

PAPER • OPEN ACCESS

Foresight of promising scientific fields in scientific social networks

To cite this article: Y A Orlova *et al* 2021 *J. Phys.: Conf. Ser.* **2060** 012007

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Foresight of promising scientific fields in scientific social networks

Y A Orlova¹, S V Pronichkin^{2,3}, I B Mamai⁴ and A V Kholstov⁵

¹Volgograd State Technical University, 28, Lenina Avenue, Volgograd, 400005, Russia

²Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, 40, Vavilov Street, Moscow, 119333, Russia

³National University of Science and Technology "MISiS", 4, Leninsky Prospect, Moscow, 119049, Russia

⁴Federal Center of Theoretical and Applied Sociology of Russian Academy of Sciences, 24/35, Krzhizhanovsky Street, Moscow, 117218, Russia

⁵N.N. Semenov Federal Research Center for Chemical Physics Russian Academy of Sciences, 4, Kosygina Street, Moscow, 119991, Russia

E-mail: pronichkin@mail.ru

Abstract. Currently, the problem of developing methods for determining promising directions, trends in scientific research and technology is the most important component of the process of forming a national science policy. The problem of finding priority or breakthrough areas of development is reduced to sorting out in one form or another and assessing possible existing areas of scientific activity. It is proposed to use a semantic network of communicative information space, in which the structuring of the subject area is based on the selection of objects and identifying features, which include not only features-properties, but also features-relationships. A module for traversing elements of a scientific social network by reference graph has been implemented, as well as an algorithm for determining the level of importance of a scientific direction according to multiple criteria.

1. Introduction

The most important feature of the existing system of forecasting and determining scientific priorities with all the variety of common methods is the inherent contradiction of conservatism and conceptual competition. Meanwhile, the latter should have generated scientific revolutions, which actually provide a scientific breakthrough in one direction or another. However, as a rule, works related to forecasting the development of fundamental science and analyzing the prospects for the practical relevance of its achievements are based on considering a kind of assessment of opportunities. This approach is predominantly retrospective and relies on a statistical or expert analysis of the existing gaps within a defined set of thematic areas, realization conditions, constraints and possible implementation timeframes.

The development of globalization and the emergence of new information and communication opportunities have led to a sharp increase in the volume of scientific and technological knowledge and a reduction in the terms of its commercialization. This, in turn, is accompanied by the evolution of procedures for selecting and forecasting priority research areas. Despite the national peculiarities of



the manifestation, the system of forecasting and identifying priority areas and trends in the development of knowledge and technology has formed as a methodological problem.

2. Methods for determining promising areas of scientific research

At present, the problem of developing methods for determining promising directions, trends in scientific research and technology has become an important component of the national scientific policy forming process of various countries [1]. No exception to this are countries that use the catch-up model in the economic development strategies shaping [2]. In formulating national policies in developing countries, the recently developed methodology of economic cycle analysis is of fundamental importance as a conceptual basis for building mathematical models linking economic growth to scientific and technological progress, to the dynamics of innovation [3].

The need to take into account the emerging numerous horizontal and vertical ties that constitute the institutional basis of the innovation environment [4] contributes to the transformation of the modern institutional structure of the emergence of innovative solutions into a complexly organized network system. In this case, innovation actors form what are known as collaborative networks, which represent a new, networked form of organization and creation of innovation. Linear models are replaced by models of network innovations, which form a new institutional environment for the creation and development of innovations, ways of innovative development of the economy.

The emergence of a new network, collaborative model of innovation is due not only to the peculiarities of the interaction of institutional actors. The territorial diversification of the technological process as the basis of modern industrial production makes a much greater contribution to the formation of network, collaborative relations. The essence of the modern technological process is the model of converting assembly components received from geographically dispersed manufacturers into a complete product. And consequently, the task of solving the rational organization of the assembly process arises from the early stages of the technological process of a product creating ready for the consumer. This task gets its solution as the result of purposeful technological conjugation of material, energy and information flows [5]. The introduction and widespread use of assembly processes automation exacerbates the need to develop network models for ensuring the control of technological production processes. At the same time, the open nature of the production processes organization also changes the very nature of the problems that accompany them.

Of considerable interest as an effective procedure for identifying promising scientific areas and interaction of various social actors is the Foresight methodology used as an algorithm for reconciling interests in determining the directions of national research and innovation systems development [6].

The Foresight method ensures the creation of communication platforms for decision-makers, for example, representatives of ministries and departments, business organizations, scientific organizations and expert communities. Government-funded research and development programs are one of the key innovation policy instruments at the national and international levels [7]. The use of Foresight methods is not accompanied by a reflection on the methodological foundations of the tools used. However, the analysis of the algorithm for carrying out the stages of the Foresight process and the structure of the foresight panels management [8] allows us to conclude that both the organizers and the participants in the examinations, at all levels of the priorities choice for scientific and innovative problems and topics, consider science as a complex communication system. Analysis of the methodology of these studies allows us to assert that the organizers and participants of foresight examinations are guided in their work by the idea of science as an interacting community of experts, researchers and politicians.

3. Features of models for determining scientific priorities

Solving the tasks of rationalizing the process of determining priorities in science and technology, including ensuring the work of experts from development institutions, is largely associated with the development and use of formalized procedures and statistical methods for researching scientific publications, patents and collections of scientific and technological documents [9].

As one of such approaches, the methods of automated intellectual analysis of document collections are used. According to the followers and developers of tools used in this direction, the use of

information technologies will allow collecting and systematizing information from voluminous, distributed and heterogeneous sources, visualizing and analyzing this information [10].

The methods of automated data mining, based on the search for statistical characteristics of documents, are based on scientific concepts of science as a body of scientific publications [11]. And as a unit of analysis, an article in a scientific journal can be accepted, which is analyzed using linguistic markers in order to identify the procedure for introducing a new subject into scientific circulation.

Considering the article as a communicative act organized according to certain rules [12], in which a new subject of knowledge for the scientific community is proposed to be considered in an original structure created by the author, containing not just the author's definition, but also a description of the method of constructing an object and correlating it with the corresponding scientific concepts, the authors of the proposed intellectual analysis of scientific articles on the basis of the statistical characteristics of syntactic-semantic models automatically select the final list of author's priorities.

The model of a scientific article formed in this way, according to the supporters of this approach, makes it possible to trace the emergence of a new term in the totality of analyzed scientific publications. This, in turn, provides for the formation of a procedure for determining the authority of a scientist / author of a publication whose publication activity has resulted in the introduction of a new term into scientific circulation.

Determining the influence degree of the author of the publication on the statistical characteristics of its citation, the developers of the methodology consider it possible to evaluate and fix the meaning that the term introduced by the author had on the intellectual activity of colleagues and changes in the conceptual foundations of not only specific studies, but also related areas of knowledge. Based on this, the method makes it possible to assess the importance of the author's institutional influence, which manifests itself in determining the prospects of scientific directions, and, accordingly, can be used in the procedures for determining scientific and technical priorities.

In addition to methods of automated intelligent analysis of unstructured data large volumes of various nature, the analysis of the scientometric and bibliometric indicators application, as well as work on procedures for the automatic evaluation of scientific results based essentially on the same methodological principles, are widespread [13]. At the same time, according to foresight experts, the list of priority selection criteria used should be supplemented with methods of analysis of scientific supply and demand.

4. Network methodology for setting scientific priorities

The development of effective methods for forecasting and determining the priorities of the science development presupposes a preliminary reflection / determination of the methodological foundations of this procedure. However, as a rule, this procedure is neglected. The point is that a seemingly natural image of science, reduced to research activity, prevails in the minds of researchers. And in connection with the naturalness of the image of science widespread in the public consciousness, there is no need for reflection on the procedure for identifying and predicting priority scientific directions. As a result, the problem of "searching" for priority or breakthrough areas of development is reduced to sorting out in one form or another and, accordingly, assessing possible existing areas of scientific activity. But, obviously, this approach is of a retrospective nature and is based on an assessment of the available scientific and technological groundwork within a certain set of thematic areas, the realization conditions, limitations and possible timing of implementation.

This feature of the scientific knowledge consideration is supported by the seemingly science studies, which usually proceeds from the postulate of the science integrity, which appears to the researcher as a certain intuitive datum. As follows from the analysis of numerous scientific literatures, most of the models used in these studies are based on the understanding of science as a research activity, which results are in the development of new knowledge.

As the result of this approach, the activities of researchers are understood as activities over objects, and it is assumed that the formation of any knowledge system begins with the allocation of objects (elements), followed by an indication of the relations connecting the elements with each other. This implies that they perform the function of elements, i.e. primary units of division of a conscious reality that is formed within the framework of a given system of knowledge, whether it is a reality accessible

to sensory perception, or consisting of ideal objects of theoretical knowledge, can only be atomic entities, each of which can be fixed, abstracting from its relationships by other entities.

According to this assumption, the primary cognitive act that a person performs in order to become aware of a certain subject area is an act in which his attention is focused on one, separately taken subject. The result of such an act is the allocation of a separate entity, isolated from other entities due to the presence of an isolating border in it, separating what is inside the border from everything that is outside.

Accordingly, the subject of research in this approach is, indeed, the direct act of incrementing new knowledge (scientific discovery). And the development of science is viewed as a time sequence of meaningfully related cognitive acts, which should embody the historical movement of the cutting edge of science.

As the result, the perceived reality is perceived as consisting of non-relative entity-objects that can be united with each other only externally, through the introduction of some additional structure that performs integrative functions. As a possible model representation of such a structure, the semantic network of the communicative information space [14] can be considered, in which the structuring of the subject area is based on the allocation of objects and identifying features, which include not only features-properties, but also features-relationships.

The cognition-structuring concept introduction allows science to be presented not as a collection of new knowledge increment acts, or a corpus of scientific publications, which are considered as an empirically observable result of incremental knowledge. Science in the context of the proposed concept is considered as a complex process of unfolding the procedure for structuring the subject space, which appears to the researcher as a space as a result of social / cognitive communication of participants in scientific cognition. This space is considered as a construct that sets the conceptual framework of the studied process of scientific cognition. The proposed model of cognitive activity is based on the procedure for structuring and representing reality in the form of a network reality.

This model of science, scientific knowledge is based on the idea of the subject space as a system of communications or interactions of the participants in these communications. And, accordingly, the elements of the science space are not forms of incremental knowledge, and not visualized forms of these knowledge increments, documents, but direct acts of information interaction, characterized by calculated communicative activity.

Thus, it is proposed to consider science as a complex communication system based on a variety of formal and informal communications. Its transformation demonstrates the progressive development from a variety of individual models of mental activity of the classical scientific era to modern complex structures and communities that unite scientists, separated by territorial and national space, institutional subordination, but connected by common tasks, research objectives and goals into a single communication network.

The change in the communication interaction paradigm in society as a result of wide access to the World Wide Web has supplemented the traditional communications of the scientific community with various forms of electronic interaction. This, in turn, led to the emergence in the virtual space of an empirically developed distributed communication infrastructure that visualizes the interaction and state of the global scientific community, as well as the relationships existing in it.

An indicator of the prospects of a scientific direction is the high interest of society in information related to a particular scientific direction and presented in a distributed communication system that visualizes the networked space of knowledge. And, accordingly, one of the ways to assess public interest, and, accordingly, the prospects of a scientific direction, can be an analysis of the statistics of search engine queries for keywords related to the chosen direction.

The analysis of the relationship between public interest and requests in the network is realized using the Socionet web application, which considers the virtual scientific community as a network interaction of many of its users, which is formed under the influence of changes in the real social space. With this application, it is possible to calculate the relative frequency of certain search terms to the total volume of search queries, as well as compare the volume of search queries for two or more search phrases.

Socionet can be viewed as a model of a network (virtual) knowledge space, each change of which as a result of the creation of articles or their editing, on the one hand, visualizes the state and interaction of the global expert community. On the other hand, it visualizes the transformation of network knowledge resulting from the structuring of news knowledge into reality, which can become the subject of analysis and research.

To solve the problem of determining scientific priorities as an actual metric of user behavioral analysis, a thematic analysis of the array of requests and editorial edits of Socionet articles was considered, which can serve as a method for determining both the interest of the expert community and the presence of public resonance on events occurring in a specific field of science.

To assess the interest of Socionet participants in various scientific disciplines and areas (topics), in the course of the study, a methodology was developed for collecting and assessing the thematic interest of participants based on an analysis of statistics of requests and edits of articles.

The TOPSIS method was used to highlight the most popular subject categories in the array of Socionet articles [15]. The method is intended for ordering multi-feature variants, which were assessed by several experts according to many criteria, based on the proximity to the reference point. Two criteria were used: K_1 is the requests number; K_2 is the edits number.

For each thematic category A_i scores were calculated for each criterion K_q . Each variant A_i is assigned a vector of assessments $y_i = (y_{i1}, \dots, y_{i2})$, where n is number of criteria. The combination of all variant assessments forms a matrix $Y = (y_{iq})_{m \times n}$, expressing the preferences of experts, where m is the experts number. In the n -dimensional vector space of estimates, two reference points were given, represented by vectors $y^+ = (y_1^+, \dots, y_n^+)$ and $y^- = (y_1^-, \dots, y_n^-)$, which correspond to the best case A^+ and worst case A^- . The best support variant A^+ has only the best grades by all criteria $y_q^{\max} = \max_{1 \leq i \leq n} y_{iq}$, and the worst support variant A^- has only the worst grades $y_q^{\min} = \min_{1 \leq i \leq n} y_{iq}$. For each variant A_i in the n -dimensional space of estimates, the distance to the best support case $d^+(A_i) = d(y_i, y^+) = \left[\sum_{q=1}^n (y_{iq} - y_q^+)^2 \right]^{\frac{1}{2}}$ and the distance to the worst support case $d^-(A_i) = d(y_i, y^-) = \left[\sum_{q=1}^n (y_{iq} - y_q^-)^2 \right]^{\frac{1}{2}}$ were calculated. For each variant A_i the indicator (1) of the relative distance from the worst support case A^- was calculated.

$$h(A_i) = \frac{d^-(A_i)}{d^+(A_i) + d^-(A_i)} \quad (1)$$

The variants were sorted in descending order of the indicator value $h(A_i)$ of the relative distance of the variant A_i from the worst support variant A^- . To collect data on the history of edits to articles in various scientific areas, a module for crawling Socionet articles by a reference graph was implemented, as well as an algorithm for determining the level of contribution of each article to the general index of publication activity by the level of its reference importance for a scientific area.

The study selected two thematic areas represented on Socionet: epidemiology and science studies. As the result of the analysis, it was revealed that for the category "Epidemiology" about 75% of events demonstrate a pronounced injective dependence of edits on the number of requests. The surge in publication activity has been observed within six months after the lockdown due to the spread of COVID-19. For events like "vaccine" and "discovery", the level of positive correlation exceeds 86%.

In the category "Science Studies", a high positive correlation was found between requests and bursts of edits to articles on the problems of this sphere of social science. First of all, this is the creation of councils on priority areas of scientific and technological development, the reorganization of the RFBR, the assessment of the effectiveness of scientific institutions and individual scientists.

Using the model of the network image of science as a basis for the analysis, considering science as a communicative space, opens up new opportunities not only for the development of effective methods for forecasting promising scientific directions for the development of scientific and technological knowledge, but also for the formation of new approaches to the analysis of scientific knowledge as an integral form structuring reality.

5. Conclusion

The tasks of identifying new scientific directions can be solved using various methods and different methodological approaches. As a rule, studies related to forecasting the development of fundamental science and analyzing the prospects for the practical relevance of its achievements are based on considering a kind of opportunities assessment. The formalized methods for identifying scientific areas are based on the idea of science as a self-sufficient research activity that produces new knowledge. The development of methods for the science development predicting based on the network model of scientific cognition, based on the concept of cognition as a procedure for structuring subject reality, will lay the foundations for analyzing the dynamics of the new subject-thematic areas formation in complex communication networks that are a source of innovative solutions.

6. Acknowledgments

The article is prepared with the financial support of the Russian Science Foundation, project № 18-78-10047.

References

- [1] Heikkinen I, Savin H, Partanen J, Seppala J and Pearce J 2020 *Technological Forecasting and Social Change* **155** 119986
- [2] Li D, Capone G and Malerba F 2018 *Research Policy* **48** 649-64
- [3] Yu Y, Ibarra J, Kumar K and Chergarova V 2020 *Technovation* **100** 102180
- [4] Iden J, Eikebrokk T and Marrone M 2020 *International Journal of Information Management* **54** 102150
- [5] Benesova A, Hirman M, Steiner F and Tupa J 2019 *Procedia Manufacturing* **38** 1691-96
- [6] Bootz J, Durance P and Monti R 2018 *Technological Forecasting and Social Change* **140** 80-83
- [7] Loe R, Melnychuk N, Murray D and Plummer R 2016 *Technological Forecasting and Social Change* **104** 78-88
- [8] Forster B and Gracht H 2014 *Technological Forecasting and Social Change* **84** 215-29
- [9] MacLeod C and Radic G 2013 *Studies in History and Philosophy of Science Part A* **44** 188-201
- [10] Gazni A 2020 *Technology in Society* **62** 101276
- [11] Bu Y, Lu W, Wu Y, Chen H and Huang Y 2021 *Information Processing & Management* **58** 102429
- [12] Wenzel M and Will M 2019 *Technological Forecasting and Social Change* **146** 517-33
- [13] Elshawi R, Sakr S, Talia D and Trunfio P 2018 *Big Data Research* **14** 1-11
- [14] Zangeneh P and McCabe B 2020 *Advanced Engineering Informatics* **46** 101164
- [15] Yu D and Pan T 2020 *Expert Systems with Applications* **168** 114238