Do first rate scientists work at first rate organizations?

Bo Yang[1], **Ronald Rousseau** [2,3], **Shuiqing Huang**[1*] and **Sulan Yan**[1]
[1]College of Information Science and Technology,
Nanjing Agricultural University, Nanjing, 210095, P.R. CHINA
[2]Institute for Education and Information Sciences, IBW, University of Antwerp (UA),
Venusstraat 35, Antwerp, B-2000, BELGIUM
[3]KU Leuven, Dept. Mathematics, Leuven B-3000, BELGIUM
e-mail: *sqhuang@njau.edu.cn (corresponding author)

ABSTRACT

In this article we study the relation between first rate scientists and first rate organizations. These notions are operationalized by counting citations (whole counts in the Web of Science) for articles published during the period 2008 to 2011. Care has been taken to obtain clean data by careful institutional name disambiguation. It is found that, especially for large fields, more than 80% of top scientists work at top organizations. This is less the case for smaller fields such as mathematics and computer science. Our research confirms the skewness (elitism) of science, also for the relation between first rate scientists and the organizations to which they belong. We find, moreover, that top institutions have relatively more excellent scientists and relatively fewer researchers with poor performance than the average level of all institutions in a field.

Keywords: First rate scientists; First rate organizations; World class universities; Skewness of science; Total citations; Tipping number (TIP)

INTRODUCTION

Although impossible to define precisely, it is common knowledge that some scientists are really exceptionally talented and some are not. If the talented ones are successful in their research they are the persons that receive Nobel prizes, Fields Medals and similar awards of high standing. Similarly some universities are known by any scientist: the Harvard, Yale, Massachusett Insititute of Technology (MIT), University of California, Oxford and Cambridge type of universities, say world class universities, and some are only of local importance. In this article we will use the general term “organization”, referring to a university, a research institute, a national research institute with many branches such as the US National Institutes of Health (NIH), the Centre National de la Recherche Scientifique (National Center for Scientific Research - CNRS) in France, the Chinese Academy of Sciences (CAS) or the Max Planck Institute in Germany. We use the term “organization” because this is the term used in Thomson Reuters’ Web of Science (WoS).

Highly visible organizations harbor many high levels, elite scientists but every now and then a top scientist works at a local or average-level institute. What is the exact relation between these two: between high level organizations and high level scientists? This is the main research question discussed in this article (made more precise further on). It is similar to the question: do journals with a high impact factor publish (all) highly cited articles? (Seglen 1994), which, of course they do not.

We also consider the complete citation distribution of scientists working at these high level organizations and compare them with all scientists publishing in the field. This comparison answers the question if all scientists working at first rate organizations are first rate scientists themselves. Common sense and e.g. the results of Abramo, Cicero, and D'Angelo (2012, 2013) suggest that this is most certainly not the case. A more precise answer is provided further on.

In the remaining sections we provide a literature review, fix the terminology and explain the methods we use. Then we show our findings, which we discuss in another section. The results of comparing scientists working at first rate organizations with the whole field is presented in the penultimate section, ending with a conclusion and some science policy advice.

LITERATURE REVIEW

On the one hand, highly visible organizations are nowadays those topping the Academic Ranking of World Universities (ARWU), the Times Higher Education (THE) World University Rankings, the Leiden Rankings, the SCImago Institutions Rankings (SIR) and similar rankings (Buela-Casal et al. 2007). These rankings have been studied intensively but have also been criticized heavily (Van Raan 2005; Van Parijs 2009; Dehon et al. 2010; Butler 2010; Bookstein et al. 2010).

On the other hand highly cited scientists are known e.g. through Thomson Reuters' lists of Highly Cited Researchers, which identify those researchers who have contributed to a significant number of the most highly cited publications in a field (based on Essential Science Indicators [ESI] data). A recent investigation based on new published data of Highly Cited Researchers (Thomson Reuters 2014) tries to investigate the distribution of these researchers across institutions around the world (Bornmann and Bauer, in press). The statistical results in this paper provide some clues on the concentration of highly cited researchers.

Aziz and Rozing (2012) proposed a profit index which measures the degree to which scientists profit from their co-authors. It is shown that the contribution of co-authors to the work of highly visible authors if often substantial. Consequently they proposed a profit adjusted h-index for a fairer impression of a scientist's academic work. Another adjustment of the h-index was proposed by Vieira and Gomes (2011). Their adjustment considers different citation cultures and the number of authors per publication. Larivière et al. (2010) studied all university professors in the province of Quebec, Canada, paying special attention to the concentration of funding, papers and citations (at the level of individual researchers). They found that each of these distributions is different, citations being the most concentrated. Moreover, they observed notable differences between disciplines: the social sciences and humanities being the most concentrated.

However, the relation between the two has rarely been investigated. A general description of the productivity of Malaysian authors and organizations, restricted to the field of library and information science is provided in Yazit and Zainab (2007). Rodriguez-Navarro (2012) claims that universities with the best reputation differ from others in terms of the concentration of outstanding scientists present, but less so in terms of the total number of papers or citations. He observes that the articles belonging to the 1% high-citation tail reveal much better (than just total number of publications or citations received) contribution of universities to the progress of science. Abramo, Cicero, and D'Angelo (2012, 2013) noted that because the Italian university system is non-competitive, one finds top researchers and unproductive ones in all universities. They further studied the impact of these two groups on the research performance of the university they belong to. As a case study they analyzed all Italian universities active in the hard sciences. They conclude that it would be preferable to allocate selective government funding directly to individual researchers.

Yet, the relation we want to discover has, to the best of our knowledge, not been studied before.

WHAT DO WE MEAN BY "FIRST RATE"? - PURPOSE OF THIS STUDY

The term "first rate" as used in the expressions "first rate organizations" and "first rate scientists" must be operationalized. Recall that the word "to operationalize" means: to use as a proxy for an idea that, by itself, cannot be defined precisely. We admit that choosing an operationalization can be done in many ways. Using the h-index in versions suitable for individuals and for organizations (Hirsch 2005; Prathap 2006; Liang and Rousseau 2009) is a possible approach; calculating average number of citations (per publication) might be another. In this paper we choose a "big is beautiful" approach and use the total number of citations received (concretely: over the period 2008 to 2011, as provided by Thomson Reuters' WoS). Finally, we selected those organizations and those scientists that belong to the top 1% (per field) in the lists resulting from ranking organizations and scientists. These are the organizations and scientists we consider to be "first rate" or the "top" ones.

Studying the relation between first rate organizations and first rate scientists is a study best placed within the sociology of science. We note further that, at least theoretically (be it not in practice), the two notions are independent. Indeed, one could imagine a field in which the top 1% of scientists all belong to small organizations and are surrounded by not-so-brilliant colleagues, while the top 1% of organizations has many scientists belonging to the top 10% - but not the top 1% - of their field. In such a hypothetical case not a single first rate scientist would belong to a first rate organization. In the remainder of this article we will estimate what the real percentages are.

METHODS

Our investigation about the relation between first rate institutes and first rate scientists is performed per field. Concretely we use fields as defined in the Essential Science Indicators (ESI). Recall that, for the ESI database, Thomson Reuters assigns each journal to one of 22 so-called broad fields, such that each journal is uniquely assigned to a single broad field. This division of journals over ESI broad fields can be considered as the result of a top-down approach (Liu and Rousseau 2010). The 22 broad fields are shown in Table 1. For practical reasons (too large or too diverse) we did not include the fields *Clinical Medicine* and *Social Sciences, general*, in our investigation working with the remaining 20 fields.

Also a time frame for our investigations must be determined. Concretely, we use articles published during the period 2008 to 2011 and collect for those articles included in the WoS Science Citation Index Explanded (SCIE) all citations as received by September 2012.

Table 1: Essential Science Indicator's Broad Fields

Agricultural Sciences	Mathematics
Biology & Biochemistry	Microbiology
Chemistry	Molecular Biology & Genetics
Clinical Medicine	Multidisciplinary
Computer Science	Neuroscience & Behavior
Economics & Business	Pharmacology & Toxicology
Engineering	Physics
Environment/Ecology	Plant & Animal Science
Geosciences	Psychology/Psychiatry
Immunology	Social Sciences, general
Materials Sciences	Space Science

Then all organizations and all scientists are ranked according to the number of received citations, using whole counts. This means that each scientist and each organization receives a credit for each citation of an article for which he/she was a co-author, or for which at least one member of the organization was a co-author. Recall that the term "first rate" is operationalized as those organizations and those scientists that belong to the top 1% in these lists. The top 1% was chosen because it is generally agreed that evaluation of the research level of e.g. organizations by using articles in the high-citation tail of the citation distribution is more accurate and reliable than other types of evaluations (Rodriguez-Navarro 2011). Simple as these definitions may seem, they lead to a well-known problem, namely that of organization and author name disambiguation (Smalheiser and Torvik 2009; Strotmann and Zhao 2012). From earlier research of one of the authors we recall the case of the ICDDR,B (International Centre for Diarrhoeal Disease Research, Bangladesh) for which we found 52 different name variations (Mahbuba et al. 2010). Now, such problems are solved as follows: a program has been written to detect all similar names, two by two. Examples of suggested terms are: UCLA and UC Berkeley (different); Zhejiang University and Zhejiang University of Technology (different); Chinese Acad Sci and CAS (same). An authority file of unique names (organizations and authors) was set up. Then all suggested similar pairs of names were checked manually by one of us (using the WoS, institutional webpages and the original publication if necessary). If the suggested terms referred to the same entity then they were both stored under the name included in the authority file. If not they were each replaced (if necessary) by the authoritative name. We point out that our procedure leads to a more exact number of different persons and organizations than available in the ESI files. Details of the procedure can be found in Huang et al. (2014).

For authors, the following rule was applied: if two scientists have the same name and same initials (as far as shown in the WoS), if they work in the same organization and in the same field, then they are considered to be the same person. Hence, immunologists Zhang B and Zhang B working at the same organization are considered to be the same person (even if in reality these names refer to Zhang Bo and Zhang Bei). Zhang X and Zhang XL are always considered to be two different persons.

Moreover we subdivided the top 1% in three groups for a more detailed analysis. As there does not exist a rule to perform such a division we chose a division with universal appeal: namely one based on the golden section. We recall that the golden section or golden ratio (often denoted as phi = ϕ) is equal to 1.61803… and 1/ (1.61803…) ≈ 0.618. Based on this ratio the third group consists of the lower 61.8% of the top 1%; the second group consists of the next $(0.618)^2$ % or 23.6% of the top 1%; while the remaining part leads to the first group (the best of the best, at least in terms of citations over this period) including 14.6% of the top 1%. We admit that this division is as arbitrary as any other, but it will serve our purpose.

One may wonder why this division was not based on Bradford's law ($1:n:n^2$), which is a well-known principle in the field of informetrics (Bradford 1934). The point is that Bradford's law refers to a division of items (all items) over sources, but here not all first rate scientists are distributed over all first rate universities, so that a Bradford-type approach is not meaningful.

FINDINGS

General

Before going into details, we first describe the answer to the general research question: to which extent do first rate scientists work at first rate organizations? Table 2 shows the main findings. We include the category *Multidisciplinary* in this table, but do not pay attention to it as it is not a field of investigation and hence less meaningful for our research. We observe several fields in which more than 80% of the top scientists work at top institutes: *chemistry*, *geosciences*, *materials sciences*, *neuroscience and behavior*, *physics*, *plant & animal science* and *psychology/psychiatry*. The smallest percentages occur in *mathematics* (45%) and *computer science* (54%).

Table 2: Numbers and Percentages of Top Organizations, Top Scientists and Percentages of Top Scientists Working at Top Organizations

Subject Field	Number of top 1% organizations	Number of top 1% scientists	Number of top 1% scientists working at a top 1% organization	Percentage of top 1% scientists working at a top 1% organization
Agricultural Sciences	338	2924	2126	72.7%
Biology & Biochemistry	526	7895	6246	79.1%
Chemistry	556	9036	7776	86.1%
Computer Science	177	1786	965	54.0%
Economics & Business	140	1076	783	72.8%
Engineering	558	6058	4760	78.4%
Environment/Ecology	388	3359	2572	76.6%
Geosciences	286	2585	2161	83.6%
Immunology	222	2495	1849	74.1%
Materials Sciences	334	3898	3225	82.7%
Mathematics	118	997	449	45.0%
Microbiology	214	2336	1534	65.7%
Molecular Biology & Genetics	329	5154	4060	78.8%
Multidisciplinary	*123*	*1660*	*989*	*59.6%*
Neuroscience & Behavior	350	4536	3802	83.8%
Pharmacology & Toxicology	327	3360	2317	69.0%
Physics	354	5635	5060	89.8%
Plant & Animal Science	609	6045	5014	82.9%
Psychology/Psychiatry	307	2658	2273	85.5%
Space Science	89	1007	701	69.6%

Considering Figure 1, we see that on the data range (x-values between 1,000 and 10,000) there is a perfect linear relation between the number of first rate scientists in a field (x-values in Figure 1) and the number of first rate scientists working at a first rate organization (y-values in Figure 1). The Pearson correlation coefficient of the data in Figure 1 is 0.994; the equation is $y = 0.88\,x – 336.1$.

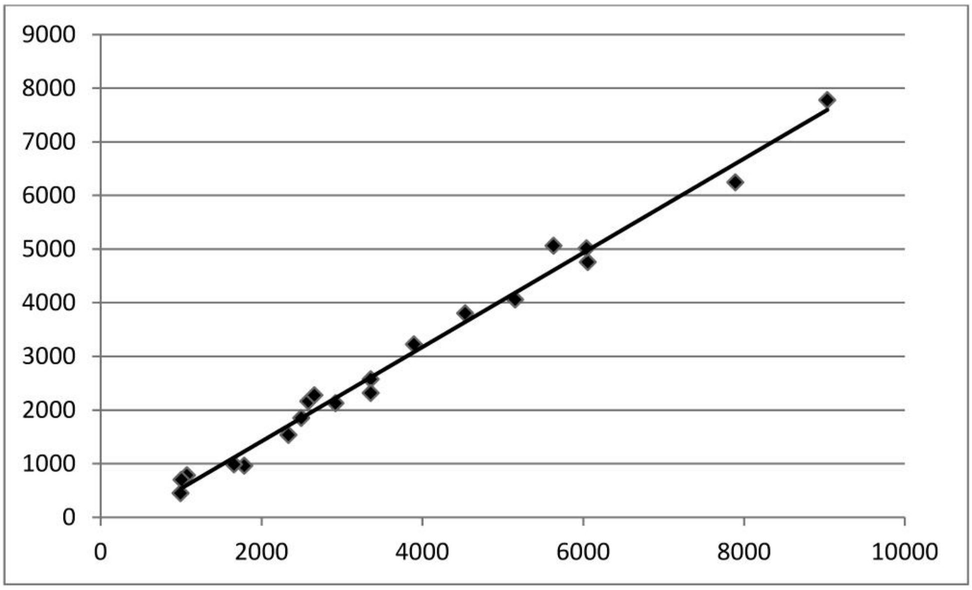

Figure 1: Relation between the Number of First Rate Scientists in a Field (Abscissa) and the Number of First Rate Scientists Working at a First Rate Organization (Ordinate).

We further observe that the larger x (the more scientists in the top 1%, hence the larger the field) the larger the percentage of first rate scientists working at first rate organizations. Indeed, using the regression equation (not the real data) we find for $x = 1000$ a y-value of 544, or 54.4% belonging to a first rate organization, while for $x = 4000$, the corresponding y-value is 3184 or 79.6 % belonging to a first rate organization. In reality there are deviations but these values illustrate the general trend. Having established that the larger the field the higher the percentage of first rate scientists working at first rate organizations we now consider the real data and see if we can detect this trend directly. We, indeed, find that the relation between the percentage of first rate scientists belonging to a first rate organization (x-values in Figure 2) and the absolute number of such scientists (y-values in Figure 2) can be described by an increasing graph with equation $y = 130.05 * e^{4.323x} – 680$ (valid for $0.4 < x < 0.9$). We obtained this equation by nonlinear regression based on the Marquardt algorithm with $R^2 = 0.46$).

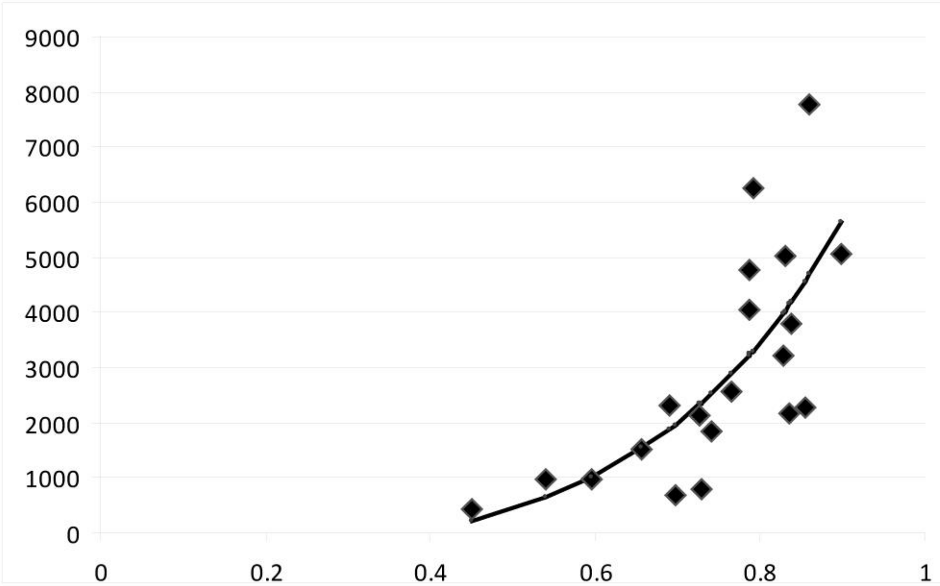

Figure 2: Nonlinear Relation between the Percentage of First Rate Scientists (Abscissa) Belonging to a First Rate Organization and the Absolute Number of Such Scientists (Ordinate).

Detailed analysis

First we explain how, based on the idea of the golden section, first rate organizations are further subdivided into A, B and C level organizations. Consider, for example, the field of *Agricultural Sciences.* There are 338 first rate organizations in this field. The first 14.6% or 49 form the group of A-level organizations, the next 23.6% or 80 lead to the group of B-level organizations and the remaining 61.8% or 209 form the C-level group. Numbers are rounded to the nearest unit and if – by rounding - the sum does not lead to the exact total number of first rate organizations, the number of C-level organizations is adapted. As these calculations are easy to perform Table 3 shows just shows the numbers of A-, B- and C-level organizations for two fields.

Table 3: Subdivision into A, B and C Levels

Fields	Number of top 1% organizations	A level	B level	C level
Agricultural Sciences	338	49	80	209
Mathematics	118	17	28	73

Table 4 shows the division of first rate scientists over the three levels of first rate organizations. The columns, e.g. for *Agricultural Science*, have the following meaning: of all first rate scientists that work at first rate organizations 66.67% work at A-level organizations, 23.23% work at B-level organizations and 10.10% work at C-level organizations. As there are much more C-level organizations than B-level ones, and much more B-level ones than A-level ones this pattern, which is generally true, points to a high concentration of top scientists in the best organizations. This observation is further illustrated in Figure 3. These percentages differ so much from a uniform distribution (the null hypothesis) that a statistical test is superfluous.

Table 4: Organizational Distribution of First Rate Authors

Subject Field	A level	B level	C level
Agricultural Sciences	66.67%	23.23%	10.10%
Biology & Biochemistry	68.06%	23.61%	8.33%
Chemistry	68.83%	23.38%	7.79%
Computer Science	59.85%	27.27%	12.88%
Economics & Business	60.87%	26.96%	12.17%
Engineering	66.67%	24.24%	9.09%
Environment/Ecology	62.22%	26.67%	11.11%
Geosciences	70.55%	21.92%	7.53%
Immunology	64.52%	25.16%	10.32%
Materials Sciences	72.36%	19.51%	8.13%
Mathematics	51.80%	25.90%	22.30%
Microbiology	61.65%	25.56%	12.78%
Molecular Biology & Genetics	67.26%	23.89%	8.85%
Multidisciplinary	71.15%	23.72%	5.14%
Neuroscience & Behavior	72.73%	19.83%	7.44%
Pharmacology & Toxicology	57.30%	29.21%	13.48%
Physics	60.87%	28.70%	10.43%
Plant & Animal Science	72.06%	20.59%	7.35%
Psychology/Psychiatry	70.07%	22.63%	7.30%
Space Science	53.25%	35.21%	11.54%

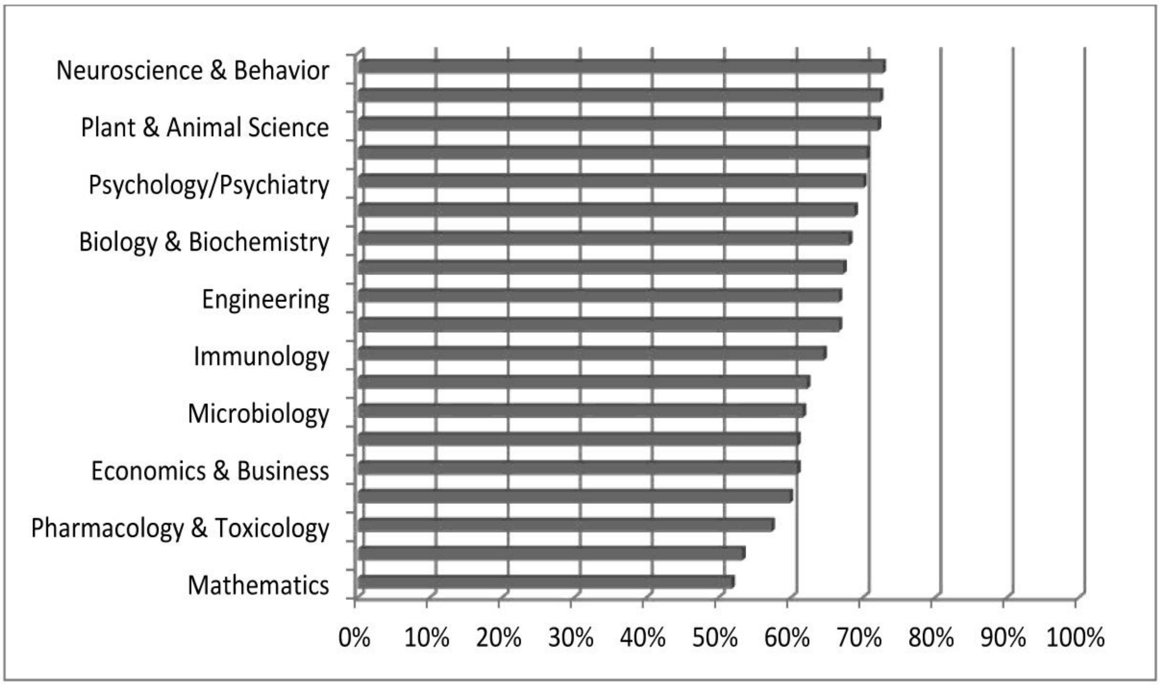

Figure 3: Percentages of First Rate Scientists Working at A-Level Organizations

DISCUSSION

The words 'elite' and 'science' go hand in hand. It is, indeed, well-known that scientific results, and especially high level results, are highly skewed among organizations and persons (Seglen 1992, 1994). In this contribution it is shown that, depending on the field, most first rate scientists work at first rate organizations. According to Abramo, Cicero, and D'Angelo (2013) this is a characteristic of competitive systems. Roughly speaking one may say that the larger the field, the higher the percentage of scientists working in A-level organizations. Leaving multidisciplinary sciences out, we have a 68% Spearman rank correlation between the number of top 1% organizations (a measure for the size of a field) and the percentage of top 1% scientists working at A-level organizations (Table 5). This correspondence is in particular true for the smaller fields.

Table 5. Relation between the Size of a Field and the Percentage of Top 1% Scientists Working at A-Level Organizations

Subject Field	Number of top 1% organizations	Percentage A level
Space Science	89	53.25
Mathematics	118	51.80
Multidisciplinary	*123*	*71.15*
Economics & Business	140	60.87
Computer Science	177	59.85
Microbiology	214	61.65
Immunology	222	64.52
Geosciences	286	70.55
Psychology/Psychiatry	307	70.07
Pharmacology & Toxicology	327	57.30
Molecular Biology & Genetics	329	67.26
Materials Sciences	334	72.36
Agricultural Sciences	338	66.67
Neuroscience & Behavior	350	72.73
Physics	354	60.87
Environment/Ecology	388	62.22
Biology & Biochemistry	526	68.06
Chemistry	556	68.83
Engineering	558	66.67
Plant & Animal Science	609	72.06

The Matthew Effect (Merton 1968, 1988) which states that the rich get richer and the poor get poorer is an explaining principle in sociology and economics. Originating in the fields of sociology and economics it has been introduced in the information sciences and linked to the cumulative advantage and success-breeds-success effects. Basically the Matthew effect is the occurrence of a positive feedback loop. This effect is the basic explanation of the ubiquity of power laws in informetrics and several other fields (Mahbuba and Rousseau 2011). We claim that it explains a large part of our observations. Yet even some "Podunk" universities, i.e. lower level universities, have great scientists (Gaston 1978, p. 122; Knudop Search Group 2008).

A GLOBAL VIEW

In this section we will compare scientists working at first rate organizations with all scientists publishing in the field. For a given field, let n be the total number of citations received by an author for articles published during the period [2008; 2011]. Authors who have published (in journals indexed by the WoS) during this period are called active authors. Let x_n be the percentage of (active) authors working at top 1% organizations with total number of citations received by these articles in the year 2011, equal to n. Similarly let y_n be the percentage of active authors in the whole field with total citations equal to n. Let MAX be the number of citations received by the most-cited scientist in a field. For each field and for n = 1, …, MAX we draw the scatterplot (x_n,y_n), but omitting points of the form (0,0). On this scatterplot, see Figure 4, we add the line y=x for visual clarity. Table 6 shows a fictitious example for n = 0, 1, 2, …, 9 = MAX, (assuming that the author with the most citations received 9 citations).

Table 6. Fictitious Example; x_n : Percentage of Authors Working at Top 1% Organizations with Total Number of Citations Equal to n; y_n Denotes the Percentage of Authors in the Whole Field with Total Citations equal to n.

n	x_n	y_n	x_n-y_n
9	0.030	0.0006	0.029
7	0.030	0.0064	0.024
6	0.060	0.0250	0.035
5	0.070	0.0120	0.058
4	0.100	0.0350	0.065
3	0.110	0.1230	-0.013
2	0.100	0.0880	0.012
1	0.200	0.2600	-0.060
0	0.300	0.4500	-0.150
SUM	1.0	1.0	0

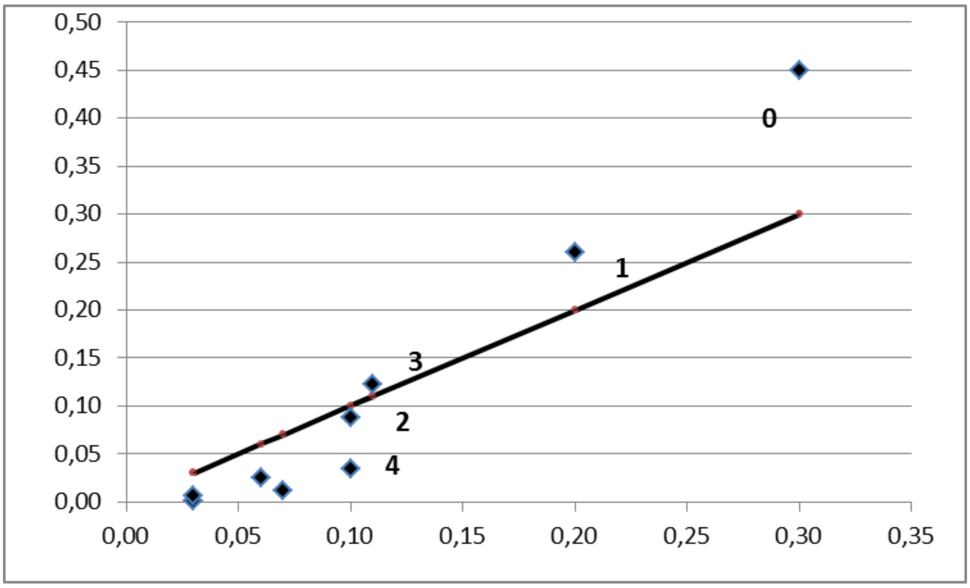

Figure 4: Scatterplot for Fictitious Example (n=0,…,4 indicated)

This fictitious example is drawn similar to real cases in the sense that for small n, data are generally situated above the line y = x, while for large n, data are generally situated below the first diagonal. We determine the largest number k such that for each m ≤ k: $x_m \le y_m$ or $(x_m,y_m) = (0,0)$. For the fictitious case we see that k = 1. For values m strictly larger than k, we have in most cases that $x_m > y_m$. The number k is referred to as the TIP (total citations tipping number). Results for real cases are shown in Table 7.

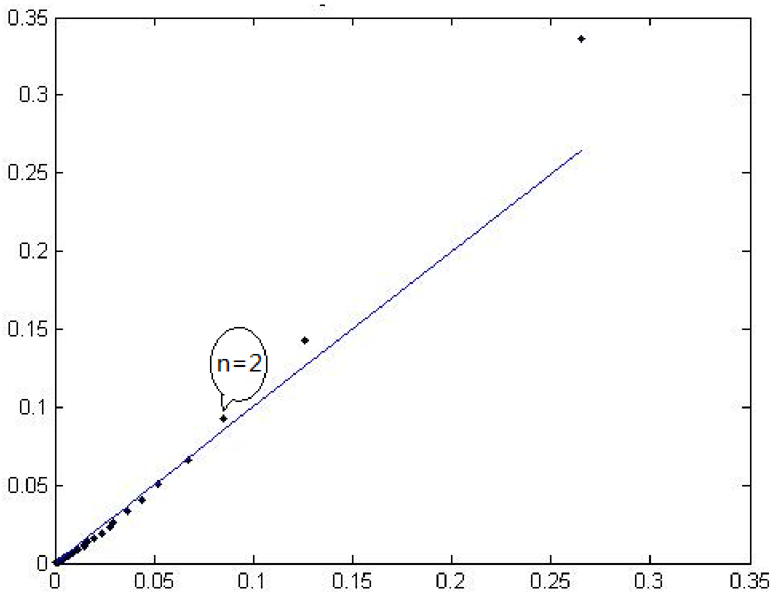

Figure 5: Scatterplot for the Field of Agricultural Sciences

Table 7: Total Citations Tipping Number (TIP) and Average Number of Total Citations per Author (denoted as ā) over the Whole Field and the Average Number of Total Citations of Authors Belonging to First-Rate Organizations, Denoted as Ā

Field	TIP	ā	Ā
Agricultural Sciences	2	6.07	9.09
Biology & Biochemistry	4	12.33	17.26
Chemistry	5	18.60	28.04
Computer Science	3	5.98	8.67
Economics & Business	3	6.27	10.99
Engineering	2	7.03	9.41
Environment/Ecology	5	10.25	15.44
Geosciences	6	13.58	21.86
Immunology	6	15.17	24.30
Materials Sciences	5	13.15	20.99
Mathematics	2	5.06	7.71
Microbiology	6	12.03	19.35
Molecular Biology & Genetics	3	10.22	16.84
Multidisciplinary	7	54.65	73.85
Neuroscience & Behavior	6	15.35	23.23
Pharmacology & Toxicology	2	9.02	13.73
Physics	9	34.94	61.27
Plant & Animal Science	3	6.89	10.05
Psychology/Psychiatry	2	10.75	16.81
Space science	19	86.22	161.27

It was found that:

a) For each subject, there is a number k, henceforward denoted as TIP (which we call the total citation tipping number) such that for n taking values from 0 to TIP the percentage of scientists with a total of n citations is larger in the whole field than among scientists working at top organizations. For n larger than TIP the opposite holds (although a few exceptions may occur). This clearly shows that the group of top organizations has relatively more excellent scientists and relatively fewer researchers with poor performance than the average level of all organizations in the field.
b) The total citations tipping number differs among fields for reasons of different citing styles. This is illustrated as follows. We calculated for each field the average number of total citations per author (denoted as $\bar{a}$) and the average number of total citations of authors belonging to first-rate organizations, denoted as $\bar{A}$ (obviously the second average is larger than the first one, see Table 7). The Pearson correlation coefficient for the total citations tipping number (TIP) and these averages are 0.913 (TIP and $\bar{a}$) and 0.945 (TIP and $\bar{A}$).
c) If among the group of scientists with 0 or 1 citation there would not be one of the top institutes (or just a tiny percentage) then the dots with labels 0 or 1 would be on the ordinate axis (or very near to it). Yet this is clearly not the case: see Figure 5; other fields have similar graphs.

We see that just like most high-impact factor journals have uncited or barely cited articles it is even more the case that first rate universities harbor "not-so-first-rate" scientists.

CONCLUSION

We studied the question if first rate scientists work at first rate organizations and found that the answer is positive, at least to a large extent. In many fields more than 80% of the top scientists indeed work at the top organizations. This is especially true for large fields such as chemistry, neuroscience, materials science, physics and plant & animal science. In mathematics and computer science this correspondence holds to a lesser degree. We assume that this difference has to do with the natural size of teams (small in mathematics and computer science) and the cost and sophistication of the equipment to perform investigations. As an aside we would like to point out that the ESI field of computer science consists mainly of the subfields: theory & methods, information systems and software engineering. It is an academic field that should not be confused with the business of consumer electronics.

We note, however, the following research limitation. Another operationalization of first rate organizations and first rate scientists would certainly yield other quantitative results. We believe, however, that they would not be qualitatively different.

As in all human and social affairs results are not completely black or white, and also small teams working at lesser known universities can occasionally obtain scientific results of the highest creative and intellectual achievement. Hence as a form of science policy advice we note that focusing too much on elite institutes may backfire (and more so in fields where teams are generally small). Yet, in general terms, this investigation confirms the skewness of science, also for the relation between first rate scientists and the organizations to which they belong.

ACKNOWLEDGEMENT

We would like to thank Qiuru Peng, Hui Lin, Xueqin Jiang, and Zengli She from the College of Information Science and Technology (Nanjing Agricultural University) for their work on data collection and verification. Work of Ronald Rousseau is supported by National Natural Science Foundation of China, grants 71173185 and 71173154, and work of Bo Yang is supported by National Social Science Foundation of China under grant 13CTQ031.

REFERENCES

Abramo, G., Cicero, T. and D'Angelo, C. A. 2012. The dispersion of research performance within and between universities as a potential indicator of the competitive intensity in higher education systems. *Journal of Informetrics*, Vol. 6, no. 2: 155-168.

Abramo, G., Cicero, T. and D'Angelo, C. A. 2013. The impact of unproductive and top researchers on overall university research performance. *Journal of Informetrics*, Vol. 7, no. 1: 166-175.

Aziz, N. A, and Rozing, M. P. 2012. Profit (p)-index: The degree to which authors profit from co-authors. *PLOS ONE* Vol. 8, no. 4: e59814.

Bookstein, F. L., Seidler, H., Fieder, M., and Winckler, G. 2010. Too much noise in the Times Higher Education rankings. *Scientometrics*, Vol. 85, no. 1: 295-299.

Bornmann, L. and Bauer, J. (in press). Which of the world's institutions employ the most highly cited researchers? An evaluation of the data from highlycited.com. *Journal of the Association for Information Science and Technology.*

Bradford, S. C. 1934. Sources of information on specific subjects. *Engineering,* Vol. 137, 85-86.

Buela-Casal, G., Gutierrez-Martinez, O., Bermudez-Sanchez, M. P. and Vadillo-Munoz, O. 2007. Comparative study of international academic rankings of universities. S*cientometrics*, Vol. 71, no. 3: 349-365.

Butler, D. 2010. University rankings smarten up. *Nature*, Vol. 464: 16-17.

Dehon, C., McCathie, A. and Verardi, V. 2010. Uncovering excellence in academic rankings: a closer look at the Shanghai ranking. *Scientometrics,* Vol. 83, no. 2: 515-524.

Gaston, J. 1978. *The reward system in British and American science*. New York: John Wiley & Sons.

Hirsch, J.E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the USA*, Vol. 102, no. 46: 16569-16572.

Huang, SQ., Yang, B., Yan, SL. and Rousseau, R. 2014. Institution name disambiguation for research assessment. *Scientometrics*, Vol. 99, no.3: 823-838.

The KNUDOP Search Group. 2008. Podunk effect. *ISSI Newsletter*, Vol. 4, no. 2: 23.

Larivière, V., Macaluso, B., Archambault, E., and Gingras, Y. 2010. Which scientific elites? On the concentration of research funds, publications and citations. *Research Evaluation*, Vol. 19, no. 1: 45-53.

Liang, L.M. and Rousseau, R. 2009. A general approach to citation analysis and an h-Index based on the standard impact factor framework. In: *ISSI 2009; 12th International Conference on Scientometrics and Informetrics* (Birger Larsen & Jacqueline Leta, eds.), Rio de Janeiro: BIREME & Federal University of Rio de Janeiro, 143-153.

Liu, Y.X. and Rousseau, R. 2010. Knowledge diffusion through publications and citations: a case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology,* Vol. 61, no. 2: 340-351.

Mahbuba, D. and Rousseau, R. 2011. The Matthew effect and a relation with concept symbols and defaults. *Annals of Library and Information Studies*, Vol. 58, no. 4: 335-345.

Mahbuba, D., Rousseau, R. and Srivastava, D. 2010. A scientometric study of health and population research in South Asia: focus on two research organizations. *Malaysian Journal of Library Information Science*, Vol. 15, no. 3: 135-147.

Merton, R. K. 1968. The Matthew effect in science. *Science,* Vol. 159: 56-63.

Merton, R. K. 1988. The Matthew effect in science II. Cumulative advantage and the symbolism of intellectual property. *ISIS,* Vol. 79, no. 4: 606-623.

Prathap, G. 2006. Hirsch-type indices for ranking institutions' scientific research output. *Current Science*, Vol. 91, no. 11: 1439.

Rodriguez-Navarro, A. 2011. A simple index for the high-citation tail of citation distribution to quantify research performance in counties and institutions. *PLOS ONE,* Vol. 6, no. 5: e20510.

Rodriguez-Navarro, A. 2012. Counting highly cited papers for university research assessment: conceptual and technical issues. *PLOS ONE*, Vol. 7, no. 10: e47210.

Seglen, P. O. 1992. The skewness of science. *Journal of the American Society for Information Science*, Vol. 43, no. 9: 628-638.

Seglen, P. O. 1994. Causal relationship between article citedness and journal impact. *Journal of the American Society for Information Science,* Vol. 45, no. 1: 1-11.

Smalheiser, N. R. and Torvik, V. I. 2009. Author name disambiguation. *Annual Review of Information Science and Technology,* Vol. 43, no. 1: 1-43.

Strotmann, A. and Zhao, DZ. 2012. Author name disambiguation: What difference does it make in author-based citation analysis? *Journal of the American Society for Information Science and Technology,* Vol. 63, no. 9: 1820-1833.

Thomson Reuters. 2014. *The world's most influential scientific minds*. Bethesda, MD, USA: Thomson Reuters.

Van Parijs, P. 2009. European higher education under the spell of university rankings. *Ethical perspectives*, Vol. 16, no. 2: 189-206.

Van Raan, A. F. J. 2005. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, Vol. 62, no. 1: 133-143.

Vieira, E.S and Gomes, J. A. N. F. 2011. An impact indicator for researchers. *Scientometrics*, Vol. 89, no. 2: 607-629.

Yazit, N. and Zainab, A. N 2007. Publication productivity of Malaysian authors and institutions in LIS. *Malaysian Journal of Library & Information Science*, Vol. 12, no. 2-3: 35-55.