KSZI [ξ] AKTÁK
MTA KIK TTO *Műhelytanulmányok*
2014/1

ANALYSIS OF CO-AUTHORSHIP PATTERNS AT THE INDIVIDUAL LEVEL

Wolfgang Glänzel

ANALYSIS OF CO-AUTHORSHIP PATTERNS AT THE INDIVIDUAL LEVEL

Wolfgang Glänzel[a,b]

[a]*KU Leuven, Centre for R&D Monitoring and Dept. MSI, Leuven, Belgium*
[b]*Library of the Hungarian Academy of Sciences, Dept. Science Policy & Scientometrics, Budapest, Hungary*

Abstract. Publication activity, citation impact and communication patterns, in general, change in the course of a scientists career. Mobility and radical changes in a scientists' research environment or profile are among the most spectacular factors that have effect on individual collaboration patterns. Although bibliometrics at this level should be applied with the utmost care, characteristic patterns of research collaboration of individual scientists and its change in the course of the career can be well depicted using bibliometric methods. A bundle of indicators and network tools are chosen to follow up the evolution and to visualise and to quantify collaboration and performance profiles of individual researchers. These methods are, however, designed to supplement expert-opinion based on other qualitative assessment, and should not be used as stand-alone evaluation tools.

1 INTRODUCTION

The evolution from "little scientometrics" to "big scientometrics" (Glänzel and Schoepflin, 1994) is characterised by two cardinal signs (Glänzel and Wouters, 2013).

In the last quarter of the 20th century, bibliometrics evolved from a sub-discipline of library and information science to an instrument for research evaluation and benchmarking called "perspective shift" (Glänzel, 2006, Wouters, 2013). As a consequence of this perspective shift, new fields of applications and challenges opened to bibliometrics, although many tools were still designed for use in the context of scientific information, information retrieval and librarianship. In other words, these became used in a context for which they were not designed (e.g., the Journal Impact Factor).

Secondly, due to the dynamics in the evaluation of research, the focus has shifted away from macro studies towards meso and micro studies of both actors and topics. More recently, the evaluation of research teams and individual scientists has become a central issue in services based on bibliometric data.

The rapid development of information technology opened bibliometrics to a broader audience. Both passive "consumers" in science policy, research management and the scientific community as well as active users and "semi-professionals" producing bibliometric indicators for various purposes gained access to the necessary data and tools. Above all, electronic communication and the Web have paved the way for some kind of democratisation of bibliometrics resulting in a rather vulgar version of democracy with anarchistic features (Glänzel and Hornbostel, 2011). This way bibliometrics has become available to practically any user, notably at the micro level.

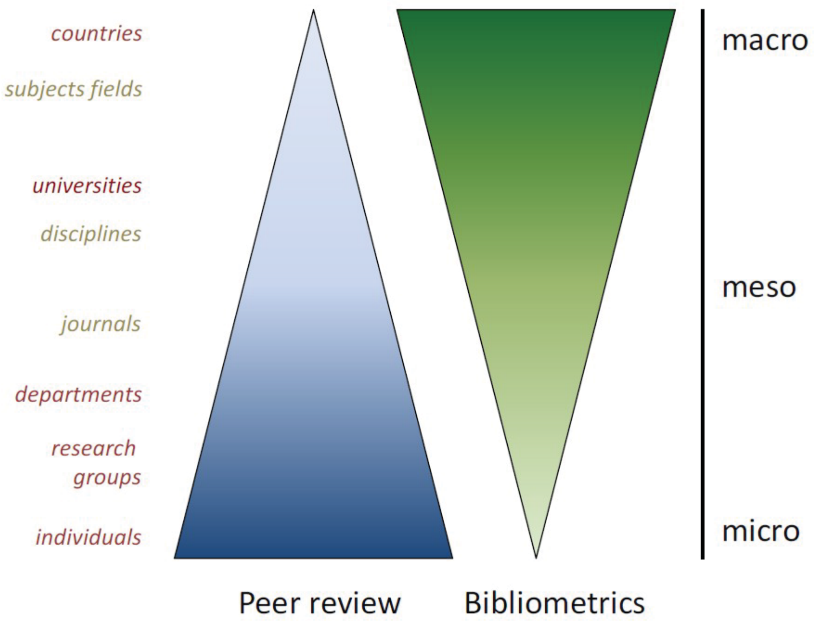

FIGURE 1. The weight of qualitative (peer evaluation) and quantitative (bibliometrics) methods as function of the aggregation level according to Glänzel and Hinze (2011)

While bibliometric macro and meso data still preserve a certain extent of anonymity, micro-level data call a spade a spade. Researchers have thus become more susceptible to the consequences of bibliometric practice since they are more and more concerned by policy use and misuse of bibliometric methods (Glänzel and Debackere, 2003). Sometimes they feel even victims of the evaluation. Bibliometric techniques should therefore always be used in a proper context, notably in combination with "qualitative methods" and special caution is always called for at this level. Figure 1 illustrates the weight of qualitative and quantitative methods in research evaluation at different levels of aggregation. Recently Glänzel and Wouters (2013) have formulated 20 recommendations for bibliometrics ("The dos and don'ts in individual-level bibliometrics"). In particular, Glänzel and Wouters recommended to use individual level bibliometrics always on the basis of the particular research portfolio of the relevant researcher. The best method to do this may be the design of individual researchers profiles combining bibliometrics with qualitative information about careers and working contexts.

As to the quantitative component of research assessment at this level, bibliometrics can be used to zoom in on a scientist's career. Here the evolution of publication activity, citation impact, mobility and changing collaboration patterns can be monitored. It is not easy to quantify the observations and the purpose is not to build indicators for possible comparison but to use bibliometric data to visually and numerically depict important aspects of the progress of a scientist's career. In what follows, I will focus on scientists' publication activity, their co-authorship patterns and the citation impact at different stages of their career.

2 METHODS AND RESULTS

2.1 *Bibliometric career analysis of individual scientists*

Although bibliometrics at the level of individuals should be applied with the utmost caution, characteristic patterns in a scientist's career can be well depicted with bibliometric methods. These methods refer to the following topics.

- Communication patterns, in general, publication activity and citation impact, in particular, change in the course of a scientists career.
- Mobility, promotion or a change in a scientist's research environment, usually results in structural changes of collaboration patterns as well.

According to the recommendations by Glänzel and Wouters (2013) the combination of bibliometrics with career analysis is one of the opportunities of quantitative science studies at the individual level. This, of course, requires the assessment on the basis of a scientist's complete oeuvre. In this context bibliometrics can be used to zoom in on various stages and phases in a scientist's career. Here the evolution of publication activity, citation impact, mobility and changing collaboration patterns can be monitored, as has been mentioned above. Some preliminary results have recently been published by Zhang and Glänzel (2012). Here I will present two examples that can be used to analyse productivity and impact patterns at different stages of a career.

The easiest way is certainly the application of *age pyramids*. This idea goes back to demographics, where population structure, composition and age of human population is quantitatively described. The population pyramid is actually an elementary tool to reflect the age structure and the growth characteristics of a given population. Here the age distribution in a human population is shown by gender in a double bar diagram: male age groups are plotted against the corresponding female groups. Usually about 5–7 paradigmatic shapes are distinguished. These reflect different paradigmatic types of growth characteristics of a given population, including expanding, stationary and contracting population models. Typical shapes are

- triangle, pagoda and bell shape (growth with high fertility and different extent of infant mortality),
- beehive shape (stationary structure with low infant mortality),
- "onion" shape (reflecting superannuation of the population).

Analogously, "age pyramids" using double bar diagrams of publication activity (at the given time period) and citation impact (based on citations received in that time period to all previously published papers) reflect important changes in the course of a scientist's career (cf. Zhang and Glänzel, 2012). Unlike in population data, the above basic types often differ for publications and citations in the bibliometric case. For instance, triangle-shaped productivity might be contrasted by an onion-shaped citation impact. In this context I would like to stress that bibliometric age pyramids reflect both subject specific peculiarities and individual "performance" patterns. This is shown using the example from four selected scientists active in different of the life sciences, natural sciences, mathematics and the social sciences. The shape of the age distribution in the natural sciences and in mathematics is expected to be more stretched and the absolute frequencies

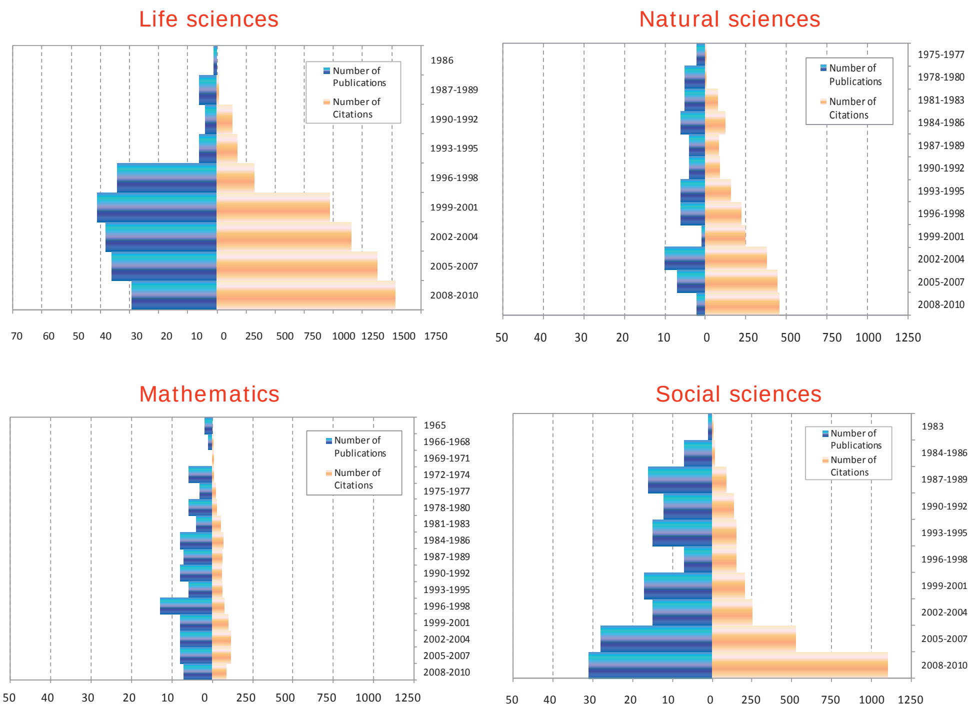

FIGURE 2. Scientometric age pyramids for four scientists according to Zhang and Glänzel (2012) (Data sourced from Thomson Reuters Web of Knowledge)

are lower than in the life sciences and the social sciences. Besides these general patterns we find interesting individual characteristics in the pyramid shapes of the four authors (see Figure 2). While the beehive in the case of the citation impact of the second author and the onion in the third case generally mirror the corresponding shapes of publication age, the onion shape of publication age is contrasted by citation triangle in the first case. The patterns for the last authors reflect a steady growth of bots productivity and impact. Some reasons for deviation productivity/impact patterns have been discussed by Zhang and Glänzel. The further analysis and interpretation of these shapes might reveal details on the relationship between the impact of recent and former research. This can be deepened by analysing the constitution of an author's highly cited papers over time. This idea goes back to the h-index sequence proposed by Liang (2006) to measure the dynamics of the h-index in a scientists career. This idea has been extended by Zhang and Glänzel (2012) to the mean age of publications of the h-core sequence, where the h-core sequence is defined analogously to the h-core for the h-index sequence. To calculate the mean age sequence of the h-core, we proceed as follows.

- The h-core is formed by those papers that have received at least h citations, where h denote the actual value of an h-index.

- The h-core sequence: we first calculated the h-index for papers published in the first year of their career, then the first two years, the first three years, and so on until the most recent year is reached.
- The mean age of publications of this h-core sequence is calculated, which expresses whether the more recent or the older publications are predominant in the respective h-core.

According to Zhang and Glänzel (2012) we can distinguish four paradigmatic patterns with case 1 representing the standard situation. A convex shape stands for "superannuation" of citation impact (i.e., mainly old papers are cited), while a concave age curve reflects an increasing number of recent papers entering the h-core.

1. A linear shape of the mean age of the h-cores plotted against time reflects steady growth of the age of most cited publications.
2. A convex shape reflects accelerated growing age of the most cited papers. This means that the "top" papers were rather published in earlier stages of the scientist's career.
3. A concave shape reflects decreasing age of highly cited papers, that is, recent papers by the author are the more cited ones.
4. "Indefinite" shape. This covers all cases not listed above.

The mean age sequence of the h-core for the same four selected scientists are plotted in Figure 3. The shape for authors #2 might be considered to be in line with the standard (linear curve), while the age sequence of author #4 corresponds to the concave case. The fluctuations in the 1970s (author #3) bear witness to quite dramatic changes in the constitution of his/her h-core. Publication activity, citation impact and their change in time are, however, not independent of collaboration and team work. Number and role of co-authors, frequency of co-authorship might strongly affect bibliometric indicators. A logical consequence is therefore the extension of individual-level bibliometrics to the analysis of collaboration patterns. This will be done in the following subsection.

2.2 *Co-authorship patterns of individual scientists*

In order to quantify the connectedness, Zhang and Glänzel (2012) have analysed "the extent of co-authorship", which denotes the number of different co-authors of the scientist under study. The objective was to analyse whether the changing size of cooperation has any influence on the authors' productivity. More recently some new indexes have been introduced to quantify this effect. The first one was proposed by Hirsch (2010). He proposed a new index called ("non-self-consistent") $\hbar$ to characterise the scientific output of a researcher that takes into account the effect of multiple co-authorship. According to its definition *A paper belongs to the $\hbar$ core of a scientist if it has $\geq \hbar$ citations and in addition belongs to the h-core of each of the co-authors of the paper.*

Schubert (2012) goes a step further. He defined a Hirsch-type index to characterise "partnership" in scientific publication output: *An actor is said to have a partnership ability index φ, if with φ of his/her n partners had at least φ joint actions each, and with the other $(n - \varphi)$ partners had no more than φ joint actions each.*

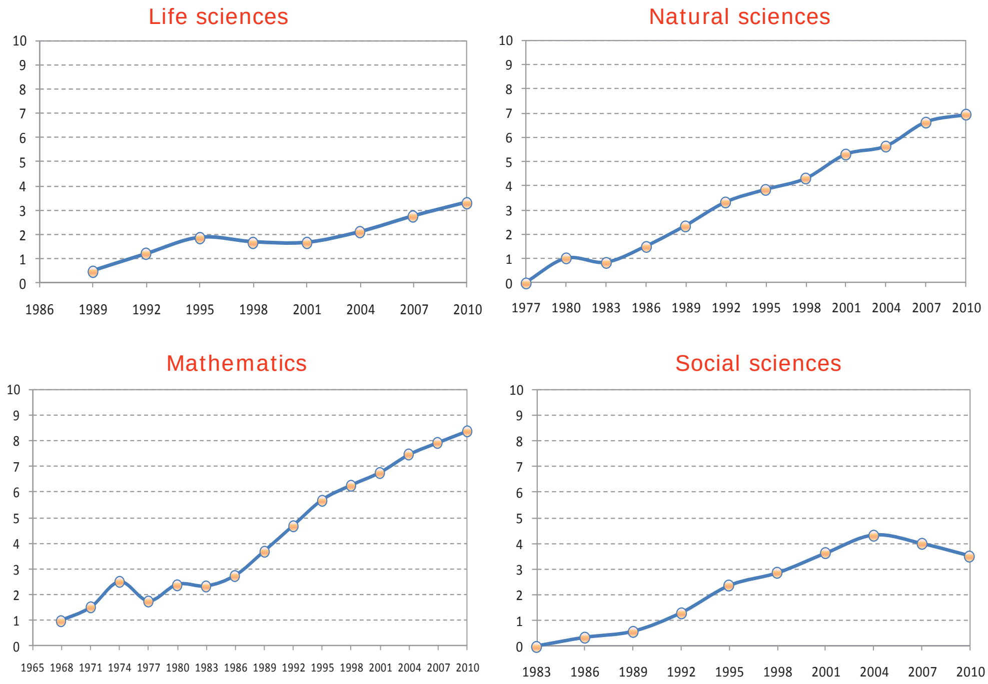

FIGURE 3. Mean age sequence of the h-core for four scientists according to Zhang and Glänzel (2012) (Data sourced from Thomson Reuters Web of Knowledge)

Similarly to the h-index that combines publication activity with citation impact, the φ-index combines two important features: publication activity with the frequency of joint activity. Some basic properties of the φ index are listed below.

- $\varphi = 0$: The author has only single-authored papers.
- $\varphi = 1$: Three cases are possible.

    1. The author has only double-authored papers with the very same co-author each (*monogamy*).
    2. The author had an arbitrary number of co-authored papers with no co-authors occurring more than once (*total promiscuity*).
    3. The author has an arbitrary number of double-authored papers with the same co-author and an arbitrary number of co-authored papers with other authors such that no co-author occurs more than once (Rousseau, 2012).

- $\varphi > 1$: In all other cases.

According to Schubert, low values reflect a scanty or inconsistent set of co-authors, while high values reflect a wide and persistent co-authorship network.

TABLE 1. Hirsch-type indicators to characterise co-authorship
(Data sourced from Thomson Reuters Web of Knowledge and retrieved in September 2013)

Author	*N*	*h*	single (all)		single (h-core)		co-authors			φ
			count	%	count	%	all	h-core	%	
A	200	32	34	17.0	6	18.8	27	14	51.9	9
B	167	30	45	26.9	1	3.3	52	25	48.1	5
D	34	11	0	0.0	0	0.0	16	9	56.3	4
E	418	44	1	0.0	0	0.0	563	169	30.0	15
G	66	19	3	4.5	1	5.3	53	25	47.2	6
J	18	8	4	22.2	0	0.0	10	7	70.0	2
K	8	4	1	12.5	0	0.0	3	3	100.0	2

After having introduced these new measures we will use Hirsch-type indices along with egocentric networks to shed some light on collaboration patterns in a scientist's career. The aim is to supplement bibliometric indicators at this level to show in how far the author's performance is related to the own and the colleagues' activity, position and impact. In order to analyse the scientists' position among their collaborators and co-authors the following set of indicators is used.

- Number of papers and h-index
- Number and share of single authored papers in all papers
- Number and share of single authored papers in the h-core
- Share of h-core co-authors
- φ index

13 collaborating authors have been chosen for illustration. The authors are anonymised and denoted by the capital letters A – M, where author "A" is chosen for the egocentric network model. Using the above-mentioned indicators in conjunction with network analysis, among others, the following questions can be answered.

- Do authors preferably work alone, in stable teams, or do they rather prefer occasional collaboration?
- Who are the collaborators and are the scientists rather 'junior', 'peers' or 'senior' partners in these relationships?

Answering these questions might help understand the scientist's own role and position in his/her research environment.

The indicator values for seven of the 13 authors can be found in Table 1. Since all authors are to a certain extent collaborators, one cannot expect serious subject-based biasses. All characteristics are therefore due to performance, (academic) age and position. Authors "A", "E" and "B" are clearly the "seniors" among the selected collaborating scientists. The number and share of co-authors as well as their share in the h-core provide important information about the co-authors' role in producing research output and in highly cited papers, in particular. The comparison of

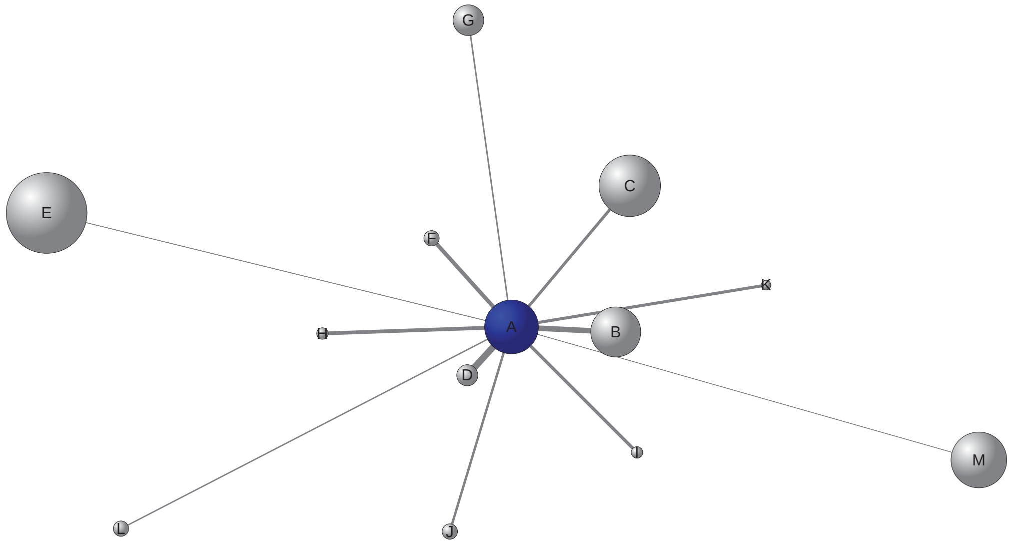

FIGURE 4. Egocentric network of an author allowing conclusions on the role of co-authors using Pajek with Kamada-Kawai layout
(Data sourced from Thomson Reuters Web of Knowledge and retrieved in September 2013)

the corresponding indicators values of author "E" and "J" allows the conclusion that above all the co-authors of "J" are responsible for "J's" high-impact papers. Usually about 50% of the co-authors contribute to high-impact papers. This alone does not point to continuous research in stable teams. The noteworthily large number of E's co-authors, which even exceeds the number of his papers, might just serve as a counter example. Collaboration with "E" is, however, not merely occasional as his large φ-values index substantiates. There is another remarkable detail: φ-values do not necessarily correlate with the number of co-authors. The comparison of "A's" large φ-index with those of "B" and "G" shows that a greater index value might be associated with a lower number of co-authors.

The above observations can be deepened by the analysis of bilateral collaboration links. Figure 4 shows the egocentric network from the viewpoint of author "A". Here all 13 selected co-authors are displayed. The size of the circles is proportional to the corresponding scientists' publication output, the thickness of lines corresponds to the strength of co-authorship links. According to publication output, "B" and "M" can therefore be considered 'peers' with respect to "A", while "D", "F", "H", "I", "J", "K", "L" are his "junior" collaborators. By contrast, "E" and "C" can be considered "seniors". Strong links with "juniors" often point to the role of supervisor and, indeed, "D", "H" and "K" were PhD students of "A". The strong link with peer "B", however, reflects long-time collaboration in a stable team and the weak link with "senior" "D", finally, reveals occasional co-author relationship.

Both exercises proposed in this subsection, the co-authorship related indicators and the network analysis, provide details that complement the "demographic" indicators described in Section 2.1 by shedding light on the scientist's position in the network of scholarly communication. A dynamic approach to capture the evolution of the scientist's position is also possible.

3 CONCLUSIONS

Bibliometric indicators and network analysis provide valuable information on the performance of individual scientists. However, this information should be considered supplementary. The emphasis in individual research assessment should always be laid on 'qualitative' methods. The added value of bibliometrics in individual-level evaluation depends on how and in what context bibliometrics is applied. The advantage of obtaining scores and numeric values, of repeating the exercise years later using the same methods, and of having the opportunity of monitoring the change of values over time should, of course, not be underestimated. Bibliometric in-depth analysis of the *evolution of publication activity and citation impact* in the course of a scientist's career can also help interpret bibliometric standard indicators at this level.

Co-authorship analysis can be used to determine the position of an author in the collaboration network and might provide important information on the scientists' own contribution to the research output reported in their CVs. In conjunction with the h-core analysis this reveals details on the extent of the scientist's real contribution to his/her research output and the citation impact these publications have achieved.

ACKNOWLEDGEMENTS

The paper is partially based on a lecture delivered at the *9th International Conference on Webometrics, Informetrics and Scientometrics and 14th COLLNET Meeting*, held in Tartu (Estonia) on August 15–17, 2013.

*

Figures 2 and 3 are reproduced from Zhang and Glänzel (2012) with permission of the publisher.

*

The author would like to thank Sarah Heeffer (ECOOM/KU Leuven) for her contribution to collecting, cleaning and processing the underlying bibliographic data.

REFERENCES

Glänzel, W., Schoepflin, U. (1994), Little Scientometrics – Big Scientometrics…and Beyond, *Scientometrics*, 30 (2-3), 375–384.

Glänzel, W., Debackere, K. (2003), On the opportunities and limitations in using bibliometric indicators in a policy relevant context, In: R. Ball (Ed.), Bibliometric Analysis in Science and Research: Applications, Benefits and Limitations, Forschungszentrum Jülich (Germany), 225–236.

Glänzel, W., *The "perspective shift" in bibliometrics and its consequences.* Keynote presented at the I. International Conference on Multidisciplinary Information Sciences & Technologies (InSciT2006): "Current Research in Information Sciences and Technologies", held in Mérida, Spain, 25-28 October 2006.

Glänzel, W., Hornbostel, S. (2011), *History and institutionalisation of scientometrics.* European Summer School for Scientometrics, Vienna, Austria, 12-16 September, 2011.

Glänzel, W., Hinze, S. (2011), *Metrics for Research Evaluation: Indicators, Methods and Mathematical Foundations.* European Summer School for Scientometrics, Vienna, Austria, 12-16 September, 2011.

Glänzel, W., Wouters, P. (2013), *The dos and don'ts in individual-level bibliometrics*, Plenary session at ISSI 2013 – The 14th International Conference on Scientometrics and Informetrics, held in Vienna, Austria, 15–19 July 2013.

Hirsch, J.E. (2010), An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, 85 (3), 741–754.

Liang, L. (2006). H-index sequence and h-index matrix: Constructions and applications. *Scientometrics*, 69 (1), 153–159.

Rousseau, R. (2012), Comments on "A Hirsch-type index of co-author partnership ability". *Scientometrics*, 91 (1), 309–310.

Schubert, A. (2012), A Hirsch-type index of co-author partnership ability. *Scientometrics*, 91 (1), 303–308.

Wouters, P. *The Citation: From Culture to Infrastructure*, In: Blaise Cronin and Cassidy Sugimoto (Eds.), Next Generation Metrics: Harnessing Multidimensional Indicators of Scholarly Performance, The MIT Press, 76–116, forthcoming.

Zhang, L., Glänzel, W. (2012), Where demographics meets scientometrics: Towards a dynamic career analysis. *Scientometrics*, 91 (2), 617–630.