The Geography of Scientific Citation

Michael Batty

19 December 2002

As the marketplace for research and education becomes ever more global, there is growing interest in international comparisons of standards and quality in different countries and places. Rankings of universities according to various criteria ranging from the quality of the educational experience imparted to their research prowess have become regular news items in English-speaking countries[1]. In some countries such as the UK where the core funding of universities is from central government, such assessments are now used routinely for resource allocation[2].

International comparisons are however difficult with few published rankings despite rapidly increasing international migration to pursue research at the graduate level. In the US, 20% of all full-time graduate students are non-US citizens[3]; in the UK, the comparable figure is 25%[4]. As there are no global rankings, most decisions to pursue research at a particular institution must be based on casual perceptions of quality, cost, and overall value for money. To examine the research quality of universities world-wide, citation indices provide a first approach to the problem[5]. The ISI's *HighlyCited* database (http://www.isihighlycited.com) which is currently composed of the top 100 or so cited individuals in 14 scientific fields is a manageable source for classifying scientists not only by their field but by their institution, their location, and the country in which they work.

Notwithstanding the limitations of this source for it excludes mathematics (other than physics), the social sciences and the humanities, and is thus biased towards the medical sciences, we have taken data from 12 of the 14 categories listed. From a detailed scrutiny of each entry, we have used data on 1222 scientists. A significant minority of the scientists cited – some 30 percent – work in research institutes, hospitals, and private firms, albeit many connected to university institutions, but to maintain comparability between cities and countries, we have retained this data.

The pattern of concentration that this analysis reveals is remarkable. 1222 scientists work in 429 institutions which are located in 232 places in 27 countries. Almost half these scientists are in 50 institutions in 5 countries, most being in the United States. In Table 1, we list the top 20 institutions in terms of the number and percent of scientists cited. These institutions contain nearly 30 percent of the scientists, and are all located in the US with the exception of University College London and the University of Cambridge. The concentration increases as we aggregate the data from institution to place and thence to country. In Figure 1, we show these aggregated data sets as Zipf plots where we have plotted the logarithm of the number of scientists for each institution, place and country, normalised by their means, against the logarithm of their normalised ranks. Collapsing each data set in this way shows quite clearly how the concentration increases as we aggregate the data into places and countries. We have fitted power laws to these plots based on $\left(P(x)/\langle x\rangle\right) \sim \left(r/M\right)^{-\alpha}$ where $P(x)$ is the number of cited scientists at rank r, $\langle x\rangle$ is the mean number of cited scientists, and M is the number of

institutions, places, or countries for each of the three respective aggregations[6]. The value of the power α is related to the degree of concentration.

In Table 2, we show the top ten countries in terms of the number of scientists and places where they work. The ratio of scientists to places provides another measure of the concentration with the implication that the larger the number of highly cited in each country, the more likely they are to be highly concentrated in a small number of places. If we normalise the data by populations, we get a slightly different picture with smaller countries like Sweden and Switzerland becoming more significant although the US still dominates. A similar analysis for institutions and places is more complicated as the choice of a population for the normalisation is uncertain. College towns begin to dominate, for example.

A particularly graphic indication of the basic pattern is illustrated in Figure 2 where we have mapped the main locations of places by circles proportional to the number of cited scientists. 40% of the most highly cited scientists work in 10 places of which 9 are in the US. These locations bear out our perceptions of where the world's top institutions are: on the west coast of the United States, the Boston-Washington megalopolis on the east coast, central London, Chicago, and interestingly in the cluster of towns around Research Triangle Park in North Carolina. We have not yet examined the local detail of where these institutions are located but casual knowledge suggests that these are even more highly clustered. For example, the institutions in Boston are all with a two mile radius of the MIT Museum whereas in London they are within a three mile radius of the British Museum. At an even more local scale in central London, for example, the majority of the scientists cited are located within half a mile of Euston station (including *Nature's* editorial offices !).

Our analysis is of course limited by the bias in the ISI data to English-speaking countries, to the medical sciences, and to full time research rather than education. Although for US institutions, there is only a 40 percent correlation with the top 50 universities in terms of doctoral programs most recently ranked by *U. S. News and World Report*, this simply indicates the fact that size is all important in the rankings produced from the ISI data[7]. The correlation in the UK with *The Times* ranking is much the same at 43 percent[8]. What this analysis reveals is a pattern of much greater concentration than we originally anticipated from other literature on the geography of the modern economy, notwithstanding the influence of history and the effects of national policy on the location of research centres[9].

We consider there are important implications from these findings for national educational policy, and the distribution of research resources, especially during a period when governments and institutions are competing every more intensely to gain and retain the best, and to build critical mass. There are issues involving the choice of the best graduate schools implied in our analysis. This analysis puts the geographical distribution of scientific wealth in perspective[10]. In a British context, it will be surprising to many academics and politicians that of the 1222 scientists cited, only 100 (about 7 percent) are located in the UK based in 38 institutions (about 9 percent) of the 429 associated with these citations.

Michael Batty, Centre for Advanced Spatial Analysis at University College London, 1-19 Torrington Place, London WC1E 6BT (email: m.batty@ucl.ac.uk)

Supplementary information is available at http://www.casa.ucl.ac.uk/citations/

References

[1] Graduate and Research Program Rankings, Education and Social Science Library, University of Illinois at Urbana-Champaign http://www.library.uiuc.edu/edx/rankgrad.htm , accessed 17 December 2002

[2] Higher Education and Research Opportunities, UK Research Assessment Exercise http://www.hero.ac.uk/rae/index.htm , accessed 17 December 2002

[3] National Center for Education Statistics, Institute of Education Sciences, US Dept. of Education, Washington DC http://nces.ed.gov/pubs2002/2002212.pdf, accessed 17 December 2002

[4] Higher Education Statistics Agency, http://www.hesa.ac.uk/holisdocs/pubinfo/stud.htm , accessed 17 December 2002

[5] Andrew Oswald, O. A crisis of quality. *Education Guardian*, Friday November 15 (2002) http://education.guardian.co.uk/higher/comment/story/0,9828,840892,00.html, accessed 17 December 2002

[6] Redner, S. How popular is your paper? An empirical study of the citation distribution. *European Physics Journal B* **4**, 131-134 (1998).

[7] *U. S. News and World Report* http://www.usnews.com/usnews/edu/college/rankings/ranknatudoc_brief.php; accessed 17 December 2002

[8] *The Times Newspaper Good University Guide* http://www.times-archive.co.uk/news/pages/tim/2000/04/14/timguggug01002.html accessed 17 December 2002

[9] Matthiessen, C. W., Schwarz, A. W. Scientific centres in Europe, *Urban Studies* **36**, 453-477 (1999)

[10] May, R. M. The scientific wealth of nations. *Science* **275**, 793-795 (1997)

Table 1: Top twenty ranking of highly cited scientists by institution

Rank	Research institution	No. of highly cited scientists	Percent highly cited scientists
1	Harvard	52	4.3
2	Stanford	36	2.9
3	U-Cal, San Diego	30	2.5
4	MIT	26	2.1
5	NIH National Cancer Institute	19	1.6
6	U-Cal, San Francisco	17	1.4
	Cornell		
8	U-Cal, Berkeley	16	1.3
	University College London UK		
10	CalTech	15	1.2
11	NIH Allergy & Infectious Diseases	13	1.1
12	Johns Hopkins	12	1.0
	University of Cambridge UK		
	U-Washington, Seattle		
	Washington U, St Louis		
16	U-Cal, Davis	11	0.9
	U-Texas Cancer Center		
18	Michigan	10	0.8
	Northwestern		
	Yale		

Table 2: Top ten ranking of highly cited scientists by country

Rank	Country	No. of highly cited	No. of places	Concentration: scientists/places	Highly cited per million population
1	US	815	90	9.06	3.16
2	UK	100	24	4.17	1.72
3	Germany	62	21	2.95	0.78
4	Canada	42	15	2.80	1.53
5	Japan	34	14	2.43	0.27
6	France	29	11	2.64	0.50
8	Switzerland	26	5	5.20	3.78
9	Sweden	17	2	8.50	1.96
10	Italy	17	10	1.7	0.29

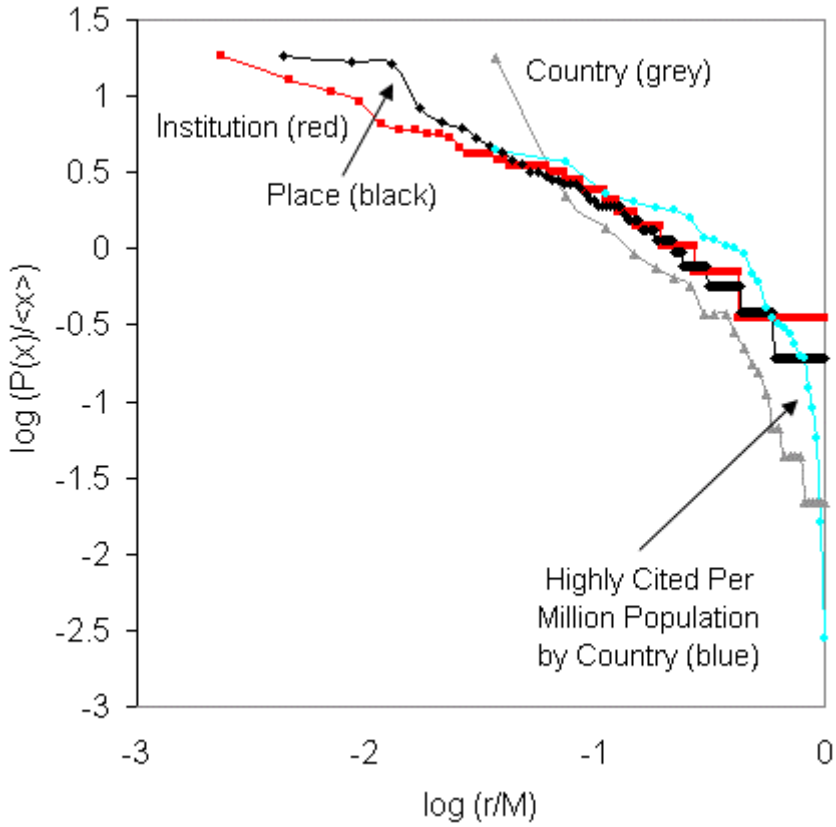

Figure 1: Rank-size distributions of highly cited scientists

The blue line is the plot for countries, normalised by population in millions, which illustrates a different pattern of concentration from the basic data. We have fitted linear plots to the basic data using $\log\left(P(x)/\langle x\rangle\right) = k - \alpha \log\left(r/M\right)$ where for institutions $\alpha = 1.049\ (R^2 = 0.962)$, for places $\alpha = 0.816\ (R^2 = 0.938)$, and for countries $\alpha = 1.997\ (R^2 = 0.949)$. All these values are significantly different from zero at the 99% level.

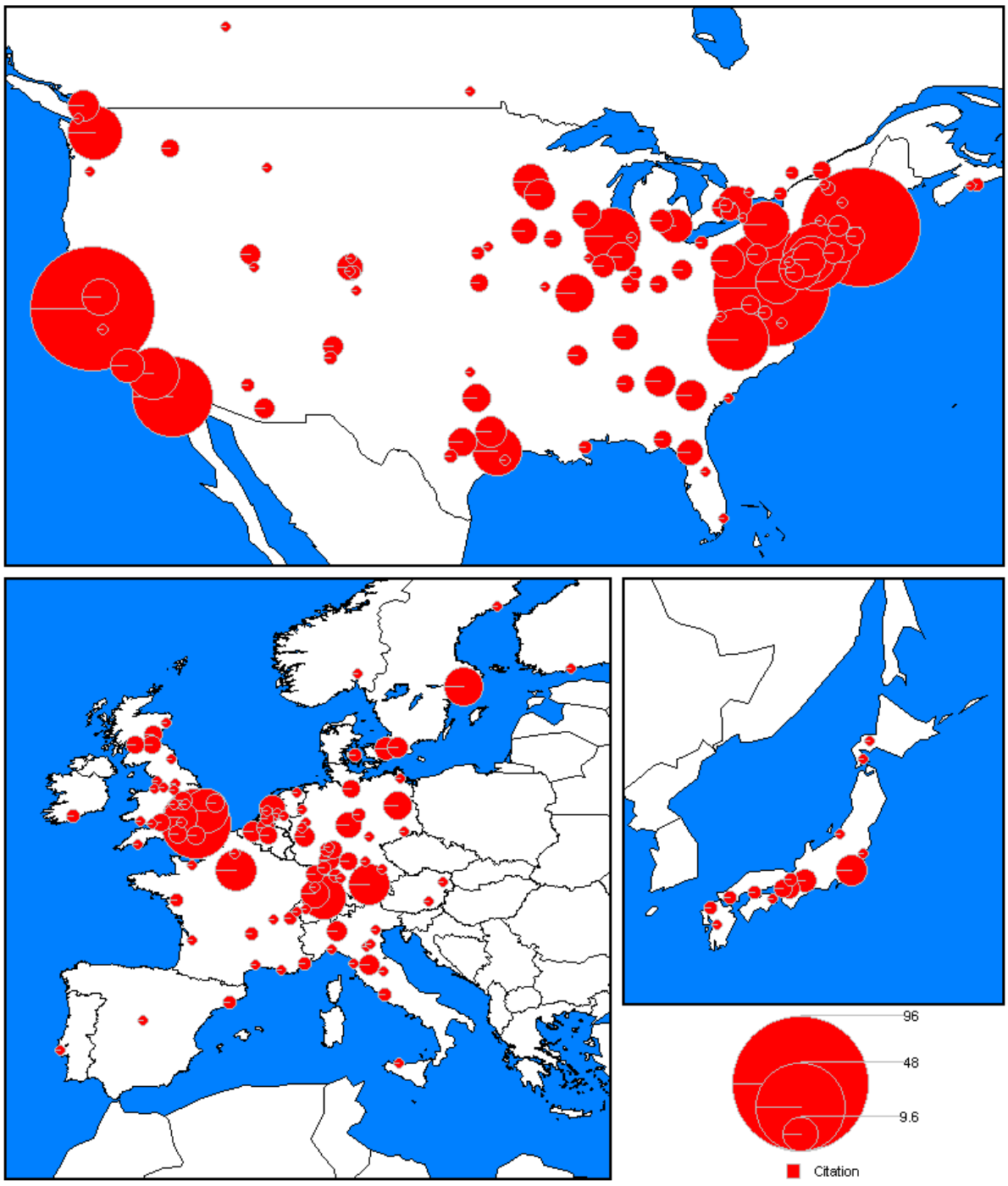

Figure 2: The geographical distribution of the highly cited