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 pre-defined 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 “topicsïn science. For the ease of readability, however, we stick with the terms “establishedänd “emerging topicsẗhroughout the remainder of the paper.

http://dx.doi.org/10.1016/j.joi.2015.09.004

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:

172). Thus, topics in a still early development stage might be characterized by a higher interdisciplinarity, which has seen an increasing trend in the last 30 years (Porter & Rafols, 2009). The combination of knowledge in turn can result in independent topics or fields (see e.g. Shafique, 2013; Alvargonzález, 2011) that can evolve in an independent way. Sometimes, this might lead again to a diminishing multidisciplinarity, which might "hinder tapping the full potential of research" in severe cases (Shafique, 2013).

The assumption that interdisciplinarity might be used as an indicator for innovation derives originally from Kuhn's definition of paradigm shifts (Kuhn, 1970: 64f). According to Kuhn (1970: 71,77), revolution in science happens when a crisis appears. Crises are evoked when present theories and methods are no longer sufficient to explain (new) observations or to fulfill the current needs (e.g. in the case of scarce resources). The solution to a crisis, a paradigm shift, can only be achieved if new theories or methods are introduced. The easiest way to do so is to be open-minded to standards in other scientific disciplines, e.g. when "scientists adopt new instruments and look in new places" (Kuhn, 1970). Thus, the transfer or adoption of knowledge across boundaries can help to turn the corner in a crisis (Thompson Klein, 2004). For example, genetic algorithms use the basic biological principles of recreation and evolutionary survival of the fittest to facilitate complex mathematical calculations.

In the context of this paper, the question arises how the interdisciplinarity of a document is measured. Mapping and evaluation of interdisciplinarity is a challenging task. Interdisciplinarity can indicate the breadth of the knowledge base of a publication or the novelty of its knowledge integration (Rafols & Meyer, 2010). Furthermore, Morillo, Bordons, & Gomez (2001) showed that the measurement of interdisciplinarity via journal categories and reference categories does not necessarily lead to similar results. We therefore test both metrics as they are not only used to measure the abstract concept of interdisciplinarity but also indirectly that of the emergence of a topic:

H1 *(:).* The chance of finding an emerging topic is on average increasing with the number of distinct scientific fields the references of a document are classified in.

H2 *(:).* The chance of finding an emerging topic in a journal is on average increasing with the number of fields a journal is classified in.

2.2. Emergence sources

Emergence sources denote crucial factors in the creation process of a publication, e.g. the authors of the respective publication as well as the available tools, methods and theories. Differences in these factors might be observable depending on the development stage of the topic. Therefore, the authors and the references of the publications are analyzed in respect to differences for emerging and established topics.

2.2.1. Number of authors and countries involved

Collaboration among scientists can be hindered by the early development stage of a topic. Scientists with new ideas might be reluctant to share their research and rather keep the idea and the expected reputation for themselves. Trust might play a more important role in emerging topics, where the involved researchers might be more cautious in regard to premature disclosure of the findings (as they give up part of the control of that knowledge, cf. Zand, 1972). Trust is especially important in risky situations (Olson & Olson, 2000), in which the disclosure of new findings – which is not supported (or driven) by a project – can be categorized. Similar observations have been made for collaboration in interdisciplinary fields (cf. Anholt, Stephen, & Copes, 2012).

In addition, knowledge exchange might be difficult in fields that are not yet properly defined. This does not only hinder the direct communication but also supportive activities as exchange programs, projects etc. To be able to write a project proposal which is accepted, either a common understanding or many preliminary studies are necessary.

Therefore, authors in emerging topics might be very conservative in respect to co-authorships. The impeded collaboration might reflect in smaller author groups. Consequently, the higher the number of authors, the smaller should be the chance for finding a new emerging topic.

H3 *(:).* The chance of finding an emerging topic is on average increasing with a decreasing number of authors.

Trust issues concerning the disclosure of new findings or methods can be higher for international collaborations, as the assessment of the partner is hindered by geographical distance. Even with the increasing use of collaboration tools that bridge this gap, trust is more often established through personal contact and/or an interchange of ideas over a long consecutive time period (cf. Handy, 1995; Rocco, 1998). Therefore, relationships that allow collaboration in such a sensitive stage of a topic are rare, especially across national borders. Enforcing the effect that already restricts collaboration among authors in general (H3), less international collaborations can be expected for emerging topics than for established ones.

In this study, the international spread of a topic is measured by the number of different countries from which the authors of a paper originate. For high values of this feature, it can be assumed that the topic is already well known across borders. This might concern in particular already established topics, as they do not have the above mentioned problems of missing trust and common ground. Emerging topics on the other hand might suffer from these issues and thus their geographical spread might be rather small.

H4 *(:)*. The chance of finding an emerging topic on average decreases for an increasing number of distinct author countries named on a document.

2.2.2. Age of references

In accordance with the saying “dwarfs standing on the shoulders of giants”, new (still small) topics might have to rely on more established, older topics. This relationship between a topic and “external” or former work can be traced by the references in the papers.

With external work, publications from other topics are denoted, which are used or adapted for the ongoing research. As Rinia et al. (2001) have shown, the citation delay for work from other disciplines is higher than for work from the same discipline. In other words, the knowledge transfer takes longer if disciplinary boundaries have to be crossed. As innovative research might make this necessary more often than traditional research, the age of the references is supposed to be higher.

Furthermore, more established, fundamental work in the own discipline might be used more often than ongoing research as the new discoveries can seldom refer to other current research issues. Thus, it can be assumed that the chance of finding a new document increases with the average age of its references.

H5 *(:)*. The chance of finding an emerging topic is on average increasing with the age of the references cited in a document.

2.3. Emergence environment

Similar to the citations, other factors prior to or following the publication process can be dependent on the acceptance of a topic. For instance, the development stage of a topic can influence its acceptance in (renowned) journals (Campanario, 2009), the opportunities for collaboration, as well as the availability of the knowledge base and thus the references. Such features allow the indirect measurement of acceptance and spread of a topic. However, they are available at the time of the publication in contrast to the citation counts. Thus, our focus lies on the identification of bibliometric characteristics that are accessible ex-ante and are derived to allow for an identification of emerging topics in real-time. To be more precise, we aim to apply exclusively bibliometric indicators that are available at the point in time when a paper is published. With the help of these indicators, documents in emerging topics could be identified, so that the respective topics can be deduced (or the documents clustered to represent these topics).

2.3.1. Size, age and JIF of journals

Chew (1991) showed that rejected publications are often resubmitted to (and finally published in) journals with a smaller size and smaller circulation paths. Furthermore, smaller journals are by tendency more specialized, while journals with a higher page count have a broader focus (cf. Michels & Schmoch, 2014). Thus, a resubmission in smaller journals can be seen as a process wherein a specialized document searches for its niche where it can be published. This might be a more important process for new emerging topics for which this niche has not yet been defined and thus is unclear for both the authors as well as the reviewers.

Turoff and Hiltz (1982) stated that both journal size as well as rejection rate have increased, which augments the demand for highly specialized journals. This is especially due to the fact that the larger journals often have to perform a balancing act among various research areas. This bears implications on the readership as well as on the set of authors which are able to publish in these journals. Most importantly, new emerging topics are unlikely to fit into the concept of a journal that aims at a broad overview of ongoing research in many topics. A publications dealing with a highly innovative topic might thus be better recognized and acknowledged in a smaller, thematically more specialized journal.

Building on the work by Chew (1991), we measure the size of a journal by the number of articles it publishes per year. The basic assumption hereby is that the more articles a journal publishes per year, the broader its focus and vice versa (see Michels & Schmoch, 2014). Good examples for journals with a broad focus are Science or Nature. In accordance with the findings of Chew (1991) and Ray, Berkwits, and Davidoff (2000), we analyze if documents dealing with new emerging topics, for which publication in general might be more difficult, are published more often in smaller (more specialized) journals. This leads to our first hypothesis.

H6 *(:)*. The chance of finding an emerging topic in a journal is on average decreasing with the size of the journal.

Similarly to the size of a journal deducing its level of specialization, the prestige or standing of a journal can be represented by the Journal Impact Factor (JIF), i.e. the number of citations it achieved divided by the number of publications published in the respective years. Its value shows how much attention articles published in the respective journal receive. This in turn reflects the readership and shows the standing of a journal. As van Dalen and Klamer (2005) pointed out the reputation of a journal can be seen as a signal for the scientific community.

It is questionable whether documents that start a new topic can be placed in journals with a high reputation, and therefore a high JIF, because the reviewers might be more critical[2] and the acceptance of new ideas lower. One of the biased reasons for rejections of papers is “avoidance of unconventional ideas” as noted by Benos et al. (2007). The Nobel prize winners

[2] The high discrepancy between reviewer scores and citation counts for conference papers has been shown by Ragone et al. (2013).

Table 1
Summary of the hypothesized effects.

Hypothesis	Feature	Influencing factor	Relationship to finding an emerging topic
H1	Fields of references	Interdisciplinarity	Positive
H2	Journal fields		Positive
H3	Number of authors	Emergence	Negative
H4	Number of countries	sources	Negative
H5	Age of references		Positive
H6	Journal size	Emergence	Negative
H7	JIF	environment	Negative
H8	Journal age		Positive

Source: Own compilation.

Hans Krebs and Barbara McClintock have been rejected in the journal Nature for their innovative papers (Kilwein, 1999; Benos et al., 2007). Benos et al. (2007) conclude that that “avoidance of avant garde and controversial topics by reviewers and editors could hamper the advance of science.” Reviewers can easier get by with such a biased reviews if the journal has many submissions (cf. H1), which is one effect of a high standing or popularity. Thus, the following hypothesis is tested in this study:

H7 *(:).* The chance of finding an emerging topic in a journal is on average decreasing with an increasing JIF.

The growth in scientific output and the spread of information makes the introduction of new, more specialized journals necessary (Turoff & Hiltz, 1982). New journals are introduced when a topic becomes so important and established that a continuation of that topic in a separate community is foreseeable. For example, the journal “International Journal of Disaster Risk Reduction” was first issued in 2012 in response to rising attention to this topic and the need for a consensus on proper definitions in the field (Alexander, 2012). However, it is evident that a topic needs to reach a certain level of establishment and dissemination so that the introduction of a new journal for it seems reasonable, i.e. the number of scientists involved and thus possible authors have reached a critical mass. Turning the argument the other way around, it is unlikely that emerging topics, i.e. topics for which only few (if any) publications already exist, can be found in newly issued journals. Therefore, documents published in younger, more recently issued journals should in the majority deal with established topics[3].

H8 *(:).* The chance of finding an emerging topic in a journal is on average increasing with the age of the journal.

To get a better overview on the hypothesized effects, Table 1 summarizes the individual features as well as their assumed relationships to emerging topics.

3. Data & methodology

A necessary precondition for testing which of the publication characteristics might help tell apart emerging from established topics is a dataset where a distinction between emerging and established topics has already been made. We therefore apply the classification of documents in emerging and established topics by the Japanese National Institute of Science and Technology Policy (NISTEP), which is based on an analysis of co-citations (cf. Saka, Igami, & Kuwahara, 2010). With the help of this classification, we are able to test the explanatory power and significance of each of the discussed characteristics with regard to identifying emerging topics as defined via the NISTEP analyses of co-citations. The characteristics that provide significant explanatory power for the differentiation between emerging and established topics can subsequently be used as a stand-alone solution to separate emerging from established topics. This enables us to generate a system of indicators for the identification of emerging topics in science, which has the clear advantage of being available at a very early-stage after a document has been published.

3.1. The data

The data employed for our analyses is based on a document collection created by NISTEP as a basis for a report on hot research topics (cf. Saka et al., 2010). In their work, Saka et al. (2010) collected the 1% most highly cited papers in 22 scientific fields for the years 2003 to 2008 from the Web of Science (WoS) database by Thomson Reuters. Only journal publications from the Science Citation Index Expanded (SCIE) and the Social Science Citation Index (SSCI) were taken into account. The resulting documents (articles and reviews) were clustered thematically on the basis of a co-citation analysis to form what Saka et al. (2010) labeled *research fronts*[4]. If a research area contained only documents from the years 2007 and 2008, Saka et al. (2010) labeled it as newly-found hot research area. Those research areas which also covered documents from former

[3] It should be noted that Dirk (1999) was not able to show a relationship between the age of a journal and the originality of published articles.

[4] The degree of co-citation for a pair of documents A and B was determined by dividing the number of papers that cited both paper A and B, NAB, by the number of papers that cited either A or B, NA + NB. All papers with a co-citation degree of at least 0.3 were clustered to form research fronts. Research fronts were clustered again by the same method with a threshold of 0.1 for the co-citation degree to form research areas.

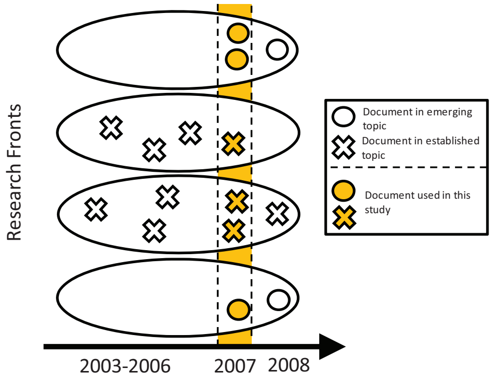

Fig. 1. Identification of emerging topics by NISTEP and usage in this study. *Source:* Own illustration based on Saka et al. (2010).

Table 2
Overview of the distribution of documents in emerging and established topics by scientific fields.

Discipline	Documents in established topics	Documents in emerging topics	Total
Engineering	356 84%	67 16%	423
Biology, environmental science & geoscience	679 92%	56 8%	735
Medicine	828 93%	64 7%	892
Chemistry	449 89%	53 11%	502
Physics, mathematics & computer science	659 92%	60 8%	719
Total	2,971 91%	300 9%	3,271

Source: Saka et al. (2010), Web of Science, own calculations.

years were classified as established topics, as their time span was longer. Fig. 1 shows four research fronts of which two, the first and the last, are classified as emerging ones since papers covering these topics were first published in 2007. The remaining two research fronts contain papers from earlier years and are thus labelled as "established topics". In this study, we used all documents from the year 2007 and applied the respective labels of belonging to an emerging or established topic from the study by Saka et al. (2010)[5]. The classification was thus transferred to the document level. Fig. 1 shows exemplarily which documents were used in our study. All relevant bibliometric information was extracted from the WoS database.

Based on this core dataset, which consists of 3271 scientific papers including their assignment to one or more of 22 scientific disciplines and the information whether a document was classified as "established" or "emerging", we collected additional data from the WoS database, which are necessary prerequisites for the calculation of our feature variables, e.g. the size of the journal in which an article had been published and the number of authors per document. Since the number of documents defined as emerging was too small for statistical inference in a fine-grained field classification, we had to aggregate the fields to a more coarse-grained classification. 19 disciplines in the original dataset were therefore aggregated to form 5 disciplines for our further analyses[6]. Table 2 provides an overview of the distribution of documents in emerging and established topics by scientific fields.

3.2. Variables & summary statistics

In this section, we briefly discuss the calculation of the feature variables used for our further analyses. The summary statistics for all variables are presented in Table 3.

Following our theoretical discussion, we use the information if the document is classified as belonging to an "established" (coded as 0) or an "emerging" (coded as 1) topic within the NISTEP database as the dichotomous dependent variable for our

[5] Due to the two year time-span between 2006 and 2008, we excluded the data for the year 2008 in our study to keep the potential errors as few as possible.
[6] The disciplines regarding the Social Sciences were dropped due to small numbers of observations even in aggregated point of view.

Table 3
Summary statistics.

Variable	Obs.	Mean	Std. dev.	Min	Max
Emerging topic	3271	0.09	0.29	0.00	1.00
Engineering	3271	0.13	0.34	0.00	1.00
Biology, environmental science & geoscience	3271	0.22	0.42	0.00	1.00
Medicine	3271	0.27	0.45	0.00	1.00
Chemistry	3271	0.15	0.36	0.00	1.00
Physics, mathematics & computer science	3271	0.22	0.41	0.00	1.00
Fields of ref.	3271	10.02	5.43	1.00	46.00
Journal fields	3271	1.72	1.20	1.00	7.00
Nr. of authors	3271	8.57	31.94	1.00	1311.00
Nr. of author countries	3271	1.76	1.65	1.00	19.00
Age of ref.	3271	5.18	2.05	0.00	15.17
Journal size	3271	1,478.57	1456.40	5.00	7266.00
JIF	3236	7.50	6.05	0.20	63.97
Journal age	3271	35.42	28.15	0.00	105.00

Source: Web of science, own calculations.

regression models. We differentiate our models by the five science fields "engineering", "biology, environmental science and geoscience", "medicine", "chemistry" and "physics, mathematics & computer science".

The features discussed in Section 2 will serve as explanatory variables. For all documents in the dataset, the following variables were calculated, which can – according to the theoretical discussion – be described as driven either by the emergence source or the environment or both in the case of the interdisciplinarity:

- Number of fields of references
- Journal fields
- Number of authors
- Number of countries of the authors
- Age of the references
- Journal size
- Journal Impact Factor (JIF)
- Journal age

Interdisciplinarity is measured via the scientific fields listed in the references and the publishing journal. The number of *fields of the references* counts all the distinct science fields the references in the document are classified in; based on the WoS classification of scientific disciplines. The frequency of the emergence of the respective fields is not taken into account. Therefore, a value of 5, for instance, might indicate 5 references to publications classified in 5 different, non-overlapping, fields or a reference to one publication assigned to 5 different fields. In any case, this value indicates the field specific diversity on which a given document is building its knowledge base. The *journal fields* variable is based on the classification of scientific disciplines in WoS. It is measured as the distinct number of scientific disciplines in which a journal is classified and thus is used to assess the interdisciplinarity of a journal or conversely its level of specialization.

The first indicator calculated for the emergence source is the *number of authors* that are named on a given document. This measure is based on the different standardized names of authors[7] in WoS and indicates the extent of collaboration that resulted in a given publication. In a similar fashion, we calculate the distinct *countries of origin of the authors* named on a publication in order to indicate if the publication is the outcome of a national or an international collaboration. For the *age of the references* of a document, the age of each reference is calculated as the difference between the year of the citation and the year of the cited publication. The values for all references of a paper are then averaged. Thus, the value indicates if the article in tendency relies rather on an older or newer knowledge base.

For the environment of the emerging topic, different features of the journal are calculated. The *size of a journal* is measured by the number of articles published in a given journal in the respective year, in our case 2007. The *JIF*, which is a citation based indicator for the evaluation of the quality of a journal, represents the prestige and reputation a journal has in the scientific community and is measured as the frequency with which articles in journals on average are cited by subsequent publications in a given period of time. In our case, the JIF is defined as the number of citations in the year 2007 divided by the number of cited publications in the period 2005–2006. The *age of the journal* is defined as the number of years since the journal first appeared in the WoS database. In theory, the first appearance in the database and the actual appearance on the market can differ. In particular, this might occur due to a belated inclusion of a journal in the Science Citation Index (SCI).

[7] Since this measure is based on standardized names, it might fail in the case of two co-authors of a single publication who share the first name initials and the last name. However, we consider this a rather rare event, which is why distortions on this indicator should be limited.

Since Thomson Reuters, however, demands the fulfillment of specific criteria of quality and quantity for the inclusion of a journal in the SCI, this can be seen as a certain selection mechanism for journals to reach a common standard[8].

3.3. Estimation method

For the multivariate analyses of the features and their power in explaining the variance between new and old research topics we employ logistic regression models as the outcome variable is dichotomous. In the logit model, the log odds of the outcome are modeled as a linear combination of the predictor variables (Long, 1997). After estimating a general model including the publication features as independent variables and controlling for field specific effects, we re-run the models for each of the five scientific fields since it can be assumed that the explanatory power of the publication features varies across disciplines.

To interpret the model coefficients, marginal effects at the means of the independent variables were calculated. They reflect the probability for a publication to fall into the "emerging topic" category as identified in the NISTEP dataset. More specifically, the marginal effect represents the effect of a one-unit change in the independent variable on the probability to belong to the "emerging topic" category of the dependent variable (coded 1), holding all other variables constant (Long & Freese, 2003). The interpretation of the coefficients is different for continuous and discrete independent variables. In the case of continuous independent variables, an infinitesimal change of the independent variable changes the probability to belong to the "emerging topic" category of the dependent variable, i.e. that the dependent variable takes the value of 1, by *X*%. For dummy variables, a change of this variable from zero to one changes the probability that the dependent variable takes the given outcome value by *X*%.

4. Empirical results

In this section, the results of our analyses will be presented. In a first step, we will provide a descriptive overview on the selected bibliometric feature variables and how well they are able to discover differences between emerging and established topics. The second step will be to test our assumptions via multivariate regression models which provide us with a more detailed overview on possible combinations of features for the early-stage identification of emerging topics.

4.1. Descriptive statistics—Box plots

Before discussing the multivariate analyses, we will first of all present a descriptive overview of the specific features and their relationship to established and emerging topics with the help of box plots (Fig. 2). The box plots show the distribution of the feature values in comparison. For the purpose of these graphs, outliers were excluded in order to allow for an easier visual comparison of the focal features of the distributions.

It is evident from the box plots that we do not find overly large differences between the two types of topics regarding the single features. The differences are largest for the journal size, where the median is smaller for emerging fields than for established fields, which is in line with H6. A similar observation can be made for the JIF (H7). Contrary to H8, however, the median for the journal age indicator is lower in the case of emerging than for the established topics. With regard to the reference-based indicators (H1 & H5) as well as for the number of authors (H3), we can only observe very slight differences between established and emerging topics. As for the journal fields (H2) as well as the number of countries (H4), the variance in these variables is comparably low, i.e. 90% of the publications in the sample have a value of 3 or lower and a median of 1. We thus find no differences between established and emerging topics for these two features, at least not in this rather coarse-grained view.

When splitting up the box plots by scientific disciplines (Fig. 3), the differences between established and emerging topics alongside the individual features become more evident. With regard to the journal size, which has been introduced as a measure of specialization, we find quite distinct distribution patterns between emerging and established topics for all disciplines, except for physics, mathematics & computer science. As hypothesized, scientific articles dealing with emerging topics generally are published in smaller, more specialized journals (H6). In a similar vein, we can interpret the feature on the number of fields of the journal, as it indicates, on the one hand, the level of specialization of the publishing journal, but can on the other hand also be seen as an estimate for the interdisciplinarity of the publication itself. However, the low variation on the variable does not leave much room for a descriptive interpretation, also when differentiating the results by scientific disciplines. Only for the engineering field – which in general seems to have a higher interdisciplinarity than the other fields under analysis – differences between emerging and established topics become visible. Contrary to our hypothesis (H2), however, emerging topics in engineering seem to be published less often in journals with a high interdisciplinarity[9]. With regard to the age of the publishing journal, the largest differences can be found in the fields of medicine and engineering. As for the JIF, we find a high variation in the values across the scientific disciplines. In the fields of engineering as well as physics,

[8] http://wokinfo.com/media/essay/journal_selection_essay-en.pdf (last accessed on 2015/02/19).

[9] As for chemistry, the distribution of the values is highly skewed. In 79% of the cases, journals are only assigned to one field. Thus, a box plot cannot be drawn in this specific case.

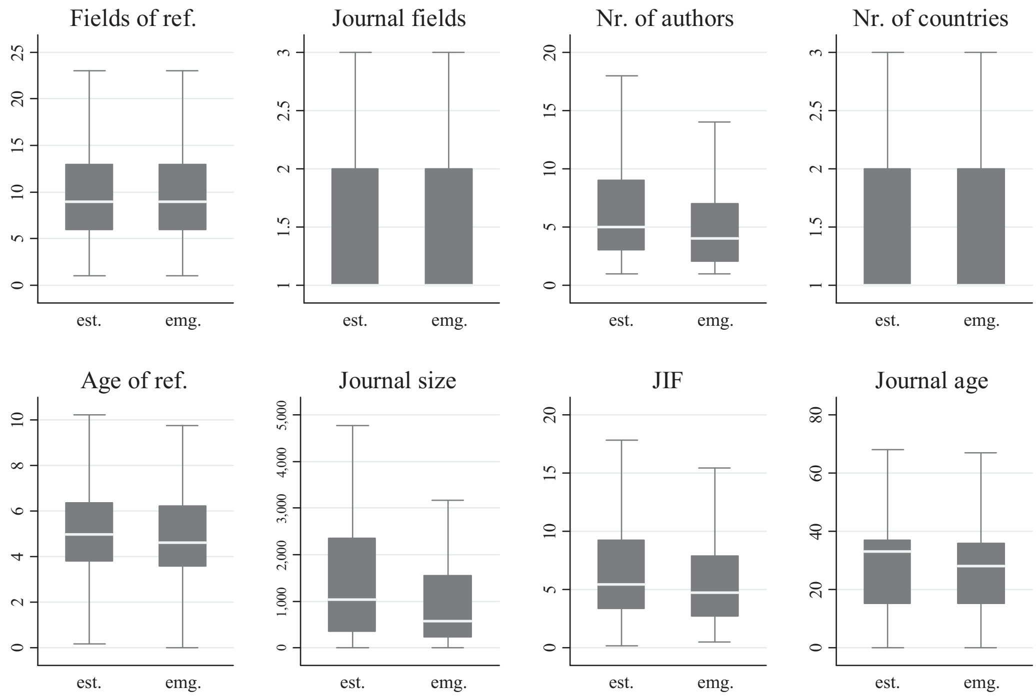

Fig. 2. Differences between established and emerging topics. *Source:* Web of Science, own calculations. *Note:* est. = established, emg. = emerging.

mathematics & computer science, for instance, the JIF is generally lower than in medicine or chemistry. Yet, in chemistry, medicine as well as physics, mathematics & computer science, we find the highest differences in the JIF values between established and emerging topics. Across these fields, emerging topics seem to be more commonly published in journals with a lower JIF, which supports our hypothesis (H7).

For the reference-based measure (H1 & H5), i.e. the age and the number of fields of the references, we find a rather differentiated picture, which does not allow for a general conclusion at this point. The range of the age of the references is similar across the disciplines, yet the differences between established and emerging fields are subject to high field-specific variations. As for the number of fields the citations are referring to, some variation across disciplines becomes evident. However, the differences between established and emerging topics seem to be minor across all fields.

Regarding the authors involved in a publication, on the other hand, differences between emerging and established topics are perceivable for all disciplines but Chemistry. In most disciplines, the number of authors per document is by trend smaller for emerging topics, which supports our hypothesis that collaboration partners are harder to find in a yet underdeveloped topic (H3). However, this observation does not reflect in the number of countries of the authors (H4), which is equally distributed for emerging and established topics, although this can once again be attributed to the rather small variance in this variable.

To get a better idea of how good the single factors are in predicting whether a topic can be categorized as established or emerging, we further applied a receiver operating characteristic (ROC) analysis. In a ROC curve, the true positive rate (often called sensitivity in ROC analysis) is plotted as a function of the false positive rate (1-specificity or true negative rate) for different cut-off points of a parameter (e.g. Metz, 1978; Pepe, Longton, & Janes, 2009; Zweig/Campbell, 1993) A perfect test has a ROC curve with 100% sensitivity and 100% specificity and the ROC area value equals 1. A value of 0.5 represents a random draw, i.e. a ROC area value of 0.5 for a given feature shows that the feature does not very well discriminate between established and emerging topics (Zweig/Campbell, 1993). The ROC summary table as well as the ROC curves for our features are shown in Table 4 and Fig. 4.

As we can see from the table as well as the curves, the single features do not very well distinguish between emerging and established topics for themselves, i.e. the values all are rather close to 0.5. In sum, the evidence from the descriptive statistics points into the direction that the single indicators as such do not help us very well with an early-stage identification of emerging topics in science. However, it might be that that an indicator system, based on the selected bibliometric indicators, might be able to provide us with a pre-selection of emerging and established topics in science. This will be tested with the help of multivariate regression models in Section 4.2. It is further evident from the box plots that we have to acknowledge the fact that there are large variations across fields, implying that a simpler general model for the early-stage identification of emerging topics might not be suitable. Therefore, besides providing a general model, we will differentiate our multivariate analyses across fields, in order to additionally provide with field-specific recommendations for the identification of emerging topics in science.

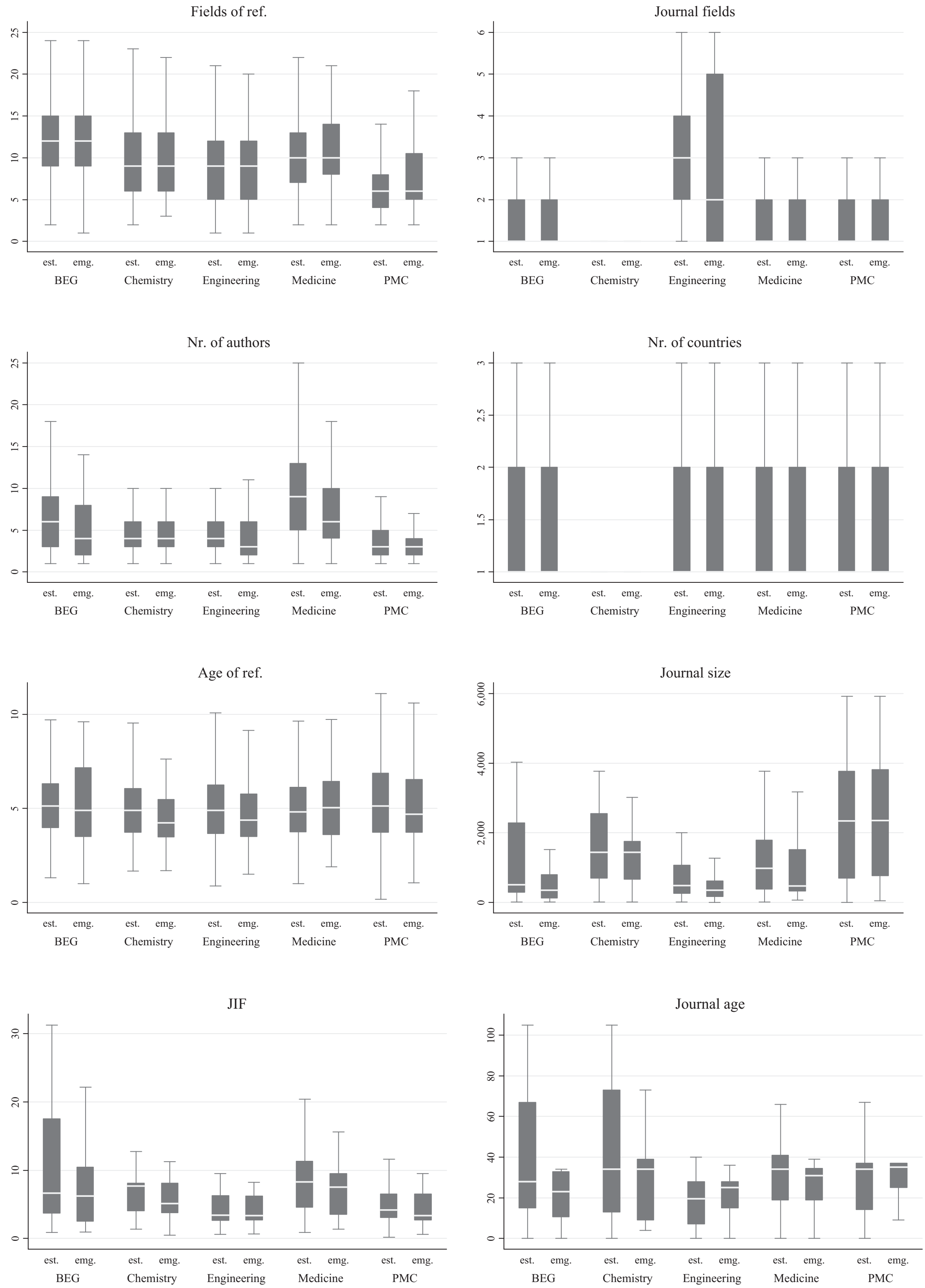

Fig. 3. Differences between established and emerging topics across scientific disciplines. *Source:* Web of Science, own calculations. *Note:* est. = established, emg. = emerging, BEG = Biology, environmental science & geoscience, PMC = Physics, mathematics & computer science.

Table 4
ROC analysis—summary statistics.

	Obs.	ROC area	S.E.	Asymptotic normal 95% conf. interval	
Fields of ref.	3295	0.506	0.017	0.472	0.540
Journal fields	3295	0.522	0.016	0.492	0.553
Nr. of authors	3295	0.403	0.017	0.371	0.435
Nr. of author countries	3295	0.446	0.014	0.419	0.472
Age of ref.	3295	0.463	0.018	0.428	0.499
Journal size	3295	0.438	0.018	0.403	0.473
JIF	3295	0.424	0.017	0.390	0.458
Journal age	3295	0.462	0.017	0.429	0.494

Source: Web of science, own calculations.

4.2. Multivariate results

As a further step towards providing indicators for the early-stage identification of emerging topics in science, we ran a set of logistic regression models with the discussed publication features as explanatory variables and the information if a document is dealing with an emerging or an established topic as the dependent variable. Analogous to the descriptive analyses, we first of all estimated a general model across all disciplines, yet controlling for the field specifities. In order to test for multicollinearity between the explanatory variables, variance inflation factors (VIFs), based on an Ordinary Least Squares (OLS) model with "emerging topic" as the dependent variable, were calculated. As a rule of thumb, VIF values above 5 indicate a high multicollinearity between the variables. Besides the field dummies, which showed VIF values between 2.12 and 2.79, the journal age variable had the highest VIF (1.70). The mean VIF for the model was 1.67. Hence, we find no multicollinearity concerns (O'Brien, 2007).

Table 5 shows the marginal effects for this model. With a value of −0.0097, the largest coefficient can be found for the number of author countries. The coefficient is significant at the 10% level. Each additional author country named on a publication thus decreases the probability to find an emerging topic by 0.9%. This provides evidence that the probability of

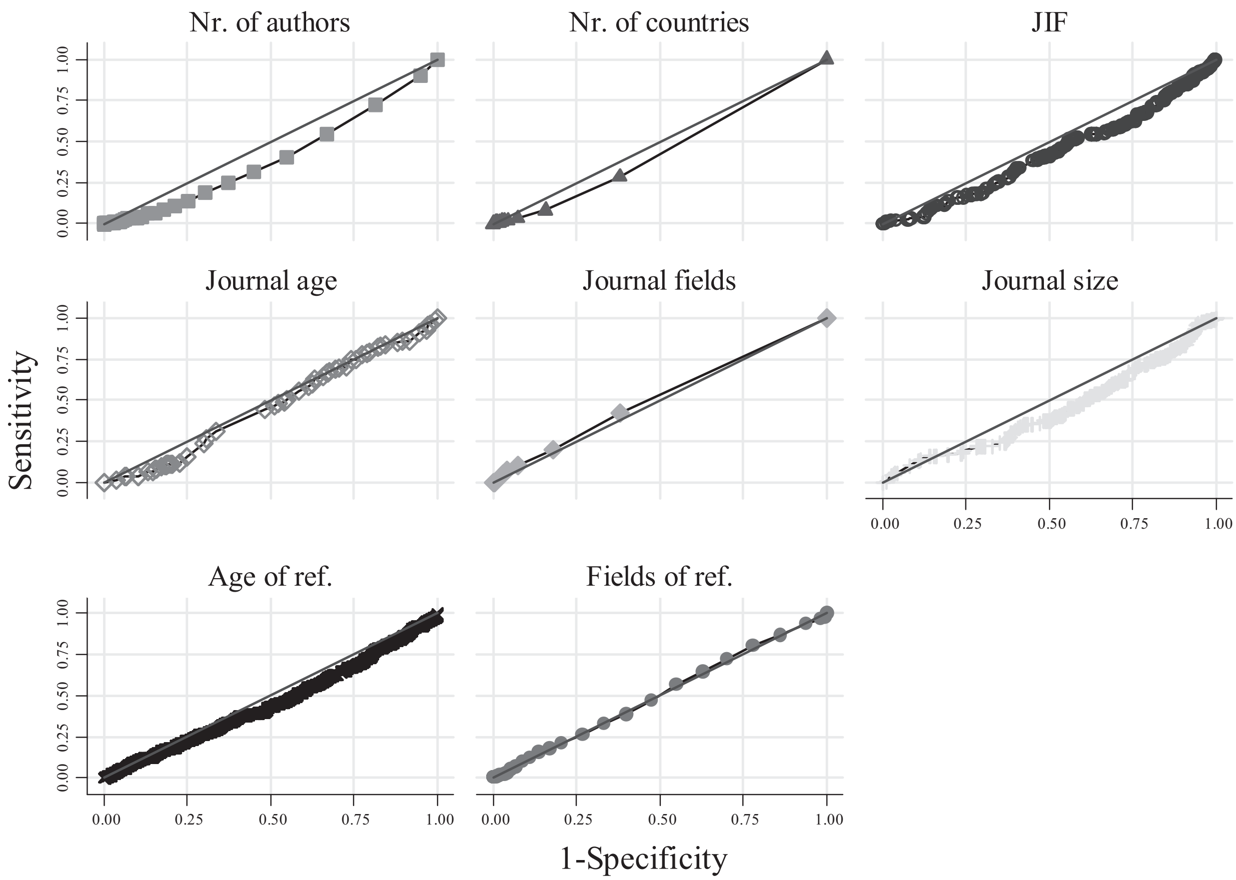

Fig. 4. ROC curves for the different features. *Source:* Web of Science, own calculations

Table 5
Logistic regression—marginal effects.

dV: Emerging topic	dy/dx	S.E.	95% Conf. interval	
Biology, environmental science & geoscience	−0.059***	0.018	−0.094	−0.024
Medicine	−0.064***	0.018	−0.099	−0.030
Chemistry	−0.035*	0.018	−0.070	0.000
Physics, mathematics & computer science	−0.057***	0.017	−0.090	−0.023
Fields of ref.	0.001	0.001	−0.001	0.003
Journal fields	−0.005	0.005	−0.015	0.005
Nr. of authors$^{\alpha}$	0.009	0.216	−0.414	0.432
Nr. of author countries	−0.010*	0.005	−0.020	0.001
Age of ref.	−0.004	0.003	−0.010	0.001
Journal size$^{\alpha}$	0.001	0.005	−0.009	0.010
JIF	−0.003**	0.001	−0.005	0.000
Journal age$^{\alpha}$	−0.370	0.261	−0.881	0.141
Number of obs.	3236			
Wald chi^2	50.1			
Prob > chi^2	0.000			
Pseudo R^2	0.026			

Source: Web of Science, own calculations.
Significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Note:* $^{\alpha}$ Coefficients, standard errors and confidence intervals multiplied by 1,000 to make effects visible. "Engineering" is the reference group for the field dummies. For dummy variables, dy/dx is for discrete change of dummy variable from 0 to 1. The number of observations is slightly lower in the model than in the summary statistics as there are some values missing on the JIF variable.

finding an emerging topic on average decreases for an increasing number of distinct author countries named on a publication, which is in correspondence with our hypothesis H4. The second largest coefficient can be observed for the distinct number of journal fields in which a journal is classified. The coefficient is negative, i.e. an increase in the number of journal fields by one unit decreases the probability to find an emerging topic by 0.5%. The negative sign of the coefficient contradicts H2, where we stated that the chance of finding an emerging topic in a journal is on average increasing with the number of fields a journal is classified in. Yet, the effect is statistically not significant. Another feature, where a comparably large coefficient can be found is the age of the references. It shows a negative value of −0.004 and is statistically not significant, but contradicts H3, where we stated that the chance of finding an emerging topic should on average increase with the age of the references cited in a document. The fourth largest coefficient, which is significant at the 5% level, can be observed for the JIF (−0.0027). The probability for documents in emerging topics to be published thus decreases with a rising impact factor of a journal, which supports the arguments of Benos et al. (2007) and is in line with our H7.

For all the other feature variables, however, we only find rather small and non-significant effects in this general model across disciplines. Yet, we observe highly significant coefficients for the field dummy variables, once again indicating that the differences between established and emerging topics are varying highly across disciplines. The individual features might therefore also have different effects depending on its field of occurrence. In a next step, we therefore re-ran our models separated by disciplines in order to find out which of the feature variables show significant effects in which of the scientific fields (Tables 6 and 7).

As expected, we find different effects of our independent variables across the disciplines. Most distinctively, there are no significant coefficients in the field of biology, environmental engineering and geoscience. The coefficients are of all variables are comparably small in size, implying that within this field, the early-stage identification of emerging topics with the help of our indicators is hardly possible. Similarly, in the field of medicine, only one significant coefficient can be observed, namely the number of authors. The negative value of the coefficient shows that within medicine, the chance of finding an emerging topic decreases by 0.8% for each additional author named on a publication. In medicine, collaboration might therefore be impeded (at least on a small scale) for emerging topics.

As for chemistry, the age of the references proves to be a valid indicator of the novelty of a topic. However, an increase in the average reference age by one unit decreases the chance of finding an emerging topic by 1.5%. This observation contradicts our hypothesis (H5). Documents in emerging topics in chemistry seem to rely by trend more on a more recent knowledge base.

When it comes to physics, mathematics and computer science as well as engineering, the indicators show a more precise picture. Here, we find statistically significant effects for two variables, namely the journal size as well as the number of fields of the references. The journal size is positively related with the development stage of a topic in physics, mathematics and computer science. Although the effect is relatively small in size – as the journal size was measured via the number of articles of a journal in a year – this contradicts H6. Furthermore, the number of fields of the references shows a positive coefficient within physics, mathematics and computer science. Thus, at least for this field, we can confirm the assumption that different fields are combined in the generation of an emerging topic.

Finally, in engineering we find three variables to be significant, implying that an early-stage indicator system based on our publication features works best within this field. Here, the journal size is negatively related to our "emerging topic" variable, i.e. an increase in article numbers decreases the chance of finding an emerging topic in this field (16.7% per 1,000 articles). The age of a journal also has a positive effect on the chance of finding an emerging topic (H8). For each additional

Table 6
Logistic regressions for the single disciplines–marginal effects I.

dV: Emerging topic	Engineering				Biology, environmental science & geoscience				Medicine			
	dy/dx	S.E.	95% Conf. interval		dy/dx	S.E.	95% Conf. interval		dy/dx	S.E.	95% Conf. interval	
Fields of ref.	0.001	0.004	−0.006	0.009	−0.002	0.002	−0.006	0.001	−0.001	0.002	−0.004	0.002
Journal fields	−0.002	0.017	−0.035	0.031	−0.015	0.012	−0.039	0.008	−0.007	0.011	−0.028	0.015
Nr. of authors	0.002*	0.001	0.000	0.003	−0.002	0.002	−0.006	0.003	−0.008***	0.002	−0.012	−0.004
Nr. of author countries	−0.021	0.025	−0.069	0.027	−0.015	0.013	−0.041	0.011	0.006	0.007	−0.007	0.019
Age of ref.	−0.009	0.008	−0.025	0.007	−0.005	0.006	−0.017	0.006	0.001	0.005	−0.008	0.010
Journal size$^{\alpha}$	−0.167***	0.041	−0.248	−0.086	−0.013	0.014	−0.040	0.014	−0.004	0.008	−0.020	0.012
JIF	−0.003	0.004	−0.011	0.005	−0.002	0.002	−0.005	0.001	−0.001	0.001	−0.004	0.002
Journal age$^{\alpha}$	4.712***	1.669	1.442	7.982	−0.460	0.546	−1.531	0.610	−0.095	0.369	−0.818	0.629
Number of obs.	420				730				886			
Wald chi^2	22.94				13.89				20.82			
Prob > chi^2	0.003				0.085				0.008			
Pseudo R^2	0.061				0.049				0.050			

Source: Web of Science, own calculations.
Significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Note:* $^{\alpha}$ Coefficients, standard errors and confidence intervals multiplied by 1000 to make effects visible.

Table 7
Logistic Regressions for the single disciplines—marginal effects II.

d*V*: Emerging topic	Chemistry				Physics, mathematics & computer science			
	dy/dx	S.E.	95% Conf. Interval		dy/dx	S.E.	95% Conf. Interval	
Fields of ref.	0.001	0.003	−0.004	0.006	0.006***	0.002	0.002	0.011
Journal fields	0.009	0.011	−0.012	0.031	0.002	0.007	−0.012	0.015
Nr. of authors	0.004	0.007	−0.009	0.017	−0.006	0.006	−0.017	0.005
Nr. of author countries	0.008	0.026	−0.043	0.059	−0.015	0.021	−0.056	0.025
Age of ref.	−0.015*	0.009	−0.032	0.002	−0.002	0.005	−0.011	0.008
Journal size$^{\alpha}$	−0.009	0.038	−0.084	0.065	0.019***	0.006	0.006	0.031
JIF	−0.004	0.005	−0.013	0.005	−0.006	0.004	−0.014	0.002
Journal age$^{\alpha}$	0.101	1.012	−1.883	2.085	0.417	0.921	−1.388	2.221
Number of obs.	500				700			
Wald chi^2	5.84				26.79			
Prob > chi^2	0.665				0.001			
Pseudo R^2	0.020				0.063			

Source: Web of Science, own calculations.
Significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. *Note:* $^{\alpha}$ Coefficients, standard errors and confidence intervals multiplied by 1000 to make effects visible.

year in the age of a journal, the probability to find an emerging topic increases by 0.5%. In engineering, emerging topics are thus on average published in older journals. In addition, the number of authors named on a publication shows a significantly positive coefficient. In engineering, documents dealing with emerging topics are thus on average published more often by larger research teams.

In sum, we can state that there are early-stage indicators for the identification of emerging topics in science. However, at the expense of a very timely availability of these indicators, we have to deal with certain inaccuracies we do not have fully under control. Using the indicators at hand thus provides us the possibility of making a certain pre-selection of documents that might with a given probability deal with an emerging topic in science. After this pre-selection via the discussed publication features, still a manual search or a further analysis, e.g. with the help of citation indicators, is indispensable in order to truly find out if a publication in fact deals with an emerging topic or not. As soon as a document is identified as such, it can further be used to search for other publications that are thematically related. In addition, it can be found that the size and significance of the coefficients differ across disciplines (Tables 6 and 7). We therefore should only perform discipline-specific analyses with the indicators that have been identified as possessing a given explanatory power in differentiating emerging from established topics.

To be more precise, specific characteristics of emerging topics involve the journal size and age, the reference age and the fields and the number of authors. Thus, all three factors, the interdisciplinarity, the emergence source as well as the emergence environment play a role in the emergence of a topic. However, the exact parameter values differed for the analyzed disciplines. This also makes it rather difficult to make clear statements about our hypotheses. Also, for two of the variables, the number of the fields a journal is classified in (H2) as well as the age of the references (H5), we find no evidence in support of our hypotheses.

In regard of the interdisciplinarity of a topic, supporting evidence can only be found within the field of physics, mathematics & computer science for the number of the fields of the references (H1). Regarding the emergence sources, the number of author countries (H4) named on a publication is negatively related to documents in emerging topics. For the second factor, the number of authors (H3), we only find support in the fields of engineering and medicine. As for the emergence environment, the JIF (H7) is negatively related to the emergence of a topic. Also, we find evidence for our hypotheses for the journal size (H6) and journal age (H8), but only in the field of engineering.

5. Conclusions

We tested various features of scientific publications in order to identify a set of early-stage indicators for emerging topics. As mentioned in the beginning, the respective set of indicators not necessarily aims to be complete. That means in particular: (1) There are certainly more characteristic features for emerging topics which we did not cover in this study (e.g. non-bibliometric features like textual features or those stemming from the CV of the authors as well as bibliometric features not covered in the Web of Science like the time between submission and publication) and (2) the application of the features does not necessarily yield a result set covering all emerging topics from a document set. Furthermore, the features are highly dependent on relative factors and comparison to other publications in a dataset. Nonetheless, they offer a useful insight in the development process of topics and their initial obstacles in the publication process.

Specific characteristics of emerging topics involved the journal size and age, the reference age and the fields and the number of authors. However, the exact parameter values differed for the analyzed disciplines. The most pronounced discipline was engineering for which we identified a smaller journal size, a higher journal age and a higher number of authors as indicators for emerging topics. Thus, publications in emerging topics were by trend published in established and specialized journals but with a higher collaboration effort than usual. In contrast, medicine was found to have a smaller number of authors, which corroborated our assumption that collaboration might be hindered in emerging topics. Thus, there were two effects apparent: In technical fields, more researchers, research groups and/or equipment were needed in order to promote an emerging topic while in other fields, communication and thus collaboration might be impeded by the novelty of a topic.

These features of emerging topics might facilitate their promotion. By a heightened awareness for particularly these topics, their development can be channeled in beneficent directions at an early stage. Foremost, strategic planning is enabled, which fosters the optimal allocation of funding, equipment and work force. The set of indicators presented here allows only for a pre-selection step of candidates for emerging topics. However, given the vast (increasing) amount of annual scientific publications, such a pre-selection could be a crucial step in research monitoring.

Yet, there are still some limitations to our study. First, it was limited by the small sample size, which only allows restricted general deductions. With a larger dataset, more profound and results could be expected. Second, the restriction to the top 1% highly cited papers might induce a bias: the pre-selection covered only documents that were at least in a time span of 1 year already acknowledged in the scientific community. This does not only restrict the dataset to a smaller size but also limits the generalizability of our results to a certain extent, as only publications that were deemed noteworthy by the scientific community are included. It cannot be traced how many of the full-range of existing documents dealing with emerging topics are lost by this pre-selection. Since many of our features are targeted towards the hindered (or more competitive) publication process of novel ideas, the results gained with a general dataset of emerging and established topics without such a limitation could also be expected to be more universally applicable and distinct.

Acknowledgements

Major parts of this paper were written during a guest stay at the National Institute of Science and Technology Policy (NISTEP), Ministry of Education, Culture, Sports, Science and Technology (MEXT), Tokyo, Japan. The data for the classification of new and old topics were provided by the NISTEP during that stay. Certain supplementary data were derived from the Science Citation Index Expanded (SCIE), the Social Science Citation Index (SSCI), the Arts and Humanities Citation Index (AHCI), and the Index to Social Sciences & Humanities Proceedings (ISSHP) (all updated June 2012) prepared by Thomson Reuters (Scientific) Inc. (TR®), Philadelphia, Pennsylvania, USA: © Copyright Thomson Reuters (Scientific) 2012. All rights reserved.

The authors would like to thank Ayaka Saka and Masatsura Igami for valuable comments on a first draft of this paper.

Author contributions

Conceived and designed the analysis: Carolin Mund; Peter Neuhäusler
Collected the data: Carolin Mund
Contributed data or analysis tools: Carolin Mund; Peter Neuhäusler
Performed the analysis: Carolin Mund; Peter Neuhäusler
Wrote the paper: Carolin Mund; Peter Neuhäusler
Other contribution: if there are no contribution this can be deleted

References

Alexander, David. (2012). Our starting point. *International Journal of Disaster Risk Reduction*, *1*, 1–4. http://dx.doi.org/10.1016/j.ijdrr.2012.06.002

Alvargonzález, David. (2011). Multidisciplinarity, interdisciplinarity, transdisciplinarity, and the sciences. *International Studies in the Philosophy of Science*, *25*, 387–403. http://dx.doi.org/10.1080/02698595.2011.623366

Anholt, R. M., Stephen, C., & Copes, R. (2012). Strategies for collaboration in the interdisciplinary field of emerging zoonotic diseases. *Zoonoses and Public Health*, *59*, 229–240. http://dx.doi.org/10.1111/j.1863-2378.2011.01449.x

Atkins, Peter W. (2003). *Galileo's finger. The ten great ideas of science*. Oxford: Oxford University Press. Available online at ⟨http://www.gbv.de/dms/goettingen/357743822.pdf⟩.

Barber, B. (1961). Resistance by scientists to scientific discovery. *Science*, *1*, 596–602. http://dx.doi.org/10.1126/science.134.3479.596

Benos, D. J., Bashari, E., Chaves, J. M., Gaggar, A., Kapoor, N., LaFrance, M., et al. (2007). The ups and downs of peer review. *AJP: Advances in Physiology Education*, *31*, 145–152. http://dx.doi.org/10.1152/advan.00104.2006

Campanario, Juan Miguel. (2009). Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates. *Scientometrics*, *81*, 549–565. http://dx.doi.org/10.1007/s11192-008-2141-5

Chew, Felix S. (1991). Fate of manuscripts rejected for publication in the AJR. *American Journal of Roentgenology*, *156*, 627–632. http://dx.doi.org/10.2214/ajr.156.3.1899764 (checked on 2/17/2014)

Costas, Rodrigo, van Leeuwen, Thed N., & van Raan, Anthony F. J. (2011). The "Mendel syndrome" in science: durability of scientific literature and its effects on bibliometric analysis of individual scientists. *Scientometrics*, *89*, 177–205. http://dx.doi.org/10.1007/s11192-011-0436-4

Cozzens, Susan E. (1989). What do citations count? The rhetoric-first model. *Scientometrics*, *15*, 437–447. http://dx.doi.org/10.1007/BF02017064

Dirk, L. (1999). A measure of originality: The elements of science. *Social Studies of Science*, *29*, 765–776. http://dx.doi.org/10.1177/030631299029005004

Franck, G. (1999). Scientific communication—A vanity fair? *Science*, *286*, 53–55. http://dx.doi.org/10.1126/science.286.5437.53

Garfield, Eugene. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*, 359–375. http://dx.doi.org/10.1007/BF02019306

Glänzel, Wolfgang, & Schoepflin, Urs. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management*, *35*, 31–44. http://dx.doi.org/10.1016/S0306-4573(98)00028-4

Handy, C. (1995). Trust and the virtual organization. *Harvard Business Review*, *73*, 40–50.

Johnson, Steven. (2013). *Wo gute Ideen herkommen. Eine kurze Geschichte der Innovation*. Bad Vilbel: Scoventa.

Kajikawa, Yuya, & Takeda, Yoshiyuki. (2009). Citation network analysis of organic LEDs. *Technological Forecasting and Social Change*, *76*, 1115–1123. http://dx.doi.org/10.1016/j.techfore.2009.04.004

Kilwein, J. H. (1999). Biases in medical literature. *Journal of Clinical Pharmacy and Therapeutics*, *24*, 393–396. http://dx.doi.org/10.1046/j.1365-2710.1999.00243.x

Kuhn, Thomas S. (1970). *The structure of scientific revolutions* (second ed., Enlarged). Chicago: The University of Chicago Press.
Larsen, Peder Olesen, & Ins, Markus. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, *84*, 575–603. http://dx.doi.org/10.1007/s11192-010-0202-z
Long, J. S. (1997). *Regression models for categorical and limited dependent variables. Advanced quantitative techniques in the social sciences series 7* Thousand Oaks, London, New Delhi: SAGE Publications.
Long, J. S., & Freese, J. (2003). *Regression models for categorical dependent variables using Stata* (revised ed.). College Station, TX: Stata Press.
MacRoberts, M. H., & MacRoberts, Barbara R. (1996). Problems of citation analysis. *Scientometrics*, *36*, 435–444. http://dx.doi.org/10.1007/BF02129604
Merton, Robert K. (1968). The Matthew effect in science. *Science*, *159*, 56–68. http://dx.doi.org/10.1126/science.159.3810.56
Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, *8*, 283–298.
Michels, Carolin, & Schmoch, Ulrich. (2012). The growth of science and database coverage. *Scientometrics*, *93*, 831–846. http://dx.doi.org/10.1007/s11192-012-0732-7
Michels, Carolin, & Schmoch, Ulrich. (2014). Impact of bibliometric studies on the publication behaviour of authors. *Scientometrics*, *98*, 369–385. http://dx.doi.org/10.1007/s11192-013-1015-7
Morillo, Fernanda, Bordons, Maria, & Gomez, Isabel. (2001). An approach to interdisciplinarity through bibliometric indicators. *Scientometrics*, *51*, 203–222. http://dx.doi.org/10.1023/A:1010529114941
O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*, 673–690. http://dx.doi.org/10.1007/s11135-006-9018-6
Olson, Gary M., & Olson, Judith S. (2000). Distance matters. *Human Computer Interaction*, *15*, 139–178.
Pepe, M. S., Longton, G., & Janes, H. (2009). Estimation and comparison of receiver operating characteristic curves. *Stata Journal*, *9*, 1–16.
Porter, Alan L., & Rafols, Ismael. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*, 719–745. http://dx.doi.org/10.1007/s11192-008-2197-2
Price, Derek de Solla. (1965). Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science Communication*, *149*, 510–515.
Rafols, Ismael, & Meyer, Martin S. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*, 263–287. http://dx.doi.org/10.1007/s11192-009-0041-y
Ragone, Azzurra, Mirylenka, Katsiaryna, Casati, Fabio, & Marchese, Maurizio. (2013). On peer review in computer science: analysis of its effectiveness and suggestions for improvement. *Scientometrics*, *97*, 317–356. http://dx.doi.org/10.1007/s11192-013-1002-z
Ray, Joel, Berkwits, Michael, & Davidoff, Frank. (2000). The fate of manuscripts rejected by a general medical journal. *The American Journal of Medicine*, *109*, 131–135. http://dx.doi.org/10.1016/S0002-9343(00)00450-2
Rinia, Ed J., van Leeuwen, Thed N., Bruins, Eppo E. W., van Vuren, Hendrik G., & van Raan, Anthony F. J. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, *51*, 293–309. http://dx.doi.org/10.1023/A:1010589300829
Rocco, E. (1998). Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In *Proceedings of the CHI'98 conference on human factors in computing systems* New York: ACM, (pp. 496–502). http://dx.doi.org/10.1145/274644.274711
Saka, A., Igami, M., & Kuwahara, T. (2010). Science map 2008. In *NISTEP report no. 139*. Tokyo: NISTEP.
Shafique, Muhammad. (2013). Thinking inside the box? Intellectual structure of the knowledge base of innovation research (1988–2008). *Strategic Management Journal*, *34*, 62–93. http://dx.doi.org/10.1002/smj.2002
Shibata, Naoki, Kajikawa, Yuya, Takeda, Yoshiyuki, & Matsushima, Katsumori. (2009). Comparative study on methods of detecting research fronts using different types of citation. *Journal of the American Society for Information Science*, *60*, 571–580. http://dx.doi.org/10.1002/asi.20994
Shibata, Naoki, Kajikawa, Yuya, Takeda, Yoshiyuki, Sakata, I., & Matsushima, Katsumori. (2009). Early detection of innovations from citation networks. In *IEEE international conference on industrial engineering and engineering management 2009 (IEEM2009) in Hong Kong* December 8–11, 2009.
Small, Henry. (2006). Tracking and predicting growth areas in science. *Scientometrics*, *68*, 595–610. http://dx.doi.org/10.1007/s11192-006-0132-y
Small, Henry, & Upham, Phineas. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, *79*, 365–375. http://dx.doi.org/10.1007/s11192-009-0424-0
Steele, Thomas W., & Stier, Jeffrey C. (2000). The impact of interdisciplinary research in the environmental sciences: a forestry case study. *Journal of the American Society for Information Science*, *51*, 476–484. http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:53.3.CO;2-7 (checked on 2/10/2014)
Stent, G. S., & Hook, E. (2002). *Prematurity in scientific discovery. Prematurity in scientific discovery: On resistance and neglect*. Berkeley, CA: University of California Press.
Stent, G. S. (1972). Prematurity and uniqueness in scientific discoveries. *Scientific American*, *227*, 84–93.
Thompson Klein, Julie. (2004). Prospects for transdisciplinarity. *Futures*, *36*, 515–526. http://dx.doi.org/10.1016/j.futures.2003.10.007
Turoff, M., & Hiltz, S. R. (1982). The electronic journal: A progress report. *Journal of the American Society for Information Science*, *33*, 195–202. http://dx.doi.org/10.1002/asi.4630330402 (checked on 2/17/2014)
Urata, Hiroaki. (1990). Information flows among academic disciplines in Japan. *Scientometrics*, *18*, 309–319. http://dx.doi.org/10.1007/BF02017767 (checked on 2/10/2014)
van Dalen, Hendrik P., & Klamer, Arjo. (2005). Is science a case of wasteful competition? *Kyklos*, *58*, 395–414. http://dx.doi.org/10.1111/j.0023-5962.2005.00294.x
van Raan, Anthony F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*, 461–466. http://dx.doi.org/10.1023/B:SCIE.0000018543.82441.f1
Zand, D. E. (1972). Trust and managerial problem solving. *Administrative Science Quarterly*, *17*, 229–239.
Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*, 561–577.