SLIS Connecting

Volume 4 | Issue 1 Article 9

2015

"Information Theory" Research Trend: A Bibliometric Approach

Maryam Asadi
Islamic Azad University

"Information Theory" Research Trend: A Bibliometric Approach

By Maryam Asadi
Islamic Azad University, Tehran, Iran

Introduction

One of the most controversial issues, in a multidisciplinary area, is *information theory. The American Heritage Science Dictionary* (2014) defines information theory as "a branch of mathematics that mathematically defines and analyzes the concept of information, statistics and probability theory, and applications include the design of systems that have to do with data transmission, encryption, compression, and information processing and deeply affected many fields." It has a fundamental contribution to make in multiple disciplines, such as mathematics, statistics, computer science, physics, neurobiology, communication theory, information science, economics, and electrical engineering. Indeed, information theory played a crucial role to achieve many pertinent scientific missions and inventions such as Voyager, Mobile phone, and Internet.

Shannon's 1948 paper, "A Mathematical Theory of Communication," in the *Bell System Technical Journal* is considered by many scholars to be the founding document of information theory. In this paper, he introduced a qualitative and quantitative model of communication as a statistical process underlying information theory (Shannon, 1948; Shannon & Weaver, 1949).

A large body of research has been published in different formats (e.g. journal, book, proceeding papers, letter, meeting abstract, etc.) in the field of information theory. Given this, identifying and measuring the effect and value of scientific outputs, such as the field of information theory, plays an important role in today's world of information overload. Bibliometric studies carried out in recent years "have provided an accurate and objective method to measure the contribution of the paper to the advancement of knowledge" and the *Web of Science* database is "the most widely accepted and frequently used database for analysis of scientific publications" (Wen & Huang, 2012).

Purpose of the Study

The purpose of this study is to carry out a bibliometric analysis and examine scientific collaborations in information theory articles gathered from the Thomson Reuters *Web of Science* (*WoS*) database. Therefore, this study focuses on the following objectives:
1) number of publications in the field of information theory and growth rate; 2) type of documents and languages used by researchers; 3) periodicals that published the most articles in this field of study; 4) countries that produced the greatest number of publications; 5) universities and institutions with the greatest number of publications; 6) the most productive and effective authors in this field of study; 7) average page count per article and average cited reference count per article; 8) most cited articles of information theory; 9) authorship and collaboration patterns in this field of study; 10) the type of collaboration between scholars in this field (institution, inter-institution, international); and 11) most common subject categories and keywords.

Methods

This research was done with a bibliometric approach. This study was limited to Thomson Reuters *Web of Science* (*WoS*) database using the following search strategy: TS="information theory". TS tag searches title, abstract, and keyword fields. Document information included name of author, author address (affiliation), title, year of publication, keywords, name of journal, and number of cited references. In total, 9,243 items were retrieved and duplicate records were excluded. Overall, 8,466 records were gathered and analyzed using Microsoft Excel software. Additional coding was manually performed based on the research objectives

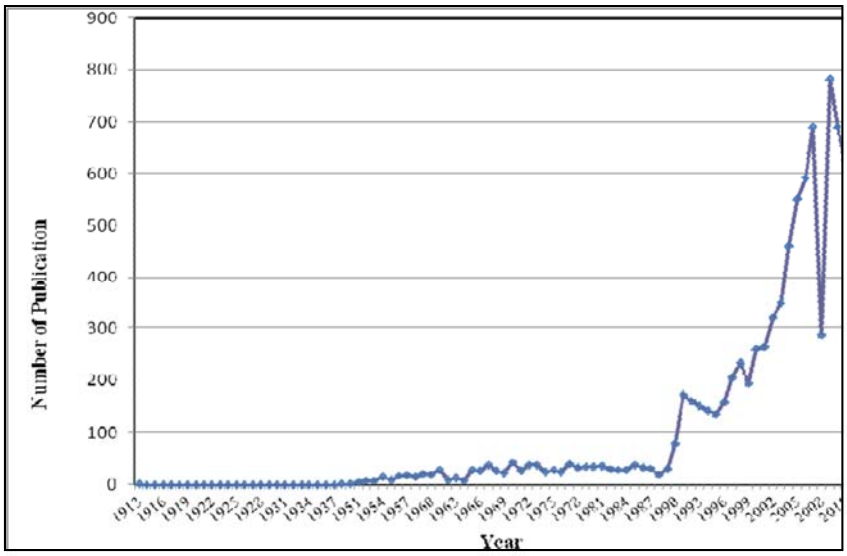

Figure1. Publications on Information Theory, 1913-2011

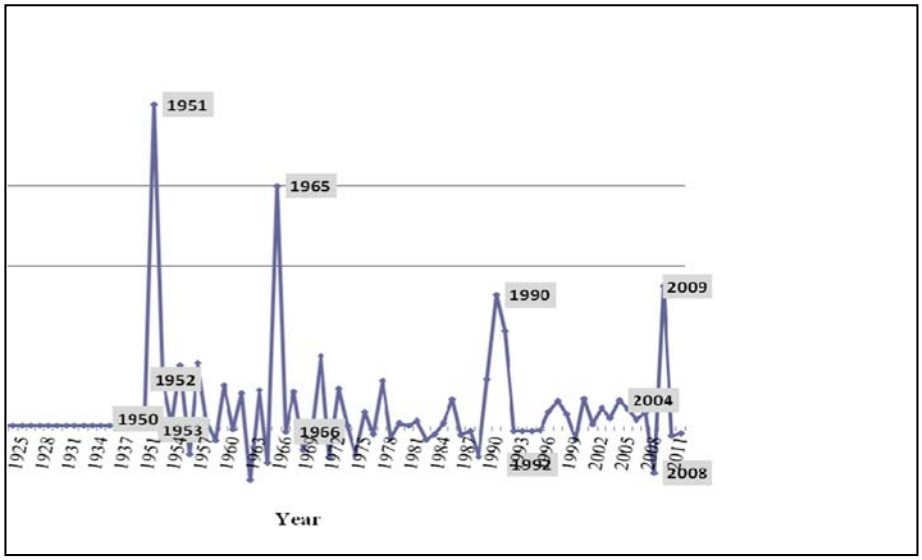

Figure 2. Growth Rate of Publications on Information Theory, 1913-2011

Results

Number of Publications and Growth Rate

Data showed that 8,466 items were published by researchers in sources covered by *Web of Science* (*WoS*) in the area of information theory.

Figure 1 indicates that the number of scientific publications in the field of information theory over the period from 1913 to 2007 increased. Despite an overall increase, there was a decrease in 2008, a sharp rise in 2009, and during 2010- 2011 bottomed out slightly.

In order to calculate trend of growth rate of publications in information theory, the following formula was used:

$$PR = = \frac{Vpresent - Vpast}{Vpast} \times 100$$

Where PR is Percent Rate, $V_{present}$ is present or future value and V_{past} is past value.

Figure 2 demonstrates the trend of growth rate of publications in information theory. As can be seen, there were fluctuations in the trend of scientific publications of information theory. In fact, there were three general trends. First, over the period from 1913 to 1950, the trend of outputs was stagnated. Secondly, in the years 1951, 1965, 1990, and 2009, there was an upward trend and in years 1952-53, 1966, 1992, 2008, and 2010-11, there was a downward trend. Finally, during 1954-1964, 1967-

1988, and 1993-2007 growth rate of outputs was uneven and did not exhibit a clear pattern.

In order to calculate the average annual growth rate of works during 89 years, the geometric mean is used. Results showed that this rate was 7.5 percent and as mentioned, breaks in the growth of publications in some years it occurred.

Types of Documents and Languages

The results of the analysis of the type of documents showed indicated that the documents were in eleven different formats including articles, proceedings, book reviews, correction, discussion, editorial material, letter, meeting abstract, news item, note, and reprint. As expected, the most frequent format was articles (60.10%), followed by proceedings (33.09%), book review (2%), and meeting abstracts (1.45%) (Table 1).

Scientific publications in the field of information theory were published in 16 languages. English documents ranked top (97.52%) followed by German (1.07%), and French (0.50%) respectively (Table 2).

Table 1. Document Type and Percentage

Article	60.10%
Proceedings Paper	33.09%
Book Review	2.00%
Meeting Abstract	1.45%
Editorial	1.39%
Letter	0.93%

Table 2. Language and Percentage

English	97.52%
German	1.07%
French	0.50%
Russian	0.35%
Czech	0.12%
Spanish	0.11%
Japanese	0.08%
Hungarian	0.06%
Italian	0.05%
Chinese	0.04%
Rumanian	0.04%
Portuguese	0.02%
Dutch	0.01%
Korean	0.01%
Polish	0.01%
Slovak	0.01%

Periodicals that Published the Most Articles

The items in this study were published in 3,424 journals and conference proceedings. *IEEE Transactions on Information Theory* published the most documents (3.59%) followed by *Physical Review A: Atomic, Molecular and Optical Physics* (2.01%), *Physica A: Statistical Mechanics and Its Applications* (1.32%), and *Physical Review E: Statistical, Nonlinear and Soft Matter Physics* (1.24%). Table 3 shows top ten journals and proceedings that published in the field of information theory.

Table 3. Publication Title and Percentage

1	*IEEE Transactions on Information Theory*	3.59%
2	*Physical Review A*	2.01%
3	*Physica A*	1.32%
4	*Physical Review E*	1.24%
5	*Electronics Letters*	1.13%
6	*Physics Letters A*	0.78%
7	*Entropy*	0.76%
8	*Physical Review Letters*	0.67%
9	*Proceedings of the US National Academy of Sciences*	0.53%
10	*Journal of Theoretical Biology*	0.52%

Countries that Produced the Greatest Percentage of Publications

Researchers of the 91 Countries had impressive contribution to this field of study. Among them, countries of USA, UK, Peoples Republic China, Canada, Germany, Japan, Spain, France, Italy, and Australia respectively, were top ten countries that published the highest number of publications during in this period in field of information theory (Table 4).

Table 4. Publications by Country

1	USA	46.92%
2	UK	8.08%
3	Peoples R China	7.36%
4	Canada	5.53%
5	Germany	5.00%
6	Japan	4.98%
7	Spain	4.55%
8	France	4.52%
9	Italy	3.71%
10	Australia	3.07%

Publications by Universities and Institutions

There were over 3,900 institutions and universities involved in research activity in the field of information theory. As can be seen in Table 5, MIT university, University of Illinois, National University La Plata, University of Waterloo, University of California, Berkeley, California Institute of Technology (CALTECH), University California, San Diego, Stanford University, Le Centre national de la recherche scientifique (CNRS), and Princeton University produced the most scientific documents in this field ranked 1-10, respectively.

Table 5. Publications by Institution

1	MIT University	8.81%
2	University of Illinois	1.04%
3	National University La Plata	0.94%
4	University of Waterloo	0.93%
5	University of California, Berkeley	0.90%
6	California Institute of Technology (CALTECH)	0.85%
7	University California, San Diego	0.83%
8	Stanford University	0.80%
9	Le Centre national de la recherche scientifique (CNRS)	0.77%
10	Princeton University	0.73%

Most Productive Authors in Information Theory

Table 6 illustrates the most productive and effective authors in the field of information theory. Findings indicated that of 8,466 documents were written by 13,590 authors. Plastino, Hayashi, Matsuda, Nalewajski, and Casas were the five most productive authors in the field of information theory; i.e. they authored the largest number of papers.

Table 6. Most Productive Authors

1	Plastino A	162
2	Hayashi Y	35
3	Nalewajski RF	34
4	Matsuda R	33
5	Casas M	28
6	Wallace R	28
7	Jumarie G	25
8	Schneider TD	25
9	Rogan PK	24
10	Szpankowski W	24

Average Number of Pages and Cited References

Characteristics of publications of information theory research (i.e. number of articles, average page count per article, and average reference count per article) in the study period 1913-2011 are summarized in Table 6. They are grouped in eight periods, including 1913, 1938, 1950-59, 1960-69, 1970-79, 1980-89, 1990-99, and 2000-11. From Table 7, it can be inferred that the number of article increased from 1 in 1913 to 5874 in 2000-2011.

The average article length and the average reference count per article increased slightly. Overall, the publications averaged 13.74 pages, 28.68 references, 16.83 times cited per article of information theory research in the study period 1913-2011.

Table 7. Publication Characteristics

Year	A	PG/A	RE/A
1913	1	8	1
1938	1	1	1
1950-1959	113	6.47	19.29
1960-1969	217	4.71	10.92
1970-1979	328	7.40	14.54
1980-1989	305	12.64	16.90
1990-1999	1627	10.24	23.97

2000-2011	5874	15.60	32.23
Total	8466	13.74	28.68

Note: A = Number of Articles; PG/A = Average Page Count; RE/A = Average Reference Count per Article

Most Cited Articles of Information Theory

Number of citations to a publication illustrates the impact of its work in the field of science. To determine articles with greatest impact, articles were ranked based on times cited (Table 8). The times cited count indicates the number of times a published paper was cited by other papers and is a measure of the impact of a work in that field of science.

Table 8: Most Cited Articles

#	Article Title	Times cited
1	Turk, M., & Pentland, A. (1991). Eignfaces for recognition. *J of Cognitive Neuroscience.*	3389
2	Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review.*	3267
3	Maes, F., et al (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*.	1834
4	Douglass, J. K. et al (1993). Noise enhancement of information. *Nature.*	805
5	Biglieri, E., et al. (1998). Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory.*	783
6	Kramer, G. et al. (2005). Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*.	721
7	Anderson, D. R.et al. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *J of Wildlife Management*.	704
8	O'Neill, R. V. et al. (1988). Indices of landscape pattern. *Landscape Ecology.*	578
8	Koetter, R., & Medard, M. (2003). An algebraic approach to network coding. *IEEE-ACM Transactions on Networking.*	527
10	Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience.*	511

Table 8 lists ten of the articles with greatest impact in information theory based on citation. We can see that Turk and Pentland (1991), Jaynes (1957), and Maes et al. (1997) have been cited 3389, 3267, and 1834 times in this period, respectively and therefore, are the most highly cited papers in the field of information theory.

Authorship Patterns in Information Theory

For the purpose of determination of authorship patterns and degree of collaboration, the author field in *Web of Science* (*WoS*) database was searched and data were classified into three groups: one author, two authors, and three or more authors; 8,466 items about information theory were statistically analyzed in Table 9, including the percentage of outputs in eight periods. Results indicate that publications with one author during 1913-1989 were dominant and publications with two and three or more authors during 1990-2011 increased sharply.

By and large, 29.60 percent of authorship is related to one author, 30.38 percent of one is related to two authors, and 40.02 percent of one is related to three or more authors that made up writing patterns in field of information theory during 89 years.

Table 9. Number of Authors per Article

Year	One auth.	Two auth.	Three or more
1913	0.01%	0.00%	0.00%
1938	0.01%	0.00%	0.00%
1950-1959	1.23%	0.11%	0.00%
1960-1969	2.10%	0.34%	0.12%
1970-1979	2.60%	0.87%	0.40%

1980-1989	1.89%	0.90%	0.82%
1990-1999	6.47%	07.08%	5.67%
2000-2011	15.28%	21.08%	33.01%
Total	29.60%	30.38%	40.02%

In order to determine proportion of single and co-authored products in this field, first, works with two authors or more classified in a group named coauthored productions. In other words, outputs that have been provided with at least two authors or more considered as collaborative. In addition, to calculate ratio of co-authorship publications to single author publications the below formula was used:

$$\text{Co-authorship ratio} = \frac{\#\text{Co-Authored Publications}}{\#\text{Single-Authored Pubs} + \#\text{Co-Authored Pubs}}$$

Ratio greater than 0.5 = more co-authored than single authored publications
Ratio of less than 0.5 = more single- authored than co-authored publications (Sutter & Kocher, 2004).

According to the formula, the ratio of co-authorship in field of information theory during 1913-2011 was 0.70. As a result, it indicates that proportion of co-authorship is more than single author production. Based on the findings, as shown in Table 9, 70.40 percent of published papers in the research period were produced collaboratively and only 29.60 percent of papers were produced individually.

Type of Collaboration

To determine the type of collaboration between scholars in this field, the papers with at least two authors or more belonging to the same organization, university or research institution were considered as institutional collaboration. The papers with at least 2 authors or more in different organizations, universities or research institution and in a country were considered as inter-institutional collaboration and the paper with at least 2 authors or more that were produced with the organizations, universities or research institutions in different countries were considered as international collaboration. The types of collaboration were grouped in eight periods, including 1913, 1938, 1950-59, 1960-69, 1970-79, 1980-89, 1990-99, and 2000-11.

Table 10 illustrates the type of collaboration in field of information theory. In 1913 and 1938 collaborated works were single author and during 1950-1959, only type of collaboration was institutional and constituted 0.15 percent and during 1960-1969, type of collaboration was institutional and inter-institutional. In period times 1970-2011 showed each three of type collaboration and institutional collaboration was dominant. Generally, the type of collaboration between scholars in the field information theory was 52.23 percent of collaborations institutional, 30.29 percent inter-institutional, and 17.48 percent of them international. Overall, institutional collaboration was dominant.

Table 10. Type of Collaboration between Iranian Scholars

Year(s)	Institutional	Inter-institutional	Inter national
1913	0.00%	0.00%	0.00%
1938	0.00%	0.00%	0.00%
1950-1959	0.15%	0.00%	0.00%
1960-1969	0.62%	0.03%	0.00%
1970-1979	1.44%	0.34%	0.03%
1980-1989	1.07%	1.16%	0.20%
1990-1999	9.73%	5.64%	2.73%
2000-2011	39.21	23.12%	14.51%
Total	52.23%	30.29%	17.48%

Distribution of Subject Categories

Distribution of subject categories via keyword analysis indicated a great diversity including 123 subject categories related to the research topic of information theory in *Web of Science*. Table 11 lists that top 10 subject categories in the 89 years. Among the top subject categories are computer science (29.93%), engineering (22.35%), and physics (18.40%) have been the most outputs and the article percentage had a significant growth.

Table 11. Top 10 Subject Categories

Subject Categories	1913	1938	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2011	Total
Computer science	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.21%	25.73%	29.93
Engineering	0.00%	0.00%	0.06%	0.07%	0.02%	0.12%	3.33%	18. 75%	22.35%
Physics	0.00%	0.00%	0.00%	0.00%	0.04%	0.02%	2.40%	15.95%	18.40%
Mathematics	0.00%	0.00%	0.06%	0.07%	0.08%	0.18%	1.41%	7.17%	8.97%
Telecommunications	0.01%	0.01%	0.14%	0.12%	0.11%	0.11%	0.48%	6.14%	7.12%
Optics	0.00%	0.00%	0.02%	0.14%	0.04%	0.06%	0.97%	5.50%	6.73%
Chemistry	0.00%	0.00%	0.02%	0.06%	0.09%	0.27%	1.13%	2.65%	4.23%
Neurosciences neurology	0.00%	0.00%	0.00%	0.00%	0.08%	0.19%	0.64%	2.70%	3.61%
Environmental sciences ecology	0.00%	0.00%	0.00%	0.00%	0.08%	0.15%	0.38%	2.34%	2.95%
Biochemistry molecular biology	0.00%	0.00%	0.00%	0.00%	0.04%	0.05%	0.33%	2.36%	2.78%

Table 12. Top 10 Keywords

Keywords	1913	1938	1950-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2011	Total
Information theory	0.01%	0.01%	1.21%	2.46%	3.65%	3.36%	6.85%	34.94%	52.7%
Entropy	0.00%	0.00%	0.00%	0.00%	0.06%	0.11%	2.51%	12.64%	15.3%
System (s)	0.00%	0.00%	0.04%	0.07%	0.37%	0.00%	2.30%	10.39%	13.2%
Model (s)	0.00%	0.00%	0.01%	0.03%	0.08%	0.17%	1.60%	8.47%	10.4%
Communication	0.00%	0.00%	0.09%	0.02%	0.10%	0.16%	0.48%	6.60%	7.5%
Quantum	0.00%	0.00%	0.00%	0.02%	0.01%	0.03%	0.59%	6.24%	6.9%
Dynamics	0.00%	0.00%	0.00%	0.03%	0.07%	0.08%	1.09%	4.42%	5.7%
Network(s)	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%	0.54%	4.52%	5.1%
Capacity	0.00%	0.00%	0.02%	0.01%	0.01%	0.00%	0.40%	4.55%	5.0%
Statistical mechanics	0.00%	0.00%	0.02%	0.03%	0.00%	0.00%	0.25%	4.54%	4.8%

As author keyword analysis offers information about research trends in the view of researchers, it has proven to be important for monitoring the development of science (Li et al., 2009). In this study, 8,930 different author keywords were used from 1913 to 2011. Forty percent (3.572) appeared once, reflecting a wide difference in research focuses. Table 12 lists the top 10 -used keywords in the 89 years, with the most frequently used keywords being Information theory (52.7%), entropy (15.3%), system (s) (13.2%), and model(s) (10.4%).

Discussion and Conclusions

This study of information theory outputs showed some significant points on the worldwide research trends and performance from 1913 to 2011. The publication outputs about information theory increased during 89 years, but there were fluctuations in trend of growth. The products in the field of information theory from 1913 to 2011 showed a growth rate of 7.5 percent. In total, there are 8,466 publications in eleven types of document and sixteen languages listed in 123 subject categories

related to the research topic of information theory indexed in the *Web of Science* database.

Findings showed that *IEEE Transactions on Information Theory* published the most documents of information theory, followed by *Physical Review A.* and *Physica A: Statistical Mechanics and It Applications*. Of the 91 countries represented in the research publications, the top three were USA, UK, and Peoples Republic China.

Results by institution indicate that MIT University published more documents than other universities and institutions followed by University of Illinois, National University La Plata. Plastino, Hayashi, Matsuda, Nalewajski, and Casas were the most productive authors in the field of information theory who authored the largest number of papers. The average 13.74 pages and 28.68 references per article indicate average publication characteristics during the time frame of this study, 1913- 2011. The articles with greatest impact (most highly cited papers in the field of information theory) were authored by Turk and Pentland (1991), Jaynes (1957), and Maes et al (1997).

During recent decades, the phenomenon of co-authorship has drawn a considerable amount of attention among sociologists of science. Researchers use formal and informal scholarly communication. The formal communication are papers published in scholarly journals that have been reviewed by peers. The informal scholarly communication can be meetings, calls between researchers, and pre and post prints (Lacy & Busch, 1983). So, scientific collaboration, in which two or more authors collaborate to create one scientific work, and in recent years, factors such as specialization and the growth of interdisciplinary research have prompted researchers to cooperate with each other. Findings of this study indicate that 29.60 percent of the authorship was by one author, 30.38 percent was two authors, and 40.02 percent was three or more authors. Therefore, the authorship pattern of one author has held steadily while the published papers by two and three or more authors has been rising as 70.40 percent of published papers in the research period were produced collaboratively and only 29.60 percent of papers were produced individually and institutional collaboration was dominant.

Research publications in the field of information theory were indexed by the keywords information theory, entropy, system, and model(s) and among the top subject categories were computer science, engineering, and physics. Hopefully, this paper provides some insight into the current state of information theory research such as the characteristics of research activities, writing patterns, publication patterns, as well as countries, institutions, journals, and authors with greatest impact. The results of this research can provide unique and detectable indicators that can serve as benchmarks for future research.

References

American Heritage Science Dictionary. (2014). Information theory. Boston: Houghton Mifflin Company. Retrieved from http://www.thefree dictionary.com/information+theory

AT&T (2015). 1948: Information Theory. *Technology Timeline*. Retrieved from http://www.corp.att.com/attlabs/reputation/timeline/

Lacy, W.B. & Busch, L. (1983). Informal scientific communication in the agricultural sciences. *Information Processing & Management*, *19*(4), 193-202.

Li, L.L., Ding, G.H., Feng, N., Wang, M.H. & Ho, Y.S. (2009). Global stem cell research trend: Bibliometric analysis as a tool for mapping of trends from 1991 to 2006. *Scientometrics*, *80*(1), 39–58.

Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379-423, 623-656. Retrieved from http://www.cs.ucf.edu/~dcm/Teaching/COP5611-Spring2012/Shannon48-MathTheoryComm.pdf

Shannon, C.E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: Univ. of Illinois Press.

Sutter, M. & Kocher, M. (2004). Patterns of co-authorship among economics departments in the USA. *Applied Economics, 36*, 327-333.

Wen, H. & Huang, Y. (2012). Trends and performance of oxidative stress research from 1991 to 2010. *Scientometrics, 91*(1), 51-63.