[ξ] KSZI [ξ] AKTÁK
MTAK TTO műhelytanulmányok
2012/2

Is scientometrics a 'metrics' indeed?

Wolfgang Glänzel – Bart Thijs – Lin Zhang

wolfgang.glanzel@econ.kuleuven.be

➲ http://www.mtakszi.hu/kszi_aktak/

IS SCIENTOMETRICS A 'METRICS' INDEED?

Wolfgang Glänzel[1,2], Bart Thijs[1], Lin Zhang[1,3]

[1] Centre for R&D Monitoring (ECOOM) and Dept. MSI, KU Leuven, Leuven, Belgium
[2] Dept. Science Policy & Scientometrics, LHAS, Budapest, Hungary
[3] Dept. Management & Economics, North China University of Water Conservancy and Electric Power, Zhengzhou, China

Abstract

In this paper the authors seek to answer the question of whether the field of scientometrics/bibliometrics shares essential characteristics of 'metrics' sciences. In order to achieve this objective, the citation network of seven selected metrics and their information environment is analysed.

1. Introduction

Whenever a discipline reaches a stage that requires the support of statistical methods, a *metrics* emerges from this discipline. Typical examples are biometrics (nowadays rather referred to as biostatistics), econometrics and scientometrics (including informetrics) as a subfield of information science. One should assume that all these metrics fields heavily rely on those fields they emerged from and on mathematical statistics and share those methods implying by and large close relationship. The objective of this study is to analyse if our field behaves like the other metrics and which of those are closest to scientometrics.

2. Data

Only 'citable papers' (article, note, letter, review and proceeding papers), which are indexed in the 1996-2010 annual volumes of Thomson Reuters' Web of Science (WoS) have been taken into account. Cited papers indexed in the Web of Science (1991-2010) and citing papers by the metrics journal papers (1996-2010) in a ten-year window are analysed.

3. Methods and results

Journals indexed in the Web of Science with titles containing the term [metric*] or [metrik*] have been selected and grouped into the following seven categories. Biometrics (1), chemometrics (2), econometrics (3), environmetrics (4), psychometrics (5), scientometrics & informetrics (6) and technometrics (7). Note that the latter discipline is focused on statistical methods in the physical, chemical and engineering sciences, and does not cover the technology-related part of bibliometrics.

Citation flow within and among these groups has been analysed. Furthermore, the relationship based on symmetrised cross-citation links using cosine similarities have been studied following the methods published by Zhang et al. (2009). Self-citation within the same group has been removed.

Finally, references and citations in the individual articles published in the metrics journals were assigned to ISI Subject Categories. Of course, only WoS-indexed references/citations could be taken into account. Group self-references and self-citations, respectively, ranged between about

50% in environmetrics and 99% in scientometrics. Therefore, these self-citations were excluded to avoid biases in measuring information flow and relationship.

First we looked at the cross-citation links among the groups. The network visualisation presented in Figure 1 is based on Pajek (Batagelj and Mrvar, 2002). The size of the circles is proportional to the number of papers assigned to the individual groups. The thickness of lines connecting the groups is proportional to the strength of the links. The results have struck us somewhat unexpectedly. The relative closeness of environmetrics to biometrics and technometrics to chemometrics, respectively seems to be plausible. However, we found the strong links between econometrics and biometrics as well as the relative isolation of scientometrics rather surprising.

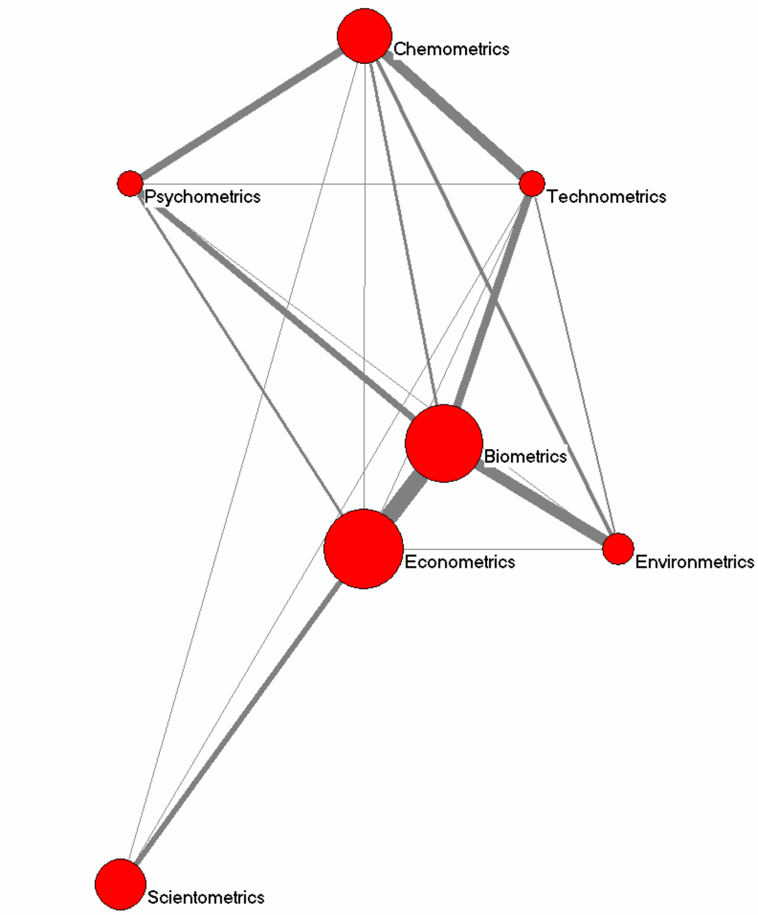

Figure 1. Cross-citation links between the metrics groups (visualisation by Pajek with Kamada–Kawai layout) [Data sourced from Thomson Reuters Web of Knowledge]

At this point it seemed to be logical and necessary, as a second step, to have a closer look at the direct references and citations among these groups. The results are shown in Figure 2. Fields numbers 1 to 7 denote biometrics, chemometrics, econometrics, environmetrics, psychometrics, scientometrics & informetrics and technometrics, respectively. The main diagonal is empty since self-citations and -references were excluded. Some of the links proved to be symmetric. This applies to the rather strong link between econometrics and biometrics but also to the somewhat weaker ones, for instance, between biometrics and environmetrics, biometrics and technometrics, and econometrics and psychometrics. The link between scientometrics and econometrics is obviously unirectional. Also the asymmetry in the relationship between chemometrics and technometrics is worth mentioning.

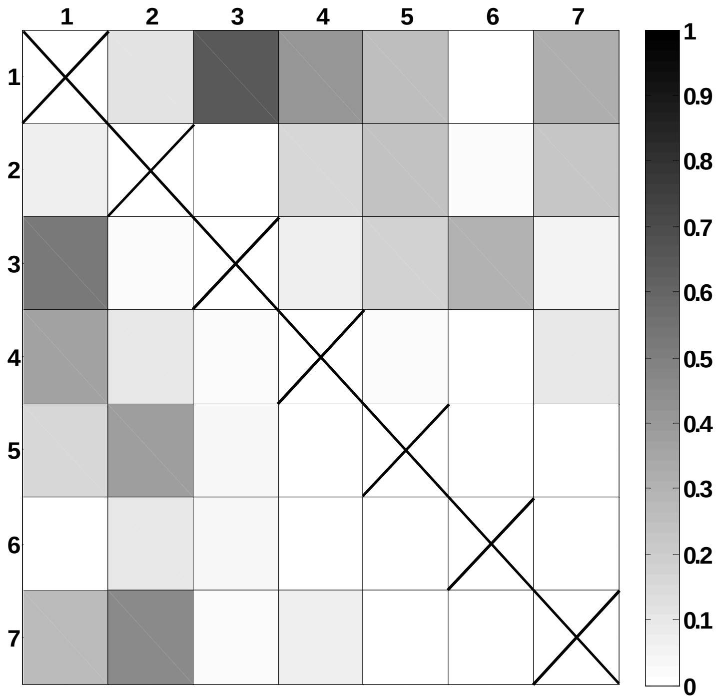

Figure 2: Strength of references and citations among the metrics groups (citing groups on the horizontal, cited groups on the vertical axis) [Data sourced from Thomson Reuters Web of Knowledge]

Table 1 shows the most important information sources outside each metrics group on the basis of ISI Subject Categories. Since journal assignment to these disciplines is not unique, figures cannot be summed up to the total. Subject categories are ranked in descending order by their share in the references in each group. Also here we find some interesting results. Metrics fields are expected to cite besides their "mother field" and closely related fields also mathematical subdisciplines, notably statistics and probability. While groups 1–5 and 7 by and large follow this pattern, scientometrics relies besides its "mother field" (information science) rather upon computer science and multidisciplinary journals. Above all, biometrics and technometrics relied largely on statistics & probability. Besides scientometrics, which is hardly rooted in mathematical subdisciplines, also chemometrics has relatively less background of mathematics or statistics. Computer science, instead of mathematical methodology, plays a more important role in the latter two metrics.

Table 1. Top 5 ISI subject categories as information "sources" for metrics journal groups

Metrics	ISI Field (Cited)	% of Refs.
Biometrics	Statistics & Probability	61.6%
Biometrics	Public, Environmental & Occupational Health	15.5%
Biometrics	Medical Informatics	11.5%
Biometrics	Medicine, Research & Experimental	11.2%
Biometrics	Genetics & Heredity	6.6%
Chemometrics	Chemistry, Analytical	31.2%
Chemometrics	Computer Science, Interdisciplinary Applications	8.1%
Chemometrics	Biochemical Research Methods	7.1%
Chemometrics	Statistics & Probability	6.9%
Chemometrics	Engineering, Chemical	6.3%
Econometrics	Economics	57.1%
Econometrics	Statistics & Probability	30.1%
Econometrics	Social Sciences, Mathematical Methods	15.1%
Econometrics	Business, Finance	13.6%
Econometrics	Mathematics, Interdisciplinary Applications	2.4%
Environmetrics	Statistics & Probability	39.1%
Environmetrics	Environmental Sciences	17.6%
Environmetrics	Meteorology & Atmospheric Sciences	10.4%
Environmetrics	Public, Environmental & Occupational Health	9.8%
Environmetrics	Ecology	5.7%
Psychometrics	Statistics & Probability	41.1%
Psychometrics	Psychology, Mathematical	24.7%
Psychometrics	Social Sciences, Mathematical Methods	23.9%
Psychometrics	Mathematics, Interdisciplinary Applications	18.8%
Psychometrics	Psychology, Experimental	13.0%
Scientometrics	Information Science & Library Science	31.8%
Scientometrics	Computer Science, Information Systems	22.6%
Scientometrics	Multidisciplinary Sciences	13.4%
Scientometrics	Management	12.7%
Scientometrics	Planning & Development	9.8%
Technometrics	Statistics & Probability	63.1%
Technometrics	Operations Research & Management Science	14.2%
Technometrics	Engineering, Industrial	12.5%
Technometrics	Engineering, Electrical & Electronic	6.3%
Technometrics	Mathematics, Interdisciplinary Applications	4.1%

Due to the interdisciplinarity of these disciplines, most of the metrics journals have multiple assignments to different WoS subject categories. For instance, the two journals in chemometrics, *Journal of Chemometrics* and *Chemometrics and Intelligent Laboratory Systems*, are assigned to six subject categories each, namely, to automation & control systems; chemistry, analytical; computer science, artificial intelligence; instruments & instrumentation; mathematics, interdisciplinary applications; statistics & probability. The question arises, then, of whether the subject assignments is properly reflected by the subject assignment of the source literature of these metrics journals. Using biometrics as an example, the subject assignments of the corresponding journals in the Web of Science are biology, mathematical & computational biology, and statistics & probability, respectively. However, after excluding the group self-citations, we do not find biology or mathematical & computational biology in the source literature, and instead, public, environmental & occupational health, medical informatics and some other medical related disciplines appear as the most important information sources besides statistics & probability. As to chemometrics, we found besides the disciplines chemistry, analytical and statistics & probability also computer science, interdisciplinary applications;

biochemical research methods and engineering, chemical that seems to be more relevant information sources than those subject categories (automation & control systems and instruments & instrumentation) to which the chemometrics journals were assigned. The observed deviations of the source literature from the actual subject assignment might also reflect some trends in the research profiles of the corresponding metrics.

Table 2. Top 5 ISI subject categories as information “targets” for metrics journal groups

Metrics	**ISI Field (Citing)**	**% of Citations**
Biometrics	Statistics & Probability	53.3%
Biometrics	Public, Environmental & Occupational Health	13.6%
Biometrics	Medical Informatics	11.0%
Biometrics	Mathematical & Computational Biology	10.7%
Biometrics	Medicine, Research & Experimental	10.1%
Chemometrics	Chemistry, Analytical	38.6%
Chemometrics	Biochemical Research Methods	9.0%
Chemometrics	Chemistry, Multidisciplinary	8.6%
Chemometrics	Spectroscopy	8.3%
Chemometrics	Engineering, Chemical	7.6%
Econometrics	Economics	61.9%
Econometrics	Statistics & Probability	18.1%
Econometrics	Business, Finance	10.6%
Econometrics	Social Sciences, Mathematical Methods	10.2%
Econometrics	Mathematics, Interdisciplinary Applications	4.3%
Environmetrics	Environmental Sciences	27.0%
Environmetrics	Statistics & Probability	25.8%
Environmetrics	Ecology	10.2%
Environmetrics	Public, Environmental & Occupational Health	8.8%
Environmetrics	Meteorology & Atmospheric Sciences	7.9%
Psychometrics	Statistics & Probability	27.7%
Psychometrics	Psychology, Mathematical	25.5%
Psychometrics	Mathematics, Interdisciplinary Applications	22.5%
Psychometrics	Social Sciences, Mathematical Methods	22.2%
Psychometrics	Psychology, Experimental	13.2%
Scientometrics	Information Science & Library Science	47.8%
Scientometrics	Computer Science, Information Systems	30.5%
Scientometrics	Management	6.4%
Scientometrics	Planning & Development	4.6%
Scientometrics	Multidisciplinary Sciences	2.8%
Technometrics	Statistics & Probability	54.3%
Technometrics	Operations Research & Management Science	21.7%
Technometrics	Engineering, Industrial	17.9%
Technometrics	Computer Science, Interdisciplinary Applications	8.7%
Technometrics	Engineering, Multidisciplinary	5.3%

In a third step, a similar analysis was applied to the metrics groups again, however, this time to the opposite direction of citation flow. In this step we were interested to learn how knowledge from these metrics was diffused to other subjects. Table 4 presents the top five subject categories citing the metrics journal groups, where, as before, the group self-citations were ignored. Statistics & probability, appears not only the main information source for most metrics, but also one of the most important disciplines to which the knowledge was transferred. Just like the case in information sources, we found one exceptional case, namely scientometrics, that had hardly any close relationship with statistics & probability as reflected by citations of either direction. Biometrics, again, has closer relationship with public, environmental & occupational health, medical informatics and other medical related disciplines. Most users of bibliometric studies and indicators are apparently active researchers in the life sciences. The direct comparison of information "sources" and "targets" and their asymmetries provides further instructive information. For instance, computer science was found to be an important information source for chemometrics, while on the other side of knowledge diffusion, spectroscopy appeared as one of the most related subjects.

4. Conclusions

A similarity of cognitive patterns was found in most of the studied metrics fields. Of course, multiple assignment of cited journals in closely related fields might here distort the picture and result in biases. A closer look at cited journals, however, reveal that journals in mathematical statistics (e.g., *Journal of the American Statistical Association* and *Annals of Statistics*) are among the most cited journals. Only scientometrics deviated from these patterns by being relatively isolated in the cross-citation network and by somewhat neglecting mathematics a methodological source. Instead of sourcing in mathematical statistics, some multidisciplinary journals including *Nature*, *Science*, *Proceedings of the National Academy of Sciences of the United States of America* serve as important information sources for scientometrics. Within the metrics journal groups, the only noticeable link that scientometric has established with another 'metric' is with econometrics.

Acknowledgement

This paper is an extended version of a poster presentation prepared for the 17th International Conference on Science and Technology Indicators (STI), 5-8 September, 2012 in Montreal, Quebec, Canada.
Lin Zhang would like to acknowledge the support from the National Natural Science Foundation of China under Grant 71103064.

References

Batagelj, V., Mrvar, A. (2002), Pajek – Analysis and visualization of large networks. *Graph Drawing*, 2265, 477–478.

Zhang, L., Glänzel, W., Liang, L. (2009), Tracing the role of individual journals in a cross-citation network based on different indicators. *Scientometrics*, 81 (3), 821–838.