

Searching for converging research using field to field citations

Reindert K. Buter · Ed. C. M. Noyons · Anthony F. J. Van Raan

Received: 26 March 2010 / Published online: 6 June 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract We define converging research as the emergence of an interdisciplinary research area from fields that did not show interdisciplinary connections before. This paper presents a process to search for converging research using journal subject categories as a proxy for fields and citations to measure interdisciplinary connections, as well as an application of this search. The search consists of two phases: a quantitative phase in which pairs of citing and cited fields are located that show a significant change in number of citations, followed by a qualitative phase in which thematic focus is sought in publications associated with located pairs. Applying this search on publications from the Web of Science published between 1995 and 2005, 38 candidate converging pairs were located, 27 of which showed thematic focus, and 20 also showed a similar focus in the other, reciprocal pair.

Keywords Interdisciplinarity · Convergence · Emergence · Non-linear growth · Journal subject categories · Citations · Research focus

Introduction

Some say that “a Jack of all Trades is a master of none”. Yet, the days in which “[...] our universities [...] [were] divided into different departments that [did] not know very much of each other” (Cassirer 1942, p. 309), appear behind us. Instead, interdisciplinary research “is spreading all over the landscape of science and technology” (Gibbons et al. 1994, p. 22) for more than one reason. For instance, knowledge migration is a fruitful mechanism by which science expands into new realms (De Mey 1982, pp. 140–145), providing an attractive opportunity for researchers to attain recognition. Also, interdisciplinary research has been encouraged by funding agencies as a problem-driven mode of research (Carayol and Thi 2005), or as a way to “attain [...] system solutions” for complex societal problems

R. K. Buter (✉) · Ed. C. M. Noyons · A. F. J. Van Raan
Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62a,
P.O. Box 905, 2300 AX Leiden, The Netherlands
e-mail: buter@cwts.leidenuniv.nl

(Persson 1999). In spite of this encouragement, the status and prestige of interdisciplinary research is not clear-cut. On one hand, Rinia et al. (2001) found no (general) evidence of a negative bias to interdisciplinary research in peer-review assessments of physics, nor in relative bibliometric indicators. On the other hand however, De Boer et al. (2006) quote researchers who attribute a perceived lack of prestige to interdisciplinary research. Also, Carayol and Thi (2005) found evidence of a lack of incentive in the academic reward system. To add to this uncertainty, Larivière and Gingras (2010) show that the amount of interdisciplinary links in a publication, which is common in many interdisciplinary publications, may influence impact. Apart from the uncertain academic reward of an interdisciplinary venture, Palmer (1999) notes that even experienced scientists feel challenged by the task of acquiring knowledge from outside their realm of expertise. Additionally, the social, political, and cultural structures of research areas create other barriers (Ruiz-Baños et al. 1999).

In all, we conclude that moving to research outside a familiar setting requires an investment in new knowledge, new vocabulary, as well as new social structures and customs, while the pay-off in terms of acceptance, let alone an increase in status and prestige among peers, appear less certain. This tension makes investigating the creation and subsequent growth and decline of interdisciplinary research areas even more interesting.

We define converging research as the emergence of a new interdisciplinary research area from fields which showed no such interdisciplinary connections before; the result we call converged research, or a convergence. Our definition is similar to that of Nordmann (2004), who speaks of mutually enabling systems and technologies in pursuit of a common goal, but our definition is more targeted to science systems. At the same time, it is more specific than that of Roco and Bainbridge (2003), who speak in more conceptual terms and identify trends on a very large scale: the “megatrends” in Roco (2002). Also, it is similar to the use of convergence as a technological phenomenon (Gambardella and Torrisi 1998; Rosenberg 1976): for example, the shared technical basis underlying convergence in industry can be compared to the generality of research methods or tools.

In this paper, we describe a process that locates converging research based on journal subject categories in the Web of Science database as proxies for fields. Citations (from one field to another) are used to measure interdisciplinary connections. Our working procedure consists of a quantitative and qualitative part: the first part locates candidates, and the second part inspects those candidates. The quantitative part uses an objective basis for the cut-off value for significant (in the sense of “noticeable larger”) growth, as suggested in our previous article (Buter et al. 2010). The qualitative part is based on a visual inspection of a tableau of graphs, as well as an inspection of publication data assembled from the converging area.

Converging research has not been the topic of many scientometric publications, although there are a lot of descriptions and analyses of emerging research, like for example Mathematical Logic (Berg and Wagner-Döbler 1996), Bioelectronics (Hinze 1994) or Nanoscience and Nanotechnology (Schummer 2004). Additionally, investigations of interdisciplinary developments have been described in for example Davidse and Van Raan (1997), Rinia et al. (2002), and Morillo et al. (2003). Also, general methodologies to find emerging patterns have been described by for example Morris (2005), Takeda and Kajikawa (2009), or Upham and Small (2010). However, like the descriptions of the emerging research areas, these general approaches are applied to a limited set of publications, or for a specific topic, and none of these publications deal with the general and broad search for converging research we describe in this paper.

Data and method

Bibliometric data

The bibliometric data used for our research, consisted of all publications in the Web of Science (WoS) database¹ published between 1995 and 2005, including the social sciences and humanities. We also used the citations from these 1995–2005 publications to WoS publications, but the cited publications were restricted to “articles”, “letters”, “notes” and “reviews” (the citing documents could be any type). Both author self-citations and journal self-citations were excluded. Author self-citations were excluded because such citations may represent other mechanisms than the use of research (de Solla Price 1981). Journal self-citations on the other hand, were excluded because our method was based on *journal* subject categories, and we found such that self-citations introduced noise in detecting interdisciplinary developments. We used all 243 journal subject categories (JSCs) provided by Thomson Scientific in 2005, which categorized all journals in the WoS into at least one and at most six categories.

Significant growth, non-linear shape and robust size

The search method we developed, had the objective to find *robust* phenomena showing *significant, non-linear growth*. In this section we describe how we implemented growth, significance, shape and robustness.

Our growth indicator used citation counts within a specific *citation window*: the range of years in which cited publications are published relative to the publication year of a citing publication. For example, for a publication from 1995 and a citation window of 10 years, we count the citations to publications from 1986 to 1995.² A small citation window focuses on recent (relative to the citing publication) developments, whereas a longer window starts to include more and more citations to “classics”. We chose a 10 year wide citation window as a good compromise. Since knowledge transfer takes time (Rinia et al. 2001), a smaller window may run the risk of not including interdisciplinary usage, while a wider window would include more classics, which we consider of limited use as they are probably too general to mark actual research.

Large differences exist in the average number of citations per publication in different fields: for instance, in our data set, the field *Genetics and Heredity* had an average of over 33, whereas *Mathematics* only had about 7. We therefore normalised citation counts from one field to another, by dividing the individual counts by the total number of citations (for a given year) from the citing field to all fields (including itself); we refer to this normalised count as the *citation share* and denote this as $c(A,B)_t$, for year t and fields A and B . For example, if *Mathematics* would give 100 citations to *Genetics and Heredity* in 1996, out of a total of 2,000 from *Mathematics* in 1996, then the citation share of *Mathematics* to *Genetics and Heredity* in 1996 would be 0.05. Next, since growth is about change, we used the difference of citation shares in subsequent years $t - 1$ and t , divided by the (absolute) value for the previous year. To this we refer to as the *growth rate* $g(A,B)_t$, and we can capture its definition in a formula as follows:

¹ This database is available to the CWTS³ under license from its publisher Thomson Scientific; it contains publication data from the year 1980 onward.

² We count whole years, so 1995–1996 counts as a 2-year citation window, even though the difference is only 1 year.

$$g(A, B)_t = (c(A, B)_t - c(A, B)_{t-1}) / |c(A, B)_{t-1}|, \text{ if } |c(A, B)_{t-1}| > 0, \text{ and } 0 \text{ otherwise.}$$

This growth rate was used to identify significant, fast-growing growth.

We wanted an objective basis for the distinction between significant and non-significant growth, because a previous version of our search process, described in Buter et al. (2010), used a strict value which was later considered too large. After some experimentation, significant growth was defined as follows. For all growth-rate time-series of the citing-cited pairs, the median³ value was calculated. The distribution of these median values appeared to be log-normal and we used this distribution characteristic to define significantly growing time-series to be those which had a median value of at least 1.5 standard deviations larger than the average of this distribution. Similar considerations were used in the definition of the RTG⁴ indicator (Buter et al. 2010), and also in Efron and Tibshirani (1993) to assess the significance of the differences between two groups of values.

To select fast-growing pairs in those with significant growth, we first experimented with methods which fit non-linear curves, as well as smoothing functions such as described by Silverman (1985). Unfortunately, the results were not useful, most probably because the time-series of growth rates were small (only ten observations for the years 1995–2005), as well as coarse (large variation of values in a time-series). For the same reasons, indicators such as developed in Egge and Rao (1992) did not yield useful results. Consequently, we resorted to a more basic approach, and devised two straightforward requirements which expressed our interest in recent, fast growth. First, the maximum value of the growth rate should fall between 2002 and 2005. Second, the sum of the citation counts in the period 2001–2005 should be double the sum of counts in the previous period (1995–2001).

A subject can be called to show a robust development if it has a “large enough” number of publications in order to be interesting. However, a requirement such as “large enough” is quite subjective and difficult to express exactly. As a result, there is a level of arbitrariness in the two robustness requirements we used, but we consider them quite acceptable: first, more than half (six) of the values in the growth rate had to be larger than 0; and second, at least 1 year in the period 2002–2005 had to have 25 citations or more. Again, some experimentation was required in order to find these values.

Assessment

The objective of the assessment was to find out more about citing and cited publications of the pairs found after applying the above requirements. Although our main concern was to find common (research) themes, we were also interested in a graphical display of citation counts, in order to evaluate our search method by verifying that our requirements did indeed result in the desired growth shapes.

The shapes of the citation counts and shares were inspected using a tableau of graphs similar to the one shown in Fig. 1 for the pair *Economics* citing *Physics, Fluids and Plasmas*. This tableau is divided vertically in two: the left contains graphs for a pair selected by the search (*Economics* citing *Physics, Fluids and Plasmas*), and the right part contains the same graphs for the *reciprocal* pair, in which the citing and cited field are exchanged (*Physics, Fluids and Plasmas* citing *Economics*). The graphs at the top show the

³ We used a median instead of a mean value, because the median is less affected by the outliers present in many of our time-series of growth rates due to the limited number of values in the series.

⁴ This number is defined as the absolute difference between the number of papers in the first and last year, divided by the total number of papers; it has an approximately normal distribution if the required number of publications is large enough.

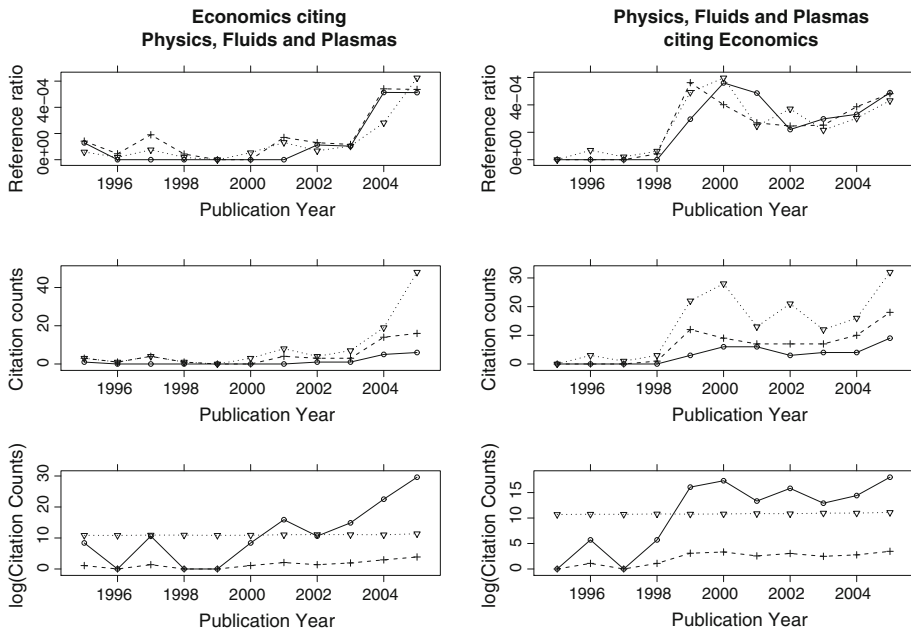


Fig. 1 Example of a tableau of graphs used to analyse the growth characteristics of a selected pair and its reciprocal pair: the graphs in the top row show the citation share developments, those in the middle row the development of the absolute number of citations, while those on the bottom row show a comparison with the (endemic) growth in the originating fields, scaled to equal units

citation share time-series. Below those, the time-series of absolute citation counts are plotted. In order to see any effects of citation delay, also plots for citation windows of different sizes (3, 5 and 10 years) are plotted. Finally, in order to rule out results that are due to a sudden increase in citations in a field as a whole,⁵ graphs are plotted at the bottom which compare the (scaled) citation counts of a pair with the (scaled) number of citations given in both originating fields.

In this paper, research focus is understood as the most specific *common* theme present in most of the publications under inspection. This is a liberal definition and it includes common themes in research subjects as well as in applied methodology. However, we do require research focus to be the most *specific* common theme, as we expect to find multiple themes in many of the pairs we find. The research focus of the pair located by our search method was leading, and alternative focus in the reciprocal pair was not considered. In order to find focus, publication content was assessed using a number of overviews, the most important of which were the following.

- A matrix of cited journals over years containing citation counts, as well one for citing journals over years.
- A list of best-cited publications, with title, journal, year of publication, and number of citations received.

⁵ For example, due to the inclusion of a new journal with a higher number of references per paper than other journals in the field.

- A list of citing publications that have the most citations to the cited field, with title, journal and publication year of both the citing publication and the cited publications.
- Lists of most active and best-cited organisations.

These overviews were also compared with those of the reciprocal pair.

Results and discussion

Search result

The distribution of the positive median growth rate values showed an approximately log-normal distribution with a mean of -2.43 and a standard deviation of 1.41 . There were 683 pairs that had a value 1.5 times the standard deviation above the mean (-0.32). Applying the fast-growth and robustness requirements reduced this to number to 38 pairs. With the sole exception of the pair *Biochemical Research Methods* citing *Statistics and Probability*, none of the 38 pairs had any journals in common. Table 1 lists all pairs. The first two columns of this table contain the names of the citing and the cited field. The column labelled **RF** deals with the research focus in the citing papers, and this column can contain four different values: F for a clear focus on a particular theme; P for a partial focus on a particular theme, with other minor themes also identified; G for a focus in a general, methodological sense (as opposed to a topical sense); and N if no focus could be found. In column **R** it is indicated whether the focus in the reciprocal (reversed) pair is the same as the focus in the pair found to converge: therefore, it only contains Y if a similar focus is found, and N if another focus is found or none at all.⁶ The other columns show indicators for the distribution of the citation counts: the total number of citations **N**, the median **Med**, the maximum **Max**, and the year of the maximum value in **Peak**.

One of the first things noticeable from Table 1, is the relatively small number of citations involved in the detected trends: an average of 251 and a median of 107 citations. So, it appears that fields with higher number of citations between them, do not show enough growth in order to be regarded significant, or do not meet some of the additional growth requirements. Figure 2 shows support for the first. This figure shows the logarithm of the median growth rate over the logarithm of the median size, for the pairs that have at least four observations. The solid diagonal line shown in this figure is a linear fit of the values, which (even though the fit is rather poor) illustrates the negative correlation between growth and size. Additionally, the dashed horizontal line indicates where the significant growth boundary of 1.5 times the standard deviation lies: only the points above that line were regarded in our search. The right-most point, corresponding to the largest median number of citations above that line corresponds to about 400. To explain this implicit limit, we again mention two possibilities. First, rapidly growing, field-surpassing developments are rather scarce, and developments are within the boundaries of a field. Second, the distribution may be dominated by smaller phenomena which show a relatively large growth-rate, although the growth in absolute number of citations is smaller.

⁶ Note that if the **Research Focus** column contains N (no focus is found in the citing publications), then the value in the **Reciprocal** column can only be N as well.

Table 1 The 38 pairs fitting all search requirements

Citing category	Cited category	RF	R	N	Med	Max	Peak
Medical Laboratory Technology	Agriculture, Multidisciplinary	F	Y	285	18	101	2005
Engineering, Industrial	Agricultural Engineering	F	Y	147	8	64	2005
Ornithology	Agriculture, Multidisciplinary	F	Y	118	5	38	2004
Genetics and Heredity	Communication	F	Y	117	6	42	2005
Immunology	Construction and Building Technology	F	Y	111	6	27	2005
Economics	Physics, Fluids and Plasmas	F	Y	98	4	48	2005
Computer Science, Software Engineering	Food Science and Technology	F	Y	81	5	27	2004
Agriculture	Soil Science,Physics,Fluids and Plasmas	F	Y	78	4	30	2005
Physics, Applied	Agriculture, Multidisciplinary	F	Y	70	3	29	2004
Materials Science, Characterization and Testing	Dentistry, Oral Surgery and Medicine	F	Y	61	2	30	2005
Imaging Science and Photographic Technology	Engineering, Geological	F	Y	47	2	31	2005
Social Sciences, Mathematical Methods	Biotechnology and Applied Microbiology	F	N	195	10	58	2005
Physics, Particles and Fields	Computer Science, Information Systems	F	N	121	4	68	2005
Remote Sensing	Engineering, Geological	F	N	55	2	36	2005
Biochemical Research Methods	Statistics and Probability	G	Y	4480	127	2111	2005
Computer Science, Theory and Methods	Neuroimaging	G	Y	646	9	248	2004
Cell Biology	Nanoscience and Nanotechnology	G	Y	452	23	164	2005
Biochemistry and Molecular Biology,Materials Science	Composites	G	Y	129	10	53	2005
Physics, Nuclear	Archaeology	G	Y	83	6	32	2004
Geosciences, Multidisciplinary	Parasitology	G	Y	72	3	28	2005
Horticulture	Allergy	G	Y	51	3	26	2005
Optics	Neuroimaging	G	N	208	12	92	2005
Engineering,Electrical and Electronic	Chemistry, Medicinal	G	N	137	9	33	2004
Physics, Fluids and Plasmas	Chemistry, Applied	P	N	187	9	73	2005
Telecommunications	Energy and Fuels	P	N	103	7	32	2005
Ornithology	Biophysics	P	N	67	2	29	2005
Materials Science, Composites	Biochemistry and Molecular Biology	N	N	266	7	155	2005
Engineering, Manufacturing	Surgery	N	N	171	11	58	2005
Automation and Control Systems	Cell Biology	N	N	166	9	55	2004
Computer Science, Theory and Methods	Urology and Nephrology	N	N	117	2	61	2004
Medicine, Research and Experimental	Applied Linguistics	N	N	112	6	36	2005
Engineering, Manufacturing	Cardiac and Cardiovascular Systems	N	N	97	4	29	2005
Materials Science, Ceramics	Microbiology	N	N	86	3	42	2005
Water Resources	Engineering, Biomedical	N	N	78	5	31	2005
Film, Radio, Television	Psychology, Experimental	N	N	68	2	29	2004

Table 1 continued

Citing category	Cited category	RF	R	N	Med	Max	Peak
Cardiac and Cardiovascular Systems	Integrative and Complementary Medicine	N	N	67	3	27	2005
Computer Science, Information Systems	Nanoscience and Nanotechnology	N	N	66	2	30	2005
International Relations	Psychiatry	N	N	65	4	33	2005

The RF column indicates the research focus, which can be either F, P, or G. The R column indicates whether or not the focus in the reciprocal pair is the same as one found in the listed pair. N contains the total number of citations between 1995 and 2005 using a 10-year wide citation window. The next columns deal with the distribution of the citations: the median and maximum value, and the year of the maximum value in the peak column

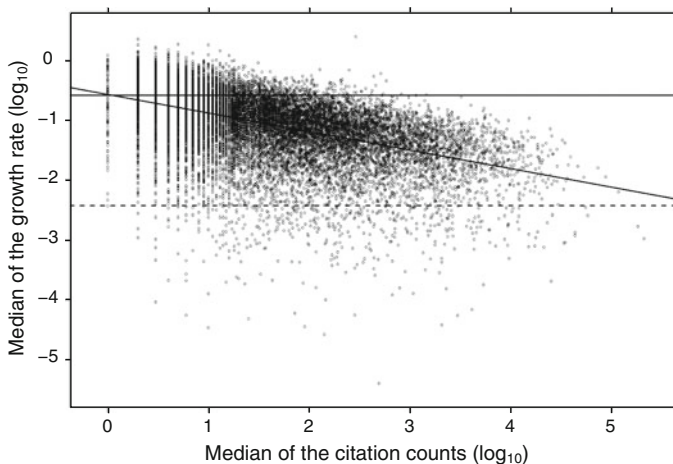


Fig. 2 The distribution of the logarithm of the median growth rate over the logarithm of the median size, for the pairs with at least four nonzero values for the growth rate series between 1995 and 2005. The diagonal line shows the trend of the distribution (linear model), the bottom, striped horizontal line shows the mean of the distribution, while horizontal line above it shows the 1.5 standard deviations above the mean

Research focus

A summary of the type of focus in the result is given in Table 2, which leaves out the 12 pairs that showed no convergence. This table shows that in half of the cases a clear focus is found, while the other half is either accompanied by other topics, or is of a more abstract nature. However, this does not appear to negatively impact the presence of the reciprocal focus.

Assessment of selected pairs

An exhaustive discussion of the content of the pairs listed in Table 1 is beyond the scope of this paper. Instead, we highlight pairs with interesting characteristics in an arbitrary order.

The first example is the pair *Economics* citing *Physics*, *Fluids* and *Plasma*, which appears to be part of the larger “Econophysics” convergence (Stanley et al. 1999). The

Table 2 A summary of the type of focus in Table 1

Focus	Found	Reciprocal
Focused	15	12
Partial	3	1
General	9	7
Total	27	20

The column Found contains the number of pairs, and Reciprocal contains the number of pairs that show the same focus in the reciprocal pairs

main source of the citations is the journal *Quantitative Finance*, and the citations are almost without exception going to *Physical Review E*. The best cited publication (counting only citations from *Economics*) is Plerou et al. (2002). On the reciprocal side, we find that most citations to *Economics* papers within the ten-year wide citation window are to publications from 2000 (as can be seen on the right half of Fig. 1). This suggests that the developments in *Physics* took place *before* those in *Economics*, which is confirmed when inspecting the content of the citing publications from those years. Also, those *Physics* papers again refer to even older publications in *Economics*, the most cited of which is Arthur (1994). We therefore conclude that this area shows an area of mutual influence and exhibits an independent, reflective nature.

Another pair, *Genetics and Heredity* citing *Communication*, is an example of societal interest, and the influence of the creation of a new journal. The focus deals with the communication of consequences of research in *Genetics*, such as ethical consequences and risks. This focus is also present in the reciprocal pair. The journal responsible for most citations is *New Genetics and Society*, which started to be covered by the WoS in 2000. Since the top-cited *Communication* publications such as Kerr et al. (1998) were already dealing with this topic, as well as citing *Genetics* publications, we could infer that the newly created journal may have provided a more focused publication stage in *Genetics and Heredity*, moving the publications away from *Communication*, while continuing to refer to relevant publications there.

The pair *Engineering, Industrial* citing *Agricultural Engineering* mainly deals with topics related to Biodiesel, and shows a large Indian presence in the research. Also, judging from the titles of the cited publications, the research in the pair shows a transfer from basic science to applied science. Therefore, we consider it as an example illustrating economical, national interests. Interestingly, the reciprocal pair shows no connection to the research in the detected pair, but instead deals with miscellaneous applied agricultural topics. Therefore, according to our definition, this interdisciplinary development cannot be regarded as converging research.

The use of topics from the humanities by the natural sciences, is visible in the pair *Physics, Nuclear* citing *Archaeology*. The research has a partial focus (which means that also other, unrelated subjects were found) on the application of physics methods to archaeological artefacts. This is also found in the reciprocal pair. However, the application is not a new development, as the journal *Archaeometry* (which plays an important role in the reciprocal pair) was already established in 1958. Also, on further inspection, a single special issue of *Nuclear Instruments and Methods in Physics Research B* on “Radiation and Archaeometry” (N1-2, V226), turns out to be the most important reason for the selection of the pair. We doubt that a single special issue may be enough to label this

example as converging research; instead, it may be an example of the import new tools from *Physics*, or alternatively, the export of specific problems to *Physics*.

As a last example we mention two related pairs: *Computer Science*, *Theory and Methods* citing *Neuroimaging*, and *Optics* citing *Neuroimaging*. Both pairs are part of the neuro-imaging and brain-imaging convergence that was also found in Buter et al. (2010), but are representative of two different (related) themes: research into computational aspects of imaging, and research into optics applied to neuro-imaging. The binding element is their cited knowledge base, because the top three cited journals is the same for both pairs: *Neuroimage*, *Human Brain Mapping* and *American Journal of Neuroradiology*.

Useful elements in assessment

We found a number of elements more informative than others in establishing a research focus. The titles of the citing and cited publications were the most important sources of research topics, as well as the spread of these topics over cited and citing publications. Important indicators for the *existence* of focus were the sizes of the journal matrices: if such a matrix contained a lot of cited (or citing) journals, then focus was usually absent. Other overviews, such as those of citing and cited affiliations, or document types of cited and cited publication, turned out less useful in this respect.

To further explore the usefulness of journal matrices as indicators for focus, we quantified the spread of citations over journals by calculating the Shannon entropy,⁷ for both the cited and the citing journals matrix. These two numbers were used as coordinates in the scatterplot in Fig. 3, where a circle indicates focus and a cross indicates no focus. Also, the size of the circle or cross is related to the total number of citations. From this figure we infer that there is a weak relation between the entropy values and the existence of a focus: below the diagonal that runs from (0,5) to (5,0), only focused pairs appear. Also, there appears no relation between the number of citations (size of a circle or cross) and the existence of focus. We consider this a useful first result and continue developing this indicator.

The nature of our results and converging research areas

The relatively small citation counts at the basis of our results challenges us to think about the nature of converging research. We hold that there are two important different developments to discern. First there is the, possibly multidisciplinary, application of problems or tools. Such a development is typically short-lived, as the application does not lead to any new or deeper insights, and the research community loses interest. In the second type of development the community keeps interested, and the research starts to show some level of independence from its “mother”-disciplines, both at the cognitive level and the social level. When successful, this development will result in an interdisciplinary or even transdisciplinary research area (Van den Besselaar and Heimeriks 2001). We consider the result of this second type of development representative for a converging research area.

Developments found by our search, will probably be the metaphorical tips of the iceberg. To establish what we have detected, more information is needed about the larger scientific surroundings, and we may have to apply background knowledge, possibly even provided by experts. To interpret the larger scientific surroundings, we need to apply or

⁷ The Shannon entropy is defined as $H = -\sum p_i \log(p_i)$, where p_i is the chance of event i , in our case the share of publications of a journal in the whole matrix.

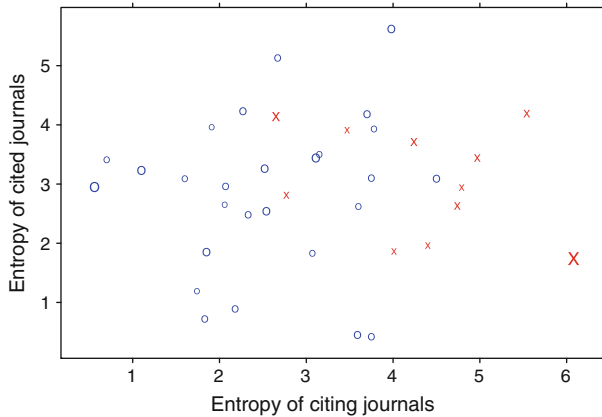


Fig. 3 The entropy of citing and cited journals used as coordinates for the 38 resulting pairs, which are represented as a circle if they were found to have a research focus, and as a cross if not. The size of a circle or cross corresponds to the number of articles in a pair. On the left imaginary line running from (0,5) to (5,0) only pairs with a research focus appear, illustrating the weak correspondence between entropy and research focus

even develop additional tools. At the social level, Berg and Wagner-Döbler (1996) hold that the structure of a research area can be seen as a combination of a “middle class” of authors around an established “prolific elite”. The elite have an important say in which themes are considered important and subsequently provide opportunity to gain status or impact for members of the “middle class”, as well as outsiders. Since citations also have a social dimension (Moed 2005), we therefore expect these structures to be visible in citation patterns, a fact which is also used in much of the work of Small (see e.g. Small and Upham (2009) for a recent example).

In the above description of example pairs, we have already established that some are indicative for larger, sustained research, like those representative of *Neuroimaging* and *Econophysics*. For the pair *Genetics and Heredity* citing *Communication* the nature is more difficult to assess, as it shows signs of an independent research area, since it has a specialized journal; however, to confirm this we would have to inspect the research in a more detail. The pair, *Physics, Nuclear* citing *Archaeology* is also challenging in this sense, because even though we regard a special issue of a journal as the expression of an accumulated (and thus sustained) interest in a specific topic, without further investigation we cannot establish whether this special interest was the start of more research.

We can also note the following with respect to the relation between the research we located and “Mode 2” research in the sense of Gibbons et al. (1994). Mode 2 knowledge production is the ability of a network of practitioners to produce knowledge, while the codification of this knowledge is of lesser importance and may even be “part of the network”. Such knowledge is difficult to capture in bibliometric terms. At the same time, Mode 2 research requires a “context of application”, which may very well be related to the research focus we tried to establish in the different phenomena.

Conclusion and future research

We described a high-level, top-down methodology for searching convergence between fields using journal subject categories as a proxy for fields and citations as a measure for

(interdisciplinary) application of research. A process was developed that consisted of two parts. In the first part, pairs of citing and cited fields were located using normalised citation counts as data and requirements with respect to growth, shape, and size. In the second part, the results were inspected, with the objective find research focus, i.e. the most specific common theme shared by most publications. We applied this process to WoS publication data for publications between 1995 and 2005. This resulted in 38 field pairs, which were inspected. After inspection, we found focus in 27 of the 38 pairs, and 20 pairs also showed similar focus in the reciprocal (reversed) pair. Also, interesting additional aspects were found in specific pairs, highlighting local, economical and societal interests, such as Biodiesel related research from India, or the ethical aspects of Genetics research.

Our search method has a number of clear advantages. First of all, it is data-driven, which makes it repeatable and applicable to new data, and even to other sources of data. Moreover, this makes the identification of converging research less dependent on the input of experts. Next, the method is easily adjustable as well as modular: search parameters such as the citation window can be tuned, and defining elements can be adapted. For instance, citation share could be calculated using the total number of citations obtained instead of given; or the journal subject categories could be replaced with another type of categorisation. Changing the required position of the peak may also be interesting, because a peak in the growth followed by enduring activity at a specific level may highlight developments that have managed to reach a steady state of activity.

Of course, the current implementation still depends on journal subject categories, which have well-known problems with respect to the delineation of related research. One problem is the coarseness, as there are only 243 categories to represent the whole of scientific research. One implication of the coarseness became apparent in the relation between size and growth shown in Fig. 2, which illustrated that the larger a category becomes, the more difficult it gets to detect emerging developments between that category and others. Another problem is that developments between fields captured in a single category, cannot be detected. By employing finer grained categorisations we may address such problems, provided the categorisations still represent (mostly) different fields. Another problem of the JSCs is that some established (interdisciplinary) fields may be covered by multiple JSCs. However, we consider this less of a problem, because if significant changes are detected between related journals in different categories, there is still something potentially interesting going on.

Another aspect of the current implementation is the sampling of reference counts on a yearly basis, which is crude and introduces large variance in the time-series. Potentially, this could be replaced by a monthly sampling; however, this would introduce other problems, such as those related to quarterly appearing journals, and those which are for example published in “Winter 2005”. Finally, the use of citations may present a problem of a more general, bibliometric nature: since most publications are published some time after the research was conducted, there is a delay between the use of the knowledge and the publication of that use, which could make it more difficult to detect trends that are in an early stage of development. Moreover, the time scales involved in the publication and citation processes differ in the various fields of science, as measured for instance by a field-specific age-distribution of references (Moed 2005). However, more research is needed before we can conclude what such effects have on the detection of interdisciplinary or (particularly) converging trends.

Future research will focus on improving the search method with respect to the points mentioned above. Additionally, the tools used to assess the results will be improved and extended, and parts of the assessment automated. The chart in Fig. 3 already provides some

clues to how this automating could take place. Maps could be useful as well, such as cognitive science maps (Buter and Noyons 2001, 2002), or network maps (Calero et al. 2006, Takeda and Kajikawa 2009). As noted above, we also need tools which are able to provide indicators for the level of independence of the research taking place in a set of publications. Once such a level of independence is established, other interesting questions could be analysed, such as the development of “spill-over”, i.e. the amount of knowledge that is developed in the convergence, but used in the originating, or other fields. A related question is whether a converged area, once it stops being active or attractive, would “dissolve” into the originating fields again, i.e. if the authors would return to their “mother”-disciplines, but keep on writing about the same issues (“annexation” by the originating field), or if the research into the topics completely stops. Finally, it may be interesting to see if some of the themes we found in the current application can be worked out in more depth and detail, not only with respect to the content of the research, but also to some additional bibliometric aspects, such as the development of the impact of converging research.

We hold that the value of our methodology lies in an interesting (apparent) paradox: scientific research is required to become more interdisciplinary in order to address complex societal, economical, technological and scientific problems, while at the same time researchers still tend to think in, organise in, and reward according to disciplinary lines. This tension provides a useful instrument, because research which does take the interdisciplinary route, is implicitly taking that tension into account, and may therefore be onto something useful or interesting. We think that our methodology can provide the basis for identifying those “Jacks of many Trades” that take up challenges and may start new convergences in order to address complex problems.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Arthur, W. B. (1994). Inductive reasoning and bounded rationality. *American Economic Review*, 84(2), 406–411.
- Berg, J., & Wagner-Döbler, R. (1996). A multidimensional analysis of scientific dynamics. *Scientometrics*, 35(3), 321–346.
- Buter, R. K., & Noyons, E. C. M. (2001). Improving the functionality of interactive bibliometric science maps. *Scientometrics*, 51(1), 55–68.
- Buter, R. K., & Noyons, E. C. M. (2002). Using bibliometric maps to visualise term distribution in scientific papers. In *International Conference on Information Visualisation*, IEEE Computer Society, Los Alamos, USA.
- Buter, R. K., Noyons, E. C. M., & van Raan, A. F. J. (2010). The identification of converging research areas using publication and citation data. *Research Evaluation*, 19(1), 19–27.
- Calero, C., Buter, R. K., Cabello Valdés, C., & Noyons, E. C. M. (2006). How to identify research groups using publication analysis: An example in the field of nanotechnology. *Scientometrics*, 66(2), 365–376.
- Carayol, N., & Thi, T. U. N. (2005). Why do academic scientists engage in interdisciplinary research? *Research Evaluation*, 14(1), 70.
- Cassirer, E. (1942). The influence of language upon the development of scientific thought. *The Journal of Philosophy*, 39(12), 309–327.
- Davidse, R., & Van Raan, A. (1997). Out of particles: Impact of CERN, DESY and SLAC research to fields other than physics. *Scientometrics*, 40(2), 171–193.
- De Boer, Y., de Gier, A., Verschuur, M., & de Wit, B. (2006). *Building Bridges: Researchers on their experiences with interdisciplinary research in the Netherlands*.

- De Mey, M. (1982). *The cognitive paradigm*. London: D. Reidel Publishing Company.
- De Solla Price, D. (1981). The analysis of square matrices of scientometric transactions. *Scientometrics*, 3(1), 55–63.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC.
- Egghe, L., & Rao, I. K. (1992). Classification of growth models based on growth rates and its applications. *Scientometrics*, 25(1), 5–46.
- Gambardella, A., & Torrisi, S. (1998). Does technological convergence imply convergence in markets? Evidence from the electronics industry. *Research Policy*, 27(5), 445–463.
- Gibbons, M., Limogenes, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge*. London: Sage Publications.
- Hinze, S. (1994). Bibliographical cartography of an emerging interdisciplinary discipline: The case of bioelectronics. *Scientometrics*, 29(3), 353–376.
- Kerr, A., Cunningham-Burley, S., & Amos, A. (1998). Drawing the line: An analysis of lay people's discussions about the new genetics. *Public Understanding of Science*, 7(2), 113–133.
- Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology*, 61(1), 126–131.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.
- Morillo, F., Bordons, M., & Gómez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13), 1237–1249.
- Morris, S. A. (2005). Manifestation of emerging specialties in journal literature: A growth model of papers, references, exemplars, bibliographic coupling, cocitation and clustering coefficient distribution. *Journal of the American Society for Information Science and Technology*, 56(12), 1250–1273.
- Nordmann, A. (2004). *Converging technologies—shaping the future of european societies*. EU HLEG Foresighting the New Technology Wave, September 14.
- Palmer, C. L. (1999). Structures and strategies of interdisciplinary science. *Journal of the American Society for Information Science*, 50(3), 242–253.
- Persson, G. (1999). Mistra's role and some lessons learned in the first five years. *Policy Sciences*, 32(4), 323–326.
- Plerou, V., Gopikrishnan, P., Gabaix, X., & Stanley, H. E. (2002). Quantifying stock-price response to demand fluctuations. *Physical Review E* 66(2), Part 2.
- Rinia, E. J., van Leeuwen, Th. N., Bruins, E. E. W., van Vuren, H. G., & van Raan, A. F. J. (2001). Citation delay in interdisciplinary knowledge exchange. *Scientometrics*, 51(1), 293–309.
- Rinia, E. J., van Leeuwen, Th. N., & van Raan, A. F. J. (2002). Impact measures of interdisciplinary research in physics. *Scientometrics*, 53(2), 241–248.
- Roco, M. C. (2002). Coherence and divergence of megatrends in science and engineering. *Journal of Nanoparticle Research*, 4(1), 9–19.
- Roco, M. C., & Bainbridge, W. S. (eds.) (2003). *Converging technologies for improving human performance*. Dordrecht: Kluwer Academic Publishers.
- Rosenberg, N. (1976). *Perspectives on technology*. New York: Cambridge University Press.
- Ruiz-Baños, R., Bailón-Moreno, R., Jiménez-Contreras, E., & Courtial, J. (1999). Structure and dynamics of scientific networks. Part I: Fundamentals of the quantitative model of translation. *Scientometrics*, 44(2), 217–234.
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3), 425–465.
- Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society Series B-Methodological*, 47(1), 1–52.
- Small, H. E., & Upham, S. P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, 79, 365–375.
- Stanley, H. E., Amaral, L. A. N., Canning, D., Gopikrishnan, P., Lee, Y., & Liu, Y. (1999). Econophysics: Can physicists contribute to the science of economics? *Physica A: Statistical Mechanics and its Applications* 269(1), 156–169.
- Takeda, Y., & Kajikawa, Y. (2009). Optics: A bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, 78(3), 543–558.
- Upham, S. P., & Small, Henry. (2010). Emerging research fronts in science and technology: Patterns of new knowledge development. *Scientometrics*, 83, 15–38.
- Van den Besselaar, P., & Heimeriks, G. (2001). Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators. In *Paper to be presented at the 8 th ISSI Conference for Scientometrics and Informetrics*, Sydney.