# An investigation of research on evolution of altruism using informetric methods and the growing hierarchical self-organizing map

**Yu-Hsiang Yang** and **Rua-Huan Tsaih**
Dept. of Management Information Systems, National Chengchi University,
Room 261036, Fl. 10, College of Commerce Building,
64, Sec. 2, Chihnan Rd., Taipei, Taiwan 116, R.O.C
e-mail: yuxiang1001@gmail.com

## ABSTRACT

*The purpose of this study was to investigate the characteristics of research related to evolution of altruism from 1971 to 2009 within the science citation index expanded (SCIE) and the social science citation index (SSCI) databases. This study showed how the growth of research related to evolution of altruism is a well known phenomenon, that statistics of the Bradford's Law identified ten core altruism-related journals, and that the altruism-related data does not fit Lotka's law. We applied Growing Hierarchical Self-Organizing Map (GHSOM), a text-mining Neural Networks tool, to obtain a hierarchical topic map. The topic map illustrated the delicate intertwining of subject areas and provided a more explicit illustration of the concepts within each subject area. Furthermore, the result of the topic map also reflects that evolutionary psychology based on neuroscience and other related discipline will play an importance role in the future exploring into the in-depth motivation of altruism.*

**Keywords:** Altruism; Growing Hierarchical Self-Organizing Map; Informetrics: Bibliometrics; Scientometrics

## INTRODUCTION

This study investigates the characteristics of articles relating to studies of altruism, from 1971 to 2009, found in the Science Citation Index Expanded (SCIE) and the Social Science Citation Index (SSCI) databases. The term "altruism" is defined as a moral principle emphasizing the importance of placing the welfare and happiness of others before that of oneself, or of sacrificing oneself for the benefit of others. This definition was offered by Auguste Comte (1798-1857), the founder of sociology (Weiner et al. 1993). It is the purist forms of prosocial behaviour occur when someone acts to help another person. It is also a traditional virtue in many cultures and a core aspect of various religions such as Buddhism, Islam, and Christianity. However, the existence of altruism represents a key problem in Darwin's theory of evolution. Survival of the fittest failed to provide a biological explanation of selfless altruism from an evolutionary perspective, or "examine the biology of selfishness and altruism" (Dawkins 2006). Early scholars sought to discover how and why selflessness could have evolved. For instance, the biological explanation was expanded to the genetic kinship theory (Hamilton 1964), reciprocal altruism (Trivers 1971), and group selection of sociobiology (Wilson 1975). Since these works were applied to many other disciplines such as economics or sociology, altruism has become an interdisciplinary issue. On the other hand, a number of scholars have delivered specific assessments of altruism in areas such as psychology (Krebs 1970; Sharabany and Bar-Tal 1982) or sociology (Piliavin and Charng 1990). Until now, no informetric analysis of altruism has been conducted, despite a plethora of research successfully applying informetric analysis to a number of multidisciplinary fields, such as Tsunami (Sagar et al. 2010), transport phenomenon (Tsay and Lin 2009), and Southeast Asian chemical engineering (Yin 2009).

In this study, we applied informetric analysis to research related to evolution of altruism in the SCIE and SSCI databases, to gain a better understanding of the quantitative aspects of recorded data and discover features of research embedded in the recorded data. Specifically, the objectives of this study were:

(a) to explore the growth of published research related to evolution of altruism;
(b) to determine the core journals that contained a substantial portion of the research related to evolution of altruism;
(c) to determine the productivity distribution of authors on this subject;
(d) to identify countries, institutions, and authors contributing the bulk of the published articles related to evolution of altruism, as well as the most cited articles;
(e) to reveal the major topics or conceptual interrelations of research related to evolution of altruism.

Standard informetric indicators such as the number of papers, number of authors, productivity by country, institutional collaboration, and most cited articles were analyzed. Lotka's Law (Nicholls 1986; Pao 1986; Potter 1981; Wolfram 2003) was applied to analyze author productivity and Bradford's Law (Wolfram 2003), to compile lists of the core journals publishing in the field of altruism. To reveal the major topics and conceptual interrelations of articles related to evolution of altruism, we adopted the Growing Hierarchical Self-Organizing Map (GHSOM) approach (Dittenbach et al. 2002; Rauber et al. 2002) to cluster the conceptual topics into a hierarchical representation of dynamic 2-dimentional interrelated structures within the data.

## LITERATURE REVIEW

"Informetrics" was used as a generic term to connote the "use and development of a variety of measures to study and analyze several properties of information in general and documents in particular" (Kawatra 2000, p.43). Obviously, informetrics covers bibliometrics and scientometrics and seeks to develop statistical, mathematical and information systematic techniques to evaluate and improve the efficiency of information services and their uses (Kawatra, 2000). Informetrics, bibliometrics and scientometrics also refer to component fields related to the research of the dynamics of disciplines as reflected in the production of their studies. Areas of study range from charting changes in the output of a scholarly field through time and across countries, to the information collection problem of maintaining control of the output, and to the publication productivity of authors, institutions and journals (Hood and Wilson 2001).

Lotka's Law (with regard to the distributed productivity of authors) was often mentioned in conjunction with Bradford's Law (about the scattering of subjects within journals). These laws are often considered the best models of research resources available in Library and

Information Sciences (Wolfram 2003). In 1926, Alfred J. Lotka first postulated the inverse square law relating the authors of published papers to the number of papers written by each author. Using data specifically represented in the decennial index of Chemical Abstracts and Auerbach's Geschichtstafeln der Physik as the name index, Lotka plotted the number of authors against the number of contributions made by each author, on a logarithmic scale. He proposed that these points were closely clustered along a straight line with a constant slope of approximately negative two. The validity of this law has been proven regarding the productivity patterns of chemists, physicists, mathematicians and econometricians (Krisciunas 1977; Nicholls 1986; Potter 1981; Wolfram 2003). Lotka's inverse square law of scientific productivity has been shown to fit data drawn from several widely varying time periods and disciplines (Allison and Stewart 1974).

Samuel C. Bradford introduced Bradford's Law in 1934. He had observed that ranked journals could be grouped into categories, called Bradford zones, and each zone contained approximately the same number of articles, with an increase in the numbers of journals in each subsequent zone. The first zone is known as the core zone containing a small number of highly productive journals. The ratio between the number of journals in subsequent zones was observed to be roughly $1 : n : n^2 : \ldots$, where n refers to Bradford multiplier (Wolfram 2003).

Since Price (1965) first proposed the possibility of dynamic mapping using the scientific method, research in bibliometrics, scientometrics and informetrics has developed techniques to analyze data sets from within publications (Leydesdorff 1987). Most early works in this field focused on identifying networks (or clusters) of authors, papers, or references. Based on the nature of words, which are important carriers of scientific concepts, ideas and knowledge (Van Raan and Tijssen 1993), co-word analysis was also adopted to identify semantic themes (Boyack et al. 2005). Co-word analysis simplifies and projects data into specific visual representations while maintaining the essential information contained within it.

Noyons (2001) suggested that informetric mapping of science appeared to have experienced a revival, due to increased interest in information technology, since the mid-1990s. Many studies, such as (Chau et al. 2006; Ding et al. 2001; Grupp and Schmoch 1992; Hassan 2003; Noyons 2001; Noyons and van Raan 1998)) have applied informetric maps using co-word analysis to visualize cognitive structures, based on scientific topics, as well as the relationships linking them. For example, clustering major topics of a large collection of documents based on their content and providing a topical landscape of a field. In particular, Noyons and van Raan (1998) adopted the Self-Organizing Map (SOM) technique (Kohonen 1982) to apply co-word approach to scientific mapping (i.e. the organization of science based topics). Self-Organizing Maps were designed according to the concept of unsupervised artificial neural networks to process high-dimensional data and provide visual results (Kohonen 1982; Kohonen et al. 2000; Noyons and van Raan 1998). However, SOM requires a predefined number of nodes (neural processing units) and implements a static architecture. These nodes result in a representation of hierarchical relations with limited capability.

Growing Hierarchical Self-Organizing Map (GHSOM) approach (Dittenbach et al. 2002; Rauber et al. 2002) was developed to overcome these limitations, and is often applied in field the information extraction (Dittenbach et al. 2002; Li and Chang 2009; Rauber et al. 2002; Shih et al. 2008; Tsaih et al. 2009). GHSOM is based on the characteristic of SOM, but it can automatically grow its own multi-layer hierarchical structure, in which each layer encompasses a number of SOMs, as shown in Figure 1. Furthermore, Shih et al (2008) and Li and Chang (2009) proposed a layered knowledge-map using the clustering of keyterms through GHSOM. This is an updated version of SOM, enabling the visualization of hierarchical topic maps.

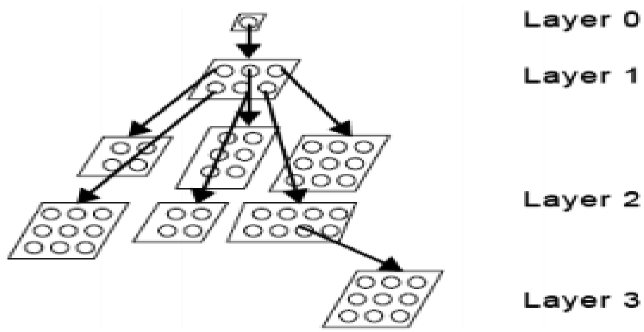



Figure 1**:** Structures of GHSOM (Rauber et al. 2002)

## DATA

The dataset used in this study come from the SCIE and SSCI databases of the Web of Science created by the Institute for Scientific Information (ISI). SCIE is a multidisciplinary index to the journal research of the sciences. It fully indexes over 6,650 major journals across 150 scientific disciplines and includes all cited references captured from indexed articles. SSCI fully indexes over 1,950 journals across 50 social sciences disciplines. It also indexes individually selected, relevant items from over 3,300 of the world's leading scientific and technical journals [1]. Although other databases such as Compendex, EngIndex/FS, or Applied Science and Technology ABS, are also available for informetric analysis, yet SCIE and SSCI databases are adopted for they are recognized as the leading English-language supplier of services providing access to the published information in the multidiscipline fields of natural sciences and social sciences.

An empirical search command was used by "(Topic=(altruism) OR Topic=("altruist* behavio*") OR Topic=("helping behavio*") OR Topic=("prosocial behavio*")) AND Topic=(evolution*)" refined by Document Type= (ARTICLE OR REVIEW) "to retrieve data related to evolution of altruism and evolution. The documents specifically included articles or reviews in the study. Book reviews, papers of proceeding, letters, notes, meeting abstracts were not taken into consideration. A total of 1,348 papers related to evolution of altruism and published between 1971 and 2009 were retrieved from the SCIE and SSCI databases.

## RESULTS

### Overview of Productivity

Figure 2 shows that a large number of research papers published in recent years (2006-2009) have been catalogued in the SCIE and SSCI databases, with distribution rates of 113 (8.4%), 138 (9.7%), 152 (11.3%) and 146 (10.6%) amongst the total number of papers,

[1] SCIE and SSCI information, retrieved August, 19, 2010 from http://images.isiknowledge.com/WOKRS49B3/help/WOS/h_database.html.

respectively. It also shows that a trend of the growth begun in 1991. Figure 3 shows that the number of citations of published altruism-related papers of each year has been increasing. Clearly, the topic of altruism has received a great deal of attention from researchers in the fields of social sciences.

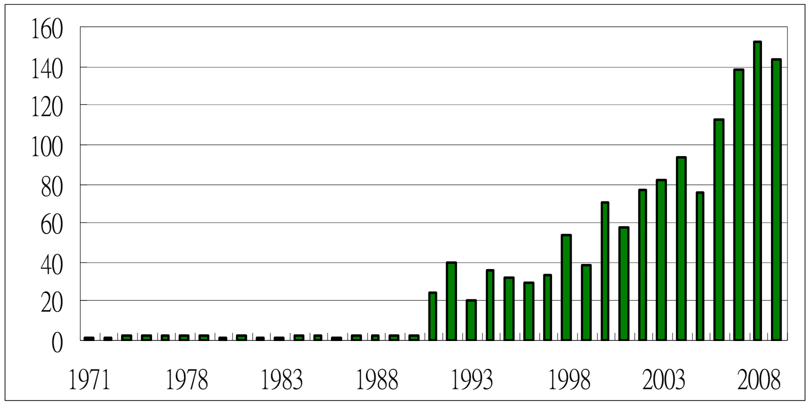



Figure 2: The number of published papers on the topic of altruism of each year from 1971 to 2009.

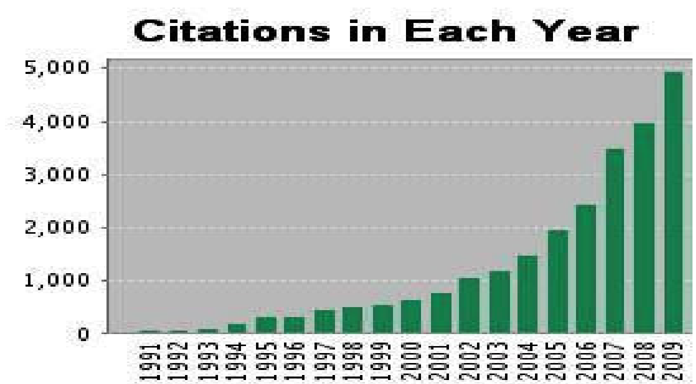



Figure 3: The number of citations of published altruism-related papers of each year

The ten countries ranked as the top countries of published altruism-related papers in the SCIE and SSCI databases are illustrated in Figure 4, which shows that the USA is the dominant country, followed by England, Canada, Germany and Switzerland. Table 1 presents a more detailed account of the top 10 academic institutions, by which indexed papers were submitted, with University of Cambridge, Harvard University, and the University of Edinburgh as the top most productive institutions. It can be observed that most of the institutions are from the USA.

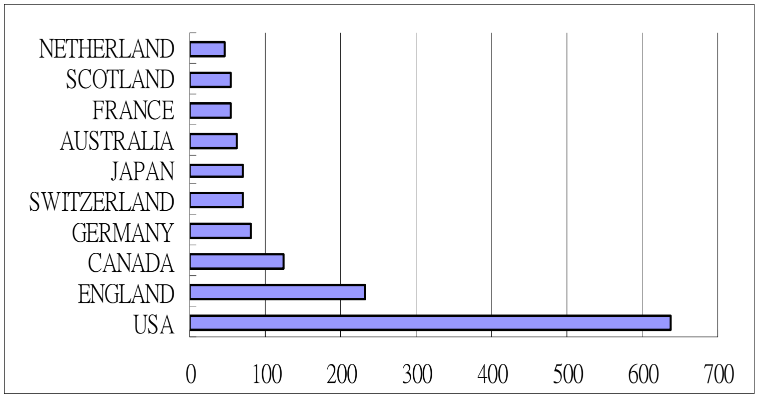



Figure 4: The top 10 most productive countries of published altruism-related papers.

Table 1: Top 10 most productive institutes for publications

| Rank | Institution | NoA[1] | % | CC%[2] | Country |
|---|---|---|---|---|---|
| 1 | Univ. Cambridge | 58 | 4.30% | 25.00% | England |
| 2 | Harvard Univ. | 49 | 3.64% | 7.69% | USA |
| 3 | Univ. Edinburgh | 34 | 2.52% | 62.96% | Scotland |
| 4 | Univ. British Columbia | 29 | 2.15% | 23.20% | Canada |
| 5 | Univ. Sheffield | 29 | 2.15% | 12.50% | England |
| 6 | Univ. Oxford | 28 | 2.08% | 4.40% | USA |
| 7 | Stanford Univ | 27 | 2.00% | 4.24% | USA |
| 8 | SUNY Binghamton | 26 | 1.93% | 4.08% | USA |
| 9 | Univ. Arizona | 26 | 1.93% | 4.08% | USA |
| 10 | Univ. Calif. Los Angeles | 23 | 1.71% | 3.61% | USA |

[1] NoA: No. of article; [2] CC %: comprising % of the country

Table 2 illustrates the output of authors who have published more than or equal to 14 papers in the altruism-related research between 1971 and 2009. The three most productive authors were Wilson, DS, West, SA, and Lehmann, L. The data indicates that the corresponding ratios for England and Scotland were much higher than the rates for the USA, indicating that among the authors in those countries, research related to evolution of altruism dominated academic research. It was observed that biology was the subject area most likely to have authors listed in Table 2.

Table 2: The most productive authors of altruism-related publication

| Author | NoA[1] | % | %C[2] | Country | Institution | Subject area |
|---|---|---|---|---|---|---|
| Wilson, DS, | 25 | 1.9% | 4% | USA | SUNY Binghamton | Biology |
| West, SA, | 20 | 1.5% | 9% | England | Univ. Oxford | Zoology |
| Lehmann, L | 18 | 1.3% | 3% | USA | Univ. Stanford | Biology |
| Nowak, MA | 18 | 1.3% | 3% | USA | Harvard Univ. | Mathematical Biol. |
| Griffin, AS | 16 | 1.2% | 7% | England | Univ. Oxford | Zoology |
| Dugatkin, LA | 15 | 1.1% | 2% | USA | Univ. Louisville | Biology |
| Gardner, A | 14 | 1.0% | 26% | Scotland | Univ. Edinburgh | Evolutionary Biol. |
| Queller, DC | 14 | 1.0% | 2% | USA | Rice Univ. | Evolutionary Biol. |

[1] NoA: No. of article; [2] % C: % at his country.

Figure 5 provides the top ten subject areas in which altruism was most widely studied, within the SCIE and SSCI databases. The most highly ranked subject area was ecology, followed by evolutionary biology and biology related to evolution of altruism.

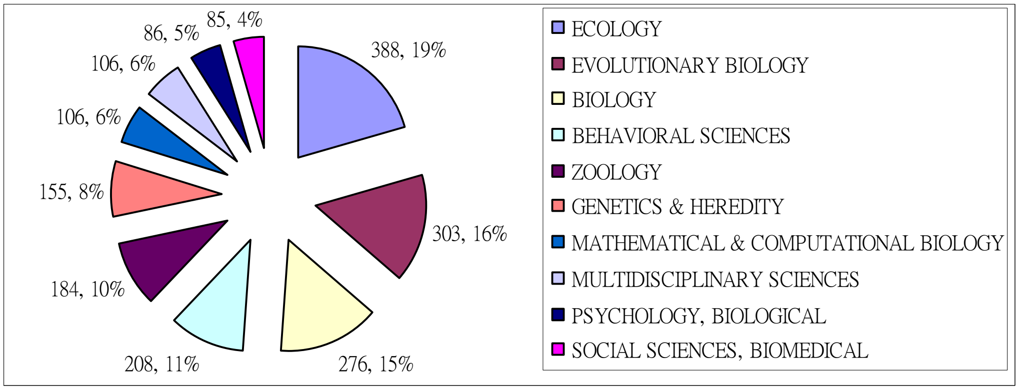



Figure 5: The top 10 subject areas for altruism-related articles

Table 3 shows the 10 altruism-related articles receiving the most citations. The results show how Trivers (1971) was an icon in altruism; however, if we take into account the average number of citations per year, the work of Fehr, E. and Gachter, S. was more influential than that of Trivers (1971). The four of the 10 articles were from *Nature*. In addition, Ernst Fehr had the two most cited altruism-related articles.

Table 3: The 10 most cited altruism-related articles (Data retrieved on September 3, 2010)

| Articles | Authors | Journal title | Year | TC[1] | ACPY[2] |
|---|---|---|---|---|---|
| Evolution of reciprocal altruism | Trivers, R. L | *Quarterly Review of Biology* | 1971 | 2,411 | 60 |
| Altruistic punishment in humans | Fehr, E. & Gachter, S. | *Nature* | 2002 | 591 | 66 |
| Evolution of indirect reciprocity by image scoring | Nowak, M. A. & Sigmund K. | *Nature* | 1998 | 447 | 34 |
| The nature of human altruism | Fehr, E. & Fischbacher, U. | *Nature* | 2003 | 362 | 45 |
| Empathy: Its ultimate and proximate bases | Preston S. D.& de Waal F. B. M. | *Behavioural and Brain Sciences* | 2002 | 361 | 45 |
| Punishment allows the evolution of cooperation ( or anything else) in sizable groups | Boyd, R. & Richerson P. J. | *Ethology and Sociobiology* | 1992 | 295 | 15 |
| Alternate routes to sociality in jays - with a theory for evolution of altruism and communal breeding | Brown J. L. | *American Zoologist* | 1974 | 285 | 8 |
| Evolution of helping behaviour in cooperatively breeding birds | Cockburn A | *Annual Review of Ecology and Systematics* | 1998 | 279 | 23 |
| Punishment in animal societies | Cluttonbrock T. H. & Parker G. A. | *Nature* | 1995 | 278 | 17 |
| A neural basis for social cooperation | Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. | *Neuron* | 2002 | 275 | 31 |

[1] TC: times cited; [2] ACPY: Average Citations per Year

## Bradford's Law and Journal Research

The 1,348 altruism-related papers referred to in this study were circulated in 353 journals. Among them, 213 journals publish only one altruism-related article. The Bradford's law has been widely employed to study journal research distribution. Brookes (1973) theorized the Bradford-Zipf's S graph to interpret the initial concave curve of the Bradford distribution as representation of three higher density of the nuclear zone. Journals in the nuclear zone constitute the core journals. Figure 6 illustrates the Bradford-Zipf plot (e.g. the cumulative number of papers for each journal against the logarithm of its ranks) for journal research related to evolution of altruism. Obviously, the Figure does not show the typical S-shape for the Bradford-Zipf plot. Nevertheless, the approximately linear portion appears after the journal ranks of about 10. The top 10 journals located within the initial concave curve portion of the Bradford-Zipf plot may be considered as the core journals (contributing 463 articles about one-third of the total as shown in Appendix 1) in the altruism-related research. The remaining altruism-related research is dispersed to 343 journals.

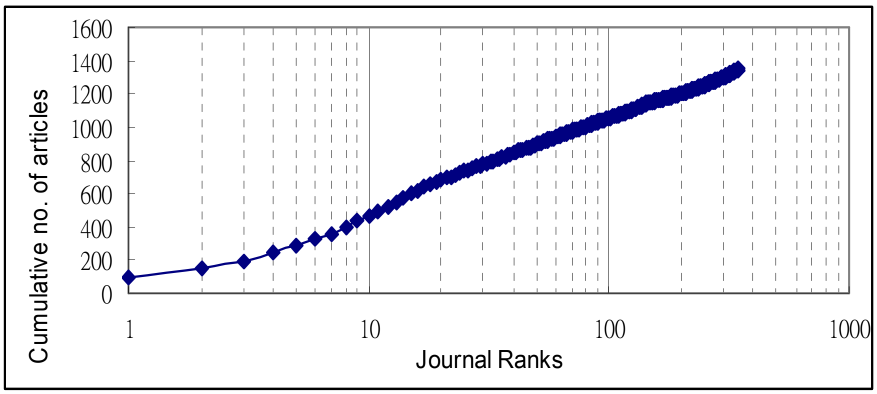



Figure 6: The Bradford-Zipf plot of journal research

Table 4 specifies the 10 leading journals that published the most altruism-related papers. According to data distribution, the papers published in these journals accounted for nearly one-third of the total. *Journal of Theoretical Biology* was top on the list, followed by *Evolution and Human Behaviour*. It was also observed that the most influential journal was *Nature*.

Table 4: Distribution of the 10 core journals

| Rank | Journal title | NoA[1] | % | TC[2] |
|---|---|---|---|---|
| 1 | *Journal of Theoretical Biology* | 102 | 7.57% | 2,114 |
| 2 | *Evolution and Human Behaviour* | 49 | 3.64% | 922 |
| 3 | *American Naturalist* | 45 | 3.34% | 1,304 |
| 4 | *Proceedings of The Royal Society B-Biological Sciences* | 44 | 3.26% | 584 |
| 5 | *Proceedings of The Royal Society of London Series B-Biological Sciences* | 44 | 3.26% | 1,621 |
| 6 | *Evolution* | 39 | 2.89% | 1,205 |
| 7 | *Proceedings of The National Academy of Sciences of The United States of America* | 38 | 2.82% | 1,119 |
| 8 | *Animal Behaviour* | 35 | 2.60% | 1061 |
| 9 | *Journal of Evolutionary Biology* | 35 | 2.60% | 769 |
| 10 | *Nature* | 32 | 2.37% | 3,785 |

[1] NoA: No. of articles; [2] TC: times cited

## Lotka's Law and Authority Productivity

We find the number of relationships between the two columns, in the column "No. of articles" and "% of Authors" in Table 5. Lotka's Law regarding author productivity can be summarized in equation (1), where $a_n$ = the number of authors publishing n papers, $a_1$= the number of authors publishing one paper, and c = a constant (in Lotka's case, c = 2) (Krisciunas 1977; Potter 1981; Wolfram 2003)

$$\mathbf{a_n = a_1/n^c, n= 1, 2, 3,..} \qquad \mathbf{(1)}$$

In computing the highest empirical values, the results of regression indicate that the constant c in equation (1) (representing the data shown in Table 5), is approximate to 2.39 and the estimated $a_1$ is 0.721. Thus, equation (1) is stated as equation (2):

$$\mathbf{a_n = 0.721 / n^{2.39}} \qquad \mathbf{(2)}$$

To verify whether the altruism-related research matched Lotka's Law, we performed a non-parametric Kolmogorov-Smirnov (K-S) goodness-of-fit test (Nicholls 1986; Pao 1986). According to the K-S test as showed in Appendix 2, the maximum difference between the observed and the estimated accumulated frequencies (Dmax) is 0.0597, and if the sampling number is greater than 35, the critical value will be $1.63 / 1953^{1/2} = 0.037$, because the total number of authors is 1,953. As the Dmax is 0.0496, this exceeds the critical value, and we conclude that the altruism-related data does not fit Lotka's law.

**Table 5.** Productivity of authors

| No. of articles | No. of authors | % of authors |
|---|---|---|
| 25 | 1 | 0.05% |
| 20 | 1 | 0.05% |
| 18 | 2 | 0.10% |
| 16 | 1 | 0.05% |
| 15 | 1 | 0.05% |
| 14 | 2 | 0.10% |
| 13 | 3 | 0.15% |
| 12 | 4 | 0.20% |
| 11 | 2 | 0.10% |
| 10 | 5 | 0.26% |
| 9 | 4 | 0.20% |
| 8 | 8 | 0.41% |
| 7 | 6 | 0.31% |
| 6 | 16 | 0.82% |
| 5 | 32 | 1.64% |
| 4 | 31 | 1.59% |
| 3 | 89 | 4.56% |
| 2 | 241 | 12.34% |
| 1 | 1504 | 77.01% |
| Total | 1953 | 100% |

## GHSOM and Topic Analysis

The process of applying GHSOM to topic analysis is illustrated in Figure 7. The three phases are: the data preprocessing phase; the clustering phase; and the interpreting phase. In the data preprocessing phase, key-terms such as titles, keywords, and subject categories are used to represent the contents of the documents. Meaningful key-terms describing the articles are extracted directly from the documents without any manual intervention. These key-terms are weighted according to a *tf* x *idf* the state-of-the-art weighting scheme shown in equation (3) (Rauber et al. 2002; Salton 1989; Shih et al. 2008; Wolfram 2003). In equation (3), $w_i(d)$ represents the weight of the ith term in document (d), $tf_i(d)$ represents the number of times the ith term appears in document (d), N (= 1348) represents the total number of altruism-related documents, and $df_i$ represents how many documents contain the ith term. The weighted value for a term will always be greater than or equal to zero. This weighting scheme assigns high values to terms considered important for describing the contents of a document and discriminating between various documents. A high weight is earned by frequent appearances of a term in a given document, with infrequent appearance of terms within the entire collection of documents. In this manner, weight assignment tends to filter out common terms. Based upon weighting values, we selected the top 219 remaining distinct key-terms for document representation. The resulting key-term vectors were used for GHSOM training.

$$\mathbf{w_i(d) = tf_i\,(d) * \log\,(\,N\,/\,df_i\,)} \qquad \mathbf{(3)}$$

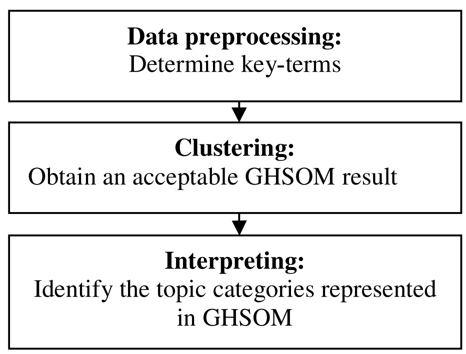



Figure 7: The three phases of the topic analysis process

In the clustering phase, the GHSOM experiment[2] was conducted through the trial and error method, using various values for breadth and depth and different normalizations to gain an acceptable GHSOM model for the analysis. The results of GHSOM are shown in Figure 8. The model comprised three layers and 46 nodes. All 1,348 altruism-related articles were clustered into a SOM of 2 x 3 nodes in layer 1, where all articles that had been clustered into the six nodes were further re-grouped into a SOM of 2 x 2 nodes in layer 2, respectively. The articles clustered into nodes 2.3, 4.1, 4.2 and 4.4 were further re-grouped into a SOM of 2 x 2 nodes in layer 3.

In the interpreting phase, for each node of GHSOM in node 1 to 6 of the first-layer and node 2.3, 4.1, 4.2 and 4.4 of the second-layer, we count the $df_i$ value of each key-term in all articles cluster them into a particular node and assigned a key-term with the highest $df_i$ value (or several key-terms if their $df_i$ values were very close), as the topic category. If there were more than five topics, we would denote it as multidisciplinary. For the remaining nodes, the utmost five important key-terms would be automatically assigned by the GHSOM using the *tf* x *idf* weighting scheme such as node 1.1, 1.2, 1.3 and so on.

[2] We used GHSOM toolbox in the Matlab R2007a® package to conduct the GHSOM experiment.

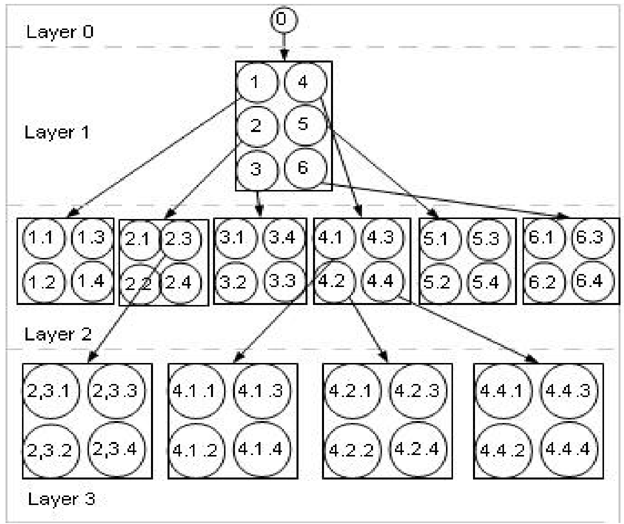



Figure 8: The GHSOM result

The results are presented in Figures 9, 10, and 11, in which the number in the parenthesis refers to the number of clustered articles. For instance, there were 144 altruism-related articles clustered into node 1, and based upon the interpretation, it was named “biology”; 226 articles in node 2 as “biology & multidiscipline”, 180 articles in node 3 as “evolutionary biology & ecology category”, 510 articles in node 4 as “multidiscipline category”, 180 articles in node 5 as “behavioural sciences & multidiscipline category”, 108 articles in node 6 as “ecology & multidiscipline category”. Based on these dominant topical clusters in the collection of altruism-related articles, further specific topics were obtained in layer 2 (Figure 10). For instance, articles in the “biology category” were further re-grouped into sub-category topics including “cooperation”, “biology”, “ecology”, “evolutionary biology” and “reciprocity” in node 1.1; the sub-category topics including “cooperation”, “biology”, “selection”, “reciprocal altruism” and “mathematical & computational biology” in node 1.2; the sub-category topics including “cooperation”, “biology”, “mathematical & computational biology”, “game” and “dynamics” in node 1.3; and sub-category topics including “cooperation”, “biology”, “reciprocal altruism”, “game” and “mechanism” in node 1.4. Articles in a number of nodes of layer 2 (that is, nodes 2.3, 4.1, 4.2, and 4.4) were further re-grouped into more specific subcategories in layer 3, as shown in Figure 11.

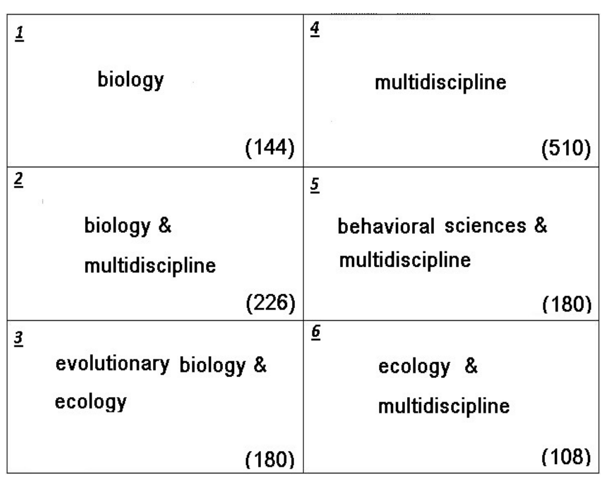
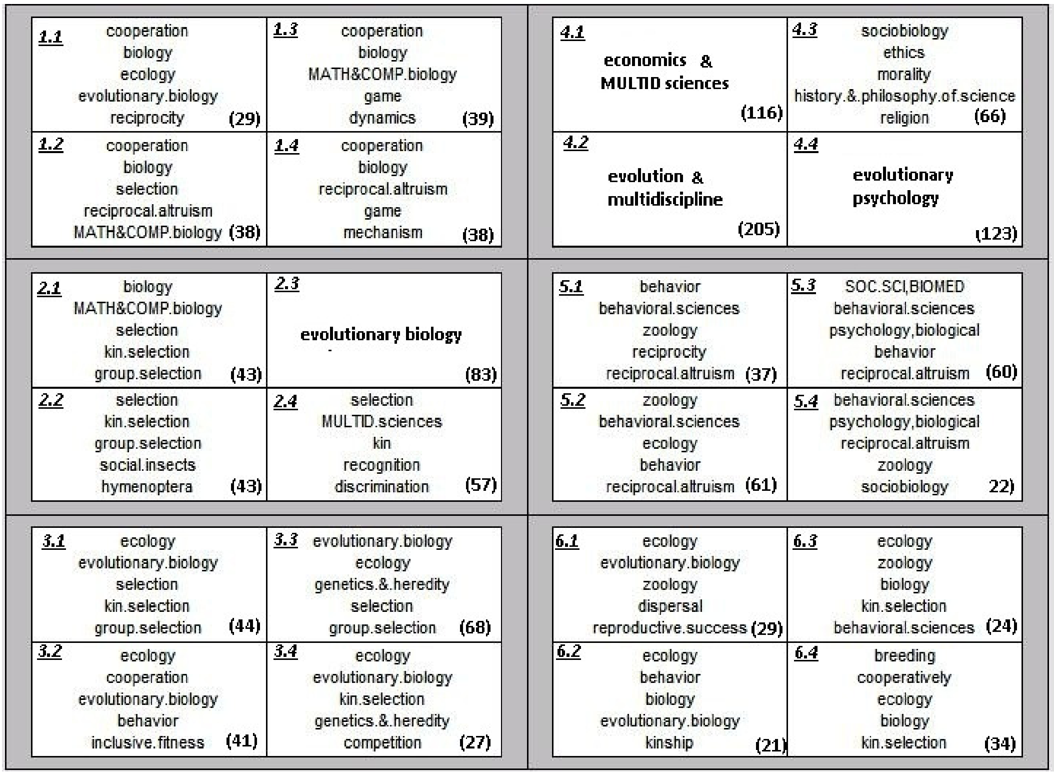


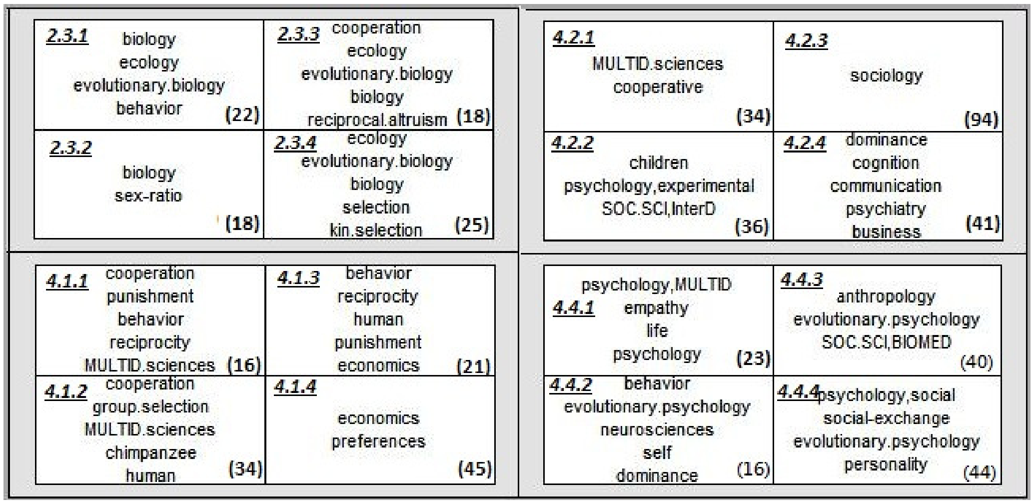

Figure 9: First-layer interpretation results of GHSOM.




MULTID is the abbreviation for multidisciplinary; SOC refers to social; SCI is science; BIOMED is biomedical; MATH is mathematical; COMP is computational

Figure 10: Second-layer interpretation result of GHSOM.

The interpretation results for the second- and third-layer of GHSOM shown in Figure 10 and 11 respectively were more delicate than those in Figure 5 were. It was observed that the interpretation results for the second-layer were more specific than in the first-layer. For instance, articles in nodes 1.1 and 1.3 belonged to the category of “biology” in node 1, but they both have further differentiations. Node 1.3 focuses on mathematical & computational biology, game and dynamics, while node 1.1 focuses on ecology, evolutionary biology and reciprocity. Another interesting observation shown in Figure 10 is that the two neighbouring nodes are much more closely related than the remote nodes. For example, articles clustered in node 4.3 related to the concept of “sociobiology”, “ethics”, “morality”, “history & philosophy of science” and “religion” at the top-right corner of Figure 10 are obviously very different from those clustered in node 3.1 related to the concept of “ecology”, “cooperation”, “evolution biology”, “behaviour” and “inclusive fitness” in the bottom-left corner of Figure 10, but they are more closely related to those in nodes 4.1, 4.2 and 4.4.




InterD is the abbreviation for interdisciplnary; SOC refers to social; SCI is science; BIOMED is biomedical.

Figure 11: Third-layer interpretation result of GHSOM.

The results of the GHSOM complied with the subject area rankings in the first layer, and provided more explicit topics implying the interrelationship of the different subject areas in the second or third layers. For example, the behavioural sciences in Figure 5 is in the node 5.1, 5.2, 5.3 and 5.4 of Figure 10, indicating that altruism-related research related to behavioural sciences was relevant to biology psychology, zoology, sociology and ecology. The first-layer interpretation results give the disciplinary map while the second- and third-layer interpretation results present topic maps indicating the relationship among different disciplines. In addition, the topic maps reflected that the evolutionary concepts were applied into multidiscipline. The terms such as “reciprocal altruism”, “kinship”, or “group selection” are penetrating with different subject areas from the evolutionary respective.

However, the evolutionary psychology sub-category in node 4.3 with 66 papers and node 4.4 with 123 altruism-related papers may indicate the new scientific frontier about altruism. Node 4.3 demonstrates such a group discussed ethics and morality from altruistic perspective in the subject areas of sociobiology, history & philosophy of science and religion, while node 4.4 deals with the evolutionary psychology sub-category. It co-exists with a number of disciplines such as multidisciplinary psychology in node 4.4.1, neurosciences in node 4.4.2 , anthropology and biomedical social science in node 4.4.3, social psychology in node 4.4.4 which imply that these studies were interdisciplinary and focused on evolutionary psychological respective.

To be more precise, the topics in nodes 4.4.1, 4.4.2, 4.4.3 and 4.4.4 explained why evolutionary psychology implies the new scientific frontier in Figure 11. For example, node 4.4.1 tells us that groups of research associated with psychology and multidisciplinary psychology were strongly related to the concept of life and empathy, which indicated that the intention of altruism could be interpreted by empathy and the significance of life. At the same time node 4.4.2 illustrates how neurosciences were adopted to explore the relationship between *altruistic behaviour* and *self*. In addition, node 4.4.3 shows that the group of biomedical social science researchers targeted altruism, which is based on the research of anthropology and evolutionary psychology. Node 4.4.4 gives us a hint that the social psychology group applies the ideas of evolutionary psychology to discuss personality and social-exchange. More specifically, the works such as Preston and de Waal (2002) dealing with empathy and Rilling et al. (2002) providing neural basis in the research related to evolution of altruism in the table 3 could explain the above suggestion, because their articles were prominently cited in research related to the behavioural or neural foundation of altruism. This implies that evolution of altruism steps closer to the inside of human, while the research focus moves from physical biology to metaphysical behavioural sciences and psychology.

## CONCLUSION

To sum up, this informetric study provided an overall picture of altruism-related articles published in the SCIE and SSCI databases. We observed a steady growth in the number of altruism-related papers between the years of 1971 and 2009. According to Bradford-Zipf's S shape of scattering with regard to scientific research, the study identified 10 core altruism-related journals, comprising one-third of the published altruism-related research. *Journal of Theoretical Biology* was top on the list in productivity, while *Nature* was the most influential journal in citation. The frequency distribution regarding author productivity did not match Lotka's Law, but it should be stressed that Lotka's inverse square law is a general, theoretical estimate of productivity, and is not a precise statistical measurement (Potter 1981). Nonetheless, its appeal as a hard and fast law of distribution is undeniable. The three most productive authors were Wilson, DS, West, SA, and Lehmann, L. The two most influential authors were Trivers, R., and Fehr, E. with regard to the number of times cited.

The GHSOM tool had all of the benefit of SOM, in providing a map from a higher dimensional input space to a lower dimensional map space, as well as providing a global orientation of independently growing maps in the individual layers of the hierarchy, which facilitated navigation across branches. The topic map illustrated the delicate intertwining of subject areas and provided a more explicit illustration of the concepts within each subject area. In this study, we found that the discovering works from evolution aspects penetrate into research of different subject areas. The in-depth exploring into the motivation of evolution of altruism will be leaded by evolutionary psychology, neurosciences and other related sciences.

In addition, a number of facts shown in this study may be due to the nature of the SCIE and SSCI databases selected for this study. For example, most of the altruism-related papers were published by USA-based institutions and the most productive countries were English speaking countries. In fact, there are many criticisms leveled at the SCIE and SSCI databases, regarding its tendency to contain a high percentage of English-language journals from English speaking countries, particularly journals from the USA and the United Kingdom, followed by other Commonwealth countries such as Canada and Australia. It is well-known that other non-native English speaking countries have greater difficulty publishing in these kinds of journals, either because of language difficulties or because their countries have their own national publication systems (Hicks 1999; Andersen 2000; Archambault et al. 2006; Barrios et al. 2008)

## REFERENCES

Allison, P., and Stewart, J. 1974. Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review,* Vol. 39, no.4: 596-606.

Boyack, K. W., Klavans, R., and Borner, K. 2005. Mapping the backbone of science. *Scientometrics,* Vol. 64, no.3: 351-374.

Brookes, B. C. 1973. Numerical methods of bibliographic analysis. *Library Trends*, Vol. 22, no.1: 18-43.

Chau, M., Huang, Z., Qin, J., Zhou, Y., and Chen, H. 2006. Building a scientific knowledge web portal: The NanoPort experience. *Decision Support Systems,* Vol. 42, no.2: 1216-1238.

Dawkins, R. 2006. *The selfish gene*. New York, USA: Oxford University Press

Ding, Y., Chowdhury, G., and Foo, S. 2001. Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing & Management,* Vol. 37, no.6: 817-842.

Dittenbach, M., Rauber, A., and Merkl, D. 2002. Uncovering hierarchical structure in data using the growing hierarchical self-organizing map. *Neurocomputing,* Vol. 48, no.1-4: 199-216.

Grupp, H., and Schmoch, U. (eds.) 1992. *Perception of scientification as measured by referencing between patents and papers*. Heidelberg: Spring Verlag.

Hamilton, W. D. 1964. The genetical theory of social behaviour, I, II. *Journal of Theoretical Biology,* Vol. 7, no.1: 17-52.

Hassan, E. 2003. Simultaneous mapping of interactions between scientific and technological knowledge bases: the case of space communications. *Journal of the American Society for Information Science and Technology,* Vol. 54, no.5: 462-468.

Hood, W. W., and Wilson, C. S. 2001. The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics,* Vol. 52, no.2: 291-314.

Kawatra, P. S. 2000. Textbook of Information Science. New Delhi: APH, Vol. 275-304.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biological cybernetics,* Vol. 43, no.1: 59-69.

Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., and Saarela, A. 2000. Self organization of a massive document collection. *IEEE Transactions on Neural Networks,* Vol. 11, no.3: 574-585.

Krebs, D. L. 1970. Altruism - examination of concept and a review of literature. *Psychological Bulletin,* Vol. 73, no.4: 258-302.

Krisciunas, K. 1977. Letter to the editor. *Journal of the ASZS,* Vol. 28, 65-66.

Leydesdorff, L. 1987. Various methods for the mapping of science. *Scientometrics,* Vol. 11, no.5: 295-324.

Li, S. T., and Chang, W. C. 2009. Design and evaluation of a layered thematic knowledge map system. *Journal of Computer Information Systems,* Vol. 49, no.2.

Nicholls, P. 1986. Empirical validation of Lotka's law. *Information Processing & Management,* Vol. 22, no.5: 417-419.

Noyons, E. 2001. Bibliometric mapping of science in a policy context. *Scientometrics,* Vol. 50, no.1: 83-98.

Noyons, E., and van Raan, A. 1998. Monitoring scientific developments from a dynamic perspective: self-organized structuring to map neural network research. *Journal of the American Society for Information Science,* Vol. 49, no.1: 68-81.

Pao, M. 1986. An empirical examination of Lotka's law. *Journal of the American Society for Information Science,* Vol. 37, no.1: 26-33.

Piliavin, J. A., and Charng, H. W. 1990. Altruism: A review of recent theory and research. *Annual Review of Sociology,* Vol. 16, no.1: 27-65.

Potter, W. G. 1981. Lotka's law revisited. *Library Trends,* Vol. 30, no.1: 21-39.

Preston, S. D., and de Waal, F. B. M. 2002. Empathy: Its ultimate and proximate bases. *Behavioural and Brain Sciences,* Vol. 25, no.1: 1-20.

Price, D. J. S. 1965. Networks of scientific papers. *Science,* Vol. 149, no.3683: 510-515.

Rauber, A., Merkl, D., and Dittenbach, M. 2002. The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. *IEEE Transactions on Neural Networks,* Vol. 13, no.6: 1331.

Rilling, J. K., Gutman, D. A., Zeh, T. R., Pagnoni, G., Berns, G. S., and Kilts, C. D. 2002. A neural basis for social cooperation. *Neuron,* Vol. 35, no.2: 395-405.

Sagar, A., Kademani, B. S., Garg, R. G., and Kumar, V. 2010. Scientometric mapping of Tsunami publications: a citation based study. *Malaysian Journal of Library & Information Science,* Vol. 15, no.1: 23-40.

Salton, G. 1989. *Automatic text processing: the transformation, analysis, and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Sharabany, R., and Bar-Tal, D. 1982. Theories of the development of altruism: Review, comparison and integration. *International Journal of Behavioural Development,* Vol. 5, no.1: 49.

Shih, J., Chang, Y., and Chen, W. 2008. Using GHSOM to construct legal maps for Taiwan's securities and futures markets. *Expert Systems with Applications,* Vol. 34, no.2: 850-858.

Trivers, R. L. 1971. Evolution of reciprocal altruism. *Quarterly Review of Biology,* Vol. 46, no.1: 35-57.

Tsaih, R., Lin, W., and Huang, S. 2009. Exploring fraudulent financial reporting with GHSOM. in H. Chen et al. (Eds.): PAISI 2009, *Lecture Notes in Computer Science,* Vol. 5477: 31-41.

Tsay, M. Y., and Lin, Y. J. 2009. Scientometric analysis of transport phenomenon literature, 1900-2007. *Malaysian Journal of Library & Information Science,* Vol. 14, no.3: 35-58.

Van Raan, A. F. J., and Tijssen, R. J. W. 1993. The neural net of neural network research. *Scientometrics,* Vol. 26, no.1: 169-192.

Weiner, E., Simpson, J., and Proffitt, M. 1993. *Oxford English Dictionary*. Clarendon.

Wilson, E. O. 1975. Sociobiology: The New Synthesis. Cambridge, USA: Harvard Univ. Press.

Wolfram, D. 2003. *Applied informetrics for information retrieval research*. Santa Barbara, USA: Greenwood Publishing Group.

Yin, C. Y. 2009. Bibliometric analysis of journal articles published by Southeast Asian chemical engineering researchers. *Malaysian Journal of Library & Information Science,* Vol. 14, no.3: 1-13.

## APPENDIX

Appendix 1: Distribution of journals according to Bradford's Law

| Zone | NoA | NoJ | AccNoJ | SToA | AccNoA |
|---|---|---|---|---|---|
| | 102 | 1 | 1 | 102 | 102 |
| | 49 | 1 | 2 | 49 | 151 |
| | 45 | 1 | 3 | 45 | 196 |
| (A) Core | 44 | 2 | 5 | 88 | 284 |
| | 39 | 1 | 6 | 39 | 323 |
| | 38 | 1 | 7 | 38 | 361 |
| | 35 | 2 | 9 | 70 | 431 |
| | 32 | 1 | 10 | 32 | 463 |
| | 29 | 1 | 11 | 29 | 492 |
| | 28 | 1 | 12 | 28 | 520 |
| | 27 | 1 | 13 | 27 | 547 |
| | 25 | 1 | 14 | 25 | 572 |
| | 24 | 1 | 15 | 24 | 596 |
| | 23 | 1 | 16 | 23 | 619 |
| | 21 | 1 | 17 | 21 | 640 |
| (B) Relevant | 15 | 2 | 19 | 30 | 670 |
| | 12 | 1 | 20 | 12 | 682 |
| | 11 | 3 | 23 | 33 | 715 |
| | 10 | 3 | 26 | 30 | 745 |
| | 8 | 4 | 30 | 32 | 777 |
| | 7 | 7 | 37 | 49 | 826 |
| | 6 | 5 | 42 | 30 | 856 |
| | 5 | 12 | | 60 | 916 |
| | 4 | 11 | 65 | 44 | 960 |
| (C) Marginal | 3 | 25 | 90 | 75 | 1035 |
| | 2 | 50 | 140 | 100 | 1135 |
| | 1 | 213 | 353 | 213 | 1348 |

NoA: No. of articles; NoJ: No. of journals; AccNoJ: Accumulated No. of journals; SToA: subtotal of articles = NoJ * NoA; AccNoA: Accumulated No. of articles.

Appendix 2: Author distribution according to Lotka's Law

| NoA | OA | Sn(X) | EVA | Fo(X) | AbV |
|---|---|---|---|---|---|
| 1 | 0.7701 | 0.7701 | 0.7207 | 0.7205 | **0.0496 (Dmax)** |
| 2 | 0.1234 | 0.8935 | 0.1374 | 0.8579 | 0.0356 |
| 3 | 0.0456 | 0.9391 | 0.0521 | 0.9101 | 0.0290 |
| 4 | 0.0159 | 0.9549 | 0.0262 | 0.9363 | 0.0187 |
| 5 | 0.0164 | 0.9713 | 0.0154 | 0.9516 | 0.0197 |
| 6 | 0.0082 | 0.9795 | 0.0099 | 0.9616 | 0.0180 |
| 7 | 0.0031 | 0.9826 | 0.0069 | 0.9684 | 0.0141 |
| 8 | 0.0041 | 0.9867 | 0.0050 | 0.9734 | 0.0132 |
| 9 | 0.0020 | 0.9887 | 0.0038 | 0.9772 | 0.0115 |
| 10 | 0.0026 | 0.9913 | 0.0029 | 0.9801 | 0.0112 |
| 11 | 0.0010 | 0.9923 | 0.0023 | 0.9825 | 0.0098 |
| 12 | 0.0020 | 0.9944 | 0.0019 | 0.9844 | 0.0100 |
| 13 | 0.0015 | 0.9959 | 0.0016 | 0.9859 | 0.0100 |
| 14 | 0.0010 | 0.9969 | 0.0013 | 0.9873 | 0.0097 |
| 15 | 0.0005 | 0.9974 | 0.0011 | 0.9884 | 0.0091 |
| 16 | 0.0005 | 0.9980 | 0.0010 | 0.9893 | 0.0086 |
| 18 | 0.0010 | 0.9990 | 0.0007 | 0.9900 | 0.0089 |
| 20 | 0.0005 | 0.9995 | 0.0006 | 0.9906 | 0.0089 |
| 25 | 0.0005 | 1.0000 | 0.0003 | 0.9909 | 0.0091 |

NoA: No. of articles; OA: Observation by author(s); Sn(X): Accumulated OA; EVA: Expected Value by Author; Fo(X): Accumulated EVA; AV: Absolute Value=|Fo(X)-Sn(X)|