SPRU
Science and Technology Policy Research

SPRU Electronic Working Paper
Number 198

Science Policy, Complex Innovation Systems and Performance Measures

J. Sylvan Katz*

* SPRU, University of Sussex
Johnson-Shoyama Graduate School of Public Policy, University of Saskatchewan
Science-Metrix 1335 A. Mont-Royal E. Montréal, QC, Canada, H2J 1Y6

The Freeman Centre (Sussex University Campus), Falmer, Brighton, BN1 9QE
United Kingdom Tel.: +44(0)1273 877943 Fax: + 44(0)1273 877977

February, 2012

University of Sussex
SPRU – Science & Technology Policy Research

SPRU – Science and Technology Policy Research
The Freeman Centre
University of Sussex
Falmer, Brighton, BN1 9QE
http://www.sussex.ac.uk/spru

Science Policy, Complex Innovation Systems and Performance Measures

J. Sylvan Katz

SPRU, University of Sussex, Freeman Centre, Brighton, East Sussex BN1 9QE, UK

Johnson-Shoyama Graduate School of Public Policy, University of Saskatchewan, Saskatoon, SK, Canada S7N 5B8

Science-Metrix 1335 A. Mont-Royal E. Montréal, QC, Canada, H2J 1Y6

The design of effective science and innovation policy is partially predicated on the notion that decision makers have reliable evidence-based performance measures[1,2]. This is an area of intense investigation as witnessed by the recent articles on *Science Metrics* in Nature [3,4] and the NSF's *The Science of Science & Innovation Policy* (SciSIP) program[5]. The study of science systems has shown they are complex, adaptive systems with emergent properties frequently characterized by power law distributions and functions[6-8]. These properties are rarely used to prepare measures that inform policy makers. A difficulty with traditional measures such as national wealth (GDP per capita), R&D intensity (GERD/GDP) and scientific impact (citations/paper) is that they are based on measures of wealth and impact (i.e. citations) that have power law distributions[9,10]. Performance measures based on population averages derived from these distributions may have large or indefinable error limits making comparisons across groups misleading or uninformative [11,12]. This article illustrates how a scaling model of a science system constructed from 1984-2007 Web of Science (WoS) data can be used to prepare measures with error limits that provide insight into the evolution and performance of a complex science system and answer policy relevant questions[13].

Complexity scientists have found a prevalence of power law distributions in natural and social systems, particularly through the study of complex networks[10,14]. These probability distributions are characterized by $p(x)=Cx^{-\alpha}$ for $x \geq x_{min}$ where α is a constant called the scaling factor. Citation networks used to prepare measures of the impact of published research are complex networks with $\alpha \approx 3.0$ [9,15].

Power law distributions are unique; they are scale-invariant. It is the only distribution that is the same on whatever scale it is examined[10]. Random and natural populations drawn from this distribution are scale-invariant too. A natural population is one that preserves the clustering, 'community' or small world structure which is unlike a population drawn from a truncated distribution[16,17].

Power law distributions with $\alpha<3$ have infinite variance, a condition that doesn't satisfy the central limit theorem, and error limits for the population averages cannot be estimated[18]. Even with this knowledge indicator developers sometimes truncate the tail of a citation distribution, model the remaining population as if it satisfies the central limit theorem and then produce population average based performance measures[19]. This action removes the highest impact and perhaps most innovative entities from consideration and distorts the information presented to decision makers. Similarly, normalized performance measures based on the ratio of population averages from these distributions cannot be assigned confidence intervals. They may have little informative value too[20]. The h-index, frequently used in the evaluation of individuals, is one of the few measures that partially accounts for the underlying distribution[21,22].

Science policy is often informed by indicators such as scientific impact which is constructed from ratios of two primary measures: citations and papers. Sometimes a collection of such ratios used to characterize groups in a system may show a scaling correlation between the numerators and denominators. For example, impact, C, measured using citations has been shown to scale with group size, P, measured using numbers of peer-reviewed publications[23,24]. This property is considered a prevalent property of a science system and the scaling factor is a measure of the Matthew effect in science[13,25]. Given $C \approx P^{\alpha}$ then $C/P \approx P^{\alpha-1}$ and if $\alpha \neq 1.0$ then we know scientific impact, C/P, scales nonlinearly with group size too.

Frequently decision makers are given reports with tables of groups (i.e. fields, subject areas, nations, institutions, etc.) ranked by their scientific impact. Shortly an example will be given of the different perspective a decision maker gets when groups are ranked by an impact measure that is adjusted for the nonlinear scaling effects of group size instead of using scientific impact.

Let's examine a model of a global science system derived from the 1984-2007 WoS and built using two scaling principles. A citation network was constructed from 10.9 million peer-reviewed[a] publications indexed between 1984 and 2002 that were cited more than 118 million times by publications indexed between 1984 and 2007. Thomson Reuters has assigned each document to one of 138 subject categories that can be aggregated into thirteen academic fields using a scheme[b] developed for the National Science Foundation[26]. Citations to publications in a given year were counted using a fixed 6-year citation window[c]. Other models can be constructed using different citation windows and document classifications.

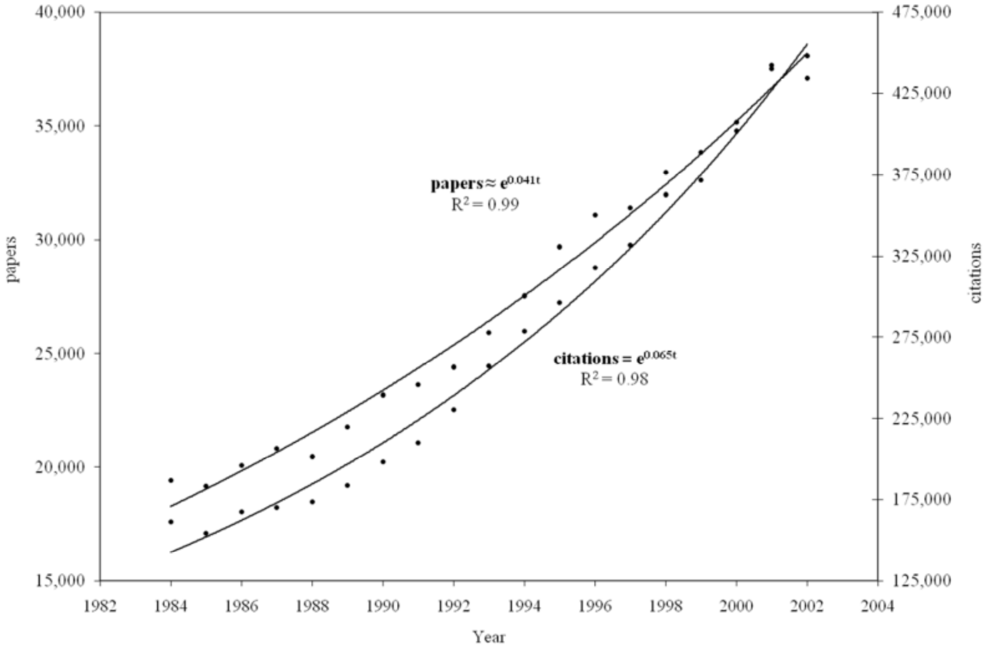
The WoS data are relatively free of measurement error. The scaling factor for a power law distribution is best determined using a maximum likelihood estimation method[10]. However, the noiseless nature of the WoS data makes ordinary least squares regression on log transformed data sufficient for estimating the parameters of the scaling correlation between citations and papers[10,27].

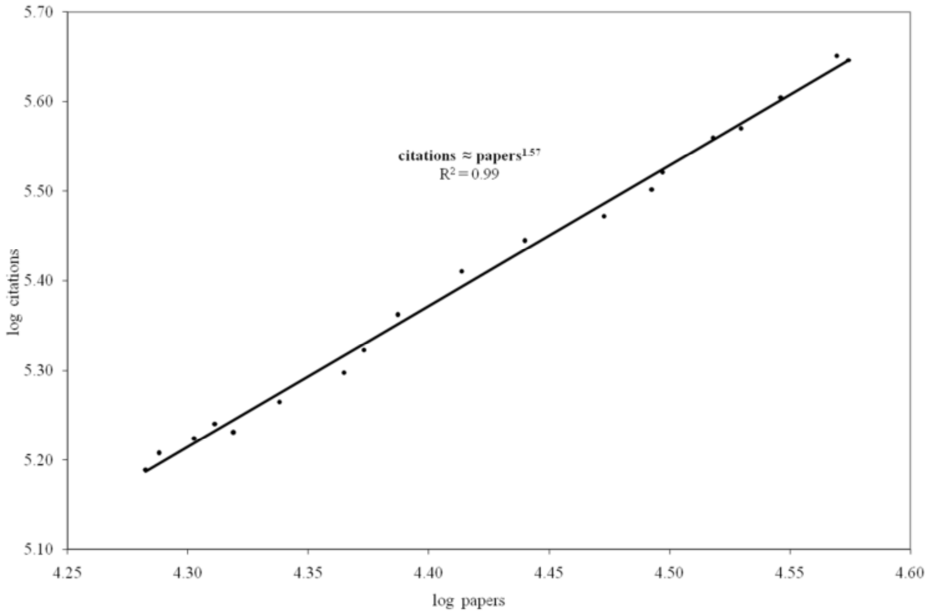
Two scaling principles are used to construct the model. The first principle is based on a power law function that exists between any pair of coupled exponential growth processes[28,29]. The scaling factor is given by the ratio of the exponents of the individual exponential processes. This is illustrated in Figure 1. Figure 1a depicts the exponential growths of Earth & Space sciences papers and 6-year citations; Figure 1b is a log-log plot of 6-year citation versus paper counts. The scaling factor predicted by the ratio of the exponential growth exponents (0.065/0.041) is 1.58. The measured α is 1.57 ± 0.03. The scaling factor indicates that given current growth rates if the Earth & Space sciences

[a] Articles, notes and review source document types that were cited at least once
[b] Recently updated by Science-Metrix, Montreal, Canada
[c] Citations in the year of publication and subsequent five years were counted. Author self-citations were not removed

community published twice as many papers their impact would be expected to increase $2^{1.57}$ or 3.0 times.

Figure 1 Exponential growth and scaling. **a** growth of earth & space papers and citations and **b** scaling function between the growths of citations and papers.





Values of α, standard errors, R^2 and scientific impact for 13 NSF fields ranked in descending order of α are given in Table 1. The scaling factor can be used as a measure of the relative growth of a field's size and its impact on the research community. This measure is called a scale-independent measure and it can be compared across fields of different sizes without additional adjustment. The relative growth of a group's impact to its growth in size is indicative of such things as the rate of innovation and utility of research to others. Notice the magnitude of the scientific impact is not a predictor of α.

Table 1 – Field scaling factor

NSF Field	α	R^2	C/P
Biology	2.92 ± 0.09	0.98	6.8
Professional Fields	2.46 ± 0.17	0.92	5.0
Psychology	2.30 ± 0.18	0.90	8.1
Humanities*	2.19 ± 0.16	0.91	4.0
Chemistry	2.12 ± 0.07	0.98	9.2
Clinical Medicine	2.06 ± 0.06	0.99	12.3
Biomedical Research	1.83 ± 0.04	0.99	20.7
Social Sciences	1.79 ± 0.04	0.99	4.6
All Fields	1.79 ± 0.04	0.99	10.7
Earth & Space	1.57 ± 0.03	1.00	10.0
Health Sciences	1.55 ± 0.05	0.98	6.5
Mathematics	1.55 ± 0.07	0.96	3.9
Engineering & Technology	1.41 ± 0.06	0.97	4.8
Physics	1.13 ± 0.05	0.97	9.5

* Publications and citations declined with time

Using Earth and Space sciences as an example, the scaling function between C and P tells us that since $C/P \approx P^{\alpha-1}$ then the growth in scientific impact scales with size and it has a scaling factor of 0.58. A doubling of size would be expected to produce ($2^{0.58}$) 1.5 times increase in scientific impact. Also, since the exponential growth of all NSF fields and their impacts are coupled in the time domain the same principle can be used to prepare a variety of scale-independent measures of the relative growths of sizes and/or impacts for pairs of fields.

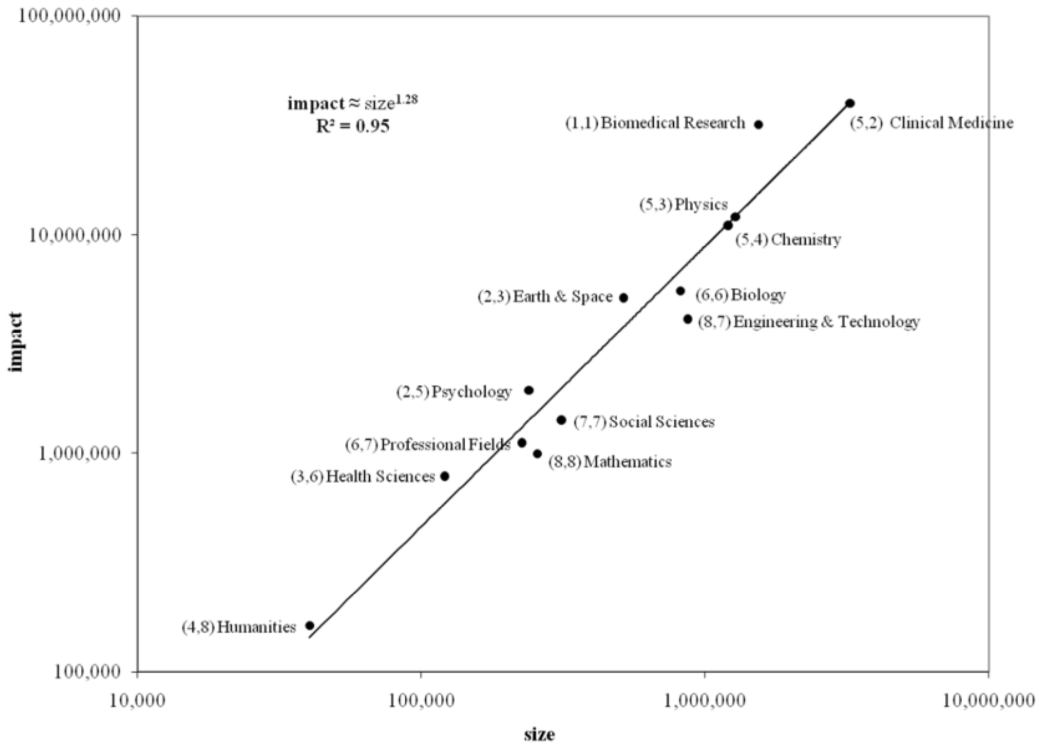
The second principle is based on empirical and simulation observations that complex weighted networks exhibit degree-strength scaling correlations[30]. Additional evidence comes from the study of web networks that show the Gini coefficient increases when nodes are aggregated into logical groups[7]. Typically a complex network such as a citation network is represented by a binary graph where the edges have a value 1 if they exist and 0 otherwise. When nodes are aggregated into groups the result is a complex network with weighted edges that exhibit a scaling function between node strength (e.g. scientific productivity) and connectivity (e.g. node degree, citations)[31,32]. The scaling factor for the function contains information about the topology of the network (e.g. small world structure) as well as the sizes of the nodes and weights of the edges.

Consider a binary citation network where papers are aggregated into groups defined by WoS subject categories or into NSF fields to produce a weighted network. Subject categories and NSF fields are natural populations selected using citation patterns and expert knowledge. The nodes represent the publications in each category/field and their sizes are the number of papers they contain. The weights of incoming and outgoing edges are the number of citations from or to other groups, respectively. Node loops are weighted by the number of citations between papers within a group. It has been found that citation networks exhibit scaling correlation between node degree (sum of incoming weights and loop weight) and node strength or size (numbers of publications)[13,24].

The second principle, illustrated in Figure 2, shows the scaling correlation between field impact and size[d]. The annual values were summed over the interval (Shapiro-Wilks test $W=0.97$ and $p=0.90$ for overall values). An examination of individual years showed α increased over time from 1.25 to about 1.32 (see insert in Figure 3) with an average value of 1.28 ± 0.09 shown in Figure 2. A doubling of field size is expected to increase the impact $2^{1.28}$ or 2.43 times.

[d] Size was measured using papers that received at least one citation. Cited and uncited, or only papers that received n or more citations could have been used, however, the parameters of the scaling function would differ.

Figure 2 – Scaling function between NSF field sizes and impacts



The regression line represents the systemic scaling tendency across all of the fields in the system. It can be used to calculate the expected impact of each field. The ratio between the observed and expected impact values is a scale-independent measure of the impact of a field relative to the systemic scaling trend. It can be used to rank the fields. The numbers in brackets separate by commas in Figure 2 are the field ranks determined using (1) the scale-independent impact measure and (2) the scientific impact, respectively. The difference in the field ranks using the scale-independent measure compared to the traditional measure is readily apparent. The ranks of 9 of 13 NSF fields changed by 2 or more positions which should give decision makers pause for thought.

Figure 3 is a scaling model of a global science system composed of two groups of elements. The first group is constructed from log-log plots of the relative growth of impact and size for the NSF fields whose scaling factors are given in Table 1. The second element is an overlay of the 1984 and 2002 systemic power regression lines.

Figure 3 – A Scaling model of a complex science system. The model is composed of the scaling functions between the growths of impact and size for NSF fields overlaid (dotted lines) with the 1984 and 2002 systemic scaling functions.

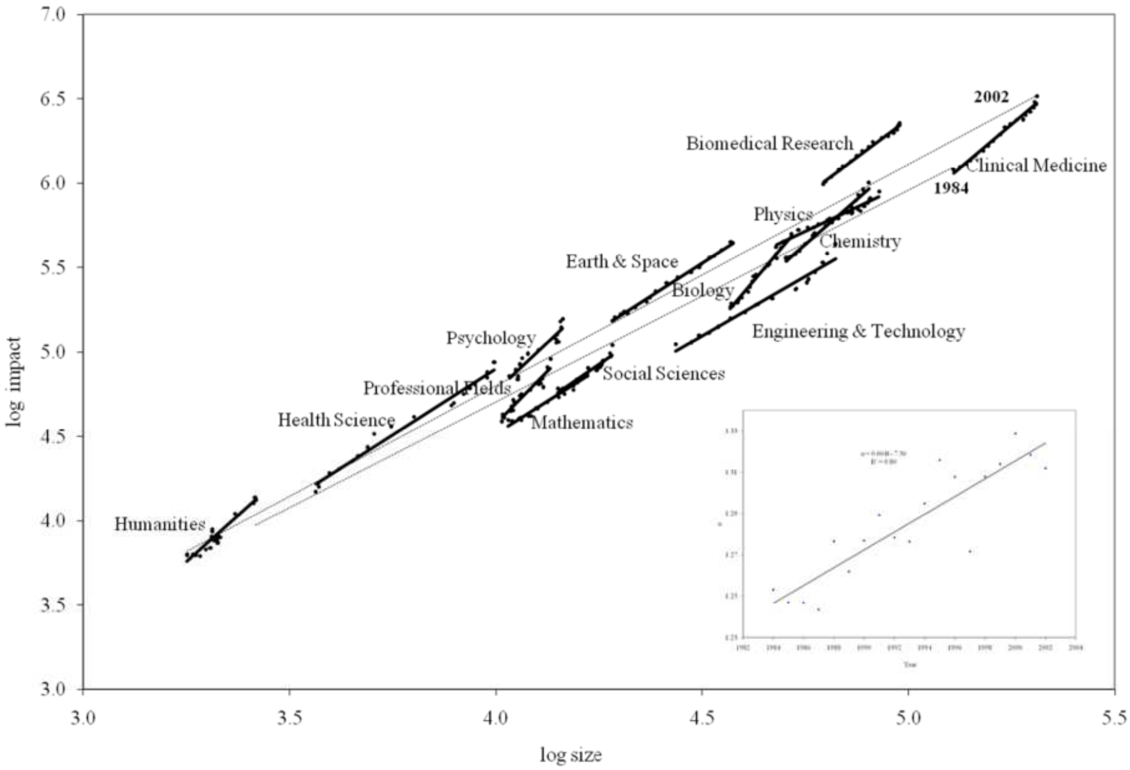


Figure 3 illustrates how a scaling model of a complex global science system can be constructed from 26 years of WoS data and a small collection of power law functions. It completely encapsulates the evolution of (1) each field's impact relative to its size and (2) each field's impact relative to the impact of other fields. Similar scaling models could be constructed for other groups such as nations and regions providing insights into the evolution and performance of national and regional science systems.

A model can be constructed for 138 WoS subject categories however it is difficult to display graphically. Also, at this level of disaggregation the statistical significance decreases. Eighty-two (60%) of the subject categories had $R^2 \geq 0.90$ and 111 (80%) categories had $R^2 \geq 0.80$. Still a significant amount of information can be gleaned from the model. For example, consider the top 20 of the 84 subfields with $R^2 \geq 0.90$ ranked in descending order of α in Table 3. Relative to the growth of its size agriculture and food science demonstrated the greatest increase in impact increasing tenfold ($2^{3.33}$) every time its size doubled. The systemic scaling factor across the 138 categories was 1.27±0.03 (R^2=0.89) determined using the method described for Figure 2.

Table 2 –Twenty high impact WoS subject categories

Subject Category	**NSF Field**	**α**	**R^2**
Agriculture & Food Science	Biology	3.33 ± 0.22	0.93
Applied Chemistry	Chemistry	2.91 ± 0.19	0.93
Geography & Regional Sciences	Social Sciences	2.80 ± 0.21	0.91
Orthopedics	Clinical Medicine	2.79 ± 0.16	0.95
Miscellaneous Biomedical Research	Biomedical Research	2.68 ± 0.13	0.96
Nutrition & Dietetics	Biomedical Research	2.50 ± 0.13	0.96
Social Studies of Medicine	Health Sciences	2.31 ± 0.13	0.95
Urology	Clinical Medicine	2.31 ± 0.08	0.98
Experimental Psychology	Psychology	2.22 ± 0.12	0.96
Geriatrics	Clinical Medicine	2.15 ± 0.10	0.96
Ecology	Biology	2.12 ± 0.07	0.98
Psychiatry	Clinical Medicine	2.07 ± 0.14	0.93
Dermatology & Venereal Diseases	Clinical Medicine	2.03 ± 0.08	0.97
Management & Business	Professional Fields	2.01 ± 0.11	0.95
Miscellaneous Clinical Medicine	Clinical Medicine	2.00 ± 0.08	0.97
Surgery	Clinical Medicine	2.00 ± 0.06	0.98
Environmental & Occupational Health	Clinical Medicine	1.98 ± 0.12	0.95
Miscellaneous Social Sciences	Social Sciences	1.96 ± 0.15	0.91
Otorhinolaryngology	Clinical Medicine	1.92 ± 0.09	0.97
Miscellaneous Biology	Biology	1.90 ± 0.05	0.99

Scaling models of innovation systems have been constructed using primary measures that are used to prepare indicators of national wealth, R&D intensity, Web visibility and scientific impact[6-8]. These models and scale-independent measures derived from them capture and quantify emerging properties of complex innovation systems that traditional measures can't. Furthermore, scale-independent measures can be assigned error limits giving users a reliable confidence interval.

Measures from scaling models answer a variety of policy relevant questions such as 'how does a primary measures such as impact, GERD and GDP that are commonly used to compare the performance of groups in a complex innovation system change relative to group sizes?' Or 'how is the performance of one group evolving relative to other groups given the systemic scaling tendency of the system?' Scaling models are easy to update. They can be used to anticipate how a complex science system might evolve if the parameters remain as measured or if they change under a variety of policy regimes. Scaling models and scale-independent indicators are useful support tools for decision makers to inform policies for complex science and innovation systems.

Acknowledgements

The authors would like to thank Eric Archambault, Viv Cothey, Diana Hicks and Juan Rogers for valuable comments on earlier drafts of the paper and Jean-Jacques Gervais, David Campbell and Hicham Sabir at Science-Metrix for assistance in preparing the citation data and statistical advice.

References

1. Husbands Fealing, K., Lane, J., Marburger, J. & Shipp, S. (eds.) *The Science of Science Policy - A Handbook* (Stanford University Press, 2011).
2. Lane, J. Let's make science metrics more scientific. **464**, 488-489 (2010).
3. Van Noorden, R. Metrics: A profusion of measures. *Nature* **465**, 864-866 (2010).
4. Abbot, A. et al. Metrics: Do metrics matter? *Nature* **465**, 860-862 (2010).
5. Lane, J. The Science of Science and Innovation Policy (SciSIP) Program at the US National Science Foundation. *Bridges* **22** (2009).
6. Katz, J. S. Indicators for complex innovation systems. *Research Policy* **35**, 893-909 (2006).
7. Katz, J. S. & Cothey, V. Web Indicators for Complex Innovation Systems. *Research Evaluation* **14**, 85-95 (2006).
8. Gao, X., Guo, X., Katz, J. S. & Guan, J. The Chinese innovation system during economic transition: A scale-independent view. *Journal of Informetrics* **4**, 618-628 (2010).

9. Clauset, A., Shalizi, C. S. & Newman, M. E. J. Power-law distributions in empirical data. *SIAM Review* **51**, 661-703 (2009).
10. Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **46**, 323-351 (2005).
11. van Raan, A. F. J. Measuring science: Capita selecta of current main issues. *In H.F. Moed, W. Glänzel, & U. Schmoch (Eds.), Handbook of quantitative science and technology research (pp. 19–50). Springer.* (2005).
12. Opthof, T. & Leydesdorff, L. Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of Informetrics* **5**, 101-113 (2011).
13. Katz, J. S. Scale independent indicators and research assessment. *Science and Public Policy* **27**, 23-36 (2000).
14. Baranger, M. Chaos, Complexity, and Entropy: A physics talk for non-physicists. *Wesleyan University Physics Dept. Colloquium*, available at http://necsi.org/projects/baranger/cce.pdf (2001).
15. Redner, S. How popular is your paper? An empirical study of the citation distribution. *The European Physical Journal B - Condensed Matter and Complex Systems* **4**, 1434-6028 (1998).
16. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 7821-7826 (2002).
17. Palla, G., Derenyi, I., Farkas, I. & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814-818 (2005).
18. Willinger, W., Alderson, D., Doyle, J. C. & Lun, L. in *Proceedings of the 2004 Winter Simulation Conference* (ed. R. G. Ingalls, M. D. R., J. S. Smith, and B. A. Peters, eds.) (2004).
19. van Raan, A. F. J. Statistical properties of bibliometric indicators: Research group indicator distributions and correlations. *Journal of the American Society for Information Science and Technology* **57**, 408-430 (2006).
20. Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S. & van Raan, A. F. J. Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics* **5**, 37-47 (2011).
21. Hirsch, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 16569-16572 (2005).
22. Peterson, G. J., Press, S. & Dill, K. A. Nonuniversal power law scaling in the probability distribution of scientific citations. *Proceedings of the National Academy of Sciences* **107**, 16023-16027 (2010).
23. Katz, J. S. The Self-Similar Science System. *Research Policy* **28**, 501-517 (1999).
24. van Raan, A. F. J. Bibliometric Statistical Properties of the 100 Largest European Research Universities: Prevalent Scaling Rules in the Science System. *Journal of the American Society for Information Science* **59**, 461-475 (2008).
25. Merton, R. K. The Matthew Effect in Science. *Science* **159**, 56-63 (1968).

26. Narin, F. Evaluative Bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity, (NSF Contract NSF C-627, Project No. 704R). (1976).
27. Warton, D. I., Wright, I. J., Falster, D. S. & Westoby, M. Bivariate line-fitting methods for allometry. *Biological Reviews* **81**, 259-291 (2006).
28. Naranan, S. Bradford's Law of Bibliography of Science: an Interpretation. **227**, 631-632 (1970).
29. Katz, J. S. Scale independent bibliometric indicators. *Measurement: Interdisciplinary Research and Perspectives* **3**, 24–28 (2005).
30. Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 3747-3752 (2004).
31. Bianconi, G. Emergence of weight-topology correlations in complex scale-free networks. *Europhysics Letters* **71**, 1029-1035 (2005).
32. Barrat, A., Barthélemy, M. & Vespignani, A. Modeling the evolution of weighted networks. *Physical Review E* **70**, 066149 (2004).