Productivity of research groups — relation between citation analysis and reputation within research communities

Junichiro Makino *

April 28, 1998

Abstract

In this paper I discuss the relation between widely used "Scientometric" measures and "reputation" of research groups within the scientific community. To this goal, I present the result of the detailed comparison of two research groups of theoretical astrophysics in post-world-war-2nd Japan. Though one of the two groups gained much higher reputation within the research community, we could not find much difference in the the macroscopic indices such as the number of publications or the average citation index. The two groups showed similar scores for these macroscopic indices. This result suggests that widely used quantitative measures of the productivity do not give meaningful measure for the actual contribution of a research group to science.

*Department of Systems Sciences, College of Arts and Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

1 Introduction

In the scientific community, quantitative measures of the productivity have become widely used in the last half century. Both scientists and policymakers use the quantitative measures like the number of papers published in peer-reviewed journals, the average citation count of the published papers, and number of papers with high citation count, or number of papers published in journals with high average citation count (impact factor) to evaluate the activities of individual scientists, departments, institutes, or nations (see, *e.g.*, [13].

The fact that these measures are well-defined and objective does not necessarily guarantee their practical significance. For example, it is clear that not all papers are equal. A paper by somebody called Junichiro Makino is counted as one paper, and the paper on the special relativity by Einstein is also counted as one. However, it is beyond any question that the latter is more important.

The citation analysis, first introduced by Garfield[5], was a major step towards constructing a more reliable measure. If a paper is cited by many other papers, it is likely to be more important than less-cited papers. In fact, large number of papers are never cited, and it is probably fair to say that these papers add rather little to the archive of the scientific knowledge.

Whether or not we can use the citation counts as a "linear" measure of the impact of the paper is not clear. Is a paper with 100 citations really 10 times more important than another paper with 10 citations? Can we add the citations for two related papers to obtain the total impact? We do not know the answer.

Many authors criticized the use of citation count as the measure for the productivity [8, 2]. However, most of them were satisfied with raising rather theoretical issues. On the other hand, other authors claimed that the citation measures agrees well with other quantitative data, suggesting that different measures are supportive to one another (see, *e.g.*, [10, 11]).

In this article, I present one example where two "quantitative" measures contradict with each other, using the result of an extensive analysis [9]. In section 2, I'll describe the data. In section 3 I'll discuss the practical and theoretical implications of our findings.

2 Observation

In this section, I'll summarize the result of [9] which are related to the citation measures. What they did is to compare two research groups in the field of theoretical astrophysics of post-world-war-II Japan (we refer two as groups A and B). These two are ideal for a number of reasons. The leaders are of similar ages and the period they were leading their groups almost completely overlap. As will be shown later, the size of the groups is also similar. In addition, both groups initially concentrated mainly on the study of the stellar structure. So they are of similar dates, similar ages, similar sizes and similar fields.

The reputations of the two groups in the Japanese astronomy community are rather different [4]. To show some quantitative support for the reputation, I summarize the number of Japanese academic prizes two groups got. For the Nishina Prize, which is the most prestigious prize for physicists and astrophysists in Japan, Group A counts five awards while group B got none. The Astronomical Society of Japan is currently offering two prizes. Five researchers from group A (or its descendants) got the prizes while only one from group B. Finally, Members of group A got several high-prestige prizes like Japan Academy Prize and Kyoto Prize, while Group B got none.

As I stated in the introduction, the number of papers is widely used as a measure of the scientific activity of a researcher or a research unit. In some cases, corrections based on the citation count or weight of the journals (the impact factors) are used. Clearly, the correction using journals is intended as an approximation to the correction using the citation count. Therefore, we decided to use the first two measures, namely the number of papers and the citation count of each papers, as the quantitative measure of the contribution of research groups.

The result was rather surprising (well, at least for me, since I had thought the citation data would support the reputations). *The number of papers published per researcher per year turned out to be essentially the same.* Strictly speaking, Group B produced 0.96 papers/(year·person), while Group A produced only 0.88, but this is not statistically significant.

It should be noted that these production rates are pretty high. Abt [1] reported that the average number of papers per person per year is 0.56 for 30 astronomy departments of universities in the U.S in the year 1992. Only one of these 30 departments (Princeton) had the score higher than 0.85.

Even when we take into account the citation count, the situation is the same. Average citation count is, again, slightly higher for Group B.

It is sometimes argued that the number of highly cited papers better reflect the quality of a research unit than the average citation count. The fraction of highly cited papers (more than 30 times) is higher for group A (8%) than for group B (4%). Since the average count is similar, this difference is compensated by the higher fraction of papers with very small citation count in group A.

Here we see some difference. However, this small difference in the count of highly cited papers does not seem satisfactory to explain the large difference in the reputations of the two groups,

3 Discussion

3.1 Which is wrong?

We assessed the validity of the now widely used "scientometric" measures, in particular the number of papers and citations. We found that they did not show much variation between two research groups we analyzed, though one group has established significantly higher reputation within the research community than the other.

This result, of course, can be interpreted in two entirely different ways: (a) the quantitative measures give the correct evaluation of the two groups and the reputation a research group within the research community has little to do with the actual contribution of the group, or (b) the quantitative measures, at least in the way we measured, have little to do with the actual contribution. In addition, there is the third possibility: (c) both are wrong.

For the two research groups we measured, I believe the hypothesis (a) is quite unlikely. One could argue that scientometric measures seems to be objective while the reputation within the research community is more like a subjective reconstruction of the history of the community. However, since we do not really know the relation between what we want to know (the "contribution" of a research group) and measurable quantities such as the number of papers, there is no reason to assume that objective measure is better.

Whether reputations have anything to do with the real contributions is a separate question. We could argue that the reputation obviously play various important roles in the dynamics of science. It reflects scientists' view of their colleagues, which is formed through various communications processes, peer-reviewed published papers being just one of them. There are a number of other channels through which scientists communicate. The most important one is personal communications with colleagues in the same institute, as vividly documented in many "anthropological" studies like [6]. Presentations in conferences, meetings, seminars, and informal discussions in meetings also play important roles. Reputations are constructed through these complex communications.

We are not arguing that reputations correctly reflect the contributions. They certainly do not. However, it is difficult to imagine that reputations have nothing to do with the actual contribution. This difficulty for me might be partly because I, being a Japanese astrophysist, share the subjective view of the community.

The fundamental problem here is that it is not clear what we mean by the word contribution and its measurement. In the following, I try to address the question of what we want to measure and why.

3.2 Practical Implications

Since we have measured just two groups, it would certainly be an over-generalization if I try to deduce any general conclusion from our result. Even so, our result has rather important implication: a naive use of "scientometric" data such as the number of papers and citation count for the evaluation of the contribution of a research group can lead to a conclusion quite different from the reputation of the group.

It is true that our measurement was rather primitive. For example, we sampled only a limited range of time and some of the most well known works of group A were published before the period we studied. In addition, we excluded the contribution of the researchers who found job elsewhere and left the group. It is quite likely that the reputation of a group or "contribution" is related not only to the scientific outcomes such as the papers, but also to more indirect products such as researchers grown up in the group. On the other hand, these rather primitive measures are actually very widely used to measure the activities of research units of all levels of aggregation, from individuals to nations.

Should we not use the paper count or citation count to assess the contribution? Or should we try to "improve" the measures further?

The improvement in the measuring method of course helps. We [9] re-analyzed the same citation data we used in section 2 in various ways, and some of them, in particular the distribution of highly cited papers over time and subfields of astrophysics, showed marked difference between two groups.

What kind of measurement we should do depends on the purpose for which we use the result of measurement. The evaluation of the contribution is necessary in various levels of the scientific enterprize. Faculty members of a university might use the citation data to select one from candidates for new position or promotion. A government might use the citation data of various national research institutes to compare them with their foreign counterparts, or to compare different institutes. The result of such comparison might then be used to make decisions on the scientific policy.

The necessary data would be different depending on the purpose, and therefore it is possible to improve the measures so that they serve the purpose we use the data. Such improvements are certainly necessary.

3.3 Theoretical Implications

I believe no scientometrician accept the view that the goal of the scientometrics is just to provide data and improve the measurement. There would be a variety of reasons not to accept that view, but the most important argument is that it is simply impossible to provide just the data without interpretation.

Leydesdorff [7] argued that the theory of citations is necessary if we want to utilize citation measures as tools. Here I'd like to argue that, if we are to measure the contribution, we need the theory of the scientific contribution, which inevitably implies *the general theory of how science proceeds.*

I'm not arguing here that such a grand theory is possible. However, as everyone knows, any measurement is theory-laden, and therefore we are applying some forms of theories, explicitly or implicitly, when we use scientometric measures. The problem here is that we do not know what theory we are applying. This is the most certain way of making mistakes, and I believe that I was able to demonstrate that it's pretty easy to make such mistakes.

As we described in section 3.2, it is certainly possible to construct more sophisticated measures so that, for example, measured quantities agree well with the reputation. However, we do not know whether the measured quantities have anything to do with the contribution.

Consider the following statement: The purpose for the faculty of a department to use the scientometric data is to obtain the objective and quantitative evaluation of the scientific contributions of the candidates for a vacant position. This statement looks perfectly reasonable. However, then, why they need objective measures? What they really want to know? In order to answer this question, we need to know how the candidate selection process operates. Candidate selection for a faculty position is just one example. Decision making is necessary at all levels of scientific communication system.

Thus, just to serve the users of scientometric data, scientometricians need a theory, or a model, which describes the interaction between the collected data and the user. However, I presume Leydesdorff's [7] claim is that such a theory is not there yet.

Without a theory as guiding principle, what we could do is to investigate various possibilities. I sincerely hope to learn more from other contributors to this special issue.

Let me end this contribution by pointing out one potential problem. The availability of "objective" measures does affect the way decision making works. The simple fact that some measures are accepted as objective is the reason why they are used. If a certain measure has become widely accepted, the researchers tend to think the goal is to achieve high score in that measure, and reputations would become linked to that measure. Thus, we see the classic problem of observation affecting the observant, and this is part of the reason why we need the theory of citation now. Primitive use of scientometric data now is now affecting the behavior of actors in various levels.

However, once established, the theory for citation, or scientific activity in general, would certainly affect the way individual scientists act. Thus, we will have to modify the theory as soon as it is established and accepted. It will be an interesting question whether there will be one "fixed point" for the theory and the behavior of the scientists.

Acknowledgments

The data used in this paper is the result of collaboration with Yuko Fujigaki, Yoshiyuki Imai, Toshikazu Ebisuzaki and Munetaka Ueno. I would like to thank to all of them, but in particular to Yuko Fujigaki, since without her strong leadership the analysis would have never take the form of [9]. I'd also thank to Yoshiyuki Imai, who did all the hard work of data collection and reduction.

References

[1] H. A. Abt,Institutional productivities, Publications of the Astronomical Society of the Pacific, 105 (1993) 794-798.

[2] H. M. Collins, The Possibilities of Science Policy, Social Studies of Science, 15 (1985) 554-558.

[3] Y. Fujigaki, Filling the gap between discussion on science and scientist's everyday activities: Applying the autopoiesis system theory to scientific knowledge, Social Science Information, 37 (1998), No.1, 5-22.

[4] J. Fukue, T. Kurokawa, The tree of Japanese Astronomers, Astronomical Herald (In Japanese), 87, (1994) 195-204.

[5] E. Garfield, Citation Indexes for Science, Science, 122 (1955) 108-111.

[6] B. Latour and S. Woolgar, Laboratory Life: The Social Construction of Scientific Facts, 2nd edition, Princeton University Press, Princeton, 1986.

[7] L. Leydesdorff *Theory of Citations?*, in this volume.

[8] M. H. MacRoberts and B. R. MacRoberts, Quantitative measures of communication in science: A study of the formal level, Social Studies of Science, 16 (1986) 151-172.

[9] J. Makino, Y. Fujigaki, and Y. Imai, Productivity of research groups — relation between citation analysis and reputation within research community (In Japanese), Japan Journal of Science, Technology and Society, 7 (1997) 85-100.

[10] B. R. Martin and J. Irvine, Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy, Research Policy, 12 (1983) 61-90.

[11] B. R. Martin and J. Irvine, Evaluating the Evaluators: A Reply to Our Critics, Social Studies of Science, 15 (1985) 558-575.

[12] D. J. de Solla Price, Networks of Scientific Papers, Science, 149 (1955) 510-515.

[13] D. Swinbanks, R. Nathan, R. Trendl, Western research assessment meets Asian cultures, Nature, 389 (1997 113-117.