

Conceptualising dimensions of bibliometric assessment: From resource allocation systems to evaluative landscapes

Fredrik Åström¹ and Björn Hammarfelt²

¹*fredrik.astrom@ub.lu.se*

Lund University, Lund University Library, P.O. Box 3, 22100 Lund (Sweden)

²*bjorn.hammarfelt@hb.se*

University of Borås, Swedish School of Library and Information Science, 501 90, Borås (Sweden)

Abstract

The purpose of this paper is to discuss the conceptualisation of bibliometric analyses in terms of the levels on which they are performed, adding contextual factors to the dimension where the size of the unit being analysed is considered. Based on empirical investigations of resource allocation systems and research evaluation practices, as well as the previous literature conceptualising bibliometric analyses, a framework based on Whitley's (2000) notion of research fields as 'reputational work organisations', is discussed. The results suggest adding a contextual 'reputational dimension' to the size-based dimension distinguishing between micro-, meso- and macro-level analyses. Furthermore, we propose that 'evaluative landscapes' (Brandter, 2017) might be a fruitful approach for further analysing how complex and multifaceted landscapes of research assessment affects the individual researcher.

Introduction

In bibliometrics and other indicator based research evaluation and funding allocation practices, a distinction is typically made between micro-, meso- and macro-level analyses (Vinkler, 1988). Albeit functional, it is however, a classification that can be perceived as one-dimensional, taking into account only the size of the unit/s being analysed; not the context in which the analysis is being performed. It can be argued that contextual factors such as the purpose of the evaluation should be considered when classifying bibliometric or indicator driven analyses, as well as taking into account if the origin of evaluation criteria for a specific situation are determined by local or national institutional policies, or if they rather relate to criteria developed within specific fields of research. Conceptualising dimensions of research evaluation, bibliometric analyses and PRFS may seem as something of an esoteric exercise, but apart from ambitions towards conceptual clarity and nuance; the conceptual issues are also a part of a wider discussion on how to navigate a situation where there is a plethora of evaluation criteria coming out of various contexts related either to institutional policies or field related assessments of reputation, at the same time as the conditions for individuals or institutions to relate to – or negotiate – the criteria varies greatly, depending on for instance the level of being established within a specific system of research (Brandtner, 2017).

The aim of this paper is to discuss the classification of bibliometric analyses in terms of levels on which the analyses are performed; to attempt a conceptualisation of analytical levels taking into account contextual dimensions other than the size of the unit analysed. The conceptual discussion is in turn related to questions on: the relation between institutional and field related evaluation criteria; the possibilities of navigating within a complex system of evaluative situations; and, how this can be understood as an 'evaluative landscape' as formulated by Brandtner (2017).

Methodological considerations

This paper is a conceptual discussion of the categorisation of bibliometric analyses in terms of the classification of levels of analysis. Thus, the specific unit of analysis for this paper is the

concepts per se, and how they have been discussed in the literature. The concepts have been analysed through a close reading of literature concerning the conceptualisation of analytical levels in bibliometrics, which in turn has been contextualised through literature concerning research evaluation and PRFS practices.

The analysis of the literature on analytical levels in bibliometrics and evaluation and funding allocation practices is informed by the theoretical framework proposed by Whitley (2000), who describes research fields as ‘reputational work organisations’, with a particular focus on Whitley’s concepts ‘mutual dependency’, the extent of which researchers are dependent on colleagues and other groups to make significant contributions to the field and thereby gaining reputation and merit; and ‘reputational autonomy’, the degree of a research fields control over standards for assessing competence and performance. In addition we make use of Brandtner’s (2017) conceptualisation of ‘evaluative landscapes’ and discuss what such an approach might offer in terms of understanding bibliometric evaluation as part of a broader context of assessment.

Conceptualising levels of analysis in bibliometrics, research evaluation and PRFS

Typically, in terms of levels of application, bibliometric analyses are categorised through the size of the unit of analysis. Analyses of individual persons or research groups are considered as micro-level analysis. If the unit of analysis is for instance a department at a university, the analysis is considered to be on the meso-level. And if the unit is a higher education institution (HEI) such as a university, or even larger units such as countries or other geopolitical regions, the analysis is considered to be on the macro-level (Vinkler, 1988). This categorisation covers all kinds of bibliometric analyses, whether performed for evaluating the outcomes of a research project or the output of a university; or if performed for the purpose of allocating funds between departments at a local faculty or between HEIs in a country based on previous performance. This categorisation makes it possible to distinguish the use of bibliometrics in for instance, on one hand, the assessment of a research group’s performance when evaluating a funding proposal as a micro-level analysis; from on the other bibliometrics being used to allocate research funds between HEIs in a country as a macro-level analysis. However, it does not make it possible to distinguish the analyses as on one hand, being part of an evaluation based on criteria coming out of a specific research field, as in the case of bibliometrics being part of a peer review assessment of a funding proposal; or on the other hand, an assessment of research performance based on indicators decided upon in an institutional policy process.

Empirical investigations of bibliometrics based research evaluation practices and PRFS

In terms of PRFS, the use of bibliometrics based allocation systems is well documented, both in terms of national systems (e.g. Hicks, 2012; Jonkers & Zacharewicz, 2016), as well as the extent to which PRFS have been applied locally (Aagaard, 2015; Hammarfelt et.al, 2016). In terms of analytical levels, as defined by for instance Vinkler (1988), we find examples of PRFS operating on all levels, from local PRFS distributing resources in-between individual scholars (micro-level) and departments or faculties (meso-level), to national systems distributing funds in-between HEIs based on bibliometric analyses (macro-level). These systems all come out of institutional policy decisions; and more often than not, does not take particularities of different research fields into account, more than possibly trying to identify means of balancing field differences when developing the indicators; and thus, reflects the goals or intentions of a local institution or national research systems, rather than for instance

the goals of a research field, that to various degrees are defined either internationally or on a more local level.

However, research evaluation and decisions on access to research funds and resources are also done on individual basis by peers of the scholars/scientists or research groups; and here, indicators of different kinds are used in support of the assessment – a practice sometimes referred to as ‘informed peer review’ (Abramo & D’Angelo, 2011) or ‘citizen bibliometrics’ (Hammarfelt & Rushforth, 2017; Leydesdorff et.al, 2016). Such practices has for instance been studied through analyses of peer assessment reports ranking candidates for academic positions (Hammarfelt, 2017; Hammarfelt & Rushforth, 2017), instructions for – and assessments of – applications for the Swedish academic title of ‘docent’/Reader (Joelsson et.al, forthcoming), and through questionnaires (Hammarfelt & Haddow, 2018). What is apparent in these studies, is how the definition of assessment criteria in these situations are related to the research fields of the applicants rather than institutional contexts, even though both the assessment criteria formulated for the decision or the promotion, and the decisions made on basis of the assessments, are local institutional matter. The assessors are typically external peers of the applicant in the same field, thus incorporating assessment criteria valid for the respective research field, rather than the institutional environment. Thus, the ‘reputational’, and disciplinary, influence over resources is considerable and may come in conflict with organisational control. The assessment criteria varies between fields, how publications are assessed, and not the least in terms of the use of indicators; to what extent they are being used at all, but also what kind of indicators are important for assessing academic merit. At the same time, it also important to note differences between research fields – and how the contribution to the field by individuals is assessed – in terms of to what extent they can be considered international or local. While some fields have assessment criteria that are essentially the same independent of research environment, some fields show variation in assessment criteria depending on national research systems (Hammarfelt & Haddow, 2018).

Thus, we have two evaluative spheres. The institutional, where assessment criteria are formulated locally or nationally, related to policies and/or organisational goals; but where there also may be great variations in terms of indicators or evaluation criteria, between national and local PRFS, in-between universities, as well as between for instance faculties at individual universities. The other sphere of evaluation is the one based on peer assessment, where the criteria for assessment are primarily coming out of how academic merit is assessed within different research fields; albeit with varying degrees of local variations depending on the field. Both within and in-between these two overlapping spheres, there may be substantial variations in terms of criteria being used, some of which might even be directly conflicting. The question is how this plurality of both evaluating bodies and evaluation practices affect individuals and organisations; and how to balance various evaluation criteria? Brandtner (2017) describes the plurality as ‘evaluative landscapes’, and goes on to describe how this can have a compromising effect on the homogenising effect of evaluative practices on organisational behaviour.

Re-conceptualising levels of analysis in bibliometrics, research evaluation and PRFS

To conceptualise the evaluative and resource allocation contexts, and re-conceptualise the notion of levels of analysis in bibliometrics, the attention is turned to Whitley (2000), describing research fields as ‘reputational work organisations’, where the assessment of academic merit is an essential part in both the definition and organisation of a field. One important aspect of this is the concept of ‘mutual dependency’, that is how dependent scholars

and scientists are on their colleagues within the field for making significant contributions, including for instance the level of specificity of standards in terms of research practices and assessment criteria. Mutual dependency is described as one dimension through which research fields can be understood; and is related to another key dimension, that of ‘task uncertainty’, that is the degree of which research outcomes can be predicted based on aspects such as the formalisation of problem formulation, levels of standards in terms of methodological familiarity and so on. Another important aspect in Whitley’s framework in relation to research evaluation practices is the concept of ‘reputational autonomy, that is, to what extent a field autonomously control standards for assessing competence and performance; and following that, also the extent to which boundaries can be – or for that matter, are important to – maintain.

In Hammarfelt’s (2017) comparison between biomedicine, economic and history research, we can see how in history and economics, reputational autonomy is maintained by a strong definition of boundaries by assessing applicants based on the field where they got their PhD degree or if the applicants have published articles in journals within the field. In biomedicine on the other hand, the background of the applicant – and the boundaries of the field in that sense – are less important. Instead, the autonomy of the field is rather maintained through a low level of task uncertainty combined with a high level of mutual dependency, regardless of the research field background of the researcher, the applicant is assessed by looking at merits in terms of the ability to perform very specific tasks using a set of well-defined methods. However, whereas the reputational autonomy of economics and history research – and in this case, they are also regarded as representatives of the social sciences and humanities (SSH) more generally – is maintained through the control over field boundaries by looking at from which field the applicants are coming from, some of the actual criteria for assessment – not the least in terms of to what extent publication related indicators play a role – seem more sensitive to developments outside the own field. While a field like biomedicine have a strongly defined set of criteria to assess those contributing to the field, and criteria substantially defined within the field; in SSH, assessment criteria seem to have a stronger relation to outside factors such as PRFS on different levels, as reflected in changes in publication patterns in SSH as exemplified by for instance Sile & Vanderstraeten (2019) and Guns et.al (2019).

Field differences as interpreted through the framework of Whitley may primarily be seen as associated with evaluations and assessments done within the field, in economics and history primarily reflecting the pursuit of reputational autonomy, in biomedicine more related to a high level of mutual dependency; and to a lesser extent in relation to assessments of performance related to PRFS. However, in local PRFS at levels below the whole university, such as at faculty or department level, there are variations in what kinds of indicators being used, taking into account publication patterns and assessment criteria in different fields (Hammarfelt et.al, 2016). And even in terms of national PRFS, differences between fields are being taken into account: either by developing performance indicators that try to balance the differences by giving different weights to different publications, such as in Norway; or by trying to change publication patterns – and thus assessment criteria – by rewarding certain types of publications, as was one the original motivations for the Swedish PRFS (Nelhans, 2013). Generally, though, PRFS – and other institutional policy related evaluations – are primarily related to Whitley’s concept of reputational autonomy; by creating a system of reward and access to resources that is not coming out of assessment criteria developed within the research fields, but rather is formulated in institutional policy processes.

Thus we find – in Brandtner’s (2017) terms – an evaluation landscape with assessments being done on all levels of aggregation, from individuals to HEIs, reflecting the dimension of size of unit of assessment as categorised by for instance Vinkler (1988). But we also find the dimension of evaluation practices and assessment criteria, either coming out of institution policy processes, or those related to specific research field practices in terms of how to decide on how someone has contributed/can be expected to contribute to the development of the field. Brandtner describes evaluation landscapes as “the collectivity of evaluation practices including rankings, ratings and rewards” (Brandtner, 2017, 201), not only describing what is going on within a field, but also being “external arbiters that motivate organizations to converge toward similar behavior.” (Brandtner, 2017, 204) However, considering the wide range of variations of evaluation practices and use of different bibliometrics based indicators both in-between and within the two dimensions described here, we find what Brandtner (2017) would describe the academic evaluation landscape as characterised by plurality; where assessment criteria are less likely to inform organisational behaviour in a field. This may also be one of the reasons for why it is difficult to study the concrete effects of specific bibliometric models of evaluation (cf. Waltman 2017)

In terms of academic research in general, according to Brandtner, the influence of evaluation practices should be relatively minor. At the same time, the ‘Whitleyan’ concepts still need to be taken into account: the level of reputational autonomy and mutual dependency of a field have a strong influence on to what extent researchers within a field are dependent on being assessed by other actors within the field (with access to resources largely coming out of funding structures more or less related to the field), or if they rather are dependent on access to resources provided on an institutional basis (e.g. time for research within a faculty position). What, in Brandtner’s terms, should be an environment less influenced by evaluation practices, becomes a balancing act between various – and varying – assessment criteria.

In conclusion

This paper suggests that, in the categorisation of bibliometric analyses, research evaluation systems and PRFS, to not only direct attention to the size of the unit being analysed, but to also add a dimension taking other contextual aspects into account: the ‘reputational’ dimension: where assessment criteria are related to the organisation of different research fields. In fact, it might be argued that PRFS and other evaluation systems and assessment procedures are better understood as part of a larger ‘evaluative landscape’ (Brandtner, 2017), in which individual researchers has to navigate in a complex setting of indicators and evaluation practices. A specific tension, describes in this paper, is the one between institutional evaluation practices, and the reputation system, which is mainly governed through discipline-based collegial norms. A further complicating factor is that an institutions or individual’s specific role and position in this ‘landscape’ may render rather different interpretations of, and reactions to, assessments procedures and evaluation systems. For example, a renowned institution may respond rather differently to measurements or ranking lists compared to less established universities. Similarly, a distinguished professor might react to the demands of assessment systems and indicators rather differently compared to a young untenured researcher. Consequently, we see the navigation of a plural evaluation landscape, where there is a balancing act between evaluations on individual, organisational and national levels, as well as between institutional and reputational assessments; and where the extent to which assessment criteria coming out if research field practices or institutional policy practices influence organisational – and individual – behaviour depends on the reputational autonomy of the field, as well as the field’s level of (within-the-field) mutual dependency.

Acknowledgments

This work was supported by Riksbankens Jubileumsfond: The Swedish Foundation for the Social Sciences and Humanities (SGO14-1153:1).

References

- Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, 42(5), 725-737.
- Abramo, G., & D'Angelo, C. A. (2011). Evaluating research: from informed peer review to bibliometrics. *Scientometrics*, 87(3), 499-514.
- Brandtner, C. (2017). Putting the world in orders: Plurality in organizational evaluation. *Sociological Theory*, 35(3), 200–227.
- Guns, R., Eykens, J. & Engels, T. (2019). To what extent do successive cohorts adopt different publications patterns? Peer review, language use and publication types in the social sciences and humanities. *Frontiers in Research Metrics and Analytics*, 3:38.
- Hammarfelt, B. (2017). Recognition and reward in the academy: Valuing publication oeuvres in biomedicine, economics and history. *Aslib Journal of Information Management*, 69(5), 607-623.
- Hammarfelt, B. & Haddow, G. (2018). Conflicting measures and values: How humanities scholars in Australia and Sweden use and react to bibliometric indicators *Journal of the Association for Information Science and Technology*, 69(7), 924-935.
- Hammarfelt, B., Nelhans, G., Eklund, P. & Åström, F. (2016). The heterogeneous landscape of bibliometric indicators: Evaluating models for allocating resources at Swedish universities. *Research Evaluation*, 25(3), 292-305.
- Hammarfelt, B. & Rushforth, AD. (2017). Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation*, 26(3), 169-180.
- Hicks D. (2012). Performance-based University Research Funding Systems. *Research Policy*, 41(2), 251-261.
- Joelsson, E., Nelhans, G & Helgesson, C-F. (forthcoming). *Hur värderas publiceringsmeriter i det svenska akademiska systemet? En undersökning av värderingen av befordran till docent med särskilt fokus på betydelsen av öppen tillgång [What is the value of publications in the Swedish academic system? A study of assessments for the promotion to 'docent' with a particular focus on open access]*. Stockholm: Kungliga Biblioteket.
- Jonkers, K. & Zacharewicz, T. (2016). *Research Performance Based Funding Systems: a Comparative Assessment*. European Union: JRC Science for Policy Report (EUR 27837 EN).
- Leydesdorff, L., Wouters, P. & Bornmann, L. (2016). Professional and citizen bibliometrics: Complementarities and ambivalences in the development and use of indicators — a state-of-the-art report. *Scientometrics*, 109(3), 2129-2150.
- Nelhans, G. (2013). *Citeringens praktiker: Det vetenskapliga publicerandet som teori, metod och forskningspolitik*. Gothenburg: University of Gothenburg (Diss.).
- Sile, L. & Vanderstraeten, R. (2019). Measuring changes in publication patterns in a context of performance-based research funding systems: the case of educational research in the University of Gothenburg (2005–2014). *Scientometrics*, 118(1), 71-91.
- Waltman, L. (2017). Special section on performance-based research funding systems. *Journal of Informetrics*, 11(3), 904.
- Vinkler, P. (1988). An attempt of surveying and classifying bibliometric indicators for scientometric purposes. *Scientometrics*, 13(5-6), 239-259.
- Whitley R. (2000). *The Intellectual and Social Organization of the Sciences*. Oxford: University Press.