

Trend Analysis of Virtual Community Productivity Based on SSCI Database by Bibliometric Methodology

¹Chien-Liang Lin, ²Wan-Chu Yen, ³Heng-Li Yang

^{1*}Department of Management Information Systems, National Cheng Chi University, Taiwan,
lin.chienliang@gamil.com

²Department of Business Administration, National Cheng Chi University, Taiwan, rbrjanet@gmail.com

³Department of Management Information Systems, National Cheng Chi University, Taiwan,
yanh@nccu.edu.tw

Abstract

As a popular phenomenon today, virtual community is significantly altering people who work, spend their free time and interact socially. Furthermore, virtual community has been receiving a growing attention in the academic literature. This study aims to investigate the current state of the academic literature regarding virtual community and analyze its knowledge base by using bibliometric method. All articles were collected from SSCI database during 1994-2011. The result indicated that the literature productivity on virtual community topic is still growing. The result also found that the author productivity distribution data related to virtual community matched Lotka's law. Moreover, the twenty-eight core journals in virtual community issue were identified utilizing the Bradford's law.

Keywords: Component, Bibliometric, Virtual community, Lotka's law, Bradford's law

1. Introduction

The development of Internet technology has led to the proliferation of virtual communities. Virtual community covers a wide range of forms of social media such as bulletin board systems (BBS), chat rooms, online forums, social networking sites and so on. Individuals can interact with each other and share common interests via those electronic means. Thus, the community networks have promoted the development of each other and reinforced the role "community" plays in the cyberspace [1, 2]. Virtual community conception was first proposed by Howard Rheingold [1]. After that, virtual community has received growing attention in the academic literature [3].

In the past, Bibliometrics is a research method used in library and information sciences. Bibliometric methods were used to serve the purposes of description, evaluation and scientific monitoring [4]. A number of studies used bibliometric methods to explain the different fields, for instance, the investigation of research on evolution of altruism [5], the research trend in RFID [6], data mining and customer relationship management [7], the research profile in Experts System with Application [8] and Information & Management [9]. Therefore, this study attempts to use bibliometric analysis to identify the current state of the academic literature regarding virtual community and analyze its knowledge base such as research authors, research countries, document types, keyword analysis and subject areas. Furthermore, Lotka's and Bradford's laws are applied to perform author productivity and core journals analyses in this field during 1994-2011 respectively, in turn discovering historical vein and forecasting research tendency in the near future.

In the following section, the related literature including that on the Bradford's Law and Lotka's Law will be reviewed. Then the applicability of the research methodology in the empirical context will be reported in Section 3, followed by an analysis and the empirical results in Section 4. The conclusion will be presented in the final section.

2. Literature Review

2.1 Bradford's Law

Bradford's law was originally proposed by Bradford [10], which was the evidence for the regularity in the distribution of scientific journals [11]. Bradford's theory explained that if scientific journals are arranged in decreasing order of productivity of articles according to a given subject, they can be divided into a "core" of journals devoted to the subject as well as radiating zones of journals contributing fewer and fewer articles [12]. Through the process, each zones with the same number of articles, that the

number of journals in each zone grew geometrically [11]. The Bradford's Law has been widely applied to study varying topic trends [24], author productivity [13], and journal distribution [14].

2.2 Lotka's Law

Lotka's law is mainly used to describe the distribution of publications within a community of scientists in a given field [15]. It is also called "the inverse square law of scientific productivity" [16]. Lotka [16] found the publication data and formulated it to predict an approximate number of authors with a certain frequency of publications. Lotka's distribution function is given by the expression:

$$f(x) = \frac{C}{x^n} \quad (1)$$

where x is the number of papers published in a period; $f(x)$ is the number of authors publishing x papers; n is a parameter to be determined from the data that taking a value close to two; and C is a normalizing constant that the sum over all x of the $f(x)$ is equal to one.

The validity of this law has been proven regarding the productivity rules of physicists and econometricians [17, 18]. Lotka's distribution can be widely used in various studies such as the time interval, varying rates of production and different counting procedures [19]. From the past theories view, Chang, Chou & Yang [13] also built up the distribution of publications of information systems field following the Lotka's law, and the result matched the generalized Lotka's law; that is, the author productivity distribution data is consistent with technology acceptance theories. Tsai & Chi [25] research supply chain management (SCM) technology trends and forecasts using bibliometric analysis from 1989 to 2009 years.

3. Data Set

In this study, all articles were accessed from the citation database of Social Science Citation Index (SSCI) on Web of Science created by the Institute for Scientific Information (ISI). Since virtual community published their first issue in 1993 [1], all articles were selected by examining the title for "virtual community" keywords from 1994 to 2011. A total of 1,228 papers in virtual community published during 1994-2011 were analyzed.

4. Analysis and Results

4.1 Descriptive Statistics

The research articles analysis was focused on the period from 1994 to 2011. As presented in Table 1, the number of articles published in academic journals was only 12 in 1994, but it has quickly increased to over 100 articles per year. To date, 213 articles have been published in 2010. If this rate continues, we can expect to see approximately 300 virtual communities academic articles published in SSCI journals.

Table 1. The tendency of literature growth of virtual community

Publication Year	Nr. of publications	%
1994	12	0.98%
1995	16	1.30%
1996	16	1.30%
1997	26	2.12%
1998	32	2.61%
1999	43	3.50%
2000	23	1.87%
2001	39	3.18%
2002	54	4.40%
2003	63	5.13%
2004	54	4.40%
2005	48	3.91 %
2006	77	6.27%
2007	96	7.82%
2008	126	10.26%
2009	146	11.89%
2010	213	17.35%
2011	144	11.73%

Apparently, the literature citation of virtual community is increasing since 2003, and citation is also increasing steadily each year. It appears that the research of virtual community is very popular and getting in the highly mature period, shown as Figure 1.

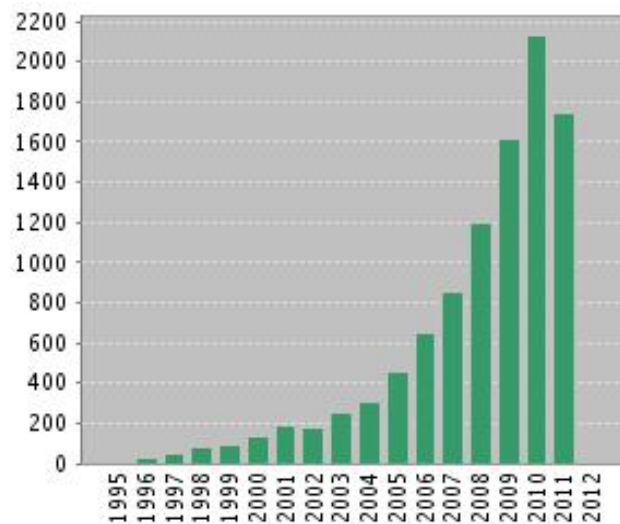


Figure 1. Citation in each year (Source: SSCI database)

The distribution of productivity countries by publishing papers regarding virtual community is listed in Table 2. The publication papers of the top three countries are USA, which is champion with 491

record counts (39.98%), followed by England and Canada, which achieved record counts of 142 (11.56%) and 90 (7.33%). Note-worthily, Taiwan is ranking as No. 4 in this research domain; it showed that Taiwan has become one of the main productivity countries within the research aspect of virtual community in the world.

Table 2. Distribution of top 10 productivity countries

Ranking	Country	Record Count	%
1	USA	491	39.98%
2	ENGLAND	142	11.56%
3	CANADA	90	7.33%
4	TAIWAN	87	7.09%
5	SPAIN	49	3.99%
6	PEOPLES R CHINA	48	3.91%
7	AUSTRALIA	45	3.66%
8	GERMANY	39	3.18%
9	NETHERLANDS	37	3.01%
10	ITALY	32	2.61%
	Others	168	13.68%

Table 3 provided an investigation into the authors who have written more than four papers related to virtual community topic during 1994-2011. Top 5 authors were sorted by the number of publications as well as their current affiliation. As the result, it indicates that the relatively more authors from non-U.S. academic institutions performed high productivity in this research domain. To our knowledge, virtual community has caught more and more attention by non-U.S. authors.

Table 3. Distribution of top 15 authors

Ranking	Author	Affiliation	Record Count	%	Subject Area
1	Lin, H.F.	National Taiwan Ocean University	7	0.57%	Knowledge Management
2	Flavian, C.	University of Zaragoza	6	0.49%	Travel Community, Brand Community
2	Guinaliu, M.	University of Zaragoza	6	0.49%	Travel Community, Brand Community
2	Riva, G.	Tech University Munich Germany	6	0.49%	Virtual Reality
2	Fueller, J.	University of Innsbruck	6	0.49%	Virtual communities

The SSCI journal comprised the majority of published virtual community document types (see Table 4). "Article" (1,100 record counts, 89.58%) is the most general publication type in the search results.

Table 4. Distribution of document types

Document Type	Record Count	%
Article	1100	89.58%
Proceedings Paper	102	8.31%
Review	44	3.58%
Book Review	44	3.58%
Editorial Material	24	1.95%
Meeting Abstract	11	0.90%
Correction	3	0.24%
Database Review	1	0.08%

Table 5 is important for researchers to gain a better understanding into the distribution of top 20 subject areas for future research trends and research directions. Among all the subject areas listed here, the top five ranking of research domains of virtual community are “Computer Science”, followed by “Information Science & Library Science”, “Business Economics”, “Psychology” and “Education & Educational Research”. The result indicated that virtual community has been considered as an important issue in Information System and Computer Science, that is, virtual community has become an interdisciplinary theme.

Table 5. Distribution of top 20 subject areas

Subject Area	Record Count	%
Computer Science	275	22.39%
Information Science & Library Science	248	20.20%
Business Economics	219	17.83%
Psychology	150	12.22%
Education & Educational Research	138	11.24%
Communication	134	10.91%
Engineering	90	7.33%
Sociology	72	5.86%
Social Sciences Other Topics	58	4.72%
Environmental Sciences Ecology	38	3.09%
Geography	34	2.77%
Public Environmental Occupational Health	34	2.77%
Operations Research Management Science	31	2.52%
Government Law	30	2.44%
Nursing	30	2.44%
Health Care Sciences Services	27	2.20%
Rehabilitation	26	2.12%
Public Administration	24	1.95%
Psychiatry	20	1.63%
Telecommunications	20	1.63%

4.2 Keyword Analysis

In addition, this research further conducted the keywords analysis. The analysis of author keywords (i.e. the keyword provided by authors) could provide the information of research trends which is concerned by most researchers. The technique of statistical analysis of keywords might be purpose exploring directions of science [20]. However, previous bibliometric studies have used the author keyword to analyze the trend of research [21]. Examining this period, we found that 2,346 author keywords were used. The top five keywords are “virtual community”, “Internet”, “e-learning”, “social network”, and “computer-mediated communication”. The detailed ranking is shown in Table 6. Through analysis we can see that virtual community has been related to many research areas.

Table 6. Top 15 frequency of author keywords user

Rank	Author keywords	Count (Sum=2346)	%
1	Virtual Community (Online Community)	187	7.97%
2	Internet (Web)	51	2.17%
3	E-learning (Online Learning, Collaborative Learning)	48	2.05%
4	Social network (Community Network, Social Networking)	39	1.66%
5	Computer-Mediated Communication	31	1.32%
6	Trust	26	1.11%
7	Virtual Organizations	20	0.85%
7	Virtual Reality	20	0.85%
9	Behavioral Intention (Continuance Intention)	18	0.77%
10	Blog (Weblog, Blogging)	17	0.72%
11	Learning Community	16	0.68%
12	Social Capital Theory	16	0.68%
13	Online Game	16	0.68%
14	Knowledge Sharing	15	0.64%
15	Loyalty	12	0.51%

Author keywords: the keyword provided by authors; %: the percentage of the author keyword.

4.3 Bradford’s Law and Core Journal

Bradford’s Law is used to estimate the exponentially diminishing returns of extending a search for references in SSCI journals. Bradford’s Law impose a formulation that if journals in a field are divided by number of articles into three zones, the number of journals in each zone will be proportional to 1: n: n² [22].

There are 580 journals has published articles related to “virtual community”. Among them, 380 journals only published one article. The number of published papers in each journal and relative ranking based on the zoning of Bradford’s Law are shown in Table 7. Table 7 also presented the publication number of three zones (64: 200: 580). It was approximately equal to 1: 3: 3² which matched the principle of Bradford’s Law. That is, A: B: C = 1: n: n². The primary journals that published virtual community articles are as follows: Computers in Human Behavior has 32 publication papers, followed by the CyberPsychology, Behavior, and Social Networking (31 record counts), Computers & Education (26 record counts) and New Media Society (23 record counts). The result indicated that the literatures of virtual community were usually published by 64 core journals.

Table 7. The distribution of virtual community literature

	Number of Articles (a)d	Number of Journal (b)	Accumulated Number of Journals (c)	Accumulated (a*b)
	32(Computers in Human Behavior)	1	1	32
	31(CyberPsychology, Behavior)	1	2	63
	26(Computers & Education)	1	3	89
	23 (New Media Society)	1	4	112
	18	1	5	130
	16	1	6	146
	15	1	7	161
	14	2	9	189
(A)	13	2	11	215
Core	12	2	13	239
	11	3	16	282
	10	1	17	292
	9	5	22	327
	8	3	25	351
	7	5	30	386
	6	6	36	422
	5	9	45	467
	4	19	64	543
(B)	3	43	107	672
Relevant	2	93	200	858
(C)	1	380	580	1,238
Marginal				

4.4 Lotka's Law and Author Productivity Analysis

In this section, we will calculate whether the virtual community follow Lotka's law. Pao [23] suggesting these steps to verify the reliability of Lotka's Law: (1) collecting data, (2) list author frequency distribution table, (3) calculating n value (slope), (4) calculating constant C value, and (5) using K-S (Kolmogorov-Smirnov) test to evaluate whether to matchd Lotka's Law.

This study examined those journal articles published during 1994-2011 on various aspects of virtual community. The least-squares method described by Pao [23] was employed. In Table 8, this study described authors' productivity in virtual community (see Table 8). There are 1,228 articles and 4,201 authors. Among these 4,201 authors, five authors published more than five articles, with the maximum count being six. Furthermore, 3,849 (81.27%) authors published only one article. The result displayed that the papers of virtual community were usually generated by single author.

Table 8. The productivity analysis of authors (a)

Record Count (x)	Author(y)	X=log(x)	Y=log(y)	XY	XX
1	3849	0.000	3.585	0	0
2	248	0.301	2.394	0.720594	0.090601
3	55	0.477	1.740	0.82998	0.227529
4	29	0.602	1.462	0.880124	0.362404
5	12	0.699	1.079	0.754221	0.488601
6	6	0.778	0.778	0.605284	0.605284
7	2	0.845	0.301	0.254345	0.714025
Total	4201	3.702	11.3390	4.0445	2.4884

In this stage, the n value was calculated by Lotka's method using the following equation (2):

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - (\sum X)^2} \quad (2)$$

Applying the values from Table 8 into equation (2), we can get the value of n = -3.6791.

In the fourth stage, the value of constant C was calculated using the following equation (3):

$$C = \frac{1}{\sum_{X=1}^{p-1} \frac{1}{X^n} + \frac{1}{(n-1)(p^{n-1})} + \frac{1}{2 * p^n} + \frac{n}{24 * (p-1)^{n+1}}}, p = 7, \quad (3)$$

$$C = \frac{1}{\sum_{X=1}^6 \frac{1}{X^n} + \frac{1}{(n-1)(7^{n-1})} + \frac{1}{2 * 7^n} + \frac{n}{24 * 6^{n+1}}}, X = 1,2,3,4,5,6.$$

C = 0.903313736.

Then, apply n = -3.6791, C = 0.903313736 into equation (1) (i.e., $f(x) = \frac{0.903313736}{x^{-3.6791}}$).

Pao [24] indicated that absolute value of n should be from 1.2 to 3.8 which were formulated by the generalized Lotka's law. In order to test whether our observed value match with the theoretical value, we further used K-S test for evaluation. Finally, according to the K-S test, Table 9 found Dmax= 0.0136 (Dmax= Absolute Value |Fo(X) - Sn(X)|), if the sampling number is bigger than 35, the threshold value is 0.025 ($\frac{1.63}{4201^{0.5}} = 0.025$). Since the value of Dmax is less than the threshold value, the result matched the generalized Lotka's law, that is, the author productivity distribution data is consistent in virtual community studies.

Table 9. The productivity analysis of authors (b)

Record Count(a)	Observation by Author(s)	Accumulated Value Sn(X)	Expected Value by Author(s) %	Accumulated Value Fo(X)	Absolute Value Fo(X)-Sn(X)
1	0.9169	0.9169	0.9033	0.9033	0.0136
2	0.0652	0.9821	0.0705	0.9738	0.0083
3	0.0090	0.9911	0.0159	0.9897	0.0014
4	0.0059	0.9970	0.0055	0.9952	0.0018
5	0.0011	0.9981	0.0024	0.9976	0.0005
6	0.0019	1.0000	0.0012	0.9989	0.0011
7	0.9169	0.9169	0.9033	0.9033	0.136

5. Conclusions and Implications

5.1 Conclusions

Virtual community topic has been received a growing attention in the academic literature. According to the analysis of this study, several findings are reported as follows. First, the literature productivity on virtual community topic is still growing. Second, the main research countries which have written the most virtual community papers is located in USA, but England, Canada and Taiwan also have potential to deliver more publications in the future. Third, virtual community has been mainly applied into some subject area such as Computer Science, Information Science & Library Science, Business Economics, Education & Educational Research, and Psychology. Fourth, the result of author keyword analysis indicated some popular related research areas were Internet, e-learning and social network. Fifth, K-S test proved that the frequency indexes of author productivity distribution follow Lotka's Law. Finally, according to Bradford's Law, the three zone ratio was approximately equal to 1: 3: 3², that mean the data was consistent with Bradford's Law. In addition, the twenty eight core journals in virtual community issues were identified.

5.2 Research Implications

In conclusion, the productivity distribution related to virtual community was explored using the bibliometric methodology at this study, the findings of which will provide the following applications: First, this research on the productivity distribution of virtual community may inform researchers and scholars of current issues and development of virtual community. Second, with regard to author productivity, our study proposed a theoretical model, based on Lotka's law, for author productivity analysis of virtual community issue, serving as reference for different areas of study in the evaluation of author productivity models. Also, in order to allow researchers to gain in-depth insights, this study aimed to report the most published institutions and keep track of the growth and trend of author productivity, by which scholars in related fields are provided with more opportunities for academic communication and technological cooperation.

5.3 Limitations and Further Research

This research also identified practical limitations that offer immediate avenues for further research. First, this research could be used for further analysis in the future such as keyword content analysis of getting the deeper comprehension with this topic. Second, only one SSCI database was considered in this research. Therefore, future research could use the different journal databases such as Social Citation Index (SCI) to retrieve more in-depth issues.

6. References

- [1] Rheingold, H., *The virtual community: Homesteading on the electronic frontier*. New York: Harper Collins, 1993.
- [2] Williams, R. L., and Cothrel, J., "Four smart ways to run online communities," *Sloan Management Review*, vol. 41, pp. 81-91, 2000.
- [3] Laine, M. O. J., "Virtual communities: A bibliometric analysis," presented at the Proceedings of the 42th Hawaii International Conference on System Sciences (HICSS), 2009.
- [4] Polanco, X., *Infométrie et ingénierie de la connaissance*. In: Noyer, J-M. Ed. *Les sciences de l'information bibliométrie scientométrie infométrie*. Presses universitaires de Rennes, Rennes., 1995.
- [5] Yang, Y. X., and Tsaih, R. H., "An Investigation of Research on Evolution of Altruism using Informetric Methods and the Growing Hierarchical Self-Organizing Map," *Malaysian Journal of Library & Information Science*, vol. 15, pp. 1-17, 2010.
- [6] Chao, C. C., Yang, J. M., & Jen, W. Y., "Determining technology trends and forecasts of RFID by a historical review and bibliometric analysis from 1991 to 2005," *Technovation*, vol. 27, pp. 268-279, 2007.
- [7] Tsai, H. H., "Research trends analysis by comparing data mining and customer relationship management through bibliometric methodology," *Scientometrics*, vol. 87, pp. 425-450, 2011.
- [8] Shiau, W. L., "A profile of information systems research published in expert systems with applications from 1995 to 2008," *Expert Systems with Applications*, vol. 38, pp. 3999-4005, 2011.

- [9] Palvia, P., Pinjani, P., & Sibley, E. H., "A profile of information systems research published in Information & Management," *Information & Management*, vol. 44, no.1, pp. 1-11, 2007.
- [10] Bradford, S. C., "Sources of information on specific subjects," *Engineering*, vol. 23, pp. 85-88, 1934.
- [11] Pulgarín, A., and Gil-Leiva, I., "Bibliometric analysis of the automatic indexing literature: 1956-2000," *Information Processing & Management*, vol. 40, pp. 365-377, 2004.
- [12] Davis, P. M., "Where to spend our e-journal money? Defining a university library's core collection through citation analysis," *portal: Libraries & the Academy*, vol. 2, pp. 155-166, 2002.
- [13] Chang, S. H., Chou, C. H., & Yang, J. M., "The literature review of technology acceptance model: A study of the bibliometric distributions," presented at the Proceedings of the 14th Pacific Asia Conference on Information Systems, 2010.
- [14] Huang, C. Y., and Ho, Y. S., "Historical research on corporate governance: A bibliometric analysis," *African Journal of Business Management*, vol. 5, pp. 276-284, 2011.
- [15] Bookstein, A., "Implications of ambiguity for scientometric measurement," *Journal of the American Society for Information Science & Technology*, vol. 52, pp. 74-79, 2001.
- [16] Lotka, A. J., "The frequency distribution of scientific productivity," *Journal of the Washington Academy of Sciences*, vol. 16, pp. 317-323, 1926.
- [17] Wolfram, D., *Applied informetrics for information retrieval research*. Santa Barbara, USA: Greenwood Publishing Group, 2003.
- [18] Nicholls, P., "Empirical validation of Lotka's law," *Information Processing & Management*, vol. 22, pp. 417-419, 1986.
- [19] Kretschmer, H., and Rousseau, R., "Author iflation leads to a beakdown of Lotka's law," *Journal of the American Society for Information Science & Technology*, vol. 52, pp. 610-614, 2001.
- [20] Garfield, E., "KeyWords Plus™ takes you beyond title words. 2. Expanded journal coverage for Current Contents on Diskette, includes social and behavioral sciences," *Current Contents*, vol. 33, pp. 5-9, 1990.
- [21] Xie, S. D., Zhang, J., & Ho, Y. S., "Assessment of world aerosol research trends by bibliometric analysis," *Scientometrics*, vol. 77, pp. 113-130, 2008.
- [22] Tsai, M. Y., *The Characteristic of Informetrics and Bibliometrics*. Taiwan: Hwa-Tai bookstore, 2003.
- [23] Pao, M. L., "Lotka's law: A testing procedure," *Information Processing & Management*, vol. 21, pp. 305-320, 1985.
- [24] Pao, M. L., *Concept of information retrieve*. Englewood, Colorado: Libraries Unlimited. , 1989.
- [25] Tsai, H. M., Chi, Y. P., "Trend Analysis of Supply Chain Management by Bibliometric Methodology", *JDCTA: International Journal of Digital Content Technology and its Applications*, Vol. 5, No. 1, pp. 285-295, 2011