DOI: 10.32347/2412-9933.2023.54.90-94
UDC 338.33

Huilin Xu
PhD student in the Department of Information Systems and Technologies,
https://orcid.org/0000-0002-1794-3270
Taras Shevchenko National University of Kyiv, Kyiv

MODEL FOR ASSESSING THE INFLUENCE OF SCIENTISTS BASED ON THE GLOBAL CITATION NETWORK AND THE HISTORY OF SCIENTIFIC RESULTS

Abstract. *The research considers the actual task of developing a model for evaluating the influence of scientists, taking into account the dynamics of changes in publishing activity. The paper formalizes the concept of the scientist's information environment, particularly the scientist's influence and the subject space. It shows the connection of the information space with choosing scientific partners. The concepts of the dynamic information space of a scientist and the subject scientific environment of a scientist are also introduced. It is established that they are accumulative in nature. As a result, a model for evaluating the influence of scientists based on the global citation network was developed, considering the age of scientific results. The introduction of the aging factor of the publication allows for increasing the sensitivity of the method for evaluating the influence of scientists. The developed model makes it possible to improve the cooperation of scientific teams, which is the basis of the analysis of their productivity.*

Keywords: method of evaluation of scientific activity; scientific project; the task of selecting a scientific partner

Introduction

Significant changes have occurred in the geography of science and scientific cooperation in recent decades. Scientific networks or networks of scientific cooperation, which previously functioned locally within universities and scientific institutes and spread over one or several countries and were more static, now cover almost the entire world. The speed of creating new information and using tools for creating network cooperation among scientists contribute to the dynamic development and diversification of scientific networks.

The development of a scientist in a specific scientific direction does not occur in isolation from the development of other scientists in this direction. Scientists publish scientific publications, cite each other, and are co-authors of some studies; they collaborate scientifically. At a specific moment of scientific activity, a scientist develops in the university educational and scientific space within the scientific subject space. At the same time, its information environment is formed. A scientist is a specialist in some field or several fields of science, who uses scientific methods in his own or collective research, develops in his information environment, and, together with his part, belongs to the relevant subject scientific space and the scientific space of a higher education institution or a scientific institution with which he is affiliated.

As shown in the study [1], to determine the advantages of subjects of scientific activity, which are included in the relevant subject scientific spaces, it is necessary to evaluate their productivity. In addition, one needs to predict future performance changes based on retrospective data for a given subject.

There are numerous models for selecting research partners. Partnership relations and principles of their management were studied in the work [2]. A multi-stage mathematical model of partner selection is described in [3]. In [4], it is proposed to use the analytical hierarchy method to select the best partners from a small set of potential partners.

Common to these studies is that the quantitative characteristics of publications influence the choice of scientific projects. Such classic indicators as Hirsch are still popular indices [5]. The probability of cooperation is higher among those with an average H-index. However, the emergence of the PageRank method [6] offered new opportunities for evaluating the scientific productivity and reputation of collective subjects. The application of the PageRank method for calculating the performance assessment of a collective of subjects of scientific activity is described in detail in [7].

The study [8] shows that an essential tool for evaluating the dynamics of scientific productivity is considering the intensity of citations. Known scientometric methods need to be sufficiently developed in evaluating the influence of scientists according to the global citation network, taking into account the age of scientific results. Therefore, developing a model for evaluating the influence of scientists, which would consider the dynamics of changes in publishing activity, is an urgent task.

The purpose of the article

Therefore, this study aims to develop a model for evaluating the influence of scientists based on the global citation network, taking into account the age of scientific results.

To achieve the goal of the research, it is necessary to solve the following problems:

1. Formalize the basic concepts of the scientist's information environment;
2. To develop a dynamic model for evaluating the influence of scientists based on the analysis of the change in the citation network of his publications.

Presenting main materials

The information environment of a scientist includes identifiers that determine his scientific activity at specific points in time, that is, scientific publications, their citations, participation in scientific and educational projects, etc. The information environment of a scientist accumulates information about his entire scientific activity.

The university scientific space consists of scientists affiliated with the relevant university. The results of the scientist's scientific activity influence the development of the university's scientific space and the university's overall rating.

The influence of a scientist is a quantitative indicator that determines the assessment of a scientist's scientific activity and productivity during a certain period of time. This quantitative indicator determines the impact of the scientific results of a scientist in the relevant subject scientific space in comparison with other scientists who belong to this space.

The project space includes influential scientists in the relevant scientific subject spaces and university scientific spaces formed according to the structure, topic, and goals of scientific or educational projects. There is no mention of possible subcontracting contracts that can be included in the project application for the performance of additional tasks by third-party organizations or employees in this task. That is, we will consider only potential project executors who work in institutions of higher education or scientific institutions and are offered to perform scientific tasks within the limits of the corresponding project.

The information space can be expanded by including other components of the scientist's scientific activity in the tuple. That is, the characteristic of extensibility is fulfilled for it. But for choosing a potential partner-executor of scientific projects, only these components are sufficient. However, to solve the problem of choosing partners for scientific cooperation and organizing scientific projects, it is essential to study the changes in the subject scientific environment of a scientist over time, as well as to understand how the intensity of citations of a scientist's scientific publications (i.e., his influence) changes over time. This is necessary so that when deciding to include a scientist in the project, there is a good picture of his scientific activity during a certain period in a specific scientific direction.

Let

$$T = (T_1, T_2, \dots, T_{N-1}, T_N).$$

Are points in time in which the scientific results of scientists are recorded; for example, the end of each calendar year T_N is the last point in time in which the results of scientific activity are recorded.

The information space of each scientist v_j should $j=\overline{1,t}$ depend on time. Accordingly, it can be transformed into the form of a tuple, the components of which are discrete time series that reflects the change in indicators of the scientist's scientific activity over time, i.e.

$$I\left(v_j, T^k\right) = \left\langle \left\{p_w^{j,k}\right\}_{w=1}^{r_j}, \left\{\widehat{C}_w^{j,k}\right\}_{w=1}^{r_j}, \left\{\breve{C}_w^{j,k}\right\}_{w=1}^{r_j}, O^{j,k} \right\rangle,$$

where $p_w^{j,k}$ are scientific publications published by the scientist v_j and recorded at the time T_k, r_j – the number of all scientific publications published by the scientist v_j at the time T_k, $k=\overline{1,N}$,

$\left\{\widehat{C}_w^{j,k}\right\}_{w=1}^{r_j}$ is the sequence of sets of publications in which publications p_w^j of a scientist are cited v_j, $w=\overline{1,r_j}$ at the moment of time T_k,

$\left\{\breve{C}_w^{j,k}\right\}_{w=1}^{r_j}$ is the sequence of sets of publications that cite the publications p_w^j of the scientist v_j, $w=\overline{1,r_j}$, at the moment of time T_k, $k=\overline{1,N}$,

$O^{j,k}$ is the scientific environment of the scientist v_j at the moment of time T_k, $k=\overline{1,N}$.

It should be noted that the scientist's information space and the subject scientific environment of the scientist have an accumulative nature. Suppose the scientist had no scientific activity before a specific period Tg. In that case, the set of his publications and citations will be empty, and, accordingly, the set of scientific subject spaces will be empty. With the first publications' appearance, these sets' accumulation will occur. If a scientist at some stage of his activity changes his scientific direction, then from this period, another direction will be included in his subject scientific environment, by which the articles will be cited. Also, the influence of a scientist will change over time. Moreover, this influence will be related to a particular state of the subject scientific environment of this scientist.

The sets $\widehat{C}_w^{j,k}$ and $\breve{C}_w^{j,k}$ for the publications p_w^j of the scientist v_j, $w=\overline{1,r_j}$ at the moment of time T_k, $k=\overline{1,N}$ form a network of citations of scientific publications. If we consider all published scientists, the network will be global; otherwise - local. All citations of one publication to another can be specified as a Markov matrix defined as

$$C=\left\{c_{xy}\right\}_{x,y=1}^{R},$$

$R=\operatorname{card}\left(\sum_{j=1}^{t} r_j\right)$ is the number of publications of scientists belonging to the citation network of scientific publications $c_{xy}\in[0,1]$ - the probability of transition from one state to another, determined by the number of citations of one publication in others. The matrix $C\geq 0$, and $\sum_{x=1}^{R} c_{xy}=1$, $y=\overline{1,R}$.

To find the influence of a scientist at the moment of time T_N, we will use the PageRank method. However, to consider changes in influence over time, the age of the citation must be included in the calculation, which can affect the final score.

Let the coefficient corresponding to the weight of p_x^j the scientist's publication V_j, based on which the rank of the publication is determined at the qth step, be denoted by b_x^q. At step (q=0), the coefficients for all scientific publications are equal and are defined as $b_x^q=\frac{1}{R}$, $y=\overline{1,R}$. All other coefficients will be calculated iteratively according to the formula:

$$b_x^{q+1}=\alpha C b_x^q+\frac{1-\alpha}{R}\mathbf{E},$$

where **E** is a unit matrix, α is an attenuation coefficient that determines the transition probability from one state (current publication) to another (another publication).

It is possible to calculate coefficients b_x^q a software approach can be applied. At the same time, we will iteratively obtain that after performing a significant number of iterations, the approximate values of the coefficients will be calculated. b_x^q To perform an iterative calculation of the coefficients b_x^q, you can use the formula for $\forall\varepsilon>0$

$$b_x^{q+1}=\sum_{y=1}^{R} b_y^q c_{yx}+\sum_{y=1}^{x-1} b_y^{q+1} c_{yx},$$

if $\left|b_x^{q+1}-b_x^q\right|<\varepsilon$, then the calculation of the coefficients b_x^q is stopped.

After reaching the condition, $\left|b_x^{q+1}-b_x^q\right|<\varepsilon$ we will get a vector of coefficients

$$\left(b_1,b_2,\ldots,b_R\right),$$

where b_z is the value of the coefficient for publication p_z^j, $b_z\in[0,1]$, $z=\overline{1,R}$, $\sum_{z=1}^{R} b_z=1$.

The scientific publication p_z^j that corresponds to the most outstanding value b_z receives the first rank. Next, the ranks of other publications are formed in order of decreasing coefficients b_z, $z=\overline{1,R}$.

Using the representation of the ranks of scientific publications, it is possible to determine the ranks or influence of scientists who are the authors of these publications. Let the subject scientific environment of some scientists V_h include publications

$$P^h=\left(p_1^h,p_2^h,\ldots,p_{r_h}^h\right),$$

with ranks

$$G\left(P^h\right)=\left(g\left(p_1^h\right),g\left(p_2^h\right),\ldots,g\left(p_{r_h}^h\right)\right),$$

where $g\left(p_w^h\right)\in\mathbb{N}$ is the publication rank p_w^h of the scientist v_h, $w=\overline{1,r_h}$.

Then the influence of the scientist v_h will be determined by the formula:

$$\gamma\left(v_h\right)=\frac{1}{r_h}\cdot\sum_{w=1}^{r_h} g\left(p_w^h\right),$$

where $\gamma\left(v_h\right)$ is the influence of the scientist v_h.

We will make changes to the PageRank method to take into account the age of scientific publications and their belonging to the relevant subject scientific space. For this purpose, at the beginning of the iterative method, it is necessary to determine the initial coefficients of scientific publications that correlate with their age. If the publication is fresh, it receives more weight, if it is old - less.

Publications $p_z^{j,N}$ published at the time T_N are not considered, because they have yet to be cited by other publications, so their impact will be low. We will consider publications that were published in periods

$$p_z^{j,N-1},\ p_z^{j,N-2},\ \ldots,\ p_z^{j,N-\pi}$$

All posts that are published in periods

$$T_{N-\pi-1},\ T_{N-\pi-2},\ldots$$

will have a zero coefficient;

π – the period for calculating the weighted coefficient corresponding to the weight of the scientific publication.

If $\tilde{\pi}=\sum_{i=1}^{\pi} i$, then the coefficients of scientific publications that were published in the period T $_{N-1}$ at step q=0 will be matched by the coefficient $\overline{b}_x^q=\frac{\pi}{\tilde{\pi}}b_x^q$. For publications published in the period T N $_{-2,}$ we will match the coefficient $\overline{b}_x^q=\frac{\pi-1}{\tilde{\pi}}b_x^q$, etc. For a scientific publication that was published in the period $T_{N-\pi}$ let's match the coefficient $\overline{b}_x^q=\frac{1}{\tilde{\pi}}b_x^q$.

Accordingly, other coefficients will be calculated iteratively according to the formula:

$$\overline{b}_x^{q+1}=\alpha C\overline{b}_x^q+\frac{1-\alpha}{R}\mathbf{E},$$

where **E** is the unit matrix, α is the extinction coefficient, the stop condition $\left|\overline{b}_x^{q+1}-\overline{b}_x^q\right|<\varepsilon$.

The introduction of the aging factor of the publication allows to increase the sensitivity of the PageRank method for evaluating the influence of scientists. As a result, by analogy with the previous method, we will receive the rank of a scientist. We will determine the ranks of scientific publications of a scientist v_h using the weighted PageRank method as follows

$$\overline{G}\left(P^h\right)=\left(\overline{g}\left(p_1^h\right),\overline{g}\left(p_2^h\right),\ldots,\overline{g}\left(p_{r_h}^h\right)\right),$$

where $\overline{g}\left(p_w^h\right)\in\mathbb{N}$ is the publication rank p_w^h of the scientist v_h , $w=\overline{1,r_h}$.

Then the influence of the scientist v_h will be determined by the formula:

$$\overline{\gamma}\left(v_h\right)=\frac{1}{r_h}\cdot\sum_{w=1}^{r_h}\overline{g}\left(p_w^h\right),$$

where $\overline{\gamma}\left(v_h\right)$ is the influence of the scientist v_h obtained by the weighted PageRank method.

Given that for solving the task of selecting a scientist for the executors or partners of a scientific project, it is important not only the productivity or influence of a potential partner-executive in general, but also his productivity within the limits of a specific subject scientific space.

Let's select for each scientist set that consist of articles belonging to each scientific subject space. The subject scientific environment of a scientist includes those spaces to which his scientific publications belong and it is part of the information space of the scientist at the moment of time T $_k$:

$$O^{j,k}=\left\{o_i\left|\Lambda\left(p_w^{j,k},o_i\right)=1,w=\overline{1,r_j},j=\overline{1,t}\right.\right\}.$$

We will form lists of scientific publications for each of the subject scientific areas to which his publications belong:

$$P_i^{j,k}=\left\{p_w^{j,k}\left|\Lambda\left(p_w^{j,k},o_i\right)=1,w=\overline{1,r_j},j=\overline{1,t}\right.\right\},$$

$$i=\overline{1,m},$$

where $P_i^{j,k}$ is the set of scientific publications of scientists v_j at the moment of time T $_k$, belonging to the subject scientific space o_i, $\operatorname{card}\left(P_i^{j,k}\right)\le r_j$.

Then the influence of a scientist v_h in a specific scientific space o_f will be determined by the formulas:

$$\gamma_f\left(v_h\right)=\frac{\sum_{p_w^{h,k}\in P_f^{h,k}} g\left(p_w^{h,k}\right)}{\operatorname{card}\left(P_f^{h,k}\right)},$$

where $\gamma_f\left(v_h\right)$ is the influence of the scientist v_h obtained by the PageRank method in the subject scientific space o_f ,

$$\overline{\gamma}_f\left(v_h\right)=\frac{\sum_{p_w^{h,k}\in P_f^{h,k}} \overline{g}\left(p_w^{h,k}\right)}{\operatorname{card}\left(P_f^{h,k}\right)},$$

where $\overline{\gamma}_f\left(v_h\right)$ is the influence of the scientist v_h obtained by the weighted PageRank method in the subject scientific space o_f .

Conclusions and prospects of further research

The paper formalizes the concept of the scientist's information environment, in particular the scientist's influence, the subject space, and shows the connection of the information space with the task of choosing scientific partners.

As a result of the development of a model for evaluating the influence of scientists according to the global citation network, taking into account the age of scientific results, it is possible to improve the cooperation of scientific teams, which is the basis of the analysis of their productivity.

Further research is possible in the part of integrating the developed model into the general methodology of selecting scientific partners.

References

1. Xu, H., Kuchansky, A. & Gladka, M. (2021). Devising an individually oriented method for selection of scientific activity subjects for implementing scientific projects based on scientometric analysis. *Eastern-European Journal of Enterprise Technologies*, 6 (3 (114)), 93–100. https://doi.org/10.15587/1729-4061.2021.248040.
2. Zhang, S. & Poulin, D. (1996). Partnership Management Within the Virtual Enterprise in a Network. International Conference on Engineering Management and Control, pp. 645–650.
3. Talluri, S. & Baker, R. C. (1996). A quantitative framework for designing efficient business process alliances. International Conference on Engineering Management and Control, pp. 656–661.
4. Chu, X. N., Tso, S. K., Zhang, W. J. & Li, Q. (2000). Partners Selection for Virtual Enterprises. Proceedings of the 3rd World Congress on Intelligent Control and Automation, pp. 164–168.
5. Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, 102(46), 16569-16572. https://www.pnas.org/doi/pdf/10.1073/pnas.0507655102.
6. Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30 (1-7), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X.
7. Lages, J., Patt, A. & Shepelyansky, D. (2016). Wikipedia ranking of world universities. *The European Physical Journal*, 89, 69. https://doi.org/10.1140/epjb/e2016-60922-0.
8. Kuchansky, A., Biloshchytskyi A., Andrashko Y., Biloshchytska S. & Faizullin, A. (2022). The Scientific Productivity of Collective Subjects Based on the Time-Weighted PageRank Method with Citation Intensity. *Publications,* 10(4), 40. https://doi.org/10.3390/publications10040040.

Received 28.05.2023

Хуейлінь Сюй
Аспірант кафедри інформаційних систем та технологій, *https://orcid.org/0000-0002-1794-3270*
Київський національний університет імені Тараса Шевченка, Київ

МОДЕЛЬ ОЦІНЮВАННЯ ВПЛИВОВОСТІ НАУКОВЦІВ ЗА ГЛОБАЛЬНОЮ МЕРЕЖЕЮ ЦИТУВАННЯ З УРАХУВАННЯМ ВІКУ НАУКОВИХ РЕЗУЛЬТАТІВ

Анотація. *У дослідженні розглядається актуальна задача розроблення моделі оцінювання впливовості науковців з врахуванням динаміки зміни публікаційної активності. У роботі формалізовано поняття інформаційного середовища науковця, зокрема впливовість науковця, предметний простір та показано зв'язок інформаційного простору із задачею вибору наукових партнерів. Також введено поняття динамічного інформаційного простору науковця і предметного наукового середовище науковця. Встановлено, що вони мають накопичувальний характер. У результаті розроблено модель оцінювання впливовості науковців за глобальною мережею цитування з урахуванням віку наукових результатів. Введення коефіцієнта старіння публікації дає змогу підвищити чутливість методу для оцінювання впливовості науковців. Розроблена модель дає змогу покращити співпрацю наукових колективів, що є основою аналізу їх продуктивності.*

Ключові слова: метод оцінювання наукової діяльності; науковий проєкт; задача вибору наукового партнера

Link to the publication

APA Xu, H. (2023). Model for assessing the influence of scientists based on the global citation network and the history of scientific results. *Management of Development of Complex Systems*, 54, 90–94, dx.doi.org\10.32347/2412-9933.2023.54.90-94.

ДСТУ Хуейлінь С. Модель оцінювання впливовості науковців за глобальною мережею цитування з урахуванням віку наукових результатів. *Управління розвитком складних систем.* Київ, 2023. № 54. С. 120 – 126, dx.doi.org\10.32347/2412-9933.2023.54.90-94.