Journal of Informetrics 9 (2015) 1018–1033

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Towards an early-stage identification of emerging topics in science—The usability of bibliometric characteristics

Carolin Mund[a], Peter Neuhäusler[b,c,*]

[a] SAP Deutschland SE & Co. KG, Hasso-Plattner-Ring 7, 69190 Walldorf, Germany
[b] Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Strasse 48, 76139 Karlsruhe, Germany
[c] Berlin University of Technology, Chair of Innovation Economics, VWS 2, Müller Breslau-Strasse, 10623 Berlin, Germany

ARTICLE INFO

Article history:
Received 2 March 2015
Received in revised form
10 September 2015
Accepted 10 September 2015
Available online 19 November 2015

Keywords:
Bibliometrics
Emerging topics
Identification
Early-stage

ABSTRACT

The assessment of research topics according to their development stage can be used for different purposes, most importantly for decisions regarding the (financial) support of research groups and regions. In this work, we try to determine the influencing factors of emerging scientific topics during their early development stage. Documents in five predefined fields are analyzed with regard to the characteristics of the involved authors, their references and journals. With the help of an assignment to emerging and established topics, the publication behavior of documents in different development stages can be compared. Foremost, indicators can be derived that can help to identify publications in emerging topics in science at an early-stage after publication.

The results show that the field differences are so pronounced that they hamper generalization. The field specific analysis, however, suggests that at least for some fields a pre-selection of emerging topics can be made. In technical fields, the involvement of larger groups of researchers is an apparent feature, while in medicine a contrary observation could be made. In addition, for the field of engineering we found that emerging topics are more often published in older but smaller journals, which indicates a high specialization of the publications.

1. Introduction

New scientific discoveries and emerging topics[1] in science are shaping the evolution of research (cf. Kuhn, 1970: 62ff). Due to various reasons, emerging topics might or might not establish themselves as independent research fields in the course of time (see van Dalen & Klamer, 2005; Campanario, 2009; Kilwein, 1999; Benos et al., 2007). Besides structural factors like scientific and technological uncertainties, path-dependencies and lock-in effects (cf. Barber, 1961; Johnson, 2013; Stent, 1972; Stent & Hook, 2002), new findings are sometimes overlooked or rejected simply because the already established knowledge seems more intuitive or persuasive (Atkins, 2003:.205)—a reaction that is not necessarily a result of the quality or potential of the finding itself (Kilwein, 1999; Benos et al., 2007; van Raan, 2004; Costas, van Leeuwen, & van

* Corresponding author at: Fraunhofer Institute for Systems and Innovation Research ISI, Competence Center Policy – Industry – Innovation, Breslauer Str. 48, 76139 Karlsruhe, Germany. Tel.: +49 0 721 6809 335.
E-mail address: peter.neuhaeusler@isi.fraunhofer.de (P. Neuhäusler).

[1] New scientific discoveries or knowledge claims do not necessarily lead to new and innovative "topics" in science. For the ease of readability, however, we stick with the terms "established" and "emerging topics" throughout the remainder of the paper.

http://dx.doi.org/10.1016/j.joi.2015.09.004
1751-1577/

Raan, 2011). A prominent example is the original paper by Gregor Johann Mendel – and subsequent papers subsumed under the label "Mendel syndrome" – that was ignored by the scientific community because of its innovativeness or "deviation" from established facts, patterns or methods (Atkins, 2003: 47; Costas et al., 2011; van Raan, 2004).

At a more general level, the Matthew effect in science shows that scientists with an already high popularity are over-proportionally acknowledged (Merton, 1968; Cozzens, 1989; MacRoberts & MacRoberts, 1996), a fact that might also hamper the dissemination of new findings especially from still relatively unknown researchers. An additional aggravating factor is the massive increase in the amount and the accessibility of annual publications over the last decade (Michels & Schmoch, 2012; Larsen & Ins, 2010). Time constraints at the readers' side make scientific work, in particular citations (Franck, 1999), more and more superficial. Advertising or signalling effects at the authors' side can be one way to increase the chance that an emerging topic is recognized. However, signalling demands willing recipients and an emerging topic might struggle with quite small audiences, once again impeding knowledge transfer: not only are few people working on the emerging topic itself, also less people might be able to grasp its (assumed) potential in general (MacRoberts & MacRoberts, 1996) or for other fields (Urata, 1990; Steele & Stier, 2000; Rinia, van Leeuwen, Bruins, van Vuren, & van Raan, 2001). Peer reviewers might therefore also fail to acknowledge innovative papers, forcing the authors to publish in less popular or lower quality journals (for examples see Campanario, 2009), which diminishes the audience even further.

In order to countervail the effect that the full potential of an emerging topic might not unfold, "early stage pointers" are needed to avoid the oversight of innovative work and emerging topics in science. Only if an emerging topic is recognized as such, the awareness in the scientific community can be raised and decisions regarding the (financial) support of research groups and regions can be made. This might further lead to the mobilization of new sources of funding or even political actions towards the promotion of a specific topic.

Our goal therefore is to identify and test features that might help to detect publications dealing with emerging topics. We focus on indicators that are computable directly after publication. One underlying assumption is that the publication process is shaped by internal and external factors. These factors differ for publications in emerging topics and those in established topics. Thus, the indicators for an emerging topic are first and foremost deviations from the publishing "norm". They can be forced upon the respective publications if review or writing processes make it necessary to publish with certain co-authors from specific countries, in certain kinds of journals or with reference to specific former work (cf. MacRoberts & MacRoberts, 1996). Due to this "forced publication behavior" it is possible to detect publications in emerging topics by these tell-tale characteristics.

The paper is structured as follows. In Section 2 we develop our theoretical arguments for the specific characteristics of publications dealing with emerging topics in science. Section 3 presents the data and describes the variables and methods used for our analyses. The descriptive and multivariate results are provided in Section 4. Finally, in Section 5 we derive our conclusions and discuss the implications of our findings.

2. Theory & hypotheses

In bibliometric studies, citations have established themselves as an indirect indicator for a paper's quality and its usefulness in particular (Garfield, 1979). They have been applied to assess the scientific landscape and its development in retrospect (see e.g. Small, 2006) and have also been applied to identify emerging topics in science (see e.g. Price, 1965; Small & Upham, 2009; Kajikawa & Takeda, 2009; Shibata, Kajikawa, Takeda, & Matsushima, 2009a; Shibata, Kajikawa, Takeda, Sakata, & Matsushima, 2009b). However, a very timely analysis for the identification of emerging topics in science is difficult to accomplish with citations as they necessarily introduce a time-lag between data availability and analysis (of approximately 3 years, see e.g. Rinia et al., 2001; Glänzel & Schoepflin, 1999). Thus, for the qualitative as well as temporal aspects, citations are hard to include in a system that identifies emerging topics in science at a very early stage.

In this paper, we differentiate between two types of indicators depending on whether they are caused by effects before or after the publication. They can also be differentiated on the influence the author has on them. Typically, the indicators before the time of publication are also choices of the authors, i.e. the authors select the publication outlet with the respective characteristics, their references etc. Contrary, the post-publication indicators are effects that are not under the control of the authors. They are rather circumstances of the (possibly hostile or maybe also competitive) environment in which a new topic is born. Thus, we distinguish between the emergence *sources* and the emergence *environment*. A third factor is the disciplinary scope that is both present in the emergence sources from which the innovative publication derives, as well as the factors in the environment, especially of the publishing journal. We therefore analyze the *interdisciplinarity* as an indicator and catalyst of topic emergence.

Possible impeding as well as fostering influence factors regarding the publication source, possible influences of its knowledge foundation as well as underlying collaboration will be analyzed. We are thereby able to deduce whether documents in emerging topics deviate in their bibliometric characteristics from those in established ones. This allows the inference of possible impediments or disruptive factors in the publication process for emerging topics.

2.1. Interdisciplinarity as an indicator and catalyst for innovation

One of the main sources for innovation is the combination of existing means and knowledge in a novel way. Exaptation, the misuse or adaptation of methods from other fields, is an illustrative example for innovation via combination (Johnson, 2013: