ÖKONOMENSTIMME

Rankings: Auf dem falschen Dampfer

http://oekonomenstimme.org/a/1217/

-
  Bruno S. Frey
-
  Margit Osterloh

Vor kurzem haben die Verantwortlichen des Handelsblatt-Rankings auf der Ökonomenstimme ihre neue Methodik vorgestellt. Dieser Beitrag kritisiert die dabei fehlende Auseinandersetzung mit den grundlegenden Problemen solcher Rankings.

Wie wichtig Publikationen und Zitierungen, Rankings und Impact Faktoren sind, weiss heute jede und jeder im Wissenschaftsbetrieb. Es gibt eine umfangreiche Diskussion zur Messung der wissenschaftlichen Produktivität (vgl. z.B. Starbuck, 2005; Oswald, 2007; Welpe et al., 2015 (Hrsg.)). Auch die *Ökonomenstimme* hat sich intensiv daran beteiligt (z.B. Frey & Osterloh, 2012; Wohlrabe, 2015; Iselin et al., 2013). Der neueste Beitrag von Gygli et al. (2017) "**Handelsblatt – Ranking 2.0: Wissenschaftlicher, flexibler, transparenter (/artikel/2017/06/handelsblatt-ranking-20-wissenschaftlicher-flexibler-transparenter/)**" beschäftigt sich mit der "richtigen" Gewichtung einzelner Zeitschriften im kommenden Handelsblatt-Ranking 2017. Weder von den Autoren noch von den Kommentierenden werden jedoch die grundsätzlichen Probleme von solchermassen ermittelten wissenschaftlichen Rankings auch nur erwähnt.

Die Messung der wissenschaftlichen Bedeutung einzelner Beiträge durch die Einstufung des Journals, in dem ein Beitrag veröffentlich wurde, ist mit fundamentalen Schwierigkeiten konfrontiert. Sie dürfen in einer wissenschaftlichen Debatte nicht einfach zur Seite geschoben werden. Das verfeinerte Messverfahren, das nun vorgeschlagen wird, ändert daran nichts:

(1) Die Bewertung einer Zeitschrift – durch die neue Einstufung von A+ bis F oder z.B. durch ihren Impact Factor – sagt nichts darüber aus, wie stark ein *einzelner* Artikel in der Wissenschaft beachtet wurde, der im betreffenden Journal veröffentlicht wurde. Die Zitierungen von Beiträgen in akademischen Zeitschriften sind äusserst schief

verteilt; sie sind durch einen "long tail" charakterisiert: Einige wenige Aufsätze werden oft zitiert, die meisten Artikel – auch in High-Impact-Journals – werden deutlich weniger häufig zitiert als die Einstufung oder der Impact Factor suggeriert. 75 Prozent der in *Nature* 2013/2014 veröffentlichten Beiträge wurden weniger als 38 mal – dem Impact-Faktor von *Nature* –zitiert, wohingegen das erfolgreichste Papier 905 Zitierungen erreichte. Bei *Science* wurden 76 Prozent der Aufsätze weniger zitiert als der Impact Factor von 35, während der erfolgreichste Artikel 694 mal zitiert wurde (Callaway, 2016).

In ökonomischen Journals sieht es ähnlich aus. Oswald (2007) hat in einer Studie für zwei A+ (*American Economic Review; Econometrica*) und einige weniger gut eingestufte Zeitschriften gezeigt, dass nach 25 Jahren in *AER* zwei (Walker; Conybeare) und in *Econometrica* drei (Skott; Hughes et al.; Shea) veröffentlichte Beiträge kein einziges Mal zitiert wurden. Hingegen wurde im gleichen Zeitraum eine Arbeit von Sandmo im weniger gut eingestuften *Journal of Public Economics* 88 mal zitiert, im *Oxford Bulletin of Economics and Statistics* ein Papier von Sen 50 mal. Selbstverständlich gibt es auch in den schlechter eingestuften Journalen einige Beiträge, die in den nachfolgenden 25 Jahren nie zitiert wurden. Es ist jedoch völlig verfehlt, die Aufsätze von Walker, Conybeare, Scott, Hughes und Shea mit dem Impact Factor der *AER* und *Econometrica* zu bewerten bzw. zu "belohnen": sie haben die wissenschaftliche Diskussion gemäss dem Kriterium "Zitierungshäufigkeit" nicht beeinflusst.

Bei schiefen Verteilungen den Durchschnitt der Zitationen als Massstab für die Qualität einer einzelnen, in diesem Journal veröffentlichten Arbeit zu nehmen ist – gelinde gesagt – skurril (Baum, 2010). Die *International Mathematical Union IMU* (2008: 14) hat ein solches Vorgehen deshalb als "breathtakingly naïve" bezeichnet. Der Chefredaktor von *Science,* Bruce Alberts, stellt in einem im Mai 2013 publizierten Leitartikel unmissverständlich fest: "Such metrics ... block innovation" (Alberts, 2013: 787).

(2) Ebenfalls wohlbekannt ist, dass jede relevante Messung zu Reaktivität und damit zu signifikanten Verzerrungen führt. Es schieben sich diejenigen an die Spitze, die am besten mit dem System spielen und zu ihren Gunsten weiter pervertieren können ("gaming the system"). Das sind in aller Regel nicht diejenigen, die am Inhalt der Forschung interessiert sind. Stattdessen werden "toxische feedback-loops" (O´Neil, 2016: 11) in Gang gesetzt. Das Gleiche gilt auf aggregierter Ebene. Die King Saud Universität hat ihr Ranking um *mehrere* hundert (!)Positionen verbessert, indem sie nach solchen Kriterien hoch eingestuften Forschern ein Honorar um die 50,000 Euro anbot, damit sie als Gegenleistung die King Saud Universität als Forschungsstätte angaben (Bhattacharjee, 2011). In China existiert ein voll entwickelter Markt für die Autorenschaft in wichtigen wissenschaftlichen Zeitschriften (Hvistendahl, 2013).

(3) Wenn es – wie beabsichtigt – schwieriger wird, Publikationen in Top-Journals durch Publikationen in geringer bewerteten Journals zu substituieren (Gygli et al., 2017: 2), dann wird dies zu einer noch stärkeren Hierarchisierung innerhalb der Forschenden führen. Zu erwarten ist dann eine noch stärkere Konzentration der Forschungsressourcen auf wenige Wissenschaftler. Das führt nicht nur zum abnehmendem Grenznutzen der wissenschaftlichen Ressourcen (Jansen et al., 2007), sondern unterdrückt auch die für die Wissenschaft essentielle Diversität, welche für neue Ideen unabdingbar ist. Dies zeigen eindrücklich Katz und Matter (2017) in einer empirischen Untersuchung für Biomedizin im amerikanischen National Institute of Health NIH.

(4) Am wichtigsten ist jedoch die durch Rankings bewirkte inzestuöse Selbstbespiegelung. Es wird nicht mehr danach gefragt, ob ein Beitrag inhaltlich von Bedeutung ist. Statt dessen bewirken Rankings eine "soziale Platzanweisung" (Mau, 2017: 73), die nicht über den wissenschaftlichen Gehalt *informieren*, sondern die Urteilsbildung *formieren*, d.h. bestimmen, nach welchen Kriterien Forschung bewertet werden soll – und das nach Kriterien, die "breathtakingly naïve" sind (siehe oben). Sie setzten zirkuläre Prioritäten innerhalb einer

geschlossenen Wissenschaftsgemeinde: die Messungen bestätigen sich selbst.

Die grundsätzlichen Probleme der Messung der Forschungs-Qualität anhand der Einstufung von Journals, in der die Forschung veröffentlicht wurde, haben viele wichtige Institutionen dazu veranlasst, solche Verfahren zu verurteilen. Besonders wichtig ist die *San Francisco Declaration on Research Assessment (DORA),* wonach der Impact Factor nicht dazu verwendet werden darf, die wissenschaftliche Qualität eines einzelnen Forschungsbeitrags oder Entscheidungen über Anstellung und Beförderung, sowie Vergabe von Forschungsgeldern zu benützen. DORA wurde von führenden Wissenschaftsgremien unterschrieben, so etwa von der *American Association for the Advancement of Science,* der *European Mathematical Society*, dem *Schweizerischen Nationalfonds, der British Library* oder *Nature Research.*

Im von vielen Ökonomen als Vorbild angesehenen amerikanischen Wissenschaftsbetrieb bezeichnen sich viele Spitzen-Universitäten heute als "reading schools"– das heisst, dass bei Berufungen eingereichte Papiere *sorgfältig gelesen* werden und nicht auf Indikatoren und Beförderungen abgestellt wird.

Was folgt daraus?

Erstens müssen die Argumente immer wieder vorgebracht werden, damit die Einsicht in die verzerrende Messung wissenschaftlicher Aktivität mehr Forschenden deutlich wird.

Zweitens muss hervorgehoben werden, dass die Beurteilung von Forschung anhand des Ranges von Journals, in denen die Forschung veröffentlicht ist, an vielen Spitzen-Universitäten als überholt gilt.

Drittens können einzelne im Scheinwerfer stehende Wissenschaftler die Zwangsjacke von Top-Journals kritisieren, so z.B. Randy Schekman, Nobelpreisträger für Medizin/Physiologie.[1] (#link1)

Viertens können Forschungsförderungsorganisationen Schritte zur Wiederherstellung von mehr Diversität in der Forschung unternehmen, welche durch das Handelsblatt-Ranking beeinträchtigt wird. Einen bemerkenswerten Schritt hat jüngst die VW-Stiftung getan: Nach einer sorgfältigen Vorauswahl werden per Los Bewilligungen von Anträgen gezogen um Anträgen eine Chance zu geben, die sonst leicht übersehen werden.[2] (#link2)

Salopp ausgedrückt: Die neue Initiative des Handelsblattes fährt den Ranking-Dampfer in die falsche Richtung und ist erst noch verspätet. Kümmern wir uns doch lieber um inhaltliche Diskurse – was die *Ökonomenstimme* vorbildlich tut.

Literatur

Alberts, B. (2013). Editorial: Impact Factor Distortions. Science, 340(6134): 787.

Baum, J. A. C. (2010). Free-Riding on Power Laws: questioning the validity of the Impact Factor as a measure of research quality in organization studies. Organization, 18(4): 449-466.

Bhattacharjee, Y. (2011). Saudi Universities offer Cash in Exchange for Academic Prestige. Science, 334(6061): 1344-1345.

Callaway, E. (2016). Beat it, impact factor! Publishing elite turns against controversial metric. Nature, 535: 210-211.

DORA (http://www.ascb.org/files/SFDeclarationFINAL.pdf) (San Francisco Declaration on Research Assessment) (2012). Abgerufen am 25. Juni 2017

Frey, B. S. & Osterloh, M. (2012). **Rankings: Unbeabsichtigte Nebenwirkungen und Alternativen (/artikel/2012/02/rankings-unbeabsichtigte-nebenwirkungen-und-alternativen/)**. Ökonomenstimme, 17. Februar 2012.

Frey, B. S. & Osterloh, M. (2015). **Zufallsauswahl in der Wissenschaft? (/artikel/2015/12/zufallsauswahl-in-der-wissenschaft/)** Ökonomenstimme, 16. Dezember 2015.

Gygli, S., Haucap, J., Ruffner, J. Sturm, J.-E. & Südekum, J. (2017). **Handelsblatt-Ranking 2.0: Wissenschaftlicher, flexibler, transparenter (/artikel/2017/06/handelsblatt-ranking-20-wissenschaftlicher-flexibler-transparenter/)**. Ökonomenstimme, 20. Juni 2017.

Hvistendahl, M. (2013). China's Publication Bazaar. Science, 342(6162): 1035-1039.

International Mathematical Union IMU (2008). Citation Statistics. A report. Corrected version, 16/12/08.

Iselin, D., Schläpfer, J. & Vogel, S. (2013). **Die publikationsstärksten Volkswirte: das Handelsblatt-VWL-Ranking 2013 (/artikel/2013/09/die-publikationsstaerksten-volkswirte-das-handelsblatt-vwl-ranking-2013/)**. Ökonomenstimme, 6. September 2013.

Jansen, D., Wald, A., Frenke, K., Schmoch, U. & Schubert, T. (2007). Drittmittel als Performanzindikator der wissenschaftlichen Forschung. Zum Einfluss von Rahmenbedingungen auf Forschungsleistung. Kölner Zeitschrift für Soziologie und Sozialpsychologie, 59: 125-149.

Katz, Y. & Matter, U. (2017). On the Biomedical Elite. Inequality and Stasis in Scientific Knowledge Production. Unpubl. Manuscript, Harvard Medical School und Faculty of Economics and Business, University of Basel.

Mau, S. (2017). Das metrische Wir. Über die Quantifizierung des Sozialen. Berlin: Suhrkamp.

O`Neil, C. (2016). Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy. New York: Allan Lane.

Oswald, A. J. (2007). An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision Makers. Economica, 74: 21-31.

Starbuck, W. H. (2005). How much better are the most prestigious journals? The statistics of academic publication. Organization Science, 16: 180-200.

The Guardian (2013). Nobel winner declares boycott of top science journals. Randy Shechtman will no longer send papers to Nature, Cell and Science as they distort scientific protest. Erschienen am 9.12.2013.

Welpe, J., Wollersheim, J., Ringelhan, S. & Osterloh, M. (Hrsg) (2015). Incentives and Performance - Governance of Research Organization. Heidelberg: Springer Verlag.

Wohlrabe, K. (2015). **Einige Anmerkungen zum FAZ-Ökonomenranking 2014 (/artikel/2015/06/einige-anmerkungen-zum-faz-oekonomenranking-2014/)**. Ökonommenstimme, 17. Juni 2015.

1 (#f-link1) in: The Guardian vom 09.12.2013
2 (#f-link2) Siehe **Volkswagenstiftung (https://www.volkswagenstiftung.de/experiment.html)** (https://www.volkswagenstiftung.de/experiment.html). Siehe auch Frey & Osterloh 2015 "Zufallsauswahl in der Wissenschaft" in der Ökonomenstimme.

©KOF ETH Zürich, 27. Jun. 2017

Links

a http://www.ascb.org/files/SFDeclarationFINAL.pdf