Collaboration - Impact on Productivity and Innovation

Proceedings of
14th International Conference on Webometrics, Informetrics and Scientometrics &
19th COLLNET Meeting 2018

December 5-8, 2018
University of Macau, Macau

Edited by

Bernd Markscheffel • Hildrun Kretschmer

Collaboration – Impact on Productivity and Innovation

Proceedings of
14th International Conference on
Webometrics, Informetrics and Scientometrics &
19th COLLNET Meeting 2018

December 5-8, 2018

University of Macau, Macau

Edited by

Bernd Markscheffel and

Hildrun Kretschmer

Collaboration – Impact on Productivity and Innovation

Proceedings of 14th International Conference on Webometrics, Informetrics and Scientometrics & 19th COLLNET Meeting 2018

Bernd Markscheffel and
Hildrun Kretschmer (Eds.)

Technische Universität Ilmenau
Fakultät für Wirtschaftswissenschaften und Medien
Institut für Wirtschaftsinformatik
P.O. Box 10 05 65
98684 Ilmenau
Germany

bernd.markscheffel@tu-ilmenau.de
kretschmer.h@onlinehome.de

URN: urn:nbn:de:gbv:ilm1-2019200405

DOI: 10.22032/dbt.39296

Publisher: Universitätsbibliothek Ilmenau
ilmedia
Postfach 10 05 65
98684 Ilmenau
Germany

Index

Editorial

The future of science will be determined primarily by the methods and tools that can be used to effectively analyze the huge amounts of data generated by the common use of digital media in scholarly communication, conversations, work processes and social structures. These characteristics, which are subsumed under the term data-intensive science, change the nature of scientific work in the most sustainable way. And most importantly, such expressions as the fact that scientific collaboration is increasingly embedded in a globally connected environment and the growth rate of scientific output in the top ten countries is increasing exponentially and this output is expressed above all by non-traditional, highly dynamic, interconnected assets such as data sets, software, ontologies, slides, videos, blog entries, are responsible for the manifold challenges for Scientometric investigations. COLLNET 2018 has met these challenges and provided an excellent opportunity to discuss the phenomena of collaboration in science, their impact on productivity, innovation, and benefits, and outcomes for individuals, institutions, and economies worldwide.

The COLLNET Conference Series and the active COLLNET community have been playing for almost two decades an outstanding role in scholarly communication for improving co-work and collaboration among researchers and practitioners all over the world. This global interdisciplinary research network is to comprise the prominent scientists, who work at present in the field of quantitative science studies.

COLLNET 2018 was also a great opportunity for both researchers and practitioners to share experiences, ideas, and research results on all aspects of Webometrics, Informetrics and Scientometrics and to open new avenues for possible research collaborations in the future.

This Procedia includes selected papers presented at the 14th International Conference on Webometrics, Informetrics and Scientometrics (WIS) & 19th COLLNET Meeting, which took place during 05 to 08 December 2018 in the University of Macau, Macau. The Conference was organized by the Department of Computer and Information Science, University of Macau, in collaboration with Digital Information Research Labs, India and the COLLNET Coordination Centre, Germany. The organizing committee and the editors are grateful to the delegates who had submitted and presented papers.

Bernd Markscheffel
Hildrun Kretschmer

Selected Papers

Article DOI: 10.22032/dbt.39309 Proceedings DOI: 10.22032/dbt.39296

Gender gaps in international research collaboration. A bibliometric approach

Dag W. Aksnes, Fredrik Niclas Piro and Kristoffer Rørstad

Dag.w.Aksnes@nifu.no

Fredrik.Piro@nifu.no

Kristoffer.Rorstad@nifu.no

Nordic Institute for studies in Innovation, Research and Education, Økernveien 9, 0653 Oslo, Norway

Abstract

This paper addresses gender differences in international research collaboration measured through international co-authorship. The study is based on a dataset consisting of 5,554 Norwegian researchers and their publication output during a three-year period (43,641 publications). Two different indicators are calculated. First, the share of researchers that *have* been involved in international collaboration measured by co-authorship, and second, the share of their publications with international co-authorship. We then develop an index which takes both these indicators into account: The Gender Difference Collaboration Index. The study shows that there are distinct gender differences in international research collaboration in Norway at an overall level. However, when the data is analyzed by scientific field, academic position and publication productivity of the researchers, the gender differences are less pronounced and in some cases, women have higher collaboration rates than men. The differences are largest for personnel in recruitment positions and for less productive researchers.

Introduction

Men and women have been shown, in numerous studies, to perform differently according to various indicators related to the process of scientific publishing. In particular, female researchers on average are less productive and publish fewer publications than men. This has been demonstrated in numerous studies (for example, Kyvik & Teigen, 1996; Piro, Aksnes & Rørstad, 2013; Sugimoto et al., 2013). The pattern seems to be universal across fields and nations, although the differences vary. As an example, Rørstad & Aksnes (2015) showed that adjusted for position and age, female researchers in Norway on average publish 10 to 20 percent less than men. The question of whether women are less cited than men has also been analyzed in several studies. Here, the results are less clear, and findings vary. As an example, a previous Norwegian study found only small gender differences (Aksnes et al., 2011), while a global analysis based on articles with first and last authors showed lower citation rates for female authors (Larivière et al. 2013). Lagging behind in terms of scientific production and impact represent a major problem, as these two factors are decisive for e.g. academic promotion and in the evaluation of research proposals among funding agencies (European Commission, 2015).

In this study, another dimension is analysed: gender differences in international collaboration. This issue has become ever more important to study, due to the steady increase worldwide in research collaboration in groups and networks, hence also growth in paper co-authorships (Leydesdorff & Wagner, 20008) and in interdisciplinary research (Lee & Bozeman, 2005). International research collaboration has been shown to be advantageous to researchers' productivity and scientific impact (e.g. Abramo, D'Angelo & Di Costa, 2009; Abramo, D'Angelo & Solazzi 2011; Adams 2012; Kyvik & Reymert, 2017; Larivière et al., 2013.

Nevertheless, the knowledge on gender differences in international research collaboration is inconclusive (Poole & Bornholt, 1998, Larivière et al. 2011, Vabø, 2012).

Expanding the knowledge gaps on gender gaps

In this study, we draw upon the methodological approach of three previous studies – presented below – using a Norwegian dataset with additional variables missing in these studies. In this way, we are able to provide a better understanding of gender differences in international research collaboration.

Larivière et al. (2013) used Web of Science (WoS) data from the period 2008-2012 to study differences in international co-authorship in 5,5 million papers with more than 27 million authorships. The dataset included information on the gender of the authors. Females were shown to be less frequently listed as first authors (roughly 2/3 of the papers had male first-authors), and less inclined to participate in international collaborations. In sum, these factors contributed to lower citation rates among women. Needless to say; such a large-scale study did not include individual data of the authors, such as academic position. The authors state (p.213) that "it is likely that many of the trends we observed can be explained by the under-representation of women among the elders of science. After all, seniority, authorship position, collaboration and citation are all highly interlinked variables".

A second study, is Abramo, D'Angelo & Murgia's (2013) analyses of international co-authorship among Italian professors, based on WoS publications from 2006 to 2010. In this study, academic discipline and institutional affiliation were taken into account, documenting gender differences in international collaboration across scientific fields (all hard sciences and economics). Interestingly, female researchers were shown to have a greater capacity to collaborate in all other collaboration forms being analysed, except for the international dimension. This study only included researchers in tenured academic positions.

A third relevant study is Uhly, Visser and Zippel's (2017) investigation of gender differences in international research collaborations in academia. This study, unlike the former two, included individual data on age (as well as academic discipline), but not academic position. This study applied a different methodological approach and was based on answers from a survey (ten countries analysed with 13,000 respondents in total), where the informants answered yes or no to the question "Do you collaborate with international colleagues?". This makes the results difficult to compare with the two former studies. As the authors state, the measurement of international collaborations is highly dependent on the survey respondents' interpretations of the question, as contrasted by use of publication data where such bias does not exist (Melin & Persson, 1996). At the same time, most studies on gender differences in research collaboration have been conducted based on surveys (Abramo, D'Angelo & Murgia, 2013).

The main result of Uhly and colleague's (2017) study is that women engage less in international collaboration than men, and that the degree of female international collaboration is dependent on a complex set of individual factors (such as partner employment status and children). The results lead the authors to conclude that 'glass fences' are apparent in "in the access to international research collaboration, as women are significantly less likely than men to participate in this elite activity" (p.761).

In our study, we aim at filling a knowledge gap in the understanding of gender differences in international research collaboration by comparing international paper co-authorship among men and women at Norwegian universities. Important dimensions of the study are:

- The application of a database which, in contrast to WoS, has complete coverage of all peer-reviewed scientific and scholarly publication output, including books, edited volumes and conference series. This means that we able to provide a better coverage of the Social Sciences and Humanities, in particular.

- We analyse the issue at the level of fields and disciplines. The importance of comparing by fields has been documented by e.g. Kyvik & Reymert (2017) and Abramo, D'Angelo & Murgia's (2013), with the latter study arguing (p. 819) that gender differences in international cooperation "could be due to certain factors that characterize each discipline, beginning from the percentage of women in the total research staff".
- We take the *academic position* of the researchers into account. Two previous Norwegian studies have found that older academic staff are less inclined than their younger colleagues to participate in international research networks (Kyvik & Reymert, 2017; Kyvik & Olsen, 2008).

In sum these factors enable us to test, first, whether there are gender differences in international collaboration, and, second, whether the differences vary by academic position (which is strongly correlated with age) and research field. In addition to this, we add a third main explanatory variable: scientific productivity, as we believe international collaboration may be more manifest among established researchers with high scientific productivity. Such a decomposed analysis based on these factors might add important knowledge to the understanding of gender differences, because while there may be gender differences at the overall level, or by *one* factor alone, it is not unlikely that the gender differences show covariation with other factors. Here, we try to isolate such factors in a multivariate analysis.

Data and methods

The study is based on the bibliographic Cristin database (The Norwegian Science Index) that has been developed as part of a current research information system for all public research institutions in Norway. The database has a complete coverage of all peer-reviewed scientific and scholarly publication output, including books, edited volumes and conference series (see Piro et al. 2013 for further details). In addition to bibliographic data on the publications, the database contains information on individual characteristics of the researchers (gender, age, and institution). The researchers were assigned to five broad domains (Social sciences, Humanities, Natural sciences, Technology and Medical/health sciences), based on the field distribution of their publication output.

The data material consists of 5,554 researchers from the four largest universities in Norway (University of Oslo, University of Bergen, University of Tromsø – The Arctic University of Norway and The Norwegian University of Science and Technology (NTNU)). The study is limited to professors, associated professors, postdocs and PhD students with at least one publication during the time period analyzed. Their publication output during the period 2015-2017, in total accounts for 43,641 publications (Table 1).

Table 1: Distribution of researchers and publications by gender fields and gender

	Number of researchers			*Number of publications*		
Major fields	*Men*	*Women*	*Total*	*Men*	*Women*	*Total*
Humanities	420	363	783	2,009	1,445	3,454
Social sciences	513	522	1,035	2,709	2,357	5,066
Natural sciences	902	408	1,310	10,815	3,016	13,831
Technology	662	183	845	6,545	1,572	8,117
Medical and health sciences	747	834	1,581	7,719	5,454	13,173
Total	3,244	2,310	5,554	29,797	13,844	43,641

Female researchers constitute 41.6 per cent of the study population, while they only account for 31.7 per cent of the publications. The female shares of the researchers vary greatly by field. It is highest in Medical and health sciences (52.8 per cent), Social sciences (50.4 per cent) and

Humanities (46.4 per cent); considerably lower in Natural sciences (31.1 per cent) and Technology (21.7 per cent). The female shares of the publication output, however, does not coincide be representation of researchers. Women publish less than men in all fields, while it is in Technology where female researchers publish most equally to men: 21.7 per cent women account for 19.4 per cent of the output, i.e. a publication output just 2.3 percentage points lower than expected based on representation of researchers. This female under-representation of the publications is moderate in Social sciences (3.9 percentage points) and Humanities (4.5); and high in Natural sciences (9.3) and Medical and health sciences (11.3).

The analyses are carried out by fields of research, academic positions, and their scientific production. The latter is a factor that we find essential when studying gender differences in international collaboration. Without a stratification of the study population to different levels of scientific production, important nuances are lost. We have split the sample in three groups based on publication volume. The first group, is the researchers with (on average) less than a publication a year (31.5 per cent of the sample), the second group is the researchers with 1-3 publications a year on average (46.3 per cent of the sample), and the third group is the bulk of very productive researchers with on average of more than 3 publications each year (22.1 per cent of the sample).

The unit for the analyses is the individual researchers. For each person we calculate whether they have published at least one publication involving international co-authorship (i.e. having co-authors affiliated with institutions in other countries) during the period. In other words, all individuals count equally as one unit in the analysis regardless of how many publications they have published. By this, we avoid that the analysis is biased towards highly productive researchers. However, such a dichotomous measure is deprived of essential information. Whilst it provides us the shares of men and women that are involved in international collaboration, we do not know anything about the degrees of internationalization among the individuals. For example, in two groups (100 men and 100 women), we may find that 54 per cent of the men have international co-authors, while 57 per cent of the women have international co-authors. Women here appear to be more international oriented than men.

If, on the other hand, the measure is the percentage of international co-authored publications, we may find that in the female group, on average 35 per cent of the publications have international co-authors, while 39 per cent of the men's publications have international co-authors. We now have two results that pull in different directions. We believe both measures are important to consider. The first is a measure of how many *individuals* that have international co-authors, while the second is a measure of how many *publications* that have international co-authors. The two factors provide complementary information on gender differences in international collaboration. What is needed is measure that takes both factors simultaneously into account. We therefore suggest a simple measure combining both *presence* and *scope* of international collaboration, which we call the Gender Difference Collaboration Index (GDCI). The GDCI is calculated as:

$$GDCI = \left(\frac{m\ int}{m} * \frac{\sum_{n=1}^{m} \left(\frac{pub\ int_n}{pub\ tot_n} \right)}{m} \right) - \left(\frac{w\ int}{w} * \frac{\sum_{n=1}^{w} \left(\frac{pub\ int_n}{pub\ tot_n} \right)}{w} \right)$$

Where *m/w* is the total number of men/women in the study sample, and *m int/w int* is the number of men/women with international collaboration. Pub tot is the total number of publications and pubs int is the number of publications with internationally collaboration. The GDCI varies between -1 (complete gender difference in favor of women) to 1 (complete gender difference in favor of men).

We first present gender differences in both sets of analysis (gender differences based on dichotomous distribution of yes or no with regard to international collaboration, and gender differences based on shares of publications with international collaboration), before we present GDCIs for each indicator in multivariate analyses.

Results

Overall, 56 per cent of the female researchers were involved in international collaboration measured by co-authorship. The corresponding figure for men was 66 percent. Thus, our study shows that overall male researchers more often are involved in international collaboration than their female colleagues. However, as expected there are large differences across domains (Figure 1). International collaboration is much more frequent in the Natural sciences, Medical and health sciences and Technology compared with Humanities and Social sciences. This holds for both genders. In the Humanities less than one third of the researchers have publications involving international collaboration. There are gender differences in all domains. The gap is largest in the Social sciences where the proportion for men is 44 per cent and 36 per cent for women. The gap is smallest in Humanities (the difference is three percentage point).

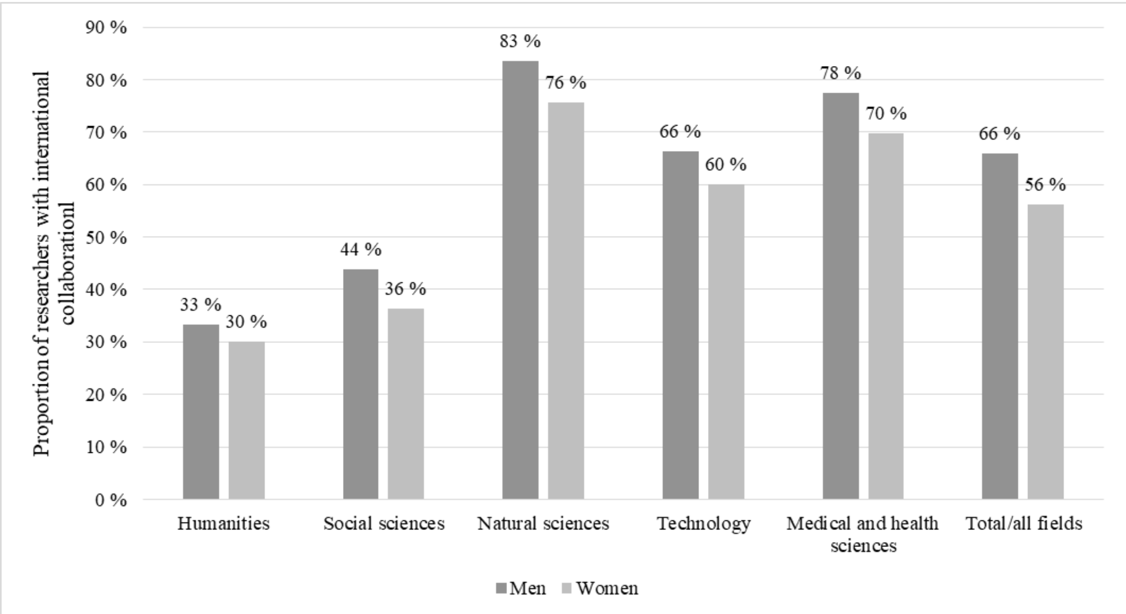


Figure 1. Proportion of researchers involved in international collaboration by fields and gender

Figure 2 shows the corresponding figures using the proportions of publications involving international collaboration as indicator. Gender differences are observed across the two different measures but now the gender differences are reduced. The most evident reduction in gender gaps is observed in Natural sciences, where a seven-percentage point higher share of men was involved in international collaboration (Figure 1), while the share of the publications that involve international collaboration is just two percentage points higher for men (Figure 2). Similar results are observed when we study academic position instead of scientific domain (not shown in figures).

In Tables 2-4 we present the results split by gender, publication volume, scientific domain and academic position simultaneously. In Tables 2-4 we only report numbers for groups with more than 20 researchers. In Table 2 we report the percentage of men/women that *have* collaborated internationally (yes or no,), while we in Table 3 report the shares of publications with international co-authors.

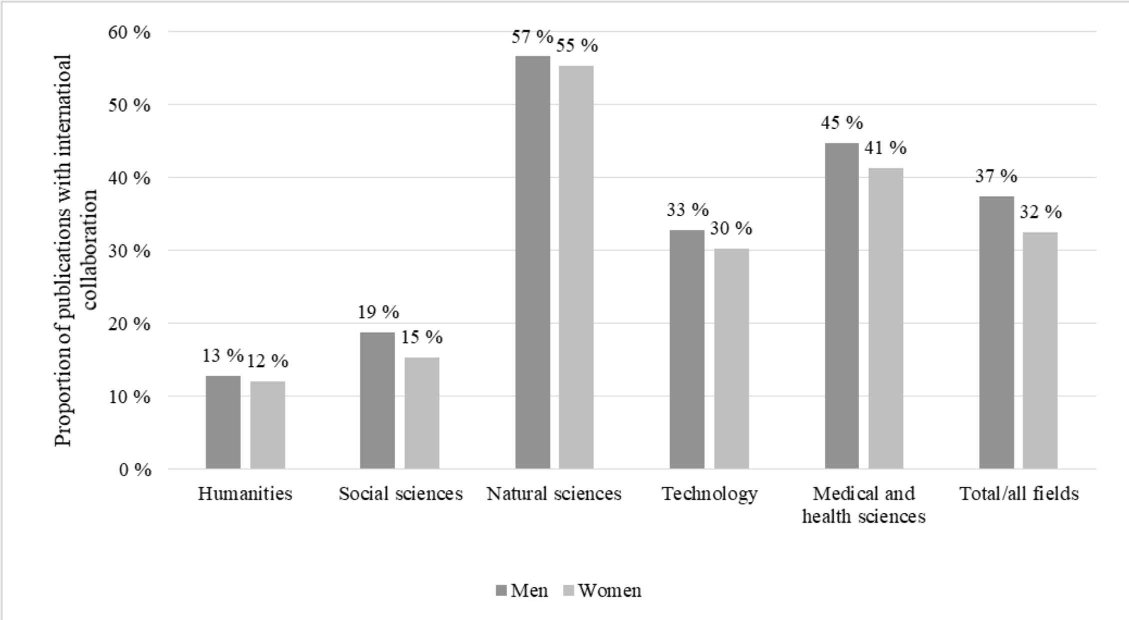


Figure 2. Average proportion of international co-authorship per individual by fields and gender

Table 2: Proportion of researchers involved in international collaboration by fields, academic position, publication productivity and gender

Fields	*1-2 publications*		*3-9 publications*		*10+ publications*		*Total*
Positions	*Men*	*Women*	*Men*	*Women*	*Men*	*Women*	
Humanities	**12 %**	**11 %**	**38 %**	**39 %**	**77 %**	**74 %**	**32 %**
Professors	13 %	17 %	38 %	35 %	73 %	74 %	38 %
Associate professors	9 %	16 %	46 %	45 %			32 %
Postdocs							27 %
PhD students	8 %	6 %					15 %
Social sciences	**20 %**	**13 %**	**50 %**	**45 %**	**78 %**	**85 %**	**40 %**
Professors	25 %	21 %	55 %	49 %	85 %	87 %	53 %
Associate professors	19 %	10 %	46 %	43 %		87 %	35 %
Postdocs				54 %			45 %
PhD students	14 %	13 %	40 %	29 %			19 %
Natural sciences	**60 %**	**59 %**	**87 %**	**83 %**	**100 %**	**98 %**	**81 %**
Professors	75 %		91 %	90 %	100 %	97 %	93 %
Associate professors	55 %		87 %	84 %	98 %		81 %
Postdocs	65 %	70 %	88 %	91 %			83 %
PhD students	55 %	56 %	79 %	70 %			65 %
Technology	**38 %**	**27 %**	**60 %**	**62 %**	**95 %**	**90 %**	**65 %**
Professors			73 %		97 %	91 %	85 %
Associate professors	21 %		60 %		93 %		64 %
Postdocs			69 %				71 %
PhD students	43 %	29 %	51 %	49 %			47 %
Medical/health sci	**43 %**	**46 %**	**79 %**	**76 %**	**98 %**	**98 %**	**73 %**
Professors	30 %		83 %	80 %	97 %	96 %	88 %
Associate professors	30 %	45 %	78 %	75 %	100 %	100 %	74 %
Postdocs		64 %	78 %	82 %			79 %
PhD students	45 %	41 %	70 %	71 %			54 %
Total	**37 %**	**33 %**	**66 %**	**63 %**	**95 %**	**93 %**	**62 %**

Table 3: Average proportion of international co-authorship per individual by fields, academic position, publication production and gender

Fields	*1-2 publications*		*3-9 publications*		*10+ publications*		*Total*
Positions	*Men*	*Women*	*Men*	*Women*	*Men*	*Women*	
Humanities	**8 %**	**8 %**	**13 %**	**13 %**	**24 %**	**26 %**	**12 %**
Professors	9 %	13 %	12 %	13 %	23 %	22 %	14 %
Associate professors	6 %	10 %	17 %	15 %			13 %
Postdocs							11 %
PhD students	8 %	5 %					9 %
Social sciences	**15 %**	**10 %**	**20 %**	**17 %**	**25 %**	**27 %**	**17 %**
Professors	18 %	16 %	20 %	18 %	29 %	32 %	21 %
Associate professors	16 %	8 %	21 %	16 %		25 %	16 %
Postdocs				18 %			16 %
PhD students	9 %	10 %	16 %	14 %			11 %
Natural sciences	**51 %**	**49 %**	**53 %**	**57 %**	**66 %**	**67 %**	**56 %**
Professors	64 %		53 %	55 %	67 %	69 %	60 %
Associate professors	43 %		53 %	54 %	57 %		51 %
Postdocs	58 %	62 %	56 %	65 %			61 %
PhD students	48 %	48 %	51 %	54 %			51 %
Technology	**34 %**	**22 %**	**27 %**	**29 %**	**40 %**	**40 %**	**32 %**
Professors			33 %		41 %	43 %	38 %
Associate professors	18 %		26 %		32 %		27 %
Postdocs			38 %				41 %
PhD students	39 %	25 %	21 %	24 %			27 %
Medical/health sci	**37 %**	**38 %**	**42 %**	**40 %**	**53 %**	**49 %**	**43 %**
Professors	20 %		43 %	43 %	52 %	49 %	47 %
Associate professors	26 %	37 %	35 %	34 %	53 %	50 %	38 %
Postdocs		53 %	48 %	45 %			50 %
PhD students	40 %	35 %	44 %	42 %			39 %
Total	**31 %**	**27 %**	**34 %**	**32 %**	**50 %**	**45 %**	**35 %**

In both tables, there is a clear association between the publication volume and international collaboration. Therefore, there is also a clear tendency that the degree of internationalization concurs with academic position, where foremost professors have the highest shares. Comparing academic fields, researchers in Humanities (32 per cent) and Social Sciences (40 per cent) have the lowest shares of international co-publications, and Technology (65 per cent), Medical and health sciences (73 per cent) and Natural sciences (81 per cent) being far more international (Table 2). The same rank order is also found when comparing shares of publications that involved international co-authorship (Table 3). Here, the lowest share is found in Humanities (12 per cent) and the highest in Natural sciences (56 per cent).

In most fields, and in most academic positions, shares of international collaboration are highest among men. There are (at the overall level, i.e. by fields not taking academic position into account) only three categories where women rank higher than men on *both* measures (Tables 2 and 3): Researchers with 1-2 publications in Medical and health sciences, researchers with 3-9 publications in Technology, and researchers with 10 or more publications in Social sciences. There are also a few categories where the two indicators show deviating patterns and one gender has the highest proportion on one indicator and lowest on the other. In Table 4 we therefore present GDCI values in all categories (with more than 20 researchers), so that we can find one unified expression of the gender inequality. In addition to GDCI values, we report size- adjusted GDCIs (summed to 100, based only on cells with n≥20, where GDCIs are adjusted for sample size, i.e. the GDCIs are multiplied by the number of respondents). This enables us to identify in which categories the origins of the gender equality can be found, and we may decompose the relative contribution of each category to the total inequality. For example, a very high gender inequality based on a very small sample, adds very little explanation to the total inequality, whereas a low/modest inequality in a very large sample, may add much explanation for the total gender inequality.

Table 4: Gender Difference Collaboration Index (GDCI) across fields, academic position and publication production

Fields	*1-2 publications*			*3-9 publications*			*10+ publications*		
Positions	*GDCI*	*n (M/W)*	*Size adj.GDCI*	*GDCI*	*n (M/W)*	*Size adj.GDCI*	*GDCU*	*n (M/W)*	*Size adj.GDCI*
Humanities	0.000	(147/159)		0.000	(226/173)		-0.006	(47/31)	-0.45 %
Professors	-0.008	(53/24)	-0.58 %	0.003	(135/69)	+0.57 %	0.009	(40/23)	+0.53 %
Associate professors	-0.010	(66/56)	-1.14 %	0.012	(63/69)	+1.48 %			
Postdocs									
PhD students	0.004	(24/65)	+0.33 %						
Social sciences	0.015	(169/216)	+5.61 %	0.024	(276/247)	+12.19 %	-0.034	(68/59)	-4.19 %
Professors	0.009	(57/28)	+0.72 %	0.023	(172/95)	+5.75 %	-0.035	(46/30)	-2.49 %
Associate professors	0.022	(59/89)	+3.05 %	0.030	(69/98)	+4.69 %			
Postdocs									
PhD students	-0.001	(44/91)	-0.13 %	0.025	(20/28)	+1.12 %			
Natural sciences	0.016	(242/164)	+6.31 %	-0.009	(391/180)	-4.99 %	0.000	(269/64)	
Professors				-0.009	(164/42)	-1.74 %	-0.003	(192/36)	-0.64 %
Associate professors				0.005	(61/38)	+0.46 %			
Postdocs	-0.056	(46/30)	-3.89 %	-0.101	(84/46)	-12.29 %			
PhD students	-0.008	(120/108)	-1.71 %	0.027	(82/54)	+3.44 %			
Technology	0.068		+13.14 %	-0.020	(298/85)	-7.44 %	0.021	(214/49)	+5.36 %
Professors							0.008	(144/22)	+1.24 %
Associate professors									
Postdocs									
PhD students	0.095	(95/34)	+11.47 %	-0.014	(138/43)	-2.37 %			
Medical/health sci	-0.019	(164/292)	-8.41 %	0.022	(317/382)	+14.93 %	0.041	(266/160)	+16.96 %
Professors				0.009	(133/82)	+1.81 %	0.039	(208/103)	+11.35 %
Associate professors	-0.088	(27/49)	-6.26 %	0.015	(74/114)	+2.64 %	0.032	(38/43)	+2.43 %
Postdocs				0.011	(46/76)	+1.26 %			
PhD students	0.036	(98/187)	+9.60 %	0.013	(64/110)	+2.12 %			
Total	0.025	(872/880)	40.62 %	0.017	(1508/1067)	+23.36 %	0.055	(864/363)	+36.02 %

The first observation in Table 4, is that it is in the group of less productive researchers (1-2 publications) that we find the highest source of gender inequality. In the two publication output groups that we consider the most important ones, the gender inequality is much higher among the most productive researchers (36 per cent of total size adjusted GDCIs) compared to the middle group (3-9 publications, 23 per cent). The common characteristics for most categories where women have higher GDCIs than men, is that the relative contribution of the females does not add much to the total numbers, as the GDCIs in favour of women are primarily based on very low samples (often in combination with low GDCIs). If we discretionary choose 5 per cent size adjusted GDCI as the threshold for important gender inequality, there are only two categories (female postdocs in Natural sciences and associate professors in Medical and health sciences with 1-2 publications) where women have substantial higher size adjusted international collaboration index than men. Among men, on the other hand, there are numerous such examples. The strongest contributions to men's higher degree of international collaboration is found for PhD students in Technology and Medical and health sciences (1-2 publications) and professors in Medical and health sciences (10 or more publications).

At a more general level, we would like to emphasise three main findings of Table 4: *First,* we find the strongest gender differences in internationalization in Medical and health sciences. Here, among the least productive researchers, women have more international collaboration, but the pattern is opposite for researchers with 3-9 publications, and the male dominance becomes even more pronounced for the most productive researchers, especially professors (17 per cent of total GDCIs).

Second, in the Natural sciences, the gender inequalities are almost completely opposite. Here, women are more international collaborative in the mid-group (3-9 publications), and at the level of the most productive researchers there are no gender differences at all. In both Humanities and the Social Sciences, women are more international than men in the most productive group, but the differences are so small, that they hardly contribute to the overall gender inequality.

Third, much of the gender imbalance stems from researchers with just 1-2 publications, and especially from researchers in recruitment positions. Male PhD students contribute to 11.5 per cent of total size adjusted GDCIs in Technology, in Medical and health sciences the corresponding figure is 9.6 per cent.

Discussions and conclusions

Our study shows that there are distinct gender differences in international research collaboration in Norway. However, women and men are not equally distributed. Women account for higher proportions of personnel with lower academic ranks and with lower publication productivity. In these groups, the propensities to collaborate internationally are lower for both genders. As a consequence, the gender differences are smaller when academic position and productivity are taken into account. Still, in the majority of categories where fields, academic positions and productivity are analysed separately, shares of international collaboration are slightly higher for men than for women.

If one wants to address solutions to reduce the gender gap in international collaboration, it is important to take both measures of international collaboration into account (how many have been involved in international collaboration, and the frequency of such collaborations), and analyse different layers that may contribute to lower international collaboration for women. Our results suggest that gender differences are particularly pronounced at an early phase of the researchers' careers, and less pronounced at later stages. At the level of fields, the gender gap is largest within Medicine and health sciences.

References

Abramo, G., C.A. D'Angelo & F. Di Costa (2009): "Research collaboration and productivity: Is there correlation?", *Higher Education* 57: 155-171.

Abramo, G., C.A. D'Angelo & M. Solazzi (2011): "The relationship between scientists' research performance and the degree of internationalization of their research", *Scientometrics* 86: 629-643.

Abramo, Giovanni, Ciriaco Andrea D'Angelo & Gianluca Murgia (2013): "Gender differences in research collaboration", *Journal of Informetrics* 7(4):811-822.

Adams, Jonathan (2012): "Collaborations: The rise of research networks", *Nature* 490: 335-336.

Aksnes, Dag W., Kristoffer Rørstad, Fredrik Niclas Piro & Gunnar Sivertsen (2011): "Are female researchers less cited? A large scale study of Norwegian researchers", *Journal of the American Society for Information Science and Technology* 62(4): 628-636.

European Commission (2015): *She Figures 2015.* Brussels: European Commission: Directorate-General for Research and Innovation.

Kyvik, Svein and Mari Teigen (1996): "Child Care, Research Collaboration, and Gender Differences in Scientific Productivity." *Science, Technology, & Human Values* 21(1): 54–71.

Kyvik, Svein and Terje Bruen Olsen (2008): "Does the aging of tenured academic staff affect the research performance of universities", *Scientometrics* 76: 439-455.

Kyvik, Svein and Ingvild Reymert (2017): "Research collaboration in groups and networks: differences across academic fields", *Scientometrics* 113(2): 951–967.

Larivière, V., E. Vignola-Gagné, C. Villeneuve, P. Gelinas, Y. Gingras (2011): “Sex differences in research funding, productivity and impact: An analysis of Quebec university professors”, *Scientometrics* 87(3): 483-498.

Larivière, Vincent, Cassidy R. Sugimoto, Ni Chaoquin, Yves Gingras & Blaine Cronin (2013): “Global gender disparities in science”. *Nature 504*: 211-213.

Lee, S. & B. Bozeman (2005): “The impact of research collaboration on scientific productivity”, *Social Studies of Science* 35: 673-702.

Leydesdorff, L. & C.S. Wagner (2008): “International collaboration in science and the formation of a core group”, *Journal of Informetrics* 2: 317-323.

Melin, G. & O. Persson (1996): “Studying research collaboration using co-authorships”, *Scientometrics* 36(3): 363-377.

Piro, Fredrik Niclas, Dag W. Aksnes og Kristoffer Rørstad (2013): “A Macro Analysis of Productivity Differences Across Fields: Challenges in the Measurement of Scientific Publishing” Journal of the American Society for Information Science and Technology (JASIST) 64(2):307-20.

Rørstad, K., & Aksnes, D. W. (2015): “Publication rate expressed by age, gender and academic position - A large-scale analysis of Norwegian academic staff”, *Journal of Informetrics,* 9(2): 317-333.

Poole, Millicent and Laurel Bornholt (1998): “Career Development of Academics: Cross-cultural and Lifespan Factors”, *International Journal of Behavioral Development* 22 (1): 103–126.

Uhly KM, Visser LM & Zippel KS (2017): “Gendered patterns in international research collaborations in academia”. *Studies in Higher Education* 42(4): 760-782.

Vabø, Agnete (2012): “Gender and International Research Cooperation”, *International Higher Education* 69: 19–20.

Article DOI: 10.22032/dbt.39326 Proceedings DOI: 10.22032/dbt.39296

Exploring the relation between press releases and media coverage of Japanese university research outputs

Masaki Nishizawa and Yuan Sun

nisizawa@nii.ac.jp, yuan@nii.ac.jp
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan

Abstract

In recent years, the number of press releases from universities has generally increased over time. How academic research is reported in the media is a big concern for management at universities and for research on the diffusion of scientific knowledge. In this study, we investigate the current situation and analyze the relation between press releases and their coverage in two major national newspapers in terms of the source article's altmetric attention score, citation, subject field, and collaboration network from 2011 to 2014.

Introduction

To chieve accountability and as one of the strategies to secure research funds and to increase university enrolment, the number of universities making an effort to publish press releases to announce research findings has grown rapidly in recent years, and the number of press releases related to top-tier universities in Japan has generally increased over time (Nishizawa and Sun, 2012). How academic research is reported in the media is a big concern for management of universities and for research on the diffusion of scientific knowledge. In our previous studies, we investigated the relation between university press releases and two major Japanese national newspapers from 2007 to 2012 (Nishizawa and Sun, 2014), and the relation between the Altmetric Attention Score (AAS) (Altmetric, 2018) of the source article in press releases and coverage in newspapers in 2012. It has been shown that research published in journals with high Eigenfactor values tend to be announced in university press releases (Nishizawa and Sun, 2016), and the AAS of publications tends to have a positive correlation with instances of being featured in newspapers (Nishizawa and Sun, 2017).

In this study, after updating the current situation of university press releases from 2005 to 2015, we identified each source article in press releases from 2011 to 2014 through their Digital Object Identifier (DOI), and investigated the relation between press releases and their coverage in two national newspapers, Mainichi Shimbun and Yomiuri Newspaper, in terms of their AAS score, citation counts, subject fields, and collaboration network.

Data and Methods

Press releases

Table 1 shows the number of press releases in the Nikkei press release (Nikkei press release, 2018) database from 2005 to 2015 that contain the query word "大学 (university)." As reported in Nishizawa and Sun (2014, 2015), the number of press releases concerning universities has increased suddenly in recent years.

Table 1: Number of the press releases found using the query word "university"

Year	*total*	*2005*	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
Press release (UNIV)	14582	1154	1097	1216	1209	1244	1299	1429	1472	1417	1375	1670

Newspaper

We drew data from the corpora of two major national newspapers (Yomiuri Shimbun and The Mainichi) in Japan (Nichigai Associates, Inc., 2018). The organization's name was extracted from the article's text body using the name identification database described previously (Nishizawa and Sun, 2016).

Correspondence between the press releases and newspapers

Our method and extraction results for articles mentioned in both newspapers and press releases were described in Nishizawa and Sun (2014). The number of articles that corresponded to the press releases from 2011 to 2014 and the newspaper article are shown in Table 2.

Altmetric Attention Score and Cited numbers in Web of Science

In this research, the DOIs of journal articles announced in press releases were picked out from the body of the press release text. When DOI information was not found in the press release text, the journal article's DOI was identified based on the article title, author information, and journal information. We found the Altmetric application-programming interface key from Altmetric.com and obtained altmetric data in JavaScript Object Notation format through an https protocol. Altmetric.com offers many altmetric indexes, and the AAS is used in this study. Similarly, the number of citations for articles was extracted from the Web of Science (WoS) (Clarivate Analytics) using DOIs.

Table 2 shows the number of DOIs identified in the press releases, the number in which AAS was obtained and the number in which the cited number was obtained, together with the number of correspondences with the newspaper article.

Table 2: Number of identified DOIs, corresponding newspaper articles, AASs, and cited numbers

Year	*DOIs in PR*	*Corr Mainichi*	*Corr Yomiuri*	*Corr Both*	*AAS hit*	*WOS hit*
2011	381	35	43	19	249	367
2012	464	57	79	34	389	452
2013	528	64	68	30	442	506
2014	465	32	33	12	388	436

Results and Discussion

Impact of corresponding newspaper articles

Figure 1 shows the frequency distribution of Log_{10} (AAS) (1a): the specific DOI, (1b): the corresponding newspaper article, (1c): without corresponding newspaper article is also shown. Similarly, Figure 2 shows the frequency distribution of Log_{10} (cited number) (2a): the specific DOI, (2b): the corresponding newspaper article, (2c): without corresponding newspaper article is also shown. Although it is a preliminary result, both (b) corresponding newspaper articles show high AAS and cited numbers, respectively. Table3 shows the result of the t-test against with/without corresponding newspaper articles for AAS ((1b: with News), (1c: without News)), and cited numbers ((2b: with News), (2c: without News)) for combined data from 2011 to 2014, respectively. As shown by the P value of the table (t-test: 95% confidence interval, two sided), the difference in the mean value is significant, and the mean value is higher when there are newspaper articles, especially in AAS.

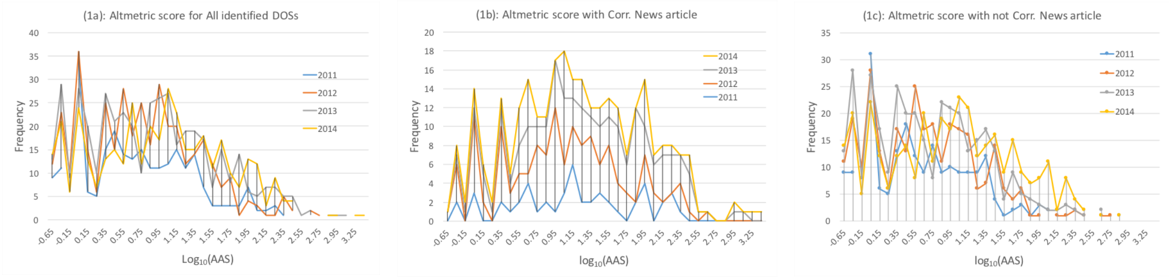


Figure 1. Distribution of $\log_{10}$ (Altmetric Attention Score)

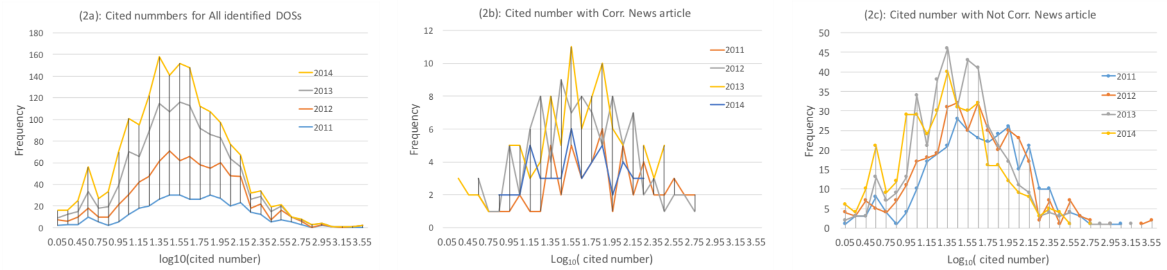


Figure 2. Distribution of $\log_{10}$ (Cited number)

Table 3: Result of t-test against with/without corresponding newspaper articles

	t-test: *log10(AAS)*		t-test: *log10(Cited number)*	
	(1b): with News	(1c): without News	(2b): with News	(2c): without News
mean	1.2047	0.7284	1.6275	1.4859
std. dev.	0.7513	0.6977	0.5387	0.4994
n	292	1183	309	1447
degree of freedom	1473		1754	
t	10.2856		4.4608	
P(T<=t) One sided	2.60E-24		4.34E-06	

Co-authorship status of the source article of Press Release

In the previous report, we have reported about the mean value and standard deviation of AAS for the co-authorship status of the article in the press release (Nishizawa and Sun, 2017). In that report, we examined the relation between the co-authorship status and the value of AAS only with data from 2012, but in this report, we extended the data from 2011 to 2014 and also examined the relation between AAS and cited numbers in the WoS.

The co-authorship status of the article announced in the press release is classified as shown in Table 4, and the mean values of AAS and the cited number for each category and their standard deviations were obtained. In the previous report, we used the correspondence author and the country of the author's institution to classify the categories, but this time we used the RA (Reprint author) term of WoS instead of the correspondence author. As RA is one person, it is not classified as a category corresponding to Japan-foreign entity corresponding to the last code 2. The co-authorship status of the article announced in the press release is classified as shown in Table 4, and the mean values of AAS and the cited number in WoS for each category, as well as the standard deviations were obtained.

Table 4: Affiliation of co-authorship of the journal articles

IntCollab code	*Int. Nat. Collab Config.*	*Remark*
0	Japanese Organization only	All authors have at least one Japanese affiliation
1	Int. Nat. Collab.: Japan-based entity	Reprint author: Japanese affiliation only
2	Int. Nat. Collab.: Japan-foreign entity	Reprint author: both Japanese and foreign affiliation *(No data)
3	Int. Nat. Collab.: Foreign-based entity	Reprint author: foreign affiliation only
4	No Japanese Organization	Reprint author: foreign affiliation only

The results are shown in Table 5. As for AAS, there is no significance in the result of the *t-test* (95% confidence interval, two sided) for the difference between the mean value of 3 and 4 of the IntCollab code. However, in other results, the mean value of AAS and the cited number showed a significant difference when co-authored internationally. However, the corresponding ratio with newspapers is slightly higher in code 4 (no Japanese organization), but no big difference is observed. As there is a difference in not only AAS but also cited number, the degree of attention of international co-authors is higher than that of co-authorship only in Japan, with respect to the source articles of press releases, and furthermore, foreign-authored papers are more when it was found that attention was high.

Table 5: Differences of AAS and cited number in WoS for the co-authorship status of the article in press release

IntCollab code	papers	with News	log10(score)			log10(cite)		
			M	SD	P-value	M	SD	P-value
0	1135	17.5%	0.676	0.688		1.423	0.489	
t-test for 0-1					5.920E-08			1.529E-11
1	442	17.0%	0.906	0.681		1.607	0.463	
t-test for 1-3					1.272E-12			6.828E-05
3	170	19.4%	1.388	0.754		1.785	0.553	
t-test for 3-4					0.2170			3.047E-03
4	24	25.0%	1.600	0.765		2.156	0.478	

Newspaper covered rate for Journals and Journal category

Table 6 shows the number of reports on journal titles in the source paper of the press release and the number of corresponding newspaper reports and their ratios for the top reported journals. In addition to Multidisciplinary Sciences journals, such as Nature and Science, the coverage rate of astronomy and geoscience magazines, such as Astrophys. J. and Nat. Geosci., and other biological systems such as Nature Genet., Neuron, Cell, and Curr. Biol is high. Meanwhile, the coverage rate is low for leading physics and chemistry journals, such as Phys. Rev. Lett. and Angew. Chem. - Int.

Table 7 shows the reported number of source papers in press releases summarized in the category of WoS, and the corresponding number of newspaper reports and the rate that they were covered. However, because the WoS has multiple field categories assigned to one journal, it is a duplicate count. Table 5 shows that the coverage rate in newspapers varies greatly depending on the field category. In addition to the interests of readers, this may be related to external factors such as Nobel laureate awards and earthquake disasters. Analysis of the strength of the correlation between these is a future task.

Conclusion

We identified the DOIs of the source articles in press releases announcing research findings sent out by Japanese universities and investigated AAS and cited number of the articles. We investigated the differences between co-authorship status, corresponding rate to newspaper, AAS value, and the cited number on the source article of the press release.

As for the corresponding rate, articles by the author of overseas organizations (No Japanese Organization) were somewhat higher, but no big difference was observed. However, the values of AAS and citations tended to be significantly higher for “Japanese Organization only,” “Int. Nat. Collab.: Japan-based entity,” “Int. Nat. Collab.: Foreign-based entity,” and “No Japanese Organization.” In the source article of the press release, there was a tendency for the author of an overseas institution to have higher attention, but this seems to not be directly related to the corresponding rate in the newspaper.

Those with high AAS and citations tended to be covered by newspapers, and those especially published in some specific journals tended to be covered by university press releases. As for future work, we will take into consideration external factors to analyze what kind of cause the press release is linked to in the newspaper publication.

Table 6: Newspaper coverage rates for Journals

rank	Journal	Press Release	with news	rate
1	Nat. Commun.	171	18	10.5%
2	Proc. Natl. Acad. Sci. U. S. A.	144	29	20.1%
3	PLoS One	104	26	25.0%
4	Nature	91	38	41.8%
5	Science	75	20	26.7%
6	Phys. Rev. Lett.	75	3	4.0%
7	Sci Rep	64	11	17.2%
8	J. Neurosci.	33	3	9.1%
9	J. Am. Chem. Soc.	32	2	6.3%
10	J. Biol. Chem.	31	2	6.5%
11	Angew. Chem.-Int. Edit.	30	1	3.3%
12	Nat. Mater.	30	2	6.7%
13	Appl. Phys. Lett.	29	1	3.4%
14	Nat. Phys.	24	2	8.3%
15	Nature Genet.	24	7	29.2%
16	Neuron	20	5	25.0%
17	Cell Reports	17		
18	Appl. Phys. Express	17	1	5.9%
19	Cell	16	6	37.5%
20	Nat. Photonics	15		
21	Curr. Biol.	15	5	33.3%
22	Astrophys. J.	14	4	28.6%
23	Nat. Neurosci.	13	1	7.7%
24	Phys. Rev. B	13		
25	Mol. Cell	13	1	7.7%
26	Nat. Geosci.	12	8	66.7%
27	Nat. Nanotechnol.	12		
28	Immunity	12	4	33.3%
29	Nat. Chem.	12	2	16.7%
30	J. Clin. Invest.	11	4	36.4%
31	Adv. Mater.	11		
32	Astrophys. J. Lett.	10	1	10.0%
33	Nat. Med.	10	3	30.0%

Table 7: Coverage rates for Journal categories

rank	Web of Science Category	journals in PR	with News	rate
1	Multidisciplinary Sciences	651	139	21.4%
2	Cell Biology	188	34	18.1%
3	Biochemistry & Molecular Biology	169	26	15.4%
4	Physics, Applied	132	5	3.8%
5	Chemistry, Multidisciplinary	122	7	5.7%
6	Neurosciences	110	16	14.5%
7	Physics, Multidisciplinary	106	5	4.7%
8	Materials Science, Multidisciplinary	101	4	4.0%
9	Chemistry, Physical	79	2	2.5%
10	Genetics & Heredity	72	16	22.2%
11	Physics, Condensed Matter	68	2	2.9%
12	Nanoscience & Nanotechnology	54	1	1.9%
13	Medicine, Research & Experimental	41	13	31.7%
14	Astronomy & Astrophysics	33	10	30.3%
15	Immunology	32	8	25.0%
16	Optics	26		
17	Geosciences, Multidisciplinary	26	12	46.2%
18	Plant Sciences	25	1	4.0%
19	Developmental Biology	24	2	8.3%
20	Ecology	24	9	37.5%
21	Evolutionary Biology	23	8	34.8%
22	Biology	22	6	27.3%
23	Cell & Tissue Engineering	19	10	52.6%
24	Microbiology	19	1	5.3%
25	Oncology	15	6	40.0%
26	Biotechnology & Applied Microbiology	15	2	13.3%
27	Biophysics	14		
28	Biochemical Research Methods	14		
29	Environmental Sciences	13	4	30.8%
30	Virology	13	1	7.7%
31	Endocrinology & Metabolism	13	6	46.2%
32	Medicine, General & Internal	13	1	7.7%
33	Parasitology	12	1	8.3%
34	Hematology	12	5	41.7%
35	Chemistry, Analytical	11		
36	Clinical Neurology	11	3	27.3%
37	Peripheral Vascular Disease	10	4	40.0%

Acknowledgments

This work was supported by JSPS KAKENHI (Grant Numbers 16K00455 and 25280121).

References

Altmetric (2018), Altmetric.com: https://www.altmetric.com (referenced March, 2018)

Clarivate Analytics (2018), https://clarivate.com/products/web-of-science/ referenced March, 2018)

Nishizawa, M. & Sun, Y. (2012), "How are universities collaborating with industry in Japan? --- An investigation based on newspaper press releases," in Proceedings on COLLNET and WIS 2012, Soul, Koria.

Nishizawa, M. & Sun, Y. (2014), "How well do newspapers describe scientific research? --- Comparison between press releases and two different national newspapers in Japan," in Proceedings on COLLNET and WIS 2014, pp.373-379, Ilmenau, Germany.

Nishizawa, M. & Sun, Y. (2016), "How is scientific research announced in Press Release? --- Focusing on its relationships with journal indicators," in Proceedings on COLLNET and WIS 2016, Nancy, Fance.

Nishizawa, M. & Sun, Y. (2017), How is academic research reported in the media? - Relationships between attention in the press and social media, in Proceedings on COLLNET and WIS 2018, Canterbury, UK

Nichigai Associates, Inc. (2018), Language resources, http://www.nichigai.co.jp/sales/corpus.html (referenced March, 2018, in Japanese)

Nikkei press release (2018), http://release.nikkei.co.jp/ (referenced March, 2018, in Japanese)

Proceedings of 14th International Conference on Webometrics, Informetrics and Scientometrics & 19th COLLNET Meeting 2018
>> Collaboration - Impact on Productivity and Innovation <<
Article DOI: 10.22032/dbt.39328 Proceedings DOI: 10.22032/dbt.39296

Bibliometric analysis of the publication output of TU Ilmenau in the period 2012-2016

Bernd Markscheffel, Ariane Hertlein and Shihao Zhao

Bernd.Markscheffel|Ariane.Hertlein|Shihao.Zhao@tu-ilmenau.de
Institut für Wirtschaftsinformatik (Business Information Systems Engineering), Department of Economic Sciences and Media, Technische Universität Ilmenau, 98684 Ilmenau, Germany

Abstract

Bibliometrics enables quantitative analysis of research achievements by applying mathematical methods to publishing behaviour whose results can be compared within certain limits. This paper uses a bibliometric analysis to examine the publication performance of scientists at the Technical University of Ilmenau. The publications in the Web of Science for the period 2012-2016 is used as the data basis. With the help of various indicators, an objective picture of the research activities in the above-mentioned period is attempted to be drawn. The results of the bibliometric analysis are used for comparisons at different levels and then presented in different rankings. All collected data are stored in a data base structured form so that it is available as a starting point for future investigations.

Introduction

The number of scientific publications is increasing exponentially, whereas the receptiveness of the individual scientists remains limited. At the same time, there is a desire to evaluate scientific achievements, for example to gain an overview of research, to recognize trends in advance and to be able to assess the efficiency of research (Tunger, 2013).

At the beginning of the 20th century, the first bibliometric analyses were carried out, which made it possible to make statements about the quantity of the various publications (Nix, 2010). These analyses make it possible to evaluate a research achievement, among other things by citations and the number of already published own publications, whereby with increasing citation rate the supposedly most important contributions within a research area can be made recognizable. This in turn can be an impulse for other scientists to take a closer look at these publications (Tunger, 2013).

Due to rising research costs, it is nowadays necessary to present oneself as an institution with strong publishing and reputation in order to a) increase one's visibility and b) effectively raise additional research funds (third-party funds) (Havemann, 2009). Furthermore, there are ranking procedures for educational institutions, such as the CWTS Leiden Ranking (CTWS, 2017) or the Shanghai Ranking (Shanghai, 2017), in which the world's most influential universities are measured by publication output, among other things.

It is therefore becoming increasingly important to monitor the quantitative research performance of one's own institution at different levels of granularity (subject area, institute, faculty) on the basis of publication output in order to be able to assess status and competitiveness and to identify conclusions for potential expansion possibilities in research (Nix, 2010).

Methodology and Indicators

For a basic understanding of bibliometric analysis, a literature review according to Webster & Watson (2002) is carried out. This includes the collection of basic information on bibliometrics, the delimitation of terms and potential indicators. The following search terms are used: *Bibliometrie*, *Kennzahlen*, *Analyse* and the corresponding English terms *bibliometrics, indicators and analysis.*

The search engines used are Google Scholar and SpringerLink, which search for the above-mentioned search terms. In order to limit the text sources acquired in this way, criteria are selected that highlight potentially important articles. These are checked by means of backward and forward analysis for further publications to be examined. The following criteria were important in the selection process:

- Basics / Basic knowledge of bibliometrics
- Indicators
- Evaluation of indicators / criticism
- Current status of the publication
- Normalization

In order to understand and carry out bibliometric analysis, the basics of bibliometrics are necessary and a selection of indicators are required, which is why the publications are examined with regard to them. Furthermore, the identified indicators must be checked for suitability with regard to the object of investigation. Advantages and disadvantages must be weighed against each other. In order to be able to analyse across departments, one need information about the normalization of the calculated indicators (see Tunger, (2013), Ball, (2006)).

The following bibliometric indicators were used in our bibliometric analysis identified as a result of the literature search and finalized as a result of a discussion of the practicability of these indicators:

- Publications per Faculty
- Publications per Institute
- Publications by chair
- Publications per Author
- Citation rate
- Average citations per author
- h-index
- g-index
- hg index
- rational h-index
- Field normalized citation rate
- Percentage of publications cited by an author
- Percentage of uncited publications by an author
- Number of publications through collaboration.

Data collection

For the evaluation of an institution, a comprehensive set of indicators is needed to carry out the complex evaluation as accurately and objectively as possible. Each of these indicators requires a specific set of data that can be used to calculate the above-mentioned indicators. The analysis of the indicators showed that two distinct objects of investigation - the Author (A) and their publications (P) - are needed. The author includes first name, surname and the assigned subject areas, while the publication contains title, date, number of citations and the names of the authors. With the help of this data, the presented indicators can be determined almost completely. Table 1 shows an exact list of which indicators require which data.

Table 1: Data required by indicators

Indicator	*Necessary data*
Publications per Faculty	A, P,
Publications per Institute	A, P,
Publications by chair	A, P,
Publications per Author	A, P
Citation rate	A, P
Average citations per author	A, P
h-index	A, P
g-index	A, P
hg index	h-Index, g-Index
rational h-index	A, P
Field normalized citation rate	A, P, average citation per publication per field
Percentage of publications cited by an author	A, P
Percentage of uncited publications by an author	A, P
Number of publications through collaboration.	A, P

It must also be taken into account that departments or institutes of different faculties may have different communication habits that cannot be directly compared with each other and that cross-disciplinary analyses are highly negligent without considering the different publication habits. It is therefore indispensable to normalise key indicators in order not to distort the overall picture. The average citation habits required for normalization (field normalized citation rate) are determined using Web of Science. The search entry is differentiated according to the respective years and filtered according to the total number of German contributions. These search results are classified into the respective scientific categories such as chemistry or mathematics.

For the period 2012 - 2016, the search in all selected databases with the search terms for the address "Ilmenau" and with the wildcards "Il*me*au" to catch possible spelling errors delivered a total of 3138 publications. The data from the Web of Science was extracted on 04-02-2017 using a web crawler and stored as an XML file. Figure 1 shows an example of the XML publication structure.

```
<Publication Id="1" Title="Transitional boundary layers in low-Prandtl-number convection"
             Date="DEC 29 2016" Citation="0" ResearchArea="Physics"
             WoSCategory="Physics, Fluids &amp; Plasmas">
  <authors>
    <Author firstname="Joerg" lastname="Schumacher" reference="1" />
    <Author firstname="Vinodh" lastname="Bandaru" reference="1" />
    <Author firstname="Ambrish" lastname="Pandey" reference="2" />
    <Author firstname="Janet D" lastname="Scheel" reference="3" />
  </authors>
  <addresses>
    <Address university="Tech Univ Ilmenau" referenceNumber="1" institute="Inst Thermo &amp;
             Fluiddynam" />
    <Address university="Indian Inst Technol" referenceNumber="2" institute="Dept Phys" />
    <Address university="Occidental Coll" referenceNumber="3" institute="Dept Phys" />
  </addresses>
  <Info>By:Schumacher, J (Schumacher, Joerg)[      1 ] ;
              Bandaru, V (Bandaru, Vinodh)[        1 ] ;
              Pandey, A (Pandey, Ambrish)[         2 ] ;
              Scheel, JD (Scheel, Janet D.)[       3 ]
  </Info>
</Publication>
```

Figure 1: XML publication sample

Because the address details in the WoS are not stored in a standardised form and can differ from publication to publication, for example it is possible that one document contains the complete details of subject areas, institute and faculty, while others contain only the reference to the "Technische Universität Ilmenau", it was necessary to check each of the publications and to precisely allocate them to the institutional units of TU Ilmenau. This mapping was achieved with a multi-stage procedure, the first source being the website of the departments of the individual institutes and faculties, in order to identify the current employees of the departments. In addition, existing electronic telephone directories for the years 2012 - 2016 were used to identify the employees for this period. As there were still a number of authors who could not be clearly assigned, an attempt was made to identify them with their institutional description using further scientific databases (SpringerLink...). All in all, 2907 publications (99.35%) with a specific assignment of author, publication and institution could be made available for analysis.

Data processing and analysis

The data for bibliometric analysis is mapped and stored in a relational database system to ensure sustainability. The powerful query and report functions simplify the evaluation of the data. The Access 2016 database consists of the two main elements *author* and *publication*. The *university* table refers to institutional elements such as faculties, institutes and departments. The semantics of the manifold m:n relationships (e.g. such that a department belongs to several institutes, while an institute can consist of several departments, staff-department, staff-publication) leads to a series of further intermediate tables, so that the final database consists of a total of 21 tables, twelve tables containing all the necessary data to store the publications and the authors. The other tables serve to store the calculated indicators.

Results

This section presents and compares distinct results of the respective indicators for the bibliometric analysis of the publication behaviour of the TU Ilmenau.
The publications are analysed on four levels of aggregation. First, the faculties as a whole are examined and the associated publications are summed up. The same is done with the institutes, the departments and the authors. Figure 2 illustrates the overall publishing rate for each faculty

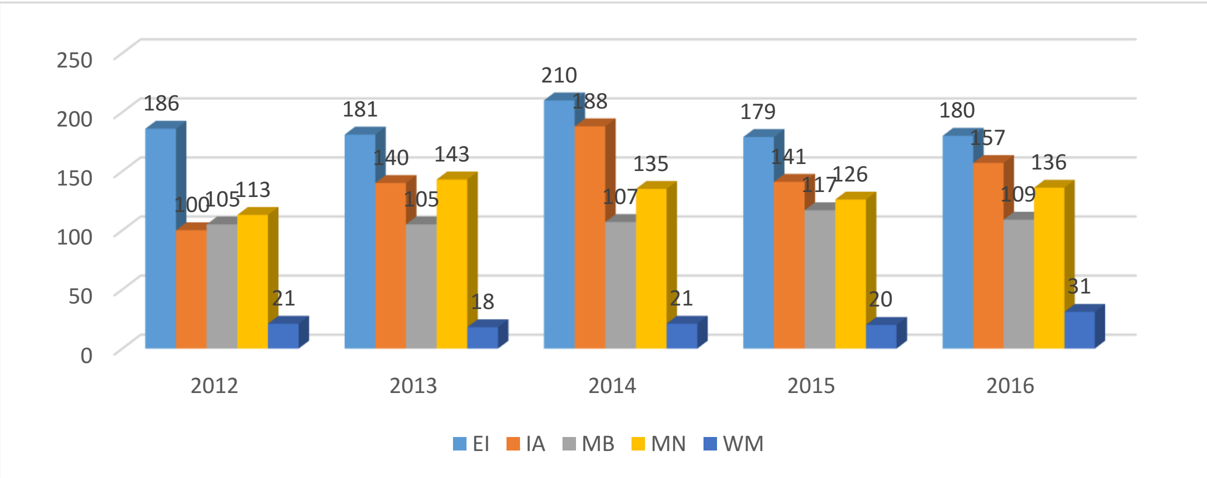


Figure 2: Publishing rate per faculty

Table 2 illustrates the top ten institute according to the publishing rate. As can be seen and as was to be expected for a Technical University, the publication landscape is dominated by the technical and scientific institutes.

Table 2: Top ten institutes (publishing rate)

	Institute	*total*
1	Institute for Information Technology	460
2	Institute for Physics	353
3	Institute for Computer and Systems Engineering	221
4	Institute for Microelectronics and Nanoelectronics	173
5	Institute for Chemistry and Biotechnology	170
6	Institute for Biomedical Engineering and Informatics	161
7	(Inter-departmental) Institute of Materials Science and Engineering	142
8	Institute for Thermodynamics and Fluid Mechanics	140
9	Institute for Theoretical Computer Science	135
10	Institute for Mathematics	130

The Technical University of Ilmenau lists 138 departments assigned to the respective institutes. There are five departments, which have published more than 100 publications during the whole period of the study.

Table 3: Top departments with more than 100 publications

	Departement	*total*
1	Electronic Measurement Engineering Group	154
2	Biomedical Engineering Group	146
3	Communications Research Laboratory	120
4	Group for Complexity Theory and Efficient Algorithms	116
5	Chemistry Group	107

A total of 2907 publications were published by 1314 authors. Table 4 lists the top five authors with the highest number of publications.

Table 4: Top five scientists (publishing rate)

	Scientist	*Total*
1	M D	110
2	M H	105
3	J H	100
4	R T	97
5	U R	80

Table 5 shows a list of the top 5 cited papers from different academic fields.

Table 5: Top five cited papers

	Paper	*Total citation*
1	Controllable Disorder Engineering in Oxygen-Incorporated MoS2 Ultrathin Nanosheets for Efficient Hydrogen Evolution	427
2	Vacancy Associates Promoting Solar-Driven Photocatalytic Activity of Ultrathin Bismuth Oxychloride Nanosheets	255
3	Binary copper oxide semiconductors: From materials towards devices	144
4	Graphene Transistors: Status, Prospects, and Problems	136
5	First-principles investigation of the size-dependent structural stability and electronic properties of O-vacancies at the ZnO polar and non-polar surfaces	120

In order to make the results from different scientific fields comparable with each other, the absolute values of the publications were normalized with the field-normalized citation rates, in keeping with the field-dependent publication habits and thus enabling an objective comparison. A result greater than 1 therefore means an above-average value for citation per publication in Germany. The citation habit for the Institute of Physics for 2013 is 13.68. The publication of the *Three-Dimensional Nanostructuring Group* of the *Institute of Physics:*"*Controllable Disorder Engineering in Oxygen-Incorporated MoS2 Ultrathin Nanosheets for Efficient Hydrogen Evolution*" thus received more than 31 times the average citation. Using this standardization approach, it was possible to compare indicators based on the publication output like h-index, g-index or hg-index of scientists throughout the university of different research areas like *Chemistry*, *Mathematics* or *Economics*.

At this point, we do not present in detail the results of the other publication-related indicators in preference to two indicators that can also be interpreted as qualitative. The ratio of cited publications to the total number of publications can be seen as a quality criterion for scientific output if we assume a certain basic quantity of publications. Of 1314 authors, 882 have at least one citation for a publication they have produced. A total of 431 scientists who have published between one and 13 publications are not cited. 219 authors have published a publication that has also been cited. 87 authors have published a publication that has been cited, as well as a publication that has not been cited. In the case of 85 scientists, exactly one cited publication is compared with 2 to 52 publications that were not cited. It is interesting to compare authors who have published more than 50 publications. The citations are put in relation to the total number of publications. Figure 19 shows this ratio for authors with a number of publications > 50.

Table 6: Scientists with the highest ratio of cited papers with more than 50 publications

	Scientist	*cited*	*uncited*	*ratio (%)*
1	H H	51	9	85
2	Y L	62	11	85
3	M K	46	10	82
4	P S	52	12	81
5	U R	61	19	76
6	A B	38	15	72

Further interesting conclusions about the way of scientific work can be made by analysing the collaboration behaviour of authors. For this purpose, the number of publications that were created in co- or multiple authorship was measured. The number of publications created through collaboration is intended to show how often employees of the TU Ilmenau cooperate. No difference is made here between cross-university and intra-university publications.

Table 7: Top five scientists in multiauthorship

	Scientist	*Total*
1	M H	105
2	R T	97
3	U R	80
4	H H	60
5	F R	57

We have further investigated these collaboration networks in further examinations. Figure 3 shows the result of a visualization of the collaboration network with VOSViewer[1].

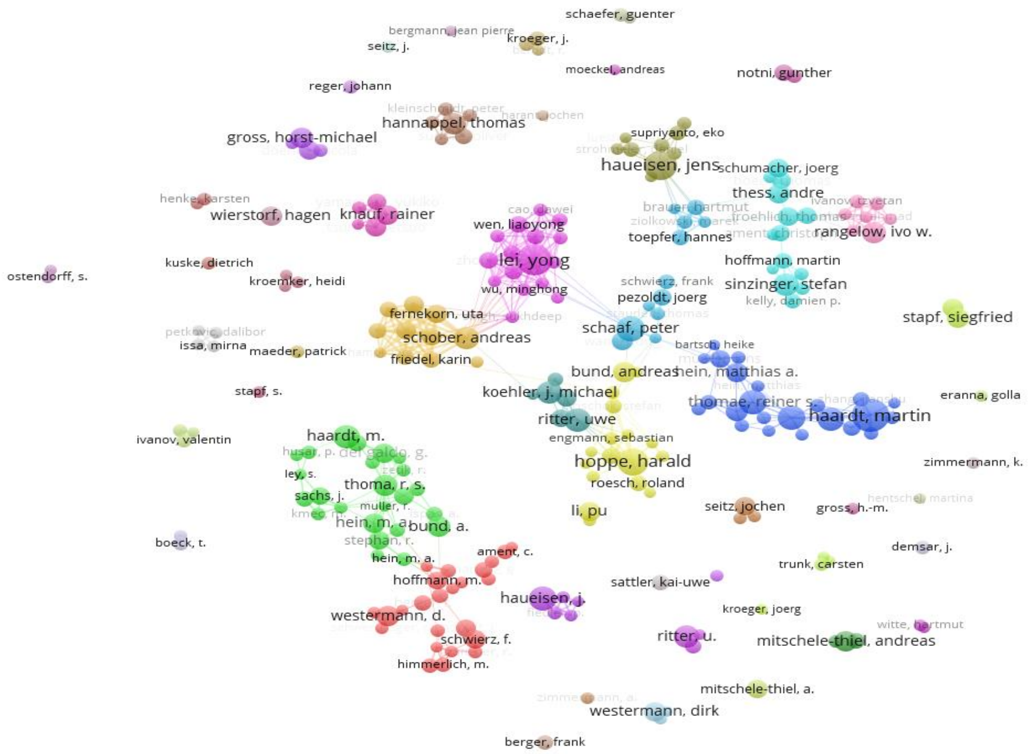


Figure 3: Co-Authorship Network of TU Ilmenau 2012 – 2016 (threshold 10)

Obstacles

The correctness of the underlying data material is a requirement for high-quality bibliometric analysis. In our specific example we had to deal with a number of error sources typical for such bibliometric investigations.

Accuracy of the data

- The data required for the bibliometric analysis were not always available in correct form. The author names were neither available on the website of the Technical University of Ilmenau, nor on the Web of Science, where they were one hundred percent error-free. Since each department maintains its own employee list online, there was no standardized form of representation and an automated extraction of the names was therefore not possible (use of different separators in the name separation). Especially with foreign names, there were often uncertainties as to which parts belong to the first name and which to the last name. As long as not all authors are registered in standardized repositories (ORCID, Researcher-ID or Scopus Author ID or similar), the solution can only consist of time-consuming manual post-processing. In our case, the author data filtered out of the website were compared with those of the telephone lists. If a person from the telephone lists did not exist with a complete first and last name in the employee list, the system searched for the last name and the first letter of the first name. If a hit was found, we manually checked whether this corresponded to the correct person. Similar problems with the assignment of names occurred in the Web of Science.

[1] The detailed outline of the results of the data science analysis of the available data sets will be the content of another publication in the pipeline.

Incorrectly written names were checked manually. For this purpose, a list was programmed to save these exceptions and then transferred it to the Access database.

- An essential requirement for the correct automated extraction of data from the Web of Science, was the correct pre-structuring with appropriate separators (bsw. comma between first and last name). If such a separator does not exist, the corresponding attribute in the raw XML file remains empty and the assignment must be completed manually. Especially often such a manual check was necessary for names from other cultures.
- For a number of publications, the authors contained references to institutional units that could not be assigned to them. In order to be able to guarantee the affiliation, all Ilmenau-related authors of a publication were first collected. The authors were then compared with the existing staff database (consisting of the data from the TU Ilmenau website and the telephone lists for the years 2012 to 2016). If there was an entry with a complete first and last name, the search was based on the number of departments. If this search yielded only one result, it was assigned to the employee. Otherwise, this data was collected and later assigned manually. A tool was programmed to select the right authors and subject areas.

Duplicate entries

- When collecting the data, it was found that publications (either as e-books or as articles available elsewhere) were published twice in the Web of Science. In order to filter these out, identical publications were searched for and the publication date as well as the names of the authors were checked. If they matched, all identical publications were combined into one. In concrete terms, this means that one publication was removed from the database and the citations were grouped.

Missing data

- Some addresses were not available in the publications, and therefore had to be added, i.e. also that no references were available. In order to correct this, the employee list was searched for hits and the author names were additionally checked by an additional Google search to ensure assignment. The author names were examined with the help of third-party sources (such as SpringerLink and Google Scholar) in which the reference to the corresponding addresses was available and an assignment of the institutional affiliation ("affiliation") could be carried out.

Summary and Future Work

With the help of a bibliometric analysis, a cross-disciplinary comparison of the publication performance of the TU Ilmenau was carried out. The available results allow the publication output of the past five years to be analysed and the strengths and weaknesses of scientists and their corresponding organisational units to be uncovered. There are considerable differences in publication behaviour between the different university tiers. A number of factors that have contributed to these results are relevant for the present results. The size of the departments and institutes can be an advantage over the smaller sized units and can have a direct influence on the position in the respective ranking, since the number of actively publishing employees can correlate with the amount of publication output within a period. By means of normalization, the indicators of specific groups were made comparable despite different citation habits. In this paper, a series of indicators were used to describe the research landscape at TU Ilmenau as objectively as possible. The Data acquisition was one of the most complex parts of the work. A number of programs were developed to convert the raw data into a usable form suitable for

bibliometric analysis. In order to provide consistent and correct data for the analysis, the data was manually checked in several stages (after each processing step) in order to minimize potential sources of error. For the future, further use of the data, they were stored in a structured form in a database.
This database can be used as a basis for further investigations. It can serve as a basis to re-examine publication behaviour at regular intervals, to draw comparisons and to work out trends and tendencies. It was also used for the work currently being completed on more content-oriented analyses, which have a more Data Science-specific character, to answer such research questions like: If there are changes in research areas, are there overlapping research areas that offer the potential for new opportunities for scientific cooperation, or how detailed the research landscape of a university can be analysed by including more content specific elements like keywords and abstracts.

References

Alonso, S.; Cabrerizo, F.J.; Herrera-Viedma, E.; Herrera, F. (2009) h-index: A Review Focused in its Variants, Computation and Standardization for Different Scientific Fields. *Journal of Informetrics*, 3(2009)4, pp. 273-289.

Ball, R. (2006). Der Hirschfaktor - ein neuer “Stern” am bibliometrischen Indikatorenhimmel. *B.I.T.-Online*, 9(2006)4, pp. 309-312.

Ball, R. (2007). Wissenschaftsindikatoren im Zeitalter digitaler Wissenschaft. *B.I.T.-Online*, 10(2007)2, pp. 121-125.

CWTS *Leiden Ranking* (2017). Retrieved December 19, 2017 from: http://www.leidenranking.com/ranking/2017/list.

Havemann, F. (2009). *Einführung in die Bibliometrie*, Berlin, 2009. 66 p.
Retrieved February 7, 2018 from
http://www.wissenschaftsforschung.de/Havemann2009Bibliometrie.pdf,

Schulze, N.; Michels, C.; Frietsch, R.; Schmoch, U.; Conchi, N. (2013). *Indikatorbericht Bibliometrische Indikatoren für den PFI Monitoring Bericht.* 2013. 85 p.
Retrieved March 7, 2018 from:
https://www.isi.fraunhofer.de/content/dam/isi/dokumente/ccp/2012/Indikatorbericht_PFI_2012.pdf.

Shanghai (2017). *Academic Ranking of World Universities* 2017. Retrieved December 19, 2017 from: http://www.shanghairanking.com/Shanghairanking-Subject-Rankings/.

Tunger, D.: *Möglichkeiten und Grenzen bibliometrischer Analysen*. Forschungszentrum Jülich, 2013

Webster, J, Watson, R.T.: Analyzing the Past to Prepare for the Future:Writing a Literature Review. *MIS Quarterly*, 2(2002)26, pp. xiii-xxiii.

Article DOI: 10.22032/dbt.39330 Proceedings DOI: 10.22032/dbt.39296

Return of Investment (ROI) in Research and Development (R&D): Towards a framework

Dr Mohammad Hassanzadeh

Associate Prof, Head, Knowledge and Information Science Dept, Faculty of Management and Economics, Tarbiat Modares University, Tehran- Iran, Tel: +98-2182883678, +98-9198771218,
hasanzadeh@modares.ac.ir

Tahereh Bozorg Bigdeli

Senior Researcher, Ayandeh Aoozane Ata, Modares S&T Park, Tehran- Iran, Tel: +98-2182883678,
bbigdelit@gmail.com

Abstract
There is a significant relationship between the development of countries and their investment on research and development (R&D). Often, investments on science is seen as a resource of national prestige in countries. An outstanding part of spending go to the universities as part of science policy utilization. For that, most of investigations on return of investments (ROI) on science focus on universities as the main authorities of scientific endeavor. The main issue is that, calculating ROI of R&D is not as simple as the industry of service sectors. In industry and service input and output of the process is clearly calculable. For example a manger in industry sector is able to calculate the cost of product and services. But in science it is not simple to calculate the cost of products because of variety and complexity of resources allocated for its production. A paper as an output of a scientific endeavor easily is not calculable in terms of financial formula. Sometimes a scientific finding save the life of millions of people and provide society with bunch of opportunities and guide the communities towards prosperous life. These are not simply calculable in term of ROI investigations. Most of scientometrics studies focuses on the determined document based outputs of the R&D systems. Since financial inputs play crucial role in progress of R&D systems, taking them in account when calculating the efficacy of this system will provide us with clearer image of R&D performance. This research aimed to focus more on financial aspects of R&D performance in universities forward utilizing some sort of knowledge economy framework. For that, we strive to formulate the expenditure has been carried out in various levels of a university which results in a unit of scientific paper, highly cited paper, fruitful actors, and etc. we have focused on the Tarbiat Modares University (TMU) a fully accredited state university for graduate studies with more than 7000 students and almost 1000 faculty members. TMU always ranks between top 3 Iranian universities.
Keywords: Return of Investment, Framework, Scientometrics

Introduction
Most of scientometrics studies focuses on the determined document based outputs of the R&D systems. Since financial inputs play crucial role in progress of R&D systems, taking them in account when calculating the efficacy of this system will provide us with clearer image of R&D performance. This research aimed to focus more on financial aspects of R&D performance in universities forward utilizing some sort of knowledge economy framework (Hassanzadeh, Akhgar and Navidi, 2014). For that, we strive to formulate the expenditure has been carried out in various levels of a university which results in a unit of scientific paper, highly cited paper, fruitful actors, and etc.
Return on Investment (ROI) is defined as the ratio of gains from investment and is used normally as an index to measure the performance and evaluate the efficiency of an investment on some project or initiative or compare the efficiency of a number of different investments.

ROI was formulated to measure the amount of return on a particular investment, against the investment's cost. Return on investment is divided by the investment costs. The result is expressed as a percentage or ratio. The classic formula is as following:

$$\text{ROI} = \frac{(\textbf{Gain from Investment} - \textbf{Cost of Investment})}{\textbf{Cost of Investment}} \times 100$$

In the formula, "Gain from Investment" refers to the proceeds obtained from the sale of the investment of interest. Because ROI is measured as a percentage, it can be easily compared with returns from other investments, allowing one to measure a variety of types of investments against one another (Investopedia: ROI, 2018).

ROI can be used in conjunction with Rate of Return, which takes in account a project's time frame. One may also use Net Present Value (NPV), which accounts for differences in the value of money over time, due to inflation. The application of NPV when calculating rate of return is often called the Real Rate of Return.

Social return on investment

Traditional ROI formula was criticized because of its deficient in calculating the real return and benefits of the investments. Some investors and businesses have taken an interest in the development of a new form of the ROI metric, called "Social Return on Investment," or SROI which intended to cover social and environmental metrics that currently do not reflected in conventional financial accounts. It was initially developed in the early 2000s and takes into account broader impacts of projects using extra-financial value. Social return normally generalizes the return measures to social context and helps understand the value proposition of certain ESG (Environmental Social & Governance) criteria used in socially responsible investing (SRI) practices. Undertakings for sustainability in terms of expenditures on infrastructures to reduce energy consumptions and other types of investments which may not be returned completely but have an immediate cost which may negatively impact traditional ROI - however, the net benefit to society and the environment could lead to a positive SROI. (Investopedia: ROI, 2018).

Social media has imposed a big change to social relationships and social network as an emerging concept deals with centrality and betweenness of actors in a network. In a social network people interact with each other by seeing and liking or mentioning posts and updates. These are new flavors of ROI that have been developed for particular purposes but are not accounted in terms of traditional formula. Similarly, marketing statistics ROI tries to identify the return attributable to advertising or marketing campaigns. So-called learning ROI relates to amount of information learned and retained as return on education or skills training. As the world progresses and the economy changes, several other niche forms of ROI are sure to be developed in the future (Investopedia: ROI, 2018). We can add other ROI indexes such as gained expertise, best practices, shared knowledge and savings because of repeated experiences.

Social return on investment (SROI) is a collection of techniques for measuring values that are not traditionally reflected in financial statements. These values may include social and environmental factors. They have been set to determine how effectively an enterprise leverage its capital and other resources to create value for the community beyond the organizational borders. While a traditional cost-benefit analysis is used to compare different investments or projects, SROI is used more to evaluate the general progress of certain developments, showing both the financial and social impact the corporation can have (Investopedia: SROI, 2018).

SROI is useful to corporations because it can improve program management through better planning and evaluation. It can also increase the corporation's understanding of its effect on the community and allow better communication regarding the value of the corporation's work (both internally and to external stakeholders). Philanthropists, venture capitalists, foundations and other non-profits may use SROI to monetize their social impact, in financial terms.

A general formula used to calculate SROI is as following:

$$\text{SROI} = \frac{(\textbf{social impact value} - \textbf{initial investment amount})}{\textbf{initial investment amount}} \times 100$$

Assigning a money value to the social impact can present problems, and various methodologies have been developed to help quantify the results. The Analytical Hierarchy Process (AHP), for example, is one method that converts and organizes qualitative information into quantitative values.

While the approach varies depending on the program that is being evaluated, there are four main elements that are needed to measure SROI:

- Inputs, or resources investments in your activity (such as the costs of running, say, a job-readiness program)
- Outputs, or the direct and tangible products from the activity (for example, the number of people trained by the program)
- Outcomes, or the changes to people resulting from the activity (i.e., new jobs, better income, improved quality of life for the individuals; increased taxes for, and reduced support from, the government)
- Impact, or the outcome less an estimate of what would have happened anyway

Research and development and ROI

Despite other businesses, research and development is a process which results in a particular product or service. Output of a research initiative may be translated into an innovative process or improvement in a social service. Sometimes a research project terminate without a tangible achievement but leaves outstanding experiences for new projects. These kind of gains always is ignored in ROI calculations. In addition to experiences, some other gains also is achieved through research and development which are important but do not counted in ROI formula. A new formula will be as following.

$$\text{R\&DROI} = \frac{(\textbf{Gain from Investment+Soclial Impact Value+National Prestige+Perpetuated gains}) - \textbf{Cost of Investment})}{\textbf{Cost of Investment}} \times 100$$

The R&D returns not only include direct returns on investments but covers social impacts of the investment and organizational and national prestige which influencing the gains from the investment. Companies with higher expenditure on research and innovation, get more attentions and supports from customers because of the potential impact of R&D on quality measures. In other hand, governments indicating their likelihood to sustainable development and science advancement by increasing R&D budgets and briefing it as a measure to increase their popularity among voters and taxpayers.

Research design

Besides depicting descriptive representation of background data for TMU, this research mainly was carried out using scientometrics techniques to analyze scientific performance of the university in terms of papers published in international peer review journal, citation to documents, citation and self-citation per paper, national and international collaboration rate and so forth. In addition we have done co-authorship analysis to depict individual and collective performance of faculty members. The data was extracted from Thomson Reuter's Web of Science (WoS) database. In the second phase, we have analyzed the expenditures of the university on R&D to calculate the cost of each scientific achievement in terms of outputs and outcomes. Applications like Bibexcel, Notepad+, SPSS and other bibliometrics analysis kits have been performed. Finally, we approached to propose an ROI model of R&D expenditures in the university.

Findings

Tarbiat Modares University as a higher education institute which provides only postgraduate studies, is the second ranked university among Iranian universities. Research policies in the university tended to focus on quality research and publishing in world class impacting journals. Findings indicated that, researchers affiliated with TMU, have been published 12394 documents in the sources indexed in WoS from the beginning of the university up to 2015. In general, among the 11 types of documents, journal papers includes almost 84 percent of publications. Besides the publications in national journals which almost all are in Persian and are counted in a separate citation database[1], more than 99 percent of documents have been published in English.

The first document published was in 1994 near to 4 years after the establishment of the university. The main increase in publishing starts from year 2000 and 2015 witnesses the most publications by TMU researchers. Publications by researchers affiliated with the university have been increased from 107 in 2000 to 2500 in 2016. This indicates more than 20 fold growth in terms of tangible research out puts.

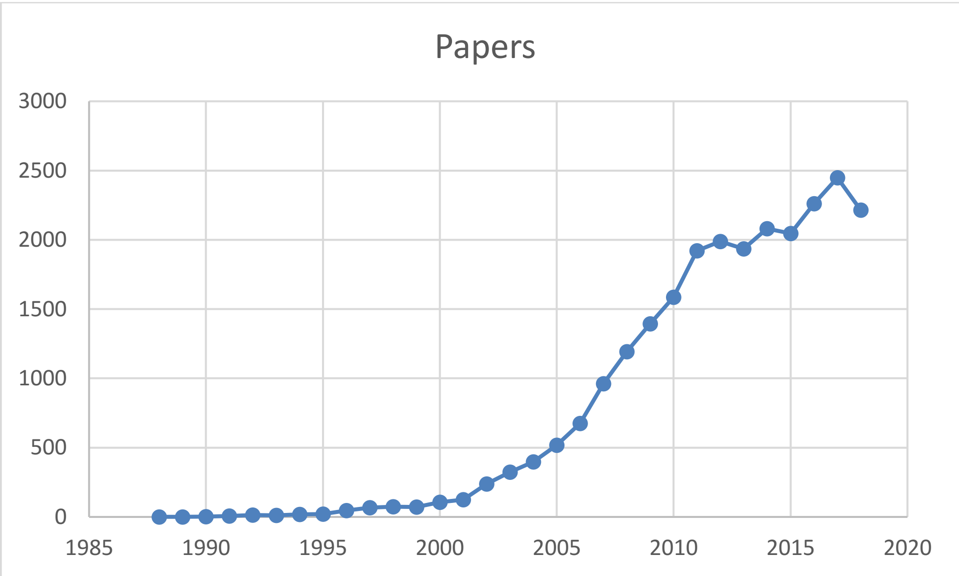


Figure 1. increase in scientific production by Tarbiat Modares University (Clarivate Analytics, 2015)

Expenditure in research affairs as an index of input measures shows an increase in terms of 1000 dollar per paper. This means that, from year 2000 to 2016 expenditure per paper has been increased and this may interpreted as decrease in researchers' productivity and based on such a conclusion, university research policy makers, may be advised to decrease expenditure on research infrastructures.

[1] Iran recently has been established a citaion analysis system under title ISC which stands for Islamic countries Science Citation in southern city Shiraz. This organization in collaboration with journal publishers in islamic countries and in accordance with Organization of Islamic Coperation (OIC) bodies strives to analyze the R&D outputs and performance.

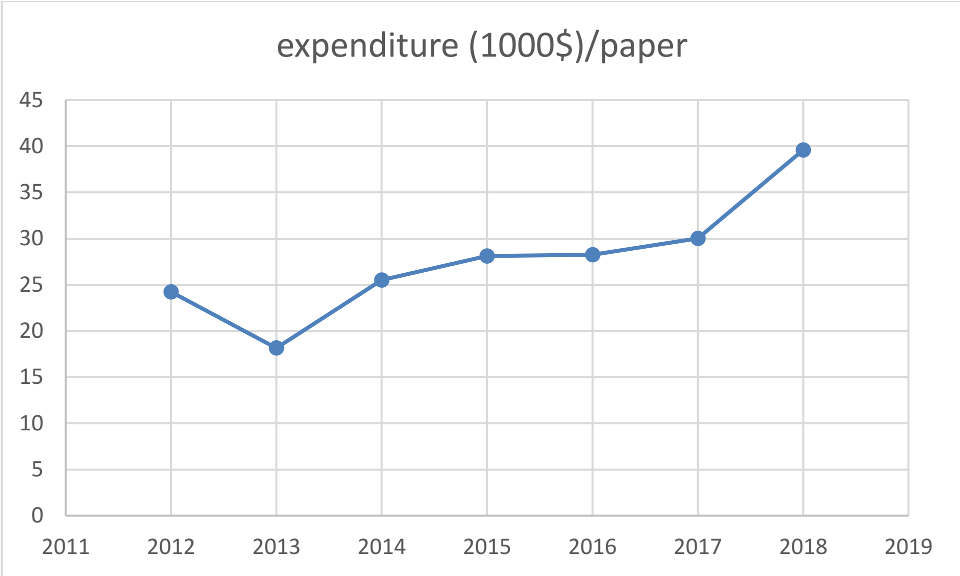


Figure 2. Mean expenditure per paper in Tarbiat Modares University (Saber and Hassanzadeh, 2016)

But based on new R&DROI formula we should consider the social impact and organizational and national prestige of the publications. Each publication in a prestigious journal, creates an opportunity to promote the university. Beyond that, findings of research which is reflected in scientific social networks may be used to solve several problems worldwide. Social impact and prestige as an element of return in research expenditures persuade policy makers in organizational and national levels to endure their support of research affairs.

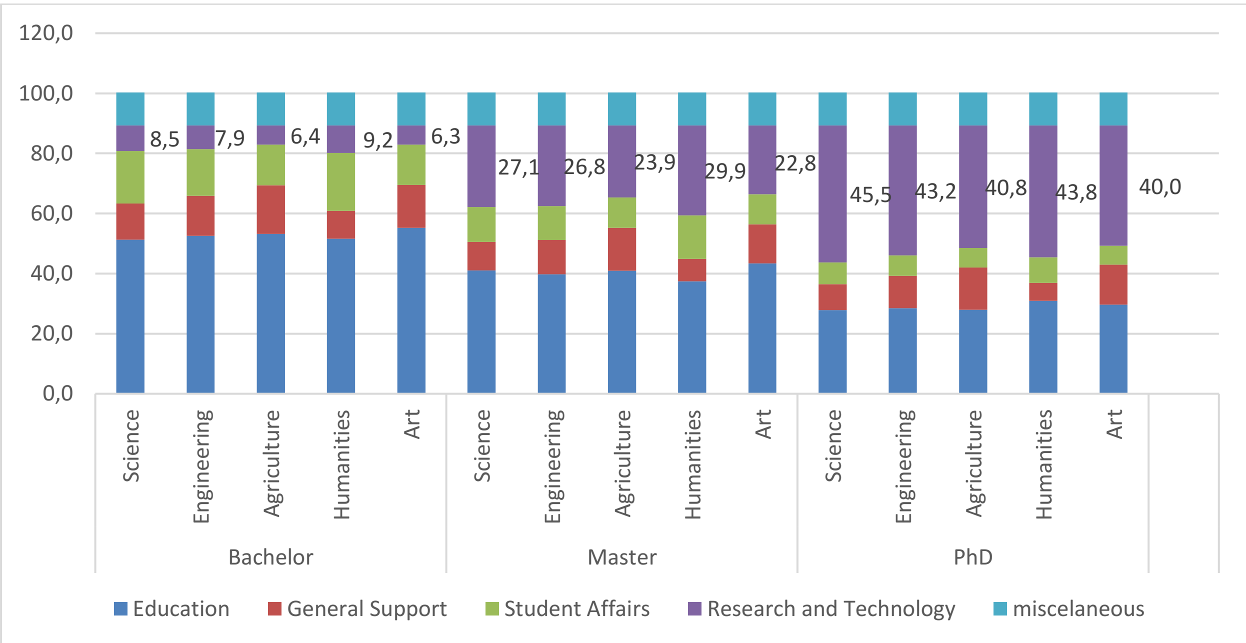


Figure 3. Share of research and technology expenditures by items, different science areas and degree

Expenditure on research tremendously varies from bachelor to PhD degrees. While in bachelor degrees only 6 to 8.5 percent of expenditures goes to research affairs, this rate hikes to more than 45 percent in PhD degrees. The increase in investment on research in higher degrees has its roots in the tendency of these courses to focus on research and scientific out puts. Students entering in post-graduate studies mainly approaching towards a research initiatives. The more university authorities engaging in research impact, the more return on investment is realized. Emphasizing on research impact will increase the return rate as well.

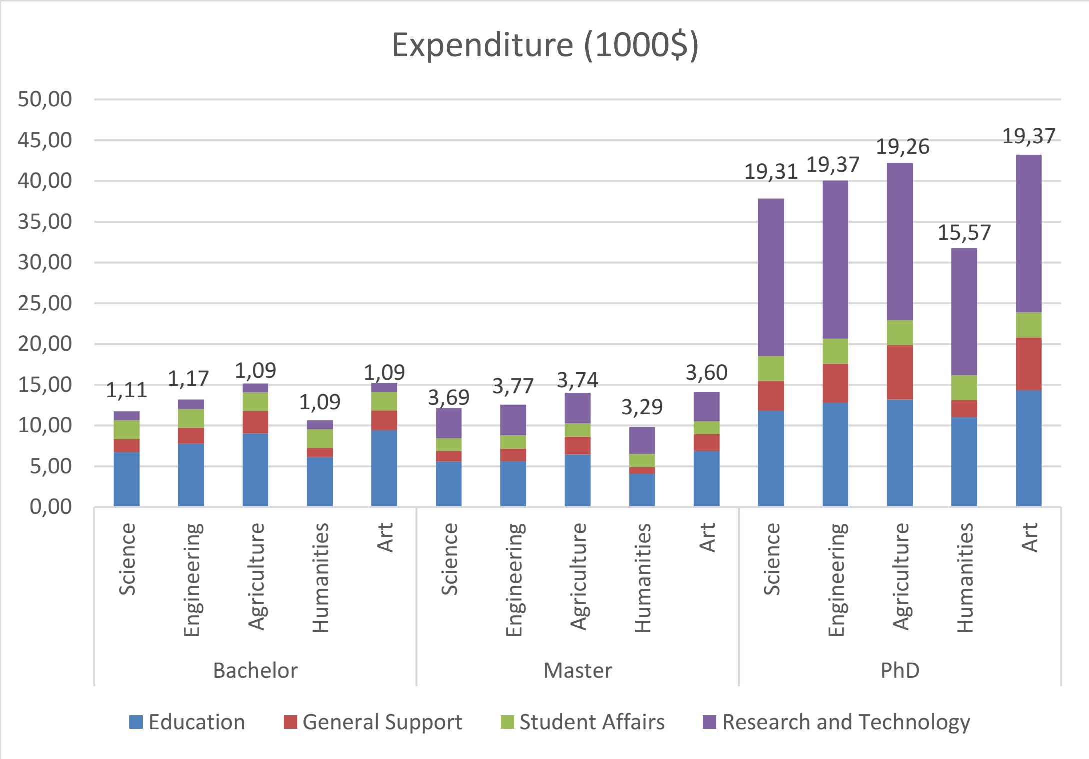


Figure 4. Expenditure on research in different science areas by degree (1000 $)

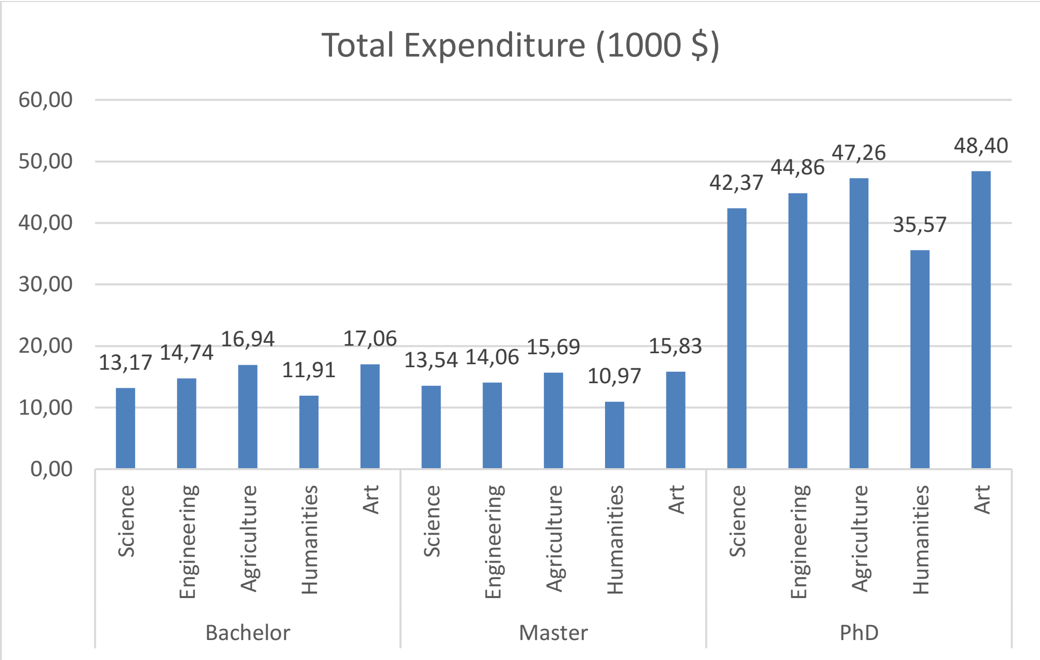


Figure 5. Total expenditures by degrees (1000$)

Expenditure on education including research and development increases moving from bachelor to PhD degree. Share of research from total expenditures increase in higher degrees. Tremendous increase in expenditures on research and development initiatives in universities only can be justified by entering the new components to the formula as explained in the previous section. By including organizational, national prestige and contribution to the global knowledge which in turn contributes to the global sustainability we will be able to calculate return of investments on research and development. Quantifying all factors which deserve to be included in the formula will result in a comprehensive ROI model.

Concluding remarks

R&D expenditures also shows the same journey as the publications. R&D expenditure well predict the scientific outputs but the main problem is that, financial investment in universities comes from various sources with various objectives. While government and social service sector expenditure on R&D targets national prestige and public responsibility on science, industry investment more intended to develop a prototype or a process kit. Formulating these heterogeneous elements in a framework will be the outstanding contribution to the scientometrics discipline.

A new model of ROI for research and development will include following components:

A. Investments
    1. Research personnel salary
    2. Technical infrastructure including scientific resources, materials, laboratories and etc.
    3. Miscellaneous costs
    4. Investments
B. Gains
    1. Direct incomes
    2. Social impact
    3. Sustainability measures
    4. Knowledge and learning
    5. Savings in future projects
    6. Skills and competencies

Since some of the mentioned components are qualitative they need to be qualified, consolidated and formulated towards a comprehensive model.

References

Hassanzadeh, Mohammad, Akhgar, Babak and NavidI, Fatemeh (2014). Integrating Knowledge and Science indicators (Kientometrics) towards a rational framework of investigations: a comparative approach. *Collnet Journal of Scientometrics and Information Management*, 8 (2), 247-256.

Investopedia. (2018). Online Resource. Entry: ROI. Retrieved: 2018/07/21.

Investopedia. (2018). Online Resource. Entry: SROI. Retrieved: 2018/07/21.

Saber, M. (2016). Investigation on the research productivity of Tarbiat Modares University. Thesis – M.A.. Tarbiat Modares University. Supervisor: Mohammad Hassanzadeh. Tehran, Iran.

Article DOI: 10.22032/dbt.39339 Proceedings DOI: 10.22032/dbt.39296

Tracing the scientific outputs of Iranian papers on Dermatology research based on publications in the Web of Science

[1]Fatemeh Makkizadeh , Esmaeal Bigdeloo MA[2] and Ali Taherizadeh BA[3]

makkizadeh@yazd.ac.ir
School of Social Sciences, Knowledge and information Science,
Yazd University, Yazd, Iran, Tel:09133516064

Abstract

Dermatology is the branch of medicine dealing with the skin, nails, hair and its diseases. It is a Specialty with both medical and surgical aspects. Systematic research plans on any subject, including dermatology. Dermatology is in need of solid data to identify the gaps in the research. This study aimed to elucidate the most important trends, directions, and gap in this subject. The data, extracted from the Institute for Scientific Information, were used to perform a bibliometric analysis of the scientific productions (1974–2016) about dermatology. Specific parameters related to dermatology were analyzed to obtain a view of the topic's structure and document relationships. Additionally, the trends and authors in the most influential publications were analyzed. The results demonstrate the growth of scientific production in this field between 1974 and 2015. 533 institutes which were involved in writing the papers, with the Tehran Univ Med Sci and Shiraz Univ Med Sci at the top of the list respectively. According to the Betweeness Centrality Indicators of terms on the maps, the most active research areas in the field are as follows: Burn, Epidemiology, Treatment, Scar, Quality of Life, Vitiligo, Wound Healing, Platelet-Rich Plasma, Mycosis Fungoides, and Cryotherapy. Through performing the first scientometric survey on dermatology research, we analyzed the characteristics of papers and the trends in scientific production. Co-word analysis revealed outstanding topics of the field, which is useful for policy makers to learn about the research status and make appropriate decisions for the promotion of scholarly products.

Keywords: Dermatology; scientometric Analysis; Scientific output

Introduction

Dermatology is the branch of medicine dealing with the skin, nails, hair and its diseases. It is a Specialty with both medical and surgical aspects (Random, 1997). In practice, dermatology includes all aspects of diseases (both internal and external factors) which affect the skin and its contents such as hair, nails, sweat glands as well as oral mucus membrane and external genital membrane. Sexually transmitted diseases are also categorized in dermatology sphere (Burns et al. 2010). There are as many as 3,000 disorders in dermatology area. Such a large number includes numerous groups with a huge variety in terms of etiology: from genetic disorders to infectious diseases, caused by environmental factors, multifactorial disorders, together with a large number of idiopathic diseases. If the outbreak and consequences caused by such diseases are taken into account, differentiation and spectrum spread of such a large number of diseases will even increase (Bickers et al, 2006, Khatami and Zartab, 2011).

Based on some statistics, at every period of time, one-fourth to one-third of people suffers at least one single skin disease (Burns et al. 2010). However, more recent studies reveal that the outbreak of skin diseases are far more both in developed and developing countries (Wolkstein, et al., 2003). Meanwhile, there is a rampant pattern of skin diseases in these countries with

Authors' affiliations:
1- School of Social Sciences, Knowledge and information Science, Yazd University, Yazd, Iran
2 - School of Social Sciences, Knowledge and information Science, Yazd University, Yazd, Iran
3 - Bachelor of English Literature; Yazd University

different dermatologic disorders. In industrial countries, those who are interested in dermatologic epidemiology have found it interesting to focus on emerging infectious diseases such as Herpes simplex, genital warts, and environmental diseases such as skin damages caused by sun exposure, malignant melanoma, and non-melanoma skin cancer. Bickers et al, in a new study, have reported that 165 million Americans suffered from Herpes complex and Herpes zoster in 2004. They added that in 2004, three out of 83 million had been diagnosed with genital warts and human papillomavirus (HPV). There were more skin disease claims across the US population in 2013 than cardiovascular disease, diabetes, or end-stage renal disease, a separate report revealed (Edison and Bruce, 2017). Skin diseases are considered as the main causes of inability in developing countries. In such territories, affliction with dermatologic diseases varies from 20% to 80% (Hay et al., 2006). Where infections and contaminations are chiefly blamed (Rosenbaum, et al., 2017).

Figueroa and et al. implemented a survey study aimed at studying dermatological community-based needs in a rural community southwest of Ethiopia. They came to the conclusion that the most common skin diseases were contaminations caused by parasites such as scabies, pédiculose and onchocerciasis (46%), followed by bacterial and fungal infections (33%) (Figueroa et al., 1998. Across the globe, there are numerous common dermatosis like acne and psoriasis. Almost all dermatologic diseases have one common characteristic: they are visible. Such property can often cause a burden for patients. The disease itself is bothering (intense itching for instance) and the appearance problem, which has been reported in 68% of patients, doubles the problem. Besides, limited activity will result in high treatment costs as well as social inabilities (Edison and Bruce, 2017, Johnson, 2004).

Overall, research and studies on dermatologic subjects are crucial due to the following reasons:

- Skin diseases are very common, affecting up to a third of the population at any one time.
- Skin diseases have serious impacts on life. They can cause physical damage, embarrassment, and social and occupational restrictions. Chronic skin diseases may cause financial constraints with repeated sick leave. Some skin conditions can be life-threatening.
- Health expenditure for skin diseases is high.

Therefore, many investigators have published articles in this field [(Rosenbaum, et al., 2017, Storan and Irvine, 2017, Bakker, 2012). However, there has been no systematic analysis of this increasing number of papers. A scientometrics method is one that measures and analyzes scientific publications related to a specific topic regarding the trends in citations, most important content, authors, and journals. A widespread use of scientometric method goes back to 1960s when Eugene Garfield finalized the construction of Science (Garfield, 1964). This method is useful for assessing the scientific advancements and motivations of researchers and determining current research directions in a specific field; such data would be extremely useful for guiding subsequent research designs as it will predict how this field would move forward (Hendrix, 2012, Hendrix, 2010). The aim of this study was to perform a scientometric analysis of articles on dermatology by Iranian researchers from 1974 to 2016 with the help of bibliometric indicators using the Institute for Scientific Information (ISI) Web of Science.

Materials and Methods

A bibliometric study was performed on the articles related to "Dermatology" published between 1974 and 2016. Our main source was chosen from the ISI Web of Science database, available at http://www.isiknowledge.com because it is one of the major sources for bibliometric, citation and other academic impact information of scientific articles in various branches of sciences. All three resources available in the ISI web of science were used for this purpose (Science Citation Index Expanded; Social Sciences Citation Index; the Arts & Humanities Citation Index). Terms, used for the searches, were chosen in accordance with Web of science category, WC=Dermatology and restricted to country, CU=Iran. Which yielded a total of 837 publications. Our search covered papers published between 1974 and 2016. We conducted the search on July 20, 2017.

Specific parameters such as the publication year, articles' language, subject distribution, first author, main journals in this field, citations of the paper by other papers, and institutional affiliations were retrieved from the ISI and analyzed with the analyze function provided by the ISI database. Each journal's impact factor was retrieved from the Journal Citation Reports available at http://scientific.thomson.com/products/jcr. All statistical analyses were performed using Microsoft Excel 2010 computer spreadsheet software and UciNet software. A Thematic structure of the dermatology field was created by including articles with a frequency of 4 up. 60 documents which had these characteristics are selected.

Results

Annual Publication Number during 1974 -2016

There were a total of 837 research articles on dermatology in the ISI Web of Science during 1974 -2016. These papers were drafted by 2359 authors, 48 countries, 533 institutions and were published in 62 journals in 2 languages. In Fig. 1, time trend of the number of articles is shown. The 77 published articles in 2016 compared to 1 articles in 1974 shows a 70-fold increase (Fig.1).

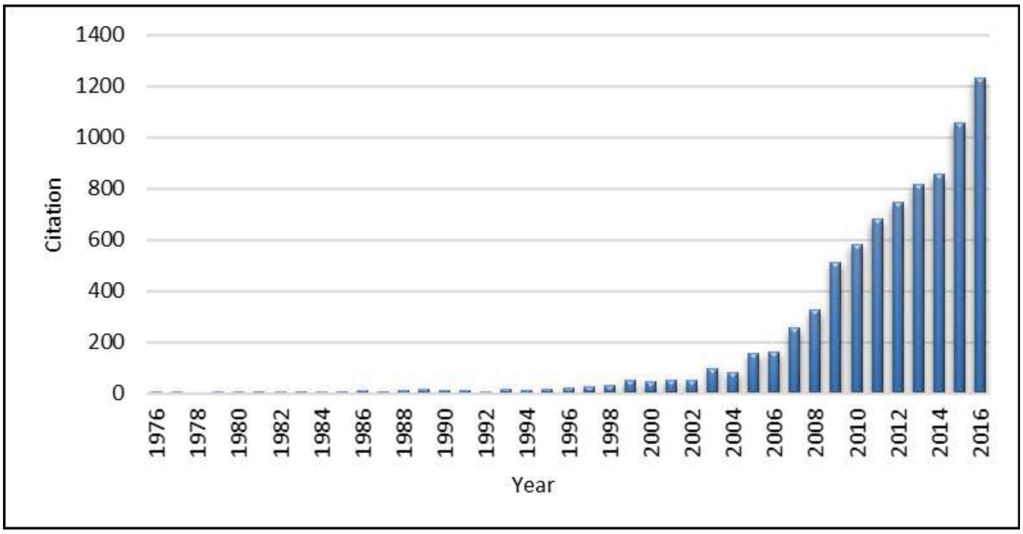


Figure 1. Trend in the number of the dermatology publications by year

Citation Profile of Articles

The number of citations cannot always be used to judge the quality of a paper, but it is a measure of its impact on Subject of research (Zhang et al., 2016). Total of citations were 8075 times. The average citations per paper (C/P) was 9.65. Table 1 shows the top 10 highly cited articles in this field.

The most frequently cited article was ‘Skin Manifestations of Mustard Gas- a Clinical- Study of 535 Patients Exposed to Mustard Gas’ published in 1992 by Momeni; Enshaeih; Meghdadi; et al. It was cited 111 times as it was after 16 years that the first published in dermatology in 1974, which vastly exceeds the citation of other articles on dermatology.

Table 1. Articles with highest number of dermatology-related citations

NO	Title	Author	Citation	Journal	IF
1	Skin Manifestations of Mustard Gas- A Clinical- Study of 535 Patients Exposed to Mustard Gas	Momeni, Az; Enshaeih, S; Meghdadi, M; Et Al	111	Archives Of Dermatology.1992;128(6):775-780	4.789
2	Prevalence, Severity, and Severity Risk Factors of Acne in High School Pupils: A Community-Based Study	Ghodsi, S. Zahra; Orawa, Helmut; Zouboulis, Christos C.	103	Journal Of Investigative Dermatology.2009;129(9):2136-2141	6.287
3	Pemphigus: Analysis of 1209 cases	Chams-Davatchi, C; Valikhani, M; Daneshpazhooh, M; et al.	93	International Journal Of Dermatology.2005;44(6):470-476	1.56
4	Cutaneous leishmaniasis: Clinical aspect	Dowlati, Y	88	Clinics In Dermatology.1996;14(5):425-431	2.253
5	Comparison Between The Efficacy Of Photodynamic Therapy And Topical Paromomycin In The Treatment Of Old World Cutaneous Leishmaniasis: A Placebo-Controlled, Randomized Clinical Trial	Asilian, A.; Davami, M.	86	Clinical And Experimental Dermatology.2006;31(5):634-637	1.589
6	Randomized controlled open-label trial of four treatment regimens for pemphigus vulgaris	Chams-Davatchi, Cheyda; Esmaili, Nafiseh; Daneshpazhooh, Maryam; et al.	82	Journal Of The American Academy Of Dermatology.2007;57(4):622-628	7.002
7	Epidemiology and mortality of burns in the South West of Iran	Panjeshahin, MR; Lari, AR; Talei, AR; et al.	77	Burns.2001;27(3):219-226	2.056
8	New Combination Of Triamcinolone, 5-Fluorouracil, And Pulsed-Dye Laser For Treatment Of Keloid And Hypertrophic Scars	Asilian, Ali; Darougheh, Afshin; Shariati, Fazlolah	73	Dermatologic Surgery.2006;32(7):907-915	2.351
9	Dermatophytoses In Iran	Khosravi, Ar; Aghamirian, Mr; Mahmoudi, M	72	Mycoses.1994;37(1-2):43-48	2.252
10	A Model For Thermal Ablation Of Biological Tissue Using Laser-Radiation	Partovi, F; Izatt, Ja; Cothren, Rm; Et Al.	72	Lasers In Surgery And Medicine.1987;7(2):141-154	2.312

Until 1999, papers about Dermatology were rarely cited. Since then, citations have consistently increased; papers about Dermatology were cited 5 times in 1999. Other peaks occurred in 2013, to 2016 (Fig. 2). The number of citations closely followed the number of publications. The number of citations per year reached its maximum in 2016 (1231 citations).

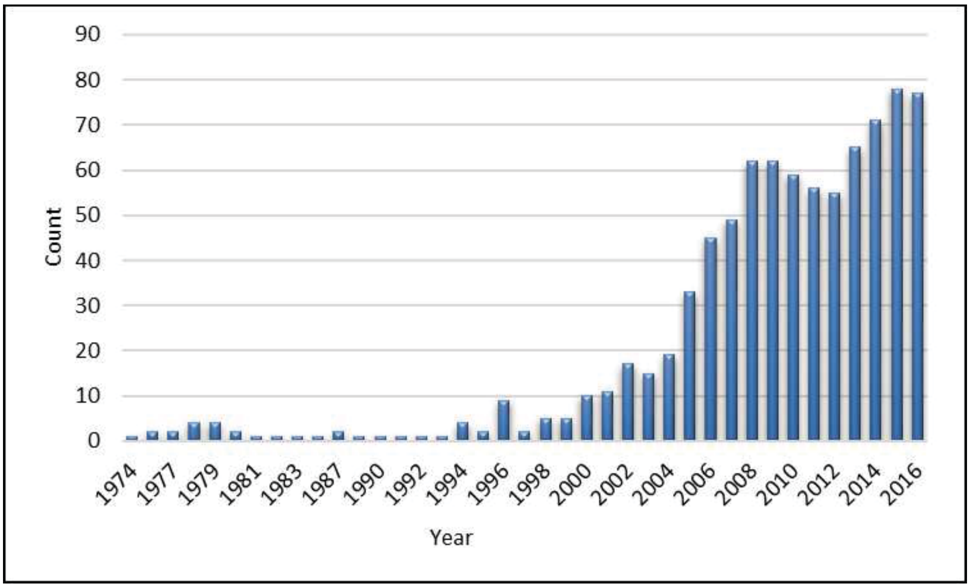


Figure 2. Citation of papers about Dermatology, published between 1974 and 2016

Subject Analysis of the articles in dermatology field

There were 2935 keywords used for the dermatology research field. Among these, only 34 keywords appeared more than ten times. The Top 60 high-frequency keywords were selected.

In the following, the map between these keywords has been considered. These are several indicators for co- word analysis that can be used in scientific maps. According to Freeman, measures of centrality are three categories: closeness, degree and betweeness (Freeman, 2004). Figure 3 shows co-word network According to Betweeness Centrality. Betweenness centrality identifies an entity's position within a network in terms of its ability to make connections to other pairs or groups in a network. Un entity with a high betweenness centrality generally: Holds a favored or powerful position in the network, represents a single point of failure, Has a greater amount of influence over what happens in a network (Scott, 2011).

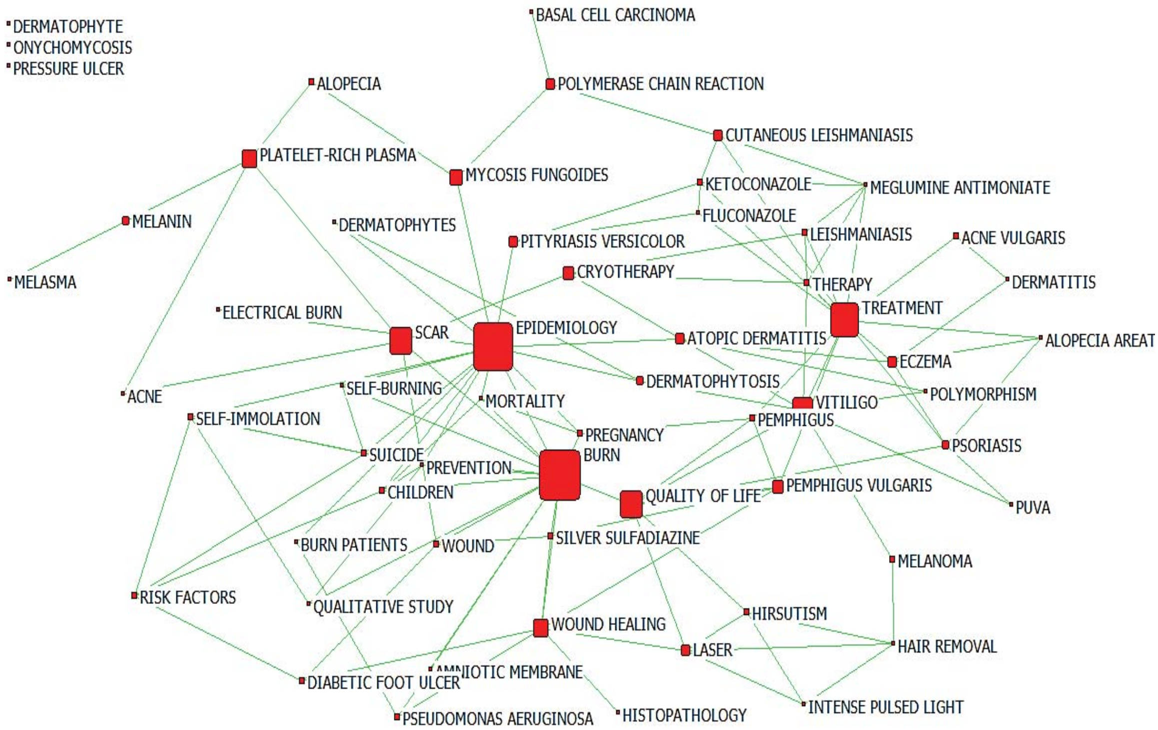


Figure 3. Co-occurrence network of the top 60 keywords from published articles of the dermatology field.

In figure 3, each square (node) indicates the keywords, and the lines represent the relationship between them. The top 10 Keywords with high betweeness centrality were as follows: Burn, Epidemiology, Treatment, Scar, Quality of Life, Vitiligo, Wound Healing, Platelet-Rich Plasma, Mycosis fungoides, and Cryotherapy. In Table 2, 10 superior figures in terms of centrality measures and frequency of keywords have been listed.

Table 2. The keywords with the highest frequency and Betweeness centrality

NO.	*Keywords*	*Frequency of Keywords*	*Keywords*	*Betweeness Centrality*
1	BURN	62	BURN	27.042
2	EPIDEMIOLOGY	25	EPIDEMIOLOGY	25.096
3	TREATMENT	20	TREATMENT	17.489
4	SCAR	6	SCAR	13.335
5	QUALITY OF LIFE	8	QUALITY OF LIFE	13.305
6	VITILIGO	22	VITILIGO	12.288
7	WOUND HEALING	24	WOUND HEALING	8.660
8	PLATELET-RICH PLASMA	4	PLATELET-RICH PLASMA	7.478
9	MYCOSIS FUNGOIDES	4	MYCOSIS FUNGOIDES	6.695
10	CRYOTHERAPY	7	CRYOTHERAPY	5.450

Journals, and Author Profiles of Publications

Table 3 shows the 10 most active authors and the authors with highest number of citation.
Dr Firooz from Tehran University of Medical Scienc (TUMS) with 45 articles had the largest number of publications in the field of dermatology research, followed by M Daneshpazhooh from TUMS (n = 41), C Chams-Davatchi from TUMS (n= 36), and Y Dowlati from TUMS (n = 31). We also quantified whether these authors published as either the cited author. Clearly, the cited author, A. Firooz still ranked first (n = 676) in the dermatology field. Considering the fact that forth author (Dowlati) is the second highest in terms of the citation (n= 591). Similarly, the information in the table shows that the ranking of authors varies in terms of the number of records and citations.

NO.	Athure	Records	Citations	NO	Athure	Records	Citations	H-index
1	Firooz A	45	676	1	Firooz A	45	676	16
2	Dowlati Y	31	591	2	Daneshpazhooh M	41	460	12
3	Chams-Davatchi C	36	546	3	Chams-Davatchi C	36	546	14
4	Daneshpazhooh M	41	460	4	Dowlati Y	31	591	12
5	Hallaji Z	17	330	5	Namazi Mr	23	225	8
6	Balighi K	19	312	6	Balighi K	19	312	9
7	Akhyani M	18	259	7	Akhyani M	18	259	9
8	Namazi Mr	23	225	8	Hallaji Z	17	330	10
9	Radmanesh M	16	140	9	Mansouri P	17	90	6
10	Mansouri P	17	90	10	Radmanesh M	16	140	8

Table 3. Most active authors and the authors with highest number of citation.

The h-index simultaneously measures the quality and quantity of the entire scientific output of a researcher, and it is one of the most commonly used indicators of research quality (Zhang et al., 2016). Consistently, we could conclude that A. Firooz, who had the highest h-index (n=16), could be considered authority in the dermatology field. But C. Chams-Davatchi is third in terms of the number of articles, but in H- index (n=14) he is the second one. Also Y. Dowlati is forth in terms of the number of articles but in H- index (n=12) is the third one. It should be underlined that all researchers come from TUMS. Although citations do not reflect the quality of a paper comprehensively, in a sense, they reflect a difference in scientific output.

Distribution of journals

All papers were published in 62 journals. The top five journals had more than 40 articles (Table 4). Approximately 38% of the WoS papers were published in these most productive top five journals (International Journal of Dermatology, Burns, Journal of the European Academy of

Dermatology, Mycoses, Clinical and Experimental Dermatology) which are considered the core journals of dermatology research under the Bradford Law (Zhang et al., 2016).

Table 4. Journals with highest number of dermatology related papers

NO.	*Journal*	*Country*	*Records*	*Citations*	*C/A*	*IF*
1	INTERNATIONAL JOURNAL OF DERMATOLOGY	USA	130	1419(2)	10.91	2.056
2	BURNS	UK	120	1467(1)	12.22	1.56
3	JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY AND VENEREOLOGY	USA	52	579(3)	11.13	3.528
4	MYCOSES	USA	50	447(4)	8.94	2.252
5	CLINICAL AND EXPERIMENTAL DERMATOLOGY	USA	41	448(5)	10.92	1.589
6	JOURNAL OF DERMATOLOGICAL TREATMENT	UK	38	300(6)	7.89	1.89
7	WOUNDS- A COMPENDIUM OF CLINICAL RESEARCH AND PRACTICE	USA	23	26(9)	1.13	1.948
8	INDIAN JOURNAL OF DERMATOLOGY	India	23	16(10)	0.69	0.99
9	INDIAN JOURNAL OF DERMATOLOGY VENEREOLOGY LEPROLOGY RESEARCH AND PRACTICE	India	22	184(7)	8.36	0.97
10	PEDIATRIC DERMATOLOGY	UK	21	139(8)	6.61	1.06

The journal that published the largest share of articles was International Journal of Dermatology (n = 130). In fact, according to the citations and C/A ratio, we can estimate that the two journals, Journal of the European Academy of Dermatology and Venereology, And Burns have the greatest influence on the field of dermatology. Moreover, although journals like, Clinical and Experimental Dermatology only published a few articles compare with the first journal, they received a high number of citations. There is no doubt that the ratio of C/A is closely linked to the quality of the articles.

Institute performances

Our results show that 533 institutes 837 papers published between 1974 and 2016. The Tehran University Med Sci, Shiraz Univ Med Sci, Isfahan Univ Med Sci, Iran Univ Med Sci were the top four most productive research institutes (table 5). Approximately 31.89 of papers with 2420 citations have been published by Tehran University of Medical Science.

Table 5. Institutions with Highest Number of Papers

NO.	*Organization*	*Count*	*Citations*
1	TEHRAN UNIV MED SCI	267	2420(1)
2	SHIRAZ UNIV MED SCI	83	751(4)
3	ISFAHAN UNIV MED SCI	60	855(2)
4	IRAN UNIV MED SCI	60	826(3)
5	SHAHID BEHESHTI UNIV MED SCI	59	197(7)
6	MASHHAD UNIV MED SCI	41	194(8)
7	UNIV TEHRAN	38	529(5)
8	ISLAMIC AZAD UNIV	34	185(9)
9	TABRIZ UNIV MED SCI	29	177(10)
10	RAZI HOSP	27	234(6)

Discussion

In this study, we have provided a supplemental evaluation of the status of dermatology in Iran. The objective of the present study was to perform a scientometrics analysis of all dermatology publications from Iranian researchers indexed in the Web of Science.
In scientometrics, quantitative statistical methods are employed to identify criteria which contribute to the growth and expansion of sciences throughout human communities. Scientometrics is a part of science sociology which is used to make scientific policies and includes quantitative studies in scientific activities and publications in that scientific area (Hood and Concepcins, 2001).

This analysis determined the current state of research and trends in studies about dermatology between 1974 and 2016. In his 1963 book "Little Science, Big Science", which is a fundamental work in scientometrics, Price argues that the number of scientific articles are doubled every fifteen years. Such a growth rate cannot be attributed on a single factor only and it can be concluded that such growth is the essence of science (Price, 1963). Evaluating the results highlights that the scientific productions in dermatologic area have shown a rising trend. The results of this research are in agreement with those yielded by Yao et al., (2014), Ramin et al., (2015), Yi et al., (2016), etc.

According to bibliometric principles, if a paper is cited more times than others, its quality is considered to be higher. In other words, the number of citations is indicative of the power and authority in the field of interest. In our study, there was no significant correlation between the JIFs and the citation frequency of articles. This can result from several factors; for example, journals with advance online publication had higher impact factors than journals without advance online publication. Thus, factors other than the quality of papers may affect the citation frequency of a paper (Ramin et al., 2015).

Based on our structural subject statistic results, the published articles mainly focused on Burn, Epidemiology and Treatment. Also Most of the highly cited papers was in burn study. The statistics indicate that burns in Iran are higher than in many other countries. So that burns in Iran are eight times the global average. Every year, 200 to 210 thousand people die in the country (Abouie et al., 2017). The probably reason for further research on this topic. These results provided a current view on the research focuses of dermatology. More importantly, subject categories can represent a suitable guide for future research directions. Although dermatology diseases affecting the quality of life and emotional status of subjects, had largely neglected this subject. Also, the other diseases that are epidemiology in Iran, such as psoriasis, have been less investigated.

Almost all (99.5 %) of the literature in the field was in English. For better international communication, English is the first language of choice for many authors. Among the top 10 journals, five were from the USA. 8 Journals out of 10 journals were from developed countries. In the field of dermatology developed countries have had a great influence on high-level science and the development of technology. Furthermore, publications from these journals were of high quality. The analysis of journals in which papers about Dermatology were published could help scholars select the appropriate journal for paper submission, thereby increasing the chance of acceptance of these journals had high citation.

According to bibliometric principles, if a paper or author is cited more times than others, its quality is considered to be higher. In other words, the number of citations is indicative of the power and authority in the field of interest. So the authors like,Dowlati Y , Hallaji Z, Balighi K who were not active among the top authors but they published high-quality articles.

Our results highlighted that, among the institutions Tehran Univ Med Sci (TUMS) and the ac tive authors " Firooz A " , "Dowlati Y ", "Chams-Davatchi C" have the most influence in this field. TUMS and its faculty members have a significant role in publishing scientific papers in the field of reproductive medicine. TUMS is the oldest and most well-known medical center in Iran, nationally as well as internationally that was established as a part of University of Tehran in 1934. TUMS, as one of the country's top research universities, accepts applications from the most qualified students. It also has the largest schools of medicine, over a hundred specialized research center and 10 teaching hospitals. Productivity of a university are mostly related to the authors who are affiliated with that university (Bazm et al. 2016). In other word, institutional centrality within high productions emerges and develops as authors affiliated with that institutions.

In conclusion, through performing the first scientometric survey on dermatology research, we analyzed the characteristics of papers and the trends in scientific production. Co-word analysis revealed outstanding topics of the field, which is useful for policy makers to learn about the research status and make appropriate decisions for the promotion of scholarly products. This study did not describe the features of journals, institutions or authors and do not compare the situation of Iran with other countries.

Limitation

We must take this limitation into account that studied data are from Web of Science. Searching other databases such as Scopus or PubMed database may lead to different results. However, In addition, performing similar researches using other scientometric techniques such as studying co-authorship and Co-occurrence network for countries, Co-occurrence network for Institutionsand other citation analysis can act as a complementary to this research.

Acknowledgement

This study was sponsored by Yazd University, Iran.

References

Abouie A, Salamati P, Hafezi-Nejad N, Rahimi-Movaghar A, Saadat S, Amin-Esmaeili M, et al (2017). Incidence and cost of non-fatal burns in Iran: a nationwide population-based study. Int J Inj Contr Saf Promot 10, 1-8.

Bakker P, Woerdenbag H, Gooskens V, Naafs B, Kaaij RVD, Wieringa N (2012). Dermatological preparations for the tropics. A formulary of dermatological preparations and background information on choices, production and dispensing. University of Groningen, the Netherlands.

Bazm S, Kalantar SM, Mirzaei M (2016). Bibliometric mapping and clustering analysis of Iranian papers on reproductive medicine in Scopus database (2010-2014). Int J Reprod Biomed 14, 371-382.

Bickers DR, Lim HW, Margolis D, et al (2006). The burden of skin diseases: 2004 a joint project of The American Academy of Dermatology Association and the Society for Investigative Dermatology. J Am Acad Dermatol 490-500.

Burns T, Breathnach S, Cox N, Griffiths C, eds. (2010). Rook's textbook of dermatology. 8th ed. Oxford: Wiley-Blackwell.

Edison K, Bruce B (2017). Commentary: Burden of skin disease report. J Am Acad Dermatol (JAAD) 76: 973 – 974.

Figueroa JI, Fuller LC, Abraha A, Hay RJ (1998) Dermatology in the southwestern Ethiopia: rational For a community approach. Int J Dermatol. 37, 752-758.

Freeman L. C (2004). The Development of Social Network Analysis: A Study in the Sociology of Science. North Charleston, SC: BookSurge.

Garfield E (1964) ''Science Citation Index''—a new dimension in indexing. Science 144, 649–654.

Hay R, Bendeck SE, Chen S, et al (2006) Skin diseases. In: Jamison DT, Breman JG, Measham AR, Alley ne G, Claeson M, Evans DB, et al, eds. Disease control priorities in developing countries. 2nd Ed. New York. Oxford University Press 707-722.

Hendrix D (2008) An analysis of bibliometric indicators, National Institutes of Health funding, and faculty size at Association of American Medical Colleges medical schools, 1997-2007. J Med Libr Assoc. 96, 324-334.

Hendrix D (2010). Tenure metrics: bibliometric education and ser-vices for academic faculty. Med Ref Serv Q 29, 183-189.

Hood W, Concepcins W (2001). The Literature of Bibliometrics, Scientometrics and Informetrics. Scientometrics 52, 291-314.

Johnson ML (2004). Defining the burden of skin disease in the United States--a historical DR Shagalov, GM Ferzli, T Wildman (2016) Perspective. J Investig Dermatol Symp Proc (DNLM) 9, 108-110.

Khatami A, Zartab H (2011) Dermatology: A public health perspective. Jdc 2, 245-253.

Price, D. J. D (1963). Little science, big science. New York: Columbia University Press 314-316.

Ramin S, Gharebaghi R, Heidary F (2015). Scientometric analysis and mapping of scientific articles on diabetic retinopathy. Med Hypothesis Discov Innov Ophthalmol 4, 81-100.

Random H (1997) Random House Webster's Unabridged Dictionary. 2th ed. Random House, New York.

Rosenbaum BE, Klein R, Hagan PG, et al (2017). Dermatology in Ghana: a retrospective review of skin disease at the Korle Bu Teaching Hospital Dermatology Clinic. Pan Afr Med J 26, 125.

Scott J (2011). Social network analysis: developments, advances, and prospects. Soc Netw Anal Min 1, 21-26.

Storan ER, Irvine AD (2017). Access to Genetic Diagnostics for Genodermatoses: Who Should Get Tested? Why? Who Pays? Pediatr Dermatol 34, 105-108.

Wolkenstein P, Grob JJ, Bastuji-Garin S, et al (2003). Societe Francaise de Dermotologie. French people and skin diseases: results of a survey using a representative sample. Arch Dermatol 139, 1614-1619.

Yao et al (2014). Scientometric trends and knowledge maps of global health systems research. Health Res Policy Syst. 12, 26.

Yi F, Yang P, Sheng H (2016). Tracing the scientific outputs in the field of Ebola research based on publications in the Web of Science. BMC research notes 9, 221.

Zhang Y, Xiao F, Lu S, Song J, Zhang C, Li J, et al (2016). Research trends and perspectives of male infertility: a bibliometric analysis of 20 years of scientific literature. Andrology 4,990-1001.

Article DOI: 10.22032/dbt.39342 Proceedings DOI: 10.22032/dbt.39296

Cooperative Network Analysis of Patent Holders in the Field of OLED Technology

[1]Jiang Chunlin and Jia Longchuan[2]

[1] *chunlinj7873@dlut.edu.cn*

Faculty of Humanities and Social Sciences and WISE Lab, Dalian University of Technology, Dalian 116024, China

[2] *jialongchuan@outlook.com*

Faculty of Humanities and Social Sciences and WISE Lab, Dalian University of Technology, Dalian 116024, China

Abstract

With the continuous development of science and technology, the number of patents continues to increase. At the same time, patent cooperation is more normal. It is particularly important to analyze the cooperation relationship among patent holders. The application of social network analysis methods solves this problem. OLED tends to gradually replace LCD. South Korea's Samsung and LG hold the majority of patents in the OLED field. How to break through has become a problem faced by Chinese companies. This paper uses the degree centrality, betweenness centrality and closeness centrality in the social network analysis method, and uses the data visualization tool Ucinet to systematically analyze the OLED technology patents from the Derwent Innovation Index. The results show that there is a clear trend of cooperation among patent holders in the OLED technology field. China's OLED enterprises should speed up the industrial chain layout, increase relevant R&D investment, and improve the R&D intensity of core technologies.

Introduction

The rapid development of the information technology industry has brought great convenience to people's lives, among which the display industry is an important pillar of the information technology industry, showing that the industry has the advantages of high added value and wide application, but also has a strong pulling effect on other industries（Wang et al. 2018）. Display technology involves all aspects of production and life, including business, communication, computer, industry and military fields, etc. Therefore, the development of display technology is of great significance to economic development. In the short 30 years of the development of display technology, it has experienced the technological change of CRT - LCD - OLED. In recent years, organic light emitting diodes (OLED) have replaced LCD gradually.

Compared with LED, OLED has many advantages: self - emission, wide viewing angle, low energy consumption, flexibility, etc. Based on the superior performance of these OLED's, the world's major display manufacturers are stepping up their industrial layout. At present, OLED technology in the world is almost monopolized by South Korea's Samsung and LG. Samsung mainly develops small screens for mobile phones, while LG mainly focuses on large screen displays. In addition, Sony in Japan and Innolux Display and AU Optronics in Taiwan are also speeding up the research and development of OLED. The mainland of China started late in this field and is still in the period of industry introduction and technology growth. The industrial chain is mainly focused on panel manufacturing, and the upstream equipment and raw materials are mostly imported from Japan and South Korea. This link is relatively weak. However, there are also many relatively strong enterprises in the mainland OLED field, such as BOE, CSOT, Visionox and so on.

Patent as a technology carrier is an important part of scientific discovery. According to statistics, the patent literature contains more than 90 % of the world's scientific and technological information (Wen et al. 2012). With the patentee, IPC, inventor, etc. in patent information as nodes, and the relationship of reference and cooperation as links, knowledge flow network can be constructed, and knowledge diffusion, knowledge transfer, and knowledge overflow can be measured more accurately. Companies often use patent cooperation to promote the improvement of their own technology level, such as LG Philips LCD Co., Ltd., which established in cooperation with Philips and LG. Visionox 's long-term OLED technology cooperation agreement with General Display, etc. These examples show that cooperation in this field is becoming more and more universal.

Literature Review

Organic Lighting Emitting Display (OLED) refers to the technology that organic semiconductor materials emit light under the action of electric field. OLED is an all-solid-state structure that actively emits light without backlight, and is called " dreamlike display technology" by the industry. OLED is one of the most promising new display technologies and also a competitive hot spot in the international high-tech field (Zhang 2011). The phenomenon of organic electroluminescence was first discovered in the 1960s, but it could not be truly industrialized due to a series of technical bottlenecks. Ching W. Tang and Van S Lyke provided breakthrough progress for OLED development in 1987 (Bernard et al. 2006). In 1990, Cambridge University discovered organic electroluminescent display technology from polymers, which greatly promoted the rapid development of OLED technology. After more than 30 years of technological evolution, OLED technology has achieved many technological breakthroughs and gradually matured. This technology has been industrialized in the field of display and lighting, and has become the most potential and promising new display technology in the future instead of CRTS and LCDS (Burroughes et al.1990).

Scholars have made a detailed analysis of OLED technology patents, including the number of patents, patentees, IPC and the distribution of patents in the country. In terms of quantity, the total number of OLED patent applications has been slightly tightened from 2005 to 2010, but the overall trend is increasing. Research shows that the number of patents varies significantly from country to country. Duan Keyu (2013) discovered through searching the VEN patent database that American scholars studied OLED at the earliest time and the research level was at the world's leading level, but the patent number advantage was surpassed by South Korea, mainly focusing on the fields of electroluminescent materials, electroluminescent power sources, electroluminescent panels and so on. As for the patentee, the patentee of OLED mainly concentrated in East Asia, Europe and America, the early research of OLED technology mainly concentrated in Europe, and the industrialization of OLED technology mainly concentrated in East Asia. Zhao Xuewu et al. (2010) analyzed OLED patent data from three patent databases of SIPO, USPTO and EPO, and found that the main OLED patent applicants are flat panel display manufacturers or flat panel technology licensing companies, and there are more companies with high patent holdings in China, Japan and South Korea. As for IPC, OLED patents mainly involve fields such as H01L (semiconductor device), H05B (electric heating), G09G (static indicating device composed of a combination of several light sources), G02F (device or device for controlling the intensity, color, phase, polarization or direction of light), etc. The main technical fields involved by different patentees are different. Luo Jiaxiu et al. (2011) compared with the patent layout of four companies (Sony, Samsung, RiTdisplay and Visionox) in the United States, found that Sony mainly concentrated on control devices and circuits, Samsung and Visionox mainly concentrated on components such as electrodes, and RiTdisplay mainly concentrated on solid-state devices.

The basic patents of OLED are mainly owned by Kodak, Cambridge Display Technology (CDT) in Britain and Universal Display Company (UDC) in the United States. Among them, the basic patents of small molecule OLED are mainly owned by Kodak, and the basic patents of polymer OLED are mainly owned by CDT and Uniax in the United States (Luo et al. 2011). These companies naturally become the main members of OLED cooperation. Kodak is an OLED research company with the largest number of core patents. In 2009, Kodak adjusted its development direction and sold OLED technology to LG, but reserved the right to use the patent. At the same time, LG set up the company Global OLED Technology in the United States to expand its influence in the Americas. In order to realize the industrialization of polymer OLED technology, Britain's Cambridge Display Technology Company (CDT) actively participates in technology diffusion. In 2007, CDT was acquired by Sumitomo Chemical and its research strength was further integrated. This kind of cooperative relationship exists widely in OLED technology field, and the social network analysis method can show the complete cooperative network relationship.

At present, scholars mainly study OLED patents from the perspective of quantitative analysis, while few scholars use social network analysis to study OLED patents. With the help of this method, the cooperative relationship between patentees can be visually analyzed, and a leading enterprise of technology can be found. In addition to regular quantitative analysis, scholars also use data visualization software to make a network so as to observe the network structure more intuitively. Common social network data visualization software includes Netdraw, Pajek, Gephi, Citespace and so on. Wang Lijie (2016) takes the Institute of Polymer Optoelectronic Materials and Devices of South China University of Technology as the research object, selects inventor information in patent information as the index to construct inventor co-occurrence matrix and Jacard co-occurrence matrix, and constructs the network diagram through Ucinet, and systematically analyzes the network structure with three indexes of degree centrality, betweenness centrality and closeness centrality.

With the support of various indicators of social network analysis and various data visualization tools, this method has been applied to patent analysis in various fields. Wang Hailong et al. (2017) searched patents in the semiconductor field through USPTO, and constructed a network based on the citation relationship between patents. An evaluation system was set up through six indicators, namely, the degree of output, the degree of input, the degree of betweenness, the degree of closeness, the effective scale and the limitation. An empirical study was conducted on the identification of patent-based technologies in the semiconductor field. Gong Jintao et al. (2013) have conducted empirical research in this field through the patents of wind power generation technology in China's patent database. They have conducted network cohesion analysis, network connectivity analysis and network centrality analysis through different theories in social network analysis. They believe that the technical research and development directors and core researchers in this field can be identified through this complex network analysis. Breschi Stefano et al. (2009) systematically analyzed the contribution of the transfer of inventors' and inventors' networks to knowledge dissemination within enterprises and within cities or states using social network analysis methods based on the original data set of patent applications filed by U.S. inventors in the European patent office.

Data

The patent data in this paper comes from Derwent Innovations Index. The formulation of keywords and the retrieval strategy adopted the scheme reported by Alan L Porter in 2008, that is, the keywords are first extracted from the literature, then the keywords are revised and the retrieval formula is determined by consulting experts in the field, and the retrieval formula is continuously optimized and adjusted through the retrieval results (Carley et al. 2013). The retrieval period is from 1967 to 2018. As of the final retrieval date of this article, 23235 patent data have been retrieved, and 23197 patent data have been obtained after cleaning. Then, we use the patentee index in the database to analyze the social network.

Results and Analysis

Overall Network of Patentees

According to the statistics on the number of patents held by the patentee, the top 10 patent holders are intercepted as shown in Table 1. OLED technology patents are mainly concentrated in South Korea, Japan, mainland of China and Taiwan of China. Four of the top 10 patent holders are from Samsung Group, which shows Samsung's strong dominance in this field, and South Korea's LG Company also has a large number of patents in OLED field. Kodak owns 435 OLED patents in the United States. Kodak is an early developer of OLED display technology and holds most of the basic technology patents. China's BOE is active in OLED technology research and development, with the third largest number of patents, after Samsung Display Co., Ltd. and LG Display Co., Ltd. But at the same time, nearly half of the technology patents in the OLED field in mainland China are utility models. Because OLED is a new industry. Chinese enterprises can increase R & D investment and strengthen international cooperation to achieve overtaking in corners.

Table 1. Patentee and its possession patent statistics

Ranking	*Patentee*	*Quantity*
1	Samsung Display Co Ltd	3267
2	LG Display Co Ltd	2669
3	BOE Technology Group Co Ltd	1311
4	Samsung Electronics Co Ltd	1244
5	Samsung Mobile Display Co Ltd	1052
6	Samsung SDI Co Ltd	716
7	LG Philips LCD Co Ltd	589
8	Eastman Kodak Co	435
9	Au Optronics Corp	362
10	Shenzhen China Star Optoelectronics Tech	329

A single quantity ranking cannot show the relationship between patent holders. To this end, the patentee data is organized into a co-occurrence matrix through the Bibexcel, and then the

co-occurrence matrix is imported into Ucinet for mapping. The result is shown in Figure 1.
In the network of OLED technology patentees, large networks coexist with small groups, which have a network structure of less than ten people. Large networks have a high density, and different nodes are closely connected. There are many important nodes that play a bridging role. These nodes represent enterprises or individuals holding important OLED patents. This reveals the current research and development status of OLED technology. Samsung, LG, Philips and other companies in Europe have mastered many basic patents and become major networks centering on them. At the same time, there are many small research and development teams in colleges and universities, which do not cooperate with large enterprises, but also have strong research and development capabilities.

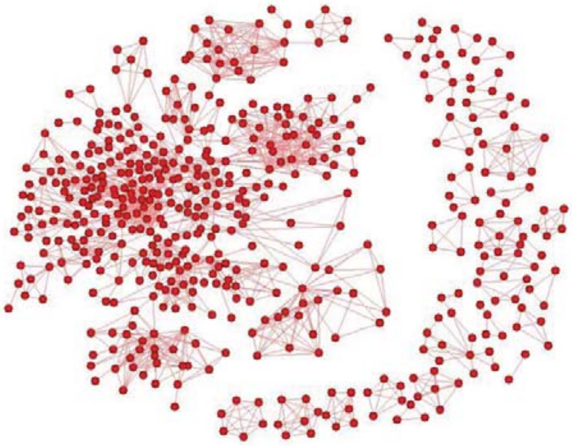


Figure 1. Co-occurrence network of Patentee

Right holder's communication ability
Centrality is an important index in the analysis of social networks. Through the evaluation of the position of the central performance node in the network, which patentees play a key role in the patent cooperation network can be evaluated. Centrality analysis includes main three parts: degree centrality, betweenness centrality and closeness centrality. Through the comparison of three types of centrality, the communication ability, control ability and independent ability of enterprises in OLED technology field can be clearly compared.
Degree centrality is the most direct measure to describe the node center in network analysis. The greater the degree centrality of a node, the higher the degree centrality of the node, and the more important the node is in the network. The greater the degree centrality of a patentee, the more central it is in the network. The degree centrality of patentees is calculated by using Ucinet and the list of the top 5 patentees is shown in Table 2. The top six patentees are subsidiaries of Samsung Group: Samsung Mobile Display Co., Ltd., Samsung Display Co., Ltd., Samsung SDI Co., Ltd., Samsung DenKan Co., Ltd., Samsung Electronics Co., Ltd. and Samsung Mobile Display Co., Ltd. Secondly, Lee's independent patentee is also from Samsung Group, which shows Samsung Group's strong monopoly power in OLED technology field, and at the same time, the group's internal ties are very close, the subsidiaries cooperate with each other, and the research and development intensity is greatly enhanced.

Table 1. Degree of patentee ranking

Ranking	*Patentee*	*Degree*
1	Samsung Mobile Display Co Ltd	1557.000
2	Samsung Display Co Ltd	1520.000
3	Samsung SDI Co Ltd	832.000
4	Samsung DenKan Kk	620.000
5	Samsung Electronics Co Ltd	406.000

Use the Netdraw to draw out the nodes with greater centrality, as shown in fig. 2. It can be seen that they are all teams with very strong scientific research strength. Apart from South Korea's Samsung scientific research team, the patentees such as Princeton University and the University of Southern California in the United States also have very close ties. There are also teams from Siemens, Osram, Merck and Philips of the Netherlands, which also play an important role in OLED patentee networks. These enterprises gradually formed patent alliances through patent cooperation, which promoted the development of their own technical level.

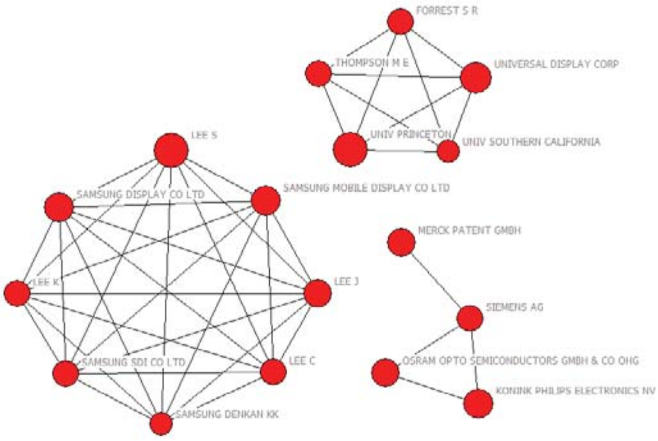


Figure 2. Network of patentee's degree

Rights holder's control ability

If a node is between other node pairs in the network and is on the only way to communicate with each other between node pairs, the node must have a very important position in the network. Therefore, mediation centrality is interpreted based on the node's control over communication. Through mediation centrality analysis, we can find the network's community bridge or communication bottleneck or cross - border. Sorting out the information about the patentee's betweenness centrality is shown in Table 3. The higher the centrality of intermediaries, the greater the patentee's control over the network. The most central intermediary is Konink Philips, which mainly produces large-sized OLED screens, while Samsung Electronics, which is in second place, mainly sells small-sized OLED screens. The two companies hold important technologies in OLED technology. In recent years, Philips of the Netherlands, in order to consolidate its position in OLED large screen displays, has entered into various cooperation with LG of South Korea, which is also in the leading position in technology, and has jointly established LG Philips Display Co., Ltd., which will cause more restrictions on the development of OLED large screen displays in China.

Table4. Betweenness of patentee ranking

Ranking	*Patentee*	*Betweenness*
1	Konink Philips Electronics Nv	23301.998
2	Samsung Electronics Co Ltd	19525.879
3	Eastman Kodak Co	14888.331
4	NXP Bv	10590.000
5	Semiconductor Energy Lab	9631.417

As shown in Figure 3, the patentee's network is drawn with the betweenness as an index. The node size represents the patentee's betweenness centrality. Samsung Electronics, Kodak and Philips play a very important control role in the network. Merck, Siemens, and other 12 companies have formed a cooperation circle structure, effectively promoting OLED technology research and development, but this has also formed a barrier to other enterprises. Within the circle structure, there are many patent holders in Samsung, including various subsidiaries and related independent patent holders. Samsung Group controls many related technical resources. Chinese OLED enterprises should deepen their cooperation with related companies in the network, strengthen the distribution of foreign patents, strive to obtain authorization for core patents or sign licensing agreements, and break through the intellectual property barriers and restrictions in international trade.

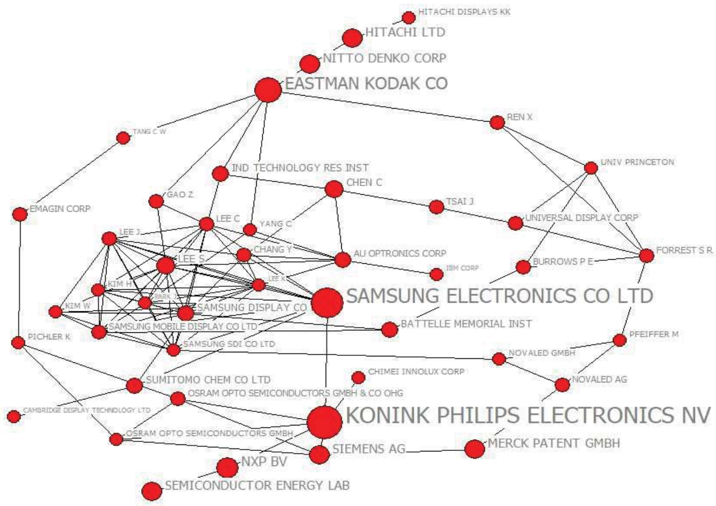


Figure 4. The network of patentee's betweenness

Independent Capacity

Closeness centrality: Closeness centrality describes the ability of actors in a network not to be " controlled" by others. It refers to the extent to which most direct paths connecting one node to all other nodes in the network are short (rather than long). The closer the center is, the smaller the node is in the core position in the network. The smaller the betweenness centrality of a patentee, the more information it can obtain. Sort out information about patentee's proximity to centrality as shown in Table 4. The top five places closeness centrality in the table belong to Samsung Group, indicating the close degree of cooperation within Samsung Group. The table shows that 11 of the top 20 patents close to centrality are individuals. It can be seen that the strength of the individual team cannot be ignored in OLED research and development. Chinese enterprises can introduce these talents or cooperate with independent teams when developing OLED. Enterprises and individuals from other countries have long-term OLED patent layout in China, becoming the patentee of important technologies in

China, and some technologies are concentrated in blank areas of OLED technology in China, which severely restricts the research and development of OLED technology in China.

Table 4. Closeness of patentee ranking

Ranking	*Patentee*	*Closeness*
1	Samsung Electronics Co Ltd	3145.000
2	Lee S	3194.000
3	Lee C	3208.000
4	Samsung Display Co Ltd	3209.000
5	Samsung Mobile Display Co Ltd	3237.000

Community Analysis

G-N algorithm is a splitting algorithm for subgroup discovery proposed by Girvan and Newman (2002). According to the description of the community, the internal nodes of the community are densely connected and the connections between the communities are relatively sparse. A few connections between subgroups will become the only way for communication traffic between subgroups. Considering some form of communication in the network and finding the edge with the highest traffic, removing the edge will get the most natural segmentation of the network. Therefore, Girvan and Newman introduced edge betweenness centrality to measure network traffic, and proposed a subgroup discovery algorithm based on edge betweenness, called G-N algorithm for short (Deng et al. 2012).

Through this clustering analysis, the main patent holder network in OLED field is divided into four communities as shown in fig. 5, in which nodes of different colors and shapes represent different communities. The blue community is the largest community in OLED, with Samsung Group as the main group, including Samsung Electronics, Samsung Display, Samsung Mobile Display and so on, as well as LG, Sumitomo and AU Optronics. The cooperation between AU Optronics and Samsung began in 2006, when the two sides signed a patent cross license. The green community is mainly European enterprises, and the representative enterprises are Konink Philips Electronics, NXP, Siemens, Osram and so on. On January 28, 2007, OSRAM and Philips reached a cross-licensing agreement for LED and OLED, which covers patents held by Philips and its U.S. subsidiary Lumileds and OSRAM and its subsidiary OSRAM Opto Semiconductor S. (Luo et al. 2011). Red community are mainly American patent holders, involving scholars, enterprises and universities, among which Princeton University, the University of Southern California, the University of Michigan, Motorola and UDC are allies. UDC cooperated with Princeton University, the University of Southern California and the University of Michigan in research and obtained Motorola's sole license to transfer. Among them, Kodak is the key node, because it holds many core patents, so it is very strict in patent licensing. At the same time, the granted enterprise must license all its own patents to Kodak free of charge. The yellow community is Chinese Taiwan enterprises and researchers. The core node is Professor Chen Chin Hsin, Deputy Secretary General of Taiwan Industrial Research Institute and China OLED Industry Alliance.

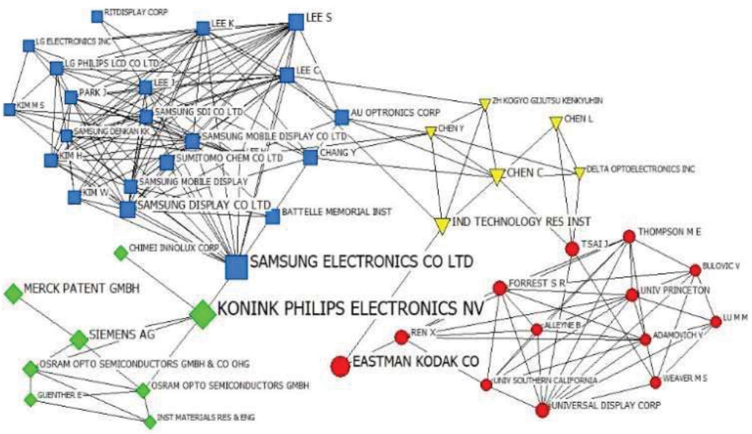


Figure 5. Patentee clustering network

Conclusion and discussion

It can be seen from the above that there is extensive patent cooperation in OLED technology, including patent alliance, patent cross licensing and other forms. South Korea holds the largest number of patents in OLED field, and each subsidiary of Samsung Group and its independent patentees within the group have close cooperation. Although the amount of patent disclosure in Europe and America is not as large as that in Asian countries, through network analysis, it is found that many enterprises in Germany, the Netherlands and the United States are still in a very important position in OLED field, such as Kodak, Philips and other enterprises, which is also proving the importance of basic technology patents controlled by European and American patent holders.

Although mainland China ranks very high in OLED patents, there are no mainland Chinese enterprises in the three central rankings in network analysis. First of all, mainland OLED companies started late, starting from the end of the value chain, mostly in a generation-by-generation processing relationship with the world's major OLED companies, and have not yet integrated into the world's major patent cooperation networks. Secondly, the major OLED enterprises have formed patent alliances, which are of great help to the technological upgrading of their internal members, but at the same time, they have caused great obstacles to the development of enterprises outside the alliance.

In order to realize the substantial breakthrough of OLED technology, Chinese enterprises should increase their scientific research intensity, move upstream into the global value chain, invest in new industries, and quickly form the results of scientific and technological transformation, and form a patent layout as soon as Possible. OLED enterprises can strengthen the cooperation with domestic and foreign display terminal products enterprises, as soon as possible to build a perfect alliance to expand the upstream and downstream Layout. Speeding up the establishment of industry standard system dominated by Chinese enterprises, such as Chinese enterprises can seize this opportunity, can significantly enhance the competitiveness of Chinese enterprises, while effectively weakening the foreign companies to China's OLED industry monopoly. In the OLED patent cooperation network, the number of enterprises in mainland China is relatively low, the relevant enterprises can take appropriate cooperative authorization mode, in order to seek their own development, fill the gaps in their technical fields, and promote industrial transformation and upgrading. The establishment of a patent cooperation network suitable for the growth of enterprises can help enterprises realize the complementarity of resources and reduce the risk challenges in the process of Innovation. At the same time, the government can provide policy concessions to promote international patent cooperation as a model of knowledge flow, in the process, enterprises should enhance the absorption of knowledge, so as to enhance their innovative ability.

References

Bernard Geffroy, Philippe Roy and Christophe Prat (2006). Organic light-emitting diode (OLED) technology: materials, devices and display technologies. *Polymer international*, 55:572-582.

Burroughes J H, Bradley D, Brown A R, et al (1990). Light-emitting diodes based on conjugated polymers. *Nature*, 348(6299):352-352.

Carley, K. M., Pfeffer J, Reminga J, Et al. (2013).Ora User's Guide 2013. *Carnegie-Mellon Univ Pittsburgh Pa Inst of Software Research Internet.*

Deng, Z. L., Gan, W. Y. (2012).Methods for discovering community structures in complex networks. *Computer Science*, 39 (S1): 103 - 108.

Duan, K. Y. (2013).Patent Analysis of OLED Display Technology. *Electronic Components and Materials*, 32 (8): 68 - 71.

Girvan M, Newman M E J (2002).Community structure in social and biological networks. *PNAS*, (99):7821-7826.

Gong, J. T., Wei, X. F. (2013). Research on patent and network based on social network analysis —A case study on the patent of wind power technology in China. *Journal of Intelligence*, (11): 38-42.

Luo, J. X., Fan, B. (2011).Analysis of Patent Licensing and Litigation of OLED Company. *China Integrated Circuit*, 20 (8): 82 - 87.

Luo, J. X., Si, Y.F., Yang, F. (2011).U.S. Patent Analysis of OLED Key Companies. *China Integrated Circuit*, 20 (5): 87 - 93.

Stefano Breschi, Francesco Lissoni(2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*,(9):439-468.

Wang, H. J., Chen, J. (2018).Strategies for China's OLED industry innovation and development under the global value chain. *Technology Economics*, 37 (6): 40 - 47.

Wang, H. L., He, F. Q, Ding K. (2017). Identification of Patented Basic Technologies Based on Social Network Analysis - Taking Semiconductor Industry as an Example. *Journal of Intelligence*, (4):78 - 84.

Wang, L.J. (2016). Patent Inventor Team and Core Inventor Analysis Based on UCINET Social Network Analysis - Taking the Research Institute of Polymer Optoelectronic Materials and Devices of South China University of Technology as an Example. *Journal of Library and Information Sciences in Agriculture*, (10):72 - 76.

Wen, T. X., Yang, Z., Liu, X. (2012). Hunan Patent Strategy Research Based on Patent Measurement Analysis. *Information studies: Theory & Application*, 35 (1): 58 - 64.

Zhang, D. Q. (2011). Opportunities and challenges for OLED technology and industry development. *Advanced Materials Industry*,(1): 12 - 14.

Zhao, X. W., Li, W. (2010).Analysis of OLED Patent Technology Development Trend. *Electronic Intellectual Property*, (3): 84 - 90.

Proceedings of 14th International Conference on Webometrics, Informetrics and Scientometrics & 19th COLLNET Meeting 2018
>> Collaboration - Impact on Productivity and Innovation <<
Article DOI: 10.22032/dbt.39352 Proceedings DOI: 10.22032/dbt.39296

Evaluation of Regional Patent Innovation efficiency and its Spatial Distribution from the Perspective of Spatial Spillover[1]

[1]Liangyu Li, Chunjuan Luan[2], Yu Wang[3]

[1] *selinamoyan@126.com*
School of Intellectual Property, Dalian University of Technology, Dalian, Liaoning, P.R.China

[2] *julielcj@163.com*
School of Intellectual Property, Dalian University of Technology, Dalian, Liaoning, P.R.China

[3] *fayebxch@163.com*
School of Business, Dalian University of Technology, Dalian, Liaoning, P.R.China

Abstract

There is considerable interest among policy-makers and academics regarding other potential arrangements to encourage the innovative activities of firms. This study aims to investigate whether R&D investments promote or impede the enhancement of patent innovation efficiency in China, and also to reveal the spatial patterns of patent innovation and its regional interdependencies and evolution, as well as its role in determining the effects of local innovation in China. The spatial autoregressive model is used to examine the effects of patent innovation efficiency. Results show that geographical proximity matters in the interregional flow of knowledge and technology; moreover, innovation in a region depends on its own R&D efforts, its innovative tradition and its human capital endowments, and the regional innovative activities has demonstrated that policies enhancing regional R&D activities are probably to get a richer effectiveness on stimulating patent innovation.

Introduction

It is widely recognized that technological innovation is a source of economic growth. Consequently, there is considerable interest among policy-makers and academics regarding other potential arrangements to encourage or facilitate the innovative activity of firms. In this context, with a rapidly growing knowledge economy and increasing economic globalization, China has proposed the 13th Five-Year Plan (2016-2020) that stimulates the innovation of science and technology continuously, improves the ability of independent innovation, and also makes innovation a core position in the process of national economic development. Therefore, independent innovation seems to have become an inevitable means for many companies to achieve lasting survival and development and a lasting competitive edge.

Recent literature on the economics of innovation and technological process has a central issue that research and development (R&D), as a form of decision-making in enterprises, has endogenized the effects of technological innovations on economic growth. According to statistics, China has made huge R&D expenditures during the past decades; in 2017, the R&D expenditures reached approximately 1750 billion RMB (Chinese Yuan), with the eastern, central and western region each accounting for 69.95%, 17.58% and 12.47%. Besides R&D expenditures, patents may also be sensitive to its filing fees, which has received relative attention to the R&D input in China . As patents involve a lot of rich and timing information during innovative activities, which are widely regarded as indicators of innovative strength a sign of great development in China’s innovative capacity, patent statistics are usually used to identify and measure innovations. While R&D expenditures are widely used as a proxy for

[1]This article is supported by the Youth Fund Project of the Ministry of Education (NO. 18YJC790083) , the Fundamental Research Funds for the Central Universities(NO. DUT17RC(3)107), National Natural Science Foundation of China (No.71774020/71473028) and Youth Fund Project of the Ministry of Education (NO. 16YJC790105)

innovation input, patent statistics could be applied to measure the output. This measure is also more easily obtainable than other proxies for outputs, such as total factor productivity (TFP) (Nagaoka, Motohashi, & Goto, 2010) . In view of the huge differences and heterogenicity across regions and provinces in China, a hypothesis could be assumed that there exists regional differentiation in patent innovation efficiency, which has important implications for policy makers to formulate the patent-related policies . However, the important question of how patents reasonably impact the continual innovation remains unsettled. Based on this background, we investigate the spatial spillover effect of technological innovations on regional economic growth, and make a further study of how patents impact the innovation incentives, and to what extent the amount of R&D input stimulates the improvement of patent innovation.

In this paper, we explore the influential relationship between patent innovation system and economic growth, and also analyses the spatial patterns of patent innovation, regional interdependencies and evolution, as well as its determinants in regional innovation in China. To have a better understanding, we studied patents usually relates to the need to protect the firms' incentive to innovate and maintain the monopoly profits to get their accumulative competitive edge under the condition of avoiding the loss of social welfare and not having free access to the protected goods. Thus the patent innovation system is getting a diverging trend among numerous patent innovation regions; the local patent innovation region is not only relevant to itself, but also relevant to the other patent innovation regions' spatial innovation spillovers. In this context, this paper analyses the spatial patterns of innovation and its spatial distribution via patent innovation efficiency, and also investigate the regional interdependencies and its evolution, so as to improve the effectiveness of patent innovation policies.

The remainder of this article is organized as follows. The paper first discusses the background and necessity to study patent innovation and spatial spillovers ; The section 2 provides the specific theoretical linkage and influential relationship of patent innovation and economic growth, where it is believed the patent innovation, instead of imitation, is definitely significant in driving long-run economic growth, in terms of patent innovation system and market institutions. The methodology for the further study underpin the logic is introduced in section 3, describing the process of data collection and variables used. In section 4, this paper makes an estimation of regional patent innovation efficiency in China by means of DEA tool, develops a spatial model to get the regional spillovers among regions, and also discusses the results of comprehensive analysis. Finally in section 5, this paper discusses the implications of overall findings for managing patent innovation policies and regional development of regional patent innovation in modern China.

Theoretical background

Current literation has a lot of discussion about the relationship between innovation and economic growth. According to the US Advisory committee, innovation is defined as follows: 'The design, invention, development and/or implementation of new or altered products, services, processes, systems, organizational structures or business models for the purpose of creating new value for customers and financial returns for the firms . In this context, patents are usually linked to the firms' inner desire of making a strategy to protect the firms' incentive, so as to invent new functional products and bring up new value for customers, and maintain the monopoly profits to get their accumulative competitive edge under the condition of avoiding the loss of social welfare and not having free access to the protected goods. Thus the technological advancement, which is generated from the patent innovation, is becoming the critical driver of long-run economic growth. Some economists have found the regular rules between patent innovation and patent policies that the optimal patent policy equates the dynamic marginal benefit with the static marginal efficiency loss(Nordhaus, 1969) , while others find patents more effective in the high-tech area(Levin et al., 1987; Cohen et al., 2000) , particularly in the chemical and pharmaceutical areas where a precise chemical formula of a specific compound can be accurately described in a patent hence reducing the likelihood of dispute over property rights . In this context, China has proposed a the National Medium and Long Term Program for Science and Technology Development(2006–2020), putting its attention precisely on high-tech areas and probably supporting with favorable policies and financial incentives. Besides, according to Besen and Raskind's study, the appropriate period of patent protection is that allows the innovator to cover the risk-adjusted cost of innovative activity . Other scholars make a further discussion of the patent protection scope, and make a specific contrast under the condition of a closed economy and an open economy respectively; the results reveal that trade-offs are less clear and less it depends on the nature of the market , and a variety of innovation capacities will be demonstrated by their differences in skill endowments and technical knowledge .

Another body literature highlights the endogenous growth theory during the study of regional innovation capability, in which the investment in Research and Development (R&D) plays a significant role in the economic growth and sustainable innovation development of countries and regions. It not only emphasizes the effect of R&D effort and knowledge stock on innovation, but also recognizes that innovation depends to some degree on the level of a regions' technological capital and their absorptive capacities . A number of economists have taken R&D as a form of decision-making in enterprises and considered it as an effectiveness of technological innovations on economic growth. Hu and Jefferson (2009) estimate a patent production function for Chinese enterprises, finding significantly low patent-R&D elasticity and claim that foreign direct investment, institution change, and other factors are behind the patent surge . Other scholars also make studies into this area, taking adventure of R&D effects on innovation, and the results reveal that: both R&D and non-R&D innovation expenditures could positively promote the productivity, the point is to make the best practice to achieve an outward shift according to firms' production frontiers .

Current literature has provided us a guidance in studying the drivers of patent innovation and its relationship with economic growth. In this context, a production–innovation system is employed to investigate the patent innovation efficiency and make a further analysis of the linkage between patent innovation and production systems. Nevertheless, considering innovation capacity in a specific area, literatures also contribute that spillovers of knowledge and information from external sources may have an inevitable impact on innovation processes and economic growth. In this context, the spatial dimension has become a critical aspect in determining how those spillovers occur and how those spillovers get interaction with each other in the local innovation process . A number of economists have investigated the spatial spillover effect of technological innovations on regional economic growth (Eaton and Kortum, 1996; Moreno et al., 2005). Some empirical findings also indicate that knowledge and information spillovers are tending to shape as clusters in spatial proximity from their respective source. Thus it can be assumed that knowledge and information spillovers could make an advantage in shaping the regional conditions for innovation activities. In this context, the framework of geographical space and spillovers lead us to get a further exploration into the question of how such spillovers become effective and what are the primary means for their diffusion . Cooperative relationships between regional actors may be an important vehicle for such spillovers. Consequently, a number of literatures demonstrate that policy could contribute to a wider and faster diffusion of knowledge and information spillovers by actively stimulating cooperative relationships (cf. Jorde and Teece, 1990) . With relevance to the importance of space for the diffusion of knowledge and information, geographic proximity to innovation producers is likely to perform in two ways: in a close region, geographic proximity to other innovative regions seems to boost the local innovation; in a public region, geographic proximity to other innovative regions could almost promote knowledge and information spillovers across borders, and the importance of regional interaction for the flow of knowledge and information is positively enhanced. Consequently, we could not only study the R&D spillovers effects on innovation, but also discuss other key determinants spillovers and make proper innovation policy through speeding up the diffusion of knowledge and information . Our analysis will therefore concentrate on the role and characteristics of patent innovation and spatial spillovers among Chinese provincial regions that go beyond merely geographical aspects.

Research methods and data

A number of studies have investigated the innovation-productivity relationship with some empirical analysis reporting on the effectiveness of innovation on firms' productivity and efficiency, using the methodology of estimating Cobb–Douglas pro-duction functions. Given the aims of our analysis, a production–innovation system is employed to investigate the patent innovation efficiency and make a further analysis of the linkage between patent innovation and production systems.

Model used for the measurement of patent innovation efficiency

Data envelopment analysis (DEA) is a nonparametric method for the estimation of production frontiers to measure the productive efficiency of decision-making units, which aims to identify the most efficient units among a set of comparable entities. Basic DEA models include the Charnese Coopere Rhodes (CCR) model (Charnes et al., 1978) and the Bankere Charnese Cooper (BCC) model (Banker et al.,1984). These DEA models have been widely used to measure the technological and economic efficiency of units (Deilmann et al., 2016). Resources can be regarded as input variables together with capital and labor; patent applications and sales revenue of new products are chosen to approximate the innovative output potential in the region, because these two targets are almost probably relevant to the significant features of invention. Hence, we employ a traditional DEA model to evaluate the innovation performances of each provincial decision-making unit.

$$\min \theta$$
$$s.t.\begin{cases} \sum_{k=1}^{n} \lambda_k x_k + s^- = \theta X_t \\ \sum_{k=1}^{n} \lambda_k y_k - s^+ = Y_t \\ \sum_{k=1}^{n} \lambda_k = 1 \\ \lambda_k \geq 0, \quad k = 1,2,\cdots,n \\ s^+ \geq 0,\ s^- \geq 0 \end{cases} \tag{1}$$

Where the notation is as follows: n, number of *DMUs*; j, other *DMUs*; m, number of inputs consumed by DMU_j, x_{ik} (i = 1,2,···,m), amount of input i consumed by DMU_j; s, number of outputs produced by DMU_j; y_{rk} , amount of output r produced by DMU_j; s^-, vector of slack variables representing the amount of input i that, if reduced, shifts the projection of *DMU* from the weakly efficient frontier to the strongly efficient frontier; s^+, vector of slack variable representing the amount of output *r* that, if increased, shifts the projection of *DMU* from the weakly efficient frontier to the strongly efficient frontier; λ_j, linear weights assigned to every single DMU_j to form a linear combination. Note that when the efficiency θ=1 and the slacks summation is zero, the unit is considered strongly efficient. If θ=1 but the slacks summation is not zero, the unit is considered weakly efficient. For any inefficient *DMU*, it is possible to find a composite *DMU* (linear combination of units) that can reduce its input level maintaining the same output level. In this study, the directional slacks-based model of inefficiency is employed to calculate regional patent innovation efficiency in China. By constructing the innovation production frontier, the term patent innovation efficiency is integrated as incorporating the extent of resources (such as capital and labor), with reliance on patent application, and sales revenue of new products involved in the creation of provinces.

The econometric model used for spatial regression

To have a better understanding of the inner heterogenicity of regional patent innovation efficiency, an econometric model is adopted to detect the geographical characteristics and its spatial interdependencies in the distribution of innovation activities. The degree of spatial dependence can be explained by Moran's I-statistic which is defined as:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_{ij} w_{ij}(X_i - \overline{X})(X_j - \overline{X})}{\sum_i (X_i - \overline{X})^2} \quad if\ i \neq j \qquad (2)$$

where X_i and X_j are the observations for regions *i* and *j* of the variable of interest, $\overline{X}$ represents the regional average, *N* represents the number of observations and w_{ij} represents a row-standardized *W* matrix of weights. There is a hypothesis based on geographical contiguity that the proximity between regions could be defined as w_{ij}, the innovative contiguity between productive sectors; if the intensity of their innovative activities are highly bonded than the average, the innovative contiguity w_{ij} could almost equal to 1. Thus the bilateral weights w_{ij} could be used to approximate the intensity of regional interdependences of patent innovation efficiency in China.

In fact, the activities of patent innovation in a region does not only have a dependency on local capacity for innovation and local economy scale, but also have a tendency of being influenced by the nearby regions, which could be explained as regional spillovers. In this study, a spatial econometric methodology is provided to discuss this problem further, and the model could be obtained according to Eq.(3)-(4)

SLM model: $I_{it} = \alpha + \rho WI_{it} + \beta_1 RDL_{it} + \beta_2 RDE_{it} + \beta_3 S_{it} + \beta_4 TD_{it} + \varepsilon_{it}$ (3)

SEM model: $I_{it} = \alpha + \beta_1 RDL_{it} + \beta_2 RDE_{it} + \beta_3 S_{it} + \beta_4 TD_{it} + \varepsilon_{it},\ \varepsilon_{it} = \lambda W\varepsilon + u_{it}$ (4)

where *I* is innovative output, *W* is the weight matrix defining the proximity of regions and the regional spillover variable, *RDE* denotes local R&D investment efforts and *RDL* is human capital endowments. Additionally, the innovative performance could also be influenced by the regional structural characteristics, such as industrial structure and external trade, and by its innovative tradition. The regional spillover term is the weighted sum of innovation efforts in nearby regions. Thus the consideration of regional spillovers will promote a richer analysis from taking different sources of public innovative efforts and its economic implications into account.

Data and determinants

In the process of making comprehensive estimation of regional patent innovation efficiency, this paper uses 2005-2015 data from *China Statistical Yearbook* and *China Statistical Yearbook of Science and Technology*. By constructing the innovation production frontier, the

term patent innovation efficiency is integrated as incorporating the extent of resources including capital and labor devoted in R&D activities, with reliance on patent application, and sales revenue of new products involved in the creation of provinces.
In the process of spatial estimation, several determinants are used to detect the proximity of regions and the regional spillover, where I represents the patent innovation efficiency, *RDE* denotes local R&D investment efforts and *RDL* is human capital endowments; *S* and *TD* represent industrial structure and external trade respectively. The original innovation database is integrated by *China Statistical Yearbook* and *China Statistical Yearbook of Science and Technology* during 2005-2015, which is classified according to the major three traditional areas in China.

Results and analysis

By using patent innovation production function, this paper first make an comprehensive evaluation including the heterogeneity in innovative capacity of different regions in China and get the basic features of the variation tendency.

Table 1. Results of comprehensive evaluation of regional patent innovation efficiency during 2005-2015 in China

	2005	*2006*	*2007*	*2008*	*2009*	*2010*	*2011*	*2012*	*2013*	*2014*	*2015*
Beijing	0.983	0.944	0.975	0.988	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Tianjin	1.000	1.000	1.000	1.000	1.000	1.000	0.892	0.851	0.945	1.000	0.808
Hebei	0.648	0.694	0.654	0.647	0.663	0.643	0.646	0.701	0.703	0.796	0.772
Shanxi	0.763	0.588	0.694	0.555	0.463	0.530	0.413	0.423	0.467	0.426	0.478
Inner Mongolia	0.731	0.735	0.552	0.457	0.476	0.341	0.379	0.289	0.372	0.321	0.354
Liaoning	0.692	0.690	0.619	0.657	0.625	0.769	0.790	0.851	0.857	0.921	0.849
Jilin	0.810	0.772	0.875	0.846	0.730	0.776	0.790	0.757	0.785	0.827	0.836
Heilongjiang	0.775	0.831	0.737	0.769	0.808	0.881	0.849	0.879	0.912	0.886	0.855
Shanghai	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990	1.000	1.000	0.980
Jiangsu	0.766	0.791	0.733	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.913
Zhejiang	1.000	1.000	1.000	1.000	1.000	1.000	0.835	0.937	1.000	1.000	0.987
Anhui	0.843	1.000	0.903	0.950	0.804	0.822	1.000	0.859	0.855	0.831	0.821
Fujian	0.812	0.790	0.626	0.655	0.645	0.642	0.647	0.689	0.755	0.641	0.782
Jiangxi	0.455	0.422	0.366	0.432	0.423	0.405	0.437	0.500	0.640	0.710	0.653
Shandong	0.728	0.722	0.733	0.762	0.784	0.902	0.700	0.658	0.964	0.921	0.824
Henan	0.568	0.616	0.589	0.612	0.601	0.604	0.506	0.528	0.525	0.752	0.710
Hubei	0.975	0.952	0.964	0.997	0.677	1.000	0.970	0.842	0.748	0.799	0.744
Hunan	1.000	1.000	1.000	1.000	0.705	0.977	0.913	0.857	0.863	0.872	0.844
Guangdong	1.000	1.000	1.000	0.956	0.970	0.986	0.981	0.976	0.964	0.980	0.954
Guangxi	0.661	0.755	0.599	0.791	0.730	0.720	0.672	0.686	0.808	0.847	0.848
Hainan	1.000	0.831	1.000	1.000	0.975	0.814	0.647	0.681	0.722	0.774	0.728
Chongqing	0.996	0.982	0.891	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Sichuan	0.418	0.499	0.526	0.608	0.712	0.820	0.796	0.859	0.831	0.831	0.826
Guizhou	0.576	0.561	0.627	0.588	0.503	0.552	0.510	0.691	0.700	0.779	0.781
Yunnan	0.690	0.542	0.565	0.601	0.550	0.606	0.640	0.648	0.634	0.695	0.777
Shan'xi	0.629	0.651	0.724	0.808	0.885	0.919	0.966	1.000	1.000	1.000	0.873
Gansu	0.649	0.681	0.734	0.748	0.773	0.644	0.658	0.687	0.734	0.756	0.763
Qinghai	0.159	0.222	0.331	0.349	0.411	0.349	0.267	0.241	0.222	0.272	0.376
Ningxia	0.466	0.452	0.520	0.478	0.498	0.522	0.321	0.367	0.593	0.738	0.495
Xinjiang	0.773	0.834	0.996	0.819	0.635	0.653	0.549	0.569	0.642	0.633	0.690

Given the availability of data and current national policy, we select 11 units in the eastern area, 8 units in the central area, 11 units in the western area and focus our attention on the time period from 2005 to 2015 to calculate the regional patent innovation efficiency of 30 provincial districts using DEA method. We discover that the regional patent innovation efficiency in southeastern coastal areas are generally higher than those of districts in central and western areas. Among regional clusters, the regional patent innovation efficiency of the Yangtze River Delta is the highest, followed by the Pearl River Delta and the Jing-Jin-Ji area. This is because the economy of the Yangtze River Delta is highly developed and its industrial structure is dominated by light industry, which results in lower innovation efficiency. The Jing-Jin-Ji area, in particular in Hebei Province, has a large number of steel-smelting and leather-processing enterprises, which are typically energy-and-emission intensive and put less attention on the innovation activities. The eastern districts have high levels of economic development, and the central districts have substantial resource and environmental carrying capacities. Therefore, districts in these two regions have got considerable patent innovation efficiency.

Based on the estimation of regional patent innovation efficiency, we are particularly interested to investigate the geographic features and regional spatial dependence using the tool of spatial econometric approaches including spatial lag model and spatial error model, with the selecting controlled variables of industrial structure and external trade. The two forms of spatial autocorrelation that are most relevant in applied empirical work are so-called substantive dependence, or dependence in the form of a spatially lagged dependent variable, and nuisance dependence, or dependence in the regression error term.

Table2. Results of spatial econometric regression

index	*OLS*	*SEM*	*SAR*
Constant	-0.152*	-0.155**	-0.241***
Iit	0.125**	0.117***	0.146***
RDLit	0.205	0.284**	0.223***
RDEit	0.365**	0.386***	0.374***
Sit	0.012	0.028**	0.019**
TDit	-0.002	-0.005	-0.004
ρ		0.241**	--
λ		--	0.307**
R-squared	0.560	0.629	0.694
Breusch-Pagan test	6.554**	11.245***	8.395**

Dependent variable: I_{it}. Note: significance indicated as ** for 5%. Period 2005–2015.

The econometric results within groups estimations are presented in Table 2. Both the human capital and R&D expenditure efforts positively determine innovation in a region. Moreover, with these controlling factors, the composition of industrial structure also plays a determinant role in the innovation in the region. These results point to the presence of a positive correlation between specialization and innovation is found in regional areas. According to other scholars' findings, we have also performed a robustness check of the main econometric results after imposing different R&D structures.

Nevertheless, from a deeper analysis of the residuals of the estimation, we detect the existence of spatial autocorrelation, which could be explained as the functions of the volume of imports between two regions and implies that the higher the volume of imports from a region, the higher the volume of innovation that is accessible for the importing region, and thus the higher the intensity of spillovers. Consequently this matrix widens the assumption largely supported by the literature (see Karlsson and Manduchi, 2001, for an empirical survey) that geographical proximity matters in the interregional flow of knowledge and technology. Moreover, both the spatial error model and spatial lag model tests reject the null hypothesis of the absence of spatial autocorrelation in the innovative activity at a 1%level of significance, which points to the necessity of revising the model specification. In this study, the SEM test has a higher value than the SLM test, pointing to a specification of the spatial dependence by means of a spatial error model. Thus, the model changes as:

$$I_{it} = \alpha + \beta_1 RDL_{it} + \beta_2 \mathrm{RDE}_{it} + \beta_3 S_{it} + \beta_4 TD_{it} + \varepsilon_{it},\ \varepsilon_{it} = \lambda \mathrm{W} \varepsilon + u_{it}$$

Given that regions trade mainly with geographically neighboring regions, interregional knowledge spillovers have significant and positive effects on local innovation. Moreover, the regional innovative activities of R&D performed by regional trade partners has demonstrated that policies enhancing regional R&D activities are probably to get a richer effectiveness on stimulating innovation.

Conclusion

The regional patent innovation efficiency in China shows a decline tendency from eastern area to western area, with three major clusters of Yangtze River Delta, Pearl River Delta and Jing-Jin-Ji area. This is because the economy of the Yangtze River Delta is highly developed and its industrial structure is dominated by light industry, which results in lower innovation efficiency. Nevertheless, this paper also detect the effect of interregional externalities on patent innovation from a temporal and spatial perspective, by means of the spatial econometric techniques. Results show that innovation in a region depends on its own R&D efforts, its innovative tradition and its human capital endowments. Moreover, the composition of industrial structure also has a positive effect on innovation. Moreover, the regional innovative activities of R&D performed by regional partners has demonstrated that policies enhancing regional R&D activities are probably to get a richer effectiveness on stimulating patent innovation.

References

Besen, S. M., Raskind, L. J. (1991). An introduction to the law and economics of intellectual property. *Journal of Economic Perspectives*, 5(1), 3–27.

Cabrer-Borras B., Serrano-Domingo G. (2007). Innovation and R&D spillover effects in Spanish regions: A spatial approach. *Research Policy*, 36, 1357-1371.

Christodoulou D., Lev B., Ma L. (2018). The productivity of Chinese patents: The role of business area and ownership type. *International Journal of Production Economics*, 199, 107-124.

Coe, D.T., Helpman, E. (1995). International R&D spillovers. *European Economic Review*, 39, 859–887.

Cohen, W.M., Nelson, R.R., Walsh, J.P.（2000). Protecting Their Intellectual Assets: Appropriability Conditions and Why US Manufacturing Firms Patent (or Not). *NBER Working Paper*, No.7552.

Dang J., Motohashi K.(2015). Patent statistics: a good indicator for innovation in China? Patent subsidy program impacts on patent quality. *China Econ,* 35, 137-155.

Deilmann, C., Lehmann, I., Reibmann, D., Hennersdorf, J. (2016). Data envelopment analysis of cities-investigation of the ecological and economic efficiency of cities using a benchmarking concept from production management. *Ecol. Indicat*, 67, 798-806.

Eaton J., Kortum S. (1996). Trade in ideas: patenting and productivity in the OECD. *J. Int. Econ*, 40 (3-4), 251-278.

Feldman, M.P.(1999). The new economics of innovation, spillovers and agglomeration: a review of empirical studies. *Economics of Innovation and New Technology*, 8, 5–25.

Fritsch M., Franke G. (2004). Innovation, regional knowledge spillovers and R&D cooperation. *Research Policy*, 33, 245–255.

Hu, A.G., Jefferson, G.H. (2009). A great wall of patents: What is behind China's recent patent explosion? *Journal of Development Economics*, 90(1), 57–68.

Jorde, T.M., Teece, D.J. (1990). Innovation and cooperation—implications for competition and antitrust. *Journal of Economic Perspectives*, 4, 75–96.

Klemperer, P. (1990). How broad should the scope of patent protection be? *Journal of Economics*, 21(1), 13–130.

Levin, R.C., Klevorick, A.K., Nelson, R.R., Winter, S.G., Gilbert, R., Griliches, Z. (1987). Appropriating the returns from industrial research and development. *Brook. Pap. Econ. Activ*, 18 (3), 783–832.

Lopez-Bazo, E., Requena, F., Serrano, G. (2006). Complementarity between local knowledge and internationalisation in regional technological progress. *Journal of Regional Science*, 46, 901–930.

Moreno, R., Paci, R., Usai, S. (2005). Spatial spillovers and innovation activity in European regions. *Environ. Plan*, 37 (10), 1793-1812.

Nordhaus, W. D. (1969). Invention, growth and welfare: A theoretical treatment of technological change. Cambridge, MA: The MIT Press.

Rodrigueza, J., Lopezba, D. (2017). Looking beyond the R&D effects on innovation: The contribution of non-R&D activities to total factor productivity growth in the EU. *Structural Change and Economic Dynamics*, 40, 37–45.

Wei Y., Zhang H., Wei J.(2015). Patent elasticity, R&D intensity and regional innovation capacity in China. *World Patent Information*, 43, 50-59.

William G., Thomas L. (1989). Assessing productivity with data envelopment analysis. *Public Product Rev*, 12(4), 361–72.

Yueh L. (2009). Patent laws and innovation in China, *International Review of Law and Economics*, 29, 304-313.

Article DOI: 10.22032/dbt.39353 Proceedings DOI: 10.22032/dbt.39296

Who fosters innovation from U.S. academic patents: A new innovation path from university to government to industry?

*Chunjuan LUAN[1, 2], Bowen SONG[1]

[1]*julielcj@163.com; CLuan@dlut.edu.cn*

Institute of Humanities & Social Sciences, Dalian University of Technology, Dalian, 116085, Ch*ina*

[2]*bowensong333@163.com*

School of Intellectual Property, Dalian University of Technology, Panjin, 124221, China

Introduction

Academic patenting represents the developing directions of future industries (Lee and Gaertner 1994, Dorner, Fryges et al. 2017), and scientific breakthroughs coming from universities can contribute to the emergence of new industries, such as in the case of biotechnology (Guerzoni, Aldridge et al. 2014). Many significant innovations impacting on our lives on a daily basis are the products of professors and students working at universities, from the internet to the nicotine patch, often in ways we don't even realize (Staff Writers 2017). However, who fosters innovation from academic patents? Fostering innovation has been recognized as a better R&D tool than mandates and funding (Snow 2017), facilitating innovation as soon as possible would be a more effective step than simply imposing mandates or increasing funding in breaking an apparent U.S. energy and climate research and development logjam. This paper aims at having an insight into leading cultivators in fostering innovations from academic patents in the leading U.S. states in terms of innovation facilitating, by analysing top Assignees in top Assignee states in respect of U.S. academic patent licenses, a proxy of technology transfer.

Data source and data process

Data in this study is from the official website of *United States Patent and Trademark Office, USPTO*, during the execution years of 1980-2016. Data of technology transfer from academic institutions are needed in this study. First, we search out academic patent assignment data during the execution years during 1980-2016, with search strategy of "Assignor name: university* OR institute* OR college* OR academy*", from the official website of *United States Patent and Trademark Office, USPTO*.

There exists a wide range of writing ways for a specific Assignee in the original patent assignment data recorded in the official website of *USPTO*. Such as there are 14 writing ways for *U.S. Navy***.** It is a time-consuming work for us trying to find out a variety of writing types for a specific Assignee among bulk of patent assignments, and then merge them into one.

Analysis and results

The leading U.S. states fostering innovation

Which states have U.S. academic technology transferred to? The leading U.S. states receiving more patent licenses from academia are considered as U.S. innovation fostering centres. It is recognized that innovation fuels economic growth, and technology transfer is a key driver of successful innovation, which helps the private sector adapt Federal research for use in the marketplace. It is well known that Silicon Valley is an established technology innovation centre in U.S. (Fleming and Frenken 2007, Henton and Held 2013), however, little investigations have been found on technology innovation fostering centres. Outcomes of our empirical analysis by employing previously unexploited data disclose the status of highly concentrated of U.S. technology innovation fostering centres.

Empirical analysis of top Assignee states uncovers that the majority of academic patents have been licensed to just a few U.S. states. There are only 7 Assignee states each getting more than 100 academic patent licenses (**Figure 1**).

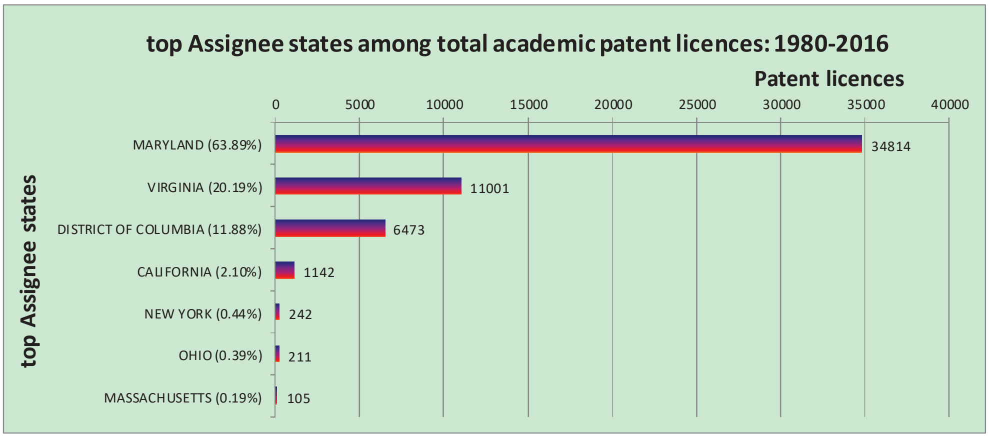


Figure 1. Top Assignee states among total academic patent licences: 1980-2016

Note: *District of Columbia, i.e., Washington District of Columbia, Washington, D.C., or WDC, the Capital of U.S., data of patent licenses is statistically paralleled with U.S. states.*

Figure 1 shows that the state of *Maryland* get the biggest number of academic patent licenses up to 34, 814, accounting for 63.89%; *Virginia* ranks 2nd with 20.19% proportion; *District of Columbia* ranks 3rd with rate of 11.88%. Other top Assignee states with more than 100 patent licenses are *California* with 2.10%, *New York* with 0.44%, *Ohio* with 0.39% and *Massachusetts* with 0.19%, respectively.

The leading fosterers/top Assignees in the leading U.S. states

The leading fosterers in Maryland

The state of *Maryland* is well known as the hometown of high-tech. Maryland gets the biggest number of academic patent licenses during the execution years of 1980-2016, up to 34, 814, accounting for 63.89% of total, being far ahead of other states. Top Assignees in Maryland, that is, academia technology innovation fosterers in Maryland, have been shown in **Figure 2**.

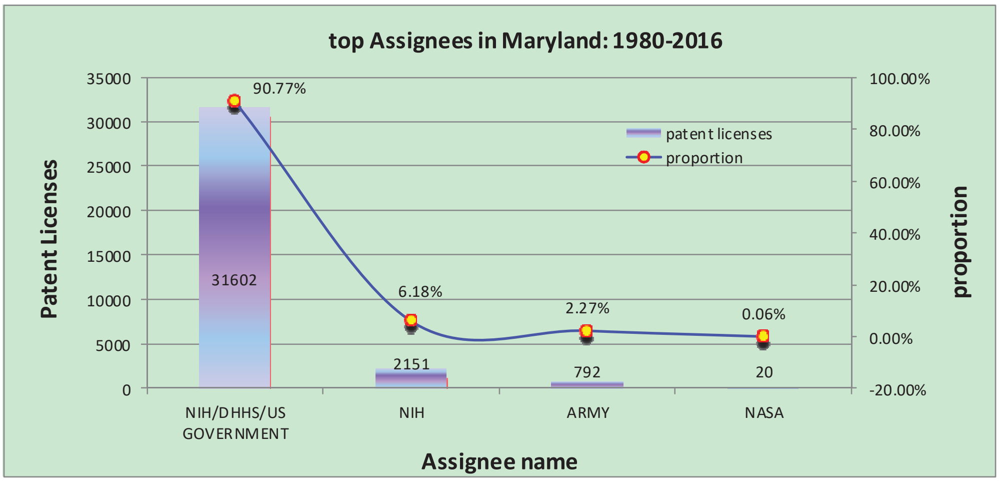


Figure 2. Top Assignees/the leading fosterers in Maryland: 1980-2016

As a whole Assignee of *NIH/DHHS/US GOV.*, abbr. of *National Institutes of Health (NIH)/ U.S. Dept. of Health and Human Services (DHHS)/ U.S. Government*, gets the majority of academic patent licenses up to 31, 602, accounting for 90.77% of total licenses transferred to Maryland, being far ahead of other top Assignees. The total proportion of other Assignees in Maryland is less than 10% and the comparative top ones are as follows: *NIH* (6.18%), *U.S. Army* (2.27%), *NASA* (0.06%), respectively.

The leading fosterers in Virginia

Virginia receives the second biggest number of academic patent licenses, 11001 items, during 1980-2016, accounting for 20.19% of all. Top Assignees in Virginia, that is, leading fosterers in Virginia, have been drawn in **Figure 3**.

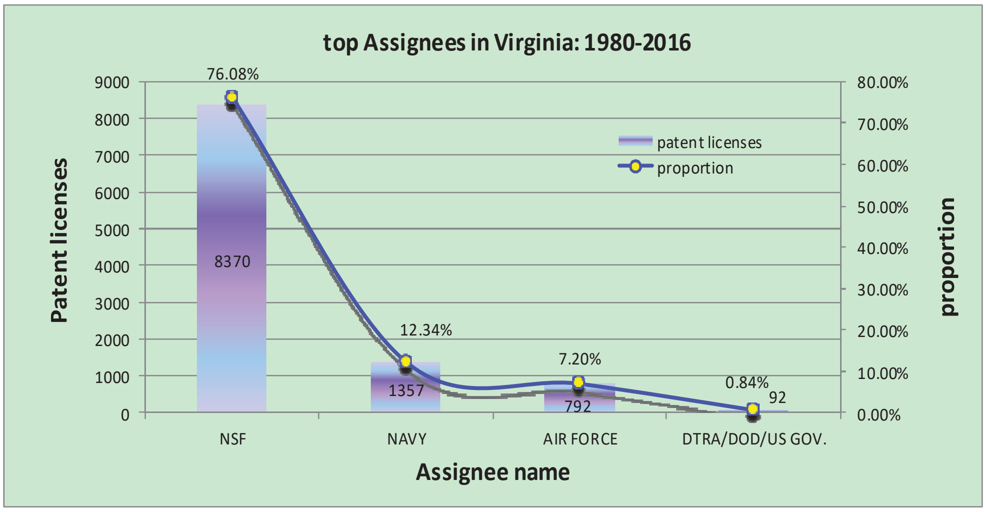


Figure 3. Top Assignees/the leading fosterers, in Virginia: 1980-2016

Figure 3 discloses that *NSF* takes the first place with 8370 academic patent licenses, accounting for 76.08% of total academic patent licenses transferred to Virginia, the most leading fosterers in this state. *U.S. Navy* takes the second place with more than 1000 academic patent licenses, accounting for more than 10% of total academic patent licenses transferred to Virginia. The innovation of *U.S. Navy* is closely related to the location of the world's largest Naval Station Norfolk in Virginia. The proportion of *U.S. Air Force* is listed at the third place.

The leading fosterers in District of Columbia

District of Columbia, i.e., Washington District of Columbia, Washington, D.C., or WDC, the Capital of U.S., data of patent licenses is statistically paralleled with U.S. states. *WDC*, together with America's fifty states composes the United States of America. WDC, as the U.S. political centre, the majority of Federal Government Agencies and foreign embassies are gathering here. WDC gets the third biggest number of academic patent licenses, 6473 items, during 1980-2016, accounting for 11.88% of total. Top Assignees in WDC, that is, leading fosterers in WDC, have been drawn in **Figure 4**.

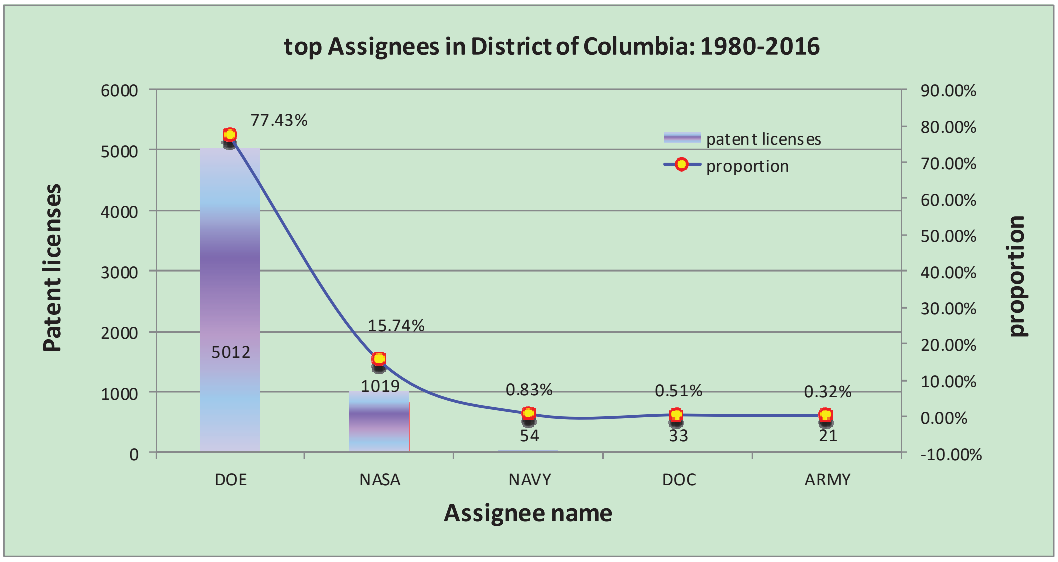


Figure 4. Top Assignees/the leading fosterers, in District of Columbia: 1980-2016

DOE, U.S. Department of Energy, with the mission to ensure America's security and prosperity by addressing its energy, environmental and nuclear challenges through transformative science and technology solutions, takes the first place with 5012 academic patent licenses, accounting for 77.43% of total academic patent licenses transferred to *WDC*, the most leading innovation fosterer in this area, being far ahead of other fosterers. *NASA* takes the second place with more than 1000 academic patent licenses, accounting for 15.74% of total academic patent licenses transferred to *WDC*.

The leading fosterers in California

California is one of the largest users of energy for it is the most populous U.S. state, and there are a few national laboratories of *DOE* locate in California, such as *Lawrence Berkeley National Laboratory at Berkeley*, California (founded in 1931); *Sandia National Laboratories at Livermore*, California (founded in 1948); *Lawrence Livermore National Laboratory at Livermore*, California (founded in 1952); *SLAC National Accelerator Laboratory at Menlo Park*, California (founded in 1962); et al. which play a significant role in fostering innovation from academic inventions, especially in the area of energy.

California ranks the fourth of academic patent licenses, total 1142 items, during 1980-2016, accounting for 2.10% of all. Top Assignees/the leading fosterers in California have been shown in **Figure 5**.

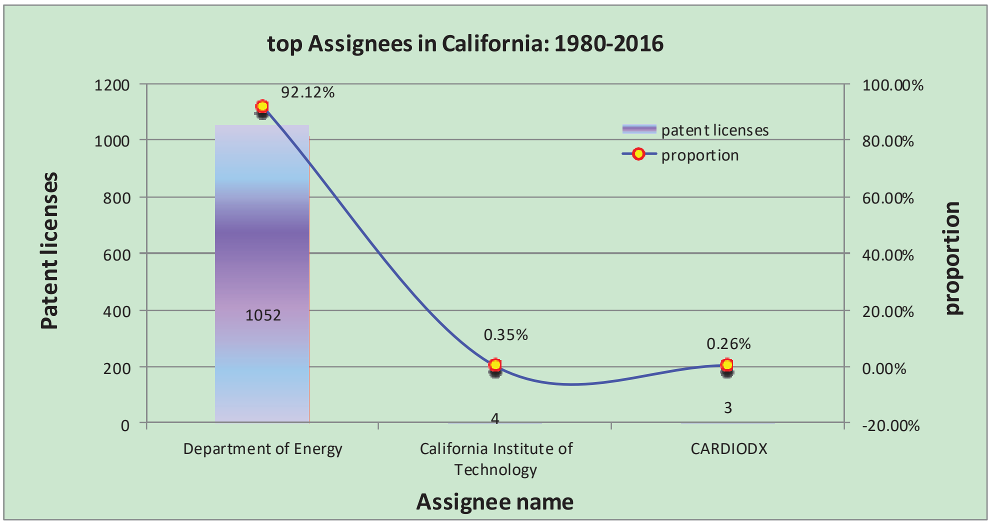


Figure 5. Top Assignees/the leading fosterers in California: 1980-2016

DOE takes the first place receiving 1052 academic patent licenses, accounting for 92.12% of total academic patent licenses transferred to *California*, the most leading fosterer in this state, being far ahead of other fosterers.

Discussions

Traditionally, industries are widely recognized that they take responsibilities for industrialization from academic inventions and further to commercialization (Lee and Gaertner 1994). However, our empirical analysis of top Assignees in top Assignee states in terms of U.S. academic patent licenses, by employing previously unexploited data disclose that it is U.S. government agencies who are the leading fosterers in fostering innovation from academic inventions. The findings of this study disclose that U.S. government has played a significant role in fostering technology innovation from academic patents instead of industries. Such results let us reflect on the government role in Triple Helix innovation system. The concept of the Triple Helix of university-industry-government relationships initiated by Etzkowitz (Etzkowitz 1996) and Etzkowitz and Leydesdorff (Etzkowitz and Leydesdorff 2000), interprets the shift from a dominating industry-government dyad in the Industrial Society to a growing triadic relationship between university-industry-government, UIG model, in the Knowledge Society (Stanford 2017).

Why the U.S. government is the academic technology fosterer rather than the industry? On one hand, it is generally acknowledged that universities are creators of new knowledge in history which have brought the revolutionary breakthroughs for human society and technology development(Guerzoni, Aldridge et al. 2014, McGrath 2015). On the other hand, general knowledge resulted by basic scientific research provides the means of answering a large number of important practical problems, though it may not give a complete specific answer to any one of them (Bush 1945). Based on such situations, the industry tends to be lack of motivations in fostering the emerging and strategic technologies coming from universities for the uncertainties and risks(Tsai, Lin et al. 2009, Tang, Murphree et al. 2016), whereas U.S. government has recognized the significance for supporting academic research from 1945 when the Second World War ended, or even earlier (Bush 1945, Hong, Lippman et al. 1995, Aizenman and Noy 2007), and further has promoted academic technology fostering via the implementations of a series of related policies (Negoita 2014, Liu and Guan 2016, Zehavi and Breznitz 2017), such as *Bayh-Dole Act of 1980, Stevenson-Wydler Technology Innovation Act of 1980, Small Business Technology Transfer Act, STTR* of 1992, *The America COMPETES Act of 2007*, et al.

What measures U.S. government have taken for boosting innovation? U.S. government takes a positive attitude and a series of steps in fostering academic inventions. Government selects and hatches cutting-edge technologies from universities; simultaneously, government adopts steps deregulating industries, encouraging emerging industries development, affording education and training for the structural unemployment, offering subsidies for sunset or uncompetitive industries, et al., for the groups losing interest due to technological change to insure the transform implemented (Miyazaki and Islam 2007, Reich 2009); government also implements public procurement and related financial policies in boosting technology innovation from academic patents. United States is acknowledged as the first and the most successful country adopting public procurement promoting innovation and emerging industries development (Aschhoff and Sofka 2009, Uyarra, Edler et al. 2014, Hellsmark and Soderholm 2017). The implementation of U.S. relevant financial policies have played a crucial role in facilitating and supporting emerging industries development in the fields of high technologies around 1970s.

U.S. government role in fostering technology innovation from academic inventions has significant policy implications and referential values to Chinese government in the process of construction of innovative country (Liu and Chen 2012, McMahon and Thorsteinsdottir 2013). A key problem for Chinese government to solve is the considerable low rate of university technology transfer (Zhang and Gallagher 2016, Zhang, Duan et al. 2016). A few related laws and regulations for promoting academic inventions transfer have been introduced in China recent year, even including a *Bayh-Dole-like law* named *the Revised Science and Technology Progress Law* implemented in 2008. However, Chinese legal system construction pertinent to university technology transfer seems playing a weak role in boosting academic technology innovation.

It is essential for Chinese government to foster academic inventions running ahead of industries. Industries tend to take a wait-and-see attitude to emerging technologies, especially academic inventions, due to the risks and uncertainties in the process of technology innovation. Government's involvement into fostering innovation includes the following steps: select emerging, cutting-edge technologies, especially from academic patents, which will play crucial roles in the future industry and society;

foster such potential technologies and promote them into industrialization and commercialization; further globalization. A number of high-tech industrializations promoted by U.S. government in advance since 1990s, have developed into mainstay industries of national economy afterwards (Doutriaux 1992, Etzkowitz 1997, Merchant 1997, Pisano and Shih 2009), further bloomed into high-end industries in 2000s.

Another effective measure Chinese government should take is public procurement in boosting innovation. Government procurement can reduce market uncertainty and risk (Keller 2011, Arve and Martimort 2016, De Clerck and Demeulemeester 2016), strengthen industries' confidence adapting new technologies and products. United States is acknowledged as the most successful country applying government procurement promoting innovation in the world (Aschhoff and Sofka 2009, Uyarra, Edler et al. 2014, Hellsmark and Soderholm 2017).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 71774020 & 71473028; Fundamental research funds for the central universities under Grant No. DUT17RW224.

References

Aizenman, J. and I. Noy (2007). Prizes for basic research: Human capital, economic might and the shadow of history. Journal of Economic Growth 12(3): 261-282.

Arve, M. and D. Martimort (2016). Dynamic Procurement under Uncertainty: Optimal Design and Implications for Incomplete Contracts. American Economic Review 106(11): 3238-3274.

Aschhoff, B. and W. Sofka (2009). Innovation on demand-Can public procurement drive market success of innovations? Research Policy 38(8): 1235-1247.

Bush, V. (1945). Science. The endless frontier. A report to the president. Washington, DC: United States Government Printing Office.

De Clerck, D. and E. Demeulemeester (2016). Creating a More Competitive PPP Procurement Market: Game Theoretical Analysis. Journal of Management in Engineering 32(6).

Dorner, M., H. Fryges and K. Schopen (2017). Wages in high-tech start-ups - Do academic spin-offs pay a wage premium? Research Policy 46(1): 1-18.

Doutriaux, J. (1992). EMERGING HIGH-TECH FIRMS - HOW DURABLE ARE THEIR COMPARATIVE START-UP ADVANTAGES. Journal of Business Venturing 7(4): 303-322.

Etzkowitz, H. (1996). The triple helix: Academic-industry-government relations - Implications for the New York regional innovation environment. Technology Link to Economic Development: Past Lessons and Future Imperatives. S. U. Raymond. New York, New York Acad Sciences. 787: 67-86.

Etzkowitz, H. (1997). From zero-sum to value-added strategies: The emergence of knowledge-based industrial policy in the states of the United States. Policy Studies Journal 25(3): 412-424.

Etzkowitz, H. and L. Leydesdorff (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations. Research Policy 29(2): 109-123.

Fleming, L. and K. Frenken (2007). The evolution of inventor networks in the silicon valley and Boston regions. Advances in Complex Systems 10(1): 53-71.

Guerzoni, M., T. T. Aldridge, D. B. Audretsch and S. Desai (2014). A new industry creation and originality: Insight from the funding sources of university patents. Research Policy 43(10): 1697-1706.

Hellsmark, H. and P. Soderholm (2017). Innovation policies for advanced biorefinery development: key considerations and lessons from Sweden. Biofuels Bioproducts & Biorefining-Biofpr 11(1): 28-40.

Henton, D. and K. Held (2013). The dynamics of Silicon Valley: Creative destruction and the evolution of the innovation habitat. Social Science Information Sur Les Sciences Sociales 52(4): 539-557.

Hong, W. K., S. M. Lippman, W. N. Hittelman and R. Lotan (1995). Retinoid chemoprevention of aerodigestive cancer: From basic research to the clinic. Clinical Cancer Research 1(7): 677-686.

Keller, L. R. (2011). Multiattribute and Intertemporal Preferences, Probability, and Stochastic Processes: Models and Assessment. Decision Analysis 8(3): 165-169.

Lee, Y. and R. Gaertner (1994). TECHNOLOGY-TRANSFER FROM UNIVERSITY TO INDUSTRY - A LARGE-SCALE EXPERIMENT WITH TECHNOLOGY DEVELOPMENT AND COMMERCIALIZATION. Policy Studies Journal 22(2): 384-399.

Liu, M. C. and S. H. Chen (2012). MNCs' offshore R&D networks in host country's regional innovation system: The case of Taiwan-based firms in China. Research Policy 41(6): 1107-1120.

Liu, N. and J. C. Guan (2016). Policy and innovation: Nanoenergy technology in the USA and China. Energy Policy 91: 220-232.

McGrath, R. G. (2015). The Academic Entrepreneur: A Biographical Sketch of Ian C. MacMillan's Contributions to Establishing the Field of Entrepreneurship. Strategic Entrepreneurship Journal 9(2): 188-204.

McMahon, D. and H. Thorsteinsdottir (2013). Pursuing endogenous high-tech innovation in developing countries: A look at regenerative medicine innovation in Brazil, China and

India. Research Policy 42(4): 965-974.

Merchant, J. E. (1997). The role of governments in a market economy: Future strategies for the high-tech industry in America. International Journal of Production Economics 52(1-2): 117-131.

Miyazaki, K. and N. Islam (2007). Nanotechnology systems of innovation - An analysis of industry and academia research activities. Technovation 27(11): 661-675.

Negoita, M. (2014). Globalization, state, and innovation: An appraisal of networked industrial policy. Regulation & Governance 8(3): 371-393.

Pisano, G. P. and W. C. Shih (2009). Restoring American Competitiveness. Harvard Business Review 87(7-8): 114-+.

Reich, R. B. (2009). Government in Your Business. Harvard Business Review 87(7-8): 94-+.

Snow, N. (2017). Fostering innovation seen as better R&D tool than mandates, funding. Oil & Gas Journal 115(4A): 23-24.

Staff Writers (2017). 100 Important Innovations That Came From University Research. http://www.onlineuniversities.com/blog/2012/08/100-important-innovations-that-came-from-university-research/.

Stanford. (2017). The Triple Helix concept. from http://triplehelix.stanford.edu/3helix_concept.

Tang, L., M. Murphree and D. Breznitz (2016). Structured uncertainty: a pilot study on innovation in China's mobile phone handset industry. Journal of Technology Transfer 41(5): 1168-1194.

Tsai, Y. Y., J. Y. Lin and L. Kurekova (2009). Innovative R&D and optimal investment under uncertainty in high-tech industries: An implication for emerging economies. Research Policy 38(8): 1388-1395.

Uyarra, E., J. Edler, J. Garcia-Estevez, L. Georghiou and J. Yeow (2014). Barriers to innovation through public procurement: A supplier perspective. Technovation 34(10): 631-645.

Zehavi, A. and D. Breznitz (2017). Distribution sensitive innovation policies: Conceptualization and empirical examples. Research Policy 46(1): 327-336.

Zhang, F. and K. S. Gallagher (2016). Innovation and technology transfer through global value chains: Evidence from China's PV industry. Energy Policy 94: 191-203.

Zhang, G. P., H. B. Duan and J. H. Zhou (2016). Investigating determinants of inter-regional technology transfer in China: a network analysis with provincial patent data. Review of Managerial Science 10(2): 345-364.

Article DOI: 10.22032/dbt.39354 Proceedings DOI: 10.22032/dbt.39296

User Interaction with online Information Resources: an Informetrics Approach

[1]Mohammad Hassanzadeh and Emran Ghorbani[2]

[1]hasanzadeh@modares.ac.ir
Associate Professor, Tarbiat Modares University
Room No. 315, Management and Ecenomics School,
Tarbiat Modares University, Tehran, Iran. P.O.Box: 14115-111
Telefax: +98-21-82883678

[2]emranghorbani2014@gmail.com
Librarian, Hamedan Central Public Library
Hamidreza mahmoodi
Tarbiat Modares University
Majid Bashirzadeh
Allameh Tabatabaei University

Abstract

This research strived to investigate and provide answer to these questions: how researchers interact with online information resources when they search for information to meet their information needs. The research also intended to address the issues and problems involved in information retrieval to provide an appropriate solution to their challenges. We have conducted a qualitative approach to the investigation. We carried out interview to collect data. In this way, we chose and interviewed the 15 most informed scholars who normally interact with information and resources.

Results indicated that the use of web-based information resources is dominants among researchers. Researchers were more likely to get the resources they needed from social networks. Consulting the databases and scientific social networks such as Google Scholar was common. Among the online information resources, the use of academic journals, specialized and public libraries website and online resources and personal repositories has been among highly commended resources.

Among the challenges that researchers have had in using information sources was that some of resource titles did not reveal information within them. Researchers were to search for hours and days to find the information they needed. The Web environment, is still far from providing a well-organized information to information seekers and users. Lack of standardized format for searching information in search engines for databases lead researchers to misinformation, and most of interviewees complained about this matter. Results also showed that researchers categorize the retrieved information into four components: 1. Explicit useful information 2. Hidden useful information 3. Explicit inappropriate information 4. Explicit Disturbing information. Findings advise information systems policy makers to adjust their propositions on the information behavior of the new generation researchers and online residents and revise their indexing and collection development guidelines.

Keywords: Online Information Resources, User Interaction, Informetrics, Explicit useful information, Hidden useful information, Explicit inappropriate information, Explicit Disturbing information.

Introduction

The universe of information and especially the Internet environment can be likened to a large store that all researchers and the public can access their information. But the way it works is not so simple because the store is so big that if on the one hand the information is not systematically classified and on the other hand, the majority of users are not familiar with the searching principles and methods. This may result in searching for hours and even for days and reaching in insufficient information, and even coming back with empty hands from this store. The importance of recognizing the information resources in accordance with the needs of the researchers is crucial to meet users' information needs. If researchers are not able to collects the information they need in a satisfactory way, the research they are carrying out, will not be satisfying.

Amid the increasing growth of data, retrieving relevant information is crucial, and online search is one of the most challenging issues on the Internet era. Online search tools try to provide capabilities that extracting the most relevant records for user needs (Haghighi Nobjari, 2014). Machine language (ML) seems to be an appropriate solution, but the natural language is very complex, and machine translation has not reached in a maturity level to help exactly relevant information retrieval (Madanker, Chandak and Chuan, 2016). The importance of understanding the sources of information required in accordance with the wishes Researchers have a special place to meet their information needs . So that if researchers are not able to get the information they need in a satisfactory way, the research they are doing will not be expected.

System-user interaction is one of the most important aspects of information retrieval (Sadooghi et al., 2011). On the one hand, the information organization has to be organized in a completely uniform, standard and transparent manner; on the other hand, users should have the skills to retrieve the information they need. Each information system has a specific basis for analysis, in which the system is based on the interpretation of information and the correspondence between documents and information requests, and thus data retrieval is carried out (Gazni, 2001). Information retrieval systems have also sought to achieve the goal of meeting the demands of users and the documents presented as search results during their history. (Hassanzadeh, 2004).

On the other hand, researchers should be able to determine the amount and type of data and information and classify the data and information they need for their research. As Nielsen and Bjorland believe: Observations that can be considered as research findings for a scientist may be considered as research background for another researcher (Nielsen and Bjorland, 2014).

This study aimed to find out how researchers interact with online information sources when they search for information, and address their information needs. A lot of research on information literacy, information seeking behavior, databases, information skills, database design problems and so on were done.

Dwyer states that students are more inclined to learning, knowledge and thinking about issues with training and frequent research. Also, by designing and conducting exercises that look for a variety of information sources (Quoted from: Babaie and Bigdeli, 2015). Mansourian and Yazdani (2015) also concluded that Kharazmi University graduate students do not use any systematic model for information seeking behavior patterns and are often not familiar with databases and how they are searched.

Hassanzadeh et al. (2016) reported a serious challenges in the information seeking behavior, which is attributed to students' lower skills. They point out that the same low skill makes students look for simple, try and error practice. On the other hand The poker (2008) Recalls the Facebook social network as one of the research tools, and the most important obstacle that makes use of these tools and other similar tools less widely known is how to use these tools efficiently (quoted in Yari, 2016). The results of the research are vital and Persian (2016) as well Suggesting that the greatest barriers to information seeking by researchers are related to the search for information sources, the use of information resources and search strategies .

On the other hand, Lewandowski (2008) concludes that users should not focus their results on a particular language when searching for information they need. But to improve the results, they have to do their search in all languages. Lynn Robinson (2013) aims to provide an insight into the information behavior of existing models using the library method and the analysis of payment information search and communication models and concluded that although most of the models in the library and information resources focus on information and intelligence users, the existing models have common elements.

Madder and colleagues (2015) also highlighted the most important parts of data retrieval, retrieval of cross-language data, multilingual information retrieval, and machine translation approaches and techniques. Using the questions they asked in a language and asking them to retrieve documents in one or more languages, they found that translation of the machine played an important role in the system. CLIR [10] and MLR [11] There.

Therefore, it can be concluded that an integrated organization with a uniform standard on the one hand as well Understanding the search method, on the other hand, will help researchers find helpful, more intimate, and faster information for researchers. This will help them on the one hand in managing time and on the other hand will lead to higher quality research. As that Russell, Chamberlain and Azzopardi [12] (2018believe that the search for the main task of legal scholars, health information professionals, and other areas has been identified.

Questions

1. What are the best practice of researchers from interacting with online information resources?
2. What is the researchers' perception of online information sources before they enter the search process?
3. What are the researchers' perceptions of online information sources after their acquisition and use?
4. What is the mechanism for improving the efficiency of online information resources from the perspective of researchers?
5. What are the challenges of online information resources for researchers?
6. What is the difference between the sources of online information and non-online information resources and how remarkable is this distinction?
7. What are the distinctive features of online information resources with non-online sources of information

Research method

This research is a fundamental research based on the Strauss and Corbin grounded theory. In this way, we chose and interviewed the 15 high-ranked scholars who naturally have more interaction with information resources.

Interview questions were drawn from the review of the research, which had the most relevancy with this research. Validity and reliability of the questions were also examined to improve the quality of results.

Findings

A) Demographic and descriptive information

Age of interviewees ranged from 26 to 40 Years, out of 15 persons, 5 was female and 11 male. Seven persons have masters Degree and 8 with PhD. Eight researchers were from information science, and 7 from other fields. The average web surfing time by these researchers was 5 hours per day and for scientific information.

Google, Google Scholar and Microsoft Academic were top search engines they had visited. Among the databases, Science Direct, Doaj, Web of science, Scopus were the mostly visited and Emerald the least used databases.

The use of synonyms, subject terms, time limits, searches based on resource formats and authors' names, and the identification of specialized databases have been among the most widely used strategies by researchers. The researchers mostly were intended to retrieve scientific papers in pdf format. To a large extent they were to use the search formulation. The web environment, library, and other researchers were the paths and resources that the researchers examined to find information through it. The level of familiarity of researchers with scientific social networks was good and LinkedIn, Research Gate, and Academia were among popular scientific social networks . The satisfaction about the organization of information on the Web was acceptable, but researchers suggested that precise categorization of information on the web environments could result in more accurate results. It was also suggested to create a specialized federated search engine that can search information sources from all databases against user queries.

B) Qualitative findings

Question: What are the best practices of researchers on interacting with online information resources?

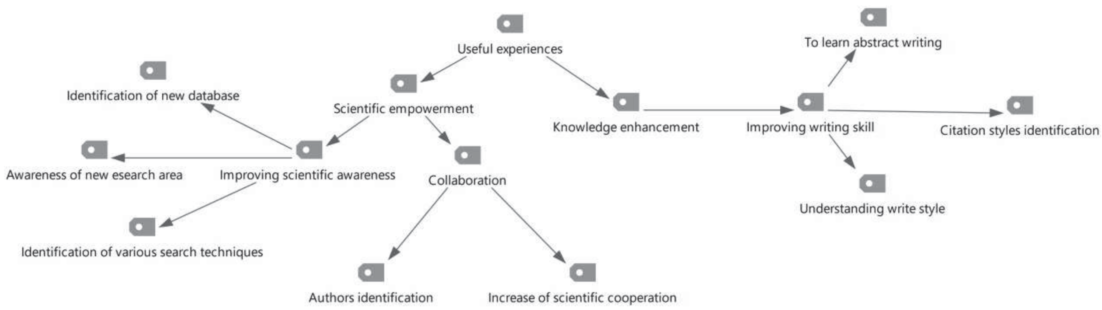


Figure 1. Best practices of the researchers

The results indicate that researchers consider online information sources as a tool for their knowledge development. They are due to meet different research techniques, aware of the new research fields and meet new databases, to strengthen their scientific consciousness. In this regard, one researcher responds to the question "What are the useful experiences of researchers with interacting with online information resources?"
Replied: "Find new management fields - Find new researcher colleagues - Increasing search skills and research methods." Online information resources are a good platform for collaborative effort. Individuals are becoming familiar with other authors through the use of online information resources, and subsequently expanding their collaborative work.
As stated above, online information resources bring knowledge to individuals and users. The researcher strengthens his writing skills while seeking and using information resources. In other words, getting acquainted with how to cite, write, and abstract is a ground for strengthening the researcher's writing skills.

Question: What are the researchers' perceptions of online information sources before they enter the search process?

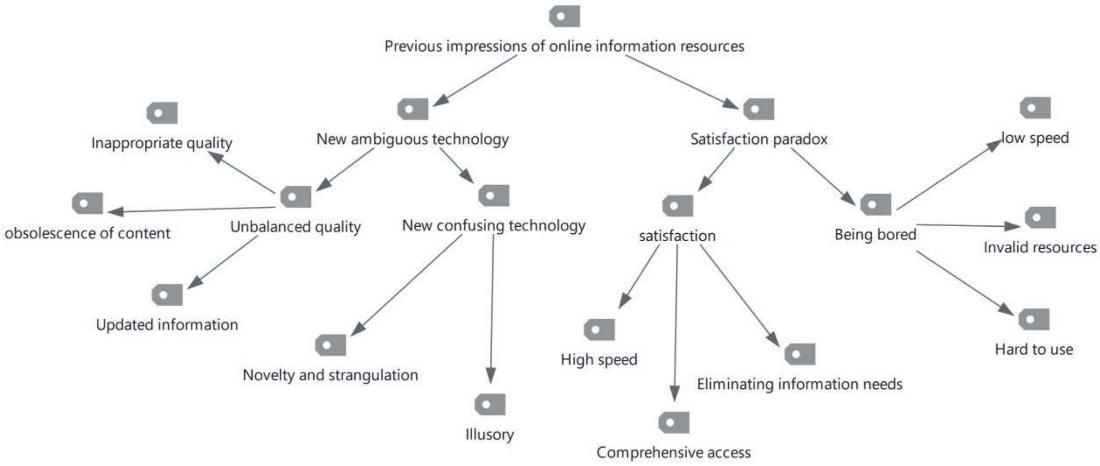


Figure 2. Researcher's impression before using online information resources

Most scholars use it as a technology challenge before they use online information resources. From the researchers' point of view, the aging of the content and, sometimes, the inadequate quality of information resources, the provision of up-to-date information along with updated resources indicates the unbalanced quality of online information resources. It seems that the use of information resources brings satisfaction to users. However, during the interviews, there is a kind of panic. The contradiction between the satisfaction and the boredom of information resources is the paradox of satisfaction. Comprehensive access, high access speeds, and high-quality access to the information are satisfaction factors for researchers. In some cases, there are some sources of invalid and difficult online resources for volatile researchers.
Question: What are the researchers' perceptions of online information sources after their access and use?

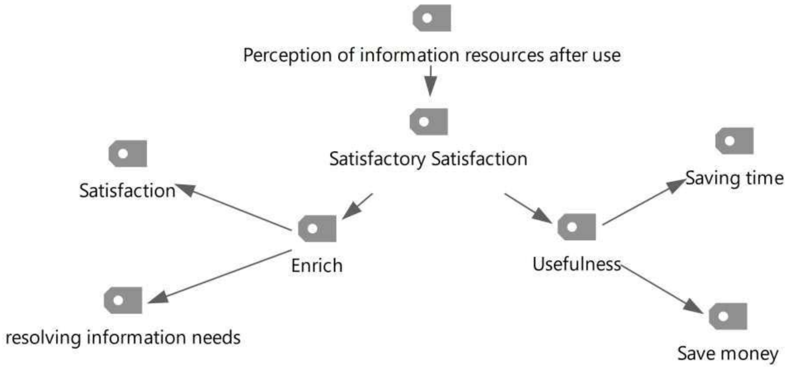


Figure 3. Researcher's impression after using information sources

The researchers stated that on the one hand, they could access online resources at the very least in the shortest possible time and at a minimal cost, and in many cases completely free of charge. On the other hand, they believed that online information resources largely solved their information needs and, therefore, had a great deal of satisfaction with these resources. The foregoing points out that online information sources seem to have some kind of usefulness and optimism.

Question: What is the mechanism for improving the efficiency of online information resources from the perspective of researchers?

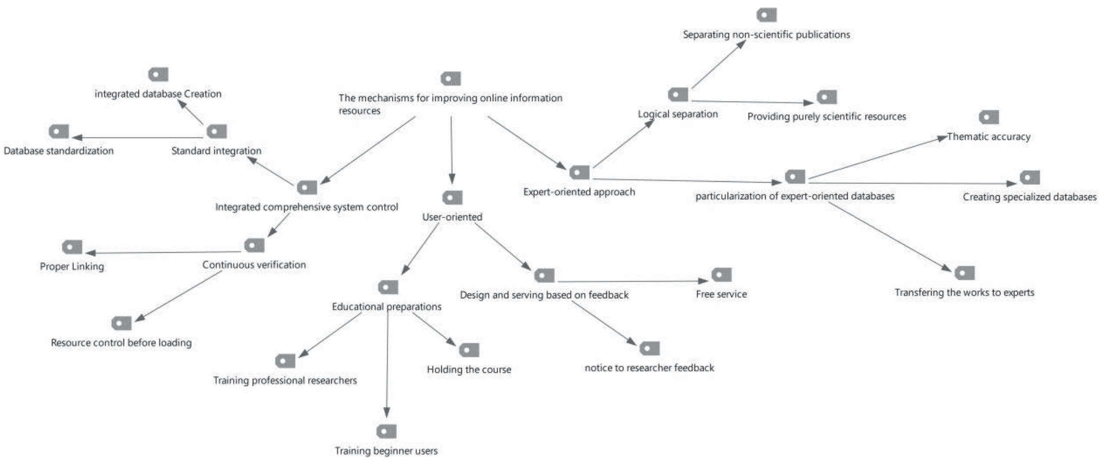


Figure 4. The mechanisms for improving online information resources

Researchers suggested expertise in single-minded military control, and user-centered information retrieval systems to increase productivity and effectiveness of online information sources. They proposed educational tools that include training beginner users, educating researchers, and holding classes and workshops. They also proposed designing information retrieval systems based on feedback from researchers and demanding the provision of free research services.

They asked specialists to provide information resources that would be devoted to the logical separation of resources by separating non-scientific publications and providing scientifically valid sources. They also called for databases to be expert-focused and resource-based, so that resources are divided into very small topics at bases to make job retrieval easier. Also, specialist databases in all disciplines were emphasized by subject specialists. The researchers called for a single standard definition for all databases, including the establishment of an integrated database of information and standardization of databases.

Question: What are the challenges of online information resources for researchers?

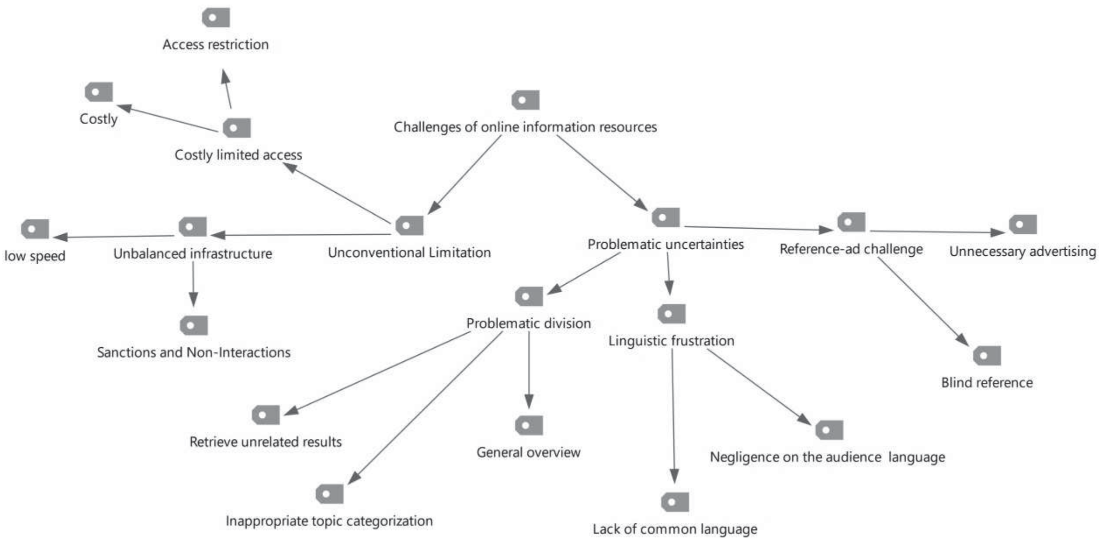


Figure 5. Online Information Resources Challenges

It seems that most researchers encounter problems when using online information resources that are not pleasant to them. The results suggest that working with online information sources sometimes leads to disappointment. Disappointment which comes from unconventional limitation and pervasive ambiguity. Unconventional constraints, unbalanced infrastructure and limited access to cost. In other words, an unbalanced infrastructure and limited access to the cost of a kind of unconventional constraint make it into mind. Access restrictions and the cost of some information resources fall under the category of cost-effective access, and the low speed of the Internet and the lack of strong international interactions are also considered to be unbalanced infrastructures. All of the aforementioned articles have been described in the description and explanation of the unconventional limitations. But the other piece of disappointment is ambiguity. The category underlying the challenge of referral - advertising, linguistic and linguistic distortion .The challenge of referrals - Advertising in some databases, including unnecessary advertisements and blind referrals, was something that was considered to be very annoying for users. Among the other challenges mentioned by the researchers was the lack of language, among which the lack of a common language and lack of attention to the language of the users was very high. False drop was one of the major issues that confused researchers and the retrieval of irrelevant results, inappropriate topic categorization, and general overview were the most important ones mentioned in this area.

Question: What can be categorized online resources from the perspective of researchers?

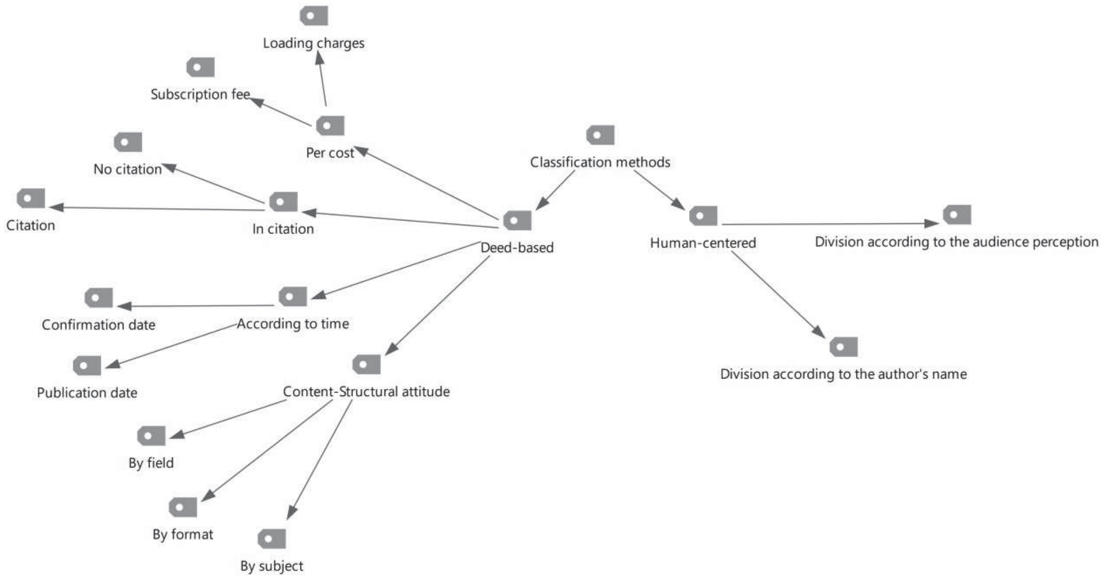


Figure 6. Information resource categorization

The resource categorization by author's name and perceived audience, which is based on the researcher-centered nature of the information resources, was the point that the researchers were considering. The breakdown by the time of approval and publication of the articles was another proposal proposed by the researchers. Citation-oriented categorization, i.e. categorization in terms of articles with cited and non-cited articles, also appears in the interviews. Categorization according to loading and subscription costs is also based on semantic convergence and nature they were placed below the cost center axis.

Question: What is the difference between the sources of online information and non-online information resources and how remarkable is this distinction?

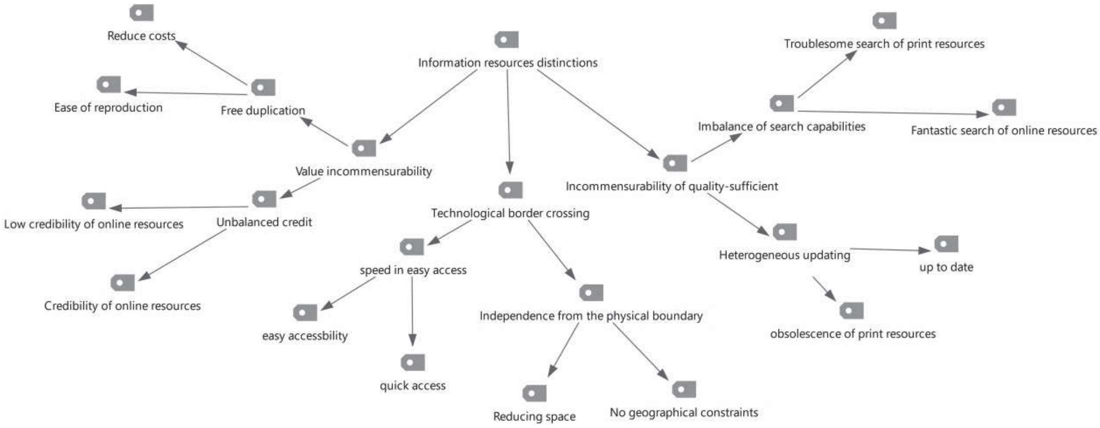


Figure 7. difference between information sources

The distinction between online sources of information and non-online sources can be sought at the cutting edge of technology and credit unequal - capability. Researchers consider ease and speed of access to these resources very satisfying. They are Maintain physical independence from the border directly to eliminate the constraints of location and space reduction are concerned that the distinction wedge with non-online online information resources is considered.

Researchers would appreciate the possibility of free downloading, given the cost and ease of replicating online information sources. They referred to interviewers with high reputation for print resources and low credit ratings for some online sources. Researchers also acknowledged that it was difficult to search for print resources and made it easy to search for online information sources. Also, the aging of print sources was considered a new defect and assumed the availability of online information sources as an advantage for these resources.

Conclusion

The challenges that researchers have had in using information sources was that some of the titles of papers and other research papers did not reveal information within them . Researchers were forced to search for hours and days to find the information they needed. The world of information, especially in the Web environment, did not have a good idea to organize information. Some profiteers made free information in cash from elsewhere. There was no standardized format for searching information in search engines for databases, and the most complaints from researchers were the same.

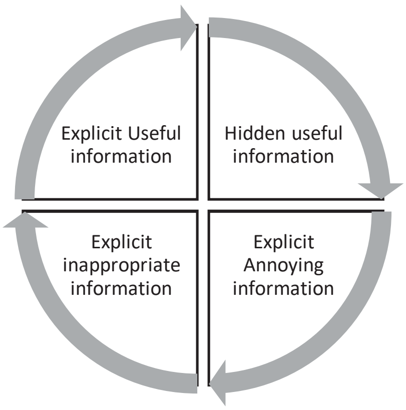


Figure 8. Four main categories of information

Explicit useful information can be easily seen and used by researchers and the general public, and will not be a major effort. The researcher may find this information sufficiently adequate and this fact my prevent them struggling more.
Hidden information is useful to most scholars, especially ordinary people, because they are not familiar with the principles of professional and advanced search, or that the title of the article is written so that many people think that they do not cover their information needs. Or, the title is not an article
Explicit inappropriate information is unprofitable, which may be thought to be the most annoying type of information for a researcher. That is, information that does not need it, but the researcher has to study them, so that he loses something. The classification of information as well as the presence of supervisors in the web environment can, to a large extent, overcome this problem to further monitor the flow of information.
Annoyingly revealing information is also information that the investigator reviews, but there is no useful information in them. In this case, the time spent by the researcher is lost, and on the other hand, he may lose a lot of useful information. For example, a scholar may want to look for information about Goethe's biography when he finds books about his subject, but after studying, he does not find any useful information.

References

Babaei, Kobra; Bigdeli, Zahid (1394). The need to know and its impact on participatory information seeking behavior of undergraduate students of the Faculty of Psychology and Educational Sciences of Allameh Tabatabai University. Human interaction and information, 2 (1), 12-20.

Bian, GW, & Chen, HH (1998). Integrating query translation and document translation in a cross-language information retrieval system. In the Conference of the Association for Machine Translation in the Americas (pp. 250-265). Springer, Berlin, Heidelberg. The

Gazni, Ali (1380). Organize information in information retrieval systems. Library and Information Science, 6 (1), 70-95.

Good work, Maliha, vital, Zuhair Vfarsy Azam. (2016). The study of barriers to information seeking from the viewpoint of faculty members of Islamic Azad University, Shiraz Branch, based on the development of Derwin's theory of meaning (theory of work). Research a Librarian, 6 (2), 61-79.

Haghighi Noobjari, Nowruz, Zahra, (2014), Information Recovery, National Conference on Engineering Sciences, New Ideas (8) , Tonekabon, Tonekabon's Elderly Higher Education Institute,

Hassanzadeh, Mohammad (2004). Effect of information retrieval models on relevancy. Journal of Informatics, 2 (1), 64-89.

hassanzadeh, Mohammad; Hasanzadeh, Reza; Safai, Zahra; Ghorbani; Omran; Dastjerdi Ali Akbar and Parvin, Somayeh (1395) . The pattern of accessing and utilizing electronic information among graduate students. Human- information interaction, 3 (1), 1-19.

Help, Shiva (1395). Social media in public libraries: Understanding uses and barriers and usage problems. Human interaction and information, 3 (1), 79-95.

Lewandowski, D. (2008). Problems with the use of web search engines to find results in foreign languages. Online information review, 32 (5), 668-672.

Madankar, M., Chandak, MB, & Chavhan, N. (2016). Information retrieval system and machine translation: a review. Procedia Computer Science, 78, 845-850.

Mansourian, Yazdan and Sangari, Mahmoud (2011). The study of the information seeking behavior of graduate students of Kharazmi students in dissertation development, human interaction and information, 2 (2), 1-7.

Nielsen, Jørn, H. and Hjørland, B., (2014). Curating research data: the potential roles of libraries and information professionals. Journal of Documentation, 70 (2), pp.221-240.

Nowkarizi, Mohsen; Kashi, Zahra and Sanat Joo, Azam. (2017). the role of librarians, Close Directory amount Svadatlaaty students education Supplementary UniversityFerdowsi Mshhddrka hsh Aztrabatla finding them. Library and Information Science, 20 (4), 30-55.

Robson, A., & Robinson, L. (2013). Building on the models of information behavior: linking information seeking and communication. Journal of documentation, 69 (2), 169-193.

Robson, A., & Robinson, L. (2013). Building on the models of information behavior: linking information seeking and communication. Journal of documentation, 69 (2), 169-193. The

Russell-Rose, T., Chamberlain, J., & Azzopardi, L. (2018). Information retrieval in the workplace: A comparison of professional search practices. Information Processing & Management, 54 (6), 1042-1057.

Sadoughi, Farahnaz, but Rahai, Ali; Vakili Mofrad, Hussein; Hafiz MohammadHassanzadeh and Bourgi, Hamid (2011). Interacting in information retrieval and evolution of its models. Journal of Science and Technology Research Institute of Iran, 159-1550.

Vanopstal, K., Buysschaert, J., Laureys G., & Vander Stichele, R. (2013). Lost in PubMed. Factors influencing the success of medical information retrieval. Expert Systems with Applications, 40 (10), 4106-4114.

Article DOI: 10.22032/dbt.39355 Proceedings DOI: 10.22032/dbt.39296

Finding copies of an image: a comparison of reverse image search systems on the WWW

Paul Nieuwenhuysen

Paul.Nieuwenhuysen@vub.be

University Library, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussel, Belgium

Abstract

Objective - The reported investigation aimed to assess the performance of services that offer search by image on the Internet / WWW to find copies of a known image. The motivation is that finding copies of an image can be useful for several purposes, while search by image is a relatively new method that allows finding such images and that is offered free of charge by a few Internet search services. Furthermore, the context of each revealed image copy may yield relevant information.
Methods - Various images have been used in empirical case studies.
Results - Finding copies is positioned in a structured view of applications of reverse image searching. The concept 'copy' is clarified and sharpened. Data have been collected regarding the performance of several contemporary services to find copies of an image, namely TinEye, Google and Yandex.
Conclusions - The reported findings have revealed significant differences in the number of copies discovered and in the precision of the search results. Therefore, all this may be helpful for users / practitioners to select and apply an appropriate image search service to reveal copies of an image and even information that is related to the image.

Keywords: TinEye, Google, Yandex.

Introduction

Finding copies of an image

Starting from a known image, it can be interesting to find duplicates (exact copies) of that image or to find near-duplicates (images that have elements in common with that image, but which are not identical).
How to detect as many as possible near-duplicates among billions of images in an efficient way is a challenging problem for scientists and engineers. Several approaches have been developed (briefly reviewed by Wang et al., 2010). An investigation showed that about 22 % of the images on the WWW had near-duplicates, and about 8 % had more than 10 (Wang et al., 2010).
One of the suitable search services on the Internet can be applied to find copies of an image. The following gives a few scenarios.

Finding copies of an image that you have created

Starting from an image that you have created or of an image affiliated with your organization, you may want to find derived images.

- In general, you may want to track how such an image is used / reused. This can reveal that your image has been copied and reused without asking permission. This means copyright infringement, unless your image has been published with an explicit message that it is free from copyright for some or all applications. This is important, for instance for professional photographers and designers of infographics (see for instance Tyrrell, 2015 or 2016).
- Even a whole web site that you have created can be copied partially or completely and can be republished on another site. This can be checked of course by a classical text

search, so that this may seem irrelevant to be mentioned in this context. However, such a text search will probably not detect a translated version of your document, while a search by image can reveal that one of your documents has been copied, translated and (re)published somewhere on the WWW. This is not purely hypothetical but realistic; I have experienced this personally and detected this by using a search by image.

- In a more positive and constructive way, finding copies or derived versions of your image allows you to assess the impact of that image on a worldwide audience, which reflects the 'value' of that image. The following are examples. Curators or owners of a collection of objects can assess the impact and reuse of photos of the physical objects in their collection, on a worldwide scale; more specifically this works well for photos of objects and landscapes (Nieuwenhuysen, 2013), for images of paintings from a museum collection (Kirton & Terras, 2013, 2014) and for digitized photographs from a collection in a university (Kelly, 2015). Scientists, academics and researchers (see for instance Kousha et al., 2010) as well as photographers, artists and designers can assess the impact and reuse on the WWW of images that they have created.
- Your image can form a component of a different image on the WWW, but it does not stop there, as digital images can also be used in the more tangible, physical reality, for instance in an illustration of a printed publication; earlier I have reported some concrete cases of digital images used on a book cover, on the cover of a music album, and on posters, all detected using search by image (Nieuwenhuysen, 2013).

Finding other versions of an interesting image

You can know an image that you consider as interesting, but

- that you did not create, so that you do not know the original version,
- that can be the original version or a derived version,
- that may give no clue about the creator/author/owner/publisher and the related copyright.

Then searching by image may yield desirable information:

- You may find other versions of that image, which are more suitable for your application and need; for instance, a version closer to the original image at a higher level of resolution or quality or integrity.
- You may find the copyright status of the image.
- You may find he author(s) or publisher or copyright holder, which can be useful to obtain more information or to discuss possible copyright linked to the image.
- You may find a copy of the image, plus also its location on some WWW page and WWW site, which can provide you with more information about the contents of the image plus related information.
- A search by text suffers from the fact that documents about the topic of your interest can be written in another language, so that they are not found by your simple search with words in only one or a few languages. On the other hand, a search by image does not suffer from this complication and difficulty. Therefore, a search by image with a source image that reflects your specific interest may be successful to reveal interesting, relevant documents on the web, independent of the text language.
- You may get a better view on the authenticity of an image that illustrates and supports the message / contents / claims of a document. The image has perhaps been copied from another site, from another context and perhaps it has even been modified / changed / doctored, to support the text, the claims of the author of the document. This

phenomenon becomes more important, because news is distributed and consumed increasingly through more informal and less expensive channels such as social media on the WWW; however, this is accompanied by an increasing number of unverified / fake / false news stories and claims. Verifying images and contents or --in other words-- detecting such untrustworthy information sources is considered now quite important, so that efforts are made to make the process of detection smooth, fast and user friendly (see for instance Elkasrawi et al., 2016; Goel, 2016).

- Related to authenticity, you may check the reality of a portrait of a person, for instance a portrait used as profile photo on a social networking site. In other words, search by image can be applied to authenticate people.

Images in information retrieval

Investigating information retrieval systems that involve images is motivated and justified by the fact that the number of images available is increasing rapidly, in parallel with the decreasing technical difficulties and costs that are related to

- digitization of hard-copy images,
- digital cameras and photography,
- publication / distribution of images through the WWW and even social interaction associated with images.

Consequently, billions of images are now available on the web. This evolution makes images more important as carriers of information. So, images have also become more important in information retrieval and discovery. More concretely, digital information systems on the Internet and WWW can help in finding ‘copies’ of an image, as well as images that are needed, or in elucidating the origin or even the contents of a known image.

A recent research paper deals with the precision of search for images on the Web with a classical text query and offers a review of searching for images with a text query (Uyar & Karapinar, 2016).

Search by image through the Internet

Finding duplicates or near-duplicates of an image is possible by applying a relatively new method of database searching, in which each query consists not of text but of an image file. The search results lead to images on the WWW and to related texts. Terms used for this method are

- Search(ing) by example
- Reverse image search(ing)
- Backwards image search(ing)
- Inside search(ing)
- Reverse image lookup = RIL
- Query by Image Content = QBIC
- Content-based information retrieval = CBIR

The company Google offers several search services and since 2011 also a search service of this type, named "reverse image search"; as Google products are popular, this is a term used in many texts about this search method. The author of this contribution prefers the term "search by image” because it is more simple and clear.

The state of the art in visual information retrieval and in particular of content-based information retrieval (CBIR) has been sketched (Marques, 2016) and has been reviewed in more detail (Tyagi, 2017).

Objectives / Aims

This contribution deals with search by image to find copies of your source image. Earlier I have reported on an investigation of finding 'copies' of an image, by using

1. the pioneering and specific system and service TinEye to search by image at https://www.tineye.com , and
2. the more recent, similar service offered by the big, popular and successful company Google since 2011 at https://images.google.com/ (Nieuwenhuysen, 2013).

Afterwards, a few additional general search services that are significant in terms of their technical capacity and number of users have started to offer also some search service to search by image. Concretely, Yandex at https://yandex.com/images/ offers searching for images in the classical way with a text query and also searching by image. The user interfaces offered by Google and Yandex are similar. Yandex is mainly active and popular in Russia: "In Russia the most popular search engine is Yandex, it shares 60.4% of the market, while Google.ru has 26.2%." (Paananen, 2012).

After a first reported investigation (Nieuwenhuysen, 2013), the systems developed by TinEye and Google have evolved, and the WWW has grown significantly in recent years. TinEye stated on their user interface, that their search service deals with about 20 billion images. Numerous documents on the WWW and some in printed format only mention or deal with search by image and some applications, but only mention in a superficial way some of the services that are available on the WWW (see for instance Tyrrell, 2015 or 2016). Furthermore, an application of search by image, can deliver various categories / types of results, as outlined below in the context of Fig. 1, but most of these publications do not make a clear distinction between these categories and write only in a general way about "similar" images (for instance Adrakatti et al., 2016).

Users and potential users of search by image may and should be interested in the performance of the respective available search services, qualitatively and quantitatively, which can be expressed in terms related to the concepts that are classical in statistics in general (see for instance Stats-Lab Dublin, 2013) and more particularly also in the theory of information retrieval, namely recall and precision. However, I am not aware of a published report of a comparative assessment of contemporary search services that offer search by image. Therefore, I have investigated / assessed / compared the performance of search by image, as offered these days by TinEye, Google, and Yandex.

Methods

In each search query by image, one source image was submitted from the computer disk as query to the chosen search service.

The images used in the tests were photos in colors.

These photos have been present on the Internet on a public access web server already for several years, as part of a classical, simple web page and site and not as part of a container file such as a PDF or Word file or of a database, so that they can be accessed and harvested/copied by human users as well as by automated harvesting robots, without difficulty, in standard ways.

Each image file was submitted as a query, keeping the same name of the file as on the Internet. In earlier reported investigations of the capacity of Google to reveal images that are semantically similar, the meaningful, informative file names were changed to neutral, insignificant names, to avoid providing / revealing information about the contents / subject of the image to the search service (Nieuwenhuysen, 2014). This is not needed in this investigation. On the contrary, by providing the file name as present on the Internet, I hope / assume that this may help the search services to reveal the original image on the WWW; in

the tests, I noted if the search service did include a link in the search results to the original web site.
Each test with a source image was executed on the same day with the various search services, to avoid measuring differences that are due to the evolution of the Internet and / or the search service. Only exceptionally a search has been repeated a few times to test the stability of the performance of the search service over time, as mentioned in the text below.
Each link that was given as a search result by a search service was manually activated and it was checked if the link leads indeed to retrieval of a document from the WWW. Most links do lead to a fetched document as expected; then this fetched document was inspected, to check if a 'copy' of the image was indeed included.
In this investigation, the relevance of a retrieved image / document / file is judged with the simplest bimodal model (relevant or irrelevant, yes or no, 0 or 1). More concretely, relevant means here that a 'copy' of the source image is indeed included, while the concept of 'copy' is clarified in a dedicated chapter of this paper. This simple judging is considered as sufficient here and has also been used in earlier similar investigations (see for instance Kousha et al., 2010; Nieuwenhuysen, 2013; Kelly, 2015; Adrakatti et al., 2016). This approach is more straightforward than in an investigation of the retrieval of semantically similar images, because then the results may contain images that are relevant in various aspects and at various levels, (see for instance Nieuwenhuysen, 2014, 2016a, 2106b).

Results and Discussion

Search by image can yield various types of search results

Investigating the applications of search by image has resulted in understanding that retrieved images (plus related information) should be considered as belonging to various types / classes / categories, as sketched in Fig. 1.

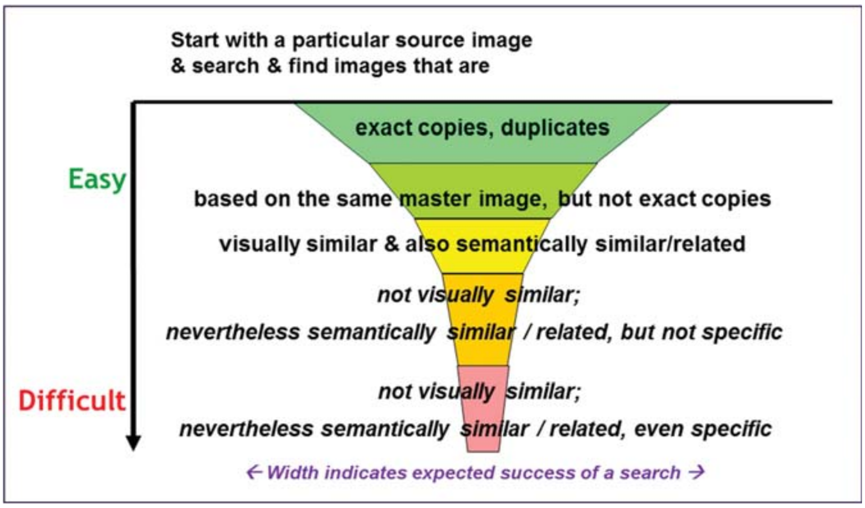


Figure 1. Search by image can yield various types of retrieved images.

The concept 'copy' of an image

This contribution deals with finding 'copies' of an image by using reverse image search. This is a relatively easy task for a computer system, as shown in Fig. 1. In the text above, the word 'copy' is used loosely, in an intuitive way. In several more superficial documents about search by image, the words "copy" and "similar" are used without definition. One result of this investigation is that it has become clear that the concept of 'copy' is not straightforward and that a sharper, structured view has been formed, as shown in Figure. 2. In this context, we mean by 'copies' those images that are related to an original / authentic master image file and thus to each other. The source image used in a search query, as well as the relevant images that are found using search by image belong to the image files represented in Figure. 2, where each circle represents an image file that includes elements from the master image on top.

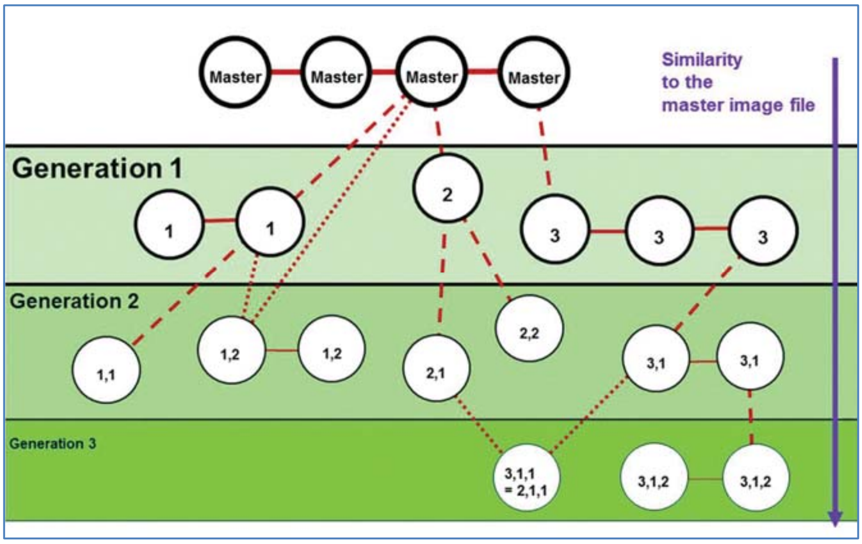


Figure. 2. The image files and their relations that form the targets of search by image, as investigated in this paper, here shown in the form of an example that is limited to three generations. A full connecting line connects exact-duplicate image files. A dashed connecting line connects an image file to a parent image file of a previous generation from which it has been derived or from which it has at least inherited some elements. A connecting line that consists of only dots indicates the less common relation between an image file and more than one image file of a previous generation, namely to several parent image files from which it has inherited some elements. (This is not only possible in theory, but also in reality; indeed, I have found an image on the web, which includes two of the images that I had created and published on my website; this example is shown in the case studies reported below.).

In view of the considerations above, not the words copy / copies, but 'copy' / 'copies' is used in this paper, to indicate in a brief and simple way the image files that are related as clarified above.

User interfaces and usability for search by image, offered by the various search services

TinEye:

This is an exceptional web search service, as it offers only search by image, while other search services offer several types of web search. Consequently, the user interface is simple and finding out how to start a search is easy and fast. TinEye offers the user various ways to rank the results: Best match / Most changed / Biggest image / Newest / Oldest.

This is a useful feature. For instance,

- to discover efficiently images that are not pure duplicates, you can select the option "Biggest change";
- to discover efficiently images that are closer to the original / authentic master image, you can select the option "Oldest".

This option is not offered by the other search services investigated.

Google
This company offers several ways to search the web. Consequently, it is less easy to find out where and how to start a search by image than with TinEye. Search by image is possible on the Google web page that offers also the more classical search for images with a text query. Here, the user interface is like the one offered by TinEye.

Yandex
This general search system offers search by image in their module / section for image search, like Google, but the search form is somewhat hidden and should be activated explicitly by the user.

Finding the original image on the WWW
The tests made in this investigation have shown that none of the search by image systems yield for each test case a link to the original source image on the WWW, even in searches that generate numerous links to 'copies'.
This agrees with my earlier findings: then TinEye gave almost no links to the original images on the WWW, while Google gave a link to the original image in most but not in all tests (Nieuwenhuysen, 2013). Also, in an investigation of reuse of digitized photos, using TinEye and Google, the original image uploads were not returned as results, except in one instance (Kelly, 2015).

Image 'copies' in complicated formats can be detected
The tests carried out in this investigation have demonstrated that search by image can even reveal / find / detect a 'copy' of a source image that is not simply present on the web as an individual image file in one of the classical formats (gif, jpg, png), but even when the image is embedded in a more complicated file format. More concrete examples:

- Google revealed a 'copy' as a part of a file in the format PDF. In an investigation of the reuse of digitized photos, Google also revealed a PDF file that includes reused images (Kelly, 2015). These observations agree with the message that at least since 2015 Google extracts images also from harvested PDF files, so that also these can be indexed and used in the search services offered by Google, which involve images (Chitu, 2015).
- Google revealed a web page that showed mainly a video including an image derived from the source image as part of a larger image.
- Yandex revealed a 'copy' in a series of slides.

So, it has become even more productive, efficient and attractive to apply searching for images.

Number of image 'copies' found by the search actions

Several empirical case studies (tests) have yielded the results that are summarized in Table1.

Table 1. Measured Number of Relevant Results.

Name of test case	*TinEye*	*Google*	*Yandex*
Bwoom	**1**	**1**	0
Congo mask	1	**11**	2
Head 1	0	**4**	0
Head 2	0	**1**	**1**
Heads	0	**5**	4

The best performances are indicated by bold numbers.
These data together with the earlier comparison of TinEye and Google (Nieuwenhuysen, 2013) can be summarized in the form of the following ranking of the performance of the search services concerning the number of image 'copies' revealed: 1. Google, 2. Yandex, 3. TinEye.
In a first attempt to use the image matching capability of TinEye as a potential tool for informetric analysis, researchers reported that "It seems likely that TinEye could only find a small fraction of the total number of copied images but it is not clear how small this fraction would typically be." (Kousha et al., 2010). All the test cases of this newer investigation show that nowadays still only a small fraction of the existing 'copies' are revealed by TinEye. In my earlier comparative assessment, TinEye gave also a smaller recall than Google (Nieuwenhuysen, 2013). In an investigation of the reuse of images of famous paintings, this was also observed: "Google Images returns a significantly larger number of results for each search than TinEye does." (Kirton & Terras, 2013, 2014). In an investigation of the reuse of digitized photos, "TinEye did not find a single reused image among the set", while Google Image Search did reveal some reuse on the web (Kelly, 2015).
A recent report mentions only one test of the performance of Google, TinEye and Yandex, using only one source image (Adrakatti et al., 2016). That publication does not refer to any earlier, published test case, and does not make an explicit distinction between retrieval of 'copies' and semantically similar images. Anyway, in the test, Google delivered many more so-called "relevant results" than TinEye and Yandex. So that single test agrees also with the findings reported here.

Precision of search results

Another aspect of performance evaluation in information retrieval is the "precision". This is also used in visual search, as reviewed recently (Tyagi, 2017, section 4.3). In this type of application, the precision is less important than the number of 'copies' found, but the measured values are given anyway as follows.
The empirical case studies (tests) of this investigation have yielded the results that are summarized in Table 2.

Table 2. Measured Precision of Search Results.

Name of test case	***TinEye***	***Google***	***Yandex***
Bwoom	**1/1 → 100%**	**1/1 →100%**	Non-Applicable
Congo mask	**1/1 →100%**	11/13 → 85%	**2/2 → 100%**
Head 1	Non-Applicable	**4/4 → 100%**	Non-Applicable
Head 2	Non-Applicable	**1/1 → 100%**	1/2 → 50%
Heads	0/1→ 0%	**5/5 → 100%**	4/18 → 22%

If a search leads to zero relevant results, then the concept of "precision" makes no sense and is Non-Applicable, as mentioned in this table.
The best performances are indicated in bold numbers.
As a summary this gives the following ranking of the performance of the search services, in terms of precision: 1. Google, 2. Yandex, 3. TinEye.
This ranking is identical to the ranking above in terms of recall. This is not a priori expected; the tests had to be carried out to come to this conclusion. For instance, changing algorithms in the search system may increase recall (which is often desirable), but can also lead to a decrease in precision (which is not desirable); so, the algorithms are developed to result in a compromise that is acceptable / satisfying for the user. The greater recall of Google is not

accompanied by a smaller precision; therefore, the greater recall is probably due to more extensive harvesting and coverage of documents on the web.

Conclusions

Finding 'copies' of a known image, using search by image

To find 'copies' of a known image, the method of search by image is suitable. Furthermore, the leading services offer a friendly user interface, deliver results quite fast, and (at least in our case studies) the best performing search service gave results with a high precision; all this is convenient, satisfactory and time saving.

Choosing a search service to find 'copies' of an image

To choose one of the services for your searches, you can consider criteria such as the usability, the number of 'copies' found and the precision of search results. This investigation leads to the following general ranking:
1. Google
2. Yandex
3. TinEye.

Google web services are not available in some countries; then the alternative web search services can be considered.

References

Adrakatti, A .F., Wodeyar, R. S., & Mulla, K. R. (2016) Search by Image: A Novel Approach to Content Based Image Retrieval System, *Int. J. Library Science*, ISSN 0975-7546, *Volume 14*; Issue No. 3. Available free of charge from https://www.researchgate.net/publication/305683970_Search_by_Image_A_Novel_Approach_to_Content_Based_Image_Retrieval_System

Chitu, Alex (2015) Google Indexes Images from PDF Files. August 9, 2015. In the blog Google Operating System - Unofficial news and tips about Google http://googlesystem.blogspot.be/2015/08/google-indexes-images-from-pdf-files.html

Elkasrawi, Sarah; Dengel, Andreas; Abdelsamad, Ahmed, & Bukhari, Syed Saqib (2016) What You See is What You Get? Automatic Image Verification for Online News Content. 12th IAPR Workshop on Document Analysis Systems (DAS), 11-14 April 2016, Electronic ISBN: 978-1-5090-1792-8, DOI: 10.1109/DAS.2016.75, Publisher: IEEE.

Goel, Sonal, Sachdeva, Niharika, Kumaraguru, Ponnurangam, Subramanyam, A.V. & Gupta, Divam (2016) PicHunt: Social Media Image Retrieval for Improved Law Enforcement. arXiv:1608.00905 [cs.MM]

Kelly, Elizabeth Joan (2015) Reverse image lookup of a small academic library digital collection. Codex: the Journal of the Louisiana Chapter of the ACRL, ISSN 2150-086X, Vol. 3, issue 2, 80-92. Available free of charge from http://journal.acrlla.org/index.php/codex/article/view/101

Kirton, Isabella & Terras, Melissa (2013) Where Do Images of Art Go Once They Go Online? A Reverse Image Lookup Study to Assess the Dissemination of Digitized Cultural Heritage. In MW2013: Museums and the Web 2013, The annual conference of Museums and the Web | April 17-20, 2013 | Portland, OR, USA. Available free of charge from http://mw2013.museumsandtheweb.com/paper/where-do-images-of-art-go-once-they-go-online-a-reverse-image-lookup-study-to-assess-the-dissemination-of-digitized-cultural-heritage/

Kirton, Isabella & Terras, Melissa (2014) Digitization and Dissemination: A Reverse Image Lookup Study to Assess the Reuse of Images of Paintings from the National Gallery's Website. *J. Digital Humanities*, Vol. 3, No. 1, Spring 2014. Available free of charge from http://journalofdigitalhumanities.org/3-1/reverse-image-lookup-paintings-digitisation-reuse/

Kousha, Kayvan; Thelwall, Mike, & Rezaie, Somayeh (2010) Can the impact of scholarly images be assessed online? An exploratory study using image identification technology. JASIST, Volume 61, Issue 9, September 2010, 1734–1744, DOI: 10.1002/asi.21370

Marques, Oge (2016) Visual Information Retrieval: The State of the Art. *IT Professional, Volume: 18*, Issue: 4, DOI: 10.1109/MITP.2016.70

Nieuwenhuysen, Paul (2013) Search by Image through the WWW: an Additional Tool for Information Retrieval. Full text published in proceedings of the international conference on Asia-Pacific Library and Information Education and Practices = A-LIEP 2013 "Issues and challenges of the information professions in the digital age" held at Pullman Khon Kaen Raja Orchid Hotel, in Khon Kaen City, Isan, Thailand, 10-12 July 2013 [online] http://aliep2013.com/index.php/table-of-contents http://aliep2013.com/images/download/pdfs1/PaperNo41.pdf

Nieuwenhuysen, Paul (2014) Search by image through the Internet: applications and limitations. In Libraries in the Transition Era: New Space –New Services – New Experience. The Proceedings of the Seventh Shanghai International Library Forum, organized by the Shanghai Library, in Shanghai Library, China, 9-11 July 2014, http://www.libnet.sh.cn/silf2014/english/index.htm, Shanghai : Shanghai Scientific and Technological Literature Press, http://www.sstlp.com, 476 pp, ISBN 978-7-5439-6289-7. 145-155.

Nieuwenhuysen, Paul (2015) Search by image through the Internet: an additional method to find information. In Transforming Libraries and Librarianship, Delhi : KBD Publication, 502 pp. Edited by Sanjay Kataria, John Paul Anbu, Shri Ram, Nirmal Kumar Swain, Naresh Singh Bhandari, 2015, ISBN: 978-81-907999-6-6, 179-194.

Nieuwenhuysen, Paul (2016a) Information discovery on the Internet, using a search query that consists of text and an image. in "Libraries: Enabling Progress", Proceedings of the 8th Shanghai International Library Forum, organized by the Shanghai Library, in Shanghai Library = SILF 2016, China, 6-8 July 2016 <http: //www.libnet.sh.cn /silf2016/English /index.htm> Shanghai : Shanghai Scientific and Technological Literature Press, http://www.sstlp.com, ISBN 978-7-5439-7063-2, 442 pp., 153-162.

Nieuwenhuysen, Paul (2016b) Information discovery on the Internet, using a search query that consists of text & an image. in A-LIEP 2016, proceedings of the 7th Asia-pacific Conference on Library and Information Education and Practice, in University of Nanjing, Nanjing = Nanking, Jiangsu province, China, Edited by Jianjun Sun, Qinghua Zhu, Christopher Khoo Shiyan Ou.. http://aliep2016.nju.edu.cn, PDF file, 99-112, Available free of charge from http://aliep2016.nju.edu.cn/files/A-LIEP%202016%20Proceedings.pdf

Paananen, Anna (2012) Comparative Analysis of Yandex and Google Search Engines. Helsinki Metropolia University of Applied Sciences, Master's Degree Information Technology, Master's Thesis, 26 May 2012. Available free of charge from https://www.theseus.fi/handle/10024/46483

Reilly, Michele, & Thompson, Santi (2016) Reverse Image Lookup: Assessing Digital Library Users and Reuses, *Journal of Web Librarianship*, DOI: 10.1080/19322909.2016.1223573

Stats-Lab Dublin (2013) Accuracy, Recall and Precision. Video available free of charge from https://www.youtube.com/watch?v=VPZiJGNX4_s

Tyagi, Vipin (2017) Content-Based Image Retrieval: Ideas, Influences, and Current Trends, Springer, Singapore, 378 pp., DOI https://doi.org/10.1007/978-981-10-6759-4, Print ISBN 978-981-10-6758-7, Online ISBN 978-981-10-6759-4

Tyrrell, Katherine (2015 or 2016) How to do a reverse image search. Available free of charge from http://www.artbusinessinfo.com/how-to-do-a-reverse-image-search.html

Uyar, Ahmet, & Karapinar, Rabia (2016) Investigating the precision of Web image search engines for popular and less popular entities. *Journal of Information Science*, April 27, 2016, DOI: 10.1177/0165551516642929

Wang, Xin-Jing, Zhang, Lei, Liu, Ming, Yi Li, & Ma, Wei-Ying (2010) ARISTA Image Search to Annotation on Billions of Web Photos, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). http://ieeexplore.ieee.org/abstract/document/5540046/ DOI: 10.1109/CVPR.2010.5540046

Article DOI: 10.22032/dbt.39356 Proceedings DOI: 10.22032/dbt.39296

Citation Analysis of Doctoral Theses of Earth Science accepted by the Manipur University during 1989-2011

[1]Bobby Phuritsabam and Gyanajeet Yumnam[2]

[1] *bobbyphster@gmail.com*
Assistant Professor (S-2)
Department of Library & Information Science,
Manipur University, Canchipur, 795 003, Manipur, India

[2] *ygyanajeet@gmail.com*
Student
Department of Library & Information Science,
Manipur University, Canchipur, 795 003, Manipur, India

Abstract
The study is to analyse the bibliographic citations which have been cited in the doctoral theses of the research scholars in Earth Science, Manipur University, during the year 1981-2011.The study is to determine the principle forms of literature used in theses; to determine country-wise, language-wise and subject-wise distribution of literature; and the applicability of Bradford's law of scattering to the pattern used in Earth Science. The bibliographical references were used at the end of each chapter and end of the doctoral theses were taken as the source of data for the study. The collected data were analysed and interpreted with table's format. Analyses of several parameters like country-wise, language-wise, subject-wise distribution of literature and the applicability of Bradford's law of scattering. Interaction with faculties of concern department was carried out. It was found that researcher emphasized more on journal source, they depend literature from UK, Netherland & USA; English was the predominant language; the total 2273 citations of journals are scattered primarily among 45 subjects and 37.3% is connected in one major subject fields; 9 (Nine) Core journal which were the source journals leads the majority of the citations.

Keywords: Citation Analysis, Bradford Law, Earth Science, country-wise, language-wise, subject-wise.

Introduction

Citation analysis is one form of Bibliometrics study. It makes use of bibliographic references which are vital part of the primary scientific communication. The technique of citation analysis involves the process of collection, counting and analysis and interpretations of citations given in various types of literature and, thereby, helps in identification of significant sources of information, individuals, institutions and other aggregates of scientific activities. Citation analysis as a tool is used to identify the core references in a subject by counting the citations appended at the end of each scientific article. The author of a paper customarily presents references as authentic source of information having research value or to substantiate the point of view of ideas expressed in the cited paper. Analysis of cited papers is used as a measure of impact of individual articles, periodicals, authors, etc. and has become accepted practice in almost all scientific communications and is a well-established part of information research.

Review of Related Literature

Kapoor (1984) made the study under the title Citation analysis of Earth Science literature. In his finding features a ranking list of 82 journals representing 76% of total citations out of the 833 title cited. It was noted that as many as 422 journals had been cited only once. It also gives country-wise and subject-wise distributions of the journals in the ranked list. Rahman and Bhattacharya (2012) investigated 162 Ph.D. theses in Botany submitted to North Bengal

University during the period 1987 to 2007, The study revealed that journals appeared to be the most preferred sources of information contributing the highest number of citations (72.87%), followed by books (13.33%), seminar/conference (4.62%) and (2.85%)). The country wise scatterings of citations showed that India occupied first position with 21.13% of the total citations, followed by USA (19.79%) and UK (7.50%).Ravichandran, Sivaprasad and Manoharan (2014) carried out the bibliometric citation analysis of four Ph.D. theses in Library and Information Science at Bharathidasan University, Tiruchi during 2008-12. In all 791 citations were analyzed. The authors noted 68.9% articles published in journals, followed by electronic resources and others with a low percentage. The researchers cited only two languages, English (98.7%) and Chinese (1.26%). For country wise citations, the authors noted USA with 33% articles, UK 27.43%, India 25.41%, Netherland 8.34% and remaining 12 countries the articles were less than 2%. The journal distribution as per Bradford's law was noted to be 5:24:103. They also noted 85% of journal citations are 15 years old. Singh and Bebi (2013) carried out a citation analysis of Ph.D. theses in Sociology submitted to university of Delhi during the year 1995-2010. The study is based on the 5766 citations taken out from 25 Ph.D. theses of Sociology. The country wise-scattering of citations reveals that 2536 (45.52%) citations were from India and it was followed by USA and UK. Kushkowski (2003) reported on a result of a study of over 9,100 citations from 629 masters and doctoral theses written between 1973 and 1992 at a Midwestern Land Grant University. The study suggests that graduate students writing theses favour current research. The study shows distinct trend in graduate students' citation pattern.

Scope

The present study is an attempts to discover and scanning of the Ph.D theses of the Earth Science department during the period 1989-2011; present study attempts on the pattern of information use by researchers in the field of Earth Science; and area of study has confined to the Ph.D theses of Earth Science declared by Manipur University for the degree of Philosophy.

Objectives

The objectives of the study are to:

(i) Study the Major source of information used in Earth Science and to determine the principal forms of literature used in theses by Earth Science doctoral students;
(ii) To determine country-wise, language-wise and subject-wise distribution of literature used by doctoral students.
(iii) Applicability of Bradford's law of scattering to the pattern used in Earth Science.

Methodology

The information were collected from Manipur University which was submitted during the period of 1989-2011 in Earth Science department were recorded the task of identifying and recording information about the individual citation. The bibliographical references which were used by the researchers for completing the theses at the end of each chapter and end of the doctoral theses were taken as the source of data for the study. The collected data was classified, tabulated, presented, analyzed and interpreted with the help of tables. The study presents analysis of several parameters like forms of literature, Country-Wise, LanguageWise, Subject-Wise and finally a list of core journals was compiled and prepared on the basis of highly cited articles of the journals in Earth Sciences.

Limitations of the study

(i) The scope of the study is restricted to the research work thesis conducted by the Department of Earth Science, Manipur University.
(ii) The analysis of the study base on journal sources.
(iii) The present study carried that available doctoral theses in University libraries and departments which taken up for detailed investigation.

Analysis of Data

Data Collected from the Thesis

A total number of 36 (thirty six) theses submitted to Manipur University during the period of 1989-2011 in Earth Science department were examined. They gave a total of citations, which is an average of 100.45 citations per thesis. The bibliography cited was sorted out according to theirs formats, i.e. books, journals, proceedings, thesis, report, records and others. Journals have a highest ranking to be cited by the researchers of Earth Science. The data collected were presented in table's aid percentages under various heading.

Table 1. Year-Wise Distribution of Thesis

Sl. No.	*Year*	*Rank*	*Thesis*	*Percentage %*
1.	2002-2007	1st	12	33.33
2.	1990-1995	2nd	10	27.78
3.	2008-2011	3rd	7	19.44
4.	1996-2001	4th	5	13.89
5.	1984-1989	5th	2	5.56
			Total = 36	100

From Table-1 it reveals that during 2002-2007, the total thesis was 12 and this period was the most producing Ph.D thesis as compare to other years and followed by the years 1990-1995, 2008-2011, 1996-2001 and 1984-1989 respectively.

Bibliographic Forms of Cited documents

The following Table-2 gives us details about the different documents cited by the scholars.

Table 2. Source-Wise Distribution of Citation

Sl. No.	*Bibliographic Forms*	*No. of Citations*	*%*	*Cumulative No. of Citations*	*Cumulative Percentage (%)*	*Rank*
1.	Books	1878	30.72	1878	30.72	2nd
2.	Journals	2273	37.18	4151	67.89	1st
3.	Proceedings	872	14.26	5023	82.16	3rd
4.	Reports	371	6.07	5394	88.22	4th
5.	Records	245	4.01	5639	92.23	5th
6.	Thesis	196	3.21	5835	95.44	6th
7.	Government Publication	161	2.63	5996	98.07	7th
8.	Maps	15	0.25	6011	98.32	9th
9.	Misc.	103	1.68	6114	100.00	8th
	Total	6114	100.00			

It is observed that from Table-2, 37.18% of the sources cited by the doctoral researchers are journals followed by books (30.72%), Proceedings (14.26%), Reports (6.07%), Records

(4.01%), Thesis (3.21%), Government (2.63%), Miscellaneous (1.68%) and Maps (0.25%) respectively.

Country-Wise Distribution of Journal

Table 3. Country-Wise Citations in Earth Science

Sl. No.	*Name of the Country*	*No. of Citations*	*Cumulative Citations*	*Percentage (%)*	*Cumulative Percentage*
1.	UK	700	700	30.796	30.796
2.	Netherlands	365	1065	16.058	46.854
3.	USA	353	1418	15.530	62.384
4.	India	234	1652	10.294	72.679
5.	Nigeria	199	1851	8.754	81.434
6.	Germany	115	1966	5.059	86.493
7.	Kenya	42	2008	1.847	88.341
8.	Australia	35	2043	1.539	89.881
9.	Switzerland	29	2072	1.275	91.157
10.	Canada	27	2099	1.187	92.344
11.	USSR	25	2124	1.099	93.444
12.	Italy	16	2140	0.703	94.148
13.	Thailand	15	2155	0.659	94.808
14.	Austria	12	2167	0.527	95.336
15.	Poland	8	2175	0.351	95.688
16.	Egypt	7	2182	0.307	95.996
17.	Germany	7	2189	0.307	96.304
18.	Japan	7	2196	0.307	96.612
19.	Finland	6	2202	0.263	96.876
20.	Ghana	6	2208	0.263	97.140
21.	Spain	5	2213	0.219	97.360
22.	Estonia	4	2217	0.175	97.536
23.	Indonesia	4	2221	0.175	97.712
24.	Lithuania	4	2225	0.175	97.888
25.	Malaysia	4	2229	0.175	98.064
26.	South Africa	4	2233	0.175	98.240
27.	Sweden	4	2237	0.175	98.416
28.	Greece	3	2240	0.131	98.548
29.	Nepal	3	2243	0.131	98.680
30.	Pakistan	3	2246	0.131	98.812
31.	Russia	3	2249	0.131	98.944
32.	Slovenia	3	2252	0.131	99.076
33.	Czech Republic	2	2254	0.087	99.164
34.	France	2	2256	0.087	99.252
35.	Iran	2	2258	0.087	99.34

36.	Ireland	2	2260	0.087	99.428
37.	Israel	2	2262	0.087	99.516
38.	Korea	2	2264	0.087	99.604
39.	Italy	2	2266	0.087	99.692
40.	Germany	1	2267	0.043	99.736
41.	Hungary	1	2268	0.049	99.780
42.	Myanmar	1	2269	0.049	99.824
43.	New Zealand	1	2270	0.043	99.868
44.	Tanzania	1	2271	0.043	99.912
45.	Turkey	1	2272	0.043	99.956
46.	UAE	1	2273	0.043	100

In this section, all the citation categorized to their country of origin to find out the most productive countries in the literature of Earth Science subject. The cited documents were analysed according to their country of origin. Table-3 indicates that 30.79% of the cited documents were from UK and the rest were from Netherlands (16.05%), USA (15.53%), India (10.29%) and remaining i.e. less than 9% were from other 42 countries. From this it can be inferred that the researchers in earth science depend much on literature emanating from UK, Netherland and USA.

Language-Wise Distribution of Journal Citations

Table 4. Language-Wise Distribution of Cited Journal Articles

Sl. No.	*Language*	*No. of Citations*	*Cumulative Citations (%)*	*Percentage (%)*	*Cumulative Percentage*
1.	English	2259	2259	99.384	99.384
2.	German	7	2266	0.307	99.692
3.	Russian	3	2269	0.131	99.824
4.	Slovene	3	2272	0.131	99.956
5.	French	1	2273	0.043	100

Further, the citations were analysed according to their language. The language-wise scattering of cited documents showed that English was the predominant language (99.384%) and the uses of other foreign language materials i.e. Russian (0.131%), Slovene (0.131%) and French (0.043%) were substantially lower. German language documents accounted for only 0.307% of citations (Table-4).

Subject-Wise Distribution of Journal

Table 5. Subject-Wise Distribution of Journal Citations

Sl. No.	*Subjects*	*Total No. of citations*	*Percentage of Citation (%)*	*Cumulative Citation*	*Cumulative Percentage (%)*
1	Geology	848	37.308	848	37.308
2	Petrology	247	10.867	1095	48.175
3	Geography	213	9.371	1308	57.546
4	General Science	152	6.687	1460	64.233
5	Environmental Sciences	117	5.147	1577	69.380
6	Sedimentology	73	3.212	1650	72.592

7	Hydrology	71	3.124	1721	75.715
8	Geophysics	44	1.936	1765	77.651
9	Others (Related Subjects)	43	1.892	1808	79.543
10	Remote Sensing	41	1.804	1849	81.347
11	Mining	40	1.760	1889	83.107
12	Ecology	34	1.496	1923	84.602
13	Medicine	34	1.496	1957	86.098
14	Agriculture	31	1.364	1988	87.462
15	Marine Science	31	1.364	2019	88.826
16	Waste Management	26	1.144	2045	89.970
17	Chemistry	20	0.880	2065	90.850
18	Regional Studies	20	0.880	2085	91.729
19	Environmental Engineering	19	0.836	2104	92.565
20	Petroleum Technology	19	0.836	2123	93.401
21	Solar and Wind	15	0.660	2138	94.061
22	Atmospheric Science	14	0.616	2152	94.677
23	Mineralogy	14	0.616	2166	95.293
24	Palaeontology	13	0.572	2179	95.865
25	Toxicology	13	0.572	2192	96.437
26	Geomorphology	11	0.484	2203	96.921
27	Climatology	6	0.264	2209	97.185
28	Forestry	6	0.264	2215	97.449
29	Psychology	6	0.264	2221	97.713
30	Civil Engineering	5	0.220	2226	97.933
31	Soil Science	5	0.220	2231	98.153
32	Biology	4	0.176	2235	98.329
33	Astronomy	4	0.176	2239	98.505
34	Environmental Chemistry	4	0.176	2243	98.681
35	Geochemistry	4	0.176	2247	98.857
36	Energy	4	0.176	2251	99.033
37	Botany	3	0.132	2254	99.165
38	Biochemistry	3	0.132	2257	99.297
39	Seismology	3	0.132	2260	99.429
40	Social Science	3	0.132	2263	99.561
41	Pharmacy	3	0.132	2266	99.693
42	Economics	2	0.088	2268	99.781
43	Sociology	2	0.088	2270	99.868
44	Pure and Applied Chemistry	2	0.088	2272	99.956
45	Statistics	1	0.044	2273	100.000
	Total	**2273**	**100.00**		

The Table-5 depicts the subject wise distribution of journals citations in Earth Science Ph.D Theses. The total 2273 citations of journals are scattered primarily among 45 subjects. However 37.3% is connected in one major subject fields i.e. Geology as majority of Ph.D theses under citation study related to this subject. Journal on the subject citation on Botany, Biochemistry, Seismology, Social Science, Pharmacy, Economics, Sociology, Pure and Applied Chemistry and Statistics are small in number and together which represents 0.968% of total number of citations.

Bradford's Law of Scattering and Analysis of Bradford's Zones

Bradford's law serves as a general guideline to librarians in determining the number of core journals in any given field. It states that journals in a single field can be divided into three zones, each zone containing the same number of citations (Bibliometric Laws, 2012):

a) Core journals on the given subject, relatively few in number that produces approximately one-third of all the articles.
b) A second zone, containing the same number of articles as the first, but a greater number of journals.
c) A third zone, containing the same number of articles as the second, but a still greater number of journals.

The mathematical relationship of the number of journals in the core to the first zone is a constant 'n' and the second zone of relationship is n^2. Bradford expressed this relationship as

1: n: n^2. The number of journals in each Bradford's zone can be calculated from multiplier constant k that is called Bradford constant using the formulation of Egghe (1986):

$$k = (e^{\gamma} \text{x} Y_m)^{1}/p$$

Where γ is Euler's number having value of 0.57772; Y_m is the number of citation of rank one journal; p is Bradford group or number of zones (p=3). From Table-7, the number of highest citation is 221. **e** = 2.718

T= Total Number of Journals

So that,

$$k = (2.718^{0.5772} X\ 221)^{1}/_{3} = (1.781\ X\ 221)^{1}/_{3} = 7.3286$$

$$r_0 = \frac{T(k-1)}{k^p - 1} = \frac{408(7.3286 - 1)}{7.3286^3 - 1} = \frac{2582.0688}{392.6072} = 6.5767$$

$$r_0 = 6.5767, r_1 = r_o Xk$$

Nucleus zone $r_0 = r_0\ X\ 1 = 6.5767$; **first zone** $r_1 = 6.5767X\ 7.3286 = 48.1980$; **second zone** $r_2 = r_0\ Xk^2 = 6.5767\ X7.3286^2 = 353.2240$

According to Bradford zones, thus identified in the form 1: n: n^2 in the present study, the relationship of each zone is 9: 46: 353.

Table 6. Bradford's Zones and Their Number of Journals

Zone	*No. of Journals*	*Cumulative No. of Journals*	*Journal %*	*No. of Citations*	*%*	*Cumulative No. of Citations*	*Cumulative %*
I	9	9	2.21	789	34.71	789	34.71
II	46	55	11.27	761	33.48	1550	68.19
III	353	408	86.52	723	31.81	2273	100.00
	408		100.00	2273	100.00		

The scattering of journals according to Bradford's described zones (on approximation) are:

Zone 1 (Core Nucleus) : 9 Journals with 789 Citations. **Zone 2 :** 46 Journals with 761Citations.
Zone 3 **:** 353 Journals with 723 Citations.

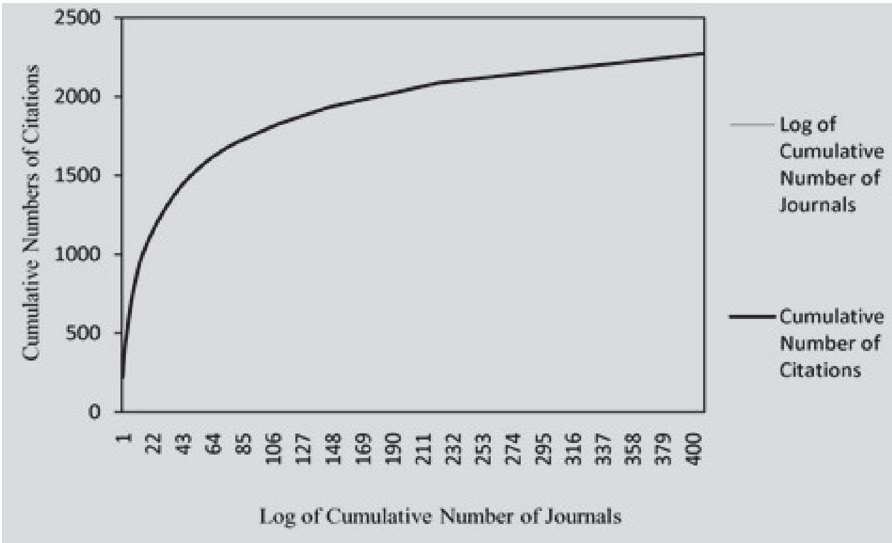


Figure 1.Bradford's Graph

Table 7. Ranking of Journals in Earth Science

Sl. No.	*Rank*	*Journal Title*	*No. of Citations*	*Cumulative Citations*	*%*	*Cumulative Percentage*	*Log N*
1	1	Journal of Sedimentary Petrology	221	221	9.723	9.72	0.000
2	2	Journal of Geology	131	352	5.763	15.49	0.693
3	3	Tectonophysics	82	434	3.608	19.09	1.099
4	4	Geochimica et Cosmochimica Acta	73	507	3.212	22.31	1.386
5	5	Journal of Geophysical Research	71	578	3.124	25.43	1.609
6	6	American Journal of Scientific Research	59	637	2.596	28.02	1.792
7	6	Himalayan Geology	59	696	2.596	30.62	1.946
8	7	Journal of Geological Society of London	48	744	2.112	32.73	2.079
9	8	Geographical Review of India	45	789	1.980	34.71	2.197
10	8	Journal of Structural Geology	45	834	1.980	36.69	2.303
11	9	Journal of Geological Society of India	42	876	1.848	38.54	2.398

12	10	Indian Journal of Earth Sciences	35	911	1.540	40.08	2.485
13	10	National Geographical Journal of India	35	946	1.540	41.62	2.565
14	11	Chemical Geology	27	973	1.188	42.81	2.639
15	12	Indian Journal of Environment and Ecoplanning	25	998	1.100	43.91	2.708
16	13	Sedimentology	23	1021	1.012	44.92	2.773
17	14	Pollution Research	22	1043	0.968	45.89	2.833
18	14	Waste Management	22	1065	0.968	46.85	2.890
19	15	Current Science	21	1086	0.924	47.78	2.944
20	16	Nature	20	1106	0.880	48.66	2.996
21	17	Geology	19	1125	0.836	49.49	3.045
22	17	Transaction American Geophysical Union	19	1144	0.836	50.33	3.091
23	18	Engineering Geology	18	1162	0.792	51.12	3.135
24	18	Quarterly Journal of Geological, Mining and Metallurgical Society of India	18	1180	0.792	51.91	3.178
25	19	Journal of Geological Society of Australia	17	1197	0.748	52.66	3.219
26	19	Journal of Geophysics	17	1214	0.748	53.41	3.258
27	20	Indian Journal of Landscape Systems and Ecological Studies	16	1230	0.704	54.11	3.296
28	20	International Journal of Remote Sensing	16	1246	0.704	54.82	3.332
29	21	3 Journals are with 21st rank	45	1291	1.980	56.80	3.367
30	22	4 Journals are with 22nd rank	56	1347	2.464	59.26	3.401
31	23	2 Journals are with 23rd rank	26	1373	1.144	60.40	3.434
32	24	4 Journals are with 24st rank	48	1421	2.112	62.52	3.466
33	25	2 Journals are with 25th rank	22	1443	0.968	63.48	3.497
34	26	3 Journals are with 26th rank	30	1473	1.320	64.80	3.526
35	27	5 Journals are with 27th rank	45	1518	1.980	66.78	3.555
36	28	8 Journals are with 28th rank	64	1582	2.816	69.60	3.584
37	29	4 Journals are with 29th rank	28	1610	1.232	70.83	3.611

38	30	11 Journals are with 30th rank	66	1676	2.904	73.74	3.638
39	31	6 Journals are with 31st rank	30	1706	1.320	75.06	3.664
40	32	30 Journals are with 32nd rank	120	1826	5.279	80.33	3.689
41	33	36 Journals are with 33rd rank	108	1934	4.751	85.09	3.714
42	34	77 Journals are with 34th rank	154	2088	6.775	91.86	3.738
43	35	85 Journals are with 35th rank	185	2273	8.139	100.00	3.761
		Total	**2273**		**100**		

Findings and Conclusion

The following are the major findings drawn from this study.

Distribution of Citations

- According to the analysis of citation in Earth Science reveals that journals appears to be the most preferred sources of information used by the researchers which occupied 37.18% of the total citations. It shows that the research scholars in earth science mainly used journals for collecting the information.
- U.K. (30.79%) attained the first rank in country wise analysis of citations in earth science.
- English language occupied the first place with 99.38% in the language-wise analysis of citations in earth science.
- It was found that the total 2273 citations of journals are scattered primarily among 45 subjects and 37.3% is connected in one major subject fields i.e. Geology as majority of Ph.D theses under citation study related to this subject.

Bradford's Law of Scattering and Analysis of Bradford's Zones

- The study reveals that 9 (Nine) Core journal which were the source journals leads the majority of the citations with a total of 789 Citation, 46 (Forty Six) journal with the total citation of 761 which were first zone of the Bradford zone and 353 journal with the citation of 723 which were the second zone of the Bradford zone of scattering respectively.

Suggestions and Conclusion

The present study is based on the Citations rendered in Ph.D theses of Earth Science submitted to Manipur University only. Analogous type of research may be carrying out covering the theses in other science subjects and social science subjects. This is to simplify the findings of the study and expand in-depth information of the literature used by the researchers. The Manipur University Library should subscribe those foreign print/e-journals and Indian journals to meet the needs of the research scholar, students and teachers. The Citations from other forms of documentation concerned i.e. e-resources, reports, conference proceedings; records, thesis, government publication etc are very less cited and used because of the insufficiency of information. In this regard it is suggested that concerned authorities should make an effort to overcome these restrictions by providing efficient library services. This would help the researchers not only in the field of earth science, but also in other fields of knowledge to get the required materials for their research purpose. University libraries in the state of Manipur should build a consortium to share their print and electronic resources available in their libraries.

References

Kapoor, S.K. (1984). Citation Analysis of Earth Science Literature. *Annals of Library Science and Documentation, 31*(1-2),56-62 Retrieved from http://nopr.niscair.res.in/bitstream/123456789/27982/1/ALIS%2031%281-2%29%2056-62.pdf

Kushkowski, Jeffrey D., Parsons, Kathy A., & Wiese, William H. (2003). Master's and Doctoral Thesis Citations: Analysis and Trends of a Longitudinal Study. *Libraries and the Academy*, *3*(3), 459-479.

Ravichandran, M., Sivaprasad, G. & Manoharan, K. (2014). Bibliometric Citations in Ph.D. Theses in Library and Information Science at Bharathidasan University, Tiruchi, *International Journal Digital Library Services, 4*(3), 218-230.

Sangam, S.L. (2015). *Scientometrics: Quantitative Methods for Library and Information Science. Bangalore: ContentCraft.*

Singh, K.P., & Bebi (2013). Citation analysis of Ph.D. theses in Sociology submitted to University of Delhi during 1995-2010. *DESIDOC Journal of Library and Information Technology, 33*(6), 489-493.

Ziaur Rahman, M.D., & Bhattacharya, U. (2012). Citation analysis of doctoral theses in Botany: A North Bengal university case study. *International Journal of Library and Information Studies, 2*(3), 1-14.

Article DOI: 10.22032/dbt.39337 Proceedings DOI: 10.22032/dbt.39296

USAGE OF OPEN ACCESS JOURNALS IN RUSSIAN PUBLICATIONS: WEB of SCIENCE, 2008-2017

Levan Mindeli[A], Vladimir Zavarukhin[A], Anna Zolotova[B], Valentina Markusova[B]

([A]The Institute for the Study of Science of the Russian Academy of Sciences, Russia),

([B]All Russian Institute for Scientific and Technical Information of the Russian Academy of Sciences)

Vzavarukhin@gmail.com; V.Zavarukhin@issras.ru; valentina.markusova@gmail.com

Abstract

The development of the Open Access (OA) journals system, its advantages, anddisadvantages are discussed. The bibliometric statistics on Russian research performance(RP) were collected from the Science Citation Index – Expanded (SCI-E) for the period 2008-2017. During this period, Russian researchers published about 34,160 articles in Gold OAjournals which share in the total Russian research performance (303,877 articles) accounts for11.2 percent. The usage pattern of Gold OA journals shows a stable growth rate of publications from 7.8% in 2008 up to 13.7% in 2017. Despite the high cost of OA publications, the Russian Academy of Sciences has the highest share (58.6%) of OA papers. We assume that this is an impact of a robust international collaboration of Russian researchers with the US (31%), Germany (29%) and other industrialized countries that cover the cost of collaborative publications. Among the funding organizations that aim to promote Russian participation in the OA system a critical role belongs to the Russian Science Foundation, the Russian Foundation for Basic Research as well as to CNRS (France), the US National Science Foundation and others. The international collaboration and government appropriations for research in universities had substantial impact on citations score: share of Gold OA highly cited articles amounted to 52% out of the Russian total RP. Leading Research Areas (RA in SCI-E) of Gold OA publications turned out to be entirely different compared with a disciplinary structure in total Russian RP. As an example, one of the most critical research areas in the world - "Scientific Technologies" ranked the third place compared to the ninth place in the total Russian RP. Russian scientists widely use the highest quality foreign journals of the Gold OA system indexed in SCI-E, the only Russian OA journal indexed in SCI-E is “Physics of Condensed Matter” which has the highest share of all Russian publications in Gold OA journals.

Introduction

Open Access system (OA) is one of the hot topics discussed by well-established publishing companies and scholars' community. As G. Eysenbach noted in 2006, the Open access system has the potential to accelerate recognition and dissemination of research findings, but its actual effects are controversial. Some researchers stated that Gold OA journals are cited more highly than printed journals (Sotudeh, 2015); other reports (Moed 2007, Bjork 2012, Solomon 2013, Wray 2016) give evidence that OA journals attract less citations. According to Solomon DJ (2012) article-processing charge (APC) is the central mechanism for funding the OA system. Many researchers discussed the high cost per page in various OA journals. This cost is much higher in well-established journals and low in developing countries. As a consequence of the low cost per page, the number of predator journals is growing. This is why information specialists and bibliometricians should play an important role in the special education program on traditional and new publishing system.

Among recent bibliometric studies of OA system we refer particularly to Prof. G. Lewison's (2015) presentation at the ISSI conference in China in 2017. The author investigated the growth of Gold OA journals and disciplinary domain that embraced this new system and which country benefited from free access to OA publications. Countries were selected according to their research performance (RP) in Web of Science (WoS) and divided according to their gross national product per capita into four groups: high income countries, upper middle income, middle income and low-income countries. All publications were assigned to five major fields of science. The interesting result is that richer countries publish less in OA journals in all domains than poor countries. As an example, the USA turned out to be below the world average in all five major fields, but Brazil was above average in all areas except for physics, and India has been above average in physics, biomedical research and engineering. Russia was above average only in physics. In January 2018 Dr. Archambault E. published a comprehensive longitudinal study on availability of Green OA journals focused on comparison of two bibliographic databases *Web of Science* and *Scopus* with the special database (designed by his company) whose goal was to facilitate retrieval of Gold and Green OA articles published in peer-reviewed journals. According to his findings, articles published in 2014 were available free in health sciences (about 60%), followed by natural sciences (55%), applied sciences (45%) in 2016. Disciplinary domains of art and humanities were significantly less available (24%) www.science-metrix.com

In the Directory of Open Access journals (DOAJ) there are 132 Russian journals. Only one of them “Physics of Condensed Matter” is indexed in the Web of Science (WoS) Core Collection and in Journal Citation reports (JCR).

It is worthwhile to notify that since 2006 Russian government took a few initiatives to reform two main Russian research bodies the Russian Academy of Sciences (RAS) and the Higher Education Sector. One of the most critical action was taken by President of the Russian Federation V. Putin (Decree 599, 2012) was assignment of a new project, denoted as Project 5-100, when special funding was transferred to a selected group of universities (Moed H., 2018).

The Russian government assigned 44 billion Rubles (around 730 million US$) for the Project 5-100 implementation for the period 2013-2016. After two tiers of competition 14 universities were selected; in a later phase, one was added. Each year, all universities were divided into 3 groups according to their results. Each university belonging to the first group receives about 960 MLN Rub.; universities in the second group receive 450-540 MLN Rub each. Finally, each university of the third group obtains about 100 MLN Rub. Meanwhile, it was a significant reduction of research personnel and budget of RAS. Bibliometric indicators play an important role in the evaluation of the efficiency of government reforms (Moed H., 2018).

Since 2016 a new option of Web of Science was introduced marking Open Access publications in Gold Open Access journals (Gold OA) and Green Open Access journals (Green OA). We set up a goal of our paper to overview trends in usage pattern of Gold Open Access journals as a tool for scientific communications and its impact on bibliometric performance indicators by Russian scholar community during the period 2008-2017.

Methods.

The primary sources of bibliometric statistics were resources produced by Clarivate Analytics: Science Citation Index –Expanded (SCI-E), Social Science Citation Index (SSCI) and Arts & Humanities Citation Index (A&H CI) that are part of the Web of Science (WoS) Core Collection. A search was performed on March 17.2018. Due to the significant difference in Russian research performance (RP) between SCI-E (about 33,000 records) and SSCI (about 1,200 records) and A&H CI (300 records) yearly, our analysis was focused on the records indexed only in SCI-E.

Since 2016 Gold OA records are marked in the WoS, we used this marking to differentiateOA and non-OA records and trace the evolution of OA usage. Open Access journals are divided into two groups: Gold OA and Green OA journals in WoS. Our analysis was focused on Gold OA journals that accounts for 95% of all OA records affiliated with Russia. Only articles and reviews (A&R) as a more critical type of publications were selected for the analysis. Bibliometric indicators were as follows: total Russian publication counts (research performance - RP), OA publications counts and it’s share, distribution of total RP and Gold OA publications by leading organizations, collaborative countries, and funding agencies; the percentage of Gold OA highly cited articles.

Special attention was paid to careful visual and manual verification of names of organizations and funding agencies. In our paper, we are focused on publication counts and do not assess citation impact, an aspect of the great importance of a fully-fledged bibliometric assessment study (Moed et.others, 2018)

Results and Discussion

A growth rate in 1.43 fold in research performance (RP) in SCI-E (1,651,875 publications) was observed in 2017 compared with 2008 (1,157,506 publications). The growth in the Gold OA publications was slightly higher - 1.73 folds. Russia occupies 15th place by RP and 29th place in Gold OA publications in SCI-E, 2017.

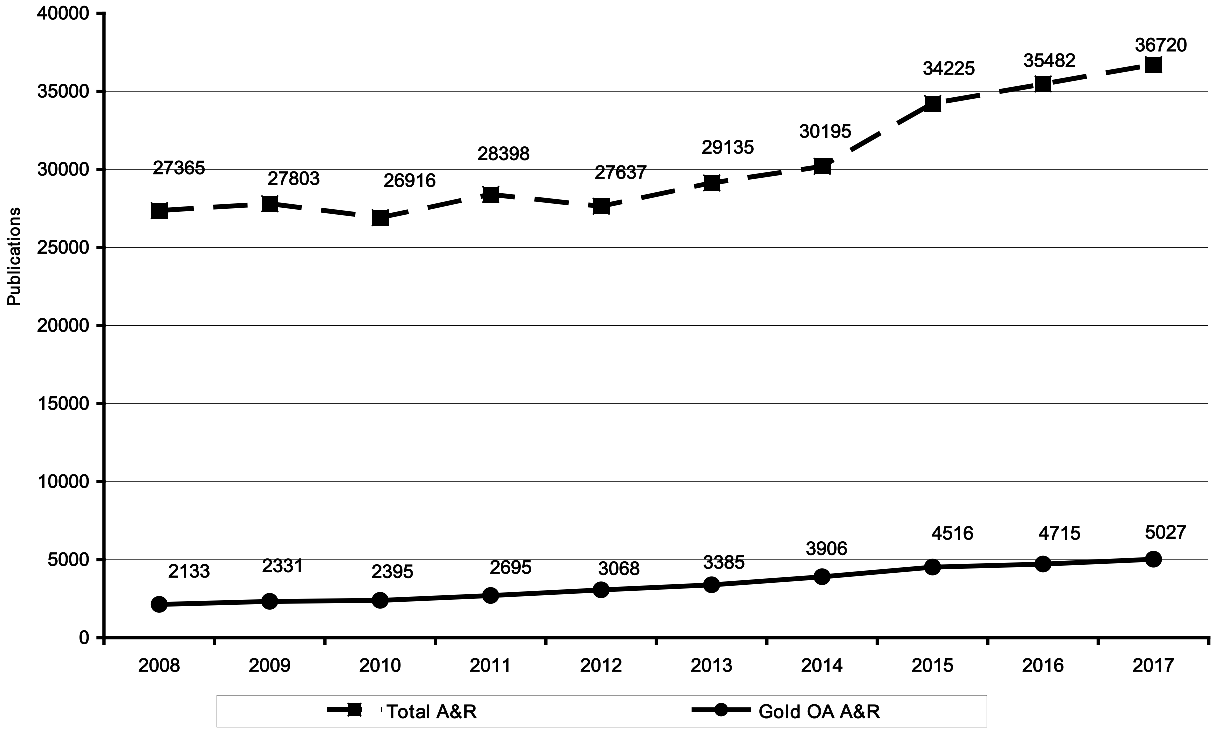
The trends of evolution of Gold OA publications share in the world in three main databases SCI-E, SSCI and A&HCI (column 2**),** separately in SCI-E ((column 3) and in total Russian RP (column 4) are presented in Table 1.

Table 1. Trends of evolution of Gold OA publications, WoS, 2008-2017.

Year	Share of Gold OA (A&R) in SCI-E, SSCI, A&HCI, (%)	Share of Gold OA (A&R) in SCI-E, (%)	Share of Gold OA (A&R) in total Russian RP in SCI-E, (%)
1	2	3	4
2008	19.7	21.08	7.79
2009	20.65	22.14	8.38
2010	21.49	23.10	8.90
2011	22.33	24.0	9.49
2012	23.65	25.44	11.10
2013	24.54	26.3	11.62
2014	25.53	27.3	12.94
2015	25.7	27.4	13.19
2016	25.44	26.74	13.27
2017	25.68	26.58	13.7

During 2008-2017 Russian researchers published more than 34,500 documents in Gold OA journals. Our finding indicates that an average share of Gold OA publication in Russia is still approximately twice less than in the world according to SCI-E in 2017. Trends on the Russian publication's growth in Gold OA journals in SCI-E are presented in Fig.1.

Figure 1. Growth trends in Russian Gold Open Access publications and total Russian research performance, SCI-E.



It was observed that the growth (in 2.36 folds) in absolute number of Russian publications in the Gold OA journals is significantly higher than the growth (in 1.3 folds) of total Russian RP during 2008-2017. There are many factors – financial, age, language barrier - that influence an author's decision whether to submit the manuscript to a regular journal or to Gold OA.

Two lists of twenty leading organizations (ranked by publications counts) in entire Russian RP and Gold OA publications were compiled with a respective share in Gold OA publications and in total Russian research performance. These data are presented in Table 2.

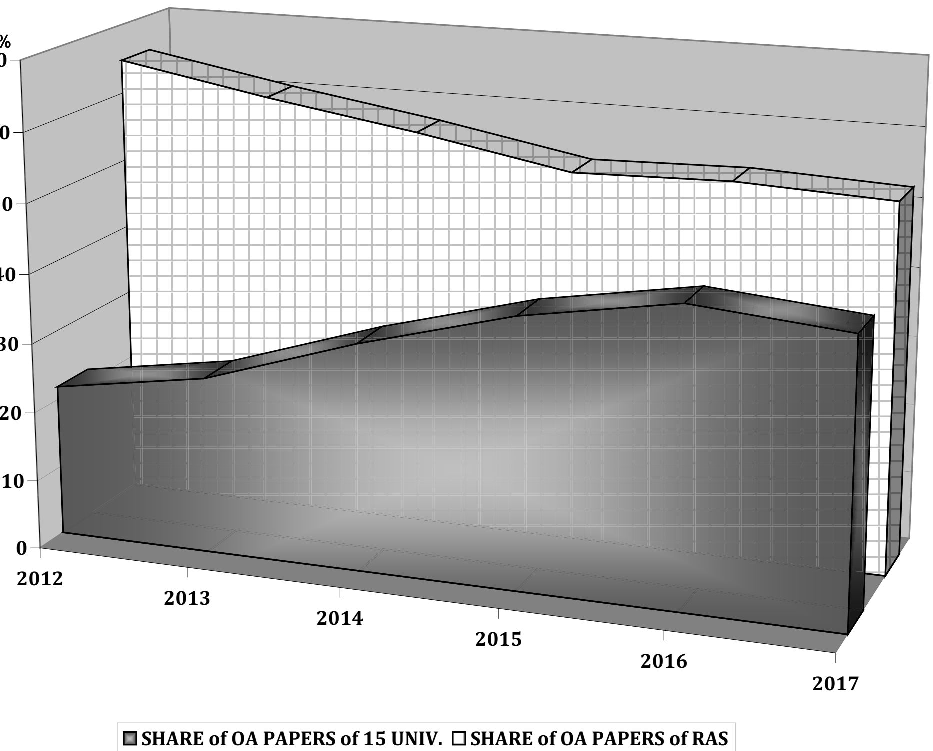
Table 2. The leading organizations by share of Golden OA publications, WoS, 2008-2017.

All publications of Russian authors			Publications only in Gold OA journals		
Rank	Organizations, combined	Share, %	Rank	Organizations, combined	Share, %
	In total 303,877 publications	100%		In total 34 ,160 publications	100%
1	Russian Academy of Sciences	56.6	1	Russian Academy of Sciences	58.6
2	Moscow State University	12.0	2	Moscow State University	17.7
3	St. Petersburg State University	4.3	3	Centre National de la Recherche Scientifique (CNRS)	12.0
4	St. Petersburg Scientific Center, Russian Academy of Sciences	4.2	4	Helmholtz Association	9.5
5	Centre National de la Recherche Scien- tifique (CNRS)	3.5	5	University of California System	9.0
6	Novosibirsk State University	3.5	6	Universite Paris Saclay Comue	8.8
7	National Research Center Kurchatov Institute	3.2	7	National Research Center Kurchatov Institute	8.6
8	Joint Institute for Nuclear Research	2.9	8	United States Department of Energy (DOE)	8.6
9	Helmholtz Association	2.7	9	Joint Institute for Nuclear Research	8.0
10	Ioffe Physical-Technical Institute, Russian Academy of Sciences	2.7	10	Alikhanov Institute of Theoretical and Experimental Physics	7.9
11	Lebedev Physical Institute, Russian Academy of Sciences	2.5	11	CNRS National Institute of Nuclear Particle Physics IN2P3	7.6
12	Moscow Institute of Physics and Technology	2.3	12	Max Planck Society	7.6
13	National Research Nuclear University MEPhI	2.1	13	Instituto Nazionale di Fisica Nucleare	7.6
14	Ural Federal University	1.9	14	Konstantinov Petersburg Nuclear Physics Institute, National Research Center Kurchatov Institute	7.2
15	Universite Paris Saclay Comue	1.8	15	Lebedev Physics Institute, Russian Academy of Sciences	7.1
16	University of California System	1.8	16	Sapienza University Rome	7.1
17	United States Department of Energy (DOE)	1.8	17	CEA (Commissariat à l'énergie atomique et aux énergies alternatives)	6.9
18	Max Planck Society	1.8	18	Consejo Superior De Invesatigaciones Cientificas CSIC	6.9
19	Russian Academy of Medical Sciences	1.8	19	European Organization for Nuclear Research (CERN)	6.5
20	Alikhanov Institute of Theoretical and Experimental Physics	1.8	20	St. Petersburg Scientific Center, Russian Academy of Sciences	6.5

As it was mentioned in many papers (Graham 1995, Karaulova 2016, Markusova, 2014) the Russian Academy of Science (RAS) is the leading basic research body. Its leading role is confirmed by high share 56.5% and 58.5% correspondently in total Russian RP and Gold OA publications during 2008-2017. Nevertheless the RAS budget was cut off, and increased slightly in 2018. No wonder that the RAS share of Gold OA publications decreased significantly from 65.8 per cent % to 53.9 % correspondently in 2012 to 2017.

As was mentioned above, the significant investment in Program 5-100 stimulated the tremendous growth of universities' publications in 2012-2017. These fifteen universities' share of Gold OA publications increased from 14.8% in 2012 to 31.54% in 2017. The impact of financial investments on the pattern of Gold OA usage is displayed in Fig.2.

Figure 2. Share of Gold OA publications by the RAS and universities included in Program 5-100, SCI-E, 2008-2017.



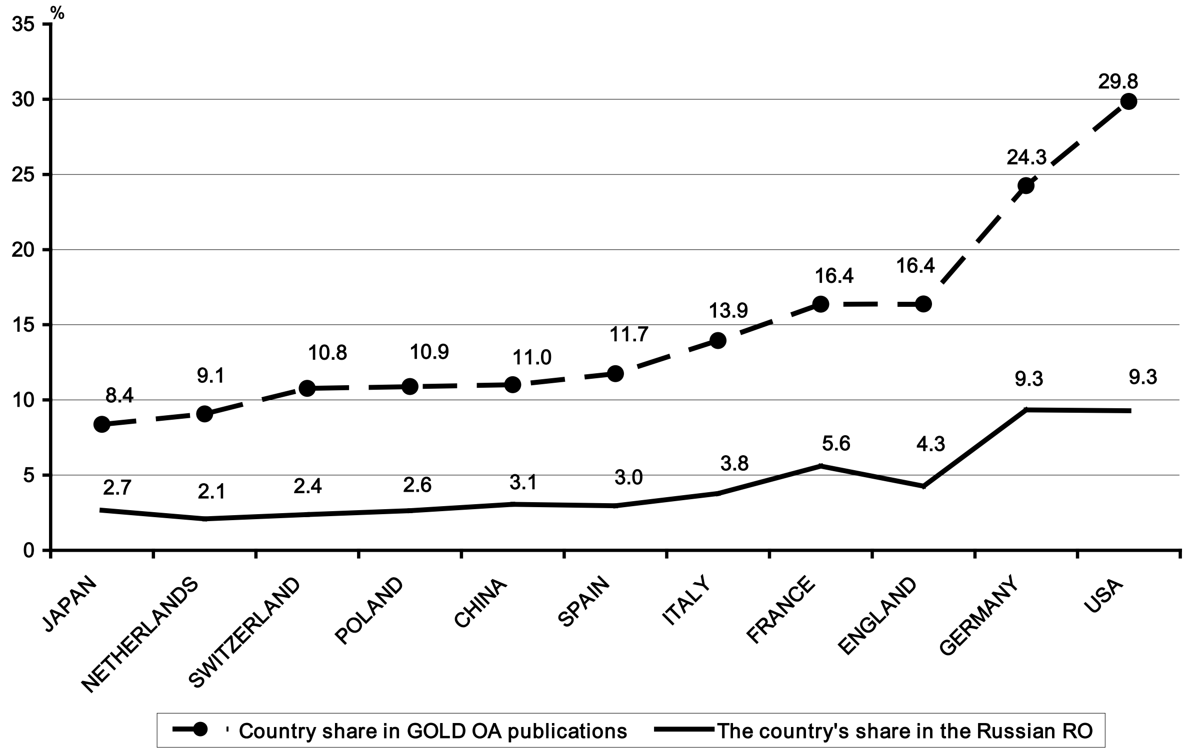
There are fourteen foreign organizations among the top twenty (Table 2) that had collaborative papers in Gold OA publications and only six of them are among the top in total RP. This finding provides indirect evidence that international scientific collaboration allow Russian researchers to overcome the financial obstacle.

Disciplinary distribution of Russians publications was traditionally focused on “hard sciences” (Markusova, 2018). To investigate the disciplinary difference /or similarity in research priorities there were selected top 50 Research Areas among Gold OA publications and total Russian research performance. Our findings show that leading Research Areas (RA) of Gold OA publications were entirely different compared with disciplinary structure in total Russian RP. As an example, one of the most critical research areas in the world - "Scientific Technologies" ranked third compared to the 9th place in the total Russian RP. Another striking result was that share of RA “Oncology is 5-fold higher than in total Russian research performance, this is partly due to strong collaboration of the Russian National Cancer Research Center with National Institutes of Health (NIH).

It is well-known that the governments of the European Union countries and the USA actively promote the OA system for disseminating knowledge obtained at taxpayers' expense. Our data demonstrate a significant influence of various funding agencies on the number of publications in Gold OA journals as compared with total Russian RP during 2008-2017. The share of funding agencies reached 79.8 % among publications in Gold OA journals and 61.8% in total RP. Among the most active Russian funding organizations is the Russian Foundation for Basic Research in collaboration with CNRS (France), the National Science Foundation (USA), the National Institutes of Health (USA) and the Russian Science Foundation (RNF) established only in 2013. An average amount of RNF grant is about $100,000 per year that is five folds more than an average grant of RFBR.

According to Wagner C.(2017) international projects account about 20% of national government spending on scientific research. International collaboration of Russian researchers have a significant impact on opportunity to be published in Gold OA journals. Our data demonstrate the higher percentage of these publications compared with the share of industrialized countries in total Russian RP. These data are presented at Figure 3.

Figure 3. Share of international collaboration in Russian Gold Open Access publications and in total Russian research performance, SCI-E, 2008-2017.



This chart demonstrates clearly the growing activity of Russian collaboration between industrial countries in Gold OA publications. The share of each of the countries - USA and Germany - has increased almost three fold compared to their share in total Russian research performance. The international collaboration had the impact on citations score: share of Gold OA highly cited articles amounted to 52% out of the Russian total RP.

Conclusions

Trends in the usage pattern of Gold OA journals revealed a stable growth rate of publications from 7.8% in 2008 up to 13.7% in 2017. Despite high Gold OA publications cost the Russian Academy of Sciences had the highest share (56.5%) of OA papers. However, a significant decrease in its share in Gold OA publications from 65.1% to 52.9% was observed correspondingly from2012 to 2017.

Due to government investments in fifteen selected universities (Program 5-100), their share in Gold OA publications doubled in 2017 compared with 2012. Our findings indicate an impact of a robust international collaboration of Russian researchers with the USA (31%), Germany (29%) and other industrialized countries that cover the cost of collaborative publications. Leading Research Areas (RA in SCI-E) of Gold OA publications were entirely different as compared with a disciplinary structure in total Russian research performance. The implications of our analysis provide a better empirical basis for science policy with respect to disseminating the results of the Russian research using Gold OA system.

Acknowledgments

Authors express the gratitude to the Russian Foundation for Basic Research for financial support (Grant №17-02-00157 and Grant N 17-02-00078) and to Clarivate Analytics for the opportunity to use the data.

References

Archambault E. (2018) *Analytical Support of Bibliometric Indicators*, info@science-metrix.com, URL: www.science-metrix.com

Bjork B.C. & Solomon D. (2012). Open access versus subscription journals: A comparison of scientific impact", *BMC Medicine* 10: 73. DOI:10.1186/1741-7015-10-73. PMC 3398850.

Eysenbach G. (2006) Citation Advantage of Open Access Articles, *PLOS Biology* 4 (5): 692-698. DOI: 10.1371/journal.pbio.0040157.

Graham L.R.(1998) What have we learned about science and technology from the Russian experience? Stanford University Press: Stanford, California, 196 p.

Karaulova M**.,** Abdullah G., Shackleton O**.,** Shapira P. (2016) Science system pass-dependencies and their influences: nanotechnology research in Russia, *Scientometrics.* .100 (3): p. 365-383. DOI 10.1007/s11192-016-1916-3

Lewison G. (2017) Open access papers: their growth over time and from different countries, and their citations".*16th International Conference on Scientometrics & Informetrics 16-20 October,* *2017. Wuhan University. Wuhan, China. ISSI Proceedings:* (1247-1255). URL: http://issi-society.org/proceedings/issi_2017/2017ISSI%20Conference%20Proceedings.pdf

Markusova V.A., Libkind A.N., Jantz M., Mindeli L.E. (2014) Research performance by Federal and National Research Universities and impact of competitive funding on their publication activity. *Collnet Journal of Scientometrics and Information Management*, 2(7): 217-229.

Markusova V.A., Libkind A.N., Noyens E., Mindeli L.E. (2014) Russian Universities Collaboration with Domestic and Foreign Funding Agencies// Proceedings Science and Technology indicators conference 2014, 3-5 September in Leiden, the Netherlands. – Edited by Ed Noyens.-CWTS- Leiden Universitiet, Leiden.- P.386-390

Moed H.F. (2007) The effect of "Open access" on citation impact: An analysis of ArXiv's condensed matter section, *Journal of the American Society for Information Science and Technology* 58: 2047-2054.

Moed H., Markusova V., Akoev M. (2018) Trends in Russian research output indexed in Scopus and Web of Science, *Scientometrics* 118 (2). URL: https://doi.org/10.1007/s11192-018-2769-8.

Pisoschi, AM, Pisoschi, CG (2016) Is open access the solution to increase the impact of scientific journals? *Scientometrics*, 109(2): 1075-1095. DOI 10.1007/s11192-016-2088-x.

Solomon D.J., Bjork B.C. (2012) A study of open access journals using article processing charges, *Journal of American Society on Information Science and Technology. Sci. Technol.*, 63(8): 1485-1495.

Solomon D.J., Laakso M. & Bjork B.C. (2013) A longitudinal comparison of citation rates and growth among open access journals, *Journal of Informetrics* 7: 642-650.

Sotudeh H., Ghasempour Z. & Yaghtin M. (2015) The citation advantage of author pays model: the case of Springer and Elsevier OA journals, *Scientometrics* 104 (2): 581-608.

Wang X.W., Liu C., Mao W.L. & Fang Z. (2015) The open access advantage considering citation, article usage and social media attention, *Scientometrics* 103(2): 555-564.

Wagner C., Koen J. (2017) Open countries have strong science, *Nature* 550 (7674): 32-33.

URN: urn:nbn:de:gbv:ilm1-2019200405

DOI: 10.22032/dbt.39296