

Feature Analysis of Suppressed Self-Citation Journals Based on JCR

Weibin Wang
*School of mangement
Harbin Institute of Technology
Harbin, China*

Tian Yu
*School of Economics and Management
Harbin Engineering University
Harbin, China*

Guang Yu*
*School of Mangement
Harbin Institute of Technology
Harbin, China
yug@hit.edu.cn*

Abstract—Science and Technology (S&T) Journals as a form of expression of scientific development achievements, S & T Journals play an important role in the process of knowledge dissemination and inheritance. However, in recent years, the alienation of citation behavior in S&T Journals will bring serious academic harm to the academic community. Therefore, Clarivate Analytics publishes JCR-suppressed journals every year while releasing JCR. Journals suppressed by JCR will also be published. Based on the type of journal self-citations pressing characterize, try looking for the typical characteristics of repression journal, for scientific research workers and S&T Journals evaluation provide certain a few suggestions.

Keywords—science and technology journals, impact factor, JCR-suppressed journals, self-citation

I. INTRODUCTION

Clarivate Analytics (formerly Thomson Reuters) published JCR (Journal Citation Reports). Clarivate Analytics published a list of suppression journals every year while releasing JCR. The types of suppression include (Type: Self and Stocking). The citations of abnormal journals are mainly reflected in the index factors such as journal impact factor (JIF), journal self-citation rate, and subject ranking. There exists serious abnormal citation behavior in the suppressed journal. Although JCR regularly publishes suppression lists every year, the data published by JCR every year is the citation data of the previous year, which will result in there is a significant lag in the JCR and repression reports, and the lag time is at least 1 year. Many researchers have submitted contributions to the journals and received their acceptance notice. Due to the lag of time, it will be discovered after the release of the JCR in the new year that the submitted papers have been removed by the JCR and will not be included in the SCI/SSCI database[1], which will seriously affect the dissemination of knowledge. In the latest JCR (2017) launched in 2018, 12 journals were suppressed, and the self-citation rate of these journals increased rapidly, which led to the rapid increase of impact factor. The self-citation rate of some journals was even as high as 90%, these journals cannot truly reflect the value of journals. For researchers, it may be important for those who are research-active to revisit the original intent of the JIF methodology and concurrently, educate the younger ones to prevent them from falling into this vicious cycle of publication gaming[1].

In this study, we used JCR-suppressed self-citation journals as research data, there is strong concealment between the stocking type journals, and data acquisition is more difficult so

in this study, JCR-suppressed journals only contains self-citation journals.

At present, in the academic world, there are two views on self-citation research: one is that self-citation should be excluded from the cites, and the other is that there is no need to exclude self-citation. Some researchers have found that self-citation does not have an excessive impact on the impact factor, and there is no significant correlation between self-citation and impact factor[2-4]. Therefore, it is not necessary to exclude self-citation in the science citation index[5-6], another part of the researchers suggested that when the citation index was published, a separate citation data was extracted[7], and Flatt proposed to establish a new citation indicator centering on scientific research transparency and using the h index as the reference — S index, calculated in the same way as the h-index, to provide a self-citation ratio[8]. Although there are some differences in these views, one thing is to reach a consensus, that is, not all self-citation exist problems, there is no need to exclude all self-citation, but it is necessary to exclude the journal coercive self-citation. In the JCR-suppresses journals, there is a clear coercive self-citation behavior. Clarivate Analytics is constantly changing in response to this coercive self-citation behavior in order to improve the journal impact factor. They began publish a series of journal citation indicator that exclude journal self-citation. Clarivate Analytics is also constantly identifying and adjusting journals included in JCR[9]. Studies have shown that journal self-citation typically account for 7% to 20% of total cites[10-12], so journals with more than 20% self-citation in the Web of Science database are considered Suspicious manipulation of journals. From our analysis, not all journals suspected of coercive self-citation have a self-citation rate of 20%. The self-citation journals suppressed by JCR were not suppressed according to the 20% standard.

The purpose of this study is to study the main features of JCR-suppressed journals and to provide contributors with certain early warning recommendations. Mainly from the following aspects: Firstly, we need to select the relevant characteristics of the journal self-citation behavior, according to the existing research of the laboratory team[13], select 15 indicators as the characteristic indicators to describe the journal self-citation behavior, use factor analysis and cluster analysis to study the characteristics of the journals. Secondly, use statistical methods to compare the suppressed journals with normal journals. Finally, summarize the main features of the JCR-suppressed journals, and serve as journal publishers and researchers, provide relevant advice.

This work was supported by the National Natural Science Foundation of China (No. 71704035), the China Postdoctoral Science Foundation (No. 2017M611361) and the Fundamental Research Funds for the Central Universities (No. HEUCFM180902)

II. DATA AND PROCESSING

A. Data

The data can be obtained from the usual Web of Science interface available from Clarivate Analytics, formerly managed by Thomson Reuters. Data derived from JCR 2017, all data come from manual collection. According to the previous research of the laboratory team on journal self-citations[14], the follow indicators were selected to describe the features of journal self-citation: Journal impact factor (X1), Journal impact factor without Journal self-cites(X2), Journal impact factor in the previous 1 year (X3), Total citation count (X4), Total self-citation count (X5), Citation count in the given year (X6), Self-citation count in the given year (X7), Citation count in the previous 2 years (X8), Self-citation count in the previous 2 years (X9), Citation count in the previous 3 years (X10), Self-citation count in the previous 3 years (X11), Total self-citations rate (X12), Self-citation rate in the given year (X13), Average self-citation rate in the previous 2 years (X14), Average self-citation rate in the previous 3 years(X15).

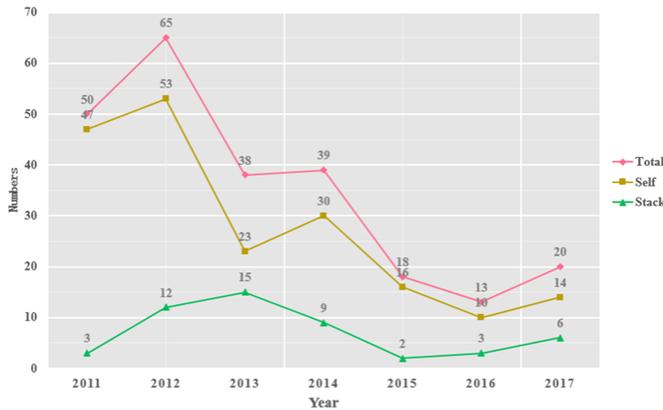


Fig. 1. JCR-suppressed journals

There were 243 suppressed journals in JCR, Select the year for 2011-2017(Fig.1),126 journals were eliminated due to missing data or type repeat, so 117 journals self-citations remained. Due to suppressed journals data cannot be obtained in the current year, So select the data from the year before the suppression. For example, 2012 was suppressed, select year 2011, Table I shows descriptive statistics of each indicator.

B. Methodology

At present, in the field of scientometrics, researchers often use factor analysis and cluster analysis to evaluate journals. In this study, the factor analysis is used to reduce the self-citation indicators of the journals, and the indicators with certain correlations are classified into one category, 15 indicators are compressed to obtain four main factors, thus summarizing that can represent self-citation journals features the categories, and then use these principal components to cluster the factor scores of each sample on the corresponding factors, and classify 117 journals. In order to find out the differences and connections between the samples, all statistical and numerical computations were performed using the spss22 and excel 2016.

TABLE I. DESCRIPTIVE STATISTICS OF INDICATORS

Indicators	Min	Max	Mean	SD
X ¹	0.25	10.72	1.62	1.44
X ²	0.05	5.63	0.61	0.65
X ³	0.17	7.79	1.16	1.01
X ⁴	106.00	4852.00	902.22	925.51
X ⁵	13.00	1971.00	358.16	421.54
X ⁶	0.00	322.00	38.11	59.34
X ₇	0.00	282.00	26.27	43.40
X ⁸	11.00	2882.00	296.55	432.95
X ⁹	1.00	1308.00	173.91	246.23
X ¹⁰	20.00	3479.00	400.04	563.96
X ¹¹	2.00	1493.00	216.24	300.43
X ¹²	0.02	0.90	0.41	0.20
X ¹³	0.00	1.00	0.59	0.34
X ¹⁴	0.05	0.96	0.58	0.21
X ¹⁵	0.06	0.96	0.54	0.20

C. Analysis of Results

1) Factor analysis: Prior to factor analysis, KMO and Bartlett test must be conducted to decide whether factor analysis was applicable. The KMO and Bartlett test are two important test methods used to test whether the original variable is suitable for factor analysis. The KMO test is used to test whether the sample data is suitable for factor analysis. Bartlett's test of sphericity is used to test for independence between indicators. Results showed that KMO test value was 0.686, which was higher than empirical value 0.5. Bartlett test value was 3309.694(P= 0.000 < 0.05). The hypothesis of not using factor analysis was rejected, indicating that factor analysis was applicable. The comprehensive KMO and Bartlett test analysis results show that the data in this study is suitable for factor analysis.

Usually, Factor analysis eigenvalue over 1, the cumulative contribution rate of the principal component corresponding to greater than 85%. Table II shows first factor F1, second factor F2, third factor F3, fourth factor F4 are 6.237, 2.954, 2.735, 1.339, respectively, both over 1, The variance contribution are 41.579%, 19.691%,18.232%, 8.927%.The cumulative variance contribution rate reached 88.428%, the information of the four common factors responding to the original variable of 88.428%, It shows that the common factors can reflect the original data, with less information lost. The common factor load matrix is established by the orthogonal rotation method (Varimax), and the rotation converges after 5 iterations, Rotated matrix is illustrated in Table II.

The first factor (F1) mainly reflects indicators X4, X5, X6, X7, X8, X9, X10, X11, which mainly explained the journal citation count and self-citation count. It mainly reflects the features of the number of journal citations. The secondly factor (F2) mainly reflects indicators X12, X14, X15, which mainly explained the Average self-citation rate and total self-citations rate, It mainly reflects the features of the self-citation rate. The third factor (F3) mainly reflects indicators X1, X2, X3, which mainly explained the indicators related to impact factors, It mainly reflects the features of the journal impact. The fourth

factor (F4) mainly reflects indicators X13, which mainly explained the self-citation rate in the given year.

TABLE II. ROTATED COMPONENT MATRIX

Indicators	Component			
	F ₁	F ₂	F ₃	F ₄
X ¹	0.350	0.115	0.897	0.039
X ²	0.317	-0.316	0.807	-0.002
X ³	0.324	0.063	0.794	0.075
X ⁴	0.878	-0.150	0.178	0.082
X ⁵	0.924	0.167	0.149	0.086
X ⁶	0.699	-0.049	0.410	0.407
X ₇	0.622	0.057	0.250	0.603
X ⁸	0.919	0.008	0.333	0.023
X ⁹	0.895	0.201	0.322	0.056
X ¹⁰	0.940	-0.005	0.283	0.014
X ¹¹	0.922	0.198	0.263	0.032
X ¹²	0.170	0.888	-0.013	0.078
X ¹³	0.008	0.269	-0.023	0.862
X ¹⁴	0.001	0.945	-0.004	0.162
X ¹⁵	0.026	0.973	-0.026	0.083
Eigenvalue	6.237	2.954	2.735	1.339
Variance%	41.579	19.691	18.232	8.927
Cumulative%	41.579	61.270	77.505	88.428

After factor analysis, 15 indicators can be reduced to 4 indicators. In the future analysis process, we will directly use these 4 indicators for analysis. The advantage of this is that it can be better clustered than the original variables. As a result, the analysis is easier, reducing the number of variables, but the information contained is not reduced.

According to the original coefficient coefficients and standardized value, the regression method is used to calculate the scores of the common factors. The score function is constructed by the weighted contribution of the four common factors of 41.589%, 18.931%, 18.264%, and 8.918% as the common factors.

$$F_1 = 0.878 * X_4 + 0.924 * X_5 + 0.699 * X_6 + 0.622 * X_7 + 0.919 * X_8 + 0.895 * X_9 + 0.940 * X_{10} + 0.922 * X_{11}$$

$$F_2 = 0.888 * X_{12} + 0.945 * X_{14} + 0.973 * X_{15}$$

$$F_3 = 0.897 * X_1 + 0.807 * X_2 + 0.794 * X_3$$

$$F_4 = 0.862 * X_{13}$$

Based on the individual factor scores and the composite factor scores, 117 rankings can be derived accordingly. The results of the factor scores are mainly used for the classification of journals in cluster analysis, so the specific ranking is not the focus of this study and will not be discussed in the study.

2) *Cluster analysis*: In order to analyzing the differences of journals in the further, 117 samples can be clustered based on four main factors. Clustering 117 samples is a problem to be discussed. First, used K-Means cluster analysis, according to the analysis of variance, we can know that the sample set must

be divided into at least 4 categories in order to be able to basically meet the requirements of significance. The variance analysis shows in Table III.

TABLE III. ANOVA

	Cluster		Error		F	Sid.
	Mean Square	df	Mean Square	df		
F ₁	20.495	3	0.482	113	42.486	.000
F ₂	1.582	3	0.985	113	1.607	.092
F ₃	24.901	3	0.365	113	68.137	.000
F ₄	24.667	3	0.371	113	66.443	.000

TABLE IV. CLUSTER ANALYSIS RESULTS

	Number	Journal sample number
Cluster1	108	3 49 84
Cluster2	3	32 63
Cluster3	2	47 70 90
Cluster4	4	others

TABLE V. MEAN AND SD OF FOUR TYPES OF JOURNALS

	Mean/SD			
	Cluster 1	Cluster 2	Cluster 3	Cluster 4
X ¹	1.36/0.75	8.42/2.50	3.53/2.50	2.59/1.17
X ²	0.53/0.33	3.40/1.94	1.96/1.44	0.77/0.51
X ³	1.01/0.70	4.79/2.60	2.66/1.54	1.77/0.97
X ⁴	772.00/709.72	1950.33/1329.91	4763.50/125.16	1701.50/1212.89
X ⁵	291.27/322.44	940.33/759.62	1886.00/120.21	963.75/477.45
X ⁶	24.32/24.25	158.00/146.77	243.00/33.94	218.00/59.91
X ₇	17.08/21.59	63.33/42.19	125.00/28.28	197.00/70.52
X ⁸	213.47/227.13	1103.33/737.17	2623.50/365.57	771.00/438.08
X ⁹	128.35/156.29	656.00/593.87	1180.00/115.97	539.00/270.81
X ¹⁰	294.85/305.28	1359.33/935.54	3476.00/4.24	982.75/557.93
X ¹¹	162.38/199.67	753.66/666.25	1468.50/34.65	641.25/303.80
X ¹²	0.39/0.19	0.48/0.14	0.40/0.01	0.65/0.19
X ¹³	0.59/0.35	0.50/0.32	0.51/0.04	0.92/0.65
X ¹⁴	0.58/0.210.20	0.58/0.22	0.45/0.02	0.72/0.13
X ¹⁵	0.53/0.20	0.54/0.18	0.42/0.01	0.69/0.14
N	108	3	2	4

It can be known from Table III, F1, F3, F4 P values were calculated, P value<0.05 was considered significant, which satisfies requirement, and the F2 significance is greater than 0.05, but for the convenience of analysis, we are still divided into four categories. After the above analysis, we conducted a system cluster analysis of the sample by clustering method. The clustering method uses Between-groups linkage and the distance test selects the Square Euclidean Distance method. Cluster results shows in Table IV.

The mean of each class can reflect the features of the each class. By calculation, we obtain the mean and standard deviation of the four classes, The compare analysis is shown in Table V. The second category is a journal with a high impact factor. The indicators reflected the impact factor are very large. We also found that the impact factor of the three samples were 10.72, 8.79, and 5.756, respectively, which was the highest among all samples. The third category is a type of journal with a high citation, and the journals with low citation rates ,but have very large indicators of citations, and the citation rate is the lowest among all samples. The fourth category is a type of journal with high citations, especially the self-citation rate is very high, even reaching 90%.The first category is a relatively balanced journal with a large sample size and average indicators in all aspects. But no matter which type of journal, their self-citation rate is greater than 50%.

3) Compare between suppressed journals and normal journals: According to the results of factor analysis and cluster analysis, citation rate indicators , citation indicators and impact factor related indicators are the main features of the suppressed journals. Therefore, it is necessary to study the changes of self-citation and impact factor in normal journals and suppressed journals. However, the above indicators are all calculated by us. It is more troublesome to identify the suppressed journals. Therefore, we choose to calculate the journals self-citation rate(SCR)and average journal impact factor percentile(AJP)in the current year.(Average JIF percentile converted JIF ranking into percentile to provide more intuitive transverse comparison between journals)

In view of Clarivate Analytics efforts in journal selection and supervision, they continue to identify and adjust the journals included in the JCR. Normal journals still account for the majority of all journals, and such journals are relatively easy to select. Through a questionnaire survey of surrounding researchers, we selected 50 self-citation normal journals for comparative study, and the selected year was 2011-2016. Suppressed journals select 67 from 117 and the selected year was 2011-2016 . SCR and AJP indicators are complete for the first 5 years of the year of suppression. Fig. 2 and Fig. 3 show that with the change of the year, the mean value of SCR and AJP in the suppressed journals is increasing year by year, the increase is very obvious, the SCR of some journals even reaches 100%.The lowest value from Group1 to Group5 SCR is also increasing year by year, and the lowest value of Group5 SCR is more than 20%. Fig. 4 and Fig. 5 show that the mean value of SCR and AJP of the normal journals remained stable, and the SCR remained at 10%, and the AJP changes remained stable.

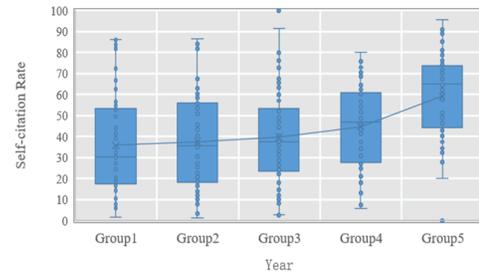


Fig. 2. JCR-suppressed Journal's SCR

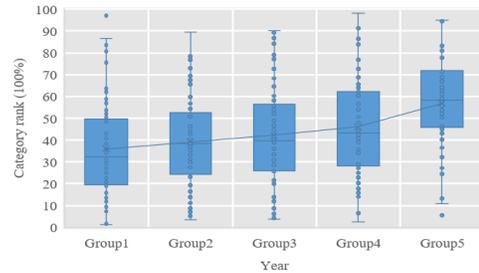


Fig. 3. JCR-suppressed Journal's AJP

Boxplot, separated for the1-5 groups, the suppressed year is Y, Group1 - Group5 represents the year Y-5 (the 5th year before the suppressed) to Y-1 (1 year before the suppressed). The horizontal line in the middle of the each box denotes the median, while the bottom and top borders of the box mark the 25th and 75th percentiles, respectively. The whiskers below and above the box mark the 5th and 95th percentiles, respectively. The points beyond the whiskers are outlier.

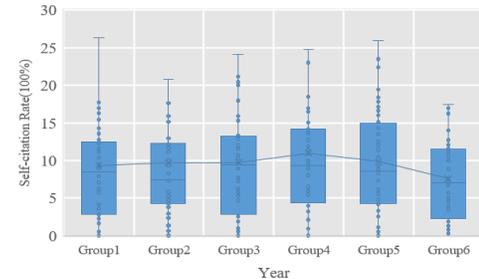


Fig. 4. Normal journal's SCR

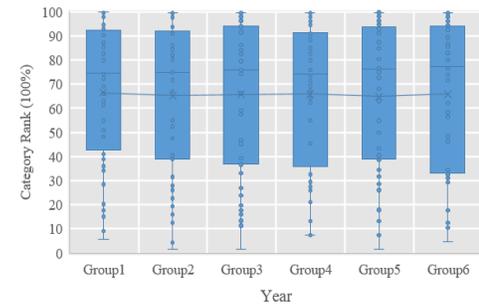


Fig. 5. Normal journal's AJP

Boxplot, separated for the1-6 groups, Group1 - Group6 represents the year 2011 – 2016. The horizontal line in the middle of the each box denotes the median, while the bottom and top borders of the box mark the 25th and 75th percentiles, respectively. The whiskers below and above the box mark the 5th and 95th percentiles, respectively. The points beyond the whiskers are outlier.

III. CONCLUSIONS AND DISCUSSIONS

According to the results of factor analysis and cluster analysis, we can divide suppressed journals into four categories: journals with high impact factor; journals with high citations but low citation rates; journals with high citation rates; Journals with more balanced indicators. Many journal publishers have improved the journal's own influence by adjusting a series of indicators such as citations, citation rates, and impact factor. This study further proves this point. In order to facilitate the easy identification of abnormal journals, we selected SCR and AJP as the main indicators to identifying normal journals and suppressed journals.

In order to maintain the accuracy, fairness, scientificity, reliability and authority of the data provided by itself, Clarivate Analytics maintains a high degree of vigilance in the detection of citation anomalies and uses repressive means to punish man-made citations. From Fig. 1, we can find that 2012 is the peak of suppression, and it is also a watershed. The number of suppressed journals is gradually decreasing. The decrease in quantity reflects to some extent that JCR's suppression policy has a certain positive effect. It can be clearly found that inhibition is an effective tool, and the suppression policy has a greater impact on the citation mode of the journal, and to some extent can suppress the high frequency self-citation number. In order to maintain their reputation, journal publishers will also begin to converge on their own violations.

Firstly, Journal self-citation rate of more than 20% in the Web of Science database are considered to be suspicious manipulated journals. It is clear that this study that the journal self-citation rate is not limited to 20%, and the threshold of suppression is defined as ambiguous, ranging from 20% to 90%. It is recommended to give clear standard and inhibition indicators in future JCR for researchers to refer to.

Secondly, through the comparative study of normal journals and suppressed journals, it can be found that SCR and AJP are greatly changed and the volatility is large, while the normal journals remain relatively stable and there is no abnormal fluctuation. so, when submitting a manuscript, researchers can properly observe the trends of SCR and AJP in journals. we need to pay attention to the journal self-citation rate is too high is not normal, but close to 0 self-citation rate is not normal, which indicates that the journal has no inheritance of knowledge.

Thirdly, JCR staff deserve recognition for the behavior of partially suppressed journals, and contributed to the creation of a better academic environment for researchers[14], but for SCR and AJP journals with huge fluctuations, it is recommended that JCR staff should strengthen Supervision of such journals.

Whether it is an editor or other person, raising the JIF by man-made journal citations is detrimental to the fairness of the

journal evaluation, and does not improve the quality of the journal. References are an important part of the scientific literature. By including each document in the list of references, the author should bear the corresponding responsibility, even if the journal editors and reviewers force them to cite papers that should not appear. the author should protect their own rights, resolutely resists this bad behavior against scientific morality [15]. Simultaneously. It is also necessary to call on all editors and publishers, researchers and journal evaluation agencies to join forces to resist violations and to maintain a healthy and sustainable development, fair competition and a vibrant research environment.

REFERENCES

- [1] Foo, Jong Yong Abdiel, "Impact of Excessive Journal Self-Citations: A Case Study on the Folia Phoniatrica et Logopaedica Journal," *Science & Engineering Ethics*, vol. 17, pp. 65-73, 2011.
- [2] Mirsaeid, Seyed Javad Ghazi, N. Motamedi, and N. R. Ghorbani. "Correlation between Self-Citation and Impact Factor in Iranian English Medical Journals in WoS and ISC: A Comparative Approach," *Iranian Journal of Public Health*, vol. 44, pp. 1234-1243, 2015.
- [3] Mimouni, Michael, et al, "Self-citation rate and impact factor in pediatrics," *Scientometrics*, vol. 198, pp. 1-6, 2016.
- [4] González, Lidia, and J. M. Campanario, "Structure of the impact factor of journals included in the Social Sciences Citation Index: Citations from documents labeled 'editorial material'," *Journal of the Association for Information Science & Technology*, vol. 58, pp. 252-262, 2010.
- [5] Aksnes D W, "A macro study of self-citation," *Scientometrics*, vol. 56, pp. 235-246, 2003.
- [6] Wolfgang Glänzel, Thijs B, "The influence of author self-citations on bibliometric macro indicators," *Scientometrics*, vol. 56, pp. 281-310, 2004.
- [7] Wolfgang Glänzel, et al, "A concise review on the role of author self-citations in information science, bibliometrics and science policy," *Scientometrics*, vol. 67, pp. 263-277, 2006.
- [8] Flatt, Justin, et al, "Improving the Measurement of Scientific Success by Reporting a Self-Citation Index," *Publications*, vol. 5, pp. 20, 2017.
- [9] Wilhite, Allen W., and E. A. Fong, "Coercive Citation in Academic Publishing," *Science*, vol. 335, pp. 542-543, 2012.
- [10] Garfield, E, "How Can Impact Factors Be Improved?," *Bmj British Medical Journal*, vol. 313, pp. 411-413, 1996.
- [11] Garfield, E, "Dispelling a few common myths about journal impact factors," *The Scientist*, vol. 3, pp. 11, 1997.
- [12] Tiew W S, "Characteristics of Self-Citation in Journal of Natural Rubber Research 1988-1997: a Ten-Year Bibliometric Study," *Malaysian Journal of Library and Information Science*, vol. 5, pp. 95-104, 2000.
- [13] Yu, Tian, G. Yu, and M. Y. Wang, "Classification method for detecting coercive self-citation in Journals," *Journal of Informetrics*, vol. 8, pp. 123-135, 2014.
- [14] Garfield, E, "The evolution of the Science Citation Index," *International Microbiology the Official Journal of the Spanish Society for Microbiology*, vol. 10, pp. 65-9, 2007.
- [15] Krell F T, "Losing the numbers game: abundant journal self-citations put journals at risk for a life without Impact Factor," *European Science Editing*, vol. 40, pp. 36-38, 2014.