CrossMark

The research trends forecasted by bibliometric methodology: a case study in e-commerce from 1996 to July 2015

Hsu-Hao Tsai[1]

Abstract This study explores e-commerce (EC) research trends and forecasts applying bibliometric analysis from 1996 to July 2015 with topic as "e-commerce" in SSCI database. The bibliometric analytical technique is used to examine the topic in SSCI journals from 1996 to July 2015, we found 5429 articles with EC. This paper surveys and classifies EC articles using the eight categories for different distribution status in order to explore how EC research trends and applications have developed in this period. Besides, the paper will perform K–S test to verify the reliability of Lotka's Law. The study provides an EC roadmap to guide future research and abstract the trend information so that EC researchers can save some time to browse sources since core knowledge will be concentrated in EC core categories. In higher quality publications, it is very common for "success breeds success" phenomenon.

Keywords E-commerce · Technology trend and forecast · Bibliometric methodology

Introduction

Electronic commerce, commonly known as e-commerce or ecommerce (EC), consists of the buying and selling of products or services over electronic systems such as the internet and other related networks. The amount of trade conducted electronically has grown extraordinarily with widespread Internet usage. The use of commerce is conducted in this way, spurring and drawing on innovations in electronic funds transfer, supply chain management, Internet marketing, online transaction processing, electronic data interchange (EDI), inventory management systems, and automated data collection systems. Modern

✉ Hsu-Hao Tsai
  simon705@mail.chihlee.edu.tw

[1] Department of Commerce Technology and Management, Chihlee University of Technology, No.313, Sec. 1, Wunhua Rd., Banciao District, New Taipei 22050, Taiwan ROC

electronic commerce typically uses the World Wide Web at least at some point in the transaction's lifecycle, although it can encompass a wider range of technologies such as e-mail as well.

A large percentage of EC is conducted entirely electronically for virtual items such as access to premium content on a website, but most EC involves the transportation of physical items in some way. Online retailers are sometimes known as e-tailers and online retail is sometimes known as e-tail. Almost all big retailers have EC presence on the world wide web.

EC that is conducted between businesses is referred to as business-to-business or B2B. B2B can be open to all interested parties or limited to specific, pre-qualified participants. EC that is conducted between businesses and consumers, on the other hand, is referred to as business-to-consumer or B2C. This is the type of EC conducted by companies such as Amazon.com.

EC is generally considered to be the sales aspect of e-business. It also consists of the exchange of data to facilitate the financing and payment aspects of the business transactions.

This paper surveys EC technology trends and forecasts using bibliometric analysis from 1996 to July 2015 with topic as "e-commerce" in SSCI database. The bibliometric analytical technique is used to examine the topic in SSCI journals from 1996 to July 2015, we found 5429 articles with EC. This paper implemented and classified EC literatures using the eight categories as: publication year, citation, country/territory, institute name, document type, language, research area and source title for different distribution status in order to explore how EC technology trends and applications have developed in this period.

For verifying the analysis result, the paper will perform by the following steps to verify the reliability of Lotka's Law: (1) Collect data (2) List author and literature distribution Table (3) Calculate n value (slope) (4) Calculate c value (5) Utilizing K–S (Kolmogorov–Smirnov, K–S) test to evaluate if matched Lotka's Law. (Potter 1988)

The paper provides an EC roadmap to guide future research and abstract the trend information so that EC researchers can save some time since core knowledge will be concentrated in EC core categories. In higher quality publications, it is very common for "success breeds success" phenomenon.

Methodology and data retrieval strategy

Methodology

One of the main areas in bibliometric research concerns the application of bibliometric laws (Luor et al. 2014). The three most commonly used laws in bibliometrics are: Lotka's law of scientific productivity, Bradford's law of scatter, and Zipf's law of word occurrence.

Lotka's Law describes the frequency of publication by authors in a given field. It states that "the number (of authors) making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent" (Tsay 2003). This means that out of all the authors in a given field, 60 % will have just one publication, and 15 % will have two publications ($1/2^2$ times 0.60). 7 % of authors will have three publications ($1/3^2$ times 0.60), and so on. According to Lotka's Law of scientific productivity, only 0.6 % of the authors in a field will produce more than ten articles. Lotka's Law, when applied to large bodies of literature over a fairly long period of time,

can be accurate in general, but not statistically exact. It is often used to estimate the frequency with which authors will appear in an online catalog (Tsay 2003).

Lotka's law is generally useful for understanding the productivity patterns of authors in a bibliography (Coille 1977; Gupta 1987; Nicholls 1989; Lour et al. 2014; Pao 1985; Rao 1980; Vlachy 1978). In this article, Lotka's Law is selected to perform bibliometric analysis to check on literature record count versus accumulated authors between 1996 and July 2015 to perform author productivity inspection for discovering historical review and collecting the results for research tendency forecast in the near future. For verifying the analysis result, the paper implements K–S test to evaluate if the result matched Lotka's Law.

Data retrieval strategy

The data retrieval strategy (shown as Fig. 1) demonstrates as below:
1. Retrieving technology trends and forecasts in eight categories by Bibliometrics.
2. Retrieving the distribution of author article production by K–S test.

Based on the above information, governments and institutes may infer collective tendencies and demands of scientific researcher in EC and formulate appropriate training strategies and policies in the future.

Research results

This research is accessing the Social Science Citation Index (SSCI) on Web of Science created by ISI. The result is summarizing those 5429 paper indexes which topic is "e-commerce" from 1996 to July 2015, shown as Fig. 2. Obviously, the literature production of EC is rising since 1998 and citation is also increasing steadily and gradually by every year. It indicates the research of EC is very popular in the highly exploration period, referred to Fig. 3. The research of EC reached the highest record in 2014.

By viewing on Table 1, they displayed that the distribution of country/territory from 1996 to July 2015, USA is a champion with 1954 record counts (35.99 %), following by People R. China, Taiwan, England and South Korea which achieved the record counts as 596(10.98 %), 433(7.98 %), 380(7.00 %) and 283(5.21 %) oppositely. For the distribution of institution name as shown in Table 1, City University of Hong Kong is a champion with
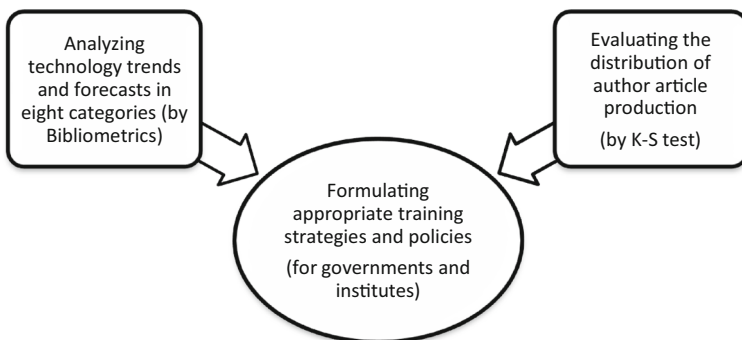

Fig. 1 Data retrieval strategy diagram

Fig. 2 The tendency chart of literature growth of EC (*source* SSCI database)
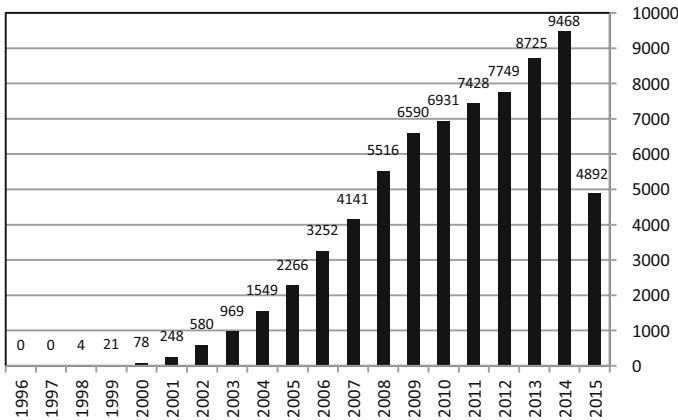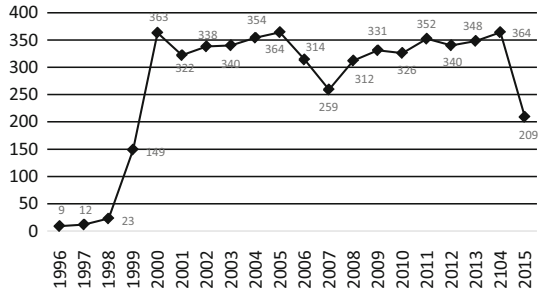

Fig. 3 Citation of EC in each year (*source* SSCI database)

83 record counts (1.53 %), following by University of Minnesota, Hong Kong Polytech University, University of Wisconsin and National University of Singapore which achieved the record counts as 60(1.11 %), 53(0.98 %), 52(0.96 %) and 50(0.92 %) oppositely.

From the Table 2, it indicated that the most popular publication document type is article (4845 record counts, 89.24 %), following by proceedings paper (756 record counts, 13.93 %) and editorial material (211 record counts, 5.98 %). The most popular language for literature is using English (5345 record counts, 98.45 %) in the research domain of EC, following by German (41 record counts, 0.76 %), see the following Table 2. The result shows that article document type and English language are still the main trend in EC research domain.

In the Table 3, it is important to summarize the trend information for EC researchers since core knowledge will be concentrated in core categories and to get understanding about the distribution of top 25 subjects in future research trends and forecasts. Focus on the right categories, researchers will catch the core research information. The top five ranking of research subjects are computer science (2829 record counts, 52.11 %), following by business economics (1764 record counts, 32.49 %), engineering (1055 record counts, 19.43 %), information science and library science (663 record counts, 12.21 %), and operations research and management science (520 record counts, 9.58 %).

Moreover, it also discovered top 25 important research sources for EC researchers. According to these core Journals, it will be easy to recognize the developing trends and

Table 1 Distribution of top 25 countries/territories and institutes from 1996 to July 2015

Ranking	Country/territory	Record count	% of 5429	Institute	Record count	% of 5429
1	USA	1954	35.99	City Univ. Hong Kong	83	1.53
2	Peoples R China	596	10.98	Univ. Minnesota	60	1.11
3	Taiwan	433	7.98	Hong Kong Polytech Univ.	53	0.98
4	England	380	7.00	Univ. Wisconsin	52	0.96
5	South Korea	283	5.21	Natl. Univ. Singapore	50	0.92
6	Canada	268	4.94	Univ. Illinois	47	0.87
7	Spain	239	4.40	Univ. Maryland	45	0.83
8	Australia	197	3.63	Georgia State Univ.	43	0.79
9	Germany	195	3.59	IBM Corp.	41	0.76
10	Italy	161	2.97	Penn State Univ.	40	0.74
11	Singapore	109	2.01	Nanyang Technol. Univ.	39	0.72
12	Netherlands	104	1.92	Purdue Univ.	39	0.72
13	India	89	1.64	Seoul Natl. Univ.	37	0.68
14	Greece	85	1.57	Natl. Sun Yat Sen Univ.	36	0.66
15	France	80	1.47	Carnegie Mellon Univ.	35	0.65
16	Switzerland	59	1.09	Univ. British Columbia	34	0.63
17	Japan	58	1.07	Univ. N Carolina	34	0.63
18	Austria	57	1.05	Korea Adv. Inst. Sci. Technol.	33	0.61
19	Turkey	57	1.05	Michigan State Univ.	33	0.61
20	Brazil	52	0.96	Univ. Michigan	33	0.61
21	Malaysia	40	0.74	Chinese Univ. Hong Kong	32	0.59
22	Israel	39	0.72	Natl. Chiao Tung Niv.	32	0.59
23	Sweden	38	0.70	Arizona State Univ.	31	0.57
24	Iran	37	0.68	Natl. Chung Cheng Univ.	31	0.57
25	Scotland	36	0.66	Univ. Hong Kong	30	0.55
	Others	64	8.436	Other	2692	8.565

Resource count from SSCI database

contents in EC. The top five ranking of research sources are *Lecture Notes in computer science* (400 record counts, 7.37 %), following by *Expert Systems with Applications* (123 record counts, 2.27 %), *International Journal of Electronic Commerce* (123 record counts, 2.27 %), *Electronic Commerce Research and Applications* (120 record counts, 2.21 %), and *Decision Support Systems* (114 record counts, 2.10 %).

The literatures productivity analysis of EC by Lotka's Law

The section will perform by the following steps to verify the reliability of Lotka's Law:

1. Collect data
2. List author and literature distribution Table
3. Calculate *n* value (slope)
4. Calculate *c* value

Table 2 Distribution of top 25 document types and languages from 1996 to July 2015

Document type	Record count	% of 5429	Language	Record count	% of 5429
Article	4845	89.24	English	5345	98.45
Proceedings paper	756	13.93	German	41	0.76
Editorial material	211	3.89	Spanish	11	0.20
Review	151	2.78	French	9	0.17
News item	142	2.62	Portuguese	5	0.09
Book review	48	0.88	Czech	4	0.07
Meeting abstract	18	0.33	Romanian	3	0.06
Letter	9	0.17	Chinese	2	0.04
Correction	4	0.07	Polish	2	0.04
Book chapter	2	0.04	Arabic	1	0.02
Software review	1	0.02	Finnish	1	0.02
Total	5429	100	Hungarian	1	0.02
			Lithuanian	1	0.02
			Russian	1	0.02
			Swedish	1	0.02
			Turkish	1	0.02
			Total	5429	100

Resource count from SSCI database

5. Utilizing K–S (Kolmogorov–Smirnov, K–S) est to evaluate if the result matched Lotka's Law

1. Collect data and

2. List author and literature distribution table:

In Table 4, it calculated the author quantity by the equality method from 9284 literatures which retrieved by index on SSCI. Thus, it is obtained altogether 6175 of authors on research aspect of EC.

3. Calculate n value (slope):

By the result of calculation on Table 5, it could bring into Lotka's Law's equation as below to calculate n value:

$$n = \frac{N \sum XY - \sum X \sum Y}{N \sum X^2 - \left(\sum X\right)^2} \quad (1)$$

We can refer Table 5 for the values in equation, then $n = -1.310626$

4. Calculate c value:

After that, we also found $c = 0.26187474$, the equation is shown as below:

$$c = \frac{1}{\sum_1^{p-1} \frac{1}{x^n} + \frac{1}{(n-1)p^{n-1}} + \frac{1}{2p^n} + \frac{n}{24(p-1)^{n+1}}} \quad (2)$$

$P = 20$, $x = 1$–19

When we got $n = -1.310626$ and $c = 0.26187474$, it explored:

Table 3 Distribution of top 25 research areas and source titles from 1996 to July 2015

Ranking	Subject area	Record count	% of 5429	Source title	Record count	% of 5429
1	Computer science	2829	52.11	Lecture Notes in Computer Science	400	7.37
2	Business economics	1764	32.49	Expert Systems with Applications	123	2.27
3	Engineering	1055	19.43	International Journal of Electronic Commerce	123	2.27
4	Information science library science	663	12.21	Electronic Commerce Research and Applications	120	2.21
5	Operations research management science	520	9.58	Decision Support Systems	114	2.10
6	Telecommunications	342	6.30	Lecture Notes in Artificial Intelligence	87	1.60
7	Psychology	156	2.87	Chemical Week	83	1.53
8	Social sciences other topics	116	2.14	Journal of Management Information Systems	74	1.36
9	Government law	96	1.77	Information Management	73	1.35
10	Mathematics	91	1.68	Industrial Management Data Systems	69	1.27
11	Communication	65	1.20	Journal of Computer Information Systems	67	1.23
12	Environmental sciences ecology	64	1.18	International Journal of Information Management	59	1.09
13	Automation control systems	60	1.11	Journal of Electronic Commerce Research	57	1.05
14	Chemistry	49	0.90	Computers in Human Behavior	50	0.92
15	Geography	47	0.87	Internet Research	47	0.87
16	Public administration	45	0.83	Journal of Organizational Computing and Electronic Commerce	46	0.85
17	Science technology other topics	42	0.77	European Journal of Operational Research	45	0.83
18	Transportation	42	0.77	Communications of the ACM	44	0.81
19	Materials science	35	0.65	Information Systems Research	44	0.81
20	Health care sciences services	27	0.50	Computer Law Security Review	43	0.79
21	Energy fuels	25	0.46	Behaviour Information Technology	41	0.76
22	Physics	23	0.42	Electronic Commerce Research	41	0.76
23	Agriculture	20	0.37	Internet Research Electronic Networking Applications and Policy	40	0.74
24	EDUcation educational research	18	0.33	Mis Quarterly	40	0.74
25	Instruments instrumentation	18	0.33	Online Information Review	36	0.66

Resource count from SSCI database

Table 4 Calculation of author productivity of EC

Record count	Author (s)	Record count	Accumulated record	Accumulated record (%)	Accumulated author (s)	Accumulated author(s) %
106	1	106	106	1.14	1	0.02
22	2	44	150	1.62	3	0.05
20	1	20	170	1.83	4	0.06
17	2	34	204	2.20	6	0.10
15	4	60	264	2.84	10	0.16
14	4	56	320	3.45	10	0.16
13	4	52	372	4.01	14	0.23
12	8	96	468	5.04	22	0.36
11	2	22	490	5.28	24	0.39
10	11	110	600	6.46	35	0.57
9	14	126	726	7.82	49	0.79
8	13	104	830	8.94	62	1.00
7	31	217	1047	11.28	93	1.51
6	53	318	1365	14.70	146	2.37
5	93	465	1830	19.71	239	3.87
4	98	392	2222	23.93	337	5.46
3	226	678	2900	31.24	563	9.12
2	776	1552	4452	47.95	1339	21.70
1	4832	4832	9284	100.00	6175	100.00

Resource calculated by authors

$$f(x) = 0.26187474/x^{1.310626} \tag{3}$$

Referring to the data from Table 6, authors with only one literature is 78.30 % (100 % − 21.70 % = 78.30 %), which is close to primitive c value 78.25 % generated by Lotka's law. After that, it can follow the calculation to get n and c value by the least squares law, carry onto the further proceeding examination for Lotka's law compliance.

According to Pao (1989) suggestion, the absolute value of n should be between 1.2 and 3.8 which formed by the generalized Lotka's Law, the result is matched the reference data by observation. The distribution chart is shown as Fig. 4.

5. Utilizing K–S (Kolmogorov–Smirnov, K–S) test to evaluate if matched Lotka's Law:

For discussing the value of n and c, we got $n = -1.310626$ and $c = 0.26187474$ generated by Lotka's Law, the result demonstrated that the EC literature author distribution and the primitive Lotka's Law are matched approximately, refer to Fig. 4. In order to examine whether the theoretical value and the observation value are tallied, the paper use K–S test to evaluate the suitability of Lotka's Law. Regarding the n and c value which gained by the formula, it is possible to calculate the expected value and the accumulation value of author, following by K–S test examination.

From Table 6, we can find $D_{\max}$ [$D_{\max}$ = ABS Value Fo(x) − Sn(x)] = 0.5407. According to K–S test, the threshold value is:

Table 5 Calculation of the exponent n for EC

Rord count (x)	Author(y)	X = log(x)	Y = log(y)	XY	XX
106	1	2.03	0.00	0.00	4.10
22	2	1.34	0.30	0.40	1.80
20	1	1.30	0.00	0.00	1.69
17	2	1.23	0.30	0.37	1.51
15	4	1.18	0.60	0.71	1.38
14	4	1.15	0.60	0.69	1.31
13	4	1.11	0.60	0.67	1.24
12	8	1.08	0.90	0.97	1.16
11	2	1.04	0.30	0.31	1.08
10	11	1.00	1.04	1.04	1.00
9	14	0.95	1.15	1.09	0.91
8	13	0.90	1.11	1.01	0.82
7	31	0.85	1.49	1.26	0.71
6	53	0.78	1.72	1.34	0.61
5	93	0.70	1.97	1.38	0.49
4	98	0.60	1.99	1.20	0.36
3	226	0.48	2.35	1.12	0.23
2	776	0.30	2.89	0.87	0.09
1	4832	0.00	3.68	0.00	0.00
Total	6175	170.06	23.02	14.44	289.99

Resource calculated by authors

$$1.63/\sqrt{6175} = 0.2097 \tag{4}$$

Because D_{max} is bigger than the K–S test threshold value, the result also indicated that the distribution of author productivity is not matched by the Lotka's Law. The consequence means the Lotka's Law is not suitable for the literature author productivity distribution in EC research domain (Potter 1981).

Conclusion

EC is one of fast growing research topics in recently years, the trend forecast of this research field by each kind of literature characteristic and author productivity distribution are in growing period. In this study, it demonstrated that the current EC literatures are still continuously to grow, the main research development facility with delivered the largest production is USA, but Peoples R China, Taiwan, England, South Korea and Canada also have potential to deliver more literatures in the future. In addition, the result shows that article document type and English language are still the main trend in EC research domain.

After checked by K–S test, the distribution of frequency indexes of author productivity in EC is not suitable for Lotka's Law. The reason why EC does not match is that the number of authors who publish only one article is too large; as a result, the difference between the observed value and the expected value becomes greater than the K–S test

Table 6 The K–S test for EC

Record count	Author(s)	Observation by author(s)	Accumulated value Sn(x)	Expected value by author(s)	Accumulated value Fo(x)	ABS value |Fo(x) − Sn(x)|
1	4832	0.7825	0.7825	0.2619	0.2619	0.5206
2	776	0.1257	0.9082	0.1056	0.3674	0.5407
3	226	0.0366	0.9448	0.0621	0.4295	0.5153
4	98	0.0159	0.9606	0.0426	0.4721	0.4886
5	93	0.0151	0.9757	0.0318	0.5038	0.4719
6	53	0.0086	0.9843	0.0250	0.5288	0.4554
7	31	0.0050	0.9893	0.0204	0.5493	0.4400
8	13	0.0021	0.9914	0.0172	0.5664	0.4250
9	14	0.0023	0.9937	0.0147	0.5812	0.4125
10	11	0.0018	0.9955	0.0128	0.5940	0.4015
11	2	0.0003	0.9958	0.0113	0.6053	0.3905
12	8	0.0013	0.9971	0.0101	0.6153	0.3817
13	4	0.0006	0.9977	0.0091	0.6244	0.3733
14	4	0.0006	0.9984	0.0082	0.6327	0.3657
15	4	0.0006	0.9990	0.0075	0.6402	0.3588
17	2	0.0003	0.9994	0.0064	0.6535	0.3459
20	1	0.0002	0.9995	0.0052	0.6701	0.3294
22	2	0.0003	0.9998	0.0046	0.6795	0.3203
106	1	0.0002	1.0000	0.0006	0.8023	0.1977

Resource calculated by authors

Fig. 4 Distribution of literature productivity of author on EC
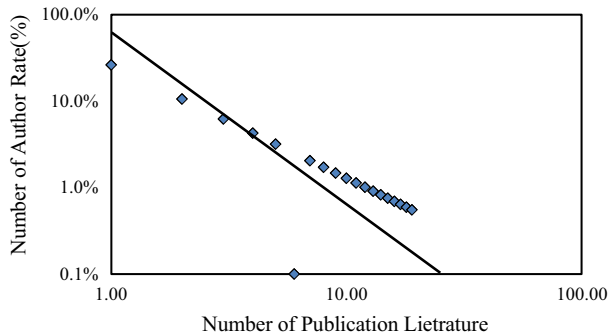
critical value. This outcome causes the EC distribution to diverge from the slope of Lotka's law.

The most relevant disciplines for EC subject categories provided by computer science, business economics, engineering, information science and library science, and operations research and management science. In addition, telecommunications, psychology, social sciences and other topics, government law and mathematics also have the potential in EC research domain.

The most important journals for EC source titles provided by *Lecture Notes in Computer Science*, *Expert Systems with Applications*, *International Journal of Electronic Commerce*, *Electronic Commerce Research and Applications*, and *Decision Support Systems*. Furthermore, *Lecture Notes in Artificial Intelligence*, *Chemical Week*, *Journal of Management Information Systems*, *Information Management*, and *Industrial Management and Data Systems*.

Focus on the right categories, the researchers may catch the valuable information and research trends. This study provides an EC roadmap to lead future research and to abstract the tendencies. Therefore, EC researchers can save time to browse sources since core knowledge will be concentrated in EC core categories. In higher quality publications, it is very common for "success breeds success" phenomenon.

Implementation

The study results can also support governments and institutes to recognize the scientific research trends and forecasts of EC, and to realize the developing scale of research in EC through analyzing the increases of the article author. The resources are always conflicted, especially for emerging and developing countries, and small and medium enterprises. According to the above information, it will be easy for governments and institutes to organize the proper training strategies and policies in the future.

References

Coille, R. C. (1977). Lotka's frequency distribution of scientific productivity. *Journal of American Society for Information Science, 28*, 366–370.

Gupta, D. K. (1987). Lotka's Law and productivity of entomological research in Nigeria for the period 1900–1973. *Scientometrics, 12*, 33–46.

Lour, T., Lu, H. P., Yu, H., & Chang, K. (2014). Trends in and contributions to entrepreneurship research: a broad review of literature from 1996 to June 2012. *Scientometrics, 99*, 353–369.

Nicholls, P. T. (1989). Bibiometric modeling processes and empirical validity of Lotka's law. *Journal of American Society for Information Science, 40*(6), 379–385.

Pao, M. L. (1985). Lotka's law, a testing procedure. *Information Processing and Management, 21*, 305–320.

Pao, M. L. (1989). *Concept of information retrieve*. Englewood, Colorado: Libraries Unlimited.

Potter, W. G. (1981). Lotka's Law revisited. *Library Trends., 30*(1), 21–39.

Potter, W. G. (1988). Of Making Many Books There Bibliometrics and Libraries. *The Journal of Academic Librarianship, 14*, 238a–238c.

Rao, I. K. R. (1980). The distribution of scientific productivity and social change. *Journal of American Society for Information Science, 31*, 111–122.

Tsay, M. Y. (2003). *The characteristic of informetric and bibliometric*. Taiwan: Hwa- Tai.

Vlachy, J. (1978). Frequency distribution of scientific performance: A bibliography of Lotka's law and related phenomena. *Scientometrics, 1*, 109–130.