

Simulation on Knowledge Diffusion in Open Source Software Community Using Multi-Agent Modeling

Yang Hu^{1,*}, Xiao-dong Zhang¹, Cai-juan Li² and Chen-jian Gong¹

¹Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China; ²Hebei Institute of Architecture and Civil Engineering, Zhangjiakou, China

Abstract: In order to improve the knowledge diffusion effect in Open Source Software Community (OSSC), we built a multi-agent-based model to simulate the whole process of knowledge diffusion in OSSC. First, by analysing real OSSC, an open source software community model was built. Second, mechanisms such as knowledge diffusion mechanism, partnership selection mechanism, knowledge learning mechanism and knowledge spillovers mechanism and sever collaboration mechanism were given. Further, by using multi-agent modeling tool and MATLAB software, the current research modeled knowledge diffusion in OSSC and investigated the effect of members' knowledge distribution on the performance of knowledge diffusion. Results show that higher initial knowledge variance in OSSC knowledge diffusion process would lead to higher knowledge efficiency, fewer knowledge diffusion path, and lower equity.

Keywords: Knowledge diffusion, multi-agent, simulation, open source software community.

1. INTRODUCTION

With the development of internet, an internet-based corporate developing community which is called open source software community (OSSC) is more and more popular and has received enormous attention in the last several years [1]. According to traditional developing organization, OSSC shows advantages such as low cost, high production and high innovative during the developing process. As a result, there are lots of successful OSSC applications such as Mozilla Firefox, Linux and Apache, which are proved to be in good robustness and applicability [2-4]. In OSSC, thousands of individuals with various kinds of knowledge backgrounds shows complex behaviors such as join, leave, collaborate and develop, which lead to a complex knowledge diffusion path. On the one hand, participants promote the software to evolve with knowledge diffusion. On the other hand, knowledge diffusion makes participants develop software easier and faster. Thus, knowledge plays an important role in the OSSC and to find how to promote the knowledge diffusion effect in the community is necessary and helpful for management. However, according to traditional software developing organization, the knowledge diffusion process is significant different in OSSC, which is in urgent need to carry out in-depth study.

The rest of the paper is organized as follows: section 2 reviews the related papers of knowledge diffusion models. Section 3 presented the model and methodology. Section 4 explores the factor that affects the diffusion with several experiments, and discusses the empirical result. Finally, Section 5 makes conclusions of this study.

2. LITERATURE REVIEW

Traditional studies use two major models to explore knowledge diffusion. The first is general model, which mainly contains Bass model and its derivative model [5-7]. These models have mostly focused on market's aggregate variables like market penetration and advertising campaigns [8], or study how mixed strategy affect the diffusion of new products [9]. General model has just keep focus on the process of knowledge diffusion in macro level, while ignores the fact that participants' adoption will be affected by each other and impress system's knowledge diffusion. The second major model for knowledge diffusion is micro-level model. This model is created to study the micro-level characteristics, like observability of innovation, compatibility of innovation, trialability of innovation and relative advantage [10-12], which may affect the adoption of consumer. However, the micro-level model has remedied general model's limitation, although it is not adequate for explaining the macro phenomena of diffusion.

In recent years, with the development of social network research, literatures on knowledge diffusion are normally combined with social network and multi-agent modeling [13-15]. Such as Cowan [13] propose a multi-agent based model which indicates agent's knowledge exchange through the link of the social network. Lin [16] extend Cowan's model to another four typical networks and find that knowledge diffuse fast in small world network. In another side, Kim [15] applies the scale-free property to the Cowan's model and confirms that the small-world network is the most efficient structure toward knowledge diffusion. By using a single-layer perceptron neural network model (SLPM), Tang [17] demonstrates that high-level knowledge holder plays a paramount role in knowledge transfer under different organizational structures. However, all of them focus only on the

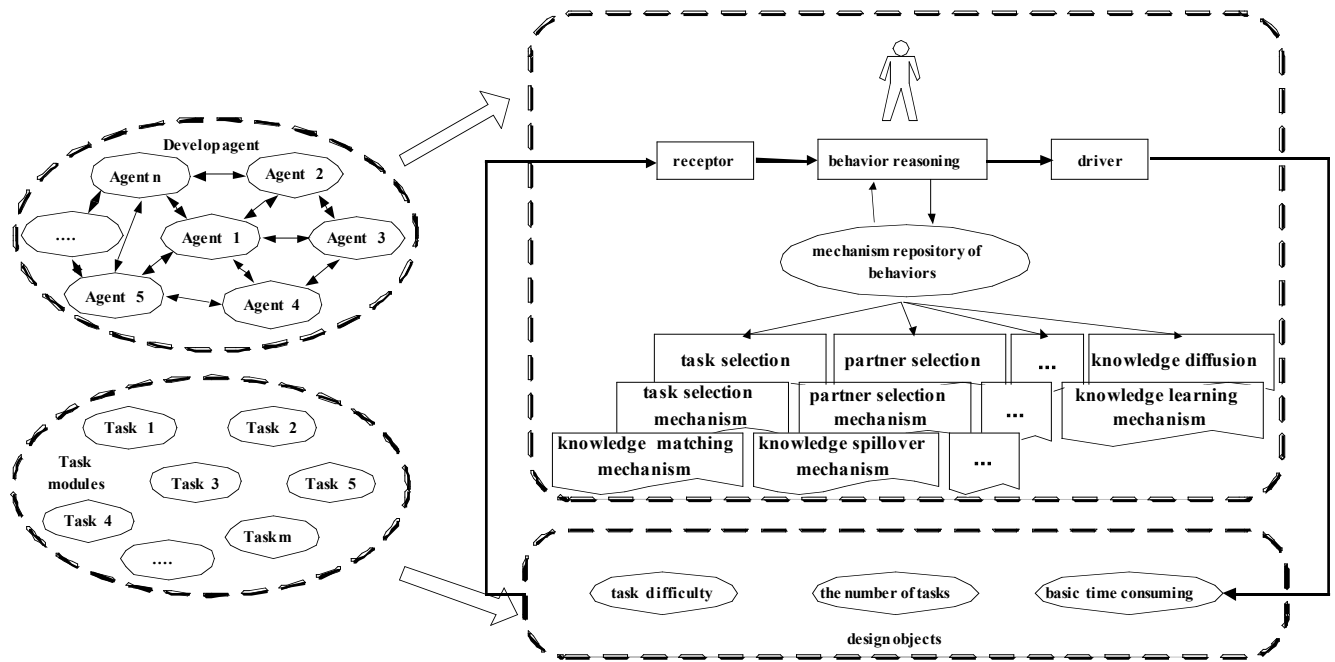


Fig. (1). Open source software community model.

influence of different network structures which originated from reconnect random probability. In another words, their research carrier is complex network-only, and merely consider the impact of agent’s behaviour and the distribution of members’ knowledge. Thus, the current research tries to apply the agent’s behaviour to the knowledge diffusion model which is based on the operation of an open source software community.

3. MODEL

In OSSC, a software project can be divided into small spices, such as task1, task2...taskm in Fingure 1. In the community, develop agent continuously chooses, collaborate and exchange knowledge to complete tasks until all modules are completed. The multi-agent simulation model is build based on above principle. As shown in Fig. (1).

The multi-agent simulation model describes the process of OSSC, which contains n develop agent and m task modules. In the model, develop agents receive information of software tasks by receptor firstly, and then analyse the received information and solution with using behavior mechanism repository. Finally, solution applied to tasks through driver and tasks’ condition changed. As a result, collaborative behavior, autonomous behavior, knowledge diffusion phenomenon, as well as the knowledge diffusion network which is produced by collaboration of tasks can be described by this model.

3.1. Develop Agent and Attributes of Task

Assuming the knowledge develop agent required to complete tasks can be divided into two categories: x and y. The knowledge attribute of agent can not exactly satisfy the knowledge level needed to complete the task. So we can constitute a knowledge space Ω of non-periodic two-

dimensional structure, which horizontal and vertical coordinates are x and y. The knowledge required to complete the task and the knowledge of develop agent are exogenously given and do not change over time. They are expressed as follows: module knowledge $k_m = [K_x^m, K_y^m]$ and agent knowledge $k_i = [k_{i,x}, k_{i,y}]$. Knowledge determines the coordinates of agent in the knowledge space, namely $x_i = k_{i,x}, y_i = k_{i,y}$. If the knowledge of agent is inability to complete the task in the develop process, agent needs to seek collaboration so that to complete task.

In the develop process, collaboration relationship can be seen as knowledge-based collaboration network. Thus $N = \{1, 2, 3...n\}$ means the collection of all the nodes and edge ij means agent i and j cooperate to complete the task successfully. Then knowledge collaborative networks can be expressed as $G, ij \in G$.

3.2. Collaboration Mechanism

When develop agent contributes to tasks, their behaviors follow these mechanisms:

- a) Collaboration occurs under the conditions: Design agent’s knowledge attribute can’t meet the task requirements.
- b) Partnership selection mechanism: In the network evolution time t , partner selection follows this rule:

$$P_{ij}^t = \frac{R_{ij}^t}{R_i^t} \tag{1}$$

At time t , P_{ij}^t means the probability of agent i choose agent j as partner. P_{ij}^t means the times of successful cooperation between agent i and agent j , R_i^t means the times of suc-

successful cooperation between agent i and other agents. The overall equation indicates that agent i prefer the agent who had given more help to him. This suggests that agent have “preferential attachment” feature when choose partner.

c) Collaborative approach: Develop agents collaborate on line and the time they spend in current task is closely related to the distance between participant’s knowledge attribution and tasks’ required knowledge attribution. Time t is given by

$$t = \begin{cases} T_{\min}, \frac{T_{\max}}{\min\{(k_{i,x} - k_x^m), (k_{i,y} - k_y^m)\}} < T_{\min} \\ \frac{T_{\max}}{\min\{(k_{i,x} - k_x^m), (k_{i,y} - k_y^m)\}}, T_{\min} < \frac{T_{\max}}{\min\{(k_{i,x} - k_x^m), (k_{i,y} - k_y^m)\}} < T_{\max} \\ T_{\max}, \frac{T_{\max}}{\min\{(k_{i,x} - k_x^m), (k_{i,y} - k_y^m)\}} > T_{\max} \end{cases} \quad (2)$$

T_{\min} is the minimum time costs which will be spend in the module. It happens when a highly over-required-knowledge participant contributes. It is the basic time that high skill level can’t reduce. T_{\max} is the maximum time costs which will be spend in the task. It happens when a just required-knowledge participant contributes. It means that participant can finish the task, though he will spend maximum time in it. $k_{i,x}$ and $k_{i,y}$ is the knowledge attribution of participant i . k_x^m and k_y^m is the required knowledge of module m .

d) Collaboration result:

① Knowledge Collaboration Network: when agent i and agent j finish the task successfully, the knowledge collaboration network will add a tie ij .

② Knowledge Spillovers Mechanism: In the network evolution time t , agent i choose agent j as partner, G_t represents collaboration network, and the cumulative amount of knowledge is expressed as follows [18]:

$$k_{i,x}^{G_t} = k_{i,x} + (\delta^{d_{G_t}(i,j)} + \theta) * \max\{0, k_{j,x} - k_{i,x}\} \quad (3)$$

$$k_{i,y}^{G_t} = k_{i,y} + (\delta^{d_{G_t}(i,j)} + \theta) * \max\{0, k_{j,y} - k_{i,y}\} \quad (4)$$

$\delta(0 \leq \delta \leq 1)$ in the equation is knowledge spillover rate. $d_{G_t}(i, j)$ is the minimum distance between i and j in network G_t , which means strangeness between i and j which would affect the knowledge spillover level. $\theta(0 \leq \theta \leq 1)$ is the basis knowledge spillover level, represents the situation that the strangeness between i and j is infinity. Equation (3) and (4) indicate that if j ’s knowledge attribution is higher than i , then agent i can get j ’s over-knowledge when they are in collaborate action. With decrease of strangeness between i and j , this collaborate team’s knowledge attribution increases in the relation of exponential function. Specially, when knowledge spillover rate (δ) is zero, cooperators’ new knowledge attribution increases fewest. And when δ is 1, cooperators’ new knowledge attribution increases fastest.

③ Knowledge Learning Mechanism: After task being completed successfully, cooperators’ knowledge property increase through knowledge learning effect (Equation 5, 6).

$$K_{i,x}^{G_t} = \begin{cases} K_{i,x}^{G_{t-1}} * (1 + E_i), & K_x^m - K_{i,x}^{G_{t-1}} \geq 0 \\ K_{i,x}^{G_{t-1}}, & K_x^m - K_{i,x}^{G_{t-1}} < 0 \end{cases} \quad (5)$$

$$K_{i,y}^{G_t} = \begin{cases} K_{i,y}^{G_{t-1}} * (1 + E_i), & K_y^m - K_{i,y}^{G_{t-1}} \geq 0 \\ K_{i,y}^{G_{t-1}}, & K_y^m - K_{i,y}^{G_{t-1}} < 0 \end{cases} \quad (6)$$

$K_{i,y}^{G_t}$ represents participant’s knowledge attribution of x in period t . $K_{i,y}^{G_t}$ is participant’s knowledge attribution of y in period t . E_i represents participants’ learning ability. These equations mean that participants will increase their knowledge attribution in the process of completing tasks.

3.3. Simulation Process

Fig. (2) shows the dynamic process of the model.

To sum up, our simulation process follows these steps:

Step 1: Simulation clock to zero and parameter initialization.

Step 2: Determine whether the status of develop agent is Free and there isn’t exist uncompleted task of last clock cycle in the task box or not. If the status of develop agent is not free, the agent will join into develop process in current clock cycle, then go to step 5. If the status of develop agent is free and there is task in the box, agent will directly does the task in the box, and then go to step 4. If the status of design agent is free and there isn’t task in the box, then agent will select task randomly.

Step 3: Comparing knowledge of agent and requirement to complete the task to determine whether send cooperation request or not. If agent needs to send cooperation request, then select partner according to partnership selection mechanism. After that if develop agent can not satisfy knowledge requirement, then agent i drop out design module and go to step 5.

Step 4: Agent starts to develop, and update agent status and task status.

Step 5: Check whether all the agents are complete traversal or not, if not, then go to step2.

Step 6: Simulation clock push forward one step.

Step 7: Check all the tasks are completed. If not, go to step 2.

Step 8: Simulation ended and output simulation parameters.

4. EXPERIMENTS AND ANALYSIS

In order to simulate the process of knowledge diffusion in OSSC, we design an experiment to reproduce it. Parameters used in the model are specified in Table 1, and three scenarios with different participants’ knowledge variance are set. Matlab 6.5 is used for implementation of the model.

The results of simulation are provided in the following figures. Changes in average knowledge stock (Fig. 3), number of finished tasks (Fig. 4), knowledge variation (Fig. 5) and percentage of skilled participants (Fig. 6) were observed with implementation of three scenarios. The average knowledge

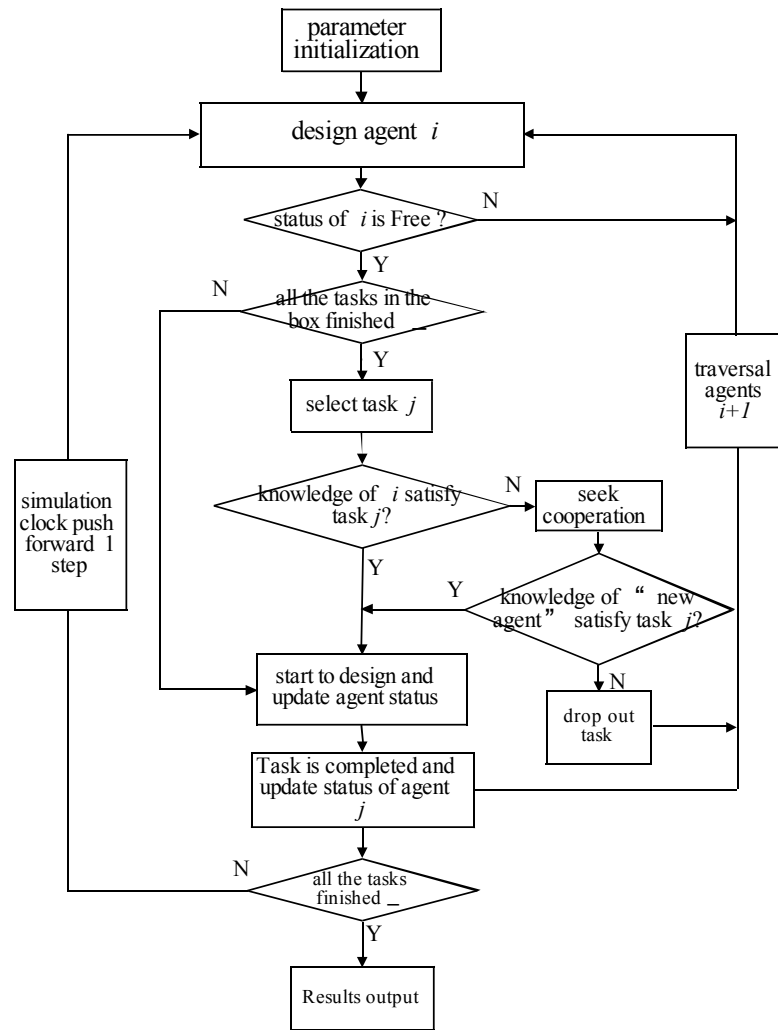


Fig. (2). Model’s simulation process.

Table 1. Parameter value.

#	Parameter	Value
1	Tasks required knowledge	$K_x^m \sim N(100,10)$; $K_y^m \sim N(100,10)$
2	Participants’ number	N=100
3	Learning ability	Pt~U(0,5%)
4	Knowledge spillover rate	$\delta = 0.5$
5	Basis knowledge spillover rate	$\theta = 0.3$
6	Minimum time cost	1
7	Maximum time cost	5
8	Society Network in period 0	Null
9	Simulation time	1000
10	Participant’s knowledge	scenario A: $k_{i,x} \sim N(100,5)$; $k_{i,y} \sim N(100,5)$
		scenario B: $k_{i,x} \sim N(100,10)$; $k_{i,y} \sim N(100,10)$
		scenario C: $k_{i,x} \sim N(100,20)$; $k_{i,y} \sim N(100,20)$

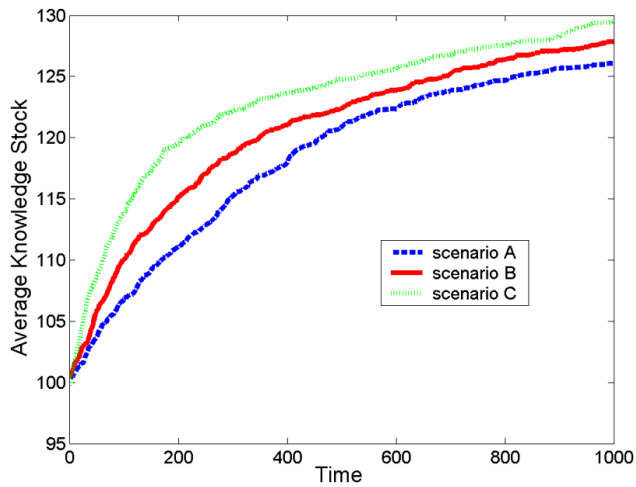


Fig. (3). Average knowledge stock of three scenarios.

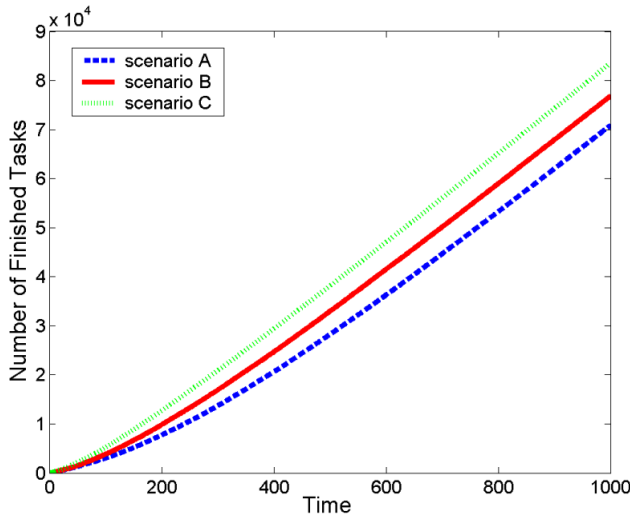


Fig. (4). Average knowledge stock of three scenarios.

stock of the community rapidly increased at the beginning, with the increase speed declining at different time. Initially, all of the three scenarios showed equal average knowledge stock. As time went on, Scenarios A shows the fast knowledge increasing speed among three scenarios, and in the whole period, the knowledge stock of scenario A is always higher than B, while knowledge stock of scenario B is always higher than C. Similar results are also happened in the changes of task accomplishment (Fig. 4), it should be the evidence of that the community’s initial knowledge distribution with higher variance could better the performance of knowledge diffusion and improve the task finish speed while community’s average knowledge is equal.

Fig. (5) shows the change of knowledge variation. It can be seen that the variation of A decrease rapidly at the beginning, and after time = 200, the knowledge variation declined smoothly. On the other side, the knowledge variation of B declined quickly before time=160, later the knowledge variation arrived at an almost steady state, which happens at scenario C’s whole period. It is because of knowledge exchange, which makes agents similar. It should be noted that high knowledge variation represents low equity of knowledge

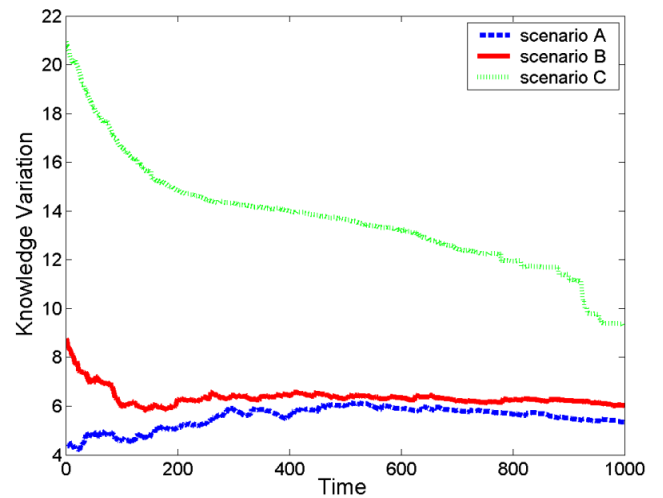


Fig. (5). Knowledge variation of three scenarios.

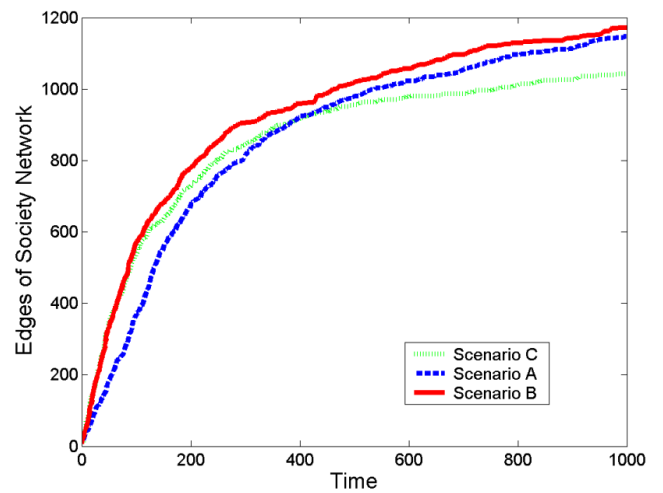


Fig. (6). network evolution of three scenarios.

diffusion. Consequently, higher knowledge variation community shows a low equity during knowledge diffusion. Fig. (6) shows the change of society network’s edge. The percentage rapidly increased at the beginning, then the increase speed reducing quickly at time = 200, the percentage achieve an almost steady but slightly increase state. In Fig. (6), the evolution of society network were shown, all three lines were increasing rapidly and got smooth at the later period. It should be noted that line C was no longer be the most advantage one. That’s because low-knowledge participants are harder to build relationship with others, normally they can only collaborate with high-knowledge participants. Thus, higher variation with more low-knowledge participants and create a less complex network. Therefore, higher knowledge variation community led to a fewer knowledge diffusion path than others.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This paper is based upon research work supported by National Natural Science Foundation of China (No.71171019), the Doctoral Program of Higher Education, China (No.20120006110034), the Central Universities under grant FRF-TP-14-055A2.

REFERENCES

- [1] Feller J, Finnegan P, Kelly D, *et al.* "Social Inclusion: Societal and Organizational Implications for Information Systems," *international working conference*, 2006, 261-278.
- [2] Linux, 2014 Linux Online - About the Linux Operating System [cited 2014, December 2] Web Link <http://www.linux.org/info/index.html>.
- [3] Mozilla 2013 About Mozilla [cited 2014, December 2] Web Link <http://www.mozilla.org/>
- [4] Apache Software Foundation 2008 The Apache HTTP Server Project [cited 2008, January 23] Web Link <http://httpd.apache.org/>.
- [5] Bass F M, "A simultaneous equation regression study of advertising and sales of cigarettes", *Journal of Marketing Research*, 1969, pp. 215 – 227.
- [6] Tseng F M, Hu Y C. "Quadratic-interval Bass model for new product sales diffusion," *Expert Systems with Applications*, 2009, 36(4): 8496-8502.
- [7] Islam T. "Household level innovation diffusion model of photovoltaic (PV) solar cells from stated preference data," *Energy Policy*, 2014, 65: 340-350.
- [8] Agarwal R, Bayus B L, "Tripsas M. Abandoning Innovation in Emerging Industries," *Customer Needs and Solutions*, 2014, 1(2): 91-104.
- [9] Tellis G J, Stremersch S, Yin E. "The international takeoff of new products: The role of economics, culture, and country innovativeness," *Marketing Science*, 2003, 22(2): 188-208.
- [10] Plouffe C R, Vandenbosch M, Hulland J. "Intermediating technologies and multi - group adoption: a comparison of consumer and merchant adoption intentions toward a new electronic payment system," *Journal of Product Innovation Management*, 2001, 18(2): 65-81.
- [11] Arts J W C, Frambach R T, Bijmolt T H A. "Generalizations on consumer innovation adoption: A meta-analysis on drivers of intention and behavior," *International Journal of Research in Marketing*, 2011, 28(2): 134-144.
- [12] Demoulin N T M, Zidda P. "Drivers of customers' adoption and adoption timing of a new loyalty card in the grocery retail market," *Journal of Retailing*, 2009, 85(3): 391-405.
- [13] Cowan R, Jonard N. "Knowledge portfolios and the organization of innovation networks," *Academy of Management Review*, 2009, 34(2): 320-342.
- [14] De Jong J P J, Von Hippel E A, Gault F, *et al.* "The diffusion of consumer-developed innovations: Patterns in Finland," Available at SSRN 2426498, 2014.
- [15] Lee H, Kim C, Cho H, *et al.* "An ANP-based technology network for identification of core technologies: A case of telecommunication technologies," *Expert Systems with Applications*, 2009, 36(1): 894-908.
- [16] Lin M, Li N. "Scale-free network provides an optimal pattern for knowledge transfer," *Physica A: Statistical Mechanics and its Applications*, 2010, 389(3): 473-480.
- [17] Mu J, Tang F, MacLachlan D L. "Absorptive and disseminative capacity: Knowledge transfer in intra-organization networks," *Expert Systems with Applications*, 2010, 37(1): 31-38.
- [18] HUANG W Q, ZHUANG X T, YAO S. "Self-organizing Evolution Model of Industry Cluster Innovation Networks and Its Simulation Analysis," *Chinese Journal of Management*, 2012, 9(10):1475-1483(in Chinese).

Received: September 16, 2014

Revised: December 23, 2014

Accepted: December 31, 2014

© Hu *et al.*; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.