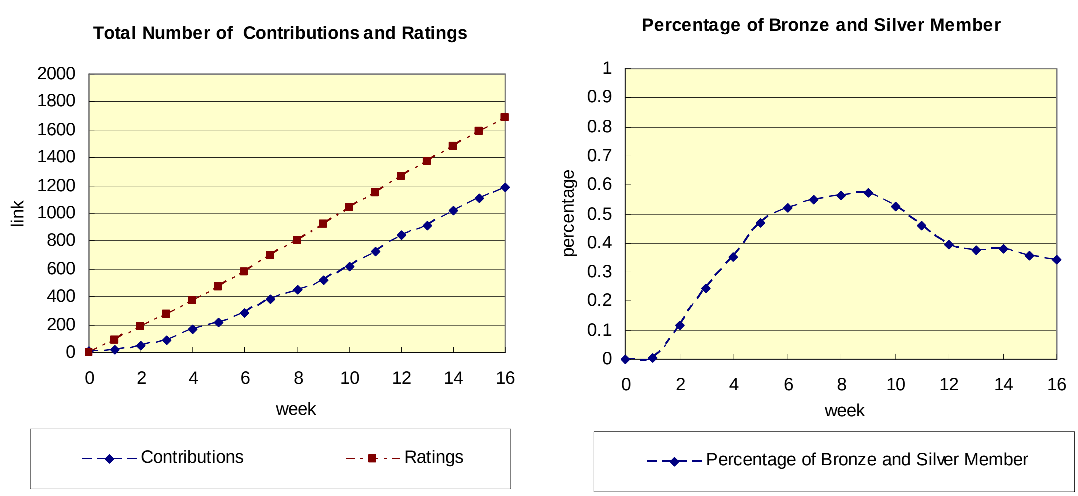


Figure 3.16. Experiment on the length of experiment (SD model)

2. Change the population size

Experiment results on changing the population size are shown in Figure 3.17. Variable "*Initial population*" is set to be 17, 32, and 50 respectively.

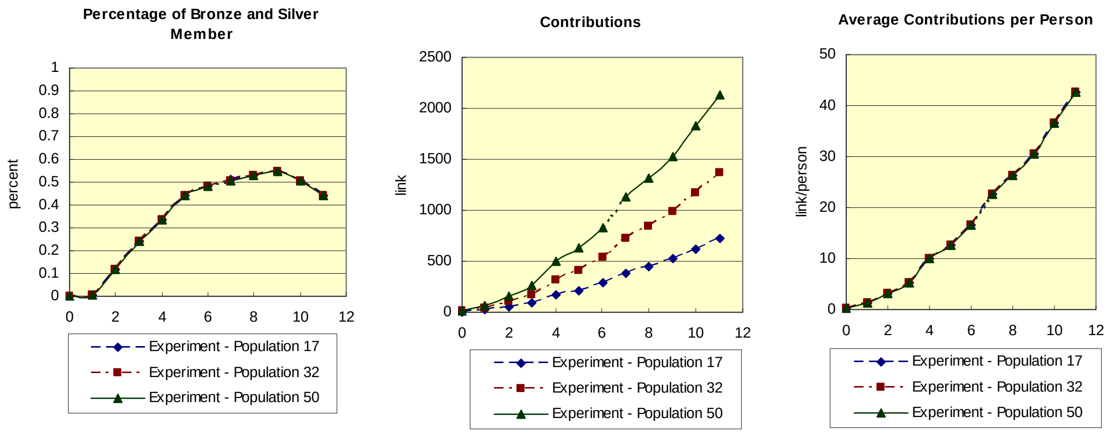


Figure 3.17. Experiment on the population (SD model)

It is shown in the diagrams that the total number of contributions is proportional to the number of participants in the community because the average numbers of contributions per person are almost the same. The percentage of bronze and silver members also stays almost the same over time. This is caused by the assumption of this system dynamics model (as mentioned in Section 3.1.2.2, page 33): all the entities in the same stock have the same behaviour. In other words, all the members in the same member group have the same share rate and weekly ratings. As a result, this system dynamics model is not good enough to model the influence of population on the user participation.

3. Change the reward unit for sharing

Figure 3.18 demonstrates the effects of reward unit for sharing on contributions and percentage of bronze and silver members. It shows that the reward unit for sharing has obvious incentive effects on contributions and the percentage of bronze and silver members.

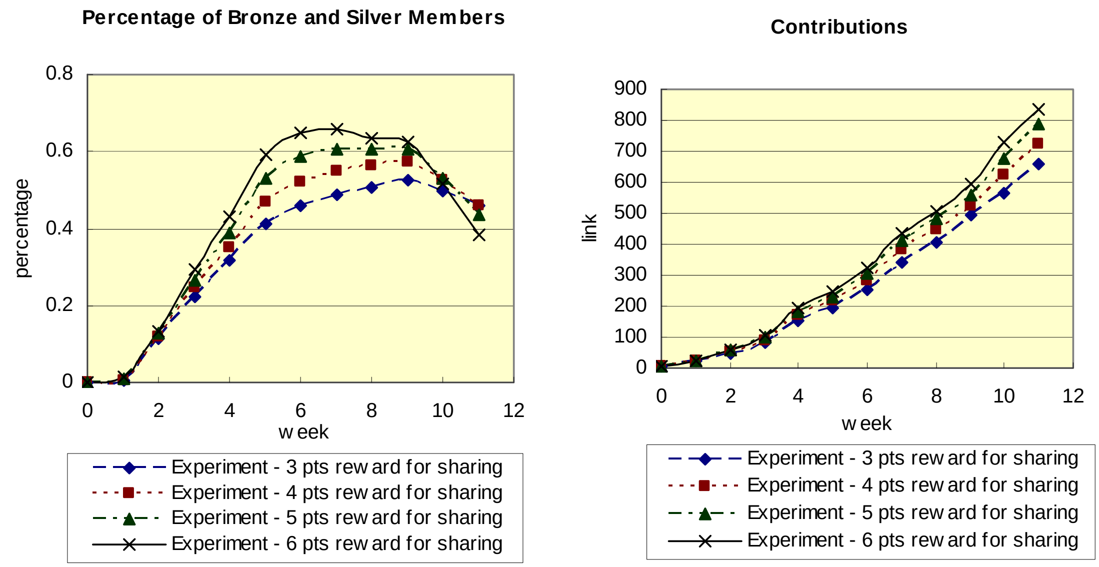


Figure 3.18. Experiment on the reward unit for sharing (SD model)

At the beginning, bronze and silver member groups comprise a higher percentage of the whole population when the reward unit for sharing is higher. However after 9 weeks, the percentage of middle level participants decreases when the reward unit for sharing increases. On the other hand, the overall contributions have a slightly increased rate when the reward unit for sharing is higher. As a result, the reward unit for sharing can be considered as an important factor in the community. It might be beneficial to introduce a higher reward unit for sharing at the beginning, and adjust it after a particular period according to the length of the experiment in the real world.

4. Change the decay rate

From Figure 3.19 it can be seen that within the first 9 weeks the decay rate for rewards has large effects on the percentage of bronze and silver members, but little effect on total number of contributions.

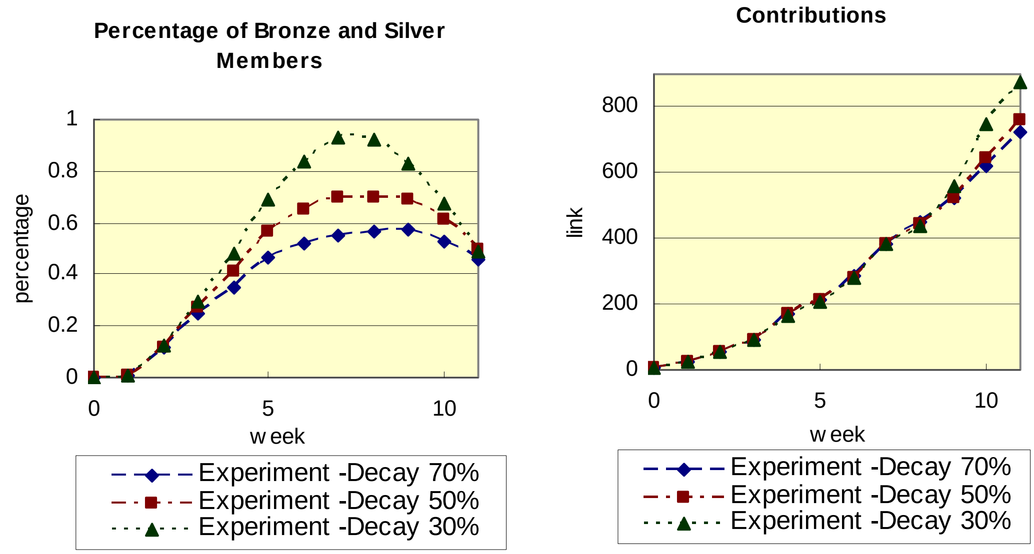


Figure 3.19. Experiment on the decay rate on rewards (SD model)

There will be more contributions shared in the community when the decay rate is lower. On the other hand, bronze and silver members constitute a larger percentage of the whole population when the decay rate is lower throughout the experiment. If the length of the experiment is longer, the percentage of the bronze and silver members will decrease faster when the decay rate is lower. Consequently, a lower value of decay rate such as 30% is not stable enough for the community in a long run.

From the experiment we can see that an appropriate value of decay rate is quite important to the community. If the decay rate is too high, participants will gradually lose their interest in the community because they lose too many activity points each week and it is too hard to reach their membership level upgrading goals. On the other side, when the decay rate is too low, participants will gradually reduce their participation in the community since they can get activity points and reach their goals easily. In the current system the value of the decay rate equals 0.7, which maintains a good percentage of bronze and silver members in the long run. For the Comtella community, the 50%

decay rate might be more efficient to motivate participants and optimize the system within one academic term, that is, 16 weeks.

5. Change the membership thresholds

Table 3.7 presents the impacts of the three membership thresholds on the contributions, the demographic structure of the Comtella population, and the percentage of the middle level participants. Similar to the experiments for the first version model, the membership upgrading threshold for common, bronze and silver members are called "*threshold1*", "*threshold2*" and "*threshold3*" for short. The unit of variation is 1 reward unit for sharing, and the number of bronze and silver members is called "*B and S members*" for short.

Table 3.7. Impact of variation by changing different thresholds (second version model)

	Common Members	Bronze Members	Silver Members	Gold Members	B and S Members	Contributions
Increase threshold1	+2.37%	-2.07%	-1.28%	-0.16%	-1.61%	+5.71%
Decrease threshold1	-3.34%	+2.25%	+1.50%	+0.51%	+1.81%	-5.84%
Increase threshold2	+0.96%	+3.51%	-1.94%	-1.69%	+0.31%	+2.18%
Decrease threshold2	-1.44%	+3.89%	-2.09%	-2.36%	-0.37%	-2.16%
Increase threshold3	+0.05%	+0.18%	+2.16%	-2.54%	+1.35%	+0.77%
Decrease threshold3	-0.06%	-0.25%	+2.49%	-2.96%	-1.56%	+58.7%

As mentioned before, we need to control the contributions, the percentage of free-riders, and the percentage of the middle level participants at the same time. The results show that when the threshold for common members decreases, there will be fewer common members and more middle level participants. Besides that, the thresholds for bronze and silver members should slightly increase in order to stimulate middle level participants and keep a large number of contributions.

Like in the experiments for the first version model, all the values in Table 3.7 are calculated at the end of the simulation. The results show the same positive or negative impacts as the experiments on the first version model in Table 3.6, so similarly in order to balance the three goals, the threshold for common members should slightly decrease, and the thresholds for bronze and silver members should increase to stimulate middle level participants.

There is also a need to further calibrate the model and investigate the optimal value of the thresholds that can motivate users in the community, which is one aspect of our future work.

3.4 Summary

This chapter described in detail a prototypical system dynamics model and an extended simulation model for a variation of the incentive mechanism of the Comtella community. The system dynamics model is presented, including the design of the model structure and first simulation results.

- I designed and implemented two different system dynamics models for two different versions of Comtella community.
- I did elementary analysis of the system dynamics simulation results. Although it is impossible to build a model that exactly reproduces the real world due to the complexity of human and social dynamics, the results from the system dynamics model are able to demonstrate the dynamics of user motivation and incentive mechanism in Comtella community to some extent.
- Experiments are launched to study the effects of several system parameters

such as the reward unit for sharing and rating, the decay rate, the population, as well as the length of the experiments. Results from the experiments provide us with further information about the factors that can motivate participants in different periods, which is quite helpful for the evaluation and improvement of incentive mechanisms in Comtella community.

CHAPTER 4

AGENT-BASED MODEL OF THE COMTELLA INCENTIVE MECHANISM

The previous chapter presented two prototypical models using a system dynamics approach to simulate the incentive process in the Comtella community. In this section, the architecture of an agent-based model for the extended version of the incentive mechanism of Comtella community is proposed. The model is developed by using the software AnyLogic.

4.1 Agent-based modeling

As mentioned in Chapter 2, the system dynamics model is well suited to study systems at a high level of aggregation; however, the agent-based model is well suited to study systems at the individual level (micro-level). In environments like online communities where there are a lot of interactions among participants and the environment itself, agent technology provides a powerful way to capture the complexity of this process. To allow comparison, the same online community is modeled using the agent-based approach and simulated in AnyLogic.

Figure 4.1 presents the way of mapping the real world into agent-based models. Compared to system dynamics modeling, modeling using the agent-based approach is decentralized. The system is modeled as a collection of autonomous agents, and there is no global system behaviour defined in the model, but only the individual behaviours of its agents. For this reason, the focus is on the individual behaviours of different types of users (or user groups) in the Comtella community. The goal is to provide insights into the user motivation process, incentive mechanism evaluation and community development.

For my study, the focuses are the user motivation process of different user groups, the individual membership levels of users, and the impact of their activities on the whole system.

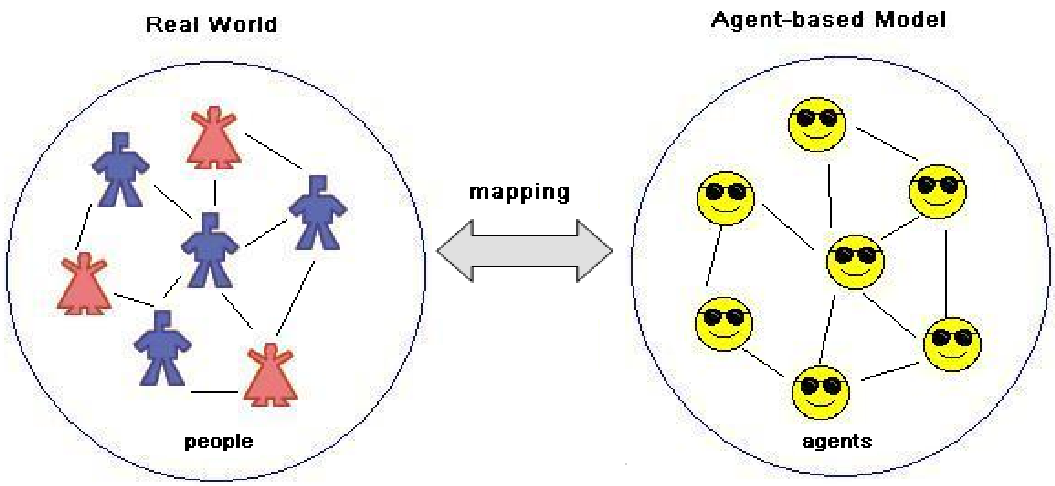


Figure 4.1. Mapping the real world in agent-based models

4.2 Implementation of the extended version of the incentive mechanism

As mentioned before in Section 2.1.2, Cheng implemented the extended version of the Comtella incentive mechanism. The system was used for a course on Ethics and Information Technology in 2004-2005 winter sessions, in the Department of Computer Science, University of Saskatchewan. Next, I will give some more details of Cheng's work (2005) on the extended version of the incentive mechanism, which was used in the design of my agent-based model.

In Cheng's design, the two individual adaptive reward units W_S and W_R represent the reward unit for sharing a new resource and giving a rating respectively. The individual reputation of the participant is measured by “activity points”, which determine the membership level of the participant and is calculated as the sum of the following four factors: the quantity of the contributions (PaperQuanCr), the quality of the contributions (PaperQualCr), the quantity of ratings (RatingQuanCr), and the quality of ratings given by the participant (RatingQualCr).

- The quantity of the contributions (PaperQuanCr)

The quantity of the contributions are calculated by the product of individual adaptive reward units for sharing new resources (W_S) and the total number of contributions shared by specific participant. The variation of the reward unit is W_S shown in Figure 2.4 (page 12). The reward unit for sharing W_S has a constant part W_{S0}. However, there are other factors that cause W_S to vary over time, such as the paper reputations, and the desired number of weekly contributions for each participant, etc.

Cheng (2005) considered those resources contributed early in each week to be more useful to the online community. After there is a mass of contributions, new contributions are less useful since they are less visible in the search result list and participants seldom click them. Consequently, a higher value of W_S should be applied to early-shared resources. Cheng used T to represent this factor, which is a function of time and also called the “time-function factor” (Figure 4.2).

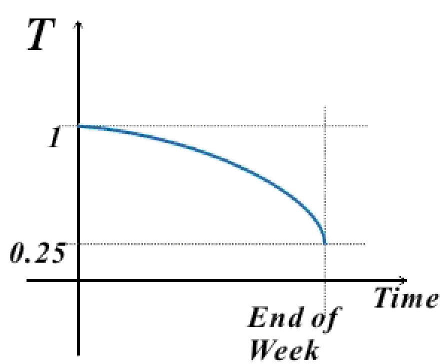


Figure 4.2. The changing of the reward factor T
(Copied from Cheng 2005 with permission)

A higher value will be assigned to the time-function factor T for the early-shared resources, because there is a strong demand for resources at that time. The time-function factor T is a declining function with a specific shape defined by Cheng (2005) and has a range of [0.25, 1]. Since the calculations of this function and the corresponding factor in the agent-based simulation are quite lengthy and they would distract from the main flow of this section, the details of the formulas are shown in Appendix E.

In addition, Cheng (2005) also stated that the number of contributions shared by the participant should depend on the paper reputation, which is determined by the ratings from other participants given to the resources shared by this participant. For this reason, each participant has a desired number of weekly contributions in Comtella. When the actual number of weekly contributions of the participant exceeds the desired number of weekly contributions, the participant will get little reward for those over-limit contributions, which generally discourages participants in the community from contributing resources beyond the limit.

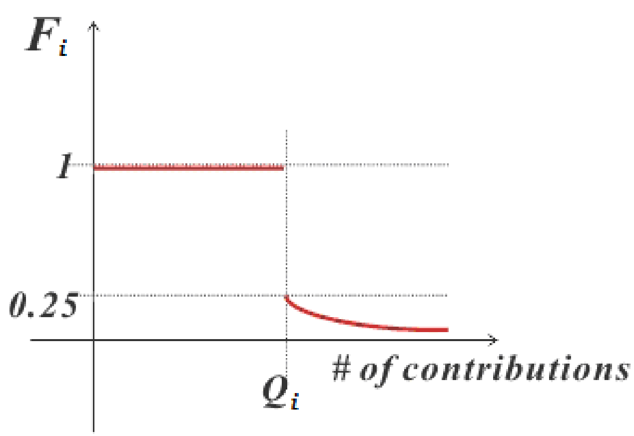


Figure 4.3. The changing of the reward factor F_I
(Copied from Cheng 2005 with permission)

Thus, the variable F_i is applied to represent this factor, which is called the "over-limit factor" and has a range of (0, 1]. The subscript i is the index of the current participant. Cheng (2005) presents the change of the over-limit factor F_i in Figure 4.3, where the variable Q_i represents the desired number of weekly contributions for the participant i. The calculation of F_i is shown as:

$$F_i = \begin{cases} 1 & when\ x \le Q_i \\ \left(\frac{1}{4}\right)^{(x-Q_i)} & when\ x > Q_i \end{cases} \tag{4.1}$$

Here, the variable x represents the weekly contributions of the participant, and the desired number Q_i depends on the paper reputation of the participant. Generally participants with higher paper reputation have a higher value of Q_i. If the paper reputation of a participant increases, the value of factor Q_i also rises, which generally encourages participants to contribute more high quality resources. Consequently, participants with high paper reputation are encouraged to contribute more resources, while the participants with low paper reputation are encouraged to contribute less.

In Cheng's implementation, the desired number of weekly contributions Q_i is determined by the summarized paper reputation (C_{ij}) of the participant i in week j, as well as the desired sum of resources desired from all users (Q). The higher the quality of the resources shared by the participant, the better paper reputation the participant gets, which means higher value of C_{ij}. As shown in Equation 4.2, the desired number of resources contributed by the user (Q_i) is proportional to the community factor Q (expressing the desired total number of contributions in the community) and the relative reputation of the user in bringing good papers, compared to the reputation of all the users (Cheng 2005). The calculations of the community factor Q and the summarized paper reputation (C_{ij}) are presented in Appendix E.

$$Q_i = Q * \frac{C_{ij}}{\sum_{k=1}^{N} C_{kj}} \tag{4.2}$$

Here, N is the number of users in the community.

Taking those two factors into consideration, the value of the reward unit for sharing (W_S) depends on the constant part W_{S0}, the time-function factor T, and the over-limit factor F_i, where the time-function factor T depends on the index of the time period, which will be explained in detail in Section 4.3.3.1:

$$W_S = W_{S0} * T * F_i \tag{4.3}$$

- The quality of contributions (PaperQualCr)

As described in the thesis of Cheng (2005), the quality of contributions is mainly determined by the weekly average ratings earned by the participant, as well as the rewards for the quantity of the contributions.

$$PaperQualCr = PaperQuanCr * AverageRating * W \tag{4.4}$$

The variable *W* in Equation 4.4 is a constant which has a different value for different periods of time (weeks). The variable "*AverageRating*" is calculated as:

$$AverageRatings = \frac{Total \quad number \quad of \quad ratings \quad earned}{Total \quad number \quad of \quad resources \quad shared} \tag{4.5}$$

- The quantity of ratings given by the participant (RatingQuanCr)

The quantity of the ratings are calculated by the product of the reward unit for rating (W_R) and the total number of ratings given by specific participant (RatingQuan). As shown in Figure 2.4, the value of W_R is constant.

$$RatingQuanCr = W_R * RatingQuan \tag{4.6}$$

- The quality of ratings given by the participant (RatingQualCr)

The quality of ratings given by the participant is calculated as the sum of the quality of each rating. As mentioned in Cheng's work (2005), the quality of each rating is measured by the difference between the value of rating and the average of all the ratings that the resource gets. In other words, a rating is considered to have high quality if there is little difference between its value and the average of all ratings received by that resource from all users. The advantage of judging the quality of ratings in this way is that the individual subjective biases can be balanced to some extent in the long run. However, this approach to evaluate the quality of ratings is based on the assumption that "the average rating of a resource can reflect the opinion of the majority and is less biased" (Cheng 2005). Thus it is good for Comtella, but not suitable for online communities with visible average rating values where the judgement of participants can be influenced or intentionally modified close to the value of the average rating.

In Cheng's implementation, the value of ratings can be either +1 or -1. He used the following equation to calculate the rewards for the quality of ratings:

$$RatingQualCr = \sum_{j=1}^{D_i} (1.5 * (1 - \left| r_{ij} - \overline{r_j} \right|)) \tag{4.7}$$

Here D_i is the total number of ratings (to different resources) given by the participant i, and $r_{ij}\,(i = 1,2,3,...,D_i)$ represents a rating given by the participant i for resource j. The variable $\overline{r_j}\,(j = 1,2,3,...,D_i)$ represents the average rating obtained from all users who have rated this resource j. Consequently, according to Equation 4.7 the range of the reward for each rating is [0, 1.5].

After presenting the details about the implementation of the extended version of the incentive mechanism in the real system, I will discuss the implementation of the agent-based model for this incentive mechanism.

4.3 Model implementation

4.3.1 AnyLogic

The model is developed by using the software AnyLogic. AnyLogic is a professional multi-method simulation tool that is quite efficient for modeling dynamics systems, even highly heterogeneous systems. It supports not only agent-based modeling, but also system dynamics, and discrete event modeling.

The reason for choosing AnyLogic for agent-based modeling is that AnyLogic is powerful and flexible. It is unique in its ability to efficiently capture the complexity and heterogeneity of dynamic systems, as well as analyze a diverse range of real-world problems (XJ Technologies 2002). First, agents in AnyLogic can be modeled at any degree of abstraction, and the model is cross-platform due to the Java technology involved in the AnyLogic language. For this reason, AnyLogic also supports the Java API and external connections, and can run as a Java applet in a web browser, which allows defining complex data structures or functions. In addition, AnyLogic has powerful data analysis tools and library support, as well as interactive GUIs with animation. AnyLogic provides a huge number of numerical methods, graphical methods, and data structure definitions, as well as powerful model control and animation support, which enable users to better investigate the model and the simulation results. The latest version of AnyLogic even supports sensitivity analysis, which allows better calibrating and validating AnyLogic models. Researchers can easily build models by using graphical shapes or controls (such as sliders, buttons, etc.); and control all the interactions among agents by giving a piece of Java code in either graphical charts or the properties of graphical shapes. An animation editor also enables one to present the simulation results in a more vivid way.

The basis of AnyLogic models is the “active object” that represents an autonomous agent. Each active object class has its own internal structure and behaviours, and they may “encapsulate other objects to any desired depth” (XJ Technologies 2002). Developing AnyLogic models means to develop the active object classes, their relationships, and the rules of interactions. In an AnyLogic agent-based model, the active object class can also encapsulate several AnyLogic objects:

- Variables: Variables represent the data units that have the potential to change continuously over time and can be shared with other active objects.

- Functions: Functions are used in active object classes (as used in Java) in order to enable reuse of code or decomposing code into simpler pieces to improve readability.
- Lookup tables: Lookup tables are used to define non-linear relationships that could not be represented by functions.
- Chart timers: Chart timers are simple ways to schedule user-defined actions.
- Statecharts: Statecharts can be used to define the activities within the active objects. It consists of states and transitions. States represent typical elements, and transitions from one state to another can be triggered by event based factors or time. Agents can take any actions (defined by Java code) after transitions.

4.3.2 Model design

Each user in the community is modeled as an autonomous participating agent. All agents in the model have unique identity numbers which are used to distinguish agents, just like the user name in the real Comtella system. Also, these participating agents possess several characteristics that bind them with other agents through the environment. These characteristics are vital to participating agents and act as fundamental social factors that connect participating agents as members in the online communities.

In the study of R. Cheng (2005), participants in the latest version Comtella system are encouraged to engage in six cooperative activities, and these activities are used to evaluate their contributions:

1. Stay online;

2. Log on the system;

3. Download resources and re-share resources;

4. Contribute new resources;

5. Comment on the resources;

6. Rate resources.

It is extremely hard to model all of these activities, since the agent-based models need detailed information at the agent-level, which is hard to obtain. First, in the real Comtella system it is very difficult to exactly record how long participants stay actively online, since the real Comtella system is session-based and the session instantiations are always implicit. Second, there were only a few comments on the resources posted in the community, which does not provide enough data for the agent-based model. Also, one of the goals of this study is to investigate the factors that influence maintaining a certain level of contributions and ratings in the community. A participant who shares or rates nothing is not helping to make the community sustainable, even if the participant logs in or downloads resources frequently. Therefore, my agent-based model only focuses on the two most beneficial cooperative activities: contributing new resources and providing ratings.

The design of the agent-based model (using AnyLogic) is shown in Figure 4.4, where inheritance arrows point at the base class and the dotted arrows show the dependencies. The arrows in the figure show the use of AnyLogic objects and tools. There are three main active object classes in the model: “User” (represents participating agent), “Paper” (represents resources), and “UserGroup” (represents the whole community). Each of these active object classes contains several parameters and variables, and some of the variables are AnyLogic objects, such as chart timers and statecharts. Arrows represent the relationships between different active object classes, and the related variable names are shown on the arrows.

The active object class “*Paper*” represents all of the resources shared in the Comtella system, and the details will be provided in Section 4.3.3.1. The active object class “*UserGroup*” represents a group of Comtella participants, which contains a set of instances of active object class “User” and “Paper”. The details will be provided in Section 4.3.3.2.

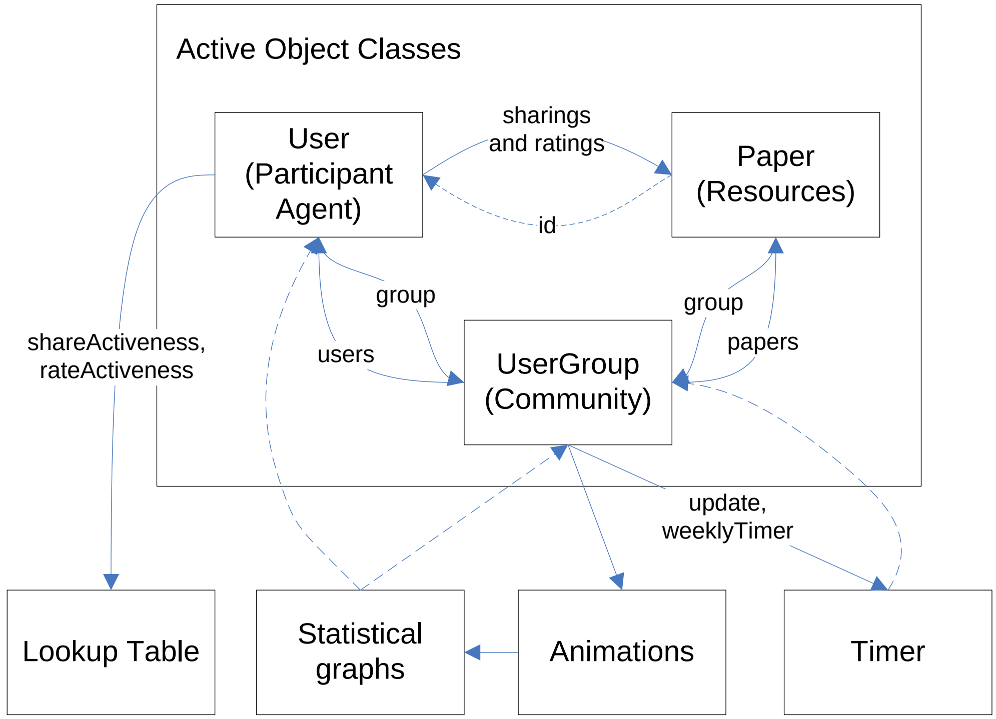


Figure 4.4. Model design (agent-based model)

The active object class "*User*" represents a single Comtella community member (participant). Compared to the system dynamics model, I divide all the participants into four groups for study, and the classification is based on the activity level instead of membership levels. The details will be provided in Section 4.3.3.3.

Based on the model design (Figure 4.4), Figure 4.5 further gives an overview of the basic model structure in AnyLogic. The properties of the active object classes and the use of AnyLogic objects are displayed.

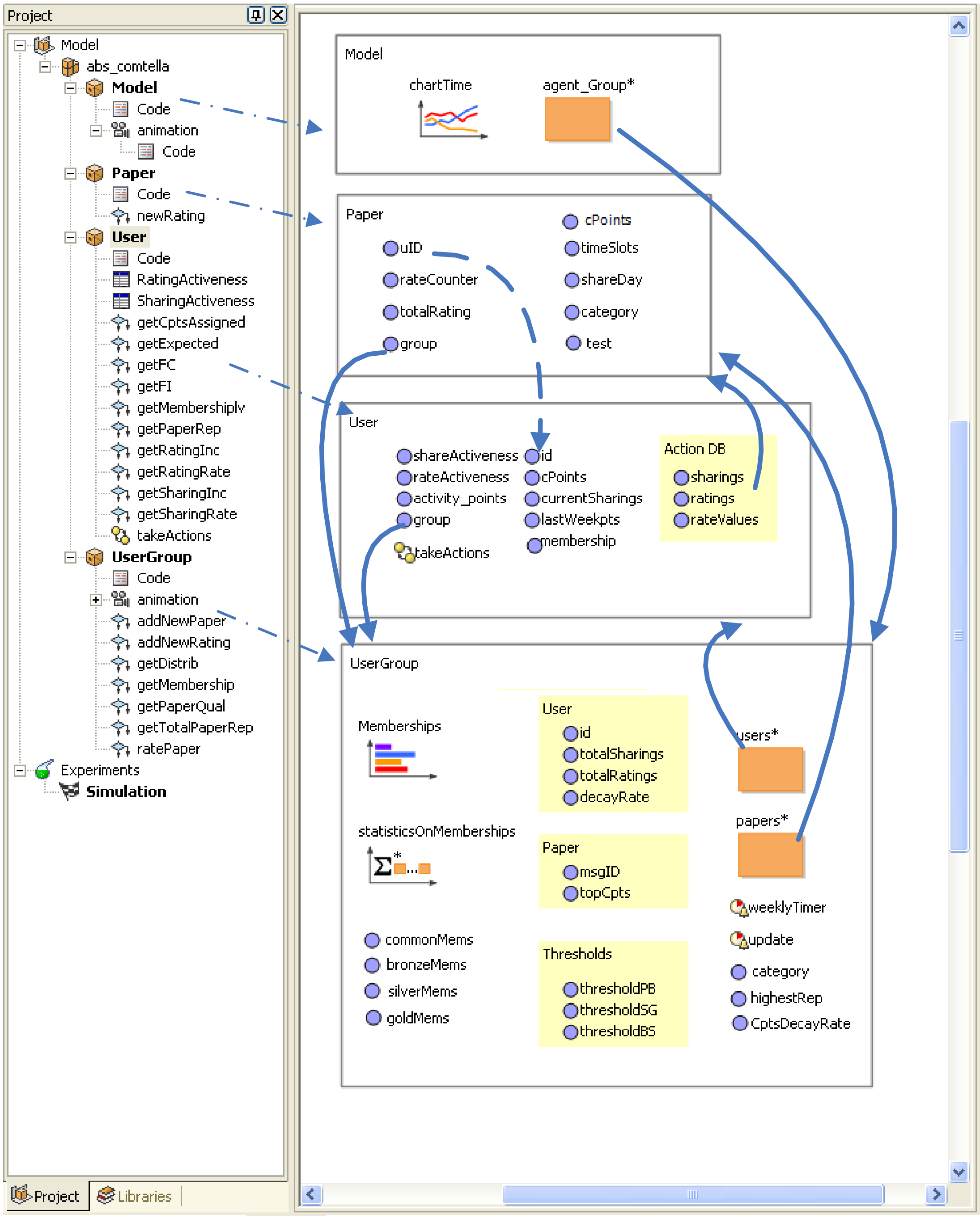


Figure 4.5. Agent-based model structure

In the figure, the left part of the view shows the tree menu of all the active object classes in the model as well as their facilities, such as lookup tables, functions, codes, and animations. The right part shows the details of each active object class, including its variables, statecharts, chart timers, statistical graphs, etc. The arrows in the figure show the Java encapsulations in the real agent-based model. For example, the model has one instance of class "*UserGroup*", which encapsulates a set of instance of class "*User*" (variable 'users'), and a set of instance of class "*Paper*" (variable 'papers').

4.3.3 Active object classes

4.3.3.1 Paper

The active object class "*Paper*" represents the contributions shared in the online community. Some important information is recorded by variables or updated during the simulation, such as average ratings and timeslots when it was contributed in a particular day. Its variables are listed in Table 4.1.

Table 4.1. Parameters and variables of the active object class "Paper"

Variable name	Type	Comment
cPoints	integer	The number of C-points that are assigned to the paper.
uID	integer	The id of each paper.
rateCounter	integer	It represents how many times this paper has been rated.
totalRating	integer	The sum of all the ratings of this paper, which is used to calculate the average rating $\overline{r_j}$ in Equation 4.7, for paper j.
group	UserGroup	The class that contains this instance of paper as well as system parameters.

timeSlots	integer	The index of the time when the paper is contributed.
share Day	integer	The index of the day when the paper is contributed.
category	integer	The index of the week when the paper is contributed.

Here, the two indexes, "*timeSlots*" and "*shareDay*", generate the index of the time period when the paper is contributed. In Comtella system, each day has its index (Monday has index 0 and Sunday has index 6), and is divided into 10 time periods. Consequently one week has 70 time periods for the hours of the day, and the variables have ranges of [1, 10] and [0, 6] respectively, which imply the index of the time period has a range of [1, 70]. For example, if participants shared one web resource on February 15, 2005 (Tuesday) at 9:25am, the value of "*Index of the time period*" is 12.

Details are presented in Table 4.2, where the columns "*Index of the day*" and "*Index of the time*" represent variables "*shareDay*" and "*timeSlots*" respectively. The values of these two variables are generated according to the distributions obtained from the database analysis.

Table 4.2. Definition of the time index

Index of the day	Definition	Index of the time	Definition
0	Monday	1	2am-8am
1	Tuesday	2	8am-10am
2	Wednesday	3	10am-12pm
3	Thursday	4	12pm-2pm

4	Friday	5	2pm-4pm
5	Saturday	6	4pm-6pm
6	Sunday	7	6pm-8pm
		8	8pm-10pm
		9	10pm-12am
		10	12am-2am (the next day)

4.3.3.2 UserGroup

It is shown in Figure 4.5 that the active object class "*UserGroup*" contains several variables and objects. The class "*UserGroup*" generates a group of participating agents, and contains all the contributions shared in the system, as well as most system parameters such as thresholds, total number of contributions or ratings, rewards decay rate, reward units, etc (Figure 4.6).

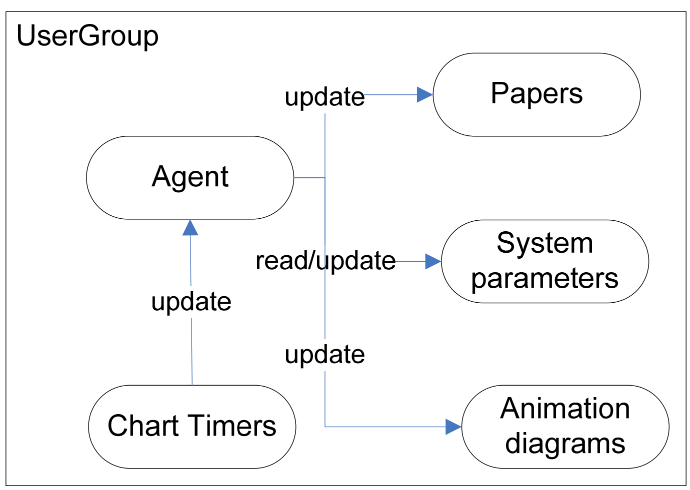


Figure 4.6. Active object class "UserGroup"

Figure 4.6 further shows the relationships among these variables and objects in the class "*UserGroup*". Agent behaviours change the properties of papers and update the system parameters. There are two chart-timers in the class "*UserGroup*". One of the chart timers schedules the weekly changes in the system, such as change the status of individual agents and reset some parameters. The other one updates the run-time simulation results, such as the statistical graphs and charts in the model. The variables of active object class "*UserGroup*" are listed in Table 4.3.

Table 4.3. Variables of the active object class "UserGroup"

Variable name	**Type**	**Comment**
id	integer	The variable used to generate the user id. Its value equals to the population of the community.
msgID	integer	The variable that is used to generate the paper id. Its value equals to the number of papers in the community.
totalSharings	integer	The total number of contributions.
totalRatings	integer	The total number of ratings.
decayRate	real	The percentage of rewards that will be lost in the next week.
CptsDecayRate	real	The percentage of C-points that will be lost in the next week.
users	User	The group of individual agents.
papers	Paper	The group of all papers shared in the community.
category	integer	The index of week in the simulation.
topCpts	integer	The highest number of C-points that are assigned to the resources.
highestRep	real	The highest paper reputation.

thresholdPB	integer	The number of activity points that are needed for common members to become bronze members.
thresholdBS	integer	The number of activity points that are needed for bronze members to become silver members.
thresholdSG	integer	The number of activity points that are needed for silver members to become gold members.
commonMems	integer	The total number of common members.
bronzeMems	integer	The total number of bronze members.
silverMems	integer	The total number of silver members.
goldMems	integer	The total number of gold members.

4.3.3.3 User

Participating agents behave individually, interact and update the environment, as well as create instances of web-resources. Its variables are listed in Table 4.4.

Table 4.4. Variables of the active object class “User”

Variable name	**Type**	**Comment**
id	integer	The index i of each participant.
shareActiveness	real	The activity level for sharing.
rateActiveness	real	The activity level for rating.
Activity_points	real	The number of activity points.
cPoints	integer	The number of C-points.
group	UserGroup	The class that contains this instance of agent as well as system parameters.

currentSharing	integer	The number of contributions shared by the participant in current week.
lastWeekpts	real	The number of activity points that are gained in the last week.
membership	integer	The membership level of each participant. Its range is [1, 4]. Here value 1 and value 4 stand for common membership level and gold membership level respectively.
sharings	Vector	An id list of the papers that are contributed by the participant.
ratings	Vector	An id list of the papers that are rated by the participant, and the number of items in this vector is the variable D_i in Equation 4.7.
rateValues	Vector	A list of rating values given by the participant, corresponding to the variable $r_{ij}\,(i = 1,2,3,\ldots,D_i)$ in Equation 4.7.

For each participant, there are three lists to record respectively the contributions shared by the participant, the resources rated by the participant, as well as the values of the ratings. These three lists track all the actions of individual participants in the system. Finally, the variable "*activity_points*" determines the membership level.

The system is determined by the actions of agents, which change the behaviours of the system. There are two main actions: sharing and rating. These actions are defined in the state chart "*take_actions*", and depend on the activity levels "*shareActiveness*" and "*rateActiveness*". In Section 4.3.4 and Section 4.3.5, the activity level variables and the state chart "*take_actions*" will be discussed in detail respectively.

Figure 4.7 shows the relationships among the activity levels, reputations, as well as weekly contributions and ratings.

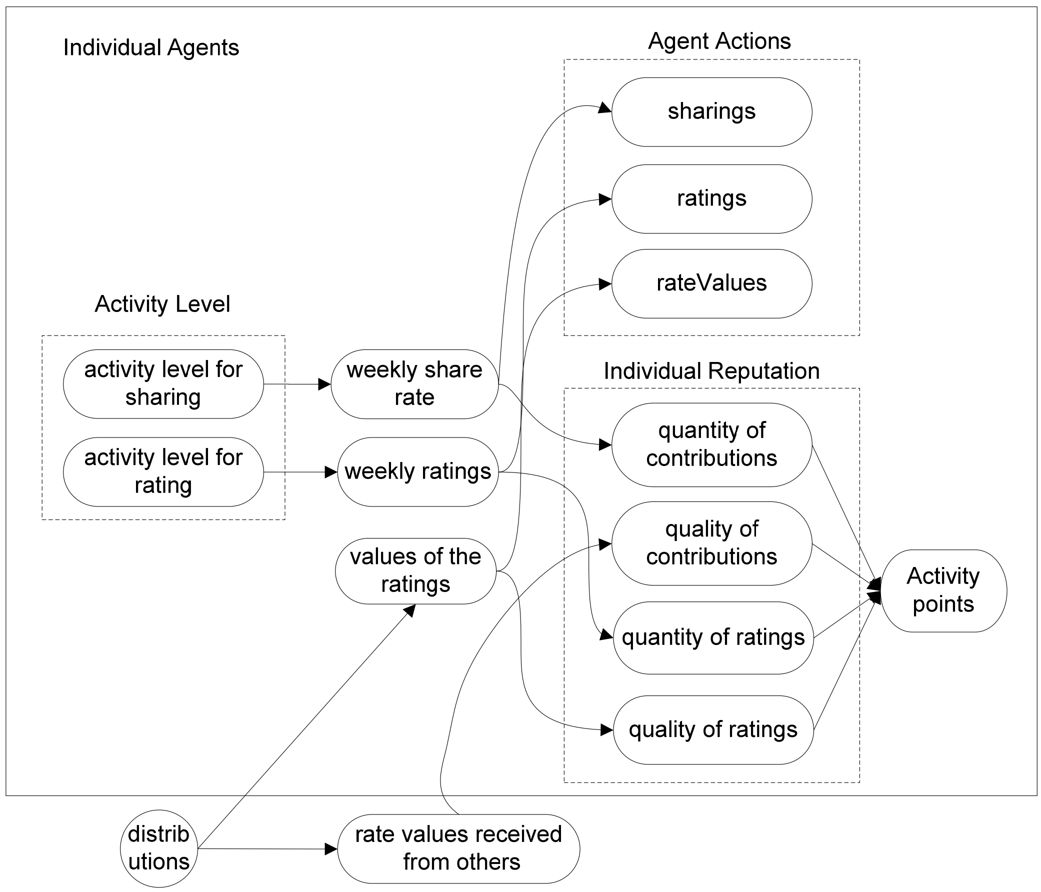


Figure 4.7. Individual model

Each agent has two activity levels, one for sharing and one for rating. The activity levels determine the number of contributions and ratings in each week and generate the actions of the agents. The actions are recorded in the three list variables: "*sharings*", "*ratings*", and "*rateValues*". At the same time, the activity points that are rewarded for the four parts of individual reputation (the quantity and quality of contributions, and the quantity and quality of ratings) are calculated using the equations mentioned in Section 4.2.

4.3.4 Activity level

In my model, activity level variables are the most important attributes that participating agents possess. All the interactions among agents depend on the activity levels, which measure how frequently the agents are willing to participate in the online community. Since this model primarily studies two cooperative activities, there are also two different activity levels that are related to these two kinds of participation (sharing and rating): activity level for sharing new resources, and activity level for rating.

Activity levels are defined by the percentage of participation at the very beginning, and will be influenced dynamically by the incentive mechanism. They determine the weekly share rates and the number of ratings. The higher the activity level, the more resources are shared or rated. For example, participants with higher activity level for sharing will contribute with higher weekly share rates. The values of activity levels vary over time within the range of [0, 1], and reflect the personalities of participants in the community. Although participants are different, I define four different levels of engagement with different ranges of activity level in the agent-based model (Table 4.5), according to analysis results on the Comtella database (Figure 4.8).

Table 4.5. Four levels of engagement in agent-based model

Type	Activity level	Percentage of participation	Description
Invisible engagement	[0, 0.25)	[0%, 1%)	Participants do not share anything in the community. Some of them might even have no interest in Comtella at all (no engagements).
Inactive engagement	[0.25, 0.5)	[1%, 5%)	Participants seldom participate (share or rate resources) in the community.
Conservative engagement	[0.5, 0.75)	[5%, 10%)	The main objective of the participation is to get enough activity points and maintain or gradually upgrade the membership levels of participants.
Competitive engagement	[0.75, 1]	[10%, 100%]	Participants are highly active, and have high probabilities to share or rate web-resources in the online community. On the other hand, sometimes they might stop sharing if they did not get what they desired.

Although this classification is based on personal opinions, Figure 4.8 is presented to justify that this classification is reasonable. In Figure 4.8, the first two diagrams show the cumulative percentages of participations in the Comtella database, and the other two diagrams show the percentages of participations for each participant in the community. For each diagram, the curve with diamond points shows the ascending percentages of contributions for each individual participant, and the other curve with square points shows the ascending percentages of ratings.

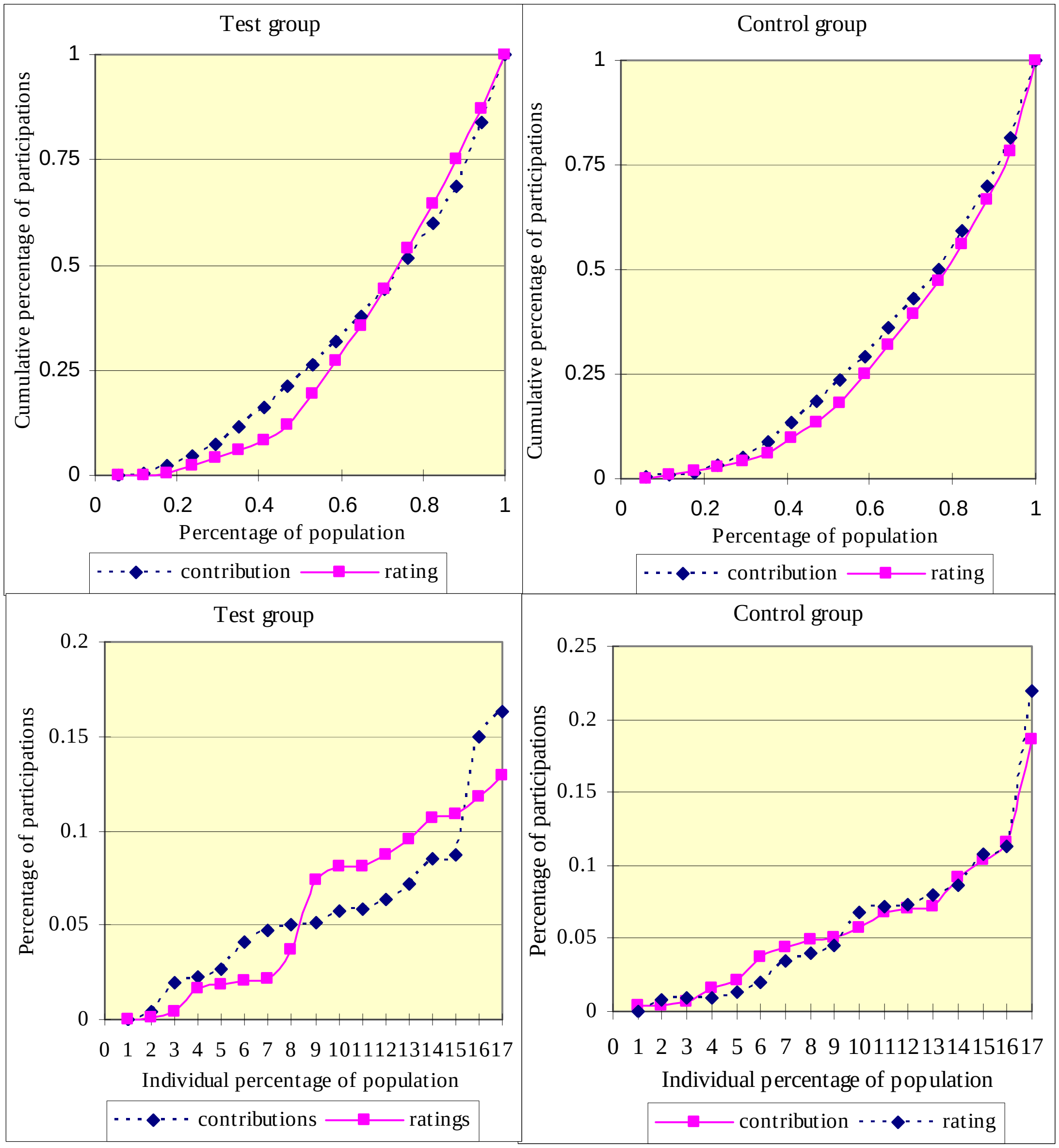


Figure 4.8. Database analysis on participations

From the four diagrams we can see that 25% of the participants share more than 50% of the resources, and on the other hand 35% of the participants share less than 10% of the resources in total. The participants who contribute less than 1% are totally inactive, and they can be considered as free-riders. Around half of the participants participate less than 5%, and they share or rate less than 25% in total. Participants who contribute or rate more than 10% can be considered as highly active users. Although there are only around 3 or 4 out of 17 participants in this class, they generate more than 40% of contributions/ratings.

As a result, percentage 1%, 5%, and 10% are used as thresholds to divide the four different levels of participation, and Figure 4.9 shows the relationship between the activity level and the percentages of participation. Through statistical analysis of the Comtella database, the distributions of participants with different levels of engagement (sharing and rating respectively) are used to generate the initial value of activity level for sharing and activity level for ratings.

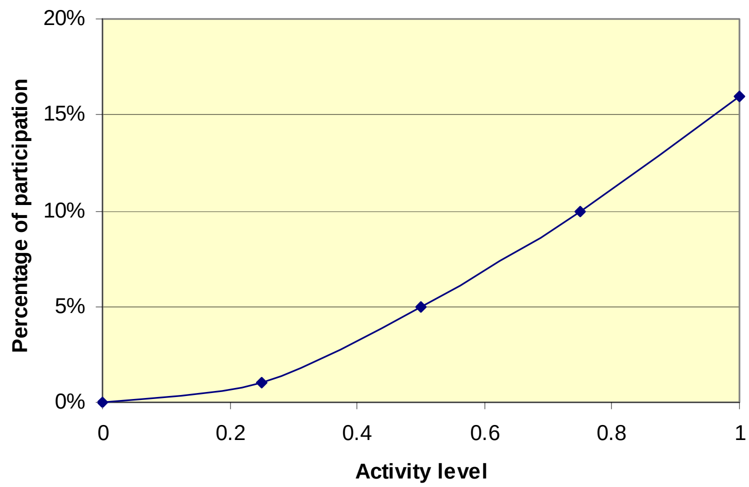


Figure 4.9. Classification of user groups based on activity levels

For the activity variables, three things need to be considered: how to initiate the activity levels, how to change their values, and how to use them to calculate weekly share rates and ratings. The next section will explain these in detail.

4.3.4.1 Initialization of activity level

According to different activity levels, the community can be broken down into four groups. Table 4.6 and Table 4.7 show the number of participants in each of the group with different activity levels for sharing. The data in these two tables are collected from the experiments on the test group and the control group respectively.

Table 4.6. Activity level for sharing (test group)

Activity level for sharing	Number of participants in the beginning	Number of participants at the end
[0, 0.25)	2	2
[0.25, 0.5)	6	5
[0.5, 0.75)	9	8
[0.75, 1]	0	2

Table 4.7. Activity level for sharing (control group)

Activity level for sharing	Number of participants in the beginning	Number of participants at the end
[0, 0.25)	2	3
[0.25, 0.5)	8	5
[0.5, 0.75)	3	6
[0.75, 1]	4	3

Based on this statistical analysis of the Comtella database, the probability distribution of activity levels for sharing are obtained and stored in a lookup table named "*SharingActiveness*". For example, the numbers of participants with different activity levels for sharing in the test group (column 2 of Table 4.6) are 2,6,9,0 respectively. Thus, the percentage can be calculated through dividing these numbers by the total number of population 17, which is 0.118, 0.353, 0.529, and 0.

In the lookup table "*SharingActiveness*", the arguments are the activity levels for sharing, and the corresponding values are their percentages of population in the very beginning of the period, or the probabilities of different activity level for sharing. Using the probability distribution of activity level for sharing in the lookup table, AnyLogic applies linear interpolation to generate the cumulative distribution function (CDF) of activity level for sharing, and uses inverse transform to get the activity level for sharing for each participating agent.

Similarly, Table 4.8 and Table 4.9 show the number of participants in each of the groups with different activity levels for rating. The data in these two tables are collected from the experiments on the test group and the control group respectively.

Based on these data from the Comtella database, the probability distribution of activity level for rating is obtained and stored in a lookup table. Same as the activity level for sharing, AnyLogic generates the cumulative distribution function (CDF) from this lookup table and uses inverse transform to get the activity level for rating for each participating agent.

Table 4.8. Activity level for rating (test group)

Activity level for rating	**Number of participants in the beginning**	**Number of participants at the end**
[0, 0.25)	6	2
[0.25, 0.5)	2	6
[0.5, 0.75)	4	5
[0.75, 1]	5	4

Table 4.9. Activity level for rating (control group)

Activity level for rating	Number of participants in the beginning	Number of participants at the end
[0, 0.25)	7	4
[0.25, 0.5)	4	5
[0.5, 0.75)	2	5
[0.75, 1]	4	3

4.3.4.2 Update of activity level

To update the activity level, several factors motivating users to participate more in the community are considered:

- Reward incentive factor

Reward incentive factors represent the incentive from activity points. It is assumed that participants will participate more if they get more rewards than in the previous week. On the other hand they will be de-motivated if they get fewer rewards than the previous week.

$$Award\ incentive\ factor = \frac{New\ awards\ acquired\ in\ current\ week}{Awards\ gained\ in\ last\ week} \tag{4.8}$$

Reward incentive factors are calculated based on Equation 4.8. Both the new rewards acquired in the current week and the rewards gained in previous week are considered. When the reward incentive factor is high enough to motivate participants, activity level will increase to some extent based on different levels of membership.

- Membership incentive factor

Obviously, the activity points needed to reach the next higher level of membership influence the motivation of the participants to contribute or rate more resources. The membership incentive factor is calculated based on the following equation. When it is low enough to motivate participants, their activity level will increase.

$$\textit{Membership incentive factor} = \frac{\textit{Threshold for higher level of membership - Current number of activity points}}{\textit{Total number of activity points needed for membership upgrading}} \tag{4.9}$$

- Chance to be motivated

AnyLogic is used to generate the Bernoulli sample values, which are used to determine the possibilities for all of the participants to be motivated. Activity level will increase when value 1 is generated, and the probability is arbitrarily chosen.

- Time factor

For the real Comtella community, normally each record in the database is related to a week. However the record in the 6^{th} week in the database contains the data for three weeks in real time. So it is assumed that in the agent-based model each participating agent has a higher weekly contribution and ratings in that time period. Similarly, the activity level will decrease a little bit at the end of the term.

4.3.4.3 Calculation of weekly contributions and ratings based on activity level

As mentioned earlier in this chapter, the weekly contributions and ratings are determined by the activity levels for sharing and rating respectively.

Based on the statistical analysis of the Comtella database, the weekly average contributions and ratings for different activity levels are listed in Table 4.10 and Table 4.11 for the test group and the control group respectively. The main difference between the test group and the control group is that the control group is tested in an environment without C-points and adaptive reward unit, as listed in Table 3.2 (page 51). There are several steps in the statistical analysis. First, the contributions and ratings are counted for each participant in the community from the historical data in the Comtella database. Then the sum of contributions and ratings for participant group with same activity levels are obtained by adding the data based on the activity levels. The individual average contributions and ratings for different activity levels are calculated by dividing the size of the corresponding participant group with the same activity level. Last, by dividing the number of weeks in the experiment, the weekly average contributions and ratings are obtained.

Table 4.10. Weekly average contributions

Activity level for sharing	**Weekly contributions (test group)**	**Weekly contributions (control group)**
[0, 0.25)	2	2
[0.25, 0.5)	5	5
[0.5, 0.75)	9	9
[0.75, 1]	14	15

Table 4.11. Weekly average ratings

Activity level for rating	**Weekly ratings (test group)**	**Weekly ratings (control group)**
[0, 0.25)	3	1
[0.25, 0.5)	7	5
[0.5, 0.75)	12	9
[0.75, 1]	14	14

From these two tables, the relationship between the activity level and weekly contributions/ratings can be easily obtained and the weekly contributions and ratings can be formulated.

For example, for the activity level for sharing in the test group, there are five numerical values that can be obtained from Table 4.10: (0, 0), (0.25, 2), (0.5, 5), (0.75, 9) and (1, 14). These numerical points constitute a curve with five points, which represents the weekly average contributions for test group (Figure 4.11).

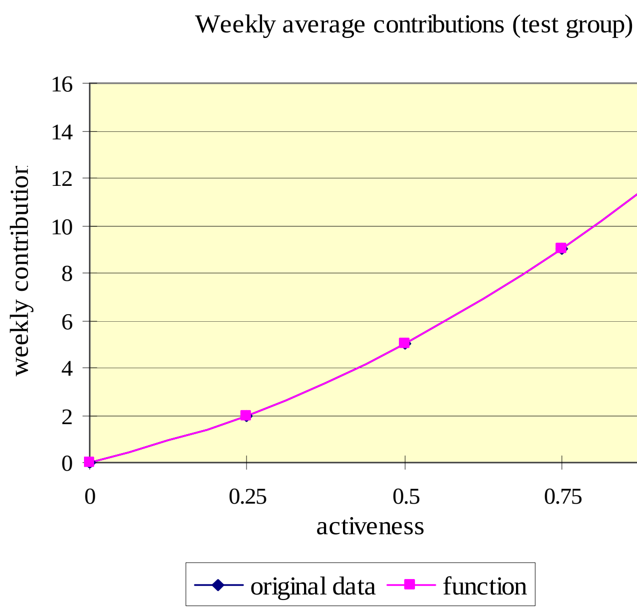


Figure 4.10. Curves of weekly average contribution (test group)

In order to map this curve, the exponential function with three parameters (A, B, and C) is used to formulate the weekly contributions:

$$Weekly\ contributions = B * A^{activeness\ \ for\ sharing} \text{-} C \qquad (4.10)$$

After the values of the three parameters are obtained, two equations are found by statistical analysis for the test group and the control group respectively, and they can be used to calculate the weekly contributions for the test group and the control group respectively as below:

$$Weekly\ contributions_{test} = \lfloor 3.825 * 4.482^{activeness\ \ for\ sharing} - 3.507 \rfloor \quad (4.11)$$

$$Weekly\ contributions_{control} = \lfloor 5.698 * 3.326^{activeness\ \ for\ sharing} - 5.490 \rfloor \quad (4.12)$$

Here, the results are changed to integers with same range [0, 14], in order to use them in AnyLogic as weekly share rates and ratings. The coefficients in the above equations are obtained by minimizing the mean squared error of the weekly contributions. This minimization was done by using Solver in Microsoft Excel. Excel Solver is an optimization tool which is quite useful in modeling. It helps to get the best set of the coefficients with the minimum difference between the function and the real curve obtained from data analysis.

Similarly, the exponential function with three parameters is used to formulate the weekly ratings:

$$Weekly\ ratings = B * A^{activeness\ \ for\ rating} - C \quad (4.13)$$

After the parameters are calibrated, the following equations are formulated to calculate weekly ratings for the test group and the control group respectively, and the results are changed to integers with same range [0, 15].

$$Weekly\ ratings_{test} = \lfloor 7.246 * 3.59^{activeness\ \ for\ sharing} - 6.792 \rfloor \quad (4.14)$$

$$Weekly\ ratings_{control} = \lfloor 13.051 * 2.34^{\ activeness\ \ for\ rating} - 12.598 \rfloor \quad (4.15)$$

4.3.5 Actions of agents

Actions will change the system. There are two main actions: share and rate. These actions are defined in state chart "*take_actions*" (Figure 4.11) and are trigged once per week.

Sharing behaviours determine the total number of contributions in the community.

When users share any web-resource, this action is performed and new instances of active object class "*Paper*" are generated based on the activity level of participants. Participants who share those web-resources will be rewarded with a number of activity points. In the model, participants in different user groups have different activity levels for sharing web-resources, which are assigned by the variable "*shareActiveness*" and determine the value of the weekly share rates. For simplicity, "*shareActiveness*" is assigned a value with range [0, 1], and there are four different sub-levels that are set according to different contribution levels of the participants (presented in proportions of the total number of contributions). An inner vector variable "*sharings*" is used to record all IDs of the web-resources that are contributed by the particular agent.

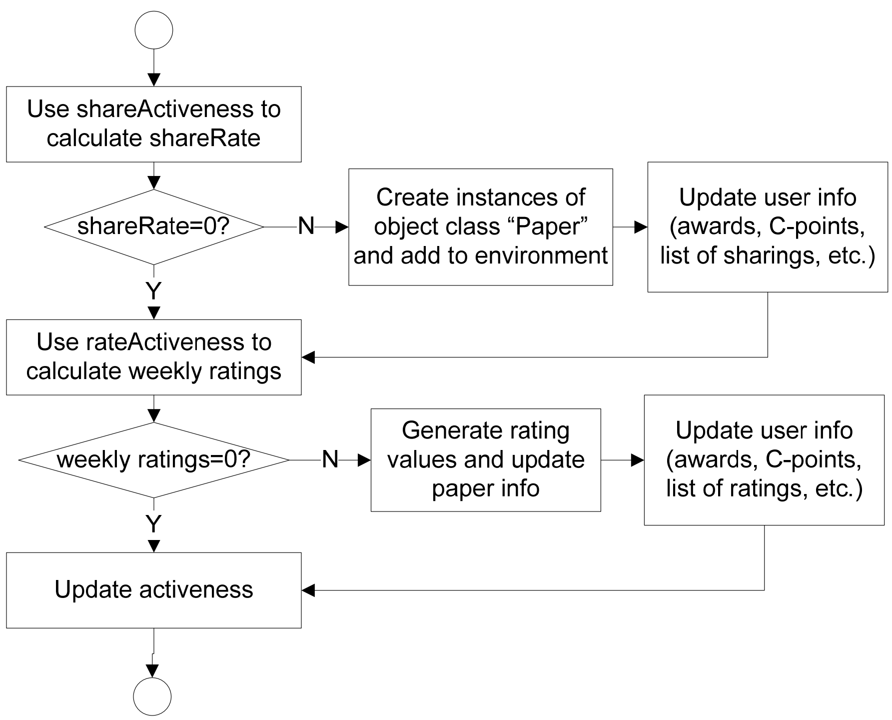


Figure 4.11. Flowchart of agent behaviours

Rating behaviours measure the quality of the resources shared in the community. In the model a variable "*rateActiveness*" is used to control rating behaviors as defined criteria, and it determines the number of weekly ratings. Similar to the variable “*shareActiveness*”, the variable “*rateActiveness*” is also assigned a value with range [0, 1], and there are four different sub-levels that are set according to different levels of ratings (presented in proportions of the total number of ratings). An inner variable “ratings” is also used to record all IDs of the web-resources that have been rated by the particular agent.

4.4 Results

As mentioned in Section 3.3.1.2, the real data for the latest version of Comtella community was collected during an 11-week comparison experiment with two groups of students in 2004-2005 winter sessions: a test group and a control group. As mentioned in Table 3.2 (page 51), the main differences between the two groups is that there is no C-points and adaptive rewards units introduced in the system for the control group. There are 17 participants in each group, and Table 4.12 lists the parameters for the test group and the control group.

Table 4.12. Parameters for agent-based model

Parameters	Test Group	Control Group
Size of Agent Group	17	17
Rewards for Sharing	4	4
Rewards for Rating	3	3
C-point rewards for rating	3	0
Initial C-points Assigned	20	0
Rewards Decay Percentage	0.7	0.7
C-points Decay Percentage	0.5	0
Common Member Upgrading Threshold	24	24
Bronze Member	32	32

Upgrading Threshold		
Silver Member Upgrading Threshold	40	40

For the control group, the distributions of agents based on different activity levels are different from the test group. Besides, the corresponding equations used to calculate weekly contributions and ratings are different from the test group as well.

Figure 4.12 shows the simulation results for these two separate systems. All participants are common members at the very beginning, and their membership levels are upgraded gradually over time.

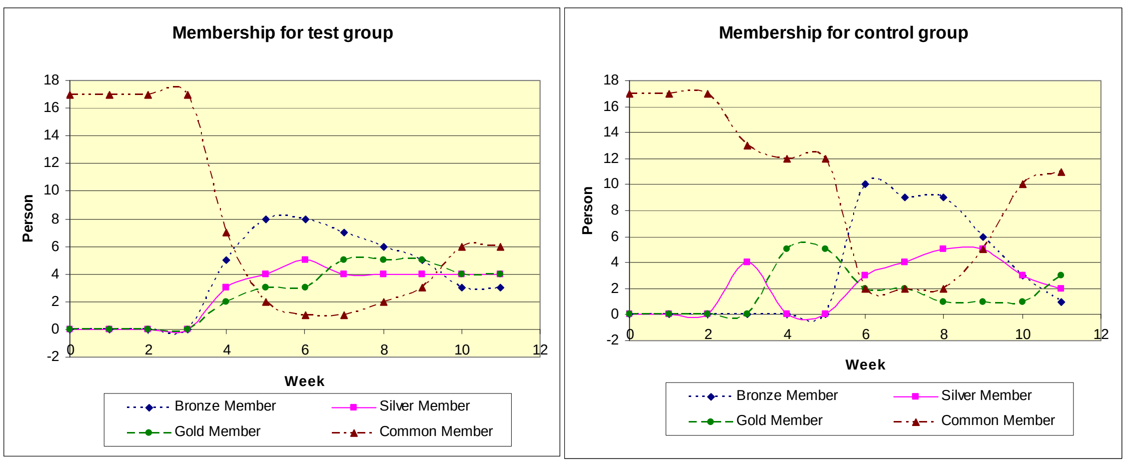


Figure 4.12. Simulation result (agent-based model)

For the test group, the population of bronze members reaches its peak value in the 5th and 6th weeks, which causes the population of common members to decrease to its lowest value in the 6th week. Afterwards the population of bronze members starts to decrease, while the common, silver and gold member groups start to increase their populations slightly. After 9 weeks the population of bronze and gold members decrease, which causes the common member group to increase its population again and reaches its peak value at 35.3 percent of the whole population in the 10th week. At the final point the population of the bronze, silver and gold members reaches 17.6, 23.5, and 23.5 percent respectively.

For the control group, the population of common members decreases to its lowest value at 11.7 percent in the 6th week, while the population of bronze members reaches its peak value. In the end, the population of common member increases to 64.7 percent of the whole population.

Comparing the two groups, the percentage of participants with low membership level (common members) is higher in the control group, which means there might be more free-riders in the control group and the number of participants with higher membership levels (bronze, silver and gold members) is higher in the test group towards the end of the experiment. Therefore, the extended version of the incentive mechanism can motivate participants in Comtella online community to some extent.

Similar to the previous system dynamics models, the simulation results are also compared with the real data, which was collected during one academic term experiment. Curves for output variables (including population of different user groups as well as total number of contributions and ratings) with trend lines are shown in Figure 4.13.

Compared to Comtella database, the simulation results for bronze members, gold members, and total number of contributions fit quite well with the real data. They have similar trend line and a similar value at the end of the simulation. The curve of total number of ratings has a little bit higher slope than the real data, but has a good fit at the end point. For common and bronze member groups, the model can only generate the identical numbers at the end of the simulation, but the trend lines have a small deviation from the 5th to the 9th week. Also, this agent-based model shows good capabilities to model emergencies (or anomalies) because the behaviours of agents are dynamic. Since the number of weeks is taken into consideration, the abrupt changes at the 5th and 6th week are captured in agent-based models. Consequently, the end values of outputs shown in the diagram have very good fits with the real data.

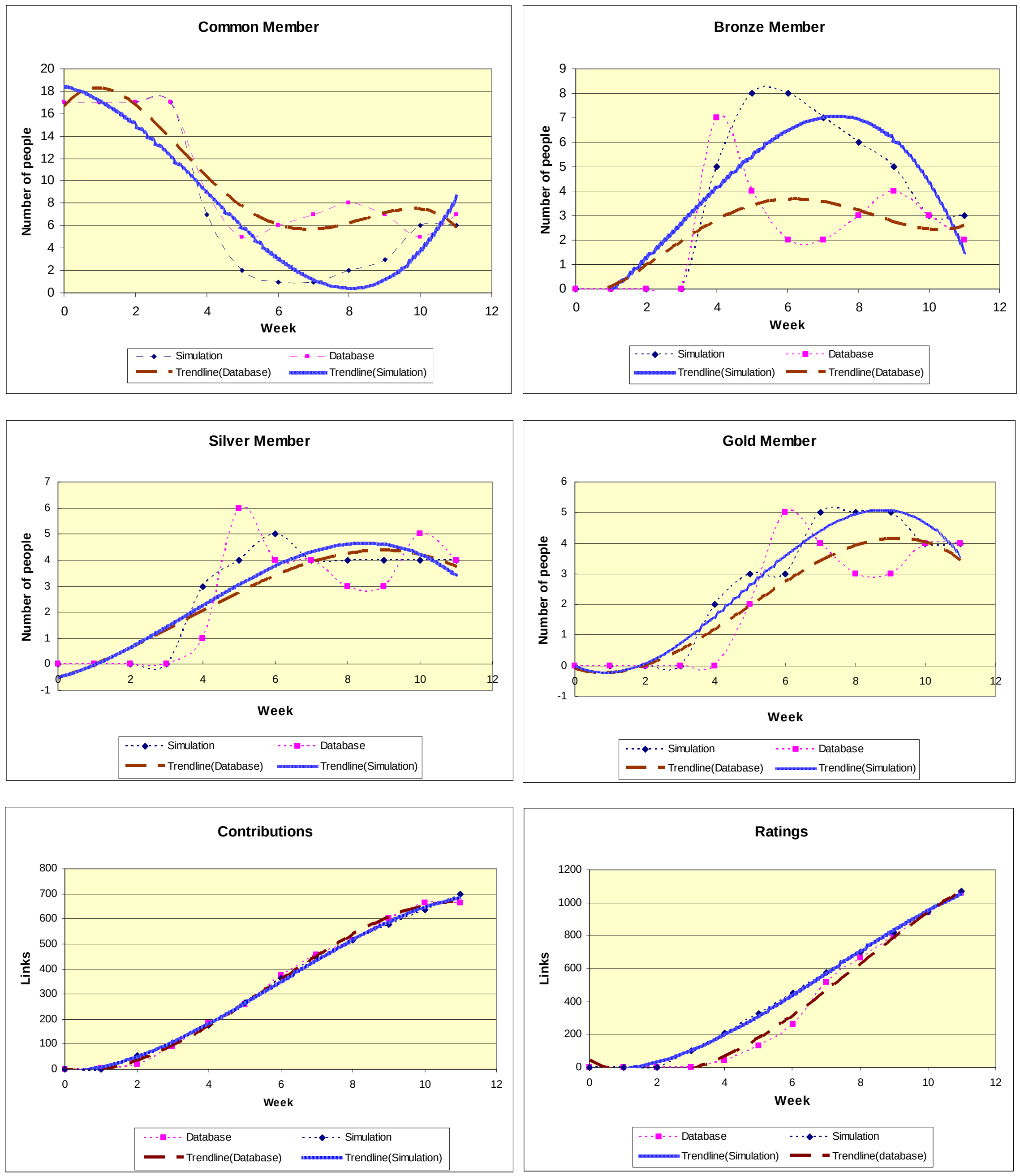


Figure 4.13. Comparisons with historical data (agent-based model)

However, the agent-level details are still needed for the variation of activity level. Randomness of agent behaviours as well as the characteristics of Comtella community (small-scale close system) also results in the numerical deviation.

Several experiments are launched to test system parameters for the test group that are important for evaluating and improving the effectiveness of current incentive mechanism. The following three experiments test the impacts of the population size, the decay rate, and the membership thresholds on the total number of contributions and the percentage of middle level participants respectively. The values of these system parameters are changed, and the diagrams to compare the results are provided.

1. Change the population size

Experiment results on changing the population size are shown in Figure 4.14. Like in the system dynamics experiments, variable *"Initial population"* is set to be 17, 32, and 50 respectively.

For the total number of contributions, the curve for population of 50 has higher increase rates on slope than the other curves within the first 9 weeks. However, the warm-up stage (as mentioned in page 33) is longer when there are 50 participants in the community. Compared to the results obtained from the same experiment in the system dynamics model, the curves of contribution in the agent-based simulation have a little bit lower slope, and the total number of contributions is not proportional to the number of participants in the community. The average number of contributions is higher in the curve for population of 17.

Although the curves of the number of common and bronze members are not accurate in numerical respects, the general impact of population on the percentage of bronze and silver members can be obtained from the trends of the curves. For the population 17, the percentage of bronze and silver members decreases slowly at the end of the experiment while other curves decreases their values more gently. Compared to the similar system dynamics experiment, the agent-based experiment yields better results regarding the changes on the percentage of the middle level participants.

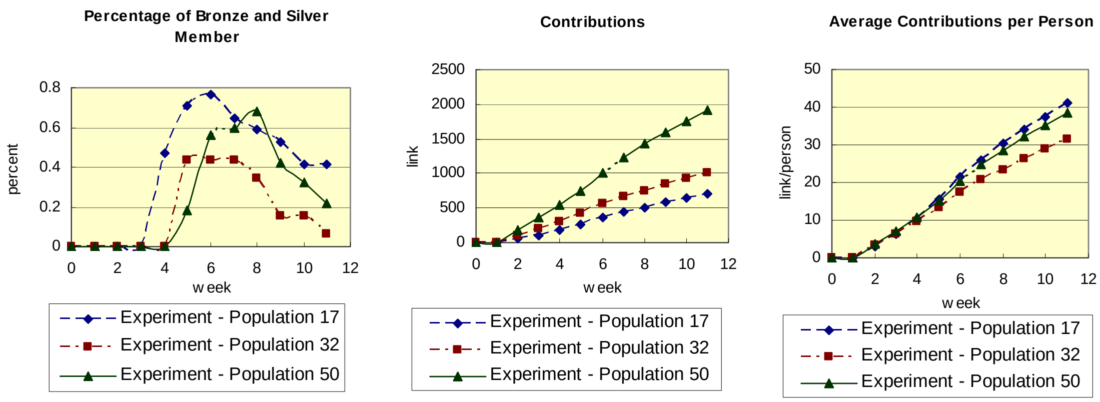


Figure 4.14. Experiment on the population (agent-based model)

2. Change the decay rate

From Figure 4.15 it can be seen that the decay rate for rewards has large effects on the percentage of bronze and silver members, but it has little effects on total number of contributions within the first 8 weeks.

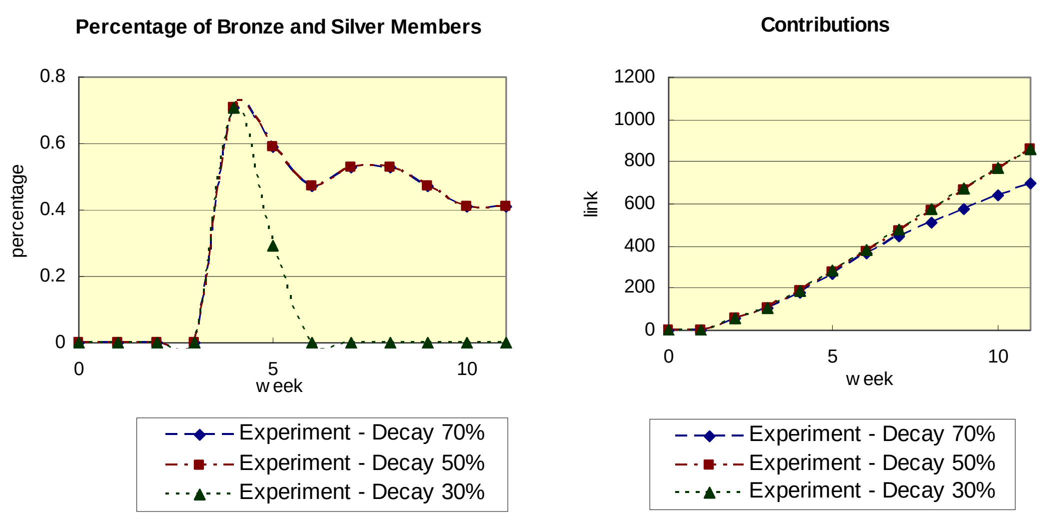


Figure 4.15. Experiment on the decay rate (agent-based model)

Bronze and silver members constitute a larger percentage of the whole population when the decay rate has a medium value of 50% or 70%. However, with the lower decay rate 30%, the number of bronze and silver members increases dramatically fast at the very beginning, and decreases sharply afterwards since they either become gold members, or totally lose interest in the community since there is no challenge to reach their goal on membership. Compared to the results obtained from the same experiment in the system dynamics model, the curve of the percentage of the middle level participants changes faster due to the ability of agent-based approach to model abrupt changes in my study. Therefore, the results are more reasonable.

Similar to the results obtained in the system dynamics experiments, after the 8th week the community gets more contributions when the decay rate has a smaller value. As a result, an appropriate value of decay rate is quite important to the community, and the 50% decay rate seems to be the best value for user motivation.

3. Change the membership thresholds

The impact of the three membership thresholds on the contributions as well as the demographic structure of the Comtella population are also investigated by changing the three thresholds respectively. The unit of variation is 1 reward unit for sharing.

Figure 4.16 shows the experiment results. Here, “threshold PB”, “threshold BS” and “threshold SG” stand for the thresholds of bronze, silver and gold membership level respectively. For these three thresholds, they all have impact on the total number of contribution and the percentage of bronze and silver members. However, although a decrease of the thresholds may result in higher number of contributions, the curves for the percentage of the bronze and silver members might become shaky in a certain period of time, which might result in a sharp increase or decrease of the middle level participants’ population. As mentioned before in page 61, participants should not be able to upgrade their memberships too fast. Consequently we could not simply lower all the thresholds to get a higher number of contributions and we also need to control the demographical structure of the Comtella population.

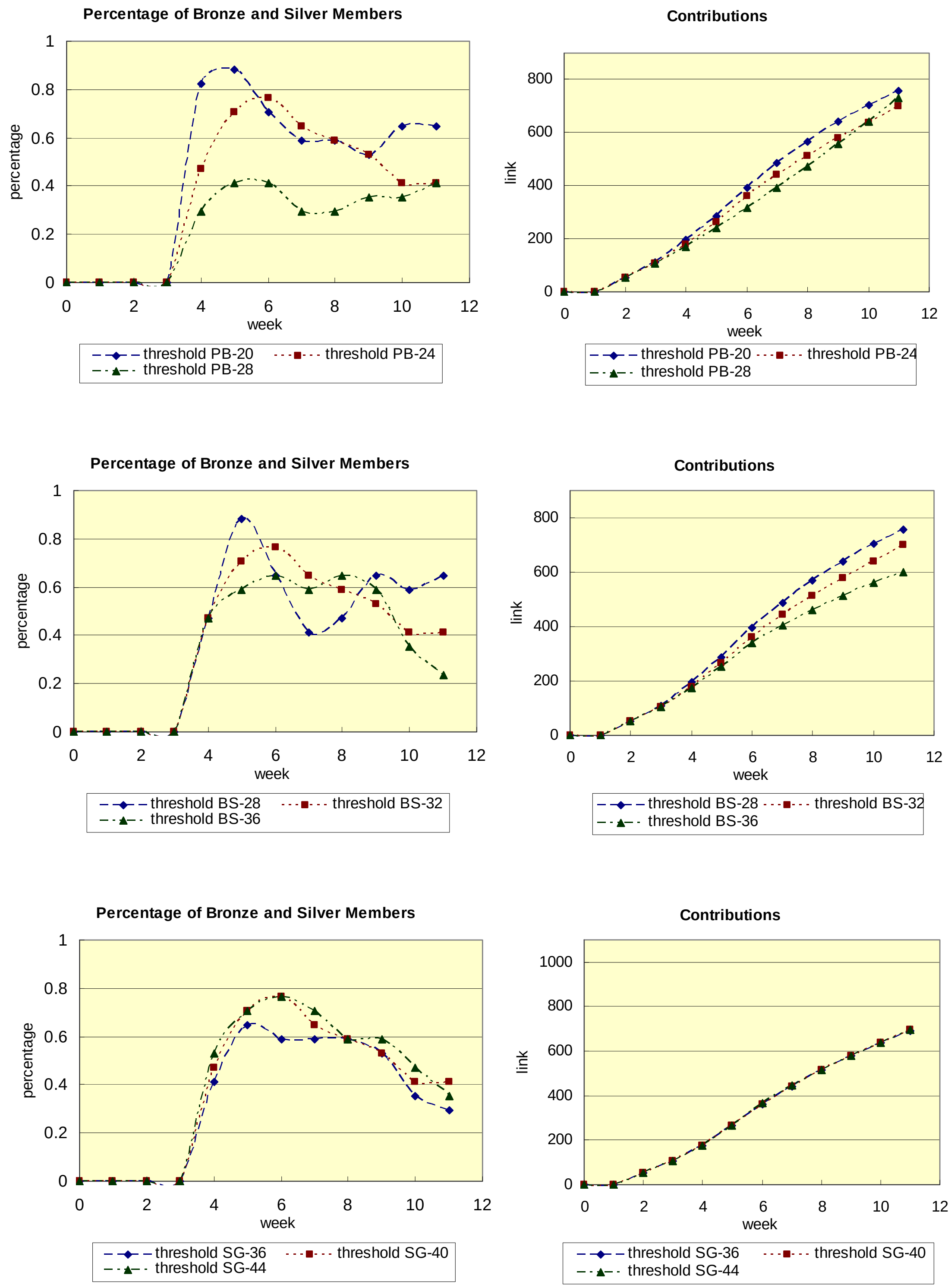


Figure 4.16. Experiment on the membership thresholds (agent-based model)

Taking the entire diagram into consideration, decreasing the bronze member thresholds will definitely benefit the community in the long run. For the silver and gold membership thresholds, the current values work fine for the Comtella community.

4.5 Summary

This chapter described in detail an agent-based model for a variation of the incentive mechanism of the Comtella community. In Section 4.1, background material of the agent-based model as well as AnyLogic is presented. Section 4.2 describes the details of the measurement on individual reputation. Details of the model implementation are explained in Section 4.3, including the detailed introduction of AnyLogic, the basic design of the agent-based model, the functions and implementation of the three important active object classes, as well as their interactions. Simulation and experiment results are discussed in Section 4.4, and future work is discussed in the next chapter.

CHAPTER 5

CONCLUSIONS ANF FUTURE WORK

In the previous two chapters, simulation models in both system dynamics approach and agent-based approach are proposed in order to study the user motivation process in the Comtella online community. This chapter presents the conclusions, discussions, and the directions for future work.

5.1 Conclusions and discussions

Lack of participation in online communities can cause their stagnation and demise. As an effective approach to encourage participation in online communities, incentive mechanisms are designed to make online communities sustainable over the long run. However, it is not easy to evaluate the effect of such incentive mechanisms for participation through experiments in the real world, which not only costs time and money, but also has some risks since inappropriate system parameters might result in unwanted behaviours and cause harm instead of benefits.

I used computer simulation models to investigate user motivation processes and evaluate the incentive mechanisms in an online community called Comtella. Comtella is a small-scale class-related system that enables participants to share web resources. Two different simulation methodologies are applied: system dynamics methodology and agent-based methodology. Furthermore, separate simulation models are developed for two different versions of the Comtella community: a model for an early version of Comtella incentive mechanism, which is called the "first version model", and a model called the "second version model" for the extended version of incentive mechanism with individual adaptive reward units, quality control, and membership decay. Different sets of parameters are set in the second version model for two different participant groups: the test group and the control group.

For the presented simulation models, several aspects mentioned in Section 3.2 are investigated:

1. How to classify the participants into different groups for study?

In the study of user participation and motivation, it is helpful to divide participants into several user groups. It is assumed that participants in the same user group have similar behaviours. User groups are identified in two different ways in my study: participants are divided into different groups based on their membership levels, or based on their activity levels for sharing and rating. The reason to consider dividing the Comtella population in different ways is that participants might behave differently even if they have the same membership level in the agent-based model. Dividing the population based on membership levels to define the individual behaviours in the agent-based model may lead to a big deviation from the observed behaviour. That is why I used the observed behaviours to define activity levels which are used to define the different behaviours of the agents. On the other hand, the system dynamics models in my study divide the Comtella population by the membership levels, because it is hard to find a trend in the database on how the activity levels of a group of users change over time, and the changes of the membership level can be obtained easily from the database.

It is quite difficult to compare the system dynamics approach and the agent-based approach: When dividing user groups based on membership levels, the demographic structure of the Comtella population is captured directly from the visualization of the community (from color, size and brightness of the star) or the Comtella database. Since it is assumed that everyone within the same group has similar behaviours, this approach is quite useful for system dynamics modeling because some part of the details on individual participants can be omitted. Besides, the information needed for model formulation such as the proportions of different groups, their impact on the total number of contributions and ratings can be obtained easily. On the other hand, dividing user groups based on activity levels seems more realistic for agent-based modeling since in agent-based models individual participants have different activity levels.

2. How do different groups impact the whole system?

As mentioned in Chapter 2, bronze and silver members are important to the whole community because they are considered as the middle level participants and they are easier to get motivated by the incentive mechanism. Normally they either have the goal to reach the higher membership levels, or they fear to lose their current membership level. On the other hand, gold and common members might lose their interest and stop participating in the community after they reach the highest membership level, or after several failures to reach to bronze membership level. As a result, the percentage of bronze and silver members needs to be maintained in order to make the community sustainable in a long run.

By performing statistical analysis of the real deployment database and using the simulation results of the system dynamics model, we see that the percentage of bronze and silver members ranges from 35% to 58% with an average value of 41.9%, and they share only about 45% of the resources and ratings in the online community. Obviously it is possible to increase the percentage of bronze and silver members.

For activity levels, it is assumed that users will participate more in the community when they have higher activity levels. Consequently those system parameters that have the ability to increase activity levels are vital to the community, such as reward unit, thresholds, decay rate, etc.

3. How users in the community behave according to the memberships based on the current incentive mechanism?

For the user groups based on different membership levels, the bronze, silver and gold member groups increase their population respectively and reach the peak value within the 4^{th} to 6^{th} week, while the number of common members decreases abruptly during the same period. Afterwards the population of the bronze, silver and gold member groups decreases a little bit, and reaches the second peak value in the 9th or 10th week.

The participants in the community start contributing resources after 3 weeks. Within the first 6 weeks the total number of contributions increases very fast, while the increase rate decreases afterwards. For the total number of ratings, its value increases gradually with a steady increase rate from the 4th week on.

4. Which elements have the most impact on users and can effectively motivate participants? How to optimize the real system using the simulation models and make the community more sustainable?

From all the experiments, we can see that the decay rate and the membership thresholds have a large impact on the contributions and the demographical structure of the Comtella population. Combining the results from all of the previous experiments, a decrease in the decay rate and the bronze member threshold, together with an increase in the gold membership thresholds might result in a better performance.

From previous experiments, a new hypothetical system with the optimal parameters is presented, and I call it as “improved system”. The parameters of the improved system are listed in Table 5.1.

Figure 5.1 presents the system dynamics as well as agent-based simulation results for the improved system. The curves with square dots are the real data from the Comtella database. From the diagrams we can see the percentage of bronze and silver members is improved, which is more efficient to maintain the user participation in a long run.

Table 5.1. Parameters for the improved system

Parameters	Value for the Current System	Value for the Improved System
Population	17	17
Decay Rate	0.7	0.5
Common Member Upgrading Threshold	24	20
Bronze Member Upgrading Threshold	32	32
Silver Member Upgrading Threshold	40	44

For the total number of contributions and ratings, the improved system also has a better performance in most cases.

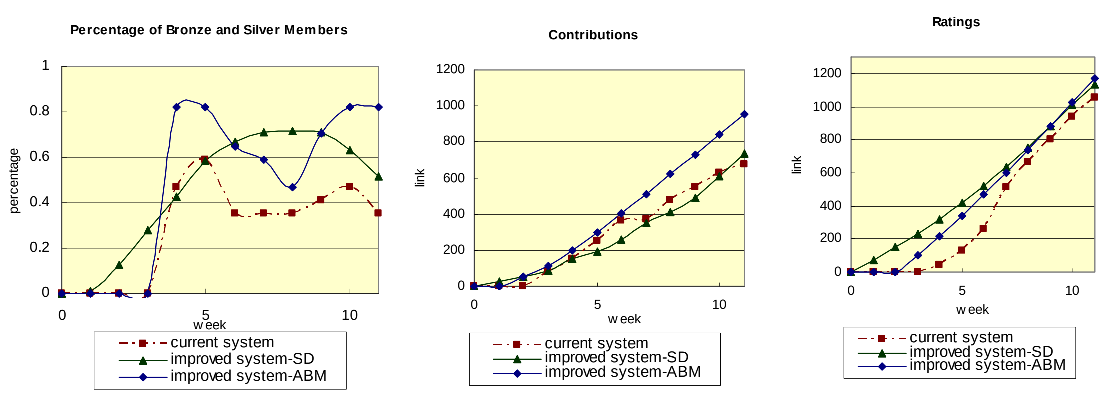


Figure 5.1. Simulation results for the improved system

To summarize, the models presented apply two different simulation approaches to model a small-scale online community called Comtella. Experiments are launched to investigate the under-contribution problem in the community, and optimized system parameters are presented.

Based on my experience in modeling the incentive mechanisms in Comtella, I would compare the relative strengths and weaknesses of the system dynamics approach and the agent-based approach, as shown in Table 5.2.

Table 5.2. Comparisons of the system dynamics approach and the agent-based approach in my study

	System Dynamics Approach	**Agent-based Approach**
Quality of the results in my study	✓	✗
Capability to model anomalies	✓	✗
Requirements on entity-level details	✓	✗
Requirements on system-level details	✓	✗
Capability to do long-scale simulation	✓	✗
Easy to calibration	✓	✗

Both of the system dynamics and the agent-based approach can help us model the Comtella online community and investigate the under-contribution problem. However, in my study the system dynamics models encountered difficulties to model anomalies due to the fixed model structures with controlled level of the simulation and aggregation, while the agent-based model was capable to model well abrupt changes in the system in my case. As a result, the agent-based simulation results are more accurate with respect to the trend line in my study.

On the other hand, since agent-based models focus on the rules of interaction among individual agents, quite a lot of details are needed to calibrate the model, while the system dynamics models only require few details on individual participants and focus more on system level parameters. As a result, the level of the simulation should be controlled carefully in order to avoid introducing extra burden on the computational requirements and the calibration. This confirms the findings of others (Osgood, 2007). When there are not enough data, assumptions are made and reasonable randomness is introduced in the model, which might result in deviations.

My study is based on the assumption that users in the community are mainly motivated by activity points, as mentioned before in page 47. During my study I found that the system dynamics approach is quite useful for systems that have any of following features:

- System has a huge number of entities
- Lack of accurate entity level details
- Need to do long-scale simulation
- System has few emergencies

In contrast, the agent-based approach is useful for systems that have any of following features:

- Lack of system level details but availability of accurate entity level details such as interaction rules
- Need to do short-scale simulation
- System has emergencies

5.2 Future work

Although the simulation and experiment results are quite positive and useful to gain insights in community improvement, there are still some directions that deserve future research:

First, how to improve the quality of resources in the community is still unknown.

No matter how I divided the user group, the question about which group produced higher-quality contributions still has no answer. In Cheng's work (2005), the system was built on the assumption that participants with higher membership levels contribute resources with higher quality. Since we expect resources with high quality instead of low quality, the weekly desired numbers of contributions are also higher for those participants with higher membership levels. However, it is really hard to evaluate whether this assumption is reasonable or not. Also, the quality control part might be inaccurate since a lot of further details are still needed from the experiments in the real world, such as how the individual activity levels change the habits of participants of spending C-points, the impact of C-points on each user group, etc. For this reason, my study in this thesis mainly focuses on the under-contribution problem in the current Comtella community. However, the collaborative rating system still needs further investigation.

Second, the current simulation models need further calibration. In my models, I am looking for a good fit between the simulation results and the real data for both the test group and the control group. One possible approach to further calibrate the model is to use an extra model to calibrate the parameters in the model. This is quite difficult since further details are still needed in the quality control part, which might have a large impact on the output data. Besides, there are too many choices of input parameters to change in calibration and it is also not easy to match all the output data generated from simulations with the real data within the acceptable criteria. As a result, model calibration is also considered to be an important direction of future work.

Third, the agent-based model needs further validation. In my research only the basic validation by comparing the results in the agent-based model with the real data is done. However, it is possible to validate the model through surveys or experiments in the real world which could be a further direction for the future work.

Finally, new ideas on incentive mechanism can be tested using current simulation models. One direction is to use dynamic membership thresholds (Sun 2005). In Sun's work, membership levels are defined dynamically by the largest gap among weekly contributions. A classification algorithm is used in the system in order to spread out the participants as much as possible. As a result, the gold members contribute more than most of other participants, and this highest membership is an exclusive class, while the other memberships are more accessible. Another direction is to use different percentage of weekly contributions as thresholds instead of fixed value of activity points. Besides, it is also possible to investigate how to generalize these models for open systems. In the current Comtella community the number of participants is fixed throughout the experiment. However for open online communities, attracting new participants is a vital aspect to investigate.

A hybrid model might also be considered into the future work, where some of the interaction rules of agents can be defined in a system dynamics way. Higher versions of AnyLogic should be appropriate for developing this hybrid model, and there might be possibilities to get some interesting results.

REFERENCES

Adar, E., & Huberman, B.A. (2000). Free Riding on Gnutella. *First Monday* 5(10), 2000.

Akkermans, H. A. (2001). Emergent Supply Networks: System Dynamics Simulation of Adaptive Supply Agent. *Proceedings of the 34th Hawaii International Conference on System Sciences*, Hawaii, 2001.

Beenen, G. et al. (2004). Using Social Psychology to Motivate Contributions to Online Communities. *Proceedings of CSCW'04*, Chicago, Illinois, Nov. 6–10, 2004.

Blanchard, A. (2004). Blogs as Virtual Communities: Identifying a Sense of Community, University of North Carolina at Charlotte (last access on Jan 9th, 2008). http://blog.lib.umn.edu/blogosphere/blogs_as_virtual.html

Bonabeau, E. (2002).Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. *Proceedings of the National Academy of Sciences (PNAS)*, vol. 99, Suppl. 3, 7280-7287, May 14, 2002.

Bretzke, H., & Vassileva, J. (2003). Motivating Cooperation in Peer to Peer Networks. *Proceedings User Modeling UM03*, Johnstown, PA, Springer Verlag LNCS, 2003.

Brook, C. & Oliver, R. (2003). Online Learning Communities: Investigating a Design Framework. Australian *Journal of Educational Technology*, 2003, 19(2), 139-160.

Cheng, R. (2005). User Motivation Mechanism for Building Sustained Online Communities. Master's thesis, University of Saskatchewan, Sep 2005 (last access on Jan 9th, 2008). http://library.usask.ca/theses/available/etd-09262005-014650/unrestricted/Thesis-RanCheng.pdf

Cheng, R. & Vassileva, J. (2005). User Motivation and Persuasion Strategy for Peer-to-peer Communities. *Proceedings of HICSS'2005 (Mini-track on Online Communities in the Digital Economy/Emerging Technologies)*, Hawaii, 2005.

Davidsson, P. (2002). Agent Based Social Simulation: A Computer Science View. *Journal of Artificial Societies and Social Simulation*, vol.5, no.1, 2002.

De Moor, A. (2005).Growing Pains: Community Modeling with System Dynamics (last access on Jan 9th, 2008). http://growingpains.blogs.com/home/2005/10/community_model.html.

De Moor, A. (2005). Towards a Design Theory for Community Information Systems. *Proceedings of the 11th International Conference on Human-Computer Interaction (HCII 2005)*, Las Vegas, July 2005. Lawrence Erlbaum Ass., Mahwah, NY.

Diker, V.G. (2004). A Dynamic Feedback Framework for Studying Growth Policies in Open Online Collaboration Communities. *Proceedings of AMCIS 2004*, NY, 2004.

Dunbar, R. (1996). *Grooming, Gossip, and the Evolution of Language*. Cambridge (Mass.), Harvard University Press, 1996.

Fabio L. B. et al. (2003). JADE - A WHITE PAPER (last access on Jan 9th, 2008). http://jade.tilab.com/papers/2003/WhitePaperJADEEXP.pdf.

Forrester, J. (1958). Industrial Dynamics: A Major Breakthrough for Decision Makers. *Harvard Business Review*, Vol. 36, No. 4, 37-66.

Forrester, J. (1961). *Industrial Dynamics*. MIT Press, Cambridge, MA, 1961.

Harris, B. & Williams, B. (2005). System Dynamics Methodology. W K Kellogg Foundation (last access on Jan 9th, 2008). http://users.actrix.co.nz/bobwill/AESSD.pdf

Homer, J. et al. (2004). The CDC's Diabetes Systems Modeling Project: Developing a New Tool for Chronic Disease Prevention and Control. *Proceedings of the 22nd International Conference of the System Dynamics Society*, July 25-29, Oxford, England, 2004.

Jensen, J., Devis, J, & Farnham, S. (2002). Finding Others Online: Reputation Systems for Social Online Spaces. *Proceedings of CHI 2002*, Minneapolis, Minnesota, 2002.

Jones, Q., Ravid, G., & Rafaeli, S. (2002). An Empirical Exploration of Mass Interaction System Dynamics: Individual Information Overload and Usenet Discourse.

Proceedings of HICSS'02, Hawaii, 2002.

Kelly, S.U., Sung, C., & Farnham, S. (2002). Designing for Improved Social Responsibility, User Participation and Content in Online Communities. *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, Minneapolis, Minnesota, USA, 2002.

Kollock, P., & Smith, M. (1996). Managing the Virtual Commons: Cooperation and Conflict in Computer Communities. In S. Herring (Ed.), *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives* (pp. 109-128). John Benjamin: Amsterdam, 1996.

Krichmar, D. M., & Preece, J. (2005). A Multilevel Analysis of Sociability, Usability, and Community Dynamics in an Online Health Community. *ACM Transactions on Computer-Human Interaction (TOCHI)*, v.12 n.2, pp.201-232, June 2005.

Leimeister, J.M., Sidiras, P., & Krcmar, H. (2004). Success Factors of Virtual Communities from the Perspective of Members and Operators: An empirical study. *Proceedings of the 37th Hawaii International Conference on System Sciences*, Hawaii, 2004.

Ludford, P. J., Cosley, D., Frankowski, D., & Terveen, L. (2004). Think Different: Increasing Online Community Participation Using Uniqueness and Group Dissimilarity. *Proceedings of the 2004 conference on Human factors in computing systems*, pp.631-638, Vienna, Austria, 2004.

Mao, Y., Vassileva, J. & Grassmann, W. (2007). A System Dynamics Approach to Study Virtual Communities, in *Proceedings IEEE HICSS'07 mini-track on Virtual Communities*, page 178a, Big Island, Hawaii, 2007.

Millen, D. R., & Patterson, J. F. (2002). Stimulating Social Engagement in a Community Network. *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, Louisiana, USA, 2002.

Mockus, A., Fielding, R.T., & Andersen, H. (2002). Two Case Studies of Open Source Software Development: Apache and Mozilla. *ACM Transactions on Software*

Engineering and Methodology, 11(3): pp. 309-346, 2002.

Nonnecke, B. and Preece, J. (2000). Lurker demographics: Counting the Silent, *Proceeding of CHI 2000*, Hague, Netherlands, 2000.

Osgood, N. (2007). Using Traditional and Agent-Based Toolsets for System Dynamics. *Proceedings of the 2007 International Conference of the System Dynamics Society and 50th Anniversay Celebration*, Boston, Massachusetts, 2007.

Parunak, H. Van Dyke, Savit, R, & Rick, L. Riolo (1998): Agent-Based Modeling vs. Equation-Based Modeling: A Case Study and User Guide. *Proceedings of Multi-agent systems and Agent-based Simulation*, Paris, 1998.

Pavlov, O.V. & K. Saeed (2004). A Resource-Based Analysis of Peer-to-Peer Technology. *System Dynamics Review*, Vol. 20, No. 3, pp. 237-262, 2004.

Pidd, M. (1993). *Computer Simulation in Management Science* (3rd Edition). John Wiley, Chichester, 1993.

Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*. Addison-Wesley, Reading, MA, 1993.

Richardson, G. P. (1991). *Feedback Thought in Social Science and System Theory*. University of Pennsylvania Press, Philadelphia, 1991.

Schuler, D. (1996). *New Community Networks: Wired for Change*. ACM Press, New York, 1996.

Simon, H. (1957). *Models of Man*. Wiley, New York, 1957, p.198.

Shirky, C. (2002). "Communities, Audiences, and Scale", at website of "Clay Shirky's Writings About the Internet: Networks, Economics, and Culture" (Created on April 6th, 2002 and last access on Jan 9th, 2008): http://shirky.com/writings/community_scale.html

Sterman, J.D. (1992). System Dynamics modeling for Project Management (last access on Jan 9th, 2008). http://web.mit.edu/jsterman/www/SDG/project.pdf

Sterman, J.D. (2000). *Business Dynamics: Systems Thinking and Modeling for a*

Complex World, Irwin McGraw-Hill, London, 2000.

Sun, L. (2005). Motivational Visualization for Resources-Sharing Online Communities. Master's thesis, University of Saskatchewan, May 2005 (last access on Jan 9th, 2008). http://library2.usask.ca/theses/available/etd-06132005-134147/unrestricted/Thesis-final.pdf

Thompson, L.F., Meriac, J.P., & Cope, J. (2002). Motivating Online Performance: The influences of goal setting and Internet self-efficacy. *Social Science Computer Review*, 20 (2), pp.149-160, 2002.

Ting, N.S. and Deters, R. (2003). 3LS—A Peer-to-Peer Network Simulator. *Proceedings of the 3rd International Conference on Peer-to-Peer Computing*, Linköping, Sweden, 2003.

Upadrashta, Y., Vassileva J., Grassmann, W. (2005). Social Networks in Peer-to-Peer Systems. *Proceedings of HICSS'2005 (Mini-track on The Peer to Peer Paradigm/Emerging Technologies)*, Hawaii, 2005.

Vassileva J. (2002). Supporting Peer-to-Peer User Communities. R. Meersman, Z. Tari et al. (Eds.) "*On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*" *Coordinated International Conferences Proceedings*, Irvine, 29 Oct - 1 Nov. 2002, Springer Verlag LNCS 2519, Springer Verlag: Berlin-Heidelberg, 230-247.

Vassileva, J. (2005). Adaptive Incentive Mechanism for Sustainable Online Community. *Proceedings of Workshop* "*Sustaining Community: The role and design of incentive mechanisms in online systems*" *at ACM Group 2005 Conference*, Florida, USA, November 2005.

Wakeland, W. W. et al. (2004). A Comparison of System Dynamics and Agent-based Simulation Applied to the Study of Cellular Receptor Dynamics. *Proceedings of the 37th Hawaii International Conference on System Sciences*, Hawaii, 2004.

Webster, A.S. & Vassileva, J. (2006). Visualizing Personal Relations in Online Communities. *Proceedings of Adaptive Hypermedia and Adaptive Web-Based*

Systems, Dublin, Ireland, 2006.

Whittaker, S., Issacs, E., & O'Day, V. (1997). Widening the Net: Workshop Report on the Theory and Practice of Physical and Network Communities. *SIGCHI Bulletin*, 29(3), p. 137.

Wikipedia (2007). Definition of online community (last access on Jan 9th, 2008): http://en.wikipedia.org/wiki/Online_communities.

XJ Technologies (2002). AnyLogic 4.0 User's Manual (last access on Jan 9th, 2008). http://www.bluechip-india.com/AnyLogic-Overview.pdf.

Yamada, K., Nakakoji, K., & Ueda, K. (2004). A Multi-agent Systems Approach to Analyze Online Community Activities. *Proceedings of ICAM'04*, 2004.

Zhang, Y.W., & Tanniru, M. (2005). An Agent-based Approach to Study Virtual Learning Communities. *Proceedings of the 38th Hawaii International Conference on System Science (HICSS'05)*, Hawaii, 2005.

APPENDIX A

CONSENT FORM

Consent Form used in CMPT 490 (2003/2004)

Title of the Study:

Impact of Persuasion Techniques and Community Visualization on User Motivation to Contribute in the Peer-to-Peer Resource Sharing Environment COMTELLA

Researchers:

Julita Vassileva, Associate Professor, Computer Science Department; 966-2073

Lingling Sun, M.Sc. Student, Computer Science Department

Ran Cheng, M.Sc. Student, Computer Science Department

The aim of this study is to investigate the usability and the motivational effects of different persuasion techniques and community visualization on participation rates in a peer-to-peer resource-sharing environment. This is accomplished through experimenting with a peer-to-peer file-sharing program (Comtella) in a 4^{th} year undergraduate students. Comtella allows a limited number of users to share links to papers that they have found with other users of the network. Depending on the level of contribution and participation different persuasion techniques will be used, including user status (level of membership) and community visualization, which is both informative and motivational. The visualization will allow seeing the other peers currently on-line, their level of contribution, areas of interest, the number of papers they are contributing and summarizing.

Potential Benefits:

The possible benefit to the participants will be a more convenient access to class resources (papers, found on the Web). It will allow the users to make use of the search results of their peers, which can lead to a synergy in the class efforts to stay current in their knowledge of the literature.

Potential Risks:

It is hard to envisage any risks or side effects of the usage of the system. The papers that are shared are publicly available on the web, and we don't envisage copyright issues. If we become aware of any negative effects during the study, we will inform immediately the participants or interrupt the study. We may discontinue a participant's involvement in the study, if they use the environment to communicate links offensive, copyrighted or inappropriate materials. In the event of a participant withdrawing from the study, his / her data will be deleted and destroyed insofar as possible.

Collection and Storage of Data:

During the study, data will be collected about the users' actions related to accessing the different views of the community visualization, for example the view of who is currently on-line, or the view of who shares papers in a particular area, or the view of who has contributed most new articles so far. This data will be correlated in anonymized form with user-participation data, in terms of number of new links found, links downloaded from other peers, number of rated and summarized papers.

All data will be stored anonymously and will be available only to the investigators involved in the study. All data about the users will be stored securely for a minimum of five years, electronic data on a password-protected computer system, and any additional on-paper data by Dr. Vassileva.

Confidentiality:

The anonymity of the collected data and the privacy of the subjects will be completely protected, and the information obtained from this data would be used only in theses, journal articles or conference publications written by the researchers. Only aggregate data will be reported in publications; the names and identities of the participants will not be published in any form.

Right to Withdraw:

Each participant is free to withdraw from the study at any time; this will not affect the participant's academic status or success in the class. In the event of a participant withdrawing from the study, his / her data will be deleted. An alternative way of participating in the class and doing the coursework through the Department E-Handin system exists and the students are free to choose this way, if they prefer so, at any point of the study. The participants will be advised of any new information that may have a bearing on the participants' decision to continue in the study.

Questions:

If you have any questions concerning the study, please feel free to ask at any point; you are also free to contact the researchers at the numbers provided above if you have questions at a later time. Any questions regarding your rights as a participant may be addressed to the Behavioural Research Ethics Board through the Office of Research Services (966-2084), or through Julita Vassileva (966-2073). If you wish to acquire information on the results of the research once the study is completed, send a request to Julita Vassileva at jiv@cs.usask.ca.

Consent to Participate:

I have read and understood the description provided above; I have been provided with an opportunity to ask questions and my questions have been answered satisfactorily. I consent to participate in the study described above, understanding that I may withdraw this consent at any time. A copy of this consent form has been given to me for my records.

Date

(Signature of Participant)

(Signature of Researcher)

APPENDIX B

COMTELLA DATABASE STRUCTURE

All the tables in the Comtella database are created from a SQL file. Here is a list of all the tables:

Table Name: accessrecord
Fields:

- messageid [int(11)]	Default =0
- uid [int(11)]	Default = 0
- accesstime [datetime]	Default = 0000-00-00 00:00:00
- sectionid [int(11)]	Default = 0

Table Name: actiondd
Fields:

- actionid [int(11)]	Primary key, auto increment
- description [text]	

Table Name: actiontrack
Fields:

- uid [int(11)]	Default =0
- starttime [datetime]	Default = 0000-00-00 00:00:00
- endtime [datetime]	Default = 0000-00-00 00:00:00
- actionid [int(11)]	Default = 0
- objective [int(11)]	Null = yes

Table Name: categorytable
Fields:

- categoryid [int(11)]	Primary Key, auto increment
- category [text]	null = yes
- expectedsum [int(11)]	Default = 0
- deadline [datetime]	Default = 0000-00-00 00:00:00
- sumdeadline [datetime]	Null = yes

Table Name: commentinfo

Fields:

- messageid [int(11)] Primary key, auto increment
- msgid [int(11)] Null = yes
- parentid [int(11)] Null = yes
- title [text]
- comment [text] Null = yes
- uid [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- sectionid [int(11)] Default = 0

Table Name: commentrating

Fields:

- messageid [int(11)] Primary key Default = 0
- uid [int(11)] Primary Key Default = 0
- rating [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00

Table Name: cpoint

Fields:

- uid [int(11)] Primary Key Default = 0
- expiredate [datetime] Primary Key
  Default = 0000-00-00 00:00:00
- amount [int(11)] Default = 0
- messageid [int(11)] Null = yes

Table Name: discussion

Fields:

- discussionid [int(11)] Primary key, auto increment
- title [text]
- content [text] Null = yes
- uid [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- categoryid [int(11)] Null = yes
- level [int(11)] Null = yes
- parentid [int(11)] Null = yes

Table Name: fakeifno4comment

Fields:

- messageid [int(11)] Primary Key Default = 0
- uid [int(11)] Primary Key Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- reason [text] Null = yes

Table Name: fakeinfo4sharing

Fields:

- messageid [int(11)] Primary Key Default = O
- uid [int(11)] Primary Key Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- reason [text] Null = yes

Table Name: invitation

Fields:

- messageid [int(11)] Default = 0
- inviteduid [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- message [text] Null = yes

Table Name: leveldd

Fields:

- levelid [int(11)] Primary Key, auto increment
- description [text]

Table Name: news

Fields:

- newsid [int(11] Primary Key, auto increment
- title [text] Null = yes
- content [text] Null = yes
- submitdate [datetime] Default = 0000-00-00 00:00:00
- uid [int(11)] Default = 0

Table Name: rating4url

Fields:

- messageid [int(11)] Primary Key Default = 0
- uid [int(11)] Primary Key Default = 0
- rating [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00

Table Name: remark

Fields:

- remarkid [int(11)] Primary Key, auto increment
- content [text]
- uid [int(11)] Default = 0
- messageid [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00

Table Name: sharedfileinfo

Fields:

- messageid [int(11)] Primary Key, auto increment
- uid [int(11)] Default = 0
- sharedurl [text]
- title [text]
- avgrating [double] Default = 0
- ratingcount [int(11)] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- lastupdate [datetime] Default = 0000-00-00 00:00:00
- categoryid [int(11)] Default = 0
- cpointamount [int(11)] Default = 0
- changecpoint [int(11)] Default = 0

Table Name: statusinfo

Fields:

- uid [int(11)] Primary Key Default = 0
- categoryid [int(11)] Primary Key Default = 0
- userlever [int(11)] Null = yes
- leftrating [int(11)] Null = yes
- expectednum [int(11)] Null = yes
- paperquan [int(11)] Null = yes
- paperquancr [double] Null = yes
- paperqual [double] Null = yes
- paperqualcr [double] Null = yes
- ratingquan [int(11)] Null = yes
- ratingquancr [double] Null = yes
- ratingqual [double] Null = yes
- ratingqualcr [double] Null = yes
- overalcr [double] Null = yes

Table Name: summaryinfo

Fields:

- summaryid [int(11)] Primary Key, auto increment
- uid [int(11)] Default = 0
- messageid [int(110] Default = 0
- submitdate [datetime] Default = 0000-00-00 00:00:00
- title [text]
- summary [text]
- approved [int(11)] Default = 0
- finalmark [float] Null = yes
- ta_comment [text] Null = yes

Table Name: summaryreview

Fields:

- summaryid [int(11)] Primary Key Default = 0

- uid [int(11)] Primary Key Default = 0
- rating [int(11)] Default = 0
- review [text]
- isfake [int(11)] Null = yes
- submitdate [datetime] Default = 0000-00-00 00:00:00

Table Name: userinfo

Fields:

- uid [int(11)] Primary Key Default = 0
- nsid [text]
- password [text]
- displayname [text]
- reputation [float] Default = 0
- privilege [int(11)] Default = 0
- class [int(11)] Null = yes
- paperrep [double] Null = yes
- ratingrep [double] Null = yes

APPENDIX C

FORMULA DERIVATION OF THE VARIABLE ‘INCETIVE RATE’ IN THE SYSTEM DYNAMICS MODEL

In the system dynamics model for extended version of the incentive mechanism, variable “*Incentive Rate*” was given by formula:

$$\begin{aligned}&Incentive\ Rate = \\ &\frac{new\ GM\ rating\ rewards\text{-}decay\ rate\ for\ GM + Awards\ for\ GM}{\left(Awards\ for\ GM\text{-}Reward\ Unit\ for\ Sharing*Expected\ Sharing\ Rewards\ for\ GM*Percentage\ for\ GM\right)}\end{aligned}$$

The basic idea is that the value of “*Incentive Rate*” should be the number of activity points of gold members in the current week divided by the number of activity points they got from the previous week:

$$\begin{aligned}IncentiveRate &= \frac{current\ awards}{awards\ got\ in\ last\ week} = \frac{Awards(t+1)}{Awards(t)} \\ &= \frac{Awards(t+1)}{Awards\ for\ GM}\end{aligned} \tag{C.1}$$

To simplify, I use several abbreviations in the following equations:

$$\begin{aligned}POR &= \frac{1\text{-}Comtella\ Gold\ Members}{Population} \\ EXP &= Expected\ Sharing\ Rewards\ for\ GM \\ RUS &= Reward\ Unit\ for\ Sharing \\ ACT &= Activeness\ Level\ of\ Gold\ Member\end{aligned} \tag{C.2}$$

Taking Equation 3.14 (page 42) and Equation 3.17 (page 43) into consideration, the following results can be obtained:

$$\begin{aligned}
&ACT = Activeness\ Level\ of\ Gold\ Member = POR * Incentive\ Rate \\
&Share\ Rate\ of\ GM = EXP * ACT = EXP * POR * Incentive\ Rate \\
&new\ awards\ by\ sharing = RUS * Share\ Rate\ of\ GM \\
&\qquad\qquad = RUS * EXP * POR * Incentive\ Rate
\end{aligned} \tag{C.3}$$

Using Equation C.2 and Equation C.3, Equation C.1 can be derived as:

$$\begin{aligned}
Incentive\ Rate &= \frac{Awards(t+1)}{Awards\ for\ GM} \\
&= \frac{Awards\ for\ GM + new\ awards}{Awards\ for\ GM} \\
&= \frac{\begin{pmatrix} Awards\ for\ GM + new\ awards\ by\ sharing \\ + new\ GM\ rating\ rewards - decay\ rate\ for\ GM \end{pmatrix}}{Awards\ for\ GM} \\
&= 1 + \frac{RUS*EXP*POR*Incentive\ Rate}{Awards\ for\ GM} \\
&\quad + \frac{new\ GM\ rating\ rewrds\text{-}decay\ rate\ for\ GM}{Awards\ for\ GM}
\end{aligned} \tag{C.4}$$

As a result, by moving the variable "*Incentive Rate*" in Equation C.4 into one side, we get the following equation, which is same as Equation 3.15:

$$\begin{aligned}
Incentive\ Rate &= \frac{\left(1 + \frac{new\ GM\ rating\ rewrds\text{-}decay\ rate\ for\ GM}{Awards\ for\ GM}\right)}{\left(1 - \frac{RUS*EXP*POR}{Awards\ for\ GM}\right)} \\
&= \frac{Awards\ for\ GM + new\ GM\ rating\ rewards\text{-}decay\ rate\ for\ GM}{Awards\ for\ GM\text{-}RUS*EXP*POR} \\
&= \frac{Awards\ for\ GM + new\ GM\ rating\ rewards\text{-}decay\ rate\ for\ GM}{\begin{pmatrix} Awards\ for\ GM\text{-}Reward\ Unit\ for\ Sharing \\ *Expected\ Sharing\ Rewards\ for\ GM*Percentage\ for\ GM \end{pmatrix}}
\end{aligned} \tag{C.5}$$

APPENDIX D

PARAMETER LIST OF THE SYSTEM DYANMICS MODEL

Parameters	Unit	Value	Comment
Initial Population	Person	17	Comtella population
Reward Unit for Sharing	Point/Link	4	The number of activity points that are rewarded for contributions
Reward Unit for Rating	Point/Link	3	The number of activity points that are rewarded for ratings
C-point rewards for rating	Point/Link	3	The number of C-points that are rewarded for ratings
Initial Cpoints Assigned	Point/Person	20	The number of C-points that are rewarded once at the very beginning of the period
Rewards Decay Percentage	1/Week	0.7	The proportion of activity points that participants will lose at the beginning of each week
C-points Decay Percentage	1/Week	0.5	The proportion of C-points that participants will lose at the beginning of each week
Possibility to assign C-points	Person*Week/Link	0.458026	The probability for participants to assign C-points
Average C-	Point/Person/Week	5.618	The average number of C-

points assigned			points that are assigned to one contribution
Common Member Upgrading Threshold	Point/Person	24	The number of activity points that are needed for common members to upgrade their membership level
Bronze Member Upgrading Threshold	Point/Person	32	The number of activity points that are needed for bronze members to upgrade their membership level
Silver Member Upgrading Threshold	Point/Person	40	The number of activity points that are needed for silver members to upgrade their membership level
Sensitivity for Common Member	1/Week	0.13	The sensitivity of common members on membership upgrading
Sensitivity for Bronze Member	1/Week	0.16	The sensitivity of bronze members on membership upgrading
Sensitivity for Silver Member	1/Week	0.21	The sensitivity of silver members on membership upgrading
Decay rate of Bronze Member	1/Week	0.25	The proportion of bronze members whose membership levels are degraded to common member
Decay rate of Silver Member	1/Week	0.075	The proportion of silver members whose membership levels are degraded to bronze member
Decay rate of Gold Member	1/Week	0.125	The proportion of gold members whose membership levels are degraded to silver member

APPENDIX E

MEMBERSHIP CALCULATIONS IN THE AGENT-BASED MODEL

As mentioned before in Chapter 4, the adaptive reward mechanism was developed by Cheng (2005). In his design, the value of the activity points (or rewards) for each participant is determined by four factors: the quantity and quality of the contributions, as well as the quantity and quality of ratings given by the participant:

$$\begin{aligned} Total\ number\ of\ awards = \quad & PaperQuanCr + PaperQualCr \\ & + RatingQuanCr + RatingQualCr \end{aligned} \tag{D.1}$$

The variables A_i, B_i, and Di stand for the total number of resources shared by the current participant, the total number of ratings earned by the current participant, as well as the total number of ratings given by the current participant respectively, where the subscript of these three variables represents the index of the participant and varies from 1 to N, where N is the Comtella population. The four parts of the participant's rewards can be calculated as follows:

The rewards for the quantity of the contributions (*PaperQuanCr*) are given by:

$$\begin{aligned} PaperQuanCr &= W_S * A_i \\ &= W_{S0} * T * F_i * A_i \end{aligned} \tag{D.2}$$

Here, the value of W_{S0} equals 4. The time-function factor T and the over-limit factor F_i are defined as follows and the parameters are defined by Cheng (2005):

$$T = \sqrt{(-0.0114 * Index\ of\ the\ time\ period + 0.817)} + 0.102 \tag{D.3}$$

$$F_i = \begin{cases} 1 & when\ x \leq Q_i \\ \left(\frac{1}{4}\right)^{(x-Q_i)} & when\ x > Q_i \end{cases} \tag{D.4}$$

For equation D.3, the "*Index of the time period*" is defined by Equation D.5. In the current Comtella system, each day is divided into 10 time periods. As a result, one week has 70 time periods. Given one specific time, the index of the time periods can be calculated by:

$$Index\ of\ the\ time\ period = Index\ of\ the\ day * 10 + Index\ of\ the\ time \tag{D.5}$$

In the above equation, "*Index of the day*" and "*Index of the time*" are defined in Table 4.2.

For Equation D.4, variable x is the current total number of resources shared by the participant, and the personal desired number of shared resources Q_i is defined by:

$$Q_i = Q * \frac{C_{ij}}{\sum_{k=1}^{N} C_{kj}} \tag{D.6}$$

In the above equation, variable Q is a community desired quantity of shared resources, and the variable C_{ij} is the paper reputation of current participant i in a particular week j. Using N as the Comtella population, the first subscript k varies from 1 to N, and $\sum_{k=1}^{N} C_{kj}$ calculates the sum of the paper reputation of all the participants in the current week j. The calculations of variable Q and C_{ij} are shown here:

$$Q = 4 * N$$

$$C_{ij} = \begin{cases} 1 & j = 1 \\ 1 + \dfrac{\sum_{k=1}^{j-1} C_{ik}}{j-1} & j > 1 \end{cases} \tag{D.7}$$

Here, the subscript i is the index of current participant, and the subscript j is the index of current week. Thus, $\sum_{k=1}^{j-1} C_{ik}$ calculates the sum of the paper reputation of current participant i in all the previous weeks.

For example, the paper reputation of participant i in the 3rd week is calculated as $C_{i3} = 1 + \dfrac{C_{i1} + C_{i2}}{2}$, where in previous weeks $C_{i2} = 1 + C_{i1}$ and $C_{i1} = 1$.

The rewards for the quality of the contributions (*PaperQualCr*) are given by:

$$PaperQualCr = PaperQuanCr * AverageRating * W \tag{D.8}$$

Here,

$$AverageRatings = \frac{Total \quad number \quad of \quad ratings \quad earned}{Total \quad number \quad of \quad resources \quad shared} = \frac{B_i}{A_i} \tag{D.9}$$

The variable W in Equation D.8 is a constant for different periods of time.

$$W = \begin{cases} 2/3 & for\ Week\ 1\text{-}3 \\ 3/4 & for\ Week\ 4 \\ 1 & for\ Week\ 5-11 \end{cases} \tag{D.10}$$

The rewards for the quantity of ratings given by the participant (*RatingQuanCr*) are given by:

$$RatingQuanCr = W_R * D_i \tag{D.11}$$

Here the value of W_R equals 3.

Last, the rewards for the quality of ratings given by the participant (*RatingQualCr*) are given by:

$$RatingQualCr = \sum_{j=1}^{D_i} (1.5 * (1 - \left| r_{ij} - \overline{r_j} \right|)) \tag{D.12}$$

Here D_i is the total number of ratings given by the participant *i*, and $r_{ij} (i = 1,2,3,..., D_i)$ represents each rating given by the participant *i* for resource *j*. The subscript *j* represents the index of the ratings given by the participant *i*, and varies from 1 to D_i. The variable $\overline{r_j} (j = 1,2,3,..., D_i)$ represents the average rating for the resource *j* obtained from all users who have rated this resource.