

Analysis and Simulation of the System Dynamics Model of the Peer Production-Taking Baidu Encyclopedia as an Example

Zhihong Li, Ming Shao, Yu Cheng

School of Business Administration, South China University of Technology, Guangzhou, China Email: <u>1083740295@qq.com</u>

Received 16 November 2014; revised 18 December 2014; accepted 26 December 2014

Copyright © 2015 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0/

ву

Abstract

Modeling the dynamical mechanism of general peer production and taking Baidu Encyclopedia as an example, the authors constructed the system dynamics model of Baidu Encyclopedia by analyzing the causality and the system flow diagram. We introduced the subsystem of the model in detail. Then, we obtained raw data by collecting data on authority media and using web crawler. Quantitative analysis was combined with experts' advice to estimate parameters. The estimated parameters were examined. Finally, we put forward dynamical mechanism of Baidu Encyclopedia based on the results of simulating.

Keywords

Peer Production, System Dynamics, Dynamical Mechanism, Baidu Encyclopedia

1. Introduction

Since 1980s and 1990s, a new production model—Peer Production has arisen (also known as Peer to Peer Production or Parallel Production), which has resulted in changes in economy and society and impacted the existing production model persistently, such as Open Source Software, Wikipedia and other knowledge products [1]. Comparing to the large quantity, the quality of knowledge products is much poorer. The number of the users is large but only few contributors make contributions to knowledge products. Knowledge products are encountering serious bottlenecks. All these problems attach great significance to transforming users to contributors in Peer Production and to improving quality of knowledge products.

Both domestic and overseas researchers have been focused on the following areas of Peer Production: partic-

How to cite this paper: Li, Z.H., Shao, M. and Cheng, Y. (2015) Analysis and Simulation of the System Dynamics Model of the Peer Production-Taking Baidu Encyclopedia as an Example. *Open Journal of Social Sciences*, **3**, 13-23. <u>http://dx.doi.org/10.4236/jss.2015.31002</u> ipants [2]-[4], evaluating indicators [5]-[7], the influence of partners on the amount and quality [8]-[12], the influence of amount and the quality on knowledge production [13]-[15], and the influence of competitors [16]-[18], etc. But few researches are concerned about dynamic mechanism of Peer Production and much of researches are qualitative analysis from the theoretical perspective. More and more scholars make researches from quantitative view, but few make use of System Dynamics. Consequently, this paper will use System Dynamics to analyze the dynamical mechanism of Baidu Encyclopedia with modeling, emulation and analysis methods. Meanwhile, it will make recommendations to promote the development of Peer Production based on the emulation results.

2. Literature Review

The definition of Peer Production in this paper is, with the mutually agreed specification, that people worldwide participate in the massive knowledge production voluntarily and collaboratively with the help of relevant tools and Internet. Peer Production constitutes knowledge productions, managers, producers, users, rules and regulations. Only few researches are about dynamics mechanism of Peer Production and most of them focus on open source software and online encyclopedia. And for the huge creativity of Peer Production, Scholars do researches on dynamical mechanism from different perspectives, for example, from the view of complexity and self-organization.

The complexity of Peer Production draws more and more attentions. Xinglu Zhao thinks Peer Production as a typical emergent phenomenon of the complex system. Based on the theory of dissipative structure, he proposes that the contradiction between knowledge supply and demand drives the development of Peer Production. The contradiction makes the system deviate from equilibrium, and then introduces negative entropy flow from outside to maintain the continuous evolution and development of the system. Feicheng Ma and colleagues describe complex adaptive characteristics of Wikipedia based on the key idea of Complex Adaptive System (CAS) theory. They define four indicators: web, content, user, organization to evaluate and analyze the ordering mechanism of Wikipedia. He proposes that key points to improve orderly evolution of Wikipedia are "the thought trusting on public wisdom" and "the mechanism promoting the information flow".

Self-organization is a major factor of Peer Production. Kevin Crowston *et al.* (2007) research on Free/Libre and Open Source Software (FLOSS) projects and point out the development team of free and open source Microsoft is self-organized. And most of the scholars do admit Wikipedia is self-organization which creates huge vitality and promotes the quantity and quality of entries for Wikipedia.

In addition, more and more scholars make researches from the quantitative perspective. Dennis M. Wilkinsonfinds that there exist strong regularities in Peer Production system: 1) The services of contributors present a characteristic of Power-law distribution and the power is positively proportional to endeavor of contributions. 2) The activity level of each topic presents logarithm distribution which is resulted by Random Strengthen mechanism, for example, the few but popular topics do make a significant contribution. All these regularities push forward the evolution of Peer Production. Most scholars do approve the contribution made by Wikipedia users to the entries growth. M. Almeida concludes that the building and maintaining of Wikipedia entries shared selfsimilar phenomenon and increasing of users results in the exponential growing of Wikipedia with detailed data. The researches of Rodrigo Almeida [10] *et al.* (2007) show that it is the growth of users that promotes Wikipedia's growth because new registrants are more enthusiastic than older ones to edit new papers. Xiaoyu Li [11] *et al.* (2009) also make analysis of Chinese Wikipedia data. They propose that the amount of users almost grows synchronously with entries in exponential type, and the edit degree and extent of entries share liner growth which results in growth of entries quality.

As open system, Peer Production is influenced by the outside world inevitably. So many scholars work on the topic about how external environment affects Peer Production. Zhicheng Luo [11] *et al.* (2008), choose the Credibility crisis and restricted access mandated by government to analyze the impact of external force showed on Peer Production. Those two factors promote the ordering of Peer Production but have no negative influence on growing trend. Fershtman and Gandal [19] (2007) investigate on participators of 71 Free and Open Source Microsoft projects to evaluate their average contributions. The results display that the contributions are positively connected with the degree of software access.

There are hardly any researches using System dynamics to research on Peer Production, Contrariwise, System Dynamics is used widely in other similar fields. Xianjin Cha [20] (2007) build a model and flow figure to research Chinese network environment by introducing some elements, such as sharing coefficient, actual sharing,

congestion impact factor, International Outlet Bandwidth, internet users, the number of websites, etc. All these elements are emulated in System Dynamics equation. Lastly, adjusting the factors is to find the change trend of new users, websites and actual sharing, which give guidance for researching Complex System.

In summary, Peer Production is a self-organizing and complex system which produces knowledge. It is related closely with all components and also influenced by external environment. So based on previous studies and System Dynamics theories, we propose the dynamics mechanism of Peer Production and set Baidu Encyclopedia as an example to establish a System Dynamics model.

3. System Dynamics Model of Baidu Encyclopedia

Baidu Encyclopedia is a typical Peer Production and online encyclopedia which edited by whole people. Baidu Encyclopedia advocates "equality, cooperation and sharing". Users can participate in editing website and sharing knowledge. The communication and cooperation among users make entries perfect, and make it a trinity knowledge sharing system to be supplement of Baidu search engine, along with Baidu Post Bar and Baidu Know.

3.1. Causality Analysis of Baidu Encyclopedia Dynamics Mechanism

Firstly, we draw a causality figure to display the dynamics mechanism of Baidu Encyclopedia based on general Peer Production motive theory and special characteristics of Baidu Encyclopedia (Figure 1).

Baidu Encyclopedia causality figure is made up of user subsystem, editor subsystem, entry create subsystem, entry quality subsystem and attraction subsystem. There exists a feedback loop of interweaving of positive and negative feedback in all subsystems.

In order to simulate the growth of users and editors realistically, we borrow ideas from the disease infection. Every day, Baidu Encyclopedia users and non-users communicate with each other, both online and offline. So users may convert non-users to be users. This flow resembles virus infection model in Communication. Users are infected with "virus", and non-users are vulnerable persons, so by communicating and infecting, non-users may turn to be users. And users may also run off to be non-users, which is analogous with curing virus. Referring to the dynamical mechanism of epidemic, we draw the causality figure of Baidu Encyclopedia users growing subsystem figure, the left bottom of the area). The middle left corner of the area is the causality figure of editors growing subsystem.

In Baidu Encyclopedia entries create and quality subsystem, the amount of editors, entries create coefficient and edit coefficient all have positive effect on new entries amount and new editing frequency. In addition, the growth of new entries or editing frequency do directly lead to the growth of entries and total editing frequency, while the difficulty of improving quality and creating more new entries are also increasing which is of obstructing. So we draw a figure to display Baidu Encyclopedia entries create subsystem (Figure 1, upper left) and entries quality subsystem (Figure 1, upper right).

In attraction subsystem, the growth of users, editors, editing frequency and the improvement of quality, the promotion do make Baidu Encyclopedia more attractive. At the same time, the demission of users and editors and competition do weaken the attraction. We refer to the idea of vicious competition model in system dynamics to reduce the causality figure of attraction subsystem (Figure 1, bottom right).

3.2. The System Flow Figure Analysis of Baidu Encyclopedia Development Motive

Infection rate 1 is a rate which evaluates the probability of users leading non-users to use Baidu Encyclopedia. Infection rate 2 is the probability rate of editors leading non-editors to edit Encyclopedia. The demission coefficient of users and editors means the probability of users and editors run off by nature. Entries create coefficient is the rate that editors create a new entry, and entries edit coefficient is the frequency or rate of editing an entry.

Based on the previous analysis, we conclude the Dynamics System flow figure of Baidu Encyclopedia. Please refer to **Figure 2**. The variables of dynamics system flow figure of Baidu Encyclopedia please refer to **Table 1**.

4. Parameter Estimation and Testing

4.1. Parameter Estimation

1) Collecting data from authoritative institution



Figure 1. Dynamics system causality figure of Baidu Encyclopedia.

Most of the variables were collected from Baidu official website, China Internet information center. Others came from authoritative media. For some missing data, we refer to Baidu Encyclopedia or other Encyclopedia and take some authority's advice into consideration, such as intermediate value method and linear difference method.

2) Collecting data by web crawler

We use web crawler to crawl online data at certain point in time. For example, we crawl some entries randomly, and calculate the average editing frequency to estimate the total editing number of all Baidu Encyclopedia entries before January 1, 2011.

3) Testing the parameters of major variables repeatedly

For the parameters need testing, we selected some probable data to estimate until the model is steady and of no signature change. After several testing, the parameters of major variables are finally fixed.

4.2. Testing the Model

The methods to test the applicability and consistency of Dynamical mechanism can be broadly divided into applicability test of the model structure, the applicability test of the model behavior, the consistency test of model structure and the actual system and the consistency test of model behavior and the actual system. Table 2 is the comparison between simulation and actual key data.

As shown in the simulation **Figure 3**, users amount and entries amount share the same "s"-shaped curve, and in the same phase of rapid growth (near the horizontal axis 48 in chart, as same below), maybe the trend will grow slowly in the future. The editor amount also displays in "S" shape curve, and grows rapidly. And for edit



Figure 2. Dynamical system flow figure of Baidu Encyclopedia.





Table 1. The variables of dynamics system flow figure of Baidu Encyclopedia.							
Variable type	Variable name						
Level variable	Users amount, editors amount, entries amount, edit frequency, the attraction						
Rate variable	The increment of users, the decrement of users, the increment of editors, the decrement of editors, the increment of entries, the editing frequency of new entries, the increased attraction, the decreased attraction						
Auxiliary variable	the number of Chinese netizen, effective contact coefficient1, non-users amount, the conversion factor of users, the increased attraction of users increment, the decreased attraction of users demission, effective contact coefficient 2, non-editors amount, the conversion factor of editors, the increased attraction of editors increment, the decreased attraction of editors demission, the difficulty of creating potential entries, the conversion factor of entries amount, the increased attraction of entries increment, entries quality, the difficulty of improving entries quality, the conversion factor of editing frequency, the increased attraction of edit frequency increment, improved effect, promotional effect, the attraction of external peer organization perceived, the competition from external peer organization						
Constant	Infection rate1, contact coefficient 1, users demission coefficient, infection rate 2, contact coefficient 2, editors demission coefficient, difficulty regulatory factor 1, entries create coefficient, entries edit coefficient, difficulty regulatory factor 2, improved coefficient, promotional coefficient, the decreased attraction of competition						

Table 2. The comparison between simulation and actual key data.

Time		2007.1.1			2008.4.23			2011.1.1	
Data type	Original value	Simulation value	Error	Original value	Simulation value	Error	Original value	Simulation value	Error
Registrant number	18^{*}	18	-	50	56.58	13.2%	Unknown	186.02	-
Editors number	1*	1	-	2	1.49	-25.5%	Unknown	11.4	-
Entries number	5.55	5.55	-	11	11.51	4.6%	29.6	30.78	4%
Total edit frequency	11.1*	11.1	-	27	30.28	12.1%	138.2 ¹	124.86	9.7%
The attraction	3	3	-	-	4.59	-	7.3	6.91	-5.3%

Remarks: besides the attraction, other values are all in million, and * are estimated. Despite the registrant number in 2008, and other two simulated data, error are between 10% and 15%. The rest are all in little error and between minus and plus 10%. The simulation coincides with the actual data by and large.

frequency, it is more likely exponential growth pattern based on the current simulation, but unknown for future trend. The trend of attraction growth is in asymptote shape.

The main purpose of this paper is not to forecast the exact value of relevant parameters, but to analyze the internal behavior model in system dynamics. So we think the model is similar to the actual system on the whole, and the dynamics model we build share the applicability and consistency which can be used as a tool for simulation and policy analysis.

5. Model Simulation and Data Analysis

5.1. The Simulation of Model

Use Vensim software to analyze the following factors: user amount, editor amount, entry amount, edit frequency and attraction. Tree Analysis results are showed in Figure 4.

The main factors that dominate the trend of users' amount are the addition and demission of users. At beginning, both addition and demission amount grow rapidly but the addition is of more which results in the vast growth and exponential trend. However in the middle and latter periods, the increased amount begins to come down after the peak, while the demission amount increases. Therefore, the growing trend of user amount begins to slow down. The growing trend of Baidu Encyclopedia users presents the "S"-shaped curve, with a typical growth ceiling characteristic.

The main factors that dominate the trend of editors amount are the addition and demission of editors. In the beginning and middle period, both the addition and demission amount grow rapidly, but the addition is of more which results in the vast growth and exponential trend. Latter, the increased amount begins to come down after

¹Selecting 8664 entries randomly (until Jan. 1st 2011), the total editing frequency is 40,476 times and the average editing frequency is 4.67.



Figure 4. The tree analysis of relevant variables simulation in Baidu Encyclopedia.

the peak, but the demission amount grows constantly and keeps at a high level. Although the increment of editors declines, the trend is gentle which leads to the growing trend for a long period comparing to the users trend. Similarly, the trend of editors' number also has a typical growth ceiling characteristic.

For entries' number, the main factor is the increased amount. The entries grow rapidly in the early and middle days, results in the exponential trend. But later, the amount falls after the peak and the growing speed slow down. So the entries' number also displays a typical growth ceiling characteristic.

The main factor that dominant editing frequency trend is the editing frequency of new entries. Until the middle-advanced period, the trend of new entries editing frequency keeps a exponential growth and the whole trend of Baidu entries grows in exponential type during most of the simulation cycle. After the peak, the new entries editing times slow down gently, but the overall trend still keeps increasing. So we can predict that the growing trend of entries editing frequency will last for a longer time than entries number. For the reason, it may contribute to the fact that the entries development depends on the number increasing at the beginning and as it develops, the main factor turns to be the entries quality or it could be related to Baidu Encyclopedia's promotions. So the growth ceiling characteristic is also a trend of editing frequency.

The main affections of the trend of Baidu Encyclopedia attractions are addition and demission of attraction. At beginning, the attraction increases rapidly and slows down after a period. The decreased amount maintains a growing trend, but the main factor is the growth trend resulting in the overall increasing trend. This trend keeps till the middle period. The growing amount keeps increasing at a high level and then decreases, and the decreased amount reaches a peak and then drops. Finally, both the increased amount and decreased amount drop at a high speed. But the increased amount still outstrip the decreased amount, which leads to the slowly growing trend and growth ceiling feature of Baidu Encyclopedia attraction trend.

5.2. Policy simulation Analysis

1) Changing some constant

As there are too many constants, we adjust the infection rate 1 (variation within plus and minus 5%) to analyze the variation of relevant constant. The results are showed in **Figure 5**. The results show that Baidu Encyclopedia user amount is very sensitive to the change of infection coefficient, and the editor amount is slightly less sensitive which indirectly lead to the change of entries number and editing times. And then the effect works on Baidu Encyclopedia attraction and finally affects other parameters.

In this paper, we do policy simulation analysis on infection coefficient, users demission coefficient, editors demission coefficient, difficulty regulatory factor, entries create factor, entries edit coefficient, improved coefficient and promotional coefficient(due to the space restriction, we don't list all the simulation figures). The simulation shows that adjusting relevant parameters can affect Baidu Encyclopedia parameters, although the sensitivity differs from each other, which proves the reliability of this model.

2) The quantity have the priority VS the quality have the priority

The knowledge production's quantity and quality are the significant measurement of Peer Production. But which weighs more important in actual case? We set a simple simulation to solve this question. Design two policies: a) increase the entries create coefficient by 10%; b) increase the entries edit coefficient by 10%. We select some key factors to compare (please refer to Table 3). From the comprehensive view, we think that adjusting the entry



Figure 5. The influence simulation of adjusting infection rate 1 on other parameters

 Table 3. Change of relevant parameters when adjusting some coefficients.

Adjusting variables parameters	Original value	Entries make coefficient + 10% (change rate)	Entries edit coefficient + 10% (change rate)
The users number of Baidu Encyclopedia	32529.1	32816.5 (+0.88%)	32768.3 (+0.74%)
The editors number of Baidu Encyclopedia	28664.4	28883.7 (+0.77%)	28848.3 (+0.64%)
Entries number of Baidu Encyclopedia	9126.04	9281.53 (+1.70%)	9137.31 (+0.12%)
Total editing times of Baidu Encyclopedia	115,706	116169 (+0.40%)	121102 (+4.66%)
The attraction of Baidu Encyclopedia	92.6706	93.0829 (+0.44%)	93.1652 (+0.53%)

Remarks: all data in the table are the final value of simulation final time. Unit is thousand.

create coefficient works more efficient and set quantity priority is a better strategy. This inference may be consistent with the realistic that most Chinese encyclopedia pursuit large entries number. Admittedly, simulation conclusion may be affected by the construction of the model which is not sufficient to be general conclusions.

3) Take the delay caused by external competition into policy analysis

There exists delay in the awareness and reaction when different Chinese Encyclopedia organizations compete with each other. The delay time is different and the following is the analysis of delay caused by external competition.

a) Changing the delay time of threat degree that external organization feels

We change the detention of the threat external organization feels, the original time is 3 months, now we change to 1 month and 6 months, and the results are showed in **Figure 6**. The sensitivity and the instant degree are weaker, as the detention is longer. And the threat degree the organization felt is lower than normal, so weakening the competitive and effectiveness of competitions or actions.

b) Changing the delay time of taking actions when external organization faces the competitions



Figure 6. The delay of threat simulation that external peer organization perceived.

The first part describes the delay of threat external organization perceived, the following analysis is focused on the delay of the reactions when organization faces competitions. We change the detention from 1 month to 3 or 6 months and observe the change of relevant parameters. When organizations defer to take reactions, the competitive activity's effectiveness is weakened, please refer to Figure 7.

6. Conclusions

We built a dynamical mechanism model to analyze the Baidu Encyclopedia. We use the model to estimate and test the parameters, to do the simulation and policy simulation analysis, and research the Peer Production theory and relevant cases. Based on these, we propose some advice to promote the development of Peer Production.

First, improve the attraction and the benign development of Peer Production.

We can make Peer Production more attractive with below methods, such as building open and sharing property mechanism, giving the users access to use and modify the knowledge product freely. And we can consummate the functions of Peer Production system to reduce the demission of users and develop the system in a benign way such as improving the user-friendliness.

Second, take the appropriate strategies to advance the evolution of Peer Production. In different periods, we should take different strategies to advance the evolution of Peer Production. In the



Figure 7. The simulation of the delay of actions external peer organizations take.

beginning, the system expands rapidly and we should take the number priority strategy to reduce the negative impact of external competition, and the strategies to improve positive feedback in early periods such as increasing the infection coefficient and organizing promotion activities. While latterly we take the strategies to weaken the negative feedback such as lower down the users and editors demission coefficient.

Third, deal with the competition from external organizations rationally and bring out a mutual benefit situation.

Faced with the competitions from external organizations, Peer Production can take use of detention of the threat and reactions to confuse competitors. This could weaken the negative effect caused by competition and avoid vicious competition. Peer Production system should take the competitors into consideration when making decisions and seek a way to realize mutual benefit.

References

- [1] Yochai, B. (2002) Coase's Penguin, or Linux and the Nature of the Firm. *Yale Law Journal*, **112**, 369-446. http://dx.doi.org/10.2307/1562247
- [2] Ye, Y. and Kishida, K. (2003) Towards an Understanding of the Motivation of Open Source Software Developers. Proceedings of the 25th International Conference on Software Engineering, Portland, 10-15 May 2003, 419-429.
- [3] Ke, W., Chen, Y.T. and Si, C.L. (2006) Incentive Mechanism in Open Source Innovation Community. Science & Technology Progress and Policy, 7, 13-16.
- [4] Raymond, E.S. (2001) The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly & Associates.
- [5] Taylor, B.J. (2008) Photobase—A Research Platform to Investigate Peer Production and Collaborative Sensing Systems. University of Massachusetts, Massachusetts.
- [6] Ma, F.C. and Xia, Y.H. (2008) Mechanism of Ordering in Wikipedia Based on the CAS Theory. *Library Tribune*, **28**, 85-92.
- [7] Messerschmitt, D.G. and Szyperski, C. (2001) Industrial and Economic Properties of Software Technology, Processes, and Value. University of California at Berkeley Computer Science Division Technical Report, California.
- [8] Lan, F. (2007) Data Analysis and Research on Open Source Software Projects. Shanghai Jiao Tong University, Shanghai.
- [9] Hu, H.H. and Zhou, Z.H. (2001) Virtualization of Knowledge Cooperation, an Analysis of the Development of Linux. *Science Research Management*, **21**, 60-67.
- [10] Almeida, R., Mozafari, B. and Cho, J. (2007) On the Evolution of Wikipedia. *The International Conference on Weblogs and Social Media*, 1-5 October 2007, 543-551.
- [11] Li, X.Y. and Luo, Z.C. (2009) Study on Evolving Tendency and Policy Environment of Chinese Language Wikipedia. *Journal of Intelligence*, **2**, 160-166.
- [12] Wikimedia Foundation (2010) Wikipedia Growth Mechanism [EB/OL]. http://en.wikipedia.org/wiki/Wikipedia:Modelling_Wikipedia's_growth
- [13] Zhao, F., Zhou, T., Zhang, L. and Ma, M.H. (2010) Research Progress on Wikipedia. Journal of the University of Electronic Science and Technology of China, 39, 321-334.
- [14] Braendle, A. (2006) Many Cooks Don't Spoil the Broth. Proceedings of Wikimania 2005-The First International Wikimedia Conference, Frankfurt, 20-23 July 2006, 215-223.
- [15] Buriol, L., Castillo, C., Donato, D., Leonardi, S. and Millozzi, S. (2006) Temporal Analysis of the Wikigraph. Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Washington DC, 1-5 May 2006, 45-51.
- [16] W3schools. The Market Share of Linux Outstrips 5% in Nov and the One of Windows Is still 90% [EB/OL]. http://www.linuxidc.com/Linux/2010-12/30446.htm
- [17] Chinese Wikipedia. Linux Vision [EB/OL]. <u>http://zh.wikipedia.org/zh/Linux%E5%8F%91%E8%A1%8C%E7%89%88</u>
- [18] Sourceforge org. Sourceforge Platform Software Library [EB/OL]. http://sourceforge.net/
- [19] von Hippel, E. (2007) Innovation by User Communities: Learning from Open-Source Software. MIT Sloan Management Review, 42, 83-89.
- [20] Cha, X.J., Yan, M. and Yan, Y.L. (2007) System Dynamics Modeling of Chinese Network Environment. *China Soft Science*, 1, 128-133.



Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.



 \checkmark