

Đilda Pečarić

The University of Zagreb. Department of Information Sciences
dpecaric@ffzg.hr

Miroslav Tuđman

The University of Zagreb. Department of Information Sciences
mtudman@ffzg.hr

Bibliometrics Approach to Scientific Communication Structure

(Bibliometryczne podejście do naukowej struktury komunikacji)

Słowa kluczowe: modele komunikacji, nauka o informacji, analiza cytowań, dysertacje doktorskie

Abstrakt: W artykule zostały przeanalizowane modele komunikacji w zbiorze cytowań z dysertacji doktorskich z dziedziny informacji naukowej na chorwackich uczelniach w latach 1978-2009. Model komunikacji naukowej jest zdeterminowany przez czynniki instytucjonalne, językowe i czasowe; skutkiem tego jest możliwe, by dokładnie opisać tradycyjny podział na komunikację pierwszo-, drugo- i trzeciorzędową za pomocą metod ilościowych. Analiza 28,188 zacytowanych fragmentów bibliograficznych za pomocą analizy cytowań dała rezultat sieci klastrow oznaczających dyscypliny naukowe (archiwistyka, bibliotekoznawstwo, muzealnictwo, nauka o komunikacji, informacja naukowa, systemy informacyjne i leksykograficzne). Zmiana słów kluczowych była zauważalna biorąc pod uwagę okres czasu, przynależność do różnych towarzystw naukowych, czynniki kulturowe i funkcje w organizacji wiedzy. Według rezultatów tej analizy wnioski mogą być oparte na wskaźnikach ilościowych: jest zatem możliwe, by uznać strukturę głębokiej komunikacji (strefa wiedzy pojęciowej, empirycznej, badawczej), poza organizacyjnymi, komunikacyjnymi i międzykulturowymi czynnikami wymiany wiedzy.

Keywords: Communication Models, Information Science, Co-Citation Analysis, Doctoral Dissertations

Abstract: Communication patterns are analyzed on the corpus of citation literature retrieved from doctoral dissertations in Information Science from 1978 to 2009 at Croatian universities. Scientific communication model is affected by institutional, language and time determinants; thereby it is possible to accurately describe traditional distribution on primary, secondary and tertiary communication by quantitative methods. Analysis of 28,188 cited bibliographic units by co-citation method resulted with network of clusters that matches scientific disciplines (archivistics, librarianship, museology, communicology, information science, information systems and lexicography). Alterations of key authors were observed according to time periods, affiliation to different scientific communities, cultural determinants, and functions in knowledge organization. According to the results of this analysis the conclusion can be drawn based on quantitative indicators: it is possible to recognize deep communication structure

(conceptual knowledge zone, empirical knowledge zone, research front zone), beside organizational, communicational and intercultural determinants of knowledge exchange.

Introduction: Bibliometric Research of Communication Patterns

At the beginning impact factor analysis and co-citation analysis are used to describe „knowledge maps“ and “cognitive structures of science”¹³. Those researches were a basis for estimate: from the influence of certain authors, scientific journals, institutions, to scientific policies and development of scientific information according to disciplines, fields and topics of interest. We want to know if quantitative indicators could be used for the recognition of phenomena and processes that could be further used for the description of communication models in scientific communication.

We used citation analysis to explore certain characteristics of scientific models that are dominant in scientific communication. This citation analysis is part of much larger bibliometrics’ research of doctoral dissertations done in Information Sciences at the Universities of Croatia. Basis for this abbreviated overview of bibliometric research are data from 170 doctoral dissertations done in period from 1978 to 2009.

About factography

Basic information about dissertations, citations, and authors are as follows.

2.1 Data about dissertations

In the period from 1978 to 2009 at Croatian universities 170 doctoral dissertations were done in seven different information sciences disciplines: 61 in information systems, 35 in information science, 28 in communicology, 26 in librarianship, 10 in archivistics and documentation, 9 in museology, and one in lexicography. The majority of doctoral dissertations was made at the Faculty of organization and informatics in Varaždin – 78, followed by the Faculty of Humanities and Social Sciences in Zagreb – 73 doctoral dissertations. According to the periods of production: 21 doctoral dissertations were made until 1989; 62 doctoral dissertations from 1990 to 1999; 87 doctoral dissertations from 2000 to 2009.

Altogether 66 mentors supervised 170 PhD candidates, 33 mentors supervised only one candidate, 9 mentors supervised two candidates, 8 mentors supervised three candidates, and also another 7 mentors supervised four candidates. Two mentors supervised five PhD candidates, two mentors supervised six candidates and 3 mentors supervised seven candidates. And finally, one mentor supervised eleven candidates and one mentor supervised thirteen candidates.

¹³ First bibliometric researches of information sciences were done by B.C. Griffith, H. Small, while in Croatia they were done by M. Tudić et al. (1984, 1988). Overview of those researches is given in H.D. White and K.W. McCain.

2.2 Data about cited documents

In 170 doctoral dissertations the total of cited bibliographic units is 28,188 (thereof 5,920 units are without authors). From 22,268 cited units with authors 653 units are self-citations (3%).

If 28,188 cited documents are 100% then: 21% of cited documents are without author; 36% of documents are cited just once; 43.3% documents are cited more than once (12,215).

Cited half-life from all documents is 7.5 years. Certain differences exist according to the type of cited documents: for monographs' cited half-life is 9.3 years, for journals it is 7.9 years, and for semi-publications it is 7.7 years. The difference exists also among disciplines. The shortest cited half-life is in information systems 4.1 years, and the longest in museology 12.5 years. For information science it is 7.0 years, for librarianship it is 7.6 years, for communicology it is 8.2 years, and for archivistics and documentation it is 9.9 years.

Co-citation analysis is used for the retrieval of citation clusters. Frequency of citation pairs is calculated. Total number of pairs, i.e. total number of co-citation is 108,228 and their distribution is as follows: one pair with frequency of 14, one pair with frequency 12, one pair with frequency 11, one pair with frequency 10; 3 pairs with frequency of 9; 12 pairs with frequency 8; 16 pairs with frequency 7; 43 pairs with frequency 6; 115 pairs with frequency 5; 280 pairs with frequency 4; 1,282 pairs with frequency 3; 12,310 pairs with frequency 2; 145,819 pairs with frequency 1.

2.3 Data about cited authors

From the overall number of citations (28,188), there are 22,268 citations with authors; that is, 13,162 is total number of cited authors. 10,053 (76.4%) authors are cited only once, and 3,109 (23.6%) authors are cited more than once. Those 3,109 authors that are cited twice or more, cover 54.9% of citations. So, the rule that small number of authors is frequently cited is repeating: 636 authors that are cited 5 or more times cover 27.3% of citations; 182 authors that are cited 10 or more times cover 14.7% of citations. The first 22 of the most frequently cited authors cover 4.2% of citations, and the first 44 authors cover 7% of citations.

Constants in Scientific Communication

Figure no 1 shows citation frequency according to the age of cited references: total frequency of all citations and frequency of most cited authors. What this graph illustrates and is also confirmed in the analysis of other data in quoted research is: the distribution curve is always the same or similar to the one shown in figure no 1, regardless of the frequency of citation, self-citation, citation according to different languages or citations that are quoted only once. This is also confirmed when the data are fragmented according to following criteria: scientific discipline, time periods or institutions at which doctoral dissertation were made.

The pattern is the following: if we know cited half-life ($t/2$), period in which 50% of documents are cited, then first 25% of documents are cited between half of

cited half-life ($t/4$). In time period $t/4$ maximum frequency from overall number of cited documents is reached. Therefore, citation curve has log-normal distribution, with maximum in time period $t/4$.

Second perceived regularity is that in time period between $t/2$ and t , i.e. second time period of cited half-life, following 25% of documents are cited. After double cited half-life, the last 25% of documents are cited, documents whose ages cannot be statistically predicted.

Figure no 1 is a good indicator about the place and space of influence of the most cited authors. The most cited authors are not among the first 50% of cited documents, which are cited until cited half-life. The biggest number of cited authors is, most often, in time period between $t/2$ and t , which means among 25% of documents that are cited after cited half-life. That greatly nullifies enforced opinion that the most valuable authors are the most recent and the most frequent authors in the area of cited half-life. This is the area of the highest density of citations in a short time period, the highest intensity and selection of the information, or in Capurro's terminology selection of "empirical knowledge"¹⁴.

Figures 2 to 4 show that most cited authors in Croatian and foreign languages form librarianship, communicology and information systems, and their distribution according to citation frequency and average age of cited references.

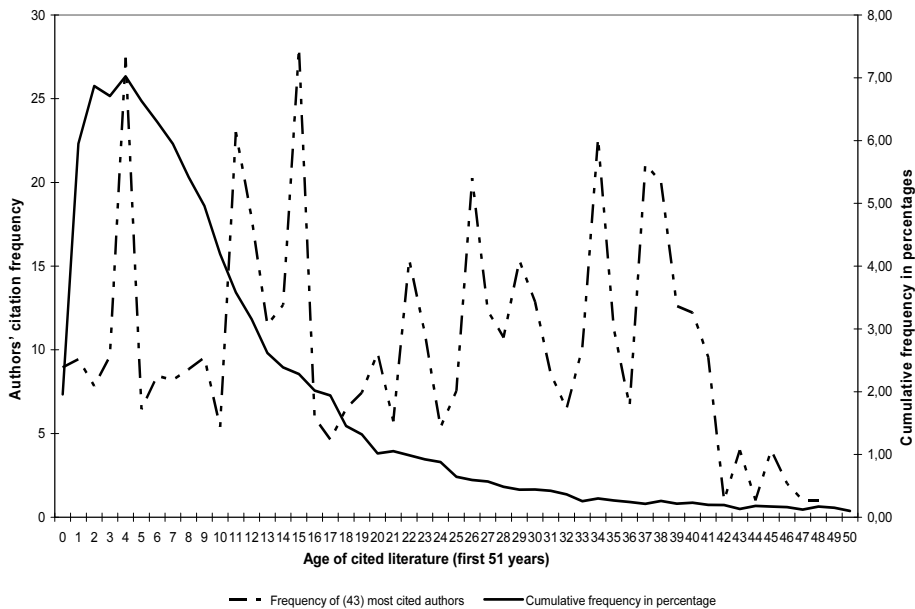
Those figures (2 to 4), once again, confirm previous conclusions. Typically, most cited authors are the ones that are cited after time of $t/2$, i.e. after cited half-life. There is no significant difference between publications cited in Croatian language and foreign languages.

Co-citation analysis appears to be insufficient for the analysis of communication patterns, because it does not take into consideration time dimension of citations or direction of communication: who gives "acknowledgement" to whom in the communication process.¹⁵ And these two determinants are important for the understanding of communication behavior.

¹⁴ According to R. Capurro (2006) primary phenomenon that information science should be researching is message, not information. According to his interpretation, message is the key phenomenon in communication process; information science should be occupied with communication in which 'message' occurs like "meaning offer", and 'information' is the process of message selection. The process of interpretation ("understanding") of messages and information is the substance of communication. That is why Capurro defines basic concepts in the following way: "data" are sets of symbols that represent empirical perceptions; "information" is a set of symbols that represent empirical knowledge; "knowledge" is a set of symbols that represent thoughts, which the individual justifiably believes to be true; "message" is a set of symbols that represent any meaningful content.

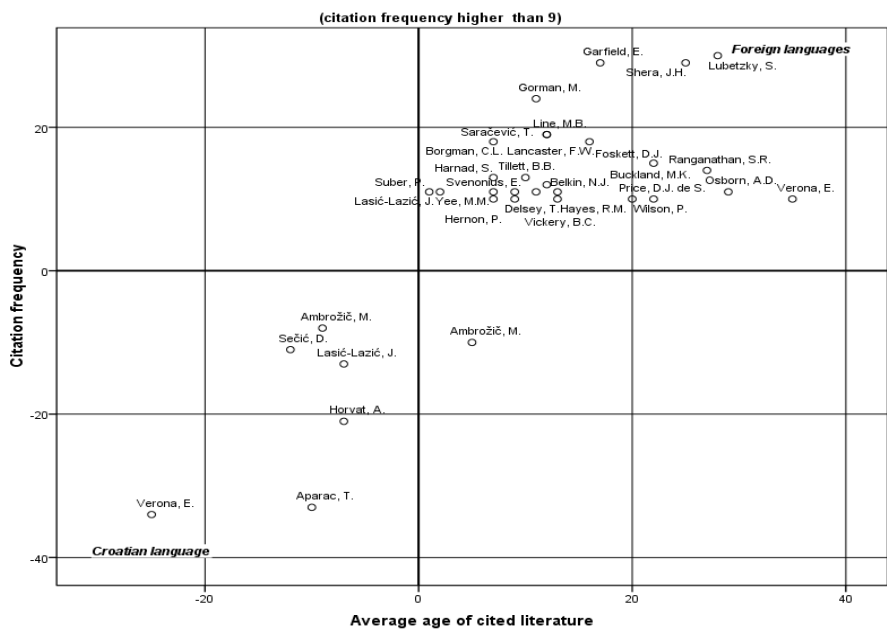
¹⁵ The difference between reference and citation is important as J. Petrak (2003) points out.

Figure 1. Distribution of overall citation frequency and the most cited authors according to the age of cited literature



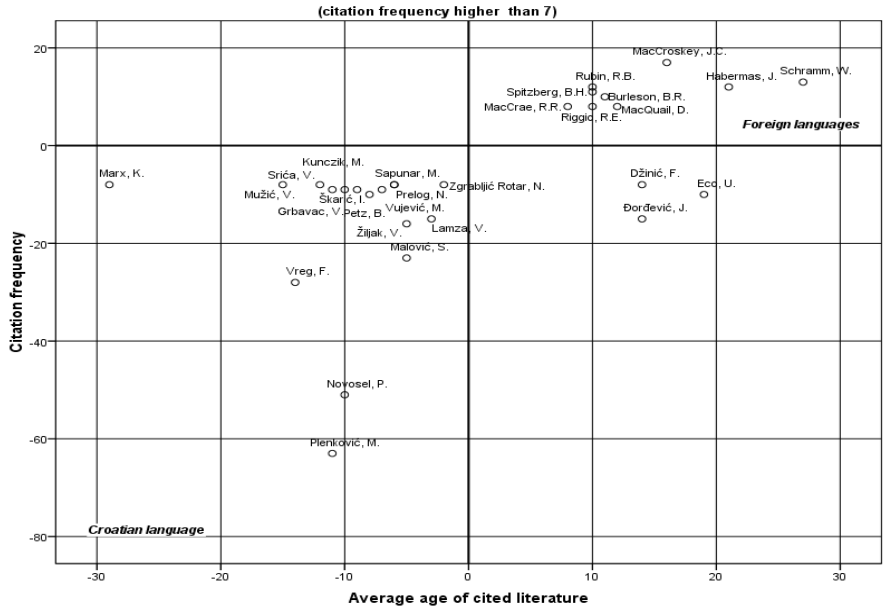
Source: self-elaboration.

Figure 2. 30 most cited authors in librarianship from 1980 to 2009



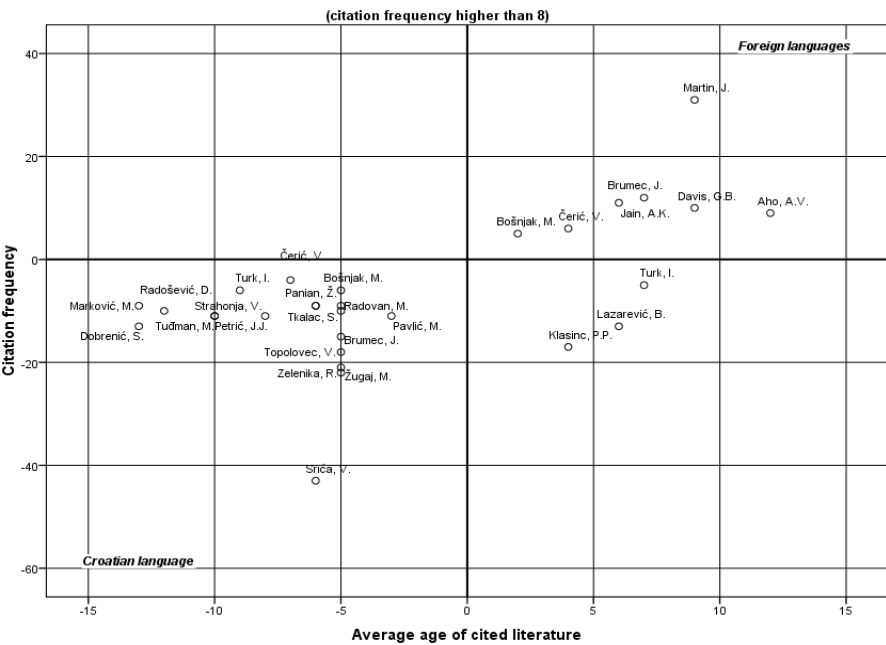
Source: self-elaboration.

Figure 3. 29 most cited authors in communicology from 1979 to 2009



Source: self-elaboration.

Figure 4. 25 most cited authors in information systems from 1980 to 2009



Source: self-elaboration.

On figures 2 to 4 cited publications published in Croatian are separated from those published in foreign languages. The fact is that cited authors make one scientific community from the standpoint of authors of doctoral dissertations. However, as there is a language barrier present¹⁶, it is not realistic to assume that a network between foreign authors and authors who publish in Croatian exists.

About Difference in Communication Models

Based on previous figures and data analysis we can notice specific patterns of behaviour that are indicated on figure no 5.

The zone of most cited authors was shown on figure no 1. It is obvious that most cited authors are cited because of their influence. Influence, by its own inner logic, is bigger if it lasts longer. That is why it does not come as a surprise that the age of authors' cited publications is older than citation half-life. This also implies another assumption: influential authors are not cited because of information, but because of the message they offer. Dominant theory that is used for solving scientific problems matters in scientific communication. In other words, members of one scientific community are bound by mutual scientific paradigm, which is defined by influential scientist. That is why we named this zone **conceptual knowledge zone**¹⁷.

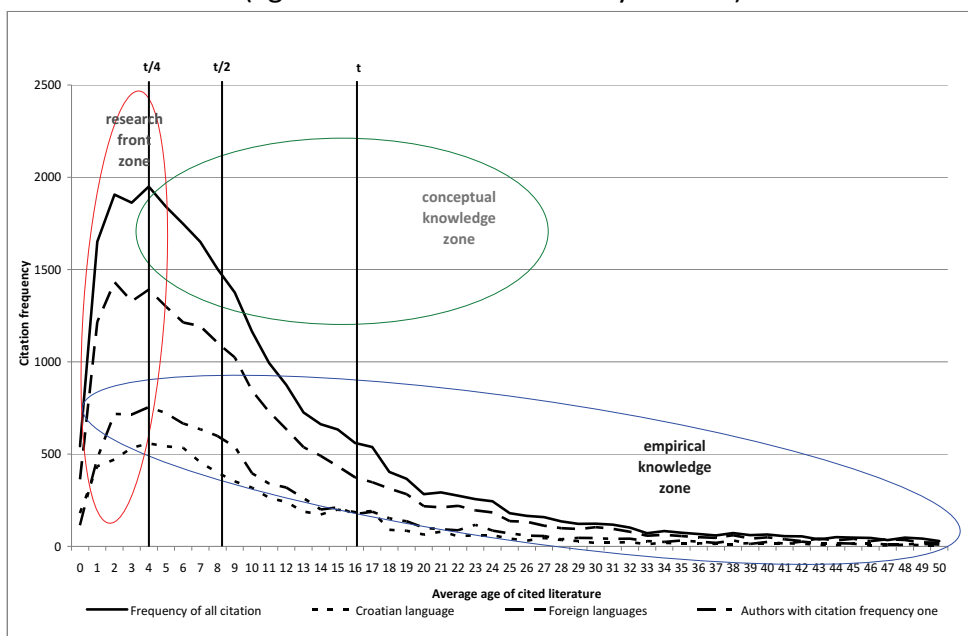
The second zone is distinguishing and regular. This zone consists of citations that are cited only once: authors and documents. Graphs 1 and 5 show their distribution which extends over the entire communication process. Furthermore, this group holds 60% (more precisely 56.7%) of citations. As these corpora of cited documents are present over the entire communication process, we named it **empirical knowledge zone**. Basic concept comes from Capurro's (2006) definition of empirical knowledge: it is information that is the result of the process of selection in communication process. The documents and authors that are cited only once have value as empirical or theoretical information. Since these citations are used only once, we assumed that their value is more empirical than conceptual.

Linear communication is dominant in these two zones. We do not think that this is exclusive or only direction of communication, but it is dominant. Among other reasons this is the dominant direction of communication because time when authors and documents are cited and time when authors published their work or when documents are published is far apart. So, when we say linear direction it means "reading" messages that originated in distant social space and time. Authors of these messages do not have the insight who is using it or who is citing it.

¹⁶ It is clear that exceptions always exist, but at this time we want to point to the existence of different communication patterns. Likewise, some authors are positioned in both areas, when their publications are published and cited in both languages, Croatian and foreign (mostly English).

¹⁷ Since doctoral dissertations were done in the time span of 30 years, we can trace the path which certain authors go through – from the source of relevant "empirical knowledge" until they become the representatives of scientific paradigm. At the same time, though, we can follow how other authors lose their influence.

Figure 5. Citation frequency and knowledge zones
(age of cited literature to 50 years old)



Source: self-elaboration.

The third zone which can be recognized from available data is **research front zone** or **personal communication zone**. This term signifies communication space and time that we marked as $t/4$, or in other words, first 25% of cited authors and documents. This zone is research front that implicates authors' understanding of the problem (and this implies routine and tacit knowledge) and communication with everyone in their surroundings relevant for the problem. This is the space of authors' bidirectional communications as a component of scientific production; whether it precedes scientific work or it is incorporated in scientific work. However, in personal communication zone empirical and conceptual knowledge overlap, they are compressed and reinterpreted. In the nature of research activities, it is typical that researched time or time of personal communication is shorter from the document citation half-time, although scientific, formal and informal communication can last much longer.

Instead of Conclusion

These three communication models, conceptual knowledge zone, empirical knowledge zone and research front zone, present an attempt to analyse process of scientific communication that we are aware of. Statistically, these models embraced most citations that are used for scientific communication. It is clear that one can find counter-arguments for proposed division, and that should be respected. However

our intention with this description – with the additional set of criteria that could describe, more precisely, existing communication differences – was to provide better understanding of scientific communication, and better understanding of bibliometric method usage in the analysis of scientific communication.

Reference

1. CAPURRO Rafael, CHAIM Zins. *Knowledge Map of Information Science. Rafael Capurro's responses to Chaim Zins* [Dokument elektroniczny]. 2006. Tryb dostępu: www.capurro.de. Stan z dnia 10.4.2009.
2. GRIFITH Belder C. (ed.). *Key papers in information science*. New York: Knowledge Industry Publications, 1980.
3. PEČARIĆ Đilda. *Razvoj informacijskih znanosti u Hrvatskoj. Bibliometrijska analiza doktorskih disertacija iz informacijskih znanosti 1978-2007*. Zagreb: doctoral dissertation, 2009.
4. PETRAK Jelka. Citati i njihova analiza. *Hrvatsko informacijsko i dokumentacijsko društvo* [Dokument elektroniczny]. 2003. Tryb dostępu: <http://www.hidd.hr/articles/citati.php>. Stan z dnia 10.4.2009.
5. SMALL Henry. Co-citation context analysis and the structure of paradigms. *Journal of documentation*. 1980, no 36, s. 183-196. ISSN 0022-0418.
6. TUĐMAN Miroslav, MILAS Milica, TUDOR-ŠILOVIĆ Neva, BORAS Damir. Bibliometric Analysis of Master Theses in Information Sciences. *Postgraduate Studies in Librarianship, Archivistics, Museology and Information Science*. 1984, 6, s. 31-40.
7. TUĐMAN Miroslav, TUDOR-ŠILOVIĆ Neva, BORAS Damir, MILAS-BRACOVIĆ Milica. A literature measure of scientific communication: Co-citation analysis of masters theses in information sciences in Yugoslavia. 1961-1984. In TUDOR-ŠILOVIĆ Neva, MIHEL Ivan (ed.). *Information research methods in library and information science*. London: Taylor Graham, 1988, s. 225-247.
8. TUĐMAN Miroslav, PEČARIĆ Đilda. Theoretical difference between impact factor and influence factor. *Journal of Information and Organizational Sciences*. 2010, no 1, s. 133-140. ISSN 1846-3312.
9. WHITE Howard D., MCCAIN Katherine W. Visualizing a discipline: An Author Co-Citation Analysis of Information Science, 1972-1995. *Journal of the American Society for Information Science*. 1998, no 4, s. 327-355. ISSN 1532-2890.