

# The Methodological Basis of Defining Research Trends and Fronts

N. A. Mazov<sup>a, b, \*</sup>, V. N. Gureev<sup>a, c, \*\*</sup>, and V. N. Glinskikh<sup>a, \*\*\*</sup>

<sup>a</sup>Trofimuk Institute of Petroleum Geology and Geophysics,  
Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630090 Russia

<sup>b</sup>State Public Scientific Technological Library, Siberian Branch, Russian Academy of Sciences, Novosibirsk, 630102 Russia

<sup>c</sup>Novosibirsk State Technical University, Novosibirsk, 630087 Russia

\*e-mail: MazovNA@ipgg.sbras.ru

\*\*e-mail: GureyevVN@ipgg.sbras.ru

\*\*\*e-mail: GlinskikhVN@ipgg.sbras.ru

Received July 15, 2020

**Abstract**—The methodological and technical aspects of identifying research fronts and trends in the development of science are considered. Based on the literature data, a comparison of scientometric methods for finding research fronts was carried out: analysis of publication activity, direct citation analysis, co-citation analysis, bibliographic coupling, and content analysis. The advantages of the combined application of various approaches are shown, the role of expert assessment and verification of the results of scientometric analysis is emphasized. We revealed topical problems associated with the detection of scientific fronts by scientometric methods and showed promising directions in their solution.

**Keywords:** scientific fronts, research fronts, scientific trends, publications, content analysis, citation analysis, bibliometrics

**DOI:** 10.3103/S0147688220040036

## INTRODUCTION

The search for scientific trends and research fronts, that is, topical or promising research, is one of the most significant problems in science policy, scientometrics, and the history and philosophy of science and is of decisive importance at the stages of planning scientific activities. The topic of scientific trends and fronts is obvious if it is dictated by socio-political, environmental, and economic factors or threats to national health [1]. These can be natural disasters, terrorist attacks [2], economic crises, or the appearance of dangerous diseases in the human population, such as the outbreak of influenza A/H1N1 pandemic in 2009 [3] or SARS Cov2 in 2019–2020. In these cases, the scientific community, states, research and funding organizations are actively and consistently involved in the search for solutions to emerging problems. The fronts of science are much less obvious in the absence of such events; they then themselves become an object of study, requiring the development and use of methodological foundations and appropriate tools to identify them.

Scientific trends and fronts, as a rule, are the object of research of science itself, and their identification is an attempt to search for new growth points, as represented by the most promising ideas and developments that are important for the further development of sci-

ence and technology. In other words, a search is carried out for changing objects of research in their relation to existing knowledge and to each other [4]. When identifying research trends and fronts it is predominantly scientometric methods that are used.

In a continuation of previous studies in the field of scientific trends in various fields of knowledge [5–7] and in the absence of reviews on the topic of detecting research fronts, we further consider the concepts of research trends and fronts, classify approaches, and describe the tools for their detection, as well as study the current issues that are pending their decision. When reviewing the literature, the Scopus and RJ “Informatika” databases of VINITI were used without restrictions on time and types of documents. The request included the following keywords: “research front”, “research trend”, and “research focus”. Additionally, sources from lists of references based on search results were used.

## A METHODOLOGY FOR IDENTIFICATION OF RESEARCH TRENDS AND FRONTS

In general, a research front is understood as the situation where the interests and needs of society coincide with the current scientific results [8]. The key object of analysis in identifying research fronts is the groups of scientific publications and their interrela-

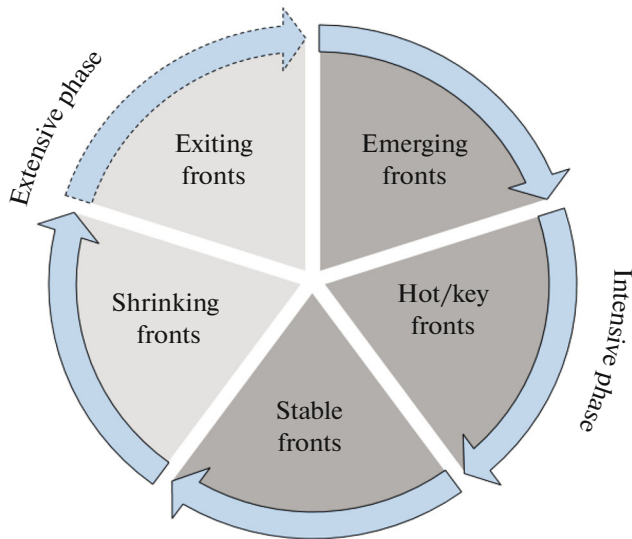


Fig. 1. Types of research fronts.

tionships. According to the classical definition of D. Price, a research front is a densely cited network of recently published papers [9]. In a more detailed definition, a research front is understood as a group of recently published articles with a common topic, which are strictly connected by a network of citations among themselves and weakly connected with publications outside the group [10]. At the same time, strong links between citations within a group are characteristic of a research front at the initial stage of its development, while at later stages, with an increase in the number of citations, including from other scientific areas, this connection weakens. The strength of citation links between publications of clusters is determined by predetermined threshold values that are unique for each scientific field. The sizes of research fronts also depend on the discipline, which usually ranges from a few publications to several dozen. As an example, in the latest report on research fronts from *Clarivate Analytics* the spread is from 2 to 50 articles [11]; sometimes a minimum threshold is set, for example, 10 publications [12].

The concept of a research trend is close in meaning to a research front. A research trend is the collective action of a group of researchers, each of which begins to pay considerable attention to a specific scientific topic: read scientific publications on this topic, refer to them, and publish the results of their own research [4]. At times the concepts of the research front and research trend are used synonymously [13].

The main types of research fronts according to the common classification of G. Small [8] are shown in Fig. 1. The method for identifying the stage of a research front involves comparing clusters of publications for two or more equal consecutive periods of time.

The *Clarivate Analytics* together with the Chinese Academy of Sciences, in its periodic reports distinguishes only two types of research fronts: key (*key hot fronts*) and incipient (*emerging fronts*) [11]. Research fronts are also revealed by the *Elsevier* company based on *SciVal* data, where the most promising topics are determined by the *Prominence* indicator.

Under the influence of various factors, the research fronts of the extensive phase can turn into an intensive one, for example, when new promising research methods appear, with increased funding for the field, when there is an urgent need to develop a topic under the influence of external factors, etc. [1, 12]. As a result of the development of a research front, according to G. Small, it can either develop into a new discipline, or be absorbed by a broader field, which adapts the achievements of a research front to a wide group of studies [8]. In the first case, this indicates the growth of a scientific front, in the second, it indicates its influence on science. As a rule, scientific fronts of interdisciplinary research develop in separate directions, while the absorbed research fronts have little to do with interdisciplinarity, but are gaining citations faster.

Study on research fronts is significant from both fundamental and applied points of view. At the theoretical level, they determine the vector of development of scientific progress and allow tracing the origin and evolution of one field or another, the division and merging of areas of knowledge, contribute to the spread of knowledge between scientific disciplines [14], and allow adjusting organizational processes when new knowledge meets traditional paradigms that dictate research topics, standards and regulations [15]. The identification of research fronts is of practical interest for a wide range of stakeholders involved in the definition of priority areas of scientific research and their funding.

To date, three main scientometric approaches are widely used to identify research trends and fronts: analysis of the dynamics of changes in scientific production, citation analysis with its varieties, and content analysis, as well as their various combinations.

#### *Analysis of Publication Activity to Identify Research Trends*

Analysis of publication activity is usually used to identify research trends, while citation analysis is used to identify research fronts [4, 16]. When analyzing scientific production, expressed by the number of publications, one resorts to models of the growth of scientific knowledge:

(1) in the first model, the growth of knowledge is considered as the cumulative development of new ideas based on previous recent scientific achievements;

(2) the second model assumes that the development of new ideas is based on the entire body of

human knowledge, and not only on recent achievements. According to this model, there is a selective choice of grounds for a new idea from all of human scientific experience;

(3) the third model is based on the theory of scientific revolutions by T. Kuhn [17] and presupposes an intensive growth of knowledge interrupted by periods of calm.

There is no consensus about which of the proposed models most closely corresponds to reality, especially since each of them, to one degree or another, explains the ongoing scientific events in various disciplines. Each of these paradigms can correspond to some mathematical model of the growth of scientific literature, for example, linear or exponential [18]. In natural science disciplines, exponential growth often prevails; when identifying scientific trends researchers therefore turn to D. Price on the exponential growth and obsolescence of scientific literature [19, 20]. The scattering law is used to identify a scientific information trend according to S. Bradford [21], which allows identification of the core of scientific journals of a given subject.

An example of a study using this method is the work to identify research trends in the field of tourism [22]. A circle of authors and organizations that form a research trend on this topic was determined according to zones of concentration and dispersion of Bradford's scientific information, as well as the analysis of the scientific productivity and authoritativeness of publications. The analysis of research trends in the field of borehole geophysics was carried out by the authors of this work: the leading positions of this field in the field of earth sciences were identified, the most productive authors were detected and the redistribution of leading positions between countries over the past 20 years was shown [7]. Further identification of research trends and fronts in the field of geophysics is extremely important, since it is associated with the search for new research areas, primarily for the creation of innovative technologies. In the field of borehole geophysics, "cheap" logging technologies will be the most demanded by both large and small service companies in the near future, which is due to the end of time of "expensive" oil.

#### *Citation Analysis to Identify Research Fronts*

The main method in identifying research fronts is citation analysis, which makes it possible to trace the growth of interest and relevance of a particular topic by the dynamics of changes in the number of citations of publications of a particular field. Citation analysis is considered more objective in comparison with expert assessment, since it takes the opinion of the entire scientific world community of scientists expressed in references [23]. The approach is based on the observation that recent scientific publications are the most cited.

Thus, the identification of thematic clusters of the most cited publications allows us to identify the research front of the corresponding discipline [9]. The response time to published papers varies across disciplines, but on average is 2–5 years, during which half of simultaneously published publications are cited [24]. Within the framework of citation analysis, where both cited and citing publications are clustered, a research front is understood as:

(a) a group of the most cited publications identified by direct citation analysis [4, 9];

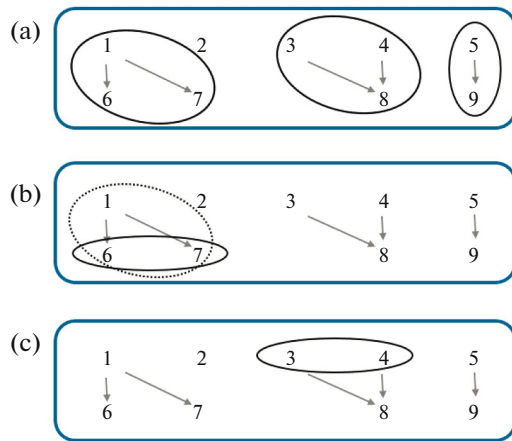
(b) a group of co-cited publications identified by co-citation analysis, positions 6 and 7 in Fig. 2b [25–27]. The cluster of a research front, in addition to co-cited publications, may include citing publications, positions 1, 6 and 7 in Fig. 2b [28]. This definition of research fronts was used by E. Garfield [29]; this approach is still implemented by the *Clarivate Analytics* in periodic reports on research fronts using *Web of Science* databases [11]. There is also a third approach, where a research front refers to publications that cited a cluster of co-cited publications, position 1 in Fig. 2b [30];

(c) a group of publications with similar references, identified by the bibliographic coupling method, positions 3 and 4 in Fig. 2b. According to this approach, the articles of a research front themselves may not have citations [2, 12, 31–33];

(d) with the joint application of the indicated approaches, a research front is understood, for example, as a group of co-cited publications plus a group of publications with similar references [30, 34–37], a group of co-cited publications plus publications citing this group [38], or several groups of publications based on the results of all three approaches [28, 39, 40]. As a rule, when used together, each method is used separately, after which the results are compared or combined. However, it is possible to build complex combined approaches: for example, clustering by bibliographic coupling of those publications in which clusters of co-cited publications are cited; this is then clustering of the first and second levels [30].

The formal similarity with the clusters of publications of research fronts is demonstrated by artificially created groups of articles united by chief editors, for example, within the framework of special issues of journals, where articles of each issue abundantly cite each other. When analyzing research fronts, groups of publications united by similar publication models are usually excluded from the analysis [8].

When describing research fronts, the concept of an intellectual base (*knowledge base*, *knowledge foundation*, *intellectual base*, or *intellectual structure*) is used, which means literature cited by publications of a research front [2, 4, 41]. Many studies demonstrate the thematic proximity of an intellectual base and research fronts [13, 31, 36, 42]. When analyzing co-citation, sometimes confusion of these concepts occurs; while some researchers understand co-cited



**Fig. 2.** The principles of clustering publications used in identifying research fronts. A, direct citation analysis; B, co-citation analysis; B, bibliographic coupling. The top row usually represents recently published publications, the bottom row represents publications of the last 2–5 years. Citation analysis can cover out-of-sample publications.

publications as a research front, others consider them as an intellectual base, and the citing publications as a front (see Figure 2B). In general, the scientometric task is to identify the points of intellectual displacement (research fronts) in the relatively stable scientific literature (intellectual base).

**Co-citation analysis** was simultaneously proposed by I.V. Marshakova and G. Small [43, 44]: two documents are considered co-cited and thematically related if they both appear in the reference list of a third document (with which the two cited documents also have a thematic relationship) and the citation rate is defined as the frequency with which two documents are cited together. Researchers usually choose a small group of publications that are highly cited within a given period of time as a basis for clustering. This could be 1 or 10% of the highly cited articles, the top 10, top 20 articles, etc.

This approach to the search for scientific fronts has a drawback associated with the nature of citation [45]. Accordingly, the ability to take new publications into account, which are often of the greatest interest in the search for scientific fronts, is limited [46]. In other words, co-citation is suitable for identifying a research front at a relatively late stage, and not at the very moment of its emergence [8]. According to one of the developers of the method of G. Small, the analysis of socializing does not cover the entirety of publications on a scientific front, but rather informs about the emergence of such a front; it is designed to do a quick screening of the scientific landscape rather than a definitive delineation of some specific area [8]. The approach does not depend on the vocabulary and language of publications.

**The bibliographic coupling method** proposed by M. Kessler [47, 48] presupposes that two works have a meaningful relationship to each other and are thematically related if they have one or more similar references. Thus, a research front consists of publications that jointly cite other publications. Since references to the analyzed papers are not important and only their reference lists are investigated, the method is free from lag (especially if it is applied not to journal publications, but to preprints) and allows one to analyze newly published papers.

The main idea of the method is as follows: (1) a separate bibliographic reference used in two publications is called one unit of coupling between these publications; (2) several publications form a linked group  $G$  if each member of the group has at least one coupling unit with the test paper  $P_0$ ; and (3) the coupling strength between  $P_0$  and any member of  $G$  is measured by the number of coupling units ( $n$ ) between them. Like co-citation analysis, the bibliographic coupling method is independent of the vocabulary and language of publications and can be automated. In comparison with the analysis of co-citation analysis, the method of bibliographic coupling is used less often to search for scientific fronts [28, 32].

One essential criterion for the study of research fronts is the choice of the citation window. The problem of choosing a citation window received full coverage in [32]: the model of a traditional static 5-year citation window was compared with a sliding overlapping citation window, as well as with the half-life of highly cited articles. Research with a static citation window was found to be the least labor-intensive; however, the most labor-intensive method with a sliding citation window helped to find more research fronts. At the same time, some of the emerging research fronts identified by the two methods did not intersect, which is why the joint use of static and sliding citation windows was recognized as the most effective.

Since the main scientometric approaches to identifying research fronts involve a procedure for clustering bibliographic data, the results of the analysis can be influenced by clustering methods and the choice of threshold values for the measure of similarity between the grouped elements [30, 31]. The object of citation analysis can be both the publications themselves and the authors of these publications, journals and, less often, subject categories [49].

Co-citation analysis is used to search for scientific fronts in various fields of knowledge: HIV/AIDS [15], scientific collaboration [13], library and information science [27]. The method of bibliographic coupling was used to study the historical development of research fronts in the field of anthrax research [12]. The joint use of methods of co-citation analysis and bibliographic coupling was carried out to search for scientific fronts in the library and information science

[36] and in the field of battery research [37]. Author's citations and content analysis of links were used to identify research fronts in the field of bacterial infections [23].

The experience of identifying research fronts not for a discipline as a whole, but for an individual organization is remarkable: in [49], the intellectual base was studied by co-citation analysis; the corpus of publications cited by the organization, on the basis of which a research fronts of the organization itself were further identified. Similar studies of the publication activity and citations of a particular organization were carried out by the authors of this work for more effective information support of scientific projects [50, 51], while the developed methods were also applicable for identifying research trends and fronts. The search for scientific fronts can also be carried out for a separate journal: for example, the *Journal of the American Society for Information Science*. Using the methods of bibliographic coupling and citation analysis, research fronts were identified and a significant closeness of the intellectual base with them was shown [31].

#### *Content Analysis to Identify Research Fronts*

Methods for semantic analysis of metadata and full texts of scientific publications, including neural network technologies [52, 53] and algorithms for detecting rapidly spreading, so-called burst terms, which express new phenomena, are widely used in identifying research fronts [2, 14, 42, 54]. Content analysis investigates the frequency of the use of words in metadata and full texts and, separately, keywords, as well as their joint occurrence in publications. Analysis of the frequency and co-occurrence of keywords is carried out:

(a) on the metadata of publications; in this case, author's or additional keywords assigned in systems are investigated (for example, *KeyWords Plus* [55, 56] extracted from lists of cited literature) and words from various subject thesauri and authoritative dictionaries (for example, *MeSH*), as well as automatically extracted keywords from titles and annotations;

(b) on full texts, where keywords and terms are also extracted and semantically analyzed using software tools.

Some researchers refer to the results of keyword co-occurrence analysis as a research focus, while the research front is considered to be the result of co-citation analysis [57].

To search for scientific fronts in the field of informatics and accounting, the content analysis method identified topics with growing and dying interest, as well as those that have lost their relevance [14]. To extract keywords, entity linking method was used that takes the context of the keyword into account. An approach based on the combined use of searching by association rules, keyword analysis and rapidly spreading terms is presented based on the example of

anticancer developments in nanomedicine [54]. Using linguistic methods for searching for the semantic similarity of texts, the identification of research fronts was described in [46]: a method of comparing phrases and fragments of identical content, not necessarily expressed by the same keywords, was presented. Cluster analysis of author's keywords was carried out to search for scientific fronts in the field of social sciences: the result of a study in five countries was a map of national science, indicating promising areas [1].

Content analysis is often combined with citation analysis methods to identify scientific fronts. Thus, research fronts in the field of artificial intelligence were identified through the combined use of methods of bibliographic coupling and content analysis of keywords [58]. Methods of bibliographic coupling (by co-authors and documents) and content analysis were used to search for scientific fronts in the field of business [41]. A co-occurrence analysis method combined with co-citation analysis has been used to find research fronts in library and information science in Spain [42]. The same two methods were used to analyze co-citation fronts in astrophysical research [59]. A more sophisticated analysis of a research fronts of the interdisciplinary direction is presented using the example of magnetic nanoparticles, where co-citation and co-word networks were studied based on a sample of the 500 most-cited publications [60].

#### THE EFFICIENCY OF DIFFERENT TYPES OF SCIENTOMETRIC ANALYSIS IN REVEALING RESEARCH FRONTS

A researcher's choice of a particular scientometric method is arbitrary in most cases, while in some situations it is necessary to correlate the method with the goals of the study and take the complexity of the calculations into account [28, 39]. Different methods are more or less applicable to one type of research front or another. Thus, the emerging research fronts are better identified by the method of bibliographic coupling, which does not have a time delay. If topological clustering is preferable for research, then citation analysis turns out to be more applicable [39]. If it is necessary to cluster based on the textual similarity of publications, content analysis has proven itself better, in which the frequency analysis of words from metadata or full texts gives better results in comparison with the frequency analysis of an author's keywords.

The choice of the approach has a significant impact on the results, as shown by the example of publications on environmental protection: the intersection of the results obtained in the co-citation analysis and the method of bibliographic coupling was only 33–41%, which in fact indicated different research fronts [30]. Comparison of methods of co-citation analysis and bibliographing coupling was carried out by M. Huang et al., who studies the methodological foundations of the search for scientific fronts [32–34].

In a series of publications, the advantages of the bibliographic coupling were shown: with its use, a greater number of fronts were identified, and several fronts were found at an earlier date [34]. The advantages of bibliographic coupling were disclosed in [39], although it was indicated that in certain narrow areas the method of direct citation analysis may be preferable, since significant publications may have few thematic links in their field but gain a large number of citations from related fields.

A comparison of direct citation analysis, co-citation analysis, and bibliographic coupling was carried out in [61] using the example of research fronts in the field of carbon nanotubes, gallium nitride, and complex network: the direct citation method showed the best results for identifying the early stages of the formation of new topics and contributed to the identification of a larger number research fronts. The next most effective methods were the method of bibliographic coupling and co-citation analysis. Another example of comparing all three methods of citation analysis is the study of scientific fronts in biomedicine, where they were additionally compared with textual analysis [28]. To test the best approach, information on grants was analyzed: since publications on a grant are thematically similar by default, a search was made for the highest concentration of publications on specific grants in each of the clusters.

#### *Weighted Approaches to Improve the Accuracy in Identifying Research Fronts*

Over time, increasingly sophisticated approaches to defining research fronts are being developed, with the goal of increasing the accuracy of clustering. One of the trends in this field is the construction of weighted citation networks. With the assignment of weight to the publications of the cluster forming scientific fronts, a series of studies was carried out by K. Fujita et al., proving the benefits of weighted citation networks [39, 40, 53]. The weight of the publication, automatically determined using neural network training technologies, takes the year of publication, the number of citations of the publication, the field of knowledge, and the strength of the links between the reference list of publications and keywords into account [39, 53]. A significant advantage of the research of this group is that various bibliometric methods are widely combined here.

The analysis of collective dynamics of knowledge networks represented by weighted citation and keyword networks, which takes both incoming and outgoing connections between network elements into account, was presented in [4], which shows the advantages of this method over the analysis of direct citation networks, since it more closely approaches identifying research trends in small areas of knowledge. For more accurate clustering, the PageRank algorithm is used to assign different weights to publications of different sig-

nificance levels: not only are the most cited publications recognized as the most significant in a cluster, but also publications cited by other equally significant publications of the cluster [35].

An analysis of links that establishes the relationship between the cited publications, taking their importance and position in the citation network into account, was carried out to search for research fronts in the field of shareholder activism: during the analyzed period, the development of this field was reconstructed by means of research fronts, including seven stages, from the theoretical origin of the concept to its practical implementation [62]. A weighted approach was used in the search for scientific fronts in chemical technology: 29 clusters were identified containing an average of 5.3 publications; for each cluster, the Price index was calculated, which quantitatively characterizes the degree of novelty of the field [38, 63].

Together with the fundamental applicability of each of the approaches in identifying research trends and fronts, the results of most studies show that the least-accurate results are obtained by the direct citation analysis, although in certain situations it shows advantages over other approaches [39, 61]. In the accuracy of its results the combination of the co-citation analysis and the bibliographic coupling is significantly superior to direct citation analysis, which does not take thematic links between publications into account [34, 39]. The most accurate results in most cases are yielded by the method of bibliographic coupling; co-citation analysis lags slightly behind. The best results are achieved with the combined use of different approaches (and, if possible, different data sets), which should take the variability of publication activity and citation models in different disciplines into account, but such approaches are more laborious and time consuming [28]. Many researchers, for example [1, 2, 64], noted the importance of involving subject experts in the qualitative assessment of the results of scientometric analysis.

#### *Software for Revealing Research Fronts*

Significant attention is paid to the study of research fronts by software developers for visualization and mapping of science [65, 66]. The visualization of bibliographic information is especially valuable for experts because it allows real-time detection of unexpected trends, gaps in scientific knowledge, statistical biases, and other important characteristics of research fronts [67]. *VOSviewer* [22, 41, 57, 68, 69] and *CiteSpace* [2, 13, 26, 42, 60] are most often used; however, ready-made programs are often seen as having limitations, since their functionality is standardized and often does not support innovative approaches [35]. Therefore, sometimes less common software products are used, for example, *Cytoscape* [15] or *BibTechMon* [37], including programs written for a specific study [12].

One of the most functional software for identifying research fronts is *CiteSpace* [2]. The capabilities of the program are presented by its developer using examples of the fields of “mass extinction” and “terrorism.” Research fronts are understood as emerging transitional clusters of ideas, expressed by small groups (several dozen positions) of co-cited publications. At the same time, the study solved the problem of identifying new fields on the basis of linguistic analysis of terms from the metadata of publications (although some researchers insist on involving experts in the designation of new fields [12, 23]).

Experience in using *VOSviewer* was presented by the scientific library of Kent State University (United States): the methods of bibliographic coupling, citation analysis and content analysis were used to identify research fronts in the field of the Internet of things [69]. Dynamic keyword analysis in *VOSviewer* allowed them show changes in research fronts in this area over time.

*The Problem of the Reliability of the Results  
of Scientometric Analysis  
in Identifying Research Fronts*

Since the definition of research fronts is based on an array of scientific publications, the question of the legitimacy of such an approach often arises. In addition to the general criticism of bibliometric approaches, there are somewhat fair statements about the devaluation of the institute of scientific publications associated with an increase in the number of duplicate works, plagiarism, and “predatory” journals, as well as the frequent absence of descriptions of research methods in publications, which prevents their reproducibility. Another criticism concerns the role of publications in rewarding a scientist for his/her work instead of spread of knowledge and a shift of the central channels of scientific communication towards “invisible colleges”. Taken together, this leads to the main question of how much one can rely on bibliometric research of publications to identify research trends and fronts.

Other problems of identifying research fronts are associated with journal articles and, more broadly, with the market for periodicals and its internal standards. As an example, reputable international journals are more willing to publish research results on popular and global topics. Accordingly, in such a limited array of publications, research fronts that are important at the regional or national levels may not be found.

The cautious attitude of reviewers and editorial boards to advanced ideas and developments, often at odds with the scientific tradition, remains an unresolved issue [70]. Modern publishing standards often imply a comprehensive coverage of a scientific problem and a description of a ready-made set of its solutions [71]. However, precisely in relation to research fronts, at the initial stages of developing a new prob-

lem, these requirements are the least feasible and force authors to bypass key issues, whose discussion is most important for understanding the essence of the problem and its causal mechanisms [64, 71]. At times, the overestimated requirements of the editors of journals for breakthrough work lead to the rejection of publications that are significant for science and society. One illustrative example is the article by A.K. Geim and K.S. Novoselov on a new material, graphene, that was rejected by *Nature*<sup>1</sup> (it was later published by *Science*).

Another problem of using journal publications as a basis for searching scientific fronts includes the time lag from the submission of the manuscript to the editorial office to its publication. This adds to the subsequent delay in distributing the journal to libraries or indexing it in bibliographic databases. On average, the delay due to the technological publishing processes is estimated at 1 year [24]. Even if we compare this period with the total time from the birth of a scientific idea to its publication, which, for example, is 4 years in medicine [59], the publication delay appear to be significant.

The databases for the selection of publications themselves have a significant impact on the identification of research fronts [27]. Most of research is based on publications indexed in *Web of Science*, and less often, *Scopus*. In addition to the delay in indexing, such systems have limitations in terms of regional and linguistic coverage of sources; the accuracy of bibliographic metadata is not always high [72]. Despite the annually expanding indexing of conference proceedings, where advanced scientific ideas are discussed much earlier than in print, international databases still tend to predominantly cover journal articles. The need for verification of automatically processed data was already noted in early works, caused by many discrepancies in the spelling of author’s names, variations in the abbreviation of the names of journals, etc. [31]. (For more detail on the problems of identifying bibliographic objects, see [73, 74].)

Some questions remain open, others are eventually answered. Thus, in recent years, reviewers have paid more attention to the transparency of the methodological part of the articles; more and more often initial data are provided in the form of appendices to publications, which significantly increase the reliability and reproducibility of the results. Ethics committees are working to improve the research and publication culture of authors, preventing unfair approaches to the publication of scientific results [75].

At the philosophical level, the role of publications in the system of scientific information and the degree of their applicability to identifying research fronts are analyzed. The transformation of the main properties of a research front into the form of bibliometric indi-

<sup>1</sup> Information from the seminar conducted by the editor of *Nature Nanotechnology* on November 28, 2017, Exhibition Center SB RAS, Novosibirsk.

cators has been substantiated, including such front characteristics as novelty, relevance, interdisciplinarity, risk factors, and a combination of fundamental and applied significance [64]. The central place of publications in scientific research fronts is proved, since in addition to the main function of information delivery, they stabilize unstable networks of various scientific practices and elements [76]. The role of scientific publications is also demonstrated in the reconstruction of the evolutionary development of science: based on the example of research fronts in scientometrics and the historical processes of the intellectual organization of knowledge in this area, their codification and structuring with a simultaneous decrease in entropy have been shown [77]. Based on the example of one area of biomedical sciences, the methodology for constructing a time scale, which allows one to visualize the development of a research front and predict the emergence of new fronts, was presented [12]. On the basis of the theory of the aging of scientific literature, the speed of dissemination of scientific ideas is investigated and the depth of research fronts was revealed [24].

The problem of publishing breakthrough articles, whose material, methodology and results differ significantly from the scientific tradition, finds its solution in the widespread dissemination of open science, the publication of preprints, the development of repositories and models of open peer review. Publication of preprints solves the lag problem. This issue is partially resolved by the development of the system of “articles in print” that are published before the formation of printed issues, as well as early indexing of such publications in bibliographic databases. One possible solution to the problem of publication lag may include the analysis of network publications, whose rate of appearance is significantly higher, as shown by the example of the search for scientific fronts in the field of XML research [78]. In this case, unlike journal databases, special systems are used, for example, *CiteSeer*. It is proposed to solve the problem of publication delay of journal articles by analyzing information about the dates of the publication process (the time of receipt of the manuscript, its approval, and publication); clustering of publications taking these dates into account gives more accurate results in identifying research fronts [59].

## CONCLUSIONS

Over a relatively short period of studying research trends and fronts, a significant complication of the methodology is noticeable: combined approaches, neural networks, a wide range of bibliographic and network databases, and special software is increasingly used. Scientometric methods show their promise due to their rapid adaptation to the changing conditions of the functioning of science and new publication models for the dissemination of scientific information.

The review of research carried out in this article shows that scientometric tools for identifying research fronts have proven themselves well as a source of reliable and objective information for subsequent expert assessment in various fields of knowledge. A wide methodological arsenal of various types of citation analysis and content analysis has been developed. The improvement of the approaches goes in the direction of specifying citation windows, objects of analysis, and identifying the advantages and disadvantages of each of the approaches, taking the types of scientific fronts and research goals into account.

We see the immediate tasks on identifying research fronts and trends as follows. The problem of the initial distrust of the scientific community in breakthrough developments, whose results or methods do not agree well with scientific tradition, awaits a solution. A scientometric solution to this problem is outlined in a broader analysis of network publications. The second task is to increase the speed of identifying new fronts, if possible at the stage of publishing preliminary data on new fields. This requires a further search for methods to neutralize the effect of publication lag.

## FUNDING

This study was carried out with the financial support of the Russian Foundation for Basic Research (project no. 19-011-00531).

## CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

## REFERENCES

1. Nederhof, A.J. and van, Wijk, E., Mapping the social and behavioral sciences world-wide: Use of maps in portfolio analysis of national research efforts, *Scientometrics*, 1997, vol. 40, no. 2, pp. 237–276.
2. Chen, C., Citespace II: Detecting and visualizing emerging trends and transient patterns in scientific literature, *J. Am. Soc. Inf. Sci. Technol.*, 2006, vol. 57, no. 3, pp. 359–377.
3. Gureyev, V.N., Mazov, N.A., Ilyicheva, T.N., and Bazhan, S.I., An informetric analysis of studies on influenza vaccines and vaccination, *OnLine J. Biol. Sci.*, 2017, vol. 17, no. 4, pp. 372–381.
4. Liu, X., Jiang, T., and Ma, F., Collective dynamics in knowledge networks: Emerging trends analysis, *J. Inf.*, 2013, vol. 7, no. 2, pp. 425–438.
5. Gureyev, V.N., Mazov, N.A., and Karpenko, L.I., Russian bioscience publications and journals in international bibliometric databases, *Ser. Rev.*, 2015, vol. 41, no. 2, pp. 77–84.
6. Ilyichev, A., Karpenko, L., Gureyev, V., and Mazov, N., Development of phage display technology: A bibliometric assessment, *OnLine J. Biol. Sci.*, 2016, vol. 16, no. 1, pp. 34–42.



7. Mazov, N.A., Gureev, V.N., and Epov, M.I., Results of scientometric analysis of the international flow of publications in the field of borehole geophysics, *Karotazhnik*, 2017, no. 12, pp. 65–86.
8. Upham, S.P. and Small, H., Emerging research fronts in science and technology: Patterns of new knowledge development, *Scientometrics*, 2010, vol. 83, no. 1, pp. 15–38.
9. Price de Solla, D.J., Networks of scientific papers, *Science*, 1965, vol. 149, no. 3683, p. 510.
10. Hsiao, C.-H., Tang, K.-Y., and Liu, J.S., Citation-based analysis of literature: A case study of technology acceptance research, *Scientometrics*, 2015, vol. 105, no. 2, pp. 1091–1110.
11. Research Fronts 2019. Clarivate Analytics, 2020. [https://discover.clarivate.com/Research-Fronts2019\\_EN](https://discover.clarivate.com/Research-Fronts2019_EN).
12. Morris, S.A., Yen, G., Wu, Z., and Asnake, B., Time line visualization of research fronts, *J. Am. Soc. Inf. Sci. Technol.*, 2003, vol. 54, no. 5, pp. 413–422.
13. Hou, J. and Chen, C., Visualization of the knowledge base and research front of scientific collaboration research, *12th International Conference on Scientometrics and Informetrics – ISSI 2009 (July 28–31, 2009, Rio de Janeiro, Brasil)*, São Paulo, 2009, vol. 1, pp. 944–945.
14. Marrone, M., Application of entity linking to identify research fronts and trends, *Scientometrics*, 2020, vol. 122, no. 1, pp. 357–379.
15. Fajardo-Ortiz, D., Lopez-Cervantes, M., Duran, L., Dumontier, M., Lara, M., Ochoa, H., and Castano, V.M., The emergence and evolution of the research fronts in HIV/AIDS research, *PLoS One*, 2017, vol. 12, no. 5.
16. Ball, R. and Tunger, D., Bibliometric analysis – a new business area for information professionals in libraries?, *Scientometrics*, 2006, vol. 66, no. 3, pp. 561–577.
17. Kuhn, Th., *The Structure of Scientific Revolutions*, University of Chicago Press, 1962.
18. Tague, J., Beheshti, J., and Reespotter, L., The law of exponential-growth – evidence, implications and forecast, *Libr. Trends*, 1981, vol. 30, no. 1, pp. 125–145.
19. Price, D.J., The exponential curve of science, *Discovery*, 1956, vol. 17, pp. 240–243.
20. Price, D.J., *Science since Babylon*, New Haven: Yale Univ. Press, 1961.
21. Bradford, S.G., *Documentation*, Washington, D.C.: Public Affairs Press, 1950.
22. Vega-Muñoz, A., Arjona-Fuentes, J.M., Ariza-Montes, A., Han, H., and Law, R., In search of “a research front” in cruise tourism studies, *Int. J. Hospitality Manage.*, 2020, vol. 85, artic. no. 102353.
23. Kanovei, A.V., Revealing the structure of the research front in a scientific area using analysis of bibliographic references, *Nauchno-Tekh. Inf., Ser. 1*, 1989, no. 9, pp. 22–25.
24. Ivanov, S.A., Investigation of the depth of scientific research fronts, *Mezhdunar. Forum Inf.*, 2003, vol. 28, no. 2, pp. 3–6.
25. Small, H. and Griffith, B.C., The structure of scientific literatures. I: Identifying and graphing specialties, *Soc. Stud. Sci.*, 1974, vol. 4, no. 1, pp. 17–40.
26. Filimonova, N.M., Morgunova, N.V., and Sinyavskii, D.A., Determination of promising areas of research of small and medium-sized businesses, *Nauchno-Tekh. Inf., Ser. 1*, 2014, no. 9, pp. 20–26.
27. Åström, F., Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004, *J. Am. Soc. Inf. Sci. Technol.*, 2007, vol. 58, no. 7, pp. 947–957.
28. Boyack, K.W. and Klavans, R., Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?, *J. Am. Soc. Inf. Sci. Technol.*, 2010, vol. 61, no. 12, pp. 2389–2404.
29. Garfield, E., Research fronts, *Curr. Contents*, 1994, vol. 41, pp. 3–7.
30. Jarneving, B., A comparison of two bibliometric methods for mapping of the research front, *Scientometrics*, 2005, vol. 65, no. 2, pp. 245–263.
31. Persson, O., The intellectual base and research fronts of JASIS 1986–1990, *J. Am. Soc. Inf. Sci. Technol.*, 1994, vol. 45, no. 1, pp. 31–38.
32. Huang, M.H. and Chang, C.P., A comparative study on three citation windows for detecting research fronts, *Scientometrics*, 2016, vol. 109, no. 3, pp. 1835–1853.
33. Huang, M.H. and Chang, C.P., Detecting research fronts in OLED field using bibliographic coupling with sliding window, *Scientometrics*, 2014, vol. 98, no. 3, pp. 1721–1744.
34. Huang, M.H. and Chang, C.P., A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation, *Scientometrics*, 2015, vol. 102, no. 3, pp. 2041–2057.
35. Xu, Y., Zhang, S., Zhang, W., Yang, S., and Shen, Y., Research front detection and topic evolution based on topological structure and the PageRank algorithm, *Symmetry*, 2019, vol. 11, no. 3, artic. no. 310.
36. Zhao, D. and Strotmann, A., The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis, *J. Assoc. Inf. Sci. Technol.*, 2014, vol. 65, no. 5, pp. 995–1006.
37. Schiebel, E., Visualization of research fronts and knowledge bases by three-dimensional areal densities of bibliographically coupled publications and co-citations, *Scientometrics*, 2012, vol. 91, no. 2, pp. 557–566.
38. Mil'man, B.L. and Gavrilova, Yu.A., Clustering of bibliographic references as a method of scientometric analysis of general chemical technology, *Nauchno-Tekh. Inf., Ser. 1*, 1990, no. 12, pp. 24–28.
39. Fujita, K., Kajikawa, Y., Mori, J., and Sakata, I., Detecting research fronts using different types of weighted citation networks, *J. Eng. Technol. Manage.*, 2014, vol. 32, pp. 129–146.
40. Fujita, K., Kajikawa, Y., Mori, J., and Sakata, I., Detecting research fronts using different types of combinational citation, *Proceedings of the 17th International Conference on Science and Technology Indicators (September 5–8, 2012, Montreal, Canada)*, 2012, vol. 1, pp. 273–284.
41. Piñeiro-Chousa, J., López-Cabarcos, M.Á., Romero-Castro, N.M., and Pérez-Pico, A.M., Innovation, en-

- preneurship and knowledge in the business scientific field: Mapping the research front, *J. Bus. Res.*, 2020, vol. 115, pp. 475–485.
42. Olmeda-Gómez, C., Ovalle-Perandones, M.-A., and Perianes-Rodríguez, A., Co-word analysis and thematic landscapes in Spanish information science literature, 1985–2014, *Scientometrics*, 2017, vol. 113, no. 1, pp. 195–217.
  43. Small, H., Co-citation in the scientific literature: A new measure of the relationship between two documents, *J. Am. Soc. Inf. Sci.*, 1973, vol. 24, no. 4, pp. 265–269.
  44. Marshakova, I.V., A system of links between documents, built on the basis of references: According to Science Citation Index, *Nauchno-Tekh. Inf., Ser. 2*, 1973, no. 6, pp. 3–8.
  45. Liu, C.L., Xu, Y.Q., Wu, H., Chen, S.S., and Guo, J.J., Correlation and interaction visualization of altmetric indicators extracted from scholarly social network activities: Dimensions and structure, *J. Med. Internet Res.*, 2013, vol. 15, no. 11, p. 17.
  46. Klimenko, S., Charnine, M., Zolotarev, O., Merkureva, N., and Khakimova, A., Semantic approach to visualization of research front of scientific papers using web-based 3D graphic, *Proceedings – Web3D 2018: 23rd International ACM Conference on 3D Web Technology (June 2018, Poznań, Poland)*, 2018, artic. no. 20.
  47. Kessler, M.M., An experimental study of bibliographic coupling between technical papers, *IEEE Trans. Inf. Theory*, 1963, vol. 9, no. 1, pp. 49–51.
  48. Kessler, M.M., Comparison of the results of bibliographic coupling and analytic subject indexing, *Am. Doc.*, 1965, vol. 16, no. 3, pp. 223–233.
  49. Miguel, S., Moya-Anegón, F., and Herrero-Solana, V., A new approach to institutional domain analysis: Multilevel research fronts structure, *Scientometrics*, 2008, vol. 74, no. 3, pp. 331–344.
  50. Gureev, V.N. and Mazov, N.A., Themes of the publications of an organization as a basis for forming an objective and optimal repertoire of scientific periodicals, *Sci. Tech. Inf. Process.*, 2013, vol. 40, no. 4, pp. 195–204.
  51. Gureyev, V.N. and Mazov, N.A., Detection of information requirements of researchers using bibliometric analyses to identify target journals, *Inf. Technol. Libr.*, 2013, vol. 32, no. 4, pp. 66–77.
  52. Zolotarev, O.V., Khakimova, A.Kh., and Sharnin, M.M., Approaches to the construction of a multilingual ontological resource for the analysis of promising areas of development of the subject area, *Mezhdunarodnaya konferentsiya CPT2019 (13–17maya 2019 g., Tsar’grad)* (International Conference CPT2019 (May 13–17, 2019, Tsargrad)), Tsargrad: Nizhegorod. Gos. Arkhit.-Stroit. Univ., 2019, pp. 298–307.
  53. Fujimagari, H. and Fujita, K., Detecting research fronts using neural network model for weighted citation network analysis, *J. Inf. Process.*, 2015, vol. 23, no. 6, pp. 753–758.
  54. Li, M. and Chu, Y., Explore the research front of a specific research theme based on a novel technique of enhanced co-word analysis, *J. Inf. Sci.*, 2017, vol. 43, no. 6, pp. 725–741.
  55. Garfield, E. and Sher, I.H., KeyWords-Plus™ – algorithmic derivative indexing, *J. Am. Soc. Inf. Sci.*, 1993, vol. 44, no. 5, pp. 298–299.
  56. Garfield, E., KeyWords Plus – ISI’s breakthrough retrieval method. 1. Expanding your searching power on current-contents on diskette, *Curr. Contents*, 1990, no. 32, pp. 5–9.
  57. Zhang, T., Chi, H., and Ouyang, Z., Detecting research focus and research fronts in the medical big data field using co-word and co-citation analysis, *The 20th IEEE International Conference on High Performance Computing and Communications (June 28–30, 2018, Exeter, United Kingdom)*, 2018, pp. 313–320.
  58. Luo, C., Zhou, L., and Wei, Q., Identification of research fronts in artificial intelligence, *2nd Asia-Pacific Conference on Intelligent Robot Systems, ACIRS 2017 (June 16–18, 2017)*, Wuhan, China, pp. 104–108.
  59. Zitt, M. and Bassecouard, E., Development of a method for detection and trend analysis of research fronts built by lexical or cocitation analysis, *Scientometrics*, 1994, vol. 30, no. 1, pp. 333–351.
  60. Liu, P., Chen, B.L., Liu, K., and Xie, H., Magnetic nanoparticles research: A scientometric analysis of development trends and research fronts, *Scientometrics*, 2016, vol. 108, no. 3, pp. 1591–1602.
  61. Shibata, N., Kajikawa, Y., Takeda, Y., and Matsushima, K., Comparative study on methods of detecting research fronts using different types of citation, *J. Am. Soc. Inf. Sci. Technol.*, 2009, vol. 60, no. 3, pp. 571–580.
  62. Ma, V.C. and Liu, J.S., Exploring the research fronts and main paths of literature: A case study of shareholder activism research, *Scientometrics*, 2016, vol. 109, no. 1, pp. 33–52.
  63. Milman, B.L. and Gavrilova, Y.A., Analysis of citation and co-citation in chemical engineering, *Scientometrics*, 1993, vol. 27, no. 1, pp. 53–74.
  64. Hörlesberger, M., Roche, I., Besagni, D., Scherngell, T., François, C., Cuxac, P., Schiebel, E., Zitt, M., and Holste, D., A concept for inferring ‘frontier research’ in grant proposals, *Scientometrics*, 2013, vol. 97, no. 2, pp. 129–148.
  65. Cobo, M.J., Lopez-Herrera, A.G., Herrera-Viedma, E., and Herrera, F., Science mapping software tools: Review, analysis, and cooperative study among tools, *J. Am. Soc. Inf. Sci. Technol.*, 2011, vol. 62, no. 7, pp. 1382–1402.
  66. Mazov, N.A. and Gureev, V.N., Software for scientometric and bibliometric research: A brief overview and comparative analysis, *Elektronnye biblioteki: Perspektivnye metody i tekhnologii, elektronnye kollektzii: Trudy XV Vserossiiskoi nauchnoi konferentsii “RCDL-2013” (14–17oktyabrya 2013 g., Yaroslavl’)* (Digital Libraries: Advanced Methods and Technologies: Proceedings of the XV All-Russian Scientific Conference RCDL-2013 (October 14–17, 2013, Yaroslavl)), Yaroslavl, 2013, pp. 122–127.
  67. Aris, A., Shneiderman, B., Qazvinian, V., and Radev, D., Visual overviews for discovering key papers and influences across research fronts, *J. Am. Soc. Inf. Sci. Technol.*, 2009, vol. 60, no. 11, pp. 2219–2228.

68. Wallace, M.L. and Ràfols, I., Institutional shaping of research priorities: A case study on avian influenza, *Res. Policy*, 2018, vol. 47, no. 10, pp. 1975-1989.
69. MacDonald, K.I. and Dressler, V., Using citation analysis to identify research fronts: A case study with the Internet of Things, *Sci. Technol. Libr.*, 2018, vol. 37, no. 2, pp. 171–186.
70. Gregory, J.G., Citation study of peripheral theories in an expanding research front, *J. Inf. Sci.*, 1983, vol. 7, no. 2, pp. 73–80.
71. Yadav, M.S., Making emerging phenomena a research priority, *J. Acad. Mark. Sci.*, 2017, vol. 46, no. 3, pp. 361–365.
72. Selivanova, I.V., Kosyakov, D.V., and Guskov, A.E., The impact of errors in the Scopus database on the research assessment, *Sci. Tech. Inf. Process.*, 2019, vol. 46, no. 3, pp. 204–212.
73. Mazov, N.A. and Gureev, V.N., The role of unique identifiers in bibliographic information systems, *Sci. Tech. Inf. Process.*, 2014, vol. 41, no. 3, pp. 206–210.
74. Jörg, B., Höllrigl, T., and Sicilia, M.-A., Entities and identities in research information systems. Infrastructures for research and innovation, *Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6–9, 2012, Prague, Czech Republic)*, 2012, pp. 185–194.
75. Gureev, V.N., Lakizo, I.G., and Mazov, N.A., Unethical authorship in scientific publications (a review of the problem), *Sci. Tech. Inf. Process.*, 2019, vol. 46, no. 4, pp. 219–232.
76. Frohmann, B., The role of the scientific paper in science information systems, *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems (October 23–25, 1998, Pittsburgh, USA)*, Medford, NJ, 2000, pp. 63–73.
77. Lucio-Arias, D. and Leydesdorff, L., An indicator of research front activity: Measuring intellectual organization as uncertainty reduction in document sets, *J. Am. Soc. Inf. Sci. Technol.*, 2009, vol. 60, no. 12, pp. 2488–2498.
78. Zhao, D. and Strotmann, A., Can citation analysis of web publications better detect research fronts?, *J. Am. Soc. Inf. Sci. Technol.*, 2007, vol. 58, no. 9, pp. 1285–1302.