

The net as a knowledge machine: How the internet became embedded in research

Eric T. Meyer

University of Oxford, UK

Ralph Schroeder

University of Oxford, UK

Josh Cows

Massachusetts Institute of Technology, USA

Corresponding author

Eric T. Meyer, Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, UK. Email: eric.meyer@oii.ox.ac.uk

Abstract

In this paper, we examine the growth of the Internet as a research topic across the disciplines, and the embedding of the Internet into the very fabric of research. While this is a trend that ‘everyone knows,’ prior to this study no work had quantified the extent to which this common sense knowledge was true, or how the embedding actually took place. Using scientometric data extracted from Scopus, we explore how the Internet has become a powerful knowledge machine which forms part of the scientific infrastructure across not just technology fields, but also right across the social sciences, sciences, and humanities.

Keywords

Internet, World Wide Web, scientometrics, disciplinarity, knowledge

Citation information

Meyer, Eric T., Schroeder, Ralph, Cows, Josh. (2016, Forthcoming). The net as a knowledge machine: How the internet became embedded in research. Accepted for publication in *New Media & Society*.

The final, definitive version of this paper has been published in *New Media & Society* (Publication details forthcoming, <http://nms.sagepub.com/>) by SAGE Publications Ltd, All rights reserved. © The Author(s)

Introduction

The World Wide Web, and more broadly the Internet, started out life as a specialized network designed to help facilitate communication among a handful of networked academic and industry researchers in fields such as physics and military research. Over the last 25 years, however, the Web has become not only a topic of study in its own right, but also a core part of the toolkit of all areas of research, including computer science, medicine, natural science, social science, humanities and many other fields and subfields. It has, in other words, become a knowledge machine (Meyer and Schroeder, 2015) which consists of digital tools and data that are networked together and able to support research both about the Web itself, but also crucially about a whole range of additional topics which are enhanced because of the availability of these networked digital tools and data. This knowledge machine in turn is leading to the emergence of new topics and new approaches to research across a range of disciplines, acting as a research technology (Schroeder, 2007) that allows methods and ideas to move outside disciplinary silos. One might ask whether the Internet is the change agent that has allowed scholars to transform research, or whether the digitization of information is responsible. We would argue that, with regard to scholarly research at least, these two are tightly (and inextricably) coupled: digitization of information happened before the rise of the Web (e.g. CD-based or laserdisc resources of the late 1980s and early 1990s), but these alone had only limited transformative power. Likewise, had the Web developed into a purely commercial space (see, for instance, early closed web-like platforms such as Prodigy or AOL) without the growth in digitization of scholarly materials in the early 2000s (Meyer and Schroeder, 2009) and the increasing academic recognition of the potential scholarly value of the traces of online behaviour (Savage and Burrows, 2007; Savage and Burrows, 2009), we may have not seen the dramatic increases in internet- and web-related research we will be discussing in this paper. Taken together, however, these and the other elements of this vast knowledge machine are proving absolutely vital engines of the research enterprise today.

This paper will take a scientometric approach to analysing the growth, expansion, and influence of the Internet and the Web on scientific disciplines. In the last 25 years, over 300,000 academic papers have been written that explicitly mention some aspect of the World Wide Web or the Internet in the title, abstract or keywords of the publication. The overall pattern of this growth is one of nearly continuous linear upward growth since 1990. Of course, the Web and the Internet are not synonymous; however, for our purposes in this article we take a practical view that these two terms are widely used in relatively interchangeable ways in normal usage, and that inclusion of both terms more accurately reflects our aims to understand how the Web and the Internet have become integral parts of academic research.

The crude data available to anyone does not help us understand how the Internet itself has penetrated a variety of disciplines. In this paper, we extracted complete data from the 25-year period from 1990-2014 to analyse this growth, and broke it down into a number of elements: the change in disciplines over time, which nations' authors are writing papers in which the Web/Internet is a significant presence in the research, how international authorship and co-authorship changes over this time period, which journals have been most active in publishing topics related to the Web over time, and a number of other measures.

This research was inspired by the call for papers for this special issue. When we started to think about the Web at 25, it occurred to us that "everyone knows" that the Internet has penetrated into many corners of both everyday life and academic practice, but to date

nobody has tried to quantify this common knowledge (with some exceptions mentioned below that take a somewhat different focus). This paper shows the actual landscape of how the Web, or more generically, the Internet, is influencing research.

Literature

Elsewhere, we (Meyer and Schroeder, 2015) have argued that the Internet has become embedded in research in various ways. There are disciplinary differences (Fry and Talja, 2007; Kling et al., 2003) in this regard, but the commonality is that research is increasingly driven by shared and distributed digital tools and data, or what came to be known as e-Research or also as cyberscience (Nentwich and König, 2012). This wider process predates e-Research and has continued since, now that e-Research has become more routine, embedded and invisible, and in its most recent incarnation has become ‘big data’ (Schroeder, 2014) – though it can be anticipated that big data too will become invisible in time.

A different way to describe this process is that the Internet has become part of a research infrastructure (Hughes, 1983; Barjak et al., 2013), a large technological system that supports many aspects of research. These processes also mean that ‘Internet’ and some related search terms (such as ‘e-health’) should be keywords mentioned in articles. However, as we have noted previously (Meyer and Schroeder, 2015), not all research that uses digital tools and data (or that can be classified as such) makes explicit that internet- or e-infrastructure has been used. This lack of explicit referencing will become a wider problem when, as can be foreseen, ‘internet’ and ‘social media’ (another keyword) research, for example, becomes invisible when research on these topics will become part of media and communications research generally, and ‘internet’ and ‘social media’ will no longer necessarily explicitly be in the title or keywords for these articles.

Methods

One of the interesting challenges in work of this kind is that constructing an accurate and reliable search term to extract articles for analysis is more difficult than one might think. Rather than rely on guesswork, we used a systematic approach starting from search terms previously used by Peng et al. (2013), supplemented by some of the terms added by Malik (2012). Peng et al. retrieved 27,340 journal articles from the Web of Science for the period 2000-2009, focusing on the social sciences involved in ‘internet studies’, while Malik retrieved 114,079 social science publications from 1990-2011 from Web of Science using a somewhat broader search query.

In order to decide the most accurate search query (set of search terms) for this study, we started with these authors’ lists (Malik, 2012; Peng et al., 2013) and focused on terms related to our broader question related to the spread of the Internet into a variety of disciplines over time. By using Malik and Peng et al. as a starting point, we are starting from a set of pre-tested terms, rather than just selecting terms using a scattershot approach. We have chosen to use Scopus rather than Web of Science for a variety of reasons, most important of which is that the data extractable via our institutional subscription includes more fields necessary to answer our research questions than does Web of Science. The scale and scope of Web of Science and Scopus, however, are broadly comparable for the time period we are interested in, although Scopus has more limited coverage for the period before 1996. Scopus is a newer service than Web of Science, and originally included literature mainly from 1996 onwards. However, Scopus has since added more data from earlier time periods,

and, given the pattern of growth of internet-related publications, the lesser coverage of the earliest period is not significantly problematic here. Also, the coverage of Scopus is generally broader than that of Web of Science, as more sources are indexed, including additional coverage in the arts & humanities. Both sources, however, do have an inherent bias towards English-language publications, which should be taken into account when interpreting our findings; as we have also only used the English version of search terms and not searched for other language equivalents, we are admittedly underrepresenting related research in areas of the world where academic publication is in languages other than English, and in journals not indexed by Scopus.

The search terms we tested are shown in Table 1. Starting with the largest term ('internet') and then proceeding through each of the remaining terms (but excluding the largest term 'internet' so as to minimize duplication; see note in Table 1), we used the Scopus 'Advanced Search' option to query all journal articles published in the 25 years from 1990-2014 inclusive containing the target term in the title, abstract, or keywords of the article (e.g. 'TITLE-ABS-KEY(internet) AND DOCTYPE(ar) AND PUBYEAR > 1989 AND PUBYEAR < 2015'). Where more than one target term appeared in a single article, it was counted only once. The results were sorted with the top-cited articles first and a sample of 20 articles for each term was downloaded for analysis, followed by the top 20 conference papers ('DOCTYPE(cp)').¹ The sampling strategy used was to take every n^{th} article, where $n=100$ for search terms yielding 1500 or more results, and $n=10$ for search terms with fewer than 1500 results. The number of citations was used as the sorting mechanism for this exercise because it yields a more random selection of publications than the alternatives allowed by the Scopus interface. Alphabetical sorting by author, for instance, in the larger samples yields too many publications by authors with a last name starting with the letter 'A', which in turn tended to be based in a narrow range of countries because of ethnic differences in naming conventions. Sorting by date is inadequate because either older or newer publications are favoured, and sorting by 'relevance' relies on a non-transparent algorithm based on the search term itself.

Next, each article was qualitatively coded based on the title, abstract and keywords in relation to the search term(s) used. One person coded the 875 articles in the sample; as a measure of the reliability of the coder, a 10% stratified random sample ($n=90$) of the first coder's results were coded by a second person who was blinded to the first codes assigned. The two coders were in agreement 90% of the time ($n=81$), with a Krippendorff's alpha score of 0.79; scores above 0.8 are considered highly reliable, and between 0.667 and 0.800 allow one to draw reasonable conclusions (Krippendorff, 2004: 429). Since the coding exercise was to determine a range of keywords to use, and not an end in itself, falling slightly below the highest threshold was not considered problematic.

Table 1. Search Term Specificity

Search topic	Search terms	Articles coded	% on-topic	% on-topic or marginal	Total articles
Myspace	myspace	40	100.0%	100.0%	422
Web 2.0	{web 2.0}	40	100.0%	100.0%	6292
Cyberspace	cyberspace or “cyber space”	40	87.5%	97.5%	2071
Facebook	facebook	40	97.5%	97.5%	5211
LinkedIn	linkedin	39	97.4%	97.4%	396
Flickr	flickr	40	95.0%	95.0%	1485
Twitter	twitter	40	95.0%	95.0%	5500
eGovernment	{e-government} or egovernment	40	40.0%	92.5%	5692
Social Media	{social media}	40	92.5%	92.5%	10253
Wiki	wiki*	40	92.5%	92.5%	7371
eBay	ebay	40	87.5%	87.5%	565
eHealth	{e-health} or ehealth	40	17.5%	87.5%	3564
Internet	internet	40	52.5%	87.5%	273882
Youtube	youtube	40	85.0%	87.5%	2008
Blog	*blog*	40	82.5%	82.5%	9161
eCommerce	{e-commerce} or ecommerce	40	80.0%	82.5%	12224
World Wide Web	{world wide web}	40	77.5%	80.0%	45754
Terms below not used in final search					
New Media	“new media”	36	66.7%	69.4%	6203
Web	web	40	42.5%	47.5%	220232
Google	google	40	40.0%	40.0%	10244
Online	online OR “on line”	40	32.5%	40.0%	277782
Social Network	“social network*”	40	27.5%	30.0%	53514

A three-category system was used in which ‘on-topic’ was taken to be those articles and conference papers which had some aspect of the Internet, broadly interpreted, as a key part of their work as indicated by the title, abstract, and keywords. ‘Off-topic’ articles were those which were captured by the search term, but were not in any meaningful way focused on the Internet. Examples include passing references to searching Google Scholar for a literature review, mentioning that a resource could be located online or on the Web at such and such an address, or pointing out that the project had a blog available. Some terms with particular problems of this nature were discarded before this stage because of their low reliability, such as ‘www’ which gathers any abstract listing a URL regardless of topic. Other publications were coded as ‘off-topic’ because the search term also yielded publications with no relation to the Internet; a key example of this is ‘social network*’ (where * indicates a wildcard, thus encompassing ‘social networks’, ‘social networking’ and so forth) which includes articles on social network analysis, a large and well-developed field of study which in the majority of cases has nothing to do with the Internet.

The third intermediate category ('marginal') is more open to interpretation. These were publications in which the Internet did not play an absolutely central role, but were arguably at least more than passingly dependent on the Internet in their content and/or methods. Two of the areas that have a particularly large number of papers than fell into this grey zone were health and government. In the health area, a number of papers that focus more on patient records systems, for instance, than on online access to those systems fall into this category, as do publications on medical research that communicated with participants online. Likewise with government, IT implementations that were designed to improve 'e-government' but were not primarily public-facing fall into this category. Note also that the term 'Internet' itself had a large number of results in this marginal category, largely due to a high number of results that coincidentally were in the health sector and were difficult to clearly classify as in or out of scope for the reasons above (the percent of publications related to the internet for the term 'Internet' increases from 52.5% to 87.5% when including these marginal cases). Some specific examples from this marginal category include descriptions of several scientific databases (such as metabolic pathways and e. coli data) that are available for download from websites, an article about building e-government applications on the Grid, articles about trust in information systems that are connected to the Internet in both the health and government sectors, and other similar topics. While these are not 'about' the Internet or the Web, it is arguable that the Internet plays a vital part in the work reported, and thus should be considered on-topic for our purposes.

In the end, we decided to include in our final search those terms which required the broader interpretation of relevance, and set our cut-off as 80% of the sample coded as relevant or marginal. This included the terms for e-government, e-health, Internet, and World Wide Web, which otherwise would have fallen below this cut-off. We have aimed for inclusivity in the final set of search terms, to cover all areas and disciplines as much as possible in the final dataset.

We include this somewhat extended discussion of our sampling method for two reasons. First, few previous studies have taken a systematic approach of this sort to decide whether to include or exclude particular search terms from their scientometric study, or if they have, they have not documented it in their methods. We feel that this approach is robust, and could be further developed and extended. Second, the accuracy and specificity of the search terms is itself interesting in the context of this article, because it suggests both how difficult it is to accurately separate the key role of the Internet in research (either as a topic or a centrally enabling technology) from the more passing references to various parts of the Internet's infrastructure that have become common as the Internet has become ubiquitous.

The final search term² yielded 334,659 documents. As can be seen, some of the tables have smaller numbers than that because of issues such as missing data, including missing country data for many authors. Where this is the case, we have noted it.

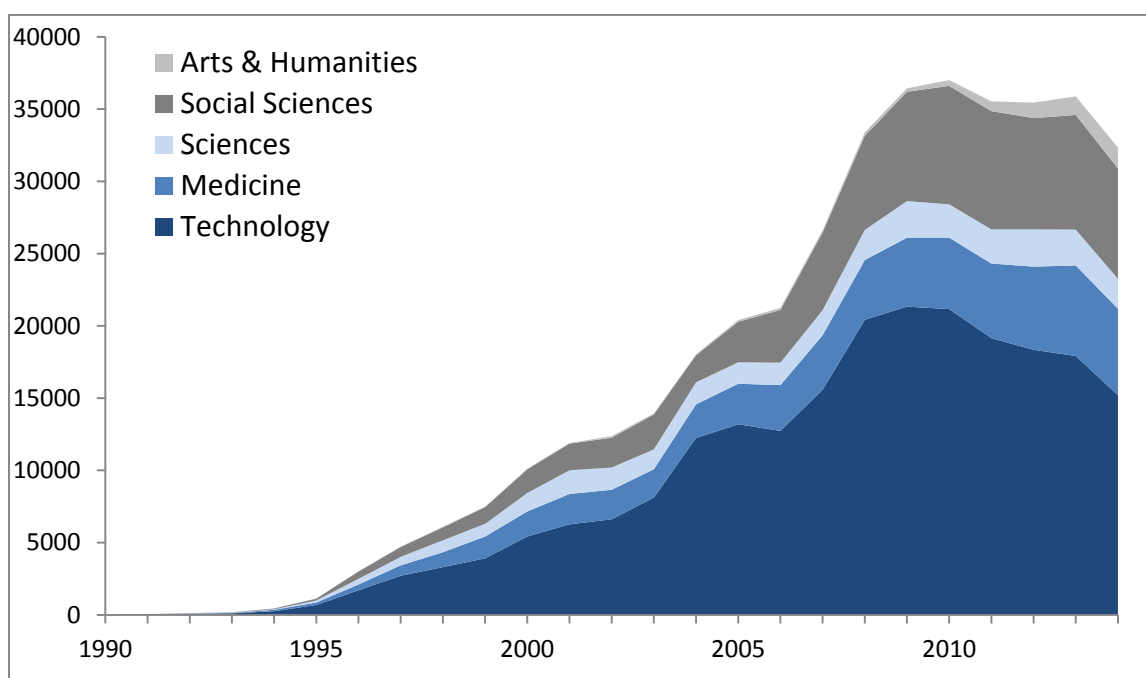
The data were downloaded from Scopus, and then processed in a database using bespoke VBA code to extract authors, author nationality, and other details.

Findings

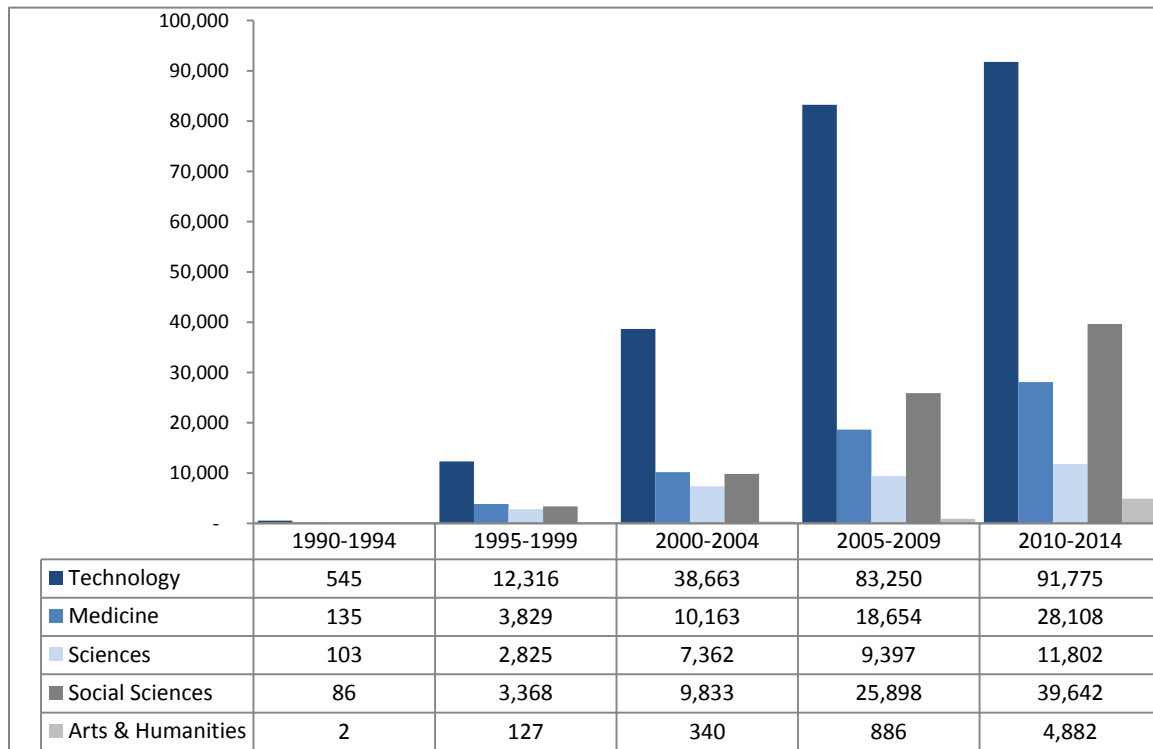
Overall trends

We first can see the overall trend of the Internet having an increasing influence on research over the last 25 years. In Figure 1, we can see the overall growth by year in five major topic areas: Technology (which includes computer science, engineering, and mathematics), Medicine (which includes medicine, genetics, immunology, neuroscience, nursing, dentistry, pharmacology, and the health professions), Sciences (which includes biology, chemistry, earth sciences, materials science, physics, astronomy, and agriculture), Social Sciences (which includes the social sciences, psychology, economics, business and decision sciences), and the Arts & Humanities.

Figure 1. The Internet in Major Research Areas (n=403,991)



The story here is clearly one of growth, and one that makes sense given our lay understanding of how the Internet started coming to public attention in 1995 with the growing availability of Netscape: there has been steady growth across all areas from 1995 to 2010. As noted, Scopus coverage is less complete for pre-1996 publications than for those after, so the early data should be considered incomplete. Less clear is the apparent flattening of growth over the last five years. While the apparent drop off in 2014 is likely attributable to not-yet-updated data in the Scopus database,³ the flattening curve from 2010-2013 is more likely to be an actual trend and not a data artefact, although it is more marked in the technology domains than in the other knowledge domains that we focus on for much of this paper.

Figure 2. The Internet in Major Research Areas, 5-year bins (n=403,991)

In Figure 2, the same data are displayed, but this time grouped into five-year bins. Another aspect of the story appears here: while the predominance of computer science, engineering and mathematics is clear throughout the 25 years in question, the social sciences overcame a late entry into engagement with the Internet to then surpass the activity in medicine and the sciences. In the 2005-2009 and 2010-2014 periods respectively, the social sciences' engagement with the Internet in research in this sample grew by 163% and 53%. It is worth noting that some of this growth could be related to the growing prevalence of some of the platforms like social media sites included in our search term being of particular interest to researchers in the social sciences within the wider context of Internet research. However, these make up a relatively small proportion of the overall sample, and the pattern holds even with the broadest search terms such as 'internet' that are relevant across the disciplines and the sample. Nonetheless, it would be remiss not to consider the changing nature of the Internet itself over the past 25 years – in particular, its adoption as a popular communication technology – and the implications of these broad trends in driving related research.

Even later to the game are the arts & humanities, which showed relatively little engagement until the most recent 5-year period, when there is a 451% increase over the previous period (4,882 publications compared with 886 in 2005-2009). One thing to note is that both Scopus and Web of Science have inherent limitations with regard to the humanities, which means that humanities publications are under-represented, compared to their actual frequency. Thus, it is difficult or impossible to compare the arts & humanities data directly to the other fields, but the marked growth within the arts &

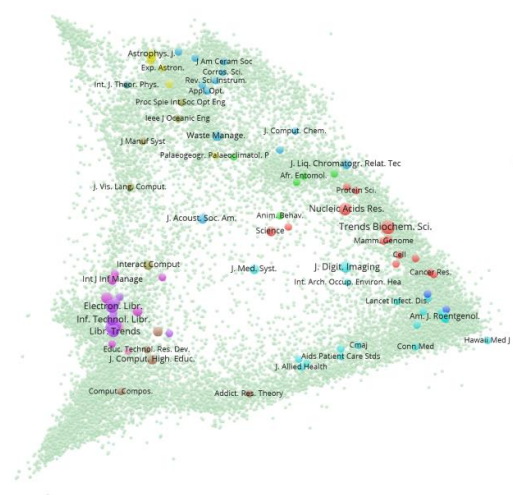
humanities data in 2010-2014 is nevertheless noteworthy and not purely due to increased data coverage.

The Internet: Spreading across the Disciplines

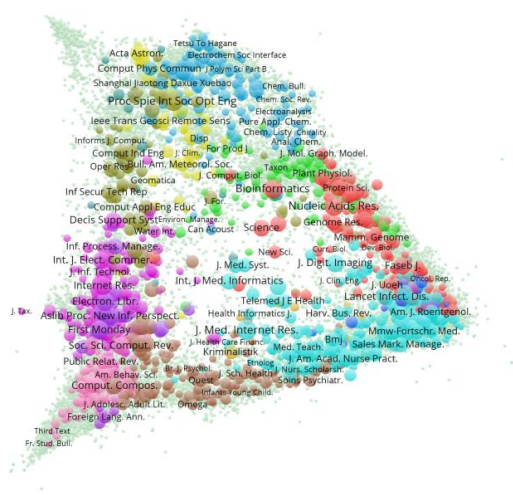
For most of the remainder of the paper, we will focus on data from the domains other than technology, which because of its relatively large size tends to obscure the results in the smaller domains. Thus, the following data exclude articles solely in the computer science, engineering, and mathematics domains (although some of the multidisciplinary journals also link into those domains). More importantly, in this paper we are primarily interested in how the Internet has influenced research across the sciences and humanities, rather than in how technological developments underpinning the Internet have grown and been communicated. The latter are also interesting questions, but beyond the scope and focus of this paper.

In Figures 3a-3e, we can see a visual representation of the spread of internet-related topics across standardized overlays of journals using the methods described in Leydesdorff et al. (2013) and Leydesdorff et al. (2015). The method Leydesdorff and his colleagues have used to create the underlying map in these figures involves downloading citation information in all 19,600 journals indexed by Scopus from 1996-2012 and analysing the matrix of all citations between journals to calculate the distance between every possible pair of journals. Journals which cite each other frequently are likely related in terms of topic and discipline, and are thus mapped closely to each other based on cosine normalized citation. The resulting map thus has all of Scopus-indexed knowledge displayed, with disciplines and fields naturally clustered in various portions of the map (described further below). As the authors explain, “interactive overlay maps enable users to project a set of documents onto a base map in terms of the journal distribution” (Leydesdorff et al., 2015: 1001). Using this standardized data provided by Leydesdorff as the underlying map, we then extracted the journal names from our dataset described above and processed them with Leydesdorff’s tools available on [\[http://www.leydesdorff.net/scopus_ovl/\]](http://www.leydesdorff.net/scopus_ovl/). These tools match our Scopus data with the standardized map of science and overlay the journals in our sample on top of the base map. The underlying map of gray dots thus represents ‘all of science,’ or at least as much of it as is represented in 19,600 journals indexed by Scopus, while the coloured dots represent the science, social science, medicine and humanities journals in which articles related to the Internet appear in each of our five 5-year periods. The maps are displayed using VOSviewer, in which “the size of each journal as a node is depicted proportionally to the $\log_4(n + 1)$ with n as the number of occurrences. (The “+1” is added to prevent single occurrences from being displayed, because $\log(1) = 0$)” (Leydesdorff et al., 2015: 1007). The colours are assigned by VOSviewer using a community detection algorithm (Waltman et al., 2010).

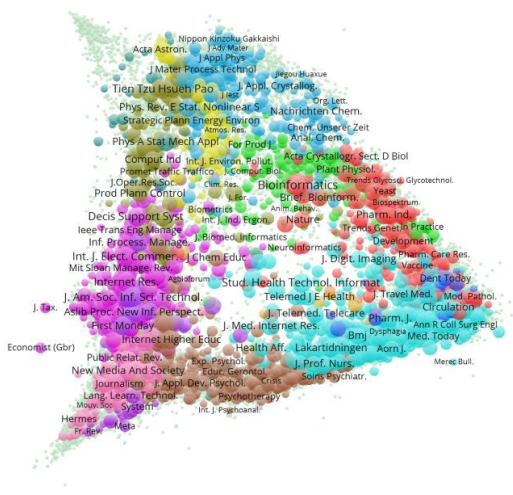
Figure 3a – 3e. Spread of Internet-influenced research in domain-based journal articles (n=81,164 in 9,995 journals) related to science, medicine, social sciences, and the humanities, 1990-2014 on standardized overlays (n=19,600 possible journals) of journals (excludes computer science, engineering and mathematics)



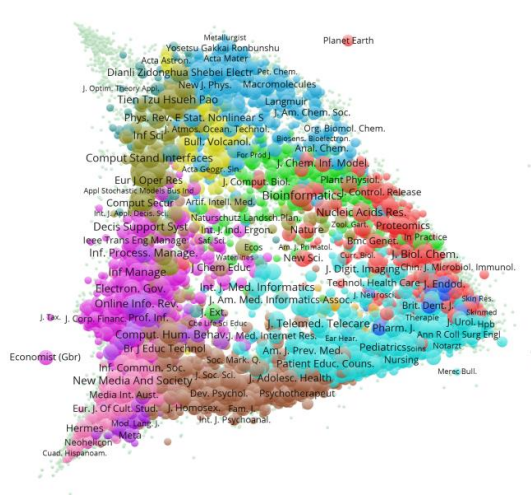
3a. 1990-1994 (n=144 articles in 93 journals)



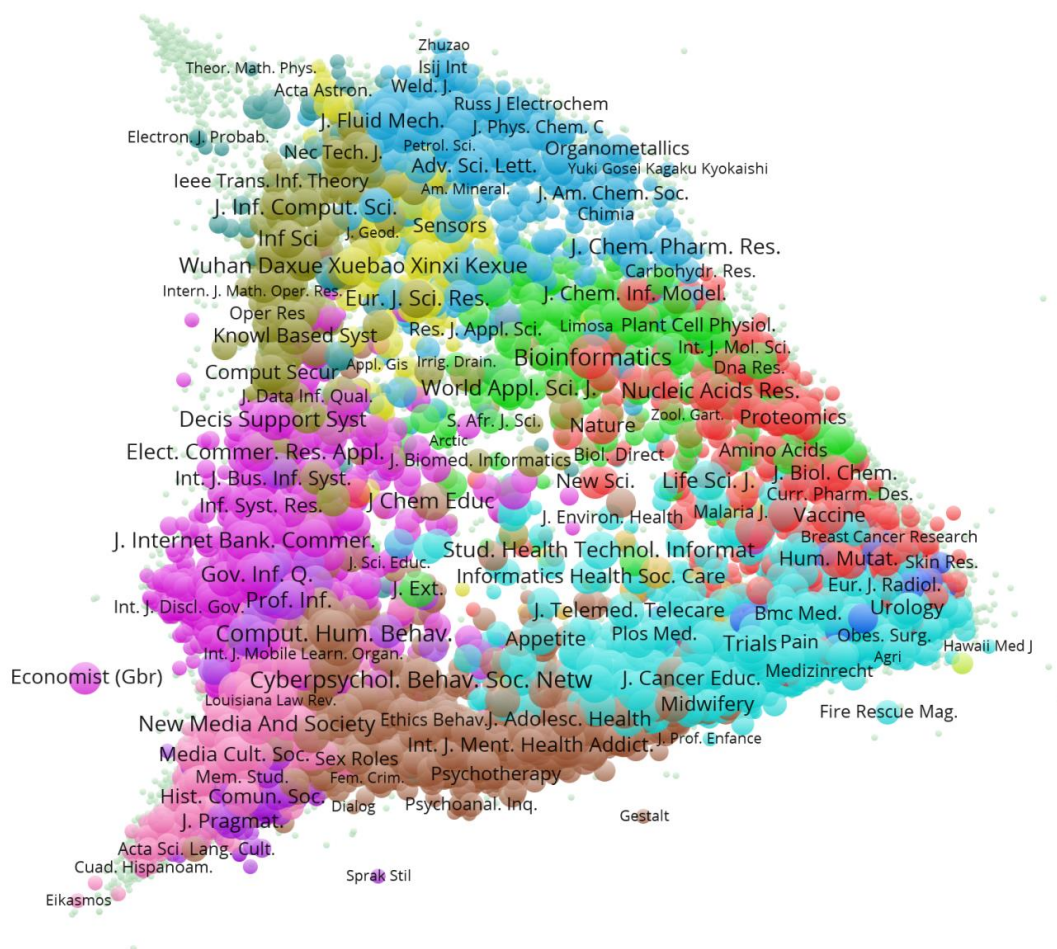
3b. 1995-1999 (n=4,918 articles in 1,518 journals)



3c. 2000-2004 (n=13,356 articles in 3,342 journals)



3d. 2005-2009 (n=25,377 articles in 5,760 journals)



3e. 2010-2014 (n=41,772 articles in 7,807 journals)

There are several narratives that become apparent in these visualizations. The first pertains to the spread of internet-related research, which originally emerged scattered quite widely across the disciplines, as shown in figure 3a. In figure 3b, we see a marked increase in publications (consistent with figures 1-2 above and tables 2-4 below), and these publications are scattered across the underlying map, which indicates that these topics from the very start were not constrained to a small range of disciplines. In each subsequent figure, we see that the coverage spreads and the publication frequency (indicated by size) intensifies in each 5-year period. Finally, in figure 3e we have the most recent 5-year period from 2010-2014, which is based on 41,772 journal articles published in 7,807 journals. By this time, there are few areas of academic publishing without some activity related to the Internet, with (roughly speaking) the sciences in the upper right, social sciences on the left, medicine on the right, and humanities in the lower left. Note that VOSviewer does not display all titles at once in the interests of making the presentation more readable; in the live VOSviewer interface one can zoom in and see the detail for any given dot (data available from the authors on request).

As indicated above, the underlying map is based on 19,600 journals in Scopus, so Internet-related publishing represents at least one article in 51% (n=9,995) of all journals in Scopus, and the number would be even higher if we were including the technology fields in our analysis, even though a high percentage of those publications are published as conference papers rather than as journal articles. Other than the Internet, has any other topic ever appeared in half of all academic journals over the course of a quarter of a century? If so, we can't think of it. This sort of penetration of the literature is both remarkable and, as far as we know, unprecedented.

We are not claiming, of course, that this penetration is evenly distributed or absolute. Many of the journals have only one or two articles in our sample. If you plot the number of articles per journal on the y-axis against the ranked order of the journals along the x-axis, there is a marked long-tail curve in each of the main areas of research. There are slight differences in the maximum value and the number of journals in each area, but essentially all four look approximately the same (which is why we haven't included them here): the same long-tail curve is shown time and again in Internet phenomena: a very tall y-intercept, with a steep drop and curve to a very long tail of thousands of journals with a small number of publications each.

In Tables 2, 3, and 4 we look at some of the top journals publishing internet-related research over the last 25 years. The topic areas of articles are determined by Scopus categorizations at the journal level, so the match is not perfect but gives a reasonably good indication of the types of topics published in a given journal. However, this is also why a few journals appear in more than one of these three tables.⁴

We have not included a similar table for the humanities because the data here are really too small in most cases to build a comparable set of data (the 20th ranked journal in terms of volume, for instance, published only 31 papers related to the Internet over this 25 year period). Generally speaking, however, the humanities data in the similar table we did not include tells a story of very little activity at all from 1990-2010, with a big increase from 2010-2014 (as is consistent with Figure 2 above).

The Internet in the Social Sciences

Table 2, which will likely be of most interest to readers of this journal, looks at journals related to the social sciences. The top twenty journals as measured by number of articles (with a secondary sort on average number of citations in the case of journals tied in the rankings) are shown here. The order of the table is based on the overall 25-year period, with the number of articles and mean number of citations within our dataset shown for the overall period, and then each 5-year period in subsequent columns. Also shown are the mean number of citations to articles in this sample for each journal, and in the overall data, the percent of articles that were cited at least once and the percent which were cited at least ten times (the i10 index, see Kozak and Bornmann, 2012). The i10 index is a more feasible indicator than an h-index for this particular study, and is an effort to highlight the proportion of papers which have been cited a reasonably large number of times.

Table 2. Top 20 Social Science-related Journals for Internet-Related Research, by total number of articles published from 1990-2014

Source title	1990-2014 Overall					1990-1994			1995-1999			2000-2004			2005-2009			2010-2014			
	Rank	n	Cited %	i10 %	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	
Computers in Human Behavior	1	891	88%	50%	24.2				87	8	68.3	10	66	70.8	2	200	38.3	1	617	12.1	
Cyberpsychology and Behavior	2	496	99%	73%	38.9				10	35	59.7	1	150	42.4	1	311	35.0				
First Monday	3	445	61%	21%	9.0				1	66	6.7	9	69	9.5	11	94	13.7	3	216	8.1	
Computers and Education	4	396	95%	67%	27.8				16	29	14.3	28	42	36.9	4	146	39.7	7	179	17.2	
Decision Support Systems	5	385	94%	64%	33.7				21	23	56.2	4	75	44.8	3	151	43.2	8	136	9.8	
New Media and Society	6	377	93%	58%	23.3			1	2.0	125	6	21.2	25	43	47.9	7	121	29.5	4	207	13.6
Internet Research	7	367	94%	55%	22.5					2	61	13.4	3	116	34.1	9	97	25.5	23	93	9.0
Government Information Quarterly	8	346	91%	59%	30.7					35	16	9.9	38	35	68.7	8	99	47.7	6	195	14.4
Cyberpsychology, Behavior, and Social Networking	9	336	85%	36%	11.2														2	336	11.2
Online Information Review	10	326	87%	39%	12.0					2	2.5	211	3	21.0	8	70	9.5	5	142	17.1	14
Electronic Library	11	290	79%	22%	6.4	1	7	1.8	4	49	4.8	6	72	6.4	10	97	7.7	54	65	5.7	
Information Communication and Society	12	273	84%	32%	10.7	1						2	14.5	24	64	18.3	5	207	7.8		
J of the American Society for Inform. Sci. and Tech.	13	248	98%	72%	35.0				2	123	39.5	6	125	30.6							
Social Science Computer Review	14	247	87%	43%	17.8				5	48	10.6	22	47	36.1	31	54	20.3	20	97	8.3	
Information Processing and Management	15	237	95%	58%	26.4				23	21	22.9	11	60	49.1	13	89	24.8	49	67	7.0	
Telecommunications Policy	16	223	89%	46%	16.8	6	3	11.0	19	25	21.2	32	38	28.5	23	64	18.9	24	93	7.5	
Information Sciences	17	217	92%	52%	17.7	1	1.0		48	12	15.8	46	31	26.7	18	78	21.2	21	96	11.7	
Journal of Information Science	18	212	86%	43%	16.9				6	45	10.6	13	59	23.0	36	51	22.8	64	56	8.8	
Electronic Commerce Research and Applications	19	208	86%	48%	19.2							68	23	41.7	16	83	21.5	17	102	9.9	
Internet and Higher Education	20	195	93%	60%	29.3				34	17	3.9	7	70	37.0	37	50	40.0	60	58	17.9	
Total articles	61309					123			3800			9722			18664			33066			
Total journals	5320					42			751			1549			2838			4206			

Again, there are a few things worth noting about this table. First, some of the journals which are most active in our sample are relative newcomers to the publishing scene. *Computers in Human Behavior* was first published in 1985 and *Cyberpsychology and Behavior* (continuing as *Cyberpsychology, Behavior, and Social Networking*) was first published in 1998. *New Media and Society*, which appears 6th overall and 4th in the most recent time period in terms of volume, published its first issue in 1999. Overall, at least 61,309 articles were published in 5,320 journals related to the social sciences from 1990-2014, with 33,066 of those articles published in the last five years alone.

Also worth noting is the general pattern of growth in prominence for certain journals over time. While *New Media and Society* experienced considerable growth in prominence during the last fifteen years (ranked 25th from 2000-2004, 7th from 2005-2009, and 4th from 2010-2014 in terms of volume), the online journal *First Monday* has been in the top 15 journals since its inception. Of course, volume isn't the only, or even necessarily a very good, measure of impact, and *New Media and Society* papers are cited more frequently (with 93% of papers cited and these receiving an average of 23.3 citations each) than *First Monday* (61% of papers cited 9.0 times on average), and both are less cited than articles in *Cyberpsychology and Behavior* (with 99% cited an average of 38.9 times each) and the *Journal of the American Society for Information Science and Technology* (JASIST) (in which 98% of papers are cited an average of 35.0 times). JASIST, it should be noted, changed names twice in this time period, which is why it only appears to have been published from 2000-2009.

The Internet in Medical Research

Table 3 reports similar data, but this time for medicine-related journals (55,322 articles in 6,495 journals). *Nucleic Acids Research*, which has published many highly-cited articles about online data banks such as PROSITE and SWISS-PROT among many others, is top-ranked in volume (n=3,272), in proportion of papers cited (98%) and in average citations per paper (97.6). This prominence is consistent across the time period. *PLoS ONE*, on the other hand, was only first published in 2006, but then immediately becomes a major outlet for internet-related medical research.

Again, we see some movement over time, as certain journals such as *Telemedicine and e-Health* become much more active in recent periods, while others such as *Studies in Health Technology and Informatics* wax and wane, ranked in the top 15 in volume for three of the time periods, but falling to 109th from 2005-2009, and always with a relatively modest 46% of papers being cited an average of 3.3 times each.

Table 3. Top 20 Medicine-related Journals for Internet-Related Research, by total number of articles published from 1990-2014

Source title	1990-2014 Overall					1990-1994			1995-1999			2000-2004			2005-2009			2010-2014		
	Rank	n	Cited %	i10 %	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites
<i>Nucleic Acids Research</i>	1	3272	98%	78%	97.6	3	5	20	1	131	98.0	1	680	177.3	1	1248	98.5	1	1208	50.2
<i>Bioinformatics</i>	2	1618	96%	62%	72.0				3	71	322.7	2	422	73.6	2	724	73.9	4	401	18.1
<i>Journal of Medical Internet Research</i>	3	1057	77%	33%	16.9				2	108	6.0	6	75	32.5	6	137	42.8	3	737	9.4
<i>PLoS ONE</i>	4	1015	84%	29%	14.1										16	57	41.3	2	958	12.1
<i>BMC Bioinformatics</i>	5	912	95%	61%	35.9							3	106	82.6	3	547	36.4	6	259	12.3
<i>AMIA Symposium</i>	6	396	60%	8%	5.1										4	355	5.3	57	41	3.9
<i>Cyberpsychology, Behavior, and Social Networking</i>	7	325	85%	30%	10.9													5	325	10.9
<i>International Journal of Medical Informatics</i>	8	298	96%	52%	18.2				9	30	11.9	8	52	21.9	8	111	25.7	16	105	9.6
<i>Journal of Telemedicine and Telecare</i>	9	295	89%	29%	9.5				26	18	12.8	5	83	11.6	7	127	9.5	36	67	5.1
<i>Telemedicine and e-Health</i>	10	288	69%	13%	6.6							465	4	10.5	13	68	12.1	7	216	3.8
<i>BMC Genomics</i>	11	263	91%	38%	13.9							300	6	4.5	11	97	21.5	10	160	9.2
<i>Studies in Health Technology and Informatics</i>	12	257	46%	2%	3.3				14	24	5.2	4	90	3.8	99	19	4.5	11	124	2.0
<i>Journal of the American Medical Informatics Association</i>	13	235	87%	51%	26.6				4	51	15.4	18	36	48.4	17	56	41.8	22	92	10.8
<i>BMC Public Health</i>	14	228	87%	29%	11.7							338	5	29.3	23	47	22.8	8	176	7.8
<i>Medical Reference Services Quarterly</i>	15	211	62%	4%	3.8		1		6	43	3.3	10	49	3.3	15	58	4.3	42	60	3.8
<i>Telemedicine journal and e-health</i>	16	204	55%	13%	6.7							162	9	16.3	65	26	6.6	9	169	5.7
<i>Medical Teacher</i>	17	191	86%	34%	12.9				122	5	7.0	20	34	15.6	12	80	16.4	28	72	7.6
<i>IEEE Engineering in Medicine and Biology</i>	18	188	28%	0%	2.1										10	104	1.8	23	84	2.4
<i>Patient Education and Counseling</i>	19	187	94%	52%	18.5		1	8		1	51.0	156	9	39.2	14	63	28.0	13	113	10.8
<i>Journal of Biological Chemistry</i>	20	175	100%	85%	24.5										9	107	28.5	34	68	18.1
Total articles	55322					121			3230			8821			16644			26506		
Total journals	6495					74			1075			2225			3434			4527		

Table 4. Top 20 Science-related Journals for Internet-Related Research, by total number of articles published from 1990-2014

Source title	1990-2014 Overall					1990-1994			1995-1999			2000-2004			2005-2009			2010-2014								
	Rank	n	Cited %	i10 %	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites	Rank	n	Mean Cites						
<i>PLoS ONE</i>	1	1001	84%	35%	14.2										4	57	41.3	1	944	12.2						
<i>Computers and Geosciences</i>	2	238	89%	35%	11.6										3	67	9.3	5	37	15.3	8	45	17.6	6	89	8.2
<i>Acta Electronica Sinica</i>	3	219	63%	7%	3.2										1			1	112	3.3	1	106	3.2			
<i>Physical Review E</i>	4	200	95%	73%	71.4													2	72	149.6	2	76	31.0	14	52	10.5
<i>Journal of Chemical Education</i>	5	165	77%	17%	6.1													25	7	12.9	12	28	9.0	13	34	6.8
<i>Physica A</i>	6	152	93%	47%	44.0	1	3.0		88	3	452.7	3	45	80.6	6	48	17.7	12	56	9.1						
<i>Proceedings of SPIE</i>	7	119	48%	26%	8.6				1	93	8.5	42	11	16.5				97	14	3.9						
<i>Journal of Lightwave Technology</i>	8	117	97%	70%	39.5				3			86.3	4	40	68.8	5	53	23.5	51	21	15.4					
<i>European Journal of Scientific Research</i>	9	107	50%	4%	3.0													2	107	3.0						
<i>Sensors</i>	10	95	66%	5%	3.9										1			7.0	5	94	3.9					
<i>Geomatics and Info. Science of Wuhan University</i>	11	95	33%	0%	1.7	1	4	2.5										4	95	1.7						
<i>Photonic Network Communications</i>	12	85	74%	29%	16.7							7	31	24.6	17	32	11.7	49	22	8.0						
<i>SIGCSE Bulletin</i>	13	83	73%	16%	5.8				2	68	4.1	43	11	14.7												
<i>The Scientific World Journal</i>	14	83	54%	2%	2.3													7	83	2.3						
<i>Vaccine</i>	15	82	93%	47%	15.8							1	7.0	84	7	26.6	47	15	22.4	10	59	12.8				
<i>Genome Biology</i>	16	78	91%	79%	70.3	1	1.0					85	7	23.5	7	47	68.5	39	24	83.3						
<i>IEEE Transactions on Instrumentation and Measurement</i>	17	77	95%	52%	18.0				19	8	27.4	31	14	34.7	10	42	13.5	105	13	5.8						
<i>Computers and Electronics in Agriculture</i>	18	74	92%	53%	16.5				1			4.0	26	16	24.8	42	17	22.6	22	39	10.2					
<i>Advanced Science Letters</i>	19	73	12%	0%	2.3													8	73	2.3						
<i>Journal of Electronics and Information Technology</i>	20	71	52%	0%	2.6										78	8	2.0	3	63	2.6						
Total articles	19190					52			1600			3349			5363			8826								
Total journals	3335					42			652			1134			1635			1945								

The Internet in Science

Finally in table 4 we look at science-related journals. Again, *PLoS ONE* makes an appearance at the top of the table (remember, journals are often classified in more than one category), with 1001 articles published, 84% of them cited an average of 14.2 times. Some of the entries here are obvious homes for technology-related publications (such as *Computers and Geosciences*), but many are mainstream science journals in disciplinary topic areas that, because of the penetration of the Internet into these scientific domains, also publish work related to the Internet. This is not ‘internet studies’, but the Internet as part of the research infrastructure of science.

The Geography of the Internet as a Knowledge Machine

We have been seeing the spread of the Internet as a topic of research spreading across the disciplines over the last 25 years, but what of its spread across the globe over the same time period? In Table 5, we have again used rank by volume, but this time sorting by the country of the corresponding author for the article.⁵ In each time period, the countries were ranked (although from the 1990-1994 period, too many journals had equal counts of between 1 and 4 articles to meaningfully rank them in relation to each other, so rank is excluded for journals with fewer than five publications).

The United States, unsurprisingly, is the most prominent country in Internet-related research, but the same can be said of many other areas of research as well. The United States has a large and prolific research culture, and so published nearly four times as many articles (38,180) as its next closest rivals, China (10,686 articles) and the United Kingdom (9,857 articles) over the 25-year period. The biggest mover in terms of absolute volume across the period is China, again unsurprisingly based on what we know about Chinese investments in research in recent years. The United Kingdom and Germany have been consistently in the top five across the last 25 years. Overall, the 119,272 total articles with valid country information for the corresponding author represent 103 countries; all countries with at least 500 publications are shown in this table.

Table 6 breaks down the county information by major research area,⁶ listing the top 10 countries in terms of number of publications across the entire 25-year time period. In this table, however, we have extracted all authors from the publications (rather than just the corresponding author as in table 5), which allows us to look at some of the differences between disciplines and countries at a more granular level. For a paper to be considered as emerging from a particular country, at least one of the authors had to be from that country; thus, individual articles with international co-authorship ‘count’ for two or more countries, but multiple authors on a paper from the same country are only counted once for a given paper. In other words, detail was extracted at the publication level, with the countries for this table interpretable as ‘at least one author from country X was an author on paper Y.’ We have also added a fifth area, multidisciplinary publications, which consists of all publications coded by Scopus as multidisciplinary as well as any that were coded for more than one of the other four areas.

Table 5. Top countries publishing domain-based Internet-related articles 1990-2014 (based on corresponding author country); countries with total n > 500 are shown

Country	1990-2014 Overall		1990-1994		1995-1999		2000-2004		2005-2009		2010-2014	
	Rank	Articles	Rank	Articles	Rank	Articles	Rank	Articles	Rank	Articles	Rank	Articles
United States	1	38180	1	121	1	3387	1	8495	1	11701	1	14476
China	2	10686		1	13	54	4	859	2	4155	2	5617
United Kingdom	3	9857	2	15	2	689	2	1814	3	3182	3	4157
Germany	4	5872	5	6	3	346	3	1051	4	1953	4	2516
Australia	5	4561		1	8	158	7	562	5	1473	5	2367
Canada	6	4292	4	8	5	205	5	754	6	1458	7	1867
Japan	7	3382	3	15	4	264	6	667	7	1230	12	1206
Spain	8	3262		1	11	58	13	289	11	978	6	1936
Taiwan	9	3147			12	57	11	333	8	1202	8	1555
France	10	3054		2	6	169	8	521	9	1103	10	1259
Italy	11	2838		3	7	164	9	458	10	1019	13	1194
Netherlands	12	2535		2	9	81	10	339	13	805	9	1308
South Korea	13	2246			22	35	12	324	12	812	14	1075
India	14	1994			30	15	18	177	14	579	11	1223
Sweden	15	1483		3	21	35	15	229	15	441	15	775
Switzerland	16	1280		2	10	76	14	237	16	405	18	560
Greece	17	1077		1	18	44	20	149	17	402	20	481
Finland	18	1051		1	19	43	19	160	18	384	21	463
Turkey	19	950			34	11	36	39	26	249	16	651
Hong Kong	20	932			23	35	17	195	19	334	26	368
Brazil	21	926			25	28	22	137	23	273	19	488
Belgium	22	887		1	14	53	26	122	20	298	24	413
Poland	23	839			20	36	27	122	25	255	23	426
Austria	24	836		3	17	45	21	137	21	280	25	371
Singapore	25	815		1	16	46	16	195	22	274	33	299
Malaysia	26	806			37	9	35	41	33	172	17	584
Israel	27	755		2	27	25	23	137	24	262	30	329
Denmark	28	687		2	24	31	25	125	30	195	28	334
New Zealand	29	687			26	26	28	106	27	234	32	321
Norway	30	660		1	32	13	30	73	28	233	27	340
Iran	31	615					48	10	32	173	22	432
Portugal	32	570			33	12	34	49	31	178	29	331
South Africa	33	564		2	28	24	29	73	35	143	31	322
Ireland	34	554		1	31	14	32	57	29	206	34	276
Others		6392		5		167		644		1956		3620
Total Articles		119272		200		6455		19680		38997		53940
Total Countries	103		28		67		93		102		103	

Table 6. Top countries publishing domain-based Internet-related articles in journals and conferences, 1990-2014

Country	Rank	Social Sciences								
		Counts		Citations			Co-authorship		Internationalism	
		Authors	Articles	% Cited	i10 Cited	Mean Cites	Mean n of Authors	% Multi-author	% Multi-country	% Multi-continent
United States	1	57612	26003	73.2%	30.6%	21.1	2.6	71.1%	15.4%	13.8%
China	2	17123	6589	32.4%	5.9%	7.1	2.9	85.7%	16.1%	11.5%
United Kingdom	3	13540	7029	73.4%	28.9%	16.1	2.5	68.0%	24.5%	15.2%
Spain	4	7487	2890	56.0%	16.0%	10.7	3.1	84.1%	21.3%	10.4%
Australia	5	7102	3567	65.3%	22.5%	14.2	2.5	73.3%	26.0%	25.0%
Germany	6	6851	3048	58.8%	17.5%	12.2	3.0	79.8%	30.1%	13.0%
Taiwan	7	6591	2702	65.7%	27.6%	18.0	2.7	82.9%	14.9%	11.7%
Canada	8	6057	3096	69.9%	26.8%	18.0	2.7	75.9%	32.0%	20.2%
Netherlands	9	4011	1821	73.9%	32.5%	19.7	3.1	81.4%	32.3%	12.7%
South Korea	10	3674	1626	64.0%	23.2%	16.6	2.9	88.2%	34.6%	30.0%
Overall total		184654	74225	61.5%	21.6%	16.2	2.5	72.4%	11.9%	8.0%
Medical Sciences										
United States	1	87108	23703	81.8%	38.8%	36.1	4.5	84.0%	16.9%	14.9%
United Kingdom	2	20771	6334	82.5%	40.0%	42.5	5.5	85.6%	36.2%	23.4%
Germany	3	14803	4265	77.2%	35.4%	34.9	5.6	90.0%	34.2%	17.1%
Canada	4	10171	3095	84.6%	38.9%	25.2	5.3	90.2%	36.6%	23.2%
Australia	5	8770	2816	81.0%	34.8%	20.3	4.7	89.7%	32.2%	30.3%
China	6	8546	2072	65.0%	25.8%	19.4	5.5	94.6%	31.6%	26.5%
France	7	7616	2088	76.9%	35.7%	31.5	6.5	88.9%	37.4%	21.6%
Japan	8	7575	1692	74.6%	35.6%	70.7	6.6	92.1%	23.5%	19.2%
Italy	9	7414	1775	77.3%	32.0%	24.0	7.5	93.9%	37.9%	18.6%
Netherlands	10	6930	1916	85.3%	37.7%	20.2	6.4	94.8%	38.7%	19.5%
Overall total		234376	54861	77.4%	33.1%	29.9	4.4	85.9%	14.7%	9.7%
Sciences										
United States	1	30590	10116	62.5%	23.6%	28.4	3.7	81.3%	19.4%	18.0%
China	2	14806	4293	48.9%	10.6%	9.1	3.9	94.5%	14.9%	11.7%
Japan	3	6155	1652	57.9%	17.4%	54.9	4.8	90.5%	18.3%	13.7%
Germany	4	5443	1910	63.1%	25.3%	22.9	4.8	86.1%	38.4%	19.6%
United Kingdom	5	5343	2151	72.0%	29.7%	32.5	4.6	82.3%	40.4%	24.3%
Italy	6	3631	1156	70.4%	28.8%	25.4	5.6	93.7%	39.6%	18.9%
France	7	3234	1058	68.0%	29.1%	27.5	5.9	90.1%	46.6%	24.4%
Canada	8	2829	1112	64.6%	23.6%	18.5	4.3	84.5%	35.6%	25.8%
South Korea	9	2809	813	60.6%	17.7%	18.4	4.5	96.1%	21.2%	17.8%
Spain	10	2610	792	68.2%	24.5%	29.2	5.8	95.7%	40.7%	18.4%
Overall total		105938	29567	58.7%	18.8%	21.0	3.6	84.9%	14.7%	9.6%
Humanities										
United States	1	4219	2007	67.8%	15.5%	8.7	2.4	59.7%	14.8%	13.0%
United Kingdom	2	977	587	63.4%	12.3%	6.5	2.2	52.5%	24.2%	14.5%
Spain	3	705	340	36.2%	4.7%	4.5	2.4	65.0%	16.8%	7.6%
Australia	4	526	293	61.1%	14.3%	8.3	2.3	54.6%	21.5%	21.2%
China	5	508	201	58.2%	11.4%	7.5	3.3	83.6%	39.3%	27.9%
Germany	6	483	238	63.9%	12.6%	8.3	2.9	65.5%	29.0%	11.3%
Canada	7	437	250	61.2%	14.4%	7.5	2.5	59.2%	28.0%	14.8%
Netherlands	8	386	167	43.1%	21.6%	10.9	3.0	75.4%	30.5%	11.4%
Taiwan	9	366	148	56.8%	10.1%	7.2	2.8	81.1%	17.6%	16.9%
Malaysia	10	355	115	42.6%	1.7%	2.8	3.2	93.9%	13.9%	6.1%
Overall total		12759	5653	56.9%	11.1%	7.3	2.3	57.1%	11.0%	7.0%
Multidisciplinary Overall (Combinations of Above Fields)										
United States	1	22911	7807	79.8%	31.5%	26.3	3.7	78.4%	17.0%	15.1%
United Kingdom	2	4615	2029	79.5%	30.0%	28.9	3.9	74.0%	33.2%	20.9%
China	3	3084	1165	58.9%	15.4%	11.3	4.4	92.7%	29.0%	23.9%
Germany	4	2821	979	75.1%	28.1%	23.6	4.9	84.0%	38.6%	18.0%
Canada	5	2458	954	80.2%	32.2%	20.4	4.6	83.5%	37.0%	23.2%
Australia	6	2309	908	77.9%	28.4%	13.8	3.7	79.2%	30.2%	29.3%
Netherlands	7	1782	647	62.1%	18.7%	21.0	4.7	83.6%	30.3%	13.8%
Spain	8	1752	618	84.1%	35.8%	19.2	4.9	90.5%	38.7%	17.6%
France	9	1422	377	71.6%	26.0%	162.2	6.2	91.0%	31.3%	24.1%
Japan	10	1392	542	67.0%	22.9%	20.3	5.0	74.2%	35.1%	19.7%
Overall total		59513	17672	77.3%	26.4%	20.3	3.4	83.2%	15.5%	10.2%

In the social sciences data in the top portion of table 6, we see the United States is still the number one country in terms of volume, with 57,612 authors contributing to 26,003 articles. The United States is also above average in terms of percentage of the papers cited (73.2% compared to the overall average for all countries of 61.5%), the papers cited at least ten times (30.6% compared to the overall average of 21.6%) and the average number of citations (21.1 per paper compared to the average of 16.2 for all the papers in the social science sample).

China, on the other hand, has reached a point of prominence in the social sciences on par with the United Kingdom; however it still lags considerably in terms of impact. Only 32.4% of Chinese social science papers related to the Internet have been cited a modest 7.2 times each, and only 5.9% were cited at least ten times. This lack of impact by authors based in China is consistent with data we have reported elsewhere with regard to e-research (Meyer and Schroeder, 2015).

The next sections of the table look at the medical sciences, sciences, and humanities in turn. Again, a few salient points about these data will suffice in lieu of more space. The United States and United Kingdom are the most consistently active publishing powerhouses across the four research areas reported here, with the United States consistently holding an unassailable top position, and the United Kingdom in the top five. Both countries also consistently outperform the average in terms of percentage of papers cited, and the average number of citations to those papers.

Other countries are less consistent across the disciplines. South Korea, for instance, is in the top ten for the social science and sciences, but not the medical sciences, humanities, or multidisciplinary categories. Spain is also a consistently regular player, but with a somewhat lower than average impact. For instance, 56.0% of Spain's social science papers and 36.2% of humanities papers are cited, and those only a modest 10.7 times and 4.5 times respectively. Japan, on the other hand, has an astonishingly high average number of citations in the medical sciences category (70.7 average citations for the 74.6% of publications cited); when we examined these data for anomalies, there were a few very highly cited (ca. 20k) articles about online resources, but also a considerable number of papers with several thousand citations each, so this performance is not just due to a few outliers.

We also looked at co-authorship across all the categories, and here we see a familiar pattern: the medical sciences are most likely to have larger numbers of co-authors (4.4 on average, with 85.9% of papers authored by more than one person), followed by the sciences (an average of 3.6 authors per paper, with 84.9% having multiple authors), followed by the social sciences (2.5 authors, 72.4% co-authored) and the humanities (2.3 authors, 57.1% co-authored). The multidisciplinary papers are somewhere in the middle of the pack (3.4 authors per paper, 83.2% of which are co-authored). While there are some interesting small differences here, there are no real standout differences between countries.

Table 7. Top Institutions publishing journal articles related to the Internet, in 1990-2004 and 2005-2014

	1990-2004		2005-2014	
	Institution	Articles	Institution	Articles
Technology, including computer science, engineering and mathematics	All Combined	20745	All Combined	51990
	IEEE	487	Tsinghua University	762
	Nippon Telegraph & Telephone	378	Beijing U of Posts and Telecommunications	633
	Tsinghua University	245	IEEE	367
	Alcatel-Lucent Bell Labs	236	Wuhan University	344
	Stanford University	169	National Chiao Tung University Taiwan	335
	IBM Thomas J. Watson Research Center	164	Harbin Institute of Technology	329
	UC Berkeley	161	Beijing Jiaotong Daxue	314
	National Chiao Tung University Taiwan	157	Shanghai Jiaotong University	306
	Hewlett Packard Laboratories	140	Northeastern University China	306
	International Business Machines	134	City University of Hong Kong	301
Social Sciences, including business, economics, and psychology	All Combined	11259	All Combined	44586
	Indiana University	82	Pennsylvania State University	296
	University of Texas at Austin	79	University of Texas at Austin	271
	National University of Singapore	74	University of Maryland	266
	Pennsylvania State University	72	Michigan State University	253
	University of Illinois at Urbana-Champaign	71	University of Toronto	225
	UC Berkeley	71	University of Florida	218
	Ohio State University	65	Ohio State University	208
	Shanghai Jiaotong University	63	University of Amsterdam	204
	University of Toronto	58	City University of Hong Kong	203
	Florida State University	57	Indiana University	202
Medicine, including genetics, neuroscience, nursing, pharmacology, and health professions	All Combined	12056	All Combined	43127
	European Bioinformatics Institute	96	VA Medical Center	455
	University of California, San Francisco	83	University of Washington Seattle	427
	VA Medical Center	81	University of Toronto	426
	Stanford University School of Medicine	77	Harvard Medical School	380
	University of Washington Seattle	75	University of California, San Francisco	375
	UCL	66	University Michigan Ann Arbor	328
	National Center for Biotechnology Inform.	61	The University of North Carolina at Chapel Hill	323
	University of California, San Diego	59	UCL	312
	Columbia University in the City of New York	58	Karolinska Institutet	311
	University of Toronto	54	University of Sydney	298
Science, including chemistry, physics, biology, earth sciences, and agriculture	All Combined	4657	All Combined	14152
	Uniwersytet Warszawski	38	Wuhan University	198
	IEEE	37	Chinese Academy of Sciences	107
	Tsinghua University	37	University of Cambridge	102
	University of Minnesota Twin Cities	28	Zhejiang University	97
	UC Davis	26	Tsinghua University	84
	University Michigan Ann Arbor	26	U of Electronic Science and Tech of China	79
	Shanghai Jiaotong University	24	Imperial College London	73
	University of Colorado at Boulder	23	Ohio State University	70
	Imperial College London	23	Stanford University	68
	Massachusetts Institute of Technology	23	UC Berkeley	68
Arts & Humanities	All Combined	414	All Combined	5005
	University of Illinois at Urbana-Champaign	5	University of Texas at Austin	38
	Ohio State University	4	University of Amsterdam	35
	Purdue University	4	Universiti Kebangsaan Malaysia	34
	University at Albany State U of New York	4	Ohio State University	33
			Pennsylvania State University	29
			University of Toronto	29
			University of Queensland	28
			University of Wisconsin Madison	28
			Monash University	26
			Linkopings Universitet	25

The final data shown in table 6 pertains to internationalism. The second to last column shows the percentage of papers with authors from more than one country, and the final column shows the percentage of papers with authors from more than one continent. One of the most notable things in these international data is the greatly increased likelihood of international co-authorship for most of the European countries. Across the disciplines, authors in many of the European countries co-author papers with people from other countries (with percentages in the 20s and 30s), but they are only somewhat more likely than non-European counterparts to co-author with people on a different continent. One interpretation of this is that it reflects a culture of scientific cooperation at the European level that is not as common in other parts of the world. One could speculate that some of this is down to geography (since many European countries are relatively small and easy to get to from their neighbours), but it is also possible that European Commission funding mechanisms that require multiple countries to partner on European grants is having the desired impact in terms of continental cooperation and collaboration. Exploring this possibility, however, would require further evidence to either support or refute in detail.

Finally, in table 7 we present data for the top institutions (for corresponding authors only) across two time periods (the first 15 years, 1990-2004, and the last 10 years, 2005-2014) and across the five main disciplinary areas. For this final part of the analysis, we have added the technology fields back in to our data. In these technology fields, we see a number of Asian institutions playing an especially prominent role, which is also true in the sciences, particularly in the latter period. The social sciences and humanities are more dominated by large American institutions, while medicine has a number of prominent medical centres in the US and Europe.

Discussion and Conclusion

The growth of web- and internet-related research has followed broadly similar trends over time in the major research areas. While Dutton (2013) has charted the emergence of the field of research related to the Internet's social implications, we take a broader view examining the impact of the Internet across the sciences, social sciences, and humanities. In the case of the social science research related to the Internet, we can see a slow take-off of this research area after the Internet becomes a widely used technology in the mid-1990s. The rise continues to the current day. In other areas, such as the technology disciplines, there is a slowed growth or even a tapering off in the last five years. What explains the plateauing? One potential explanation is that the Internet has become dispersed among a wider range of technologies that have superseded it in appeal to researchers in technology domains. So, for example, the Internet has arguably become displaced somewhat in importance by newer web platforms or by smartphones, both of which are of course also an outgrowth of the Internet and part and parcel of it. As mentioned at the beginning of this article, it is also possible that as the Web and Internet have become so thoroughly normal parts of the everyday practice of research (i.e. part of the research infrastructure, which tends to be invisible except when it breaks) that authors have started to mention it less prominently in the titles and abstracts of their publications.

This process of becoming so normal as not to merit mention may already be in play, although to confirm this would require further research.

Yet this trend is not confined only to technology domains. In terms of social science research, too, the early focus on computer-mediated communication has thus given way to a range of subfields. To take just one example, one area of interest is how research is supported by the Internet. But this specialism has itself undergone a number of terminological changes: from e-Research and its various cognates (cyberinfrastructure, e-Science, e-Humanities and the like) to the current focus on ‘big data’ and ‘computational social science’ (Schroeder, 2014). These various specialisms have meant that the label of ‘Internet’ research has become outgrown by various other labels including various ‘e-’ and ‘cyber’ prefixed labels. What we can see here is a simultaneous differentiation of research and a moving of the research front from one area to the next as the technology develops. Future research in this area might therefore use this more specific terminology for research using the Internet, to investigate this trend on a more granular level.

Just as we describe research using the Internet as a research infrastructure that is becoming at once more differentiated and more specialized, perhaps the same applies to the increasing invisibility of the Internet as a societal infrastructure. As society’s use of the Internet becomes taken for granted, it is no longer discussed as a separate entity but moves into the background or into various niches while other terms move into the foreground. These include (again), web, or life online, or other technologies such as smartphones in which internet technology is embedded in the background. Further, these newer technologies then become the subject of new research areas, as when smart- or mobile phone research has new journals devoted to it which eclipse, as a research area, the older journals devoted to computer mediated communication.

In terms of distribution of research by country, we see, unsurprisingly, the rise of China from laggard to being second only to the US in terms of volume of research, although it still lags in terms of impact. We also acknowledge, as indicated earlier in our methods section, that Scopus coverage of journals from certain areas of the world is less complete than its coverage of North American and Western European journals, a problem compounded by the fact that our search strategy relied only on English terms. It is likely that there is additional research being published in this area in countries not captured via these methods. While we are attuned to this recurring form of exclusion of certain authors and languages, the inability to capture such publications in our sample should not be understood as intended to diminish the importance of such work.

What will the next 25 years of Internet-related research bring? One obvious change that has taken place in the last five years is that the Internet has been overtaken by mobile phones in terms of the number of worldwide users. This trend is set to continue. An interesting point to note here is that mobile (or cell) phone owners use the Internet often without even being aware of the fact. This underscores our point about the infrastructure fading into the background. However, will mobile phones impact academic research in the same way that the Web and Internet have? The Web, and the Internet technologies that underlie it, are extensible – generative, to use the terminology suggested

by Zittrain (2008). It is this generativity that underpins the widespread growth across scholarly disciplines shown in this paper, as both historians who consult digital manuscripts and biologists who share genomic data rely on the same basic underlying generative Internet and the Web interfaces that make it usable. Mobile phone operating systems such as Android and iOS that rely more heavily on a closed app-based ecosystem to access the Internet may prove to have more limited uses across the academic landscape (although many interesting examples already exist); we will have to wait to see how the next 25 years play out to know for certain. It may develop that the Web sees other equally widespread ways of sharing, communicating, and accessing information spring up alongside it, or possibly replace it altogether.

Dutton (2013) characterized the Internet as a ‘network of networks’. Perhaps more accurately it could be described in the face of the future as an emerging penetration – and disappearance - of digital networks into everyday life. Or, as Ling (2012) has called it in relation to mobile phones, ‘taken-for-granted-ness’. Still, as the Web turns 25, its impact on the academic landscape is undeniable, even as the younger technologies it has spawned may be poised to steal the spotlight.

Funding

This work was supported in part by the Alfred P. Sloan Foundation [grant number 2012-06-17].

References

- Barjak F, Eccles K, Meyer ET, et al. (2013) The Emerging Governance of e-Infrastructure. *Journal of Computer-Mediated Communication* 18(2): 113-136.
- Dutton WH (2013) Internet Studies: The Foundation of a Transformative Field. In: Dutton WH (ed) *The Oxford Handbook of Internet Studies*. Oxford: Oxford University Press, 1-26.
- Fry J and Talja S (2007) The intellectual and social organization of academic fields and the shaping of digital resources. *Journal of Information Science* 33(2): 115-133.
- Hughes TP (1983) *Networks of Power: Electrification in Western Society, 1880-1930*. Baltimore, MD: Johns Hopkins University Press.
- Kling R, McKim G and King A (2003) A Bit More to IT: Scholarly Communication Forums as Socio-Technical Interaction Networks. *Journal of the American Society for Information Science and Technology* 54(1): 46-67.
- Kozak M and Bornmann L (2012) A New Family of Cumulative Indexes for Measuring Scientific Performance. *PLOS ONE* 7(10): e47679.
- Krippendorff K (2004) Reliability in content analysis. *Human Communication Research* 30(3): 411-433.
- Leydesdorff L, Moya-Anegón Fd and Guerrero-Bote VP (2015) Journal Maps, Interactive Overlays, and the Measurement of Interdisciplinarity on the Basis of

- Scopus Data (1996-2012). *Journal of the Association for Information Science and Technology* 66(5): 1001-1016.
- Leydesdorff L, Rafols I and Chen C (2013) Interactive Overlays of Journals and the Measurement of Interdisciplinarity on the basis of Aggregated Journal-Journal Citations. *Journal of the American Society of Information Science & Technology* 64(12): 2573-2586.
- Ling R (2012) *Taken for Grantedness: The Embedding of Mobile Communication into Society*. Cambridge, MA: MIT Press.
- Malik MM (2012) *Networks of collaboration and field emergence in 'Internet Studies'*. Thesis for the degree MSc in Social Science of the Internet, Oxford Internet Institute, University of Oxford, Oxford, UK. Available at: http://mominmalik.com/Malik_MSc_Thesis.pdf.
- Meyer ET and Schroeder R (2009) The World Wide Web of Research and Access to Knowledge. *Journal of Knowledge Management Research and Practice* 7(3): 218-233.
- Meyer ET and Schroeder R (2015) *Knowledge Machines: Digital Transformations of the Sciences and Humanities*. Cambridge, MA: MIT Press.
- Nentwich M and König R (2012) *Cyberscience 2.0: Research in the Age of Digital Social Networks*. Frankfurt: Campus Verlag.
- Peng T-Q, Zhang L, Zhong Z-J, et al. (2013) Mapping the landscape of Internet Studies: Text mining of social science journal articles 2000–2009. *new media & society* 15(5): 644-664.
- Savage M and Burrows R (2007) The Coming Crisis of Empirical Sociology. *Sociology* 41(5): 885-899.
- Savage M and Burrows R (2009) Some Further Reflections on the Coming Crisis of Empirical Sociology. *Sociology* 43(4): 762-772.
- Schroeder R (2007) *Rethinking Science, Technology, and Social Change*. Stanford: Stanford University Press.
- Schroeder R (2014) Big Data: Towards a More Scientific Social Science and Humanities? In: Graham M and Dutton WH (eds) *Society and the Internet*. Oxford: Oxford University Press.
- Waltman L, van Eck NJ and Noyons ECM (2010) A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics* 4(4): 629-635.
- Zittrain J (2008) *The future of the internet--and how to stop it*. New Haven: Yale University Press.

Notes

¹ The samples for the coding were extracted from Scopus by sorting on the top cited articles for a search term separately for articles (doctype(AR)) and conference papers (doctype(CP)), and then taking every nth article, where n=100 for samples > 1500, and n=10 for samples < 1500. Some samples are thus smaller because with some terms the

journal articles or conference papers yielded slightly fewer than 20 articles using this nth article sampling strategy. Except for the term ‘internet’ all other search terms were for the term(s) but also excluding the word internet (“AND NOT internet” in the search term) to correct for the relatively large size of that sample and minimize the number of papers that would thus be double-counted. Of the 909 articles coded, 5 were retrieved as part of more than one search term using this sampling strategy, thus constituting less than 1% of the sample.

² TITLE-ABS-KEY (internet OR {web 2.0} OR myspace OR facebook OR linkedin OR twitter OR flickr OR {social media} OR Wiki* OR cyberspace OR "cyber space" OR ebay OR youtube OR *blog* OR {e-government} OR egovernment OR {e-health} OR ehealth OR {e-commerce} OR ecommerce OR {world wide web}) AND pubyear > 1989 AND pubyear < 2015 AND (LIMIT-TO(DOCTYPE,"ar") OR LIMIT-TO(DOCTYPE,"cp")) AND (LIMIT-TO(SRCTYPE,"j") OR LIMIT-TO(SRCTYPE,"p"))

³ Even though this project was carried out halfway through 2015, publication databases continue to be updated regularly, so 2014 data could still increase in absolute volume.

⁴ Note that there are slight differences that appear when the same journal appears in multiple areas (see, for instance, PLoS One). These are due to small inaccuracies in the original Scopus data, which we have chosen not to change.

⁵ Excludes computer science, engineering, and mathematics

⁶ Excludes computer science, engineering, and mathematics