

New Measures of Journal Impact Based on the Number of Citations and PageRank

Wataru Souma¹, Irena Vodenska², Lou Chitkushev³

¹College of Science and Technology, Nihon University, Funabashi, 274-8501 (Japan)
souma.wataru@nihon-u.ac.jp

²Metropolitan College, Boston University, Boston, MA 02215 (USA)
Center for Polymer Studies, Boston University, Boston, MA 02215 (USA)
vodenska@bu.edu

³Metropolitan College, Boston University, Boston, MA 02215 (USA)
lrc@bu.edu



*Journal of Digital
Information Management*

ABSTRACT: *The number of citations has been used for measuring the significance of a paper. Moreover, we have the following question: which paper is the most important if there are some papers with the same number of citations? Some measures have been introduced to answer this question: one of them is PageRank. We use the Science Citation Index Expanded from 1981 to 2015 to calculate the number of citations and the Google number in the citation network consisting of 34,666,719 papers and 591,321,826 citations. We clarify the positive linear relationship between the number of citations and the Google number, as well as extract some outliers from this positive linear relationship. These outliers are considered to be extremely prestige papers. Furthermore, we calculate the mean values of the number of citations and the Google number for all journals, construct a new measure of journal influence, and extract extremely prestige journals. This new measure has a positive and medium correlation with the impact factor, Eigenfactor score, and SCImago Journal Rank.*

Subject Categories and Descriptors:

[H.5.3 Group and Organization Interfaces]; Web-based interaction: [H.3.3 Information Search and Retrieval]; Search process

General Terms: Page Rank, Journal Evaluation

Keywords: Page Rank, Journal Impact, Citation Impact, Google Page Algorithm, Impact Factor

Received: 30 August 2019, Revised 4 November 2019, Accepted 18 November 2019

Review Metrics: Review Scale: 0-6, Review Score: 5.35, Inter-reviewer consistency: 94%

DOI: 10.6025/jdim/2020/18/1/11-20

1. Introduction

The number of citations is the most frequently used measure to quantify the significance of papers. Meanwhile, we have the following question: which paper is the most important if there are some papers with the same number of citations? Some measures have been introduced to answer this question: one of them is PageRank that was proposed by Brin and Page (1999).

Bollen et al. (2006) described the Institute for Scientific Information impact factor (IF), which is defined as the mean number of citations that a journal receives over two years, as a metric of popularity, and PageRank as a metric of prestige. Chen et al. (2007) calculated the number of citations and the Google number for all papers in the Physical Review family of journals from 1893 to 2003. They observed a linear relationship between the number of citations and the Google number. Additionally, they discovered that some outliers from this linear relationship were universally familiar to physicists, especially papers in which

the ranking of PageRank is substantially high and that of citation is slightly high. Thus, they called these papers scientific “gems.” This structure was also studied by Maslov and Redner (2008). Ma et al. (2008) confirmed this structure in the citation networks of biochemistry and molecular biology.

These previous studies have investigated the citation networks of some selected scientific fields, but no study has been conducted on the application of the concept of PageRank to all papers in all fields of science. Thus, one aim of the present study is to obtain prestige papers (Souma & Jibu, 2018) in all fields of science. Additionally, by employing the number of citations and the Google number of each paper contained in each journal, this study aims to obtain the mean values of the number of citations and the Google number for each journal. By utilizing the relationship between the mean values of the number of citations and the Google number, we propose a new measure of journal influence.

The remainder of this paper is organized as follows: In the next section, we explain the data used in this study and illustrate some basic stochastic properties, i.e., the yearly change of the number of published papers, the size distribution of connected components in the citation networks, and the distribution of the number of citations. Subsequently, we show the relationship between the number of citations and the Google number of each paper and present some extremely prestige papers. Furthermore, we define a new measure of journal influence and compare it with the IF, Eigenfactor score, and SCImago Journal Rank. The last section is devoted to the summary and discussion.

2. Data and Definition of the Google Number

We use the Science Citation Index Expanded (SCIE) provided by Clarivate Analytics Co., Ltd. This dataset contains bibliographic information of scientific papers

published from 1900 to the present. However, due to the limited research grants of the authors, we utilized the dataset from 1981 to 2015 in this study. This dataset contains 34,666,719 papers and 591,321,826 citations.

Let $n(t)$ denote the number of papers published in the year t . Figure 1 depicts the change of $n(t)$. This figure shows that $n(t)$ almost monotonically increased from 1981 to 2013, whereas it decreased after 2013. However, this behavior of $n(t)$ is fake. This is because the dataset was developed at the beginning of 2016 and partially contained papers published in 2014 and 2015. It took a few years for all the published papers to be included in the SCIE.

By considering papers as nodes and citations from a citing paper to a cited paper as directed links, we can consider the dataset of citations as a directed network. We refer to this network as the citation network. The citation network comprises numerous connected components. We denote the number of nodes contained in connected components by c and the frequency of c by $F(c)$, which is depicted in Figure 2. This figure illustrates the giant weakly connected component (GWCC). The GWCC comprises 34,428,322 nodes, which are 99.3% of the total number of papers contained in the dataset, and 591,177,607 directed links, which are 99.98% of the total number of citations contained in the dataset. We focus on the GWCC below.

The number of citations of paper i , k_i , corresponds to the number of in-degrees, k_i^k , of node i in the citation network. Figure 3 depicts a double-logarithmic scale plot of the rank size distribution, $R(k)$. The right-tail part of the distribution decreases almost monotonically. This implies that this part follows a power-law distribution: $R(k) \propto k^{-\mu}$, where the exponent, μ , denotes the power law exponent or the Pareto exponent that originated from the name of an Italian economist, Vilfredo Pareto.

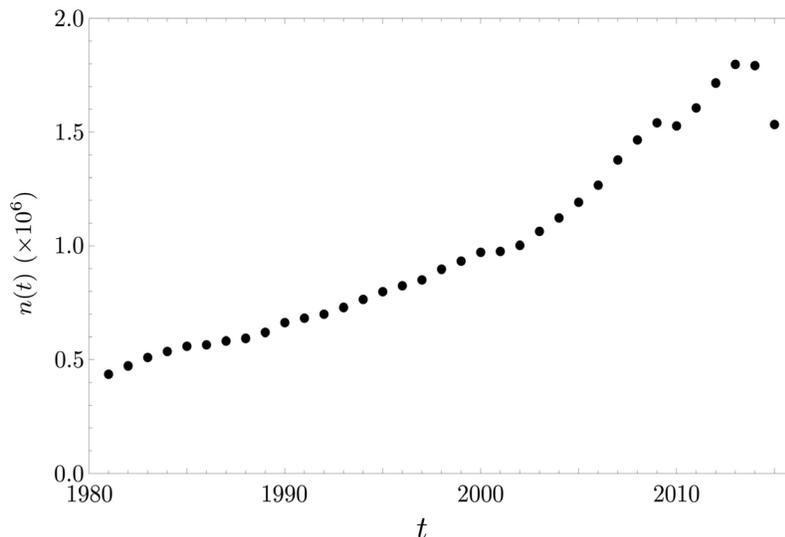


Figure 1. Yearly change of the number of published papers. The abscissa denotes the year, t , whereas the ordinate denotes the number of papers, $n(t)$, published in the year t (Souma & Jibu, 2018)

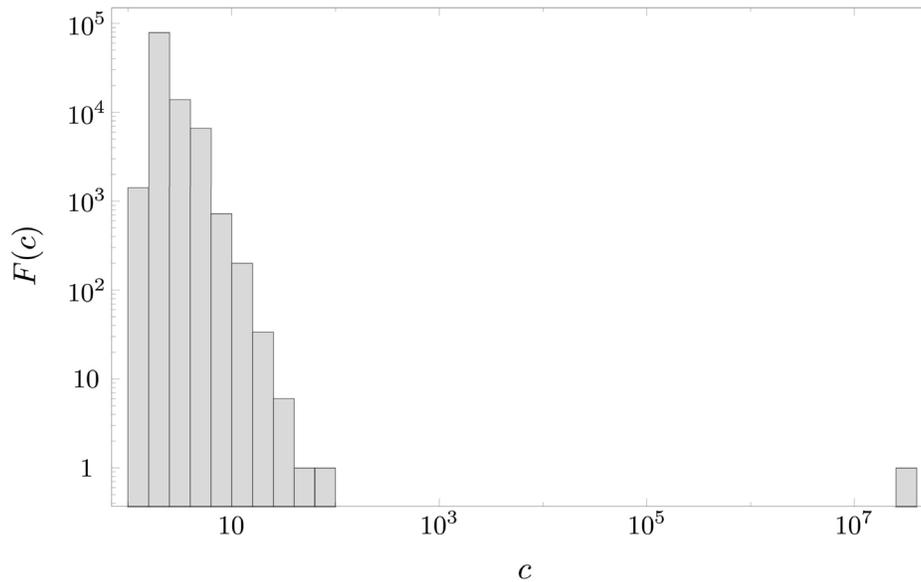


Figure 2. Size distribution of connected components. The abscissa denotes the size of the connected component, c , and the ordinate denotes the frequency, $F(c)$ (Souma & Jibu, 2018)

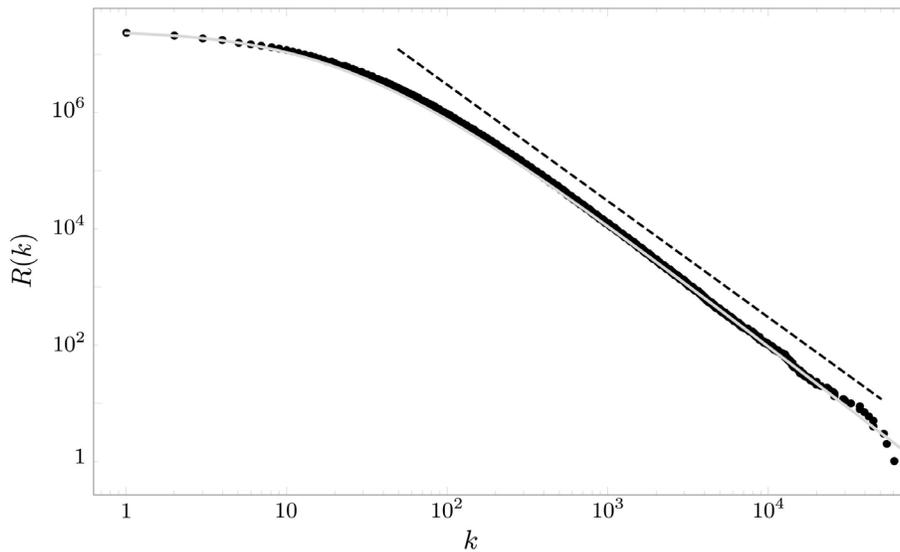


Figure 3. Rank-size distribution of the number of citations. The abscissa denotes the number of citations, k , whereas the ordinate denotes the rank-size distribution, $R(k)$ (Souma & Jibu, 2018)

The dashed line in Figure 3 represents the reference line, which is the power-law distribution with $\mu = 2$: $R(k) \propto k^{-2}$. The light gray line in Figure 3 denotes the best fit by the generalized beta distribution of the second kind (GB2) (or called the beta prime distribution), which is given by

$$f(k; a, b, \mu, \nu) = \frac{ak^{a\mu-1}}{b^{a\mu}B(\mu, \nu)} \left[1 + \left(\frac{k}{b}\right)^a \right]^{-(\mu+\nu)} \quad (1)$$

where $\beta(\mu, \nu)$ denotes the following beta function:

$$\beta(\mu, \nu) = \int_0^1 x^{\mu-1}(1-x)^{\nu-1} dx. \quad (2)$$

The light gray line corresponds to Equation (1) with $a = 0.7$, $b = 15.2$, $\mu = 2.0$, and $\nu = 3.0$.

Brin and Page (1999) were the first to propose the so-called PageRank for obtaining the appropriate ranking of a web page in the World Wide Web (WWW). PageRank of paper i is derived from the Google number, g_i , defined by the following recursion formula (Chen et al., 2007):

$$g_i = (1-d) \sum_{inn j} \frac{g_j}{k_j} + \frac{d}{N} \quad (3)$$

Here, $N = 34,428,322$ denotes the total number of papers contained in the GWCC. The sum is taken over the neighboring nodes j , which are link points to node i . In Equation (3), d denotes a free parameter that controls the convergence and effectiveness of the recursion calculation. In the original calculation of PageRank, $d = 0.15$ was adopted for the case of the WWW (Brin & Page, 1999). Mean

while, $d = 0.5$ was adopted for the case of the citation network (Chen et al., 2007). Following Chen et al. (2007), we use $d = 0.5$ in this study.

Correlation between the Number of Citations and the Google Number

First, we examine the correlation between the number of citations, k_i , and the Google number, g_i , of each paper and present some extremely prestige papers. Next, we calculate the mean values of the number of citations, K_j , and the Google number, G_j , for each journal and present some extremely prestige journals. Furthermore, we define a new measure of journal influence by utilizing the ranking of K_j and G_j and compare it with the IF, Eigenfactor score, and SCImago Journal Rank.

Paper-Level Investigation

The double-logarithmic scale scatter plot of k_i and g_i is depicted in Figure 4. In this figure, each black dot represents each paper. The cyan solid line represents the average value, $\langle g \rangle$, which is calculated for bins with logarithmically equal width against k . This figure shows that the graph of $\langle g \rangle$ versus k is smooth and increases linearly, especially in the range $k \geq 500$. Therefore, the number of citations and the Google number are almost similar measures for characterizing the significance of papers. This result demonstrates that prestige (the high Google number) is proportional to popularity (the high number of citations) in numerous cases and indicates the existence of outliers, i.e., extremely prestige papers.

We denote the ranking of the number of citations of paper

i by $r_{k,i}$ and that of PageRank (the ranking of the Google number) of paper i by $r_{g,i}$. By employing these two rankings, we define the average ranking as follows:

$$r_i = \frac{1}{2} (r_{k,i} + r_{g,i}) \quad (4)$$

We consider this as the new measure of paper influence. The green dots with the number depicted in Figure 4 represent the top 10 papers with a high value of r_i . These papers are located in the upper right part of the distribution by definition.

The numbers assigned to green dots in Figure 4 correspond to the rankings in Table 1. This table was arranged in descending order of r_i with the name of the first author and the published year; therefore, we can find them in the list of references. One of the unique natures of Table 1 is that the main subjects of the papers are biochemistry and molecular biology. This characteristic is the same as what is generally told, i.e., important papers are published in the fields of biology and chemistry. Another unique nature of Table 1 is that the years of the published papers are old except for Sheldrick (2008).

Figure 4 shows that there are papers in which $r_{g,i}$ is substantially high while $r_{k,i}$ is slightly high. We call these papers extremely prestige papers. These papers are located far above the cyan solid line in Figure 4. We introduce the ratio, $f_i = r_{k,i} / r_{g,i}$ to extract extremely prestige papers. Table 2 presents the top 10 extremely prestige papers based on the constraint $f_i > 10$. Here, we obtain the

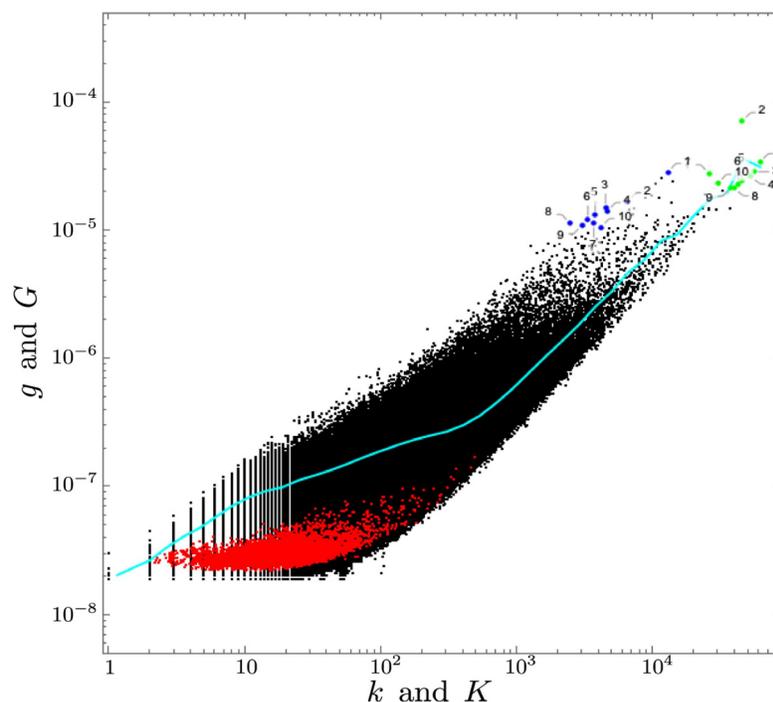


Figure 4. Scatter plot of the number of citations, k , and that of the Google number, g , for each paper (black dots). The cyan solid line represents the average value calculated for bins with logarithmically equal width against k . The scatter plot of the averaged value of the number of citations, K , and that of the Google number, G , for each journal (red dots)

Ranking	r_i	$r_{k,i}$	$r_{g,i}$	1st author (year)	Main subject
1	1.5	1	2	Chomczynski (1987)	Biochemistry & Molecular Biology
2	2.5	4	1	Sheldrick (2008)	Chemistry
3	3	2	4	Becke (1993)	Chemistry
4	5	3	7	Lee (1988)	Materials Science
5	7	5	9	Perdew (1996)	Physics
6	9.5	6	13	Thompson (1994)	Biochemistry & Molecular Biology
7	10.5	15	6	Bland (1986)	General & Internal Medicine
8	11	7	15	Altschul (1997)	Biochemistry & Molecular Biology
9	11	8	14	Altschul (1990)	Biochemistry & Molecular Biology
10	11	11	11	Otwinowski (1997)	Biochemistry & Molecular Biology

Table 1. Top 10 papers ordered by r_i

Ranking	$r_{g,i}$	$r_{k,i}$	f_i	1st author (year)	1st subject
1	5	64	12.8	Kennedy (1995)	Computer Science
2	22	240	10.9	Alamouti (1998)	Engineering
3	25	516	20.6	Akyildiz (2002)	Computer Science
4	33	481	14.6	Pawlak (1982)	Information Science
5	36	784	21.8	Akyildiz (2002)	Engineering
6	43	998	23.2	Gruber (1994)	Computer Science
7	48	828	17.3	Gupta (2000)	Computer Science
8	49	1916	39.1	Floyd (1993)	Computer Science
9	53	1247	23.5	Bianchi (2000)	Engineering
10	60	609	10.1	Haykin (2005)	Engineering

Table 2. Top 10 extremely prestige papers ordered by $r_{g,i}$ under the constraint $f_i > 10$

value of this constraint via repeated trial and error method. Therefore, there is no special scientific reason for this value. Thus, in our future study, we hope to find a special scientific reason for the value of the constraint.

In Table 2, PageRank, $r_{g,i}$, is arranged in ascending order. In this table, we write the name of the first author and the published year; therefore, we can find them in the list of references and also write the main subject of the papers. These 10 extremely prestige papers are represented by blue dots with numbers in Figure 4. These numbers correspond to the ranking in Table 2. This table demonstrates that extremely prestige papers mainly relate to computer science and engineering. This characteristic is different from that of Table 1.

Journal Level Investigation

We define the mean values of the number of citations, K_j ,

and the Google number, G_j , as follows:

$$K_j = \frac{1}{n_j} \sum_{i \in j} k_i, G_j = \frac{1}{n_j} \sum_{i \in j} g_i \quad (5)$$

where n_j denotes the number of papers contained in journal j . Here, $i \in j$ means that paper i is contained in journal j . The value of n_j depends on the journal. Herein, the journals considered to be appropriate for investigation are published yearly and contain at least one paper. We investigate the data of the SCIE for a period of 35 years (from 1981 to 2015), so we decide to ignore journals with $n_i < 35$. The total number of journals utilized is 15,533. The scatter plots of K_j and G_j are depicted as red dots in Figures 4 and 5, respectively. The red dots are distributed in the lower-left part, as illustrated in Figure 4. This means that the discrepancy between the journals is less than that between the papers from the perspectives of K_j and G_j .

However, as future work, it is also logical to calculate variances $\sigma_{K_j}^2$ and $\sigma_{G_j}^2$ for each journal and compare them between the journals.

In Figure 5, the black solid line represents the average value, $\langle G \rangle$, which is calculated for bins with logarithmically equal width against K . This figure depicts that the graph of $\langle G \rangle$ versus K is smooth and increases linearly, especially in the range $K \geq 50$. Thus, similar to the case of the paper-level investigation, it is expected that the extremely prestige journals should be distributed far above the black solid line in Figure 5.

We denote the ranking of the number of citations of journal j as R_{K_j} and PageRank (the ranking of the Google number) of journal j as R_{G_j} . By employing these two rankings, we define the following new measure, i.e., the average ranking:

$$R_j = \frac{1}{2} (R_{K_j} + R_{G_j}) \quad (6)$$

We consider this new measure of journal influence. Top 10 journals are shown as green dots with numbers in Figure 5. These numbers correspond to the ranking in Table 3, showing that the main subjects of the journals are mainly biology and chemistry.

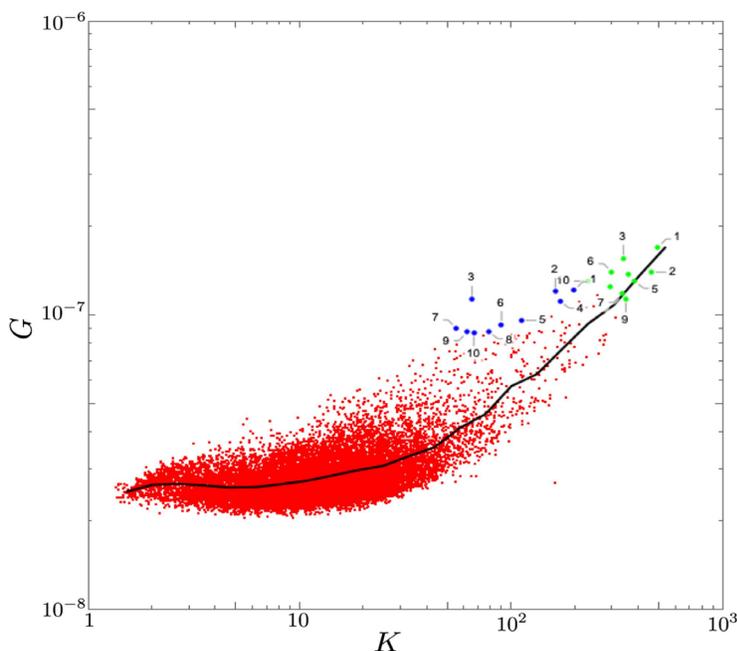


Figure 5. Scatter plot of the mean values of the number of citations, K_j and the Google number, G_j (red dots). The black solid line represents the average value calculated for bins with logarithmically equal width against K . Green dots correspond to the papers listed in Table 3. Blue dots correspond to the papers listed in Table 5

<i>Ranking</i>	R_j	R_{K_j}	R_{G_j}	<i>Journal</i>
1	1	1	1	Annual Review of Biochemistry
2	2.5	2	3	Annual Review of Immunology
3	4	6	2	Microbiological Reviews
4	4.5	4	5	Reviews of Modern Physics
5	4.5	3	6	Annual Review of Cell Biology
6	6.5	9	4	Annual Review of Plant Physiology and Plant Molecular Biology
7	9	7	11	Physiological Reviews
8	9	10	8	Annual Review of Ecology and Systematics
9	9	5	13	Annual Review of Neuroscience
10	13.5	20	7	Progress in Inorganic Chemistry

Table 3. Top 10 journals with constraint $n_j > 30$

Moreover, these rankings strongly depend on the lower bound of n_j . Table 4 is obtained by considering journals with $n_j > 10000$. This table shows that most journals belong

to medicine. This strong dependence on the lower bound of n_j suggests, as a future study, that we must precisely investigate the relationship between n_j and journal influence.

<i>Ranking</i>	R_j	R_{Kj}	R_{Gj}	<i>Journal</i>
1	27.5	13	42	Cell
2	76	39	113	Science
3	79.5	43	116	Nature
4	84	61	107	New England Journal of Medicine
5	84	61	107	Journal of Experimental Medicine
6	127	72	182	Journal of Clinical Investigation
7	187	176	198	JAMA-Journal of the American Medical Association
8	188.5	205	172	Annals of Internal Medicine
9	218.5	112	325	Journal of Cell Biology
10	222.5	101	344	EMBO Journal

Table 4. Top 10 journals with constraint $n_j > 10000$

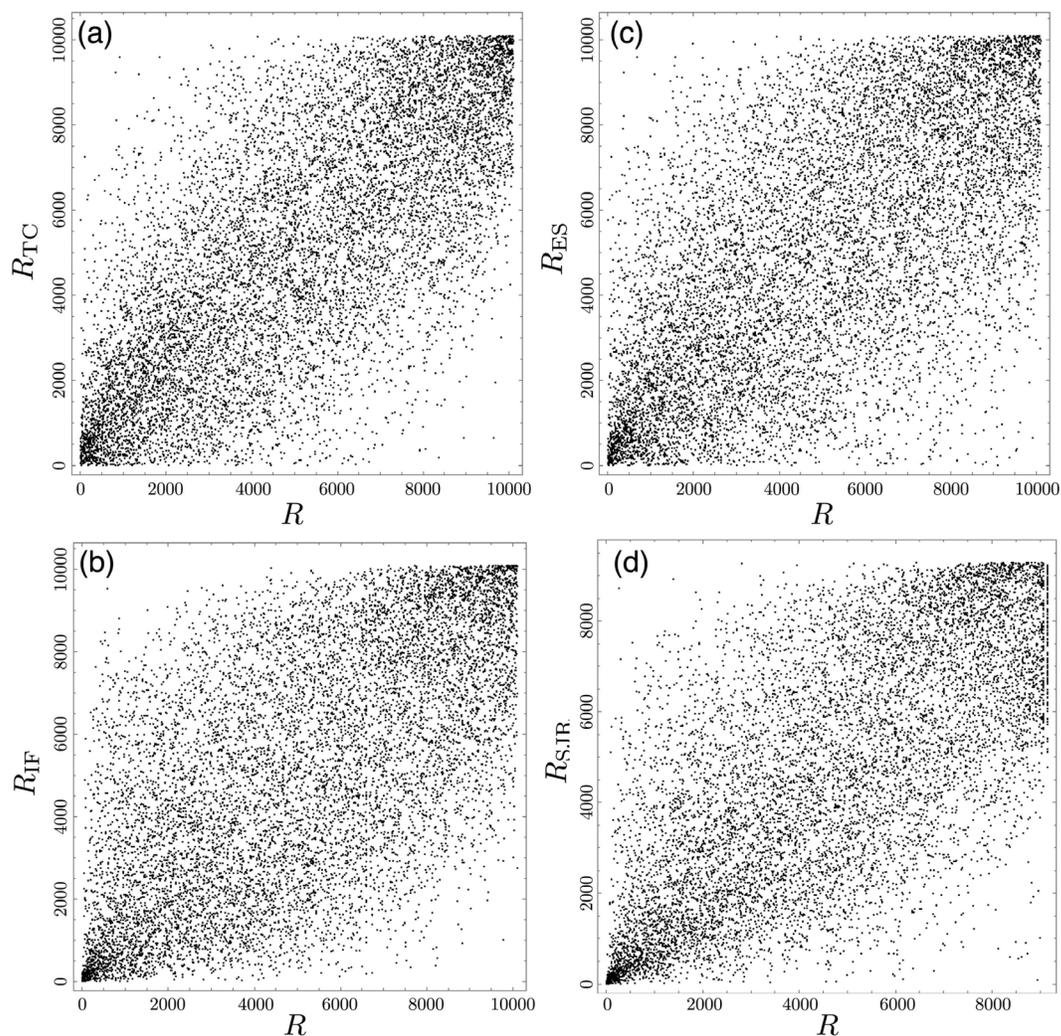


Figure 6. Scatter plots of R and (a) R_{TC} , (b) R_{IF} , (c) R_{ES} , and (d) R_{SJR}

It is also important to investigate the correlation between R and other rankings of journals. We can obtain the following rankings from the 2015 Journal Citation Report (JCR, 2015): the total citations, R_{TC} ; the journal IF, R_{IF} ; and the Eigenfactor score, R_{ES} . We compared our list of journals and those provided by JCR and obtained the ranking list of 10,106 journals. The ranking of the SCImago journal, R_{SJR} is also useful and available from the 2015 SCImago Journal and Country Rank (SJR, 2015). We compared our list of journals and those provided by SJR and obtained the ranking list of 9,283 journals. Figure 5 illustrates the ranking correlation between R and (a) R_{TC} , (b) R_{IF} , (c) R_{ES} and (d) R_{SJR} . This figure shows the positive correlation between R and other rankings. To confirm this observation numerically, we respectively calculated Spearman's rank correlation coefficient as follows: (a) $\rho=0.7231$, (b) $\rho=0.6335$, (c) $\rho=0.6354$, and (d) $\rho=0.6963$. Meanwhile, Spearman's rank correlation coefficient between R_{IF} and R_{TC} is $\rho=0.7521$, whereas that between R_{IF} and R_{ES} is $\rho=0.7905$. Thus, the correlation between our new measure R and other previously introduced measures is weaker than that between the previously introduced measures.

Figure 5 demonstrates that there are papers in which R_{Gj} is substantially high while R_{Kj} is slightly high. Such a journal is called an extremely prestige journal. These journals are located far above the black solid line shown in Figure 5. Similar to the paper-level investigation, we extract these extremely prestige journals by introducing the constraint defined by $F_j = R_{Kj}/R_{Gj}$. Table 5 presents the top 10 extremely prestige journals based on constraint $F_j > 3$. We obtained this constraint via repeated trial and error method; therefore, there is no essential scientific reason for this, which is similar to the paper-level investigation. We arranged the journals in descending order of R_{Gj} as illustrated in Table 5. These 10 extremely prestige journals are shown by blue dots with numbers in Figure 5.

These numbers correspond to the rankings shown in Table 5. This table lists various journals of numerous research fields. Albeit not as much as in the case of extremely prestige papers, there were numerous journals on information science.

4. Summary and Discussion

This study utilized a dataset containing 34,666,719 papers and 591,321,826 citations to investigate the relationship between the number of citations of paper i , k_i , and the Google number of paper i , g_i . We assigned the ranking of the number of citations, $r_{k,i}$ and PageRank, $r_{g,i}$ to each paper and define the new measure of paper influence, r_i . We extracted the top 10 papers with a high ranking of r_i and obtained that the main subjects of these papers were biochemistry and molecular biology.

Additionally, we obtained a positive linear relationship between k_i and g_i and extracted papers in which $r_{g,i}$ was substantially high while $r_{k,i}$ was slightly high by considering the constraint $r_{k,i}/r_{g,i} > 10$. We called such a paper an extremely prestige paper. We extracted the top 10 extremely prestige papers and obtained that the main subjects of these papers were computer science and engineering. Meanwhile, in the future, we hope to construct the new index that can extract extremely prestige papers.

We calculated the mean values of the number of citations of journal j , K_j , and the Google number of journal j , G_j . We assigned the ranking of the number of citations of journal j , R_{Kj} and that of PageRank of journal j , R_{Gj} and defined the new measure of journal influence by R_j . We compared R_j with the IF, Eigenfactor score, and SCImago Journal Rank and obtained that Spearman's rank correlation coefficient is $\rho=0.630.72$.

Ranking	R_{Gj}	R_{Kj}	F_j	Journal
1	9	28	3.1	Advances in Physics
2	10	52	5.2	Annual Review of Fluid Mechanics
3	14	326	23.3	Computing Surveys
4	15	47	3.1	Advances in Organometallic Chemistry
5	22	108	4.9	Annual Review of Biophysics and Bioengineering
6	26	159	6.1	Advances in Parasitology
7	29	469	16.2	ACM Transactions on Computer Systems
8	30	226	7.5	Advances in Agronomy
9	33	364	11.0	IEEE Personal Communications
10	34	312	9.2	Advances in Cryptology - Eurocrypt 2000

Table 5. Top 10 extremely prestige journals

We extracted the top 10 journals with a high ranking R_i and observed that the main subjects of these journals were biology and chemistry. Meanwhile, this result strongly depends on the number of papers included in the journal, n_j . Similar to the investigation for papers, we extracted extremely prestige journals by considering the constraint $R_{K_j}/R_{G_j} > 3$ and observed that the main subject of these extremely prestige journals was computer science. We remark that the construction of the new index that can extract extremely prestige journals is left as a future study.

We used 35 years (from 1981 to 2015) of data of the SCIE to construct the new measure of journal influence. Therefore, the recent two or three years of data were used in developing the IF, Eigenfactor score, and SCImago Journal Rank. Thus, the new measure proposed by this study did not incorporate the effect of the trend. Hence, incorporation of the effect of the trend to the new measure is also left for future study.

Although we considered extremely prestige papers and journals, if we regard interdisciplinary as the most important factor, we would calculate the betweenness centrality and investigate the correlation with the number of citations. It is also significant to define indices by integrating the number of citations, the Google number, and the betweenness centrality.

Acknowledgments

This work is supported by Nihon University College of Science and Technology Grants-in Aid 2012 and 2016.

References

[1] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer Networks*, 38, 393-422.

[2] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., Cayirci, E. (2002). A survey on sensor networks. *IEEE Communications Magazine*, 40, 102-114.

[3] Alamouti, S. M. (1998). A simple transmit diversity technique for wireless communications. *IEEE Journal on selected areas in communications*, 16, 1451-1458.

[4] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-410.

[5] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389-3402.

[6] Becke, A. D. (1993). Density-functional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98, 5648-5652.

[7] Bianchi, G. (2000). Performance analysis of the IEEE 802.11 distributed coordination function. *IEEE Journal on*

Selected Areas in Communications, 18, 535-547.

[8] Bland, J. M., Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327, 307-310.

[9] Bollen, J., Rodriguez, M. A., Van de Sompel, H. (2006). Journal status. *Scientometrics*, 69, 669-687.

[10] Brin, S., Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30, 107-117.

[11] Chen, P., Xie, H., Maslov, S., Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, 1, 8-15.

[12] Chomczynski, P., Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*, 162, 156-159.

[13] Floyd, S., Jacobson, V. (1993). Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1, 397-413.

[14] Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge. Acquisition*. 5, 199-220.

[15] Gupta, P., Kumar, P. R. (2000). The capacity of wireless networks. *IEEE Transactions on information theory*, 46, 388-404.

[16] Haykin, S. (2005). Cognitive radio: Brain-empowered wireless communications. *IEEE journal on selected areas in communications*. 23, 201-220.

[17] Journal Citation Reports. From: <https://jcr.incites.thomsonreuters.com/JCRLandingPageAction.action>

[18] Kennedy, J., Eberhart, R. (1995). Particle swarm optimization. In: *Proceedings of the 1995 IEEE international conference on neural networks (Perth, Australia)*. 1942-1948.

[19] Lee, C., Yang, W., Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37, 785.

[20] Ma, N., Guan, J., Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing & Management*, 44, 800-810.

[21] Maslov, S., Redner, S. (2008). Promise and pitfalls of extending google's pagerank algorithm to citation networks. *Society for Neuroscience*, 28, 11103-11105.

[22] Otwinowski, Z., Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, 276, 307-326.

[23] Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, 11, 341-356.

[24] Perdue, J. P., Burke, K., Ernzerhof, M. (1996). Generalized gradient approximation made simple. *Physical Review Letters*, 77, 3865.

[25] Sheldrick, G. M. (2008). A short history of SHELX. *Acta crystallographica section A: Foundations of Crystallography*, 64, 112-122.

[26] Shimago Journal & Country Rank. From: <https://www.scimagojr.com/journalrank.php?year=2015>.

[27] Souma, W., Jibu, M. (2018). Progress of Studies of

Citations and PageRank. *In: Scientometrics* (p. 213-231). IntechOpen.

[28] Thompson, J. D., Higgins, D. G., Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673-4680.