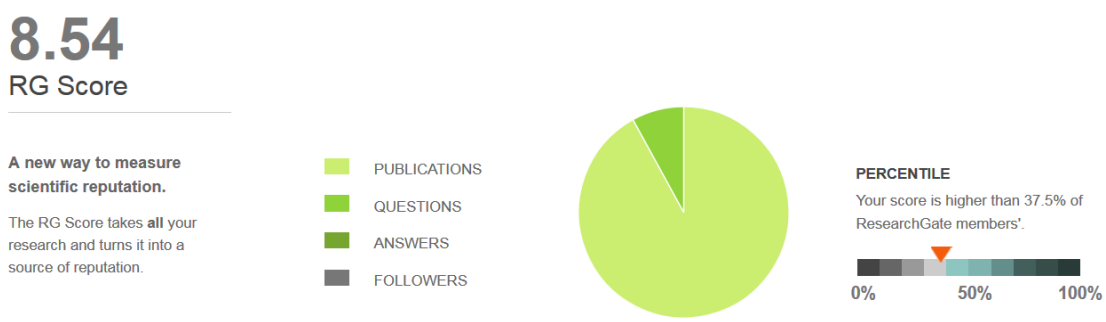


ResearchGate Score: un buen ejemplo del uso incorrecto de un indicador métrico

De acuerdo con ResearchGate, una red social académica prominente, el ResearchGate Score (RG) es: "una nueva manera de medir la reputación científica". Pero, RG es una medida oscura u opaca, porque se desconoce tanto sus componentes individuales como su peso en el valor global del indicador. Carece por ello, de transparencia y reproducibilidad, una de las recomendaciones realizadas por el Manifiesto de Leiden para el uso de indicadores métricos en la evaluación del impacto de la investigación y la actividad científica. Con el propósito de descubrir sus componentes individuales y su influencia en el indicador a nivel global, Peter Kraker, Katie Jordin y Elisabeth Lex, a partir de un proceso de ingeniería inversa, hallaron un peso significativo de los "puntos de impacto" - una métrica similar al factor de impacto de la revista, ampliamente desacreditado. La transparencia estructural y funcional de una métrica es la clave para evaluar su comportamiento en un contexto y descubrir sus sesgos, que son inherentes a esta clase de indicadores.

According to ResearchGate, the academic social networking site, their RG Score is "a new way to measure your scientific reputation". With such high aims, Peter Kraker, Katy Jordan and Elisabeth Lex take a closer look at the opaque metric. By reverse engineering the score, they find that a significant weight is linked to 'impact points' – a similar metric to the widely discredited journal impact factor. Transparency in metrics is the only way scholarly measures can be put into context and the only way biases – which are inherent in all socially created metrics – can be uncovered.

Launched in 2008, ResearchGate was one of the earlier academic social networks on the Web. The platform revolves around research papers, a question and

answering system, and a job board. Researchers are able to create a profile that showcases their publication record and their academic expertise. Other users are then able to follow these profiles and are notified of any updates. In recent years, ResearchGate has become more aggressive in marketing its platform via e-mail. In default settings, ResearchGate sends between 4 and 10 emails per week, depending on the activity in your network. The high numbers of messages prove to be very successful for ResearchGate: according to a study by Nature from 2014, ResearchGate is the most well-known social network among researchers; 35% of surveyed researchers say that they signed up for ResearchGate "because they received an e-mail". It may come as no surprise that this strategy has since been adopted by many of ResearchGate's competitors, including Academia.edu and Mendeley.

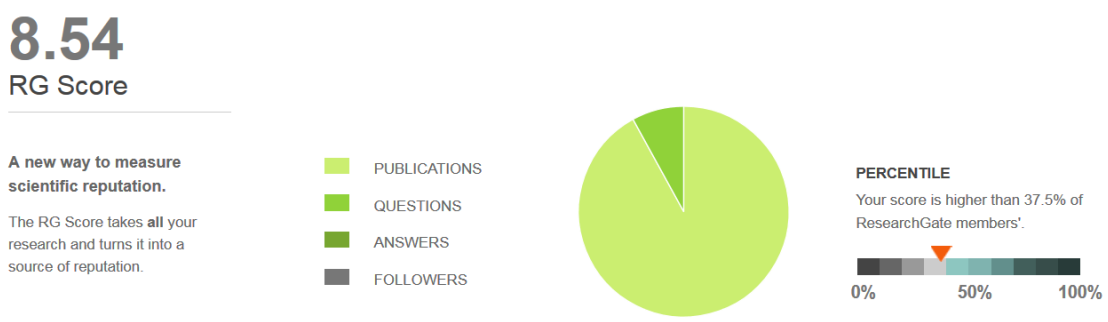
One of the focal points in ResearchGate's e-mails is a researcher's latest ResearchGate Score (RG Score). Updated weekly, the RG Score is a single number that is attached to a researcher's profile. According to ResearchGate, the score includes the research outcomes that you share on the platform, your interactions with other members, and the reputation of your peers (i.e., it takes into consideration publications, questions, answers, followers). The RG Score is displayed on every profile alongside the basic information about a researcher. ResearchGate has received substantial financial backing from venture capitalists and Bill Gates, but it is not clear how the platform will generate revenue; the possibility of the score being linked to financial value warrants further exploration and critical assessment.

The results of our evaluation of the RG Score were rather discouraging: while there are some innovative ideas in the way ResearchGate approached the measure, we also found that the RG Score ignores a number of fundamental bibliometric guidelines and that ResearchGate makes basic mistakes in the way the score is calculated. We deem these shortcomings to be so problematic that the RG Score should not be considered as a measure of scientific reputation in its current form. The measure comes with bold statements: according to the site, the RG Score is "a new way to measure your scientific reputation"; it was designed to "help you measure and leverage you're standing within the scientific community". With such high aims, it seemed to be appropriate to take a closer look at the RG Score and to

evaluate its capability as a measure of scientific reputation. We based our evaluation on well-established bibliometric guidelines for research metrics, and an empirical analysis of the score. The results were presented at a recent workshop on Analyzing and Quantifying Scholarly Communication on the Web (ASCW'15 – introductory post here) in a position paper and its discussion.

In-transparency and irreproducibility over time

One of the most apparent issues of the RG Score is that it is in-transparent. ResearchGate does present its users with a breakdown of the individual parts of the score, i.e., publications, questions, answers, followers (also shown as a pie-chart), and to what extent these parts contribute to your score. Unfortunately, that is not enough information to reproduce one's own score. For that you would need to know the exact measures being used as well as the algorithm used for calculating the score. These elements are, however, unknown.

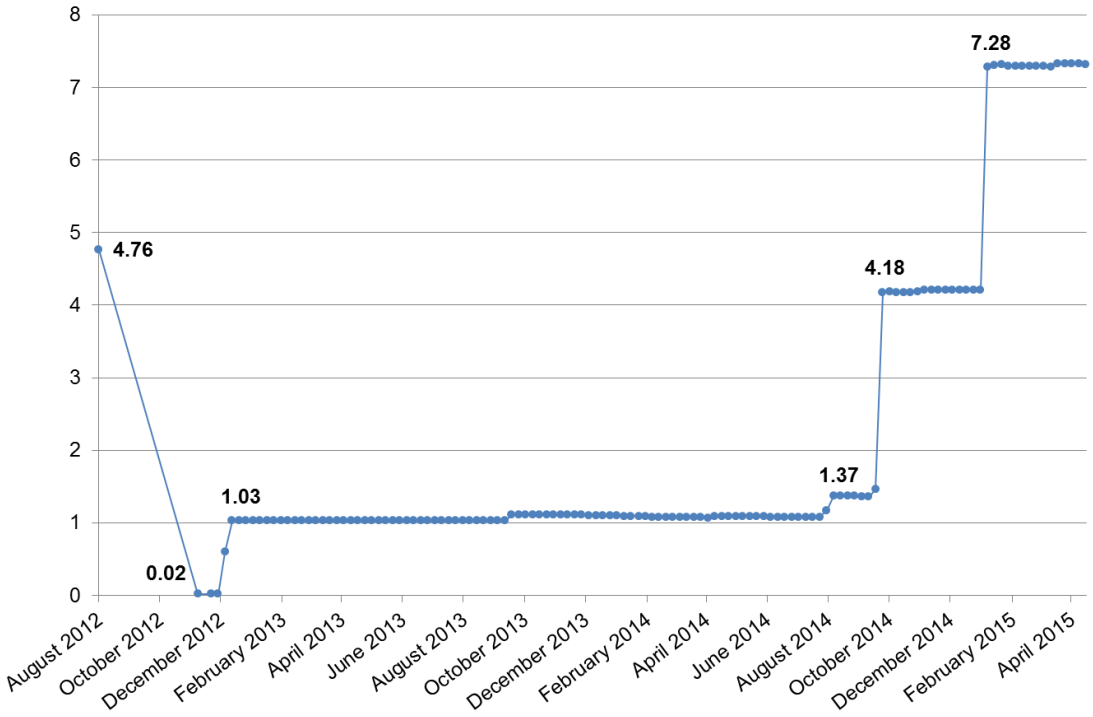


ResearchGate thus creates a sort of black-box evaluation machine that keeps researchers guessing, which actions are taken into account when their reputation is measured. This is exemplified by the many questions in ResearchGate's own question and answering system pertaining to the exact calculation of the RG Score. There is a prevalent view in the bibliometric community that transparency and openness are important features of any metric. One of the principles of the Leiden Manifesto states for example: "Keep data collection and analytical processes open, transparent and simple", and it continues: "Recent commercial entrants should be held to the same standards; no one should accept a black-box evaluation machine."

Transparency is the only way measures can be put into context and the only way biases – which are inherent in all socially created metrics – can be uncovered. Furthermore, in-transparency makes it very hard for outsiders to detect gaming of the system. In ResearchGate for example, contributions of others (i.e., questions and answers) can be anonymously down voted. Anonymous down voting has been criticized in the past as it often happens without explanation. Therefore, online networks such as Reddit have started to moderate down votes.

Further muddying the water, the algorithm used to calculate the RG Score is changing over time. That in itself is not necessarily a bad thing. The Leiden Manifesto states that metrics should be regularly scrutinized and updated, if needed. Also, ResearchGate does not hide the fact that it modifies its algorithm and the data sources being considered along the way. The problem with the way that ResearchGate handles this process is that it is not transparent and that there is no way to reconstruct it. This makes it impossible to compare the RG Score over time, further limiting its usefulness.

As an example, we have plotted Peter's RG Score from August 2012 to April 2015. Between August 2012, when the score was introduced, and November 2012 his score fell from an initial 4.76 in August 2012 to 0.02. It then gradually increased to 1.03 in December 2012 where it stayed until September 2013. It should be noted that Peter's behavior on the platform has been relatively stable over this timeframe. He has not removed pieces of research from the platform or un-followed other researchers. So what happened during that timeframe? The most plausible explanation is that ResearchGate adjusted the algorithm – but without any hint as to why and how that has happened, it leaves the researcher guessing. In the Leiden Manifesto, there is one firm principle against this practice: "Allow those evaluated to verify data and analysis".



An attempt at reproducing the ResearchGate Score

In order to learn more about the composition of the RG Score, we tried to reverse engineer the score. There are several pieces of profile information which could potentially contribute to the score; at the time of the analysis, these included 'impact points' (calculated using impact factors of the journals an individual has published in), 'downloads', 'views', 'questions', 'answers', 'followers' and 'following'. Looking at the pie charts of RG Score breakdowns, academics that have a RG Score on their profile can therefore be thought of as including several subgroups:

- Those whose score is based only on their publications.
- Scores based on question and answer activity.
- Scores based on followers and following.
- Scores based on a combination of any of the three.

For our initial analysis, we focused on the first group: we constructed a small sample of academics (30), which have a RG Score and only a single publication on their profile. This revealed a strong correlation between impact points (which, for a single paper academic, is simply the Journal Impact Factor (JIF) of that one papers' journal). Interestingly, the correlation is not linear but logarithmic. Why ResearchGate chooses to transform the 'impact points' in this way is not clear. Using the natural log of impact points will have the effect of diminishing returns for those with the highest impact points, so it could be speculated that the natural log is used to encourage less experienced academics.

We then expanded the sample to include examples from two further groups of academics: 30 academics who have a RG Score and multiple publications; and a further 30 were added who have a RG Score, multiple publications, and have posted at least one question and answer. Multiple regression analysis indicated that RG Score was significantly predicted by a combination of number of views, natural logs of impact points, answers posted and number of publications. Impact points proved to be very relevant; for this exploratory sample at least, impact points accounted for a large proportion of the variation in the data (68%).

Incorporating the journal impact factor to evaluate individual researchers

Our analysis shows that the RG Score incorporates the Journal Impact Factor to evaluate individual researchers. The JIF, however, was not introduced as a measure to evaluate individuals, but as a measure to guide libraries' purchasing decisions of journals. Over the years, it has also been used for evaluating individual

researchers. But there are many good reasons why this is a bad practice. For one, the distribution of citations within a journal is highly skewed; one study found that articles in the most cited half of articles in a journal were cited 10 times more often than articles in the least cited half. As the JIF is based on the mean number of citations, a single paper with a high number of citations can therefore considerably skew the metric.

In addition, the correlation between JIF and individual citations to articles has been steadily decreasing since the 1990s, meaning that it says less and less about individual papers. Furthermore, the JIF is only available for journals; therefore it cannot be used to evaluate fields that favor other forms of communication, such as computer science (conference papers) or the humanities (books). But even in disciplines that communicate in journals, there is a high variation in the average number of citations which is not accounted for in the JIF. As a result, the JIF is rather problematic when evaluating journals; when it comes to single contributions it is even more questionable.

There is a wide consensus among researchers on this issue: the San Francisco Declaration of Research Assessment (DORA) that discourages the use of the Journal Impact Factor for the assessment of individual researchers has garnered more than 12,300 signees at the time of writing. It seems puzzling that a score that claims to be "a new way to measure your scientific reputation" would go down that way.

Final words

There are a number of interesting ideas in the RG Score: including research outputs other than papers (e.g. data, slides) is definitely a step into the right direction, and the idea of considering interactions when thinking about academic reputation has some merit. However, there is a mismatch between the goal of the RG Score and use of the site in practice. Evidence suggests that academics that use ResearchGate tend to view it as an online business card or curriculum vitae, rather than a site for active interaction with others. Furthermore, the score misses any activities that take place outside of ResearchGate; for example, Twitter is more frequently the site for actively discussing research.

The extensive use of the RG Score in marketing e-mails suggests that it was meant to be a marketing tool that drives more traffic to the site. While it may have succeeded in this department, we found several critical issues with the RG Score, which need to be addressed before it can be seen as a serious metric.

ResearchGate seems to have reacted to the criticisms surrounding the RG Score. In September, they introduced a new metric named "Reads". "Reads", which is defined as the sum of views and downloads of a researcher's work, is now the main focus of their e-mails and the metric is prominently displayed in a researcher's profile. At the same time, ResearchGate has decided to keep the score, albeit in a smaller role. It is still displayed in every profile and it is also used as additional information in many of the site's features, e.g. recommendations.

Finally, it should be pointed out that the RG Score is not the only bad metric out there. With metrics becoming ubiquitous in research assessment, as evidenced in the recent HEFCE report "The Metric Tide", we are poised to see the formulation of many more. With these developments in mind, it becomes even more important for us bibliometric researchers to inform our stakeholders (such as funding agencies and university administrators) about the problems with individual metrics. So if you have any concerns with a certain metric, don't hesitate to share it with us, write about it – or even nominate it for the Bad Metric prize.

Fuente: Kraker P, Jordan K, Let E. The ResearchGate Score: a good example of a bad metric. The London School of Economics and Political Science; 2015. Disponible en: http://blogs.lse.ac.uk/impactofsocialsciences/2015/12/09/the-researchgate-score-a-good-example-of-a-bad-metric/