System of Document Connections Based on References

Nauchn-Techn.Inform. Ser.2, 1973 (6): 3-8

Irena Marshakova-Shaikevich - ishaikev@mail.ru

A new method of formal classification of documents is introduced. This method of prospective coupling is based on quantitative analysis of references in *Science Citation Index*.

Analysis of bibliographic references is becoming a tool of studying information processes in science, of classification of documents. In order to classify documents in a field of knowledge it is important to study all references in the documents, connected with the given direction of research. The numerically expressed data of citing practices within the field should be followed by further conceptual analysis.

The proposed method is the logical opposite of the method of bibliographic coupling. From the mathematical point of view citation network is a set of documents with the relation of citing imposed on it. In other words it is a union of a set of citing papers and a set of cited papers. A citation network is a potential base for various classification of member-papers. The method of bibliographic coupling, worked out by M. Kessler of MTI gained a great popularity in scientometrics and document classification. The method is based on one main principle: two publications are connected if they have a number of common references to previous literature. The technical apparatus of the method is simple:
1) a reference, used in both publications, is called a unit of their connection;
2) a few articles constitute a group G, if every member of the group has at least one connection with the rest of the group;
3) the strength of connection between an article P and any member of G is measured by the number of common references.

M. Kessler based his bibliographic coupling on the assumption that references in technical papers give the possibility to the author to show his or her intellectual environment. Thus two papers having similar references are kindred. The gist of Kessler's method is the idea that connection of documents depends on the number of common cited documents. Once established this connection does not change with time. No matter how many newcomers will join the corpus of papers under study, the connection will remain constant. Accordingly Kessler's bibliographic coupling may be called RETROSPECTIVE. Any new paper is connected with previous literature and this connection will remain stable in future. New coming papers easily find their place in the whole

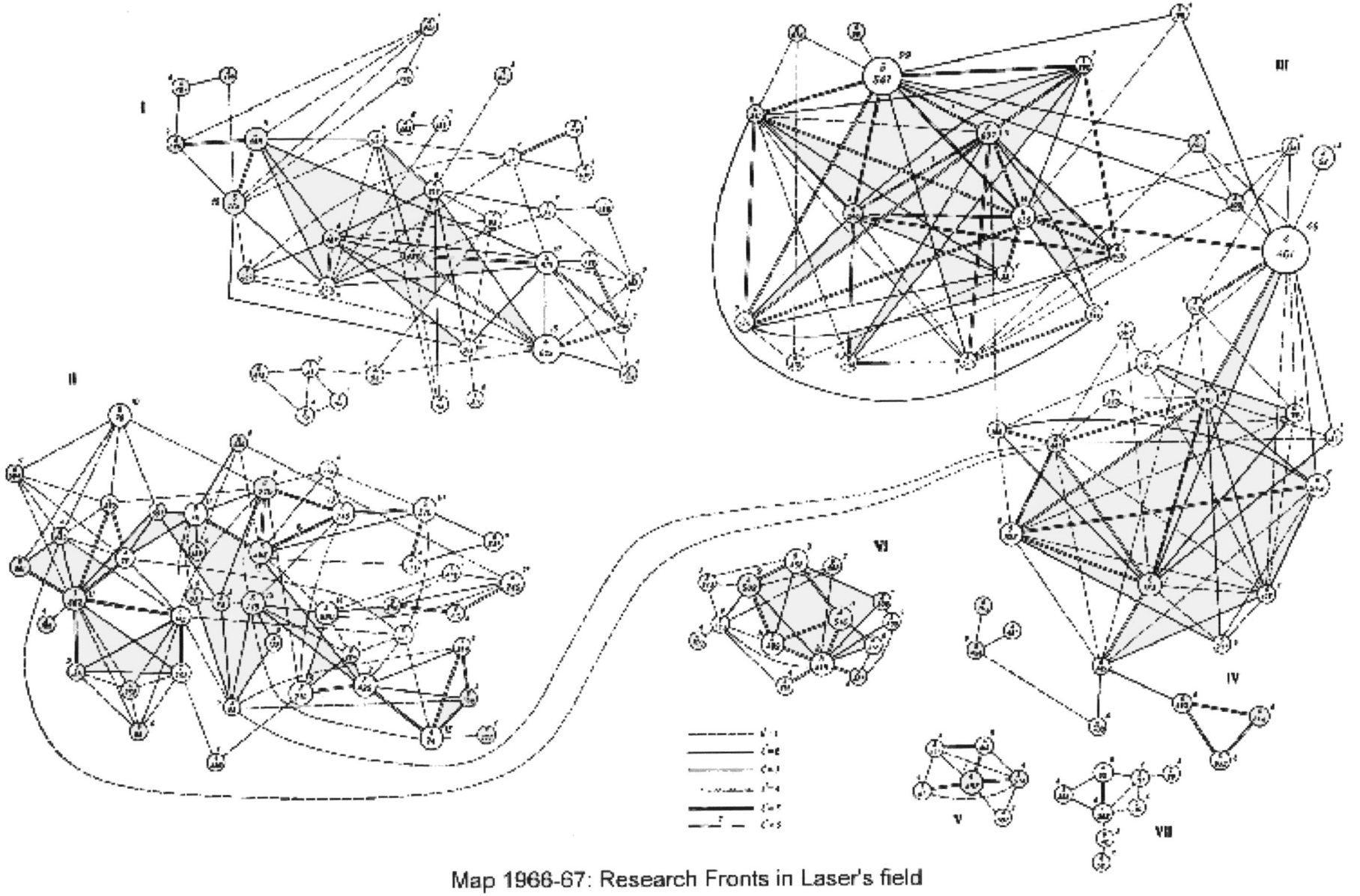
framework of literature.

Interest in the problem of changing classifications calls for a slightly different aim: - to measure documents' connection by common citing papers. Two documents ('a' and 'b') are considered connected if there is a number of other documents, each citing both 'a' and 'b'. At the moment of publishing one cannot say anything if 'a' and 'b' are connected, the picture becomes clear when new papers appear, citing both 'a' and 'b'. The connection measured in this way may be called PROSPECTIVE. The prospective connection of documents entirely depends on the development of science, which makes it interesting for the analysis of evolution of classifications and for scientometrics, in general. The technical apparatus of the present method of prospective coupling is more complicated than that of bibliographic coupling.

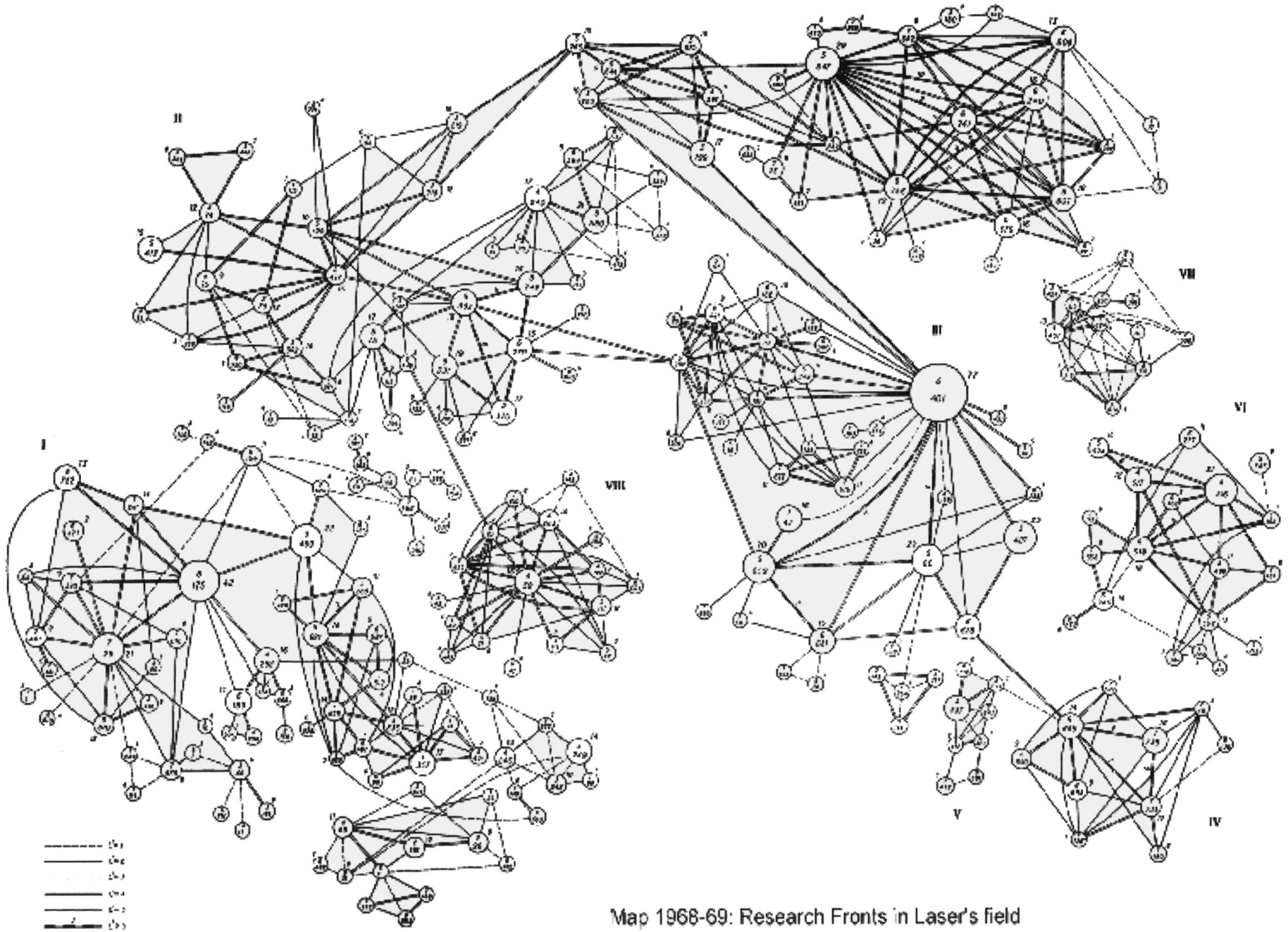
The development of the method included two steps: first a mathematical classificatory model was elaborated and then a procedure of eliciting groups of documents by means of prospective coupling was worked out. For this purpose the documents of the archive under study are divided into cited (basic) documents and citing (prospective) documents. The number of citations to a basic document 'a' is called its volume, the number of references in a prospective document 'x' is called its weight. The calculation of the strength of connection between basic documents 'a' and 'b' (the ultimate aim of the present study) takes into account their respective volumes as well as the weights of prospective documents.

The present procedure of classification was tested on the corpus of documents (8380 publications) under heading 'Generators of stimulated radiation' (lasers) from VINITI abstracts journal 'Physics' for 1961-1969. The period witnessed a spectacular growth of publications on lasers: only seven documents in 1960, 125 documents in 1961 and 1525 documents in 1969. *Science Citation Index* served as the source of references.

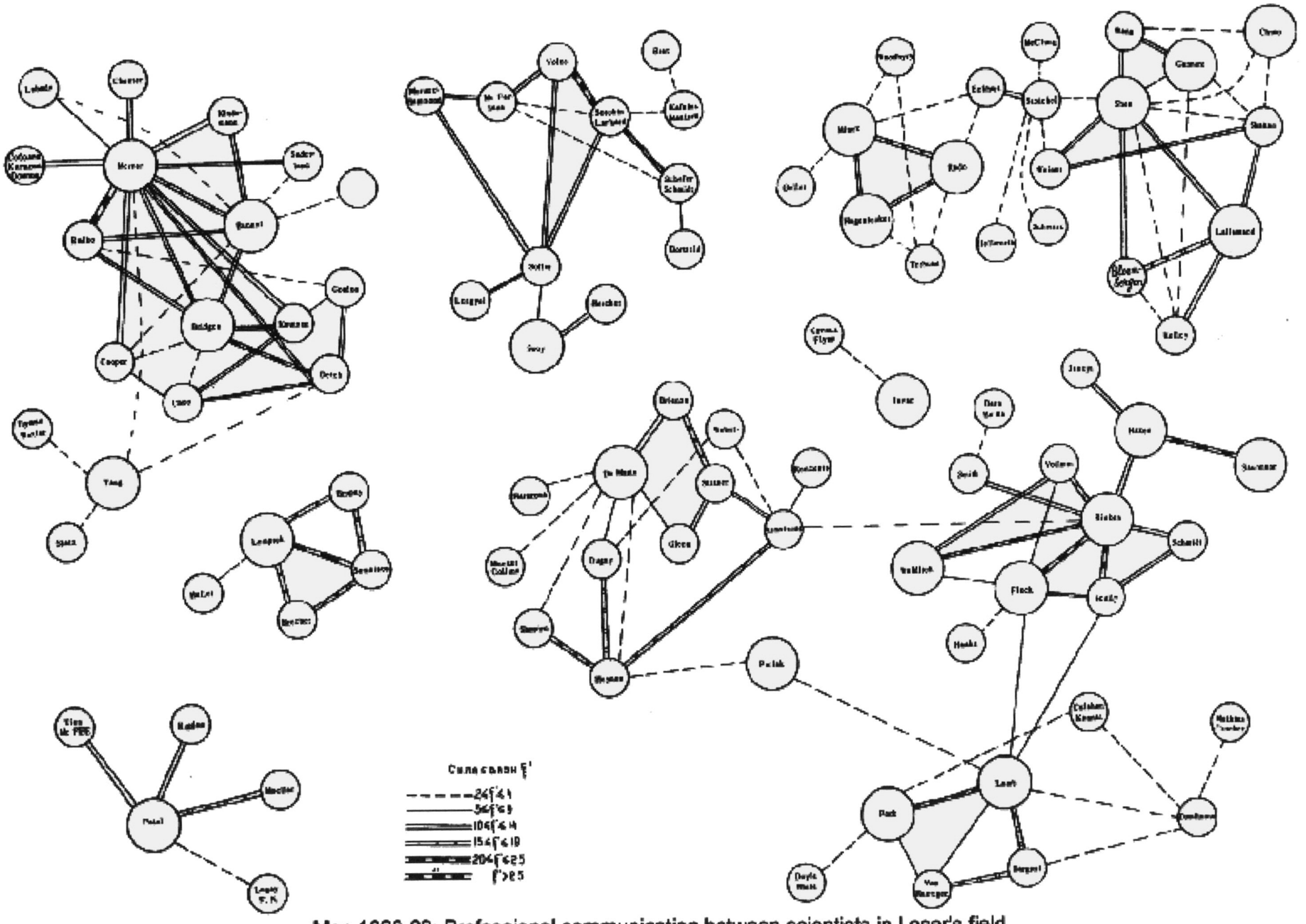
The experiment was carried out in two separate stages
(1) for 1966-1967 and (2) for 1968-1969.
On the basis of the formal mathematical model and resulting procedure 13 autonomous groups were found for 1966-1967 and 26 groups for 1968-1969 (see map1966-67 and map 1968-69) : those groups are easily interpreted and labeled e.g. Laser's theory of laser; Quantum theory of OKG; Laser in magnetic field; Stimulation of laser radiation; Self-focusing of laser beams; Brillouin scattering; Interaction of oscillations; Zeeman effect in lasers; Giant pulse laser; Research on distribution of laser radiation; Generation of optical harmonics; Solid-state laser; Organic dyer laser; Molecular gas lasers; Liquid lasers; Ion lasers.



Map 1966-67: Research Fronts in Laser's field



Map 1968-69: Research Fronts in Laser's field



Map 1968-69: Professional communication between scientists in Laser's field

Later on the idea of prospective coupling was extended by the author to constructing networks of scientists [see Irina Marshakova. Citation Networks in Information Science // Scientometrics, Vol.3 No.1 (1981) 13-26 and I.V.Marshakova . THE SCI SYSTEM AS A MEAN OF MONITORING OF SCIENCE DEVELOPMENT. Moscow, Nauka, 1988 - 288P (in Russian)]] Using the same terminology one can speak of prospective communities of authors (groups of researches linked by prospective coupling). The 1968-69 map, which included 442 publications, served as a source for establishing links between authors. The productivity (number of publications) of each author in 1968-69 was taken into consideration, when the sum of mutual prospective links was calculated. On the basis of those normalized links between authors a n-measured symmetrical matrix was built. Graphic representation of that matrix was published as a map of author-to-author links (see map Professional communication between authors in the Laser's field 1968-69). There is a difference between maps of co-cited publications and maps of co-cited authors, which means that cognitive and social structures of a field of knowledge differ as well. For a more extensive treatment of the subject see [I.V.Marshakova . THE SCI SYSTEM AS A MEAN OF MONITORING OF SCIENCE DEVELOPMENT. Moscow, Nauka, 1988 - 288P (in Russian)]

The author thanks Dr Aneta Drabek for creating electronic copies of the co-citation maps of science.