Scientometrics Today: A Methodological Overview

Ludmila Ivancheva[1]

2008-09-16

Abstract

The paper considers the current state of scientometrics, in methodological context, as a branch of the discipline "science of science". They are discussed: its fields of study /science by itself in epistemological sense – its general system development, disciplinary structure and interrelations, research front dynamics, etc.; the process of scientific knowledge production – quantitative characteristics of the research potential, communications in science, research productivity, evaluation of scientists and research institutions, research collaboration, a structure of research communities and networks, etc.; and the macro-environment of scientific research – science policy, innovation processes, globalization/; the wider thematic scope of scientometrics, including monitoring of research production, assessment of scientific contributions, dynamical studies of science /in cognitive and social aspect/, revealing emerging research problems, determining of research elites, modelling of science processes and phenomena, study of science-industry relations, etc.; the specific research instruments, objects /and indicators, related to them – "input" and "output" ones/, the empirical basis, and explanatory abilities of scientometrics. Some classifications of scientometric methods are also introduced.

Further, the methodological peculiarities, limitations, and problems of scientometric research are presented, such as multi-factorial dependence, permanent rearrangements of the factors, subjectivity of the measuring results, non-additivity of the empirical data, availability of two types of variables – indicators and latent variables, theoretical importance of the empirical indicators, etc. The interrelations and interactions between scientometrics and other research disciplines /as sociology of science, information sciences, philosophy of science, economics, linguistics, etc./ are also considered in the paper. Some modern tendencies in scientometric studies are highlighted, too.

At the end, the conclusion is made, that the scientometrics becomes an especially perspective part of the general "science of science", and a powerful tool of the research and innovation policy.

1 Introduction

Nowadays the scientometrics, studying mainly the quantitative aspects of science (in cognitive, as well as in social context), has strengthen its position as a significant component of the general Science of science, and it appears to be a completed disciplinary field with clearly outlined subjects of research, specific set of good elaborated research methods and techniques, a significant concerning size and geographical scope research community, numerous research institutions, constituted regular conferences and its own printed organ – the prestigious international journal *Scientometrics*.

2 Scope of research, research instruments, empirical basis and explanatory abilities of the scientometric approach

In general, scientometrics relates to the following subjects of research:

- *science by itself* in epistemological sense – its general system development, disciplinary structure and interrelations, research front dynamics, etc.; main research instruments: "mapping" by "bibliographic coupling", co-citation or co-word analysis, different types of mathematical models;
- the process of *scientific knowledge production* – quantitative characteristics of the research potential, communications in science, research productivity, evaluation of scientists and research institutions, research collaboration, a structure of research communities and networks, etc.; main research instruments: statistical processing (inclusive multidimensional analyses) of the number of sci-

1 Centre for Science Studies and History of Science, Bulgarian Academy of Sciences, Sofia, Bulgaria, l_ivancheva at dir dot bg

entists, publications, citations, co-authorships, grants, projects, etc.;

- *macro-environment* of scientific research – science policy, innovation processes, globalization, etc.; main research instruments: statistical analyses based on financial, patent, and publication indicators.

The wider thematic scope of scientometrics includes issues as: quantitative studies of scientists, projects, funding of research, research infrastructure, etc.; quantitative studies of publications, patents, and citations by institutions, countries, languages, co-authorships, thematic fields, etc.; investigations and monitoring of individual, institutional, or state research production; identification of relations between different research disciplines; studies of cognitive structure of science, or of different research disciplines; studies of structure of research communities; studies of the internationalization of science; dynamical studies of science, tracing the development of a given scientific field, research community, institution, etc.; revealing of emerging research problems; evaluation of research production and scientific contributions (of scientists, institutions, regions, countries, etc.); determining of research elites; assessing the impact factor of scientific journals; modelling of science processes and phenomena, based on mathematical methods; studies of science-industry relations, and of innovation processes; scientific prediction and foresight.

The objects of study in scientometric research (as well as the indicators, related to them), are two main types: "input" ones, connected with the research process – scientists, financial parameters, infrastructure and organization entities, research programs, etc.; and "output" ones, related to the research products – implemented projects, registered discoveries, patents, publications (or their components), as well as the citations of them. The scientific documents (articles, monographs, conference reports, patent descriptions, etc.) and their derivative elements are the major subject of *bibliometrics* as an important part of the considered research approach.

The main set of research instruments, applied in scientometrics, comprises observation, measuring, mathematical processing, comparison, classification, generalization, visualization, and interpretation of data.

Scientometrics deals with three fundamental types of information (Dou 1994):

- *operational information* (of little scope, but very precise and detailed information – for example, about the publication activity of small groups of scientists or separate research units, about the status of a concrete research field, etc.);
- *tactical information* (of wider scope, more generalized information, used for instance in studies of interdisciplinary relations, or of the dynamics of virtual research communities);
- *strategical information* (of much wider scope, more comprehensive and more summarized information, providing data for generalized indicators as the contribution of a given country to the world scientific information flow, or to the rate of growing of scientific discoveries).

An important feature of scientometrics is the possibility to be carried out synchronous, as well as diachronous analyses. It enables the tracing of the dynamic changes of the objects of study – for instance, revealing the shift of the research front in a given scientific discipline or the changes in the research productivity, the evolution of some research networks, etc.

Another significant characteristic of scientometrics is that it enables predictive studies, and also strategic prognostications. Further, as a research field it incorporates empirical, as well as theoretical types of research (dealing mostly with mathematical modelling). To this effect, scientometric studies enable the identification of various scientific facts and regularities, difficultly reachable by other modes of research. This concerns the problem of quantification – the application of mathematical means of research and analysis itself is not only a token of methodological maturity of the scientific discipline, but it provides also a possibility for creation of models, rich in content and enabling the revealing of new phenomena and processes – herein is their heuristic role. A characteristic example in this connection is the revealing the structure of research networks or scientific disciplines, the identification of emerging research fields, the revealing of similarities and differences in research conduct or policies, ascertaining the degree of coherence between different spheres (for instance, academic science and industry), prediction of future development of science, etc.

A broadly accepted empirical source of information for scientometrics are the databases of the Institute for Scientific Information (ISI, Philadelphia, USA) – Thomson Scientific: *Science Citation Index, Social Sciences Citation Index, Arts and Humanities Citation Index*, *Essential Science Indicators, Journal Citation Reports,* etc. In recent years similar information resources and services have been provided by the information system *Scopus* as well.

3 Classification of the scientometric methods

There are some classifications of the scientometric methods and models, submitted mainly by representatives of the Russian school of scientometrics.

So, for instance, Haitun (1983) divides the scientometric methods into several classes: *statistical method* with measures – number of discoveries, number of journals, number of institutions, number of scientists, frequency of co-authorships, and some others; a method of *publication counting* with a measure – number of research products (articles, monographs, patent descriptions, reports, etc.); *citation index* with a measure – number of citations; *text analyses* (content analysis, thesaurus and slang method) with measures – different text entities.

According to Yablonskij (1977), the studies of science by quantitative methods could be classified as following: 1) scientometric studies with objective to be gathered and statistically processed some empirical data (*empirical line*); 2) theoretically-mathematical generalization of scientometric research data with the purpose of revealing the regularities and developing mathematical models of science (*theoretical line*); 3) elaboration of quantitative methods of scientific and technical prognostication and planning of research activities, directed to practical application of the science products (*normative line*). The author classifies the models, elaborated within the framework of the theoretical line, in several groups: dynamical ones (comprising the time parameter), diffusion (epidemic and physical models), synchronously-statistical, and structural (in a cognitive, social, and organizational sense). By the integration of these models they could be revealed interesting features and regularities of science, of importance for the prediction of its future development, for evaluation of research productivity, for assessing the rate of aging of scientific information, etc.

Another Russian researcher, concerned by the problem of application and systemizing of the quantitative methods in science studies, is I. Marshakova (1988). She considers the bibliometric approach, which presupposes quantification of the documental information flows. According to her, there are two fundamental types of quantification:

- tracing the dynamics of the separate objects of study (publications, authors, their distribution among journals or research fields, etc.);
- establishing of relations between the objects, their correlations, classifications, etc. – a “structural method”.

The second type is connected with the revealing the structural (the qualitative) aspect of the state of science. Its major research instrument is the co-citation analysis, finding out prospective relations between the publications.

Another kind of quantification is this, based on a lexical analysis, implemented by:

- distributive-statistical analysis of the texts;
- method of the biology taxonometry.

A more modern classification is this of W. Glänzel (2003). He divides the scientometrics and its methods into the following structural entities:

- dynamical scientometrics, handling with the construction of comprehensive models of growing of the scientific knowledge, the aging of the scientific information, the development of the citation processes, etc.;
- structural scientometrics, corresponding mainly with the problem “mapping of the cognitive structure of scientific knowledge”, based on methods as co-citation, bibliographic coupling or co-word analysis;
- evaluative scientometrics, with a subject – the assessment in the sphere of scientific research, and for the purposes of the science policy;
- prognostic scientometrics, drawing visions about the development of the science processes in the future.

4 Scientometrics – methodological peculiarities, limitations, and problems

A lot of authors indicate the following principal specialties of the quantitative analysis of complex objects as science, which contain a social element:

1. Greater than in the natural sciences multi-factoriality of the phenomena. This requires the analysis of these phenomena to be mostly multidimensional. The evaluation also should be done on the bases of more than one criterion.
2. Permanent rearrangement of the factors according to their degree of impact. Therefore not one of them could be ignored, except in the cases of some synchronous descriptions of the situation.
3. Some subjectivity in the results of measuring. This means that studying phenomena, related to social factors, it is more difficult to reveal reproducible

characteristics of the objects.

4. Less applicability of mathematical formula. In principle, social sciences differ from the natural sciences mainly in the kind of their variables – oftentimes they are non-additive ones. In other words, regarding the variety of values of the measured variables, here often it could not be applied the operation *addition*, which is at the base of all other mathematical operations. The non-additive variables are in fact qualitative ones, and namely this is a quite significant cause for the comparatively low effectiveness of applying quantitative methods in the social sciences, as the "science of science". This problem could be solved to a great extent, avoiding the use of closed scales (limited from above, for instance – assessment marks), and applying mainly the so-called open scales, such as the scientometric ones (Haitun 1983).

Another peculiarity of the scientometric analyses is the fact that they operate with two types of variables – indicators and latent variables. The latent variable (or parameter) is only the notion of the researcher about some measurable characteristic of the object, so, for instance, "quality of the scientists", "research contribution", "content adequacy", etc. The indicators, in contrary, are observable at first hand, empirically directly measurable variables – for example number of publications, number of citations, number of co-words, etc., by which we judge indirectly of the state or dynamics of the corresponding latent variable. In general, the two quantities are connected stochastically. To every indicator's value it corresponds a probabilistic distribution of the possible values of the latent parameter, and vise versa. Namely the problem of the interrelation between the indicators and the latent variables remains a central one in the social sciences, and in particular – in scientometrics (Haitun 1983). It is found out that the operational determination of the latent variable according to one indicator only is poorly productive. If we want to have an exacter description of the reality, we should define the operational variables by means of some groups of interrelated indicators. That's why, in the quantitative studies in the field of scientometrics it is of great importance the elaboration of some suitable *systems* of indicators.

A substantial characteristic of the scientometric analyses is the necessity the data to be comparable, and the results – reproducible. This presupposes, especially in comparative studies, the use of similar concerning form and range information databases, the application of objectivated and well algorithmized methods, and a correct interpretation of the obtained results, considering the possible random and systematic errors.

A topic of the day it has become also the theoretical importance of the scientometric indicators, as well as the limitations in the application of different data bases and indicators (Woolgar 1991). According to Weingart (2005), the bibliometric indicators can be used only as an auxiliary instrument for peer review. Such an extreme position is quite debatable. Of course, the scientomertics as a set of methodological tools has numerous disadvantages and limitations (inaccessibility or incompleteness of the databases, impossibility of considering in the mass some social and subjective factors, non-additivity of the scientometric data, availability of some systematic errors in the methodology – for instance, an omitted elimination of the self-citations in some citation analyses, etc.). So, it would be better the scientometric approach to be combined with some other research methods, in order to achieve more objective and correct research outputs.

5 Relations to other disciplinary fields

Scientometrics interacts actively with disciplinary fields as sociology of science, information sciences, philosophy of sciences (see for instance Geisler 2005), history of science, economics, linguistics, etc. It could be said that to some extent the scientometrics integrates different approaches to studies of science as a complex object, enriching them with a broad arsenal of specific quantitative empirical and analytical techniques, getting in this way to new significant results and conclusions. For example, the sociology of science with a typical object of its studies (networks of scientists – authors of publications), and the bibliography and information sciences, whose element are some aspects of the citation theory, converge by the scientometric methodology (co-citation analyses) on revealing of the cognitive organization of knowledge and its dynamics (that, on their part, are components of the history and philosophy of science).

In contrast to the most sociological methods (inquiring, interviewing, etc.), in scientometric research we deal with relatively materialized phenomena, that gives more exactness and reliability to the results of the analysis. Besides, the scientometric approach relates to the system of science in the mass. Every other method in comparison with it is fragmentary in a sense. For example, they could be delivered and analyzed social and psychological data, but it would be not enough to describe the studied field as a whole. In addition to that, scientometric analyses have a large source foundation – in circulation they are involved some worldwide databases. It provides the opportunity various ways of analyses to be used, revealing specific interrelations and characteristics of scientific communications. So, the quantitative enlargement of the information basis leads to a new qualitative output. In consequence of this we receive some new semantic information, necessary in the process of studying and governing science and research activities (Marshakova-Shejkevich 2002).

But it does not mean that the interaction between both approaches (scientometric and sociological) would not be beneficial. On purposes of some essential scientometric studies – for instance for indicating of concrete interrelations and configurations among authors, institutions, science communities, etc., the scientometric data could be processed by various sociological methods, some of them taken, for example, from the theory of the social networks. By such a study we could make conclusions concerning some typical sociological categories as behaviour, hierarchies, group structures, and outlining of an elite (Brut & Minor 1983). Other methods in sociology (inquiry surveys, analysis of behaviour data, etc.) could also be very useful (for instance, for investigating some aspect of the mechanism of creation of new knowledge), as complementary to the scientometric techniques.

Besides documents, authors, institutions, financial indicators, etc., in some moment the scientometricians have begun to study also networks of language elements, applying methods and techniques from information sciences (Hesse, 1980), or to create new ones (as the co-word analysis). But while the information studies concern only the documents by themselves, without considering the users, the scientometrics pays attention to the interrelations between the information sources and the user – for instance in the citation analysis, in the research of co-authorships, etc. (see Egghe 1994).

Often the scientometrics, with its specific methods of analysis, happens to corroborate or to reject some concepts, elaborated by the research instruments of other sciences. So, the analysis of some socio-economic indicators about the development of the science system (connected mainly with the funding of research activities, with the human potential, etc.) leads sometimes to conclusions, which do not fit with the data from scientometric studies, based on output indicators as number of publications and their citations.

A typical illustration in this respect we meet also in the interrelation between philosophy of science and scientometrics. A purely speculative theory of scientific revolutions, propounded by Kuhn (1970), finds an explicit support (and in some aspects it gets even further developed) by means of some specific scientometric models (see Bailon-Moreno et al. 2005, 2007). Such a model, based on the theory of bifurcations, creates A. Yablonskij (1986) as well. So he finds out that the stability in the static status of “normal science” and the instability in a moment of “scientific revolution” are equally necessary for the development of science. The evolution of the problems, exhausting the given paradigm, and the generation of new problems, threatening this paradigm, ultimately set the paradigm (and together with it – the science community as well) from a stable position into an instable one. As a result of the emerging of some new in principle and usually unpredictable scientific findings (creative “fluctuation”) it happens a saltatorial change of the paradigm. Such kind of processes, according to the author, can be described well by the thermodynamic theory of the open systems, after which the hierarchical organization of the developing object is a consequence of the evolutionary exchange of structures with a growing rate of complexity. Correlating the existence of different levels of organization with the sequence of instabilities, this theory indicates that a state with a given complexity could have “a memory” about previous instabilities, every of them with its own contribution to revealing of a new characteristic, significant for the preservation of the final position. In other words, according to Yablonskij, the knowledge is not “drawn” from the environment, additively adding to the previous knowledge, but it has been created as a result of the development of science, being transformed by the research community, which in the given case proves to be “a non-linear transducer” (if using the cybernetic terminology) of the unorganized information into organized scientific knowledge.

In general, scientometrics directs valences to lots of disciplinary fields that benefits all of them and is leading to more system and reliable investigation of their common subject – the science.

6 Some up-to-date tendencies in scientometrics

The scientometric studies have been facilitated to a great extent by the progress in the field of information and communication technologies, providing unexpected before opportunities for access, exchange, and processing of scientific information. It gets possible to incorporate scientometric data in large expert information systems for providing the policy and decision makers with strategic information, such as Web of Knowledge. Other direction of development is the enlargement of the scope of scientometric analysis, enrolling new types of sources – for instance business information, websites, presentations, etc. Scientometrics has to answer more adequately the growing market demand for such kind of information. In this respect, the so-called "webometrics" undergoes a rapid progress.

Another tendency is the creation of complex "hybrid" indicators, combining purely scientometric information and economic, social, in particular – demographic type of data, such as the so-called "factor of scientific development", incorporated indicators as number of publications of the scientists of a given country, indexed in the databases of ISI – USA, Philadelphia; number of population of the country; the total number of the world population for the corresponding year (Dikusar 1999). Lately it is noted a drive towards more strongly and systematic commitment of scientometric methods to the qualitative approaches (Leydesdorff 2001).

It advances also the process of "mapping" the networks of authors, disciplinary fields, institutions, journals, etc. Besides the classical methods as cluster analysis, factor analysis, graph theory, multidimensional scaling, etc., it has been applied more largely the theory of neural networks as well, with its characteristic "algorithm of Cohonen" (Campanario 1995; Guerrero & Anegyn 2001).

One of the modern trends in the development of scientometrics is the study of the relation academic science – industry, and of science – industry – government relations (see Cassiman et al. 2007). The external context of the scientific research, i.e. the relation between the process of production of scientific knowledge and other social domains as the governance and the industry, activates indicators other than publications, patents, and citations (Leydesdorff & Meyer 2007).

Scientometrics gets the role of a method of analysis of the production and dissemination of knowledge in the innovation systems. It could be used for an identification of the actors in the research intensive innovation system, the degree of technological specialization of particular companies, the networks and characteristics of the partnerships (Sandstrom et al. 2000). So, for example, the correspondence factor analysis is applied for investigating the correlation between countries and technological spheres, and for depicting the tendencies in patenting (Dore et al. 2000). The scientometric studies are focused also on the identification of new technology branches (for instance, in the field of bio- and nanotechnologies). It is looking for some new indicator of technology development as well, such as the so-called Literature-based Innovation Output data – LBIO data) (Panne 2007).

7 Conclusion

Worldwide scientometrics is becoming a more powerful instrument of science policy, determining to a great extent the way of a project and institutional funding by assessment of priorities, perspectives, and capacity.

As a whole, scientometrics becomes a very perspective research field in the general studies of science, providing powerful and effective instruments for analyses and evaluations in the sphere of science as a significant accelerator of the economic growth and social prosperity, helping to realize the Lisbon strategy for establishing a knowledge-based society.

References

Bailon-Moreno, E. et al. (2005). The unified scientometric model. Fractality and transfractality. *Scientometrics*, 63 (2), 231-257.

Bailon-Moreno, E. et al. (2007). The pulsing structure of science: Ortega y Gasset, Saint Matthew, fractality and transfractality . *Scientometrics*, 71 (1), 3-24.

Burt, R. Minor, M. (Eds.). (2003). Applied Network Analysis. Sage, Beverly Hills.

Campanario, J. (1995). Using neural networks to study networks of scientific journals. *Scientometrics*, 33 (1), 23-40.

Cassiman, B., Glenisson.P., Van Looy, B. (2007). Measuring industry-science links through inventor-author relations: A profiling methodology. *Scientometrics*, 70 (2), 379-391.

Dikusar, A. (1999). The interrelation between the process of social-economic and scientific development of the society (in Rus.). *Naukovedenie*, (2), 51-74.

Dore, J. et al. (2000). Multidimensional analysis of trends in patent activity. *Scientometrics*, 47 (3), 475-492.

Dou, H. (1994). In which business we are? *Scientometrics*, 30 (2), 401-405.

Egghe, L. (1994). Bridging the gaps: conceptual discussion discussion on informetrics. *Scientometrics*, 30 (1), 35-47.

Geisler, E. (2005). The measurement of scientific activity: research directions in linking philosophy of science and metrics of science and technology outputs. *Scientometrics*, 62 (1), 269-284.

Glänzel, W. (2003). Bibliometrics as a research field. A course on theory and application of bibliometric indicators. Course Handouts.

Guerrero, V., Anegyn, F. (2001). Reduction of the dimension of document space using the fuzzified output of a Kohonen network. *Journal of the Am. Society for Inf. Sci. and Tech*., 52 (14), 1234-1241.

Haitun, S. (1983). Scientometrics: state and outlook (in Rus.). Nauka, Moscow.

Hesse, M. (1980). Revolutions and Reconstructions in the Philosophy of Science. Harvester Press, London.

Kuhn, T. (1970). The Structure of scientific Revolutions. Chicago.

Leydesdorff, L. (2001). The Challenge of scientometrics. The Development, Measurement, and Self-Organization of Scientific Communications. Universal Publishers.

Leydesdorff, L., Meyer, M. (2007). The scientometrics of a Tripple Helix of universities – industry – government relations (Introduction to the topical issue). *Scientometrics*, 70 (2), 207-222.

Marshakova, I. (1988). System of citing scientific literature as an instrument of observation of the development of science (in Rus.). Nauka, Moscow.

Marshakova – Shejkevich, I. (2002). The contribution of Russia to the development of world science: a bibliometric evaluation (in Rus.). *Otechestvenie zapiski*, (7), 16-27.

Panne, G. (2007). Issues in measuring innovation. *Scientometrics*, 71 (3), 495-507.

Sandstrom, A., Petterson, I., Nilsson, A. (2000). Knowledge production and knowledge flows in the Swedish biotechnology innovation system. *Scientometrics*, 48 (2), 179-201.

Woolgar, S. (1991). Beyond the citation debate: towards a sociology of measurement technologies and their use in science policy. *Science and Public Policy*, 18, 319-326.

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics*, 62 (1), 117-131.

Yablonskij, A. (1977). Models and methods of mathematical studies of science: an analytical review (in Rus.). Nauka, Moscow.

Yablonskij, A. (1986). Mathematical models in science studies (in Rus.). Nauka, Moscow.