Clustering of Scientific Activity of Faculty Staff Based on the Results of Publication Activity

O. A. Zyateva[1(✉)], E. A. Pitukhin[1], and M. P. Astafyeva[2]

[1] Petrozavodsk State University, Petrozavodsk, Russia
olga_zyateva@mail.ru, eugene@petrsu.ru
[2] Dubna State University, Protvino, Russia
marines2000@mail.ru
https://petrsu.ru

Abstract. The article is devoted to the study of the publication activity of the faculty staff, which is reflected in the world's leading scientometric systems. The report presents tools for comparative analysis of two main scientometric indicators – the Hirsch index in the RSCI system and the number of publications indexed in Scopus and Web of Science. Significant features that characterize the faculty staff and their publication activity were determined, on the basis of which the clustering of scientific activities was carried out. As a result of cluster analysis, the faculty staff was divided into eight meaningfully interpreted clusters with similar characteristics of employees and their publication activity. A structural model of teacher transitions between selected clusters is presented. The drivers of the University's scientific activity are only two clusters out of eight, in which the publication activity is most pronounced. Employees in the other clusters should also fall into these two groups. From a practical point of view, the presented results can be useful for management to assess the current state of scientific activity and its improvement. This approach is universal and can be carried out in any educational organization of higher education, regardless of its specifics. This will help universities to conduct self-analysis and manage the performance of departments, in particular, to improve the results of publication activity.

Keywords: Data Mining · Clustering · Statistical analysis · Scientific activity · Publication activity · Scientometric indicators · H-index · Scopus and Web of Science

1 Introduction

To date, there are a large number of indicators that allow to assess the effectiveness of the University's scientific activities to some extent [1, 2]. One of the main indicators that characterize scientific activity is the indicator of publication activity [3, 4], in particular, the number of publications of the organization indexed by international databases [5–7]. This is primarily due to the fact that the total indicator for the country depends on the publication activity of each University, which in turn gives an idea of the country's place in the scientific world [8, 9]. The most significant are publications in the scientometric databases Scopus and Web of Science (hereinafter – WofS).

R. Silhavy et al. (Eds.): CoMeSySo 2020, AISC 1295, pp. 771–778, 2020.
https://doi.org/10.1007/978-3-030-63319-6_71

Analysis of the dynamics of publications by Russian authors and determination of the place of Russian science at the world level are presented in [10, 11]. Issues related to the assessment of scientific activity of universities and their employees, as well as ways to improve it both in General [12, 13] and in certain areas [14, 15], remain relevant. Foreign scientists also pay great attention to the issues of evaluating the scientific potential of the University depending on various factors [16–20].

As part of the main work functions, publication activities are not mandatory for faculty staff. It is obvious that all employees work with different productivity in the field of scientific activity, and, as a result, have different indicators of publication activity in Scopus and Web of Science. Accordingly, the contribution of each employee to the overall rating of the University is different. Thus, the University administration faces an important problem – how to most effectively increase the number of publications indexed in Scopus and Web of Science.

In this regard, the task of analyzing the publication activity of the University's faculty staff and identifying features that can be grouped by teachers who have the maximum number of publications in the Scopus and Web of Science databases is relevant. This step is necessary for the University administration to make effective management decisions aimed at increasing the indicators of scientific activity.

2 Materials and Methods

As information base of research were collected and processed data on the main indicators characterizing the scientific activity of the University staff – the parameters of their publication activity and the distribution of the number of publications by year over the past 15 years, existing in the bibliographic database of the Russian science citation index (hereinafter RSCI) and the distribution of articles indexed in the databases Scopus and Web of Science over the same period as 14.04.2020. It is assumed that if the publication was simultaneously indexed in both databases, then in our view the publication was counted twice.

The study used Data Mining methods, the Highlight Exceptions tool, which uses the Microsoft clustering algorithm [21], as well as statistical and comparative analysis. The statistical analysis software used the Excel table processor and the SQL Analysis Services Excel Add-In, which provided clustering algorithms.

3 Analysis of Publication Activity Indicators

We will conduct a study of the indicators of publication activity of the faculty staff on the example of one of the leading universities in the North-West of Russia – Petrozavodsk state University (PetrSU).

The distribution of the total number of PetrSU employees by academic degrees, as well as the results of grouping employees of educational institutions by the growth of the Hirsch index in the RSCI (hereinafter – HI) and the average number of publications in Scopus and Web of Science over the past 2 years is shown in Table 1. The institutions are arranged in descending order of number of publications in the databases:

Scopus/WofS for 2019 from left to right, namely the Institute of mathematics and information technology (IMIT), Institute of physics and technology (FTI), the Medical Institute (MI), Institute of forest, mountain and building Sciences (IFMBS), Institute of biology, ecology and technology (IBEAT), Institute of Philology (IF), Institute of history, political and social Sciences (IHPSS), Institute of foreign languages (IFL), Institute of Economics and law (IEP), the Institute of physical culture, sport and tourism (IPCST) and the Institute of pedagogy and psychology (IPP).

Table 1. Distribution of educational institutions by the value of the H-index and the number of publications in Scopus and Web of Science.

Institute	IMIT	FTI	MI	IFMBS	IBEAT	IF	IHPSS	IFL	IEP	IPCST	IPP	PetrSU
Number of faculty staff:	63	62	107	81	58	66	53	55	46	52	64	707
Candidate of science	32	43	62	58	39	32	37	23	26	18	43	413
Doctor of science	9	7	29	17	15	10	9	2	10	1	4	113
HI growth (ΔH)	0,7	0,4	0,77	0,26	0,46	1,12	0,73	2,2	0,71	0,44	0,78	0,59
Growth in the number of Scopus/WofS publications (ΔSW)	0,53	0,43	0,43	0,79	0,31	2,75	1,2	0,16	0,38	0,75	1	0,50
ΔSW/ΔH	0,75	1,07	0,56	3,09	0,69	2,44	1,64	0,07	0,53	1,69	1,29	0,84
Number of Scopus/WofS publications in 2019	91	54	45	34	13	10	8	5	4	3	1	268
Percentage of doctor of science with Scopus/WoS publications	0,8	0,7	0,6	0,6	0,5	0,1	1	0,5	0,1	0	0	0,51
Percentage of candidates of science with Scopus/WofS publications	0,3	0,5	0,1	0,1	0,1	0,1	0,1	0,1	0	0,1	0,05	0,15
Percentage of non-degree faculty staff with Scopus/WofS publications	0,1	0,1	0,06	0	0	0,04	0,3	0,03	0	0	0	0,04

The largest division is the Medical Institute, which comprises 15.1% of the total number of employees. The largest number with a degree (93%) in the institutions IBEAT and IFMBS, least of all in the IFL (45%) and IPCST (36%). From the presented

data, we can see a significant increase in the average value of these H-index for all institutions. The largest growth is observed in IFL (320%) and IF (212%), the smallest – in IFMBS (126%) and FTI (140%). The situation with the publications in Scopus and Web of Science develops in a different way. Despite the fact that all institutions have increased the number of publications over the past two years, the leaders are MI, FTI and IMIT. Their contribution is 76% to the total number of PetrSU publications.

In 2019, Scopus and Web of Science publications have the highest number of publications (34%) from the total number of IMIT, and the lowest (0.4%) from IPP.

The main contribution to the publication activity in terms of degrees is made by doctors of science in technical fields (IMIT, FTI, IFMBS). The Humanities are less active, perhaps due to a relatively small number of magazines that correspond to their interests. A large variation is observed among doctors of science and employees without a degree. Low activity also occurs in candidates of science.

Analysis of growth rates showed that the number of institutions (IF, IHPSS, IPP, IPCST, IFMBS) has set for itself the primary goal to increase the number of publications in Scopus and WofS, FTI – working to increase both indicators. In other institutions, there is a stronger growth of publications in the RSCI, but despite this, it is impossible to conclude that they put this as a priority, given the significant initial data on the number of publications in comparison with other institutions.

4 The Clustering Results

As a result of the analysis of the publication activity of the faculty, a large differentiation was found, both in the context of educational institutions and in the context of academic degrees. In this regard, the problem arises of dividing the entire set of teachers into stable groups with similar characteristics of publication activity in publications indexed in the Scopus and Web of Science databases. As a result of the research, all teachers were divided into eight clusters. Each cluster is characterized by the average values of 7 attributes – age, H-index and length of service, as well as academic degree, academic title, number of publications and citations of articles in the RSCI.

As a result of clustering, the main variable features of clustering were two factors: the employee age factor and a publication activity factor such as the H-index.

At the same time, it is worth noting that another factor of SW publication activity – the average number of Scopus and Web of Science publications per 1 PPP-correlates well with the average value of the H-index for the obtained clusters (see Fig. 1).

Based on the selected clusters, a structural model of transitions between clusters is proposed, which is based on the age factor. It is assumed that employees from more "young" age clusters have a chance, with the appropriate desire and opportunities, to engage in science, and thus strengthen their publication activity, and, consequently, improve their scientometric indicators.

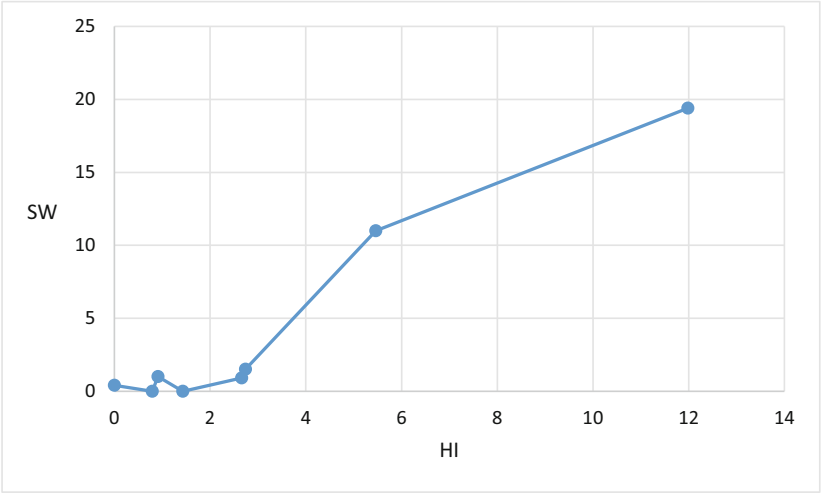

Fig. 1. Dependence of the average number of Scopus and Web of Science publications (SW) on the average value of the Hirsch index (HI)

A structural model of teacher transitions between selected clusters is shown in Fig. 2. Transitions between them are divided into three types: the traditional scenario – the employee does not change his attitude to publication activity (bold arrow), the optimistic scenario – the employee actively publishes and strengthens his position (thin arrow) and the negative scenario – the employee ceases to engage in scientific activities (dotted arrow).

5 Discussions

The first cluster consists of those employees who reduce the average value of the H-index in the whole University (mostly there is no academic degree, the H-index is zero, the average age is about 50 years).

The second cluster includes teachers who stopped actively publishing after defending their PhD thesis (mostly candidates of science with the academic title of associate Professor, the average value of the H-index is 1.43, the average value of age and experience – 50 and 20 years, respectively).

The third cluster includes inactive young scientists who most likely do not plan to get a degree or title in the near future (half are candidates of science, the average value of the H-index is 0.91, the average age is 35 years).

The fourth cluster is the mainstay of the University. Employees are actively engaged in publishing activities that are interesting to others (all have a degree, more than 80% have an academic title, the average value of the H-index is 5.46, the average age is 53 years).

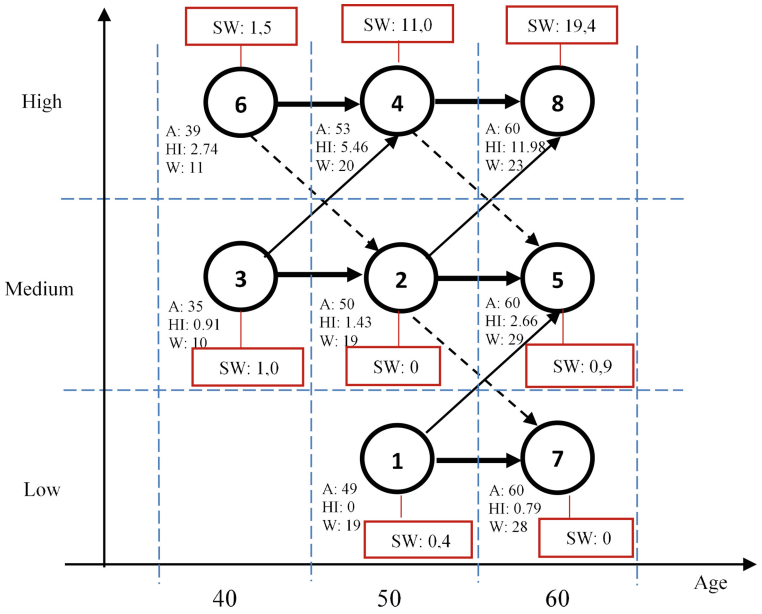

Fig. 2. Structural model of faculty staff transitions between selected clusters (A - average age, HI-average value of the H-index, W-length of service, SW-average number of Scopus and Web of Science publications per 1 faculty staff)

The fifth cluster includes scientists who, having both a degree and a scientific title, continue their publishing activities (the average age is 60 years, experience is more than 30 years, the average Hirsch index is 2.66).

The sixth cluster consists of young scientists who are actively engaged in scientific activities and may be preparing to apply for a scientific title (mainly candidates of science without a scientific title, average age 39 years and experience of 11 years, average H-index 2.74).

In the seventh cluster, there are employees who were active in publishing before receiving the academic title, after which they stopped doing it (the average age is 60 years, experience is more than 28 years, the average Hirsch index is 0.79).

The eighth cluster consists of honorary professors who, despite their age, are still engaged in scientific activities and publish their results (the average age is 61 years, the average value of the Hirsch index is 11.98, mostly doctors of science).

It is interesting to observe which clusters' employees are most actively published in publications indexed in the information and analytical systems of scientific citation Web of Science or Scopus. For this purpose, the average number of Scopus and Web of Science publications per 1 faculty staff in each cluster (SW) was calculated (see Fig. 2). Teachers of 4 and 8 clusters confirm their status – the present and future of the University. They own 79% of the total number of publications. A quarter of the employees from cluster 6 are actively engaged in their activities, thereby strengthening their positions and have the opportunity to become the mainstay of the University in the

future (transition to clusters 4 and 8). Teachers from clusters 1, 3, and 5 are inactive. This is also confirmed in publications indexed in the Web of Science or Scopus databases. The data shown in the table shows that there are no representatives of clusters 2 and 7. Perhaps this is due to the fact that after receiving a scientific title, they stopped actively engaged in science, as evidenced by a fairly large average age and a small average value of the Hirsch index, so it is not appropriate to talk about publications in highly rated journals.

6 Conclusion

As a result of the research, seven features were identified, according to which the faculty staff was divided into eight meaningfully interpreted clusters with similar characteristics of publication activity. A structural model of teacher transitions between selected clusters is presented, which allows building individual growth trajectories for those faculty staff who are trying to engage in scientific activities productively. This is especially true for teachers of the sixth cluster.

The proposed approach will help universities conduct self-analysis, adjust the indicators of scientific activity of institutions and employees, taking into account their characteristics. This will undoubtedly have a positive impact on the image of universities and will strengthen their position in the rankings.

References

1. Butvilo, A.I., Korzhov, S.T., Krivonozhenko, A.F., Syunyov, V.S.: Institutes integrated interdisci inary research as a tool of intensification of research activities of the modern University. Univ. Manag. Pract. Anal. **5**(99), 30–38 (2015)
2. Zyateva, O., Pitukhin, E., Peshkova, I.: University performance indicators impact on their ranking position. In: 8th International Conference on Education and New Learning Technologies, pp. 8751–8759. INTED, Spain (2016)
3. Gilyarevskii, R.S.: Publication activity as an indicator of scientific performance. Sci. Tech. Inf. Proc. **41**(3), 170–177 (2014)
4. Klochkov, Y.: Analysis of the publication activity of university researchers. In: Proceedings - 2019 Amity International Conference on Artificial Intelligence. AICAI, pp. 74–79 (2019)
5. Efimova, I.N.: Ratings as comparative assessment of universities for solving management tasks at national level. Manag. Issues **4**(6), 222–231 (2013)
6. Kincharova, A.V.: Methodology of international university rankings: analysis and criticism. Univ. Manag. Pract. Anal. **2**(90), 70–80 (2014)
7. Kim, I.N.: About the measures promoting successful formation of publishing career of the teacher of higher education institution. Eng. Educ. **18**, 64–78 (2015)
8. Smolentseva, A.: Globalization and the research mission of universities in Russia. In: Higher Education in the BRICS Countries, pp. 399–421. Tomsk (2015)
9. Rodionov, D.G., Rudskaia, I.A., Kushneva, O.A.: The importance of the university world rankings in the context of globalization. Life Sci. J. **11**(10), 442–446 (2014)
10. Kotsemir, M.N.: Publication activity of Russian researches in leading international scientific journals. Acta Naturae **4**(2), 15–34 (2012)

11. Moskovkin, V.M., Zdorovtsev, A.D., Munenge, S., Peresypkin, A.P.: Comparative analysis of the publication activity level of the leading Russian Universities conducted in reliance on Web of Science and Scopus databases. Glob. J. Pure Appl. Math. **11**(6), 5121–5133 (2015)
12. Alfred, R., Tanakinjal, G.H., Obit, J.H.: Too few too far: research productivity assessment in Malaysia based on the H-index analysis. In: 6th International Technology, Education and Development Conference, pp. 1925–1934. Spain (2012)
13. Zyateva, O., Pitukhin, E., Peshkova, I.: Research on university key performance indicators: comparative analysis. In: 7th International Conference on Education and New Learning Technologies, pp. 4010–4020. INTED, Spain (2015)
14. Filonov, M.R.: Results of scientific research activity of the national university of science and technology misis (Nust misis) for 2017. Izvestiya Vysshikh Uchebnykh Zavedenij. Chernaya Metallurgiya. **61**, 503–509 (2018)
15. Evdokimov, V.I., Goryachkina, T.G., Leonov, B.I.: Analysis of publication activity in the field of biomedical engineering in 2003–2012. Biomed. Eng. **48**, 1–8 (2014)
16. Bonaccorsi, A., Ancaiani, A., Secondi, L., Setteducati, E.: Exploring the role of third-party research in Italian universities. In: 4th International Conference of Education, Research and Innovation, pp. 1566–1576. Spain (2011)
17. Rodionov, D.G., Fersman, N.G., Kushneva, O.A.: Credibility of the Russian higher education in the world: problems and solutions. Life Sci. J. **11**(11s), 43–47 (2014)
18. Karfaa, M., Haq, F.: The challenges of strategic management in higher education: the case of a Malaysian private university. In: 7th International Technology, Education and Development Conference, pp. 1779–1782. Spain (2013)
19. Berezkina, N.Y., Sikorskaya, O.N., Khrenova, G.S.: Analyzing the publication activities of scientists of the national academy of sciences of Belarus. Sci. Tech. Inf. Process. **39**, 164–168 (2012)
20. Krasovska, O., Andrushchenko, V., Velichko, I.: Scientific cooperation in basic research and higher education. Cult. Psychol. Educ. **7**, 53–58 (2018)
21. LNCS Homepage. https://docs.microsoft.com/ru-ru/sql/analysis-services/highlight-exceptions-table-analysis-tools-for-excel?view=sql-server-2014. Accessed 21 Mar 2020